



Cursus Intégré  
2004-2005

Rappels de statistique mathématique  
*Réponses question par question des travaux  
dirigés n° 2*

Guillaume Lacôte  
Bureau **E03**

✉ [Guillaume.Lacote@ensae.fr](mailto:Guillaume.Lacote@ensae.fr)

☞ <http://ensae.no-ip.com/SE222/>

**Exercice corrigé 1**

On considère  $n$  systèmes dont les durées de vie  $X_1, \dots, X_n$  suivent indépendamment une même loi de densité  $f$ . On observe uniquement les durées de vie  $Y_1, \dots, Y_r$  des  $r$  premiers systèmes tombant en panne.

☞ Q1 Ecrire la loi du  $r$ -uplet  $(Y_1, \dots, Y_r)$ , puis celle de la variable  $Y_r$ .

Loi de  $(Y_1, \dots, Y_r)$  : il existe au moins deux méthodes.

Méthode rapide mais pas naturelle :

L'idée ici est de considérer tous les réordonnements de  $r$  variables parmi  $n$  telles que les  $r$  premières soient croissantes et plus petites que les  $n - r$  suivantes.

Plus précisément, soit  $\mathcal{A}_n^r$  l'ensemble des parties à  $r \leq n$  éléments d'un ensemble à  $n \geq 1$  éléments.

Alors pour  $y_1 < \dots < y_r$  et  $dy_1, \dots, dy_r$ , avec  $0 < dy_i < \frac{y_{i+1} - y_i}{2}$

$$\begin{aligned} & \mathbb{P}(Y_1 \in [y_1, y_1 + dy_1[, \dots, Y_r \in [y_r, y_r + dy_r]) \\ &= \sum_{\pi \in \mathcal{A}_n^r} \mathbb{P} \left( \begin{array}{l} X_{\pi_1} \in [y_1, y_1 + dy_1[ \wedge \dots \wedge X_{\pi_r} \in [y_r, y_r + dy_r[ \\ \wedge X_{\pi_1} < \dots < X_{\pi_r} \wedge \forall i \notin \pi, X_i > y_r + dy_r \end{array} \right) \\ & \quad \text{où } \pi_k \text{ est le } k\text{-ième élément de la partie } \pi \text{ à } r \text{ éléments parmi } \llbracket 1, n \rrbracket \\ &= |\mathcal{A}_n^r| \times \begin{cases} \mathbb{P}(Z_1 \in [y_1, y_1 + dy_1]) \cdots \mathbb{P}(Z_r \in [y_r, y_r + dy_r]) \\ \times \mathbb{1}_{y_1 < \dots < y_r} \\ \times \mathbb{P}(Z_{r+1} > y_r + dy_r) \cdots \mathbb{P}(Z_n > y_r + dy_r) \end{cases} \\ & \quad \text{car les } [y_i, y_i + dy_i[ \text{ sont disjoints, et les } Z_i \text{ sont indépendantes car les } X_i \text{ le sont} \\ &= \frac{n!}{(n-r)!} \times \begin{cases} (F(y_1 + dy_1) - F(y_1)) \cdots (F(y_r + dy_r) - F(y_r)) \\ \times \mathbb{1}_{y_1 < \dots < y_r} \\ \times (1 - F(y_r + dy_r)) \cdots (1 - F(y_r + dy_r)) \end{cases} \end{aligned}$$

En divisant et en faisant successivement tendre  $dy_r \rightarrow 0^+, \dots, dy_1 \rightarrow 0^+$  il vient <sup>1</sup>

<sup>1</sup>Précisément :

$$\frac{1}{dy_1} \mathbb{P}(Y_1 \in [y_1, y_1 + dy_1[, \dots, Y_r \in [y_r, y_r + dy_r]) \xrightarrow{dy_1 \rightarrow 0^+} \frac{n!}{(n-r)!} \times \begin{cases} f(y_1) (F(y_2 + dy_2) - F(y_2)) \cdots (F(y_r + dy_r) - F(y_r)) \\ \times (1 - F(y_r + dy_r))^{n-r} \end{cases}$$

puis

$$\frac{1}{dy_1 \cdots dy_{r-1}} \mathbb{P}(Y_1 \in [y_1, y_1 + dy_1[, \dots, Y_r \in [y_r, y_r + dy_r]) \xrightarrow{dy_1 \rightarrow 0^+, \dots, dy_{r-1} \rightarrow 0^+} \frac{n!}{(n-r)!} \begin{cases} f(y_1) \cdots f(y_{r-1}) (F(y_r + dy_r) - F(y_r)) \\ \times (1 - F(y_r + dy_r))^{n-r} \end{cases}$$

$$\forall (y_1, \dots, y_r) \in \mathbb{R}^r, l(y_1, \dots, y_r) = \frac{n!}{(n-r)!} \mathbb{1}_{y_1 < \dots < y_r} f(y_1) \dots f(y_r) (1 - F(y_r))^{n-r}$$

Méthode simple mais longue :

L'idée est ici de calculer d'abord la loi de  $(Y_1, \dots, Y_n)$ , puis d'en tirer celle de  $(Y_1, \dots, Y_r)$  au moyen de la relation  $f_X(x) = \int_{\mathcal{Y}} f_{X,Y}(x, y) dy$ .

Donnons-nous donc  $\mathcal{A}$  mesurable quelconque.

Alors

$$((Y_1, \dots, Y_n) \in \mathcal{A}) \Leftrightarrow (\exists \sigma / (X_{\sigma(1)}, \dots, X_{\sigma(n)} \in \mathcal{A} \text{ et } X_{\sigma(1)} \leq \dots \leq X_{\sigma(n)})$$

et donc presque sûrement

$$((Y_1, \dots, Y_n) \in \mathcal{A}) \Leftrightarrow (\exists \sigma / (X_{\sigma(1)}, \dots, X_{\sigma(n)} \in \mathcal{A} \text{ et } X_{\sigma(1)} < \dots < X_{\sigma(n)})$$

de sorte que

$$\mathbb{P}((Y_1, \dots, Y_n) \in \mathcal{A}) = \sum_{\sigma \in \mathcal{S}_n} \mathbb{P}((X_{\sigma(1)}, \dots, X_{\sigma(n)}) \in \mathcal{A} \wedge X_{\sigma(1)} < \dots < X_{\sigma(n)})$$

où  $\sigma \in \mathcal{S}_n$  désigne une permutation des  $n$  indices telle que le  $n$ -uplet soit ordonné en plus d'être dans  $\mathcal{A}$ . Or pour toute  $\sigma \in \mathcal{S}_n$  on a

$$\begin{aligned} & \mathbb{P}((X_{\sigma(1)}, \dots, X_{\sigma(n)}) \in \mathcal{A} \wedge X_{\sigma(1)} < \dots < X_{\sigma(n)}) \\ &= \int_{\mathbb{R}^n} \mathbb{1}_{(y_1, \dots, y_n) \in \mathcal{A}} \mathbb{1}_{y_1 < \dots < y_n} f_{X_{\sigma(1)}}(y_1) \dots f_{X_{\sigma(n)}}(y_n) dy_1 \dots dy_n \\ &= \int_{\mathcal{A}} \mathbb{1}_{y_1 < \dots < y_n} f(y_1) \dots f(y_n) dy_1 \dots dy_n \quad \text{car les } (X_i)_i \text{ sont i.i.d} \end{aligned}$$

et donc comme  $|\mathcal{S}_n| = n!$

$$\mathbb{P}((Y_1, \dots, Y_n) \in \mathcal{A}) = n! \int_{\mathcal{A}} \mathbb{1}_{y_1 < \dots < y_n} f(y_1) \dots f(y_n) dy_1 \dots dy_n$$

On en conclut que

$$f_{Y_1, \dots, Y_n}(y_1, \dots, y_n) = n! \mathbb{1}_{y_1 < \dots < y_n} f(y_1) \dots f(y_n)$$

et finalement

$$\frac{1}{dy_1 \dots dy_r} \mathbb{P}(Y_1 \in [y_1, y_1 + dy_1[, \dots, Y_r \in [y_r, y_r + dy_r]) \xrightarrow{dy_1 \rightarrow 0^+; \dots; dy_r \rightarrow 0^+} \frac{n!}{(n-r)!} f(y_1) \dots f(y_r) (1 - F(y_r + dy_r))^{n-r}$$

On a alors

$$\begin{aligned}
 f_{Y_1, \dots, Y_{n-1}}(y_1, \dots, y_{n-1}) &= n! \mathbb{1}_{y_1 < \dots < y_{n-1}} f(y_1) \dots f(y_{n-1}) \int_{\mathbb{R}} \mathbb{1}_{y_{n-1} < y_n} f(y_n) dy_n \\
 &= n! \mathbb{1}_{y_1 < \dots < y_{n-1}} f(y_1) \dots f(y_{n-1}) \int_{y_{n-1}}^{+\infty} f(y_n) dy_n \\
 &= n! \mathbb{1}_{y_1 < \dots < y_{n-1}} f(y_1) \dots f(y_{n-1}) (1 - F(y_{n-1}))
 \end{aligned}$$

De façon similaire

$$\begin{aligned}
 f_{Y_1, \dots, Y_{n-2}}(y_1, \dots, y_{n-2}) &= n! \mathbb{1}_{y_1 < \dots < y_{n-2}} f(y_1) \dots f(y_{n-2}) \int_{y_{n-2}}^{+\infty} f(y_{n-1}) (1 - F(y_{n-1})) dy_{n-1} \\
 &= n! \mathbb{1}_{y_1 < \dots < y_{n-2}} f(y_1) \dots f(y_{n-2}) \left[ -\frac{1}{2} (1 - F(u))^2 \right]_{y_{n-2}}^{+\infty} \\
 &= \frac{n!}{2!} \mathbb{1}_{y_1 < \dots < y_{n-2}} f(y_1) \dots f(y_{n-2}) (1 - F(y_{n-2}))^2
 \end{aligned}$$

On en conclut donc par récurrence que

$$\forall (y_1, \dots, y_r) \in \mathbb{R}^r, l(y_1, \dots, y_r) = \frac{n!}{(n-r)!} \mathbb{1}_{y_1 < \dots < y_r} f(y_1) \dots f(y_r) (1 - F(y_r))^{n-r}$$

Loi de  $Y_r$  seul :

On a :

$$\begin{aligned}
 &f_{Y_2, \dots, Y_r}(y_2, \dots, y_r) \\
 &= \frac{n!}{(n-r)!} f(y_2) \dots f(y_r) \dots f(y_r) (1 - F(y_r))^{n-r} \mathbb{1}_{y_2 < \dots < y_r} \int_{\mathbb{R}} \mathbb{1}_{y_1 < y_2} f(y_1) dy_1 \\
 &= \frac{n!}{(n-r)!} f(y_2) \dots f(y_r) \dots f(y_r) (1 - F(y_r))^{n-r} \mathbb{1}_{y_2 < \dots < y_r} \int_{-\infty}^{y_2} f(y_1) dy_1 \\
 &= \frac{n!}{(n-r)!} f(y_2) \dots f(y_r) \dots f(y_r) (1 - F(y_r))^{n-r} \mathbb{1}_{y_2 < \dots < y_r} F(y_2)
 \end{aligned}$$

Par suite

$$\begin{aligned}
 &f_{Y_3, \dots, Y_r}(y_3, \dots, y_r) \\
 &= \frac{n!}{(n-r)!} f(y_3) \dots f(y_r) \dots f(y_r) (1 - F(y_r))^{n-r} \mathbb{1}_{y_3 < \dots < y_r} \int_{\mathbb{R}} \mathbb{1}_{y_2 < y_3} F(y_2) f(y_2) dy_2 \\
 &= \frac{n!}{2!(n-r)!} f(y_3) \dots f(y_r) \dots f(y_r) (1 - F(y_r))^{n-r} \mathbb{1}_{y_3 < \dots < y_r} F(y_3)^2
 \end{aligned}$$

Et par une récurrence immédiate il vient

$$\forall y_r \in \mathbb{R}, l(y_r) = \frac{n!}{(n-r)!(r-1)!} F(y_r)^{r-1} f(y_r) (1 - F(y_r))^{n-r}$$

On notera la densité du plus petit des  $x_i$ , à savoir

$$l(y_1) = n(1 - F(y_1))^{n-1} f(y_1)$$

et symétriquement celle du plus grand

$$l(y_n) = nF(y_n)^{n-1} f(y_n)$$

On suppose ici que la loi des  $X_i$  est exponentielle de densité

⇒ Q2

$$f(x) = \frac{1}{\theta} \exp\left(\frac{-(x - \alpha)}{\theta}\right) \mathbf{1}_{[\alpha, +\infty[}(x)$$

où  $\theta > 0$  et  $\alpha \geq 0$  sont des paramètres inconnus.

Ecrire la loi du  $r$ -uplet  $(Y_1, \dots, Y_r)$  dans ce cas.

On a pour tout  $x \in \mathbb{R}$

$$F(x) = \left(1 - e^{-\frac{x-\alpha}{\theta}}\right) \mathbf{1}_{\alpha < x}$$

et donc d'après le résultat précédent

$$\begin{aligned} f_{Y_1, \dots, Y_r}(y_1, \dots, y_r) &= \frac{n!}{(n-r)!} \mathbf{1}_{y_1 < y_2 < \dots < y_r} \left( \prod_{i=1}^r \frac{1}{\theta} e^{-\frac{y_i - \alpha}{\theta}} \mathbf{1}_{\alpha < y_i} \right) \left( e^{-\frac{y_r - \alpha}{\theta}} \mathbf{1}_{\alpha < y_r} \right)^{n-r} \\ &= \frac{n!}{(n-r)!} \frac{1}{\theta^r} e^{-\frac{1}{\theta} \sum_{i=1}^r (y_i - \alpha) - (n-r) \frac{y_r - \alpha}{\theta}} \cdot \mathbf{1}_{\alpha < y_1 < y_2 < \dots < y_r} \\ &= \frac{n!}{(n-r)!} \frac{1}{\theta^r} e^{-r \frac{\bar{y} - \alpha}{\theta} - (n-r) \frac{y_r - \alpha}{\theta}} \cdot \mathbf{1}_{\alpha < y_1 < y_2 < \dots < y_r} \end{aligned}$$

où  $\bar{y} = \frac{1}{r} \sum_{i=1}^r y_i$ .

⇒ Q3

Trouver une statistique exhaustive pour les paramètres  $\alpha$  et  $\theta$ .

On a

$$f_{Y_1, \dots, Y_r}(y_1, \dots, y_r) = \frac{n!}{(n-r)!} \underbrace{\frac{1}{\theta^r} e^{-\frac{1}{\theta}(r\bar{y} + (n-r)y_r)} e^{-n\alpha}}_{\text{dépend de } \theta, \alpha \text{ et que de } r\bar{y} + (n-r)y_r} \underbrace{\mathbf{1}_{\alpha < y_1}}_{\text{dépend de } \alpha \text{ et que de } y_1} \underbrace{\mathbf{1}_{y_1 < y_2 < \dots < y_r}}_{\text{ne dépend que de } y}$$

Soit donc  $S_\theta(y_1, \dots, y_r) = r\bar{y} + (n-r)y_r$  et  $S_\alpha(y_1, \dots, y_r) = y_1$ . Alors en vertu du théorème de factorisation  $(S_\theta, S_\alpha)$  est une statistique exhaustive pour le modèle paramétré par  $\theta, \alpha$ .

En outre lorsque  $\alpha$  est connu  $S_\theta$  est une statistique exhaustive pour le modèle paramétré par  $\theta$ , de même que  $S_\alpha$  lorsque  $\theta$  est connu dans le modèle paramétré par  $\alpha$ .

<b>Exercice corrigé 2</b>
---------------------------

Calculer l'information de Fisher dans les modèles statistiques suivants :

une loi de Poisson de paramètre $\lambda$ :
---

☞ Q1

$$\mathbb{P}(X = k) = e^{-\lambda} \frac{\lambda^k}{k!} \quad \text{pour } k \in \mathbb{N}$$

L'information de Fisher pour  $n$  observations i.i.d. est égale à  $n$  fois l'information de Fisher pour une observation. On se contente donc de calculer l'information de Fisher dans le cas  $n = 1$ .

$$\begin{aligned} L_1(k; \lambda) &= \exp^{-\lambda} \frac{\lambda^k}{k!} \\ \ln L_1(k; \lambda) &= -\lambda + k \ln \lambda - \ln(k!) \\ \frac{\partial \ln L_1}{\partial \lambda}(k, \lambda) &= -1 + \frac{k}{\lambda} \\ \frac{\partial^2 \ln L_1}{\partial \lambda^2}(k, \lambda) &= -\frac{k}{\lambda^2} \end{aligned}$$

Or  $\mathbb{E}(k) = \lambda$ , et donc  $I_1(\lambda) = \frac{1}{\lambda}$  (et  $I_n(\lambda) = \frac{n}{\lambda}$ )

une loi de Pareto de paramètres $\alpha$ et $\theta$ avec $\alpha > 1$ et $\theta > 0$ , de densité :
---

☞ Q2

$$f(x) = \frac{\alpha - 1}{\theta} \left(\frac{\theta}{x}\right)^\alpha \mathbf{1}_{x \geq \theta}$$

$$L = \frac{\alpha}{\theta} \left(\frac{\theta}{x}\right)^\alpha \mathbf{1}_{[\theta, +\infty[}(x)$$

La vraisemblance n'est pas dérivable en  $\theta$  : l'information de Fisher n'est pas définie pour ce modèle. Si  $\theta$  est une constante connue et non un paramètre à estimer, on peut calculer l'information de Fisher pour le modèle 'réduit' (paramétré par  $\alpha$ ) :

$$\begin{aligned} \frac{\partial \ln L}{\partial \alpha} &= 1/(\alpha - 1) + \ln \theta - \ln x, \\ \frac{\partial^2 \ln L}{\partial \alpha^2} &= -\frac{1}{(\alpha - 1)^2} \end{aligned}$$

D'où  $I_1(\alpha) = \frac{1}{(\alpha-1)^2}$ .

une loi de Weibull de paramètres $\alpha$ et $\theta$ avec $\alpha > 0$ et $\theta > 0$ de densité :
--

☞ Q3

$$f(x) = \alpha \theta x^{\alpha-1} e^{-\theta x^\alpha}$$

On a

$$\begin{aligned}\ln L &= \ln \alpha + \ln \theta + (\alpha - 1) \ln x - \theta x^\alpha \\ \frac{\partial^2 \ln L}{\partial \theta^2} &= -\frac{1}{\theta^2} \\ \frac{\partial^2 \ln L}{\partial \theta \partial \alpha} &= -x^\alpha \ln x, \\ \frac{\partial^2 \ln L}{\partial \alpha^2} &= -\frac{1}{\alpha^2} - \theta x^\alpha (\ln x)^2.\end{aligned}$$

d'où

$$I_1(\alpha, \theta) = \begin{pmatrix} \frac{1}{\alpha^2} + \theta \mathbb{E}(X^\alpha (\ln X)^2) & \mathbb{E}(X^\alpha \ln X) \\ \mathbb{E}(X^\alpha \ln X) & \frac{1}{\theta^2} \end{pmatrix}$$

Il reste alors à exprimer  $\mathbb{E}(X^\alpha (\ln X)^i)$  pour  $i \in \{1, 2\}$ ; ce calcul peut être fait au moyen des intégrales eulériennes.

Remarquons pour ce faire que pour  $p \in \mathbb{N}^*$  on a

$$\begin{aligned}\mathbb{E}(X^\alpha \ln(X)^p) &= \int_0^{+\infty} \theta x^\alpha \ln^p(x) \alpha x^{\alpha-1} e^{-\theta x^\alpha} dx \\ &= \frac{1}{\alpha^p \theta} \int_0^{+\infty} \ln^p\left(\frac{u}{\theta}\right) u e^{-u} du \quad \text{en posant } u = \theta x^\alpha\end{aligned}$$

En particulier pour  $p \in \{1, 2\}$  on a

$$\begin{aligned}\mathbb{E}(X^\alpha \ln(X)) &= \frac{1}{\alpha \theta} \int_0^{+\infty} \ln\left(\frac{u}{\theta}\right) u e^{-u} du \\ \mathbb{E}(X^\alpha \ln(X)^2) &= \frac{1}{\alpha^2 \theta} \int_0^{+\infty} \ln\left(\frac{u}{\theta}\right)^2 u e^{-u} du\end{aligned}$$

Or la fonction  $\Gamma$  d'Euler, définie par  $\Gamma : \begin{pmatrix} \mathbb{R}^{+*} & \rightarrow & \mathbb{R} \\ x & \mapsto & \int_0^{+\infty} t^{x-1} e^{-t} dt \end{pmatrix}$ , est de classe  $C^\infty$  et vérifie  $\forall p \in \mathbb{N}, \forall x \in \mathbb{R}^{+*}, \Gamma^{(p)}(x) = \int_0^{+\infty} \ln^p(t) t^{x-1} e^{-t} dt$ . En conséquence

$$\begin{aligned}\mathbb{E}(X^\alpha \ln(X)) &= \frac{1}{\alpha \theta} \left( \Gamma'(2) - \ln(\theta) \Gamma(2) \right) \\ \mathbb{E}(X^\alpha \ln(X)^2) &= \frac{1}{\alpha^2 \theta} \left( \Gamma''(2) - 2 \ln(\theta) \Gamma'(2) + \ln(\theta)^2 \Gamma(2) \right)\end{aligned}$$

Enfin, on peut montrer que  $\frac{\Gamma'(u)}{\Gamma(u)} = -\gamma - \frac{1}{u} + \sum_{n=1}^{\infty} \left( \frac{1}{n} - \frac{1}{n+u} \right)$  (où  $\gamma$  est la constante d'Euler définie par  $\gamma = \lim_{N \rightarrow +\infty} \sum_{n=1}^N \frac{1}{n} - \ln N$ ) ce dont on tire par dérivation que  $\frac{\Gamma''(u)\Gamma(u) - \Gamma'(u)^2}{\Gamma(u)^2} = \frac{1}{u^2} + \sum_{n=1}^{\infty} \frac{1}{(n+u)^2}$ , ceci pour tout  $u \in \mathbb{R}^{+*}$ ; les séries figurant dans ces expressions sont alors calculables lorsque  $u$  est entier.

En particulier lorsque  $x = 2$  il vient :

$$\begin{aligned}\Gamma(2) &= 1! \\ &= 1 \\ \Gamma'(2) &= -\gamma - \frac{1}{2} + \sum_{n=1}^{\infty} \left( \frac{1}{n} - \frac{1}{n+2} \right) \\ &= 1 - \gamma \\ \Gamma''(2) &= \frac{1}{4} + \sum_{n=1}^{\infty} \frac{1}{(n+2)^2} + \Gamma'(2)^2 \\ &= \frac{1}{4} + \left( \frac{\pi^2}{6} - 1 - \frac{1}{4} \right) + (1 - \gamma)^2 \\ &= \frac{\pi^2}{6} - 2\gamma + \gamma^2\end{aligned}$$

En définitive on a donc

$$\begin{aligned}\mathbb{E}(X^\alpha \ln(X)) &= \frac{1 - \gamma - \ln \theta}{\theta \alpha} \\ \mathbb{E}(X^\alpha (\ln X)^2) &= \frac{1}{\alpha^2 \theta} \left( \frac{\pi^2}{6} - 2\gamma + \gamma^2 - 2(1 - \gamma) \ln(\theta) + \ln(\theta)^2 \right)\end{aligned}$$

de sorte que

$$I_1(\alpha, \theta) = \left( \frac{\frac{\theta}{\alpha^2} + \frac{1}{\alpha^2} \left( \frac{\pi^2}{6} - 2\gamma + \gamma^2 - 2(1 - \gamma) \ln(\theta) + \ln(\theta)^2 \right)}{\frac{1 - \gamma - \ln \theta}{\theta \alpha}} \mid \frac{\frac{1 - \gamma - \ln \theta}{\theta \alpha}}{\frac{1}{\theta^2}} \right)$$

Remarquons que dans le sous-modèle dans lequel  $\alpha$  est connu et  $\theta$  est inconnu on a simplement

$$I_1(\theta) = \frac{1}{\theta^2}$$

⇒ Q4 loi uniforme sur  $[0, \theta]$  avec  $\theta > 0$  inconnu.

La vraisemblance du modèle est

$$f_{X_1, \dots, X_n}(x_1, \dots, x_n; \theta) = \frac{1}{\theta^n} \mathbb{1}_{\min x_i \geq 0} \mathbb{1}_{\max x_i \leq \theta}$$

donc la log-vraisemblance s'écrit, pour  $x \in [0, \theta]^n$

$$L_{X_1, \dots, X_n}(x_1, \dots, x_n; \theta) = -n \ln \theta$$

L'information de Fisher n'est pas définie, car le modèle n'est pas homogène (à  $x$  fixé, la log-vraisemblance  $\theta \mapsto L_{X_1, \dots, X_n}(x_1, \dots, x_n; \theta)$  n'est pas définie (donc pas dérivable) au point  $\theta =$



$\max_i x_i$ ). Sur  $]0, \max_i x_i[$  on peut néanmoins calculer par analogie<sup>2</sup>

$$\begin{aligned} I_n(\theta) &= \mathbb{E} \left( \left( \frac{\partial L_{X_1, \dots, X_n}(x; \theta)}{\partial \theta} \right)^2 \right) \\ &= \int_{\mathbb{R}^n} \left( -\frac{n}{\theta} \right)^2 \frac{dx_1}{\theta} \dots \frac{dx_n}{\theta} \\ &= \frac{n^2}{\theta^2} \end{aligned}$$

Remarquons que

$$\begin{aligned} \mathbb{E} \left( -\frac{\partial^2 L_{X_1, \dots, X_n}(x; \theta)}{\partial \theta^2} \right) &= \mathbb{E} \left( \frac{n}{\theta^2} \right) \\ &= \frac{n}{\theta^2} \\ &\neq \frac{n^2}{\theta^2} \end{aligned}$$

En effet, l'hypothèse de “dérivabilité deux fois sous le signe  $\int_{\mathcal{X}}$ ” qui stipule que

$$\frac{\partial}{\partial \theta} \int_{\mathcal{X}} \ln f_{X_1, \dots, X_n}(x) dx = \int_{\mathcal{X}} \frac{\partial \ln f_{X_1, \dots, X_n}(x)}{\partial \theta} dx$$

et que

$$\frac{\partial^2}{\partial \theta^2} \int_{\mathcal{X}} \ln f_{X_1, \dots, X_n}(x) dx = \int_{\mathcal{X}} \frac{\partial^2 \ln f_{X_1, \dots, X_n}(x)}{\partial \theta^2} dx$$

n'est pas vérifiée ici.  $I_n(\theta)$  est toujours par définition la variance du score, mais n'est plus égale à la courbure espérée de la log-vraisemblance.

Rappel : Lorsque cette hypothèse est vérifiée on a d'une part

$$\begin{aligned} \mathbb{E} \left( \frac{\partial \ln f_{X_1, \dots, X_n}(x; \theta)}{\partial \theta} \right) &= \mathbb{E} \left( \frac{\frac{\partial f_{X_1, \dots, X_n}(x; \theta)}{\partial \theta}}{f_{X_1, \dots, X_n}(x; \theta)} \right) \\ &= \int_{\mathbb{X}} \frac{\partial f_{X_1, \dots, X_n}(x; \theta)}{\partial \theta} dx \\ &= \frac{\partial}{\partial \theta} \left( \theta \mapsto \int_{\mathbb{X}} f_{X_1, \dots, X_n}(x; \theta) dx \right)_{(\theta)} \\ &\quad \text{d'après la première partie de l'hypothèse} \\ &= \frac{\partial}{\partial \theta} (\theta \mapsto 1)_{(\theta)} \\ &= 0 \end{aligned}$$

<sup>2</sup>L'objet  $I_n(\theta)$  existe pour  $\theta \in ]0, \max_i x_i[$ , mais n'a plus les propriétés habituelles de l'information de Fisher (à commencer par un problème de définition), et c'est pourquoi on s'abstient habituellement de la définir dans ce cas.

c'est-à-dire que le score est centré, et d'autre part en dérivant de nouveau

$$\begin{aligned}
 \mathbb{E} \left( \frac{\partial^2 \ln f_{X_1, \dots, X_n}(x, \theta)}{\partial \theta^2} \right) &= \mathbb{E} \left( \frac{\partial}{\partial \theta} \left( \frac{\frac{\partial f_{X_1, \dots, X_n}(x; \theta)}{\partial \theta}}{f_{X_1, \dots, X_n}(x, \theta)} \right) \right) \\
 &= \mathbb{E} \left( \frac{\frac{\partial^2 f_{X_1, \dots, X_n}(x; \theta)}{\partial \theta^2}}{f_{X_1, \dots, X_n}(x, \theta)} - \frac{\left( \frac{\partial f_{X_1, \dots, X_n}(x; \theta)}{\partial \theta} \right)^2}{(f_{X_1, \dots, X_n}(x; \theta))^2} \right) \\
 &= \mathbb{E} \left( \frac{\frac{\partial^2 f_{X_1, \dots, X_n}(x; \theta)}{\partial \theta^2}}{f_{X_1, \dots, X_n}(x, \theta)} - \left( \frac{\partial \ln f_{X_1, \dots, X_n}(x; \theta)}{\partial \theta} \right)^2 \right) \\
 &= \int_{\mathcal{X}} \frac{\partial^2 f_{X_1, \dots, X_n}(x, \theta)}{\partial \theta^2} dx - I_n(\theta) \\
 &= \frac{\partial^2}{\partial \theta^2} \left( \theta \mapsto \int_{\mathcal{X}} f_{X_1, \dots, X_n}(x, \theta) dx \right)_{(\theta)} - I_n(\theta) \\
 &\quad \text{d'après la deuxième partie de l'hypothèse} \\
 &= -I_n(\theta)
 \end{aligned}$$

### Exercice corrigé 3

On dispose de  $n$  observations  $y_1, \dots, y_n$  sur les durées de vie de certains composants industriels. On suppose que les variables aléatoires  $Y_1, \dots, Y_n$  associées sont i.i.d, de densité  $f(t) = \frac{1}{\theta} e^{-\frac{t}{\theta}} \mathbf{1}_{t \geq 0}$ .

☞ Q1

Soit  $F$  la fonction de répartition des  $Y_i$ . On cherche à estimer la “fonction de survie” de chaque composant  $\bar{F}(t) = 1 - F(t)$ .  
Calculer  $\bar{F}(t)$  en fonction de  $t$  et  $\theta$ .

$\bar{F}(t)$  désigne la probabilité qu'un composant donné survive jusqu'à la date  $t$  et mesure ainsi sa fiabilité. On a

$$\begin{aligned}
 F(t) &= \mathbb{P}(Y_1 \leq t) \\
 &= \int_0^t \frac{1}{\theta} e^{-\frac{u}{\theta}} du \\
 &= 1 - e^{-\frac{t}{\theta}}
 \end{aligned}$$

en conséquence de quoi

$$\bar{F}(t) = e^{-\frac{t}{\theta}}.$$

☞ Q2

Calculer l'estimateur du maximum de vraisemblance de  $\theta$  et en déduire un estimateur convergent  $\hat{F}(t)$  de  $\bar{F}(t)$ . Que peut-on dire du biais de  $\hat{F}(t)$  ?

La log-vraisemblance du modèle s'écrit pour  $\min_i y_i \geq 0$

$$\ln L(y_1, \dots, y_n; \theta) = -n \ln \theta - \frac{1}{\theta} \sum_{i=1}^n y_i$$

Or si  $\hat{\theta}_{(y)}$  maximise la vraisemblance il vérifie  $\frac{\partial(\ln L)}{\partial \theta}(\hat{\theta}_{(y)}) = 0$ , soit  $-\frac{n}{\hat{\theta}_{(y)}} + \frac{1}{(\hat{\theta}_{(y)})^2} \sum_{i=1}^n y_i = 0$  et donc

$$\hat{\theta} : (y_1, \dots, y_n) \mapsto \frac{1}{n} \sum_{i=1}^n y_i$$

(la réciproque étant immédiate).

Définissons donc  $\hat{F}(t) : y \mapsto e^{-t/\hat{\theta}_{(y)}}$ .

Comme  $\hat{\theta}(Y) \xrightarrow[n \rightarrow +\infty]{p.s.} \theta$  et que  $\phi_t : (u \mapsto e^{-\frac{t}{u}})$  est absolument continue en  $u$ , on a

$$\forall t, \hat{F}(t) \xrightarrow[n \rightarrow +\infty]{p.s.} \bar{F}(t)$$

En revanche à distance finie, même si  $\mathbb{E}(\hat{\theta}(Y)) = \theta$  pour comparer  $\mathbb{E}(\hat{F}(t)_{(Y)})$  à  $F(t)$  il n'est **pas** possible d'utiliser de l'inégalité de Jensen selon laquelle

Si  $\phi : (\mathbb{R} \rightarrow \mathbb{R})$  est strictement convexe, et si  $X$  est non-dégénérée (i.e.  $\mathbb{V}(X) \neq 0$ ), alors  $\mathbb{E}(\phi(X)) > \phi(\mathbb{E}(X))$ .

car  $\phi_t : (u \mapsto e^{-\frac{t}{u}})$  n'est **pas** convexe, puisque  $\frac{\partial^2 \phi_t}{\partial u^2}(u) = \frac{t}{u^3} (\frac{t}{u} - 2) e^{-\frac{t}{u}}$  qui n'est pas toujours positif.

Calculons donc le biais de  $\hat{F}(t)$  : on a

$$\begin{aligned} \mathbb{E}(\hat{F}(t)) &= \mathbb{E}\left(e^{-\sum_{i=0}^n \frac{nt}{Y_i}}\right) \\ &= \mathbb{E}\left(\sum_{k=0}^{+\infty} \frac{1}{k!} \left(-\frac{nt}{\sum_{i=0}^n Y_i}\right)^k\right) \\ &= \sum_{k=0}^{+\infty} \frac{1}{k!} (-nt)^k \mathbb{E}\left(\frac{1}{(\sum_{i=0}^n Y_i)^k}\right) \quad \text{par Fubini} \end{aligned}$$

Or pour tout  $k \geq 1$  fixé,  $\phi_k : \begin{pmatrix} \mathbb{R}^+ & \rightarrow & \mathbb{R}^+ \\ x & \mapsto & \frac{1}{x^k} \end{pmatrix}$  est convexe (car  $\mathcal{C}^\infty$  et de dérivée  $\phi'_k : (x \mapsto -\frac{k}{x^{k+1}})$ , donc  $\phi''_k(x) = \frac{k(k+1)}{x^{k+2}} > 0$ ) et donc d'après l'inégalité stricte de Jensen  $\mathbb{E}\left(\frac{1}{(\sum_{i=0}^n Y_i)^k}\right) > \frac{1}{(\mathbb{E}(\sum_{i=0}^n Y_i))^k}$  de sorte que finalement

$$\mathbb{E}(\hat{F}(t)) > \sum_{k=0}^{+\infty} \frac{1}{k!} (-nt)^k \frac{1}{(\mathbb{E}(\sum_{i=0}^n Y_i))^k}$$

Or  $(Y_i)_i \underset{iid}{\rightsquigarrow} \mathcal{E}\left(\frac{1}{\theta}\right)$  et donc (voir TD 1, exercice 2)  $(\sum_{i=0}^n Y_i) \underset{iid}{\rightsquigarrow} \Gamma\left(n, \frac{1}{\theta}\right)$  de sorte qu'en particulier  $\mathbb{E}\left(\sum_{i=0}^n Y_i\right) = n\theta$

Par suite

$$\begin{aligned}\mathbb{E}\left(\widehat{F}(t)\right) &> \sum_{k=0}^{+\infty} \frac{1}{k!} (-nt)^k \frac{1}{(n\theta)^k} \\ &= \sum_{k=0}^{+\infty} \frac{1}{k!} \frac{(-t)^k}{\theta^k} \\ &= e^{-\frac{t}{\theta}} \\ &= \bar{F}(t)\end{aligned}$$

Ainsi,  $\widehat{F}(t)$  est un estimateur biaisé de  $\bar{F}(t)$ .

Notons que bien entendu  $\mathbb{E}\left(\widehat{F}(t)_{(Y)}\right) \xrightarrow[n \rightarrow +\infty]{} e^{-t} = \bar{F}(t)$ , ce qui découle de ce que  $\forall t, \widehat{F}(t) \xrightarrow[n \rightarrow +\infty]{p.s.} F(t)$ .

UNIVERSITÉ  
DE  
LILLE  
SCIENCE  
ET  
TECHNIQUE

☞ Q3 Calculer la loi limite de  $\sqrt{n}(\hat{F}(t) - \bar{F}(t))$ .

On a  $\sqrt{n}(\hat{\theta} - \theta) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}(0, I_1(\theta)^{-1})$ , avec

$$\begin{aligned} I_1(\theta) &= \frac{1}{n} I_n(\theta) \\ &= -\frac{1}{n} \mathbb{E} \left( \frac{\partial^2 \ln L}{\partial \theta^2} \right) \\ &= -\frac{1}{n} \left( \frac{n}{\theta^2} - \frac{2n}{\theta^2} \right) \\ &= \frac{1}{\theta^2} \end{aligned}$$

En posant  $g : (\theta \mapsto e^{-\frac{t}{\theta}})$  de dérivée  $g'(\theta) = \frac{t}{\theta^2} e^{-\frac{t}{\theta}}$  on a par la delta-méthode

$$\begin{aligned} \sqrt{n} \left( g(\hat{\theta}) - g(\theta) \right) &\xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N} \left( 0, g'(\theta) \cdot I_1(\theta)^{-1} \cdot g'(\theta) \right) \\ &= \mathcal{N} \left( 0, \left( \frac{t}{\theta^2} e^{-\frac{t}{\theta}} \right) \theta^2 \left( \frac{t}{\theta^2} e^{-\frac{t}{\theta}} \right) \right) \\ &= \mathcal{N} \left( 0, \frac{t^2}{\theta^2} e^{-\frac{2t}{\theta}} \right) \end{aligned}$$

Soit  $T$  la variable aléatoire définie par :

$$T = \begin{cases} 1, & \text{si } Y_1 > t \\ 0, & \text{sinon.} \end{cases}$$

On note par ailleurs  $S(Y) = Y_1 + \dots + Y_n$ .

☞ Q4

(a) Déterminer la loi de  $Y_1$  conditionnellement à  $S$ .

Cherchons dans un premier temps la densité  $f_{S_n(Y)}(s)$  de la loi de la variable  $S_n(Y) = Y_1 + \dots + Y_n$ .

- On a tout d'abord  $f_{S_1(Y)}(s) = f_{Y_1}(s) = \frac{1}{\theta} e^{-\frac{s}{\theta}} \mathbf{1}_{s \geq 0}$ .
- $Y_1$  et  $Y_2$  étant indépendante, on a alors

$$\begin{aligned} f_{S_2(Y)}(s) &= \int_{\mathbb{R}^2} f_{Y_1}(y_1) f_{Y_2}(y_2) \mathbf{1}_{y_1+y_2=s} dy_1 dy_2 \\ &= \int_{\mathbb{R}} \mathbf{1}_{y_1 \geq 0} \mathbf{1}_{s-y_1 \geq 0} f_{Y_1}(y_1) f_{Y_2}(s-y_1) dy_1 \\ &= \mathbf{1}_{s \geq 0} \int_0^s \left( \frac{1}{\theta} e^{-\frac{y_1}{\theta}} \right) \left( \frac{1}{\theta} e^{-\frac{s-y_1}{\theta}} \right) dy_1 \\ &= \mathbf{1}_{s \geq 0} s \frac{e^{-\frac{s}{\theta}}}{\theta^2} \end{aligned}$$

– puis de même

$$\begin{aligned}
 f_{S_3(Y)}(s) &= \int_{\mathbb{R}^3} f_{Y_1}(y_1) f_{Y_2}(y_2) f_{Y_3}(y_3) \mathbb{1}_{y_1+y_2+y_3=s} dy_1 dy_2 dy_3 \\
 &= \int_{\mathbb{R}^2} \left( \int_{\mathbb{R}^2} f_{Y_1}(y_1) f_{Y_2}(y_2) \mathbb{1}_{y_1+y_2=s_1} dy_1 dy_2 \right) f_{Y_3}(y_3) \mathbb{1}_{s_1+y_3=s} ds_1 dy_3 \\
 &= \int_{\mathbb{R}} \mathbb{1}_{y_3 \geq 0} \mathbb{1}_{s-y_3 \geq 0} f_{S_2(Y)}(s-y_3) f_{Y_3}(y_3) dy_3 \\
 &= \int_{\mathbb{R}} f_{S_2(Y)}(s-y_3) f_{Y_3}(y_3) dy_3 \\
 &= \mathbb{1}_{s \geq 0} \int_0^s \left( \frac{s-y_3}{\theta^2} e^{-\frac{s-y_3}{\theta}} \right) \left( \frac{1}{\theta} e^{-\frac{y_3}{\theta}} \right) dy_3 \\
 &= \mathbb{1}_{s \geq 0} \frac{s^2}{2} \frac{e^{-\frac{s}{\theta}}}{\theta^3}
 \end{aligned}$$

– (récurrence ...)

– et finalement

$$f_{S_n(Y)}(s) = \mathbb{1}_{s \geq 0} \frac{s^{n-1}}{(n-1)!} \frac{e^{-\frac{s}{\theta}}}{\theta^n}$$

*Remarque* : Ce résultat peut bien-sûr être obtenu directement en constatant que  $S(Y) \sim \Gamma(n, \frac{1}{\theta})$  (voir TD 1, exercice 2).

On en tire alors la loi conditionnelle  $f(y, s)$  de la variable  $(Y, S(Y))$  : pour  $y_1 \in [0, s]$  on a

$$\begin{aligned}
 f(y_1, s) &= \frac{f_{Y_1, Y_1+\dots+Y_n}(y_1, s)}{f_{Y_1+\dots+Y_n}(s)} \\
 &= \frac{f_{Y_1, Y_2+\dots+Y_n}(y_1, s-y_1)}{f_{Y_1+\dots+Y_n}(s)} \\
 &= \frac{f_{Y_1}(y_1) f_{Y_2+\dots+Y_n}(s-y_1)}{f_{Y_1+\dots+Y_n}(s)} \\
 &\quad \text{car les } Y_i \text{ sont indépendantes et donc } Y_1 \text{ est indépendante de } Y_2 + \dots + Y_n \\
 &= \frac{\left( \frac{(s-y_1)^{n-2}}{(n-2)!} \frac{e^{-\frac{s-y_1}{\theta}}}{\theta^{n-1}} \right) \left( \frac{e^{-\frac{y_1}{\theta}}}{\theta} \right)}{\frac{s^{n-1}}{(n-1)!} \frac{e^{-\frac{s}{\theta}}}{\theta^n}}
 \end{aligned}$$

et donc en définitive

$$f(y_1, s) = (n-1) \frac{(s-y_1)^{n-2}}{s^{n-1}} \mathbb{1}_{0 \leq y_1 \leq s}$$

Calculer

(b)

$$T^*(Y) = \mathbb{E}(T|S(Y))$$

Comment s'appelle cet estimateur ?

Calculons tout d'abord que  $T^*(Y)$ .

On a successivement, pour  $t \in \mathbb{R}^+$  et  $s \geq t$

$$\begin{aligned} \mathbb{E}(T|S(Y) = s) &= \mathbb{P}(Y_1 > t | S(Y) = s) \\ &= \int_{\mathbb{R}} \left( (n-1) \frac{(s-y)^{n-2}}{s^{n-1}} \mathbf{1}_{0 \leq y \leq s} \right) \mathbf{1}_{t \leq y} dy \\ &= \int_t^s (n-1) \frac{(s-y)^{n-2}}{s^{n-1}} dy \\ &= \frac{1}{s^{n-1}} [-(s-y)^{n-1}]_t^s \\ &= \left(1 - \frac{t}{s}\right)^{n-1} \end{aligned}$$

En d'autres termes,

$$T^*(Y) = \left(1 - \frac{t}{S(Y)}\right)^{n-1}$$

$T^*$  est l'amélioré de Rao-Blackwell de  $T$  pour la statistique exhaustive  $S$ .

- (c) Montrer que  $T^*$  est l'estimateur sans biais de  $\bar{F}(t)$  optimal (parmi les estimateurs sans biais).

On a successivement

$$\begin{aligned} \mathbb{E}(\mathbb{E}(T|S(Y))) &= \mathbb{E}(T) \\ &= \mathbb{E}(\mathbf{1}_{Y_1 > t}) \\ &= \mathbb{P}(Y_1 > t) \\ &= \bar{F}(t) \end{aligned}$$

(on retrouve ainsi que l'amélioré  $T^*$  de  $T$ , qui estime sans biais  $\bar{F}(t)$ , est lui-même un estimateur sans biais de  $\bar{F}(t)$ ).

Enfin,  $T^*(Y) = \mathbb{E}(T|S(Y))$  est fonction d'une statistique exhaustive complète (car  $T$  est une statistique canonique dans un modèle exponentiel), et est sans biais, donc d'après le théorème de Lehman-Scheffé  $T^*$  est optimal parmi les estimateurs sans biais de  $\bar{F}(t)$ . On vérifie d'ailleurs que

$$\begin{aligned} \mathbb{V}(T) &= \mathbb{V}(\mathbb{E}(T|S(Y))) + \mathbb{E}(\mathbb{V}(T|S(Y))) \\ &> \mathbb{V}(\mathbb{E}(T|S(Y))) \\ &= \mathbb{V}(T^*(Y)) \end{aligned}$$



(d) Peux-on dire si  $T^*$  est efficace (à distance finie) ?

Discuter de l'efficacité de  $T^*$ , c'est comparer sa variance à la borne de Fréchet-Darmois-Cramer-Rao  $\frac{\partial \phi_t}{\partial \theta}(\theta)' I_1(\theta)^{-1} \frac{\partial \phi_t}{\partial \theta}(\theta)$ .

Cependant le calcul de  $[\mathbb{E}(\mathbb{E}(T|S(Y)))^2 - \mathbb{E}((T|S(Y))^2)]$  est ardu : il est nécessaire d'explicitier l'intégrale et d'y soustraire la borne FDCR pour conclure (un tel calcul est laissé à la sagacité du lecteur).

Constatant néanmoins que le modèle est exponentiel et régulier, et comme  $T^*$  n'est **pas** une fonction affine de la statistique exhaustive  $S$  comme il devrait l'être s'il était efficace (cf cours de P. Doukhan, Théorème 4.3), on peut en conclure que  $T^*$  n'est **pas** efficace.

### Exercice corrigé 4

On étudie une variable aléatoire  $X$ , de densité  $f(\cdot, \theta)$  ( $f$  est  $C^1$ ).

→ Q1 Quelle est la fonction score du modèle, notée  $S(X; \theta)$  ?

Donner l'expression de l'information de Fisher  $I_X(\theta)$ .

On a

$$S(x; \theta) = \frac{\partial \ln L}{\partial \theta}(x, \theta)$$

En outre, si on peut intervertir  $\int_{\mathcal{X}}$  et  $\frac{\partial}{\partial \theta}$  alors le score est centré :  $\mathbb{E}(S(x; \theta)) = 0$  (voir exercice 2)

L'information de Fisher pour une observation est définie par

$$I_1(\theta) = \mathbb{E} \left( \frac{\partial \ln L}{\partial \theta}(x, \theta) \frac{\partial \ln L}{\partial \theta'}(x, \theta) \right)$$

En outre, si on peut intervertir  $\int_{\mathcal{X}}$  avec  $\frac{\partial}{\partial \theta}$  et  $\frac{\partial^2}{\partial \theta^2}$  alors

$$I_1(\theta) = \mathbb{E} \left( -\frac{\partial^2 \ln L}{\partial \theta^2}(x, \theta) \right)$$

(voir exercice 2)

En fait, on ne parvient pas à observer  $X$ , mais seulement  $Y$  définie par :

$$Y = \begin{cases} 1, & \text{si } X \geq s \\ 0, & \text{si } X < s \end{cases}$$

→ Q2 où  $s$  est un seuil connu.

On suppose que l'on peut intervertir  $\int_X$  et  $\frac{\partial}{\partial \theta}$ ; donner la fonction score du modèle, notée  $S_Y(y; \theta)$ .

En déduire que

$$S_Y(y; \theta) = \mathbb{E}(S_X(X; \theta) | Y = y)$$

La variable aléatoire  $Y$  ne prend que deux valeurs, et suit donc une loi de Bernoulli; en outre on a  $(Y = 1) \Leftrightarrow (X \geq s)$  et donc  $\mathbb{P}(Y = 1) = \mathbb{P}(X \geq s)$ .

Notons donc  $p_s(\theta) = \mathbb{P}(X \geq s)$  : la densité de  $Y$  est

$$f_Y(y; \theta) = p_s(\theta)^y (1 - p_s(\theta))^{1-y} \mathbf{1}_{y \in \{0,1\}}$$

de sorte que pour  $y \in \{0, 1\}$

$$\ln L_Y(y; \theta) = y \ln p_s(\theta) + (1 - y) \ln(1 - p_s(\theta))$$

Le score est alors

$$\begin{aligned} S_Y(y; \theta) &= y \frac{\partial \ln p_s}{\partial \theta}(\theta) + (1 - y) \frac{\partial \ln(1 - p_s)}{\partial \theta}(\theta) \\ &= y \frac{\partial}{\partial \theta} \left( t \mapsto \ln \left( \int_s^{+\infty} f_X(x, t) dx \right) \right)_{(\theta)} + (1 - y) \frac{\partial}{\partial \theta} \left( t \mapsto \ln \left( \int_{-\infty}^s f_X(x, t) dx \right) \right)_{(\theta)} \\ &= y \frac{\frac{\partial}{\partial \theta} \left( t \mapsto \int_s^{+\infty} f_X(x, t) dx \right)_{(\theta)}}{\int_s^{+\infty} f_X(x, \theta) dx} + (1 - y) \frac{\frac{\partial}{\partial \theta} \left( t \mapsto \int_{-\infty}^s f_X(x, t) dx \right)_{(\theta)}}{\int_{-\infty}^s f_X(x, \theta) dx} \\ &= y \frac{\int_s^{+\infty} \frac{\partial f_X}{\partial \theta}(x; \theta) dx}{\int_s^{+\infty} f_X(x; \theta) dx} + (1 - y) \frac{\int_{-\infty}^s \frac{\partial f_X}{\partial \theta}(x; \theta) dx}{\int_{-\infty}^s f_X(x; \theta) dx} \end{aligned}$$

où l'un seulement des deux termes est non-nul puisque  $y \in \{0, 1\}$ .

Calculons alors le score pour le modèle en  $X$  : on a pour  $y \in \{0, 1\}$

$$\mathbb{E}(S_X(X, \theta) | Y = y) = \int_{\mathbb{R}} S_X(x, \theta) f_{X|Y=y}(x) dx$$

Or

$$S_X(x; \theta) = \frac{1}{f_X(x; \theta)} \frac{\partial f_X}{\partial \theta}(x; \theta)$$

et en outre

$$\begin{aligned} f_X(x; \theta | Y = 1) &= \frac{\mathbb{P}(X = x \wedge Y = 1)}{\mathbb{P}(Y = 1)} \\ &= \frac{f_X(x; \theta) \cdot \mathbf{1}_{X \geq s}}{\int_s^{+\infty} f_X(x; \theta) dx} \end{aligned}$$

tandis que

$$\begin{aligned}
 f_X(x; \theta | Y = 0) &= \frac{\mathbb{P}(X = x \wedge Y = 0)}{\mathbb{P}(Y = 0)} \\
 &= \frac{f_X(x; \theta) \cdot \mathbf{1}_{x < s}}{\int_{-\infty}^s f_X(x; \theta) dx}
 \end{aligned}$$

donc en définitive

$$\begin{aligned}
 \mathbb{E}(S_X(X, \theta) | Y = y) &= \int_{\mathbb{R}} S_X(x, \theta) f_{X|Y=y}(x) dx \\
 &= \int_{\mathbb{R}} S_X(x, \theta) (y f_{X|Y=1}(x) + (1 - y) f_{X|Y=0}(x)) dx \\
 &= y \int_{\mathbb{R}} S_X(x, \theta) f_{X|Y=1}(x) dx + (1 - y) \int_{\mathbb{R}} S_X(x, \theta) f_{X|Y=0}(x) dx \\
 &= y \int_{\mathbb{R}} \left( \frac{1}{f_X(x; \theta)} \frac{\partial f_X}{\partial \theta}(x; \theta) \right) \times \frac{f_X(x; \theta) \cdot \mathbf{1}_{x \geq s}}{\int_s^{+\infty} f_X(x; \theta) dx} dx \\
 &+ (1 - y) \int_{\mathbb{R}} \left( \frac{1}{f_X(x; \theta)} \frac{\partial f_X}{\partial \theta}(x; \theta) \right) \times \frac{f_X(x; \theta) \cdot \mathbf{1}_{x < s}}{\int_{-\infty}^s f_X(x; \theta) dx} dx \\
 &= y \frac{\int_s^{+\infty} \frac{\partial f_X}{\partial \theta}(x; \theta) dx}{\int_s^{+\infty} f_X(x; \theta) dx} + (1 - y) \frac{\int_{-\infty}^s \frac{\partial f_X}{\partial \theta}(x; \theta) dx}{\int_{-\infty}^s f_X(x; \theta) dx} \\
 &= \boxed{S_Y(y; \theta)}
 \end{aligned}$$

☞ Q3

En déduire alors que  $I_X(\theta) \gg I_Y(\theta)$ , où  $I_Y(\theta)$  est l'information de Fisher associée à  $Y$  (l'inégalité s'entend au sens des matrices symétriques).  
 Quelle interprétation pouvez-vous donner à l'inégalité ci-dessus ?

L'information de Fisher d'un modèle étant la variance du score correspondant, on a

$$\begin{aligned}
 I_Y(\theta) &= \mathbb{V}(S(Y; \theta)) \\
 &= \mathbb{V}(\mathbb{E}(S(X; \theta) | Y))
 \end{aligned}$$

Or pour toutes variables aléatoires  $A$  et  $B$ , à supposer que ces termes existent on a

$$\begin{aligned}
 \mathbb{E}(A) &= \mathbb{E}(\mathbb{E}_B(A|B)) \\
 \mathbb{V}(A) &= \underbrace{\mathbb{E} \left( \underbrace{\mathbb{V}_B(A|B)}_{\substack{\text{variance de } A \text{ sachant } B \\ \text{ce qui dépend de } B}} \right)}_{\text{variance espérée de } A} + \underbrace{\mathbb{V} \left( \underbrace{\mathbb{E}_B(A|B)}_{\substack{\text{espérance de } A \text{ sachant } B \\ \text{ce qui dépend de } B}} \right)}_{\text{variance due à } B}
 \end{aligned}$$

En l'occurrence il vient

$$\begin{aligned} I_X(\theta) &= \mathbb{V}(S(X; \theta)) \\ &= \mathbb{V}(\mathbb{E}(S(X; \theta)|Y)) + \mathbb{E}(\mathbb{V}(S(X; \theta)|Y)) \\ &= I_Y(\theta) + \mathbb{E}(\mathbb{V}(S(X; \theta)|Y)) \\ &\gg I_Y(\theta) \end{aligned}$$

En d'autres termes, l'information véhiculée par  $X$  est strictement plus importante que celle véhiculée par  $Y$ , ce qui est conforme à l'intuition.