PC 9 : intervalles de confiance

1 Exercice corrigé

Exercice 1. (Loi normale) Soient $x_1, \dots, x_n$ des réalisations indépendantes et de même loi de densité

$$f_\theta(x) = \frac{1}{\sqrt{2\pi\theta^2}} \exp\left(-\frac{x^2}{2\theta^2}\right) \qquad \text{avec } \theta > 0 \text{ inconnu.}$$

1. On veut estimer $\tau = \theta^2$. Proposer un estimateur $\hat{\tau}$ de τ par la méthode des moments et étudier sa loi.
2. Construire un intervalle de confiance au niveau $1-\alpha$ de la forme $[S_1, S_2]$ tel que
$$\mathbb{P}(\tau < S_1) = \mathbb{P}(\tau > S_2) = \alpha/2.$$
3. Donner la loi asymptotique de $\hat{\tau}$ et en déduire un intervalle de confiance asymptotique.
4. Application numérique : Lorsque $n = 10$, $\hat{\tau} = 2$ et $\alpha = 0.05$, comparer l'intervalle de confiance obtenu à la question 2 (non asymptotique) avec celui obtenu à la question 3 (asymptotique). Lequel est préférable. Quelle conclusion si $n = 1000$?

Solution.

1. Les X_i suivent une loi normale $\mathcal{N}(0, \theta^2)$. On sait donc que $\mathrm{Var}(X_i) = \mathbb{E}(X_i^2) = \theta^2 = \tau$. Pour estimer τ, on propose l'estimateur suivant :
$$\hat{\tau} = \frac{1}{n} \sum_{i=1}^n X_i^2.$$
On a alors $\frac{n\hat{\tau}}{\tau} = \sum_{i=1}^n \left(\frac{X_i}{\theta}\right)^2$. Or comme pour tout i, $\frac{X_i}{\theta} \sim \mathcal{N}(0,1)$ et comme les v.a. X_i sont i.i.d., on en déduit que
$$\frac{n\hat{\tau}}{\tau} \sim \chi^2_n.$$
Cette caractérisation de la loi de $\hat{\tau}$ suffit, cependant on peut remarquer qu'une loi du χ^2_n est aussi une loi $\gamma(\frac{n}{2}, \frac{1}{2})$. Ainsi,
$$\hat{\tau} \sim \Gamma\left(\frac{n}{2}, \frac{n}{2\tau}\right).$$
2. On note $F_{\chi^2_n}$ la f.d.r. de la loi du χ^2_n et $F^{-1}_{\chi^2_n}$ la fonction quantile (qui est ici l'inverse de $F_{\chi^2_n}$ car cette dernière est une bijection). On peut alors écrire :
$$\mathbb{P}\left(F^{-1}_{\chi^2_n}(\alpha/2) \le \frac{n\hat{\tau}}{\tau} \le F^{-1}_{\chi^2_n}(1-\alpha/2)\right) = 1-\alpha,$$
donc
$$\mathbb{P}\left(\frac{n\hat{\tau}}{F^{-1}_{\chi^2_n}(1-\alpha/2)} \le \tau \le \frac{n\hat{\tau}}{F^{-1}_{\chi^2_n}(\alpha/2)}\right) = 1-\alpha.$$
En posant
$$S_1 = \frac{n\hat{\tau}}{F^{-1}_{\chi^2_n}(1-\alpha/2)} \quad \text{et} \quad S_2 = \frac{n\hat{\tau}}{F^{-1}_{\chi^2_n}(\alpha/2)},$$
on en déduit que $\mathbb{P}(\tau < S_1) = \mathbb{P}(\tau > S_2) = \alpha/2$:
$$\begin{aligned}\mathbb{P}\big(\tau < S_1\big) &= \mathbb{P}\Big(\tau < \frac{n\hat{\tau}}{F^{-1}_{\chi^2_n}(1-\alpha/2)}\Big) = \mathbb{P}\Big(F^{-1}_{\chi^2_n}(1-\alpha/2) < \frac{n\hat{\tau}}{\tau}\Big) \\ &= 1 - F_{\chi^2_n}\big(F^{-1}_{\chi^2_n}(1-\alpha/2)\big) = \alpha/2.\end{aligned}$$
$$\mathbb{P}\big(\tau > S_2\big) = \mathbb{P}\Big(\tau > \frac{n\hat{\tau}}{F^{-1}_{\chi^2_n}(\alpha/2)}\Big) = \mathbb{P}\Big(F^{-1}_{\chi^2_n}(\alpha/2) > \frac{n\hat{\tau}}{\tau}\Big) = F_{\chi^2_n}\big(F^{-1}_{\chi^2_n}(\alpha/2)\big) = \alpha/2.$$

3. Les X_i suivent une loi normale et admettent donc des moments à tous les ordres, en particulier à l'ordre 4. Les v.a. X_i^2 sont donc dans $\mathbb{L}^2$ et satisfont le TLC :
$$\sqrt{n}\frac{\hat{\tau}-\tau}{\sqrt{\operatorname{Var}(X_1^2)}} = \sqrt{n}\frac{\frac{1}{n}\sum_{i=1}^n X_i^2 - \theta^2}{\sqrt{\operatorname{Var}(X_1^2)}} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1).$$
De plus, $\operatorname{Var}(X_1^2) = \theta^4 \operatorname{Var}((X_1/\theta)^2) = 2\theta^4 = 2\tau^2$ car la variance d'une χ_1^2 est égale à 2.
On peut alors construire un IC asymptotique pour τ. La convergence en loi vue plus haut donne, en notant Φ la fonction de répartition de la $\mathcal{N}(0,1)$,
$$\begin{aligned}1-\alpha &= \lim_{n\to\infty} \mathbb{P}\left(\Phi^{-1}(\alpha/2) \le \sqrt{n}\frac{\hat{\tau}-\tau}{\sqrt{2\tau^2}} \le \Phi^{-1}(1-\alpha/2)\right) \\ &= \lim_{n\to\infty} \mathbb{P}\left(1+\Phi^{-1}(\alpha/2)/\sqrt{n/2} \le \hat{\tau}/\tau \le 1+\Phi^{-1}(1-\alpha/2)/\sqrt{n/2}\right) \\ &= \lim_{n\to\infty} \mathbb{P}\left(\frac{\hat{\tau}}{1+\Phi^{-1}(1-\alpha/2)/\sqrt{n/2}} \le \tau \le \frac{\hat{\tau}}{1-\Phi^{-1}(1-\alpha/2)/\sqrt{n/2}}\right),\end{aligned}$$
où on a utilisé que $\Phi^{-1}(\gamma) = -\Phi^{-1}(1-\gamma)$.
4. Pour $n = 10$, on a $F_{\chi_n^2}^{-1}(\alpha/2) \simeq 3.25$, $F_{\chi_n^2}^{-1}(1-\alpha/2) \simeq 20.48$ et $\Phi^{-1}(1-\alpha/2) \simeq 2$. L'intervalle non asymptotique (question 2) associé est $[0.97, 6.16]$ (environ), alors que l'intervalle asymptotique est $[1.05, 18.95]$ (environ). On remarque que l'intervalle de confiance asymptotique est différent de l'intervalle de confiance non asymptotique. Ceci est normal car $n = 10$ est trop petit pour que l'approche asymptotique soit valide (la convergence n'a pas eu lieu).
Lorsque $n = 1000$, on trouve les intervalles $[1.836, 2.188]$ (non asymptotique) et $[1.839, 2.192]$. La différence a quasi disparu, les intervalles sont quasi identiques, les deux approches sont valides.
Remarque : On préférera l'IC non-asymptotique, car il a été construit avec la vraie loi, et est donc plus précis. On a recours aux intervalles asymptotiques, lorsqu'on n'arrive pas à construire d'intervalle non asymptotiques ou lorsque la taille d'échantillon est vraiment très grande.

2 Intervalles de confiance

Exercice 2. Soit la variable aléatoire
$$Y = \mathbb{1}_{\{\theta>\xi\}},$$
où $\theta \in \mathbb{R}$ et ξ est une variable aléatoire de loi $\mathcal{N}(0,1)$. On dispose d'un échantillon $Y_1, \dots, Y_n$ des réalisations i.i.d. de Y.

1. Soit Φ la fonction de répartition de la loi $\mathcal{N}(0,1)$. Montrer que $\hat{\theta}_n = \Phi^{-1}(\frac{1}{n}\sum_{i=1}^n Y_i)$ est l'estimateur de maximum de vraisemblance de θ. Cet estimateur est-il consistant ?
2. Chercher la loi limite de l'estimateur $\hat{\theta}_n$ quand $n \to \infty$.
3. Soit $0 < \alpha < 1$. Proposer un intervalle de confiance asymptotique de niveau $1-\alpha$.

Exercice 3. On effectue une enquête, durant une épidémie de grippe, dans le but de connaître la proportion p de personnes présentant ensuite des complications graves. On observe un échantillon représentatif de 400 personnes et pour un tel échantillon 40 personnes ont présenté des complications.

1. Utiliser l'inégalité de Bienamyé-Tchebychev pour construire un intervalle de confiance pour p au niveau 95%.
2. On désire que la valeur estimée $\widehat{p}$ diffère de la proportion inconnue exacte p de moins de 0.005 avec une probabilité d'au moins 95%. Quel sera l'effectif d'un tel échantillon ?
3. Selon cette approche, quel est le niveau de confiance du même intervalle qu'à la question précédente avec un effectif de $n = 400$? Quelle conclusion peut-on en tirer ?

Exercice 4. (Intervalles de confiance asymptotiques) Soit $(X_n)_{n\ge1}$ une suite de variables aléatoires indépendantes et de même loi. On suppose que X_1 est de carré intégrable, de moyenne m et de variance $\sigma^2 > 0$. Notons
$$\bar{X}_n = \frac{1}{n}\sum_{k=1}^n X_k, \qquad \widehat{\sigma}_n^2 = \frac{1}{n-1}\sum_{k=1}^n (X_k - \bar{X}_n)^2.$$

1. L'estimateur $\widehat{\sigma}_n^2$ de σ est-il sans biais ?
   Indication. On pourra démontrer que $\frac{n-1}{n}\widehat{\sigma}_n^2 = \frac{1}{n}\sum_{k=1}^n X_k^2 - \bar{X}_n^2$.
2. Proposer un intervalle de confiance asymptotique pour m au niveau $1-\alpha$.
3. Considérons un échantillon $(X_1, \ldots, X_n)$ représentant des temps d'attente d'un transport, que l'on suppose indépendantes de même loi exponentielle de paramètre $\lambda > 0$ inconnu. Donner un intervalle de confiance asymptotique pour $\theta = \frac{1}{\lambda}$ au niveau $1-\alpha$. Comment obtenir un intervalle de confiance asymptotique pour λ au niveau $1-\alpha$?

Exercice 5. Soient $X_1, \ldots, X_n$ des variables aléatoires i.i.d. d'une loi dont la densité par rapport à la mesure de Lebesgue sur $\mathbb{R}$ est donnée par

$$f(x;\theta) = \frac{2}{\sqrt{\pi\theta}} \exp(-x^2/\theta)\, \mathbb{1}_{\{x>0\}},$$

où $\theta > 0$ est un paramètre inconnu. On observe une réalisation $(x_1, \ldots, x_n)$ du vecteur aléatoire $(X_1, \ldots, X_n)$. On désigne par α un réel donné dans $[0,1]$ et on note $\hat{m}_2 = \frac{1}{n}\sum_{i=1}^n X_i^2$ le moment empirique d'ordre 2.

1. Déterminer l'estimateur du maximum de vraisemblance $\hat{\theta}$ de θ.
2. Déterminer la loi de la variable $X_1/\sqrt{\theta}$. En déduire la loi de $n\hat{\theta}/\theta$.
3. Trouver des réels a et b tels que $[\hat{\theta}/a, \hat{\theta}/b]$ soit un intervalle de confiance de niveau $1-\alpha$ pour θ.

Exercice 6 (Modèle de Poisson)**.** Soit $(X_1, \ldots, X_n)$ un échantillon i.i.d. de la loi de Poisson $\mathcal{P}(\lambda)$ de paramètre $\lambda > 0$ inconnu. Notons $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$ et $\hat{\sigma}_n^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X}_n)^2$. On rappelle que $\bar{X}_n$ est un estimateur consistant et asymptotiquement normal, plus précisément on a $\sqrt{n}\frac{\bar{X}_n-\lambda}{\sqrt{\lambda}} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1)$. De même, $\hat{\sigma}_n^2$ est un estimateur consistant de λ.

1. En utilisant le lemme de Slutsky, montrer que $\hat{\sigma}_n^2$ est asymptotiquement normal. (On utilisera, sans le démontrer, que $\mathbb{E}_\lambda[(X_1-\lambda)^4] = \lambda + 3\lambda^2$).
2. Quel estimateur de λ est à privilégier, $\bar{X}_n$ ou $\hat{\sigma}_n^2$?
3. Montrer qu'on peut obtenir les résultats de convergence suivants
   (i) $\sqrt{n}\dfrac{\bar{X}_n - \lambda}{\sqrt{\bar{X}_n}} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1)$
   (ii) $\sqrt{n}\dfrac{\bar{X}_n - \lambda}{\sigma_n} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1)$
   (iii) $\sqrt{n}\left(g(\bar{X}_n) - g(\lambda)\right) \xrightarrow{\mathcal{L}} \mathcal{N}(0,1)$ pour un choix approprié de la fonction g à préciser.
4. Déterminer les intervalles de confiances de niveau asymptotique $1-\alpha$ correspondants. Lequel est le meilleur ?

3 Révisions - exercices supplémentaires

Exercice 7 (Détrucage d'une pièce)**.** On dispose d'une pièce pour laquelle la probabilité $p \in]0,1[$ de tomber sur pile n'est pas nécessairement égale à $1/2$ et n'est pas connue. On veut cependant simuler un jeu à deux issues équilibré. On utilise alors l'algorithme suivant (imaginé par von Neumann). On lance la pièce deux fois successivement.
- Si on obtient (P, F) (pile puis face), on considère qu'on gagne.
- Si on obtient (F, P) (face puis pile), on considère qu'on perd.
- Sinon, on relance la pièce deux fois successivement.

On note T la variable aléatoire qui décrit le nombre de lancers après lequel l'algorithme s'arrête, et R la variable aléatoire qui décrit le résultat de l'algorithme (gagné ou perdu).

1. Démontrer que, pour tout $k \geq 1$,

$$\mathbb{P}(T = 2k) = (p^2 + (1-p)^2)^{k1-}2p(1-p)$$

   et en déduire l'algorithme se termine presque sûrement, c'est à dire que $\mathbb{P}(T < 1) = 1$.

2. Démontrer que l'algorithme fait bien gagner ou perdre avec probabilité 1/2 chacun.
3. Combien de lancers faut-il faire en moyenne pour que l'algorithme renvoie un résultat ?

Exercice 8 (Vecteurs gaussiens et espérance conditionnelle)**.** Soit $X = (X_1, X_2, X_3)$ un vecteur gaussien centré de matrice de covariance

$$\Gamma = \begin{pmatrix} 3 & 1 & 0 \\ 1 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

1. Donner la loi de (X_1, X_2) et de X_3. Que peut-on dire de X_3 et de (X_1, X_2) ?
2. Montrer que pour tout $a \in \mathbb{R}$ le vecteur $(X_2, X_2 + aX_1)$ est un vecteur gaussien.
3. En choisissant a de sorte que X_2 et $X_2 + aX_1$ soient indépendants, calculer $E[X_1|X_2]$.

Exercice 9 (Loi d'un produit)**.** Soit (X, Y) un vecteur aléatoire de loi uniforme sur $[0, 1]^2$. Déterminer la loi de $Z = XY$.