DST1 – Statistique

Polycopié et calculatrice autorisés. Tout autre document interdit.
Durée 2h
Date : 8 novembre 2021

Dans tout le sujet pour tout $\alpha \in]0,1[$, q_α désigne le quantile d'ordre α de la loi normale $\mathcal{N}(0,1)$. Approximations à 10^{-2} près : $q_{0.9} = 1.28$, $q_{0.95} = 1.64$, $q_{0.975} = 1.96$.

Exercice 1 (questions de cours (/5))

1. Nous observons $X_1, \dots, X_n \sim_{i.i.d.} X$ et nous souhaitons estimer le paramètre θ de la loi de X. Nous supposons qu'il existe $k \geq 1$ tel que

$$\theta = \mathbb{E}[X^k]^2$$

et nous supposons d'autre part que

$$\mathbb{E}[X^k] \neq 0 \text{ et que } \mathrm{Var}(X^k) < +\infty.$$

Écrire la variance asymptotique de l'estimateur des moments de θ en fonction de $\mathrm{Var}(X^k)$ et θ.

Solution: Nous avons $\theta = h(\mu_k)$ avec $\mu_k = \mathbb{E}[X^k]$ et $h(t) = t^2$. La variance asymptotique de l'estimateur des moments

$$\widehat{\theta}^{Mom} = \widehat{\mu}_k^2 = \left(\frac{1}{n}\sum_{i=1}^n X_i^k\right)^2$$

est

$$\mathrm{Var}^{(n)}(\widehat{\theta}^{Mom}) = \frac{h'(\mu_k)^2 \mathrm{Var}(X^k)}{n} = \frac{4\mu_k^2 \mathrm{Var}(X^k)}{n} = \frac{4\theta \mathrm{Var}(X^k)}{n}.$$

2. Parmi les deux graphiques de la Figure 1, lequel est le tracé de la fonction de répartition d'une loi à densité ?

Solution: Figure (b) (la fonction de répartition d'une variable à densité est continue).

3. On lance une pièce autant de fois que nécessaire jusqu'à ce qu'elle tombe sur Pile. Écrivez l'univers des possibles Ω associé à cette expérience aléatoire.

Solution: On pourra poser par exemple $\Omega = \{P, F\}^{\mathbb{N}^*}$.

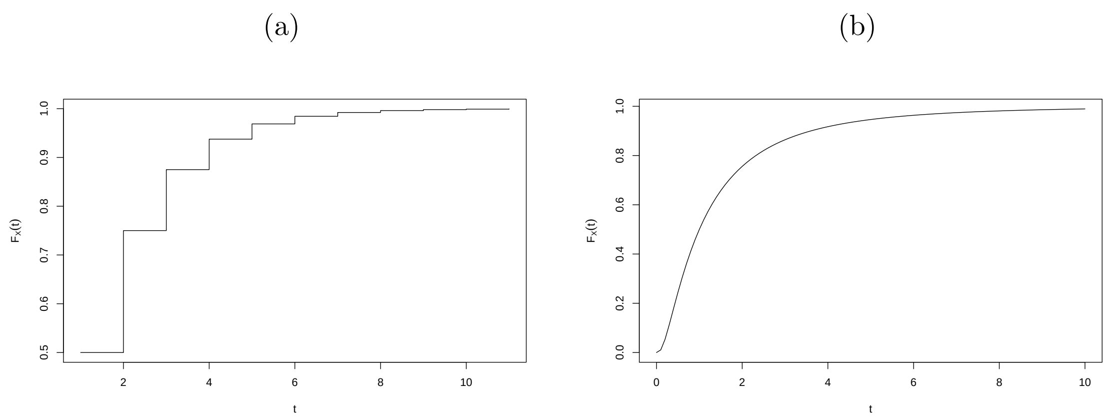

Figure 1: Tracé de deux fonctions de répartition (exercice 1, question 2)

4. Soit $\mathbb{P}$ une mesure de probabilité sur l'ensemble $\Omega = \{1, 2, 3\}$ telle que

$$\mathbb{P}(\{1\}) = 1/2 \text{ et } \mathbb{P}(\{2\}) = 1/3.$$

Que vaut $\mathbb{P}(\{3\})$?

Solution: On doit avoir

$$\mathbb{P}(\Omega) = \mathbb{P}(\{1\}) + \mathbb{P}(\{2\}) + \mathbb{P}(\{3\}) = 1.$$

Donc

$$\mathbb{P}(\{3\}) = 1 - \mathbb{P}(\{1\}) - \mathbb{P}(\{2\}) = 1 - \frac{1}{2} - \frac{1}{3} = \frac{1}{6}.$$

5. On observe $X_1, \ldots, X_n \sim_{i.i.d} \mathcal{N}(\theta, 0.5)$. On pose $\widehat{\theta} = 1$. Est-ce que $\widehat{\theta}$ est un estimateur de θ?

Solution: Oui car c'est une fonction des observations, même si ce n'est pas un estimateur convergent de θ dès que $\theta \neq 1$.

Exercice 2 (intervalle de confiance (/2))

Sur 1664 personnes interrogées dans une commune, 66 prennent les transports en commun pour se rendre à leur travail. Donner un intervalle de confiance asymptotique au niveau de risque 5% et un intervalle de confiance asymptotique au niveau de risque 10% pour la proportion de personnes de la commune se rendant en transport en commun au travail.

Solution: La proportion de personnes prenant les transports en commun parmi celles interrogées est

$$\widehat{p}(x) = \frac{66}{1664}.$$

Un intervalle de confiance de niveau de risque α pour la proportion p dans la population totale est donné par

$$\left[\widehat{p} \pm q_{1-\alpha/2}\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}\right].$$

Sa réalisation sur les données est

$$IC_{1-\alpha}^{(n)} = \left[\widehat{p}(x) \pm q_{1-\alpha/2}\sqrt{\frac{\widehat{p}(x)(1-\widehat{p}(x))}{n}}\right].$$

Au niveau 5% cela donne environ

$$IC_{95\%}^{(n)} = \left[\frac{66}{1664} \pm 1.96\sqrt{\frac{\frac{66}{1664}(1-\frac{66}{1664})}{1664}}\right] = [0.030; 0.050].$$

Au niveau 10% cela donne environ

$$IC_{90\%}^{(n)} = \left[\frac{66}{1664} \pm 1.64\sqrt{\frac{\frac{66}{1664}(1-\frac{66}{1664})}{1664}}\right] = [0.031; 0.048].$$

Exercice 3 (loi géométrique et loi exponentielle (/7))

Soit $X \sim \mathcal{E}(\lambda)$ c'est-à-dire que la densité de X est la fonction

$$f_X(t) = \lambda e^{-\lambda t}\mathbf{1}_{\{t\geq 0\}}.$$

Nous définissons la partie entière $\lfloor x \rfloor$ d'un nombre réel x comme le plus grand entier inférieur à x, autrement dit $\lfloor x \rfloor = k \in \mathbb{N}$ si et seulement si $k \leq x < k+1$ (par exemple $\lfloor 3.14 \rfloor = 3$). Nous posons

$$Y = \lfloor X \rfloor + 1.$$

Nous modélisons la durée de vie d'un appareil électrique d'une certaine marque par une loi exponentielle de paramètre inconnu λ. Pour estimer λ, nous nous munissons de n appareils électriques de la marque et nous notons X_i la durée de vie (en heures) du i-ème appareil. Nous supposons

$$X_1, \ldots, X_n \sim_{i.i.d.} X.$$

N'ayant pas la possibilité d'observer les appareils en continu mais seulement une fois par heure, nous avons accès seulement à la durée de vie à l'heure près, c'est-à-dire que nous observons $Y_i = \lfloor X_i \rfloor + 1$ et non X_i. L'objectif de l'exercice est de proposer un estimateur de λ pour lequel nous pouvons calculer un intervalle de confiance.

1. Soit F_X la fonction de répartition de X. Montrer que, pour tout $t \geq 0$,

$$F_X(t) = 1 - e^{-\lambda t}.$$

Solution: F_X est une primitive de f_X qui tend vers 0 quand t tend vers $-\infty$ et vers 1 quand t tend vers $+\infty$, c'est donc bien la fonction de répartition de X.

2. Montrer que Y suit une loi géométrique de paramètre $p = 1 - e^{-\lambda}$. On rappelle que, pour une loi géométrique de paramètre p, $\mathbb{E}[Y] = 1/p$, $\mathrm{Var}(Y) = (1-p)/p^2$.

Solution:

$$\begin{aligned}
\mathbb{P}(Y = k) &= \mathbb{P}(\lfloor X \rfloor = k-1) = \mathbb{P}(k-1 \leq X < k) = F_X(k) - F_X(k-1) \\
&= 1 - e^{-\lambda k} - (1 - e^{-\lambda(k-1)}) = e^{-\lambda(k-1)} - e^{-\lambda k} \\
&= (1 - e^{-\lambda})e^{-\lambda(k-1)}.
\end{aligned}$$

3. Proposer un estimateur des moments $\widehat{\lambda}^{Mom}$ de λ à partir de l'observation de $Y_1, \ldots, Y_n$.

Solution: D'après la question précédente $\mathbb{E}[Y] = 1/(1 - e^{-\lambda})$ donc

$$1 - e^{-\lambda} = \frac{1}{\mathbb{E}[Y]} \text{ i.e. } e^{-\lambda} = 1 - \frac{1}{\mathbb{E}[Y]} \text{ i.e. } \lambda = -\ln\left(1 - \frac{1}{\mathbb{E}[Y]}\right).$$

Nous proposons donc comme estimateur des moments

$$\widehat{\lambda}^{Mom} = -\ln\left(1 - \frac{1}{\frac{1}{n}\sum_{i=1}^n Y_i}\right).$$

4. Montrer que

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^n Y_i - \frac{1}{1 - e^{-\lambda}}\right) \xrightarrow[n \to +\infty]{\mathcal{L}} \mathcal{N}\left(0, \frac{e^{-\lambda}}{(1 - e^{-\lambda})^2}\right).$$

Solution: Nous appliquons le TCL à $Y_1, \ldots, Y_n$ en notant que $\mathrm{Var}(Y) = (1-p)/p^2 = e^{-\lambda}/(1 - e^{-\lambda})^2$.

5. En déduire que l'estimateur des moments $\widehat{\lambda}^{Mom}$ est asymptotiquement normal et donner sa variance asymptotique.

Solution: On applique la méthode Δ à la fonction $g(x) = 1 - \ln(1 - 1/x)$. La fonction g est de classe $\mathcal{C}^1$ sur $]1, +\infty[$ et, pour tout $x > 1$,

$$g'(x) = -\frac{\frac{1}{x^2}}{1 - \frac{1}{x}} = \frac{1}{x(1-x)} \neq 0$$

, donc par 4.

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

et

$$\left(g'\left(1/(1-e^{-\lambda})\right)\right)^2 \frac{e^{-\lambda}}{(1-e^{-\lambda})^2} = \left(\frac{1-e^{-\lambda}}{1-\frac{1}{1-e^{-\lambda}}}\right)^2 \frac{e^{-\lambda}}{(1-e^{-\lambda})^2} = \frac{e^{-\lambda}}{\frac{e^{-2\lambda}}{(1-e^{-\lambda})^2}} = e^{\lambda}(1-e^{-\lambda})^2.$$

La variance asymptotique de $\widehat{\lambda}^{Mom}$ est donc

$$\frac{e^{\lambda}(1-e^{-\lambda})^2}{n} = \frac{p^2}{n(1-p)} = \frac{1}{n\mathbb{E}[Y]^2(1-1/\mathbb{E}[Y])}.$$

6. Montrer que, pour tout $\alpha \in]0,1[$,

$$\lim_{n\to+\infty} \mathbb{P}\left(\lambda \in \left[\widehat{\lambda}^{Mom} \pm q_{1-\alpha/2}\sqrt{\frac{1}{n\widehat{\mu}^2(1-1/\widehat{\mu})}}\right]\right) = 1 - \alpha,$$

où $\widehat{\mu} = \frac{1}{n}\sum_{i=1}^n Y_i$.

7. Nous observons les durées suivantes

i	1	2	3	4	5	6	7	8	9	10
Y_i	0	1	1	1	1	4	2	1	4	4

Donner un intervalle de confiance au niveau 95% de λ.

Solution:

$$\widehat{\lambda}^{Mom} = -\ln\left(1 - \frac{10}{19}\right) \approx 0.74.$$

D'après la question 6., un intervalle de confiance est

$$[1.27; 2.21].$$

Exercice 4 (estimateur des moments, loi continue (/6))

Soit X une variable aléatoire de densité

$$f(x) = \begin{cases} \frac{1}{2a} & \text{si } x \in [0,a] \\ \frac{1}{2(1-a)} & \text{si } x \in]a,1] \\ 0 & \text{sinon.} \end{cases}$$

1. Montrer que X admet des moments de tous ordres et que, pour tout $k \geq 1$,

$$\mathbb{E}[X^k] = \frac{1}{2(k+1)} \left(a^k + \frac{1-a^{k+1}}{1-a} \right).$$

Solution: X est à support dans un intervalle fini, donc elle admet des moments de tous ordres et

$$\begin{aligned} \mathbb{E}[X^k] &= \int_0^a \frac{x^k}{2a} dx + \int_a^1 \frac{x^k}{2(1-a)} dx = \frac{1}{2a} \left[\frac{x^{k+1}}{k+1} \right]_{x=0}^{x=a} + \frac{1}{2(1-a)} \left[\frac{x^{k+1}}{k+1} \right]_{x=a}^{x=1} \\ &= \frac{a^{k+1}}{2a(k+1)} + \frac{1}{2(1-a)(k+1)} \left(1 - a^{k+1} \right) \\ &= \frac{1}{2(k+1)} \left(a^k + \frac{1-a^{k+1}}{1-a} \right) \end{aligned}$$

2. Nous observons $X_1, \ldots, X_n \sim_{i.i.d.} X$. Montrer que

$$\widehat{a}^{Mom} = \frac{2}{n} \sum_{i=1}^n X_i - \frac{1}{2}$$

est un estimateur des moments de a.

Solution: D'après 1., nous avons, pour $k = 1$,

$$\mu_1 = \mathbb{E}[X] = \frac{2a+1}{4} = \frac{a}{2} + \frac{1}{4}.$$

Donc

$$a = 2 \left(\mu_1 - \frac{1}{4} \right) = 2\mu_1 - \frac{1}{2}.$$

Un estimateur des moments est donc

$$\widehat{a}^{Mom} = 2\widehat{\mu}_1 - \frac{1}{2} = \frac{2}{n} \sum_{i=1}^n X_i - \frac{1}{2}.$$

3. Montrer que $\widehat{a}^{Mom}$ est un estimateur convergent de a.

4. Montrer que $\widehat{a}^{Mom}$ est asymptotiquement normal et préciser sa variance asymptotique. On admettra que $\mathrm{Var}(X) = \frac{a^2}{12} - \frac{a}{12} + \frac{5}{48}$.

Solution: D'après 1., X admet un moment d'ordre 2 donc par le TCL,

$$\sqrt{n}\left(\widehat{\mu}_1 - \mu_1\right) \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}(0, \mathrm{Var}(X)).$$

On en déduit

$$\sqrt{n}\left(\widehat{a}^{Mom} - a\right) \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}(0, 4\mathrm{Var}(X)).$$

La variance asymptotique de $\widehat{a}^{Mom}$ est donc

$$\frac{4\mathrm{Var}(X)}{n} = \frac{1}{n}\left(\frac{a^2}{3} - \frac{a}{3} + \frac{1}{12}\right).$$

5. Définir un estimateur convergent $\widehat{\sigma}(X)$ de l'écart-type de X, $\sigma(X)$.

Solution:

Solution 1 On utilise l'écart-type empirique (corrigé ou non) qui est convergent d'après le cours

$$\widehat{\sigma}(X) = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \widehat{\mu}_1)^2}.$$

Solution 2 On définit un estimateur, dit *plug-in*, à partir de $\widehat{a}^{Mom}$,

$$\widehat{\sigma}(X) = \sqrt{\frac{\left(\widehat{a}^{Mom}\right)^2}{12} - \frac{\widehat{a}^{Mom}}{12} + \frac{5}{48}},$$

qui est convergent car $\widehat{a}^{Mom}$ l'est (par la loi des grands nombres et la continuité de la fonction $x \mapsto 2x - \frac{1}{2}$) et la fonction $x \mapsto a^2/12 - a/12 + 5/48$ est continue.

6. En déduire un intervalle de confiance asymptotique de niveau de confiance $1 - \alpha$ pour a.

Solution: D'après 3.

$$\frac{\sqrt{n}}{4\sigma(X)}\left(\widehat{a}^{Mom}-a\right) \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}(0,1).$$

Par le théorème de Slutsky, comme $\widehat{\sigma}(X)$ converge en probabilité vers une constante, nous avons aussi

$$\frac{\sqrt{n}}{4\widehat{\sigma}(X)}\left(\widehat{a}^{Mom}-a\right) \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}(0,1).$$

Donc, pour tout $\alpha \in]0,1[$,

$$\mathbb{P}\left(\left|\frac{\sqrt{n}}{4\widehat{\sigma}(X)}\left(\widehat{a}^{Mom}-a\right)\right| \leq q_{1-\alpha/2}\right) \xrightarrow[n\to+\infty]{} \mathbb{P}(|Z| \leq q_{1-\alpha/2}) = 1-\alpha,$$

avec $Z \sim \mathcal{N}(0,1)$.

Donc un intervalle de confiance asymptotique de niveau de confiance $1-\alpha$ pour a est donné par

$$\left[\widehat{a}^{Mom} \pm \frac{4\widehat{\sigma}(X)}{\sqrt{n}} q_{1-\alpha/2}\right].$$