Série 10

Exercice 1. Soit l'échantillon aléatoire $Y_1, Y_2, \ldots, Y_n \overset{iid}{\sim} \mathcal{U}(0,\theta)$, où $\theta > 0$.

(i). On propose d'estimer θ par la valeur maximale de l'échantillon aléatoire : $M_n = \max(Y_1, Y_2, \ldots, Y_n)$. **Remarque** : on peut montrer que c'est l'estimateur du maximum de vraisemblance.

  (a) Trouver la fonction de densité de M_n.

  (b) Calculer le biais de l'estimateur M_n.

  (c) Proposer un estimateur non-biaisé de θ, $\tilde{\theta}_1$ de la forme $\tilde{\theta}_1 = aM_n$ pour $a \in \mathbb{R}$. Quelle est sa variance ?

(ii). On propose un deuxième estimateur de θ, notons $\hat{\theta}_2$, basé sur la méthode des moments.

  (a) trouver $\hat{\theta}_2$.

  (b) Calculer son erreur quadratique moyenne.

(iii). Parmi les estimateurs $\hat{\theta}_1$ et $\hat{\theta}_2$, lequel choisiriez-vous et pourquoi ?

(iv). Montrer que $I_n = [M_n, \alpha^{-1/n} \times M_n)$ est un intervalle de confiance de niveau $1-\alpha$ pour θ.

Exercice 2. On veut estimer la valeur de p à l'aide d'un échantillon de m observations $Y_1, \ldots, Y_m$ indépendantes et suivant une loi Bernoulli $\mathcal{B}(1,p)$ avec $0 < p < 1$ inconnu. On définit les statistiques suivantes :

$$\hat{p}_1 = \frac{\sum_{i=1}^m Y_i}{m} \quad \text{et} \quad \hat{p}_2 = \frac{\sum_{i=1}^m Y_i + 1/2}{m+1}.$$

Calculer l'erreur quadratique moyenne des deux estimateurs $\hat{p}_1$ et $\hat{p}_2$. Lequel des deux estimateurs est préférable ?

Exercice 3. Soit $X_1, \ldots, X_n$ un échantillon tiré de la distribution ayant comme fonction de densité $f(x;\theta) = \theta x^{\theta-1}$ pour $x \in (0,1)$ et $f(x;\theta) = 0$ autrement (série de la semaine passée). On a vu dans la série précédante que l'estimateur du maximum de vraisemblance était $\hat{\theta}_{ML} = -\frac{n}{\sum_{i=1}^n \log(X_i)}$.

(i). Trouver la distribution asymptotique de $\hat{\theta}_{ML}$ à l'aide de l'information de Fisher.

(ii). Maintenant, on va vérifier ce résultat dans cet exemple. Trouver la fonction de répartition de $Y_i = -\log(X_i)$, sa fonction de densité, son espérance et sa variance. Utiliser le théorème centrale limite pour déduire la distribution asymptotique de $\overline{Y}$ et la méthode delta pour trouver la distribution asymptotique de $\hat{\theta}_{ML}$.

(iii). On considère maintenant $\widehat{\theta}_{MOM} = \overline{X}_n/(1-\overline{X}_n)$. Montrer que $\mathbb{E}_\theta[X_1] = \theta/(\theta+1)$ et $\text{var}_\theta[X_i] = \theta/[(\theta+2)(\theta+1)^2]$, et que

$$\widehat{\theta}_{MOM} \overset{app}{\sim} \mathcal{N}\left(\theta, \frac{\theta(\theta+1)^2}{(\theta+2)n}\right)$$

Comparer avec la distribution asymptotique de $\widehat{\theta}_{ML}$. Lequel est préférable ?

Exercice 4. Trouver l'information de Fisher $I_n(\theta)$ et l'information observée $J_n(\widehat{\theta}_{ML})$ dans le cas $X_1, X_2, \ldots, X_n \overset{iid}{\sim} B(1,p)$ avec $p \in (0,1)$ et dans le cas $X_1, \ldots, X_n \sim$Poiss(λ).

Exercice 5. Une entreprise d'automobiles veut publier des informations sur la consommation d'un nouveau modèle de voiture. Pour 12 voitures de ce modèle on mesure la consommation d'essence, en litres, pour parcourir 100 km, et on obtient les résultats suivants :

$$14.60, 11.21, 11.56, 11.37, 13.68, 15.07, 11.06, 16.58, 13.37, 15.98, 12.07, 13.22.$$

(a) Supposez que la consommation d'essence pour une voiture de ce modèle est une variable aléatoire $X \sim \mathcal{N}(\mu, 3.5)$ avec μ inconnu, et que les données récoltées sont indépendantes. Donnez un intervalle de confiance à 95 % pour μ (la moyenne empirique de l'échantillon est $\bar{x} = 13.31$).

On va maintenant abandonner l'hypothèse de variance connue (ce qui est plus réaliste). On considère toujours un modèle normal, mais cette fois on suppose que $X \sim \mathcal{N}(\mu, \sigma^2)$ avec μ et σ^2 inconnus. Les données récoltées sont toujours considérées indépendantes.

(b) Donnez un intervalle de confiance à 95 % pour μ (la variance empirique de l'échantillon est $\frac{1}{11}\sum_{i=1}^{12}(x_i - \bar{x})^2 = 3.69$).

(c) Donnez un intervalle de confiance à 90 % pour μ. Commentez la différence avec la partie (b).
Le seuil à utiliser est : $s_{10} = 1.796$ pour $d = 11$.

(d) Que l'entreprise peut-elle faire pour obtenir un intervalle de confiance à 95 % plus étroit que celui obtenu dans la partie (b) ?