UQÀM

Devoir 1 - Éléments de correction

STT 1000 - Automne 2025

Devoir à rendre le Mercredi 1er Octobre 2025 à 9h00 (au début du cours magistral)
Un rendu manuscrit par personne étudiante

Exercice 1 (40%).

1. Calculez (en détaillant) les fonctions de répartition:

   (a) D'une variable aléatoire de loi $\mathscr{U}(\{0,\dots,n\})$, $n \in \mathbb{N}$;

   Solution Comme le support de la loi est $\{\mathbf{0}, 1, \dots, n\}$, on a que pour $X \sim \mathscr{U}(\{0,\dots,n\})$:

   $$\mathbb{P}(X = k) = \frac{1}{n+1}\mathbb{1}_{\{0,\dots,n\}}(k).$$

   Donc, la fonction de répartition est donnée par, pour $k \in \{0,\dots,n\}$:

   $$F_X(k) = \sum_{j=0}^{k} \mathbb{P}(X = j) = \frac{k+1}{n+1}.$$

   Ainsi, pour $t \in \mathbb{R}$, on a:

   $$F_X(t) = \begin{cases} 0 & \text{si } t < 0, \\ \frac{\lfloor t \rfloor + 1}{n+1} & \text{si } 0 \le t \le n, \\ 1 & \text{si } t > n. \end{cases}$$

   Car X ne peut prendre que des valeurs entières. Ainsi, si t n'est pas entier (par ex. $t = 3.7$), il faut sommer jusqu'à l'entier le plus grand qui soit inférieur où égal à t, c'est-à-dire la partie entière inférieure $\lfloor t \rfloor$.

   (b) D'une variable aléatoire de loi $\mathscr{U}([a,b])$, $a, b \in \mathbb{R}$;

   Solution La densité d'une loi $\mathscr{U}([a,b])$ est $f(t) = \frac{1}{b-a}\mathbb{1}_{[a,b]}(t)$. Ainsi, sa fonction de répartition est donnée par:

   $$F(t) = \begin{cases} 0 & \text{si } t < a, \\ \int_a^t f(t)dt = \frac{t-a}{b-a} & \text{si } a \le t \le b, \\ 1 & \text{si } t > b \end{cases}$$

   (c) D'une variable aléatoire de loi $\mathscr{E}(\lambda)$, $\lambda > 0$.

Solution La densité d'une loi $\mathscr{E}(\lambda)$ est $f(t) = \lambda e^{-\lambda x} \mathbb{1}_{[0,\infty)}(t)$. Ainsi, sa fonction de répartition est donnée par:

$$F(t) = \begin{cases} 0 & \text{si } t < 0, \\ \int_a^t f(t)dt = 1 - e^{-\lambda t} & \text{si } t \geq 0, \end{cases}$$

Rappel: La fonction de répartition d'une variable aléatoire X est définie par:

$$F_X(t) = \mathbb{P}(X \leq t) = \begin{cases} \int_{-\infty}^t f_X(t)dt, \text{si } X \text{ est continue,} \\ \sum_{0 \leq j \leq t} \mathbb{P}(X = j), \text{si } X \text{ est discrète.} \end{cases}$$

2. Soit X une variable aléatoire continue de fonction de répartition F_X. On définit:

$$T = F_X(X).$$

(a) Quel est le support de la variable aléatoire T ?

Solution F_X est une fonction de répartition et ne peut renvoyer que des valeurs dans $[0, 1]$. Donc, la variable aléatoire $T = F_X(X)$ ne peut produire que des observations entre $[0, 1]$. Son support est donc $[0, 1]$.

(b) Quelle est sa fonction de répartition ?
Rappel:

$$\mathbb{P}(F_X(X) \leq t) = \mathbb{P}(X \leq F_X^{-1}(t))$$

Et comme X est continue, sa fonction de quantile $F_X^{-1}(t) = Q_X(t)$ est bien définie.

Solution Pour $t \in [0, 1]$, on a:

$$F_T(s) = \mathbb{P}(T \leq s) = \mathbb{P}(F_X(X) \leq s)$$

Comme X est continue, F_X admet une unique inverse F_X^{-1}. Donc,

$$\begin{aligned} F_T(s) &= \mathbb{P}(F_X(X) \leq s) \\ &= \mathbb{P}(X \leq F_X^{-1}(s)) \\ &= F_X\left(F_X^{-1}(s)\right) = s \end{aligned}$$

(c) En déduire sa loi.

Solution Comme la fonction de répartition caractérise la loi d'une variable aléatoire, et que l'on reconnaît la fonction de répartition d'une loi Uniforme de paramétrée par l'intervalle $[0, 1]$ (Question (b)), on peut conclure que:

$$X \sim \mathscr{U}([0, 1]).$$

3. Soit $X \sim \text{Bin}(n, p)$. On souhaite approcher cette loi en utilisant le **Théorème de la limite centrale (TLC)**.

(a) Rappelez par quelle loi peut-on approcher celle de X, et donner ses paramètres.

Solution Si $X \sim \mathcal{B}(n,p)$, alors :

$$X \approx \mathcal{N}\big(\mu, \sigma^2\big), \quad \mu = np, \quad \sigma^2 = np(1-p).$$

(b) Supposons que $p = 0.4$ et $n = 30$.

i. Vérifiez que nous sommes bien dans le régime asymptotique.

Solution Critères du cours:

$$np - 3\sqrt{np(1-p)} > 0 \quad \text{et} \quad n > np + 3\sqrt{np(1-p)}.$$

Ces deux critères sont respectés, on peut conclure que nous sommes dans le régime asymptotique.

ii. A l'aide d'une table de probabilités, déterminez le plus petit entier a tel que

$$\mathbb{P}(X \leq a) \approx 0.9505$$

à l'aide du TLC.

Solution On cherche a tel que $\mathbb{P}(X \leq a) \approx 0.9505$. On a que:

$$\begin{aligned}
\mathbb{P}(X \leq a) &= \mathbb{P}\left(\frac{X-np}{\sqrt{p(1-p)}} \leq \frac{a-np}{\sqrt{p(1-p)}}\right) \\
&\simeq \mathbb{P}\left(Z \leq \frac{a+0.5-np}{\sqrt{p(1-p)}}\right) && \text{avec } Z \sim \mathcal{N}(0,1) \\
&= \Phi\left(\frac{a+0.5-np}{\sqrt{p(1-p)}}\right)
\end{aligned}$$

Or, d'après la table de probabilité, on a que $\Phi(t) = 0.9505$ si $t = 1,65$. Ainsi, on a l'équation:

$$\frac{a+0.5-np}{\sqrt{p(1-p)}} = 1,65 \quad \Longrightarrow \quad a = \left(1,65 \times \sqrt{p(1-p)}\right) + np - 0.5$$

En replaçant avec les valeurs, on tombe sur un nombre dont l'entier le plus proche est 16, et on accepte $a = 16$ comme solution.

(c) Supposons maintenant que $n = 30$ et $a = 7$. A l'aide d'une table de probabilités, déterminez approximativement la valeur de p pour que

$$\mathbb{P}(X \leq a) \approx 0.9505$$

à l'aide du TLC.

Solution Avec les mêmes arguments que ci-dessus, on retombe sur l'équation

$$\frac{a+0.5-np}{\sqrt{p(1-p)}} = 1,65 \quad \Longrightarrow \quad \underbrace{(n^2 + 1.65^2)}_{\alpha} p^2 - \underbrace{(2n(a+0.5) + 1.65^2)}_{\beta} p + \underbrace{(a+0.5)^2}_{\gamma} = 0.$$

> On a donc un polynôme du second degré de la forme $\alpha \times p^2 + \beta \times p + \gamma$ à annuler.
> A l'aide du déterminant, on trouve deux solutions, dont une qui est inadmissible, car renvoie $-1,65$ au problème initial.

4. Soient X et Y deux variables aléatoires, et $a, b \in \mathbb{R}$.

   (a) Calculez

   $$\text{Cov}(X+a, Y+b) \quad \text{et} \quad \text{Cov}(X-a, Y-b)$$

   > **Solution** On se rend compte que $\text{Cov}(X+a, Y+b) = \text{Cov}(X-a, Y-b) = \text{Cov}(X, Y)$

   (b) Que pouvez-vous conclure ?

   > **Solution** Ajouter ou retirer des constantes à deux variables aléatoires ne change pas leur covariance. Autrement dit, **la covariance est invariante par translation**.

Exercice 2 (30%). On considère l'échantillon observé suivant :

$$\boxed{8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 40, 60}$$

On soupçonne que la valeur maximale 60 est aberrante.

1. Calculer la médiane, ainsi que les quartiles Q_1 et Q_3.

   > **Solution** En reprenant **la formule du cours**, on retrouve:
   >
   > $$M = 19, \quad Q_1 = 12.5, \quad Q_3 = 25.5$$

2. Déterminer l'écart interquartile. En déduire l'intervalle de valeur non-aberrantes selon la règle "$1,5 \times \text{IQR}$".

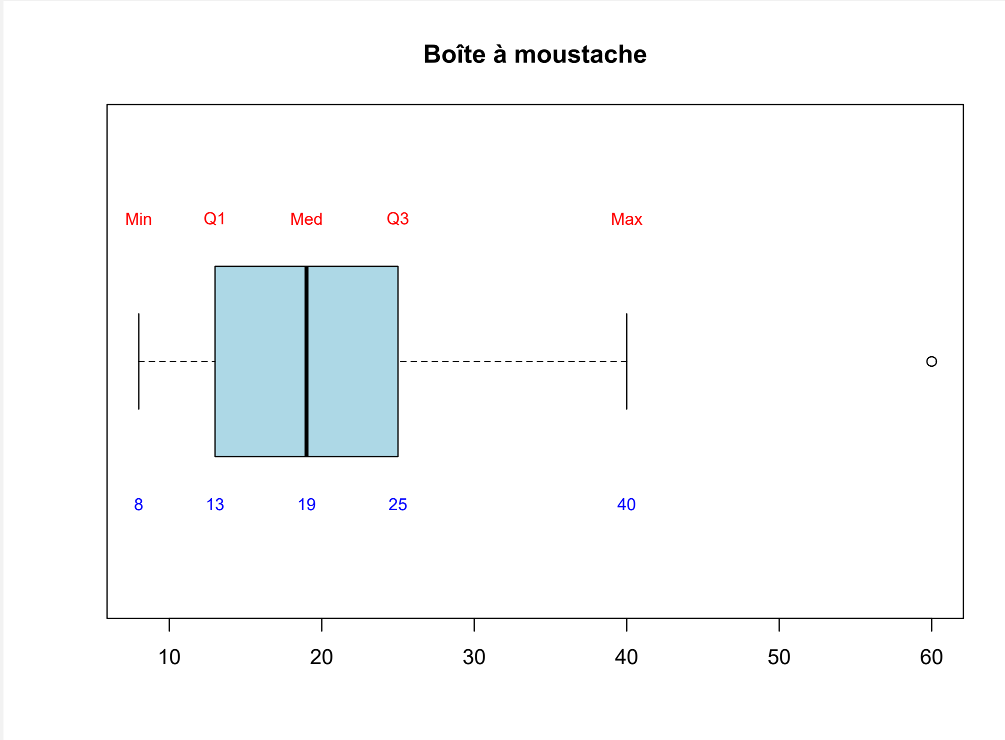
   > **Solution** L'écart inter-quartile est donné par $I_{IQ} = Q_3 - Q_1 = 13$, et l'intervalle de valeur non-aberrantes est
   >
   > $$[Q_1 - 1.5 \times I_{IQ}, Q_3 + 1.5 \times I_{IQ}] = [-7, 45].$$

3. Identifier les valeurs aberrantes de l'échantillon, et déterminer si 60 est effectivement aberrante.

   > **Solution** La seule valeur aberrante de cet échantillon observé est 60.

4. Tracer la boîte à moustache correspondant à cet échantillon observé.

   > **Solution**



Exercice 3 (30%)**.** Voici une série de six températures enregistrées en Arizona et exprimées en degrés Fahrenheit :

$$x = (x_1, x_2, x_3, x_4, x_5, x_6) = (50°F,\ 59°F,\ 68°F,\ 77°F,\ 86°F,\ 95°F)$$

1. Calculez la moyenne $\overline{x}$ et la variance s_x^2 de cet échantillon observé.

> **Solution** On trouve:
> $$\overline{x}_n = 72.5, \quad \text{et } s_x^2 = 238.5$$

2. Exprimez les températures en degrés Celsius à l'aide de la transformation

$$y_i = -\frac{160}{9} + \frac{5}{9}x_i,$$

et donnez le nouvel échantillon $y = (y_1, y_2, y_3, y_4, y_5, y_6)$.

> **Solution** On retrouve $y = (10, 15, 20, 25, 30, 35)$.

3. Calculez $\overline{y}$ et la variance s_y^2 du nouvel échantillon observé :
   (a) A partir des données exprimées en degrés Celsius;
   (b) En utilisant les formules $\overline{y} = a + b\overline{x}$ et $s_y^2 = b^2 s_x^2$ lorsque $y = a + bx$.

> **Solution** Dans les deux cas, on retombe sur le même résultat:
> $$\overline{y}_n = 22.5, \quad \text{et } s_y^2 = 87.5$$

4. Centrez et réduisez les deux échantillons observés x et y, et vérifiez que les deux échantillons observés résultants sont identiques. On notera $z = (z_1, z_2, z_3, z_4, z_5, z_6)$ cet échantillon observé centré-réduit.

5. Calculez la moyenne $\overline{z}$ et l'écart-type $s_z$de z.

> **Solution** La moyenne de l'échantillon centré-réduit est 0, et on écart-type est 1.

> **Solution** Solution du troisième exercice
