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
http://ensae.no-ip.com/SE222/

Exercice corrigé 1

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

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

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

Méthode rapide mais pas naturelle :

L'idée ici est de considérer tous les réordonnancements de r variables parmi n telles c 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 ≥ éléments.
Alors pour $y_1 < \ldots < y_r$ et $dy_1, \ldots, dy_r$, avec $0 < dy_i < \frac{y_{i+1} - y_i}{2}$

$$\mathbb{P}\left(Y_1 \in [y_1, y_1 + dy_1[, \ldots, Y_r \in [y_r, y_r + dy_r[\right)$$

$$= \sum_{\pi \in \mathcal{A}_n^r} \mathbb{P}\left(\begin{array}{l} X_{\pi_1} \in [y_1, y_1 + dy_1[\wedge \ldots \wedge X_{\pi_r} \in [y_r + y_r + dy_r[\\ \wedge X_{\pi_1} < \cdots < X_{\pi_r} \wedge \forall i \notin \pi, X_i > y_r + dy_r \end{array} \right)$$

où π_k est le k-ième élément de la partie π à r éléments parmi $[\![1, n]\!]$

$$= |\mathcal{A}_n^r| \times \left\{ \begin{array}{ll} & \mathbb{P}(Z_1 \in [y_1, y_1 + dy_1[) \cdots \mathbb{P}(Z_r \in [y_r + y_r + dy_r[) \\ \times & \mathbb{1}_{y_1 < \cdots < y_r} \\ \times & \mathbb{P}(Z_{r+1} > y_r + dy_r) \cdots \mathbb{P}(Z_n > y_r + dy_r) \end{array} \right.$$

car les $[y_i, y_i + dy_i[$ sont disjoints, et les Z_i sont indépendantes car les X_i le son

$$= \frac{n!}{(n-r)!} \times \left\{ \begin{array}{ll} & (F(y_1 + dy_1) - F(y_1)) \cdots (F(y_r + dy_r) - F(y_r)) \\ \times & \mathbb{1}_{y_1 < \cdots < y_r} \\ \times & (1 - F(y_r + dy_r)) \cdots (1 - F(y_r + dy_r)) \end{array} \right.$$

En divisant et en faisant successivement tendre $dy_r \to 0^+, \ldots, dy_1 \to 0^+$ il vient [1]

[1] Précisément :

$$\frac{1}{dy_1} \mathbb{P}\left(Y_1 \in [y_1, y_1 + dy_1[, \ldots, Y_r \in [y_r, y_r + dy_r[\right) \xrightarrow[dy_1 \to 0^+]{} \frac{n!}{(n-r)!} \times \left\{ \begin{array}{ll} & f(y_1)\ (F(y_2 + dy_2) - F(y_2)) \cdots (F(y_r + \\ \times & (1 - F(y_r + dy_r))^{n-r} \end{array} \right.$$

puis

$$\frac{1}{dy_1 \cdots dy_{r-1}} \mathbb{P}\left(Y_1 \in [y_1, y_1 + dy_1[, \ldots, Y_r \in [y_r, y_r + dy_r[\right) \xrightarrow[dy_1 \to 0^+ ; \ldots ; dy_{r-1} \to 0^+]{} \frac{n!}{(n-r)!} \left\{ \begin{array}{l} f(y_1) \cdots f(y_{r-1})\ (F(y_r + dy \\ \times (1 - F(y_r + \end{array} \right.$$

$$\boxed{\forall (y_1,\ldots,y_r)\in\mathbb{R}^r,\ l(y_1,\ldots,y_r)=\tfrac{n!}{(n-r)!}\mathbb{1}_{y_1<\ldots<y_r}f(y_1)\ldots f(y_r)\left(1-F(y_r)\right)^{n-r}}$$

Méthode simple mais longue :

L'idée est ici de calculer d'abord la loi de $(Y_1,\ldots,Y_n)$, puis d'en tirer celle de $(Y_1,\ldots,Y_r)$ 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

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

et donc presque sûrement

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

de sorte que

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

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

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

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

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

On en conclut que

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

et finalement

$$\frac{1}{dy_1\cdots dy_r}\mathbb{P}\left(Y_1\in[y_1,y_1+dy_1[,\ldots,Y_r\in[y_r,y_r+dy_r[\right)\xrightarrow[dy_1\to0^+\,;\,\ldots\,;\,dy_r\to0^+]{}\frac{n!}{(n-r)!}f\left(y_1\right)\cdots f\left(y_r\right)\,\left(1-F(y_r+dy_r)\right)^{n-r}$$

On a alors

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

De façon similaire

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

On en conclut donc par récurrence que

$$\boxed{\forall (y_1,\ldots,y_r)\in\mathbb{R}^r,\ l(y_1,\ldots,y_r)=\tfrac{n!}{(n-r)!}\mathbb{1}_{y_1<\ldots<y_r}f(y_1)\ldots f(y_r)\left(1-F(y_r)\right)^{n-r}}$$

Loi de Y_r seul :

On a :

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

Par suite

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

Et par une récurrence immédiate il vient

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

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

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

et symétriquement celle du plus grand

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

☞ Q2 On suppose ici que la loi des X_i est exponentielle de densité

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

où $\theta > 0$ et $\alpha \geq 0$ sont des paramètres inconnus.
Ecrire la loi du r-uplet $(Y_1, ..., Y_r)$ dans ce cas.

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

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

et donc d'après le résultat précédent

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

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

☞ Q3 Trouver une statistique exhaustive pour les paramètres α et θ.

On a

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

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

En outre lorsque α est connu S_θ est une statistique exhaustive pour le modèle paramétré par θ, de même que S_α lorsque θ est connu dans le modèle paramétré par α.

Exercice corrigé 2

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

☞ Q1 une loi de Poisson de paramètre λ :

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

L'information de Fisher pour n observations i.i.d. est égale à n fois l'information de Fisher po
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 $I_n(\lambda) = \frac{n}{\lambda}$)

☞ Q2 une loi de Pareto de paramètres α et θ avec $\alpha > 1$ et $\theta > 0$, de densité :

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

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

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

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

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

☞ Q3 une loi de Weibull de paramètres α et θ avec $\alpha > 0$ et $\theta > 0$ de densité :

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

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

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

Il reste alors à exprimer $\mathbb{E}\left(X^\alpha(\ln X)^i\right)$ 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

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

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

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

Or la fonction Γ d'Euler, définie par $\Gamma : \begin{pmatrix} \mathbb{R}^{+*} & \to & \mathbb{R} \\ x & \mapsto & \int_0^{+\infty} t^{x-1}\, e^{-t} dt \end{pmatrix}$, est de classe C^∞ 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

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

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 γ est la constante d'Euler définie par $\gamma = \lim_{N\to+\infty} \sum_{n=1}^N \frac{1}{n} - \ln N$) 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

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

de sorte que

$$\boxed{I_1(\alpha,\theta) = \left(\begin{array}{c|c} \frac{\theta}{\alpha^2} + \frac{1}{\alpha^2}\left(\frac{\pi^2}{6} - 2\gamma + \gamma^2 - 2\,(1-\gamma)\,\ln(\theta) + \ln(\theta)^2\right) & \frac{1-\gamma-\ln\theta}{\theta\,\alpha} \\ \hline \frac{1-\gamma-\ln\theta}{\theta\,\alpha} & \frac{1}{\theta^2} \end{array}\right)}$$

Remarquons que dans le sous-modèle dans lequel α est connu et θ est inconnu on a simplem

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

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

La vraisemblance du modèle est

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

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

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

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

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

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

Remarquons que

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

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

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

et que

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

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

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

[2]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

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

Exercice corrigé 3

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

☞ 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 θ.

$\bar{F}(t)$ désigne la probabilité qu'un composant donné survive jusqu'à la date t et mesure ainsi 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 $\boxed{\bar{F}(t) = e^{-\frac{t}{\theta}}}$.

☞ Q2 Calculer l'estimateur du maximum de vraisemblance de θ 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 $min_i y_i \geq 0$

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

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

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

(la réciproque étant immédiate).
Définissons donc $\widehat{F}(t) : y \mapsto e^{-t/\widehat{\theta}_{(y)}}$.
Comme $\widehat{\theta}(Y) \xrightarrow[n\to+\infty]{p.s.} \theta$ et que $\phi_t : \left(u \mapsto e^{-\frac{t}{u}}\right)$ est absolument continue en u, on a

$$\boxed{\forall t,\ \widehat{F}(t) \xrightarrow[n\to+\infty]{p.s.} \overline{F}(t)}$$

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

> Si $\phi : (\mathbb{R} \to \mathbb{R})$ 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 $\phi_t : \left(u \mapsto e^{-\frac{t}{u}}\right)$ n'est **pas** convexe, puisque $\frac{\partial^2 \phi_t}{\partial u^2}(u) = \frac{t}{u^3}\left(\frac{t}{u} - 2\right) e^{-\frac{t}{u}}$ qui n'est pas toujours positif.
Calculons donc le biais de $\widehat{F}(t)$: on a

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

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

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

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

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

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

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

On a $\sqrt{n}(\widehat{\theta} - \theta) \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}\left(0, I_1(\theta)^{-1}\right)$, 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 $g : \left(\theta \mapsto e^{-\frac{t}{\theta}}\right)$ de dérivée $g'(\theta) = \frac{t}{\theta^2}e^{-\frac{t}{\theta}}$ on a par la delta-méthode

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

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 $S(Y) = Y_1 + \cdots + Y_n$.

☞ 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 dela variable $S_n(Y)$
$Y_1 + \ldots + Y_n$.

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

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

– puis de même

$$\begin{aligned}
f_{S_3(Y)}(s) &= \int_{\mathbb{R}^3} f_{Y_1}(y_1)f_{Y_2}(y_2)f_{Y_3}(y_3)\mathbb{1}_{y_1+y_2+y_3=s}\ dy_1dy_2dy_3 \\
&= \int_{\mathbb{R}^2}\left(\int_{\mathbb{R}^2} f_{Y_1}(y_1)f_{Y_2}(y_2)\mathbb{1}_{y_1+y_2=s_1}\ dy_1dy_2\right) f_{Y_3}(y_3)\mathbb{1}_{s_1+y_3=s}\ ds_1dy_3 \\
&= \int_{\mathbb{R}} \mathbb{1}_{y_3\geq 0}\mathbb{1}_{s-y_3\geq 0} f_{S_2(Y)}(s-y_3)f_{Y_3}(y_3)dy_3 \\
&= \int_{\mathbb{R}} f_{S_2(Y)}(s-y_3)f_{Y_3}(y_3)\,dy_3 \\
&= \mathbb{1}_{s\geq 0}\int_0^s \left(\frac{s-y_3}{\theta^2}e^{-\frac{s-y_3}{\theta}}\right)\left(\frac{1}{\theta}e^{-\frac{y_3}{\theta}}\right)\,dy_3 \\
&= \mathbb{1}_{s\geq 0}\,\frac{s^2}{2}\,\frac{e^{-\frac{s}{\theta}}}{\theta^3}
\end{aligned}$$

– (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 $S(Y) \rightsquigarrow \Gamma(n, \frac{1}{\theta})$ (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

$$\begin{aligned}
f(y_1, s) &= \frac{f_{Y_1,Y_1+\cdots+Y_n}(y_1, s)}{f_{Y_1+\cdots+Y_n}(s)} \\
&= \frac{f_{Y_1,Y_2+\cdots+Y_n}(y_1, s-y_1)}{f_{Y_1+\cdots+Y_n}(s)} \\
&= \frac{f_{Y_1}(y_1)\ f_{Y_2+\cdots+Y_n}(s-y_1)}{f_{Y_1+\cdots+Y_n}(s)}
\end{aligned}$$

car les Y_i sont indépendantes et donc Y_1 est indépendante de $Y_2 + \cdots + Y_n$

$$= \frac{\left(\frac{(s-y_1)^{n-2}}{(n-2)!}\frac{e^{-\frac{s-y_1}{\theta}}}{\theta^{n-1}}\right)\left(\frac{e^{-\frac{y_1}{\theta}}}{\theta}\right)}{\frac{s^{n-1}}{(n-1)!}\frac{e^{-\frac{s}{\theta}}}{\theta^n}}$$

et donc en définitive

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

(b) Calculer

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

Comment s'appelle cet estimateur ?

Calculons tout d'abord que $T^*(Y)$.
On a sucessivement, pour $t \in R^+$ et $s \geq t$

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

En d'autres termes,

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

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 b

On a successivement

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

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

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

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

(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 $\left[\mathbb{E}\left(\mathbb{E}\left(T|S(Y)\right)\right)^2 - \mathbb{E}\left(\left(T|S(Y)\right)^2\right)\right]$ est ardu : il est nécessaire d'expliciter 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

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

En outre, si on peut intervertir $\int_{\mathcal{X}}$ et $\frac{\partial}{\partial\theta}$ alors le score est centré : $\mathbb{E}\left(S(x;\theta)\right) = 0$ (voir exercice 2)
L'information de Fisher pour une observation est définie par

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

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 $\int_{\mathcal{X}}$ et $\frac{\partial}{\partial\theta}$; donner la fonction score du modèle notée $S_Y(y;\theta)$.
En déduire que

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

La variable aléatoire Y ne prend que deux valeurs, et suit donc une loi de Bernoulli ; en ou
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

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

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

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

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 $y \in \{0,1\}$

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

Or

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

et en outre

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

tandis que

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

donc en définitive

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

☞ Q3 En déduire alors que $I_X(\theta) >> I_Y(\theta)$, 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

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

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

$$\mathbb{E}(A) = \mathbb{E}\left(\mathbb{E}_B(A|B)\right)$$

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

En l'occurence il vient

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

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