Statistiques

Décembre 2022, 1h25

Préambule : Le sujet est composé de cinq exercices indépendants. La qualité de la rédaction sera prise en compte. Toutes les réponses seront données sur la copie (ne pas rendre le sujet).

Exercice 1

1. Donner la définition de la convergence en moyenne quadratique (ou convergence L_2).
2. Expliquer les liens entre ce mode de convergence et le compromis biais/variance en théorie de l'estimation.
3. On considère $X_1, \ldots, X_n$ n variables aléatoires indépendantes qui suivent une loi de Poisson de paramètre $\lambda > 0$.
   (a) Calculer l'estimateur des moments $\widehat{\lambda}$ de λ.
   (b) Calculer le biais et la variance de $\widehat{\lambda}$.
   (c) Est-ce que $\widehat{\lambda}$ converge en moyenne quadratique vers λ ? Justifier.

Exercice 2

Soit $(X_n)_n$ une suite de variables aléatoires dont la loi est donnée par

$$\mathbf{P}(X_n = 1) = 1 - \frac{1}{n+1} \quad \text{et} \quad \mathbf{P}(X_n = \sqrt{n}) = \frac{1}{n+1}.$$

1. Est-ce que $(X_n)_n$ converge en probabilités vers 0 ? Justifier.
2. Est-ce que $(X_n)_n$ converge en probabilités vers 1 ? Justifier.
3. Est-ce que $(X_n)_n$ converge dans L_1 ? Si oui, justifier en indiquant la limite.

Exercice 3

Soit $X_1, \ldots, X_{n_1}$ n_1 variables aléatoires indépendantes de loi $\mathcal{N}(\mu_1, \sigma^2)$.

1. Enoncer le théorème de Cochran.
2. En déduire un intervalle de confiance de niveau 95% pour μ_1.
3. Même question pour un intervalle de confiance de niveau 90% pour σ^2.
4. On considère de plus n_2 variables aléatoires indépendantes $Y_1, \ldots, Y_{n_2}$ (et également indépendantes des $X_i, i = 1, \ldots, n_1$) de loi $\mathcal{N}(\mu_2, \sigma^2)$. On pose

$$S^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

   avec

$$S_1^2 = \frac{1}{n_1 - 1} \sum_{i=1}^{n_1} (X_i - \bar{X}_{n_1})^2 \quad \text{et} \quad S_2^2 = \frac{1}{n_2 - 1} \sum_{i=1}^{n_2} (Y_i - \bar{Y}_{n_2})^2.$$

   (a) Quelle est la loi de $(n_1 + n_2 - 2)S^2/\sigma^2$? Justifier.
   (b) Quelle est la loi de $\bar{X}_{n_1} - \bar{Y}_{n_2}$? Justifier.
   (c) Déduire des deux questions précédentes un intervalle de confiance de niveau $1 - \alpha$ pour $\mu_1 - \mu_2$.
   (d) On suppose maintenant que les deux populations n'ont pas la même variance. C'est-à-dire que les X_i suivent une loi $\mathcal{N}(\mu_1, \sigma_1^2)$ et les Y_j une loi $\mathcal{N}(\mu_2, \sigma_2^2)$. Calculer un intervalle de confiance de niveau $1 - \alpha$ pour σ_1^2/σ_2^2.

Exercice 4

On considère $X_1, \ldots, X_n$ n variables aléatoires dont la loi admet pour densité

$$f_\mu(x) = \frac{1}{\mu} x^{-(\frac{1}{\mu}+1)}, \quad x \geq 1, \quad 0 < \mu < 1.$$

1. Calculer l'espérance de X_1.
2. Montrer que $\log(X_1)$ suit une loi exponentielle dont on précisera le paramètre.
3. Calculer l'estimateurs du maximum de vraisemblance de μ. On le notera $\widehat{\mu}$.
4. $\widehat{\mu}$ est-il VUMSB ?

Exercice 5

On considère $x_1, \dots, x_n$ n nombres réels fixés entre 0 et 1. Soit $Y_1, \dots, Y_n$ n variables aléatoires indépendantes telles que Y_i suit une loi $\mathcal{N}(\beta x_i, 1)$ où $\beta \in \mathbb{R}$ désigne le paramètre inconnu du modèle.

1. Soit $y_1, \dots, y_n$ des réels. Calculer la vraisemblance $L(y_1, \dots, y_n; \beta)$ du modèle.
2. En déduire que l'estimateur du maximum de vraisemblance est donné par
$$\widehat{\beta} = \frac{\sum_{i=1}^n x_i Y_i}{\sum_{i=1}^n x_i^2}.$$
3. Calculer le biais et la variance de $\widehat{\beta}$.
4. On suppose dans cette question que $x_i = i/n, i = 1, \dots, n$.
   (a) Est-ce que l'estimateur du maximum de vraisemblance converge en moyenne quadratique vers β ?
   (b) Est-ce que l'estimateur du maximum de vraisemblance converge en probabilité vers β ?

   Pour ces deux dernières questions, on pourra utiliser que
$$\sum_{i=1}^n i^2 = \frac{n(n+1)(2n+1)}{6}.$$