Numéro d'ordre : **370**

UNIVERSITÉ BLAISE PASCAL
(UFR Sciences et Technologies)

Laboratoire de Mathématiques – UMR 6620

Habilitation à Diriger des Recherches

présentée par

Sergueï Dachian

Tome II – Textes des Articles

Sujet de l'Habilitation :

Quelques Contributions à la Statistique des Processus,
à la Théorie des Champs Aléatoires
et à la Statistique des Champs Aléatoires

Soutenue le 12 décembre 2012 devant le jury composé de :

Président	Prof. M. Hoffmann, Université Paris-Dauphine (Paris, France)
Rapporteurs	Prof. R. Fernández, Université d'Utrecht (Utrecht, Pays-Bas) Prof. R. Höpfner, Université Johannes Gutenberg (Mayence, Allemagne) Prof. E. Moulines, Télécom ParisTech (Paris, France)
Examinateurs	Prof. D. Dereudre, Université Lille 1 (Lille, France) Prof. P. Druilhet, Université Blaise Pascal (Clermont-Ferrand, France) Prof. A. Guillin, Université Blaise Pascal (Clermont-Ferrand, France) Prof. Yu.A. Kutoyants, Université du Maine (Le Mans, France)

Table des matières

Première Partie

Statistique des Processus

Statistical Inference for Stochastic Processes **6:** 1–14, 2003.

Estimation of Cusp Location by Poisson Observations

S. DACHIAN
Laboratoire de Mathématiques Appliquées, Université Blaise Pascal (Clermont-Ferrand 2), 63177 Aubière Cedex, France. e-mail: serguei.dachian@math.univ-bpclermont.fr

(Received in final form 5 May 2001; Accepted 30 December 2001)

Abstract. We consider an inhomogeneous Poisson process X on $[0, T]$. The intensity function of X is supposed to be regular on $[0, T]$ except at the point θ, in which it has a singularity (a cusp) of order p. We suppose that we know the shape of the intensity function, but not the location (given by the parameter θ) of the point of cusp. We consider the problem of estimation of this location (shift) parameter θ based on n observations of the process X. We study the maximum likelihood estimator and the Bayesian estimators. We show that these estimators are consistent, their rate of convergence is $n^{1/(2p+1)}$, they have different limit distributions, and the Bayesian estimators are asymptotically efficient.

AMS Mathematics Subject Classification (2000): 62M05.

Key words: inhomogeneous Poisson process, cusp, parameter estimation, Bayesian estimators, maximum likelihood estimator, consistency, limit distribution, convergence of moments, asymptotic efficiency.

1. Introduction

In this paper, we consider the following parameter estimation problem. Suppose we observe n realizations of a Poisson process X on some fixed interval $[0, T]$. The intensity function of the process is supposed to be of the form $S_\theta(t) = s(t - \theta)$, where $s(\cdot)$ is some known strictly positive function and $\theta \in (0, T)$ is some unknown parameter.

In the case where the problem is regular (e.g. if the function $s(\cdot)$ has a bounded derivative) the model is locally asymptotically normal (LAN), and both the maximum likelihood estimator (MLE) and the Bayesian estimators (BE) consistent, asymptotically normal and asymptotically efficient (see, e.g. [5, 6]). Here we deal with the case where the intensity function $S_\theta(\cdot)$ is regular everywhere on $[0, T]$ except at the point θ, in which it has a singularity (a cusp) of order p. More precisely, we suppose that

$$S_\theta(t) = \begin{cases} a|t - \theta|^p + \psi(t - \theta) & \text{if } t \leqslant \theta, \\ b|t - \theta|^p + \psi(t - \theta) & \text{if } t \geqslant \theta, \end{cases}$$

where $a^2 + b^2 > 0$, and $\psi(\cdot)$ is regular. If the singularity is of order higher than $1/2$, then, in spite of the non-regularity of the intensity function in θ, the Fisher

information is finite. That is why the case $p > 1/2$ can be treated as the regular one, and in the present work we study the case $0 < p < 1/2$ only. In this case we study the asymptotic behavior of the MLE and the BE and we prove that the rate of convergence of the estimators is faster than in the regular case ($n^{1/(2p+1)}$), the estimators are consistent, they have different limit distributions, and the BE are asymptotically efficient. We verify as well the convergence of moments.

A similar problem of parameter estimation is the problem of estimation of a cusp location of the density for the i.i.d. model of observations. This problem was first considered by Prakasa Rao in [7]. Further developments were carried out by Ermakov in [1] and by Ibragimov and Khasminskii [2–4]. An exhaustive exposition of the results can be found in Chapter 6 of their book [4], but one can also refer to the previous works [2, 3] of the authors. More precisely, in Chapter 6 of [4], the problem of estimation of a shift parameter θ by n independent observations of a random variable is considered in three different situations (three types of singularities). Our type of parameter estimation problem corresponds to the case where the density $f_\theta(x)$ of the observed random variable has a singularity 'of the second type' at the point θ, that is,

$$f_\theta(x) = \begin{cases} h(x-\theta)\ \exp\{a(x-\theta)\,|x-\theta|^p\} & \text{if } x \leqslant \theta, \\ h(x-\theta)\ \exp\{b(x-\theta)\,|x-\theta|^p\} & \text{if } x \geqslant \theta \end{cases}$$

with some regularity conditions on functions $h(\cdot)$, $a(\cdot)$ and $b(\cdot)$. The asymptotic behavior of the MLE and of a wide class of BE obtained for this (i.i.d.) model is similar to the one obtained here for the model of Poisson observations. Particularly, the rate of convergence of the estimators is $n^{1/(2p+1)}$, and the BE are asymptotically efficient.

Finally, let us mention here that for the study of the asymptotic behavior of the estimators we use the method of Ibragimov and Khasminskii presented by the authors in [4] (see as well [6], where this method is applied to inhomogeneous Poisson process).

2. Main Results

Suppose we observe n realizations $(X_1, \dots, X_n) = X^n$ of the Poisson process $X = \{X(t),\ 0 \leqslant t \leqslant T\}$ of intensity function $S_\theta(t) = s(t-\theta)$, where θ is some unknown parameter, $\theta \in \Theta = (\alpha, \beta) \subseteq (0, T)$, and $s(\cdot)$ is some known strictly positive function on $[-T, T]$. We suppose that the function $s(\cdot)$ can be written as $s(t) = d(t)|t|^p + \psi(t)$, where $0 < p < 1/2$,

$$d(t) = \begin{cases} a & \text{if } t < 0, \\ b & \text{if } t > 0, \end{cases}$$

$a^2 + b^2 > 0$, and the function $\psi(\cdot)$ is Hölder continuous of order higher than $p + 1/2$, that is, $|\psi(x) - \psi(y)| \leqslant L\,|x-y|^\varkappa$ for all $x, y \in [-T, T]$ with some

fixed constants $L > 0$ and $\varkappa > p + 1/2$. Our aim is to estimate the parameter θ and to study the asymptotic behavior of estimators as $n \to \infty$.

The likelihood ratio in our problem can be written (see, e.g. [6]) as

$$L(\theta, \theta_1, X^n) = \exp\left\{ \sum_{i=1}^{n} \int_0^T \ln \frac{S_\theta(t)}{S_{\theta_1}(t)} \, \mathrm{d}X_i(t) - \right.$$
$$\left. - n \int_0^T \left[\frac{S_\theta(t)}{S_{\theta_1}(t)} - 1 \right] S_{\theta_1}(t) \, \mathrm{d}t \right\},$$

where θ_1 is some fixed value of θ.

As usually, introduce the MLE $\widehat{\theta}_n$ as one of the solutions of the equation

$$L(\widehat{\theta}_n, \theta_1, X^n) = \sup_{\theta \in \Theta} L(\theta, \theta_1, X^n),$$

and the BE $\widetilde{\theta}_n$ for prior density q (supposed to be positive and continuous) and quadratic loss function as

$$\widetilde{\theta}_n = \int_\alpha^\beta \theta \, q(\theta | X^n) \, \mathrm{d}\theta,$$

where the posterior density

$$q(\theta | X^n) = L(\theta, \theta_1, X^n) \, q(\theta) \left(\int_\alpha^\beta L(\theta, \theta_1, X^n) \, q(\theta) \, \mathrm{d}\theta \right)^{-1}.$$

To describe the properties of these estimators we need to introduce the stochastic process

$$Z(u) = \exp\left\{ W^{p+1/2}(u) - \tfrac{1}{2} |u|^{2p+1} \right\}, \quad u \in \mathbb{R}.$$

Here and in the sequel $W^H(\cdot)$ denotes a standard fractional Brownian motion (FBM) of the Hurst parameter H, that is a Gaussian random process with zero mean and the covariance function

$$\mathbf{E}\big[W^H(u_1) W^H(u_2) \big] = \tfrac{1}{2} \big[|u_1|^{2H} + |u_2|^{2H} - |u_1 - u_2|^{2H} \big].$$

We introduce also the random variables ξ and ζ by the equations

$$Z(\xi) = \sup_{u \in \mathbb{R}} Z(u),$$

and

$$\zeta = \int_{-\infty}^{+\infty} u Z(u) \, \mathrm{d}u \left(\int_{-\infty}^{+\infty} Z(u) \, \mathrm{d}u \right)^{-1}.$$

Let us note here, that ξ is well defined since with probability one the process $Z(u)$ attains its maximum in a unique point (see, e.g. [1]).

Finally, we put

$$I_p(a,b)=\int_{-\infty}^{+\infty}[d(x-1)|x-1|^p-d(x)|x|^p]^2\,\mathrm{d}x,$$

and we introduce the constant

$$\gamma=\left(\frac{I_p(a,b)}{\psi(0)}\right)^{1/(2p+1)}.$$

Note that $0<I_p(a,b)<+\infty$ since $a^2+b^2>0$ and $p<1/2$, and that it has the following representation (see Section VI.4 of [4])

$$\begin{aligned}I_p(a,b) &= \frac{\Gamma(1+p)\,\Gamma(\frac{1}{2}-p)}{2^{2p}\sqrt{\pi}\,(2p+1)}\,[a^2+b^2-2ab\,\cos(\pi p)]\\ &= \mathrm{B}(p+1\,,\,p+1)\left[\frac{a^2+b^2}{\cos(\pi p)}-2ab\right].\end{aligned}$$

Now we can finally state the main results of this paper.

THEOREM 1. *Under the maid assumptions, the following lower bound on the risks of all estimators holds: for any $\theta_0\in\Theta$ we have*

$$\lim_{\delta\to 0}\lim_{n\to\infty}\inf_{\overline{\theta}_n}\sup_{|\theta-\theta_0|<\delta}\mathbf{E}_\theta(n^{1/(2p+1)}(\overline{\theta}_n-\theta))^2\geqslant\frac{\mathbf{E}\zeta^2}{\gamma^2},$$

where inf *is taken over all possible estimators $\overline{\theta}_n$ of θ.*

This theorem leads us to introduce the following definition.

DEFINITION 2. We say that the estimator $\overline{\theta}_n$ is asymptotically efficient if

$$\lim_{\delta\to 0}\lim_{n\to\infty}\sup_{|\theta-\theta_0|<\delta}\mathbf{E}_\theta(n^{1/(2p+1)}(\overline{\theta}_n-\theta))^2=\frac{\mathbf{E}\zeta^2}{\gamma^2}$$

for any $\theta_0\in\Theta$.

For the MLE we have the following theorem.

THEOREM 3. *The MLE $\widehat{\theta}_n$ has uniformly in $\theta\in\mathbf{K}$ (for any compact $\mathbf{K}\subset\Theta$) the following properties*:

– *$\widehat{\theta}_n$ is consistent, that is,*
$$\widehat{\theta}_n\xrightarrow{\mathbf{P}_\theta}\theta\,(\textit{convergence in probability}),$$
– *the limit distribution of $\widehat{\theta}_n$ is ξ/γ, that is,*
$$n^{1/(2p+1)}(\widehat{\theta}_n-\theta)\Longrightarrow\xi/\gamma\,(\textit{convergence in law}),$$
– *for any $k>0$ we have*
$$\lim_{n\to\infty}\mathbf{E}_\theta\left|n^{1/(2p+1)}(\widehat{\theta}_n-\theta)\right|^k=\frac{\mathbf{E}|\xi|^k}{\gamma^k}\,.$$

And for the BE we have the following theorem.

THEOREM 4. *The BE $\widetilde{\theta}_n$ have uniformly in $\theta \in \mathbf{K}$ (for any compact $\mathbf{K} \subset \Theta$) the following properties*:

– *$\widetilde{\theta}_n$ is consistent, that is,*

$$\widetilde{\theta}_n \xrightarrow{\mathbf{P}_\theta} \theta \,(\textit{convergence in probability}),$$

– *the limit distribution of $\widetilde{\theta}_n$ is ζ/γ, that is,*

$$n^{1/(2p+1)}(\widetilde{\theta}_n - \theta) \Longrightarrow \zeta/\gamma \ (\textit{convergence in law}),$$

– *for any $k > 0$ we have*

$$\lim_{n\to\infty} \mathbf{E}_\theta \left| n^{1/(2p+1)}(\widetilde{\theta}_n - \theta) \right|^k = \frac{\mathbf{E}|\zeta|^k}{\gamma^k},$$

and, moreover, $\widetilde{\theta}_n$ is asymptotically efficient.

To prove the above stated theorems we apply the method of Ibragimov and Khasminskii (see [4]). For this we denote $\theta_u = \theta + un^{-1/(2p+1)}$ for all $u \in U_n = (n^{1/(2p+1)}(\alpha - \theta)\,, n^{1/(2p+1)}(\beta - \theta))$, we introduce the normalized likelihood ratio process as

$$Z_n(u) = L(\theta_u, \theta, X^n), \quad u \in U_n,$$

the stochastic process $Z_\gamma(u)$ as

$$Z_\gamma(u) = Z(\gamma u) = \exp\{\gamma^{p+1/2}\, W^{p+1/2}(u) - \tfrac{1}{2}\,\gamma^{2p+1}\,|u|^{2p+1}\}, \quad u \in \mathbb{R},$$

and we establish (the proofs are in the next section) the following three lemmas.

LEMMA 5. *The finite-dimensional distributions of $Z_n(u)$ converge to the finite-dimensional distributions of $Z_\gamma(u)$ uniformly in $\theta \in \mathbf{K}$ (for any compact $\mathbf{K} \subset \Theta$).*

LEMMA 6. *For any compact $\mathbf{K} \subset \Theta$ there exists some positive constant C such that*

$$\mathbf{E}_\theta |Z_n^{1/2}(u_1) - Z_n^{1/2}(u_2)|^2 \leqslant C\,|u_1 - u_2|^{2p+1}$$

for all $u_1, u_2 \in U_n$, $\theta \in \mathbf{K}$ and $n \in \mathbb{N}$.

LEMMA 7. *For any compact $\mathbf{K} \subset \Theta$ there exists some positive constant c such that*

$$\mathbf{E}_\theta Z_n^{1/2}(u) \leqslant \exp\{-c\,|u|^{2p+1}\}$$

for all $u \in U_n$, $\theta \in \mathbf{K}$ and $n \in \mathbb{N}$.

Using these lemmas and applying Theorems 1.9.1, 1.10.1 and 1.10.2 of [4] we get Theorems 1, 3, and 4, respectively.

3. Proofs of the Lemmas

For simplicity of exposition, the proofs will be carried out in the case $a = b$. The general case proofs are similar. For convenience of notation, all throughout this section C and c denote generic positive constants which can differ from formula to formula (and even in the same formula), and we put $\nu = 1/(2p+1)$ and $\Gamma = \gamma^{p+1/2}$.

In order to prove Lemma 5 we will study the convergence of the two-dimensional (the general case can be considered similarly) distributions $(\ln Z_n(u_1), \ln Z_n(u_2))$ of the process

$$\begin{aligned}\ln Z_n(u) &= \sum_{i=1}^{n} \int_0^T \ln \frac{S_{\theta_u}(t)}{S_\theta(t)}\, \mathrm{d}X_i(t) - n \int_0^T \left[\frac{S_{\theta_u}(t)}{S_\theta(t)} - 1\right] S_\theta(t)\, \mathrm{d}t \\ &= \sum_{i=1}^{n} \int_0^T f\, \mathrm{d}X_i(t) - n \int_0^T g\, S_\theta(t)\, \mathrm{d}t,\end{aligned}$$

where we denote

$$f = f(\theta, t, u, n) = \ln \frac{S_{\theta_u}(t)}{S_\theta(t)} \quad \text{and} \quad g = g(\theta, t, u, n) = \frac{S_{\theta_u}(t)}{S_\theta(t)} - 1.$$

The characteristic function of the vector $(\ln Z_n(u_1), \ln Z_n(u_2))$ can be written as (see, e.g. Lemma 1.1 of [6])

$$\begin{aligned}\phi_n(\lambda_1, \lambda_2) &= \mathbf{E}_\theta \exp\{\mathrm{i}\lambda_1 \ln Z_n(u_1) + \mathrm{i}\lambda_2 \ln Z_n(u_2)\} \\ &= \exp\left\{n \int_0^T \left[\mathrm{e}^{\mathrm{i}\lambda_1 f_1 + \mathrm{i}\lambda_2 f_2} - 1 - \mathrm{i}\lambda_1 g_1 - \mathrm{i}\lambda_2 g_2\right] S_\theta(t)\, \mathrm{d}t\right\},\end{aligned}$$

and hence

$$\begin{aligned}\ln \phi_n(\lambda_1, \lambda_2) &= n \int_0^T [\mathrm{e}^F - 1 - F] S_\theta(t)\, \mathrm{d}t + \\ &\quad + \mathrm{i}\lambda_1 n \int_0^T [f_1 - g_1] S_\theta(t)\, \mathrm{d}t + \\ &\quad + \mathrm{i}\lambda_2 n \int_0^T [f_2 - g_2] S_\theta(t)\, \mathrm{d}t, \qquad (1)\end{aligned}$$

where we denote $f_j = f(\theta, t, u_j, n)$ and $g_j = g(\theta, t, u_j, n)$ for $j = 1, 2$, and $F = \mathrm{i}\lambda_1 f_1 + \mathrm{i}\lambda_2 f_2$.

To study this expression, let us at first establish the two following properties.

(a) For any fixed u, we have $\lim_{n\to\infty} g(\theta, t, u, n) = 0$ uniformly in $\theta \in \mathbf{K}$ and $t \in [0, T]$.

(b) We have

$$\lim_{n\to\infty} n \int_0^T g_1\, g_2\, S_\theta(t)\, \mathrm{d}t = \frac{1}{2}\,\Gamma^2[|u_1|^{2p+1} + |u_2|^{2p+1} - |u_1 - u_2|^{2p+1}],$$

and particularly (taking $u_1 = u_2 = u$)

$$\lim_{n\to\infty} n \int_0^T g^2\, S_\theta(t)\, \mathrm{d}t = \Gamma^2 |u|^{2p+1}.$$

To prove (a), let us consider separately two cases. First, if $|t - \theta| \geqslant n^{-\nu/2}$, for n sufficiently large we have

$$\begin{aligned}
|g(\theta, t, u, n)| &= \frac{|s(t - \theta_u) - s(t - \theta)|}{s(t - \theta)} \\
&\leqslant C\left|a\,|t - \theta_u|^p + \psi(t - \theta_u) - a\,|t - \theta|^p - \psi(t - \theta)\right| \\
&\leqslant C\left||t - \theta - un^{-\nu}|^p - |t - \theta|^p\right| + \\
&\quad + C\,|\psi(t - \theta - un^{-\nu}) - \psi(t - \theta)| \\
&\leqslant C\,|t - \theta|^p \left|\left|1 - \frac{u}{(t - \theta)n^\nu}\right|^p - 1\right| + C\,|un^{-\nu}|^\varkappa \\
&\leqslant C\,\frac{|u|}{|t - \theta|^{1-p}\, n^\nu} + C\,\frac{|u|^\varkappa}{n^{\nu\varkappa}} \leqslant \frac{C(u)}{n^c}.
\end{aligned}$$

Finally, if $|t - \theta| \leqslant n^{-\nu/2}$, for n sufficiently large we get similarly

$$\begin{aligned}
|g(\theta, t, u, n)| &\leqslant C\left||t - \theta - un^{-\nu}|^p - |t - \theta|^p\right| + C\,|un^{-\nu}|^\varkappa \\
&\leqslant C\,(|t - \theta| + |un^{-\nu}|)^p + C\,|t - \theta|^p + C\,\frac{|u|^\varkappa}{n^{\nu\varkappa}} \\
&\leqslant C\,(n^{-\nu/2} + |u|\,n^{-\nu})^p + C\,n^{-\nu\,p/2} + C\,\frac{|u|^\varkappa}{n^{\nu\varkappa}} \leqslant \frac{C(u)}{n^c}.
\end{aligned}$$

So, (a) is proved.

To prove (b), let us write

$$n \int_0^T g_1\, g_2\, S_\theta(t)\, \mathrm{d}t = nI_1 + nI_2 + nI_3 + nI_4$$

with

$$
\begin{aligned}
nI_1 &= n\int_0^T \frac{[a\,|t-\theta_{u_1}|^p - a\,|t-\theta|^p][a\,|t-\theta_{u_2}|^p - a\,|t-\theta|^p]}{s(t-\theta)}\,\mathrm{d}t,\\
nI_2 &= n\int_0^T \frac{[\psi(t-\theta_{u_1}) - \psi(t-\theta)][\psi(t-\theta_{u_2}) - \psi(t-\theta)]}{s(t-\theta)}\,\mathrm{d}t,\\
nI_3 &= n\int_0^T \frac{[a\,|t-\theta_{u_1}|^p - a\,|t-\theta|^p][\psi(t-\theta_{u_2}) - \psi(t-\theta)]}{s(t-\theta)}\,\mathrm{d}t,\\
nI_4 &= n\int_0^T \frac{[a\,|t-\theta_{u_2}|^p - a\,|t-\theta|^p][\psi(t-\theta_{u_1}) - \psi(t-\theta)]}{s(t-\theta)}\,\mathrm{d}t.
\end{aligned}
$$

In order to study I_1, let us fix a sequence (A_n) such that $A_n \to +\infty$ and $A_n n^{-\nu} \to 0$ and separate the integral I_1 in three parts: integral J_1 over $(0\,,\theta - A_n n^{-\nu})$, integral J_2 over $(\theta - A_n n^{-\nu}\,,\theta + A_n n^{-\nu})$ and integral J_3 over $(\theta + A_n n^{-\nu}\,,T)$.

For J_2 we get

$$
\begin{aligned}
nJ_2 &= na^2 \int_{\theta - A_n\, n^{-\nu}}^{\theta + A_n\, n^{-\nu}} \frac{[|t-\theta_{u_1}|^p - |t-\theta|^p][|t-\theta_{u_2}|^p - |t-\theta|^p]}{s(t-\theta)}\,\mathrm{d}t\\
&= na^2 \int_{-A_n\, n^{-\nu}}^{A_n\, n^{-\nu}} \frac{[|y-u_1\, n^{-\nu}|^p - |y|^p][|y-u_2\, n^{-\nu}|^p - |y|^p]}{a|y|^p + \psi(y)}\,\mathrm{d}y\\
&\simeq n\,\frac{a^2}{\psi(0)} \int_{-A_n}^{A_n} [|z-u_1|^p - |z|^p][|z-u_2|^p - |z|^p]\, n^{-\nu(2p+1)}\,\mathrm{d}z\\
&= \frac{a^2}{2\,\psi(0)} \left\{ \int_{-A_n}^{A_n} [|z-u_1|^p - |z|^p]^2\,\mathrm{d}z + \int_{-A_n}^{A_n} [|z-u_2|^p - |z|^p]^2\,\mathrm{d}z - \right.\\
&\quad \left. - \int_{-A_n}^{A_n} [|z-u_1|^p - |z-u_2|^p]^2\,\mathrm{d}z \right\},
\end{aligned}
$$

where the symbol '$\simeq$' means equality of limits, and is true since $A_n n^{-\nu} \to 0$ and the function $a|y|^p + \psi(y)$ is continuous in 0. It is easy to see that

$$
\lim_{n\to\infty} \int_{-A_n}^{A_n} [|z-u|^p - |z|^p]^2\,\mathrm{d}z = |u|^{2p+1} \int_{-\infty}^{+\infty} [|x-1|^p - |x|^p]^2\,\mathrm{d}x.
$$

Hence we have clearly

$$
\lim_{n\to\infty} nJ_2 = \tfrac{1}{2}\,\Gamma^2[|u_1|^{2p+1} + |u_2|^{2p+1} - |u_1-u_2|^{2p+1}]. \tag{2}
$$

To study J_3, let us at first note that

$$
\begin{aligned}
n\int_{\theta + A_n n^{-\nu}}^{T} [|t-\theta_u|^p - |t-\theta|^p]^2\,\mathrm{d}t &\leqslant n\int_{A_n\, n^{-\nu}}^{+\infty} [|y-un^{-\nu}|^p - |y|^p]^2\,\mathrm{d}y\\
&= \int_{A_n}^{+\infty} [|z-u|^p - |z|^p]^2\,\mathrm{d}z \to 0,
\end{aligned}
$$

since $A_n \to +\infty$ and the integral is convergent. Therefore, using Cauchy–Schwarz inequality, we easily get

$$\begin{aligned} |nJ_3| &= \left| na^2 \int_{\theta + A_n n^{-\nu}}^{T} \frac{[|t-\theta_{u_1}|^p - |t-\theta|^p][|t-\theta_{u_2}|^p - |t-\theta|^p]}{s(t-\theta)} \, \mathrm{d}t \right| \\ &\leqslant C \sqrt{n \int [|t-\theta_{u_1}|^p - |t-\theta|^p]^2 \, \mathrm{d}t \times n \int [|t-\theta_{u_2}|^p - |t-\theta|^p]^2 \, \mathrm{d}t}, \end{aligned}$$

and hence $\lim_{n\to\infty} nJ_3 = 0$. Similarly, $\lim_{n\to\infty} nJ_1 = 0$, and combining with (2) we obtain

$$\lim_{n\to\infty} nI_1 = \tfrac{1}{2}\,\Gamma^2[|u_1|^{2p+1} + |u_2|^{2p+1} - |u_1 - u_2|^{2p+1}].$$

So, to verify (b) it remains to show that

$$\lim_{n\to\infty} nI_2 = \lim_{n\to\infty} nI_3 = \lim_{n\to\infty} nI_4 = 0.$$

For this, it is sufficient to remark that

$$\begin{aligned} n \int_0^T [\psi(t-\theta_u) - \psi(t-\theta)]^2 \, \mathrm{d}t &\leqslant nC\,|un^{-\nu}|^{2\varkappa} \\ &= C\,|u|^{2\varkappa}\, n^{-c} \to 0, \end{aligned}$$

and apply Cauchy–Schwarz inequality. So, (b) is proved.

Now, using (a), (b) and the representation (1) we can easily terminate the proof of Lemma 5. Indeed, (a) and (b) imply clearly that

$$\lim_{n\to\infty} \int_0^T g_1^k\, g_2^l\, S_\theta(t) \, \mathrm{d}t = 0$$

in the case $k + l \geqslant 3$, and hence, using (1), we have

$$\begin{aligned} \ln \phi_n(\lambda_1, \lambda_2) &\simeq -\frac{1}{2}\,\mathrm{i}\,\lambda_1\, n \int_0^T g_1^2\, S_\theta(t) \, \mathrm{d}t - \frac{1}{2}\,\mathrm{i}\,\lambda_2\, n \int_0^T g_2^2\, S_\theta(t) \, \mathrm{d}t + \\ &\quad + \frac{1}{2}\, n \int_0^T F^2\, S_\theta(t) \, \mathrm{d}t \\ &\simeq -\frac{1}{2}\,\mathrm{i}\,\lambda_1\, n \int_0^T g_1^2\, S_\theta(t) \, \mathrm{d}t - \frac{1}{2}\,\mathrm{i}\,\lambda_2\, n \int_0^T g_2^2\, S_\theta(t) \, \mathrm{d}t - \\ &\quad - \frac{1}{2}\,\lambda_1^2\, n \int_0^T g_1^2\, S_\theta(t) \, \mathrm{d}t - \frac{1}{2}\,\lambda_2^2\, n \int_0^T g_2^2\, S_\theta(t) \, \mathrm{d}t - \\ &\quad - \lambda_1\,\lambda_2\, n \int_0^T g_1\, g_2\, S_\theta(t) \, \mathrm{d}t, \end{aligned}$$

where the symbol '$\simeq$', as before, means equality of limits. So, using (b), we get finally

$$\lim_{n\to\infty}\phi_n(\lambda_1,\lambda_2) \;=\; \exp\Big\{-\frac{1}{2}\,\mathrm{i}\,\lambda_1\,\Gamma^2|u_1|^{2p+1}-\frac{1}{2}\,\mathrm{i}\,\lambda_2\,\Gamma^2|u_2|^{2p+1}- -\frac{1}{2}\,\lambda_1^2\,\Gamma^2|u_1|^{2p+1}-\frac{1}{2}\,\lambda_2^2\,\Gamma^2|u_2|^{2p+1}- -\lambda_1\lambda_2\,\Gamma^2\,\frac{|u_1|^{2p+1}+|u_2|^{2p+1}-|u_1-u_2|^{2p+1}}{2}\Big\}.$$

The last expression is clearly the characteristic function of the two-dimensional distribution $(\ln Z_\gamma(u_1),\ln Z_\gamma(u_2))$ of the process

$$\ln Z_\gamma(u)=\Gamma\,W^{p+1/2}(u)-\tfrac{1}{2}\,\Gamma^2\,|u|^{2p+1},$$

and hence the two-dimensional distributions of $Z_n(u)$ converge to the two-dimensional distributions of $Z_\gamma(u)$. The case of higher-dimensional distributions can be treated similarly. The uniformity in θ on any compact set $\mathbf{K}\subset\Theta$ is evident. Lemma 5 is proved.

Now let us prove Lemma 6. For $|u_1-u_2|\geqslant 1$ the assertion is evident since for all θ and n we have

$$\mathbf{E}_\theta|Z_n^{1/2}(u_1)-Z_n^{1/2}(u_2)|^2\leqslant 4\leqslant 4\,|u_1-u_2|^{2p+1}.$$

Suppose now that $|u_1-u_2|\leqslant 1$. Using Lemma 1.5 of [6] we can write

$$\begin{aligned}\mathbf{E}_\theta\left|Z_n^{1/2}(u_1)-Z_n^{1/2}(u_2)\right|^2 &\leqslant n\int_0^T\left[\sqrt{S_{\theta_{u_1}}(t)}-\sqrt{S_{\theta_{u_2}}(t)}\,\right]^2\,\mathrm{d}t\\ &\leqslant Cn\int_0^T[S_{\theta_{u_1}}(t)-S_{\theta_{u_2}}(t)]^2\,\mathrm{d}t\\ &\leqslant Cn\int_0^T[|t-\theta_{u_1}|^p-|t-\theta_{u_2}|^p]^2\,\mathrm{d}t+\\ &\quad+Cn\int_0^T[\psi(t-\theta_{u_1})-\psi(t-\theta_{u_2})]^2\,\mathrm{d}t\\ &= CnI_1+CnI_2\end{aligned}$$

with evident notations.

For the first integral we have clearly

$$\begin{aligned}nI_1 &= n\int_0^T[|t-\theta-u_1\,n^{-\nu}|^p-|t-\theta-u_2\,n^{-\nu}|^p]^2\,\mathrm{d}t\\ &\leqslant \int_{-\infty}^{+\infty}[|z-u_1|^p-|z-u_2|^p]^2\,\mathrm{d}z\\ &= |u_1-u_2|^{2p+1}\int_{-\infty}^{+\infty}[|x-1|^p-|x|^p]^2\,\mathrm{d}x=C\,|u_1-u_2|^{2p+1}.\end{aligned}$$

For the second one, taking into account that $|u_1 - u_2| \leqslant 1$ we get

$$\begin{aligned} nI_2 &= n\int_0^T [\psi(t-\theta-u_1\,n^{-\nu}) - \psi(t-\theta-u_2\,n^{-\nu})]^2\,\mathrm{d}t \\ &\leqslant Cn(|u_1-u_2|\,n^{-\nu})^{2\varkappa} \leqslant Cn(|u_1-u_2|\,n^{-\nu})^{2p+1} \\ &= C|u_1-u_2|^{2p+1}. \end{aligned}$$

So, in the case $|u_1 - u_2| \leqslant 1$, for all θ and n we get finally

$$\mathbf{E}_\theta|Z_n^{1/2}(u_1) - Z_n^{1/2}(u_2)|^2 \leqslant CnI_1 + CnI_2 \leqslant C\,|u_1-u_2|^{2p+1}.$$

Lemma 6 is proved.

It remains to verify Lemma 7. Using Lemma 1.5 of [6], for any n, θ and $u \in U_n$ we can write

$$\mathbf{E}_\theta Z_n^{1/2}(u) \leqslant \exp\big\{-\tfrac{1}{2}\,n\,F(u\,n^{-\nu})\big\},$$

where for all $u \in (\alpha-\theta\,,\,\beta-\theta)$ we denote

$$\begin{aligned} F(u) &= \int_0^T \Big[\sqrt{S_{\theta+u}(t)} - \sqrt{S_\theta(t)}\Big]^2\,\mathrm{d}t \geqslant c\int_0^T [S_{\theta+u}(t) - S_\theta(t)]^2\,\mathrm{d}t \\ &= c\int_{-\theta}^{T-\theta} [|y-u|^p - |y|^p]^2\,\mathrm{d}y + c\int_{-\theta}^{T-\theta} [\psi(y-u)-\psi(y)]^2\,\mathrm{d}y + \\ &\quad + c\int_{-\theta}^{T-\theta} [|y-u|^p - |y|^p][\psi(y-u)-\psi(y)]\,\mathrm{d}y \\ &= cI_1 + cI_2 \pm c\,|I_3| \end{aligned}$$

with evident notations.

For the first integral we have

$$cI_1 \leqslant C\int_{-\infty}^{+\infty} [|y-u|^p - |y|^p]^2\,\mathrm{d}y = C\,|u|^{2p+1},$$

and

$$\begin{aligned} cI_1 &= c\,|u|^{2p+1}\,\mathrm{sign}(u)\int_{-\theta/u}^{(T-\theta)/u} [|z-1|^p - |z|^p]^2\,\mathrm{d}z \\ &\geqslant c\,|u|^{2p+1}\int_0^1 [|z-1|^p - |z|^p]^2\,\mathrm{d}z = c\,|u|^{2p+1}, \end{aligned}$$

since for $u \in (0, \beta-\theta)$ we have $-\theta/u < 0$ and $(T-\theta)/u > 1$, and for $u \in (\alpha-\theta, 0)$ we have $-\theta/u > 1$ and $(T-\theta)/u < 0$.

For the second integral we get clearly $cI_2 \leqslant C|u|^{2\varkappa}$, and hence, using Cauchy–Schwarz inequality, we obtain $c\,|I_3| \leqslant C|u|^{p+1/2+\varkappa}$ for the last integral, and finally

$$F(u) \geqslant c\,|u|^{2p+1} - C|u|^{p+1/2+\varkappa} = c\,|u|^{2p+1}(1 - C\,|u|^{\varkappa-p-1/2}) \geqslant c\,|u|^{2p+1}$$

for all u such that $|u| \leqslant \delta$, where $\delta > 0$ is some fixed constant.

On the other hand, we have also

$$\inf_{|u| \geqslant \delta} F(u) = c > 0,$$

since otherwise we should have $S_{\theta+u^*}(t) = S_\theta(t)$ for some fixed u^* and almost all $t \in [0,\, T]$, which is impossible. Hence, for all $|u| \geqslant \delta$ we can write

$$F(u) \geqslant c \geqslant c\,\frac{|u|^{2p+1}}{T^{2p+1}} = c\,|u|^{2p+1}.$$

So, for all θ and $u \in (\alpha - \theta\,,\, \beta - \theta)$ we have

$$F(u) \geqslant c\,|u|^{2p+1},$$

and hence for all n, θ and $u \in U_n$ we can write

$$\mathbf{E}_\theta Z_n^{1/2}(u) \leqslant \exp\big\{-\tfrac{1}{2}\,nF(un^{-\nu})\big\} \leqslant \exp\{-c\,|u|^{2p+1}\}.$$

Lemma 7 is proved.

4. Concluding Remarks

1. For simplicity of exposition, in this paper we considered the Bayesian estimators and the notion of asymptotic efficiency in the case of quadratic loss function. In fact, the results hold for a larger class of loss functions (see [4] for more details).
2. Again for simplicity of exposition, we considered the case where the unknown parameter θ is a shift parameter, that is we supposed that $S_\theta(t) = s(t - \theta)$. In fact, the results hold in a more general situation, for example, when the intensity function is strictly positive and can be written as
   $$S_\theta(t) = d(t - \theta)|t - \theta|^p + \Psi(\theta, t),$$
   where $0 < p < 1/2$, $a^2 + b^2 > 0$, and the function $\Psi(\theta, t)$ is continuous, and uniformly in t Hölder continuous (of order higher than $p + 1/2$) with respect to θ. It is not difficult to obtain for this case the same results as those presented above. The only difference is the constant γ, which now depends on θ and is given by
   $$\gamma = \gamma(\theta) = \left(\frac{I_p(a, b)}{\Psi(\theta, \theta)}\right)^{1/(2p+1)}.$$

3. Like in Chapter 6 of [4], one can consider a situation when the intensity function has several cusps of the same order. More precisely, we suppose that $t_1 < \cdots < t_r$ with $t_r - t_1 < T$, the unknown parameter $\theta \in \Theta = (\alpha,\ \beta) \subseteq (-t_1\ ,\ T - t_r)$, and the intensity function is strictly positive and can be written as

$$S_\theta(t) = \sum_{i=1}^{r} d_i(t - \theta - t_i)|t - \theta - t_i|^p + \Psi(\theta, t),$$

where $0 < p < 1/2$,

$$d_i(x) = \begin{cases} a_i & \text{if } x < 0, \\ b_i & \text{if } x > 0, \end{cases}$$

$a_i^2 + b_i^2 > 0$, and the function $\Psi(\theta, t)$ is continuous, and uniformly in t Hölder continuous (of order higher than $p + 1/2$) with respect to θ. It is not difficult to obtain for this problem the same results as those presented above. The only difference is the constant γ, which now depends on θ and is given by

$$\gamma = \gamma(\theta) = \left(\sum_{i=1}^{r} \frac{I_p(a_i, b_i)}{S_\theta(\theta + t_i)} \right)^{1/(2p+1)}.$$

4. One can also consider similar problems of parameter estimation for the model of spatial Poisson observations. An interesting and simple example is the following. Let us consider a two-dimensional Poisson process whose intensity function has a cusp of order p along a circle of unknown radius $\rho \in (\alpha,\ \beta) \subseteq (0,\ R)$ centered at the origin. More precisely, we suppose that the intensity function is strictly positive and can be written in polar coordinates as

$$S_\rho(r, \varphi) = d(r - \rho)|r - \rho|^p + \Psi(\rho, r, \varphi),$$

where $0 < p < 1/2$, $a^2 + b^2 > 0$, and the function Ψ is continuous, and uniformly in r and φ Hölder continuous (of order higher than $p+1/2$) with respect to ρ. We observe n realizations of this Poisson process on the disk of radius R centered at the origin, and we want to estimate ρ. This problem can arise in image reconstruction theory, when we are given an optical detector counting the photoelectrons emitted by a rough surface, and we want to estimate the radius of a 'crater', whose location is known. It is not difficult to obtain for this model the same results as those presented above. The only difference is the constant γ, which now depends on ρ and is given by

$$\gamma = \gamma(\rho) = \left(\rho\, I_p(a, b) \int_0^{2\pi} \frac{1}{\Psi(\rho, \rho, \varphi)}\, \mathrm{d}\varphi \right)^{1/(2p+1)}.$$

Acknowledgements

The author would like to express his gratitude to Yu. A. Kutoyants for fruitful discussions as well as the referee for helpful suggestions and comments.

References

1. Ermakov, M. S.: On the asymptotic behavior of statistical estimates for samples having a density with singularities, *Theory Probab. and Appl.* **21** (1976), 649–651.
2. Ibragimov, I. A. and Khasminskii, R. Z.: On the asymptotic behavior of generalised Bayes' estimator, *Dokl. Akad. Nauk SSSR* **198**(3) (1970), 520–523.
3. Ibragimov, I. A. and Khasminskii, R. Z.: The asymptotic behavior of the shift parameter for samples with a continuous density with singularities, *Zap. Nauchn. Seminar. LOMI Akad. Nauk SSSR* **44** (1974), 67–93.
4. Ibragimov, I. A. and Khasminskii, R. Z.: *Statistical Estimation. Asymptotic Theory*, Springer-Verlag, New York, 1981.
5. Kutoyants, Yu. A.: *Parameter Estimation for Stochastic Processes*, Armenian Academy of Sciences, Yerevan, 1980 (in Russian), translation of revised version, Heldermann-Verlag, Berlin, 1984.
6. Kutoyants, Yu. A.: *Statistical Inference for Spatial Poisson Processes*, Lect. Notes Statist. Vol. 134, Springer-Verlag, New York, 1998.
7. Prakasa Rao, B. L. S.: Estimation of the location of the cusp of a continuous density, *Ann. Math. Statist.* **20**(1) (1968), 76–87.

ELSEVIER

Journal of Statistical Planning and Inference 117 (2003) 153–166

journal of statistical planning and inference

www.elsevier.com/locate/jspi

On cusp estimation of ergodic diffusion process

Sergueï Dachian[a], Yury A. Kutoyants[b,*]

[a] *Laboratoire de Mathématiques Appliquées, Université Blaise Pascal, 63177 Aubière Cedex, France*
[b] *Laboratoire de Statistique et Processus, France Telecom Recherche et Developpement, Université du Maine, Avenue Olivier Messiaen, 72085 Le Mans Cedex 9, France*

Received 28 November 2001; accepted 11 May 2002

Abstract

The properties of the maximum-likelihood (MLE) and Bayesian (BE) estimators of the parameter of ergodic diffusion process are studied in the situation when the trend coefficient has a cusp, i.e., it admits the representation $S(\vartheta,x)=d(x-\vartheta)|x-\vartheta|^p+h(x-\vartheta)$, where $p\in(0,\frac{1}{2})$, $d(x)=a$ for $x<0$, $d(x)=b$ for $x>0$, and the function $h(\cdot)$ is regular. This problem of estimation is not regular (Fisher information is equal to infinity), and it is shown that the rate of convergence of the estimators is $T^{1/(2p+1)}$, the estimators MLE and BE have different limit laws, and the BE is asymptotically optimal.

MSC: 62M05

Keywords: Ergodic diffusion process; Cusp estimation; Singular estimation problem; Maximum-likelihood estimator; Bayesian estimator; Limit distribution

1. Introduction

Let us consider the problem of parameter estimation by the observations of diffusion process

$$\mathrm{d}X_t=S(\vartheta,X_t)\,\mathrm{d}t+\sigma(X_t)\,\mathrm{d}W_t,\quad X_0,\ 0\leqslant t\leqslant T, \tag{1}$$

where $\vartheta\in\Theta=(\alpha,\beta)$ with $-\infty<\alpha<\beta<+\infty$ is some unknown one-dimensional parameter. The trend coefficient $S(\vartheta,x)=s(x-\vartheta)$, where the function $s(\cdot)$ is regular everywhere except 0, and has a cusp in 0. More precisely, we suppose that

* Corresponding author. Tel.: +33-2-4383-3219; fax: +33-2-4383-3579.
E-mail addresses: serguei.dachian@math.univ-bpclermont.fr (S. Dachian), kutoyants@univ-lemans.fr (Y.A. Kutoyants).

($\mathscr{J}$) *The function* $\sigma(\cdot)$ *is strictly positive and continuous, and the function* $S(\vartheta,x)$ *admits the representation*

$$S(\vartheta,x)=\begin{cases} a|x-\vartheta|^p+h(x-\vartheta) & \text{if } x\leqslant\vartheta,\\ b|x-\vartheta|^p+h(x-\vartheta) & \text{if } x\geqslant\vartheta,\end{cases}$$

where $p\in(0,\frac{1}{2})$, $a\neq 0$, $b\neq 0$, *and the function* $h(\cdot)$ *satisfies Hölder condition of order* $\mu>p+\frac{1}{2}$.

Therefore, in this parameter estimation problem the usual regularity conditions are not fulfilled, Fisher information is equal infinity, and to describe the asymptotic ($T\to\infty$) properties of the maximum-likelihood estimator (MLE) and the Bayes estimators (BE) we need a special study. For this we use general results by Ibragimov and Khasminskii (1981).

For the i.i.d. model of observations, a similar problem of parameter estimation for the densities with singularities was studied by Parakasa Rao (1968) and Ibragimov and Khasminskii (1981). More precisely, in Ibragimov and Khasminskii (1981, Chapter VI) the problem of estimation of a shift parameter ϑ by n independent observations of a random variable is considered in three different situations (three types of singularities). Our type of parameter estimation problem corresponds to the case when the density $f(\vartheta,x)$ of the observed random variables has a singularity "of the second type" at the point ϑ, that is

$$f(\vartheta,x)=\begin{cases} h(x-\vartheta)\exp\{a(x-\vartheta)|x-\vartheta|^p\} & \text{if } x\leqslant\vartheta,\\ h(x-\vartheta)\exp\{b(x-\vartheta)|x-\vartheta|^p\} & \text{if } x\geqslant\vartheta\end{cases}$$

with some regularity condition on functions $h(\cdot)$, $a(\cdot)$ and $b(\cdot)$. The asymptotic behavior of the MLE and of a wide class of BE obtained for this (i.i.d.) model is similar to those obtained here for the ergodic diffusion process model. Particularly, the rate of convergence of the estimators is $n^{1/(2p+1)}$, and the BE are asymptotically optimal.

Another similar problem of parameter estimation was studied in Dachian (2001) for the model of Poisson observations. More precisely, the problem of estimation of a shift parameter ϑ by n independent observations of the Poisson process of intensity $S_\vartheta(t)$ on a fixed interval $[0,T]$ was considered in the case when the intensity has a cusp in the point ϑ, that is

$$S_\vartheta(t)=\begin{cases} a|t-\vartheta|^p+\psi(t-\vartheta), & \text{if } t\leqslant\vartheta,\\ b|t-\vartheta|^p+\psi(t-\vartheta), & \text{if } t\geqslant\vartheta\end{cases}$$

with some regularity conditions on function $\psi(\cdot)$. Again, the results obtained for this model are very similar to those presented here, the rate of convergence of the estimators is $n^{1/(2p+1)}$, and the BE are asymptotically optimal.

We also suppose that

($\mathscr{A}_0$) *The functions* $\sigma(\cdot),\sigma(\cdot)^{-1}$ *and* $S(\vartheta,\cdot)$ *have polynomial majorants and*

$$\lim_{|x|\to\infty}\sup_{\vartheta\in\Theta}\operatorname{sgn}(x)\,\frac{S(\vartheta,x)}{\sigma(x)^2}<0.$$

By this condition process (1) has ergodic properties with invariant density

$$f(\vartheta,x)=\frac{1}{G(\vartheta)\sigma(x)^2}\exp\left\{2\int_{\vartheta}^{x}\frac{S(\vartheta,v)}{\sigma(v)^2}\,\mathrm{d}v\right\},\quad x\in\mathbb{R},$$

where

$$G(\vartheta)=\int_{\mathbb{R}}\frac{1}{\sigma(x)^2}\exp\left\{2\int_{\vartheta}^{x}\frac{S(\vartheta,v)}{\sigma(v)^2}\,\mathrm{d}v\right\}\mathrm{d}x$$

is the normalizing constant.

We consider the problem of estimation ϑ in the asymptotics of large samples, i.e., we have the continuous-time observations $X^T=\{X_t,\ 0\leqslant t\leqslant T\}$ and describe the properties of the MLE and BE as $T\to\infty$.

Remind that the likelihood ratio in this problem is

$$L(\vartheta,\vartheta_1,X^T)=\exp\left\{\int_0^T\frac{S(\vartheta,X_t)-S(\vartheta_1,X_t)}{\sigma(X_t)^2}\,\mathrm{d}X_t-\frac{1}{2}\int_0^T\frac{S(\vartheta,X_t)^2-S(\vartheta_1,X_t)^2}{\sigma(X_t)^2}\,\mathrm{d}t\right\}.$$

Here ϑ_1 is some fixed value.

The MLE $\hat{\vartheta}_T$ and BE (for quadratic loss function) $\tilde{\vartheta}_T$ are defined by the usual relations

$$L(\hat{\vartheta}_T,\vartheta_1,X^T)=\sup_{\vartheta\in\Theta}L(\vartheta,\vartheta_1,X^T)\tag{2}$$

and

$$\tilde{\vartheta}_T=\int_{\Theta}\theta q(\theta|X^T)\,\mathrm{d}\theta,\quad q(\vartheta|X^T)=\frac{q(\vartheta)L(\vartheta,\vartheta_1,X^T)}{\int_{\Theta}q(\theta)L(\theta,\vartheta_1,X^T)\,\mathrm{d}\theta}.\tag{3}$$

We suppose that the prior density $q(\cdot)$ is a positive and continuous on Θ function.

To describe the asymptotics of these estimators we need the following quantities. Let us put $H=p+\frac{1}{2}$ (the Hurst parameter) and introduce the fractional Brownian motion $W^H(\cdot)$, i.e., the Gaussian random process with zero mean and the covariance function

$$\mathbf{E}W^H(u_1)W^H(u_2)=\tfrac{1}{2}[|u_1|^{2H}+|u_2|^{2H}-|u_1-u_2|^{2H}]$$

and the stochastic process

$$Z(u)=\exp\{W^H(u)-\tfrac{1}{2}|u|^{2H}\},\quad u\in\mathbb{R}.$$

Further, let us define two random variables $\hat{u}$ and $\tilde{u}$ by relations

$$Z(\hat{u})=\sup_{u\in R}Z(u),$$

$$\tilde{u}=\frac{\int_{\mathbb{R}}uZ(u)\,\mathrm{d}u}{\int_{\mathbb{R}}Z(v)\,\mathrm{d}v}.$$

We introduce as well the function

$$\Gamma_\vartheta^2 = \frac{1}{G(\vartheta)\sigma^4(\vartheta)} \int_{-\infty}^{+\infty} [d(x-1)|x-1|^p - d(x)|x|^p]^2 \, \mathrm{d}x, \tag{4}$$

where

$$d(x) = \begin{cases} a & \text{if } x < 0, \\ b & \text{if } x > 0. \end{cases}$$

Note that $\Gamma_\vartheta^2 < \infty$ since $p < \frac{1}{2}$, and that it has the following representation (see Ibragimov and Khasminskii, 1981, Section VI.4):

$$\Gamma_\vartheta^2 = \frac{1}{G(\vartheta)\sigma(\vartheta)^4} \frac{\Gamma(1+p)\Gamma(\frac{1}{2}-p)}{2^{2p}\sqrt{\pi}(2p+1)} [a^2 + b^2 - 2ab\cos(\pi p)]$$

or equally

$$\Gamma_\vartheta^2 = \frac{\mathrm{B}(p+1, p+1)}{G(\vartheta)\sigma(\vartheta)^4} \left[\frac{a^2+b^2}{\cos(\pi p)} - 2ab \right].$$

Finally, we put $\gamma_\vartheta = \Gamma_\vartheta^{1/H}$.

2. Main results

The first result concerns the lower minimax bound.

Theorem 1. *Suppose that the conditions* $(\mathscr{A}_0)$ *and* $(\mathscr{J})$ *are fulfilled. Then, for any* $\vartheta_0 \in \Theta$,

$$\lim_{\delta\to 0} \lim_{T\to\infty} \inf_{\bar\vartheta_T} \sup_{|\vartheta-\vartheta_0|<\delta} T^{1/H} \mathbf{E}_\vartheta (\bar\vartheta_T - \vartheta)^2 \geqslant \frac{\mathbf{E}\tilde u^2}{\gamma_{\vartheta_0}^2},$$

where inf *is taken over all estimators* $\bar\vartheta_T$.

The proof of this theorem is based on the asymptotic behavior of the Bayesian estimators, so we discuss it a bit later. The more general result can be found in (Ibragimov and Khasminskii, 1981, Section I.9).

This inequality allows us to define the asymptotically efficient estimators as follows:

Definition 2. Let the conditions $(\mathscr{A}_0)$ and $(\mathscr{J})$ be fulfilled. We call an estimator $\bar\vartheta_T$ asymptotically efficient if, for any $\vartheta_0 \in \Theta$

$$\lim_{\delta\to 0} \lim_{T\to\infty} \sup_{|\vartheta-\vartheta_0|<\delta} T^{1/H} \mathbf{E}_\vartheta (\bar\vartheta_T - \vartheta)^2 = \frac{\mathbf{E}\tilde u^2}{\gamma_{\vartheta_0}^2}.$$

The properties of the estimators are described in the following theorem:

Theorem 3. *Let the conditions* $(\mathscr{A}_0)$ *and* $(\mathscr{J})$ *be fulfilled, then the MLE and BE are, uniformly on compacts* $\mathbf{K} \subset \Theta$, *consistent, have the following limits in distribution*:

$$\mathscr{L}_\vartheta\{T^{1/2H}(\hat{\vartheta}_T - \vartheta)\} \Rightarrow \mathscr{L}_\vartheta\left\{\frac{\hat{u}}{\gamma_\vartheta}\right\},$$

$$\mathscr{L}_\vartheta\{T^{1/2H}(\tilde{\vartheta}_T - \vartheta)\} \Rightarrow \mathscr{L}_\vartheta\left\{\frac{\tilde{u}}{\gamma_\vartheta}\right\}$$

and for any $k > 0$ *we have the convergence*

$$T^{k/2H}\mathbf{E}_\vartheta|\hat{\vartheta}_T - \vartheta|^k \to \mathbf{E}_\vartheta\left|\frac{\hat{u}}{\gamma_\vartheta}\right|^k,$$

$$T^{k/2H}\mathbf{E}_\vartheta|\tilde{\vartheta}_T - \vartheta|^k \to \mathbf{E}_\vartheta\left|\frac{\tilde{u}}{\gamma_\vartheta}\right|^k.$$

Moreover, the BE are asymptotically efficient.

3. Proofs

For simplicity of exposition, the proofs will be carried out in the case $a = b$. The general case proofs are similar. For convenience of notations, throughout this section C and c denote generic positive constants which can differ from formula to formula and even in the same formula.

As we are going to apply the general results by Ibragimov and Khasminskii (1981), we have to establish several properties of the likelihood ratio process

$$Z_T(u) = L(\vartheta_u, \vartheta, X^T), \quad u \in U_T = (T^\gamma(\alpha - \vartheta),\ T^\gamma(\beta - \vartheta)),$$

where $\gamma = \frac{1}{2}H$ and $\vartheta_u = \vartheta + u/T^\gamma$. These properties will be described below, in the Lemmas 5–7. But before, let us establish the following:

Lemma 4. *Let the conditions* $(\mathscr{A}_0)$ *and* $(\mathscr{J})$ *be fulfilled. Then*

1. *For any* $u_1, u_2 \in \mathbb{R}$, *uniformly on compacts* $\mathbf{K} \subset \Theta$, *the limit of the integral*

$$TI = T\int_{\mathbb{R}} \frac{[S(\vartheta_{u_1}, x) - S(\vartheta, x)][S(\vartheta_{u_2}, x) - S(\vartheta, x)]}{\sigma(x)^2} f(\vartheta, x)\,\mathrm{d}x$$

is equal to

$$\tfrac{1}{2}\,\Gamma_\vartheta^2[|u_1|^{2H} + |u_2|^{2H} - |u_1 - u_2|^{2H}].$$

Particularly,

$$\lim_{T\to\infty} T\int_{\mathbb{R}} \left(\frac{S(\vartheta_u, x) - S(\vartheta, x)}{\sigma(x)}\right)^2 f(\vartheta, x)\,\mathrm{d}x = \Gamma_\vartheta^2\,|u|^{2H}. \tag{5}$$

2. *There exists a constant* $C > 0$, *such that*

$$\sup_{\vartheta\in\mathbf{K}} T\int_{\mathbb{R}}\left(\frac{S(\vartheta_{u_1},x)-S(\vartheta_{u_2},x)}{\sigma(x)}\right)^2 f(\vartheta_{u_2},x)\,\mathrm{d}x \leqslant C|u_2-u_1|^{2H} \tag{6}$$

for all $T > 1$ *and* $u_1,u_2\in\mathbb{R}$ *such that* $|u_1-u_2|<1$.

3. *There exists a constant* $c_* = c_*(\mathbf{K}) > 0$, *such that*

$$\int_{\mathbb{R}}\left(\frac{S(\vartheta+u,x)-S(\vartheta,x)}{\sigma(x)}\right)^2 f(\vartheta,x)\,\mathrm{d}x \geqslant c_*|u|^{2H} \tag{7}$$

for all $\vartheta\in\mathbf{K}$ *and* $u\in(\alpha-\vartheta,\beta-\vartheta)$.

Proof. We start with 1. Let us write

$$TI = TI_1 + TI_2 + TI_3 + TI_4$$

with

$$I_1 = \int_{\mathbb{R}} \frac{[a|x-\vartheta_{u_1}|^p - a|x-\vartheta|^p][a|x-\vartheta_{u_2}|^p - a|x-\vartheta|^p]}{\sigma(x)^2} f(\vartheta,x)\,\mathrm{d}x,$$

$$I_2 = \int_{\mathbb{R}} \frac{[h(x-\vartheta_{u_1}) - h(x-\vartheta)][h(x-\vartheta_{u_2}) - h(x-\vartheta)]}{\sigma(x)^2} f(\vartheta,x)\,\mathrm{d}x,$$

$$I_3 = \int_{\mathbb{R}} \frac{[a|x-\vartheta_{u_1}|^p - a|x-\vartheta|^p][h(x-\vartheta_{u_2}) - h(x-\vartheta)]}{\sigma(x)^2} f(\vartheta,x)\,\mathrm{d}x,$$

$$I_4 = \int_{\mathbb{R}} \frac{[a|x-\vartheta_{u_2}|^p - a|x-\vartheta|^p][h(x-\vartheta_{u_1}) - h(x-\vartheta)]}{\sigma(x)^2} f(\vartheta,x)\,\mathrm{d}x.$$

In order to study I_1, let us fix a function $A(T)$ such that $A(T)\to+\infty$ and $A(T)/T^\gamma \to 0$ and write the integral I_1 as a sum of two: integral J_1 over the interval $L = (\vartheta - A(T)/T^\gamma, \vartheta + A(T)/T^\gamma)$, and integral J_2 over the set $M = \mathbb{R}\setminus L$.

For J_1 we have

$$\begin{aligned} TJ_1 &= Ta^2\int_{\vartheta-A(T)/T^\gamma}^{\vartheta+A(T)/T^\gamma} \frac{[|x-\vartheta_{u_1}|^p - |x-\vartheta|^p][|x-\vartheta_{u_2}|^p - |x-\vartheta|^p]}{\sigma(x)^2} f(\vartheta,x)\,\mathrm{d}x \\ &= Ta^2\int_{-A(T)/T^\gamma}^{A(T)/T^\gamma} \frac{[|y-u_1/T^\gamma|^p - |y|^p][|y-u_2/T^\gamma|^p - |y|^p]}{\sigma(y+\vartheta)^2} f(\vartheta,y+\vartheta)\,\mathrm{d}y \\ &\simeq T\,\frac{a^2 f(\vartheta,\vartheta)}{\sigma(\vartheta)^2}\int_{-A(T)}^{A(T)} [|z-u_1|^p - |z|^p][|z-u_2|^p - |z|^p]\,T^{-\gamma(2p+1)}\,\mathrm{d}z \end{aligned}$$

$$= \frac{a^2 f(\vartheta,\vartheta)}{2\sigma(\vartheta)^2}\left\{\int_{-A(T)}^{A(T)}[|z-u_1|^p-|z|^p]^2\,\mathrm{d}z+\int_{-A(T)}^{A(T)}[|z-u_2|^p-|z|^p]^2\,\mathrm{d}z\right.$$

$$\left.-\int_{-A(T)}^{A(T)}[|z-u_1|^p-|z-u_2|^p]^2\,\mathrm{d}z\right\},$$

where the symbol "$\simeq$" means equality of limits, and is true since $A(T)/T^\gamma \to 0$ and the functions $f(\vartheta,\cdot)$ and $\sigma(\cdot)$ are continuous in 0. It is easy to see that

$$\lim_{T\to\infty}\int_{-A(T)}^{A(T)}[|z-u|^p-|z|^p]^2\,\mathrm{d}z = |u|^{2H}\int_{-\infty}^{+\infty}[|x-1|^p-|x|^p]^2\,\mathrm{d}x.$$

Hence, we have clearly

$$\lim_{T\to\infty} TJ_1 = \tfrac{1}{2}I_\vartheta^2[|u_1|^{2H}+|u_2|^{2H}-|u_1-u_2|^{2H}]. \tag{8}$$

To study J_2, let us at first note that

$$T\int_M[|x-\vartheta_u|^p-|x-\vartheta|^p]^2\,\mathrm{d}x \leqslant 2T\int_{A(T)/T^\gamma}^{+\infty}[|y-u/T^\gamma|^p-|y|^p]^2\,\mathrm{d}y$$

$$= 2\int_{A(T)}^{+\infty}[|z-u|^p-|z|^p]^2\,\mathrm{d}z \to 0,$$

since $A(T)\to+\infty$ and the integral is finite. Hence, using Cauchy–Schwarz inequality, we easily get

$$|TJ_2| = \left|Ta^2\int_M\frac{[|x-\vartheta_{u_1}|^p-|x-\vartheta|^p][|x-\vartheta_{u_2}|^p-|x-\vartheta|^p]}{\sigma(x)^2}f(\vartheta,x)\,\mathrm{d}x\right|$$

$$\leqslant C\sqrt{T\int_M[|x-\vartheta_{u_1}|^p-|x-\vartheta|^p]^2\,\mathrm{d}x\times T\int_M[|x-\vartheta_{u_2}|^p-|x-\vartheta|^p]^2\,\mathrm{d}x},$$

therefore $\lim_{T\to\infty} TJ_2 = 0$, and combining with (8),

$$\lim_{T\to\infty} TI_1 = \tfrac{1}{2}I_\vartheta^2[|u_1|^{2H}+|u_2|^{2H}-|u_1-u_2|^{2H}].$$

So, it remains to show that

$$\lim_{T\to\infty} TI_2 = \lim_{T\to\infty} TI_3 = \lim_{T\to\infty} TI_4 = 0.$$

For this, it is sufficient to remark that

$$T\int_{\mathbb{R}}\frac{[h(x-\vartheta_u)-h(x-\vartheta)]^2}{\sigma(x)^2}f(\vartheta,x)\,\mathrm{d}x \leqslant TC\,|u/T^\gamma|^{2\mu}\int_{\mathbb{R}}\frac{f(\vartheta,x)}{\sigma(x)^2}\,\mathrm{d}x$$

$$\leqslant C|u|^{2\mu}T^{-(\mu-H)/H}\to 0$$

and apply Cauchy–Schwarz inequality. So, part 1 is proved.

To verify part 2 we write

$$
\begin{aligned}
&T\int_{\mathbb{R}}\left(\frac{S(\vartheta_{u_1},x)-S(\vartheta_{u_2},x)}{\sigma(x)}\right)^2 f(\vartheta_{u_2},x)\,\mathrm{d}x\\
&\quad\leqslant CT\int_{\mathbb{R}}[|x-\vartheta_{u_1}|^p-|x-\vartheta_{u_2}|^p]^2\,\mathrm{d}x\\
&\qquad+CT\int_{\mathbb{R}}[h(x-\vartheta_{u_1})-h(x-\vartheta_{u_2})]^2\frac{f(\vartheta_{u_2},x)}{\sigma(x)^2}\,\mathrm{d}x\\
&\quad=CTI_1+CTI_2
\end{aligned}
$$

with evident notations.

For the first integral, we have clearly

$$
\begin{aligned}
TI_1&=T\int_{\mathbb{R}}[|y-u_1/T^\gamma|^p-|y-u_2/T^\gamma|^p]^2\,\mathrm{d}y\\
&=\int_{\mathbb{R}}[|z-u_1|^p-|z-u_2|^p]^2\,\mathrm{d}z\\
&=|u_1-u_2|^{2H}\int_{-\infty}^{+\infty}[|x-1|^p-|x|^p]^2\,\mathrm{d}x\\
&=C|u_1-u_2|^{2H}.
\end{aligned}
$$

For the second one, taking into account that $|u_1-u_2|<1$ and $T>1$, we get

$$
\begin{aligned}
TI_2&\leqslant TC|(u_1-u_2)/T^\gamma|^{2\mu}\int_{\mathbb{R}}\frac{f(\vartheta_{u_2},x)}{\sigma(x)^2}\,\mathrm{d}x\\
&\leqslant TC|(u_1-u_2)/T^\gamma|^{2H}=C|u_1-u_2|^{2H}.
\end{aligned}
$$

So, we get finally

$$
\begin{aligned}
&T\int_{\mathbb{R}}\left(\frac{S(\vartheta_{u_1},x)-S(\vartheta_{u_2},x)}{\sigma(x)}\right)^2 f(\vartheta_{u_2},x)\,\mathrm{d}x\\
&\quad\leqslant CTI_1+CI_2\leqslant CT|u_1-u_2|^{2H}.
\end{aligned}
$$

To prove part 3 we first write

$$
\begin{aligned}
F(u)&=\int_{\mathbb{R}}[S(\vartheta+u,x)-S(\vartheta,x)]^2\,\frac{f(\vartheta,x)}{\sigma(x)^2}\,\mathrm{d}x\\
&=c\int_{\mathbb{R}}[|x-\vartheta-u|^p-|x-\vartheta|^p]^2\,\frac{f(\vartheta,x)}{\sigma(x)^2}\,\mathrm{d}x\\
&\quad+c\int_{\mathbb{R}}[h(x-\vartheta-u)-h(x-\vartheta)]^2\,\frac{f(\vartheta,x)}{\sigma(x)^2}\,\mathrm{d}x
\end{aligned}
$$

$$\pm c\int_{\mathbb{R}}[|x-\vartheta-u|^p-|x-\vartheta|^p][h(x-\vartheta-u)-h(x-\vartheta)]\frac{f(\vartheta,x)}{\sigma(x)^2}\,\mathrm{d}x$$
$$= cI_1 + cI_2 \pm cI_3$$

with evident notations. For the first integral we have

$$cI_1 \leqslant C\int_{-\infty}^{+\infty}[|y-u|^p-|y|^p]^2\,\mathrm{d}y = C|u|^{2H}$$

and

$$cI_1 \geqslant c\int_{\alpha}^{\beta}[|x-\vartheta-u|^p-|x-\vartheta|^p]^2\,\frac{f(\vartheta,x)}{\sigma(x)^2}\,\mathrm{d}x$$
$$\geqslant c\int_{\alpha}^{\beta}[|x-\vartheta-u|^p-|x-\vartheta|^p]^2\,\mathrm{d}x$$
$$= c|u|^{2H}\mathrm{sign}(u)\int_{(\alpha-\vartheta)/u}^{(\beta-\vartheta)/u}[|z-1|^p-|z|^p]^2\,\mathrm{d}z$$
$$\geqslant c\,|u|^{2H}\int_0^1[|z-1|^p-|z|^p]^2\,\mathrm{d}z = c\,|u|^{2H},$$

since for $u\in(0,\beta-\vartheta)$ we have $(\alpha-\vartheta)/u<0$ and $(\beta-\vartheta)/u>1$, and for $u\in(\alpha-\vartheta,0)$ we have $(\alpha-\vartheta)/u>1$ and $(\beta-\vartheta)/u<0$.

For the second integral we get clearly $cI_2\leqslant C|u|^{2\mu}$, and hence, using Cauchy–Schwarz inequality, we obtain $|cI_3|\leqslant C|u|^{H+\mu}$ for the last integral, and finally

$$F(u)\geqslant c|u|^{2H}-C|u|^{H+\mu}=c|u|^{2H}(1-C|u|^{\mu-H})\geqslant c_1|u|^{2H}$$

for all u such that $|u|\leqslant\delta$ where $\delta>0$ is some fixed constant.

On the other hand, we have also

$$\inf_{|u|\geqslant\delta}F(u)=c_2>0,$$

since otherwise we should have $S(\vartheta+u^*,x)=S(\vartheta,x)$ for some fixed u^* and almost all $x\in\mathbb{R}$, which is impossible. Hence, for all $|u|\geqslant\delta$ we can write

$$F(u)\geqslant c_2\geqslant c_2\,\frac{|u|^{2H}}{(\beta-\alpha)^{2H}}=c_3\,|u|^{2H}.$$

So, for all ϑ and $u\in(\alpha-\vartheta,\beta-\vartheta)$ we have

$$F(u)\geqslant c_*|u|^{2H}$$

with $c_*=\min(c_1,c_3)$. Therefore, Lemma 4 is proved. □

Now let us turn to the properties of the likelihood ratio process $Z_T(\cdot)$. Put $Z_\vartheta(u)=Z(\gamma_\vartheta u)$, $u\in\mathbb{R}$, i.e.,

$$Z_\vartheta(u)=\exp\left\{\Gamma_\vartheta W^H(u)-\frac{1}{2}\,\Gamma_\vartheta^2|u|^{2H}\right\},\quad u\in\mathbb{R}.$$

Lemma 5. *Let the conditions* $(\mathscr{A}_0)$ *and* $(\mathscr{J})$ *be fulfilled. Then the marginal distributions of the likelihood ratio* $Z_T(\cdot)$ *converge to the marginal distributions of the stochastic process* $Z_\vartheta(\cdot)$ *and this convergence is uniform in* ϑ *on the compacts* $\mathbf{K} \subset \Theta$.

Proof. As before, we put $\vartheta_u = \vartheta + u/T^\gamma$. The function $Z_T(\cdot)$ can be written as

$$\ln Z_T(u) = \int_0^T \frac{S(\vartheta_u, X_t) - S(\vartheta, X_t)}{\sigma(X_t)}\,\mathrm{d}W_t - \frac{1}{2}\int_0^T \left(\frac{S(\vartheta_u, X_t) - S(\vartheta, X_t)}{\sigma(X_t)}\right)^2 \mathrm{d}t.$$

Using local time $\Lambda_T(\vartheta, x)$ of this diffusion process (Karatzas and Shreve, 1991), we can write the second integral as

$$\begin{aligned}
&\int_0^T \left(\frac{S(\vartheta_u, X_t) - S(\vartheta, X_t)}{\sigma(X_t)}\right)^2 \mathrm{d}t \\
&= 2\int_{\mathbb{R}} \frac{[S(\vartheta_u, x) - S(\vartheta, x)]^2}{\sigma(x)^4}\,\Lambda_T(\vartheta, x)\,\mathrm{d}x \\
&= T\int_{\mathbb{R}} \left(\frac{S(\vartheta_u, x) - S(\vartheta, x)}{\sigma(x)}\right)^2 f(\vartheta, x)\,\mathrm{d}x \\
&+ T\int_{\mathbb{R}} \left(\frac{[S(\vartheta_u, x) - S(\vartheta, x)}{\sigma(x)}\right)^2 \left(\frac{2\Lambda_T(\vartheta, x)}{T\sigma(x)^2} - f(\vartheta, x)\right) \mathrm{d}x.
\end{aligned}$$

For the random function

$$\eta_T(\vartheta, x) = \sqrt{T}\left(\frac{2\Lambda_T(\vartheta, x)}{T\sigma(x)^2} - f(\vartheta, x)\right)$$

and any $m \geqslant 2$, under condition $(\mathscr{A}_0)$ we have the estimate

$$\sup_{\vartheta \in \Theta} \mathbf{E}_\vartheta |\eta_T(\vartheta, x)|^m \leqslant C_m \mathrm{e}^{-c_m|x|} \tag{9}$$

with some positive constants C_m, c_m (see Kutoyants, 2001, Proposition 1.6). Hence, we can write

$$\begin{aligned}
&\mathbf{E}_\vartheta \left| \int_{\mathbb{R}} \left(\frac{S(\vartheta_u, x) - S(\vartheta, x)}{\sigma(x)}\right)^2 \eta_T(\vartheta, x)\,\mathrm{d}x \right| \\
&\leqslant \int_{\mathbb{R}} \left(\frac{S(\vartheta_u, x) - S(\vartheta, x)}{\sigma(x)}\right)^2 \mathbf{E}_\vartheta |\eta_T(\vartheta, x)|\,\mathrm{d}x \\
&\leqslant C\int_{\mathbb{R}} \left(\frac{S(\vartheta_u, x) - S(\vartheta, x)}{\sigma(x)}\right)^2 \mathrm{e}^{-c_2|x|/2}\,\mathrm{d}x.
\end{aligned}$$

For the last integral, according to (6) we have

$$T^{1/2}\int_{\mathbb{R}}\left(\frac{S(\vartheta_u,x)-S(\vartheta,x)}{\sigma(x)}\right)^2 \mathrm{e}^{-c_2|x|/2}\,\mathrm{d}x \leqslant CT^{-1/2}|u|^{2H} \to 0.$$

Hence (see (5))

$$P-\lim_{T\to\infty}\int_0^T\left(\frac{S(\vartheta_u,X_t)-S(\vartheta,X_t)}{\sigma(X_t)}\right)^2 \mathrm{d}t = \Gamma_\vartheta^2|u|^{2H}$$

and by the central limit theorem (see Kutoyants, 1984, Theorem 3.3.3) the stochastic integral is asymptotically normal:

$$\int_0^T \frac{S(\vartheta_u,X_t)-S(\vartheta,X_t)}{\sigma(X_t)}\,\mathrm{d}W_t \Rightarrow \mathcal{N}(0,\Gamma_\vartheta^2|u|^{2H}).$$

Therefore, we have the convergence of one-dimensional distributions of $Z_T(u)$ to those of $Z_\vartheta(u)$. The proof of the convergence of the multi-dimensional distributions is based on part 1 of Lemma 4 and the mentioned central limit theorem. It is quite similar to the given one, so we omit it. □

Lemma 6. *Let the conditions* $(\mathscr{A}_0)$ *and* $(\mathscr{J})$ *be fulfilled. Then, for any compact* $\mathbf{K}\subset\Theta$, *there exist some constant* $C>0$ *such that*

$$\mathbf{E}_\vartheta|Z_T^{1/2}(u_1)-Z_T^{1/2}(u_2)|^2 \leqslant C|u_1-u_2|^{2H}$$

for all $T>1$, $\vartheta\in\mathbf{K}$ *and* $u_1,u_2\in U_T$.

Proof. For $|u_1-u_2|\geqslant 1$ the assertion is evident since for all ϑ and T we have

$$\mathbf{E}_\vartheta|Z_T^{1/2}(u_1)-Z_T^{1/2}(u_2)|^2 \leqslant 4 \leqslant 4|u_1-u_2|^{2H}.$$

Suppose now that $|u_1-u_2|<1$. Remind that the stochastic process

$$V(t)=\left(\frac{Z_t(u_2)}{Z_t(u_1)}\right)^{1/2}, \quad 0\leqslant t\leqslant T,$$

by the Itô formula admits the representation (with $\mathbf{P}_{\vartheta_{u_1}}$ probability 1)

$$V(T)=1-\frac{1}{8}\int_0^T V(t)\delta(X_t)^2\,\mathrm{d}t-\frac{1}{4}\int_0^T V(t)\delta(X_t)\,\mathrm{d}W_t,$$

where

$$\delta(x)=\frac{S(\vartheta_{u_2},x)-S(\vartheta_{u_1},x)}{\sigma(x)}.$$

Hence

$$\begin{aligned}\mathbf{E}_\vartheta|Z_T^{1/2}(u_1)-Z_T^{1/2}(u_2)|^2 &= 2-2\mathbf{E}_{\vartheta_{u_1}}V(T)\\ &\leqslant \frac{1}{4}\int_0^T \mathbf{E}_{\vartheta_{u_1}}V(t)\,\delta(X_t)^2\,\mathrm{d}t\end{aligned}$$

$$\leqslant \frac{1}{2}\int_0^T \mathbf{E}_{\vartheta_{u_2}}\delta(X_t)^2\,\mathrm{d}t + \frac{1}{2}\int_0^T \mathbf{E}_{\vartheta_{u_1}}\delta(X_t)^2\,\mathrm{d}t$$

$$= \frac{T}{2}\int_{\mathbb{R}} \delta(x)^2[f(\vartheta_{u_2},x) + f(\vartheta_{u_1},x)]\,\mathrm{d}x \leqslant C|u_2-u_1|^{2H},$$

where we used estimate (6). The lemma is proved. □

Lemma 7. *Let the conditions* $(\mathscr{A}_0)$ *and* $(\mathscr{J})$ *be fulfilled. Then, for any compact* $\mathbf{K} \subset \Theta$, *there exist some constant* $\kappa > 0$ *and some function* $C(N)$ *defined for all* $N > 0$, *such that*

$$\sup_{\vartheta\in\mathbf{K}} \mathbf{P}_\vartheta\{Z_T(u) \geqslant \mathrm{e}^{-\kappa|u|^{2H}}\} \leqslant \frac{C(N)}{|u|^N}.$$

Proof. We follow the proof of Lemma 2.4 in Kutoyants (2001). Below $0 < c_1 < 1$ and

$$\delta(u,x) = \frac{S(\vartheta_u,x) - S(\vartheta,x)}{\sigma(x)}.$$

We have

$$\mathbf{P}_\vartheta\{Z_T(u) \geqslant \mathrm{e}^{-\kappa|u|^{2H}}\}$$

$$= \mathbf{P}_\vartheta\left\{c_1\int_0^T \delta(u,X_t)\,\mathrm{d}W_t - \frac{c_1}{2}\int_0^T \delta(u,X_t)^2\,\mathrm{d}t \geqslant -c_1\kappa|u|^{2H}\right\}$$

$$\leqslant \mathbf{P}_\vartheta\left\{c_1\int_0^T \delta(u,X_t)\,\mathrm{d}W_t - \frac{c_1^2}{2}\int_0^T \delta(u,X_t)^2\,\mathrm{d}t \geqslant c_1\kappa|u|^{2H}\right\}$$

$$+ \mathbf{P}_\vartheta\left\{\frac{c_1-c_1^2}{2}\left[TF\left(\frac{u}{T^\gamma}\right) - \int_0^T \delta(u,X_t)^2\,\mathrm{d}t\right]\right.$$

$$\left.\geqslant \frac{c_1-c_1^2}{2}TF\left(\frac{u}{T^\gamma}\right) - 2c_1\kappa|u|^{2H}\right\}$$

$$\leqslant \mathrm{e}^{-c_1\kappa|u|^{2H}} + \mathbf{P}_\vartheta\left\{\frac{c_1-c_1^2}{2}\left[\int_0^T [\mathbf{E}_\vartheta\delta(u,X_t)^2 - \delta(u,X_t)^2]\,\mathrm{d}t\right]\right.$$

$$\left.\geqslant \left(\frac{c_1-c_1^2}{2}c_* - 2c_1\kappa\right)|u|^{2H}\right\},$$

where we used estimate (7). Let us denote

$$h(u,x) = \mathbf{E}_\vartheta\delta(u,X_t)^2 - \delta(u,x)^2$$

and put

$$\kappa = \frac{c_1 - c_1^2}{8c_1}\, c_*.$$

Then, for any $M > 1$, the last probability can be estimated as follows:

$$\mathbf{P}_\vartheta \left\{ \int_0^T h(u, X_t)\, \mathrm{d}t > \frac{c_*}{2}\, |u|^{2H} \right\}$$

$$\leqslant \left(\frac{2}{c_* |u|^{2H}} \right)^{2M} \mathbf{E}_\vartheta \left(\int_0^T h(u, X_t)\, \mathrm{d}t \right)^{2M}$$

$$\leqslant C|u|^{-4MH} \left(\mathbf{E}_\vartheta \left(\int_{X_0}^{X_T} \frac{H(u,x)}{\sigma(x)}\, \mathrm{d}x \right)^{2M} + T^M \mathbf{E}_\vartheta H(u,\xi)^{2M} \right),$$

where ξ is a random variable with the density $f(\vartheta, \cdot)$ and

$$H(u,x) = \frac{2}{\sigma(x) f(\vartheta, x)} \int_{-\infty}^{x} h(u,v) f(\vartheta, v)\, \mathrm{d}v.$$

Remind that $T\mathbf{E}_\vartheta \delta(u,\xi)^2 \leqslant C|u|^{2H}$. The similar estimate is valid for the function

$$T^M \mathbf{E}_\vartheta H(u,\xi)^{2M} \leqslant CT^{-M}|u|^{4MH}.$$

Hence, using $T^\gamma > |u|(\beta - \alpha)^{-1}$, we finally obtain

$$\mathbf{P}_\vartheta \{ Z_T(u) \geqslant \mathrm{e}^{-\kappa|u|^{2H}} \} \leqslant \frac{C}{T^M} \leqslant \frac{C(M)}{|u|^{M/\gamma}}.$$

The properties of the likelihood ratio described in the Lemmas 5–7 allow us to cite Theorems I.10.1 and I.10.2 by Ibragimov and Khasminskii (1981), where the general results concerning the consistency, limit distributions and convergence of moments of the MLE and BE are established. Further, the Theorem 1 now follows from the limit behavior of the Bayes estimators and Ibragimov and Khasminskii, 1981 (Theorem I.9.1). □

4. Concluding remarks

Like in Ibragimov and Khasminskii (1981, Chapter VI), one can consider a situation when the trend coefficient has several cusps. For example, we can consider the situation when the trend coefficient $S(\vartheta, x) = s(x - \vartheta)$, where the function $s(\cdot)$ is regular everywhere except at points $x_1, \ldots, x_r$, and has cusps of order p in this points. More precisely, we suppose that

$$S(\vartheta, x) = \sum_{i=1}^{r} d_i(x - \vartheta - x_i)|x - \vartheta - x_i|^p + h(x - \vartheta),$$

where

$$d_i(x) = \begin{cases} a_i & \text{if } x < 0, \\ b_i & \text{if } x > 0, \end{cases}$$

$p \in (0, \frac{1}{2})$, $a_i \neq 0$, $b_i \neq 0$, and the function $h(\cdot)$ satisfies Hölder condition of order $\mu > p + \frac{1}{2}$.

In this situation, we obtain exactly the same results as the ones presented above. The only difference is the constant Γ^2_ϑ, which is now given by

$$\Gamma^2_\vartheta = \sum_{i=1}^{r} \Gamma^2_{\vartheta,i}$$

with $\Gamma^2_{\vartheta,i}$ defined as in (4), but using $d_i(\cdot)$ in place of $d(\cdot)$. Indeed, if we introduce r independent fractional Brownian motions W_i^H, $i = 1, \ldots, r$, then it is not difficult to establish that the likelihood ratio process $Z_T(\cdot)$ converge to the stochastic process

$$Z_\vartheta(u) = \exp\left\{ \sum_{i=1}^{r} \Gamma_{\vartheta,i} W_i^H(u) - \frac{1}{2} |u|^{2H} \sum_{i=1}^{r} \Gamma^2_{\vartheta,i} \right\}$$

$$= \exp\left\{ \Gamma_\vartheta W^H(u) - \frac{1}{2} \Gamma^2_\vartheta |u|^{2H} \right\},$$

as well as the analogues of the Lemmas 6 and 7.

The problem considered here belongs to the class of problems described in Kutoyants (2001), where the observations X^T can be replaced by the observations $Y^T = \{Y_t, 0 \leqslant t \leqslant T\}$ with $Y_t = X_t \chi_{\{X_t \in [\alpha,\beta]\}}$. The MLE and BE constructed by Y^T will have the same asymptotic properties as if the whole observations X^T were used. These estimators are defined by the same relations (2) and (3), where the likelihood ratio $L(\vartheta, \vartheta_1, X^T)$ is replaced by

$$\bar{L}(\vartheta, \vartheta_1, Y^T) = \exp\left\{ \int_0^T \frac{S(\vartheta, X_t) - S(\vartheta_1, X_t)}{\sigma(X_t)^2} \chi_{\{X_t \in \mathbb{B}\}} \, \mathrm{d}X_t \right.$$

$$\left. - \frac{1}{2} \int_0^T \frac{S(\vartheta, X_t)^2 - S(\vartheta_1, X_t)^2}{\sigma(X_t)^2} \chi_{\{X_t \in \mathbb{B}\}} \, \mathrm{d}t \right\}$$

with the *window* $\mathbb{B} = [\alpha, \beta]$. The analysis of proof of the Theorem 3 (and especially of Lemma 7) shows that all the properties of the likelihood ratio established here do not change if we take $\bar{L}(\cdot)$ in the place of $L(\cdot)$. The details can be found in Kutoyants (2001).

References

Dachian, S., 2001. Estimation of cusp location by Poisson observations. Statist. Inf. Stochastic Process., to appear.

Ibragimov, I.A., Khasminskii, R.Z., 1981. Statistical Estimation: Asymptotic Theory. Springer, New York.

Karatzas, I., Shreve, S.E., 1991. Brownian Motion and Stochastic Calculus. Springer, New York.

Kutoyants, Yu.A., 1984. Parameter Estimation for Stochastic Processes. Heldermann, Berlin.

Kutoyants, Yu.A., 2001. Statistical Inference for Ergodic Diffusion Processes, forthcoming book.

Parakasa Rao, B.L.S., 1968. Estimation of the location of the cusp of a continuous density. Ann. Math. Statist. 20 (1), 76–87.

doi: 10.1111/j.1467-9469.2005.00484.x

Hypotheses Testing: Poisson Versus Self-exciting

SERGUEÏ DACHIAN

Laboratoire de Mathématiques, Université Blaise Pascal

YURY A. KUTOYANTS

Laboratoire de Statistique et Processus, Université du Maine

ABSTRACT. We consider the problem of hypotheses testing with the basic simple hypothesis: observed sequence of points corresponds to the stationary Poisson process with known intensity. The alternatives are stationary self-exciting point processes. We consider one-sided parametric and one-sided non-parametric composite alternatives and construct locally asymptotically uniformly most powerful tests. The results of numerical simulations of the tests are presented.

Key words: hypotheses testing, Poisson process, self-exciting process, uniformly most powerful test

1. Introduction

Let $\{t_1, t_2, \dots\}$ be a sequence of events of a stationary point process $X = \{X_t, t \geq 0\}$ (X_t is a counting process). The simplest stationary point process is, of course, the Poisson process with a constant intensity $S > 0$, i.e. the increments of X on disjoint intervals are independent and distributed according to the Poisson law

$$\mathbf{P}\{X_t - X_s = k\} = \frac{S^k (t-s)^k}{k!} \mathrm{e}^{-S(t-s)}, \quad 0 \leq s < t, \quad k = 0, 1, \dots .$$

Therefore, if we have a stationary sequence of events it is interesting to check first of all if this model (Poisson process) corresponds well to the observations. The importance of this problem was discussed by Cox & Lewis (1966, Section 6.3).

The alternatives close to the basic hypothesis corresponds to the case when the non-poissonian behaviour is due to the small perturbations of the Poisson process and are most interesting to test. For 'far alternatives' any reasonable test has power function close to 1 and the comparison of tests seems less important. Let us consider the problem of small signals detection by the tests of fixed size $\varepsilon \in (0, 1)$. Using the terminology of statistical radiotechnics we say that there is at least two types of close alternatives: the first one corresponds to small 'signal–noise ratio' (signals of small energy) and the second, when the amplitude of the signal can be small, but the total energy because of the sufficiently long time of observation is comparable with the noise energy (see, e.g. Kutoyants, 1976). For the first class of alternatives the approach of *locally optimal tests*, which provides the optimality of the power function at the small vicinity of the basic hypothesis (the values of the power function are close to ε) was developed (see, e.g. Capon, 1961) and for the second class of *contiguous alternatives* the optimality of the test for a wider class of close alternatives (the values of the power function are in $(\varepsilon, 1)$) was proved (Pitman's, 1948 approach; Le Cam's, 1956 theory).

For stationary point processes with Poisson hypothesis and stationary alternatives Davies (1977) proposed the *locally optimal (efficient)* or *asymptotically locally efficient* test. This test

is based on the comparison of the derivative of the log-likelihood ratio with some threshold. See Daley & Vere-Jones (2003, Section 13.1), where the approach of Davies was discussed.

In the present note we suppose that we have observations of the point process $X^T = \{X_t, 0 \le t \le T\}$ on the interval $[0, T]$ and consider two problems of hypotheses testing in the asymptotics of large samples ($T \to \infty$). In both problems the basic hypothesis is simple: the observed process is standard Poisson with known constant intensity $S_* > 0$. The composite alternatives are: the observed process is a realization of self-exciting point process [sometimes called Hawkes (1972) process] within the first case intensity function depending on one-dimensional parameter and in the second case the intensity function belonging to a wider (non-parametric) class of functions. We follow the mentioned above Pitman-Le Cam approach. We start with the 'locally asymptotically uniformly most powerful test' (LAUMPT) in the parametric case and the main result of the presented work is the LAUMPT where the optimality is shown for sufficiently large class of local non-parametric alternatives. The similar results for diffusion processes can be found in Iacus & Kutoyants (2001) (small noise asymptotics) and Kutoyants (2003) (ergodic processes).

2. Preliminaries

Recall several facts from the theory of point processes [the details can be found in, e.g. Liptser & Shiryaev (2001, Ch. 18)]. Let $(\Omega, \mathfrak{F}, \mathbf{P})$ be a probability space and let $\{\mathfrak{F}_t, t \ge 0\}$ be a non-decreasing family of right continuous σ-algebras $\mathfrak{F}_s \subset \mathfrak{F}_t \subset \mathfrak{F}$ for any $0 \le s < t$. We denote by $t_1, t_2, \ldots$, a sequence of Markov stopping times adapted to $\{\mathfrak{F}_t, t \ge 0\}$ (i.e. $\{\omega : t_i \le t\} \in \mathfrak{F}_t$ for all $t \ge 0$). Let X_t be the number of *events* t_i up to time t, i.e. $X = \{X_t, \mathfrak{F}_t, t \ge 0\}$ is a random process such that

$$X_t = \sum_{i \ge 1} \chi_{\{t_i \le t\}}, \quad t \ge 0,$$

where $\chi_{\{A\}}$ is the indicator-function of the event A.

We assume that $\mathbf{E}X_t < \infty$ (there is no accumulation points on any bounded interval). The process X admits a unique (up to stochastic equivalence) decomposition (Doob–Meyer decomposition)

$$X_t = \mathcal{A}_t + \mathcal{M}_t, \tag{1}$$

where $\mathcal{M} = \{\mathcal{M}_t, \mathfrak{F}_t, t \ge 0\}$ is a martingale and $\mathcal{A} = \{\mathcal{A}_t, \mathfrak{F}_t, t \ge 0\}$ is predictable increasing process (Liptser & Shiryayev, 2001, Theorem 18.1). We suppose that the compensator $\mathcal{A}$ is absolutely continuous

$$\mathcal{A}_t = \int_0^t S(v, \omega)\, \mathrm{d}v, \quad t \ge 0,$$

where $\mathcal{S} = \{S(t, \omega), \mathfrak{F}_t, t \ge 0\}$ is called intensity function. We suppose that (1) is the *minimal representation* of the point process, i.e. $S(t, \omega)$ is predictable with respect to the filtration generated by the counting process X and we write $S(t, \omega) = S(t, X)$. To describe a point process it is sufficient to specify its intensity function. We study in this work a special class of point processes with intensity functions that can be written as stochastic integrals with respect to the past of the underlying point process.

In the particular case when $\mathcal{S}$ is deterministic, the process X is (inhomogeneous) Poisson process with intensity function $S(v, \omega) = S(v)$. In this case

$$\mathbf{P}\{X_t - X_s = k\} = \frac{\left[\int_s^t S(v)\, \mathrm{d}v\right]^k}{k!} \exp\left\{-\int_s^t S(v)\, \mathrm{d}v\right\}$$

for any $t > s \geq 0$ and $k = 0, 1, \ldots$. If the assumption of the independence of increments is no more valid, then $\mathcal{S}$ is no more deterministic and X can be a stationary point process [see Brillinger (1975) and Daley & Vere-Jones (2003) and references therein for wide classes of such processes and their applications in real problems].

Recall that the distribution $\mathbf{P}_S^{(T)}$ of the point process in the space of its realizations $(\mathscr{D}(0, T), \mathfrak{B}_T)$ is entirely characterized by its intensity function $\mathcal{S}$. The likelihood ratio formula (w.r.t. Poisson process of constant intensity S_*) has the following form [see Liptser & Shiriyev, 2001, Theorem 19.10]

$$L(X^T) = \exp\left\{\int_0^T \ln\frac{S(t,\omega)}{S_*}\,\mathrm{d}X_t - \int_0^T [S(t,\omega) - S_*]\,\mathrm{d}t\right\},$$

where we suppose that the intensity $S(t, \omega)$ is left continuous function and

$$\mathbf{P}\left\{\int_0^T S(t,\omega)\,\mathrm{d}t < \infty\right\}$$

under all alternatives studied in this work.

3. One-sided parametric alternative

Suppose that we observe a trajectory $X^T = \{X_t, 0 \leq t \leq T\}$ of point process of intensity function $S_T(\vartheta) = \{S(\vartheta, t, \omega), 0 \leq t \leq T\}$. If $\vartheta = 0$, then $S(0, t, \omega) \equiv S_*$, i.e. this point process is a homogeneous Poisson process of intensity $S_* > 0$. Under alternative $\vartheta > 0$ and $S_T(\vartheta)$ is the intensity function of self-exciting point process. As usual in such problems, we consider contiguous alternatives (Pitman's, 1948 alternatives; Roussas, 1972), hence we change the variable $\vartheta = u/\sqrt{T}$ and test the following two hypotheses

$$\mathscr{H}_0: \quad u = 0$$
$$\mathscr{H}_1: \quad u > 0.$$

We denote $\mathbf{E}_0$ the mathematical expectation under the hypothesis $\mathscr{H}_0$, and $\mathbf{E}_u$ under (simple) alternative $\vartheta = u/\sqrt{T}$.

Let us fix $\varepsilon \in (0, 1)$ and denote by $\mathscr{K}_\varepsilon$ the class of test functions $\phi_T(X^T)$ of asymptotic size ε, i.e. for $\phi_T \in \mathscr{K}_\varepsilon$ we have

$$\lim_{T\to\infty} \mathbf{E}_0 \phi_T(X^T) = \varepsilon. \tag{2}$$

As usual, $\phi_T(X^T)$ is the probability to accept the hypothesis $\mathscr{H}_1$ having observations X^T. The corresponding power function is

$$\beta_T(u, \phi_T) = \mathbf{E}_u \phi_T(X^T), \quad u \geq 0.$$

We introduce the asymptotic optimality of tests with the help of the following definition Le Cam (1956).

Definition 1
A test $\phi_T^*(\cdot)$ *is called* locally asymptotically uniformly most powerful in the class $\mathscr{K}_\varepsilon$ *if for any other test* $\phi_T(\cdot) \in \mathscr{K}_\varepsilon$ *and any constant* $K > 0$ *we have*

$$\lim_{T\to\infty} \inf_{0<u\leq K} [\beta_T(u, \phi_T^*) - \beta_T(u, \phi_T)] \geq 0.$$

Our goal is to construct locally asymptotically uniformly most powerful test in class $\mathscr{K}_\varepsilon$.

Self-exciting type processes were introduced by Hawkes (1972) and defined by intensity function of the following form

$$S(t,\omega)=S_*+\int_0^{t-} g(t-s)\,\mathrm{d}X_s=S_*+\sum_{t_i<t} g(t-t_i), \tag{3}$$

where $S_*>0$, t_i are the events of the point process and the function $g(\cdot)\geq 0$ satisfies the condition

$$\rho=\int_0^{\infty} g(t)\,\mathrm{d}t<1. \tag{4}$$

Recall that according to this representation of the intensity function, the distribution of t_1 is exponential at rate S_* and for all $n\geq 1$

$$\mathbf{P}\{t_{n+1}>t|t_1,\ldots,t_n\}=\exp\left(-S_*t-\int_0^t\sum_{i=1}^{X_s} g(s-t_i)\,\mathrm{d}s\right).$$

Note that $\Lambda(t)=\mathbf{E}X_t$ is solution of the equation

$$\begin{aligned}\Lambda(t)&=\mathbf{E}\int_0^t S(v,\omega)\,\mathrm{d}v=S_*t+\mathbf{E}\int_0^t\int_0^{v-} g(v-s)\,\mathrm{d}X_s\,\mathrm{d}v\\&=S_*t+\int_0^t\int_0^v g(v-s)\Lambda'(s)\,\mathrm{d}v\,\mathrm{d}s.\end{aligned}$$

In stationary case the intensity $S(t,\omega)$, is a stationary process

$$S(t,\omega)=S_*+\int_{-\infty}^{t-} g(t-s)\,\mathrm{d}X_s$$

and

$$\Lambda(t)=\frac{S_*}{1-\rho}t\equiv\mu t.$$

The spectral density of this process is

$$f(\lambda)=\frac{\mu}{2\pi|1-G(\lambda)|^2},$$

where

$$G(\lambda)=\int_0^{\infty}\mathrm{e}^{\mathrm{i}\lambda t}g(t)\,\mathrm{d}t,\quad \rho=G(0).$$

Example 1. Let $g(t)=\alpha\mathrm{e}^{-\gamma t}$, where $\alpha>0,\gamma>0$ and $\alpha/\gamma<1$. Then the point process X with intensity function

$$S(t,\omega)=S_*+\alpha\sum_{t_i\leq t}\mathrm{e}^{-\gamma(t-t_i)}$$

is self-exciting with the rate

$$\mu=\frac{S_*\gamma}{\gamma-\alpha}.$$

Example 2. The function $g(\cdot)$ can be chosen in such a way that the spectral density of the point process will be rational

$$f(\lambda)=\frac{\mu}{2\pi}\frac{|Q(\mathrm{i}\lambda)|^2}{|P(\mathrm{i}\lambda)|^2},$$

where $Q(z)=z^p+a_1z^{p-1}+\cdots+a_p$ and $P(z)=z^p+b_1z^{p-1}+\cdots+b_p$. It is supposed that $P(\cdot)$ and $Q(\cdot)$ have no zeroes in common and no zeroes in the closed right half plane [see Pham (1981), where the asymptotic properties of the maximum likelihood estimator (MLE) for this model are described].

We assume that the observed process is either Poisson with constant intensity S_* or self-exciting with *contiguous* intensity function

$$S(\vartheta,t,\omega)=S_*+\vartheta_T\int_0^t h(t-s)\,\mathrm{d}X_s.$$

Contiguous means that the likelihood ratio is asymptotically non-degenerate. The function $h(\cdot)$ is supposed to be known, bounded and

$$h(\cdot)\in\mathcal{L}^1_+(\mathbb{R}_+)=\left\{f(\cdot)\geq 0:\int_0^\infty f(t)\,\mathrm{d}t<\infty\right\}.$$

To have contiguous alternatives we choose, as usual in regular problems, $\vartheta_T=u/\sqrt{T}$, i.e.

$$S(u,t,\omega)=S_*+\frac{u}{\sqrt{T}}\int_0^t h(t-s)\,\mathrm{d}X_s,\quad u\geq 0.$$

Note that for any $h(\cdot)\in\mathcal{L}^1_+(\mathbb{R}_+)$ and any $u\leq K$ for sufficiently large T the condition (4) is fulfilled for the corresponding function $g(\cdot)=uT^{-1/2}h(\cdot)$. This leads us to the following one-sided hypotheses testing problem:

$$\mathscr{H}_0:\quad u=0,\qquad \text{(Poisson process)}$$
$$\mathscr{H}_1:\quad u>0,\qquad \text{(self-exciting process)}.$$

This model corresponds to 'small self-exciting perturbations' of the Poisson process of intensity S_*.

Note that as we use the LAN approach (Le Cam, 1956), we study the behaviour of the tests statistics under hypothesis only (Poisson process with constant intensity) and do not use the stationarity of the self-exciting processes under alternatives. The limit of the power function is obtained using LAN and Le Cam's Third Lemma.

Let us denote

$$\Delta_T(X^T)=\frac{1}{S_*\sqrt{T\ \mathrm{I}^*_h}}\int_0^T\int_0^{t-}h(t-s)\,\mathrm{d}X_s[\mathrm{d}X_t-S_*\,\mathrm{d}t].$$

Here

$$\int_0^{t-}h(t-s)\,\mathrm{d}X_s=\sum_{t_i<t}h(t-t_i)$$

(limit from the left of the integral, i.e. the term with $s_i=t$ is excluded) and

$$\mathrm{I}^*_h=\int_0^\infty h(t)^2\,\mathrm{d}t+S_*\left(\int_0^\infty h(t)\,\mathrm{d}t\right)^2$$

is the Fisher information of the problem. Throughout this paper we denote by z_ε the $1-\varepsilon$ quantile of the Gaussian law $\mathcal{N}(0,1)$.

Theorem 1

Let $h(\cdot)\in\mathcal{L}^1_+\left(\mathbb{R}_+\right)$ and bounded. Then the test

$$\hat{\phi}_T(X^T)=\chi_{\{\Delta_T(X^T)>z_\varepsilon\}}$$

is locally asymptotically uniformly most powerful in the class $\mathscr{K}_\varepsilon$ and for any $u>0$ its power function

$$\beta_T(u,\hat{\phi}_T)\longrightarrow\hat{\beta}(u)=\mathbf{P}\{\zeta>z_\varepsilon-u\sqrt{\mathrm{I}_h^*}\},\tag{5}$$

where $\zeta\sim\mathcal{N}(0,1)$.

Proof. First note that the family of measures $\{\mathbf{P}_\vartheta^{(T)},\vartheta>0\}$ under hypothesis $\mathscr{H}_0$ is LAN at the point $\vartheta=0$, i.e. the random function $Z_T(u)=L(u/\sqrt{T},X^T)$ admits the representation (see Kutoyants, 1984, Theorem 4.5.3)

$$\begin{aligned}Z_T(u)&=\exp\left\{\int_0^T\ln\left(1+\frac{u}{S_*\sqrt{T}}\int_0^{t-}h(t-s)\,\mathrm{d}X_s\right)\mathrm{d}X_t\right.\\&\qquad\left.-\frac{u}{\sqrt{T}}\int_0^T\int_0^t h(t-s)\,\mathrm{d}X_s\,\mathrm{d}t\right\}\\&=\exp\left\{u\sqrt{\mathrm{I}_h^*}\Delta_T(X^T)-\frac{u^2}{2}\mathrm{I}_h^*+r_T(u,X^T)\right\},\end{aligned}$$

where

$$\mathcal{L}_0\left\{\Delta_T(X^T)\right\}\Longrightarrow\mathcal{N}(0,1)\tag{6}$$

and $r_T\left(u_T,X^T\right)\to0$ for any bounded sequence $\{u_T\}$.

To verify (6) we check the following two conditions:

1 *Lindeberg condition* for stochastic integral: for any $\delta>0$

$$\lim_{T\to\infty}\frac{1}{T}\mathbf{E}_0\int_0^T H_t^2\chi_{\{|H_t|>\delta\sqrt{T}\}}\,\mathrm{d}t=0,$$

2 the *law of large numbers*:

$$\mathbf{P}_0-\lim_{T\to\infty}\frac{1}{S_*T}\int_0^T H_t^2\,\mathrm{d}t=\mathrm{I}_h^*.\tag{7}$$

Here, we denote

$$H_t=\int_0^{t-}h(t-s)\,\mathrm{d}X_s.$$

By these conditions the stochastic integral $\Delta_T(X^T)$ is asymptotically normal. The proof of the corresponding central limit theorem can be found, say, in Kutoyants (1984, Theorem 4.5.4) (of course, this theorem is a particular case of general CLT for martingales).

To check these conditions we introduce an independent Poisson process $\{X_t,t\le0\}$ of intensity S_* and replace H_t by

$$H_t^*=\int_{-\infty}^{t-}h(t-s)\,\mathrm{d}X_s.$$

It is easy to see that for the process $H_t^*,t\ge0$ we have

$$\mathbf{E}_0H_t^*=S_*\int_0^\infty h(v)\,\mathrm{d}v$$

and

$$\mathbf{E}_0([H_t^*-\mathbf{E}_0H_t^*][H_s^*-\mathbf{E}_0H_s^*])=S_*\int_{\max(0,s-t)}^\infty h(v+t-s)h(v)\,\mathrm{d}v.$$

Note as well that

$$\begin{aligned}
&\mathbf{P}_0\left\{\frac{1}{\sqrt{T}}\int_0^T [H_t^*-H_t][\mathrm{d}X_t - S_*\,\mathrm{d}t] > v\right\}\\
&\quad\le \frac{S_*}{Tv^2}\int_0^T \mathbf{E}_0\left(\int_{-\infty}^0 h(t-s)\,\mathrm{d}X_s\right)^2 \mathrm{d}t\\
&\quad= \frac{S_*^2}{Tv^2}\int_0^T\left[\int_t^\infty h(v)^2\,\mathrm{d}v + S_*\left(\int_t^\infty h(v)\,\mathrm{d}v\right)^2\right]\mathrm{d}t \longrightarrow 0,
\end{aligned}$$

as $T\to\infty$.

Now the process $H_t^*, t\ge 0$ is second-order stationary and

$$\mathbf{E}_0(H_t^*)^2 = S_*\int_0^\infty h(t)^2\,\mathrm{d}t + S_*^2\left(\int_0^\infty h(t)\,\mathrm{d}t\right)^2 = \mathbf{E}_0(H_0^*)^2 < \infty.$$

Hence

$$\mathbf{E}_0\left(H_t^{*2}\chi_{\{|H_t^*|>\delta\sqrt{T}\}}\right) = \mathbf{E}_0\left(H_0^{*2}\chi_{\{|H_0^*|>\delta\sqrt{T}\}}\right)\longrightarrow 0,$$

as $T\to\infty$ and

$$\lim_{T\to\infty}\frac{1}{T}\int_0^T \mathbf{E}_0\left(H_t^{*2}\chi_{\{|H_t^*|>\delta\sqrt{T}\}}\right)\mathrm{d}t = 0.$$

The law of large numbers (7) will follow from the convergence:

$$\begin{aligned}
M_T &= \mathbf{E}_0\left(\frac{1}{T}\int_0^T H_t^{*2}\,\mathrm{d}t - \mathbf{E}_0(H_0^*)^2\right)^2\\
&= \frac{1}{T^2}\int_0^T\int_0^T \mathbf{E}_0(H_t^{*2}-\mathbf{E}_0(H_0^*)^2)(H_s^{*2}-\mathbf{E}_0(H_0^*)^2)\,\mathrm{d}t\,\mathrm{d}s \longrightarrow 0.
\end{aligned}$$

To prove it we need the following elementary result.

Lemma 1

Let $X=\{X_t, t\in A\}$ be a Poisson process of constant intensity $S>0$ on $A\subset\mathbb{R}$, and let $f(\cdot), g(\cdot)\in \mathcal{L}^k(A)=\left\{f(\cdot): \int_A |f(t)|^k\,\mathrm{d}t<\infty\right\}$, $k=1,\dots,4$. Then

$$\begin{aligned}
&\mathrm{Cov}\left(\left(\int_A f(v)\,\mathrm{d}X_v\right)^2, \left(\int_A g(v)\,\mathrm{d}X_v\right)^2\right)\\
&=4\int_A f(v)S\,\mathrm{d}v\int_A g(v)S\,\mathrm{d}v\int_A f(v)g(v)S\,\mathrm{d}v\\
&\quad+2\left(\int_A f(v)g(v)S\,\mathrm{d}v\right)^2 + \int_A f^2(v)g^2(v)S\,\mathrm{d}v\\
&\quad+2\int_A f(v)S\,\mathrm{d}v\int_A f(v)g^2(v)S\,\mathrm{d}v + 2\int_A g(v)S\,\mathrm{d}v\int_A f^2(v)g(v)S\,\mathrm{d}v.
\end{aligned}$$

Proof. Using well-known properties of the Poisson processes (see, e.g. Kutoyants, 1998, Lemma 1.1), we obtain the moment generating function

$$\begin{aligned}
\phi(\lambda,\mu) &= \mathbf{E}_0\exp\left\{\lambda\int_A f(v)\,\mathrm{d}X_v + \mu\int_A g(v)\,\mathrm{d}X_v\right\}\\
&= \exp\left\{\int_A (\mathrm{e}^{\lambda f(v)+\mu g(v)}-1)\,S\,\mathrm{d}v\right\}.
\end{aligned}$$

Recall that

$$\mathrm{Cov}\left(\left(\int_A f(v)\,\mathrm{d}X_v\right)^2,\left(\int_A g(v)\,\mathrm{d}X_v\right)^2\right)$$
$$=\left.\frac{\partial^4\phi(\lambda,\mu)}{\partial\lambda^2\partial\mu^2}\right|_{\lambda=0,\,\mu=0}-\left.\frac{\partial^2\phi\,(\lambda,0)}{\partial\lambda^2}\right|_{\lambda=0}\left.\frac{\partial^2\phi(0,\mu)}{\partial\mu^2}\right|_{\mu=0}.$$

Therefore the proof of the lemma follows from direct calculations.

Now we can write

$$R(t,s)=\mathbf{E}_0((H_t^*)^2-\mathbf{E}_0(H_0^*)^2)((H_s^*)^2-\mathbf{E}_0(H_0^*)^2)$$
$$=4a^2K(t,s)+2K(t,s)^2+S_*\int_{-\infty}^{t\wedge s}h(t-v)^2h(s-v)^2\,\mathrm{d}v$$
$$+2aS_*\int_{-\infty}^{t\wedge s}[h(t-v)^2h(s-v)+h(t-v)h(s-v)^2]\,\mathrm{d}v,$$

where we put

$$a=S_*\int_0^\infty h(y)\,\mathrm{d}y$$

and (for $\tau=t-s$)

$$K(t,s)=S_*\int_{|\tau|}^\infty h(y)h(y-|\tau|)\,\mathrm{d}y=K(\tau).$$

Further, as the function $h(\cdot)$ is bounded, we have the estimate

$$R(t,s)\le CK(\tau).$$

Hence

$$M_T=\frac{1}{T^2}\int_0^T\int_0^T R(t,s)\,\mathrm{d}t\,\mathrm{d}s\le\frac{C}{T^2}\int_0^T\int_0^T K(t,s)\,\mathrm{d}t\,\mathrm{d}s$$
$$\le\frac{C}{T}\int_{-T}^T K(\tau)\,\mathrm{d}\tau.$$

For the function $K(\cdot)$ we have

$$\int_{-T}^T K(\tau)\,\mathrm{d}\tau=S_*\int_{-T}^T\int_{|\tau|}^\infty h(y)h(y-|\tau|)\,\mathrm{d}y\,\mathrm{d}\tau\le 2S_*\left(\int_0^\infty h(y)\,\mathrm{d}y\right)^2.$$

Hence $M_T\to 0$ and we have the law of large numbers (7).

The property $\hat\phi_T(\cdot)\in\mathscr{K}_\varepsilon$ follows from the mentioned above asymptotic normality of the statistic $\Delta_T(X^T)$.

Note as well that the convergence (5) follows from

$$\mathcal{L}_u\{\Delta_T(X^T)\}\Longrightarrow\mathcal{N}(u\sqrt{\mathrm{I}_h^*},1)$$

[see the Third Lemma of Le Cam (van der Vaart, 1998, p. 90)].

The asymptotic optimality of the test follows as well from the general theory (see, e.g. Le Cam, 1956 or Roussas, 1972), because if we replace $\mathscr{H}_1$ by any simple alternative $\mathscr{H}_*:u=u_*$, then the test

$$\bar\phi_T(X^T)=\chi_{\{L(u_*/\sqrt{T},\,X^T)>b_\varepsilon\}}$$

is the most powerful. Here

$$b_\varepsilon = \exp\left\{u_* z_\varepsilon \sqrt{\mathrm{I}_h^*} - \frac{1}{2}u_*^2 \mathrm{I}_h^*\right\}(1+o(1)).$$

It is easy to see that $\bar{\phi}_T(\cdot) \in \mathscr{K}_\varepsilon$ and the power function

$$\beta_T(u_*, \bar{\phi}_T) \to \hat{\beta}(u_*).$$

Therefore the test $\hat{\phi}_T(\cdot)$ is asymptotically as good as the likelihood ratio test for any simple alternative.

Remark 1. Note that the statistic $\Delta_T(X^T)$ can be written as follows

$$\Delta_T(X^T) = \frac{1}{S^*\sqrt{T\mathrm{I}_h^*}} \sum_{0\le t_j\le T} \sum_{t_i<t_j} h(t_j - t_i) - \frac{1}{\sqrt{T\mathrm{I}_h^*}} \sum_{0\le t_j\le T} \int_0^{T-t_j} h(v)\,\mathrm{d}v,$$

where t_i are the events of the observed process.

Remark 2. By a similar way we can consider the problem of contiguous hypotheses testing when under the hypothesis $\mathscr{H}_0$ the observed process is self-exciting too. For example, let $h(\vartheta, t) \ge 0, t \ge 0$ be a smooth function of $\vartheta \in \Theta$, such that for all $\vartheta \in \Theta$ the condition

$$\int_0^\infty h(\vartheta, t)\,\mathrm{d}t < 1$$

holds. Then with the help of this function we introduce a family of self-exciting processes with intensity functions

$$S(\vartheta, t, \omega) = S_* + \int_{-\infty}^t h(\vartheta, t-s)\,\mathrm{d}X_s.$$

Recall that these are stationary processes.

Now we can test the hypotheses

$$\mathscr{H}_0: \quad \vartheta = \vartheta_0,$$
$$\mathscr{H}_1: \quad \vartheta > \vartheta_0$$

by the observations $X^T = \{X_t, 0 \le t \le T\}$. Suppose as well that the function $h(\vartheta, \cdot)$ is two times continuously differentiable on ϑ at the point $\vartheta = \vartheta_0$ and the derivatives $\dot{h}(\vartheta, \cdot), \ddot{h}(\vartheta, \cdot)$ satisfy suitable conditions of integrability. Let us denote

$$\xi_t(\vartheta) = \int_0^{t-} h(\vartheta, t-s)\,\mathrm{d}X_s, \quad \dot{\xi}_t(\vartheta) = \int_0^{t-} \frac{\partial h(\vartheta, t-s)}{\partial \vartheta}\,\mathrm{d}X_s,$$

and put

$$\Delta_T(\vartheta_0, X^T) = \frac{1}{\sqrt{T}} \int_0^T \frac{\dot{\xi}_t(\vartheta_0)}{S_* + \xi_t(\vartheta_0)} [\mathrm{d}X_t - S_*\,\mathrm{d}t - \xi_t(\vartheta_0)\,\mathrm{d}t].$$

Then it can be easily shown that the test

$$\hat{\phi}_T(X^T) = \chi_{\{\Delta_T(\vartheta_0, X^T) > c_\varepsilon\}},$$

where $c_\varepsilon = z_\varepsilon \sqrt{\mathrm{I}_h(\vartheta_0)}$ is chosen from the condition $\hat{\phi}_T \in \mathscr{K}_\varepsilon$ is locally asymptotically uniformly most powerful in the class $\mathscr{K}_\varepsilon$. Here $\mathrm{I}_h(\vartheta_0)$ is the Fisher information

$$\mathrm{I}_h(\vartheta_0) = \mathbf{E}_{\vartheta_0}\left(\frac{\dot{\xi}(\vartheta_0)^2}{S_* + \xi(\vartheta_0)}\right),$$

where $\dot{\xi}(\vartheta_0)$ and $\xi(\vartheta_0)$ are 'stationary random variables' related to the limit distribution of the vector $\dot{\xi}_t(\vartheta_0), \xi_t(\vartheta_0)$.

4. Testing of dependence

Suppose that we have two sequences of events $0<t_1<t_2<\cdots<t_N<T$ and $0<s_1<s_2<\cdots<s_M<T$ with corresponding counting processes $X^T=\{X_t, 0\le t\le T\}$ and $Y^T=\{Y_t, 0\le t\le T\}$. The first process is Poisson with constant known intensity function $S_X(t,\omega)=S_X>0$ and the intensity function of the second process can be written as

$$S_Y(t,\omega)=S_Y+\int_{-\infty}^{t} r(t-s)\,\mathrm{d}X_s,$$

where $r(\cdot)\in\mathcal{L}^1(\mathbb{R}_+)$. Therefore, if $r(t)\equiv 0$, then the observed processes are standard (independent) Poisson processes of intensities S_X and S_Y respectively (Hypothesis $\mathscr{H}_0$). For the other values of $r(\cdot)$ we have dependent point processes.

We suppose that the dependence between these two processes, if exists, is weak, i.e. the function $r(\cdot)$ is sufficiently small and we can apply the local approach. As before we suppose that $r(t)=\vartheta_T h(t)$, where $h(\cdot)\in\mathcal{L}^1(\mathbb{R}_+)$ and $\vartheta_T=u/\sqrt{T}\to 0$.

$\mathscr{H}_0$: $u=0$, (independent Poisson processes)

$\mathscr{H}_1$: $u>0$, (depending processes).

Introduce the statistic

$$\begin{aligned}\Delta_T(X^T,Y^T)&=\frac{1}{S_Y\sqrt{T\mathrm{I}_h}}\int_0^T\int_0^{t-} h(t-s)\,\mathrm{d}X_s[\mathrm{d}Y_t-S_Y\,\mathrm{d}t]\\&=\frac{1}{S_Y\sqrt{T\mathrm{I}_h}}\sum_{0\le s_j\le T}\sum_{t_j<s_i} h(s_j-t_i)-\frac{1}{\sqrt{T\mathrm{I}_h}}\sum_{0\le t_j\le T}\int_0^{T-t_j} h(v)\,\mathrm{d}v,\end{aligned}$$

where

$$\mathrm{I}_h=\frac{S_X}{S_Y}\left(\int_0^\infty h(t)^2\,\mathrm{d}t+S_X\left(\int_0^\infty h(t)\,\mathrm{d}t\right)^2\right).$$

Proposition 1
Let $h(\cdot)\in\mathcal{L}^1_+(\mathbb{R}_+)$ and bounded. Then the test

$$\hat{\phi}_T(X^T,Y^T)=\chi_{\{\Delta_T(X^T,Y^T)>z_\varepsilon\}}$$

is locally asymptotically uniformly most powerful in the class $\mathscr{K}_\varepsilon$ and for any $u>0$ its power function

$$\beta_T(u,\hat{\phi}_T)\longrightarrow\hat{\beta}(u)=\mathbf{P}\{\zeta>z_\varepsilon-u\sqrt{I_h}\},\tag{8}$$

where $\zeta\sim\mathcal{N}(0,1)$.

Proof. The proof is quite close to the given above proof of the theorem 1, and hence is omitted.

Remark 3. The similar problem can be considered for the couple of mutually exciting point processes with intensity functions

$$S_X(t,\omega)=S_X+\int_{-\infty}^{t} r_{XY}(t-s)\,\mathrm{d}Y_s,\ \ S_Y(t,\omega)=S_Y+\int_{-\infty}^{t} r_{YX}(t-s)\,\mathrm{d}X_s,$$

where $r_{XY}(\cdot), r_{YX}(\cdot)\in\mathcal{L}^1(\mathbb{R}_+)$. Therefore, if $r_{XY}(t)\equiv 0$ and $r_{YX}(t)\equiv 0$, then the observed processes are standard (independent) Poisson processes of intensities $S_X>0$ and $S_Y>0$ respec-

tively (Hypothesis $\mathscr{H}_0$). Under alternative there exists a weak dependence of these processes through their intensity functions.

5. One-sided non-parametric alternative

In all the above considered problems the alternatives are one-sided parametric. It is possible to describe similar asymptotically uniformly most powerful tests even in some non-parametric situations. Using the minimax approach we can consider the least favourable model in the derivation of the upper bound on the powers of all tests, but, of course, for special classes of intensities. This approach sometimes is called semiparametric and the rate of convergence of alternatives is $\sqrt{T}$.

As before, we suppose that under hypothesis $\mathscr{H}_0$ the observed point process $X^T=\{X_t, 0\le t\le T\}$ is standard Poisson with known intensity function $S(t)=S_*>0$ and under alternative $\mathscr{H}_1$ it is self-exciting point process with intensity function

$$S(t,\omega)=S_*+\int_{-\infty}^{t} g(t-s)\,\mathrm{d}X_s,\quad 0\le t\le T,$$

where $g(\cdot)$ is now *unknown* function. We suppose as well that

$$\int_0^\infty g(t)\,\mathrm{d}t<1, \tag{9}$$

hence the process X^T is stationary. To describe the class of local non-parametric alternatives we rewrite this intensity function as

$$S(t,\omega)=S_*+\frac{1}{\sqrt{T}}\int_{-\infty}^{t} u(t-s)\,\mathrm{d}X_s,\quad 0\le t\le T,$$

where the function $u(\cdot)$ belongs to the set $\mathscr{U}_r$ defined below. Let us denote by $\mathcal{C}_+^b$ the set of non-negative functions bounded by the same constant and introduce the set

$$\mathscr{U}_r=\left\{u(\cdot)\in\mathcal{C}_+^b:\quad \int_0^\infty u(t)\,\mathrm{d}t=r, \operatorname{supp} u(\cdot) \text{ is bounded}\right\}.$$

Note, that for any $r>0$ and $T>r^2$ the condition (9) is fulfilled.

Therefore, we consider the following hypotheses testing problem

$$\mathscr{H}_0:\qquad u(\cdot)\equiv 0,$$

$$\mathscr{H}_1:\qquad u(\cdot)\in\mathscr{U}_r,\quad r>0.$$

The power function of a test ϕ_T depends on the function $u(\cdot)$ and we write it as

$$\beta_T(u,\phi_T)=\mathbf{E}_u\phi_T(X^T),$$

where $u=u(\cdot)\in\mathscr{U}_r$ with some $r>0$. We want to apply an approach similar to the minimax one in the estimation theory. More precisely, we seek to maximize the minimal power of test on the class $\mathscr{U}_r$. However, for any test $\phi_T\in\mathscr{K}_\varepsilon$ we have

$$\inf_{u(\cdot)\in\mathscr{U}_r}\beta_T(u,\phi_T)\le\varepsilon,$$

since for any $T>0$ we can take a function from $\mathscr{U}_r$ equal 0 on $[0,T]$. Hence we introduce the set

$$\mathscr{U}_{r,N}=\{u(\cdot)\in\mathscr{U}_r:\quad \operatorname{supp} u(\cdot)\subset[0,N]\},$$

denote

$$B_T(r,N,\phi_T)=\inf_{u(\cdot)\in\mathscr{U}_{r,N}}\beta_T(u,\phi_T)$$

and give the following.

Definition 2
A test $\phi_T^*(\cdot)$ *is called* locally asymptotically uniformly most powerful in the class $\mathscr{K}_\varepsilon$ *if for any other test* $\phi_T(\cdot)\in\mathscr{K}_\varepsilon$ *and any* $K>0$ *we have*

$$\lim_{N\to\infty}\lim_{T\to\infty}\inf_{0\le r\le K}[B_T(r,N,\phi_T^*)-B_T(r,N,\phi_T)]\ge 0.$$

Let us introduce the decision function

$$\hat\phi_T(X^T)=\chi_{\{\delta_T(X^T)>z_\varepsilon\}},\quad \delta_T(X^T)=\frac{X_T-S_*T}{\sqrt{S_*T}}.$$

Theorem 2
The test $\hat\phi_T$ *is locally asymptotically uniformly most powerful in the class* $\mathscr{K}_\varepsilon$ *and for any* $u(\cdot)\in\mathscr{U}_r$ *its power function*

$$\beta_T(u,\hat\phi_T)\longrightarrow\hat\beta(u)=\mathbf{P}\left\{\zeta>z_\varepsilon-r\sqrt{S_*}\right\},\tag{10}$$

where $\zeta\sim\mathcal{N}(0,1)$.

Proof. Let us fix a simple alternative $u(\cdot)\in\mathscr{U}_r$, then the likelihood ratio $L_T\left(\frac{u(\cdot)}{\sqrt{T}},X^T\right)=Z_T\left(u(\cdot)\right)$ admits (under hypothesis $\mathscr{H}_0$) the representation (see the proof of the theorem 1)

$$\begin{aligned}Z_T\left(u(\cdot)\right)&=\exp\left\{\int_0^T\ln\left(1+\frac{1}{S_*\sqrt{T}}\int_0^{t-}u(t-s)\,\mathrm{d}X_s\right)\mathrm{d}X_t\right.\\&\qquad\left.-\frac{1}{\sqrt{T}}\int_0^T\int_0^t u(t-s)\,\mathrm{d}X_s\,\mathrm{d}t\right\}\\&=\exp\left\{\Delta_T(u,X^T)-\frac{1}{2}\mathrm{I}(u)+r_T(u,X^T)\right\},\end{aligned}$$

where

$$\Delta_T(u,X^T)=\frac{1}{S_*\sqrt{T}}\int_0^T\int_0^{t-}u(t-s)\,\mathrm{d}X_s[\,\mathrm{d}X_t-S_*\,\mathrm{d}t],$$

$$\mathrm{I}(u)=\int_0^\infty u(t)^2\,\mathrm{d}t+S_*\left(\int_0^\infty u(t)\,\mathrm{d}t\right)^2=\int_0^\infty u(t)^2\,\mathrm{d}t+S_*r^2$$

and

$$\mathcal{L}_0\{\Delta_T\left(u,X^T\right)\}\Longrightarrow\mathcal{N}(0,\mathrm{I}(u)),\quad r_T(u,X^T)\to 0.$$

Moreover, these last two convergences are uniform on $u(\cdot)\in\mathscr{U}_{r,N}$, $0\le r\le K$ for any $K>0$. Hence the likelihood ratio test

$$\bar\phi_T(X^T)=\chi_{\{Z_T(u(\cdot))>d_\varepsilon\}},$$

with

$$d_\varepsilon=\exp\left\{z_\varepsilon\sqrt{\mathrm{I}(u)}-\frac{\mathrm{I}(u)}{2}\right\}$$

is the most powerful in the class $\mathscr{K}_\varepsilon$ for any two simple hypotheses and its power function

$$\beta(u, \bar{\phi}_T) \longrightarrow \mathbf{P}\left\{\zeta > z_\varepsilon - \mathrm{I}(u)^{1/2}\right\}, \quad \zeta \sim \mathcal{N}(0,1).$$

It is easy to see that

$$\inf_{u(\cdot)\in\mathscr{U}_{r,N}} \mathrm{I}(u) = S_* r^2 + \frac{r^2}{N}$$

because

$$r^2 = \left(\int_0^N u(t)\,\mathrm{d}t\right)^2 \le N \int_0^N u(t)^2\,\mathrm{d}t$$

with equality on the 'least favourable alternative'

$$u^*(t) = (r/N)\chi_{\{0\le t\le N\}}.$$

Hence

$$\inf_{u(\cdot)\in\mathscr{U}_{r,N}} \mathbf{P}\{\zeta > z_\varepsilon - \mathrm{I}(u)^{1/2}\} = \mathbf{P}\{\zeta > z_\varepsilon - r\sqrt{S_* + N^{-1}}\}.$$

Now we study the power function of the test $\hat{\phi}_T$. Let us denote

$$U_t = \int_0^{t-} u(t-s)\,\mathrm{d}X_s, \quad \pi_t = X_t - S_* t,$$

then

$$\Delta_T(u, X^T) = \frac{1}{S_*\sqrt{T}} \int_0^T U_t\,\mathrm{d}\pi_t, \quad \delta_T(X^T) = \frac{1}{\sqrt{S_* T}} \int_0^T \mathrm{d}\pi_t$$

and

$$\mathbf{E}_0\Delta_T(u, X^T) = 0, \quad \mathbf{E}_0\Delta_T(u, X^T)^2 = \mathrm{I}(u), \quad \mathbf{E}_0\delta_T(X^T) = 0,$$
$$\mathbf{E}_0\delta_T(X^T)^2 = 1, \quad \mathbf{E}_0(\delta_T(X^T)\Delta_T(u, X^T)) = r\sqrt{S_*}.$$

Hence, under hypothesis $\mathscr{H}_0$, we have

$$\mathcal{L}_0\{\Delta_T(u, X^T), \delta_T(X^T)\} \Longrightarrow \mathcal{N}(\mathbf{0}, \mathbf{R}),$$

where $\mathbf{R}$ is covariance matrix of the vector (Δ_T, δ_T) described above. Therefore $\hat{\phi}_T \in \mathscr{K}_\varepsilon$, and using Le Cam's Third Lemma (van der Vaart, 1998) we obtain that under alternative $u(\cdot) \in \mathscr{U}_r$

$$\delta_T(X^T) \Longrightarrow \mathcal{N}(r\sqrt{S_*}, 1).$$

For the power function we have

$$\beta(u, \hat{\phi}_T) \longrightarrow \mathbf{P}\{\zeta > z_\varepsilon - r\sqrt{S_*}\}.$$

It can be shown that this convergence is uniform over $u(\cdot) \in \mathscr{U}_{r,N}$, $0 \le r \le K$ for any $K > 0$ and this proves the theorem.

6. Simulations

The main results (theorems 1 and 2) of this work are 'asymptotic in nature' and it is interesting to see the properties of the tests for the moderate values of T. This can be performed by Monte-Carlo simulations.

6.1. Parametric alternative

To illustrate theorem 1 we take $S_*=1$ and $h(t)=1/2\,\mathrm{e}^{-t/2}$ (see example 1). This yields

$$S(u,t,\omega)=1+\frac{u}{2\sqrt{T}}\sum_{t_i\le t}\mathrm{e}^{-(t-t_i)/2},\quad u\ge 0,\ \ 0\le t\le T.$$

In this case

$$\Delta_T(X^T)=\frac{1}{\sqrt{5T}}\sum_{0\le t_j\le T}\sum_{t_i<t_j}\mathrm{e}^{-(t_j-t_i)/2}-\frac{2}{\sqrt{5T}}\left(X_T-\sum_{0\le t_j\le T}\mathrm{e}^{-(T-t_j)/2}\right),$$

where t_i are the events of the observed process, and the test $\hat{\phi}_T^{\varepsilon}$ given by

$$\hat{\phi}_T^{\varepsilon}=\hat{\phi}_T(X^T)=\chi_{\{\Delta_T(X^T)>z_\varepsilon\}},$$

is locally asymptotically uniformly most powerful in the class $\mathscr{K}_\varepsilon$.

In Fig. 1 we represent the size of the test $\hat{\phi}_T^{0.05}$ as a function of $T\in[0,1000]$. This size is given by

$$\alpha(T)=\mathbf{P}_0\{\Delta_T(X^T)>z_{0.05}\},\quad 1\le T\le 1000$$

and is obtained by simulating $M=10^7$ trajectories on $[0,T]$ of Poisson process of constant intensity $S(t,\omega)=1$ and calculating empirical frequency of accepting the alternative hypothesis.

In Fig. 2 we represent the power function of the test $\hat{\phi}_T^{0.05}$ given by

$$\beta_T(u,\hat{\phi}_T^{0.05})=\mathbf{P}_u\{\Delta_T(X^T)>z_{0.05}\},\quad 0\le u\le 5$$

for $T=100$, 300 and 1000, as well as the limiting (Gaussian) power function given by

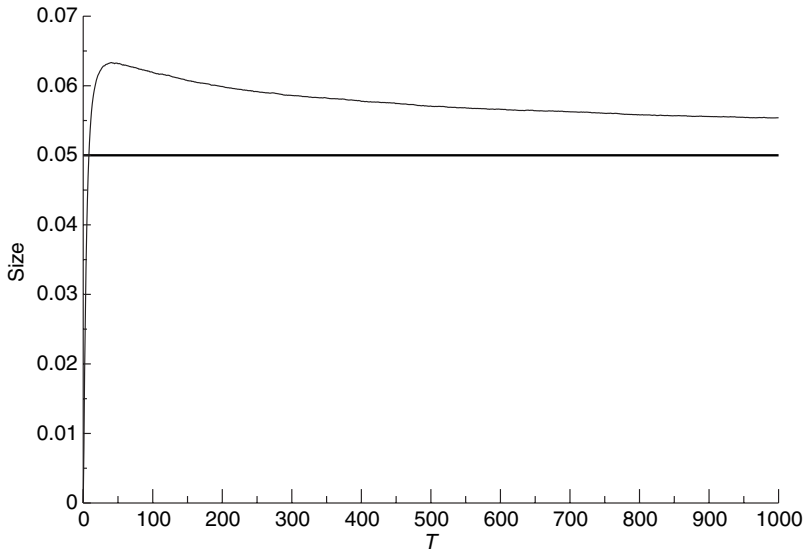

Fig. 1. Test size.

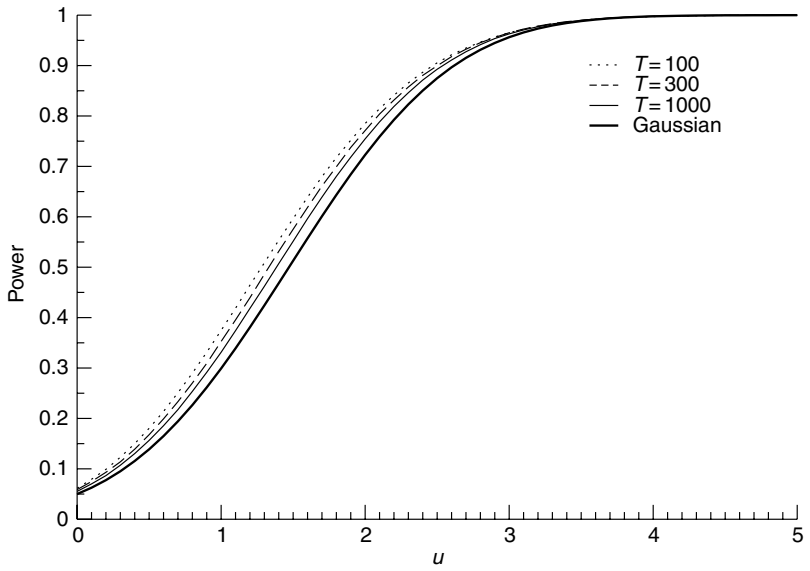

Fig. 2. Test power.

$$\hat{\beta}(u)=\mathbf{P}\{\zeta>z_{0.05}-u\sqrt{5}/2\}=\frac{1}{\sqrt{2\pi}}\int_{z_{0.05}-u\sqrt{5}/2}^{\infty}\mathrm{e}^{-v^2/2}\,\mathrm{d}v,\quad 0\le u\le 5.$$

The function β_T is obtained by simulating (for each value of u) $M=10^6$ trajectories on $[0,T]$ of self-exciting process of intensity $S(u,t,\omega)$ and calculating empirical frequency of accepting the alternative hypothesis.

Now let us consider the $\tilde{\phi}_T^{\varepsilon}$ given by

$$\tilde{\phi}_T^{\varepsilon}=\tilde{\phi}_T(X^T)=\chi_{\{\Delta_T(X^T)>z\}},$$

where the threshold z is chosen so that this test is of exact size ε. The choice of this threshold z as a function of $\varepsilon\in[0,0.25]$ is shown in Figs 3 and 4 for $T=100$, 300 and 1000, as well as the Gaussian threshold z_{ε}. The values of z are obtained by simulating $M=10^7$ trajectories on $[0,T]$ of Poisson process of constant intensity $S(t,\omega)=1$ and calculating empirical $1-\varepsilon$ quantiles of Δ_T.

For example to obtain test of exact size 0.05 one needs take $z\simeq 1.78$ for $T=100$ ($z\simeq 1.74$ for $T=300$, $z\simeq 1.70$ for $T=1000$) against $z_{\varepsilon}\simeq 1.64$ for Gaussian case.

6.2. Non-parametric alternative

To illustrate the non-parametric alternatives we take intensity functions corresponding to $S_*=1$ and $u(t)=(r/N)\chi_{\{0\le t\le N\}}$, i.e.

$$S(t,\omega)=1+\frac{r}{N\sqrt{T}}\sum_{t_i<t}\chi_{\{t-t_i\le N\}},\quad 0\le t\le T,$$

where t_i are the events of the observed process. This choice of $u(\cdot)$ allows us to compare the power function of our locally asymptotically uniformly most powerful test

$$\hat{\phi}_T^{\varepsilon}(X^T)=\chi_{\{X_T>z_{\varepsilon}\sqrt{T}+T\}},$$

with the asymptotic power

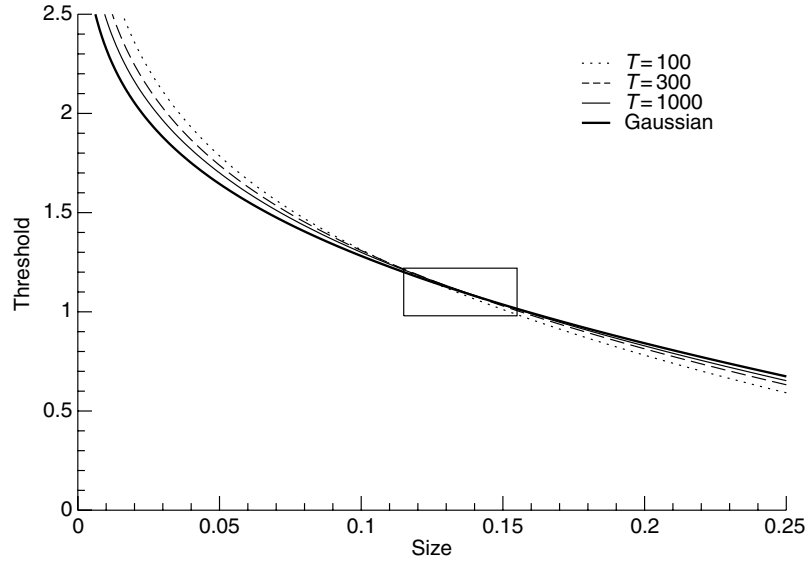

Fig. 3. Threshold choice.

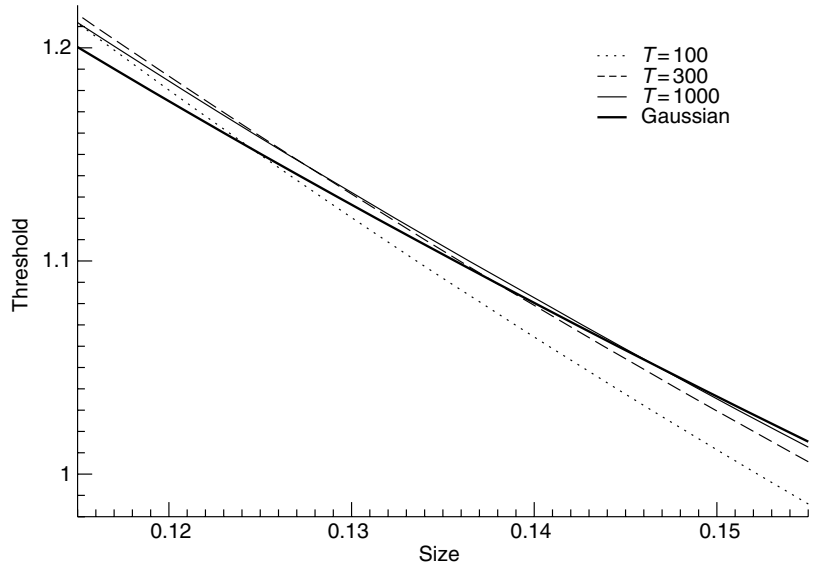

Fig. 4. Threshold choice (zoom).

$$\beta(r)=\frac{1}{\sqrt{2\pi}}\int_{z_\varepsilon-r}^{\infty}\mathrm{e}^{-v^2/2}\,\mathrm{d}v,$$

of Neyman–Pearson test for the least favourable alternatives.

Note that under $\mathscr{H}_0$, X_T is Poisson random variable with parameter T, therefore the size of the test $\hat{\phi}_T^{\varepsilon}$, as well as the threshold giving a test of exact size ε, can be calculated directly (without resort to Monte-Carlo simulations).

We represent the power function of the test $\hat{\phi}_T^{0.05}$ given by

$$\beta_T(u,\hat{\phi}_T)=\mathbf{P}_u\{X_T>z_{0.05}\sqrt{T}+T\},\quad 0\le r\le 5,$$

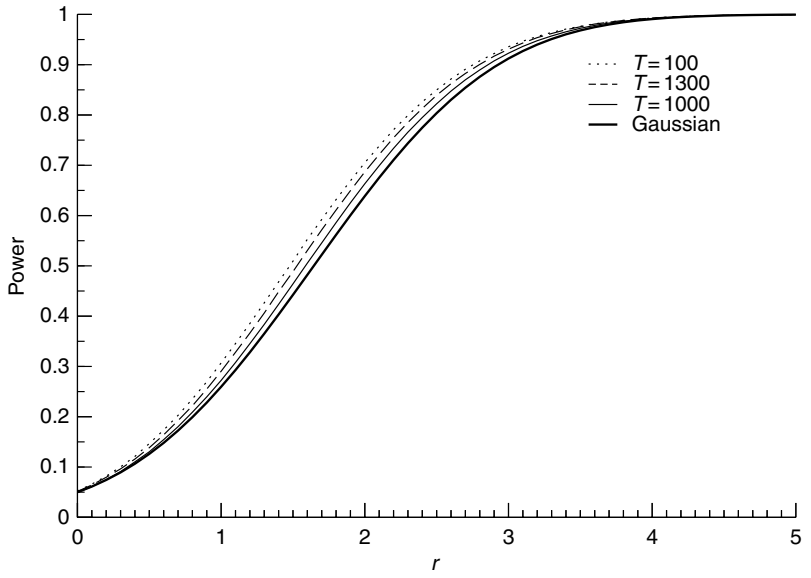

Fig. 5. Test power ($N=5$).

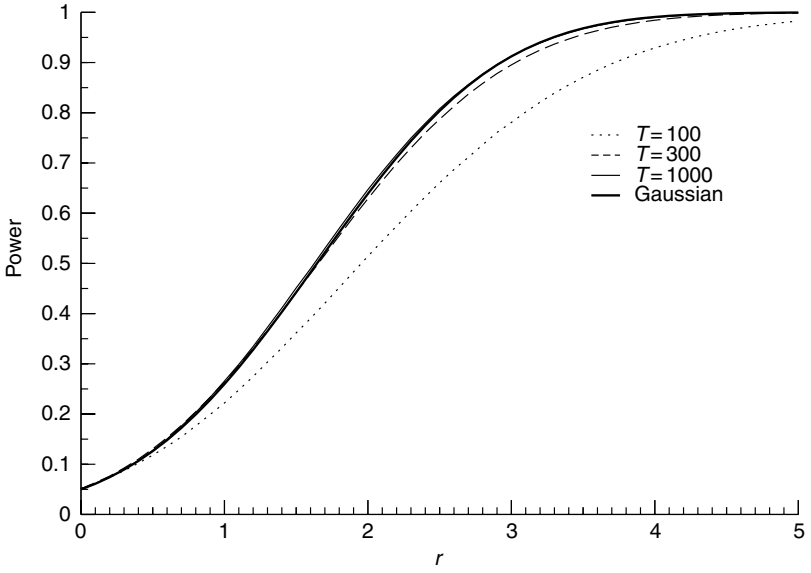

Fig. 6. Test power ($N=50$).

for $T=100$, 300 and 1000 as well as the limiting (Gaussian) function $\beta(r)$, $0\leq r\leq 5$. In Figs 5 and 6 we take $N=5$ and $N=50$ respectively. The function β_T is obtained by simulating (for each value of r and N) $M=10^6$ trajectories on $[0,T]$ of self-exciting process of intensity $S(t,\omega)$ and calculating empirical frequency of accepting the alternative hypothesis.

We see that if $1\ll N\ll T$, then the power function converge to the limiting function (e.g. if $N=50$ and $T=1000$, the power function almost coincides with the limiting one). If N and T are of the same order (e.g. if $N=50$ and $T=100$) then the power function of the test can be essentially smaller. This example confirms the importance of use of functions with bounded support and of the order of limits in definition 2.

7. Discussions

The constructed tests are asymptotically optimal for parametric (section 3) and non-parametric (section 5) alternatives. It seems that these are just the first results in this field and it is interesting to develop the construction of the asymptotically optimal tests for wider classes of alternatives. Particularly, it is interesting to study 'smooth alternatives' like

$$\mathcal{H}_1: \qquad \int_0^\infty u^{(k)}(t)^2 \, \mathrm{d}t > r,$$

where $r > 0$. Note that the test $\hat{\phi}_T$ is no more uniformly consistent in this situation.

Acknowledgements

We are grateful to an anonymous referee and the associate editor for useful comments.

References

Brillinger, R. (1975). Statistical inference for stationary point processes. In *Stochastic processes and related topics, 1* (ed. M. L. Puri), 55–99. Academic Press, New York, NY.

Capon, J. (1961). On asymptotic efficiency of locally optimum detectors, *IRE Trans. Inf. Theory* **7**, 67–71.

Cox, D. R. & Lewis, P. A. W. (1966). *Statistical analysis of a series of events*. Methuen, London.

Daley, D. J. & Vere-Jones, D. (2003). *An introduction to the theory of point processes*, Vol. I, 2nd edn. Springer-Verlag, New York, NY.

Davies, R. (1977). Testing the hypothesis that a point process is Poisson. *Adv. Appl. Probab.* **9**, 724–746.

Hawkes, A. G. (1972). Spectra for some mutually exciting point processes with associated variable. In *Stochastic point processes* (ed. P. A. W. Lewis), 261–271. Wiley, New York, NY.

Iacus, S. & Kutoyants, Yu. A. (2001). Semiparametric hypotheses testing for dynamical systems with small noise. *Math. Methods Statist.* **10**, 105–120.

Kutoyants, Y. A. (1976). On asymptotic theory of signal detection. *Radiotekhnika i Elektronika* **21**, 1458–1466.

Kutoyants, Y. A. (1984). *Parameter estimation for stochastic processes*. Heldermann, Berlin.

Kutoyants, Y. A. (1998). *Statistical inference for spatial Poisson processes*. Springer-Verlag, New York, NY.

Kutoyants, Y. A. (2003). *Statistical inference for ergodic diffusion processes*. Springer-Verlag, New York, NY.

Le Cam, L. (1956). Locally asymptotically normal families of distributions. *Univ. Calif. Publ. Statist.* **3**, 27–98.

Liptser, R. S. & Shiryayev, A. N. (2001). *Statistics of random processes. II. Applications*, 2nd edn. Springer-Verlag, New York, NY.

Pham, D. T. (1981). Estimation of the spectral parameters of a stationary point process. *Ann. Statist.* **9**, 615–627.

Pitman, E. J. G. (1948). *Lecture notes on nonparametric statistical inference*. Lectures given for the University of North Carolina, Institute of Statistics, NC.

Roussas, G. G. (1972). *Contiguity of probability measures*. Cambridge University Press, Cambridge.

van der Vaart, A. W. (1998). *Asymptotic statistics*. Cambridge University Press, Cambridge.

Received October 2004, in final form August 2005

Yury A. Kutoyants, Laboratoire de Statistique et Processus, Université du Maine, 72085 Le Mans Cedex 9, France.
E-mail: kutoyants@univ-lemans.fr

27

On the Goodness-of-Fit Tests for Some Continuous Time Processes

Sergueï Dachian[1] **and Yury A. Kutoyants**[2]

[1] *Laboratoire de Mathématiques, Université Blaise Pascal, Aubière, France*
[2] *Laboratoire de Statistique et Processus, Université du Maine, Le Mans, France*

Abstract: We present a review of several results concerning the construction of the Cramér–von Mises and Kolmogorov–Smirnov type goodness-of-fit tests for continuous time processes. As the models we take a stochastic differential equation with small noise, ergodic diffusion process, Poisson process, and self-exciting point processes. For every model we propose the tests which provide the asymptotic size α and discuss the behaviour of the power function under local alternatives. The results of numerical simulations of the tests are presented.

Keywords and Phrases: Hypotheses testing, diffusion process, Poison process, self-exciting process, goodness-of-fit tests

27.1 Introduction

The goodness-of-fit tests play an important role in classical mathematical statistics. Particularly, the tests of Cramér–von Mises, Kolmogorov–Smirnov, and chi-squared are well studied and allow us to verify the correspondence of the mathematical models to the observed data [see, e.g., Durbin (1973) or Greenwood and Nikulin (1996)]. A similar problem, of course, exists for the continuous-time stochastic processes. The diffusion and Poisson processes are widely used as mathematical models of many evolution processes in biology, medicine, physics, financial mathematics, and in many other fields. For example, some theory can propose a diffusion process

$$\mathrm{d}X_t = S_*\left(X_t\right)\,\mathrm{d}t + \sigma\,\mathrm{d}W_t, \qquad X_0, \qquad 0 \le t \le T$$

as an appropriate model for description of the real data $\{X_t, 0 \le t \le T\}$ and we can try to construct an algorithm to verify if this model corresponds well to these data. The model here is totally defined by the trend coefficient $S_*(\cdot)$,

which is supposed (if the theory is true) to be known. We do not discuss here the problem of verification if the process $\{W_t, 0 \le t \le T\}$ is Wiener. This problem is much more complicated and we suppose that the *noise is white Gaussian*. Therefore we have a basic hypothesis defined by the trend coefficient $S_*(\cdot)$ and we have to test this hypothesis against any other alternative. Any other means that the observations come from stochastic differential equation

$$\mathrm{d}X_t = S(X_t)\,\mathrm{d}t + \sigma\,\mathrm{d}W_t, \qquad X_0, \qquad 0 \le t \le T,$$

where $S(\cdot) \ne S_*(\cdot)$. We propose some tests which are in some sense similar to the Cramér–von Mises and Kolmogorov–Smirnov tests. The advantage of classical tests is that they are distribution-free; that is, the distribution of the underlying statistics does not depend on the basic model and this property allows us to choose the *universal thresholds* which can be used for all models.

For example, if we observe n independent identically distributed random variables $(X_1, \dots, X_n) = X^n$ with distribution function $F(x)$ and the basic hypothesis is simple, $F(x) \equiv F_*(x)$, then the Cramér–von Mises W_n^2 and Kolmogorov–Smirnov D_n statistics are

$$W_n^2 = n\int_{-\infty}^{\infty}\left[\hat{F}_n(x) - F_*(x)\right]^2 \mathrm{d}F_*(x), \qquad D_n = \sup_x\left|\hat{F}_n(x) - F_*(x)\right|,$$

respectively. Here

$$\hat{F}_n(x) = \frac{1}{n}\sum_{j=1}^{n} 1_{\{X_j < x\}}$$

is the empirical distribution function. Let us denote by $\{W_0(s), 0 \le s \le 1\}$ a Brownian bridge, that is, a continuous Gaussian process with

$$\mathbf{E}W_0(s) = 0, \qquad \mathbf{E}W_0(s)\,W_0(t) = t \wedge s - st.$$

Then the limit behaviour of these statistics can be described with the help of this process as follows.

$$W_n^2 \Longrightarrow \int_0^1 W_0(s)^2\,\mathrm{d}s, \qquad \sqrt{n}D_n \Longrightarrow \sup_{0\le s\le 1}|W_0(s)|.$$

Hence the corresponding Cramér–von Mises and Kolmogorov–Smirnov tests

$$\psi_n(X^n) = 1_{\{W_n^2 > c_\alpha\}}, \qquad \phi_n(X^n) = 1_{\left\{\sqrt{n}D_n > d_\alpha\right\}}$$

with constants c_α, d_α defined by the equations

$$\mathbf{P}\left\{\int_0^1 W_0(s)^2\,\mathrm{d}s > c_\alpha\right\} = \alpha, \qquad \mathbf{P}\left\{\sup_{0\le s\le 1}|W_0(s)| > d_\alpha\right\} = \alpha$$

are of asymptotic size α. It is easy to see that these tests are distribution-free [the limit distributions do not depend of the function $F_*(\cdot)$] and are consistent against any fixed alternative [see, e.g., Durbin (1973)].

It is interesting to study these tests for a *nondegenerate set of alternatives*, that is, for alternatives with limit power function less than 1. It can be realized on the close nonparametric alternatives of the special form making this problem asymptotically equivalent to the *signal in Gaussian noise* problem. Let us put

$$F(x) = F_*(x) + \frac{1}{\sqrt{n}} \int_{-\infty}^{x} h(F_*(y)) \, \mathrm{d}F_*(y),$$

where the function $h(\cdot)$ describes the alternatives. We suppose that

$$\int_0^1 h(s) \, \mathrm{d}s = 0, \qquad \int_0^1 h(s)^2 \, \mathrm{d}s < \infty.$$

Then we have the following convergence [under a fixed alternative, given by the function $h(\cdot)$],

$$W_n^2 \Longrightarrow \int_0^1 \left[\int_0^s h(v) \, \mathrm{d}v + W_0(s) \right]^2 \mathrm{d}s,$$
$$\sqrt{n} D_n \Longrightarrow \sup_{0 \le s \le 1} \left| \int_0^s h(v) \, \mathrm{d}v + W_0(s) \right|.$$

We see that this problem is asymptotically equivalent to the following *signal in Gaussian noise* problem,

$$\mathrm{d}Y_s = h_*(s) \, \mathrm{d}s + \mathrm{d}W_0(s), \qquad 0 \le s \le 1. \tag{27.1}$$

Indeed, if we use the statistics

$$W^2 = \int_0^1 Y_s^2 \, \mathrm{d}s, \qquad D = \sup_{0 \le s \le 1} |Y_s|$$

then under hypothesis $h(\cdot) \equiv 0$ and alternative $h(\cdot) \neq 0$ the distributions of these statistics coincide with the limit distributions of W_n^2 and $\sqrt{n} D_n$ under the hypothesis and alternative, respectively.

Our goal is to see how such kinds of tests can be constructed in the case of continuous-time models of observation and particularly in the cases of some diffusion and point processes. We consider the diffusion processes with small noise, ergodic diffusion processes, and Poisson processes with Poisson and self-exciting alternatives. For the first two classes we just show how Cramér–von Mises and Kolmogorov–Smirnov type tests can be realized using some known results and for the last models we discuss this problem in detail.

27.2 Diffusion Process with Small Noise

Suppose that the observed process is the solution of the stochastic differential equation

$$\mathrm{d}X_t = S\left(X_t\right)\,\mathrm{d}t + \varepsilon\,\mathrm{d}W_t, \qquad X_0 = x_0, \qquad 0 \le t \le T, \tag{27.2}$$

where $W_t, 0 \le t \le T$ is a Wiener process [see, e.g., Liptser and Shiryayev (2001)]. We assume that the function $S\left(x\right)$ is two times continuously differentiable with bounded derivatives. These are not the minimal conditions for the results presented below, but this assumption simplifies the exposition. We are interested in the statistical inference for this model in the asymptotics of small noise: $\varepsilon \to 0$. The statistical estimation theory (parametric and nonparametric) was developed in Kutoyants (1994).

Recall that the stochastic process $X^\varepsilon = \{X_t, 0 \le t \le T\}$ converges uniformly in $t \in [0,T]$ to the deterministic function $\{x_t, 0 \le t \le T\}$, which is a solution of the ordinary differential equation

$$\frac{\mathrm{d}x_t}{\mathrm{d}t} = S(x_t), \qquad x_0, \qquad 0 \le t \le T. \tag{27.3}$$

Suppose that the function $S_*\left(x\right) > 0$ for $x \ge x_0$ and consider the following problem of hypotheses testing,

$$\begin{aligned} &\mathcal{H}_0 : S\left(x\right) = S_*(x), \qquad x_0 \le x \le x_T^*\\ &\mathcal{H}_1 : S\left(x\right) \ne S_*(x), \qquad x_0 \le x \le x_T^*, \end{aligned}$$

where we denoted by x_t^* the solution of equation (27.3) under hypothesis $\mathcal{H}_0$:

$$x_t^* = x_0 + \int_0^t S_*\left(x_v^*\right)\,\mathrm{d}v, \qquad 0 \le t \le T.$$

Hence, we have a simple hypothesis against the composite alternative.

The Cramér–von Mises $\left(W_\varepsilon^2\right)$ and Kolmogorov–Smirnov (D_ε) type statistics for this model of observations can be

$$W_\varepsilon^2 = \left[\int_0^T \frac{\mathrm{d}t}{S_*\left(x_t^*\right)^2}\right]^{-2} \int_0^T \left(\frac{X_t - x_t^*}{\varepsilon\, S_*\left(x_t^*\right)^2}\right)^2 \mathrm{d}t,$$

$$D_\varepsilon = \left[\int_0^T \frac{\mathrm{d}t}{S_*\left(x_t^*\right)^2}\right]^{-1/2} \sup_{0\le t\le T}\left|\frac{X_t - x_t^*}{S_*\left(x_t^*\right)}\right|.$$

It can be shown that these two statistics converge (as $\varepsilon \to 0$) to the following functionals,

$$W_\varepsilon^2 \Longrightarrow \int_0^1 W\left(s\right)^2\,\mathrm{d}s, \qquad \varepsilon^{-1} D_\varepsilon \Longrightarrow \sup_{0\le s\le 1}\left|W\left(s\right)\right|,$$

where $\{W(s), 0 \leq s \leq 1\}$ is a Wiener process [see Kutoyants (1994)]. Hence the corresponding tests

$$\psi_\varepsilon(X^\varepsilon) = 1_{\{W_\varepsilon^2 > c_\alpha\}}, \qquad \phi_\varepsilon(X^\varepsilon) = 1_{\{\varepsilon^{-1} D_\varepsilon > d_\alpha\}}$$

with the constants c_α, d_α defined by the equations

$$\mathbf{P}\left\{\int_0^1 W(s)^2 \, \mathrm{d}s > c_\alpha\right\} = \alpha, \qquad \mathbf{P}\left\{\sup_{0\leq s\leq 1} |W(s)| > d_\alpha\right\} = \alpha \tag{27.4}$$

are of asymptotic size α. Note that the choice of the thresholds c_α and d_α does not depend on the hypothesis (distribution-free). This situation is quite close to the classical case mentioned above.

It is easy to see that if $S(x) \neq S_*(x)$, then $\sup_{0\leq t\leq T} |x_t - x_t^*| > 0$ and $W_\varepsilon^2 \to \infty$, $\varepsilon^{-1} D_\varepsilon \to \infty$. Hence these tests are consistent against any fixed alternative. It is possible to study the power function of this test for local (contiguous) alternatives of the following form,

$$\mathrm{d}X_t = S_*(X_t)\, \mathrm{d}t + \varepsilon \, \frac{h(X_t)}{S_*(X_t)} \, \mathrm{d}t + \varepsilon \, \mathrm{d}W_t, \qquad 0 \leq t \leq T.$$

We describe the alternatives with the help of the (unknown) function $h(\cdot)$. The case $h(\cdot) \equiv 0$ corresponds to the hypothesis $\mathcal{H}_0$. One special class of such nonparametric alternatives for this model was studied in Iacus and Kutoyants (2001).

Let us introduce the composite (nonparametric) alternative

$$\mathcal{H}_1 : h(\cdot) \in \mathcal{H}_\rho,$$

where

$$\mathcal{H}_\rho = \left\{h(\cdot) : \int_{x_0}^{x_T} h(x)^2 \, \mu(\mathrm{d}x) \geq \rho\right\}.$$

To choose the alternative we have to make precise the "natural for this problem" distance described by the measure $\mu(\cdot)$ and the rate of $\rho = \rho_\varepsilon$. We show that the choice

$$\mu(\mathrm{d}x) = \frac{\mathrm{d}x}{S_*(x)^3}$$

provides for the test statistic the following limit,

$$W_\varepsilon^2 \longrightarrow \int_0^1 \left[\int_0^s h_*(v)\, \mathrm{d}v + W(s)\right]^2 \mathrm{d}s,$$

where we denoted

$$h_*(s) = u_T^{1/2} h\left(x^*_{u_T s}\right), \qquad u_T = \int_0^T \frac{\mathrm{d}s}{S_*(x_s^*)^2}.$$

We see that this problem is asymptotically equivalent to the *signal in white Gaussian noise* problem:

$$\mathrm{d}Y_s = h_* (s)\, \mathrm{d}s + \mathrm{d}W (s), \qquad 0 \le s \le 1, \tag{27.5}$$

with the Wiener process $W(\cdot)$. It is easy to see that even for fixed $\rho > 0$ without further restrictions on the smoothness of the function $h_*(\cdot)$, *uniformly good* testing is impossible. For example, if we put

$$h_n (x) = c\, S_* (x)^3 \cos [n (x - x_0)]$$

then for the power function of the test we have

$$\inf_{h(\cdot) \in \mathcal{H}_\rho} \beta (\psi_\varepsilon, h) \le \beta (\psi_\varepsilon, h_n) \longrightarrow \alpha.$$

The details can be found in Kutoyants (2006). The construction of the uniformly consistent tests requires a different approach [see Ingster and Suslina (2003)].

Note as well that if the diffusion process is

$$\mathrm{d}X_t = S (X_t)\, \mathrm{d}t + \varepsilon \sigma (X_t)\, \mathrm{d}W_t, \qquad X_0 = x_0, \qquad 0 \le t \le T,$$

then we can put

$$W_\varepsilon^2 = \left[\int_0^T \left(\frac{\sigma (x_t^*)}{S_* (x_t^*)} \right)^2 \mathrm{d}t \right]^{-2} \int_0^T \left(\frac{X_t - x_t^*}{\varepsilon\, S_* (x_t^*)^2} \right)^2 \mathrm{d}t$$

and have the same results as above [see Kutoyants (2006)].

27.3 Ergodic Diffusion Processes

Suppose that the observed process is the one-dimensional diffusion process

$$\mathrm{d}X_t = S (X_t)\, \mathrm{d}t + \mathrm{d}W_t, \qquad X_0, \qquad 0 \le t \le T, \tag{27.6}$$

where the trend coefficient $S(x)$ satisfies the conditions of the existence and uniqueness of the solution of this equation and this solution has ergodic properties; that is, there exists an invariant probability distribution $F_S(x)$, and for any integrable w.r.t. this distribution function $g(x)$ the law of large numbers holds

$$\frac{1}{T} \int_0^T g (X_t)\, \mathrm{d}t \longrightarrow \int_{-\infty}^{\infty} g (x)\, \mathrm{d}F_S (x).$$

These conditions can be found, for example, in Kutoyants (2004).

Recall that the invariant density function $f_S(x)$ is defined by the equality

$$f_S(x) = G(S)^{-1} \exp\left\{2\int_0^x S(y)\ \mathrm{d}y\right\},$$

where $G(S)$ is the normalising constant.

We consider two types of tests. The first one is a direct analogue of the classical Cramér–von Mises and Kolmogorov–Smirnov tests based on empirical distribution and density functions and the second follows the considered-above (small noise) construction of tests.

The invariant distribution function $F_S(x)$ and this density function can be estimated by the *empirical distribution function* $\hat{F}_T(x)$ and by the *local time type* estimator $\hat{f}_T(x)$ defined by the equalities

$$\hat{F}_T(x) = \frac{1}{T}\int_0^T 1_{\{X_t<x\}}\,\mathrm{d}t, \qquad \hat{f}_T(x) = \frac{2}{T}\int_0^T 1_{\{X_t<x\}}\,\mathrm{d}X_t,$$

respectively. Note that both of them are unbiased,

$$\mathbf{E}_S\hat{F}_T(x) = F_S(x), \qquad \mathbf{E}_S\hat{f}_T(x) = f_S(x),$$

admit the representations

$$\eta_T(x) = -\frac{2}{\sqrt{T}}\int_0^T \frac{F_S(X_t\wedge x) - F_S(X_t)\,F_S(x)}{f_S(X_t)}\,\mathrm{d}W_t + o(1),$$
$$\zeta_T(x) = -\frac{2f_S(x)}{\sqrt{T}}\int_0^T \frac{1_{\{X_t>x\}} - F_S(X_t)}{f_S(X_t)}\,\mathrm{d}W_t + o(1),$$

and are $\sqrt{T}$ asymptotically normal (as $T\to\infty$)

$$\eta_T(x) = \sqrt{T}\left(\hat{F}_T(x) - F_S(x)\right) \Longrightarrow \mathcal{N}\left(0, d_F(S,x)^2\right),$$
$$\zeta_T(x) = \sqrt{T}\left(\hat{f}_T(x) - f_S(x)\right) \Longrightarrow \mathcal{N}\left(0, d_f(S,x)^2\right).$$

Let us fix a simple (basic) hypothesis

$$\mathcal{H}_0 : S(x) \equiv S_*(x).$$

Then to test this hypothesis we can use these estimators for construction of the Cramér–von Mises and Kolmogorov–Smirnov type test statistics

$$W_T^2 = T\int_{-\infty}^{\infty}\left[\hat{F}_T(x) - F_{S_*}(x)\right]^2\,\mathrm{d}F_{S_*}(x),$$
$$D_T = \sup_x\left|\hat{F}_T(x) - F_{S_*}(x)\right|,$$

and

$$V_T^2 = T\int_{-\infty}^{\infty}\left[\hat{f}_T(x) - f_{S_*}(x)\right]^2 \mathrm{d}F_{S_*}(x),$$
$$d_T = \sup_x\left|\hat{f}_T(x) - f_{S_*}(x)\right|,$$

respectively. Unfortunately, all these statistics are not distribution-free even asymptotically and the choice of the corresponding thresholds for the tests is much more complicated. Indeed, it was shown that the random functions $(\eta_T(x), x \in R)$ and $(\zeta_T(x), x \in R)$ converge in the space $(\mathcal{C}_0, \mathfrak{B})$ (of continuous functions decreasing to zero at infinity) to the zero mean Gaussian processes $(\eta(x), x \in R)$ and $(\zeta(x), x \in R)$, respectively, with the covariance functions [we omit the index S_* of functions $f_{S_*}(x)$ and $F_{S_*}(x)$ below]:

$$R_F(x,y) = \mathbf{E}_{S_*}[\eta(x)\eta(y)]$$
$$= 4\mathbf{E}_{S_*}\left(\frac{[F(\xi\wedge x) - F(\xi)F(x)][F(\xi\wedge y) - F(\xi)F(y)]}{f(\xi)^2}\right)$$
$$R_f(x,y) = \mathbf{E}_{S_*}[\zeta(x)\zeta(y)]$$
$$= 4f(x)f(y)\,\mathbf{E}_{S_*}\left(\frac{\left[1_{\{\xi>x\}} - F(\xi)\right]\left[1_{\{\xi>y\}} - F(\xi)\right]}{f(\xi)^2}\right).$$

Here ξ is a random variable with the distribution function $F_{S_*}(x)$. Of course,

$$d_F(S,x)^2 = \mathbf{E}_S\left[\eta(x)^2\right], \qquad d_f(S,x)^2 = \mathbf{E}_S\left[\zeta(x)^2\right].$$

Using this weak convergence it is shown that these statistics converge in distribution (under hypothesis) to the following limits (as $T \to \infty$),

$$W_T^2 \Longrightarrow \int_{-\infty}^{\infty}\eta(x)^2\,\mathrm{d}F_{S_*}(x), \qquad T^{1/2}D_T \Longrightarrow \sup_x|\eta(x)|,$$
$$V_T^2 \Longrightarrow \int_{-\infty}^{\infty}\zeta(x)^2\,\mathrm{d}F_{S_*}(x), \qquad T^{1/2}d_T \Longrightarrow \sup_x|\zeta(x)|.$$

The conditions and the proofs of all these properties can be found in Kutoyants (2004), where essentially different statistical problems were studied, but the calculus is quite close to what we need here.

Note that the Kolmogorov–Smirnov test for ergodic diffusion was studied in Fournie (1992) [see as well Fournie and Kutoyants (1993) for further details], and the weak convergence of the process $\eta_T(\cdot)$ was obtained in Negri (1998).

The Cramér–von Mises and Kolmogorov–Smirnov type tests based on these statistics are

$$\Psi_T\left(X^T\right) = 1_{\left\{W_T^2 > C_\alpha\right\}}, \qquad \Phi_T\left(X^T\right) = 1_{\left\{T^{1/2}D_T > D_\alpha\right\}},$$
$$\psi_T\left(X^T\right) = 1_{\left\{V_T^2 > c_\alpha\right\}}, \qquad \phi_T\left(X^T\right) = 1_{\left\{T^{1/2}d_T > d_\alpha\right\}}$$

with appropriate constants.

The contiguous alternatives can be introduced in the following way,

$$S(x) = S_*(x) + \frac{h(x)}{\sqrt{T}}.$$

Then we obtain for the Cramér–von Mises statistics the limits [see Kutoyants (2004)]:

$$W_T^2 \Longrightarrow \int_{-\infty}^{\infty} \left[2\mathbf{E}_{S_*} \left(\left[1_{\{\xi<x\}} - F_{S_*}(x) \right] \int_0^{\xi} h(s)\ \mathrm{d}s \right) + \eta(x) \right]^2 \mathrm{d}F_{S_*}(x),$$

$$V_T^2 \Longrightarrow \int_{-\infty}^{\infty} \left[2f_{S_*}(x)\mathbf{E}_{S_*} \int_{\xi}^{x} h(s)\ \mathrm{d}s + \zeta(x) \right]^2 \mathrm{d}F_{S_*}(x).$$

Note that the transformation $Y_t = F_{S_*}(X_t)$ simplifies the writing, because the diffusion process Y_t satisfies the differential equation

$$\mathrm{d}Y_t = f_{S_*}(X_t)\left[2S_*(X_t)\,\mathrm{d}t + \mathrm{d}W_t \right], \qquad Y_0 = F_{S_*}(X_0)$$

with reflecting bounds in 0 and 1 and (under hypothesis) has uniform on $[0,1]$ invariant distribution. Therefore,

$$W_T^2 \Longrightarrow \int_0^1 V(s)^2\,\mathrm{d}s, \qquad T^{1/2}D_T \Longrightarrow \sup_{0\le s\le 1} |V(s)|,$$

but the covariance structure of the Gaussian process $\{V(s), 0 \le s \le 1\}$ can be quite complicated.

To obtain an asymptotically distribution-free Cramér–von Mises type test we can use another statistic, which is similar to that of the preceding section. Let us introduce

$$\tilde{W}_T^2 = \frac{1}{T^2}\int_0^T \left[X_t - X_0 - \int_0^t S_*(X_v)\ \mathrm{d}v \right]^2 \mathrm{d}t.$$

Then we have immediately (under hypothesis)

$$\tilde{W}_T^2 = \frac{1}{T^2}\int_0^T W_t^2\,\mathrm{d}t = \int_0^1 W(s)^2\ \mathrm{d}s,$$

where we put $t = sT$ and $W(s) = T^{-1/2}W_{sT}$. Under the alternative we have

$$\begin{aligned}\tilde{W}_T^2 &= \frac{1}{T^2}\int_0^T \left[W_t + \frac{1}{\sqrt{T}}\int_0^t h(X_v)\,\mathrm{d}v \right]^2 \mathrm{d}t \\ &= \frac{1}{T}\int_0^T \left[\frac{W_t}{\sqrt{T}} + \frac{t}{T}\,\frac{1}{t}\int_0^t h(X_v)\,\mathrm{d}v \right]^2 \mathrm{d}t.\end{aligned}$$

The stochastic process X_t is ergodic, hence

$$\frac{1}{t}\int_0^t h\left(X_v\right)\mathrm{d}v \longrightarrow \mathbf{E}_{S_*} h\left(\xi\right) = \int_{-\infty}^{\infty} h\left(x\right) f_{S_*}\left(x\right)\mathrm{d}x \equiv \rho_h$$

as $t \to \infty$. It can be shown [see Section 2.3 in Kutoyants (2004), where we have the similar calculus in another problem] that

$$\tilde{W}_T^2 \Longrightarrow \int_0^1 \left[\rho_h\, s + W\left(s\right)\right]^2 \mathrm{d}s.$$

Therefore the power function of the test $\psi\left(X^T\right) = 1_{\left\{\tilde{W}_T^2 > c_\alpha\right\}}$ converges to the function

$$\beta_\psi\left(\rho_h\right) = \mathbf{P}\left(\int_0^1 \left[\rho_h\, s + W\left(s\right)\right]^2 \mathrm{d}s > c_\alpha\right).$$

Using standard calculus we can show that for the corresponding Kolmogorov–Smirnov type test the limit will be

$$\beta_\phi\left(\rho_h\right) = \mathbf{P}\left(\sup_{0\le s\le 1}\left|\rho_h\, s + W\left(s\right)\right| > c_\alpha\right).$$

These two limit power functions are the same as in the next section devoted to self-exciting alternatives of the Poisson process. We calculate these functions with the help of simulations in Section 27.5 below.

Note that if the diffusion process is

$$\mathrm{d}X_t = S\left(X_t\right)\,\mathrm{d}t + \sigma\left(X_t\right)\,\mathrm{d}W_t, \qquad X_0, \qquad 0 \le t \le T,$$

but the functions $S\left(\cdot\right)$ and $\sigma\left(\cdot\right)$ are such that the process is ergodic then we introduce the statistics

$$\hat{W}_T^2 = \frac{1}{T^2\,\mathbf{E}_{S_*}\left[\sigma\left(\xi\right)^2\right]}\int_0^T \left[X_t - X_0 - \int_0^t S_*\left(X_v\right)\,\mathrm{d}v\right]^2 \mathrm{d}t.$$

Here ξ is a random variable with the invariant density function

$$f_{S_*}\left(x\right) = \frac{1}{G\left(S_*\right)\sigma\left(x\right)^2}\,\exp\left\{2\int_0^x \frac{S_*\left(y\right)}{\sigma\left(y\right)^2}\,\mathrm{d}y\right\}.$$

This statistic under hypothesis is equal to

$$\begin{aligned}\hat{W}_T^2 &= \frac{1}{T^2\,\mathbf{E}_{S_*}\left[\sigma\left(\xi\right)^2\right]}\int_0^T \left[\int_0^t \sigma\left(X_v\right)\mathrm{d}W_v\right]^2 \mathrm{d}t \\ &= \frac{1}{T\,\mathbf{E}_{S_*}\left[\sigma\left(\xi\right)^2\right]}\int_0^T \left[\frac{1}{\sqrt{T}}\int_0^t \sigma\left(X_v\right)\mathrm{d}W_v\right]^2 \mathrm{d}t.\end{aligned}$$

The stochastic integral by the central limit theorem is asymptotically normal

$$\eta_t = \frac{1}{\sqrt{t\mathbf{E}_{S_*}\left[\sigma\left(\xi\right)^2\right]}} \int_0^t \sigma\left(X_v\right) \mathrm{d}W_v \Longrightarrow \mathcal{N}\left(0,1\right)$$

and moreover it can be shown that the vector of such integrals converges in distribution to the Wiener process

$$\left(\eta_{s_1 T}, \ldots, \eta_{s_k T}\right) \Longrightarrow \left(W\left(s_1\right), \ldots, W\left(s_k\right)\right)$$

for any finite collection of $0 \leq s_1 < s_2 < \cdots < s_k \leq 1$. Therefore, under mild regularity conditions it can be proved that

$$\hat{W}_T^2 \Longrightarrow \int_0^1 W\left(s\right)^2 \, \mathrm{d}s.$$

The power function has the same limit,

$$\beta_\psi\left(\rho_h\right) = \mathbf{P}\left(\int_0^1 \left[\rho_h \, s + W\left(s\right)\right]^2 \mathrm{d}s > c_\alpha\right).$$

but with

$$\rho_h = \frac{\mathbf{E}_{S_*} h\left(\xi\right)}{\sqrt{\mathbf{E}_{S_*}\left[\sigma\left(\xi\right)^2\right]}}.$$

Similar consideration can be done for the Kolmogorov–Smirnov type test too.

We see that both tests cannot distinguish the alternatives with $h\left(\cdot\right)$ such that $\mathbf{E}_{S_*} h\left(\xi\right) = 0$. Note that for ergodic processes usually we have $\mathbf{E}_S S\left(\xi\right) = 0$ and $\mathbf{E}_{S_* + h/\sqrt{T}}\left[S_*\left(\xi\right) + T^{-1/2} h\left(\xi\right)\right] = 0$ with corresponding random variables ξ, but this does not imply $\mathbf{E}_{S_*} h\left(\xi\right) = 0$.

27.4 Poisson and Self-Exciting Processes

The Poisson process is one of the simplest point processes and before taking any other model it is useful first of all to check the hypothesis that the observed sequence of events, say, $0 < t_1, \ldots, t_N < T$ corresponds to a Poisson process. It is natural in many problems to suppose that this Poisson process is periodic of known period, for example, many daily events, signal transmission in optical communication, season variations, and so on. Another model of point processes frequently used as well is the self-exciting stationary point process introduced

in Hawkes (1972). As any stationary process it can also describe the periodic changes due to the particular form of its spectral density.

Recall that for the Poisson process $X_t, t \geq 0$ of intensity function $S(t), t \geq 0$ we have (X_t is the counting process)

$$\mathbf{P}\{X_t - X_s = k\} = (k!)^{-1} (\Lambda(t) - \Lambda(s))^k \exp\{\Lambda(s) - \Lambda(t)\},$$

where we suppose that $s < t$ and put

$$\Lambda(t) = \int_0^t S(v)\, \mathrm{d}v.$$

The self-exciting process $X_t, t \geq 0$ admits the representation

$$X_t = \int_0^t S(s, X)\, \mathrm{d}s + \pi_t,$$

where $\pi_t, t \geq 0$ is a local martingale and the intensity function

$$S(t, X) = S + \int_0^t g(t-s)\, \mathrm{d}X_s = S + \sum_{t_i < T} g(t - t_i).$$

It is supposed that

$$\rho = \int_0^\infty g(t)\, \mathrm{d}t < 1.$$

Under this condition the self-exciting process is a stationary point process with the rate

$$\mu = \frac{S}{1-\rho}$$

and the spectral density

$$f(\lambda) = \frac{\mu}{2\pi |1 - G(\lambda)|^2}, \qquad G(\lambda) = \int_0^\infty e^{\mathrm{i}\lambda t} g(t)\, \mathrm{d}t$$

[see Hawkes (1972) or Daley and Vere-Jones (2003) for details].

We consider two problems: Poisson against another Poisson and Poisson against a close self-exciting point process. The first one is to test the simple (basic) hypothesis

$$\mathcal{H}_0 : S(t) \equiv S_*(t), \qquad t \geq 0$$

where $S_*(t)$ is a known periodic function of period τ, against the composite alternative

$$\mathcal{H}_1 : S(t) \neq S_*(t), \qquad t \geq 0,$$

but $S(t)$ is always τ-periodic.

Let us denote $X_j(t) = X_{\tau(j-1)+t} - X_{\tau(j-1)}$, $j = 1, \ldots, n$, suppose that $T = n\tau$, and put

$$\hat{\Lambda}_n(t) = \frac{1}{n}\sum_{j=1}^{n} X_j(t).$$

The corresponding goodness-of-fit tests of Cramér–von Mises and Kolmogorov–Smirnov type can be based on the statistics

$$W_n^2 = \Lambda_*(\tau)^{-2} n \int_0^\tau \left[\hat{\Lambda}_n(t) - \Lambda_*(t)\right]^2 \mathrm{d}\Lambda_*(t),$$
$$D_n = \Lambda_*(\tau)^{-1/2} \sup_{0\le t\le\tau} \left|\hat{\Lambda}_n(t) - \Lambda_*(t)\right|.$$

It can be shown that

$$W_n^2 \Longrightarrow \int_0^1 W(s)^2\,\mathrm{d}s, \qquad \sqrt{n}\,D_n \Longrightarrow \sup_{0\le s\le 1} |W(s)|,$$

where $\{W(s), 0 \le s \le 1\}$ is a Wiener process [see Kutoyants (1998)]. Hence these statistics are asymptotically distribution-free and the tests

$$\psi_n\left(X^T\right) = 1_{\{W_n^2 > c_\alpha\}}, \qquad \phi_n\left(X^T\right) = 1_{\left\{\sqrt{n}D_n > d_\alpha\right\}}$$

with the constants c_α, d_α taken from Equations (27.4), are of asymptotic size α.

Let us describe the close contiguous alternatives which asymptotically reduce this problem to the *signal in the white Gaussian noise* model (27.5). We put

$$\Lambda(t) = \Lambda_*(t) + \frac{1}{\sqrt{n\Lambda_*(\tau)}} \int_0^t h(u(v))\ \mathrm{d}\Lambda_*(v), \qquad u(v) = \frac{\Lambda_*(v)}{\Lambda_*(\tau)}.$$

Here $h(\cdot)$ is an arbitrary function defining the alternative. Then if $\Lambda(t)$ satisfies this equality we have the convergence

$$W_n^2 \Longrightarrow \int_0^1 \left[\int_0^s h(v)\,\mathrm{d}v + W(s)\right]^2 \mathrm{d}s.$$

This convergence describes the power function of the Cramér–von Mises type test under these alternatives.

The second problem is to test the hypothesis

$$\mathcal{H}_0 : S(t) = S_*, \qquad t \ge 0$$

against nonparametric close (contiguous) alternative

$$\mathcal{H}_1 : S(t) = S_* + \frac{1}{\sqrt{T}} \int_0^t h(t-s)\,\mathrm{d}X_t, \qquad t \ge 0.$$

We consider the alternatives with the functions $h(\cdot) \geq 0$ having compact support and bounded.

We have $\Lambda_*(t) = S_*\, t$ and for some fixed $\tau > 0$ we can construct the same statistics

$$W_n^2 = \frac{n}{S_*\tau^2} \int_0^\tau \left[\hat{\Lambda}_n(t) - S_*\, t\right]^2 \mathrm{d}t, \qquad D_n = (S_*\, \tau)^{-1/2} \sup_{0\leq t\leq\tau} \left|\hat{\Lambda}_n(t) - S_*\, t\right|.$$

Of course, they have the same limits under hypothesis

$$W_n^2 \Longrightarrow \int_0^1 W(s)^2\, \mathrm{d}s, \qquad \sqrt{n} D_n \Longrightarrow \sup_{0\leq s\leq 1} |W(s)|.$$

To describe their behaviour under any fixed alternative $h(\cdot)$ we have to find the limit distribution of the vector

$$\mathbf{w}_n = \left(w_n\left(t_1\right), \ldots, w_n\left(t_k\right)\right), \qquad w_n(t_l) = \frac{1}{\sqrt{S_*\tau\, n}} \sum_{j=1}^n \left[X_j(t_l) - S_* t_l\right],$$

where $0 \leq t_l \leq \tau$. We know that this vector under hypothesis is asymptotically normal

$$\mathcal{L}_0\{\mathbf{w}_n\} \Longrightarrow \mathcal{N}(\mathbf{0}, \mathbf{R})$$

with covariance matrix

$$\mathbf{R} = (R_{lm})_{k\times k}, \qquad R_{lm} = \tau^{-1} \min(t_l, t_m).$$

Moreover, it was shown in Dachian and Kutoyants (2006) that for such alternatives the likelihood ratio is locally asymptotically normal; that is, the likelihood ratio admits the representation

$$Z_n(h) = \exp\left\{\Delta_n(h, X^n) - \frac{1}{2}\mathrm{I}(h) + r_n(h, X^n)\right\},$$

where

$$\Delta_n(h, X^n) = \frac{1}{S_*\sqrt{\tau n}} \int_0^{\tau n} \int_0^{t-} h(t-s)\ \mathrm{d}X_s\ \left[\mathrm{d}X_t - S_*\mathrm{d}t\right],$$

$$\mathrm{I}(h) = \int_0^\infty h(t)^2\, \mathrm{d}t + S_* \left(\int_0^\infty h(t)\, \mathrm{d}t\right)^2$$

and

$$\Delta_n(h, X^n) \Longrightarrow \mathcal{N}(0, \mathrm{I}(h)), \qquad r_n(h, X^n) \to 0. \tag{27.7}$$

To use the third Le Cam's lemma we describe the limit behaviour of the vector $(\Delta_n(h, X^n), \mathbf{w}_n)$. For the covariance $\mathbf{Q} = (Q_{lm}), l, m = 0, 1, \ldots, k$ of this vector we have

$$\mathbf{E}_0 \Delta_n(h, X^n) = 0, \qquad Q_{00} = \mathbf{E}_0 \Delta_n(h, X^n)^2 = \mathrm{I}(h)\,(1 + o(1)).$$

Furthermore, let us denote $\mathrm{d}\pi_t = \mathrm{d}X_t - S_*\mathrm{d}t$ and $H(t) = \int_0^{t-} h(t-s)\,\mathrm{d}X_s$; then we can write

$$\begin{aligned} Q_{0l} &= \mathbf{E}_0\left[\Delta_n(h, X^n)\, w_n(t_l)\right] \\ &= \frac{1}{nS_*^{3/2}\tau}\,\mathbf{E}_0\left(\sum_{j=1}^{n}\int_{\tau(j-1)}^{\tau j} H(t)\,\mathrm{d}\pi_t \sum_{i=1}^{n}\int_{\tau(i-1)}^{\tau(i-1)+t_l}\mathrm{d}\pi_t\right) \\ &= \frac{1}{n\tau\sqrt{S_*}}\sum_{j=1}^{n}\int_{\tau(j-1)}^{\tau(j-1)+t_l}\mathbf{E}_0 H(t)\;\mathrm{d}t = \frac{t_l}{\tau}\,\sqrt{S_*}\int_0^{\infty} h(t)\,\mathrm{d}t\;(1+o(1)), \end{aligned}$$

because

$$\mathbf{E}_0 H(t) = S_*\int_0^{t-} h(t-s)\,\mathrm{d}s = S_*\int_0^{\infty} h(s)\,\mathrm{d}s$$

for the large values of t [such that $[0,t]$ covers the support of $h(\cdot)$].

Therefore, if we denote

$$\bar{h} = \int_0^{\infty} h(s)\,\mathrm{d}s$$

then

$$Q_{0l} = Q_{l0} = \frac{t_l}{\tau}\,\sqrt{S_*}\,\bar{h}.$$

The proof of Theorem 1 in Dachian and Kutoyants (2006) can be applied to the linear combination of $\Delta_n(h, X^n)$ and $w_n(t_1), \ldots, w_n(t_k)$ and this yields the asymptotic normality

$$\mathcal{L}_0\Big(\Delta_n(h, X^n), \mathbf{w}_n\Big) \Longrightarrow \mathcal{N}(\mathbf{0}, \mathbf{Q}).$$

Hence by the third lemma of Le Cam we obtain the asymptotic normality of the vector $\mathbf{w}_n$,

$$\mathcal{L}_h\Big(\mathbf{w}_n\Big) \Longrightarrow \mathcal{L}\left(W(s_1) + s_1\sqrt{S_*}\,\bar{h}, \ldots, W(s_k) + s_k\sqrt{S_*}\,\bar{h}\right),$$

where we put $t_l = \tau\, s_l$. This weak convergence together with the estimates such as

$$\mathbf{E}_h\left|w_n(t_1) - w_n(t_2)\right|^2 \le C\,|t_1 - t_2|$$

provides the convergence (under alternative)

$$W_n^2 \Longrightarrow \int_0^1 \left[\sqrt{S_*}\,\bar{h}\,s + W(s)\right]^2 \mathrm{d}s.$$

We see that the limit experiment is of the type

$$dY_s = \sqrt{S_*}\,\bar{h}\,\mathrm{d}s + \mathrm{d}W(s), \qquad Y_0 = 0, \qquad 0 \le s \le 1.$$

The power $\beta(\psi_n, h)$ of the Cramér–von Mises type test $\psi_n(X^n) = 1_{\{W_n^2 > c_\alpha\}}$ is a function of the real parameter $\rho_h = \sqrt{S_*}\,\bar{h}$,

$$\beta\left(W_n, h\right) = \mathbf{P}\left(\int_0^1 \left[\rho_h\, s + W\left(s\right)\right]^2 \mathrm{d}s > c_\alpha\right) + o\left(1\right) = \beta_\psi\left(\rho_h\right) + o\left(1\right).$$

Using the arguments of Lemma 6.2 in Kutoyants (1998) it can be shown that for the Kolmogorov–Smirnov type test we have the convergence

$$\sqrt{n} D_n \Longrightarrow \sup_{0\le s\le 1} \left|\rho_h\, s + W\left(s\right)\right|.$$

The limit power function is

$$\beta_\phi\left(\rho_h\right) = \mathbf{P}\left(\sup_{0\le s\le 1} \left|\rho_h\, s + W\left(s\right)\right| > d_\alpha\right).$$

These two limit power functions are obtained by simulation in the next section.

27.5 Simulation

First, we present the simulation of the thresholds c_α and d_α of our Cramér–von Mises and Kolmogorov–Smirnov type tests. Because these thresholds are given by the equations (27.4), we obtain them by simulating 10^7 trajectories of a Wiener process on [0,1] and calculating empirical $1-\alpha$ quantiles of the statistics

$$W^2 = \int_0^1 W(s)^2\, \mathrm{d}s \quad \text{and} \quad D = \sup_{0\le s\le 1} \left|W\left(s\right)\right|,$$

respectively. Note that the distribution of W^2 coincides with the distribution of the quadratic form

$$W^2 = \sum_{k=1}^{\infty} \frac{\zeta_k^2}{\left(\pi k\right)^2}, \qquad \zeta_k \text{ i.i.d. } \sim \mathcal{N}\left(0,1\right)$$

and both distributions are extensively studied [see (1.9.4(1)) and (1.15.4) in Borodin and Salmienen (2002)]. The analytical expressions are quite complicated and we would like to compare by simulation c_α and d_α with the real (finite time) thresholds giving the tests of exact size α, that is, c_α^T and d_α^T given by equations

$$\mathbf{P}\left\{W_n^2 > c_\alpha^T\right\} = \alpha \quad \text{and} \quad \mathbf{P}\left\{\sqrt{n} D_n > d_\alpha^T\right\} = \alpha,$$

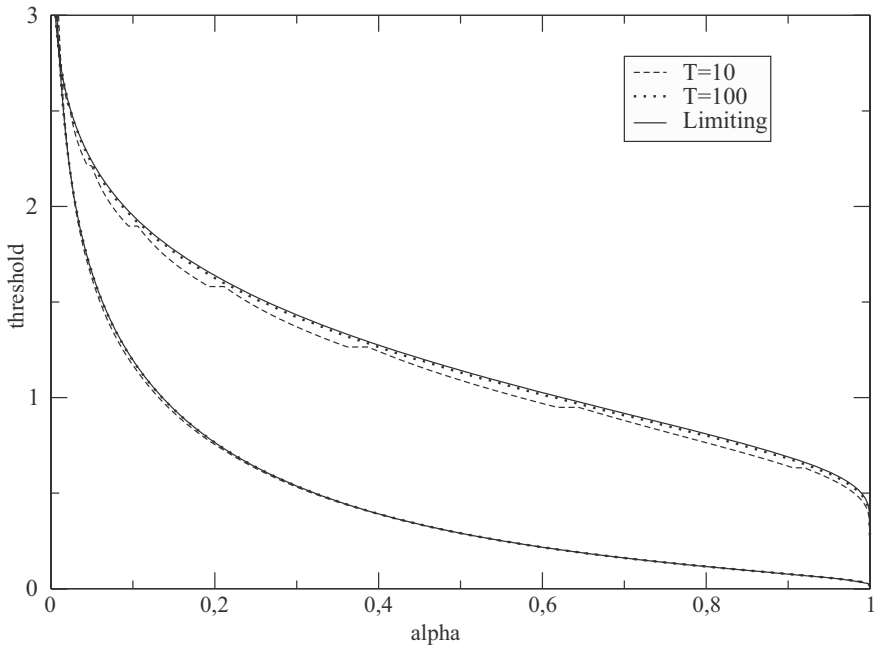

Figure 27.1. Threshold choice.

respectively. We choose $S^* = 1$ and obtain c_α^T and d_α^T by simulating 10^7 trajectories of a Poisson process of intensity 1 on [0,T] and calculating empirical $1-\alpha$ quantiles of the statistics W_n^2 and $\sqrt{n}D_n$. The thresholds simulated for $T = 10$, $T = 100$, and for the limiting case are presented in Figure 27.1. The lower curves correspond to the Cramér–von Mises type test, and the upper ones to the Kolmogorov–Smirnov type test. As we can see, for $T = 100$ the real thresholds are already indistinguishable from the limiting ones, especially in the case of the Cramér–von Mises type test.

It is interesting to compare the asymptotics of the Cramér–von Mises and Kolmogorov–Smirnov type tests with the locally asymptotically uniformly most powerful (LAUMP) test

$$\hat{\phi}_n\left(X^n\right) = 1_{\{\delta_T > z_\alpha\}}, \qquad \delta_T = \frac{X_{n\tau} - S_* n\tau}{\sqrt{S_* n\tau}}$$

proposed for this problem in Dachian and Kutoyants (2006). Here z_α is the $1-\alpha$ quantile of the standard Gaussian law, $\mathbf{P}\left(\zeta > z_\alpha\right) = \alpha$, $\zeta \sim \mathcal{N}\left(0,1\right)$. The limit power function of $\hat{\phi}_n$ is

$$\beta_{\hat{\phi}}\left(\rho_h\right) = \mathbf{P}\left(\rho_h + \zeta > z_\alpha\right).$$

In Figure 27.2 we compare the limit power functions $\beta_\psi\left(\rho\right)$, $\beta_\phi\left(\rho\right)$, and $\beta_{\hat{\phi}}\left(\rho\right)$. The last one can clearly be calculated directly, and the first two are obtained by

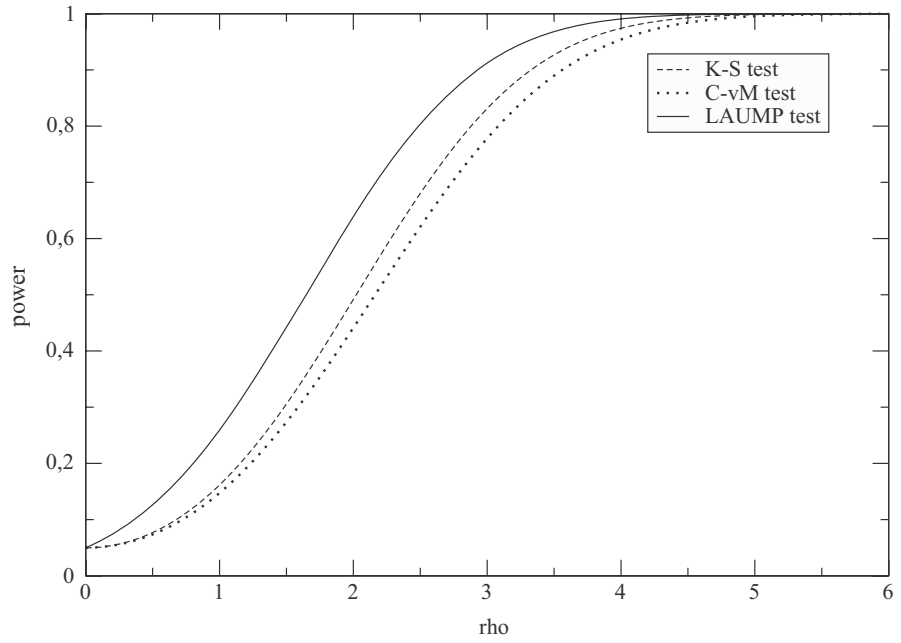

Figure 27.2. Limit power functions.

simulating 10^7 trajectories of a Wiener process on [0,1] and calculating empirical frequencies of the events

$$\left\{\int_0^1 \left[\rho\, s + W\left(s\right)\right]^2 \mathrm{d}s > c_\alpha\right\} \quad \text{and} \quad \left\{\sup_{0\le s\le 1} \left|\rho\, s + W\left(s\right)\right| > d_\alpha\right\},$$

respectively.

The simulation shows the exact (quantitative) comparison of the limit power functions. We see that the power of the LAUMP test is higher than the two others and this is of course evident. We see also that the Kolmogorov–Smirnov type test is more powerful that the Cramér–von Mises type test.

References

1. Borodin, A. N. and Salmienen, R. (2002). *Handbook of Brownian Motion–Facts and Formulae, (2nd ed.)*, Birkhäuser, Basel.

2. Dachian, S. and Kutoyants, Yu. A. (2006). Hypotheses testing: Poisson versus self-exciting, *Scandinavian Journal of Statistics*, **33**, 391–408.

3. Daley, D. J. and Vere-Jones, D. (2003). *An Introduction to the Theory of Point Processes. vol. I. (2nd ed.)*, Springer-Verlag, New York.

4. Durbin, J. (1973). *Distribution Theory for Tests Based on the Sample Distribution Function*, SIAM, Philadelphia.

5. Fournie, E. (1992). Un test de type Kolmogorov-Smirnov pour processus de diffusions ergodic, *Rapport de Recherche*, **1696**, INRIA, Sophia-Antipolis.

6. Fournie, E. and Kutoyants, Yu. A. (1993). Estimateur de la distance minimale pour des processus de diffusion ergodiques, *Rapport de Recherche*, **1952**, INRIA, Sophia-Antipolis.

7. Greenwood, P. E. and Nikulin, M. (1996). *A Guide to Chi-Squared Testing*, John Wiley & Sons, New York.

8. Hawkes, A. G. (1972). Spectra for some mutually exciting point processes with associated variable, In *Stochastic Point Processes*, (Ed., P. A. W. Lewis), John Wiley & Sons, New York.

9. Iacus, S. and Kutoyants, Yu. A. (2001). Semiparametric hypotheses testing for dynamical systems with small noise, *Mathematical Methods of Statistics*, **10**, 105–120.

10. Ingster, Yu. I. and Suslina, I. A. (2003). *Nonparametric Goodness-of-Fit Testing Under Gaussian Models*, Springer-Verlag, New York.

11. Kutoyants, Yu. A. (1994). *Identification of Dynamical Systems with Small Noise,* Kluwer, Dordrecht.

12. Kutoyants, Yu. A. (1998). Statistical inference for spatial Poisson processes, In *Lecture Notes in Statistics*, **134**, Springer-Verlag, New York.

13. Kutoyants, Yu. A. (2004). *Statistical Inference for Ergodic Diffusion Processes,* Springer-Verlag, London.

14. Kutoyants, Yu. A. (2006). Goodness-of-fit tests for perturbed dynamical systems, in preparation.

15. Liptser, R. S. and Shiryayev, A. N. (2001). *Statistics of Random Processes. II. Applications,* (2nd ed.) Springer-Verlag, New York.

16. Negri, I. (1998). Stationary distribution function estimation for ergodic diffusion process, *Statistical Inference for Stochastic Processes,* **1**, 1, 61–84.

Journal of Statistical Planning and Inference 139 (2009) 1668 – 1684

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

journal homepage: www.elsevier.com/locate/jspi

Hypotheses testing: Poisson versus stress-release

Sergueï Dachian[a], Yury A. Kutoyants[b,*]

[a] *Laboratoire de Mathématiques, Université Blaise Pascal, 63177 Aubière Cedex, France*
[b] *Laboratoire de Statistique et Processus, Université du Maine, 72085 Le Mans Cedex 9, France*

ARTICLE INFO

Available online 24 May 2008

MSC:
62M05

Keywords:
Poisson process
Self-correcting process
Stress-release process
Hypotheses testing
Wald test
Likelihood ratio test
Score-function test
Unit-root AR process

ABSTRACT

We consider the problem of hypotheses testing with the basic simple hypothesis: observed sequence of points corresponds to stationary Poisson process with known intensity against a composite one-sided parametric alternative that this is a stress-release point process. The underlying family of measures is locally asymptotically quadratic and we describe the behavior of score-function, likelihood ratio and Wald tests in the asymptotics of large samples. The results of numerical simulations are presented.

1. Introduction

Poisson process plays a key role in describing of the reliability systems (see, for example, Rigdon and Basu, 2000). The freedom in the choice of intensity function of inhomogeneous Poisson process allows to apply this model to a wide range of applied problems. One of the main characteristics of this process is the independence of the increments on disjoint intervals and the main statistical advantage is the possibility to use the likelihood ratio analysis. The statistical inference for the other point processes is essentially more difficult because the likelihood ratio formula is rarely available (in closed form). There are at least two exceptions. The first one concerns the self-exciting point processes and the second is the stress-release point processes. In these both cases the increments are not independent but the intensity function, being random process, is measurable with respect to the observations and therefore we have the opportunity to use the likelihood ratio analysis. Note that these three types of point processes (Poisson, self-exciting and stress-release) cover a large class of stationary point processes. We suppose that the problem of the choice of the type of point process is quite important and the most interesting is the testing in the region, where these models are statistically close and the large samples analysis is non-degenerate (contiguous alternatives).

The model of *self-correcting* (also called *stress-release*) point process was proposed in Isham and Westcott (1979) to describe a stationary sequence of events $\{t_1, t_2, \ldots\}$ which automatically corrects the intensity function. Note that essentially similar model was introduced in Knopov (1971) and in Vere-Jones (1978) to describe the seismic activity (see Ogata and Vere-Jones, 1984; Lu et al., 1999). This is an elementary stochastic version of the *elastic rebound theory* of earthquake formation. This model is used in storage and insurance applications too. Roughly speaking, the stress level (intensity function) increases deterministically between the events and at the instant of event it is reduced (released). We suppose that this model corresponds well to the behavior of certain technical systems and can be applied in the study of reliability of such models.

* Corresponding author.
E-mail address: kutoyants@univ-lemans.fr (Y.A. Kutoyants).

0378-3758/$ - see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.jspi.2008.05.025

To introduce this processes we denote by $X=\{X_t, t\geqslant 0\}$ the counting process, i.e., X_t is equal to the number of events on the time interval $[0,t]$. Recall that for a stationary Poisson process with a constant intensity $S>0$ the increments of X on disjoint intervals are independent and distributed according to Poisson law

$$\mathbf{P}\{X_t - X_s = k\} = \frac{S^k(t-s)^k}{k!}\mathrm{e}^{-S(t-s)}, \quad 0\leqslant s<t, \ \ k=0,1,\dots\ .$$

Particularly,

$$\mathbf{P}\{X_{t+\mathrm{d}t} - X_t > 0\} = S\,\mathrm{d}t(1+\mathrm{o}(1)).$$

For stress-release point process we have

$$\mathbf{P}\{X_{t+\mathrm{d}t} - X_t > 0|\mathscr{F}_t\} = S(t,X_t)\,\mathrm{d}t(1+\mathrm{o}(1)),$$

where $\mathscr{F}_t$ is the σ-field generated by $\{X_s, 0\leqslant s\leqslant t\}$ and the intensity function

$$S(t,X_t) = a\,\psi(at - X_t), \quad t\geqslant 0.$$

Here $a>0$ and the function $\psi(\cdot)$ satisfies the following conditions:

1. $0\leqslant\psi(x)<\infty$ for any $x\in\mathbb{R}$,
2. there exists a positive constant c such that $\psi(x)\geqslant c$ for any $x>0$,
3. $\underline{\lim}_{x\to\infty}\psi(x)>1$ and $\overline{\lim}_{x\to-\infty}\psi(x)<1$.

Self-correcting processes are called as well stress-release processes (see Daley and Vere-Jones, 2003, p. 239). This class of processes is widely used as a good mathematical model for non-Poissonian sequences of events. This model was found especially attractive in the description of earthquakes.

Example 1. Let

$$S(t,X_t) = \exp\{\alpha + \beta(t - \varrho X_t)\},$$

where $\beta>0$, $\varrho>0$. It is easy to see that conditions 1–3 are fulfilled and the point process with such intensity function is stress-release.

This model was studied by many authors (see the references in Daley and Vere-Jones, 2003). Particularly it was shown that under mild conditions there exists an invariant measure μ and the law of large numbers (LLN)

$$\frac{1}{T}\int_0^T h(St - X_t)\,\mathrm{d}t \longrightarrow \int h(y)\mu(\mathrm{d}y) \tag{1}$$

is valid (see Vere-Jones and Ogata, 1984; Hayashi, 1986; Zheng, 1991). Here $h(\cdot)$ is a continuous, integrable (w.r.t. μ) function and $S>0$ is the rate of the point process. For the model of Example 1 we have the LLN if $\rho>0$ and $\beta>0$.

As the stress-release model is an alternative for the stationary Poisson process, it is natural and important to test these two hypotheses by the observations $\{t_1,t_2,\dots\}$ on the time interval $[0,T]$, i.e., to test

$$S(t,X_t) = S \quad \text{versus} \quad S(t,X_t) = a\psi(at - X_t).$$

Remind that the likelihood ratio in this problem has the following form:

$$L(X^T) = \exp\left\{\int_0^T \ln\frac{a\psi(at - X_{t-})}{S}[\mathrm{d}X_t - S\,\mathrm{d}t] - \int_0^T\left[\frac{a\psi(at-X_t)}{S} - 1 - \ln\frac{a\psi(at-X_t)}{S}\right]S\,\mathrm{d}t\right\},$$

where X_{t-} is the limit from the left of X_t at the point t (see, for example, Liptser and Shiryayev, 2001). Therefore, if the function $a\psi(\cdot)/S$ is separated from 1 then the second integral in this representation tends to infinity and there are many consistent tests. Hence it is more interesting to compare tests in the situations when the alternatives are *contiguous*, i.e., the corresponding sequence of measures are contiguous. This corresponds well to Pitman's approach in hypotheses testing (see Pitman, 1948). We can have such situations if $\psi(\cdot) = S + \mathrm{o}(1)$ with special rates $\mathrm{o}(1)$. In this work we consider one of such models defined by the intensity function $S(t,X_t) = S\psi(\vartheta(St - X_t))$ where ϑ is a *small parameter* and $\psi(0)=1$. We suppose that the function $\psi(\cdot)$ is smooth and we can write

$$\int_0^T[\psi(\vartheta(St-X_t)) - 1 - \ln\psi(\vartheta(St-X_t))]S\,\mathrm{d}t = \frac{\vartheta^2\dot{\psi}(0)^2 S}{2}\int_0^T(St-X_t)^2\mathrm{d}t(1+\mathrm{o}(1)).$$

It is easy to see that the rate $\vartheta=\vartheta_T\to 0$ under hypothesis $S(t,X_t)=S$ is $\vartheta_T\sim T^{-1}$ because

$$\frac{1}{ST^2}\int_0^T (St-X_t)^2\mathrm{d}t=\int_0^1 W_T(s)^2\,\mathrm{d}s\Longrightarrow\int_0^1 W(s)^2\,\mathrm{d}s,$$

where $W_T(s)=(ST)^{-1/2}(STs-X_{Ts})\Rightarrow W(s)$, and $\{W(s),0\leqslant s\leqslant 1\}$ is Wiener process. Note that we put $a=S$, otherwise

$$\begin{aligned}\frac{\dot\psi(0)^2\vartheta_T^2}{2}\int_0^T (at-X_t)^2\,\mathrm{d}t&=\frac{\dot\psi(0)^2\vartheta_T^2}{2}\int_0^T\left((a-S)t+\sqrt{ST}\frac{St-X_t}{\sqrt{ST}}\right)^2\mathrm{d}t\\&=\frac{\dot\psi(0)^2\vartheta_T^2}{2}T\int_0^1((a-S)\nu T+\sqrt{ST}W_T(\nu))^2\,\mathrm{d}\nu\\&=\frac{\dot\psi(0)^2}{6}\vartheta_T^2(a-S)^2T^3(1+\mathrm{o}(1)).\end{aligned}$$

Therefore, if $a\neq S$, then we have to take $\vartheta_T=uT^{-3/2}$ and to test the simple hypothesis $\mathscr{H}_0: u=0$ against $\mathscr{H}_1: u>0$. In this case the family of measures is LAN and the usual construction provides us *asymptotically uniformly most powerful test* (see, for example, Roussas, 1972). Note that according to (1) for any fixed alternative $\vartheta>0$ we have the convergence

$$\frac{1}{T}\int_0^T (St-X_t)^2\mathrm{d}t\longrightarrow\int y^2\mu(\mathrm{d}y)$$

which, of course, requires another normalization.

Therefore we consider the problem of hypotheses testing when under hypothesis $\mathscr{H}_0$ the intensity function is a known constant $S>0$ (Poisson process) and the alternative $\mathscr{H}_1$ is one-sided composite: stress-release process with intensity function $S(t,X_t)=S\psi(\vartheta_T(St-X_t))$, where for convenience of notation we put $\vartheta_T=u/S\dot\psi(0)T$ (we suppose that $\dot\psi(0)>0$). In this case the corresponding likelihood ratio $Z_T(u)$ converges to the limit process

$$Z(u)=\exp\left\{-u\int_0^1 W(s)\,\mathrm{d}W(s)-\frac{u^2}{2}\int_0^1 W(s)^2\,\mathrm{d}s\right\},$$

i.e., the family of measures is *locally asymptotically quadratic* (LAQ) (see, for example, Le Cam and Yang, 2000). We study three tests: *score-function test*, *likelihood ratio test*, *Wald test* and compare their power functions with the power function of the *Neyman–Pearson test*. Note that we calculate all limits under hypothesis (Poisson process) and we obtain the limit distributions of the underlying statistics under alternative (stress-release process) with the help of Le Cam's Third Lemma. Therefore we do not use directly conditions 1–3 given above.

The similar limit likelihood ratio process arises in the problem of hypotheses testing $u=0$ against $u>0$ for the time series

$$X_j=\left(1-\frac{u}{n}\right)X_{j-1}+\varepsilon_j,\quad j=1,\dots,n\to\infty,$$

where ε_j are i.i.d. random variables, $\mathbf{E}\varepsilon_j=0,\mathbf{E}\varepsilon_j^2=\sigma^2$. The asymptotic properties of tests are described under hypothesis and alternatives in Chan and Wei (1987) and Phillips (1987). Particularly, the limits of the power functions are given with the help of Ornstein–Uhlenbeck process

$$\mathrm{d}Y_s=-uY_s\,\mathrm{d}s+\mathrm{d}W_s,\quad Y_0=0,\ \ 0\leqslant s\leqslant 1.$$

Then these limit powers were compared in Swensen (1997).

For the model of Example 1 the power function (for local alternatives) was studied in Ogata and Vere-Jones (1984) and in Luschgy (1993, 1994). The limit likelihood ratio and tests are similar to that of the mentioned above time series problem. Remind as well that in Feigin (1979) it was noted that the same limit likelihood ratio arises in the problem of testing the simple hypothesis $u=0$ against one-sided alternative $u>0$ by observations

$$\mathrm{d}X_t=-\frac{u}{T}X_t\,\mathrm{d}t+\mathrm{d}W_t,\quad X_0=0,\ \ 0\leqslant t\leqslant T\to\infty.$$

In our case we obtain similar limit expressions for the likelihood ratio and power functions and compare the errors of tests. The analytical considerations give us an asymptotic (for large values of u) ordering of the tests. The numerical simulations of the tests show that for the small values of ε and for the moderate values of u the power functions of the likelihood ratio and Wald tests are indistinguishable (from the point of view of numerical simulations) of the Neyman–Pearson envelope. This interesting property was noticed (for $\varepsilon=0.05$) in Eliott et al. (1996) on the base of 2×10^3 simulations. In our work we obtain similar result having 10^7 simulations and we observe for the larger values of ε that the asymptotic ordering of the tests holds already for the moderate values of u.

Note finally that a similar problem of hypotheses testing in the situation when the alternative process is self-exciting (see Hawkes, 1972) was considered in Dachian and Kutoyants (2006).

2. Score-function test

We observe a trajectory $X^T = \{X_t, 0 \leqslant t \leqslant T\}$ of a point process of intensity function $S(\cdot, X_t)$ and consider the problem of testing the simple hypothesis against close one-sided composite alternative

$$\mathscr{H}_0 : S(t, X_t) = S_*, \tag{2}$$

$$\mathscr{H}_1 : S(t, X_t) = S_* \psi(\vartheta_T [S_* t - X_t]), \quad \vartheta_T > 0, \tag{3}$$

where ϑ_T is a small parameter, the value S_* and the function $\psi(\cdot)$ are known. The problem is regular in the following sense.

Condition $\mathscr{A}$. *The function $\psi(x), x \in \mathbb{R}$ is positive, continuously differentiable at the point $x = 0$, $\psi(0) = 1$ and $\dot{\psi}(0) > 0$.*

The rate of convergence $\vartheta_T \to 0$ is chosen such that the likelihood ratio $L(\vartheta_T, X^T)$ is asymptotically non-degenerate. In the case $\dot{\psi}(0) < 0$ we need to change just one sign in the test. This leads us to the reparametrization

$$\vartheta_T = \frac{u}{S_* \dot{\psi}(0) T}, \quad u \geqslant 0$$

and to the corresponding hypotheses testing problem

$$\mathscr{H}_0 : u = 0, \tag{4}$$

$$\mathscr{H}_1 : u > 0. \tag{5}$$

Therefore, we observe a Poisson process of intensity S_* under hypothesis $\mathscr{H}_0$ and the point process under alternative $\mathscr{H}_1$ has intensity function

$$S(t, X_t) = S_* + \frac{u}{T}(S_* t - X_t) + o(T^{-1/2}).$$

Let us fix $\varepsilon \in (0, 1)$ and denote by $\mathscr{K}_\varepsilon$ the class of test functions $\phi_T(X^T)$ of asymptotic size ε, i.e., for $\phi_T \in \mathscr{K}_\varepsilon$ we have

$$\lim_{T \to \infty} \mathbf{E}_0 \phi_T(X^T) = \varepsilon.$$

As usual, $\phi_T(X^T)$ is the probability to accept the hypothesis $\mathscr{H}_1$ having observations X^T. The corresponding power function is

$$\beta_T(u, \phi_T) = \mathbf{E}_u \phi_T(X^T), \quad u \geqslant 0.$$

Let us introduce the statistic

$$\Delta_T(X^T) = \frac{1}{S_* T} \int_0^T (S_* t - X_{t-})[\mathrm{d}X_t - S_* \, \mathrm{d}t] = \frac{X_T - (X_T - S_* T)^2}{2 S_* T}. \tag{6}$$

The last equality follows from the elementary representation (see, for example, Kutoyants, 1984, Lemma 4.2.1) for the centered Poisson process $\pi_t = X_t - S_* t$:

$$\pi_T^2 = 2 \int_0^T \pi_{t-} \, \mathrm{d}\pi_t + \pi_T + S_* T$$

which obviously is equivalent to

$$\frac{1}{T} \int_0^T \pi_{t-} \, \mathrm{d}\pi_t = \frac{\pi_T^2 - X_T}{2T}.$$

Define as well two random variables

$$\Delta(W) = \frac{1}{2}(1 - W(1)^2) = -\int_0^1 W(s) \mathrm{d}W(s), \quad \mathrm{J}(W) = \int_0^1 W(s)^2 \mathrm{d}s,$$

where $\{W(s), 0 \leqslant s \leqslant 1\}$ is standard Wiener process.

Remind that the likelihood ratio in this problem has the form

$$L\left(\frac{u}{\gamma T}, X^T\right) = \exp\left\{ \int_0^T \ln \psi\left(\frac{u}{\gamma T}(S_* t - X_{t-})\right) [\mathrm{d}X_t - S_* \, \mathrm{d}t] \right.$$
$$\left. - \int_0^T \left[\psi\left(\frac{u}{\gamma T}(S_* t - X_t)\right) - 1 - \ln \psi\left(\frac{u}{\gamma T}(S_* t - X_t)\right) \right] S_* \, \mathrm{d}t \right\}, \tag{7}$$

where $\gamma = S_* \dot{\psi}(0)$ (see, for example, Liptser and Shiryayev, 2001).

Therefore the direct differentiation w.r.t. u at the point $u=0$ gives us the introduced above statistic

$$\frac{\partial}{\partial u}\ln L\left(\frac{u}{\gamma T},X^T\right)\Big|_{u=0}=\Delta_T(X^T).$$

Below we denote

$$a_\varepsilon=\frac{1-z^2_{(1-\varepsilon)/2}}{2}\quad\text{and}\quad h(u)=\sqrt{\frac{2u}{1-\mathrm{e}^{-2u}}},$$

where z_a is $1-a$ quantile of standard Gaussian law, i.e., $\mathbf{P}(\zeta>z_a)=a$, for $\zeta\sim\mathcal{N}(0,1)$.

We have the following result.

Theorem 1. *Let the Condition $\mathcal{A}$ be fulfilled, then the score-function test*

$$\phi^*_T(X^T)=\mathbb{1}_{\{\Delta_T(X^T)>a_\varepsilon\}} \tag{8}$$

belongs to the class $\mathcal{K}_\varepsilon$ and for any $u_*>0$ its power function

$$\beta_T(u_*,\phi^*_T)\to\beta^*(u_*)=\mathbf{P}\{|\zeta|\leqslant h(u_*)z_{(1-\varepsilon)/2}\}. \tag{9}$$

Proof. Under hypothesis $\mathcal{H}_0$ the value X_T is a Poissonian random variable with parameter S_*T. Therefore we have immediately

$$\frac{X_T}{S_*T}\longrightarrow 1,\quad\frac{X_T-ST}{\sqrt{S_*T}}\Longrightarrow W(1)\sim\mathcal{N}(0,1)$$

and $\Delta_T(X^T)\Longrightarrow\Delta(W)$ as $T\to\infty$. Hence

$$\mathbf{P}_0\{\Delta_T(X^T)>a_\varepsilon\}\longrightarrow\mathbf{P}\left\{\Delta(W)>\frac{1-z^2_{(1-\varepsilon)/2}}{2}\right\}=\mathbf{P}\{|\zeta|<z_{(1-\varepsilon)/2}\}=\varepsilon.$$

This provides $\phi^*_T\in\mathcal{K}_\varepsilon$.

To study the power $\beta_T(u_*,\phi^*_T)$ we would like to use the Third Le Cam Lemma (see Le Cam and Yang, 2000; Strasser, 1985). Therefore we need first to show the joint weak convergence

$$\mathcal{L}_0(\Delta_T,l_T(u))\Longrightarrow\mathcal{L}\left(\Delta(W),u\Delta(W)-\frac{u^2}{2}\mathrm{J}(W)\right), \tag{10}$$

where $l_T(u)=\ln L(u/\gamma T,X^T)$.

To verify (10) we denote

$$l^*_T(u)=u\Delta_T(X^T)-\frac{u^2}{2}\mathrm{J}_T(X^T),$$

where

$$\mathrm{J}_T(X^T)=\frac{1}{S_*T^2}\int_0^T(S_*t-X_t)^2\,\mathrm{d}t$$

and show that

$$\mathcal{L}_0(l^*_T(u))\Longrightarrow\mathcal{L}\left(u\Delta(W)-\frac{u^2}{2}\mathrm{J}(W)\right). \tag{11}$$

Then (10) will follow from the convergence

$$l^*_T(u_T)-l_T(u_T)\to 0 \tag{12}$$

for any bounded sequence u_T. □

Lemma 1.

$$\mathcal{L}_0\{\Delta_T(X^T),\mathrm{J}_T(X^T)\}\Longrightarrow\left(-\int_0^1 W(s)\,\mathrm{d}W(s),\int_0^1 W(s)^2\,\mathrm{d}s\right). \tag{13}$$

Proof. Let us put $W_T(s)=(S_*T)^{-1/2}\pi_{sT}$, $s\in[0,1]$. Then

$$\mathbf{E}_0 W_T(s)=0,\quad \mathbf{E}_0[W_T(s_1)W_T(s_2)]=\min(s_1,s_2)$$

and we have

$$\mathrm{J}_T(X^T)=\frac{1}{S_*T^2}\int_0^T \pi_t^2\,\mathrm{d}t=\int_0^1 W_T(s)^2\,\mathrm{d}s.$$

Using the standard arguments we verify (well-known fact) that for any collection $\{s_1,\dots,s_k\}$ we have the weak convergence (as $T\to\infty$) of the vectors

$$(W_T(s_1),\dots,W_T(s_k))\Longrightarrow(W(s_1),\dots,W(s_k)).$$

Moreover the following estimate holds:

$$(\mathbf{E}_0|W_T(s_1)^2-W_T(s_2)^2|)^2\leqslant\mathbf{E}_0|W_T(s_1)-W_T(s_2)|^2\mathbf{E}_0|W_T(s_1)+W_T(s_2)|^2\leqslant 4|s_2-s_1|.$$

Hence (see Gikhman and Skorokhod, 1969, Section IX.7) we have the convergence (in distribution) of integrals

$$\int_0^1 W_T(s)^2\,\mathrm{d}s\Longrightarrow\int_0^1 W(s)^2\,\mathrm{d}s$$

and

$$\Delta_T(X^T)=\frac{1-W_T(1)^2}{2}(1+\mathrm{o}(1))\Longrightarrow\frac{1-W(1)^2}{2}=-\int_0^1 W(s)\,\mathrm{d}W(s).$$

It is easy to see that we have the same time the joint convergence too because from the given above proof it follows that for any λ_1,λ_2:

$$\lambda_1 W_T(1)^2+\lambda_2\int_0^1 W_T(s)^2\,\mathrm{d}s\Longrightarrow\lambda_1 W(1)^2+\lambda_2\int_0^1 W(s)^2\,\mathrm{d}s.$$

Therefore Lemma 1 is proved. □

Our goal now is to establish a slightly more strong than (12) relation

$$l_T(u_T)=u_T\Delta_T(X^T)(1+\mathrm{o}(1))-\frac{u_T^2}{2}\int_0^1 W_T(s)^2\,\mathrm{d}s(1+\mathrm{o}(1)),\tag{14}$$

where $\mathrm{o}(1)\to 0$ for any sequence $u_T\in\mathbb{U}_T$ with $\mathbb{U}_T=\{u:0\leqslant u<\sqrt{S_*T}/\ln T\}$.

We can write

$$\begin{aligned}l_T^*(u)-l_T(u)&=\int_0^T\left[-\frac{uW_T\left(\frac{t}{T}\right)}{\sqrt{S_*T}}-\ln\psi\left(\frac{-uW_T\left(\frac{t}{T}\right)}{\dot\psi(0)\sqrt{S_*T}}\right)\right]\mathrm{d}\pi_t\\&\quad-\int_0^T\left[\frac{u^2W_T\left(\frac{t}{T}\right)^2}{2S_*T}-\psi\left(\frac{-uW_T\left(\frac{t}{T}\right)}{\dot\psi(0)\sqrt{S_*T}}\right)+1+\ln\psi\left(\frac{-uW_T\left(\frac{t}{T}\right)}{\dot\psi(0)\sqrt{S_*T}}\right)\right]S_*\,\mathrm{d}t\\&\equiv u\delta_{1,T}-\frac{u^2}{2}\delta_{2,T}\end{aligned}$$

with obvious notation. Remind that $u>0$. Using Lenglart inequality (see, for example, Liptser and Shiryayev, 2001) we obtain for the first term

$$\mathbf{P}_0\{|\delta_{1,T}|>a\}\leqslant\frac{b}{a}+\mathbf{P}_0\left\{\int_0^1\left[W_T(s)+\frac{\sqrt{S_*T}}{u}\ln\psi\left(\frac{-uW_T(s)}{\dot\psi(0)\sqrt{S_*T}}\right)\right]^2\mathrm{d}s>b\right\}$$

for any $a>0$ and $b>0$. Now expanding the functions $\psi(\cdot)$ we obtain

$$\psi\left(\frac{-uW_T(s)}{\dot\psi(0)\sqrt{S_*T}}\right)=1-\frac{uW_T(s)}{\dot\psi(0)\sqrt{S_*T}}\dot\psi\left(\frac{-\tilde uW_T(s)}{\dot\psi(0)\sqrt{S_*T}}\right),$$

where $\tilde{u} \leqslant u$. Introduce the set

$$\mathbb{C}_T = \left\{ \omega : \sup_{0 \leqslant s \leqslant 1} |W_T(s)| \leqslant \dot{\psi}(0)\sqrt{\ln T} \right\}$$

and note that for $\omega \in \mathbb{C}_T$ we have the estimate

$$\sup_{u \in \mathbb{U}_T} \sup_{0 \leqslant s \leqslant 1} \frac{u|W_T(s)|}{\dot{\psi}(0)\sqrt{S_* T}} \leqslant \frac{1}{\sqrt{\ln T}}.$$

Hence for all $u \in \mathbb{U}_T$ on this set we can write

$$\sup_{0 \leqslant s \leqslant 1} \left| \dot{\psi}(0) - \dot{\psi}\left(\frac{-\tilde{u} W_T(s)}{\dot{\psi}(0)\sqrt{S_* T}} \right) \right| \leqslant \sup_{|v| \leqslant (\ln T)^{-1/2}} |\dot{\psi}(0) - \dot{\psi}(v)| = h_T \to 0$$

as $T \to \infty$ because the derivative is continuous at the point $v = 0$.

Let us denote $u_s = u W_T(s)/\dot{\psi}(0)\sqrt{S_* T}$. Using the expansion of the logarithm

$$\ln(\psi(-u_s)) = \ln(1 - u_s \dot{\psi}(-\tilde{u}_s)) = -\frac{u_s \dot{\psi}(-\tilde{u}_s)}{1 - \tilde{\tilde{u}}_s \dot{\psi}(-\tilde{u}_s)}$$

we obtain the following estimate:

$$\begin{aligned} &\mathbf{P}_0 \left\{ \int_0^1 \left[W_T(s) + \frac{\sqrt{S_* T}}{u} \ln \psi(-u_s) \right]^2 \mathrm{d}s > b \right\} \\ &\quad \leqslant \mathbf{P}_0\{\mathbb{C}_T^c\} + \mathbf{P}_0 \left\{ \int_0^1 W_T(s)^2 \left(1 - \frac{\dot{\psi}(-\tilde{u}_s)}{\dot{\psi}(0)(1 - \tilde{\tilde{u}}_s \dot{\psi}(-\tilde{u}_s))} \right)^2 \mathrm{d}s > b, \mathbb{C}_T \right\}. \end{aligned}$$

Remind that $W_T(s)$ is martingale, hence by Doob inequality we have

$$\mathbf{P}_0\{\mathbb{C}_T^c\} \leqslant \mathbf{P}_0\{|W_T(1)| > \dot{\psi}(0)\sqrt{\ln T}\} \leqslant \frac{1}{\dot{\psi}(0)^2 \ln T}.$$

For the second probability after elementary estimates we obtain

$$\begin{aligned} &\mathbf{P}_0 \left\{ \int_0^1 W_T(s)^2 \left(1 - \frac{\dot{\psi}(-\tilde{u}_s)}{\dot{\psi}(0)(1 - \tilde{\tilde{u}}_s \dot{\psi}(-\tilde{u}_s))} \right)^2 \mathrm{d}s > b, \mathbb{C}_T \right\} \\ &\quad \leqslant \mathbf{P}_0 \left\{ C \int_0^1 W_T(s)^2 \, \mathrm{d}s \left(h_T^2 + \frac{1}{\ln T} \right) > b \right\} \leqslant \frac{C}{2b} \left(h_T^2 + \frac{1}{\ln T} \right) \end{aligned}$$

with some constant $C > 0$. Recall that by Tchebyshev inequality

$$\mathbf{P}_0 \left\{ \int_0^1 W_T(s)^2 \, \mathrm{d}s > A \right\} \leqslant \frac{1}{2A}.$$

Therefore, if we take $b = a^2$ then for any $a > 0$

$$\mathbf{P}_0\{|\delta_{1,T}| > a\} \longrightarrow 0$$

as $T \to \infty$.

The similar arguments allow to prove the convergence

$$\mathbf{P}_0\{|\delta_{2,T}| > a\} \longrightarrow 0$$

too.

Therefore, the likelihood ratio $Z_T(u) = L(u/\gamma T, X^T), u \geqslant 0$ is (under hypothesis $\mathscr{H}_0$) LAQ (see, for example, Le Cam and Yang, 2000), because

$$Z_T(u) \Longrightarrow Z(u) = \exp\left\{ -u \int_0^1 W(s) \, \mathrm{d}W(s) - \frac{u^2}{2} \int_0^1 W(s)^2 \, \mathrm{d}s \right\}. \tag{15}$$

Moreover, we have the convergence $l_T^*(u_T) - l_T(u_T) \to 0$ for any bounded sequence of $u_T \in \mathbb{U}_T$. Note that the random function $Z(u)$ is the likelihood ratio in the hypotheses testing problem

$$\mathscr{H}_0 : u = 0,$$

$$\mathscr{H}_1 : u > 0$$

by observations of Ornstein–Uhlenbeck process

$$\mathrm{d}Y(s) = -uY(s)\,\mathrm{d}s + \mathrm{d}W(s), \quad Y(0) = 0, \ \ 0 \leqslant s \leqslant 1 \tag{16}$$

under hypothesis $u = 0$.

This limit for the likelihood ratio under alternative can be obtained directly as follows. Let us denote

$$Y_T(s) = \frac{X_{sT} - sS_*T}{\sqrt{S_*T}}, \quad 0 \leqslant s \leqslant 1.$$

Then using the representation

$$X_t = S\int_0^t \psi(\vartheta_T[S_*r - X_r])\,\mathrm{d}r + M_t,$$

where M_t is local martingale and expansion of the function $\psi(\cdot)$ at the vicinity of 0 we obtain the equation

$$Y_T(s) = -u\int_0^s \frac{\dot\psi(g_v)}{\dot\psi(0)} Y_T(v)\,\mathrm{d}v + V_T(s), \quad Y_T(0) = 0, \ \ 0 \leqslant s \leqslant 1,$$

where $V_T(s)$ is local martingale and $g_v = (-\tilde u/\dot\psi(0)\sqrt{S_*T})Y_T(v) \to 0$. The central limit theorem for local martingales provides the convergence $V_T(s) \Longrightarrow W(s)$. Hence process (16) is the limit (in distribution) of $Y_T(s)$. Moreover from (7) we have

$$\Delta_T(X^T) = \frac{Y_T(1)}{2\sqrt{S_*T}} + \frac{1 - Y_T(1)^2}{2} \Longrightarrow \frac{1 - Y(1)^2}{2}.$$

This limit of the statistic $\Delta_T(X^T)$ follows from the Third Le Cam Lemma as well. Particularly, for any continuous bounded function $H(\cdot)$:

$$\mathbf{E}_u H(\Delta_T(X^T)) = \mathbf{E}_0[Z_T(u)H(\Delta_T(X^T))] \longrightarrow \mathbf{E}_0[Z(u)H(\Delta(W))] = \mathbf{E}_u H(\Delta(Y)),$$

where

$$\Delta(Y) = -\int_0^1 Y(s)\,\mathrm{d}Y(s) = \frac{1 - Y(1)^2}{2}.$$

Hence under alternative ($\vartheta_T = u_*/\gamma T$) we have the convergence

$$\beta_T(u_*, \phi_T^*) \longrightarrow \mathbf{P}_{u_*}\left\{|Y(1)| \leqslant z_{(1-\varepsilon)/2}\right\} = \mathbf{P}\left\{|W(1)| \leqslant z_{(1-\varepsilon)/2}\sqrt{\frac{2u_*}{1 - \mathrm{e}^{-2u_*}}}\right\}$$

because

$$Y(1) = \int_0^1 \mathrm{e}^{-u(1-s)}\,\mathrm{d}W(s) \sim \mathscr{N}\left(0, \frac{1 - \mathrm{e}^{-2u_*}}{2u_*}\right).$$

This proves (9).

Theorem 1 is *asymptotic in nature*, and it is interesting to see the powers of the score-function test for the moderate values of T and especially to compare them with the limit power functions. This can be done using numerical simulations.

We consider the model of Example 1 with $S_* = 1$ and $\psi(t) = \mathrm{e}^t$. This yields the intensity function

$$S(u, t, X_t) = \exp\left(\frac{u}{T}[t - X_t]\right), \quad u \geqslant 0, \ \ 0 \leqslant t \leqslant T.$$

In Fig. 1 we represent the power function of the score-function test ϕ_T^* of asymptotic size 0.05 given by

$$\beta_T(u, \phi_T^*) = \mathbf{P}_u\{\Delta_T(X^T) > a_{0.05}\}, \quad 0 \leqslant u \leqslant 20$$

for $T = 100$, 300 and 1000, as well as the limiting power function $\beta^*(\cdot)$ given by formula (9).

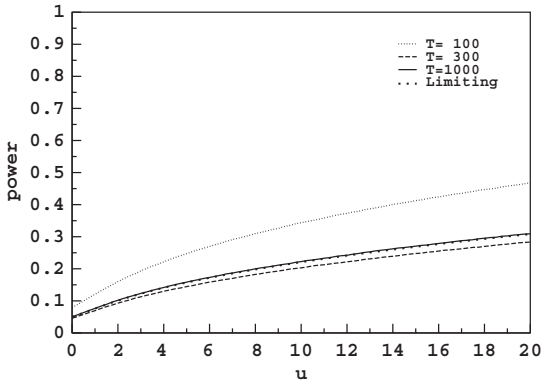

Fig. 1. Power of the score-function test.

The function $\beta_T(\cdot,\phi_T^*)$ is estimated in the following way. We simulate (for each value of u) $M=10^6$ trajectories X_j^T, $j=1,\dots,M$ of stress-release process of intensity $S(u,t,X_t)$ and calculate $\Delta_j=\Delta_T(X_j^T)$. Then we calculate the empirical frequency of accepting the alternative hypothesis

$$\frac{1}{M}\sum_{j=1}^{M}\mathbb{1}_{\{\Delta_j>a_{0.05}\}}\approx\beta_T(u,\phi_T^*).$$

Note that for $T=1000$ the limiting power function is practically attained. Note also that for $T=100$ the size of the test is 0.079 which explains the position of the corresponding curve.

Remind that score-function test is locally optimal (see Capon, 1961).

3. The likelihood ratio test and the Wald test

Let us study two other well-known tests: the *likelihood ratio test* $\bar{\phi}_T$ based on the maximum of the likelihood ratio function and the *Wald test* $\hat{\phi}_T$ based on the MLE $\hat{\vartheta}_T$.

Remind that the log-likelihood ratio formula is

$$\ln L(\vartheta,X^T)=\int_0^T\ln\psi(\vartheta(S_*t-X_{t-}))[\mathrm{d}X_t-S_*\,\mathrm{d}t]-\int_0^T[\psi(\vartheta(S_*t-X_{t-}))-1-\ln\psi(\vartheta(S_*t-X_{t-}))]S_*\,\mathrm{d}t$$

and the likelihood ratio test is based on the statistic

$$\delta_T(X^T)=\sup_{\vartheta\in\Theta}L(\vartheta,X^T),$$

where Θ is the set of values of ϑ under alternative. The test is given by the decision function

$$\bar{\phi}_T(X^T)=\mathbb{1}_{\{\delta_T(X^T)>\tilde{b}_\varepsilon\}},$$

where the threshold $\tilde{b}_\varepsilon$ is chosen from the condition $\bar{\phi}_T\in\mathcal{K}_\varepsilon$.

Note that $\delta_T(X^T)=L(\hat{\vartheta}_T,X^T)$ as well, where $\hat{\vartheta}_T$ is the maximum likelihood estimator of the parameter ϑ.

The reparametrization $\vartheta=\vartheta_T=u/\gamma T$ reduces the problem (2)–(3) to (4)–(5) and we have to precise the region of *local alternatives*. In the traditional approach of *locally asymptotically uniformly most powerful tests* (regular case, see Roussas, 1972), in order to check the optimality of a test ϕ_T we compare the power function $\beta_T(u,\phi_T)$ with the power function of the Neyman–Pearson test on the compacts $0\leqslant u\leqslant K$ for any $K>0$. For these values of u the alternatives are always *contiguous*. To consider the similar class of alternatives in our case is not reasonable because the constant $\tilde{b}_\varepsilon$ became dependent of K. Indeed if we take the test function

$$\bar{\phi}_T(X^T)=\mathbb{1}_{\{\sup_{0<u\leqslant K}Z_T(u)>\tilde{b}_\varepsilon\}},\quad Z_T(u)=L\left(\frac{u}{\gamma T},X^T\right),$$

then the condition $\bar{\phi}_T\in\mathcal{K}_\varepsilon$ implies $\tilde{b}_\varepsilon=\tilde{b}_\varepsilon(K)$. Therefore we suppose that $K=K_T=\sqrt{S_*T}/\ln T\to\infty$.

Finally, we have the following hypotheses testing problem:

$$\mathscr{H}_0 : u = 0, \tag{17}$$

$$\mathscr{H}_1 : u = u_* \in \mathbb{U}_T. \tag{18}$$

Therefore, to study

$$\bar{\phi}_T(X^T) = \mathbb{1}_{\{\sup_{u\in\mathbb{U}_T} Z_T(u) > \tilde{b}_\varepsilon\}}$$

we need to describe the asymptotics of its errors under hypothesis $\mathscr{H}_0$ and alternatives $\mathscr{H}_1$ with $\vartheta = u_*/\gamma T$, $u_* \in \mathbb{U}_T$.

Below

$$\Lambda(W) = \frac{\Delta(W)}{\sqrt{2\mathrm{J}(W)}}.$$

Theorem 2. *Let us suppose that condition $\mathscr{A}$ is fulfilled and the value b_ε is solution of the equation*

$$\mathbf{P}(\Lambda(W) > b_\varepsilon) = \varepsilon. \tag{19}$$

Then the test $\bar{\phi}_T$ with $\tilde{b}_\varepsilon = e^{b_\varepsilon^2}$ belongs to $\mathscr{K}_\varepsilon$ and its power function converges to the following limit:

$$\beta(u_*, \bar{\phi}_T) \longrightarrow \hat{\beta}(u_*) = \mathbf{P}\{\Lambda(Y_{u_*}) > b_\varepsilon\},$$

where

$$\Lambda(Y_{u_*}) = \frac{\Delta(Y_{u_*})}{\sqrt{2\mathrm{J}(Y_{u_*})}} = \frac{1 - Y_{u_*}(1)^2}{\sqrt{8\mathrm{J}(Y_{u_*})}}$$

and $Y_{u_*} = \{Y_{u_*}(s), 0 \leqslant s \leqslant 1\}$ is Ornstein–Uhlenbeck process (16) *with $u = u_*$.*

Proof. The log-likelihood process $l_T(u) = \ln Z_T(u)$ admits (under hypothesis $\mathscr{H}_0$) the representation (14)

$$_T(u) = u\Delta_T(X^T)(1 + \delta_{1,T}) - \frac{u^2}{2}\mathrm{J}_T(X^T)(1 + \delta_{2,T}), \tag{20}$$

where $\delta_{i,T} \to 0$ uniformly on $u \in \mathbb{U}_T$. Hence

$$\Lambda_T(X^T)^2 \equiv \sup_{u\in\mathbb{U}_T} l_T(u) \Longrightarrow \frac{\Delta(W)^2}{2\mathrm{J}(W)}$$

and we have

$$\mathbf{E}_0\bar{\phi}_T(X^T) = \mathbf{P}_0\left\{\sup_{u\in\mathbb{U}_T} l_T(u) > b_\varepsilon^2\right\} \longrightarrow \mathbf{P}(\Lambda(W) > b_\varepsilon) = \varepsilon.$$

Let us fix an alternative $u = u_*$. We have the convergence

$$\mathscr{L}_0\{\Lambda_T(X^T), l_T(u_*)\} \Longrightarrow \mathscr{L}\left\{\Lambda(W), u_*\Delta(W) - \frac{u_*^2}{2}\mathrm{J}(W)\right\}. \tag{21}$$

Convergence (21) allows us to apply Third Le Cam's Lemma as follows: for any bounded continuous function $H(\cdot)$:

$$\mathbf{E}_{u_*}H(\Lambda_T(X^T)) = \mathbf{E}_0[Z_T(u_*)H(\Lambda_T(X^T))] \longrightarrow \mathbf{E}_0[Z(u_*)H(\Lambda(W))] = \mathbf{E}_{u_*}H(\Lambda(Y_{u_*})).$$

Hence

$$\beta(u_*, \bar{\phi}_T) = \mathbf{P}_{u_*}\left\{\sup_{u\in\mathbb{U}_T} l_T(u) > b_\varepsilon^2\right\} \longrightarrow \mathbf{P}_{u_*}\{\Lambda(Y_{u_*}) > b_\varepsilon\}.$$

This completes the proof of Theorem 2. □

Let us note that the threshold b_ε is given implicitly as the solution of equation (19). In the following table we give some values of b_ε obtained using numerical simulations.

ε	0.01	0.02	0.03	0.04	0.05	0.1
b_ε	1.814	1.636	1.524	1.440	1.373	1.144

These thresholds are obtained by simulating $M = 10^7$ trajectories on $[0,1]$ of a standard Wiener process, calculating for each of them the quantity $\Lambda(W)$ and taking $(1-\varepsilon)M$-th greatest between them.

The next test usually studied in such hypotheses testing problems is the Wald test

$$\hat{\phi}_T(X^T) = \mathbb{1}_{\{\gamma T\hat{\vartheta}_T \geqslant c_\varepsilon\}},$$

where $\hat{\vartheta}_T$ is the maximum likelihood estimator of ϑ.

Below

$$\Gamma(W) = \frac{\Lambda(W)}{\mathrm{J}(W)}.$$

Theorem 3. *Let us suppose that condition $\mathscr{A}$ is fulfilled and the value c_ε is solution of the equation*

$$\mathbf{P}(\Gamma(W) > c_\varepsilon) = \varepsilon. \tag{22}$$

Then the test $\hat{\phi}_T$ belongs to $\mathscr{K}_\varepsilon$ and its power function for any alternative u_* converges to the following limit:

$$\beta(u_*, \hat{\phi}_T) \longrightarrow \hat{\beta}(u_*) = \mathbf{P}\{\Gamma(Y_{u_*}) > c_\varepsilon\},$$

where

$$\Gamma(Y_{u_*}) = \frac{\Lambda(Y_{u_*})}{\mathrm{J}(Y_{u_*})} = -u_* + \frac{\int_0^1 Y_{u_*}(s)\,\mathrm{d}W(s)}{\mathrm{J}(Y_{u_*})}$$

and Y_{u_*} is the same as in Theorem 2.

Proof. The proof follows immediately from representation (20), because

$$\begin{aligned}\mathbf{P}_0^{(T)}\{\gamma T\hat{\vartheta}_T \geqslant c_\varepsilon\} &= \mathbf{P}_0^{(T)}\left\{\sup_{0\leqslant u\leqslant c_\varepsilon} Z_T(u) < \sup_{u>c_\varepsilon, u\in\mathbb{U}_T} Z_T(u)\right\}\\ &\longrightarrow \mathbf{P}_0\left\{\sup_{0\leqslant u\leqslant c_\varepsilon} Z(u) < \sup_{u>c_\varepsilon} Z(u)\right\} = \mathbf{P}\{\Gamma(W) > c_\varepsilon\} = \varepsilon\end{aligned}$$

and (under alternative $u = u_*$)

$$\begin{aligned}\mathbf{P}_{u_*}^{(T)}\{\gamma T\hat{\vartheta}_T \geqslant c_\varepsilon\} &= \mathbf{P}_{u_*}^{(T)}\left\{\sup_{0\leqslant u\leqslant c_\varepsilon} Z_T(u) < \sup_{u>c_\varepsilon, u\in\mathbb{U}_T} Z_T(u)\right\}\\ &\longrightarrow \mathbf{P}_{u_*}\left\{\sup_{0\leqslant u\leqslant c_\varepsilon} Z(u) < \sup_{u>c_\varepsilon} Z(u)\right\} = \mathbf{P}\{\Gamma(Y_{u_*}) > c_\varepsilon\} = \hat{\beta}(u_*). \qquad \square\end{aligned}$$

As above, the threshold c_ε is given implicitly as the solution of equation (22). In the following table we give some values of c_ε obtained using numerical simulations.

ε	0.01	0.02	0.03	0.04	0.05	0.1
c_ε	13.692	11.224	9.803	8.806	8.042	5.719

These thresholds are obtained by simulating $M = 10^7$ trajectories on $[0,1]$ of a standard Wiener process, calculating for each of them the quantity $\Gamma(W)$ and taking $(1-\varepsilon)M$-th greatest between them.

4. Comparison of the tests

Remind that all these three tests ϕ^*_T, $\bar{\phi}_T$ and $\hat{\phi}_T$ in regular (LAN) case are asymptotically equivalent to the Neyman–Pearson test $\phi^0_{u,T}$ (with known alternative u) and hence are asymptotically uniformly most powerful. In our singular situation all of them have different asymptotic behavior and therefore it is interesting to compare their limit power functions

$$\beta^*(u)=\mathbf{P}_u\{\Delta(Y_u)>a_\varepsilon\},\quad \bar{\beta}(u)=\mathbf{P}_u\left\{\frac{\Delta(Y_u)}{\sqrt{2\mathrm{J}(Y_u)}}>b_\varepsilon\right\},$$

$$\hat{\beta}(u)=\mathbf{P}_u\left\{\frac{\Delta(Y_u)}{\mathrm{J}(Y_u)}>c_\varepsilon\right\},\quad \beta^0(u)=\mathbf{P}_u\left\{u\Delta(Y_u)-\frac{u^2}{2}\mathrm{J}(Y_u)>d_\varepsilon\right\}$$

of course, under condition that all of them belong to $\mathscr{K}_\varepsilon$. Our goal is to compare these quantities for the large values of u.

We have to study the distribution of the vector $(\Delta(Y_u),\mathrm{J}(Y_u))$, where

$$\Delta(Y_u)=-\int_0^1 Y_u(s)\,\mathrm{d}Y_u(s),\quad \mathrm{J}(Y_u)=\int_0^1 Y_u(s)^2\,\mathrm{d}s,$$

where Y_u is solution of the equation

$$\mathrm{d}Y_u(s)=-uY_u(s)\,\mathrm{d}s+\mathrm{d}W(s),\quad Y_u(0)=0,\ \ 0\leqslant s\leqslant 1.$$

Let us introduce the stochastic process $y_v=\sqrt{u}Y_u(v/u), 0\leqslant v\leqslant u$ (this transformation was introduced in Luschgy, 1994). Then we can write

$$\mathrm{d}y_v=-y_v\,\mathrm{d}v+\mathrm{d}w_v,\quad y_0=0,\ \ 0\leqslant v\leqslant u,$$

where $w_v=\sqrt{u}W(v/u)$ is a Wiener process and

$$\Delta(Y_u)=-u^{-1}\int_0^u y_v\,\mathrm{d}y_v\equiv\frac{\Delta_u}{u},\quad \mathrm{J}(Y_u)=u^{-2}\int_0^u y_v^2\,\mathrm{d}v\equiv\frac{\mathrm{J}_u}{u^2}$$

in obvious notation. Further, the process y_v is ergodic with the density of the invariant law $f(y)=\mathrm{e}^{-y^2}/\sqrt{\pi}$. Hence $\mathrm{J}_u\to\infty$ and

$$\frac{1}{u}\int_0^u y_v^2\,\mathrm{d}v\longrightarrow\frac{1}{2}.$$

Note that the distribution of the process y_v does not depend on u.

The constant $d_\varepsilon=d_\varepsilon(u)$ because it is defined by the equation

$$\mathbf{P}_0\left\{u\Delta(W)-\frac{u^2}{2}\mathrm{J}(W)>d_\varepsilon\right\}=\varepsilon.$$

For the large values of u this constant can be approximated as follows. We have (under hypothesis $\mathscr{H}_0$) as $u\to\infty$:

$$\mathbf{P}_0\left\{u\Delta(W)-\frac{u^2}{2}\mathrm{J}(W)>d_\varepsilon(u)\right\}=\mathbf{P}_0\left\{\int_0^1 W(s)^2\mathrm{d}s<-\frac{2d_\varepsilon(u)}{u^2}+\frac{2\Delta(W)}{u}\right\}$$
$$\longrightarrow\mathbf{P}_0\left\{\int_0^1 W(s)^2\mathrm{d}s<e_\varepsilon\right\}=\varepsilon,$$

where the constant e_ε is defined by the last equality. For example, if we take $\varepsilon=0.05$ then the numerical simulation gives us the value $e_{0.05}=0.056$. Therefore $d_\varepsilon(u)=-0.5e_\varepsilon u^2(1+\mathrm{o}(1))$. If we suppose that ε is small and try to solve the equation

$$\int_0^{e_\varepsilon} f_{\mathrm{J}}(x)\,\mathrm{d}x=\varepsilon,$$

where $f_{\mathrm{J}}(x)$ is the density function of the integral $\mathrm{J}(W)$, then we can easily see that $f_{\mathrm{J}}(0)=0$ and all its derivatives $f^{(k)}_{\mathrm{J}}(0)=0$, $k=1,2,\ldots$. Hence to see an approximative solution we need to calculate the large deviation probability of the following form (below $r=s/\sqrt{e_\varepsilon}$, $E=e_\varepsilon^{-1/2}\to\infty$):

$$\mathbf{P}_0\left\{e_\varepsilon^{-1}\int_0^1 W(s)^2\mathrm{d}s<1\right\}=\mathbf{P}_0\left\{\int_0^E W(r)^2\mathrm{d}r<1\right\}.$$

Below we put $d_\varepsilon(u)=-0.5e_\varepsilon u^2$.

We have the relations

$$\beta^*(u) = \mathbf{P}\{\Delta_u > u\, a_\varepsilon\} = \mathbf{P}\left\{\int_0^u y_v \,\mathrm{d}w_v < \mathrm{J}_u - a_\varepsilon u\right\},$$

$$\bar\beta(u) = \mathbf{P}\left\{\frac{\Delta_u}{\sqrt{2\mathrm{J}_u}} > b_\varepsilon\right\} = \mathbf{P}\left\{\int_0^u y_v \,\mathrm{d}w_v < \mathrm{J}_u - b_\varepsilon\sqrt{2\mathrm{J}_u}\right\},$$

$$\hat\beta(u) = \mathbf{P}\left\{\frac{\Delta_u}{\mathrm{J}_u} > \frac{c_\varepsilon}{u}\right\} = \mathbf{P}\left\{\int_0^u y_v \,\mathrm{d}w_v < \mathrm{J}_u - \frac{c_\varepsilon}{u}\mathrm{J}_u\right\},$$

$$\beta^0(u) = \mathbf{P}\left\{\Delta_u - \frac{\mathrm{J}_u}{2} > d_\varepsilon\right\} = \mathbf{P}\left\{\int_0^u y_v \,\mathrm{d}w_v < \frac{1}{2}\mathrm{J}_u + \frac{e_\varepsilon}{2}u^2\right\}.$$

Therefore the large values of u ($\mathrm{J}_u \sim u/2$):

$$\frac{1}{2}\mathrm{J}_u + \frac{e_\varepsilon}{2}u^2 > \mathrm{J}_u - \frac{c_\varepsilon}{u}\mathrm{J}_u > \mathrm{J}_u - b_\varepsilon\sqrt{2\mathrm{J}_u} > \mathrm{J}_u - a_\varepsilon u,$$

and finally

$$\beta^*(u) < \bar\beta(u) < \hat\beta(u) < \beta^0(u).$$

These inequalities are in accord with Swensen (1997).

Note that for small values of ε the constant a_ε is close to 0.5 (e.g., $a_{0.05} = 0.498$, $a_{0.01} = 0.49992$) and in this asymptotics the power of score-function test is

$$\beta^*(u) = \mathbf{P}\left\{\int_0^u y_v \,\mathrm{d}w_v < (0.5 - a_\varepsilon)\,u(1 + \mathrm{o}(1))\right\}.$$

Hence one can expect that in this case the score-function test has essentially smaller power than the others.

Now let us turn to numerical simulations of the limiting power functions. We aim to obtain the limiting power functions of all the three tests, as well as the Neyman–Pearson envelope, for the moderate values of u ($u \leqslant 15$).

Note that for the score-function test $\beta^*(u)$ can be computed directly using (9). However, the limiting power functions of the likelihood ratio and of the Wald tests are written as probabilities of some events related to Ornstein–Uhlenbeck process and can be obtained using numerical simulations.

For the likelihood ratio test we have

$$\bar\beta(u) = \mathbf{E}_u \mathbb{1}_{\{\Delta(Y_u) > b_\varepsilon\}} = \mathbf{E}_0 Z(u) \mathbb{1}_{\{\Delta(W) > b_\varepsilon\}},$$

where

$$Z(u) = \exp\left\{u\Delta(W) - \frac{u^2}{2}\mathrm{J}(W)\right\}.$$

So we simulate $M = 10^7$ trajectories $W_j = \{W_j(s),\ 0 \leqslant s \leqslant 1\}, j = 1, \dots, M$ of a standard Wiener process and calculate for each of them the quantities $\Delta_j = \Delta(W_j)$, $\mathrm{J}_j = \mathrm{J}(W_j)$, $\Lambda_j = \Delta_j/\mathrm{J}_j$ and (for each value of u) $Z_j(u) = \exp\{u\Delta_j - (u^2/2)\mathrm{J}_j\}$. Then we calculate the empirical mean

$$\frac{1}{M}\sum_{j=1}^M Z_j(u) \mathbb{1}_{\{\Lambda_j > b_\varepsilon\}} \approx \bar\beta(u).$$

For the Wald test we have similarly

$$\frac{1}{M}\sum_{j=1}^M Z_j(u) \mathbb{1}_{\{\Gamma_j > c_\varepsilon\}} \approx \hat\beta(u),$$

where $\Gamma_j = \Delta_j/\sqrt{2\mathrm{J}_j}$.

Finally, in order to compute the Neyman–Pearson envelope, we first approximate (for each value of u) the quantity $d_\varepsilon = d_\varepsilon(u)$ by the $(1-\varepsilon)M$-th greatest between the quantities $\ln Z_j(u)$, and then calculate

$$\frac{1}{M}\sum_{j=1}^M Z_j(u) \mathbb{1}_{\{\ln Z_j(u) > d_\varepsilon(u)\}} \approx \beta^0(u).$$

The results of these simulations for $\varepsilon = 0.05$ are presented in Fig. 2.

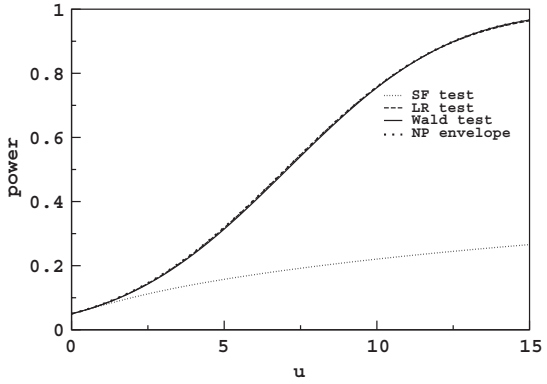

Fig. 2. Limiting powers for $\varepsilon = 0.05$.

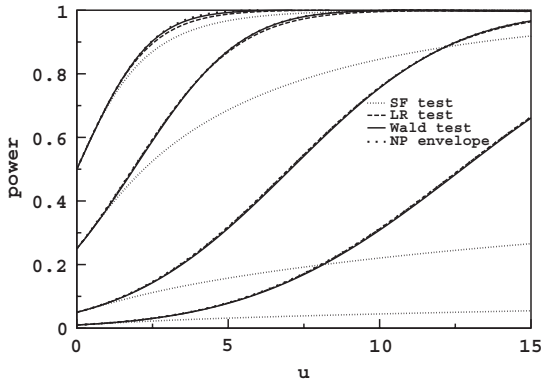

Fig. 3. Limiting powers for different values of ε.

Let us note here that in this case the power functions of the likelihood ratio test and of the Wald test are indistinguishable (from the point of view of numerical simulations) from the Neyman–Pearson envelope. This quite surprising fact was already mentioned in Eliott et al. (1996), who showed the similar pictures having 2×10^3 simulations. As we see from Fig. 2, with 10^7 simulations the curves are still indistinguishable. The situation is, however, different for bigger values of ε. The results of simulations for $\varepsilon = 0.01, 0.05, 0.25$ and 0.5 are presented in Fig. 3.

One can note that for big values of ε (e.g., $\varepsilon = 0.5$) the powers became more distinguishable, and that the asymptotically established ordering of the tests holds already for these moderate values of u. Note also that for the small values of ε (e.g., $\varepsilon = 0.01$ and 0.05) the curve of score-function test is essentially lower as expected.

5. Discussion

Remark 1. Note that alternatives $u = u_T \to \infty$ with $\vartheta_{u_T} \to 0$ are local but not contiguous. That means that the corresponding sequences of measures $(\mathbf{P}^{(T)}_{\vartheta_{u_T}}, \mathbf{P}^{(T)}_0), T \to \infty$ are not contiguous. Particularly, the second integral in the likelihood ratio formula tends to infinity:

$$\int_0^T [\psi(\vartheta_{u_T}(S_* t - X_{t-})) - 1 - \ln \psi(\vartheta_{u_T}(S_* t - X_{t-}))] S_* \, \mathrm{d}t \longrightarrow \infty.$$

In such situation the power function of any reasonable test tends to 1 and to compare tests we have to use, say, the large deviation principle. For example, the likelihood ratio test ϕ^*_T is consistent for the *local far alternatives* $\vartheta = v/\sqrt{S_* T}$, $v \in [v, V]$ where $0 < v < V < \infty$.

Indeed, under mild regularity conditions we can write

$$\begin{aligned}\mathbf{E}_v\phi_T^*(X^T)&=\mathbf{P}_0\left\{\sup_{v<v<V}L\left(\frac{v}{\sqrt{S_*T}},X^T\right)>c_\varepsilon\right\}\\&=\mathbf{P}_0\left\{\sup_{v<v<V}\left[\sqrt{S_*T}\int_0^1\ln\psi(vW_T(s))\,\mathrm{d}W_T(s)\right.\right.\\&\quad\left.\left.-S_*T\int_0^1[\psi(vW_T(s))-1-\ln\psi(vW_T(s))]\,\mathrm{d}s\right]>\ln c_\varepsilon\right\}\\&=\mathbf{P}_0\left\{\sup_{v<v<V}\left[\frac{1}{\sqrt{S_*T}}\int_0^1\ln\psi(vW_T(s))\,\mathrm{d}W_T(s)\right.\right.\\&\quad\left.\left.-\int_0^1[\psi(vW_T(s))-1-\ln\psi(vW_T(s))]\,\mathrm{d}s\right]>\frac{\ln c_\varepsilon}{S_*T}\right\}\\&\longrightarrow\mathbf{P}\left\{\inf_{v<v<V}\int_0^1[\psi(vW(s))-1-\ln\psi(vW(s))]\,\mathrm{d}s>0\right\}=1\end{aligned}$$

because the function $g(y)=y-1-\ln y>0$ for $y\neq1$ and $g(y)=0$ iff $y=1$.

Remark 2. Note that we can construct asymptotically uniformly most powerful test if we change the statement of the problem in the following way. Let us fix some $D>0$ and introduce the stopping time

$$\tau_D=\inf\left\{\tau:\int_0^\tau(S_*t-X_t)^2S_*\,\mathrm{d}t\geqslant D^2\right\}.$$

Then we consider the problem of testing hypotheses

$$\mathscr{H}_0:S(t,X_t)=S_*,$$

$$\mathscr{H}_1:S(t,X_t)=S_*\psi(\vartheta_D[S_*t-X_t]),\quad\vartheta_D=\frac{u}{\dot\psi(0)D}>0$$

by observations $X^{\tau_D}=\{X_t,0\leqslant t\leqslant\tau_D\}$ in the asymptotics $D\to\infty$. Now the likelihood ratio $Z_{\tau_D}(u)=L(u/\dot\psi(0)D,X^D)$ will be LAN:

$$Z_{\tau_D}(u)\Longrightarrow\exp\left\{u\zeta-\frac{u^2}{2}\right\},\quad\zeta\sim\mathscr{N}(0,1)$$

and the test $\hat\phi_{\tau_D}=\mathbb{1}_{\{\Delta_{\tau_D}(X^{\tau_D})>z_\varepsilon\}}$ where

$$\Delta_{\tau_D}(X^{\tau_D})=\frac{1}{D}\int_0^{\tau_D}(S_*t-X_{t-})[\mathrm{d}X_t-S_*\,\mathrm{d}t]$$

is locally asymptotically uniformly most powerful.

The proof follows from the central limit theorem for stochastic integrals and the standard arguments (for LAN families).

Remark 3. Note that these problems of hypotheses testing are similar to the corresponding problems of hypotheses testing for diffusion processes. In particular, let the observed process $X^T=\{X_t,0\leqslant t\leqslant T\}$ be diffusion

$$\mathrm{d}X_t=\psi(-\vartheta_T X_t)\,\mathrm{d}t+\sigma\,\mathrm{d}W_t,\quad X_0=0,\ \ 0\leqslant t\leqslant T,$$

where the function $\psi(0)=0$ is continuously differentiable at the point 0 and $\dot\psi(0)>0$. If we consider two hypotheses: $\vartheta=0$ and $\vartheta>0$ then the reparametrization

$$\vartheta_T=\frac{u\sigma}{\dot\psi(0)T}$$

provides local contiguous alternatives, i.e., the log-likelihood ratio in the problem

$$\mathscr{H}_0:u=0,$$

$$\mathscr{H}_1:u>0$$

has the limit

$$\ln L\left(\frac{u\sigma}{\dot\psi(0)T},X^T\right)\Longrightarrow-u\int_0^1W(s)\,\mathrm{d}W(s)-\frac{u^2}{2}\int_0^1W(s)^2\,\mathrm{d}s.$$

The score-function test based on the statistic

$$\Delta_T^*(X^T) = -\frac{1}{T}\int_0^T X_t\,\mathrm{d}X_t,$$

the likelihood ratio test and the Wald test have the same asymptotic properties as those described in Theorems 1–3.

For example, if $\psi(x)=x$, then we have the Wiener process (under hypothesis $\mathscr{H}_0$) against ergodic Ornstein–Uhlenbeck process under alternative $\mathscr{H}_1$.

Remark 4. We supposed above that the derivative of the function $\psi(x)$ at the point $x=0$ is not equal to 0, but sometimes it can be interesting to study the score-function and the likelihood ratio tests in the situations when the first $k-1$ derivatives with $k\geqslant 2$ are null.

Let us consider a stress-release process $X^T=\{X_t, 0\leqslant t\leqslant T\}$ with intensity function $S_*\psi(\vartheta(S_*t-X_t))$ such that $\psi(0)=1$ $\dot{\psi}(0)=0$ and $\ddot{\psi}(\cdot)\neq 0$ $(k=2)$. In this case the modifications have to be the following. Suppose that $\ddot{\psi}(0)>0$. To have LAQ family at the point $\vartheta=0$ we chose the reparametrization $\vartheta=\vartheta_u$:

$$\vartheta_u = \sqrt{\frac{2u}{\ddot{\psi}(0)}}(S_*T)^{-3/4},$$

which provides the limit

$$\ln L(\vartheta_u, X^T) \Longrightarrow u\int_0^1 W(s)^2\,\mathrm{d}W(s) - \frac{u^2}{2}\int_0^1 W(s)^4\,\mathrm{d}s.$$

Then in the hypotheses testing problem

$$\mathscr{H}_0 : u=0,$$

$$\mathscr{H}_1 : u>0$$

the score-function test $\hat{\psi}(X^T)=\mathbb{1}_{\{\Delta_T(X^T)>c_\varepsilon\}}$ is based on the statistic

$$\Delta_T(X^T) = \frac{1}{(S_*T)^{3/2}}\int_0^T (S_*t-X_t)^2[\mathrm{d}X_t - S_*\,\mathrm{d}t].$$

It is easy to see that under $\mathscr{H}_0$

$$\Delta_T(X^T) \Longrightarrow \frac{W(1)^3}{3} - \int_0^1 W(s)\,\mathrm{d}s.$$

Hence to chose the threshold c_ε we have to solve the following equation:

$$\frac{4}{3}\iint_{x^3-y>3c} \exp\left\{-2x^2+2xy-\frac{2}{3}y^2\right\}\mathrm{d}x\,\mathrm{d}y = \varepsilon$$

because $(W(1), 3\int_0^1 W(s)\,\mathrm{d}s)$ is Gaussian vector.

The cases $k>2$ can be treated in a similar way.

References

Capon, J., 1961. On asymptotic efficiency of locally optimum detectors. IRE Trans. Inform. Theory 7, 67–71.
Chan, N.H., Wei, C.Z., 1987. Asymptotic inference for nearly nonstationary AR(1) processes. Ann. Statist. 15, 1050–1063.
Dachian, S., Kutoyants, Yu.A., 2006. Hypotheses testing: Poisson versus self-exciting. Scand. J. Statist. 33, 391–408.
Daley, D.J., Vere-Jones, D., 2003. An Introduction to the Theory of Point Processes, vol. I, second ed. Springer, New York.
Eliott, G., Rothenberg, T.J., Stock, J.H., 1996. Efficient tests for an autoregressive unit root. Econometrika 64 (4), 813–836.
Feigin, P.D., 1979. Some comments concerning a curios singularity. J. Appl. Probab. 16, 440–444.
Gikhman, I.I., Skorokhod, A.V., 1969. Introduction to Theory of Random Processes. W.B. Saunders, Philadelphia.
Hawkes, A.G., 1972. Spectra for some mutually exciting point processes with associated variable. In: Lewis, P.A.W. (Ed.), Stochastic Point Processes. Wiley, New York.
Hayashi, T., 1986. Laws of large numbers in self-correcting point processes. Stochastic Process. Appl. 23, 319–326.
Isham, V., Westcott, M., 1979. A self-correcting point processes. Stochastic Process. Appl. 8, 335–347.
Knopov, L., 1971. A stochastic model for the occurrence of main sequence events. Rev. Geophys. Space Phys. 9, 175–178.
Kutoyants, Yu.A., 1984. Parameter Estimation for Stochastic Processes. Heldermann, Berlin.
Le Cam, L., Yang, G.L., 2000. Asymptotics in Statistics. second ed. Springer, New York.
Liptser, R.S., Shiryayev, A.N., 2001. Statistics of Random Processes. II. Applications. second ed. Springer, New York.

Lu, C., Harte, D., Bebbington, M., 1999. A linked stress release model for historical Japanese earthquakes: coupling among major seismic regions. Earth Planets Space 51, 907–916.
Luschgy, H., 1993. On a singularity occurring in a self-correcting point process model. Ann. Inst. Statist. Math. 45 (3), 445–452.
Luschgy, H., 1994. Asymptotic inference for semimartingale models with singular parameter points. J. Stochastic Plan. Inference 39, 155–186.
Ogata, Y., Vere-Jones, D., 1984. Inference for earthquake models: a self-correcting model. Stochastic Process. Appl. 17, 337–347.
Phillips, P.C.B., 1987. Toward a unified asymptotic theory for autoregression. Biometrika 74, 535–547.
Pitman, E.J.G., 1948. Lecture notes on nonparametric statistical inference. Lectures given for the University of North Carolina, Institute of Statistics.
Rigdon, S.E., Basu, A.P., 2000. Statistical Methods for Reliability of Repairable Systems. Wiley, New York.
Roussas, G.G., 1972. Contiguity of Probability Measures. Cambridge University Press, Cambridge.
Strasser, H., 1985. Mathematical Theory of Statistics. Walter de Gruyter, Berlin.
Swensen, A.R., 1997. Contiguity in nonstationary time series. In: Festschift for Lucien Le Cam.Springer, New York, pp. 377–384.
Vere-Jones, D., 1978. Earthquake prediction—a statistician's view. J. Phys. Earth 26, 129–146.
Vere-Jones, D., Ogata, Y., 1984. On the moments of a self-correcting point processes. J. Appl. Probab. 21, 335–342 Corr: 1987. ibid 24, 291.
Zheng, X., 1991. Ergodic theorems for stress release processes. Stochastic Process. Appl. 37, 239–258.

Journal of Statistical Planning and Inference 140 (2010) 2682–2692

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

journal homepage: www.elsevier.com/locate/jspi

On limiting likelihood ratio processes of some change-point type statistical models

Sergueï Dachian *

Laboratoire de Mathématiques, Université Blaise Pascal, 63177 Aubière CEDEX, France

ARTICLE INFO

Article history:
Received 24 September 2009
Received in revised form
2 March 2010
Accepted 11 March 2010
Available online 20 March 2010

MSC:
62F99
62M99

Keywords:
Non-regularity
Change-point
Limiting likelihood ratio process
Bayesian estimators
Maximum likelihood estimator
Limiting distribution
Limiting variance
Asymptotic efficiency

ABSTRACT

Different change-point type models encountered in parametric statistical inference give rise to different limiting likelihood ratio processes. In this paper we consider two such likelihood ratios. The first one is an exponential functional of a two-sided Poisson process driven by some parameter, while the second one is an exponential functional of a two-sided Brownian motion. We establish that for sufficiently small values of the parameter, the Poisson type likelihood ratio can be approximated by the Brownian type one. As a consequence, several statistically interesting quantities (such as limiting variances of different estimators) related to the first likelihood ratio can also be approximated by those related to the second one. Finally, we discuss the asymptotics for large values of the parameter and illustrate the results by numerical simulations.

1. Introduction

In this work we are interested by the asymptotic study of non-regular parametric statistical models. It is well known that in regular case the classical estimators (the maximum likelihood estimator and the Bayesian estimators) are consistent, asymptotically normal (with rate $1/\sqrt{n}$) and asymptotically efficient. In non-regular cases the situation essentially changes: usually these estimators are consistent, have different limiting distributions with a rate better than $1/\sqrt{n}$ and only the Bayesian estimators are asymptotically efficient. An exhaustive exposition of the parameter estimation theory in both regular and non-regular cases is given in the classical book by Ibragimov and Khasminskii (1981). They have developed a general theory of estimation based on the analysis of renormalized likelihood ratio. The approach consists in proving first that the renormalized likelihood ratio (with a properly chosen renormalization rate) weekly converges to a non-degenerate limit. Thereafter, the properties of the estimators are deduced. Finally, based on the estimators, one can also construct confidence intervals, tests, and so on. Note that this approach also provides the convergence of moments, allowing one to deduce equally the asymptotics of some statistically important quantities, such as the mean square errors of the estimators.

* Tel.: +33 4 73 40 53 74.
E-mail address: Serguei.Dachian@math.univ-bpclermont.fr

0378-3758/$ - see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.jspi.2010.03.030

More precisely, consider an observation X^n from the distribution $\mathbf{P}_\theta^n$. Suppose one have found a renormalization rate $\varphi_n \to 0$, such that the renormalized likelihood ratio process

$$Z_n(u) = \frac{d\mathbf{P}_{\theta+u\varphi_n}^n}{d\mathbf{P}_\theta^n}(X^n), \quad u \in \mathbb{R},$$

converges weekly (in a suitable functional space) to some non-degenerate limiting likelihood ratio process $Z(u)$, $u \in \mathbb{R}$. Then, one can usually deduce that the Bayesian estimators (here we consider quadratic loss function only) and the maximum likelihood estimator are consistent, converge with rate φ_n, and their limiting distributions are given by the random variables

$$\zeta = \frac{\int_{\mathbb{R}} uZ(u)\,du}{\int_{\mathbb{R}} Z(u)\,du} \quad \text{and} \quad \xi = \operatorname*{argsup}_{u\in\mathbb{R}} Z(u)$$

respectively. So, the quantiles of these random variables can be used to construct confidence intervals and tests based on the estimators. The second moments $B = \mathbf{E}\zeta^2$ and $M = \mathbf{E}\xi^2$ of these variables (also called limiting variances of the estimators) are also important, since usually the convergence of moments is also shown, and so the mean square errors of the Bayesian estimators and of the maximum likelihood estimator are $B\varphi_n^2(1+o(1))$ and $M\varphi_n^2(1+o(1))$, respectively. Moreover, usually it can also be shown that the Bayesian estimators are asymptotically efficient (have the smallest possible limiting variance), and so the quantity $E=B/M$ can be used as a measure of the (relative) asymptotic efficiency of the maximum likelihood estimator.

In regular models the renormalization rate is usually $\varphi_n = 1/\sqrt{n}$ and the limiting likelihood ratio is the same for different models (LAN property). In non-regular cases the rates are usually better (for example, in change-point situation $\varphi_n = 1/n$) and the limiting likelihood ratio processes can be different in different models. In this paper we consider two such limiting likelihood ratios arising in various change-point type models encountered in statistical inference.

The first one is the random process Z_ρ on $\mathbb{R}$ defined by

$$\ln Z_\rho(x) = \begin{cases} \rho\Pi_+(x) - x & \text{if } x \geq 0, \\ -\rho\Pi_-(-x) - x & \text{if } x \leq 0, \end{cases} \tag{1}$$

where $\rho > 0$, and Π_+ and Π_- are two independent Poisson processes on $\mathbb{R}_+$ with intensities $1/(e^\rho - 1)$ and $1/(1-e^{-\rho})$, respectively. We also consider the random variables

$$\zeta_\rho = \frac{\int_{\mathbb{R}} xZ_\rho(x)\,dx}{\int_{\mathbb{R}} Z_\rho(x)\,dx} \quad \text{and} \quad \xi_\rho = \operatorname*{argsup}_{x\in\mathbb{R}} Z_\rho(x) \tag{2}$$

related to this process, as well as their second moments $B_\rho = \mathbf{E}\zeta_\rho^2$ and $M_\rho = \mathbf{E}\xi_\rho^2$ and the quantity $E_\rho = B_\rho/M_\rho$.

The process Z_ρ (up to a linear time change) arises in various change-point type statistical models as the limiting likelihood ratio process. The main such model is the below detailed model of i.i.d. observations in the situation when their density has a jump (is discontinuous). Probably the first general result about this model goes back to Chernoff and Rubin (1956). Later, it was exhaustively studied by Ibragimov and Khasminskii (1981, Chapter 5) (see also their previous works Ibragimov and Khasminskii, 1970 and Ibragimov and Khasminskii, 1972).

Example 1. Consider the problem of estimation of the location parameter θ based on the observation $X^n = (X_1, \dots, X_n)$ of the i.i.d. sample from the density $f(x-\theta)$, where the known function f is smooth enough everywhere except at 0, and in 0 we have

$$0 \neq \lim_{x\uparrow 0} f(x) = a \neq b = \lim_{x\downarrow 0} f(x) \neq 0.$$

Denote $\mathbf{P}_\theta^n$ the distribution (corresponding to the parameter θ) of the observation X^n. As $n \to \infty$, the normalized likelihood ratio process of this model defined by

$$Z_n(u) = \frac{d\mathbf{P}_{\theta+u/n}^n}{d\mathbf{P}_\theta^n}(X^n) = \prod_{i=1}^n \frac{f\left(X_i - \theta - \frac{u}{n}\right)}{f(X_i - \theta)}$$

converges weakly in the space $\mathcal{D}_0(-\infty, +\infty)$ (the Skorohod space of functions on $\mathbb{R}$ without discontinuities of the second kind and vanishing at infinity) to the process $Z_{a,b}$ on $\mathbb{R}$ defined by

$$\ln Z_{a,b}(u) = \begin{cases} \ln\left(\frac{a}{b}\right)\Pi_b(u) - (a-b)u & \text{if } u \geq 0, \\ -\ln\left(\frac{a}{b}\right)\Pi_a(-u) - (a-b)u & \text{if } u \leq 0, \end{cases}$$

where Π_b and Π_a are two independent Poisson processes on $\mathbb{R}_+$ with intensities b and a, respectively. The limiting distributions of the Bayesian estimators and of the maximum likelihood estimator are given by

$$\zeta_{a,b}=\frac{\int_{\mathbb{R}}uZ_{a,b}(u)\,du}{\int_{\mathbb{R}}Z_{a,b}(u)\,du}\quad\text{and}\quad\xi_{a,b}=\operatorname*{argsup}_{u\in\mathbb{R}}Z_{a,b}(u)$$

respectively. The convergence of moments also holds, and the Bayesian estimators are asymptotically efficient. So, $\mathbf{E}\zeta^2_{a,b}$ and $\mathbf{E}\xi^2_{a,b}$ are the limiting variances of these estimators, and $\mathbf{E}\zeta^2_{a,b}/\mathbf{E}\xi^2_{a,b}$ is the asymptotic efficiency of the maximum likelihood estimator.

Now let us note, that up to a linear time change, the process $Z_{a,b}$ is nothing but the process Z_ρ with $\rho=\left|\ln\left(\frac{a}{b}\right)\right|$. Indeed, by putting $u=x/(a-b)$ we get

$$\ln Z_{a,b}(u)=\begin{cases}\ln\left(\frac{a}{b}\right)\Pi_b\left(\frac{x}{a-b}\right)-x & \text{if } \frac{x}{a-b}\geq 0,\\ -\ln\left(\frac{a}{b}\right)\Pi_a\left(-\frac{x}{a-b}\right)-x & \text{if } \frac{x}{a-b}\leq 0,\end{cases}=\ln Z_\rho(x)=\ln Z_\rho((a-b)u).$$

So, we have

$$\zeta_{a,b}=\frac{\zeta_\rho}{a-b}\quad\text{and}\quad\xi_{a,b}=\frac{\xi_\rho}{a-b},$$

and hence

$$\mathbf{E}\zeta^2_{a,b}=\frac{B_\rho}{(a-b)^2},\quad\mathbf{E}\xi^2_{a,b}=\frac{M_\rho}{(a-b)^2}\quad\text{and}\quad\frac{\mathbf{E}\zeta^2_{a,b}}{\mathbf{E}\xi^2_{a,b}}=E_\rho.$$

Some other models where the process Z_ρ arises occur in the statistical inference for inhomogeneous Poisson processes, in the situation when their intensity function has a jump (is discontinuous). In Kutoyants (1998, Chapter 5) (see also his previous work Kutoyants, 1984) one can find several examples, one of which is detailed below.

Example 2. Consider the problem of estimation of the location parameter $\theta\in]\alpha,\beta[$, $0<\alpha<\beta<\tau$, based on the observation X^T on $[0,T]$ of the Poisson process with τ- periodic strictly positive intensity function $S(t+\theta)$, where the known function S is smooth enough everywhere except at points $t^*+\tau k$, $k\in\mathbb{Z}$, with some $t^*\in[0,\tau]$, in which we have

$$0\neq\lim_{t\uparrow t^*}S(t)=S_-\neq S_+=\lim_{t\downarrow t^*}S(t)\neq 0.$$

Denote $\mathbf{P}^T_\theta$ the distribution (corresponding to the parameter θ) of the observation X^T. As $T\to\infty$, the normalized likelihood ratio process of this model defined by

$$Z_T(u)=\frac{d\mathbf{P}^T_{\theta+\frac{u}{T}}}{d\mathbf{P}^T_\theta}(X^T)=\exp\left\{\int_0^T\ln\frac{S_{\theta+u/T}(t)}{S_\theta(t)}dX(t)-\int_0^T[S_{\theta+u/T}(t)-S_\theta(t)]\,dt\right\}$$

converges weakly in the space $\mathcal{D}_0(-\infty,+\infty)$ to the process Z_{τ,S_-,S_+} on $\mathbb{R}$ defined by

$$\ln Z_{\tau,S_-,S_+}=\begin{cases}\ln\left(\frac{S_+}{S_-}\right)\Pi_{S_-}\left(\frac{u}{\tau}\right)-(S_+-S_-)\frac{u}{\tau} & \text{if } u\geq 0,\\ -\ln\left(\frac{S_+}{S_-}\right)\Pi_{S_+}\left(-\frac{u}{\tau}\right)-(S_+-S_-)\frac{u}{\tau} & \text{if } u\leq 0,\end{cases}$$

where Π_{S_-} and Π_{S_+} are two independent Poisson processes on $\mathbb{R}_+$ with intensities S_- and S_+, respectively. The limiting distributions of the Bayesian estimators and of the maximum likelihood estimator are given by

$$\zeta_{\tau,S_-,S_+}=\frac{\int_{\mathbb{R}}uZ_{\tau,S_-,S_+}(u)\,du}{\int_{\mathbb{R}}Z_{\tau,S_-,S_+}(u)\,du}\quad\text{and}\quad\xi_{\tau,S_-,S_+}=\operatorname*{argsup}_{u\in\mathbb{R}}Z_{\tau,S_-,S_+}(u)$$

respectively. The convergence of moments also holds, and the Bayesian estimators are asymptotically efficient. So, $\mathbf{E}\zeta^2_{\tau,S_-,S_+}$ and $\mathbf{E}\xi^2_{\tau,S_-,S_+}$ are the limiting variances of these estimators, and $\mathbf{E}\zeta^2_{\tau,S_-,S_+}/\mathbf{E}\xi^2_{\tau,S_-,S_+}$ is the asymptotic efficiency of the maximum likelihood estimator.

Now let us note, that up to a linear time change, the process Z_{τ,S_-,S_+} is nothing but the process Z_ρ with $\rho=|\ln(\frac{S_+}{S_-})|$. Indeed, by putting $u=\tau x/(S_+-S_-)$ we get

$$Z_{\tau,S_-,S_+}(u)=Z_\rho(x)=Z_\rho\left(\frac{S_+-S_-}{\tau}u\right).$$

So, we have

$$\zeta_{\tau,S_-,S_+}=\frac{\tau\zeta_\rho}{S_+-S_-}\quad\text{and}\quad\zeta_{\tau,S_-,S_+}=\frac{\tau\xi_\rho}{S_+-S_-},$$

and hence

$$\mathbf{E}\zeta^2_{\tau,S_-,S_+}=\frac{\tau^2 B_\rho}{(S_+-S_-)^2},\quad \mathbf{E}\xi^2_{\tau,S_-,S_+}=\frac{\tau^2 M_\rho}{(S_+-S_-)^2}\quad\text{and}\quad\frac{\mathbf{E}\zeta^2_{\tau,S_-,S_+}}{\mathbf{E}\xi^2_{\tau,S_-,S_+}}=E_\rho.$$

The second limiting likelihood ratio process considered in this paper is the random process

$$Z_0(x)=\exp\{W(x)-\tfrac{1}{2}|x|\},\quad x\in\mathbb{R},\tag{3}$$

where W is a standard two-sided Brownian motion. We also consider the random variables

$$\zeta_0=\frac{\int_{\mathbb{R}}xZ_0(x)\,dx}{\int_{\mathbb{R}}Z_0(x)\,dx}\quad\text{and}\quad\xi_0=\operatorname*{argsup}_{x\in\mathbb{R}}Z_0(x)\tag{4}$$

related to this process, as well as their second moments $B_0=\mathbf{E}\zeta_0^2$ and $M_0=\mathbf{E}\xi_0^2$ and the quantity $E_0=B_0/M_0$.

The models where the process Z_0 arises occur in various fields of statistical inference. A well-known example is the below detailed model of a discontinuous signal in a white Gaussian noise exhaustively studied by Ibragimov and Khasminskii (1981, Chapter 7.2) (see also their previous work Ibragimov and Khasminskii, 1975), but one can also cite change-point type models of dynamical systems with small noise (see Kutoyants, 1984 and Kutoyants (1994, Chapter 5)), those of ergodic diffusion processes (see Kutoyants, 2004, Chapter 3), a change-point type model of delay equations (see Küchler and Kutoyants, 2000), an i.i.d. change-point type model (see Deshayes and Picard, 1984), a model of a discontinuous periodic signal in a time inhomogeneous diffusion (see Höpfner and Kutoyants, 2009), and so on.

Example 3. Consider the problem of estimation of the location parameter $\theta\in]\alpha,\beta[$, $0<\alpha<\beta<1$, based on the observation X^ε on $[0,1]$ of the random process satisfying the stochastic differential equation

$$dX^\varepsilon(t)=\frac{1}{\varepsilon}S(t-\theta)\,dt+dW(t),$$

where W is a standard Brownian motion, and S is a known function having a bounded derivative on $]-1,0[\cup]0,1[$ and satisfying

$$\lim_{t\uparrow 0}S(t)-\lim_{t\downarrow 0}S(t)=r\neq 0.$$

Denote $\mathbf{P}^\varepsilon_\theta$ the distribution (corresponding to the parameter θ) of the observation X^ε. As $\varepsilon\to 0$, the normalized likelihood ratio process of this model defined by

$$Z_\varepsilon(u)=\frac{d\mathbf{P}^\varepsilon_{\theta+\varepsilon^2u}}{d\mathbf{P}^\varepsilon_\theta}(X^\varepsilon)=\exp\left\{\frac{1}{\varepsilon}\int_0^1[S(t-\theta-\varepsilon^2u)-S(t-\theta)]\,dW(t)\quad-\frac{1}{2\varepsilon^2}\int_0^1[S(t-\theta-\varepsilon^2u)-S(t-\theta)]^2\,dt\right\}$$

converges weakly in the space $\mathcal{C}_0(-\infty,+\infty)$ (the space of continuous functions vanishing at infinity equipped with the supremum norm) to the process $Z_0(r^2u)$, $u\in\mathbb{R}$. The limiting distributions of the Bayesian estimators and of the maximum likelihood estimator are $r^{-2}\zeta_0$ and $r^{-2}\xi_0$, respectively. The convergence of moments also holds, and the Bayesian estimators are asymptotically efficient. So, $r^{-4}B_0$ and $r^{-4}M_0$ are the limiting variances of these estimators, and E_0 is the asymptotic efficiency of the maximum likelihood estimator.

Let us also note that Terent'yev (1968) determined explicitly the distribution of ξ_0 and calculated the constant $M_0=26$. These results were taken up by Ibragimov and Khasminskii (1981, Chapter 7.3), where by means of numerical simulation they equally showed that $B_0=19.5\pm 0.5$, and so $E_0=0.73\pm 0.03$. Later in Golubev (1979), Golubev expressed B_0 in terms of the second derivative (with respect to a parameter) of an improper integral of a composite function of modified Hankel and Bessel functions. Finally in Rubin and Song (1995), Rubin and Song obtained the exact values $B_0=16\zeta(3)$ and $E_0=8\zeta(3)/13$, where ζ is Riemann's zeta function defined by

$$\zeta(s)=\sum_{n=1}^{\infty}\frac{1}{n^s}.$$

The random variables ζ_ρ and ξ_ρ and the quantities B_ρ, M_ρ and E_ρ, $\rho>0$, are much less studied. One can cite Pflug (1993) for some results about the distribution of the random variables

$$\operatorname*{argsup}_{x\in\mathbb{R}_+}Z_\rho(x)\quad\text{and}\quad\operatorname*{argsup}_{x\in\mathbb{R}_-}Z_\rho(x)$$

related to ξ_ρ.

In this paper we establish that the limiting likelihood ratio processes Z_ρ and Z_0 are related. More precisely, we show that as $\rho\to 0$, the process $Z_\rho(y/\rho)$, $y\in\mathbb{R}$, converges weakly in the space $\mathcal{D}_0(-\infty,+\infty)$ to the process Z_0. So, the random variables $\rho\zeta_\rho$ and $\rho\xi_\rho$ converge weakly to the random variables ζ_0 and ξ_0, respectively. We show equally that the convergence of moments of these random variables holds, so in particular $\rho^2B_\rho\to 16\zeta(3)$, $\rho^2M_\rho\to 26$ and $E_\rho\to 8\zeta(3)/13$. Besides their theoretical interest, these results allow one, for example, to construct tests and confidence intervals on the base of the distributions of ζ_0 and ξ_0 (rather than on the base of much less known distributions of ζ_ρ and ξ_ρ) in models having the process Z_ρ with a small ρ as a limiting likelihood ratio. Also, the limiting variances of the estimators and the asymptotic

efficiency of the maximum likelihood estimator can be approximated as

$$B_\rho \approx \frac{16\zeta(3)}{\rho^2}, \quad M_\rho \approx \frac{26}{\rho^2} \quad \text{and} \quad E_\rho \approx \frac{8\zeta(3)}{13}$$

in such models.

These are the main results of the present paper, and they are presented in Section 2, where we also briefly discuss the second possible asymptotics $\rho \to +\infty$ and present some numerical simulations of the quantities B_ρ, M_ρ and E_ρ for $\rho \in]0,\infty[$. Finally, the proofs are carried out in Section 3.

2. Main results and numerical simulations

Consider the process $X_\rho(y) = Z_\rho(y/\rho)$, $y \in \mathbb{R}$, where $\rho > 0$ and Z_ρ is defined by (1). Note that

$$\frac{\int_{\mathbb{R}} y X_\rho(y)\,dy}{\int_{\mathbb{R}} X_\rho(y)\,dy} = \rho\zeta_\rho \quad \text{and} \quad \operatorname*{argsup}_{y\in\mathbb{R}} X_\rho(y) = \rho\xi_\rho,$$

where the random variables ζ_ρ and ξ_ρ are defined by (2). Remind also the process Z_0 on $\mathbb{R}$ defined by (3) and the random variables ζ_0 and ξ_0 defined by (4). Recall finally the quantities $B_\rho = \mathbf{E}\zeta_\rho^2$, $M_\rho = \mathbf{E}\xi_\rho^2$, $E_\rho = B_\rho/M_\rho$, $B_0 = \mathbf{E}\zeta_0^2 = 16\zeta(3)$, $M_0 = \mathbf{E}\xi_0^2 = 26$ and $E_0 = B_0/M_0 = 8\zeta(3)/13$. Now we can state the main result of the present paper.

Theorem 1. *The process X_ρ converges weakly in the space $\mathcal{D}_0(-\infty,+\infty)$ to the process Z_0 as $\rho \to 0$. In particular, the random variables $\rho\zeta_\rho$ and $\rho\xi_\rho$ converge weakly to the random variables ζ_0 and ξ_0, respectively. Moreover, for any $k > 0$ we have*

$$\rho^k \mathbf{E}\zeta_\rho^k \to \mathbf{E}\zeta_0^k \quad \textit{and} \quad \rho^k \mathbf{E}\xi_\rho^k \to \mathbf{E}\xi_0^k,$$

and in particular $\rho^2 B_\rho \to 16\zeta(3)$, $\rho^2 M_\rho \to 26$ and $E_\rho \to 8\zeta(3)/13$.

This theorem will be proved in the next section, but before let us discuss the second possible asymptotics $\rho \to +\infty$. One can show that in this case, the process Z_ρ converges weakly in the space $\mathcal{D}_0(-\infty,+\infty)$ to the process $Z_\infty(x) = e^{-x}\mathbb{1}_{\{x>\eta\}}$, $x \in \mathbb{R}$, where η is a negative exponential random variable with $\mathbf{P}\{\eta < t\} = e^t$, $t \le 0$. So, the random variables ζ_ρ and ξ_ρ converge weakly to the random variables

$$\zeta_\infty = \frac{\int_{\mathbb{R}} x Z_\infty(x)\,dx}{\int_{\mathbb{R}} Z_\infty(x)\,dx} = \eta + 1 \quad \text{and} \quad \xi_\infty = \operatorname*{argsup}_{x\in\mathbb{R}} Z_\infty(x) = \eta$$

respectively. One can equally show that, moreover, for any $k > 0$ we have

$$\mathbf{E}\zeta_\rho^k \to \mathbf{E}\zeta_\infty^k \quad \text{and} \quad \mathbf{E}\xi_\rho^k \to \mathbf{E}\xi_\infty^k,$$

and in particular, denoting $B_\infty = \mathbf{E}\zeta_\infty^2$, $M_\infty = \mathbf{E}\xi_\infty^2$ and $E_\infty = B_\infty/M_\infty$, we finally have $B_\rho \to B_\infty = \mathbf{E}(\eta+1)^2 = 1$, $M_\rho \to M_\infty = \mathbf{E}\eta^2 = 2$ and $E_\rho \to E_\infty = 1/2$.

Let us note that these convergences are natural, since the process Z_∞ can be considered as a particular case of the process Z_ρ with $\rho = +\infty$ if one admits the convention $+\infty \cdot 0 = 0$.

Note also that the process Z_∞ (up to a linear time change) is the limiting likelihood ratio process of Model 1 (Model 2) in the situation when $a \cdot b = 0$ ($S_- \cdot S_+ = 0$). In this case, the variables $\zeta_\infty = \eta + 1$ and $\xi_\infty = \eta$ (up to a multiplicative constant) are the limiting distributions of the Bayesian estimators and of the maximum likelihood estimator, respectively. In particular, $B_\infty = 1$ and $M_\infty = 2$ (up to the square of the above multiplicative constant) are the limiting variances of these estimators, and the Bayesian estimators being asymptotically efficient, $E_\infty = 1/2$ is the asymptotic efficiency of the maximum likelihood estimator.

To conclude this section, let us present some numerical simulations of the quantities B_ρ, M_ρ and E_ρ for $\rho \in]0,\infty[$. Besides giving approximate values of these quantities, the simulation results illustrate both the asymptotics

$$B_\rho \sim \frac{B_0}{\rho^2}, \quad M_\rho \sim \frac{M_0}{\rho^2} \quad \text{and} \quad E_\rho \to E_0 \quad \text{as} \quad \rho \to 0,$$

with $B_0 = 16\zeta(3) \approx 19.2329$, M_0=26 and $E_0 = 8\zeta(3)/13 \approx 0.7397$, and

$$B_\rho \to B_\infty, \quad M_\rho \to M_\infty \quad \text{and} \quad E_\rho \to E_\infty \quad \text{as} \quad \rho \to \infty,$$

with $B_\infty = 1$, $M_\infty = 2$ and $E_\infty = 0.5$.

First, we simulate the events $x_1, x_2, \ldots$ of the Poisson process Π_+ (with the intensity $1/(e^\rho - 1)$), and the events $x'_1, x'_2, \ldots$ of the Poisson process Π_- (with the intensity $1/(1-e^{-\rho})$).

Then we calculate

$$\zeta_\rho = \frac{\int_{\mathbb{R}} x Z_\rho(x)\,dx}{\int_{\mathbb{R}} Z_\rho(x)\,dx} = \frac{\sum_{i=1}^\infty x_i e^{\rho i - x_i} + \sum_{i=1}^\infty e^{\rho i - x_i} - \sum_{i=1}^\infty x'_i e^{\rho - \rho i + x'_i} + \sum_{i=1}^\infty e^{\rho - \rho i + x'_i}}{\sum_{i=1}^\infty e^{\rho i - x_i} + \sum_{i=1}^\infty e^{\rho - \rho i + x'_i}}$$

and

$$\xi_\rho = \operatorname*{argsup}_{x\in\mathbb{R}} Z_\rho(x) = \begin{cases} x_k & \text{if } \rho k - x_k > \rho - \rho\ell + x'_\ell, \\ -x'_\ell & \text{otherwise,} \end{cases}$$

where

$$k = \operatorname*{argmax}_{i\geq 1}(\rho i - x_i) \quad \text{and} \quad \ell = \operatorname*{argmax}_{i\geq 1}(\rho - \rho i + x'_i),$$

so that

$$x_k = \operatorname*{argsup}_{x\in\mathbb{R}_+} Z_\rho(x) \quad \text{and} \quad -x'_\ell = \operatorname*{argsup}_{x\in\mathbb{R}_-} Z_\rho(x).$$

Finally, repeating these simulations 10^7 times (for each value of ρ), we approximate $B_\rho = \mathbf{E}\zeta_\rho^2$ and $M_\rho = \mathbf{E}\xi_\rho^2$ by the empirical second moments, and $E_\rho = B_\rho/M_\rho$ by their ratio.

The results of the numerical simulations are presented in Figs. 1 and 2. The $\rho\to 0$ asymptotics of B_ρ and M_ρ can be observed in Fig. 1, where besides these functions we also plotted the functions $\rho^2 B_\rho$ and $\rho^2 M_\rho$, making apparent the constants $B_0 \approx 19.2329$ and $M_0 = 26$.

In Fig. 2 we use a different scale on the vertical axis to better illustrate the $\rho\to\infty$ asymptotics of B_ρ and M_ρ, as well as both the asymptotics of E_ρ. Note that the function E_ρ appear to be decreasing, so we can conjecture that bigger is ρ, smaller is the efficiency of the maximum likelihood estimator, and so this efficiency is always between $E_\infty = 0.5$ and $E_0 \approx 0.7397$.

3. Proofs

The results concerning the random variable ζ_ρ are direct consequence of Ibragimov and Khasminskii (1981, Theorem 1.10.2) and the following three lemmas.

Lemma 2. *The finite-dimensional distributions of the process X_ρ converge to those of Z_0 as $\rho\to 0$.*

Lemma 3. *For all $\rho > 0$ and all $y_1, y_2 \in \mathbb{R}$ we have*

$$\mathbf{E}|X_\rho^{1/2}(y_1) - X_\rho^{1/2}(y_2)|^2 \leq \tfrac{1}{4}|y_1 - y_2|.$$

Lemma 4. *For any $c \in]0, 1/8[$ we have*

$$\mathbf{E}X_\rho^{1/2}(y) \leq \exp(-c|y|)$$

for all sufficiently small ρ and all $y \in \mathbb{R}$.

Note that these lemmas are not sufficient to establish the weak convergence of the process X_ρ in the space $\mathcal{D}_0(-\infty, +\infty)$ and the results concerning the random variable ξ_ρ. However, the increments of the process $\ln X_\rho$ being independent, the convergence of its restrictions (and hence of those of X_ρ) on finite intervals $[A,B] \subset \mathbb{R}$ (that is, convergence in the Skorohod space $\mathcal{D}[A,B]$ of functions on [A,B] without discontinuities of the second kind) follows from Gihman and Skorohod (1974, Theorem 6.5.5), Lemma 2 and the following lemma.

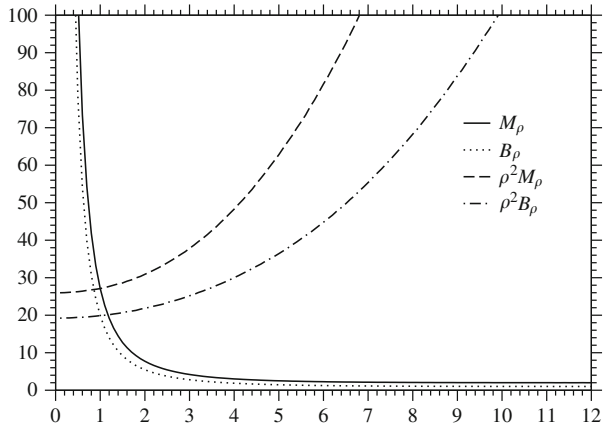

Fig. 1. B_ρ and M_ρ ($\rho\to 0$ asymptotics).

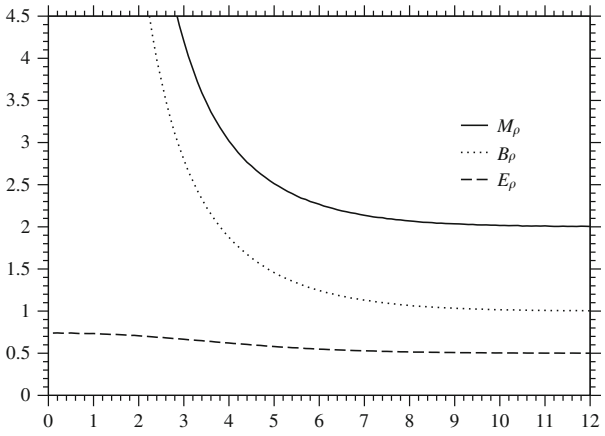

Fig. 2. B_ρ, M_ρ ($\rho\to\infty$ asymptotics) and E_ρ (both asymptotics).

Lemma 5. *For any $\varepsilon>0$ we have*

$$\lim_{h\to 0}\lim_{\rho\to 0}\sup_{|y_1-y_2|<h}\mathbf{P}\{|\ln X_\rho(y_1)-\ln X_\rho(y_2)|>\varepsilon\}=0.$$

Now, Theorem 1 follows from the following estimate on the tails of the process X_ρ by standard argument (see, for example, Ibragimov and Khasminskii, 1981).

Lemma 6. *For any $b\in]0,3/40[$ we have*

$$\mathbf{P}\left\{\sup_{|y|>A}X_\rho(y)>e^{-bA}\right\}\le 4e^{-bA}$$

for all sufficiently small ρ and all $A>0$.

So, it remains to prove the lemmas. We start with Lemma 2. Note that the restrictions of the process $\ln X_\rho$ (as well as those of the process $\ln Z_0$) on $\mathbb{R}_+$ and on $\mathbb{R}_-$ are mutually independent processes with stationary and independent increments. So, to obtain the convergence of all the finite-dimensional distributions, it is sufficient to show the convergence of one-dimensional distributions only, that is,

$$\ln X_\rho(y)\Rightarrow \ln Z_0(y)=W(y)-\frac{|y|}{2}=\mathcal{N}\left(-\frac{|y|}{2},|y|\right)$$

for all $y\in\mathbb{R}$. Here and in the sequel "$\Rightarrow$" denotes the weak convergence of the random variables, and $\mathcal{N}(m,V)$ denotes a "generic" random variable distributed according to the normal law with mean m and variance V.

Let $y>0$. Then, noting that $\Pi_+\left(\frac{y}{\rho}\right)$ is a Poisson random variable of parameter $\lambda=y/\rho(e^\rho-1)\to\infty$, we have

$$\ln X_\rho(y)=\rho\Pi_+\left(\frac{y}{\rho}\right)-\frac{y}{\rho}=\rho\sqrt{\frac{y}{\rho(e^\rho-1)}}\frac{\Pi_+\left(\frac{y}{\rho}\right)-\lambda}{\sqrt{\lambda}}+\frac{y}{e^\rho-1}-\frac{y}{\rho}=\sqrt{y}\sqrt{\frac{\rho}{e^\rho-1}}\frac{\Pi_+\left(\frac{y}{\rho}\right)-\lambda}{\sqrt{\lambda}}-y\frac{e^\rho-1-\rho}{\rho(e^\rho-1)}\Rightarrow\mathcal{N}\left(-\frac{y}{2},y\right),$$

since

$$\frac{\rho}{e^\rho-1}=\frac{\rho}{\rho+o(\rho)}\to 1,\quad \frac{e^\rho-1-\rho}{\rho(e^\rho-1)}=\frac{\rho^2/2+o(\rho^2)}{\rho(\rho+o(\rho))}\to\frac{1}{2}$$

and

$$\frac{\Pi_+\left(\frac{y}{\rho}\right)-\lambda}{\sqrt{\lambda}}\Rightarrow\mathcal{N}(0,1).$$

Similarly, for $y<0$ we have

$$\ln X_\rho(y)=-\rho\Pi_-\left(\frac{-y}{\rho}\right)-\frac{y}{\rho}=\rho\sqrt{\frac{-y}{\rho(1-e^{-\rho})}}\frac{\lambda'-\Pi_-\left(\frac{-y}{\rho}\right)}{\sqrt{\lambda'}}-\frac{-y}{1-e^{-\rho}}-\frac{y}{\rho}$$

$$= \sqrt{-y}\sqrt{\frac{\rho}{1-e^{-\rho}}}\,\frac{\lambda'-\Pi_-\left(\frac{-y}{\rho}\right)}{\sqrt{\lambda'}} + y\frac{e^{-\rho}-1+\rho}{\rho(1-e^{-\rho})} \Rightarrow \mathcal{N}\left(\frac{y}{2},-y\right),$$

and so Lemma 2 is proved.

Now we turn to the proof of Lemma 4 (we will prove Lemma 3 just after). For $y>0$ we can write

$$\mathbf{E}X_\rho^{1/2}(y) = \mathbf{E}\exp\left(\frac{\rho}{2}\Pi_+\left(\frac{y}{\rho}\right)-\frac{y}{2\rho}\right) = \exp\left(-\frac{y}{2\rho}\right)\mathbf{E}\exp\left(\frac{\rho}{2}\Pi_+\left(\frac{y}{\rho}\right)\right).$$

Note that $\Pi_+\left(\frac{y}{\rho}\right)$ is a Poisson random variable of parameter $\lambda = y/\rho(e^\rho-1)$ with moment generating function $M(t)=\exp(\lambda(e^t-1))$. So, we get

$$\mathbf{E}X_\rho^{1/2}(y) = \exp\left(-\frac{y}{2\rho}\right)\exp\left(\frac{y}{\rho(e^\rho-1)}(e^{\rho/2}-1)\right) = \exp\left(-\frac{y}{2\rho}+\frac{y}{\rho(e^{\rho/2}+1)}\right) = \exp\left(-y\frac{e^{\rho/2}-1}{2\rho(e^{\rho/2}+1)}\right)$$
$$= \exp\left(-y\frac{e^{\rho/4}-e^{-\rho/4}}{2\rho(e^{\rho/4}+e^{-\rho/4})}\right) = \exp\left(-y\frac{\tanh(\rho/4)}{2\rho}\right).$$

For $y<0$ we obtain similarly

$$\mathbf{E}X_\rho^{1/2}(y) = \mathbf{E}\exp\left(-\frac{\rho}{2}\Pi_-\left(\frac{-y}{\rho}\right)-\frac{y}{2\rho}\right) = \exp\left(-\frac{y}{2\rho}\right)\exp\left(\frac{-y}{\rho(1-e^{-\rho})}(e^{-\rho/2}-1)\right)$$
$$= \exp\left(-\frac{y}{2\rho}+\frac{y}{\rho(1+e^{-\rho/2})}\right) = \exp\left(y\frac{1-e^{-\rho/2}}{2\rho(1+e^{-\rho/2})}\right) = \exp\left(y\frac{\tanh(\rho/4)}{2\rho}\right).$$

Thus, for all $y\in\mathbb{R}$ we have

$$\mathbf{E}X_\rho^{1/2}(y) = \exp\left(-|y|\frac{\tanh(\rho/4)}{2\rho}\right), \tag{5}$$

and since

$$\frac{\tanh(\rho/4)}{2\rho} = \frac{\rho/4+o(\rho)}{2\rho} \to \frac{1}{8}$$

as $\rho\to 0$, for any $c\in]0,1/8[$ we have $\mathbf{E}X_\rho^{1/2}(y)\le\exp(-c|y|)$ for all sufficiently small ρ and all $y\in\mathbb{R}$. Lemma 4 is proved.

Further we verify Lemma 3. We first consider the case $y_1,y_2\in\mathbb{R}_+$ (say $y_1\ge y_2$). Using (5) and taking into account the stationarity and the independence of the increments of the process $\ln X_\rho$ on $\mathbb{R}_+$, we can write

$$\mathbf{E}|X_\rho^{1/2}(y_1)-X_\rho^{1/2}(y_2)|^2 = \mathbf{E}X_\rho(y_1)+\mathbf{E}X_\rho(y_2)-2\mathbf{E}X_\rho^{1/2}(y_1)X_\rho^{1/2}(y_2) = 2-2\mathbf{E}X_\rho(y_2)\mathbf{E}\frac{X_\rho^{1/2}(y_1)}{X_\rho^{1/2}(y_2)}$$
$$= 2-2\mathbf{E}X_\rho^{1/2}(y_1-y_2) = 2-2\exp\left(-|y_1-y_2|\frac{\tanh(\rho/4)}{2\rho}\right) \le |y_1-y_2|\frac{\tanh(\rho/4)}{\rho} \le \frac{1}{4}|y_1-y_2|.$$

The case $y_1,y_2\in\mathbb{R}_-$ can be treated similarly.

Finally, if $y_1y_2\le 0$ (say $y_2\le 0\le y_1$), we have

$$\mathbf{E}|X_\rho^{1/2}(y_1)-X_\rho^{1/2}(y_2)|^2 = 2-2\mathbf{E}X_\rho^{1/2}(y_1)\mathbf{E}X_\rho^{1/2}(y_2) = 2-2\exp\left(-|y_1|\frac{\tanh(\rho/4)}{2\rho}-|y_2|\frac{\tanh(\rho/4)}{2\rho}\right)$$
$$= 2-2\exp\left(-|y_1-y_2|\frac{\tanh(\rho/4)}{2\rho}\right) \le \frac{1}{4}|y_1-y_2|,$$

and so Lemma 3 is proved.

Now let us check Lemma 5. First let $y_1,y_2\in\mathbb{R}_+$ (say $y_1\ge y_2$) such that $\Delta=|y_1-y_2|<h$. Then

$$\mathbf{P}\{|\ln X_\rho(y_1)-\ln X_\rho(y_2)|>\varepsilon\} \le \frac{1}{\varepsilon^2}\mathbf{E}|\ln X_\rho(y_1)-\ln X_\rho(y_2)|^2 = \frac{1}{\varepsilon^2}\mathbf{E}|\ln X_\rho(\Delta)|^2 = \frac{1}{\varepsilon^2}\mathbf{E}\left|\rho\Pi_+\left(\frac{\Delta}{\rho}\right)-\Delta\right|^2$$
$$= \frac{1}{\varepsilon^2}\left(\rho^2(\lambda+\lambda^2)+\frac{\Delta^2}{\rho^2}-2\lambda\Delta\right) = \frac{1}{\varepsilon^2}(\beta(\rho)\Delta+\gamma(\rho)\Delta^2) < \frac{1}{\varepsilon^2}(\beta(\rho)h+\gamma(\rho)h^2),$$

where $\lambda=\Delta/\rho(e^\rho-1)$ is the parameter of the Poisson random variable $\Pi_+(\Delta/\rho)$,

$$\beta(\rho) = \frac{\rho}{(e^\rho-1)} = \frac{\rho}{\rho+o(\rho)} \to 1$$

and

$$\gamma(\rho) = \frac{1}{(e^\rho-1)^2}+\frac{1}{\rho^2}-\frac{2}{\rho(e^\rho-1)} = \left(\frac{1}{\rho}-\frac{1}{e^\rho-1}\right)^2 = \left(\frac{e^\rho-1-\rho}{\rho(e^\rho-1)}\right)^2 = \left(\frac{\rho^2/2+o(\rho^2)}{\rho(\rho+o(\rho))}\right)^2 \to \frac{1}{4}$$

as $\rho\to 0$. So, we have

$$\lim_{\rho\to 0}\sup_{|y_1-y_2|<h}\mathbf{P}\{|\ln X_\rho(y_1)-\ln X_\rho(y_2)|>\varepsilon\}\le\lim_{\rho\to 0}\frac{1}{\varepsilon^2}(\beta(\rho)h+\gamma(\rho)h^2)=\frac{1}{\varepsilon^2}\left(h+\frac{h^2}{4}\right),$$

and hence

$$\lim_{h\to 0}\lim_{\rho\to 0}\sup_{|y_1-y_2|<h}\mathbf{P}\{|\ln X_\rho(y_1)-\ln X_\rho(y_2)|>\varepsilon\}=0,$$

where the supremum is taken only over $y_1,y_2\in\mathbb{R}_+$.

For $y_1,y_2\in\mathbb{R}_-$ such that $\Delta=|y_1-y_2|<h$ one can obtain similarly

$$\mathbf{P}\{|\ln X_\rho(y_1)-\ln X_\rho(y_2)|>\varepsilon\}\le\frac{1}{\varepsilon^2}\mathbf{E}|\ln X_\rho(y_1)-\ln X_\rho(y_2)|^2=\frac{1}{\varepsilon^2}(\beta'(\rho)\Delta+\gamma'(\rho)\Delta^2)<\frac{1}{\varepsilon^2}(\beta'(\rho)h+\gamma'(\rho)h^2),$$

where

$$\beta'(\rho)=\frac{\rho}{(1-e^{-\rho})}=\frac{\rho}{\rho+o(\rho)}\to 1$$

and

$$\gamma'(\rho)=\left(\frac{e^{-\rho}-1+\rho}{\rho(1-e^{\rho})}\right)^2=\left(\frac{\rho^2/2+o(\rho^2)}{\rho(\rho+o(\rho))}\right)^2\to\frac{1}{4}$$

as $\rho\to 0$, which will yield the same conclusion as above, but with the supremum taken over $y_1,y_2\in\mathbb{R}_-$.

Finally, for $y_1y_2\le 0$ (say $y_2\le 0\le y_1$) such that $|y_1-y_2|<h$, using the elementary inequality $(a-b)^2\le 2(a^2+b^2)$ we get

$$\mathbf{P}\{|\ln X_\rho(y_1)-\ln X_\rho(y_2)|>\varepsilon\}\le\frac{1}{\varepsilon^2}\mathbf{E}|\ln X_\rho(y_1)-\ln X_\rho(y_2)|^2\le\frac{2}{\varepsilon^2}(\mathbf{E}|\ln X_\rho(y_1)|^2+\mathbf{E}|\ln X_\rho(y_2)|^2)$$
$$=\frac{2}{\varepsilon^2}(\beta(\rho)y_1+\gamma(\rho)y_1^2+\beta'(\rho)|y_2|+\gamma'(\rho)|y_2|^2)\le\frac{2}{\varepsilon^2}((\beta(\rho)+\beta'(\rho))h+(\gamma(\rho)+\gamma'(\rho))h^2),$$

which again will yield the desired conclusion. Lemma 5 is proved.

It remains to verify Lemma 6. Clearly,

$$\mathbf{P}\left\{\sup_{|y|>A}X_\rho(y)>e^{-bA}\right\}\le\mathbf{P}\left\{\sup_{y>A}X_\rho(y)>e^{-bA}\right\}+\mathbf{P}\left\{\sup_{y<-A}X_\rho(y)>e^{-bA}\right\}.$$

In order to estimate the first term, we need two auxiliary results.

Lemma 7. *For any $c\in]0,3/32[$ we have*

$$\mathbf{E}X_\rho^{1/4}(y)\le\exp(-c|y|)$$

for all sufficiently small ρ and all $y\in\mathbb{R}$.

Lemma 8. *For all $\rho>0$ the random variable*

$$\eta_\rho=\sup_{t\in\mathbb{R}_+}(\Pi_\lambda(t)-t),$$

where Π_λ is a Poisson process on $\mathbb{R}_+$ with intensity $\lambda=\rho/(e^\rho-1)\in]0,1[$, verifies

$$\mathbf{E}\exp\left(\frac{\rho}{4}\eta_\rho\right)\le 2.$$

The first result can be easily obtained following the proof of Lemma 4, so we prove the second one only. For this, let us remind that according to Shorack and Wellner (1986, Proposition 1 on page 392) (see also Pyke, 1959), the distribution function $F_\rho(x)=\mathbf{P}\{\eta_\rho<x\}$ of η_ρ is given by

$$1-F_\rho(x)=\mathbf{P}\{\eta_\rho\ge x\}=(1-\lambda)e^{\lambda x}\sum_{n>x}\frac{(n-x)^n}{n!}(\lambda e^{-\lambda})^n$$

for $x>0$, and is zero for $x\le 0$. Hence, for $x>0$ we have

$$1-F_\rho(x)\le(1-\lambda)e^{\lambda x}\sum_{n>x}\frac{(n-x)^n}{\sqrt{2\pi}\mathrm{n}n^ne^{-n}}(\lambda e^{-\lambda})^n=\frac{1-\lambda}{\sqrt{2\pi}}e^{\lambda x}\sum_{n>x}\frac{1}{\sqrt{n}}\left(1-\frac{x}{n}\right)^n(\lambda e^{1-\lambda})^n$$
$$\le\frac{1-\lambda}{\sqrt{2\pi}}e^{\lambda x}\sum_{n>x}e^{-x}\frac{(\lambda e^{1-\lambda})^n}{\sqrt{n}}\le\frac{1-\lambda}{\sqrt{2\pi}}e^{(\lambda-1)x}(\lambda e^{1-\lambda})^x\sum_{n>x}\frac{(\lambda e^{1-\lambda})^{n-x}}{\sqrt{n-x}}$$
$$=\frac{1-\lambda}{\sqrt{2\pi}}\lambda^x\sum_{k>0}\frac{(\lambda e^{1-\lambda})^k}{\sqrt{k}}\le\frac{1-\lambda}{\sqrt{2\pi}}\lambda^x\int_{\mathbb{R}_+}\frac{(\lambda e^{1-\lambda})^t}{\sqrt{t}}\,dt=\frac{1-\lambda}{\sqrt{2\pi}}\lambda^x\frac{\Gamma(1/2)}{\sqrt{-\ln(\lambda e^{1-\lambda})}}=\frac{1-\lambda}{\sqrt{-2\ln(\lambda e^{1-\lambda})}}\left(\frac{\rho}{e^\rho-1}\right)^x$$

$$\leq \left(\frac{\rho e^{-\rho/2}}{e^{\rho/2}-e^{-\rho/2}}\right)^x = \left(\frac{\rho e^{-\rho/2}}{2\sinh(\rho/2)}\right)^x \leq e^{-\rho x/2},$$

where we used Stirling inequality and the inequality $1-\lambda \leq \sqrt{-2\ln(\lambda e^{1-\lambda})}$, which is easily reduced to the elementary inequality $\ln(1-\mu) \leq -\mu-\mu^2/2$ by putting $\mu = 1-\lambda$. So, we can finish the proof of Lemma 8 by writing

$$\begin{aligned}\mathbf{E}\exp\left(\frac{\rho}{4}\eta_\rho\right) &= \int_{\mathbb{R}} e^{\rho x/4} dF_\rho(x) = [e^{\rho x/4}(F_\rho(x)-1)]_{-\infty}^{+\infty} - \frac{\rho}{4}\int_{\mathbb{R}} e^{\rho x/4}(F_\rho(x)-1)\,dx \\ &= \frac{\rho}{4}\int_{\mathbb{R}_-} e^{\rho x/4}\,dx + \frac{\rho}{4}\int_{\mathbb{R}_+} e^{\rho x/4}(1-F_\rho(x))\,dx \leq 1 + \frac{\rho}{4}\int_{\mathbb{R}_+} e^{-\rho x/4}\,dx = 2.\end{aligned}$$

Now, let us get back to the proof of Lemma 6. Using Lemma 8 and taking into account the stationarity and the independence of the increments of the process $\ln X_\rho$ on $\mathbb{R}_+$, we obtain

$$\begin{aligned}\mathbf{P}\left\{\sup_{y>A} X_\rho(y) > e^{-bA}\right\} &\leq e^{bA/4}\mathbf{E}\sup_{y>A} X_\rho^{1/4}(y) = e^{bA/4}\mathbf{E}X_\rho^{1/4}(A)\mathbf{E}\sup_{y>A}\frac{X_\rho^{1/4}(y)}{X_\rho^{1/4}(A)} = e^{bA/4}\mathbf{E}X_\rho^{1/4}(A)\mathbf{E}\sup_{z>0} X_\rho^{1/4}(z) \\ &= e^{bA/4}\mathbf{E}X_\rho^{1/4}(A)\mathbf{E}\sup_{z>0}\left(\exp\left(\frac{\rho}{4}\Pi_+(z/\rho) - \frac{z}{4\rho}\right)\right) = e^{bA/4}\mathbf{E}X_\rho^{1/4}(A)\mathbf{E}\exp\left(\sup_{t>0}\left(\frac{\rho}{4}\left(\Pi_{\frac{\rho}{e^\rho-1}}(t)-t\right)\right)\right) \\ &= e^{bA/4}\mathbf{E}X_\rho^{1/4}(A)\mathbf{E}\exp\left(\frac{\rho}{4}\eta_\rho\right) \leq 2e^{bA/4}\mathbf{E}X_\rho^{1/4}(A).\end{aligned}$$

Hence, taking $b \in]0,3/40[$, we have $5b/4 \in]0,3/32[$ and, using Lemma 7, we finally get

$$\mathbf{P}\left\{\sup_{y>A} X_\rho(y) > e^{-bA}\right\} \leq 2e^{bA/4}\exp\left(-\frac{5b}{4}A\right) = 2e^{-bA}$$

for all sufficiently small ρ and all $A > 0$, and so the first term is estimated.

The second term can be estimated in the same way, if we show that for all $\rho > 0$ the random variable

$$\eta'_\rho = \sup_{t\in\mathbb{R}_+}(-\Pi_{\lambda'}(t)+t) = -\inf_{t\in\mathbb{R}_+}(\Pi_{\lambda'}(t)-t),$$

where $\Pi_{\lambda'}$ is a Poisson process on $\mathbb{R}_+$ with intensity $\lambda' = \rho/(1-e^{-\rho}) \in]0,1[$, verifies

$$\mathbf{E}\exp\left(\frac{\rho}{4}\eta'_\rho\right) \leq 2.$$

For this, let us remind that according to Pyke (1959) (see also Cramér, 1954), η'_ρ is an exponential random variable with parameter r, where r is the unique positive solution of the equation $\lambda'(e^{-r}-1)+r = 0$. In our case, this equation becomes

$$\frac{\rho}{1-e^{-\rho}}(e^{-r}-1)+r = 0,$$

and $r = \rho$ is clearly its solution. Hence η'_ρ is an exponential random variable with parameter ρ, which yields

$$\mathbf{E}\exp\left(\frac{\rho}{4}\eta'_\rho\right) = \frac{4}{3} < 2,$$

and so Lemma 6 is proved.

References

Chernoff, H., Rubin, H., 1956. The estimation of the location of a discontinuity in density. In: Proceedings of 3rd Berkeley Symposium, vol. 1, pp. 19–37.

Cramér, H., 1954. On some questions connected with mathematical risk. Univ. California Publ. Statist. 2, 99–123.

Deshayes, J., Picard, D., 1984. Lois asymptotiques des tests et estimateurs de rupture dans un modèle statistique classique. Ann. Inst. H. Poincaré Probab. Statist. 20 (4), 309–327.

Gihman, I.I., Skorohod, A.V., 1974. The Theory of Stochastic Processes, vol. I. Springer-Verlag, New York.

Golubev, G.K., 1979. Computation of the efficiency of the maximum-likelihood estimator when observing a discontinuous signal in white noise. Probl. Inf. Transm. 15 (3), 61–69.

Höpfner, R., Kutoyants, Yu.A., 2009. Estimating discontinuous periodic signals in a time inhomogeneous diffusion, preprint, 2009. ⟨http://www.mathematik.uni-mainz.de/∼ hoepfner/ssp/zeit.html⟩.

Ibragimov, I.A., Khasminskii, R.Z., 1970. The asymptotic behavior of generalized Bayesian estimates. Dokl. Akad. Nauk SSSR 194, 257–260.

Ibragimov, I.A., Khasminskii, R.Z., 1972. The asymptotic behavior of statistical estimates for samples with a discontinuous density. Mat. Sb. 87(129) (4), 554–558.

Ibragimov, I.A., Khasminskii, R.Z., 1975. Estimation of a parameter of a discontinuous signal in a white Gaussian noise. Probl. Inform. Transm. 11 (3), 31–43.

Ibragimov, I.A., Khasminskii, R.Z., 1981. Statistical Estimation. Asymptotic Theory. Springer-Verlag, New York.

Küchler, U., Kutoyants, Yu.A., 2000. Delay estimation for some stationary diffusion-type processes. Scand. J. Statist. 27 (3), 405–414.

Kutoyants, Yu.A., 1984. Parameter estimation for stochastic processes. Armenian Academy of Sciences, Yerevan, 1980 (in Russian), translation of revised version, Heldermann-Verlag, Berlin.

Kutoyants, Yu.A., 1994. Identification of dynamical systems with small noise. Mathematics and its Applications, vol. 300. Kluwer Academic Publishers Group, Dordrecht.

Kutoyants, Yu.A., 1998. Statistical inference for spatial poisson processes. Lecture Notes in Statistics, vol. 134, Springer-Verlag, New York.

Kutoyants, Yu.A., 2004. Statistical inference for ergodic diffusion processes. Springer Series in Statistics. Springer-Verlag, London.

Pflug, G.Ch., 1993. On an argmax-distribution connected to the Poisson process. In: Mandl, P., Hušková, H. (Eds.), Proceedings of the Fifth Prague Conference on Asymptotic Statistics, pp. 123–130.

Pyke, R., 1959. The supremum and infimum of the Poisson process. Ann. Math. Statist. 30, 568–576.
Rubin, H., Song, K.-S., 1995. Exact computation of the asymptotic efficiency of maximum likelihood estimators of a discontinuous signal in a Gaussian white noise. Ann. Statist. 23 (3), 732–739.
Shorack, G.R., Wellner, J.A., 1986. Empirical Processes with Applications to Statistics. John Wiley & Sons Inc., New York.
Terent'yev, A.S., 1968. Probability distribution of a time location of an absolute maximum at the output of a synchronized filter. Radioengineering and Electronics 13 (4), 652–657.

Statistics, Vol. 45, No. 5, October 2011, 509–523

Taylor & Francis
Taylor & Francis Group

Estimation of the location of a 0-type or ∞-type singularity by Poisson observations

Sergueï Dachian*

Laboratoire de Mathématiques, Université Blaise Pascal, 63177 Aubière Cedex, France

(*Received 2 November 2006; final version received 21 January 2010*)

We consider an inhomogeneous Poisson process X on $[0, T]$. The intensity function of X is supposed to be strictly positive and smooth on $[0, T]$ except at the point θ, in which it has either a 0-type singularity (tends to 0 like $|x|^p$, $p \in (0, 1)$), or an ∞-type singularity (tends to ∞ like $|x|^p$, $p \in (-1, 0)$). We suppose that we know the shape of the intensity function, but not the location of the singularity. We consider the problem of estimation of this location (shift) parameter θ based on n observations of the process X. We study the Bayesian estimators and, in the case $p > 0$, the maximum-likelihood estimator. We show that these estimators are consistent, their rate of convergence is $n^{1/(p+1)}$, they have different limit distributions, and the Bayesian estimators are asymptotically efficient.

Keywords: inhomogeneous Poisson process; singularity; parameter estimation; Bayesian estimators; maximum-likelihood estimator; consistency; limit distribution; convergence of moments; asymptotic efficiency

2000 Mathematics Subject Classification: 62M05

1. Introduction

Inhomogeneous Poisson process is one of the simplest point processes (see, e.g. [1]). However, due to the large choice of intensity functions, this model is rich enough and is widely used in many applied statistical problems, such as optical communications, reliability, biology, medicine, image treatment, and so on (see, e.g. [2–5]).

The diversity of applications is also due to the possibility of using the likelihood ratio analysis. In parameter estimation problems, the large samples theory is quite close to the one of the classical (i.i.d.) statistics. In particular, let us consider the problem of estimation of the parameter θ by n independent observations on some fixed interval $[0, T]$ of an inhomogeneous Poisson process $X = \{X(t),\ 0 \leqslant t \leqslant T\}$ of intensity function $S_\theta(t)$. Let us mention that this problem is equivalent to the one of estimation of the parameter by one observation on a growing interval of a periodic inhomogeneous Poisson process. If the problem is regular (the model is locally asymptotically normal), then both the maximum-likelihood estimator (MLE) $\hat{\theta}_n$ and the

*Email: Serguei.Dachian@math.univ-bpclermont.fr

ISSN 0233-1888 print/ISSN 1029-4910 online

DOI: 10.1080/02331881003768784
http://www.informaworld.com

Bayesian estimators (BE) $\tilde{\theta}_n$ are consistent, asymptotically normal:

$$\sqrt{n}(\hat{\theta}_n - \theta) \Longrightarrow \mathcal{N}(0, I(\theta)^{-1}), \quad \sqrt{n}(\tilde{\theta}_n - \theta) \Longrightarrow \mathcal{N}(0, I(\theta)^{-1}),$$

and asymptotically efficient (see, e.g. [6,7]). Here $I(\theta)$ is the Fisher information given by

$$I(\theta) = \int_0^T \frac{\dot{S}_\theta^2(t)}{S_\theta(t)}\,\mathrm{d}t,$$

where $S_\theta(t)$ is the intensity function and $\dot{S}_\theta(t) = (\partial/\partial\theta)S_\theta(t)$.

If the problem is not regular, then the properties of estimators essentially change. For example, if $S_\theta(\cdot)$ is smooth everywhere on $[0, T]$ except at the point θ, in which it has a jump (consider for instance $S_\theta(t) = s(t-\theta)$, where $s(\cdot)$ is discontinuous in 0), then the MLE and BE are still consistent, but converge at a faster rate:

$$n(\hat{\theta}_n - \theta) \Longrightarrow \xi_1, \quad n(\tilde{\theta}_n - \theta) \Longrightarrow \xi_2,$$

have different limit distributions (ξ_1 and ξ_2 are different with $\mathbf{E}\xi_1^2 > \mathbf{E}\xi_2^2$), and the BE are asymptotically efficient (see, e.g. [6,7]).

In this paper, we deal with the case where the intensity function $S_\theta(\cdot)$ is smooth everywhere on $[0, T]$ except at the point θ, in which it has a singularity of order p. The cusp-type singularities were already studied in the preceding paper [8]. Here we consider 0-type and ∞-type singularities. More precisely, we suppose that $S_\theta(t) = s(t-\theta)$, where $s(\cdot)$ is some known strictly positive function on $[-T, T] \setminus \{0\}$ and $\theta \in (0, T)$ is some unknown parameter, and that we have the following representation:

$$S_\theta(t) = s(t-\theta) = \begin{cases} a|t-\theta|^p + \psi(t-\theta) & \text{if } t < \theta, \\ b|t-\theta|^p + \psi(t-\theta) & \text{if } t > \theta, \end{cases}$$

where $a, b > 0$, $p > -1$ (to guarantee the finiteness of intensity measure), and $\psi(\cdot)$ is smooth.

If $\psi(0) \neq 0$ and $p > 1/2$ then, in spite of the singularity of the intensity function in θ, the Fisher information is finite, and so this case can be treated as the regular one.

If $\psi(0) \neq 0$ and $0 < p < 1/2$, we say that the intensity function has a cusp at θ. This is the case treated in [8] (where instead of $a, b > 0$ it was supposed $a^2 + b^2 > 0$ only). There it was shown that the MLE and the BE are consistent, converge at the rate $n^{1/(2p+1)}$ (which is faster than in the regular case but slower than in discontinuous case):

$$n^{1/(2p+1)}(\hat{\theta}_n - \theta) \Longrightarrow \eta_1, \quad n^{1/(2p+1)}(\tilde{\theta}_n - \theta) \Longrightarrow \eta_2,$$

have different limit distributions, and the BE are asymptotically efficient. The convergence of moments was equally verified.

If $\psi(0) = 0$ and $p > 1$ then, as above, the Fisher information is finite and this case can be treated as the regular one.

If $\psi(0) = 0$ and $0 < p < 1$, we say that the intensity function has a 0-type singularity at θ. In this case, we study the asymptotic behaviour of the MLE and the BE, and we prove that the estimators are consistent, converge at the rate $n^{1/(p+1)}$ (which is again intermediate between the regular and discontinuous case rates), have different limit distributions, and the BE are asymptotically efficient. We also verify the convergence of moments.

If $-1 < p < 0$ we say that the intensity function has a ∞-type singularity at θ. In this case, we study the asymptotic behaviour of the BE only (MLE makes no sense in this case). We prove that the estimators are consistent, converge at the rate $n^{1/(p+1)}$ (which is even faster

than in discontinuous case), and are asymptotically efficient. We verify as well the convergence of moments.

Let us note that the jump can also be considered as a singularity by taking $p = 0$ and $a \neq b$, which explains that the rates are slower for $p > 0$ and faster for $p < 0$.

Let us also mention that our results are similar to those obtained by Ibragimov and Khasminskii for the problem of estimation of a singularity location of the density for the i.i.d. model of observations. An exhaustive exposition of the results can be found in Chapter 6 of their book [9], but one can also refer to their previous works [10,11]. The asymptotic behaviour of the MLE and of a wide class of BE obtained for this (i.i.d.) model is similar to the one obtained here for the model of Poisson observations. Particularly, the rate of convergence of the estimators is $n^{1/(p+1)}$, and the BE are asymptotically efficient.

Finally, let us note that for the study of the asymptotic behaviour of the estimators we use the method of Ibragimov and Khasminskii presented in their book [9] (see as well [7], where this method is applied to inhomogeneous Poisson process).

2. Main results

Suppose that we observe n realizations $(X_1, \dots, X_n) = X^n$ of the Poisson process $X = \{X(t), 0 \leqslant t \leqslant T\}$ of intensity function $S_\theta(t) = s(t - \theta)$, where θ is some unknown parameter, $\theta \in \Theta = (\alpha, \beta) \subseteq (0, T)$, and $s(\cdot)$ is some known strictly positive function on $[-T, T] \setminus \{0\}$. We suppose that the function $s(\cdot)$ can be written in the form $s(t) = d(t)|t|^p + \psi(t)$, where $p \in (-1, 0) \cup (0, 1)$,

$$d(t) = \begin{cases} a & \text{if } t < 0, \\ b & \text{if } t > 0, \end{cases}$$

$a, b > 0$, and the function $\psi(\cdot)$ is Hölder continuous on $[-T, T]$ of order higher than $(p + 1)/2$, that is $|\psi(x) - \psi(y)| \leqslant L\,|x - y|^{\varkappa}$ for all $x, y \in [-T, T]$ with some fixed constants $L > 0$ and $\varkappa > (p + 1)/2$. In the case $p > 0$, we suppose equally that $\psi(0) = 0$. Our aim is to estimate the parameter θ and to study the asymptotic behaviour of estimators as n goes to infinity.

The likelihood ratio in our problem can be written (see, e.g. [7]) as

$$L(\theta, \theta_1, X^n) = \exp\left\{ \sum_{i=1}^{n} \int_0^T \ln \frac{S_\theta(t)}{S_{\theta_1}(t)}\, \mathrm{d}X_i(t) - n \int_0^T \left[\frac{S_\theta(t)}{S_{\theta_1}(t)} - 1 \right] S_{\theta_1}(t)\, \mathrm{d}t \right\},$$

where θ_1 is some fixed value of θ.

As usual, introduce the MLE $\hat{\theta}_n$ as one of the solutions of the equation

$$L(\hat{\theta}_n, \theta_1, X^n) = \sup_{\theta \in \Theta} L(\theta, \theta_1, X^n)$$

and the BE $\tilde{\theta}_n$ for prior density q (supposed to be positive and continuous) and quadratic loss function as

$$\tilde{\theta}_n = \int_\alpha^\beta \theta\, q(\theta | X^n)\, \mathrm{d}\theta,$$

where the posterior density

$$q(\theta | X^n) = L(\theta, \theta_1, X^n)\, q(\theta) \left(\int_\alpha^\beta L(\theta, \theta_1, X^n)\, q(\theta)\, \mathrm{d}\theta \right)^{-1}.$$

Note that the MLE makes no sense in the case $p < 0$, since in this case the likelihood equals infinity in any point θ which is event of one of the Poisson processes $X_1, \dots, X_n$.

To describe the properties of these estimators, we need to introduce the stochastic process

$$Z(u) = \exp\left\{ p \int_{-\infty}^{+\infty} \ln\left|1 - \frac{u}{z}\right| \pi(\mathrm{d}z) + \ln\frac{a}{b} \int_0^u Y(\mathrm{d}z) \right.$$
$$- \int_{-\infty}^{+\infty} \left[\left|1 - \frac{u}{z}\right|^p - 1 - p \ln\left|1 - \frac{u}{z}\right| \right] d(z)|z|^p \, \mathrm{d}z$$
$$\left. - \frac{a-b}{p+1} |u|^{p+1} \operatorname{sign}(u) \right\}, \quad u \in \mathbb{R}.$$

Here and in the sequel Y denotes a Poisson process on $\mathbb{R}$ of intensity function $S_0(z) = d(z)|z|^p$, and π is its centred version: $\pi = Y - \mathbf{E}Y$.

We introduce also the random variable ζ, and in the case $p > 0$ the random variable ξ by the equations

$$\zeta = \int_{-\infty}^{+\infty} u \, Z(u) \, \mathrm{d}u \left(\int_{-\infty}^{+\infty} Z(u) \, \mathrm{d}u \right)^{-1}$$

and

$$Z(\xi) = \sup_{u \in \mathbb{R}} Z(u).$$

Let us note here that ξ is well defined in the case $p > 0$, since in this case with probability one the process $Z(u)$ attains its maximum in a unique point (see, e.g. [12]).

Now we can finally state the main results of this paper.

THEOREM 1 *Under the made assumptions, the following lower bound on the risks of all estimators holds: for any $\theta_0 \in \Theta$ we have*

$$\lim_{\delta \to 0} \varliminf_{n \to \infty} \inf_{\bar{\theta}_n} \sup_{|\theta - \theta_0| < \delta} \mathbf{E}_\theta (n^{1/(p+1)}(\bar{\theta}_n - \theta))^2 \geq \mathbf{E}\zeta^2,$$

where inf *is taken over all possible estimators $\bar{\theta}_n$ of θ.*

This theorem leads us to introduce the following definition.

DEFINITION 2 *We say that the estimator $\bar{\theta}_n$ is asymptotically efficient if*

$$\lim_{\delta \to 0} \lim_{n \to \infty} \sup_{|\theta - \theta_0| < \delta} \mathbf{E}_\theta (n^{1/(p+1)}(\bar{\theta}_n - \theta))^2 = \mathbf{E}\zeta^2$$

for any $\theta_0 \in \Theta$.

For the BE we have the following theorem.

THEOREM 3 *The BE $\tilde{\theta}_n$ have uniformly in $\theta \in \mathbf{K}$ (for any compact $\mathbf{K} \subset \Theta$) the following properties*:

- *$\tilde{\theta}_n$ is consistent, that is*

$$\tilde{\theta}_n \xrightarrow{\mathbf{P}_\theta} \theta \text{ (convergence in probability)},$$

- *the limit distribution of $\tilde{\theta}_n$ is ζ, that is*

$$n^{1/(p+1)}(\tilde{\theta}_n - \theta) \Longrightarrow \zeta \text{ (convergence in law)},$$

- *for any $k > 0$ we have*

$$\lim_{n\to\infty} \mathbf{E}_\theta |n^{1/(p+1)}(\tilde{\theta}_n - \theta)|^k = \mathbf{E}|\zeta|^k$$

and, moreover, $\tilde{\theta}_n$ is asymptotically efficient.

And for the MLE (in the case $p > 0$) we have the following theorem.

THEOREM 4 *Let $p \in (0, 1)$. The MLE $\hat{\theta}_n$ has uniformly in $\theta \in \mathbf{K}$ (for any compact $\mathbf{K} \subset \Theta$) the following properties*:

- *$\hat{\theta}_n$ is consistent, that is*

$$\hat{\theta}_n \xrightarrow{\mathbf{P}_\theta} \theta \text{ (convergence in probability)},$$

- *the limit distribution of $\hat{\theta}_n$ is ξ, that is*

$$n^{1/(p+1)}(\hat{\theta}_n - \theta) \Longrightarrow \xi \text{ (convergence in law)},$$

- *for any $k > 0$ we have*

$$\lim_{n\to\infty} \mathbf{E}_\theta |n^{1/(p+1)}(\hat{\theta}_n - \theta)|^k = \mathbf{E}|\xi|^k.$$

To prove the above stated theorems, we apply the method of Ibragimov and Khasminskii [9]. For this we denote $\theta_u = \theta + u\, n^{-1/(p+1)}$ for all $u \in U_n = (n^{1/(p+1)}(\alpha - \theta), n^{1/(p+1)}(\beta - \theta))$, we introduce the normalized likelihood ratio process as

$$Z_n(u) = L(\theta_u, \theta, X^n), \quad u \in U_n,$$

and we establish (the proofs are in the next section) the following three lemmas.

LEMMA 5 *The finite-dimensional distributions of $Z_n(u)$ converge to those of $Z(u)$ uniformly in $\theta \in \mathbf{K}$ (for any compact $\mathbf{K} \subset \Theta$).*

LEMMA 6 *For any compact $\mathbf{K} \subset \Theta$ there exists some positive constant C such that*

$$\mathbf{E}_\theta |Z_n^{1/2}(u_1) - Z_n^{1/2}(u_2)|^2 \leqslant C\, |u_1 - u_2|^{p+1}$$

for all $u_1, u_2 \in U_n$, $\theta \in \mathbf{K}$ and n sufficiently large.

LEMMA 7 *For any compact $\mathbf{K} \subset \Theta$ there exists some positive constant c such that*

$$\mathbf{E}_\theta Z_n^{1/2}(u) \leqslant \exp\{-c\, |u|^{p+1}\}$$

for all $u \in U_n$, $\theta \in \mathbf{K}$, and $n \in \mathbb{N}$.

Using these lemmas and applying Theorems 1.9.1, 1.10.2 and 1.10.1 of [9], we get Theorems 1, 3, and 4, respectively.

3. Proofs of the lemmas

For convenience of notation, all throughout this section C and c denote generic positive constants which can differ from formula to formula (and even in the same formula), and we put $\nu = 1/(p+1)$.

First of all let us fix some $\delta > 0$ such that $c\,d(t)|t|^p \leqslant s(t) \leqslant C\,d(t)|t|^p$ on $(-\delta, \delta)$, and $s(t) \geqslant c$ on $[-T, T] \setminus (-\delta/4, \delta/4)$. To do so, we note that

$$|\psi(t)| \leq |\psi(0)| + C|t|^{\varkappa} = C|t|^p \left(|t|^{\varkappa - p} + c|\psi(0)||t|^{-p}\right) \leqslant \min\{a, b\}\,|t|^p/2$$

for $t \in (-\delta, \delta)$, since $\varkappa - p > (1-p)/2 > 0$ and $\psi(0) = 0$ in the case $p > 0$. It follows clearly

$$s(t) = d(t)|t|^p + \psi(t) \geqslant \left(d(t) - \frac{\min\{a, b\}}{2}\right)|t|^p \geqslant \frac{d(t)|t|^p}{2}$$

and $s(t) \leqslant 2\,d(t)|t|^p$. Finally, on the compact set $[-T, T] \setminus (-\delta/4, \delta/4)$ we have $s(t) \geqslant c$ since the function $s(\cdot)$ is continuous.

Now, let us fix some sequence (A_n) such that $A_n \to +\infty$ sufficiently slow. More precisely, we suppose that $A_n\, n^{-\nu} \to 0$ and we will give some additional conditions below. We split the interval $[0, T]$ into three parts:

$$E_1 = \{t : |t - \theta| < A_n\, n^{-\nu}\} = (\theta - A_n\, n^{-\nu}, \theta + A_n\, n^{-\nu}),$$
$$E_2 = \{t : A_n\, n^{-\nu} < |t - \theta| < \delta\} = (\theta - \delta, \theta - A_n\, n^{-\nu}) \cup (\theta + A_n\, n^{-\nu}, \theta + \delta),$$
$$E_3 = \{t : \delta < |t - \theta|\} = (0, \theta - \delta) \cup (\theta + \delta, T).$$

In order to prove Lemma 5, we will only study the convergence of the one-dimensional (the general case can be considered similarly) distributions of the process

$$\ln Z_n(u) = \sum_{i=1}^n \int_0^T \ln \frac{S_{\theta_u}(t)}{S_\theta(t)}\,\mathrm{d}X_i(t) - n\int_0^T \left[\frac{S_{\theta_u}(t)}{S_\theta(t)} - 1\right] S_\theta(t)\,\mathrm{d}t$$
$$= \sum_{i=1}^n \int_0^T f\,\mathrm{d}X_i(t) - n\int_0^T g\,S_\theta(t)\,\mathrm{d}t,$$

where we denote

$$f = f(\theta, t, u, n) = \ln \frac{S_{\theta_u}(t)}{S_\theta(t)} \quad \text{and} \quad g = g(\theta, t, u, n) = \frac{S_{\theta_u}(t)}{S_\theta(t)} - 1.$$

The characteristic function of the random variable $\ln Z_n(u)$ can be written as (see, e.g., [7], Lemma 1.1)

$$\phi_n(\lambda) = \mathbf{E}_\theta \exp\{i\,\lambda\,\ln Z_n(u)\} = \exp\left\{n\int_0^T [\mathrm{e}^{i\,\lambda\,f} - 1 - i\,\lambda\,g] S_\theta(t)\,\mathrm{d}t\right\},$$

and hence

$$\ln \phi_n(\lambda) = n\int_0^T [\mathrm{e}^{i\,\lambda\,f} - 1 - i\,\lambda\,f] S_\theta(t)\,\mathrm{d}t + i\,\lambda\,n\int_0^T [f - g] S_\theta(t)\,\mathrm{d}t. \tag{1}$$

To study this expression, let us at first establish the two following properties.

(a) For any fixed u, we have $\lim_{n\to\infty} g(\theta, t, u, n) = 0$ uniformly in $\theta \in \mathbf{K}$ and $t \in E_2 \cup E_3$.

(b) We have

$$\lim_{n\to\infty} n \int_{E_2\cup E_3} g^2\, S_\theta(t)\,\mathrm{d}t = 0.$$

To prove (a), we set $y = t - \theta \in (E_2 - \theta) \cup (E_3 - \theta)$ and we write

$$\begin{aligned}
|g(\theta, t, u, n)| &= \left| \frac{s(t-\theta_u)}{s(t-\theta)} - 1 \right| \\
&= \left| \frac{d(y - u\,n^{-\nu})\,|y - un^{-\nu}|^p + \psi(y - u\,n^{-\nu}) - d(y)|y|^p - \psi(y)}{s(y)} \right| \\
&= \frac{|C\,(|y - u\,n^{-\nu}|^p - |y|^p) + \psi(y - u\,n^{-\nu}) - \psi(y)|}{s(y)} \\
&\leqslant C\,\frac{||y - u\,n^{-\nu}|^p - |y|^p|}{s(y)} + \frac{|\psi(y - u\,n^{-\nu}) - \psi(y)|}{s(y)} \\
&= M_1 + M_2
\end{aligned}$$

with evident notations.

For $y \in E_2 - \theta$ we have

$$M_1 \leqslant C\,\frac{||y - u\,n^{-\nu}|^p - |y|^p|}{c|y|^p} = C\,\left|\left|1 - \frac{u}{y\,n^\nu}\right|^p - 1\right| \leqslant \frac{C\,|u|}{|y|\,n^\nu} \leqslant \frac{C\,|u|}{A_n} \longrightarrow 0,$$

$$M_2 \leqslant \frac{C\,|u\,n^{-\nu}|^\varkappa}{c} = C\,|u|^\varkappa\, n^{-\nu\varkappa} \longrightarrow 0 \quad \text{if } p < 0,$$

$$M_2 \leqslant \frac{C\,|u\,n^{-\nu}|^\varkappa}{c|y|^p} \leqslant \frac{C\,|u|^\varkappa\, n^{-\nu\varkappa}}{(A_n\, n^{-\nu})^p} = C\,|u|^\varkappa\, \frac{n^{-\nu(\varkappa-p)}}{A_n^p} \longrightarrow 0 \quad \text{if } p > 0.$$

Finally for $y \in E_3 - \theta$, using the Hölder continuity of $s(\cdot)$, we have

$$|g(\theta, t, u, n)| \leqslant \frac{C\,|u\,n^{-\nu}|^\varkappa}{c} = C\,|u|^\varkappa\, n^{-\nu\varkappa} \longrightarrow 0.$$

So, (a) is proved.

To prove (b), we first note that

$$\begin{aligned}
n\int_{E_3} g^2\, S_\theta(t)\,\mathrm{d}t &= n\int_{E_3} \frac{(S_{\theta_u}(t) - S_\theta(t))^2}{S_\theta(t)}\,\mathrm{d}t = n\int_{E_3-\theta} \frac{(s(y - u\,n^{-\nu}) - s(y))^2}{s(y)}\,\mathrm{d}y \\
&\leqslant C\,n\int_{E_3-\theta} |u\,n^{-\nu}|^{2\varkappa}\,\mathrm{d}y = C\,|u|^{2\varkappa}\, n^{-(2\nu\varkappa-1)} \longrightarrow 0
\end{aligned}$$

since $2\nu\varkappa - 1 > 0$. To conclude the proof it remains to show that

$$n\int_{\theta+A_n\,n^{-\nu}}^{\theta+\delta} g^2\, S_\theta(t)\,\mathrm{d}t + n\int_{\theta-\delta}^{\theta-A_n\,n^{-\nu}} g^2\, S_\theta(t)\,\mathrm{d}t \longrightarrow 0.$$

For the first term we have

$$n\int_{\theta+A_n\,n^{-\nu}}^{\theta+\delta} g^2\, S_\theta(t)\,\mathrm{d}t = n\int_{A_n\,n^{-\nu}}^{\delta} \frac{(s(y - u\,n^{-\nu}) - s(y))^2}{s(y)}\,\mathrm{d}y$$

$$
\begin{aligned}
&\leqslant n \int_{A_n n^{-\nu}}^{\delta} \frac{(s(y - u\,n^{-\nu}) - s(y))^2}{c|y|^p}\,\mathrm{d}y \\
&= C\,n \int_{A_n n^{-\nu}}^{\delta} \frac{(|y - u\,n^{-\nu}|^p - |y|^p)^2}{|y|^p}\,\mathrm{d}y \\
&\quad + C\,n \int_{A_n n^{-\nu}}^{\delta} \frac{(\psi(y - u\,n^{-\nu}) - \psi(y))^2}{|y|^p}\,\mathrm{d}y \\
&\quad + C\,n \int_{A_n n^{-\nu}}^{\delta} \frac{(\psi(y - u\,n^{-\nu}) - \psi(y))(|y - u\,n^{-\nu}|^p - |y|^p)}{|y|^p}\,\mathrm{d}y \\
&= n\,J_1 + n\,J_2 + n\,J_3
\end{aligned}
$$

with evident notations. Further

$$
n\,J_1 = C\,n \int_{A_n}^{\delta n^{\nu}} \frac{(|z - u|^p - |z|^p)^2\,n^{-2\nu p}}{|z|^p\,n^{-\nu p}} n^{-\nu}\,\mathrm{d}z \leqslant C \int_{A_n}^{+\infty} \frac{(|z - u|^p - |z|^p)^2}{|z|^p}\,\mathrm{d}z \longrightarrow 0
$$

since

$$
\frac{(|z - u|^p - |z|^p)^2}{|z|^p} = |z|^p \left(\left| 1 - \frac{u}{z} \right|^p - 1 \right)^2 \sim |z|^p \left(\frac{C}{z} \right)^2 = C\,|z|^{p-2}
$$

and $p - 2 < -1$. Similarly

$$
n\,J_2 \leqslant C\,n \int_{A_n n^{-\nu}}^{\delta} \frac{|u\,n^{-\nu}|^{2\varkappa}}{|y|^p}\,\mathrm{d}y \leqslant C\,|u|^{2\varkappa}\,n^{-(2\nu\varkappa - 1)} \int_0^{\delta} |y|^{-p}\,\mathrm{d}y \longrightarrow 0
$$

since $2\nu\varkappa - 1 > 0$ and $-p > -1$. Finally

$$
|n\,J_3| \leqslant C \sqrt{(n\,J_1)\,(n\,J_2)} \longrightarrow 0
$$

by Cauchy–Schwarz inequality, and so the first term converges to 0.

The second term can be treated in the same way. So, (b) is proved.

Now let us return to the study of the characteristic function $\phi_n(\cdot)$. Using Equation (1) we can write

$$
\ln \phi_n = \varphi_1 + \varphi_2 + \varphi_3,
$$

where we set

$$
\varphi_k = n \int_{E_k} [\mathrm{e}^{i\,\lambda\,f} - 1 - i\,\lambda\,f]\,S_\theta(t)\,\mathrm{d}t + i\,\lambda\,n \int_{E_k} [f - g] S_\theta(t)\,\mathrm{d}t.
$$

For φ_3 we get

$$
\begin{aligned}
\varphi_3 &= n \int_{E_3} \left[\mathrm{e}^{i\,\lambda\,f} - 1 - i\,\lambda\,f\right] S_\theta(t)\,\mathrm{d}t + i\,\lambda\,n \int_{E_3} [f - g]\,S_\theta(t)\,\mathrm{d}t \\
&\simeq \frac{1}{2}\,n \int_{E_3} (i\,\lambda\,f)^2\,S_\theta(t)\,\mathrm{d}t - \frac{1}{2}\,i\,\lambda\,n \int_{E_3} g^2\,S_\theta(t)\,\mathrm{d}t \\
&\simeq -\frac{1}{2}\,\lambda^2\,n \int_{E_3} g^2\,S_\theta(t)\,\mathrm{d}t - \frac{1}{2}\,i\,\lambda\,n \int_{E_3} g^2\,S_\theta(t)\,\mathrm{d}t \longrightarrow 0,
\end{aligned}
$$

where the symbol '$\simeq$' means equality of limits.

In the same way, we get $\varphi_2 \to 0$, and it remains to study φ_1. For this we set $y_u = y - u\,n^{-\nu}$, $\alpha(y) = d(y)|y|^p$,

$$\beta(y) = \frac{s(y)}{\alpha(y)} = 1 + \frac{\psi(y)}{d(y)|y|^p}$$

for $y \in [-T, T] \setminus \{0\}$, and $\beta(0) = 1$. Note that the function $\beta(\cdot)$ is clearly Hölder continuous of order

$$\mu = \begin{cases} \varkappa - p & \text{if } p > 0, \\ \min\{\varkappa, -p\} & \text{if } p < 0. \end{cases}$$

We have

$$\begin{aligned}
\varphi_1 &= n \int_{E_1} \left[\mathrm{e}^{i\,\lambda\,f} - 1 - i\,\lambda\,f\right] S_\theta(t)\,\mathrm{d}t + i\,\lambda\,n \int_{E_1} [f - g]\,S_\theta(t)\,\mathrm{d}t \\
&= n \int_{-A_n\,n^{-\nu}}^{A_n\,n^{-\nu}} \left[\left(\frac{\alpha(y_u)\,\beta(y_u)}{\alpha(y)\,\beta(y)} \right)^{i\,\lambda} - 1 - i\,\lambda\,\ln \frac{\alpha(y_u)\,\beta(y_u)}{\alpha(y)\,\beta(y)} \right] \alpha(y)\,\beta(y)\,\mathrm{d}y \\
&\quad - i\,\lambda\,n \int_{-A_n\,n^{-\nu}}^{A_n\,n^{-\nu}} \left[\frac{\alpha(y_u)\,\beta(y_u)}{\alpha(y)\,\beta(y)} - 1 - \ln \frac{\alpha(y_u)\,\beta(y_u)}{\alpha(y)\,\beta(y)} \right] \alpha(y)\,\beta(y)\,\mathrm{d}y \\
&\simeq n \int_{-A_n\,n^{-\nu}}^{A_n\,n^{-\nu}} \left[\frac{\alpha^{i\,\lambda}(y_u)}{\alpha^{i\,\lambda}(y)} - 1 - i\,\lambda\,\ln \frac{\alpha(y_u)}{\alpha(y)} \right] \alpha(y)\,\mathrm{d}y \\
&\quad + n \int_{-A_n\,n^{-\nu}}^{A_n\,n^{-\nu}} \frac{\alpha^{i\,\lambda}(y_u)}{\alpha^{i\,\lambda}(y)} \left(\frac{\beta^{i\,\lambda}(y_u)}{\beta^{i\,\lambda}(y)} - 1 \right) \alpha(y)\,\mathrm{d}y \\
&\quad - i\,\lambda\,n \int_{-A_n\,n^{-\nu}}^{A_n\,n^{-\nu}} \left[\frac{\alpha(y_u)}{\alpha(y)} - 1 - \ln \frac{\alpha(y_u)}{\alpha(y)} \right] \alpha(y)\,\mathrm{d}y \\
&\quad - i\,\lambda\,n \int_{-A_n\,n^{-\nu}}^{A_n\,n^{-\nu}} \alpha(y_u) \left(\frac{\beta(y_u)}{\beta(y)} - 1 \right) \mathrm{d}y \\
&= n\,I_1 + n\,I_2 - i\,\lambda\,n\,I_3 - i\,\lambda\,n\,I_4
\end{aligned}$$

with evident notations.

Using the Hölder continuity of $\beta(\cdot)$, we get

$$\begin{aligned}
|n\,I_4| &\leqslant n \int_{-A_n\,n^{-\nu}}^{A_n\,n^{-\nu}} \alpha(y_u)\,\frac{|\beta(y_u) - \beta(y)|}{\beta(y)}\,\mathrm{d}y \leqslant n\,C\,|u\,n^{-\nu}|^\mu \int_{-A_n\,n^{-\nu}}^{A_n\,n^{-\nu}} \frac{\alpha(y_u)}{\beta(y)}\,\mathrm{d}y \\
&\simeq C\,|u|^\mu\,n^{1-\nu\mu} \int_{-A_n\,n^{-\nu}}^{A_n\,n^{-\nu}} \alpha(y_u)\,\mathrm{d}y = C\,|u|^\mu\,n^{1-\nu\mu} \int_{(-A_n-u)\,n^{-\nu}}^{(A_n-u)\,n^{-\nu}} d(x)|x|^p\,\mathrm{d}x \\
&= C\,|u|^\mu\,n^{1-\nu\mu} \left[\frac{a}{p+1}(A_n + u)^{p+1} + \frac{b}{p+1}(A_n - u)^{p+1} \right] n^{-\nu(p+1)} \\
&\leqslant C\,|u|^\mu\,(A_n + |u|)^{p+1} n^{-\nu\mu} \longrightarrow 0
\end{aligned}$$

if (A_n) is chosen, so that $A_n\,n^{-\nu^2\mu} \to 0$.

Similarly, noting that $\beta^{i\,\lambda}(\cdot)$ is also Hölder continuous of order μ, we get

$$|n\,I_2| \leqslant n \int_{-A_n\,n^{-\nu}}^{A_n\,n^{-\nu}} \frac{|\alpha^{i\,\lambda}(y_u)|}{|\alpha^{i\,\lambda}(y)|}\,\frac{|\beta^{i\,\lambda}(y_u) - \beta^{i\,\lambda}(y)|}{|\beta^{i\,\lambda}(y)|}\,\alpha(y)\,\mathrm{d}y$$

$$= n \int_{-A_n n^{-\nu}}^{A_n n^{-\nu}} |\beta^{i\lambda}(y_u) - \beta^{i\lambda}(y)|\, \alpha(y)\, \mathrm{d}y \leqslant n\, C\, |u\, n^{-\nu}|^{\mu} \int_{-A_n n^{-\nu}}^{A_n n^{-\nu}} \alpha(y)\, \mathrm{d}y$$
$$= C\, |u|^{\mu}\, n^{1-\nu\mu}\, A_n^{p+1}\, n^{-\nu(p+1)} = C\, |u|^{\mu}\, A_n^{p+1}\, n^{-\nu\mu} \longrightarrow 0$$

under the same condition on the choice of (A_n).

For $n\, I_3$ we can write

$$n\, I_3 = n \int_{-A_n n^{-\nu}}^{A_n n^{-\nu}} \left[\frac{d(y_u)|y_u|^p}{d(y)|y|^p} - 1 - \ln \frac{d(y_u)|y_u|^p}{d(y)|y|^p} \right] d(y)|y|^p\, \mathrm{d}y$$
$$= n \int_{-A_n}^{A_n} \left[\frac{d(z-u)}{d(z)} \left|1 - \frac{u}{z}\right|^p - 1 - \ln \left(\frac{d(z-u)}{\mathrm{d}(z)} \left|1 - \frac{u}{z}\right|^p \right) \right] \frac{d(z)|z|^p}{n^{\nu(p+1)}}\, \mathrm{d}z$$
$$\longrightarrow \int_{-\infty}^{\infty} \left[\frac{d(z-u)}{d(z)} \left|1 - \frac{u}{z}\right|^p - 1 - \ln \left(\frac{d(z-u)}{d(z)} \left|1 - \frac{u}{z}\right|^p \right) \right] d(z)|z|^p\, \mathrm{d}z.$$

Note that the last integral is finite, since

$$\frac{d(z-u)}{d(z)} = \left(\frac{a}{b}\right)^{\mathrm{sign}(u)\, \mathbb{1}_{[u^-, u^+]}(z)},$$

and hence the integrand behaves as $C|z|^{p-2}$ as $z \to \infty$.

Finally, for $n\, I_1$, we have

$$n\, I_1 = n \int_{-A_n n^{-\nu}}^{A_n n^{-\nu}} \left[\frac{d^{i\lambda}(y_u)|y_u|^{i\lambda p}}{d^{i\lambda}(y)|y|^{i\lambda p}} - 1 - i\,\lambda\, \ln \frac{d(y_u)|y_u|^p}{d(y)|y|^p} \right] d(y)|y|^{\mathrm{p}}\, \mathrm{d}y$$
$$= n \int_{-A_n}^{A_n} \left[\left(\frac{d(z-u)}{d(z)} \left|1 - \frac{u}{z}\right|^p \right)^{i\lambda} - 1 - i\,\lambda\, \ln \left(\frac{d(z-u)}{d(z)} \left|1 - \frac{u}{z}\right|^p \right) \right] \frac{d(z)|z|^p}{n^{\nu(p+1)}}\, \mathrm{d}z$$
$$\longrightarrow \int_{-\infty}^{\infty} \left[\left(\frac{d(z-u)}{d(z)} \left|1 - \frac{u}{z}\right|^p \right)^{i\lambda} - 1 - i\,\lambda\, \ln \left(\frac{d(z-u)}{d(z)} \left|1 - \frac{u}{z}\right|^p \right) \right] d(z)|z|^p\, \mathrm{d}z,$$

where the last integral is finite as before.

So we get

$$\ln \phi_n \longrightarrow \mathcal{L} = \int_{-\infty}^{\infty} \left[\left(\frac{d(z-u)}{d(z)} \right)^{i\lambda} \left|1 - \frac{u}{z}\right|^{i\lambda p} - 1 - i\,\lambda\, p\, \ln \left|1 - \frac{u}{z}\right| \right] d(z)|z|^p\, \mathrm{d}z$$
$$- i\,\lambda \int_{-\infty}^{\infty} \left[\frac{d(z-u)}{d(z)} \left|1 - \frac{u}{z}\right|^p - 1 - p \ln \left|1 - \frac{u}{z}\right| \right] d(z)|z|^p\, \mathrm{d}z.$$

To terminate the proof of Lemma 5, it remains to show that $\mathcal{L} = \ln \phi$, where $\phi(\cdot)$ is the characteristic function of $\ln Z(u)$.

Recall that

$$\ln Z(u) = p \int_{-\infty}^{+\infty} \ln \left|1 - \frac{u}{z}\right| \pi(\mathrm{d}z) + \ln \frac{a}{b} \int_0^u Y(\mathrm{d}z)$$
$$- \int_{-\infty}^{+\infty} \left[\left|1 - \frac{u}{z}\right|^p - 1 - p \ln \left|1 - \frac{u}{z}\right| \right] d(z)|z|^p\, \mathrm{d}z$$
$$- \frac{a-b}{p+1}\, |u|^{p+1}\, \mathrm{sign}(u)$$
$$= K_1 + K_2 - K_3 - K_4$$

with evident notations.

Hence

$$
\begin{aligned}
\ln \phi(\lambda) &= \ln \mathbf{E} \exp\{i\,\lambda\,\ln Z(u)\} \\
&= \ln \mathbf{E} \exp\{i\,\lambda\,K_1 + i\,\lambda\,K_2\} - i\,\lambda\,K_3 - i\,\lambda\,K_4 \\
&= \ln\,\mathbf{E} \exp\left\{ i\,\lambda \int_{-\infty}^{+\infty} \left[p\,\ln\left|1-\frac{u}{z}\right| + \ln\left(\frac{a}{b}\right)\,\operatorname{sign}(u)\,\mathbb{1}_{[u^-,u^+]}(z)\right] \pi(\mathrm{d}z)\right\} \\
&\quad + i\,\lambda\,\ln\frac{a}{b}\int_0^u d(z)|z|^p\,\mathrm{d}z - i\,\lambda\,K_3 - i\,\lambda\,K_4 \\
&= \int_{-\infty}^{+\infty} \left[\exp\left\{ i\,\lambda\,p \ln\left|1-\frac{u}{z}\right| + i\,\lambda \ln\left(\frac{a}{b}\right)\operatorname{sign}(u)\,\mathbb{1}_{[u^-,u^+]}(z)\right\} - 1 \right. \\
&\quad \left. - i\,\lambda\,p\,\ln\left|1-\frac{u}{z}\right| - i\,\lambda \ln\left(\frac{a}{b}\right)\operatorname{sign}(u)\,\mathbb{1}_{[u^-,u^+]}(z)\right] d(z)|z|^p\,\mathrm{d}z \\
&\quad + i\,\lambda\,\ln\frac{a}{b}\int_0^u d(z)|z|^p\,\mathrm{d}z - i\,\lambda\,K_3 - i\,\lambda\,K_4 \\
&= \int_{-\infty}^{+\infty} \left[\left|1-\frac{u}{z}\right|^{i\,\lambda\,p} \left(\frac{d(z-u)}{d(z)}\right)^{i\,\lambda} - 1 - i\,\lambda\,p\,\ln\left|1-\frac{u}{z}\right| \right] d(z)|z|^p\,\mathrm{d}z \\
&\quad - i\,\lambda\,K_3 - i\,\lambda\,K_4 \\
&= \mathcal{L}(\lambda) + i\,\lambda \int_{-\infty}^{+\infty} \left|1-\frac{u}{z}\right|^p \left(\frac{d(z-u)}{d(z)} - 1\right) d(z)|z|^p\,\mathrm{d}z - i\,\lambda\,K_4 \\
&= \mathcal{L}(\lambda) + i\,\lambda \int_{-\infty}^{+\infty} |z-u|^p\,(a-b)\,\operatorname{sign}(u)\,\mathbb{1}_{[u^-,u^+]}(z)\,\mathrm{d}z - i\,\lambda\,K_4 \\
&= \mathcal{L}(\lambda) + i\,\lambda\,(a-b)\int_0^u |z-u|^p\,\mathrm{d}z - i\,\lambda\,\frac{a-b}{p+1}\,|u|^{p+1}\operatorname{sign}(u) = \mathcal{L}(\lambda).
\end{aligned}
$$

So, the convergence of the one-dimensional distributions is proved. The case of higher-dimensional distributions can be treated similarly. The uniformity in θ on any compact set $\mathbf{K} \subset \Theta$ is evident. Lemma 5 is proved.

Now let us prove Lemma 6. For $|u_1 - u_2| \geqslant 1$ the assertion is evident since for all θ and n we have

$$
\mathbf{E}_\theta \left| Z_n^{1/2}(u_1) - Z_n^{1/2}(u_2)\right|^2 \leqslant 4 \leqslant 4\,|u_1 - u_2|^{2p+1}.
$$

Suppose now that $|u_1 - u_2| \leqslant 1$. Denoting $\Delta = u_1 - u_2$ and using Lemma 1.5 of [7] we can write

$$
\begin{aligned}
\mathbf{E}_\theta \left| Z_n^{1/2}(u_1) - Z_n^{1/2}(u_2)\right|^2 &\leqslant n \int_0^T \left[\sqrt{S_{\theta_{u_1}}(t)} - \sqrt{S_{\theta_{u_2}}(t)}\,\right]^2 \mathrm{d}t \\
&= n \int_0^T \left[\sqrt{s(t-\theta-u_1\,n^{-\nu})} - \sqrt{s(t-\theta-u_2\,n^{-\nu})}\right]^2 \mathrm{d}t \\
&= n \int_{-\theta-u_2\,n^{-\nu}}^{T-\theta-u_2\,n^{-\nu}} \left[\sqrt{s(y-\Delta\,n^{-\nu})} - \sqrt{s(y)}\right]^2 \mathrm{d}y \\
&= n \int_{-\theta-u_2\,n^{-\nu}}^{T-\theta-u_2\,n^{-\nu}} \frac{[s(y-\Delta\,n^{-\nu}) - s(y)]^2}{[\sqrt{s(y-\Delta\,n^{-\nu})} + \sqrt{s(y)}]^2}\,\mathrm{d}y \\
&= n\,I_1 + n\,I_2,
\end{aligned}
$$

where I_1 and I_2 are the integrals of the same function over the interval $(-\delta/2, \delta/2)$ and over the set $E = (-\theta - u_2 n^{-\nu}, T - \theta - u_2 n^{-\nu}) \setminus (-\delta/2, \delta/2)$, respectively.

On the set E we have $|y| \geqslant \delta/2$, and hence $|y - \Delta n^{-\nu}| \geqslant \delta/4$ for n sufficiently large. Recall that on the set $\{y : |y| \geqslant \delta/4\}$ the function $s(\cdot)$ is separated from zero and Hölder continuous of order $\varkappa$. So, for n sufficiently large we get

$$n\, I_2 \leqslant n \int_E \frac{|\Delta n^{-\nu}|^{2\varkappa}}{c}\, \mathrm{d}y \leqslant C\, n |\Delta n^{-\nu}|^{p+1} = C\, |u_1 - u_2|^{p+1}.$$

Further, for the first integral we have

$$\begin{aligned}
n\, I_1 &\leqslant C\, n \int_{-\delta/2}^{\delta/2} \frac{[d(y - \Delta n^{-\nu})|y - \Delta n^{-\nu}|^p - d(y)|y|^p]^2}{[\sqrt{s(y - \Delta n^{-\nu})} + \sqrt{s(y)}]^2}\, \mathrm{d}y \\
&\quad + C\, n \int_{-\delta/2}^{\delta/2} \frac{[\psi(y - \Delta n^{-\nu}) - \psi(y)]^2}{s(y)}\, \mathrm{d}y \\
&\leqslant C\, n \int_{-\delta/2}^{\delta/2} \frac{[d(y - \Delta n^{-\nu})|y - \Delta n^{-\nu}|^p - d(y)|y|^p]^2}{1/2[\sqrt{d(y - \Delta n^{-\nu})}|y - \Delta n^{-\nu}|^{p/2} + \sqrt{d(y)}|y|^{p/2}]^2}\, \mathrm{d}y \\
&\quad + C\, n \int_{-\delta/2}^{\delta/2} \frac{|\Delta n^{-\nu}|^{2\varkappa}}{1/2\, d(y)|y|^p}\, \mathrm{d}y \\
&\leqslant C\, n \int_{-\delta/2}^{\delta/2} \left[\sqrt{d(y - \Delta n^{-\nu})}|y - \Delta n^{-\nu}|^{p/2} - \sqrt{d(y)}|y|^{p/2}\right]^2 \mathrm{d}y \\
&\quad + C\, n\, |\Delta n^{-\nu}|^{p+1} \int_{-\delta/2}^{\delta/2} \frac{1}{d(y)} |y|^{-p}\, \mathrm{d}y \\
&\leqslant C\, n\, |\Delta n^{-\nu}|^{p+1} \int_{-\infty}^{\infty} [\tilde{d}(z-1)|z-1|^{p/2} - \tilde{d}(z)|z|^{p/2}]^2\, \mathrm{d}z + C\, |\Delta|^{p+1} \\
&= C\, |\Delta|^{p+1} = C\, |u_1 - u_2|^{p+1}.
\end{aligned}$$

Here in the last integral we have denoted $\tilde{d}(z) = \sqrt{d(z\,\Delta)}$ and noticed that the integrand behaves as $C|z|^{p-2}$ as $z \to \infty$.

So, in the case $|u_1 - u_2| \leqslant 1$, for all θ and n sufficiently large we get finally

$$\mathbf{E}_\theta |Z_n^{1/2}(u_1) - Z_n^{1/2}(u_2)|^2 \leqslant C\, n\, I_1 + C\, n\, I_2 \leqslant C\, |u_1 - u_2|^{p+1}.$$

Lemma 6 is proved.

It remains to verify Lemma 7. Using Lemma 1.5 of [7], for any n, θ and $u \in U_n$ we can write

$$\mathbf{E}_\theta Z_n^{1/2}(u) \leqslant \exp\left\{-\frac{1}{2}\, n\, F(u\, n^{-\nu})\right\},$$

where for all $u \in (\alpha - \theta, \beta - \theta) \subset (-T, T)$ we denote

$$F(u) = \int_0^T \left[\sqrt{S_{\theta+u}(t)} - \sqrt{S_\theta(t)}\right]^2 \mathrm{d}t.$$

First we suppose $|u| \leqslant \delta/2$. Since $\theta \in \mathbf{K} \subset (0, T)$, we have

$$\begin{aligned}
F(u) &= \int_0^T \left[\sqrt{s(t - \theta - u)} - \sqrt{s(t - \theta)}\right]^2 \mathrm{d}t \\
&= \int_{-\theta}^{T-\theta} \left[\sqrt{s(y - u)} - \sqrt{s(y)}\right]^2 \mathrm{d}y \geqslant \int_{-\varepsilon}^{\varepsilon} \left[\sqrt{s(y - u)} - \sqrt{s(y)}\right]^2 \mathrm{d}y,
\end{aligned}$$

where we can take $0 < \varepsilon < \delta/2$. Hence $|y| \leqslant \delta/2 < \delta$ and $|y-u| \leqslant \delta$, and so we get

$$\begin{aligned}
F(u) &\geqslant \int_{-\varepsilon}^{\varepsilon} \frac{[s(y-u)-s(y)]^2}{[\sqrt{s(y-u)}+\sqrt{s(y)}]^2}\,\mathrm{d}y \\
&\geqslant \int_{-\varepsilon}^{\varepsilon} \frac{[(d(y-u)|y-u|^p - d(y)|y|^p) + (\psi(y-u)-\psi(y))]^2}{[\sqrt{2\,d(y-u)}|y-u|^{p/2}+\sqrt{2\,d(y)}|y|^{p/2}]^2}\,\mathrm{d}y \\
&= c\int_{-\varepsilon}^{\varepsilon}\left[\sqrt{d(y-u)}|y-u|^{p/2}-\sqrt{d(y)}|y|^{p/2}\right]^2\,\mathrm{d}y \\
&\quad + c\int_{-\varepsilon}^{\varepsilon}\frac{[\psi(y-u)-\psi(y)]^2}{[\sqrt{d(y-u)}|y-u|^{p/2}+\sqrt{d(y)}|y|^{p/2}]^2}\,\mathrm{d}y \\
&\quad + c\int_{-\varepsilon}^{\varepsilon}\frac{(\sqrt{d(y-u)}|y-u|^{p/2}-\sqrt{d(y)}|y|^{p/2})(\psi(y-u)-\psi(y))}{\sqrt{d(y-u)}|y-u|^{p/2}+\sqrt{d(y)}|y|^{p/2}}\,\mathrm{d}y \\
&= I_1 + I_2 \pm |I_3|
\end{aligned}$$

with evident notations.

For the first integral we have

$$I_1 = C|u|^p \int_{-\varepsilon/|u|}^{\varepsilon/|u|}\left[\sqrt{d(u(z-1))}|z-1|^{p/2}-\sqrt{d(u\,z)}|z|^{p/2}\right]^2\,\mathrm{d}z,$$

and so $c|u|^p \leqslant I_1 \leqslant C|u|^p$ since the last integral can be bounded from above and from below by the integral of the same function over $\mathbb{R}$ and over $(-\varepsilon/T, \varepsilon/T)$, respectively.

For the second integral we get

$$I_2 \leqslant C|u|^{2\varkappa}\int_{-\varepsilon}^{\varepsilon}\frac{1}{[\sqrt{d(y)}|y|^{p/2}]^2}\,\mathrm{d}y = C|u|^{2\varkappa}.$$

Using Cauchy–Schwarz inequality, we obtain $|I_3| \leqslant C\sqrt{I_1\,I_2} \leqslant C|u|^{\varkappa+(p+1)/2}$ for the last integral, and finally

$$F(u) \geqslant c\,|u|^{p+1} - C|u|^{\varkappa+(p+1)/2} = c\,|u|^{p+1}\left(1 - C\,|u|^{\varkappa-(p+1)/2}\right) \geqslant c\,|u|^{p+1}$$

for u sufficiently small, that is for $|u| \leqslant \sigma$, where $\sigma > 0$ is some fixed constant.

On the other hand, we have also

$$\inf_{|u|\geqslant\sigma} F(u) = c > 0,$$

since otherwise we should have $S_{\theta+u^*}(t) = S_\theta(t)$ for some fixed u^* and almost all $t \in [0, T]$, which is impossible. Hence, for all $|u| \geqslant \sigma$ we can write

$$F(u) \geqslant c \geqslant c\,\frac{|u|^{p+1}}{T^{p+1}} = c\,|u|^{p+1}.$$

So, for all θ and $u \in (\alpha-\theta, \beta-\theta)$ we have

$$F(u) \geqslant c\,|u|^{p+1},$$

and hence for all n, θ, and $u \in U_n$ we can write

$$\mathbf{E}_\theta Z_n^{1/2}(u) \leqslant \exp\left\{-\frac{1}{2}\,n\,F(u\,n^{-\nu})\right\} \leqslant \exp\{-c\,|u|^{p+1}\}.$$

Lemma 7 is proved.

4. Concluding remarks

(1) For simplicity of exposition, in this paper we considered the Bayesian estimators and the notion of asymptotic efficiency in the case of quadratic loss function. In fact, the results hold for a larger class of loss functions (see [9] for more details).

(2) Again for simplicity of exposition, we considered the case where the unknown parameter θ is a shift parameter, that is $S_\theta(t) = s(t-\theta)$. In fact, the results hold in a more general situation, for example when the intensity function is strictly positive (except possibly in θ) and can be written as

$$S_\theta(t) = d(t-\theta)|t-\theta|^p + \Psi(\theta, t),$$

where $p \in (-1, 0) \cup (0, 1)$, the function $d(\cdot)$ is as before, and the function $\Psi(\theta, t)$ is continuous, and uniformly in t Hölder continuous (of order higher than $(p+1)/2$) with respect to θ. In the case $p > 0$ we suppose equally that $\Psi(\theta, \theta) = 0$. It is not difficult to obtain for this case the same results as those presented above.

(3) Like in Chapter 6 of [9], one can consider a situation when the intensity function has several singularities of the same order. More precisely, we suppose that $t_1 < \cdots < t_r$ with $t_r - t_1 < T$, the unknown parameter $\theta \in \Theta = (\alpha, \beta) \subseteq (-t_1, T - t_r)$, and the intensity function is strictly positive and can be written as

$$S_\theta(t) = \sum_{i=1}^{r} d_i(t-\theta-t_i)\,|t-\theta-t_i|^p + \Psi(\theta, t),$$

where $p \in (-1, 0) \cup (0, 1)$,

$$d_i(x) = \begin{cases} a_i & \text{if } x < 0, \\ b_i & \text{if } x > 0, \end{cases}$$

$a_i, b_i > 0$, and the function $\Psi(\theta, t)$ is continuous, and uniformly in t Hölder continuous (of order higher than $(p+1)/2$) with respect to θ. In the case $p > 0$ we suppose equally that $\Psi(\theta, \theta + t_i) = 0$. It is not difficult to obtain for this problem the results similar to those presented above. The difference is that now one needs to introduce the process Z (and hence the random variables ζ and ξ) in a slightly different manner. More precisely, for each $i = 1, \ldots, n$, one should introduce a process Z_i in the same manner (but using constants a_i and b_i instead of a and b) as Z was introduced. Further one should consider the process Z defined by

$$Z(u) = \prod_{i=1}^{r} Z_i(u),$$

where the processes Z_i are independent.

References

[1] D.J. Daley and D. Vere-Jones, *An Introduction to the Theory of Point Processes*, 2nd ed., Springer, New York, 2003.

[2] A.F. Karr, *Statistical models and methods in image analysis: A survey*, in *Statistical Inference in Stochastic Processes*, N.U. Prabhu and I.V. Basawa, eds., Marcel Dekker, New York, 1989.

[3] A.F. Karr, *Point Processes and Their Statistical Inference*, 2nd ed., Marcel Dekker, New York, 1991.

[4] D.R. Snyder and M.I. Miller, *Random Point Processes in Time and Space*, Springer, New York, 1991.

[5] W.A. Thompson, *Point Processes Models with Applications to Safety and Reliability*, Chapman and Hall, New York, 1988.

[6] Yu.A. Kutoyants, *Parameter Estimation for Stochastic Processes*, Armenian Academy of Sciences, Yerevan, 1980 (in Russian); translation of revised version, Heldermann-Verlag, Berlin, 1984.

[7] Yu.A. Kutoyants, *Statistical Inference for Spatial Poisson Processes*, Lecture Notes Statistics, 134, Springer-Verlag, New York, 1998.
[8] S. Dachian, *Estimation of cusp location by Poisson observations*, Stat. Inference Stoch. Process. 6(1) (2003), pp. 1–14.
[9] I.A. Ibragimov and R.Z. Khasminskii, *Statistical Estimation. Asymptotic Theory*, Springer-Verlag, New York, 1981.
[10] I.A. Ibragimov and R.Z. Khasminskii, *The asymptotic behavior of generalized Bayesian estimates*, Dokl. Akad. Nauk SSSR 194 (1970), pp. 257–260.
[11] I.A. Ibragimov and R.Z. Khasminskii, *Asymptotic behavior of statistical estimators of the location parameter for samples with continuous density with singularities*, Zap. Nauchn. Sem. Leningrad. Otdel. Mat. Inst. Steklov. (LOMI) 41 (1974), pp. 67–93, 166–167.
[12] M.S. Ermakov, *On the asymptotic behavior of statistical estimates for samples having a density with singularities*, Theory Probab. Appl. 21 (1976), pp. 649–651.

Stat Inference Stoch Process (2011) 14:255–271
DOI 10.1007/s11203-011-9059-x

On compound Poisson processes arising in change-point type statistical models as limiting likelihood ratios

Sergueï Dachian · Ilia Negri

Received: 29 July 2010 / Accepted: 7 February 2011 / Published online: 30 June 2011

Abstract Different change-point type models encountered in statistical inference for stochastic processes give rise to different limiting likelihood ratio processes. In a previous paper of one of the authors it was established that one of these likelihood ratios, which is an exponential functional of a two-sided Poisson process driven by some parameter, can be approximated (for sufficiently small values of the parameter) by another one, which is an exponential functional of a two-sided Brownian motion. In this paper we consider yet another likelihood ratio, which is the exponent of a two-sided compound Poisson process driven by some parameter. We establish, that similarly to the Poisson type one, the compound Poisson type likelihood ratio can be approximated by the Brownian type one for sufficiently small values of the parameter. We equally discuss the asymptotics for large values of the parameter and illustrate the results by numerical simulations.

Keywords Compound Poisson process · Non-regularity · Change-point · Limiting likelihood ratio process · Bayesian estimators · Maximum likelihood estimator · Limiting distribution · Limiting mean squared error · Asymptotic relative efficiency

Mathematics Subject Classification (2000) 62F99 · 62M99

1 Introduction

In this work we are interested by the asymptotic study of non-regular parametric statistical models encountered in statistical inference for stochastic processes. An exhaustive

S. Dachian
Laboratoire de Mathématiques UMR6620, Université Blaise Pascal, Clermont Université,
Aubière Cedex, 63177, France
e-mail: Serguei.Dachian@math.univ-bpclermont.fr

I. Negri (✉)
Department of Information Technology and Mathematical Methods, University of Bergamo,
Dalmine, 24044 BG, Italy
e-mail: ilia.negri@unibg.it

exposition of the parameter estimation theory in both regular and non-regular cases is given in the classical book (Ibragimov and Khasminskii 1981). They have developed a general theory of estimation based on the analysis of renormalized likelihood ratio. Their approach consists in proving first that the renormalized likelihood ratio (with a properly chosen renormalization rate) weekly converges to some non-degenerate limit: the limiting likelihood ratio process. Thereafter, the properties of the estimators (namely their rate of convergence and limiting distributions) are deduced. Finally, based on the estimators, one can also construct confidence intervals, tests, and so on. Note that this approach also provides the convergence of moments, allowing one to deduce equally the asymptotics of some statistically important quantities, such as the mean squared errors of the estimators.

It is well known that in the regular case the limiting likelihood ratio is given by the LAN property and is the same for different models (the renormalization rate being usually $1/\sqrt{n}$). So, the classical estimators—the maximum likelihood estimator and the Bayesian estimators—are consistent, asymptotically normal (usually with rate $1/\sqrt{n}$) and asymptotically efficient.

In non-regular cases the situation essentially changes: the renormalization rate is usually better (for example, $1/n$ in change-point type models), but the limiting likelihood ratio can be different in different models. So, the classical estimators are still consistent, but may have different limiting distributions (though with a better rate) and, in general, only the Bayesian estimators are asymptotically efficient.

In Dachian (2010) a relation between two different limiting likelihood ratios arising in change-point type models was established by one of the authors. More precisely, it was shown that the first one, which is an exponential functional of a two-sided Poisson process driven by some parameter, can be approximated (for sufficiently small values of the parameter) by the second one, defined by

$$Z_0(x) = \exp\left\{W(x) - \frac{1}{2}\,|x|\right\}, \quad x \in \mathbb{R}, \tag{1}$$

where W is a standard two-sided Brownian motion. In this paper we consider yet another limiting likelihood ratio process arising in change-point type models and show that it is related to Z_0 in a similar way.

1.1 The process $Z_{\gamma,f}$

We introduce the random process $Z_{\gamma,f}$ on $\mathbb{R}$ as the exponent of a two-sided compound Poisson process given by

$$\ln Z_{\gamma,f}(x) = \begin{cases} \sum_{k=1}^{\Pi_+(x)} \ln \frac{f(\varepsilon_k^+ + \gamma)}{f(\varepsilon_k^+)}, & \text{if } x \geq 0, \\ \sum_{k=1}^{\Pi_-(-x)} \ln \frac{f(\varepsilon_k^- - \gamma)}{f(\varepsilon_k^-)}, & \text{if } x \leq 0, \end{cases} \tag{2}$$

where $\gamma > 0$, f is a strictly positive density of some random variable ε with mean 0 and variance 1, Π_+ and Π_- are two independent Poisson processes of intensity 1 on $\mathbb{R}_+$, $\varepsilon_k^\pm$ are independent random variables with density f which are also independent of $\Pi_\pm$, and we use the convention $\sum_{k=1}^{0} a_k = 0$. We equally introduce the random variables

$$\begin{aligned} \zeta_{\gamma,f} &= \frac{\int_{\mathbb{R}} x\, Z_{\gamma,f}(x)\, dx}{\int_{\mathbb{R}} Z_{\gamma,f}(x)\, dx}, \\ \xi^-_{\gamma,f} &= \inf\left\{z : Z_{\gamma,f}(z) = \sup_{x\in\mathbb{R}} Z_{\gamma,f}(x)\right\}, \end{aligned} \tag{3}$$

$$\xi^{+}_{\gamma,f} = \sup\left\{z : Z_{\gamma,f}(z) = \sup_{x\in\mathbb{R}} Z_{\gamma,f}(x)\right\},$$
$$\xi^{\alpha}_{\gamma,f} = \alpha\,\xi^{-}_{\gamma,f} + (1-\alpha)\,\xi^{+}_{\gamma,f}, \quad \alpha \in [0,1],$$

related to this process, as well as their second moments $B_{\gamma,f} = \mathbf{E}\zeta^2_{\gamma,f}$ and $M^{\alpha}_{\gamma,f} = \mathbf{E}(\xi^{\alpha}_{\gamma,f})^2$.

An important particular case of this process is the one where the density f is Gaussian, that is, $\varepsilon \sim \mathcal{N}(0,1)$. In this case we will omit the index f and write Z_γ instead of $Z_{\gamma,f}$, ξ^{α}_{γ} instead of $\xi^{\alpha}_{\gamma,f}$, and so on. Note that since

$$\ln\frac{f(\varepsilon\pm\gamma)}{f(\varepsilon)} = \mp\gamma\varepsilon - \frac{\gamma^2}{2} \sim \mathcal{N}(-\gamma^2/2, \gamma^2),$$

the process Z_γ is symmetric and has Gaussian jumps.

The process $Z_{\gamma,f}$, up to a linear time change, arises in some non-regular, namely change-point type, statistical models as the limiting likelihood ratio process, and the variables $\zeta_{\gamma,f}$ and $\xi^{\alpha}_{\gamma,f}$ as the limiting distributions of the Bayesian estimators and of the appropriately chosen maximum likelihood estimator, respectively. The maximum likelihood estimator being not unique in the underlying models, the appropriate choice here is a linear combination with weights α and $1-\alpha$ of its minimal and maximal values. Moreover, the quantities $B_{\gamma,f}$ and $M^{\alpha}_{\gamma,f}$ are the limiting mean squared errors (sometimes also called limiting variances) of these estimators and, the Bayesian estimators being asymptotically efficient, the ratio $E^{\alpha}_{\gamma,f} = B_{\gamma,f}/M^{\alpha}_{\gamma,f}$ is the asymptotic relative efficiency of this maximum likelihood estimator.

The examples include the two-phase regression model and the threshold autoregressive (TAR) model. The linear case of the former was studied by Koul and Qian (2002), while the non-linear one was investigated by Ciuperca (2004). Concerning the TAR model, the first results were obtained by Chan (1993) where he studies the least squares estimator, which is, in the Gaussian case, equivalent to the maximum likelihood estimator. A more recent and exhaustive study was performed by Chan and Kutoyants (submitted) (for the Gaussian TAR model) and in Chan and Kutoyants (to appear). Note finally that in both models, the parameter γ of the limiting likelihood ratio is related to the jump size of the model.

1.2 The process Z_0

On the other hand, many change-point type statistical models encountered in various fields of statistical inference for stochastic processes rather have as limiting likelihood ratio process, up to a linear time change, the process Z_0 defined by (1). In this case, the limiting distributions of the Bayesian estimators and of the maximum likelihood estimator are given by

$$\zeta_0 = \frac{\int_{\mathbb{R}} x\,Z_0(x)\,dx}{\int_{\mathbb{R}} Z_0(x)\,dx} \quad\text{and}\quad \xi_0 = \operatorname*{argsup}_{x\in\mathbb{R}} Z_0(x), \tag{4}$$

respectively, while the limiting mean squared errors of these estimators are $B_0 = \mathbf{E}\zeta^2_0$ and $M_0 = \mathbf{E}\xi^2_0$. The Bayesian estimators are still asymptotically efficient, and the asymptotic relative efficiency of the maximum likelihood estimator is $E_0 = B_0/M_0$.

A well-known example is the model of a discontinuous signal in a white Gaussian noise exhaustively studied by Ibragimov and Khasminskii (1975) and Ibragimov and Khasminskii (1981, Chap. 7.2), but one can also cite change-point type models of dynamical systems with small noise considered by Kutoyants (1984) and Kutoyants (1994, Chap. 5), those of ergodic diffusion processes examined by Kutoyants (2004, Chap. 3), a change-point type model of

a delay equation analyzed by Küchler and Kutoyants (2000), a model of a discontinuous periodic signal in a time inhomogeneous diffusion explored by Höpfner and Kutoyants (2010), a change-point type model of a threshold diffusion process investigated by Kutoyants (to appear), and so on.

Let us also note that Terent'yev (1968) determined the Laplace transform of $\mathbf{P}(|\xi_0| > t)$ and calculated the constant $M_0 = 26$. Moreover, the explicit expression of the density of ξ_0 was later successively provided by Bhattacharya and Brockwell (1976), Yao (1987) and Fujii (2007). Regarding the constant B_0, Ibragimov and Khasminskii (1981, Chap. 7.3) showed by means of numerical simulation that $B_0 = 19.5 \pm 0.5$, and so $E_0 = 0.73 \pm 0.03$. Later Golubev (1979) expressed B_0 in terms of the second derivative (with respect to a parameter) of an improper integral of a composite function of modified Hankel and Bessel functions. Finally Rubin and Song (1995) obtained the exact values $B_0 = 16\,\zeta(3)$ and $E_0 = 8\,\zeta(3)/13$, where ζ is Riemann's zeta function defined by $\zeta(s) = \sum_{n=1}^{\infty} 1/n^s$.

1.3 The results of the present paper

In this paper we establish that the limiting likelihood ratio processes $Z_{\gamma,f}$ and Z_0 are related. More precisely, under some regularity assumptions on f, we show that as $\gamma \to 0$, the process $Z_{\gamma,f}(y/I\gamma^2)$, $y \in \mathbb{R}$, (where I is the Fisher information related to f) converges weakly in the space $\mathcal{D}_0(-\infty, +\infty)$ (the Skorohod space of functions on $\mathbb{R}$ without discontinuities of the second kind and vanishing at infinity) to the process Z_0. Hence, the random variables $I\gamma^2\zeta_{\gamma,f}$ and $I\gamma^2\xi^{\alpha}_{\gamma,f}$ converge weakly to the random variables ζ_0 and ξ_0, respectively. We show equally that the convergence of moments of these random variables holds and so, in particular, $I^2\gamma^4 B_{\gamma,f} \to 16\,\zeta(3)$, $I^2\gamma^4 M^{\alpha}_{\gamma,f} \to 26$ and $E^{\alpha}_{\gamma,f} \to 8\,\zeta(3)/13$. Besides their theoretical interest, these results have also some practical implications. For example, they allow to construct tests and confidence intervals on the base of the distributions of ζ_0 and ξ_0 (rather than on the base of those of $\zeta_{\gamma,f}$ and $\xi^{\alpha}_{\gamma,f}$, which depend on the density f and are not known explicitly) in models having the process $Z_{\gamma,f}$ with a small γ as a limiting likelihood ratio. Also, the limiting mean squared errors of the estimators and the asymptotic relative efficiency of the maximum likelihood estimator can be approximated as

$$B_{\gamma,f} \approx \frac{16\,\zeta(3)}{I^2\gamma^4}, \quad M^{\alpha}_{\gamma,f} \approx \frac{26}{I^2\gamma^4} \quad \text{and} \quad E^{\alpha}_{\gamma,f} \approx \frac{8\,\zeta(3)}{13}$$

in such models.

These are the main results of the present paper, and they are presented in Sect. 2, where we also briefly discuss the second possible asymptotics $\gamma \to +\infty$ and present some numerical simulations of the quantities B_γ, M^{α}_{γ} and E^{α}_{γ} for $\gamma \in]0, \infty[$. Finally, the proofs of the necessary lemmas are carried out in Sect. 3.

Concluding the introduction let us note that a preliminary exposition (in the particular Gaussian case) of the results of the present paper can be found in Dachian and Negri (2009) and (2010).

2 Asymptotics of $Z_{\gamma,f}$

Let $\gamma > 0$, and let f be a strictly positive density of some random variable ε with mean 0 and variance 1.

Stat Inference Stoch Process (2011) 14:255–271

2.1 Regularity assumptions

We will always suppose that $\sqrt{f}$ is continuously differentiable in L^2, that is, there exists $\psi \in L^2$ satisfying $\int_{\mathbb{R}} \left(\sqrt{f(x+h)}-\sqrt{f(x)}-h\,\psi(x)\right)^2 dx = o(h^2)$ and $\int_{\mathbb{R}} (\psi(x+h)-\psi(x))^2 dx = o(1)$, as well as that $\|\psi\| > 0$.

Note that under this assumptions, the model of i.i.d. observations with density $f(x+\theta)$ is, in particular, LAN at $\theta = 0$ with Fisher information $I = 4\,\|\psi\|^2 = 4\int_{\mathbb{R}} \psi^2(x)\,dx$ (see, for example, Ibragimov and Khasminskii (1981, Chap. 2.1)) and so, using characteristic functions, we have

$$\lim_{n\to\infty} \left(\mathbf{E} e^{it \ln \frac{f(\varepsilon+u/\sqrt{n})}{f(\varepsilon)}}\right)^n = e^{i\left(-\frac{Iu^2}{2}\right)t-\frac{1}{2}Iu^2t^2}$$

and, more generally,

$$\lim_{\gamma\to 0} \left(\mathbf{E} e^{it \ln \frac{f(\varepsilon+\gamma)}{f(\varepsilon)}}\right)^{1/\gamma^2} = e^{i\left(-\frac{I}{2}\right)t-\frac{1}{2}It^2} \tag{5}$$

for all $t \in \mathbb{R}$.

Note also, that only the convergence (5) will be needed in our considerations. So, one can rather assume it directly, or make any other regularity assumptions sufficient for it as, for example, Hájek's conditions: f is differentiable and the Fisher information $I = \int_{\mathbb{R}} f^{-1}(x)\left(f'(x)\right)^2 dx$ is finite and strictly positive (see, for example, Ibragimov and Khasminskii (1981, Chap. 2.2)).

Note finally, that in the Gaussian case the regularity assumptions clearly hold and we have $I = 1$.

2.2 The asymptotics $\gamma \to 0$

Let us consider the process $X_{\gamma,f}(y) = Z_{\gamma,f}(y/I\gamma^2)$, $y \in \mathbb{R}$, where $Z_{\gamma,f}$ is defined by (2). Note that

$$\frac{\int_{\mathbb{R}} y\,X_{\gamma,f}(y)\,dy}{\int_{\mathbb{R}} X_{\gamma,f}(y)\,dy} = I\gamma^2\zeta_{\gamma,f}\,,$$

$$\inf\left\{z : X_{\gamma,f}(z) = \sup_{y\in\mathbb{R}} X_{\gamma,f}(y)\right\} = I\gamma^2\xi^-_{\gamma,f}$$

and

$$\sup\left\{z : X_{\gamma,f}(z) = \sup_{y\in\mathbb{R}} X_{\gamma,f}(y)\right\} = I\gamma^2\xi^+_{\gamma,f}\,,$$

where the random variables $\zeta_{\gamma,f}$ and $\xi^{\pm}_{\gamma,f}$ are defined by (3). Remind also the process Z_0 on $\mathbb{R}$ defined by (1) and the random variables ζ_0 and ξ_0 defined by (4). Recall finally the quantities $B_{\gamma,f} = \mathbf{E}\zeta^2_{\gamma,f}$, $M^{\alpha}_{\gamma,f} = \mathbf{E}(\xi^{\alpha}_{\gamma,f})^2$, $E^{\alpha}_{\gamma,f} = B_{\gamma,f}/M^{\alpha}_{\gamma,f}$, as well as $B_0 = \mathbf{E}\zeta_0^2 = 16\,\zeta(3)$, $M_0 = \mathbf{E}\xi_0^2 = 26$ and $E_0 = B_0/M_0 = 8\,\zeta(3)/13$. Now we can state the main result of the present paper.

Theorem 1 *The process $X_{\gamma,f}$ converges weakly in the space $\mathcal{D}_0(-\infty,+\infty)$ to the process Z_0 as $\gamma \to 0$. In particular, the random variable $I\gamma^2\zeta_{\gamma,f}$ converges weakly to the random*

variable ζ_0 and, for any $\alpha \in [0, 1]$, the random variable $I\gamma^2\xi^\alpha_{\gamma,f}$ converges weakly to the random variable ξ_0. Moreover, for any $k > 0$ we have

$$I^k\gamma^{2k}\,\mathbf{E}\zeta^k_{\gamma,f} \to \mathbf{E}\zeta^k_0 \quad \text{and} \quad I^k\gamma^{2k}\,\mathbf{E}(\xi^\alpha_{\gamma,f})^k \to \mathbf{E}\xi^k_0.$$

In particular, $I^2\gamma^4 B_{\gamma,f} \to 16\,\zeta(3)$, $I^2\gamma^4 M^\alpha_{\gamma,f} \to 26$ and $E^\alpha_{\gamma,f} \to 8\,\zeta(3)/13$.

The results concerning the random variable $\zeta_{\gamma,f}$ are direct consequence of Ibragimov and Khasminskii (1981, Theorem 1.10.2) and the following three lemmas.

Lemma 2 *The finite-dimensional distributions of the process $X_{\gamma,f}$ converge to those of Z_0 as $\gamma \to 0$.*

Lemma 3 *For any $C > 1/4$ we have*

$$\mathbf{E}\left|X^{1/2}_{\gamma,f}(y_1) - X^{1/2}_{\gamma,f}(y_2)\right|^2 \le C\,|y_1 - y_2|$$

for all sufficiently small γ and all $y_1, y_2 \in \mathbb{R}$.

Lemma 4 *For any $c \in \,]\,0\,,\ 1/8\,[$ we have*

$$\mathbf{E}X^{1/2}_{\gamma,f}(y) \le \exp\left(-c\,|y|\right)$$

for all sufficiently small γ and all $y \in \mathbb{R}$.

Note that these lemmas are not sufficient to establish the weak convergence of the process $X_{\gamma,f}$ in the space $\mathcal{D}_0(-\infty, +\infty)$ and the results concerning the random variable $\xi^\alpha_{\gamma,f}$. However, the increments of the process $\ln X_{\gamma,f}$ being independent, the convergence of its restrictions (and hence of those of $X_{\gamma,f}$) on finite intervals $[A, B] \subset \mathbb{R}$ (that is, convergence in the Skorohod space $\mathcal{D}[A, B]$ of functions on $[A, B]$ without discontinuities of the second kind) follows from Gihman and Skorohod (1974, Theorem 6.5.5), Lemma 2 and the following lemma.

Lemma 5 *For any $\delta > 0$ we have*

$$\lim_{h\to 0}\ \lim_{\gamma\to 0}\ \sup_{|y_1-y_2|<h} \mathbf{P}\left\{\left|\ln X_{\gamma,f}(y_1) - \ln X_{\gamma,f}(y_2)\right| > \delta\right\} = 0.$$

Now, Theorem 1 follows from the following estimate on the tails of the process $X_{\gamma,f}$ by standard argument (see, for example, Ibragimov and Khasminskii (1981)).

Lemma 6 *For any $b \in \,]\,0\,,\ 1/12\,[$ we have*

$$\mathbf{P}\left\{\sup_{|y|>A} X_{\gamma,f}(y) > e^{-bA}\right\} \le 4\,e^{-bA}$$

for all sufficiently small γ and all $A > 0$.

The proofs of all these lemmas will be given in Sect. 3.

Stat Inference Stoch Process (2011) 14:255–271

2.3 The asymptotics $\gamma \to +\infty$

Now let us discuss the second possible asymptotics $\gamma \to +\infty$. It can be shown that in this case, the process $Z_{\gamma,f}$ converges weakly in the space $\mathcal{D}_0(-\infty,+\infty)$ to the process $Z_\infty(x) = \mathbb{1}_{\{-\eta<x<\tau\}}$, $x \in \mathbb{R}$, where η and τ are two independent exponential random variables with parameter 1. So, the random variables $\zeta_{\gamma,f}$, $\xi^-_{\gamma,f}$, $\xi^+_{\gamma,f}$ and $\xi^\alpha_{\gamma,f}$ converge weakly to the random variables

$$\zeta_\infty = \frac{\int_{\mathbb{R}} x\, Z_\infty(x)\, dx}{\int_{\mathbb{R}} Z_\infty(x)\, dx} = \frac{\tau-\eta}{2},$$

$$\xi^-_\infty = \inf\left\{z : Z_\infty(z) = \sup_{x\in\mathbb{R}} Z_\infty(x)\right\} = -\eta,$$

$$\xi^+_\infty = \sup\left\{z : Z_\infty(z) = \sup_{x\in\mathbb{R}} Z_\infty(x)\right\} = \tau$$

and

$$\xi^\alpha_\infty = \alpha\,\xi^-_\infty + (1-\alpha)\,\xi^+_\infty = (1-\alpha)\,\tau - \alpha\,\eta,$$

respectively. It can be equally shown that, moreover, for any $k > 0$ we have

$$\mathbf{E}\zeta^k_{\gamma,f} \to \mathbf{E}\zeta^k_\infty \quad \text{and} \quad \mathbf{E}(\xi^\alpha_{\gamma,f})^k \to \mathbf{E}(\xi^\alpha_\infty)^k.$$

In particular, denoting $B_\infty = \mathbf{E}\zeta^2_\infty$, $M^\alpha_\infty = \mathbf{E}(\xi^\alpha_\infty)^2$ and $E^\alpha_\infty = B_\infty/M^\alpha_\infty$, we finally have

$$B_{\gamma,f} \to B_\infty = \mathbf{E}\left(\frac{\tau-\eta}{2}\right)^2 = \frac{1}{2},$$

$$M^\alpha_{\gamma,f} \to M^\alpha_\infty = \mathbf{E}\left((1-\alpha)\,\tau - \alpha\,\eta\right)^2 = 6\left(\alpha - \frac{1}{2}\right)^2 + \frac{1}{2} \tag{6}$$

and

$$E^\alpha_{\gamma,f} \to E^\alpha_\infty = \frac{1}{12\left(\alpha-\frac{1}{2}\right)^2 + 1}. \tag{7}$$

Let us note that these convergences are natural, since the process Z_∞ can be considered as a particular case of the process $Z_{\gamma,f}$ with $\gamma = +\infty$ under natural conventions $f(\varepsilon \pm \infty) = 0$ and $\ln 0 = -\infty$.

Note also, that Z_∞ is the limiting likelihood ratio process in the problem of estimating the parameter θ by i.i.d. uniform observations on $[\theta, \theta+1]$. So, in this problem, the variables ζ_∞ and ξ^α_∞ are the limiting distributions of the Bayesian estimators and of the appropriately chosen maximum likelihood estimator, respectively, while B_∞ and M^α_∞ are the limiting mean squared errors of these estimators and, the Bayesian estimators being asymptotically efficient, E^α_∞ is the asymptotic relative efficiency of this maximum likelihood estimator.

Finally observe, that the formulae (6) and (7) clearly imply that in the latter problem (as well as in any problem having Z_∞ as limiting likelihood ratio) the best choice of the maximum likelihood estimator is $\alpha = 1/2$, and that the so chosen maximum likelihood estimator is asymptotically efficient. This choice was also suggested for TAR model (which has limiting likelihood ratio Z_γ) by Chan and Kutoyants (submitted). For large values of γ this suggestion is confirmed by our asymptotic results. However, we see that for small values of γ the choice of α will not be so important, since the limits in Theorem 1 do not depend on α.

2.4 Numerical simulations

Here we present some numerical simulations (in the Gaussian case) of the quantities B_γ, M_γ^α and E_γ^α for $\gamma \in]0, \infty[$. Besides giving approximate values of these quantities, the simulation results illustrate both the asymptotics

$$B_\gamma = \frac{B_0}{\gamma^4} + o(\gamma^{-4}), \quad M_\gamma^\alpha = \frac{M_0}{\gamma^4} + o(\gamma^{-4}) \quad \text{and} \quad E_\gamma^\alpha \to E_0 \quad \text{as} \quad \gamma \to 0,$$

with $B_0 = 16\,\zeta(3) \approx 19.2329$, $M_0 = 26$ and $E_0 = 8\,\zeta(3)/13 \approx 0.7397$, and

$$B_\gamma \to B_\infty, \quad M_\gamma^\alpha \to M_\infty^\alpha \quad \text{and} \quad E_\gamma^\alpha \to E_\infty^\alpha \quad \text{as} \quad \gamma \to \infty,$$

with $B_\infty = 0.5$, $M_\infty^\alpha = 6\,(\alpha - 0.5)^2 + 0.5$ and $E_\infty^\alpha = 1/\left(12\,(\alpha - 0.5)^2 + 1\right)$.

First, we simulate the events $x_1^+, x_2^+, \ldots$ of the Poisson process Π_+ and the events $x_1^-, x_2^-, \ldots$ of the Poisson process Π_- (both of intensity 1), as well as the partial sums $S_1^+, S_2^+, \ldots$ of the i.i.d. $\mathcal{N}(0, 1)$ sequence $\varepsilon_1^+, \varepsilon_2^+, \ldots$ and the partial sums $S_1^-, S_2^-, \ldots$ of the i.i.d. $\mathcal{N}(0, 1)$ sequence $\varepsilon_1^-, \varepsilon_2^-, \ldots$. For convenience we also put $x_0^+ = x_0^- = S_0^+ = S_0^- = 0$.

Then we calculate

$$\begin{aligned}
\zeta_\gamma &= \frac{\int_{\mathbb{R}} x\, Z_\gamma(x)\, dx}{\int_{\mathbb{R}} Z_\gamma(x)\, dx} \\
&= \frac{\sum_{i=0}^\infty \frac{1}{2}\, e^{S_i^+} \left((x_{i+1}^+)^2 - (x_i^+)^2 \right) - \sum_{i=0}^\infty \frac{1}{2}\, e^{S_i^-} \left((x_{i+1}^-)^2 - (x_i^-)^2 \right)}{\sum_{i=0}^\infty e^{S_i^+} (x_{i+1}^+ - x_i^+) + \sum_{i=0}^\infty e^{S_i^-} (x_{i+1}^- - x_i^-)}, \\
\xi_\gamma^- &= \inf \left\{ z : Z_\gamma(z) = \sup_{x \in \mathbb{R}} Z_\gamma(x) \right\} = \begin{cases} x_k^+, & \text{if } S_k^+ > S_\ell^-, \\ -x_{\ell+1}^-, & \text{otherwise,} \end{cases} \\
\xi_\gamma^+ &= \sup \left\{ z : Z_\gamma(z) = \sup_{x \in \mathbb{R}} Z_\gamma(x) \right\} = \begin{cases} x_{k+1}^+, & \text{if } S_k^+ \geq S_\ell^-, \\ -x_\ell^-, & \text{otherwise,} \end{cases}
\end{aligned}$$

and

$$\xi_\gamma^\alpha = \alpha\, \xi_\gamma^- + (1 - \alpha)\, \xi_\gamma^+,$$

where

$$k = \operatorname*{argmax}_{i \geq 0} S_i^+ \quad \text{and} \quad \ell = \operatorname*{argmax}_{i \geq 0} S_i^-,$$

and we use the values $1/2$, $1/4$ and 0 for α. Note that in this Gaussian case (due to the symmetry of the process Z_γ) the random variable $\xi_\gamma^{1-\alpha}$ has the same law as the variable $-\xi_\gamma^\alpha$, that's why we use for α only values less or equal than $1/2$.

Finally, repeating these simulations 10^7 times (for each value of γ), we approximate $B_\gamma = \mathbf{E}\zeta_\gamma^2$ and $M_\gamma^\alpha = \mathbf{E}(\xi_\gamma^\alpha)^2$ by the empirical second moments, and $E_\gamma^\alpha = B_\gamma / M_\gamma^\alpha$ by their ratio.

The results of the numerical simulations are presented in Figs. 1, 2, 3. The $\gamma \to 0$ asymptotics of the limiting mean squared errors is illustrated in Fig. 1, where we rather plotted the functions $\gamma^4 B_\gamma$ and $\gamma^4 M_\gamma^\alpha$, making apparent the constants $B_0 \approx 19.2329$ and $M_0 = 26$. One can observe here that the choice $\alpha = 1/2$ is the best one, though its advantage diminishes as γ approaches 0 and seems negligible for $\gamma < 1$.

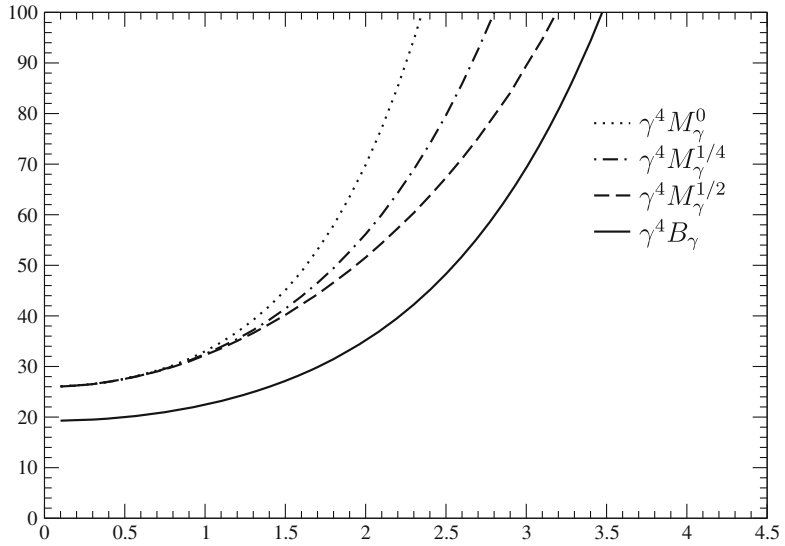

Fig. 1 $\gamma^4 B_\gamma$ and $\gamma^4 M_\gamma^\alpha$ ($\gamma \to 0$ asymptotics)

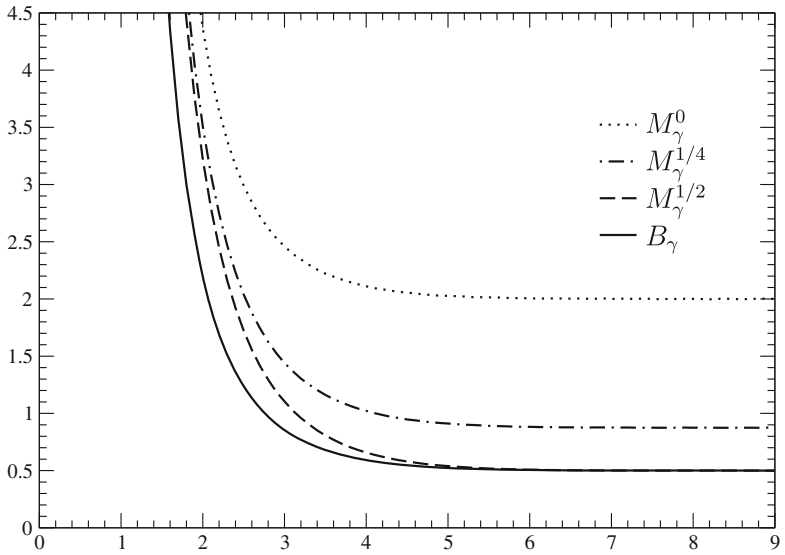

Fig. 2 B_γ and M_γ^α ($\gamma \to \infty$ asymptotics)

In Fig. 2 we illustrate the $\gamma \to \infty$ asymptotics of the limiting mean squared errors by plotting the functions B_γ and M_γ^α themselves. Here the advantage of the choice $\alpha = 1/2$ is obvious, and one can observe that for $\gamma > 5$ this choice makes negligible the loss of efficiency resulting from the use of the maximum likelihood estimator instead of the asymptotically efficient Bayesian estimators.

Finally, in Fig. 3 we illustrate the behavior both at 0 and at ∞ of the asymptotic relative efficiency of the maximum likelihood estimators by plotting the functions E_γ^α. All the observations made above can be once more noticed in this figure. Note also that as γ increases from 0 to ∞, the asymptotic relative efficiency seems first to decrease from $E_0 \approx 0.7397$ for all the maximum likelihood estimators, before increasing back to E_∞^α for the maximum likelihood estimators with α close to the optimal value $1/2$.

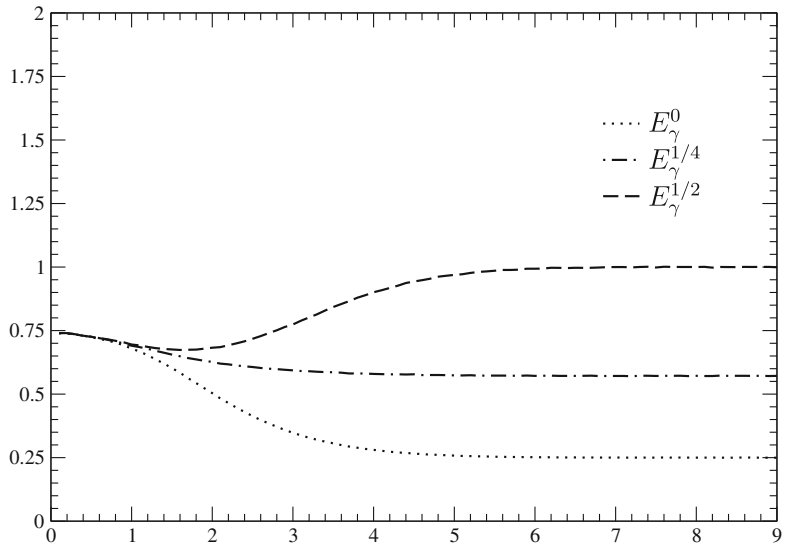

Fig. 3 E_γ^α (both asymptotics)

3 Proofs of the Lemmas

For the sake of clarity, for each lemma we will first give the proof in the particular Gaussian case (in which it is more explicit) and then explain how it can be extended to the general one.

3.1 Proof of Lemma 2

Note that the restrictions of the process $\ln X_\gamma(y) = \ln Z_\gamma(y/\gamma^2)$, $y \in \mathbb{R}$, (as well as those of the process $\ln Z_0$) on $\mathbb{R}_+$ and on $\mathbb{R}_-$ are mutually independent processes with stationary and independent increments. So, to obtain the convergence of all the finite-dimensional distributions, it is sufficient to show the convergence of one-dimensional distributions only, that is, the weak convergence of $\ln X_\gamma(y)$ to

$$\ln Z_0(y) = W(y) - \frac{|y|}{2} \sim \mathcal{N}\left(-\frac{|y|}{2}, |y|\right)$$

for all $y \in \mathbb{R}$. Moreover, these processes being symmetric, it is sufficient to consider $y \in \mathbb{R}_+$ only.

The characteristic function $\varphi_\gamma(t)$ of $\ln X_\gamma(y)$ is

$$\begin{aligned}
\varphi_\gamma(t) &= \mathbf{E}\, e^{it \ln X_\gamma(y)} = \mathbf{E}\, e^{-it\gamma \sum_{k=1}^{\Pi_+(y/\gamma^2)} \varepsilon_k^+ - it\frac{\gamma^2}{2}\Pi_+(y/\gamma^2)} \\
&= \mathbf{E}\,\mathbf{E}\left(e^{-it\gamma \sum_{k=1}^{\Pi_+(y/\gamma^2)} \varepsilon_k^+ - it\frac{\gamma^2}{2}\Pi_+(y/\gamma^2)} \,\middle|\, \mathscr{F}_{\Pi_+} \right) \\
&= \mathbf{E}\left(e^{-it\frac{\gamma^2}{2}\Pi_+(y/\gamma^2)} \prod_{k=1}^{\Pi_+(y/\gamma^2)} \mathbf{E}\, e^{-it\gamma\varepsilon_k^+} \right) \\
&= \mathbf{E}\, e^{-it\frac{\gamma^2}{2}\Pi_+(y/\gamma^2) - \frac{t^2\gamma^2}{2}\Pi_+(y/\gamma^2)} = \mathbf{E}\, e^{-\frac{\gamma^2}{2}(it+t^2)\Pi_+(y/\gamma^2)}
\end{aligned}$$

where we have denoted $\mathscr{F}_{\Pi_+}$ the σ-algebra related to the Poisson process Π_+, used the independence of ε_k^+ and Π_+ and recalled that $\mathbf{E}\, e^{it\varepsilon} = e^{-t^2/2}$.

Then, noting that $\Pi_+\left(y/\gamma^2\right)$ is a Poisson random variable of parameter y/γ^2 with moment generating function $\mathbf{E}\, e^{t\,\Pi_+(y/\gamma^2)} = \exp\left(\frac{y}{\gamma^2}\left(e^t-1\right)\right)$, we get

$$\begin{aligned}\ln\varphi_\gamma(t) &= \frac{y}{\gamma^2}\left(e^{-\frac{\gamma^2}{2}(it+t^2)}-1\right) = \frac{y}{\gamma^2}\left(-\frac{\gamma^2}{2}\left(it+t^2\right)+o\left(\gamma^2\right)\right)\\ &= -\frac{y}{2}\left(it+t^2\right)+o(1) \to -\frac{y}{2}\left(it+t^2\right) = \ln\mathbf{E}\, e^{it\ln Z_0(y)}\end{aligned}$$

as $\gamma \to 0$ and so, in the Gaussian case Lemma 2 is proved.

In the general case, proceeding similarly we get

$$\begin{aligned}\varphi_\gamma(t) &= \mathbf{E}\, e^{it\ln X_{\gamma,f}(y)} = \mathbf{E}\, e^{it\sum_{k=1}^{\Pi_+\left(y/I\gamma^2\right)}\ln\frac{f(\varepsilon_k^+ +\gamma)}{f(\varepsilon_k^+)}}\\ &= \mathbf{E}\left(\left(\mathbf{E}\, e^{it\ln\frac{f(\varepsilon+\gamma)}{f(\varepsilon)}}\right)^{\Pi_+(y/I\gamma^2)}\right) \to e^{i\left(-\frac{I}{2}\right)t-\frac{1}{2}yt^2} = \mathbf{E}\, e^{it\ln Z_0(y)}\end{aligned}$$

by dominated convergence theorem, since

$$\left(\mathbf{E}\, e^{it\ln\frac{f(\varepsilon+\gamma)}{f(\varepsilon)}}\right)^{1/\gamma^2} \to e^{i\left(-\frac{I}{2}\right)t-\frac{1}{2}It^2}$$

by (5), and $\gamma^2\,\Pi_+(y/I\gamma^2)$ converges clearly to y/I in L^2 (and hence in probability).

3.2 Proof of Lemma 4

Now we turn to the proof of Lemma 4 (we will prove Lemma 3 just after). For $y > 0$ we have

$$\begin{aligned}\mathbf{E}X_\gamma^{1/2}(y) &= \mathbf{E}\,\mathbf{E}\left(e^{-\frac{\gamma}{2}\sum_{k=1}^{\Pi_+\left(y/\gamma^2\right)}\varepsilon_k^+ - \frac{\gamma^2}{4}\Pi_+(y/\gamma^2)} \,\middle|\, \mathscr{F}_{\Pi_+}\right)\\ &= \mathbf{E}\, e^{-\frac{\gamma^2}{4}\Pi_+(y/\gamma^2)+\frac{\gamma^2}{8}\Pi_+(y/\gamma^2)} = \mathbf{E}\, e^{-\frac{\gamma^2}{8}\Pi_+(y/\gamma^2)}\\ &= \exp\left(\frac{y}{\gamma^2}\left(e^{-\frac{\gamma^2}{8}}-1\right)\right).\end{aligned}$$

The process X_γ being symmetric, we have

$$\mathbf{E}X_\gamma^{1/2}(y) = \exp\left(\frac{|y|}{\gamma^2}\left(e^{-\frac{\gamma^2}{8}}-1\right)\right) \tag{8}$$

for all $y \in \mathbb{R}$ and, since

$$\frac{1}{\gamma^2}\left(e^{-\frac{\gamma^2}{8}}-1\right) = \frac{1}{\gamma^2}\left(-\frac{\gamma^2}{8}+o(\gamma^2)\right) \to -\frac{1}{8}$$

as $\gamma \to 0$, for any $c \in \left]0\,,\,1/8\right[$ we have $\mathbf{E}X_\gamma^{1/2}(y) \le \exp\left(-c\,|y|\right)$ for all sufficiently small γ and all $y \in \mathbb{R}$. So, in the Gaussian case Lemma 4 is proved.

In the general case, equality (8) becomes $\mathbf{E}X_{\gamma,f}^{1/2}(y) = \exp\left(|y|\,(I_\gamma - 1)/I\gamma^2\right)$ with

$$I_\gamma = \mathbf{E}\sqrt{\frac{f(\varepsilon+\gamma)}{f(\varepsilon)}} \le \sqrt{\mathbf{E}\,\frac{f(\varepsilon+\gamma)}{f(\varepsilon)}} = 1.$$

Recall the convergence (5) of characteristic functions and note that I_γ^{1/γ^2} are the corresponding moment generating functions at point $1/2$. The convergence of these moment generating functions (at any point smaller than 1) follows from the fact that for all γ they are equal 1 at point 1 (which provides uniform integrability). Thus we have $I_\gamma^{1/\gamma^2} \to e^{-I/8}$, which implies $(\ln I_\gamma)/\gamma^2 \to -I/8$, and so $(I_\gamma - 1)/I\gamma^2 \to -1/8$.

3.3 Proof of Lemma 3

First we consider the case $y_1, y_2 \in \mathbb{R}_+$ (say $y_1 \ge y_2$). Using (8) and taking into account the stationarity and the independence of the increments of the process $\ln X_\gamma$ on $\mathbb{R}_+$, we can write

$$\begin{aligned}
\mathbf{E}\left|X_\gamma^{1/2}(y_1) - X_\gamma^{1/2}(y_2)\right|^2 &= \mathbf{E}X_\gamma(y_1) + \mathbf{E}X_\gamma(y_2) - 2\,\mathbf{E}X_\gamma^{1/2}(y_1)X_\gamma^{1/2}(y_2)\\
&= 2 - 2\,\mathbf{E}X_\gamma(y_2)\,\mathbf{E}\,\frac{X_\gamma^{1/2}(y_1)}{X_\gamma^{1/2}(y_2)}\\
&= 2 - 2\,\mathbf{E}X_\gamma^{1/2}\left(|y_1 - y_2|\right)\\
&= 2 - 2\exp\left(\frac{|y_1 - y_2|}{\gamma^2}\left(e^{-\frac{\gamma^2}{8}} - 1\right)\right)\\
&\le -2\,\frac{|y_1 - y_2|}{\gamma^2}\left(e^{-\frac{\gamma^2}{8}} - 1\right) \le \frac{1}{4}\,|y_1 - y_2|\,.
\end{aligned}$$

The process X_γ being symmetric, we have the same result for the case $y_1, y_2 \in \mathbb{R}_-$. Finally, if $y_1 y_2 \le 0$ (say $y_2 \le 0 \le y_1$), we have

$$\begin{aligned}
\mathbf{E}\left|X_\gamma^{1/2}(y_1) - X_\gamma^{1/2}(y_2)\right|^2 &= 2 - 2\,\mathbf{E}X_\gamma^{1/2}(y_1)\,\mathbf{E}X_\gamma^{1/2}(y_2)\\
&= 2 - 2\exp\left(\frac{|y_1|}{\gamma^2}\left(e^{-\frac{\gamma^2}{8}} - 1\right) + \frac{|y_2|}{\gamma^2}\left(e^{-\frac{\gamma^2}{8}} - 1\right)\right)\\
&= 2 - 2\exp\left(\frac{|y_1 - y_2|}{\gamma^2}\left(e^{-\frac{\gamma^2}{8}} - 1\right)\right)\\
&\le \frac{1}{4}\,|y_1 - y_2|\,,
\end{aligned}$$

and so, in the Gaussian case we obtain even more than the assertion of Lemma 3.

In the general case, proceeding similarly we get

$$\mathbf{E}\left|X_{\gamma,f}^{1/2}(y_1) - X_{\gamma,f}^{1/2}(y_2)\right|^2 \le -2\,\frac{|y_1 - y_2|}{I\gamma^2}(I_\gamma - 1)$$

and, since $-2(I_\gamma - 1)/I\gamma^2 \to 1/4$, the proof is concluded.

3.4 Proof of Lemma 5

First let $y_1, y_2 \in \mathbb{R}_+$ (say $y_1 \geq y_2$) such that $\Delta = |y_1 - y_2| < h$. Then, noting that conditionally to $\mathscr{F}_{\Pi_+}$ the random variable

$$\ln X_\gamma(\Delta) = -\gamma \sum_{k=1}^{\Pi_+(\Delta/\gamma^2)} \varepsilon_k^+ - \frac{\gamma^2}{2}\Pi_+(\Delta/\gamma^2)$$

is Gaussian with mean $-\frac{\gamma^2}{2}\Pi_+(\Delta/\gamma^2)$ and variance $\gamma^2\Pi_+(\Delta/\gamma^2)$, we get

$$\begin{aligned}
\mathbf{P}\left\{\left|\ln X_\gamma(y_1) - \ln X_\gamma(y_2)\right| > \delta\right\} &\leq \frac{1}{\delta^2}\,\mathbf{E}\left|\ln X_\gamma(y_1) - \ln X_\gamma(y_2)\right|^2 \\
&= \frac{1}{\delta^2}\,\mathbf{E}\left|\ln X_\gamma(\Delta)\right|^2 \\
&= \frac{1}{\delta^2}\,\mathbf{E}\,\mathbf{E}\left(\left(\ln X_\gamma(\Delta)\right)^2 \,\middle|\, \mathscr{F}_{\Pi_+}\right) \\
&= \frac{1}{\delta^2}\,\mathbf{E}\left(\gamma^2\Pi_+(\Delta/\gamma^2) + \frac{\gamma^4}{4}\left(\Pi_+(\Delta/\gamma^2)\right)^2\right) \\
&= \frac{1}{\delta^2}\left(\Delta + \frac{\gamma^4}{4}\left(\frac{\Delta}{\gamma^2} + \frac{\Delta^2}{\gamma^4}\right)\right) \\
&= \frac{1}{\delta^2}\left((1+\gamma^2/4)\Delta + \Delta^2/4\right) \\
&< \frac{1}{\delta^2}\left(\beta(\gamma)\,h + h^2/4\right)
\end{aligned}$$

where $\beta(\gamma) = 1 + \gamma^2/4 \to 1$ as $\gamma \to 0$. So, we have

$$\begin{aligned}
\lim_{\gamma\to 0}\sup_{|y_1-y_2|<h} \mathbf{P}\left\{\left|\ln X_\gamma(y_1) - \ln X_\gamma(y_2)\right| > \delta\right\} &\leq \lim_{\gamma\to 0}\frac{1}{\delta^2}\left(\beta(\gamma)\,h + h^2/4\right) \\
&= \frac{1}{\delta^2}\left(h + \frac{h^2}{4}\right),
\end{aligned}$$

and hence

$$\lim_{h\to 0}\lim_{\gamma\to 0}\sup_{|y_1-y_2|<h} \mathbf{P}\left\{\left|\ln X_\gamma(y_1) - \ln X_\gamma(y_2)\right| > \delta\right\} = 0,$$

where the supremum is taken only over $y_1, y_2 \in \mathbb{R}_+$.

The process X_γ being symmetric, we have the same conclusion with the supremum taken over $y_1, y_2 \in \mathbb{R}_-$.

Finally, for $y_1 y_2 \leq 0$ (say $y_2 \leq 0 \leq y_1$) such that $|y_1 - y_2| < h$, using the elementary inequality $(a-b)^2 \leq 2(a^2+b^2)$ we get

$$\begin{aligned}
\mathbf{P}\left\{\left|\ln X_\gamma(y_1) - \ln X_\gamma(y_2)\right| > \delta\right\} &\leq \frac{1}{\delta^2}\,\mathbf{E}\left|\ln X_\gamma(y_1) - \ln X_\gamma(y_2)\right|^2 \\
&\leq \frac{2}{\delta^2}\left(\mathbf{E}\left|\ln X_\gamma(y_1)\right|^2 + \mathbf{E}\left|\ln X_\gamma\left(|y_2|\right)\right|^2\right) \\
&= \frac{2}{\delta^2}\left(\beta(\gamma)y_1 + y_1^2/4 + \beta(\gamma)\,|y_2| + |y_2|^2/4\right) \\
&< \frac{2}{\delta^2}\left(\beta(\gamma)h + h^2/4\right),
\end{aligned}$$

which again yields the desired conclusion. So, in the Gaussian case Lemma 5 is proved.

Another way to prove this lemma, is to notice first that the weak convergence of $\ln X_\gamma(y)$ to $\ln Z_0(y)$ (established in Lemma 2) is uniform with respect to $y \in K$ for any compact $K \subset \mathbb{R}$. Indeed, the uniformity of the convergence of the characteristic functions in the proof of Lemma 2 is obvious, and so one can apply, for example, Theorem 7 from Appendix I of Ibragimov and Khasminskii (1981), whose remaining conditions are easily checked.

Second, using this uniformity we obtain

$$\lim_{\gamma\to 0} \sup_{|y_1-y_2|<h} \mathbf{P}\left\{\left|\ln X_\gamma(y_1) - \ln X_\gamma(y_2)\right| > \delta\right\} = \lim_{\gamma\to 0} \sup_{|y|<h} \mathbf{P}\left\{\left|\ln X_\gamma(y)\right| > \delta\right\}$$
$$= \sup_{|y|<h} \mathbf{P}\left\{|\ln Z_0(y)| > \delta\right\}$$

where the supremum is taken over $y_1, y_2 \in \mathbb{R}$ such that $y_1 y_2 \geq 0$, and

$$\lim_{\gamma\to 0} \sup_{|y_1-y_2|<h} \mathbf{P}\left\{\left|\ln X_\gamma(y_1) - \ln X_\gamma(y_2)\right| > \delta\right\} \leq 2 \sup_{|y|<h} \mathbf{P}\left\{|\ln Z_0(y)| > \frac{\delta}{2}\right\}$$

where the supremum is taken over $y_1, y_2 \in \mathbb{R}$ such that $y_1 y_2 \leq 0$.

Finally, reminding that $\ln Z_0(y) \sim \mathcal{N}(-|y|/2, |y|)$ and denoting Φ the distribution function of the standard Gaussian law, we get

$$\mathbf{P}\left\{|\ln Z_0(y)| > \delta\right\} = \Phi\left(-\frac{\delta}{\sqrt{|y|}} + \frac{\sqrt{|y|}}{2}\right) + 1 - \Phi\left(\frac{\delta}{\sqrt{|y|}} + \frac{\sqrt{|y|}}{2}\right)$$
$$\leq \Phi\left(-\frac{\delta}{\sqrt{h}} + \frac{\sqrt{h}}{2}\right) + 1 - \Phi\left(\frac{\delta}{\sqrt{h}}\right)$$

for $|y| < h$. The last expression does not depend on y and clearly converges to 0 as $h \to 0$, so the assertion of the lemma follows.

It remains to observe that this second proof does not use any particularity of the process X_γ and, hence, is trivially extendable to the general case.

3.5 Proof of Lemma 6

Taking into account the symmetry of the process $\ln X_\gamma$, as well as the stationarity and the independence of its increments on $\mathbb{R}_+$, we obtain

$$\mathbf{P}\left\{\sup_{|y|>A} X_\gamma(y) > e^{-bA}\right\} \leq 2\,\mathbf{P}\left\{\sup_{y>A} X_\gamma(y) > e^{-bA}\right\}$$
$$\leq 2\,e^{bA/2}\,\mathbf{E}\sup_{y>A} X_\gamma^{1/2}(y) \tag{9}$$
$$= 2\,e^{bA/2}\,\mathbf{E}X_\gamma^{1/2}(A)\,\mathbf{E}\sup_{y>A}\frac{X_\gamma^{1/2}(y)}{X_\gamma^{1/2}(A)}$$
$$= 2\,e^{bA/2}\,\mathbf{E}X_\gamma^{1/2}(A)\,\mathbf{E}\sup_{z>0} X_\gamma^{1/2}(z).$$

In order to estimate the last factor we write

$$\mathbf{E}\sup_{z>0} X_\gamma^{1/2}(z) = \mathbf{E}\exp\left(\frac{1}{2}\sup_{z>0}\left(-\gamma\sum_{k=1}^{\Pi_+(z/\gamma^2)}\varepsilon_k^+ - \frac{\gamma^2}{2}\Pi_+(z/\gamma^2)\right)\right)$$
$$= \mathbf{E}\exp\left(\frac{1}{2}\sup_{n\in\mathbb{N}}\left(-\gamma\sum_{k=1}^{n}\varepsilon_k^+ - \frac{n\gamma^2}{2}\right)\right).$$

Now, let us observe that the random walk $S_n = -\sum_{k=1}^n \varepsilon_k^+$, $n \in \mathbb{N}$, has the same law as the restriction on $\mathbb{N}$ of a standard Brownian motion W. So,

$$\mathbf{E}\sup_{z>0} X_\gamma^{1/2}(z) = \mathbf{E}\exp\left(\frac{1}{2}\sup_{n\in\mathbb{N}}\left(\gamma W(n) - n\gamma^2/2\right)\right)$$
$$= \mathbf{E}\exp\left(\frac{1}{2}\sup_{n\in\mathbb{N}}\left(W(n\gamma^2) - n\gamma^2/2\right)\right)$$
$$\le \mathbf{E}\exp\left(\frac{1}{2}\sup_{t>0}\left(W(t) - t/2\right)\right) = \mathbf{E}\exp\left(\frac{1}{2}M\right)$$

with an evident notation. It is known that the random variable M is exponential of parameter 1 (see, for example, Borodin and Salminen (2002)) and hence, using its moment generating function $\mathbf{E}\,e^{tM} = (1-t)^{-1}$, we get

$$\mathbf{E}\sup_{z>0} X_\gamma^{1/2}(z) \le 2. \tag{10}$$

Finally, taking $b \in \,]0,\,1/12[$ we have $3b/2 \in \,]0,\,1/8[$ and, combining (9), (10) and using Lemma 4, we finally obtain

$$\mathbf{P}\left\{\sup_{|y|>A} X_\gamma(y) > e^{-bA}\right\} \le 4\,e^{bA/2}\,\exp\left(-\tfrac{3b}{2}A\right) = 4\,e^{-bA}$$

for all sufficiently small γ and all $A > 0$, which concludes the proof in the Gaussian case.

In the general case the proof is almost the same. Note that we have no longer the symmetry of the process $X_{\gamma,f}$, so we need to consider the cases $y > A$ and $y < -A$ separately. Besides that, the only difference is in the derivation of the bound (10). Here we get

$$\mathbf{E}\sup_{z>0} X_{\gamma,f}^{1/2}(z) = \mathbf{E}\exp\left(\frac{1}{2}M\right),$$

where M is the supremum of the random walk $S_n = \sum_{k=1}^n X_k$, $n \in \mathbb{N}$, with $X_k = \ln\frac{f(\varepsilon_k^+ + \gamma)}{f(\varepsilon_k^+)}$. Note that

$$\mathbf{E}\,e^{X_1} = \mathbf{E}\,\frac{f(\varepsilon+\gamma)}{f(\varepsilon)} = 1,$$

and so, the cummulant generating function $k(t) = \ln(\mathbf{E}\,e^{tX_1})$ of X_1 admits a strictly positive zero $t_0 = 1$. Hence, by the well-known Cramér-Lundberg bound on the tail probabilities of M (see, for example, Theorem 5.1 from Chapter XIII of Asmussen (2003)), we have

$$\mathbf{P}(M > x) \le e^{-t_0 x} = e^{-x}$$

for all $x > 0$. Finally, denoting F the distribution function of M and using this bound we obtain

$$\begin{aligned}\mathbf{E}\exp\left(\frac{1}{2}M\right) &= \int_{\mathbb{R}} e^{x/2}\,dF(x)\\ &= \left[e^{x/2}\,(F(x)-1)\right]_{-\infty}^{+\infty} - \frac{1}{2}\int_{\mathbb{R}} e^{x/2}\,(F(x)-1)\;dx\\ &= \frac{1}{2}\int_{\mathbb{R}_-} e^{x/2}\,dx + \frac{1}{2}\int_{\mathbb{R}_+} e^{x/2}\,(1-F(x))\;dx\\ &\le 1 + \frac{1}{2}\int_{\mathbb{R}_+} e^{-x/2}\,dx = 2,\end{aligned}$$

which concludes the proof.

Acknowledgements We would like to thank the anonymous referees for their useful suggestions and comments.

References

Asmussen S (2003) Applied probability and queues, 2nd ed., Applications of Mathematics 51. Springer, New York

Bhattacharya PK, Brockwell PJ (1976) The minimum of an additive process with applications to signal estimation and storage theory. Z Wahrsch verw Geb 37(1):51–75

Borodin AN, Salminen P (2002) Handbook of Brownian motion—facts and formulae. Probability and its applications. Birkhäuser, Basel

Chan KS (1993) Consistency and limiting distribution of the least squares estimator of a threshold autoregressive model. Ann Stat 21(1):520–533

Chan NH, Kutoyants Yu A (submitted) On parameter estimation of threshold autoregressive models. http://arxiv.org/abs/1003.3800

Chan NH, Kutoyants YuA (to appear) Recent developments of threshold estimation for nonlinear time series. J Jpn Stat Soc http://www.univ-lemans.fr/sciences/statist/download/Kutoyants/PrepCh_Kut2.pdf

Ciuperca G (2004) Maximum likelihood estimator in a two-phase nonlinear random regression model. Stat Decis 22(4):335–349

Dachian S (2010) On limiting likelihood ratio processes of some change-point type statistical models. J Stat Plann Inference 140(9):2682–2692

Dachian S, Negri I (2009) On compound Poisson type limiting likelihood ratio process arising in some change-point models, WP n.11/MS, Dep. IIMM, 2009. http://hdl.handle.net/10446/529

Dachian S, Negri I (2010) On Gaussian compound poisson type limiting likelihood ratio process. In: Proceedings of the 45th scientific meeting of the italian statistical society, 2010. http://homes.stat.unipd.it/mgri/SIS2010/Program/contributedpaper/737-1258-1-DR.pdf

Fujii T (2007) A note on the asymptotic distribution of the maximum likelihood estimator in a non-regular case. Stat Probab Lett 77(16):1622–1627

Gihman II, Skorohod AV (1974) The theory of stochastic processes I. Springer, New York

Golubev GK (1979) Computation of the efficiency of the maximum-likelihood estimator when observing a discontinuous signal in white noise. Probl Inf Transm 15(3):61–69

Höpfner R, Kutoyants YuA (2010) Estimating discontinuous periodic signals in a time inhomogeneous diffusion. Stat Inference Stoch Process 13(3):193–230

Ibragimov IA, Khasminskii RZ (1975) Estimation of a parameter of a discontinuous signal in a white Gaussian noise. Probl Inf Transm 11(3):31–43

Ibragimov IA, Khasminskii RZ (1981) Statistical estimation asymptotic theory. Springer, New York

Koul HL, Qian L (2002) Asymptotics of maximum likelihood estimator in a two-phase linear regression model. J Stat Plann Inference 108(1–2):99–119

Küchler U, Kutoyants YuA (2000) Delay estimation for some stationary diffusion-type processes. Scand J Stat 27(3):405–414

Kutoyants YuA (1984) Parameter estimation for stochastic processes. Armenian Academy of Sciences, Yerevan, 1980 (in Russian), translation of revised version. Heldermann, Berlin

Kutoyants YuA (1994) Identification of dynamical systems with small noise, Mathematics and its Applications 300. Kluwer, Dordrecht

Kutoyants YuA (2004) Statistical inference for ergodic diffusion processes. Springer Series in Statistics. Springer, London

Kutoyants YuA (to appear) On identification of the threshold diffusion processes. Ann Inst Stat Math http://arxiv.org/abs/1003.3539

Rubin H, Song K-S (1995) Exact computation of the asymptotic efficiency of maximum likelihood estimators of a discontinuous signal in a Gaussian white noise. Ann Stat 23(3):732–739

Terent'yev AS (1968) Probability distribution of a time location of an absolute maximum at the output of a synchronized filter. Radioeng Electron 13(4):652–657

Yao Y-C (1987) Approximating the distribution of the maximum likelihood estimate of the change-point in a sequence of independent random variables. Ann Stat 15(3):1321–1328

Deuxième Partie

Théorie des Champs Aléatoires

INCLUSION–EXCLUSION DESCRIPTION OF RANDOM FIELDS

S. Yu. Dashian, B. S. Nahapetian

**Izvestiya Natsionalnoi Akademii Nauk Armenii. Matematika,
Vol. 30, No. 6, 1995**

The inclusion-exclusion approach towards construction of random fields on the ν-dimensional integer lattice is described. Comparison with classical Gibbs description is presented.

§1. INTRODUCTION

The inclusion-exclusion approach was successfully applied in the framework of point process theory by Ambartzumian and Sukiasian [1]. The main result of this approach is the following theorem.

THEOREM A. *(R. V. Ambartzumian, H. S. Sukiasian) Let a system* $\{f(x_1, ..., x_n)\}$, $x_i \in \mathbb{R}^d$ *of nonnegative symmetrical functions be given satisfying the condition* $f(x_1, ..., x_n) < b^n$, $n = 1, 2, ...$ *for some* $b > 0$. *If for almost all* $x_1, ..., x_n \in \mathbb{R}^d$ *and all convex* $D \subset \mathbb{R}^d$ *the following inequalities hold*

$$1 + \sum_{n=1}^{\infty} \frac{(-1)^n}{n!} \int_D \cdots \int_D f(y_1, ..., y_n)\, dy_1 \cdots dy_n \geq 0,$$

$$f(x_1, ..., x_m) + \sum_{n=1}^{\infty} \frac{(-1)^n}{n!} \int_D \cdots \int_D f(x_1, ..., x_m, y_1, ..., y_n)\, dy_1 \cdots dy_n \geq 0, \quad m > 0,$$

then there exists a point process P, *such that at the continuity points, the values of* f *coincide with the densities of* P.

The purpose of this paper is to apply the same approach towards construction of random fields on the integer lattice $\mathbb{Z}^\nu$. Special attention is paid to the classical Gibbs random fields. The paper describes the main facts of the proposed approach, gives some examples and points at a broad class of non Gibbsian random fields. We note

that non Gibbsian random fields in statistical physics now receive intensive consideration (see, for example, [2] – [5]).

§2. RANDOM FIELDS AND P–FUNCTIONS

We consider random fields on the integer ν-dimensional lattice $\mathbf{Z}^\nu$, $\nu \geq 1$. For simplicity the phase space X we assume to consist of two points: $X = \{0, 1\}$. Denote by $\mathcal{E}$ the set of all finite subsets of $\mathbf{Z}^\nu$ and let

$$X^\Lambda = \{x_t,\ t \in \Lambda\}, \quad x_t \in X, \quad t \in \Lambda, \quad \Lambda \in \mathcal{E}$$

be the set of all configurations (realizations) on Λ. Each element $x \in X^\Lambda$ is uniquely determined by the subset of Λ, where the configuration x assumes the value 1 (in physical terminology this is the subset occupied by the particles). Therefore any configuration on Λ we will identify with corresponding subset of Λ. A probability distribution on X^Λ we denote by $P_\Lambda = \{P_\Lambda(x),\ x \subseteq \Lambda\}$, $\Lambda \in \mathcal{E}$. For $\Lambda = \emptyset$ there exists only one probability distribution $P_\emptyset(\emptyset) = 1$. For $\Lambda \in \mathcal{E}$ and $I \subseteq \Lambda$ denote

$$(P_\Lambda)_I(x) = \sum_{J \subseteq \Lambda \setminus I} P_\Lambda(x \cup J), \quad x \subseteq I.$$

Definition 1. A set of probability distributions $P = \{P_\Lambda,\ \Lambda \in \mathcal{E}\}$ is called *consistent in Kolmogorov sense*, if for any $\Lambda \in \mathcal{E}$ and $I \subseteq \Lambda$ $(P_\Lambda)_I(x) = P_I(x)$, $x \subseteq I$.

It is well known that any set of probability distributions consistent in Kolmogorov sense determines some probability measure on $X^{\mathbf{Z}^\nu}$ equivalently some random field. In the inclusion-exclusion approach the Kolmogorov's consistency condition is replaced by some nonnegativity condition imposed on certain finite sums with alternating signs of summands.

Let $\mathcal{B}$ be the Banach space of all bounded functions defined on $\mathcal{E}$ with the norm

$$||b|| = \sup_{\Lambda \in \mathcal{E}} \frac{1}{2n(\Lambda)} \sum_{J \subset \Lambda} |b_J|, \quad b = \{b_j,\ J \in \mathcal{E}\} \in \mathcal{B},$$

where $n(\Lambda)$ is some numeration of elements from $\mathcal{E}$.

Definition 2. A function $f = \{f_J,\ J \in \mathcal{E}\}$, $f \in \mathcal{B}$ we call *a P-function*, if $f_\emptyset = 1$ and for any $\Lambda \in \mathcal{E}$ and $x \subseteq \Lambda$

$$\sum_{J \subseteq x} (-1)^{|x \setminus J|}\ f_{\Lambda \setminus J} \geq 0, \tag{1}$$

where $|\cdot|$ denotes the number of points in a finite set.

Note that the space $\mathcal{B}^P \subset \mathcal{B}$ of all P-function is a convex closed subset of $\mathcal{B}$. It is not difficult to show that $\mathcal{B}^P$ is compact.

THEOREM 1. *A system $P = \{P_\Lambda, \Lambda \in \mathcal{E}\}$ of probability distributions is consistent in Kolmogorov sense if and only if there exists a P-function f such that*

$$P_\Lambda(x) = \sum_{J \subseteq x} (-1)^{|x \setminus J|} f_{\Lambda \setminus J}, \quad P_\Lambda(\emptyset) = f_\Lambda, \quad x \subseteq \Lambda, \quad \Lambda \in \mathcal{E}. \tag{2}$$

PROOF: *Necessity.* Let $P = \{P_\Lambda, \Lambda \in \mathcal{E}\}$ be a system of probability distributions consistent in Kolmogorov sense. Put $f_\Lambda = P_\Lambda(\emptyset)$, $\Lambda \in \mathcal{E}$. Clearly $0 \le f_\Lambda \le 1$, $f_\emptyset = P_\emptyset(\emptyset) = 1$. Further we have

$$\sum_{J \subseteq x} (-1)^{|x \setminus J|} f_{\Lambda \setminus J} = \sum_{J \subseteq x} (-1)^{|x \setminus J|} P_{\Lambda \setminus J}(\emptyset) = \sum_{J \subseteq x} (-1)^{|x \setminus J|} (P_\Lambda)_{\Lambda \setminus J}(\emptyset) =$$
$$= \sum_{J \subseteq x} (-1)^{|x \setminus J|} \sum_{\widetilde{J}: \widetilde{J} \subseteq J} P_\Lambda(\widetilde{J}) = \sum_{\widetilde{J} \subseteq x} P_\Lambda(\widetilde{J}) \sum_{J: \widetilde{J} \subseteq J \subseteq x} (-1)^{|x \setminus J|} = P_\Lambda(x).$$

At the last step we used the relation

$$\sum_{A: B \subset A \subset C} (-1)^{|C \setminus A|} = \begin{cases} 1, & B = C, \\ 0, & B \ne C. \end{cases} \tag{3}$$

Sufficiency. Let f be a P-function. We put

$$P_\Lambda(x) = \sum_{J \subseteq x} (-1)^{|x \setminus J|} f_{\Lambda \setminus J}, \quad x \subseteq \Lambda, \quad \Lambda \in \mathcal{E}$$

and show that $P = \{P_\Lambda, \Lambda \in \mathcal{E}\}$ is a family of probability distributions consistent in Kolmogorov sense. We have

$$\sum_{x \subseteq \Lambda} P_\Lambda(x) = \sum_{J \subseteq x} (-1)^{|x \setminus J|} f_{\Lambda \setminus J} = \sum_{J \subseteq \Lambda} f_{\Lambda \setminus J} \sum_{x: J \subseteq x \subseteq \Lambda} (-1)^{|x \setminus J|} = f_\emptyset = 1,$$

i.e. P is a system of probability distributions. Let us verify that it is consistent. For any $\Lambda \in \mathcal{E}$ and $I \subseteq \Lambda$ we can write

$$(P_\Lambda)_I(x) = \sum_{J \subseteq \Lambda \setminus I} \sum_{\widetilde{J} \subseteq x \cup J} (-1)^{|x \cup J \setminus \widetilde{J}|} f_{\Lambda \setminus \widetilde{J}} = \sum_{J \subseteq \Lambda \setminus I} \sum_{\widetilde{J}_1 \subseteq x} (-1)^{|x \setminus \widetilde{J}_1|} \sum_{\widetilde{J}_2 \subseteq J} (-1)^{|J \setminus \widetilde{J}_2|} f_{\Lambda \setminus (\widetilde{J}_1 \cup \widetilde{J}_2)} =$$
$$= \sum_{\widetilde{J}_1 \subseteq x} (-1)^{|x \setminus \widetilde{J}_1|} \sum_{\widetilde{J}_2 \subseteq \Lambda \setminus I} f_{\Lambda \setminus (\widetilde{J}_1 \cup \widetilde{J}_2)} \sum_{J: \widetilde{J}_2 \subseteq J \subseteq \Lambda \setminus I} (-1)^{|J|} = \sum_{\widetilde{J}_1 \subseteq x} (-1)^{|x \setminus \widetilde{J}_1|} f_{I \setminus J_1} = P_I(x).$$

§3. EXAMPLES OF P-FUNCTIONS

Example 1. Let f be a P-function. For any $B \in \mathcal{E}$ such that $f_B > 0$ consider the function $f^{(B)} = \{f_{B \cup J}(f_B)^{-1}, J \in \mathcal{E}\}$. It is not difficult to see that $f^{(B)}$ again is a P-function. The realizations of corresponding random fields may assume the value 1 only outside B.

Example 2. Let P be a random field with independent components and $f_t = P_{\{t\}}(\emptyset)$, $t \in \mathbf{Z}^\nu$. The corresponding P-function is

$$f = \left\{ \prod_{t \in J} f_t, \ J \in \mathcal{E} \right\}.$$

The case $f_t \equiv q$, $t \in \mathbf{Z}^\nu$, $0 \le q \le 1$ corresponds to Bernoulli random field.

Example 3. Suppose $p(x)$, $x \in [0,1]$ is a probability density and $f = \{q^{|J|}, J \in \mathcal{E}\}$ is a Bernoulli P-function. The function

$$b = \left\{ \int_0^1 q^{|J|} p(q)\, dq, \quad J \in \mathcal{E} \right\}$$

is a P-function of the corresponding mixture of the Bernoulli random fields. In case $p(x) = \tau x^{\tau-1}$, $\tau > 0$ the corresponding P-function is

$$b = \left\{ \tau \int_0^1 q^{|\Lambda|+\tau-1}\, dq = \frac{\tau}{|\Lambda|+\tau}, \quad \Lambda \in \mathcal{E} \right\}.$$

In this case the set of finite dimensional distributions is

$$P_\Lambda(x) = \frac{\tau}{|\Lambda|+\tau} \prod_{i=1}^{|\Lambda|} \frac{\tau}{|\Lambda|+\tau-i}, \quad \Lambda \in \mathcal{E}.$$

In §7 we will demonstrate that this random field is non Gibbsian.

Example 4. Suppose $\beta_{t,s}$, $t, s \in \mathbf{Z}^\nu$ is a family of nonnegative numbers such that $\sum\limits_{t \in \mathbf{Z}^\nu} \beta_{t,s} < \infty$. Consider

$$f = \left\{ \exp\left[- \sum_{t,s \in J} \beta_{t,s} \right], \quad J \in \mathcal{E} \right\}.$$

It is a P-function, which is the discrete analog of Ambartzumian–Sukiasian [1] point random field in $\mathbf{R}^d$. We briefly remind the main result of [1]. The following sequence of functions was considered:

$$f(x_1) \equiv \alpha, \quad f(x_1, ..., x_n) = \alpha^n \prod_{\{i,j\} \subset \{1,...,n\}} h(x_i, x_j), \quad n = 2, 3, ..., \tag{4}$$

where $0 \le h(x,y) \le 1$ is a symmetrical function in $\mathbf{R}^d \times \mathbf{R}^d$, $\alpha > 0$ is a parameter (intensity), while the product is taken over all possible two-subsets of $\{1, ..., n\}$. Under the convergence condition

$$\sup_x \int_{\mathbf{R}^d} [1 - h(x,y)]\, dy < \infty$$

there exists a point process P in $\mathbb{R}^d$ for which (4) present the so-called *absolutely densities*, i.e. for every sequence $x_1, ..., x_n$

$$P(dx_1, ..., dx_n) = f(x_1, ..., x_n)\, dx_1 \cdots dx_n.$$

The proof of this result was obtained in [1] using Theorem A. It seems possible to prove the corresponding result for P-functions.

Example 5. Here we describe some P-functions occurring in the Gibbs random fields theory. For any nonempty $A \in \mathcal{E}$ let us fix an arbitrary point $t_A \in A$ and define a partially ordering in $\mathcal{E}$. Assume that $B < A$, $B, A \in \mathcal{E}$ if there exists a sequence $B = B_1, B_2, ..., B_n = A$ such that $B_{i-1} = B_i \setminus t_{B_i}$, $i = 2, ..., n$. A sequence

$$\beta = \{B_1, T_1; ...; B_n, T_n\}, \quad B_i, T_i \in \mathcal{E}, \quad T_i \cap B_i = t_{B_i}, \quad i = 1, ..., n, \quad B_1 \le A, \quad B_i \le B_{i-1} \cup T_{i-1}, \quad i = 2, ..., n$$

we call *a path beginning at A and of length n*. The set of all paths the beginning at A and of length n we denote by $B_A^{(n)}$, $A \in \mathcal{E}$. Now let $K = (K_J,\, J \in \mathcal{E})$ be a function such that

$$\sum_{\substack{J: t \in J \in \mathcal{E} \\ |J| = n}} |K_J| < \alpha \lambda^n, \quad \lambda(1 + \sqrt{\alpha})^2 < 1, \quad \lambda, \alpha > 0.$$

Then the function

$$f_J = 1 + \sum_n \sum_{\beta \in B_J^{(n)}} (-1)^n K_{T_1} \cdots K_{T_n}, \quad J \in \mathcal{E}$$

presents a P-function (see, for example, [9], [10]).

§4. RANDOM FIELDS AND Q-FUNCTIONS

In this section we construct P-functions by the principle used in the theory of Gibbs random fields.

Definition 3. A function $\theta = \{\theta_J,\, J \in \mathcal{E}\}$ we call *a Q-function*, if $\theta_J > 0$, $J \in \mathcal{E}$, $\theta_\emptyset = 1$ and for any $x \in \mathcal{E}$

$$\sum_{J \subseteq x} (-1)^{|x \setminus J|}\, \theta_J \ge 0. \tag{5}$$

Unlike the P-functions, the Q-functions have simple constructive description.

THEOREM 2. *A function $\theta = \{\theta_J,\, J \in \mathcal{E}\}$ is a Q-function if and only if there exists a function $H = \{H_S,\, S \in \mathcal{E}\}$, $H_S \ge 0$, $S \in \mathcal{E}$, $H_\emptyset = 1$ such that for each $J \in \mathcal{E}$, $\theta_J = \sum\limits_{S \subseteq J} H_S$.*

PROOF: *Necessity.* Let $\theta = \{\theta_J,\, J \in \mathcal{E}\}$ be a Q-function. Put

$$H_S = \sum_{J \subseteq S} (-1)^{|S \setminus J|}\, \theta_J, \quad S \in \mathcal{E}. \tag{6}$$

Since θ is a Q-function, according to definition of H_S, we have $H_S \geq 0$, $S \in \mathcal{E}$ and $H_\emptyset = 1$. We can write

$$\sum_{S \subseteq \Lambda} H_S = \sum_{S \subseteq \Lambda} \sum_{J \subseteq S} (-1)^{|S \setminus J|}\, \theta_J = \sum_{J \subseteq \Lambda} \theta_J \sum_{S : J \subseteq S \subseteq \Lambda} (-1)^{|S \setminus J|} = \theta_\Lambda, \quad \Lambda \in \mathcal{E}.$$

Sufficiency. Let $\theta_\Lambda = \sum\limits_{S \subseteq \Lambda} H_S$. Clearly $\theta_J > 0$, $J \in \mathcal{E}$, $\theta_\emptyset = 1$. Finally

$$\sum_{J \subseteq S} (-1)^{|S \setminus J|}\, \theta_J = \sum_{J \subseteq S} (-1)^{|S \setminus J|} \sum_{\widetilde{J} \subseteq J} H_{\widetilde{J}} = \sum_{\widetilde{J} \subseteq S} H_{\widetilde{J}} \sum_{J : \widetilde{J} \subseteq J \subseteq S} (-1)^{|S \setminus J|} = H_S > 0, \quad S \in \mathcal{E}.$$

Theorem 2 is proved.

THEOREM 3. *Let $\theta = \{\theta_\Lambda,\, \Lambda \in \mathcal{E}\}$ be a Q-function. Suppose $\Lambda \uparrow \mathbf{Z}^\nu$ is a sequence of increasing subsets, such that for any $J \in \mathcal{E}$ the following limit exists:*

$$\lim_{\Lambda \uparrow \mathbf{Z}^\nu} \frac{\theta_{\Lambda \setminus J}}{\theta_\Lambda} = f_J. \tag{7}$$

Then $f = \{f_J,\, J \in \mathcal{E}\}$ is a P-function.

PROOF: We have for any $I \in \mathcal{E}$

$$\sum_{J \subseteq x} (-1)^{|x \setminus J|}\, f_{I \setminus J} = \lim_{\Lambda \uparrow \mathbf{Z}^\nu} \sum_{J \subseteq x} (-1)^{|x \setminus J|}\, \frac{\theta_{\Lambda \setminus (I \setminus J)}}{\theta_\Lambda} = \lim_{\Lambda \uparrow \mathbf{Z}^\nu} \frac{1}{\theta_\Lambda} \sum_{J \subseteq x} (-1)^{|x \setminus J|}\, \theta_{(\Lambda \setminus I) \cup J}, \quad x \subseteq I.$$

Since θ is a Q-function, for $\Lambda \in \mathcal{E}$, $I \subseteq \Lambda$, $S \subset \Lambda \setminus I$ and $x \subseteq I$ we have $\sum\limits_{J \subseteq x \cup S} (-1)^{|x \cup S \setminus J|}\, \theta_J \geq 0$, and hence

$$D \equiv \sum_{S \subseteq \Lambda \setminus I} \sum_{J \subseteq x \cup S} (-1)^{|x \cup S \setminus J|}\, \theta_J \geq 0.$$

It follows that

$$D = \sum_{S \subseteq \Lambda \setminus I} \sum_{J_1 \subseteq x} (-1)^{|x \setminus J_1|} \sum_{J_2 \subseteq \Lambda \setminus I} \theta_{J_1 \cup J_2} \sum_{J_2 : J_2 \subseteq S \subseteq \Lambda \setminus I} (-1)^{|S \setminus J_2|} = \sum_{J_1 \subseteq x} (-1)^{|x \setminus J_1|}\, \theta_{(\Lambda \setminus I) \cup J_1} \geq 0.$$

Theorem 3 is proved.

We give characterization of Q-systems to be used in §§5,6 below.

Definition 4. A family of probability distributions $Q = \{Q_\Lambda,\ \Lambda \in \mathcal{E}\}$, $Q_\Lambda(\emptyset) > 0$, $Q_\emptyset(\emptyset) = 1$ is called *consistent in Dobrushin sense*, if for any $\widetilde{\Lambda}, \Lambda \in \mathcal{E}$, $\Lambda \cap \widetilde{\Lambda} = \emptyset$

$$Q_{\Lambda\cup\widetilde{\Lambda}}(x) = \frac{Q_{\Lambda\cup\widetilde{\Lambda}}(\emptyset)}{Q_\Lambda(\emptyset)}\, Q_\Lambda(x), \quad x \subseteq \Lambda. \tag{8}$$

Note that one can equivalently rewrite (8) as follows

$$Q_{\Lambda\cup\widetilde{\Lambda}}(x) = Q_\Lambda(x)\big(Q_{\Lambda\cup\widetilde{\Lambda}}\big)_{\widetilde{\Lambda}}(\emptyset). \tag{9}$$

THEOREM 4. *A system $Q = \{Q_\Lambda,\ \Lambda \in \mathcal{E}\}$, $Q_\Lambda(\emptyset) > 0$, $Q_\emptyset(\emptyset) = 1$ of probability distributions is consistent in Dobrushin sense if and only if there exists a Q-function $\theta = \{\theta_J,\ J \in \mathcal{E}\}$ such that for any $\Lambda \in \mathcal{E}$*

$$Q_\Lambda(x) = \frac{1}{\theta_\Lambda} \sum_{J\subseteq x} (-1)^{|x\setminus J|}\, \theta_J, \quad x \subseteq \Lambda. \tag{10}$$

PROOF: *Necessity.* Let $Q = \{Q_\Lambda,\ \Lambda \in \mathcal{E}\}$, $Q_\Lambda(\emptyset) > 0$, $Q_\emptyset(\emptyset) = 1$ be a set of probability distributions consistent in Dobrushin sense. Set $\theta_\Lambda = [Q_\Lambda(\emptyset)]^{-1}$, $\Lambda \in \mathcal{E}$. We can write

$$1 = \sum_{S\subseteq J} Q_J(S) = \sum_{S\subseteq J} \frac{Q_J(\emptyset)}{Q_\Lambda(\emptyset)} Q_\Lambda(S) = \frac{Q_J(\emptyset)}{Q_\Lambda(\emptyset)} \sum_{S\subseteq J} Q_\Lambda(S).$$

From this we get

$$\theta_J = \theta_\Lambda \sum_{S\subseteq J} Q_\Lambda(S).$$

Therefore

$$\sum_{J\subseteq x} (-1)^{|x\setminus J|}\, \theta_J = \theta_\Lambda \sum_{J\subseteq x} (-1)^{|x\setminus J|} \sum_{S\subseteq J} Q_\Lambda(S) = \theta_\Lambda Q_\Lambda(x),$$

and we obtain (10).

Sufficiency. Let there exist a Q-function $\theta = \{\theta_J,\ J \in \mathcal{E}\}$ such that (10) holds. By definition, $Q_\Lambda(x) \geq 0$, $Q_\emptyset(\emptyset) = 1$. Further we have

$$\sum_{x\subseteq\Lambda} Q_\Lambda(x) = \frac{1}{\theta_\Lambda} \sum_{x\subseteq\Lambda} \sum_{J\subseteq x} (-1)^{|x\setminus J|}\, \theta_J = \frac{1}{\theta_\Lambda} \sum_{J\subseteq\Lambda} \theta_J \sum_{x: J\subseteq x\subseteq\Lambda} (-1)^{|x\setminus J|} \theta_J = 1,$$

i.e. the system (10) is a set of probability distributions. Let us verify that it is consistent in Dobrushin sense. First we note that $Q_\Lambda(\emptyset) = \dfrac{1}{\theta_\Lambda}$, $\Lambda \in \mathcal{E}$. Also

$$Q_{\Lambda\cup\widetilde{\Lambda}}(x) = \frac{1}{\theta_{\Lambda\cup\widetilde{\Lambda}}} \sum_{J\subseteq x} (-1)^{|x\setminus J|}\, \theta_J = \frac{\theta_\Lambda}{\theta_{\Lambda\cup\widetilde{\Lambda}}} Q_\Lambda(x) = \frac{Q_{\Lambda\cup\widetilde{\Lambda}}(\emptyset)}{Q_\Lambda(\emptyset)} Q_\Lambda(x), \quad z \subseteq \Lambda.$$

Theorem 4 is proved.

Remark 1. If $\theta = \{\theta_J,\ J \in \mathcal{E}\}$ is a Q-function, then

$$f^{(\Lambda)} = \left\{ f_J = \frac{\theta_{\Lambda \setminus J}}{\theta_\Lambda}, \quad J \in \mathcal{E} \right\}$$

is a P-function with parameter set $\Lambda \in \mathcal{E}$. The probability distributions for the corresponding random fields have the form

$$P_I^{(\Lambda)}(x) = (P_\Lambda)_{\Lambda \cap I}(x), \quad x \subset I,$$

$$(P_\Lambda)_\emptyset(x) = \begin{cases} 1, & x = \emptyset, \\ 0, & x \neq \emptyset. \end{cases}$$

This is a random field *in finite volume* Λ.

Remark 2. Since the set of all P-functions is compact, one can choose a convergent sequence of P-functions $f^{(\Lambda_k)} \to f$ (as $k \to \infty$), where $\Lambda_k \uparrow \mathbf{Z}^\nu$, $k \to \infty$ is some increasing sequence of subsets. A random field with P-function f is called *limiting* for random fields in finite volumes.

§5. Q-FUNCTIONS WITH BOUNDARY CONDITIONS AND CONDITIONAL DISTRIBUTIONS FOR RANDOM FIELDS

Let $P = \{P_\Lambda,\ \Lambda \in \mathcal{E}\}$ be a random field. According to well known martingale convergence theorem for any $\Lambda \in \mathcal{E}$, $x \subset \Lambda$, $\overline{x} \subset \mathbf{Z}^\nu \setminus \Lambda$ there exists the following limit

$$Q_\Lambda^{\overline{x}}(x) = \lim_{\widetilde{\Lambda} \uparrow \mathbf{Z}^\nu \setminus \Lambda} \frac{P_{\Lambda \cup \widetilde{\Lambda}}(x \cup \overline{x}_{\widetilde{\Lambda}})}{P_{\widetilde{\Lambda}}(\overline{x}_{\widetilde{\Lambda}})} \quad \text{a.s.}, \tag{11}$$

where $\overline{x}_{\widetilde{\Lambda}} = \overline{x} \cap \widetilde{\Lambda}$. For each $\Lambda \in \mathcal{E}$ the quantity (11) defines some probability distribution, which we call *conditional distribution on* Λ *with boundary condition* $\overline{x} \subset \mathbf{Z}^\nu \setminus \Lambda$ (see [6], [7]). The family of conditional distributions depending on $\Lambda \in \mathcal{E}$ and the boundary conditions $\overline{x}$:

$$Q = \left\{ Q_\Lambda^{\overline{x}},\ \Lambda \in \mathcal{E},\ \overline{x} \subset \mathbf{Z}^\nu \setminus \Lambda \right\} \tag{11'}$$

is called *conditional distribution* of the random field P.

Definition 5. A system of probability distributions (11') is called *consistent in Dobrushin sense*, if for any $\widetilde{\Lambda}, \Lambda \in \mathcal{E}$, $\Lambda \cap \widetilde{\Lambda} = \emptyset$ and any $x \subset \Lambda$, $y \subset \widetilde{\Lambda}$, $\overline{x} \subset \mathbf{Z}^\nu \setminus (\Lambda \cup \widetilde{\Lambda})$

$$Q_{\Lambda \cup \widetilde{\Lambda}}^{\overline{x}}(x \cup y) = Q_\Lambda^{\overline{x} \cup y}(x) \left(Q_{\Lambda \cup \widetilde{\Lambda}}^{\overline{x}} \right)_{\widetilde{\Lambda}} (y). \tag{12}$$

Definition 6. A function $\theta(\overline{x}) = \{\theta_J^{\overline{x}_{J^c}},\ J \in \mathcal{E}\}$, $\theta_\emptyset^{\overline{x}} = 1$, $\overline{x} \subset \mathbf{Z}^\nu$, $\overline{x}_{J^c} = \overline{x} \cap (\mathbf{Z}^\nu \setminus J)$ is called *Q-function with boundary condition* $\overline{x}$, if for any $x \in \mathcal{E}$

$$\sum_{J \subseteq x} (-1)^{|x \setminus J|}\ \theta_J^{\overline{x}_{J^c}} \geq 0.$$

Any function $H(\overline{x}) = \{H_J^{\overline{x}_{J^c}},\ J \in \mathcal{E}\}$, $H_\emptyset^{\overline{x}} = 1$, $\overline{x} \subset \mathbf{Z}^\nu$ with nonnegative values we call *a H-function with boundary condition* $\overline{x}$.

THEOREM 5. *A function* $\theta(\overline{x}) = \{\theta_J^{\overline{x}_{J^c}},\ J \in \mathcal{E}\}$, $\theta_\emptyset^{\overline{x}} = 1$, $\overline{x} \subset \mathbf{Z}^\nu$ *is a Q-function with boundary condition* $\overline{x}$, *if and only if there exists a H-function* $H(\overline{x})$ *such that for any* $\Lambda \in \mathcal{E}$

$$\theta_\Lambda^{\overline{x}_{\Lambda^c}} = \sum_{J \subset \Lambda} H_J^{\overline{x}_{J^c}}.$$

Definition 7. A system $\theta = \{\theta(\overline{x}),\ \overline{x} \subset \mathbf{Z}^\nu\}$ of Q-functions depending on boundary conditions is called *consistent*, if the corresponding system $H = \{H(\overline{x}),\ \overline{x} \subset \mathbf{Z}^\nu\}$ of H-functions has the following property: for any $J_1, J_2 \in \mathcal{E}$ and $\overline{x} \subset \mathbf{Z}^\nu$

$$H_{J_1 \cup J_2}^{\overline{x}_{(J_1 \cup J_2)^c}} = H_{J_1}^{\overline{x}_{(J_1 \cup J_2)^c}} H_{J_2}^{J_1 \cup \overline{x}_{(J_1 \cup J_2)^c}}.$$

THEOREM 6. *A system of conditional distributions (11') is consistent in Dobrushin sense if and only if there exists a consistent system of Q-functions* $\theta = \{\theta(\overline{x}),\ \overline{x} \subset \mathbf{Z}^\nu\}$ *such that*

$$Q_\Lambda^{\overline{x}}(x) = \frac{1}{\theta_\Lambda^{\overline{x}_{\Lambda^c}}} \sum_{J \subseteq x} (-1)^{|x \setminus J|}\ \theta_J^{\overline{x}_{J^c}}, \quad x \subset \Lambda.$$

The proofs of Theorems 5,6 are similar to those for Theorems 2,4. The latter correspond to the case of empty boundary conditions.

§6. *Q*-FUNCTIONS FOR GIBBS RANDOM FIELDS

A measurable function Φ defined on $\mathcal{E}$ we call *a potential* if

$$\sup_{a \in \mathbf{Z}^\nu} \sum_{J : a \in J \in \mathcal{E}} |\Phi(J)| < \infty. \tag{13}$$

The potential energy of the configuration $x \subset \Lambda$, $\Lambda \in \mathcal{E}$ with boundary condition $\overline{x} \subset \mathbf{Z}^\nu \setminus \Lambda$ is defined by the expression

$$U^{\overline{x}}(x) = \sum_{\emptyset \neq J \subset x} \sum_{\widetilde{J} \subset \overline{x}} \Phi(J \cup \widetilde{J}).$$

The Q-functions with boundary condition $\overline{x} \subset \mathbf{Z}\!\!\mathbf{Z}^\nu$ for Gibbs random fields have the form

$$\mathbf{Z}\!\!\mathbf{Z}(\overline{x}) = \left\{ \mathbf{Z}\!\!\mathbf{Z}_\Lambda^{\overline{x}_{\Lambda^c}} = \sum_{x \subset \Lambda} \exp\left[-U^{\overline{x}_{\Lambda^c}}(x)\right], \quad \Lambda \in \mathcal{E} \right\}.$$

The corresponding Gibbs conditional distributions are

$$Q_\Lambda^{\overline{x}_{\Lambda^c}}(x) = \left(\mathbf{Z}\!\!\mathbf{Z}_\Lambda^{\overline{x}_{\Lambda^c}}\right)^{-1} \exp\left[-U^{\overline{x}_{\Lambda^c}}(x)\right], \quad x \subset \Lambda.$$

Any element of close convex hull of the set of limiting Gibbs distributions is called *a Gibbs random field* (see [7], [8]). Note also that any H-system corresponding to a Gibbs random field has the form

$$\left\{ H_S^{\overline{x}} = \exp\left[-U^{\overline{x}}(S)\right], \quad S \in \mathcal{E}, \quad \overline{x} \subset \mathbf{Z}\!\!\mathbf{Z}^\nu \setminus S \right\}.$$

§7. NON GIBBSIAN RANDOM FIELDS

Below we construct some non Gibbsian random fields.

PROPOSITION 1. *Let $\theta = \{\theta_J^{\overline{x}},\ J \in \mathcal{E},\ \overline{x} \subset \mathbf{Z}\!\!\mathbf{Z}^\nu \setminus J\}$ be a consistent system of Q-functions and $H = \{H_J^{\overline{x}},\ J \in \mathcal{E},\ \overline{x} \subset \mathbf{Z}\!\!\mathbf{Z}^\nu \setminus J\}$ be the corresponding system of nonnegative functions (H-system). Let $R(\overline{x})$, $\overline{x} \subset \mathbf{Z}\!\!\mathbf{Z}^\nu$ be a function such that $R(\overline{x}_1) = R(\overline{x}_2)$ if $\overline{x}_1 = \overline{x}_2$ up to a finite number of lattice points. Then the system*

$$H_R = \left\{ \left(H_J^{\overline{x}}\right)^{R(\overline{x})}, \quad J \in \mathcal{E}, \quad \overline{x} \subset \mathbf{Z}\!\!\mathbf{Z}^\nu \setminus J \right\}$$

determines some consistent θ-system of Q-functions, which we denote by θ_R.

PROOF: For any $J_1, J_2 \in \mathcal{E}$ and $\overline{x} \subset \mathbf{Z}\!\!\mathbf{Z}^\nu \setminus \{J_1 \cup J_2\}$ we can write

$$\left(H_{J_1 \cup J_2}^{\overline{x}}\right)^{R(\overline{x})} = \left(H_{J_1}^{\overline{x}} H_{J_2}^{\overline{x} \cup J_1}\right)^{R(\overline{x})} = \left(H_{J_1}^{\overline{x}}\right)^{R(\overline{x})} \left(H_{J_2}^{\overline{x} \cup J_1}\right)^{R(\overline{x})} = \left(H_{J_1}^{\overline{x}}\right)^{R(\overline{x})} \left(H_{J_2}^{\overline{x} \cup J_1}\right)^{R(\overline{x} \cup J_1)}.$$

PROPOSITION 2. *Let $\theta = \{\theta_J,\ J \in \mathcal{E},\ \overline{x} \subset \mathbf{Z}\!\!\mathbf{Z}^\nu \setminus J\}$ be a Gibbsian system of Q-functions. Then the corresponding θ_R is non Gibbsian system of Q-functions.*

PROOF: Since θ is Gibbsian, then the corresponding H system has the form

$$H = \left\{ \exp\left[-U^{\overline{x}}(x)\right], \quad x \subset \Lambda, \quad \overline{x} \subset \mathbf{Z}\!\!\mathbf{Z}^\nu \setminus \Lambda, \quad \Lambda \in \mathcal{E} \right\}.$$

Hence

$$H_R = \left\{ \exp\left[-U^{\overline{x}}(x)R(\overline{x})\right], \quad x \subset \Lambda, \quad \overline{x} \subset \mathbf{Z}^\nu \setminus \Lambda, \quad \Lambda \in \mathcal{E} \right\}.$$

According to Proposition 1, H_R determines some θ_R, which is consistent and hence in turn determines a random field. Let us show that this random field is non Gibbsian, i.e. there is no potential $\widetilde{\Phi}$ such that

$$U^{\overline{x}}(x)R(\overline{x}) = \sum_{J \subset x} \sum_{\emptyset \neq \widetilde{J} \subset \overline{x}} \widetilde{\Phi}(J \cup \widetilde{J}). \tag{14}$$

Suppose the contrary is true, i.e. (14) holds. For $x = \emptyset$ we have

$$\left(\Phi(\emptyset) + \sum_{\emptyset \neq J \subset \overline{x}} \Phi(J) \right) R(\overline{x}) = \widetilde{\Phi}(\emptyset) + \sum_{\emptyset \neq J \subset \overline{x}} \widetilde{\Phi}(J).$$

Therefore, if $\overline{x} = \emptyset$, then $\Phi(\emptyset)R(\emptyset) = \widetilde{\Phi}(\emptyset)$, and if $\overline{x} = \{t\}$, then

$$[\Phi(\emptyset) + \Phi(t)]R(t) = \widetilde{\Phi}(\emptyset) + \widetilde{\Phi}(t), \quad [\Phi(\emptyset) + \Phi(t)]R(\emptyset) = \widetilde{\Phi}(\emptyset) + \widetilde{\Phi}(t), \quad \Phi(t)]R(\emptyset) = \widetilde{\Phi}(t).$$

In the same way we find that $\Phi(J)R(\emptyset) = \widetilde{\Phi}(t)$ for any $J \in \mathcal{E}$. Hence

$$H^{\overline{x}}(x)R(\overline{x}) = H^{\overline{x}}(x)R(\emptyset) \quad \text{or} \quad R(\overline{x}) = R(\emptyset).$$

But the last relation is not valid if $\overline{x}$ is infinite.

Now we demonstrate that the random field of Example 3 is non Gibbsian. The conditional distributions in question are

$$Q^{\overline{x}}_\Lambda(x) = p^{|x|}(\overline{x})(1 - p(\overline{x}))^{|\Lambda| - |x|}, \quad x \subset \Lambda, \quad \overline{x} \subset \mathbf{Z}^\nu \setminus \Lambda, \quad \Lambda \in \mathcal{E},$$

where

$$p(\overline{x}) = \lim_{|\Lambda| \uparrow \infty} \frac{|\overline{x}|}{|\Lambda|} \quad \text{a.s.}$$

We rewrite $Q^{\overline{x}}_\Lambda(x)$ as

$$Q^{\overline{x}}_\Lambda(x) = \frac{\left(\frac{p(\overline{x})}{1 - p(\overline{x})}\right)^{|x|}}{(1 - p(\overline{x}))^{-|\Lambda|}}.$$

Thus the norming factor should be $Z_\Lambda(\overline{x}) = (1 - p(\overline{x}))^{-|\Lambda|}$, and the potential energy is

$$U^{\overline{x}}(x) = |x| \ln \frac{p(\overline{x})}{1 - p(\overline{x})}.$$

According to Proposition 2, this potential energy fails to generate a Gibbsian random field in classical sense.

R E F E R E N C E S

1. R. V. Ambartzumian, H. S. Sukiasian, "Inclusion-exclusion and point processes", Acta Appl. Math., vol. 22, pp. 15 – 31, 1991.
2. R. H. Schonmann, "Projections of Gibbs measures may be non Gibbsian", CMP, vol. 124, pp. 1 – 7, 1989.
3. R. B. Israel, "Banach algebras and Kadanoff transformations", in Random Fields (Estergom 1979), North–Holland, vol. 7, pp. 593 – 608, 1981.
4. J. L. Lebovitz, C. Mayes, "The effect of an external field on an interface, entropic repulsion", J. Stat. Phys., vol. 46, pp. 39 – 49, 1987.
5. A. van Enter, R. Fernandez, A. Sokal, "Regularity properties and pathologies of position – space renormalization – group transformations: scope and limitations of Gibbsian theory", J. Stat. Phys., vol. 72, no. 5,6, 1993.
6. R. L. Dobrushin, "The description of a random field by means of conditional probabilities and conditions of its regularity", Th. Prob. Appl., vol. 13, pp. 197 – 224, 1968.
7. R. L. Dobrushin, "Gibbsian random fields for lattice systems with pairwise interactions", Funct. Anal. Appl., vol. 2, pp. 292 – 301, 1968.
8. R. L. Dobrushin, "The problem of uniqueness of a Gibbs random field and the problem of phase transition", Funct. Anal. Appl., vol. 2, pp. 302 – 312, 1968.
9. V. A. Malyshev, "Cluster expansion in lattice models of statistical physics and the quantum theory of fields", Russian Math. Surveys, vol. 35, no. 2, pp. 3 – 53, 1980.
10. V. A. Malyshev, R. A. Minlos, Gibbs Random Fields. Method of Cluster Expansions [in Russian], Nauka, Moscow, 1985.

19 November 1995

Institute of Mathematics
National Academy of Sciences of Armenia

Markov Processes Relat. Fields **7**, 193–214 (2001)

Markov Processes and **Related Fields**

Description of Random Fields by Means of One-Point Conditional Distributions and Some Applications

S. Dachian[1] **and B.S. Nahapetian**[2]

[1] Laboratoire de Mathématiques Appliquées, Université Blaise Pascal (Clermont-Ferrand 2), 63177 Aubière Cedex, France. E-mail: sdachian@ucfma.univ-bpclermont.fr
[2] Institute of Mathematics, National Academy of Sciences of Armenia, 24-b, Av. M. Bagramian, 375019 Yerevan, Armenia. E-mail: nahapet@instmath.sci.am

Received May 26, 2000, revised July 28, 2000

Abstract. The problem of description of random fields by means of one-point conditional distributions is considered. A necessary and sufficient condition for a given system of one-point distributions with boundary conditions to be a subsystem of some specification is given. A sufficient condition for existence of random fields with given one-point conditional distributions, as well as some applications concerning Gibbs description of random fields, non-Gibbsian random fields and martingale-difference random fields are presented.

Keywords: random fields, conditional probabilities, one-point systems, Gibbs distributions

AMS Subject Classification: 82A05, 60G60

1. Introduction

The description of a random field by means of its conditional distribution was given by Dobrushin in his fundamental works [5–7]. In [5], Dobrushin gave sufficient conditions for existence and for uniqueness of a random field with given specification (consistent system of distributions in finite volumes with boundary conditions). The existence condition was imposed on the whole specification, while the uniqueness one was imposed only on its subsystem consisting of one-point distributions. Discussing this fact, Dobrushin noted that under some strict positivity conditions, the whole specification can be determined by its subsystem consisting only of one-point distributions, and stated the problem of

finding consistency conditions for a given system of one-point distributions with boundary conditions to be a subsystem of some specification. The answer to this problem would not only permit one to reformulate Dobrushin's theory in terms of one-point conditional distributions, but also to develop the theory in some directions.

In this paper we propose a solution to that problem, by giving necessary and sufficient consistency conditions for a given system of one-point distributions with boundary conditions to be a subsystem of some specification. So, instead of specifications we introduce a new object of consideration: one-point systems. We give sufficient conditions for existence of a random field with given one-point conditional distributions, obtain some new conditions for a Gibbs description of random fields without the usual assumption of strict positivity of its conditional probabilities, propose some scheme for constructing non-Gibbsian random fields and a simple method for constructing martingale-difference random fields. Another application of this result concerning nonparametric identification of random fields was given in [2].

Note that in this paper we consider so-called weakly positive or vacuum specifications with finite state spaces (for example lattice gas models of statistical physics). Generalizations to the case of vacuum systems with more general state spaces are possible.

Note also that a preliminary exposition of our results was given in [4].

2. Preliminaries

2.1. Random fields and specifications

We consider random fields on the ν-dimensional integer lattice $\mathbf{Z}^\nu$, i.e., probability measures P on $(\mathcal{X}^{\mathbf{Z}^\nu}, \mathcal{F}^{\mathbf{Z}^\nu})$ where $(\mathcal{X}, \mathcal{F})$ is some *state space*, i.e., space of values of a single variable. Usually the space $\mathcal{X}$ is assumed to be endowed with some topology $\mathcal{T}$, and $\mathcal{F}$ is assumed to be the Borel σ-algebra for this topology.

In this work we concentrate on the case when $\mathcal{X}$ is finite, $\mathcal{T}$ is the discrete topology and $\mathcal{F}$ is the total σ-algebra, that is $\mathcal{F} = \mathcal{T} = \exp(\mathcal{X})$.

For any $S \subset \mathbf{Z}^\nu$ let us consider the space $\mathcal{X}^S$ of all configurations on S. If $S = \emptyset$, we assume that the space $\mathcal{X}^\emptyset = \{\boldsymbol{\emptyset}\}$ where $\boldsymbol{\emptyset}$ is understood as an empty configuration. A probability distribution on $\mathcal{X}^S$ is denoted by P_S.

For any $T, S \subset \mathbf{Z}^\nu$ such that $T \subset S$ and any configuration $\mathbf{x} = \{x_t,\ t \in S\}$ on S, we denote by $\mathbf{x}_T$ the *subconfiguration* (*restriction*) of $\mathbf{x}$ on T defined by $\mathbf{x}_T = \{x_t,\ t \in T\}$. For any $T, S \subset \mathbf{Z}^\nu$ such that $T \cap S = \emptyset$ and any configurations $\mathbf{x}$ on T and $\mathbf{y}$ on S we denote by $\mathbf{x} \oplus \mathbf{y}$ a configuration on $T \cup S$ equal to $\mathbf{x}$ on T and to $\mathbf{y}$ on S.

Denote by $\mathcal{E}$ the set of all finite subsets of $\mathbf{Z}^\nu$, i.e., let $\mathcal{E} = \{\Lambda \subset \mathbf{Z}^\nu : |\Lambda| < \infty\}$ where $|\Lambda|$ is the number of points of the set Λ. If $\Lambda \in \mathcal{E} \setminus \{\emptyset\}$, we can write

$\mathsf{P}_\Lambda = \{\mathsf{P}_\Lambda(\mathbf{x}),\ \mathbf{x} \in \mathcal{X}^\Lambda\}$. For convenience of notations we agree that for $\Lambda = \emptyset$ there exists only one probability distribution $\mathsf{P}_\emptyset(\emptyset) = 1$.

For each $\Lambda \in \mathcal{E}$ and $I \subset \Lambda$ we write

$$(\mathsf{P}_\Lambda)_I(\mathbf{x}) = \sum_{\mathrm{y} \in \mathcal{X}^{\Lambda \setminus I}} \mathsf{P}_\Lambda(\mathbf{x} \oplus \mathbf{y}), \quad \mathbf{x} \in \mathcal{X}^I, \tag{2.1}$$

to denote the restriction (or marginal) $(\mathsf{P}_\Lambda)_I$ of P_Λ on I.

Any random field P on $\mathbf{Z}^\nu$ can be described in terms of its *finite-dimensional distributions* $\{\mathsf{P}_\Lambda,\ \Lambda \in \mathcal{E}\}$ which are consistent in the sense that for any $\Lambda \in \mathcal{E}$ and $I \subset \Lambda$ we have $\big(\mathsf{P}_\Lambda\big)_I = \mathsf{P}_I$.

For all $\Lambda \in \mathcal{E}$ there exist for P_{Λ^c}-almost all $\overline{\mathbf{x}} \in \mathcal{X}^{\Lambda^c}$ the following limits

$$\mathbf{q}_\Lambda^{\overline{\mathrm{x}}}(\mathbf{x}) = \lim_{\widetilde{\Lambda} \uparrow \Lambda^c} \frac{\mathsf{P}_{\Lambda \cup \widetilde{\Lambda}}(\mathbf{x} \oplus \overline{\mathbf{x}}_{\widetilde{\Lambda}})}{\mathsf{P}_{\widetilde{\Lambda}}(\overline{\mathbf{x}}_{\widetilde{\Lambda}})}\ , \quad \mathbf{x} \in \mathcal{X}^\Lambda.$$

Any system

$$\boldsymbol{\mathcal{Q}} = \{\mathsf{Q}_\Lambda^{\overline{\mathrm{x}}},\ \Lambda \in \mathcal{E} \text{ and } \overline{\mathbf{x}} \in \mathcal{X}^{\Lambda^c}\}$$

of probability distributions such that for all $\Lambda \in \mathcal{E}$ we have $\mathsf{Q}_\Lambda^{\overline{\mathrm{x}}} = \mathbf{q}_\Lambda^{\overline{\mathrm{x}}}$ for P_{Λ^c}-almost all $\overline{\mathbf{x}} \in \mathcal{X}^{\Lambda^c}$ is called *conditional distribution* of the random field P.

A conditional distribution $\boldsymbol{\mathcal{Q}}$ of a random field P satisfies P-almost surely the condition

$$\mathsf{Q}_{\Lambda \cup \widetilde{\Lambda}}^{\overline{\mathrm{x}}}(\mathbf{x} \oplus \mathbf{y}) = \mathsf{Q}_\Lambda^{\overline{\mathrm{x}} \oplus \mathrm{y}}(\mathbf{x})\, (\mathsf{Q}_{\Lambda \cup \widetilde{\Lambda}}^{\overline{\mathrm{x}}})_{\widetilde{\Lambda}}(\mathbf{y})$$

where $\Lambda, \widetilde{\Lambda} \in \mathcal{E}$, $\Lambda \cap \widetilde{\Lambda} = \emptyset$, $\mathbf{x} \in \mathcal{X}^\Lambda$, $\mathbf{y} \in \mathcal{X}^{\widetilde{\Lambda}}$ and $\overline{\mathbf{x}} \in \mathcal{X}^{\left(\Lambda \cup \widetilde{\Lambda}\right)^c}$.

Definition 2.1. A system

$$\boldsymbol{\mathcal{Q}} = \{\mathsf{Q}_\Lambda^{\overline{\mathrm{x}}}, \quad \Lambda \in \mathcal{E} \text{ and } \overline{\mathbf{x}} \in \mathcal{X}^{\Lambda^c}\}$$

of probability distributions is called *specification* if for any $\Lambda, \widetilde{\Lambda} \in \mathcal{E}$ such that $\Lambda \cap \widetilde{\Lambda} = \emptyset$ and for any $\mathbf{x} \in \mathcal{X}^\Lambda$, $\mathbf{y} \in \mathcal{X}^{\widetilde{\Lambda}}$ and $\overline{\mathbf{x}} \in \mathcal{X}^{\left(\Lambda \cup \widetilde{\Lambda}\right)^c}$ we have

$$\mathsf{Q}_{\Lambda \cup \widetilde{\Lambda}}^{\overline{\mathrm{x}}}(\mathbf{x} \oplus \mathbf{y}) = \mathsf{Q}_\Lambda^{\overline{\mathrm{x}} \oplus \mathrm{y}}(\mathbf{x})\, (\mathsf{Q}_{\Lambda \cup \widetilde{\Lambda}}^{\overline{\mathrm{x}}})_{\widetilde{\Lambda}}(\mathbf{y}). \tag{2.2}$$

One of the main goals of random field theory is to study the set of all random fields having a given specification as a conditional distribution, and particularly to find conditions on the specification, sufficient for existence and uniqueness of such random fields. The best known conditions of such type are the quasilocality of a specification for existence, and Dobrushin's uniqueness condition for uniqueness.

Definition 2.2. Let $S \subset \mathbf{Z}^\nu$ and let $\mathbf{g} = \{g^{\mathrm{x}},\ \mathbf{x} \in \mathcal{X}^S\}$ be an arbitrary real-valued function on $\mathcal{X}^S$. We say that the function $\mathbf{g}$ is *local* if there exist some finite $\Lambda \subset S$ such that $g^{\mathrm{x}} = g^{\mathrm{y}}$ for all $\mathbf{x}, \mathbf{y} \in \mathcal{X}^S$ satisfying $\mathbf{x}_\Lambda = \mathbf{y}_\Lambda$, and it is *quasilocal* if it is a uniform limit of local functions.

Note that quasilocality of $\mathbf{g}$ is equivalent to its continuity with respect to the topology $\mathcal{T}^S$, or also to the following condition:

$$\sup_{\mathrm{x}\in\mathcal{X}^S} |g^{\mathrm{x}^a_I} - g^{\mathrm{x}}| \xrightarrow[I\uparrow S]{} 0, \tag{2.3}$$

where $\mathbf{x}^a_I$ denotes the configuration on S equal to $\mathbf{x}_I$ on I and to some fixed $a \in \mathcal{X}$ in all points of $S \setminus I$.

Definition 2.3. A specification $\boldsymbol{\mathcal{Q}}$ is called (*quasi*)*local* if for all $\Lambda \in \mathcal{E}$ and $\mathbf{x} \in \mathcal{X}^\Lambda$ the function $\mathsf{Q}^{\overline{\mathrm{x}}}_\Lambda(\mathbf{x})$ is (quasi)local as a function of $\overline{\mathbf{x}} \in \mathcal{X}^{\Lambda^{\mathrm{c}}}$.

Definition 2.4. We say that a specification $\boldsymbol{\mathcal{Q}}$ satisfies *Dobrushin's uniqueness condition* if it is quasilocal and we have

$$\frac{1}{2} \sup_{t\in\mathrm{Z}^\nu} \sum_{s\in\mathrm{Z}^\nu\setminus t} \sup_{\overline{\mathrm{x}},\overline{\mathrm{y}}} \sum_{x\in\mathcal{X}} \left|\mathsf{Q}^{\overline{\mathrm{x}}}_t(x_t) - \mathsf{Q}^{\overline{\mathrm{y}}}_t(x_t)\right| < 1, \tag{2.4}$$

where the second sup is taken over all pairs $\overline{\mathbf{x}}, \overline{\mathbf{y}} \in \mathcal{X}^{\mathrm{Z}^\nu\setminus t}$ such that we have $\overline{\mathbf{x}}_{\mathrm{Z}^\nu\setminus\{s,t\}} = \overline{\mathbf{y}}_{\mathrm{Z}^\nu\setminus\{s,t\}}$.

Here and in the sequel, for convenience of notations, we write t for the one-point set $\{t\}$, and x_t for the configuration taking value $x \in \mathcal{X}$ on the set t.

Now we can state the following theorem (see [5]).

Theorem 2.1. *Let $\boldsymbol{\mathcal{Q}}$ be some fixed specification.*

1) *If $\boldsymbol{\mathcal{Q}}$ is quasilocal, then there exists a random field P having $\boldsymbol{\mathcal{Q}}$ as a conditional distribution.*

2) *If $\boldsymbol{\mathcal{Q}}$ satisfies Dobrushin's uniqueness condition, then the random field P having $\boldsymbol{\mathcal{Q}}$ as a conditional distribution is unique.*

2.2. Gibbsian specifications

The best known examples of specifications are Gibbsian specifications. These specifications have the following (Gibbsian) form:

$$\mathsf{Q}^{\overline{\mathrm{x}}}_\Lambda(\mathbf{x}) = \frac{\exp(-U^{\overline{\mathrm{x}}}_\Lambda(\mathbf{x}))}{\sum_{\mathrm{y}\in\mathcal{X}^\Lambda} \exp(-U^{\overline{\mathrm{x}}}_\Lambda(\mathbf{y}))}\ , \quad \Lambda \in \mathcal{E},\ \mathbf{x} \in \mathcal{X}^\Lambda,\ \overline{\mathbf{x}} \in \mathcal{X}^{\Lambda^{\mathrm{c}}}.$$

Here the function $\boldsymbol{\mathcal{U}}_\Lambda = \{U^{\overline{\mathrm{x}}}_\Lambda(\mathbf{x}),\ \mathbf{x} \in \mathcal{X}^\Lambda,\ \overline{\mathbf{x}} \in \mathcal{X}^{\Lambda^{\mathrm{c}}}\}$ is called *Hamiltonian in* Λ, the set of functions $\boldsymbol{\mathcal{U}} = \{\boldsymbol{\mathcal{U}}_\Lambda,\ \Lambda \in \mathcal{E}\}$ is called *system of Hamiltonians*, and is assumed to be given by the formula

$$U^{\overline{\mathrm{x}}}_\Lambda(\mathbf{x}) = \sum_{J\,:\,\emptyset\neq J\subset\Lambda} \sum_{\widetilde{J}\in\mathcal{E}\,:\,\widetilde{J}\subset\Lambda^{\mathrm{c}}} \Phi(\mathbf{x}_J \oplus \overline{\mathbf{x}}_{\widetilde{J}}), \quad \Lambda \in \mathcal{E},\ \mathbf{x} \in \mathcal{X}^\Lambda,\ \overline{\mathbf{x}} \in \mathcal{X}^{\Lambda^{\mathrm{c}}}, \tag{2.5}$$

where $\mathbf{\Phi} = \{\Phi(\mathbf{x}),\ \mathbf{x} \in \mathcal{X}^J,\ J \in \mathcal{E} \setminus \{\emptyset\}\}$ is some function taking values in $\mathbf{R} \cup \{+\infty\}$ and called (*interaction*) *potential*. Here and in the sequel we use the convention that any sum over an empty space of indexes is equal to 0, i.e., $U^{\overline{\mathrm{x}}}_{\emptyset}(\emptyset) = 0$ for all $\overline{\mathbf{x}} \in \mathcal{X}^{\mathbf{Z}^\nu}$.

Let us put

$$u^{\overline{\mathrm{x}}}_{\Delta}(x_t) \ = \sum_{\widetilde{J} \subset \Delta \setminus t} \Phi(x_t \oplus \overline{\mathbf{x}}_{\widetilde{J}}), \quad t \in \mathbf{Z}^\nu, \Delta \in \mathcal{E},\ x \in \mathcal{X},\ \overline{\mathbf{x}} \in \mathcal{X}^{\mathbf{Z}^\nu \setminus t}.$$

In order for the system of Hamiltonians $\boldsymbol{\mathcal{U}}$ to be well-defined, potentials are always supposed to be such that the limit

$$u^{\overline{\mathrm{x}}}(x_t) = \lim_{\Delta \uparrow \mathbf{Z}^\nu} u^{\overline{\mathrm{x}}}_{\Delta}(x_t) \tag{2.6}$$

exists and is in $\mathbf{R} \cup \{+\infty\}$ for all $t \in \mathbf{Z}^\nu$, $x \in \mathcal{X}$ and $\overline{\mathbf{x}} \in \mathcal{X}^{\mathbf{Z}^\nu \setminus t}$. Since $u^{\overline{\mathrm{x}}}(x_t) = U^{\overline{\mathrm{x}}}_t(x_t)$ for all $t \in \mathbf{Z}^\nu$, $x \in \mathcal{X}$ and $\overline{\mathbf{x}} \in \mathcal{X}^{\mathbf{Z}^\nu \setminus t}$, we call the system $\mathbf{u} = \{u^{\overline{\mathrm{x}}}(x_t),\ t \in \mathbf{Z}^\nu,\ x \in \mathcal{X},\ \overline{\mathbf{x}} \in \mathcal{X}^{\mathbf{Z}^\nu \setminus t}\}$ *one-point Hamiltonian*.

Potentials satisfying (2.6) are called *convergent*. Usually the potentials are supposed to be *uniformly convergent*, i.e., potential and Hamiltonians are supposed to be finite, and the convergence in (2.6) (and hence in (2.5)) is supposed to be uniform with respect to $\overline{\mathbf{x}}$. Note that Gibbsian specifications with uniformly convergent potentials are clearly quasilocal.

2.3. Vacuum systems

Let us start by introducing the notion of so-called vacuum potentials. Fix some element $\theta \in \mathcal{X}$ which will be called *vacuum* and let $\mathcal{X}^*$ denote $\mathcal{X} \setminus \theta$.

Definition 2.5. A potential $\mathbf{\Phi} = \{\Phi(\mathbf{x}),\ \mathbf{x} \in \mathcal{X}^J,\ J \in \mathcal{E} \setminus \{\emptyset\}\}$ is called *vacuum potential* (with the vacuum θ) if we have $\Phi(\mathbf{x}) = 0$ for all $\mathbf{x} \in \mathcal{X}^J$ such that there exists some $t \in J$ satisfying $x_t = \theta$.

The class of vacuum potentials corresponds to so-called lattice gas models of statistical physics. Note that for an arbitrary uniformly convergent potential and any $\theta \in \mathcal{X}$, one can find a unique (not necessarily uniformly convergent) vacuum potential giving the same specification as the initial one (see, for example, [9]). In physical terminology $x_t = \theta$ means that this site is not occupied by any particle, while all other values represent different types of particles.

Consider an arbitrary configuration $\mathbf{x} \in \mathcal{X}^S$, $S \subset \mathbf{Z}^\nu$. Denote by T the set of sites occupied by particles, i.e., $T = \{t \in S,\ x_t \neq \theta\} \subset S$. Clearly, we have $\mathbf{x} = \mathbf{x}^\theta_T$, and hence the configuration $\mathbf{x}$ is uniquely determined by its subconfiguration $\mathbf{x}_T \in \mathcal{X}^{*T}$. In the sequel we will not distinguish between this two configurations and will write, for example, $\mathbf{x} \in \mathcal{X}^{*T}$, $T \subset S$ for a configuration $\mathbf{x}$ on S.

Now we can rewrite all the above formulas in these notations. For example, the quasilocality condition (2.3) becomes

$$\sup_{\mathrm{x}\in\mathcal{X}^{*T},\ T\subset S} |g^{\mathrm{x}_I} - g^{\mathbf{x}}| \xrightarrow[I\uparrow S]{} 0.$$

The Gibbsian form is

$$\mathsf{Q}^{\overline{\mathrm{x}}}_{\Lambda}(\mathbf{x}) = \frac{\exp(-U^{\overline{\mathrm{x}}}(\mathbf{x}))}{\sum_{\mathrm{y}\in\mathcal{X}^{\Lambda}} \exp(-U^{\overline{\mathrm{x}}}(\mathbf{y}))}\ , \quad \Lambda\in\mathcal{E},\ \mathbf{x}\in\mathcal{X}^{*I},\ I\subset\Lambda,\ \overline{\mathbf{x}}\in\mathcal{X}^{*S},\ S\subset\Lambda^{\mathrm{c}}.$$

The Hamiltonian $\boldsymbol{\mathcal{U}} = \{U^{\overline{\mathrm{x}}}(\mathbf{x}),\ \mathbf{x}\in\mathcal{X}^{*I},\ I\in\mathcal{E},\ \overline{\mathbf{x}}\in\mathcal{X}^{*S},\ S\subset I^{\mathrm{c}}\}$ corresponding to a potential $\boldsymbol{\Phi} = \{\Phi(\mathbf{x}),\ \mathbf{x}\in\mathcal{X}^{*J},\ J\in\mathcal{E}\setminus\{\emptyset\}\}$ is given by the formula

$$U^{\overline{\mathrm{x}}}(\mathbf{x}) \ = \sum_{J\,:\,\emptyset\neq J\subset I}\ \sum_{\widetilde{J}\in\mathcal{E}\,:\,\widetilde{J}\subset S} \Phi(\mathbf{x}_J\oplus\overline{\mathbf{x}}_{\widetilde{J}}).$$

Note that the Hamiltonian no longer depends on Λ. In fact, vacuum condition implies that for all $\Lambda\in\mathcal{E}$ satisfying $I\subset\Lambda\subset S^{\mathrm{c}}$ we get the same Hamiltonian (adding empty sites does not change the energy of a configuration). The relation (2.6) can be rewritten as

$$u^{\overline{\mathrm{x}}}(x_t) = \lim_{\Delta\uparrow\mathbb{Z}^{\nu}} u^{\overline{\mathrm{x}}_{\Delta}}(x_t).$$

Let us finally note here, that in the vacuum case we clearly have $U^{\overline{\mathrm{x}}}(\emptyset)=0$ for all $\overline{\mathbf{x}}\in\mathcal{X}^{*S}$, $S\subset\mathbf{Z}^{\nu}$, and hence we have $\mathsf{Q}^{\overline{\mathrm{x}}}_{\Lambda}(\emptyset)>0$ for all $\Lambda\in\mathcal{E}$ and $\overline{\mathbf{x}}\in\mathcal{X}^{*S}$, $S\subset\Lambda^{\mathrm{c}}$. Here $\emptyset$ is nothing but the configuration θ^{Λ} identically equal to θ on Λ. This remark leads us to introduce the following

Definition 2.6. A specification

$$\boldsymbol{\mathcal{Q}} = \{\mathsf{Q}^{\overline{\mathrm{x}}}_{\Lambda},\ \Lambda\in\mathcal{E},\ \overline{\mathbf{x}}\in\mathcal{X}^{*S},\ S\subset\Lambda^{\mathrm{c}}\}$$

is called *vacuum specification* (with the vacuum θ) if for all $\Lambda\in\mathcal{E}$ and $\overline{\mathbf{x}}\in\mathcal{X}^{*S}$, $S\subset\Lambda^{\mathrm{c}}$, we have

$$\mathsf{Q}^{\overline{\mathrm{x}}}_{\Lambda}(\emptyset) > 0. \tag{2.7}$$

Sometimes vacuum specifications are also called *weakly positive specifications*, and the condition (2.7) is called *"essentiality" of vacuum.*

Note finally that for vacuum specifications the consistency condition (2.2) can be rewritten in an equivalent form

$$\mathsf{Q}^{\overline{\mathrm{x}}}_{\Lambda\cup\widetilde{\Lambda}}(\mathbf{x}\oplus\mathbf{y}) = \frac{\mathsf{Q}^{\overline{\mathrm{x}}}_{\Lambda\cup\widetilde{\Lambda}}(\mathbf{x})}{\mathsf{Q}^{\overline{\mathrm{x}}\oplus\mathrm{x}}_{\widetilde{\Lambda}}(\emptyset)}\ \mathsf{Q}^{\overline{\mathrm{x}}\oplus\mathrm{x}}_{\widetilde{\Lambda}}(\mathbf{y}). \tag{2.8}$$

3. One-point systems

In this section we propose a description of vacuum specifications, based on the notion of a consistent system of functions (called one-point system), which is closely related to one-point distributions. Particularly, the answer to Dobrushin's problem for vacuum specifications follows immediately from this description.

For simplicity, we will first consider the case when $\mathcal{X} = \{0,1\}$, the vacuum $\theta = 0$ and $\mathcal{X}^* = \{1\}$, i.e., there exists just one type of particles, corresponding to 1. Hence, any configuration $\mathbf{x}$ on $S \subset \mathbf{Z}^\nu$ is identified with a subset T of S where the configuration $\mathbf{x}$ takes the value 1. In the sequel we will not distinguish between these two notions and will write, for example, $\mathbf{x} \subset S$ for a configuration $\mathbf{x}$ on S.

With these notations we have $\mathbf{x}_\Lambda = \mathbf{x} \cap \Lambda$, $\mathbf{x} \oplus \mathbf{y} = \mathbf{x} \cup \mathbf{y}$, the Gibbsian form is

$$\mathsf{Q}_\Lambda^{\overline{\mathrm{x}}}(\mathbf{x}) = \frac{\exp(-U^{\overline{\mathrm{x}}}(\mathbf{x}))}{\sum_{\mathrm{y} \subset \Lambda} \exp(-U^{\overline{\mathrm{x}}}(\mathbf{y}))}\ , \quad \Lambda \in \mathcal{E},\ \mathbf{x} \subset \Lambda,\ \overline{\mathbf{x}} \subset \Lambda^{\mathrm{c}},$$

the Hamiltonian $\boldsymbol{\mathcal{U}} = \{U^{\overline{\mathrm{x}}}(\mathbf{x}),\ \mathbf{x} \in \mathcal{E} \text{ and } \overline{\mathbf{x}} \subset \mathbf{x}^{\mathrm{c}}\}$ corresponding to a potential $\boldsymbol{\Phi} = \{\Phi(J),\ J \in \mathcal{E} \setminus \{\emptyset\}\}$ is given by the formula

$$U^{\overline{\mathrm{x}}}(\mathbf{x}) \ = \sum_{J\,:\,\emptyset \neq J \subset \mathrm{x}} \ \sum_{\widetilde{J} \in \mathcal{E}\,:\,\widetilde{J} \subset \overline{\mathrm{x}}} \Phi(J \cup \widetilde{J}),$$

and the condition of essentiality of vacuum is just $\mathsf{Q}_\Lambda^{\overline{\mathrm{x}}}(\emptyset) > 0$ for all $\Lambda \in \mathcal{E}$ and $\overline{\mathbf{x}} \subset \Lambda^{\mathrm{c}}$.

Before introducing description of specifications by means of one-point systems, we need some preliminary results.

3.1. Description of specifications by means of *H*-systems

Here we show that a vacuum specification can be described by means of a system of functions (H-system), satisfying some consistency conditions. A more detailed exposition of the results concerning H-systems can be found in [3] and [4].

Definition 3.1. A system $\boldsymbol{\mathcal{H}} = \{H_{\mathrm{x}}^{\overline{\mathrm{x}}},\ \mathbf{x} \in \mathcal{E} \text{ and } \overline{\mathbf{x}} \subset \mathbf{x}^{\mathrm{c}}\}$ such that $H_{\mathrm{x}}^{\overline{\mathrm{x}}} \geq 0$ for all $\mathbf{x} \in \mathcal{E}$, $\overline{\mathbf{x}} \subset \mathbf{x}^{\mathrm{c}}$, and $H_\emptyset^{\overline{\mathrm{x}}} = 1$ for all $\overline{\mathbf{x}} \subset \mathbf{Z}^\nu$, is called *$H$-system*.

An H-system $\boldsymbol{\mathcal{H}}$ is called *consistent* if it satisfies the following condition: for any $\mathbf{x}, \mathbf{y} \in \mathcal{E}$ such that $\mathbf{x} \cap \mathbf{y} = \emptyset$ and any $\overline{\mathbf{x}} \subset (\mathbf{x} \cup \mathbf{y})^{\mathrm{c}}$ we have

$$H_{\mathrm{x}\cup\mathrm{y}}^{\overline{\mathrm{x}}} = H_{\mathrm{x}}^{\overline{\mathrm{x}}}\, H_{\mathrm{y}}^{\overline{\mathrm{x}}\cup\mathrm{x}}. \tag{3.1}$$

Theorem 3.1. *A system* $\boldsymbol{\mathcal{Q}} = \{\mathsf{Q}^{\overline{\mathrm{x}}}_{\Lambda},\ \Lambda \in \mathcal{E} \text{ and } \overline{\mathbf{x}} \subset \Lambda^{\mathrm{c}}\}$ *is a vacuum specification if and only if there exists a consistent* H*-system* $\boldsymbol{\mathcal{H}}$ *such that for any* $\Lambda \in \mathcal{E}$ *and any* $\overline{\mathbf{x}} \subset \Lambda^{\mathrm{c}}$ *we have*

$$\mathsf{Q}^{\overline{\mathrm{x}}}_{\Lambda}(\mathbf{x}) = \frac{H^{\overline{\mathrm{x}}}_{\mathrm{x}}}{\sum_{\mathrm{y}\subset\Lambda} H^{\overline{\mathrm{x}}}_{\mathrm{y}}}\ , \quad \mathbf{x} \subset \Lambda. \tag{3.2}$$

Proof. 1) Necessity. Let $\boldsymbol{\mathcal{Q}} = \{\mathsf{Q}^{\overline{\mathrm{x}}}_{\Lambda},\ \Lambda \in \mathcal{E} \text{ and } \overline{\mathbf{x}} \subset \Lambda^{\mathrm{c}}\}$ be a specification with $\mathsf{Q}^{\overline{\mathrm{x}}}_{\Lambda}(\emptyset) > 0$ for all $\Lambda \in \mathcal{E}$ and $\overline{\mathbf{x}} \subset \Lambda^{\mathrm{c}}$. For all $\mathbf{x} \in \mathcal{E}$ and $\overline{\mathbf{x}} \subset \mathbf{x}^{\mathrm{c}}$, we use the notation

$$H^{\overline{\mathrm{x}}}_{\mathrm{x}} = \sum_{J\subset\mathrm{x}} (-1)^{|\mathrm{x}\setminus J|} \frac{1}{\mathsf{Q}^{\overline{\mathrm{x}}}_{J}(\emptyset)}.$$

Let us show that (3.2) holds. For any $\Lambda \in \mathcal{E}$, $J \subset \Lambda$ and $\overline{\mathbf{x}} \subset \Lambda^{\mathrm{c}}$ we can write

$$1 = \sum_{\mathrm{y}\subset J} \mathsf{Q}^{\overline{\mathrm{x}}}_{J}(\mathbf{y}) = \sum_{\mathrm{y}\subset J} \frac{\mathsf{Q}^{\overline{\mathrm{x}}}_{J}(\emptyset)}{\mathsf{Q}^{\overline{\mathrm{x}}}_{\Lambda}(\emptyset)}\, \mathsf{Q}^{\overline{\mathrm{x}}}_{\Lambda}(\mathbf{y}) = \frac{\mathsf{Q}^{\overline{\mathrm{x}}}_{J}(\emptyset)}{\mathsf{Q}^{\overline{\mathrm{x}}}_{\Lambda}(\emptyset)} \sum_{\mathrm{y}\subset J} \mathsf{Q}^{\overline{\mathrm{x}}}_{\Lambda}(\mathbf{y}),$$

and hence

$$\frac{1}{\mathsf{Q}^{\overline{\mathrm{x}}}_{J}(\emptyset)} = \frac{1}{\mathsf{Q}^{\overline{\mathrm{x}}}_{\Lambda}(\emptyset)} \sum_{\mathrm{y}\subset J} \mathsf{Q}^{\overline{\mathrm{x}}}_{\Lambda}(\mathbf{y}).$$

Therefore

$$H^{\overline{\mathrm{x}}}_{\mathrm{x}} = \sum_{J\subset\mathrm{x}} (-1)^{|\mathrm{x}\setminus J|} \frac{1}{\mathsf{Q}^{\overline{\mathrm{x}}}_{J}(\emptyset)} = \frac{1}{\mathsf{Q}^{\overline{\mathrm{x}}}_{\Lambda}(\emptyset)} \sum_{J\subset\mathrm{x}} (-1)^{|\mathrm{x}\setminus J|} \sum_{\mathrm{y}\subset J} \mathsf{Q}^{\overline{\mathrm{x}}}_{\Lambda}(\mathbf{y}) = \frac{\mathsf{Q}^{\overline{\mathrm{x}}}_{\Lambda}(\mathbf{x})}{\mathsf{Q}^{\overline{\mathrm{x}}}_{\Lambda}(\emptyset)}, \tag{3.3}$$

and taking into account the equality

$$\sum_{\mathrm{y}\subset\Lambda} H^{\overline{\mathrm{x}}}_{\mathrm{y}} = \frac{1}{\mathsf{Q}^{\overline{\mathrm{x}}}_{\Lambda}(\emptyset)}$$

we obtain (3.2).

Obviously, (3.3) implies that $\boldsymbol{\mathcal{H}} = \{H^{\overline{\mathrm{x}}}_{\mathrm{x}},\ \mathbf{x} \in \mathcal{E} \text{ and } \overline{\mathbf{x}} \subset \mathbf{x}^{\mathrm{c}}\}$ is an H-system, and it remains to verify its consistency. For any $\mathbf{x}, \mathbf{y} \in \mathcal{E}$ such that $\mathbf{x} \cap \mathbf{y} = \emptyset$ and any $\overline{\mathbf{x}} \subset (\mathbf{x} \cup \mathbf{y})^{\mathrm{c}}$, using (3.3) and (2.8) we can write

$$H^{\overline{\mathrm{x}}}_{\mathrm{x}}\, H^{\overline{\mathrm{x}}\cup\mathrm{x}}_{\mathrm{y}} = \frac{\mathsf{Q}^{\overline{\mathrm{x}}}_{\mathrm{x}\cup\mathrm{y}}(\mathbf{x})}{\mathsf{Q}^{\overline{\mathrm{x}}}_{\mathrm{x}\cup\mathrm{y}}(\emptyset)}\ \frac{\mathsf{Q}^{\overline{\mathrm{x}}\cup\mathrm{x}}_{\mathrm{y}}(\mathbf{y})}{\mathsf{Q}^{\overline{\mathrm{x}}\cup\mathrm{x}}_{\mathrm{y}}(\emptyset)} = \frac{\mathsf{Q}^{\overline{\mathrm{x}}}_{\mathrm{x}\cup\mathrm{y}}(\mathbf{x}\cup\mathbf{y})}{\mathsf{Q}^{\overline{\mathrm{x}}}_{\mathrm{x}\cup\mathrm{y}}(\emptyset)} = H^{\overline{\mathrm{x}}}_{\mathrm{x}\cup\mathrm{y}}$$

which concludes the proof of necessity.

2) Sufficiency. Let $\boldsymbol{\mathcal{H}} = \{H^{\overline{\mathrm{x}}}_{\mathrm{x}},\ \mathbf{x} \in \mathcal{E} \text{ and } \overline{\mathbf{x}} \subset \mathbf{x}^{\mathrm{c}}\}$ be a consistent H-system. For all $\Lambda \in \mathcal{E}$, $\mathbf{x} \subset \Lambda$ and $\overline{\mathbf{x}} \subset \Lambda^{\mathrm{c}}$ put

$$\mathsf{Q}^{\overline{\mathrm{x}}}_{\Lambda}(\mathbf{x}) = \frac{H^{\overline{\mathrm{x}}}_{\mathrm{x}}}{\sum_{\mathrm{y}\subset\Lambda} H^{\overline{\mathrm{x}}}_{\mathrm{y}}}.$$

Let us prove that $\boldsymbol{\mathcal{Q}} = \{\mathsf{Q}^{\overline{\mathrm{x}}}_{\Lambda},\ \Lambda \in \mathcal{E} \text{ and } \overline{\mathbf{x}} \subset \Lambda^{\mathrm{c}}\}$ is a specification. Obviously Q is a system of probability distributions in finite volumes with boundary conditions. It remains to verify the consistency condition (2.8). We have

$$\begin{aligned}\mathsf{Q}^{\overline{\mathrm{x}}}_{\Lambda\cup\widetilde{\Lambda}}(\mathbf{x}\cup\mathbf{y}) &= \frac{H^{\overline{\mathrm{x}}}_{\mathrm{x}\cup\mathrm{y}}}{\sum_{\mathrm{z}\subset\Lambda\cup\widetilde{\Lambda}} H^{\overline{\mathrm{x}}}_{\mathrm{z}}} = \frac{H^{\overline{\mathrm{x}}}_{\mathrm{x}}\, H^{\overline{\mathrm{x}}\cup\mathrm{x}}_{\mathrm{y}}}{\sum_{\mathrm{z}\subset\Lambda\cup\widetilde{\Lambda}} H^{\overline{\mathrm{x}}}_{\mathrm{z}}} \\ &= \frac{H^{\overline{\mathrm{x}}}_{\mathrm{x}}}{\sum_{\mathrm{z}\subset\Lambda\cup\widetilde{\Lambda}} H^{\overline{\mathrm{x}}}_{\mathrm{z}}} \frac{H^{\overline{\mathrm{x}}\cup\mathrm{x}}_{\mathrm{y}}}{\sum_{\mathrm{z}\subset\widetilde{\Lambda}} H^{\overline{\mathrm{x}}\cup\mathrm{x}}_{\mathrm{z}}} \sum_{\mathrm{z}\subset\widetilde{\Lambda}} H^{\overline{\mathrm{x}}\cup\mathrm{x}}_{\mathrm{z}} = \frac{\mathsf{Q}^{\overline{\mathrm{x}}}_{\Lambda\cup\widetilde{\Lambda}}(\mathbf{x})}{\mathsf{Q}^{\overline{\mathrm{x}}\cup\mathrm{x}}_{\widetilde{\Lambda}}(\emptyset)}\, \mathsf{Q}^{\overline{\mathrm{x}}\cup\mathrm{x}}_{\widetilde{\Lambda}}(\mathrm{y}).\end{aligned}$$

The theorem is proved. □

Remark 3.1. Let $\boldsymbol{\mathcal{H}}$ be a consistent H-system. For all $\mathbf{x} \in \mathcal{E}$ and $\overline{\mathbf{x}} \subset \mathbf{x}^{\mathrm{c}}$ denote $U^{\overline{\mathrm{x}}}(\mathbf{x}) = -\ln H^{\overline{\mathrm{x}}}_{\mathrm{x}}$. Then the system $\boldsymbol{\mathcal{U}} = \{U^{\overline{\mathrm{x}}}(\mathbf{x}),\ \mathbf{x} \in \mathcal{E} \text{ and } \overline{\mathbf{x}} \subset \mathbf{x}^{\mathrm{c}}\}$ satisfies the following consistency property: for all $\mathbf{x}, \mathbf{y} \in \mathcal{E}$ such that $\mathbf{x} \cap \mathbf{y} = \emptyset$ and all $\overline{\mathbf{x}} \subset (\mathbf{x}\cup\mathbf{y})^{\mathrm{c}}$ we have

$$U^{\overline{\mathrm{x}}}(\mathbf{x}\cup\mathbf{y}) = U^{\overline{\mathrm{x}}}(\mathbf{x}) + U^{\overline{\mathrm{x}}\cup\mathrm{x}}(\mathbf{y}). \tag{3.4}$$

Now we can rewrite (3.2) in the form

$$\mathsf{Q}^{\overline{\mathrm{x}}}_{\Lambda}(\mathbf{x}) = \frac{\exp(-U^{\overline{\mathrm{x}}}(\mathbf{x}))}{\sum_{\mathrm{y}\subset\Lambda}\exp(-U^{\overline{\mathrm{x}}}(\mathbf{y}))}, \quad \Lambda \in \mathcal{E},\ \mathbf{x} \subset \Lambda,\ \overline{\mathbf{x}} \subset \Lambda^{\mathrm{c}},$$

which is the usual Gibbsian form with Hamiltonian $\boldsymbol{\mathcal{U}}$. But in our case, the system $\boldsymbol{\mathcal{U}}$ is an arbitrary system satisfying (3.4), and in general does not have an explicit form in terms of some potential.

3.2. Description of specifications by means of one-point systems

As we have already seen, consistent H-systems are convenient tool for description of vacuum specifications. Here we will show that one can describe specifications by means of more simple systems, namely by means of one-point systems.

Definition 3.2. A system $\mathbf{h} = \{h^{\overline{\mathrm{x}}}_t,\ t \in \mathbf{Z}^\nu \text{ and } \overline{\mathbf{x}} \subset \mathbf{Z}^\nu \setminus t\}$ is called *one-point system* if for all $t \in \mathbf{Z}^\nu$ and $\overline{\mathbf{x}} \subset \mathbf{Z}^\nu \setminus t$ we have $h^{\overline{\mathrm{x}}}_t \geq 0$ and for all $s, t \in \mathbf{Z}^\nu$ and $\overline{\mathbf{x}} \subset \mathbf{Z}^\nu \setminus \{s,t\}$ we have

$$h^{\overline{\mathrm{x}}}_s\, h^{\overline{\mathrm{x}}\cup s}_t = h^{\overline{\mathrm{x}}}_t\, h^{\overline{\mathrm{x}}\cup t}_s. \tag{3.5}$$

The following lemma shows that these one-point systems correspond one-to-one to consistent H-systems.

Lemma 3.1. *A system $\boldsymbol{\mathcal{H}} = \{H^{\overline{\mathrm{x}}}_{\mathrm{x}},\ \mathbf{x} \in \mathcal{E} \text{ and } \overline{\mathbf{x}} \subset \mathbf{x}^{\mathrm{c}}\}$ is a consistent H-system if and only if there exists a one-point system $\mathbf{h}$ such that for all $\mathbf{x} \in \mathcal{E}$ and $\overline{\mathbf{x}} \subset \mathbf{x}^{\mathrm{c}}$ we have*

$$H^{\overline{\mathrm{x}}}_{\mathrm{x}} = h^{\overline{\mathrm{x}}}_{t_1}\, h^{\overline{\mathrm{x}}\cup t_1}_{t_2}\ \cdots\ h^{\overline{\mathrm{x}}\cup t_1\cup\cdots\cup t_{n-1}}_{t_n}$$

where $n = |\mathbf{x}|$ and $t_1, \dots, t_n$ is an arbitrary enumeration of elements of the set $\mathbf{x}$. Particularly, for all $t \in \mathbf{Z}^\nu$ and $\overline{\mathbf{x}} \subset \mathbf{Z}^\nu \setminus t$ we have $H_t^{\overline{\mathrm{x}}} = h_t^{\overline{\mathrm{x}}}$.

Proof. 1) Necessity. Let $\boldsymbol{\mathcal{H}} = \{H_{\mathrm{x}}^{\overline{\mathrm{x}}},\ \mathbf{x} \in \mathcal{E} \text{ and } \overline{\mathbf{x}} \subset \mathbf{x}^{\mathrm{c}}\}$ be a consistent H-system and put $h_t^{\overline{\mathrm{x}}} = H_t^{\overline{\mathrm{x}}} \geq 0$ for all $t \in \mathbf{Z}^\nu$ and $\overline{\mathbf{x}} \subset \mathbf{Z}^\nu \setminus t$. Since H-system $\boldsymbol{\mathcal{H}}$ is consistent, using (3.1) we obtain

$$H_{\{s,t\}}^{\overline{\mathrm{x}}} = H_s^{\overline{\mathrm{x}}}\, H_t^{\overline{\mathrm{x}}\cup s} = h_s^{\overline{\mathrm{x}}}\, h_t^{\overline{\mathrm{x}}\cup s}.$$

In the same manner $H_{\{s,t\}}^{\overline{\mathrm{x}}} = h_t^{\overline{\mathrm{x}}}\, h_s^{\overline{\mathrm{x}}\cup t}$, and hence $\mathbf{h}$ is a one-point system. Again using (3.1) we obtain easily

$$H_{\mathrm{x}}^{\overline{\mathrm{x}}} = H_{t_1}^{\overline{\mathrm{x}}}\, H_{\{t_2,\dots,t_n\}}^{\overline{\mathrm{x}}\cup t_1} = H_{t_1}^{\overline{\mathrm{x}}}\, H_{t_2}^{\overline{\mathrm{x}}\cup t_1}\, H_{\{t_3,\dots,t_n\}}^{\overline{\mathrm{x}}\cup t_1\cup t_2} = \dots = h_{t_1}^{\overline{\mathrm{x}}}\, h_{t_2}^{\overline{\mathrm{x}}\cup t_1} \cdots h_{t_n}^{\overline{\mathrm{x}}\cup t_1\cup\dots\cup t_{n-1}}$$

which concludes the proof of the necessity.

2) Sufficiency. Let $\mathbf{h} = \{h_t^{\overline{\mathrm{x}}},\ t \in \mathbf{Z}^\nu \text{ and } \overline{\mathbf{x}} \subset \mathbf{Z}^\nu \setminus t\}$ be a one-point system and for all $\mathbf{x} \in \mathcal{E}$ and $\overline{\mathbf{x}} \subset \mathbf{x}^{\mathrm{c}}$ put

$$H_{\mathrm{x}}^{\overline{\mathrm{x}}} = h_{t_1}^{\overline{\mathrm{x}}}\, h_{t_2}^{\overline{\mathrm{x}}\cup t_1}\ \cdots\ h_{t_n}^{\overline{\mathrm{x}}\cup t_1\cup\dots\cup t_{n-1}} \geq 0. \tag{3.6}$$

First of all let us verify that this definition is correct, i.e., that it does not depend on the enumeration of the set $\mathbf{x}$. To this end, let us fix some enumeration $t_1, \dots, t_n$ and let $\boldsymbol{\varphi} = \{\varphi(1), \dots, \varphi(n)\}$ and $\boldsymbol{\psi} = \{\psi(1), \dots, \psi(n)\}$ be two permutations of the set $\{1, \dots, n\}$. We need to show that

$$h_{t_{\varphi(1)}}^{\overline{\mathrm{x}}} h_{t_{\varphi(2)}}^{\overline{\mathrm{x}}\cup t_{\varphi(1)}} \cdots h_{t_{\varphi(n)}}^{\overline{\mathrm{x}}\cup t_{\varphi(1)}\cup\dots\cup t_{\varphi(n-1)}} = h_{t_{\psi(1)}}^{\overline{x}} h_{t_{\psi(2)}}^{\overline{\mathrm{x}}\cup t_{\psi(1)}} \cdots h_{t_{\psi(n)}}^{\overline{\mathrm{x}}\cup t_{\psi(1)}\cup\dots\cup t_{\psi(n-1)}}. \tag{3.7}$$

It is well known that any permutation of the set $\{1, \dots, n\}$ can be decomposed in a product of transpositions of nearest neighbours, and hence it suffices to consider only the case where $\boldsymbol{\psi} = \boldsymbol{\varphi} \circ (k, k+1)$ with some $k \in \{1, \dots, n-1\}$, i.e., $\boldsymbol{\psi} = \{\varphi(1), \dots, \varphi(k-1), \varphi(k+1), \varphi(k), \varphi(k+2), \dots, \varphi(n)\}$. But in this case (3.7) is reduced to

$$\begin{aligned} & h_{t_{\varphi(k)}}^{\overline{\mathrm{x}}\cup t_{\varphi(1)}\cup\dots\cup t_{\varphi(k-1)}}\, h_{t_{\varphi(k+1)}}^{\overline{\mathrm{x}}\cup t_{\varphi(1)}\cup\dots\cup t_{\varphi(k-1)}\cup t_{\varphi(k)}} \\ &\quad = \quad h_{t_{\varphi(k+1)}}^{\overline{\mathrm{x}}\cup t_{\varphi(1)}\cup\dots\cup t_{\varphi(k-1)}}\, h_{t_{\varphi(k)}}^{\overline{\mathrm{x}}\cup t_{\varphi(1)}\cup\dots\cup t_{\varphi(k-1)}\cup t_{\varphi(k+1)}}, \end{aligned}$$

which is an evident consequence of (3.5). So, $\boldsymbol{\mathcal{H}}$ is an H-system, and it remains to check its consistency. Let us take some $\mathbf{x} = \{t_1, \dots, t_n\} \in \mathcal{E}$ and $\mathbf{y} = \{s_1, \dots, s_m\} \in \mathcal{E}$ such that $\mathbf{x} \cap \mathbf{y} = \emptyset$ and some $\overline{\mathbf{x}} \subset (\mathbf{x} \cup \mathbf{y})^{\mathrm{c}}$. We have $\mathbf{x} \cup \mathbf{y} = \{t_1, \dots, t_n, s_1, \dots, s_m\}$ and hence, using the definition (3.6) of the H-system $\boldsymbol{\mathcal{H}}$, we get

$$\begin{aligned} H_{\mathrm{x}}^{\overline{\mathrm{x}}} &= h_{t_1}^{\overline{\mathrm{x}}}\, h_{t_2}^{\overline{\mathrm{x}}\cup t_1} \cdots h_{t_n}^{\overline{\mathrm{x}}\cup t_1\cup\dots\cup t_{n-1}}, \\ H_{\mathrm{y}}^{\overline{\mathrm{x}}\cup\mathrm{x}} &= h_{s_1}^{\overline{\mathrm{x}}\cup\mathrm{x}}\, h_{s_2}^{\overline{\mathrm{x}}\cup\mathrm{x}\cup s_1} \cdots h_{s_m}^{\overline{\mathrm{x}}\cup\mathrm{x}\cup s_1\cup\dots\cup s_{m-1}}, \\ H_{\mathrm{x}\cup\mathrm{y}}^{\overline{\mathrm{x}}} &= h_{t_1}^{\overline{\mathrm{x}}} \cdots h_{t_n}^{\overline{\mathrm{x}}\cup t_1\cup\dots\cup t_{n-1}}\, h_{s_1}^{\overline{\mathrm{x}}\cup t_1\cup\dots\cup t_n} \cdots h_{s_m}^{\overline{\mathrm{x}}\cup t_1\cup\dots\cup t_n\cup s_1\cup\dots\cup s_{m-1}}, \end{aligned}$$

and hence (3.1) holds. The theorem is proved. □

Now we can formulate the main result of this section.

Theorem 3.2. *A system* $\boldsymbol{\mathcal{Q}} = \{\mathsf{Q}^{\overline{\mathrm{x}}}_{\Lambda},\ \Lambda \in \mathcal{E}$ *and* $\overline{\mathbf{x}} \subset \Lambda^{\mathrm{c}}\}$ *is a vacuum specification if and only if there exists a one-point system* $\mathbf{h}$ *such that for any* $t \in \mathbf{Z}^{\nu}$ *and any* $\overline{\mathbf{x}} \subset \mathbf{Z}^{\nu} \setminus t$ *we have*

$$\mathsf{Q}^{\overline{\mathrm{x}}}_{t}(\emptyset) = \frac{1}{1 + h^{\overline{\mathrm{x}}}_{t}} \quad \textit{and} \quad \mathsf{Q}^{\overline{\mathrm{x}}}_{t}(t) = \frac{h^{\overline{\mathrm{x}}}_{t}}{1 + h^{\overline{\mathrm{x}}}_{t}}. \tag{3.8}$$

Proof. 1) Necessity. Let $\boldsymbol{\mathcal{Q}} = \{\mathsf{Q}^{\overline{\mathrm{x}}}_{\Lambda},\ \Lambda \in \mathcal{E}$ and $\overline{\mathbf{x}} \subset \Lambda^{\mathrm{c}}\}$ be a vacuum specification. Consider the H-system $\boldsymbol{\mathcal{H}}$ corresponding to $\boldsymbol{\mathcal{Q}}$, and let $\mathbf{h}$ be the one-point system corresponding to $\boldsymbol{\mathcal{H}}$. This $\mathbf{h}$ is the desired one-point system, since relations (3.8) follow immediately from (3.2). The necessity is proved.

2) Sufficiency. Let $\mathbf{h} = \{h^{\overline{\mathrm{x}}}_{t},\ t \in \mathbf{Z}^{\nu}$ and $\overline{\mathbf{x}} \subset \mathbf{Z}^{\nu} \setminus t\}$ be a one-point system. Consider the H-system $\boldsymbol{\mathcal{H}}$ corresponding to $\mathbf{h}$, and let $\boldsymbol{\mathcal{Q}}$ be the vacuum specification corresponding to $\boldsymbol{\mathcal{H}}$. Again (3.8) follows from (3.2), and so the theorem is proved. □

Let us note that this theorem answers Dobrushin's problem, by showing when a system of one-point distributions with boundary conditions is a subsystem of some specification. In fact, a necessary and sufficient condition for that is condition (3.5) which can be rewritten, using the obvious relation

$$h^{\overline{\mathrm{x}}}_{t} = \frac{\mathsf{Q}^{\overline{\mathrm{x}}}_{t}(t)}{\mathsf{Q}^{\overline{\mathrm{x}}}_{t}(\emptyset)}, \tag{3.9}$$

as follows:

$$\mathsf{Q}^{\overline{\mathrm{x}}}_{s}(s)\,\mathsf{Q}^{\overline{\mathrm{x}}\cup s}_{t}(t)\,\mathsf{Q}^{\overline{\mathrm{x}}}_{t}(\emptyset)\,\mathsf{Q}^{\overline{\mathrm{x}}\cup t}_{s}(\emptyset) = \mathsf{Q}^{\overline{\mathrm{x}}}_{t}(t)\,\mathsf{Q}^{\overline{\mathrm{x}}\cup t}_{s}(s)\,\mathsf{Q}^{\overline{\mathrm{x}}}_{s}(\emptyset)\,\mathsf{Q}^{\overline{\mathrm{x}}\cup s}_{t}(\emptyset).$$

Remark 3.2. Let $\mathbf{h}$ be a one-point system. For all $t \in \mathbf{Z}^{\nu}$ and $\overline{\mathbf{x}} \subset \mathbf{Z}^{\nu} \setminus t$ denote $u^{\overline{\mathrm{x}}}(t) = -\ln h^{\overline{\mathrm{x}}}_{t}$. The system $\mathbf{u} = \{u^{\overline{\mathrm{x}}}(t), t \in \mathbf{Z}^{\nu}$ and $\overline{\mathbf{x}} \subset \mathbf{Z}^{\nu} \setminus t\}$ is a one-point Hamiltonian, which in general does not have an explicit form in terms of some potential.

Finally, let us give an example of one-point system, based on a simple idea which will be used in Section 4 for constructing non-Gibbsian random fields.

Example 3.1. Let $\mathbf{h} = \{h^{\overline{\mathrm{x}}}_{t},\ t \in \mathbf{Z}^{\nu}$ and $\overline{\mathbf{x}} \subset \mathbf{Z}^{\nu} \setminus t\}$ be a non-negative system such that $h^{\overline{\mathrm{x}}_1}_{t} = h^{\overline{\mathrm{x}}_2}_{t}$ if $\overline{\mathbf{x}}_1 = \overline{\mathbf{x}}_2$ up to a finite number of lattice points. Then $\mathbf{h}$ is clearly a one-point system.

3.3. Generalizations to the case of arbitrary finite state space

All the results obtained above can be straightforwardly generalized to the case of an arbitrary finite state space $\mathcal{X}$. As always we suppose that there is some fixed element $\theta \in \mathcal{X}$ which is called vacuum and we use $\mathcal{X}^*$ to denote $\mathcal{X} \setminus \theta$.

Definition 3.3. A system $\boldsymbol{\mathcal{H}} = \{H_{\mathrm{x}}^{\overline{\mathrm{x}}},\ \mathbf{x} \in \mathcal{X}^{*I},\ I \in \mathcal{E},\ \overline{\mathbf{x}} \in \mathcal{X}^{*S},\ S \subset I^{\mathrm{c}}\}$ such that $H_{\mathrm{x}}^{\overline{\mathrm{x}}} \geq 0$ for all $\mathbf{x} \in \mathcal{X}^{*I}$, $I \in \mathcal{E}$, $\overline{\mathbf{x}} \in \mathcal{X}^{*S}$, $S \subset I^{\mathrm{c}}$, and $H_{\emptyset}^{\overline{\mathrm{x}}} = 1$ for all $\overline{\mathbf{x}} \in \mathcal{X}^{*S},\ S \subset \mathbf{Z}^{\nu}$, is called *$H$-system*.

An H-system $\boldsymbol{\mathcal{H}}$ is called *consistent* if it satisfies the following condition: for any $\mathbf{x} \in \mathcal{X}^{*I}$, $I \in \mathcal{E}$, $\mathbf{y} \in \mathcal{X}^{*J}$, $J \in \mathcal{E}$ such that $I \cap J = \emptyset$ and any $\overline{\mathbf{x}} \in \mathcal{X}^{*S}$, $S \subset (I \cup J)^{\mathrm{c}}$ we have

$$H_{\mathrm{x}\oplus\mathrm{y}}^{\overline{\mathrm{x}}} = H_{\mathrm{x}}^{\overline{\mathrm{x}}}\, H_{\mathrm{y}}^{\overline{\mathrm{x}}\oplus\mathrm{x}}.$$

Theorem 3.3. *A system $\boldsymbol{\mathcal{Q}} = \{\mathsf{Q}_{\Lambda}^{\overline{\mathrm{x}}},\ \Lambda \in \mathcal{E},\ \overline{\mathbf{x}} \in \mathcal{X}^{*S},\ S \subset \Lambda^{\mathrm{c}}\}$ is a vacuum specification if and only if there exists a consistent H-system $\boldsymbol{\mathcal{H}}$ such that for any $\Lambda \in \mathcal{E}$ and $\overline{\mathbf{x}} \in \mathcal{X}^{*S}$, $S \subset \Lambda^{\mathrm{c}}$, we have*

$$\mathsf{Q}_{\Lambda}^{\overline{\mathrm{x}}}(\mathbf{x}) = \frac{H_{\mathrm{x}}^{\overline{\mathrm{x}}}}{\sum_{\mathrm{y}\in\mathcal{X}^{\Lambda}} H_{\mathrm{y}}^{\overline{\mathrm{x}}}}, \quad \mathbf{x} \in \mathcal{X}^{*I},\ I \subset \Lambda.$$

Definition 3.4. A system

$$\mathbf{h} = \{h_t^{\overline{\mathrm{x}}}(x),\ t \in \mathbf{Z}^{\nu},\ x \in \mathcal{X}^*,\ \overline{\mathbf{x}} \in \mathcal{X}^{*S},\ S \subset \mathbf{Z}^{\nu} \setminus t\}$$

is called *one-point system* if for all $t \in \mathbf{Z}^{\nu}$, $x \in \mathcal{X}^*$ and $\overline{\mathbf{x}} \in \mathcal{X}^{*S}$, $S \subset \mathbf{Z}^{\nu} \setminus t$ we have $h_t^{\overline{\mathrm{x}}}(x) \geq 0$, and for all $s, t \in \mathbf{Z}^{\nu}$, $x, y \in \mathcal{X}^*$ and $\overline{\mathbf{x}} \in \mathcal{X}^{*S}$, $S \subset \mathbf{Z}^{\nu} \setminus \{s,t\}$ we have

$$h_s^{\overline{\mathrm{x}}}(y)\, h_t^{\overline{\mathrm{x}}\oplus y_s}(x) = h_t^{\overline{\mathrm{x}}}(x)\, h_s^{\overline{\mathrm{x}}\oplus x_t}(y).$$

Lemma 3.2. *A system $\boldsymbol{\mathcal{H}} = \{H_{\mathrm{x}}^{\overline{\mathrm{x}}},\ \mathbf{x} \in \mathcal{X}^{*I},\ I \in \mathcal{E},\ \overline{\mathbf{x}} \in \mathcal{X}^{*S},\ S \subset I^{\mathrm{c}}\}$ is a consistent H-system if and only if there exists a one-point system $\mathbf{h}$ such that for all $\mathbf{x} \in \mathcal{X}^{*I}$, $I \in \mathcal{E}$ and $\overline{\mathbf{x}} \in \mathcal{X}^{*S}$, $S \subset I^{\mathrm{c}}$ we have*

$$H_{\mathrm{x}}^{\overline{\mathrm{x}}} = h_{t_1}^{\overline{\mathrm{x}}}(x_{t_1})\, h_{t_2}^{\overline{\mathrm{x}}\oplus x_{t_1}}(x_{t_2}) \cdots h_{t_n}^{\overline{\mathrm{x}}\oplus x_{t_1}\oplus\cdots\oplus x_{t_{n-1}}}(x_{t_n})$$

where $n = |I|$ and $t_1, \ldots, t_n$ is an arbitrary enumeration of elements of the set I. Particularly, for all $t \in \mathbf{Z}^{\nu}$, $x \in \mathcal{X}^*$ and $\overline{\mathbf{x}} \in \mathcal{X}^{*S}$, $S \subset \mathbf{Z}^{\nu} \setminus t$, we have $H_{x_t}^{\overline{\mathrm{x}}} = h_t^{\overline{\mathrm{x}}}(x)$.

Theorem 3.4. *A system $\boldsymbol{\mathcal{Q}} = \{\mathsf{Q}_{\Lambda}^{\overline{\mathrm{x}}},\ \Lambda \in \mathcal{E},\ \overline{\mathbf{x}} \in \mathcal{X}^{*S},\ S \subset \Lambda^{\mathrm{c}}\}$ is a vacuum specification if and only if there exists a one-point system $\mathbf{h}$ such that for any $t \in \mathbf{Z}^{\nu}$ and any $\overline{\mathbf{x}} \in \mathcal{X}^{*S}$, $S \subset \mathbf{Z}^{\nu} \setminus t$, we have*

$$\mathsf{Q}_t^{\overline{\mathrm{x}}}(\theta) = \frac{1}{1 + \sum_{y\in\mathcal{X}^*} h_t^{\overline{\mathrm{x}}}(y)} \quad \text{and} \quad \mathsf{Q}_t^{\overline{\mathrm{x}}}(x_t) = \frac{h_t^{\overline{\mathrm{x}}}(x)}{1 + \sum_{y\in\mathcal{X}^*} h_t^{\overline{\mathrm{x}}}(y)},\ x \in \mathcal{X}^*.$$

Finally relation (3.9) becomes

$$h_t^{\overline{\mathrm{x}}}(x) = \frac{\mathsf{Q}_t^{\overline{\mathrm{x}}}(x_t)}{\mathsf{Q}_t^{\overline{\mathrm{x}}}(\theta)}\ , \quad x \in \mathcal{X}^*.$$

4. Applications

In this section we propose some applications of one-point systems.

4.1. Description of random fields by means of one-point systems

As we have already mentioned in Section 2, quasilocality of a specification guarantees existence of random fields having this specification as a conditional distribution. Using the results of Section 3, we can describe a vacuum specification by means of a one-point system. We have the following

Theorem 4.1. *A vacuum specification $\boldsymbol{\mathcal{Q}}$ is quasilocal if and only if the corresponding one-point system $\mathbf{h} = \{h_t^{\overline{\mathrm{x}}},\ t \in \mathbf{Z}^\nu$ and $\overline{\mathbf{x}} \subset \mathbf{Z}^\nu \setminus t\}$ is quasilocal as a function of $\overline{\mathbf{x}}$ for all $t \in \mathbf{Z}^\nu$, i.e., if we have*

$$\sup_{\overline{\mathrm{x}} \subset \mathrm{Z}^\nu \setminus t} |h_t^{\overline{\mathrm{x}}_I} - h_t^{\overline{\mathrm{x}}}| \xrightarrow[I \uparrow \mathrm{Z}^\nu]{} 0, \quad t \in \mathbf{Z}^\nu.$$

Proof. Recall that quasilocality means continuity with respect to the topology $\mathcal{T}^{\mathrm{Z}^\nu \setminus t}$. If $\boldsymbol{\mathcal{Q}}$ is quasilocal, then using (3.9) we get clearly the quasilocality of $\mathbf{h}$. Now, if $\mathbf{h}$ is quasilocal, then by (3.6) we get the quasilocality of the corresponding H-system $\boldsymbol{\mathcal{H}}$, and then by (3.2) the quasilocality of $\boldsymbol{\mathcal{Q}}$. □

Now we can state a theorem about existence and uniqueness of a random field with a given one-point system.

Theorem 4.2. *Let $\mathbf{h} = \{h_t^{\overline{\mathrm{x}}},\ t \in \mathbf{Z}^\nu$ and $\overline{\mathbf{x}} \subset \mathbf{Z}^\nu \setminus t\}$ be a one-point system.*

1) *If $\mathbf{h}$ is quasilocal, then there exists a random field P having $\mathbf{h}$ as a one-point system.*

2) *If, moreover, $\mathbf{h}$ satisfies the condition*

$$\sup_{t \in \mathrm{Z}^\nu} \sum_{s \in \mathrm{Z}^\nu \setminus t} \sup_{\overline{\mathrm{x}} \subset \mathrm{Z}^\nu \setminus \{s,t\}} \frac{|h_t^{\overline{\mathrm{x}}} - h_t^{\overline{\mathrm{x}} \cup s}|}{(1 + h_t^{\overline{\mathrm{x}}})\,(1 + h_t^{\overline{\mathrm{x}} \cup s})} < 1, \tag{4.1}$$

then the random field P having $\mathbf{h}$ as a one-point system is unique.

The condition (4.1) is nothing but Dobrushin's uniqueness condition rewritten in terms of one-point systems. It is obtained by replacing $\boldsymbol{\mathcal{Q}}$ by its values (expressed in terms of $\mathbf{h}$) in (2.4).

4.2. Gibbsian one-point systems

The problem of characterization of Gibbsian random fields was considered in many works (see, for example, [1, 9, 11, 15]). One of the most interesting aspects of this problem is the description of the class of specifications which are Gibbsian with a potential satisfying some given conditions. For example, let us mention the following result.

Theorem 4.3. *A specification $\boldsymbol{\mathcal{Q}}$ is Gibbsian with uniformly convergent potential if and only if it is quasilocal and strictly positive.*

Since vacuum specifications can be described by means of one-point systems, the above mentioned problem can be reduced to a similar problem for one-point systems.

Definition 4.1. A one-point system $\mathbf{h} = \{h_t^{\overline{\mathrm{x}}},\ t \in \mathbf{Z}^\nu \text{ and } \overline{\mathbf{x}} \subset \mathbf{Z}^\nu \setminus t\}$ is called Gibbsian with potential satisfying some given condition, if the corresponding specification $\boldsymbol{\mathcal{Q}}$ is Gibbsian with a potential satisfying this condition.

Combining Theorems 4.1 and 4.3 we obtain the following result.

Theorem 4.4. *A one-point system* $\mathbf{h}$ *is Gibbsian with uniformly convergent potential if and only if it is quasilocal and strictly positive.*

Note that this result can be also proved directly, taking into account that the uniform convergence of potential is equivalent to the quasilocality of corresponding Hamiltonians, and hence to the quasilocality and strict positivity of $\mathbf{h}$.

Let us now describe a wider class of one-point systems which are Gibbsian with (not necessarily uniformly) convergent vacuum potentials. For convenience we call such one-point systems Gibbsian.

Theorem 4.5. *A one-point system* $\mathbf{h} = \{h_t^{\overline{\mathrm{x}}},\ t \in \mathbf{Z}^\nu \text{ and } \overline{\mathbf{x}} \subset \mathbf{Z}^\nu \setminus t\}$ *is Gibbsian if and only if the following two conditions are satisfied:*

(h1) *for all $t \in \mathbf{Z}^\nu$ and $\overline{\mathbf{x}} \subset \mathbf{Z}^\nu \setminus t$ we have $\lim\limits_{I \uparrow \mathrm{Z}^\nu} h_t^{\overline{\mathrm{x}}_I} = h_t^{\overline{\mathrm{x}}}$,*

(h2) *for all $t \in \mathbf{Z}^\nu$ and $\overline{\mathbf{x}} \subset \mathbf{Z}^\nu \setminus t$ we have $h_t^{\overline{\mathrm{x}}} = 0$ if there exists $T \in \mathcal{E}$ such that $h_t^{\overline{\mathrm{x}}_T} = 0$.*

Proof. 1) Necessity. We suppose that the one-point system $\mathbf{h}$ is Gibbsian, i.e., that for all $t \in \mathbf{Z}^\nu$ and $\overline{\mathbf{x}} \subset \mathbf{Z}^\nu \setminus t$ we have $h_t^{\overline{\mathrm{x}}} = \exp(-u^{\overline{\mathrm{x}}}(t))$ with

$$u^{\overline{\mathrm{x}}}(t) = \sum_{V \in \mathcal{E}\,:\, V \subset \overline{\mathrm{x}}} \Phi(V \cup t)$$

where $\mathbf{\Phi}$ is some convergent potential. We need to check conditions (h1) and (h2). The first condition follows obviously from the fact that the potential $\mathbf{\Phi}$ is convergent. To check the second one, let us take some $t \in \mathbf{Z}^\nu$ and $\overline{\mathbf{x}} \subset \mathbf{Z}^\nu \setminus t$ and suppose that there exists $T \in \mathcal{E}$ such that $h_t^{\overline{\mathbf{x}}_T} = 0$. We need to show that $h_t^{\overline{\mathbf{x}}} = 0$. We have

$$u^{\overline{\mathbf{x}}_T}(t) = -\ln(h_t^{\overline{\mathbf{x}}_T}) = +\infty = \sum_{V \in \mathcal{E}\,:\, V \subset \overline{\mathbf{x}}_T} \Phi(V \cup t) \;=\; \sum_{V \subset \overline{\mathbf{x}}_T} \Phi(V \cup t).$$

But the last sum contains a finite number of summands and hence at least one of them is equal to $+\infty$. This implies that for any $I \in \mathcal{E}$ such that $I \supset T$ we have $u^{\overline{\mathbf{x}}_I}(t) = +\infty$, and since $\mathbf{\Phi}$ is convergent we have also $u^{\overline{\mathbf{x}}}(t) = +\infty$, and hence $h_t^{\overline{\mathbf{x}}} = \exp(-u^{\overline{\mathbf{x}}}(t)) = 0$ which concludes the proof of the necessity.

2) Sufficiency. We suppose that the one-point system $\mathbf{h}$ satisfies conditions (h1) and (h2) and that $\mathbf{u}$ is the corresponding one-point Hamiltonian. Let us consider the potential $\mathbf{\Phi}$ defined by

$$\Phi(J) = \begin{cases} +\infty, & \text{if } \forall\ \ell \in J \text{ we have } u^{J\setminus\ell}(\ell) = +\infty, \\ \displaystyle\sum_{R \subset J\setminus\ell} (-1)^{|(J\setminus\ell)\setminus R|} u^R(\ell), & \text{if } \exists\ \ell \in J \text{ such that } u^{J\setminus\ell}(\ell) \in \mathbf{R}. \end{cases}$$

Note that the last sum is well defined since the number of summands is finite and by (h2) all the summands are finite. We can also show that this definition is correct, i.e., that if $u^{J\setminus\ell}(\ell), u^{J\setminus s}(s) \in \mathbf{R}$ then

$$\sum_{R \subset J\setminus\ell} (-1)^{|(J\setminus\ell)\setminus R|} u^R(\ell) = \sum_{R \subset J\setminus s} (-1)^{|(J\setminus s)\setminus R|} u^R(s).$$

Indeed, we have

$$\begin{aligned} \sum_{R \subset J\setminus\ell} (-1)^{|(J\setminus\ell)\setminus R|} u^R(\ell) \;&=\; \sum_{R \subset J\setminus\{\ell,s\}} (-1)^{|(J\setminus\ell)\setminus R|} u^R(\ell) \\ &\quad + \sum_{R \subset J\setminus\{\ell,s\}} (-1)^{|(J\setminus\ell)\setminus(R\cup s)|} u^{R\cup s}(\ell) \\ &=\; \sum_{R \subset J\setminus\{\ell,s\}} (-1)^{|(J\setminus\ell)\setminus R|} (u^R(\ell) - u^{R\cup s}(\ell)), \end{aligned}$$

and in the same manner

$$\sum_{R \subset J\setminus s} (-1)^{|(J\setminus s)\setminus R|} u^R(s) = \sum_{R \subset J\setminus\{\ell,s\}} (-1)^{|(J\setminus s)\setminus R|} (u^R(s) - u^{R\cup\ell}(s)).$$

Since all the terms in these sums are finite, using (3.5) we see that the sums are term by term equal.

It remains to check that the potential $\mathbf{\Phi}$ indeed corresponds to our one-point system $\mathbf{h}$, i.e., that

$$u^{\overline{\mathrm{x}}}(t) = \sum_{V\in\mathcal{E}\,:\,V\subset\overline{\mathrm{x}}} \Phi(V\cup t) \tag{4.2}$$

for all $t \in \mathbf{Z}^\nu$ and $\overline{\mathbf{x}} \subset \mathbf{Z}^\nu \setminus t$. Since condition (h1) holds, it is sufficient to verify this relation only in the case when $\overline{\mathbf{x}} \in \mathcal{E}$. Let us at first suppose that the left-hand side of (4.2) is finite. In this case by (h1) we have $u^V(t) < +\infty$ for all $V \subset \overline{\mathbf{x}}$. Then by definition of $\mathbf{\Phi}$ we have

$$\Phi(V\cup t) = \sum_{R\subset V} (-1)^{|V\setminus R|} u^R(t),$$

and hence the right-hand side of (4.2) equals

$$\sum_{V\subset\overline{\mathrm{x}}} \sum_{R\subset V} (-1)^{|V\setminus R|} u^R(t) = u^{\overline{\mathrm{x}}}(t).$$

Now let us consider the case when the left-hand side of (4.2) is infinite, i.e., when $u^{\overline{\mathrm{x}}}(t) = +\infty$. We need to show that the right-hand side of (4.2) is also infinite. Two cases are possible:

- We have $u^{\emptyset}(t) = +\infty$. In this case by the definition of $\mathbf{\Phi}$ we obtain $\Phi(t) = +\infty$, and since $\Phi(t)$ is one of the summands in the right-hand side of (4.2), the latter is infinite.
- We have $u^{\emptyset}(t) \in \mathbf{R}$. In this case clearly there exists a $V \subset \overline{\mathbf{x}}$ such that $V \neq \emptyset$, $u^V(t) = +\infty$, and for all $\ell \in V$ we have $u^{V\setminus\ell}(t) \in \mathbf{R}$. Hence, for all $\ell \in V$ we can write

  $$u^{V\setminus\ell}(t) + u^{(V\setminus\ell)\cup t}(\ell) = u^{V\setminus\ell}(\ell) + u^V(t) = u^{V\setminus\ell}(\ell) + (+\infty) = +\infty.$$

  But $u^{V\setminus\ell}(t) \in \mathbf{R}$, and hence we have $u^{(V\setminus\ell)\cup t}(\ell) = u^{(V\cup t)\setminus\ell}(\ell) = +\infty$ for all $\ell \in V$. Clearly we have also $u^{(V\cup t)\setminus t}(t) = u^V(t) = +\infty$. Thus, by definition of $\mathbf{\Phi}$ we have $\Phi(V\cup t) = +\infty$, and hence the right-hand side of (4.2) is infinite.

□

4.3. Non-Gibbsian one-point systems and random fields

In Theorem 4.5 we have seen necessary and sufficient conditions for a one-point system to be Gibbsian. Now we will describe a simple scheme for constructing non-Gibbsian one-point systems and non-Gibbsian random fields. We need the following

Lemma 4.1. *Let* $\mathbf{h} = \{h_t^{\overline{\mathrm{x}}},\ t \in \mathbf{Z}^\nu \text{ and } \overline{\mathbf{x}} \subset \mathbf{Z}^\nu \setminus t\}$ *be a one-point system and* $\mathcal{R} = \{R(\overline{\mathbf{x}}),\ \overline{\mathbf{x}} \subset \mathbf{Z}^\nu\}$ *be a real-valued strictly positive function such that* $R(\overline{\mathbf{x}}_1) = R(\overline{\mathbf{x}}_2)$ *if* $\overline{\mathbf{x}}_1 = \overline{\mathbf{x}}_2$ *up to a finite number of lattice points. Then the system*

$$\mathbf{h}_{\mathcal{R}} = \{(h_t^{\overline{\mathrm{x}}})^{R(\overline{\mathrm{x}})},\ t \in \mathbf{Z}^\nu \text{ and } x \subset \mathbf{Z}^\nu \setminus t\}$$

is also a one-point system.

Proof. For any $s, t \in \mathcal{E}$ and $\overline{\mathbf{x}} \subset \mathbf{Z}^\nu \setminus \{s,t\}$ we can write

$$\begin{aligned}(h_t^{\overline{\mathrm{x}}})^{R(\overline{\mathrm{x}})}\,(h_s^{\overline{\mathrm{x}}\cup t})^{R(\overline{x}\cup t)} &= (h_t^{\overline{\mathrm{x}}})^{R(\overline{\mathbf{x}})}\,(h_s^{\overline{\mathrm{x}}\cup t})^{R(\overline{\mathrm{x}})} = (h_t^{\overline{\mathrm{x}}}\,h_s^{\overline{\mathrm{x}}\cup t})^{R(\overline{\mathrm{x}})} \\ &= (h_s^{\overline{\mathrm{x}}}\,h_t^{\overline{\mathrm{x}}\cup s})^{R(\overline{\mathrm{x}})} = (h_s^{\overline{\mathrm{x}}})^{R(\overline{\mathrm{x}})}\,(h_t^{\overline{\mathrm{x}}\cup s})^{R(\overline{\mathrm{x}}\cup s)}\end{aligned}$$

which concludes the proof. □

Remark 4.1. We require the function $\mathcal{R}$ to be real-valued and strictly positive only in order for the system $\mathbf{h}_{\mathcal{R}}$ to be well-defined. But the lemma holds under less restrictive conditions. For example, if the system $\mathbf{h}$ is strictly positive, which is equivalent to say that the corresponding Hamiltonian $\boldsymbol{\mathcal{U}}$ is finite, we can consider $\mathcal{R}$ to be any real-valued function, and if the system $\mathbf{h}$ is less or equal than 1 (respectively greater or equal than 1), which is equivalent to say that the Hamiltonian $\boldsymbol{\mathcal{U}}$ is strictly positive (respectively strictly negative), we can allow $\mathcal{R}$ to take the value $+\infty$ (respectively $-\infty$). Here and in the sequel we admit that $1^{\pm\infty} = 0^0 = 1$, or equivalently $(\pm\infty)\cdot 0 = 0\cdot(\pm\infty) = 0$.

Proposition 4.1. *Let* $\mathbf{h} = \{h_t^{\overline{\mathrm{x}}},\ t \in \mathbf{Z}^\nu \text{ and } \overline{\mathbf{x}} \subset \mathbf{Z}^\nu \setminus t\}$ *be a strictly positive Gibbsian one-point system, and let* $\mathcal{R} = \{R(\overline{\mathbf{x}}),\ \overline{\mathbf{x}} \subset \mathbf{Z}^\nu\}$ *be a real-valued function such that* $R(\overline{\mathbf{x}}_1) = R(\overline{\mathbf{x}}_2)$ *if* $\overline{\mathbf{x}}_1 = \overline{\mathbf{x}}_2$ *up to a finite number of lattice points. The following three conditions are equivalent:*

1) *the one-point system* $\mathbf{h}_{\mathcal{R}}$ *is non-Gibbsian;*

2) *there exists at least one pair* $t \in \mathbf{Z}^\nu$ *and* $\overline{\mathbf{x}} \subset \mathbf{Z}^\nu \setminus t$ *such that* $R(\overline{\mathbf{x}}) \neq R(\emptyset)$ *and that* $h_t^{\overline{\mathrm{x}}} \neq 1$*;*

3) *the function* $\mathcal{R} = \{R(\overline{\mathbf{x}}),\ \overline{\mathbf{x}} \subset \mathbf{Z}^\nu\}$ *is not constant on the set* $\mathfrak{N}$ *defined by* $\mathfrak{N} = \{\overline{\mathbf{x}} \subset \mathbf{Z}^\nu \mid \exists\, t \in \mathbf{Z}^\nu \text{ such that } t \notin \overline{\mathbf{x}} \text{ and } h_t^{\overline{\mathrm{x}}} \neq 1\}$.

Proof. First let us show that 2) ⇒ 1). Since $\mathbf{h}$ is Gibbsian, it has the form

$$\mathbf{h} = \{\exp(-u^{\overline{\mathrm{x}}}(t)),\ t \in \mathbf{Z}^\nu \text{ and } \overline{\mathbf{x}} \subset \mathbf{Z}^\nu \setminus t\}$$

where the one-point Hamiltonian $\mathbf{u}$ is finite and given by some convergent potential $\boldsymbol{\Phi} = \{\Phi(J),\ J \in \mathcal{E} \setminus \{\emptyset\}\}$. Hence

$$\mathbf{h}_{\mathcal{R}} = \{\exp(-u^{\overline{\mathrm{x}}}(t)\,R(\overline{\mathbf{x}})),\ t \in \mathbf{Z}^\nu \text{ and } \overline{\mathbf{x}} \subset \mathbf{Z}^\nu \setminus t\}.$$

We need to show that the specification determined by $\mathbf{h}_{\mathcal{R}}$ is non-Gibbsian, i.e., that there is no convergent potential $\widetilde{\mathbf{\Phi}} = \{\widetilde{\Phi}(J),\ J \in \mathcal{E} \setminus \{\emptyset\}\}$ such that

$$u^{\overline{\mathrm{x}}}(t)\, R(\overline{\mathbf{x}}) \ = \sum_{\widetilde{J} \in \mathcal{E}\,:\, \widetilde{J} \subset \overline{\mathbf{x}}} \widetilde{\Phi}(t \cup \widetilde{J}), \quad t \in \mathbf{Z}^{\nu},\ \overline{\mathbf{x}} \subset \mathbf{Z}^{\nu} \setminus t. \tag{4.3}$$

Suppose that the contrary is true, i.e., that (4.3) holds. In this case we would clearly have

$$u^{\overline{\mathrm{x}}}(t)\, R(\overline{\mathbf{x}}) = \lim_{I \uparrow \mathbf{Z}^{\nu}} u^{\overline{\mathrm{x}}_I}(t)\, R(\overline{\mathbf{x}}_I) = R(\emptyset) \lim_{I \uparrow \mathbf{Z}^{\nu}} u^{\overline{\mathrm{x}}_I}(t) = R(\emptyset)\, u^{\overline{\mathrm{x}}}(t)$$

for any $t \in \mathbf{Z}^{\nu}$ and $\overline{\mathbf{x}} \subset \mathbf{Z}^{\nu} \setminus t$. But the last relation contradicts condition 2).

The implication 1) $\Rightarrow$ 2) is easy to see, since if 2) does not hold, then for all $t \in \mathbf{Z}^{\nu}$ and $\overline{\mathbf{x}} \subset \mathbf{Z}^{\nu} \setminus t$ we have $R(\overline{\mathbf{x}}) = R(\emptyset)$ or $h_t^{\overline{\mathrm{x}}} = 1$, and hence $(h_t^{\overline{\mathrm{x}}})^{R(\overline{\mathrm{x}})} = (h_t^{\overline{\mathrm{x}}})^{R(\emptyset)}$. But the last specification is clearly Gibbsian, since got by multiplying a Gibbsian potential by a constant.

The implication 3) $\Rightarrow$ 2) is evident.

For the proof of 2) $\Rightarrow$ 3) note that since there exists a pair $t \in \mathbf{Z}^{\nu}$ and $\overline{\mathbf{x}} \subset \mathbf{Z}^{\nu} \setminus t$ such that $R(\overline{\mathbf{x}}) \neq R(\emptyset)$ and $h_t^{\overline{\mathrm{x}}} \neq 1$, clearly we have $\overline{\mathbf{x}} \in \mathfrak{N}$. On the other hand, $\mathfrak{N}$ contains at least one finite set $\mathbf{y}$, since otherwise we would have $h_s^{\mathrm{x}} = 1$, and hence $u^{\overline{x}}(s) = 0$, for all $\mathbf{x} \in \mathcal{E}$ and $s \in \mathbf{x}^{\mathrm{c}}$, which is possible if and only if $\mathbf{\Phi} \equiv 0$ on $\mathcal{E} \setminus \{\emptyset\}$ which contradicts $h_t^{\overline{\mathrm{x}}} \neq 0$. So we will have $R(\overline{\mathbf{x}}) \neq R(\emptyset) = R(\mathbf{y})$, which shows that $\mathcal{R}$ is not constant on $\mathfrak{N}$. □

Remark 4.2. 1) Clearly, as in Lemma 4.1 we can allow $\mathcal{R}$ to take the value $+\infty$ or $-\infty$ under suitable conditions.

2) If $\mathbf{h}_{\mathcal{R}}$ is a one-point system of some random field P, then this random field is non-Gibbsian (i.e., does not have any Gibbsian one-point system) if and only if the function $\mathcal{R} = \{R(\overline{\mathbf{x}}),\ \overline{\mathbf{x}} \subset \mathbf{Z}^{\nu}\}$ is not P-almost surely constant on $\mathfrak{N}$.

Proposition 4.1 allows to construct non-Gibbsian one-point systems and, if the existence is known, non-Gibbsian random fields. Note that non-Gibbsian random fields constructed this way are not quasilocal, in the sense that they do not have any quasilocal conditional distribution.

As an application let us give an example of a non-Gibbsian one-point system. For any $p \in (0,1)$ let us denote by $\mathfrak{I}^p$ the set of all $\overline{\mathbf{x}} \subset \mathbf{Z}^{\nu}$ such that the following limit exists:

$$\lim_{I \uparrow \mathbf{Z}^{\nu}} \frac{|\overline{\mathbf{x}}_I|}{|I|} = p(\overline{\mathbf{x}}) = p,$$

and put $\overline{\mathfrak{I}} = \mathcal{X}^{\mathbf{Z}^{\nu}} \setminus \Big(\bigcup_{p \in (0,1)} \mathfrak{I}^p \Big)$ and

$$h_t^{\overline{\mathrm{x}}} = \begin{cases} 0, & \text{if } \overline{\mathbf{x}} \in \overline{\mathfrak{I}}, \\ \dfrac{p(\overline{\mathbf{x}})}{1 - p(\overline{\mathbf{x}})}, & \text{otherwise.} \end{cases}$$

Clearly this is a one-point system, since

$$h_t^{\overline{\mathrm{x}}}\, h_s^{\overline{\mathrm{x}}\cup t} = \frac{p(\overline{\mathbf{x}})}{1-p(\overline{\mathbf{x}})}\,\frac{p(\overline{\mathbf{x}}\cup t)}{1-p(\overline{\mathbf{x}}\cup t)} = \frac{p(\overline{\mathbf{x}})}{1-p(\overline{\mathbf{x}})}\,\frac{p(\overline{\mathbf{x}}\cup s)}{1-p(\overline{\mathbf{x}}\cup s)} = h_s^{\overline{\mathrm{x}}}\, h_t^{\overline{\mathrm{x}}\cup s}.$$

Now, let us remark that the system $\mathbf{h}$ can be rewritten in the form $h_t^{\overline{\mathrm{x}}} = (\widetilde{h}_t^{\overline{\mathrm{x}}})^{R(\overline{\mathrm{x}})}$ where $\widetilde{h}_t^{\overline{\mathrm{x}}} = e^{-1}$ is the Gibbsian one-point system corresponding to the potential $\mathbf{\Phi} = \{\Phi(J) = \mathbf{1}_{\{|J|=1\}},\ J \in \mathcal{E}\setminus\{\emptyset\}\}$, and the function $\mathcal{R}$ is given by

$$R(\overline{\mathbf{x}}) = \begin{cases} +\infty, & \text{if } \overline{\mathbf{x}} \in \overline{\mathfrak{I}}, \\ -\ln\dfrac{p(\overline{\mathbf{x}})}{1-p(\overline{\mathbf{x}})}, & \text{otherwise.} \end{cases}$$

Clearly condition 3) of Proposition 4.1 is satisfied, and hence $\mathbf{h}$ is non-Gibbsian.

Note that this one-point system corresponds to a well known example of a non-Gibbsian random field. In fact, for any $p \in (0,1)$ let us consider the Bernoulli random field B^p with parameter p. Since B^p is concentrated on $\mathfrak{I}^p$, we have B^p-almost surely the equality

$$h_t^{\overline{\mathrm{x}}} = \frac{p}{1-p} = \frac{\mathsf{B}_t^p(1)}{\mathsf{B}_t^p(0)}.$$

Hence, taking into account (3.9) we see that all the fields B^p, $p \in (0,1)$, have $\mathbf{h}$ as a one-point system. So, any mixture of these fields also has $\mathbf{h}$ as a one-point system. Let us consider such a non-trivial ($\mathsf{P} \neq \mathsf{B}^p$, $p \in (0,1)$) mixture P and show that it is non-Gibbsian. Indeed, in this case the set $\mathfrak{N}$ is equal to $\{\overline{\mathbf{x}} \subsetneq \mathbf{Z}^\nu\}$, and the function $\mathcal{R}$ is not P-almost surely constant on $\mathfrak{N}$, since it takes different values on different $\mathfrak{I}^p$-s. So, according to part 2) of Remark 4.2 the random field P is non-Gibbsian. Note that this fact can also be obtained using a general result from [8, Section 4.5.1].

To conclude, let us note that by direct calculation one can easily obtain the explicit form of the specification $\boldsymbol{\mathcal{Q}}$ determined by $\mathbf{h}$:

$$\mathsf{Q}_\Lambda^{\overline{\mathrm{x}}}(\mathbf{x}) = \begin{cases} \mathbf{1}_{\{\mathrm{x}=\emptyset\}}, & \text{if } \overline{\mathbf{x}} \in \overline{\mathfrak{I}}, \\ (p(\overline{\mathbf{x}}))^{|\mathrm{x}|}\,(1-p(\overline{\mathbf{x}}))^{|\Lambda\setminus\mathrm{x}|}, & \text{otherwise.} \end{cases}$$

4.4. Martingale-difference one-point systems and random fields

A random field $\{\xi_t,\ t \in \mathbf{Z}^\nu\}$ taking values in $\mathcal{X} \subset \mathbf{R}$ is called martingale-difference if for all $t \in \mathbf{Z}^\nu$ we have

$$\mathsf{E}(\xi_t \mid \xi_s,\ s \in \mathbf{Z}^\nu \setminus t) = 0$$

almost surely. This definition implies that if for all $\Lambda \in \mathcal{E}$ we put $S_\Lambda = \sum_{t\in\Lambda} \xi_t$, then for all finite $\widetilde{\Lambda} \subset \Lambda \subset \mathbf{Z}^\nu$ the following martingale equality holds:

$$\mathsf{E}(S_\Lambda \mid \xi_t,\ t \in \widetilde{\Lambda}) = S_{\widetilde{\Lambda}}.$$

Martingale-difference random fields are interesting for several reasons. For example, for such random fields one can develop a limit theory similar to one developed in [10] for martingale-difference random processes (see, for example, [12, 14]). Besides, such random fields are of interest in some problems of statistical physics. One of the reasons is the asymptotically normal behaviour of the total spin at the critical point for martingale-difference models, another one is that to any lattice model with finite spin one can correspond some martingale-difference model, so that probabilities of the total spin in one model can be expressed in terms of probabilities of the other one by means of some combinatorial formulas (see [13]). This gives a possibility to study the models of statistical physics by means of martingale theory.

Below we give a simple method for constructing martingale-difference random fields using one-point systems.

Let $\mathcal{X} \subset \mathbf{R}$ be a finite space with a fixed vacuum θ.

Definition 4.2. A one-point system

$$\mathbf{h} = \{h_t^{\overline{\mathrm{x}}}(x),\ t \in \mathbf{Z}^\nu,\ x \in \mathcal{X}^*,\ \overline{\mathbf{x}} \in {\mathcal{X}^*}^S,\ S \subset \mathbf{Z}^\nu \setminus t\}$$

is called martingale-difference if we have

$$\sum_{x \in \mathcal{X}^*} x\, h_t^{\overline{\mathrm{x}}}(x) = 0$$

for all $t \in \mathbf{Z}$ and $\overline{\mathbf{x}} \in {\mathcal{X}^*}^S$, $S \subset \mathbf{Z}^\nu \setminus t$.

We need the following general result about one-point systems.

Lemma 4.2. *Consider finite state spaces $\mathcal{X}$ and $\mathcal{W}$ with fixed vacuums $\theta \in \mathcal{X}$ and $\theta' \in \mathcal{W}$ respectively and write, as usually, $\mathcal{X}^* = \mathcal{X} \setminus \theta$ and $\mathcal{W}^* = \mathcal{W} \setminus \theta'$. Let $\mathbf{h} = \{h_t^{\overline{\mathrm{x}}}(x),\ t \in \mathbf{Z}^\nu,\ x \in \mathcal{X}^*,\ \overline{\mathbf{x}} \in {\mathcal{X}^*}^S,\ S \subset \mathbf{Z}^\nu \setminus t\}$ be a one-point system, and let φ be a function $\varphi :\ \mathcal{W} \longrightarrow \mathcal{X}$ such that $\varphi(w) = \theta$ if and only if $w = \theta'$. Then the system*

$$\mathbf{h}_\varphi = \{h_t^{\varphi(\overline{\mathrm{w}})}(\varphi(w)),\ t \in \mathbf{Z}^\nu,\ w \in \mathcal{W}^*,\ \overline{\mathbf{w}} \in {\mathcal{W}^*}^S,\ S \subset \mathbf{Z}^\nu \setminus t\},$$

where $\varphi(\overline{\mathbf{w}}) = \{\varphi(\overline{w}_s),\ s \in S\}$, is also a one-point system.

Proof. First of all, let us note that since $\varphi(w) = \theta$ if and only if $w = \theta'$, the system $\mathbf{h}_\varphi$ is well defined. Further, for all $s, t \in \mathbf{Z}^\nu$, $w, v \in \mathcal{W}^*$ and $\overline{\mathbf{w}} \in {\mathcal{W}^*}^S$, $S \subset \mathbf{Z}^\nu \setminus \{s,t\}$ we have

$$\begin{aligned} h_s^{\varphi(\overline{\mathrm{w}})}(\varphi(v))\, h_t^{\varphi(\overline{\mathrm{w}} \oplus v_s)}(\varphi(w)) &= h_s^{\varphi(\overline{\mathrm{w}})}(\varphi(v))\, h_t^{\varphi(\overline{\mathrm{w}}) \oplus \varphi(v)_s}(\varphi(w)) \\ &= h_t^{\varphi(\overline{\mathrm{w}})}(\varphi(w))\, h_s^{\varphi(\overline{\mathrm{w}}) \oplus \varphi(w)_t}(\varphi(v)) \\ &= h_t^{\varphi(\overline{\mathrm{w}})}(\varphi(w))\, h_s^{\varphi(\overline{\mathrm{w}} \oplus w_t)}(\varphi(v)) \end{aligned}$$

which concludes the proof. □

Now we can state the following

Theorem 4.6. *Suppose the conditions of the preceding lemma are satisfied, and suppose moreover* $\mathcal{W} \subset \mathbf{R}$, $\theta' = 0$ *and*

$$\sum_{w \in \varphi^{-1}(x)} w = 0 \quad \textit{for all} \quad x \in \mathcal{X}^*.$$

Then the one-point system $\mathbf{h}_\varphi$ *is martingale-difference.*

Proof. We need to show that for all $t \in \mathbf{Z}$ and $\overline{\mathrm{w}} \in \mathcal{W}^{*S}$, $S \subset \mathbf{Z}^\nu \setminus t$, we have $\sum_{w \in \mathcal{W}^*} w\, h_t^{\varphi(\overline{\mathrm{w}})}(\varphi(w)) = 0$. Indeed,

$$\begin{aligned} \sum_{w \in \mathcal{W}^*} w\, h_t^{\varphi(\overline{\mathrm{w}})}(\varphi(w)) &= \sum_{x \in \mathcal{X}^*} \sum_{w \in \varphi^{-1}(x)} w\, h_t^{\varphi(\overline{\mathrm{w}})}(\varphi(w)) \\ &= \sum_{x \in \mathcal{X}^*} h_t^{\varphi(\overline{\mathrm{w}})}(x) \sum_{w \in \varphi^{-1}(x)} w \;=\; 0 \end{aligned}$$

which concludes the proof. □

It follows clearly from Theorem 4.6 that if the one-point system $\mathbf{h}_\varphi$ defines some random field, then this field is martingale-difference. The existence of such fields can be guaranteed, for example, by quasilocality of the one-point system $\mathbf{h}$, since quasilocality of $\mathbf{h}$ implies obviously the quasilocality of $\mathbf{h}_\varphi$. For example, if we consider the ferromagnetic Ising model, $\mathcal{W} = \{-1,0,1\}$ and $\varphi(w) = 2\,|w| - 1$, we get the martingale-difference model considered in [13].

References

[1] M.B. Averintsev (1970) A way of describing random fields with a discrete argument. *Problems Inform. Transmission* **6**, 169–175.

[2] S. Dachian (1998) Nonparametric estimation for Gibbs random fields specified through one-point systems. *Statistical Inference for Stochastic Processes* **1**, 245–264.

[3] S. Dachian and B.S. Nahapetian (1995) Inclusion-exclusion description of random fields. *J. Contemp. Math. Anal. Armen. Acad. Sci.* **30** (6), 50–61.

[4] S. Dachian and B.S. Nahapetian, (1998) An approach towards description of random fields. Preprint Université du Maine, no. 98–5.

[5] R.L. Dobrushin (1968) The description of a random field by means of conditional probabilities and conditions of its regularity. *Theory Probab. and Appl.* **13**, 197–224.

[6] R.L. Dobrushin (1968) Gibbsian random fields for lattice systems with pairwise interaction. *Funct. Anal. Appl.* **2**, 292–301.

[7] R.L. Dobrushin (1968) The problem of uniqueness of a Gibbs random field and the problem of phase transitions. *Funct. Anal. Appl.* **2**, 302–312.

[8] A.C.D van Enter, R. Fernández and A.D. Sokal (1993) Regularity properties and pathologies of position-space renormalization-group transformations: scope and limitations of Gibbsian theory. *J. Stat. Phys.* **72** (5–6), 879–1167.

[9] H.-O. Georgii (1988) *Gibbs Measures and Phase Transitions.* Walter de Gruyter, Berlin.

[10] P. Hall and C.C. Heyde (1980) *Martingale Limit Theory and its Application.* Acad. Press.

[11] O.K. Kozlov (1974) Gibbsian description of a system of random quantities. *Problems Inform. Transmission* **10**, 258–265.

[12] B.S. Nahapetian (1995) Billingsley – Ibragimov theorem for martingale-difference Gibbs random fields and its applications to some models of classical statistical physics. *C. R. Acad. Sci. Paris, Ser. I* **320**, 1539–1544.

[13] B.S. Nahapetian (1997) Models with even potentials and the behaviour of total spin at the critical point. *Commun. Math. Phys.* **189** (2), 513–519.

[14] B.S. Nahapetian and A.N. Petrosian (1992) Martingale-difference Gibbs random fields and central limit theorem. *Ann. Acad. Sci. Fennicae, Ser. A–I* **17**, 105–110.

[15] W.G. Sullivan (1973) Potentials for almost Markovian random fields. *Commun. Math. Phys.* **33**, 61–74.

Journal of Statistical Physics, Vol. 117, Nos. 1/2, October 2004 (© 2004)

Description of Specifications by Means of Probability Distributions in Small Volumes under Condition of Very Weak Positivity

S. Dachian[1] and B. S. Nahapetian[2]

Received July 18, 2003; accepted April 28, 2004

The problem of description of specifications by means of probability distributions in small volumes with infinite boundary conditions is considered. The description of specifications by means of n-specifications (consistent systems of probability distributions in volumes of cardinality bounded by n with infinite boundary conditions) is established under the condition of very weak positivity. Particular attention is paid to the most important case $n=1$ which requires special considerations.

KEY WORDS: Consistency conditions; specification; n-specification; positivity conditions; weak positivity; very weak positivity.

1. INTRODUCTION

The notion of specification—consistent system of probability distributions in finite volumes with infinite boundary conditions—is a basic one in the theory of random fields and in mathematical statistical physics. The importance of this notion is that the description of random fields in terms of specifications turned out to be a powerful tool for the development of the theory of random fields (see, for example, ref. 1). Besides, the specifications admitting Gibbsian description represent the mathematical

[1]Laboratoire de Mathématiques Appliquées, Université Blaise Pascal (Clermont-Ferrand 2), 63177 Aubière Cedex, France; e-mail: Serguei.Dachian@math.univ-bpclermont.fr

[2]Institute of Mathematics, National Academy of Sciences of Armenia, 24-b, Av. M. Bagramian, 375019 Yerevan, Armenia; e-mail: nahapet@instmath.sci.am

background for the study of systems of statistical physics. The problem of Gibbsian description of specifications was a subject of consideration of many authors (see, for example, refs. 2–5).

The theory of description of random fields by means of specifications was constructed by Dobrushin in his fundamental works (refs. 6–8). Particularly, the conditions of existence and uniqueness of random fields described by a given specification were obtained in ref. 6.

In the latter work, while commenting the uniqueness condition, Dobrushin touched upon the problem of restoration of specifications by means of their one-point elements. Several years ago, in a private conversation with one of the authors Dobrushin pointed out the importance of a closely related problem: the problem of description of specifications by means of consistent systems of one-point probability distributions with infinite boundary conditions. However, at that time no consistency conditions on one-point probability distributions were known.

These two problems of Dobrushin were solved by the authors in refs. 9 and 10 under the condition of weak positivity (as well as under the condition of strict positivity). In particular, consistency conditions under which a system of one-point probability distributions with infinite boundary conditions describes a specification were established in ref. 10 under the condition of weak positivity. There it was also shown that the weak positivity condition is coordinating, that is, a specification is weakly positive if and only if its subsystem consisting of one-point elements is weakly positive. It was equally proved that under the condition of weak positivity, the quasilocality property is heritable, that is, a weakly positive specification is quasilocal if and only if its subsystem consisting of one-point elements is quasilocal.

Let us note here that the consistency conditions established in refs. 9 and 10 were mentioned as properties of strictly positive conditional probabilities of Markov random fields in ref. 11. Note also that some results concerning the problem of restoration of strictly positive specifications can be found in refs. 1 and 5. In ref. 12 the attempt to solve the problems of restoration and description of specifications in some non-positive cases was undertaken, but sufficiently full results were obtained only in one-dimensional case and under more complicated conditions.

In the present work the results of ref. 10 are extended to essentially more general situation. First, instead of consistent systems of one-point probability distributions with infinite boundary conditions we consider more general systems: so-called n-specifications, that is, consistent systems of probability distributions in small volumes (volumes of cardinality bounded by n) with infinite boundary conditions. But the principal

difference is that the results are obtained under so-called very weak positivity condition which is essentially weaker than the conditions used in ref. 10.

Note that the results of the present work allow one to formulate the condition of existence of random fields described by a given specification in terms of the latter's one-point elements only, that is, exactly in the same terms as the well-known Dobrushin's uniqueness condition. So, it becomes possible to formulate the problem of description of random fields directly in terms of 1-specifications.

Note in addition, that the results of the present work will be probably useful in the recently emerged theory of non-Gibbsian random fields which are now intensively studied (see, for example, ref. 13).

Note finally, that the methods used in the present work are new and considerably differ from those used in ref. 10.

2. PRELIMINARIES

We denote by $\mathbb{Z}^\nu$ the ν-dimensional integer lattice and by $\mathscr{E}$ the set of all finite subsets of $\mathbb{Z}^\nu$, that is, $\mathscr{E}=\{\Lambda\subset\mathbb{Z}^\nu:|\Lambda|<\infty\}$, where $|\Lambda|$ is the cardinality (the number of points) of the set Λ. For convenience of notations, we will omit braces for one-point sets, that is, will write a instead of $\{a\}$. For any $n\in\mathbb{N}\cup\infty=\{1,2,\dots,\infty\}$ we equally denote $\mathscr{E}_n=\{\Lambda\in\mathscr{E}:|\Lambda|\leqslant n\}$. Clearly, for $n=\infty$ we have $\mathscr{E}_\infty=\mathscr{E}$.

Let $(\mathscr{X},\mathscr{F})$ be some measurable *state space*. Usually $\mathscr{X}$ is assumed to be endowed with some topology $\mathscr{T}$, and $\mathscr{F}$ is assumed to be the Borel σ-algebra for this topology. In the present work we concentrate on the case when $\mathscr{X}$ is finite, $\mathscr{T}$ is the discrete topology and $\mathscr{F}$ is the total σ-algebra, that is, $\mathscr{F}=\mathscr{T}=\mathrm{part}(\mathscr{X})$.

For any $T\subset\mathbb{Z}^\nu$ we consider the space $\mathscr{X}^T$ of all configurations on T. For $T=\emptyset$ we assume that $\mathscr{X}^\emptyset=\{\boldsymbol{\emptyset}\}$, where $\boldsymbol{\emptyset}$ is understood as an empty configuration. For any $T,S\subset\mathbb{Z}^\nu$ such that $T\subset S$ and any configuration $\boldsymbol{x}=\{x_t,\ t\in S\}$ on S we denote $\boldsymbol{x}_T$ the *subconfiguration (restriction)* of $\boldsymbol{x}$ on T defined by $\boldsymbol{x}_T=\{x_t,\ t\in T\}$. For any $T,S\subset\mathbb{Z}^\nu$ such that $T\cap S=\emptyset$ and any configurations $\boldsymbol{x}$ on T and $\boldsymbol{y}$ on S we denote $\boldsymbol{x}\boldsymbol{y}$ the *concatenation* of $\boldsymbol{x}$ and $\boldsymbol{y}$, that is, the configuration on $T\cup S$ equal to $\boldsymbol{x}$ on T and to $\boldsymbol{y}$ on S. For any $a\in\mathscr{X}$, $T\subset\mathbb{Z}^\nu$ and $\boldsymbol{x}\in\mathscr{X}^T$, the notation $\boldsymbol{x}\equiv a$ will mean $x_t=a$ for any $t\in T$, and the notation $\boldsymbol{x}\ni a$ will mean $x_t=a$ for some $t\in T$.

Let $\Lambda\in\mathscr{E}$. We denote a probability distribution $\{\mathbf{P}_\Lambda(\boldsymbol{x}),\ \boldsymbol{x}\in\mathscr{X}^\Lambda\}$ on $\mathscr{X}^\Lambda$ by $\mathbf{P}_\Lambda$. Note that in the case $\Lambda=\emptyset$ there exists only one probability distribution defined by $\mathbf{P}_\emptyset(\boldsymbol{\emptyset})=1$. For any $I\subset\Lambda$ we denote $(\mathbf{P}_\Lambda)_I$ the

restriction (or marginal distribution) of $\mathbf{P}_\Lambda$ on I, defined by

$$(\mathbf{P}_\Lambda)_I(\boldsymbol{x}) = \sum_{\boldsymbol{y} \in \mathscr{X}^{\Lambda \setminus I}} \mathbf{P}_\Lambda(\boldsymbol{x}\boldsymbol{y}).$$

Finally, let us recall Dobrushin's consistency condition and the notion of specification, introduced in ref. 6.

Definition 1. Let $\Lambda \in \mathscr{E}$. Any system $\{\mathbf{Q}_\Lambda^{\overline{\boldsymbol{x}}},\ \overline{\boldsymbol{x}} \in \mathscr{X}^{\mathbb{Z}^\nu \setminus \Lambda}\}$ of probability distributions on $\mathscr{X}^\Lambda$ indexed by infinite boundary conditions will be called Λ-*kernel* and denoted by $\mathbf{Q}_\Lambda^\bullet$.

Definition 2. Let $\Lambda \in \mathscr{E}$ and $I \subset \Lambda$. We will say that a Λ-kernel $\mathbf{Q}_\Lambda^\bullet$ is *consistent in Dobrushin's sense* with an I-kernel $\mathbf{Q}_I^\bullet$ (and vice versa), if

$$\mathbf{Q}_\Lambda^{\overline{\boldsymbol{x}}}(\boldsymbol{x}\boldsymbol{y}) = (\mathbf{Q}_\Lambda^{\overline{\boldsymbol{x}}})_{\Lambda \setminus I}(\boldsymbol{x})\ \mathbf{Q}_I^{\overline{\boldsymbol{x}}\boldsymbol{x}}(\boldsymbol{y})$$

$$\text{for any } \boldsymbol{x} \in \mathscr{X}^{\Lambda \setminus I},\ \boldsymbol{y} \in \mathscr{X}^I \text{ and } \overline{\boldsymbol{x}} \in \mathscr{X}^{\mathbb{Z}^\nu \setminus \Lambda}.$$

Definition 3. A family $\{\mathbf{Q}_\Lambda^\bullet, \Lambda \in \mathscr{E}\}$ of Λ-kernels indexed by $\Lambda \in \mathscr{E}$ will be called *specification*, if $\mathbf{Q}_\Lambda^\bullet$ and $\mathbf{Q}_I^\bullet$ are consistent in Dobrushin's sense for any $\Lambda \in \mathscr{E}$ and $I \subset \Lambda$.

The main goal of this work is the description of specifications by means of probability distributions in small volumes with infinite boundary conditions, more precisely, by means of n-specifications.

3. NOTION OF *n*-SPECIFICATION AND POSITIVITY POINTS

Recall that specifications are families of Λ-kernels in finite volumes. Let us consider smaller systems: families of Λ-kernels in volumes with bounded size.

Definition 4. Let $n \in \mathbb{N}$. Any family $\{\mathbf{Q}_\Lambda^\bullet, \Lambda \in \mathscr{E}_n\}$ of Λ-kernels indexed by $\Lambda \in \mathscr{E}_n$ will be called *n-system*.

In order to describe specifications, n-systems must satisfy some consistency conditions which should at least be properties of n-systems contained in specifications. So, let us introduce the following notion of n-specification.

Definition 5. Let $n \in \mathbb{N} \setminus 1$. An n-system $\{\mathbf{Q}_\Lambda^\bullet, \Lambda \in \mathscr{E}_n\}$ will be called *n-specification*, if $\mathbf{Q}_\Lambda^\bullet$ and $\mathbf{Q}_I^\bullet$ are consistent in Dobrushin's sense for any $\Lambda \in \mathscr{E}_n$ and $I \subset \Lambda$.

Note that the n-systems contained in specifications are indeed n-specifications. Note also, that in Definitions 4 and 5 one can include the case $n=\infty$, and that ∞-specifications defined this way will be clearly nothing else but specifications.

Remark equally, that we did not yet define the 1-specifications, which would be the most interesting for our purpose, since they are the smallest. Why we did not do it? The matter is that if we introduce the notion of 1-specification in the way of Definition 5, then it would be degenerate, since for 1-systems Dobrushin's consistency conditions become identities. So, in order to define the notion of 1-specification, it is necessary to find some "internal consistency conditions" (that is, some relations between one-point probabilities only), which should be properties of 1-systems contained in specifications. Such properties are given in Theorem 8, but before formulating it let us introduce the notion of positivity point, which will play an important role all along this paper.

Definition 6. Let $\Lambda \in \mathscr{E}$, let $T \subset \mathbb{Z}^\nu \setminus \Lambda$ and $\bar{\boldsymbol{x}} \in \mathscr{X}^{\mathbb{Z}^\nu \setminus \Lambda \setminus T}$, and let $\mathbf{Q}^\bullet_\Lambda$ be a Λ-kernel. A configuration $\boldsymbol{u} \in \mathscr{X}^\Lambda$ is called *positivity point (p.p.) of* $\mathbf{Q}^\bullet_\Lambda$ *under boundary condition (b.c.) varying on* T *and equal to* $\bar{\boldsymbol{x}}$ *outside*, if for any $\boldsymbol{\alpha} \in \mathscr{X}^T$, we have $\mathbf{Q}^{\bar{\boldsymbol{x}}\boldsymbol{\alpha}}_\Lambda(\boldsymbol{u}) > 0$.

Let us formulate immediately one of the most important properties of positivity points.

Theorem 7. Let $J, I \in \mathscr{E}$ such that $J \cap I = \emptyset$, put $\Lambda = J \cup I$, let $T \subset \mathbb{Z}^\nu \setminus \Lambda$ and $\bar{\boldsymbol{x}} \in \mathscr{X}^{\mathbb{Z}^\nu \setminus \Lambda \setminus T}$, and let $\mathbf{Q}^\bullet_J$, $\mathbf{Q}^\bullet_I$ and $\mathbf{Q}^\bullet_\Lambda$ be a J-kernel, an I-kernel and a Λ-kernel. Suppose $\mathbf{Q}^\bullet_\Lambda$ is consistent in Dobrushin's sense both with $\mathbf{Q}^\bullet_J$ and $\mathbf{Q}^\bullet_I$. If $\boldsymbol{u}$ is a p.p. of $\mathbf{Q}^\bullet_J$ under b.c. varying on $I \cup T$ and equal to $\bar{\boldsymbol{x}}$ outside, $\boldsymbol{v}$ is a p.p. of $\mathbf{Q}^\bullet_I$ under b.c. varying on $J \cup T$ and equal to $\bar{\boldsymbol{x}}$ outside, then the concatenation $\boldsymbol{uv}$ is a p.p. of $\mathbf{Q}^\bullet_\Lambda$ under b.c. varying on T and equal to $\bar{\boldsymbol{x}}$ outside.

This theorem will be proved in Section 6, as well as the following theorem presenting the above mentioned properties of 1-systems contained in specifications.

Theorem 8. If $\mathcal{Q} = \{\mathbf{Q}^\bullet_\Lambda, \Lambda \in \mathscr{E}_2\}$ is 2-specification, then

$$\mathbf{Q}^{\bar{\boldsymbol{x}}v}_t(x)\, \mathbf{Q}^{\bar{\boldsymbol{x}}x}_s(y)\, \mathbf{Q}^{\bar{\boldsymbol{x}}y}_t(u)\, \mathbf{Q}^{\bar{\boldsymbol{x}}u}_s(v) = \mathbf{Q}^{\bar{\boldsymbol{x}}u}_s(y)\, \mathbf{Q}^{\bar{\boldsymbol{x}}y}_t(x)\, \mathbf{Q}^{\bar{\boldsymbol{x}}x}_s(v)\, \mathbf{Q}^{\bar{\boldsymbol{x}}v}_t(u)$$
$$\text{for any } t, s \in \mathbb{Z}^\nu,\ x \in \mathscr{X}^t,\ y, v \in \mathscr{X}^s \text{ and } \bar{\boldsymbol{x}} \in \mathscr{X}^{\mathbb{Z}^\nu \setminus t \setminus s}, \tag{1}$$

and for any p.p. u of $\mathbf{Q}^\bullet_t$ under b.c. varying on s and equal to $\bar{\boldsymbol{x}}$ outside.

Remarks: (1) This theorem remains valid if any one of x, y, u, v is supposed to be a positivity point.

(2) In the formulation of the theorem we could take $\mathcal{Q}$ to be n-specification for some $n \in (\mathbb{N} \setminus 1) \cup \infty$.

(3) In this theorem $\mathcal{Q}$ is arbitrary, and the conditions are imposed on the arguments of the relation (1) only. A weaker version of the theorem was already established by the authors in ref. 10 under some additional conditions on $\mathcal{Q}$. Note also, that it is not possible to obtain the relation (1) without any condition at all. Indeed, as shows the following example this relation may not hold in general.

Example 9. Let the state space $\mathcal{X} = \{0, 1, 2, 3\}$ and consider the ∞-system $\mathcal{Q} = \{\mathbf{Q}^{\bullet}_{\Lambda}, \Lambda \in \mathcal{E}\}$ defined by

$$\mathbf{Q}^{\overline{x}}_{\Lambda}(x) = \begin{cases} \mathbb{1}_{\{x \equiv 0\}} & \text{if } |\Lambda| \geqslant 2 \\ \mathbb{1}_{\{x=0\}} & \text{if } \overline{x} \ni 0 \\ 1/5 & \text{if } \overline{x} \equiv 1 \text{ and } x \in \{0, 1, 2\} \\ 2/5 & \text{if } \overline{x} \equiv 1 \text{ and } x = 3 \\ 1/4 & \text{if } \overline{x} \not\ni 0 \text{ and } \overline{x} \not\equiv 1 \end{cases} \qquad \text{if } |\Lambda| = 1.$$

It is not difficult to verify that $\mathcal{Q}$ is a specification. Further, if for some arbitrary $t, s \in \mathbb{Z}^{\nu}$, we take $\overline{x} \in \mathcal{X}^{\mathbb{Z}^{\nu} \setminus s \setminus t}$ such that $\overline{x} \equiv 1$, and put $x = 2$, $u = 3$, $y = 1$ and $v = 2$, the relation (1) will clearly fail.

Now, in view of Theorem 8 we can introduce the following notion of 1-specification.

Definition 10. A 1-system $\{\mathbf{Q}^{\bullet}_{\Lambda}, \Lambda \in \mathcal{E}_1\}$ is called 1-*specification*, if

$$\mathbf{Q}^{\overline{x}v}_{t}(x)\, \mathbf{Q}^{\overline{x}x}_{s}(y)\, \mathbf{Q}^{\overline{x}y}_{t}(u)\, \mathbf{Q}^{\overline{x}u}_{s}(v) = \mathbf{Q}^{\overline{x}u}_{s}(y)\, \mathbf{Q}^{\overline{x}y}_{t}(x)\, \mathbf{Q}^{\overline{x}x}_{s}(v)\, \mathbf{Q}^{\overline{x}v}_{t}(u)$$
$$\text{for any } t, s \in \mathbb{Z}^{\nu}, x \in \mathcal{X}^{t}, y, v \in \mathcal{X}^{s} \text{ and } \overline{x} \in \mathcal{X}^{\mathbb{Z}^{\nu} \setminus t \setminus s}, \tag{2}$$

and for any p.p. u of $\mathbf{Q}^{\bullet}_{t}$ under b.c. varying on s and equal to $\overline{x}$ outside.

Note, that like the case $n \geqslant 2$, the 1-systems contained in specifications will be 1-specifications. So, for any $n \in \mathbb{N}$, the restriction of a specification on $\mathcal{E}_n$ is nothing but an n-specification. Description of specifications by means of n-specifications is in some sense an inverse operation to this restriction.

4. PROBLEM OF DESCRIPTION OF SPECIFICATIONS BY MEANS OF *n*-SPECIFICATIONS

The problems of this type was firstly considered by the authors in refs. 9 and 10. In these works, the problem of description of specifications by means of n-specifications was solved for $n=1$ under the condition of "strict positivity", as well as under the condition of "weak positivity".

4.1. Strict Positivity

The strict positivity is the simplest positivity condition for n-systems.

Definition 11. Let $n \in \mathbb{N} \cup \infty$. An n-system $\{\mathbf{Q}^{\bullet}_{\Lambda}, \Lambda \in \mathscr{E}_n\}$ will be called *strictly positive*, if for any $\Lambda \in \mathscr{E}_n$ each configuration $\boldsymbol{x} \in \mathscr{X}^{\Lambda}$ is a p.p. of $\mathbf{Q}^{\bullet}_{\Lambda}$ under b.c. varying on $\mathbb{Z}^{\nu} \setminus \Lambda$ and equal to $\emptyset$ outside.

Remark that Definition 11 simply means, that for any $\Lambda \in \mathscr{E}_n$, any $\boldsymbol{x} \in \mathscr{X}^{\Lambda}$ and any $\overline{\boldsymbol{x}} \in \mathscr{X}^{\mathbb{Z}^{\nu} \setminus \Lambda}$ we have $\mathbf{Q}^{\overline{\boldsymbol{x}}}_{\Lambda}(\boldsymbol{x}) > 0$.

The strictly positive specifications are widely studied and used in mathematical statistical physics. For example, the specifications admitting Gibbsian description with a real-valued potential are strictly positive.

Note also, that under the condition of strict positivity, the consistency conditions (2) from Definition 10 of 1-specification become

$$\mathbf{Q}_t^{\overline{\boldsymbol{x}}v}(x)\,\mathbf{Q}_s^{\overline{\boldsymbol{x}}x}(y)\,\mathbf{Q}_t^{\overline{\boldsymbol{x}}y}(u)\,\mathbf{Q}_s^{\overline{\boldsymbol{x}}u}(v) = \mathbf{Q}_s^{\overline{\boldsymbol{x}}u}(y)\,\mathbf{Q}_t^{\overline{\boldsymbol{x}}y}(x)\,\mathbf{Q}_s^{\overline{\boldsymbol{x}}x}(v)\,\mathbf{Q}_t^{\overline{\boldsymbol{x}}v}(u)$$

$$\text{for any } t, s \in \mathbb{Z}^{\nu}, x, u \in \mathscr{X}^t, y, v \in \mathscr{X}^s \text{ and } \overline{\boldsymbol{x}} \in \mathscr{X}^{\mathbb{Z}^{\nu} \setminus t \setminus s}.$$

Let us now explain the nature and point out several consequences of the problem of description of specifications by means of n-specifications using as example the results obtained in refs. 9 and 10.

The main result is that any strictly positive 1-specification $\boldsymbol{q}$ *describes* a specification, that is, there exists a unique specification containing $\boldsymbol{q}$.

The second result is that the strict positivity condition is *coordinating*, that is, a specification $\mathcal{Q}$ is strictly positive if and only if the 1-specification contained in $\mathcal{Q}$ is strictly positive. Let us note here that the necessity is trivial, and the sufficiency becomes evident in view of considerations of the present work due to Theorem 7.

Note that these two results imply also that any strictly positive specification $\mathcal{Q}$ can be restored by the 1-specification contained in it (that is, any specification containing the same 1-specification is necessarily equal to $\mathcal{Q}$) and allow us to conclude that the description is a one-to-one correspondence between strictly positive 1-specifications and strictly positive specifications.

The third result is that under the condition of strict positivity the quasilocality property is *heritable*, that is, a strictly positive specification $\mathcal{Q}$ is quasilocal if and only if the 1-specification contained in $\mathcal{Q}$ is quasilocal.

This result together with the first one allow us to formulate the condition of existence of random fields described by a given specification in terms of the latter's one-point elements only, that is, exactly in the same terms as the well-known Dobrushin's uniqueness condition, and so, it becomes possible to formulate the problem of description of random fields directly in terms of 1-specifications.

Note in addition, that as it will become clear from the subsequent considerations of this work, these results can be extended to the case of arbitrary $n \in \mathbb{N}$.

Now we want to consider the problem of description outside of the scope of strict positivity condition. First of all let us notice that under no condition at all this description does not hold.

4.2. Counterexample

Let us fix some $n \in \mathbb{N}$. If the description of specifications by means of n-specifications held under no condition at all, then any n-specification would describe a specification. The following example shows that it is not true.

Example 12. Let $\mathscr{X} = \{0, 1\}$, denote $F(\boldsymbol{x})$ the function which counts the number of elements equal to 1 in a configuration $\boldsymbol{x}$ on $T \subset \mathbb{Z}^\nu$ and consider the ∞-system $\mathcal{Q} = \{\mathbf{Q}^\bullet_\Lambda, \Lambda \in \mathscr{E}\}$ defined by

$$\mathbf{Q}^{\overline{\boldsymbol{x}}}_\Lambda(\boldsymbol{x}) = \begin{cases} \mathbb{1}_{\{\boldsymbol{x}\equiv 0\}} & \text{if } F(\overline{\boldsymbol{x}}) = 0 \\ \mathbb{1}_{\{\boldsymbol{x}\equiv 1\}} & \text{if } F(\overline{\boldsymbol{x}}) \geqslant 1 \quad \text{if } |\Lambda| \geqslant 2 \\ \mathbb{1}_{\{\boldsymbol{x}=0\}} & \text{if } F(\overline{\boldsymbol{x}}) = 0 \\ 1/2 & \text{if } F(\overline{\boldsymbol{x}}) = 1 \\ \mathbb{1}_{\{\boldsymbol{x}=1\}} & \text{if } F(\overline{\boldsymbol{x}}) \geqslant 2 \quad \text{if } |\Lambda| = 1. \end{cases}$$

It is not difficult to verify that $\mathcal{Q}$ is a specification. However, the n-specification $\boldsymbol{q}_n$ contained in $\mathcal{Q}$ does not describe a specification, since, for example, the ∞-system $\widehat{\mathcal{Q}} = \{\widehat{\mathbf{Q}}^\bullet_\Lambda, \Lambda \in \mathscr{E}\}$ defined by

$$\widehat{\mathbf{Q}}^{\overline{\boldsymbol{x}}}_\Lambda(\boldsymbol{x}) = \begin{cases} \mathbb{1}_{\{\boldsymbol{x}\equiv 1\}} & \text{if } |\Lambda| \geqslant n+2, \\ \mathbf{Q}^{\overline{\boldsymbol{x}}}_\Lambda(\boldsymbol{x}) & \text{if } |\Lambda| \leqslant n+1 \end{cases}$$

is also a specification containing $\boldsymbol{q}_n$.

So, it becomes evident that in order for the description of specifications by means of n-specifications to hold, some kind of positivity condition is necessary. The strict positivity is the most restrictive positivity condition, since it does not permit zeros at all. A weaker positivity condition is the "weak positivity" which was already studied by the authors in refs. 9 and 10.

4.3. Weak Positivity

The weak positivity condition for n-systems is formulated as follows.

Definition 13. Let $n \in \mathbb{N} \cup \infty$. An n-system $\{\mathbf{Q}^{\bullet}_{\Lambda}, \Lambda \in \mathscr{E}_n\}$ will be called *weakly positive*, if there exist some element $\theta \in \mathscr{X}$ (called *vacuum*), such that for any $\Lambda \in \mathscr{E}_n$ the configuration $\boldsymbol{x} \equiv \theta$ is a p.p. of $\mathbf{Q}^{\bullet}_{\Lambda}$ under b.c. varying on $\mathbb{Z}^{\nu} \setminus \Lambda$ and equal to $\boldsymbol{\emptyset}$ outside.

Clearly, this condition on n-systems is really weaker than the strict positivity one. It remains really weaker when applied to n-specifications too. For instance, the n-specification contained in the specification $\mathcal{Q}$ from Example 9 is weakly positive but not strictly positive.

Weakly positive specifications are well known in mathematical statistical physics. For example, the specifications admitting Gibbsian description with a vacuum potential (which may take infinite values) are weakly positive.

Note also, that under the condition of weak positivity, the consistency conditions (2) from Definition 10 of 1-specification have a simpler equivalent form given in the following proposition. The proof of this proposition is quite similar to those of Proposition 18 (see Section 6) and will be omitted.

Proposition 14. A weakly positive 1-system $\{\mathbf{Q}^{\bullet}_{\Lambda}, \Lambda \in \mathscr{E}_1\}$ is 1-specification if and only if

$$\mathbf{Q}_t^{\overline{\boldsymbol{x}} v^{\circ}}(x)\, \mathbf{Q}_s^{\overline{\boldsymbol{x}} x}(y)\, \mathbf{Q}_t^{\overline{\boldsymbol{x}} y}(u^{\circ})\, \mathbf{Q}_s^{\overline{\boldsymbol{x}} u^{\circ}}(v^{\circ}) = \mathbf{Q}_s^{\overline{\boldsymbol{x}} u^{\circ}}(y)\, \mathbf{Q}_t^{\overline{\boldsymbol{x}} y}(x)\, \mathbf{Q}_s^{\overline{\boldsymbol{x}} x}(v^{\circ})\, \mathbf{Q}_t^{\overline{\boldsymbol{x}} v^{\circ}}(u^{\circ})$$

$$\text{for any } t, s \in \mathbb{Z}^{\nu},\ x \in \mathscr{X}^t,\ y \in \mathscr{X}^s,\ \overline{\boldsymbol{x}} \in \mathscr{X}^{\mathbb{Z}^{\nu} \setminus t \setminus s},$$

$$\text{and for } u^{\circ} \in \mathscr{X}^t \text{ such that } u^{\circ} = \theta \text{ and } v^{\circ} \in \mathscr{X}^s \text{ such that } v^{\circ} = \theta.$$

As we have already mentioned, the problem of description of specifications by means of n-specifications under the condition of weak positivity was solved for $n = 1$ in refs. 9 and 10. There it was shown, that any weakly positive 1-specification describes a specification. It was equally shown, that the weak positivity condition is coordinating, and under this condition the

quasilocality property is heritable. Moreover, as it will become clear from the subsequent considerations of this work, these results can be extended to the case of arbitrary $n \in \mathbb{N}$.

So, the further study of the problem of description of specifications by means of n-specifications reduces to determination of a weaker (in ideal case the weakest) positivity condition, under which this description holds. Such a condition is the very weak positivity condition obtained in the present work.

4.4. Very Weak Positivity

Since the positivity points used in Definition 10 of 1-specification are positivity points under boundary condition varying on one-point sets only, it seems natural to consider the following positivity condition.

Definition 15. Let $n \in \mathbb{N} \cup \infty$. An n-system $\{\mathbf{Q}^\bullet_\Lambda, \Lambda \in \mathscr{E}_n\}$ will be called *too weakly positive*, if for any $\Lambda \in \mathscr{E}_n$, any $s \in \mathbb{Z}^\nu \setminus \Lambda$ and any $\overline{\boldsymbol{x}} \in \mathscr{X}^{\mathbb{Z}^\nu \setminus \Lambda \setminus s}$, there exists a p.p. of $\mathbf{Q}^\bullet_\Lambda$ under b.c. varying on s and equal to $\overline{\boldsymbol{x}}$ outside.

However, in accordance with its name, this condition is too weak in order to solve the problem of description. Indeed, a too weakly positive n-specification not necessarily describes a specification (for $n = 1$ it is sufficient to consider the 1-specification $\boldsymbol{q}_1$ from Example 12, and a similar example can be easily constructed for arbitrary $n \in \mathbb{N}$). Moreover, the too weak positivity condition is not coordinating (for instance, the specification $\boldsymbol{Q}$ from Example 12 is not too weakly positive). But what is the matter?

The weak positivity and strict positivity conditions were shown to be coordinating by concatenating positivity points thanks to Theorem 7. But for the too weak positivity condition this approach does not work: if we concatenate two positivity points under boundary conditions varying on one-point sets, we obtain a positivity point under fixed (varying on the empty set) boundary condition. So, we need to modify (strengthen) the condition of too weak positivity in order to be able to correctly concatenate positivity points. This leads us to introduce the following positivity condition.

Definition 16. Let $n \in \mathbb{N} \cup \infty$. An n-system $\{\mathbf{Q}^\bullet_\Lambda, \Lambda \in \mathscr{E}_n\}$ will be called *very weakly positive*, if for any $\Lambda \in \mathscr{E}_n$, any $V \in \mathscr{E}$ such that $V \subset \mathbb{Z}^\nu \setminus \Lambda$ and any $\overline{\boldsymbol{x}} \in \mathscr{X}^{\mathbb{Z}^\nu \setminus \Lambda \setminus V}$, there exists some p.p. $\boldsymbol{u} = \theta(\Lambda, V, \overline{\boldsymbol{x}})$ of $\mathbf{Q}^\bullet_\Lambda$ under b.c. varying on V and equal to $\overline{\boldsymbol{x}}$ outside.

Clearly, this condition on n-systems is really weaker than the weak positivity one. As shows the following example, it remains really weaker when applied to n-specifications too.

Example 17. Let $\mathscr{X} = \{0, 1\}$, let F be the function used in Example 12 and consider the ∞-system $\mathcal{Q} = \{\mathbf{Q}^{\bullet}_{\Lambda}, \Lambda \in \mathscr{E}\}$ defined by

$$\mathbf{Q}^{\overline{x}}_{\Lambda}(x) = \begin{cases} \mathbb{1}_{\{x \equiv 0\}} & \text{if } F(\overline{x}) = \infty, \\ \mathbb{1}_{\{x \equiv 1\}} & \text{if } F(\overline{x}) < \infty. \end{cases}$$

It is not difficult to verify that $\mathcal{Q}$ is a specification, and that the n-specification contained in $\mathcal{Q}$ is very weakly positive but not weakly positive.

Note also, that as well as in the weakly positive case, under the condition of very weak positivity, the consistency conditions (2) from Definition 10 of 1-specification have a simpler equivalent form given in the following proposition which will be proved in Section 6.

Proposition 18. A very weakly positive 1-system $\{\mathbf{Q}^{\bullet}_{\Lambda}, \Lambda \in \mathscr{E}_1\}$ is 1-specification if and only if

$$\mathbf{Q}^{\overline{x}v^{\circ}}_{t}(x)\, \mathbf{Q}^{\overline{x}x}_{s}(y)\, \mathbf{Q}^{\overline{x}y}_{t}(u^{\circ})\, \mathbf{Q}^{\overline{x}u^{\circ}}_{s}(v^{\circ}) = \mathbf{Q}^{\overline{x}u^{\circ}}_{s}(y)\, \mathbf{Q}^{\overline{x}y}_{t}(x)\, \mathbf{Q}^{\overline{x}x}_{s}(v^{\circ})\, \mathbf{Q}^{\overline{x}v^{\circ}}_{t}(u^{\circ})$$
$$\text{for any } t, s \in \mathbb{Z}^{\nu}, x \in \mathscr{X}^{t}, y \in \mathscr{X}^{s}, \overline{x} \in \mathscr{X}^{\mathbb{Z}^{\nu} \setminus t \setminus s}, \tag{3}$$
$$\text{and for } u^{\circ} = \theta(t, s, \overline{x}) \text{ and } v^{\circ} = \theta(s, t, \overline{x}).$$

In Section 5 we present the main results of this paper which establish the description of specifications by means of n-specifications under the condition of very weak positivity.

5. MAIN RESULTS AND THEIR PROOFS

The main results of this work consist of the following three theorems.

The first one is that any very weakly positive n-specification describes a specification.

Theorem 19. Let $n \in \mathbb{N}$, and let $\boldsymbol{q}$ be a very weakly positive n-specification. Then there exists a unique specification containing $\boldsymbol{q}$.

The second one is that the very weak positivity condition is coordinating.

Theorem 20. Let $n \in \mathbb{N}$, let $\mathcal{Q}$ be a specification, and let q be the n-specification contained in $\mathcal{Q}$. Then $\mathcal{Q}$ is very weakly positive if and only if q is very weakly positive.

The third one is that under the condition of very weak positivity, the quasilocality property is heritable.

Theorem 21. Let $n \in \mathbb{N}$, let $\mathcal{Q}$ be a very weakly positive specification, and let q be the n-specification contained in $\mathcal{Q}$. Then $\mathcal{Q}$ is quasilocal if and only if q is quasilocal.

The proof of the second theorem is evident, since the necessity is trivial, and the sufficiency directly follows from Theorem 7. The third theorem will become clear in view of the proof of the first one. The proof of the latter will be given in the end of this section and needs some auxiliary results which are of independent interest too. These results are given below and will be proved in Section 6.

Proposition 22. Let $\Lambda \in \mathscr{E}$ and $I \subset \Lambda$. A Λ-kernel $\mathbf{Q}^{\bullet}_{\Lambda}$ and an I-kernel $\mathbf{Q}^{\bullet}_{I}$ are consistent in Dobrushin's sense if and only if

$$\mathbf{Q}^{\overline{x}}_{\Lambda}(x\,y)\;\mathbf{Q}^{\overline{x}x}_{I}(v) = \mathbf{Q}^{\overline{x}}_{\Lambda}(x\,v)\;\mathbf{Q}^{\overline{x}x}_{I}(y) \quad \text{for any } x \in \mathscr{X}^{\Lambda \setminus I},\ y, v \in \mathscr{X}^{I} \text{ and } \overline{x} \in \mathscr{X}^{\mathbb{Z}^{\nu} \setminus \Lambda}. \tag{4}$$

The equivalent form given in this proposition looks simpler than the original form of Dobrushin's consistency condition and will be intensively used in our considerations.

Proposition 23. Let $n \in (\mathbb{N} \setminus 1) \cup \infty$. An n-system $\mathcal{Q} = \{\mathbf{Q}^{\bullet}_{\Lambda}, \Lambda \in \mathscr{E}_n\}$ will be n-specification if and only if $\mathbf{Q}^{\bullet}_{\Lambda}$ and $\mathbf{Q}^{\bullet}_{\Lambda \setminus t}$ are consistent in Dobrushin's sense for any $\Lambda \in \mathscr{E}_n$ and $t \in \Lambda$.

This proposition considerably reduces the set of Dobrushin's consistency conditions needed in order to check if an n-system is n-specification.

The next and final theorem establish a general and useful property of n-specifications.

Theorem 24. Let $n \in (\mathbb{N} \setminus 1) \cup \infty$ and let $\mathcal{Q} = \{\mathbf{Q}^{\bullet}_{\Lambda}, \Lambda \in \mathscr{E}_n\}$ be an n-system.

(1) If $\mathcal{Q}$ is n-specification, then

$$\begin{gathered}\mathbf{Q}_A^{\bar{x}u_B}(x_A)\,\mathbf{Q}_B^{\bar{x}x_A}(x_B)\,\mathbf{Q}_C^{\bar{x}x_D}(u_C)\,\mathbf{Q}_D^{\bar{x}u_C}(u_D)\\ =\mathbf{Q}_D^{\bar{x}u_C}(x_D)\,\mathbf{Q}_C^{\bar{x}x_D}(x_C)\,\mathbf{Q}_B^{\bar{x}x_A}(u_B)\,\mathbf{Q}_A^{\bar{x}u_B}(u_A)\\ \text{for any } A,B,C,D \text{ such that}\\ A\cup B=C\cup D\in\mathscr{E}_n \text{ and } A\cap B=C\cap D=\emptyset,\\ \text{and for any } x,u\in\mathscr{X}^{A\cup B} \text{ and } \bar{x}\in\mathscr{X}^{\mathbb{Z}^\nu\setminus A\setminus B} \text{ such that}\\ u_C \text{ is a p.p. of } \mathbf{Q}_C^\bullet \text{ under b.c. varying on } D \text{ and equal to } \bar{x} \text{ outside.}\end{gathered} \tag{5}$$

In particular

$$\begin{gathered}\mathbf{Q}_t^{\bar{x}v}(x)\;\mathbf{Q}_{\Lambda\setminus t}^{\bar{x}x}(y)\;\mathbf{Q}_\Lambda^{\bar{x}}(uv)=\mathbf{Q}_\Lambda^{\bar{x}}(xy)\;\mathbf{Q}_{\Lambda\setminus t}^{\bar{x}x}(v)\;\mathbf{Q}_t^{\bar{x}v}(u)\\ \text{for any } \Lambda\in\mathscr{E}_n,\ t\in\Lambda,\ x,u\in\mathscr{X}^t,\ y,v\in\mathscr{X}^{\Lambda\setminus t} \text{ and } \bar{x}\in\mathscr{X}^{\mathbb{Z}^\nu\setminus\Lambda}.\end{gathered} \tag{6}$$

(2) Conversely, if (6) is fulfilled, then $\mathcal{Q}$ is n-specification.

This theorem contains in particular the results of Theorem 8 and at the same time characterizes n-specifications.

Now, we can at last prove the above stated theorem about description of specifications.

Proof of Theorem 19. Let $n\in\mathbb{N}$, and let $q=\{\mathbf{q}_\Lambda^\bullet,\Lambda\in\mathscr{E}_n\}$ be a very weakly positive n-specification.

In order to prove the theorem it is sufficient to show, that there exist a unique $(n+1)$-specification $\mathcal{Q}$ containing q. Indeed, in this case $\mathcal{Q}$ is clearly very weakly positive too, and so we can conclude the proof by means of iteration.

First we prove the uniqueness: if there exists an $(n+1)$-specification $\mathcal{Q}=\{\mathbf{Q}_\Lambda^\bullet,\Lambda\in\mathscr{E}_{n+1}\}$ containing q, then it is the unique $(n+1)$-specification containing q. For each $\Lambda\in\mathscr{E}$ let us fix some point $\ell\in\Lambda$. If $|\Lambda|\leqslant n$, then clearly

$$\mathbf{Q}_\Lambda^{\bar{x}}(x)=\mathbf{q}_\Lambda^{\bar{x}}(x)\,. \tag{7}$$

Now let $|\Lambda|=n+1$ and $\bar{x}\in\mathscr{X}^{\mathbb{Z}^\nu\setminus\Lambda}$, and let $u\in\mathscr{X}^\Lambda$ be the configuration defined by $u_t=\theta(t,\Lambda\setminus t,\bar{x})$. Using (6) we have

$$\mathbf{Q}_\Lambda^{\bar{x}}(x)=\mathbf{Q}_\Lambda^{\bar{x}}(u)\,\frac{\mathbf{q}_\ell^{\bar{x}u_{\Lambda\setminus\ell}}(x_\ell)\;\mathbf{q}_{\Lambda\setminus\ell}^{\bar{x}x_\ell}(x_{\Lambda\setminus\ell})}{\mathbf{q}_\ell^{\bar{x}u_{\Lambda\setminus\ell}}(u_\ell)\;\mathbf{q}_{\Lambda\setminus\ell}^{\bar{x}x_\ell}(u_{\Lambda\setminus\ell})}\,. \tag{8}$$

Since $\sum_{y\in\mathscr{X}^\Lambda} \mathbf{Q}_\Lambda^{\overline{x}}(y)=1$, we get finally

$$\mathbf{Q}_\Lambda^{\overline{x}}(u)=\left(\sum_{y\in\mathscr{X}^\Lambda}\frac{\mathbf{q}_\ell^{\overline{x}u_{\Lambda\setminus\ell}}(y_\ell)\,\mathbf{q}_{\Lambda\setminus\ell}^{\overline{x}y_\ell}(y_{\Lambda\setminus\ell})}{\mathbf{q}_\ell^{\overline{x}u_{\Lambda\setminus\ell}}(u_\ell)\,\mathbf{q}_{\Lambda\setminus\ell}^{\overline{x}y_\ell}(u_{\Lambda\setminus\ell})}\right)^{-1}. \tag{9}$$

So, any $(n+1)$-specification containing $\boldsymbol{q}$ have necessarily the explicit form given by the formulas (7), (8) and (9), and hence the uniqueness is proved.

To conclude the prove of the theorem, it remains to verify that the $(n+1)$-system $\mathcal{Q}=\{\mathbf{Q}_\Lambda^\bullet,\ \Lambda\in\mathscr{E}_{n+1}\}$ defined by (7), (8) and (9) is indeed an $(n+1)$-specification. Applying Proposition 23 and taking into account that $\boldsymbol{q}$ is n-specification, it is sufficient to verify Dobrushin's consistency condition for $\mathbf{Q}_\Lambda^\bullet$ and $\mathbf{q}_{\Lambda\setminus t}^\bullet$ with $|\Lambda|=n+1$ only. Further, according to Proposition 22 this condition becomes

$$\mathbf{Q}_\Lambda^{\overline{x}}(xy)\,\mathbf{q}_{\Lambda\setminus t}^{\overline{x}x}(v)=\mathbf{Q}_\Lambda^{\overline{x}}(xv)\,\mathbf{q}_{\Lambda\setminus t}^{\overline{x}x}(y). \tag{10}$$

For the case $t=\ell$, using (8) we obtain

$$\begin{aligned}\mathbf{Q}_\Lambda^{\overline{x}}(xy)\,\mathbf{q}_{\Lambda\setminus\ell}^{\overline{x}x}(v) &= \mathbf{Q}_\Lambda^{\overline{x}}(u)\,\frac{\mathbf{q}_\ell^{\overline{x}u_{\Lambda\setminus\ell}}(x)\,\mathbf{q}_{\Lambda\setminus\ell}^{\overline{x}x}(y)}{\mathbf{q}_\ell^{\overline{x}u_{\Lambda\setminus\ell}}(u_\ell)\,\mathbf{q}_{\Lambda\setminus\ell}^{\overline{x}x}(u_{\Lambda\setminus\ell})}\,\mathbf{q}_{\Lambda\setminus\ell}^{\overline{x}x}(v)\\ &= \mathbf{Q}_\Lambda^{\overline{x}}(u)\,\frac{\mathbf{q}_\ell^{\overline{x}u_{\Lambda\setminus\ell}}(x)\,\mathbf{q}_{\Lambda\setminus\ell}^{\overline{x}x}(v)}{\mathbf{q}_\ell^{\overline{x}u_{\Lambda\setminus\ell}}(u_\ell)\,\mathbf{q}_{\Lambda\setminus\ell}^{\overline{x}x}(u_{\Lambda\setminus\ell})}\,\mathbf{q}_{\Lambda\setminus\ell}^{\overline{x}x}(y)\\ &= \mathbf{Q}_\Lambda^{\overline{x}}(xv)\,\mathbf{q}_{\Lambda\setminus\ell}^{\overline{x}x}(y),\end{aligned}$$

and so (10) is verified. Now, for the case of arbitrary $t\in\Lambda$, it is sufficient to show that the right-hand side of (8) does not depend on the choice of ℓ and apply the same argument.

This property is true due to the following chain of equalities

$$\begin{aligned}\frac{\mathbf{q}_\ell^{\overline{x}u_{\Lambda\setminus\ell}}(x_\ell)\,\mathbf{q}_{\Lambda\setminus\ell}^{\overline{x}x_\ell}(x_{\Lambda\setminus\ell})}{\mathbf{q}_\ell^{\overline{x}u_{\Lambda\setminus\ell}}(u_\ell)\,\mathbf{q}_{\Lambda\setminus\ell}^{\overline{x}x_\ell}(u_{\Lambda\setminus\ell})} &= \frac{\mathbf{q}_\ell^{\overline{x}u_{\Lambda\setminus\ell}}(x_\ell)\,\mathbf{q}_t^{\overline{x}x_\ell u_{\Lambda\setminus\ell\setminus t}}(x_t)\,\mathbf{q}_{\Lambda\setminus\ell\setminus t}^{\overline{x}x_\ell x_t}(x_{\Lambda\setminus\ell\setminus t})}{\mathbf{q}_\ell^{\overline{x}u_{\Lambda\setminus\ell}}(u_\ell)\,\mathbf{q}_t^{\overline{x}x_\ell u_{\Lambda\setminus\ell\setminus t}}(u_t)\,\mathbf{q}_{\Lambda\setminus\ell\setminus t}^{\overline{x}x_\ell x_t}(u_{\Lambda\setminus\ell\setminus t})}\\ &= \frac{\mathbf{q}_t^{\overline{x}u_{\Lambda\setminus t}}(x_t)\,\mathbf{q}_\ell^{\overline{x}x_t u_{\Lambda\setminus t\setminus\ell}}(x_\ell)\,\mathbf{q}_{\Lambda\setminus t\setminus\ell}^{\overline{x}x_t x_\ell}(x_{\Lambda\setminus t\setminus\ell})}{\mathbf{q}_t^{\overline{x}u_{\Lambda\setminus t}}(u_t)\,\mathbf{q}_\ell^{\overline{x}x_t u_{\Lambda\setminus t\setminus\ell}}(u_\ell)\,\mathbf{q}_{\Lambda\setminus t\setminus\ell}^{\overline{x}x_t x_\ell}(u_{\Lambda\setminus t\setminus\ell})}\\ &= \frac{\mathbf{q}_t^{\overline{x}u_{\Lambda\setminus t}}(x_t)\,\mathbf{q}_{\Lambda\setminus t}^{\overline{x}x_t}(x_{\Lambda\setminus t})}{\mathbf{q}_t^{\overline{x}u_{\Lambda\setminus t}}(u_t)\,\mathbf{q}_{\Lambda\setminus t}^{\overline{x}x_t}(u_{\Lambda\setminus t})}.\end{aligned}$$

The validity of these equalities in the case $n \geqslant 2$ is guarantied by Theorem 24. For $n = 1$ the first and the third equalities are trivial, and the second one follows from the definition of 1-specification. So, the theorem is proved. ∎

6. PROOF OF AUXILIARY RESULTS

Proof of Theorem 7. Let us suppose the contrary: there exists some $\boldsymbol{\alpha} \in \mathscr{X}^T$ such that $\mathbf{Q}_\Lambda^{\bar{x}\boldsymbol{\alpha}}(\boldsymbol{uv}) = 0$. Since $\mathbf{Q}_\Lambda^\bullet$ is consistent in Dobrushin's sense with $\mathbf{Q}_I^\bullet$, according to Proposition 22 we can write

$$\mathbf{Q}_\Lambda^{\bar{x}\boldsymbol{\alpha}}(\boldsymbol{uv})\, \mathbf{Q}_I^{\bar{x}\boldsymbol{\alpha}\boldsymbol{u}}(\boldsymbol{y}) = \mathbf{Q}_\Lambda^{\bar{x}\boldsymbol{\alpha}}(\boldsymbol{uy})\, \mathbf{Q}_I^{\bar{x}\boldsymbol{\alpha}\boldsymbol{u}}(\boldsymbol{v}).$$

Taking into account that $\boldsymbol{v}$ is a positivity point, we have $\mathbf{Q}_I^{\bar{x}\boldsymbol{\alpha}\boldsymbol{u}}(\boldsymbol{v}) > 0$, and hence $\mathbf{Q}_\Lambda^{\bar{x}\boldsymbol{\alpha}}(\boldsymbol{uy}) = 0$ for any $\boldsymbol{y} \in \mathscr{X}^I$.

Similarly, for any $\boldsymbol{y} \in \mathscr{X}^I$, from the relation

$$\mathbf{Q}_\Lambda^{\bar{x}\boldsymbol{\alpha}}(\boldsymbol{uy})\, \mathbf{Q}_J^{\bar{x}\boldsymbol{\alpha}\boldsymbol{y}}(\boldsymbol{x}) = \mathbf{Q}_\Lambda^{\bar{x}\boldsymbol{\alpha}}(\boldsymbol{xy})\, \mathbf{Q}_J^{\bar{x}\boldsymbol{\alpha}\boldsymbol{y}}(\boldsymbol{u}),$$

we get $\mathbf{Q}_\Lambda^{\bar{x}\boldsymbol{\alpha}}(\boldsymbol{xy}) = 0$ for any $\boldsymbol{x} \in \mathscr{X}^J$.

So $\mathbf{Q}_\Lambda^{\bar{x}\boldsymbol{\alpha}}(z) = 0$ for any $z \in \mathscr{X}^\Lambda$, which contradicts the fact that $\mathbf{Q}_\Lambda^{\bar{x}\boldsymbol{\alpha}}$ is a probability distribution. ∎

Proof of Theorem 8. This theorem clearly follows from the first assertion of Theorem 24 by substituting $A = C = t$, $B = D = s$, $\boldsymbol{x} = xy$ and $\boldsymbol{u} = uv$. ∎

Proof of Proposition 18. The necessity is trivial. In order to prove the sufficiency, let us first show that

$$\mathbf{Q}_t^{\bar{x}v^\circ}(x)\, \mathbf{Q}_s^{\bar{x}x}(y)\, \mathbf{Q}_t^{\bar{x}y}(u)\, \mathbf{Q}_s^{\bar{x}u}(v^\circ) = \mathbf{Q}_s^{\bar{x}u}(y)\, \mathbf{Q}_t^{\bar{x}y}(x)\, \mathbf{Q}_s^{\bar{x}x}(v^\circ)\, \mathbf{Q}_t^{\bar{x}v^\circ}(u)$$
$$\text{for any } t, s \in \mathbb{Z}^\nu,\ x, u \in \mathscr{X}^t,\ y \in \mathscr{X}^s,\ \bar{\boldsymbol{x}} \in \mathscr{X}^{\mathbb{Z}^\nu \setminus t \setminus s}, \tag{11}$$
$$\text{and for } v^\circ = \theta(s, t, \bar{\boldsymbol{x}}).$$

Using (3) we obtain

$$\mathbf{Q}_t^{\bar{x}v^\circ}(x)\, \mathbf{Q}_s^{\bar{x}x}(y)\, \mathbf{Q}_t^{\bar{x}y}(u^\circ)\, \mathbf{Q}_s^{\bar{x}u^\circ}(v^\circ) = \mathbf{Q}_s^{\bar{x}u^\circ}(y)\, \mathbf{Q}_t^{\bar{x}y}(x)\, \mathbf{Q}_s^{\bar{x}x}(v^\circ)\, \mathbf{Q}_t^{\bar{x}v^\circ}(u^\circ),$$
$$\mathbf{Q}_t^{\bar{x}v^\circ}(u)\, \mathbf{Q}_s^{\bar{x}u}(y)\, \mathbf{Q}_t^{\bar{x}y}(u^\circ)\, \mathbf{Q}_s^{\bar{x}u^\circ}(v^\circ) = \mathbf{Q}_s^{\bar{x}u^\circ}(y)\, \mathbf{Q}_t^{\bar{x}y}(u)\, \mathbf{Q}_s^{\bar{x}u}(v^\circ)\, \mathbf{Q}_t^{\bar{x}v^\circ}(u^\circ).$$

Suppose $\mathbf{Q}_s^{\bar{x}u^\circ}(y) > 0$. Then, if we cross-wise multiply these two equalities and cancel identical strictly positive terms, we get the necessary relation. Now suppose $\mathbf{Q}_s^{\bar{x}u^\circ}(y) = 0$. From the same equalities we get clearly $\mathbf{Q}_t^{\bar{x}v^\circ}(x)\, \mathbf{Q}_s^{\bar{x}x}(y) = 0$ and $\mathbf{Q}_t^{\bar{x}v^\circ}(u)\, \mathbf{Q}_s^{\bar{x}u}(y) = 0$, and so the property (11) is proved.

Further, using (11) we obtain

$$\mathbf{Q}_t^{\bar{x}v^\circ}(x)\, \mathbf{Q}_s^{\bar{x}x}(y)\, \mathbf{Q}_t^{\bar{x}y}(u)\, \mathbf{Q}_s^{\bar{x}u}(v^\circ) = \mathbf{Q}_s^{\bar{x}u}(y)\, \mathbf{Q}_t^{\bar{x}y}(x)\, \mathbf{Q}_s^{\bar{x}x}(v^\circ)\, \mathbf{Q}_t^{\bar{x}v^\circ}(u),$$
$$\mathbf{Q}_t^{\bar{x}v^\circ}(x)\, \mathbf{Q}_s^{\bar{x}x}(v)\, \mathbf{Q}_t^{\bar{x}v}(u)\, \mathbf{Q}_s^{\bar{x}u}(v^\circ) = \mathbf{Q}_s^{\bar{x}u}(v)\, \mathbf{Q}_t^{\bar{x}v}(x)\, \mathbf{Q}_s^{\bar{x}x}(v^\circ)\, \mathbf{Q}_t^{\bar{x}v^\circ}(u),$$

and so, applying once more the same argument we can conclude the proof of the proposition. ∎

Proof of Proposition 22. First suppose that $\mathbf{Q}_\Lambda^\bullet$ and $\mathbf{Q}_I^\bullet$ are consistent in Dobrushin's sense. Then

$$\begin{aligned}\mathbf{Q}_\Lambda^{\bar{x}}(\boldsymbol{x}\boldsymbol{y})\, \mathbf{Q}_I^{\bar{x}x}(\boldsymbol{v}) &= (\mathbf{Q}_\Lambda^{\bar{x}})_{\Lambda\setminus I}(\boldsymbol{x})\, \mathbf{Q}_I^{\bar{x}x}(\boldsymbol{y})\, \mathbf{Q}_I^{\bar{x}x}(\boldsymbol{v}) \\ &= (\mathbf{Q}_\Lambda^{\bar{x}})_{\Lambda\setminus I}(\boldsymbol{x})\, \mathbf{Q}_I^{\bar{x}x}(\boldsymbol{v})\, \mathbf{Q}_I^{\bar{x}x}(\boldsymbol{y}) \\ &= \mathbf{Q}_\Lambda^{\bar{x}}(\boldsymbol{x}\boldsymbol{v})\, \mathbf{Q}_I^{\bar{x}x}(\boldsymbol{y}),\end{aligned}$$

and so we have (4).

Now suppose (4). For any $\boldsymbol{v} \in \mathscr{X}^I$ we can write

$$\mathbf{Q}_\Lambda^{\bar{x}}(\boldsymbol{x}\boldsymbol{y})\, \mathbf{Q}_I^{\bar{x}x}(\boldsymbol{v}) = \mathbf{Q}_\Lambda^{\bar{x}}(\boldsymbol{x}\boldsymbol{v})\, \mathbf{Q}_I^{\bar{x}x}(\boldsymbol{y}).$$

Summing over $\boldsymbol{v}$ we obtain

$$\mathbf{Q}_\Lambda^{\bar{x}}(\boldsymbol{x}\boldsymbol{y}) = \sum_{\boldsymbol{v}\in\mathscr{X}^I} \mathbf{Q}_\Lambda^{\bar{x}}(\boldsymbol{x}\boldsymbol{v})\, \mathbf{Q}_I^{\bar{x}x}(\boldsymbol{y}) = (\mathbf{Q}_\Lambda^{\bar{x}})_{\Lambda\setminus I}(\boldsymbol{x})\, \mathbf{Q}_I^{\bar{x}x}(\boldsymbol{y}),$$

and so $\mathbf{Q}_\Lambda^\bullet$ and $\mathbf{Q}_I^\bullet$ are consistent in Dobrushin's sense. ∎

Proof of Proposition 23. The necessity is trivial. In order to prove the sufficiency, it is sufficient to show that the consistency in Dobrushin's sense is transitive, that is, if $J \subset I \subset \Lambda \in \mathscr{E}$, and if a Λ-kernel $\mathbf{Q}_\Lambda^\bullet$ is consistent with an I-kernel $\mathbf{Q}_I^\bullet$ which in its turn is consistent with a J-kernel $\mathbf{Q}_J^\bullet$, then $\mathbf{Q}_\Lambda^\bullet$ and $\mathbf{Q}_J^\bullet$ are also consistent.

Let $\boldsymbol{x} \in \mathscr{X}^{\Lambda \setminus J}$, let $\boldsymbol{y}, \boldsymbol{v} \in \mathscr{X}^J$ and let $\overline{\boldsymbol{x}} \in \mathscr{X}^{\mathbb{Z}^\nu \setminus \Lambda}$. Since $\mathbf{Q}_I^\bullet$ is consistent with $\mathbf{Q}_J^\bullet$, using Proposition 22, we have

$$\mathbf{Q}_I^{\overline{\boldsymbol{x}}\boldsymbol{x}_{\Lambda \setminus I}}(\boldsymbol{x}_{I \setminus J}\boldsymbol{y})\, \mathbf{Q}_J^{\overline{\boldsymbol{x}}\boldsymbol{x}}(\boldsymbol{v}) = \mathbf{Q}_I^{\overline{\boldsymbol{x}}\boldsymbol{x}_{\Lambda \setminus I}}(\boldsymbol{x}_{I \setminus J}\boldsymbol{v})\, \mathbf{Q}_J^{\overline{\boldsymbol{x}}\boldsymbol{x}}(\boldsymbol{y}).$$

Hence

$$\begin{aligned}&(\mathbf{Q}_\Lambda^{\overline{\boldsymbol{x}}})_{\Lambda \setminus I}(\boldsymbol{x}_{\Lambda \setminus I})\, \mathbf{Q}_I^{\overline{\boldsymbol{x}}\boldsymbol{x}_{\Lambda \setminus I}}(\boldsymbol{x}_{I \setminus J}\boldsymbol{y})\, \mathbf{Q}_J^{\overline{\boldsymbol{x}}\boldsymbol{x}}(\boldsymbol{v})\\ &\quad = (\mathbf{Q}_\Lambda^{\overline{\boldsymbol{x}}})_{\Lambda \setminus I}(\boldsymbol{x}_{\Lambda \setminus I})\, \mathbf{Q}_I^{\overline{\boldsymbol{x}}\boldsymbol{x}_{\Lambda \setminus I}}(\boldsymbol{x}_{I \setminus J}\boldsymbol{v})\, \mathbf{Q}_J^{\overline{\boldsymbol{x}}\boldsymbol{x}}(\boldsymbol{y}).\end{aligned}$$

Further, since $\mathbf{Q}_\Lambda^\bullet$ is consistent with $\mathbf{Q}_I^\bullet$ we obtain

$$\mathbf{Q}_\Lambda^{\overline{\boldsymbol{x}}}(\boldsymbol{x}\boldsymbol{y})\, \mathbf{Q}_J^{\overline{\boldsymbol{x}}\boldsymbol{x}}(\boldsymbol{v}) = \mathbf{Q}_\Lambda^{\overline{\boldsymbol{x}}}(\boldsymbol{x}\boldsymbol{v})\, \mathbf{Q}_J^{\overline{\boldsymbol{x}}\boldsymbol{x}}(\boldsymbol{y}),$$

and so, applying once more Proposition 22 we can conclude the proof of the proposition. ∎

Proof of Theorem 24. In order to carry out the proof we need the following two simple lemmas.

Lemma 25. Let $I, V \in \mathscr{E}$ such that $I \cap V = \emptyset$, put $\Lambda = I \cup V$, let $\overline{\boldsymbol{x}} \in \mathscr{X}^{\mathbb{Z}^\nu \setminus \Lambda}$, and let a Λ-kernel $\mathbf{Q}_\Lambda^\bullet$ be consistent in Dobrushin's sense with an I-kernel $\mathbf{Q}_I^\bullet$. If $\boldsymbol{u}$ is a p.p. of $\mathbf{Q}_I^\bullet$ under b.c. varying on V and equal to $\overline{\boldsymbol{x}}$ outside, then there exists a configuration $\boldsymbol{\gamma} \in \mathscr{X}^V$ such that $\mathbf{Q}_\Lambda^{\overline{\boldsymbol{x}}}(\boldsymbol{u}\boldsymbol{\gamma}) > 0$.

Proof. Let us suppose the contrary: for any configuration $\boldsymbol{\gamma} \in \mathscr{X}^V$ we have $\mathbf{Q}_\Lambda^{\overline{\boldsymbol{x}}}(\boldsymbol{u}\boldsymbol{\gamma}) = 0$. Since $\mathbf{Q}_\Lambda^\bullet$ is consistent in Dobrushin's sense with $\mathbf{Q}_I^\bullet$, for any $\boldsymbol{\alpha} \in \mathscr{X}^I$ and any $\boldsymbol{\gamma} \in \mathscr{X}^V$ according to Proposition 22 we can write

$$\mathbf{Q}_\Lambda^{\overline{\boldsymbol{x}}}(\boldsymbol{u}\boldsymbol{\gamma})\, \mathbf{Q}_I^{\overline{\boldsymbol{x}}\boldsymbol{\gamma}}(\boldsymbol{\alpha}) = \mathbf{Q}_\Lambda^{\overline{\boldsymbol{x}}}(\boldsymbol{\alpha}\boldsymbol{\gamma})\, \mathbf{Q}_I^{\overline{\boldsymbol{x}}\boldsymbol{\gamma}}(\boldsymbol{u}),$$

and hence, taking into account that $\boldsymbol{u}$ is a positivity point we obtain the equality $\mathbf{Q}_\Lambda^{\overline{\boldsymbol{x}}}(\boldsymbol{\alpha}\boldsymbol{\gamma}) = 0$.

So $\mathbf{Q}_\Lambda^{\overline{\boldsymbol{x}}}(z) = 0$ for any $z \in \mathscr{X}^\Lambda$, which contradicts the fact that $\mathbf{Q}_\Lambda^{\overline{\boldsymbol{x}}}$ is a probability distribution. ∎

Lemma 26. Let $\Lambda \in \mathscr{E}$ and $I \subset \Lambda$, let $\overline{\boldsymbol{x}} \in \mathscr{X}^{\mathbb{Z}^\nu \setminus \Lambda}$, and let a Λ-kernel $\mathbf{Q}_\Lambda^\bullet$ be consistent in Dobrushin's sense with an I-kernel $\mathbf{Q}_I^\bullet$. If for some $\boldsymbol{x} \in \mathscr{X}^{\Lambda \setminus I}$ and $\boldsymbol{y}, \boldsymbol{v} \in \mathscr{X}^I$ we have $\mathbf{Q}_\Lambda^{\overline{\boldsymbol{x}}}(\boldsymbol{x}\boldsymbol{y}) = 0$ and $\mathbf{Q}_\Lambda^{\overline{\boldsymbol{x}}}(\boldsymbol{x}\boldsymbol{v}) > 0$, then $\mathbf{Q}_I^{\overline{\boldsymbol{x}}\boldsymbol{x}}(\boldsymbol{y}) = 0$.

Proof. Since $\mathbf{Q}_\Lambda^\bullet$ is consistent in Dobrushin's sense with $\mathbf{Q}_I^\bullet$, according to Proposition 22 we can write

$$\mathbf{Q}_\Lambda^{\bar{x}}(xy)\,\mathbf{Q}_I^{\bar{x}x}(v)=\mathbf{Q}_\Lambda^{\bar{x}}(xv)\,\mathbf{Q}_I^{\bar{x}x}(y),$$

and so, taking into account that $\mathbf{Q}_\Lambda^{\bar{x}}(xy)=0$ and $\mathbf{Q}_\Lambda^{\bar{x}}(xv)>0$, we obtain immediately $\mathbf{Q}_I^{\bar{x}x}(y)=0$. ∎

Now we turn to the proof of Theorem 24. First let us suppose that $\mathcal{Q}$ is n-specification and prove the property (5). For convenience of notations let us denote $\Lambda=A\cup B=C\cup D$. According to Proposition 22, we have

$$\mathbf{Q}_\Lambda^{\bar{x}}(x)\,\mathbf{Q}_B^{\bar{x}x_A}(u_B)=\mathbf{Q}_\Lambda^{\bar{x}}(x_Au_B)\,\mathbf{Q}_B^{\bar{x}x_A}(x_B).$$

Multiplying this equality by $\mathbf{Q}_A^{\bar{x}u_B}(u_A)$ and using Proposition 22 on the right hand side, we obtain

$$\mathbf{Q}_\Lambda^{\bar{x}}(x)\,\mathbf{Q}_B^{\bar{x}x_A}(u_B)\,\mathbf{Q}_A^{\bar{x}u_B}(u_A)=\mathbf{Q}_\Lambda^{\bar{x}}(u)\,\mathbf{Q}_A^{\bar{x}u_B}(x_A)\,\mathbf{Q}_B^{\bar{x}x_A}(x_B). \tag{12}$$

In the same way we have

$$\mathbf{Q}_\Lambda^{\bar{x}}(x)\,\mathbf{Q}_C^{\bar{x}x_D}(u_C)\,\mathbf{Q}_D^{\bar{x}u_C}(u_D)=\mathbf{Q}_\Lambda^{\bar{x}}(u)\,\mathbf{Q}_D^{\bar{x}u_C}(x_D)\,\mathbf{Q}_C^{\bar{x}x_D}(x_C). \tag{13}$$

Suppose first $\mathbf{Q}_\Lambda^{\bar{x}}(x)>0$ and $\mathbf{Q}_\Lambda^{\bar{x}}(u)>0$. Then, if we cross-wise multiply the equalities (12) and (13) and cancel identical strictly positive terms, we get the relation claimed in (5).

Suppose now $\mathbf{Q}_\Lambda^{\bar{x}}(x)=0$ and $\mathbf{Q}_\Lambda^{\bar{x}}(u)>0$. Then from (12) and (13) we have $\mathbf{Q}_A^{\bar{x}u_B}(x_A)\,\mathbf{Q}_B^{\bar{x}x_A}(x_B)=0$ and $\mathbf{Q}_D^{\bar{x}u_C}(x_D)\,\mathbf{Q}_C^{\bar{x}x_D}(x_C)=0$ correspondingly, and so, the necessary relation is still valid. Similar considerations show that it remains valid for the case $\mathbf{Q}_\Lambda^{\bar{x}}(x)>0$ and $\mathbf{Q}_\Lambda^{\bar{x}}(u)=0$.

Suppose finally $\mathbf{Q}_\Lambda^{\bar{x}}(x)=0$ and $\mathbf{Q}_\Lambda^{\bar{x}}(u)=0$. Since u_C is a positivity point, due to Lemma 25 there exists some configuration $y\in\mathcal{X}^D$ such that $\mathbf{Q}_\Lambda^{\bar{x}}(u_Cy)>0$. The latter inequality together with $\mathbf{Q}_\Lambda^{\bar{x}}(u)=0$ implies according to Lemma 26 that $\mathbf{Q}_D^{\bar{x}u_C}(u_D)=0$, and so, the left hand side of the relation claimed in (5) vanishes. It remains to show that the right hand side of this relation vanishes too. Indeed, if $\mathbf{Q}_\Lambda^{\bar{x}}(u_Cx_D)=0$ then taking into consideration that $\mathbf{Q}_\Lambda^{\bar{x}}(u_Cy)>0$ and using Lemma 26 we obtain $\mathbf{Q}_D^{\bar{x}u_C}(x_D)=0$, and if $\mathbf{Q}_\Lambda^{\bar{x}}(u_Cx_D)>0$ then taking into account that $\mathbf{Q}_\Lambda^{\bar{x}}(x)=0$ we get $\mathbf{Q}_C^{\bar{x}x_D}(x_C)=0$.

So, the property (5) is established. In order to prove (6) it is sufficient now to put $A=t$, $B=\Lambda\setminus t$, $C=\emptyset$, $D=\Lambda$, $\boldsymbol{x}=x\boldsymbol{y}$ and $\boldsymbol{u}=u\boldsymbol{v}$ in (5), and note that $\boldsymbol{u}_\emptyset=\boldsymbol{\emptyset}$ is indeed a p.p. of $\mathbf{Q}^{\bullet}_{\emptyset}$ under b.c. varying on Λ and equal to $\overline{\boldsymbol{x}}$ outside.

It remains to prove the second part of the theorem. Suppose (6) is fulfilled, take some $\Lambda\in\mathscr{E}_n$, $t\in\Lambda$, $x\in\mathscr{X}^t$, $\boldsymbol{y},\boldsymbol{v}\in\mathscr{X}^{\Lambda\setminus t}$ and $\overline{\boldsymbol{x}}\in\mathscr{X}^{\mathbb{Z}^\nu\setminus\Lambda}$, and let us show that

$$\mathbf{Q}^{\overline{\boldsymbol{x}}}_{\Lambda}(x\boldsymbol{y})\,\mathbf{Q}^{\overline{\boldsymbol{x}}x}_{\Lambda\setminus t}(\boldsymbol{v})=\mathbf{Q}^{\overline{\boldsymbol{x}}}_{\Lambda}(x\boldsymbol{v})\,\mathbf{Q}^{\overline{\boldsymbol{x}}x}_{\Lambda\setminus t}(\boldsymbol{y}). \tag{14}$$

Suppose first $\mathbf{Q}^{\overline{\boldsymbol{x}}\boldsymbol{v}}_{t}(x)>0$. Then, taking $u=x$ in (6) and canceling the term $\mathbf{Q}^{\overline{\boldsymbol{x}}\boldsymbol{v}}_{t}(x)$ we obtain (14).

Suppose now $\mathbf{Q}^{\overline{\boldsymbol{x}}\boldsymbol{y}}_{t}(x)>0$. Then, interchanging the positions of $\boldsymbol{y}$ and $\boldsymbol{v}$ in (6), taking $u=x$ and canceling the term $\mathbf{Q}^{\overline{\boldsymbol{x}}\boldsymbol{y}}_{t}(x)$ we obtain (14).

Suppose finally $\mathbf{Q}^{\overline{\boldsymbol{x}}\boldsymbol{v}}_{t}(x)=0$ and $\mathbf{Q}^{\overline{\boldsymbol{x}}\boldsymbol{y}}_{t}(x)=0$. Taking in consideration the first equality, we can show that the left hand side of the relation (14) vanishes. Indeed, since $\mathbf{Q}^{\overline{\boldsymbol{x}}\boldsymbol{v}}_{t}$ is probability distribution, we can chose $u\in\mathscr{X}^t$ such that $\mathbf{Q}^{\overline{\boldsymbol{x}}\boldsymbol{v}}_{t}(u)>0$, and using (6) we clearly obtain $\mathbf{Q}^{\overline{\boldsymbol{x}}}_{\Lambda}(x\boldsymbol{y})\,\mathbf{Q}^{\overline{\boldsymbol{x}}x}_{\Lambda\setminus t}(\boldsymbol{v})=0$. Similarly, the second equality implies that the right hand side of the relation (14) vanishes, and so this relation is proved.

Now, in order to conclude the proof of the theorem it remains to apply consecutively Propositions 22 and 23. ▮

REFERENCES

1. H.-O. Georgii, *Gibbs measures and phase transitions*, (de Gruyter, Berlin, 1988).
2. M. B. Averintsev, A way of describing random fields with a discrete argument, *Probl. Inf. Transm.* **6**:169–175 (1970).
3. M. B. Averintsev, Gibbsian description of random fields whose conditional probabilities may vanish, *Probl. Inf. Transm.* **11**:326–334 (1975).
4. O. K. Kozlov, Gibbsian description of a system of random quantities, *Probl. Inf. Transm.* **10**:258–265 (1974).
5. W. G. Sullivan, Potentials for almost Markovian random fields, *Comm. Math. Phys.* **33**:61–74 (1973).
6. R. L. Dobrushin, The description of a random field by means of conditional probabilities and conditions of its regularity, *Theory Probab. Appl.* **13**:197–224 (1968).
7. R. L. Dobrushin, Gibbsian random fields for lattice systems with pairwise interaction, *Funct. Anal. Appl.* **2**:292–301 (1968).
8. R. L. Dobrushin, The problem of uniqueness of a Gibbs random field and the problem of phase transitions, *Funct. Anal. Appl.* **2**:302–312 (1968).
9. S. Dachian and B. S. Nahapetian, *An approach towards description of random fields*, Preprint 98–5, Laboratoire de Statistique et Processus, (Université du Maine, 1998).
10. S. Dachian, B. S. Nahapetian, Description of random fields by means of one-point conditional distributions and some applications, *Markov Process. Related Fields* **7**:193–214 (2001).

11. M. B. Averintsev, Description of Markovian random fields by Gibbsian conditional probabilities, *Theory Probab. Appl*. **17**:20–33 (1972).
12. R. G. Flood and W. G. Sullivan, Consistency of random field specifications, *Z. Wahrscheinlichkeitstheorie verw. Gebiete* **53**:147–156 (1980).
13. A. C. D. Van Enter, R. Fernández, A. D. Sokal, Regularity properties and pathologies of position-space renormalization-group transformations: scope and limitations of Gibbsian theory, *J. Statist. Phys*. **72**:879–1167 (1993).

Markov Processes Relat. Fields **15**, 81–104 (2009)

Markov Processes and **Related Fields**

On Gibbsianness of Random Fields

S. Dachian[1] **and B.S. Nahapetian**[2]

[1] Laboratoire de Mathématiques, Université Blaise Pascal (Clermont-Ferrand 2), 63177 Aubière CEDEX, France. E-mail: Serguei.Dachian@math.univ-bpclermont.fr
[2] Institute of Mathematics, National Academy of Sciences of Armenia, 24-b Av. M. Bagramian, 375019 Yerevan, Armenia. E-mail: nahapet@instmath.sci.am

Received September 10, 2007, revised September 17, 2008

Abstract. The problem of characterization of Gibbs random fields is considered. Various Gibbsianness criteria are obtained using the earlier developed one-point framework which in particular allows to describe random fields by means of either one-point conditional or one-point finite-conditional distributions. The main outcome are the criteria in terms of one-point finite-conditional distribution. On the basis of one of the criteria a probabilistically explicit definition of Gibbs random field is given and the development of an alternative approach to the Gibbs theory is started.

Keywords: Gibbsianness, Gibbs random fields, Gibbsian specifications, one-point conditional distribution, one-point finite-conditional distribution

AMS Subject Classification: Primary 60G60, Secondary 60K35

Introduction

The classes of processes considered in the theory of random processes are usually characterized by some properties of their finite-dimensional or conditional distributions. However in practice, the study of a particular class usually goes through some representation theorem expressing processes in terms of simple and convenient objects, such as transition matrices for Markov chains, characteristic functions for processes with independent increments, spectral functions for stationary processes, and so on.

The situation is quite different for the class of Gibbs random fields. Historically, instead of being characterized by some properties of their finite-dimensional or conditional distributions, Gibbs random fields have been defined directly by the well-known representation of their conditional distributions in terms of

potentials. And only afterwards the problem of internal characterization of Gibbs random fields was considered.

It was shown by Kozlov [20] and Sullivan [24] that Gibbs random fields (with uniformly convergent potentials) can be characterized by strict positivity and quasilocality of their conditional distributions. More precisely, in order for a random field to be Gibbsian, its conditional distribution (the system of finite-volume conditional probabilities with conditions on the entire exterior) must have a version which is a strictly positive quasilocal specification. As we see, this criterion imposes conditions on an object (conditional distribution) which is neither unambiguously defined (is defined up to a set of probability zero), nor constructive (its elements are indexed by infinite-dimensional boundary conditions) and, in addition, does not always determine the random field uniquely (phase transitions). In our opinion, it is preferable that a characterization be in terms of an object which does not have these features.

As a matter of fact, such characterization already exists for the subclass of Gibbs random fields with real-valued finite-range potentials. It was shown by Averintsev [1–3] and Sullivan [23] that these random fields are characterized by strict positivity and Markov properties. Note that in the strictly positive case, the Markov property can be formulated in terms of one-point finite-conditional distribution (the system of single-site conditional probabilities with finite-volume conditions). This object is defined unambiguously and in constructive manner (its elements are ratios of finite-dimensional probabilities). Moreover, according to Dalalyan and Nahapetian [7], it uniquely determines (can be identified with) the random field.

The aim of the present work is to characterize Gibbs random fields by some properties of their one-point finite-conditional distributions in the general case of uniformly convergent potentials. It is worth mentioning that such characterization is very natural in light of and was made possible due to the one-point framework developed in some recent papers. Namely, an approach towards description of random fields by means of one-point conditional distributions (the system of single-site conditional probabilities with conditions on the entire exterior) was developed by the authors in [4–6] (see also Fernández and Maillard [14, 15]). Later on, a closely related and in some way complementary description of random fields based on one-point finite-conditional distributions was proposed in [7].

The main outcome of the present work are random field Gibbsianness criteria in terms of one-point finite-conditional distribution. These criteria deal with an unambiguously defined constructive object and allow us to start to develop an alternative approach to the Gibbs theory by giving (on the basis of one of the criteria) a probabilistically explicit definition of Gibbs random field.

The plan of the paper is as follows. The necessary notations and prerequisites are given in Section 1, the one-point framework is presented in Section 2,

the random field Gibbsianness criteria are established in Section 3 and the alternative approach to the Gibbs theory is introduced in Section 4.

1. Preliminaries

In this section we briefly recall some necessary notions and facts from the theory of Gibbs random fields.

1.1. Random fields

We consider random fields on the ν-dimensional integer lattice $\mathbb{Z}^\nu$ (or, more generally, on any countable set $\mathbb{L}$), i.e. probability measures $\mathbf{P}$ on $(\mathscr{X}^{\mathbb{Z}^\nu}, \mathscr{F}^{\mathbb{Z}^\nu})$ where $(\mathscr{X}, \mathscr{F})$ is some measurable space of values on single sites (*state space*). Usually the space $\mathscr{X}$ is assumed to be endowed with some topology $\mathscr{T}$, and $\mathscr{F}$ is assumed to be the Borel σ-algebra for this topology. In this work we concentrate on the case when $\mathscr{X}$ is finite, $\mathscr{T}$ is the discrete topology, and $\mathscr{F}$ is the total σ-algebra, that is, $\mathscr{F} = \mathscr{T} = \mathrm{part}(\mathscr{X})$.

For any $S \subset \mathbb{Z}^\nu$, we denote by $\mathscr{E}(S)$ the set of all finite subsets of S, that is, we put $\mathscr{E}(S) = \{\Lambda \subset S : |\Lambda| < \infty\}$ where $|\Lambda|$ is the number of points of the set Λ. For convenience of notation we will omit braces for one-point sets, that is, will write t instead of $\{t\}$. We put also $\mathscr{E}^*(S) = \mathscr{E}(S) \setminus \{\varnothing\}$. For $S = \mathbb{Z}^\nu$ we write $\mathscr{E} = \mathscr{E}(\mathbb{Z}^\nu)$ and $\mathscr{E}^* = \mathscr{E}^*(\mathbb{Z}^\nu)$.

For any $S \subset \mathbb{Z}^\nu$, the space $\mathscr{X}^S$ is the space of all configurations on S. If $S = \varnothing$, we assume that the space $\mathscr{X}^\varnothing = \{\boldsymbol{\varnothing}\}$ where $\boldsymbol{\varnothing}$ is the empty configuration. For any $T, S \subset \mathbb{Z}^\nu$ such that $T \subset S$ and any configuration $\boldsymbol{x} = \{x_t,\ t \in S\}$ on S, we denote by $\boldsymbol{x}_T$ the *subconfiguration* (*restriction*) of $\boldsymbol{x}$ on T defined by $\boldsymbol{x}_T = \{x_t,\ t \in T\}$. For any $T, S \subset \mathbb{Z}^\nu$ such that $T \cap S = \varnothing$ and any configurations $\boldsymbol{x}$ on T and $\boldsymbol{y}$ on S, we denote by $\boldsymbol{xy}$ the *concatenation* of $\boldsymbol{x}$ and $\boldsymbol{y}$, that is, the configuration on $T \cup S$ equal to $\boldsymbol{x}$ on T and to $\boldsymbol{y}$ on S. For any configuration $\boldsymbol{x} \in \mathscr{X}^S$, the set $S \subset \mathbb{Z}^\nu$ will be called *support* of $\boldsymbol{x}$ and we will write $S = \mathfrak{S}(\boldsymbol{x})$. For any $\Lambda \in \mathscr{E}$, we denote

$$\widetilde{\mathscr{X}^\Lambda} = \bigcup_{\widetilde{\Lambda} \in \mathscr{E}^*(\Lambda^c)} \mathscr{X}^{\widetilde{\Lambda}}$$

the space of all configurations with non-empty finite support contained in the exterior of Λ.

For any $S \subset \mathbb{Z}^\nu$, a probability distribution on $\mathscr{X}^S$ will be denoted by $\mathbf{P}_S$. Note that if $S = \varnothing$ there exists only one probability distribution $\mathbf{P}_\varnothing(\boldsymbol{\varnothing}) = 1$. For any $T, S \subset \mathbb{Z}^\nu$ such that $T \subset S$ and any $\mathbf{P}_S$, we denote by $(\mathbf{P}_S)_T$ the *marginal distribution* (*restriction*) of $\mathbf{P}_S$ on T. If $\Lambda \in \mathscr{E}$ and $I \subset \Lambda$, we can

write $\mathbf{P}_\Lambda = \{\mathbf{P}_\Lambda(\boldsymbol{x}),\ \boldsymbol{x} \in \mathscr{X}^\Lambda\}$ and

$$(\mathbf{P}_\Lambda)_I(\boldsymbol{x}) = \sum_{\boldsymbol{y} \in \mathscr{X}^{\Lambda \setminus I}} \mathbf{P}_\Lambda(\boldsymbol{x}\boldsymbol{y}), \quad \boldsymbol{x} \in \mathscr{X}^I.$$

Any random field $\mathbf{P}$ on $\mathbb{Z}^\nu$ is uniquely determined by (can be identified with) the system $\{\mathbf{P}_\Lambda,\ \Lambda \in \mathscr{E}\}$ of its *finite-dimensional distributions* which are consistent in the sense that for any $\Lambda \in \mathscr{E}$ and $I \subset \Lambda$ we have $(\mathbf{P}_\Lambda)_I = \mathbf{P}_I$.

Finally, a random field $\mathbf{P}$ will be called *strictly positive* if for any $\Lambda \in \mathscr{E}$ the finite-dimensional distribution $\mathbf{P}_\Lambda$ is *strictly positive*, that is, $\mathbf{P}_\Lambda(\boldsymbol{x}) > 0$ for all $\boldsymbol{x} \in \mathscr{X}^\Lambda$. The set of all strictly positive random fields will be denoted $\mathcal{P}$.

1.2. Finite-conditional and conditional distributions of random fields

Let $\mathbf{P}$ be some random field. For any $\Lambda \in \mathscr{E}$, we denote by $\mathbf{P}_{\mathscr{E}^*(\Lambda^{\mathrm{c}})}$ the measure on $\widetilde{\mathscr{X}^\Lambda}$ whose projection on $\mathscr{X}^{\widetilde{\Lambda}}$ is $\mathbf{P}_{\widetilde{\Lambda}}$ for any $\widetilde{\Lambda} \in \mathscr{E}^*(\Lambda^{\mathrm{c}})$, that is, $\mathbf{P}_{\mathscr{E}^*(\Lambda^{\mathrm{c}})}$ is the direct sum of the measures $\mathbf{P}_{\widetilde{\Lambda}}$.

For all $\Lambda \in \mathscr{E}$, the ratios

$$\boldsymbol{q}_\Lambda^{\widetilde{\boldsymbol{x}}}(\boldsymbol{x}) = \frac{\mathbf{P}_{\Lambda \cup \mathfrak{S}(\widetilde{\boldsymbol{x}})}(\boldsymbol{x}\widetilde{\boldsymbol{x}})}{\mathbf{P}_{\mathfrak{S}(\widetilde{\boldsymbol{x}})}(\widetilde{\boldsymbol{x}})}\ , \quad \boldsymbol{x} \in \mathscr{X}^\Lambda,$$

exist for $\mathbf{P}_{\mathscr{E}^*(\Lambda^{\mathrm{c}})}$-almost all $\widetilde{\boldsymbol{x}} \in \widetilde{\mathscr{X}^\Lambda}$. Any system

$$\widetilde{\boldsymbol{\mathcal{Q}}} = \{\ \mathbf{Q}_\Lambda^{\widetilde{\boldsymbol{x}}},\ \Lambda \in \mathscr{E} \text{ and } \widetilde{\boldsymbol{x}} \in \widetilde{\mathscr{X}^\Lambda}\}$$

of probability distributions such that for every $\Lambda \in \mathscr{E}$ we have $\mathbf{Q}_\Lambda^{\widetilde{\boldsymbol{x}}} = \boldsymbol{q}_\Lambda^{\widetilde{\boldsymbol{x}}}$ for $\mathbf{P}_{\mathscr{E}^*(\Lambda^{\mathrm{c}})}$-almost all $\widetilde{\boldsymbol{x}} \in \widetilde{\mathscr{X}^\Lambda}$ will be called *finite-conditional distribution* of the random field $\mathbf{P}$. The subsystem of $\widetilde{\boldsymbol{\mathcal{Q}}}$ consisting of single-site distributions $\big(|\Lambda| = 1\big)$ will be called *one-point finite-conditional distribution* of $\mathbf{P}$. Note that in general a random field may have many *versions* both of finite-conditional and one-point finite-conditional distributions. However, for strictly positive random fields these distributions are uniquely determined and strictly positive (consist of strictly positive elements only). Note also that it is not difficult to check that if a random field $\mathbf{P}$ has a strictly positive version of (one-point) finite-conditional distribution, then $\mathbf{P}$ is necessarily strictly positive itself.

Further, for all $\Lambda \in \mathscr{E}$, the limits

$$\boldsymbol{q}_\Lambda^{\overline{\boldsymbol{x}}}(\boldsymbol{x}) = \lim_{\widetilde{\Lambda} \uparrow \mathbb{Z}^\nu \setminus \Lambda} \boldsymbol{q}_\Lambda^{\overline{\boldsymbol{x}}_{\widetilde{\Lambda}}}(\boldsymbol{x})\ , \quad \boldsymbol{x} \in \mathscr{X}^\Lambda,$$

exist for $\mathbf{P}_{\Lambda^{\mathrm{c}}}$-almost all $\overline{\boldsymbol{x}} \in \mathscr{X}^{\Lambda^{\mathrm{c}}}$. Any system

$$\boldsymbol{\mathcal{Q}} = \{\ \mathbf{Q}_\Lambda^{\overline{\boldsymbol{x}}}, \quad \Lambda \in \mathscr{E} \text{ and } \overline{\boldsymbol{x}} \in \mathscr{X}^{\Lambda^{\mathrm{c}}}\}$$

of probability distributions such that for every $\Lambda \in \mathscr{E}$ we have $\mathbf{Q}_\Lambda^{\overline{\boldsymbol{x}}} = \boldsymbol{q}_\Lambda^{\overline{\boldsymbol{x}}}$ for $\mathbf{P}_{\Lambda^c}$-almost all $\overline{\boldsymbol{x}} \in \mathscr{X}^{\Lambda^c}$ will be called *conditional distribution* of the random field $\mathbf{P}$. The subsystem of $\boldsymbol{\mathcal{Q}}$ consisting of single-site distributions will be called *one-point conditional distribution* of $\mathbf{P}$. Note that in general a random field $\mathbf{P}$ may have many *versions* both of conditional and one-point conditional distributions (even if $\mathbf{P}$ is strictly positive). Note also that if a random field $\mathbf{P}$ has a strictly positive (consisting of strictly positive elements only) version of (one-point) conditional distribution, then $\mathbf{P}$ is necessarily strictly positive itself. For the case of conditional distribution this fact is well-known, while for the case of one-point conditional distribution we refer to the Proposition 3.2 below.

Concluding this section let us emphasize that random field's (one-point) finite-conditional distribution contains more information about the random field than its (one-point) conditional distribution. Indeed, the latter can be clearly deduced from the former (by passing to the limit), while the converse is not so clear. Moreover, it is not true in general, since in the strictly positive case, the (one-point) finite-conditional distribution determines the random field uniquely (see Section 2.2), while the (one-point) conditional distribution does not always do so (phase transitions). All this becomes particularly apparent in the Markov case, when (one-point) conditional distributions can be considered as subsystems of (one-point) finite-conditional distributions. Indeed, let $\mathbf{P}$ be a Markov random field and let $\partial\Lambda$ denote the neighborhood of the set Λ. As we have $\mathbf{Q}_\Lambda^{\overline{\boldsymbol{x}}} = \mathbf{Q}_\Lambda^{\overline{\boldsymbol{x}}_{\partial\Lambda}}$, the elements of the (one-point) conditional distribution of $\mathbf{P}$ can be considered as elements of the (one-point) finite-conditional distribution of $\mathbf{P}$. However, not all the elements of the latter correspond to the elements of the former, but only the elements $\mathbf{Q}_\Lambda^{\widetilde{\boldsymbol{x}}}$ such that $\mathfrak{S}(\widetilde{\boldsymbol{x}}) \supset \partial\Lambda$.

1.3. Description of random fields by means of conditional distributions

The well-known description of random fields by means of conditional distributions introduced by Dobrushin in [8–10] is carried out in terms of specifications. A system

$$\boldsymbol{\mathcal{Q}} = \left\{ \mathbf{Q}_\Lambda^{\overline{\boldsymbol{x}}}, \quad \Lambda \in \mathscr{E} \text{ and } \overline{\boldsymbol{x}} \in \mathscr{X}^{\Lambda^c} \right\}$$

of probability distributions is called *specification* if

$$\mathbf{Q}_\Lambda^{\overline{\boldsymbol{x}}}(\boldsymbol{x}\boldsymbol{y}) = \left(\mathbf{Q}_\Lambda^{\overline{\boldsymbol{x}}}\right)_{\Lambda\setminus I}(\boldsymbol{x})\ \mathbf{Q}_I^{\overline{\boldsymbol{x}}\boldsymbol{x}}(\boldsymbol{y}) \tag{1.1}$$

$$\text{for all } \Lambda \in \mathscr{E},\ I \subset \Lambda,\ \boldsymbol{x} \in \mathscr{X}^{\Lambda\setminus I},\ \boldsymbol{y} \in \mathscr{X}^I \text{ and } \overline{\boldsymbol{x}} \in \mathscr{X}^{\Lambda^c}.$$

Note that any version of conditional distribution of a random field $\mathbf{P}$ satisfies a condition somewhat weaker than (1.1), where $\mathbf{P}_{\Lambda^c}$-almost all (and not necessarily all) $\overline{\boldsymbol{x}} \in \mathscr{X}^{\Lambda^c}$ are considered. However, any random field possesses at least one version of conditional distribution being a specification (see [18, 21] and [22]).

One of the main goals of Dobrushin's theory is to study the set of all random fields *compatible* with a given specification, that is, having this specification as a version of conditional distribution. The best-known sufficient conditions for existence and for uniqueness of random fields compatible with a given specification are quasilocality and Dobrushin's uniqueness conditions respectively. The first one will play an important role in our considerations, so we recall it below.

Let $S \subset \mathbb{Z}^\nu$. A real-valued function g on $\mathscr{X}^S$ is called *quasilocal* if

$$\lim_{\Lambda \uparrow S} \sup_{\boldsymbol{x},\boldsymbol{y} \in \mathscr{X}^S \colon \boldsymbol{x}_\Lambda = \boldsymbol{y}_\Lambda} \big| g(\boldsymbol{x}) - g(\boldsymbol{y}) \big| = 0,$$

or equivalently if g is a uniform limit of functions depending only on values of configuration on finite sets of sites (*local* functions). Note also that the quasilocality is nothing but continuity with respect to the topology $\mathscr{T}^S$ and, taking into account that $\mathscr{X}^S$ is compact, the strict positivity and uniform nonnullness conditions are equivalent for quasilocal functions.

A specification $\boldsymbol{\mathcal{Q}} = \big\{ \mathbf{Q}_\Lambda^{\overline{\boldsymbol{x}}},\ \Lambda \in \mathscr{E} \text{ and } \overline{\boldsymbol{x}} \in \mathscr{X}^{\Lambda^{\mathrm{c}}} \big\}$ is called (*quasi*)*local* if for any $\Lambda \in \mathscr{E}$ and $\boldsymbol{x} \in \mathscr{X}^\Lambda$ the function $\overline{\boldsymbol{x}} \mapsto \mathbf{Q}_\Lambda^{\overline{\boldsymbol{x}}}(\boldsymbol{x})$ on $\mathscr{X}^{\Lambda^{\mathrm{c}}}$ is (quasi)local.

Finally, a specification will be called *strictly positive* if all its elements are strictly positive.

1.4. Gibbs random fields and Gibbsian specifications

The main object of consideration of the present paper are Gibbs random fields. They are defined in terms of Gibbsian specifications, which in turn are defined in terms of potentials.

Any function Φ on $\widetilde{\mathscr{X}^{\varnothing}}$ taking values in $\mathbb{R} \cup \{+\infty\}$ is called (*interaction*) *potential*. A potential Φ is called *convergent* if it is real-valued and the series

$$\sum_{\widetilde{J} \in \mathscr{E}(t^{\mathrm{c}})} \Phi\big(x\, \overline{\boldsymbol{x}}_{\widetilde{J}}\big) \tag{1.2}$$

converge for all $t \in \mathbb{Z}^\nu$, $x \in \mathscr{X}^t$ and $\overline{\boldsymbol{x}} \in \mathscr{X}^{t^{\mathrm{c}}}$.

A potential Φ is called *uniformly convergent* if it is convergent and the convergence in (1.2) is uniform with respect to $\overline{\boldsymbol{x}}$.

A potential Φ is called *finite-range potential* if for any $t \in \mathbb{Z}^\nu$ there exist only a finite number of sets $\widetilde{J} \in \mathscr{E}(t^{\mathrm{c}})$ such that $\Phi \not\equiv 0$ on $\mathscr{X}^{t \cup \widetilde{J}}$. Note that any real-valued finite-range potential is uniformly convergent.

For an arbitrary convergent potential Φ one can construct the specification $\boldsymbol{\mathcal{Q}} = \{\mathbf{Q}_\Lambda^{\overline{\boldsymbol{x}}},\ \Lambda \in \mathscr{E} \text{ and } \overline{\boldsymbol{x}} \in \mathscr{X}^{\Lambda^{\mathrm{c}}}\}$ given by Gibbs formulae

$$\mathbf{Q}_\Lambda^{\overline{\boldsymbol{x}}}(\boldsymbol{x}) = \frac{\exp\big(-U_\Lambda^{\overline{\boldsymbol{x}}}(\boldsymbol{x})\big)}{\sum_{\boldsymbol{y} \in \mathscr{X}^\Lambda} \exp\big(-U_\Lambda^{\overline{\boldsymbol{x}}}(\boldsymbol{y})\big)}, \quad \Lambda \in \mathscr{E},\ \boldsymbol{x} \in \mathscr{X}^\Lambda,\ \overline{\boldsymbol{x}} \in \mathscr{X}^{\Lambda^{\mathrm{c}}}, \tag{1.3}$$

where

$$U_\Lambda^{\overline{\boldsymbol{x}}}(\boldsymbol{x}) = \sum_{J\colon \varnothing \neq J \subset \Lambda} \sum_{\widetilde{J} \in \mathscr{E}(\Lambda^{\mathrm{c}})} \Phi\big(\boldsymbol{x}_J\, \overline{\boldsymbol{x}}_{\widetilde{J}}\big), \quad \Lambda \in \mathscr{E},\ \boldsymbol{x} \in \mathscr{X}^{\Lambda},\ \overline{\boldsymbol{x}} \in \mathscr{X}^{\Lambda^{\mathrm{c}}}. \tag{1.4}$$

The specification $\boldsymbol{\mathcal{Q}}$ is called *Gibbsian with potential* Φ. Any random field compatible with $\boldsymbol{\mathcal{Q}}$ is called *Gibbs random field with potential* Φ.

The problem of characterization of the class of Gibbsian specifications with potentials satisfying some given conditions was subject of consideration of many authors: one can refer to Averintsev [1–3] and Sullivan [23] (see also Grimmett [19]) for real-valued finite-range potentials, Kozlov [20] and Sullivan [24] for uniformly convergent potentials, and the authors' works [4, 5] for more general potentials (which in particular can assume the value $+\infty$). Such characterizations are useful since they yield characterizations (in terms of conditional distribution) of the classes of corresponding Gibbs random fields.

In this paper we consider uniformly convergent potentials only, so Gibbsian specifications and Gibbs random fields with uniformly convergent potentials will be called shortly *Gibbsian specifications* and *Gibbs random fields*. The best-known characterization of the class of Gibbsian specifications is given by the following criterion (see, for example, [17]).

Criterion 1.1 (Kozlov – Sullivan). *A specification is Gibbsian if and only if it is quasilocal and strictly positive.*

Concerning (the subclass of) Gibbsian specifications with real-valued finite-range potentials, let us recall that they are characterized by strict positivity and locality. So, Gibbs random fields with real-valued finite-range potentials are characterized by strict positivity and Markov properties. Let us note that if the first property (strict positivity) holds, the second one (Markov) allows various equivalent formulations, one of which uses only single-site conditional probabilities with finite-volume conditions (see, for example, [25]). So, one has an internal characterization of Gibbs random fields with real-valued finite-range potentials in terms of one-point finite-conditional distribution. Establishment of a similar characterization in the general case of uniformly convergent potentials is not so straightforward. However, it was made possible due to and is very natural in light of the recently developed one-point framework which is presented in the following section.

2. One-point framework

In this section we briefly recall the main results of the authors' works [4–6] and of the paper by Dalalyan and Nahapetian [7], concerning the problems of description of random fields by means of one-point conditional and one-point finite-conditional distributions correspondingly.

2.1. Description of specifications and random fields by means of one-point conditional distributions

The idea that it is possible to describe and study random fields by means of one-point conditional distributions goes back to Dobrushin [8, 11]. Some steps in this direction were made by Sullivan [23] and Flood and Sullivan [16]. We can also mention Theorem 1.33 from the book by Georgii [17] concerning the problem of restoration of specifications by means of their single-site elements. However, the realization of Dobrushin's idea goes through a more important problem: the problem of description of specifications by means of systems of single-site probability distributions indexed by infinite boundary conditions (one-point systems) consistent in some sense. This problem was treated much later, the main difficulty residing in finding appropriate consistency conditions.

For the case of finite state space (considered in this paper), Dobrushin's idea was realized by the authors in [4, 5] under the weak positivity condition (as well as under the strict positivity condition) and in [6] under the newly-introduced very weak positivity condition. The case of a general (not necessarily finite) state space was studied by Fernández and Maillard in [14] under an alternative nonnullness condition and in [15] under the extension to this case of the very weak positivity condition. However, some important issues were left open in these papers. In particular, the necessity of the consistency conditions proposed in the first paper was not considered. Moreover, it is not difficult to see that except for some particular cases (for example, the strictly positive case), this necessity fails. Concerning the second paper, perhaps the most important issue, the equivalence between the compatibility with the original one-point system and the compatibility with the full specification constructed from it, is established within some class of random fields only.

Let us now briefly recall the main results of the authors' works [4–6]. In these papers, under wide positivity assumptions (*very weak positivity*), necessary and sufficient conditions for a system $\{\mathbf{Q}_t^{\overline{\boldsymbol{x}}},\ t \in \mathbb{Z}^\nu \text{ and } \overline{\boldsymbol{x}} \in \mathscr{X}^{t^c}\}$ of probability distributions to be contained in some specification were established. A system satisfying these conditions was called 1-specification. It was equally shown that the specification containing the given 1-specification is uniquely determined by some explicit formulae involving only the elements of this 1-specification. Moreover, since these formulae make use of finite number of elementary operations, the entire specification is quasilocal if and only if the 1-specification is, and the set of random fields compatible with the 1-specification coincides with the set of random fields compatible with the entire specification. Hence, whole Dobrushin theory can be reformulated in terms of 1-specifications, and so one can speak about description of random fields by means of one-point conditional distributions. The same applies to the results about characterization of Gibbsian specifications.

Below, we give some more details in the particular strictly positive case.

The definition of strictly positive 1-specification can be formulated in the following way: a system

$$\boldsymbol{\mathcal{Q}} = \left\{ \mathbf{Q}_t^{\overline{\boldsymbol{x}}},\ t \in \mathbb{Z}^\nu \text{ and } \overline{\boldsymbol{x}} \in \mathscr{X}^{t^{\mathrm{c}}} \right\}$$

of strictly positive probability distributions will be called 1-*specification* if

$$\mathbf{Q}_t^{\overline{\boldsymbol{x}}v}(x)\,\mathbf{Q}_s^{\overline{\boldsymbol{x}}x}(y)\,\mathbf{Q}_t^{\overline{\boldsymbol{x}}y}(u)\,\mathbf{Q}_s^{\overline{\boldsymbol{x}}u}(v) = \mathbf{Q}_s^{\overline{\boldsymbol{x}}u}(y)\,\mathbf{Q}_t^{\overline{\boldsymbol{x}}y}(x)\,\mathbf{Q}_s^{\overline{\boldsymbol{x}}x}(v)\,\mathbf{Q}_t^{\overline{\boldsymbol{x}}v}(u) \tag{2.1}$$
$$\text{for all } \ t,s \in \mathbb{Z}^\nu,\ \ x,u \in \mathscr{X}^t,\ \ y,v \in \mathscr{X}^s \ \text{ and } \ \overline{\boldsymbol{x}} \in \mathscr{X}^{\{t,s\}^{\mathrm{c}}}.$$

Further, a 1-specification $\boldsymbol{\mathcal{Q}} = \{\mathbf{Q}_t^{\overline{\boldsymbol{x}}},\ t \in \mathbb{Z}^\nu \text{ and } \overline{\boldsymbol{x}} \in \mathscr{X}^{t^{\mathrm{c}}}\}$ is called (*quasi*)*local* if for any $t \in \mathbb{Z}^\nu$ and $x \in \mathscr{X}^t$ the function $\overline{\boldsymbol{x}} \mapsto \mathbf{Q}_t^{\overline{\boldsymbol{x}}}(x)$ on $\mathscr{X}^{t^{\mathrm{c}}}$ is (quasi)local. Finally, a random fields $\mathbf{P}$ is called *compatible* with a 1-specification if the latter is a version of one-point conditional distribution of $\mathbf{P}$.

The above mentioned explicit formulae determining the elements of the specification $\boldsymbol{\mathcal{Q}} = \{\mathbf{Q}_\Lambda^{\overline{\boldsymbol{x}}},\ \Lambda \in \mathscr{E} \text{ and } \overline{\boldsymbol{x}} \in \mathscr{X}^{\Lambda^{\mathrm{c}}}\}$ containing the given strictly positive 1-specification have the following form: for all $\Lambda \in \mathscr{E}$ and $\overline{\boldsymbol{x}} \in \mathscr{X}^\Lambda$ one has

$$\mathbf{Q}_\Lambda^{\overline{\boldsymbol{x}}}(\boldsymbol{x}) = \frac{\mathbf{Q}_{t_1}^{\overline{\boldsymbol{x}}\boldsymbol{u}_{\{t_2,\dots,t_n\}}}(x_{t_1})\,\mathbf{Q}_{t_2}^{\overline{\boldsymbol{x}}\boldsymbol{x}_{\{t_1\}}\boldsymbol{u}_{\{t_3,\dots,t_n\}}}(x_{t_2})\cdots}{\mathbf{Q}_{t_1}^{\overline{\boldsymbol{x}}\boldsymbol{u}_{\{t_2,\dots,t_n\}}}(u_{t_1})\,\mathbf{Q}_{t_2}^{\overline{\boldsymbol{x}}\boldsymbol{x}_{\{t_1\}}\boldsymbol{u}_{\{t_3,\dots,t_n\}}}(u_{t_2})\cdots} \times \frac{\mathbf{Q}_{t_n}^{\overline{\boldsymbol{x}}\boldsymbol{x}_{\{t_1,\dots,t_{n-1}\}}}(x_{t_n})}{\mathbf{Q}_{t_n}^{\overline{\boldsymbol{x}}\boldsymbol{x}_{\{t_1,\dots,t_{n-1}\}}}(u_{t_n})} \times C, \quad \boldsymbol{x} \in \mathscr{X}^\Lambda,$$

where C is the normalizing factor. Here some fixed configuration $\boldsymbol{u} \in \mathscr{X}^\Lambda$ and some enumeration $t_1, \dots, t_n$ of elements of Λ are chosen arbitrary. Note that the right hand side of these formulae does not depend on this choice (correctness of the formulae) thanks to consistency condition (2.1).

Note also, that these formulae imply that the specification containing a strictly positive 1-specification is necessarily strictly positive itself. Recall that the quasilocality is also "heritable". Now Kozlov – Sullivan Criterion 1.1 can be clearly reduced to the following one, already obtained by the authors in [4, 5].

Criterion 2.1. *A specification is Gibbsian if and only if the* 1*-specification contained in it is quasilocal and strictly positive.*

Since the uniform convergence of potential assures the quasilocality of the 1-specification expressed by Gibbs formulae, one can also obtain the following corollary of Criterion 2.1.

Criterion 2.2. *A specification is Gibbsian if and only if the* 1*-specification contained in it admits the representation given by Gibbs formulae* (1.3) *and* (1.4) *with some uniformly convergent potential.*

2.2. Description of random fields by means of one-point finite-conditional distributions

Now we turn to the problem of description of random fields by means of one-point finite-conditional distributions considered in [7]. This description is closely related (and in some way complementary) to the one presented in the previous section.

First, let us note that the necessary and sufficient conditions for a system $\widetilde{\boldsymbol{q}} = \{\widetilde{\mathbf{Q}}_t^{\boldsymbol{x}},\ t \in \mathbb{Z}^\nu \text{ and } \widetilde{\boldsymbol{x}} \in \widetilde{\mathscr{X}^t}\}$ of probability distributions to be contained in some system $\widetilde{\boldsymbol{\mathcal{Q}}} = \{\widetilde{\mathbf{Q}}_\Lambda^{\boldsymbol{x}},\ \Lambda \in \mathscr{E} \text{ and } \widetilde{\boldsymbol{x}} \in \widetilde{\mathscr{X}^\Lambda}\}$ of probability distributions satisfying

$$\begin{aligned} &\widetilde{\mathbf{Q}}_\Lambda^{\boldsymbol{x}}(\boldsymbol{xy}) = \widetilde{\mathbf{Q}}_{\Lambda\setminus I}^{\boldsymbol{x}}(\boldsymbol{x})\, \widetilde{\mathbf{Q}}_I^{\boldsymbol{xx}}(\boldsymbol{y}) \qquad (2.2)\\ &\text{for all } \Lambda \in \mathscr{E},\ I \subset \Lambda,\ \boldsymbol{x} \in \mathscr{X}^{\Lambda\setminus I},\ \boldsymbol{y} \in \mathscr{X}^I \text{ and } \widetilde{\boldsymbol{x}} \in \widetilde{\mathscr{X}^\Lambda} \end{aligned}$$

are the following:

$$\begin{aligned} &\widetilde{\mathbf{Q}}_t^{\boldsymbol{x}}(x)\, \widetilde{\mathbf{Q}}_s^{\boldsymbol{x}x}(y) = \widetilde{\mathbf{Q}}_s^{\boldsymbol{x}}(y)\ \widetilde{\mathbf{Q}}_t^{\boldsymbol{x}y}(x) \qquad (2.3)\\ &\text{for all } t, s \in \mathbb{Z}^\nu,\ x \in \mathscr{X}^t,\ y \in \mathscr{X}^s \text{ and } \widetilde{\boldsymbol{x}} \in \widetilde{\mathscr{X}^{\{t,s\}}}. \end{aligned}$$

Note also that if $\widetilde{\boldsymbol{q}}$ is the one-point finite-conditional ($\widetilde{\boldsymbol{\mathcal{Q}}}$ is the finite-conditional) distribution of some strictly positive random field, then it necessarily satisfies the condition (2.3) (the condition (2.2)). However, in order for a strictly positive system $\widetilde{\boldsymbol{q}}$ satisfying (2.3) ($\widetilde{\boldsymbol{\mathcal{Q}}}$ satisfying (2.2)) to be the one-point finite-conditional (the finite-conditional) distribution of some strictly positive random field one needs some additional conditions. It turns out that such conditions are the following:

$$\begin{aligned} &\mathbf{Q}_t^v(x)\, \mathbf{Q}_s^x(y)\, \mathbf{Q}_t^y(u)\, \mathbf{Q}_s^u(v) = \mathbf{Q}_s^u(y)\, \mathbf{Q}_t^y(x)\, \mathbf{Q}_s^x(v)\, \mathbf{Q}_t^v(u) \qquad (2.4)\\ &\text{for all } t, s \in \mathbb{Z}^\nu,\ x, u \in \mathscr{X}^t \text{ and } y, v \in \mathscr{X}^s. \end{aligned}$$

More precisely, in [7] it was shown that the strict positivity of elements and the fulfillment of the conditions (2.3) and (2.4) are necessary and sufficient for a system $\{\widetilde{\mathbf{Q}}_t^{\boldsymbol{x}},\ t \in \mathbb{Z}^\nu \text{ and } \widetilde{\boldsymbol{x}} \in \widetilde{\mathscr{X}^t}\}$ of probability distributions to be the one-point finite-conditional distribution of some strictly positive random field. It was equally shown that this random field is uniquely determined by this system. In particular, a strictly positive random field is uniquely determined by (can be identified with) its one-point finite-conditional distribution, and so one can speak about description of random fields by means of one-point finite-conditional distributions.

3. Random field Gibbsianness criteria

In this section we turn to the main subject of the present work: the problem of internal characterization of Gibbs random fields. The main results: random field Gibbsianness criteria in terms of one-point finite-conditional distribution will be established in Section 3.3. Before that, random field Gibbsianness criteria in terms of conditional and one-point conditional distribution will be obtained in the next two sections by means of transformation and subsequent improvement of Criteria 1.1, 2.1, 2.2.

3.1. Random field Gibbsianness criteria in terms of conditional distribution

Combining the definition of Gibbs random field with Criterion 1.1 one gets the following well-known characterization: a random field is a Gibbs random field if and only if it has a version of conditional distribution which is a quasilocal and strictly positive specification. This criterion can be improved in the following way.

Criterion 3.1. *A random field is a Gibbs random field if and only if it has a version of conditional distribution which is quasilocal and strictly positive.*

Since the strict positivity of a version of conditional distribution implies the strict positivity of the random field, the criterion is immediately deduced from the following proposition which is of general interest.

Proposition 3.1. *If a strictly positive random field has a quasilocal version of conditional distribution, the latter is unique and is necessarily a specification.*

Proof. Let $\mathbf{P}$ be a strictly positive random field. First, note that the measure $\mathbf{P}$ is everywhere dense, that is, $\mathbf{P}(A) > 0$ for any non-empty open set $A \in \mathscr{T}^{\mathbb{Z}^\nu} \setminus \{\varnothing\}$. Indeed, since such a set A necessarily contains a non-empty cylinder subset A', which in turn contains a subset $\{\overline{\boldsymbol{x}} \in \mathscr{X}^{\mathbb{Z}^\nu} : \overline{\boldsymbol{x}}_\Lambda = \boldsymbol{x}^\circ\}$ where $\Lambda \in \mathscr{E}^*$ and $\boldsymbol{x}^\circ \in \mathscr{X}^\Lambda$, we have $\mathbf{P}(A) \geqslant \mathbf{P}(A') \geqslant \mathbf{P}_\Lambda(\boldsymbol{x}^\circ) > 0$. An important evident property of everywhere dense measures is the following: if a continuous function is equal to zero almost everywhere (with respect to such a measure), then it is equal to zero everywhere.

Now, suppose $\{\mathbf{Q}_\Lambda^{\overline{\boldsymbol{x}}},\ \Lambda \in \mathscr{E} \text{ and } \overline{\boldsymbol{x}} \in \mathscr{X}^{\Lambda^c}\}$ and $\{\boldsymbol{q}_\Lambda^{\overline{\boldsymbol{x}}},\ \Lambda \in \mathscr{E} \text{ and } \overline{\boldsymbol{x}} \in \mathscr{X}^{\Lambda^c}\}$ are two quasilocal versions of conditional distribution of $\mathbf{P}$. Hence, for any $\Lambda \in \mathscr{E}$ and $\boldsymbol{x} \in \mathscr{X}^\Lambda$, the function

$$\overline{\boldsymbol{x}} \longmapsto \mathbf{Q}_\Lambda^{\overline{\boldsymbol{x}}}(\boldsymbol{x}) - \boldsymbol{q}_\Lambda^{\overline{\boldsymbol{x}}}(\boldsymbol{x})$$

on $\mathscr{X}^{\Lambda^c}$ is quasilocal and equal to zero $\mathbf{P}_{\Lambda^c}$-almost everywhere. Since quasilocality is nothing but continuity and the measure $\mathbf{P}_{\Lambda^c}$ is everywhere dense, this function is equal to zero everywhere. So, the uniqueness is proved.

Finally, suppose $\boldsymbol{\mathcal{Q}} = \{\mathbf{Q}_\Lambda^{\overline{\boldsymbol{x}}},\ \Lambda \in \mathscr{E} \text{ and } \overline{\boldsymbol{x}} \in \mathscr{X}^{\Lambda^c}\}$ is (the unique) quasilocal version of conditional distribution of $\mathbf{P}$. For any $\Lambda \in \mathscr{E}$, $I \subset \Lambda$, $\boldsymbol{x} \in \mathscr{X}^{\Lambda\setminus I}$ and $\boldsymbol{y} \in \mathscr{X}^I$ consider the function

$$\overline{\boldsymbol{x}} \longmapsto \mathbf{Q}_\Lambda^{\overline{\boldsymbol{x}}}(\boldsymbol{x}\boldsymbol{y}) - (\mathbf{Q}_\Lambda^{\overline{\boldsymbol{x}}})_{\Lambda\setminus I}(\boldsymbol{x})\, \mathbf{Q}_I^{\overline{\boldsymbol{x}}\boldsymbol{x}}(\boldsymbol{y})$$

on $\mathscr{X}^{\Lambda^c}$. This function is clearly quasilocal and, as it follows from the properties of conditional probabilities, is equal to zero $\mathbf{P}_{\Lambda^c}$-almost everywhere. Hence it is equal to zero everywhere, and so $\boldsymbol{\mathcal{Q}}$ is a specification. □

Let us note that Criterion 3.1 was as a matter of fact obtained in [24] using a different approach.

3.2. Random field Gibbsianness criteria in terms of one-point conditional distribution

Criterion 3.1 characterizes Gibbs random fields in terms of conditional distribution. However, in view of Section 2.1, it should be possible to do it in terms of one-point conditional distribution. Indeed, combining the definition of Gibbs random field with Criterion 2.1 and taking into account the results of Section 2.1, one gets the following characterization: a random field is a Gibbs random field if and only if it has a version of one-point conditional distribution which is a quasilocal and strictly positive 1-specification. As in the preceding section we can improve this criterion in the following way.

Criterion 3.2. *A random field is a Gibbs random field if and only if it has a version of one-point conditional distribution which is quasilocal and strictly positive.*

The criterion is immediately deduced from the following two propositions which are of general interest.

Proposition 3.2. *If a random field* $\mathbf{P}$ *has a strictly positive version of one-point conditional distribution, then* $\mathbf{P}$ *is strictly positive itself.*

Proof. Let us suppose that the random field $\mathbf{P}$ is not strictly positive. In this case we can find some $\Lambda \in \mathscr{E}^*$, $t \in \Lambda$ and $\boldsymbol{z} \in \mathscr{X}^\Lambda$ such that $\mathbf{P}_\Lambda(\boldsymbol{z}) = 0$ and $\mathbf{P}_{\Lambda\setminus t}(\boldsymbol{z}_{\Lambda\setminus t}) > 0$ (recall that $\mathbf{P}_\varnothing(\varnothing) = 1$). Now denote

$$A = \big\{\overline{\boldsymbol{x}} \in \mathscr{X}^{t^c} : \overline{\boldsymbol{x}}_{\Lambda\setminus t} = \boldsymbol{z}_{\Lambda\setminus t}\big\}.$$

Obviously $\mathbf{P}_{t^c}(A) = \mathbf{P}_{\Lambda\setminus t}(\boldsymbol{z}_{\Lambda\setminus t}) > 0$. Introduce also

$$B = \bigcap_{\widetilde{\Lambda}\in\mathscr{E}(t^c)} \big\{\overline{\boldsymbol{x}} \in \mathscr{X}^{t^c} : \mathbf{P}_{\widetilde{\Lambda}}(\overline{\boldsymbol{x}}_{\widetilde{\Lambda}}) > 0\big\}.$$

Since B is a countable intersection of sets of probability 1, we have $\mathbf{P}_{t^{\mathrm{c}}}(B) = 1$. So, it comes $\mathbf{P}_{t^{\mathrm{c}}}(A \cap B) > 0$.

For all $\overline{\boldsymbol{x}} \in A \cap B$ and all $\widetilde{\Lambda} \in \mathscr{E}(t^{\mathrm{c}})$ such that $\widetilde{\Lambda} \supset \Lambda \setminus t$, we have

$$\boldsymbol{q}_t^{\overline{\boldsymbol{x}}_{\widetilde{\Lambda}}}(\boldsymbol{z}_t) = \frac{\mathbf{P}_{t\cup\widetilde{\Lambda}}(\boldsymbol{z}_t\overline{\boldsymbol{x}}_{\widetilde{\Lambda}})}{\mathbf{P}_{\widetilde{\Lambda}}(\overline{\boldsymbol{x}}_{\widetilde{\Lambda}})} = 0.$$

Hence, for all $\overline{\boldsymbol{x}} \in A \cap B$ we get

$$\lim_{\widetilde{\Lambda}\uparrow\mathbb{Z}^\nu\setminus t} \boldsymbol{q}_t^{\overline{\boldsymbol{x}}_{\widetilde{\Lambda}}}(\boldsymbol{z}_t) = 0$$

which contradicts the existence of a strictly positive version of one-point conditional distribution of $\mathbf{P}$. □

Proposition 3.3. *If a strictly positive random field has a quasilocal version of one-point conditional distribution, the latter is unique and is necessarily a* 1*-specification.*

Proof. The uniqueness is proved following exactly the same argument as in the proof of Proposition 3.1.

To prove the second assertion, suppose $\{\mathbf{Q}_t^{\overline{\boldsymbol{x}}},\ t \in \mathbb{Z}^\nu \text{ and } \overline{\boldsymbol{x}} \in \mathscr{X}^{t^{\mathrm{c}}}\}$ is (the unique) quasilocal version of one-point conditional distribution of a strictly positive random field $\mathbf{P}$. For any $t, s \in \mathbb{Z}^\nu$, $x, u \in \mathscr{X}^t$ and $y, v \in \mathscr{X}^s$ consider the function

$$\overline{\boldsymbol{x}} \longmapsto \mathbf{Q}_t^{\overline{\boldsymbol{x}}v}(x)\,\mathbf{Q}_s^{\overline{\boldsymbol{x}}x}(y)\,\mathbf{Q}_t^{\overline{\boldsymbol{x}}y}(u)\,\mathbf{Q}_s^{\overline{\boldsymbol{x}}u}(v) - \mathbf{Q}_s^{\overline{\boldsymbol{x}}u}(y)\,\mathbf{Q}_t^{\overline{\boldsymbol{x}}y}(x)\,\mathbf{Q}_s^{\overline{\boldsymbol{x}}x}(v)\,\mathbf{Q}_t^{\overline{\boldsymbol{x}}v}(u)$$

on $\mathscr{X}^{\{t,s\}^{\mathrm{c}}}$. Applying the reasoning used in the proof of Proposition 3.1, it clearly comes that this function is equal to zero everywhere. □

Concluding this section let us note that combining the definition of Gibbs random field with Criterion 2.2 and taking into account the results of Section 2.1, one also has the following characterization.

Criterion 3.3. *A random field is a Gibbs random field if and only if it has a version of one-point conditional distribution admitting the representation given by Gibbs formulae* (1.3) *and* (1.4) *with some uniformly convergent potential.*

3.3. Random field Gibbsianness criteria in terms of one-point finite-conditional distribution

Now we can establish random field Gibbsianness criteria in terms of one-point finite-conditional distribution, which are precisely the main outcome of the present paper. The first such criterion is the following.

Criterion 3.4. *A random field is a Gibbs random field if and only if it is strictly positive and its one-point finite-conditional distribution* $\{\boldsymbol{q}_t^{\widetilde{\boldsymbol{x}}},\ t \in \mathbb{Z}^\nu$ *and* $\widetilde{\boldsymbol{x}} \in \widetilde{\mathscr{X}^t}\}$ *satisfy one of the following equivalent conditions:*

(A) *the limits*

$$\lim_{\Lambda \uparrow \mathbb{Z}^\nu \setminus t} \boldsymbol{q}_t^{\overline{\boldsymbol{x}}_\Lambda}(x), \quad t \in \mathbb{Z}^\nu,\ x \in \mathscr{X}^t,\ \overline{\boldsymbol{x}} \in \mathscr{X}^{t^{\mathrm{c}}},$$

exist, are uniformly nonnull with respect to $\overline{\boldsymbol{x}}$*, and the convergence is uniform with respect to* $\overline{\boldsymbol{x}}$*,*

(B) *the limits*

$$\lim_{\Lambda \uparrow \mathbb{Z}^\nu \setminus t} \boldsymbol{q}_t^{\overline{\boldsymbol{x}}_\Lambda}(x), \quad t \in \mathbb{Z}^\nu,\ x \in \mathscr{X}^t,\ \overline{\boldsymbol{x}} \in \mathscr{X}^{t^{\mathrm{c}}},$$

exist, are strictly positive, and the convergence is uniform with respect to $\overline{\boldsymbol{x}}$*.*

Proof. The sufficiency is quite evident. Indeed, the strictly positive limits supposed to exist form a strictly positive version of one-point conditional distribution of the random field. The uniformity of convergence guarantees that this version is quasilocal and so, the sufficiency follows from Criterion 3.2. Let us also note that at the same time this quasilocality clearly yields the equivalence of the conditions (A) and (B).

Now let us turn to the proof of the necessity. Let $\mathbf{P}$ be a Gibbs random field. According to Criterion 3.2 it has a quasilocal and strictly positive version $\boldsymbol{\mathcal{Q}} = \{\mathbf{Q}_t^{\overline{\boldsymbol{x}}},\ t \in \mathbb{Z}^\nu$ and $\overline{\boldsymbol{x}} \in \mathscr{X}^{t^{\mathrm{c}}}\}$ of one-point conditional distribution. So, according to Proposition 3.2, the random field $\mathbf{P}$ is strictly positive, and to conclude the proof it is sufficient to show that

$$\lim_{\Lambda \uparrow \mathbb{Z}^\nu \setminus t} \sup_{\overline{\boldsymbol{x}} \in \mathscr{X}^{t^{\mathrm{c}}}} \left| \boldsymbol{q}_t^{\overline{\boldsymbol{x}}_\Lambda}(x) - \mathbf{Q}_t^{\overline{\boldsymbol{x}}}(x) \right| = 0$$

for all $t \in \mathbb{Z}^\nu$ and $x \in \mathscr{X}^t$.

For this we need the following inequality due to Sullivan:

$$\inf_{\overline{\boldsymbol{y}} \in \mathscr{X}^{t^{\mathrm{c}}} :\, \overline{\boldsymbol{y}}_\Lambda = \boldsymbol{z}} \mathbf{Q}_t^{\overline{\boldsymbol{y}}}(x) \leqslant \boldsymbol{q}_t^{\boldsymbol{z}}(x) \leqslant \sup_{\overline{\boldsymbol{y}} \in \mathscr{X}^{t^{\mathrm{c}}} :\, \overline{\boldsymbol{y}}_\Lambda = \boldsymbol{z}} \mathbf{Q}_t^{\overline{\boldsymbol{y}}}(x) \tag{3.1}$$

for all $t \in \mathbb{Z}^\nu$, $\Lambda \in \mathscr{E}^*(t^{\mathrm{c}})$, $x \in \mathscr{X}^t$ and $\boldsymbol{z} \in \mathscr{X}^\Lambda$. This inequality is clearly valid since

$$\boldsymbol{q}_t^{\boldsymbol{z}}(x) = \frac{\mathbf{P}_{t \cup \Lambda}(x\boldsymbol{z})}{\mathbf{P}_\Lambda(\boldsymbol{z})} = \frac{1}{\mathbf{P}_\Lambda(\boldsymbol{z})} \int\limits_{\{\overline{\boldsymbol{y}} \in \mathscr{X}^{t^{\mathrm{c}}} :\, \overline{\boldsymbol{y}}_\Lambda = \boldsymbol{z}\}} \mathbf{Q}_t^{\overline{\boldsymbol{y}}}(x)\, \mathbf{P}_{t^{\mathrm{c}}}(d\overline{\boldsymbol{y}}).$$

Taking this inequality into account, it remains to verify that

$$\lim_{\Lambda \uparrow \mathbb{Z}^\nu \setminus t} \sup_{\overline{\boldsymbol{x}} \in \mathscr{X}^{t^{\mathrm{c}}}} \left| \inf_{\overline{\boldsymbol{y}} \in \mathscr{X}^{t^{\mathrm{c}}} :\, \overline{\boldsymbol{y}}_\Lambda = \overline{\boldsymbol{x}}_\Lambda} \mathbf{Q}_t^{\overline{\boldsymbol{y}}}(x) - \mathbf{Q}_t^{\overline{\boldsymbol{x}}}(x) \right| = 0$$

and

$$\lim_{\Lambda\uparrow\mathbb{Z}^\nu\setminus t}\ \sup_{\overline{\boldsymbol{x}}\in\mathscr{X}^{t^{\mathrm{c}}}}\Big|\sup_{\overline{\boldsymbol{y}}\in\mathscr{X}^{t^{\mathrm{c}}}\colon\overline{\boldsymbol{y}}_\Lambda=\overline{\boldsymbol{x}}_\Lambda}\mathbf{Q}_t^{\overline{\boldsymbol{y}}}(x)-\mathbf{Q}_t^{\overline{\boldsymbol{x}}}(x)\Big|=0$$

for all $t\in\mathbb{Z}^\nu$ and $x\in\mathscr{X}^t$. To show the first one we write

$$\sup_{\overline{\boldsymbol{x}}\in\mathscr{X}^{t^{\mathrm{c}}}}\Big|\inf_{\overline{\boldsymbol{y}}\in\mathscr{X}^{t^{\mathrm{c}}}\colon\overline{\boldsymbol{y}}_\Lambda=\overline{\boldsymbol{x}}_\Lambda}\mathbf{Q}_t^{\overline{\boldsymbol{y}}}(x)-\mathbf{Q}_t^{\overline{\boldsymbol{x}}}(x)\Big|\leqslant\sup_{\overline{\boldsymbol{x}}\in\mathscr{X}^{t^{\mathrm{c}}}}\ \sup_{\overline{\boldsymbol{y}}\in\mathscr{X}^{t^{\mathrm{c}}}\colon\overline{\boldsymbol{y}}_\Lambda=\overline{\boldsymbol{x}}_\Lambda}\big|\,\mathbf{Q}_t^{\overline{\boldsymbol{y}}}(x)-\mathbf{Q}_t^{\overline{\boldsymbol{x}}}(x)\big|$$

and use the quasilocality of $\boldsymbol{\mathcal{Q}}$. The second one is proved similarly. □

At first sight, Criterion 3.4 deals only with the one-point finite-conditional distribution. However, in fact it imposes conditions equally on its limit, that is, on the one-point conditional distribution. The following and last criterion really deals only with the one-point finite-conditional distribution. Before formulating it, let as agree that in the sequel when we use the notation $\boldsymbol{x}_T$ we presume that only configurations $\boldsymbol{x}$ such that $\mathfrak{S}(\boldsymbol{x})\supset T$ are considered.

Criterion 3.5. *A random field is a Gibbs random field if and only if it is strictly positive, its one-point finite-conditional distribution* $\{\boldsymbol{q}_t^{\widetilde{\boldsymbol{x}}},\ t\in\mathbb{Z}^\nu$ *and* $\widetilde{\boldsymbol{x}}\in\widetilde{\mathscr{X}^t}\}$ *is uniformly nonnull* (*consists of uniformly nonnull with respect to* $\widetilde{\boldsymbol{x}}$ *elements*) *and one of the following equivalent conditions holds:*

(C) *for any* $t\in\mathbb{Z}^\nu$ *and* $x\in\mathscr{X}^t$ *one has*

$$\lim_{\Lambda\uparrow\mathbb{Z}^\nu\setminus t}\ \sup_{\widetilde{\boldsymbol{x}},\widetilde{\boldsymbol{y}}\in\widetilde{\mathscr{X}^t}\colon\widetilde{\boldsymbol{x}}_\Lambda=\widetilde{\boldsymbol{y}}_\Lambda}\big|\boldsymbol{q}_t^{\widetilde{\boldsymbol{x}}}(x)-\boldsymbol{q}_t^{\widetilde{\boldsymbol{y}}}(x)\big|=0,$$

(D) *for any* $t\in\mathbb{Z}^\nu$ *and* $x\in\mathscr{X}^t$ *one has*

$$\lim_{\Lambda\uparrow\mathbb{Z}^\nu\setminus t}\ \sup_{J\in\mathscr{E}^*(t^{\mathrm{c}})}\ \sup_{\widetilde{\boldsymbol{x}},\widetilde{\boldsymbol{y}}\in\mathscr{X}^J\colon\widetilde{\boldsymbol{x}}_\Lambda=\widetilde{\boldsymbol{y}}_\Lambda}\big|\boldsymbol{q}_t^{\widetilde{\boldsymbol{x}}}(x)-\boldsymbol{q}_t^{\widetilde{\boldsymbol{y}}}(x)\big|=0,$$

(E) *for any* $t\in\mathbb{Z}^\nu$ *and* $x\in\mathscr{X}^t$ *one has*

$$\lim_{\Lambda\uparrow\mathbb{Z}^\nu\setminus t}\ \sup_{\widetilde{\boldsymbol{x}}\in\widetilde{\mathscr{X}^t}}\big|\boldsymbol{q}_t^{\widetilde{\boldsymbol{x}}}(x)-\boldsymbol{q}_t^{\widetilde{\boldsymbol{x}}_\Lambda}(x)\big|=0.$$

Proof. First we concentrate on the condition (E). Clearly

$$\sup_{\widetilde{\boldsymbol{x}}\in\widetilde{\mathscr{X}^t}}\big|\boldsymbol{q}_t^{\widetilde{\boldsymbol{x}}}(x)-\boldsymbol{q}_t^{\widetilde{\boldsymbol{x}}_\Lambda}(x)\big|=\sup_{I\in\mathscr{E}(t^{\mathrm{c}})\colon I\supset\Lambda}\ \sup_{\overline{\boldsymbol{x}}\in\mathscr{X}^{t^{\mathrm{c}}}}\big|\boldsymbol{q}_t^{\overline{\boldsymbol{x}}_I}(x)-\boldsymbol{q}_t^{\overline{\boldsymbol{x}}_\Lambda}(x)\big|,$$

and so the condition (E) is nothing but the Cauchy condition for the existence of the uniform limits considered in Criterion 3.4. The sufficiency now clearly

follows from Criterion 3.4 since the Cauchy principle yields the existence of the uniform limits, and the uniform nonnullness of one-point finite-conditional distribution guarantees their strict positivity. The necessity also follows from Criterion 3.4 since the condition (E) is ensured by the Cauchy principle, and the uniform nonnullness of one-point finite-conditional distribution can be easily obtained from (3.1) and the condition (A) (use the first inequality of (3.1) and the uniform nonnullness of limits considered in the condition (A)).

It remains to check the equivalence of the conditions (C), (D) and (E). The implications (C) $\Rightarrow$ (D) and (C) $\Rightarrow$ (E) are trivial since

$$\sup_{J\in\mathscr{E}^*(t^{\mathrm{c}})}\sup_{\widetilde{\boldsymbol{x}},\widetilde{\boldsymbol{y}}\in\mathscr{X}^J:\,\widetilde{\boldsymbol{x}}_\Lambda=\widetilde{\boldsymbol{y}}_\Lambda}\left|\boldsymbol{q}_t^{\widetilde{\boldsymbol{x}}}(x)-\boldsymbol{q}_t^{\widetilde{\boldsymbol{y}}}(x)\right|\leqslant\sup_{\widetilde{\boldsymbol{x}},\widetilde{\boldsymbol{y}}\in\widetilde{\mathscr{X}^t}:\,\widetilde{\boldsymbol{x}}_\Lambda=\widetilde{\boldsymbol{y}}_\Lambda}\left|\boldsymbol{q}_t^{\widetilde{\boldsymbol{x}}}(x)-\boldsymbol{q}_t^{\widetilde{\boldsymbol{y}}}(x)\right|$$

and

$$\sup_{\widetilde{\boldsymbol{x}}\in\widetilde{\mathscr{X}^t}}\left|\boldsymbol{q}_t^{\widetilde{\boldsymbol{x}}}(x)-\boldsymbol{q}_t^{\widetilde{\boldsymbol{x}}_\Lambda}(x)\right|\leqslant\sup_{\widetilde{\boldsymbol{x}},\widetilde{\boldsymbol{y}}\in\widetilde{\mathscr{X}^t}:\,\widetilde{\boldsymbol{x}}_\Lambda=\widetilde{\boldsymbol{y}}_\Lambda}\left|\boldsymbol{q}_t^{\widetilde{\boldsymbol{x}}}(x)-\boldsymbol{q}_t^{\widetilde{\boldsymbol{y}}}(x)\right|.$$

Similarly, the inequality

$$\sup_{\widetilde{\boldsymbol{x}},\widetilde{\boldsymbol{y}}\in\widetilde{\mathscr{X}^t}:\,\widetilde{\boldsymbol{x}}_\Lambda=\widetilde{\boldsymbol{y}}_\Lambda}\left|\boldsymbol{q}_t^{\widetilde{\boldsymbol{x}}}(x)-\boldsymbol{q}_t^{\widetilde{\boldsymbol{y}}}(x)\right|\leqslant 2\sup_{\widetilde{\boldsymbol{x}}\in\widetilde{\mathscr{X}^t}}\left|\boldsymbol{q}_t^{\widetilde{\boldsymbol{x}}}(x)-\boldsymbol{q}_t^{\widetilde{\boldsymbol{x}}_\Lambda}(x)\right|$$

yields the implications (E) $\Rightarrow$ (C). To prove the last implication (D) $\Rightarrow$ (C), we need the following lemma.

Lemma 3.1. *Let* $\{\boldsymbol{q}_I^{\widetilde{\boldsymbol{x}}},\ I\in\mathscr{E}$ *and* $\widetilde{\boldsymbol{x}}\in\widetilde{\mathscr{X}^I}\}$ *be the finite-conditional distribution of some strictly positive random field. Then the set*

$$A=\left\{\overline{\boldsymbol{x}}\in\mathscr{X}^{\mathbb{Z}^\nu}:\lim_{\Lambda\uparrow\mathbb{Z}^\nu\setminus I}\boldsymbol{q}_I^{\overline{\boldsymbol{x}}_\Lambda}(x)\ \text{ exists for every }\ I\in\mathscr{E}\ \text{ and }\ x\in\mathscr{X}^I\right\}$$

is of probability 1 *and possesses the following property: if* $\overline{\boldsymbol{x}}\in A$ *then* $\boldsymbol{z}\overline{\boldsymbol{x}}_{J^{\mathrm{c}}}\in A$ *for all* $J\in\mathscr{E}$ *and* $\boldsymbol{z}\in\mathscr{X}^J$.

Proof. Since the set A is a countable intersection of sets of probability 1, it is also of probability 1. It remains to show that if $\overline{\boldsymbol{x}}\in A$ then $\overline{\boldsymbol{y}}=z\overline{\boldsymbol{x}}_{t^{\mathrm{c}}}\in A$ for all $t\in\mathbb{Z}^\nu$ and $z\in\mathscr{X}^t$, that is, $\lim_{\Lambda\uparrow\mathbb{Z}^\nu\setminus I}\boldsymbol{q}_I^{\overline{\boldsymbol{y}}_\Lambda}(\boldsymbol{x})$ exists for every $I\in\mathscr{E}$ and $\boldsymbol{x}\in\mathscr{X}^I$. This is trivial if $t\in I$ (since in this case $\overline{\boldsymbol{y}}_\Lambda=\overline{\boldsymbol{x}}_\Lambda$) and clearly follows from the relation

$$\boldsymbol{q}_I^{\overline{\boldsymbol{y}}_\Lambda}(\boldsymbol{x})=\boldsymbol{q}_I^{z\overline{\boldsymbol{x}}_{\Lambda\setminus t}}(\boldsymbol{x})=\frac{\boldsymbol{q}_{t\cup I}^{\overline{\boldsymbol{x}}_{\Lambda\setminus t}}(z\boldsymbol{x})}{\left(\boldsymbol{q}_{t\cup I}^{\overline{\boldsymbol{x}}_{\Lambda\setminus t}}\right)_t(z)},\quad\Lambda\ni t,$$

otherwise. □

Returning to the proof of the implication (D) $\Rightarrow$ (C), let us fix some $t \in \mathbb{Z}^\nu$ and $x \in \mathscr{X}^t$, denote

$$f(\Lambda) = \sup_{J \in \mathscr{E}^*(t^c)} \sup_{\widetilde{\boldsymbol{x}}, \widetilde{\boldsymbol{y}} \in \mathscr{X}^J : \widetilde{\boldsymbol{x}}_\Lambda = \widetilde{\boldsymbol{y}}_\Lambda} \left| \boldsymbol{q}_t^{\widetilde{\boldsymbol{x}}}(x) - \boldsymbol{q}_t^{\widetilde{\boldsymbol{y}}}(x) \right|,$$

and for any $\varepsilon > 0$ choose (according to the condition (D)) some $\Lambda_\varepsilon \in \mathscr{E}$ such that $|f(\Lambda)| < \varepsilon$ for all $\Lambda \in \mathscr{E}$, $\Lambda \supset \Lambda_\varepsilon$.

First we will show that $\lim_{\Lambda \uparrow \mathbb{Z}^\nu \setminus t} \boldsymbol{q}_t^{\overline{\boldsymbol{x}}_\Lambda}(x)$ exists for every $\overline{\boldsymbol{x}} \in \mathscr{X}^{t^c}$. Let us take some $\overline{\boldsymbol{x}}^\circ \in A$ (according to the lemma, the set A is of probability 1 and so is not empty) and consider $\overline{\boldsymbol{y}} = \overline{\boldsymbol{x}}_{\Lambda_\varepsilon} \overline{\boldsymbol{x}}^\circ_{\Lambda_\varepsilon^c} \in A$. So, we can find some $\Lambda'_\varepsilon \in \mathscr{E}$ such that

$$|\boldsymbol{q}_t^{\overline{\boldsymbol{y}}_I}(x) - \boldsymbol{q}_t^{\overline{\boldsymbol{y}}_J}(x)| < \varepsilon$$

for all $I, J \in \mathscr{E}$, $I \supset \Lambda'_\varepsilon$, $J \supset \Lambda'_\varepsilon$. Thus, for all $I, J \in \mathscr{E}$ such that $I \supset \Lambda_\varepsilon \cup \Lambda'_\varepsilon$ and $J \supset \Lambda_\varepsilon \cup \Lambda'_\varepsilon$ we can write

$$\begin{aligned} \left| \boldsymbol{q}_t^{\overline{\boldsymbol{x}}_I}(x) - \boldsymbol{q}_t^{\overline{\boldsymbol{x}}_J}(x) \right| &\leqslant \left| \boldsymbol{q}_t^{\overline{\boldsymbol{x}}_I}(x) - \boldsymbol{q}_t^{\overline{\boldsymbol{y}}_I}(x) \right| + \left| \boldsymbol{q}_t^{\overline{\boldsymbol{y}}_I}(x) - \boldsymbol{q}_t^{\overline{\boldsymbol{y}}_J}(x) \right| + \left| \boldsymbol{q}_t^{\overline{\boldsymbol{y}}_J}(x) - \boldsymbol{q}_t^{\overline{\boldsymbol{x}}_J}(x) \right| \\ &< f(\Lambda_\varepsilon) + \varepsilon + f(\Lambda_\varepsilon) < 3\varepsilon, \end{aligned}$$

and hence $\lim_{\Lambda \uparrow \mathbb{Z}^\nu \setminus t} \boldsymbol{q}_t^{\overline{\boldsymbol{x}}_\Lambda}(x)$ exists according to Cauchy principle.

Further, for every $\overline{\boldsymbol{x}} \in \mathscr{X}^{t^c}$ consider the set $V(\overline{\boldsymbol{x}}) = \{\overline{\boldsymbol{y}} \in \mathscr{X}^{t^c} : \overline{\boldsymbol{y}}_{\Lambda_\varepsilon} = \overline{\boldsymbol{x}}_{\Lambda_\varepsilon}\}$. Clearly these sets are either mutually disjoint or coinciding, and there is only a finite number k (more precisely $k = |\mathscr{X}^{\Lambda_\varepsilon}|$) of different sets among them. Hence there exists a finite collection $\overline{\boldsymbol{x}}^1, \ldots, \overline{\boldsymbol{x}}^k \in \mathscr{X}^{t^c}$ such that

$$\mathscr{X}^{t^c} = \bigcup_{i=1}^k V(\overline{\boldsymbol{x}}^i).$$

(This fact equally follows from the compactness of $\mathscr{X}^{t^c}$.) So, using Cauchy principle we can find some $\Lambda''_\varepsilon \in \mathscr{E}$ such that $|\boldsymbol{q}_t^{\overline{\boldsymbol{x}}^i_I}(x) - \boldsymbol{q}_t^{\overline{\boldsymbol{x}}^i_J}(x)| < \varepsilon$ for all $i = 1, \ldots, k$ and all $I, J \in \mathscr{E}$, $I \supset \Lambda''_\varepsilon$, $J \supset \Lambda''_\varepsilon$.

Now, let the set $\Lambda \in \mathscr{E}$ be such that $\Lambda \supset \Lambda_\varepsilon \cup \Lambda''_\varepsilon$, the sets $I, J \in \mathscr{E}$ be such that $I \supset \Lambda$ and $J \supset \Lambda$, and the configurations $\widetilde{\boldsymbol{x}} \in \mathscr{X}^I$ and $\widetilde{\boldsymbol{y}} \in \mathscr{X}^J$ be such that $\widetilde{\boldsymbol{x}}_\Lambda = \widetilde{\boldsymbol{y}}_\Lambda$. Clearly, we can find some $i \in \{1, \ldots, k\}$ such that $\overline{\boldsymbol{x}}^i_{\Lambda_\varepsilon} = \widetilde{\boldsymbol{x}}_{\Lambda_\varepsilon} = \widetilde{\boldsymbol{y}}_{\Lambda_\varepsilon}$, and thus we may write

$$\begin{aligned} \left| \boldsymbol{q}_t^{\widetilde{\boldsymbol{x}}}(x) - \boldsymbol{q}_t^{\widetilde{\boldsymbol{y}}}(x) \right| &\leqslant \left| \boldsymbol{q}_t^{\widetilde{\boldsymbol{x}}}(x) - \boldsymbol{q}_t^{\overline{\boldsymbol{x}}^i_I}(x) \right| + \left| \boldsymbol{q}_t^{\overline{\boldsymbol{x}}^i_I}(x) - \boldsymbol{q}_t^{\overline{\boldsymbol{x}}^i_J}(x) \right| + \left| \boldsymbol{q}_t^{\overline{\boldsymbol{x}}^i_J}(x) - \boldsymbol{q}_t^{\widetilde{\boldsymbol{y}}}(x) \right| \\ &< f(\Lambda_\varepsilon) + \varepsilon + f(\Lambda_\varepsilon) < 3\varepsilon \end{aligned}$$

which shows that the condition (C) holds. □

Concluding this section let us note that "multi-point" analogues of Criteria 3.4 and 3.5 formulated in terms of the whole finite-conditional distribution are of course valid. Concerning the first one, we would like to mention that its necessity statement was as a matter of fact contained in the proof of Lemma 1 of [24], whose argument we follow while proving Criterion 3.4. As to the second one, let us mention that the part utilizing the analogue of the condition (D) can be deduced from Theorems 1 and 2 of [20]. It should be pointed out that the author does not provide the proof of the sufficiency statement of Theorem 2 (leaving it, as he says, to the reader). However, our considerations show that the proof of this statement is neither intuitive, nor technically simple. Moreover, the validity of the statement seems doubtful in the settings of [20] where the state space is not supposed to be finite or even compact.

4. Some further development

The random field Gibbsianness criteria presented in the previous section are formulated either in terms of (one-point) conditional distribution, or in terms of (one-point) finite-conditional distribution. These two types of criteria are complementary, however, the second type criteria deal with an unambiguously defined constructive object and allow us to take a different look on and try to develop an alternative approach to the Gibbs theory. In this section we undertake some introductory steps in this direction.

First let us note, that roughly speaking, Criterion 3.4 asserts that aside from positivity considerations, Gibbs random fields are characterized by the uniform convergence of their one-point finite-conditional distribution (to the one-point conditional one), while only a weaker (almost sure) convergence is guaranteed for an arbitrary random field. In our opinion, this is perhaps the most comprehensible characterization of Gibbs random fields, on the basis of which the following probabilistically explicit definition of Gibbs random field can be given.

Definition 4.1. A random field $\mathbf{P}$ is called *Gibbs random field* if

1) for any $\Lambda \in \mathscr{E}$ and $\boldsymbol{x} \in \mathscr{X}^{\Lambda}$ one has $\mathbf{P}_{\Lambda}(\boldsymbol{x}) > 0$,

2) the limits

$$\lim_{\Lambda \uparrow \mathbb{Z}^{\nu} \setminus t} \frac{\mathbf{P}_{t \cup \Lambda}(x \overline{\boldsymbol{x}}_{\Lambda})}{\mathbf{P}_{\Lambda}(\overline{\boldsymbol{x}}_{\Lambda})}, \quad t \in \mathbb{Z}^{\nu},\ x \in \mathscr{X}^{t},\ \overline{\boldsymbol{x}} \in \mathscr{X}^{t^{c}}, \tag{4.1}$$

exist, are strictly positive, and the convergence is uniform with respect to $\overline{\boldsymbol{x}}$.

Note that if $\mathbf{P}$ is a Gibbs random field, then the limits (4.1) form a version of one-point conditional distribution of $\mathbf{P}$ and, moreover, their multi-point analogues exist and form a version of conditional distribution of $\mathbf{P}$. We call these

versions *canonical*. Note also, that the canonical (one-point) conditional distribution of a Gibbs random field is the only quasilocal version of its (one-point) conditional distribution. Note also that now, the Criterion 3.4 turns into the following theorem about representation of canonical conditional distribution of Gibbs random fields.

Theorem 4.1. *If* $\mathbf{P}$ *is a Gibbs random field, then the canonical* (*one-point*) *conditional distribution of* $\mathbf{P}$ *admits the representation given by Gibbs formulae* (1.3) *and* (1.4) *with some uniformly convergent potential.*

Conversely, if a random field $\mathbf{P}$ *has a version of* (*one-point*) *conditional distribution admitting the representation given by Gibbs formulae* (1.3) *and* (1.4) *with some uniformly convergent potential, then* $\mathbf{P}$ *is a Gibbs random field, and this version is canonical.*

The set $\mathscr{G}$ of all Gibbs random fields is not empty since, as it follows immediately from the above definition, it contains the set $\mathscr{M}$ of all strictly positive Markov random fields. On the other hand, as shows the following example, not all strictly positive random fields are Gibbsian.

Example 4.1. Let $\mathscr{X} = \{0,1\}$ and consider the random field $\mathbf{P}$ given by

$$\mathbf{P}_\Lambda(\boldsymbol{x}) = \frac{1}{(|\Lambda|+1)C_{|\Lambda|}^{|\boldsymbol{x}|}}, \quad \Lambda \in \mathscr{E},\ \boldsymbol{x} \in \mathscr{X}^\Lambda,$$

where $|\boldsymbol{x}| = |\{t \in \Lambda : x_t = 1\}|$ is the number of "particles" (ones) in the configuration $\boldsymbol{x}$. (This random field describes the situation when the number of particles in the volume Λ is distributed uniformly on the discrete interval $[\![0, \Lambda]\!]$ and, given that this number is k, all k-particle configurations are conditionally equiprobable.) First, for all $t \in \mathbb{Z}^\nu$, $\overline{\boldsymbol{x}} \in \mathscr{X}^{t^c}$ and $\Lambda \in \mathscr{E}^*(t^c)$ we have

$$\boldsymbol{q}_t^{\overline{\boldsymbol{x}}_\Lambda}(1) = \frac{\mathbf{P}_{t\cup\Lambda}(1\,\overline{\boldsymbol{x}}_\Lambda)}{\mathbf{P}_\Lambda(\overline{\boldsymbol{x}}_\Lambda)} = \frac{|\overline{\boldsymbol{x}}_\Lambda|+1}{|\Lambda|+2}.$$

Further, for any $p \in [0,1]$ let us denote by $\mathfrak{I}_t^p$ the set of all $\overline{\boldsymbol{x}} \in \mathscr{X}^{t^c}$ such that

$$\exists\ \lim_{\Lambda\uparrow t^c} \frac{|\overline{\boldsymbol{x}}_\Lambda|}{|\Lambda|} = p(\overline{\boldsymbol{x}}) = p,$$

and put $\overline{\mathfrak{I}}_t = \mathscr{X}^{t^c} \setminus \big(\bigcup_{p\in[0,1]} \mathfrak{I}_t^p \big)$. Now we see that the limits (4.1) do not exist for $\overline{\boldsymbol{x}} \in \overline{\mathfrak{I}}_t$ and are not all strictly positive for $\overline{\boldsymbol{x}} \in \mathfrak{I}_t^0 \cup \mathfrak{I}_t^1$. Each one of these facts yields the non-Gibbsianness of $\mathbf{P}$.

Let us note that the random field $\mathbf{P}$ considered in this example is the uniform mixture of Bernoulli random fields $\mathbf{B}^p$, $p \in (0,1)$. Indeed, for all $\Lambda \in \mathscr{E}$ and

$\boldsymbol{x} \in \mathscr{X}^\Lambda$, we have

$$\mathbf{P}_\Lambda(\boldsymbol{x}) = \frac{|\boldsymbol{x}|!\,(|\Lambda| - |\boldsymbol{x}|)!}{(|\Lambda|+1)!} = \mathrm{B}\big(|\boldsymbol{x}|+1, |\Lambda| - |\boldsymbol{x}| + 1\big)$$
$$= \int\limits_0^1 p^{|\boldsymbol{x}|}(1-p)^{|\Lambda|-|\boldsymbol{x}|}\,dp.$$

Now, the non-Gibbsianness of $\mathbf{P}$ also follows from the general fact that any non-trivial mixture of Bernoulli random fields is non Gibbsian (see the authors' work [5], as well as Section 4.5.1 of [12] and Section 4 of [13] for the case of finite or countably infinite mixtures). Let us give two other examples of such mixtures where the finite-dimensional distributions are explicit and permit to check the non-Gibbsianness directly.

Example 4.2. Let $\tau > 0$ and consider the random field $\mathbf{P}$ which is the mixture with the density $\tau p^{\tau-1}$ of Bernoulli random fields $\mathbf{B}^p$, $p \in (0,1)$, that is,

$$\mathbf{P}_\Lambda(\boldsymbol{x}) = \int\limits_0^1 p^{|\boldsymbol{x}|}(1-p)^{|\Lambda|-|\boldsymbol{x}|}\tau p^{\tau-1}\,dp$$
$$= \tau\mathrm{B}\big(|\boldsymbol{x}|+\tau, |\Lambda| - |\boldsymbol{x}| + 1\big), \quad \Lambda \in \mathscr{E},\ \boldsymbol{x} \in \mathscr{X}^\Lambda.$$

(This is a generalization of the previous example, the latter being obtained for $\tau = 1$.) Here, for all $t \in \mathbb{Z}^\nu$, $\overline{\boldsymbol{x}} \in \mathscr{X}^{t^{\mathrm{c}}}$ and $\Lambda \in \mathscr{E}^*(t^{\mathrm{c}})$ we have

$$\boldsymbol{q}_t^{\overline{\boldsymbol{x}}_\Lambda}(1) = \frac{\mathbf{P}_{t\cup\Lambda}(1\,\overline{\boldsymbol{x}}_\Lambda)}{\mathbf{P}_\Lambda(\overline{\boldsymbol{x}}_\Lambda)} = \frac{|\overline{\boldsymbol{x}}_\Lambda| + \tau}{|\Lambda| + \tau + 1},$$

and so, all the considerations of the previous example hold.

Example 4.3. Let $\alpha, p_1, p_2 \in (0,1)$ such that $p_1 \neq p_2$ and consider the random field $\mathbf{P}$ which is the mixture of Bernoulli random fields $\mathbf{B}^{p_1}$ and $\mathbf{B}^{p_2}$ with the coefficients α and $\beta = 1 - \alpha$, that is,

$$\mathbf{P}_\Lambda(\boldsymbol{x}) = \alpha p_1^{|\boldsymbol{x}|}(1-p_1)^{|\Lambda|-|\boldsymbol{x}|} + \beta p_2^{|\boldsymbol{x}|}(1-p_2)^{|\Lambda|-|\boldsymbol{x}|}, \quad \Lambda \in \mathscr{E},\ \boldsymbol{x} \in \mathscr{X}^\Lambda.$$

Here, for all $t \in \mathbb{Z}^\nu$, $\overline{\boldsymbol{x}} \in \mathscr{X}^{t^{\mathrm{c}}}$ and $\Lambda \in \mathscr{E}^*(t^{\mathrm{c}})$ we have

$$\boldsymbol{q}_t^{\overline{\boldsymbol{x}}_\Lambda}(1) = \frac{\mathbf{P}_{t\cup\Lambda}(1\,\overline{\boldsymbol{x}}_\Lambda)}{\mathbf{P}_\Lambda(\overline{\boldsymbol{x}}_\Lambda)} = \frac{\alpha p_1 + \beta p_2 \exp\{|\Lambda| H_\Lambda(\overline{\boldsymbol{x}}_\Lambda)\}}{\alpha + \beta \exp\{|\Lambda| H_\Lambda(\overline{\boldsymbol{x}}_\Lambda)\}},$$

where

$$H_\Lambda(\overline{\boldsymbol{x}}_\Lambda) = \frac{|\overline{\boldsymbol{x}}_\Lambda|}{|\Lambda|}\ln\frac{p_2}{p_1} + \Big(1 - \frac{|\overline{\boldsymbol{x}}_\Lambda|}{|\Lambda|}\Big)\ln\frac{1-p_2}{1-p_1}.$$

Clearly, there exist a configuration $\overline{\boldsymbol{z}} \in \overline{\mathfrak{I}}_t$ such that $|\overline{\boldsymbol{z}}_\Lambda|/|\Lambda|$ is oscillating between 0 and 1, that is,

$$\liminf_{\Lambda\uparrow t^{\mathrm{c}}} \frac{|\overline{\boldsymbol{z}}_\Lambda|}{|\Lambda|} = 0 \quad \text{and} \quad \limsup_{\Lambda\uparrow t^{\mathrm{c}}} \frac{|\overline{\boldsymbol{z}}_\Lambda|}{|\Lambda|} = 1.$$

Then, since $\ln(p_2/p_1)$ and $\ln((1-p_2)/(1-p_1))$ are of opposite signs, $|\Lambda|H_\Lambda(\overline{\boldsymbol{z}}_\Lambda)$ is oscillating between $-\infty$ and $+\infty$. Hence, the limits (4.1) do not exist for $\overline{\boldsymbol{x}} = \overline{\boldsymbol{z}}$, and so the random field $\mathbf{P}$ is non-Gibbsian.

The preceding examples show that $\mathscr{G} \subsetneq \mathscr{P}$. However, $\mathscr{G}$ is dense in $\mathscr{P}$ with respect to the topology of weak convergence ($\mathbf{P}^n \to \mathbf{P}$ if $\mathbf{P}^n_\Lambda(\boldsymbol{x}) \to \mathbf{P}_\Lambda(\boldsymbol{x})$ for all $\Lambda \in \mathscr{E}$ and $\boldsymbol{x} \in \mathscr{X}^\Lambda$. Indeed, let $\mathbf{P} \in \mathscr{P}$, let $\mathbf{B}^p$ be a Bernoulli random field (with some $p \in (0,1)$, and let $\Lambda_n \in \mathscr{E}$, $n \in \mathbb{N}$, such that $\Lambda_n \uparrow \mathbb{Z}^\nu$ as $n \to \infty$. For each n, consider the random field $\mathbf{P}^n$ such that its restrictions on Λ_n and Λ_n^{c} are independent and given by $(\mathbf{P}^n)_{\Lambda_n} = (\mathbf{P})_{\Lambda_n}$ and $(\mathbf{P}^n)_{\Lambda_n^{\mathrm{c}}} = (\mathbf{B}^p)_{\Lambda_n^{\mathrm{c}}}$. Clearly, $\mathbf{P}^n \in \mathscr{M}$ and $\mathbf{P}^n \to \mathbf{P}$ as $n \to \infty$. So, we have shown that $\mathscr{M}$ (and hence, a fortiori, $\mathscr{G}$) is dense in $\mathscr{P}$. Note that a similar (though limited to the scope of translation invariant random fields) statement for the case of a general (not necessarily finite) state space can be found in Section 4.5.6 of [12].

As we have seen above, the mixtures and the limits of Gibbs random fields can be non-Gibbsian. However, this is no longer the case if we consider Gibbs random fields having the same canonical (one-point) conditional distribution. Indeed, we will see below that we have even more: for any $\mathbf{P}^\circ \in \mathscr{G}$, the set $\mathscr{G}(\mathbf{P}^\circ)$ of all Gibbs random fields having the same canonical (one-point) conditional distribution as $\mathbf{P}^\circ$ (which is equivalently the set of all random fields consistent with the canonical (one-point) conditional distribution of $\mathbf{P}^\circ$) is convex and closed.

First we show the convexity. Let $\mathbf{P}^1, \mathbf{P}^2 \in \mathscr{G}(\mathbf{P}^\circ)$, let $\{\boldsymbol{q}_t^{\overline{\boldsymbol{x}}},\ t \in \mathbb{Z}^\nu$ and $\overline{\boldsymbol{x}} \in \mathscr{X}^{t^{\mathrm{c}}}\}$ be their common canonical one-point conditional distribution, let $\alpha \in (0,1)$ and put $\mathbf{P} = \alpha\,\mathbf{P}^1 + \beta\,\mathbf{P}^2$, where $\beta = 1-\alpha$. The strict positivity of $\mathbf{P}$ is evident. Further, for all $t \in \mathbb{Z}^\nu$ and $x \in \mathscr{X}^t$, we have

$$\begin{aligned}
&\sup_{\overline{\boldsymbol{x}} \in \mathscr{X}^{t^{\mathrm{c}}}} \left| \frac{\mathbf{P}_{t\cup\Lambda}(x\overline{\boldsymbol{x}}_\Lambda)}{\mathbf{P}_\Lambda(\overline{\boldsymbol{x}}_\Lambda)} - \boldsymbol{q}_t^{\overline{\boldsymbol{x}}}(x) \right| \\
&\quad = \sup_{\overline{\boldsymbol{x}} \in \mathscr{X}^{t^{\mathrm{c}}}} \left| \frac{\alpha\,\mathbf{P}^1_\Lambda(\overline{\boldsymbol{x}}_\Lambda)\big(\,\mathbf{P}^1_{t\cup\Lambda}(x\overline{\boldsymbol{x}}_\Lambda)/\,\mathbf{P}^1_\Lambda(\overline{\boldsymbol{x}}_\Lambda) - \boldsymbol{q}_t^{\overline{\boldsymbol{x}}}(x)\big)}{\alpha\,\mathbf{P}^1_\Lambda(\overline{\boldsymbol{x}}_\Lambda) + \beta\,\mathbf{P}^2_\Lambda(\overline{\boldsymbol{x}}_\Lambda)} \right. \\
&\qquad \left. + \frac{\beta\,\mathbf{P}^2_\Lambda(\overline{\boldsymbol{x}}_\Lambda)\big(\,\mathbf{P}^2_{t\cup\Lambda}(x\overline{\boldsymbol{x}}_\Lambda)/\,\mathbf{P}^2_\Lambda(\overline{\boldsymbol{x}}_\Lambda) - \boldsymbol{q}_t^{\overline{\boldsymbol{x}}}(x)\big)}{\alpha\,\mathbf{P}^1_\Lambda(\overline{\boldsymbol{x}}_\Lambda) + \beta\,\mathbf{P}^2_\Lambda(\overline{\boldsymbol{x}}_\Lambda)} \right| \\
&\quad \leqslant \sup_{\overline{\boldsymbol{x}} \in \mathscr{X}^{t^{\mathrm{c}}}} \left| \frac{\mathbf{P}^1_{t\cup\Lambda}(x\overline{\boldsymbol{x}}_\Lambda)}{\mathbf{P}^1_\Lambda(\overline{\boldsymbol{x}}_\Lambda)} - \boldsymbol{q}_t^{\overline{\boldsymbol{x}}}(x) \right| + \sup_{\overline{\boldsymbol{x}} \in \mathscr{X}^{t^{\mathrm{c}}}} \left| \frac{\mathbf{P}^2_{t\cup\Lambda}(x\overline{\boldsymbol{x}}_\Lambda)}{\mathbf{P}^2_\Lambda(\overline{\boldsymbol{x}}_\Lambda)} - \boldsymbol{q}_t^{\overline{\boldsymbol{x}}}(x) \right|,
\end{aligned}$$

and hence $\mathbf{P} \in \mathscr{G}(\mathbf{P}^\circ)$.

Now let us verify that the set $\mathscr{G}(\mathbf{P}^\circ)$ is closed. Let $\mathbf{P}^n \in \mathscr{G}(\mathbf{P}^\circ)$, $n \in \mathbb{N}$, such that $\mathbf{P}^n \to \overline{\mathbf{P}}$ as $n \to \infty$, and let $\{\boldsymbol{q}_\Lambda^{\overline{\boldsymbol{x}}},\ \Lambda \in \mathscr{E} \text{ and } \overline{\boldsymbol{x}} \in \mathscr{X}^{\Lambda^c}\}$ be the common canonical conditional distribution of the elements of $\mathscr{G}(\mathbf{P}^\circ)$. To show that $\overline{\mathbf{P}}$ is strictly positive, let us note that for all $\Lambda \in \mathscr{E}$ and $\boldsymbol{x} \in \mathscr{X}^\Lambda$ we have

$$\mathbf{P}_\Lambda(\boldsymbol{x}) \geqslant \inf_{\overline{\boldsymbol{x}} \in \mathscr{X}^{\Lambda^c}} \boldsymbol{q}_\Lambda^{\overline{\boldsymbol{x}}}(\boldsymbol{x})$$

for any $\mathbf{P} \in \mathscr{G}(\mathbf{P}^\circ)$. Hence,

$$\inf_{n \in \mathbb{N}} \mathbf{P}_\Lambda^n(\boldsymbol{x}) \geqslant \inf_{\mathbf{P} \in \mathscr{G}(\mathbf{P}^\circ)} \mathbf{P}_\Lambda(\boldsymbol{x}) \geqslant \inf_{\overline{\boldsymbol{x}} \in \mathscr{X}^{\Lambda^c}} \boldsymbol{q}_\Lambda^{\overline{\boldsymbol{x}}}(\boldsymbol{x}) > 0,$$

and so $\overline{\mathbf{P}}_\Lambda(\boldsymbol{x}) > 0$. Further, according to our definition of Gibbs random field, for all $t \in \mathbb{Z}^\nu$ and $x \in \mathscr{X}^t$, we have

$$\lim_{\Lambda \uparrow \mathbb{Z}^\nu \setminus t} \sup_{\overline{\boldsymbol{x}} \in \mathscr{X}^{t^c}} \left| \frac{\mathbf{P}_{t \cup \Lambda}(x \overline{\boldsymbol{x}}_\Lambda)}{\mathbf{P}_\Lambda(\overline{\boldsymbol{x}}_\Lambda)} - \boldsymbol{q}_t^{\overline{\boldsymbol{x}}}(x) \right| = 0$$

for any $\mathbf{P} \in \mathscr{G}(\mathbf{P}^\circ)$. Moreover, taking a closer look on the necessity part of the proof of Criterion 3.4 one can see that all the estimates therein are based on the canonical one-point conditional distribution only, and so, it becomes clear that the above convergence is also uniform with respect to $\mathbf{P}$, that is,

$$\lim_{\Lambda \uparrow \mathbb{Z}^\nu \setminus t} \sup_{\mathbf{P} \in \mathscr{G}(\mathbf{P}^\circ)} \sup_{\overline{\boldsymbol{x}} \in \mathscr{X}^{t^c}} \left| \frac{\mathbf{P}_{t \cup \Lambda}(x \overline{\boldsymbol{x}}_\Lambda)}{\mathbf{P}_\Lambda(\overline{\boldsymbol{x}}_\Lambda)} - \boldsymbol{q}_t^{\overline{\boldsymbol{x}}}(x) \right| = 0.$$

Hence, for all sufficiently large Λ, the quantity

$$\sup_{\overline{\boldsymbol{x}} \in \mathscr{X}^{t^c}} \left| \frac{\mathbf{P}_{t \cup \Lambda}^n(x \overline{\boldsymbol{x}}_\Lambda)}{\mathbf{P}_\Lambda^n(\overline{\boldsymbol{x}}_\Lambda)} - \boldsymbol{q}_t^{\overline{\boldsymbol{x}}}(x) \right|$$

can be made smaller than an arbitrary $\varepsilon > 0$ for all $n \in \mathbb{N}$. For any such Λ, by passing to the limit as $n \to \infty$, we get

$$\sup_{\overline{\boldsymbol{x}} \in \mathscr{X}^{t^c}} \left| \frac{\overline{\mathbf{P}}_{t \cup \Lambda}(x \overline{\boldsymbol{x}}_\Lambda)}{\overline{\mathbf{P}}_\Lambda(\overline{\boldsymbol{x}}_\Lambda)} - \boldsymbol{q}_t^{\overline{\boldsymbol{x}}}(x) \right| \leqslant \varepsilon,$$

and so, $\overline{\mathbf{P}} \in \mathscr{G}(\mathbf{P}^\circ)$.

In conclusion we would like to recall that this section is only a first step in the development of the above mentioned alternative approach. In our opinion, the realization of the potential of this approach will make a real contribution to the Gibbs theory, namely in such problems as uniqueness, decay of correlations and limit theorems. For example, some results concerning the problem of decay of correlations can already be found in [7].

References

[1] M.B. Averintsev (1970) A way of describing random fields with a discrete argument. *Problems Inform. Transmission* **6**, 169–175.

[2] M.B. Averintsev (1972) Description of Markovian random fields by Gibbsian conditional probabilities. *Theory Probab. and Appl.* **17**, 20–33.

[3] M.B. Averintsev (1975) Gibbsian distribution of random fields whose conditional probabilities may vanish. *Problems Inform. Transmission* **11**, 326–334.

[4] S. Dachian and B.S. Nahapetian (1998) An approach towards description of random fields. Université du Maine, preprint no. 98-5, Mai 1998.

[5] S. Dachian and B.S. Nahapetian (2001) Description of random fields by means of one-point conditional distributions and some applications. *Markov Processes Relat. Fields* **7**, 193–214.

[6] S. Dachian and B.S. Nahapetian (2004) Description of specifications by means of probability distributions in small volumes under condition of very weak positivity. *J. Stat. Phys.* **117** (1–2), 281–300.

[7] A. Dalalyan and B.S. Nahapetian (2008) Description of random fields by means of one-point finite-conditional distributions. Submitted.

[8] R.L. Dobrushin (1968) The description of a random field by means of conditional probabilities and conditions of its regularity. *Theory Probab. and Appl.* **13**, 197–224.

[9] R.L. Dobrushin (1968) Gibbsian random fields for lattice systems with pairwise interaction. *Funct. Anal. Appl.* **2**, 292–301.

[10] R.L. Dobrushin (1968) The problem of uniqueness of a Gibbs random field and the problem of phase transitions. *Funct. Anal. Appl.* **2**, 302–312.

[11] R.L. Dobrushin (1970) Prescribing a system of random variables by conditional distributions. *Theory Probab. and Appl.* **15**, 458–486.

[12] A.C.D van Enter, R. Fernández and A.D. Sokal (1993) Regularity properties and pathologies of position-space renormalization-group transformations: scope and limitations of Gibbsian theory. *J. Stat. Phys.* **72** (5–6), 879–1167.

[13] A.C.D van Enter and J. Lörinczi (1996) Robustness of the non-Gibbsian property: some examples. *J. Phys. A* **29** (10), 2465–2473.

[14] R. Fernández and G. Maillard (2004) Chains with complete connections and one-dimensional Gibbs measures. *Electron. J. Probab.* **9**, 145–176.

[15] R. Fernández and G. Maillard (2006) Construction of a specification from its singleton part. *ALEA Lat. Am. J. Probab. Math. Stat.* **2**, 297–315.

[16] R.G. Flood and W.G. Sullivan (1980) Consistency of random field specifications. *Z. Wahrsch. Verw. Geb.* **53**, 147–156.

[17] H.-O. Georgii (1988) *Gibbs Measures and Phase Transitions.* De Gruyter, Berlin.

[18] S. Goldstein (1978) A note on specifications. *Z. Wahrsch. Verw. Geb.* **46**, 45–51.

[19] G.R. Grimmett (1973) A theorem about random fields. *Bull. London Math. Soc.* **5**, 81–84.

[20] O.K. Kozlov (1974) Gibbs description of a system of random variables. *Problems Inform. Transmission* **10**, 258–265.

[21] C. Preston (1980) Construction of specifications. In: *Quantum Fields – Algebras*, Processes (Bielefeld symposium 1978), L. Streit (ed.) Springer, Wien – New York, 269–292.

[22] A.D. Sokal (1981) Existence of compatible families of proper regular conditional probabilities. *Z. Wahrsch. Verw. Geb.* **56**, 537–548.

[23] W.G. Sullivan (1973) Finite range random fields and energy fields. *J. Math. Anal. Appl.* **44**, 710–724.

[24] W.G. Sullivan (1973) Potentials for almost Markovian random fields. *Commun. Math. Phys.* **33**, 61–74.

[25] P. Suomela (1976) *Construction of Nearest Neighbour Systems.* Ann. Acad. Scient. Fennicae, Series A, I. Mathematica Dissertationes **10**.

Troisième Partie

Statistique des Champs Aléatoires

Statistical Inference for Stochastic Processes **1:** 245–264, 1998.

Nonparametric Estimation for Gibbs Random Fields Specified Through One-Point Systems

S. DACHIAN
Laboratoire de Statistique et Processus, Université du Maine, Av. Olivier Messiaen, 72085 Le Mans Cedex 9, France

Abstract. The problem of nonparametric estimation for Gibbs random fields is considered. The field is supposed to be specified through a translation invariant quasilocal one-point system. An estimator of one-point system is constructed by the method of sieves, and its exponential and $\mathbf{L}^p$ consistencies are proved in different setups. The results hold regardless of non-uniqueness and translation invariance breaking.

Mathematics Subject Classification (1991): 60K35, 62M40.

Key words: Gibbs random fields, quasilocality, one-point systems, nonparametric estimation, method of sieves, consistency.

1. Introduction

This paper is devoted to nonparametric estimation problem for a class of random fields defined on the ν-dimensional integer lattice $\mathbb{Z}^\nu$, $\nu \geqslant 1$ and taking values in the state space $\mathcal{X} = \{0, 1\}$. An approach towards description of such fields was introduced in our joint papers with B. S. Nahapetian [2–4]. The main idea of this approach is to describe random fields by specifications (just as in Gibbs random field theory) but to define specifications by some systems of real numbers like Q-functions, H-functions, Q-systems, H-systems and one-point systems, rather than by interaction potentials. This approach permits one also to describe non-Gibbsian random fields and provides a parametrization of random fields suitable for statistical inference.

In this paper we consider the problem of nonparametric estimation of a one-point system. We construct an estimator by combining the idea of approximating a ratio of conditional probabilities by a ratio of some empirical conditional frequencies with the main idea of the method of sieves (introduced by U. Grenander [10]): approximation of infinite-dimensional parameters by finite-dimensional ones. We prove exponential consistency and $\mathbf{L}^p$-consistency, for all $p \in (0, \infty)$, of our sieve estimator in different setups.

Let us note here that for maximum likelihood estimators F. Comets in [1] also gets exponential consistency in the parametric case and in a classical Gibbsian setup using the theory of large deviations.

Note too that in general the problem of estimation for Gibbs random fields is complicated by such classical phenomenons of Gibbs random fields theory as non-uniqueness (presence of a phase transition) and translation invariance breaking. In our work the results are established regardless of this aspects of Gibbs random fields theory.

Parametric statistical inference for Gibbs random fields is now quite well developed in classical Gibbsian setup. The actual state of the theory is well presented in the monograph by X. Guyon [11] and references therein. For more information see [1, 8, 12–14].

As to nonparametric inference, it seems to be less well investigated. We can mention here a preprint by C. Ji [12]. He considers a classical Gibbsian setup where the random filed is described by an exponentially decreasing pair-interaction potential. For this model he studies the sieve estimator of 'local characteristics'. The proof presented there needs some rectification. Our work is similar to [12] in that our one-point system is in fact something similar to local characteristics, and in that we study the sieve estimator. But unlike [12], our setup is much more general and in our case we estimate the object (one-point system) which itself describes the random field.

Let us finally note here that the results of this paper were presented with more detail in [2]. All the results hold in the case of arbitrary finite state space $\mathcal{X}$. See [2] for more details about this case.

2. Preliminary Results

In this section we recall some basic notions of the Gibbs random fields theory and introduce the notion of one-point systems.

2.1. RECAPITULATION OF GIBBS RANDOM FIELDS THEORY

We consider random fields on the ν-dimensional integer lattice $\mathbb{Z}^\nu$, i.e., probability measures on $(\Omega, \mathcal{F}) = (\mathcal{X}^{\mathbb{Z}^\nu}, \mathcal{F}_0^{\mathbb{Z}^\nu})$. For simplicity the state space $\mathcal{X}$ (space of values of a single variable) is assumed to consist of two points and be endowed with the total σ-algebra (the σ-algebra consisting fo all subsets of $\mathcal{X}$), i.e., $(\mathcal{X}, \mathcal{F}_0) = (\{0, 1\}, \exp\{0, 1\})$. Denote $\mathcal{E} = \{\Lambda \subset \mathbb{Z}^\nu\colon\ |\Lambda| < \infty\}$ the set of all fintie subsets of $\mathbb{Z}^\nu$. Here and in the sequel $|\Lambda|$ is the number of points of the set Λ. For any $\Lambda \in \mathcal{E}$ let

$$\mathcal{X}^\Lambda = \{\mathbf{x} = \{x_t, t \in \Lambda\} : x_t \in \mathcal{X} \text{ for all } t \in \Lambda\}$$

be the set of all configurations (realizations) on Λ. Clearly, each element $\mathbf{x} \in \mathcal{X}^\Lambda$ is uniquely determined by the subset of Λ where the configuration $\mathbf{x}$ assumes the value 1 (in physical terminology this subset is occupied by particles). Therefore, we can identify any configuration $\mathbf{x}$ on Λ with the corresponding subset X of Λ. In the sequel we will not make any distinction between these two notions and we will

write $\mathbf{x} \subset \Lambda$ for a configuration $\mathbf{x}$ on Λ. Note that $\mathcal{E}$ is countable. Note too that by definition $\mathcal{F}$ is the smallest σ-algebra on Ω containing the cyclinder events

$$\{\mathbf{x} \in \Omega : \mathbf{x}_\Lambda \in A\}, \quad \Lambda \in \mathcal{E}, A \in \mathcal{F}_0^\Lambda.$$

Here and in the sequel $\mathbf{x}_\Lambda = \mathbf{x} \cap \Lambda \subset \Lambda$ is the subconfiguration (restriction) on Λ of the configuration $\mathbf{x}$.

A probability distribution on $\mathcal{X}^\Lambda$, $\Lambda \in \mathcal{E}$, is denoted by

$$\mathbf{P}_\Lambda = \{\mathbf{P}_\Lambda(\mathbf{x}), \mathbf{x} \subset \Lambda\}.$$

For $\Lambda = \emptyset$, we consider that there exits only one probability distribution $\mathbf{P}_\emptyset(\emptyset) = 1$.

It is well known that a probability measure on $\mathcal{X}^{\mathbb{Z}^\nu}$ (or, equivalently, a random field on $\mathbb{Z}^\nu$) can be described in terms of its *finite-dimensional distributions* which are consistent in the sense that $(\mathbf{P}_\Lambda)_I = \mathbf{P}_I$. Here and in the sequel the probability distribution $(\mathbf{P}_\Lambda)_I$ on $\mathcal{X}^I$ is the restriction of $\mathbf{P}_\Lambda$ on I.

Let us now introduce the concept of conditional distribution of a random field.

For this we need the notion of convergence of nets (sequences) of real numbers indexed by elements of $\mathcal{E}$. Let $\{a_\Lambda, \Lambda \in \mathcal{E}\}$ be a real-valued function on $\mathcal{E}$ and $T \subset \mathbb{Z}^\nu$ be an infinite subset of $\mathbb{Z}^\nu$. We say that $\lim_{\Lambda \uparrow T} a_\Lambda = a_T$ if for any sequence $\Lambda_n \in \mathcal{E}$ such that $\Lambda_n \uparrow T$ we have the convergence $\lim_{n\to\infty} a_{\Lambda_n} = a_T$.

Let $\mathbf{P}$ be a random field. It is well known that for any $\Lambda \in \mathcal{E}$ there exist for $\mathbf{P}_{\Lambda^c}$'-almost all $\overline{\mathbf{x}} \subset \Lambda^c = \mathbb{Z}^\nu \backslash \Lambda$ the following limits:

$$\mathbf{q}_\Lambda^{\overline{\mathbf{x}}}(\mathbf{x}) = \lim_{\tilde{\Lambda} \uparrow \Lambda^c} \frac{\mathbf{P}_{\Lambda \cup \tilde{\Lambda}}(\mathbf{x} \cup \overline{\mathbf{x}}_{\tilde{\Lambda}})}{\mathbf{P}_{\tilde{\Lambda}}(\overline{\mathbf{x}}_{\tilde{\Lambda}})}, \quad \mathbf{x} \subset \Lambda.$$

Any system

$$\mathcal{Q} = \{\mathbf{Q}_\Lambda^{\overline{\mathbf{x}}}, \quad \Lambda \in \mathcal{E} \text{ and } \overline{\mathbf{x}} \subset \Lambda^c\}$$

of probability distributions in various finite volumes Λ with various boundary conditons $\overline{\mathbf{x}}$ on Λ^c such that for all $\Lambda \in \mathcal{E}$ we have $\mathbf{Q}_\Lambda^{\overline{\mathbf{x}}} = \mathbf{q}_\Lambda^{\overline{\mathbf{x}}}$ for $\mathbf{P}_{\Lambda^c}$-almost all $\overline{\mathbf{x}} \subset \Lambda^c$ is called *conditional distribution* of the random field $\mathbf{P}$.

It is also well known that any conditonal distribution $\mathcal{Q}$ of a random field $\mathbf{P}$ satisfies $\mathbf{P}$-almost surely the condition

$$\mathbf{Q}_{\Lambda \cup \tilde{\Lambda}}^{\overline{\mathbf{x}}}(\mathbf{x} \cup \mathbf{y}) = \mathbf{Q}_\Lambda^{\overline{\mathbf{x}} \cup \mathbf{y}}(\mathbf{x})(\mathbf{Q}_{\Lambda \cup \tilde{\Lambda}}^{\overline{\mathbf{x}}})_{\tilde{\Lambda}}(\mathbf{y}) \tag{1}$$

where $\Lambda, \tilde{\Lambda} \in \mathcal{E}$, $\Lambda \cap \tilde{\Lambda} = \emptyset$, $\mathbf{x} \subset \Lambda$, $\mathbf{y} \subset \tilde{\Lambda}$ and $\overline{\mathbf{x}} \subset (\Lambda \cup \tilde{\Lambda})^c$. In fact, this is nothing but the elementary formula

$$\mathbf{P}(A \cap B|C) = \mathbf{P}(A|B \cap C)\mathbf{P}(B|C) \tag{2}$$

written for our case.

Now let us consider an arbitrary system

$$\mathcal{Q} = \{\mathbf{Q}_\Lambda^{\overline{\mathbf{x}}}, \quad \Lambda \in \mathcal{E} \text{ and } \overline{\mathbf{x}} \subset \Lambda^c\}$$

of probability distributions in finite volumes with boundary conditions. If we want this system to be a conditional distribution of some random field $\mathbf{P}$, then we need to suppose that it $\mathbf{P}$-almost surely satisfies the condition (1). However, we do not know the random field $\mathbf{P}$ *a priori*. Therefore we need to require that the condition (1) holds always, rather than almost surely. This leads us to introduce the following

DEFINITION 1. A system

$$\mathcal{Q} = \{\mathbf{Q}_\Lambda^{\overline{\mathbf{x}}}, \quad \Lambda \in \mathcal{E} \text{ and } \overline{\mathbf{x}} \subset \Lambda^c\}$$

of probability distributions in finite volumes with boundary conditions is called specification, if for any $\Lambda, \tilde{\Lambda} \in \mathcal{E}$ such that $\Lambda \cap \tilde{\Lambda} = \emptyset$ and for any $\mathbf{x} \subset \Lambda$, $\mathbf{y} \subset \tilde{\Lambda}$ and $\overline{\mathbf{x}} \subset (\Lambda \cup \tilde{\Lambda})^c$ we have

$$\mathbf{Q}_{\Lambda\cup\tilde{\Lambda}}^{\overline{\mathbf{x}}}(\mathbf{x} \cup \mathbf{y}) = \mathbf{Q}_\Lambda^{\overline{\mathbf{x}}\cup\mathbf{y}}(\mathbf{x})(\mathbf{Q}_{\Lambda\cup\tilde{\Lambda}}^{\overline{\mathbf{x}}})_{\tilde{\Lambda}}(\mathbf{y}).$$

In Gibbs random field theory a random field is described through a specification $\mathcal{Q} = \{\mathbf{Q}_\Lambda^{\overline{\mathbf{x}}}, \Lambda \in \mathcal{E} \text{ and } \overline{\mathbf{x}} \subset \Lambda^c\}$. Usually the specification is assumed to have so-called Gibbsian form, written in terms of some physical quantities like interaction potentials, but we do not suppose that here. Any random field having the specification $\mathcal{Q}$ as a conditional distribution is called a *Gibbs random field* for $\mathcal{Q}$. Note that we use the traditional term 'Gibbs' even though $\mathcal{Q}$ does not necessarily have Gibbsian form. The main question of the Gibbs random field theory is the study of the set $\mathcal{G} = \mathcal{G}(\mathcal{Q})$ of all Gibbs random fields for $\mathcal{Q}$. Is this set empty of not? If it is not empty, is it a singleton or not, i.e., is the field having $\mathcal{Q}$ as a conditonal distribution unique or not? In the non-uniqueness case, what can be said about the structure of this set? Another interesting question is the following. Suppose that $\mathcal{Q}$ is translation invariant (i.e., invariant with respect to shift operators on $\mathbb{Z}^\nu$ or, in other words, stationary). Are all the random fields from $\mathcal{G}(\mathcal{Q})$ translation invariant or not? In the latter case what can be said about the subset $\mathcal{G}_{\text{t.i}} = \mathcal{G}_{\text{t.i}}(\mathcal{Q})$ of translation invariant random field?

These questions are answered in a general setup, when the specification $\mathcal{Q}$ is not supposed to have Gibbsian form, but rather is supposed to be *quasilocal*. In this case the sets are non-empty and the structure of the sets can be studied. Note, that it is possible to have *non-uniqueness* ($|\mathcal{G}| > 1, |\mathcal{G}_{\text{t.i.}}| > 1$) and *translation invariance breaking* ($\mathcal{G} \neq \mathcal{G}_{\text{t.i.}}$). Note also, that in our work the results are established regardless of these phenomena of Gibbs random field theory, since they hold uniformly on $\mathcal{G}$. For detailed exposition of Gibbs random field theory see the works of R. L. Dobrushin [5–7] for the Gibbsian case and the excellent book of H.-O. Georgii [9] for the general case.

DEFINITION 2. Let $\mathbf{g} = \{g^{\mathbf{x}}, \mathbf{x} \subset \mathbb{Z}^\nu\}$ be an arbitrary real-valued function on Ω.

(1) We say that the function $\mathbf{g}$ is *local* if it is $\mathcal{F}_0^\Lambda$ measurable for some $\Lambda \in \mathcal{E}$, i.e., if it depends only on the restriction $\mathbf{x}_\Lambda$ of $\mathbf{x}$ on Λ or, equivalently, if we have $g^{\mathbf{x}} = g^{\mathbf{x}_\Lambda}$ for all $\mathbf{x} \in \Omega$.

(2) We say that the function $\mathbf{g}$ is *quasilocal* if we have $\lim_{I \uparrow \mathbb{Z}^\nu} g^{\mathbf{x}_I} = g^{\mathbf{x}}$ uniformly on $\mathbf{x} \in \Omega$, i.e.,

$$\sup_{\mathbf{x} \in \Omega} |g^{\mathbf{x}_I} - g^{\mathbf{x}}| \underset{I \uparrow \mathbb{Z}^\nu}{\longrightarrow} 0.$$

DEFINITION 3. A specification $\mathcal{Q}$ is called (quasi)local if for all $\Lambda \in \mathcal{E}$ and $\mathbf{x} \subset \Lambda$ the function $\{\mathbf{Q}_\Lambda^{\overline{\mathbf{z}}_{\Lambda^c}}(\mathbf{x}), \overline{\mathbf{z}} \in \Omega\}$ is (quasi)local, i.e., if for all $\Lambda \in \mathcal{E}$ and $\mathbf{x} \subset \Lambda$ the quantity

$$\varphi_{\mathbf{x},\Lambda}(I) = \sup_{\overline{\mathbf{x}} \subset \Lambda^c} |\mathbf{Q}_\Lambda^{\overline{\mathbf{x}}_I}(\mathbf{x}) - \mathbf{Q}_\Lambda^{\overline{\mathbf{x}}}(\mathbf{x})|$$

tends to 0 as $I \uparrow \mathbb{Z}^\nu$ (for the quasilocal case) or equals 0 for I sufficiently large (for the local case). A random field $\mathbf{P}$ is called (quasi)local if it has a (quasi)local conditional distribution.

2.2. DESCRIPTION OF SPECIFICATIONS BY MEANS OF ONE-POINT SYSTEMS

As we have already seen the notion of the specification plays central role in the (Gibbs) random fields theory. In [2–4] an approach towards description of specifications was developed. In this approach speficiations are described by some systems of real numbers like Q- functions, Q-systems, H-systems and one-point systems, rather than by interaction potentials. Here we recall some results about description of specifications by consistent H-systems and one-point systems. Details and proofs can be found in [2] or in [4].

DEFINITION 4. A system $\mathcal{H} = \{H_{\mathbf{x}}^{\overline{\mathbf{x}}}, \mathbf{x} \in \mathcal{E} \text{ and } \overline{\mathbf{x}} \subset \mathbf{x}^c\}$ is called *H-system* if $H_{\mathbf{x}}^{\overline{\mathbf{x}}} \geqslant 0$ for all $\mathbf{x} \in \mathcal{E}$ and $\overline{\mathbf{x}} \subset \mathbf{x}^c$ and $H_{\emptyset}^{\overline{\mathbf{x}}} = 1$ for all $\overline{\mathbf{x}} \subset \mathbb{Z}^\nu$. This H-system is called *consistent* if it satisfies the following condition: for any $\mathbf{x}, \mathbf{y} \in \mathcal{E}$ such that $\mathbf{x} \cap \mathbf{y} = \emptyset$ and any $\overline{\mathbf{x}} \subset (\mathbf{x} \cup \mathbf{y})^c$ we have

$$H_{\mathbf{x} \cup \mathbf{y}}^{\overline{\mathbf{x}}} = H_{\mathbf{x}}^{\overline{\mathbf{x}}} H_{\mathbf{y}}^{\overline{\mathbf{x}} \cup \mathbf{x}}.$$

H-systems let one describe specifications in the following way.

THEOREM 5. *A system $\mathcal{Q} = \{Q_\Lambda^{\overline{\mathbf{x}}}, \Lambda \in \mathcal{E}$ and $\overline{\mathbf{x}} \subset \Lambda^c\}$ is a specification satisfying $\mathbf{Q}_\Lambda^{\overline{\mathbf{x}}}(\emptyset) > 0$ for all $\Lambda \in \mathcal{E}$ and $\overline{\mathbf{x}} \subset \Lambda^c$ if and only if there exists a consistent H-system $\mathcal{H} = \{H_{\mathbf{x}}^{\overline{\mathbf{x}}}, \mathbf{x} \in \mathcal{E}$ and $\overline{\mathbf{x}} \subset \mathbf{x}^c\}$ such that for any $\Lambda \in \mathcal{E}$ and $\overline{\mathbf{x}} \subset \Lambda^c$ we have*

$$\mathbf{Q}_\Lambda^{\overline{\mathbf{x}}}(\mathbf{x}) = \frac{H_{\mathbf{x}}^{\overline{\mathbf{x}}}}{\Sigma_{\mathbf{y} \subset \Lambda} H_{\mathbf{y}}^{\overline{\mathbf{x}}}}, \quad \mathbf{x} \subset \Lambda. \tag{3}$$

Now let us introduce the following

DEFINITION 6. A system $\mathbf{h} = \{h_t^{\overline{\mathbf{x}}}, t \in \mathbb{Z}^\nu \text{ and } \overline{\mathbf{x}} \subset \mathbb{Z}^\nu \backslash t\}$ is called *one-point system* if for all $t \in \mathbb{Z}^\nu$ and $\overline{\mathbf{x}} \subset \mathbb{Z}^\nu \backslash t$ we have $H_t^{\overline{\mathbf{x}}} \geqslant 0$ and for all $s, t \in \mathbb{Z}^\nu$ and $\overline{\mathbf{x}} \subset \mathbb{Z}^\nu \backslash \{s, t\}$ we have

$$h_s^{\overline{\mathbf{x}}} h_t^{\overline{\mathbf{x}} \cup s} = h_t^{\overline{\mathbf{x}}} h_s^{\overline{\mathbf{x}} \cup t}.$$

Here and in the sequel we write $\mathbb{Z}^\nu \backslash t$ as a shorthand notation for $\mathbb{Z}^\nu \backslash \{t\}$, $\overline{\mathbf{x}} \cup t$ as a shorthand notation for $\overline{\mathbf{x}} \cup \{t\}$ and in general we omit braces for one-point sets.

As the following theorem shows, these one-point systems correspond one-to-one to consistent H-systems. In fact they are nothing but one-point subsystems of consistent H-systems and hence, just like H-systems, describe specifications.

THEOREM 7. *A system $\mathcal{H} = \{H_{\mathbf{x}}^{\overline{\mathbf{x}}}, \mathbf{x} \in \mathcal{E} \text{ and } \overline{\mathbf{x}} \subset \mathbf{x}^c\}$ is a consistent H-system if and only if there exists a one-point system $\mathbf{h} = \{h_t^{\overline{\mathbf{x}}}, t \in \mathbb{Z}^\nu \text{ and } \overline{\mathbf{x}} \subset \mathbb{Z}^\nu \backslash t\}$ such that for all $\mathbf{x} \in \mathcal{E}$ and $\overline{\mathbf{x}} \subset \mathbf{x}^c$ we have*

$$H_{\mathbf{x}}^{\overline{\mathbf{x}}} = h_{t_1}^{\overline{\mathbf{x}}} h_{t_2}^{\overline{\mathbf{x}} \cup t_1} \cdots h_{t_n}^{\overline{\mathbf{x}} \cup t_1 \cup \cdots \cup t_{n-1}} \tag{4}$$

where $n = |\mathbf{x}|$ and $t_1, \ldots, t_n$ is some arbitrary enumeration of elements of the set $\mathbf{x}$. Particularly, for all $t \in \mathbb{Z}^\nu$ and $\overline{\mathbf{x}} \subset \mathbb{Z}^\nu \backslash t$ we have $H_t^{\overline{\mathbf{x}}} = h_t^{\overline{\mathbf{x}}}$.

So, a specifcation $\boldsymbol{\mathcal{Q}}$ satisfying $Q_\Lambda^{\overline{\mathbf{x}}}(\emptyset) > 0$ can be described by some one-point system $\mathbf{h}$. Let us note that such specifications are some-times called *vacuum specifications*. Note also that we include the Gibbsian case.

Clearly the quasilocality of $\mathbf{h}$ is equivalent to the quasilocality of corresponding specification $\boldsymbol{\mathcal{Q}}$.

Finally, let us turn us turn to the translation invariant case. Obviously a specification $\boldsymbol{\mathcal{Q}}$ is translation invariant if and only if the corresponding one-point system $\mathbf{h}$ is translation invariant, i.e., if we have $h_t^{\overline{\mathbf{x}}} = h_{t+s}^{\overline{\mathbf{x}}+s}$ for all $t, s \in \mathbb{Z}^\nu$ and $\overline{\mathbf{x}} \subset \mathbb{Z}^\nu \backslash t$. In this case, clearly one needs to know only the subsystem $\{h^{\overline{\mathbf{x}}}, \overline{\mathbf{x}} \subset \mathbb{Z}^\nu \backslash \mathbf{0}\}$, where $h^{\overline{\mathbf{x}}} = h_{\mathbf{0}}^{\overline{\mathbf{x}}}$ and $\mathbf{0}$ is the origin of $\mathbb{Z}^\nu$. This subsystem is the object of statistical interest of this work. Since it determines the whole one-point system we will use the same notation $\mathbf{h}$ for it. Condition of the quasilocality in this case will be written in the form

$$\gamma(I) = \sup_{\overline{\mathbf{x}} \subset \mathbb{Z}^\nu \backslash \mathbf{0}} |h^{\overline{\mathbf{x}}_I} - h^{\overline{\mathbf{x}}}| \underset{I \uparrow \mathbb{Z}^\nu}{\longrightarrow} 0.$$

3. The Nonparamteric Estimation Problem

In this section we present the statistical model, construct an estimator and state the main results of the paper: exponential and $\mathbf{L}^p$- consistencies of this estimator.

3.1. STATISTICAL MODEL

We write $\mathcal{H} = \{\mathbf{h}\colon\ \mathbf{h}$ is quasilocal and translation invariant$\}$. To any $\mathbf{h} \in \mathcal{H}$ we associate some specification $\boldsymbol{\mathcal{Q}}$ and hence some sets $\mathcal{G}(\mathbf{h}) = \mathcal{G}(\boldsymbol{\mathcal{Q}})$ and $\mathcal{G}_{\text{t.i.}}(\mathbf{h}) = \mathcal{G}_{\text{t.i.}}(\boldsymbol{\mathcal{Q}})$ of Gibbs random fields.

Suppose $\mathbf{h} \in \mathcal{H}$ is some unknown translation invariant quasilocal one-point system. We observe a realisation of some random field $\mathbf{P} \in \mathcal{G}(\mathbf{h})$ in the observation window Λ_n. Here and in the sequel Λ_n denotes the symmetric cube with the side size n centred at the origin $\mathbf{0}$ of $\mathbb{Z}^\nu$ (without loss of generality we assume that n is odd). So, based on the data $\mathbf{x}_n = \mathbf{x}_{\Lambda_n} \subset \Lambda_n$ generated by some random field $\mathbf{P} \in \mathcal{G}(\mathbf{h})$ we want to estimate $\mathbf{h}$. More formally, the statistical model is

$$\{\Omega, \mathcal{F}, \mathbf{P} \in \mathcal{G}(\mathbf{h}), \mathbf{h} \in \mathcal{H}^{\exp}_{A,B}\}$$

where $0 < A \leqslant B < \infty$ are some constants and $\mathcal{H}^{\exp}_{A,B}$ is the space of one-point systems satisfying the following conditions.

(C1) $\mathbf{h} \in \mathcal{H}$, i.e., $\mathbf{h}$ is quasilocal and translation invariant.
(C2) For all $\overline{\mathbf{x}} \subset \mathbb{Z}^\nu \backslash \mathbf{0}$ we have $A \leqslant h^{\overline{\mathbf{x}}} \leqslant B$.
(C3) Let ρ be the supremum norm on $\mathbb{Z}^\nu$ and put

$$\varphi(d) = \sup_{I\colon\ \rho(I^c \backslash \mathbf{0}, \mathbf{0}) \geqslant d} \ \sup_{\overline{\mathbf{x}} \subset \mathbb{Z}^\nu \backslash \mathbf{0}} |h^{\overline{\mathbf{x}}_I} - h^{\overline{\mathbf{x}}}|.$$

We call the function $\varphi(\cdot)$ *rate of quasilocality* and we suppose that $\varphi(d) \leqslant c\, e^{-a\, d^{\nu+\delta}}$ where c, a and δ are some positive constants.

Note that c, a and δ are not supposed to be known *a priori* and may differ for different $\mathbf{h} \in \mathcal{H}^{\exp}_{A,B}$.

Let us remark here that our statistical model is a bit unusual, in the sense that the probability measure $\mathbf{P}$ is not determined by the parameter $\mathbf{h}$. Rather, $\mathbf{h}$ determines some set $\mathcal{G}(\mathbf{h})$ of probability measures. The observations come from an arbitrary element of this set but we are not interested in this element, the only object of interest is the parameter $\mathbf{h}$ itself. That is, we want to identify the class $\mathcal{G}(\mathbf{h})$ corresponding to (unknown) one-point system $\mathbf{h}$, and not a particular element of this class. In fact, this is the reason for which our results hold regardless of non-uniqueness and translation invariance breaking. In some sense, if $|\mathcal{G}(\mathbf{h})| > 1$, then $\mathbf{P} \in \mathcal{G}(\mathbf{h})$ can be viewed as $\mathbf{P} = \mathbf{P}(\mathbf{h}, \mu)$, and only $\mathbf{h}$ is the parameter of interest (something like semiparametric statistical problem), while all our considerations will be performed on conditional distributions, the latter ones depending only on $\mathbf{h}$, and not on μ.

Any real-valued random function $\overline{\mathbf{h}}_n = \{\overline{h}_n^{\overline{\mathbf{x}}}, \overline{\mathbf{x}} \subset \mathbb{Z}^\nu \backslash \mathbf{0}\}$ constructed from $\mathbf{x}_n$ is said to be an *estimator* of $\overline{\mathbf{h}}$. The distance between the estimator $\overline{\mathbf{h}}_n$ and the true value $\mathbf{h}$ is measured in the supremum norm:

$$\|\overline{\mathbf{h}}_n - \mathbf{h}\| = \sup_{\overline{\mathbf{x}} \subset \mathbb{Z}^\nu \backslash \mathbf{0}} |\overline{h}_n^{\overline{\mathbf{x}}} - h^{\overline{\mathbf{x}}}|.$$

The estimator $\overline{\mathbf{h}}_n$ is said to be *consistent*, if for any $\mathbf{h} \in \mathcal{H}^{\exp}_{A,B}$ we have $\|\overline{\mathbf{h}}_n - \mathbf{h}\| \longrightarrow_{n\to\infty} 0$ in probability, uniformly over $p \in \mathcal{G}(\mathbf{h})$, i.e., if for any $\mathbf{h} \in \mathcal{H}^{\exp}_{A,B}$ and any $\varepsilon > 0$ we have

$$\sup_{\mathbf{P} \in \mathcal{G}(\mathbf{h})} \mathbf{P}(\|\overline{\mathbf{h}}_n - \mathbf{h}\| > \varepsilon) \underset{n\to\infty}{\longrightarrow} 0.$$

If the last relation holds uniformly on $\mathbf{h} \in \mathcal{H}^{\exp}_{A,B}$, then the estimator $\overline{\mathbf{h}}_n$ is said to be *uniformly consistent*.

The estimator $\overline{\mathbf{h}}_n$ is said to be $\mathbf{L}^p$*-consistent* for some $p \in (0, \infty)$, if for any $\mathbf{h} \in \mathcal{H}^{\exp}_{A,B}$ we have $\|\overline{\mathbf{h}}_n - \mathbf{h}\| \longrightarrow_{n\to\infty} 0$ in $\mathbf{L}^p$, uniformly over $\mathbf{P} \in \mathcal{G}(\mathbf{h})$, i.e., if for any $\mathbf{h} \in \mathcal{H}^{\exp}_{A,B}$ we have

$$\sup_{\mathbf{P} \in \mathcal{G}(\mathbf{h})} \mathbf{E}\|\overline{\mathbf{h}}_n - \mathbf{h}\|^p \underset{n\to\infty}{\longrightarrow} 0.$$

If the last relation holds uniformly on $\mathbf{h} \in \mathcal{H}^{\exp}_{A,B}$, then the estimator $\overline{\mathbf{h}}_n$ is said to be *uniformly* $\mathbf{L}^p$*-consistent*.

Let us finally note here that, if the random field corresponding to a one-point system $\mathbf{h}$ is unique, then the statistical model, the identifiability and all the notions of consistency regain their classical statistical sense.

3.2. CONSTRUCTION OF THE SIEVE ESTIMATOR

Let us first note that by (3) we have

$$\mathbf{Q}_t^{\overline{\mathbf{x}}}(t) = \frac{H_t^{\overline{\mathbf{x}}}}{\Sigma_{\mathbf{y} \subset t} H_{\mathbf{y}}^{\overline{\mathbf{x}}}} = \frac{H_t^{\overline{\mathbf{x}}}}{H_\emptyset^{\overline{\mathbf{x}}} + H_t^{\overline{\mathbf{x}}}} = \frac{h_t^{\overline{\mathbf{x}}}}{1 + h_t^{\overline{\mathbf{x}}}},$$

$$\mathbf{Q}_t^{\overline{\mathbf{x}}}(\emptyset) = \frac{H_\emptyset^{\overline{\mathbf{x}}}}{\Sigma_{\mathbf{y} \subset t} H_y^{\overline{\mathbf{x}}}} = \frac{H_\emptyset^{\overline{\mathbf{x}}}}{H_\emptyset^{\overline{\mathbf{x}}} + H_t^{\overline{\mathbf{x}}}} = \frac{1}{1 + h_t^{\overline{\mathbf{x}}}},$$

and hence

$$h^{\overline{\mathbf{x}}} = h_{\mathbf{0}}^{\overline{\mathbf{x}}} = \frac{\mathbf{Q}_{\mathbf{0}}^{\overline{\mathbf{x}}}(\mathbf{0})}{\mathbf{Q}_{\mathbf{0}}^{\overline{\mathbf{x}}}(\emptyset)} = \frac{\mathbf{Q}_{\mathbf{0}}^{\overline{\mathbf{x}}}(1)}{\mathbf{Q}_{\mathbf{0}}^{\overline{\mathbf{x}}}(0)}. \tag{5}$$

The main idea of the estimator is to take some $k = k(n)$ and approximate $h^{\overline{\mathbf{x}}}$ by the ratio of the conditional probabilities with condition in the volume Λ_k^*.

For this we use the formula (5) and we approximate the conditional probabilities $\mathbf{Q}_{\mathbf{0}}^{\overline{\mathbf{x}}}(x)$, $x \in \{0,1\}$, by the probabilities $\mathbf{P}_{\mathbf{0}|\Lambda_k^*}(x|\overline{\mathbf{x}}_{\Lambda_k^*})$ of observing x in $\mathbf{0}$ given that $\overline{\mathbf{x}}_{\Lambda_k^*}$ is observed on Λ_{k^*}. The volume Λ_k is called a *sieve* and $k = k(n)$ is called *sieve size* and is supposed to grow fast enough. In fact, using the total probability formula and the quasilocality condition, we have

$$\begin{aligned}\mathbf{P}_{\mathbf{0}|\Lambda_k^*}(x|\overline{\mathbf{x}}_{\Lambda_k^*}) &= \int_{\mathcal{X}^{\Lambda_k^c}} \mathbf{Q}_{\mathbf{0}}^{\overline{\mathbf{x}}_{\Lambda_k^*}\cup\overline{\mathbf{y}}}(x)\, \mathbf{P}_{\Lambda_k^c|\Lambda_k^*}(\mathrm{d}\overline{\mathbf{y}}|\overline{\mathbf{x}}_{\Lambda_k^*}) \\ &\approx \mathbf{Q}_{\mathbf{0}}^{\overline{\mathbf{x}}}(x) \int_{\mathcal{X}^{\Lambda_k^c}} \mathbf{P}_{\Lambda_k^c|\Lambda_k^*}(\mathrm{d}\overline{\mathbf{y}}|\overline{\mathbf{x}}_{\Lambda_k^*}) = \mathbf{Q}_{\mathbf{0}}^{\overline{\mathbf{x}}}(x).\end{aligned}$$

On the other hand, if k grows much slower than n, then the probabilities $\mathbf{P}_{\mathbf{0}|\Lambda_k^*}(x|\overline{\mathbf{x}}_{\Lambda_k^*})$ in their turn can be estimated by empirical conditional frequency of the value x observed in some point $t \in \Lambda_n$ given that $\overline{\mathbf{x}}_{\Lambda_k^*} + t$ is observed on the set $\Lambda_k^* + t$.

More precisely, let $\mathbf{x}(n)$ be the periodization on $\mathbb{Z}^\nu$ of the observation $\mathbf{x}_n$, that is, $(\mathbf{x}(n))_{\Lambda_n + nt} = \mathbf{x}_n + nt$ for all $t \in \mathbb{Z}^\nu$. Note that equivalently periodization can be viewed as wrapping the observation $\mathbf{x}_n$ on a torus. Now, for every $\overline{\mathbf{x}} \subset \mathbb{Z}^\nu \backslash \mathbf{0}$, let us put

$$A^1 = \{\mathbf{y} \subset \mathbb{Z}^\nu\colon\ \mathbf{y}_{\Lambda_k} = \overline{\mathbf{x}}_{\Lambda_k^*} \cup \mathbf{0}\} \quad \text{and} \quad A^0 = \{\mathbf{y} \subset \mathbb{Z}^\nu\colon\ \mathbf{y}_{\Lambda_k} = \overline{\mathbf{x}}_{\Lambda_k^*}\}.$$

Let us also put

$$N^1 = \sum_{t \in \Lambda_n} \mathbb{1}_{\{\mathbf{x}(n) - t \in A^1\}} \quad \text{and} \quad N^0 = \sum_{t \in \Lambda_m} \mathbb{1}_{\{\mathbf{x}(n) - t \in A^0\}}.$$

Clearly, N^1 and N^0 are the total numbers of subconfigurations of $\mathbf{x}_n$ of the 'form' Λ_k and equal to $\overline{\mathbf{x}}_{\Lambda_k^*} \cup \mathbf{0}$ and $\overline{\mathbf{x}}_{\Lambda_k^*}$, respectively.

Now we define the *sieve estimator* $\hat{h}_n$ by

$$\hat{h}_n^{\overline{x}} = \begin{cases} N^1/N^0, & \text{if } N^0 > 0 \text{ and } N^1 > 0, \\ A, & \text{if } N^1 = 0, \\ B, & \text{if } N^0 = 0 \text{ (and } N^1 > 0). \end{cases}$$

Note that the cases $N^0 = 0$ and $N^1 = 0$ are asymptotically not important. Moreover, we could have not considered at all the second case, that is, we could have put the estimator still to be $N^1/N^0 = 0$. Our definition of the estimator pursues rather practical aims, and is motivated by the following reasons: $N^0 = 0$ means that $\mathbf{Q}_{\mathbf{0}}^{\overline{\mathbf{x}}}(0) \approx 0$ and hence $h^{\overline{\mathbf{x}}}$ is 'large', while $N^1 = 0$ means that $\mathbf{Q}_{\mathbf{0}}^{\overline{\mathbf{x}}}(1) \approx 0$ and hence $h^{\overline{\mathbf{x}}}$ is 'small'; but we know *a priori* that $A \leqslant h^{\overline{\mathbf{x}}} \leqslant B$.

Let us note here, that the idea of using empirical conditional frequencies to construct estimators, as well as some results on consistency of estimators of such type for parametric models in the classical Gibbsian setup, can be found in [8, 11–14].

3.3. ASYMPTOTIC PROPERTIES OF THE ESTIMATOR

Note that the definition of the sieve estimator depends on the choice of k. Choosing k too large may results in insufficient number of repetitions of the subconfiguration $\overline{\mathbf{x}}_{\Lambda_k^*}$ in $\mathbf{x}_n$, i.e., one can too often have $N^0 = 0$ or $N^1 = 0$. On the other hand, choosing k too small may result in poor quality of the approximation $\mathbf{Q}_0^{\overline{\mathbf{x}}}(x) \approx \mathbf{P}_{\mathbf{0}|\Lambda_k^*}(x|\overline{\mathbf{x}}_{\Lambda_k^*})$. The following theorems show a 'good' choice of k. Let us denote

$$b^* = \max\{\ln(1+B), \ln(1+B) - \ln A\} \quad \text{and} \quad d^* = \nu/(2b^*).$$

Now we can formulate the following theorems.

THEOREM 8 (Exponential consistency of the sieve estimator). *Assume that* $\mathbf{h} \in \mathcal{H}_{A,B}^{\exp}$ *and* $\hat{\mathbf{h}}_n$ *is the sieve estimator with* $k = [(d \ln n)^{1/\nu}]$ *and* $d \in (0, d^*)$. *Then, for any* $\mathbf{h} \in \mathcal{H}_{A,B}^{\exp}$ *and any* $\varepsilon > 0$, *there exist some positive constant* $\alpha > 0$ *and some* $n_0 \in \mathbb{N}$ *such that*

$$\sup_{\mathbf{P} \in \mathcal{G}(\mathbf{h})} \mathbf{P}(\|\hat{\mathbf{h}}_n - \mathbf{h}\| > \varepsilon) \leqslant e^{-\alpha n^{\nu - 2db^*}/\ln n}$$

for all $n > n_0$, *i.e.., the estimator* $\hat{\mathbf{h}}_n$ *is exponentially consistent.*

THEOREM 9 ($\mathbf{L}^p$-consistency of the sieve estimator). *Assume that* $\mathbf{h} \in \mathcal{H}_{A,B}^{\exp}$ *and* $\hat{\mathbf{h}}_n$ *is the sieve estimator with* $k = [(d \ln n)^{1/\nu}]$ *and* $d \in (0, d^*)$, *and fix some* $p \in (0, \infty)$. *Then, for any* $\mathbf{h} \in \mathcal{H}_{A,B}^{\exp}$ *and for sufficiently large values of* n, *we have*

$$\sup_{\mathbf{P} \in \mathcal{G}(\mathbf{h})} (\mathbf{E}\|\hat{\mathbf{h}}_n - \mathbf{h}\|^p)^{1/p} \leqslant n^{-(\nu/2 - db^* - \sigma)}$$

where σ *is an arbitrary small positive constant,* i.e., *the estimator* $\hat{\mathbf{h}}_n$ *is* $\mathbf{L}^p$*-consistent.*

These theorems are the main results of this paper. The next section is devoted to their proof.

Let us finally note here, that the consistencies of the sieve estimator proved in the Theorems 8 and 9 can be trivially straightened to be uniform, if we consider a narrower class of one-point systems by fixing not only the constants A and B from the condition (**C2**), but also the constants a, c, and δ from the condition (**C3**), that is considering the class $\tilde{\mathcal{H}} = \tilde{\mathcal{H}}(A, B, a, c, \delta)$ defined by conditions (**C1**), (**C2**) and (**C3**) with some *a priori* fixed constants A, B, a, c and δ.

Note also that all the bounds on the rates of consistency obtained in the previous section are 'slowed' by the constant d from the definition of the sieve size k. Hence, one can consider getting rid of the terms containing d by slightly modifying the choice of the sieve size k. In fact, in the case of the space $\tilde{\mathcal{H}}$, by putting $k = [(\ln n)^{1/(\nu+\delta/2)}]$, one can get uniform exponential consistency with the rate

$$\sup_{\mathbf{h} \in \tilde{\mathcal{H}}} \sup_{\mathbf{P} \in \mathcal{G}(\mathbf{h})} \mathbf{P}(\|\hat{\mathbf{h}}_n - \mathbf{h}\| > \varepsilon) \leqslant e^{-\alpha\kappa(n)n^{\nu}}$$

where $\kappa(n)$ is some slowly varying function, and $\mathbf{L}^p$-consistency with the rate

$$\sup_{\mathbf{h}\in\widetilde{\mathcal{H}}}\sup_{\mathbf{P}\in\mathcal{G}(\mathbf{h})}(\mathbf{E}\|\hat{\mathbf{h}}_n-\mathbf{h}\|^p)^{1/p}\leqslant n^{-(\nu/2-\sigma)}.$$

The proofs are similar. The rates in this case are almost same as in the case of parametric estimation, when the unknown one-point system is supposed to be local. For more detailed discussion and for the parametric case see [2].

4. Proof of Theorems 8 and 9

Throughout the proof C, α and n_0 denote generic positive constants which can differ from formula to formula (and even in the same formula).

The main component of the proof of Theorem 8 is the so-called 'conditional mixing lemma'.

LEMMA 10 (Conditional mixing). *Let* $\mathbf{h}\in\mathcal{H}^{\exp}_{A,B}$, $\mathbf{P}\in\mathcal{G}(\mathbf{h})$ *and let* $\varphi(\cdot)$ *be the corresponding rate of quasilocality also let* $L=L(n)\in\mathbb{N}$ *and let the sets* $R_1=R_1(n),\dots,R_L=R_L(n)$ *be finite subsets of* $\mathbb{Z}^\nu$ *such that* $\rho(R_{\ell_1},R_{\ell_2})\geqslant\beta_n$ *for* $\ell_1\neq\ell_2$ *where* $\beta_n\longrightarrow_{n\to\infty}\infty$ *and*

$$\lim_{n\to\infty}\max_{1\leqslant\ell\leqslant L}|R_\ell|\,\varphi(\beta_n)=0.$$

Denote $\mathcal{R}=\mathbb{Z}^\nu\backslash(R_1\cup\cdots\cup R_L)$ *and let* $u_\ell\colon\ \mathcal{X}^{R_\ell}\to\mathbb{R}$, $\ell=1,\dots,L$, *be some bounded measurable functions. Then*

$$\mathbf{E}_{R_1\cup\cdots\cup R_L|\mathcal{R}}\left(\prod_{\ell=1}^{L}u_\ell(\mathbf{x}_{R_\ell})|\mathbf{x}_{\mathcal{R}}\right)$$

$$=\left(\prod_{\ell=1}^{L}\mathbf{E}_{R_\ell|\mathcal{R}}(u_\ell(\mathbf{x}_{\mathrm{R}_\ell})|\mathbf{x}_{\mathcal{R}})\right)(1+\delta_n)^L \tag{6}$$

where $\mathbf{E}_{R_\ell|\mathcal{R}}$ *is the expectation with respect to* $\mathbf{P}_{R_\ell|\mathcal{R}}$ *and*

$$\delta_n=O\left(\max_{1\leqslant\ell\leqslant L}|R_\ell|\varphi(\beta_n)\right). \tag{7}$$

Proof. First of all let us note that if $\mathbf{x}_t=\mathbf{y}_t$ for all t such that $\rho(t,\mathbf{0})\geqslant d$ then by (**C2**) and (**C3**) we have

$$\left|\ln\frac{h^{\mathbf{y}}}{h^{\mathbf{x}}}\right|=|\ln h^{\mathbf{y}}-\ln h^{\mathbf{x}}|\leqslant C|h^{\mathbf{y}}-h^{\mathbf{x}}|\leqslant C\varphi(d).$$

Now suppose $K_1=K_1(n)$, $K_2=K_2(n)$ and $K_3=K_3(n)$ form a disjoint decomposition of $\mathbb{Z}^\nu$ such that $K_1\in\mathcal{E}$ and $\rho(K_1,K_2)\geqslant\beta_n$. Then, using translation invariance and the formula (4), we get

$$\left|\ln\frac{H^{\mathbf{x}_{K_3}\cup\mathbf{x}'_{K_2}}_{\mathbf{x}_{K_1}}}{H^{\mathbf{x}_{K_3}\cup\mathbf{x}_{K_2}}_{\mathbf{x}_{K_1}}}\right|\leqslant C|\mathbf{x}_{K_1}|\varphi(\beta_n)\leqslant C|K_1|\varphi(\beta_n)$$

for all $\mathbf{x}, \mathbf{x}' \subset \mathbb{Z}^\nu$. If, moreover, $|K_1|\varphi(\beta_n) \longrightarrow_{n\to\infty} 0$, then clearly

$$\left| \frac{H_{\mathbf{x}_{K_1}}^{\mathbf{x}_{K_3}\cup\mathbf{x}'_{K_2}}}{H_{\mathbf{x}_{K_1}}^{\mathbf{x}_{K_3}\cup\mathbf{x}_{K_2}}} - 1 \right| = O(|K_1|\varphi(\beta_n)).$$

Now we can see that for all $\mathbf{x}, \mathbf{x}' \subset \mathbb{Z}^\nu$

$$\begin{aligned}
\frac{\mathbf{Q}_{K_1}^{\mathbf{x}_{K_3}\cup\mathbf{x}'_{K_2}}(\mathbf{x}_{K_1})}{Q_{K_1}^{\mathbf{x}_{K_3}\cup\mathbf{x}_{K_2}}(\mathbf{x}_{K_1})} &= \frac{H_{\mathbf{x}_{K_1}}^{\mathbf{x}_{K_3}\cup\mathbf{x}'_{K_2}} \Sigma_{S\subset K_1} H_S^{\mathbf{x}_{K_3}\cup\mathbf{x}_{K_2}}}{H_{\mathbf{x}_{K_1}}^{\mathbf{x}_{K_3}\cup\mathbf{x}_{K_2}} \Sigma_{J\subset K_1} H_J^{\mathbf{x}_{K_3}\cup\mathbf{x}'_{K_2}}} \\
&= \frac{H_{\mathbf{x}_{K_1}}^{\mathbf{x}_{K_3}\cup\mathbf{x}'_{K_2}}}{H_{\mathbf{x}_{K_1}}^{\mathbf{x}_{K_3}\cup\mathbf{x}_{K_2}}} \sum_{S\subset K_1} \frac{H_S^{\mathbf{x}_{K_3}\cup\mathbf{x}'_{K_2}}}{\Sigma_{J\subset K_1} H_J^{\mathbf{x}_{K_3}\cup\mathbf{x}'_{K_2}}} \frac{H_S^{\mathbf{x}_{K_3}\cup\mathbf{x}_{K_2}}}{H_S^{\mathbf{x}_{K_3}\cup\mathbf{x}'_{K_2}}} \\
&= \left[\left(\frac{H_{\mathbf{x}_{K_1}}^{\mathbf{x}_{K_3}\cup\mathbf{x}'_{K_2}}}{H_{\mathbf{x}_{K_1}}^{\mathbf{x}_{K_3}\cup\mathbf{x}_{K_2}}} - 1 \right) + 1 \right] \times \\
&\quad \times \sum_{S\subset K_1} \mathbf{Q}_{K_1}^{\mathbf{x}_{K_3}\cup\mathbf{x}'_{K_2}}(S) \left[\left(\frac{H_S^{\mathbf{x}_{K_3}\cup\mathbf{x}_{K_2}}}{H_S^{\mathbf{x}_{K_3}\cup\mathbf{x}'_{K_2}}} - 1 \right) + 1 \right] \\
&= \left(\frac{H_{\mathbf{x}_{K_1}}^{\mathbf{x}_{K_3}\cup\mathbf{x}'_{K_2}}}{H_{\mathbf{x}_{K_1}}^{\mathbf{x}_{K_3}\cup\mathbf{x}_{K_2}}} - 1 \right) \sum_{S\subset K_1} \mathbf{Q}_{K_1}^{\mathbf{x}_{K_3}\cup\mathbf{x}'_{K_2}}(S) \left(\frac{H_S^{\mathbf{x}_{K_3}\cup\mathbf{x}_{K_2}}}{H_S^{\mathbf{x}_{K_3}\cup\mathbf{x}'_{K_2}}} - 1 \right) + \\
&\quad + \left(\frac{H_{\mathbf{x}_{K_1}}^{\mathbf{x}_{K_3}\cup\mathbf{x}'_{K_2}}}{H_{\mathbf{x}_{K_1}}^{\mathbf{x}_{K_3}\cup\mathbf{x}_{K_2}}} - 1 \right) + \\
&\quad + \sum_{S\subset K_1} \mathbf{Q}_{K_1}^{\mathbf{x}_{K_3}\cup\mathbf{x}'_{K_2}}(S) \left(\frac{H_S^{\mathbf{x}_{K_3}\cup\mathbf{x}_{K_2}}}{H_S^{\mathbf{x}_{K_3}\cup\mathbf{x}'_{K_2}}} - 1 \right) + 1 \\
&= \Delta_n + 1 \qquad (8)
\end{aligned}$$

where $\Delta_n = O(|K_1|\varphi(\beta_n))$. Using the last formula and the total probability formula we get for all $\ell = 1, \dots, n$

$$\begin{aligned}
&\mathbf{P}_{R_\ell|\mathcal{R}\cup R_1\cup\cdots\cup R_{\ell-1}}(\mathbf{x}_{R_\ell}|\mathbf{x}_\mathcal{R} \cup \mathbf{x}_{R_1} \cup \cdots \cup \mathbf{x}_{R_{\ell-1}}) \\
&\quad = \mathbf{P}_{R_\ell|\mathcal{R}}(\mathbf{x}_{R_\ell}|\mathbf{x}_\mathcal{R})(1+\delta_n)
\end{aligned}$$

where δ_n satisfies (7). Multiplying this relations over $\ell = 1, \dots, n$ we get

$$\mathbf{P}_{R_1\cup\cdots\cup R_L|\mathcal{R}}(\mathbf{x}_{R_1} \cup \cdots \cup \mathbf{x}_{R_L}|\mathbf{x}_\mathcal{R}) = \left(\prod_{\ell=1}^{L} \mathbf{P}_{R_\ell|\mathcal{R}}(\mathbf{x}_{R_\ell}|\mathbf{x}_\mathcal{R}) \right) (1+\delta_n)^L$$

which implies (6). The lemma is proved.

The next lemma gives us a uniform lower bound for the conditional probabilities $\mathbf{Q}_\Lambda^{\overline{\mathbf{x}}}(\mathbf{x})$ and for the probabilities $\mathbf{P}_\Lambda(\mathbf{x})$.

LEMMA 11. *Let* $\mathbf{P} \in \mathcal{G}(\mathbf{h})$ *for some* $\mathbf{h}$ *satisfying the condition (***C2***). Then, uniformly on* $\mathbf{x} \subset \Lambda$ *and* $\overline{\mathbf{x}} \subset \Lambda^c$*, we have*

$$\mathbf{Q}_\Lambda^{\overline{\mathbf{x}}}(\mathbf{x}) \geqslant e^{-b^*|\Lambda|} \quad \text{and} \quad \mathbf{P}_\Lambda(\mathbf{x}) \geqslant e^{-b^*|\Lambda|}$$

where $b^* = \max\{\ln(1+B), \ln(1+B) - \ln A\}$.

Proof. The second assertion clearly follows from the first one using the total probability formula. By the same formula and properties of conditional distributions the first assertion clearly can be derived from the bound $\mathbf{Q}_{\mathbf{0}}^{\overline{\mathbf{x}}}(x) \geqslant e^{-b^*}$ for all $\overline{\mathbf{x}} \subset \mathbb{Z}^\nu \backslash \mathbf{0}$ and $x \in \{0, 1\}$. But by (**C2**) we have

$$\mathbf{Q}_{\mathbf{0}}^{\overline{\mathbf{x}}}(1) = \frac{h^{\overline{\mathbf{x}}}}{1+h^{\overline{\mathbf{x}}}} \geqslant \frac{A}{1+B} \quad \text{and} \quad \mathbf{Q}_{\mathbf{0}}^{\overline{\mathbf{x}}}(0) = \frac{1}{1+h^{\overline{\mathbf{x}}}} \geqslant \frac{1}{1+B}$$

and hence

$$\mathbf{Q}_{\mathbf{0}}^{\overline{\mathbf{x}}}(x) \geqslant \min\left\{\frac{A}{1+B}, \frac{1}{1+B}\right\} = e^{\min\{\ln A - \ln(1+B), -\ln(1+B)\}} = e^{-b^*}.$$

The lemma is proved.

In order to use the conditional mixing lemma, let us decompose Λ_n in the following way. For technical reasons suppose $n = 2\,m\,k$ for some $m \in \mathbb{N}$. Then Λ_n is partitioned into $m^\nu = n^\nu/(2k)^\nu$ cubes $D_1, \ldots, D_{m^\nu}$ with side $2k$. Each D_i contains $(2k)^\nu$ lattice sites. We order sites of each D_i in the same arbitrary way. Hence, every $t \in \Lambda_n$ can be referred to as a pair (i, j), $i = 1, \ldots, m^\nu$, $j = 1, \ldots, (2k)^\nu$, which means the j-th site in the cube D_i. In the sequel we will use both the notations t and (i, j) for points of Λ_n.

If we define

$$Y_{ij}^0 = \mathbb{1}_{\{\mathbf{x}(n)-(i,j)\in A^0\}} \quad \text{and} \quad Y_{ij}^1 = \mathbb{1}_{\{\mathbf{x}(n)-(i,j)\in A^1\}}$$

and

$$N_j^0 = \sum_{i=1}^{m^\nu} Y_{ij}^0 \quad \text{and} \quad N_j^1 = \sum_{i=1}^{m^\nu} Y_{ij}^1,$$

then N^0 and N^1 from the definition of the sieve estimator will have the form

$$N^0 = \sum_{j=1}^{(2k)^\nu} N_j^0 \quad \text{and} \quad N^1 = \sum_{j=1}^{(2k)^\nu} N_j^1.$$

Note that all Y_{ij}^0, Y_{ij}^1, N_j^0, N_j^1, N^0 and N^1 depend on n, on $\overline{\mathbf{x}}_{\Lambda_k^*}$ and on the observation $\mathbf{x}_n$.

Now, for any $\overline{\mathbf{x}} \subset \mathbb{Z}^\nu \backslash \mathbf{0}$, we can write

$$\begin{aligned}
|\hat{h}_n^{\overline{\mathbf{x}}} - h^{\overline{\mathbf{x}}}| &\leqslant |h^{\overline{\mathbf{x}}_{\Lambda_k^*}} - h^{\overline{\mathbf{x}}}| + |\hat{h}_n^{\overline{\mathbf{x}}} - h^{\mathbf{x}_{\Lambda_k^*}}| \\
&= |h^{\mathbf{x}_{\Lambda_k^*}} - h^{\overline{\mathbf{x}}}| + \mathbb{1}_{\{N^0=0 \text{ or } N^1=0\}} |\hat{h}_n^{\overline{\mathbf{x}}} - h^{\overline{\mathbf{x}}_{\Lambda_k^*}}| + \\
&\quad + \mathbb{1}_{\{N^0>0, N^1>0\}} \left| \frac{\sum_{j=1}^{(2k)^\nu} N_j^1}{N^0} - h^{\overline{\mathbf{x}}_{\Lambda_k^*}} \right| \\
&\leqslant |h^{\overline{\mathbf{x}}_{\Lambda_k^*}} - h^{\overline{\mathbf{x}}}| + \mathbb{1}_{\{N^0=0\}} |B - h^{\overline{\mathbf{x}}_{\Lambda_k^*}}| + \mathbb{1}_{\{N^1=0\}} |A - h^{\overline{\mathbf{x}}_{\Lambda_k^*}}| + \\
&\quad + \mathbb{1}_{\{N^0>0, N^1>0\}} \sum_{j=1}^{(2k)^\nu} \left| \frac{N_j^1}{N^0} - \frac{N_j^0}{N^0} h^{\overline{\mathbf{x}}_{\Lambda_k^*}} \right| \\
&= |h^{\overline{\mathbf{x}}_{\Lambda_k^*}} - h^{\overline{\mathbf{x}}}| + \mathbb{1}_{\{N^0=0\}} |B - h^{\overline{\mathbf{x}}_{\Lambda_k^*}}| + \mathbb{1}_{\{N^1=0\}} |A - h^{\overline{\mathbf{x}}_{\Lambda_k^*}}| + \\
&\quad + \sum_{j=1}^{(2k)^\nu} \mathbb{1}_{\{N^0>0, N^1>0, N_j^0=0\}} \frac{N_j^1}{N^0} + \\
&\quad + \sum_{j=1}^{(2k)^\nu} \mathbb{1}_{\{N_j^0>0, N^1>0\}} \frac{1}{N^0} |N_j^1 - N_j^0 h^{\mathbf{x}_{\Lambda_k^*}}| \\
&= D_n^1(\overline{\mathbf{x}}) + D_n^2(\overline{\mathbf{x}}) + D_n^3(\overline{\mathbf{x}}) + D_n^4(\overline{\mathbf{x}}) + D_n^5(\overline{\mathbf{x}}) \qquad (9)
\end{aligned}$$

with evident notations.

First of all, by (**C3**) we have

$$\|D_n^1(\cdot)\| = \sup_{\overline{\mathbf{x}} \subset \mathbb{Z}^\nu \backslash \mathbf{0}} |h^{\overline{\mathbf{x}}_{\Lambda_k^*}} - h^{\overline{\mathbf{x}}}| \leqslant \varphi(k) \leqslant c\, e^{-a\, k^{\nu+\delta}} \underset{n\to\infty}{\longrightarrow} 0$$

and hence

$$\mathbf{P}(\|D_n^1(\cdot)\| > \varepsilon/5) = 0 \qquad (10)$$

for $n \geqslant n_0$.

To estimate the remaining summands we need the following lemma.

LEMMA 12. *Denote* $\Gamma(n) = n^{-d\,b^*}$, $\lambda_n = \Gamma(n) m^\nu = n^{\nu - d\,b^*}/(2k)^\nu$ *and fix some* $r \in \{0, 1\}$. *Then, for any* $\varepsilon \in (0, 1)$, *there exist some positive constant* $\alpha > 0$ *and some* $n_0 \in \mathbb{N}$ *such that*

$$\mathbf{P}\left(\frac{N_j^r}{\lambda_n} < 1 - \varepsilon \right) \leqslant e^{-\alpha\, n^{\nu - 2\,d\,b^*} / \ln n},$$

uniformly on $n > n_0$, $j = 1, \ldots, (2k)^\nu$ *and* $\overline{\mathbf{x}}_{\Lambda_k^*} \subset \Lambda_k^*$.

Proof. For definitioness let us take $r = 0$. We denote by V_{ij} a cube with side k centred at (i, j), $i = 1, \ldots, m^\nu$, $j = 1, \ldots, (2k)^\nu$, and let $\mathcal{V}_j = \mathbb{Z}^\nu \backslash (V_{1j} \cup \cdots \cup$

$V_{m^\nu j}$). Note that Y_{ij}^0 depends only on the restriction of our periodized observation $\mathbf{x}(n)$ on the set V_{ij} and that for $i_1 \neq i_2$ we have $\rho(V_{i_1 j}, V_{i_2 j}) \geqslant 2k - k = k$. So, for any $\lambda > 0$, it follows from the conditional mixing lemma that

$$\mathbf{E}(e^{-\lambda N_j^0} | \mathbf{x}_{\mathcal{V}_j}) = (1+\delta_n)^{m^\nu} \prod_{i=1}^{m^\nu} \mathbf{E}(e^{-\lambda Y_{ij}^0} | \mathbf{x}_{\mathcal{V}_j}) \tag{11}$$

with $\delta_n = O(k^\nu \varphi(k)) = O(d \ln n \, c \, e^{-\alpha k^{\nu+\delta}}) = o(n^{-\beta})$ for all $\beta > 0$.

Clearly, using the Lemma 11, definition of Y_{ij}^0 and total probability formula, we have

$$\mathbf{E}(Y_{ij}^0 | \mathbf{x}_{\mathcal{V}_j}) \geqslant e^{-b^* |\Lambda_k|} \geqslant e^{-b^* d \ln n} = \Gamma(n).$$

Furthermore, using Taylor expansion formula, we get

$$\begin{aligned} \mathbf{E}(e^{-\lambda Y_{ij}^0} | \mathbf{x}_{\mathcal{V}_j}) &= e^{-\lambda \mathbf{E}(Y_{ij}^0 | \mathbf{x}_{\mathcal{V}_j})} \mathbf{E}(e^{-\lambda (Y_{ij}^0 - \mathbf{E}(Y_{ij}^0 | \mathbf{x}_{\mathcal{V}_j}))} | \mathbf{x}_{\mathcal{V}_j}) \\ &\leqslant e^{-\lambda \Gamma(n)} \left(1 + \frac{\lambda^2}{2} e^{\lambda}\right) \leqslant \exp\left[-\lambda \left(\Gamma(n) - \frac{\lambda}{2} e^{\lambda}\right)\right]. \end{aligned} \tag{12}$$

Finally, combining (11), (12), and using Chebychev's inequality and the total probability formula, for sufficiently large values of n we get

$$\begin{aligned} \mathbf{P}\left(\frac{N_j^0}{\lambda_n} < 1 - \varepsilon\right) &\leqslant e^{\lambda(1-\varepsilon)\lambda_n} \mathbf{E}\, e^{-\lambda N_j^0} \\ &\leqslant e^{\lambda(1-\varepsilon)\Gamma(n) m^\nu} \exp\left[-\lambda \left(\Gamma(n) - \frac{\lambda}{2} e^{\lambda}\right) m^\nu\right] (1+\delta_n)^{m^\nu} \\ &\leqslant C \exp\left[-\lambda m^\nu \left(\varepsilon \Gamma(n) - \frac{\lambda}{2} e^{\lambda}\right)\right]. \end{aligned}$$

Now, choosing $\lambda = \varepsilon \Gamma(n)/e = \varepsilon \, n^{-d\, b^*}/e < 1$, for sufficiently large values of n we get

$$\begin{aligned} \mathbf{P}\left(\frac{N_j^0}{\lambda_n} < 1 - \varepsilon\right) &\leqslant C \exp\left[-\frac{\varepsilon \, n^{-d\, b^*}}{e} \frac{n^\nu}{2^\nu \, d \, \ln n} \left(\varepsilon \, n^{-d\, b^*} - \frac{\varepsilon \, n^{-d\, b^*}}{2}\right)\right] \\ &\leqslant e^{-\alpha \, n^{\nu - 2\, d\, b^*} / \ln n} \end{aligned}$$

with an arbitrary $\alpha < \varepsilon^2/(2^{\nu+1} \, e \, d)$. The lemma is proved.

Using this lemma we clearly get

$$\mathbf{P}(N_j^r = 0) \leqslant \mathbf{P}\left(\frac{N_j^r}{\lambda_n} < 1 - \varepsilon\right) \leqslant e^{-\alpha \, n^{\nu - 2\, d\, b^*} / \ln n}$$

for all $j = 1, \dots, (2k)^\nu$, $r \in \{0, 1\}$ and for sufficiently large values of n. Therefore, for sufficiently large values of n, we have

$$\begin{aligned}\mathbf{P}(\|D_n^2(\cdot)\| > \varepsilon/5) &= \mathbf{P}\left(\sup_{\overline{\mathbf{x}} \subset \mathbb{Z}^\nu \setminus \mathbf{0}} |D_n^2(\overline{\mathbf{x}})| > \varepsilon/5\right) \\ &\leqslant \sum_{\overline{\mathbf{x}}_{\Lambda_k^*} \subset \Lambda_k^*} \mathbf{P}(N^0 = 0) \leqslant e^{-\alpha n^{\nu - 2db^*/\ln n}}, \end{aligned} \tag{13}$$

where we take into account that N^0 depends only on $\overline{\mathbf{x}}_{\Lambda_k^*}$, and hence the supremum over $\overline{\mathbf{x}} \subset \mathbb{Z}^\nu \setminus \mathbf{0}$ is in fact a maximum over $\overline{\mathbf{x}}_{\Lambda_k^*} \subset \Lambda_k^*$, i.e., a maximum over $2^{|\Lambda_k^*|} \leqslant 2^{d \ln n}$ elements.

In exactly the same way we have

$$\mathbf{P}(\|D_n^3(\cdot)\| > \varepsilon/5) \leqslant e^{-\alpha n^{\nu - 2db^*/\ln n}}, \tag{14}$$

and similarly we get

$$\begin{aligned}\mathbf{P}(\|D_n^4(\cdot)\| > \varepsilon/5) &= \mathbf{P}\left(\sup_{\overline{\mathbf{x}} \subset \mathbb{Z}^\nu \setminus \mathbf{0}} |D_n^4(\overline{\mathbf{x}})| > \varepsilon/5\right) \\ &\leqslant \sum_{\overline{\mathbf{x}}_{\Lambda_k^*} \subset \Lambda_k^*} \sum_{j=1}^{(2k)^\nu} \mathbf{P}(N_j^0 = 0) \leqslant e^{-\alpha n^{\nu - 2db^*/\ln n}}. \end{aligned} \tag{15}$$

Finally, the last summand is estimated by the following lemma.

LEMMA 13. *For any $\varepsilon \in (0, 1)$ there exist some positive constant $\alpha > 0$ and some $n_0 \in \mathbb{N}$ such that*

$$\mathbf{P}(\|D_n^5(\cdot)\| > \varepsilon/5) \leqslant e^{-\alpha n^{\nu - 2db^*/\ln n}} \tag{16}$$

for all $n > n_0$.

Proof. As before, it is sufficient to show that

$$\mathbf{P}\left(N_j^0 > 0, \frac{1}{N^0}|N_j^1 - N_j^0 h^{\overline{\mathbf{x}}_{\Lambda_k^*}}| > \frac{\varepsilon}{5(2k)^\nu}\right) \leqslant e^{-\alpha n^{\nu - 2db^*/\ln n}}.$$

We have obviously

$$\begin{aligned}&\mathbf{P}\left(N_j^0 > 0, \frac{1}{N^0}|N_j^1 - N_j^0 h^{\overline{\mathbf{x}}_{\Lambda_k^*}}| > \frac{\varepsilon}{5(2k)^\nu}\right) \\ &\leqslant \mathbf{P}\left(\left|\sum_{i=1}^{m^\nu}(Y_{ij}^1 - Y_{ij}^0 h^{\overline{\mathbf{x}}_{\Lambda_k^*}})\right| > \frac{\varepsilon N^0}{5(2k)^\nu}\right) \\ &\leqslant \mathbf{P}\left(\sum_{j=1}^{(2k)^\nu} \frac{N_j^0}{\lambda_n} \leqslant (1-\varepsilon)(2k)^\nu\right) + \mathbf{P}\left(\left|\sum_{i=1}^{m^\nu} W_{ij}\right| \geqslant \tau\lambda_n\right), \end{aligned}$$

where $\tau = \varepsilon(1-\varepsilon)/5$ and $W_{ij} = Y_{ij}^1 - Y_{ij}^0\, h^{\overline{\mathbf{x}}_{\Lambda_k^*}}$. The estimate of the first term easily follows from the preceding lemma. To estimate the second one let us at first note that using translation invariance, total probability formula and the formulas (2), (5) and (8) we have

$$\begin{aligned}
\mathbf{E}(Y_{ij}^0|\mathbf{x}_{\mathcal{V}_j})\, h^{\overline{\mathbf{x}}_{\Lambda_k^*}} &= \mathbf{P}_{\Lambda_k|\mathcal{V}_j-(i,j)}(\overline{\mathbf{x}}_{\Lambda_k^*}|\mathbf{x}_{\mathcal{V}_j} - (i,j))\, h^{\overline{\mathbf{x}}_{\Lambda_k^*}} \\
&= \mathbf{Q}_{\mathbf{0}}^{\overline{\mathbf{x}}_{\Lambda_k^*}\cup(\mathbf{x}_{\mathcal{V}_j}-(i,j))}(0)\, \mathbf{P}_{\Lambda_k^*|\mathcal{V}_j-(i,j)}(\overline{\mathbf{x}}_{\Lambda_k^*}|\mathbf{x}_{\mathcal{V}_j} - (i,j)) \times \\
&\quad \times \mathbf{Q}_{\mathbf{0}}^{\overline{\mathbf{x}}_{\Lambda_k^*}}(1)/\mathbf{Q}_{\mathbf{0}}^{\overline{\mathbf{x}}_{\Lambda_k^*}}(0) \\
&= (1+\rho_n)\, \mathbf{P}_{\Lambda_k^*|\mathcal{V}_j-(i,j)}(\overline{\mathbf{x}}_{\Lambda_k^*}|\mathbf{x}_{\mathcal{V}_j} - (i,j))\, \mathbf{Q}_{\mathbf{0}}^{\mathbf{x}_{\Lambda_k^*}}(1) \\
&= (1+\rho_n)^2\, \mathbf{P}_{\Lambda_k^*|\mathcal{V}_j-(i,j)}(\overline{\mathbf{x}}_{\Lambda_k^*}|\mathbf{x}_{\mathcal{V}_j} - (i,j)) \times \\
&\quad \times \mathbf{Q}_{\mathbf{0}}^{\overline{\mathbf{x}}_{\Lambda_k^*}\cup(\mathbf{x}_{\mathcal{V}_j}-(i,j))}(1) \\
&= \mathbf{E}(Y_{ij}^1|\mathbf{x}_{\mathcal{V}_j})(1+\rho_n),
\end{aligned}$$

where $\rho_n = O(\varphi(k)) = O(c\, e^{-\alpha\, k^{\nu+\delta}}) = o(n^{-\beta})$ for all $\beta > 0$.

The last equality clearly implies that

$$\mathbf{E}(W_{ij}|\mathbf{x}_{\mathcal{V}_j}) = \mathbf{E}(Y_{ij}^1|\mathbf{x}_{\mathcal{V}_j}) - \mathbf{E}(Y_{ij}^0|\mathbf{x}_{\mathcal{V}_j})\, h^{\overline{\mathbf{x}}_{\Lambda_k^*}} = O(\rho_n)$$

and hence, for any $\lambda > 0$, using the fact that $-B \leqslant W_{ij} \leqslant 1$ and Taylor expansion, we get

$$\begin{aligned}
\mathbf{E}(e^{\lambda\, W_{ij}^0}|\mathbf{x}_{\mathcal{V}_j}) &= e^{\lambda\, \mathbf{E}(W_{ij}^0|\mathbf{x}_{\mathcal{V}_j})}\, \mathbf{E}(e^{\lambda(W_{ij}^0 - \mathbf{E}(W_{ij}^0|\mathbf{x}_{\mathcal{V}_j}))}|\mathbf{x}_{\mathcal{V}_j}) \\
&\leqslant e^{\lambda O(\rho_n)}\left[1 + \frac{\lambda^2(B+1)^2}{2} e^{\lambda(B+1)}\right] \\
&\leqslant \exp\left[\lambda O(\rho_n) + \frac{\lambda^2(B+1)^2}{2} e^{\lambda(B+1)}\right].
\end{aligned}$$

Finally, using Chebychev's inequality, total probability formula and conditional mixing lemma, we get

$$\begin{aligned}
&\mathbf{P}\left(\sum_{i=1}^{m^\nu} W_{ij} \geqslant \tau\lambda_n\right) \\
&\quad \leqslant e^{-\lambda\tau\lambda_n}\, \mathbf{E}\, \exp\left(\lambda \sum_{i=1}^{m^\nu} W_{ij}\right) \\
&\quad \leqslant e^{-\lambda\tau\Gamma(n)m^\nu}\, \mathbf{E}\left(\prod_{i=1}^{m^\nu} \mathbf{E}(\lambda e^{W_{ij}}|\xi_{\mathcal{V}_j})\right)(1+\delta_n)^{m^\nu}
\end{aligned}$$

$$\leqslant C\, e^{-\lambda\tau\, n^{-d\,b^*}\, m^\nu} \left\{\exp\left[\lambda O(\rho_n) + \frac{\lambda^2 (B+1)^2}{2} e^{\lambda(B+1)}\right]\right\}^{m^\nu}$$

$$\leqslant C\ \exp\left\{-\lambda\, m^\nu \left[\tau n^{-d\,b^*} - \frac{(B+1)^2}{2}\lambda e^{\lambda(B+1)} - O(\rho_n)\right]\right\}.$$

Now, choosing $\lambda = \tau n^{-d\,b^*}/((B+1)^2\, e^{B+1}) < 1$, we get

$$\mathbf{P}\left(\sum_{i=1}^{m^\nu} W_{ij} \geqslant \tau\lambda_n\right) \ \leqslant\ C\ \exp\left[-\frac{\tau\, n^{-d\,b^*}}{(B+1)^2\, e^{B+1}} \frac{n^\nu}{2^\nu\, d\ \ln\, n} \frac{\tau n^{-d\,b^*}}{2}\right]$$

$$\leqslant\ e^{-\alpha\, n^{\nu-2\,d\,b^*}/\ln\, n}$$

with an arbitrary $\alpha < \tau^2/(2^{\nu+1}(B+1)^2\, e^{B+1}\, d)$.

By the same argument we have

$$\mathbf{P}\left(-\sum_{i=1}^{m^\nu} W_{ij} \geqslant \tau\lambda_n\right) \leqslant e^{-\alpha\, n^{\nu-2\,d\,b^*}/\ln\, n}$$

which concludes the proof of the lemma.

Now, combining (10), (13)–(16) and taing into account the inequality (9), we get the assertion of the Theorem 8. The uniformity on $\mathbf{P} \in \mathcal{G}(\mathbf{h})$ is trivial. The Theorem 8 is proved.

Let us note, that the details of the proof clearly give rise to explicit expression for the constant α. For example, if $\varepsilon \in (0, 1)$, then one can take an arbitrary

$$\alpha < \frac{\tau^2}{2^{\nu+1}(B+1)^2\, e^{B+1}\, d}.$$

Note, too, that taking a closer look on the proof we can give a 'more precise' bound on the rate of consistency, explicitly showing the dependence of the rate on ε. That is, for $\varepsilon \in (0, 1/2)$, we have the bound

$$\sup_{\mathbf{P}\in\mathcal{G}(\mathbf{h})} \mathbf{P}(\|\hat{\mathbf{h}}_n - \mathbf{h}\| > \varepsilon) \leqslant \mathbb{1}_{\{6cn^{-a\,d(d\,\ln\,n)^{\delta/\nu}} > \varepsilon\}} +$$

$$+ \psi_n \exp\{-\alpha\varepsilon^2\, n^{\nu-2\,d\,b^*}/\ln\, n + O(\rho_n)\beta\varepsilon n^{\nu-d\,b^*}/\ln\, n\}$$

where

$$\alpha = \frac{1}{25\cdot 2^{\nu+3}(B+1)^2\, e^{B+1}\, d}, \quad \beta = \frac{1}{5\cdot 2^{\nu+1}(B+1)^2\, e^{B+1}\, d}$$

and the sequence ψ_n is given by $\psi_n = 2^{d\,\ln\,n}(2^\nu\, d\ \ln\, n + 1)(2^\nu\, d\ \ln\, n + 2)$.

Considering $\varepsilon = \varepsilon_n = n^{-(\nu/2-\sigma)}$ with an arbitrary small positive constant σ and using the above bound we can easily prove Theorem 9. Indeed, we have

$$\begin{aligned}
\mathbf{E}\|\hat{\mathbf{h}}_n - \mathbf{h}\|^p &= \int_{\|\hat{\mathbf{h}}_n - \mathbf{h}\| > \varepsilon_n} \|\hat{\mathbf{h}}_n - \mathbf{h}\|^p \, \mathrm{d}\,\mathbf{P} + \int_{\|\hat{\mathbf{h}}_n - \mathbf{h}\| \leqslant \varepsilon_n} \|\hat{\mathbf{h}}_n - \mathbf{h}\|^p \, \mathrm{d}\,\mathbf{P} \\
&\leqslant (\max\{n^\nu, B\} + B)^p \, \mathbf{P}(\|\hat{\mathbf{h}}_n - \mathbf{h}\| > \varepsilon_n) + \varepsilon_n^p \\
&\leqslant C\, n^{-(\nu/2-\sigma)p}
\end{aligned}$$

for sufficiently large values of n, where we use the fact that $\mathbf{h}$ is bounded by B and $\hat{\mathbf{h}}$ by $\max\{n^\nu, B\}$. the assertion of Theorem 9 follows trivially.

Acknowledgements

The author is grateful to F. Comets and X. Guyon, whose remarks and suggestions have helped to improve this work. Special thanks should be addressed to B. S. Nahapetian whose fruitful advice and ideas have made this work possible. Finally, the author would like to express his gratitude to Yu. A. Kutoyants for helpful discussions.

References

1. Comets, F.: 'On consistency of a class of estimators for exponential families of Markov random fields on the lattice', *Ann. Stat.* **20** (1992), 455–468.
2. Dachian, S.: 'Une approche vers la description et l'identification d'une classe de champs aléatoires', Thése de Doctorat de l'Université Paris 6, January 1999. (http://www.univ-lemans.fr/sciences/statist/download/serge).
3. Dachian, S. and Nahapetian, B. S.: 'Inclusion-exclusion description of random fields', *J. Contemp. Math. Anal. Armen. Acad. Sci.* **30**, No. 6, (1995), 50–61.
4. Dachian, S. and Nahapetian, B. S.: 'An approach towards description of random fields', Université du Maine, Preprint No. 98–5, May 1998.
5. Dobrushin, R. L.: 'The description of a random field by means of conditional probabilities and conditions of its regularity', *Theory Probab. Appl.* **13** (1968), 197–224.
6. Dobrushin, R. L.: 'Gibbsian random fields for lattice systems with pairwise interaction', *Funct. Anal. Appl.* **2** (1968), 292–301.
7. Dobrushin, R. L.: 'The problem of uniqueness of a Gibbs random field and the problem of phase transitions', *Funct. Anal. Appl.* **2** (1968), 302–312.
8. Geman, S. and Graffigne, C.: 'Markov random fields image models and their applications to computer vision', in: A. M. Gleason (Ed.), *Proc. Int. Cong. Math. 1986*, American Math. Society. NY, 1987.
9. Georgii, H.-O.: *Gibbs Measures and Phase Transitions*, Berlin: de Gruyter, 1988.
10. Grenander, U.: *Abstract Inference*, New York: Wiley, 1981.
11. Guyon, X.: *Random Fields on a Network. Modelling, Statistics and Applications*, New York: Springer-Verlag, 1995.
12. Ji, C.: 'Sieve estimators for pair-interaction potentials and local characteristics in Gibbs random fields', University of North Carolina, Preprint No. 2037, October 1990.

13. Possolo, A.: 'Estimation of binary Markov random fields', *Tech. Report No.* 77, Dept. of Stat., Univ. of Washington, 1986.
14. Possolo, A.: 'Subsampling a random field', in: A. Possolo (Ed.), *Spat. Stat. and Imaging*, IMS, L. N.-M. S., 1991.

Biostatistics (2012), **13**, 2, *pp.* 241–255
doi:10.1093/biostatistics/kxr043
Advance Access publication on November 30, 2011

Classification method for disease risk mapping based on discrete hidden Markov random fields

MYRIAM CHARRAS-GARRIDO*, DAVID ABRIAL, JOCELYN DE GOËR
Unité d'Epidémiologie Animale, UR346, Centre INRA de Clermont-Ferrand-Theix, 63122 Saint-Genès-Champanelle, France
myriam.charras-garrido@clermont.intra.fr

SERGUEÏ DACHIAN
Laboratoire de Mathématiques, Université Clermont-Ferrand 2, 24 avenue des Landais, 63177 Aubière, France

NATHALIE PEYRARD
Unité de Biométrie et Intelligence Artificielle, UR875, Centre INRA de Toulouse Chemin de Borde Rouge, 31326 Castanet Tolosan, France

SUMMARY

Risk mapping in epidemiology enables areas with a low or high risk of disease contamination to be localized and provides a measure of risk differences between these regions. Risk mapping models for pooled data currently used by epidemiologists focus on the estimated risk for each geographical unit. They are based on a Poisson log-linear mixed model with a latent intrinsic continuous hidden Markov random field (HMRF) generally corresponding to a Gaussian autoregressive spatial smoothing. Risk classification, which is necessary to draw clearly delimited risk zones (in which protection measures may be applied), generally must be performed separately. We propose a method for direct classified risk mapping based on a Poisson log-linear mixed model with a latent discrete HMRF. The discrete hidden field (HF) corresponds to the assignment of each spatial unit to a risk class. The risk values attached to the classes are parameters and are estimated. When mapping risk using HMRFs, the conditional distribution of the observed field is modeled with a Poisson rather than a Gaussian distribution as in image segmentation. Moreover, abrupt changes in risk levels are rare in disease maps. The spatial hidden model should favor smoothed out risks, but conventional discrete Markov random fields (e.g. the Potts model) do not impose this. We therefore propose new potential functions for the HF that take into account class ordering. We use a Monte Carlo version of the expectation–maximization algorithm to estimate parameters and determine risk classes. We illustrate the method's behavior on simulated and real data sets. Our method appears particularly well adapted to localize high-risk regions and estimate the corresponding risk levels.

Keywords: Animal epidemiology; Disease mapping; Expectation-maximization (EM) algorithm; Image segmentation.

*To whom correspondence should be addressed.

1. INTRODUCTION

Efficient disease control requires an understanding of the determinants and dynamics of the disease in question. In situations where little initially is known, the first questions to ask are: Where are the populations at high risk located? Are these locations structured in space? If so, how? Can they be linked to environmental factors? Disease mapping models of epidemiological risk provide tools to localize high-risk areas and identify potential sources of a disease. A comparison of disease maps with the spatial distribution of factors suspected of causing the disease can help to identify which are most significant.

Epidemiological data are frequently count data aggregated at the level of spatial units (e.g. administrative zones): For each unit, the available information is the number of observed cases and the population size. The risk in a given spatial unit is the probability that an arbitrary individual in the population of the unit is contaminated. Classical risk mapping methods are based on spatial models. Spatial correlations account for spatial structure in the unknown or unobserved factors affecting the risk level and/or for spatial transmission of infectious diseases. These methods are inspired by statistical methods for image restoration or denoising (see, e.g. Li, 2001). The mathematical framework is that of hidden Markov random fields (HMRF). The "hidden image" to be restored is the risk value at each spatial location, and the observed image is the set of counts of observed cases. The numbers of cases usually are modeled by Poisson distributions. Most statistical methods for risk mapping of aggregated data (see, e.g. Mollié, 1999; Pascutto *and others*, 2000) dedicated to noncontagious diseases are based on a Poisson log-linear mixed model initially proposed by Besag *and others* (1991). This model is based on a hierarchical Bayesian approach where the latent intrinsic risk field (parameter of the Poisson distribution) is represented by a Markov random field (MRF) with continuous state space modeled by a Gaussian autoregressive spatial smoothing. Recent developments in risk mapping concern spatiotemporal mapping (see Knorr-Held and Richardson, 2003; Robertson *and others*, 2010) and multivariate disease mapping (see Knorr-Held *and others*, 2002; MacNab, 2010). These procedures produce a precise real-valued estimation of the different risks in each spatial unit.

Most applications of disease mapping involve real-valued disease risk maps. However, in some cases, such as animal epidemiology (see, e.g. Abrial *and others*, 2005), a coarser spatial representation of risk is needed in which locations with similar risk values are grouped. Such a representation with clearly delimited areas at risk can help decision makers interpret the risk structure and is important to determine protection measures such as culling, movement restriction, mass vaccination, etc. These areas at risk can be viewed as clusters as in Knorr-Held and Rasser (2000), but we prefer to interpret them as classes of risk, as in Green and Richardson (2002) or Alfo *and others* (2009), because geographically separated areas can have similar risks and be grouped in the same class. There consequently are fewer classes than clusters, facilitating interpretation by decision makers. Until present, an additional postprocessing step is generally conducted to define the risk classes either manually (involving the difficult definition of the risk range of each class) or using automated statistical classification methods (e.g. Fraley and Raftery, 2007). In either case, the classification step, which is of major interest in animal epidemiology, is not part of the initial risk mapping procedure and is performed separately. In 2 recent papers, risk classification is part of a single procedure: Green and Richardson (2002) present a model that is based on hierarchical Bayesian approaches with the latent risk field modeled by a Gaussian autoregressive spatialsmoothing, while Alfo *and others* (2009) present a method based on discrete HMRF models estimated by an EM-type algorithm using mode field approximation.

In this article, we propose another approach to risk modeling that integrates an automatic and unsupervised classification of locations into a few risk classes. This method relies on a discrete HMRF model in which spatial correlations are embedded in the map of classes. A representative risk associated with each class is estimated during the procedure. Many biologists are unfamiliar with the Bayesian context used

for estimation in conventional risk mapping methods, such as Besag *and others* (1991), Mollié (1999), or Pascutto *and others* (2000), and they hesitate to use it. They are most interested in a practical interpretation of each parameter of the model, including hyperprior distributions and hyperparameters. To limit the hyperprior stacking up that occurs in hierarchical Bayesian approaches, we chose to investigate an alternative estimation of the different parameters by maximum likelihood (ML) using an EM-type algorithm (see McLachlan and Krishnan, 2008). More generally, we aimed to produce a model that can be interpreted easily by epidemiologists and designed easily based on the context. In particular, the spatial correlation is controlled by the potential function of order 2 that is quite simple, interpretable, and flexible. At the end of the estimation procedure, the classification is directly estimated from the observed count data, without any intermediate estimation of individual risks at each location, by computation of the iterated conditional modes (ICM) estimator (see Besag, 1986). The output is a disease map where the color of each spatial unit represents the associated risk class. We present our model for classified disease mapping from count data in Section 2. The estimation classification method is described in Section 3. In Section 4, we illustrate the performance of our method first on simulated data and then to produce a spatial classification of the risk of bovine spongiform encephalopathy (BSE) in France.

2. The Disease Mapping Model

We present here our discrete HMRF for the mapping of disease risk classes. The numbers of observed cases y_i of disease in the spatial units $i = 1, \ldots, n$ form the observed field $\mathbf{y} = (y_i)_{i=1,\ldots,n}$ and are associated to the random field $\mathbf{Y} = (Y_i)_{i=1,\ldots,n}$. We represent the unknown risk classes by the random field $\mathbf{X} = (X_i)_{i=1,\ldots,n}$, usually referred to as the hidden field (HF). We will see that the spatial correlation characterizing disease maps is fully embedded in $\mathbf{X}$.

2.1 *The observed field*

In the HMRF framework, the observed field $\mathbf{Y}$ is linked to the HF $\mathbf{X}$ assuming conditional independence given a realization $\mathbf{x}$ (no spatial correlation appears at this stage):

$$P(\mathbf{Y} = \mathbf{y}|\mathbf{x}, \theta) = \prod_{i=1}^{n} P(Y_i = y_i|x_i, \theta) = \exp\left[\sum_{i=1}^{n} \log P(Y_i = y_i|x_i, \theta)\right],$$

with, for a discrete HMRF, $\theta = (\theta_k)_{k=1,\ldots,K}$. In risk mapping, θ_k is a representative risk associated with class k ($k = 1, \ldots, K$), e.g. the absolute risk that corresponds to a probability of infection. For a rare and noncontagious disease, the case we consider, the distribution of Y_i is usually modeled by a Poisson distribution with expectation equal to $n_i\theta_{x_i}$ (the mean number of cases in unit i), with n_i denoting the population size: $Y_i|x_i \sim \mathcal{P}(n_i\theta_{x_i})$, with $P(Y_i = y_i|x_i, \theta) = \exp(-n_i\theta_{x_i})(n_i\theta_{x_i})^{y_i}/y_i!$, for $x_i = 1, \ldots, K$ and $i = 1, \ldots, n$. This discrete distribution of the Y_is is the first main difference with the discrete HMRF models commonly used in image segmentation where a Gaussian distribution often is considered.

In practice, epidemiologists usually prefer to study the relative risk r_{x_i}, corresponding to the ratio between a local and an overall risk rather than θ_{x_i}. It allows an easier comparison of different risk maps as the scale of the r_{x_i}s is always the same; in contrast, the range of θ_{x_i} can vary considerably between 2 data sets. For a unique population (without any structure), these 2 risks are equivalent.

2.2 *The HF*

We now will describe the modeling of the HF $\mathbf{X}$, which involves the spatial correlation characterizing disease maps. We consider an MRF (see, e.g. Li, 2001) with potential functions of order 1 and 2.

The distribution of $\mathbf{X}$ then is expressed as

$$P(\mathbf{X} = \mathbf{x}|\alpha, \beta) = \gamma \exp\left(-\sum_{i=1}^{n} \varphi_1(x_i|\alpha) + \beta \sum_{i=1}^{n} \sum_{j \in V_i} \varphi_2(x_i, x_j)\right), \tag{2.1}$$

where V_i is the set of neighbors of i. In our illustrations (Section 4), the territory of France is divided into 1264 hexagons and adjacent hexagons are defined as neighbors.

The use of a potential function of order 1, φ_1, parameterized by a parameter α, is one way to tune the proportions of the different risk classes in $\mathbf{X}$. The term φ_2 accounts for spatial correlation and the parameter β fixes the balance between these terms. In other words, it controls the importance of the spatial smoothing. In MRF models, only potential functions of order 1 and 2 usually are considered to facilitate computation and the interpretation of these functions. The same is true in spatial epidemiology, where the joint influence of 2 (or more) neighbors on a geographical unit seems much more difficult to interpret than the influence of a single neighbor. Since we observed good results in our experiments with only φ_1 and φ_2, we did not consider larger orders.

We compared 2 choices for φ_1. First, we considered a potential function with no a priori, $\varphi_1(k|\alpha) = 0$, for $k = 1, \ldots, K$, so that only the data and the spatial structure of the classes guide the final distribution of classes. We then considered the general case $\varphi_1(k|\alpha) = \alpha_k$, with $\alpha_1 = 0$ as the potential functions are defined up to a constant.

We aim to represent situations with smooth variations of risk in space (as can be expected from epidemiological data) through the use of the spatial correlation term φ_2. It is highly unlikely that a very low-risk area would be observed immediately next to a very high-risk area. Extreme risk areas logically would be separated by a smooth gradation of risks. This means that $\varphi_2(x_i, x_j)$ should not only favor configurations where neighboring locations are in the same class, as with the classical Potts model (Wu, 1982) used in Alfo *and others* (2009) which penalizes equally all pair configurations where $x_i \neq x_j$. φ_2 also should decrease with the distance between x_i and x_j.

We propose 2 expressions of φ_2 that take these specific features of risk mapping into account. First, we assume an ordering of the classes in the sense that $\theta_k < \theta_{k+1}$. Then φ_2 should penalize pairs of classes k and k' at neighboring sites according to their distance: The closer the 2 classes, the higher the probability that this configuration would be observed. We propose the 2 following φ_2 based on 2 distances between the classes of neighboring units that take into account a gradual correlation. The first, referred to as grad1, is based on the absolute distance: $\varphi_2(x_i, x_j) = 1 - |x_i - x_j|/(K-1)$. The second, referred to as grad2, is based on the squared distance: $\varphi_2(x_i, x_j) = 1 - (x_i - x_j)^2/(K-1)^2$. While the correlation here only depends on the distance between the risk classes of 2 adjacent units, more complex forms could be possible in other contexts. For example, asymmetric φ_2 can be introduced to model a dissemination influenced by an ecological gradient; and φ_2 can differ in different regions of the map according to geographical barriers. φ_2 is easy to interpret and to construct according to the intended application and is one of the main advantages of our approach.

3. MODEL ESTIMATION

For an epidemiologist, the first output of interest is the risk map, i.e. the values of the risk classes x_i, followed by the values of the risk θ_{x_i} (or r_{x_i}) associated with the x_is. The classification procedure that determines the risk classes is detailed in Section 3.2. To obtain the x_is, we must estimate the different parameters of the model: The risks θ_k (or r_k), the classes proportions α_k ($k = 1, \ldots, K$), and the smoothing strength β. The estimation procedure is presented in Section 3.1 below. As mentioned previously, we chose to apply an ML procedure through an EM-type algorithm instead of a Bayesian procedure to avoid hyperpriors that are difficult for epidemiologists to interpret.

3.1 *Parameter estimation*

Without loss of generality, we present here the method for the parameterization of the model with the absolute risks θ_k: (θ, α, β). The formulas remain valid if relative risks r_k are used in the place of θ_k. The model parameters are estimated with missing data (the HF $\mathbf{X}$). To evaluate the ML estimator, we therefore used the EM algorithm (Dempster *and others*, 1977). However, the complexity of the HMRF model renders impossible the computation of the expectation of the complete log-likelihood in the expectation step of the EM algorithm. To overcome this problem, we applied the Monte Carlo EM (MCEM) solution proposed by Wei and Tanner (1990) which relies on a Monte Carlo method to generate realizations of the HF $\mathbf{X}$. In the expectation step, we also need to compute the *a priori* distribution of the HF, $P(\mathbf{X} = \mathbf{x}|\alpha, \beta)$. The computation of this probability is too complex to be obtained directly. To overcome this problem, we used the notion of pseudolikelihood proposed by Besag (1974) which consists of an approximation of the probability of the MRF $\mathbf{X}$ by a product of local conditional probabilities of the X_is given their neighborhood: $P(\mathbf{X} = \mathbf{x}|\alpha, \beta) \approx \prod_{i=1}^{n} P(X_i = x_i|\mathbf{x}_{V_i}, \alpha, \beta)$, where $\mathbf{x}_{V_i} = \{x_j, j \in V_i\}$ denotes the value of neighboring units. Computing this approximated probability is simple and fast. The iteration $q + 1$ of the MCEM algorithm with the use of pseudolikelihood is the following:

Monte Carlo expectation step. Generate T_{q+1} realizations $\mathbf{x}^{(q+1),(1)}, \ldots, \mathbf{x}^{(q+1),(T_{q+1})}$ of the HF according to the conditional probability distribution of the missing data $P(\mathbf{X} = \mathbf{x}|\mathbf{y}, \theta^{(q)}, \alpha^{(q)}, \beta^{(q)})$ using one iteration of the Gibbs sampling procedure. The expectation of the complete data log-likelihood then is approximated by

$$\hat{Q}_{T_{q+1}}(\theta, \alpha, \beta|\theta^{(q)}, \alpha^{(q)}, \beta^{(q)})$$

$$= \frac{1}{T_{q+1}} \sum_{t=1}^{T_{q+1}} \sum_{i=1}^{n} [\log P(Y_i = y_i|x_i^{(q+1),(t)}, \theta) + \log P(X_i = x_i^{(q+1),(t)}|\mathbf{x}_{V_i}^{(q+1),(t)}, \alpha, \beta)].$$

Maximization step. Update the parameters to $(\theta^{(q+1)}, \alpha^{(q+1)}, \beta^{(q+1)})$ by maximizing the function $\hat{Q}_{T_{q+1}}(\theta, \alpha, \beta|\theta^{(q)}, \alpha^{(q)}, \beta^{(q)})$ according to (θ, α, β).

Remark that the θ_k may be ordered, but this constraint is not included in this estimation procedure. We only specified ordered initial values of the risks θ_k^0 as the design of the potential function φ_2 that favors ordered situations generally suffices to maintain this order of the risks during the estimation algorithm. A more complete description of the estimation procedure is presented in Section A of the supplementary material available at *Biostatistics* online.

3.2 *Classification*

The EM algorithm (and the Monte Carlo version we used) provides an estimation of the model parameters but does not assign a class to each site. Conditional probabilities of each class k at each site i are computed during the EM procedure. These are used to estimate the x_is (in order to classify the different geographical units) using the maximum a posteriori (MAP) rule. The MAP estimate is the realization of the HF with the highest probability conditional to the observed data. Unfortunately, the MAP estimate cannot be computed directly for a Gibbs distribution as it involves the computation of the conditional distribution of the hidden variables for all possible hidden maps. We chose to use the ICM algorithm that can be considered (Besag, 1986) as an approximation and an improvement of the MAP. During ICM, the class of each geographical unit i is iteratively fixed to the mode of the conditional distribution of X_i knowing the observed data y_i

and the current value of the neighbors $\mathbf{x}_{V_i}$, i.e. solving:

$$x_i = \arg \max_{k=1,\ldots,K} P(X_i = k | x_1, \ldots, x_{i-1}, x_{i+1}^s, \ldots, x_n^s, \mathbf{y}, \hat{\theta}, \hat{\alpha}, \hat{\beta})$$

where $\hat{\theta}$, $\hat{\alpha}$, and $\hat{\beta}$ are estimations of the model parameters (we use an average of the latest values of the MCEM algorithm), and $\mathbf{x}^s$ is the initialization of the ICM procedure (the last simulated map during the MCEM algorithm). ICM can be iterated but, in practice, we observe that only one application generally is sufficient.

4. ILLUSTRATIONS

We first apply our method on a few examples to understand the behavior of the model and to compare it with the method currently used by practitioners, e.g. the model of Besag *and others* (1991) with posterior classification. We then present the performance of our method on intensive simulations. Finally, we apply the model to a real data set concerning BSE in France.

4.1 *Illustration on simulated representative data sets*

We simulated different data sets using the cattle population in France (see Figure 3(a)) as n_is and 3 risk maps we built manually. In the first case (Figure 1(a)), we defined 3 large risk regions for a rare disease with $\theta_1 = 10^{-5}$, $\theta_2 = 10^{-4}$, and $\theta_3 = 10^{-3}$. In the second case (Figure 1(b)), we defined 5 smaller risk regions, with $\theta_1 = 10^{-5}$, $\theta_2 = 5 \times 10^{-5}$, $\theta_3 = 10^{-4}$, $\theta_4 = 5 \times 10^{-4}$, and $\theta_5 = 10^{-3}$. In the third case (Figure 1(c)), we drew a continuous North–South (NS) gradient for the risks based on the hexagon centroids going from 10^{-5} in the South to 10^{-3} in the North.

Using the population n_i, the true classes x_i and the associated risk values θ_k, we simulated numbers of cases y_i from Poisson distributions $\mathcal{P}(n_i \theta_{x_i})$. Figures 1(d–f) present examples of observed number of cases (y_i maps) obtained for the 3-classes, 5-classes, and NS gradient risk maps. We first applied our procedure to these data sets: A discrete HMRF model with φ_2 grad2 estimated with MCEM algorithm. We applied the 2 forms of φ_1: The complete procedure estimating both α and β, and the simplest case of $\alpha = 0$ estimating only the smoothing parameter β. We then compared our model to one currently applied by practitioners: The model of Besag *and others* (1991), a Gaussian conditional autoregressive (CAR) smoothing with a single local spatial effect and no global effect (for details, see Section B of the supplementary material available at *Biostatistics* online). We obtained worse results when using also a global effect. To classify this continuous estimation of the risks, we applied the clustering procedure based on a Bayesian regularization for Gaussian mixture (Fraley and Raftery, 2007). For the 3-classes simulation, Figures 2(a) and (b) show that our classification procedure performs well estimating almost exactly the outline of the high risk region. The 3 estimated risks are also very close to the true values. The low risk region is not as well retrieved. In particular, when estimating both α and β (Figure 2(b)), there is an important region in the South-East (SE) that is classified incorrectly in the low-risk class. However, this region has a very low population density and, despite the medium risk level, no disease cases have been observed there. This important lack of information explains the misclassification of the region. The CAR model presented in Figure 2(c) detects that the North-East has a higher risk than the Centre and the South-West (SW). However, the real pattern is not retrieved and the highest risks are estimated in regions with very small populations. In addition, all the estimated risks are overestimated.

For the 5-classes simulation, Figures 2(d) and (e) show that our procedure performs well, roughly retrieving the true pattern and estimating quite well the outlines of the high-risk areas. In comparison, the general pattern and the outlines of these regions are retrieved more approximatively by the CAR model as

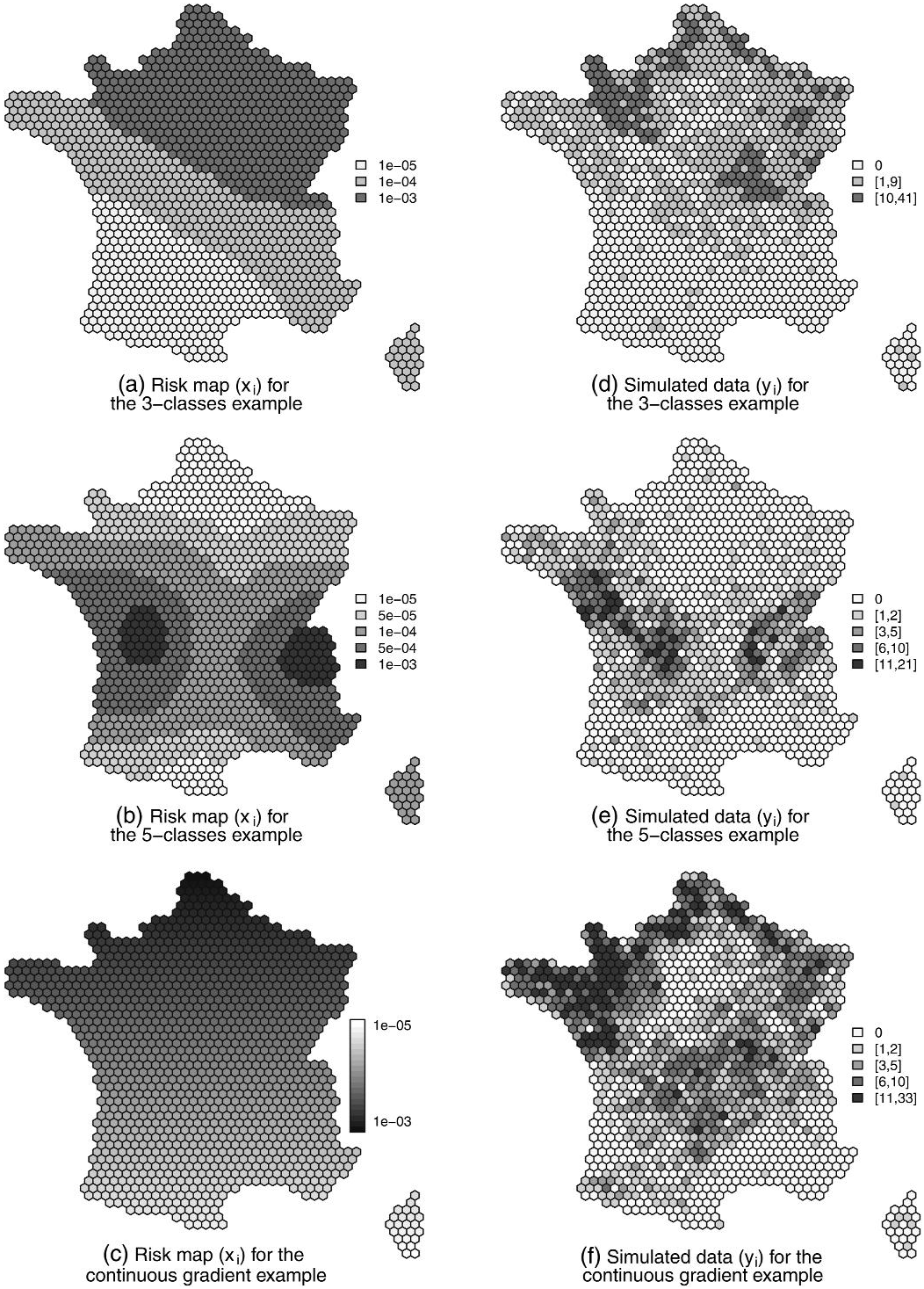

Fig. 1. Simulated data sets: Arbitrary generated risk maps (x_is) used to simulate data sets and examples of cases maps (y_is) for the 3-classes, the 5-classes, and the continuous gradient examples.

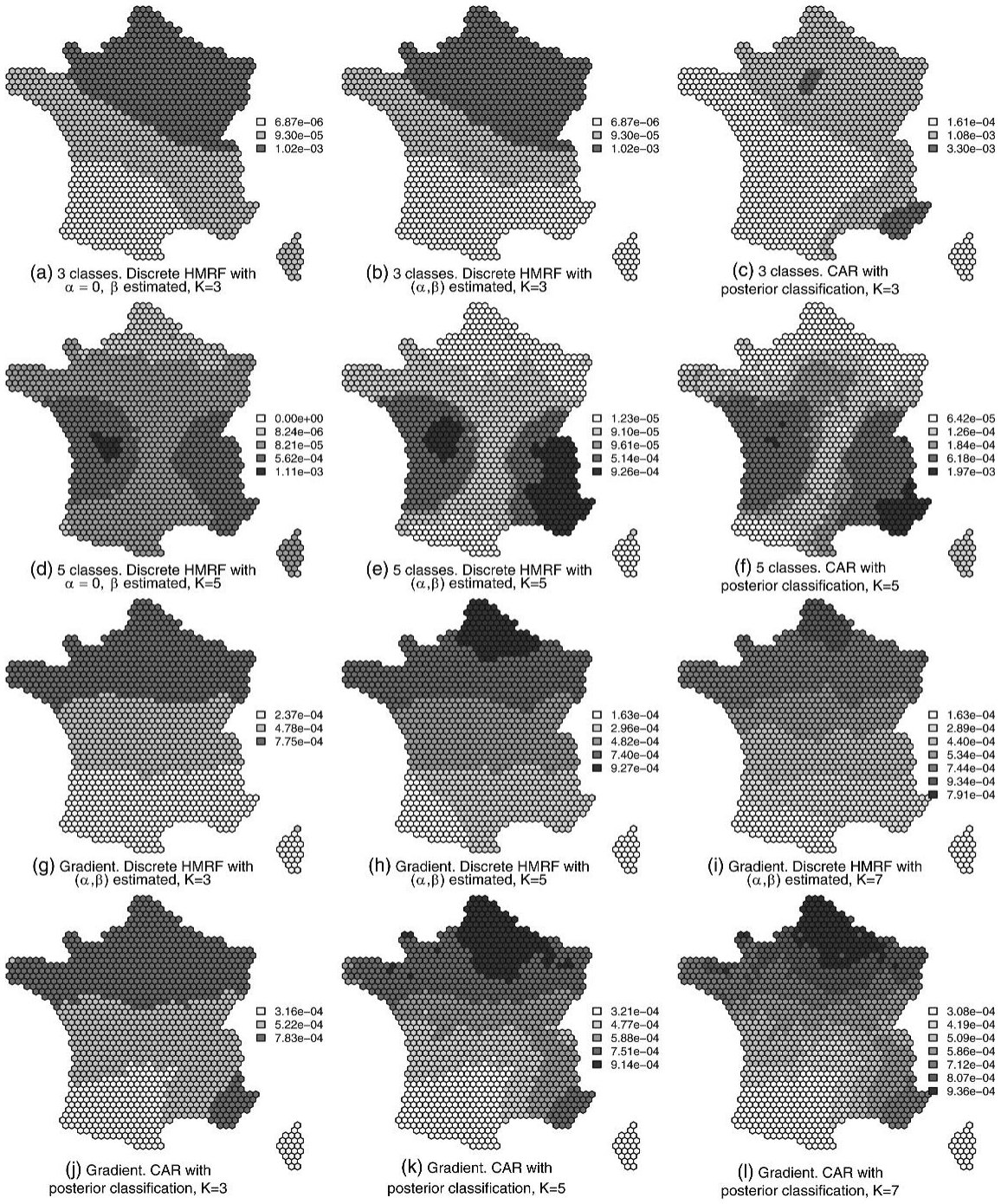

Fig. 2. Results on simulated examples for 3-classes, 5-classes, and continuous gradient with our discrete HMRF method and the CAR model with posterior classification.

shown in Figure 2(f). In particular, the highest risk is again estimated in the SE region where the population is very small. In all cases, the estimation of this eastern high-risk region extends too far into the south, probably because, due to the small population in the SE, the data do not contain enough information to

detect changes in risk level. It should be noted that the incorrect extension of the SE high-risk region into the south is much less important with our method. The 2 highest risk levels are very well estimated by our procedure, coming closer to the true values than the CAR estimates, which are slightly overestimated. In all the procedures, the 3 other true classes essentially are represented by 2 colors. One color matches approximately the lowest true risk region, with the corresponding estimates: $\hat{\theta}_2 = 8.24\times10^{-6}$ for our method with $\alpha = 0$, $\hat{\theta}_1 = 1.23\times10^{-5}$ for our complete method and $\hat{\theta}_1 = 6.42\times10^{-5}$ for the CAR model. The estimated risk values are closest to the true one ($\theta_1 = 10^{-5}$) with our procedure and the outlines of the lowest risk region are more clearly delimited and exact. The other colors encompass the low and medium true risk regions ($\theta_2 = 5\times10^{-5}$, $\theta_3 = 10^{-4}$). With our method, the estimates corresponding to this global region lie between the true ones: $\hat{\theta}_3 = 8.21\times10^{-5}$ when $\alpha = 0$, $\hat{\theta}_2 = 9.1\times10^{-5}$, and $\hat{\theta}_3 = 9.61\times10^{-5}$ for the complete procedure. In contrast, for the CAR model, the values $\hat{\theta}_2 = 1.26\times10^{-4}$ and $\hat{\theta}_3 = 1.84\times10^{-4}$ clearly overestimate these risks.

Experiments were made on these data sets with a number of classes K entered in the procedure that were different from the true one (see Section B of the supplementary material available at *Biostatistics* online). The good classification rate with our method is always equivalent to or higher than with the CAR model, and when K is higher than the true one, the discrete HMRF method can estimate at the end of the procedure less classes than initially asked.

Figures 2(g–l) present the estimated risk maps obtained for the NS gradient simulated data set for $K = 3$, 5, and 7 classes with our method and the CAR model. As already observed, the CAR model estimates high risks in the SE region with a very small population (see Figures 2(j–l)). With this model, the real NS pattern is not really retrieved. Moreover, the different risk regions are not clearly delimited, with some isolated units having a different risk from surrounding areas. In contrast, with our model (see Figures 2(g–i)), the risk regions are delimited very clearly and the NS pattern is retrieved very well. Only Figure 2(h) presents an incorrect curving delimitation in the SW which may be due to the lack of population in the SE, leading again to an overvaluation of the risk in this region. However, this phenomenon is very limited with our method compared to the CAR model. With regard to risk values, the lowest are overestimated by all methods but less by our model (between 1.63×10^{-4} and 2.37×10^{-4}) than by the CAR model (around 3×10^{-4}); and the estimates of the highest risks are similar with both methods, close to the true values and slightly underestimated (around 9.3×10^{-4}).

4.2 *Intensive simulation study*

We now present the performance of our method on intensive simulations. For the 3-classes and the 5-classes risk maps presented in Figures 1(a) and (b), we simulated 100 data sets (y_i maps as in Figures 1(d) and (e)) from the Poisson distribution $\mathcal{P}(n_i\theta_{x_i})$. We display the results obtained with our method for the true number of classes ($K = 3$ or 5) in different cases: With no spatial smoothing, i.e. when $\beta = 0$; in the simplified case where $\alpha = 0$ and β is estimated; with an intermediate procedure, α being fixed to the values obtained when $\beta = 0$; and with the complete model estimating both α and β. When β is estimated, we explored the 2 proposed φ_2. According to the number of classes in estimated maps (Class est.), we present the number of simulations (Sim. nb.), the mean value of the distribution of the absolute distance between the true and the estimated risk classes $d_i = |\hat{x}_i - x_i|$, and the median of the estimated risk values $\hat{\theta}_k$.

For the 3-classes simulations, as shown in Table 1, we observe that the result is quite similar with the 2 proposed φ_2. Our method usually retrieves the true number of classes, a little more often for φ_2 grad2. When the model returns the true number of classes, except when there is no spatial smoothing ($\beta = 0$), the good classification rate (corresponding to $d = 0$) is very high; important classification errors ($d = 2$, i.e. when the true x_i and estimated $\hat{x}_i$ classes are very different) are almost never observed; and the estimated risk values are very close to the true ones.

Table 1. *Distribution of the absolute distance between the true and the estimated risk classes for the* 3*-classes intensive simulations and median of estimated risk values*

Method	Class estimated	Sim. nb.	$d =$ 0	1	2	Median value of $\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\theta}_3$
$\beta = 0$	3	91	65	31.1	4	2.2×10^{-5}	1.1×10^{-4}	10^{-3}
	2	9	57.7	35.4	6.9	7×10^{-5}	9.5×10^{-4}	10^{-3}
$\alpha = 0$, φ_2 grad1	3	100	94.7	5.3	0	10^{-5}	9.7×10^{-5}	10^{-3}
$\alpha = 0$, φ_2 grad2	3	99	96.1	3.9	0	10^{-5}	9.7×10^{-5}	10^{-3}
α fixed, φ_2 grad1	3	84	78.9	21	0	6×10^{-5}	1.2×10^{-4}	10^{-3}
	2	16	65.2	34.7	0.1	6.9×10^{-5}	6.7×10^{-5}	10^{-3}
α fixed, φ_2 grad2	3	97	83.8	16.2	0	1.3×10^{-5}	10^{-4}	10^{-3}
	2	2	64.5	35.4	0.2	7×10^{-5}	0	10^{-3}
φ_2 grad1	3	88	87.9	12.1	0	1.2×10^{-5}	10^{-4}	10^{-3}
	2	12	64.5	35.4	0.1	6.9×10^{-5}	0	10^{-3}
φ_2 grad2	3	97	85.4	14.6	0	1.2×10^{-5}	10^{-4}	10^{-3}
	2	2	64.3	35.4	0.3	6.8×10^{-5}	2×10^{-4}	10^{-3}

For the 5-classes experiments, as shown in Table 2, the differences between the tested versions of our model are more distinct. We see clearly that the spatial smoothing is very important in such a context since important classification errors ($d \geqslant 3$) are observed when $\beta = 0$, even when the true number of classes is retrieved. We also notice that φ_2 grad2 performs a little better than φ_2 grad1. For example, when α and β are estimated, φ_2 grad1 gives the true number of classes in 65% of the simulations, whereas φ_2 grad2 gives it in 84%. In terms of classification, the best results are obtained when α and β are estimated (less errors in the number of classes and high rate of good classification) or when $\alpha = 0$ (high rate of good classification, even if the number of classes estimated is not the true one). When α is fixed, the results seem worse, but they are difficult to interpret since the important representation of $d = 1$ when the number of classes in estimated maps is not the true one can be due to a shift in the numbering of classes when some are not represented. We observe that the risk values are generally well estimated, especially for the highest risks. In general, when some classes disappear they correspond to those with the lowest risk, possibly because when no or few cases are observed the differences between low risks are difficult to discriminate. In this case, there is limited underestimation of other risks since these regions may be integrated into other ones.

4.3 *Illustration on a real data set: BSE in France*

BSE is a noncontagious neurodegenerative disease in cattle. This sudden and unexpected disease threatened bovine production in Europe and has been studied intensively (e.g. Abrial *and others*, 2005 for spatial analyses). In our data set, the numbers of cases shown in Figure 3(b) occurred in France between July 2001 and December 2005. We compared the following methods: Our discrete HMRF for φ_2 grad 2 with only β estimated ($\alpha = 0$), or α and β estimated, and the CAR model with and without a posterior classification. Following current practice, we considered a fixed number of classes, taking $K = 5$ to obtain the following risk levels: very low, low, medium, high, and very high.

As shown in Figures 3(c–f), the risk pattern is roughly the same with all methods, with high risks in Brittany (West), the Center, the Alps (East), and in the SW, and low risks in the South-Center, the North-East, and on Corsica island. The main differences between our method and the CAR model concern

Table 2. *Distribution of the absolute distance between the true and the estimated risk classes for the* 5*-classes intensive simulations and median of estimated risk values*

Method	Class estimated	Sim. nb.	$d=$ 0	1	2	3	4	Median value of $\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\theta}_3$	$\hat{\theta}_4$	$\hat{\theta}_5$
$\beta = 0$	5	37	36.4	33.3	21.8	4.4	0.6	3.1×10^{-5}	8.9×10^{-5}	10^{-4}	5.2×10^{-4}	1.1×10^{-3}
	4	62	35.5	39.4	20.5	3.3	0.3	3×10^{-5}	8.9×10^{-5}	9.4×10^{-5}	5.1×10^{-4}	1.1×10^{-3}
	3	1	33.3	28	31.2	6.5	0.9	6.4×10^{-5}	6.7×10^{-5}	7×10^{-5}	4.7×10^{-4}	1.1×10^{-3}
$\alpha = 0$	5	8	58.7	36.2	4.8	0	0	4.8×10^{-5}	6×10^{-6}	8.1×10^{-5}	5.5×10^{-4}	1.2×10^{-3}
and	4	43	60.2	34.8	4.4	0	0	0	6×10^{-6}	8.3×10^{-5}	5.3×10^{-4}	1.2×10^{-3}
φ_2 grad1	3	22	57.3	30.3	13.1	0	0	0	0	$7.3 \times 10{-}5$	5.7×10^{-4}	1.4×10^{-3}
	2	25	56.6	30.3	13.1	0	0	0	0	7.3×10^{-5}	5.8×10^{-4}	1.1×10^{-4}
$\alpha = 0$	5	9	60.5	34.3	5.8	0	0	5×10^{-5}	9×10^{-6}	8.5×10^{-5}	5.2×10^{-4}	1.1×10^{-3}
and	4	49	59.4	35.2	4.7	0	0	0	6×10^{-6}	8.2×10^{-5}	5.3×10^{-4}	1.2×10^{-3}
φ_2 grad2	3	17	57	30	13.1	0	0	0	4.5×10^{-6}	7.4×10^{-5}	5.7×10^{-4}	1.3×10^{-3}
	2	23	56.8	30.1	13.1	0	0	0	0	7.2×10^{-5}	5.8×10^{-4}	0
α fixed	5	13	26.2	41.5	22.2	1	0	5.6×10^{-5}	5×10^{-4}	2.6×10^{-4}	4.8×10^{-4}	9.6×10^{-4}
and	4	9	24.4	55.9	8.7	0.1	0	3.2×10^{-5}	10^{-4}	5.6×10^{-4}	5×10^{-4}	8.6×10^{-4}
φ_2 grad1	3	11	13.1	29.4	41.1	13.3	0	7×10^{-5}	5.8×10^{-4}	9.6×10^{-4}	7×10^{-5}	0
	2	58	7.1	55.1	24.6	13.2	0	0	0	0	7.1×10^{-5}	5.7×10^{-4}
α fixed	5	26	47.6	36.4	12.9	0	0	4.3×10^{-5}	1.1×10^{-4}	1.1×10^{-4}	5.4×10^{-4}	9.4×10^{-4}
and	4	4	15.7	40.7	33.9	0	0	6.9×10^{-5}	$5.3 \times 10{-}4$	5.2×10^{-4}	9.7×10^{-4}	9.9×10^{-4}
φ_2 grad2	3	8	13.1	20.2	51.7	0	0	7.2×10^{-5}	5.7×10^{-4}	1.5×10^{-3}	0	7×10^{-4}
	2	56	6.9	54.3	25.4	0	0	0	0	0	7.1×10^{-5}	5.7×10^{-4}
φ_2 grad1	5	65	58.7	34.3	6.1	0	0	1.8×10^{-5}	3.9×10^{-5}	8.7×10^{-5}	5.4×10^{-4}	9×10^{-4}
	4	28	52.6	38.4	7.2	0	0	1.3×10^{-5}	3.6×10^{-5}	8.2×10^{-5}	5×10^{-4}	9.3×10^{-4}
φ_2 grad2	5	84	59.3	33	5.8	0	0	1.5×10^{-5}	6.1×10^{-5}	9.1×10^{-5}	5.3×10^{-4}	9×10^{-4}
	4	12	58	36.6	5.8	0	0	0	1.7×10^{-5}	8.4×10^{-5}	5.3×10^{-4}	9.3×10^{-4}

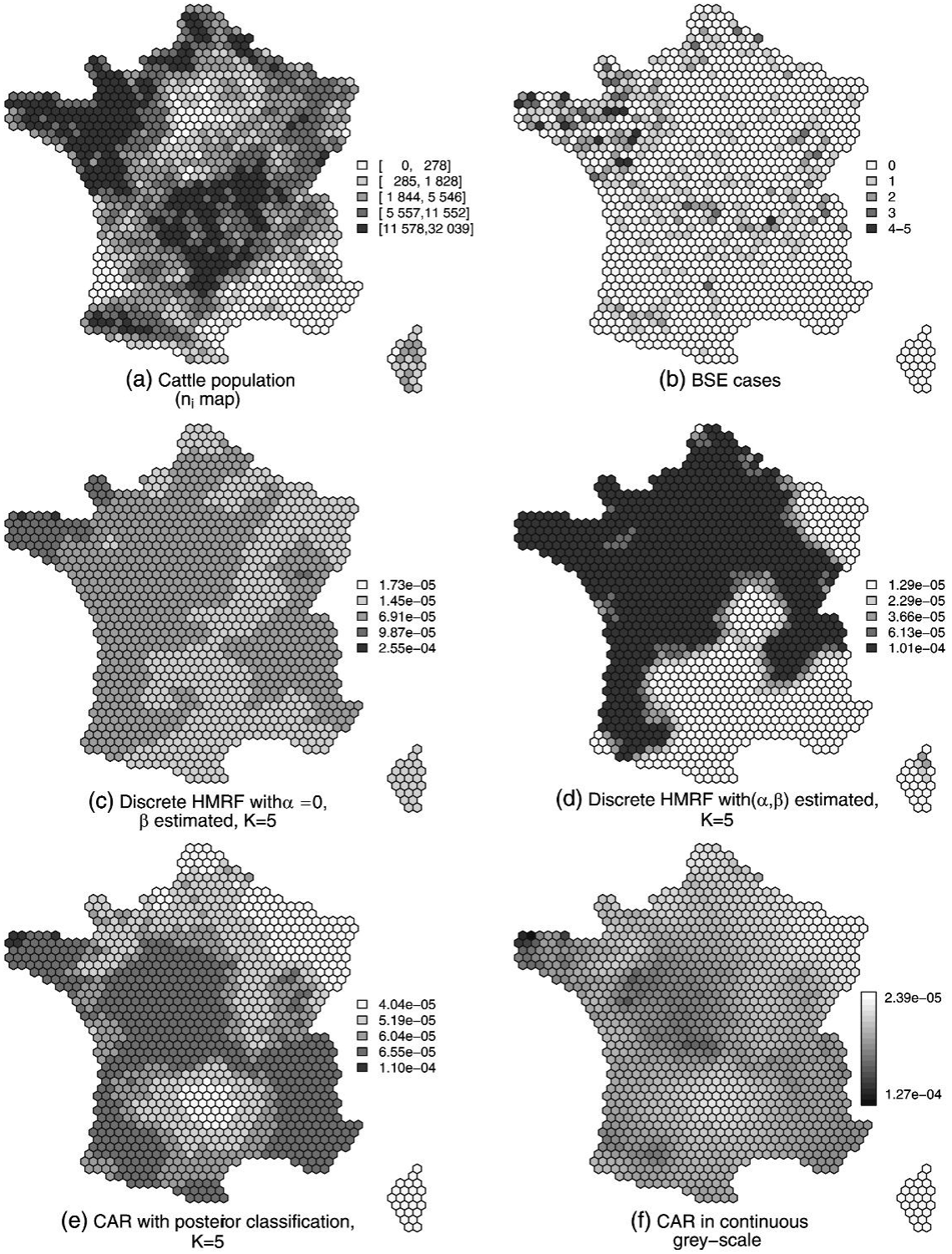

Fig. 3. Real data set: BSE in France. Population map, number of cases during the study and estimated risk maps.

the South around the Mediterranean that has a low risk with our method, and the North that has a lower risk with the CAR model. The apparent differences between maps must be interpreted in the light of the estimated risk values used. The lowest risk estimated by the CAR model ($\hat{\theta}_1 \approx 4\times10^{-5}$) is higher than the low risk estimated by our method when $\alpha = 0$ ($\hat{\theta}_2 = 1.45\times10^{-5}$) and the very low risk estimated by our complete method ($\hat{\theta}_1 \approx 1.3\times10^{-5}$). In contrast, the estimations of the highest risks are smaller in the CAR model ($\hat{\theta}_5 = 1.1\times10^{-4}$ and $\hat{\theta}_4 = 6.55\times10^{-5}$) than in our method ($\hat{\theta}_5 = 2.55\times10^{-4}$ and $\hat{\theta}_4 = 9.87\times10^{-5}$ when $\alpha = 0$, $\hat{\theta}_5 = 1.01\times10^{-4}$ for the complete procedure). It also should be noted that Figure 3(f) is difficult to interpret as it does not highlight the important spatial structures as risk clusters. It illustrates why epidemiologists prefer the type of representations given in Figures 3(b–e) built with a posterior classification and is why we propose an integrated classification method. The maps produced by our discrete HMRF using φ_2 grad1 compared to Figures 3(c) and (d) show only a few differences located in the SW where, as already noted, the low population imply a more difficult estimation.

5. Discussion

In this article, we propose to adapt a method relying on discrete HRMF modeling to risk mapping to produce an automated, integrated, and unsupervised method for the classification of geographical units into risk classes. The main differences between this method and the CAR model for disease mapping are the latent discrete HMRF and the EM procedure for ML estimation. The main differences between our model and HMRF models used for image segmentation consist first in replacing the usual Gaussian distribution by a discrete Poisson distribution to link the HF of risk classes $\mathbf{X}$ to data $\mathbf{Y}$, and second in specific potential functions of order 2 for a spatial correlation taking into account a smooth spatial gradation between risk classes.

Our discrete HMRF-based method provides risk maps that are coherent with the CAR model but with fewer classification errors and more clearly delimited zones at risk. The best results in terms of estimated number of classes and classification errors are obtained for the discrete HMRF when estimating only β (α being set to zero) and for the complete procedure (α and β estimated). In practice, for computational reasons, we suggest using the version with $\alpha = 0$, which is more rapid. The simulations show that our model is particularly adapted to determining high-risk regions (both to precisely localize these regions and to estimate the associated risk level) that are of principal interest in practice for the eventual imposition of control procedures. Low-risk regions are more difficult to determine, especially when they are in regions with small populations, and our method tends to underestimate low risks; however, such regions are less important for decision makers. The CAR model leads to more classification errors and tends to overestimate all risks. Our experiments suggest that the CAR model is not adapted to rare diseases in very heterogeneous populations as it tends to estimate high risks in regions with very small populations.

The main advantage of our discrete HMRF method is that all the parameters are easy to interpret and the model can be adapted easily to different epidemiological situations. In particular, the interpretation of the potential function of order 2 in terms of neighborhood interaction enables a simple definition of the smoothing according to the intended application. The definition of the neighborhood used, in this paper simply based on geographical proximity, can be adapted to different contexts. For example, a dissymmetry due to an ecological gradient such as wind dissemination could be introduced. Another strength of our model is that the key classification step is integrated into the model instead of being a separate procedure as in the method currently used by animal epidemiologists.

Beyond the obvious advantages of drawing risk classes as visual tools, another strength is the spatial analysis of disease risk. The risk can depend on different explicative variables that influence its spatial repartition. If the effect of known explicative variables is taken into account when building the risk map,

the analysis of the remaining spatial structure in the risk map can help identify other, unsuspected factors implied in the epidemiological risk. Such a method already has been considered in conventional risk mapping models. The same approach could be envisaged easily in the discrete HMRF framework by introducing the effect of covariates.

SUPPLEMENTARY MATERIAL

Supplementary material is available at http://biostatistics.oxfordjournals.org.

ACKNOWLEDGMENTS

We sincerely thank the Associate Editor and the referee for their careful reading, constructive comments, and relevant remarks that helped to improve the presentation of the method. We also are very grateful to M. Reichstadt, S. Cat, and M. Simon for their help on grid computing. *Conflict of Interest:* None declared.

FUNDING

The Institut National de la Recherche Agronomique, Santé Animale Department, France.

REFERENCES

ABRIAL, D., CALAVAS, D., JARRIGE, N. AND DUCROT, C. (2005). Poultry, pig and the risk of BSE following the feed ban in France —a spatial analysis. *Veterinary Research* **36**, 615–628.

ALFO, M., NIEDDU, L. AND VICARI, D. (2009). Finite mixture models for mapping spatially dependent disease counts. *Biometrical Journal* **52**, 84–97.

BESAG, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society Series B* **2**, 192–236.

BESAG, J. (1986). On the statistical analysis of dirty pictures. *Journal of the Royal Statistical Society Series B* **48**, 259–302.

BESAG, J., YORK, J. AND MOLLIÉ, A. (1991). Bayesian image restoration, with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics* **43**, 1–59.

DEMPSTER, A. P., LAIRD, N. M. AND RUBIN, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society Series B* **39**, 1–38.

FRALEY, C. AND RAFTERY, A. E. (2007). Bayesian regularization for normal mixture estimation and model-based clustering. *Journal of Classification* **24**, 155–181.

GREEN, P. J. AND RICHARDSON, S. (2002). Hidden Markov models and disease mapping. *Journal of the American Statistical Association* **97**, 1055–1070.

KNORR-HELD, L. AND RASSER, G. (2000). Bayesian detection of clusters and discontinuities in disease maps. *Biometrics* **56**, 13–21.

KNORR-HELD, L., RASSER, G. AND BECKER, N. (2002). Disease mapping of stage-specific cancer incidence data. *Biometrics* **58**, 492–501.

KNORR-HELD, L. AND RICHARDSON, S. (2003). A hierarchical model for space-time surveillance data on meningococcal disease incidence. *Journal of the Royal Statistical Society Series B* **52**, 169–183.

LI, S. Z. (2001). *Markov Random Field Modeling in Image Analysis*. Computer Science Workbench. London: Springer.

MACNAB, Y. C. (2010). On Gaussian Markov random fields and Bayesian disease mapping. *Statistical Methods in Medical Research* **20**, 49–68.

MCLACHLAN, G. J. AND KRISHNAN, T. (2008). *The EM Algorithm and Extensions*, 2nd edition. Hoboken, NJ: Wiley, p. 360.

MOLLIÉ, A. (1999). Bayesian and empirical Bayes approaches to disease mapping. In: Lawson, A., Biggeri, A. and Bohning, D. (editors), *Disease Mapping and Risk Assessment for Public Health*. Chichester: Wiley, pp. 15–29.

PASCUTTO, C., WAKEFIELD, J., BEST, N., RICHARDSON, S., BERNARDINELLI, L., STAINES, A. AND ELLIOTT, P. (2000). Statistical issues in the analysis of disease mapping data. *Statistics in Medicine* **19**, 2493–2519.

ROBERTSON, C., NELSON, T. A., MACNAB, Y. C. AND LAWSON, A. B. (2010). Review of methods for space-time disease surveillance. *Spatial and Spatio-temporal Epidemiology* **1**, 105–116.

WEI, G. C. G. AND TANNER, M. A. (1990). A Monte Carlo implementation of the EM algorithm and the poor man's data augmentation algorithms. *Journal of the American Statistical Association* **85**, 411.

WU, F. Y. (1982). The Potts model. *Reviews of Modern Physics* **54**, 235–268.

[*Received July 17, 2009; revised October 25, 2011; accepted for publication October 26, 2011*]