ENSAE

ECOLE NATIONALE
DE LA STATISTIQUE
ET DE L'ADMINISTRATION
ECONOMIQUE

Cursus Intégré
2004-2005

Rappels de statistique mathématique
Réponses question par question des travaux dirigés n°1 à 8

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

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

Table des matières

1 Travaux Dirigés n°1

Exercice corrigé 1

On dispose d'observations Y_i relatives au comportement de remboursement ou de non-remboursement d'emprunteurs :

$$Y_i = \begin{cases} 1 & \text{si l'emprunteur } i \text{ rembourse son crédit} \\ 0 & \text{si l'emprunteur } i \text{ est défaillant} \end{cases}$$

Afin de modéliser ce phénomène, on suppose l'existence d'une variable aléatoire Y_i^* normale, d'espérance m et de variance σ^2, qu'on appellera "capacité de remboursement de l'individu i", telle que :

$$Y_i = \begin{cases} 1 & \text{si } Y_i^* \geq 0 \\ 0 & \text{si } Y_i^* < 0 \end{cases}$$

☞ Q1 On note Φ la fonction de répartition de la loi $\mathcal{N}(0,1)$.
Exprimer la loi de Y_i en fonction de Φ.

Par définition $Y_i^* \rightsquigarrow \mathcal{N}(m, \sigma^2)$,
donc $\frac{Y_i^* - m}{\sigma} \rightsquigarrow \mathcal{N}(0,1)$
et par suite

$$\begin{aligned} \mathbb{P}\left(Y_i^* \geq 0\right) &= \mathbb{P}\left(\frac{Y_i^* - m}{\sigma} \geq -\frac{m}{\sigma}\right) \\ &= 1 - \Phi\left(-\frac{m}{\sigma}\right) \\ &= \Phi\left(\frac{m}{\sigma}\right) \end{aligned}$$

puisque la distribution de la loi normale est symétrique. Ainsi

$$Y_i \rightsquigarrow \mathcal{B}\left(1, \Phi\left(\frac{m}{\sigma}\right)\right)$$

☞ Q2 Les paramètres m et σ^2 sont-ils identifiables ?

Le couple (m, σ^2) n'est clairement pas identifiable, car

$$\mathcal{B}\left(1, \Phi\left(\frac{m}{\sigma}\right)\right) = \mathcal{B}\left(1, \Phi\left(\frac{2m}{2\sigma}\right)\right)$$

En revanche, le rapport $\frac{m}{\sigma}$ l'est (en vertu du théorème de factorisation).

Exercice corrigé 2

Un système S fonctionne en utilisant deux machines de types différents. Les durées de vie X_1 et X_2 des deux machines suivent des lois exponentielles de paramètres λ_1 et λ_2. Les variables aléatoires X_1 et X_2 sont supposées indépendantes.

☞ Q1

Montrer que $$X \rightsquigarrow \mathcal{E}(\lambda) \Leftrightarrow \forall x \in \mathbb{R},\ \mathbb{P}(X > x) = \exp(-\lambda x)$$

- Si $X \rightsquigarrow \mathcal{E}(\lambda)$, alors $\forall x > 0,\ \mathbb{P}(X > x) = \int_x^{+\infty} \lambda \exp(-\lambda t)dt = e^{-\lambda x}$
- La réciproque s'obtient par dérivation de l'égalité précédente :
  $\forall x \in \mathbb{R}, \quad f(x)$ existe et vaut $\lambda e^{-\lambda x}\mathbb{1}_{\mathbb{R}^+}(x)$,
  donc toute variable aléatoire réelle vérifiant l'inégalité précédente admet une densité qui est celle de $\mathcal{E}(\lambda)$.

☞ Q2

Calculer la probabilité pour que le système ne tombe pas en panne avant la date t. En déduire la loi de la durée de vie Z du système. Calculer la probabilité pour que la panne du système soit due à une défaillance de la machine 1.

- On a successivement

$$\begin{aligned}\mathbb{P}(Z > t) &= \mathbb{P}(X_1 > t \wedge X_2 > t) \quad \text{par définition de } Z\\ &= \mathbb{P}(X_1 > t)\mathbb{P}(X_2 > t) \quad \text{par indépendance}\\ &= e^{-(\lambda_1+\lambda_2)t}\end{aligned}$$

  Donc Z suit une loi $\mathcal{E}(\lambda_1 + \lambda_2)$.
- Par définition

$$\begin{aligned}\mathbb{P}(X_2 > X_1) &= \int\int \mathbb{1}_{(v>u)}.\lambda_1 e^{-\lambda_1 u}\lambda_2 e^{-\lambda_2 v}dudv\\ &= \int_0^{+\infty} e^{-\lambda_2 t}\lambda_1 e^{-\lambda_1 t}dt\\ &= \frac{\lambda_1}{\lambda_1+\lambda_2}\end{aligned}$$

☞ Q3 On dispose de n systèmes $S_1, ..., S_n$ identiques dont on observe les durées de vie $Z_1, ...Z_n$.

(a)

Ecrire le modèle statistique correspondant et la vraisemblance des observations. A-t-on suffisamment d'information pour estimer λ_1 et λ_2 ?

Il s'agit de n observations i.i.d de loi $\mathcal{E}(\lambda_1 + \lambda_2)$.
Modèle : $\{\mathbb{R}^{+n}, \mathcal{E}(\lambda_1+\lambda_2)^{\otimes n}, (\lambda_1, \lambda_2) \in \mathbb{R}^{+2}\}$
Vraisemblance : $L_n = (\lambda_1 + \lambda_2)^n e^{-(\lambda_1+\lambda_2)\sum Z_i}$
Bien entendu seule la somme $(\lambda_1 + \lambda_2)$ est identifiable.

(b) Si on observe à la fois les durées de vie des systèmes et la cause de la défaillance (machine 1 ou 2), écrire le modèle statistique correspondant et la vraisemblance des observations. A-t-on alors suffisamment d'information pour estimer λ_1 et λ_2 ?

Cette fois, on dispose de la variable supplémentaire $S_i = \mathbb{1}_{(X_2 > X_1)}$. Alors

$$\begin{aligned}
\mathbb{P}(Z > t \,|\, S_i = 1) &= \mathbb{P}(X_2 > X_1 > t) \\
&= \int_{\mathbb{R}} \int_{\mathbb{R}} \mathbb{1}_{(u > v > t)} \lambda_1 e^{-\lambda_1 u} \lambda_2 e^{-\lambda_2 v} du dv \\
&= \int_t^{+\infty} \left(\int_v^{+\infty} \lambda_1 e^{-\lambda_1 u} du \right) \lambda_2 e^{-\lambda_2 v} dv \\
&= \int_t^{+\infty} \left(e^{-\lambda_1 v} \right) \lambda_2 e^{-\lambda_2 v} dv \\
&= \frac{\lambda_1}{\lambda_1 + \lambda_2} \exp(-(\lambda_1 + \lambda_2)t)
\end{aligned}$$

De même, $\mathbb{P}(Z > t \,|\, S_i = 0) = \frac{\lambda_2}{\lambda_1 + \lambda_2} \exp(-(\lambda_1 + \lambda_2)t)$
Les densités correspondantes s'en déduisent (au signe près) par dérivation selon t :

$$\begin{aligned}
L_n &= \prod_{s_i = 1} \{\lambda_1 e^{-(\lambda_1 + \lambda_2) z_i}\} \prod_{s_i = 0} \{\lambda_2 e^{-(\lambda_1 + \lambda_2) z_i}\} \\
&= \lambda_1^{n_1} \lambda_2^{n_2} e^{-(\lambda_1 + \lambda_2) \sum_{i=1}^n z_i}
\end{aligned}$$

où $n_1 = |\{i \,/\, s_i = 1\}|$ et $n_2 = |\{i \,/\, s_2 = 1\}| = n - n_1$.
Le couple (λ_1, λ_2) est donc identifiable, et on peut donc envisager d'estimer à la fois λ_1 et λ_2.

☞ Q4 Dans cette question, on considère un seul système S utilisant une machine de type 1 et une machine de type 2, mais on suppose que l'on dispose d'un stock de n_1 machines de type 1, de durées de vie $X_1^1, ..., X_1^{n_1}$, et d'un stock de n_2 machines de type 2, de durées de vie $X_2^1, ..., X_2^{n_2}$. Quand une machine tombe en panne, on la remplace par une machine du même type, tant que le stock de machines de ce type n'est pas épuisé. Quand cela arrive, on dit que le système S lui-même est en panne. On note toujours Z la durée de vie du système.
Le cas $n_1 = n_2 = 0$ correspond donc à la première question (pas de stock).

(a) Donner la loi de la somme de n variables indépendantes qui suivent une loi exponentielle de même paramètre λ.

La loi $\Gamma(k, \lambda)$ admet pour densité $f(x) = \frac{\lambda^k}{(k-1)!} x^{k-1} e^{-\lambda x} \mathbb{1}_{x>0}$.
En particulier, toute loi $\mathcal{E}(\lambda)$ est une loi $\Gamma(1, \lambda)$.
On a en outre : *la somme de deux variables indépendantes de lois* $\Gamma(k, \lambda)$ *et* $\Gamma(l, \lambda)$ *suit une loi* $\Gamma(k + l, \lambda)$.

Cette propriété se prouve par récurrence sur k : [1]
soit $U \rightsquigarrow \Gamma(k,\lambda)$ et $V \rightsquigarrow \Gamma(1,\lambda)$; alors $W = U + V$ a pour densité :

$$\begin{aligned}
f_W(w) &= \int_{u+v=w} \frac{\lambda^k}{(k-1)!} u^{k-1} e^{-\lambda u} \lambda e^{-\lambda v} du dv \\
&= \int_0^{+\infty} \frac{\lambda^k}{(k-1)!} u^{k-1} e^{-\lambda u} \lambda e^{-\lambda(w-u)} du \\
&= \frac{\lambda^{k+1}}{(k-1)!} e^{-\lambda w} \int_0^{+\infty} u^{k-1} du = \frac{\lambda^{k+1}}{k!} e^{-\lambda w}
\end{aligned}$$

de sorte que $W \rightsquigarrow \Gamma(k+1,\lambda)$.
Ainsi la somme de $n \geq 1$ variables indépendantes qui suivent une loi exponentielle de même paramètre λ suit une loi $\Gamma(n,\lambda)$.

(b) Ecrire Z en fonction des X_j^i et en déduire $P(Z \geq t)$ en fonction de certaines lois gamma, dont on précisera les paramètres.

On a $Z = \min\left(\sum_{i=0}^{n_1} X_1^i, \sum_{j=0}^{n_2} X_2^j\right)$ où $(\sum_{i=0}^{n_1} X_1^i) \rightsquigarrow \Gamma(n_1+1,\lambda_1)$ et $(\sum_{i=0}^{n_2} X_2^i) \rightsquigarrow \Gamma(n_2+1,\lambda_2)$.
Donc

$$\begin{aligned}
\mathbb{P}(Z \geq t) &= \mathbb{P}\left(\left(\sum_{i=0}^{n_1} X_1^i \geq t\right) \text{ et } \left(\sum_{i=0}^{n_2} X_2^i \geq t\right)\right) \\
&= \mathbb{P}\left(\sum_{i=0}^{n_1} X_1^i \geq t\right) \times \mathbb{P}\left(\sum_{i=0}^{n_2} X_2^i \geq t\right) \quad \text{par indépendance} \\
&= \left(\int_t^{+\infty} \frac{\lambda_1}{n_1!} e^{-\lambda_1 x} (\lambda_1 x)^{n_1} dx\right) \times \left(\int_t^{+\infty} \frac{\lambda_2}{n_2!} e^{-\lambda_2 x} (\lambda_2 x)^{n_2} dx\right)
\end{aligned}$$

On note alors N le nombre de machines (des deux types) sorties du stocks quand le système tombe en panne, et Z_0 la durée écoulée avant la première panne d'une machine. On note Z_i la durée écoulée entre la i-ème panne et la $(i+1)$-ème panne. La durée de vie totale du système est donc :

$$Z = \sum_{i=0}^{N} Z_i$$

La $(N+1)$-ème panne est donc la panne fatale au système.

[1]Elle se montre aussi à l'aide des fonctions caractéristiques.

(c) Montrer que les variables Z_i sont i.i.d. et donner leur loi.
On pourra utiliser (après l'avoir démontré) le résultat suivant :
Si X est une variable aléatoire de loi exponentielle de paramètre λ, alors

$$\mathbb{P}(X \geq s+t | X \geq s) = \mathbb{P}(X \geq t) = e^{-\lambda t}$$

(on dit que X est "sans mémoire").

Si $X \rightsquigarrow \mathcal{E}(\lambda)$, alors :

$$\begin{aligned} \mathbb{P}(X \geq s+t | X \geq t) &= \frac{\mathbb{P}(X \geq s+t)}{\mathbb{P}(X \geq t)} \\ &= \frac{\exp(-\lambda(s+t))}{\exp(-\lambda t)} \\ &= e^{-\lambda s} \\ &= \mathbb{P}(X \geq s) \end{aligned}$$

Cette propriété caractérise la loi exponentielle.
En l'occurence, $Z_0 = \min(X_1^0, X_2^0) \rightsquigarrow \mathcal{E}(\lambda_1 + \lambda_2)$
Supposons que la première panne ait lieu au temps t (donc $Z_0 = t$) ; supposons (sans perte de généralité) qu'elle touche une machine de type 1. On a alors :

$$(Z_1 | Z_0 = t \wedge S_0 = 1) = (\min(X_1^1, X_2^0) | X_2^0 > t)$$

où X_1^1 désigne le temps de fonctionnement d'une nouvelle machine de type 1 (et suit donc une $\mathcal{E}(\lambda_1)$), et X_2^0 désigne le temps jusqu'à la première panne de l'"ancienne" machine de type 2.
Or $(X_2^0 | X_2^0 > t)$ suit une loi $\mathcal{E}(\lambda_2)$ (la loi exponentielle est "sans mémoire"), et on a a nouveau

$$(Z_1 | Z_0 = t \wedge S_0 = 1) \rightsquigarrow \mathcal{E}(\lambda_1 + \lambda_2)$$

Un raisonnement similaire dans le cas d'une panne de type 2 montre que la loi de Z_1 est indépendante de Z_0 et S_0 :

$$Z_i \rightsquigarrow \mathcal{E}(\lambda_1 + \lambda_2)(i.i.d)$$

(d) Préciser l'ensemble des valeurs possibles pour la variable N et en donner la loi.

Remarque : Attention à la convention de l'exercice : il y a $(n_1 + 1)$ machines de type 1 : n_1 sont en stock, et une est déjà présente dans le système. Il y a donc panne générale quand le stock est épuisé pour un type de machine, **et qu'en plus la dernière machine du même type tombe en panne**.
On a $\min(n_1, n_2) \leq N \leq n_1 + n_2 - 1$. En outre N ne peut prendre la valeur $k \in [\![\min(n_1, n_2, n_1 + n_2 - 1]\!]$ que dans deux cas :

– les n_1 machines de types 1 sont sorties du stock, $(k-n_1)$ machines de type 2 sont sorties du stock, et la dernière machine de type 1 tombe en panne (rappelons qu'il y a (n_1+1) machines de type 1).
– cas symétrique en échangeant type 1 et type 2.

Notons $p_1 = \frac{\lambda_1}{\lambda_1+\lambda_2}$ et$p_2 = \frac{\lambda_2}{\lambda_1+\lambda_2}$. Une nouvelle panne a une probabilité p_1 de concerner une machine de type 1, p_2 de type 2.

On a alors

$$\forall k \geq n_1, n_2, \quad \mathbb{P}(N=k) = p_1 C_k^{n_1} p_1^{n_1} p_2^{k-n_1} + p_2 C_k^{n_2} p_2^{n_2} p_1^{k-n_2}$$

Ainsi, si par exemple $n_1 \leq k < n_2$, on a $\mathbb{P}(N=k) = p_1 C_k^{n_1} p_1^{n_1} p_2^{k-n_1}$.

(e) On admet que N et les Z_i sont indépendantes. Calculer $\mathbb{E}\left(Z|N\right)$ en fonction de N, λ_1 et λ_2. Donner l'expression de $\mathbb{E}\left(Z\right)$ en fonction de $\mathbb{E}\left(N\right)$, λ_1 et λ_2.

On a

$$\begin{aligned}
\mathbb{E}\left(Z|N=n\right) &= \mathbb{E}\left(\sum_{i=0}^{n} Z_i\right) \\
&= (n+1)\mathbb{E}\left(Z_1\right) \\
&= \frac{n+1}{\lambda_1+\lambda_2}
\end{aligned}$$

Or $\mathbb{E}\left(Z\right) = \mathbb{E}\left(\mathbb{E}\left(Z|N\right)\right)$, donc finalement

$$\mathbb{E}\left(Z\right) = \frac{\mathbb{E}\left(N\right)+1}{\lambda_1+\lambda_2}$$

Exercice corrigé 3

Ecrire la vraisemblance et déterminer une statistique exhaustive pour un échantillon de n observations i.i.d. de lois :

☞ Q1 loi de Poisson de paramètre λ :

$$P(X=k) = e^{-\lambda}\frac{\lambda^k}{k!}, \;\; k \in \mathbb{N}^*$$

Rappel : la statistique T est dite *exhaustive ssi* la loi conditionnelle $\mathbb{P}_\theta\left(X \in \cdot | T\left(X\right)\right)$ est indépendante de θ.

Théorème de factorisation : T est exhaustive *ssi*

$$\exists g_\theta, h : (\mathbb{R} \to \mathbb{R}) \;\; / \; \forall x \in \mathbb{R}, \;\; p_\theta(x) = g_\theta\left(T\left(x\right)\right) h(x)$$

où g_θ dépend de θ mais pas h.

On retrouve les statistiques exhaustives en écrivant la vraisemblance du modèle et en appliquant le théorème de factorisation.

Dans le cas présent le paramètre du modèle est $\theta \in \Theta = \mathbb{R}$. On a pour $x \in \mathbb{R}^n$

$$L_n(x) = e^{-n\lambda} \frac{\lambda^{\sum_{i=1}^n x_i}}{\prod_{i=1}^n x_i!}$$

D'où une statistique exhaustive : $S(x_1, ..., x_n) = \sum_{i=1}^n x_i$.

☞ Q2

> loi de Pareto de paramètres α et θ avec $\alpha > 1$, $\theta > 0$ de densité :
>
> $$f(x) = \frac{\alpha - 1}{\theta} \left(\frac{\theta}{x}\right)^\alpha \mathbb{1}_{[\theta;+\infty[}(x)$$

Le paramètre du modèle étant $(\alpha, \theta) \in \Theta = \mathbb{R}^2$, on a

$$L_n(x) = \left(\frac{\alpha - 1}{\theta}\right)^n \frac{\theta^{n\alpha}}{\exp(\alpha \sum_{i=1}^n \log(x_i))} \mathbb{1}_{[\theta,+\infty]}(\min x_i)$$

D'où **une** statistique exhaustive :

$$S(x_1, ..., x_n) = \left(\sum_{i=1}^n \log(x_i), \min x_i\right)$$

☞ Q3

> loi de Weibull de paramètre α et θ avec $\alpha > 0, \theta > 0$ de densité :
>
> $$f(x) = \alpha\theta x^{\alpha-1} e^{-\theta x^\alpha} \mathbb{1}_{[0;+\infty[}(x)$$

Cette fois le paramètre du modèle est $(\alpha, \theta) \in \Theta = \mathbb{R}^{+*2}$. On a

$$L_n(x) = \alpha^n \theta^n \left(\prod_{i=1}^n x_i\right)^{\alpha-1} \exp\left(-\theta \sum_{i=1}^n x_i^\alpha\right) \mathbb{1}_{\mathbb{R}^+}(\min_i x_i)$$

Dans ce cas, on ne peut exhiber de statistique exhaustive autre que $T(X_1, \ldots, X_n) = (X_1, \ldots, X_n)$. En revanche, si on estimait uniquement θ (α constante connue), alors $T(x) = \sum_{i=1}^n x_i^\alpha$ serait une statistique exhaustive.

☞ Q4

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

La densité[2] de la loi $\mathcal{U}_{[0,\theta]}$ est

$$f_{\mathcal{U}_{[0,\theta]}}(x) = \frac{1}{\theta}\mathbb{1}_{x\geq 0}\mathbb{1}_{x\leq\theta}$$

et on a donc

$$\begin{aligned} L_n\left(x_1,\ldots,x_n\right) &= \frac{1}{\theta^n}\mathbb{1}_{x_1,\ldots,x_n\geq 0}\mathbb{1}_{x_1,\ldots,x_n\leq\theta} \\ &= \frac{1}{\theta^n}\mathbb{1}_{\min x_i\geq 0}\mathbb{1}_{\max x_i\leq\theta} \end{aligned}$$

et une statistique exhaustive est donc $T(x) = \max_{i\in[\![1,n]\!]} x_i$

Exercice corrigé 4

On veut compter le nombre θ de poissons dans un lac fermé. Pour cela, on tire un poisson au hasard, on le marque et on le remet dans le lac. On tire un second poisson. S'il est déjà marqué, on en prend note et on le remet dans le lac. Sinon, on le marque à son tour et on le remet dans le lac. Et ainsi de suite.

On tire n poissons selon la procédure ci-dessus. Au n-ième tirage, l'observation consiste en une variable aléatoire Y_n qui vaut 1 si le poisson est déjà marqué, 0 sinon. Par définition, on a $Y_1 = 0$. Le but de l'exercice est de montrer que :

$$R_n = \sum_{i=1}^{n} Y_i$$

est une statistique exhaustive pour θ.

☞ Q1

Montrer que :

$$\mathbb{P}\left(Y_n = y_n, ..., Y_1 = y_1\right) = \left\{\begin{array}{ll} & \mathbb{P}\left(Y_n = y_n | Y_{n-1} = y_{n-1},\ldots,Y_1 = y_1\right) \\ \times & \mathbb{P}\left(Y_{n-1} = y_{n-1} | Y_{n-2} = y_{n-2},\ldots,Y_1 = y_1\right) \\ \vdots & \\ \times & \mathbb{P}\left(Y_1 = y_1\right) \end{array}\right.$$

[2]En toute rigueur le modèle $\left(\mathcal{U}_{[0,\theta]}\right)_{\theta\in\mathbb{R}^{+*}}$ n'est pas dominé (il n'y a pas une unique mesure μ qui domine toutes les probabilités $\mathcal{U}_{[0,\theta]}$).

Mais pour tout $M > 0$, le sous-modèle $\left(\mathcal{U}_{[0,\theta]}\right)_{\theta\in]0,M]}$ est dominé (par la mesure particulière qu'est la probabilité de $\mathcal{U}_{[0,M]}$). On exhibe donc ici une statistique $T(X)$ qui est exhaustive pour tout sous-modèle : elle est donc telle que

$$\forall M, \forall\theta\in]0,M], \text{ la loi de } X|T(X) \text{ ne dépend pas de } \theta$$

et donc

$$\forall\theta > 0, \text{ la loi de } X|T(X) \text{ ne dépend pas de } \theta$$

et donc $T(X)$ est bien exhaustive dans le modèle originel.

On applique $n-1$ fois la *formule du conditionnement* $\mathbb{P}(\mathcal{A}\cap\mathcal{B}) = \mathbb{P}(\mathcal{A}|\mathcal{B})\mathbb{P}(\mathcal{B})$:
Ici :

$$
\begin{aligned}
& \mathbb{P}(Y_n = y_n, \dots, Y_1 = y_1) \\
= \; & \mathbb{P}(Y_n = y_n | Y_{n-1} = y_{n-1}, \dots, Y_1 = y_1) \times \mathbb{P}(Y_{n-1} = y_{n-1}, \dots, Y_1 = y_1) \\
\vdots \; & \text{(par récurrence)} \\
= \; & \left\{ \begin{array}{ll} & \mathbb{P}\left(Y_n = y_n | Y_{n-1} = y_{n-1}, \dots, Y_1 = y_1\right) \\ \times & \mathbb{P}\left(Y_{n-1} = y_{n-1} | Y_{n-2} = y_{n-2}, ..., Y_1 = y_1\right) \\ \vdots & \\ \times & \mathbb{P}\left(Y_1 = y_1\right) \end{array} \right.
\end{aligned}
$$

☞ Q2

> Montrer que la loi conditionnelle de Y_n sachant $R_{n-1} = r_{n-1}$ est une Bernoulli de paramètre :
>
> $$\frac{n - r_{n-1} - 1}{\theta}$$
>
> et en déduire que la vraisemblance est proportionnelle à :
>
> $$\prod_{i=1}^{n} \frac{\left(\theta - i + 1 + r_{i-1}\right)^{1-y_i}}{\theta} \tag{1}$$

Soit p le nombre de poissons distincts qui ont été tirés au cours des n premiers tirages. Supposons que $R_n = r_n$: il y a eu r_n retirages d'un poisson déjà tiré au moins une fois, donc $p + r_n = n$. Ainsi après $n-1$ tirages le nombre de poissons distincts tirés est $n-1-r_{n-1}$, et la proportion de poissons qui ont été tirés (et sont donc marqués) est

$$\boxed{\frac{n-1-r_{n-1}}{\theta}}$$

Par conséquent

$$\mathbb{P}\left(Y_n = y_n | R_{n-1} = r_{n-1}\right) = \left(\frac{n-1-r_{n-1}}{\theta}\right)^{y_n} \left(1 - \frac{n-1-r_{n-1}}{\theta}\right)^{1-y_n}$$

Donc d'après le résultat précédent la vraisemblance s'écrit

$$\begin{aligned}
& \mathbb{P}(Y_1 = y_1, \ldots, Y_n = y_n; \theta) \\
= & \prod_{i=1}^{n} \mathbb{P}(Y_i = y_i | Y_{i-1} = y_{i-1}, \ldots, Y_1 = y_1) \\
= & \prod_{i=1}^{n} \mathbb{P}\left(Y_i = y_i | R_{i-1} = r_{i-1}, Y_{i-1} = (r_{i-1} - y_{i-1} - \cdots - y_1), Y_{i-2} = y_{i-2}, \ldots, Y_1 = y_1\right) \\
= & \prod_{i=1}^{n} \mathbb{P}(Y_i = y_i | R_{i-1} = r_{i-1}) \quad \text{car la loi de } Y_i | R_{i-1}, Y_{i-1}, \ldots, Y_1 \text{ ne dépend pas de } Y_1, \ldots, Y_{i-1} \\
= & \prod_{i=1}^{n} \mathbb{P}\left(Y_i = y_i | Y_{i-1} = y_{i-1}, \ldots, Y_1 = y_1\right) \\
= & \prod_{i=1}^{n} \left(\left(\frac{i-1-r_{i-1}}{\theta} \right)^{y_i} \left(\frac{\theta - i + 1 + r_{i-1}}{\theta} \right)^{1-y_i} \right) \\
= & \prod_{i=1}^{n} \left(\frac{1}{\theta} (i-1-r_{i-1})^{y_i} (\theta - i + 1 + r_{i-1})^{1-y_i} \right)
\end{aligned}$$

qui est proportionnel à

$$\boxed{\prod_{i=1}^{n} \frac{1}{\theta} (\theta - i + 1 + r_{i-1})^{1-y_i}}$$

par le facteur $\prod_{i=1}^{n} (i-1-r_{i-1})^{y_i}$ qui ne dépend pas de θ et n'a donc pas d'incidence sur la détermination d'une statistique exhaustive (d'après le théorème de factorisation).[3]

☞ Q3

> Montrer que l'expression `(1)` se réécrit :
>
> $$\frac{1}{\theta^n} \frac{(\theta - 1)!}{(\theta - n - 1 + r_n)!}$$

On a successivement

$$\prod_{i=1}^{n} \frac{(\theta - i + 1 + r_{i-1})^{1-y_i}}{\theta} = \frac{1}{\theta^n} \prod_{i=1}^{n} (\theta - (i - 1 - r_{i-1}))^{1-y_i}$$

[3] Ce facteur aurait bien entendu une influence pour le calcul de l'information de Fisher ou du maximum de vraisemblance.

Or $\forall i \in [\![1,n]\!]$, $(y_i = 1)$:$(\theta - (i-1-r_{i-1}))^{1-y_i} = 1$, donc

$$= \frac{1}{\theta^n} \times \prod_{\begin{cases} 1 \leq i \leq n \\ y_i = 1 \end{cases}} 1 \times \prod_{\begin{cases} 1 \leq i \leq n \\ y_i = 0 \end{cases}} (\theta - (i-1-r_{i-1}))$$

$$= \frac{1}{\theta^n} \prod_{\begin{cases} 1 \leq i \leq n \\ j_i = i-1-r_{i-1} \\ y_i = 0 \end{cases}} (\theta - j_i)$$

Or $\phi : i \mapsto i-1-r_{i-1}$ associe au nombre de tirages le nombre de poissons distincts tirés. Donc restreinte à $\mathcal{I}_0 = \{i \leq n \,/\, y_i = 0\}$ (l'ensemble des i tels que le i-ième poisson tiré est un nouveau poisson pas encore marqué), c'est une bijetion de $\mathcal{I}_0$ sur $[\![\phi(1), \phi(i_0)]\!]$ où $i_0 = \max \mathcal{I}_0$. Donc

$$\prod_{\begin{cases} 1 \leq i \leq n \\ j_i = i-1-r_{i-1} \\ y_i = 0 \end{cases}} (\theta - j_i) = \prod_{1 \leq j \leq \phi(i_0)} (\theta - j)$$

Remarquons enfin que par définition de i_0, $\forall j > i_0, j \notin \mathcal{I}_0$ et donc $y_j = 1$ de sorte que

$$\phi(i_0) = i_0 - 1 - r_{i_0} = (i_0 + (j - i_0)) - 1 - (r_{i_0} + (j - i_0)) = j - 1 - r_{i_0 + (j - i_0)} = j - 1 - r_j$$

de sorte que $\phi(i_0) = n - 1 - r_n$ et en définitive la vraisemblance est proportionnelle à

$$\frac{1}{\theta^n} \prod_{1 \leq j \leq n-1-r_n} (\theta - j) = \boxed{\frac{1}{\theta^n} \frac{(\theta-1)!}{(\theta-n-1+r_n)!}}$$

☞ Q4 $\boxed{\text{En déduire que } R_n \text{ est une statistique exhaustive pour } \theta.}$

On a alors

$$l(y, \theta) = g_\theta(R_n(y)) \times h(y)$$

où

$$h(y) = \prod_{i=1}^{n} (i - 1 - R_{i-1}(y))^{y_i}$$

qui est indépendant de θ, et

$$g_\theta(R_n(y)) = \frac{1}{\theta^n} \frac{(\theta - 1)!}{(\theta - n - 1 + R_n(y))!}$$

qui ne dépend de y qu'au travers de $R_n(y)$.

Donc d'après le théorème de factorisation R_n est une statistique exhaustive. Ainsi pour estimer le nombre de poissons dans le lac, la connaissance de tous les tirages est superflue et le seul nombre total de poissons tirés plus d'une fois suffit.

2 Travaux Dirigés n°2

Exercice corrigé 1

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

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

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

Méthode rapide mais pas naturelle :

L'idée ici est de considérer tous les réordonnancements de r variables parmi n telles que les r premières soient croissantes et plus petites que les $n-r$ suivantes.
Plus précisément, soit $\mathcal{A}_n^r$ l'ensemble des parties à $r \leq n$ éléments d'un ensemble à $n \geq 1$ éléments.
Alors pour $y_1 < \ldots < y_r$ et $dy_1, \ldots, dy_r$, avec $0 < dy_i < \frac{y_{i+1}-y_i}{2}$

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

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

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

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

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

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

En divisant et en faisant successivement tendre $dy_r \to 0^+, \ldots, dy_1 \to 0^+$ il vient [4]

[4]Précisément :

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

puis

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

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

<u>Méthode simple mais longue :</u>

L'idée est ici de calculer d'abord la loi de $(Y_1,\ldots,Y_n)$, puis d'en tirer celle de $(Y_1,\ldots,Y_r)$ au moyen de la relation $f_X(x)=\int_{\mathcal{Y}} f_{X,Y}(x,y)dy$.
Donnons-nous donc $\mathcal{A}$ mesurable quelconque.
Alors

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

et donc presque sûrement

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

de sorte que

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

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

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

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

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

On en conclut que

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

---

et finalement

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

On a alors

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

De façon similaire

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

On en conclut donc par récurrence que

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

Loi de Y_r seul :

On a :

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

Par suite

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

Et par une récurrence immédiate il vient

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

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

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

et symétriquement celle du plus grand

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

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

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

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

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

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

et donc d'après le résultat précédent

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

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

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

On a

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

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

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

Exercice corrigé 2

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

☞ Q1

une loi de Poisson de paramètre λ :

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

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

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

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

☞ Q2

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

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

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

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

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

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

☞ Q3

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

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

On a

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

d'où

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

Il reste alors à exprimer $\mathbb{E}\left(X^{\alpha}(\ln X)^i\right)$ pour $i \in \{1,2\}$; ce calcul peut être fait au moyen des intégrales eulériennes.
Remarquons pour ce faire que pour $p \in \mathbb{N}^*$ on a

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

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

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

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

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

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

En particulier lorsque $x = 2$ il vient :

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

En définitive on a donc

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

de sorte que

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

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

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

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

La vraisemblance du modèle est

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

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

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

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

$\max_i x_i$). Sur $]0, \max_i x_i[$ on peut néanmoins calculer par analogie[5]

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

Remarquons que

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

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

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

et que

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

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

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

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

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

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

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

Exercice corrigé 3

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

☞ Q1 Soit F la fonction de répartition des Y_i. On cherche à estimer la "fonction de survie" de chaque composant $\bar{F}(t) = 1 - F(t)$.
Calculer $\bar{F}(t)$ en fonction de t et θ.

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

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

en conséquence de quoi $\boxed{\bar{F}(t) = e^{-\frac{t}{\theta}}}$.

☞ Q2 Calculer l'estimateur du maximum de vraisemblance de θ et en déduire un estimateur convergent $\hat{F}(t)$ de $\bar{F}(t)$. Que peut-on dire du biais de $\hat{F}(t)$?

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Soit T la variable aléatoire définie par :

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

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

☞ Q4

(a) $\boxed{\text{Déterminer la loi de } Y_1 \text{ conditionnellement à } S.}$

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

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

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

– puis de même

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

– (récurrence ...)
– et finalement

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

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

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

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

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

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

et donc en définitive

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

(b) Calculer

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

Comment s'appelle cet estimateur ?

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

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

En d'autres termes,

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

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

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

On a successivement

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

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

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

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

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

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

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

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

Exercice corrigé 4

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

☞ Q1 Quelle est la fonction score du modèle, notée $S(X; \theta)$?
Donner l'expression de l'information de Fisher $I_X(\theta)$.

On a

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

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

L'information de Fisher pour une observation est définie par

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

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

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

(voir exercice 2)

☞ Q2

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

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

où s est un seuil connu.

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

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

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

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

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

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

Le score est alors

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

où l'un seulement des deux termes est non-nul puisque $y \in \{0,1\}$.
Calculons alors le score pour le modèle en X : on a pour $y \in \{0,1\}$

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

Or

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

et en outre

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

tandis que

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

donc en définitive

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

☞ Q3 En déduire alors que $I_X(\theta) >> I_Y(\theta)$, où $I_Y(\theta)$ est l'information de Fisher associée à Y (l'inégalité s'entend au sens des matrices symétriques).
Quelle interprétation pouvez-vous donner à l'inégalité ci-dessus ?

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

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

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

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

En l'occurence il vient

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

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

3 Travaux Dirigés n°3

Exercice corrigé 1

Soit Y un vecteur aléatoire de taille N et X une matrice aléatoire à N lignes et K colonnes. Soit $\theta \to S(\theta)$ une application de classe C^1 définie sur un voisinage de 0 dans $\mathbb{R}$, à valeurs dans l'ensemble des matrices symétriques définies positives de taille N. On suppose que $S(0) = I_N$ et on note : $A = \frac{\partial S}{\partial \theta}(0)$.
On rappelle par ailleurs que

$$\begin{aligned}\frac{\partial \ln(|S|)}{\partial \theta} &= Tr\left(S^{-1}S'\right)\\ \frac{\partial S^{-1}}{\partial \theta} &= -S^{-1}S'S^{-1}\end{aligned}$$

On considère le modèle linéaire (conditionnel) gaussien suivant :

$$\mathbb{P}_Y^X = \mathcal{N}\left(Xb, S(\theta)\right)$$

où $b \in \mathbb{R}^K$ et $\theta \in \mathbb{R}$ sont les paramètres inconnus.

☞ Q1 $\boxed{\text{Ecrire la vraisemblance du modèle.}}$

On se donne Y variable aléatoire sur $\mathcal{M}_{N,1}(\mathbb{R})$ et X sur $\mathcal{M}_{N,K}(\mathbb{R})$.
La vraisemblance (conditionnelle) du modèle normal s'écrit

$$f(Y|X,b,\theta) = \frac{1}{(2\pi)^{\frac{N}{2}}\sqrt{|S(\theta)|}}\exp\left(-\frac{1}{2}(Y-Xb)'\cdot S(\theta)^{-1}\cdot(Y-Xb)\right)$$

Note : il s'agit bien d'un nombre réel car $(Y - Xb) \in \mathcal{M}_{N,1}(\mathbb{R})$.
La log-vraisemblance conditionnelle s'écrit donc

$$\ln L(Y|X,b,\theta) = -\frac{N}{2}\ln(2\pi) - \frac{1}{2}\ln|S(\theta)| - \frac{1}{2}(Y-Xb)'\cdot S(\theta)^{-1}\cdot(Y-Xb)$$

☞ Q2 $\boxed{\text{Ecrire le vecteur score (de taille } K+1\text{) et vérifier qu'il est d'espérance nulle.}}$

Le vecteur score $\mathcal{S}(b,\theta) \in \mathcal{M}_{K+1,1}(\mathbb{R})$ se décompose entre les composantes suivant b, soit $\mathcal{S}_b(b) \in \mathcal{M}_{K,1}(\mathbb{R})$, et celle suivant θ, soit $\mathcal{S}_\theta(\theta) \in \mathbb{R}$, que nous calculons alternativement.
– Composantes suivant b :

$$\begin{aligned}\mathcal{S}_b(b) &= \frac{\partial \ln L(Y|X,\cdot,\theta)}{\partial b}(b)\\ &= -\frac{1}{2}\left(\frac{\partial}{\partial b}\left(b \mapsto (Y-Xb)'\cdot S(\theta)^{-1}\cdot(Y-Xb)\right)\right)(b)\\ &= -\frac{1}{2}\left(\frac{\partial}{\partial b}\left(b \mapsto Y'S(\theta)^{-1}Y - b'X'S(\theta)^{-1}Y - Y'S(\theta)^{-1}Xb + b'X'S(\theta)^{-1}Xb\right)\right)(b)\end{aligned}$$

Rappelons que pour $A \in \mathcal{M}_{n,p}(\mathbb{R})$ on a $\frac{\partial}{\partial b}\left(b \mapsto b'A\right) = A$ et $\frac{\partial}{\partial b}\left(b \mapsto A'b\right) = A'$, donc pour Σ symétrique

$$\frac{\partial}{\partial b}\left(b \mapsto b'\Sigma b\right)(b) = 2\Sigma b$$

Donc

$$\begin{aligned}
\mathcal{S}_b(b) &= -\frac{1}{2}\left(0 - X'S(\theta)^{-1}Y - X'S(\theta)^{-1}Y + 2(X'S(\theta)^{-1}X)b\right) \\
&= X' \cdot S(\theta)^{-1} \cdot (Y - Xb)
\end{aligned}$$

– Composante suivant θ :

$$\begin{aligned}
\mathcal{S}_\theta(\theta) &= \frac{\partial \ln L(Y|X, b, \cdot)}{\partial \theta}(\theta) \\
&= -\frac{1}{2}\left(\frac{\partial}{\partial \theta}\left(\theta \mapsto \ln|S(\theta)| + (Y - Xb)' \cdot S(\theta)^{-1} \cdot (Y - Xb)\right)\right)(\theta)
\end{aligned}$$

Or

$$\left(\frac{\partial}{\partial \theta}\left(\theta \mapsto \ln|S(\theta)|\right)\right)(\theta) = Tr\left(S(\theta)^{-1}\frac{\partial S}{\partial \theta}(\theta)\right)$$

et

$$\left(\frac{\partial}{\partial \theta}\left(\theta \mapsto S(\theta)^{-1}\right)\right)(\theta) = -S(\theta)^{-1} \cdot \frac{\partial S}{\partial \theta}(\theta) \cdot S(\theta)^{-1}$$

La première assertion découle en effet directement du théorème 2 du chapitre 3 de *"Matrix Differential Calculus"*, Jan R. Magnus et Heinz Neudecker, qui stipule que si F : $\left(\mathcal{M}_{n,p}(\mathbb{R}) \to \mathcal{S}_k^+(\mathbb{R})\right)$ est k fois différentiable, alors $\ln|F|$ l'est et admet pour différentielle $d\ln|F| = Tr\,(F^{-1})dF$ (il suffit de poser $n = p = 1$ et $F = S$). Elle peut même être redémontrée directement, en notant λ_i^A la i-ième valeur propre d'une matrice A (sans ordre de multiplicité) :

$$\begin{aligned}
\left(\frac{\partial}{\partial \theta}\left(\theta \mapsto \ln|S(\theta)|\right)\right)(\theta) &= \left(\frac{\partial}{\partial \theta}\left(\theta \mapsto \ln\left(\Pi_{i=1}^n \lambda_i^{S(\theta)}\right)\right)\right)(\theta) \\
&= \sum_{i=1}^n \left(\frac{\partial}{\partial \theta}\left(\theta \mapsto \left(\ln \lambda_i^{S(\theta)}\right)\right)\right)(\theta) \\
&= \sum_{i=1}^n \left(\frac{\partial \lambda_i^{S(\theta)}}{\partial \theta}(\theta)\right)\left(\lambda_i^{S(\theta)}\right)^{-1} \\
&= \sum_{i=1}^n \left(\lambda_i^{\frac{\partial S(\theta)}{\partial \theta}(\theta)}\right)\left(\lambda_i^{S(\theta)^{-1}}\right) \\
&= Tr\left(\left(\frac{\partial S(\theta)}{\partial \theta}(\theta)\right) \times \left(S(\theta)^{-1}\right)\right)
\end{aligned}$$

Une preuve de la seconde assertion est également proposée dans le théorème 3 de ce même ouvrage, ou peut aussi être obtenue en développant l'égalité

$$\frac{\partial}{\partial \theta}\left(S(\theta) \times S(\theta)^{-1}\right) = \frac{\partial I_N}{\partial \theta} = (0)$$

Il en ressort que

$$\begin{aligned}\mathcal{S}_\theta(\theta) &= -\frac{1}{2}Tr\left(S(\theta)^{-1}\frac{\partial S}{\partial \theta}(\theta)\right) + \frac{1}{2}\underbrace{(Y-Xb)'\left(S(\theta)^{-1}\cdot\frac{\partial S}{\partial \theta}(\theta)\cdot S(\theta)^{-1}\right)(Y-Xb)}_{\in\mathbb{R}} \\ &= -\frac{1}{2}Tr\left(\left(S(\theta)^{-1}\frac{\partial S}{\partial \theta}(\theta)S(\theta)^{-1}\right)(S(\theta)-(Y-Xb)'(Y-Xb))\right)\end{aligned}$$

Reste à montrer que le score est centré :
- On a d'un part $\mathbb{E}(Y|X,b) = Xb$,
  en conséquence de quoi $\mathbb{E}(Y-Xb|X,b) = (0)$
  de sorte que $\mathbb{E}(\mathcal{S}_b(b)|X,b) = (0)$: les composantes selon b sont centrées.
- Et d'autre part $\mathbb{E}_\theta((Y-Xb)(Y-Xb)') = S(\theta)$,
  de sorte que $\mathbb{E}(\mathcal{S}_\theta(\theta)|X,b) = 0$.

En conséquence, $\mathbb{E}_{b,\theta}(\mathcal{S}(b,\theta)|X,b,\theta) = (0)$: le vecteur score est centré.

☞ Q3 | Calculer la matrice d'information du modèle en $(b,\theta=0)$.
Le résultat s'exprime <u>très</u> simplement en fonction de X et A.

Le calcul de la matrice d'information de Fisher est encore plus convivial, car c'est un prétexte au calcul des dérivées secondes de la log-vraisemblance (le calcul alternatif de la variance du score calculé est laissé en exercice).

- Dérivée seconde selon (b,b) :

$$\frac{\partial^2 \ln L(Y|X,\cdot,\cdot)}{\partial b \partial b'}(\theta,b) = -X'\cdot S(\theta)^{-1}\cdot X$$

  Donc, comme $S(0) = I_N$,

$$\mathbb{E}_{b,\theta=0}\left(\frac{\partial^2 \ln L(Y|X,\cdot,\cdot)}{\partial b \partial b'}(\theta,b)\right) = -X'X$$

- Dérivée seconde selon (b,θ) :

$$\frac{\partial^2 \ln L(Y|X,\cdot,\cdot)}{\partial b \partial \theta}(\theta,b) = -X'\cdot S(\theta)^{-1}\frac{\partial S}{\partial \theta}(\theta)S(\theta)^{-1}\cdot(Y-Xb)$$

  Donc, comme $\mathbb{E}_{b,\theta}(Y-Xb) = 0$,

$$\mathbb{E}_{b,\theta=0}\left(\frac{\partial^2 \ln L(Y|X,\cdot,\cdot)}{\partial b \partial \theta}(\theta,b)\right) = (0)$$

- Dérivée seconde selon (θ,b) :
  Loins de nous lancer dans la dérivation selon b de $\frac{\partial \ln L(Y|X,\cdot,\cdot)}{\partial \theta}$, constatons astucieusement que $(b,\theta)\mapsto\mathcal{S}(b,\theta)$ est de classe C^2 sur un voisinage de 0 : elle est en effet la composée d'une fonction de classe C^∞, à savoir
  $\begin{pmatrix} \mathbb{R}\times\mathcal{M}_{N,N}(\mathbb{R}) & \to & \mathbb{R} \\ (x,A) & \mapsto & -\frac{N}{2}\ln 2 - \frac{1}{2}x - \frac{1}{2}(Y-Xb)'\cdot A\cdot(Y-Xb)\end{pmatrix}$, et des deux fonctions

$(M \mapsto \ln|M|)$ et $(M \mapsto M^{-1})$ qui sont de classe C^∞ sur une partie dense d'un voisinage de 0. Cauchy nous assure alors de l'égalité des dérivées croisées pour θ voisin de 0 : $\forall(b,\theta)$, $\frac{\partial^2 \ln L(Y|X,\cdot,\cdot)}{\partial b \partial \theta}(\theta,b) = \frac{\partial^2 \ln L(Y|X,\cdot,\cdot)}{\partial \theta \partial b}(\theta,b)$, ce qui permet de conclure que

$$\mathbb{E}_{b,\theta=0}\left(\frac{\partial^2 \ln L(Y|X,\cdot,\cdot)}{\partial\theta\partial b}(\theta,b)\right) = (0)$$

– Dérivée seconde selon (θ,θ) :

$$\begin{aligned}
&\mathbb{E}_{b,\theta}\left(\frac{\partial^2 \ln L(Y|X,\cdot,\cdot)}{\partial\theta^2}(\theta,b)\right)\\
= \;& \mathbb{E}_{b,\theta}\left(\frac{\partial}{\partial\theta}\left(\theta \mapsto -\frac{1}{2}Tr\left(\underbrace{\left(S(\theta)^{-1}\frac{\partial S}{\partial\theta}(\theta)S(\theta)^{-1}\right)}_{\psi(\theta)}\underbrace{(S(\theta)-(Y-Xb)'(Y-Xb))}_{h(\theta)}\right)\right)\right)\\
= \;& -\frac{1}{2}\mathbb{E}_{b,\theta}\left(Tr\left(\frac{\partial\psi}{\partial\theta}(\theta)h(\theta) + \psi(\theta)\frac{\partial h}{\partial\theta}(\theta)\right)\right)\\
= \;& -\frac{1}{2}Tr\left(\mathbb{E}_{b,\theta}\left(\frac{\partial\psi}{\partial\theta}(\theta)h(\theta)\right)\right) - \frac{1}{2}\mathbb{E}_{b,\theta}\left(Tr\left(\psi(\theta)\frac{\partial h}{\partial\theta}(\theta)\right)\right)\\
= \;& -\frac{1}{2}Tr\left(\mathbb{E}_{b,\theta}\left(\frac{\partial\psi}{\partial\theta}(\theta)\underbrace{\mathbb{E}_b\left(h(\theta)|\theta\right)}_{=0}\right)\right) - \frac{1}{2}\mathbb{E}_{b,\theta}\left(Tr\left(\psi(\theta)\frac{\partial S}{\partial\theta}(\theta)\right)\right)\\
= \;& -\frac{1}{2}Tr\left(\psi(\theta)\frac{\partial S}{\partial\theta}(\theta)\right)
\end{aligned}$$

Donc

$$\mathbb{E}_{b,\theta}\left(-\frac{\partial^2 \ln L(Y|X,\cdot,\cdot)}{\partial\theta^2}(\theta,b)\right) = \frac{1}{2}Tr\left(S(\theta)^{-1}\;\frac{\partial S}{\partial\theta}(\theta)\;S(\theta)^{-1}\;\frac{\partial S}{\partial\theta}(\theta)\right)$$

et par suite, sachant que $S(0) = I_N$ et en notant $A = \frac{\partial S}{\partial\theta}(0)$

$$\mathbb{E}_{b,\theta=0}\left(-\frac{\partial^2 \ln L(Y|X,\cdot,\cdot)}{\partial\theta^2}(\theta,b)\right) = \frac{1}{2}Tr\left(A^2\right)$$

On a donc finalement :

$$\boxed{I(b,0) = \left(\begin{array}{c|c} X'X & \begin{matrix}0\\ \dots\\ 0\end{matrix} \\ \hline 0\dots0 & \frac{1}{2}Tr\left(A^2\right)\end{array}\right)}$$

☞ Q4

> Dans cette question, on suppose que $S(\theta)$ est la matrice de terme général :
> $$S(\theta)_{i,j} = \left(\theta^{|i-j|}\right)_{1\leq i,j\leq N}$$
> Calculer $I_{b,\theta}(b,0)$.

On a $S(\theta) = \left(\theta^{|i-j|}\right)_{(i,j)\in[\![1,N]\!]^2}$.

Or $\frac{\partial}{\partial\theta}\left(\theta\mapsto\theta^k\right)(\theta) = \begin{cases} 0 & \text{si } k=0 \\ 1 & \text{si } k=1 \\ \theta^{k-1} & \text{sinon}\end{cases}$ En particulier $A = \frac{\partial S}{\partial\theta}(0) = \left(\mathbb{1}_{|i-j|=1}\right)_{(i,j)\in[\![1,N]\!]^2}$.

En conséquence

$$A^2 = \begin{pmatrix} 1 & \star & \star & \dots & \star & \star \\ \star & 2 & \star & \dots & \star & \star \\ \star & \star & 2 & \dots & \star & \star \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \star & \star & \star & \dots & 2 & \star \\ \star & \star & \star & \dots & \star & 1 \end{pmatrix}$$

et donc $Tr\left(A^2\right) = 2(N-1)$.
Ainsi,

$$\boxed{I_{b,\theta}(b,0) = \begin{pmatrix} X'X & 0 \\ 0 & N-1 \end{pmatrix}}$$

Exercice corrigé 2

☞ Q1

> Calculer l'estimateur du maximum de vraisemblance (e.m.v.) $\hat{p}$ de p dans le modèle
> $$X_i \underset{iid}{\leadsto} \mathcal{B}(1,p)$$
> et calculer la loi limite de $\sqrt{n}\,(\hat{p}-p)$.

On a

$$f_X(x,p) = p^{\sum_{i=1}^n x_i}(1-p)^{n-\sum_{i=1}^n x_i}\mathbb{1}_{x\in\{0,1\}}$$

d'où on tire que

$$\frac{\partial \ln f_X}{\partial p}(x,p) = \frac{\sum_{i=1}^n x_i}{p} - \frac{n-\sum_{i=1}^n x_i}{1-p}$$

en conséquence de quoi un estimateur $\widehat{p}$ de p vérifie

$$(1-\widehat{p})\sum_{i=1}^n x_i = \widehat{p}(n-\sum_{i=1}^n x_i)$$

soit $\widehat{p} = \frac{1}{n}\sum_{i=1}^n x_i = \overline{X}$ (la réciproque étant immédiate).

On a par ailleurs

$$\frac{\partial^2 \ln f_X}{\partial p^2}(x,\theta) = -\frac{1}{p^2}\sum_{i=1}^n x_i - \frac{n-\sum_{i=1}^n x_i}{(1-p)^2}$$

d'où $I_n(\theta) = \frac{n}{p(1-p)}$ et donc d'après le Théorème Central Limite il vient

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

☞ Q2

> Calculer l'e.m.v. $(\hat{m}, \hat{\sigma}^2)$ de (m, σ^2) dans le modèle
>
> $$X_i \underset{iid}{\leadsto} \mathcal{N}\left(m, \sigma^2\right)$$
>
> et donner la loi limite du vecteur
>
> $$\sqrt{n}\begin{pmatrix} \hat{m}-m \\ \hat{\sigma^2}-\sigma^2 \end{pmatrix}$$

Le paramètre à estimer est ici (m,σ^2) (et pas (m,σ) : on s'abstiendra d'écrire σ^4 mais plutôt $(\sigma^2)^2$). On a

$$f_X(x,m,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}^n} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^n (x_i-m)^2}$$

et donc

$$\frac{\partial \ln f_X}{\partial m}(x,m,\sigma^2) = \frac{1}{\sigma^2}\sum_{i=1}^n (x_i - m)$$

Par ailleurs,

$$\frac{\partial \ln f_X}{\partial(\sigma^2)}(x,m,\sigma^2) = -\frac{n}{2\sigma^2} + \frac{1}{2\left(\sigma^2\right)^2}\sum_{i=1}^n (x_i-m)^2$$

L'estimateur du maximum de vraisemblance $\begin{pmatrix} \widehat{m} \\ \widehat{\sigma^2} \end{pmatrix}$ de $\begin{pmatrix} m \\ \sigma^2 \end{pmatrix}$ vérifie alors pour tout x

$$\frac{\partial \ln f_X}{\partial \begin{pmatrix} m \\ \sigma^2 \end{pmatrix}}\left(\begin{pmatrix} \widehat{m}(x) \\ \widehat{\sigma^2}(x) \end{pmatrix}\right) = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

ce dont on tire

$$\left\{ \begin{array}{r} \widehat{m}(x) = x \\ \frac{1}{\sigma^2(x)}\sum_{i=1}^n \left(x_i - \widehat{m}(x)\right) = 0 \end{array} \right.$$

soit encore

$$\begin{pmatrix} \widehat{m} \\ \widehat{\sigma^2} \end{pmatrix} : x \mapsto \begin{pmatrix} \overline{x} \\ \frac{1}{n}\sum_{i=1}^n (x_i - \overline{x})^2 \end{pmatrix}$$

On a enfin

$$\begin{aligned}
\frac{\partial^2 \ln f_X}{\partial m^2}(x,m,\sigma^2) &= -\frac{n}{\sigma^2} \\
\frac{\partial^2 \ln f_X}{\partial m \partial(\sigma^2)}(x,m,\sigma^2) &= \left(-\frac{1}{(\sigma^2)^2}\right)\sum_{i=1}^n (x_i - m) \\
\frac{\partial^2 \ln f_X}{\partial(\sigma^2)^2}(x,m,\sigma^2) &= \frac{n}{2\left(\sigma^2\right)^2} + \frac{1}{2}\sum_{i=1}^n (x_i - m)^2 \left(-\frac{2}{(\sigma^2)^3}\right)
\end{aligned}$$

De $\mathbb{E}\left(X_i - m\right) = 0$ et $\mathbb{E}\left((X_i - m)^2\right) = \sigma^2$ on tire finalement que

$$\begin{aligned}
\mathbb{E}\left(\frac{\partial^2 \ln f_X}{\partial(\sigma^2)^2}(X;m,\sigma^2)\right) &= \frac{n}{2\left(\sigma^2\right)^2} - \frac{1}{2}\frac{2}{\left(\sigma^2\right)^3}\left(\sigma^2\right) \\
&= -\frac{n}{2\left(\sigma^2\right)^2}
\end{aligned}$$

de sorte que

$$I_n(m,\sigma^2) = \begin{pmatrix} \frac{n}{\sigma^2} & 0 \\ 0 & \frac{n}{2(\sigma^2)^2} \end{pmatrix}$$

et donc en définitive

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

Les variables aléatoires (asymptotiquement normales) $\widehat{m}$ et $\widehat{\sigma^2}$ sont de covariance asymptotique nulle, donc[6] sont asymptotiquement indépendantes.

☞ Q3

> Calculer l'e.m.v. $(\hat{a}, \hat{b})$ de (a,b) dans le modèle
>
> $$X_i \underset{iid}{\leadsto} \mathcal{U}([a,b]) \quad \text{loi uniforme sur } [a,b]$$
>
> Donner la loi limite du vecteur
>
> $$n\begin{pmatrix} \hat{a} - a \\ b - \hat{b} \end{pmatrix}$$

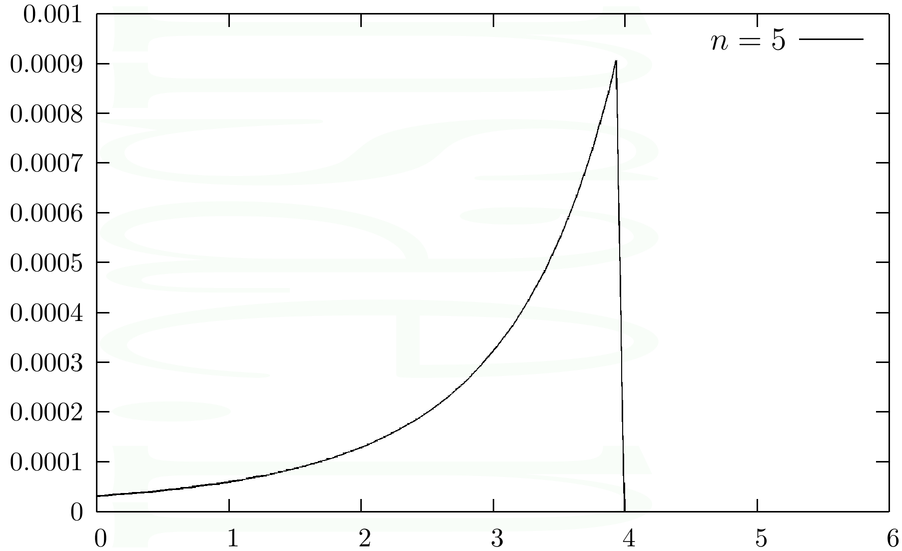
La vraisemblance s'écrit :

$$L_X(x;a,b) = \frac{1}{(b-a)^n}\mathbb{1}_{\min x_i \geq a}\mathbb{1}_{\max x_i \leq b}$$

et n'est pas dérivable en a et b (disconstinuités en $a = \min x_i$ et $b = \max x_i$). Les conditions de régularité habituelles ne sont pas vérifiées, et le point $(\widehat{a}(x), \widehat{b}(x)$ qui maximise la vraisemblance ne vérifie pas les conditions du premier ordre.

Il peut néanmoins être déterminé directement : en effet pour $x \in \mathbb{R}^n$ et $b \in \mathbb{R}$ fixés, $\begin{pmatrix}]-\infty, b[& \to & \mathbb{R}^+ \\ a & \mapsto & \frac{1}{(b-a)^n}\mathbb{1}_{\min x_i \geq a}\mathbb{1}_{\max x_i \leq b} \end{pmatrix}$ est identiquement nulle si $\max_i x_i > b$, et sinon admet pour graphe

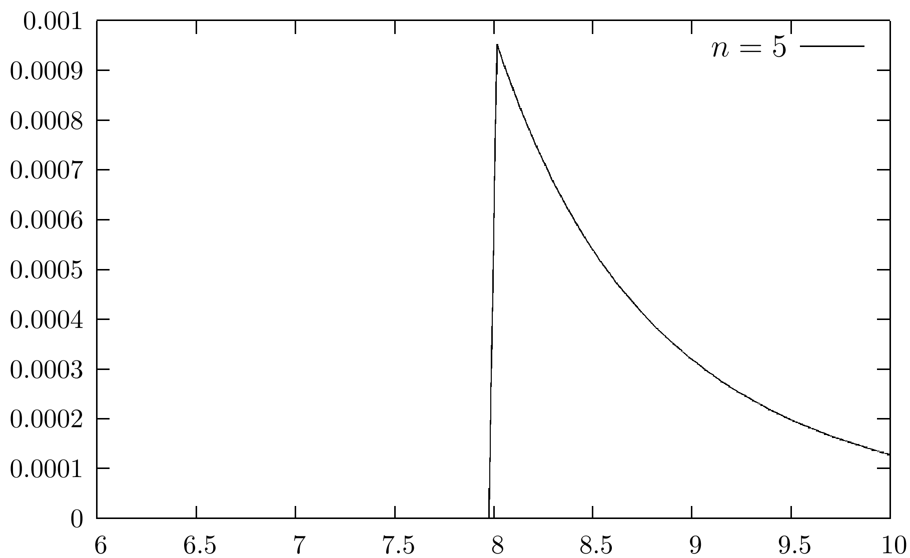
[6] Attention, propriété propre à la loi normale



Graphe de $a \mapsto L_X(x; a, b)$ *lorsque* $b = 8 > \max_i x_i$ *et* $\min_i x_i = 4$

Elle est donc dans tous les cas maximale en $\widehat{a}(x) = \min_i x_i$ (qui ne dépend pas de b).

De même pour $x \in \mathbb{R}^n$ et $a \in \mathbb{R}$ fixés, $\begin{pmatrix}]a, +\infty[& \to & \mathbb{R}^+ \\ b & \mapsto & \frac{1}{(b-a)^n} \mathbb{1}_{\min x_i \geq a} \mathbb{1}_{\max x_i \leq b} \end{pmatrix}$ est identiquement nulle si $\min_i x_i < a$, et sinon admet pour graphe



Graphe de $b \mapsto L_X(x; a, b)$ *lorsque* $a = 4 \leq \min_i x_i$ *et* $\max_i x_i = 8$

Elle est donc dans tous les cas maximale en $\widehat{b}(x) = \max_i x_i$ (qui ne dépend pas de a).
Par conséquent l'estimateur du maximum de vraisemblance de (a, b) est $x \mapsto (\min_i x_i, \max_i x_i)$.
Cependant cette statistique ne vérifie pas les propriétés habituelles de l'e.m.v., à commencer par la normalité asymptotique. La loi limite du couple $(\widehat{a}, \widehat{b})$ doit donc être retrouvée 'à la main'.

Cherchons tout d'abord pour $\alpha > 0$ la loi de $n^\alpha(\widehat{a} - a)$:

$$\begin{aligned}
\mathbb{P}(n^\alpha(\widehat{a}(X) - a) > t) &= \mathbb{P}(\min X_i > a + \frac{t}{n^\alpha}) \\
&= \mathbb{P}(X_1 > a + \frac{t}{n^\alpha})^n \\
&= \left(1 - \frac{t}{n^\alpha}(b-a)\right)^n \\
&= e^{n\ln\left(1-\frac{t}{n^\alpha}(b-a)\right)} \\
&\xrightarrow[n\to+\infty]{} \begin{cases} 0 & \text{si } \alpha < 1 \\ e^{-t/(b-a)} & \text{si } \alpha = 1 \\ +\infty & \text{si } \alpha > 1 \end{cases}
\end{aligned}$$

En particulier, $\mathbb{P}(n(\widehat{a}(X) - a) > t) \xrightarrow[n\to+\infty]{} e^{-t/(b-a)}$ et donc la fonction de survie de $n(\widehat{a} - a) > t$ tend quand $n \to +\infty$ vers la fonction de survie d'une loi exponentielle $\mathcal{E}\left(\frac{1}{b-a}\right)$. Rappelons que la convergence de la fonction de répartition en tout point t de continuité (et donc de la loi de survie) entraîne la convergence en loi. On montre de la même façon (attention au sens de l'inégalité) que

$$n(b - \widehat{b}(X)) \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{E}(1/(b-a))$$

Calculons enfin la loi jointe de $n\begin{pmatrix} \widehat{a} - a \\ b - \widehat{b} \end{pmatrix}$: pour ce faire calculons de même

$$\begin{aligned}
\mathbb{P}\left(n(\widehat{a}(X) - a) > t \wedge n(b - \widehat{b}(X)) > u\right) &= \mathbb{P}\left((\min_i X_i) > \frac{t}{n} + a \ \wedge \ (\max_i X_i) < b - \frac{u}{n}\right) \\
&= \int_{\mathbb{R}}\int_{\mathbb{R}} f_{(\min_i X_i, \max_i X_i)}(y,z) \ \mathbb{1}_{y > \frac{t}{b} + a \ \wedge \ z < b - \frac{u}{n}} \ dydz
\end{aligned}$$

La loi du couple $(\min_i X_i, \max_i X_i)$ s'obtient alors à partir de celle du n-uplet ordonné $(X_{(1)}, \ldots, X_{(n)})$, à savoir

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

(Voir TD 2, exercice 1) en intégrant successivement (d'après le théorème de projection) suivant y_{n-1}, puis $y_{n-2}, \ldots$, et ainsi de suite jusqu'à intégrer selon y_2. Il vient en tout état de cause

$$f_{(\min_i X_i, \max_i X_i)}(y,z) = \frac{n!}{(n-2)!} f(y) \left(F_X(z) - F_X(y)\right)^{n-2} f(z) \mathbb{1}_{y<z}$$

En conséquence,

$$\begin{aligned}
& \mathbb{P}\left(n(\widehat{a}(X)-a) > t \ \wedge \ n(b-\widehat{b}(X)) > u\right) \\
= \quad & \int_{a+\frac{t}{n}}^{b-\frac{u}{n}} \left(\int_{y}^{b-\frac{u}{n}} \frac{n!}{(n-2)!} \left(\frac{1}{b-a}\mathbb{1}_{[a,b]}(y)\right) \left(\frac{z-a}{b-a} - \frac{y-a}{b-a}\right)^{n-2} \left(\frac{1}{b-a}\mathbb{1}_{[a,b]}(z)\right) dz \right) dy \\
\vdots \quad & \\
= \quad & \left(1 - \frac{1}{n}\frac{y+z}{b-a}\right)^n \\
\xrightarrow[n\to+\infty]{} \quad & e^{-\frac{y+z}{b-a}} \\
= \quad & e^{-\frac{y}{b-a}} \times e^{-\frac{z}{b-a}}
\end{aligned}$$

qui est la loi jointe de deux exponentielles **indépendantes**.

Autre méthode (beaucoup plus simple) :

On peut directement déterminer la loi jointe de $n\begin{pmatrix} \widehat{a}(X)-a \\ b-\widehat{b}(X) \end{pmatrix}$:

$$\begin{aligned}
\mathbb{P}\left(n(\widehat{a}(X)-a) > t \wedge n(b-\widehat{b}(X)) < u\right) \quad = \quad & \mathbb{P}\left((\min_i X_i) > \frac{t}{n} + a \ \wedge \ (\max_i X_i) < b - \frac{u}{n}\right) \\
= \quad & \mathbb{P}\left(\forall i \in [\![1,n]\!],\, X_i \in \left]\frac{t}{n} + a \,,\, b - \frac{u}{n}\right[\right) \\
= \quad & \mathbb{P}\left(X_1 \in \left]\frac{t}{n} + a \,,\, b - \frac{u}{n}\right[\right)^n \\
= \quad & \left(\int_{\frac{t}{n}+a}^{b-\frac{u}{n}} \frac{1}{b-a} dx\right)^n \\
= \quad & \left(\int_{\frac{t}{n}+a}^{b-\frac{u}{n}} \frac{1}{b-a} dx\right)^n \\
= \quad & \left(1 - \frac{1}{n}\frac{t+u}{b-a}\right)^n \\
\xrightarrow[n\to+\infty]{} \quad & e^{-\frac{t+u}{b-a}} \\
= \quad & e^{-\frac{t}{b-a}} \times e^{-\frac{u}{b-a}}
\end{aligned}$$

se dont on déduit tout d'abord que les $\widehat{a}(X)-a$ et $b-\widehat{b}(X)$ sont indépendantes, et en outre qu'elles suivent une même loi exponentielle de paramètre $\frac{1}{b-a}$, ce qui s'écrit simplement

$$\boxed{n\begin{pmatrix} \widehat{a}(X)-a \\ b-\widehat{b}(X) \end{pmatrix} \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{E}\left(\tfrac{1}{b-a}\right)^{\otimes 2}}$$

Exercice corrigé 3

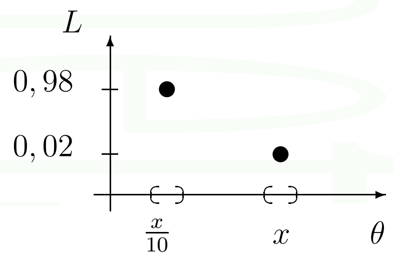
Exemple tiré de *Basu D.(1988) Stastical Information and Likelihood, Springer-Verlag, N.Y.*
Dans une urne contenant 1000 tickets, 20 sont marqués θ et 980 sont marqués 10θ.

☞ Q1 **Donner l'estimateur du maximum de vraisemblance $\widehat{\theta}$ de θ lorsque l'on tire un unique ticket de valeur X, et montrer que $P(\widehat{\theta} = \theta) = 0.98$.**

La vraisemblance s'écrit

$$L(x;\theta) = 0.98 \cdot \mathbb{1}_{10\theta}(x) \; + \; 0.02 \cdot \mathbb{1}_{\theta}(x)$$

Le graphe de $\theta \mapsto L(x;\theta)$ est alors, à x fixé



La vraisemblance est donc manifestement maximale en $\theta = \frac{x}{10}$.
Donc $\widehat{\theta} = \frac{x}{10}$ et $\mathbb{P}(\widehat{\theta}(X) = \theta) = \mathbb{P}\left(\frac{X}{10} = \theta\right) = \mathbb{P}(X = 10\theta) = \boxed{0.98}$

☞ Q2 **On renumérote les tickets marqués 10θ par $a_i\theta$ ($1 \leq i \leq 980$) où les a_i sont des réels connus, deux-à-deux distincts, et compris dans l'intervalle $[10, 10.1]$. Donner le nouvel estimateur du maximum de vraisemblance $\widetilde{\theta}$ et montrer que $P(\widetilde{\theta} = \theta) = 0.02$. Ce résultat vous semble-t-il paradoxal ?**

La vraisemblance s'écrit cette fois

$$L(x;\theta) = 0.02 \cdot \mathbb{1}_{\theta}(x) + \sum_{i=1}^{980} 0.001 \cdot \mathbb{1}_{a_i\theta}(x)$$

et est maximale en $\theta = x$.
En conséquence, $\mathbb{P}(\widetilde{\theta} = \theta) = \mathbb{P}(X = \theta) = 0.02$ On a même $\mathbb{P}(\widetilde{\theta} \geq 10\theta) = \mathbb{P}(X \geq 10\theta) = 1 - \mathbb{P}(X = \theta) = 0.98$: on est presque certain de se tromper d'un facteur 10 !
Ce résultat est paradoxal, puisqu'il incite à penser qu'une même méthode d'estimation (en l'occurence, la maximisation de la vraisemblance) peut conduire, face à deux problèmes quasi-identiques, à un estimation satisfaisante dans un cas mais très décevante dans l'autre.
Plus inquiétant encore (sauf bien-sûr pour l'étudiant de l'ENSAE, qui est vif d'esprit et averti), supposons que les a_i soient inconnus et tirés indépendamment selon une même loi aléatoire $\mathcal{L}$, d'espérance 10.5 et (pour simplifier) à support borné. Alors l'e.m.v. reprendrait à nouveau une valeur beaucoup plus satisfaisante ($\widehat{\theta} = \frac{x}{10.5}$) : en d'autres termes, l'information supplémentaire "connaissance des a_i" se traduit par une estimation beaucoup moins bonne que si on ne disposait pas de cette information !

L'explication de ce paradoxe tient à la discontinuité de la vraisemblance, et à ce que l'estimateur par maximum de vraisemblance n'est justifié qu'asymptotiquement (lorsque $n \to +\infty$) ; cet exemple donne en fait un cas extrême d'écart entre la loi à distance finie et la loi asymptotique (ici $n = 1$, mais il faut noter qu'on peut aisément construire d'autres paradoxes de ce type pour un nombre quelconque d'observations n).

Cet exemple illustre le danger des méthodes dont la justification repose uniquement sur un comportement asymptotique, alors que le nombre de données dont on dispose est toujours fini en pratique.

4 Travaux Dirigés n°4

Exercice corrigé 1

Soit f la densité de la loi exponentielle de paramètre $\frac{1}{\theta}$ translatée de α

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

☞ Q1 Donner les e.m.v. $\hat{\alpha}$ et $\hat{\theta}$ de α et θ.

On obtient : $\widehat{\alpha} = \min x_i$, $\widehat{\theta} = \overline{x} - \min x_i$.

☞ Q2 Calculer la loi (à distance finie) de $n(\hat{\alpha} - \alpha)$.

On a :

$$\mathbb{P}(n(\widehat{\alpha} - \alpha) \geq t) = \mathbb{P}(\min x_i \geq \alpha + t/n) = \left(\exp^{-t/n\theta}\right)^n = \exp(-t/\theta)$$

et donc $n(\widehat{\alpha} - \alpha) \rightsquigarrow \mathcal{E}\left(\frac{1}{\theta}\right)$.

☞ Q3 Déterminer la loi limite de $\sqrt{n}(\hat{\theta} - \theta)$.

On ne peut pas appliquer le théorème de la normalité asymptotique à $\widehat{\theta}$, car $\widehat{\theta}$ n'est que l'une des des deux composantes de l'estimateur du maximum de vraisemblance $(\widehat{\alpha}, \widehat{\theta})$ dont les composantes ne sont pas indépendantes, et qui dans ce cas particulier ne vérifie pas les conditions de régularité habituelles (notamment dérivabilité de la log-vraisemblance).

Il faut retrouver la loi limite d'une autre façon. Ici :

$$\sqrt{n}(\widehat{\theta} - \theta) = \sqrt{n}\left(\overline{x} - (\theta + \alpha)\right) + \sqrt{n}(\alpha - \widehat{\alpha})$$

Comme $\mathbb{E}(X) = \theta + \alpha$, $\mathbb{V}(X) = \theta^2$, il vient d'après le théorème central limite :

$$\sqrt{n}\left(\overline{x} - (\theta + \alpha)\right) \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}(0, \theta^2)$$

De plus,

$$\sqrt{n}(\alpha - \widehat{\alpha}) = \frac{-1}{\sqrt{n}} n(\widehat{\alpha} - \alpha) \xrightarrow{P} 0$$

puisque $n(\widehat{\alpha} - \alpha) \rightsquigarrow \mathcal{E}\left(\frac{1}{\theta}\right)$. Rappelons que :

$$X_n \xrightarrow[n\to+\infty]{\mathcal{L}} X, Y_n \xrightarrow{P} a{:}X_n + Y_n \xrightarrow[n\to+\infty]{\mathcal{L}} X + a$$

si a est une constante. Donc :

$$\sqrt{n}(\widehat{\theta} - \theta) \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}(0, \theta^2)$$

☞ Q4

> Rappeler l'expression de la loi de la statistique d'ordre $(X_{(1)}, \dots, X_{(N)})$ en fonction de f. En déduire la loi du n-uplet
>
> $$(nX_{(1)}, (n-1)(X_{(2)} - X_{(1)}), \dots, 2(X_{(n-1)} - X_{(n-2)}), X_{(n)} - X_{(n-1)})$$
>
> En déduire que $\hat{\theta}$ et $\hat{\alpha}$ sont indépendants (à distance finie).

La densité de l'échantillon ordonné $Y = (X_{(1)}, ..., X_{(n)})$ s'écrit

$$f_Y(y_1, ..., y_n) = n! f(y_1)...f(y_n) \mathbb{1}_{y_1 < ... < y_n}$$

En d'autres termes,

$$\begin{aligned} f_{X_{(1)},...,X_{(n)}(y_1,...,y_n)} &= n! \mathbb{1}_{y_1 < ... < y_n} \frac{1}{\theta^n} \left(\Pi_{i=1}^n e^{-\frac{y_i - \alpha}{\theta}} \right) \mathbb{1}_{min_i y_i \geq \alpha} \\ &= \frac{n!}{\theta^n} \mathbb{1}_{\alpha \leq y_1 < ... < y_n} e^{-\sum_{i=1}^n \frac{y_i - \alpha}{\theta}} \end{aligned}$$

Soit alors

$$\phi : \begin{pmatrix} \mathbb{R}^n & \rightarrow & \mathbb{R}^n \\ (y_1, \dots, y_n) & \mapsto & (ny_1, (n-1)(y_2 - y_1), \dots, (y_n - y_{n-1})) \end{pmatrix}$$

Alors ϕ est de clasee C^∞ et en outre

$$f_{\phi(Y)}(z_1, \dots, z_n) = \left| Jac\phi^{-1} \right| f_Y(\phi^{-1}(z_1, \dots, z_n))$$

(ceci se montre en effectuant le changement de variable $z = \phi(y)$ dans l'intégrale $\int_A f_Y(y)$, pour toute partie mesurable A de $\mathbb{R}^n$).
Or $\phi(Y)_1 = nY_1$, donc

$$\begin{aligned} \frac{\partial \phi(\cdot)_1}{\partial Y_1} &= n \\ \frac{\partial \phi(\cdot)_1}{\partial Y_i} &= 0, \quad \forall i \geq 2 \end{aligned}$$

puis $\phi(Y)_2 = (n-1)(Y_2 - Y_1)$, et donc

$$\begin{aligned} \frac{\partial \phi(\cdot)_2}{\partial Y_1} &= -(n-1) \\ \frac{\partial \phi(\cdot)_2}{\partial Y_2} &= (n-1), \quad \forall i \geq 2 \\ \frac{\partial \phi(\cdot)_2}{\partial Y_i} &= 0, \quad \forall i \geq 3 \end{aligned}$$

de sorte qu'en définitive

$$Jac(\phi) = \begin{pmatrix} n & 0 & 0 & \dots & 0 & 0 \\ -(n-1) & (n-1) & 0 & \dots & 0 & 0 \\ 0 & -(n-2) & (n-2) & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \dots & \vdots & \vdots \\ 0 & 0 & \dots & \dots & -1 & 1 \end{pmatrix}$$

En conséquence, $|Jac(\phi^{-1})| = |Jac(\phi)^{-1}| = |Jac(\phi)|^{-1} = \frac{1}{n!}$
D'autre part, $\mathbb{1}_{y_1 \geq \alpha} = \mathbb{1}_{\phi(y)_1 \geq n\alpha}$, puis pour $i \geq 2$, $\mathbb{1}_{y_i < y_{i+1}} = \mathbb{1}_{\phi(y)_i > 0}$.
Enfin, une récurrence immédiate montre que $\forall i \in [\![1, n]\!],\ \phi^{-1}(z)_i = \sum_{j=1}^{i} \frac{z_j}{n-j+1}$, de sorte qu'en définitive

$$\boxed{f_Z(z_1, ..., z_n) = \theta^{-n} \exp^{-(\sum z_i - n\alpha)/\theta} \mathbb{1}_{z_1 \geq n\alpha} \mathbb{1}_{z_2 > 0} ... \mathbb{1}_{z_n > 0}}$$

Autrement dit, les Z_i sont indépendants, avec $Z_1 \rightsquigarrow n\alpha + \mathcal{E}\left(\frac{1}{\theta}\right)$, et $Z_i \rightsquigarrow \mathcal{E}\left(\frac{1}{\theta}\right)$ pour $i \geq 2$.
Finalement, constatant que $\widehat{\alpha} = \frac{1}{n} Z_1$ et $\widehat{\theta} = \frac{1}{n} \sum_{i=2}^{n} Z_i$, on conclut que $\widehat{\alpha}$ et $\widehat{\theta}$ sont indépendants.

Exercice corrigé 2

Cet exercice présente les bases de l'estimation bayésienne.

☞ Q1 Soit $(\Omega, \mathcal{A}, \mathcal{P})$ un espace probabilisé, A un événement non-vide et $(H_1, \ldots, H_n)$ un *système complet d'hypothèses incompatibles non-vides* (c'est-à-dire une partition de Ω).
Exprimer $\mathbb{P}(H_i|A)$ en fonction des probabilités de $H_1, \ldots, H_n$, de celles de A conditionnellement à H_i et inversement, mais pas de $\mathbb{P}(A)$.

On a par définition de la probabilité conditionnelle

$$\forall i \in [\![1, n]\!],\ \mathbb{P}(A|H_i) = \frac{\mathbb{P}(A \cap H_i)}{\mathbb{P}(H_i)}$$

et donc

$$\begin{aligned}
\mathbb{P}(A) &= \mathbb{P}(A \cap \Omega) \\
&= \mathbb{P}(A \cap (\cup_{i=1}^{n} H_i)) \quad \text{car } (H_1, \ldots, H_n) \text{ est complet} \\
&= \sum_{i=1}^{n} \mathbb{P}(A \cap H_i) \quad \text{car } (H_1, \ldots, H_n) \text{ est disjoint} \\
&= \sum_{i=1}^{n} \mathbb{P}(A|H_i)\, \mathbb{P}(H_i) \quad \text{car chaque } H_i \text{ est non-vide}
\end{aligned}$$

(cette formule est dite *des probabilités totales*).
Par ailleurs on a

$$\mathbb{P}(A|H_i)\, \mathbb{P}(H_i) = \mathbb{P}(A \cap H_i) = \mathbb{P}(H_i|A)\, \mathbb{P}(A)$$

et donc

$$\begin{aligned}
\mathbb{P}(H_i|A) &= \frac{\mathbb{P}(H_i)\, \mathbb{P}(A|H_i)}{\mathbb{P}(A)} \\
&= \boxed{\frac{\mathbb{P}(H_i)\mathbb{P}(A|H_i)}{\sum_{i=1}^{n} \mathbb{P}(H_i)\mathbb{P}(A|H_i)}}
\end{aligned}$$

Cette relation exprime la probabilité d'un événement H_i, une fois connue la réalisation de A, en fonction de sa probabilité a priori.

☞ Q2

> Soit $(X_1, \ldots, X_n) \underset{iid}{\leadsto} \mathcal{L}_\theta$, de paramètre inconnu $\theta \in \Theta$.
> On suppose que θ suit une loi *a priori* de densité π_0 sur Θ.
> Donner la densité $\pi_{\cdot|x}$ de la loi *a posteriori* de θ conditionnellement à l'observation de $X = x$.

Dans le cas où $\Theta = \{\theta_i \,/\, i \in \mathbb{N}\}$ est dénombrable il suffit d'appliquer le résultat précédent à $A = \{(x_1, \ldots, x_n)\}$ et $\forall i \in \mathbb{N}$, $H_i = \{\theta_i\}$.
Dans le cas général le calcul est similaire : la formule des probabilités totales s'écrit pour toute $\mathcal{A} \subset \mathbb{R}^n$ mesurable

$$\mathbb{P}(X \in \mathcal{A}) = \int_\Theta \mathbb{P}\left(X \in \mathcal{A} | \theta = \theta_0\right) \mathbb{P}\left(\theta = \theta_0\right) d\theta_0$$

et par suite pour toute $\mathcal{B} \subset \Theta$ mesurable, à supposer que $\mathbb{P}(X \in \mathcal{A}) \neq 0$

$$\mathbb{P}\left(\theta \in \mathcal{B} | X \in \mathcal{A}\right) = \frac{\mathbb{P}\left(\theta \in \mathcal{B}\right) \mathbb{P}\left(X \in \mathcal{A} | \theta \in \mathcal{B}\right)}{\mathbb{P}\left(X \in \mathcal{A}\right)}$$

et donc

$$\boxed{\pi_{\cdot|x}\left(\theta | X = x\right) = \frac{f_{\mathcal{L}_\theta^{\otimes n}}(x_1, \ldots, x_n)\, \pi_0(\theta)}{\int_\Theta f_{\mathcal{L}_\theta^{\otimes n}}(x_1, \ldots, x_n)\, \pi_0(\theta) d\theta}}$$

☞ Q3

> Définissons pour tout estimateur $\widehat{\theta}(x)$ de θ la *fonction de risque quadratique*
>
> $$R_\nu\left(\widehat{\theta}\right) = \int_\Theta \mathbb{E}_\theta \left(\widehat{\theta}(X) - \theta\right)^2 d\nu(\theta))$$
>
> où ν désigne une loi quelconque sur Θ (par exemple $d\nu = \pi_0 d\lambda$).
> On appelle *estimateur bayésien* de θ l'estimateur $\widehat{\theta}$ qui minimise le risque associé.
> Montrer que l'estimateur bayésien de θ associé au risque R_{π_0} est
>
> $$\widehat{\theta}(x) = \mathbb{E}_{\pi_{\cdot|x}}\left(\theta\right) = \int_\Theta \theta \pi_{\cdot|x}(\theta) d\theta$$

Attention aux notations : $\mathbb{E}_\theta\left(\ldots \widehat{\theta}(X) \ldots \theta \ldots\right)$ désigne l'espérance sur la variable X, espérance qui dépend de θ, et non pas une espérance sur la variable θ. En outre l'énoncé suppose ici que Θ est de dimension 1 (sans quoi il faudrait noter $\mathbb{E}_\theta \left\|\widehat{\theta}(x) - \theta\right\|^2$).
Soit $\mathcal{X} = \mathbb{R}^n$.
<u>*Première méthode :*</u>
On cherche l'élément $T : \mathcal{X} \to \Theta$ (qui est une *fonction*) qui minimise

$$\begin{aligned} R_{\pi_0}(T) &= \int_\Theta \int_\mathcal{X} \left(T(x) - \theta\right)^2 f_{\mathcal{L}_\theta}(x) \pi_0(\theta) dx d\theta \\ &= \int_\mathcal{X} \left(\int_\Theta \left(T(x) - \theta\right)^2 f_{\mathcal{L}_\theta}(x) \pi_0(\theta) d\theta\right) dx \quad \text{d'après Fubini} \end{aligned}$$

Or la fonction

$$\phi : \begin{pmatrix} \mathcal{X}^2 & \rightarrow & \Theta \\ (x,t) & \mapsto & \int_\Theta (t-\theta)^2 f_{\mathcal{L}_\theta}(x)\pi_0(\theta)d\theta \end{pmatrix}$$

est positive, donc il suffit de minimiser point-par-point :

$$\widehat{\theta} = argmin_{T:\mathcal{X}\rightarrow\Theta} \int_{\mathcal{X}} \phi\left(T(x)\right) dx = \begin{pmatrix} \mathcal{X} & \rightarrow & \Theta \\ x & \mapsto & argmin_{t\in\Theta}\, \phi(x,t) \end{pmatrix}$$

Soit donc $x \in \mathcal{X}$ **fixé**, et calculons le nombre $\widehat{\theta}(x) = argmin_{t\in\Theta}\, \phi(x,t)$. On a

$$\begin{aligned} \phi(x,t) &= \int_\Theta (t-\theta)^2 f_{\mathcal{L}_\theta}(x)\phi_0(\theta)d\theta \\ &= \underbrace{\left(\int_\Theta f_{\mathcal{L}_\theta}(x)\phi_0(\theta)d\theta\right)}_{a>0} t^2 + \underbrace{\left(-2\int_\Theta \theta f_{\mathcal{L}_\theta}(x)\phi_0(\theta)d\theta\right)}_{b} t + \underbrace{\left(\int_\Theta \theta^2 f_{\mathcal{L}_\theta}(x)\phi_0(\theta)d\theta\right)}_{c} \end{aligned}$$

Donc $\phi(x,t) = at^2 + bt + c$ est une parabole en t (à x fixé), et est donc extrêmale en $-\frac{b}{2a}$; comme $a > 0$ elle est *minimale* en $-\frac{b}{2a}$ et donc

$$\begin{aligned} argmin_{t\in\Theta}\, \phi(x,t) &= -\frac{b}{2a} \\ &= \frac{\int_\Theta \theta f_{\mathcal{L}_\theta}(x)\phi_0(\theta)d\theta}{\int_\Theta f_{\mathcal{L}_\theta}(x)\phi_0(\theta)d\theta} \\ &= \int_\Theta \theta \frac{f_{\mathcal{L}_\theta}(x)\phi_0(\theta)}{\int_\Theta f_{\mathcal{L}_\theta}(x)\phi_0(\theta)d\theta} d\theta \end{aligned}$$

Par conséquent,

$$\boxed{\forall x \in \mathcal{X},\ \widehat{\theta}(x) = \int_\Theta \theta \pi_{\cdot|x}(\theta)d\theta}$$

☞ Q4 | Donner l'estimateur bayésien de $\theta \in [a,b]$ lorsque $\pi_0 = \mathcal{U}_{[a,b]}$ en fonction de la densité de la loi $\mathcal{L}_\theta$ de X.
Expliciter cet estimateur lorsque $\mathcal{L}_\theta$ est la loi exponentielle $\mathcal{E}(\theta)$ de paramètre inconnu $\theta > 0$.

Dans le cas d'une loi a priori uniforme on a simplement

$$\begin{aligned} \widehat{\theta}(x) &= \int_{[a,b]} \theta \pi_{\cdot|x}(\theta)d\theta \\ &= \int_{[a,b]} \theta \frac{f_{\mathcal{L}_\theta}(x_1,\ldots,x_n)\frac{1}{b-a}}{\int_\Theta f_{\mathcal{L}_\theta}(x_1,\ldots,x_n)\frac{1}{b-a}d\theta} d\theta \\ &= \frac{\int_{[a,b]} \theta f_{\mathcal{L}_\theta}(x)d\theta}{\int_{[a,b]} f_{\mathcal{L}_\theta}(x)d\theta} \end{aligned}$$

Dans le cas où $\mathcal{L}_\theta = \mathcal{E}(\theta)$ la densité de l'échantillon indépendant $(X_1, \ldots, X_n)$ est

$$\begin{aligned} f_{\mathcal{E}(\theta)^{\otimes n}}(X_1 = x_1, \ldots, X_n = x_n) &= \Pi_{i=1}^n \left(\theta e^{-\theta x_i} \mathbb{1}_{x_i \geq 0}\right) \\ &= \theta^n e^{-\theta(x_1 + \cdots + x_n)} \mathbb{1}_{\min_i x_i \geq 0} \end{aligned}$$

et donc pour $(x_1, \ldots, x_n) \in (\mathbb{R}^+)^n$ il vient

$$\begin{aligned} \widehat{\theta}(x) &= \frac{\int_{[a,b]} \theta^{n+1} e^{-\theta(x_1+\cdots+x_n)} d\theta}{\int_{[a,b]} \theta^n e^{-\theta(x_1+\cdots+x_n)} d\theta} \\ &= \frac{\left[\theta^{n+1} \frac{e^{-\theta \sum_{i=1}^n x_i}}{-\sum_{i=1}^n x_i}\right]_a^b - \int_{[a,b]} (n+1)\theta^n \frac{e^{-\theta \sum_{i=1}^n x_i}}{-\sum_{i=1}^n x_i} d\theta}{\int_{[a,b]} \theta^n e^{-\theta(x_1+\cdots+x_n)} d\theta} \quad \text{par parties} \\ &= \frac{1}{\sum_{i=1}^n x_i} \left(n + 1 - \frac{b^{n+1} e^{-b \sum_{i=1}^n x_i} - a^{n+1} e^{-a \sum_{i=1}^n x_i}}{\int_{[a,b]} \theta^n e^{-\theta(x_1+\cdots+x_n)} d\theta}\right) \end{aligned}$$

Ainsi $\boxed{\widehat{\theta}(x) = \frac{1}{\sum_{i=1}^n x_i} \left(n + 1 - \frac{\overline{\theta}^{n+1} e^{-\overline{\theta} \sum_{i=1}^n x_i}}{\int_{[0,\overline{\theta}]} \theta^n e^{-\theta(x_1+\cdots+x_n)} d\theta}\right)}$ lorsque $\Theta = [0, \overline{\theta}]$ par exemple.

L'estimateur du maximum de vraisemblance de θ, par comparaison, est $\widehat{\theta}^{emv}(x) = \frac{n}{\sum_{i=1}^n x_i}$.

Exercice corrigé 3

Un des premiers exemples d'utilisation de la Statistique Bayésienne remonte à Laplace, en 1786. Celui-ci décida de répondre à la question suivante : au regard du nombre observé n_g de naissances masculines parmi n naissances à Paris, peut-on dire si la probabilité p qu'un enfant qui naisse soit un garçon est supérieure à $\frac{1}{2}$?

☞ Q1 Laplace munit le paramètre p d'une loi a priori uniforme sur l'intervalle $[0, 1]$. Ce choix vous semble-t-il naturel ?

Laplace considère en fait de ne pas détenir d'information sur p autre que celle apportée par les observations : en d'autres termes, il n'a pas d' "a priori" sur p, ce qu'il modélise par une loi a priori uniforme. Cela revient intuitivement à donner une "probabilité égale" à toutes les valeurs possibles de p. Mais cette intuition est en fait trompeuse : si on change la paramétrisation du modèle (en remplaçant p par $q = p^2$ par exemple), on se rend compte alors que la loi a priori sur q n'est plus uniforme ! Cette reparamétrisation ne nous a pourtant apporté aucune information supplémentaire. Ce paradoxe illustre la difficulté de bien choisir une loi a priori, surtout dans un cadre non-informatif (c'est à dire lorsqu'aucune information n'est disponible a priori). Cependant, nous verrons en 4) qu'il est assez simple de se restreindre au moins à un choix limité de lois a priori raisonnables.

☞ Q2

Exprimer alors la loi a posteriori de p, puis exprimer la probabilité $\mathbb{P}\left(p > \frac{1}{2}|N_g = n_g\right)$ sous forme d'un rapport d'intégrales.
Remarque : Laplace, obtint, pour $n = 493472$ et $n_g = 251527$, une probabilité $\mathbb{P}\left(p > \frac{1}{2}|N_g = n_g\right) \simeq 1 - 1.15.10^{-42}$ et en conclut que p était très vraisemblablement plus grand que $\frac{1}{2}$.

La densité de la loi a posteriori $\pi(\theta|x_1, ..., x_n)$ du paramètre θ s'écrit sous la forme (voir exercice 2) :

$$\pi(\theta|x_1, ..., x_n) = \frac{\pi(\theta)\mathbb{P}(x_1, ..., x_n|\theta)}{\int_\Theta \pi(\theta)\mathbb{P}(x_1, ..., x_n|\theta)d\theta}$$

Il vient

$$\begin{aligned}\pi(p|N_g = n_g) &= \frac{\pi(p)\mathbb{P}(N_g = n_g|p)}{\int_0^1 \pi(p)\mathbb{P}(N_g = n_g|p)dp} \\ &= \frac{p^{n_g}(1-p)^{n-n_g}}{\int_0^1 p^{n_g}(1-p)^{n-n_g}dp}\end{aligned}$$

Par conséquent

$$\begin{aligned}\mathbb{P}(p > \frac{1}{2}|N_g = n_g) &= \int_{[0,1]} \frac{\pi(p)\mathbb{P}(N_g = n_g|p)}{\int_0^1 \pi(p)\mathbb{P}(N_g = n_g|p)dp}\mathbb{1}_{p>\frac{1}{2}}dp \\ &= \frac{\int_{\frac{1}{2}}^1 p^{n_g}(1-p)^{n-n_g}dp}{\int_0^1 p^{n_g}(1-p)^{n-n_g}dp}\end{aligned}$$

☞ Q3

Donner l'espérance et la variance de la loi a posteriori en fonction de la vrai valeur p_0 du paramètre p. Quelles sont leurs limites lorsque $n \to +\infty$?
Rappel : la loi Beta $\mathcal{B}(\alpha, \beta)$ de paramètres $\alpha > 0$, $\beta > 0$ admet pour densité :

$$f_{\alpha,\beta}(x) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)}\mathbb{1}_{[0,1]}(x)$$

où $B(\alpha, \beta) = \int_0^1 x^{\alpha-1}(1-x)^{\beta-1}dx$.
Son espérance est $\mathbb{E}(X) = \frac{\alpha}{\alpha+\beta}$ et sa variance $\mathbb{V}(X) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$.

Notons π_p la densité de la loi a posteriori : π_p est la densité de la loi $\mathcal{B}(n_g + 1, n - n_g + 1)$. Son espérance et sa variance sont donc :

$$\begin{aligned}\mathbb{E}_{\pi_p}[p] &= \frac{n_g + 1}{n + 2} \\ \mathbb{V}_{\pi_p}[p] &= \frac{(n_g + 1)(n - n_g + 1)}{(n + 2)^2(n + 3)}\end{aligned}$$

Or $n_g = \sum_{i=1}^n g_i$ où $g_i = \mathbb{1}_{\text{l'enfant } i \text{ est un garçon}} \underset{iid}{\rightsquigarrow} \mathcal{B}(1, p_0)$.

Donc en vertu de la loi forte des grands nombres $\frac{n_g}{n} = \frac{1}{n}\sum_{i=1}^{n} g_i \xrightarrow[n\to+\infty]{p.s.} p_0$ et donc

$$\begin{array}{ccc} \mathbb{E}_{\pi_p}[p] & \xrightarrow[n\to+\infty]{p.s.} & p_0 \\ \mathbb{V}_{\pi_p}[p] & \underset{n\to+\infty}{\overset{p.s.}{\rightsquigarrow}} & \dfrac{p_0(1-p_0)}{n} \end{array}$$

En d'autres termes, lorsque $n \to +\infty$ la loi a posteriori se "rétrécie" (sa variance tend vers 0) et se concentre autour de la vraie valeur p_0. Ce résultat est intuitif : plus on dispose d'observations, plus faible est l'incertidude sur la valeur du paramètre p.

☞ Q4 | Ces limites sont-elles modifiées si on choisit pour loi a priori sur p une loi $\mathcal{B}(\alpha,\beta)$ quelconque ? Pourquoi est-il judicieux malgré tout de se restreindre au moins à la classe des lois a priori telles que $\alpha=\beta$?

Si on change la loi a priori uniforme (qui correspond d'ailleurs à une loi $\mathcal{B}(1,1)$) pour une loi a priori plus générale $\mathcal{B}(\alpha,\beta)$, on obtient une loi a posteriori :

$$\begin{aligned} \pi(p|N_g &= n_g) \propto p^{\alpha-1}(1-p)^{\beta-1}p^{n_g}(1-p)^{n-n_g}1_{[0,1]}(p) \\ &\propto p^{\alpha+n_g-1}(1-p)^{\beta+n-n_g-1}1_{[0,1]}(p) \end{aligned}$$

(où $\propto$ signifie "proportionnel à")

La loi a posteriori est donc une loi $\mathcal{B}(n_g+\alpha, n-n_g+\beta)$. On vérifie alors facilement que $\mathbb{E}_{\pi_p}[p]$ et $\mathbb{V}_{\pi_p}[p]$ conservent le même comportement asymptotique que dans le cas précédent. En effet, plus grand est le nombre d'observations, plus le poids de l'a priori est faible, et ce poids devient nul asymptotiquement.

Cependant, la Statistique Bayésienne ne repose pas sur des justifications asymptotiques. A distance finie (c'est à dire pour un nombre fini d'observations), une loi a priori mal choisie peut sérieusement déformer la loi a posteriori, et mener à une inférence complètement érronée. Ainsi, au vu du probème posé (situer p dans l'un des deux intervalles $[0,\frac{1}{2}]$ ou $[\frac{1}{2},1]$), on n'a aucune raison de favoriser *a priori* l'une de ces deux régions, seules les observations pouvant nous permettre de les départager. Il est donc naturel de se restreindre au moins aux lois a priori symétriques ($\alpha=\beta$). Parmi celles-ci, on peut aussi écarter les lois de très faible variance (ie telles que α prend de fortes valeurs, ex : $\mathcal{B}(10,10)$), qui privilégient inutilement un voisinage trop restreint du point $\frac{1}{2}$.

Finalement, la loi a priori uniforme proposée par Laplace semble donc un choix raisonnable parmi d'autres, au moins pour le problème posé. Il est en fait possible de proposer une loi plus satisfaisante (dite loi a priori de Jeffreys, ici $\mathcal{B}(1/2,1/2)$) qui permette de s'affranchir du paradoxe de la reparamétrisation présenté en 1).

5 Travaux Dirigés n°5

Exercice corrigé 1

On souhaite évaluer et analyser le phénomène du chômage. Pour cela, on dispose de n observations sur les durées $y_i, 1 \leq i \leq n$, pendant lesquelles des individus sont restés sans emploi. On suppose dans la suite que les variables aléatoires correspondantes $(Y_i)_{i \in [\![1,n]\!]}$ sont i.i.d. et suivent la loi de Weibull de paramètres a et b. On rappelle que cette loi est continue sur $\mathbb{R}^+$ et admet la fonction de répartition pour $y > 0$

$$F(y; a, b) = 1 - \exp\left(-ay^b\right)$$

On définit la fonction de survie par

$$S(y) = 1 - F(y)$$

et la fonction de hasard par $h(y) = \frac{f(y)}{S(y)}$.

■ Partie 1 **Généralités**

☞ Q1 Donner l'expression de la fonction de hasard du modèle.

Par définition on a pour $y \in \mathbb{R}^+$

$$\begin{aligned} f(y; a, b) &= \frac{\partial F}{\partial \cdot}(y, a, b) \\ &= aby^{b-1}e^{-ay^b} \end{aligned}$$

Il en découle que

$$h(y; a, b) = aby^{b-1}\mathbb{1}_{y \geq 0}$$

☞ Q2 Quelle est en terme de chômage l'interprétation de la fonction de hasard ?
Expliquer alors pourquoi il est important de considérer le cas particulier où cette fonction est constante.
Pour quelles valeurs des paramètres, la fonction de hasard est-elle constante ?
Quelles sont alors les lois des durées de chômage ?

La fonction de hasard $h(y)$ s'écrit comme le rapport $\frac{f(y)}{S(y)}$. Or

$$\begin{aligned} f(y) &= \frac{\partial F(\cdot; a, b)}{\partial y}(y) \\ &= \lim_{dy \to 0^+} \frac{\mathbb{P}\left(Y \in [y, y + dy[\right)}{dy} \end{aligned}$$

et par ailleurs $S(y) = 1 - F(y) = \mathbb{P}(Y \geq y)$.

Autrement dit

$$\begin{aligned} h(y) &= \lim_{dy \to 0^+} \frac{1}{dy} \frac{\mathbb{P}\left(Y \in [y, y+dy[\right)}{\mathbb{P}(Y \geq y)} \\ &= \lim_{dy \to 0^+} \frac{1}{dy} \frac{\mathbb{P}\left(Y \in [y, y+dy[\wedge Y \geq y\right)}{\mathbb{P}(Y \geq y)} \\ &= \lim_{dy \to 0^+} \frac{1}{dy} \mathbb{P}\left(NON(Y \geq y+dy) \mid Y \geq y\right) \end{aligned}$$

En d'autres termes $h(y)dy$ mesure la probabilité, pour un individu encore au chômage à la date y, de **sortir du chômage** peu après y (au sens “après $y+dy$”, ceci pour dy “voisin” de 0).

Supposer que h est constante revient à supposer que la probabilité de sortir d'une période de chômage de durée dy ne dépend pas de sa date y, c'est-à-dire qu'il n'y a pas d'*effet mémoire*.

Sachant que

$$\begin{aligned} h(y) &= \frac{aby^{b-1}e^{-aby^b}}{e^{-ay^b}} \\ &= aby^{b-1} \end{aligned}$$

cette condition est vérifiée **si, et seulement si** $b = 1$.

Lorsque cette condition est vérifiée, on constate que la loi de durée du chômage s'écrit simplement pour $y \geq 0$

$$f(y; a, 1) = ae^{-ay}$$

dont on remarque qu'il s'agit d'une loi exponentielle de paramètre a.

☞ Q3 | Etudier l'évolution de la fonction de hasard en fonction de a, puis en fonction de b. |

Etudions les variations de $h(y; \cdot, \cdot)$ en fonctionde a et b à $y \geq 0$ donné.

Il vient

– Selon a :

$$\begin{aligned} \frac{\partial h}{\partial a}(y, a, b) &= by^{b-1} \\ &> 0 \quad \text{car } b > 0 \end{aligned}$$

Ainsi la situation du chômage s'améliore (au sens où la probabilité de sortie s'accroît) lorsque a s'accroît.

– Selon b :

$$\begin{aligned} \frac{\partial h}{\partial b}(y, a, b) &= ay^{b-1} + ab\ln(y)y^{b-1} \\ &= ay^{b-1}\left(1 + b\ln(y)\right) \end{aligned}$$

En conséquence

– Si $y \in]0, e^{-\frac{1}{b}}[$, soit si $b > -\frac{1}{\ln(y)}$, la situation du chômage *se détériore* lorsque b croît.

– Si $y \in]e^{-\frac{1}{b}}, +\infty[$, soit si $b < -\frac{1}{\ln(y)}$, la situation du chômage *s'améliore* lorsque b croît.

■ Partie 2 **Estimation contrainte**

On suppose dans cette partie $b = 1$. Le modèle est alors uniquement paramétré par a.

☞ Q1 Le modèle est-il exponentiel ? Si oui, expliciter une statistique exhaustive.

La vraisemblance du modèle est donnée pour $(y_1, \ldots, y_n) \in \mathbb{R}^n$ par

$$\begin{aligned} L(y; a, 1) &= \Pi_{i=1}^n \left(a e^{-a y_i} \mathbb{1}_{y_i \geq 0} \right) \\ &= \left(a^n \mathbb{1}_{\min_i y_i \geq 0} \right) \cdot e^{-a \left(\sum_{i=1}^n \cdot y_i \right)} \end{aligned}$$

On constate donc que le modèle est exponentiel, de statistique exhaustive $T(Y) = \sum_{i=1}^n Y_i$.

☞ Q2 Déterminer le vecteur du score et vérifier directement qu'il est centré.

La log-vraisemblance est alors pour $(y_1, \ldots, y_n) \in \mathbb{R}^{+n}$

$$\ln L(y; a, 1) = n \ln a - a \sum_{i=1}^n y_i$$

Le score s'écrit donc

$$\begin{aligned} S_a(y; a, 1) &= \frac{\partial \ln L}{\partial a}(y; a, 1) \\ &= \frac{n}{a} - \sum_{i=1}^n y_i \end{aligned}$$

Or

$$\begin{aligned} \mathbb{E}(y_i) &= \int_{\mathbb{R}} y_i \cdot a e^{-a y_i} \mathbb{1}_{y_i \geq 0} d y_i \\ &= -\int_0^{+\infty} (-au) e^{-au} du \\ &= +\frac{1}{a} \left[e^{-au} \right]_0^{+\infty} \\ &= \frac{1}{a} \end{aligned}$$

Il s'ensuit que

$$\mathbb{E}(S_a(y; a, 1)) = \frac{n}{a} - \sum_{i=1}^n \mathbb{E}(y_i) = 0$$

de sorte que le vecteur score est bien centré (ce qui est une propriété générale, voir TD 2, exercice 2)

☞ Q3 Quel est l'estimateur du maximum de vraisemblance $\hat{a}_0$ de a ?
Est-il sans biais, y a-t-il surestimation ou sous-estimation systématique ?

De la condition nécessaire $\frac{\partial \ln L}{\partial a}(\widehat{a_0}) = 0$ on tire $\widehat{a_0} = \frac{1}{\bar{y}}$; réciproquement $\widehat{a_0}$ maximise bien la log-vraisemblance car celle-ci est concave.

Par ailleurs, on a

$$\begin{aligned} \mathbb{E}\left(\widehat{a_0}\right) &= \mathbb{E}\left(\frac{1}{\bar{y}}\right) \\ &> \frac{1}{\mathbb{E}\left(\bar{y}\right)} \quad \text{(d'après l'inégalité de Jensen)} \\ &= a \end{aligned}$$

Autrement dit, $\widehat{a_0}$ est biaisé et surestime systématiquement la vraie valeur a.

On pourrait même calculer le biais exact, en s'appuyant sur le fait que $(Y_1 + \ldots + Y_n) \rightsquigarrow \Gamma(n, a)$ et en calculant explicitement $\int_{\mathbb{R}} \frac{n}{s} f_{\Gamma(a)}(s) ds = \frac{n}{n-1} a$; on en déduirait que $\frac{n-1}{n}\widehat{a_0}$ est un estimateur sans biais de a.

☞ Q4 Déterminer la variance asymptotique de cet estimateur $\hat{a}_0$

On a

$$I_n(a) = -\mathbb{E}\left(\frac{\partial^2 \ln L}{\partial a^2} = -\frac{n}{a^2}\right)$$

et donc

$$\boxed{\mathbb{V}_{as}\left(\widehat{a_0}\right) = \frac{a^2}{n}}$$

(Attention, c'est bien $I_n(a)^{-1}$ et non pas $I_1(a)^{-1}$ car on étudie la loi asymptotique de $(\widehat{a_0} - \mathbb{E}\left(\widehat{a_0}\right))$ et non pas celle de $\sqrt{n}(\widehat{a_0} - \mathbb{E}\left(\widehat{a_0}\right)))$

■ Partie 3 **Estimation non contrainte**

On considère maintenant le cas où a et b peuvent a priori prendre toutes valeurs positives.

☞ Q1 Le modèle est-il exponentiel avec une statistique exhaustive dont la taille est indépendante du nombre n d'observations ?
Si oui, expliciter une telle statistique.

La vraisemblance du modèle est donnée pour $(y_1, \ldots, y_n) \in \mathbb{R}^n$ par

$$\begin{aligned} L\left(y; a, b\right) &= \Pi_{i=1}^n \left(ab{y_i}^{b-1} e^{-a{y_i}^b} \mathbb{1}_{y_i \geq 0}\right) \\ &= \left((ab)^n \mathbb{1}_{\min_i y_i \geq 0}\right) \cdot \left(\Pi_{i=1}^n {y_i}^{b-1}\right) \cdot e^{-\sum_{i=1}^n \left(a \cdot {y_i}^b\right)} \end{aligned}$$

On constate donc que le modèle n'est **pas** exponentiel, car on ne peut dissocier les termes en y_i du paramètre (a, b) là où ils interviennent (à savoir $\sum_{i=1}^n {y_i}^b$). Donc en vertu du théorème de factorisation $(Y_1, \ldots, Y_n)$ est exhaustive et minimale, et il n'existe donc pas de statistique exhaustive comportant moins de n composantes.

☞ Q2 Ecrire les équations de vraisemblance. Sont-elles résolubles sous forme analytique ?

La log-vraisemblance s'écrit pour $(y_1, \ldots, y_n) \in {\mathbb{R}^+}^n$

$$\ln L\,(y; a, b) = n \ln a + n \ln b + (b-1) \sum_{i=1}^{n} \ln y_i - a \sum_{i=1}^{n} {y_i}^b$$

On a

$$\begin{cases} \frac{\partial \ln L}{\partial a}(y; a, b) & = & \frac{n}{a} - \sum_{i=1}^{n} {y_i}^b \\ \frac{\partial \ln L}{\partial b}(y; a, b) & = & \frac{n}{b} + \sum_{i=1}^{n} \ln y_i - a \sum_{i=1}^{n} \ln y_i \; {y_i}^b \end{cases}$$

et donc l'estimateur $(\widehat{a}, \widehat{b})$ de (a, b) vérifie

$$\begin{cases} \widehat{a} & = & \frac{n}{\sum_{i=1}^{n} {y_i}^{\widehat{b}}} \\ \widehat{a} \sum_{i=1}^{n} \ln y_i \; {y_i}^{\widehat{b}} - \frac{n}{\widehat{b}} & = & \sum_{i=1}^{n} \ln y_i \end{cases}$$

On constate que $\widehat{b}$, et par suite $\widehat{a}$, n'est pas exprimable sous forme analytique. Pour autant l'étude de la fonction $\left(\begin{array}{ccc} \mathbb{R}^+ & \rightarrow & \mathbb{R} \\ b & \mapsto & \frac{1}{b} + \alpha - \frac{\sum_{i=1}^{n} \beta_i y_i^b}{\sum_{i=1}^{n} y_i^b} \end{array} \right)$ assure de l'existence et de l'unicité de $\widehat{b}$, et donc de $\widehat{a}$.

☞ Q3 Donner la forme de la variance asymptotique de l'estimateur du maximum vraisemblance $\begin{pmatrix} \hat{a} \\ \hat{b} \end{pmatrix}$ du paramètre $\begin{pmatrix} a \\ b \end{pmatrix}$.

On a successivement

$$\begin{aligned} \frac{\partial^2 \ln L}{\partial a^2}(y; a, b) &= -\frac{n}{a^2} \\ \frac{\partial^2 \ln L}{\partial a \partial b}(y; a, b) &= -\sum_{i=1}^{n} \left(\ln y_i {y_i}^b \right) \\ \frac{\partial^2 \ln L}{\partial b^2}(y; a, b) &= -\frac{n}{b^2} - a \sum_{i=1}^{n} \left((\ln y_i)^2 {y_i}^b \right) \end{aligned}$$

En conséquence

$$\mathbb{V}_{as}\left(\begin{pmatrix} \widehat{a} \\ \widehat{b} \end{pmatrix} \right) = \begin{pmatrix} \frac{n}{a^2} & n\mathbb{E}\left({y_1}^b \ln y_1 \right) \\ n\mathbb{E}\left({y_1}^b \ln y_1 \right) & \frac{n}{b^2} + na\mathbb{E}\left((\ln y_i)^2 {y_i}^b \right) \end{pmatrix}^{-1}$$

Il reste alors à exprimer $\mathbb{E}\left(y_1^b \ln(y_1)^i \right)$ pour $i \in \{1, 2\}$; ce calcul peut être fait au moyen des intégrales eulériennes.

Remarquons pour ce faire que si Z suit une loi de Weibull de paramètre (α, β), de densité $f_{\alpha,\beta}(x) = \alpha\, \theta\, x^{\alpha-1}\, \mathrm{e}^{-\theta x^\alpha} \mathbb{1}_{x>0}$, alors pour $p \in \mathbb{N}^*$ on a

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

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

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

Or (TD 2, exercice 2)

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

Ceci permet alors de conclure en posant $\theta = a$ et $\alpha = b$ que[7]

$$\boxed{\mathbb{V}_{as}\left(\begin{pmatrix} \widehat{a} \\ \widehat{b} \end{pmatrix}\right) = \begin{pmatrix} \frac{n}{a^2} & n\frac{1-\gamma-\ln a}{ab} \\ n\frac{1-\gamma-\ln a}{ab} & \frac{n}{b^2}\left(1+\frac{\pi^2}{6} - 2\gamma + \gamma^2 - 2(1-\gamma)\ln a + (\ln a)^2\right) \end{pmatrix}^{-1}}$$

☞ Q4 En déduire la variance asymptotique de $\hat{a}$ lorsque $b=1$.
Comparer alors les estimateurs $\hat{a}$ et $\hat{a}_0$ lorsque $b=1$.
Quelle conclusion en tirer ?

Remarquons que pour toute matrice A, $A^{-1} = \frac{1}{|A|}Com(A)'$ où $Com(A)$ désigne la comatrice de A ; par suite

$$I_n(a,1)^{-1} = \frac{1}{|I_n(a,1)|}\begin{pmatrix} n + na\mathbb{E}\left((\ln y_1)^2 y_1\right) & -n\mathbb{E}\left(y_1 \ln y_1\right) \\ -n\mathbb{E}\left(y_1 \ln y_1\right) & \frac{n}{a^2} \end{pmatrix}$$

Par ailleurs,

$$|I_n(a,1)| = \frac{n}{a^2}\left(n + na\mathbb{E}\left((\ln y_1)^2 y_1\right)\right) - \left(n\mathbb{E}\left(y_1\ln y_1\right)\right)^2$$

En conséquence,

$$\begin{aligned}
\mathbb{V}_{as}\left(\widehat{a}\right) &= \frac{n + na\mathbb{E}\left((\ln y_1)^2 y_1\right)}{|I_n(a,1)|} \\
&= \frac{n + na\mathbb{E}\left((\ln y_1)^2 y_1\right)}{\frac{n}{a^2}\left(n + na\mathbb{E}\left((\ln y_1)^2 y_1\right)\right) - \left(n\mathbb{E}\left(y_1\ln y_1\right)\right)^2} \\
&= \frac{a^2}{n}\left(1 + \frac{\frac{a^2}{n}\left(n\mathbb{E}\left(y_1\ln y_1\right)\right)^2}{\left(n + na\mathbb{E}\left((\ln y_1)^2 y_1\right)\right) - \frac{a^2}{n}\left(n\mathbb{E}\left(y_1\ln y_1\right)\right)^2}\right) \\
&\geq \frac{a^2}{n} \\
&= \mathbb{V}_{as}\left(\widehat{a_0}\right)
\end{aligned}$$

[7]Bien que la tentation soit forte, nous résisterons ici au plaisir d'inverser cette matrice.

Ainsi estimer sous contrainte permet d'obtenir un estimateur *plus efficace* qu'estimer sans contrainte.

☞ Q5 Quelle démarche pourrait-on proposer pour étudier la distribution de l'estimateur $\begin{pmatrix} \hat{a} \\ \hat{b} \end{pmatrix}$ lorsque l'échantillon est de petite taille, par exemple $n = 10$?

Lorsque l'échantillon est de petit taille, une simulation permettrait d'étudier la distribution de $\begin{pmatrix} \widehat{a} \\ \widehat{b} \end{pmatrix}$, de la façon suivante :

- Observant $(y_1, \ldots, y_n)$, on en tire $\begin{pmatrix} \widehat{a^0} \\ \widehat{b^0} \end{pmatrix}$ estimateur du maximum de vraisemblance de $\begin{pmatrix} a^0 \\ b^0 \end{pmatrix}$.
- Soit alors $\mathcal{W}^0$ la loi d Weibull de paramètre $\widehat{a^0}, \widehat{b^0}$; on tire alors un échantillon indépendant de taille N selon $\mathcal{W}^0$, soit $({y^0}_1, \ldots, {y^0}_N)$.
- de cet échantillon on tire alors $\begin{pmatrix} \widehat{a^1} \\ \widehat{b^1} \end{pmatrix}$, et ainsi de suite.

Bien-entendu, à la limite lorsque $N \to +\infty$ on a $\begin{pmatrix} \widehat{a^1} \\ \widehat{b^1} \end{pmatrix} = \begin{pmatrix} \widehat{a^0} \\ \widehat{b^0} \end{pmatrix}$; il en va de même lorsque le nombre d'itération devient arbitrairement grand.
L'idée serait donc plutôt de se limiter à une ou deux itérations, mais de **répéter le tirage** pour obtenir la distribution empirique de $\begin{pmatrix} \widehat{a^1} \\ \widehat{b^1} \end{pmatrix}$.

Partie 4 Cas indépendant - non équidistribué

On considère maintenant le cas de T observations $Y_1, \ldots, Y_T$ indépendantes, de lois respectives :

$$F(y; e^{\alpha t}, 1)\ ,\ t \in [\![1, T]\!]\ ,\ \alpha \in \mathbb{R}$$

☞ Q1 Déterminer la vraisemblance du modèle $\mathcal{L}$, et vérifier qu'elle est concave en α à $(y_1, \ldots, y_T)$ fixé.
En déduire l'équation caractérisant l'estimateur du maximum de vraisemblance $\widehat{\alpha_T}$ de α.

La vraisemblance du modèle est donnée pour $(y_1, \ldots, y_n) \in \mathbb{R}^n$ par

$$\begin{aligned} L(y; \alpha) &= \Pi_{t=1}^T e^{\alpha t} e^{-e^{\alpha t} y_t} \left(\mathbb{1}_{y_t \geq 0}\right) \\ &= \mathbb{1}_{\inf_{t \in [\![1,T]\!]} y_t \geq 0} e^{\alpha \sum_{t=1}^T t} e^{-\sum_{t=1}^T e^{\alpha t} y_t} \end{aligned}$$

Une condition nécéssaire sur l'estimateur $\widehat{\alpha}$ de α est donc

$$\sum_{t=1}^{T} \left(t y_t e^{t\widehat{\alpha}} \right) = \frac{T(T+1)}{2}$$

condition dont on vérifie qu'elle est suffisante puisque la log-vraisemblance est strictement concave car de dérivée seconde

$$\frac{\partial^2 \ln L}{\partial \alpha^2}(y;\alpha) = -\sum_{t=1}^{T} t^2 y_t e^{\alpha t} < 0 \quad \forall \alpha \in \mathbb{R}^+$$

☞ Q2

On note $u_t = y_t - e^{-\hat{\alpha}_T t}$. Donner l'interprétation de u_t.

On a

$$\begin{aligned} u_t &= y_t - e^{-\widehat{\alpha} t} \\ &= y_t - \mathbb{E}_{\mid \alpha = \widehat{\alpha}}(y_t) \end{aligned}$$

Ainsi u_t représente l'*erreur de la prévision empirique* de y_t, différence entre la réalisation réelle de y_t et la meilleure prévision compte-tenu du modèle estimé.

☞ Q3

Montrer que l'équation de la vraisemblance correspond à la condition d'orthogonalité de $(u_1, \ldots, u_T)$ et de $1, \ldots, T$ pour un certain produit scalaire que l'on précisera.

On a

$$\begin{aligned} \frac{\partial \ln L}{\partial \alpha}(y,\alpha) &= \frac{T(T+1)}{2} - \sum_{t=1}^{T} t y_t e^{\alpha t} \\ &= \sum_{t=1}^{T} \left(t - t y_t e^{\alpha t}\right) \\ &= \sum_{t=1}^{T} t\left(1 - y_t e^{\alpha t}\right) \\ &= \sum_{t=1}^{T} t e^{\alpha t}\left(e^{-\alpha t} - y_t\right) \end{aligned}$$

de sorte que

$$\frac{\partial \ln L}{\partial \alpha}(y,\widehat{\alpha}) = \sum_{t=1}^{T} e^{\widehat{\alpha} t} t u_t$$

Posons donc

$$< \cdot | \cdot >_\rho : \begin{pmatrix} (\mathbb{R}^T)^2 & \to & \mathbb{R}^+ \\ ((x_t)_t, (y_t)_t) & \mapsto & \sum_{t=1}^{T} \rho^t x_t y_t \end{pmatrix}$$

Alors pour tout $\rho \in \mathbb{R}^{+*}$, $< \cdot | \cdot >_\rho$ est un produit scalaire, et en outre

$$\frac{\partial \ln L}{\partial \alpha}(y,\widehat{\alpha}) = 0 \quad \Leftrightarrow \quad < (1, \ldots, T) | u) >_{e^{\widehat{\alpha}}} = 0$$

Exercice corrigé 2

On étudie entre les dates 0 et T un groupe de n individus sans emploi à la date 0, et on cherche à modéliser les durées de chômage $(T_i)_{i\in[\![1,n]\!]}$.

En pratique, on observe les durées de chômage en mois. Plus précisément, on ne dispose pas de la variable continue T_i, mais seulement de la variable discrète T_i^* donnée par

$$T_i^* = \lceil T_i \rceil$$

Autrement dit, la variable T_i^* vaut $t+1$ si l'individu i a retrouvé du travail entre le t-ème et $(t+1)$-ième mois.

En outre entre t et $t+1$, on suppose que :
- l'individu i reçoit N_t^i offres d'emploi, où N_t^i est une suite de variables i.i.d. de loi de Poisson $\mathcal{P}(\lambda)$;
- si l'individu i est toujours au chômage à la date t, et si parmi les N_t^i offres qu'il reçoit, l'une au moins offre un salaire supérieur à une constante ξ_i, propre à l'individu (appelée salaire de réserve), alors l'individu n'est plus au chômage : $(T_i^* = t+1)$;
- les salaires des offres d'emploi sont tirés indépendamment des dates d'arrivée des offres et de leur nombre dans une loi de fonction de répartition F.

On suppose dans un premier temps pour simplifier que $T = +\infty$.

☞ Q1 $\boxed{\text{Calculer } \mathbb{P}\left(T_i^* = t+1 \,|T_i > t\right) \text{ en fonction de } F,\ \xi_i,\ \lambda\ .}$

On suppose pour simplier que jusqu'à la question **4**, $T = +\infty$. Sinon, il faudrait alourdir les expressions des densités qui suivent en adjoignant l'indicatrice $\mathbb{1}_{t_i \leq T}$ adéquate.

On a pour $t \in [\![1, T-1]\!]$

$$\begin{aligned}
\mathbb{P}(T_i^* = t+1 \,|\, T_i > t) &= \mathbb{P}(\text{L'individu } i \text{ reçoit au moins une offre intéressante entre } t \text{ et } t+1) \\
&= \mathbb{P}(N_t^i \geq 1 \ \wedge \ \text{Une des } N_t^i \text{ offres propose un salaire} \geq \xi_i) \\
&= \sum_{k=1}^{+\infty} \mathbb{P}(N_t^i = k \ \wedge \ \text{Une des } k \text{ offres propose un salaire} \geq \xi_i) \\
&= \sum_{k=1}^{+\infty} \mathbb{P}(N_t^i = k) \cdot \mathbb{P}(\text{Une des } k \text{ offres propose un salaire} \geq \xi_i) \\
&= \sum_{k=1}^{+\infty} \mathbb{P}(N_t^i = k) \cdot (1 - \mathbb{P}(\text{Toutes les } k \text{ offres proposent un salaire} < \xi_i)) \\
&= \sum_{k=1}^{+\infty} \left(\frac{\lambda^k}{k!} e^{-\lambda}\right) \cdot \left(1 - F(\xi_i)^k\right) \\
&= \sum_{k=1}^{+\infty} \frac{\lambda^k}{k!} e^{-\lambda} \ - \ e^{-\lambda} \sum_{k=1}^{+\infty} \frac{(\lambda F(\xi_i))^k}{k!} \\
&= e^{-\lambda}\left(e^{\lambda} - 1\right) \ - \ e^{-\lambda}\left(e^{\lambda F(\xi_i)} - 1\right) \\
&= 1 - e^{\lambda(F(\xi_i) - 1)}
\end{aligned}$$

☞ Q2 En déduire la vraisemblance de $(T_1^*, \ldots, T_n^*)$.

On a pour $t \in [\![1, T-1]\!]$

$$\begin{aligned}
\mathbb{P}(T_i^* = t+1) &= \mathbb{P}(T_i^* = t+1 \ \wedge \ T_i > t) \quad \text{car } T_i^* - 1 < T_i \leq T_i^* \\
&= \mathbb{P}(T_i^* = t+1 \,|\, T_i > t) \cdot \mathbb{P}(T_i > t) \\
&= \mathbb{P}(T_i^* = t+1 \,|\, T_i > t) \cdot (1 - \mathbb{P}(T_i \leq t)) \\
&= \mathbb{P}(T_i^* = t+1 \,|\, T_i > t) \cdot \left(1 - \sum_{k=1}^{t} \mathbb{P}(T_i^* = k)\right)
\end{aligned}$$

ce qui s'écrit encore, en notant $u_t = \mathbb{P}(T_i^* = t)$ et $\rho = 1 - e^{\lambda(F(\xi_i)-1)}$

$$u_{t+1} = \rho\left(1 - \sum_{k=1}^{t} u_k\right)$$

avec $u_1 = \mathbb{P}(T_i^* = 1) = \mathbb{P}(T_i^* = 1 \,|\, T_i > 0) = \rho$.
Alors, $u_2 = \rho(1-\rho)$, puis $u_3 = \rho(1 - \rho - \rho(1-\rho)) = \rho(1-\rho)^2$.
Soit donc $t \geq 3$ tel que $\forall t' \leq t, u_{t'} = \rho(1-\rho)^{t'-1}$; alors

$$\begin{aligned}
u_{t+1} &= \rho\left(1 - \sum_{t'=1}^{t} u_{t'}\right) \\
&= \rho\left(1 - \sum_{t'=1}^{t} \rho(1-\rho)^{t'-1}\right) \\
&= \rho\left(1 - \rho\sum_{t'=0}^{t-1} (1-\rho)^{t'}\right) \\
&= \rho\left(1 - \rho\frac{1-(1-\rho)^t}{1-(1-\rho)}\right) \\
&= \rho\,(1-\rho)^t
\end{aligned}$$

Ainsi $\forall t \in [\![1, T-1]\!],\ u_{t+1} = \rho(1-\rho)^t$, ce qui s'écrit encore

$$\boxed{\forall t \in [\![1, T-1]\!],\ \mathbb{P}(T_i^* = t+1) = \left(1 - e^{\lambda(F(\xi_i)-1)}\right) e^{\lambda t(F(\xi_i)-1)}}$$

<u>*Autre méthode :*</u> On a directement pour $t \in [\![1, T-1]\!]$

$$\begin{aligned}
\mathbb{P}(T_i^* = t+1) &= \mathbb{P}(T_i^* = t+1 \ \wedge \ T_i > t) \quad \text{car } T_i^* - 1 < T_i \leq T_i^* \\
&= \mathbb{P}(T_i^* = t+1 \,|\, T_i > t) \cdot \mathbb{P}(T_i > t) \\
&= \mathbb{P}(T_i^* = t+1 \,|\, T_i > t) \cdot \left(\prod_{j=0}^{t-1} \mathbb{P}(T_i > j+1 \,|\, T_i > j) \right) \\
&= \mathbb{P}(T_i^* = t+1 \,|\, T_i > t) \cdot \left(\prod_{j=0}^{t-1} (1 - \mathbb{P}(T_i \leq j+1 \,|\, T_i > j)) \right) \\
&= \mathbb{P}(T_i^* = t+1 \,|\, T_i > t) \cdot \left(\prod_{j=0}^{t-1} (1 - \mathbb{P}(T_i^* = j+1 \,|\, T_i > j)) \right) \\
&= \left(1 - e^{\lambda(F(\xi_i)-1)}\right) \cdot \left(\prod_{j=0}^{t-1} \left(e^{\lambda(F(\xi_i)-1)}\right) \right) \\
&= \left(1 - e^{\lambda(F(\xi_i)-1)}\right) e^{\lambda t(F(\xi_i)-1)}
\end{aligned}$$

Les comportements des individus étant censément indépentants, les variables $(T_i^*)_{i \in [\![1,n]\!]}$ sont statistiquement indépendantes et par conséquent

$$L_{T_1^*,\ldots,T_n^*}(t_1, \ldots, t_n \ ; \ F, \lambda) = \Pi_{i=1}^n \left(1 - e^{\lambda(F(\xi_i)-1)}\right) e^{\lambda(F(\xi_i)-1)\cdot(t_i-1)} \mathbb{1}_{\min_i t_i \geq 1}$$

☞ Q3

> On suppose que tous les individus ont le même salaire de réserve : $\xi_i = \xi$ et que ce salaire de réserve commun ξ est connu, ainsi que la fonction de répartition F.
> Trouver l'estimateur du maximum de vraisemblance de λ.
> Montrer directement que cet estimateur est convergent et asymptotiquement efficace quand $n \to +\infty$.

Soit $(t_1, \ldots, t_n) \in [\![1, T]\!]^n$, et soit $\phi : \left(\begin{array}{ccc} [0,1] & \to & \mathbb{R}^+ \\ \eta & \mapsto & (1-\eta)^n \eta^{\sum_{i=1}^n (t_i - 1)} \end{array} \right)$.

Alors $\ln \phi$ est C^∞, et de dérivée $\frac{\partial \ln \phi}{\partial \eta}(\eta) = -\frac{n}{1-\eta} + \frac{1}{\eta} \sum_{i=1}^n (t_i - 1)$; donc elle est maximale en $\widehat{\eta} = 1 - \frac{1}{\bar{t}}$ où $\bar{t} = \frac{1}{n} \sum_{i=1}^n t_i$.

Or $L_{T_1^*,\ldots,T_n^*}(t_1, \ldots, t_n \ ; \ F, \lambda) = \phi\left(e^{\lambda(F(\xi)-1)}\right)$, donc par conséquent $\widehat{\lambda}$ vérifie $e^{\widehat{\lambda}(F(\xi)-1)} = \widehat{\eta} = 1 - \frac{1}{\bar{t}}$, c'est-à-dire

$$\boxed{\widehat{\lambda} = \frac{1}{1-F(\xi)} \ln(1 + \frac{1}{\bar{t}-1})}$$

On a d'après la loi de grands nombres

$$\begin{aligned}
\lim_{p.s.,n\infty} \bar{T} &= \mathbb{E}\left(T_i^*\right) \\
&= \sum_{t=1}^{+\infty} \mathbb{P}\left(T_i^* = t\right) \cdot t \\
&= \sum_{t=1}^{+\infty} (1-a)ta^{t-1}, \quad \text{en notant } a = e^{\lambda(F(\xi)-1)} \\
&= (1-a) \cdot \begin{pmatrix}]-1,1[& \to & \mathbb{R} \\ x & \mapsto & \sum_{t=1}^{+\infty} tx^{t-1} \end{pmatrix} (a) \\
&= (1-a) \cdot \frac{\partial \left(x \mapsto \sum_{t=0}^{+\infty} x^t\right)}{\partial x}(a), \quad \text{car la série converge absolument sur le disque ouvert} \\
&= (1-a) \cdot \frac{\partial \left(x \mapsto \frac{1}{1-x}\right)}{\partial x}(a) \\
&= (1-a) \cdot \frac{1}{(1-a)^2} \\
&= \frac{1}{1 - e^{\lambda(F(\xi)-1)}}
\end{aligned}$$

Remarque : Pour calculer $\sum_{k=0}^{+\infty} P(k)t^k$, il suffit de décomposer le polynôme $P(\mathbb{X})$ dans la base $1, \mathbb{X}-1, (\mathbb{X}-1)(\mathbb{X}-2), \ldots$ ce qui conduit à

$$\begin{aligned}
\sum_{k=0}^{+\infty} P(k)t^k &= \sum_{k=0}^{+\infty} \left(\sum_{i=0}^{d} \alpha_i \underbrace{(k-1)\cdots(k-i)}_{i \text{ termes}} \right) t^k \\
&= \sum_{i=0}^{d} \sum_{k=i}^{+\infty} \alpha_i \frac{\partial^i \left(x \mapsto x^k\right)}{\partial x^i}_{(t)} \\
&= \sum_{i=0}^{d} \alpha_i \frac{\partial^i \left(x \mapsto \frac{1}{1-x}\right)}{\partial x^i}_{(t)} \\
&= \sum_{i=0}^{d} \alpha_i \, \frac{(-1)^i i!}{(1-t)^{i+1}}
\end{aligned}$$

Par suite

$$\begin{aligned}
\widehat{\lambda} \xrightarrow[n\to+\infty]{p.s.} & \frac{1}{1-F(\xi)} \ln \left(1 + \frac{1}{\frac{1}{1-e^{\lambda(F(\xi)-1)}} - 1}\right) \\
= & \frac{1}{1-F(\xi)} \ln \left(1 + \frac{1 - e^{\lambda(F(\xi)-1)}}{e^{\lambda(F(\xi)-1)}}\right) \\
= & \frac{1}{1-F(\xi)} \ln \left(\frac{1}{e^{\lambda(F(\xi)-1)}}\right) \\
= & \lambda
\end{aligned}$$

ce qui prouve que $\widehat{\lambda}$ est un estimateur convergent de λ.

Cherchons enfin la loi limite de $(\widehat{\lambda}-\lambda)$; pous ce faire cherchons tout d'abord celle de $(\bar{T}-\mathbb{E}\left(\bar{T}\right))$. On a, en notant $a = e^{\lambda(F(\xi)-1)}$

$$\begin{aligned}
\mathbb{V}\left(T_i^*\right) &= \mathbb{E}\left(T_i^{*2}\right) - \mathbb{E}\left(T_i^*\right)^2 \\
&= (1-a)\sum_{t=1}^{+\infty} t^2 a^{t-1} - \frac{1}{(1-a)^2} \\
&= (1-a)\left(a\sum_{t=2}^{+\infty} t(t-1)a^{t-2} + \sum_{t=1}^{+\infty} t a^{t-1}\right) - \frac{1}{(1-a)^2} \\
&= (1-a)\left(a\frac{2}{(1-a)^3} + \frac{1}{(1-a)^2}\right) - \frac{1}{(1-a)^2}, \quad \text{car les séries convergent absolument} \\
&= \frac{1+a}{(1-a)^2} - \frac{1}{(1-a)^2} \\
&= \frac{a}{(1-a)^2}
\end{aligned}$$

D'après le Théorème Central Limite on a donc

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

donc d'après le théorème de Slutsky

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

où $g : \begin{pmatrix}]0,1[& \to & \mathbb{R} \\ x & \mapsto & \frac{1}{1-F(\xi)}\ln(1+\frac{1}{x-1}) \end{pmatrix}$.

On a $\forall x, g'(x) = \frac{1}{1-F(\xi)}\frac{1}{x(x-1)}$ et donc $g'\left(\frac{1}{1-a}\right) = \frac{1}{1-F(\xi)}\frac{(1-a)^2}{a}$ de sorte que finalement

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

☞ Q4

En pratique, l'enquête se termine à la fin du T-ième mois, $T < +\infty$: à cette date, certains individus sont encore au chômage ; on n'observe donc que :

$$\begin{aligned}
T_i^{**} &= T_i^* \text{ si } T_i^* \leq T \\
&= T+1 \text{ si } T_i^* > T
\end{aligned}$$

Ecrire la vraisemblance des observations $(T_1^{**}, \ldots, T_n^{**})$.
Donner l'estimateur du maximum de vraisemblance de λ.
<u>*Remarque*</u> : On suppose toujours le salaire de réserve commun ξ et la fonction de répartition F connue.

On a pour $t \in [\![1, T-1]\!]$

$$\begin{aligned}
\text{“}T_i^{**} = t+1\text{”} &= \begin{cases} \text{“}T_i^* = t+1\text{”} & \text{si “}T_i^* \leq T\text{”} \\ \text{“}T+1 = t+1\text{”} & \text{si “}T_i^* > T\text{”} \end{cases} \\
&= \begin{cases} \text{soit} & \text{“}T_i^* \leq T \text{ et } T_i^* = t+1\text{”} \\ \text{soit} & \text{“}T_i^* > T \text{ et } T+1 = t+1\text{”} \end{cases} \\
&= \begin{cases} \text{soit} & \text{“}t+1 \leq T \text{ et } T_i^* = t+1\text{”} \\ \text{soit} & \text{“}T_i^* = T+1 \text{ et } T+1 = t+1\text{ ”} \\ \text{soit} & \text{“}T_i^* = T+1 \text{ et } T+1 > t+1\text{ ”} \end{cases} \\
&= \begin{cases} \text{soit} & \text{“}t+1 \leq T \text{ et } T_i^* = t+1\text{”} \\ \text{soit} & \text{“}T_i^* = t+1 \text{ et } T+1 = t+1\text{ ”} \\ \text{soit} & \underbrace{\text{“}T_i^* = T+1 \text{ et } T+1 > t+1\text{”}}_{\text{événement impossible}} \end{cases}
\end{aligned}$$

Donc

$$\mathbb{P}\left(T_i^{**} = t+1\right) = \begin{cases} \mathbb{P}\left(T_i^* = t+1\right) & \text{si } t+1 \leq T \\ \mathbb{P}\left(T_i^* > T\right) & \text{si } t+1 = T+1 \\ 0 & \text{si } t+1 > T+1 \end{cases}$$

Or $\mathbb{P}\left(T_i^* = t+1\right) = \left(1 - e^{\lambda(F(\xi)-1)}\right) e^{\lambda t(F(\xi)-1)}$; et par ailleurs

$$\begin{aligned}
\mathbb{P}\left(T_i^* > T\right) &= \sum_{\tau=T+1}^{+\infty} \mathbb{P}\left(T_i^* = \tau\right) \\
&= \sum_{\tau=T}^{+\infty} \left(1 - e^{\lambda(F(\xi)-1)}\right) \left(e^{\lambda(F(\xi)-1)}\right)^{\tau} \\
&= e^{\lambda(F(\xi)-1)\cdot T}
\end{aligned}$$

En conséquence on a

$$\begin{aligned}
& L_{T_1^{**},\ldots,T_n^{**}}(t_1+1, \ldots, t_n+1\,;\, F, \lambda) \\
= \; & \Pi_{i\in[\![1,n]\!]} \mathbb{P}\left(T_i^{**} = t_i+1\right) \\
= \; & \left(\Pi_{t_i+1\leq T} \mathbb{P}\left(T_i^* = t_i+1\right)\right) \cdot \left(\Pi_{t_i+1>T} \mathbb{P}\left(T_i^* > T\right) \mathbb{1}_{t_i+1=T+1}\right) \\
= \; & \left(\Pi_{t_i+1\leq T} \left(1 - e^{\lambda(F(\xi)-1)}\right) e^{\lambda t_i(F(\xi)-1)}\right) \left(\Pi_{t_i\geq T} \left(e^{\lambda(F(\xi)-1)}\right)^T \cdot \mathbb{1}_{t_i+1=T+1}\right) \\
= \; & \left(\left(1 - e^{\lambda(F(\xi)-1)}\right)^{|\{i/t_i+1\leq T\}|} \cdot \left(e^{\lambda(F(\xi)-1)}\right)^{\sum_{t_i+1\leq T} t_i}\right) \left(\left(e^{\lambda(F(\xi)-1)}\right)^{\sum_{t_i\geq T} t_i}\right) \\
= \; & \left(\left(1 - e^{\lambda(F(\xi)-1)}\right)^{|\{i/t_i<T\}|}\right) \left(\left(e^{\lambda(F(\xi)-1)}\right)^{\sum_{\forall t_i} t_i}\right) \\
= \; & \left(\left(1 - e^{\lambda(F(\xi)-1)}\right)^{|\{i/t_i<T\}|}\right) \left(\left(e^{\lambda(F(\xi)-1)}\right)^{n\bar{t}}\right)
\end{aligned}$$

On en tire que

$$\frac{\partial \ln L_{T_1^{**},\ldots,T_n^{**}}}{\partial \lambda}(t_1, \ldots, t_n\;;\; F, \lambda) = -\frac{|\{i/t_i < T\}|}{1 - e^{\lambda(F(\xi)-1)}} + \frac{n(\bar{t}-1)}{e^{\lambda(F(\xi)-1)}}$$

de sorte que

$$e^{\widehat{\lambda(F(\xi)-1)}} = \frac{n(\bar{t}-1)}{|\{i/t_i < T\}| + n(\bar{t}-1)}$$

et donc

$$\boxed{\widehat{\lambda^{censure}} = \frac{1}{F(\xi)-1} \ln\left(\frac{n(\bar{t}-1)}{|\{i/t_i<T\}|+n(\bar{t}-1)}\right)}$$

Comme en l'absence de censure ($T = +\infty$) on a $\widehat{\lambda} = \frac{1}{F(\xi)-1} \ln\left(1 - \frac{1}{\bar{t}}\right)$ on a

$$\begin{aligned}
\widehat{\lambda^{censure}} - \widehat{\lambda} &= \frac{1}{F(\xi)-1}\left(\ln\left(\frac{\bar{t}-1}{\frac{|\{i/t_i<T\}|}{n} + \bar{t} - 1}\right) - \ln\left(\frac{\bar{t}-1}{\bar{t}}\right)\right) \\
&= -\frac{1}{F(\xi)-1} \ln\left(1 + \frac{|\{i/t_i < T\}|}{n}\right)
\end{aligned}$$

☞ Q5

On suppose désormais que F s'écrit :

$$F(x) = \left(1 - e^{-\gamma(\xi-\xi_0)}\right) \mathbb{1}_{\xi \geq \xi_0}$$

où γ est un paramètre inconnu à estimer et ξ_0 est connu.
Le couple (λ, γ) est-il identifiable ?

La vraisemblance s'écrit cette fois

$$L_{T_1^{**},\ldots,T_n^{**}}(t_1, \ldots, t_n \ ; \ \gamma, \lambda) = \left(1 - e^{-\lambda e^{\gamma(\xi-\xi_0)}}\right)^n \left(e^{-\lambda e^{\gamma(\xi-\xi_0)}}\right)^{n\bar{t}} \mathbb{1}_{\xi\geq\xi_0}$$

On constate alors que $\phi : ((\lambda,\gamma) \mapsto \lambda r^\gamma)$ n'est jamais injective (car $\forall r \in \mathbb{R}^{+*}, \phi(1,1) = r = \phi\left(\frac{1}{2}, \sqrt{r}\right)$), de sorte que la vraisemblance $\left((\lambda,\gamma) \mapsto L_{T_1^{**},\ldots,T_n^{**}}(t_1,\ldots,t_n \ ; \ \gamma,\lambda)\right)$ ne l'est pas non plus, de sorte que le paramètre (λ, γ) n'est **pas** identifiable. D'où l'intérêt des calculs précédents …

Exercice corrigé 3

■ Partie 1 <u>**Préliminaire**</u>

☞ Q1

On rappelle que la loi exponentielle de paramètre λ admet la densité $f(y,\lambda) = \begin{cases} \lambda e^{(-\lambda y)}, \text{ si } y \geq 0 \\ 0 \text{ sinon.} \end{cases}$
Soient n variables aléatoires $Y_1 ... Y_n$ i.i.d. de densité $f(\cdot, \lambda)$.
Montrer que $\sum_{i=1}^{n} Y_i$ suit une loi de densité $\lambda^n \frac{y^{n-1}}{(n-1)!} e^{-\lambda y} \mathbb{1}_{[0,+\infty]}(y)$ (on pourra commencer par le montrer pour $n = 2$ puis procéder par récurrence).

L'assertion est triviale lorsque $n = 1$.
Soit donc $n \geq 1$ tel que l'assertion soit vraie.
Soient $Y_1, \ldots, Y_{n+1}$ i.i.d. de densité $f_{\mathcal{E}}(\cdot, \lambda)$.
Définissons pour $k \in [\![1, n+1]\!]$ $S_k = \sum_{i=1}^k Y_i$.
Constatant que $S_{n+1} = S_n + Y_{n+1}$, définissons $\phi : \begin{pmatrix} \mathbb{R}^{+*2} & \longrightarrow & \mathbb{R}^{+*}\times]0,1[\\ (x,y) & \longmapsto & (x+y, \frac{x}{x+y})) \end{pmatrix}$. Alors ϕ est un C^∞-difféomorphisme, de jacobien $\left| \begin{pmatrix} \text{y} & \text{x} \\ \text{1-y} & \text{x} \end{pmatrix}^{-1} \right| = \frac{1}{x}$.
Or $L_{S_n, Y_{n+1}}(s, y) = \left(\lambda^n \frac{s^{n-1}}{(n-1)!} e^{-\lambda s} \mathbb{1}_{s>0}\right) \cdot \left(\lambda e^{-\lambda y} \mathbb{1}_{y>0}\right)$, par hypothèse de récurrence et puisque S_n et Y_{n+1} sont indépendantes (car les $(Y_i)_i$ sont toutes indépendantes).
En conséquence, puisque $\phi^{-1}(u, v) = (uv, u(1-v))$, $L_{\phi(S_n, Y_{n+1})}(u, v) = \lambda^{n+1} \frac{(uv)^{n-1}}{(n-1)!} e^{-\lambda u} u \mathbb{1}_{u>0\ \wedge\ v\in]0,1[}$
.
En particulier d'après le théorème de projection $L_{(\phi(S_n, Y_{n+1}))_1}(u) = \lambda^{n+1} \frac{u^n}{n!} e^{-\lambda u} \mathbb{1}_{u>0}$, c'est-à-dire

$$L_{Y_1 + \cdots + Y_{n+1}}(u) = \lambda^{n+1} \frac{u^n}{n!} e^{-\lambda u} \mathbb{1}_{u>0}$$

de sorte que l'assertion est vraie pour tout $n \in \mathbb{N}^*$.

Dans tout le problème, Z_i et C_i, pour $i \in [\![1, n]\!]$ désignent des variables aléatoires indépendantes suivant des lois exponentielles de paramètres respectifs $\lambda, \mu \in \mathbb{R}^{+*}$.

■ Partie 2 **Observation parfaite**

On dispose d'un échantillon d'observations $(z_i, c_i)_{i\in[\![1,n]\!]}$

☞ Q1

Ecrire le modèle statistique correspondant. S'agit-il d'une famille exponentielle ? Si oui, peut-on exhiber une statistique exhaustive ?

Le modèle statistique s'écrit $\left\{ \left(\mathbb{R}^{+2}\right)^n\ ,\ (\mathcal{E}\cdot\mathcal{E})^{\otimes n} \right\}$.
On a de plus pour $i \in [\![1, n]\!]$ $L_{Z_i, C_i}(z_i, c_i\ ;\ \lambda, \mu) = \lambda\mu e^{-\lambda z_i - \mu c_i}$, et donc

$$L_{Z_1,\ldots,Z_n,C_1,\ldots,C_n}(z_1, \ldots, z_n, c_1, \ldots, c_n\ ;\ \lambda, \mu) = (\lambda\mu)^n e^{-\lambda \sum_{i=1}^n z_i - \mu \sum_{i=1}^n c_i} \mathbb{1}_{\min_i z_i \geq 0} \mathbb{1}_{\min_i c_i \geq 0}$$

En particulier le modèle est exponentiel et une statistique exhaustive est

$$(z_1, \ldots, z_n, c_1, \ldots, c_n) \mapsto \left(\sum_{i=1}^n z_i, \sum_{i=1}^n c_i \right)$$

☞ Q2

Quel est l'estimateur du maximum de vraisemblance $\begin{pmatrix} \hat{\lambda} \\ \hat{\mu} \end{pmatrix}$ du paramètre $\begin{pmatrix} \lambda \\ \mu \end{pmatrix}$?

On a immédiatement $\frac{\partial \ln L_{Z_1,\ldots,Z_n,C_1,\ldots,C_n}}{\partial \lambda} = \frac{n}{\lambda} - \sum_{i=1}^n z_i$, et par suite un estimateur de λ et $\widehat{\lambda} = \frac{1}{\bar{z}}$ (la log-vraisemblance étant concave, elle est bien maximale en $\bar{z}$). De même un estimateur de μ et $\widehat{\mu} = \frac{1}{\bar{c}}$.

☞ Q3 | Déterminer la loi asymptotiquement de $\begin{pmatrix} \hat{\lambda} \\ \hat{\mu} \end{pmatrix}$.

On calcule successivement

$$\begin{aligned} \frac{\partial^2 \ln L_{Z_1,\ldots,Z_n,C_1,\ldots,C_n}}{\partial \lambda^2} &= \frac{n}{\lambda^2} \\ \frac{\partial^2 \ln L_{Z_1,\ldots,Z_n,C_1,\ldots,C_n}}{\partial \lambda \partial \mu} &= 0 \\ \frac{\partial^2 \ln L_{Z_1,\ldots,Z_n,C_1,\ldots,C_n}}{\partial \mu^2} &= \frac{n}{\mu^2} \end{aligned}$$

de sorte que $I_1(\lambda,\mu) = \begin{pmatrix} \frac{1}{\lambda^2} & 0 \\ 0 & \frac{1}{\mu^2} \end{pmatrix}$ d'inverse $I_1(\lambda,\mu)^{-1} = \begin{pmatrix} \lambda^2 & 0 \\ 0 & \mu^2 \end{pmatrix}$ et donc d'après le Théorème Central Limite

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

☞ Q4 | Déterminer la loi (à distance finie) de $\begin{pmatrix} \hat{\lambda} \\ \hat{\mu} \end{pmatrix}$.
Est-il biaisé à distance finie ?
Calculer sa matrice de variance-covariance.

Déterminons la loi de $\widehat{\lambda}$: soit donc $h : (\mathbb{R} \to \mathbb{R})$ mesurable bornée, et calculons $\mathbb{E}\left(h(\widehat{\lambda})\right)$. On a

$$\begin{aligned} \mathbb{E}\left(h(\widehat{\lambda})\right) &= \mathbb{E}\left(h\left(\frac{n}{\sum_{i=1}^n z_i}\right)\right) \\ &= \int_{\mathbb{R}} h\left(\frac{n}{s}\right) \lambda^n \frac{s^{n-1}}{(n-1)!} e^{-\lambda s} \mathbb{1}_{s>0}\, ds \\ &= -\int_{0^+}^{+\infty} h(u) \lambda^n \frac{\left(\frac{n}{u}\right)^{n-1}}{(n-1)!} e^{-\lambda \frac{n}{u}} \mathbb{1}_{u>0} \left(-\frac{n}{u^2}\right) du \end{aligned}$$

En conséquence, $L_{\widehat{\lambda}}(u) = \frac{\lambda^n}{n!}\left(\frac{n}{u}\right)^{n+1} e^{-\lambda\frac{n}{u}} \mathbb{1}_{u>0}$.
De façon similaire, $L_{\widehat{\mu}}(v) = \frac{\mu^n}{n!}\left(\frac{n}{v}\right)^{n+1} e^{-\mu\frac{n}{v}} \mathbb{1}_{v>0}$.
Ainsi,

$$L_{\widehat{\lambda},\widehat{\mu}}(l,m) = \frac{(\lambda\mu)^n}{(n!)^2}\left(\frac{n^2}{lm}\right)^{n+1} e^{-n\frac{\lambda}{l} - n\frac{\mu}{m}}$$

Par ailleurs on a pour $n \geq 2$

$$\begin{aligned}
\mathbb{E}\left(\widehat{\lambda}\right) &= \int_{\mathbb{R}} l \frac{\lambda^n}{n!}\left(\frac{n}{l}\right)^{n+1} e^{-\lambda\frac{n}{l}} \mathbb{1}_{l>0} dl \\
&= \int_{0^+}^{+\infty} \frac{\lambda^n}{(n-1)!}\left(\frac{n}{l}\right)^{n} e^{-\lambda\frac{n}{l}} dl \\
&= \int_{0^+}^{+\infty} \frac{\lambda^n}{(n-1)!} u^{n-2} e^{-\lambda u} n du \\
&= \frac{n\lambda}{(n-1)} \int_{0^+}^{+\infty} \frac{1}{(n-2)!} \lambda^{n-2} u^{n-2} e^{-\lambda u}\, \lambda du \\
&= \frac{n\lambda}{(n-1)} \underbrace{\int_{0^+}^{+\infty} \frac{1}{(n-2)!} v^{n-2} e^{-v} dv}_{\int_{0^+}^{+\infty} f_{\Gamma(n-2,\lambda)}(v)dv = 1} \\
&= \frac{n}{n-1}\lambda
\end{aligned}$$

de sorte que

$$\mathbb{E}\left(\begin{pmatrix} \widehat{\lambda} \\ \widehat{\mu} \end{pmatrix}\right) = \frac{n}{n-1}\begin{pmatrix} \lambda \\ \mu \end{pmatrix}$$

Autrement dit, $\begin{pmatrix} \widehat{\lambda} \\ \widehat{\mu} \end{pmatrix}$ est biaisé à distance finie.

Enfin pour $n \geq 3$

$$\begin{aligned}
\mathbb{E}\left(\widehat{\lambda}^2\right) &= \int_{\mathbb{R}} l^2 \frac{\lambda^n}{n!}\left(\frac{n}{l}\right)^{n+1} e^{-\lambda\frac{n}{l}} \mathbb{1}_{l>0} dl \\
&= -\int_{0^+}^{+\infty} \left(\frac{n}{u}\right)^{2} \frac{\lambda^n}{n!} u^{n+1} e^{-\lambda u} \left(-\frac{n}{u^2}\right) du \\
&= \frac{n^2\lambda^2}{(n-1)(n-2)} \int_{0^+}^{+\infty} \frac{1}{(n-3)!} \lambda^{n-3} u^{n-3} e^{-\lambda u}\, \lambda du \\
&= \frac{n^2\lambda^2}{(n-1)(n-2)} \underbrace{\int_{0^+}^{+\infty} \frac{1}{(n-3)!} v^{n-3} e^{-v} dv}_{\int_{0^+}^{+\infty} f_{\Gamma(n-3,\lambda)}(v)dv = 1} \\
&= \frac{n^2}{(n-1)(n-2)}\lambda^2
\end{aligned}$$

et donc

$$\begin{aligned}
\mathbb{V}\left(\widehat{\lambda}\right) &= \mathbb{E}\left(\widehat{\lambda}^2\right) - \mathbb{E}\left(\widehat{\lambda}\right)^2 \\
&= \frac{n^2}{(n-1)(n-2)}\lambda^2 - \left(\frac{n}{n-1}\lambda\right)^2 \\
&= \frac{n^2(n-1)-(n-2)n^2}{(n-1)^2(n-2)}\lambda^2 \\
&= \frac{n^2}{(n-1)^2(n-2)}\lambda^2
\end{aligned}$$

(on vérifie que l'e.m.v. est asymptotiquement efficace : $\mathbb{V}\left(\widehat{\lambda}\right) \underset{n\infty}{\sim} \frac{1}{n}\lambda^2$).

Autre méthode (plus rapide) :

$$\begin{aligned}
\mathbb{E}\left(\widehat{\lambda}\right) &= \int_{\mathbb{R}} \frac{n}{s}\lambda^n \frac{s^{n-1}}{(n-1)!}e^{-\lambda s}ds \\
&= \frac{n}{n-1}\lambda \int_{\mathbb{R}} \lambda^{n-1}\frac{s^{n-2}}{(n-2)!}e^{-\lambda s}ds \\
&= \frac{n}{n-1}\lambda \int_{\mathbb{R}} \underbrace{\lambda^{n-1}\frac{s^{n-2}}{(n-2)!}e^{-\lambda s}}_{f_{\Gamma(n-1,\lambda)}(s)} ds \\
&= \frac{n}{n-1}\lambda
\end{aligned}$$

De même

$$\begin{aligned}
\mathbb{E}\left(\widehat{\lambda}^2\right) &= \frac{n^2}{(n-1)(n-2)}\lambda^2 \int_{\mathbb{R}} \underbrace{\lambda^{n-2}\frac{s^{n-3}}{(n-3)!}e^{-\lambda s}}_{f_{\Gamma(n-2,\lambda)}(s)} ds \\
&= \frac{n^2}{(n-1)(n-2)}\lambda^2
\end{aligned}$$

Enfin, remarquant que toutes les $Z_1, \ldots, Z_n, C_1, \ldots, C_n$ sont indépendantes, on a $\mathbb{C}ov\left(\widehat{\lambda}, \widehat{\mu}\right) = 0$ de sorte que

$$\boxed{\mathbb{V}\left(\begin{pmatrix} \widehat{\lambda} \\ \widehat{\mu} \end{pmatrix}\right) = \frac{n^2}{n-1}\left(\frac{1}{n-2} - \frac{1}{n-1}\right)\begin{pmatrix} \lambda^2 & 0 \\ 0 & \mu^2 \end{pmatrix}}$$

☞ Q5 Proposer des estimateurs sans biais optimaux de λ et μ, si possible efficaces.

Posons $\widehat{\lambda}^* = \frac{n-1}{n}\widehat{\lambda}$ et $\widehat{\mu}^* = \frac{n}{n-1}\widehat{\mu}$.

D'après le résultat précédent, $\begin{pmatrix} \widehat{\lambda}^* \\ \widehat{\mu}^* \end{pmatrix}$ est un estimateur sans biais de $\begin{pmatrix} \lambda \\ \mu \end{pmatrix}$. Or $T(Z,C) =$

$\left(\frac{1}{n}\sum_{i=1}^{n} Z_i, \frac{1}{n}\sum_{i=1}^{n} C_i\right)$ est une statistique exhaustive complète (car canonique dans un modèle exponentielle) ; en conséquence, d'après le théorème de Lehman-Scheffé, $\begin{pmatrix} \widehat{\lambda}^* \\ \widehat{\mu}^* \end{pmatrix}$ est optimal parmi les estimateurs sans biais de $\begin{pmatrix} \lambda \\ \mu \end{pmatrix}$.

Par ailleurs,

$$\begin{aligned}
\mathbb{C}ov\left(\begin{pmatrix} \widehat{\lambda}^* \\ \widehat{\mu}^* \end{pmatrix}\right) &= \left(\frac{n-1}{n}\right)^2 \mathbb{C}ov\left(\begin{pmatrix} \widehat{\lambda} \\ \widehat{\mu} \end{pmatrix}\right) \\
&= \frac{1}{n-2}\begin{pmatrix} \lambda^2 & 0 \\ 0 & \mu^2 \end{pmatrix}
\end{aligned}$$

Or la borne FDCR du modèle est $\frac{1}{n}\begin{pmatrix} \lambda^2 & 0 \\ 0 & \mu^2 \end{pmatrix}$.

En conséquence, $\begin{pmatrix} \widehat{\lambda}^* \\ \widehat{\mu}^* \end{pmatrix}$ n'est **pas** un estimateur *efficace* de $\begin{pmatrix} \lambda \\ \mu \end{pmatrix}$.

■ Partie 3 **Observation imparfaite**

On suppose dans cette seconde partie que les seules observations disponibles portent sur $X_i = \text{Min}\left(Z_i, C_i\right),\ i = 1, ..., n$.

☞ Q1 Calculer la fonction de répartition de la variable X_i pour $i \in [\![1, n]\!]$

On a pour $t \in \mathbb{N}$

$$\begin{aligned}
\mathbb{P}\left(X_i > t\right) &= \mathbb{P}\left(Z_i > t\ \wedge\ C_i > t\right) \\
&= \mathbb{P}\left(Z_i > t\right) \cdot \mathbb{P}\left(C_i > t\right) \quad \text{par indépendance} \\
&= \left(\int_t^{+\infty} \lambda e^{-\lambda u} du\right)\left(\int_t^{+\infty} \lambda e^{-\mu v} dv\right) \\
&= e^{-(\lambda+\mu)t}
\end{aligned}$$

de sorte que $X_i \rightsquigarrow \mathcal{E}(\lambda + \mu)$.

☞ Q2 Ecrire le modèle statistique correspondant et déterminer les fonctions identifiables du paramètre $\begin{pmatrix} \lambda \\ \mu \end{pmatrix}$.

Le modèle statistique s'écrit ici $\{\mathbb{R}^{+n}\ ,\ \mathcal{E}^{\otimes n}\}$.

Sa vraisemblance étant $L_{X_1,\ldots,X_n}(x_1, \ldots, x_n\ ;\ \lambda, \mu) = (\lambda + \mu)^n e^{-(\lambda+\mu)\sum_{i=1}^{n} x_i}$, la seule fonction identifiable de (λ, μ) est $\lambda + \mu$.

☞ Q3

Quels sont les estimateurs du maximum de vraisemblance de $\gamma = \lambda + \mu$ fondés : i) sur $(X_i)_{i\in[\![1,n]\!]}$; ii) sur $(Z_i, C_i)_{i\in[\![1,n]\!]}$? Est-il naturel que ces estimateurs soient différents ?

Cherchons à estimer $\gamma = \lambda + \mu$.

(i) La maximisation de $L_{X_1,\dots,X_n}(x_1,\dots,x_n\ ;\ \gamma) = (\gamma)^n e^{-(\gamma)\sum_{i=1}^n x_i}$ conduit à l'estimateur $\widehat{\gamma_X} = \frac{1}{\bar{x}} = \frac{n}{\sum_{i=1}^n \min(z_i,c_i)}$;

(ii) tandis que celle de $L_{Z_1,\dots,Z_n,C_1,\dots,C_n}(z_1,\dots,z_n,c_1,\dots,c_n\ ;\ \lambda,\mu)$ conduit à estimer λ et μ individuellement, ce dont on tire une estimation convergente de $\lambda+\mu$, à savoir $\widehat{\gamma_{Z,C}} = \widehat{\lambda}+\widehat{\mu} = \frac{n}{\sum_{i=1}^n Z_i} + \frac{n}{\sum_{i=1}^n C_i}$

Bien entendu il n'y a aucune raison pour que ces estimateurs, fondés sur des observations différentes (c'est-à-dire, en définitive, définis dans des modèles statistiques différents), soient les mêmes. Comme il est montré ci-après ils ont même des variances asymptotiques différentes.

☞ Q4

Comparer les propriétés asymptotiques de ces estimateurs.

On a alternativement

(i) $\frac{\partial \ln L_{X_1,\dots,X_n}}{\partial\gamma}(x_1,\dots,x_n\ ;\ \gamma) = \frac{n}{\gamma} - \sum_{i=1}^n x_i$,
puis $\frac{\partial^2 \ln L_{X_1,\dots,X_n}}{\partial\gamma^2}(x_1,\dots,x_n\ ;\ \gamma) = -\frac{n}{\gamma^2}$,
donc

$$\sqrt{n}\,(\widehat{\gamma_X} - \gamma) \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}(O, (\lambda+\mu)^2)$$

(ii) Soit $g : \begin{pmatrix} \mathbb{R}^{+*2} & \to & \mathbb{R}^{+*2} \\ (x,y) & \mapsto & (x+y,y) \end{pmatrix}$ qui est de classe C^∞, et de jacobienne $\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$. Alors d'après le théorème de Slutsky

$$\sqrt{n}\left(g\begin{pmatrix} \widehat{\lambda} \\ \widehat{\mu} \end{pmatrix} - g\begin{pmatrix} \lambda \\ \mu \end{pmatrix}\right) \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \frac{\partial g}{\partial(x,y)}(\lambda,\mu)\begin{pmatrix} \lambda^2 & 0 \\ 0 & \mu^2 \end{pmatrix}\frac{\partial g}{\partial(x,y)'}(\lambda,\mu)\right)$$

En particulier selon la première composante,[8]

$$\sqrt{n}\,(\widehat{\gamma_{Z,C}} - \gamma) \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}(O, \lambda^2+\mu^2)$$

On constate que $(\lambda+\mu)^2 \geq \lambda^2+\mu^2$, avec égalité *ssi* $\lambda = 0$ ou $\mu = 0$ (situations dégénérées et sans intérêt). Autrement dit, asympotiquement $\widehat{\gamma^{Z,C}}$ est préférable à $\widehat{\gamma^X}$, ce que l'on écrira sous la forme publicitaire

$$\boxed{\widehat{\lambda}+\widehat{\mu} \gg \widehat{\lambda+\mu}}$$

Ce résultat est bien-sûr naturel dans la mesure où le premier estimateur est fondé sur une information plus fine que le second.

[8] Il n'est pas nécessaire pour pratiquer la delta méthode que g soit inversible, et il aurait suffit ici de considérer $g : \begin{pmatrix} \mathbb{R}^{+*}2 & \to & \mathbb{R}^{+*} \\ (x,y) & \mapsto & x+y \end{pmatrix}$.

■ Partie 4 **Conclusion**

☞ Q1

Déduire des parties **I** et **II** l'expression de l'espérance conditionnelle

$$\mathrm{E}\left(\frac{1}{\sum_{i=1}^n \min(Z_i, C_i)}\middle| \sum_{i=1}^n Z_i, \sum_{i=1}^n C_i\right)$$

Pour calculer $\mathbb{E}\left(\frac{1}{X_1+\cdots+X_n} \mid Z_1+\cdots+Z_n, C_1+\cdots+C_n\right)$, il semble naturel de déterminer tout d'abord

$$\begin{aligned} & L_{X_1+\cdots+X_n \mid Z_1+\cdots+Z_n, C_1+\cdots+C_n} \\ = \ & \frac{L_{X_1+\cdots+X_n, Z_1+\cdots+Z_n, C_1+\cdots+C_n}}{L_{Z_1+\cdots+Z_n, C_1+\cdots+C_n}} \end{aligned}$$

La difficulté ici est que X_i n'est **pas** indépendante de (Z_i, C_i), puisque $X_i = \min(Z_i, C_i)$, ce qui complique singulièrement le calcul.
En fait, le résultat s'exprime très simplement et peut être démontré de la façon suivante, astucieuse quoique moins naturelle.
Notons en effet $\widehat{\theta} = \mathbb{E}\left(\frac{1}{X_1+\cdots+X_n} \mid Z_1+\cdots+Z_n = z, C_1+\cdots+C_n = c\right)$; alors d'après la partie **II** $\widehat{\theta}$ est un estimateur sans biais de $\frac{\lambda+\mu}{n-1}$. La statistique en $\widehat{\theta}(X_1,\ldots,X_n)$ étant régulière (car le modèle est exponentiel), d'après le théorème de Lehman-Scheffé $\widehat{\theta}$ est en outre optimal.
Mais d'après la partie **I**, $\frac{\widehat{\lambda+\mu}}{n}$ est un autre estimateur sans biais de $\frac{\lambda+\mu}{n-1}$, lui aussi optimal.
En conséquence, $\widehat{\theta}$ et $\widehat{\theta} - \frac{\widehat{\lambda+\mu}}{n}$ sont décorréllés, ainsi que $\frac{\widehat{\lambda}}{n}$ et $(\widehat{\theta} - \frac{\widehat{\mu}}{n}) - \frac{\widehat{\lambda}}{n}$, et que $\frac{\widehat{\mu}}{n}$ et $(\widehat{\theta} - \frac{\widehat{\lambda}}{n}) - \frac{\widehat{\mu}}{n}$, et ce en vertu du lemme suivant :

Si $\widehat{a}$ et $\widehat{b}$ sont deux estimateurs sans biais optimaux du même paramètre, alors $\mathbb{C}ov\left(\widehat{a}, \widehat{a-b}\right) = 0$

En effet, définissons pour $t \in \mathbb{R}$ $c_t = (1-t)\widehat{a} + t\widehat{b} = \widehat{a} + t\left(\widehat{a} - \widehat{b}\right)$.
Alors $\forall t \in \mathbb{R},\ \mathbb{V}(c_t) \geq \mathbb{V}\left(\widehat{b}\right)$ puisque $\widehat{b}$ est optimal ;
ceci s'écrit $\forall t \in \mathbb{R},\ \mathbb{V}(\widehat{a}) + 2t\mathbb{C}ov\left(\widehat{a}, \widehat{a} - \widehat{b}\right) + t^2\mathbb{V}\left(\widehat{a} - \widehat{b}\right)$ qui est une inégalité polynômiale en t qui ne p
de celui-ci est négatif ou nul, ce dont on tire le résultat.

Par suite,

$$\begin{aligned} \mathbb{V}\left(\widehat{\theta}\right) &= \mathbb{V}\left(\frac{\widehat{\lambda}}{n}\right) + \mathbb{V}\left(\frac{\widehat{\mu}}{n}\right) \\ &= \mathbb{V}\left(\frac{\widehat{\lambda} + \widehat{\mu}}{n}\right) \quad \text{car } \widehat{\lambda} \text{ et } \widehat{\mu} \text{ sont indépendants} \end{aligned}$$

Dans ces conditions, les deux estimateurs $\widehat{\theta}$ et $\frac{\widehat{\lambda}+\widehat{\mu}}{n}$ qui ont même espérance, même variance et sont exactement corrélés, sont égaux presque sûrement :[9]

$$\widehat{\theta} = \frac{\widehat{\lambda}+\widehat{\mu}}{n} \text{p.s.}$$

c'est-à-dire

$$\boxed{\mathbb{E}\left(\tfrac{1}{\sum_{i=1}^n \min(Z_i,C_i)} \mid \sum_{i=1}^n Z_i \ \wedge\ \sum_{i=1}^n C_i\right) = \tfrac{1}{\sum_{i=1}^n Z_i} + \tfrac{1}{\sum_{i=1}^n C_i} \quad \text{p.s.}}$$

[9]Ce résultat est généralement connu comme l'*unicité presque sûre* d'un estimateur optimal fonction d'une statistique totale.

6 Travaux Dirigés n°6

Exercice corrigé 1

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

☞ Q1

> Donner $\mathbb{E}(X_i)$ et $\mathbb{V}(X_i)$.
> Calculer $\phi(t) = \mathbb{E}\left(e^{tX_i}\right)$.
> En déduire que
>
> $$\begin{aligned} \mathbb{E}(X) &= \lambda \\ \mathbb{E}\left(X^2\right) &= \lambda(1+\lambda) \\ \mathbb{E}\left(X^3\right) &= \lambda\left(1+3\lambda+\lambda^2\right) \\ \mathbb{E}\left(X^4\right) &= \lambda\left(1+7\lambda+6\lambda^2+\lambda^3\right) \end{aligned}$$

– 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

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

ce dont on tire

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

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

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

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

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

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

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

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 $n \in \mathbb{N}$

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

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 $P_n(\mathbb{X}) = \sum_{k=0}^{n} a_n^k\mathbb{X}^k$

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

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 pose $\widehat{\lambda_1}(x) = \frac{1}{n}\sum_{i=1}^n x_i = \bar{x}$ et $\widehat{\lambda_2}(x) = \frac{1}{n}\sum_{i=1}^n (x_i - \bar{x})^2$.
Par quelle méthode d'estimation ont été obtenus $\widehat{\lambda_1}$ et $\widehat{\lambda_2}$?

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) Les estimateurs $\widehat{\lambda_1}$ et $\widehat{\lambda_2}$ estiment-ils λ sans biais ?
Proposer un autre estimateur sans biais de λ, que l'on notera $\widehat{\lambda_3}$.

On a

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

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) Donner la loi (à distance finie) de $\widehat{\lambda_1}$.

<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

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

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) Donner la loi asymptotique jointe de $\begin{pmatrix}\widehat{\lambda_1} \\ \widehat{\lambda_2}\end{pmatrix}$, puis celle de $\begin{pmatrix}\widehat{\lambda_1} \\ \widehat{\lambda_3}\end{pmatrix}$.

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

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

Enfin, posons $g : \begin{pmatrix} \mathbb{R}^2 & \to & \mathbb{R}^2 \\ \begin{pmatrix} u \\ v \end{pmatrix} & \mapsto & \begin{pmatrix} u \\ v-u^2 \end{pmatrix} \end{pmatrix}$, 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

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

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) Soit $\widehat{\lambda_\infty}=\alpha\widehat{\lambda_1}+\beta\widehat{\lambda_3}$, $(\alpha,\beta)\in\mathbb{R}^{+2}$.
Donner la valeur de (α,β) telle que $\widehat{\lambda_\infty}$ soit le meilleur estimateur asymptotiquement sans biais de λ.
Ce résultat est-il surprenant (donner l'e.m.v. de λ) ?

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 $n\to+\infty$

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

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 Dans cette question, on se place dans le cas où $n = 2$.

(a) Calculer les erreurs quadratiques moyennes pour les trois estimateurs $\widehat{\lambda_1}$, $\widehat{\lambda_2}$ et $\widehat{\lambda_3}$.

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) $\boxed{\text{Déterminer en fonction de } \lambda \text{ lequel de ces estimateurs est le meilleur selon la critère de l'erreur quadratique moyenne.}}$

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) $\boxed{\text{Comparer explicitement l'EQM de } \hat{\lambda}_3 \text{ avec la borne FDCR.}}$

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 $\lambda > 0$.

Exercice corrigé 2

On considère une population de n individus infectés par un virus ; on étudie leurs durées d'incubation $(T_i)_{i\in[\![1,n]\!]}$, 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 $\lambda_i e^{-\lambda_i t}\mathbb{1}_{t\geq 0}$) ;
- la famille $(\Lambda_i)_{i\in[\![1,n]\!]}$ est identiquement distribuée selon la loi $\mathcal{L}$ de densité

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

  où $\alpha > 0$ et $r > 2$;
- les couples (T_i, Λ_i) sont indépendants entre eux.

☞ Q1 $\boxed{\text{Donner la vraisemblance de } (T_1, \ldots, T_n).}$

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

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

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 Calculer, lorsqu'il existe, le moment d'odre k $\mathbb{E}\left(T_i^k\right)$.

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 $\frac{n!}{p!(n-p)!}$)

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

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

Calculer l'information de Fisher du modèle. Dans le cas où α est connu, calculer l'estimateur du maximm de vraisemblance de r. Que se passe-t-il si α et r sont tous deux inconnus ?

– La vraisemblance s'écrit pour $t_1, \ldots, t_n > 0$

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

et donc

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

puis

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

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 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 ?

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

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

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 $g : \begin{pmatrix} \mathbb{R}^{+*} & \to & \mathbb{R}^{+*} \\ s & \mapsto & 1 + \frac{\alpha}{s} \end{pmatrix}$; 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

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

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 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.

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

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

En l'occurence on pose

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

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

(b) Donner la loi limite du vecteur

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

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

On a successivement

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

de sorte que d'après le Théorème Central Limite

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

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

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

En définitive on a[10]

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

Exercice corrigé 3

On considère les réalisations de T variables aléatoires i.i.d. $Y_1, Y_2, \ldots, Y_T$, issues d'une loi de Poisson de paramètre λ inconnu.
On s'intéresse au test de l'hypothèse :

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

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.

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 $H_a = \neg H_0$.

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

– <u>Test de *Wald* :</u>
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

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

– <u>Test du *score* :</u>
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

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

Donc

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

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

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

– <u>Test du *rapport des maxima de vraisemblance* :</u>
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

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

et donc par la loi forte des grands nombres

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

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

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

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 | 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 $\alpha \in [0,1]$: $W_\alpha = \{(y_1,\ldots,y_T) \ / \ \xi_T^{test}(y_1,\ldots,y_T) > q_{1-\alpha}^{\chi^2_1}\}$

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}.

- <u>Test de *Wald* :</u>
  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

$$n\frac{\left(\widehat{\lambda}-\lambda\right)^2}{\lambda} \xrightarrow[n\to+\infty]{\mathcal{L}} \chi^2_1$$

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

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

  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

On veut tester $\lambda_0 = 1$ au seuil $\alpha = 5\%$.
Décider de chacun des tests dans le cas où l'échantillon observé $(y_1, \ldots, y_T)$ 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.

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 $q^{\chi^2_1}_{1-5\%}$
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%.

7 Travaux Dirigés n°7

Exercice corrigé 1

Pour étudier l'arrivée des appels dans un central téléphonique, on comptabilise lors de 200 observations consécutives, le nombre d'appels observés par seconde, ce qui produit les résultats suivants :

Nombre d'appels par seconde	Effectifs observés
0	6
1	15
2	40
3	42
4	37
5	30
6	10
7	9
8	5
9	3
10	2
11	1

On suppose les arrivées des appels indépendantes, et en outre que la probabilité élémentaire λdt qu'il arrive un appel entre les instants t et $t + dt$ est indépendante de la date t.
Autrement dit

- le nombre N_t d'appels observés sur l'intervalle de temps $[0, t]$ suit une loi de Poisson de paramètre λt.
- plus généralement $N_{t+s} - N_t$ est indépendant de N_t et suit une loi de Poisson de paramètre λs.

☞ Q1

(a) Tester l'adéquation de la loi empirique à la famille des lois de Poisson.

<u>*Rappel :*</u>
Essentiellement, si $(X_i)_{i\in[\![1,n]\!]} \underset{iid}{\rightsquigarrow} \mathcal{L}$ loi discrète sur $[\![1, m]\!]$, et si $p_k = \mathbb{P}(X_i = k)$ et $N_k = |\{i/X_i = k\}|$ est le nombre de réalisations de $k \in [\![1, m]\!]$, alors (en appliquant le Théorème Central Limite centré réduit à $(N_1, \dots, N_m)$ puis en l'"élevant au carré") on a

$$\xi_n^{\text{adéq}} = \sum_{k=1}^{m} \frac{(N_k - np_k)^2}{np_k} \xrightarrow[n\to+\infty]{\mathcal{L}} \chi^2_{m-1}$$

(cf cours de Paul Doukhan, théorème 9.2) On en déduit une région de test asymptotique de niveau $\alpha \in]0, 1[$: $W_n^{\text{adéq}} = \{\xi_n^{\text{adéq}} > q_{1-\alpha}^{\chi^2_{m-1}}\}$. A distance finie on s'assurera avant d'appliquer ce test asymptotique que $\forall k \in [\![1, m]\!]$, $np_k > 5$, quitte à fusionner plusieurs classes : cette condition est requise pour que l'approximation d'une binômiale en normale soit raisonnable. Enfin, dans le cas d'une loi continue on se ramènera au cas précédent en discrétisant l'espace des réalisations en m classes.

<u>*Pourquoi ne pas utilise le test de Kolmogorov-Smirnov ?*</u> (cf théorème 9.3)
D'une certaine façon ce test se fonde sur la distance en $\|\|_\infty$ tandis que le test du χ^2 se fonde sur la distance en $\|\|_2$ entre la loi théorique et la loi empirique. De plus, ce test utilise une statistique dont on ne sait pas exprimer la loi limite (mais on sait la tabuler) ; à l'inverse le test du χ^2 utilise une statistique dont la loi limite est bien connue mais nécessite de discrétiser l'univers continu en classes ce qui introduit une erreur. De toute façon ce test est inapplicable ici car la loi de Poisson est discrète, tandis que le test du χ^2 est bien adapté.
Pour tester l'adéquation à une famille de lois paramétrée, une idée naturelle est de tester l'adéquation simple à la loi la plus vraisemblable, c'est-à-dire à la loi dont le paramètre est l'estimateur du maximum de vraisemblance.
Considérons donc le modèle statistique $\left(\mathbb{N}^{200}, \tau, \left(\mathcal{P}(\lambda)^{\otimes 200}\right)_{\lambda \in \mathbb{R}^{+*}}\right)$. Dans ce modèle la vraisemblance s'écrit $L_{X_1,\ldots,X_n}(x_1,\ldots,x_n;\lambda) = \prod_{i=1}^n e^{-\lambda}\frac{\lambda^{x_i}}{x_i!}$ et l'estimateur du maximum de vraisemblance est donc $\widehat{\lambda} = \bar{X}$; on a numériquement

$$\begin{aligned}\bar{x} &= \frac{6\cdot 0+15\cdot 1+40\cdot 2+42\cdot 3+37\cdot 4+30\cdot 5+10\cdot 6+9\cdot 7+5\cdot 8+9\cdot 3+2\cdot 10+1\cdot 11}{200} \\ &\simeq 3,7\end{aligned}$$

En conséquence, si la famille $(X_i)_i$ suivait bien une loi de Poisson, celle-ci serait de paramètre $\widehat{\lambda} \simeq 3,7$ et par conséquent on aurait

$$\begin{aligned}\forall k \in [\![0,11]\!],\ \widehat{p_k} &= \mathbb{P}_{\widehat{\lambda}}(X_i = k) \\ &= e^{-\widehat{\lambda}}\frac{\widehat{\lambda}^k}{k!} \\ &\simeq e^{-3,7}\frac{3,7^k}{k!}\end{aligned}$$

et il ne reste plus qu'à pratiquer une test d'adéquation simple à la loi $\mathcal{P}\left(\widehat{\lambda}\right)$.
Pour ce faire, afin de préserver la légitimité du test asymptotique à distance finie on s'assure que $\forall k \in [\![0,11]\!]$, $N_k < 5$, ce qui conduit à fusionner les quatre dernières modalités selon le tableau suivant :

Modalité	Effectif observé
$X_i = 0$	$N_0 = 6$
$X_i = 1$	$N_1 = 15$
$X_i = 2$	$N_2 = 40$
$X_i = 3$	$N_3 = 42$
$X_i = 4$	$N_4 = 37$
$X_i = 5$	$N_5 = 30$
$X_i = 6$	$N_6 = 10$
$X_i = 7$	$N_7 = 9$
$X_i \geq 8$	$N_8 = 11$

Par conséquent $\xi_{200}^{\text{adéq}} = \sum_{k=0}^{8}\frac{(N_k - n\widehat{p_k})^2}{n\widehat{p_k}} \simeq 7,58$. Or si $(X_i)_i$ suit bien une certaine loi de poisson, alors $\xi_{200}^{\text{adéq}}$ suit une loi du χ^2 à $(9-1) - \dim(\mathbb{R}^{+*}) = 7$ degrés de liberté (puisqu'on n'a conservé que 9 classes). Le fractile de niveau 5% valant $q_{95\%}^{\chi_7^2} \simeq 14,1$, on **accepte** finalement au seuil 5% l'hypothèse d'adéquation de la vrai loi inconnue de $(X_i)_i$ à la famille des lois de Poisson ; on peut en outre affirmer que $(X_i)_i$ suit vraisemblablement une loi de Poisson de paramètre $3,7$.

(b) Quel est le niveau limite permettant d'accepter l'adéquation à la loi de Poisson ?

Le niveau limite entre l'acceptation et le rejet de l'hypothèse pour l'échantillon observé est par définition la *p-value*, qui est telle que $\xi_{200}^{\text{adéq}} = q_{p-value}^{\chi_7^2}$; on a ici $p - value \simeq 19\%$, ce qui est assez élevé : l'hypothèse est rejetée au seuil 20%, mais acceptée à tout niveau $\alpha \leq 18\%$.

☞ Q2 A une date antérieure le nombre d'appels sur un intervalle de temps $[0,t]$ suivait une loi de Poisson de paramètre 4.
Tester la stabilité du comportement entre les deux dates.

On cherche cette fois à tester l'adéquation simple de la loi des $(X_i)_i$ à la loi de Poisson de paramètre 4.

De la même façon que dans le cas précédent les quatre dernières classes doivent être regroupées, et on calcule successivement $p_k = e^{-4} \cdot \frac{4^k}{k!}$ pour $k \in [\![0,8]\!]$, puis $\xi_{200}^{\text{adéq}} = \sum_{k=0}^{8} \frac{(N_k - np_k)^2}{np_k} \simeq 16,8$, qui est censée suivre une loi du χ^2 à $(9-1) = 8$ degrés de liberté (il n'y a pas de paramètre à estimer ici, donc $\dim(\Theta) = 0$). Comme $q_{99\%}^{\chi_8^2} \simeq 20,09$ on **accepte** l'hypothèse de stabilité au niveau 1%.

☞ Q3 Que donnerait un test d'adéquation à une loi de Poisson de paramètre $3,7$? Conclure.
Table de la loi de Poisson :

	$\mathbb{P}_{3,7}(X=k)$	$\mathbb{P}_4(X=k)$
0	0.0247	0,0183
1	0.0915	0,0733
2	0.1684	0,1465
3	0.2087	0,1957
4	0.1930	0,1954
5	0.1428	0,1563
6	0.0881	0,1042
7	0.0465	0,0595
8	0.0215	0,0298
9	0.0099	0,0132
10	0.0046	0,0053
11	0.0021	0,0019
12	0.0001	0,0006
13	< 0.0001	0,0002
14	< 0.0001	0,0001

De façon similaire, on reprend $p_k = e^{-3,7}\frac{3,7^k}{k!}$ pour $k \in [\![0,8]\!]$, ce dont on tire $\xi_{200}^{\text{adéq}} \simeq 7,58$, qui est censée suivre une loi du χ^2 à 8 degrés de liberté (là encore, $\lambda = 3,7$ est donné a priori et il n'y a pas de paramètre à estimer). Bien évidemment, 3,7 étant la valeur de l'estimateur du maximum de vraisemblance dans le modèle statistique d'une famille de lois de Poisson de paramètre inconnu, à laquelle l'adéquation a déjà été acceptée au niveau α, l'adéquation simple à la loi de paramètre $\widehat{\lambda} = 3,7$ est **nécessairement** acceptée au même niveau α puisque $1 - F_{\chi^2_8}(x) > 1 - F_{\chi^2_7}(x)$ pour tout x.

En définitive, en notant $p^{\chi^2}(k)$ la p-value de la loi du χ^2 à k degrés de liberté on a :

$\xi_{200}^{\text{adéq}}$	Adéquation à la famile $\mathcal{P}(\lambda)_\lambda$	Adéquation simple à la loi $\mathcal{P}(3,7)$
$[0, q^{\chi^2}(7)[$	acceptée	acceptée
$]q^{\chi^2}(7), q^{\chi^2}(8)[$	rejetée	acceptée
$]q^{\chi^2}(8), +\infty[$	rejetée	rejetée

Exercice corrigé 2

On s'intéresse à la proportion des ménages équipés d'un magnétoscope. Pour cela, on tire de manière équiprobable avec remise un échantillon de n ménages, et on observe pour chaque ménage $i \in [\![1,n]\!]$ la variable

$$y_i = \begin{cases} 1 \text{ si le ménage } i \text{ est équipé} \\ 0 \text{ sinon.} \end{cases}$$

☞ Q1 Réaliser un test de l'hypothèse nulle : “la proportion des ménages équipés n'excède pas 20 %”.

Soit p la vraie probabilité qu'un ménage soit équipé en magnétoscope (on identifiera la proportion de ménages équipés à p). On a immédiatement $\widehat{p}(y_1, \ldots, y_n) = \bar{y}$, pour lequel $\sqrt{n}\,(\widehat{p} - p) \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}\left(0, p(1-p)\right)$

En appliquant la même méthode que pour l'exercice exercice 3 , on montre qu'on accepte H_0 : “$p \leq p_0 = 20\%$” *ssi* $\bar{y} < k_\alpha$ avec

$$k_\alpha = p_0 + q_{1-\alpha}^{\mathcal{N}(0,1)}\sqrt{\frac{\bar{y}((1-\bar{y})}{n}}$$

☞ Q2 On se demande si la probabilité qu'un ménage i soit équipé n'est pas fonction d'une variable (scalaire) x_i donnée (revenu, âge du chef de famille...).

On définit à cet effet un modèle statistique conditionnel à X_i de la façon suivante :

$$\mathbb{P}(y_i = 1 \mid x_i) = \frac{e^{a+bx_i}}{1+e^{a+bx_i}}$$

où a et b sont deux paramètres réels inconnus.

On cherche alors à tester $\underline{H_0\ :}\ b = 0$.

(a) Calculer le score et la matrice d'information de Fisher du modèle pour n observations.

La vraisemblance du modèle s'écrit pour $(y_1,\dots,y_n)\in\mathbb{R}^{+n}$ et $(y_1,\dots,y_n)\in\mathbb{R}^n$

$$L_{Y_1,\dots,Y_n|X_1=x_1,\dots,X_n=x_n}\left(y_1,\dots,y_n;a,b\right)=\prod_{i=1}^{n}\frac{\left(e^{a+bx_i}\right)^{y_i}}{1+e^{a+bx_i}}$$

Le score s'écrit donc

$$\frac{\partial L_{Y_1,\dots,Y_n|X_1=x_1,\dots,X_n=x_n}}{\partial\begin{pmatrix}a\\b\end{pmatrix}}\left(y_1,\dots,y_n;a,b\right)=\begin{pmatrix}\sum_{i=1}^{n}\left(y_i-\frac{e^{a+bx_i}}{1+e^{a+bx_i}}\right)\\ \sum_{i=1}^{n}\left(y_i-\frac{e^{a+bx_i}}{1+e^{a+bx_i}}\right)x_i\end{pmatrix}$$

et l'information de Fisher est donc

$$I_n(a,b)=\sum_{i=1}^{n}\mathbb{E}\left(\frac{e^{a+bx_i}}{\left(1+e^{a+bx_i}\right)^2}\begin{pmatrix}1 & x_i\\ x_i & {x_i}^2\end{pmatrix}\right)$$

qu'on ne peut expliciter davantage sans hypothèse sur la distribution de $(X_i)_i$.

(b) Expliciter le modèle contraint par H_0.
Calculer les estimateurs du maximum de vraisemblance $\widehat{a^0}$ et $\widehat{b^0}$ de a et b dans ce modèle, puis évaluer le score en $\left(\widehat{a^0},\widehat{b^0}\right)$.

Sous l'hypothèse H_0 : "$b=0$" il vient

$$\forall x_i\in\mathbb{R},\ \mathbb{P}_{H_0}\left(Y_i=1|X_i=x_i\right)=\frac{e^a}{1+e^a}$$

c'est-à-dire que $(Y_i|X_i)$ suit une binômiale de paramètre $p_0=\frac{e^a}{1+e^a}$ indépendant de x_i.
La vraisemblance conditionnelle sous H_0 est donc

$$L^{H_0}_{Y_1,\dots,Y_n|X_1=x_1,\dots,X_n=x_n}\left(y_1,\dots,y_n;a,b\right)=\frac{e^{a\sum_{i=1}^{n}y_i}}{\left(1+e^a\right)^n}$$

et un estimateur $\widehat{a^0}$ de a vérifie donc, sous H_0, $\frac{e^{\widehat{a^0}}}{1+e^{\widehat{a^0}}}=\bar{y}$, soit $\widehat{a^0}=\ln\frac{\bar{y}}{1-\bar{y}}$. On a bien entendu $\widehat{b^0}=0$.
Il vient alors

$$\frac{\partial L^{H_0}_{Y_1,\dots,Y_n|X_1=x_1,\dots,X_n=x_n}}{\partial\begin{pmatrix}a\\b\end{pmatrix}}\left(y_1,\dots,y_n;\widehat{a^0},\widehat{b^0}\right)=\begin{pmatrix}0\\ \sum_{i=1}^{n}\left(y_i-\bar{y}\right)x_i\end{pmatrix}$$

(c) Donner un estimateur $\widehat{I_1^{H_0}}$ de I_1, convergent sous H_0 et fonction des estimateurs contraints des paramètres.

La matrice d'information de Fisher sous H_0 est

$$I_n(a,b) =_{|H_0} \frac{e^a}{(1+e^a)^2}\mathbb{E}\left(\sum_{i=1}^{n}\begin{pmatrix} 1 & x_i \\ x_i & {x_i}^2 \end{pmatrix}\right)$$

Or $\frac{e^{\widehat{a^0}}}{1+e^{\widehat{a^0}}} = \bar{y}$, donc $\frac{e^{\widehat{a^0}}}{\left(1+e^{\widehat{a^0}}\right)^2} = \bar{y}(1-\bar{y})$ et par conséquent (en notant $\bar{x^2} = \frac{1}{n}\sum_{i=1}^{n} {x_i}^2$)

$$\widehat{I_1^{H_0}} = \bar{y}(1-\bar{y})\begin{pmatrix} 1 & \bar{x} \\ \bar{x} & \bar{x^2} \end{pmatrix}$$

est un estimateur convergent de I_1 si H_0 est vraie.

(d) Exprimer la statistique du test du score de l'hypothèse H_0.
Quelle est sa loi asymptotique sous H_0 ? Sur quelle corrélation repose le test ?

On a par définition

$$\xi_n^S = \frac{1}{n}\widehat{S^0}'\left(\widehat{I_1^{H_O}}\right)^{-1}\widehat{S^0}$$

en notant $\widehat{S^0} = \frac{\partial L_{Y_1,\ldots,Y_n|X_1=x_1,\ldots,X_n=x_n}}{\partial(a,b)'}(y_1,\ldots,y_n;\widehat{a^0},\widehat{b^0})$.
Or

$$\begin{aligned}\left(\widehat{I_1^{H_O}}\right)^{-1} &=_{|H_0} \frac{1}{\bar{y}(1-\bar{y})}\cdot\left(\begin{pmatrix} 1 & \bar{x} \\ \bar{x} & \bar{x^2} \end{pmatrix}\right)^{-1} \\ &=_{|H_0} \frac{1}{\bar{y}(1-\bar{y})}\cdot\frac{1}{\bar{x^2}-(\bar{x})^2}\cdot\begin{pmatrix} \bar{x^2} & -\bar{x} \\ -\bar{x} & 1 \end{pmatrix}\end{aligned}$$

Donc

$$\begin{aligned}&\xi_n^S(y_1,\ldots,y_n) \\ =_{|H_0}\ & \frac{1}{n}\left(0,\sum_{i=1}^{n}(y_i-\bar{y})\,x_i\right)\times\left(\frac{1}{\bar{y}(1-\bar{y})}\frac{1}{\bar{x^2}-(\bar{x})^2}\begin{pmatrix} \bar{x^2} & -\bar{x} \\ -\bar{x} & 1 \end{pmatrix}\right)\times\begin{pmatrix} 0 \\ \sum_{i=1}^{n}(y_i-\bar{y})\,x_i \end{pmatrix} \\ =_{|H_0}\ & \frac{1}{n}\left(\sum_{i=1}^{n}(y_i-\bar{y})\,x_i\right)\cdot\frac{1}{\bar{y}(1-\bar{y})}\frac{1}{\bar{x^2}-(\bar{x})^2}\cdot\left(\sum_{i=1}^{n}(y_i-\bar{y})\,x_i\right) \\ =_{|H_0}\ & \frac{1}{\bar{y}(1-\bar{y})}\cdot\frac{\left(\sum_{i=1}^{n}(y_i-\bar{y})\,x_i\right)^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\end{aligned}$$

Par construction, si H_0 est vraie alors

$$\xi_n^S \xrightarrow[n\to+\infty]{\mathcal{L}} \chi_1^2$$

ce qui permet de tester effectivement H_0.
Remarquons que ξ_n^S est une fonction de $(X_i)_{i\in[\![1,n]\!]}$ uniquement à travers sa variance et sa covariance avec $(Y_i)_{i\in[\![1,n]\!]}$; en outre elle est croissante avec cette dernière, comme le suggère l'intuition : dire que la loi de Y_i conditionnellement à X_i dépend effectivement de X_i (ce qui revient à dire "$\neg H_0$"), c'est dire que les observations Y_i sont significativement corrélées avec celles de X_i, et inversement.

Exercice corrigé 3

Un industriel reçoit K lots de n pièces, avec la garantie que la proportion p de pièces défectueuses est la même dans chaque lot et inférieure à 5%. Pour vérifier que la garantie est exacte, on tire avec remise des pièces dans chaque lot, jusqu'à obtenir une pièce défectueuse par lot. Soit Y_k, la variable aléatoire désignant le nombre de tirages nécessaires dans le lot k.

☞ Q1 | Calculer la loi de Y_k.
Calculer $\mathbb{E}(Y_k)$ et $\mathbb{V}(Y_k)$.

Dans la pratique la variable Y_k est tronquée, c'est-à-dire l'opérateur n'attend pas plus de $N \geq 1$ tirages successifs qu'une pièce soit défectueuse. Les calculs pourraient être menés directement dans le cas où $N = +\infty$, ce qui revient à supprimer tous les termes en facteur de " x^N " dans les lignes qui suivent.
On a pour $k \in [\![1, K]\!]$ et $n \in [\![1, N]\!]$, en notant $M = \sum_{n=N+1}^{+\infty} p(1-p)^{n-1}$

$$\begin{aligned}\mathbb{P}(Y_k = n) &= \mathbb{P}\left(\text{"Les } n-1 \text{ premières pièces tirées sont conformes et la } n\text{-ième est défectueuse"}\right)\\ &= \frac{1}{M}(1-p)^{n-1}\cdot p \quad \text{(le tirage s'effectue avec remise)}\end{aligned}$$

Le tirage s'effectuant avec remise, le nombre de tirages nécessaires n pour obtenir une pièce défectueuse n'est pas borné (avec toutefois $\mathbb{P}(n = +\infty) = 0$).

$$
\begin{aligned}
\mathbb{E}\left(Y_k\right) &= \sum_{n=0}^{N} n\mathbb{P}(Y_k = n) \\
&= \frac{1}{M}\sum_{n=0}^{N} np(1-p)^{n-1} \\
&= \frac{1}{M}p\sum_{n=0}^{N}\left(\frac{\partial}{\partial x}(x \mapsto x^n)\right)_{(1-p)} \\
&= \frac{1}{M}p\left(\left(\frac{\partial}{\partial x}(x \mapsto \sum_{n=0}^{N} x^n)\right)_{(1-p)}\right) \\
&= \frac{1}{M}p\left(\left(\frac{\partial}{\partial x}(x \mapsto \frac{1-x^{N+1}}{1-x})\right)_{(1-p)}\right) \\
&= \frac{1}{M}p\left(x \mapsto \frac{-(N+1)x^N(1-x)+(1-x^{N+1})}{(1-x)^2}\right)_{(1-p)} \\
&= \frac{1}{M}p\left(x \mapsto \frac{1+Nx^{N+1}-(N+1)x^N}{(1-x)^2}\right)_{(1-p)} \\
&= \frac{1}{M}\frac{1}{p}\left(1-(1-p)^N\left((N+1)p+1-p\right)\right) \\
&\underset{N\infty}{\longrightarrow} \frac{1}{p}
\end{aligned}
$$

et d'autre part que

$$\begin{aligned}
\mathbb{E}\left({Y_k}^2\right) &= \sum_{n=0}^{N} n^2 \mathbb{P}(Y_k = n) \\
&= \frac{1}{M} \sum_{n=0}^{N} n^2 p (1-p)^{n-1} \\
&= \frac{1}{M} p \left((1-p) \sum_{n=2}^{N} n(n-1)(1-p)^{n-2} + \sum_{n=1}^{N} n(1-p)^{n-1} \right) \\
&= \frac{1}{M} p \left((1-p) \left(\sum_{n=2}^{N} \frac{\partial^2}{\partial x^2} (x \mapsto x^n) \right)_{(1-p)} + \left(\sum_{n=1}^{N} \frac{\partial}{\partial x} (x \mapsto x^n) \right)_{(1-p)} \right) \\
&= \frac{1}{M} p \left((1-p) \left(\frac{\partial^2}{\partial x^2} (x \mapsto \frac{1-x^{N+1}}{1-x}) \right)_{(1-p)} + \left(\frac{\partial}{\partial x} (x \mapsto \frac{1-x^{N+1}}{1-x}) \right)_{(1-p)} \right) \\
&= \frac{1}{M} \left\{ \begin{array}{l} p(1-p) \left(x \mapsto \frac{(N(N+1)x^N - (N+1)Nx^{N-1})(1-x) + 2(1+Nx^{N+1} - (N+1)x^N)}{(1-x)^3} \right)_{(1-p)} \\ +p \left(x \mapsto \frac{1+Nx^{N+1} - (N+1)x^N}{(1-x)^2} \right)_{(1-p)} \end{array} \right. \\
&= \frac{1}{M} \left\{ \begin{array}{l} p(1-p) \left(x \mapsto \frac{2 - N(N+1)x^{N-1} + 2N^2 X^N - N(N-1)x^{N+1}}{(1-x)^3} \right)_{(1-p)} \\ +p \left(x \mapsto \frac{1+Nx^{N+1} - (N+1)x^N}{(1-x)^2} \right)_{(1-p)} \end{array} \right. \\
&= \frac{1}{p^2} \left\{ \begin{array}{l} 2(1-p) - N(N+1)(1-p)^N + 2N^2(1-p)^{N+1} - N(N+1)(1-p)^{N+2} \\ \ldots + p - (N+1)p(1-p)^N - (1-p)^{N+1} \end{array} \right. \\
&= \frac{1}{M} \frac{2 - p - (1-p)^N \left(N(N+1) - 2N^2(1-p) + N(N-1)(1-p)^2 + (N+1)p \right)}{p^2} \\
&\xrightarrow[N\infty]{} \frac{2-p}{p^2}
\end{aligned}$$

de sorte que

$$\begin{aligned}
\mathbb{V}(Y_k) &= \mathbb{E}\left({Y_k}^2\right) - \left(\mathbb{E}(Y_k)\right)^2 \\
&= \frac{1}{M} \frac{1}{p^2} \left\{ \begin{array}{l} 1-p \\ -(1-p)^N \left[N(N+1) - 2N^2(1-p) + N(N-1)(1-p)^2 - ((N+1)p + 2 - 2p) \right] \\ -(1-p)^{2N} \left[N(N+1) - 2N^2(1-p) + N(N-1)(1-p)^2 + (N+1)p \right]^2 \end{array} \right. \\
&\xrightarrow[N\infty]{} \frac{1-p}{p^2}
\end{aligned}$$

On se placera dans toute la suite dans le cas où $N = +\infty$

☞ Q2

Proposer un test de Wald de l'hypothèse : $\underline{H_0} : p = 5\%$

Le test de Wald se fonde sur la normalité asymptotique de l'estimateur du maximum de vraisemblance, qui vaut ici $\widehat{p} = \frac{1}{\bar{y}}$. Ecrire le Théorème Central Limite implique donc de l'écrire en $\bar{y}$ puis d'appliquer le théorème de Slutsky à $g : (x \mapsto \frac{1}{x})$. On en tirerait alors la statistique de Wald habituelle, dont on saurait qu'elle suit asymptotiquement une loi du χ^2 à un degré de liberté.
Pour autant, comme g est bijective, nous préférons appliquer le Théorème Central Limite à $\bar{y}$ (comparé à $g^{-1}(p_0) = \frac{1}{p_0}$), et reconstruire de façon analogue à la construction de la statistique de Wald une statistique qui suit un χ^2 à un degré de liberté.
On a en effet d'après le Théorème Central Limite

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

donc

$$\sqrt{K}\frac{\bar{y} - \frac{1}{p}}{\sqrt{\frac{1-p}{p^2}}} \xrightarrow[K\infty]{\mathcal{N}} (0,1)$$

et donc

$$K\frac{\left(\bar{y} - \frac{1}{p}\right)^2}{\frac{1-p}{p^2}} \xrightarrow[K\infty]{} \chi_1^2$$

Enfin, en approchant la variance $\frac{1-p}{p^2}$ par $\frac{1-\frac{1}{\bar{y}}}{\frac{1}{\bar{y}^2}} = \bar{y}(\bar{y}-1)$ il vient

$$\xi_n^W(y_1, \ldots, y_n) = K\frac{\left(\bar{y} - \frac{1}{p_0}\right)^2}{\bar{y}(\bar{y}-1)} \xrightarrow[K\infty]{H_0} \chi_1^2$$

☞ Q3

> Soit $A_k = \left\{\bar{Y} < k_\alpha\right\}$, pour k_α donné ; calculer $\mathbb{P}_p(A_k)$ en fonction de $p = p_0$.
> Montrer que $\sup_{p\leq 5\%}\left(\lim_{K\infty}\mathbb{P}_p(A_k)\right) = \lim_{K\infty}\mathbb{P}_{p=5\%}(A_k)$.
> En déduire un test asymptotique de l'hypothèse :
>
> $\underline{H_0} : p \leq 5\%$

On cherche à tester l'hypothèse H_0 : "$p \leq p_0 = 5\%$" au seuil $\alpha \in [0,1[$ en comparant $\widehat{p} = \bar{y}$ à un certain fractile k_α. L'idée est donc d'exploiter la normalité asymptotique de l'estimateur $\widehat{p}$ pour en déduire la forme explicite (et tabulable) de k_α
Considérons donc $k_\alpha \in \mathbb{R}$; on a

$$\begin{aligned}
\mathbb{P}_p(A_k) &= \mathbb{P}_p(\bar{y} < k_\alpha) \\
&= \mathbb{P}_p\left(\sqrt{K}\frac{\bar{y} - \frac{1}{p}}{\sqrt{\bar{y}(\bar{y}-1)}} < \sqrt{K}\frac{k_\alpha - \frac{1}{p}}{\sqrt{\bar{y}(\bar{y}-1)}}\right)
\end{aligned}$$

<u>*Remarque :*</u> Même si cela ne change rien à la méthode, rien n'implique que k_α ne dépende pas de la taille K de l'échantillon. On est donc amené à exprimer k_α en tant que fonction de K, supposée sufissamment régulière.

Ainsi, si $k_\alpha : (\mathbb{N} \mapsto \mathbb{R})$ est une fonction de K qui admet une limite finie lorsque $K \to +\infty$, on a

$$\mathbb{P}_p\left(A_k\right) \xrightarrow[K\infty]{} \Phi\left(\lim_{K\infty} \sqrt{K}\frac{k_\alpha(K) - \frac{1}{p}}{\sqrt{\bar{y}(\bar{y}-1)}}\right)$$

où Φ est la fonction de répartition de la loi normale centrée réduite ; il reste déterminer $\sup_{p\leq p_0} \lim_{K\infty} \mathbb{P}_p\left(A_k\right)$.
Or si $p_2 \leq p_1 \leq p_0$, on a

$$\sqrt{K}\frac{k_\alpha(K) - \frac{1}{p_2}}{\sqrt{\bar{y}(\bar{y}-1)}} \leq \sqrt{K}\frac{k_\alpha(K) - \frac{1}{p_1}}{\sqrt{\bar{y}(\bar{y}-1)}}$$

et par conséquent

$$\Phi\left(\lim_{K\infty} \sqrt{K}\frac{k_\alpha(K) - \frac{1}{p_2}}{\sqrt{\bar{y}(\bar{y}-1)}}\right) \leq \Phi\left(\lim_{K\infty} \sqrt{K}\frac{k_\alpha(K) - \frac{1}{p_1}}{\sqrt{\bar{y}(\bar{y}-1)}}\right)$$

soit encore

$$\mathbb{P}_{p_2}\left(A_k\right) \leq \mathbb{P}_{p_1}\left(A_k\right)$$

Il s'ensuit que

$$\begin{aligned}
\sup_{p\leq p_0} \lim_{K\infty} \mathbb{P}_p\left(A_k\right) &= \lim_{K\infty} \sup_{p\leq p_0} \mathbb{P}_p\left(A_k\right) \quad \text{par uniforme continuité} \ldots \\
&= \lim_{K\infty} \mathbb{P}_{p=p_0}\left(A_k\right) \\
&= \Phi\left(\lim_{K\infty} \sqrt{K}\frac{k_\alpha(K) - \frac{1}{p}}{\sqrt{\bar{y}(\bar{y}-1)}}\right)
\end{aligned}$$

Pour K asymptotiquement grand et $\alpha \in [0, 1[$ on accepte donc H_0 *ssi*

$$\sqrt{K}\frac{k_\alpha(K) - \frac{1}{p_0}}{\sqrt{\bar{y}(\bar{y}-1)}} > q_\alpha^{\mathcal{N}(0,1)}$$

où $q_\alpha^{\mathcal{N}(0,1)}$ désigne le fractile de niveau α de la loi normale centrée réduite.
L'idée est alors de définir k_α de façon à ce que $\Phi(T(k_\alpha)) = q_\alpha^{\mathcal{N}(0,1)}$, où $T(x) = \sqrt{K}\frac{x-\frac{1}{p}}{\sqrt{\bar{y}(\bar{y}-1)}}$; de cette façon, tester si $p \leq p_0$ au seuil α, c'est-à-dire comparer $\sup_{p\leq p_0} \mathbb{P}_{p,k\infty}\left(A_k\right)$ à α, revient juste à comparer $\bar{y}$ à k_α.
Plus précisément, H_0 est acceptée *ssi* $\bar{y} > k_\alpha(K)$ avec

$$\boxed{k_\alpha(K) = \frac{1}{p_0} + q_\alpha^{\mathcal{N}(0,1)}\sqrt{\frac{\bar{y}(\bar{y}-1)}{K}}}$$

(et on vérifie a posteriori que $\lim_{K\infty} k_\alpha(K) = \frac{1}{p_0}$ existe).

Exercice corrigé 4

On dispose de n observations i.i.d. d'un couple de variables aléatoires positives scalaires (W_i, X_i).
Le but de l'exercice est de suggérer une procédure pour tester l'hypothèse H_0 selon laquelle la loi conditionnelle de W_i sachant X_i est une loi de Pareto, de densité :

$$f(W|X,b) = \frac{bX}{W_0}\left(\frac{W_0}{W}\right)^{bX+1} \mathbb{1}_{W \geq W_0}$$

pour $b > 0$ et $W_0 > 0$

☞ Q1 On suppose dans un premier temps que les X_i sont tous égaux à $X \in \mathbb{R}$ connu, et que W_0 est connu, de sorte que le seul paramètre inconnu du modèle est $\alpha = bX + 1$.

(a) Calculer l'espérance et la variance de W, notées m et v.
A quelle condition sur α, supposée vérifiée par la suite, les moments m et v existent-ils ?

On a

$$\begin{aligned} m &= \mathbb{E}(W) \\ &= \int_{\mathbb{R}} \frac{bx}{W_0}\left(\frac{W_0}{W}\right)^{\alpha} \mathbb{1}_{W \geq W_0} W dW \\ &= (\alpha-1)(\alpha-1){W_0}^{\alpha-1} \int_{W_0}^{+\infty} W^{1-\alpha} dW \\ &= (\alpha-1){W_0}^{\alpha-1}\left[\frac{W^{2-\alpha}}{2-\alpha}\right]_{W_0}^{+\infty} \\ &= \boxed{\tfrac{\alpha-1}{\alpha-2} W_0} \end{aligned}$$

aussitôt que $\alpha > 2$, et par ailleurs

$$\begin{aligned} \mathbb{E}\left(W^2\right) &= \int_{\mathbb{R}} \frac{bx}{W_0}\left(\frac{W_0}{W}\right)^{\alpha} \mathbb{1}_{W \geq W_0} W^2 dW \\ &= (\alpha-1){W_0}^{\alpha-1} \int_{W_0}^{+\infty} W^{2-\alpha} dW \\ &= (\alpha-1){W_0}^{\alpha-1}\left[\frac{W^{3-\alpha}}{3-\alpha}\right]_{W_0}^{+\infty} \\ &= \frac{\alpha-1}{\alpha-3}{W_0}^2 \end{aligned}$$

d'où

$$\begin{aligned} v &= \mathbb{E}\left(W^2\right) - (\mathbb{E}(W))^2 \\ &= \left(\frac{\alpha-1}{\alpha-3} - \left(\frac{\alpha-1}{\alpha-2}\right)^2\right){W_0}^2 \\ &= \boxed{\tfrac{\alpha-1}{(\alpha-2)^2(\alpha-3)}{W_0}^2} \end{aligned}$$

aussitôt que $\alpha > 3$.
En toute rigueur, il faudrait encore prouver que lorsque $\alpha < 2$,
$\mathbb{E}(W^2) - (\mathbb{E}(W))^2 = (+\infty) - (+\infty)$ diverge effectivement, en écrivant

$$\begin{aligned}
\mathbb{E}\left(W^2\right) - (\mathbb{E}(W))^2 &= \int_{\mathbb{R}} W^2 f(W|\alpha)dW - \left(\int_{\mathbb{R}} W f(W|\alpha)dW\right)^2 \\
&= \lim_{M\infty} \left(\left(\int_{W_0}^M W^2 f(W|\alpha)dW\right) - \left(\int_{W_0}^M W f(W|\alpha)dW\right)^2\right) \\
&= \lim_{M\infty} \left(\left(\int_{W_0}^M W^2 f(W|\alpha)dW\right) - \left(\int_{W_0}^M W f(W|\alpha)dW\right)\left(\int_{W_0}^M \omega f(\omega|\alpha)d\omega\right)\right) \\
&= \lim_{M\infty} \left(\int_{W_0}^M \left(W^2 - W\left(\int_{W_0}^M \omega f(\omega|\alpha)d\omega\right)\right) f(W|\alpha)dW\right) \\
&= \lim_{M\infty} \left(\int_{W_0}^M \left(W^2 - W\left[\frac{W^{2-\alpha}}{2-\alpha}\right]_{W_0}^M\right) f(W|\alpha)dW\right) \\
&= \lim_{M\infty} \left(\int_{W_0}^M W\left(W - \frac{M^{2-\alpha}}{2-\alpha} + \frac{{W_0}^{2-\alpha}}{2-\alpha}\right)(\alpha-1){W_0}^{\alpha-1}W^{-\alpha}dW\right) \\
&= (\alpha-1){W_0}^{\alpha-1} \lim_{M\infty} \left(\int_{W_0}^M W^{2-\alpha} - \left(\frac{M^{2-\alpha}}{2-\alpha} + \frac{{W_0}^{2-\alpha}}{2-\alpha}\right) W^{1-\alpha}dW\right) \\
&\underset{M\infty}{\approx} (\alpha-1){W_0}^{\alpha-1}\left(\frac{M^{3-\alpha}}{3-\alpha} - \frac{M^{4-2\alpha}}{(2-\alpha)^2}\right) \\
&\underset{M\infty}{\longrightarrow} \pm\infty \quad \text{selon le signe de } \alpha - 1
\end{aligned}$$

Ainsi les deux premiers moments existent *ssi* $\alpha = Xb + 1 > 3$, condition supposée vérifiée par la suite.

(b) $\boxed{\text{Déterminer l'estimateur du maximum de vraisemblance } \widehat{\alpha} \text{ de } \alpha.}$

On a pour $(w_1, \ldots, w_n) \in [W_0, +\infty[^{\,n}$

$$\begin{aligned}
L_{W_1,\ldots,W_n}(w_1, \ldots, w_n; \alpha) &= \prod_{i=1}^n \frac{\alpha-1}{W_0}\left(\frac{W_0}{w_i}\right)^\alpha \\
&= (\alpha-1)^n {W_0}^{n(\alpha-1)} \frac{1}{\prod_{i=1}^n {w_i}^\alpha}
\end{aligned}$$

Par suite l'estimateur du maximisation de vraisemblance $\widehat{\alpha}$ de α vérifie

$$\frac{n}{\widehat{\alpha}-1} - \sum_{i=1}^n \ln \frac{w_i}{W_0} = 0$$

de sorte que

$$\boxed{\widehat{\alpha} = 1 + \frac{n}{\sum_{i=1}^{n} \ln \frac{w_i}{W_0}}}$$

(et on vérifie que cette condition est bien suffisante)

(c) Déterminer le score et l'information de Fisher du modèle.
En déduire $\mathbb{E}(\ln W)$ et $\mathbb{E}((\ln W)^2)$.

On a pour une observation $w \in [W_0, +\infty[$

$$\begin{aligned} S_1(w;\alpha) &= \frac{\partial \ln L_{W_1}}{\partial \alpha}(w;\alpha) \\ &= \frac{1}{\alpha - 1} - \ln \frac{w}{W_0} \end{aligned}$$

Or le score est centré, ce qui s'écrit $\mathbb{E}\left(\frac{1}{\alpha-1} - \ln \frac{W}{W_0}\right) = 0$, soit

$$\boxed{\mathbb{E}(\ln W) = \frac{1}{\alpha-1} + \ln W_0}$$

De façon similaire on a

$$\begin{aligned} I_1(\alpha) &= -E\frac{\partial^2 \ln L_{W_1}}{\partial \alpha^2}(w;\alpha) \\ &= \frac{1}{(\alpha-1)^2} \end{aligned}$$

Or $I_1(\alpha) = \mathbb{V}(S_1(W;\alpha)) = \mathbb{E}(S_1(W;\alpha)^2)$, donc finalement

$$\mathbb{V}(\ln W) = \frac{1}{(\alpha-1)^2}$$

et

$$\boxed{\mathbb{E}\left((\ln W)^2\right) = \frac{1}{(\alpha-1)^2} + \left(\frac{1}{\alpha-1} + \ln W_0\right)^2}$$

(d) Soit $s = \frac{1}{n}\sum_{i=1}^{n}\left(\ln(W_i) - \overline{\ln W}\right)^2$, où $\overline{\ln W}$ est la moyenne empirique des logarithmes de W_i.
On se propose de fonder un test de H_0 sur la différence :

$$s - \frac{1}{\left(\widehat{\alpha} - 1\right)^2}$$

Justifier un tel test.

Notons $Y_i = \ln W_i$ pour $i \in [\![1, n]\!]$; alors

$$s = \frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \bar{Y}\right)^2$$

est la variance empirique de $(\ln W_1, \ldots, \ln W_n)$.
Or si H_0 est vraie, alors $\mathbb{V}(\ln W_i) = \frac{1}{(\alpha-1)^2}$ dont un estimateur convergent est $\frac{1}{(\widehat{\alpha}-1)^2}$, d'où l'idée de construire une statistique de test fondée sur $s - \frac{1}{(\widehat{\alpha}-1)^2}$.
Un tel test est habituellement appelé *test d'indépendance de Fisher.*

(e) Ecrire le Théorème Central Limite pour $\left(\ln W_i, (\ln W_i)^2\right)_{i\in[\![1,N]\!]}$, et tester l'hypothèse H_0 (On donnera la forme de la matrice de variance-covariance sans pour autant la calculer explicitement).

On a

$$\sqrt{n}\left(\begin{pmatrix} \overline{\ln W} \\ \overline{(\ln W)^2} \end{pmatrix} - \begin{pmatrix} \mathbb{E}(\ln W) \\ \mathbb{E}\left((\ln W)^2\right) \end{pmatrix}\right) \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \Omega\right)$$

avec $\Omega = \begin{pmatrix} \mathbb{V}(\ln W) & \mathbb{C}ov\left(\ln W, (\ln W)^2\right) \\ \mathbb{C}ov\left(\ln W, (\ln W)^2\right) & \mathbb{V}\left((\ln W)^2\right) \end{pmatrix}$ que l'on n'explicitera pas davantage.

Soit alors $g : \begin{pmatrix} \mathbb{R}^2 & \to & \mathbb{R} \\ (m, v) & \mapsto & v - m^2 \end{pmatrix}$; alors d'après le théorème de Slutsky

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

avec $M(\alpha) = \frac{\partial g}{\partial \alpha}(\alpha)' \times \Omega \times \frac{\partial g}{\partial \alpha}(\alpha) \in \mathbb{R}^+$.
Enfin, en considérant l'estimateur convergent $\widehat{M(\alpha)} = M(\widehat{\alpha})$ de $M(\alpha)$ on en tire que

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

de sorte que finalement si H_0 est vraie

$$\boxed{\xi_n^F = n\frac{\left(s - \frac{1}{(\alpha-1)^2}\right)^2}{M(\widehat{\alpha})} \xrightarrow[n\to+\infty]{\mathcal{L}} \chi_1^2}$$

☞ Q2 On suppose toujours W_0 connu, mais les X_i prennent désormais des valeurs a priori distinctes.

(a) Calculer l'estimateur du maximum de vraisemblance de b.

On a pour $(w_1, \ldots, w_n, x_1, \ldots, x_n) \in [W_0, +\infty[^{2n}$

$$\begin{aligned} L_{W_1,\ldots,W_n|X_1=x_1,\ldots,X_n=x_n}(w_1, \ldots, w_n, x_1, \ldots, x_n; b) &= \prod_{i=1}^{n} \frac{bx_i}{W_0} \left(\frac{W_0}{w_i}\right)^{bx_i+1} \\ &= b^n \left(\prod_{i=1}^{n} x_i\right) {W_0}^{b\sum_{i=1}^{n} x_i} \frac{1}{\prod_{i=1}^{n} {w_i}^{bx_i+1}} \end{aligned}$$

Par suite l'estimateur du maximum de vraisemblance $\widehat{b}$ de b vérifie

$$\frac{n}{\widehat{b}} + \sum_{i=1}^{n} x_i \ln W_0 - \sum_{i=1}^{n} x_i \ln w_i = 0$$

ce dont on tire

$$\boxed{\widehat{b} = \frac{n}{\sum_{i=1}^{n} x_i \ln \frac{w_i}{W_0}}}$$

(et on vérifie que cette condition est bien suffisante)

(b) Montrer que : $\mathbb{E}\left(X^2 \left(\ln \frac{W_0}{W}\right)^2 - \frac{2X}{b} \ln \frac{W_0}{W}\right) = 0.$
Comment testeriez-vous alors l'hypothèse H_0 ?

Le score s'écrit cette fois pour une observation

$$\begin{aligned} S_1(w|X = x; b) &= \frac{\partial \ln L_{W_1|X=x}}{\partial b}(w; b) \\ &= \frac{1}{b} - x \ln \frac{w}{W_0} \end{aligned}$$

Il est centré, et donc

$$\boxed{\mathbb{E}\left(X \ln \frac{W}{W_0}\right) = \frac{1}{b}}$$

Par ailleurs l'information de Fisher s'écrit

$$\begin{aligned} I_1(b) &= -E\frac{\partial^2 \ln L_{W_1|X_1}}{\partial b^2}(w; b) \\ &= \frac{1}{b^2} \end{aligned}$$

Or

$$\begin{aligned}
I_1(b) &= \mathbb{V}\left(S_1(W|X;b)\right) \\
&= \mathbb{E}\left(S_1(W|X;b)^2\right) \\
&= \mathbb{E}\left(\left(\frac{1}{b} - X \ln \frac{W}{W_0}\right)^2\right) \\
&= \mathbb{E}\left(\frac{1}{b^2} - 2\frac{X}{b}\ln\frac{W}{W_0} + X^2\left(\ln\frac{W}{W_0}\right)^2\right) \\
&= \frac{1}{b^2} + \mathbb{E}\left(X^2\left(\ln\frac{W}{W_0}\right)^2 - 2\frac{X}{b}\ln\frac{W}{W_0}\right)
\end{aligned}$$

Par conséquent,

$$\boxed{\mathbb{E}\left(X^2\left(\ln\tfrac{W}{W_0}\right)^2 - 2\tfrac{X}{b}\ln\tfrac{W}{W_0}\right) = 0}$$

De façon similaire au cas où les X_i étaient constants, on cherche alors à tester H_0 en considérant l'estimateur empirique de cette expression, à savoir

$$\frac{1}{n}\sum_{i=1}^{n} {x_i}^2\left(\ln\frac{w_i}{W_0}\right)^2 - \frac{1}{b}\frac{2}{n}\sum_{i=1}^{n} x_i \ln\frac{w_i}{W_0}$$

Il suffit alors, en s'appuyant sur la loi asymptotique de $\begin{pmatrix} X\ln\frac{W}{W_0} \\ X^2\left(\ln\frac{W}{W_0}\right)^2\end{pmatrix}$ obtenue par le Théorème Central Limite, d'appliquer le théorème de Slutsky en considérant $g : \begin{pmatrix} \mathbb{R}^2 & \rightarrow & \mathbb{R} \\ (m,v) & \mapsto & v - \frac{2}{b}m \end{pmatrix}$, ce qui conduit à

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

où $t = \frac{1}{n}\sum_{i=1}^{n} {x_i}^2\left(\ln\frac{w_i}{W_0}\right)^2 - \frac{2}{n^2}\left(\sum_{i=1}^{n} x_i\ln\frac{w_i}{W_0}\right)^2$ (en substituant $\widehat{b}$ à b)
et $N(b) = \frac{\partial g}{\partial b}(b)' \times \mathbb{V}\left(\begin{pmatrix} X\ln\frac{W}{W_0} \\ X^2\left(\ln\frac{W}{W_0}\right)^2\end{pmatrix}\right) \times \frac{\partial g}{\partial b}(b)$.
On en tire en définitive la statistique de test

$$\xi_n^F = n\frac{\left(t - \frac{1}{b^2}\right)^2}{N(\widehat{b})} \xrightarrow[n\to+\infty]{\mathcal{L}} \chi_1^2$$

si H_0 est vraie.

☞ Q3 On suppose enfin que W_0 est inconnu.

(a) Calculer l'estimateur du maximum de vraisemblance de W_0.

On a pour $(w_1, \ldots, w_n, x_1, \ldots, x_n) \in \mathbb{R}^{2n}$

$$L_{W_1,\ldots,W_n|X_1=x_1,\ldots,X_n=x_n}(w_1, \ldots, w_n; W_0, b) = b^n \left(\prod_{i=1}^{n} \frac{x_i}{(w_i)^{bx_i+1}} \right) {W_0}^{b\sum_{i=1}^{n} x_i} \mathbb{1}_{\min_i w_i \geq W_0}$$

En particulier l'estimateur du maximum de vraisemblance $\widehat{W_0}$ de W_0 est

$$\boxed{\widehat{W_0} = \min_i w_i}$$

(b) Peut-on adapter la procédure de test précédemment mise-en-œuvre ?

La difficulté est que le modèle statistique n'est plus régulier ; en particulier, la log-vraisemblance n'est plus dérivable, donc ni le score ni la matrice d'information de Fisher ne sont définis. La procédure de test utilisée précédemment, fondées sur leurs propriétés statistiques, n'est donc plus applicable.

8 Travaux Dirigés n°8

Exercice corrigé 1

Soit un échantillon $X_1, \ldots, X_n$ de variables aléatoires i.i.d telles que :
– X_i a une probabilité α de valoir a, et
– une probabilité $(1-\alpha)$ de suivre une loi normale $\mathcal{N}(0,1)$,
ce qui s'écrit encore :

$$X_1 = a \cdot \mathbb{1}_{Z_1=1} + Y_1 \cdot \mathbb{1}_{Z_1=0}$$

où $Z_1 \rightsquigarrow \mathcal{B}(1,\alpha)$ est indépendante de $Y_1 \rightsquigarrow \mathcal{N}(0,1)$.

☞ Q1 Montrer que $(\delta_a + \lambda)^{\otimes n}$ est une mesure dominante pour le modèle considéré.

Posons $\mu = \delta_a + \lambda$ et $\mu_n = \mu^{\otimes n}$.

Remarque : nous restreindrons les démonstrations au cas où une seule observation X_1 est disponible, dans la mesure où car la généralisation à n observations i.i.d. se réduit à un passage au produit immédiat :

$$dP_{X_1,\ldots,X_n} = \prod_{i=1}^{n} dP_{X_i}$$

$$\frac{dP_{X_1,\ldots,X_n}}{d\mu_n} = \prod_{i=1}^{n} \frac{dP_{X_i}}{d\mu}$$

Remarque préliminaire

Pour $\mathcal{A}$ ensemble mesurable, on a

$$\begin{aligned}\mathbb{P}(X_1 \in \mathcal{A}) &= \mathbb{P}(X_1 \in \mathcal{A}|X_1 = a)\mathbb{P}(X_1 = a) + \mathbb{P}(X_1 \in \mathcal{A}|X_1 \neq a)\mathbb{P}(X_1 \neq a)\\ &= \alpha\mathbb{P}(X_1 \in \mathcal{A}|X_1 = a) + (1-\alpha)\int_{\mathcal{A}} \varphi(x)d\lambda(x)\end{aligned}$$

où φ désigne la densité de la loi normale centrée réduite.

Or $\mathbb{P}(X_1 \in \mathcal{A}|X_1 = a)$ vaut 1 si $a \in \mathcal{A}$ et 0 sinon, de sorte que

$$\mathbb{P}(X_1 \in \mathcal{A}) = \alpha\mathbb{1}_{a\in\mathcal{A}} + (1-\alpha)\int_{\mathcal{A}} \varphi(x)d\lambda(x) \tag{2}$$

Or μ est une mesure dominante d'une variable X_1 ssi :

$$\forall \mathcal{A} \text{ mesurable, } \mu(\mathcal{A}) = 0{:}P_{X_1}(\mathcal{A}) = 0$$

En l'occurence pour tout $\mathcal{A}$ mesurable :
– $\mu(\mathcal{A}) = 0{:}\delta_a(\mathcal{A}) = 0$
– $\delta_a(\mathcal{A}) = 0{:}a \notin \mathcal{A}$, donc d'après 2 $\lambda(\mathcal{A}) = 0{:}P_{X_1}(\mathcal{A}) = 0$

Donc $(\delta_a + \lambda)^{\otimes n}$ est une mesure dominante.

☞ Q2

> Montrer que :
>
> $$\frac{dP_{X_1,\ldots,X_n}(x_1,\ldots,x_n)}{d(\delta_a+\lambda)^{\otimes n}} = \prod_{i=1}^{n}(\alpha \mathbb{1}_a(x_i) + (1-\alpha)(1-\mathbb{1}_a(x_i))\phi(x_i))$$
>
> où ϕ est la densité d'une loi normale $\mathcal{N}(0,1)$.

On a $\mathbb{1}_{a\in\mathcal{A}} = \int_{\mathcal{A}} \delta_a(x)$,
donc en reportant dans l'équation 2 il vient

$$P(X_1 \in \mathcal{A}) = \int_{\mathcal{A}} (\alpha\delta_a(x) + (1-\alpha)\varphi(x)d\lambda(x))$$

d'où on tire facilement

$$P(X_1 \in \mathcal{A}) = \int_{\mathcal{A}} (\alpha\mathbb{1}_a(x) + (1-\alpha)\varphi(x)(1-\mathbb{1}_a(x)))\, d\mu(x)$$

de sorte qu'en définitive

$$\frac{dP_{X_1}(x_1)}{d(\delta_a+\lambda)} = \alpha\mathbb{1}_a(x_1) + (1-\alpha)(1-\mathbb{1}_a(x_1))\phi(x_1)$$

d'où le résultat par indépendance de $(X_1,\ldots,X_n)$.

Exercice corrigé 2

Cet exercice s'inspire librement des travaux de Dov Samet, 2003.[11]

On s'intéresse pour tout couple (X,Y) de variables aléatoires réelles à la propriété $\mathcal{P}$ suivante :

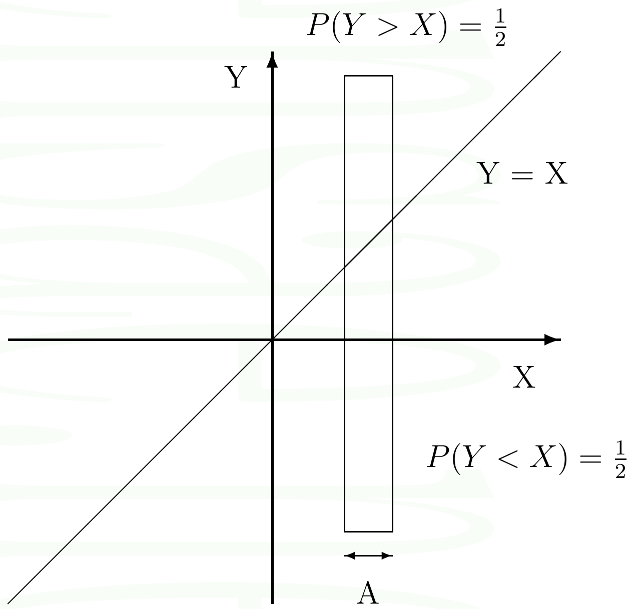
$$(\mathcal{P})\ \ \forall\mathcal{A} \text{ mesurable}, \begin{cases} \mathbb{P}(X<Y \text{ et } X\in\mathcal{A}) = \mathbb{P}(X>Y \text{ et } X\in\mathcal{A}) = \frac{1}{2}\mathbb{P}(X\in\mathcal{A}) \\ \\ \mathbb{P}(Y<X \text{ et } Y\in\mathcal{A}) = \mathbb{P}(Y>X \text{ et } Y\in\mathcal{A}) = \frac{1}{2}\mathbb{P}(Y\in\mathcal{A}) \end{cases}$$

☞ Q1

> Comment s'interprête la propriété $\mathcal{P}$?

Si un couple de variables aléatoires réelles (X,Y) vérifie $\mathcal{P}$, alors au voisinage de tout X, la probabilité est la même pour Y d'être au-dessus ou au-dessous de X, comme illustré sur la figure suivante :

[11] Résultats présentés à la 14$^{\text{ième}}$ conférence internationale de Théorie des jeux, Stony Brook 2003. Voir `Http ://www.sunysb.edu/gametheory/Conf03/twopuzzles.pdf` pour un exposé plus complet.



■ Partie 1 "***Devinez quel est le plus grand !***"

On s'intéresse au problème suivant, présenté originellement par Blackwell (1951) :

> Deux nombres réels sont tirés aléatoirement, et chacun est placé dans une enveloppe. L'une d'elle est (indépendamment) tirée au hasard et vous est présentée. Vous devez deviner, au vu du nombre qu'elle contient, s'il s'agit du plus grand ou le plus petit des deux. Sauriez-vous deviner juste plus d'une fois sur deux en moyenne ?

Pour simplifier la présentation, supposons que vous gagnez +1 si vous avez deviné juste et −1 sinon ; ce problème revient à savoir si vous pouvez vous garantir un gain espéré strictement positif.

☞ Q1 Soient X et Y les deux nombres tirés aléatoirement, et supposons que (X, Y) vérifie $\mathcal{P}$. Quel est votre meilleur gain espéré ? Interprêter.

Soit $\sigma : \mathbb{R} \to \{-1, +1\}$ votre stratégie, et soient (x, y) une réalisation du tirage.
- Si x vous est présenté :
  Si $x > y$, votre gain est $\sigma(x)$; si $x < y$, votre gain est $-\sigma(x)$. Votre gain espéré, sachant que x vous est présenté, est donc nul.
- Si y vous est présenté, il en va de même.

Le choix de l'enveloppe étant indépendant du tirage de X et de Y, votre gain espéré est toujours nul, quelle que soit votre stratégie.

Comme l'hypothèse $\mathcal{P}$ semble "raisonnable", l'intuition suggère qu'à ce jeu vous ne pourrez jamais obtenir un gain espéré strictement positif. Cette intuition est cependant erronée.

☞ Q2 Cherchons néanmoins à construire une stratégie gagnante.

Pour ce faire, donnons-nous $s \in \mathbb{R}$ et soit $\sigma_s : \begin{pmatrix} \mathbb{R} & \to & \{-1,1\} \\ x & \mapsto & \begin{cases} 1 & \text{si } x > s \\ -1 & \text{sinon} \end{cases} \end{pmatrix}$ la *stratégie de seuil* associée à s.

(a) Soit $(x \neq y)$ une réalisation du tirage ; le nombre qui vous est présenté est soit x, soit y. Quel est votre gain espéré en jouant selon σ_s sachant que $x < s$ et $y < s$?

On a dans tous les cas $\sigma(x) = \sigma(y) = -1$ car $x < s$ et $y < s$.
Or si $x < y$, le gain espéré est

$$\mathbb{P}(\text{le nombre présenté est } x) \cdot 1 + \mathbb{P}(\text{le nombre présenté est } y) \cdot (-1) = 0$$

puisque ces deux alternatives sont équiprobables et que le choix de l'enveloppe est indépendant du tirage de X et de Y.
Il en va de même lorsque $x > y$, de sorte que le gain espéré, sachant que $x < s$ et $y < s$, est nul.

(b) Quel est votre gain espéré sachant que $x < s < y$?

Si le nombre présenté est x, on a $\sigma(x) = -1$ car $x < s$; or $x < y$, donc la réponse est juste et le gain est $+1$.
Si le nombre présenté est y, on a $\sigma(y) = +1$ car $y > s$; or $y > x$, donc la réponse est encore juste et le gain est toujours $+1$.
Ainsi dans tous les cas le gain espéré, sachant que $x < s < y$, est $+1$. Il en va de même bien-sûr sachant que $y < s < x$.

(c) En déduire que si votre gain espéré (sur tous les tirages possibles) est strictement positif.

D'après les résultats précédents le gain espéré en jouant σ_s est donc

$$\boxed{g(s) = \mathbb{P}(X < s < Y) + \mathbb{P}(Y < s < X)}$$

Or ces deux probabilités ne peuvent être simultanément nulles pour tout s : en effet

$$\begin{aligned}
\int_{\mathbb{R}} \mathbb{P}(X < s < Y) + \mathbb{P}(Y < s < X) ds &= \int_{\mathbb{R}} \left(\int_{\mathbb{R}^2} \left(\mathbb{1}_{x<s<y} + \mathbb{1}_{y<s<x}\right) f_{X,Y}(x,y) dxdy \right) ds \\
&= \int_{\mathbb{R}^2} \left(\int_{\mathbb{R}} \mathbb{1}_{x<s<y} ds + \int_{\mathbb{R}} \mathbb{1}_{y<s<x} ds \right) f_{X,Y}(x,y) dxdys \\
&= \int_{\mathbb{R}^2} \left(\mathbb{1}_{x<y}(y-x) + \mathbb{1}_{y<x}(y-x)\right) f_{X,Y}(x,y) dxdys \\
&= \int_{\mathbb{R}^2} |x-y| \, \mathbb{1}_{x \neq y} f_{X,Y}(x,y) dxdys \\
&= \mathbb{E}\left(|X-Y|\right)
\end{aligned}$$

Si $\mathbb{P}(X \neq Y) = 0$ alors $\mathbb{P}(X = Y) = 1$ et le jeu devient trivial : la stratégie qui prétend que le nombre tiré est le plus grand est presque sûrement toujours gagnante ; en particulier lorsque $s = 0$ $g(0) > 0$.
Et si $\mathbb{P}(X \neq Y) \neq 0$ alors $\mathbb{E}(|X - Y|) > 0$, et donc $\int_{\mathbb{R}} g(s)ds > 0$.
Donc dans tous les cas

$$\boxed{\exists s \in \mathbb{R} \ / \ g(s) > 0}$$

En fait, si (X, Y) est distribué de telle sorte que toute partie de mesure non-nulle a une probabilité non-nulle, alors *n'importe quel* seuil s assure un gain espéré strictement positif.

☞ Q3 En déduire qu'aucune paire de variables aléatoires réelles (X, Y) ne peut vérifier $\mathcal{P}$.

Si une telle distribution sur $\mathbb{R}^2$ existait, alors d'après la question (1) dans le jeu de Blackwell associé nul ne saurait gagner plus que zéro en moyenne, ce qui est pourtant possible d'après la question (2).

Partie 2 ***Le jeu des enveloppes***

On s'intéresse désormais au célèbre jeu dit *des enveloppes*, introduit par Kraitchik (1953) sous la forme suivante :

> On considère deux enveloppes contenant une certaine somme d'argent, mais l'une contenant le double de l'autre ; chacune a autant de chances que l'autre de contenir la plus grosse somme. L'une d'elles vous est donnée au hasard, et vous devez choisir entre la prendre, ou prendre l'autre.
> A supposer qu'elle contienne x, l'autre a une chance sur deux de contenir $2x$ et une chance sur deux de contenir $\frac{x}{2}$; donc le contenu espéré de l'autre enveloppe est $\frac{5}{4}x > x$, de sorte que vous avez intérêt à changer d'enveloppe. Mais le même raisonnement s'applique aussi à la nouvelle enveloppe, de sorte que vous avez à nouveau intérêt à changer d'enveloppe, et ainsi de suite ...
> Que faire dans cette situation ?

Résoudre ce paradoxe a été le prétexte à une littérature foisonnante ; voici une solution.

Notons X la somme (variable aléatoire réelle) contenue dans l'enveloppe qui vous est présentée, et Y celle contenue dans l'autre.

☞ Q1 Quel sens donner à l'expression $\mathcal{H}_x$: "l'autre a une chance sur deux de contenir $2x$ et une chance sur deux de contenir $\frac{x}{2}$" ?

La formulation du paradoxe suggère que l'on se donne une distribution de probabilité a priori sur les deux sommes X et Y ; cette distribution est nécessairement telle que $\mathbb{P}(X = 2Y \text{ ou } Y = 2X) = 1$.

Etant donné $x \in \mathbb{R}$ et à supposer que $X = x$, l'hypothèse $\mathcal{H}_x$ s'exprime en termes mathématiques sous la forme

$$\mathbb{P}(Y = 2x \,|\, X = x) = \mathbb{P}\left(Y = \frac{x}{2} \,|\, X = x\right) = \frac{1}{2}$$

☞ Q2 $\boxed{\text{En déduire que si cette hypothèse est vraie, alors la distribution de } (X,Y) \text{ vérifie } \mathcal{P}.}$

L'hypothèse $\mathcal{H}_y$ analogue à la précédente en $Y = y \in \mathbb{R}$ s'écrit

$$\mathbb{P}(X = 2y \,|\, Y = y) = \mathbb{P}\left(X = \frac{y}{2} \,|\, Y = y\right) = \frac{1}{2}$$

Dans l'énoncé du paradoxe, l'hypothèse selon laquelle “chacune a autant de chances que l'autre de contenir la plus grosse somme” se traduit alors en $\mathcal{H}_x$ quelle que soit la somme x de la première enveloppe, et $\mathcal{H}_y$ quelle que soit la somme y de la deuxième enveloppe.
En d'autres termes

$$\begin{array}{l} \forall x \in \mathbb{R},\ \mathbb{P}(Y = 2x \,|\, X = x) = \mathbb{P}\left(Y = \frac{x}{2} \,|\, X = x\right) = \frac{1}{2} \\ \forall y \in \mathbb{R},\ \mathbb{P}(X = 2y \,|\, Y = y) = \mathbb{P}\left(X = \frac{y}{2} \,|\, Y = y\right) = \frac{1}{2} \end{array}$$

Or pour tout $x \in \mathbb{R}$ l'événement $Y = 2x$ est égal à l'événement $Y > x$, et l'événement $Y = \frac{x}{2}$ à $Y < x$. Par suite

$$\begin{array}{l} \forall x \in \mathbb{R},\ \mathbb{P}(Y > x \,|\, X = x) = \mathbb{P}(Y < x \,|\, X = x) = \frac{1}{2} \\ \forall y \in \mathbb{R},\ \mathbb{P}(X > y \,|\, Y = y) = \mathbb{P}(X < y \,|\, Y = y) = \frac{1}{2} \end{array}$$

Par conséquent pour toute partie mesurable $\mathcal{A}$ de $\mathbb{R}$ telle que $\mathbb{P}(X \in A) \neq 0$ et $\mathbb{P}(Y \in \mathcal{A}) \neq 0$ il vient

$$\begin{array}{l} \mathbb{P}(Y > X | X \in \mathcal{A}) = \mathbb{P}(Y < X | X \in \mathcal{A}) = \frac{1}{2} \\ \mathbb{P}(X > Y | Y \in \mathcal{A}) = \mathbb{P}(X < Y | Y \in \mathcal{A}) = \frac{1}{2} \end{array}$$

de sorte que (X, Y) vérifie la propriété $\mathcal{P}$.

☞ Q3 $\boxed{\text{Conclure.}}$

Ainsi, à supposer que l'on se donne une distribution a priori sur les sommes d'argent de chacune des deux enveloppes, il n'est pas possible de traduire l'hypothèse “chacune a autant de chances que l'autre de contenir la plus grosse somme” comme il est suggéré, car la distribution vérifierait alors la propriété $\mathcal{P}$ ce qui n'est pas possible.

La démarche proposée, bien qu'intuitive, n'est donc pas raisonnable : en particulier il est faux qu'après un nombre quelconque de changements d'enveloppe on ait “à nouveau intérêt à changer d'enveloppe”.

Exercice corrigé 3

Soient $X_1, \ldots, X_n$ n variables aléatoires i.i.d dans $\mathbb{R}^k$, de loi paramétrique $\mathcal{P}_\theta$ de densité

$$\frac{d\mathcal{P}_\theta(x)}{d\lambda} = f(x-\theta)$$

$\theta \in \mathbb{R}^k$ est appelé *paramètre de position*.
On cherche à étudier les estimateurs T du paramètre θ. On se restreint pour cela à la classe des estimateurs *équivariants*, c'est-à-dire des estimateurs T vérifiant :

$$\forall c \in \mathbb{R}^k, T(X_1+c, \ldots, X_n+c) = T(X_1, \ldots, X_n) + c$$

Cette restriction est naturelle : en effet, pour tout $c \in \mathbb{R}^k$, $X_i + c \rightsquigarrow \mathcal{P}_{\theta+c}$, et on ne peut admettre qu'un simple changement d'échelle puisse mener à une estimation différente.

☞ Q1 Montrer que :

$$\text{T équivariant} \Leftrightarrow \begin{array}{l} \exists T_1 : (\mathbb{R}^{k-1} \to \mathbb{R}^k) \ / \ \forall (x_1, \ldots, x_n) \in \mathbb{R}^k \\ T(x_1, \ldots, x_n) = T_1(x_2 - x_1, \ldots, x_n - x_1) + x_1 \end{array}$$

Il suffit de poser $c = -x_1$ dans la définition de l'équivariance.

☞ Q2 Soit $W(x,y) = w(x-y)$ une fonction de perte. Montrer que, si un estimateur T est équivariant, le risque associé, pour la perte W, ne dépend pas de θ.
Un estimateur équivariant T qui minimise le risque $R_w(T,0)$ (c'est-à-dire au point $\theta = 0$) est alors un estimateur optimal (parmi les équivariants) pour la fonction de perte W.

Le risque associé vérifie :

$$\begin{aligned} R_w(T,\theta) &= E_\theta[w(T-\theta)] \\ &= \int_{(\mathbb{R}^k)^n} w(T(x_1, \ldots, x_n) - \theta) \prod_{\iota=1}^n f(x_i - \theta) dx_i \\ &= \int_{(\mathbb{R}^k)^n} w(T(x_1 - \theta, \ldots, x_n - \theta)) \prod_{\iota=1}^n f(x_i - \theta) dx_i \\ &= \int_{(\mathbb{R}^k)^n} w(T(u_1, \ldots, u_n)) \prod_{\iota=1}^n f(u_i) du_i \\ &= R_w(T,0) \end{aligned}$$

en effectuant un changement de variable $u_i = x_i - \theta$.

☞ Q3 Soit la fonction

$$\psi(x) = \int_{\mathbb{R}^k} w(x-u) \prod_{i=1}^n f(X_i - u) du$$

et l'estimateur T^* défini par

$$\psi(T^*) = \inf_{x \in \mathbb{R}^k} \psi(x)$$

T^* est appelé *l'estimateur de Pitman* (on suppose ici que w est telle que l'équation $\psi(x^*) = \inf_{x\in\mathbb{R}^k} \psi(x)$ admette une solution et une seule).

(a) $\boxed{\text{Montrer que } T^* \text{ est équivariant.}}$

Soit $c \in \mathbb{R}^k$; on a

$$\begin{aligned}\psi(x+c) &= \int w(x+c-u)\prod_{i=1}^n f(X_i-u)du \\ &= \int w(x-v)\prod_{i=1}^n f(X_i-c-v)dv\end{aligned}$$

en effectuant le changement de variable $v = u - c$.
En particulier $\psi(x+X_1) = \int w(x-v)f(-v)\prod_{i=2}^n f\left((X_i-X_1)-v\right)dv$
Donc $\psi(x+X_1)$ ne dépend que de x et des variables aléatoires $X_2-X_1,\ldots,X_n-X_1$: soit donc T_1 tel que $\arg\min_{x\in\mathbb{R}^k}\psi(x+X_1) = T_1(X_2-X_1,\ldots,X_n-X_1)$.
Or $\forall\phi, \forall c\in\mathbb{R}^k,\quad \arg\inf\left(x\mapsto\phi(x+c)\right) = \left(\arg\inf\left(x\mapsto\phi(x)\right)\right)-c$.
Donc en définitive

$$T_1(X_2-X_1,\ldots,X_n-X_1) = (\arg\inf\psi) - X_1 = T^* - X_1$$

Autrement dit T^* est équivariant d'après la question précédente.

(b) Montrer que, pour toute fonction φ telle que $\mathbb{E}\left(\|\varphi(X_1,\ldots,X_n)\|\right) < +\infty$, on a :

$$E_\theta[\varphi(X_1,...,X_n)|X_2-X_1,...,X_n-X_1] = \int_{\mathbb{R}^k}\varphi(X_1+\theta-u,...,X_n+\theta-u)\frac{\prod_{i=1}^n f(X_i-u)}{\int_{\mathbb{R}^k}\prod_{i=1}^n f(X_i-v)dv}du$$

$$\begin{aligned}&E_\theta[\varphi(X_1,\ldots,X_n)|X_2-X_1=y_2,\ldots,X_n-X_1=y_n] \\ =\ &\int\varphi(x_1,x_1+y_2,\ldots,x_1+y_n)d\mathbb{P}(X_1=x_1|X_2-X_1=y_2,\ldots,X_n-X_1=y_n) \\ =\ &\int\varphi(x_1,x_1+y_2,\ldots,x_1+y_n)\frac{f(x_1-\theta)\prod_{i=2}^n f(x_1+y_i-\theta)}{\left[\int f(x_1-\theta)\prod_{i=2}^n f(x_1+y_i-\theta)dx_1\right]}dx_1\end{aligned}$$

Effectuons le changement de variables $u = \theta - x_1 + c$ (c constante quelconque) dans les deux intégrales :

$$\begin{aligned}&E_\theta[\varphi(X_1,\ldots,X_n)|X_2-X_1=y_2,\ldots,X_n-X_1=y_n] \\ =\ &\int\varphi(c+\theta-u,c+y_2+\theta-u,\ldots)\frac{f(c-u)\prod_{i=2}^n f(c+y_i-u)}{\left[\int f(c-u)\prod_{i=2}^n f(c+y_i-u)du\right]}du\end{aligned}$$

La relation suivante est alors vraie pour tout $c \in \mathbb{R}^k$:

$$\begin{aligned} & E_\theta[\varphi(X_1,\ldots,X_n)|X_2-X_1,\ldots,X_n-X_1] \\ = & \int \varphi(c+\theta-u, c+X_2-X_1+\theta-u,\ldots)\frac{f(c-u)\prod_{i=2}^n f(c+X_i-X_1-u)}{\left[\int f(c-u)\prod_{i=2}^n f(c+X_i-X_1-u)du\right]}du \end{aligned}$$

elle est donc vraie aussi pour $c = X_1$:

$$\begin{aligned} & E_\theta[\varphi(X_1,\ldots,X_n)|X_2-X_1,\ldots,X_n-X_1] \\ = & \int \varphi(X_1+\theta-u, X_2+\theta-u,\ldots)\frac{\prod_{i=1}^n f(X_i-u)}{\left[\int f(X_1-u)\prod_{i=2}^n f(X_i-u)du\right]}du \end{aligned}$$

(c) $\boxed{\text{Montrer finalement que } T^* \text{ est optimal pour } w.}$

Soit T un estimateur équivariant quelconque. On va montrer que T^* domine T. Pour cela, il suffit de prouver que $R_w(T,0) \geq R_w(T^*,0)$, selon le résultat de la question (2).

$$\begin{aligned} R_w(T,0) &= \mathbb{E}_0[w(T)] \\ &= \mathbb{E}_0[\mathbb{E}_0[w(T)|X_2-X_1,\ldots,X_n-X_1]] \\ &= \mathbb{E}_0\left[\int_{\mathbb{R}} w(T-u)\frac{\prod_{i=1}^n f(X_i-u)}{\int_{\mathbb{R}}\prod_{i=1}^n f(X_i-u)du}du\right] \end{aligned}$$

Or par définition de T^*, pour tous $x \in \mathbb{R}^n$ et $t \in \mathbb{R}$

$$\int_{\mathbb{R}} w(T^*(x)-u)\prod_{i=1}^n f(x_i-u)du \leq \int_{\mathbb{R}} w(t-u)\prod_{i=1}^n f(x_i-u)du$$

donc

$$\begin{aligned} R_w(T,0) &\geq \mathbb{E}_0\left[\int_{\mathbb{R}} w(T^*(X)-u)\frac{\prod_{i=1}^n f(X_i-u)}{\int_{\mathbb{R}}\prod_{i=1}^n f(X_i-u)du}du\right] \\ &= R_w(T^*,0) \end{aligned}$$

Donc T^* domine T.

☞ Q4 $\boxed{\text{Donner l'expression de } T^* \text{ lorsque } w(x-y) = \|x-y\|^2.}$

Si $w(x-y) = \|x-y\|^2$, ψ vaut :

$$\psi(x) = \int_{\mathbb{R}^k} \|x-u\|^2 \prod_{i=1}^n f(X_i-u)du$$

et le gradient $\psi'(x)$ vaut : $\psi'(x) = \int_{\mathbb{R}^k} 2(x-u)\ \prod_{i=1}^n f(X_i-u)du$. T^* est défini par l'équation $\psi'(T^*) = 0$, qui se réduit ici à :

$$T^* = \frac{\int_{\mathbb{R}^k} u\prod_{i=1}^n f(X_i-u)du}{\int_{\mathbb{R}^k}\prod_{i=1}^n f(X_i-u)du}$$

☞ Q5

> Dans ce cas, que vaut T^* lorsque $X_i \rightsquigarrow \mathcal{N}(\theta, 1)$?
>
> Et lorsque $X_i \rightsquigarrow \mathcal{U}_{[-\frac{1}{2}+\theta, \frac{1}{2}+\theta]}$?

Lorsque $X_i \sim \mathcal{N}(0,1)$, on a

$$T^* = \frac{1}{n}\sum_{i=1}^{n} X_i$$

Si $X_i \sim \mathcal{U}(0,1)$, il vient

$$T^* = \frac{\min(X_i) + \max(X_i)}{2}$$

Exercice corrigé 4

On considère une loi multinomiale à K modalités ($K \geq 3$) de probabilités $p_1, p_2, \ldots, p_K$. On dispose de n observations indépendantes issues de cette loi. On notera N_k le nombre d'observations de la modalité $k \in [\![1, K]\!]$.

On veut tester l'hypothèse : $\underline{H_0}$: "$p_1 + p_2 = \frac{1}{2}$"

☞ Q1

> Donner les estimateurs du maximum de vraisemblance contraints $(\widehat{p_k^0})_k$ et non contraints $(\widehat{p_k})_k$ de $p_1, \ldots, p_K$.

Les observations étant les effectifs N_k de chaque classe $k \in [\![1, K]\!]$, la vraisemblance s'écrit

$$L_{N_1,\ldots,N_K}(n_1, \ldots, n_K; p_1, \ldots, p_K) \quad = \quad \frac{n!}{n_1! \cdots n_K!}\prod_{k=1}^{K} {p_k}^{n_k} \quad \text{en notant } n = n_1 + \cdots + n_K$$

La maximisation sans hypothèse conduit aux conditions nécessaires du premier ordre, en n'omettant la contrainte $p_1 + \cdots + p_K = 1$

$$\begin{array}{l} \max_{p_1,\ldots,p_K} \left(\sum_{k=1}^{K} n_k \ln p_k\right) \\ \text{s.c. } \{p_1 + \cdots + p_K = 1 \end{array}$$

Le lagrangien associé s'écrit

$$\mathcal{L}(p_1, \ldots, p_K, \lambda) = \sum_{k=1}^{K} n_k \ln p_k + \lambda(1 - \sum_{k=1}^{K} p_k)$$

et il vient

$$\forall k \in [\![1, K]\!],\ \frac{n_k}{\widehat{p_k}} - \lambda = 0$$

Or en sommant de 1 à K il vient

$$n = \sum_{k=1}^{K} n_k = \lambda \sum_{k=1}^{K} p_k = \lambda$$

de sorte que finalement

$$\boxed{\forall k \in [\![1, K-1]\!],\ \widehat{p_k} = \frac{n_k}{n}}$$

(et on vérifie que cette condition est également suffisante).
La maximisation sous l'hypothèse H_0 correspond au programme de maximisation sous contrainte

$$\begin{array}{l} \max_{p_1,\ldots,p_K} \left(\sum_{k=1}^K n_k \ln p_k\right) \\ \text{s.c.} \left\{ \begin{array}{l} p_1 + \cdots + p_K = 1 \\ p_1 + p_2 = \frac{1}{2} \end{array} \right. \end{array}$$

Le lagrangien associé s'écrit

$$\mathcal{L}(p_1, \ldots, p_K, \lambda, \mu) = \sum_{k=1}^K n_k \ln p_k + \lambda(1 - \sum_{k=1}^K p_k) + \mu(\frac{1}{2} - p_1 - p_2)$$

et il vient

$$\left\{ \begin{array}{l} \frac{\partial \mathcal{L}}{\partial p_1} = \frac{n_1}{p_1} - \lambda - \mu \\ \frac{\partial \mathcal{L}}{\partial p_2} = \frac{n_2}{p_2} - \lambda - \mu \\ \frac{\partial \mathcal{L}}{\partial p_k} = \frac{n_k}{p_k} - \lambda, \quad \forall k \in [\![3, K]\!] \\ \frac{\partial \mathcal{L}}{\partial \lambda} = 1 - \sum_{k=1}^K p_k \\ \frac{\partial \mathcal{L}}{\partial \mu} = \frac{1}{2} - p_1 - p_2 \end{array} \right.$$

d'où finalement

$$\boxed{\left\{ \begin{array}{l} \widehat{p_1^0} = \frac{n_1}{2(n_1+n_2)} \\ \widehat{p_2^0} = \frac{n_2}{2(n_1+n_2)} \\ \widehat{p_k^0} = \frac{n_k}{2(n_3+\cdots+n_K)}, \quad \forall k \in [\![3, K]\!] \end{array} \right.}$$

☞ Q2 Calculer la statistique ξ^W du test de Wald de l'hypothèse H_0.

Pour déterminer la statistique de Wald, appliquons le Théorème Central Limite à $\begin{pmatrix} \widehat{p_1} \\ \vdots \\ \widehat{p_K} \end{pmatrix}$; pour ce faire, calculons tout d'abord sa variance.
On a en effet pour $k \in [\![1, K]\!]$, $\widehat{p_k} = \frac{n_k}{n} = \frac{1}{n}\sum_{i=1}^n \mathbb{1}_{X_i \in C_k}$ où C_k désigne la k-ième classe ; or

$$\mathbb{V}\left(\mathbb{1}_{X_i \in C_k}\right) = p_k \cdot 1 - (p_k)^2 = p_k(1 - p_k)$$

et

$$\mathbb{C}ov\left(\mathbb{1}_{X_i \in C_k}, \mathbb{1}_{X_j \in C_l}\right) = \mathbb{E}\left(\mathbb{1}_{X_i \in C_k} \cdot \mathbb{1}_{X_j \in C_l}\right) - \mathbb{E}\left(\mathbb{1}_{X_i \in C_k}\right)\mathbb{E}\left(\mathbb{1}_{X_j \in C_l}\right) = -p_k p_l \mathbb{1}_{i=j}$$

de sorte que

$$\sqrt{n}\begin{pmatrix} \widehat{p_1} - p_1 \\ \vdots \\ \widehat{p_K} - p_K \end{pmatrix} \xrightarrow[n \to +\infty]{\mathcal{L}} \mathcal{N}\left(0, \begin{pmatrix} p_1(1-p_1) & -p1p_2 & \cdots & -p_1 p_K \\ -p_2 p_1 & p_2(1-p_2) & \cdots & -p_2 p_K \\ \vdots & \vdots & \ddots & \vdots \\ -p_K p_1 & -p_K p_2 & \cdots & p_K(1-p_K) \end{pmatrix}\right)$$

Posons finalement $g : \begin{pmatrix} \mathbb{R}^K & \longrightarrow & \mathbb{R} \\ x_1, \ldots, x_K & \longmapsto & x_1 + x_2 \end{pmatrix}$; alors d'après le théorème de Slutsky, si H_0 est vraie on a

$$\begin{array}{rcl} \sqrt{n}\left(\widehat{p_1} + \widehat{p_2} - \frac{1}{2}\right) & \xrightarrow[n\to+\infty]{\mathcal{L}} & \mathcal{N}\left(0, \left(\frac{\partial g}{\partial p_1, \ldots, p_K}\right)' \mathbb{V}\left(\left(\begin{array}{c} \widehat{p_1} \\ \vdots \\ \widehat{p_K} \end{array}\right)\right)\left(\frac{\partial g}{\partial p_1, \ldots, p_K}\right)\right) \\ & = & \mathcal{N}\left(0, (p_1 + p_2)(1 - p_1 - p_2)\right) \end{array}$$

d'où en centrant et en réduisant, puis en élevant au carré

$$\boxed{\xi_n^W = n\frac{\left(\widehat{p_1}+\widehat{p_2}-\frac{1}{2}\right)^2}{(\widehat{p_1}+\widehat{p_2})(1-\widehat{p_1}-\widehat{p_2})} \xrightarrow[n\to+\infty]{\mathcal{L}} \chi_1^2}$$

☞ Q3

(a) En constatant que $\widehat{p_3^0}$ est asymptotiquement efficace sous H_0, montrer que

$$\mathbb{C}ov_{as}\left(\widehat{p_3^0}, \widehat{p_3} - \widehat{p_3^0}\right) = 0$$

et en déduire que

$$\mathbb{V}_{as}\left(\widehat{p_3} - \widehat{p_3^0}\right) = \mathbb{V}_{as}\left(\widehat{p_3}\right) - \mathbb{V}_{as}\left(\widehat{p_3^0}\right)$$

On sait que $\widehat{p_3^0} \xrightarrow[n\to+\infty]{p.s.} p_3$ et que $\widehat{p_3} \xrightarrow[n\to+\infty]{p.s.} p_3$.
Soit donc pour $\lambda \in \mathbb{R}$ $p_3(\lambda) = \lambda\widehat{p_3^0} + (1-\lambda)\widehat{p_3} \xrightarrow[n\to+\infty]{p.s.} p_3$.
Alors

$$\begin{array}{rcl} \mathbb{V}_{as}\left(p_3(\lambda)\right) & = & \mathbb{V}_{as}\left(\widehat{p_3^0} + \lambda\left(\widehat{p_3^0} - \widehat{p_3}\right)\right) \\ & = & \mathbb{V}_{as}\left(\widehat{p_3^0}\right) + 2\lambda\mathbb{C}ov_{as}\left(\widehat{p_3^0}, \widehat{p_3^0} - \widehat{p_3}\right) + \lambda^2\mathbb{V}_{as}\left(\widehat{p_3^0} - \widehat{p_3}\right) \\ & \geq & \mathbb{V}_{as}\left(\widehat{p_3^0}\right) \end{array}$$

car $\widehat{p_3^0}$ est asymptotiquement efficace, car sa variance est égale à la borne FDCR $I(p_3)^{-1}$.
Par conséquent le polynôme du second degré $P(\mathbb{X}) = \mathbb{V}_{as}\left(\widehat{p_3^0} - \widehat{p_3}\right)\mathbb{X}^2 + 2\mathbb{C}ov_{as}\left(\widehat{p_3^0}, \widehat{p_3} - \widehat{p_3^0}\right)\mathbb{X}$ est de signe constant. En particulier,

$$\mathbb{C}ov_{as}\left(\widehat{p_3^0}, \widehat{p_3^0} - \widehat{p_3}\right) = 0$$

et donc

$$\begin{array}{rcl} \mathbb{V}_{as}\left(\widehat{p_3}\right) & = & \mathbb{V}_{as}\left(\widehat{p_3^0} - \left(\widehat{p_3^0} - \widehat{p_3}\right)\right) \\ & = & \mathbb{V}_{as}\left(\widehat{p_3^0}\right) + \mathbb{V}_{as}\left(\widehat{p_3^0} - \widehat{p_3}\right) \end{array}$$

ce qui s'écrit encore

$$\boxed{\mathbb{V}_{as}\left(\widehat{p_3}-\widehat{p_3^0}\right)=\mathbb{V}_{as}\left(\widehat{p_3}\right)-\mathbb{V}_{as}\left(\widehat{p_3^0}\right)}$$

(b) Calculer la loi limite de $\left(\widehat{p_3^0}-p_3\right)$, et en déduire $\mathbb{V}_{as}\left(\widehat{p_3^0}\right)$.

La loi asymptotique de $\left(\widehat{p_3}-\widehat{p_3^0}\right)$ se déduit de celle de $\begin{pmatrix}\widehat{p_1}\\ \vdots \\ \widehat{p_K}\end{pmatrix}$ déterminée précédemment.

Soit en effet $g : \begin{pmatrix}\mathbb{R}^K & \rightarrow & \mathbb{R}\\ (x_1,\ldots,x_K) & \mapsto & \frac{x_3}{2(1-x_1-x_2)}\end{pmatrix}$; ainsi $g\left(\widehat{p_3}\right)=\widehat{p_3^0}$. Alors

$$\begin{aligned}\frac{\partial g}{\partial x_1}(p_1,\ldots,p_K) &= \frac{p_3}{2\left(1-p_1-p_2\right)^2}\\ &= 2p_3\end{aligned}$$

$$\begin{aligned}\frac{\partial g}{\partial x_2}(p_1,\ldots,p_K) &= \frac{p_3}{2\left(1-p_1-p_2\right)^2}\\ &= 2p_3\end{aligned}$$

$$\frac{\partial g}{\partial x_3}(p_1,\ldots,p_K) = 1$$

$$\frac{\partial g}{\partial x_k}(p_1,\ldots,p_K) = 0,\quad \forall k\in [\![3,K]\!]$$

Donc d'après le théorème de Slutsky

$$
\begin{array}{rcl}
\sqrt{n}\left(\widehat{p_3^0}-p_3\right) & \xrightarrow[n\to+\infty]{\mathcal{L}} & \mathcal{N}\left(0, \begin{pmatrix} 2p_3 \\ 2p_3 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}' \begin{pmatrix} p_1(1-p_1) & -p_1p_2 & \cdots & -p_1p_K \\ -p_2p_1 & p_2(1-p_2) & \cdots & -p_2p_K \\ \vdots & \vdots & \ddots & \vdots \\ -p_Kp_1 & -p_Kp_2 & \cdots & p_K(1-p_K) \end{pmatrix} \begin{pmatrix} 2p_3 \\ 2p_3 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}\right) \\
& = & \mathcal{N}\left(0, \begin{pmatrix} 2p_3 \\ 2p_3 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}' \times \begin{pmatrix} 2p_3p_1(1-p_1)-2p_1p_2p_3-p_1p_3 \\ -2p_1p_2p_3+2p_3p_2(1-p_2)-p_2p_3 \\ -2{p_3}^2p_1-2{p_3}^2p_2+p_3(1-p_3) \\ 0 \\ \vdots \\ 0 \end{pmatrix}\right) \\
& = & \mathcal{N}\left(0, \left\{ \begin{array}{c} 4{p_3}^2p_1(1-p_1)-4p_1p_2{p_3}^2-2p_1{p_3}^2 \\ -4p_1p_2{p_3}^2+4{p_3}^2p_2(1-p_2)-2p_2{p_3}^2 \\ -2{p_3}^2p_1-2{p_3}^2p_2+p_3(1-p_3) \end{array}\right.\right) \\
& = & \mathcal{N}\left(0, \left\{ \begin{array}{c} [4p_1(1-p_1)-4p_1p_2-2p_1-4p_1p_2+ \\ 4p_2(1-p_2)-2p_2-2p_1-2p_2]\,{p_3}^2 \\ -{p_3}^2+p_3 \end{array}\right.\right) \\
& = & \mathcal{N}\left(0, \left\{-4(p_1+p_2)^2{p_3}^2-{p_3}^2+p_3\right.\right) \\
& = & \boxed{\mathcal{N}\left(0, p_3(1-2p_3)\right)} \quad \text{sous } H_0 : \text{“ } p_1+p_2=\frac{1}{2} \text{ ”}
\end{array}
$$

Ainsi $\boxed{\mathbb{V}_{as}\left(\widehat{p_3^0}\right)=p_3(1-2p_3)}$.

(c) Donner un estimateur $\widehat{\mathbb{V}_{as}\left(\widehat{p_3}-\widehat{p_3^0}\right)}$ convergent sous H_0 de $\mathbb{V}_{as}\left(\widehat{p_3}-\widehat{p_3^0}\right)$.
En déduire l'expression de la statistique du test d'Hausman de H_0

$$\xi^H = n\frac{\left(\widehat{p_3}-\widehat{p_3^0}\right)^2}{\widehat{\mathbb{V}_{as}\left(\widehat{p_3}-\widehat{p_3^0}\right)}}$$

On a

$$
\begin{array}{rcl}
\mathbb{V}_{as}\left(\widehat{p_3}-\widehat{p_3^0}\right) & = & \mathbb{V}_{as}\left(\widehat{p_3}\right)-\mathbb{V}_{as}\left(\widehat{p_3^0}\right) \\
& = & p_3(1-p_3)-p_3(1-2p_3) \\
& = & \boxed{{p_3}^2}
\end{array}
$$

Posons donc $\widehat{\mathbb{V}_{as}\left(\widehat{p_3}-\widehat{p_3^0}\right)} = \widehat{p_3}^2 \xrightarrow[n\to+\infty]{p.s.} \mathbb{V}_{as}\left(\widehat{p_3}-\widehat{p_3^0}\right)$. On a par ailleurs

$$\begin{aligned}
\widehat{p_3}-\widehat{p_3^0} &= \frac{n_3}{n} - \frac{n_3}{2(n-n_1-n2)} \\
&= \frac{n_3}{n}\left(1-\frac{1}{2(1-\frac{n_1}{n}-\frac{n_2}{n}}\right) \\
&= \widehat{p_3}\left(1-\frac{1}{2(1-\widehat{p_1}-\widehat{p_2})}\right) \\
&= \widehat{p_3}\frac{1-2\widehat{p_1}-2\widehat{p_2}}{2-2\widehat{p_1}-2\widehat{p_2}}
\end{aligned}$$

de sorte que

$$\begin{aligned}
\xi_n^H &= n\frac{\widehat{p_3}^2\left(\frac{1-2\widehat{p_1}-2\widehat{p_2}}{2-2\widehat{p_1}-2\widehat{p_2}}\right)^2}{\widehat{p_3}^2} \\
&= \boxed{n\left(\frac{1-2\widehat{p_1}-2\widehat{p_2}}{2-2\widehat{p_1}-2\widehat{p_2}}\right)^2} \\
&\xrightarrow[n\to+\infty]{\mathcal{L}} \chi_1^2
\end{aligned}$$

(d) Vérifier que le test d'Hausman est convergent.

Si H_0 est fausse, alors $p_1+p_2 \neq \frac{1}{2}$ et donc

$$\begin{aligned}
\frac{1}{n}\xi_n^H &= \left(\frac{1-2\widehat{p_1}-2\widehat{p_2}}{2-2\widehat{p_1}-2\widehat{p_2}}\right)^2 \\
&\xrightarrow[n\to+\infty]{p.s.} \left(\frac{p_1+p_2-\frac{1}{2}}{1-p_1-p_2}\right) \\
&> 0
\end{aligned}$$

de sorte que

$$\xi_n^H \xrightarrow[n\to+\infty]{\mathcal{L}} (+\infty)$$

Ainsi, si H_0 est fausse alors $\mathbb{P}\left(\xi_n^H < q_{1-\alpha}^{\chi_1^2}\right) \xrightarrow[n\to+\infty]{} \mathbb{P}\left((+\infty) < q_{1-\alpha}^{\chi_1^2}\right) = 0$:
le test d'Hausman est donc convergent.

(e) Montrer que les statistiques ξ^W et ξ^H sont asymptotiquement équivalentes sous H_0.

On a enfin

$$\begin{aligned}
\frac{\xi_n^H}{\xi_n^W} &= \frac{n\left(\frac{1-2\widehat{p_1}-2\widehat{p_2}}{2-2\widehat{p_1}-2\widehat{p_2}}\right)^2}{n\frac{\left(\widehat{p_1}+\widehat{p_2}-\frac{1}{2}\right)^2}{(\widehat{p_1}+\widehat{p_2})(1-\widehat{p_1}-\widehat{p_2})}} \\
&= \frac{\widehat{p_1}+\widehat{p_2}}{1-\widehat{p_1}-\widehat{p_2}} \\
&\xrightarrow[n\to+\infty]{p.s.} (1)
\end{aligned}$$

Ainsi, les deux statistiques de test ξ_n^H et ξ_n^W sont asymptotiquement équivalentes.

Exercice corrigé 5

La mise en œuvre de nombreuses techniques d'analyse, et notamment la quasi-totalité des techniques d'estimation et de test de modèles statistiques, nécessitent d'*engendrer* des réalisations x d'une variable aléatoire réelle X de loi $\mathcal{L}$ donnée.
L'objet de cet exercice est de présenter quelques techniques permettant à un ordinateur, machine essentiellement déterministe, d'engendrer de telles variables. Il s'agit donc de construire pour toute loi $\mathcal{L}$ une fonction récursive $f^{\mathcal{L}} : \mathbb{N} \to I$ qui énumère des réalisations d'un variable $X \rightsquigarrow \mathcal{L}$, *i.e.* telle que la distribution empirique des $\left(f^{\mathcal{L}}(n)\right)_{n\in\mathbb{N}}$ converge vers la vraie distribution $F^{\mathcal{L}}$.

☞ Q1 Cherchons tout d'abord à engendrer une variable aléatoire **entière** uniforme.
Soit $N \in \mathbb{N}^*$ et considérons $\mathcal{L} = \mathcal{U}_{[\![0,N-1]\!]}$.

(a)

> Définissons pour $\phi : [\![0, N-1]\!] \to [\![0, N-1]\!]$ et $s \in [\![0, N-1]\!]$
>
> $$f_{\phi,s} : \left(\begin{array}{ccc} [\![0, N-1]\!] & \to & [\![0, N-1]\!] \\ t & \mapsto & \phi^t(s) \end{array}\right)$$
>
> Montrer que $f_{\phi,s}$ est périodique à partir d'un certain rang ; on note $T_{\phi,s}$ sa période.
> En quel sens peut-on dire que $f_{\phi,s}$ "génère" une variable uniformément distribuée sur $[\![0, N-1]\!]$?

Comme $[\![0, N-1]\!] \subsetneq \mathbb{N}$ $f_{\phi,s}$ n'est pas injective, soient $i < j$ les deux premiers entiers tels que $f_{\phi,s}(i) = f_{\phi,s}(j)$, et posons $T = j - i \geq 1$. Alors $f_{\phi,s}(i+T) = f_{\phi,s}(i)$, donc $f_{\phi,s}(i+1+T) = f_{\phi,s}(i+1)$ et par récurrence $\forall k \geq i, f_{\phi,s}(k+T) = f_{\phi,s}(k)$ et donc $f_{\phi,s}$ et périodique au-delà de i.
Remarquons que la suite $(f_{\phi,s}(k))_{k\in\mathbb{N}}$ est *déterministe* et pas aléatoire ; ll s'agit de déterminer si elle peut être la suite des *réalisations* d'une variable aléatoire uniformément distribuée sur $[\![0, N-1]\!]$.
Définissons $g(k) = \lim_{N\to+\infty} \frac{1}{N}|\{i \in \mathbb{N}/y(i) = k\}|$ la fréquence empirique de $k \in [\![0, N-1]\!]$; alors $f_{\phi,s}$ "génère" une variable distribuée selon la loi $\mathcal{L}$ sur $[\![0, N-1]\!]$ si la fréquence empirique de tout k est égale à sa probabilité selon $\mathcal{L}$.
En particulier, $f_{\phi,s}$ est uniformément distribuée *ssi* $\forall k \in [\![0, N-1]\!] f_{\phi,s}(k) = \frac{1}{N-1}$.

(b) En quel sens peut-on dire que $f_{\phi,s}$ est d'autant meilleure que $T_{\phi,s}$ est grande ?

On a $T \leq N$: en effet pour tout $a \in \mathbb{N}$, $t \mapsto f_{\phi,s}(a+t)$ ne peut pas être injective de $[\![0,N]\!]$ vers $[\![0,N-1]\!]$, donc il existe $i < j \in [\![0,N]\!]$ tel que $f_{\phi,s}(a+j) = f_{\phi,s}(a+i)$, et par suite $T_{\phi,s} \leq N$.

Or si $T_{\phi,s} < N$, alors $f_{\phi,s}([i,+\infty[) = \{f_{\phi,s}(i), f_{\phi,s}(i+1), \ldots, f_{\phi,s}(i+T_{\phi,s}-1)\}$, donc $|f_{\phi,s}([i,+\infty[)| \leq T_{\phi,s} < N$, et donc $f_{\phi,s}([i,+\infty[) \subsetneq [\![0,N-1]\!]$. Ainsi l'un au moins des entiers de $[\![0,N-1]\!]$ n'est plus jamais atteint au-delà de i, et sa fréquence empirique est donc nulle : $f_{\phi,s}$ ne "génère" pas une loi uniforme sur $[\![0,N-1]\!]$.

Remarquons réciproquement que si $f_{\phi,s}$ est périodique de période exactement N, alors elle est uniforme.

Ainsi, un "bon générateur" de nombres pseudo-aléatoires sur $[\![0,N-1]\!]$ est périodique de période N.

(c) Soient $a \in [\![2,N-1]\!]$, $c \in [\![0,N-1]\!]$ et $p \in [\![2,N-1]\!]$ et définissons $\psi_{a,c,p}$: $\begin{pmatrix} \mathbb{N} & \to & [\![0,N-1]\!] \\ n & \mapsto & (an+c) \mod p \end{pmatrix}$.
A quelles conditions $f_{\psi_{a,c,p},s}$ est-elle un bon générateur uniforme sur $[\![0,N-1]\!]$?

Par définition de la division euclidienne, $f_{\psi_{a,c,p},s}$ est à valeur dans $[\![0,p-1]\!]$ donc pour qu'elle soit un bon générateur sur $[\![0,N-1]\!]$ il est nécessaire que $p = N$.

Par ailleurs dans $\mathbb{Z}/p\mathbb{Z}$ la suite $(a^n)_{n\in\mathbb{N}}$ est périodique et sa période divise p. Donc si $c = 0$ et si p est premier, $f_{\psi_{a,c,p},s}$ est de période 1 ou p, donc p (car $a \geq 2$), et est donc un bon générateur uniforme sur $[\![0,N-1]\!]$ (en fait il suffit que p soit premier avec a, et il n'est pas nécessaire que $c = 0$).

Ces générateurs ont été introduits par Lehmer et sont encore souvent utilisés[12]. On sait notamment que $2^{31} - 1 = 2147483647$ est premier (vérifié par Euler en 1750), ainsi que $2^{61} - 1$ (Pervouchine 1883) et $2^{127} - 1$ (Lucas, 1876). [13]

☞ Q2 Cherchons à engendrer une variable aléatoire réelle uniforme sur $[0,1]$.

[12] cf Knuth, D.E., 1981 ; The Art of Computer Programming, Volume 2 Seminumerical Algorithms, Addison-Wesley, Reading Mass., 688 pages, ISBN 0-201-03822-6

[13] Parmi les nombres sous la forme $2^n - 1$ seuls ceux où $n \in \{2,3,5,7,13,17,19,31,61,89,107,127\}$ sont premiers, et non pas seulement pour $n \in \{2,3,5,7,13,17,19,31,67,127,257\}$ comme prétendu incorrectement par Marin Mersenne dans la préface de Cogitata Physica-Mathematica (1644). Parmi les $2^{2^n-1} - 1$ courants en informatique, ceux pour $n \in \{2,3,5,7\}$ sont premiers.

(a) Soit $(x_n)_{n\in\mathbb{N}} \in [0,1]^{\mathbb{N}}$, et définissons pour $f : [0,1] \to \mathbb{R}$ continue et $n \in \mathbb{N}^*$

$$S_n^x(f) = \frac{1}{n}\sum_{k=1}^{n} f(x_k)$$

On dit de x qu'elle vérifie le *critère de Weyl* si pour toute f continue

$$S_n^x(f) \xrightarrow[n\to+\infty]{} \int_0^1 f$$

Comment s'interprète ce critère ?

Il s'agit d'une sorte de critère ergodique : la suite x est suffisamment "équi-répartie" et "équilibrée" pour que sommer aux points chargés par x ou sommer en tout point de $[0,1]$ soit équivalent.

(b) Soit $r \in \mathbb{R}$, et définissons $x = (frac(rk))_{k\in\mathbb{N}}$ où $frac(u) = u - E(u) \in [0,1[$ désigne la partie fractionnaire de $u \in \mathbb{R}$.
Montrer que x vérifie le critère de Weyl *ssi* r est irrationnel.

Supposons que r est rationnel.

– Supposons tout d'abord que $f_p = x \mapsto e^{2\pi ipx}$, $p \in \mathbb{Z}$. Par définition $rk = frac(rk) + E(rk)$ et donc

$$\begin{aligned} S_n^x(f_p) &= \frac{1}{n}\sum_{k=1}^{n} e^{2\pi ipx_k} \\ &= \frac{1}{n}\sum_{k=1}^{n} e^{(2p\pi i)\,rk} \text{ car } e^{(2p\pi i)E(rk)} = 1 \\ &= \frac{1}{n}\sum_{k=1}^{n} \left(e^{(2p\pi i)\,r}\right)^k \end{aligned}$$

Comme r est irrationnel, si $p \neq 0$, alors $e^{(2p\pi i)r} \neq 1$ et donc

$$\begin{aligned} S_n^x(f_p) &= \frac{1}{n}\frac{e^{(2p\pi i)r} - \left(e^{(2p\pi i)r}\right)^{n+1}}{1 - e^{(2p\pi i)r}} \\ &= \frac{1}{n}e^{(2p\pi i)r\frac{n+1}{2}}\frac{\sin(\pi nr)}{\sin(\pi r)} \end{aligned}$$

et donc

$$|S_n^x(f_p)| \leq \frac{1}{n}\frac{1}{|\sin(\pi r)|} \xrightarrow[n\to\infty]{} 0$$

soit

$$S_n^x(f_p) \xrightarrow[n\to\infty]{} \int_{[0,1]} e^{2\pi piu}du$$

Lorsque $p = 0$, on a bien-sûr $S_n^x(f_0) = 1 = \int_{[0,1]} f_0$.

– Le résultat reste donc vrai par linéarité de l'intégrale pour tout polynôme trigonométrique.
– Soit enfin f continue sur $[0,1]$; d'après le théorème de Dirichlet f est limite *uniforme* d'une suite $(P_l)_{l\in\mathbb{N}}$ de polynômes trigonométriques. Donc

$$\begin{aligned}
S_n^x(f) &= \frac{1}{n}\sum_{k=1}^{n} f(x_k) \\
&= \frac{1}{n}\sum_{k=1}^{n} \lim_{l\to\infty} P_l(x_k) \\
&= \lim_{k\to\infty} S_{nb}^x(P_k) \quad \text{car la somme } \sum_{k=1}^{n} \text{ est finie} \\
&= \lim_{k\to\infty} \int_{[0,1]} P_l \quad \text{d'après le résultat précédent} \\
&= \int_{[0,1]} \lim_{k\to\infty} P_l \quad \text{cqr la convergence de } (P_l)_{l\in\mathbb{N}} \text{ est uniforme} \\
&= \int_{[0,1]} f
\end{aligned}$$

ce qui achève la preuve.
Dans le cas en revanche où $r = \frac{a}{b}$ est rationnel, alors $(rk)_{k\in\mathbb{N}} = \{0, \frac{1}{b}, \ldots, \frac{b-1}{b}\}$: la suite x ne prend que b valeurs distinctes et le résultat est faux, par exemple lorsque f est la fonction affine par morceaux telle que

$$\left\{\begin{array}{rl} f\left(\frac{k}{b}\right) & = 1 \\ f\left(\frac{k+\frac{1}{2}}{b}\right) & = 2 \end{array}\right.$$

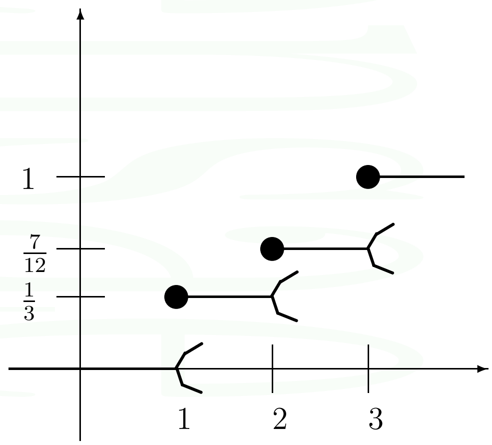
car alors $S_n^x(f) = 1 < \frac{3}{2} = \int_{[0,1]} f$.

☞ Q3 On cherche désormais à engendrer des réalisations d'une variable aléatoire de loi quelconque.

(a)

> Soit $\mathcal{L}$ la loi de densité $f_{\mathcal{L}} = \frac{1}{3}\delta_1 + \frac{1}{4}\delta_2 + \frac{5}{12}\delta_3$.
> Tracer la fonction de répartition $F_{\mathcal{L}}$ de $\mathcal{L}$.
> En déduire son inverse généralisée
>
> $$F^{-1} : \left(\begin{array}{ccc} [0,1] & \to & \mathbb{R} \\ s & \mapsto & \inf\{x \in \mathbb{R} \,/\, F_{\mathcal{L}}(x) \geq s\} \end{array}\right)$$
>
> Soit $U \rightsquigarrow \mathcal{U}_{[0,1]}$ et définissons $X = F_{\mathcal{L}}^{-1}(U)$; quelle est la loi de X ?

Le graphe de $F_{\mathcal{L}}$ est le suivant :



On a donc

$$F_{\mathcal{L}}^{-1}(t) = \begin{cases} -\infty & \text{si } 0 = t \\ 1 & \text{si } 0 < t < \frac{1}{3} \\ 2 & \text{si } \frac{1}{3} \leq t < \frac{7}{12} \\ 3 & \text{si } \frac{7}{12} \leq t \leq 1 \end{cases}$$

Par conséquent lorsque $U \rightsquigarrow \mathcal{U}_{[0,1]}$, $X = F_{\mathcal{L}}^{-1}(U)$ ne charge positivement que $\{1, 2, 3\}$ et on a

$$\begin{array}{rclcl} \mathbb{P}(X=1) & = & \mathbb{P}\left(0 \leq U < \frac{1}{3}\right) & = & \frac{1}{3} \\ \mathbb{P}(X=2) & = & \mathbb{P}\left(\frac{1}{3} \leq U < \frac{7}{12}\right) & = & \frac{1}{4} \\ \mathbb{P}(X=3) & = & \mathbb{P}\left(\frac{7}{12} \leq U < 1\right) & = & \frac{5}{12} \end{array}$$

de sorte que $X \rightsquigarrow \mathcal{L}$.

(b) Soit alors $F_{\mathcal{L}}$ la fonction de répartition supposée continue et strictement croissante d'une loi $\mathcal{L}$ quelconque.
On définit de même $X = F_{\mathcal{L}}^{-1}(U)$ pour $U \rightsquigarrow \mathcal{U}_{[0,1]}$; quelle est la loi de X ?

On a pour tout $t \in [0, 1]$

$$\begin{aligned} \mathbb{P}(X \leq t) &= \mathbb{P}\left(F_{\mathcal{L}}^{-1}(U) \leq t\right) \\ &= \mathbb{P}(U \leq F_{\mathcal{L}}(t)) \quad \text{car } F_{\mathcal{L}} \text{ est croissante} \\ &= F_{\mathcal{L}}(t) \quad \text{car } U \rightsquigarrow U_{[0,1]} \end{aligned}$$

Par conséquent $X \rightsquigarrow \mathcal{L}$.

Remarquons que $F_{\mathcal{L}}$ étant strictement croissante, c'est sa continuité qui nous assure qu'elle est bijective.

(c) Proposer un algorithme qui génère des réalisations successives d'une variable X de loi exponentielle $\mathcal{E}(\lambda)$ de densité $f_{\mathcal{E}_\lambda}(x) = \lambda e^{-\lambda x} \mathbb{1}_{[0,+\infty[}(x)$ pour $\lambda > 0$.

$$F_{\mathcal{E}(\lambda)}(u) = \int_0^u \lambda e^{-\lambda x} dx = \left(1 - e^{-\lambda u}\right)$$

Soit donc $U \rightsquigarrow \mathcal{U}_{[0,1]}$; alors $-\frac{1}{\lambda}\ln(1-U) \rightsquigarrow \mathcal{E}(\lambda)$; or par ailleurs $(1-U) \rightsquigarrow \mathcal{U}_{[0,1]}$, et donc $-\frac{1}{\lambda}\ln(U) \rightsquigarrow \mathcal{E}(\lambda)$.
On génère donc u_n réalisation de $U \rightsquigarrow \mathcal{U}_{[0,1]}$ et on renvoit $x_n = -\frac{1}{\lambda}\ln(u_n)$.

(d) Soit $(X,Y) \rightsquigarrow \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \sigma^2 I_2\right)$.
Soient $R^2 = X^2 + Y^2$ et $\theta = \arctan\frac{Y}{X}$.
Montrer que R^2 et θ sont indépendantes, et donner leurs lois.
En déduire un algorithme de génération d'une variable gaussienne.

Soit $G : \mathbb{R}^2 \to R^n$ absolument continue ; alors

$$\begin{aligned}
\mathbb{E}\left(G(R^2,\theta)\right) &= \int_{\mathbb{R}^2} G\left(x^2+y^2, \arctan\frac{y}{x}\right)\frac{1}{\sqrt{2\pi\sigma^2}^2}e^{-\frac{x^2+y^2}{2\sigma^2}}dyxy \\
&= \int_{\mathbb{R}^+\times[0,2\pi]} G\left(\rho^2,\theta\right)\frac{1}{2\pi\sigma^2}e^{-\frac{\rho^2}{2\sigma^2}}\rho d\rho d\theta
\end{aligned}$$

par le changement de variables en coordonnées pôlaires [14]. En particulier si G est un produit $G(a,b) = G_1(a)G_2(b)$ alors

$$\mathbb{E}\left(G_1(R^2)G_2(\theta)\right) = \left(\frac{1}{\sigma^2}\int_{\mathbb{R}^+}\underbrace{G_1(\rho^2)e^{-\frac{\rho^2}{2\sigma^2}}\rho}_{\text{ne dépend que de }\rho}d\rho\right)\left(\frac{1}{2\pi}\int_{[0,2\pi}\underbrace{G_2(\theta)}_{\text{que de }\theta}d\theta\right)$$

Par conséquent, les variables R^2 et θ sont indépendantes.
En outre on a pour toute G_1 absolument continue

$$\begin{aligned}
\mathbb{E}\left(G_1(R^2)\right) &= \int_{\mathbb{R}^+}\frac{1}{\sigma^2}G_1(\rho^2)e^{-\frac{\rho^2}{2\sigma^2}}\rho d\rho \\
&= \int_{\mathbb{R}^+}\frac{1}{2\sigma^2}G_1(u)e^{-\frac{1}{2\sigma^2}u}du
\end{aligned}$$

[14] Soit $f : \mathbb{R}^2 \to E$ mesurable absolument continue, et soit à calculer $\int_{\mathcal{A}\subset\mathbb{R}^2} f(x,y)dxdy$. Posons

$$\phi : \left(\begin{array}{ccc} \mathbb{R}^*\times\mathbb{R} & \to & \mathbb{R}^+\times\mathbb{R} \\ \begin{pmatrix} x \\ y \end{pmatrix} & \mapsto & \begin{pmatrix} x^2+y^2 \\ \begin{cases} \arctan\left(\frac{y}{x}\right) & \text{si } x>0 \\ \arctan\left(\frac{y}{x}\right)+\pi & \text{si } x<0 \end{cases} \end{pmatrix} \end{array}\right)$$

ϕ est différentiable et de jacobienne

$$Jac(\phi)_{(x,y)} = \begin{pmatrix} 2x & 2y \\ -\frac{y}{x^2}\frac{1}{1+\left(\frac{y}{x}\right)^2} & \frac{1}{x}\frac{1}{1+\left(\frac{y}{x}\right)^2} \end{pmatrix}$$

donc de jacobien

$$\left|Jac(\phi)_{(x,y)}\right| = 2x\frac{1}{x}\frac{1}{1+\left(\frac{y}{x}\right)^2} - 2y\left(-\frac{y}{x^2}\frac{1}{1+\left(\frac{y}{x}\right)^2}\right) = 2$$

de sorte que

$$\int_{\mathcal{A}\subset\mathbb{R}^2} f(x,y)dxdy = \int_{\phi(\mathcal{A})} f\left(\phi^{-1}(\rho^2,\theta)\right)2\left(\frac{1}{2}\rho d\rho\right)d\theta$$

et donc $R^2 \rightsquigarrow \mathcal{E}(\frac{1}{2\sigma^2})$.
Par ailleurs pour toute G_2 absolument continue

$$\mathbb{E}\left(G_2(\theta)\right) = \int_{[0,2\pi[} G_2(\theta)\frac{1}{2\pi}d\theta$$

et donc $\theta \rightsquigarrow \mathcal{U}_{[0,2\pi]}$
Pour générer une réalisation d'une variable qui suivrait la loi $\mathcal{N}(m,\sigma^2)$, on génère donc $\frac{\theta_n}{2\pi}$ uniformément sur $[0,1]$ et r_n^2 selon la loi $\mathcal{E}(\frac{1}{\sigma^2})$ et on renvoit le nombre

$$\boxed{x_n = m + \sqrt{r_n^2}\cos\left(2\pi\left(\frac{\theta_n}{2\pi}\right)\right)}$$

(e) Même question si X suit une loi de Weibull $\mathcal{W}(\alpha,\beta)$ de densité $f_{\mathcal{W}(\alpha,\beta)}(x) = \alpha\beta x^{\beta-1}e^{-\alpha x^\beta}\mathbb{1}_{x\geq 0}$.

$$F_{\mathcal{W}(\alpha,\beta)}(y) = \int_0^y f_{\mathcal{W}(\alpha,\beta)}(x)dx = 1 - e^{\alpha y^\beta}$$

et donc on génère u_n réalisation de $U \rightsquigarrow \mathcal{U}_{[0,1]}$ et on renvoit

$$\boxed{x_n = \left(-\frac{1}{\alpha}\ln\left(u_n\right)\right)^{\frac{1}{\beta}}}$$

(f) Même question si X suit une loi de Poisson $\mathcal{P}(\lambda)$ de densité $f_{\mathcal{P}(\lambda)}(k) = \frac{\lambda^k}{k!}e^{-\lambda}$.
On pourra comparer $x_n = u_1 \times \cdots \times u_n$, où chaque u_n est une réalisation de $U \rightsquigarrow U_{[0,1]}$, à $e^{-\lambda}$.

La densité $f_{\mathcal{P}(\lambda)}$ s'écrit sur $\mathbb{R}$ comme une somme infinie de masses de Diracq $f_{\mathcal{P}(\lambda)}(s) = \sum_{k=0}^{+\infty}\frac{\lambda^k}{k!}e^{-\lambda}\delta_k(s)$ donc un algorithme consiste à générer u uniformément sur $[0,1]$ et à le placer dans l'un des intervalles $[S_n, S_{n+1}[$ où $S_n = \sum_{k=0}^{n}\frac{\lambda^k}{k!}e^{-\lambda}$.
Le calcul des $(S_n)_{n\in\mathbb{N}}$ peut néanmoins être dissuasif en pratique ; il est possible en fait de s'en passer.
Soient en effet $(Y_n)_{n\in\mathbb{N}} \underset{iid}{\rightsquigarrow} \mathcal{E}(\mu)$; alors $\forall n \in \mathbb{N}^*$, $Y_1 + \cdots + Y_n \rightsquigarrow \Gamma(n,\mu)$ (voir TD 1, exercice 2), et donc

$$\mathbb{P}\left(Y_1 + \cdots + Y_n > 1\right) = \int_{\frac{1}{\mu}}^{+\infty}\frac{t^{n-1}}{(n-1)!}e^{-t}dt$$

Définissons alors $\nu(Y) = \min\{n \,/\, Y_1 + \cdots + Y_n > 1\} - 1$; alors

$$\begin{aligned}
\mathbb{P}\left(\nu(Y) = n\right) &= \mathbb{P}\left(Y_1 + \cdots + Y_{n+1} > 1\right) - \mathbb{P}\left(Y_1 + \cdots + Y_n > 1\right) \\
&= \int_{\frac{1}{\mu}}^{+\infty}\frac{t^n}{n!}e^{-t}dt - \int_{\frac{1}{\mu}}^{+\infty}\frac{t^{n-1}}{(n-1)!}e^{-t}dt \\
&= \frac{1}{\mu^n n!}e^{-\frac{1}{\mu}}
\end{aligned}$$

et donc $\nu(Y) \rightsquigarrow \mathcal{P}\left(\frac{1}{\mu}\right)$.

Ainsi, engendrer une loi de Poisson de paramètre λ revient à engendrer une succession de lois exponentielles d'espérance $\frac{1}{\lambda}$. Or $\mathcal{E}\left(\frac{1}{\lambda}\right) = \mathcal{W}\left(\frac{1}{\lambda}, 1\right)$ donc d'après la question précédente si $U \rightsquigarrow \mathcal{U}_{[0,1]}$, alors $-\frac{1}{\lambda}\ln U \rightsquigarrow \mathcal{E}\left(\frac{1}{\lambda}\right)$. Par conséquent posons $U_i = 1 - e^{-\frac{1}{\lambda}Y_i}$; U_i est uniforme, et en outre

$$\begin{aligned} Y_1 + \cdots + Y_n > 1 &\Leftrightarrow -\frac{1}{\lambda}\sum_{k=1}^{n} \ln U_k > 1 \\ &\Leftrightarrow U_1 \cdots U_n > e^{-\lambda} \\ &\Leftrightarrow X_n > e^{-\lambda} \end{aligned}$$

où $X_n = U_1 \cdots U_n$

Définissons donc finalement $N(X) = \min\{n / X_n < e^{-\lambda}\} - 1$; alors $N(X) = \nu(X)$, donc $N(X) \rightsquigarrow \mathcal{P}(\lambda)$.

Pour générer une loi $\mathcal{P}(\lambda)$, il suffit donc de générer une suite $u_1, \ldots, u_n, \ldots$ uniformes sur $[0,1]$, et de renvoyer le premier entier $n-1$ tel que $u_1 \cdots u_n > e^{-\lambda}$.

Un tel algorithme s'écrit : [15]

$n \leftarrow 0$ et $x \leftarrow 1$

Tant que $x > e^{-\lambda}$

    Générer u_n uniformément sur $[0,1]$

    $x \leftarrow x \times u_n$ et $n \leftarrow n+1$

Renvoyer n

[15] Pour un exposé plus complet, voir D. Knuth, *The Art of Computer Programming*, vol. 2, p 132.