MAP 311: Aléatoire PC 8

Marc Lelarge

20 juin 2016

Exercice 1

Soit $(X_n)_{n\geq 1}$ une suite de v.a. indépendantes et de même loi. On suppose que X_1 est de carré intégrable, de moyenne m et de variance $\sigma^2 > 0$. On définit:

$$\hat{m}_n = \frac{X_1 + \cdots + X_n}{n}, \quad \hat{\sigma}_n^2 = \frac{1}{n-1}\sum_{k=1}^n (X_k - \hat{m}_n)^2.$$

1. Jusitifier que $\sqrt{n}\frac{\hat{m}_n - m}{\sigma}$ converge en loi vers $\mathcal{N}(0,1)$.
2. En déduire un intervalle de confiance asymptotique pour m au niveau 95% (en supposant σ connu).
3. Montrer que $\hat{\sigma}_n$ converge presque sûrement vers σ. L'estimateur $\hat{\sigma}_n^2$ est-il sans biais? On pourra montrer que $\frac{n-1}{n}\hat{\sigma}_n^2 = \frac{1}{n}\sum_{k=1}^n X_k^2 - \hat{m}_n^2$.
4. Montrer que
$$\sqrt{n}\frac{\hat{m}_n - m}{\hat{\sigma}_n} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1).$$
5. En déduire un intervalle de confiance asymptotique pour m au niveau 95%.

Exercice 2:

Le but de cet exercice est d'estimer un temps d'attente qui suit une loi exponentielle de paramètre $\lambda > 0$ inconnu. Soit $(E_i)_{i\geq 1}$ une suite de v.a.r. i.i.d. de loi exponentielle de paramètre $\lambda > 0$ inconnu. On pose $\hat{m}_n = \frac{E_1 + \cdots + E_n}{n}$ et

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

1. Donner un intervalle de confiance asymptotique pour $\frac{1}{\lambda}$ au niveau 95%. En déduire un intervalle de confiance asymptotique pour λ au niveau 95%
2. Appliquer la méthode delta avec la fonction $f(x) = 1/x$.

Exercice 3: Soit $(X_n)_{n\geq 1}$ des v.a.r. i.i.d. de loi uniforme sur $[0,\theta]$.

1. Étudier la convergence, le biais, la variance, l'écart quadratique moyen, et la fluctuation asymptotique de l'estimateur $m_n = 2(X_1 + \cdots + X_n)/n$ de θ. Construire un intervalle de confiance asymptotique;
2. Montrer que l'estimateur $M_n = \max(X_1, \ldots, X_n)$ de θ a pour densité $(n/\theta)(x/\theta)^{n-1}\mathbf{1}_{[0,\theta]}(x)$ puis calculer $\mathbb{E}(M_k)$ et $Var(M_k)$. Comparer avec l'estimateur moyenne empirique m_n;

3. Montrer que $M_n \xrightarrow{\mathbb{P}} \theta$. Montrer de deux manières différentes que la convergence est p.s.;
4. Montrer que $W_n = n(M_n/\theta - 1)$ converge en loi quand $n \to \infty$, et déterminer la limite. Construire un intervalle de confiance asymptotique.
5. Montrer que M_n est l'estimateur de maximum de vraisemblance de θ.

Exercice 4: Théorème de Cochran

Soit X un vecteur colonne aléatoire de $\mathbb{R}^n$ de loi $\mathcal{N}(m, \sigma^2 I_n)$, 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 $Y_k := \mathbf{P}_k X$ la projection orthogonale de X sur E_k.

1. Montrer que les projections $(Y_1, \ldots, Y_p)$ sont indépendantes et $Y_k \sim \mathcal{N}(\mathbf{P}_k m, \sigma^2 \mathbf{P}_k)$;
2. Monter que $\|Y_1 - \mathbf{P}_1 m\|^2, \ldots, \|Y_p - \mathbf{P}_p m\|^2$ sont indépendantes et $\sigma^{-2}\|Y_k - \mathbf{P}_k m\|^2 \sim \chi^2(d_k)$.

Exercice 5: Échantillons gaussiens

Soit $(X_1, \ldots, X_n)$ un échantillon de loi $\mathcal{N}(m, \sigma^2)$. On lui associe la *moyenne empirique* et la *variance empirique* définies respectivement par

$$\overline{X}_n = \frac{1}{n}\sum_{k=1}^n X_k \qquad \text{et} \qquad \sigma_n^2 = \frac{1}{n-1}\sum_{k=1}^n (X_k - \overline{X}_n)^2.$$

Monter que les variables aléatoires $\overline{X}_n$ et σ_n^2 sont indépendantes avec

$$\overline{X}_n \sim \mathcal{N}\left(m, \frac{\sigma^2}{n}\right) \qquad \text{et} \qquad \frac{(n-1)}{\sigma^2}\sigma_n^2 \sim \chi^2(n-1).$$

Exercice 6: Test d'aquéquation du chi-deux

Il s'agit d'un test non-paramétrique classique.

1. Soient $p = (p_1, \ldots, p_k)$, $q = (q_1, \ldots, q_k)$ des lois discrètes avec $p_i > 0$ pour tout $1 \leq i \leq k$, et
$$D(p, q) := \sum_{i=1}^k \frac{(p_i - q_i)^2}{p_i}.$$
Montrer que $D(p, q) \geq 0$ avec égalité ssi $p = q$. S'agit-il d'une distance?
2. Soit $X_1, \ldots, X_n$ un échantillon d'une loi discrète inconnue $q = (q_1, \ldots, q_k)$ sur $\{1, \ldots, k\}$. On souhaite savoir si cette loi q est égale à une loi discrète de référence $p = (p_1, \ldots, p_k)$ (qui est connue). Pour cela, on introduit les hypothèses statistiques antagonistes:
$$H_0 : p = q, \quad H_1 : p \neq q.$$
On introduit la statistique de test $S_n = S(X_1, \ldots, X_n)$ définie par
$$S_n := nD(p, \widehat{q}) = n\sum_{i=1}^k \frac{(p_i - \widehat{q}_i)^2}{p_i} = n\sum_{i=1}^k \frac{(n_i - N_i)^2}{n_i}$$
où

- $\widehat{q} = (N_1/n, \dots, N_k/n)$ est l'estimateur empirique de q;
- $N_i := \mathrm{card}\{1 \leq m \leq n : X_m = i\} = \sum_{m=1}^{n} \mathbf{1}_{\{X_m=i\}}$ est l'effectif de i dans l'échantillon;
- $n_i := np_i$ est l'effectif théorique sous H_0.

En pratique, on calcule les n_i et les N_i (comptage) puis S_n. Montrer que

- si H_0 est fausse alors $S_n \to +\infty$ presque sûrement quand $n \to \infty$;
- si H_0 est vraie alors $S_n \to \chi^2(k-1)$ en loi quand $n \to \infty$.

Indication: établir que

$$S_n = \left\| \frac{1}{\sqrt{n}} \sum_{m=1}^{n} V_m \right\|_2^2 \quad \text{où} \quad V_{m,i} := \frac{\mathbf{1}_{\{X_m=i\}} - p_i}{\sqrt{p_i}}.$$

3. Fixons un seuil de tolérance $0 < \alpha < 1$, typiquement $\alpha = 5/100$, appelé *niveau du test*, et considérons la région $\mathcal{R}_\alpha = [0, a]$ où a est le quantile $1 - \alpha$ de la loi $\chi^2(k-1)$, appelée *région d'acceptation du test*. Au vu de $X_1, \dots, X_n$, on décide comme suit:

   - Si $S_n \in \mathcal{R}_\alpha$, on accepte l'hypothèse H_0;
   - Si $S_n \in \mathcal{R}_\alpha$, on rejette l'hypothèse H_0.

   De manière résumée la décision prise avec le test est $T_n := H_{\mathbf{1}_{S_n \notin \mathcal{R}_\alpha}}$. Montrer que

   - Si H_0 est vraie alors la probabilité de rejeter à tort H_0 tend vers α quand $n \to \infty$;
   - Si H_0 est fausse, alors la probabilité d'accepter à tort H_0 tend vers 0 quand $n \to \infty$.

   On parle d'*erreurs de première et de seconde espèce du test* ($\to$ niveau et puissance).