Petite Classe 8 - MAP 311 - Groupes 1 et 12

Josselin Garnier

19 juin 2017 - Amphi Curie

Exercice 1. (Suite et fin de l'exercice 3 de la PC 7) On observe un échantillon $(X_1, \dots, X_n)$ de n variables aléatoires i.i.d. de loi exponentielle de paramètre $\theta > 0$ inconnu.

1. Déterminer l'estimateur du maximum de vraisemblance $\widehat{\theta}_n = \theta_n(X_1, \dots, X_n)$ de θ et montrer qu'il converge presque sûrement vers θ lorsque $n \to \infty$.
2. Cet estimateur est-il sans biais ?
   Indication. On rappelle que la densité de la loi $\Gamma(a, \lambda)$ est $\frac{1}{\Gamma(a)} \lambda^a x^{a-1} e^{-\lambda x} \mathbf{1}_{x>0}$.
3. Démontrer que $\sqrt{n}(\widehat{\theta}_n - \theta)$ converge en loi vers une loi qu'on déterminera.
   Indication. On pourra utiliser la "méthode Delta".
4. Construire un intervalle de confiance asymptotique à 95% pour θ.

Exercice 2. Soit $(X_k)_{k\geq 1}$ des variables aléatoires indépendantes de même loi uniforme sur $[0, \theta]$ (avec θ inconnu).

1. Montrer que l'estimateur du maximum de vraisemblance $\widehat{\theta}_n = \widehat{\theta}_n(X_1, \dots, X_n)$ de θ est $\widehat{\theta}_n = \max(X_1, \dots, X_n)$.
2. Montrer que $W_n = n\big(1 - \frac{\widehat{\theta}_n}{\theta}\big)$ converge en loi quand $n \to \infty$ et déterminer sa limite.
3. Construire un intervalle de confiance asymptotique à 95% pour θ.

Exercice 3. (Intervalles de confiance asymptotiques avec le TCL) Soit $(X_n)_{n\geq 1}$ 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 $\sigma^2 > 0$. On pose

$$\overline{X}_n = \frac{X_1 + \cdots + X_n}{n}, \qquad \widehat{\sigma}_n^2 = \frac{1}{n-1} \sum_{k=1}^n (X_k - \overline{X}_n)^2.$$

1. Justifier que $\sqrt{n}\frac{\overline{X}_n - m}{\sigma}$ converge en loi vers $\mathcal{N}(0, 1)$, une gaussienne centrée réduite.
2. En déduire un intervalle de confiance asymptotique pour m au niveau 95% (en supposant σ connu).
3. Montrer que $\widehat{\sigma}_n$ converge presque sûrement vers σ. L'estimateur $\widehat{\sigma}_n^2$ est-il sans biais ?
4. Montrer que
$$\sqrt{n}\frac{\overline{X}_n - m}{\widehat{\sigma}_n} \quad \xrightarrow[n\to\infty]{\text{loi}} \quad \mathcal{N}(0, 1).$$
5. En déduire un intervalle de confiance asymptotique pour m au niveau 95% (en supposant σ inconnu).

Exercice 4. Le but de cet exercice est d'estimer le temps d'attente du RER B, qui suit une loi exponentielle de paramètre $\lambda > 0$ inconnu. Soit $(E_i)_{i\geq 1}$ une suite de variables aléatoires indépendantes de même loi exponentielle de paramètre $\lambda > 0$ inconnu. On pose $\widehat{\lambda}_n = \frac{E_1+\cdots+E_n}{n}$ et

$$\widehat{\sigma}_n^2 = \frac{1}{n-1}\sum_{k=1}^{n}(E_k - \widehat{\lambda}_n)^2.$$

1. Donner un intervalle de confiance asymptotique pour $\frac{1}{\lambda}$ au niveau 95%.
2. Comment obtenir un intervalle de confiance asymptotique pour λ au niveau 95% ?

Exercice 5. (Stabilisation de la variance) On dispose d'un échantillon $X_1, \ldots, X_n$ de variables aléatoires indépendantes de même loi de Bernoulli de paramètre $0 < \theta < 1$.

1. On note $\overline{X}_n = \frac{X_1+\cdots+X_n}{n}$ la moyenne empirique des X_i. Que donnent la loi forte des grands nombres et le TCL?
2. Trouver une fonction g telle que $\sqrt{n}\big(g(\overline{X}_n)-g(\theta)\big)$ converge en loi vers une loi gaussienne centrée réduite.
3. On note $z_{1-\alpha/2}$ le quantile d'ordre $1-\alpha/2$ de la loi normale ($\mathbb{P}(Z \geq z_{1-\alpha/2}) = \alpha/2$ si Z est une loi normale centrée réduite). En déduire un intervalle de confiance asymptotique $\widehat{I}_{n,\alpha}$ (qui dépend de $z_{1-\alpha/2}$, n et $\overline{X}_n$) tel que $\lim_{n\to\infty}\mathbb{P}(\theta \in \widehat{I}_{n,\alpha}) = 1-\alpha$.

Exercice 6. Soient $(X_i)_{1\leq i\leq n}$ des variables aléatoires réelles indépendantes de même loi, de carré intégrable, d'espérance m et de variance σ^2. On suppose que la moyenne empirique $\overline{X}_n$ et la variance empirique V_n, définies par

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i, \qquad V_n = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2,$$

sont indépendantes. Le but de cet exercice est de démontrer que la loi de X_i est une loi gaussienne $\mathcal{N}(m, \sigma^2)$.

1. On suppose dans les questions 1 à 4 que $m = 0$. Calculer $\mathbb{E}((n-1)V_n)$ en fonction de σ^2. Montrer que pour tout $t \in \mathbb{R}$:
$$\mathbb{E}\big((n-1)V_n e^{itn\overline{X}_n}\big) = (n-1)\psi(t)^n\sigma^2,$$
où on note $\psi(t)$ la fonction caractéristique de X_1.
2. En développant V_n dans l'égalité précédente, vérifier que :
$$\mathbb{E}\big((n-1)V_n e^{itn\overline{X}_n}\big) = -(n-1)\psi''(t)\psi(t)^{n-1} + (n-1)\psi'(t)^2\psi(t)^{n-2}.$$
3. En déduire que, sur un voisinage ouvert de 0, ψ est solution de l'équation différentielle :
$$\frac{\psi''}{\psi} - \left(\frac{\psi'}{\psi}\right)^2 = -\sigma^2, \qquad \psi(0) = 1, \qquad \psi'(0) = 0.$$

4. En déduire que la loi des variables X_i est la loi gaussienne $\mathcal{N}(0, \sigma^2)$.
   Indication : Considérer $\phi = \psi'/\psi$.

5. Que peut-on dire si l'on ne suppose plus $m = 0$?

Exercice 7. 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. Donner un intervalle de confiance pour p au risque 5%.

2. On désire que la valeur estimée $\widehat{p}$ diffère de la proportion inconnue exacte p de moins de 0,005 avec une probabilité égale à 95%. Quel sera l'effectif d'un tel échantillon ?

3. Quel devrait être le risque pour obtenir le même intervalle qu'à la question précédente en conservant l'effectif $n = 400$? Quelle conclusion peut-on en tirer ?

Pour aller plus loin

Rappel (théorème de Cochran, extension de la proposition 6.2.8 du poly). Soit $\boldsymbol{X}$ un vecteur colonne aléatoire de $\mathbb{R}^n$ de loi $\mathcal{N}(\boldsymbol{m}, \sigma^2\mathbf{I}_n)$ (avec $\boldsymbol{m} \in \mathbb{R}^n$, $\sigma > 0$) et $\mathbb{R}^n = E_1 \oplus \cdots \oplus E_p$ une décomposition de $\mathbb{R}^n$ en somme directe de p sous-espaces vectoriels orthogonaux de dimensions $d_1, \ldots, d_p$ avec $d_1 + \cdots + d_p = n$. Soit $\mathbf{P}_k$ la matrice du projecteur orthogonal sur E_k et $\boldsymbol{Y}_k = \mathbf{P}_k\boldsymbol{X}$ la projection orthogonale de $\boldsymbol{X}$ sur E_k. Alors :

1. Les vecteurs aléatoires $(\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_p)$ sont indépendants et $\boldsymbol{Y}_k$ suit la loi $\mathcal{N}(\mathbf{P}_k\boldsymbol{m}, \sigma^2\mathbf{P}_k)$.

2. Les variables aléatoires $(\|\boldsymbol{Y}_i - \mathbf{P}_i\boldsymbol{m}\|^2)_{1\leq i\leq p}$ sont indépendantes et $\|\boldsymbol{Y}_k - \mathbf{P}_k\boldsymbol{m}\|^2/\sigma^2$ suit la loi $\chi^2_{d_k}$.

Exercice 8. On considère que la réponse d'un appareil de mesure à un signal déterministe ξ est égale à $a\xi$ plus un bruit gaussien centré de variance b, où $(a, b) \in \mathbb{R} \times \mathbb{R}^{+*}$. On se propose d'étalonner l'appareil (c'est-à-dire estimer les valeurs de a et b) en envoyant une suite $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$ de signaux connus. On note $Y_i = ax_i + \sqrt{b}U_i$ la réponse au i-ième signal où on suppose que les coordonnées du vecteur $\boldsymbol{U} = (U_1, U_2, \ldots, U_n)$ sont des variables aléatoires gaussiennes centrées réduites indépendantes. On note $\boldsymbol{Y} = (Y_1, Y_2, \ldots, Y_n)$,

$$\widehat{A}_n = \frac{\sum_{i=1}^n x_iY_i}{\sum_{i=1}^n x_i^2} \qquad \text{et} \qquad \widehat{B}_n = \frac{\sum_{i=1}^n (Y_i - x_i\widehat{A}_n)^2}{n-1}.$$

1. Donner la loi de $\widehat{A}_n$. A quelle condition sur la suite $(x_i)_{i\geq 1}$ a-t-on $\mathbb{E}((\widehat{A}_n - a)^2) \to 0$ lorsque $n \to \infty$?

On complète $\boldsymbol{e}_1 = \frac{\boldsymbol{x}}{\|\boldsymbol{x}\|}$ en une base orthonormée $(\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n)$ de $\mathbb{R}^n$. Notons $\mathbf{P}$ la projection orthogonale sur $E_1 = \mathrm{Vect}(\boldsymbol{e}_1)$ et $\mathbf{Q}$ la projection orthogonale sur $E_2 = \mathrm{Vect}(\boldsymbol{e}_2, \ldots, \boldsymbol{e}_n)$.

2. Déterminer $\mathbf{P}\boldsymbol{Y}$ et $\mathbf{Q}\boldsymbol{Y}$. En déduire que $\widehat{A}_n$ et $\widehat{B}_n$ sont des variables aléatoires réelles indépendantes. Donner l'espérance et la variance de $\widehat{B}_n$.

3. Montrer qu'il existe une constante c (dépendant de $\boldsymbol{x}$) telle que la variable aléatoire $c\frac{\widehat{A}_n - a}{\sqrt{\widehat{B}_n}}$ suive une loi de Student.

4. Donner un intervalle de confiance à 95% pour le paramètre a, suivant que l'on connaît la valeur de b ou non.