Correction des exercices du Chapitre 01

Correction de l'exercice 01

a) **Soit A un événement tel qu'il y ait au moins deux coïncidences entre**

$$X_1, X_2, \ldots, X_n \quad A_{ij} = \{X_i = X_j\}.$$

$$P\{A_{ij}\} = 0 \text{ if } i \neq j.$$

$$1 - P\{X_{1,n} < X_{2,n} < \cdots < X_{n,n}\} = P\{A\} \leqslant \sum_{i \neq j} P\{A_{ij}\} = 0.$$

b)

$$p_n = n! P\{X_1 < X_2 < \cdots < X_n\} = n! \sum_{k_1=0}^{\infty} (1-p) p^{k_1} \sum_{k_2=k_1+1}^{\infty} (1-p) p^{k_2} \cdots \sum_{k_n=k_{n-1}+1}^{\infty} (1-p) p^{k_n}.$$

$$p_n = \frac{n!(1-p)^n p^{n(n-1)/2}}{\prod_{k=1}^{n} (1-p)^k} = \frac{n! p^{n(n-1)/2}}{(1+p)(1+p+p^2)\cdots(1+p+\cdots+p^{n-1})}.$$

En particulier,

$$p_2 = \frac{2p}{1+p} \quad \text{and} \quad p_3 = \frac{6p^3}{(1+p)(1+p+p^2)}.$$

c) Les espérances et les variances de R(k)

$$\mathrm{E}R(k) = \frac{n+1}{2}, \mathrm{Var}\, R(k) = \frac{(n-1)^2}{12}, 1 \leqslant k \leqslant n.$$

Correction de l'exercice 02

1. Comme les X_i ont des densités par rapport à Lebesgues, on a $X_i \neq X_j$ $\lambda-$p.p.. Alors p.p.

$$f(X_{(1)}, \dots, X_{(n)}) = \sum_{\sigma \in \mathcal{P}(n)} f(X_{\sigma(1)}, \dots, X_{\sigma(n)}) I(X_{\sigma(1)} < \cdots < X_{\sigma(n)}).$$

Soit $\sigma \in \mathcal{P}(n)$. Comme les X_i sont i.i.d., on voit que $(X_{\sigma(1)}, \dots, X_{\sigma(n)})^\top \sim (X_1, \dots, X_n)^\top$. Alors, pour tout $f \in \mathcal{C}_b(\mathbb{R}^n)$,

$$\mathbb{E} f(X_{\sigma(1)}, \dots, X_{\sigma(n)}) I(X_{\sigma(1)} < \cdots < X_{\sigma(n)}) = \mathbb{E} f(X_1, \dots, X_n) I(X_1 < \cdots < X_n)$$
$$= \int_{\mathcal{R}^n} f(x_1, \dots, x_n) \Big(\Pi_{i=1}^n f(x_i) \Big) I(x_1 < \cdots < x_n) dx_1 \cdots dx_n.$$

On en déduit que la loi de $(X_{(1)}, \dots, X_{(n)})$ admet une densité par rapport à Lebesgue donnée par

$$f(x_1, \dots, x_n) = n! \Big(\Pi_{i=1}^n f(x_i) \Big) I(x_1 < \cdots < x_n).$$

2. On calcul la fonction de répartition de $X_{(k)}$. Soit $t \in \mathcal{R}$,

$$\mathbb{P}[X_{(k)} \leq t] = \mathbb{P}\big[\exists I \subset \{1, \dots, n\} : |I| \geq k, \forall i \in I, X_i \leq t\big] = \mathbb{P}[M \geq k]$$

où $M = \sum_{i=1}^n I(X_i \leq t)$ est une multinomiale de paramétre n et $\mathbb{P}[X_1 \leq t] = F(t)$. On a donc

$$\mathbb{P}[X_{(k)} \leq t] = \sum_{j=k}^n \binom{n}{j} F(t)^j (1 - F(t))^{n-j}.$$

Comme F est absolument continue la cdf de $X_{(k)}$ l'est aussi. Donc $X_{(k)}$ admet une densité par rapport à Lebesgues donnée par :

$$f(t) = \sum_{j=k}^n \binom{n}{j} \left(j f(t) F(t)^{j-1} (1 - F(t))^{n-j} + (n-j) F(t)^j (-f(t)) (1 - F(t))^{n-j-1} \right)$$
$$= \frac{n!}{(k-1)!(n-k)!} F(t)^{k-1} (1 - F(t))^{n-k}.$$

3. La fonction de répartition de $X_{(1)}$ vérifie :

$$1 - F_{X_{(1)}}(t) = \mathbb{P}[X_{(1)} > t] = \mathbb{P}[X_1 > t, \dots, X_n > t] = \Big(\mathbb{P}[X_1 > t] \Big)^n = \big(1 - F(t)\big)^n.$$

La fonction de répartition de $X_{(n)}$ est donnée par :

$$F_{X_{(n)}}(t) = \mathbb{P}[X_{(n)} \leq t] = \mathbb{P}[X_1 \leq t, \dots, X_n \leq t] = \Big(\mathbb{P}[X_1 \leq t] \Big)^n = \big(F(t)\big)^n.$$

Pour la fonction de répartition du couple $(X_{(1)}, X_{(n)})$, on calcul la répartition du couple $(X_{(1)}, X_{(n)})$ dans le quadrant inférieur droit. On a pour tout x, y réels :

$$\mathbb{P}[X_{(1)} > x, X_{(n)} \leq y] = \mathbb{P}[x < X_1 \leq y, \dots, x < X_n \leq y]$$
$$= \Big(\mathbb{P}[x < X_1 \leq y] \Big)^n = I(x \leq y) \big(F(y) - F(x)\big)^n.$$

On a :

$$\mathbb{P}[X_{(1)} > x, X_{(n)} \leq y] + \mathbb{P}[X_{(1)} \leq x, X_{(n)} \leq y] = \mathbb{P}[X_{(n)} \leq y] = F(y)^n.$$

Alors,

$$F(x,y) = \mathbb{P}[X_{(1)} \leq x, X_{(n)} \leq y] = F(y)^n - I(x \leq y)\big(F(y) - F(x)\big)^n.$$

La densité de $(X_{(1)}, X_{(n)})$ est donnée par

$$f(x,y) = \frac{\partial^2 F}{\partial x \partial y}(x,y) = n(n-1)I(x \leq y)f(x)f(y)\big(F(y) - F(x)\big)^{n-2}.$$

La loi de la statistique $W = X_{(n)} - X_{(1)}$ est donnée par ce qui suit. Soit $f \in \mathcal{C}_b(\mathcal{R})$, on a

$$\begin{aligned}
\mathbb{E}f(W) &= \int_{\mathcal{R}^2} f(y-x)d\mathbb{P}^{(X_{(1)}, X_{(n)})}(x,y) \\
&= n(n-1)\int_{\mathcal{R}^2} f(y-x)I(x \leq y)\big(F(y) - F(x)\big)^{n-2}dxdy \\
&= \int_0^\infty f(u)\Big(n(n-1)\int_{\mathcal{R}} \big(F(u+x) - F(x)\big)^{n-2}dx\Big)du.
\end{aligned}$$

Alors W a pour densité

$$u \mapsto I(u \geq 0)n(n-1)\int_{\mathcal{R}} \big(F(u+x) - F(x)\big)^{n-2}dx.$$

Les variables $X_{(1)}$ et $X_{(n)}$ sont indépendantes si et seulement si pour tout x et y, on a

$$\begin{aligned}
F(y)^n - I(x \leq y)\big(F(y) - F(x)\big)^n &= \mathbb{P}[X_{(1)} \leq x, X_{(n)} \leq y] \\
&= \mathbb{P}[X_{(1)} \leq x]\mathbb{P}[X_{(n)} \leq y] = \Big(1 - (1 - F(x))^n\Big)F(y)^n.
\end{aligned}$$

Il faut donc $I(x \leq y)\big(F(y) - F(x)\big)^n = \big(F(y) - F(y)F(x)\big)^n$ pour tout x, y. Ce qui n'est pas vrai en générale.

Correction de l'exercice 03

1) On a

$$F_{X_{(n)}}(x) = P(X_{(n)} \leq x) = P\left(\cap_{i=1}^n \{X_i \leq x\}\right) = \prod_{i=1}^n P(X_i \leq x) = F^n(x),$$

où l'avant dernière égalité est justifiée par l'indépendance entre les v.a.r. $X_1, \ldots, X_n$. Cette fonction étant dérivable (puisque F l'est) sur $\mathbb{R}^+$, la densité de X_n est :

$$f_{X_{(n)}}(x) = nF^{n-1}(x)f(x).$$

2) On a

$$P(X_{(1)} > x) = P\left(\cap_{i=1}^n \{X_i > x\}\right) = \prod_{i=1}^n P(X_i > x) = (1 - F(x))^n,$$

où l'avant dernière égalité est ici aussi justifiée par l'indépendance entre les v.a.r. $X_1, \ldots, X_n$. D'où

$$F_{X_{(1)}}(x) = 1 - (1 - F(x))^n$$

et

$$f_{X_{(n)}}(x) = -n(1 - F(x))^{n-1}(-f(x)) = n(1 - F(x))^{n-1}f(x).$$

3) Supposons dans un premier temps que $x_1 \leq x_n$. On peut écrire :

$$\begin{aligned} P(X_{(1)} \leq x_1, X_{(n)} \leq x_n) &= P(X_{(n)} \leq x_n) - P(X_{(1)} > x_1, X_{(n)} \leq x_n) \\ &= F^n(x_n) - P\left(\cap_{i=1}^n \{X_i \in]x_1, x_n]\}\right) = F^n(x_n) - (F(x_n) - F(x_1))^n . \end{aligned}$$

En dérivant deux fois, on obtient

$$\frac{\partial F_{X_{(1)},X_{(n)}}}{\partial x_1}(x_1, x_n) = n\left(F(x_n) - F(x_1)\right)^{n-1} f(x_1)$$

$$\text{et} \quad \frac{\partial^2 F_{X_{(1)},X_{(n)}}}{\partial x_1 \partial x_n}(x_1, x_n) = n(n-1)\left(F(x_n) - F(x_1)\right)^{n-2} f(x_1)f(x_n).$$

Maintenant si $x_1 > x_n$, on a

$$P(X_{(1)} \leq x_1, X_{(n)} \leq x_n) = P(X_{(n)} \leq x_n) = F^n(x_n)$$

qui en dérivant par rapport à x_1 et x_n s'annule. On a donc la densité du couple $(X_{(1)}, X_{(n)})$:

$$f_{(X_{(1)},X_{(n)})}(x_1, x_n) = n(n-1)\left(F(x_n) - F(x_1)\right)^{n-2} f(x_1)f(x_n)\mathbb{1}_{\{x_1 \leq x_n\}}.$$

Disposant de la densité du couple $(X_{(1)}, X_{(n)})$, pour trouver la densité de la v.a.r. $R = X_{(n)} - X_{(1)}$, on peut dans un premier calculer la densité du couple (Q, R), où $Q = X_{(1)}$, et ensuite calculer la loi marginale de la seconde coordonnée de ce couple.

Le calcul de la loi du couple (Q, R) s'effectue facilement grâce à la formule du changement de variable. Prenons la fonction $\varphi(u, v) = (u, v - u)$ qui est évidemment un $\mathcal{C}^1$-difféomorphisme de fonction réciproque $\varphi^{-1}(y_1, y_2) = (y_1, y_1 + y_2)$. Le Jacobien de φ^{-1} est égal à 1. Ainsi la formule du changement de variable nous donne :

$$\begin{aligned} f_{Q,R}(q,r) &= f_{X_{(1)},X_{(n)}}\left(\varphi^{-1}(q,r)\right)|J_{\varphi^{-1}}(q,r)|1\!\!1_{Im\varphi}(q,r) \\ &= n(n-1)\left(F(q+r)-F(q)\right)^{n-2} f(q)f(q+r)1\!\!1_{\mathbb{R}\times\mathbb{R}^+}(q,r). \end{aligned}$$

La densité marginale de R est donc :

$$f_R(r) = \int_{R+} n(n-1))\left(F(q+r)-F(q)\right)^{n-2} f(q)f(q+r)dq.$$

Sa f.d.r. est alors :

$$F_R(r) = \int_0^r f_R(x)dx.$$

4) On a

$$N_y = \sum_{i=1}^n 1\!\!1_{X_i \le y}.$$

Les v.a.r. $1\!\!1_{X_i \le y}$, pour $i = 1, \ldots, n$, étant i.i.d. de loi de Bernoulli de paramètre $F(y)$, la loi de N_y est une Binomiale de paramètres n et $F(y)$, i.e.

$$N_y \sim B(n, F(y)).$$

Par ailleurs, on a l'égalité entre les événements :

$\{N_y \ge k\} = \{$Il y a un nombre supérieur ou égal à k de X_i inférieurs à $y\} = \{X_{(k)} \le y\}$.

Ainsi, il vient :

$$F_{X_{(k)}}(x) = P(X_{(k)} \le x) = P(N_x \ge k) = \sum_{i=k}^n C_n^i \left(F(x)\right)^i \left(1-F(x)\right)^{n-i}.$$

5) Calculons la probabilité $P(X_{(k)} \in]x, x+dx])$. Différents événements disjoints peuvent donner l'événement dont on veut calculer la probabilité :

- $k-1$ variables "tombent" dans l'intervalle $]-\infty, x]$, 1 variable dans l'intervalle $]x, x+dx]$ et $n-k$ dans l'intervalle $]x+dx, +\infty[$;
- $k-2$ variables "tombent" dans l'intervalle $]-\infty, x]$, 2 variables dans l'intervalle $]x, x+dx]$ et $n-k$ dans l'intervalle $]x+dx, +\infty[$;
- $k-1$ variables "tombent" dans l'intervalle $]-\infty, x]$, 2 variables dans l'intervalle $]x, x+dx]$ et $n-k-1$ dans l'intervalle $]x+dx, +\infty[$;
- $k-3$ variables 'tombent" dans l'intervalle $]-\infty, x[$, 3 variables dans l'intervalle $]x, x+dx]$ et $n-k$ dans l'intervalle $]x+dx, +\infty]$;
- etc...

Le premier événement s'écrit :

$$\begin{aligned} &\{X_{(1)}, \ldots, X_{(k-1)} \text{ sont inférieurs à } x, \\ &X_{(k)} \text{ est dans l'intervalle }]x, x+dx] \\ &\text{ et } X_{(k+1)}, \ldots, X_{(n)} \text{ sont supérieurs à } x+dx\} \end{aligned}$$

La probabilité $P_1(dx)$ de cet événement s'obtient aisément en remarquant que l'on est dans la situation d'un tirage d'une loi multinomiale à trois résultats possibles. Ainsi:

$$P_1(dx) = \frac{n!}{(k-1)!1!(n-k)!} \left(F(x)\right)^{k-1} \left(F(x+dx) - F(x)\right) \left(1 - F(x+dx)\right)^{n-k}.$$

D'où on tire :

$$\lim_{dx \to 0} \frac{P_1(dx)}{dx} = \frac{n!}{(k-1)!(n-k)!} \left(F(x)\right)^{k-1} f(x) \left(1 - F(x)\right)^{n-k}.$$

Regardons maintenant les probabilités des autres événements ci-dessus. Pour chacun d'entre eux, il y a au moins deux variables X_i qui se trouvent dans l'intervalle $]x, x+dx]$. La probabilité de ces événements contiendra donc un terme de la forme $(F(x+dx) - F(x))^m$ avec $2 \leq m \leq n$. Ces termes divisés par dx tendront alors vers 0 quand dx tend vers 0. Ainsi toutes les probabilités des événements autres que le premier de la liste précédente divisées par dx ont une limite qui tend vers 0 quand dx tend vers 0. On a donc :

$$f_{X_{(k)}}(x) = \lim_{dx \to 0} \frac{P\left(X_{(k)} \in]x, x+dx]\right)}{dx} = \frac{n!}{(k-1)!(n-k)!} \left(F(x)\right)^{k-1} f(x) \left(1 - F(x)\right)^{n-k}.$$

6) Comme $F(x)$ et $1 - F(x)$ sont dans $[0, 1]$, on peut écrire :

$$\mathbb{E}|X_{(k)}| = \int_{\mathbb{R}} |x| f_{X_{(k)}}(x) dx \leq \frac{n!}{(k-1)!(n-k)!} \int_{\mathbb{R}} |x| f(x) dx = \frac{n!}{(k-1)!(n-k)!} \mathbb{E}|X|$$

dont on tire aisément le résultat voulu.

7) Notons Σ_n l'ensemble des permutations sur l'ensemble $\{1, 2, \dots, n\}$. Soit B un borélien de $\mathbb{R}^n$. On a :

$$\begin{aligned} P\left((X_{(1)}, \dots, X_{(n)}) \in B\right) &= \sum_{\sigma \in \Sigma_n} P\left(\{(X_{\sigma(1)}, \dots, X_{\sigma(n)}) \in B\} \cap \{X_{\sigma(1)} < \cdots < X_{\sigma(n)}\}\right) \\ &= \sum_{\sigma \in \Sigma_n} \int_B \mathbb{1}_{u_1 < u_2 < \cdots < u_n} \left(\prod_{i=1}^n f(u_i)\right) du_1 \cdots du_n \\ &= \int_B n! \left(\prod_{i=1}^n f(u_i)\right) \mathbb{1}_{u_1 < u_2 < \cdots < u_n} du_1 \cdots du_n. \end{aligned}$$

Cette égalité étant vraie pour tout borélien B de $\mathbb{R}^n$, on en déduit que

$$f_{X_{(1)}, \dots, X_{(n)}}(u_1, \dots, u_n) = n! \left(\prod_{i=1}^n f(u_i)\right) \mathbb{1}_{u_1 < u_2 < \cdots < u_n}.$$

Correction de l'exercice 04

Dans ce qui suit, X_1, … X_n est un échantillon aléatoire simple et $Y_1 \leq \ldots \leq Y_n$ sont les statistiques d'ordre.

1 Soit $X_1, \ldots, X_n$ un échantillon aléatoire simple d'une population exponentielle de paramètre β.

a) Montrer que Y_1 est de loi exponentielle de paramètre β/n.

$G(y) = P(Y_1 \leq y) = 1 - P(Y_1 > y) = 1 - P(X_1 > y, \ldots, X_n > y) = 1 - \left[e^{-y/\beta}\right]^n =$

$1 - e^{-ny/\beta}$, ce qui est bien la fonction de répartition d'une exponentielle de paramètre β/n.

b) Montrer que la densité de Y_n est

$$g(y_n) = \frac{n}{\beta} e^{-y_n/\beta}[1 - e^{-y_n/\beta}]^{n-1}$$

$G(y) = P(Y_n \leq y) = P(X_1 \leq y, \ldots, X_n \leq y) = \left[1 - e^{-y/\beta}\right]^n$. La fonction de densité est $g(y) = G'(y) = g(y) = \frac{n}{\beta} e^{-y/\beta}[1 - e^{-y/\beta}]^{n-1}$

2 Soit $X_1, \ldots, X_n$ un échantillon aléatoire simple d'une population uniforme sur (0 ; 1).

a) Déterminer les distributions de Y_1 et de Y_n.

La fonction de répartition de X est $F(x) = x$ sur (0 ; 1).

$G_1(y) = P(Y_1 \leq y) = 1 - P(Y_1 > y) = 1 - P(X_1 > y, \ldots, X_n > y) = 1 - [1 - F(y)]^n$
$= 1 - [1-y]^n$

$G_n(y) = P(Y_n \leq y) = P(X_1 \leq y, \ldots, X_n \leq y) = [F(y)]^n = y^n$.

Pour trouver les fonctions de densité il suffit de dériver.

b) Déterminer la distribution de la médiane $\tilde{x}$

Nous supposerons que $n = 2m+1$. Donc la médiane est la donnée de rang $m+1$. La fonction de densité est donc

$$h(\tilde{x}) = \frac{n!}{m!m!}[F(\tilde{x})]^m f(\tilde{x})[1 - F(\tilde{x})]^m = \frac{n!}{m!m!}[\tilde{x}]^m [1 - \tilde{x}]^m$$

c) Déterminer l'espérance et la variance de Y_1

La fonction de densité de Y_1 est $g(y) = n(1-y)^{n-1}$ sur (0,1). L'espérance de Y_1 est $\int_0^1 x\left[n(1-x)^{n-1}\right]dx = \frac{1}{(n+1)}$.