Cursus Intégré
2004-2005

Rappels de statistique mathématique
Corrigé des travaux dirigés n°6

Guillaume Lacôte
Bureau **E03**
Guillaume.Lacote@ensae.fr
http://ensae.no-ip.com/SE222/

Version du 20050421-15h38, révisée le 21 avril 2005

Corrigé de l'exercice 1

☞ Q1 – On a

$$\begin{aligned}
\mathbb{E}\left(X_i\right) &= \sum_{k\in\mathbb{N}}\left(\left(e^{-\lambda}\frac{\lambda^k}{k!}\right)k\right)\\
&= 0+\lambda\underbrace{\left(\sum_{k\in\mathbb{N}^*}e^{-\lambda}\frac{\lambda^{k-1}}{(k-1)!}\right)}_{\sum_{l\in\mathbb{N}}\mathbb{P}(X_1=l)}\\
&= \boxed{\lambda}
\end{aligned}$$

et de même

$$\begin{aligned}
\mathbb{E}\left({X_i}^2\right) &= \sum_{k\in\mathbb{N}^*}\left(\left(e^{-\lambda}\frac{\lambda^k}{k!}\right)k^2\right)\\
&= \left(\sum_{k\in\mathbb{N}^*}\left(e^{-\lambda}\frac{\lambda^k}{k!}k(k-1)\right)\right)+\left(\sum_{k\in\mathbb{N}^*}\left(e^{-\lambda}\frac{\lambda^k}{k!}k\right)\right)\\
&= \left(\underbrace{\sum_{k\geq 2}e^{-\lambda}\frac{\lambda^{k-2}}{(k-2)!}}_{\sum_{l\in\mathbb{N}}\mathbb{P}(X_1=l)}\right)\lambda^2+\left(\underbrace{\sum_{k\geq 1}\frac{\lambda^{k-1}}{(k-1)!}}_{\sum_{l\in\mathbb{N}}\mathbb{P}(X_1=l)}\right)\lambda\\
&= \lambda^2+\lambda
\end{aligned}$$

ce dont on tire

$$\boxed{\mathbb{V}\left(X_i\right)=\lambda}$$

– Soit $\phi:\begin{pmatrix}\mathbb{R} & \rightarrow & \mathbb{R}^{+*}\\ t & \mapsto & \mathbb{E}\left(e^{tX_i}\right)\end{pmatrix}$; ϕ est appelée la *fonction génératrice* de la variable aléatoire X_i.
Alors pour $t\in\mathbb{R}$

$$\begin{aligned}
\phi(t) &= \sum_{k=0}^{+\infty}\mathbb{P}\left(X_i=k\right)e^{tk}\\
&= \sum_{k=0}^{+\infty}e^{-\lambda}\frac{\lambda^k}{k!}e^{tk}\\
&= e^{-\lambda}\sum_{k=0}^{+\infty}\frac{\left(\lambda e^t\right)^k}{k!}\\
&= e^{\lambda(e^t-1)}
\end{aligned}$$

En particulier, ϕ est de classe $C^{+\infty}$.

L'idée ici est que pour $\rho > 0$, la série $\begin{pmatrix}]-\rho,+\rho[& \rightarrow & \mathbb{R}^{+*} \\ t & \mapsto & \sum_{k\in\mathbb{N}} \frac{\left(\lambda e^t\right)^k}{k!} \end{pmatrix}$ est normalement convergente ; donc sur tout voisinage de 0 on peut intervertir $\frac{\partial}{\partial t}$ et $\sum_{\mathbb{N}}$, de sorte que

$$\begin{aligned}
\phi^{(n)}(0) &= \left(\frac{\partial}{\partial t}\left(t \mapsto e^{-\lambda}\sum_{k\in\mathbb{N}}\frac{\left(\lambda e^t\right)^k}{k!}\right)\right)_{(0)} \\
&= e^{-\lambda}\sum_{k\in\mathbb{N}}\left(\frac{\partial}{\partial t}\left(t\mapsto\frac{\left(\lambda e^t\right)^k}{k!}\right)_{(0)}\right) \\
&= e^{-\lambda}\sum_{k\in\mathbb{N}}\left(\left(t\mapsto k^n\frac{\left(\lambda e^t\right)^k}{k!}\right)_{(0)}\right) \\
&= e^{-\lambda}\sum_{k\in\mathbb{N}}k^n\frac{\lambda^k}{k!} \\
&= \sum_{k\in\mathbb{N}}k^n\mathbb{P}\left(X_i = k\right) \\
&= \mathbb{E}\left({X_i}^n\right)
\end{aligned}$$

Ce résultat est en fait vrai plus généralement pour toute densité suffisamment régulière (*i.e.* telle que ϕ soit développable en séries entières et somme de sa série) : en effet on a alors

$$\phi(t) = \sum_{n=0}^{+\infty}\frac{\phi^{(n)}(0)}{n!}t^n$$

Or par ailleurs on a toujours

$$\begin{aligned}
\phi(t) &= \mathbb{E}\left(e^{tX_i}\right) \\
&= \mathbb{E}\left(\sum_{n=0}^{+\infty}\frac{\left(tX_i\right)^n}{n!}\right) \\
&= \sum_{n=0}^{+\infty}\frac{t^n\mathbb{E}\left({X_i}^n\right)}{n!} \quad \text{sous réserve que tous les moments existent} \\
&= \sum_{n=0}^{+\infty}\frac{\mathbb{E}\left({X_i}^n\right)}{n!}t^n
\end{aligned}$$

ce dont on déduit que

$$\boxed{\forall n\in\mathbb{N},\ \phi^{(n)}(0) = \mathbb{E}\left({X_i}^n\right)}$$

– Par ailleurs, on a successivement :

$$
\begin{aligned}
\phi^{'}(t) &= \phi(t)\lambda e^t \\
\phi^{''}(t) &= \phi(t)\lambda e^t + \phi^{'}(t)\lambda e^t \\
&= (\lambda e^t + \left(\lambda e^t\right)^2)\phi(t) \\
\phi^{'''}(t) &= \left(\lambda e^t + 2\left(\lambda e^t\right)^2\right)\phi(t) + \left(\lambda e^t + \left(\lambda e^t\right)^2\right)\lambda e^t \phi^{'}(t) \\
&= \left(\lambda e^t + 3\left(\lambda e^t\right)^2 + \left(\lambda e^t\right)^3\right)\phi(t) \\
\phi^{''''}(t) &= \left(\lambda e^t + 6\left(\lambda e^t\right)^2 + 3\left(\lambda e^t\right)^3\right)\phi(t) + \left(\lambda e^t + 3\left(\lambda e^t\right)^2 + \left(\lambda e^t\right)^3\right)\phi^{'}(t) \\
&= \left(\lambda e^t + 7\left(\lambda e^t\right)^2 + 6\left(\lambda e^t\right)^3 + \left(\lambda e^t\right)^4\right)\phi(t)
\end{aligned}
$$

Plus généralement, on montre qu'il existe une suite de polynômes $(P_n)_{n\in\mathbb{N}} \in \mathbb{R}[\mathbb{X}]^{\mathbb{N}}$ telle que $\forall n \in \mathbb{N},\ \phi^{(n)}(t) = P_n(\lambda e^t)\phi(t)$.
On a en effet $P_0 = (1)$, $P_1 = \mathbb{X}$ puis pour tout $n \in \mathbb{N}$

$$
\begin{aligned}
\phi^{(n+1)}(t) &= \left(P_n^{'}\left(\lambda e^t\right)\lambda e^t\right)\phi(t) + P_n\left(\lambda e^t\right)\phi^{'}(t) \\
&= \left(P_n^{'}\left(\lambda e^t\right)\lambda e^t\right)\phi(t) + P_n\left(\lambda e^t\right)\left(\lambda e^t \phi(t)\right) \\
&= \left(\mathbb{X}P_n^{'}(\mathbb{X}) + \mathbb{X}P_n(\mathbb{X})\right)_{(\lambda e^t)}\phi(t)
\end{aligned}
$$

Par suite

$$\forall n \in \mathbb{N},\ P_{n+1}(\mathbb{X}) = \mathbb{X}\left(P_n(\mathbb{X}) + P_n^{'}(\mathbb{X})\right)$$

de sorte que, en notant $P_n(\mathbb{X}) = \sum_{k=0}^{n} a_n^k \mathbb{X}^k$

$$\forall n \in \mathbb{N}, \forall k \in [\![0, n]\!],\ a_{n+1}^k = k a_n^k + a_n^{k-1} \quad \text{(avec la convention } a_n^{-1} = 0)$$

ce qui permet de calculer les $(a_n^k)_{n,k}$ et d'en déduire $\mathbb{E}\left(X_i{}^n\right) = \phi^{(n)}(0)$. Puisque $\phi(0) = 1$, on peut donc facilement calculer *pour tout* $n \in \mathbb{N}^*$ le moment d'ordre n

$$\boxed{\mathbb{E}\left(T_i{}^k\right) = P_n(\lambda)}$$

– Dans le cas présent on vérifie immédiatement que

$$
\begin{cases}
\mathbb{E}(X) &= \lambda \\
\mathbb{E}(X^2) &= \lambda(1+\lambda) \\
\mathbb{E}(X^3) &= \lambda(1+3\lambda+\lambda^2) \\
\mathbb{E}(X^4) &= \lambda(1+7\lambda+6\lambda^2+\lambda^3)
\end{cases}
$$

☞ Q2 (a) On constate que $\widehat{\lambda_1}$ et $\widehat{\lambda_2}$ ont été obtenus par la méthode des moments, qui consiste pour estimer λ à inverser l'expression du moment théorique d'une variable X en fonction de λ, et de l'appliquer au moment empirique.

En l'occurence $\widehat{\lambda^1}(x) = \bar{x}$ est le premier moment empirique, qui converge p.s. vers le premier moment théorique $\mathbb{E}(X_1) = \lambda$. Donc $\widehat{\lambda^1}$ est l'estimateur fondé sur le premier moment de X. De même $\widehat{\lambda^2}(x) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$ est le moment centré d'ordre 2 empirique, qui converge vers le moment centré théorique d'ordre 2 $\mathbb{V}(X_1) = \lambda$. Donc $\widehat{\lambda^2}$ est l'estimateur fondé sur le second moment centré de X.

Les questions suivantes s'assurent que $\widehat{\lambda_1}$ et $\widehat{\lambda_2}$ sont bien des estimateurs de λ et étudient leurs propriétés à distance (in-)finie.

(b) On a

$$\begin{aligned}
\mathbb{E}\left(\widehat{\lambda_1}(X)\right) &= \mathbb{E}\left(\bar{X}\right) \\
&= \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left(X_i\right) \\
&= \boxed{\lambda}
\end{aligned}$$

puis

$$\begin{aligned}
\mathbb{E}\left(\widehat{\lambda_2}(X)\right) &= \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left(\left(X_i - \bar{X}\right)^2\right) \\
&= \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left({X_i}^2 - 2X_i\bar{X} + \bar{X}^2\right) \\
&= \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left({X_i}^2\right) - 2\mathbb{E}\left(\bar{X}^2\right) + \mathbb{E}\left(\bar{X}^2\right) \\
&= \mathbb{E}\left({X_1}^2\right) - \mathbb{E}\left(\bar{X}^2\right) \\
&= \mathbb{E}\left({X_1}^2\right) - \left(\mathbb{V}\left(\bar{X}\right) + \mathbb{E}\left(\bar{X}\right)^2\right) \\
&= \mathbb{E}\left({X_1}^2\right) - \left(\frac{1}{n}\mathbb{V}\left(X_1\right) + \mathbb{E}\left(X_1\right)^2\right) \\
&= \left(\lambda^2 + \lambda\right) - \left(\frac{1}{n}\lambda + \lambda^2\right) \\
&= \boxed{\tfrac{n-1}{n}\lambda}
\end{aligned}$$

Ainsi, $\widehat{\lambda_1}$ n'est pas biaisé mais $\widehat{\lambda_2}$ l'est, et on peut proposer $\widehat{\lambda_3} = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2$

(c) <u>*Première méthode :*</u> Cherchons la loi de $\sum_{i=1}^{n} X_i$.

Une façon naturelle serait de montrer qu'une somme de variables indépendantes qui suivent chacune un loi de Poisson de paramètre λ_i suit elle-même un loi de Poisson de paramètre

$\sum_{i=1}^{n} \lambda_i$; ceci se montre par récurrence soit en calculant explicitement

$$\mathbb{P}(X_1 + \cdots + X_n = k) = \sum_{s=0}^{k} \mathbb{P}(X_1 + \cdots + X_{n-1} = s) \cdot \mathbb{P}(X_n = k - s)$$

soit par convolution.
Une autre façon est de calculer la fonction génératrice de la variable $\sum_{i=1}^{n} X_i$, et de reconnaître une fonction génératrice connue : la bijectivité du passage entre une densité et une fonction génératrice permet alors de conclure que $\sum_{i=1}^{n} X_i$ suit bien la loi dont on a reconnu la fonction génératrice.
En l'occurence

$$\begin{aligned}
\phi_{\sum_{i=1}^{n} X_i}(t) &= \mathbb{E}\left(e^{t\sum_{i=1}^{n} X_i}\right) \\
&= \mathbb{E}\left(\Pi_{i=1}^{n} e^{tX_i}\right) \\
&= \Pi_{i=1}^{n} \mathbb{E}\left(e^{tX_i}\right) \quad \text{par indépendance des } \left(e^{tX_i}\right)_i \\
&= \Pi_{i=1}^{n} \phi_{X_i}(t) \\
&= e^{n\lambda\left(e^t - 1\right)} \\
&= \phi_{\mathcal{P}(n\lambda)}(t)
\end{aligned}$$

ce dont on conclut que $\sum_{i=1}^{n} X_i \rightsquigarrow \mathcal{P}(n\lambda)$.
En définitive la loi de $\widehat{\lambda_1}$ est caractérisée par

$$\forall k \in \mathbb{Q},\ \mathbb{P}\left(\widehat{\lambda_1} = k\right) = \mathbb{P}\left(\sum_{i=1}^{n} X_i = nk\right) = e^{-n\lambda} \frac{(n\lambda)^{nk}}{(nk)!} \mathbb{1}_{nk \in \mathbb{N}}$$

(on notera que la statistique $\widehat{\lambda_1}$ n'est **pas** entière a priori, ce qui justifie l'indicatrice).

<u>*Autre Méthode :*</u> On peut aussi directement calculer pour $t \in \mathbb{N}$

$$\begin{aligned}
\mathbb{P}\left(\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{t}{n}\right) &= \mathbb{P}\left(X_1 + \cdots + X_n = t\right) \\
&= \sum_{x_1+\cdots+x_n=t} \mathbb{P}\left(X_1 = x_1 \wedge \cdots \wedge X_n = x_n\right) \\
&= \sum_{x_1+\cdots+x_n=t} \prod_{i=1}^{n} \mathbb{P}\left(X_i = x_i\right) \quad \text{car les } (X_i)_i \text{ sont indépendants} \\
&= \sum_{x_1+\cdots+x_n=t} \prod_{i=1}^{n} \frac{e^{-\lambda}\lambda^{x_i}}{x_i!} \\
&= \sum_{x_1+\cdots+x_n=t} \frac{e^{-n\lambda}\lambda^{x_1+\cdots+x_n}}{x_1!\cdots x_n!} \\
&= \frac{e^{-n\lambda}\lambda^t}{t!} \sum_{x_1+\cdots+x_n=t} \frac{t!}{x_1!\cdots x_n!} \\
&= \frac{e^{-n\lambda}\lambda^t}{t!} \left(1 + \cdots + 1\right)^t \\
&= \boxed{\frac{e^{-n\lambda}(n\lambda)^t}{t!}}
\end{aligned}$$

(d) Pour déterminer la loi limite jointe de $\begin{pmatrix} \widehat{\lambda}_1 \\ \widehat{\lambda}_2 \end{pmatrix}$, on serait tenté de poser $Z_i = \begin{pmatrix} X_i \\ \left(X_i - \bar{X}\right)^2 \end{pmatrix}$ et d'appliquer le Théorème Central Limite à Z_i ; la difficulté est que les Z_i ne sont plus i.i.d dans ce cas, du fait de $\bar{X}$. C'est pourquoi on pose $Y_i = \begin{pmatrix} X_i \\ {X_i}^2 \end{pmatrix}$, et on constate que

$$\sqrt{n}\left(Y_i - \mathbb{E}\left(Y_i\right)\right) \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \mathbb{V}\left(Y_i\right)\right)$$

ce qui s'écrit

$$\sqrt{n}\left(\begin{pmatrix} \bar{X} \\ \bar{X^2} \end{pmatrix} - \begin{pmatrix} \lambda \\ \lambda(1+\lambda) \end{pmatrix}\right) \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \mathbb{V}\left(X_i\right) & \mathbb{C}ov\left(X_i, {X_i}^2\right) \\ \mathbb{C}ov\left(X_i, {X_i}^2\right) & \mathbb{V}\left({X_i}^2\right) \end{pmatrix}\right)$$

Par ailleurs on calcule successivement

$$\begin{array}{rcl}\mathbb{V}(X_i) & = & \lambda \\ \mathbb{V}\left(X_i^2\right) & = & \mathbb{E}\left(X_i^4\right) - \mathbb{E}\left(X_i^2\right) \\ & = & \lambda(1+6\lambda+4\lambda^2) \\ \mathbb{C}ov\left(X_i, X_i^2\right) & = & \mathbb{E}\left(X_i^3\right) - \mathbb{E}(X_i)\mathbb{E}\left(X_i^2\right) \\ & = & \lambda(1+2\lambda)\end{array}$$

Enfin, posons $g : \left(\begin{array}{ccc}\mathbb{R}^2 & \to & \mathbb{R}^2 \\ \begin{pmatrix} u \\ v \end{pmatrix} & \mapsto & \begin{pmatrix} u \\ v-u^2 \end{pmatrix}\end{array}\right)$, de sorte que $\frac{1}{n}\sum_{i=1}^n Z_i = g\left(\frac{1}{n}\sum_{i=1}^n Y_i\right)$, car $\frac{1}{n}\sum_{i=1}^n \left(X_i - \bar{X}\right)^2 = \frac{1}{n}\sum_{i=1}^n X_i^2 - \left(\bar{X}\right)^2$.
Alors g est de classe C^∞, et de plus $Jac\, g\begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ -2u & 1 \end{pmatrix}$, donc d'après le théorème de Slutsky

$$\sqrt{n}\left(g\begin{pmatrix} \frac{1}{n}\sum_{i=1}^n X_i \\ \frac{1}{n}\sum_{i=1}^n X_i^2 \end{pmatrix} - g\begin{pmatrix} \lambda \\ \lambda(1+\lambda) \end{pmatrix}\right)$$
$$\xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ -2\lambda & 1 \end{pmatrix} \times \begin{pmatrix} \lambda & \lambda(1+2\lambda) \\ \lambda(1+2\lambda) & \lambda(1+\lambda) \end{pmatrix} \times \begin{pmatrix} 1 & -2\lambda \\ 0 & 1 \end{pmatrix}\right)$$

soit

$$\sqrt{n}\left(\begin{pmatrix} \widehat{\lambda_1} \\ \widehat{\lambda_2} \end{pmatrix} - \begin{pmatrix} \lambda \\ \lambda \end{pmatrix}\right) \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \lambda & \lambda \\ \lambda & \lambda(1+2\lambda) \end{pmatrix}\right)$$

Enfin, remarquant que

$$\begin{pmatrix} \widehat{\lambda_1} \\ \widehat{\lambda_3} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & \frac{n}{n-1} \end{pmatrix}\begin{pmatrix} \widehat{\lambda_1} \\ \widehat{\lambda_2} \end{pmatrix} \xrightarrow[n\to+\infty]{\mathcal{L}} \begin{pmatrix} \lambda \\ \lambda \end{pmatrix}$$

il vient

$$\sqrt{n}\left(\begin{pmatrix} \widehat{\lambda_1} \\ \widehat{\lambda_3} \end{pmatrix} - \begin{pmatrix} \lambda \\ \frac{n}{n-1}\lambda \end{pmatrix}\right) \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \lambda & \lambda \\ \lambda & \lambda(1+2\lambda) \end{pmatrix}\right)$$

et comme par ailleurs $\frac{n}{n-1}\lambda \xrightarrow[n\to+\infty]{\mathcal{L}} \lambda$ on conclut que

$$\sqrt{n}\left(\begin{pmatrix} \widehat{\lambda_1} \\ \widehat{\lambda_3} \end{pmatrix} - \begin{pmatrix} \lambda \\ \lambda \end{pmatrix}\right) \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \lambda & \lambda \\ \lambda & \lambda(1+2\lambda) \end{pmatrix}\right)$$

(e) On a $\mathbb{E}\left(\widehat{\lambda_\infty}(X)\right) = \alpha\mathbb{E}\left(\widehat{\lambda_1}(X)\right) + \beta\mathbb{E}\left(\widehat{\lambda_3}(X)\right) = (\alpha+\beta)\lambda$,
donc $\widehat{\lambda_\infty}$ est sans biais *ssi* $\alpha + \beta = 1$.

Par ailleurs, $\widehat{\lambda_\infty}$ est le meilleur estimateur sans biais de λ s'il est de variance minimale ; or pour $n \to +\infty$

$$\begin{aligned}
\mathbb{V}\left(\widehat{\lambda_\infty}(X)\right) &= \alpha^2\mathbb{V}\left(\widehat{\lambda_1}(X)\right) + 2\alpha(1-\alpha)\mathbb{C}ov\left(\widehat{\lambda_1}(X), \widehat{\lambda_3}(X)\right) + (1-\alpha)^2\mathbb{V}\left(\widehat{\lambda_3}(X)\right) \\
&= \begin{cases} \left(\mathbb{V}\left(\widehat{\lambda_1}(X)\right) + \mathbb{V}\left(\widehat{\lambda_3}(X)\right) - 2\mathbb{C}ov\left(\widehat{\lambda_1}(X), \widehat{\lambda_3}(X)\right)\right)\alpha^2 \\ \qquad +2\left(\mathbb{C}ov\left(\widehat{\lambda_1}(X), \widehat{\lambda_3}(X)\right) - \mathbb{V}\left(\widehat{\lambda_3}\right)(X)\right)\alpha \\ \qquad\qquad + \left(\mathbb{V}\left(\widehat{\lambda_3}\right)(X)\right) \end{cases}
\end{aligned}$$

Ce polynôme de degré 2 en α a un cœfficient $\mathbb{V}\left(\widehat{\lambda_1}\right)(X)+\mathbb{V}\left(\widehat{\lambda_3}\right)(X)-2\mathbb{C}ov\left(\widehat{\lambda_1}(X), \widehat{\lambda_3}\right)(X) = \mathbb{V}\left(\widehat{\lambda_1}(X) - \widehat{\lambda_3}(X)\right) \geq 0$ en α^2, donc est maximal aux bornes et minimal au milieu des racines, soit en

$$\begin{aligned}
\alpha^* &= -\frac{2\left(\mathbb{C}ov\left(\widehat{\lambda_1}(X), \widehat{\lambda_3}(X)\right) - \mathbb{V}\left(\widehat{\lambda_3}(X)\right)\right)}{2 \times \left(\mathbb{V}\left(\widehat{\lambda_1}(X)\right) + \mathbb{V}\left(\widehat{\lambda_3}(X)\right) - 2\mathbb{C}ov\left(\widehat{\lambda_1}(X), \widehat{\lambda_3}(X)\right)\right)} \\
&= -\frac{2n\left(\lambda - \lambda(1+2\lambda)\right)}{2n\left(\lambda + \lambda(1+2\lambda) - 2\lambda\right)} \\
&= -\frac{2\left(\lambda - \lambda - 2\lambda^2\right)}{2\left(\lambda + \lambda + 2\lambda^2 - 2\lambda\right)} \\
&= 1
\end{aligned}$$

Ainsi, $\widehat{\lambda_\infty}^* = \widehat{\lambda_1}$, de sorte que $\widehat{\lambda_1} = (x \mapsto \bar{x}) = \widehat{\lambda_{emv}}$ est le meilleur estimateur asymptotiquement sans biais de λ.

☞ Q3 (a) Rappel : l'erreur quadratique moyenne de l'estimateur $\widehat{\theta}$ de θ est définie comme $\mathbb{E}\left(\left(\widehat{\theta} - \theta\right)^2\right)$.

– On a $\widehat{\lambda_1} = \frac{1}{2}(X_1 + X_2)$, de sorte que

$$\begin{aligned}
\mathbb{E}\left(\left(\widehat{\lambda_1} - \lambda\right)^2\right) &= \mathbb{E}\left(\widehat{\lambda_1}^2\right) - \lambda^2 \\
&= \frac{1}{4}\left(\mathbb{E}\left({X_1}^2\right) + 2\mathbb{C}ov\left(X_1, X_2\right) + \mathbb{E}\left({X_2}^2\right)\right) - \lambda^2 \\
&= \frac{1}{4}\left(2\lambda(1+\lambda) + 2\lambda^2\right) - \lambda^2 \\
&= \frac{\lambda}{2}
\end{aligned}$$

– De façon similaire

$$
\begin{aligned}
\widehat{\lambda_2} &= \frac{1}{2}\left(\left(X_1 - \frac{X_1+X_2}{2}\right)^2 + \left(X_2 - \frac{X_1+X_2}{2}\right)^2\right) \\
&= \frac{1}{2}\left({X_1}^2 - X_1(X_1+X_2) + \frac{(X_1+X_2)^2}{4} + {X_2}^2 - X_2(X_1+X_2) + \frac{(X_1+X_2)^2}{4}\right) \\
&= \frac{1}{4}\left({X_1}^2 + {X_2}^2 - 2X_1X_2\right)
\end{aligned}
$$

de sorte que

$$
\mathbb{E}\left(\left(\widehat{\lambda_2} - \lambda\right)^2\right) = \mathbb{E}\left(\widehat{\lambda_2}^2\right) - 2\lambda\mathbb{E}\left(\widehat{\lambda_2}\right) + \lambda^2
$$

Or $\mathbb{E}\left(\widehat{\lambda_2}\right) = \frac{\lambda}{2}$, et en outre

$$
\begin{aligned}
\mathbb{E}\left(\widehat{\lambda_2}^2\right) &= \frac{1}{16}\mathbb{E}\left(\left({X_1}^2 + {X_2}^2 - 2X_1X_2\right)^2\right) \\
&= \frac{1}{16}\mathbb{E}\left(({X_1}^2 + {X_2}^2)^2 - X_1X_2({X_1}^2 + {X_2}^2) + 4{X_1}^2{X_2}^2\right) \\
&= \frac{1}{16}\mathbb{E}\left({X_1}^4 + 2{X_1}^2{X_2}^2 + {X_2}^4 + -{X_1}^3X_2 - X_1{X_2}^3 + 4{X_1}^2{X_2}^2\right) \\
&= \frac{1}{8}\left(\mathbb{E}\left({X_1}^4\right) + 3\mathbb{E}\left({X_1}^2\right)\mathbb{E}\left({X_2}^2\right) - 4\mathbb{E}\left({X_1}^3\right)\mathbb{E}\left(X_2\right)\right) \quad \text{par indépendance} \\
&= \frac{1}{8}\left(\lambda + 6\lambda^2\right)
\end{aligned}
$$

et donc

$$
\begin{aligned}
\mathbb{E}\left(\left(\widehat{\lambda_2} - \lambda\right)^2\right) &= \frac{1}{8}\left(\lambda + 6\lambda^2\right) - 2\lambda\left(\frac{\lambda}{2}\right) + \lambda^2 \\
&= \frac{3}{4}\lambda^2 + \frac{1}{8}\lambda
\end{aligned}
$$

– Enfin, $\widehat{\lambda_3} = 2\widehat{\lambda_2}$ et donc

$$
\begin{aligned}
\mathbb{E}\left(\left(\widehat{\lambda_3} - \lambda\right)^2\right) &= \mathbb{E}\left(\widehat{\lambda_3}^2\right) - \lambda^2 \\
&= 4\mathbb{E}\left(\widehat{\lambda_2}^2\right) - \lambda^2 \\
&= 2\lambda^2 + \frac{1}{2}\lambda
\end{aligned}
$$

(b) Au sens de l'erreur quadratique moyenne, $\widehat{\lambda_1}$ est toujours préférable à $\widehat{\lambda_3}$.
De plus,

$$\begin{aligned}
\widehat{\lambda_2} \text{ est préférable à } \widehat{\lambda_3} &\Leftrightarrow \mathbb{E}\left(\left(\widehat{\lambda_2}-\lambda\right)^2\right) < \mathbb{E}\left(\left(\widehat{\lambda_3}-\lambda\right)^2\right) \\
&\Leftrightarrow \left(2\lambda^2+\frac{1}{2}\lambda\right) - \left(\frac{3}{4}\lambda^2+\frac{1}{8}\lambda\right) > 0 \\
&\Leftrightarrow \left(2-\frac{3}{4}\right)\lambda^2 + \left(\frac{1}{2}-\frac{1}{8}\right)\lambda > 0 \\
&\Leftrightarrow \lambda > -\frac{3}{8} \quad \text{cd qui est toujours vrai car } \lambda > 0
\end{aligned}$$

et de façon analogue

$$\begin{aligned}
\widehat{\lambda_2} \text{ est préférable à } \widehat{\lambda_1} &\Leftrightarrow \mathbb{E}\left(\left(\widehat{\lambda_2}-\lambda\right)^2\right) < \mathbb{E}\left(\left(\widehat{\lambda_1}-\lambda\right)^2\right) \\
&\Leftrightarrow \left(\frac{\lambda}{2}\right) - \left(\frac{3}{4}\lambda^2+\frac{1}{8}\lambda\right) > 0 \\
&\Leftrightarrow \lambda < \frac{1}{2}
\end{aligned}$$

(c) On a $L(k,\lambda) = e^{-\lambda}\frac{\lambda^k}{k!}$, donc $\frac{\partial \ln L}{\partial \lambda}(k,\lambda) = -1 + \frac{k}{\lambda}$ et par suite $\frac{\partial^2 \ln L}{\partial \lambda}(k,\lambda) = -\frac{k}{\lambda^2}$. donc

$$\begin{aligned}
I_1(\lambda) &= -\mathbb{E}\left(\frac{\partial^2 \ln L}{\partial \lambda}(k,\lambda)\right) \\
&= \sum_{k=0}^{+\infty} \frac{k}{\lambda^2} e^{-\lambda}\frac{\lambda^k}{k!} \\
&= \frac{e^{-\lambda}}{\lambda}\sum_{k=1}^{+\infty}\frac{\lambda^{k-1}}{(k-1)!} \\
&= \frac{1}{\lambda}
\end{aligned}$$

de sorte que la borne FDCR est $\frac{\lambda}{2}$. En particulier, $\widehat{\lambda_3}$ est efficace *ssi* $2\lambda^2 + \frac{1}{2}\lambda = \frac{\lambda}{2}$, c'est-à-dire jamais puisque $\lambda > 0$.

⋆
⋆ ⋆

Corrigé de l'exercice 2

☞ Q1 Connaissant les lois de $T_i|\Lambda_i$ et de Λ_i, cherchons celle de T_i : on a pour $t \in \mathbb{R}$

$$\begin{aligned} f_{T_i}(t) &= \int_{\mathbb{R}} f_{T_i|\Lambda_i=\lambda}(t) f_{\Lambda_i}(\lambda) d\lambda \\ &= \int_{\mathbb{R}} \left(\lambda e^{-\lambda t} \mathbb{1}_{t\geq 0}\right) \left(\frac{\alpha^r}{\Gamma(r)} \lambda^{r-1} e^{-\alpha\lambda} \mathbb{1}_{\lambda \geq 0}\right) d\lambda \\ &= \frac{\alpha^r}{\Gamma(r)} \left(\int_0^{+\infty} \lambda^r e^{-(\alpha+t)\lambda} d\lambda\right) \mathbb{1}_{t\geq 0} \end{aligned}$$

Or par définition de la fonction Γ

$$\int_0^{+\infty} \lambda^r (\alpha+t)^{r+1} e^{-(\alpha+t)\lambda} d\lambda = \Gamma(r+1)$$

donc

$$f_{T_i}(t) = \frac{\alpha^r}{\Gamma(r)} \frac{\Gamma(r+1)}{(\alpha+t)^{r+1}} \mathbb{1}_{t\geq 0}$$

Or $\forall x \in \mathbb{R}^{+*}, \Gamma(x+1) = x\Gamma(x)$, donc finalement

$$f_{T_i}(t) = \frac{r\alpha^r}{(\alpha+t)^{r+1}} \mathbb{1}_{t\geq 0}$$

Les $(T_i)_i$ étant indépendants il vient en conséquence

$$\boxed{L_{T_1,\ldots,T_n}(t_1,\ldots,t_n;\alpha,r) = r^n \alpha^{nr} \frac{1}{\left(\Pi_{i=1}^n (\alpha+t_i)\right)^{r+1}} \mathbb{1}_{\min_i t_i \geq 0}}$$

☞ Q2 Pour $k \in \mathbb{N}^*$ on a

$$\begin{aligned} \mathbb{E}\left(T_i^k\right) \text{ existe} &\Leftrightarrow \int_0^{+\infty} t^k \frac{r\alpha^r}{(\alpha+t)^{r+1}} dt \quad \text{converge} \\ &\Leftrightarrow r+1-k > 1 \\ &\Leftrightarrow k < r \end{aligned}$$

Soit donc $k \in [\![1, r-1]\!]$, et notons $M_k^r = \mathbb{E}\left(T_i^k\right) < +\infty$. Alors

$$\begin{aligned} M_k^r &= \int_0^{+\infty} \left(\frac{r\alpha^r}{(\alpha+t)^{r+1}}\right) \left(t^k\right) dt \\ &\quad \text{par parties : } u = -\frac{\alpha^r}{(\alpha+t)^r} \text{ et } v = t^k \\ &= \left[\left(-\frac{\alpha^r}{(\alpha+t)^r} t^k\right)\right]_0^{+\infty} - \int_0^{+\infty} -\frac{\alpha^r}{(\alpha+t)^r} k t^{k-1} dt \\ &= +\frac{k\alpha}{r-1} \int_0^{+\infty} \frac{(r-1)\alpha^{r-1}}{(\alpha+t)^{(r-1)+1}} t^{k-1} dt \\ &= \frac{k\alpha}{r-1} M_{k-1}^{r-1} \end{aligned}$$

Donc par une récurrence immédiate pour $r \geq 2$ et $k < r$

$$\begin{aligned} M_k^r &= \alpha^k \frac{k!}{\Pi_{i=1}^k (r-i)} M_0^{r-k} \\ &= \alpha^k \frac{k!}{\frac{(r-1)!}{(r-k-1)!}} \quad \text{car } M_0^{r-k} = 1 \end{aligned}$$

c'est-à-dire (en notant $\mathcal{C}_n^p$ le cœfficient binômial $\frac{n!}{p!(n-p)!}$)

$$\boxed{\mathbb{E}\left(T_i{}^k\right) = \frac{\alpha^k}{\mathcal{C}_{r-1}^k}}$$

On notera en particulier que lorsque $r > 2$

$$\begin{aligned} \mathbb{E}\left(T_i\right) &= \frac{\alpha}{r-1} \\ \mathbb{E}\left(T_i{}^2\right) &= \frac{2\alpha^2}{(r-1)(r-2)} \\ \mathbb{V}\left(T_i\right) &= \frac{2\alpha^2}{(r-1)(r-2)} - \left(\frac{\alpha}{r-1}\right)^2 \\ &= \frac{r\alpha^2}{(r-1)^2(r-2)} \end{aligned}$$

☞ Q3 – La vraisemblance s'écrit pour $t_1, \dots, t_n > 0$

$$\ln L_{T_1,\dots,T_n}(t_1, \dots, t_n; \alpha, r) = n \ln r + nr \ln \alpha - (r+1) \sum_{i=1}^n \ln(\alpha + t_i)$$

et donc

$$\begin{aligned} \frac{\partial \ln L_{T_1,\dots,T_n}}{\partial r}(t_1, \dots, t_n; \alpha, r) &= \frac{n}{r} + n \ln \alpha - \sum_{i=1}^n \ln(\alpha + t_i) \\ \frac{\partial \ln L_{T_1,\dots,T_n}}{\partial \alpha}(t_1, \dots, t_n; \alpha, r) &= \frac{nr}{\alpha} - (r+1) \sum_{i=1}^n \frac{1}{\alpha + t_i} \end{aligned}$$

puis

$$\begin{aligned} \frac{\partial^2 \ln L_{T_1,\dots,T_n}}{\partial r^2}(t_1, \dots, t_n; \alpha, r) &= -\frac{n}{r^2} \\ \frac{\partial^2 \ln L_{T_1,\dots,T_n}}{\partial r \partial \alpha}(t_1, \dots, t_n; \alpha, r) &= \frac{n}{\alpha} - \sum_{i=1}^n \frac{1}{\alpha + t_i} \\ \frac{\partial^2 \ln L_{T_1,\dots,T_n}}{\partial \alpha^2}(t_1, \dots, t_n; \alpha, r) &= -\frac{nr}{\alpha^2} + (r+1) \sum_{i=1}^n \frac{1}{(\alpha + t_i)^2} \end{aligned}$$

Pour calculer la matrice d'information de Fisher il reste à calculer

$$\begin{aligned}
\mathbb{E}\left(\frac{1}{(\alpha+t_i)^p}\right) &= \int_0^{+\infty} \frac{r\alpha^r}{(\alpha+t_i)^{r+p-1}}dt \\
&= r\alpha^r\left[-\frac{1}{(r+p)}(\alpha+t_i)^{-(r+p-1)-1}\right]_0^{+\infty} \\
&= \frac{r}{(r+p)\alpha^p}
\end{aligned}$$

En conséquence on a

$$\begin{aligned}
\mathbb{E}\left(\frac{\partial^2 \ln L_{T_1,\ldots,T_n}}{\partial r^2}(t_1,\ldots,t_n;\alpha,r)\right) &= -\frac{n}{r^2} \\
\mathbb{E}\left(\frac{\partial^2 \ln L_{T_1,\ldots,T_n}}{\partial r\partial\alpha}(t_1,\ldots,t_n;\alpha,r)\right) &= \frac{n}{\alpha}-\frac{nr}{(r+1)\alpha} \\
&= \frac{n}{\alpha(r+1)} \\
\mathbb{E}\left(\frac{\partial^2 \ln L_{T_1,\ldots,T_n}}{\partial \alpha^2}(t_1,\ldots,t_n;\alpha,r)\right) &= -\frac{nr}{\alpha^2}+n(r+1)\frac{r}{(r+2)\alpha^2} \\
&= -\frac{nr}{(r+2)\alpha^2}
\end{aligned}$$

et finalement

$$\boxed{I_1(r,\alpha) = \begin{pmatrix} \frac{1}{r^2} & -\frac{1}{\alpha(r+1)} \\ -\frac{1}{\alpha(r+1)} & \frac{r}{(r+2)\alpha^2} \end{pmatrix}}$$

– Lorsque α est connu, le modèle est paramétré par r uniquement et la vraisemblance s'écrit pour $t_1,\ldots,t_n>0$

$$\ln L_{T_1,\ldots,T_n}(t_1,\ldots,t_n;r) = n\ln r + nr\ln\alpha - (r+1)\sum_{i=1}^n \ln(\alpha+t_i)$$

et donc l'estimateur du maximume de vraisemblance $\widehat{r}$ de r vérifie

$$\frac{n}{\widehat{r}} + n\ln\alpha - \sum_{i=1}^n \ln(\alpha+t_i) = 0$$

d'où on tire

$$\widehat{r} = \frac{1}{\frac{1}{n}\sum_{i=1}^n \ln(\alpha+t_i) - \ln\alpha}$$

(la réciproque étant immédiate par concavité).

– Mais dans le cas où le modèle est paramétré (α, r) l'estimateur du maximum de vraisemblance $(\widehat{\alpha}, \widehat{r})$ du paramètre vérifie

$$\left\{\begin{array}{rcl} \widehat{r} & = & \frac{n}{\sum_{i=1}^n \ln(\widehat{\alpha}+t_i) - n\ln\widehat{\alpha}} \\ \widehat{r} & = & \frac{\sum_{i=1}^n \frac{1}{\widehat{\alpha}+t_i}}{\frac{n}{\widehat{\alpha}} - \sum_{i=1}^n \frac{1}{\widehat{\alpha}+t_i}} \end{array}\right.$$

dont la résolution semble hors de portée (*i.e.* $\widehat{\alpha}$ n'admet pas d'expression analytique, sauf dans le cas dégénéré où $t = (0)$).

☞ Q4 Constatant que $\bar{t} \xrightarrow[n\to+\infty]{p.s.} \mathbb{E}(T_i) = \frac{\alpha}{r-1}$, posons $\widehat{r} = 1 + \frac{\alpha}{\bar{t}}$; alors $\widehat{r}$ est un estimateur convergent (presque sûr) de r puisque $\left(u \mapsto \frac{1}{u}\right)$ est continue.

Cependant, à distance finie on observe que

$$\begin{array}{rcll} \mathbb{E}(\widehat{r}) & = & \mathbb{E}\left(1 + \frac{\alpha}{\bar{T}}\right) & \\ & > & 1 + \frac{\alpha}{\mathbb{E}(\bar{T})} & \text{d'après Jensen} \\ & = & r & \end{array}$$

de sorte que $\widehat{r}$ est biaisé et sur-estime systématiquement r (et n'est a fortiori pas efficace à distance finie).

Enfin, appliquant le Théorème Central Limite à $\bar{T}$ il vient

$$\sqrt{n}\left(\bar{T} - \mathbb{E}(\bar{T})\right) \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}(0, \mathbb{V}(T_1))$$

car $T_1, \ldots, T_n$ sont indépendants.

Soit alors $g : \left(\begin{array}{ccc} \mathbb{R}^{+*} & \to & \mathbb{R}^{+*} \\ s & \mapsto & 1 + \frac{\alpha}{s} \end{array}\right)$; g est de classe C^∞ et donc d'après Slutsky

$$\sqrt{n}\left(g(\bar{T}) - g\left(\mathbb{E}(\bar{T})\right)\right) \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}\left(0, g'\left(\frac{\alpha}{r-1}\right)^2 \left(\frac{r\alpha^2}{(r-1)^2(r-2)}\right)\right)$$

soit

$$\sqrt{n}\left(\widehat{r} - r\right) \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}\left(0, \frac{r(r-1)^2}{r-2}\right)$$

Constatant finalement que $\frac{r(r-1)^2}{r-2} > r^2$ pour $r > 2$, on conclut que $\widehat{r}$ n'est **pas** asymptotiquement efficace.

☞ Q5 (a) Observant la réalisation empirique $\left(\begin{array}{c} m_1 \\ m_2 \end{array}\right) = \left(\begin{array}{c} \frac{1}{n}\sum_{i=1}^n t_i \\ \frac{1}{n}\sum_{i=1}^n {t_i}^2 \end{array}\right)$ des moments théoriques $\left(\begin{array}{c} \mathbb{E}(T_i) \\ \mathbb{E}({T_i}^2) \end{array}\right) = \left(\begin{array}{c} \frac{\alpha}{r-1} \\ \frac{2\alpha^2}{(r-1)(r-2)} \end{array}\right)$, on détermine $(\widetilde{\alpha}, \widetilde{r})$ de façon à ce que

$$\left(\begin{array}{c} \frac{1}{n}\sum_{i=1}^n t_i \\ \frac{1}{n}\sum_{i=1}^n {t_i}^2 \end{array}\right) = \left(\begin{array}{c} \frac{\widetilde{\alpha}}{\widetilde{r}-1} \\ \frac{2\widetilde{\alpha}^2}{(\widetilde{r}-1)(\widetilde{r}-2)} \end{array}\right)$$

En l'occurence on pose

$$\boxed{\left\{\begin{array}{lcl}\widetilde{\alpha} & = & \frac{\bar{t}\ \bar{t^2}}{\bar{t^2}-2(\bar{t})^2}\\ \widetilde{r} & = & \frac{2\left(\bar{t^2}-(\bar{t})^2\right)}{\bar{t^2}-2(\bar{t})^2}\end{array}\right.} \quad \text{en notant} \left\{\begin{array}{lcl}\bar{t} & = & \frac{1}{n}\sum_{i=1}^n t_i\\ \bar{t^2} & = & \frac{1}{n}\sum_{i=1}^n {t_i}^2 \neq (\bar{t})^2\end{array}\right.$$

ce qui assure que $(\widetilde{\alpha},\widetilde{r})$ est un estimateur convergent (presque sûr) de (α, r).

(b) On a successivement

$$\begin{array}{rcl}\mathbb{V}(T_i) & = & \dfrac{r\alpha^2}{(r-1)^2(r-2)}\\[2ex] \mathbb{V}\left({T_i}^2\right) & = & \mathbb{E}\left({T_i}^4\right)-\left(\mathbb{E}\left({T_i}^2\right)\right)^2\\ & = & \dfrac{24\alpha^4}{(r-1)(r-2)(r-3)(r-4)}-\dfrac{4\alpha^4}{(r-1)^2(r-2)^2}\\ & = & \dfrac{4r\alpha^4(5r-11)}{(r-1)^2(r-2)^2(r-3)(r-4)}\\[2ex] \mathbb{C}ov\left(T_i,{T_i}^2\right) & = & \mathbb{E}\left({T_i}^3\right)-\mathbb{E}(T_i)\mathbb{E}\left({T_i}^2\right)\\ & = & \dfrac{6\alpha^3}{(r-1)(r-2)(r-3)}-\dfrac{2\alpha^3}{(r-1)^2(r-2)}\\ & = & \dfrac{4r\alpha^3}{(r-1)^2(r-2)(r-3)}\end{array}$$

de sorte que d'après le Théorème Central Limite

$$\sqrt{n}\begin{pmatrix}\bar{T}-\mathbb{E}(T_i)\\ \bar{T^2}-\mathbb{E}(T^2)\end{pmatrix}\xrightarrow[n\to+\infty]{\mathcal{L}}\mathcal{N}\left(\begin{pmatrix}0\\0\end{pmatrix},\begin{pmatrix}\frac{r\alpha^2}{(r-1)^2(r-2)} & \frac{4r\alpha^3}{(r-1)^2(r-2)(r-3)}\\ \frac{4r\alpha^3}{(r-1)^2(r-2)(r-3)} & \frac{4r\alpha^4(5r-11)}{(r-1)^2(r-2)^2(r-3)(r-4)}\end{pmatrix}\right)$$

Posons donc $g:\begin{pmatrix}{\mathbb{R}^{+*}}^2 & \to & {\mathbb{R}^{+*}}^2\\ (m_1,m_2) & \mapsto & \left(\frac{m_1m_2}{m_2-2{m_1}^2},\frac{2(m_2-m_1)}{m_2-2{m_1}^2}\right)\end{pmatrix}$;

g est de classe C^∞ et de Jacobienne

$$Jac\ g(m_1,m_2)=\begin{pmatrix}\frac{m_2(m_2+2{m_1}^2)}{(m_2-2{m_1}^2)^2} & \frac{4m_1m_2}{(m_2-2{m_1}^2)^2}\\ \frac{-2{m_1}^3}{(m_2-2{m_1}^2)^2} & \frac{-2{m_1}^2}{(m_2-2{m_1}^2)^2}\end{pmatrix}$$

En définitive on a[1]

$$\sqrt{n}\left(\begin{pmatrix}\widetilde{\alpha}\\ \widetilde{r}\end{pmatrix}-\begin{pmatrix}\alpha\\ r\end{pmatrix}\right)\xrightarrow[n\to+\infty]{\mathcal{L}}\mathcal{N}\left(\begin{pmatrix}0\\0\end{pmatrix},\left[\begin{array}{l}\begin{pmatrix}\frac{m_2(m_2+2{m_1}^2)}{(m_2-2{m_1}^2)^2} & \frac{-2{m_1}^3}{(m_2-2{m_1}^2)^2}\\ \frac{4m_1m_2}{(m_2-2{m_1}^2)^2} & \frac{-2{m_1}^2}{(m_2-2{m_1}^2)^2}\end{pmatrix}\\ \times\begin{pmatrix}\frac{r\alpha^2}{(r-1)^2(r-2)} & \frac{4r\alpha^3}{(r-1)^2(r-2)(r-3)}\\ \frac{4r\alpha^3}{(r-1)^2(r-2)(r-3)} & \frac{4r\alpha^4(5r-11)}{(r-1)^2(r-2)^2(r-3)(r-4)}\end{pmatrix}\\ \times\begin{pmatrix}\frac{m_2(m_2+2{m_1}^2)}{(m_2-2{m_1}^2)^2} & \frac{4m_1m_2}{(m_2-2{m_1}^2)^2}\\ \frac{-2{m_1}^3}{(m_2-2{m_1}^2)^2} & \frac{-2{m_1}^2}{(m_2-2{m_1}^2)^2}\end{pmatrix}\end{array}\right.\right)$$

[1]Nous ne chercherons pas à simplifier davantage cette expression malgré l'intérêt manifeste que cela présenterait ...

⋆
⋆ ⋆

Corrigé de l'exercice 3

☞ Q1 *Rappel :* Etant donné un modèle statistique paramétré par θ, et une hypotèse H_0 sur θ, construire un *test* de niveau $\alpha \in [0,1[$, c'est déterminer une statistique dont la loi $\mathcal{L}_0$, si H_0 est vraie, est :
- complètement déterminée
- "sympathique" (*i.e.*, bien connue et tabulée)

Effectuer le test revient alors à comparer à $1-\alpha$ la vraisemblance de la réalisation de la statistique, au moyen de la table de la loi $\mathcal{L}_0$.

On appelle alors *erreur de première espèce* α du test considéré la probabilité de rejeter H_0 alors qu'elle est vraie, et *erreur de seconde espèce* β la probabilité d'accepter H_0 alors qu'elle est fausse. Si R désigne la zone de rejet (ensemble des observations conduisant à $\neg H_0$), on a $\alpha = \mathbb{P}_{H_0}(W)$ et $\beta = 1 - \mathbb{P}_{\neg H_0}(W)$ (qui est donc difficilement calculable). Enfin, on définit habituellement $\rho = 1-\beta$ la *puissance* du test.

On cherche ici à tester l'hypothèse H_0 : " $\lambda = \lambda_0$" contre $H_a = \neg H_0$.

- Test de *Wald* :

  Constatant que si $\widehat{\lambda}$ est un estimateur asymptotiquement efficace de λ, alors

  $$\sqrt{n}\left(\widehat{\lambda}-\lambda_0\right) \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}\left(0, I_1(\lambda_0)^{-1}\right)$$

  si H_0 est vraie, et donc

  $$n\left(\widehat{\lambda}-\lambda_0\right)' I_1(\lambda_0)\left(\widehat{\lambda}-\lambda_0\right) \xrightarrow[n\to+\infty]{\mathcal{L}} \chi^2_k$$

  où k est le nombre de contraintes (supposées linéaires) imposées par l'hypothèse H_0 ; dans le cas présent, on a $k=1$. En se fondant sur l'approximation convergente $I_1\left(\widehat{\lambda}\right)$ de $I_1(\lambda_0)$ on pose donc

  $$\xi_n^W = n\left(\widehat{\lambda}-\lambda_0\right)' I_1\left(\widehat{\lambda}\right)\left(\widehat{\lambda}-\lambda_0\right) \xrightarrow[n\to+\infty]{\mathcal{L}} \chi^2_k$$

- Test du *score* :

  La normalité asymtpotique de l'estimateur du maximum de vraisemblance permet d'en déduire celle du score grâce au théorème de Slutsky :

  $$\sqrt{n}\left(S(\lambda)-S(\lambda_0)\right) \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}\left(0, \frac{\partial g}{\partial \lambda}(\lambda_0)^2 I_1(\lambda_0)^{-1}\right)$$

  où $g : x \mapsto \frac{\partial \ln \mathcal{L}}{\partial \lambda}(x)$ D'après le théorème central limite

  $$\frac{1}{n} g\left(\widehat{\lambda}\right)^2 \xrightarrow[n\to+\infty]{p.s.} \mathbb{E}\left(g(\lambda_0)^2\right) = -I_1(\lambda_0)$$

de sorte que finalement

$$\sqrt{n}\,(S(\lambda)-S(\lambda_0)) \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}\left(0,(nI_1(\lambda_0))^2\,I_1(\lambda_0)^{-1}\right)$$

Donc

$$\frac{\sqrt{n}}{n}\,(S(\lambda)-S(\lambda_0))\,I_1(\widehat{\lambda}) \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}\,(0,1)$$

et par conséquent

$$\frac{1}{n}S(\lambda_0)'I_1(\lambda_0)^{-1}S(\lambda_0) \xrightarrow[n\to+\infty]{\mathcal{L}} \chi^2_k$$

On définit donc

$$\xi_n^S=\frac{1}{n}S\,(\lambda_0)'\,I_1\,(\lambda_0)^{-1}\,S\left(\widehat{\lambda}\right) \xrightarrow[n\to+\infty]{\mathcal{L}} \chi^2_k$$

– Test du *rapport des maxima de vraisemblance* :
  Sous réserve que la vraisemblance soit suffisamment régulière et que le modèle statistique soit homogène et dominé on a en développant à l'ordre deux

$$\begin{aligned}\ln\mathcal{L}_{X_1,\ldots,X_n}(x_1,\ldots,x_n;\lambda) \;=\; &\ln\mathcal{L}_{X_1,\ldots,X_n}(x_1,\ldots,x_n;\lambda_0) + \frac{\partial\ln\mathcal{L}_{X_1,\ldots,X_n}}{\partial\lambda}(x_1,\ldots,x_n;\lambda_0)\,(\lambda-\lambda_0)\\ &+\frac{1}{2}\frac{\partial^2\ln\mathcal{L}_{X_1,\ldots,X_n}}{\partial\lambda^2}(x_1,\ldots,x_n;\lambda_0)\,(\lambda-\lambda_0)^2 + o_{\mathbb{P}_{\lambda_0}}\left((\lambda-\lambda_0)^2\right)\end{aligned}$$

et donc par la loi forte des grands nombres

$$\ln\mathcal{L}_{X_1,\ldots,X_n}(x_1,\ldots,x_n;\lambda)-\ln\mathcal{L}_{X_1,\ldots,X_n}(x_1,\ldots,x_n;\lambda_0) \xrightarrow[n\to+\infty]{p.s.} \mathbb{E}\,(S(\lambda_0))-\frac{1}{2}I_n(\lambda_0)\,(\lambda-\lambda_0)^2+o(\lambda-\lambda_0)^2$$

et comme

$$\sqrt{n}\left(\widehat{\lambda}-\lambda_0\right) \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}\left(0,I_1(\lambda_0)^{-1}\right)$$

il s'ensuite que

$$2\,(\ln\mathcal{L}_{X_1,\ldots,X_n}(x_1,\ldots,x_n;\lambda_0)-\ln\mathcal{L}_{X_1,\ldots,X_n}(x_1,\ldots,x_n;\lambda)) \xrightarrow[n\to+\infty]{\mathcal{L}} \chi^2_k$$

On pose donc

$$\xi_n^{mv}=2\left(\ln L^{\neg H_0}\left(\widehat{\lambda}\right)-\ln L^{H_0}\,(\lambda_0)\right) \xrightarrow[n\to+\infty]{\mathcal{L}} \chi^2_k$$

Pour chacun de ces tests $\xi_n^{\cdots}$, on accepte finalement l'hypothèse H_0 au niveau α *ssi*

$$\xi_n^{\cdots}(y_1,\ldots,y_n)<q_{1-\alpha}^{\chi^2_k}$$

où $q_{1-\alpha}^{\chi^2_k}$ désigne le fractile de niveau $1-\alpha$ de la loi du χ^2 à k degrés de liberté.
Notons que ces trois tests sont asymptotiquement équivalents.

☞ Q2 On a pour $(y_1,\ldots,y_n)\in\mathbb{N}^n$

$$L(y_1,\ldots,y_n;\lambda)=\Pi_{i=1}^n\left(e^{-\lambda}\frac{\lambda_i^y}{y_i!}\right)$$

dont on tire que $\widehat{\lambda}=\bar{y}$; il reste à calculer explicitement ξ^W, ξ^S et ξ^{mv}.

– Test de *Wald* :
On sait d'après le Théorème Central Limite que

$$\sqrt{n}\left(\widehat{\lambda}-\lambda\right) \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}(0,\lambda)$$

car $I_(\lambda)=\frac{1}{\lambda}$, donc sous forme centrée réduite il vient

$$\sqrt{n}\frac{\widehat{\lambda}-\lambda}{\sqrt{\lambda}} \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}(0,1)$$

et donc

$$n\frac{\left(\widehat{\lambda}-\lambda\right)^2}{\lambda} \xrightarrow[n\to+\infty]{\mathcal{L}} \chi_1^2$$

et comme $\widehat{\lambda}=\bar{y} \xrightarrow[n\to+\infty]{p.s.} \lambda$ on pose

$$\xi^W(y_1,\ldots,y_n) = n\frac{(\bar{y}-\lambda_0)^2}{\bar{y}}$$

De cette façon, si H_0 est vraie alors $\xi^W(y_1,\ldots,y_n) \xrightarrow[n\to+\infty]{p.s.} \chi_1^2$, tandis que sinon $\xi^W(y_1,\ldots,y_n) \xrightarrow[n\to+\infty]{p.s.} (+\infty)$.

– Test du *score* :
On a

$$\ln L(y_1,\ldots,y_n;\lambda) = -n\lambda + \left(\sum_{i=1}^n y_i\right)\ln(\lambda) - \sum_{i=1}^n \ln(y_i!)$$

d'où on tire

$$\frac{\partial \ln L}{\partial \lambda}(y_1,\ldots,y_n;\lambda_0) = -n + \frac{n\bar{y}}{\lambda_0}$$

Le théorème central limite appliqué au vecteur score s'écrit

$$\frac{1}{\sqrt{n}}\left(\frac{\partial \ln L}{\partial \lambda}(y_1,\ldots,y_n;\lambda_0)\right)\sqrt{\lambda_0} \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}\left(0,\frac{1}{\lambda_0}\right)$$

si H_0 est vraie, auquel cas

$$\xi_n^S = n\frac{(\bar{y}-\lambda_0)^2}{\lambda_0} \xrightarrow[n\to+\infty]{\mathcal{L}} \chi_1^2$$

– Test du *rapport des maxima de vraisemblance* :
On a

$$\ln L^{\neg H_0}(y_1,\ldots,y_n;\widehat{\lambda}) = -n\bar{y} + \left(\sum_{i=1}^n y_i\right)\ln(\bar{y}) - \sum_{i=1}^n \ln(y_i!)$$

et

$$\ln L^{H_0}(y_1,\ldots,y_n;\lambda_0) = -n\lambda_0 + \left(\sum_{i=1}^n y_i\right)\ln(\lambda_0) - \sum_{i=1}^n \ln(y_i!)$$

et donc finalement

$$\xi^{mv}(y_1, \ldots, y_n) = 2n\left(\bar{y}\left(\ln \bar{y} - 1\right) + \lambda_0 - \ln \lambda_0\right) \xrightarrow[n\to+\infty]{\mathcal{L}} \chi^2_1$$

si H_0 est vraie.

On en tire alors $W_\alpha^{test} = \{(y_1, \ldots, y_T) \;/\; \xi_T^{test}(y_1, \ldots, y_T) > q_{1-\alpha}^{\chi_1^2}\}$ pour $test \in \{W, S, mv\}$.

☞ Q3 Application numérique : $\lambda_0 = 1$, $q_{1-5\%}^{\chi_1^2} = 3,84$

Test	$n = 100$ $\bar{y} = 1,2$	$n = 200$ $\bar{y} = 1,1$
Wald	$\xi^W_{100} = 3,33$ H_0 acceptée	$\xi^W_{200} = 1,82$ H_0 acceptée
Score	$\xi^S_{100} = 4$ H_0 **rejetée**	$\xi^S_{200} = 2$ H_0 acceptée
M.V.	$\xi^{mv}_{100} = 3,76$ H_0 acceptée	$\xi^{mv}_{200} = 1,94$ H_0 acceptée

Il apparaît que pour $n = 200$, H_0 est unanimement acceptée au seuil de 5% ; la situation est moins nette lorsque $n = 100$:

– à la fois parce que le test du score conduit à rejeter H_0 au contraire des deux autres tests ;

– et parce que même pour ceux-là la statistique est proche du fractile $q_{1-5\%}^{\chi_1^2}$

Sachant que ces tests devraient être asymptotiquement équivalents (donc en particulier conduire à la même décision, pour une observation donnée), on peut en conclure hardiment que $n = 100$ n'est **pas** "asymptotiquement grand".

Quant à décider de l'acceptation, une première solution serait d'affiner ou d'élargir le seuil d'acceptation, par exemple en choisissant $\alpha = 10\%$ (ce qui conduirait à accepter unanimement H_0) ou à l'inverse $\alpha = 1\%$ (ce qui conduirait à rejeter unanimement H_0) ; mais modifier a posteriori le problème posé de façon à savoir le résoudre est une démarche un peu discutable ...

On remarquera plutôt que l'issue d'un test pour $n_2 > n_1$ est toujours préférable à celle pour n_1 puisque ce test n'est justifié qu'asymptotiquement ; comme en outre ces tests sont asymptotiquement équivalents, à partir d'un certain rang il seront tous unanimes, et c'est cettte décision qu'il faut retenir. Dans le cas présent, on retiendra donc l'issue (unanime) des tests pour $n = 200$, à savoir l'**acceptation** de H_0 au seuil 5%.

⋆

⋆ ⋆