Mini-Projet statistiques

Athlem Jouidi

Exercice 1 :

1)

2) a) La fonction de répartition de loi normale $X \sim N(\mu, \sigma^2)$ s'écrit en utilisant la fonction $erf()$:

$$F(x) = \frac{1}{2}\left(1 + erf\left(\frac{x-\mu}{\sigma\sqrt{2}}\right)\right)$$

$$F(x) = y.$$

$$\Leftrightarrow \quad erf\left(\frac{x-\mu}{\sigma\sqrt{2}}\right) = 2y - 1$$

comme la fonction erf est inversible :

$$\Leftrightarrow \quad \frac{x-\mu}{\sigma\sqrt{2}} = erf^{-1}(2y-1)$$

$$\Leftrightarrow \quad x = \sigma\sqrt{2}\, erf^{-1}(2y-1) + \mu$$

alors $F^{-1}(y) = \sigma\sqrt{2}\, erf^{-1}(2y-1) + \mu$

pour $\sigma = 1$ $\mu = 0$

$$F^{-1}(y) = \sqrt{2}\, erf^{-1}(2y-1) = F^{-1}(p, 0, 1)$$

$$\Leftrightarrow \quad \boxed{F^{-1}(p, \mu, \sigma^2) = \mu + \sigma \, F^{-1}(p, 0, 1)}$$

b) D'après la courbe QQ-plot de loi normale, on remarque :

- pour les quantiles observés $\in [0, 2000]$: un écart faible entre la courbe des quantiles empiriques et la courbe des quantiles observées $\Rightarrow$ la QQ-plot est presque linéaire.

$\Rightarrow$ Le modèle est adapté pour l'observation dans cette région.

⊛ Pour les quantiles observés $\in [2000, 6000]$; la courbe QQ-plot diverge et ~~~~ prend une forme convexe. L'écart entre la courbes des quantiles empiriques et des quantiles observées sera plus important.

$\Rightarrow$ le modèle n'est plus adaptés pour cette région d'observation.

3) a) La fonction de répartition de la fonction exponentielle est : $F(p, \delta) = 1 - e^{-\delta x}$

$$F(x) = y$$

$$\Leftrightarrow \quad 1 - e^{-\delta x} = y$$

$$\Leftrightarrow \quad 1 - y = e^{-\delta x}$$

$$\Leftrightarrow \quad x = -\frac{1}{\delta} \ln(1-y)$$

$$\Leftrightarrow \quad F^{-1}(p, \delta) = -\frac{1}{\delta} \ln(1-y)$$

pour $\delta = 1$

$$F^{-1}(p, 1) = -\ln(1-y)$$

$$\Rightarrow \quad \boxed{F^{-1}(p, \delta) = \frac{1}{\delta} F^{-1}(p, 1)}$$

e) D'après la courbe des quantiles : on remarque que pour la majorité des quantiles observées $\in [0, 5000]$, la courbe QQ-plot a presque une forme linéaire et l'écart entre la courbe QQ-plot et la courbe des quantiles empiriques est presque faible.

Donc on peut déduire que le modèle est plus adaptés pour l'observation des que le modèle normal.

Exercice 2

$(X_1, \ldots, X_n)$ n variables aléatoires, $\forall i \in \{1, \ldots, n\}$ $X_i \sim \mathcal{E}(\lambda)$; $\lambda > 0$

$$F_\lambda(x) = 1 - P(X > x) = \begin{cases} 1 - e^{-\lambda x} & \text{si } x > 0 \\ 0 & \text{sinon} \end{cases}$$

La densité $p_\lambda(x) = \dfrac{\partial F_\lambda(x)}{\partial x} = \begin{cases} \lambda e^{-\lambda x} & \text{si } x > 0 \\ 0 & \text{sinon} \end{cases}$

soit $X \sim \mathcal{E}(\lambda) \Rightarrow \begin{cases} E(x) = \dfrac{1}{\lambda} \\ var(X) = \dfrac{1}{\lambda^2} \end{cases}$

; on cherche à estimer $g_1(\lambda) = \dfrac{1}{\lambda}$

1) soit $T(X) = \dfrac{1}{n} \sum_{i=1}^{n} X_i$

$$E(T(X)) = E\left(\frac{1}{n} \sum_{i=1}^{n} X_i\right) = \frac{1}{n} E\left(\sum_{i=1}^{n} X_i\right)$$

$$\underset{iid}{=} \frac{n}{n} E(X_1) \overset{X \sim \mathcal{E}(\lambda)}{=} \frac{1}{\lambda}$$

biais $E\,T(X) = E(T(X)) - g_1(\lambda) = 0$

$\Rightarrow$ $T(X)$ est non biaisée

$$EQM(T(X)) = var(T(X)) = var\left(\frac{1}{n} \sum_{i=1}^{n} X_i\right)$$

$$= \frac{1}{n^2} var\left(\sum_{i=1}^{n} X_i\right) \underset{X_i\ iid}{=} \frac{n}{n^2} var(X_1)$$

$$= \frac{1}{n} \frac{1}{\lambda^2}$$

$$\boxed{EQM(T(X)) = \frac{1}{n\lambda^2}}$$

$T_1(X)$ est un estimateur UVMB $\Leftrightarrow$ $var(T_1(X)) = \dfrac{g'_1(\lambda)}{I_n(\frac{1}{\lambda})}$

* calcule I_n

on pose $\theta = \frac{1}{\lambda}$

$I_n(\theta) = n I_1(\theta)$

$I_1(\theta) = - E\left(\frac{\partial^2}{\partial\theta^2} \log(P(X,\theta))\right)$

$\log(P(X,\theta)) = \log\left(\frac{1}{\theta} e^{-\frac{x}{\theta}}\right) = -\log(\theta) - \frac{x}{\theta}$

$\frac{\partial^2}{\partial\theta^2} \log(P(X,\theta)) = + \frac{1}{\theta^2} \quad \frac{2x}{\theta^3}$

$\frac{1}{\theta} = \lambda$

$\Rightarrow - E(-\lambda^2 - 2\lambda^3 x) = -(+\lambda^2 - 2\lambda^2) = \lambda^2$

$\boxed{I_n(\theta) = n I_1(\theta) = n\lambda^2}$

conclusion : Puisque $g'_1(\lambda) = 1$ on aura bien

$\frac{g'_1(\lambda)}{I_n(\lambda)} = var(T_1(X))$

$\Rightarrow$ $T_1(X)$ est un estimateur UVMB.

2) $\alpha > 0$

$T_{1,\alpha}(X) = \alpha T_1(X)$

$E(T_{1,\alpha}(X)) = \alpha E(T_1(X)) = \frac{\alpha}{\lambda}$

biais $(T_{1,\alpha}(X)) = E(T_{1,\alpha}(X)) - g_1(\lambda) = \frac{\alpha - 1}{\lambda}$

$T_{1,\alpha}$ est biaisée.

$$R(\delta, T_{1,\alpha}) = var(T_{1,\alpha}) + biais^2(T_{1,\alpha})$$
$$= var(\alpha T_1(X)) + biais^2(T_{1,\alpha})$$

$$R(\delta, T_{1,\alpha}) = \alpha^2 \, var(T_1) - \frac{(\alpha-1)^2}{\delta^2}$$

pour que : $R(\delta, T_{1,\alpha}) < R(\delta, T_1)$

$$\Leftrightarrow \alpha^2 \, var(T_1) + \frac{(\alpha-1)^2}{\delta^2} < var(\delta, T_1) \; ; \; \begin{matrix} \alpha > 0 \\ \delta > 0 \end{matrix}$$

$$\Leftrightarrow \begin{cases} \alpha^2 < 1 & \Rightarrow 0 < \alpha < 1 \\ \dfrac{(\alpha-1)^2}{\delta^2} = 0 & \Rightarrow \alpha - 1 = 0 \Rightarrow \alpha = 1 \end{cases}$$

$$\Leftrightarrow \alpha \in [0, 1]$$

pour $\alpha = 1$; on revient au cas initial $T_{1,\alpha} = T_1$

3) Estimation de $g_2(\delta) = \frac{\ln(2)}{\delta}$

on pose $\theta = \frac{\ln(2)}{\delta}$

pour $\varphi(x) = x$ on a :

$$\varphi(\theta) = E(\varphi(x)) = E(x) = \frac{1}{\delta} = \frac{\theta}{\ln(2)}$$

$$\Rightarrow \varphi(\theta) = \frac{\theta}{\ln(2)}$$

φ est continue et bijective $\Rightarrow$

$$\varphi^{-1}(x) = \ln(2)\, x$$

on pose comme estim une éspérance théorique :

$$\hat{\Phi}_n = \frac{1}{n} \sum_{i=1}^{n} \varphi(X_i) = \frac{1}{n} \sum_{i=1}^{n} X_i \in \Phi(\Theta)$$

L'estimateur $\hat{\theta}$ de θ est celui la solution de l'équation

$$\hat{\Phi}_n = \Phi(\theta)$$

$\Rightarrow \hat{\theta} = \Theta^{-1}(\hat{\Phi}_n)$

$$\boxed{\hat{\theta} = \frac{\ln(2) \sum_{i=1}^{n} x_i}{n}}$$

5) Comparaison mediane empirique et mediane estimée.

médiane empirique = 89,9

médiane estimée = 503,45

Les deux valeurs de la médiane sont très éloignées.

Ce résultat est prévisible, car la mediane empirique est fausse car le nombre des observations est très petit ($n = 55$) pour s'approcher de la loi exponentielle.

6) Les courbes du risque quadratique sont décroissantes, en fonction du nombre de l'échantillon.

Exercice 3 :

1) $X \sim \mathcal{E}(\lambda)$

La fonction caractéristique :

$$\phi_x(t) = E(e^{itx}) = \frac{\lambda}{\lambda - it}$$

Soit $T(x) = \sum_{i=1}^{n} T(x)$

$$\phi_{T(x)}(t) = E\left(e^{\sum_{j=1}^{n} x_i it}\right) = E\left(\prod_{j=1}^{n} e^{itx_j}\right)$$

$$\underset{=}{iid} \prod_{j=1}^{n} E(e^{itx_j}) = \prod_{j=1}^{n} \frac{\lambda}{\lambda - it} = \frac{\lambda^n}{(\lambda - it)^n}$$

$\Rightarrow \phi_{T(x)}(t) = \dfrac{\lambda^n}{(\lambda - it)^n} \Rightarrow$ il s'agit bien de la fonction caractéristique de loi Gamma

$\Rightarrow \cancel{T}(x) \sim G(n, \lambda)$

Le rapport de vraisemblance :

$$Z(T(x)) = \frac{P_{H_1}(T(x))}{P_{H_0}(T(x))}$$

$$= \frac{\lambda_1^n \, e^{-\lambda_1 T(x)}}{\lambda_0^n \, e^{\lambda_0 T(x)}} \qquad \lambda_1 > \lambda_0$$

$$Z(T(x)) = \left(\frac{\lambda_1}{\lambda_0}\right)^n e^{(\lambda_0 - \lambda_1) T(x)}$$

Le test de Nayman-Pearson de niveau α s'écrit :

$\mathbb{1}_{\{Z(x) \geq \alpha\}}$; $Z(x)$ est monotone en $(\lambda_0 - \lambda) T(x)$

alors $\mathbb{1}_{\{Z(x) > \alpha\}} = \mathbb{1}_{\{(\lambda_0 - \lambda) T(x) > c\}}$

$$= \mathbb{1}_{\{T(x) \leq \frac{c}{\lambda_0 - \lambda}\}} \qquad \text{puisque } \lambda_0 - \lambda < 0$$

$$= \mathbb{1}_{\{T(x) \leq t_\alpha\}}$$

$P_{\lambda_0}(\delta(T(x)) = 1) = \alpha$

$P_{\lambda_0}(T(x) < t_\alpha) = \alpha$

alors t_α est donnée par la quantile de loi gamma

$\Rightarrow \boxed{t_\alpha = q_{G(m,\lambda_0)}(\alpha)}$

2) p-valeur $= \inf\{\alpha \in [0,1],$ tel que $T(x) > q_{m,\lambda_0}(\alpha)\}$

$= [1 - F(T(x))]$ où F est la fonction de repartition de la loi $G(m, \lambda_0)$

3) voir fichier.R

4) $E\left(\frac{T(x)}{m}\right) = \frac{1}{m}\frac{m}{\lambda_0} = \frac{1}{\lambda_0}$

$var\left(\frac{T(x)}{m}\right) = \frac{1}{m^2}(T(x)) = \frac{1}{m^2}\sum_{i=1}^{m}(X_i) = \frac{\sigma^2}{m}$

$\Rightarrow \frac{\frac{T(x)}{m} - \frac{1}{\lambda_0}}{\sigma/\sqrt{m}} \longrightarrow \mathcal{N}(0,1)$

5)

6) On voit que la fonction quantile est croissante par rapport à alpha et decroissante par rapport à lampda
donc si on dimine α la valeur augmente.
Alors pour garder le même region de décision on doit diminier α.