Corrigé de l'examen final

Exercice 1

Une urne contient 3 boules numérotées de 1 à 3. On tire sucessivement et avec remise n boules (on remet la boule tirée dans l'urnie après chaque tirage). On suppose que les tirages sont aléatoires et indépendants. Soit X_i le numéro de la $i^{\text{ème}}$ boule tirée.

1. Quelle est la loi de X_i, pour $1 \leq i \leq n$?

   Solution : Il est clair que les variables aléatoires X_i ont toutes *la même loi*, uniforme sur l'ensemble fini $\{1;2;3\}$. Autrement dit, leur loi est donnée par

k	1	2	3
$\mathbb{P}(X_i = k)$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$

2. Calculer $\mathbb{E}[X_i]$ et $\mathbb{V}\text{ar}[X_i]$, pour $1 \leq i \leq n$.

   Solution : Ayant toutes la même loi, les X_i ont aussi *la même espérance et la même variance.* Pour tout i entre 1 et n,

$$\begin{aligned} \mathbb{E}[X_i] &= \frac{1}{3} + 2 \times \frac{1}{3} + 3 \times \frac{1}{3} = 2, \\ \mathbb{E}\left[X_i^2\right] &= \frac{1}{3} + 4 \times \frac{1}{3} + 9 \times \frac{1}{3} = \frac{14}{3}, \\ \mathbb{V}\text{ar}[X_i] &= \frac{14}{3} - 4 = \frac{2}{3}. \end{aligned}$$

3. Soit $Z = \max(X_1, X_2)$. Déterminer la loi de Z.

   Solution : Comme X_1 et X_2 ne peuvent prendre que les valeurs $1, 2$ et 3, il en va de même pour Z. On trouve, en étudiant les différents cas possibles et en utilisant l'indépendance de X_1 et X_2,

$$\begin{aligned} \mathbb{P}(Z = 1) &= \mathbb{P}(X_1 = 1, X_2 = 1) = \left(\frac{1}{3}\right)^2 = \frac{1}{9}, \\ \mathbb{P}(Z = 2) &= \mathbb{P}(X_1 = 1, X_2 = 2) + \mathbb{P}(X_1 = 2, X_2 = 1) + \mathbb{P}(X_1 = 2, X_2 = 2) = 3 \times \frac{1}{9} = \frac{1}{3}, \\ \mathbb{P}(Z = 3) &= 1 - \mathbb{P}(Z = 1) - \mathbb{P}(Z = 2) = 1 - \frac{1}{9} - \frac{1}{3} = \frac{5}{9}, \\ \mathbb{E}[Z] &= \frac{1}{9} + 2 \times \frac{1}{3} + 3 \times \frac{5}{9} = \frac{22}{9}. \end{aligned}$$

4. Soit $\overline{X} = \frac{X_1 + X_2 + \cdots + X_n}{n}$. Calculer $\mathbb{E}\left[\overline{X}\right]$ et $\mathbb{V}\text{ar}\left[\overline{X}\right]$.

   Solution : On commence par utiliser la linéarité de l'espérance :

$$\mathbb{E}\left[\overline{X}\right] = \frac{1}{n}\mathbb{E}[X_1 + \cdots + X_n] = \frac{1}{n}(\mathbb{E}[X_1] + \cdots + \mathbb{E}[X_n]) = \frac{1}{n} n \mathbb{E}[X_1] = \mathbb{E}[X_1] = 2$$

   On utilise ensuite le fait que pour toute variables aléatoire X et tout réel constant λ, $\mathbb{V}\text{ar}[\lambda X] = \lambda^2 \, \mathbb{V}\text{ar}[X]$, ce qui donne

$$\mathbb{V}\text{ar}\left[\frac{1}{n}(X_1 + \cdots + X_n)\right] = \frac{1}{n^2} \mathbb{V}\text{ar}[X_1 + \cdots + X_n].$$

   Enfin, *comme les X_i sont indépendantes*, on a

$$\mathbb{V}\text{ar}[X_1 + \cdots + X_n] = \mathbb{V}\text{ar}[X_1] + \cdots + \mathbb{V}\text{ar}[X_n] = n \, \mathbb{V}\text{ar}[X_1] = \frac{2}{3} n.$$

Finalement,

$$\mathbb{V}\mathrm{ar}\left[\overline{X}\right] = \frac{1}{n^2} \times \frac{2}{3}n = \frac{2}{3n}.$$

5. Quelle loi continue peut-on utiliser pour approcher la loi de $\overline{X}$ si n est suffisamment grand ? (Préciser les paramètres de la loi.)

   Solution : D'après le théorème central limite, si n est suffisamment grand [1] on peut approcher la loi de la moyenne $\overline{X}$ de variables aléatoires indépendantes de même loi par la loi normale, de même espérance et de même variance que $\overline{X}$.

   Ici, on peut donc dire que la loi de $\overline{X}$ est « proche » de la loi $\mathcal{N}\left(2, \left(\sqrt{\frac{2}{3n}}\right)^2\right)$.

6. En déduire les probabilités $\mathbb{P}\left(\overline{X} \leq 2,16\right)$ et $\mathbb{P}\left(1,84 \leq \overline{X} \leq 2,16\right)$ pour $n = 100$.

   Solution : On fait donc ici l'approximation que $X \sim \mathcal{N}\left(2, \left(\sqrt{\frac{2}{3\times 100}}\right)^2\right)$. On a donc

$$\mathbb{P}\left(\overline{X} \leq 2,16\right) = \mathbb{P}\left(\frac{\overline{X}-2}{\sqrt{\frac{2}{3\times 100}}} \leq \frac{2,16}{\sqrt{\frac{2}{3\times 100}}}\right) \approx \mathbb{P}\left(N \leq 1,96\right),$$

   où $N \sim \mathcal{N}(0,1)$. Par conséquent, d'après la table de la fonction de répartition de la loi normale centrée réduite,

$$\mathbb{P}\left(\overline{X} \leq 2,16\right) \approx 0,975.$$

   Par un calcul similaire, on trouve

$$\mathbb{P}\left(1,84 \leq \overline{X} \leq 2,16\right) \approx \mathbb{P}\left(-1,96 \leq N \leq 1,96\right).$$

   En utilisant la symétrie de la loi normale centrée réduite, on a

$$\begin{aligned}\mathbb{P}\left(-1,96 \leq N \leq 1,96\right) &= \mathbb{P}\left(N \leq 1,96\right) - \mathbb{P}\left(N < -1,96\right) = \mathbb{P}\left(N \leq 1,96\right) - \mathbb{P}\left(N > 1,96\right)\\ &= \mathbb{P}\left(N \leq 1,96\right) - \left(1 - \mathbb{P}\left(N \leq 1,96\right)\right) = 2 \times \mathbb{P}\left(N \leq 1,96\right) - 1.\end{aligned}$$

   Finalement, on trouve

$$\mathbb{P}\left(1,84 \leq \overline{X} \leq 2,16\right) \approx 0,950.$$

Exercice 2 Soit X une variable aléatoire de densité f définie par :

$$f(x) = \begin{cases} ax^{-\lambda} & \text{si } x \geq 1, \\ 0 & \text{sinon.} \end{cases} \qquad \text{avec } \lambda > 1.$$

1. Déterminer la valeur de a en fonction de λ.

   Solution : Il faut que f soit positive et d'intégrale 1. Pour que f soit positive, il faut que a le soit aussi. Ensuite, on calcule

$$\int_1^{\infty} x^{-\lambda}\,\mathrm{d}x = \left[\frac{1}{1-\lambda}x^{1-\lambda}\right]_0^{+\infty} = \frac{1}{1-\lambda} - \lim_{x\to\infty} \frac{1}{1-\lambda}x^{1-\lambda}$$

   Comme $\lambda > 1$, on a $1-\lambda < 0$ donc cette dernière limite est 0. Finalement, on a

$$\int_1^{+\infty} x^{-\lambda}\,\mathrm{d}x = \frac{1}{\lambda - 1},$$

   de sorte que $a = \lambda - 1$ pour que f soit d'intégrale 1, et alors a est automatiquement positif.

1. En pratique, les statisticiens utilisent cette approximation dès que $n \geq 30$.

2. Déterminer la fonction de répartition de X.

   Solution : La fonction de répartition de X est la primitive de f dont la limite en $-\infty$ est 0. Comme la densité f est nulle sur l'intervalle $]-\infty\,;1[$, la fonction de répartition F aussi, et pour $x \geq 1$,

$$F(x) = \mathbb{P}(X \leq x) = \int_1^x f(t)\,\mathrm{d}t = \left[(\lambda-1)\left(\frac{t^{-\lambda+1}}{-\lambda+1}\right)\right]_1^x = \left[-t^{-\lambda+1}\right]_1^x = 1 - x^{1-\lambda}.$$

3. Quelles sont les valeurs de λ pour lesquelles $\mathbb{V}\mathrm{ar}[X]$ existe ? Calculer alors $\mathbb{E}[X]$ et $\mathbb{V}\mathrm{ar}[X]$.

   Solution : Pour que $\mathbb{V}\mathrm{ar}[X]$ existe, il faut (et il suffit) que l'espérance $\mathbb{E}[X^2]$ soit finie. Or,

$$\mathbb{E}\left[X^2\right] = \int_1^{+\infty} x^2 f(x)\,\mathrm{d}x = (\lambda-1)\int_1^{+\infty} x^{2-\lambda}\,\mathrm{d}x.$$

   Cette dernière intégrale est finie si et seulement si l'exposant $-\lambda+2 < -1$, c'est-à-dire $\lambda > 3$. Par des calculs de primitives et de limites similaires à ceux de la question 1, on trouve

$$\begin{aligned} \mathbb{E}[X] &= \int_1^{+\infty} x f(x)\,\mathrm{d}x = (\lambda-1)\int_1^{+\infty} x^{-\lambda+1}\,\mathrm{d}x = \frac{\lambda-1}{\lambda-2}, \\ \mathbb{E}\left[X^2\right] &= (\lambda-1)\int_1^{+\infty} x^{2-\lambda}\,\mathrm{d}x = \frac{\lambda-1}{\lambda-3}, \\ \mathbb{V}\mathrm{ar}[X] &= \frac{\lambda-1}{\lambda-3} - \left(\frac{\lambda-1}{\lambda-2}\right)^2. \end{aligned}$$

Exercice 3

Soit (X,Y) un couple de variables aléatoires de densité définie par :

$$f(x,y) = \begin{cases} xe^{-2y} & \text{si } x \in [0\,;2] \text{ et } y \geq 0, \\ 0 & \text{sinon.} \end{cases}$$

1. Justifier par un calcul que Y suit la loi exponentielle de paramètre 2 et que la fonction f_X suivante est une densité de X :

$$f_X(x) = \begin{cases} \frac{x}{2} & \text{si } x \in [0\,;2], \\ 0 & \text{sinon.} \end{cases}$$

   Solution : À partir de la densité jointe, on peut calculer la densité f_Y de Y : pour tout $y \geq 0$,

$$f_Y(y) = \int_0^2 xe^{-2y}\,\mathrm{d}x = e^{-2y}\left[\frac{x^2}{2}\right]_0^2 = e^{-2y}\left(2^2 - 0^2\right) = 2e^{-2y}.$$

   Ainsi, Y suit la loi exponentielle de paramètre 2. La densité de X vaut 0 en-dehors de l'intervalle $[0\,;2]$, et pour $x \in [0\,;2]$,

$$f_X(x) = \int_0^\infty xe^{-2y}\,\mathrm{d}y = x\int_0^\infty e^{-2y}\,\mathrm{d}y = x\left[\frac{1}{-2}e^{-2y}\right]_0^{+\infty} = \frac{1}{2}x.$$

2. Justifier que X et Y sont indépendantes.
Solution : Pour tous réels $x \in [0\,;2]$ et $y \in [0\,;+\infty[$, on a
$$f_X(x)f_Y(y) = \frac{1}{2}x \times 2e^{-2y} = xe^{-2y} = f(x,y),$$
donc les variables aléatoires X et Y sont indépendantes.
3. Calculer $\mathbb{P}(X > 1)$, $\mathbb{P}(X < 1, Y > 2)$.
Solution : Par définition de la fonction de densité,
$$\mathbb{P}(X > 1) = \int_1^2 \frac{1}{2}x \,\mathrm{d}x = \left[\frac{1}{4}x^2\right]_1^2 = \frac{3}{4}.$$
Comme X et Y sont indépendantes, on a
$$\mathbb{P}(X < 1, Y > 2) = \mathbb{P}(X < 1)\,\mathbb{P}(Y > 2) = \left(1 - \frac{3}{4}\right)\mathbb{P}(Y > 2).$$
Enfin, comme $Y \sim \mathcal{E}(2)$,
$$\mathbb{P}(Y > 2) = e^{-2\times 2} = e^{-4},$$
donc $\mathbb{P}(X < 1, Y > 2) = \frac{e^{-4}}{4}$.
4. Calculer $\mathbb{E}[X]$, $\mathbb{V}\mathrm{ar}[X]$, $\mathbb{E}[Y]$ et $\mathbb{V}\mathrm{ar}[Y]$.
Solution :
$$\mathbb{E}\;[X] = \int_0^2 x \times \frac{1}{2}\,\mathrm{d}x = \left[\frac{1}{6}x^3\right]_0^2 = \frac{4}{3}, \tag{1}$$
$$\mathbb{E}\left[X^2\right] = \int_0^2 x^2 \times \frac{1}{2}x\,\mathrm{d}x = \left[\frac{1}{8}x^4\right]_0^2 = \frac{16}{4} = 2, \tag{2}$$
$$\mathbb{V}\mathrm{ar}[X] = 2 - \frac{16}{9} = \frac{2}{9}. \tag{3}$$
Comme $Y \sim \mathcal{E}(2)$, on a $\mathbb{E}[Y] = \frac{1}{2}$ et l'écart-type pour la loi exponentielle est égal à son espérance donc $\sigma[Y] = \frac{1}{2}$ donc $\mathbb{V}\mathrm{ar}[Y] = \frac{1}{4}$.
5. Calculer $\mathbb{E}[XY]$ et $\mathbb{E}[XY^2]$.
Solution : Par indépendance,
$$E[XY] = \mathbb{E}[X]\,\mathbb{E}[Y] = \frac{4}{3} \times \frac{1}{2} = \frac{2}{3}.$$
Toujours par indépendance et en utilisant que
$$\mathbb{E}\left[Y^2\right] = \mathbb{V}\mathrm{ar}[Y] + (\mathbb{E}[Y])^2,$$
$$\mathbb{E}\left[XY^2\right] = \mathbb{E}[X]\,\mathbb{E}\left[Y^2\right] = \frac{4}{3}\left(\frac{1}{4} + \left(\frac{1}{2}\right)^2\right) = \frac{2}{3}.$$

Exercice 4

Soit $(X_1, X_2, \ldots, X_n)$ un échantillon d'une variable aléatoire de loi $\mathcal{N}(m, \sigma^2)$, $m = 9$, les paramètres m et σ^2 étant inconnus.

1. Quelle est la loi de $\overline{X} = \frac{X_1 + \cdots + X_n}{n}$? (Préciser les paramètres de la loi).
Solution : Une combinaison linéaire de variables aléatoires normales *indépendantes* est encore normale. Pour connaître ses paramètres, il suffit de calculer son espérance et sa variance. Par le même calcul qu'à la question 4 de l'exercice 1, on a

$$\mathbb{E}\left[\overline{X}\right] = \mathbb{E}\left[X\right] = m,$$
$$\mathbb{V}\text{ar}\left[\overline{X}\right] = \frac{\sigma^2}{n}.$$

2. Supposons que les données de l'échantillon sont :

$$2,62; 4,78; 1,48; 0.20; 0,16; 3.98; 2,65; 2,61; 5,84.$$

Déterminer une estimation de m et σ^2. Ces estimateurs sont-ils sans biais ? Calculer le biais et la variance de l'estimateur de m. Quelle est la loi de l'estimateur de m ?
Solution : On sait que la moyenne empirique est un estimateur sans biais de l'espérance m, donc ici, la réalisation de $\overline{X}$ est $\frac{24,32}{9}$, ce qui donne comme estimation de m, $2,7022$. La variance de cet estimateur est $\frac{\sigma^2}{9}$ car la loi de $\overline{X}$ est $\mathcal{N}\left(m, \frac{\sigma^2}{9}\right)$.
Un estimateur de σ^2 est

$$\widehat{s_n}^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \overline{X}^2$$

Ici, une estimation de σ^2 est donc

$$10,64 - 2,7022^2 = 3,3369.$$

L'estimateur précédent est biaisé. Sa version sans biais est

$$\widehat{\sigma_n}^2 = \frac{n}{n-1}\widehat{s_n}^2.$$

Aussi, une réponse qui était aussi acceptée est $3,7541$.