



Cursus Intégré
2004-2005

Rappels de statistique mathématique
Énoncé et corrigé des travaux dirigés n°6

Guillaume Lacôte

 Bureau **E03**

✉ Guillaume.Lacote@ensae.fr

☞ <http://ensae.no-ip.com/SE222/>

Énoncé de l'exercice 1

On considère un échantillon X_1, \dots, X_n i.i.d. tiré de la loi de Poisson de paramètre λ .

- ☞ Q1 Donner $\mathbb{E}(X_i)$ et $\mathbb{V}(X_i)$.
Calculer $\phi(t) = \mathbb{E}(e^{tX_i})$.
En déduire que

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

- ☞ Q2 (a) On pose $\hat{\lambda}_1(x) = \frac{1}{n} \sum_{i=1}^n x_i = \bar{x}$ et $\hat{\lambda}_2(x) = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$.
Par quelle méthode d'estimation ont été obtenus $\hat{\lambda}_1$ et $\hat{\lambda}_2$?
- (b) Les estimateurs $\hat{\lambda}_1$ et $\hat{\lambda}_2$ estiment-ils λ sans biais ?
Proposer un autre estimateur sans biais de λ , que l'on notera $\hat{\lambda}_3$.
- (c) Donner la loi (à distance finie) de $\hat{\lambda}_1$.
- (d) Donner la loi asymptotique jointe de $\begin{pmatrix} \hat{\lambda}_1 \\ \hat{\lambda}_2 \end{pmatrix}$, puis celle de $\begin{pmatrix} \hat{\lambda}_1 \\ \hat{\lambda}_3 \end{pmatrix}$.
- (e) Soit $\hat{\lambda}_\infty = \alpha \hat{\lambda}_1 + \beta \hat{\lambda}_3$, $(\alpha, \beta) \in \mathbb{R}^{+2}$.
Donner la valeur de (α, β) telle que $\hat{\lambda}_\infty$ soit le meilleur estimateur asymptotiquement sans biais de λ .
Ce résultat est-il surprenant (donner l'e.m.v. de λ) ?
- ☞ Q3 Dans cette question, on se place dans le cas où $n = 2$.
- (a) Calculer les erreurs quadratiques moyennes pour les trois estimateurs $\hat{\lambda}_1$, $\hat{\lambda}_2$ et $\hat{\lambda}_3$.
- (b) Déterminer en fonction de λ lequel de ces estimateurs est le meilleur selon la critère de l'erreur quadratique moyenne.
- (c) Comparer explicitement l'EQM de $\hat{\lambda}_3$ avec la borne FDCR.

Corrigé de l'exercice 1

☞ Q1 – On a

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

et de même

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

ce dont on tire

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

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

Alors pour $t \in \mathbb{R}$

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

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

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

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

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

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

ce dont on déduit que

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

– Par ailleurs, on a successivement :

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

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 $n \in \mathbb{N}$

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

Par suite

$$\forall n \in \mathbb{N}, P_{n+1}(\mathbb{X}) = \mathbb{X} (P_n(\mathbb{X}) + P'_n(\mathbb{X}))$$

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

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

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

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

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

Les questions suivantes s'assurent que $\hat{\lambda}_1$ et $\hat{\lambda}_2$ sont bien des estimateurs de λ et étudient leurs propriétés à distance (in-)finie.

(b) On a

$$\begin{aligned} \mathbb{E}(\hat{\lambda}_1(X)) &= \mathbb{E}(\bar{X}) \\ &= \frac{1}{n} \sum_{i=1}^n \mathbb{E}(X_i) \\ &= \boxed{\lambda} \end{aligned}$$

puis

$$\begin{aligned} \mathbb{E}(\hat{\lambda}_2(X)) &= \frac{1}{n} \sum_{i=1}^n \mathbb{E}((X_i - \bar{X})^2) \\ &= \frac{1}{n} \sum_{i=1}^n \mathbb{E}(X_i^2 - 2X_i\bar{X} + \bar{X}^2) \\ &= \frac{1}{n} \sum_{i=1}^n \mathbb{E}(X_i^2) - 2\mathbb{E}(\bar{X}^2) + \mathbb{E}(\bar{X}^2) \\ &= \mathbb{E}(X_1^2) - \mathbb{E}(\bar{X}^2) \\ &= \mathbb{E}(X_1^2) - (\mathbb{V}(\bar{X}) + \mathbb{E}(\bar{X})^2) \\ &= \mathbb{E}(X_1^2) - \left(\frac{1}{n}\mathbb{V}(X_1) + \mathbb{E}(X_1)^2\right) \\ &= (\lambda^2 + \lambda) - \left(\frac{1}{n}\lambda + \lambda^2\right) \\ &= \boxed{\frac{n-1}{n}\lambda} \end{aligned}$$

Ainsi, $\hat{\lambda}_1$ n'est pas biaisé mais $\hat{\lambda}_2$ l'est, et on peut proposer $\hat{\lambda}_3 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$

(c) Première méthode : 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

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

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'occurrence

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

ce dont on conclut que $\sum_{i=1}^n X_i \sim \mathcal{P}(n\lambda)$.

En définitive la loi de $\hat{\lambda}_1$ est caractérisée par

$$\forall k \in \mathbb{Q}, \mathbb{P}\left(\hat{\lambda}_1 = k\right) = \mathbb{P}\left(\sum_{i=1}^n X_i = nk\right) = e^{-n\lambda} \frac{(n\lambda)^{nk}}{(nk)!} \mathbf{1}_{nk \in \mathbb{N}}$$

(on notera que la statistique $\hat{\lambda}_1$ n'est **pas** entière a priori, ce qui justifie l'indicatrice).

Autre Méthode : On peut aussi directement calculer pour $t \in \mathbb{N}$

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

- (d) Pour déterminer la loi limite jointe de $\begin{pmatrix} \hat{\lambda}_1 \\ \hat{\lambda}_2 \end{pmatrix}$, on serait tenté de poser $Z_i = \begin{pmatrix} X_i \\ (X_i - \bar{X})^2 \end{pmatrix}$ 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

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

ce qui s'écrit

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

Par ailleurs on calcule successivement

$$\begin{aligned}\mathbb{V}(X_i) &= \lambda \\ \mathbb{V}(X_i^2) &= \mathbb{E}(X_i^4) - \mathbb{E}(X_i^2)^2 \\ &= \lambda(1 + 6\lambda + 4\lambda^2) \\ \text{Cov}(X_i, X_i^2) &= \mathbb{E}(X_i^3) - \mathbb{E}(X_i)\mathbb{E}(X_i^2) \\ &= \lambda(1 + 2\lambda)\end{aligned}$$

Enfin, posons $g : \left(\begin{array}{c} \mathbb{R}^2 \rightarrow \mathbb{R}^2 \\ \begin{pmatrix} u \\ v \end{pmatrix} \mapsto \begin{pmatrix} u \\ v - u^2 \end{pmatrix} \end{array} \right)$, de sorte que

$$\frac{1}{n} \sum_{i=1}^n Z_i = g\left(\frac{1}{n} \sum_{i=1}^n Y_i\right), \text{ car } \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 = \frac{1}{n} \sum_{i=1}^n X_i^2 - (\bar{X})^2.$$

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

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

soit

$$\sqrt{n} \left(\begin{pmatrix} \hat{\lambda}_1 \\ \hat{\lambda}_2 \end{pmatrix} - \begin{pmatrix} \lambda \\ \lambda \end{pmatrix} \right) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N} \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \lambda & \lambda \\ \lambda & \lambda(1 + 2\lambda) \end{pmatrix} \right)$$

Enfin, remarquant que

$$\begin{pmatrix} \hat{\lambda}_1 \\ \hat{\lambda}_3 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & \frac{n}{n-1} \end{pmatrix} \begin{pmatrix} \hat{\lambda}_1 \\ \hat{\lambda}_2 \end{pmatrix} \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \begin{pmatrix} \lambda \\ \lambda \end{pmatrix}$$

il vient

$$\sqrt{n} \left(\begin{pmatrix} \hat{\lambda}_1 \\ \hat{\lambda}_3 \end{pmatrix} - \begin{pmatrix} \lambda \\ \frac{n}{n-1} \lambda \end{pmatrix} \right) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N} \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \lambda & \lambda \\ \lambda & \lambda(1 + 2\lambda) \end{pmatrix} \right)$$

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

$$\sqrt{n} \left(\begin{pmatrix} \hat{\lambda}_1 \\ \hat{\lambda}_3 \end{pmatrix} - \begin{pmatrix} \lambda \\ \lambda \end{pmatrix} \right) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N} \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \lambda & \lambda \\ \lambda & \lambda(1 + 2\lambda) \end{pmatrix} \right)$$

- (e) On a $\mathbb{E}(\widehat{\lambda}_\infty(X)) = \alpha \mathbb{E}(\widehat{\lambda}_1(X)) + \beta \mathbb{E}(\widehat{\lambda}_3(X)) = (\alpha + \beta)\lambda$,
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 $n \rightarrow +\infty$

$$\begin{aligned} \mathbb{V}(\widehat{\lambda}_\infty(X)) &= \alpha^2 \mathbb{V}(\widehat{\lambda}_1(X)) + 2\alpha(1-\alpha) \text{Cov}(\widehat{\lambda}_1(X), \widehat{\lambda}_3(X)) + (1-\alpha)^2 \mathbb{V}(\widehat{\lambda}_3(X)) \\ &= \begin{cases} \left(\mathbb{V}(\widehat{\lambda}_1(X)) + \mathbb{V}(\widehat{\lambda}_3(X)) - 2\text{Cov}(\widehat{\lambda}_1(X), \widehat{\lambda}_3(X)) \right) \alpha^2 \\ + 2 \left(\text{Cov}(\widehat{\lambda}_1(X), \widehat{\lambda}_3(X)) - \mathbb{V}(\widehat{\lambda}_3(X)) \right) \alpha \\ + \mathbb{V}(\widehat{\lambda}_3(X)) \end{cases} \end{aligned}$$

Ce polynôme de degré 2 en α a un coefficient $\mathbb{V}(\widehat{\lambda}_1(X)) + \mathbb{V}(\widehat{\lambda}_3(X)) - 2\text{Cov}(\widehat{\lambda}_1(X), \widehat{\lambda}_3(X)) = \mathbb{V}(\widehat{\lambda}_1(X) - \widehat{\lambda}_3(X)) \geq 0$ en α^2 , donc est maximal aux bornes et minimal au milieu des racines, soit en

$$\begin{aligned} \alpha^* &= -\frac{2 \left(\text{Cov}(\widehat{\lambda}_1(X), \widehat{\lambda}_3(X)) - \mathbb{V}(\widehat{\lambda}_3(X)) \right)}{2 \times \left(\mathbb{V}(\widehat{\lambda}_1(X)) + \mathbb{V}(\widehat{\lambda}_3(X)) - 2\text{Cov}(\widehat{\lambda}_1(X), \widehat{\lambda}_3(X)) \right)} \\ &= -\frac{2n(\lambda - \lambda(1 + 2\lambda))}{2n(\lambda + \lambda(1 + 2\lambda) - 2\lambda)} \\ &= -\frac{2(\lambda - \lambda - 2\lambda^2)}{2(\lambda + \lambda + 2\lambda^2 - 2\lambda)} \\ &= 1 \end{aligned}$$

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 $\mathbb{E} \left((\widehat{\theta} - \theta)^2 \right)$.

– On a $\widehat{\lambda}_1 = \frac{1}{2}(X_1 + X_2)$, de sorte que

$$\begin{aligned} \mathbb{E} \left((\widehat{\lambda}_1 - \lambda)^2 \right) &= \mathbb{E}(\widehat{\lambda}_1^2) - \lambda^2 \\ &= \frac{1}{4} \left(\mathbb{E}(X_1^2) + 2\text{Cov}(X_1, X_2) + \mathbb{E}(X_2^2) \right) - \lambda^2 \\ &= \frac{1}{4} (2\lambda(1 + \lambda) + 2\lambda^2) - \lambda^2 \\ &= \frac{\lambda}{2} \end{aligned}$$

– De façon similaire

$$\begin{aligned}\widehat{\lambda}_2 &= \frac{1}{2} \left(\left(X_1 - \frac{X_1 + X_2}{2} \right)^2 + \left(X_2 - \frac{X_1 + X_2}{2} \right)^2 \right) \\ &= \frac{1}{2} \left(X_1^2 - X_1(X_1 + X_2) + \frac{(X_1 + X_2)^2}{4} + X_2^2 - X_2(X_1 + X_2) + \frac{(X_1 + X_2)^2}{4} \right) \\ &= \frac{1}{4} (X_1^2 + X_2^2 - 2X_1X_2)\end{aligned}$$

de sorte que

$$\mathbb{E} \left((\widehat{\lambda}_2 - \lambda)^2 \right) = \mathbb{E} (\widehat{\lambda}_2^2) - 2\lambda \mathbb{E} (\widehat{\lambda}_2) + \lambda^2$$

Or $\mathbb{E} (\widehat{\lambda}_2) = \frac{\lambda}{2}$, et en outre

$$\begin{aligned}\mathbb{E} (\widehat{\lambda}_2^2) &= \frac{1}{16} \mathbb{E} \left((X_1^2 + X_2^2 - 2X_1X_2)^2 \right) \\ &= \frac{1}{16} \mathbb{E} \left((X_1^2 + X_2^2)^2 - 2X_1X_2(X_1^2 + X_2^2) + 4X_1^2X_2^2 \right) \\ &= \frac{1}{16} \mathbb{E} (X_1^4 + 2X_1^2X_2^2 + X_2^4 - 2X_1^3X_2 - 2X_1X_2^3 + 4X_1^2X_2^2) \\ &= \frac{1}{8} (\mathbb{E} (X_1^4) + 3\mathbb{E} (X_1^2)\mathbb{E} (X_2^2) - 4\mathbb{E} (X_1^3)\mathbb{E} (X_2)) \quad \text{par indépendance} \\ &= \frac{1}{8} (\lambda + 6\lambda^2)\end{aligned}$$

et donc

$$\begin{aligned}\mathbb{E} \left((\widehat{\lambda}_2 - \lambda)^2 \right) &= \frac{1}{8} (\lambda + 6\lambda^2) - 2\lambda \left(\frac{\lambda}{2} \right) + \lambda^2 \\ &= \frac{3}{4}\lambda^2 + \frac{1}{8}\lambda\end{aligned}$$

– Enfin, $\widehat{\lambda}_3 = 2\widehat{\lambda}_2$ et donc

$$\begin{aligned}\mathbb{E} \left((\widehat{\lambda}_3 - \lambda)^2 \right) &= \mathbb{E} (\widehat{\lambda}_3^2) - \lambda^2 \\ &= 4\mathbb{E} (\widehat{\lambda}_2^2) - \lambda^2 \\ &= 2\lambda^2 + \frac{1}{2}\lambda\end{aligned}$$

(b) Au sens de l'erreur quadratique moyenne, $\hat{\lambda}_1$ est toujours préférable à $\hat{\lambda}_3$.

De plus,

$$\begin{aligned} \hat{\lambda}_2 \text{ est préférable à } \hat{\lambda}_3 &\Leftrightarrow \mathbb{E} \left((\hat{\lambda}_2 - \lambda)^2 \right) < \mathbb{E} \left((\hat{\lambda}_3 - \lambda)^2 \right) \\ &\Leftrightarrow \left(2\lambda^2 + \frac{1}{2}\lambda \right) - \left(\frac{3}{4}\lambda^2 + \frac{1}{8}\lambda \right) > 0 \\ &\Leftrightarrow \left(2 - \frac{3}{4} \right) \lambda^2 + \left(\frac{1}{2} - \frac{1}{8} \right) \lambda > 0 \\ &\Leftrightarrow \lambda > -\frac{3}{8} \quad \text{cd qui est toujours vrai car } \lambda > 0 \end{aligned}$$

et de façon analogue

$$\begin{aligned} \hat{\lambda}_2 \text{ est préférable à } \hat{\lambda}_1 &\Leftrightarrow \mathbb{E} \left((\hat{\lambda}_2 - \lambda)^2 \right) < \mathbb{E} \left((\hat{\lambda}_1 - \lambda)^2 \right) \\ &\Leftrightarrow \left(\frac{\lambda}{2} \right) - \left(\frac{3}{4}\lambda^2 + \frac{1}{8}\lambda \right) > 0 \\ &\Leftrightarrow \lambda < \frac{1}{2} \end{aligned}$$

(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 $\frac{\partial^2 \ln L}{\partial \lambda^2}(k, \lambda) = -\frac{k}{\lambda^2}$. donc

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

de sorte que la borne FDCR est $\frac{\lambda}{2}$. En particulier, $\hat{\lambda}_3$ est efficace ssi $2\lambda^2 + \frac{1}{2}\lambda = \frac{\lambda}{2}$, c'est-à-dire jamais puisque $\lambda > 0$.

*
* *

Énoncé de l'exercice 2

On considère une population de n individus infectés par un virus; on étudie leurs durées d'incubation $(T_i)_{i \in \llbracket 1, n \rrbracket}$, dont on suppose qu'elle est observable.

Pour modéliser l'hétérogénéité de la population, on suppose qu'on peut caractériser chaque individu i par un "facteur de risque" inobservable, réalisation de la variable aléatoire Λ_i , de telle sorte que :

- la loi de T_i , conditionnellement à Λ_i est la loi exponentielle de paramètre λ_i (de densité $\lambda_i e^{-\lambda_i t} \mathbf{1}_{t \geq 0}$);
- la famille $(\Lambda_i)_{i \in \llbracket 1, n \rrbracket}$ est identiquement distribuée selon la loi \mathcal{L} de densité

$$f(\lambda) = \frac{\alpha^r}{\Gamma(r)} \lambda^{r-1} e^{-\alpha \lambda} \mathbf{1}_{\mathbb{R}^+}$$

où $\alpha > 0$ et $r > 2$;

- les couples (T_i, Λ_i) sont indépendants entre eux.

☞ Q1 Donner la vraisemblance de (T_1, \dots, T_n) .

☞ Q2 Calculer, lorsqu'il existe, le moment d'ordre k $\mathbb{E}(T_i^k)$.

☞ Q3 Calculer l'information de Fisher du modèle.

Dans le cas où α est connu, calculer l'estimateur du maximum de vraisemblance de r .

Que se passe-t-il si α et r sont tous deux inconnus?

☞ Q4 On suppose α connu.

Déterminer au moyen de la méthode des moments un estimateur convergent de r .

Cet estimateur est-il sans biais? Est-il asymptotiquement efficace?

☞ Q5 On suppose α et r inconnus.

(a) En utilisant les deux premiers moments de T_i , trouver des estimateurs convergents $\tilde{\alpha}$ et \tilde{r} de α et r .

(b) Donner la loi limite du vecteur

$$\sqrt{n} \begin{pmatrix} \bar{T} - \mathbb{E}(T) \\ \bar{T}^2 - \mathbb{E}(T)^2 \end{pmatrix}$$

En déduire la loi asymptotique du vecteur $(\tilde{\alpha}, \tilde{r})$.

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 $t \in \mathbb{R}$

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

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

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

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

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

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

$$L_{T_1, \dots, T_n}(t_1, \dots, t_n; \alpha, r) = r^n \alpha^{nr} \frac{1}{(\prod_{i=1}^n (\alpha + t_i))^{r+1}} \mathbf{1}_{\min_i t_i \geq 0}$$

⇨ Q2 Pour $k \in \mathbb{N}^*$ on a

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

Soit donc $k \in \llbracket 1, r-1 \rrbracket$, et notons $M_k^r = \mathbb{E}(T_i^k) < +\infty$. Alors

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

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

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

c'est-à-dire (en notant C_n^p le coefficient binomial $\frac{n!}{p!(n-p)!}$)

$$\mathbb{E}(T_i^k) = \frac{\alpha^k}{C_{r-1}^k}$$

On notera en particulier que lorsque $r > 2$

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

⇨ 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

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

puis

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

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

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

et finalement

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

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

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

et donc l'estimateur du maximum de vraisemblance \hat{r} de r vérifie

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

d'où on tire

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

(la réciproque étant immédiate par concavité).

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

$$\begin{cases} \hat{r} &= \frac{n}{\sum_{i=1}^n \ln(\hat{\alpha} + t_i) - n \ln \hat{\alpha}} \\ \hat{\alpha} &= \frac{\sum_{i=1}^n \frac{1}{\hat{\alpha} + t_i}}{\frac{n}{\hat{\alpha}} - \sum_{i=1}^n \frac{1}{\hat{\alpha} + t_i}} \end{cases}$$

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

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

Cependant, à distance finie on observe que

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

de sorte que \hat{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

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

car T_1, \dots, T_n sont indépendants.

Soit alors $g : \begin{pmatrix} \mathbb{R}^{+*} & \rightarrow & \mathbb{R}^{+*} \\ s & \mapsto & 1 + \frac{\alpha}{s} \end{pmatrix}$; g est de classe C^∞ et donc d'après Slutsky

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

soit

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

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

⇒ Q5 (a) Observant la réalisation empirique $\begin{pmatrix} m_1 \\ m_2 \end{pmatrix} = \begin{pmatrix} \frac{1}{n} \sum_{i=1}^n t_i \\ \frac{1}{n} \sum_{i=1}^n t_i^2 \end{pmatrix}$ des moments théoriques

$$\begin{pmatrix} \mathbb{E}(T_i) \\ \mathbb{E}(T_i^2) \end{pmatrix} = \begin{pmatrix} \frac{\alpha}{r-1} \\ \frac{2\alpha^2}{(r-1)(r-2)} \end{pmatrix}, \text{ on détermine } (\tilde{\alpha}, \tilde{r}) \text{ de façon à ce que}$$

$$\begin{pmatrix} \frac{1}{n} \sum_{i=1}^n t_i \\ \frac{1}{n} \sum_{i=1}^n t_i^2 \end{pmatrix} = \begin{pmatrix} \frac{\tilde{\alpha}}{\tilde{r}-1} \\ \frac{2\tilde{\alpha}^2}{(\tilde{r}-1)(\tilde{r}-2)} \end{pmatrix}$$

En l'occurrence on pose

$$\boxed{\begin{cases} \tilde{\alpha} &= \frac{\bar{t} \bar{t}^2}{\bar{t}^2 - 2(\bar{t})^2} \\ \tilde{r} &= \frac{2(\bar{t}^2 - (\bar{t})^2)}{\bar{t}^2 - 2(\bar{t})^2} \end{cases}} \text{ en notant } \begin{cases} \bar{t} &= \frac{1}{n} \sum_{i=1}^n t_i \\ \bar{t}^2 &= \frac{1}{n} \sum_{i=1}^n t_i^2 \neq (\bar{t})^2 \end{cases}$$

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

(b) On a successivement

$$\begin{aligned} \mathbb{V}(T_i) &= \frac{r\alpha^2}{(r-1)^2(r-2)} \\ \mathbb{V}(T_i^2) &= \mathbb{E}(T_i^4) - (\mathbb{E}(T_i^2))^2 \\ &= \frac{24\alpha^4}{(r-1)(r-2)(r-3)(r-4)} - \frac{4\alpha^4}{(r-1)^2(r-2)^2} \\ &= \frac{4r\alpha^4(5r-11)}{(r-1)^2(r-2)^2(r-3)(r-4)} \\ \text{Cov}(T_i, T_i^2) &= \mathbb{E}(T_i^3) - \mathbb{E}(T_i)\mathbb{E}(T_i^2) \\ &= \frac{6\alpha^3}{(r-1)(r-2)(r-3)} - \frac{2\alpha^3}{(r-1)^2(r-2)} \\ &= \frac{4r\alpha^3}{(r-1)^2(r-2)(r-3)} \end{aligned}$$

de sorte que d'après le Théorème Central Limite

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

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

g est de classe C^∞ et de Jacobienne

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

En définitive on a¹

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

*
* *

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

Énoncé de l'exercice 3

On considère les réalisations de T variables aléatoires i.i.d. Y_1, Y_2, \dots, Y_T , issues d'une loi de Poisson de paramètre λ inconnu.

On s'intéresse au test de l'hypothèse :

$$H_0 : \lambda = \lambda_0 \text{ contre } H_a : \lambda \neq \lambda_0$$

où λ_0 est un réel donné.

☞ Q1 Pour tester ce type d'hypothèse, on dispose de trois tests asymptotiques usuels : test de **Wald**, du **score** et du **rapport de maximum de vraisemblance**. Rappeler rapidement leur principe et définir les statistiques de test sur lesquelles ils s'appuient.

☞ Q2 Pratiquer explicitement chacun de ces tests, en calculant l'expression de la statistique de test ξ_T^{test} et en définissant la région critique au seuil $\alpha \in [0, 1]$: $W_\alpha = \{(y_1, \dots, y_T) / \xi_T^{test}(y_1, \dots, y_T) > q_{1-\alpha}^{\chi_1^2}\}$

☞ Q3 On veut tester $\lambda_0 = 1$ au seuil $\alpha = 5\%$.

Décider de chacun des tests dans le cas où l'échantillon observé (y_1, \dots, y_T) est tel que :

- $T = 100$ et $\bar{y} = 1, 2$;
- $T = 200$ et $\bar{y} = 1, 1$.

Discuter de l'acceptation de H_0 et expliquer pourquoi certains tests ne conduisent pas à la même décision.

Conclure.

Corrigé de l'exercice 3

☞ Q1 Rappel : Étant donné un modèle statistique paramétré par θ , et une hypothèse 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 fautive. 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 $\hat{\lambda}$ est un estimateur asymptotiquement efficace de λ , alors

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

si H_0 est vraie, et donc

$$n (\hat{\lambda} - \lambda_0)' I_1(\lambda_0) (\hat{\lambda} - \lambda_0) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \chi_k^2$$

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 $I_1(\hat{\lambda})$ de $I_1(\lambda_0)$ on pose donc

$$\xi_n^W = n (\hat{\lambda} - \lambda_0)' I_1(\hat{\lambda}) (\hat{\lambda} - \lambda_0) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \chi_k^2$$

– Test du score :

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

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

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

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

de sorte que finalement

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

Donc

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

et par conséquent

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

On définit donc

$$\xi_n^S = \frac{1}{n} S(\lambda_0)' I_1(\lambda_0)^{-1} S(\hat{\lambda}) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \chi_k^2$$

– 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

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

et donc par la loi forte des grands nombres

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

et comme

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

il s'ensuit que

$$2 (\ln \mathcal{L}_{X_1, \dots, X_n}(x_1, \dots, x_n; \lambda_0) - \ln \mathcal{L}_{X_1, \dots, X_n}(x_1, \dots, x_n; \lambda)) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \chi_k^2$$

On pose donc

$$\xi_n^{mv} = 2 \left(\ln L^{-H_0}(\hat{\lambda}) - \ln L^{H_0}(\lambda_0) \right) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \chi_k^2$$

Pour chacun de ces tests ξ_n^{\dots} , on accepte finalement l'hypothèse H_0 au niveau α ssi

$$\xi_n^{\dots}(y_1, \dots, y_n) < q_{1-\alpha}^{\chi_k^2}$$

où $q_{1-\alpha}^{\chi_k^2}$ 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 $(y_1, \dots, y_n) \in \mathbb{N}^n$

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

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

– Test de Wald :

On sait d'après le Théorème Central Limite que

$$\sqrt{n} (\hat{\lambda} - \lambda) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}(0, \lambda)$$

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

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

et donc

$$n \frac{(\hat{\lambda} - \lambda)^2}{\lambda} \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \chi_1^2$$

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

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

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

$$\xi^W(y_1, \dots, y_n) \xrightarrow[n \rightarrow +\infty]{p.s.} (+\infty).$$

– Test du score :

On a

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

d'où on tire

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

Le théorème central limite appliqué au vecteur score s'écrit

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

si H_0 est vraie, auquel cas

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

– Test du rapport des maxima de vraisemblance :

On a

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

et

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

et donc finalement

$$\xi^{mv}(y_1, \dots, y_n) = 2n(\bar{y}(\ln \bar{y} - 1) + \lambda_0 - \ln \lambda_0) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \chi_1^2$$

si H_0 est vraie.

On en tire alors $W_\alpha^{test} = \{(y_1, \dots, y_T) / \xi_T^{test}(y_1, \dots, y_T) > q_{1-\alpha}^{\chi_1^2}\}$ 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	$\xi_{100}^W = 3,33$ H_0 acceptée	$\xi_{200}^W = 1,82$ H_0 acceptée
Score	$\xi_{100}^S = 4$ H_0 rejetée	$\xi_{200}^S = 2$ H_0 acceptée
M.V.	$\xi_{100}^{mv} = 3,76$ H_0 acceptée	$\xi_{200}^{mv} = 1,94$ 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 ils seront tous unanimes, et c'est cette 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%.

*
* *