Étude de Cas : Contrôle Continu N°2

Responsable : H. Li

17 novembre 2020

Consignes

— L'échange en dehors du binôme est interdit (Utilisation de la détection de plagiat Compilatio)
— L'organisation du travail au sein d'un binôme relève de la seule responsabilité de ses membres
— L'épreuve dure 60 minutes
— Vous disposez de 30 minutes supplémentaires pour numériser vos copies et les rassembler en un seul document PDF (en utilisant par exemple https://smallpdf.com/fr/jpeg-en-pdf) et rendre votre travail dans le dépôt Moodle prévu à cet effet ; vous rendrez deux documents nommés NOM1-NOM2.pdf et NOM1-NOM2.R
— Voici le code qui indique les moyens de répondre aux questions :
♣ → sur la copie
♥ → code dans le fichier script
♠ → commentaire dans le fichier script (dont résultats numériques)

Notations

— La fonction log désigne le logarithme néperien.

Indication Pour la construction d'un intervalle de confiance asymptotique en utilisant l'approximation gaussienne, par exemple avec $Z \approx \mathcal{N}(0\,;\,1)$, on utilise le fait que

$$P\left(-C_q \leq Z \leq C_q\right) \simeq 1-\alpha$$

avec C_q le quantile d'ordre $1-\alpha/2$ d'une loi $\mathcal{N}(0;1)$.

Exercice 1. Soit $(X_1, \ldots, X_n)$ un échantillon i.i.d. d'une loi de densité

$$f_X(x \mid \theta) = \sqrt{\frac{\theta}{2\pi x}} \exp\left\{-\frac{\theta x}{2}\right\} \mathbf{1}_{\{x \geq 0\}}$$

avec $\theta > 0$.

1. Calculer l'estimateur $\hat{\theta}_{MV}$ du maximum de vraisemblance de θ. ♣
2. Simuler un échantillon de taille 500 de densité $f_X(x \mid 2)$ par la commande :
   `> x=rchisq(500,1)/2`
   Tracer l'histogramme de x et superposer la densité $f_X(x \mid 2)$. ♥
3. Simuler $M = 5000$ réalisations d'un échantillon de taille $n_1 = 10$ de densité $f_X(x \mid 2)$ ♥. Comparer les qualités et les défauts des estimateurs suivants ♥ ♠ :

$$\hat{\theta}_1 = \sqrt{\frac{3n}{2\sum_{i=1}^n X_i^2}}\,; \qquad \hat{\theta}_2 = \sqrt{\frac{8}{9S_c^2}}\,;$$

$$\hat{\theta}_3 = \frac{n}{\sum_{i=1}^n X_i}\,; \qquad \hat{\theta}_4 = \frac{n-2}{\sum_{i=1}^n X_i}$$

Exercice 2. On conserve la densité de l'Exercice 1, à savoir

$$f_X(x \mid \theta) = \sqrt{\frac{\theta}{2\pi x}} \exp\left\{-\frac{\theta x}{2}\right\} \mathbf{1}_{\{x \geq 0\}}$$

mais on va maintenant construire des intervalles de confiance pour θ sur la base d'échantillons de taille $n_2 = 200$ pour lesquels $\theta = 2$.

1. Que vaut $\mathbf{E}\left[\sqrt{X}\right]$? ♣
2. On peut montrer que $\mathbf{E}[X] = \dfrac{1}{\theta}$ et $\mathrm{Var}(X) = \dfrac{2}{\theta^2}$.
   (a) Appliquer le théorème central limite à l'échantillon i.i.d. $\{X_i, 1 \leq i \leq n\}$ dont la loi de chacune des variables aléatoires est de densité $f_X(x \mid \theta)$. ♣
   (b) Construire un intervalle de confiance asymptotique ♣ basé sur le fait que
   $$\sqrt{n}\frac{\overline{X} - \theta^{-1}}{\overline{X}\sqrt{2}} \xrightarrow{\mathcal{L}} \mathcal{N}(0;1).$$
3. On peut montrer d'autre part que
   $$\sqrt{n}\frac{\theta\overline{X} - 1}{\sqrt{2}} \xrightarrow{\mathcal{L}} \mathcal{N}(0;1)$$
   En déduire un intervalle de confiance asymptotique pour θ. ♣
4. Simuler $M = 10000$ réalisations des intervalles de niveau de confiance $1-\alpha = 90\%$ suivants pour θ sur la base d'échantillons de taille $n_2 = 200$ ♥
   $$IC_1(\theta) = \left[\frac{1}{\overline{X}}\left(1 - \frac{\sqrt{2}C_q}{\sqrt{n}}\right) ; \frac{1}{\overline{X}}\left(1 + \frac{\sqrt{2}C_q}{\sqrt{n}}\right)\right];$$
   $$IC_2(\theta) = \left[\frac{C_1}{\sum_{i=1}^n X_i} ; \frac{C_2}{\sum_{i=1}^n X_i}\right];$$
   $$IC_3(\theta) = \left[\frac{1}{\overline{X}} \times \frac{1}{1 + \dfrac{C_q\sqrt{2}}{\sqrt{n}}} ; \frac{1}{\overline{X}} \times \frac{1}{1 - \dfrac{C_q\sqrt{2}}{\sqrt{n}}}\right];$$
   $$IC_4(\theta) = \left[\frac{\sqrt{2}}{\sqrt{S_c^2 + \dfrac{C_q\overline{X}^2\sqrt{56}}{\sqrt{n}}}} ; \frac{\sqrt{2}}{\sqrt{S_c^2 - \dfrac{C_q\overline{X}^2\sqrt{56}}{\sqrt{n}}}}\right]$$
   où C_q est le quantile d'ordre $1 - \alpha/2$ d'une loi $\mathcal{N}(0;1)$ et C_1 et C_2 sont respectivement les quantiles d'ordre $\alpha/2$ et $1 - \alpha/2$ d'une loi $\chi^2(n_2)$.
5. Comparer les taux de couverture (pourcentage de fois où l'intervalle contient la vraie valeur de θ) et les longueurs moyennes des intervalles $IC_j(\theta)$, $1 \leq j \leq 4$. Commenter. ♥ ♠