Corrigé 13

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,

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

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

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

(iv). Par la formule de Bayes,

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

(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 $\Pr(R_1) = \Pr(R_2) = 1/2$. Alors $\Pr(R_1\cap R_2) \neq \Pr(R_1)\cdot\Pr(R_2)$, donc les événements ne sont pas indépendants (mais ils sont indépendants conditionnellement au dé que l'on utilise, en effet $\Pr(R_1\cap R_2|A) = \Pr(R_1|A)\cdot\Pr(R_2|A)$ 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 $x = 1$, $z = 1$, $x + z = 1$.

(i)

$$1 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{(X,Z)}(x,z)\,dx\,dz = \int_0^1 \left(\int_{1-x}^1 \frac{x+z-1}{c}dz\right)dx = \int_0^1 \frac{(x-1)x + \frac{1-(1-x)^2}{2}}{c}dx$$
$$= \int_0^1 \frac{x^2}{2c}dx = \frac{1}{6c},$$

donc $c = \frac{1}{6}$.

(ii) Non, parce que l'ensemble où la densité conjointe est non-nulle n'est pas un rectangle et donc la densité conjointe ne peut pas être écrite comme le produit de deux fonctions.

(iii)

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

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

(iv) Calculons

$$\mathbb{E}(XZ) = \int_{-\infty}^{\infty} xz f_{(X,Z)}(x,z)\,dx\,dz = \int_0^1 \left(\int_{1-x}^1 6xz(x+z-1)dz \right) dx$$
$$= \int_0^1 (3x(x-1)(1-(1-x)^2) + 2x(1-(1-x)^3))dx = \frac{11}{20}$$

et

$$\mathbb{E}\,X = \mathbb{E}\,Z = \int_{-\infty}^{\infty} x f_X(x)dx = \int_0^1 3x^3 dx = \frac{3}{4}.$$

Alors

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

(v) La formule générale est

$$F_{(X,Z)}(x,z) = P(X \le x, Z \le z) = \int_{-\infty}^{x} \int_{-\infty}^{z} f_{(X,Z)}(u,v)\,du\,dv$$

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

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

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

(vi) On peut calculer

$$F_X(x) = \int_{-\infty}^{x} f_X(u)du = \begin{cases} 0, & \text{si } x \le 0, \\ \int_0^x 3u^2 du = x^3, & \text{si } x \in (0,1), \\ 1, & \text{si } x \ge 1 \end{cases}$$

et $F_Z(z) = F_X(z)$. On peut également calculer $F_X(x) = F_{(X,Z)}(x,\infty)(\equiv \lim_{z\to\infty} F_{(X,Z)}(x,z))$ et $F_Z(z) = F_{(X,Z)}(\infty, z)$; ceci donne le même résultat.

(vii) On a

$$P(X \le \tfrac{3}{4} | Z \le \tfrac{3}{4}) = \frac{F_{(X,Z)}(\frac{3}{4}, \frac{3}{4})}{F_Z(\frac{3}{4})} = \frac{8}{27}.$$

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

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

Exercice 4.

(a) On voit que $\mathbb{E}\,X = 0$ (la densité est symétrique autour de 0) et

$$\mathsf{Var}\,X = \mathbb{E}(X^2) = \int_{-1}^{1} \frac{3}{4} x^2 (1 - x^2) dx = 1/5.$$

En utilisant le TCL, on obtient

$$P(-5 \leq X_1 + \cdots + X_{100} \leq 5) = P\left(\frac{-5}{\sqrt{100 \cdot 1/5}} \leq \frac{X_1 + \cdots + X_{100}}{\sqrt{100 \cdot 1/5}} \leq \frac{5}{\sqrt{100 \cdot 1/5}} \right)$$
$$\doteq 2\Phi(\sqrt{5}/2) - 1 = 0.736.$$

(b) On cherche x tel que

$$P(X_1 + \cdots + X_{100} > x) \doteq 0.1.$$

Approximons (en utilisant le TCL)

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

Il faut que

$$1 - \Phi\left(\frac{x}{\sqrt{100 \cdot 1/5}} \right) = 0.1.$$

Ceci donne

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

Exercice 5.

(i) La vraisemblance pour le paramètre α égale

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

son logarithme est

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

La dérivée de L égale

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

La solution de l'équation $L'(\alpha) = 0$ est

$$\hat{\alpha}_{\mathrm{ml}} = \frac{n}{\sum_{i=1}^{n} \log X_i}.$$

La deuxième dérivée de L,

$$L''(\alpha) = -\frac{n}{\alpha^2},$$

est négative, donc la fonction L est concave. Alors la solution $\hat{\alpha}_{\mathrm{ml}}$ est le maximum global de L.

(ii) Par Théorème 5 dans les notes du cours, la loi approximative (pour n grand) de $\hat{\alpha}$ est $\mathrm{N}(\alpha, J(\hat{\alpha}_{\mathrm{ml}})^{-1})$ où $J(\alpha) = -L''(\alpha) = n/\alpha^2$, donc $\hat{\alpha}_{\mathrm{ml}} \dot{\sim} \mathrm{N}(\alpha, \hat{\alpha}_{\mathrm{ml}}^2/n)$.

(iii) Par le même théorème, on obtient l'intervalle de confiance à 95% :

$$(\hat{\alpha}_{ml} - \Phi^{-1}(0.975)\sqrt{J(\hat{\alpha}_{ml})^{-1}}, \hat{\alpha}_{ml} + \Phi^{-1}(0.975)\sqrt{J(\hat{\alpha}_{ml})^{-1}})$$

(iv) Dans l'exercice 2 on a trouvé que $\mathbb{E}\, X = \frac{\alpha}{\alpha - 1}$. L'équation

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

donne

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

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

(a) Il faut tester l'hypothèse H_0: $\mu = \mu_0$ contre l'alternative H_0: $\mu \neq \mu_0$ où $\mu_0 = 38$. La statistique

$$T = \sqrt{n}\frac{\bar{X} - \mu_0}{\sigma}$$

mesure la distance entre les observations et l'hypothèse. Si la valeur de $|T|$ est significativement grande, on rejette H_0. La statistique T suit approximativement la loi $\mathrm{N}(0, 1)$ par le TCL. On ne connait pas l'écart-type σ mais on peut le remplacer par l'estimateur S, donc on utilise

$$T = \sqrt{n}\frac{\bar{X} - \mu_0}{S}.$$

L'approximation normale pour la loi de T reste valable. Donc le test rejette H_0 si $|T| > \Phi^{-1}(1 - \alpha/2)$. Numériquement on a

$$t_{obs} = \sqrt{183}\frac{37.87 - 38}{\sqrt{0.25}} = -3.517$$

et $\Phi^{-1}(1 - 0.01/2) = 2.576$. Alors on rejette l'hypothèse H_0 et on accepte l'alternative que le vrai contenu d'alcool est différent de 38.

Dans le calcul précédent on n'a pas supposé de forme concrète de la distribution de $X_1, \ldots, X_n$. On peut probablement supposer que la distribution soit normale car il s'agit d'une mesure de la valeur μ avec erreur aléatoire qui est peut-être symétriquement distribuée autour de 0 (donc plus ou moins normale) et l'écart-type 0.5 est relativement petit par rapport à la longueur de l'intervalle $[0, 100]$ où les observations peuvent se trouver. Alors, il faut tester H_0: $\mu = \mu_0$ contre H_1: $\mu \neq \mu_0$ pour un échantillon de n observations de la loi $\mathrm{N}(\mu, \sigma^2)$ avec σ^2. La statistique pour ce problème est de nouveau

$$T = \sqrt{n}\frac{\bar{X} - \mu_0}{S}$$

qui maintenant suit la loi de Student avec $n - 1$ degrés de liberté. On rejette H_0 et accepte H_1 quand $|T| > t_{n-1;1-\alpha/2}$. Finalement, on voit que la seule différence entre cette méthode et la méthode basée sur le TCL est le quantile que l'on utilise. Mais quand le nombre d'observations est grand, les deux quantiles sont très proches (la loi normale $\mathrm{N}(0, 1)$ est la limite de la loi de Student t_k pour $k \to \infty$). Le quantile $t_{182;0.995}$ est 2.603 (ceci n'est pas dans les tableaux), donc l'hypothèse H_0 est aussi rejetée.

(b) L'intervalle cherché basé sur le TCL est

$$(\bar{x} - \Phi^{-1}(1 - \alpha/2)\tfrac{s}{\sqrt{n}}, \bar{x} + \Phi^{-1}(1 - \alpha/2)\tfrac{s}{\sqrt{n}}) = (37.775, 37.965).$$

Si on suppose que les données sont normales, l'intervalle est

$$(\bar{x} - t_{n-1;1-\alpha/2}\tfrac{s}{\sqrt{n}}, \bar{x} + t_{n-1;1-\alpha/2}\tfrac{s}{\sqrt{n}}) = (37.774, 37.966).$$

Exercice 7. Voici le corrigé tiré 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 :

$$\begin{aligned} P(Y=0) &= C_0^3\left(\frac{1}{6}\right)^0\left(\frac{5}{6}\right)^3 = 0{,}579 \\ P(Y=1) &= C_1^3\left(\frac{1}{6}\right)^1\left(\frac{5}{6}\right)^2 = 0{,}347 \\ P(Y=2) &= C_2^3\left(\frac{1}{6}\right)^2\left(\frac{5}{6}\right)^1 = 0{,}069 \\ P(Y=3) &= C_3^3\left(\frac{1}{6}\right)^3\left(\frac{5}{6}\right)^0 = 0{,}005. \end{aligned}$$

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 X^2 d'ajustement vaut 24,9 qui est à comparer au quantile $\chi^2_{3;0{,}995} = 12{,}84$. On a donc une forte évidence pour rejeter H_0 et affirmer que les dés sont truqués.

Notation : $C_i^j = \binom{j}{i} = \frac{j!}{i!(j-i)!}$.