MSIAM M1: Probability and Statistics

Solutions to [the most difficult] exercises

1 TD 1

Exercise 1.2. *Provide an example of asymmetric density with $\alpha = 0$.*

Proof. Consider a discrete random variable X, such that $\mathbb{P}(X = 2) = \mathbb{P}(X = -1) = \frac{1}{4}, \mathbb{P}(X = \sqrt{7}) = \frac{1}{4\sqrt{7}}$, and $\mathbb{P}(X = 0) = \frac{1}{2} - \frac{1}{4\sqrt{7}}$. First, note that this r.v. is asymmetric. Indeed, $\mathbb{E}[X] = \frac{1}{2} - \frac{1}{4} - \sqrt{7}\frac{1}{4\sqrt{7}} = 0$. However,

$$F_X(0) = \frac{3}{4} \quad \neq \quad 1 - F_X(0) = \frac{1}{4}.$$

Now, compute the third moment of X:

$$\mathbb{E}[X^3] = 2 - \frac{1}{4} - 7\sqrt{7}\frac{1}{4\sqrt{7}} = 0.$$

Thus, the skewness parameter is 0. □

Exercise 1.4. *Let (ξ_n) and (η_n) be two sequences of r.v. Prove the following statements:*

1°. *If $a \in \mathbb{R}$ is a constant, then when $n \to \infty$:*

$$\xi_n \xrightarrow{D} a \quad \Leftrightarrow \quad \xi_n \xrightarrow{P} a$$

2°. *(**Slutsky's theorem**.) If $\xi_n \xrightarrow{D} a$ and $\eta_n \xrightarrow{D} \eta$ when $n \to \infty$, where $a \in \mathbb{R}$ and η is a random variable, then*

$$\xi_n + \eta_n \xrightarrow{D} a + \eta, \quad \text{as } n \to \infty.$$

Show that if a is a general random variable, these two relations do not hold (construct a counterexample).

3°. *If $\xi_n \xrightarrow{D} a$ and $\eta_n \xrightarrow{D} \eta$ when $n \to \infty$, where $a \in \mathbb{R}$ and η is a random variable, then*

$$\xi_n \eta_n \xrightarrow{D} a\eta, \quad \text{as } n \to \infty.$$

Would this result continue to hold if we suppose that a is a general random variable?

Proof. 1°. (Anatolii's proof) By the definition of convergence in probability, we want to show that $\lim_{n\to\infty} \mathbb{P}(|\xi_n - a| \geq \varepsilon) = 0$. Let us consider two continuous functions, defined as follows:

$$f_\varepsilon(x) = \begin{cases} 1, & x \in \left(-\infty, a - \frac{3\varepsilon}{2}\right] \bigcup \left[a + \frac{3\varepsilon}{2}, \infty\right) \\ 0, & x \in [a - \varepsilon, a + \varepsilon], \end{cases}$$

$$g_\varepsilon(x) = \begin{cases} 1, & x \in (-\infty, a - \varepsilon] \bigcup [a + \varepsilon, \infty) \\ 0, & x \in \left[a - \frac{\varepsilon}{2}, a + \frac{\varepsilon}{2}\right], \end{cases}$$

Functions f and g are "smoothed" versions of an indicator function, which check if x contains outside of balls of radius ε and $\frac{\varepsilon}{2}$ respectively. Then,

$$\int_{-\infty}^{\infty} f_\varepsilon(x) dF_n(x) \leq \mathbb{P}(|\xi_n - a| \geq \varepsilon) \leq \int_{-\infty}^{\infty} g_\varepsilon(x) dF_n(x),$$

where F_n is a cumulative distribution function of ξ_n. Both left and right sides converge to 0 as $n \to \infty$, since $f_\varepsilon(a) = g_\varepsilon(a) = 0$. Consequently, $\lim_{n\to\infty} \mathbb{P}(|\xi_n - a| \geq \varepsilon) = 0$. It gives the result.

1°. (Alternative proof) By the definition of convergence in probability, we want to show that $\lim_{n\to\infty} \mathbb{P}(|\xi_n - a| \geq \varepsilon) = 0$. Fix some $\varepsilon > 0$. Denote by $B_\varepsilon(a)$ be the open ball of radius ε around point a, and $\bar{B}_\varepsilon(a)$ its complement. Then

$$\mathbb{P}(|\xi_n - a| \geq \varepsilon) = \mathbb{P}\left(\xi_n \in \bar{B}_\varepsilon(a)\right)$$

Then we can observe that:

$$\lim_{n\to\infty} \mathbb{P}(|\xi_n - a| \geq \varepsilon) \leq \limsup_{n\to\infty} \mathbb{P}(|\xi_n - a| \geq \varepsilon) =$$
$$\limsup_{n\to\infty} \mathbb{P}\left(\xi_n \in \bar{B}_\varepsilon(a)\right) = \mathbb{P}\left(a \in \bar{B}_\varepsilon(a)\right) = 0$$

By definition, it means that the sequence converges to a in probability.

2°a. It is a direct consequence of 1°. What we need to show is the following:

$$\mathbb{P}\left(\xi_n + \eta_n \leq t\right) \longrightarrow \mathbb{P}\left(a + \eta \leq t\right) = \mathbb{P}\left(\eta \leq t - a\right).$$

Consider the following event:

$$\{\xi_n + \eta_n - a \le t - a\} = \\ \{\xi_n + \eta_n - a \le t - a, |\xi_n - a| \le \varepsilon\} \bigcup \{\xi_n + \eta_n - a \le t - a, |\xi_n - a| > \varepsilon\}$$

Note that the probability of the second event tends to 0 as $n \to \infty$ (because of the result from 1°). Then, consider

$$\{\eta_n \le t - a - \varepsilon\} \subseteq \{\xi_n + \eta_n - a \le t - a, |\xi_n - a| \le \varepsilon\} \subseteq \{\eta_n \le t - a + \varepsilon\}$$

As a consequence,

$$\mathbb{P}\left(\eta_n \le t - a + \varepsilon\right) \longrightarrow \mathbb{P}\left(\eta \le t - a + \varepsilon\right) \to \mathbb{P}\left(\eta \le t - a\right) \text{ as } \varepsilon \to 0.$$

$2°b$. **(Counterexample)** Consider a sequence ξ_n of Bernoulli variables, such that $\mathbb{P}(\xi_n = 1) = \mathbb{P}(\xi_n = 0) = \frac{1}{2}$, and consider $\eta_n = 1 - \xi_n$. It is easy to see that $\xi_n \xrightarrow{D} \xi$ and $\eta_n \xrightarrow{D} \eta$, where ξ and η are again Bernoulli variables taking values 0 and 1 with equal probability. Obviously, $\xi_n + \eta_n = 1$ is not converging in law to the variable $\xi + \eta$, which is taking values $0, 1, 2$ with probabilities $\frac{1}{4}, \frac{1}{2}, \frac{1}{4}$ respectively.

3°. First note that

$$\xi_n \eta_n = (\xi_n - a)\eta_n + a\eta_n.$$

Note that $a\eta_n \xrightarrow{D} a\eta$ because $\forall a > 0, x \in \mathbb{R}, \quad \mathbb{P}\left(a\eta_n \le x\right) = \mathbb{P}\left(\eta_n \le \frac{x}{a}\right) \to \mathbb{P}\left(\eta \le \frac{x}{a}\right) = \mathbb{P}\left(a\eta \le x\right)$. Also, $\forall C < \infty$

$$\{|\eta_n(\xi_n - a)| > \varepsilon\} \subseteq \{|\eta_n| > C\} \bigcup \left\{|\xi_n - a| > \frac{\varepsilon}{C}\right\},$$

thus

$$\mathbb{P}\left(|\eta_n(\xi_n - a)| > \varepsilon\right) \le \mathbb{P}\left(|\eta_n| > C\right) + \mathbb{P}\left(|\xi_n - a| > \frac{\varepsilon}{C}\right).$$

Note that $\mathbb{P}\left(|\xi_n - a| > \frac{\varepsilon}{C}\right)$ converges to 0 as $n \to \infty$ due to 1°. Now it only remains to note that $\mathbb{P}\left(|\eta_n| > C\right) \to \mathbb{P}\left(|\eta| > C\right) < \frac{\delta}{4}$ for $C = C(\delta)$ sufficiently large.

$\square$

Exercise 1.9. *Let $\xi_1, \ldots, \xi_n$ be independent r.v. and let*

$$\xi_{\min} = \min(\xi_1, \ldots, \xi_n), \quad \xi_{\max} = \max(\xi_1, \ldots, \xi_n).$$

1. *Show that*

$$\mathbb{P}(\xi_{\min} \ge x) = \prod_{i=1}^{n} \mathbb{P}(\xi_i \ge x), \quad \mathbb{P}(\xi_{\max} < x) = \prod_{i=1}^{n} \mathbb{P}(\xi_i < x)$$

2. *Suppose, furthermore, that* $\xi_1, \dots, \xi_n$ *are identically distributed with uniform distribution* $\mathcal{U}[0,a]$. *Compute* $\mathbb{E}[\xi_{\min}], \mathbb{E}[\xi_{\max}], Var[\xi_{\min}], Var[\xi_{\max}]$

Proof. We consider the variable $\xi_{\max}$. The proof for $\xi_{\min}$ is identical.

1. Thanks to the independence of $\xi_1, \dots, \xi_n$ we have:

$$\mathbb{P}\left(\max_{i=1,\dots,n} \xi_i\right) = \mathbb{P}(\xi_1 < x, \dots, \xi_n < x) = \mathbb{P}(\xi_1 < x) \dots \mathbb{P}(\xi_n < x).$$

2. Since $\xi_i \sim \mathcal{U}[0,a]$, we have the following c.d.f. $F^*(x)$ of $\xi_{\max}$:

$$F^*(x) = \prod_{i=1}^{n} \frac{x}{a} = \left(\frac{x}{a}\right)^n \quad \forall 0 \le x \le a.$$

From that, we can easily derive the density of $\xi_{\max}$:

$$f^*(x) = \frac{nx^{n-1}}{a^n}.$$

Then, it is easy to obtain the expressions for the first and the second moments:

$$\mathbb{E}[\xi_{\max}] = \int_0^a xn\frac{x^{n-1}}{a^n}dx = \frac{an}{n+1}$$
$$\mathbb{E}[\xi_{\max}^2] = \int_0^a x^2 n\frac{x^{n-1}}{a^n}dx = \frac{a^2 n}{n+2}.$$

From that we compute the varaince:

$$Var[\xi_{\max}] = \mathbb{E}[\xi_{\max}^2] - (\mathbb{E}[\xi_{\max}])^2 = a^2\frac{n}{(n+1)^2(n+2)}.$$

It gives the statement.

□

Exercise 1.10. *Let* $\xi_1, \dots, \xi_n$ *be i.i.d. Bernoulli r.v. with*

$$\mathbb{P}(\xi_i = 0) = 1 - \lambda_i \Delta, \quad \mathbb{P}(\xi_i = 1) = \lambda_i \Delta,$$

where $\lambda_i > 0$ *and* $\Delta > 0$ *is small. Show that*

$$\mathbb{P}\left(\sum_{i=1}^{n} \xi_i = 1\right) = \left(\sum_{i=1}^{n} \lambda_i\right)\Delta + O(\Delta^2), \quad \mathbb{P}\left(\sum_{i=1}^{n} \xi_i > 1\right) = O(\Delta^2).$$

Proof. Note that

$$\{\xi_1 + \cdots + \xi_n = 1\} = \bigcup_{i=1}^{n} \{\xi_i = 1, \xi_{j \neq i} = 0\}.$$

Since all the variables are independent, the following holds:

$$\begin{aligned}
\mathbb{P}(\xi_1 + \cdots + \xi_n = 1) &= \sum_{i=1}^{n} \mathbb{P}\left(\xi_i = 1, \xi_{j \neq i} = 0\right) \\
&= \sum_{i=1}^{n} \mathbb{P}(\xi_i = 1) \prod_{j \neq i} \mathbb{P}\left(\xi_j = 0\right) = \sum_{i=1}^{n} \lambda_i \Delta \prod_{i \neq j} (1 - \lambda_j \Delta) \\
&= \sum_{i=1}^{n} \lambda_i \Delta + O(\Delta^2).
\end{aligned}$$

What about the second statement, note that

$$\mathbb{P}\left(\sum_{i=1}^{n} \xi_i > 1\right) = 1 - \mathbb{P}\left(\sum_{i=1}^{n} \xi_i = 0\right) - \mathbb{P}\left(\sum_{i=1}^{n} \xi_i = 1\right)$$

Let us compute the second term:

$$\mathbb{P}\left(\sum_{i=1}^{n} \xi_i = 0\right) = \prod_{i=1}^{n} \mathbb{P}(\xi_i = 0) = \prod_{i=1}^{n} \mathbb{P}(1 - \lambda_i \Delta) = 1 - \sum_{i=1}^{n} \lambda_i \Delta + O(\Delta^2).$$

That, in addition with the result for $\mathbb{P}(\xi_1 + \cdots + \xi_n = 1)$, gives the statement of the exercise. □

Exercise 1.11. 1. *Prove that* $\inf_{a \in \mathbb{R}} \mathbb{E}[(\xi - a)^2]$ *is attained for* $a = \mathbb{E}[\xi]$ *and so*

$$\inf_{a \in \mathbb{R}} \mathbb{E}[(\xi - a)^2] = Var[\xi].$$

2. *Let* ξ *be a nonnegative r.v. with c.d.f.* F *and finite expectation. Prove that*

$$\mathbb{E}[\xi] = \int_0^\infty (1 - F(x)) dx.$$

3. *Show, using the result from 2. that if* M *is the median of the c.d.f.* F *of* ξ*,*

$$\inf_{a \in \mathbb{R}} \mathbb{E}[|\xi - a|] = \mathbb{E}[|\xi - M|].$$

Proof. 1. Trivial (write an expression as a polynom depending on a, take the derivative w.r.t a, find zeroes).

2. Note that by the statement of the exercise, we have

$$\int_t^\infty x dF(x) \to 0 \quad t \to \infty.$$

As $\int_t^\infty x dF(x) \geq t(1-F(t))$, it implies that $t(1-F(t)) \to 0, t \to \infty$. Now we can use the integration by part formula, which results in:

$$\begin{aligned}\mathbb{E}[x] = \int_0^\infty x dF(x) &= -\int_0^\infty x d(1-F(x)) \\ &= -x((1-F(x))\Big|_0^\infty + \int_0^\infty (1-F(x))dx = \int_0^\infty (1-F(x))dx\end{aligned}$$

3. The previous formula actually gives the remaining result. First, we note that $\mathbb{P}(|\xi - a| > x) = \mathbb{P}(\xi > x + a) + \mathbb{P}(\xi > -x + a)$, thus

$$\begin{aligned}\mathbb{E}(|\xi - a|) = \int_0^\infty \mathbb{P}(|\xi - a| > x)\, dx &= \int_0^\infty \mathbb{P}(\xi > x+a)dx + \int_0^\infty \mathbb{P}(\xi > -x+a)\, dx \\ &= \int_a^\infty \mathbb{P}(\xi > z)dz - \int_a^\infty \mathbb{P}(\xi < z)dz\end{aligned}$$

The result can be obtained by computing the derivative w.r.t. a.

□

Exercise 1.12. *Let X_1 and X_2 be two independent r.v. with exponential distribution $\mathscr{E}(\lambda)$. Show that* $\min(X_1, X_2)$ *and* $|X_1 - X_2|$ *are r.v. with distributions, respectively,* $\mathscr{E}(2\lambda)$ *and* $\mathscr{E}(\lambda)$.

Proof. The first result is the direct consequence of Exercise 10. For the second result, consider a r.v. $\zeta = X_1 - X_2$. As both variables X_1 and X_2 are independent, we can use the Fubini theorem, and find the c.d.f of ζ as follows:

$$\begin{aligned}F_\zeta(z) = \mathbb{P}(\zeta < z) &= \int_{x\geq 0, y\geq 0, x-y\leq z} dF(x)dF(y) = \int_{x,y\geq 0} \mathbb{1}_{x-y\leq z} dF(x)dF(y) \\ &= \int_0^\infty dF(x)\left[\int_0^\infty \mathbb{1}_{y\geq x-z} dF(y)\right] \\ &= \int_0^\infty dF(x)\left[\mathbb{1}_{x-z\geq 0}\int_{x-z}^\infty dF(y) + \mathbb{1}_{x-z<0}\int_0^\infty dF(y)\right] \\ &= \int_0^\infty dF(x)\,[\mathbb{1}_{x\geq z}(1-F(x-z)) + \mathbb{1}_{x<z}]\end{aligned}$$

Then, two cases are possible:

$z < 0$:

$$F_\zeta(z) = \int_0^\infty dF(x)(1-F(x-z)) = e^{\lambda z}\lambda \int_0^\infty e^{-2\lambda x}dx = \frac{1}{2}e^{\lambda z}$$

$z \geq 0$:

$$F_\zeta(z) = \int_0^z dF(x) + \int_z^\infty dF(x)(1 - F(x-z)) = F(z) + \lambda \int_z^\infty e^{\lambda(z-x)} e^{-\lambda x} dx$$
$$= (1 - e^{-\lambda z}) + \frac{1}{2} e^{-\lambda z} = 1 - \frac{e^{-\lambda z}}{2}$$

It only remains to note that $F_{|\zeta|}(x) = F_\zeta(x) - F_\zeta(-x)$ for all $x \geq 0$. □

Exercise 1.14. *Suppose that r.v. $\xi_1, \dots, \xi_n$ are mutually independent and identically distributed with the c.d.f. F. For $x \in \mathbb{R}$, let us define the random variable $\hat{F}_n(x) = \frac{1}{n}\mu_n$, where μ_n is the number of $\xi_1, \dots, \xi_n$ which satisfy $\xi_k \leq x$. Show that for any x,*

$$\hat{F}_n(x) \xrightarrow{P} F(x).$$

*The function $\hat{F}_n(x)$ is called **the empirical distribution function**.*

Proof. Consider a sequence of random variables $\zeta_1, \dots, \zeta_n$ such that $\zeta_i = \mathbb{1}_{\xi_k \leq x}$. Note that $\{\zeta_i\}_{i=1,n}$ is a sequence of i.i.d. Bernoulli random varaibles with the probability of success $F(x)$. Observe that

$$\hat{F}_n(x) = \frac{1}{n} \sum_{i=1}^n \zeta_i.$$

$n\hat{F}_n(x)$ is a Binomial random variable with the expectation and variance being $F(x)$ and $\frac{F(x)(1-F(x))}{n}$ respectively. Then, by Chebyshev's inequality, we have the following result $\forall \varepsilon > 0$:

$$\mathbb{P}\left(|F_n(x) - F(x)| \geq \varepsilon\right) \leq \frac{F(x)(1-F(x))}{n\varepsilon^2}$$

The right part converges to 0 as $n \to \infty$, which gives the result. □

2 TD 2

Exercise 2.1. *Two random vectors $X \in \mathbb{R}^p$ and $Y \in \mathbb{R}^q$ are independent iff the characteristic function $\phi_Z(u)$ of the vector $Z = (X,Y)^T$ can be represented, for any $u = (a,b)^T$, $a \in \mathbb{R}^p$ and $b \in \mathbb{R}^q$, as*

$$\phi_Z(u) = \phi_X(a)\phi_Y(b)$$

Proof. The necessity is evident (i.e., if we can represent the characteristic function as a product, the variables are independent). Let us show the sufficiency in the continuous case (assuming that the common density (X,Y) exists). The density of $f_Z(x,y)$ of Z, $x \in \mathbb{R}^p$ and $y \in \mathbb{R}^q$ is given by

$$\begin{aligned} f_Z(x,y) &= (2\pi)^{-(p+q)} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-iu^T z}\phi_Z(u)du \\ &= \left[(2\pi)^{-p}\int_{-\infty}^{\infty} e^{-ia^T x}\phi_X(a)da\right]\left[(2\pi)^{-q}\int_{-\infty}^{\infty} e^{-ib^T y}\phi_Y(b)db\right] \\ &= f_X(x)f_Y(y) \end{aligned}$$

□

Exercise 2.2. *Let the joint density of r.v.'s X and Y satisfy*

$$f(x,y) = \frac{1}{2\pi}e^{-\frac{x^2}{2}}e^{-\frac{y^2}{2}}\left[1 + xy\mathbb{1}_{-1\le x,y\le 1}.\right]$$

What is the distribution of X, of Y?

Proof. To find a marginal density of Y we only need to integrate the joint density on the whole space:

$$\begin{aligned} \int_{-\infty}^{\infty} f(x,y)dx &= \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-\frac{x^2}{2}}e^{-\frac{y^2}{2}}\left[1 + xy\mathbb{1}_{-1\le x,y\le 1}dx\right] \\ &= \frac{1}{2\pi}e^{-\frac{y^2}{2}}\left[\int_{-\infty}^{\infty} e^{-\frac{x^2}{2}}dx + y\int_{-1}^{1} xe^{-\frac{x^2}{2}}dx\right] \\ &= \frac{1}{2\pi}e^{-\frac{y^2}{2}}\left[\sqrt{2\pi}+0\right] = \frac{1}{\sqrt{2\pi}}e^{-\frac{y^2}{2}} \end{aligned}$$

Thus, Y follows a standard normal distribution. The proof for X is analogous.

□

Exercise 2.3. *Consider $X \sim \mathcal{N}_p(\mu,\Sigma)$, where $\mu \in \mathbb{R}^p$ and $\Sigma \in \mathbb{R}^{p\times p}$ Prove that any linear transformation of a normal vector is again a normal vector: if $Y = AX + c$ where $A \in \mathbb{R}^{q\times p}$ and $c \in \mathbb{R}^q$ are some fixed matrix and vector (non-random),*

$$Y \sim \mathcal{N}_q\left(A\mu + c, A\Sigma A^T\right)$$

Proof. Note that any projection of Y is a normal univariate random variable. So, indeed for all $b \in \mathbb{R}^q$ the following holds:

$$b^T Y = b^T AX + b^T c = a^T X + d,$$

with $a = A^T b$ and $d = b^T c$. Using the Theorem 2.2 from course we deduce that Y is $q-$variate normal vector. Its mean and covariance matrix are given by:

$$\mathbb{E}[Y] = A\mu + c, \quad Var(Y) = A\Sigma A^T.$$

□

Exercise 2.11. *Given 2 independent r.v. X_1 and X_2 with exponential distribution with parameters λ_1 and λ_2. Find the distribution $Z = \frac{X_1}{X_2}$. Compute $\mathbb{P}(X_1 < X_2)$.*

Proof. Let us compute the following probability:

$$\begin{aligned}
\mathbb{P}(Z \geq t) = \mathbb{P}(Z \geq t) = &\iint_{\{(x_1,x_2):x_1 \geq tx_2 \geq 0\}} \lambda_1 \lambda_2 e^{-\lambda_1 x_1} e^{-\lambda_2 x_2} dx_1 dx_2 \\
&= \lambda_2 \int_0^\infty e^{-\lambda_2 x_2} \left(\int_{tx_2}^\infty e^{-\lambda_1 x_1} d(\lambda_1 x_1) \right) dx_2 = \lambda_2 \int_0^\infty e^{-\lambda_2 x_2} e^{-\lambda_1 t x_2} dx_2 \\
&= \lambda_2 \int_0^\infty e^{-x_2(\lambda_1 t + \lambda_2)} dx_2 = \frac{\lambda_2}{\lambda_1 t + \lambda_2}
\end{aligned}$$

Then,

$$F_Z(t) = \frac{\lambda_1 t}{\lambda_1 t + \lambda_2}$$

Then, it is easy to compute $\mathbb{P}(X_1 < X_2)$, which is given as:

$$\mathbb{P}(X_1 < X_2) = \mathbb{P}(Z < 1) = \frac{\lambda_1}{\lambda_1 + \lambda_2}.$$

□

Exercise 2.15. *Show that if ϕ is a characteristic function of some r.v., then $\phi^\star$, $|\phi|^2$ and $Re(\phi)$ are also characteristic functions (of certain r.v.).*

Hint: for $Re(\phi)$ consider 2 independent random variables X and Y, where Y takes values -1 and 1 with probabilities $\frac{1}{2}$, X has characteristic function ϕ, then compute the characteristic function of XY.

Proof.

- For the complex conjugate:

$$\phi^\star = \mathbb{E}[\cos(tX) - i\sin(tX)] = \mathbb{E}[\cos(-tX) + i\sin(-tX)] = \mathbb{E}\left[e^{it(-X)}\right]$$

Thus, we see that $\phi^\star$ is a characteristic function of the variable $-X$.

- Note that

$$|\phi(t)|^2 = \phi(t)\phi^\star(t) = \mathbb{E}[e^{itX}]\mathbb{E}[e^{-itX'}] = \mathbb{E}[e^{it(X-X')}],$$

where X' is a r.v. with the same distribution as X, independent of X. Then, the function $|\phi(t)|^2$ is a c.f. of a variable $X - X'$, whose c.d.f. is given by a convolution:

$$F(t) = \int_{-\infty}^{\infty} (1 - F(u - x - 0))\, dFu$$

- Note that:

$$Re(\phi) = \frac{\phi + \phi^\star}{2}$$

We have seen previously that $\phi^\star$ is the characteristic function of the variable $-X$. Consider a variable Y taking 1 and -1 with probability $\frac{1}{2}$ (independently of X). Let us write the characteristic function of the product (using the result of the exercise 2.1):

$$\mathbb{E}[\exp(itXY)] = \frac{\mathbb{E}\left[\left(e^{itX} + e^{-itX}\right)\right]}{2} = \frac{\phi + \phi^\star}{2} = Re(\phi)$$

□

Exercise 2.17. *Let (X, Y) be a random vector with density*

$$f(x, y) = C \exp\left(-x^2 + xy - \frac{y^2}{2}\right).$$

1. *Show that (X, Y) is a normal vector. Compute the expectation, the covariance matrix and the characteristic function of (X, Y). Compute the correlation coefficient ρ_{XY} of X and Y.*

2. *What is the distribution of X? Of Y? Of $2X - Y$?*

3. *Show that X and $Y - X$ are independent random variables with the same distribution.*

Proof. 1. The fact that it is normal is (more or less) obvious. We only need to find a constant C to obtain the density. Note that

$$C \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \exp\left(-x^2 + xy - \frac{y^2}{2}\right) dx dy = 2\pi C.$$

Since the double integral over the density must be equal to 1, $C = \frac{1}{2\pi}$. Let us proceed to computing the expectation and so on.

$$\mathbb{E}[X] = \frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x \exp\left(-x^2 + xy - \frac{y^2}{2}\right) dxdy = 0$$

Idem for $\mathbb{E}[Y]$. Then, the mean vector is given by $(0,0)^T$. Let us compute the second moments:

$$\mathbb{E}\left[X^2\right] = \frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x^2 \exp\left(-x^2 + xy - \frac{y^2}{2}\right) dxdy = 1$$

In a similar way we obtain $\mathbb{E}\left[Y^2\right] = 2$ and $\mathbb{E}[XY] = 1$. Thus, the covariance matrix is given as:

$$\Sigma = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}$$

Then, we can find the correlation coefficient by computing $\rho = \frac{2}{\sqrt{2}\sqrt{4}} = \frac{1}{\sqrt{2}}$. Characteristic function is given by $\exp\left(-\frac{1}{2}(z^T\Sigma z)\right)$.

2. To find the marginal density function of Y, we have to integrate the joint density, thus we have:

$$f(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \exp\left(-x^2 + xy - \frac{y^2}{2}\right) dy = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$
$$f(y) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \exp\left(-x^2 + xy - \frac{y^2}{2}\right) dx = \frac{1}{2\sqrt{\pi}} e^{-\frac{y^2}{4}}$$

Easy to see that X is a standard normal variable, while Y is a centered normal with the variance 2. In order to find the distribution of $2X - Y$ we can use the characteristic functions. Note that $\phi_X(t) = \exp\left(-t^2\right)$ and $\phi_Y(t) = \exp\left(-\frac{t^2}{2}\right)$

$$\begin{aligned}\phi_{2X-Y}(t) = \mathbb{E}\left[\exp(-it(2X - Y))\right] = \\ \frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \exp(-it(2X - Y)) f(x,y) dxdy = \frac{1}{2\sqrt{\pi}} e^{-t^2}\end{aligned}$$

Then, by Theorem 2.1. from course, $2X - Y$ is a again a normal variable with mean 0 and variance 2.

Analogously, we can compute the mean and the variance by linear algebra (knowing that $2X - Y$ follows normal distribution). $\mu_{2X-Y} = 0$, and

$$\begin{aligned} Var(2X - Y) &= Var(2X + (-Y)) = Var(2X) + Var(-Y) + 2Cov(2X, -Y) \\ &= 4Var(X) + Var(Y) - 2Cov(X,Y) = 4 + 2 - 2 = 4. \end{aligned}$$

3. Note that the vector $Z = (X, Y - X)$ is a linear transformation of a normal vector (X, Y). More precisely,

$$Z = \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} X \\ Y \end{pmatrix}$$

Then, Z follows a normal distribution with the mean 0 and the covariance matrix given by

$$\Sigma_Z = \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix}^T = I_2$$

Since Σ_Z is an identity matrix, both X and $Y - X$ are distributed by the same law and are independent.

□

Exercise 2.19. *Let ξ and η be independent r.v. with uniform distribution $U[0,1]$. Prove that*

$$X = \sqrt{-2\ln\xi}\cos(2\pi\eta), \quad Y = \sqrt{-2\ln\xi}\cos(2\pi\eta)$$

satisfy $Z = (X, Y)^T \sim N_2(0, I)$.

Hint: Let $(X, Y)^T \sim N_2(0, I)$. Change to the polar coordinates.

Proof. Recall that we can switch to the polar coordinates by applying the following transformation:

$$X = r\cos(\varphi)$$
$$Y = r\sin(\varphi).$$

Recall that the density function of the standard bivariate normal distribution is given by:

$$f(x, y) = \frac{1}{2\pi}\exp\left(-\frac{x^2 + y^2}{2}\right).$$

We can verify that in the polar coordinates the density function of the normal bivariate distribution satisfies:

$$f_{\rho,\phi}(r, \varphi) = \frac{re^{-r^2/2}}{2\pi}\mathbb{1}_{0\le\varphi<2\pi}.$$

Thus, we see that ρ and ϕ are independent. □

3 TD 3

Exercise 3.1. *We have X and Z, 2 r.v., independent with exponential distribution, $X \sim \mathscr{E}(\lambda)$, $Z \sim \mathscr{E}(1)$. Let $Y = X + Z$. Compute the regression function $g(y) = \mathbb{E}[X|Y = y]$.*

Proof. Note that

$$\mathbb{E}\left[X|Y = y\right] = \mathbb{E}\left[X|X + Z = y\right] = \mathbb{E}\left[X|X = y - Z\right]$$

Then, we can use the law of total expectation

$$\mathbb{E}_Z\left[\mathbb{E}_X\left[X|X = y - Z\right]\right] = \mathbb{E}_Z\left[y - Z\right] = y - 1$$

□

Exercise 3.10. *Consider the joint density function of X and Y given by:*

$$f(x, y) = \frac{6}{7}\left(x^2 + \frac{xy}{2}\right), \; 0 \le x \le 1, \; 0 \le y \le 2.$$

1. *Verify that f is a joint density.*
2. *Find the density of X, the conditional density $f_{Y|X}(y|x)$.*
3. *Compute $P\left(Y > \frac{1}{2}|X < \frac{1}{2}\right)$.*

Proof. 1. In order to verify that $f(x, y)$ is a joint density, we need to compute the double integral and see if it equals 1

$$\frac{6}{7}\int_0^1\int_0^2\left(x^2 + \frac{xy}{2}\right)dydx = \frac{6}{7}\int_0^1\left(2x^2 + \left(\frac{xy^2}{4}\Big|_0^2\right)\right)dx = \frac{6}{7}\int_0^1\left(2x^2 + x\right)dx$$
$$= \frac{6}{7}\left(\frac{2}{3}x^3 + \frac{x^2}{2}\right)\Big|_0^1 = 1$$

2. Using partly the computations from the first step, we obtain the density of X, given as

$$f_X(x) = \frac{6}{7}(2x^2 + x).$$

Then, we can compute the conditional density $f_{Y|X}(y|x)$ as follows:

$$f_{Y|X}(y|x) = \begin{cases} \frac{1}{2}\frac{2x+y}{2x+1} & \text{if } x \in (0,1), \\ 0 & \text{otherwise.} \end{cases}$$

3. First, note that

$$\mathbb{P}\left(X < \frac{1}{2}\right) = \frac{6}{7}\int_0^{\frac{1}{2}} (2x^2 + x)dx = \frac{6}{7}\left(\frac{2}{3}x^3 + \frac{x^2}{2}\right)\Big|_0^{\frac{1}{2}} = \frac{6}{7}\cdot\frac{5}{24} = \frac{5}{28}.$$

Then, we can compute the probability as follows:

$$\begin{aligned}
\mathbb{P}\left(Y > \frac{1}{2}, X < \frac{1}{2}\right) &= \mathbb{P}\left(Y > \frac{1}{2}\Big| X < \frac{1}{2}\right)\mathbb{P}\left(X < \frac{1}{2}\right) \\
&= \frac{5}{56}\int_0^{\frac{1}{2}}\left(\int_{\frac{1}{2}}^{2}\frac{2x+y}{2x+1}dy\right)dx = \frac{5}{56}\int_0^{\frac{1}{2}}\left(\frac{3}{2}\frac{2x}{2x+1} + \frac{1}{2}\left(\frac{y^2}{2x+1}\right)\Big|_{1/2}^{2}\right)dx \\
&= \frac{1}{2}\cdot\frac{5}{56}\int_0^{1/2}\frac{24x+15}{2x+1}dx = \frac{5}{112}\left(6 + 3\int_0^{1/2}\frac{dx}{2x+1}\right) = \\
&\frac{5}{112}\left(6 + \frac{3}{2}\log(2x+1)\Big|_0^{1/2}\right) = \frac{5}{112}\left(6 + \frac{3}{2}\log(2)\right) \approx 0.376
\end{aligned}$$

□

Exercise 3.11. *Let X and N be r.v. such that N is valued in $\{1,2,\dots\}$, and $\mathbb{E}(|X|) < \infty, \mathbb{E}(N) < \infty$. Consider the sequence $X_1, X_2, \dots,$ of independent r.v. with the same distribution as X. Show the Wald identity: if N is independent of X_i, then*

$$\mathbb{E}\left(\sum_{i=1}^{N} X_i\right) = \mathbb{E}(N)\mathbb{E}(X).$$

Proof. (Done during the lecture). This statement is easily verified by the formula of total expectation:

$$\mathbb{E}\left[\sum_{i=1}^{N} X_i\right] = \mathbb{E}\left[\mathbb{E}\left[\sum_{i=1}^{N} X_i \Big| N\right]\right] = \mathbb{E}\left[N\mathbb{E}[X]\right] = \mathbb{E}[N]\,\mathbb{E}[X].$$

□

Exercise 3.12. *Suppose that the salary of an individual satisfies $Y^\star = Xb + \sigma\varepsilon$, where $\sigma > 0, b \in \mathbb{R}, X$ is a r.v. with bounded second order moments corresponding to the capacities of the individual and ε is independent of X standard normal variable, $\varepsilon \sim \mathcal{N}(0,1)$. If $Y^\star$ is larger than the SMIC value S, the received salary is $Y = Y^\star$, otherwise it is equal to S. Compute $\mathbb{E}[Y|X]$. Is this expectation linear?*

Proof. Let us start with computing the following conditional probability:

$$\mathbb{P}\left[Y^\star > S | X = x\right] = \mathbb{P}\left[Xb + \sigma\varepsilon > S | X = x\right] = \mathbb{P}\left[\varepsilon > \frac{S - bx}{\sigma}\right] = 1 - F_{\mathcal{N}(0,1)}\left(\frac{S - bx}{\sigma}\right),$$

where $F_{\mathcal{N}(0,1)}\left(\frac{S-bx}{\sigma}\right)$ is a c.d.f. of a standard normal distribution. Then,

$$\mathbb{E}[Y|X] = Xb\left(1 - F_{\mathcal{N}(0,1)}\left(\frac{S-bx}{\sigma}\right)\right) + SF_{\mathcal{N}(0,1)}\left(\frac{S-bx}{\sigma}\right)$$

Easy to note that the expectation is linear. □

Exercise 3.13. *Let X, Y_1 and Y_2 be independent r.v., with Y_1 and Y_2 being normal $\mathcal{N}(0,1)$ and*

$$Z = \frac{Y_1 + XY_2}{\sqrt{1+X^2}}.$$

Using the conditional distribution $\mathbb{P}(Z < u|X = x)$ show that $Z \sim \mathcal{N}(0,1)$.

Proof. Done during the lecture □

Exercise 3.16. *Let $X_1, ..., X_n$ be i.i.d. r.v. with density f which is continuous except at a finite number of points. Let $S = \max(X_1, ..., X_n)$ and $I = \min(X_1, ..., X_n)$. We assume that $n > 2$.*
1^o. Identify the distributions of S, I, and (S, I), the conditional distribution of I given $S = s$ (we admit that the distribution of (S, I) possesses a density $f(s, i)$).
2^o. Apply these results in the case of the uniform distribution $U[0,1]$ of X_i; compute $E[I|S]$, conditional distribution of X_1 given $S = s$, and $E[X_1|S]$ in this case.

Proof. For the first part, recall the exercise 1.9 from the first TD. Then, the cumulative distribution function of S verifies:

$$F_S(s) = \mathbb{P}(X_i \leq s, i = 1, \ldots, n) = (F(s))^n,$$

where F is a distribution function of X_i (since they are all identically distributed). We can then write the density as follows

$$f_S(s) = nf(s)F^{n-1}(s).$$

For the I we have the following:

$$F_I(i) = 1 - [1 - F(i)]^n, \quad f_I(i) = nf(i)[1 - F(i)]^{n-1}$$

Finally, $F_{S,I}(s, i) = \mathbb{P}(I \leq i, S \leq s)$, and for any $i \leq s$ we have

$$F_{S,I}(s, i) = \mathbb{P}(S \leq s) - \mathbb{P}(S \leq s, I > i) = F^n(s) - [F(s) - F(i)]^n$$

Then, since the existence of the density function is given, we can compute it as follows:

$$f_{S|I}(s, i) = n(n-1)f(s)f(i)[F(s) - F(i)]^{n-2}\mathbb{1}_{i\leq s}$$

Therefore, the law $\mathbb{P}(I \le i|S=s)$ has a density, given by

$$f_{I|S}(i|s) = \frac{f_{S,I}(s,i)}{f_S(s)} = (n-1)\frac{f(i)}{F(s)}\left[1-\frac{F(i)}{F(s)}\right]^{n-2} \mathbb{1}_{i\le s}.$$

2^o. If $X_1,...,X_n$ follow the law $U[0,1]$, we have the following:

$$f_S(s) = ns^{n-1}\mathbb{1}_{\{0\le s\le 1\}}, \quad f_I(s) = n(1-i)^{n-1}\mathbb{1}_{\{0\le i\le 1\}},$$

and $f_{I|S}(i|s) = \frac{n-1}{s}\left(1-\frac{i}{s}\right)^{n-2}\mathbb{1}_{\{0\le i\le s\}}$. So, the conditional law $F(I\le i|S=s)$ is the law of the minimum of $n-1$ i.i.d. random variables following law $U[0,s]$. We can immediately compute that $E[I|S] = S/n$.

The computation of $E[X_1|S]$ is more intricate. Note that the law of the pair (X_1, S) does not have density with respect to Lebesgue measure, because, in particular $P(S = X_1) = \frac{1}{n}$. Then, we cannot apply the formula for computing the expectation in a bivariate case. Instead, we can proceed in the following manner: note that a necessary and sufficient condition for a family of probabilities $F(y|x)$ to be a conditional law of Y given $X=x$, is defined as follows. For any measurable and bounded functions $g(\cdot)$ and $h(\cdot)$

$$E[g(X)h(Y)] = \int g(x)\left[\int h(y)dF(y|x)\right]dF_X(x).$$

In our case, we have

$$\begin{aligned}
E(g(S)h(X_1)) &= E(g(S)h(X_1)\mathbb{1}_{\{X_1\le \max(X_2,...,X_n)\}}) + E(g(S)h(X_1)\mathbb{1}_{\{X_1>\max(X_2,...,X_n)\}}) \\
&= E(g(\max(X_2,...,X_n))h(X_1)\mathbb{1}_{\{X_1\le \max(X_2,...,X_n)\}}) \\
&+ E(g(X_1)h(X_1)\mathbb{1}_{\{X_1>\max(X_2,...,X_n)\}}) \\
&= \int_0^1\int_0^1 g(z)h(x)(n-1)z^{n-2}\mathbb{1}_{\{x\le z\}}dxdz \\
&+ \int_0^1\int_0^1 g(x)h(x)(n-1)z^{n-2}\mathbb{1}_{\{z<x\}}dxdz \\
&= \int_0^1 g(z)(n-1)z^{n-2}\left[\int_0^z h(x)dx\right]dz + \int_0^1 g(x)h(x)\left[\int_0^x (n-1)z^{n-2}dz\right]dx \\
&= \int_0^1 g(z)\left[\int_0^z h(x)\frac{n-1}{nz}dx\right]nz^{n-1}dz + \int_0^1 \frac{g(x)h(x)}{n}nx^{n-1}dx \\
&= \int_0^1 g(z)\left[\int_0^z h(x)\frac{n-1}{nz}dx + \frac{h(z)}{n}\right]dF_S(z),
\end{aligned}$$

so the conditional law of X_1 given $S=z$ is

$$dF(x|z) = \frac{n-1}{nz}\mathbb{1}_{\{0\le x<z\}} + \frac{1}{n}\delta(z-x).$$

where $\delta(u)$ is the Dirac measure at point 0 and we have

$$E[X_1|S=s] = n - \frac{1}{ns}\int_0^s x dx + \frac{s}{n} = \frac{(n-1)s+2s}{2n} = \frac{n+1}{2n}s.$$

□

Exercise 3.17. *Let $Z=(Z_1,Z_2,Z_3)^T$ be a normal vector, with density f given by,*

$$f(z_1,z_2,z_3) = \frac{1}{4(2\pi)^{3/2}}\exp\left(-\frac{6z_1^2+6z_2^2+8z_3^2+4z_1z_2}{32}\right).$$

What is the distribution of (Z_2,Z_3) given $Z_1=z_1$?

Proof. As the vector is Gaussian, we will apply Theorem 3.3 from the cours (Gauss-Markov). For that, we have to find the mean and the covariance vector of Z. After some computations, we obtain the mean vector $(0,0,0)^T$ and the covariance matrix given by:

$$\Sigma_Z = \begin{pmatrix} 3 & -1 & 0 \\ -1 & 3 & 0 \\ 0 & 0 & 2 \end{pmatrix}$$

Then, the resulting conditional distribution is given by a normal vector with the mean and covariance matrix given, respectively, by

$$\mu_{z_2,z_3|z_1} = (-1,0)^T\frac{1}{3}z_1 = (-\frac{z_1}{3},0)^T$$

$$\Sigma_{z_2,z_3|z_1} = \begin{pmatrix} 3 & 0 \\ 0 & 2 \end{pmatrix} - (-\frac{1}{3},0)^T(-1,0) = \begin{pmatrix} 3 & 0 \\ 0 & 2 \end{pmatrix} - \frac{1}{3}(-1,0)^T(-1,0) = \begin{pmatrix} \frac{8}{3} & 0 \\ 0 & 2 \end{pmatrix}$$

□

4 TD 4

Exercise 4.3. *Let $X_1, \dots, X_n$ be i.i.d. r.v. variables with exponential distribution with density $f(x) = \theta \exp(-\theta x) \mathbb{1}_{\{x>0\}}$.*
1^o. What is the distribution of $\bar{X}$? Calculate $\mathbb{E}[1/\bar{X}]$ and $Var[1/\bar{X}]$. Show that $\mathbb{E}[1/\bar{X}]$ converges to θ as $n \to \infty$. Establish the relationship

$$\mathbb{E}\left[\left(\frac{1}{\bar{X}} - \theta\right)^2\right] = Var\left(\frac{1}{\bar{X}}\right) + \left(\mathbb{E}\left[\frac{1}{\bar{X}}\right] - \theta\right)^2,$$

and conclude that

$$\mathbb{E}\left[\left(\frac{1}{\bar{X}} - \theta\right)^2\right] \to 0$$

when $n \to \infty$.
2^o. Show that $1/\bar{X}$ converges in probability to θ. What is the limit distribution of $\sqrt{n}(\bar{X} - \frac{1}{\theta})$? Is the variance of this distribution equal to $\lim_{n\to\infty} nVar(1/\bar{X})$?

Proof. To obtain the first formula, simply observe that

$$\mathbb{E}\left[\left(\frac{1}{\bar{X}} - \theta\right)^2\right] = \mathbb{E}\left[\left(\frac{1}{\bar{X}} - \mathbb{E}\left[\frac{1}{\bar{X}}\right] + \mathbb{E}\left[\frac{1}{\bar{X}}\right] - \theta\right)^2\right] = Var\left(\frac{1}{\bar{X}}\right) + \left(\mathbb{E}\left[\frac{1}{\bar{X}}\right] - \theta\right)^2.$$

Then, note that the sum of n independent i.i.d. exponential variables follows Erlang distribution, so that

$$f_{n\bar{X}}(x) = \frac{\theta^n x^{n-1}}{(n-1)!} e^{-\theta x}.$$

Then,

$$\mathbb{E}\left[\frac{1}{n\bar{X}}\right] = \frac{\theta}{n-1}, \quad \mathbb{E}\left[\frac{1}{(n\bar{X})^2}\right] = \frac{\theta^2}{(n-1)(n-2)},$$

and $\mathbb{E}[1/\bar{X}] = \theta \frac{n}{n-1}$, $Var(1/\bar{X}) = \frac{\theta^2 n^2}{(n-1)^2(n-2)}$. By the central limit theorem we have

$$\sqrt{n}\left(\frac{1}{\bar{X}} - \theta\right) \xrightarrow{\mathscr{D}} \mathcal{N}(0, \theta^2).$$

□

Exercise 4.5 (or 4.3). *Let $X_1, \dots, X_n$ be i.i.d. r.v. with common c.d.f. F. We suppose that F admits a density f with respect to the Lebesgue measure. We define* order statistics $(X_{(1)}, \dots, X_{(n)})$ *where $X_{(i)}$ are X_i sorted in ascending order:*

$$X_{(i)} \in \{X_1, \dots, X_n\}, \quad X_{(1)} \le X_{(2)} \le \dots \le X_{(n)}.$$

1^o. What is the density $f_k(x)$ of $X_{(k)}$. Compute the c.d.f. $G_k(x)$ of $X_{(k)}$.
2^o. What is the distribution of the couple $(X_{(1)}, X_{(n)})$ and the distribution of the statistic $W = X_{(n)} - X_{(1)}$ (W is called the range*). Are variables $X_{(1)}$ et $X_{(n)}$ independent?*
2^o. What are the distributions of the random variables

$$Y_k = F(X_{(k)}) \textit{ and } Z_k = G_k(X_{(k)}).$$

Proof. 1^o. First, note

$$\mathbb{P}\left(X_{(k)} < x\right) = \sum_{i=k}^{n} C_n^k F^i(x)(1-F(x))^{n-i}.$$

By integrating by parts, we obtain:

$$\begin{aligned}
\frac{n!}{(k-1)!(n-k)!}\int_0^u t^{k-1}(1-t)^{n-k}dt &= \frac{n!}{k!(n-k)!}u^k(1-u)^{n-k} + \frac{n!}{k!(n-k-1)!}u^k(1-u)^{n-k-1} \\
&= \frac{n!}{k!(n-k)!}u^k(1-u)^{n-k} + \ldots + C_n^{k+i}u^{k+i}(1-u)^{n-k-i} + \ldots + u^n \\
&= \sum_{i=k}^{n} C_n^i u^i (1-u)^{n-i}.
\end{aligned}$$

Then,

$$\mathbb{P}(X_{(k)} < x) = \frac{n!}{(k-1)!(n-k)!}\int_0^{F(x)} t^{k-1}(1-t)^{n-k}dt$$

and

$$\begin{aligned}
f_{X_{(k)}}(x) = \frac{d}{dx}P(X_{(k)} < x) &= \frac{n!}{(k-1)!(n-k)!}\frac{d}{dx}\left(\int_0^{F(x)} t^{k-1}(1-t)^{n-k}dt\right) \\
&= \frac{n!}{(k-1)!(n-k)!}f(x)F^{k-1}(x)(1-F(x))^{n-k} \\
&= kC_n^k f(x)F^{k-1}(x)(1-F(x))^{n-k}.
\end{aligned}$$

2^o. As in Exercise 3.16, we have: $F_{X_{(n)},X_{(1)}}(y,x) = P(X_{(1)} \le x, X_{(n)} \le y)$, and for $x \le y$

$$F_{X_{(n)},X_{(1)}}(y,x) = P(X_{(n)} \le y) - P(X_{(n)} \le y, X_{(1)} > x) = F^n(y) - [F(y) - F(x)]^n.$$

A joint density of $X_{(n)}$ and $X_{(1)}$ is

$$f_{X_{(n)},X_{(1)}}(y,x) = n(n-1)f(y)f(x)[F(y)-F(x)]^{n-2}I\{x \le y\}.$$

The density of W is thus

$$f_W(w) = n(n-1)\int f(x+w)f(x)[F(x+w)-F(x)]^{n-2}dx I\{w \ge 0\},$$

and its cumulative distribution function is given by

$$F_W(w) = n(n-1)\int_0^w \int f(x+u)f(x)[F(x+u)-F(x)]^{n-2}dxdu =$$
$$n\int f(x)[F(x+w)-F(x)]^{n-1}dx.$$

□

Exercise 5.8. *Let X_i, $i = 1,\dots,n$, be i.i.d. r.v. with density $\frac{\theta}{x^{\theta+1}}\mathbb{1}_{\{x\geq 1\}}$, où $\theta > 0$.*
1^o. Construct Method of Moments and Maximum Likelihood estimates of θ.
2^o. Compute corresponding Fisher information $I(\theta)$ and study the limit distribution of the MLE. Compare $(nI(\theta))^{-1}$ with the asymptotic variance of the Maximum Likelihood estimator.

Proof. For the method of moments:

- First, write $\mathbb{E}[X]$ as a function of θ. (Hint: if everything is OK, it should be equal to $\frac{\theta}{\theta+1}$).
- On the other hand, we know that the natural estimator of $\mathbb{E}[X]$ is $\frac{1}{n}\sum_{i=1}^n X_i$.
- Deduce θ from this expression!

For the Maximum Likelihood Estimator: first, let us write a joint distribution of the sample (also known as a likelihood function). It is given by

$$L_n(\theta; y) = \prod_{i=1}^n \frac{\theta}{X_i^{\theta+1}}.$$

Note that it is easier to consider the logarithm of $L_n(\theta; y)$ in order to obtain the optimal value of θ:

$$\log L_n(\theta; y) = \log \prod_{i=1}^n \frac{\theta}{X_i^{\theta+1}} = n\log\theta - \sum_{i=1}^n \log X_i^{\theta+1}$$

Then, to find a value of θ which maximizes the likelihood, we have to solve the following equation:

$$\frac{d}{d\theta}\log L_n(\theta; y) = 0 \quad \Rightarrow \quad \theta^\star = \frac{1}{\frac{1}{n}\sum_{i=1}^n \ln X_i}$$

The Maximum Likelihood Estimator is given as follows:

$$\widehat{\theta}_n^{ML} = \frac{1}{\frac{1}{n}\sum_{i=1}^n \ln X_i}.$$

Note that the random variables $Y_i = \ln X_i$ follow the exponential law: $f(y,\theta) = \theta e^{-\theta y}$, $y \geq 0$. Starting from this moment, we can simply use the computations from the Exercise 4.1. □

Exercise 5.10. *Let $X_1, \dots, X_n$ be i.i.d. r.v. with the density*

$$\theta^2 x \exp(-\theta x) \mathbb{1}_{\{x \geq 0\}}.$$

- *Compute the Method of Moments estimator $\tilde{\theta}$ of θ.*
- *Compute Maximum Likelihood estimator $\hat{\theta}$ along with its quadratic risk. Propose an unbiased estimate of θ and compare it to $\hat{\theta}$.*
- *What is the asymptotic distribution of $\sqrt{n}(\hat{\theta} - \theta)$?*

Proof. **MME of θ.**

- Compute the expectation of X, obtain $\mathbb{E}[X] = \frac{2}{\theta}$.
- Conclude that $\tilde{\theta} = \frac{2}{\bar{X}}$.

MLE of θ.

- $L_n(\theta; y) = \prod_{i=1}^n \theta^2 X_i \exp(-\theta X_i)$
- $\log L_n(\theta; y) = 2n \log\theta + \sum_{i=1}^n \log X_i - \theta \sum_{i=1}^n X_i$
- Find the maximum:

$$\frac{d}{d\theta} \log L_n(\theta; y) = \frac{2n}{\theta} - \sum_{i=1}^n X_i = 0 \quad \Rightarrow \quad \widehat{\theta} = \frac{2}{\bar{X}}$$

- Is this estimator biased?

MLE of θ is $\widehat{\theta}_n^{ML} = \frac{2}{\bar{X}}$. Here, $n\bar{X}$ follows χ^2_{4n} with $4n$ degrees of freedom (X_1 follows χ_2 with 4 degrees of freedom). □

Exercise 5.11. *Let $X_1, \dots, X_n$ be i.i.d. r.v. taking values 0, 1 and 2 with probabilities $p/2$, $p/2$, $1-p$. In this exercise, we denote n_0, n_1 et n_2 the number of 0, of 1 and of 2 in the sample.*
1^o. What is the interval Θ of variation of p? Propose MME $\hat{p}_1$ of p and compute its quadratic risk. Express $\hat{p}_1$ as a function of n_0, n_1, n_2 and n.
2^o. Express the MLE $\hat{p}_2$ as function of n_0, n_1 and n_2. Notice that $n_k = \sum_{i=1}^n I\{X_i = k\}$, $k = 0, 1, 2$, compute its quadratic risk and compare it to that of $\hat{p}_1$.

Proof. **MLE for p.**

- $L_n(\theta; y) = \left(\frac{p}{2}\right)^{n_0} \left(\frac{p}{2}\right)^{n_1} (1-p)^{n_2}$
- $\log L_n(\theta; y) = (n_0 + n_1)\big(\log(p) - \log(2)\big) + n_2 \log(1-p)$

- $\hat{p}_2 = \frac{n_0+n_1}{n}$
- Compute $\mathbb{E}\left[\hat{p}_2\right]$ and $Var\left[\hat{p}_2\right]$.

□

Exercise 5.12. *Let $X_1,\ldots,X_n$ be i.i.d. r.v. with uniform distribution $U[0,\theta]$. Determine the MLE $\hat{\theta}$ of the unknown parameter, compute its bias and variance. Consider estimates of the form $c\hat{\theta}$, $c \in \mathbb{R}$, and identify the value of c which results in the smallest quadratic risk.*

Proof. MLE of θ is $\max_n X_{(n)}$. So, (see. Exercice 1.9, part 2) the bias and the variance of the estimator are given as follows:

$$\mathbb{E}_{\theta^*}\left[\widehat{\theta}_n^{ML}\right] = \theta\frac{n}{n+1}, \quad b_n = -\frac{\theta}{n+1}, \quad \sigma_n^2(\widehat{\theta}_n^{MV}) = \frac{\theta^2 n}{(n+1)^2(n+2)}.$$

The risk of MLE is given by

$$R_n(\theta,\widehat{\theta}_n^{MV}) = b_n^2 + \sigma_n^2 = \frac{2\theta}{(n+1)(n+2)}.$$

□

Exercise 5.14. *Suppose that an observation sample of size n from the Poisson distribution is available. Our objective is to estimate the unknown parameter θ of the distribution. Describe two estimators of θ and a confidence interval for θ asymptotically at level 0.99 in the case of $n = 200$.*

Proof. We will use the usual statistics $\bar{X}$ to construct the confidence intervals. Note that $n\bar{X} \sim \mathscr{P}(n\theta)$ under $\mathbb{P}_\theta$ (here $\mathscr{P}(\lambda)$ is a Poisson distribution with parameter λ). It is easy to see that the function $G_\theta(k) = P_\theta(n\bar{X} \le k)$ is monotone in θ: indeed, let η be a Poisson variable with parameter λ. Thus,

$$\mathbb{P}_\lambda(\eta \le k) = \sum_{i=0}^{k} \frac{\lambda^i}{i!} e^{-\lambda}.$$

However,

$$\left(\mathbb{P}_\lambda(\eta \le k)\right)'_\lambda = \sum_{i=0}^{k}\left(\frac{\lambda^i}{i!} - \frac{\lambda^{i-1}}{(i-1)!}\right)e^{-\lambda} = -\frac{\lambda^k}{k!}e^{-\lambda} < 0.$$

Then, the function $G_\theta(k)$ is monotonously decreasing in λ for all $k \in \mathbb{N}$.

Now, denote $F_\lambda(k)$ the distribution function for the Poisson law with parameter λ, $b_{\alpha/2}$ and $b_{1-\alpha/2}$ defined as follows:

$$b_{\alpha/2} = \theta \text{ such that } F_{n\theta}(n\bar{X}) = \frac{\alpha}{2}, \text{ and } b_{1-\alpha/2} = \theta' \text{ such that } F_{n\theta'}(n\bar{X}) = 1 - \frac{\alpha}{2},$$

where $[a]$ is an integer part of a. Thus, $C(X^n) = [b_{1-\alpha/2}, b_{\alpha/2}]$ is a confidence interval in θ on the level $1-\alpha$.

Alternatively, we can construct an asymptotic confidence interval with an asymptotic level $1-\alpha$: as $\widehat{\theta}_n^{ML} = \bar{X}$, $I(\theta) = \frac{1}{\theta}$,

$$C^a(X^n) = \left[\bar{X} - \sqrt{\frac{\bar{X}}{n}} q^N_{1-\alpha/2},\ \bar{X} + \sqrt{\frac{\bar{X}}{n}} q^N_{1-\alpha/2} \right]$$

$\square$