Statistics (1): Estimation

Marco Banterlé, Christian Robert and Judith Rousseau

Practicals

2014-2015 L3, MIDO, Université Paris Dauphine

Table des matières

1 Random variables, probability, expectation

Exercice 1.1 ***Reminders*** Let X be an integrable random vaiable. Are the following propositions true/false :

— $E(1/X) = 1/E(X)$?
— $E(X^2) \geq E(X)^2$?
— If X is symetric with respect to 0 then $E[X] = 0$?
— $E(XY) = E(X)E(Y)$?

Brief justification is expected.

Exercice 1.2 ***The multinomial distribution*** Consider a population divided into k categories according to the proportions $p_1, ..., p_k$, with $0 \leq p_i \leq 1$ and $\sum_{i=1}^{k} p_i = 1$. We draw n individuals with relacement. Denote by N_i the number of individuals belonging to category i, among the n individuals.

1. Determine the distribution of $(N_1, ..., N_k)$.
2. What is the marginal distribution of N_j for all $j = 1, ..., k$? Compute the expectation of N_j, $j \leq k$.
3. Compute
$$P[N_1 = n_1 | N_2 = n_2]$$

Exercice 1.3 ***Exponential random variables : memoryless variables***

Let T be a real random variable. Assume that T satisfies $P(T > 0) > 0$ together with the following condition :

$$\forall (s,t) \in R_+^{*2},\ P(T > t + s) = P(T > s)P(T > t)$$

1. Show that for all $t > 0$, $P(T > t) > 0$.
2. Consider the following application $f :]0, +\infty[\to \mathbb{R}, t \mapsto \ln P(T > t)$. Show that for all positive real $t > 0$
$$f(t) = tf(1)$$
3. Show that it implies that T is an exponential random variable.

Exercice 1.4

1. Let X be a random variable following the uniform distribution on $\left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$. Determine the distribution of $Y = \tan X$.
2. Consider the same question for a general random variable X having a distribution absolutely continuous distribution with respect to Lebesgue measure, with density f, continuous everywhere (apart possibly at a finite number of points).
Apply this result to the case of an exponential random variable X with parameter $\theta > 0$.

Exercice 1.5 Let X be a real random variable with density

$$f_X(x) = \left(\frac{10}{x^2}\right) \mathbb{I}_{x>10} \,.$$

Compute $\mathbb{P}(X > 20)$, $F_X(x)$ and $\mathbb{E}(X)$.

Exercice 1.6

1. Let Y be a real random variable and Z a random variable, independent of Y, such that
$$\mathbb{P}(Z = 1) = \mathbb{P}(Z = -1) = \frac{1}{2} \,.$$
Show that the law of $X = ZY$ is symmetric and compute its characteristic function according to Φ_Y (characteristic function of Y). If the random variable Y admits the density f_Y on $\mathbb{R}$, what is the law of X ?
2. Let X be a random variable following the standard Laplace law :
$$f_X(x) = \frac{1}{2} \exp(-|x|) \,.$$
Show that, for every real t, we have that,
$$\Phi_X(t) = \frac{1}{1+t^2} \,.$$

2 bivariate random variables, change of variables and independence

Exercice 2.1 Let $X = (X_1, X_2) \in \mathbf{R}^2$ be a random Gaussian vector with density with respect to Lebesgue measure on $\mathbf{R}^2$:

$$f(x_1, x_2) = \frac{e^{-(x_1^2+x_2^2)/2}}{2\pi}, \quad \forall x_1, x_2 \in \mathbf{R}$$

Let $g : \mathbf{R} \setminus \{(0,0)\} \to \mathbf{R}^{+*} \times [0, 2\pi)$ inversible such that

$$g^{-1}(r, \theta) = (r\cos(\theta), r\sin(\theta)), \quad g(X_1, X_2) = (R, \Theta)$$

Determine the distribution of (R, Θ).

Exercice 2.2 Recall that for $p > 0$, we note

$$\Gamma(p) = \int_0^{+\infty} e^{-t}\, t^{p-1}\, dt$$

and denote $\Gamma(p,\theta)$ the distribution on $\mathbf{R}_+$ with density with respect to Lebesgue measure

$$t \mapsto \frac{\theta^p}{\Gamma(p)}\, e^{-\theta\, t}\, t^{p-1}\, \mathbf{1}_{\{t>0\}}.$$

We also denote, for $p, q > 0$, Beta(p, q) the Beta distribution on $[0,1]$ whosedensity with respect to Lebesgue measure is given by

$$t \mapsto \frac{1}{B(p,q)}\, t^{p-1}\, (1-t)^{q-1}\, \mathbf{1}_{\{t\in]0,1[\}}$$

whee $B(p,q)$ is a normalising constant.

Assume that X and Y are 2 independent random variables following $\Gamma(p,\theta)$ and $\Gamma(q,\theta)$ respectively.

1. Show that $X+Y$ and $\frac{X}{X+Y}$ are independent and distributed according to a $\Gamma(p+q,\theta)$ et Beta(p, q), respecively. Deduce the expression of $B(p,q)$ in terms of the Γ functions.
2. Show that the distribution of X/Y is independent of θ and determine its density.

Exercice 2.3 ***Chi-square and Student*** Let $(X_n)_n$ be independent Gaussian random variables $\mathcal{N}(0,1)$. Define for all $n \in \mathbb{N}^*$

$$Y_n = \sum_{j=1}^{n} X_j^2, \quad T'_n = \frac{X_{n+1}}{\sqrt{Y_n}}, \quad T_n = \sqrt{n}T'_n.$$

1. Chi-square : Show by induction that the distribution of Y_n is absolutely continuous wrt Lebesgue measure with density f_{Y_n} :

$$\forall y \in \mathbf{R} \quad f_{Y_n}(y) = \mathbf{1}_{y>0} \frac{2^{-n/2}}{\Gamma(n/2)} y^{n/2-1} \exp\left(-\frac{y}{2}\right)$$

2. Compute the expectation and the variance of Y_n.
3. Student : Compute the density of the distribution of T'_n with respect to Lebesgue measure and show that the distribution of T_n has a density wrt Lebesgue measure given by

$$\forall y \in \mathbf{R} \quad f_{T_n}(y) = \frac{\Gamma((n+1)/2)}{\sqrt{n\pi}\Gamma(n/2)} \left(1+\frac{y^2}{n}\right)^{-(n+1)/2}$$

4. When is T_n integrable ? T_n ?

Exercice 2.4

Consider n random variables $X_1, X_2, \ldots, X_n$ iid from a distribution with cdf F. We assume that F is strictly monotone and continuous.

Give the distribution of $Z = -2\sum_{i=1}^{n} \log(F(X_i))$.

Exercice 2.5 Let (X,Y) be a couple of real random variable with density

$$f_{(X,Y)}(x,y) = \frac{3\sqrt{3}}{4\pi}\exp\{-3/2(x^2+y^2-xy)\}\,.$$

Compute $f_X(x)$, $f_Y(y)$, $f_{X|Y=y}(x)$ and $\mathbb{E}(X|Y)$.

Exercice 2.6 Define

$$X_n = \begin{cases} n^{1/2} & U \leq 1/n \\ 0 & \text{otherwise} \end{cases}.$$

where $U \sim \mathcal{U}([0,1])$.

Study the convergence L^2, $\mathbb{P}$ et $\mathcal{L}$ of the sequence (X_n).

Exercice 2.7

1. Compute the moment-generating function of the uniform law on $\{2,\dots,12\}$
2. Let X_1 and X_2 be independent random variables with values in $\{1,\dots,6\}$. Studying the roots of the polynomial $G_{X_1}G_{X_2}$ (where G_{X_i} is the moment-generating function of X_i), show that the law of the random variable X_1+X_2 cannot be the uniform law on $\{2,\dots,12\}$ (we should find that for every real s and every $i \in \{1,2\}$, there is a polynomial ϕ_i of odd degree with real coefficients such that for every strictly positive real s, we have $G_{X_i}(s) = s\phi_i(s)$).
3. Can we build two 'loaded dices' (two independent random variables) such that every outcome of their sum between 2 and 12 is equally likely ?

Exercice 2.8 Let $U_1,\dots,U_n$ be an n-sample of a $\mathcal{U}([a,b])$. Define $M_n = \max(U_1,\dots,U_n)$ and $N_n = \min(U_1,\dots,U_n)$.

1. Show that M_n converges in probability to b.
2. Show that N_n converges in L^2 norm toward a.
3. Show that $n(b-M_n)$ converges in distribution to a random variable that follow an exponential law with parameter $\frac{1}{b-a}$.

3 Standard distributions

Exercice 3.1 Consider a sequence X_i of iid Poisson variables with parameter $\lambda_i > 0$ denoted $\mathcal{P}(\lambda_i)$:

$$P[X_i = k] = e^{-\lambda_i}\frac{\lambda_i^k}{k!}, k \in \mathbb{N}$$

1. Show that, for all $p \in \mathbb{N}^*$

$$\sum_{i=1}^{p} X_i \sim \mathcal{P}(\sum_{i=1}^{p} \lambda_i)$$

2. For every $n \geq 1$ take the same iid sample $X_1, ..., X_n$ as above and denote N_n the number of zero realisations among those X_i's. Deduce the distribution of N_n.

Exercice 3.2 Let (X, Y) be a couple of real random variable with density

$$f_{(X,Y)}(x, y) = \frac{1}{756}(x^2 + xy)\mathbb{I}_{0<x<6}\mathbb{I}_{0<y<6} \,.$$

Compute $f_X(x)$, $f_{Y|X=x}(y)$, $\mathbb{P}(X < Y)$ and $\mathbb{E}(Y|X = x)$.

Exercice 3.3 Let $(U_n)_{n\geq 0}$ be a sequence of independent and identically distributed random variables with Bernoulli law of parameter $p \in]0, 1[$, for every positive integer n, we denote $Y_n = U_n U_{n+1}$ and $S_n = Y_1 + \cdots + Y_n$

1. For every integer $n \geq 0$, what is the law of Y_n ?
2. At what condition on the integers n and m such that $0 \leq n < m$ we get that the random variables Y_n and Y_m are independent ?
3. Compute $\mathbb{E}[Y_n Y_m]$, then compute $\mathbb{E}\left[\frac{S_n}{n}\right]$.
4. Show that there exists a real constant C such that $\mathbb{V}[S_n] \leq Cn$.
5. Establish that the sequence $\left(\frac{S_n}{n}\right)_{n\geq 0}$ converges in probability toward a constant (and specify this limit).

Exercice 3.4 Consider $a > 0$ and $\lambda > 0$. Define the distribution $\mathcal{G}a(a, \lambda)$ as associated with the following density function :

$$f_{a,\lambda}(x) = \frac{\lambda^a}{\Gamma(a)} \exp(-\lambda x) x^{a-1} \mathbb{I}_{x\geq 0} \,.$$

1 Check that this function defines a probability density function.

2 Derive the expectation of this law.

Let $V_1, \ldots, V_n$ be independent random variables with distribution $\mathcal{E}(\lambda)$.

3 Show, through recursion, that the law of the sum $V_1 + \cdots + V_n$ is the $\mathcal{G}a_{n,\lambda}$ law.

Let X and Y be two independent random variables with distribution $\mathcal{G}a(a, \lambda)$.

4 Derive the law of λX.

5 Show that the random variables $X+Y$ and $\frac{X}{Y}$ are independent and compute their laws.

6 Show that the random variables $X+Y$ and $\frac{X}{X+Y}$ are independent. Compute the law of the random variable $\frac{X}{X+Y}$.

Let $X, Y : \Omega \to \mathbf{R}$ be two independent random variables with law $\mathcal{G}a_{a,\lambda}$ and $\mathcal{G}a_{b,\lambda}$ respectively.

7 Derive the law of the random variable $X+Y$.

Let $Z_1, \ldots, Z_n$ be iid random variables with $\mathcal{N}(0,1)$ distribution.

8 Show that the random variable Z_1^2 follow a $\mathcal{G}a\left(\frac{1}{2}, \frac{1}{2}\right)$ distribution.

9 Show that the law of the random variable $Z_1^2 + \cdots + Z_n^2$ follow a $\mathcal{G}a\left(\frac{n}{2}, \frac{1}{2}\right)$ distribution, also called the $\chi^2(n)$ distribution.

Exercice 3.5

1. Let X and Y be two independent random variables with law $\mathcal{N}(m_1, \sigma_1^2)$ and $\mathcal{N}(m_2, \sigma_2^2)$ respectively, what is the law of $X+Y$?
2. Let $X_1, \ldots, X_n$ be iid random variables with distribution $\mathcal{N}(m, \sigma^2)$, what is the law of the random variable $\overline{X_n} = \frac{X_1 + \cdots + X_n}{n}$?
3. Show that the random variable $\frac{\sqrt{n}}{\sigma}(\overline{X_n} - m)$ follow the $\mathcal{N}(0,1)$ law.
4. Define $\alpha \in]0,1[$, show that there exists a unique positive real number ϕ_α such that,

$$\int_{-\phi_\alpha}^{\phi_\alpha} \frac{\exp(-\frac{x^2}{2})}{t} \sqrt{2\pi} dx = 1 - \alpha$$

5. Derive that exists an interval $I_\alpha = [m-t, m+t]$ with a real t such that $\mathbb{P}(\overline{X_n} \in I_\alpha) = 1 - \alpha$.
6. Show that for every real strictly positive number ϵ, we have that

$$\lim_{n \to +\infty} \mathbb{P}(|\overline{X_n} - m| > \epsilon) = 0 \,.$$

Exercice 3.6 Consider a sequence $(X_n)_{n \geq 1}$ of real random variables. For $n \in \mathbb{N}^*$, X_n follow the exponential law with parameter n. Define

$$Y_n = \sin\left([X_n]\frac{\pi}{2}\right),$$

where $[X_n]$ is the integer part of X_n.

1. Find the distribution of the random variable Y_n, and compute $\mathbb{E}(Y_n)$.

2. Show that the sequence $(Y_n)_{n\geq 1}$ converges in distribution toward a constant random variable Y ; specify the law of Y .
3. Check the convergence in probability of the sequence $(Y_n)_{n\geq 1}$.

Exercice 3.7 1. Derive the characteristic function of the uniform law in $[-1, 1]$.

2. For every n, define the random variable X_n with

$$\mathbb{P}(X_n = 1/2^n) = \mathbb{P}(X_n = -1/2^n) = \frac{1}{2}$$

and suppose that X_n are mutually independent. Define $S_n = \sum_{i=1}^{n} X_i$. Show that (S_n) converges in distribution toward a random variable S and precise its law.

Exercice 3.8 Normalisation of the asymptotic variance.
Let (X_n) be a sequence of i.i.d. random variables with law $\mathbb{P}$. Suppose that $\mathbb{E}(X_1^2) < \infty$ such that the Central Limit Theorem applies :

$$\sqrt{n}(\overline{X}_n - \mu) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma^2)$$

for $\sigma^2 = \text{Var}\,(X_1) > 0$.

1. Making use of Slutsky's theorem, derive a sequence of random variables (a_n) function of $X_1, \ldots, X_n$ such that

$$\sqrt{n/a_n}(\overline{X}_n - \mu) \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1).$$

2. Suppose now that σ^2 is a function of μ. Applying Slutsky's method, derive the function ϕ such that $\sqrt{n}(\phi(\overline{X}_n) - \phi(\mu)) \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1)$.
3. Find (a_n) et ϕ in the particular cases $\mathbb{P} = \mathcal{B}(p)$ with $0 < p < 1/2$ and $\mathbb{P} = \mathcal{E}(\lambda)$ with $\lambda > 0$.

Exercice 3.9 A system works using two different machines serially. The life expectancy X_1 and X_2 of the two machines follow the exponential distribution with parameters λ_1 and λ_2. The random variables X_1 and X_2 are supposed independent.

1. Show that

$$X \sim \mathcal{E}(\lambda) \text{ iff } \forall x > 0,\ \mathbb{P}(X > x) = \exp(-\lambda x)$$

2. Compute the probability that the system do not break down before time t. Infer the law of the survival time Z of the system.
3. Compute the probability that a breakdown is due to machine 1.
4. Let $I = 1$ if the breakdown is due to machine 1, $I = 0$ otherwise. Compute $\mathbb{P}(Z > t, I = \delta)$, for every $t \geq 0$ and $\delta \in \{0, 1\}$. Show that Z et I are independent.
5. Suppose that we have n identical systems that function independently and we observe their survival times $Z_1, ...Z_n$. Write the corresponding parametric model.