UCLouvain

Faculté des sciences

Ecole de physique

LMAT1271 - Calcul des probabilités et analyse statistique
Projet de fin d'année

Auteurs:

Gorski Romain, Navarro Morales Emiliano, Willaert Maxime

22 mai 2022

1 Exercice 1 :

a) **Fonction quantile :**

On commence par calculer la fonction quantile pour la distribution étudiée dans l'exercice. Pour ce faire, on calculera la fonction $\mathbf{Q_{\theta_1,\theta_2}(t)}$ telle que $\mathbf{P(X \leq x_t) = t}$

$$P(X \leq x_t) = \int_{-\infty}^{x_t} f_{\theta_1,\theta_2}(x)dx = t$$

faisant appel à la condition (dans l'énoncé) que $x \geq \theta_2$ et en explicitant f

$$\begin{aligned}
&= \int_{\theta_2}^{x_t} \left(\theta_1 \theta_2^{\theta_1}\right) \frac{1}{x^{\theta_1+1}} dx \\
&= \left(\theta_1 \theta_2^{\theta_1}\right) \int_{\theta_2}^{x_t} x^{-(\theta_1+1)} dx \\
&= -\theta_2^{\theta_1}\, x^{-\theta_1} \Big|_{\theta_2}^{x_t} \\
&= \theta_2^{\theta_1} \left(\theta_2^{-\theta_1} - x_t^{-\theta_1}\right)
\end{aligned}$$

On emploie alors la condition que $P(X \leq x_t) = t$ afin de déterminer la fonction Q(t)

$$t = 1 - \left(\frac{\theta_2}{x_t}\right)^{\theta_1}$$

$$\implies \boxed{Q_{\theta_1,\theta_2}(t) = \theta_2 (1-t)^{-\frac{1}{\theta_1}}}$$

b) **Indice de Gini :** On commence par calculer l'espérance de X

$$\begin{aligned}
E[X] &= \int_{-\infty}^{\infty} x f_{\theta_1,\theta_2}(x) dx \\
&= \int_{\theta_2}^{\infty} \left(\theta_1 \theta_2^{\theta_1}\right) x^{-\theta_1} dx \\
&= \left(\theta_1 \theta_2^{\theta_1}\right) \frac{x^{1-\theta_1}}{(1-\theta_1)} \Big|_{\theta_2}^{\infty} \\
&= \frac{\theta_1 \theta_2}{\theta_1 - 1}
\end{aligned}$$

Puis nous utilisons les expressions de $E[X]$ et $Q(t)$ pour calculer l'indice de Gini

$$G_{\theta_1,\theta_2}(t) = 2\int_0^1 p - \frac{\int_0^p Q(t)dt}{E[X]} dp$$

$$= 2\int_0^1 p - \frac{\theta_1 - 1}{\theta_1\theta_2}\left(\int_0^p \theta_2(1-t)^{-\frac{1}{\theta 1}}dt\right) dp$$

$$= 2\int_0^1 p + \frac{\theta_1 - 1}{\theta_1}\left[\frac{(1-t)^{1-\frac{1}{\theta_1}}}{1-\frac{1}{\theta 1}}\right]_0^p dp$$

$$= 2\int_0^1 (1-p)^{1-\frac{1}{\theta_1}} - (1-p)dp$$

$$= \frac{1}{2\theta_1 - 1}$$

c) **Méthode du maximum de vraisemblance pour G :**

Nous commençons par appliquer la méthode du maximum de vraisemblance aux paramètres θ_1 et θ_2 pour obtenir leurs estimateurs $\hat{\theta}_{1,MLE}$ et $\hat{\theta}_{2,MLE}$.
Pour alléger la notation, nous noterons dans ce paragraphe $\theta = (\theta_1, \theta_2)$ et $\hat{\theta} = (\hat{\theta}_1, \hat{\theta}_2) = (\hat{\theta}_{1,MLE}, \hat{\theta}_{2,MLE})$.
Soit un échantillon $(x_i)_{i=1,\dots,n}$ obtenu à partir de nos variables iid $(X_i)_{i=1,\dots,n}$. L'estimateur $\hat{\theta}$ doit maximiser la fonction de vraisemblance

$$L(\theta) = \prod_{i=1}^n f_\theta(x_i)$$

ou, de manière équivalente, la fonction $\ln L(\theta)$. Nous cherchons ce maximum dans le domaine $\Theta =]2, +\infty[\times]0, +\infty[$.
Commençons par déterminer $\hat{\theta_2}$. Pour cela détaillons la forme de $L(\theta)$:

$$L(\theta) = \begin{cases} \frac{\left(\theta_1\theta_2^{\theta_1}\right)^n}{\left(\prod_{i=1}^n x_i\right)^{\theta_1+1}} & \text{si } \theta_2 \leq \min_{i=1}^n x_i \\ 0 & \text{sinon} \end{cases}$$

Nous voyons immédiatement que $\hat{\theta}_2$ doit se trouver dans l'intervalle $]0, \min_i x_i]$. Nous remarquons ensuite que pour $\theta_1 \in]0, \infty[$ et $0 < \theta_2^a < \theta_2^b \leq \min_i x_i$, nous avons $L(\theta_1, \theta_2^a) < L(\theta_1, \theta_2^b)$, nous en concluons

$$\hat{\theta}_{2,MLE} = min_{i=1}^n X_i$$

Nous déterminons ensuite $\hat{\theta}_1$ en fixant $\theta_2 = \hat{\theta}_2 = \min_i x_i$ et en résolvant l'équation

$$0 = \partial_{\theta_1} \ln L(\theta)|_{(\hat{\theta}_1, \hat{\theta_2})}$$
$$n\left(\frac{1}{\hat{\theta}_1} + \ln(\hat{\theta}_2)\right) - \ln\left(\prod_i x_i\right)$$

Ce qui nous donne

$$\hat{\theta}_{1,MLE} = \frac{1}{\ln\left(\frac{\prod_{i=1}^n X_i}{\hat{\theta}_2}\right)} = \frac{1}{\ln\left(\frac{\prod_{i=1}^n X_i}{\min_{i=1}^n X_i}\right)}$$

Enfin, puisque $G = h(\theta_1)$ avec $h(x) = \frac{1}{2x-1}$, nous avons

$$\hat{G}_{MLE} = h(\hat{\theta}_{1,MLE}) = \left[\frac{2}{\ln\left(\frac{\prod_{i=1}^n X_i}{\min_{i=1}^n X_i}\right)} - 1\right]^{-1}$$

d) **Méthode des moments pour G :**

Commençons par exprimer les paramètres $\theta_{1,2}$ en fonctions des moments $\mu_1 = E[X]$ et $\mu_2 = E[X^2]$ de X. Nous avons déjà

$$\mu_1 = \frac{\theta_1 \theta_2}{\theta_1 - 1}$$

Il nous reste à calculer μ_2:

$$\mu_2 = E[X^2] = \int_{\theta_2}^{\infty} \frac{\theta_1 \theta_2^{\theta_1}}{x^{\theta_1 - 1}} dx$$

Comme $\theta_1 > 2$, cette intégrale est bien définie et nous obtenons

$$\mu_2 = \frac{\theta_1 \theta_2^2}{\theta_1 - 2}$$

Nous devons donc résoudre le système d'équations

$$\begin{cases} \mu_1 = \frac{\theta_1 \theta_2}{\theta_1 - 1} \\ \mu_2 = \frac{\theta_1 \theta_2^2}{\theta_1 - 2} \end{cases}$$

En $\theta_{1,2}$. La première équation nous permet d'exprimer θ_2 en fonction de θ_1:

$$\theta_2 = \frac{\mu_1(\theta_1 - 1)}{\theta_1}$$

En injectant cette expression dans la seconde équation, nous obtenons une équation du second degré en θ_1:

$$0 = (\mu_2 - \mu_1^2)\theta_1^2 - 2(\mu_2 - \mu_1^2)\theta_1 - \mu_1^2$$

donnant les solutions

$$\theta_1 = 1 \pm \sqrt{\frac{\mu_2}{\mu_2 - \mu_1^2}}$$

En ajoutant la condition $\theta_1 > 2$, nous retenons la solution:

$$\theta_1 = 1 + \sqrt{\frac{\mu_2}{\mu_2 - \mu_1^2}}$$

Ce qui nous donne, pour θ_2:

$$\theta_2 = \frac{\mu_2 - \sqrt{\mu_2(\mu_2 - \mu_1^2)}}{\mu_1}$$

Finalement pour obtenir les estimateurs $\hat{\theta}_{1,MME}$ et $\hat{\theta}_{2,MME}$, nous remplaçons les μ_k $k = 1, 2$ par leurs estimateurs $m_k := n^{-1} \sum_{i=0}^{n} X_i^k$, ce qui nous donne:

$$\begin{cases} \hat{\theta}_{1,MME} = 1 + \sqrt{\frac{m_2}{m_2 - m_1^2}} \\ \hat{\theta}_{2,MME} = \frac{m_2 - \sqrt{m_2(m_2 - m_1^2)}}{m_1} \end{cases}$$

Finalement, puisque G est fonction de θ_1, nous pouvons à nouveau construire l'estimateur de G à partir de l'estimateur de θ_1:

$$\hat{G}_{MME} = \frac{1}{2\sqrt{\frac{m_2}{m_2 - m_1^2}} + 1}$$

e) **Calcul des estimateurs MLE et MME pour l'indice de Gini pour un échantillon de 20**

f) **Calcul des estimateurs MLE et MME pour l'indice de Gini pour un échantillon de 20, 1000 fois**

g) **Estimation du biais, de la variance et de l'erreur quadratique moyenne**

h) **Même procédure pour différentes tailles d'échantillons**

i)

2 Exercice 2 : Régression

a)

b)

c)