Semaine 7

Exercice 45.

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

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

Exercice 46.

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

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

et en la posant égale à zéro, nous obtenons

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

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

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

(iii) Nous savons que

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

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

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

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

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

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

Or

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

L'estimateur $\hat{\lambda}_n$ n'atteint donc (tout juste) pas la borne de Cramér-Rao.
Quant à θ, l'information de Fisher $I(\theta)$ est

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

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

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

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

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

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

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

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

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

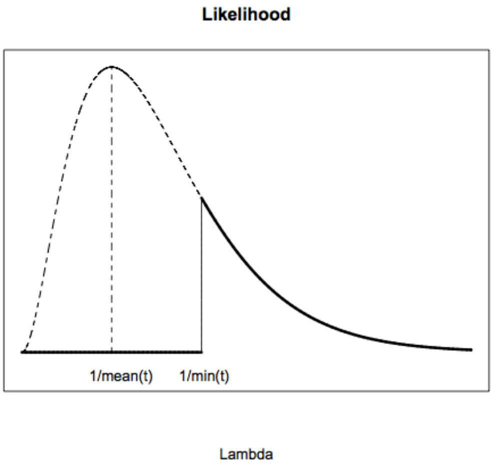
$$\frac{\partial\ell_n}{\partial\lambda}=\frac{n}{\lambda}-n\bar{t}.$$

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

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

Malheureusement, puisque $\bar{t}>t_{(1)}$, $\frac{1}{\bar{t}}<\frac{1}{t_{(1)}}$; notre solution ne satisfait donc pas à la condition $\lambda\ge 1/t_{(1)}$ et la vraisemblance vaut zéro. Puisque ℓ_n (et donc L_n) est décroissante sur $[1/t_{(1)},\infty[$, le maximum sera atteint au premier point où la vraisemblance ne s'annule pas (voir le graphique ci-dessous). L'estimateur est donc $\widehat{\lambda}_n=1/t_{(1)}$.

Remarque. Il se peut que $\bar{t} = t_{(1)}$, mais même dans ce cas l'estimateur sera $1/t_{(1)} = 1/\bar{t}$. Cette particularité n'arrive cependant qu'avec probabilité zéro, à moins que $n = 1$.



Exercice 48.

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

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

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

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

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

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

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

Ainsi le meilleur estimateur de cette forme est

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

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

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

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

Exercice 50.

(i) Remarquons que

$$\ell_n(\theta) = \log f(X_1, \ldots, X_n; \theta) = \eta(\theta) \sum_{i=1}^n T(X_i) - nd(\theta) + \sum_{i=1}^n S(X_i);$$
$$\ell_n'(\theta) = \eta'(\theta) \sum_{i=1}^n T(X_i) - nd'(\theta) = n(\eta'(\theta)\overline{T} - d'(\theta));$$
$$\ell_n''(\theta) = \eta''(\theta) \sum_{i=1}^n T(X_i) - nd''(\theta) = n(\eta''(\theta)\overline{T} - d''(\theta)).$$

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

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

tel que requis.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

avec les derivées partielles

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

En posant les dérivés de égales à zéro, nous obtenons

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

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

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

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