Université A/Mira de Béjaia
Faculté des Sciences Exactes
Département de Mathématiques
2ème année Licence STID.

2019/2020

Corrigé de la série N°1 - Statistique inférentielle

Exercice 1

Soit $(X_1, ..., X_n)$ un n-échantillon aléatoire simple issu d'une variable aléatoire X de moyenne μ et d'écart-type σ. On suppose que la v.a X admet un moment centré d'ordre 4 :

$$\mu_4^{'} = \mathbb{E}(X - E(X))^4.$$

Le moment centré d'ordre 2 (variance empirique) de l'échantillon est :

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

– Montrer que :

$$Var(S^2) = \frac{n-1}{n^3}\left((n-1)\mu_4^{'} - (n-3)\sigma^4\right).$$

Solution.
On a :

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

D'autre part :

$$\begin{aligned}
\sum_{i,j}(X_i - X_j)^2 &= \sum_{i,j}\left(X_i^2 - 2X_iX_j + X_j^2\right) \\
&= \sum_{i,j}X_i^2 - 2\sum_i\sum_j X_iX_j + \sum_i\sum_j X_j^2 \\
&= 2n\sum_i X_i^2 - 2\sum_i X_i\sum_j X_j \\
&= 2n\sum_i X_i^2 - 2(n\overline{X})(n\overline{X}) \\
&= 2n^2S^2
\end{aligned}$$

On peut donc calculer la variance de S^2 en utilisant la relation suivante :

$$Var(S^2) = cov\left(S^2, S^2\right) = \frac{1}{(2n^2)^2}\sum_{i,j,k}cov\left((X_i - X_j)^2, (X_k - X_l)^2\right).$$

On calcule alors les différentes covariances selon la forme des facteurs :
- de la forme $cov((X_i - X_j)^2, (X_k - X_l)^2)$ avec i, j, k, l tous différents,
- de la forme $cov((X_i - X_j)^2, (X_k - X_j)^2)$ avec i, j, k différents,
- de la forme $cov((X_i - X_j)^2, (X_i - X_j)^2)$ avec i, j différents.

On remarque que si $i = j$ ou $k = l$, alors on obtient une covariance avec zéro (de la forme $cov(0, (X_k - X_l)^2))$ ou $cov((X_i - X_j)^2, 0)$ qui est nulle.

Commençons par le calcul de $cov((X_i - X_j)^2, (X_i - X_j)^2)$ avec $i \neq j$.

$$cov((X_i - X_j)^2, (X_i - X_j)^2) = E((X_i - X_j)^4) - [E((X_i - X_j)^2)]^2.$$

On introduit la moyenne μ dans le calcul de l'espérance :

$$\begin{aligned}
(X_i - X_j)^4 &= [(X_i - \mu) - (X_j - \mu)]^4 \\
&= (X_i - \mu)^4 - 4(X_i - \mu)(X_j - \mu)^3 + 6(X_i - \mu)^2(X_j - \mu)^2 - 4(X_i - \mu)^3(X_j - \mu) \\
&\quad +(X_j - \mu)^4 \\
E\left((X_i - X_j)^4\right) &= 2\mu'_4 - 8\mu'_3\mu'_1 + 6\mu'^2_2 \\
&= 2\mu'_4 + 6\sigma^4 \quad \text{car } \mu_1 = 0 \text{ et } \mu'_2 = \sigma^2.
\end{aligned}$$

$$\begin{aligned}
(X_i - X_j)^2 &= [(X_i - \mu) - (X_j - \mu)]^2 \\
&= (X_i - \mu)^2 - 2(X_i - \mu)(X_j - \mu) + (X_j - \mu)^2 \\
E\left((X_i - X_j)^2\right) &= 2\mu'_2 = 2\sigma^2.
\end{aligned}$$

Ainsi, pour $i \neq j$,

$$cov((X_i - X_j)^2, (X_i - X_j)^2) = 2\mu'_4 + 2\sigma^4.$$

Continuons par le calcul de $cov((X_i - X_j)^2, (X_k - X_j)^2)$ avec i, j, k différents.

$$\begin{aligned}
cov((X_i - X_j)^2, (X_k - X_j)^2) &= E((X_i - X_j)^2(X_k - X_j)^2) - [E((X_i - X_j)^2)E((X_k - X_j)^2] \\
&= E((X_i - X_j)^2(X_k - X_j)^2) - (2\sigma^2)^2.
\end{aligned}$$

$$
\begin{aligned}
&(X_i - X_j)^2(X_k - X_j)^2 \\
&= \left[(X_i-\mu)^2 - 2(X_i-\mu)(X_j-\mu) + (X_j-\mu)^2\right]\left[(X_k-\mu)^2 - 2(X_k-\mu)(X_j-\mu) + (X_j-\mu)^2\right] \\
&= (X_i-\mu)^2(X_k-\mu)^2 - 2(X_i-\mu)(X_j-\mu)(X_k-\mu)^2 + (X_j-\mu)(X_k-\mu)^2 \\
&\quad -2(X_i-\mu)^2(X_k-\mu)(X_j-\mu) + 4(X_i-\mu)(X_k-\mu)(X_j-\mu)^2 - 2(X_k-\mu)(X_j-\mu)^3 \\
&\quad +(X_i-\mu)^2(X_j-\mu)^2 - 2(X_i-\mu)(X_j-\mu)^3 + (X_j-\mu)^4 \\
&= 3(\acute{\mu_2})^2 + \acute{\mu_4}
\end{aligned}
$$

Ainsi, pour i, j, k différents,

$$cov((X_i - X_j)^2, (X_k - X_j)^2) = \acute{\mu_4} - \sigma^4.$$

Le dernier cas est rapidement calculé : si i, j, k, l sont différents, alors, par indépendance des X_i,

$$cov((X_i - X_j)^2, (X_k - X_l)^2) = 0.$$

Il reste à compter le nombre de termes dans chaque cas présenté.

- $cov((X_i - X_j)^2, (X_k - X_l)^2)$ est un terme de la forme $cov(X_i - X_j)^2, (X_i - X_j)^2)$ lorsque $(k = i, l = j)$ ou $(k = j, l = i)$ avec $i \neq j$, soit $2n(n-1)$ termes.
- $cov((X_i - X_j)^2, (X_k - X_l)^2)$ est un terme de la forme $cov(X_i - X_j)^2, (X_k - X_j)^2)$ lorsque ($l = j$ ou $l = i$) et k, i, j différents ou ($k = i$ ou $k = j$) et l, i, j différents, soit $(2+2)n(n-1)(n-2)$ termes.

$$
\begin{aligned}
\sum_{i,j,k,l} cov &= 2n(n-1)(2\acute{\mu_4} + 2\sigma^4) + 4n(n-1)(n-2)(\acute{\mu_4} - \sigma^4) \\
&= 4n(n-1)^2\left[\acute{\mu_4} - \frac{n-3}{n-1}\sigma^4\right]
\end{aligned}
$$

Exercice 2

Soit $(X_1, ..., X_n)$ un n-échantillon aléatoire simple issu d'une variable aléatoire X. les réalisations $x_1, x_2, ..., x_n$ peuvent être réordonnées en $y_1, y_2, ..., y_n$ où $y_1 < y_2 < ... < y_n$. Les y_i constituent une permutation particulière des x_i. Les y_i sont des réalisations du n-uplet de variables aléatoires $(Y_1, Y_2, ..., Y_n)$ qui constitue l'échantillon ordonné de X.

1. Calculer la loi de $Y_n = \sup_{1 \leq i \leq n} X_i$.
2. Calculer la loi de $Y_1 = \inf_{1 \leq i \leq n} X_i$.
3. Déterminer la loi du couple (Y_1, Y_n). En déduire celle de l'étendue $R = Y_n - Y_1$.
4. Soit N_z le nombre de X_i inférieurs à z. Quelle est la loi de N_z?

Solution :

Soit F (resp. f) la fonction de répartition (resp. la densité) de la variable aléatoire X. Soient H_k (resp. h_k) les fonctions de répartition (resp. les densités) de Y_k.

1. • **Loi de Y_n.**

$$H_n(y) = [F(y)]^n, h_n(y) = nF(y)^{n-1}f(y).$$

2. • **Loi de Y_1.**

$$H_1(y) = 1 - [1 - F(y)]^n, h_1(y) = n[1 - F(y)]^{n-1} f(y).$$

3. • **Loi du couple** (Y_1, Y_n).

Supposons dans un premier temps que $y_1 \leq y_n$. On peut écrire :

$$\begin{aligned}
\mathbb{P}(Y_1 \leq y_1, Y_n \leq y_n) &= \mathbb{P}(Y_n \leq y_n) - \mathbb{P}(Y_1 > y_1, Y_n \leq y_n) \\
&= [F(y_n)]^n - \mathbb{P}\left(\bigcap_{i=1}^{n} \{X_i \in]y_1, y_n]\}\right) \\
&= [F(y_n)]^n - (F(y_n) - F(y_1))^n .
\end{aligned}$$

En dérivant deux fois, on obtient

$$\frac{\partial F_{Y_1,Y_n}(y_1, y_n)}{\partial y_1} = n(F(y_n) - F(y_1))^{n-1} f(y_1)$$

et

$$\frac{\partial^2 F_{Y_1,Y_n}(y_1, y_n)}{\partial y_1 \partial y_n} = n(n-1)(F(y_n) - F(y_1))^{n-2} f(y_1) f(y_n).$$

La densité de probabilité du couple (Y_1, Y_n) est

$$h_{Y_1,Y_n}(y_1, y_n) = n(n-1)(F(y_n) - F(y_1))^{n-2} f(y_1) f(y_n) 1_{y_1 \leq y_n}.$$

• **Loi de** $R = Y_n - Y_1$.

Disposant de la densité du couple (Y_1, Y_n), pour trouver la densité de la v.a.r. $R = Y_n - Y_1$, on peut dans un premier temps calculer la densité du couple (Q, R), où $Q = Y_1$, et ensuite calculer la loi marginale de la seconde coordonnée de ce couple. Le calcul de la loi du couple (Q,R) s'effectue facilement grâce à la formule du changement de variable. Prenons la fonction $g(q, r) = (g_1(q, r), g_2(q, r)) = (q, r - q)$ qui est évidemment

$$\begin{cases} q = y_1 \\ r = y_n - y_1 \end{cases} \Longrightarrow \begin{cases} y_1 = q = \varphi_1(q, r) \\ y_n = r + q = \varphi_2(q, r) \end{cases}$$

La matrice jacobienne de la transformation est donnée par :

$$J = \begin{pmatrix} \frac{\partial \varphi_1(q,r)}{\partial q} & \frac{\partial \varphi_1(q,r)}{\partial r} \\ \frac{\partial \varphi_2(q,r)}{\partial r} & \frac{\partial \varphi_2(q,r)}{\partial q} \end{pmatrix}$$

Le déterminant de la matrice J est égal à 1.

La densité de probabilité du couple (Q, R) est donnée par :

$$\begin{aligned}
\psi_{Q,R}(q, r) &= h_{Y_1,Y_n}(q, r + q) \mid J \mid 1_{\mathbb{R} \times \mathbb{R}^+}. \\
&= n(n-1)(F(q + r) - F(q))^{n-2} f(q) f(q + r) 1_{\mathbb{R} \times \mathbb{R}^+}.
\end{aligned}$$

La densité marginale de R est :

$$\begin{aligned}
\phi_R(r) &= \int_{\mathbb{R}^+} \psi_{Q,R}(q, r) dq. \\
&= \int_{\mathbb{R}^+} n(n-1)(F(q + r) - F(q))^{n-2} f(q) f(q + r) dq.
\end{aligned}$$

La fonction de répartition de la variable R est donnée par :

$$F_R(r) = \int_0^r \phi_R(x) dx.$$

4. **Loi de N_z**. On a

$$N_z = \sum_{i=1}^{n} 1_{X_i \le z}.$$

Les v.a.r. $1_{X_i \le z}$, pour $i = \overline{1,n}$, étant i.i.d. de loi de Bernoulli de paramètre $F(z)$, la loi de N_z est une Binomiale de paramètres n et $F(z)$, i.e. $N_z \rightsquigarrow \mathcal{B}(n, F(z))$.

Exercice 3 Pour les modèles suivants, donner la vraisemblance associé à l'observation d'un échantillon aléatoire simple $(X_1, ..., X_n)$.

1. Modèle gaussien.
2. Modèle uniforme
3. Modèle exponentiel.

Solution.

1. • **La vraisemblance pour un modèle gaussien**.
On considère le modèle gaussien de loi $\left\{\mathcal{N}(\mu, \sigma^2) : \mu \in \mathbb{R}, \sigma > 0\right\}$. La vraisemblance de l'échantillon $(x_1, ..., x_n)$ est alors :

$$\begin{aligned}
\mathcal{L}(x_1, ..., x_n; \mu, \sigma^2) &= \prod_{i=1}^{n} f_{(\mu,\sigma)}(x_i) \\
&= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2}\left(\frac{x_i - \mu}{\sigma}\right)^2\right\} \\
&= \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n \exp\left\{-\frac{1}{2}\sum_{i=1}^{n}\left(\frac{x_i - \mu}{\sigma}\right)^2\right\}
\end{aligned}$$

2. • **La vraisemblance pour un modèle de loi uniforme**.
On considère le modèle uniforme de loi : $\left\{\mathcal{U}_{[0,\theta]} : \theta > 0\right\}$. La densité de la v.a. (variable générique) X est :

$$f_\theta(x) = \frac{1}{\theta} 1_{[0,\theta]}(x).$$

La vraisemblance de l'échantillon $(x_1, ..., x_n)$ est alors :

$$\begin{aligned}
\mathcal{L}(x_1, ..., x_n, \theta) &= \prod_{i=1}^{n} f_\theta(x_i) \\
&= \frac{1}{\theta^n} \prod_{i=1}^{n} 1_{[0,\theta]}(x_i) \\
&= \frac{1}{\theta^n} 1_{[0,+\infty[}\left(\inf_{1\le i\le n} x_i\right) 1_{]-\infty,\theta]}\left(\sup_{1\le i\le n} x_i\right) \\
&= \frac{1}{\theta^n}\left(1_{\{\inf x_i \ge 0\}}\right)\left(1_{\{\sup x_i \le \theta\}}\right).
\end{aligned}$$

3. • **La vraisemblance pour un modèle exponentiel**.
On considère le modèle exponentiel de loi : $\{exp(\lambda) : \lambda > 0\}$. La densité de la v.a. générique X est :

$$f_\lambda(x) = \lambda e^{-\lambda x}, x \geq 0.$$

La vraisemblance de l'échantillon $(x_1, ..., x_n)$ est alors :

$$\begin{aligned}
\mathcal{L}(x_1, ..., x_n, \lambda) &= \prod_{i=1}^{n} f_\lambda(x_i) \\
&= \prod_{i=1}^{n} \lambda e^{-\lambda x_i} \\
&= \lambda^n e^{-\lambda \sum_{i=1} x_i}.
\end{aligned}$$