CORRIGÉ 14

Exercice 1. Définissons d'abord les événements

$$A = \{\text{on joue avec le dé A}\}, \quad B = \{\text{on joue avec le dé B}\},$$

$$R = \{\text{le jet donne rouge}\}, \quad R_j = \{\text{le jème jet donne rouge}\}.$$

(i). Par les probabilités totales,

$$\Pr(R) = \Pr(R|A)\Pr(A) + \Pr(R|B)\Pr(B) = \frac{2}{3}\cdot\frac{1}{2} + \frac{1}{3}\cdot\frac{1}{2} = \frac{1}{2}.$$

(ii). Par définition des probabilités conditionnelles et la formule des probabilités totales, on obtient

$$\begin{aligned}\Pr(R_3|R_1\cap R_2) = \frac{\Pr(R_1\cap R_2\cap R_3)}{\Pr(R_1\cap R_2)} &= \frac{\Pr(R_1\cap R_2\cap R_3|A)\Pr(A) + \Pr(R_1\cap R_2\cap R_3|B)\Pr(B)}{\Pr(R_1\cap R_2|A)\Pr(A) + \Pr(R_1\cap R_2|B)\Pr(B)} \\ &= \frac{(2/3)^3\cdot(1/2) + (1/3)^3\cdot(1/2)}{(2/3)^2\cdot(1/2) + (1/3)^2\cdot(1/2)} = \frac{3}{5}.\end{aligned}$$

Notons que nous avons utilisé le fait que les événements R_1, R_2 et R_3 sont indépendants conditionnellement à A et à B. Nous utilisons de nouveau implicitement le même type d'argument dans les questions suivantes.

(iii). Par définition des probabilités conditionnelles et la formule des probabilités totales,

$$\begin{aligned}\Pr(R_3|R_1\cup R_2) &= \frac{\Pr((R_1\cup R_2)\cap R_3)}{\Pr(R_1\cup R_2)} \\ &= \frac{\Pr((R_1\cup R_2)\cap R_3|A)\Pr(A) + \Pr((R_1\cup R_2)\cap R_3|B)\Pr(B)}{\Pr(R_1\cup R_2|A)\Pr(A) + \Pr(R_1\cup R_2|B)\Pr(B)} \\ &= \frac{\Pr((R_1^c\cap R_2^c)^c\cap R_3|A)\Pr(A) + \Pr((R_1^c\cap R_2^c)^c\cap R_3|B)\Pr(B)}{\Pr((R_1^c\cap R_2^c)^c|A)\Pr(A) + \Pr((R_1^c\cap R_2^c)^c|B)\Pr(B)} \\ &= \frac{(1-(\frac{1}{3})^2)\cdot\frac{2}{3}\cdot\frac{1}{2} + (1-(\frac{2}{3})^2)\cdot\frac{1}{3}\cdot\frac{1}{2}}{(1-(\frac{1}{3})^2)\cdot\frac{1}{2} + (1-(\frac{2}{3})^2)\cdot\frac{1}{2}} = \frac{7}{13}.\end{aligned}$$

(iv). Par la formule de Bayes, on a

$$\Pr(A|R_1\cap R_2) = \frac{\Pr(R_1\cap R_2|A)\Pr(A)}{\Pr(R_1\cap R_2|A)\Pr(A) + \Pr(R_1\cap R_2|B)\Pr(B)} = \frac{(\frac{2}{3})^2\cdot\frac{1}{2}}{(\frac{2}{3})^2\cdot\frac{1}{2} + (\frac{1}{3})^2\cdot\frac{1}{2}} = \frac{4}{5}.$$

(v). Par la formule de Bayes,

$$\begin{aligned}\Pr(A|R_1\cup R_2) &= \frac{\Pr(R_1\cup R_2|A)\Pr(A)}{\Pr(R_1\cup R_2|A)\Pr(A) + \Pr(R_1\cup R_2|B)\Pr(B)} \\ &= \frac{\Pr((R_1^c\cap R_2^c)^c|A)\Pr(A)}{\Pr((R_1^c\cap R_2^c)^c|A)\Pr(A) + \Pr((R_1^c\cap R_2^c)^c|B)\Pr(B)} \\ &= \frac{(1-(\frac{1}{3})^2)\cdot\frac{1}{2}}{(1-(\frac{1}{3})^2)\cdot\frac{1}{2} + (1-(\frac{2}{3})^2)\cdot\frac{1}{2}} = \frac{8}{13}.\end{aligned}$$

(vi). Par les probabilités totales,

$$\Pr(R_1 \cap R_2) = \Pr(R_1 \cap R_2|A)\Pr(A) + \Pr(R_1 \cap R_2|B)\Pr(B) = (2/3)^2 \cdot (1/2) + (1/3)^2 \cdot (1/2) = 5/18.$$

En (a), on a trouvé $\Pr(R_1) = \Pr(R_2) = 1/2$. Ainsi, $\Pr(R_1 \cap R_2) \neq \Pr(R_1) \cdot \Pr(R_2)$, ce qui montre que les événements R_1 et R_2 ne sont pas indépendants. Cependant, comme indiqué précédemment, ils sont indépendants conditionnellement à A et à B : $\Pr(R_1 \cap R_2|A) = \Pr(R_1|A) \cdot \Pr(R_2|A)$ et $\Pr(R_1 \cap R_2|B) = \Pr(R_1|B) \cdot \Pr(R_2|B)$.

Exercice 2. (i). On a

$$1 = \int_{-\infty}^{\infty} f_X(x)dx = \int_1^{\infty} \frac{c}{x^{1+\alpha}}dx = \frac{c}{\alpha},$$

donc $c = \alpha$.

(ii).

$$F_X(x) = \int_{-\infty}^{x} f_X(t)dt = \begin{cases} \int_1^x \frac{\alpha}{t^{\alpha+1}} = 1 - x^{-\alpha} & \text{pour } x > 1, \\ 0 & \text{pour } x \leq 1. \end{cases}$$

(iii).

$$\Pr(X > bz | X > z) = \frac{\Pr(X > bz, X > z)}{\Pr(X > z)} = \begin{cases} \frac{\Pr(X>z)}{\Pr(X>z)} = 1 & \text{pour } b \leq 1, \\ \frac{\Pr(X>bz)}{\Pr(X>z)} = \frac{(bz)^{-\alpha}}{z^{-\alpha}} = b^{-\alpha} & \text{pour } b > 1. \end{cases}$$

(iv).

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x f_X(x)dx = \int_1^{\infty} \alpha x^{-\alpha}dx = \begin{cases} \frac{\alpha}{\alpha-1} & \text{pour } \alpha > 1, \\ \infty & \text{pour } \alpha \leq 1. \end{cases}$$

(v).

$$\mathbb{E}[X^2] = \int_{-\infty}^{\infty} x^2 f_X(x)dx = \int_1^{\infty} \alpha x^{-\alpha+1}dx = \frac{\alpha}{\alpha - 2},$$

alors

$$\mathrm{Var}[X] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \frac{\alpha}{(\alpha-2)(\alpha-1)^2}.$$

Exercice 3. Pour les calculs suivants, il est utile de dessiner l'ensemble où la densité conjointe est non-nulle. Il s'agit du triangle borné par les droites $x = 1$, $z = 1$ et $x + z = 1$.

(i). On a

$$\begin{aligned} 1 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Z}(x,z)\mathrm{d}x\ \mathrm{d}z &= \int_0^1 \left(\int_{1-x}^1 \frac{x+z-1}{c}\mathrm{d}z \right) \mathrm{d}x \\ &= \int_0^1 \frac{(x-1)x + \frac{1-(1-x)^2}{2}}{c}\mathrm{d}x \\ &= \int_0^1 \frac{x^2}{2c}\mathrm{d}x \\ &= \frac{1}{6c}. \end{aligned}$$

Ainsi, on obtient $c = \frac{1}{6}$.

(ii). Non. En effet, la densité conjointe ne peut pas être écrite comme le produit de deux fonctions car l'ensemble où la densité conjointe est non-nulle n'est pas un rectangle (ou l'union de rectangles). Cela est corroboré par notre intuition : on voit bien, par exemple, que l'événement $X = x$ fournit beaucoup d'information sur Z puisque cela impose que $Z > 1 - x$.

(iii). On a

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Z}(x,z)\mathrm{d}z = \begin{cases} 0 & \text{pour } x \leq 0 \text{ ou } x \geq 1, \\ \int_{1-x}^{1} 6(x+z-1)\mathrm{d}z = 3x^2 & \text{pour } x \in (0,1). \end{cases}$$

Etant donné que les variables X et Z jouent un rôle symétrique, on en déduit que $f_Z(z) = 3z^2$ pour $z \in (0,1)$ et $f_Z(z) = 0$ sinon.

(iv). On a

$$\begin{aligned} \mathbb{E}[XZ] &= \int_{-\infty}^{\infty} xz f_{X,Z}(x,z)\mathrm{d}x \ \mathrm{d}z \\ &= \int_0^1 \left(\int_{1-x}^1 6xz(x+z-1)\mathrm{d}z \right) \mathrm{d}x \\ &= \int_0^1 \left[3x(x-1)(1-(1-x)^2) + 2x(1-(1-x)^3) \right] \mathrm{d}x = \cdots = \frac{11}{20}. \end{aligned}$$

Par ailleurs,

$$\mathbb{E}[X] = \mathbb{E}[Z] = \int_{-\infty}^{\infty} x f_X(x)\mathrm{d}x = \int_0^1 3x^3 \mathrm{d}x = \frac{3}{4}.$$

Finalement, on obtient

$$\mathrm{Cov}(X,Z) = \mathbb{E}[XZ] - \mathbb{E}[X]\mathbb{E}[Z] = -\frac{1}{80}.$$

(v). La formule générale est

$$F_{X,Z}(x,z) = \Pr(X \leq x, Z \leq z) = \int_{-\infty}^{x} \int_{-\infty}^{z} f_{X,Z}(u,v)\mathrm{d}u \ \mathrm{d}v.$$

On a écrit l'intégrande avec les variables u et v car les lettres x et z sont déjà utilisées dans les bornes d'intégration. Il faut calculer cette intégrale séparément pour les différentes valeurs de x et z. Pour $x < 0$ ou $z < 0$ ou $x + z < 1$, on a $F_{X,Z}(x,z) = 0$ car on intègre 0. Pour x et z tels que $x < 1$, $z \leq 1$ et $x + z > 1$, on a

$$\begin{aligned} F_{X,Z}(x,z) &= \int_{1-z}^{x} \left(\int_{1-u}^{z} 6(u+v-1)\mathrm{d}v \right) \mathrm{d}u \\ &= \int_{1-z}^{x} \left[6(u-1)(z-1+u) + 3z^2 - 3(1-u)^2 \right] \mathrm{d}u \\ &= x^3 - 1 + 3z - 3z^2 + z^3 + 3zx^2 - 3x^2 + 3z^2x + 3x - 6zx. \end{aligned}$$

Pour $x \geq 1$ et $z \in [0,1]$, $F_{X,Z}(x,z) = F_{X,Z}(1,z)$. Pour $z > 1$ et $x \in [0,1)$, $F_{X,Z}(x,z) = F_{X,Z}(x,1)$. Finalement, $F_{X,Z}(x,z) = 1$ pour $x \geq 1$ et $z > 1$.

(vi). On a

$$F_X(x) = \int_{-\infty}^{x} f_X(u)\mathrm{d}u = \begin{cases} 0 & \text{si } x \leq 0, \\ \int_0^x 3u^2 \mathrm{d}u = x^3 & \text{si } x \in (0,1), \\ 1 & \text{si } x \geq 1. \end{cases}$$

Par ailleurs, on a $F_Z(z) = F_X(z)$. Rappelons qu'il est également possible de déterminer ces fonctions de répartition marginales en utilisant les formules $F_X(x) = \lim_{z\to\infty} F_{X,Z}(x,z)$ et $F_Z(z) = \lim_{x\to\infty} F_{X,Z}(x,z)$. Ceci donne évidemment le même résultat.

(vii). On a

$$\Pr\left(X \leq \frac{3}{4} \middle| Z \leq \frac{3}{4}\right) = \frac{F_{X,Z}\left(\frac{3}{4}, \frac{3}{4}\right)}{F_Z\left(\frac{3}{4}\right)} = \frac{8}{27}.$$

(viii). Pour x tel que $f_X(x) > 0$, i.e. pour $x \in (0,1)$, on a

$$f_{Z|X}(z|x) = \frac{f_{X,Z}(x,z)}{f_X(x)} = \begin{cases} \frac{6(x+z-1)}{3x^2} & \text{si } z \in (1-x, 1], \\ 0 & \text{sinon.} \end{cases}$$

Exercice 4. (i). Etant donné que la densité est symétrique autour de 0, on obtient directement $\mathbb{E}[X] = 0$. On a donc

$$\mathrm{Var}(X) = \mathbb{E}\left[X^2\right] = \int_{-1}^{1} \frac{3}{4} x^2 (1 - x^2) \mathrm{d}x = \cdots = \frac{1}{5}.$$

En utilisant le TCL, on obtient

$$\begin{aligned} \Pr(-5 \leq X_1 + \cdots + X_{100} \leq 5) &= \Pr\left(\frac{-5}{\sqrt{\frac{1}{5}}\sqrt{100}} \leq \frac{X_1 + \cdots + X_{100}}{\sqrt{\frac{1}{5}}\sqrt{100}} \leq \frac{5}{\sqrt{\frac{1}{5}}\sqrt{100}}\right) \\ &\approx 2\Phi\left(\frac{\sqrt{5}}{2}\right) - 1 \\ &= 0.736. \end{aligned}$$

(ii). On cherche x tel que

$$\Pr(X_1 + \cdots + X_{100} > x) = 0.1.$$

En utilisant le TCL, on obtient

$$\Pr(X_1 + \cdots + X_{100} > x) = 1 - \Pr\left(\frac{X_1 + \cdots + X_{100}}{\sqrt{\frac{1}{5}}\sqrt{100}} < \frac{x}{\sqrt{\frac{1}{5}}\sqrt{100}}\right) \approx 1 - \Phi\left(\frac{x}{\sqrt{\frac{1}{5}}\sqrt{100}}\right).$$

Ainsi, on cherche x tel que

$$1 - \Phi\left(\frac{x\sqrt{5}}{10}\right) = 0.1,$$

ce qui donne $x = \Phi^{-1}(0.9)\frac{10}{\sqrt{5}} = 5.73$.

Exercice 5. (i). La vraisemblance pour le paramètre α est donnée par

$$L(\alpha) = \prod_{i=1}^{n} \frac{\alpha}{x_i^{1+\alpha}}.$$

Ainsi,

$$\ell(\alpha) = \log(L(\alpha)) = n\log(\alpha) - \sum_{i=1}^{n}(1+\alpha)\log(x_i).$$

On a donc

$$\ell'(\alpha) = \frac{n}{\alpha} - \sum_{i=1}^{n}\log(x_i),$$

ce qui donne

$$\ell'(\alpha) = 0 \Leftrightarrow \alpha = \frac{n}{\sum_{i=1}^{n}\log(x_i)}.$$

Etant donné que $\ell''(\alpha) = -\frac{n}{\alpha^2}$, on a que $\ell''(\alpha) < 0$ pour tout $n \in \mathbb{N}\backslash\{0\}$ et $\alpha > 0$. Ainsi, l'estimateur du maximum de vraisemblance est donné par $\hat{\alpha}_{ML} = \frac{n}{\sum_{i=1}^{n}\log(x_i)}$.

(ii). Le Théorème 5 du cours nous donne que $\hat{\alpha}_{ML} \approx \mathcal{N}\left(\alpha, J(\hat{\alpha}_{ML})^{-1}\right)$, où $J(\alpha) = -\ell''(\alpha) = \frac{n}{\alpha^2}$, $\alpha > 0$, et $\approx$ signifie "suit approximativement". Finalement, on obtient que $\hat{\alpha}_{ML} \approx \mathcal{N}\left(\alpha, \frac{\hat{\alpha}_{ML}^2}{n}\right)$.

(iii). Le même théorème nous donne qu'un intervalle de confiance au niveau approximatif de 95% est :

$$\left[\hat{\alpha}_{ML} - \Phi^{-1}(0.975)\sqrt{J(\hat{\alpha}_{ML})^{-1}}, \hat{\alpha}_{ML} + \Phi^{-1}(0.975)\sqrt{J(\hat{\alpha}_{ML})^{-1}}\right],$$

c'est-à-dire

$$\left[\hat{\alpha}_{ML} - 1.96\frac{\hat{\alpha}_{ML}}{\sqrt{n}}, \hat{\alpha}_{ML} + 1.96\frac{\hat{\alpha}_{ML}}{\sqrt{n}}\right].$$

(iv). Pour $\alpha > 1$, on sait d'après le (iv) de l'Exercice 2 que $\mathbb{E}[X] = \frac{\alpha}{\alpha-1}$. Ainsi, $\hat{\alpha}_{MOM}$ vérifie

$$\frac{\alpha}{\alpha - 1} = \bar{X},$$

i.e.

$$\hat{\alpha}_{MOM} = \frac{\bar{X}}{\bar{X} - 1}.$$

Exercice 6. On a un échantillon $X_1, \ldots, X_n$ d'une distribution normale d'espérance μ et de variance σ^2. On s'intéresse au paramètre μ.

(i). Il faut tester l'hypothèse $H_0 : \mu = \mu_0$ contre l'alternative $H_1 : \mu \neq \mu_0$, où $\mu_0 = 38\%$. On a $X_1, \ldots, X_n \stackrel{iid}{\sim} \mathcal{N}(\mu, \sigma^2)$ où σ^2 est inconnue. En notant S_n l'estimateur classique de σ, on utilise donc la statistique

$$T = \sqrt{n}\frac{\bar{X}_n - \mu_0}{S_n},$$

qui, d'après le cours, suit la distribution t_{n-1} sous H_0. Ici, on souhaite effectuer un test au niveau de signification de 1%. Ainsi, on rejette H_0 si $|T| > s_1(t_{n-1})$, où $s_1(t_{n-1})$ est tel que $\Pr(|t_{n-1}| > s_1(t_{n-1})) = \frac{1}{100}$. Ici, on rejette donc H_0 si

$$\left|\sqrt{183}\frac{\bar{X}_{183} - \mu_0}{S_{183}}\right| > s_1(t_{182}).$$

On a

$$\sqrt{183}\frac{\bar{x}_{183} - 0.38}{s_{183}} = -3.517 \quad \text{et} \quad s_1(t_{182}) = 2.603.$$

On rejette donc H_0 et on peut dire, avec une probabilité d'erreur de 1%, que le vrai contenu d'alcool moyen μ est différent de la valeur déclarée. Compte tenu du nombre élevé de degrés de liberté, dans la plupart des tables de la loi de Student, la valeur de $s_1(t_{182})$ n'est pas accessible. En revanche, celle-ci s'obtient immédiatement avec un logiciel adapté aux statistiques comme R.

(ii). En utilisant le fait que

$$\sqrt{n}\frac{\bar{X}_n - \mu}{S_n} \sim t_{n-1},$$

on obtient que

$$\left[\bar{X}_n - s_1(t_{n-1})\frac{S_n}{\sqrt{n}}, \bar{X}_n - s_1(t_{n-1})\frac{S_n}{\sqrt{n}}\right]$$

est un intervalle de confiance au niveau 99%. Cet intervalle se réécrit dans notre cas

$$\left[\bar{X}_{183} - s_1(t_{182})\frac{S_{183}}{\sqrt{183}}, \bar{X}_{183} - s_1(t_{182})\frac{S_{183}}{\sqrt{183}}\right],$$

et sa réalisation vaut $[37.774\%, 37.966\%]$.

Remarque 1. Si on n'avait pas supposé que la distribution des X_i était normale, on aurait pu procéder de la même façon pour (i) et (ii). En effet, comme n est grand, il est possible de montrer en utilisant le TCL (démonstration hors programme et omise ici) que

$$\sqrt{n}\frac{\bar{X}_n - \mu}{S_n} \approx t_{n-1}.$$

Mais, contrairement au cadre de l'exercice, il s'agit alors d'une approximation. Le niveau du test et celui de l'intervalle de confiance ne seraient donc pas exactement de 1% et 99% respectivement.

Exercice 7. Voici le corrigé adapté de Cantoni et al. (2009, ex. 8.40) :

Si un dé est régulier, la probabilité d'obtenir un 6 est égale à 1/6. Soit Y, la variable aléatoire qui décrit le nombre de 6 lors du lancer des 3 dés. Y est distribuée selon la loi binomiale de paramètres 3 et 1/6. Sous l'hypothèse nulle que les dés sont réguliers, on a les probabilités suivantes :

$$
\begin{aligned}
\mathrm{pr}(Y=0) &= \binom{3}{0}\left(\frac{1}{6}\right)^0\left(\frac{5}{6}\right)^3 = 0.579,\\
\mathrm{pr}(Y=1) &= \binom{3}{1}\left(\frac{1}{6}\right)^1\left(\frac{5}{6}\right)^2 = 0.347,\\
\mathrm{pr}(Y=2) &= \binom{3}{2}\left(\frac{1}{6}\right)^2\left(\frac{5}{6}\right)^1 = 0.069,\\
\mathrm{pr}(Y=3) &= \binom{3}{3}\left(\frac{1}{6}\right)^3\left(\frac{5}{6}\right)^0 = 0.005.
\end{aligned}
$$

Puisque le joueur a joué 100 fois, les fréquences espérées pour $Y = 0, 1, 2$ et 3 sont 57.9, 34.7, 6.9 et 0.5 respectivement. La statistique du test de χ^2 d'ajustement vaut 24.9 qui est à comparer au quantile $\chi^2_{3;0.95} = 7.815$. On a donc une forte évidence pour rejecter H_0 et affirmer que les dés sont truqués.

Rappel : Le coefficient binomial est donné par $\binom{j}{i} = \frac{j!}{i!(j-i)!}$.