

PC 3 – Random vectors & Convergence

Gamma distribution

Exercise 1

(Gamma distribution). One says that X has Gamma distribution with parameters $p > 0$ et $\theta > 0$, denoted by $\gamma(p, \theta)$, if its density is given by

$$f(x) = \frac{\theta^p}{\Gamma(p)} \exp(-\theta x) x^{p-1} \mathbb{1}_{[0, +\infty[}(x)$$

The associated characteristic function is given by

$$\Phi_X(t) = \frac{1}{(1 - it/\theta)^p}, \quad t \in \mathbb{R}.$$

Here $\Gamma(\cdot)$ denotes the Gamma function defined as

$$\forall \alpha > 0, \quad \Gamma(\alpha) = \int_0^\infty x^{\alpha-1} \exp(-x) dx, \quad \Gamma(\alpha + 1) = \alpha \Gamma(\alpha), \quad \Gamma(1/2) = \sqrt{\pi}$$

1. Compute $\mathbb{E}[X^k]$ for $k \geq 1$. Deduce that $\mathbb{E}[X] = p/\theta$ and $\text{Var}(X) = p/\theta^2$.
2. Let $a > 0$. Show that $X/a \sim \gamma(p, a\theta)$.
3. Let X and Y be two independent random variables with Gamma distribution $\gamma(p_1, \theta)$ and $\gamma(p_2, \theta)$, respectively. Show that $X + Y \sim \gamma(p_1 + p_2, \theta)$.
4. Let Z have standard normal distribution $\mathcal{N}(0, 1)$. What is the distribution of Z^2 ?
5. Let X_1, \dots, X_n be n i.i.d. random variables aléatoires with exponential distribution $\text{Exp}(\theta)$. Determine the distribution of the sum $S_n = X_1 + \dots + X_n$. Compute $\mathbb{E}[S_n]$ and $\text{Var}(S_n)$.
6. Let X_1, \dots, X_n be n i.i.d. random variables aléatoires with standard normal distribution $\mathcal{N}(0, 1)$. Determine the distribution of the sum $S'_n = X_1^2 + \dots + X_n^2$. Compute $\mathbb{E}[S'_n]$ and $\text{Var}(S'_n)$.

Solution 1 1. We have:

$$\begin{aligned} \mathbb{E}[X^k] &= \frac{\theta^p}{\Gamma(p)} \int_0^\infty x^k e^{-\theta x} x^{p-1} dx \\ &= \frac{\theta^p}{\Gamma(p)} \frac{\Gamma(k+p)}{\theta^{k+p}} \\ &= \frac{(k+p-1) \times \dots \times p}{\theta^k} \end{aligned}$$

Therefore,

$$\mathbb{E}[X] = p/\theta$$

and

$$\text{Var}(X) = p/\theta^2$$

2. For $a > 0$, we have:

$$\begin{aligned}\Phi_{X/a}(t) &= \mathbb{E}[e^{itX/a}] \\ &= \mathbb{E}[e^{i\frac{t}{a}X}] \\ &= \frac{1}{(1 - i\frac{t}{a\theta})^p}\end{aligned}$$

Therefore $X/a \sim \gamma(p, a\theta)$ identifying random variables from characteristic functions.

3. We have

$$\begin{aligned}\Phi_{X+Y}(t) &= \Phi_X(t)\Phi_Y(t) \\ &= \frac{1}{(1 - i\frac{t}{a\theta})^{p_1}} \frac{1}{(1 - i\frac{t}{a\theta})^{p_2}} \\ &= \frac{1}{(1 - i\frac{t}{a\theta})^{p_1+p_2}}\end{aligned}$$

Which concludes.

4. The distribution of Z is exactly a Chi-squared distribution χ_1^2 with one degree of liberty:

$$f_{Z^2}(x) = \frac{1}{\sqrt{2\pi x}e^{-x/2}}$$

5. For X following an exponential distribution,

$$\Phi_X(t) = \frac{1}{1 - it/\theta}$$

Therefore,

$$\Phi_{S_n}(t) = \left(\frac{1}{1 - it/\theta} \right)^n$$

And finally $S_n \sim \gamma(n, \theta)$ with same expected value and variance than in 1.

6.

Random vectors

Exercise 2

Denote

$$f(x, y) = ce^{-x} \mathbb{1}_{|y| \leq x}$$

1. Find c such that f is a probability density function of a pair (X, Y) of random variables.
2. Compute the marginal distributions of X and Y .
3. Conclude on the independence of X and Y .

Solution 2 1. We have:

$$\begin{aligned}
f \text{ is a density} &\iff \int_{\mathbb{R}^2} ce^{-x} \mathbb{1}_{|y| \leq x} d(x, y) = 1 \\
&\iff \int_{\mathbb{R}} \int_{\mathbb{R}^+} ce^{-x} \mathbb{1}_{|y| \leq x} dx dy = 1 \\
&\iff \int_{\mathbb{R}^+} 2xce^{-x} dx = 1 \\
&\iff 2c \int_{\mathbb{R}^+} xe^{-x} dx = 1 \\
&\iff 2c = 1 \iff c = \frac{1}{2}
\end{aligned}$$

2. Moreover,

$$f_X(x) = xe^{-x}$$

And

$$f_Y(y) = \frac{1}{2}e^{-y}$$

3. We finally have:

$$f(x, y) \neq f_X(x)f_Y(y)$$

and the random variables therefore are not independent.

Exercise 3

Let X and Y be two random variables taking their values in \mathbb{N} . Consider the joint probability mass function of (X, Y) given by

$$\mathbb{P}[(X = i) \cap (Y = j)] = \frac{a}{2^{i+j}}, i, j \in \mathbb{N}, a \in \mathbb{R}.$$

1. Compute a .
2. Give the marginal distributions of X and Y .
3. Are X and Y independent?

Solution 3 1. We have:

$$\sum_{i,j=0}^{\infty} \frac{a}{2^{i+j}} = a \left(\sum_{i=0}^{\infty} \frac{1}{2^i} \right)^2 = a \cdot 2 \cdot 2 = 4a$$

Therefore, $4a = 1$ and finally $a = \frac{1}{4}$.

2. We have:

$$\begin{aligned}
\mathbb{P}[X = i] &= \sum_{j=0}^{\infty} \mathbb{P}[(X = i) \cap (Y = j)] \\
&= \sum_{j=0}^{\infty} \frac{1}{4 \cdot 2^i \cdot 2^j} \\
&= \frac{1}{2^{i+1}}
\end{aligned}$$

In the same way:

$$\mathbb{P}[Y = i] = \frac{1}{2^{i+1}}$$

3. We have:

$$\mathbb{P}[(X = i) \cap (Y = j)] = \frac{1}{2^{i+j+2}} = \left(\frac{1}{2^{i+1}}\right) \left(\frac{1}{2^{j+1}}\right) = \mathbb{P}[X = i] \mathbb{P}[Y = j]$$

And the random variables are therefore independent.

Exercise 4

Denote

$$f(x, y) = a(x^2 + y^2) \mathbb{1}_{(x, y) \in [-1, 1]^2}.$$

1. Find a such that f is a probability density. We denote (X, Y) the pair of random variables with joint distribution f .
2. Compute the marginal distributions of X and Y .
3. Compute the covariance of X and Y .
4. Are X and Y independent?

Solution 4 1. We have

$$\int_{[-1, 1]^2} x^2 + y^2 = \pi$$

Therefore $a = \frac{1}{\pi}$.

2.

Exercise 5

Let $\mathbf{X} = (X_1, X_2, X_3)$ be a random vector with the following covariance matrix

$$\text{Cov}(\mathbf{X}) = \begin{pmatrix} 2 & 1 & 3 \\ 1 & 5 & 6 \\ 3 & 6 & 9 \end{pmatrix}$$

1. Give the variance of X_2 and the covariance between X_1 and X_3 .
2. Compute the variance of $Z = X_3 - \alpha_1 X_1 - \alpha_2 X_2$ for $\alpha_1, \alpha_2 \in \mathbb{R}$.
3. Deduce that X_3 is almost surely a linear combination of X_1 and X_2 .
4. More generally, let \mathbf{Y} be a random vector. Give a necessary and sufficient condition on the covariance matrix of \mathbf{Y} ensuring that one of the components of \mathbf{Y} is almost surely a linear combination of the components of \mathbf{Y} .

Solution 5

Convergence

Exercise 6

Let $\{X_i\}_{i \geq 0}$ be a sequence of i.i.d. Bernoulli variables with parameter θ .

1. Show that $\sqrt{n}(\bar{X}_n - \theta) \xrightarrow{d} \mathcal{N}(0, \theta(1 - \theta))$, where $\bar{X}_n = n^{-1} \sum_{i=1}^n X_i$.
2. Show that $\bar{X}_n(1 - \bar{X}_n) \xrightarrow{P} \theta(1 - \theta)$.
3. Show that $\sqrt{n}(\bar{X}_n - \theta)^2 \xrightarrow{P} 0$.
4. Determine the limit distribution of $\sqrt{n}(\bar{X}_n(1 - \bar{X}_n) - \theta(1 - \theta))$.

Solution 6 1. Par le TCL, on a $\sqrt{n}(\bar{X}_n - \theta) = \sqrt{n}(\bar{X}_n - \mathbb{E}[X_1]) \xrightarrow{d} \mathcal{N}(0, \text{Var}(X_1)) = \mathcal{N}(0, \theta(1 - \theta))$.
 2. Par la LGN, on a $\bar{X}_n \xrightarrow{P} \mathbb{E}[X_1] = \theta$. La fonction $h(x) = x(1 - x)$ étant continue, on obtient par le théorème de continuité, $\bar{X}_n(1 - \bar{X}_n) = h(\bar{X}_n) \xrightarrow{P} h(\theta) = \theta(1 - \theta)$.

3. On a

$$\sqrt{n}(\bar{X}_n - \theta)^2 = \underbrace{\sqrt{n}(\bar{X}_n - \theta)}_{\xrightarrow{d} \mathcal{N}(0, \theta(1 - \theta))} \underbrace{(\bar{X}_n - \theta)}_{\xrightarrow{P} 0} \xrightarrow{d} 0 \times \mathcal{N}(0, \theta(1 - \theta)) = 0$$

La convergence en loi vers une constante est équivalente à la convergence en probabilité, d'où le résultat.

4. On écrit

$$\begin{aligned} \sqrt{n}(\bar{X}_n(1 - \bar{X}_n) - \theta(1 - \theta)) &= \sqrt{n}((\bar{X}_n - \theta)(1 - \bar{X}_n) + \theta(1 - \bar{X}_n) - \theta(1 - \theta)) \\ &= \sqrt{n}((\bar{X}_n - \theta)(1 - \bar{X}_n) - \theta(\bar{X}_n - \theta)) \\ &= \underbrace{\sqrt{n}(\bar{X}_n - \theta)}_{\xrightarrow{d} \mathcal{N}(0, \theta(1 - \theta))} \underbrace{(1 - \bar{X}_n - \theta)}_{\xrightarrow{P} 1 - 2\theta} \\ &\xrightarrow{d} (1 - 2\theta)\mathcal{N}(0, \theta(1 - \theta)) = \mathcal{N}(0, (1 - 2\theta)^2 \theta(1 - \theta)), \end{aligned}$$

par le lemme de Slutsky.

Exercise 7

Let $(X_n)_{n \geq 1}$ be a sequence of i.i.d. square-integrable random variables with mean m and variance $\sigma^2 > 0$. Denote $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$ and $\hat{\sigma}_n^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2$.

1. Show that $\hat{\sigma}_n^2$ converges in probability to σ^2 as $n \rightarrow \infty$.
2. Determine the limit distribution of $\sqrt{n}(\bar{X}_n - m) / \hat{\sigma}_n$.

Solution 7 Commençons par étudier le comportement limite de $\hat{\sigma}_n^2$ quand $n \rightarrow +\infty$.

$$\begin{aligned} (n - 1)\hat{\sigma}_n^2 &= \sum_{k=1}^n (X_k - \bar{X}_n)^2 \\ &= \sum_{k=1}^n (X_k - m)^2 + 2 \sum_{k=1}^n (X_k - m)(m - \bar{X}_n) + n(m - \bar{X}_n)^2 \\ &= \sum_{k=1}^n (X_k - m)^2 - n(m - \bar{X}_n)^2. \end{aligned}$$

Donc

$$\begin{aligned}\frac{n-1}{n}\hat{\sigma}_n^2 &= \frac{1}{n} \sum_{k=1}^n (X_k - m)^2 - (m - \bar{X}_n)^2 \\ &\xrightarrow{p.s.} \mathbb{E}[(X_1 - m)^2] - 0 = \text{Var}(X_1) =: \sigma^2,\end{aligned}$$

où la limite est donnée par la loi des grands nombres. Par suite, $\hat{\sigma}_n \rightarrow \sigma$ presque sûrement. Notons $Z_n = \sqrt{n}(\bar{X}_n - m)$ qui converge en loi, d'après le théorème limite central vers une variable aléatoire gaussienne $Z \sim \mathcal{N}(0, \sigma^2)$. D'après le lemme de Slutsky, le couple $(Z_n, \hat{\sigma}_n^{-1})$ converge en loi vers (Z, σ^{-1}) . En particulier, la fonction produit étant continue, $\frac{Z_n}{\hat{\sigma}_n} \xrightarrow{d} Z/\sigma \sim \mathcal{N}(0, 1)$

Exercise 8

(Poisson model). Let (X_1, \dots, X_n) be an i.i.d. sample from the Poisson distribution with unknown parameter $\lambda > 0$. Denote $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$.

1. Show that \bar{X}_n is an unbiased estimator of λ , that is $\mathbb{E}[\bar{X}_n] = \lambda$.
2. Show that \bar{X}_n converges in probability to λ when n tends to infinity.
3. Determine the limit distribution of $\sqrt{n}(\bar{X}_n - \lambda) / \sqrt{\bar{X}_n}$.
4. Find an appropriate function g such that $\sqrt{n}(g(\bar{X}_n) - g(\lambda)) \xrightarrow{d} \mathcal{N}(0, 1)$.

Solution 8 1. On rappelle, pour $X \sim \mathcal{P}(\lambda)$, $\lambda > 0$, on a $\mathbb{E}[X] = \text{Var}[X] = \lambda$. Alors l'estimateur \bar{X}_n est alors sans biais ($\mathbb{E}[\bar{X}_n] = \lambda$), consistant en vertu de la LFGN ($\bar{X}_n \rightarrow \mathbb{E}[X_1] = \lambda$ p.s), et enfin, \bar{X}_n est asymptotiquement normal par le TCL:

$$\sqrt{n}(\bar{X}_n - \lambda) = \sqrt{n}(\bar{X}_n - \mathbb{E}[X_1]) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \text{Var}[X_1]) = \mathcal{N}(0, \lambda), \quad \text{lorsque } n \rightarrow \infty$$

2. En utilisant la question 1 et le lemme de Slutsky, on obtient

$$\sqrt{n} \left(\frac{\bar{X}_n - \lambda}{\sqrt{\bar{X}_n}} \right) = \underbrace{\sqrt{n} \left(\frac{\bar{X}_n - \lambda}{\sqrt{\lambda}} \right)}_{\xrightarrow{\mathcal{L}} \mathcal{N}(0, 1)} \underbrace{\frac{\sqrt{\lambda}}{\sqrt{\bar{X}_n}}}_{\xrightarrow{P} \frac{\sqrt{\lambda}}{\sqrt{\mathbb{E}[X_1]}=1}} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1).$$

3. D'après la delta méthode, pour toute fonction g continument dérivable sur \mathbb{R}_+ , on a

$$\sqrt{n}(g(\bar{X}_n) - g(\lambda)) \xrightarrow{\mathcal{L}} \mathcal{N}(0, (g'(\lambda))^2 \text{Var}[X])$$

Nous cherchons donc une fonction g telle que la variance limite vaut 1. Ce qui veut dire

$$(g'(\lambda))^2 \text{Var}[X] = 1 \Leftrightarrow (g'(\lambda))^2 = \frac{1}{\lambda}$$

On peut alors choisir $g(u) = 2\sqrt{u}$ avec dérivée $g'(u) = 1/\sqrt{u}$ et on obtient

$$\sqrt{n} \left(2\sqrt{\bar{X}_n} - 2\sqrt{\lambda} \right) \xrightarrow{\mathcal{L}} \mathcal{N} \left(0, \left(\frac{1}{\sqrt{\lambda}} \right)^2 \lambda \right) = \mathcal{N}(0, 1).$$

Exercise 9

Define the random variable

$$Y = \mathbb{1}\{\theta > X\}$$

where $\theta \in \mathbb{R}$ and X is a random variable with standard normal distribution $\mathcal{N}(0, 1)$. We observe a sample Y_1, \dots, Y_n of i.i.d. realizations of Y and suppose that parameter θ is unknown. Denote by Φ the cumulative distribution function of the standard normal distribution $\mathcal{N}(0, 1)$. An estimator $\hat{\theta}_n$ of θ is given by

$$\hat{\theta}_n = \Phi^{-1}(\bar{Y}_n)$$

where $\bar{Y}_n = \frac{1}{n} \sum_{i=1}^n Y_i$

1. Determine the distribution of Y .
2. Study the convergence in probability of $\hat{\theta}_n$ towards θ when n tends to infinity.
3. Study the limit distribution of $\sqrt{n}(\hat{\theta}_n - \theta)$.

Solution 9 1. Comme Y prend ses valeurs dans $\{0, 1\}$, Y suit une loi de Bernoulli avec paramètre $\mathbb{P}(Y = 1) = \mathbb{P}(\theta > X) = \Phi(\theta)$

2. Puisque $\frac{1}{n} \sum_{i=1}^n Y_i \rightarrow \mathbb{E}[Y_1] = \Phi(\theta)$ p.s et Φ^{-1} est une fonction continue, on a $\hat{\theta}_n = \Phi^{-1}(\frac{1}{n} \sum_{i=1}^n Y_i) \rightarrow \Phi^{-1}(\Phi(\theta)) = \theta$ p.s. Donc $\hat{\theta}_n$ est consistant pour θ .

3. En vertu du TCL (car $\mathbb{E}[Y_1^2] < \infty$), on a $\sqrt{n}(\frac{1}{n} \sum_{i=1}^n Y_i - \Phi(\theta)) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \text{Var}([Y_1]) = \mathcal{N}(0, \Phi(\theta)(1 - \Phi(\theta)))$. La fonction $\Phi^{-1}(\theta)$ est continument dérivable avec dérivée $(\Phi^{-1})'(\theta) = 1/\varphi(\Phi^{-1}(\theta))$. On obtient par la delta-méthode

$$\begin{aligned} \sqrt{n}(\hat{\theta}_n - \theta) &= \sqrt{n} \left(\Phi^{-1} \left(\frac{1}{n} \sum_{i=1}^n Y_i \right) - \Phi^{-1}(\Phi(\theta)) \right) \xrightarrow{\mathcal{L}} \mathcal{N}(0, ((\Phi^{-1})'(\Phi(\theta)))^2 \Phi(\theta)(1 - \Phi(\theta))) \\ &= \mathcal{N} \left(0, \frac{\Phi(\theta)(1 - \Phi(\theta))}{\varphi^2(\theta)} \right) \end{aligned}$$