EXERCICES ET SOLUTIONS

à faire chez vous

Exercice 1. Considérons une boîte contenant 6 boules blanches, 3 boules rouges et une boule bleue. Nous tirons de façon aléatoire une boule de la boîte. Soit X une variable aléatoire prenant la valeur 1 si la boule pigée est blanche, 5 si la boule est rouge et 10 si la boule est bleue.

(a) Trouver la fonction de masse de X.

(b) Trouver la fonction de répartition de X.

(c) Représenter graphiquement la fonction trouvée en (b).

Solution 1. (a) La probabilité que X égale à 1 est la probabilité qu'on tire une des 6 boules blanches. Puisqu'il y a $6+3+1=10$ boules dans la boîte, cette probabilité vaut $6/10$. Ainsi $\mathbb{P}(X=1)=6/10$. Le même raisonnement nous amène à

$$f_X(x)=\mathbb{P}(X=x)=\begin{cases}\frac{6}{10} & x=1\\ \frac{3}{10} & x=5\\ \frac{1}{10} & x=10\\ 0 & \text{sinon.}\end{cases}$$

(b) Nous trouvons la fonction de répartition grâce à un calcul direct :

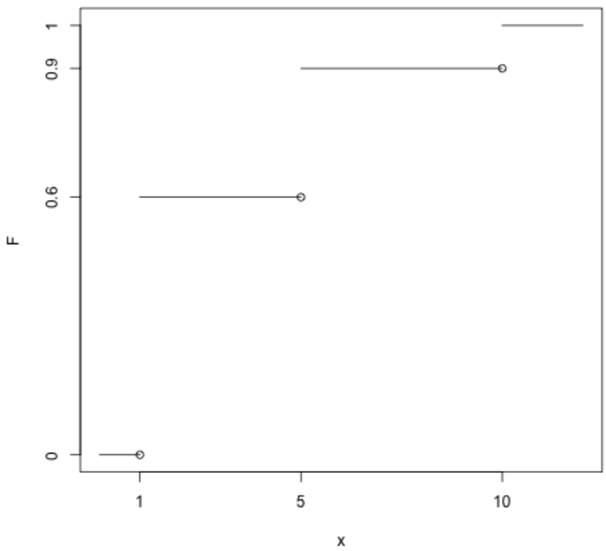
$$F_X(x)=\mathbb{P}(X\leq x)=\begin{cases}0 & x<1\\ \frac{6}{10} & 1\leq x<5\\ \frac{6}{10}+\frac{3}{10}=\frac{9}{10} & 5\leq x<10\\ 1 & x\geq 10.\end{cases}$$

(c) Voici la représentation graphique de la fonction de répartition. Remarquer la continuité à droite !

Exercice 2. On tire trois boules (sans remise) au hasard d'une boîte contenant $n_1=6$ boules rouges et $n_2=4$ boules vertes. Soit X la variable aléatoire représentant le nombre de boules rouges parmi les trois boules pigées. Calculer l'espérance et la variance de X.

Solution 2. Nous utilisons les formules pour un tirage sans remise qui se trouvent au chapitre 2.5 du livre du cours de Probabilités. Évidemment X ne peut prendre que les valeurs $0,1,2$ et 3, avec

$$\mathbb{P}(X=x)=\begin{cases}\frac{\binom{4}{3}\binom{6}{0}}{\binom{10}{3}}=\frac{1}{30} & x=0\\ \frac{\binom{4}{2}\binom{6}{1}}{\binom{10}{3}}=\frac{9}{30} & x=1\\ \frac{\binom{4}{1}\binom{6}{2}}{\binom{10}{3}}=\frac{15}{30} & x=2\\ \frac{\binom{4}{0}\binom{6}{3}}{\binom{10}{3}}=\frac{5}{30} & x=3.\end{cases}$$



On voit bien que ces quatre probabilités somment à 1. Calculons

$$\mathbb{E}[X] = 0\frac{1}{30} + 1\frac{9}{30} + 2\frac{15}{30} + 3\frac{5}{30} = \frac{54}{30} = \frac{9}{5};$$
$$\mathbb{E}[X^2] = 0\frac{1}{30} + 1\frac{9}{30} + 4\frac{15}{30} + 9\frac{5}{30} = \frac{114}{30} = \frac{19}{5};$$

et donc $\mathrm{Var}[X] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = 14/25$. Une autre méthode, sans devoir évaluer $\mathbb{E}[X^2]$, serait de calculer

$$\mathrm{Var}[X] = \mathbb{E}\left[\left(X - \frac{9}{5}\right)^2\right] = 30^{-1}\left(1\left(0-\frac{9}{5}\right)^2 + 9\left(1-\frac{9}{5}\right)^2 + 15\left(2-\frac{9}{5}\right)^2 + 5\left(3-\frac{9}{5}\right)^2\right)$$
$$= \frac{81 + 9\cdot 16 + 15\cdot 1 + 5\cdot 36}{30\cdot 25} = \frac{14}{25}.$$

Exercice 3. Dénotons respectivement par μ et $\sigma^2 > 0$, l'espérance et la variance de la variable aléatoire X. Déterminer $\mathbb{E}\left[\frac{X-\mu}{\sigma}\right]$ et $\mathbb{E}\left[\left(\frac{X-\mu}{\sigma}\right)^2\right]$.

Solution 3. Nous appliquons la linéarité de l'espérance pour calculer

$$\mathbb{E}\left[\frac{X-\mu}{\sigma}\right] = \frac{1}{\sigma}\,\mathbb{E}[X-\mu] = \frac{1}{\sigma}(\mathbb{E}[X] - \mathbb{E}[\mu]) = \frac{1}{\sigma}(\mu-\mu) = 0,$$

car l'espérance de la constante μ égale μ. Aussi

$$\mathbb{E}\left[\left(\frac{X-\mu}{\sigma}\right)^2\right] = \frac{1}{\sigma^2}\,\mathbb{E}[(X-\mu)^2] = \frac{1}{\sigma^2}\,\mathrm{Var}[X] = 1.$$

par définition de la variance.
Ainsi, pour n'importe quelle variable aléatoire X de variance finie et non nulle, la variable aléatoire $Z = (X-\mu)/\sigma$ a une espérance nulle et une variance égale à 1.

Exercice 4. Soient X et Y deux variables aléatoires indépendantes et soient $M_X, M_Y : \mathbb{R} \to \mathbb{R}$ leurs fonctions génératrices de moments respectives. Montrer que la fonction génératrice des moments de la variable aléatoire $Z = X + Y$ est égale à

$$M_Z(t) = M_X(t) \cdot M_Y(t).$$

Solution 4. Puisque X et Y sont indépendantes, $\exp(tX)$ et $\exp(sY)$ sont indépendantes pour chaque $s, t \in \mathbb{R}$. Par conséquent $\mathbb{E}[\exp(tX)\exp(sY)] = \mathbb{E}[\exp(tX)] \cdot \mathbb{E}[\exp(sY)]$. Prendre $s = t$ pour obtenir

$$M_Z(t) = \mathbb{E}[\exp(t(X+Y))] = \mathbb{E}[\exp(tX)\exp(tY)] = \mathbb{E}[\exp(tX)]\cdot\mathbb{E}[\exp(tY)] = M_X(t)\cdot M_Y(t) \in (0, \infty].$$

Exercice 5. Soit Y une variable aléatoire dont la fonction de densité est donnée par

$$g(y) = \begin{cases} cy^2 & \text{si } -1 < y < 1 \\ 0 & \text{sinon.} \end{cases}$$

(a) Déterminer la valeur de la constante c afin que $g(y)$ satisfasse les propriétés d'une fonction de densité.

(b) Trouver la fonction de répartition de Y.

(c) Trouver $\mathbb{P}(0 < Y < 1)$, $\mathbb{P}(0 < Y \leq 3)$ et $\mathbb{P}(Y = 0)$. **Remarque.** On peut répondre à cette question sans calculer aucune intégrale !

(d) Trouver $\mathbb{E}[Y]$ et $\mathrm{Var}[Y]$.

Solution 5. (a) L'intégrale d'une fonction de densité vaut forcément 1. Donc

$$1 = \int_{-\infty}^{\infty} g(y)\,\mathrm{d}y = c\int_{-1}^{1} y^2\,\mathrm{d}y = c\left(\frac{1}{3} - \frac{-1}{3}\right) = \frac{2c}{3}.$$

Ainsi $c = 3/2$.

(b) La fonction de répartition F_Y se trouve en prenant l'intégrale de g. Pour $y \in]-1, 1[$ nous avons

$$F_Y(y) = \mathbb{P}(Y \leq y) = \int_{-1}^{y} cu^2\,\mathrm{d}u = c\left(\frac{y^3}{3} - \frac{-1}{3}\right) = \frac{y^3+1}{2}.$$

Par conséquent $F_Y(y) = \min(1, \max(0, (y^3+1)/2))$, c'est-à-dire

$$F_Y(y) = \mathbb{P}(Y \leq y) = \begin{cases} 0 & y \leq -1 \\ \frac{y^3+1}{2} & -1 < y < 1 \\ 1 & y \geq 1. \end{cases}$$

(c) Puisque Y est une variable aléatoire continue, pour chaque $y \in \mathbb{R}$, on a $\mathbb{P}(Y = y) = 0$. Par exemple si $y = 0$ on a pour $\varepsilon \in]0, 1[$ que

$$0 \leq \mathbb{P}(Y = 0) \leq \mathbb{P}(-\varepsilon < Y \leq \varepsilon) = \mathbb{P}(Y \leq \varepsilon) - \mathbb{P}(Y \leq -\varepsilon) = \frac{\varepsilon^3+1}{2} - \frac{-\varepsilon^3+1}{2} = \varepsilon^3.$$

En laissant $\varepsilon \to 0$ nous voyons effectivement que $\mathbb{P}(Y = 0) = 0$.
Donc $\mathbb{P}(0 < Y < 1) = \mathbb{P}(0 < Y \leq 1) = F_Y(1) - F_Y(0) = 1 - 1/2 = 1/2$ et $\mathbb{P}(0 < Y \leq 3) = F_Y(3) - F_Y(0) = 1 - 1/2 = 1/2$.
En fait la densité de Y est symétrique et nulle à l'extérieur de $[-1, 1]$, ce qui implique $1 = \mathbb{P}(-1 < Y < 1) = 2\mathbb{P}(0 < Y < 1)$.

(d) On peut noter que l'espérance de Y est nulle puisque c'est une variable aléatoire dont la densité est symétrique ; autrement, calculons

$$\mathbb{E}[Y] = \int_{-\infty}^{\infty} y g(y)\,\mathrm{d}y = \int_{-1}^{1} c y^3\,\mathrm{d}y = \frac{3}{2}\left(\frac{1^4}{4} - \frac{1^4}{4}\right) = 0;$$
$$\mathrm{Var}[Y] = \mathbb{E}[Y^2] - (\mathbb{E}[Y])^2 = \mathbb{E}[Y^2] = \int_{-\infty}^{\infty} y^2 g(y)\,\mathrm{d}y = \frac{3}{2}\left(\frac{1^5}{5} - \frac{(-1)^5}{5}\right) = \frac{3}{5}.$$

Exercice 6. Soit X une variable aléatoire dont la fonction de densité est donnée par

$$f(x) = \begin{cases} \frac{1}{10}\exp\left(\frac{-x}{10}\right) & \text{si } 0 < x < \infty \\ 0 & \text{sinon.} \end{cases}$$

(a) Trouver la fonction génératrice des moments $M_X(t)$ de X.

(b) En utilisant $M_X(t)$ ou $R_X(t) = \ln(M_X(t))$, déterminer la moyenne et la variance de X.

Solution 6. (a) Calculons

$$M_X(t) = \mathbb{E}[\exp(tX)] = \int_{-\infty}^{\infty} \exp(tx) f(x)\,\mathrm{d}x = \frac{1}{10}\int_0^{\infty} \exp(x(t - 1/10))\,\mathrm{d}x.$$

Cette intégrale est certainement infinie si $t \geq 1/10$. Dans le cas contraire, nous pouvons joyeusement conclure que

$$M_X(t) = \frac{1}{10}\left(\frac{1}{10} - t\right)^{-1} = \frac{1}{1-10t}; \qquad R_X(t) = \ln(M_X(t)) = -\ln(1-10t).$$

(b) Par les propriétés de la fonction génératrice des moments,

$$\mathbb{E}[X] = M_X'(0) = \left.\frac{10}{(1-10t)^2}\right|_{t=0} = 10; \qquad \mathbb{E}[X] = R_X'(0) = \left.\frac{10}{1-10t}\right|_{t=0} = 10;$$
$$\mathbb{E}[X^2] = M_X''(0) = \left.\frac{200}{(1-10t)^3}\right|_{t=0} = 200; \qquad \mathrm{Var}[X] = R_X''(0) = \left.\frac{100}{(1-10t)^2}\right|_{t=0} = 100;$$
$$\mathrm{Var}[X] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = 100.$$

Remarque. X est une variable aléatoire exponentielle de paramètre $\lambda = 1/10$.

Exercice 7. Montrer que si $X = \sum_{i=1}^n Y_i$ où $Y_i \overset{iid}{\sim} \mathrm{Bern}(p)$, alors $X \sim \mathrm{Bin}(n,p)$.

Solution 7. Puisque les Y_i ne prennent que les valeurs 0 et 1, X ne peut prendre comme valeur que les entiers entre 0 et n. Mais $X = x$ si et seulement si exactement x des Y_i valent 1. Pour chaque $I \subseteq \{1,\dots,n\}$ de cardinalité x, $\mathbb{P}(Y_i = 1 \text{ pour } i \in I \text{ et } Y_i = 0 \text{ pour } i \notin I) = p^x(1-p)^{n-x}$, en raison de l'indépendance des Y_i. L'événement $X = x$ est donc l'union (disjointe) sur tous les I de cardinalité x possibles, il y en a donc $\binom{n}{x}$. Ainsi

$$\mathbb{P}(X = x) = \binom{n}{x} p^x (1-p)^{n-x}, \qquad x = 0,\dots,n.$$

On peut aussi utiliser la fonction génératrice des moments. En effet, par l'exercice 4 de la série 1 et la formule du binôme, on a que

$$M_X(t) = (M_{Y_1}(t))^n = ((1-p)+pe^t)^n = \sum_{x=0}^{n} \binom{n}{x} p^x (1-p)^{n-x} e^{tx}.$$

Cette dernière est par définition $\mathbb{E}[e^{tZ}]$ où $Z \sim Binom(n,p)$. Ainsi $X \sim Binom(n,p)$.

Exercice 8. Soit $\{Y_i\}_{i\geq 1}$ une collection infinie de variables aléatoires, où $Y_i \overset{iid}{\sim} \text{Bern}(p)$. Soit $T = \min\{k \in \mathbb{N} : Y_k = 1\} - 1$, montrer que $T \sim \text{Geom}(p)$.

Solution 8. Il est évident que T ne prend que des valeurs dans $\{0\} \cup \mathbb{N}$. Remarquons que $T+1 = x+1$ si et seulement si $Y_1 = Y_2 = \cdots = Y_x = 0$ et $Y_{x+1} = 1$ et cet événement a une probabilité (grâce à l'indépendance des Y_i)

$$\mathbb{P}(Y_{x+1} = 1) \prod_{i=1}^{x} \mathbb{P}(Y_i = 0) = (1-p)^x p.$$

Ainsi $T \sim \text{Geom}(p)$.

Exercice 9. Montrer que si $X = \sum_{i=1}^{r} Y_i$ où $Y_i \overset{iid}{\sim} \text{Geom}(p)$, alors $X \sim \text{NegBin}(r,p)$.

Solution 9. La fonction génératrice des moments de Y_i est

$$M_{Y_i}(t) = \frac{p}{1-(1-p)e^t}, \quad t < -\log(1-p).$$

Puisque les Y_i sont indépendantes, la fonction génératrice des moments de $X = \sum_{i=1}^{r} Y_i$ est

$$M_X(t) = \prod_{i=1}^{r} M_{Y_i}(t) = \left(\frac{p}{1-(1-p)e^t}\right)^r = \frac{p^r}{[1-(1-p)e^t]^r}, \quad t < -\log(1-p),$$

et donc $X \sim \text{NegBin}(r,p)$

Exercice 10. Soient $X_i \overset{iid}{\sim} \text{Poisson}(\lambda)$. Montrer que $Y = \sum_{i=1}^{n} X_i \sim \text{Poisson}(n\lambda)$.

Solution 10. Nous allons montrer que si $X \sim Poisson(\lambda)$ et $Y \sim Poisson(\mu)$ sont indépendantes pour $\lambda, \mu \geq 0$ alors $X+Y \sim Poisson(\lambda+\mu)$. L'énoncé sera donc achevé par récurrence. Pour x entier on a (car X et Y ne prennent que les valeurs dans $\{0\} \cup \mathbb{N}$)

$$\mathbb{P}(X+Y=x) = \sum_{k=0}^{x} \mathbb{P}(X=k, Y=x-k) = \sum_{k=0}^{x} e^{-\lambda} e^{-\mu} \frac{\lambda^k \mu^{x-k}}{k!(x-k)!} = e^{-(\lambda+\mu)} \frac{(\lambda+\mu)^x}{x!} \sum_{k=0}^{x} \binom{x}{k} \frac{\lambda^k \mu^{x-k}}{(\lambda+\mu)^x}.$$

Cette dernière somme vaut 1 par la formule du binôme. Par conséquent $X+Y \sim Poisson(\lambda+\mu)$.

Exercice 11. Soient $X \sim \text{Poisson}(\lambda)$ et $Y \sim \text{Poisson}(\mu)$ indépendantes. Montrer que la distribution conditionnelle de X sachant $X+Y=k$ est $\text{Bin}(k, \lambda/(\lambda+\mu))$.

Solution 11. Il est clair que les valeurs possibles de X sachant $X+Y=k$ sont $0, 1, \ldots, k$. Pour un tel x, en utilisant l'exercice précédent,

$$\mathbb{P}(X=x|X+Y=k) = \frac{\mathbb{P}(X=x, Y=k-x)}{\mathbb{P}(X+Y=k)} = e^{-\lambda}e^{-\mu}\frac{\lambda^x \mu^{k-x}}{x!(k-x)!}e^{\lambda+\mu}\frac{k!}{(\lambda+\mu)^k} = \binom{k}{x}p^x(1-p)^{k-x},$$

où $p = \lambda/(\lambda+\mu)$. L'énoncé est donc demontré.

Exercice 12. Soient $X \sim \text{Exp}(\lambda)$ et $t \geq 0$. Montrer que $\mathbb{P}[X \geq x+t|X>t] = \mathbb{P}[X \geq x]$.

Solution 12. Nous avons par calcul direct que $\mathbb{P}(X>t) = e^{-\lambda t}$. De plus, lorsque $x>0$, l'événement $\{X \geq x+t\}$ est inclus dans $\{X > t\}$. Il s'en suit que

$$\mathbb{P}(X \geq x+t|X>t) = \frac{e^{-\lambda(x+t)}}{e^{-\lambda t}} = e^{-\lambda x} = \mathbb{P}(X \geq x).$$

Si $x \leq 0$ l'égalité est évidente, car les deux côtés valent 1.

Exercice 13. Soient X et Y des variables aléatoires indépendantes qui suivent des distributions exponentielles d'intensité λ_1 et λ_2 respectivement. Montrer que $Z = \min\{X, Y\}$ est une variable aléatoire exponentielle d'intensité $\lambda_1 + \lambda_2$.
Bonus. Montrer que $\mathbb{P}(Z = X) = \lambda_1/(\lambda_1+\lambda_2)$.

Solution 13. Soit $x \geq 0$. Grâce à l'indépendence de X et Y,

$$\mathbb{P}(\min(X,Y) > x) = \mathbb{P}(X>x, Y>x) = \mathbb{P}(X>x)\mathbb{P}(Y>x) = e^{-\lambda_1 x}e^{-\lambda_2 x} = e^{-(\lambda_1+\lambda_2)x}.$$

Il en découle que $\min(X,Y) \sim Exp(\lambda_1+\lambda_2)$.
Bonus. Nous avons que

$$\mathbb{P}(Z=X) = \mathbb{P}(\min(X,Y)=X) = \mathbb{P}(X \leq Y).$$

Les variables X et Y étant indépendantes, la densité conjointe de (X, Y) est donnée par

$$f_{X,Y}(x,y) = f_X(x)f_Y(y) = \lambda_1 e^{-\lambda_1 x}\lambda_2 e^{-\lambda_2 y},$$

de sorte que

$$\begin{aligned}
\mathbb{P}(X \leq Y) &= \int_0^\infty \int_x^\infty \lambda_1\lambda_2 e^{-\lambda_1 x}e^{-\lambda_2 y}dydx \\
&= \int_0^\infty \lambda_1 e^{-\lambda_1 x}\left[-e^{-\lambda_2 y}\right]_x^\infty dx \\
&= \int_0^\infty \lambda_1 e^{-\lambda_1 x}e^{-\lambda_2 x}dx \\
&= \frac{\lambda_1}{\lambda_1+\lambda_2}.
\end{aligned}$$

Exercice 14. Montrer que $X \sim \chi^2_2$ si et seulement si $X \sim \text{Exp}(1/2)$.

Solution 14. La fonction de densité d'une variable aléatoire Gamma(r, λ) pour $r = 1$ est

$$\lambda e^{-\lambda x}, \qquad x \geq 0.$$

Donc la distribution $Exp(\lambda)$ est la même que la distribution Gamma$(1, \lambda)$.
La distribution χ^2_2 n'est que la distribution Gamma$(1, 1/2)$ qui est donc la même distribution que $Exp(1/2)$.

Exercice 15. Montrer que les distributions suivantes constituent des familles Exponentielles (peut-être lorsqu'un de leurs paramètres est fixé) :

(i) La distribution de Poisson.
(ii) La distribution géométrique.
(iii) La distribution binomiale négative.
(iv) La distribution exponentielle.
(v) La distribution gamma.
(vi) La distribution khi carré.

Solution 15. Rappelons qu'une famille de distributions est une famille exponentielle si sa fonction de masse/densité admet la représentation :

$$f(x) = \exp\left\{\sum_{i=1}^{k} \phi_i T_i(x) - \gamma(\phi_1, \ldots, \phi_k) + S(x)\right\}, \qquad x \in \mathcal{X}. \tag{1}$$

Noter que dans les exemples suivants, les paramétrisations ne sont pas uniques.

(i) Si $X \sim Pois(\lambda)$, alors

$$\begin{aligned} f(x;\lambda) &= \frac{e^{-\lambda}\lambda^x}{x!} \\ &= \exp\left(\ln\left(\frac{e^{-\lambda}\lambda^x}{x!}\right)\right) \\ &= \exp\left(-\lambda + x\ln(\lambda) - \ln(x!)\right). \end{aligned}$$

En posant $\phi = \ln(\lambda)$, $T(x) = x$, $\gamma(\phi) = e^{\phi}$ et $S(x) = -\ln(x!)$ et en notant que le support de f (donné par $\mathcal{X} = \{0\} \cup \mathbb{N}$) ne dépend pas de ϕ, nous obtenons bien que $f(x;\lambda)$ est de la forme (1).

(ii) Si $X \sim Geom(p)$, alors

$$\begin{aligned} f(x;p) &= (1-p)^x p \\ &= \exp\left(x\ln(1-p) + \ln(p)\right). \end{aligned}$$

En posant $\phi = \ln(1-p)$, $T(x) = x$, $\gamma(\phi) = -\ln(1-e^{\phi})$ et $S(x) = 0$ et en notant que le support de f (donné par $\mathcal{X} = \{0\} \cup \mathbb{N}$) ne dépend pas de ϕ, nous obtenons bien que $f(x;p)$ est de la forme (1).

(iii) Si $X \sim NegBin(r,p)$, alors

$$\begin{aligned} f(x;r,p) &= \binom{x+r-1}{x}(1-p)^x p^r \\ &= \exp\left(\ln\binom{x+r-1}{x} + x\ln(1-p) + r\ln(p)\right). \end{aligned}$$

En fixant r et en posant $\phi = \ln(1-p)$, $T(x) = x$, $\gamma(\phi) = -r\ln(1-e^{\phi})$ et $S(x) = \ln\binom{x+r-1}{x}$ et en notant que le support de f (donné par $\mathcal{X} = \{0\} \cup \mathbb{N}$) ne dépend pas de ϕ, nous obtenons bien que $f(x;p)$ est de la forme (1).
Si r est inconnu, la famille binomiale négative n'est pas une famille exponentielle.

(iv) Si $X \sim Exp(\lambda)$, alors pour $x \geq 0$,

$$\begin{aligned} f(x;\lambda) &= \lambda e^{-\lambda x} \\ &= \exp\left(\ln(\lambda) - \lambda x\right). \end{aligned}$$

En posant $\phi = \lambda$, $T(x) = -x$, $\gamma(\phi) = -\ln(\phi)$ et $S(x) = 0$ et en notant que le support de f (donné par $\mathcal{X} = [0,\infty)$) ne dépend pas de ϕ, nous obtenons bien que $f(x;\lambda)$ est de la forme (1).

(v) Si $X \sim Gamma(r,\lambda)$, alors pour $x \geq 0$,

$$\begin{aligned} f(x;r,\lambda) &= \frac{\lambda^r}{\Gamma(r)} x^{r-1} e^{-\lambda x} \\ &= \exp\left(\ln\left(\frac{\lambda^r}{\Gamma(r)}\right) + (r-1)\ln(x) - \lambda x\right) \\ &= \exp\left(r\ln(\lambda) - \ln(\Gamma(r)) + r\ln(x) - \ln(x) - \lambda x\right) \end{aligned}$$

Noter qu'ici $k = 2$, contrairement aux exercices précédents où k était égal à 1. En posant $\phi = (\phi_1,\phi_2) = (\lambda, r)$, $T_1(x) = -x$, $T_2(x) = \ln(x)$, $\gamma(\phi) = -\phi_2\ln(\phi_1) + \ln(\Gamma(\phi_2))$ et $S(x) = -\ln(x)$ et en notant que le support de f (donné par $\mathcal{X} = [0,\infty)$) ne dépend pas de ϕ, nous obtenons bien que $f(x;r,\lambda)$ est de la forme (1). Noter que nous aurions aussi pu poser $\phi = (\phi_1,\phi_2) = (\lambda, r-1)$, $T_1(x) = -x$, $T_2(x) = \ln(x)$, $\gamma(\phi) = -(\phi_2+1)\ln(\phi_1) + \ln(\Gamma(\phi_2+1))$ et $S(x) = 0$.

(vi) Si $X \sim \chi^2_k$, alors $X \sim Gamma(k/2, 1/2)$. Ainsi, il suffit de poser $r = k/2$ et $\lambda = 1/2$ dans les équations du problème (v), afin d'obtenir que $\phi = k/2$, $T(x) = \ln(x)$, $\gamma(\phi) = -\phi\ln(1/2) + \ln(\Gamma(\phi))$ et $S(x) = -\ln(x) - x/2$ nous donne la représentation (1).

Exercice 16. Soit $Y \sim Unif(0,1)$ et soit F une fonction de répartition. Montrer que la fonction de répartition de la variable aléatoire $X = F^{-1}(Y)$ est F, où $F^{-1}(y) = \inf\{t \in \mathbb{R} : F(t) \geq y\}$.

Solution 16. Soit F_X la fonction de répartition de X, montrons que $F_X = F$. Nous avons

$$F_X(x) = \mathbb{P}(X \leq x) = \mathbb{P}(F^{-1}(Y) \leq x).$$

Il suffit donc de montrer que $F^{-1}(Y) \leq x \iff Y \leq F(x)$, car $Y \sim Unif(0,1)$ et donc $\mathbb{P}(F^{-1}(Y) \leq x) = \mathbb{P}(Y \leq F(x)) = F(x)$.

Si $Y \leq F(x)$ alors x appartient à l'ensemble $\{t \in \mathbb{R} : F(t) \geq Y\}$ et x est donc plus grand que l'infimum de cet ensemble, $F^{-1}(Y)$. Donc $Y \leq F(x)$ implique que $F^{-1}(Y) \leq x$.
Si $Y > F(x)$ alors, F étant continue à droite, il existe $\varepsilon > 0$ tel que $Y > F(x+\varepsilon)$. Ainsi (puisque F est croissante) $F^{-1}(Y) = \inf\{t \in \mathbb{R} : F(t) \geq Y\} \geq x + \varepsilon > x$. Donc $F^{-1}(Y) \leq x$ implique que $Y \leq F(x)$. La démonstration est ainsi achevée.

Exercice 17. Soit $X \sim N(\mu, \sigma^2)$, montrer que la fonction de densité de $Y = e^X$ est donnée par

$$f_Y(y) = \frac{1}{y\sigma\sqrt{2\pi}} \exp\left(\frac{-(\ln y - \mu)^2}{2\sigma^2}\right), \quad 0 < y < \infty.$$

Solution 17. Nous avons $Y = g(X) = e^X$ avec $X \sim N(\mu, \sigma^2)$. Par le lemme 1.30 des notes de cours nous savons que $\mathcal{Y} = g(\mathcal{X}) = g((-\infty, \infty)) = (0, \infty)$ et que

$$f_Y(y) = \left|\frac{\mathrm{d}}{\mathrm{d}y} g^{-1}(y)\right| f_X(g^{-1}(y)), \quad y \in (0, \infty),$$

où

$$g^{-1}(y) = \ln(y) \text{ et donc } \frac{\mathrm{d}}{\mathrm{d}y} g^{-1}(y) = \frac{1}{y} > 0, \text{ puisque } y > 0,$$

et

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\}.$$

Nous obtenons finalement que

$$f_Y(y) = \frac{1}{y}\frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(\frac{\ln(y)-\mu}{\sigma}\right)^2\right\}, \quad y \in (0, \infty).$$

Exercice 18. Prouver le théorème sur les transformations multidimensionnelles (page 45 des diapositives du cours) en utilisant la formule de changement de variables dans une intégrale.

Solution 18. Pour n'importe quel $A \subset \mathcal{Y}^n$, on a

$$P(Y \in A) = \int_A f_{\boldsymbol{Y}}(\boldsymbol{y})\,\mathrm{d}\boldsymbol{y}.$$

Mais on a aussi que

$$\begin{aligned} P(Y \in A) = P(g^{-1}(Y) \in g^{-1}(A)) &= P(X \in g^{-1}(A)) \\ &= \int_{g^{-1}(A)} f_{\boldsymbol{X}}(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} \\ &= \int_A f_{\boldsymbol{X}}(g^{-1}(\boldsymbol{y}))\left|\det J_{g^{-1}}(\boldsymbol{y})\right|\,\mathrm{d}\boldsymbol{y}, \end{aligned}$$

où on a utilisé la formule de changement de variables dans une intégrale. Donc, pour chaque $A \subset \mathcal{Y}^n$,

$$\int_A f_{\boldsymbol{Y}}(\boldsymbol{y})\,\mathrm{d}\boldsymbol{y} = \int_A f_{\boldsymbol{X}}(g^{-1}(\boldsymbol{y}))\left|\det J_{g^{-1}}(\boldsymbol{y})\right|\,\mathrm{d}\boldsymbol{y}$$

et on conclut que

$$f_{\boldsymbol{Y}}(\boldsymbol{y}) = f_{\boldsymbol{X}}(g^{-1}(\boldsymbol{y}))\left|\det J_{g^{-1}}(\boldsymbol{y})\right|, \quad \forall \boldsymbol{y} \in \mathcal{Y}^n.$$

Exercice 19. Soient $X \sim N(\mu_1, \sigma_1^2)$ et $Y \sim N(\mu_2, \sigma_2^2)$ indépendantes. Montrer que $X+Y \sim N(\mu_1+\mu_2, \sigma_1^2+\sigma_2^2)$.

Solution 19. D'après le corollaire 1.34, la fonction de densité de $X+Y$ en $z \in \mathbb{R}$ est

$$f_{X+Y}(z) = \int_{\mathbb{R}} \frac{1}{\sigma_1\sqrt{2\pi}} \exp\left[\frac{(z-v-\mu_1)^2}{-2\sigma_1^2}\right] \frac{1}{\sigma_2\sqrt{2\pi}} \exp\frac{(v-\mu_2)^2}{-2\sigma_2^2}\,\mathrm{d}v.$$

Nous allons faire en sorte que l'élément dans l'exponentielle serait $-(v-\mu)^2/2\sigma^2$ de sorte à pouvoir évaluer cette intégrale. On a

$$\begin{aligned}
-\frac{1}{2}\left[\frac{(z-v-\mu_1)^2}{\sigma_1^2}+\frac{(v-\mu_2)^2}{\sigma_2^2}\right] &= -\frac{1}{2\sigma_1^2\sigma_2^2}\left[v^2(\sigma_1^2+\sigma_2^2)-2v(\sigma_2^2(z-\mu_1)+\sigma_1^2\mu_2)+\sigma_2^2(z-\mu_1)^2+\sigma_1^2\mu_2^2\right]\\
&= -\frac{\sigma_2^2+\sigma_1^2}{2\sigma_1^2\sigma_2^2}\left[v^2-2v\frac{\sigma_2^2(z-\mu_1)+\sigma_1^2\mu_2}{\sigma_2^2+\sigma_1^2}+\frac{\sigma_2^2(z-\mu_1)^2+\sigma_1^2\mu_2^2}{\sigma_2^2+\sigma_1^2}\right]\\
&= -\frac{\sigma_2^2+\sigma_1^2}{2\sigma_1^2\sigma_2^2}\left[v-\frac{\sigma_2^2(z-\mu_1)+\sigma_1^2\mu_2}{\sigma_2^2+\sigma_1^2}\right]^2\\
&\quad -\frac{\sigma_2^2+\sigma_1^2}{2\sigma_1^2\sigma_2^2}\left[\frac{\sigma_2^2(z-\mu_1)^2+\sigma_1^2\mu_2^2}{\sigma_2^2+\sigma_1^2}-\left(\frac{\sigma_2^2(z-\mu_1)+\sigma_1^2\mu_2}{\sigma_2^2+\sigma_1^2}\right)^2\right].
\end{aligned}$$

Donc

$$\begin{aligned}
f_{X+Y}(z) = \frac{1}{2\pi\sigma_1\sigma_2}&\exp\left\{-\frac{\sigma_2^2+\sigma_1^2}{2\sigma_1^2\sigma_2^2}\left[\frac{\sigma_2^2(z-\mu_1)^2+\sigma_1^2\mu_2^2}{\sigma_2^2+\sigma_1^2}-\left(\frac{\sigma_2^2(z-\mu_1)+\sigma_1^2\mu_2}{\sigma_2^2+\sigma_1^2}\right)^2\right]\right\}\times\\
&\int_{\mathbb{R}}\exp\left\{-\frac{\sigma_2^2+\sigma_1^2}{2\sigma_1^2\sigma_2^2}\left[v-\frac{\sigma_2^2(z-\mu_1)+\sigma_1^2\mu_2}{\sigma_2^2+\sigma_1^2}\right]^2\right\}\mathrm{d}v.
\end{aligned}$$

L'expression dans la dernière intégrale est liée à la densité d'une variable aléatoire normale d'une certaine moyenne et de variance $\Sigma^2 = \sigma_1^2\sigma_2^2/(\sigma_2^2+\sigma_1^2)$. Elle vaut donc $\sqrt{2\pi\Sigma^2}$ et on obtient

$$\begin{aligned}
f_{X+Y}(z) &= \frac{1}{\sqrt{2\pi(\sigma_2^2+\sigma_1^2)}}\exp\left\{-\frac{\sigma_2^2+\sigma_1^2}{2\sigma_1^2\sigma_2^2}\left[\frac{\sigma_2^2(z-\mu_1)^2+\sigma_1^2\mu_2^2}{\sigma_2^2+\sigma_1^2}-\left(\frac{\sigma_2^2(z-\mu_1)+\sigma_1^2\mu_2}{\sigma_2^2+\sigma_1^2}\right)^2\right]\right\}\\
&= \frac{1}{\sqrt{2\pi(\sigma_2^2+\sigma_1^2)}}\exp\frac{(z-\mu_1)^2+\mu_2^2-2(z-\mu_1)\mu_2}{-2(\sigma_2^2+\sigma_1^2)} = \frac{1}{\sqrt{2\pi(\sigma_2^2+\sigma_1^2)}}\exp\frac{(z-\mu_1-\mu_2)^2}{-2(\sigma_2^2+\sigma_1^2)},
\end{aligned}$$

qui est bien la densité d'une loi $N(\mu_1+\mu_2, \sigma_1^2+\sigma_2^2)$.

Exercice 20. Soit Z_1 une variable aléatoire normale standard et Z_2 une variable aléatoire χ_n^2 où $n \geq 1$, tels que Z_1 et Z_2 sont indépendantes. À l'aide du théorème 1 du cours (le théorème 1.33 à la page 28 des notes du cours), trouver la densité de la variable aléatoire T, où $T = Z_1/\sqrt{Z_2/n}$. *Indice :* définir $g(Z_1, Z_2) = (T, V) = (T, Z_2)$ pour trouver la densité conjointe de T et V. La densité de T se trouve en intégrant par rapport à V (penser à la distribution Gamma).

Remarque. La loi de T s'appelle la loi t de Student avec n degrés de liberté. Elle est très utilisée en statistique et on verra plus tard dans le cours pourquoi. Dans la plupart des cas, n est un nombre entier, mais la distribution est définie pour n'importe quel n réel.

Solution 20. Soient

$$Z_1 \sim \mathcal{N}(0,1), \qquad Z_2 \sim \chi^2_n \qquad \text{et} \qquad T = \frac{Z_1}{\sqrt{Z_2/n}}.$$

Considérons la transformation $g : \mathbb{R} \times \mathbb{R}_+ \setminus \{0\} \to \mathbb{R} \times \mathbb{R}_+ \setminus \{0\}$

$$g : (Z_1, Z_2) \mapsto (T, V) = \left(\frac{Z_1}{\sqrt{Z_2/n}}, Z_2 \right).$$

La fonction inverse est

$$g^{-1} : (T, V) \mapsto \left(T\sqrt{\frac{V}{n}}, V \right), \qquad T \in \mathbb{R}, \quad V \in \mathbb{R}_+ \setminus \{0\},$$

ayant pour Jacobian

$$J_{g^{-1}} = \begin{pmatrix} \sqrt{V/n} & \star \\ 0 & 1 \end{pmatrix} \quad \Rightarrow \quad \det(J_{g^{-1}}(t,v)) = \sqrt{\frac{v}{n}}.$$

L'idée de la preuve est d'utiliser le théorème 1.33 afin de trouver la fonction de densité conjointe de (T, V) et d'ensuite obtenir la fonction de densité marginale de T en intégrant cette densité par rapport à V.
Rappelons que Z_1 et Z_2 sont des variables aléatoires indépendantes et donc

$$f_{(Z_1,Z_2)}(z_1,z_2) = f_{Z_1}(z_1) f_{Z_2}(z_2) = \frac{1}{2^{\frac{n+1}{2}} \pi^{\frac{1}{2}} \Gamma\left(\frac{n}{2}\right)} z_2^{\frac{n}{2}-1} e^{-\frac{1}{2}(z_2+z_1^2)}.$$

La fonction de densité conjointe de (T, V) est alors donnée par

$$\begin{aligned} f_{(T,V)}(t,v) &= f_{(Z_1,Z_2)}(g^{-1}(t,v)) |\det(J_{g^{-1}}(t,v))| \\ &= \frac{1}{2^{\frac{n+1}{2}} \pi^{\frac{1}{2}} \Gamma\left(\frac{n}{2}\right)} v^{\frac{n}{2}-1} e^{-\frac{1}{2}(v + v\frac{t^2}{n})} \cdot \left(\frac{v}{n}\right)^{\frac{1}{2}} \\ &= \frac{1}{2^{\frac{n+1}{2}} \sqrt{\pi n} \Gamma\left(\frac{n}{2}\right)} \cdot v^{\frac{n-1}{2}} e^{-\frac{v}{2}(1+\frac{t^2}{n})}. \end{aligned}$$

Nous pouvons maintenant intégrer par rapport à v afin d'obtenir la fonction de densité de T :

$$f_T(t) = \frac{1}{2^{\frac{n+1}{2}} \Gamma\left(\frac{n}{2}\right) \sqrt{n\pi}} \int_0^\infty e^{-\frac{v}{2}\left(\frac{t^2}{n}+1\right)} v^{\frac{n-1}{2}} \, \mathrm{d}v.$$

En posant

$$y = \frac{v}{2}\left(\frac{t^2}{n} + 1\right),$$

nous obtenons

$$v = \frac{2y}{\left(\frac{t^2}{n}+1\right)} \qquad \text{et} \qquad \mathrm{d}v = \frac{2}{\left(\frac{t^2}{n}+1\right)},$$

et donc

$$\begin{aligned}
f_T(t) &= \frac{1}{2^{\frac{n+1}{2}}\Gamma\left(\frac{n}{2}\right)\sqrt{n\pi}} \cdot \int_0^\infty e^{-y} \cdot \left[(2y)\left(\frac{t^2}{n}+1\right)^{-1}\right]^{\frac{n-1}{2}} \cdot 2\left(\frac{t^2}{n}+1\right)^{-1} \mathrm{d}y \\
&= \frac{1}{2^{\frac{n+1}{2}}\Gamma\left(\frac{n}{2}\right)\sqrt{n\pi}} \cdot \left(\frac{t^2}{n}+1\right)^{-\frac{n+1}{2}} \cdot 2^{\frac{n+1}{2}} \cdot \int_0^\infty y^{\frac{n-1}{2}} e^{-y}\,\mathrm{d}y \\
&= \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} \cdot \frac{1}{\sqrt{n\pi}} \cdot \left(\frac{t^2}{n}+1\right)^{-\frac{n+1}{2}} \cdot \int_0^\infty \frac{1}{\Gamma\left(\frac{n+1}{2}\right)} \cdot y^{\frac{n+1}{2}-1} e^{-y}\,\mathrm{d}y \\
&= \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} \cdot \frac{1}{\sqrt{n\pi}} \cdot \left(\frac{t^2}{n}+1\right)^{-\frac{n+1}{2}}.
\end{aligned}$$

où l'intégrale de l'avant dernière ligne est égale à 1, car c'est l'intégrale de la fonction de densité d'une distribution $\Gamma(n/2, 1)$.

Autre façon de trouver la densité conjointe (portez attention à la nouvelle notation)

Soient

$$Z \sim \mathcal{N}(0,1), \qquad V \sim \chi^2_n \qquad \text{et} \qquad T = \frac{Z}{\sqrt{V/n}}.$$

L'idée dans ce qui suit est de trouver la fonction de densité conjointe de (T, V) en utilisant la densité conditionnelle de $T|V = v$.

La distribution conditionnelle de T sachant $V = v$ est normale de moyenne 0 et de variance n/v. Nous pouvons alors calculer la densité conjointe de la façon suivante :

$$\begin{aligned}
f_{(T,V)}(t,v) =& f_{T|V}(t|V=v) f_V(v) \\
&= \frac{1}{\sqrt{2\pi}}\left(\frac{v}{n}\right)^{\frac{1}{2}} e^{-\frac{1}{2}t^2\frac{v}{n}} \frac{1}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)} v^{\frac{n}{2}-1} e^{-\frac{v}{2}} \\
&= \frac{1}{2^{\frac{n+1}{2}}\Gamma\left(\frac{n}{2}\right)\sqrt{n\pi}} e^{-\frac{v}{2}\left(\frac{t^2}{n}+1\right)} v^{\frac{n-1}{2}}.
\end{aligned}$$

Ensuite l'on procède comme avant pour trouver la densité de T.

***Exercice 21.** Montrer que la distribution exponentielle est l'unique distribution sans mémoire. Plus précisément, soit X une variable aléatoire telle que $\mathbb{P}(X > 0) > 0$ et

$$\mathbb{P}(X > t + s | X > t) = \mathbb{P}(X > s), \qquad \forall t, s \geq 0.$$

Montrer qu'il existe un $\lambda > 0$ tel que $X \sim \text{Exp}(\lambda)$.

Indice : Soit $G(t) = \mathbb{P}(X > t)$. L'absence de mémoire implique que $G(t+s) = G(t)G(s)$ pour $t, s \geq 0$ (pourquoi ?). Poser $g(t) = -\ln G(t)$ et $\lambda = g(1)$. Montrer que $g(t) = t\lambda$ pour chaque $t > 0$ rationnel. En déduire (avec justification !) que $g(t) = t\lambda$ pour chaque $t \geq 0$. Quel est le signe de λ ? Enfin, montrer que $\lambda < \infty$ en utilisant le fait que $G(0) > 0$ et la continuité à droite de G.

Solution 21. Les hypothèses impliquent que

$$G(t+s) = G(t)G(s), \qquad \forall t, s \geq 0,$$

au moins lorsque $G(t) > 0$. Or, si $G(t) = 0$ l'égalité est évidante, car G est décroissante et nonnegative.
En termes de $g(x) = -\ln G(x)$, cette égalité s'écrit

$$g(t+s) = g(t) + g(s), \qquad \forall t, s \geq 0.$$

(À noter que cette égalité tient, et a un sens, même si $g = \infty$, puisque $g(x) \in [0, \infty]$ pour chaque $x \geq 0$.)
Soit $\lambda = g(1)$, alors $g(2) = 2\lambda$ et par récurrence $g(n) = n\lambda$ pour n entier. Par récurrence encore $g\left(\frac{k}{n}\right) = kg\left(\frac{1}{n}\right)$ pour des entiers n, k. En posant $k = n$ nous obtenons $\lambda = g(1) = ng\left(\frac{1}{n}\right)$, et donc $g\left(\frac{k}{n}\right) = \frac{k}{n}\lambda$, c'est-à-dire que $g(q) = q\lambda$ pour chaque $q > 0$ rationnel. Pour $t > 0$ réel, prenons une suite de rationnels $q_n \searrow t$. En utilisant la continuité à droite de g (qui résulte de celle de G),

$$g(t) = \lim_{n\to\infty} g(q_n) = \lim_{n\to\infty} q_n\lambda = t\lambda.$$

(Nous aurions pu utiliser le fait que G, et par conséquent g, est monotone, sans utiliser la continuité à droite.)
Ainsi $G(t) = \exp(-t\lambda)$ pour chaque t. Puisque $G(t) \to 0$ lorsque $t \to \infty$, forcément $\lambda > 0$ et la fonction qui vaut 0 pour $t < 0$ et $1 - G(t)$ pour $t \geq 0$ est bien la fonction de répartition d'une variable aléatoire exponentielle de paramètre λ. Il est impossible que $\lambda = \infty$, puisque G est continue à droite et $G(0) > 0$.
Remarque 1. Nous n'avons même pas supposé ni que X soit une variable aléatoire continue, ni que $\mathbb{P}(X \geq 0) = 1$!
Remarque 2. Il existe des fonctions « sans mémoire » qui ne sont pas de la forme $G(t) = e^{-\lambda t}$. Ces fonctions, évidemment, ne sont pas continues à droite ni monotones. Leur existence requiert une base de $\mathbb{R}$ sur $\mathbb{Q}$ dont la construction nécessite (une version faible de) l'axiome du choix.

Exercice 22. Rappelons que pour un échantillon $x_1, \dots, x_n$ la moyenne échantillonnale est définie par

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

et la médiane échantillonnale par

$$M = \begin{cases} x_{\left(\frac{n+1}{2}\right)}, & \text{si } n \text{ est impair,} \\ \frac{x_{\left(\frac{n}{2}\right)} + x_{\left(\frac{n}{2}+1\right)}}{2}, & \text{sinon.} \end{cases}$$

Montrer que

(i) la fonction $f(\gamma) = \sum_{i=1}^{n}(x_i - \gamma)^2$ atteint son minimum (uniquement) en $\bar{x}$.

(ii) la fonction $g(\gamma) = \sum_{i=1}^{n}|x_i - \gamma|$ atteint son minimum en M. **Attention :** g n'est pas dérivable au point γ si $\gamma = x_i$ pour un i quelconque.

Solution 22.

(i) La dérivée de f est

$$\frac{d}{d\gamma}f(\gamma) = -2\sum_{i=1}^{n}(x_i - \gamma).$$

En la mettant égale à zéro, on trouve

$$\sum_{i=1}^{n} x_i - n\gamma = 0 \Rightarrow \gamma = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x}.$$

Puisque $f''(\gamma) = 2n > 0$, $\bar{x}$ est le minimum global de f.

(ii) On peut écrire

$$g(\gamma) = \sum_{i=1}^{n} |x_i - \gamma| = \sum_{i=1}^{n} |x_{(i)} - \gamma|.$$

La fonction g est dérivable pour chaque $\gamma \in \mathbb{R}\backslash\{x_{(1)}, \ldots, x_{(n)}\}$.

— Quand $\gamma \in (-\infty, x_{(1)})$, on a $g(\gamma) = \sum_{i=1}^{n}(x_{(i)} - \gamma)$ et donc $g'(\gamma) = \sum_{i=1}^{n} -1 = -n$.

— Quand $\gamma \in (x_{(n)}, \infty)$, on a $g(\gamma) = \sum_{i=1}^{n} -(x_{(i)} - \gamma)$ et donc $g'(\gamma) = \sum_{i=1}^{n} 1 = n$.

— Quand $\gamma \in (x_{(j)}, x_{(j+1)})$, $j = 1, \ldots, n-1$, on a

$$g(\gamma) = \sum_{i=1}^{j} -(x_{(i)} - \gamma) + \sum_{i=j+1}^{n} (x_{(i)} - \gamma)$$

et donc $g'(\gamma) = \sum_{i=1}^{j} 1 + \sum_{i=j+1}^{n} -1 = j - (n-j) = 2j - n$.

Distinguons les deux cas suivants :

1. n pair :
   — $g'(\gamma) < 0$ quand $\gamma \in (-\infty, x_{(1)})$ ou $\gamma \in (x_{(j)}, x_{(j+1)})$ avec $j = 1, \ldots, \frac{n}{2} - 1$.
   — $g'(\gamma) = 0$ quand $\gamma \in (x_{(\frac{n}{2})}, x_{(\frac{n}{2}+1)})$.
   — $g'(\gamma) > 0$ quand $\gamma \in (x_{(n)}, \infty)$ ou $\gamma \in (x_{(j)}, x_{(j+1)})$ avec $j = \frac{n}{2} + 1, \ldots, n-1$.
   Puisque g est continue, chaque point en $[x_{(\frac{n}{2})}, x_{(\frac{n}{2}+1)}]$ est un minimum de g et en particulier $M = \frac{x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)}}{2}$ est un minimum.
2. n impair :
   — $g'(\gamma) < 0$ quand $\gamma \in (-\infty, x_{(1)})$ ou $\gamma \in (x_{(j)}, x_{(j+1)})$ avec $j = 1, \ldots, \frac{n+1}{2} - 1$.
   — $g'(\gamma) > 0$ quand $\gamma \in (x_{(n)}, \infty)$ ou $\gamma \in (x_{(j)}, x_{(j+1)})$ avec $j = \frac{n+1}{2}, \ldots, n-1$.
   Puisque g est continue, $M = x_{(\frac{n+1}{2})}$ est l'unique minimum de g.

Remarque : il est possible que $x_{(k)} = x_{(k+1)}$ pour un certain k (c'est-à-dire qu'on observe la même valeur plusieurs fois), mais la preuve reste valide même dans ce cas.

Exercice 23.

(i) Calculez la moyenne $\bar{x}$ et la médiane M des données suivantes :

9.2	11.5	9.7	11.0	8.5
9.8	10.0	12.1	10.5	10.1

(ii) Refaire votre calcul quand la valeur 12.1 est remplacé par 48.6.

(iii) Comparez les valeurs de $\bar{x}$ et M dans les parties (i) et (ii). Qu'est-ce que vous notez ? Expliquez vos observations.

Solution 23.

(i) Nous obtenons $\bar{x} = 10.24$ et $M = 10.05$.

(ii) Maintenant nous obtenons $\bar{x} = 13.89$ et $M = 10.05$.

(iii) On observe que dans la partie (i) les valeurs de $\bar{x}$ et de M sont similaires, tandis que dans la partie (ii) la valeur de $\bar{x}$ a beaucoup changé à cause de la valeur atypique 48.6. En même temps, la valeur de M n'a pas changé. On note que la moyenne $\bar{x}$ est plus susceptible aux valeurs aberrantes que la médiane M. En fait, dans la partie (ii), $\bar{x}$ est plus grande que chaque observation sauf la valeur extrême 48.6. À cause de cette valeur, la moyenne n'est pas un très bon résumé de la position de cet échantillon. En revanche, la médiane n'est pas affectée par cette valeur extrême.

Exercice 24. Soit $x_1, \ldots, x_n$ un échantillon. Est-ce que c'est possible que la moyenne de cet échantillon est égal la médiane de cet échantillon, mais l'échantillon n'est pas symétrique. Trouvez un exemple.

Solution 24. Considerons l'échantillon :

$$-2, -2, 0, 1, 3$$

La moyenne et la médiane sont égal 0, mais l'échantillon n'est pas symétrique autour de 0. **Remarque** pour ceux qui ont besoin d'une définition mathématique formelle de la symétrie. L'échantillon $x_1, \ldots, x_n$ s'appelle symétrique autour de $a \in \mathbb{R}$, si

$$\{x_1, \ldots, x_n\} = \{-(x_1 - a) + a, \ldots, -(x_n - a) + a\}.$$

L'égalité est comprise comme l'égalité des ensembles.

Exercice 25 (exercice 17)**.** Montrer qu'une formule équivalente pour la variance empirique est $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^n x_i^2 - \bar{x}^2$. Expliquer pourquoi cette formule peut être plus utile.

Solution 25. Nous écrivons :

$$\begin{aligned} n\widehat{\sigma}^2 = \sum_{i=1}^n (x_i - \bar{x})^2 \quad &= \quad \sum_{i=1}^n x_i^2 + \sum_{i=1}^n \bar{x}^2 - 2\sum_{i=1}^n x_i\bar{x} \\ &= \quad \sum_{i=1}^n x_i^2 + \sum_{i=1}^n \bar{x}^2 - 2\bar{x}\sum_{i=1}^n x_i \\ &= \quad \sum_{i=1}^n x_i^2 + n\bar{x}^2 - 2n\bar{x}^2 \\ &= \quad \sum_{i=1}^n x_i^2 - n\bar{x}^2. \end{aligned}$$

Cette formule est plus pratique, car elle demande de calculer les carrés de $n+1$ nombres et une différence, au lieu de devoir calculer n différences, et puis n carrés, comme dans la formule originale.

Exercice 26 (exercice 18)**.** Soit un échantillon $x_1, \dots, x_n$. Quels sont la médiane M et les quartiles Q_1 et Q_3 quand $n = 12$, 13, 14 ou 15 ?
**Bonus (c'est un peu fastidieux)* : trouver des formules générales (pour n quelconque) pour le premier et troisième quartile, Q_1 et Q_3. *Indice* : ces formules seront de la forme

$$\begin{cases} ? & n \equiv 0 \mod 4 \\ ? & n \equiv 1 \mod 4 \\ ? & n \equiv 2 \mod 4 \\ ? & n \equiv 3 \mod 4. \end{cases}$$

Solution 26. Si $n = 12$ alors $M = (x_{(6)} + x_{(7)})/2$, $Q_1 = x_{(4)}$ et $Q_3 = x_{(9)}$.
Si $n = 13$ alors $M = x_{(7)}$, $Q_1 = x_{(4)}$ et $Q_3 = x_{(10)}$.
Si $n = 14$ alors $M = (x_{(7)} + x_{(8)})/2$, $Q_1 = (x_{(4)} + x_{(5)})/2$ et $Q_3 = (x_{(10)} + x_{(11)})/2$.
Si $n = 15$ alors $M = x_{(8)}$, $Q_1 = (x_{(4)} + x_{(5)})/2$ et $Q_3 = (x_{(11)} + x_{(12)})/2$.
Pour n quelconque, on obtient les formules

$$Q_1 = \begin{cases} x_{\left(\frac{n}{4}+1\right)} & n \equiv 0 \mod 4 \\ x_{\left(\frac{n-1}{4}+1\right)} & n \equiv 1 \mod 4 \\ \frac{1}{2}\left(x_{\left(\frac{n-2}{4}+1\right)} + x_{\left(\frac{n-2}{4}+2\right)}\right) & n \equiv 2 \mod 4 \\ \frac{1}{2}\left(x_{\left(\frac{n-3}{4}+1\right)} + x_{\left(\frac{n-3}{4}+2\right)}\right) & n \equiv 3 \mod 4, \end{cases} \quad Q_3 = \begin{cases} x_{\left(\frac{3n}{4}\right)} & n \equiv 0 \mod 4 \\ x_{\left(\frac{3(n-1)}{4}+1\right)} & n \equiv 1 \mod 4 \\ \frac{1}{2}\left(x_{\left(\frac{3(n-2)}{4}+1\right)} + x_{\left(\frac{3(n-2)}{4}+2\right)}\right) & n \equiv 2 \mod 4 \\ \frac{1}{2}\left(x_{\left(\frac{3(n-3)}{4}+2\right)} + x_{\left(\frac{3(n-3)}{4}+3\right)}\right) & n \equiv 3 \mod 4. \end{cases}$$

Exercice 27 (exercice 19)**.** Les données suivantes représentent les charges maximales (en tonnes) supportées par les câbles fabriqués par une usine :

10.1	12.2	9.3	12.4	13.7	11.1	13.3
10.8	11.6	10.1	11.2	11.4	11.8	7.1
12.2	12.6	9.2	14.2	10.5		

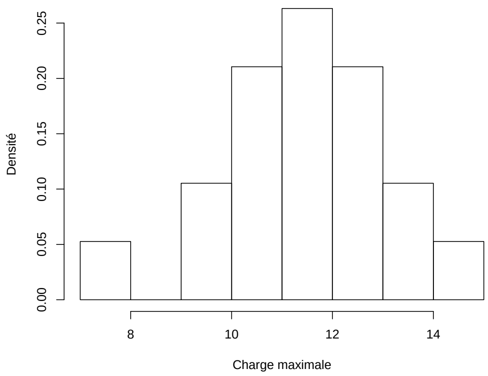
(i) Représenter les données sous la forme d'un histogramme dont la largeur des intervalles est égale à $h = 1$ et l'origine est égale à $\kappa = 10$. Refaire l'histogramme avec $h = 2$ et $\kappa = 11$ et comparer les deux figures.

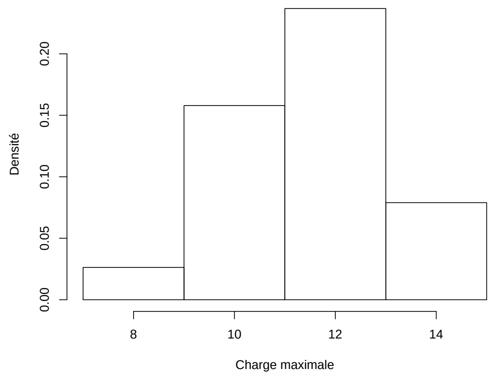
(ii) Quelle est approximativement la valeur de la charge que les trois quarts des câbles peuvent supporter ?

(iii) Donner le troisième quartile.

(iv) Tracer une boîte à moustaches. Parmi les données, y a-t-il des valeurs aberrantes ? Dans ce diagramme, où visualise-t-on la valeur déterminée au point (ii) ?

Solution 27.

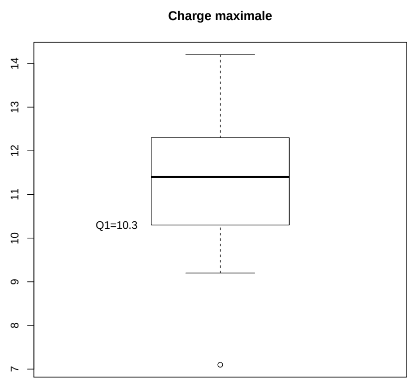




(i) À gauche : $h = 1$, $\kappa = 10$; à droite : $h = 2$, $\kappa = 11$.
Les deux histogrammes donnent plus ou moins le même message : la distribution est unimodale et légèrement asymétrique à gauche. Le premier histogramme a une plus grande “résolution”, mais avec plus de variabilité. Par exemple, on peux déduire la location du mode plus précisément avec le premier histogramme, mais il y a un intervalle vide entre 8 et 9.

(ii) Il s'agit du premier quartile de l'échantillon, Q_1. Ici $n = 19$ et donc la médiane est $M = x_{(10)}$. Le premier quartile est donc défini comme étant la médiane du sous-échantillon $x_{(1)}, \ldots, x_{(10)}$, il est donc donné par $(x_{(5)} + x_{(6)})/2 = 10.3$.

(iii) Le troisième quartile est défini comme étant la médiane du sous-échantillon $x_{(10)}, \ldots, x_{(19)}$, il est donc donné par $(x_{(14)} + x_{(15)})/2 = 12.3$.



(iv) Voir le graphique ci-dessous. La valeur 7.1 est une valeur aberrante et le premier quartile Q_1 détermine la borne inférieure de la boîte.

Exercice 28 (exercices 70–71)**. (Il serait utile de lire la section 6.5 des notes de cours avant de commencer cet exercice.)**

(i) Soit $X \sim \text{Exp}(\lambda)$ où $\lambda > 0$. Montrer que le $\alpha-$quantile de X est

$$q_\alpha = F_X^-(\alpha) = -\log(1-\alpha)/\lambda,$$

pour $0 < \alpha < 1$.

(ii) Les fonctions quantile déterminent les distributions : soient X et Y des variables aléatoires quelconques avec des fonctions de répartition F_X et F_Y. Supposons que $F_X^-(\alpha) = F_Y^-(\alpha)$ pour tout $\alpha \in]0,1[$. Montrer que $F_X = F_Y$.

Solution 28.

(i) La fonction de répartition de la loi exponentielle de paramètre λ est donnée par

$$F_X(x) = 1 - \exp(-\lambda x), \qquad x \geq 0.$$

Puisque cette fonction est continue et strictement croissante sur son support $[0,\infty)$, nous obtenons que $q_\alpha = F_X^-(\alpha) = F_X^{-1}(\alpha)$ et donc

$$\alpha = F_X(q_\alpha) = 1 - \exp(-\lambda q_\alpha) \Longrightarrow q_\alpha = \frac{-\ln(1-\alpha)}{\lambda}.$$

(ii) Supposons par l'absurde que $F_Y(t) < F_X(t)$ pour un certain $t \in \mathbb{R}$. Il existe un $\varepsilon > 0$ tel que $F_Y(t+\varepsilon) < F_X(t)$, car F_Y est continue à droite. Il existe un α tel que $F_Y(t+\varepsilon) < \alpha < F_X(t)$. Visiblement $\alpha \in]0,1[$ et par les définitions de F_X^- et F_Y^- nous avons

$$F_X^-(\alpha) \leq t < t+\varepsilon \leq F_Y^-(\alpha),$$

ce qui contredit l'hypothèse $F_X^- = F_Y^-$ sur $]0,1[$. En supposant qu'il existe un $t \in \mathbb{R}$ tel que $F_X(t) < F_Y(t)$, on arrive à une contradiction semblable.

Exercice 29 (exercice 20)**.** Le tableau suivant contient les résultats des matchs de rugby à XV des onzième et douzième journées (novembre 2014) du championnat français de rugby de première ("Top 14") et deuxième ("Pro D2") division. L'équipe jouant à domicile est celle notée à gauche du tiret.

Top 14		D2	
Montpellier – Brive	10–25	Albi – Agen	22–9
Castres – Toulon	22–14	Béziers – Aurillac	14–19
Clermont – Stade Français	51–9	Colomiers – Pau	50–10
Grenoble – Lyon	34–30	Montauban – Tarbes	31–13
Oyonnax – La Rochelle	37–9	Biarritz – Massy	21–3
Racing Métro – Bayonne	27–10	Dax – Narbonne	12–3
Bordeaux Bègles – Toulouse	20–21	Perpignan – Bourgoin	42–0
		Carcassonne – Mont-de-Marsan	17–28
Toulon – Clermont	27–19	Biarritz – Agen	42–18
Castres – Racing Métro	9–14	Albi – Carcassonne	34–22
La Rochelle – Bayonne	19–19	Aurillac – Colomiers	20–13
Lyon – Montpellier	23–20	Bourgoin – Montauban	14–20
Oyonnax – Bordeaux Bègles	28–23	Massy – Dax	50–13
Toulouse – Grenoble	22–25	Mont-de-Marsan – Béziers	32–18
Stade Français – Brive	20–17	Narbonne – Tarbes	36–23
		Pau – Perpignan	22–19

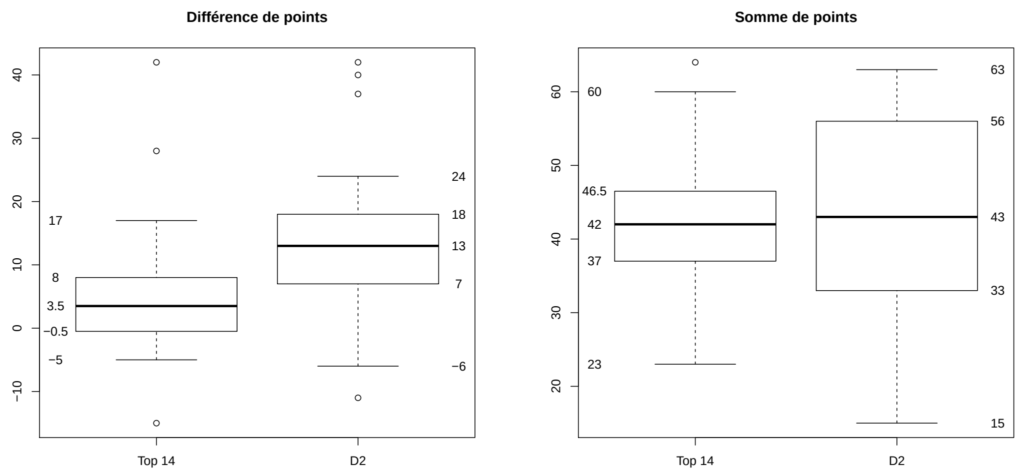
(i) Nous voulons comparer le comportement des équipes en première et en deuxième division. Pour ce faire, calculer pour chacune des divisions quelques statistiques pertinentes (la moyenne, la médiane, les quartiles et l'écart interquartile) pour la différence de points entre le club jouant à domicile et le club visiteur et pour le somme des points par match.

(ii) Représenter côte à côte, sous forme de deux boîtes à moustaches, la somme de points par match en première et en deuxième divison. Faire de même pour la différence de points. Quelles conclusions peut-on en tirer ?

Solution 29. (i) Voici les tables :

Différence de points			Nombre total de points		
	Top 14	D2		Top 14	D2
Moyenne	6.7	14.2	Moyenne	43.1	43.1
Médiane	3.5	13	Médiane	42	43
Q_1	-0.5	7	Q_1	37	33
Q_3	8	18	Q_3	46.5	56
EIQ	8.5	11	EIQ	9.5	23
W_1	-5	-6	W_1	23	15
W_2	17	24	W_2	60	63

(ii) Voici les graphiques :



En regardant le premier graphique ci-dessus, il semble que dans les deux ligues l'équipe jouant à domicile gagne plus souvent. En plus, l'avantage du terrain est nettement plus prononcé en D2. Il y a une proportion importante de valeurs aberrantes (4 sur 16, 3 sur 14), ce qui pourrait suggérer que les ligues ne sont pas équilibrées.

En regardant le second graphique, on ne peut pas dire qu'une certaine ligue est plus défensive que l'autre. En revanche, la variation entre les matchs semble être plus grande en D2. Il est intéressent de noter que la valeur aberrante correspond au match Grenoble–Lyon, un classique du championnat de France, d'autant plus que la plupart des équipes de rugby à XV viennent du sud de la France.

Exercice 30 (exercice 21)**.** Soient $X_1, \ldots, X_n \overset{iid}{\sim} Unif(0, \theta)$. Montrer que $T(X_1, \ldots, X_n) = X_{(n)}$ est une statistique exhaustive pour θ, et trouver sa distribution d'échantillonage.

Solution 30. Pour chaque i, la fonction de densité de X_i est

$$f_{X_i}(x_i; \theta) = \frac{1}{\theta}\mathbf{1}\{x_i \in [0, \theta]\}.$$

Ainsi, les X_i étant indépendantes, la fonction de densité conjointe est

$$f_{X_1,\ldots,X_n}(x_1, \ldots, x_n; \theta) = \frac{1}{\theta^n}\prod_{i=1}^{n}\mathbf{1}\{x_i \in [0, \theta]\} = \frac{1}{\theta^n}\mathbf{1}\{x_{(n)} \leq \theta\}\mathbf{1}\{x_{(1)} \geq 0\}.$$

Par le théorème 2.3 (p. 48), nous avons que $T(X_1, \ldots, X_n) = X_{(n)}$ est une statistique exhaustive pour θ.
Il est évident que $\mathbb{P}(X_{(n)} \leq 0) = 0$ et $\mathbb{P}(X_{(n)} \leq \theta) = 1$. Pour $0 < t < \theta$, X_i étant indépendantes, on a

$$F_T(t; \theta) = \mathbb{P}\left(X_{(n)} \leq t\right) = \mathbb{P}\left(\bigcap_{i=1}^{n}\{X_i \leq t\}\right) = \prod_{i=1}^{n}\mathbb{P}(X_i \leq t) = \left(\frac{t}{\theta}\right)^n.$$

En prenant la dérivée, il s'en suit que $T = X_{(n)}$ est une variable aléatoire continue avec densité

$$f_T(t; \theta) = n\frac{t^{n-1}}{\theta^n}, \qquad t \in [0, \theta].$$

Exercice 31 (exercice 22)**.** Soient $X_1, \ldots, X_n \overset{iid}{\sim} Pois(\lambda)$. Montrer que $T(X_1, \ldots, X_n) = \sum_{i=1}^{n} X_i$ est une statistique exhaustive pour λ, et trouver sa distribution d'échantillonage.

Solution 31. Pour chaque i, la fonction de masse de X_i est

$$f_{X_i}(x_i; \lambda) = \frac{\lambda^{x_i}}{x_i!}e^{-\lambda}\mathbf{1}\{x_i \in \mathcal{X}\}, \qquad \mathcal{X} = \{0, 1, 2, \ldots\}.$$

Ainsi, les X_i étant indépendantes, la fonction de masse conjointe est

$$f_{X_1,\ldots,X_n}(x_1, \ldots, x_n; \lambda) = \prod_{i=1}^{n}\frac{\lambda^{x_i}}{x_i!}e^{-\lambda}\mathbf{1}\{x_i \in \mathcal{X}\} = \lambda^{\sum_{i=1}^{n} x_i}e^{-n\lambda}\left(\prod_{i=1}^{n}\frac{1}{x_i!}\right)\mathbf{1}\{x_i \in \mathcal{X} \ \ \forall i\}.$$

Par le théorème 2.3 (p. 48), nous avons que $T(X_1, \ldots, X_n) = \sum_{i=1}^{n} X_i$ est une statistique exhaustive pour λ.
D'après l'exercice 4, série 1, la distribution de T est $Poisson(n\lambda)$, c'est-à-dire $f_T(t; \lambda) = e^{-n\lambda}(n\lambda)^t/t!$ pour $t = 0, 1, 2, \ldots$.

Exercice 32 (le théorème 2.9 (p. 54) du livre)**.** Prouver que si $X_1, \ldots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$, alors

$$\frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t_{n-1}$$

où t_{n-1} représente la distribution de Student avec $n - 1$ degrés de libérte.

Solution 32. Let

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

We need to show that $f_T(t)$, the density function of T, is as in Definition 2.10 (p.54). Observe that T can be written as a function of two independent random variables $Z = (\bar{X} - \mu)/(\sigma/\sqrt{n})$ and $V = (n-1)S^2/\sigma^2$ (by Proposition 2.7, p.51) :

$$T = \frac{Z}{\sqrt{\frac{V}{n-1}}}$$

Since,

$$f_{Z,V}(z,v) = f_Z(z)f_V(v) = \frac{1}{2^{\frac{n}{2}}\pi^{\frac{1}{2}}\Gamma(\frac{n-1}{2})} v^{\frac{n-1}{2}-1} e^{-\frac{1}{2}(v+z^2)}$$

we can apply Theorem 1.33 (p.28) to the transformation

$$g : (Z, V) \mapsto (T, V) = \left(\frac{Z}{\sqrt{V/(n-1)}}, V \right)$$

to get

$$f_{T,V}(t,v) = \frac{1}{2^{n/2}\sqrt{\pi(n-1)}\Gamma(\frac{n-1}{2})} \cdot v^{\frac{n-2}{2}} e^{-\frac{v}{2}(1+\frac{t^2}{n-1})}$$

And as a consequence,

$$f_T(t) = \frac{1}{2^{n/2}\sqrt{\pi(n-1)}\Gamma(\frac{n-1}{2})} \cdot \int_0^\infty v^{\frac{n-2}{2}} e^{-\frac{v}{2}\left(1+\frac{t^2}{n-1}\right)} \, dv$$

Substitute

$$y = \frac{v}{2}\left(\frac{t^2}{n-1} + 1\right)$$

for v and integrate to obtain the marginal distribution of T. The conclusion follows.

Exercice 33 (Une preuve alternative de la proposition 2.7, p. 51)**.** Soient $X_1, \ldots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$. Définir

$$\begin{aligned}
\mathbf{a}_1 &= \frac{1}{\sqrt{n}}(1, 1, \ldots, 1)', \\
\mathbf{a}_2 &= \frac{1}{\sqrt{2}}(1, -1, 0, \ldots, 0)', \\
\mathbf{a}_3 &= \frac{1}{\sqrt{6}}(1, 1, -2, 0, \ldots, 0)', \\
&\vdots \\
\mathbf{a}_n &= \frac{1}{\sqrt{n(n-1)}}(1, 1, \ldots, 1, -(n-1))'.
\end{aligned}$$

(i) Définir la $n \times n$ matrice $\mathbf{A} = [\mathbf{a}_1 : \mathbf{a}_2 : \ldots : \mathbf{a}_n]$. Montrer que $\mathbf{A}$ est un matrice orthogonale, c'est-à-dire $\mathbf{A}^T\mathbf{A} = \mathbf{A}\mathbf{A}^T = \mathbf{I}_n$, où $\mathbf{I}_n$ est la $n \times n$ matrice d'identité.

(ii) Définir la transformation $Y_i = \mathbf{a}_i'(\mathbf{X} - \mathbf{m}), i = 1, 2, \ldots, n$, où $\mathbf{X} = (X_1, X_2, \ldots, X_n)'$ et $\mathbf{m} = (\mu, \mu, \ldots, \mu)'$. Trouvez la densité conjointe de $Y_1, Y_2, \ldots, Y_n$. Sont-ils indépendants ? Quelle est la distribution de Y_i pour chaque i ?

(iii) Montrer que

$$Y_1 = \sqrt{n}(\bar{X} - \mu) \quad \& \quad \sum_{i=2}^{n} Y_i^2 = (n-1)S^2.$$

Indice : Puisque $\mathbf{A}$ est une matrice orthogonale, $\sum_{i=1}^{n} Y_i^2 = \sum_{i=1}^{n}(X_i - \mu)^2$.

(iv) Utilisez la partie (iii) pour montrer que $\bar{X}$ et S^2 sont indépendants. Montrer aussi que $\bar{X} \sim N(\mu, \sigma^2/n)$ et $(n-1)S^2/\sigma^2 \sim \chi^2_{n-1}$.

Solution 33.

(i) It is easy to show that $\mathbf{a}_1'\mathbf{a}_1 = 1$. Observe that for any $2 \le j \le n$, the lth term of $\mathbf{a}_j$ equals $[j(j-1)]^{-1/2}$ if $l = 1, 2, \ldots, (j-1)$, equals $-[(j-1)/j]^{1/2}$ if $l = j$, and equals 0 if $l > j$. So, direct calculation shows that

$$\mathbf{a}_j'\mathbf{a}_j = 1 \quad \& \quad \mathbf{a}_j'\mathbf{a}_1 = [nj(j-1)]^{-1/2}\left\{\sum_{l=1}^{j-1} 1 - (j-1)\right\} = 0$$

for all $j = 2, 3, \ldots, n$. Further, for any $2 \le j < k \le n$,

$$\mathbf{a}_j'\mathbf{a}_k = [jk(j-1)(k-1)]^{-1/2}\left\{\sum_{l=1}^{j-1} 1 - (j-1)\right\} = 0.$$

Thus, $\mathbf{A}^T\mathbf{A} = \mathbf{I}_n$. Since $\mathbf{A}$ is a $n \times n$ matrix, this also implies that $\mathbf{A}^T = \mathbf{A}^{-1}$ and $\mathbf{A}\mathbf{A}^T = \mathbf{I}_n$.

(ii) In matrix notation, $\mathbf{Y} = \mathbf{A}^T(\mathbf{X} - \mathbf{m})$, where $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_n)'$. Thus, the inverse transformation is given by $\mathbf{X} = \mathbf{A}\mathbf{Y} + \mathbf{m}$, which is a linear transformation. So, $X_i = \mathbf{b}_i'\mathbf{Y} + \mu$ for all $i = 1, 2, \ldots, n$, where $\mathbf{b}_i'$ is the ith row of $\mathbf{A}$. Also, the Jacobian of the inverse transformation is $\mathbf{A}$. Since $\mathbf{A}$ is an orthogonal matrix, $|\det(\mathbf{A})| = 1$. Define $\mathbf{y} = (y_1, y_2, \ldots, y_n)'$. The joint distribution of $Y_1, Y_2, \ldots, Y_n$ is given by

$$\begin{aligned}
f_{Y_1,Y_2,\ldots,Y_n}(y_1, y_2, \ldots, y_n) &= \left(\frac{1}{2\pi\sigma^2}\right)^{n/2} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(\mathbf{b}_i'\mathbf{y} + \mu - \mu)^2\right\} \times |\det(\mathbf{A})| \\
&= \left(\frac{1}{2\pi\sigma^2}\right)^{n/2} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}\mathbf{y}'\mathbf{b}_i\mathbf{b}_i'\mathbf{y}\right\} \\
&= \left(\frac{1}{2\pi\sigma^2}\right)^{n/2} \exp\left\{-\frac{1}{2\sigma^2}\mathbf{y}'\left(\sum_{i=1}^{n}\mathbf{b}_i\mathbf{b}_i'\right)\mathbf{y}\right\} \\
&= \left(\frac{1}{2\pi\sigma^2}\right)^{n/2} \exp\left\{-\frac{1}{2\sigma^2}\mathbf{y}'\left(\mathbf{A}^T\mathbf{A}\right)\mathbf{y}\right\} \\
&= \left(\frac{1}{2\pi\sigma^2}\right)^{n/2} \exp\left\{-\frac{1}{2\sigma^2}\mathbf{y}'\mathbf{y}\right\} \\
&= \left(\frac{1}{2\pi\sigma^2}\right)^{n/2} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n} y_i^2\right\}.
\end{aligned}$$

So,

$$f_{Y_1,Y_2,\ldots,Y_n}(y_1, y_2, \ldots, y_n) = \prod_{i=1}^{n} \left(\frac{1}{2\pi\sigma^2}\right)^{1/2} \exp\left\{-\frac{y_i^2}{2\sigma^2}\right\},$$

which implies that $Y_1, Y_2, \ldots, Y_n$ are independent random variables. Further, $Y_i \sim N(0, \sigma^2)$ for each $i = 1, 2, \ldots, n$.

(iii) By definition of Y_1, it follows that $Y_1 = n^{-1/2} \sum_{i=1}^{n}(X_i - \mu) = \sqrt{n}(\bar{X} - \mu)$.
Observe that

$$\sum_{i=1}^{n} Y_i^2 = \mathbf{Y}'\mathbf{Y} = (\mathbf{X} - \mathbf{m})'\mathbf{A}\mathbf{A}^T(\mathbf{X} - \mathbf{m}) = (\mathbf{X} - \mathbf{m})'(\mathbf{X} - \mathbf{m}) = \sum_{i=1}^{n}(X_i - \mu)^2.$$

Thus,

$$\sum_{i=2}^{n} Y_i^2 = \sum_{i=1}^{n} Y_i^2 - Y_1^2 = \sum_{i=1}^{n}(X_i - \mu)^2 - n(\bar{X} - \mu)^2 = \sum_{i=1}^{n}(X_i - \bar{X})^2 = (n-1)S^2.$$

(iv) Note that $\bar{X}$ is a function of Y_1 only and S^2 is a function of $Y_2, Y_3, \ldots, Y_n$. Since Y_1 is independent of $Y_2, Y_3, \ldots, Y_n$, it follows that $\bar{X}$ and S^2 are independent.
Since $Y_1 \sim N(0, \sigma^2)$, it follows that $\bar{X} \sim N(\mu, \sigma^2/n)$. Further, $(n-1)S^2/\sigma^2 = \sum_{i=2}^{n}(Y_i/\sigma)^2$. As $Y_i/\sigma \overset{iid}{\sim} N(0,1)$ for $i = 2, 3, \ldots, n$, the latter sum is that of $(n-1)$ independent χ_1^2 random variables. So, $(n-1)S^2/\sigma^2 \sim \chi_{n-1}^2$.

Exercice 34. Soient $X_1, X_2, \ldots, X_n \overset{iid}{\sim} N(0, 1)$. Montrer que $X_i^2 / \sum_{j=1}^{n} X_j^2$ et $\sum_{j=1}^{n} X_j^2$ sont indépendants pour chaque $i = 1, 2, \ldots, n$.

Solution 34. First observe that the random variables $Y_i = X_i^2, i = 1, 2, \ldots, n$, are independent, and each one of them have a χ_1^2 distribution. So, the joint density of $Y_1, Y_2, \ldots, Y_n$ is given by

$$f_{Y_1,Y_2,\ldots,Y_n} = \frac{c_n \exp\{-\sum_{i=1}^{n} y_i/2\}}{(y_1 y_2 \ldots y_n)^{1/2}},$$

where c_n is a constant depending on n.
Define the transformation $Z_1 = \sum_{i=1}^{n} Y_i, Z_2 = Y_2/Z_1, \ldots, Z_n = Y_n/Z_1$. The inverse transformation is given by $Y_1 = Z_1(1 - \sum_{i=2}^{n} Z_i), Y_2 = Z_1 Z_2, \ldots, Y_n = Z_1 Z_n$. The Jacobian of the inverse transformation is given by

$$\begin{bmatrix} 1 - \sum_{i=2}^{n} z_i & -z_1 & -z_1 & -z_1 & \ldots & -z_1 \\ z_2 & z_1 & 0 & 0 & \ldots & 0 \\ z_3 & 0 & z_1 & 0 & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ldots & \vdots \\ z_n & 0 & 0 & 0 & \ldots & z_1 \end{bmatrix} = \begin{bmatrix} 1 - \sum_{i=2}^{n} z_i & \mathbf{a}' \\ \mathbf{b} & z_1 \mathbf{I}_{n-1} \end{bmatrix},$$

where $\mathbf{a} = (-z_1, -z_1, \ldots, -z_1)'$ and $\mathbf{b} = (z_2, z_3, \ldots, z_n)'$. Using the formula for the determinant of a partitioned matrix, it follows that the determinant of the Jacobian of the inverse transformation is $\det(z_1 \mathbf{I}_{n-1}) \times \{1 - \sum_{i=2}^{n} z_i - \mathbf{a}' z_1^{-1} \mathbf{I}_{n-1} \mathbf{b}\} = z_1^{n-1}$.

So, the joint density of $Z_1, Z_2, \ldots, Z_n$ is given by

$$\begin{aligned} f_{Z_1,Z_2,\ldots,Z_n}(z_1,z_2,\ldots,z_n) &= \frac{c_n \exp\{-z_1/2\} z_1^{n-1}}{\left(z_1^n[1-\sum_{i=2}^n z_i] z_2 z_3 \ldots z_n\right)^{1/2}} \\ &= \frac{c_n \; z_1^{n/2-1} \exp\{-z_1/2\}}{\left([1-\sum_{i=2}^n z_i] z_2 z_3 \ldots z_n\right)^{1/2}}. \end{aligned}$$

Thus, Z_1 is independent of $Z_2, Z_3, \ldots, Z_n$. Since $Y_1/Z_1 = 1 - \sum_{i=2}^n Y_i/Z_1 = 1 - \sum_{i=2}^n Z_i$, it follows that Y_1/Z_1 is also independent of Z_1.

Exercice 35 (exercice 23)**.** Soient $X_1, \ldots, X_n \overset{iid}{\sim} f$, où f est de la forme d'une famille exponentielle, exprimée dans la paramétrisation usuelle comme $f(x) = \exp\left[\eta(\theta)T(x) - d(\theta) + S(x)\right]$, $\theta \in \Theta \subseteq \mathbb{R}$ ouvert. Montrer que :

(i) Si η est k-fois continûment dérivable ($k \geq 1$) et inversible avec la dérivée jamais nulle, alors d est aussi k-fois continûment dérivable.

(ii) Si η est deux fois continûment dérivable et inversible avec la dérivée jamais nulle, alors

$$\mathbb{E}[\tau(X_1,\ldots,X_n)] = n\frac{d'(\theta)}{\eta'(\theta)} \quad \& \quad \mathrm{Var}[\tau(X_1,\ldots,X_n)] = n\frac{d''(\theta)\eta'(\theta) - d'(\theta)\eta''(\theta)}{[\eta'(\theta)]^3},$$

où $\tau(X_1, \ldots, X_n) = \sum_{i=1}^n T(X_i)$.

Indice : utiliser le théorème de la fonction inverse (théorème 6.2, p. 162).

Solution 35.

(i) Nous avons que $f(x) = \exp\left[\eta(\theta)T(x) - d(\theta) + S(x)\right] = \exp\left[\phi T(x) - \gamma(\phi) + S(x)\right]$ où $\eta(\theta) = \phi$ et $\gamma(\phi) = \gamma(\eta(\theta)) = d(\theta)$ avec $d = \gamma \circ \eta$. Puisque $\eta(\theta)$ est dérivable k fois par l'hypothèse, d le sera aussi, à condition que γ soit suffisamment dérivable. D'après la proposition 2.11 (p. 57), il suffit d'établir que

$$\Phi = \eta(\Theta) = \{\phi \in \mathbb{R} : \text{il existe un } \theta \in \Theta \text{ tel que } \phi = \eta(\theta)\}$$

est un ouvert ; il en résulte que γ est infiniment dérivable.

Pour ce faire, nous devons montrer que pour chaque $\phi_0 \in \Phi$, il existe un $\delta > 0$ tel que

$$]\phi_0 - \delta, \phi_0 + \delta[\subseteq \Phi = \eta(\Theta).$$

On remarque tout d'abord que, puisque $\phi_0 \in \Phi$, forcément $\phi_0 = \eta(\theta_0)$ pour un certain $\theta_0 \in \Theta$. Maintenant, on va utiliser les deux faits suivants :

(i) Sous l'hypothèse Θ est ouvert, il existe donc $\epsilon > 0$ tel que $]\theta_0 - \epsilon, \theta_0 + \epsilon[\subseteq \Theta$.

(ii) La dérivée η' est continue et $\eta'(\theta_0) \neq 0$. Le théorème de fonction inverse implique que η^{-1} est continue (en fait, continûment dérivable) sur un intervalle ouvert I contenant $\phi_0 = \eta(\theta_0)$.

Ceci montre que η est un homéomorphisme local et donc η est une application ouverte, et la preuve est achevée. Ceux qui n'aiment pas la topologie devraient se contenter de l'argument élémentaire suivant : η^{-1} étant continue sur $I \ni \eta(\theta_0) = \phi_0$, il existe un $\delta > 0$ tel que

$$|\phi - \phi_0| < \delta \implies |\underbrace{\eta^{-1}(\phi)}_{=\theta \text{ (disons)}} - \underbrace{\eta^{-1}(\phi_0)}_{=\theta_0}| < \epsilon$$

de sorte que $]\phi_0 - \delta, \phi_0 + \delta[\subseteq I$ et ϵ est défini par (i).

Pour résumer : il existe un $\delta > 0$ tel que pour chaque $\phi \in]\phi_0 - \delta, \phi_0 + \delta[$ il existe $\theta \in]\theta_0 - \epsilon, \theta_0 + \epsilon[\subseteq \Theta$ pour lequel $\phi = \eta(\theta)$, et donc $]\phi_0 - \delta, \phi_0 + \delta[\subseteq \Phi = \eta(\Theta)$. Or ϕ_0 est arbitraire ; ceci montre donc que Φ est ouvert et donc γ est infiniment dérivable. Il s'en suit que si η est k fois dérivable, alors $d = \gamma \circ \eta$ l'est aussi.

Remarque. La fonction $\eta(\theta) = \theta^3$ est bijective et dérivable, mais sa dérivée s'annule en zéro.

(ii) Par la proposition 2.11, nous savons que $\mathbb{E}[\tau(X_1, \dots, X_n)] = n\gamma'(\phi)$ où

$$\gamma'(\phi) = \gamma'(\eta(\theta)) = \frac{(\gamma \circ \eta)'(\theta)}{\eta'(\theta)} = \frac{d'(\theta)}{\eta'(\theta)},$$

car $(f \circ g)'(x) = f'(g(x))g'(x)$. Ainsi, $\mathbb{E}[\tau(X_1, \dots, X_n)] = n\frac{d'(\theta)}{\eta'(\theta)}$.

Par la proposition 2.11, nous savons aussi que $\mathrm{Var}[\tau(X_1, \dots, X_n)] = n\gamma''(\phi)$ où

$$\gamma''(\phi) = \gamma''(\eta(\theta)) = \frac{(\gamma'(\eta(\theta)))'}{\eta'(\theta)} = \left(\frac{d'(\theta)}{\eta'(\theta)}\right)' \frac{1}{\eta'(\theta)} = \frac{d''(\theta)\eta'(\theta) - d'(\theta)\eta''(\theta)}{[\eta'(\theta)]^3}.$$

Ainsi, $\mathrm{Var}[\tau(X_1, \dots, X_n)] = n\frac{d''(\theta)\eta'(\theta) - d'(\theta)\eta''(\theta)}{[\eta'(\theta)]^3}$.

Exercice 36 (loi des événements rares, exercice 24)**.** Soit $\{X_n\}_{n\geq 1}$ une séquence de variables aléatoires $\mathrm{Bin}(n, p_n)$, telle que $p_n = \lambda/n$, pour une certaine constante $\lambda > 0$. Montrer que $X_n \xrightarrow{d} Y$, où $Y \sim \mathrm{Poisson}(\lambda)$.

Indice : (1) montrer que pour $k \in \mathbb{N} \cup \{0\}$, $\mathbb{P}(X_n = k) \to \mathbb{P}(Y = k)$. (2) Déduire que $\mathbb{P}(X_n \leq k) \to \mathbb{P}(Y \leq k)$. (3) Conclure.

Solution 36. Suivant l'indice, montrons que $f_{X_n}(x) \overset{n\to\infty}{\longrightarrow} f_Y(x), \forall x \in \{0\} \cup \mathbb{N}$. Rappelons que (pour $n \geq \lambda$)

$$f_Y(x) = e^{-\lambda}\frac{\lambda^x}{x!} \quad \text{et} \quad f_{X_n}(x) = \binom{n}{x} p_n^x (1 - p_n)^{n-x} = \binom{n}{x}\left(\frac{\lambda}{n}\right)^x \left(1 - \frac{\lambda}{n}\right)^{n-x}.$$

Nous pouvons réécrire $f_{X_n}(x)$ de la façon suivante :

$$\begin{aligned} f_{X_n}(x) &= \frac{n!}{x!(n-x)!}\frac{\lambda^x}{n^x}\left(1 - \frac{\lambda}{n}\right)^n \left(1 - \frac{\lambda}{n}\right)^{-x} \\ &= \frac{n(n-1)\cdot \ldots \cdot (n-x+1)}{n^x}\frac{\lambda^x}{x!}\left(1 - \frac{\lambda}{n}\right)^n \left(1 - \frac{\lambda}{n}\right)^{-x} \\ &= \left(\frac{n}{n}\frac{n-1}{n} \cdot \ldots \cdot \frac{n-x+1}{n}\right)\left(1 - \frac{\lambda}{n}\right)^{-x} \frac{\lambda^x}{x!}\left(1 - \frac{\lambda}{n}\right)^n. \end{aligned}$$

Le terme dans la première parenthèse contient le produit d'un nombre fixe $x < \infty$ de termes qui convergent vers 1 lorsque $n \to \infty$, il converge donc vers 1. La deuxième parenthèse converge aussi vers 1, puisque x est constant. Finalement, la troisième parenthèse converge vers $e^{-\lambda}$. Nous obtenons donc que

$$\binom{n}{x}\left(\frac{\lambda}{n}\right)^x\left(1-\frac{\lambda}{n}\right)^{n-x} \overset{n\to\infty}{\longrightarrow} e^{-\lambda}\frac{\lambda^x}{x!}, \quad \forall x \in \{0\} \cup \mathbb{N}.$$

Ensuite, on remarque que pour $k \in \mathbb{N} \cup \{0\}$,

$$\mathbb{P}(X_n \leq k) = \sum_{x=0}^{k} \mathbb{P}(X_n = x) \to \sum_{x=0}^{k} \mathbb{P}(Y = x) = \mathbb{P}(Y \leq k), \qquad n \to \infty.$$

Les deux égalités viennent du fait que X_n ainsi que Y prennent des valeurs uniquement dans $\mathbb{N} \cup \{0\}$.
Pour $t < 0$, $\mathbb{P}(X_n \leq t) = 0 = \mathbb{P}(Y \leq t)$. Pour $t \geq 0$, posons $k = [t] = \max\{n \in \mathbb{Z} : n \leq t\}$. Alors

$$\mathbb{P}(X_n \leq t) = \mathbb{P}(X_n \leq k) \to \mathbb{P}(Y \leq k) = \mathbb{P}(Y \leq t), \qquad n \to \infty.$$

Par définition, on a donc $X_n \xrightarrow{d} Y$.

Exercice 37 (de la distribution exponentielle à la géométrique et inversement)**.**

(i). Soit $X \sim \text{Exp}(\lambda)$ pour $\lambda > 0$. Montrer que $\lfloor X \rfloor \sim \text{Geom}(p)$ pour un p approprié à trouver. (On définit $\lfloor t \rfloor = \max\{n \in \mathbb{Z} :\ n \leq t\}$, pour $t \in \mathbb{R}$.)

(ii). Soit $\{X_n\}_{n=1}^{\infty}$ une suite de variables aléatoires avec $X_n \sim \text{Geom}\left(\frac{\lambda}{n}\right)$ et soit $Z \sim \text{Exp}(\lambda)$, pour un certain $\lambda > 0$. Montrer que $\frac{X_n}{n} \xrightarrow{d} Z$, lorsque $n \to \infty$.

Solution 37.

(i). Soit $Y = \lfloor X \rfloor$, où $X \sim Exp(\lambda)$. Alors, on calcule, pour chaque $k \in \mathbb{N} \cup \{0\}$:

$$\begin{aligned}
\mathbb{P}(Y = k) &= \mathbb{P}(k \leq X < k+1) \\
&= \int_k^{k+1} \lambda e^{-\lambda x} dx \\
&= \left[-e^{-\lambda x}\right]_k^{k+1} \\
&= e^{-\lambda k} - e^{-\lambda(k+1)} \\
&= e^{-\lambda k}(1 - e^{-\lambda}) \\
&= (1-p)^k p
\end{aligned}$$

où on a défini $p = 1 - e^{-\lambda}$. Nous avons que $p \in (0,1)$, car $\lambda > 0$. Alors $Y \sim \text{Geom}(p)$.

(ii). Pour montrer la convergence en loi, il faut calculer la fonction de répartition de X_n/n. Pour chaque $t \in \mathbb{R}$, $t \geq 0$, nous avons :

$$\begin{aligned}
\mathbb{P}\left(\frac{X_n}{n} > t\right) &= \mathbb{P}(X_n > nt) \\
&= \sum_{k=\lfloor nt \rfloor+1}^{\infty} \left(1-\frac{\lambda}{n}\right)^k \frac{\lambda}{n} \\
&= \frac{\lambda}{n} \frac{\left(1-\frac{\lambda}{n}\right)^{\lfloor nt \rfloor+1}}{1-\left(1-\frac{\lambda}{n}\right)} \\
&= \left(1-\frac{\lambda}{n}\right)^{nt} \left(1-\frac{\lambda}{n}\right)^{r_n}, \qquad r_n = \lfloor nt \rfloor + 1 - nt,
\end{aligned}$$

où $0 < r_n \leq 1$ pour tout $n \in \mathbb{N}$.
Puisque $\left(1-\frac{\lambda}{n}\right) \leq 1$, on a que $\left(1-\frac{\lambda}{n}\right) \leq \left(1-\frac{\lambda}{n}\right)^{r_n} \leq 1$ pour tout $n \in \mathbb{N}$, et donc $\left(1-\frac{\lambda}{n}\right)^{r_n} \to 1$ lorsque $n \to \infty$ (on utilise le théorème de deux gendarmes). Puis, $\left(1-\frac{\lambda}{n}\right)^{nt} \to e^{-\lambda t}$, lorsque $n \to \infty$, parce que $t \in \mathbb{R}$ est fixé.
En résumé, nous avons montré que pour tout $t \in \mathbb{R}$,

$$F_{X_n/n}(t) = 1 - \mathbb{P}\left(\frac{X_n}{n} > t\right) \to 1 - e^{-\lambda t}$$

lorsque $n \to \infty$ (c'est évident quand $t < 0$). Le membre à droite est la fonction de répartition d'une variable aléatoire $Exp(\lambda)$, et donc $X_n/n \xrightarrow{d} Z$, où $Z \sim Exp(\lambda)$.

Exercice 38 (exercice 25)**.** On dit qu'une suite de variables aléatoires X_n converge vers une variable aléatoire Y *en probabilité* (p. 60) si

$$\forall \epsilon > 0 \qquad \lim_{n\to\infty} \mathbb{P}[|X_n - Y| > \epsilon] = 0.$$

Dans ce cas on écrit $X_n \xrightarrow{p} Y$.

Soit $\{X_n\}_{n=1}^{\infty}$ une suite de variables aléatoires avec

$$X_n = (-1)^n X, \quad \mathbb{P}(X=-1) = \mathbb{P}(X=1) = \frac{1}{2}.$$

Montrer que $X_n \xrightarrow{d} X$, mais que $X_n \stackrel{p}{\nrightarrow} X$.

Solution 38. La variable aléatoire X est discrète avec fonction de masse :

$$f_X(x) = \mathbb{P}(X = x) = \begin{cases} 1/2 & \text{si } x = 1 \\ 1/2 & \text{si } x = -1 \\ 0 & \text{sinon.} \end{cases}$$

Si n est pair, nous avons que $X_n = X$ et donc $f_{X_n} = f_X$. Si n est impair nous avons :

$$f_{X_n}(x) = \mathbb{P}(X_n = x) = \mathbb{P}(-X = x) = \mathbb{P}(X = -x) = \begin{cases} 1/2 & \text{si } x = 1 \\ 1/2 & \text{si } x = -1 \\ 0 & \text{sinon.} \end{cases}$$

Nous avons donc montré que $f_{X_n} = f_X$, quelque soit n. Il s'en suit que $X_n \xrightarrow{d} X$.
Notons que pour n pair, nous avons que $\forall \epsilon > 0$:

$$\mathbb{P}(|X_n - X| > \epsilon) = \mathbb{P}(0 > \epsilon) = 0.$$

Par contre, si n est impair, nous avons que

$$\mathbb{P}(|X_n - X| > \epsilon) = \mathbb{P}(|-2X| > \epsilon) = \mathbb{P}(2 > \epsilon) = 1,$$

pour $0 < \epsilon < 2$. Ainsi la séquence $\{\mathbb{P}(|X_n - X| > \epsilon)\}_{n\geq 1}$ est de la forme $\{0, 1, 0, 1, \dots\}$, elle ne converge donc pas et on peut conclure que $X_n \overset{p}{\nrightarrow} X$.

Exercice 39 (exercise 27)**.** Soient $X_1, \dots, X_n \overset{iid}{\sim} Pois(\lambda)$, où $\lambda \in (0, \infty)\backslash\{1\}$ et considérons la probabilité $\pi = \mathbb{P}(X_i = 1) = \lambda e^{-\lambda}$. Nous voulons estimer π par $\hat{\pi}_n = \hat{\lambda}_n e^{-\hat{\lambda}_n}$ où $\hat{\lambda}_n = \frac{1}{n}\sum_{i=1}^n X_i$. Montrer que

$$\frac{\sqrt{n}(\hat{\pi}_n - \pi)}{\sqrt{\hat{\lambda}_n} e^{-\hat{\lambda}_n}(1 - \hat{\lambda}_n)} \xrightarrow{d} Y,$$

où $Y \sim N(0, 1)$. Indication : vous aurez besoin du théorème limite central, de la méthode delta, de la loi faible des grands nombres ainsi que du théorème de Slutsky.

Solution 39. Puisque les variables aléatoires $X_1, \dots, X_n$ sont iid de moyenne $\mathbb{E}[X_i] = \lambda$ et $\text{Var}[X_i] = \lambda < \infty$, nous avons par le théorème limite central que

$$\sqrt{n}(\hat{\lambda}_n - \lambda) \xrightarrow{d} Y,$$

avec $Y \sim N(0, \lambda)$. Définissons $g : \mathbb{R} \to \mathbb{R}$ par $g(x) = xe^{-x}$. Par la méthode delta, nous obtenons que

$$\sqrt{n}(g(\hat{\lambda}_n) - g(\lambda)) \xrightarrow{d} Y \cdot g'(\lambda),$$

où $g(\hat{\lambda}_n) = \hat{\pi}_n$, $g(\lambda) = \pi$ et $g'(\lambda) = e^{-\lambda}(1 - \lambda)$. Ainsi,

$$\sqrt{n}(\hat{\pi}_n - \pi) \xrightarrow{d} Y_1,$$

avec $Y_1 \sim N(0, \lambda e^{-2\lambda}(1 - \lambda)^2)$.
De plus, par la loi faible des grands nombres, nous savons que $\hat{\lambda}_n \xrightarrow{p} \lambda$. Soit $h : \mathbb{R} \times \mathbb{R}_+ \to \mathbb{R}$ une fonction définie telle que $h(x, y) = \frac{x}{\sqrt{y}e^{-y}(1-y)}$. Par le théorème de Slutsky, nous obtenons que

$$\frac{\sqrt{n}(\hat{\pi}_n - \pi)}{\sqrt{\hat{\lambda}_n} e^{-\hat{\lambda}_n}(1 - \hat{\lambda}_n)} = h(\sqrt{n}(\hat{\pi}_n - \pi), \hat{\lambda}_n) \xrightarrow{d} h(Y_1, \lambda) = \frac{Y_1}{\sqrt{\lambda}e^{-\lambda}(1 - \lambda)} = W,$$

avec $W \sim N(0, 1)$, ce qui conclut la preuve.

Exercice 40 (exercice 28)**.** Soient $x_1, \dots, x_n$ des réalisations indépendantes d'une variable aléatoire X ayant une fonction de densité f continue. Soit $y \in \mathbb{R}$, montrer que la fonction $hist_{x_1,\dots,x_n}(y)$ converge en probabilité vers $f(y)$, lorsque $n \to \infty$, $h_n \to 0$ et $nh_n \to \infty$. Indication : le nombre d'observation tombant dans l'intervalle I_{j_n}, donné par $N_n = \sum_{i=1}^n 1_{\{x_i \in I_{j_n}\}}$, suit une loi $\text{Bin}(n, p_n)$ où $p_n = \int_{I_{j_n}} f(x)\text{d}x$. Vous aurez besoin d'utiliser le fait que

$$\left|\frac{N_n}{nh_n} - f(y)\right| \leq \left|\frac{N_n}{nh_n} - \frac{p_n}{h_n}\right| + \left|\frac{p_n}{h_n} - f(y)\right|,$$

ainsi que l'inégalité de Chebyshev (lemme 6.4, p. 163).

Solution 40. Supposons sans perte de généralité que $y \in I_{j_n}$ et dénotons $\sum_{i=1}^{n} 1_{\{x_i \in I_{j_n}\}}$ par $N_n \sim Binom(n, p_n)$. Nous avons alors

$$\begin{aligned} |hist_{x_1,\ldots,x_n}(y) - f(y)| &= \left| \frac{N_n}{nh_n} - f(y) \right| \\ &\leq \left| \frac{N_n}{nh_n} - \frac{p_n}{h_n} \right| + \left| \frac{p_n}{h_n} - f(y) \right|, \end{aligned} \tag{2}$$

où $p_n = \int_{I_{j_n}} f(x)\mathrm{d}x$. Notons que puisque f est continue, nous avons que $\forall \delta > 0, \exists \rho > 0$ tel que $|f(x) - f(y)| \leq \delta$ si $|x - y| \leq \rho$. Puisque $h_n \to 0$ il existe N_δ tel que pour $n > N_\delta$, $h_n \leq \rho$. La longueur de I_{j_n} est h_n, donc pour chaque $n > N_\delta$ on a

$$h_n(f(y) - \delta) = \int_{I_{j_n}} (f(y) - \delta)\,\mathrm{d}x \leq \int_{I_{j_n}} f(x)\,\mathrm{d}x \leq \int_{I_{j_n}} (f(y) + \delta)\,\mathrm{d}x = h_n(f(y) + \delta).$$

Ainsi, pour chaque $n > N_\delta$ on a $|p_n/h_n - f(y)| \leq \delta$. Ceci est vrai pour chaque $\delta > 0$, et on conclut que le deuxième terme de l'expression (2) converge vers 0 lorsque $n \to \infty$.
De plus, par l'inégalité de Chebyshev (lemme 6.4, p. 163 du livre),

$$\mathbb{P}\left(\left| \frac{N_n}{nh_n} - \frac{p_n}{h_n} \right| > \epsilon \right) = \mathbb{P}(|N_n - np_n| > nh_n\epsilon) \leq \frac{np_n(1-p_n)}{(nh_n\epsilon)^2} = \frac{p_n(1-p_n)}{nh_n^2\epsilon^2}.$$

Nous obtenons donc

$$\begin{aligned} \lim_{n\to\infty} \mathbb{P}(|hist_{x_1,\ldots,x_n}(y) - f(y)| > \epsilon) &\leq \lim_{n\to\infty} \left[\mathbb{P}\left(\left| \frac{N_n}{nh_n} - \frac{p_n}{h_n} \right| > \frac{\epsilon}{2} \right) + \mathbb{P}\left(\left| \frac{p_n}{h_n} - f(y) \right| > \frac{\epsilon}{2} \right) \right] \\ &\leq \lim_{n\to\infty} \frac{4p_n(1-p_n)}{nh_n^2\epsilon^2} \\ &= 4 \lim_{n\to\infty} \frac{p_n}{h_n} \cdot \lim_{n\to\infty} \frac{1-p_n}{nh_n\epsilon^2} = 0. \end{aligned}$$

***Exercice 41** (exercice 26)**.** Prouver le lemme 2.20 (p. 60) du livre.
(*L'étoile est la notation standard dans les livres de mathématiques pour des exercices plus difficiles.*)

Solution 41.
($\Rightarrow$) Notons tout d'abord que $X_n \xrightarrow{d} c$ signifie que $\forall x \neq c$

$$\mathbb{P}(X_n \leq x) \overset{n\to\infty}{\longrightarrow} \mathbb{P}(c \leq x) = \begin{cases} 1 & \text{si } x \geq c \\ 0 & \text{si } x < c. \end{cases}$$

Nous pouvons maintenant calculer :

$$\begin{aligned} \mathbb{P}(|X_n - c| > \epsilon) &= \mathbb{P}(X_n > c + \epsilon) + \mathbb{P}(X_n < c - \epsilon) \\ &\leq 1 - \mathbb{P}(X_n \leq c + \epsilon) + \mathbb{P}(X_n \leq c - \epsilon) \\ &\overset{n\to\infty}{\longrightarrow} 1 - \mathbb{P}(c \leq c + \epsilon) + \mathbb{P}(c \leq c - \epsilon) \\ &= 1 - 1 + 0 \\ &= 0. \end{aligned}$$

Nous venons de montrer que $X_n \xrightarrow{p} c$.
($\Leftarrow$) Rappelons tout d'abord que lorsque $A \subseteq B$, alors $\mathbb{P}(A) \leq \mathbb{P}(B)$. Soit $\epsilon > 0$ et $x \neq c$, nous avons :

$$\begin{aligned} \mathbb{P}(X_n \leq x) &= \mathbb{P}(X_n \leq x, |X_n - c| > \epsilon) + \mathbb{P}(X_n \leq x, |X_n - c| \leq \epsilon) \\ &\leq \mathbb{P}(|X_n - c| > \epsilon) + \mathbb{P}(c \leq x + \epsilon) \\ &\xrightarrow{n\to\infty} \mathbb{P}(c \leq x + \epsilon). \end{aligned} \tag{3}$$

L'inégalité vient du fait que l'événement $\{X_n \leq x, |X_n - c| > \epsilon\}$ est inclus dans l'événement $\{|X_n - c| > \epsilon\}$ et que l'événement $\{X_n \leq x, |X_n - c| \leq \epsilon\}$ est inclus dans l'événement $\{c \leq x + \epsilon\}$. Quant à elle la dernière ligne est une conséquence du fait que $X_n \xrightarrow{p} c$.
De façon similaire nous obtenons que :

$$\begin{aligned} \mathbb{P}(c \leq x - \epsilon) &= \mathbb{P}(c \leq x - \epsilon, |X_n - c| > \epsilon) + \mathbb{P}(c \leq x - \epsilon, |X_n - c| \leq \epsilon) \\ &\leq \mathbb{P}(|X_n - c| > \epsilon) + \mathbb{P}(X_n \leq x), \end{aligned}$$

ce qui implique

$$\begin{aligned} \mathbb{P}(X_n \leq x) &\geq \mathbb{P}(c \leq x - \epsilon) - \mathbb{P}(|X_n - c| > \epsilon) \\ &\xrightarrow{n\to\infty} \mathbb{P}(c \leq x - \epsilon). \end{aligned} \tag{4}$$

En combinant les équations (3) et (4) et le fait que ϵ soit arbitraire, nous obtenons finalement que $X_n \xrightarrow{d} c$.

Exercice 42 (exercise 29)**.** Nous allons traiter la question de l'existence d'estimateurs non biaisés.
Soit $Y \sim \mathrm{Bin}(n, p)$, où $p \in]0, 1[$.

(i) Montrer que Y/n est un estimateur non biaisé pour p.

(ii) Montrer qu'il n'existe pas d'estimateur non biaisé pour $1/p$.

(iii) Montrer qu'il n'existe pas d'estimateur non biaisé pour le paramètre naturel $\phi = \log\left(\frac{p}{1-p}\right)$.
Remarque : ϕ s'appelle le *log odds ratio* ou de manière moins anglophone le *log du rapport des chances*.

Solution 42.

(i) L'estimateur Y/n est non-biaisé car

$$\mathbb{E}\left(\frac{Y}{n}\right) = \frac{np}{n} = p.$$

(ii) On cherche une fonction U telle que

$$\frac{1}{p} = \mathbb{E}_p[U(Y)] = \sum_{k=0}^{n} \binom{n}{k} U(k) p^k (1-p)^{n-k}, \qquad \forall p \in]0, 1[.$$

Or, le membre droite de l'équation est un polynôme alors que le membre gauche ne l'est pas. Ainsi, une telle fonction U ne peut pas exister. (Un autre raisonnement serait de dire que la limite du membre gauche de l'équation lorsque $p \searrow 0$ est ∞.)

(iii) Pareil qu'en (ii) : supposons que $V(Y)$ soit un estimateur non biaisé de ϕ, c'est-à-dire que $\mathbb{E}_p(V(Y)) = \phi$. Nous avons alors

$$\sum_{k=0}^{n} \binom{n}{k} V(k) p^k (1-p)^{n-k} = E_p[V(Y)] = \phi = \log\left(\frac{p}{1-p}\right).$$

Le polynôme ci-dessus est de degré inférieur ou égale à n, tandis que ϕ n'est pas un polynôme de degré fini, nous obtenons donc une contradiction.

Exercice 43. Soient $X_1, \ldots, X_n \overset{iid}{\sim}$ Poisson(λ). Définissons les estimateurs $\widehat{\lambda}_n = \overline{X}_n = \sum_{i=1}^n X_i/n$ et $S_n^2 = (n-1)^{-1}\sum_{i=1}^n (X_i - \overline{X}_n)^2$.
Montrer que $\operatorname{Var} S_n^2 \geq \operatorname{Var} \widehat{\lambda}_n$.
Indice : la borne de Cramér–Rao peut s'averer utile.

Solution 43. Remarquons que $\overline{X}_n$ est un estimateur non biaisé pour λ, puisque

$$\mathbb{E}_\lambda(\overline{X}_n) = \frac{n\lambda}{n} = \lambda.$$

On va montrer que $\overline{X}_n$ atteint la borne de Cramér-Rao. Il suffit de calculer le logarithme de la loi de probabilité de Poisson, et de dériver :

$$I(\lambda) = \mathbb{E}\left(\frac{\partial \log f_\lambda(X)}{\partial \lambda}\right)^2 = \mathbb{E}\left(\frac{X}{\lambda} - 1\right)^2 = \frac{\mathbb{E}\, X^2}{\lambda^2} - 2\frac{\mathbb{E}\, X}{\lambda} + 1 = \frac{\lambda^2 + \lambda - 2\lambda^2 + \lambda^2}{\lambda^2} = \frac{1}{\lambda}.$$

Ainsi $I(\lambda) = 1/\lambda$. Comme $\overline{X}_n$ est un estimateur non biaisé de λ, la borne de Cramér–Rao est

$$\operatorname{Var}_\lambda(\overline{X}_n) \geq \frac{1}{nI(\lambda)} = \frac{\lambda}{n}.$$

Or $\operatorname{Var}_\lambda(\overline{X}_n) = \operatorname{Var}(X)/n = \lambda/n$, donc $\overline{X}_n$ atteint cette borne.
Pour S_n^2, on effectue la manipulation suivante (voir l'exercice 4, série 3) :

$$\begin{aligned} S_n^2 &= \frac{1}{n-1}\sum_{i=1}^n X_i^2 - 2X_i\overline{X} + (\overline{X})^2 \\ &= \frac{1}{n-1}\left(\sum_{i=1}^n X_i^2 - n(\overline{X})^2\right) \\ &= \frac{1}{n-1}\left(\sum_{i=1}^n X_i^2 - \frac{(\sum_{i=1}^n X_i)^2}{n}\right). \end{aligned}$$

Puisque $X_i \sim Poi(\lambda)$,

$$\mathbb{E}_\lambda(X_i^2) = \operatorname{Var}_\lambda(X_i) + (\mathbb{E}_\lambda(X_i))^2 = \lambda + \lambda^2.$$

D'après exercice 2, série 4 (exercice 22 du livre), on sait que $Z = \sum_{i=1}^n X_i \sim Poi(n\lambda)$, et donc $\mathbb{E}_\lambda(Z^2) = \operatorname{Var}_\lambda(Z) + (\mathbb{E}_\lambda Z)^2 = n\lambda + (n\lambda)^2$.

Par la linéarité de l'espérance, on écrit

$$\begin{aligned}
\mathbb{E}_\lambda(S_n^2) &= \frac{1}{n-1}\left(\sum_{i=1}^n \mathbb{E}_\lambda(X_i^2) - \frac{1}{n}\mathbb{E}_\lambda\left(\sum_{i=1}^n X_i\right)^2\right) \\
&= \frac{1}{n-1}\left(n(\lambda+\lambda^2) - \frac{n\lambda + n^2\lambda^2}{n}\right) \\
&= \frac{1}{n-1}\left((n-1)\lambda + n\lambda^2 - n\lambda^2\right) \\
&= \lambda.
\end{aligned}$$

Autrement dit, S_n^2 est lui aussi un estimateur non biaisé pour λ. Puisque $\overline{X}_n$ atteint la borne de Cramér-Rao, on sait que

$$\operatorname{Var}\widehat{\lambda}_n \leq \operatorname{Var} S_n^2.$$

Un calcul exacte de la variance de S_n^2 est possible, en utilisant $\operatorname{Var} S_n^2 = \mathbb{E}[S_n^2]^2 - [\mathbb{E}S_n^2]^2$, mais fastidieux.

Exercice 44. Soient $X_1, \dots, X_n \overset{iid}{\sim} \operatorname{Exp}(\lambda)$, où $n > 2$.

(i) Montrer que l'estimateur $\hat{\lambda}_n = (\overline{X})^{-1}$ est consistent pour λ.

(ii) Montrer que $\mathbb{E}_\lambda(\hat{\lambda}_n) = \lambda n/(n-1)$, et trouver un estimateur $\hat{\lambda}_n^{\mathrm{NB}}$ non biaisé de λ.
Indice : utiliser le fait que $Z = \sum_{i=1}^n X_i \sim \operatorname{Gamma}(n, \lambda)$.

(iii) Montrer que $\operatorname{Var}_\lambda(\hat{\lambda}_n) = n^2\lambda^2 / \big((n-1)^2 (n-2)\big)$.

(iv) L'estimateur $\hat{\lambda}_n^{\mathrm{NB}}$ atteint-il la borne inférieure de Cramér–Rao ?

Solution 44.

(i) On a que $X_i,\ i = 1, \dots, n$ sont i.i.d, et $\mathbb{E}_\lambda(X_i) = \frac{1}{\lambda}$. Donc, par la loi des grands nombres,

$$\overline{X} = \frac{1}{n}X_i \xrightarrow{p} \frac{1}{\lambda}.$$

En utilisant la theoréme de l'application continue avec la fonction $g(x) = \frac{1}{x}$, on a

$$\overline{X}^{-1} \xrightarrow{p} \lambda,$$

ce qui dit justement que $\widehat{\lambda}_n = \overline{X}^{-1}$ est consistent pour λ.

(ii) On utilise le fait que $Z = \sum_{i=1}^n X_i \sim \operatorname{Gamma}(n, \lambda)$, et on écrit $\widehat{\lambda}_n = n/Z$. Ainsi

$$\begin{aligned}
\mathbb{E}_\lambda(\hat{\lambda}_n) &= n\int_0^\infty \frac{1}{z} f_{\lambda,n}(z)dz = n\int_0^\infty \frac{1}{z}\cdot\frac{1}{\Gamma(n)}\lambda \mathrm{e}^{-\lambda z}(\lambda z)^{n-1}dz \\
&= \frac{n\Gamma(n-1)}{\Gamma(n)}\lambda\int_0^\infty \frac{1}{\Gamma(n-1)}\lambda \mathrm{e}^{-\lambda z}(\lambda z)^{n-2}dz = \frac{n}{n-1}\lambda.
\end{aligned}$$

L'estimateur $\hat{\lambda}_n^{NB} = \frac{n-1}{n}\hat{\lambda}_n$ est donc non biaisé. La dernière égalité vient du fait que l'intégrale vaut 1, et que $\Gamma(x) = (x-1)\Gamma(x-1)$ pour tout $x > 1$.

(iii) Calculons

$$\begin{aligned}\mathbb{E}_\lambda(\hat{\lambda}_n^2) &= n^2\int_0^\infty \frac{1}{z^2} f_{\lambda,n}(z)dz = n^2\int_0^\infty \frac{1}{z^2}\cdot\frac{1}{\Gamma(n)}\lambda \mathrm{e}^{-\lambda z}(\lambda z)^{n-1}dz \\ &= \frac{n^2\Gamma(n-2)}{\Gamma(n)}\lambda^2\int_0^\infty \frac{1}{\Gamma(n-2)}\lambda \mathrm{e}^{-\lambda z}(\lambda z)^{n-3}dz = \frac{n^2}{(n-1)(n-2)}\lambda^2.\end{aligned}$$

Ainsi

$$\begin{aligned}\mathrm{Var}_\lambda(\hat{\lambda}_n) &= \mathbb{E}_\lambda(\hat{\lambda}_n^2) - [\mathbb{E}_\lambda(\hat{\lambda}_n)]^2 \\ &= \frac{n^2}{(n-1)(n-2)}\lambda^2 - \frac{n^2}{(n-1)^2}\lambda^2 \\ &= \frac{n^2}{(n-1)^2(n-2)}\lambda^2.\end{aligned}$$

(iv) L'information de Fisher $I(\lambda)$ est

$$\begin{aligned}I(\lambda) &= \mathbb{E}\left[\left\{\frac{\partial}{\partial\lambda}\log\left(\lambda\exp\left(-\lambda X_1\right)\right)\right\}^2\right] \\ &= \mathbb{E}\left[\left\{\frac{1}{\lambda} - X_1\right\}^2\right] \\ &= \mathbb{E}\left[\frac{1}{\lambda^2} - \frac{2}{\lambda}X_1 + X_1^2\right] = \frac{1}{\lambda^2},\end{aligned}$$

car $X_1 \sim Exp(\lambda)$ dont l'éspérance est $1/\lambda$ et la variance $1/\lambda^2$. La borne de Cramér–Rao est donc $(nI(\lambda))^{-1} = \lambda^2/n$.
Comme $\mathrm{Var}_\lambda(\hat{\lambda}_n^{\mathrm{NB}}) = \lambda^2/(n-2) > \lambda^2/n$, l'estimateur $\hat{\lambda}_n^{\mathrm{NB}}$ n'atteint (tout juste) pas la borne de Cramér-Rao.

Exercice 45. Soient $X_1, \ldots, X_n \overset{iid}{\sim} \mathrm{Poisson}(\lambda)$.

(i) Montrer que l'estimateur du maximum de vraisemblance $\hat{\lambda}_n$ de λ est consistant et non-biaisé.

(ii) Donner un estimateur (par exemple une simple modification de $\hat{\lambda}_n$) qui est consistant, mais néanmoins biaisé.

Solution 45.

(i) On a $\hat{\lambda}_n = n^{-1}\sum_{i=1}^n X_i$. Par la loi faible des grands nombres, $\hat{\lambda}_n \xrightarrow{p} \mathbb{E}X_1 = \lambda$ et donc $\hat{\lambda}_n$ est consistant. Puisque $\mathbb{E}[\hat{\lambda}_n] = \lambda$, $\hat{\lambda}_n$ est en plus non-biaisé.

(ii) L'estimateur $\tilde{\lambda}_n = \hat{\lambda}_n + \frac{1}{n}$ est un exemple d'estimateur biaisé mais consistant. L'estimateur $\frac{n+1}{n}\hat{\lambda}_n$ en est un autre.

Exercice 46 (exercice 31)**.** Soient $X_1, \ldots, X_n \overset{iid}{\sim} \mathrm{Exp}(\lambda)$, où $n > 2$.

(i) Trouver l'estimateur du maximum de vraisemblance $\hat{\lambda}_n$.

(ii) Déterminer l'estimateur du maximum de vraisemblance $\hat{\theta}_n^{\mathrm{MV}}$ et la borne de Cramér–Rao associés au paramètre $\theta = 1/\lambda$. Peut-on utiliser la proposition 3.17 ?

(iii) Comparer $\widehat{\lambda}_n$ et $\widehat{\theta}_n^{\mathrm{MV}}$ avec les bornes de Cramér–Rao correspondantes. *Attention* : quand l'estimateur est biaisé, le nominateur de la borne de Cramér–Rao n'est pas 1.

Solution 46.

(i) En dérivant la fonction de log vraisemblance (par rapport à λ)

$$\ell_n(\lambda) = \log(\lambda^n e^{-\lambda \sum_{i=1}^n x_i}) = n \log \lambda - \lambda \sum_{i=1}^n x_i,$$

et en la posant égale à zéro, nous obtenons

$$\hat{\lambda}_n = \frac{n}{\sum_{i=1}^n x_i} = \frac{1}{\overline{X}_n}.$$

La fonction ℓ_n étant concave, il s'agit bien d'un maximum.

(ii) Nous pouvons en effet utiliser la proposition 3.17, puisque $\lambda \mapsto \theta = \frac{1}{\lambda}$ sur $]0, \infty[$ est une fonction bijective de λ. Donc $\hat{\theta}_n^{\mathrm{MV}} = 1/\widehat{\lambda}_n = \bar{X}_n$. C'est un estimateur non biaisé de θ.

(iii) Nous savons que

$$\mathbb{E}_\lambda(\hat{\lambda}_n) = \frac{n}{n-1}\lambda \qquad \Longrightarrow \qquad \beta(\lambda) = \mathbb{E}_\lambda \widehat{\lambda}_n - \lambda = \frac{\lambda}{n-1}.$$

L'information de Fisher $I(\lambda)$ est

$$\begin{aligned} I(\lambda) &= \mathbb{E}\left[\left\{\frac{\partial}{\partial \lambda} \log\left(\lambda \exp\left(-\lambda X_1\right)\right)\right\}^2\right] \\ &= \mathbb{E}\left[\left\{\frac{1}{\lambda} - X_1\right\}^2\right] \\ &= \mathbb{E}\left[\frac{1}{\lambda^2} - \frac{2}{\lambda}X_1 + X_1^2\right] = \frac{1}{\lambda^2}, \end{aligned}$$

car $X_1 \sim Exp(\lambda)$ dont l'espérance est $1/\lambda$ et la variance $1/\lambda^2$. La borne de Cramér–Rao est donc

$$\frac{(\beta'(\lambda)+1)^2}{nI(\lambda)} = \frac{(1+1/(n-1))^2}{n/\lambda^2} = \frac{n^2\lambda^2}{n(n-1)^2} = \frac{n\lambda^2}{(n-1)^2}.$$

(Au fait, la borne de Cramér–Rao correspondante à aT est a^2 fois la borne de Cramér–Rao correspondante à T, si $a \in \mathbb{R}$; on aurait donc pu utiliser le fait que la borne de Cramér–Rao pour $\widehat{\lambda}_n^{\mathrm{NB}} = (n-1)\widehat{\lambda}_n/n$ est λ^2/n.)

Or

$$\mathrm{Var}_\lambda(\hat{\lambda}_n) = \frac{n^2}{(n-1)^2(n-2)}\lambda^2 = \frac{n\lambda^2}{(n-1)^2}\frac{n}{n-2} > \frac{n\lambda^2}{(n-1)^2}.$$

L'estimateur $\hat{\lambda}_n$ n'atteint donc (tout juste) pas la borne de Cramér-Rao.

Quant à θ, l'information de Fisher $I(\theta)$ est

$$\begin{aligned} I(\theta) &= \mathbb{E}\left[\left\{\frac{\partial}{\partial\theta}\log\left(\frac{1}{\theta}\exp\left(-\frac{1}{\theta}X_1\right)\right)\right\}^2\right] \\ &= \mathbb{E}\left[\left\{\frac{X_1}{\theta^2}-\frac{1}{\theta}\right\}^2\right] \\ &= \frac{1}{\theta^2}\mathbb{E}\left[\frac{X_1^2}{\theta^2}-\frac{2}{\theta}X_1+1\right]=\frac{1}{\theta^2}. \end{aligned}$$

La borne de Cramér–Rao est donc $\theta^2/n = \mathrm{Var}_\theta(\hat{\theta}_n^{\mathrm{MV}})$, donc $\hat{\theta}_n^{\mathrm{MV}}$ atteint la borne de Cramér-Rao.

Exercice 47 (exercice 33)**.** Un malheureux époux bavarde souvent à son téléphone portable afin d'oublier ses misères. On sait que la longueur de ses jasettes téléphoniques suit une loi exponentielle de paramètre $\lambda > 0$. Longtemps gênée par les conversation de son époux, la femme de ce monsieur malchanceux se mit à mesurer la longueur de celles-ci ; ayant un nombre infini d'observations, elle connaît la valeur précise du paramètre λ.
Lors d'une dispute avec son mari et afin d'avoir un argument plus concret, la femme montra à son époux un échantillon $t_1, \ldots, t_n$ des longueurs de n de ses conversations téléphoniques, et ce, afin de lui prouver qu'il placote au téléphone de manière excessive.
L'homme, tout méfiant, ne croit guère sa femme ; connaissant celle avec laquelle il vit déjà depuis quelques décennies, il la soupçonne d'avoir choisi l'échantillon de manière aléatoire, mais uniquement à partir des conversations qui duraient plus longtemps que la moyenne (théorique) de la longueur des conversations. En supposant ceci, le bavard s'attaque au problème d'estimer le paramètre λ, dont seule son épouse connaît la valeur véritable.
Trouver l'estimateur de maximum de vraisemblance de λ à partir de l'échantillon $t_1, \ldots, t_n$, mais sous l'hypothèse que le monsieur a raison. *Attention* : comme à l'exemple 3.20 (du livre), le support de la distribution dépend de l'état de la nature, c'est-à-dire de la vraie valeur de λ.

Solution 47. L'espérance de la durée des conversations est $1/\lambda$. Lorsque les soupçons du monsieur sont justifiés, la fonction de répartition de la distribution qui génère l'échantillon $t_1, \ldots, t_n$ est

$$F_T(t) = \mathbb{P}\left[Y \le t | Y > \frac{1}{\lambda}\right], \qquad t \ge \frac{1}{\lambda},$$

où $Y \sim Exp(\lambda)$. Grâce à l'absence de mémoire de la distribution exponentielle (cf. exercice 6, série 1), on a pour $t \ge \lambda^{-1}$

$$F_T(t) = 1 - \mathbb{P}\left[Y > t | Y > \frac{1}{\lambda}\right] = 1 - \mathbb{P}\left[Y > t - \frac{1}{\lambda}\right] = 1 - e^{-\lambda(t-1/\lambda)} = 1 - e^{1-\lambda t}.$$

La densité de la variable aléatoire T est donc $f(t;\lambda) = \lambda e^{1-\lambda t}\mathbf{1}\{t \ge \lambda^{-1}\}$. La vraisemblance à partir d'un échantillon $t_1, \ldots, t_n$ s'écrit

$$\begin{aligned} L_n(\lambda;(t_i)) = \prod_{i=1}^n f(t_i;\lambda) &= \lambda^n e^{n-\lambda\sum_{i=1}^n t_i}\prod_{i=1}^n \mathbf{1}\{t_i \ge 1/\lambda\} \\ &= \lambda^n e^{n-\lambda\sum_{i=1}^n t_i}\mathbf{1}\{\lambda \ge 1/t_{(1)}\}, \qquad t_{(1)} = \min\{t_1,\ldots,t_n\}, \end{aligned}$$

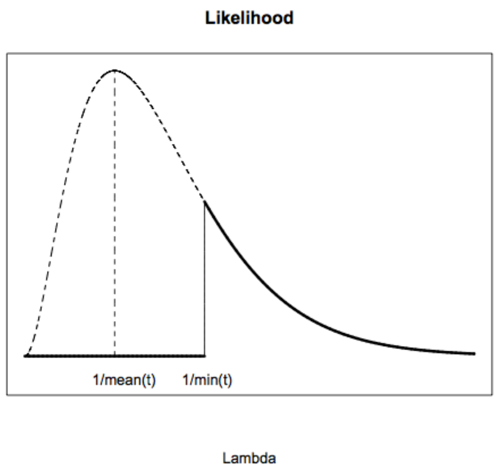
puisque $\prod_{i=1}^n \mathbf{1}\{t_i \geq 1/\lambda\} = 1$ si et seulement si $t_{(1)} \geq 1/\lambda$ si et seulement si $\lambda \geq 1/t_{(1)}$.
Afin de maximiser cette fonction, faisons comme si la fonction indicatrice n'était pas là et dérivons $\ell_n(\lambda; (t_i)) = n\log(\lambda) + n - n\lambda\bar{t}$:

$$\frac{\partial \ell_n}{\partial \lambda} = \frac{n}{\lambda} - n\bar{t}.$$

En posant cette dernière équation égale à zéro, nous obtenons :

$$\frac{\partial \ell_n}{\partial \lambda} = 0 \iff \hat{\lambda} = \frac{1}{\bar{t}}.$$

Malheureusement, puisque $\bar{t} > t_{(1)}$, $\frac{1}{\bar{t}} < \frac{1}{t_{(1)}}$; notre solution ne satisfait donc pas à la condition $\lambda \geq 1/t_{(1)}$ et la vraisemblance vaut zéro. Puisque ℓ_n (et donc L_n) est décroissante sur $[1/t_{(1)}, \infty[$, le maximum sera atteint au premier point où la vraisemblance ne s'annule pas (voir le graphique ci-dessous). L'estimateur est donc $\widehat{\lambda}_n = 1/t_{(1)}$.
Remarque. Il se peut que $\bar{t} = t_{(1)}$, mais même dans ce cas l'estimateur sera $1/t_{(1)} = 1/\bar{t}$. Cette particularité n'arrive cependant qu'avec probabilité zéro, à moins que $n = 1$.



Exercice 48 (exercice 35)**.** Soient $X_1, \ldots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$ où les deux paramètres sont inconnus ($n > 1$). On peut estimer σ^2 par

$$S_n^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X})^2,$$

ou bien par l'estimateur de maximum de vraisemblance $\widehat{\sigma}_n^2 = (n-1)S_n^2/n$ (cf. l'exemple 3.16, p. 75).

(i) Lequel de ces estimateurs est meilleur au sens de l'erreur quadratique moyenne ?
Indication : on a $(n-1)S_n^2/\sigma^2 \sim \chi^2_{n-1}$ (cf. proposition 2.7, p. 51).

(ii) Considérons les estimateurs de la forme aS_n^2 où $a \in \mathbb{R}$. Quelle est la meilleure valeur de a au sens de l'erreur quadratique moyenne ?

Solution 48.

(i) Puisque l'espérance d'une variable aléatoire χ^2_{n-1} est $n-1$ et sa variance est $2(n-1)$, $\mathbb{E}[S_n^2] = \sigma^2$ et $EQM(S_n^2, \sigma^2) = \mathrm{Var}[S_n^2] = 2\sigma^4/(n-1)$.
Puisque $\widehat{\sigma}_n^2 = (n-1)S_n^2/n$, nous avons $\mathbb{E}[\widehat{\sigma}_n^2] = (n-1)\sigma^2/n$ et $\mathrm{Var}[\widehat{\sigma}_n^2] = 2(n-1)\sigma^4/n^2$. Ainsi

$$EQM(\widehat{\sigma}_n^2, \sigma^2) = \left(\frac{n-1}{n}\sigma^2 - \sigma^2\right)^2 + \frac{2(n-1)}{n^2}\sigma^4 = \frac{2n-1}{n^2}\sigma^4 < \frac{2}{n-1}\sigma^4,$$

puisque $\sigma^4 > 0$ et $(2n-1)/n^2 < 2/n < 2/(n-1)$. On remarque que même si $\widehat{\sigma}_n^2$ est biaisé et S_n^2 ne l'est pas, ce dernier a une erreur quadratique moyenne plus élevée.

(ii) Ici l'espérance est $a\sigma^2$ et la variance $2a^2\sigma^4/(n-1)$ de sorte que l'erreur quadratique moyenne vaille

$$(a\sigma^2 - \sigma^2)^2 + \frac{2a^2}{n-1}\sigma^4 = \sigma^4\left((a-1)^2 + \frac{2a^2}{n-1}\right) = \frac{\sigma^4}{n-1}\left((a^2-2a+1)(n-1) + 2a^2\right).$$

C'est une parabole convexe en fonction de a dont l'unique minimum est la racine de l'équation

$$0 = 2a(n-1) + 4a - 2(n-1) = 2a(n+1) - 2(n-1) \quad \Longrightarrow \quad a = \frac{n-1}{n+1}.$$

Ainsi le meilleur estimateur de cette forme est

$$\frac{n-1}{n+1}S_n^2 = \frac{1}{n+1}\sum_{i=1}^n (X_i - \overline{X})^2.$$

Exercice 49. Soient $X_1, \ldots, X_n \overset{iid}{\sim} Unif(0, \theta)$, où $\theta > 0$. Soit $\widehat{\theta}_n$ l'estimateur de maximum de vraisemblance. Trouver $\widehat{\theta}_n$ et montrer que $n(\theta - \widehat{\theta}_n)$ converge en distribution vers une distribution à trouver.

Solution 49. L'estimateur de maximum de vraisemblance est $\widehat{\theta}_n = X_{(n)}$ (cf. l'exemple 3.20, p. 77). On trouve pour $x \geq 0$,

$$\begin{aligned}\mathbb{P}(n(\theta - \widehat{\theta}_n) \leq x) = \mathbb{P}\left(X_{(n)} \geq \theta - \frac{x}{n}\right) &= 1 - \mathbf{1}\{x \leq n\theta\}\left(1 - \frac{x}{\theta n}\right)^n \\ &\to 1 - \exp\left(-\frac{x}{\theta}\right), \quad n \to \infty.\end{aligned}$$

Ainsi $n(\theta - \widehat{\theta}_n) \xrightarrow{d} Exp(1/\theta)$.

Exercice 50 (exercices 36 et 37)**.**

(i) Considérons la représentation usuelle d'une famille exponentielle

$$f(x; \theta) = \exp(\eta(\theta)T(x) - d(\theta) + S(x)), \qquad x \in \mathcal{X}, \quad \theta \in \Theta,$$

où $\Theta \subseteq \mathbb{R}$ est un ouvert et η est deux fois continûment dérivable et inversible avec la dérivée jamais nulle. Soient $X_1, \ldots, X_n \overset{iid}{\sim} f(x; \theta)$. Montrer que

$$\mathbb{E}\left[\frac{\partial}{\partial\theta}\log f(X_1, \ldots, X_n; \theta)\right] = 0, \qquad \text{et}$$

$$\mathbb{E}\left[\left(\frac{\partial}{\partial\theta}\log f(X_1,\ldots,X_n;\theta)\right)^2\right] = -\mathbb{E}\left[\frac{\partial^2}{\partial\theta^2}\log f(X_1,\ldots,X_n;\theta)\right]. \tag{5}$$

Indication : ce n'est pas pour rien qu'on a fait l'exercice 35.

(ii) * Soit $f(x;\theta)$ un modèle paramétrique régulier (pas forcément une famille exponentielle !) tel que

$$\mathcal{X} = \{x \in \mathbb{R} : f(x;\theta) > 0\}$$

ne dépend pas de θ, et que f est doublement dérivable par rapport à θ. Soient en plus $X_1,\ldots,X_n \overset{iid}{\sim} f(x;\theta)$. Montrer que l'égalité (5) est équivalente à une condition de régularité qui dit que l'on peut interchanger la dérivée et l'intégrale.

Indication : il faut absolument se rendre compte que pour chaque fonction $g : \mathbb{R}^n \to \mathbb{R}$,

$$\mathbb{E}[g(X)] = \int_{\mathcal{X}^n} g(\vec{x})f(\vec{x};\theta)\,\mathrm{d}\vec{x} \qquad \text{quand cette intégrale existe} \quad (\vec{x} = (x_1,\ldots,x_n) \in \mathbb{R}^n).$$

Solution 50.

(i) Remarquons que

$$\ell_n(\theta) = \log f(X_1,\ldots,X_n;\theta) = \eta(\theta)\sum_{i=1}^n T(X_i) - nd(\theta) + \sum_{i=1}^n S(X_i);$$

$$\ell_n'(\theta) = \eta'(\theta)\sum_{i=1}^n T(X_i) - nd'(\theta) = n(\eta'(\theta)\overline{T} - d'(\theta));$$

$$\ell_n''(\theta) = \eta''(\theta)\sum_{i=1}^n T(X_i) - nd''(\theta) = n(\eta''(\theta)\overline{T} - d''(\theta)).$$

Par l'exercice 35, $\mathbb{E}[\ell_n'(\theta)] = n(\eta'(\theta)\mathbb{E}[\overline{T}] - d'(\theta)) = 0$ et

$$\mathbb{E}[(\ell_n'(\theta))^2] = \operatorname{Var}[\ell'(\theta)] = n^2(\eta'(\theta))^2\operatorname{Var}[\overline{T}] = n\frac{d''(\theta)\eta'(\theta) - d'(\theta)\eta''(\theta)}{\eta'(\theta)};$$

$$\mathbb{E}[\ell_n''(\theta)] = n(\eta''(\theta)\mathbb{E}[\overline{T}] - d''(\theta)) = n\left(\eta''(\theta)\frac{d'(\theta)}{\eta'(\theta)} - d''(\theta)\right) = n\frac{d'(\theta)\eta''(\theta) - d''(\theta)\eta'(\theta)}{\eta'(\theta)},$$

tel que requis.

(ii) Soit $\ell_n(\theta; X_1,\ldots,X_n) = \log f(X_1,\ldots,X_n;\theta)$. Afin d'alléger la notation (souvent quelque peu fastidieuse en statistiques), nous allons simplement écrire f et ℓ_n. Lorsqu'on prend une dérivée, cela se fait toujours par rapport à θ. (En fait il n'a souvent pas de sens de dériver par rapport à x, par exemple lorsque l'espace $\mathcal{X}$ est discret.) Avec cette notation, la question est : est-ce que $\mathbb{E}[\ell_n''] = -\mathbb{E}[(\ell_n')^2]$?

Dérivons : $\ell_n' = f'/f$ et $\ell_n'' = (f''f - f'f')/f^2$. Par conséquent, $\mathbb{E}[(\ell_n')^2] = -\mathbb{E}[\ell_n'']$ si et seulement si

$$\int_{\mathcal{X}^n}\frac{(f')^2}{f}\,\mathrm{d}\vec{x} = \int_{\mathcal{X}^n}(\ell_n')^2 f\,\mathrm{d}\vec{x} = \mathbb{E}[(\ell_n')^2] = -\mathbb{E}[\ell_n'']$$

$$= -\int_{\mathcal{X}^n}\left(\frac{f''}{f} - \frac{(f')^2}{f^2}\right)f\,\mathrm{d}\vec{x} = \int_{\mathcal{X}^n}\frac{(f')^2}{f}\,\mathrm{d}\vec{x} - \int_{\mathcal{X}^n}f''\,\mathrm{d}\vec{x}.$$

De manière équivalente, $0 = \int_{\mathcal{X}^n} f'' \,\mathrm{d}\vec{x}$ ou bien :

$$\frac{\partial^2}{\partial\theta^2}\int_{\mathcal{X}^n} f(\vec{x};\theta)\,\mathrm{d}\vec{x} = \frac{\partial^2}{\partial\theta^2}1 = 0 = \int_{\mathcal{X}^n} f''\,\mathrm{d}\vec{x} = \int_{\mathcal{X}^n}\frac{\partial^2}{\partial\theta^2} f\,\mathrm{d}\vec{x},$$

car $f(\vec{x};\theta)$ est une fonction de densité pour n'importe quel θ. En d'autres mots, $\mathbb{E}[\ell_n''(\theta)] = -\mathbb{E}[(\ell_n'(\theta)^2)]$ est équivalent au fait de pouvoir interchanger la dérivée seconde et l'intégrale comme le font nos amis les physiciens.

Exercice 51. Soit la variable aléatoire X, dont la densité est donnée par

$$f(x;\theta) = \begin{cases} \theta x^{\theta-1}, & \text{si } 0 < x \leq 1\,; \\ 0, & \text{sinon,} \end{cases}$$

où $\theta > 0$ est un paramètre inconnu. Trouver, sans calculer aucune intégrale, $\mathbb{E}[\log X]$ et $\mathbb{E}[(\log X)^2]$.
Remarque. Cette méthode est beaucoup moins laborieuse que de calculer explicitement

$$\int_0^1 \theta x^{\theta-1}\log x\,\mathrm{d}x \qquad \text{et} \qquad \int_0^1 \theta x^{\theta-1}(\log x)^2\,\mathrm{d}x.$$

Solution 51. Il s'agit bien d'une famille exponentielle, où

$$\begin{aligned} \ell_1(\theta) &= \log\theta + (\theta-1)\log X; \\ \ell_1'(\theta) &= \frac{1}{\theta} + \log X; \\ \ell_1''(\theta) &= -\frac{1}{\theta^2}. \end{aligned}$$

Or, $\mathbb{E}[\ell_1'(\theta)] = 0$, et par conséquent $\mathbb{E}[\log X] = -1/\theta$. De plus, d'après l'exercice 1,

$$\frac{1}{\theta^2} = -\mathbb{E}[\ell_1''(\theta)] = \mathbb{E}[(\ell_1'(\theta))^2] = \frac{1}{\theta^2} + \frac{2\mathbb{E}[\log X]}{\theta} + \mathbb{E}[(\log X)^2] = \frac{1}{\theta^2} - \frac{2}{\theta^2} + \mathbb{E}[(\log X)^2],$$

donc $\mathbb{E}[(\log X)^2] = 2\theta^{-2}$.

Exercice 52. Soit $X \sim \mathrm{Exp}(\lambda)$, où $\lambda > 0$. Montrer que $Y = aX \sim \mathrm{Exp}(\lambda/a)$ pour $a > 0$.

Solution 52. La densité de X est $f_X(x) = \lambda e^{-\lambda x}\mathbf{1}\{x \geq 0\}$ et grâce au corollaire 1.31 (p. 27) la densité de aX est $a^{-1}f_X(x/a) = (\lambda/a)e^{-(\lambda/a)x}\mathbf{1}\{x \geq 0\}$. Par miracle, il s'agit de la densité d'une variable aléatoire exponentielle de paramètre λ/a.

Exercice 53. Nous avons montré une sorte de théorème centrale limite pour les familles exponentielles (théorème 3.23, p. 81 ; corollaire 3.27, p. 84). Nous verrons dans cet exercice deux exemples de ce qui se passe en dehors du cadre des familles exponentielles.
Considérons $\widehat{\lambda}_n$, l'estimateur de l'exercice 47. Trouver une suite de nombres réels a_n telle que $a_n(\lambda - \widehat{\lambda}_n)$ converge en distribution vers une distribution non dégénérée.
Indication : utiliser l'exercice 13 et l'exercice 52.

Solution 53. L'estimateur de maximum de vraisemblance est $\widehat{\lambda}_n = 1/t_{(1)}$ (cf. exercice 47). Or $t_{(1)} - 1/\lambda = \tilde{t}_{(1)} \sim Exp(n\lambda)$, où $\tilde{t} = t - 1/\lambda \sim Exp(\lambda)$ (cf. exercice 50).

Solution ≪ intélligente ≫. Par l'exercice 52, $n(t_{(1)} - 1/\lambda) \sim Exp(\lambda)$. Appliquons la méthode delta avec $g(t) = -1/t$ et encore une fois l'exercice 3 pour conclure

$$n(\lambda - \widehat{\lambda}_n) = n(\lambda - 1/t_{(1)}) = n(g(t_{(1)}) - g(1/\lambda)) \xrightarrow{d} Exp(\lambda)\lambda^2 \sim Exp(1/\lambda).$$

Solution ≪ brute-force ≫. On peut calculer la distribution exacte de $a_n(\lambda - \widehat{\lambda}_n)$, puisque c'est une fonction de $t_{(1)} - 1/\lambda$ dont on connaît la distribution : soit $x \geq 0$.

$$\begin{aligned}\mathbb{P}\left(a_n(\lambda - \widehat{\lambda}_n) \leq x\right) &= \mathbb{P}\left(\widehat{\lambda}_n \geq \lambda - \frac{x}{a_n}\right) \\ &= \mathbb{P}\left(t_{(1)} \leq \frac{a_n}{a_n\lambda - x}\right) \\ &= \mathbb{P}\left(t_{(1)} - \frac{1}{\lambda} \leq \frac{x}{\lambda(a_n\lambda - x)}\right) \\ &= 1 - \exp\left(\frac{-nx}{a_n\lambda - x}\right), \qquad \text{ou } 1 \text{ si } x \geq a_n\lambda.\end{aligned}$$

On aimerait que la limite de cette probabilité soit une fonction qui dépend de x. Si $a_n/n \to 0$ l'exponentielle converge vers 0 et donc la probabilité converge vers 1, et ce, quelque soit la valeur de x. Il faut donc que $a_n \geq O(n)$ et en particulier $a_n \to \infty$, ce qui implique que pour x fixé, $x < a_n\lambda$ pour n suffisamment grand. On a

$$\lim_{n\to\infty} 1 - \exp\left(\frac{-nx}{a_n\lambda - x}\right) = 1 - \exp\left(\lim_{n\to\infty} \frac{-nx}{a_n\lambda - x}\right) = 1 - \exp\left(\frac{-x}{\lambda} \lim_{n\to\infty} \frac{n}{a_n}\right),$$

car $a_n \to \infty$ donc λx devient négligeable lorsque $n \to \infty$. Si $a_n/n \to \infty$ la limite est 0 qui ne dépend pas de x. Il faut donc que $\lim a_n/n \in]0, \infty[$, et on peut choisir par exemple $a_n = n$.
Remarque. Puisque $\lambda \geq \widehat{\lambda}_n$, nous ne pouvons pas nous attendre à ce que la distribution limite de $a_n(\lambda - \widehat{\lambda}_n)$ soit normale ; en effet, n'importe quelle distribution limite est forcément non-négative ! De même pour $a_n(\theta - \widehat{\theta}_n)$.

Exercice 54. (i). Soit $X = (x_1, \ldots, x_n)^T$ une image de dimension 1. Supposons que l'on puisse uniquement observer une version de cette image sur laquelle il y a du bruit numérique, i.e, que l'on observe $Y = (y_1, \ldots, y_n)^T$, où chaque pixel s'écrit comme

$$y_i = x_i + \varepsilon_i,$$

où $\varepsilon_i \overset{iid}{\sim} \mathcal{N}(0, \sigma^2)$. Trouver une estimation de l'image originale X par la méthode du maximum de vraisemblance.

(ii). Supposons maintenant que l'on vous donne une information supplémentaire sur l'allure de l'image : l'image est en fait une ligne, où chaque pixel satisfait la relation

$$y_i = a + bx_i + \varepsilon_i.$$

Calculer l'estimateur du maximum de vraisemblance des paramètres a et b.

Solution 54. (i). Etant donné l'image originale x_i, les y_i suivent la distribution du bruit, et donc $y_i \sim N(x_i, \sigma^2)$. La log-vraisemblance de $\{x_i\}$ est donc donnée par

$$\ell_n(x_i; y_i, \sigma^2) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - x_i)^2.$$

En posant les dérivés de la log-vraisemblance égales à zéro, nous obtenons

$$\hat{x_i} = y_i.$$

Alors, nous avons montré que si on a une seule observation y_i par pixel, la vraisemblance ne nous donne aucune d'information supplémentaire sur l'image.

(ii). Maintenant, les y_i sont indépendants et sont distribués comme

$$y_i \sim \mathcal{N}(a + bx_i, \sigma^2).$$

La log-vraisemblance de $\{a,\ b\}$ est donc

$$\ell_n(a, b; y) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - a - bx_i)^2$$

avec les derivées partielles

$$\begin{aligned}\frac{\partial \ell_n}{\partial a} &= -2\sum_i \frac{(y_i - a - bx_i)}{\sigma^2}\\ \frac{\partial \ell_n}{\partial b} &= -2\sum_i \frac{(y_i - a - bx_i)}{\sigma^2}x_i.\end{aligned}$$

En posant les dérivés de égales à zéro, nous obtenons

$$\begin{aligned}\frac{\partial \ell_n}{\partial a} = 0 \quad &\Leftrightarrow \quad \sum_i \frac{y_i}{\sigma^2} - \frac{na}{\sigma^2} - \frac{b}{\sigma^2}\sum_i x_i = 0\\ \frac{\partial \ell_n}{\partial b} = 0 \quad &\Leftrightarrow \quad \sum_i \frac{x_i y_i}{\sigma^2} - a\sum_i \frac{x_i}{\sigma^2} - b\sum_i \frac{x_i^2}{\sigma^2} = 0.\end{aligned}$$

Soit $\overline{x} = n^{-1}\sum_{i=1}^{n} x_i$, et soient $\overline{x^2}$, $\overline{y}$ et $\overline{xy}$ definis de la même maniére. On trouve finalement que

$$\begin{aligned}\hat{a} &= \frac{\overline{x^2}\overline{y} - \overline{x}\cdot\overline{xy}}{\overline{x^2} - \overline{x}^2}\\ \hat{b} &= \frac{\overline{xy} - \overline{x}\cdot\overline{y}}{\overline{x^2} - \overline{x}^2}\end{aligned}$$

Finalement, observez que $(\hat{a}, \hat{b})$ minimisent la somme des carrés résiduels $\sum_i (y_i - a - bx_i)^2$, et donc nous les appelons aussi les *estimateurs des moindres carrés*.

Exercice 55. Soient $x_1, \ldots, x_n \overset{iid}{\sim} \text{Gamma}(r, 1)$. Trouver l'estimateur des moments $\widehat{r}^{\text{mom}}$ de r, et la loi limite de

$$\sqrt{n}\left(\widehat{r}^{\text{mom}} - \frac{(\log\Gamma(\cdot))'(\widehat{r}^{\text{mom}}) - \overline{\log}X}{(\log\Gamma(\cdot))''(\widehat{r}^{\text{mom}})} - r\right)$$

où $\overline{\log}X = n^{-1}\sum \log x_i$.

Solution 55. On trouve $\widehat{r}^{\text{mom}}$ en résolvent l'équation

$$\overline{x}_n = \mathbb{E}_{\widehat{r}^{\text{mom}}} x_1 = \widehat{r}^{\text{mom}}.$$

Grâce au théorème limit centrale, $\sqrt{n}(\widehat{r}^{\text{mom}} - r) \to N(0, \text{Var}(x_1)) = N(0, r)$. En définissant

$$\beta_n = \widehat{r}^{\text{mom}} - \frac{\ell_n'(\widehat{r}^{\text{mom}})}{\ell_n''(\widehat{r}^{\text{mom}})} = \widehat{r}^{\text{mom}} - \frac{(\log\Gamma)'(\widehat{r}^{\text{mom}}) - \overline{\log}X}{(\log\Gamma)''(\widehat{r}^{\text{mom}})},$$

et remarquant que Γ est suffisamment régulière, on obtient

$$\sqrt{n}(\beta_n - r) \to N\left(0, \frac{1}{I(r)}\right) = N\left(0, \frac{1}{(\log\Gamma)''(r)}\right).$$

Exercice 56. Soit $X_1, \ldots, X_n$ un échantillon i.i.d. tiré d'une distribution de densité

$$f(x;\theta) = \begin{cases} 3\theta^3 x^{-4}, & \text{si } x \geq \theta, \\ 0, & \text{sinon,} \end{cases}$$

où $\theta > 0$.

(i) Trouver l'estimateur $\hat{\theta}_n^{\text{MoM}}$ de θ par la méthode des moments.

(ii) Trouver l'estimateur du maximum de vraisemblance $\hat{\theta}_n^{\text{MV}}$ de θ.

(iii) Montrer que $\hat{\theta}_n^{\text{MoM}}$ est non-biaisé, tandis que $\hat{\theta}_n^{\text{MV}}$ est un estimateur biaisé.

(iv) Calculer l'erreur quadratique moyenne de $\hat{\theta}_n^{\text{MoM}}$ et de $\hat{\theta}_n^{\text{MV}}$. Quel estimateur est le meilleur au sens de l'erreur quadratique moyenne ?

Solution 56.

(i) On a

$$\mathbb{E}[X_1] = m(\theta) = \int_\theta^\infty 3\theta^3 x^{-3}\,\mathrm{d}x = \frac{3}{2}\theta.$$

On obtient donc l'équation et la solution suivantes :

$$\frac{3}{2}\hat{\theta}_n^{\text{MoM}} = \frac{1}{n}\sum_{i=1}^n X_i \quad \Longleftrightarrow \quad \hat{\theta}_n^{\text{MoM}} = \frac{2}{3n}\sum_{i=1}^n X_i.$$

(ii) La vraisemblance de θ est

$$L_n(\theta) = \prod_{i=1}^n 3\theta^3 X_i^{-4}\mathbf{1}\{X_i \geq \theta\} = 3^n\theta^{3n}\prod_{i=1}^n X_i^{-4}\mathbf{1}\{X_{(1)} \geq \theta\}.$$

Pour $\theta \in]0, X_{(1)}]$, la vraisemblance est une fonction strictement croissante et pour $\theta \in]X_{(1)}, \infty[$, $L_n(\theta) = 0$. Il s'en suit que $\hat{\theta}_n^{\text{MV}} = X_{(1)} = \min(X_1, \ldots, X_n)$.

(iii) Pour $\hat{\theta}_n^{\text{MoM}}$, on a

$$\mathbb{E}\left[\hat{\theta}_n^{\text{MoM}}\right] = \frac{2}{3n}\sum_{i=1}^{n}\mathbb{E}[X_i] = \frac{2}{3n}\cdot n\cdot\frac{3}{2}\theta = \theta$$

et $\hat{\theta}_n^{\text{MoM}}$ est donc non-biasié.
Pour trouver l'espérance de $\hat{\theta}_n^{\text{MV}} = X_{(1)}$, il faut tout d'abord trouver la distribution de $X_{(1)}$: pour $t \geq \theta$,

$$\mathbb{P}(X_{(1)} \leq t) = 1 - \mathbb{P}(X_{(1)} > t) = 1 - \prod_{i=1}^{n}\mathbb{P}(X_i > t) = 1 - \prod_{i=1}^{n}\int_t^\infty 3\theta^3 x^{-4}\,\mathrm{d}x = 1 - \left(\frac{\theta}{t}\right)^{3n}.$$

Pour $t < \theta$ cette probabilité vaut 0. Ainsi, la fonction de densité de $X_{(1)}$ est

$$f_{X_{(1)}}(t) = \frac{d}{dt}\left(1 - \left(\frac{\theta}{t}\right)^{3n}\right) = 3n\theta^{3n}t^{-3n-1}, \quad t \in [\theta, \infty),$$

et on obtient

$$\mathbb{E}\left[\hat{\theta}_n^{\text{MV}}\right] = \mathbb{E}[X_{(1)}] = \int_\theta^\infty 3n\theta^{3n}t^{-3n}\,\mathrm{d}t = \frac{3n}{3n-1}\theta.$$

Le biais de $\hat{\theta}_n^{\text{MV}}$ est donc biais $\left[\hat{\theta}_n^{\text{MV}}\right] = \mathbb{E}\left[\hat{\theta}_n^{\text{MV}}\right] - \theta = \frac{1}{3n-1}\theta \neq 0$.

(iv) On a

$$\mathbb{E}[X_1^2] = \int_\theta^\infty 3\theta^3 x^{-2}\,\mathrm{d}x = 3\theta^2,$$

de sorte que $\mathrm{Var}[X_1] = \mathbb{E}[X_1^2] - \mathbb{E}[X_1]^2 = \frac{3}{4}\theta^2$. Puisque les X_i sont iid, on a

$$\mathrm{Var}\left[\hat{\theta}_n^{\text{MoM}}\right] = \frac{4}{9n^2}\sum_{i=1}^{n}\mathrm{Var}[X_i] = \frac{1}{3n}\theta^2$$

et l'erreur quadratique moyenne est

$$\mathrm{EQM}\left[\hat{\theta}_n^{\text{MoM}}\right] = \mathrm{Var}\left[\hat{\theta}_n^{\text{MoM}}\right] = \frac{1}{3n}\theta^2,$$

où on a utilisé le fait que $\hat{\theta}_n^{\text{MoM}}$ est non-biaisé.
Pour $X_{(1)}$, on obtient

$$\mathbb{E}[X_{(1)}^2] = \int_\theta^\infty 3n\theta^{3n}t^{-3n+1}\,\mathrm{d}t = \frac{3n}{3n-2}\theta^2.$$

Donc

$$\mathrm{Var}\left[\hat{\theta}_n^{\text{MV}}\right] = \mathbb{E}[X_{(1)}^2] - \mathbb{E}[X_{(1)}]^2 = \frac{3n}{3n-2}\theta^2 - \left(\frac{3n}{3n-1}\right)^2\theta^2$$

et on obtient

$$\begin{aligned}\mathrm{EQM}\left[\hat{\theta}_n^{\text{MV}}\right] = \text{biais}\left[\hat{\theta}_n^{\text{MV}}\right]^2 + \mathrm{Var}\left[\hat{\theta}_n^{\text{MV}}\right] &= \frac{1}{(3n-1)^2}\theta^2 + \frac{3n}{3n-2}\theta^2 - \left(\frac{3n}{3n-1}\right)^2\theta^2 \\ &= \frac{2}{(3n-1)(3n-2)}\theta^2.\end{aligned}$$

Par un calcul standard, on obtient que

$$\mathrm{EQM}\left[\hat{\theta}_n^{\mathrm{MV}}\right] < \mathrm{EQM}\left[\hat{\theta}_n^{\mathrm{MoM}}\right] \quad \Longleftrightarrow \quad n \geq 2.$$

De plus, quand $n = 1$, on a $\mathrm{EQM}\left[\hat{\theta}_n^{\mathrm{MV}}\right] > \mathrm{EQM}\left[\hat{\theta}_n^{\mathrm{MoM}}\right]$. Donc, pour $n = 1$, l'estimateur $\hat{\theta}_n^{\mathrm{MoM}}$ est meilleur que $\hat{\theta}_n^{\mathrm{MV}}$, mais pour chaque $n \geq 2$ l'estimateur de maximum de vraisemblance est meilleur.

Exercice 57. Soit $X_1, \ldots, X_n$ un échantillon i.i.d. tiré de la distribution uniforme sur $[0, \theta]$ où le paramètre $\theta > 0$ est inconnue. Dans les exercices précédentes on a trouvé l'estimateur du maximum de vraisemblance $\hat{\theta}_n^{\mathrm{MV}} = X_{(n)}$.

(i) Trouver l'estimateur $\hat{\theta}_n^{\mathrm{MoM}}$ de θ par la méthode des moments. Montrer qu'il est non-biaisé.

(ii) Modifier l'estimateur $\hat{\theta}_n^{\mathrm{MV}}$, par example en multipliant par un constant, pour le rendre non-biaisé. Dénoter cet estimateur $\hat{\theta}_n^{\mathrm{MV,modif}}$.

(iii) Calculer l'erreur quadratique moyenne de $\hat{\theta}_n^{\mathrm{MoM}}$ et de $\hat{\theta}_n^{\mathrm{MV,modif}}$. Quel estimateur est le meilleur au sens de l'erreur quadratique moyenne ?

(iv) Commenter la vitesse de convergence de l'erreur quadratique moyenne de ces deux estimateur.

Solution 57. (i) Selon la définition du méthode des moments :

$$\frac{X_1 + \cdots + X_n}{n} = \frac{\hat{\theta}_n^{\mathrm{MoM}}}{2}$$

Alors, $\hat{\theta}_n^{\mathrm{MoM}} = 2\bar{X}$. De plus, $\mathbb{E}[2\bar{X}] = 2 \cdot \frac{n\theta}{n} = \theta$, donc l'estimateur est non-biaisé.

(ii) Rappelle que la fonction de répartition de $X_{(n)}$, c'est $F_{X_{(n)}}(x) = (x/\theta)^n$ pour $x \in [0, \theta]$. Donc, $f_{X_{(n)}}(x) = \frac{n}{\theta}(x/\theta)^{n-1}$ pour $x \in [0, \theta]$. Calculer l'espérance :

$$\mathbb{E}[\hat{\theta}_n^{\mathrm{MV}}] = \mathbb{E}[X_{(n)}] = \int_0^\theta x \frac{n}{\theta}\left(\frac{x}{\theta}\right)^{n-1} dx = \frac{n\theta}{n+1}$$

Alors on peut mettre, $\hat{\theta}_n^{\mathrm{MV,modif}} = \left(1 + \frac{1}{n}\right) X_{(n)}$.

(iii) Les deux estimateur sont non-biaisés, alors :

$$\begin{aligned}
\mathrm{MSE}\left[\hat{\theta}_n^{\mathrm{MV,modif}}\right] = \mathrm{Var}\left[\hat{\theta}_n^{\mathrm{MV,modif}}\right] &= \mathbb{E}\left[\left(\hat{\theta}_n^{\mathrm{MV,modif}}\right)^2\right] - \theta^2 \\
&= \int_0^\theta \left[\left(1 + \frac{1}{n}\right) x\right]^2 \frac{n}{\theta}\left(\frac{x}{\theta}\right)^{n-1} dx - \theta^2 \\
&= \frac{\theta^2}{n(n+2)} \\
\mathrm{MSE}\left[\hat{\theta}_n^{\mathrm{MoM}}\right] = \mathrm{Var}\left[\hat{\theta}_n^{\mathrm{MoM}}\right] &= \frac{4}{n^2} \cdot n\mathrm{Var}\left[X_1\right] \\
&= \frac{4}{n^2} \cdot n\left[\int_0^\theta x^2\left(\frac{1}{\theta}\right) dx - \left(\frac{\theta}{2}\right)^2\right] \\
&= \frac{\theta^2}{3n}
\end{aligned}$$

(iv) L'erreur quadratique moyenne de l'EMV, MSE $\left[\hat{\theta}_n^{\text{MV,modif}}\right]$, converge vers 0 avec la vitesse quadratique grâce à n^{-2}. Au contraire, l'erreur quadratique moyenne de l'estimateur par le méthode des moments, MSE $\left[\hat{\theta}_n^{\text{MoM}}\right]$, converge vers 0 avec la vitesse linéaire (grâce à n^{-1}).

Exercice 58. Soit $X_1, \ldots, X_n$un échantillon i.i.d. tiré de la distribution binomial avec les deux paramètres m et p inconnues. Trouver $\hat{m}$, $\hat{p}$ les estimateurs des m et p par la méthode des moments. Montrer que cela peut arriver que $\hat{m} \notin \{0, 1, \ldots\}$ ou $\hat{p} \notin (0, 1)$.

Solution 58. Le système des equations pour les deux premiers moments :

$$\bar{X} = mp, \qquad \frac{1}{n}\sum_{i=1}^{n} X_i^2 = mp(1-p) + m^2p^2.$$

Donc :

$$\hat{p} = \frac{\bar{X}}{\hat{m}}, \qquad \hat{m} = \frac{(\bar{X})^2}{\bar{X} - \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2}.$$

Si la moyenne échantillonnale est plus petite que la variance échantillonnale, $\hat{m}$ et $\hat{p}$ sont négatives.

***Exercice 59.** (un exercice théorétique) Soit $f(x;\theta) = \exp(T(x)\eta(\theta) - d(\theta) + S(x))$ une famille exponentielle non dégénérée, où l'espace des paramètres Θ est ouvert, et soit $x_1, x_2, \ldots, x_n$ un échantillon iid tiré de $f(x;\theta_0)$ pour un certain θ_0. Soit α_n n'importe quel estimateur tel que $\sqrt{n}(\alpha_n - \theta_0) \to V$ pour une variable aléatoire V. Imaginons qu'on cherche à approximer l'estimateur de maximum de vraisemblance $\widehat{\theta}_n$ avec une seule itération de Newton–Raphson,

$$\beta_n = \alpha_n - \frac{\ell_n'(\alpha_n)}{\ell_n''(\alpha_n)}.$$

En supposant que $\eta \in C^3(\Theta)$, montrer que

$$\sqrt{n}(\beta_n - \theta_0) \to N\left(0, \frac{1}{I(\theta_0)}\right),$$

où $I(\theta_0)$ est l'information de Fisher, et commenter ce résultat.
Indice : faire une développement de Taylor d'ordre 2 de ℓ_n' autour de θ_0, et remarque que cette fonction (aléatoire !) est une somme de variables aléatoires iid.

Solution 59. On écrit

$$\ell_n'(\alpha_n) = \ell_n'(\theta_0) + (\alpha_n - \theta_0)\ell_n''(\theta_0) + \frac{1}{2}(\alpha_n - \theta_0)^2 \ell_n'''(\alpha_n^*),$$

où $|\alpha_n^* - \theta_0| \leq |\alpha_n - \theta_0|$, de sorte que

$$\begin{aligned}\sqrt{n}(\beta_n - \theta_0) &= \sqrt{n}(\alpha_n - \theta_0) + \sqrt{n}(\beta_n - \alpha_n) \\ &= -\frac{n^{-1/2}\ell_n'(\theta_0)}{n^{-1}\ell_n''(\alpha_n)} + \sqrt{n}(\alpha_n - \theta_0)\left(1 - \frac{n^{-1}\ell_n''(\theta_0)}{n^{-1}\ell_n''(\alpha_n)} - \frac{1}{2}(\alpha_n - \theta_0)\frac{n^{-1}\ell_n'''(\alpha_n^*)}{n^{-1}\ell_n''(\alpha_n)}\right).\end{aligned}$$

Le fait que $\sqrt{n}(\alpha_n - \theta_0) \xrightarrow{d} V$ implique que $\alpha_n \xrightarrow{p} \theta_0$ (voir le corollaire 3.26, p 84), et donc aussi $\alpha_n^* \xrightarrow{p} \theta_0$. Rappelons que :

$$\begin{aligned}
\mathbb{E}T(X_1) &= \frac{d'(\theta_0)}{\eta'(\theta_0)} \\
\mathrm{Var}(T(X_1)) &= \frac{d''(\theta_0)\eta'(\theta_0) - d'(\theta_0)\eta''(\theta_0)}{[\eta'(\theta_0)]^3} > 0 \\
I(\theta_0) &= \frac{d''(\theta_0)\eta'(\theta_0) - d'(\theta_0)\eta''(\theta_0)}{\eta'(\theta_0)} > 0
\end{aligned}$$

Puisque $\eta \in C^3(\mathbb{R})$, l'exercice 35 montré que $d \in C^3(\mathbb{R})$. Donc, la loi faible des grands nombres, les théorèmes de Slutsky, de l'application continue, et centrale limite impliquent les convergences (en distribution) suivantes :

$$\begin{aligned}
n^{-1/2}\ell_n'(\theta_0) = n^{1/2}\left(\eta'(\theta_0)\overline{T}_n - d'(\theta_0)\right) &\to N\left(0, [\eta'(\theta_0)]^2\mathrm{Var}(T(X_1))\right) = N\left(0, I(\theta_0)\right); \\
n^{-1}\ell_n''(\theta_0) = \eta''(\theta_0)\overline{T}_n - d''(\theta_0) &\to \eta''(\theta_0)\frac{d'(\theta_0)}{\eta'(\theta_0)} - d''(\theta_0) = -I(\theta_0) < 0; \\
n^{-1}\ell_n''(\alpha_n) = \eta''(\alpha_n)\overline{T}_n - d''(\alpha_n) &\to \eta''(\theta_0)\frac{d'(\theta_0)}{\eta'(\theta_0)} - d''(\theta_0) = -I(\theta_0); \\
n^{-1}\ell_n'''(\alpha_n^*) = \eta'''(\alpha_n^*)\overline{T}_n - d'''(\alpha_n^*) &\to -\eta'''(\theta_0)\frac{d'(\theta_0)}{\eta'(\theta_0)} - d'''(\theta_0).
\end{aligned}$$

On en déduit, avec l'aide du théorème de Slutsky

$$\begin{aligned}
-\frac{n^{-1/2}\ell_n'(\theta_0)}{n^{-1}\ell_n''(\alpha_n)} &\to \frac{N(0, I(\theta))}{I(\theta)} = N\left(0, \frac{1}{I(\theta)}\right); \\
1 - \frac{n^{-1}\ell_n''(\theta_0)}{n^{-1}\ell_n''(\alpha_n)} &\to 1 - \frac{-I(\theta_0)}{-I(\theta_0)} = 0; \\
\frac{1}{2}(\alpha_n - \theta_0)\frac{n^{-1}\ell_n'''(\alpha_n^*)}{n^{-1}\ell_n''(\alpha_n)} &\to \frac{1}{2} \cdot 0 \cdot \frac{-\eta'''(\theta_0)\frac{d'(\theta_0)}{\eta'(\theta_0)} - d'''(\theta_0)}{-I(\theta)} = 0; \\
\sqrt{n}(\alpha_n - \theta_0)\left(1 - \frac{n^{-1}\ell_n''(\theta_0)}{n^{-1}\ell_n''(\alpha_n)} - \frac{1}{2}(\alpha_n - \theta_0)\frac{n^{-1}\ell_n'''(\alpha_n^*)}{n^{-1}\ell_n''(\alpha_n)}\right) &\to V \cdot (0 - 0) = 0.
\end{aligned}$$

et encore une fois par Slutsky $\sqrt{n}(\beta_n - \theta_0) \xrightarrow{d} N(0, 1/I(\theta_0))$. Ainsi, β_n a la même loi asymptotique que l'estimateur du maximum de vraisemblance (donc quasiment optimale), même si on a utilisé une seule itération de Newton–Raphson !

Exercice 60 (exercice 40)**.** Pour chacun des scénarios suivants, trouver les hypothèses à tester ainsi que les deux types d'erreurs qu'on peut commettre. Sur la base de ces informations, décider quelle hypothèse devrait être l'hypothèse nulle H_0 et laquelle devrait être l'alternative H_1.

(i) Une physicienne travaille sur une expérience dont le but est de détecter des particules de matières noires. Elle aimerait tester si ses données indiquent la présence de matière noire.

(ii) Un fêtard voudrait savoir s'il est en mesure de conduire après un apéro. Il aimerait donc tester si le taux d'alcool dans son sang est supérieur à celui autorisé par la loi.

(iii) Barack Obama et Mitt Romney étaient les deux candidats principaux à l'élection présidentielle de 2012 aux États-Unis. Le directeur de campagne de M. Obama aimerait savoir si M. Obama est en tête dans l'état d'Iowa afin de décider s'il doit allouer ou non plus de ressources financières pour la campagne dans cet état. Il faut donc tester si M. Obama est en tête dans l'état d'Iowa. De quelle façon le test changerait-il si on était à la place du directeur de campagne de M. Romney ?

(iv) Un scientifique travaillant pour une compagnie pharmaceutique a pu développer un nouveau médicament afin de réduire la pression artérielle trop élevée. Il voudrait tester si le médicament produit l'effet attendu.

Solution 60. À noter que le choix des hypothèses nulles dans cette exercice est quelque peu subjectif. Les choix ci-dessous reflètent cependant ce qui est habituellement fait en pratique dans les domaines considérés.

(i) Les expériences concernant la matière noire sont habituellement des expériences de décompte modélisés par des lois de Poisson. Soient μ le nombre moyen de particules dénombrées pendant l'expérience, b le nombre moyen de particules dénombrées pendant l'expérience lorsqu'il n'y a pas présence de matières noires et s le nombre moyen de particules de matières noires dénombrées lors de l'expérience. Les deux hypothèses sont $\mu = b$, i.e. qu'il n'y pas d'indication de la présence de matière noire et $\mu = b + s$, c'est-à-dire qu'il y a une indication de matière noire. On fait une fausse découverte si on affirme qu'il y a présence de matière noire lorsqu'en fait il n'y en a pas. On peut «rater une découverte» si on affirme qu'il n'y a pas d'indication de la présence de matière noire lorsqu'en fait il y en a une. Faire une fausse découverte est considéré comme une très grave erreur (cf. l'affaire des «faster-than-light neutrinos» au CERN, qui a provoqué la démission du chairman de l'expérience OPERA [1]). On teste donc :

$$H_0 : \mu = b$$
$$H_1 : \mu = b + s$$

(ii) Soient μ le vrai taux d'alcool dans le sang et μ_0 la limite légale. Les hypothèses à tester sont $\mu \le \mu_0$, c'est-à-dire qu'on peut conduire en toute légalité, et $\mu > \mu_0$, c'est-à-dire qu'on n'est pas autorisé à conduire. Si on pense que $\mu > \mu_0$ lorsqu'en fait $\mu \le \mu_0$, on peut décider inutilement de ne pas conduire et de rentrer chez soi en transport en commun/taxi/à pied. Si on pense que $\mu \le \mu_0$ lorsqu'en fait $\mu > \mu_0$, on va conduire sous l'influence d'alcool et ainsi risquer d'avoir une amende ; ou pire encore, de provoquer un accident. Il est clair que la dernière erreur peut avoir des conséquences beaucoup plus sérieuses que la première, ainsi on devrait tester :

$$H_0 : \mu > \mu_0$$
$$H_1 : \mu \le \mu_0$$

(iii) Soit O le nombre d'habitants d'Iowa qui ont l'intention de voter pour M. Obama et soit R le nombre d'habitants d'Iowa qui ont l'intention de voter pour M. Romney. Les deux

1. http://www.lescienze.it/news/2012/03/30/news/opera_ereditatos_point_of_view-938232/

hypothèses sont $O > R$, c'est-à-dire que M. Obama est en tête, et $O \leq R$, c'est-à-dire que M. Obama est en train de perdre (ou qu'il y a égalité). Si le directeur de campagne de M. Obama pense que M. Obama est en train de perdre lorsqu'en fait $O > R$, il décidera de dépenser inutilement plus d'argent dans l'Iowa. S'il pense que M. Obama est en tête alors qu'en fait $O \leq R$, il décidera de ne pas dépenser d'argent supplémentaire dans l'Iowa, ce qui peut avoir pour conséquence la défaite de M. Obama dans cet état. Cette dernière erreur est certainement la plus grave, on devrait donc tester :

$$H_0 : O \leq R$$
$$H_1 : O > R$$

Pour le directeur de campagne de M. Romney, les hypothèses seront inversées :

$$H_0 : R \leq O$$
$$H_1 : R > O$$

(iv) Afin de vérifier l'efficacité du médicament, on devra faire une étude clinique avec des patients souffrant de pression artérielle élevée (il ne sera certainement pas difficile d'en trouver, puisqu'on estime que plus que 20% de la population a une pression artérielle élevée). Dans cette étude, il y aura un groupe appelé «traitement» à qui on administrera le nouveau médicament et un groupe appelé «contrôle» à qui on administrera un placebo. Soit p_T la moyenne des pressions artérielles du groupe traitement et soit p_C la moyenne des pressions artérielles du groupe contrôle. Les deux hypothèses sont $p_T = p_C$, c'est-à-dire que le médicament ne fonctionne pas, et $p_T < p_C$, i.e. que le médicament réduit la pression artérielle. Lorsque $p_T = p_C$ on pourra déclarer, à tort, que le médicament fonctionne et lorsque $p_T < p_C$, on pourra penser à tort que le médicament n'est pas efficace. Dans le premier cas, un médicament inefficace pourrait se retrouver sur le marché, entrainant potentiellement d'important effets secondaires tandis que dans le deuxième cas, le développement d'un médicament efficace pourrait être arrêté. Puisque nous voulons être certains que les médicaments que nous utilisons sont efficaces à traiter les maladies, nous devrions choisir :

$$H_0 : p_T = p_C$$
$$H_1 : p_T < p_C$$

Exercice 61 (**tests d'hypothèses intuitifs**, exercice 48)**.** Le but de cet exercice est de donner une motivation intuitive aux tests d'hypothèses. Soient $X_1, \ldots, X_n$ iid avec la fonction de densité

$$f_X(x) = \frac{1}{48}\lambda^5 x^{3/2} e^{-\lambda\sqrt{x}}, \qquad x > 0,$$

où $\lambda > 0$ est un paramètre. On aimerait tester l'hypothèse $H_0 : \lambda = \lambda_0$ vs. $H_1 : \lambda = \lambda_1$, où $\lambda_0 > \lambda_1$.

(i) Trouver l'estimateur du maximum de vraisemblance $\widehat{\lambda}_n$.

(ii) Comme expliqué au chapitre 3 du livre, $\widehat{\lambda}_n$ est un bon estimateur. Ainsi, il est en un certain sens naturel de rejeter H_0 si λ_0 n'est pas « compatible »avec $\widehat{\lambda}_n$. Dans notre cas, cela voudrait dire : rejeter H_0 lorsque $\widehat{\lambda}_n$ est petit. (Si $\widehat{\lambda}_n > \lambda_0$, on préferera certainement H_0 et non H_1.) Quelle forme prendra donc la fonction de test ? Donner-la à une constante D près.

(iii) Maintenant, il faut trouver la fonction de test précise. Pour cela, il faudrait choisir une borne en dessous de laquelle on juge $\widehat{\lambda}_n$ suffisamment petit pour rejeter H_0. Pour un seuil $\alpha \in]0,1[$ donné, on voudrait que la probabilité de commettre une erreur de type I soit α. À partir de là, décrire la relation entre α et D.

(iv) Nous voilà un test au niveau α. On peut ensuite se demander s'il est le meilleur test. Aurons-nous pu faire mieux, c'est-à-dire trouver un test au niveau α mais plus puissant ? Montrer que la réponse est négative, en montrant que notre fonction de test est exactement la même que celle décrite par le lemme de Neyman–Pearson. (On peut supposer que la valeur Q du lemme existe ; ce résultat sera démontrée ultérieurement.)

(v) Trouver une formule, la plus simple possible, pour la fonction de test $\delta(X_1, \dots, X_n)$. *Indice :* $\widehat{\lambda}_n$ contient une somme dont chaque élément suit une distribution qu'on a déjà vu.

Solution 61. (i) On procède comme d'habitude :

$$L_n(\lambda) = (48)^{-n}\lambda^{5n} \exp\left(-\lambda \sum_{i=1}^n \sqrt{X_i}\right)\left(\prod_{i=1}^n X_i\right)^{3/2}$$

$$\ell_n(\lambda) = 5n \log \lambda - \lambda \sum_{i=1}^n \sqrt{X_i} + \frac{3}{2}\sum_{i=1}^n \log X_i - n \log 48$$

$$\ell_n'(\lambda) = \frac{5n}{\lambda} - \sum_{i=1}^n \sqrt{X_i}, \qquad \ell_n''(\lambda) = \frac{-5n}{\lambda^2} < 0,$$

d'où on trouve aisément

$$\widehat{\lambda}_n = \frac{5n}{\sum_{i=1}^n \sqrt{X_i}}.$$

(ii) Par définition, on rejette H_0 si et seulement si la fonction de test $\delta(X_1, \dots, X_n)$ est égale à 1. L'énoncé suggère qu'on la rejette si et seulement si $\widehat{\lambda}_n$ est inférieur à un seuil quelconque, D. Ainsi, la fonction de test est de la forme $\delta(X_1, \dots, X_n) = \mathbf{1}\{\widehat{\lambda}_n \leq D\}$.

(iii) La probabilité de commettre une erreur de type I est la probabilité de rejeter H_0 lorsqu'elle est vraie. Ainsi, on obtient l'équation suivante :

$$\alpha = \mathbb{P}_{\lambda_0}(\delta(X_1, \dots, X_n) = 1) = \mathbb{P}_{\lambda_0}(\widehat{\lambda}_n \leq D),$$

où les probabilités (qui dépendent bien sûr de λ !) sont calculées pour $\lambda = \lambda_0$. Si G_n est la fonction de distribution de la variable aléatoire $\widehat{\lambda}_n$ quand $\lambda = \lambda_0$, alors la solution est le α-quantile de G_n : $D = G_n^{-}(\alpha) = G_n^{-1}(\alpha)$ car G_n est strictement croissante et continue.

(iv) Le rapport de vraisemblance est

$$\Lambda_n(X_1, \dots, X_n) = \frac{L_n(\lambda_1)}{L_n(\lambda_0)} = \left(\frac{\lambda_1}{\lambda_0}\right)^{5n} \exp\left[(\lambda_0 - \lambda_1)\sum_{i=1}^n \sqrt{X_i}\right].$$

Puisque $\lambda_0 > \lambda_1$, on voit que

$$\mathbf{1}\{\Lambda_n \geq Q\} = \mathbf{1}\left\{\sum_{i=1}^n \sqrt{X_i} \geq \frac{\log\left[Q\left(\frac{\lambda_0}{\lambda_1}\right)^{5n}\right]}{\lambda_0 - \lambda_1}\right\} = \mathbf{1}\left\{\widehat{\lambda}_n \leq \frac{5n(\lambda_0 - \lambda_1)}{\log\left[Q\left(\frac{\lambda_0}{\lambda_1}\right)^{5n}\right]}\right\}.$$

Ce qui est important ici n'est pas ces expressions atroces, mais le fait que la fonction de test du rapport de vraisemblance est elle aussi de la forme $\mathbf{1}\{\widehat{\lambda}_n \leq D'\}$, qui est exactement la même forme de δ. Par le lemme de Neyman–Pearson, Q (et donc D') est tel que

$$\alpha = \mathbb{P}_{\lambda_0}(\Lambda_n \geq Q) = \mathbb{P}_{\lambda_0}(\widehat{\lambda}_n \leq D').$$

Il s'en suit que $D' = G_n^{-1}(\alpha) = D$ et donc $\mathbf{1}\{\Lambda_n \geq Q\} = \delta$. Ainsi, notre test intuitif est optimal !

(v) À l'expression de $\widehat{\lambda}_n$ ainsi qu'à celle de Λ_n, le seul élément aléatoire est $\sum \sqrt{X_i}$. Essayons donc de trouver la fonction distribution de $Y = \sqrt{X_1}$. C'est une tranformation de X_1 dont l'inverse est $X_1 = Y^2$. Ainsi

$$f_Y(y) = f_X(y^2)2y = \frac{1}{48}\lambda^5 y^3 e^{-\lambda y} 2y = \frac{1}{24}\lambda^5 y^4 e^{-\lambda y}.$$

Même si on ne se souvient pas que $\Gamma(5) = (5-1)! = 24$, on reconnaît ici la loi Gamma$(5, \lambda)$. (D'ici, on peut déduire que soit $\Gamma(5) = 24$, soit ceux qui ont écrit cet exercice se sont trompés.) Il s'en suit que $\sum_{i=1}^n \sqrt{X_i} \sim$ Gamma$(5n, \lambda)$ (on peut voir cela en utilisant la fonction génératrice des moments). Sous H_0, $\lambda = \lambda_0$. À partir de là on peut trouver les valeurs de D et Q, mais on peut se simplifier la vie en remarquant que la fonction de test est également de la forme $\mathbf{1}\{\sum \sqrt{X_i} \geq D''\}$. Pour que $\mathbb{P}_{\lambda_0}(\sum \sqrt{X_i} \geq D'') = \alpha$, il faudrait que D'' soit le $(1-\alpha)$-quantile de la distribution de $\sum \sqrt{X_i}$ sous H_0, à savoir Gamma$(5n, \lambda_0)$. Ainsi, la fonction de test optimal au seuil α est

$$\mathbf{1}\left\{\sum_{i=1}^n \sqrt{X_i} \geq \text{Gamma}_{5n,\lambda_0,1-\alpha}\right\}.$$

Le message à retenir ici est que ce qui est important est la fonction de test, et non pas sa représentation. Par exemple, si on observe $X_1, \ldots, X_{10}$ et on veut tester $H_0 : \lambda = 1$ vs. $H_1 : \lambda = 0.5$, il est plus simple d'utiliser $\mathbf{1}\{\sum \sqrt{X_i} \geq 62.17\}$ que d'utiliser $\mathbf{1}\{50/\sum \sqrt{X_i} \leq 0.804\}$!

Exercice 62 (exercice 41). Soit $X_1, \ldots, X_n$ un échantillon iid provenant d'une distribution $N(\mu, 1)$. On va tester l'hypothèse nulle $H_0 : \mu = 0$ vs. l'hypothèse alternative $H_1 : \mu \neq 0$ en utilisant la statistique de test

$$T_n(X_1, \ldots, X_n) = \bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i,$$

et la fonction de test

$$\delta(X_1, \ldots, X_n) = \begin{cases} 1, & \text{si } |T_n(X_1, \ldots, X_n)| \geq Q, \\ 0, & \text{sinon,} \end{cases}$$

où $Q > 0$.

(i) Trouver la probabilité de commettre une erreur de type I.

(ii) Trouver la probabilité de commettre une erreur de type II.

(iii) Comment se comportent ces deux probabilités lorsqu'on augmente la valeur de Q ?

(iv) Trouver la plus petite valeur de Q pour laquelle le seuil de signification du test est $\alpha \in]0,1[$. Quelle est cette valeur lorsque $\alpha = 0.05$ et $n = 10$? Trouver le supremum de la probabilité de commettre une erreur de type II pour cette valeur de Q.

Solution 62.

(i) En utilisant la proposition 2.7 (p. 51), on trouve que sous H_0 la statistique de test $T_n = T_n(X_1, \ldots, X_n)$ suit une loi $N(0, 1/n)$. Ainsi, $\sqrt{n}T_n \sim N(0,1)$ et la probabilité de commettre une erreur de type I est

$$\begin{aligned}\mathbb{P}_0(\delta = 1) &= \mathbb{P}_0(|T_n| \geq Q) = \mathbb{P}_0(T_n \leq -Q) + \mathbb{P}_0(T_n \geq Q)\\ &= \mathbb{P}_0(\sqrt{n}T_n \leq -\sqrt{n}Q) + \mathbb{P}_0(\sqrt{n}T_n \geq \sqrt{n}Q) = 2\Phi(-\sqrt{n}Q),\end{aligned}$$

où $\mathbb{P}_0$ est la probabilité sous H_0 et Φ est la fonction de répartition de $N(0,1)$, et on a utilisé $\Phi(-z) = 1 - \Phi(z)$. Le graphique 1 donne cette probabilité en fonction de Q pour $n = 10$.

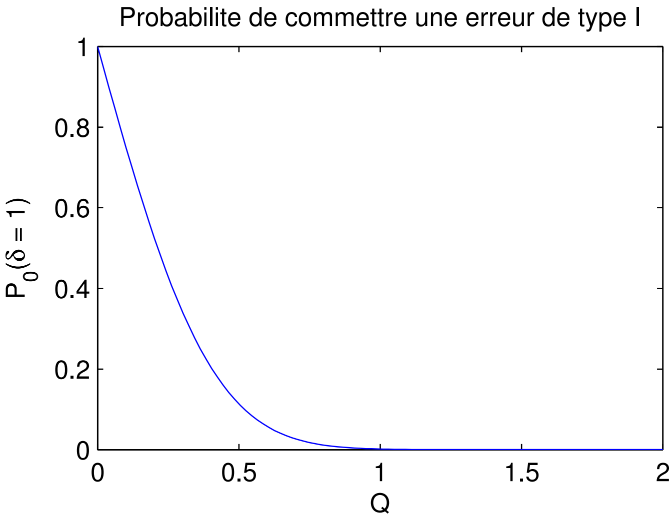


FIGURE 1 – La probabilité de commettre une erreur de type I en fonction de Q pour $n = 10$.

(ii) En utilisant la même proposition 2.7, on trouve que sous H_1 la statistique de test T_n suit la loi $N(\mu, 1/n)$, où $\mu \neq 0$. Il s'en suit que $\sqrt{n}(T_n - \mu) \sim N(0,1)$ et la probabilité de commettre une erreur de type II est

$$\begin{aligned}g(\mu) &= \mathbb{P}_\mu(\delta = 0) = \mathbb{P}_\mu(|T_n| < Q) = \mathbb{P}_\mu(-Q < T_n < Q)\\ &= \mathbb{P}_\mu(\sqrt{n}(-Q - \mu) < \sqrt{n}(T_n - \mu) < \sqrt{n}(Q - \mu))\\ &= \Phi(\sqrt{n}(Q - \mu)) - \Phi(\sqrt{n}(-Q - \mu))\end{aligned}$$

avec $\mu \neq 0$.

(iii) On remarque que Φ est continue, strictement croissante et tend vers 0 lorsque $z \to -\infty$, vers 1 lorsque $z \to \infty$.

On en déduit que, en fonction de Q, la probabilité de commettre une erreur de type I est une fonction strictement décroissante tandis que la probabilité de commettre une erreur de type II est une fonction strictement croissante. Cela veut dire qu'en réduisant l'erreur de type I, on va forcément augmenter l'erreur de type II. Par ailleurs ces probabilités convergent vers 0 et 1 lorsque $Q \to \infty$.

(iv) Comme la probabilité de commettre une erreur de type I est une fonction continue et strictement décroissante, la valeur de Q demandée est la solution de l'équation

$$\alpha = \mathbb{P}_0(\delta = 1) = 2\Phi(-\sqrt{n}Q).$$

La solution est $Q = -\frac{1}{\sqrt{n}}\Phi^{-1}(\frac{\alpha}{2}) = -\frac{1}{\sqrt{n}}z_{\alpha/2} = \frac{1}{\sqrt{n}}z_{1-\alpha/2}$, où $z_\beta = \Phi^{-1}(\beta)$ est le β-quantile de $N(0,1)$. Donc, pour $\alpha = 0.05$ et $n = 10$, on trouve

$$Q = \frac{1}{\sqrt{10}}z_{0.975} \approx 0.62.$$

Cela veut dire que l'on rejette H_0 au niveau 0.05 si $|T| \geq 0.62$.
Le dérivée de $g(\mu)$ (par rapport à μ) est

$$g'(\mu) = -\sqrt{n}\Phi'(\sqrt{n}(Q-\mu)) + \sqrt{n}\Phi'(\sqrt{n}(-Q-\mu)) = -\sqrt{n}\phi(\sqrt{n}(Q-\mu)) + \sqrt{n}\phi(\sqrt{n}(-Q-\mu)),$$

où ϕ est la fonction de densité de $N(0,1)$. En mettant la dérivée egale à zero, on trouve

$$\begin{aligned}
&\phi(\sqrt{n}(Q-\mu)) = \phi(\sqrt{n}(-Q-\mu)) \\
\Longleftrightarrow\ &\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{n}{2}(Q-\mu)^2\right) = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{n}{2}(-Q-\mu)^2\right) \\
\Longleftrightarrow\ &(Q-\mu)^2 = (-Q-\mu)^2 \\
\Longleftrightarrow\ &\mu = 0.
\end{aligned}$$

Il est aisé de voir que ceci correspond à un maximum. Ainsi, $\sup_{\mu \neq 0} g(\mu) = g(0) = \Phi(\sqrt{n}Q) - \Phi(-\sqrt{n}Q) = 1 - 2\Phi(-\sqrt{n}Q) = 1 - \alpha = 0.95$. La fonction $g(\mu)$ avec $Q = 0.62$ et $n = 10$ est illustrée dans le figure 2.

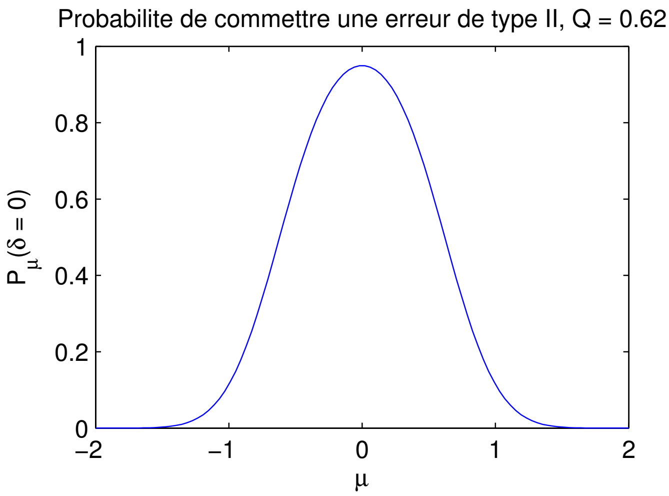


Figure 2 – La probabilité de commettre une erreur de type II en fonction de μ avec $Q = 0.62$.

Exercice 63 (exercice 45)**.** Soient $X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} N(\mu, \sigma^2)$ avec $\sigma^2 > 0$ connue. Trouver le test le plus puissant pour tester $H_0 : \mu = \mu_0$ vs. $H_1 : \mu = \mu_1$ avec $\mu_0 < \mu_1$ à un seuil de signification $\alpha \in (0, 1)$.

Solution 63. La loi gaussienne avec une variance connue fait partie d'une famille exponentielle à 1-paramètre avec $\eta(\mu) = \mu/\sigma^2$ et $T(x) = x$. Puisque η est croissante, on peut utiliser l'exemple 4.14 (p. 108) pour déduire que la fonction de test pour le test le plus puissant est

$$\delta = \mathbf{1}\{\tau_n > q_{1-\alpha}\},$$

avec $\tau_n = \sum_{i=1}^n X_i$ et $q_{1-\alpha}$ le $(1-\alpha)$-quantile de τ_n sous H_0. Lorsque $\mu = \mu_0$, nous avons que $\tau_n \sim \mathcal{N}(n\mu_0, n\sigma^2)$, ce qui implique

$$Z = \frac{\tau_n - n\mu_0}{\sqrt{n\sigma^2}} \sim N(0, 1).$$

Puisque τ_n est une variable aléatoire continue, nous pouvons calculer $q_{1-\alpha}$ à partir de

$$\begin{aligned} 1 - \alpha &= \mathbb{P}_{\mu_0}\left(\tau_n \leq q_{1-\alpha}\right) \\ &= \mathbb{P}_{\mu_0}\left(\frac{\tau_n - n\mu_0}{\sqrt{n\sigma^2}} \leq \frac{q_{1-\alpha} - n\mu_0}{\sqrt{n\sigma^2}}\right) \\ &= \mathbb{P}_{\mu_0}\left(Z \leq \frac{q_{1-\alpha} - n\mu_0}{\sqrt{n\sigma^2}}\right) = \Phi\left(\frac{q_{1-\alpha} - n\mu_0}{\sqrt{n\sigma^2}}\right). \end{aligned}$$

Nous obtenons alors

$$q_{1-\alpha} = \sqrt{n\sigma^2}\Phi^{-1}(1-\alpha) + n\mu_0 = \sqrt{n\sigma^2} z_{1-\alpha} + n\mu_0,$$

où $z_{1-\alpha}$ est le $(1-\alpha)$-quantile d'une loi $N(0, 1)$. La fonction de test est donc donnée par

$$\delta = \mathbf{1}\{\tau_n > q_{1-\alpha}\} = \mathbf{1}\left\{\frac{\tau_n - n\mu_0}{\sqrt{n\sigma^2}} > z_{1-\alpha}\right\} = \mathbf{1}\left\{\frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} > z_{1-\alpha}\right\}.$$

Exercice 64 (exercice 46)**.** Pour un échantillon $X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \text{Bernoulli}(p)$, on veut tester

$$H_0 : \ p = 0.49 \quad \text{vs} \quad H_1 : \ p = 0.51.$$

Déterminez approximativement la taille de l'échantillon pour laquelle la probabilité de commettre une erreur de type I et la probabilité de commettre une erreur de type II sont approximativement égales à 0.01. Utilisez une fonction de test qui rejette H_0 si $\sum_i X_i$ est grande. *Indice : Utilisez le théorème centrale limite pour approximer la distribution de* $n^{-1}\sum_{i=1}^n X_i$ *par une loi normale. Vous avez aussi besoin du fait que* $z_{0.99} \approx 2.33$*, où* $z_{0.99}$ *est le 0.99-quantile de la loi* $N(0, 1)$.

Solution 64. Nous utilisons la statistique de test $\tau_n = \sum_{i=1}^n X_i \sim Bin(n, p)$. Nous allons utiliser l'approximation normale de la loi binomiale, i.e. que nous approximons la distribution de $Z = \frac{\tau_n - np}{\sqrt{np(1-p)}}$ par une loi $\mathcal{N}(0, 1)$. Nous voulons n et Q tel que

$$\begin{aligned} \mathbb{P}_{p_0}(\tau_n > Q) &= \alpha \\ \mathbb{P}_{p_1}(\tau_n > Q) &= 1 - \alpha, \end{aligned}$$

où $\alpha = 0.01$. Les deux dernières équations sont équivalentes à

$$\mathbb{P}\left(Z > \frac{Q - n \cdot 0.49}{\sqrt{n \cdot 0.49 \cdot 0.51}}\right) = 0.01$$
$$\mathbb{P}\left(Z > \frac{Q - n \cdot 0.51}{\sqrt{n \cdot 0.51 \cdot 0.49}}\right) = 0.99,$$

i.e. que nous devons résoudre

$$\frac{Q - n \cdot 0.49}{\sqrt{n \cdot 0.49 \cdot 0.51}} = 2.33$$
$$\frac{Q - n \cdot 0.51}{\sqrt{n \cdot 0.51 \cdot 0.49}} = -2.33,$$

ce qui nous donne $n = 13567$ et $Q = 6783.5$.

Exercice 65 (exercice 47)**.** Soient $X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} Unif(0, \theta)$ et considérez $H_0 : \theta = \theta_0$ et $H_1 : \theta = \theta_1$ avec $\theta_1 < \theta_0$.

(i) Trouvez le test le plus puissant de H_0 vs. H_1 à un seuil de signification $\alpha = (\theta_1/\theta_0)^n$. Considérez le comportement de ce seuil, comme fonction de θ_0, θ_1 et n. Quelle est la puissance de ce test ? Est-ce qu'on peut définir un test optimal de type Neyman–Pearson pour d'autres valeurs de α ?

(ii) Considérez un test (pas nécessairement optimal) de seuil de signification $\alpha < (\theta_1/\theta_0)^n$ qui rejette H_0 quand $X_{(n)} < k$. Trouvez la valeur appropriée de k. Quelle est la puissance de ce test ?

Solution 65.

(i) La vraisemblance est

$$L_n(\theta; X_1, \ldots, X_n) = \prod_{i=1}^{n} \frac{1}{\theta} \mathbf{1}\{0 \leq X_i \leq \theta\} = \frac{1}{\theta^n} \mathbf{1}\{X_{(n)} \leq \theta\}.$$

Grâce au lemme de Neyman-Pearson (lemme 4.11, p. 106), nous savons que la statistique de test optimale pour un seuil α est

$$\Lambda_n(\mathbf{X}) = \frac{L_n(\theta_1)}{L_n(\theta_0)} = \left(\frac{\theta_0}{\theta_1}\right)^n \mathbf{1}\{X_{(n)} \leq \theta_1\} = \begin{cases} \left(\frac{\theta_0}{\theta_1}\right)^n & X_{(n)} \leq \theta_1 \\ 0 & X_{(n)} > \theta_1, \end{cases}$$

lorsqu'il existe une valeur $Q > 0$ satisfaisant $\mathbb{P}_{\theta_0}(\Lambda_n \geq Q) = \alpha$. Lorsque $X_{(n)} \leq \theta_1$, $\Lambda_n(\mathbf{X}) = \left(\frac{\theta_0}{\theta_1}\right)^n$, sinon $\Lambda_n(\mathbf{X}) = 0$. Ainsi, pour chaque $Q \in]0, (\theta_0/\theta_1)^n]$ (par exemple $Q = 1$) nous avons $\Lambda_n \geq Q$ si et seulement si $X_{(n)} \leq \theta_1$. Rejeter H_0 lorsque $\Lambda_n \geq Q$ est donc équivalent à la rejeter lorsque $X_{(n)} \leq \theta_1$, et la fonction de test devient donc $\delta = \mathbf{1}\{X_{(n)} \leq \theta_1\}$. La probabilité de commettre une erreur de type I est alors

$$\mathbb{P}_{\theta_0}(\delta = 1) = \mathbb{P}_{\theta_0}(X_{(n)} \leq \theta_1) = \left(\frac{\theta_1}{\theta_0}\right)^n = \alpha.$$

C'est exactement le seuil demandé, ainsi nous avons bien défini un test le plus puissant au seuil $\alpha = (\theta_1/\theta_0)^n$. Ce seuil est croissant en tant que fonction de θ_1 et décroissant en tant que fonction de θ_0 et de n. La puissance est $\mathbb{P}_{\theta_1}(\delta = 1) = \mathbb{P}_{\theta_1}(X_{(n)} \leq \theta_1) = 1$. De plus, il n'est pas possible d'utiliser le lemme de Neyman–Pearson afin de créer des tests PP pour d'autres valeurs de α.

(ii) Nous cherchons la valeur de k telle que

$$\alpha = \mathbb{P}_{\theta_0}(X_{(n)} \leq k) = \left(\frac{k}{\theta_0}\right)^n,$$

ce qui donne $k = \theta_0 \alpha^{1/n} < \theta_1$.
La puissance de ce test est

$$\mathbb{P}_{\theta_1}(X_{(n)} \leq \theta_0 \alpha^{1/n}) = \alpha \left(\frac{\theta_0}{\theta_1}\right)^n < 1.$$

Il est possible de montrer que ce test est en fait un test PP pour $\alpha < (\theta_1/\theta_0)^n$.
Remarque : il est naturel de baser le test sur $X_{(n)}$, puisque c'est une statistique exhaustive pour θ.

Exercice 66 (exercice 49)**.** Un laboratoire de traitement d'images a développé une nouvelle méthode pour scanner le cerveau. Le laboratoire prétend qu'ils sont capables de scanner le cerveau en moins de 20 minutes. Voici un échantillon de temps de 12 scans de cerveau :

$$\mathbb{X} = \{21, 18, 19, 16, 18, 24, 22, 19, 24, 26, 18, 21\}.$$

(i) Supposons que la durée de scan suit $\mathcal{N}(\mu, 3^2)$. Testez si la durée moyenne de scan est moins de 20 minutes, i.e., testez $H_0 : \ \mu \leq \mu_0$ vs $H_1 : \ \mu > \mu_0$ avec $\mu_0 = 20$ à un seuil de signification $\alpha = 0.05$.

(ii) Pourriez-vous faire la même analyse sachant que la variance de la loi normale est inconnue ? *Indice : Utilisez* $\delta = \mathbf{1}\left(\frac{\sqrt{n}(\bar{X}-\mu_0)}{S} \geq t_{n-1,1-\alpha}\right)$ *comme fonction de test. Ici* $t_{n-1,1-\alpha}$ *est le* $1-\alpha$ *quantile de la loi Student avec* $n-1$ *degrés de liberté.*

Solution 66.

(i) En utilisant le théorème 4.16 (p. 112), nous rejetons H_0 lorsque $\tau = \sum_i X_i$ est grand, i.e. lorsque $\tau \geq q_{1-\alpha}$, où $q_{1-\alpha}$ est le $(1-\alpha)$-quantile de la distribution de τ avec $\mu = \mu_0 = 20$. En utilisant le même raisonnement que dans l'exercice 1, on arrive à la fonction de test

$$\delta = \mathbf{1}\{\tau \geq q_{1-\alpha}\} = \mathbf{1}\left\{\frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \geq z_{1-\alpha}\right\}.$$

La fonction précédente évaluée aux valeurs données dans l'énoncé nous donne

$$\delta = \mathbf{1}\{0.577 \geq 1.645\} = 0.$$

Il n'y a donc pas d'évidence, à un seuil de signification de 5%, nous permettant de rejeter l'affirmation de la compagnie.

(ii) Pour $n = 12$ et $\alpha = 0.05$, $t_{n-1,1-\alpha} = 1.796$. Nous rejetons donc si

$$\tau = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} \geq 1.796,$$

où $S^2 = \sum_{i=1}^{n}(X_i - \bar{X})^2/(n-1)$. Dans notre cas, la valeur de τ est

$$\tau = \frac{20.5 - 20}{\frac{3.03}{\sqrt{12}}} = 0.572,$$

nous ne pouvons donc pas rejeter H_0.

Exercice 67 (exercice 50)**.** Soient $Y_1, \ldots, Y_4$ des variables aleatoires independantes et identiquement distribues selon une loi normale $\mathcal{N}(\mu, 4^2)$. On veut montrer que μ est plus grand que $\mu_0 = 10$. Par conséquent, on effectue un test au niveau $\alpha = 5\%$ de l'hypothèse nulle $H_0 : \mu \leq 10$.

(i) Calculez la puissance du test pour des vraies valeurs de μ égales à 13 et 11.

(ii) Si la vraie valeur de μ est égale à 13, quelle chance a-t-on de la détecter ?

(iii) Pour augmenter la chance de détection, déterminez le nombre d'observations nécessaires pour obtenir une puissance de 90% dans le cas $\mu = 13$.

Solution 67.

(i) En utilisant le même test que dans l'exercice précédent, on trouve que la puissance vaut :

$$\beta(\mu) = \mathbb{P}_\mu\left(\frac{\bar{Y} - \mu_0}{\sigma/\sqrt{n}} \geq z_{1-\alpha}\right) = \mathbb{P}_\mu\left(\frac{\bar{Y} - \mu}{\sigma/\sqrt{n}} \geq z_{0.95} + \frac{\mu_0 - \mu}{\sigma/\sqrt{n}}\right)$$
$$= 1 - \Phi\left(z_{0.95} + \frac{\mu_0 - \mu}{\sigma/\sqrt{n}}\right), \quad \mu > \mu_0.$$

Ce qui nous donne les valeurs suivantes :

μ	13	11
$\beta(\mu)$	0.44	0.12

(ii) On cherche n tel que $1 - \Phi\left(z_{0.95} + \frac{\mu_0 - \mu}{\sigma/\sqrt{n}}\right) = 0.9$; c'est-à-dire :

$$n = \frac{16}{9}\left(z_{0.95} - z_{0.10}\right)^2$$

Dans notre cas, $n = 15.22$. Il faut donc 16 observations.

Exercice 68 (exercice 51, **test apparié**)**.** Une compagnie pharmaceutique veut vérifier si son nouveau produit amaigrissant ABALGRA est efficace. Pour ce faire, le poids (en kilo) de 10 hommes choisis de façon aléatoire a été recueilli juste avant la première prise du médicament ainsi qu'à la fin du traitement, 7 semaines plus tard. Soit X_i le poids du i^{e} homme avant le traitement et soit Y_i son poids à la fin du traitement. On peut donc supposer que X_i sont iid, puisque les différentes personnes ont été choisies au hasard. De même pour Y_i, car chaque personne a reçu le même traitement. Soient $\mu_1 = \mathbb{E}X_i$ et $\mu_2 = \mathbb{E}Y_i$.
On s'intéresse donc aux différences $d_i = Y_i - X_i$. Celles-ci sont indépendantes et on suppose qu'elles suivent une loi normale $\mathcal{N}(\mu_2 - \mu_1, 5)$. Tester à l'aide des données du tableau ci-dessous si le médicament semble entraîner une perte de poids au seuil $\alpha = 0.05$.

i	1	2	3	4	5	6	7	8	9	10
X_i	55.5	75	63.8	54.7	62.7	71	68.3	56	74.4	65
Y_i	52.8	73.7	62.7	55	59.3	70.2	67.1	55.4	71.9	65.2

Remarque. Puisque X_1 et Y_1 proviennent de la même personne, il est irréaliste de les supposer indépendantes. Dans ce contexte, on parle d'un *test apparié* (angl. « paired test »).
Bonus. Expliquer le nom ABALGRA.

Solution 68. Nous voulons tester $H_0 : \mu_2 - \mu_1 \geq 0$ contre $H_1 : \mu_2 - \mu_1 < 0$. En posant $\mu = \mu_2 - \mu_1$, nous obtenons un test unilatéral classique avec $\mu_0 = 0$. Nous savons donc par le théorème 10 (p. 96) que le test uniformément le plus puissant est de la forme $\mathbf{1}\{\tau_n(d_1, \ldots, d_n) < q_\alpha\}$, où $\tau_n(d_1, \ldots, d_n) = \sum_{i=1}^n d_i$ et

$$\alpha = \mathbb{P}_{\mu_0}(\tau_n < q_\alpha) = \mathbb{P}\left(\frac{\tau_n/n - \mu_0}{\sigma/\sqrt{n}} < \frac{q_\alpha/n - \mu_0}{\sigma/\sqrt{n}}\right).$$

Le terme de gauche de la deuxième égalité suit une loi $\mathcal{N}(0,1)$ sous H_0. Ainsi

$$\frac{q_\alpha/n - \mu_0}{\sigma/\sqrt{n}} = z_\alpha.$$

Nous obtenons finalement

$$\mathbf{1}\{\tau_n(d_1, \ldots, d_n) < q_\alpha\} = \mathbf{1}\left\{\frac{\overline{d}_n - \mu_0}{\sigma/\sqrt{n}} < z_\alpha\right\}.$$

Dans notre cas, $\overline{d}_n = -1.31$, $\sigma^2 = 5$, $n = 10$, $\alpha = 0.05$ et $z_{0.05} = -1.645$, ce qui donne la fonction de test

$$\mathbf{1}\left\{\frac{-1.31}{\sqrt{5/10}} < -1.645\right\} = \mathbf{1}\{-1.85 < -1.645\} = 1.$$

On rejette donc l'hypothèse nulle.

Exercice 69 (exercice 52, **test de variance pour la loi gaussienne**).

(i) Soit $X_1, \ldots, X_n$ un échantillon iid tiré d'une distribution normale $\mathcal{N}(\mu, \sigma^2)$, où les paramètres μ et σ^2 sont inconnus. Montrer que la fonction de test du test du rapport de vraisemblance pour les hypothèses $H_0 : \sigma^2 = \sigma_0^2$ et $H_1 : \sigma^2 \neq \sigma_0^2$ à un seuil α est de la forme $\mathbf{1}\{W > c_1\} + \mathbf{1}\{W < c_2\}$, où $W = (1/\sigma_0^2)\sum_{i=1}^n (X_i - \bar{X})^2$ et où c_1 et c_2 sont tels que $c_1^{-n} e^{c_1} = c_2^{-n} e^{c_2}$.
Indice : écrire le rapport de vraisemblance comme une fonction de W et étudier la forme de cette fonction.

(ii) En pratique, on choisit c_1 et c_2 tel que $\mathbb{P}_{H_0}(W > c_1) = \mathbb{P}_{H_0}(W < c_2) = \alpha/2$. (Le test obtenu n'est donc pas un test du rapport de vraisemblance.) Trouver les valeurs de c_1 et c_2 lorsque $\alpha = 0.05$, et effectuer ce test pour $\sigma_0^2 = 4$ sur les données suivantes :

$$0.449,\ -3.421,\ -2.841,\ 0.829,\ -0.941,\ 1.789,\ 0.889,\ 1.109,\ 0.969,\ 1.169$$

(Noter que $\bar{X} = 0$.)

Solution 69.

(i) Les hypothèses du test sont $H_0 : (\mu,\sigma^2) \in \Theta_0$ vs. $H_1 : (\mu,\sigma^2) \in \Theta_1$, où $\Theta_0 = \{(\mu,\sigma^2) : \mu \in \mathbb{R}, \sigma^2 = \sigma_0^2\}$ et $\Theta_1 = \{(\mu,\sigma^2) : \mu \in \mathbb{R}, \sigma^2 > 0, \sigma^2 \neq \sigma_0^2\}$. Nous obtenons donc que

$$\sup_{(\mu,\sigma^2)\in\Theta_1} L_n(\mu,\sigma^2) = L_n(\widehat{\mu}_n, \widehat{\sigma}_n^2),$$

où $(\widehat{\mu}_n, \widehat{\sigma}_n^2) = (\overline{X}_n, n^{-1}\sum_{i=1}^n (X_i - \overline{X}_n)^2)$ est l'estimateur du maximum de vraisemblance de (μ,σ^2) et

$$\sup_{(\mu,\sigma^2)\in\Theta_0} L_n(\mu,\sigma^2) = L_n(\widehat{\mu}_n, \sigma_0^2),$$

où $\widehat{\mu}_n = \overline{X}_n$ est l'estimateur du maximum de vraisemblance de μ lorsque la variance est connue. Le rapport de vraisemblance est

$$\begin{aligned}\Lambda_n(X_1,\ldots,X_n) = \frac{L_n(\widehat{\mu}_n,\widehat{\sigma}_n^2)}{L(\widehat{\mu}_n,\sigma_0^2)} &= \left(\frac{n\sigma_0^2}{\sum_{i=1}^n (X_i - \overline{X}_n)^2}\right)^{n/2} \exp\left(\frac{\sum_{i=1}^n (X_i-\overline{X}_n)^2}{2\sigma_0^2} - \frac{n}{2}\right) \\ &= \left(\frac{n}{W}\right)^{n/2} \exp\left(\frac{-n}{2} + \frac{W}{2}\right) = \sqrt{\left(\frac{n}{e}\right)^n W^{-n} e^W},\end{aligned}$$

où $W = (1/\sigma_0^2)\sum_{i=1}^n (X_i - \overline{X}_n)^2 \sim \chi^2_{n-1}$ sous H_0. Nous avons donc

$$\Lambda_n(X_1,\ldots,X_n) > Q \Leftrightarrow \sqrt{\left(\frac{n}{e}\right)^n W^{-n}e^W} > Q \Leftrightarrow W^{-n}e^W > Q',$$

où Q' est tel que $\mathbb{P}_{H_0}(W^{-n}e^W > Q') = \alpha$. Posons $f(w) = w^{-n}e^w$, et analysons cette fonction. Nous avons

$$f'(w) = w^{-n-1}e^w\,(w-n) \qquad \begin{cases} < 0 & 0 < w < n \\ = 0 & w = n \\ > 0 & w > n \end{cases}$$

Nous obtenons donc que $\Lambda_n(X_1,\ldots,X_n) > Q \Leftrightarrow W > c_1$ ou $W < c_2$, où c_1 et c_2 sont telles que $f(c_1) = f(c_2)$ et telles que $\mathbb{P}_{H_0}(W > c_1) + \mathbb{P}_{H_0}(W < c_2) = \alpha$ (voir graphique). Ceci nous donne un système à deux équations deux inconnus compliqué à résoudre.

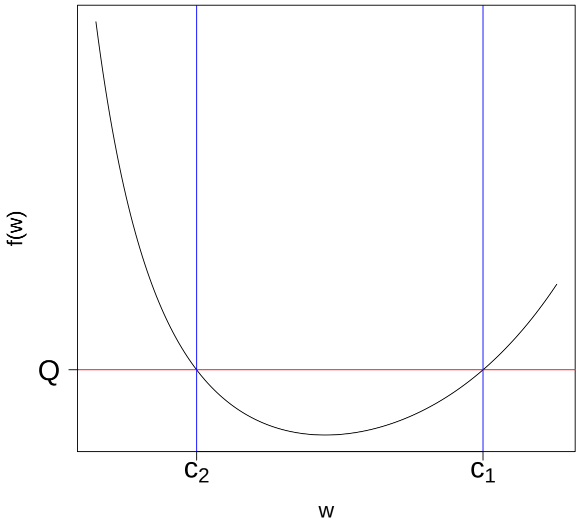
(ii) En supposant que c_1 et c_2 sont telles que

$$\mathbb{P}_{H_0}(W > c_1) = \alpha/2 \text{ et } \mathbb{P}_{H_0}(W < c_2) = \alpha/2,$$

nous obtenons que $c_1 = \chi^2_{n-1,1-\alpha/2}$ et $c_2 = \chi^2_{n-1,\alpha/2}$, puisque $W \sim \chi^2_{n-1}$ sous H_0. Lorsque $\alpha = 0.05$, nous obtenons $c_1 = 19$ et $c_2 = 2.7$. Les données nous donnent $W = 7.27$. Nous ne rejetons pas l'hypothèse nulle puisque $\delta(X_1,\ldots,X_{10}) = \mathbf{1}\{7.27 > 19\} + \mathbf{1}\{7.27 < 2.7\} = 0$.

Exercice 70. La brasserie québécoise Unibroue produit des bières mondialement reconnues [2]. Elle souhaite vérifier si les bouteilles de bière qu'elle produit contiennent bien 341 ml, comme indiqué à l'étiquette. En effet, si la quantité était inférieure à 341 ml, la brasserie risquerait

2. http ://www.unibroue.com/fr/unibroue/medals



un mécontentement de la part de sa fidèle clientèle, ainsi que des problèmes juridiques. En revanche, une quantité supérieure à 341 entrainerait des pertes financières. Afin d'effectuer cette vérification, la quantité de bière dans $n = 100$ bouteilles a été mesurée, et les valeurs $x_1, \ldots, x_n$ ont été observées. On suppose que les observations x_i sont indépendantes et tirées d'une loi normale $\mathcal{N}(\mu, \sigma^2)$ dont les deux paramètres sont inconnus. Les observations obtenues sont de moyenne $\bar{x} = 337$ et de variance échantillonnale $S^2 = 40$. Tester à un niveau $\alpha = 0.05$ si les bouteilles produites contiennent en moyenne 341 ml.
Indice : consulter l'exemple 4.22 (p. 119).
Est-ce que la conclusion changerait si n était égal à 10 ?

Solution 70. Nous voulons faire le test bilatéral $H_0 : \mu = 341$ contre $H_1 : \mu \neq 341$. Puisque les données sont tirées d'une loi normale, il suffit d'appliquer le résultat trouvé à l'exemple 4.22. La fonction de test est donc

$$\delta(X_1, \ldots, X_n) = \mathbf{1}\left\{\frac{|337 - 341|}{\sqrt{40/100}} > t_{99,0.975}\right\} = \mathbf{1}\left\{\sqrt{40} > 1.984\right\} = 1.$$

Nous rejetons donc l'hypothèse nulle au niveau 0.05.
Par contre, si l'échantillon était de taille 10, la fonction de test serait

$$\delta(X_1, \ldots, X_n) = \mathbf{1}\left\{\frac{|337 - 341|}{\sqrt{40/10}} > t_{9,0.975}\right\} = \mathbf{1}\{2 > 2.262\} = 0,$$

on ne rejetterait donc pas l'hypothèse nulle.

Exercice 71 (exercice 54)**.**

(i) Soit $X_1, \ldots, X_n$ un échantillon tiré d'une distribution de Poisson de paramètre θ. Nous voulons tester $H_0 : \theta = \theta_0$ vs. $H_1 : \theta \neq \theta_0$. Trouver un test du rapport de vraisemblance approximatif permettant de tester ces deux hypothèses.
Indice : utiliser le théorème 4.23.

(ii) Supposons que nous ayons observé $n = 100$ observations de moyenne $\bar{x} = 2.1$. Tester à un seuil de signification $\alpha = 0.05$ les hypothèses H_0 et H_1 définies ci-dessus pour $\theta_0 = 2$.

Solution 71.

(i) La fonction de vraisemblance

$$L_n(\theta) = \frac{e^{-n\theta}\theta^{\sum_{i=1}^n X_i}}{\prod_{i=1}^n X_i!},$$

est maximisée en $\widehat{\theta}_n = \overline{X}_n$. Le rapport de vraisemblance est donc

$$\Lambda_n(X_1, \ldots, X_n) = \frac{L_n(\overline{X}_n)}{L_n(\theta_0)} = e^{n(\theta_0 - \overline{X}_n)} \left(\frac{\overline{X}_n}{\theta_0}\right)^{\sum_{i=1}^n X_i}.$$

Par le théorème 4.23, nous savons que

$$2n\left(\theta_0 - \overline{X}_n + \overline{X}_n \log \frac{\overline{X}_n}{\theta_0}\right) = 2 \log \Lambda_n(X_1, \ldots, X_n) \xrightarrow{d} \chi_1^2,$$

sous H_0. Un test approximatif peut donc être défini par la fonction de test

$$\mathbf{1}\{2 \log \Lambda_n(X_1, \ldots, X_n) > \chi^2_{1,1-\alpha}\} = \mathbf{1}\left\{2n\left(\theta_0 - \overline{X}_n + \overline{X}_n \log \frac{\overline{X}_n}{\theta_0}\right) > \chi^2_{1,1-\alpha}\right\}.$$

(ii) Nous avons

$$\mathbf{1}\left\{2 \cdot 100\left(2 - 2.1 + 2.1 \log \frac{2.1}{2}\right) > \chi^2_{1,0.95}\right\} = \mathbf{1}\{0.49 > 3.84\} = 0,$$

nous ne rejetons donc pas l'hypothèse nulle.

Exercice 72 (exercice 55)**.** Soit un échantillon iid $X_1, \ldots, X_n$ issu d'une loi $N(0, \sigma^2)$ où la variance σ^2 est inconnue. Construire un test de Wald approximatif (de niveau α) afin de tester l'hypothèse $H_0 : \sigma^2 = \sigma_0^2$ versus $H_1 : \sigma^2 \neq \sigma_0^2$ pour $\sigma_0^2 > 0$ fixé. Comparer avec le test du rapport de vraisemblance.

Solution 72. La fonction de densité s'écrit

$$f(x; \sigma^2) = \exp\left\{-\frac{1}{2}\ln(2\pi\sigma^2) - \frac{x^2}{2\sigma^2}\right\}; \qquad \theta = \sigma^2 > 0, \qquad x \in \mathbb{R},$$

de sorte que $\eta(\theta) = -1/(2\theta)$ et $d(\theta) = \frac{1}{2}\ln(2\pi\theta)$. L'estimateur du maximum de vraisemblance est $\widehat{\sigma}_n^2 = \widehat{\theta}_n = n^{-1}\sum_{i=1}^n X_i^2$. Le calcul

$$\eta'(\theta) = \frac{1}{2\theta^2}; \quad \eta''(\theta) = -\frac{1}{\theta^3}; \quad d'(\theta) = \frac{1}{2\theta}; \quad d''(\theta) = -\frac{1}{2\theta^2},$$

implique que

$$\widehat{J}_n = n\frac{d''(\widehat{\theta}_n)\eta'(\widehat{\theta}_n) - d'(\widehat{\theta}_n)\eta''(\widehat{\theta}_n)}{\eta'(\widehat{\theta}_n)} = n\widehat{\theta}_n^{-2}/2.$$

Afin de tester l'hypothèse nulle $H_0 : \sigma^2 = \sigma_0^2$, M. Wald vérifiera si

$$Q < \widehat{J}_n(\widehat{\theta}_n - \theta_0)^2 = \frac{n}{2}\left(1 - \frac{\sigma_0^2}{\widehat{\sigma}_n^2}\right)^2.$$

D'après le théorème 4.26, la distribution approximative du membre à droite est χ_1^2 sous H_0. Ainsi (voir la remarque 4.27), le test de Wald approximatif rejette H_0 si et seulement si

$$\frac{n}{2}\left(1 - \frac{\sigma_0^2}{\widehat{\sigma}_n^2}\right)^2 > \chi^2_{1,1-\alpha}.$$

Remarque. On connaît la loi exacte de $\widehat{J}_n(\widehat{\theta}_n - \theta_0)^2$ sous H_0. En effet, $n\widehat{\sigma}_n^2/\sigma_0^2 \sim \chi_n^2$. À partir de là on peut trouver le test exacte, mais il n'aura pas une forme explicite ; cela ressemble à ce qui se passe dans l'exercice 4 de la série 10.

Rapport de vraisemblance. Puisque la moyenne $\mu = 0$ est connue, l'estimateur du maximum de vraisemblance est différent qu'à l'exercice 1 de cette série et il faut refaire le calcul :

$$\begin{aligned}
L_n(\sigma^2) &= \left(\frac{1}{2\pi\sigma^2}\right)^{n/2} \exp\left(-\frac{\sum_{i=1}^n X_i^2}{2\sigma^2}\right) \\
\ell_n(\sigma^2) &= -\frac{n}{2}\log 2\pi\sigma^2 - \frac{\sum_{i=1}^n X_i^2}{2\sigma^2} \\
\ell_n'(\sigma^2) &= -\frac{n}{2\sigma^2} + \frac{\sum_{i=1}^n X_i^2}{2\sigma^4} \qquad \Longrightarrow \qquad \widehat{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^n X_i^2 \\
\ell_n''(\widehat{\sigma}_n^2) &= \frac{n}{2\widehat{\sigma}_n^4} - \frac{\sum_{i=1}^n X_i^2}{\widehat{\sigma}_n^6} = -\frac{n}{2}\widehat{\sigma}_n^4 < 0.
\end{aligned}$$

Le rapport de vraisemblance est

$$\Lambda_n = \frac{L_n(\widehat{\sigma}_n^2)}{L_n(\sigma_0^2)} = \left(\frac{\sigma_0^2}{\widehat{\sigma}_n^2}\right)^{n/2} \exp\left(\frac{n}{2}\frac{\widehat{\sigma}_n^2}{\sigma_0^2}\right)\exp\left(-\frac{n}{2}\right).$$

Quand H_0 est vraie $2\log\Lambda_n \xrightarrow{d} \chi_1^2$ (théorème 4.23).
Le test approximatif est donc (voir la remarque 4.24)

$$n\left[\frac{\widehat{\sigma}_n^2}{\sigma_0^2} - \log\frac{\widehat{\sigma}_n^2}{\sigma_0^2} - 1\right] > \chi^2_{1,1-\alpha}.$$

En faisant l'approximation de Taylor $\log x \approx \log 1 + (x-1) - (x-1)^2/2$ pour $x \approx 1$ on voit que les deux tests sont proches.

Exercice 73 (exercice 56)**.** Soit un échantillon iid $X_1, \ldots, X_n$ issu d'une loi Bernoulli de paramètre p inconnu. Construire un test de Wald approximatif (de niveau α) afin de tester l'hypothèse $H_0 : p = p_0$ versus $H_1 : p \neq p_0$ pour $p_0 \in]0,1[$ fixé. Comparer avec le test de rapport du vraisemblance.

Solution 73. La fonction de masse s'écrit

$$f(x;p) = \exp\{x \ln p + (1-x)\ln(1-p)\} = \exp\{x[\ln p - \ln(1-p)] + \ln(1-p)\}, \quad x \in \{0,1\},$$

de sorte que $\eta(p) = \ln p - \ln(1-p)$ et $d(p) = -\ln(1-p)$. L'estimateur du maximum de vraisemblance est $\widehat{p}_n = n^{-1}\sum_{i=1}^n X_i = \overline{X}_n$ et $n\overline{X}_n \sim Binom(n,p)$. Calculons

$$\eta'(p) = \frac{1}{p(1-p)}; \quad \eta''(p) = \frac{2p-1}{p^2(1-p)^2}; \quad d'(p) = \frac{1}{1-p}; \quad d''(p) = \frac{1}{(1-p)^2},$$

d'où

$$\widehat{J}_n = n\frac{d''(\widehat{\theta}_n)\eta'(\widehat{\theta}_n) - d'(\widehat{\theta}_n)\eta''(\widehat{\theta}_n)}{\eta'(\widehat{\theta}_n)} = \frac{n}{\widehat{p}_n(1-\widehat{p}_n)}.$$

Le test de Wald rejette l'hypothèse $H_0 : p = p_0$ si et seulement si

$$Q < \frac{n}{\widehat{p}_n(1-\widehat{p}_n)}(\widehat{p}_n - p_0)^2 \xrightarrow{\text{d}} \chi_1^2 \quad (\text{sous } H_0, \text{ théorème 4.26}).$$

Le test de Wald approximatif est donc (remarque 4.27)

$$n\frac{(\widehat{p}_n - p_0)^2}{\widehat{p}_n(1-\widehat{p}_n)} > \chi^2_{1,1-\alpha}.$$

Le test du rapport de vraisemblance rejette si $(\widehat{p}_n/p_0)^{n\widehat{p}_n}[(1-\widehat{p}_n)/(1-p_0)]^{n-n\widehat{p}_n}$ est grand, et 2 fois le logarithme de cette quantité converge en distribution vers une variable aléatoire χ_1^2 quand H_0 est vraie (théorème 4.23).
Le test approximatif est donc (remarque 4.24)

$$2n\left[\widehat{p}_n \log\frac{\widehat{p}_n}{p_0} + (1-\widehat{p}_n)\log\frac{1-\widehat{p}_n}{1-p_0}\right] > \chi^2_{1,1-\alpha}.$$

Exercice 74 (exercice 53, **test non apparié**)**.** Soit un échantillon $X_1, \ldots, X_n$, $Y_1, \ldots, Y_m$ de $n+m$ variables aléatoires indépendantes, où $X_i \overset{iid}{\sim} \mathcal{N}(\mu_1, \sigma^2)$ et $Y_i \overset{iid}{\sim} \mathcal{N}(\mu_2, \sigma^2)$, où σ^2 est inconnue (mais la même pour les X et les Y). Le but de cet exercice est de trouver le test du rapport de vraisemblance permettant de tester $H_0 : \mu_1 = \mu_2$ contre $H_1 : \mu_1 \neq \mu_2$.

(i) Définir la fonction de vraisemblance du paramètre $\theta = (\mu_1, \mu_2, \sigma^2)$.

(ii) En remarquant que $\Theta_0 = \{(\mu, \mu, \sigma^2) : -\infty < \mu < \infty, 0 < \sigma^2 < \infty\}$ et que $\Theta_1 = \{(\mu_1, \mu_2, \sigma^2) : -\infty < \mu_1 \neq \mu_2 < \infty, 0 < \sigma^2 < \infty\}$, montrer que

$$\sup_{\theta\in\Theta_0} L(\theta) = \left(\frac{e^{-1}}{2\pi\hat{\sigma}^2_{\Theta_0}}\right)^{(m+n)/2},$$

où $\hat{\sigma}^2_{\Theta_0} = \frac{1}{n+m}\left(\sum_{i=1}^n (X_i - \hat{\mu})^2 + \sum_{j=1}^m (Y_j - \hat{\mu})^2\right)$, avec $\hat{\mu} = \frac{1}{n+m}\left(\sum_{i=1}^n X_i + \sum_{j=1}^m Y_j\right)$.
Montrer aussi que

$$\sup_{\theta\in\Theta_1} L(\theta) = \left(\frac{e^{-1}}{2\pi\hat{\sigma}^2_{\Theta_1}}\right)^{(m+n)/2},$$

où $\hat{\sigma}^2_{\Theta_1} = \frac{1}{n+m}\left(\sum_{i=1}^n (X_i - \bar{X})^2 + \sum_{j=1}^m (Y_j - \bar{Y})^2\right)$.

(iii) En utilisant le fait que $\sum_{i=1}^{n}(X_i-\hat{\mu})^2=\sum_{i=1}^{n}(X_i-\bar{X})^2+\frac{nm^2(\bar{X}-\bar{Y})^2}{(n+m)^2}$ et que $\sum_{j=1}^{m}(Y_j-\hat{\mu})^2=\sum_{j=1}^{m}(Y_j-\bar{Y})^2+\frac{mn^2(\bar{X}-\bar{Y})^2}{(n+m)^2}$, montrer que

$$\Lambda(X_1,\ldots,X_n,Y_1,\ldots,Y_m)=\left(1+\frac{t^2}{m+n-2}\right)^{(n+m)/2},$$

où

$$t=\frac{\sqrt{\frac{nm}{n+m}}(\bar{X}-\bar{Y})}{\sqrt{\frac{1}{n+m-2}[(n-1)S_X^2+(m-1)S_Y^2]}},$$

avec $S_X^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2$ et $S_Y^2=\frac{1}{m-1}\sum_{j=1}^{m}(Y_j-\bar{Y})^2$.

(iv) En utilisant le fait que le test de niveau α dont la fonction de test est donnée par $\mathbf{1}\{\Lambda(X_1,\ldots,X_n,Y_1,\ldots,Y_m)>Q\}$ est le même que celui dont la fonction de test est $\mathbf{1}\{|t|>Q'\}$ où Q' est tel que $\sup_{\theta\in\Theta_0}\mathbb{P}_\theta(|t|>Q')=\alpha$, énoncer le test du rapport de vraisemblance, i.e. trouver la loi de t sous H_0 et par le fait même la valeur de Q'.
Indice : si $A\sim\chi_a^2$ et $B\sim\chi_b^2$ sont indépendantes, alors $A+B\sim\chi_{a+b}^2$.

Solution 74.

(i) La fonction de vraisemblance est

$$\begin{aligned}L(\mu_1,\mu_2,\sigma^2)&=\prod_{i=1}^{n}f(X_i;\mu_1,\sigma^2)\prod_{j=1}^{m}f(Y_j;\mu_2,\sigma^2)\\&=\left(\frac{1}{2\pi\sigma^2}\right)^{(n+m)/2}\exp\left(-\frac{1}{2\sigma^2}\left[\sum_{i=1}^{n}(X_i-\mu_1)^2+\sum_{j=1}^{m}(Y_j-\mu_2)^2\right]\right).\end{aligned}$$

(ii) Lorsque $\theta\in\Theta_0$, nous sommes dans la situation bien connue d'un échantillon iid, de taille $n+m$, tiré d'une $\mathcal{N}(\mu,\sigma^2)$. Nous avons donc que le supremum de $L(\theta)$ est atteint en

$$\widehat{\theta}=(\widehat{\mu},\widehat{\mu},\widehat{\sigma}^2_{\Theta_0})=\left(\frac{n\overline{X}+m\overline{Y}}{n+m},\frac{n\overline{X}+m\overline{Y}}{n+m},\frac{1}{n+m}\left(\sum_{i=1}^{n}(X_i-\widehat{\mu})^2+\sum_{j=1}^{m}(Y_j-\widehat{\mu})^2\right)\right),$$

il est donc égal à

$$\begin{aligned}L(\widehat{\mu},\widehat{\mu},\widehat{\sigma}^2_{\Theta_0})&=\left(\frac{1}{2\pi\widehat{\sigma}^2_{\Theta_0}}\right)^{(n+m)/2}\exp\left(-\frac{1}{2\widehat{\sigma}^2_{\Theta_0}}\left[\sum_{i=1}^{n}(X_i-\widehat{\mu})^2+\sum_{j=1}^{m}(Y_j-\widehat{\mu})^2\right]\right)\\&=\left(\frac{1}{2\pi\widehat{\sigma}^2_{\Theta_0}}\right)^{(n+m)/2}\exp\left(-\frac{n+m}{2}\right)\\&=\left(\frac{e^{-1}}{2\pi\widehat{\sigma}^2_{\Theta_0}}\right)^{(n+m)/2}.\end{aligned}$$

Lorsque $\theta \in \Theta_1$, nous devons maximiser la fonction de vraisemblance trouvée en (i) par rapport à (μ_1, μ_2, σ^2). En dérivant la fonction de log-vraisemblance par rapport à chacun des paramètres et en posant les 3 expressions obtenues égales à 0, nous obtenons que le supremum de $L(\theta)$ est atteint en

$$\widehat{\theta} = (\widehat{\mu}_1, \widehat{\mu}_2, \widehat{\sigma}^2_{\Theta_1}) = \left(\overline{X}, \overline{Y}, \frac{1}{n+m}\left(\sum_{i=1}^{n}(X_i - \overline{X})^2 + \sum_{j=1}^{m}(Y_j - \overline{Y})^2\right)\right),$$

il est donc égal à

$$\begin{aligned}
L(\widehat{\mu}_1, \widehat{\mu}_2, \widehat{\sigma}^2_{\Theta_1}) &= \left(\frac{1}{2\pi\widehat{\sigma}^2_{\Theta_1}}\right)^{(n+m)/2} \exp\left(-\frac{1}{2\widehat{\sigma}^2_{\Theta_1}}\left[\sum_{i=1}^{n}(X_i - \overline{X})^2 + \sum_{j=1}^{m}(Y_j - \overline{Y})^2\right]\right) \\
&= \left(\frac{1}{2\pi\widehat{\sigma}^2_{\Theta_1}}\right)^{(n+m)/2} \exp\left(-\frac{n+m}{2}\right) \\
&= \left(\frac{e^{-1}}{2\pi\widehat{\sigma}^2_{\Theta_1}}\right)^{(n+m)/2}.
\end{aligned}$$

(La matrice hessienne évaluée en $(\widehat{\mu}_1, \widehat{\mu}_2, \widehat{\sigma}^2_{\Theta_1})$ est diagonale et définie négative.)

(iii) Remarquons tout d'abord qu'utilisant les deux identités fournies dans la question, nous obtenons que $\widehat{\sigma}^2_{\Theta_0}$ peut s'écrire de la façon suivante :

$$\widehat{\sigma}^2_{\Theta_0} = \widehat{\sigma}^2_{\Theta_1} + \frac{mn(\overline{X} - \overline{Y})^2}{(m+n)^2}.$$

Le rapport de vraisemblance est donc

$$\begin{aligned}
\Lambda(X_1, \ldots, X_n, Y_1, \ldots, Y_m) = \frac{L(\widehat{\mu}_1, \widehat{\mu}_2, \widehat{\sigma}^2_{\Theta_1})}{L(\widehat{\mu}, \widehat{\mu}, \widehat{\sigma}^2_{\Theta_0})} &= \left(\frac{\widehat{\sigma}^2_{\Theta_0}}{\widehat{\sigma}^2_{\Theta_1}}\right)^{(n+m)/2} \\
&= \left(1 + \frac{mn(\overline{X} - \overline{Y})^2}{\widehat{\sigma}^2_{\Theta_1}(m+n)^2}\right)^{(n+m)/2} = (1+s)^{(n+m)/2},
\end{aligned}$$

où

$$\begin{aligned}
s &= \frac{mn(\overline{X}-\overline{Y})^2}{\widehat{\sigma}^2_{\Theta_1}(m+n)^2} \\
&= \frac{mn(\overline{X}-\overline{Y})^2}{(m+n)\left[\sum_{i=1}^n (X_i-\overline{X})^2+\sum_{j=1}^m (Y_j-\overline{Y})^2\right]} \\
&= \frac{\frac{mn(\overline{X}-\overline{Y})^2}{(m+n)}}{\frac{n+m-2}{n+m-2}\left[\sum_{i=1}^n (X_i-\overline{X})^2+\sum_{j=1}^m (Y_j-\overline{Y})^2\right]} \\
&= \frac{\left(\frac{\sqrt{\frac{mn}{(m+n)}}(\overline{X}-\overline{Y})}{\sqrt{\frac{1}{n+m-2}\left[\sum_{i=1}^n (X_i-\overline{X})^2+\sum_{j=1}^m (Y_j-\overline{Y})^2\right]}}\right)^2}{n+m-2} \\
&= \frac{\left(\frac{\sqrt{\frac{nm}{n+m}}(\overline{X}-\overline{Y})}{\sqrt{\frac{1}{n+m-2}[(n-1)S_X^2+(m-1)S_Y^2]}}\right)^2}{n+m-2}.
\end{aligned}$$

(iv) Nous devons trouver la distribution de t sous H_0, c'est-à-dire lorsque $\mu_1=\mu_2=\mu$. Pour ce faire, nous allons réécrire t de la façon suivante :

$$t = \frac{\frac{1}{\sigma}\sqrt{\frac{nm}{n+m}}(\overline{X}-\overline{Y})}{\sqrt{\frac{1}{n+m-2}\left[\frac{(n-1)S_X^2}{\sigma^2}+\frac{(m-1)S_Y^2}{\sigma^2}\right]}} = \frac{Z_1}{\sqrt{\frac{Z_2}{n+m-2}}},$$

où σ^2 est la variance inconnue de notre échantillon.

Analysons premièrement Z_1. Nous savons que sous H_0, $\overline{X} \sim \mathcal{N}(\mu,\sigma^2/n)$ et $\overline{Y} \sim \mathcal{N}(\mu,\sigma^2/m)$, puisque ces deux variables aléatoires sont indépendantes, $\overline{X}-\overline{Y} \sim \mathcal{N}(0,\sigma^2(n+m)/nm)$. Ainsi $Z_1 \sim \mathcal{N}(0,1)$.

Analysons maintenant le dénominateur. Nous savons par la proposition 2.7 que $(n-1)S_X^2/\sigma^2 \sim \chi^2_{n-1}$ et que $(m-1)S_Y^2/\sigma^2 \sim \chi^2_{m-1}$. En utilisant l'indice fourni dans la question, nous obtenons que $Z_2=(n-1)S_X^2/\sigma^2+(m-1)S_Y^2/\sigma^2 \sim \chi^2_{n+m-2}$, puisque les deux variables aléatoires sont indépendantes. La proposition 2.7 nous dit également que Z_1 et Z_2 sont indépendantes, et d'après l'exercice 6 de la série 2, $t \sim t_{n+m-2}$. Le test du rapport de vraisemblance est donc défini par la fonction de test suivante :

$$\delta(X_1,\dots,X_n,Y_1,\dots,Y_m) = \mathbf{1}\{|t| > t_{n+m-2,1-\alpha/2}\}.$$

Exercice 75 (*exercice 57)**.** Soient $X_1,\dots,X_n \overset{iid}{\sim} f(x;\theta)$. Supposons que l'on veut tester $H_0:\theta=\theta_0$ versus $H_1:\theta\neq\theta_0$ en utilisant la fonction de test δ_α de la forme

$$\delta_\alpha(T(X_1,\dots,X_n)) = \mathbf{1}\{T(X_1,\dots,X_n) > q_{1-\alpha}\} \text{ ou } \delta_\alpha(T(X_1,\dots,X_n)) = \mathbf{1}\{T(X_1,\dots,X_n) < q_\alpha\},$$

où q_α est le α-quantile de G_0, la fonction de distribution de $T(X_1,\dots,X_n)$ quand $\theta=\theta_0$. Supposons que G_0 est une fonction continue. Montrer que sous H_0, la valeur-p suit la distribution uniforme sur $[0,1]$.
Indice : utiliser le lemme 4.30.

Solution 75. D'après le lemme 4.30, il suffit de montrer que $G_0(T(X_1, \dots, X_n)) \sim U[0,1]$ sous H_0. Or, sous H_0, G_0 est la fonction de distribution de $T(X_1, \dots, X_n)$, supposée continue. Lorsque Z est une variable aléatoire avec fonction de distribution continue G (cf. l'exercice 2, série 2)

$$\mathbb{P}(G(Z) \geq u) = \mathbb{P}(Z \geq G^{-1}(u)) = 1 - G(G^{-1}(u)) = 1 - u, \qquad u \in]0,1[,$$

où $G^{-1}(u) = \inf\{t : G(t) \geq u\}$ et les deux dernières égalités découlent de la continuité de G. Ainsi $G_0(T(X_1, \dots, X_n)) \sim U[0,1]$ et $1 - G_0(T(X_1, \dots, X_n)) \sim U[0,1]$ si H_0 est vraie.

Exercice 76 (exercice 60, **intervalle bilatéral optimal**). Afin de construire un intervalle de confiance bilatéral pour la moyenne d'une distribution normale (dont la variance est connue), nous avons choisi $z_{\alpha/2}$ et $z_{1-\alpha/2}$ comme bornes de l'intervalle (cf. exemple 5.3). L'on peut se demander pourquoi ne pas choisir par exemple $z_{\alpha/3}$ et $z_{1-2\alpha/3}$.
Il est vrai qu'on aime les intervalles plus « naturels » ou symétriques, mais la raison de ce choix est la suivante :

(i) Soient $Z \sim N(0,1)$ et $\alpha \in]0,1[$. Montrer que l'intervalle $I = [L, U]$ ayant la plus petite longueur et tel que $\mathbb{P}(I \ni Z) \geq 1 - \alpha$ est donné par $L = z_{\alpha/2}$ et $U = z_{1-\alpha/2}$.

(ii) Soient $X_1, \dots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$ où la variance σ^2 est connue. Trouver l'intervalle $I_n = [A_n, B_n]$ ayant la plus petite longueur et tel que $\mathbb{P}(I_n \ni \mu) \geq 1 - \alpha$.

(iii) *Peut-on généraliser ce résultat ?

Solution 76. (i) Il faut résoudre le problème suivant :

$$\min U - L \qquad \text{t.q.} \qquad \Phi(U) - \Phi(L) \geq 1 - \alpha \qquad (U, L \in \mathbb{R}).$$

Puisque Φ est une fonction croissante, la contrainte peut s'écrire $U \geq \Phi^{-1}(1 - \alpha + \Phi(L))$. Pour un L donné, il faut choisir le U le plus petit qui satisfait la contrainte. Ainsi, notre problème se réduit à trouver

$$\min g(L) = \Phi^{-1}(1 - \alpha + \Phi(L)) - L, \qquad L \in \mathbb{R}.$$

Notons cependant que $\Phi(L) \leq \Phi(U) - 1 + \alpha < \alpha$ et le domaine de g est $] - \infty, \Phi^{-1}(\alpha)[$. De plus, $g(L) \to \infty$ lorsque $L \to -\infty$ ou lorsque $L \to \Phi^{-1}(\alpha)$, et $g \geq 0$. Le minimum de g sera donc atteint à un point intérieur du domaine de g. Celle-ci est dérivable par le théorème de la fonction inverse (car Φ est strictement croissante et continûment dérivable).
La dérivée de g s'annule si et seulement si

$$1 = \frac{\Phi'(L)}{\Phi'(\Phi^{-1}(1 - \alpha + \Phi(L)))} = \frac{\Phi'(L)}{\Phi'(U)} = \frac{\exp(-L^2/2)}{\exp(-U^2/2)},$$

c'est-à-dire lorsque $L = \pm U$. Or, Φ est croissante et $\Phi(U) - \Phi(L) = 1 - \alpha > 0$, donc forcément $L < U$. On a donc $L = -U$ et par symétrie $\Phi(U) = 1 - \Phi(L)$, donc

$$1 - \alpha = \Phi(U) - \Phi(L) = 1 - 2\Phi(L) \Longrightarrow \Phi(L) = \frac{\alpha}{2} \Longrightarrow \Phi(U) = 1 - \frac{\alpha}{2}.$$

Le but de la discussion ci-dessus était de montrer qu'il s'agit d'un minimum sans devoir calculer la dérivée seconde de g . À noter qu'il est facile dans ce cas de montrer que $g''(\Phi^{-1}(\alpha/2)) > 0$ et donc qu'il s'agit bel et bien d'un minimum.

Remarque. Le choix $L = \Phi^{-1}(\alpha)$ correspond à $U = \infty$ et donne l'intervalle de confiance unilatéral à gauche. Le choix $L = -\infty$ correspond à $U = \Phi^{-1}(1-\alpha)$ et donne l'intervalle unilatéral à droite.

(ii) Comme dans l'exemple 5.3, posons $Z_n = \sqrt{n}(\overline{X}_n - \mu)/\sigma \sim N(0,1)$ et remarquons que

$$\mathbb{P}[A_n \leq \mu \leq B_n] = \mathbb{P}(\overline{X}_n - B_n \leq \overline{X}_n - \mu \leq \overline{X}_n - A_n) = \mathbb{P}\left[\frac{\sqrt{n}}{\sigma}(\overline{X}_n - B_n) \leq Z_n \leq \frac{\sqrt{n}}{\sigma}(\overline{X}_n - A_n)\right].$$

Il faut minimiser $B_n - A_n$, ce qui équivaut à minimiser $(\sqrt{n}/\sigma)(\overline{X}_n - A_n) - (\sqrt{n}/\sigma)(\overline{X}_n - B_n)$, mais sous la contrainte que cette probabilité soit au moins $1-\alpha$. Par la partie (i), la solution est

$$\left[\frac{\sqrt{n}}{\sigma}(\overline{X}_n - B_n), \frac{\sqrt{n}}{\sigma}(\overline{X}_n - A_n)\right] = [z_{\alpha/2}, z_{1-\alpha/2}] = [-z_{1-\alpha/2}, z_{1-\alpha/2}].$$

Ainsi, la solution de notre problème est

$$[A_n, B_n] = \left[\overline{X}_n - z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}, \overline{X}_n + z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\right],$$

qui est donc l'intervalle de confiance basé sur $\overline{X}_n$ de seuil (supérieure ou égale à) $(1-\alpha)$ ayant la plus petite longueur.

(iii) Le même résultat est valable lorsque Z suit une loi ayant une densité symétrique f, qui est strictement décroissante sur $\mathbb{R}_+$. C'est-à-dire, le résultat est valable si

— pour chaque $x \in \mathbb{R}$, $f(x) = f(-x)$;

— pour chaque $0 < x < y$, $f(x) > f(y)$.

Par exemple, ceci est bien le cas si $Z \sim \mathbf{t}_k$ pour $k > 0$. Ainsi, même si la variance σ^2 est inconnue, en la remplaçant par l'estimateur S^2, on obtiendra l'intervalle de confiance ayant la plus petite longueur.

Remarque. Sous ces conditions, on peut montrer que l'intervalle $[L, U]$ est l'ensemble (mesurable) F ayant la mesure de Lebesgue la plus petite et tel que $\mathbb{P}(F \ni Z) \geq 1-\alpha$. Il est donc inutile de chercher (par exemple) une union d'intervalles.

Exercice 77 (exercice 61, **différence de moyennes**).

(i) Soient $X_1, \ldots, X_n \overset{\text{iid}}{\sim} N(\mu_X, \sigma^2)$ et $Y_1, \ldots, Y_n \overset{\text{iid}}{\sim} N(\mu_Y, \sigma^2)$ deux échantillons indépendants, où μ_X, μ_Y et σ^2 sont inconnus. Trouver un intervalle de confiance bilatéral pour le paramètre $\theta = \mu_X - \mu_Y$ avec un seuil de confiance $1-\alpha$.

(ii) On veut comparer la durée d'efficacité de deux nouveaux médicaments, M_1 et M_2, contre la lombalgie [3]. On a donc administré chaque médicament à un groupe de 15 patients, et ensuite mesuré (en heures) la période sans douleur après la prise du médicament. On obtient la moyenne du temps d'efficacité $\overline{X_1} = 7.5$ pour M_1 et $\overline{X_2} = 6.3$ pour M_2. On

3. C'est ce qu'a eu Pierre Brochant dans le film *le dîner des cons*. Il n'est pas le seul : on estime qu'entre 40 et 70% de la population en sera touché au cours de la vie.

a aussi les écart-types estimés $S_1 = 1.1$ et $S_2 = 1.3$ pour M_1 et M_2 respectivement. En supposant que les observations des groupes 1 et 2 sont indépendantes et suivent des lois $N(\mu_1, \sigma^2)$ et $N(\mu_2, \sigma^2)$ respectivement, donner l'intervalle de confiance à 95% pour la différence $\mu_1 - \mu_2$. Que peut-on constater sur l'efficacité relative de M_1 et M_2 ?

Solution 77. (i) D'après l'exercice 74 (avec $n = m$), la variable aléatoire

$$T = \frac{\sqrt{\frac{n^2}{n+n}}\left(\overline{X} - \mu_X - \overline{Y} + \mu_Y\right)}{\sqrt{\frac{1}{n+n-2}[(n-1)S_X^2 + (n-1)S_Y^2]}} = \frac{\sqrt{n}\left(\overline{X} - \overline{Y} - \theta\right)}{\sqrt{S_X^2 + S_Y^2}} = \frac{\overline{X} - \overline{Y} - \theta}{\sqrt{\frac{1}{n}(S_X^2 + S_Y^2)}}$$

suit une loi $\mathbf{t}_{2n-2}$ pour chaque $\theta \in \mathbb{R}$. (Parce que $S_X^2 = S_{X-c}^2$ pour chaque constante $c \in \mathbb{R}$.) Ainsi, $T = g(X_1, \ldots, X_n, Y_1, \ldots, Y_n, \theta)$ est un pivot (la continuité par rapport à θ est évidente). À partir de là, on n'a qu'à faire les manipulations habituelles :

$$\begin{aligned}
1 - \alpha &= \mathbb{P}_\theta\left[t_{2n-2,\alpha/2} \leq \frac{\bar{X} - \bar{Y} - \theta}{\sqrt{\frac{1}{n}(S_X^2 + S_Y^2)}} \leq t_{2n-2,1-\alpha/2}\right] \\
&= \mathbb{P}_\theta\left[t_{2n-2,\alpha/2}\sqrt{\frac{1}{n}(S_X^2 + S_Y^2)} \leq \bar{X} - \bar{Y} - \theta \leq t_{2n-2,1-\alpha/2}\sqrt{\frac{1}{n}(S_X^2 + S_Y^2)}\right] \\
&= \mathbb{P}_\theta\left[\bar{X} - \bar{Y} - t_{2n-2,1-\alpha/2}\sqrt{\frac{1}{n}(S_X^2 + S_Y^2)} \leq \theta \leq \bar{X} - \bar{Y} - t_{2n-2,\alpha/2}\sqrt{\frac{1}{n}(S_X^2 + S_Y^2)}\right] \\
&= \mathbb{P}_\theta\left[\bar{X} - \bar{Y} - t_{2n-2,1-\alpha/2}\sqrt{\frac{1}{n}(S_X^2 + S_Y^2)} \leq \theta \leq \bar{X} - \bar{Y} + t_{2n-2,1-\alpha/2}\sqrt{\frac{1}{n}(S_X^2 + S_Y^2)}\right].
\end{aligned}$$

On conclut que $\left[\bar{X} - \bar{Y} - t_{2n-2,1-\alpha/2}\sqrt{\frac{1}{n}(S_X^2 + S_Y^2)}, \bar{X} - \bar{Y} + t_{2n-2,1-\alpha/2}\sqrt{\frac{1}{n}(S_X^2 + S_Y^2)}\right]$ est un intervalle de confiance pour $\theta = \mu_X - \mu_Y$ avec un seuil $1 - \alpha$.

Remarque. On peut définir $Z_i = X_i - Y_i \overset{\text{iid}}{\sim} N(\theta, 2\sigma^2)$ (ce qui par ailleurs aurait été plus compliqué si $m \neq n$, c'est-à-dire si le nombre de X_i n'était pas égal au nombre de Y_i) de sorte que

$$\frac{\overline{X} - \overline{Y} - \theta}{\sqrt{\frac{1}{n}S_Z^2}} = \sqrt{n}\frac{\overline{Z} - \theta}{\sqrt{S_Z^2}} = \frac{\sqrt{\frac{n}{2\sigma^2}}(\overline{Z} - \theta)}{\sqrt{\frac{1}{n-1}}\sqrt{\frac{(n-1)S_Z^2}{2\sigma^2}}} \sim \mathbf{t}_{n-1}, \qquad S_Z^2 = \sum_{i=1}^{n}(Z_i - \overline{Z})^2,$$

car le nominateur suit une loi $N(0, 1)$ et le dénominateur est $V/\sqrt{n-1}$ où $V \sim \chi_{n-1}^2$, et les deux sont indépendantes. Ainsi on obtient l'intervalle de confiance

$$\left[\bar{X} - \bar{Y} - t_{n-1,1-\alpha/2}\sqrt{\frac{1}{n}S_Z^2}, \bar{X} - \bar{Y} + t_{n-1,1-\alpha/2}\sqrt{\frac{1}{n}S_Z^2}\right].$$

Cet intervalle sera probablement plus grand que celui d'avant : puisque $S_Z^2 \to 2\sigma^2$ et $S_X^2 + S_Y^2 \to 2\sigma^2$, on s'attend à ce que les deux aient une taille similaire (ils ont en tous cas la même espérance et la même variance). Or pour β fixé, la fonction $k \mapsto t_{k,\beta}$ est décroissante. Notre deuxième intervalle aura donc tendance à être plus grand, puisqu'on

utilise t_{n-1} au lieu de t_{2n-2}. Intuitivement, on a utilisé n données (les différences $X_i - Y_i$) au lieu d'en utiliser $2n$.
En revanche, le premier intervalle est moins général que le deuxième : ce dernier suppose uniquement que les différences Z_i sont iid, alors que dans le premier cas on a supposé que toutes les X_i sont indépendantes de toutes les Y_i, une supposition plus forte. Dans le cas apparié (exercice 6, série 10), on ne peut utiliser que le deuxième intervalle !

(ii) En utilisant le résultat de la partie a), on obtient l'intervalle de confiance à 95% :

$$\left[\overline{X}_1 - \overline{X}_2 - t_{28,0.975}\sqrt{\frac{1}{15}(S_1^2 + S_2^2)}, \overline{X}_1 - \overline{X}_2 + t_{28,0.975}\sqrt{\frac{1}{15}(S_1^2 + S_2^2)}\right] \approx [0.30, 2.10].$$

On peut donc conclure que le temps de l'efficacité de M_1 est meilleur que celui de M_2 par 18–126 minutes avec un seuil de confiance 95%.

Exercice 78 (*exercice 62)**.** Soient $T_k \sim \mathbf{t}_k$ et soit $Z \sim N(0,1)$. Montrer que $T_k \xrightarrow{d} Z$ lorsque $k \to \infty$.
Indice : s'inspirer des exemples 5.3 et 5.7.

Solution 78. Soient $X_1, \dots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$ avec $\mu \in \mathbb{R}$ et $\sigma^2 > 0$ inconnus. Supposons qu'on aimerait trouver un intervalle de confiance pour μ (donc σ^2 est un paramètre de nuisance). D'après l'exemple 5.7, on sait que

$$\frac{\overline{X}_n - \mu}{S_n/\sqrt{n}} \to N(0,1), \qquad \overline{X}_n = \frac{1}{n}\sum_{i=1}^n X_i, \quad S_n^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X}_n)^2.$$

Or ici les X_i sont normales ; on connaît donc la distribution exacte de $T_n = \sqrt{n}(\overline{X}_n - \mu)/S_n$: d'après le théorème 2.9, elle est $\mathbf{t}_{n-1}$. L'énoncé est donc démontré.

Exercice 79 (exercice 63)**.** En utilisant la même notation que celle de la proposition 5.8 du cours, prouver que le tableau suivant contient les intervalles de confiance approximatifs avec seuil $(1-\alpha)$ pour θ :

Confiance approximative $1-\alpha$	$L(X_1, \dots, X_n)$	$U(X_1, \dots, X_n)$
Bilatéral	$\hat{\theta}_n - z_{1-\alpha/2}\hat{J}_n^{-1/2}$	$\hat{\theta}_n + z_{1-\alpha/2}\hat{J}_n^{-1/2}$
Unilatéral à gauche	$\hat{\theta}_n - z_{1-\alpha}\hat{J}_n^{-1/2}$	$+\infty$
Unilatéral à droite	$-\infty$	$\hat{\theta}_n + z_{1-\alpha}\hat{J}_n^{-1/2}$

Indice : si $Z_n \xrightarrow{d} Z$, où Z est une variable aléatoire continue, alors $\mathbb{P}[Z_n = a] \to 0$ pour chaque $a \in \mathbb{R}$.

Solution 79. Montrons tout d'abord le résultat dans l'indice.
Soient F_n et F les fonctions de répartition de Z_n et Z respectivement. Pour tout $\epsilon > 0$,

$$0 \le \mathbb{P}(Z_n = a) \le \mathbb{P}(a - \epsilon < Z_n \le a) = F_n(a) - F_n(a-\epsilon) \to F(a) - F(a-\epsilon),$$

lorsque $n \to \infty$, puisque F est continue. Ainsi

$$\limsup_{n\to\infty} \mathbb{P}(Z_n = a) \leq F(a) - F(a-\epsilon) \to 0, \qquad \epsilon \to 0,$$

car F est continue. Ceci prouve le résultat cherché.
Vérifions le cas bilatéral :

$$\begin{aligned}
&\mathbb{P}[\hat{\theta}_n - z_{1-\alpha/2}\hat{J}_n^{-1/2} \leq \theta \leq \hat{\theta}_n + z_{1-\alpha/2}\hat{J}_n^{-1/2}] \\
=&\mathbb{P}[-z_{1-\alpha/2} \leq \hat{J}_n^{1/2}(\hat{\theta}_n - \theta) \leq z_{1-\alpha/2}] \\
=&\mathbb{P}[z_{\alpha/2} \leq \hat{J}_n^{1/2}(\hat{\theta}_n - \theta) \leq z_{1-\alpha/2}] \\
=&F_n(z_{1-\alpha/2}) - F_n(z_{\alpha/2}) + \mathbb{P}[\hat{J}_n^{1/2}(\hat{\theta}_n - \theta) = z_{\alpha/2}],
\end{aligned}$$

où F_n est la fonction de répartition de $\hat{J}_n^{1/2}(\hat{\theta}_n - \theta)$ et où on a utilisé le fait que $z_{\alpha/2} = -z_{1-\alpha/2}$. Par la proposition 5.8 (p. 123), on sait que $F_n(x) \to \Phi(x)$ pour chaque $x \in \mathbb{R}$, où $\Phi(x)$ est la fonction de répartition de $N(0,1)$. De plus, par la proposition ci-dessus

$$\mathbb{P}[\hat{J}_n^{1/2}(\hat{\theta}_n - \theta) = z_{\alpha/2}] \to 0, \qquad n \to \infty.$$

Donc

$$\mathbb{P}[\hat{\theta}_n - z_{1-\alpha/2}\hat{J}_n^{-1/2} \leq \theta \leq \hat{\theta}_n + z_{1-\alpha/2}\hat{J}_n^{-1/2}] \to \Phi(z_{1-\alpha/2}) - \Phi(z_{\alpha/2}) = 1 - \alpha/2 - \alpha/2 = 1 - \alpha.$$

Il s'en suit que $[\hat{\theta}_n - z_{1-\alpha/2}\hat{J}_n^{-1/2}, \hat{\theta}_n + z_{1-\alpha/2}\hat{J}_n^{-1/2}]$ est un intervalle de confiance approximatif avec seuil $1-\alpha$.
De la même façon, on trouve que

$$\mathbb{P}[\hat{\theta}_n - z_{1-\alpha}\hat{J}_n^{-1/2} \leq \theta] = \mathbb{P}[\hat{J}_n^{1/2}(\hat{\theta}_n - \theta) \leq z_{1-\alpha}] = F_n(z_{1-\alpha}) \to \Phi(z_{1-\alpha}) = 1 - \alpha$$

et que

$$\begin{aligned}
\mathbb{P}[\theta \leq \hat{\theta}_n + z_{1-\alpha}\hat{J}_n^{-1/2}] &= 1 - \mathbb{P}[\theta > \hat{\theta}_n + z_{1-\alpha}\hat{J}_n^{-1/2}] \\
&= 1 - \mathbb{P}[\hat{J}_n^{1/2}(\hat{\theta}_n - \theta) < z_\alpha] \\
&= 1 - F_n(z_\alpha) + \mathbb{P}[\hat{J}_n^{1/2}(\hat{\theta}_n - \theta) = z_\alpha] \\
&\to 1 - \Phi(z_\alpha) = 1 - \alpha.
\end{aligned}$$

Donc, $[\hat{\theta} - z_{1-\alpha}\hat{J}_n^{-1/2}, +\infty]$ et $[-\infty, \hat{\theta} + z_{1-\alpha}\hat{J}_n^{-1/2}]$ sont les intervalles de confiance unilatéraux approximatifs avec seuil $1-\alpha$.

Exercice 80 (*exercice 64, **pivot général**).

(i) Soient $X_1, \ldots, X_n \overset{\text{iid}}{\sim} f(x;\theta)$ et $T_n(X_1, \ldots, X_n)$ une statistique continue. Soit $Y_n = F_{T_n}(T_n;\theta)$, où $F_{T_n}(t;\theta) = \mathbb{P}_\theta[T_n \leq t]$ est la fonction de répartition de T_n. Supposons que $F_{T_n}(t;\theta)$ est pour chaque t une fonction continue de θ. Montrer que $Y_n \sim U(0,1)$ et donc que Y_n est un pivot. Comment peut on utiliser ce résultat pour trouver un intervalle de confiance pour θ ?

(ii) Soit $f(x;\theta) = e^{-(x-\theta)}\mathbf{1}_{[\theta,\infty)}(x)$. Utiliser la partie a) et la statistique $T_n = \min\{X_1, \ldots, X_n\}$ pour trouver un intervalle de confiance pour θ avec un seuil $1-\alpha$.

Solution 80. (i) Nous avons déjà montré que $Y_n \sim U(0,1)$. Ainsi, pour chaque θ et chaque $\alpha \in]0,1[$ on a

$$1-\alpha = \mathbb{P}_\theta[\alpha/2 \leq Y_n \leq 1-\alpha/2] = \mathbb{P}_\theta[\alpha/2 \leq F_{T_n}(T_n;\theta) \leq 1-\alpha/2].$$

Donc, l'ensemble

$$S = \{\theta \in \Theta : \alpha/2 \leq F_{T_n}(T_n;\theta) \leq 1-\alpha/2\}$$

est une région de confiance pour θ avec un seuil $1-\alpha$. Si S est un intervalle, on a trouvé un intervalle de confiance pour θ. Cela est le cas par exemple quand $F_{T_n}(t;\theta)$ est une fonction monotone de θ pour chaque t. Si c'est une fonction croissante et T_n est une variable aléatoire continue, alors

$$S = \{\theta \in \Theta : q_{\alpha/2}(\theta) \leq T_n \leq q_{1-\alpha/2}(\theta)\},$$

où q_β est le β-quantile de la distribution de T_n.
Si $T_n = \tau_n$ est la statistique exhaustive d'une famille exponentielle à 1-paramètre, S est la région de confiance au seuil $(1-\alpha)$ obtenue en inversant le test dont la fonction de test est

$$\delta(X_1,\ldots,X_n) = \mathbf{1}\{q_{\alpha/2} \leq \tau_n(X_1,\ldots,X_n) \leq q_{1-\alpha/2}\}.$$

(ii) Trouvons la fonction de répartition de $T_n = \min\{X_1,\ldots,X_n\}$. Ceci se fait facilement en utilisant $\mathbb{P}(T > t) = \mathbb{P}(X_1 > t)^n$ pour chaque t. On peut éviter quelques calculs en remarquant que $X_i - \theta \overset{\text{iid}}{\sim} Exp(1)$ et donc $T_n - \theta \overset{\text{iid}}{\sim} Exp(n)$, de sorte que pour $t \geq \theta$, $1 - F_{T_n}(t;\theta) = \mathbb{P}_\theta[T_n - \theta > t - \theta] = \exp\{-n(t-\theta)\}$. Ainsi

$$F_{T_n}(t;\theta) = \begin{cases} 1-\exp\{-n(t-\theta)\} & t \geq \theta \\ 0 & t < \theta, \end{cases}$$

est décroissante en θ (vérifier les deux cas !) et donc l'ensemble S est un intervalle $[L,U]$. Les bornes sont tels que $F_{T_n}(T_n;L) = 1-\alpha/2$ et $F_{T_n}(T_n;U) = \alpha/2$; autrement dit

$$1-e^{-n(T_n-L)} = 1-\alpha/2, \qquad 1-e^{-n(T_n-U)} = \alpha/2.$$

La solution est

$$[L,U] = \left[T_n + \frac{1}{n}\log(\alpha/2), T_n + \frac{1}{n}\log(1-\alpha/2)\right]$$

qui est par construction un intervalle de confiance pour θ avec un seuil $1-\alpha$:

$$\mathbb{P}([L,U] \ni \theta) = \mathbb{P}\left(\alpha/2 \leq Y_n \leq 1-\alpha/2\right) = 1-\alpha.$$

On remarque que les deux logarithmes sont négatifs ; on sait que $T_n \geq \theta$!

Exercice 81 (exercice 65)**.** Soient $X_1,\ldots,X_n \overset{\text{iid}}{\sim} N(\mu,\sigma^2)$, où σ^2 est connu. Trouver une expression pour l'intervalle de confiance unilatéral à gauche avec seuil $1-\alpha$ pour μ.

Solution 81. Grâce à la proposition 5.17, nous savons qu'il faut inverser un test uniformément le plus puissant pour la paire d'hypothèses

$$H_0 : \mu \leq \mu_0 \qquad \text{vs} \qquad H_1 : \mu > \mu_0.$$

Le théorème 4.16 dit que ce test aura la fonction de test

$$\delta(X_1, \dots, X_n; \mu_0) = \mathbf{1}(\tau_n(X_1, \dots, X_n) > q_{1-\alpha}(\mu_0)),$$

où $\tau_n = \sum_{i=1}^n X_i = n\overline{X}_n$ est la statistique exhaustive et $q_{1-\alpha}(\mu_0)$ est le $(1-\alpha)$-quantile de la distribution de τ_n sous H_0, à savoir $N(\mu_0 n, n\sigma^2)$. Il est élémentaire que $q_{1-\alpha}(\mu_0) = n\mu_0 + \sqrt{n}\sigma z_{1-\alpha}$, où $z_{1-\alpha}$ est le $(1-\alpha)$-quantile d'une loi $N(0,1)$. La région de confiance pour μ est la collection de tous les μ_0 pour lequels on ne rejette pas l'hypothèse nulle avec les données $X_1, \dots, X_n$, soit

$$\begin{aligned} R(X_1, \dots, X_n) &= \{\mu_0 : \tau_n \leq q_{1-\alpha}(\mu_0)\} \\ &= \left\{\mu_0 : \frac{\tau_n - n\mu_0}{\sqrt{n}\sigma} \leq z_{1-\alpha}\right\} \\ &= \{\mu_0 : \mu_0 \geq \overline{X}_n - z_{1-\alpha}\sigma/\sqrt{n}\} \\ &= \left[\overline{X}_n - z_{1-\alpha}\frac{\sigma}{\sqrt{n}}, +\infty\right[. \end{aligned}$$

Les conditions sont satisfaites, puisque c'est une famille exponentielle avec $\eta(\mu) = \mu/\sigma^2$ strictement croissante, τ_n est une variable aléatoire continue, et sa loi $\mathbb{P}_\mu[\tau_n \leq t] = \Phi((t - n\mu)/\sigma\sqrt{n})$ est continue en μ.
La borne inférieure est donc construite à partir de l'estimateur de maximum de vraisemblance, en laissant une marge d'erreur pour compenser le fait que celui-ci est aléatoire. La taille de cette marge d'erreur dépend de α, σ et n comme expliqué à la page 136.

Exercice 82 (exercice 66)**.** Soient $X_1, \dots, X_n \overset{\text{iid}}{\sim} \text{Bern}(p)$. Avec l'aide d'une statistique exhaustive $\tau_n(X_1, \dots, X_n)$ pour p, trouver une expression pour l'intervalle de confiance unilatéral à gauche pour p avec seuil approximatif $1-\alpha$, en inversant le test

$$H_0 : p \leq p_0 \qquad \text{vs} \qquad H_1 : p > p_0.$$

Utiliser une fonction de test qui rejette H_0 lorsque τ_n prend une valeur (strictement) plus grande qu'une certaine valeur critique. Les bornes de cet intervalle ne seront malheureusement pas si explicitent qu'à l'exercice précédente.
Indice : suivre la proposition 5.14. Hélas, une des conditions de cette proposition n'est pas satisfaite (laquelle ?). Ainsi, pour la plupart des valeurs de p, la probabilité de couverture de l'intervalle sera seulement approximativement $1-\alpha$.

Solution 82. L'intervalle cherché peut être obtenu en inversant le test

$$H_0 : p \leq p_0 \qquad \text{vs} \qquad H_1 : p > p_0.$$

Notre test est basé sur $\tau_n = n\overline{X}_n = \sum_{i=1}^n X_i \sim Binom(n,p)$, qui est une statistique exhaustive pour p. On rejette H_0 lorsque $\tau_n > C(p_0)$.

Ainsi, notre intervalle de confiance cherché est détérminé par la région

$$R(X_1, \dots, X_n) = \{p_0 : \tau_n \leq C(p_0)\}.$$

Par le théorème 4.16, le test optimal est obtenu quand $C(p_0) = q_{1-\alpha}(p_0)$, à condition qu'il existe un $q_{1-\alpha}$ tel que $\mathbb{P}_{p_0}[\tau_n \leq q_{1-\alpha}] = 1 - \alpha$. Or, τ_n étant une variable aléatoire discrète, un tel $q_{1-\alpha}$ existe uniquement pour certaines valeurs de α. En particulier, nous ne pouvons pas avoir un seuil de test qui sera exactement α pour n'importe quel α (et donc la probabilité que notre intervalle de confiance contienne p ne sera pas exactement $1 - \alpha$).
On choisit donc $C(p_0)$ de sorte que le seuil du test soit le plus proche possible de α, sans en être plus grand. Ainsi, $C(p_0)$ est détérminé par les deux inégalités suivantes :

$$\mathbb{P}_{p_0}(\tau_n \leq C(p_0)) \geq 1 - \alpha \tag{6}$$

$$\mathbb{P}_{p_0}(\tau_n \leq C(p_0) - 1) < 1 - \alpha. \tag{7}$$

L'inégalité (**??**) dit que le test a un seuil $\leq \alpha$. L'inégalité (**??**) dit que $C(p_0)$ est le nombre entier minimal tel que le test a un seuil $\leq \alpha$. On note que le $(1-\alpha)$-quantile de $Binom(n, p_0)$ satisfait ces deux propriétés (voir la définition 6.6). Donc,

$$C(p_0) = F^{-}_{\tau_n, p_0}(1 - \alpha) = \inf \left\{ t \in \mathbb{R} : \sum_{k=0}^{\lfloor t \rfloor} \binom{n}{k} p_0^k (1 - p_0)^{n-k} \geq 1 - \alpha \right\}.$$

Le fait que $C(p_0)$ est croissante en p_0 n'est pas évidente de l'équation ci-dessus, mais cela résulte de la (preuve de la) proposition 5.17. De toute façon, c'est une fonction en escalier (continue à gauche) telle que $C(0) = 0$ et $C(1) = n$.
La région de confiance qui résulte de l'inversion du test, $\{p_0 : \tau_n \leq C(p_0)\}$, est un intervalle de la forme $(L, 1]$ dont la borne inférieure est

$$L = \inf\{p_0 : \tau_n \leq C(p_0)\} = \inf\{p_0 : \overline{X}_n \leq n^{-1} F^{-}_{\tau_n, p_0}(1 - \alpha)\}.$$

Malheureusement ces expressions n'ont pas une forme plus explicite, mais il est encore facile de calculer la borne L avec un ordinateur.