Filière 3EA
1^{re} année

Statistique - TD1

Moyenne et variance empiriques, manupulations de concepts de bases

1 Indépendance de la moyenne et de la variance empiriques (cas Gaussien)

Rappels et complément de cours donnés en début de séance

La loi normale à n dimension est une loi d'un vecteur aléatoire. On dira qu'un vecteur $X = \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix}$ suit une loi normale à n dimensions (ou encore est un vecteur Gaussien) d'espérance $\underline{m} = (m_1, ..., m_n)^T \in \mathbb{R}^n$ et matrice de covariance $\Lambda \in S_n^+(\mathbb{R})$ (ensemble des matrices symétriques définies positives) si sa loi est absolument continue par rapport à la mesure de Lebesgue, de densité :

$$f : (\underline{x}) = (x_1, ..., x_n)^T \mapsto \frac{1}{(2\pi)^{\frac{n}{2}} \sqrt{det(\Lambda)}} \exp\left(-\frac{1}{2}(\underline{x} - m)^T \Lambda^{-1} (\underline{x} - m)\right).$$

On le notera $X \sim \mathcal{N}(m; \Lambda)$.
Une manière équivalente de les définir est de demander que toute combinaison linéaire de ces composantes soit une variable réelle Gaussienne.
<u>Lois marginales :</u> Les lois marginales d'un vecteur normal sont normales. En fait, l'on a même que $Y = \begin{pmatrix} X_1 \\ \vdots \\ X_k \end{pmatrix}$ avec $k < n$, alors $Y \sim \mathcal{N}\left(\begin{pmatrix} m_1 \\ \vdots \\ m_k \end{pmatrix}; [\Lambda_{i,j}]_{(i,j) \in [\![1,k]\!]^2}\right)$.
<u>Transformation linéaire :</u> Considérons $P \in M_{k,n}(\mathbb{R})$ une matrice de rang $k < n$, $a \in \mathbb{R}^k$ et X un vecteur Gaussien. Toute transformation linéaire ou affine d'un vecteur Gaussien est encore un vecteur Gaussien. Plus précisément, pour $U = PX + a$, nous avons que

$$U \sim \mathcal{N}\left(Pm + a; P\Lambda P^T\right).$$

Nous allons voir dans le début de cet exercice une manière de construire un vecteur Gaussien : prendre des variables aléatoires Gaussiennes **indépendantes** et en former un vecteur par concaténation.
On considère un vecteur aléatoire à valeurs dans $\mathbb{R}^3$, de composantes X_1, X_2 et X_3 dans la base canonique. On suppose que X_1, X_2 et X_3 sont des variables aléatoires réelles (VAR) indépendantes de loi gaussienne $\mathcal{N}(0, 1)$.

a) Quelle est la densité du triplet (X_1, X_2, X_3) ? En déduire qu'il s'agit bien d'un vecteur Gaussien.

Comme les variables aléatoires sont indépendantes, la densité du triplet est le produit des densités :

$$f(x_1, x_2, x_3) = \prod_{i=1}^{3} f_i(x_i) = \prod_{i=1}^{3} \frac{1}{\sqrt{2\pi}} e^{-\frac{x_i^2}{2}} = \frac{1}{\sqrt{2\pi}^3} e^{-\frac{1}{2}(x_1^2 + x_2^2 + x_3^2)}.$$

et l'on reconnait alors la densité de la loi $\mathcal{N}(0_{\mathbb{R}^3}, I_3)$.
Soit M la matrice de changement de base orthonormée telle que :

$$M^T := \begin{pmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 \\ \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} & -\frac{2}{\sqrt{6}} \end{pmatrix}$$

Il est possible de vérifier (on ne demande pas de le faire) que la matrice M est une matrice orthogonale qui vérifie donc $M^{-1} = M^T$. On note X le vecteur colonne de composantes $(X_i)_{i=1,2,3}$ et Y le vecteur colonne $M^T X$ de composantes $(Y_i)_{i=1,2,3}$.

b) Quelle est la densité du triplet (Y_1, Y_2, Y_3) ?

Il s'agit d'une transformation affine d'un vecteur Gaussien. Y est donc également Gaussien, avec

$$Y \sim \mathcal{N}(M^T 0_{\mathbb{R}^3}; M^T M = I_3),$$

et a la même densité que X.

c) Déterminer les lois de Y_1, Y_2 et Y_3.

Il s'agit des coordonnées d'un vecteur Gaussien centré et réduit, et suivent donc une loi normale centrée et réduite unidimensionnelle.

d) Soit $\overline{X} = \frac{1}{3}\sum_{i=1}^{3} X_i$. Vérifier que :

$$\sum_{i=1}^{3} X_i^2 = \sum_{i=1}^{3} Y_i^2.$$

$$Y_2^2 + Y_3^2 = \sum_{i=1}^{3} X_i^2 - 3\overline{X}.$$

En déduire que les variables aléatoires $\overline{X}$ et $S_*^2 := \frac{1}{2}\sum_{i=1}^{3}(X_i - \overline{X})^2$ sont indépendantes.

La première égalité est une conséquence immédiate du fait que M préserve la norme. En effet,

$$\sum_{i=1}^{3} Y_i^2 = Y^T Y = (M^T X)^T M^T X = X^T M M^T X = X^T X = \sum_{i=1}^{3} X_i^2.$$

Pour montre la deuxième égalité à partir de la première, il s'agit de montrer que $Y_1^2 = 3\overline{X}^2$. Or

$$Y_1 = \frac{1}{\sqrt{3}} X_1 + \frac{1}{\sqrt{3}} X_2 + \frac{1}{\sqrt{3}} X_3 = \sqrt{3}\ \overline{X}.$$

Pour l'indépendance, il suffit de remarquer en développant $2S_*^2$ que

$$S_*^2 = \frac{1}{2}(Y_2^2 + Y_3^2).$$

Le lemme de coalition permet alors de conclure que S_*^2 et $\overline{X} = \frac{1}{\sqrt{3}} Y_1$ sont indépendants.

e) Donner les lois de $\overline{X}$ et de $2S_*^2$.

Comme combinaison linéaire de coordonnés d'un vecteur Gaussien, $\overline{X} \sim \mathcal{N}(0, \frac{1}{3})$. De plus, la forme trouvée dans la question précédente comme somme de carré de deux variables normales centrées et réduites, $2S_*^2 \sim \chi_2^2$.

2 Loi uniforme sur $[0, \theta]$

Nous reprenons l'exemple de cours où les x_i sont 100 réalisations de VAR X_i indépendantes et de même loi uniforme sur $[0, \theta]$. On sait de plus que $\min_{1 \le i \le 100} x_i = 0,25$ et $\max_{1 \le i \le 100} x_i = 24,71$. On pose

$$Z_n = \max_{1 \le i \le 100} Xi \qquad Y_n = \min_{1 \le i \le 100} Xi.$$

a) Calculer la fonction de répartition puis la densité de Z_n. En déduire l'espérance et la variance de Z_n.

On se donne $t \in \mathbb{R}$, on aura alors

$$\mathbb{P}(Z_n \le t) = \mathbb{P}(X_1 \le t \cap ... \cap X_n \le t) = \prod_{i=1}^{n} \mathbb{P}(X_i \le t) = \begin{cases} 0 & \text{si } t \le 0 \\ (\frac{t}{\theta})^n & \text{si } 0 \le t \le \theta \\ 1 & \text{sinon.} \end{cases}$$

où l'on a utilisé pour la dernière égalité que $\int_{-\infty}^{t} \frac{1}{\theta} 1_{[0;\theta]}(x)dx = \begin{cases} 0 & \text{si } t \le 0 \\ \frac{t}{\theta} & \text{si } 0 \le t \le \theta \\ 1 & \text{sinon.} \end{cases}$.

Nous en déduisons en dérivant la densité de Z_n :

$$f_{Z_n}(t) = \begin{cases} \frac{n}{\theta^n} t^{n-1} & \text{si } 0 \le t \le \theta \\ 0 & \text{sinon.} \end{cases}$$

Donc,

$$\mathbb{E}(Z_n) = \int_0^\theta t \frac{n}{\theta^n} t^{n-1} dt = \frac{n}{\theta^n} \frac{\theta^{n+1}}{n+1} = \frac{n}{n+1} \theta.$$

En particulier, Z_n est un estimateur biaisé de θ, mais asymptotiquement sans biais.
Pour calculer la variance, nous allons utiliser que $V(Z_n) = \mathbb{E}(Z_n^2) - \mathbb{E}(Z_n)^2$ pour simplifier le calcul.

$$\mathbb{E}(Z_n^2) = \int_0^\theta t^2 \frac{n}{\theta^n} t^{n-1} dt = \frac{n}{\theta^n} \frac{\theta^{n+2}}{n+2} = \frac{n}{n+2} \theta^2,$$

et donc

$$V(Z_n) = \frac{n}{n+2} \theta^2 - \frac{n^2}{(n+1)^2} \theta^2 = \frac{n}{(n+1)^2 (n+2)} \theta^2.$$

b) Calculer la fonction de répartition de Y_n. En déduire que sa loi admet la densité :

$$y \mapsto \frac{n}{\theta} (1 - \frac{y}{\theta})^{n-1} 1_{[0,\theta]}(y).$$

En déduire l'espérance et la variance de Y_n.

Le calcul de la fonction de répartition est très similaire à celui de celle de Y_n. La différence réside en le fait qu'il est plus facile de décrire l'événement "le minimum est plus grand que t" que l'événement "le minimum est plus petit que t".

$$\begin{aligned} \mathbb{P}(Y_n \le t) &= 1 - \mathbb{P}(Y_n > t), \\ &= 1 - \mathbb{P}(X_1 > t \cap ... \cap X_n > t), \\ &= 1 - \prod_{i=1}^{n} \mathbb{P}(X_i > t) = 1 - \prod_{i=1}^{n} (1 - \mathbb{P}(X_i \le t)), \\ &= 1 - \prod_{i=1}^{n} (1 - F_{X_i}(t)) = \begin{cases} 0 & \text{si } t \le 0 \\ 1 - (1 - \frac{t}{\theta})^n & \text{si } 0 \le t \le \theta \\ 1 & \text{sinon.} \end{cases} \end{aligned}$$

Un calcul de dérivée montre bien que Y_n admet pour densité

$$y \mapsto \frac{n}{\theta}(1-\frac{y}{\theta})^{n-1}1_{[0,\theta]}(y).$$

Pour le calcul de l'espérance, nous allons utiliser l'indication.

$$\begin{aligned}
\mathbb{E}[Y_n] &= \int_0^\theta t\frac{n}{\theta}(1-\frac{t}{\theta})^{n-1}dt,\\
&= n\theta\int_0^1 u(1-u)^{n-1}du \quad (u=\frac{t}{\theta}),\\
&= n\theta\beta(2,n) = n\theta\frac{1!(n-1)!}{(n+1)!},\\
&= \frac{\theta}{n+1}.
\end{aligned}$$

De la même manière,

$$\begin{aligned}
\mathbb{E}[Y_n^2] &= \int_0^\theta t^2\frac{n}{\theta}(1-\frac{t}{\theta})^{n-1}dt,\\
&= n\theta^2\int_0^1 u^2(1-u)^{n-1}du \quad (u=\frac{t}{\theta}),\\
&= n\theta^2\beta(3,n) = n\theta\frac{2!(n-1)!}{(n+2)!},\\
&= \frac{2\theta}{(n+1)(n+2)}.
\end{aligned}$$

Donc

$$V(Y_n) = \frac{2\theta}{(n+1)(n+2)} - \left(\frac{\theta}{n+1}\right)^2 = \frac{n\theta^2}{(n+1)^2(n+2)}.$$

c) On veut trouver un nouvel estimateur de θ. On pose $T_n = Z_n + Y_n$. Montrer que T_n est sans biais pour θ. En utilisant la formule générale (sans indépendance) de la variance d'une somme, puis en appliquant l'inégalité de Schwarz au terme de covariance, montrer qu'asymptotiquement quand $n\to\infty$, on a $V(T_n)\leq O(\frac{1}{n^2})$. Donner la valeur de l'estimation de θ associée à T_n pour les valeurs mesurées

L'estimateur T_n est bien sans biais :

$$\begin{aligned}
\mathbb{E}[T_n] &= \mathbb{E}[Z_n] + \mathbb{E}[Y_n],\\
&= \frac{n\theta}{n+1} + \frac{\theta}{n+1} = \theta.
\end{aligned}$$

Pour majorer la variance, nous allons utiliser que

$$V(Z_n+Y_n) = V(Z_n) + V(Y_n) + 2Cov(Z_n,Y_n).$$

Cette égalité et l'inégalité de Cauchy-Schwarz montre que

$$\begin{aligned}
V(T_n) = V(Z_n+Y_n) &= V(Z_n) + V(Y_n) + 2Cov(Z_n,Y_n),\\
&\leq V(Z_n) + V(Y_n) + 2\sqrt{V(Z_n)V(Y_n)},\\
&\leq \frac{n\theta^2}{(n+1)^2(n+2)} + \frac{n\theta^2}{(n+1)^2(n+2)} + 2\frac{n\theta^2}{(n+1)^2(n+2)},\\
&= \frac{4n\theta^2}{(n+1)^2(n+2)} = O(\frac{1}{n^2}).
\end{aligned}$$

Pour les valeurs mesurées, $t_n = \min\{x_i\} + \max\{x_i\} = 0,25 + 24,71. = 25,96$. Remarque : Comme T_n est sans biais et que sa variance tend vers 0, il s'agit d'un estimateur convergent (notion que nous verrons prochainement en cours)

Filière 3EA
1^{re} année

Statistique - TD2
Manipulation d'estimations dans le cas continue

1 Estimateurs de l'écart type

1.1 Un premier estimateur

Soient $x_1, ..., x_n$ les réalisations respectives de $X_1, ..., X_n$ variables aléatoires indépendantes et de même loi gaussienne $\mathcal{N}(0, \sigma^2)$.

a) Déterminer l'information de Fisher apportée par ce n-échantillon sur le paramètre σ.

Par définition, l'information de Fisher vérifie

$$I_n(\sigma) = nI_1(\sigma) = n\left(-\mathbb{E}(\frac{\partial^2}{\partial\sigma^2}\ln(f(x,\theta))\right),$$

où f est la vraisemblance d'une variable :

$$f(x\sigma) = \frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{x^2}{2\sigma^2}}.$$

On calcule donc successivement, pour $x \in \mathbb{R}$ et $\sigma \in \mathbb{R}_+$ donnés :

$$\ln(f(x,\sigma) = -\ln(\sqrt{2\pi}) - \ln(\sigma) - \frac{x^2}{2\sigma^2},$$

$$\frac{\partial}{\partial\sigma}\ln(f(x,\sigma) = -\frac{1}{\sigma} + \frac{x^2}{\sigma^3},$$

$$\frac{\partial^2}{\partial\sigma^2}\ln(f(x,\sigma) = \frac{1}{\sigma^2} - \frac{3x^2}{\sigma^4}.$$

Nous prenons alors l'espérance, en utilisant que $\mathbb{E}[X^2] = V(X) + \mathbb{E}[X]^2$.

$$\begin{aligned}
I_n(\sigma) &= n\left(-\mathbb{E}(\frac{\partial^2}{\partial\sigma^2}\ln(f(x,\theta))\right),\\
&= n\left(-\frac{1}{\sigma^2} + \frac{3}{\sigma^4}\mathbb{E}[X^2]\right),\\
&= n\left(-\frac{1}{\sigma^2} + \frac{3\sigma^2}{\sigma^4}\right),\\
&= \frac{2n}{\sigma^2}.
\end{aligned}$$

b) En déduire un minorant de la variance de tout estimateur sans biais de σ.

Nous déduisons directement de la question précédente et du théorème de Cramer-Rao un tel minorant :

$$BCR = \frac{1}{I_n(\sigma)} = \frac{\sigma^2}{2n}.$$

c) Déterminer a de façon que :

$$\tilde{\sigma} = \frac{a}{n}\sum_{i=1}^{n}|X_i|$$

soit un estimateur sans biais de σ.

Nous avons besoin pour répondre à la question de calculer l'espérance de $\tilde{\sigma}$

$$\mathbb{E}[\tilde{\sigma}] = \frac{a}{n}\sum_{i=1}^{n}\mathbb{E}(|X_i|).$$

Nous avons donc un calcul d'intégrale à effectuer :

$$\begin{aligned}\mathbb{E}(|X|) &= \int_{-\infty}^{+\infty}|x|\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{x^2}{2\sigma^2}}dx,\\ &= 2\int_{0}^{+\infty}x\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{x^2}{2\sigma^2}}dx,\\ &= \frac{2\sigma}{sqrt2\pi}\int^{+\infty}e^{-u}du \qquad (u = \frac{x^2}{2\sigma^2}),\\ &= \sigma\sqrt{\frac{2}{\pi}},\end{aligned}$$

et alors

$$\mathbb{E}[\tilde{\sigma}] = \frac{a}{n}n\sigma\sqrt{\frac{2}{\pi}} = a\sigma\sqrt{\frac{2}{\pi}}.$$

Pour avoir un estimateur sans biais, il suffira de prendre $a = \sqrt{\frac{\pi}{2}}$.

d) Avec cette valeur de a, l'estimateur $\tilde{\sigma}$ est-il convergent ? efficace ?

Pour répondre à ces questions, nous allons avoir besoin de calculer la variance de notre estimateur. Par indépendance mutuelle des variables (X_i),

$$V(\tilde{\sigma}) = \frac{a^2}{n^2}V(\sum_{i=1}^{n}|X_i|) = \frac{a^2}{n^2}\sum_{i=1}^{n}V(|X_i|) = \frac{a^2}{n}V(|X|).$$

En utilisant alors la question précédente et la formule de König-Huygens,

$$\begin{aligned}V(|X|) &= \mathbb{E}[|X|^2] - \mathbb{E}[|X|]^2 = \mathbb{E}[X^2] - \left(\sigma\sqrt{\frac{2}{\pi}}\right)^2,\\ &= \sigma^2 - \frac{2}{\pi}\sigma^2 = \frac{\pi-2}{\pi}\sigma^2.\end{aligned}$$

Donc en regroupant, nous aurons

$$V(\tilde{\sigma}) = \frac{\pi}{2n}\frac{\pi-2}{\pi}\sigma^2 = (\pi-2)\frac{\sigma^2}{2n}.$$

Comme $V(\tilde{\sigma}) \neq BCR$, l'estimateur n'est pas efficace. En revanche, comme il est sans biais et vérifie $V(\tilde{\sigma}_n) \to 0$, il est convergent.

1.2 Illustrations numériques et comparaison d'estimateurs

On considère une loi gaussienne $\mathcal{N}(0;4)$. On veut effectuer des comparaisons entre l'estimateur précédent, l'estimateur du maximum de vraisemblance : $\hat{\sigma} = \left(\frac{1}{n}\sum_{i=1}^n X_i^2\right)^2$ et avec la racine carrée de l'estimateur sans biais de la variance $s_* = \left(\frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X})^2\right)^{\frac{1}{2}}$

a) Quel protocole proposez-vous pour évaluer ces estimateurs ?

b) Comme il s'agit de trois variables aléatoires, on choisit de simuler n réalisations d'une loi normale centrée et de variance 4, pour diverses valeurs de n, puis d'effectuer les calculs correspondants. Commentez ce tableau.

n	10	50	100	200	500	1000
$\tilde{\sigma}$	2.6655	2.2528	2.0063	1.9844	2.0146	1.9515
$\hat{\sigma}$	2.7737	2.3343	2.0966	2.0034	2.0336	1.9708
s_*	2.5647	2.3544	2.1072	1.9859	2.0345	1.9718

FIGURE 1 – Valeurs des trois estimations pour chaque n.

On arrive bien à voir que les trois estimateurs convergent vers la bonne valeur 2 : on arrive de mieux en mieux à estimer l'écart-type.

c) On a répété 100 fois l'expérience précédente. Pour chaque valeur de n, on a calculé la moyenne des 100 estimations obtenues et leur variance reportée, en chiffres plus petits, sous la moyenne. Au vu de ce tableau, quel estimateur choisiriez-vous ? Justifier votre réponse.

n	10	20	50	100
$\tilde{\sigma}$	2.0361 0.2863	2.0331 0.1257	1.9744 0.0511	2.0110 0.0217
$\hat{\sigma}$	1.9727 0.2298	2.0086 0.1147	1.9688 0.0448	2.0031 0.0209
s_*	1.9854 0.2424	2.0018 0.1259	1.9655 0.0466	2.0048 0.0211

FIGURE 2 – Moyenne et variance des estimations en fonction de n, sur 100 répétitions.

Tout d'abord, notons que les variations d'efficacité entre les trois estimateurs sont très faibles. S'il faut choisir un seul estimateur, l'on prendra celui qui minimise l'erreur quadratique empirique pour la valeur de n qui nous intéresse. Par exemple, pour $n = 100$, comme

$$r_{\tilde{\sigma}} = (2.0110 - 2)^2 + 0.0217^2 \approx 5.9189 \times 10^{-4},$$

$$r_{\hat{\sigma}} = (2.031 - 2)^2 + 0.0209^2 \approx 4.4642 \times 10^{-4},$$

et

$$r_{s_*} = (2.0048 - 2)^2 + 0.0211^2 \approx 4.6825 \times 10^{-4},$$

l'on choisira $\hat{\sigma}$ l'estimateur du maximum de vraisemblance.

2 Estimateur du maximum de vraisemblance - Loi inverse-gamma

On dit qu'une VAR X suit une loi inverse-gamma, notée $IG(r,\lambda)$, si et seulement si elle admet la densité :

$$x \mapsto f(x) = \frac{\lambda^r}{\Gamma(r)} \frac{1}{x^{r+1}} \exp\left(-\frac{\lambda}{x}\right) 1_{\mathbb{R}_+^*}(x), \quad r > 0, \lambda > 0.$$

Une telle VAR peut s'écrire $X = \frac{1}{Y}$ avec Y qui suit une loi $\gamma(r,\lambda)$. Il est possible de montrer que

$$\mathbb{E}[X] = \frac{\lambda}{r-1} \qquad V(X) = \frac{\lambda^2}{(r-1)^2(r-2)}.$$

On considère un n-échantillon $x_1,...,x_n$ constitué des n réalisations respectives de $X_1,...,X_n$ VAR indépendantes et de même loi que X. On supposera dans cet exercice que r est connu.

a) On veut estimer λ. Écrire la log-vraisemblance du n-échantillon.

Nous avons affaire à une loi absolument continue par rapport à la mesure de Lebesgue, la vraisemblance est donc sa densité. Par indépendance des variables, nous obtenons

$$f_n(x_1,...,x_n,\lambda) = \prod_{i=1}^n \frac{\lambda^r}{\Gamma(r)} \frac{1}{x_i^{r+1}} \exp\left(-\frac{\lambda}{x_i}\right).$$

Et donc la log-vraisemblance est :

$$\ln(f_n) = nr\ln(\lambda) - n\ln(\Gamma(r)) - \sum_{i=1}^n (r+1)\ln(x_i) - \sum_{i=1}^n \frac{\lambda}{x_i}.$$

b) En déduire que l'estimateur du maximum de vraisemblance du paramètre λ est :

$$\lambda_n^{MV} = rn\left(\sum_{i=1}^n \frac{1}{X_i}\right)^{-1}.$$

L'on fera bien attention à justifier qu'on obtient bien un maximum en ce point.

Pour trouver le maximum de la vraisemblance, nous allons commencer par rechercher les points critiques de la log-vraisemblance, c'est-à-dire les points d'annulation de

$$\frac{\partial}{\partial\lambda}\ln(f_n) = \frac{nr}{\lambda} - \sum_{i=1}^n \frac{1}{x_i}.$$

On trouve bien un seul point critique, égal à

$$rn/\left(\sum_{i=1}^n \frac{1}{x_i}\right).$$

Pour montrer qu'il s'agit bien d'un maximum de la vraisemblance, nous allons montrer que la log-vraisemblance est concave (et admet donc un unique maximum, réalisé en son point critique). Pour cela, il suffit de remarquer que

$$\frac{\partial^2}{\partial\lambda^2}\ln(f_n) = -\frac{nr}{\lambda^2} < 0.$$

<u>Point dangereux :</u> Attention, il ne suffit pas de vérifier que la dérivée seconde est strictement négative uniquement en le point critique. En effet, cela montrerait uniquement qu'il s'agit d'un maximum local, et non d'un maximum global.

c) En s'appuyant sur les résultats de l'addition de VAR gammas indépendantes, dire quelle est la loi de :

$$Z = \frac{1}{\sum_{i=1}^{n} \frac{1}{X_i}}.$$

Calculer alors l'espérance et la variance de λ_n^{MV}. Cet estimateur est-il convergent ?

D'après l'énoncé, si $X_i \sim IG(r, \lambda)$, alors $\frac{1}{X_i} \sim \gamma(r, \lambda)$. Donc $\sum_{i=1}^{n} \frac{1}{X_i} \sim \gamma(nr, \lambda)$, et finalement

$$Z = \frac{1}{\sum_{i=1}^{n} \frac{1}{X_i}} \sim IG(nr, \gamma).$$

Donc

$$\mathbb{E}[\lambda_n^{MV}] = rn\frac{\lambda}{rn-1},$$

$$V(\lambda_n^{MV}) = (rn)^2\frac{\lambda^2}{(rn-1)^2(rn-2)}.$$

En particulier, l'estimateur λ_n^{MV} est asymptotiquement sans biais et vérifie $V(\lambda_n^{MV}) \to 0$. Il est donc convergent.

d) Calculer la quantité d'information de Fisher apportée par la VAR X sur le paramètre λ. En déduire simplement la quantité d'information apportée par le n-échantillon $X_1, ..., X_n$ sur le paramètre λ. L'estimateur $\hat{\lambda}_n = \frac{rn-1}{rn}\lambda_n^{MV}$ est-il un estimateur efficace du paramètre λ ?

Pour calculer l'information de Fischer, nous pouvons utiliser la dérivée seconde de la log-vraisemblance, calculé précédemment.

$$I_n(\lambda) = nI_1(\lambda) = -n\mathbb{E}[\frac{\partial^2}{\partial\lambda^2}\ln(f_1)] = n\frac{r}{\lambda}.$$

Enfin,

$$V(\hat{\lambda}_n) = \frac{(rn-1)^2}{(rn)^2}V(\lambda_n^{MV}) = \frac{\lambda^2}{rn-2} > \frac{\lambda^2}{rn} = BCR.$$

donc il n'est pas efficace.
En revanche, comme $\frac{BCR}{V(\hat{\lambda}_n)} = \frac{rn-2}{rn} \to 1$, il est bien asymptotiquement efficace.

Filière 3EA
1^{re} année

Statistique - TD3

Estimations cas continu (suite) et Test de Neyman-Pearson

1 Estimateur du maximum de vraisemblance - Loi de Pareto

On appelle loi de Pareto de paramètres $a > 0$ et $\theta > 0$ la loi qui a pour fonction de répartition F

$$F(x, a, \theta) := \begin{cases} 1 - (\frac{a}{x})^\theta & \text{si } x \geq a \\ 0 & \text{sinon.} \end{cases}$$

Elle a donc comme densité contre la mesure de Lebesgue la fonction

$$x \mapsto f(x, a, \theta) := \frac{\theta}{a}(\frac{a}{x})^{\theta+1} 1_{[a;+\infty[}(x).$$

Cette loi a été introduite pour modéliser la distribution de revenus supérieurs à un seuil donné.

a) Calculer la quantité d'information de Fisher apportée par une observation X_i sur le paramètre θ. On la notera $I_1(\theta)$.

Comme la variable est continue par rapport à la mesure de Lebesgue, la vraisemblance d'un unique tirage est égale à sa densité :

$$L(x, \theta) = \frac{\theta}{a}(\frac{a}{x})^{\theta+1}.$$

Donc une Log-vraisemblance égale à

$$\ln(L(x, \theta)) = \ln(\theta) - \ln(a) + (\theta + 1)\ln(\frac{a}{x}).$$

Ceci donne donc les dérivées partielles par rapport au paramètre

$$\frac{\partial}{\partial\theta}\ln(L(x, \theta)) = \frac{1}{\theta} + \ln(\frac{a}{x}),$$

$$\frac{\partial^2}{\partial\theta^2}\ln(L(x, \theta)) = -\frac{1}{\theta^2}.$$

Donc nous avons l'information de Fisher :

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

Remarque : Si l'on s'intéresse à l'information de Fisher pour le paramètre $u = \frac{1}{\theta}$, cela modifiera la variable par laquelle nous prenons les dérivées partielles.

b) Montrer que si la variable aléatoire réelle (VAR) X suit une loi de Pareto de paramètres a et θ alors la variable aléatoire réelle $Y = \ln(\frac{X}{a})$ suit une loi exponentielle $\mathcal{E}(\theta)$ de paramètre θ. Rappeler l'espérance et la variance d'une telle loi.

Il s'agit d'effectuer un changement de variables C^1. Nous pouvons vérifier que le changement inverse est

$$X = ae^Y.$$

Grace à la formule du changement de variables, nous trouvons alors que Y est une variable absolument continue, de densité

$$\begin{aligned} g(y) &= \frac{\theta}{a}(\frac{a}{ae^y})^{\theta+1}1_{[a;+\infty[}(ae^y)|ae^y|, \\ &= \theta e^{-\theta y}1_{[0;+\infty[}(y). \end{aligned}$$

Nous reconnaissons la densité de la loi exponentielle de paramètre θ. Nous rappelons

$$\mathbb{E}[Y] = \frac{1}{\theta} \qquad V(Y) = \frac{1}{\theta^2}.$$

c) On considère la statistique $T_n = \frac{1}{n}\sum_{i=1}^n \ln(\frac{X_i}{a})$, où les X_i sont des variables aléatoires indépendantes et de même loi que X. Montrer que T_n est un estimateur sans biais et efficace de $u = \frac{1}{\theta}$. Pour cela, admettre que dans cet exemple particulier, $I_1(u) = \frac{1}{I_1(\theta)}$.

Comme

$$\begin{aligned} \mathbb{E}[T_n] &= \frac{1}{n}\sum_{i=1}^n \mathbb{E}[\ln(\frac{X_i}{a})] \quad \text{(linéarité de l'espérance)}, \\ &= \frac{1}{n}\sum_{i=1}^n \frac{1}{\theta} = \frac{n}{n\theta} = \frac{1}{\theta}. \end{aligned}$$

T_n est sans biais.
De plus, nous avons :

$$\begin{aligned} V(T_n) &= \frac{1}{n^2}V\left(\sum_{i=1}^n \ln(\frac{X_i}{a})\right) \quad \text{(propriété de la variance)}, \\ &= \frac{1}{n^2}\sum_{i=1}^n V(\ln(\frac{X_i}{a})) \quad \text{(indépendances des observations)}, \\ &= \frac{1}{n^2}n\frac{1}{\theta^2} = \frac{1}{n\theta^2}. \end{aligned}$$

Or, l'indication nous donne que $I_n(u) = nI_1(u) = \frac{n}{I_1(\theta)} = n\theta^2$.
Nous avons bien que $V(T_n) = \frac{1}{I_n(u)}$, et donc que l'estimateur T_n est efficace.
Remarque : Nous obtenons au passage que T_n est convergent, puisque sans biais et de variance tendant vers 0.

d) (*Pour aller plus loin, à faire chez vous*) Montrer la relation admise sur l'information de Fisher, et montrer que T_n est en fait l'estimateur au maximum de vraisemblance.

2 Test d'hypothèses de Neyman et Pearson - Loi inverse-gamma

La puissance de l'écho reçu par un radar altimétrique permet d'analyser l'état de l'océan. Ceci permet également de déceler la présence de pluie qui diminue cette puissance.
On a mesuré la puissance de $n = 25$ échos reçus et on a obtenu $x_1, ..., x_n$ telles que $\sum_{k=1}^{n} \frac{1}{x_i} = 7,70$.
La puissance d'un écho est souvent modélisée par une loi inverse-gamma. On rappelle qu'une VAR X suit une loi inverse-gamma, notée $IG(r, \lambda)$, si et seulement si elle admet la densité :

$$x \mapsto f(x) = \frac{\lambda^r}{\Gamma(r)} \frac{1}{x^{r+1}} \exp\left(-\frac{\lambda}{x}\right) 1_{\mathbb{R}_+^*}(x), \quad r > 0, \lambda > 0.$$

Une telle VAR peut s'écrire $X = \frac{1}{Y}$ avec Y qui suit une loi $\gamma(r, \lambda)$. Il est possible de montrer que

$$\mathbb{E}[X] = \frac{\lambda}{r-1} \qquad V(X) = \frac{\lambda^2}{(r-1)^2(r-2)}.$$

Une connaissance préalable du phénomène et des appareils de mesure nous permet de fixer la valeur de r à 4. On veut mettre en œuvre le test de l'hypothèse H0 : absence de pluie, contre H1 : présence de pluie.

a) Montrer que le test peut se ramener à la situation suivante. On observe un échantillon $(X_1, ..., X_n)$ de loi $IG(r, \lambda)$ où λ est inconnu et on teste :

$$H_0 : \lambda = \lambda_0 \qquad \text{contre} \qquad H_1 : \lambda = \lambda_1$$

pour un certain $\lambda_1 < \lambda_0$.

D'après l'énoncé, en présence de pluie (hypothèse alternative H_1) le signal reçu est atténué,et donc en moyenne plus faible. En absence de pluie (hypothèse nulle H_0), ceci ne se produit pas. Or, selon le modèle retenu de loi inverse gamma, la puissance moyenne des observations acquises est proportionnelle au paramètre λ par la relation

$$\mathbb{E}[X] = \frac{\lambda}{r-1}.$$

Nous aurons donc bien un test de la forme

$$H_0 : \lambda = \lambda_0 \qquad \text{contre} \qquad H_1 : \lambda = \lambda_1 < \lambda_0.$$

b) À l'aide du théorème de Neyman et Pearson, montrer que le test le plus puissant de niveau α fixé repose sur la statistique $T_n = \sum_{k=1}^{n} \frac{1}{X_k}$. Énoncer la règle de décision correspondante.

Selon le Théorème de Neyman-Pearson, la statistique du test le plus puissant est celui pour lequel nous rejetons H_0 lorsque le rapport des vraisemblances dépasse une constante (dépendant de α et de la loi) :

$$R = \frac{L(X_1, ..., X_n, \lambda_1)}{L(X_1, ..., X_n, \lambda_0)} > C_{1,\alpha}.$$

Nous allons utiliser que les observations sont des réalisations indépendantes d'une variable absolument continue pour écrire que

$$L(X_1, ..., X_n, \lambda_i) = \prod_{k=1}^{n} \frac{\lambda_i^r}{\Gamma(r)} \frac{1}{X_k^{r+1}} \exp(-\frac{\lambda_i}{X_k}).$$

Raisonnons par équivalence, en notant $C_{i,\alpha}$ les constantes successives qui apparaissent :

$$\begin{aligned}
\frac{L(X_1,...,X_n,\lambda_1)}{L(X_1,...,X_n,\lambda_0)} > C_{1,\alpha} &\iff \left(\frac{\lambda_1}{\lambda_0}\right)^{nr} \exp\left[(\lambda_0-\lambda_1)\sum_{k=1}^{n}\frac{1}{X_k}\right] > C_{1,\alpha}, \\
&\iff \exp\left[(\lambda_0-\lambda_1)\sum_{k=1}^{n}\frac{1}{X_k}\right] > C_{2,\alpha}, \\
&\iff (\lambda_0-\lambda_1)\sum_{k=1}^{n}\frac{1}{X_k} > C_{2,\alpha} (\text{croissance de} \ln \text{ et de } \exp), \\
&\iff \sum_{k=1}^{n}\frac{1}{X_k} > C_{3,\alpha} \ (\text{car } \lambda_0 > \lambda_1).
\end{aligned}$$

Observez que le sens de l'inégalité est préservé dans le raisonnement par équivalence à l'aide de fonction croissante et de multiplication par une constante positive.
Le test repose donc sur une règle de la forme

$$Si\ T_n > \mu_\alpha, \text{ on rejette } H_0.$$

dans laquelle $T_n = \sum_{k=1}^{n} \frac{1}{X_k}$ est la statistique de test.

c) Sous l'hypothèse H_0, quelle est la loi de la statistique de test ? Montrer que le risque α de première espèce peut s'écrire :

$$\alpha = \mathbb{P}(T_n > \mu_\alpha) = \int_{2\mu_\alpha\lambda_0} h(u)du,$$

où $h(.)$ représente la densité de probabilité d'un khi-deux à $2rn$ degrés de liberté et μ_α la valeur critique de niveau α.

Sous H_0, la variable $Y_i = \frac{1}{X_i}$ suit la loi $\gamma(r, \lambda_0)$. Donc, comme vu dans le cours de Probabilité, ou encore à l'aide des fonctions caractéristiques,

$$T_n \sim \gamma(nr, \lambda_0).$$

Le changement de variable $u = 2\lambda_0 x$ permet de montrer que

$$\begin{aligned}
\alpha &= \mathbb{P}(\text{rejeter } H_0 | H_0 \text{ vraie}), \\
&= \int_{\mu_\alpha}^{+\infty} \frac{\lambda_0^{nr}}{\Gamma(nr)} x^{rn-1} e^{-\lambda_0 x} dx, \\
&= \int_{2\lambda_0\mu_\alpha}^{+\infty} \frac{\lambda_0^{nr}}{\Gamma(nr)} \left(\frac{u}{2\lambda_0}\right)^{rn-1} e^{-u/2} dx, \\
&= \int_{2\lambda_0\mu_\alpha}^{+\infty} \frac{1}{2^{rn}\Gamma(nr)} u^{rn-1} e^{-u/2} dx, \\
&= 1 - F_{\chi^2_{2rn}}(2\lambda_0\mu_\alpha),
\end{aligned}$$

où $F_{\chi^2_{2rn}}$ est la fonction de répartition de la loi χ^2_{2rn}. On peut alors utiliser la table de cette loi pour avoir le seuil exact, $\mu_\alpha = \frac{1}{2\lambda_0} q_{1-\alpha}$ avec $q_{1-\alpha}$ le quantile d'ordre $1-\alpha$ d'un χ^2_{2rn}.
Un raisonnement similaire permettrait alors de calculer la puissance de ce test.

d) En appliquant le théorème central limite à la suite $\frac{1}{X_1}, ..., \frac{1}{X_n}$ montrer que, sous l'hypothèse H_0, on a :

$$\frac{\sum_{i=1}^{n} \frac{1}{X_k} - \frac{nr}{\lambda_0}}{\frac{\sqrt{nr}}{\lambda_0}} \underset{n\to+\infty}{\to} \mathcal{N}(0,1).$$

On utilise cette loi asymptotique pour faire le test et on fixe $\lambda_0 = 25$. On rappelle que le quantile d'ordre 95% d'une loi $\mathcal{N}(0,1)$ vaut approximativement $F^{-1}(0,95.) = 1,64$. Quelle hypothèse retenez-vous au risque $\alpha = 5\%$?

Puisque T_n est une somme de variables indépendantes de même loi de variance finie, nous pouvons utiliser le Théorème Central Limite afin d'obtenir un seuil approché (valable asymptotiquement pour $n \to +\infty$). Nous avons sous H_0 que

$$\mathbb{E}[T_n] = \frac{rn}{\lambda_0} \qquad V(T_n) = \frac{rn}{\lambda_0^2}.$$

Nous pouvons donc conclure que la convergence en loi proposée par l'énoncé est alors correcte :

$$\frac{\sum_{i=1}^{n} \frac{1}{X_k} - \frac{nr}{\lambda_0}}{\frac{\sqrt{nr}}{\lambda_0}} \underset{n\to+\infty}{\to} \mathcal{N}(0,1).$$

Ceci permet l'approximation :

$$\begin{aligned}
\alpha &= \mathbb{P}(T_n > \mu_\alpha | H_0), \\
&= \mathbb{P}\Big(\frac{T_n - \frac{rn}{\lambda_0}}{\frac{\sqrt{rn}}{\lambda_0}} > \frac{\mu_\alpha - \frac{rn}{\lambda_0}}{\frac{\sqrt{rn}}{\lambda_0}} | H_0\Big), \\
&\approx \mathbb{P}\Big(Z > \frac{\mu_\alpha - \frac{rn}{\lambda_0}}{\frac{\sqrt{rn}}{\lambda_0}} | Z \sim \mathcal{N}(0,1)\Big), \\
&= 1 - \Phi_{\mathcal{N}(0,1)}\Big(\frac{\mu_\alpha - \frac{rn}{\lambda_0}}{\frac{\sqrt{rn}}{\lambda_0}}\Big).
\end{aligned}$$

Avec les valeurs numériques fournit de $\Phi^{-1}_{\mathcal{N}(0,1)}(0.95) = 1.64$, nous obtenons (par continuité de l'inverse de la fonction de répartition) l'approximation : $\frac{\mu_\alpha - \frac{rn}{\lambda_0}}{\frac{\sqrt{rn}}{\lambda_0}} \approx 1.64$.

Avec nos valeurs numériques, nous aurons alors $\mu_\alpha \approx \frac{\sqrt{rn}}{\lambda_0} 1.64 + \frac{rn}{\lambda_0} \approx 4,66$.
Puisque ici $T_n = 7,70 > 4,66$ (le seuil approché), nous rejetons H_0.

Filière 3EA
1^{re} année

Statistique - TD4
Régression linéaire multiple et estimation discrète

1 Régression linéaire multiple et influence du choix du modèle

Dans cet exercice, nous allons considérer des observations Y_1, ..., Y_n d'une variable Y qui dépend linéairement de deux paramètres X_1 et X_2, à l'ajout d'un bruit additif près. Nous aurons donc n de loi donnée par les relations

$$Y_k := aX_{k,1} + bX_{k,2} + B_k,$$

avec $B_1, ..., B_n \overset{i.i.d.}{\sim} \mathcal{N}(0, \sigma^2)$

a) Écrire l'estimateur des coefficients a et b obtenue par la méthode des moindres carrés, en faisant l'hypothèse de rang plein.

On a

$$\begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix} = \begin{pmatrix} X_{1,1} & X_{1,2} \\ \vdots & \vdots \\ X_{n,1} & X_{n,2} \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} + \begin{pmatrix} B_1 \\ \vdots \\ B_n \end{pmatrix}.$$

En notant $\mathbf{X} := \begin{pmatrix} X_{1,1} & X_{1,2} \\ \vdots & \vdots \\ X_{n,1} & X_{n,2} \end{pmatrix}$, la méthode des moindres carrés donne

$$\begin{pmatrix} \hat{a} \\ \hat{b} \end{pmatrix} = \left(\mathbf{X}^T\mathbf{X}\right)^{-1} \mathbf{X}^T \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix}.$$

Or, en notant $\underline{x_i} := \begin{pmatrix} x_{1,i} \\ \vdots \\ x_{n,i} \end{pmatrix}$ les vecteurs des paramètres d'influences, nous avons

$$\mathbf{X}^T\mathbf{X} := \begin{pmatrix} ||\underline{x_1}||^2 & \langle \underline{x_1}, \underline{x_2} \rangle \\ \langle \underline{x_1}, \underline{x_2} \rangle & ||\underline{x_2}||^2 \end{pmatrix}.$$

Le critère de rang plein est donc équivalent à ce que $||\underline{x_1}||^2||\underline{x_2}||^2 \neq \langle \underline{x_1}, \underline{x_2} \rangle^2$, ce qui avec Cauchy-Schwarz revient à demander que $\underline{x_1}$ et $\underline{x_2}$ ne soit pas colinéaire.
Sous cette condition,

$$\mathbf{X}^T\mathbf{X} := \frac{1}{||\underline{x_1}||^2||\underline{x_2}||^2 - \langle \underline{x_1}, \underline{x_2} \rangle^2} \begin{pmatrix} ||\underline{x_2}||^2 & -\langle \underline{x_1}, \underline{x_2} \rangle \\ -\langle \underline{x_1}, \underline{x_2} \rangle & ||\underline{x_1}||^2 \end{pmatrix}.$$

Comme de plus

$$\mathbf{X}^T \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix} := \begin{pmatrix} \underline{x_1}^T \underline{Y} \\ \underline{x_2}^T \underline{Y} \end{pmatrix},$$

nous arrivons aux estimateurs

$$\hat{a} := \frac{||\underline{x_2}||^2 \underline{x_1}^T \underline{Y} - \langle \underline{x_1}, \underline{x_2} \rangle \underline{x_2}^T \underline{Y}}{||\underline{x_1}||^2 ||\underline{x_2}||^2 - \langle \underline{x_1}, \underline{x_2} \rangle^2},$$

$$\hat{b} := \frac{-\langle \underline{x_1}, \underline{x_2} \rangle \underline{x_1}^T \underline{Y} + ||\underline{x_1}||^2 \underline{x_2}^T \underline{Y}}{||\underline{x_1}||^2 ||\underline{x_2}||^2 - \langle \underline{x_1}, \underline{x_2} \rangle^2}.$$

b) Les estimateurs obtenus sont-ils sans biais ?

En notant $q = ||\underline{x_1}||^2 ||\underline{x_2}||^2 - \langle \underline{x_1}, \underline{x_2} \rangle^2$,

$$\begin{aligned}
\mathbb{E}(\hat{a}) &= \frac{1}{q} \left[||\underline{x_2}||^2 \mathbb{E}(\sum_{k=1}^n a x_{k,1}^2 + \sum_{k=1}^n b x_{k,1} x_{k,2} + \sum_{k=1}^n x_{k,1} B_k) \right. \\
&\quad \left. - \langle \underline{x_1}, \underline{x_2} \rangle \mathbb{E}(\sum_{k=1}^n a x_{k,2} x_{k,1} + \sum_{k=1}^n b x_{k,2}^2 + \sum_{k=1}^n x_{k,2} B_k) \right], \\
&= \frac{1}{q} \left[qa + 0 \cdot b + \mathbb{E}(\sum_{k=1}^n ||\underline{x_2}||^2 \sum_{k=1}^n x_{k,1} B_k - \langle \underline{x_1}, \underline{x_2} \rangle \sum_{k=1}^n x_{k,2} B_k) \right], \\
&= a.
\end{aligned}$$

Donc $\hat{a}$ est un estimateur sans biais de a.
Un calcul similaire montre que $\hat{b}$ est un estimateur sans biais de b.

Une autre possibilité est d'utiliser la linéarité de l'espérance vectorielle, pour montrer que

$$\begin{aligned}
\begin{pmatrix} \mathbb{E}[\hat{a}] \\ \mathbb{E}[\hat{b}] \end{pmatrix} &= \mathbb{E} \left[\begin{pmatrix} \hat{a} \\ \hat{b} \end{pmatrix} \right], \\
&= \mathbb{E} \left[\left(\mathbf{X}^T \mathbf{X} \right)^{-1} \mathbf{X}^T \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix} \right], \\
&= \left(\mathbf{X}^T \mathbf{X} \right)^{-1} \mathbf{X}^T \mathbb{E} \left[\begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix} \right], \\
&= \left(\mathbf{X}^T \mathbf{X} \right)^{-1} \mathbf{X}^T \mathbb{E} \left[\mathbf{X} \begin{pmatrix} a \\ b \end{pmatrix} + \begin{pmatrix} B_1 \\ \vdots \\ B_n \end{pmatrix} \right], \\
&= \left(\mathbf{X}^T \mathbf{X} \right)^{-1} \mathbf{X}^T \mathbf{X} \begin{pmatrix} a \\ b \end{pmatrix} + 0, \\
&= \begin{pmatrix} a \\ b \end{pmatrix}.
\end{aligned}$$

c) Calculer la matrice de covariance de $\hat{\theta} := \begin{pmatrix} \hat{a} \\ \hat{b} \end{pmatrix}$.

Dans quel cas $\hat{a}$ et $\hat{b}$ sont-ils décorrélés ?
L'on pourra utiliser la relation donnant la matrice de covariance de $Z = AY$ en fonction de celle de Y et de la matrice A.

Notons $Q_X := \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T$, ce qui permet de réécrire

$$\begin{aligned} Cov(\hat{\theta}) &= Cov(Q_X\underline{Y}) \\ &= Q_X Cov(\underline{Y}) Q_X^T. \end{aligned}$$

Or $Cov(Y) = Cov(\mathbf{X}\theta + B) = Cov(B) = \sigma^2 I_n$, donc

$$Cov(\hat{\theta}) = \sigma^2 Q_X Q_X^T = \sigma^2\left(\mathbf{X}^T\mathbf{X}\right)^{-1}.$$

$\hat{a}$ et $\hat{b}$ sont décorrélés si et seulement si leur matrice de covariance est diagonale, c'est-à-dire ici que $\langle \underline{x_1}, \underline{x_2}\rangle = 0$.

d) Plaçons-nous dans la même situation que précédemment, et modifions le choix de l'estimateur. Nous allons prendre l'estimateur des moindres carrées d'un modèle ignorant l'influence du paramètre explicatif X_2 (soit par méconnaissance du système, ou par choix du fait de la difficulté de la mesure). Ainsi, ceci revient à prendre l'estimateur $\hat{\alpha}$ des moindres carrées du modèle

$$Z_k = \alpha X_{k,1} + \epsilon,$$

et de l'appliquer aux mesures (y_k) et $(x_{k,1})$. Est-ce que $\hat{\alpha}$ est un estimateur sans biais de a ? Quelle est sa variance (la comparer à la variance de $\hat{a}$) ?

Un simple calcul montre que

$$\begin{aligned} \hat{\alpha} &= (\underline{x_1}^T\underline{x_1})^{-1}\underline{x_1}^T\underline{Y}, \\ &= \frac{1}{||\underline{x_1}||^2}\left(a\sum_{k=1}^n x_{k,1}^2 + b\sum_{k=1}^n x_{k,1}x_{k,2} + \sum_{k=1}^n x_{k,1}B_k\right), \\ &= a + \frac{\langle \underline{x_1}, \underline{x_2}\rangle}{||X_1||^2}b + \frac{1}{||\underline{x_1}||^2}\sum_{k=1}^n x_{k,1}B_k. \end{aligned}$$

En passant à l'espérance, l'on obtient que $\mathbb{E}(\hat{\alpha}) = a + \frac{\langle \underline{x_1}, \underline{x_2}\rangle}{||X_1||^2}b$. Ainsi, $\hat{\alpha}$ est un estimateur sans biais si et seulement si $\langle \underline{x_1}, \underline{x_2}\rangle = 0$.
Enfin, en utilisant l'indépendance dans leur ensemble des bruits, l'on retrouve que

$$V(\hat{\alpha}) = \frac{1}{||\underline{x_1}||^4}\sum_{k=1}^n x_{k,1}^2\sigma = \frac{\sigma}{||\underline{x_1}||^2} = V(\hat{a}).$$

e) (**Pour aller plus loin**) Proposer un cheminement similaire à la question précédente dans le cas où l'on prend l'estimateur des moindres carrés d'un modèle ajoutant un paramètre explicatif sans influence.

2 Mobilisation contre la réforme des retraites

On désire évaluer la mobilisation contre la réforme des retraites à Toulouse. Pour cela, on s'intéresse au nombre de manifestants qui traversent le pont Saint-Pierre par seconde lors d'une manifestation. On compte, à plusieurs reprises pendant la manifestation, le nombre de manifestants Ni qui passent pendant T_i secondes sur le pont et on note les résultats obtenus sur n mesures (i = 1,...,n). Le nombre N_i est modélisé comme une variable aléatoire discrète suivant une loi de Poisson de paramètre λT_i et on suppose que les différentes mesures sont issues de variables aléatoires indépendantes.
On rappelle que pour une variable aléatoire X suivant une loi de Poisson de paramètre μ, on a : $\mathbb{P}[X = k] = \frac{\mu^k}{k!}e^{-\mu}$ pour $k \in \mathbb{N}$, $\mathbb{E}[X] = Var(X) = \mu$, $\mathbb{E}[e^{itX}] = \varphi_{Poisson}(t) = e^{\mu(e^{it}-1)}$).

2.1 Estimation du paramètre λ

À partir de l'observation de N_i, pour i = 1, . . . , n et de la connaissance exacte des T_i pour i = 1,...,n :

a) Déterminer l'estimateur $\hat{\lambda}^{MV}$ du maximum de vraisemblance de λ.

Calculons la vraisemblance.

$$\mathbb{P}(N_1 = k_1, ..., N_n = k_n) = \prod_{i=1}^{n} \frac{(\lambda T_i)^{k_i}}{k_i!} e^{-\lambda T_i}.$$

D'où

$$\ln(\mathbb{P}) = \ln(\prod \frac{T_i^{k_i}}{k_i!} + \sum_{i=1}^{n} k_i \ln(\lambda) - \lambda \sum_{i=1}^{n} T_i,$$

et donc

$$\frac{\partial \ln \mathbb{P}}{\partial \lambda} = \frac{1}{\lambda} \sum_{i=1}^{n} k_i - \sum_{i=1}^{n} T_i$$

Nous reconnaissons une fonction décroissante en λ (donc la log-vraisemblance est une fonction concave), qui ne s'annule qu'en $\sum_{i=1}^{n} N_i / \sum_{i=1}^{n} T_i$.
Donc

$$\hat{\lambda}^{MV} = \frac{\sum_{i=1}^{n} N_i}{\sum_{i=1}^{n} T_i}$$

b) $\hat{\lambda}^{MV}$ est-il sans biais ?

$$\mathbb{E}[\hat{\lambda}^{MV}] = \frac{\sum_{i=1}^{n} \lambda T_i}{\sum_{i=1}^{n} T_i} = \lambda$$

Donc l'estimateur est sans biais.

c) Donner une condition sur les T_i pour que l'estimateur $\hat{\lambda}^{MV}$ soit convergent.
On rappelle que l'estimateur du maximum de vraisemblance est asymptotiquement sans biais, convergent et asymptotiquement efficace dans le cas de variables aléatoires indépendantes et de même loi. Est-ce le cas ici ?

$$Var(\hat{\lambda}^{MV}) = \frac{1}{(\sum_{i=1}^{n} T_i)^2} \sum_{i=1}^{n} Var(N_i) = \frac{1}{(\sum_{i=1}^{n} T_i)^2} \sum_{i=1}^{n} \lambda T_i = \frac{\lambda}{\sum_{i=1}^{n} T_i}$$

Donc si $\sum_{i=1}^n T_i \overset{n\to+\infty}{\to} +\infty$, $\hat{\lambda}^{MV}$ est convergent.
Le résultat général ne peut pas s'appliquer ici, car si les N_i sont bien indépendantes, elles ne sont pas identiquement distribuées.

d) $\hat{\lambda}^{MV}$ est-il efficace ?

Reprenons la dérivée de la log-vraisemblance, afin d'obtenir

$$\frac{\partial^2 \ln \mathbb{P}}{\partial \lambda^2} = -\frac{1}{\lambda^2}\sum_{i=1}^n k_i$$

En prenant l'espérance, l'on obtient alors l'information de Fisher :

$$I(\lambda) = \mathbb{E}[-\frac{\partial^2 \ln \mathbb{P}}{\partial \lambda^2}] = \frac{\sum_{i=1}^n \lambda T_i}{\lambda^2}$$

Ainsi, $\hat{\lambda}^{MV}$ est efficace puisqu'il atteind la borne de Cramer-Rao :

$$Var(\hat{\lambda}^{MV}) = \frac{\lambda}{\sum T_i} = \frac{1}{I(\lambda)}$$

e) Application au cas où $T_i = iT$, pour i = 1,...,n, où T est une constante fixée.

$\sum_{i=1}^n T_i = \frac{n(n+1}{2}T \to +\infty$, donc $\hat{\lambda}^{MV} = \frac{2}{n(n+1)T}\sum_{i=1}^n N_i$ est sans biais, convergeant et efficace.

2.2 Test paramétrique sur λ

À partir de l'observation des N_i, pour i = 1, . . . , n et de la connaissance exacte des T_i pour i = 1,...,n, on souhaite valider le comptage réalisé par les syndicats ou celui réalisé par le ministère de l'Intérieur.

a) On exprime cette problématique sous la forme d'un test binaire d'hypothèses simples :

$$H_0 : \lambda = \lambda_0 \;\text{ contre } H_1 : \lambda = \lambda_1 > \lambda_0.$$

À quelle hypothèse (H0 ou H1) peut-on associer le comptage des syndicats (respectivement celui du ministère de l'Intérieur) en prenant en compte les tendances usuelles de minimisation et d'exagération ?

Comme λ est égal à la valeur moyenne, nous pouvons associer H_1 au comptage des syndicats.

b) Construire le test optimal de Neyman-Pearson : déterminer la statistique exhaustive (après simplifications) et la règle de décision.

Le rapport des vraisemblances est

$$\frac{\mathbb{P}_{H_1}}{\mathbb{P}_{H_0}}(N_1 = k_1, ..., N_n = k_n) = \prod_{i=1}^n \left(\frac{\lambda_1}{\lambda_0}\right)^{k_1} e^{-(\lambda_1)-\lambda_0)T_i}$$

Et la règle de rejet $\frac{\mathbb{P}_{H_1}}{\mathbb{P}_{H_0}} > C_\alpha$ est équivalente à

$$\begin{aligned}\mathbb{P}_{H_0} > C_\alpha &\iff \sum_{i=1}^n k_i \ln(\frac{\lambda_1}{\lambda_0}) > \tilde{C_\alpha} \\ &\iff S := \sum_{i=1}^n k_i > \mu_\alpha\end{aligned}$$

c) Quelle est la loi de $S = \sum_{i=1}^{n} N_i$ pour n fixé ?

Il s'agit d'une somme de loi de Poisson indépendantes. Redémontrons le résultat avec la fonction caractéristique. Soit $t \in \mathbb{R}$,

$$\varphi_S(t) = \prod_{i=1}^{n} \varphi_{N_i}(t) = e^{\lambda(\sum T_i)(e^{it}-1)}$$

Et l'on reconnait la fonction caractéristique d'une loi de Poisson de paramètre $\lambda(\sum T_i)$.

d) Pour $\lambda_0 = 0.1$ et $\lambda_1 = 0.2$, un risque α de l'ordre de 5% et $\sum_{i=1}^{n} T_i = 170$, calculer le seuil μ_α. (Le risque β sera estimé aux questions suivantes.)

Sous H_0, S suit une loi de Poisson de paramètre 17.
Au vu de la table,

$$\sum_{k=24}^{+\infty} \mathbb{P}(Poisson(17) = k) = 0.041 \qquad \sum_{k=25}^{+\infty} \mathbb{P}(Poisson(17) = k) = 0.063,$$

et l'on prend donc $\mu_\alpha = 24$

e) La table de la loi de Poisson est donnée pour des valeurs de paramètres allant jusqu'à $\lambda = 20$. Au-delà de cette valeur, on considère que la loi de Poisson peut être approchée par une loi normale. Refaites le calcul précédent en utilisant cette approximation (même si le paramètre de la loi de Poisson utilisé est inférieur, mais proche, de 20).

Nous approximons donc une loi de Poisson de paramètre 17 par une loi $\mathcal{N}(17, 17)$. Nous voulons donc

$$\alpha = 1 - \phi(\frac{\mu_\alpha - 17}{\sqrt{17}})$$

Ce qui donne avec la table

$$\frac{\mu_\alpha - 17}{\sqrt{17}} = 1.645$$

soit

$$\mu_\alpha = 23.78$$

f) En déduire le calcul de β.

On trouve sous H_1 une loi de Poisson de paramètre 34. Avec l'approximation, l'on trouve

$$\beta = \phi(\frac{\mu_\alpha - 34}{\sqrt{34}}) \approx 0.05$$

g) On a relevé les cinq mesures suivantes :

T_i	10	10	30	50	70
N_i	2	5	7	11	14

Que concluez-vous ?

On trouve $\sum_i N_i = 39$, avec $39 > \mu_\alpha$. On rejette donc $H_0 : \lambda = 0.1$ au risque et à la puissance $\alpha = 5\% \approx \beta$.
Remarquons que nous avons $\hat{\lambda}^{MV} = \frac{39}{170} \approx 0.229$ avec une variance $Var(\hat{\lambda}^{MV}) = \frac{\lambda}{\sum T_i} \approx 0.001$ très faible, ce qui aurait donné une bonne intuition de la réponse.

h) (**Pour aller plus loin**) En utilisant le relevé des valeurs et l'estimation du maximum de vraisemblance de λ, donner un intervalle de confiance à 95% de la valeur du paramètre λ (c'est-à-dire un intervalle aléatoire dépendant des mesures contenant λ avec probabilité 95%).

k \ λ	17	24	34
0	,0000	,0000	,0000
1	,0000	,0000	,0000
2	,0000	,0000	,0000
3	,0000	,0000	,0000
4	,0002	,0000	,0000
5	,0007	,0000	,0000
6	,0021	,0000	,0000
7	,0054	,0000	,0000
8	,0126	,0002	,0000
9	,0261	,0004	,0000
10	,0491	,0011	,0000
11	,0847	,0025	,0000
12	,1350	,0054	,0000
13	,2009	,0107	,0000
14	,2808	,0198	,0001
15	,3715	,0344	,0002
16	,4677	,0563	,0005
17	,5640	,0871	,0010
18	,6550	,1283	,0020
19	,7363	,1803	,0037
20	,8055	,2426	,0068
21	,8615	,3139	,0116
22	,9047	,3917	,0191
23	,9367	,4728	,0302
24	,9594	,5540	,0460
25	,9748	,6319	,0674
26	,9848	,7038	,0953
27	,9912	,7677	,1306
28	,9950	,8225	,1733
29	,9973	,8679	,2235
30	,9986	,9042	,2804
31	,9993	,9322	,3427
32	,9996	,9533	,4089
33	,9998	,9686	,4772
34	,9999	,9794	,5454
35	1,0000	,9868	,6117
36	1,0000	,9918	,6744
37	1,0000	,9950	,7319
38	1,0000	,9970	,7834
39	1,0000	,9983	,8283
40	1,0000	,9990	,8664
41	1,0000	,9995	,8981
42	1,0000	,9997	,9237
43	1,0000	,9998	,9439
44	1,0000	,9999	,9596
45	1,0000	1,0000	,9714
46	1,0000	1,0000	,9801

	0	**0,01**	**0,02**	**0,03**	**0,04**	**0,05**	**0,06**	**0,07**	**0,08**	**0,09**
0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7703	0,7734	0,7764	0,7793	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8906	0,8925	0,8943	0,8962	0,8980	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441
1,6	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
1,8	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,9767
2	0,9772	0,9778	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817
2,1	0,9821	0,9826	0,9830	0,9834	0,9838	0,9842	0,9846	0,9850	0,9854	0,9857
2,2	0,9861	0,9864	0,9868	0,9871	0,9875	0,9878	0,9881	0,9884	0,9887	0,9890
2,3	0,9893	0,9896	0,9898	0,9901	0,9904	0,9906	0,9909	0,9911	0,9913	0,9916
2,4	0,9918	0,9920	0,9922	0,9925	0,9927	0,9929	0,9931	0,9932	0,9934	0,9936
2,5	0,9938	0,9940	0,9941	0,9943	0,9945	0,9946	0,9948	0,9949	0,9951	0,9952
2,6	0,9953	0,9955	0,9956	0,9957	0,9959	0,9960	0,9961	0,9962	0,9963	0,9964
2,7	0,9965	0,9966	0,9967	0,9968	0,9969	0,9970	0,9971	0,9972	0,9973	0,9974
2,8	0,9974	0,9975	0,9976	0,9977	0,9977	0,9978	0,9979	0,9979	0,9980	0,9981
2,9	0,9981	0,9982	0,9982	0,9983	0,9984	0,9984	0,9985	0,9985	0,9986	0,9986

FIGURE 3 – Fonctions de répartition des lois de Poisson $\mathbb{P}_\lambda(X \leq k)$ (gauche) et normale centrée réduite $\mathbb{P}_{\mathcal{N}(0,1)}(X \leq y)$ (droite).

Rappel : La fonction de répartition de la loi $\mathcal{N}(0,1)$ vérifie

$$\Phi_{\mathcal{N}(0,1)}(y) = \int_{-\infty}^{y} \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}} dx, \qquad \Phi_{\mathcal{N}(0,1)}(-y) = 1 - \Phi_{\mathcal{N}(0,1)}(y)$$