

Exercice 1 Dans cet exercice il est demandé de donner la ou les bonnes réponses. Seules les réponses justifiées seront validées. Il n'y a pas de points négatifs. Sauf mention contraire, n est un entier naturel supérieur à 2.

1. Soient $X_1, \dots, X_n$ i.i.d. de loi $\text{Beta}(1, \lambda)$, $\lambda \in \mathbb{R}_+^*$, et de réalisations $x_1, \dots, x_n$. L'estimateur du maximum de vraisemblance de λ associé à $(x_1, \dots, x_n)$ est donné par

- (a) $\hat{\lambda}_n = 1/(1 - \bar{x}_n)$, (c) $\hat{\lambda}_n = -\log [\prod_{i=1}^n (1 - x_i)] / n$,
(b) $\hat{\lambda}_n = -\sum_{i=1}^n \log(1 - x_i)/n$, (d) $\hat{\lambda}_n = -n / \sum_{i=1}^n \log(1 - x_i)$.

La bonne proposition est (d) car la log-vraisemblance vaut

$$\ell(\lambda \mid x_1, \dots, x_n) = n \log \underbrace{\frac{\Gamma(1+\lambda)}{\Gamma(1)\Gamma(\lambda)}}_{\lambda} + (\lambda - 1) \sum_{i=1}^n \log(1 - x_i).$$

La log-vraisemblance est deux fois dérivable. Comme la loi $\text{Beta}(1, \lambda)$ appartient à une famille exponentielle minimale régulière, l'unique maximum de la vraisemblance est donné par

$$\frac{\partial}{\partial \lambda} \ell(\lambda \mid x_1, \dots, x_n) = \frac{n}{\lambda} + \sum_{i=1}^n \log(1 - x_i). \quad \text{Donc} \quad \hat{\lambda}(x_1, \dots, x_n) = -n / \sum_{i=1}^n \log(1 - x_i).$$

2. Soit un vecteur aléatoire $(X_1, \dots, X_n)$ tel que pour $i, j \in \llbracket 1, n \rrbracket$, $i \neq j$, X_i indépendant de X_j et X_i suit la loi $\mathcal{N}(\mu, \sigma_i^2)$, où $\mu \in \mathbb{R}$ est inconnu et les $\sigma_i \in \mathbb{R}_+^*$ sont supposés connus, de réalisation $(x_1, \dots, x_n)$. L'estimateur du maximum de vraisemblance de μ associé à $(x_1, \dots, x_n)$ est donné par

- (a) $\hat{\mu}_n = (\sigma_1^{-2}x_1 + \dots + \sigma_n^{-2}x_n)/(\sigma_1^{-2} + \dots + \sigma_n^{-2})$, (c) $\sum_{i=1}^n \sigma_i^2(x_i - \hat{\mu}_n)^{-1} = 0$,
(b) $\hat{\mu}_n = (x_1 + \dots + x_n)/n$, (d) $\hat{\mu}_n = (\sigma_1^2x_1 + \dots + \sigma_n^2x_n)/(\sigma_1^2 + \dots + \sigma_n^2)$.

La bonne proposition est (a) car la log-vraisemblance vaut

$$\ell(\mu \mid x_1, \dots, x_n) = -1/2 \sum_{i=1}^n \{\sigma_i^{-2}(x_i - \mu)^2 + \log(2\pi\sigma_i^2)\}$$

Elle est deux fois dérivable et les points critiques sont obtenus par

$$\frac{\partial}{\partial \lambda} \ell(\lambda \mid x_1, \dots, x_n) = \sum_{i=1}^n \{\sigma_i^{-2}(x_i - \mu) = \sum_{i=1}^n \{\sigma_i^{-2}x_i + \mu \sum_{i=1}^n \{\sigma_i^{-2} = 0$$

Comme il s'agit d'une famille exponentielle minimale régulière, (a) est l'unique maximum de la vraisemblance.

3. Soient $X_1, \dots, X_n$ i.i.d. suivant une distribution discrète qui prend avec probabilité $1/3$ chacune des valeurs $\{\theta - 1, \theta, \theta + 1\}$, $\theta \in \mathbb{R}$. La statistique des observations extrêmes $S(X_1, \dots, X_n) = (X_{(1)}, X_{(n)})$ est

- (a) incomplète et non-exhaustive, (c) complète et exhaustive,
 (b) complète et non-exhaustive, (d) incomplète et exhaustive.

La bonne proposition est (d). La vraisemblance associée à $(x_1, \dots, x_n)$ vaut

$$L(\theta \mid x_1, \dots, x_n) = \prod_{i=1}^n \mathbb{I}_{\{\theta-1, \theta, \theta+1\}}(x_i)/3.$$

Si $x_{(n)} - x_{(1)} = 2$, $\theta = x_{(1)} + 1 = x_{(n)} - 1$; si $x_{(n)} - x_{(1)} = 1$, la vraisemblance est non-nulle et identique en $x_{(1)}$ et en $x_{(1)} + 1$, et si $x_{(n)} - x_{(1)} = 0$, la vraisemblance est non-nulle et identique en $x_{(1)} - 1$, en $x_{(1)}$ et en $x_{(1)} + 1$. La statistique $(X_{(1)}, X_{(n)})$ détermine donc la vraisemblance. Elle n'est pas complète puisque $(X_{(1)} - X_{(n)}) = (X_{(1)} - \theta) - (X_{(n)} - \theta)$ est libre (c.f. Théorème de Basu).

4. Soient $X_1, \dots, X_n$ i.i.d. de loi géométrique $\mathcal{G}(e^{-\theta_0})$, avec $\theta_0 \in \mathbb{R}_+^*$ inconnu, i.e., telle que pour $k \in \mathbb{N}$, $\mathbb{P}(X_1 = k) = e^{-\theta_0}(1 - e^{-\theta_0})^{k-1}$. L'information de Fisher sur θ_0 fournie par $\sum_{i=1}^n X_i$ vaut

- (a) $(1 - e^{-\theta_0})/n$, (b) $n/(e^{\theta_0} - 1)$, (c) $n/(1 - e^{-\theta_0})$, (d) $ne^{-\theta_0}/(1 - e^{-\theta_0})^2$.

La bonne proposition est (c). On peut écrire

$$\mathbb{P}(X_1 = k) = \frac{e^{-\theta_0}}{1 - e^{-\theta_0}} \exp\{\log(1 - e^{-\theta_0})k\}$$

ce qui implique que la loi géométrique définit une famille exponentielle de statistique naturelle $S(X) = X$. Pour une observation, la fonction score et la dérivée seconde valent

$$\frac{\partial}{\partial \theta_0} \ell(\theta_0 \mid x) = -1 + \frac{x-1}{1 - e^{-\theta_0}} \quad \text{et} \quad \frac{\partial^2}{\partial \theta_0^2} \ell(\theta_0 \mid x) = -\frac{(x-1)e^{-\theta_0}}{(1 - e^{-\theta_0})^2}$$

. Donc

$$-\mathbb{E}_{\theta_0} \left[\frac{\partial^2}{\partial \theta_0^2} \ell(\theta_0 \mid X) \right] = \frac{e^{-\theta_0}}{(1 - e^{-\theta_0})^2} \mathbb{E}_{\theta_0}[X - 1] = \frac{e^{-\theta_0}}{(1 - e^{-\theta_0})^2} (e^{\theta_0} - 1) = \frac{1}{1 - e^{-\theta_0}}.$$

5. Soient $X_1, \dots, X_n$ i.i.d. de loi normale $\mathcal{N}(1, \sigma^2)$, $\sigma \in \mathbb{R}_+^*$. Si $\theta = \sigma^2$ est le paramètre naturel, quelle est la bonne fonction de score ?

- (a) $\frac{1}{\theta^3} \sum_{i=1}^n X_i^2 - \frac{n}{2\theta^2}$, (b) $\frac{1}{\theta^{3/2}} \sum_{i=1}^n (X_i - 1)^2 - \frac{n}{2\theta}$, (c) $\frac{1}{2\theta^2} \sum_{i=1}^n (X_i - 1)^2 - \frac{n}{2\theta}$, (d) $\frac{n}{\theta^2} \overline{X_n}^2 - \frac{n}{2\theta}$.

La bonne proposition est (c). La fonction score vaut

$$\frac{\partial}{\partial \theta_0} \ell(\theta_0 \mid x_1, \dots, x_n) = \sum_{i=1}^n \frac{\partial}{\partial \theta_0} \log f(x_i \mid \theta) = \sum_{i=1}^n \frac{\partial}{\partial \theta_0} \left[\frac{(x_i - 1)^2}{2\theta} - \frac{\log(2\pi\theta)}{2} \right] = \sum_{i=1}^n \frac{(x_i - 1)^2}{2\theta^2} - \frac{n}{2\theta}.$$

6. Si $X_1, \dots, X_n$ sont i.i.d. de loi $\text{Beta}(\beta + 1, 1)$, $\beta \in \mathbb{R}_+^*$, un estimateur sans biais de $\beta/(1 + \beta)$ est donné par

- (a) $1 + (n + 1) [\sum_{i=1}^n \log(X_i)]^{-1}$, (c) $1 + (n - 1) [\sum_{i=1}^n \log(X_i)]^{-1}$,
 (b) $1 + n [\sum_{i=1}^n \log(X_i)]^{-1}$, (d) $1 + n^{-1} [\sum_{i=1}^n \log(X_i)]$.

La bonne proposition est (d). Pour une loi $\text{Beta}(\beta + 1, 1)$, de densité proportionnelle à x^β sur $(0, 1)$, la fonction $\log(x)^{-1}$ n'est pas intégrable en $x = 1$, ce qui élimine les propositions (a) à (c) (pour $n = 1$). Par

ailleurs, une intégration par partie conduit à

$$\int_0^1 \log(x) (\beta + 1) x^\beta dx = - \int_0^1 \frac{1}{x} x^{\beta+1} dx$$

donc

$$\mathbb{E}_\beta[\log X] = \frac{-1}{\beta + 1} \quad \text{et} \quad \mathbb{E}_\beta \left[1 + n^{-1} \sum_{i=1}^n \log(X_i) \right] = 1 - \frac{1}{\beta + 1} = \frac{\beta}{\beta + 1}.$$

7. Soit $(x_1, \dots, x_n)$ un échantillon *i.i.d.* de distribution F donnée par $X_1 \sim \mathcal{N}(\mu_0, 1/2)$. Si le modèle statistique imposé est $\{\mathcal{N}(\mu, 1) \mid \mu \in \mathbb{R}\}$, quelle affirmation sur la vraisemblance $L(\mu \mid x_1, \dots, x_n)$ du modèle est correcte ?

- (a) elle vaut $\prod_{i=1}^n [e^{-(x_i - \mu)^2} / \sqrt{2\pi}]$, (c) $\int_{\mathbb{R}} L(\mu \mid x_1, \dots, x_n) d\mu = 1$,
(b) $\int_{\mathbb{R}^n} L(\mu \mid x_1, \dots, x_n) dF(x_1, \dots, x_n) = 1$, (d) $\int_{\mathbb{R}^n} L(\mu \mid x_1, \dots, x_n) d(x_1, \dots, x_n) = 1$.

La bonne proposition est (d). Puisque la vraisemblance $L(\mu \mid x_1, \dots, x_n)$ (en tant que fonction de θ) est définie comme la densité de l'échantillon $(x_1, \dots, x_n)$ (en tant que fonction de cet échantillon), son intégrale (d) en $(x_1, \dots, x_n)$ vaut 1. Son espérance (b) ne vaut pas un, ni son intégrale en μ , (c), tandis qu'il manque un 2 dans le dénominateur de l'exponentielle en (a).

8. Soit $\mathbf{x}$ un vecteur de n observations supposées être des réalisations de la loi de Cauchy de paramètre $(\theta, 1)$, $\theta \in \mathbb{R}$. La médiane de l'échantillon est un estimateur de θ dont on souhaite estimer le biais par une approche bootstrap non-paramétrique. Quel code permet d'obtenir une réalisation bootstrap de la médiane ?

- (a) `median(rcauchy(n, median(x)))` (d) `median(pcauchy(n, median(x)))`
(b) `sample(x, replace = TRUE)` (e) `rcauchy(n, median(x))`
(c) `median(sample(x, replace = FALSE))` (f) `median(sample(x, replace = TRUE))`

La bonne proposition est (f). Une approche bootstrap non-paramétrique est associée à la mesure empirique, obtenue par tirage avec remise depuis $\mathbf{x}$, donc `sample(x, replace = TRUE)` est la bonne commande R pour produire un échantillon bootstrap. On prend ensuite la médiane.

9. Soit $\mathbf{x}$ un vecteur de n observations. Supposons que l'on travaille avec le modèle statistique associé à la loi Gamma $\mathcal{G}a(1, 2)$, quelle ligne de commande permet d'obtenir le premier quartile de cette loi ?

- (a) `rgamma(0.25, 1, 2)` (c) `qgamma(0.25, 1, 2)` (e) `quantile(x, 0.25)`
(b) `quantile(x, 0.25, 1, 2)` (d) `pgamma(0.25, 1, 2)`

La bonne proposition est (c). La fonction quantile d'une loi standard est obtenue avec le préfixe `q` attaché au radical de la loi, soit `verb!gamma!`. La fonction `verb!quantile!` retourne les quantiles empiriques sur un échantillon.

10. Pour $X_1, \dots, X_n$ *i.i.d.* de loi $\mathcal{N}(0, \lambda)$, une statistique libre de λ est

- (a) $X_2 X_3$, (b) $\mathbb{I}_{\bar{X}_n \geq 0}$, (c) $(\bar{X}_n)^2$, (d) X_2 / X_3 , (e) $-\bar{X}_n$.

Les bonnes propositions sont (b) et (d). Une variable $\mathcal{N}(0, \lambda)$ s'écrit $\lambda^{1/2} \epsilon$ avec $\epsilon \sim \mathcal{N}(0, 1)$. En remplaçant les X_i dans les propositions, (a), (c) et (e) dépendent de λ tandis qu'il s'élimine dans (b) et (d).

Exercice 2 La loi binomiale négative est une loi discrète dont la densité par rapport à la mesure de comptage sur $\mathbb{N}$ est

$$f(x | p) = \binom{x+r-1}{x} p^r (1-p)^x, \quad \text{avec } r \in \mathbb{R}_+^* \quad \text{et } p \in]0, 1[.$$

Dans cet exercice, on suppose que le paramètre r est connu et fixé, et le paramètre p est inconnu. On note $X_1, \dots, X_n$ des variables aléatoires *i.i.d.* suivant la loi binomiale négative de paramètre (r, p) et $x_1, \dots, x_n$ un n -échantillon pour cette loi.

1. Montrer que cette distribution constitue une famille exponentielle. Donner un paramètre naturel, une statistique naturelle, et démontrer que la représentation est minimale et régulière.

La densité s'écrit

$$f(x | p) = c(p)h(x) \exp[\log(1-p)x], \quad \text{avec } \begin{cases} c(p) = p^r > 0, & \text{pour } p \in]0; 1[, \\ h(x) = \binom{x+r-1}{x} > 0, & \text{pour } x \in \mathbb{N}. \end{cases}$$

C'est donc une famille exponentielle avec statistique naturelle $T(x) = x$ et un paramètre naturel $\theta = \eta(p) = \log(1-p)$. La représentation est minimale car T est une statistique de dimension 1 non constante *p.s.* L'espace naturel des paramètres est $\Theta = \{\theta \in \mathbb{R} \mid a(\theta) > 0\}$ où $a(\theta) = c(\eta^{-1}(\theta))$ est la fonction de partition de la forme canonique. On a donc $\theta = \{\theta \in \mathbb{R} \mid e^\theta < 1\} = \mathbb{R}_-^*$ est un ouvert et la représentation est régulière.

2. Montrer que la fonction génératrice des moments de cette loi vaut

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

On utilise les propriétés de la famille exponentielle $\theta = \log(1-p)$. Comme $T(X_1) = X_1$, la fonction génératrice des moments de la loi binomiale négative est, pour t tel que $\theta + t \in \Theta$ (*i.e.*, $\log(1-p) + t < 0$),

$$M_{T(X)}(t) = M_X(t) = \frac{a(\theta)}{a(\theta+t)} = \left[\frac{1 - \exp(\theta)}{1 - \exp(\theta+t)} \right]^r = \left[\frac{p}{1 - e^t(1-p)} \right]^r.$$

3. Montrer que l'estimateur suivant converge en probabilité vers p

$$\hat{p}_n = \frac{nr}{nr + \sum_{k=1}^n X_k}.$$

Comme $T(X_1) = X_1$, le résultat sur l'espérance de la statistique naturelle d'une famille exponentielle s'écrit

$$\mathbb{E}[X_1] = \mathbb{E}[T(X_1)] = -\frac{\partial}{\partial \theta} \log a(\theta) = \frac{re^\theta}{1-e^\theta} = \frac{r(1-p)}{p}.$$

On en déduit que $X_1, \dots, X_n$ est une suite de variables aléatoires *i.i.d.* intégrables et la loi faible des grands nombres donne

$$\frac{1}{n} \sum_{k=1}^n X_k \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} \frac{r(1-p)}{p}.$$

La fonction $g : x \mapsto x/(x+r)$ est continue au voisinage de $r(1-p)/p$ pour $p \in]0; 1[$, donc

$$\hat{p}_n = g\left(\frac{1}{n} \sum_{k=1}^n X_k\right) \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} g\left(\frac{r(1-p)}{p}\right) = p.$$

4. Trouver une suite c_n indépendante de p telle que

$$c_n(\hat{p}_n - p) \xrightarrow[n \rightarrow +\infty]{d} \mathcal{N}(0, 1).$$

Comme à la question précédente, on peut utiliser les propriétés des familles exponentielles pour calculer $\mathbb{V}\text{ar}[X_1]$

$$\mathbb{V}\text{ar}[X_1] = \mathbb{V}\text{ar}[T(X_1)] = -\frac{\partial^2}{\partial \theta^2} \log a(\theta) = \frac{re^\theta}{(1-e^\theta)^2} = \frac{r(1-p)}{p^2}.$$

Donc, $X_1, \dots, X_n$ sont *i.i.d.* de carré intégrable et le théorème central limite donne

$$\sqrt{n} \left(\frac{1}{n} \sum_{k=1}^n X_k - \frac{r(1-p)}{p} \right) \xrightarrow[n \rightarrow +\infty]{d} \mathcal{N}\left(0, \frac{r(1-p)}{p^2}\right).$$

La fonction $g : x \mapsto x/(x+r)$ est dérivable au voisinage de $r(1-p)/p$ et $g'(r(1-p)/p) = -p^2/r \neq 0$ pour $p \in]0; 1[$. La méthode delta conduit à

$$\sqrt{n} \left[g\left(\frac{1}{n} \sum_{k=1}^n X_k\right) - g\left(\frac{r(1-p)}{p}\right) \right] = \sqrt{n}(\hat{p}_n - p) \xrightarrow[n \rightarrow +\infty]{d} \mathcal{N}\left(0, \frac{r(1-p)}{p^2} \frac{p^4}{r^2}\right) \equiv \mathcal{N}\left(0, \frac{p^2(1-p)}{r}\right).$$

La fonction $x \mapsto \sqrt{r}/(x\sqrt{1-x})$ étant continue sur $]0; 1[$ et $\hat{p}_n \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} p$, le théorème de Slutsky donne

$$\frac{\sqrt{r}\sqrt{n}}{\hat{p}_n\sqrt{1-\hat{p}_n}}(\hat{p}_n - p) \xrightarrow[n \rightarrow +\infty]{d} \frac{\sqrt{r}}{p\sqrt{1-p}} \mathcal{N}\left(0, \frac{p^2(1-p)}{r}\right) \equiv \mathcal{N}(0, 1) \quad \text{et donc} \quad c_n = \frac{\sqrt{rn}}{\hat{p}_n\sqrt{1-\hat{p}_n}}.$$

5. Donner la vraisemblance de p associée à un n -échantillon $(x_1, \dots, x_n)$ et montrer que l'estimateur du maximum de vraisemblance de p est égal à $\hat{p}_n$.

Travaillant avec un échantillon *i.i.d.*, la vraisemblance de p associée à $(x_1, \dots, x_n)$ est

$$L(p \mid x_1, \dots, x_n) = \prod_{k=1}^n f(x_1, \dots, x_n \mid \theta) = p^{nr} (1-p)^{\sum_{k=1}^n x_k} \prod_{k=1}^n \binom{x_k + r - 1}{x_k}.$$

La fonction $p \mapsto \log L(p \mid x_1, \dots, x_n)$ est deux fois dérivable et ses points critiques sont solutions de

$$\frac{\partial}{\partial p} \log L(p \mid x_1, \dots, x_n) = \frac{nr}{p} - \frac{1}{1-p} \sum_{k=1}^n x_k = \frac{1}{p(1-p)} \left[nr - p \left(nr + \sum_{k=1}^n x_k \right) \right] = 0.$$

On en déduit que $\hat{p}_n$ est un point critique. Comme la loi binomiale négative formant une famille exponentielle minimale régulière, on en déduit que $\hat{p}_n$ est l'unique estimateur du maximum de vraisemblance.

6. Montrer que l'erreur quadratique moyenne pour $\hat{p}_n$ est **strictement** minorée par $\text{Var}[\hat{p}_n]$, *i.e.*, $\mathbb{E}_p [(\hat{p}_n - p)^2] > \text{Var}[\hat{p}_n]$. Sachant que l'on dispose d'un vecteur d'observations $\mathbf{x}$, écrire un code R donnant une estimation de cette erreur quadratique moyenne utilisant k itérations d'un bootstrap non-paramétrique.

On a $\mathbb{E}[(\hat{p}_n - p)^2] = \text{Var}[\hat{p}_n] + (\mathbb{E}[\hat{p}_n] - p)^2$. La fonction $g : x \mapsto r/(r+x)$ est strictement convexe sur $\mathbb{R}_+^*$

$$\mathbb{E}[\hat{p}_n] = \mathbb{E} \left[g \left(\frac{1}{n} \sum_{k=1}^n X_k \right) \right] > g \left(\mathbb{E} \left[\frac{1}{n} \sum_{k=1}^n X_k \right] \right) = g(\mathbb{E}[X_1]) = p.$$

Donc $\mathbb{E}[\hat{p}_n] - p < 0$ et donc $\mathbb{E}_p [(\hat{p}_n - p)^2] > \text{Var}[\hat{p}_n]$.

```
a <- numeric(k)
for (i in 1:k)
  a[i] <- mean(sample(x, replace = TRUE))
mse <- mean((r/(r + a) - r / (r + mean(x)))**2)
```

7. Montrer que $S(X_1, \dots, X_n) = \sum_{i=1}^n X_i$ suit une loi binomiale négative de paramètre (nr, p) .

La fonction génératrice d'une somme de variables aléatoires *i.i.d.* est le produit des fonctions génératrices de ces variables aléatoires. On a donc :

$$M_{S(X_1, \dots, X_n)}(t) = M_X(t)^n = \left[\frac{1}{1 - e^t(1-p)} \right]^{nr}$$

qui est la fonction génératrice des moments d'une loi binomiale négative de paramètre (nr, p) .

8. Montrer que $\hat{\theta}_n = (nr - 1)/[S(X_1, \dots, X_n) + nr - 1]$ est un estimateur sans biais de p .

$S(X_1, \dots, X_n) = \sum_{i=1}^n X_i$ suit une loi binomiale négative de paramètre (nr, p) . Le théorème de transfert conduit à

$$\begin{aligned}\mathbb{E}[\hat{\theta}_n] &= \sum_{k=0}^{\infty} \frac{nr-1}{k+nr-1} \binom{k}{k+nr-1} p^{nr}(1-p)^k = p \sum_{k=0}^{\infty} \frac{nr-1}{k+nr-1} \frac{(k+nr-1)!}{k!(nr-1)!} p^{nr-1}(1-p)^k \\ &= p \sum_{k=0}^{\infty} \binom{k}{k+nr-2} p^{nr-1}(1-p)^k\end{aligned}$$

La dernière somme vaut 1 car il s'agit de l'intégrale (par rapport à la mesure de comptage sur $\mathbb{N}$) de la densité d'une loi binomiale négative de paramètre $(nr-1, p)$. L'estimateur est donc sans biais.

9. Montrer que $\hat{\theta}_n$ est de variance minimale parmi les estimateurs sans biais de p .

Indication. Si une famille exponentielle paramétrée par $\theta \in \Theta$ et de statistique naturelle $T(X)$ est régulière alors, pour un échantillon *i.i.d.* $(X_1, \dots, X_n)$ de cette famille, $\sum_{k=1}^n T(X_i)$ est une statistique complète de θ .

La loi binomiale négative formant une famille exponentielle régulière de statistique naturelle $T(X) = X$, on en déduit que $S(X_1, \dots, X_n)$ est une statistique complète (indication) et exhaustive (théorème de factorisation) de θ . $\hat{\theta}_n$ étant un estimateur sans biais, le théorème de Lehmann-Scheffé donne que $\mathbb{E}[\hat{\theta}_n | S(X_1, \dots, X_n)]$ est l'unique estimateur sans biais de variance minimale. Or par construction $\hat{\theta}_n$ est une fonction mesurable de $S(X_1, \dots, X_n)$ et on a donc $\mathbb{E}[\hat{\theta}_n | S(X_1, \dots, X_n)] = \hat{\theta}_n$.

Exercice 3 Soit $\mathbf{X} = (X_1, \dots, X_n)$ un échantillon *i.i.d.* de loi $\mathcal{N}(\mu, \sigma^2)$, avec $\mu \in \mathbb{R}$ et $\sigma \in \mathbb{R}_+^*$. On rappelle que la loi de Student $\mathcal{T}(n-1)$ est historiquement (*c.f.* Student, 1908) définie comme la loi du rapport

$$T(\mathbf{X}) \stackrel{\text{def}}{=} \sqrt{n} \frac{(\bar{X}_n - \mu)}{S_n}, \quad \text{où} \quad \bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i \quad \text{et} \quad S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$$

sont respectivement la moyenne et la variance empiriques associées à l'échantillon.

1. En utilisant (obligatoirement) le théorème de Basu, montrer que si $(X_1, \dots, X_n)$ est un échantillon *i.i.d.* de loi $\mathcal{N}(\mu, 1)$, alors $\bar{X}_n$ et S_n sont indépendantes. En déduire que ce résultat reste vrai pour un échantillon *i.i.d.* de loi $\mathcal{N}(\mu, \sigma)$.

Indication. On pourra utiliser l'indication de la question 9. de l'Exercice 2.

A $\sigma = 1$ fixé, la statistique S_n est libre puisque $X_i - \bar{X}_n = (X_i - \mu) - (\bar{X}_n - \mu)$. Par ailleurs, puisque $\mathcal{N}(\mu, 1)$ définit une famille exponentielle régulière de statistique naturelle $\bar{X}_n$, cette statistique est exhaustive minimale et complète. Le théorème de Basu implique donc que $\bar{X}_n$ et S_n sont indépendantes. Passer au cas de la loi $\mathcal{N}(\mu, \sigma)$ revient à remplacer $\bar{X}_n$ et S_n par $\sigma \bar{X}_n$ et σS_n . Si les premiers sont indépendants, les seconds le sont aussi par transformation mesurable.

2. Montrer, en utilisant une transformation affine des X_i , $i \in \llbracket 1, n \rrbracket$, que quelque soit $(\mu, \sigma) \in \mathbb{R} \times \mathbb{R}_+^*$, la statistique $T(\mathbf{X})$ a une seule et même loi de probabilité. Peut-on en déduire que $T(\mathbf{X})$ est une statistique libre ?

On a

$$T(\mathbf{X}) = \sqrt{n} \frac{(\bar{X}_n - \mu)}{S_n} = \sqrt{n} \frac{(\bar{X}_n - \mu)/\sigma}{S_n/\sigma} \quad \text{et} \quad \sqrt{n} \frac{\bar{X}_n - \mu}{\sigma} \sim \mathcal{N}(0, 1), \quad (n-1) \frac{S_n^2}{\sigma^2} \sim \chi^2(n-1).$$

Donc le rapport a une loi qui ne dépend ni de μ ni de σ . Par contre, $T(\mathbf{X})$ n'est pas libre car ce n'est pas une statistique, comme elle est aussi fonction de μ .

3. Montrer (ou admettre) que la loi de $T(\mathbf{X})$ est celle du rapport A_1/A_2 où A_1 et A_2 sont deux variables aléatoires indépendantes telles que A_1 soit de loi $\mathcal{N}(0, 1)$ et $A_2 = \sqrt{B_2/(n-1)}$ avec B_2 de loi $\chi^2(n-1)$. En déduire que, pour $\nu \in \mathbb{N}^*$, la densité de la loi de Student $\mathcal{T}(\nu)$ s'exprime, pour $t \in \mathbb{R}$, comme

$$f(t; \nu) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi} \Gamma(\nu/2)} (1 + t^2/\nu)^{-\frac{\nu+1}{2}}, \quad \text{où } \Gamma(\cdot) \text{ désigne la fonction Gamma.}$$

À partir de la décomposition précédente, on pose $A_1 = \sqrt{n}(\bar{X}_n - \mu)/\sigma$ et $A_2 = S_n/\sigma$. On a alors que A_1 est de loi $\mathcal{N}(0, 1)$ et

$$(n-1)A_2^2 = \sum_{i=1}^n \left(\frac{X_i - \bar{X}_n}{\sigma} \right)^2 = \sum_{i=1}^n \overbrace{\left(\frac{X_i - \mu}{\sigma} \right)^2}^{A_i \sim \chi^2(1)} - n \overbrace{\left(\frac{\bar{X}_n - \mu}{\sigma} \right)^2}^{A_1^2 \sim \chi^2(1)}$$

est indépendant de A_1 . Cela implique que $(n-1)A_2^2 \sim \chi^2(n-1)$. (Ce résultat s'appelle le théorème de Cochran.) La loi de $T(\mathbf{X})$ conditionnellement à B_2 est une loi Normale $\mathcal{N}(0, (n-1)/B_2)$. Par conséquent la densité de la loi jointe de $(B_2, T(\mathbf{X}))$ est donnée par

$$\frac{b_2^{n-1/2-1}}{2^{n-1/2} \Gamma(\frac{n-1}{2})} e^{-b_2/2} \times \frac{b_2^{1/2}}{\sqrt{2\pi(n-1)}} \exp\left(-\frac{b_2 t^2}{2(n-1)}\right) = \frac{b_2^{n-1/2-1/2}}{2^{n-1/2} \Gamma(\frac{n-1}{2}) \sqrt{2\pi(n-1)}} \exp\left(-\frac{b_2}{2} - \frac{b_2 t^2}{2(n-1)}\right). \quad (1)$$

La loi marginale de $T(\mathbf{X})$ s'obtient en intégrant (1) en b_2 :

$$\int_0^\infty b_2^{n-1/2-1/2} \exp\left(-\frac{b_2}{2} - \frac{b_2 t^2}{2(n-1)}\right) db_2 = \Gamma\left(\frac{n}{2}\right) 2^{n/2} \left[1 + \frac{t^2}{(n-1)}\right]^{-n/2}$$

ce qui conduit à l'expression donnée dans la question.

On considère alors l'extension de la loi ci-dessus en la famille de lois indexée par (μ, τ) , de densité

$$g(t; \nu, \mu, \tau) = \frac{\Gamma(\frac{\nu+1}{2})}{\tau \sqrt{\nu\pi} \Gamma(\nu/2)} \left(1 + \frac{(t - \mu)^2}{\nu\tau^2}\right)^{-\frac{\nu+1}{2}} \quad t \in \mathbb{R} \quad (2)$$

Cette nouvelle famille, notée $\mathcal{T}(\nu, \mu, \tau^2)$, est toujours appelée loi de Student. **Dans toute la suite de l'exercice, le paramètre ν , appelé degré de liberté de la loi, est supposé fixe et connu.**

4. Montrer que, si T est de loi $\mathcal{T}(\nu, \mu, \tau^2)$, alors $(T - \mu)/\tau$ est de loi $\mathcal{T}(\nu)$.

Il suffit de considérer le changement de variable $z = (t - \mu)/\tau$ qui réalise un $\mathcal{C}^1$ -difféomorphisme de $\mathbb{R}$ dans $\mathbb{R}$, de Jacobien $1/\tau$, qui conduit à une densité

$$\tilde{g}(z; \nu, \mu, \tau) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi} \Gamma(\nu/2)} \left(1 + \frac{z^2}{\nu}\right)^{-\frac{\nu+1}{2}}$$

qui est indépendante de (μ, τ) et redonne la densité de la loi $\mathcal{T}(\nu)$.

5. En déduire que si l'échantillon *i.i.d.* $\mathbf{X} = (X_1, \dots, X_n)$ est de loi $\mathcal{N}(\mu, \tau^2)$, alors le rapport $\sqrt{n} \bar{X}_n / S_n$ **ne suit pas** une loi $\mathcal{T}(n-1, \mu, \tau^2)$ (mais une autre loi appelée Student décentrée).

Le rapport

$$\frac{1}{\tau} \left(\sqrt{n} \bar{X}_n / S_n - \mu \right) = \frac{\sqrt{n} \bar{X}_n / \tau}{S_n / \tau} - \frac{\mu}{\tau}$$

dépend de τ si $\mu \neq 0$, donc ne peut suivre la loi $\mathcal{T}(\nu)$.

6. Montrer que la famille associée aux densités (2), de paramètre (μ, τ) n'est pas une famille exponentielle. Est-ce que l'estimateur du maximum de vraisemblance de (μ, τ) peut être une statistique exhaustive pour n'importe quelle taille $m \in \mathbb{N}^*$ d'un échantillon $(T_1, \dots, T_m)$ de $\mathcal{T}(\nu, \mu, \tau^2)$?

Indication. On ne calculera pas l'estimateur du maximum de vraisemblance de (μ, τ) .

Quand on considère la densité (2), l'expression $\log [1 + (t - \mu)^2 / (\nu \tau^2)]$ ne peut se décomposer en un produit d'une fonction de (μ, τ) et d'une fonction de t . Une autre façon de voir cette impossibilité est de considérer la vraisemblance, puissance d'un polynôme de degré $2n$ qui ne peut se paramétrer par un vecteur de taille constante en n . Par conséquent, comme la famille n'est pas de type exponentielle, l'estimateur du maximum de vraisemblance de (μ, τ) , qui est de dimension deux, ne peut pas être une statistique exhaustive.

7. En admettant que l'information de Fisher sur (μ, τ) apportée par T de loi $\mathcal{T}(\nu, \mu, \tau^2)$ vaut

$$\mathfrak{I}_\nu(\mu, \tau) = \begin{bmatrix} \frac{\nu+1}{(\nu+3)\tau^2} & 0 \\ 0 & \frac{\nu}{2(\nu+3)\tau^4} \end{bmatrix},$$

montrer que X de loi $\mathcal{N}(\mu, \tau^2)$ est plus informatif que T sur (μ, τ) .

L'information de Fisher sur (μ, τ) apportée par X vaut

$$\mathfrak{I}_\infty(\mu, \tau) = \begin{bmatrix} \frac{1}{\tau^2} & 0 \\ 0 & \frac{1}{2\tau^4} \end{bmatrix},$$

(ce qui correspond au cas $\nu = +\infty$). La différence des matrices, $\mathfrak{I}_\infty(\mu, \tau) - \mathfrak{I}_\nu(\mu, \tau)$ est définie positive, donc X est plus informative que T .

8. Donner un argument calculatoire pour caractériser la complexité du calcul de l'estimateur du maximum de vraisemblance de (μ, τ) fondé sur un m -échantillon $(T_1, \dots, T_m)$ de $\mathcal{T}(\nu, \mu, \tau^2)$, $m \in \mathbb{N}^*$.

La vraisemblance $L(\mu, \tau \mid t_1, \dots, t_m)$ est une puissance du polynôme de degré $2m \prod_{i=1}^m (1 + (t_i - \mu)^2 / (\nu \tau^2))$ à une puissance de τ près. A τ fixé, obtenir l'estimateur du maximum de vraisemblance de μ exige donc de déterminer tous les zéros d'un polynôme de degré $2m - 1$ afin de déterminer le minimum global.

9. Soit $T = \mu + \tau \sqrt{\nu} A_1 / \sqrt{A_2}$, avec A_1 de loi $\mathcal{N}(0, 1)$, indépendant de A_2 , et A_2 de loi $\chi^2(\nu)$. Donner

- (a) la densité de la loi conditionnelle de T conditionnellement à A_2 comme une densité normale ;
- (b) la densité de la loi jointe de (T, A_2) , $f^c(t, a_2; \nu, \mu, \tau)$;
- (c) la densité de la loi conditionnelle de A_2 conditionnellement à T , $f^\ell(a_2; t, \nu, \mu, \tau)$ comme une densité Gamma.

Indication. On ne redémontrera pas que T est de loi $\mathcal{T}(\nu, \mu, \tau^2)$.

- (a) Déjà déterminée dans la question 3 : $T \mid A_2 = a_2 \sim \mathcal{N}(\mu, \tau^2 \nu / a_2)$
- (b) Déjà déterminée dans la question 3 :

$$f^c(t, a_2; \nu, \mu, \tau) = \frac{1}{2^{\nu/2} \Gamma(\nu/2)} a_2^{\nu/2-1} \exp\left(-\frac{a_2}{2}\right) \times \frac{a_2^{1/2}}{\sqrt{2\pi\nu\tau^2}} \exp\left(-\frac{a_2(t-\mu)^2}{2\nu\tau^2}\right).$$

- (c) La densité de la loi conditionnelle de A_2 conditionnellement à T est alors donnée par $f^\ell(a_2; t, \nu, \mu, \tau) \propto$

$f^c(t, a_2; \nu, \mu, \tau)$, en tant que fonction de a_2 , donc proportionnelle à

$$a_2^{\nu/2-1+1/2} \exp\left(-\frac{a_2}{2} - \frac{a_2(t-\mu)^2}{2\nu\tau^2}\right)$$

qui correspond à une loi Gamma de paramètres $(\nu+1)/2$ et $[1 + (t-\mu)^2/(\nu\tau^2)]/2$.

Admis. Si A_2 est de loi $\mathcal{G}(a, b)$, alors $\mathbb{E}_{a,b}[A_2] = a/b$ et $\mathbb{E}_{a,b}[\log A_2] = \psi(a) - \log b$, où $\psi(\cdot)$ est une fonction spéciale, dite *digamma*.

10. En vue d'implémenter l'algorithme EM pour trouver l'estimateur du maximum de vraisemblance, associé à un m -échantillon $(T_1, \dots, T_m)$ de $\mathcal{T}(\nu, \mu, \tau^2)$, $m \in \mathbb{N}^*$, montrer que, pour $i \in \llbracket 1, m \rrbracket$,

$$\begin{aligned} Q_i(\mu, \tau; t_i, \mu^0, \tau^0) &\stackrel{\text{def}}{=} \mathbb{E}_{\mu^0, \tau^0} [\log f^c(t, A_{2i}; \nu, \mu, \tau) \mid T_i = t_i] \\ &= -\frac{1}{2} \log[2\pi] - \log \Gamma(\nu/2) - \frac{\nu}{2} \log(2) - \log(\tau) + \frac{\nu-1}{2} \mathbb{E}_{\mu^0, \tau^0} [\log(A_{2i}) \mid T_i = t_i] \\ &\quad - \left\{ \frac{1}{2} + \frac{(t-\mu)^2}{2\nu\tau^2} \right\} \mathbb{E}_{\mu^0, \tau^0} [A_{2i} \mid T_i = t_i] \end{aligned}$$

où l'espérance conditionnelle est calculée sous la densité $f^\ell(a_{2i}; t_i, \nu, \mu^0, \tau^0)$ (ce qui correspond à l'étape E).

$$\begin{aligned} &\mathbb{E}_{\mu^0, \tau^0} [\log f^c(t, A_{2i}; \nu, \mu, \tau) \mid T_i = t_i] \\ &= \mathbb{E}_{\mu^0, \tau^0} \left[\log \left\{ \frac{A_{2i}^{\nu+1/2-1}}{2^{\nu+1/2} \Gamma(\nu/2)} \frac{1}{\sqrt{\nu\tau}} \exp \left[-\frac{A_{2i}}{2} - \frac{A_{2i}(t-\mu)^2}{2\nu\tau^2} \right] \right\} \mid T_i = t_i \right] \\ &= -\frac{\log(2\pi)}{2} - \log \Gamma(\nu/2) - \frac{\log(\nu)}{2} - \log(\tau) + \frac{\nu-1}{2} \mathbb{E}_{\mu^0, \tau^0} [\log(A_{2i}) \mid T_i = t_i] - \left[\frac{1}{2} + \frac{(t-\mu)^2}{2\nu\tau^2} \right] \mathbb{E}_{\mu^0, \tau^0} [A_{2i} \mid T_i = t_i] \end{aligned}$$

Les seuls termes qui dépendent de (μ, τ) et comptent pour l'optimisation sont

$$-\log(\tau) - \left\{ \frac{1}{2} + \frac{(t-\mu)^2}{2\nu\tau^2} \right\} \mathbb{E}_{\mu^0, \tau^0} [A_{2i} \mid T_i = t_i]$$

11. Montrer que la solution de $(\mu^1, \tau^1) \stackrel{\text{def}}{=} \arg \max_{\mu, \tau} \sum_{i=1}^m Q_i(\mu, \tau; t_i, \mu^0, \tau^0)$ est

$$\mu^1 = \frac{\sum_{i=1}^m t_i \mathbb{E}_{\mu^0, \tau^0} [A_{2i} \mid T_i = t_i]}{\sum_{i=1}^m \mathbb{E}_{\mu^0, \tau^0} [A_{2i} \mid T_i = t_i]} \quad \text{et} \quad \tau^1 = \frac{1}{m\nu} \sum_{i=1}^m (t_i - \mu)^2 \mathbb{E}_{\mu^0, \tau^0} [A_{2i} \mid T_i = t_i]$$

(ce qui correspond à l'étape M).

Minimiser

$$m \log(\tau) + \sum_{i=1}^m \left\{ \frac{1}{2} + \frac{(t_i - \mu)^2}{2\nu\tau^2} \right\} \mathbb{E}_{\mu^0, \tau^0} [A_{2i} \mid T_i = t_i]$$

en (μ, τ) conduit aux équations

$$\sum_{i=1}^m (t_i - \mu) \mathbb{E}_{\mu^0, \tau^0} [A_{2i} \mid T_i = t_i] = 0 \quad \text{et} \quad \frac{m}{\tau} - \sum_{i=1}^m \frac{(t_i - \mu)^2}{\nu\tau^3} \mathbb{E}_{\mu^0, \tau^0} [A_{2i} \mid T_i = t_i] = 0$$

ce qui conduit bien aux solutions proposées (en vérifiant le second ordre).