Discrete Random Variables

Antoine Ayache
University of Lille

Contents

1 Some generalities

Definition 1.1. *A real-valued random variable is a random numerical quantity related to some random experiment.*

Usually one denotes this random quantity by X, Y, Z or T, and one denotes by $X(\Omega)$, $Y(\Omega)$, $Z(\Omega)$ or $T(\Omega)$ the set of all its possible values. Notice that Ω is the probability space corresponding to the random experiment we are interested in.

Definition 1.2. *One says that a random variable X is discrete when $X(\Omega)$ only consists in isolated values, as for instance $X(\Omega) = \{1, 2, 3, 4, 5, 6\} = \{1, \ldots, 6\}$; in the latter case the random variable X can not take a value strictly between 1 and 2, 2 and 3, and so on.*

Example 1.1. *One tosses a fair coin* 3 *times and one denotes by* X *the random total number of Heads. The probability space associated to this random experiment is:*

$$\Omega = \{(H,H,H);(T,H,H);(H,T,H);(H,H,T);(T,T,H);(T,H,T);(H,T,T);(T,T,T)\}.$$

One mentions that the elementary event $\omega = (H,T,H)$ *means that one obtains a Head at the first toss and a Tail at second one and a Head at the third one. What is the meaning of each one of the seven other elementary events ?*

The set of all the possible values of X *is* $X(\Omega) = \{0,1,2,3\}$. *More precisely:*

ω	(H,H,H)	(T,H,H)	(H,T,H)	(H,H,T)	(T,T,H)	(T,H,T)	(H,T,T)	(T,T,T)
$X(\omega)$	3	2	2	2	1	1	1	0

Definition 1.3. (Probability distribution) *To each value* $i \in X(\Omega)$ *of a discrete random variable* X *one associates the probability* $\mathbb{P}(X = i)$ *which is the probability of the event* $\{X = i\}$. *The collection of all the probabilities associated to all the values of* X *is called the probability distribution of* X.

Example 1.1 (revisited) *1) Determine the probability distribution of the random variable* X.
2) Calculate the four probabilities $\mathbb{P}(X < 2)$, $\mathbb{P}(X \le 2)$, $\mathbb{P}(1 \le X \le 3)$ *and* $\mathbb{P}(1 < X \le 3)$.

Proposition 1.1. *Let* X *be a discrete random variable.*

(a) For all $i \in X(\Omega)$ *and* $j \in X(\Omega)$, *the events* $\{X = i\}$ *and* $\{X = j\}$ *are incompatible as soon as* $i \neq j$.

(b) The sum of all the probabilities associated to all the values of X *is equal to* 1, *more formally, one has*

$$\sum_{i \in X(\Omega)} \mathbb{P}(X = i) = 1\,. \tag{1.1}$$

Definition 1.4. (cumulative function) *The cumulative function of a random variable* X *is the function denoted by* F_X *and defined, for all real number* t, *as:*

$$F_X(t) = \mathbb{P}(X < t)\,. \tag{1.2}$$

For each real numbers a *and* b *such that* $a < b$, *one has*

$$\mathbb{P}(a \le X < b) = F_X(b) - F_X(a)\,.$$

Problem 1.1. *Let* F_X *be the cumulative function of the random variable* X *in Example 1.1 calculate* $F_X(0)$, $F_X(0,1)$, $F_X(1,2)$, $F_X(2,4)$ *and* $F_X(3,1)$.

Proposition 1.2. *Generally speaking the cumulative function F_X of a random variable X is always with values in the interval $[0,1]$. Moreover, it satisfies the following two important properties.*

(a) F_X is a non-decreasing function, that is for all real numbers t_1 and t_2 satisfying $t_1 \le t_2$, one has $F_X(t_1) \le F_X(t_2)$.

(b) One has $\lim_{t\to-\infty} F_X(t) = 0$ and $\lim_{t\to+\infty} F_X(t) = 1$.

Definition 1.5. (Expectation, Variance and Standard Deviation) *Let X be a discrete random variable.*

(a) The expectation of X is denoted by $\mathbb{E}(X)$ and defined as the weighted mean of all the values of X, that is

$$\mathbb{E}(X) = \sum_{i\in X(\Omega)} i\,\mathbb{P}(X=i)\,. \tag{1.3}$$

(b) The variance of X is denoted by $\mathrm{Var}(X)$ and defined as

$$\mathrm{Var}(X) = \mathbb{E}\big[\big(X-\mathbb{E}(X)\big)^2\big] = \mathbb{E}(X^2) - E(X)^2\,. \tag{1.4}$$

(c) The standard deviation of X is denoted by $\sigma(X)$ and defined as

$$\sigma(X) = \sqrt{\mathrm{Var}(X)}\,. \tag{1.5}$$

Roughly speaking $\sigma(X)$ allows to measure "the average distance" between the values of X and the expectation $\mathbb{E}(X)$.

Proposition 1.3 (Markov's inequality). *Let X be an arbitrary random variable. Then, for all strictly positive real number u, one has*

$$\mathbb{P}\big(|X|>u\big) \le \frac{\mathbb{E}(|X|)}{u}\,, \tag{1.6}$$

where $|X|$ denotes the absolute value of X. Recall that for, any real number x, its absolute value $|x|$ is the nonnegative real number such that $|x|=x$ if $x\ge 0$ and $|x|=-x$ if $x<0$ (for instance $|-2,3|=2,3$ and $|4,6|=4,6$).

Proposition 1.4 (Chebyshev's inequality). *Let X be a random variable whose expectation $\mathbb{E}(X)$ and variance $\mathrm{Var}(X)$ exist. Then, for all strictly positive real number u, one has*

$$\mathbb{P}\Big(\big|X-\mathbb{E}(X)\big|>u\Big) \le \frac{\mathrm{Var}(X)}{u^2}\,, \tag{1.7}$$

where $\left|X - \mathbb{E}(X)\right|$ *denotes the absolute value of* $X - \mathbb{E}(X)$. *Notice that one can derive from (1.7) that, for all strictly positive real number* u, *one has*

$$\mathbb{P}\Big(X - u \leq \mathbb{E}(X) \leq X + u\Big) \geq 1 - \frac{\mathrm{Var}(X)}{u^2} \tag{1.8}$$

as well as

$$\mathbb{P}\Big(\mathbb{E}(X) - u \leq X \leq \mathbb{E}(X) + u\Big) \geq 1 - \frac{\mathrm{Var}(X)}{u^2}\,. \tag{1.9}$$

Remark 1.1. *The Chebyshev's inequality is of great importance for constructing a confidence interval for an unknown parameter of a probability distribution starting from observed data related with this distribution. More precisely, let us denote, for instance, by* p *this unknown parameter. A confidence interval for* p *is a random interval* $I = [a, b]$ *such that the probability that* p *belongs to* I *is (very) close to* 1 *(100%), namely one has that*

$$\mathbb{P}(a \leq p \leq b) \geq \alpha\,, \tag{1.10}$$

where the quantity α, *which is called the level of confidence, is close to* 1 *(typically one has* $\alpha =$ *95% or 98% or 99%). Notice the smaller is* $b - a$ *the diameter (or length) of the confidence interval* I *the better is the approximation of the unknown parameter* p *provided by this interval.*

Example 1.1 (revisited) *3) Calculate the expectation* $\mathbb{E}(X)$.
4) Calculate the variance $\mathrm{Var}(X)$.
5) Calculate the standard deviation $\sigma(X)$.

Remark 1.2. *Let* $Y_1, \ldots, Y_n$ *be* n *random variables and let* $a_1, \ldots, a_n, b$ *be* $n + 1$ *real numbers.*

$$\mathbb{E}(a_1 Y_1 + \ldots + a_n Y_n + b) = a_1 \mathbb{E}(Y_1) + \ldots + a_n \mathbb{E}(Y_n) + b \tag{1.11}$$

Example 1.1 (revisited) *6) Antony and John play together a game whose rules are the following. A coin is tossed three times:*

(i) if the result of the first toss is "tail" then Antony gives to John 1 Euro, otherwise John gives him 0,5 Euro;

(ii) if the result of the second toss is "tail" then Antony gives to John 1 Euro, otherwise John gives him 2,5 Euro;

(ii) if the result of the third toss is "tail" then Antony gives to John 0,5 Euro, otherwise John gives him nothing.

For all $k = 1, 2, 3$, one denotes by Y_k the random variable which is equal to 1 if the result of the k-th toss is "head" and which is equal to 0 otherwise. One denotes by G_a the total random amount of money which is given by John to Antony. One denotes by G_j the total random amount of money which is given by Antony to John.

Find the formula which connects G_a to Y_1 and Y_2, then use this formula for computing the expectation $\mathbb{E}(G_a)$.

Find the formula which connect G_j to Y_1, Y_2 and Y_3, then use this formula for computing the expectation $\mathbb{E}(G_j)$.

Do you think that in this game one of those two players has an advantage over the other ? (justify your answer)

Definition 1.6. *One says that two random variables Y_1 and Y_2 are independent, if for all $i_1 \in Y_1(\Omega)$ and $i_2 \in Y_2(\Omega)$ the events $\{Y_1 = i_1\}$ and $\{Y_1 = i_2\}$ are independent.*

One says that n random variables $Y_1, \ldots, Y_n$ are pairwise independent if any two of them are independent.

One says that n random variables $Y_1, \ldots, Y_n$ are mutually independent, if for all $i_1 \in Y_1(\Omega), \ldots, i_n \in Y_n(\Omega)$ the events $\{Y_1 = i_1\}, \ldots, \{Y_n = i_n\}$ are mutually independent.

Remark 1.3. *Let Y_1 and Y_2 be two independent random variables, and let a_1, a_2 and b be three real numbers then*

$$\mathrm{Var}(a_1 Y_1 + a_2 Y_2 + b) = a_1^2 \,\mathrm{Var}(Y_1) + a_2^2 \,\mathrm{Var}(Y_2) \tag{1.12}$$

More generally, let $Y_1, \ldots, Y_n$ be n pairwise independent random variables and let $a_1, \ldots, a_n, b$ be $n+1$ real numbers then

$$\mathrm{Var}(a_1 Y_1 + \ldots + a_n Y_n + b) = a_1^2 \,\mathrm{Var}(Y_1) + \ldots + a_n^2 \mathrm{Var}(Y_n)\,. \tag{1.13}$$

Observe that there is no "b" in the right-hand side of the equalities (1.12) and (1.13).

Example 1.1 (revisited) *7) Calculate the standard deviations $\sigma(G_a)$ and $\sigma(G_j)$ of the random variables G_a and G_j. Comment your result.*

Definition 1.7. *Let X and Y be two random variables, the covariance of X and Y is denoted by $\mathrm{Cov}(X, Y)$ and defined as:*

$$\mathrm{Cov}(X, Y) = \mathbb{E}\big[\big(X - \mathbb{E}(X)\big)\big(X - \mathbb{E}(Y)\big)\big] = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)\,. \tag{1.14}$$

Observe that one always has $\mathrm{Cov}(X, Y) = \mathrm{Cov}(Y, X)$ and $\mathrm{Cov}(X, X) = \mathrm{Var}(X)$.

Remark 1.4. *Let X be a random variable, let $Y_1, \ldots, Y_n$ be n other random variables and let $a_1, \ldots, a_n, b$ be $n+1$ real numbers. One has*

$$\mathrm{Cov}\big(X, a_1 Y_1 + \ldots + a_n Y_n + b\big) = a_1 \mathrm{Cov}(X, Y_1) + \ldots + a_n \mathrm{Cov}(X, Y_n) \tag{1.15}$$

and

$$\mathrm{Cov}\big(a_1 Y_1 + \ldots + a_n Y_n + b, X\big) = a_1 \mathrm{Cov}(Y_1, X) + \ldots + a_n \mathrm{Cov}(Y_n, X)\,. \tag{1.16}$$

Observe that there is no "b" in the right-hand side of the equalities (1.15) and (1.16).

Remark 1.5. *When two random variables X and Y are independent, then one necessarily has that $\mathrm{Cov}(X,Y) = 0$. The reciprocal is not necessarily true; namely, one may have $\mathrm{Cov}(X,Y) = 0$ for two dependent random variables X and Y.*

Definition 1.8. *Let X and Y be two random variables whose standard deviations $\sigma(X)$ and $\sigma(Y)$ do not vanish. The correlation coefficient of X and Y is denoted by $\rho(X,Y)$ and defined as:*

$$\rho(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sigma(X)\sigma(Y)}\,. \tag{1.17}$$

Remark 1.6. *The value taken by the correlation coefficient $\rho(X,Y)$ is always between -1 and 1.*

Example 1.1 (revisited) *8) Calculate the correlation coefficient $\rho(G_a, G_j)$.*

2 Discrete uniform distribution

Definition 2.1. *Let an arbitrary integer $n \geq 1$. One says that a random variable X has a discrete uniform distribution over the the set $\{1, \ldots, n\}$, if $X(\Omega) = \{1, \ldots, n\}$ and for all $i \in X(\Omega)$, one has*

$$\mathbb{P}(X = i) = \frac{1}{n}\,.$$

Example 2.1. *One tosses a fair dice and one denotes by X the obtained result. Then the random variable X has a discrete uniform distribution over the set $\{1,2,3,4,5,6\} = \{1, \ldots, 6\}$.*

Proposition 2.1. *Generally speaking when X has a discrete uniform distribution over the the set $\{1, \ldots, n\}$, then its expectation and variance are given by the formulas:*

$$\mathbb{E}(X) = \frac{n+1}{2} \quad \textit{and} \quad \mathrm{Var}(X) = \frac{n^2-1}{12}$$

Thus in the case of the random variable X in Exemple 2.1, one has

$$\mathbb{E}(X) = \frac{n+1}{2} = \frac{6+1}{2} = 3,5$$

and

$$\mathrm{Var}(X) = \frac{n^2-1}{12} = \frac{6^2-1}{12} = \frac{35}{12} = 2,92$$

3 Bernoulli distribution and binomial distribution

Definition 3.1. *One says that a random variable Y has Bernoulli distribution of parameter $p \in [0,1]$, if $Y(\Omega) = \{0,1\}$ (Y can only take two values 0 and 1) and*

$$\mathbb{P}(Y = 1) = p$$

This implies that $\mathbb{P}(Y = 0) = 1 - p$. Usually, one sets $q = 1 - p$.

One mention in passing that the random variables Y_1, Y_2 and Y_3, we have introduced in part 6) of Exemple 1.1, have the same Bernoulli distribution of parameter $p = 0,5$.

Proposition 3.1. *Generally speaking when Y has a Bernoulli distribution of parameter $p \in [0,1]$, then its expectation and variance are given by the formulas:*

$$\mathbb{E}(Y) = p \quad \textit{and} \quad \mathrm{Var}(Y) = p(1-p) = pq$$

Problem 3.1. *Give the proof of Proposition 3.1.*

Definition 3.2. *One says that a random variable X has a binomial distribution of parameters $n \in \mathbb{N}^*$ ($\mathbb{N}^*$ being the set of the strictly positive integers) and $p \in [0,1]$, if $X(\Omega) = \{0,1,\ldots,n\}$ and, for all $k \in X(\Omega)$,*

$$\mathbb{P}(X = k) = C_n^k \, p^k \, q^{n-k}\,, \tag{3.1}$$

where $q = 1 - p$.

Proposition 3.2. *Suppose that n independent trials, each of which results in a "success" with probability p and in a "failure" with probability $q = 1 - p$, are to be performed. If X represents the number of successes that occur in the n trials, then this random variable X has a binomial distribution of parameters n and p.*

Thus, it turns out that a random variable X has binomial distribution of parameters n and p if and only if it can be expressed as

$$X = Y_1 + \ldots + Y_n\,,$$

where $Y_1, \ldots, Y_n$ are mutually independent Bernoulli random variables of the same parameter p.

The following proposition is a consequence of Proposition 3.2, Remark 1.2, Remark 1.3 and Proposition 3.1.

Proposition 3.3. *Generally speaking when X has a binomial distribution of parameters $n \in \mathbb{N}^*$ and $p \in [0,1]$, then its expectation and variance are given by the formulas:*

$$\mathbb{E}(X) = np \quad \text{and} \quad \mathrm{Var}(X) = np(1-p) = npq$$

Remark 3.1. *Notice that the useful inequality*

$$x(1-x) \leq \frac{1}{4}, \quad \text{for all } x \in [0,1], \tag{3.2}$$

implies that the variance of a random variable X having a binomial distribution of arbitrary parameters $n \in \mathbb{N}^*$ and $p \in [0,1]$ always satisfies

$$\mathrm{Var}(X) \leq \frac{n}{4}. \tag{3.3}$$

Proposition 3.4. *Let X_1 and X_2 be two independent random variables such that X_1 has a binomial distribution with parameters $n = n_1$ and $p = p_0$, and X_2 has a binomial distribution with parameters $n = n_2$ and $p = p_0$. Then the random variable $X_1 + X_2$ has a binomial distribution with parameters $n = n_1 + n_2$ and $p = p_0$.*

4 Poisson distribution

Definition 4.1. *Let λ be a strictly positive real number, that is $\lambda \in]0, +\infty[$. One says that a random variable X has a Poisson distribution of parameter λ, if $X(\Omega) = \mathbb{N}$="the set of the nonnegative integers", and, for all $k \in X(\Omega)$, one has*

$$\mathbb{P}(X = k) = \frac{e^{-\lambda}\,\lambda^k}{k!}.$$

Proposition 4.1. *Generally speaking when X has a Poisson distribution of parameter $\lambda \in]0, +\infty[$, then its expectation and variance are given by the formulas:*

$$\mathbb{E}(X) = \lambda \quad \text{and} \quad \mathrm{Var}(X) = \lambda$$

Proposition 4.2. *Let X_1 and X_2 be two independent random variables such that X_1 has a Poisson distribution with parameter $\lambda = \lambda_1$ and X_2 has a Poisson distribution with parameter $\lambda = \lambda_2$. Then the random variable $X_1 + X_2$ has a Poisson distribution with parameter $\lambda = \lambda_1 + \lambda_2$.*

Proposition 4.3. *One of the main reasons for which Poisson distribution has a great importance is that, under some conditions, it allows to approximate binomial distribution. More precisely let Z be a random variable having a binomial distribution whose parameters n and p satisfy the conditions:*

$$n \geq 30 \quad \textit{and} \quad p \leq 0{,}1 \quad \textit{and} \quad np \leq 15\,, \tag{4.1}$$

then the binomial distribution of Z can be approximated by a Poisson distribution of parameter $\lambda = np$, namely, for all $k \in \{0, 1, \ldots, n\}$, one has

$$\mathbb{P}(Z = k) = C_n^k\, p^k (1-p)^{n-k} \simeq \frac{e^{-np}(np)^k}{k!}\,.$$

5 Hypergeometric distribution

Definition 5.1. *Let N, M and n be 3 nonnegative integers such that $0 \leq n \leq N$ and $0 \leq M \leq N$, one says that a random variable X has an hypergeometric distribution of parameters N, M and n, if $X(\Omega)$ is the set of the integers k such that $0 \leq k \leq M$ and $0 \leq n-k \leq N-M$ and, for all $k \in X(\Omega)$, one has*

$$\mathbb{P}(X = k) = \frac{C_M^k\, C_{N-M}^{n-k}}{C_N^n}\,. \tag{5.1}$$

Concrete meaning:

- *N is the size of the (statistical) population we are interested in.*
- *M is the number of individuals in this population who have some feature $\mathcal{F}$.*
- *$p = M/N$ is the proportion of individuals in the population who have the feature $\mathcal{F}$.*
- *n is the size of the sample which is chosen at random* **without replacement** *from the population; one does not take into account the order in which the n individuals forming the sample are chosen.*
- *X is the random number of individuals in the sample who have the feature $\mathcal{F}$.*

It is worth mentioning that hypergeometric distribution plays an important role in quality controls and in opinion polls.

Proposition 5.1. *Generally speaking when X has an hypergeometric distribution of parameters N, M and n, then its expectation and variance are given by the formulas:*

$$\mathbb{E}(X) = n\Big(\frac{M}{N}\Big) = np$$

and

$$\mathrm{Var}(X) = n\Big(\frac{N-n}{N-1}\Big)\Big(\frac{M}{N}\Big)\Big(1-\frac{M}{N}\Big) = n\Big(\frac{N-n}{N-1}\Big)p(1-p)\,.$$

Proposition 5.2. *If N and M are large compared to n, and $p = M/N$ is not close to 0 or 1, then the hypergeometric distribution of parameters N, M and n can be approximated by the binomial distribution of parameter n and $p = M/N$. More precisely, for all $k \in X(\Omega)$, one has*

$$\mathbb{P}(X=k) = \frac{C_M^k\, C_{N-M}^{n-k}}{C_N^n} \simeq C_n^k\Big(\frac{M}{N}\Big)^k\Big(1-\frac{M}{N}\Big)^{n-k}.$$

6 Geometric distribution

Definition 6.1. *One says that a random variable X has a geometric distribution of parameter $p \in]0,1[$, if $X(\Omega) = \mathbb{N}^* =$"the set of the strictly positive integers", and for $k \in X(\Omega)$, one has*

$$\mathbb{P}(X=k) = q^{k-1}\,p\,, \tag{6.1}$$

where $q = 1-p$.

Example 6.1. *Suppose that independent trials, each having probability p of being a success, are performed until a success occurs. If we let X be the number of trials required until the first success then X has a geometric distribution of parameter p.*

Proposition 6.1. *Generally speaking when X has a geometric distribution of parameter $p \in]0,1[$, then its expectation and variance are given by the formulas:*

$$\mathbb{E}(X) = \frac{1}{p} \quad \textit{and} \quad \mathrm{Var}(X) = \frac{q}{p^2}$$

Remark 6.1. *In the frame of geometric distribution the formula*

$$\sum_{k=0}^{n} x^k = 1 + x + \ldots + x^n = \frac{1-x^{n+1}}{1-x}\,,$$

which holds for all nonnegative integer n and real number $x \neq 1$, is very useful.

7 Problems

Problem 7.1. *Pieces of metal manufactured in a workshop can have at most 3 defects (too heavy piece, too long, or too wide). One chooses at random a piece of metal in this workshop and denotes by D its number of defects. One assumes that the probability distribution of the random variable D satisfies $\mathbb{P}(D = 0) = 90\%$, $\mathbb{P}(D = 1) = 3\%$ and $\mathbb{P}(D = 3) = 1,5\%$. Imperfect pieces can still be used and thus they can be sold. The more defects there are in a piece the lower is its price. More precisely the price X in euros of a piece of metal is connected to its number of defects D through the formula $X = 1 - 0,3D$.*
1) Calculate the expectation and the standard deviation of the random variable X.
2) Determine the probability distribution of X.
3) One chooses at random 2 pieces. Their 2 prices in euros are denoted by X_1 and X_2; one assumes that the random variables X_1 and X_2 are independent.
(a) Calculate the following probabilities: $\mathbb{P}\big(\{X_1 = 1\} \cup \{X_2 = 1\}\big)$, $\mathbb{P}(X_1 + X_2 = 2)$, $\mathbb{P}(X_1 + X_2 = 1,70)$ and $\mathbb{P}(X_1 = X_2)$.
(b) Calculate the expectation $\mathbb{E}(X_1 - X_2)$ and the standard deviation $\sigma(X_1 - X_2)$.
(c) Calculate the covariance $\mathrm{Cov}(X_1 + X_2, X_1 - X_2)$

Problem 7.2. *Two products A and B are commercialized by a society. One chooses at random a day in the next week. The random variables X denotes the number of items of the product A which will be sold by the society during this day. The random variables Y denotes the number of items of the product B which will be sold by the society during this day. The random variables Z denotes the number of items of the two products A and B together which will be sold by the society during this day. Statistical surveys have show that the standard deviations of X, Y and Z satisfy the equalities: $\sigma(X) = 10,2$; $\sigma(Y) = 13,5$ and $\sigma(Z) = 20,3$. Calculate the correlation coefficient of X and Y.*

Problem 7.3. *A seller has planed to visit one after the other 6 potential buyers for proposing to each one of them to buy a single product. One assumes that for each one of them the probability to buy is 20% and he takes this decision independently of the others. The random variable Z denotes the total number of products which will be sold thanks to the 6 visits.*
1) Determine the probability distribution of Z.
2) Calculate $\mathbb{E}(Z)$, $\mathrm{Var}(Z)$ and $\sigma(Z)$.
3) Calculate the probabilities $\mathbb{P}(Z = 2)$, $\mathbb{P}(Z = 4)$, $\mathbb{P}(Z > 0)$, $\mathbb{P}(Z \leq 5)$, $\mathbb{P}(0 < Z < 6)$ and $\mathbb{P}(1 < Z \leq 4)$.

Problem 7.4. *The probability that a motor manufactured in some factory be defective is 1%. This factory has delivered 100 motors which have been chosen at random. One denotes by X the random number of defective motors among them.*

1) What is the probability distribution of the random variable X ?
2) By using the probability distribution of X, calculate the probabilities $\mathbb{P}(X = 1)$ and $\mathbb{P}(X > 2)$.
3) By using an approximation of the probability distribution of X, calculate the probability $\mathbb{P}(2 \leq X < 5)$.

Problem 7.5. *An association intends to organize a travel by bus which will take place on March 12, 2023. It has to support a fix cost of 630 euros whatever the number of travelers might be. 70 persons have bought their tickets for this travel; the association has sold each ticket at 25 euros. The probability that one of these persons abandons the travel is 5%; in this case the association reimburses him 80% of the price of his ticket. One assumes that each person who decides to abandon the travel takes his decision independently of the other persons who have planed to make it. At last, in order to simplify the problem, one assumes that the association does not resell the tickets of the persons who abandon the travel.*

One denotes by R the random number of the persons who will abandon the travel. One denotes by B the random profit that the association will have thanks to this travel; observe that it is not completely excluded that $B < 0$, in his case B will be a loss.
1) Determine the probability distribution of the random variable R, calculate its expectation, its variance and its standard deviation.
2) a) Calculate the probability that no person abandon the travel.
b) Calculate the probability that at least 3 persons abandon the travel.
3) a) Give a formula for calculating B from R.
b) Calculate the expectation, the variance and the standard deviation of B.
4) By using an approximation of the probability distribution of R by a well-chosen distribution, find an estimate of the probability that the profit B be strictly bigger than 1000 euros.

Problem 7.6. *Mister Smith is the owner of 3 cinemas A, B, and C in 3 different small cities. The independent random variables X_1, X_2 and X_3 respectively denote the numbers of tickets per week sold by the cinemas A, B and C. One assumes they have Poisson distributions of parameters $\lambda_1 = 223$, $\lambda_2 = 254$ and $\lambda_3 = 279$.*

The random variable Y denotes the total number of tickets per week which are sold by the 3 cinemas together. Determine the probability distribution of Y and calculate its expectation and standard deviation.

Problem 7.7. *A safety engineer knows from statistical surveys that industrial accidents in her plant happen independently of each other, and there is a 25% probability of such an accident being caused by failure of employees to follow instructions. She decides to look at the accident reports until she finds one that*

shows an accident caused by failure of employees to follow instructions. The random variable X denotes the number of reports the safety engineer must examine until she finds a report showing an accident caused by employee failure to follow instructions.
1) Determine the probability distribution of the random variable X, calculate its expectation, its variance and its standard deviation.
2) Calculate the probabilities $\mathbb{P}(X = 5)$, $\mathbb{P}(X < 20)$ and $\mathbb{P}(2 \leq X \leq 10)$ by making as less calculations as possible.

Problem 7.8. *A light bulb has been selected at random. The random variable T denotes its lifetime measured in number of days. One assumes that T has a geometric distribution of parameter $p \in]0, 1[$. Show that, for any arbitrary strictly positive integers m and n the conditional probability $\mathbb{P}(T \geq m + n - 1/T \geq m)$ satisfies*

$$\mathbb{P}(T \geq m + n - 1/T \geq m) = \mathbb{P}(T \geq n)\,.$$

Give a concrete interpretation of this result.

Problem 7.9. *A coin having probability p of coming up head is successively flipped until the 5-th head appears. The random variable X denotes the numbers of flips required to this end.*
1) Determine $X(\Omega)$ the set of all the possible values of X.
2) Show that, for all $n \in X(\Omega)$, one has

$$\mathbb{P}(X = n) = \mathbb{P}\bigg(\Big\{Y_n = 1\Big\} \cap \Big\{\sum_{k=1}^{n-1} Y_k = 4\Big\}\bigg)\,,$$

where, for every $k \in \{1, \ldots, n-1, n\}$, one denotes by Y_k the Bernoulli random variable such that $Y_k = 1$ if the outcome of the k-th flip is head and $Y_k = 0$ otherwise.
3) By using the fact that the two random variables Y_n and $\sum_{k=1}^{n-1} Y_k$ are independent, calculate the probability $\mathbb{P}(X = n)$ in terms of n and p

Problem 7.10. *You are president of an on-campus special events organization which consists of 18 women and 15 men. You choose at random 7 persons among them for forming a committee. The random variable Z denotes the number of men in this committee.*
1) Determine $Z(\Omega)$ the set of all the possible values of Z.
2) Determine the probability distribution of Z.
3) Calculate the expectation, the variance and the standard deviation of Z.

Problem 7.11. *A supermarket has a stock of* 10000 *tin cans,* 550 *among them are of bad quality. Someone chooses at random a sample of* 20 *tin cans in this stock. The random variable* Z *denotes the number of tin cans of bad quality in this sample. Find an approximative value for the probability* $\mathbb{P}(Z \leq 2)$.

Problem 7.12 (On Markov's inequality and Chebishev's inequality (see Propositions 1.3 and 1.4)). *The random variable* X *denotes the number of items which will be produced in some factory during the next week.*
1) If one knows that the expectation $\mathbb{E}(X) = 500$*, then what can be said about the probability* $\mathbb{P}(X > 1000)$ *?*
2) If one knows that the expectation $\mathbb{E}(X) = 500$ *and the standard deviation* $\sigma(X) = 10$*, then what can be said about the probability* $\mathbb{P}(400 \leq X \leq 600)$ *?*

Problem 7.13. *One tosses a fair coin* 400 *times; the random variable* X *denotes the total number of heads. Show that*

$$\mathbb{P}(X \geq 170) \geq \frac{8}{9}\,.$$

Problem 7.14. *Let* B *be the same random variable as in Problem 7.5. Answer to the following questions by making as less calculations as possible.*
1) Show that the probability $\mathbb{P}(B \leq 0)$ *is strictly less than* $6,4\%$.
2) Show that the probability $\mathbb{P}(985 \leq B \leq 1115)$ *is strictly greater than* $68,5\%$.
3) Show that the probability $\mathbb{P}(B \geq 750)$ *is strictly greater than* $98,5\%$.

Problem 7.15. *The goal of this problem is to estimate the unknown probability* θ *that an individual, who is chosen at random in a very large population has the disease* $\mathcal{D}$*. A test is used for detecting it. For an individual having the disease there is a* 95% *probability that the test gives a positive result. For an individual not having the disease there is a* $2,5\%$ *probability that the test gives a positive result.*
1) One denotes by θ_* *the probability that the test gives a positive result for an individual selected at random in the population. Calculate* θ_* *in terms of* θ*. Then calculate* θ *in terms of* θ_*
2) One focuses on a sample of 500 *individuals who have been chosen at random and independently of each other in the population. Each one of these individuals has passed the test. The random variable* X *denotes the total number of positive results in the sample*

a) Determine the probability distribution of X.

b) In fact, it has been observed that $X = 31$*. Determine a confidence interval for* θ_* *at the level of confidence* $\alpha = 95\%$*, then use it for determining a confidence interval for* θ *at the level of confidence* $\alpha = 95\%$.

3) If one wants to construct at the level of confidence $\alpha = 95\%$ *a confidence* *interval for* θ *with a very small diameter equals to* $10^{-4} = 0,0001$*, then what should be the minimal size of the corresponding sample of individuals selected from the population ?*