TD 2 : Estimation, intervalles de confiance et tests

Exercice 1 (Comparaison d'estimateurs) On considère un échantillon i.i.d. $(X_i)_{1\leq i\leq n}$ tel que $\mathbb{E}(X_1) = m$ et $\mathrm{Var}(X_1) = \sigma^2$. On suppose σ^2 connue, $m \in \mathbb{R}$ étant ici le paramètre inconnu. On propose deux estimateurs de m :

$$\overline{X}_n \quad \text{et} \quad Z_n = \frac{1}{2}(X_n + X_{n-1}).$$

1. Montrer que $\overline{X}_n$ et Z_n sont sans biais.
2. Qui de $\overline{X}_n$ ou Z_n choisiriez-vous pour approcher m ?
3. Que pensez-vous de l'estimateur $W_n = 0$?

Exercice 2 (Stabilisation de la variance versus "plug-in") Soient $X_1, \ldots, X_n$ des variables aléatoires indépendantes de même loi de Poisson de paramètre $\theta > 0$. On note $\widehat{\theta}_n = \overline{X}_n = \sum_{i=1}^n X_i/n$.

1. Donner le TCL vérifié par $\widehat{\theta}_n$.
2. Trouver une fonction $g : \mathbb{R}_+ \to \mathbb{R}$ telle que $\sqrt{n}(g(\widehat{\theta}_n) - g(\theta))$ tende en loi vers une gaussienne centrée réduite. En déduire un intervalle de confiance asymptotique à 95% pour θ.
3. Par la méthode de plug-in, obtenir d'une autre façon un intervalle de confiance asymptotique à 95% pour θ.
4. Vérifier que les deux résultats sont équivalents.

Exercice 3 (Loi normale) On observe $X_1, \ldots, X_n$ indépendantes et de même loi. On suppose qu'il existe $\theta > 0$ tel que cette loi admet la densité

$$f_\theta(x) = \frac{1}{\sqrt{2\pi\theta^2}} \exp\left(-\frac{x^2}{2\theta^2}\right).$$

Indication : $F^{-1}_{\chi^2_{10}}(0.025) \simeq 3.25$, $F^{-1}_{\chi^2_{10}}(0.975) \simeq 20.48$, $F^{-1}_{\chi^2_{10}}(0.95) \simeq 18.31$, $F_{\chi^2_{10}}(40/3) \simeq 0.79$ et $\Phi^{-1}(0.975) \simeq 2$.

1. On veut estimer $\tau = \theta^2$. Proposer un estimateur $\widehat{\tau}$ de τ et étudier sa loi.
2. Soit $\alpha \in]0,1[$. 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 $\widehat{\tau}$ et en déduire un intervalle de confiance asymptotique.
4. Lorsque $n = 10$, $\widehat{\tau} = 2$ et $\alpha = 0.05$, comparer l'intervalle de confiance obtenu à la question 2 (non asymptotique) avec celui obtenu à la question 3 (asymptotique).

5. Constuire un test de niveau α de
$$H_0 : \tau \leq 3 \text{ contre } H_1 : \tau > 3.$$
Pour un seuil de 5%, lorsque $n = 10$ et qu'on observe $\widehat{\tau}(\omega) = 4$, rejette-t-on l'hypothèse nulle ? Calculer la p-valeur du test.

Exercice 4 (Loi exponentielle) On considère un échantillon i.i.d. $(X_i)_{1\leq i\leq n}$ avec $n \geq 3$ et X_1 de densité
$$f_\theta(x) = \theta \exp(-\theta x)\mathbb{1}_{[0,+\infty[}(x),$$
où $\theta > 0$ est le paramètre inconnu.

1. On propose d'estimer θ par $Y_n = \frac{n}{\sum_{i=1}^n X_i}$.
   (a) Montrer que la v.a. Y_n est bien définie.
   (b) Expliquer pourquoi il est logique de choisir Y_n comme estimateur de θ.
   (c) Déterminer la loi limite de $\sqrt{n}(Y_n - \theta)$.
   (d) Donner la loi de $\sum_{i=1}^n X_i$. En déduire la valeur de $\mathbb{E}[(Y_n - \theta)^2]$.
2. Pour estimer θ, on propose d'utiliser $Z_n = \frac{n-1}{n} Y_n$.
   (a) Z_n vérifie-t-il des propriétés de convergence similaires à celles de Y_n ?
   (b) Qui de Y_n ou Z_n choisiriez-vous pour estimer θ ?
3. Soit $\alpha \in]0, 1[$. Donner un intervalle de confiance bilatère de niveau asymptotique $1 - \alpha$ pour θ.
4. Proposer un test de niveau asymptotique α pour tester $H_0 : \theta \geq 1$ contre $H_1 : \theta < 1$.
5. Proposer un test de niveau asymptotique α pour tester $H_0 : \theta = 1$ contre $H_1 : \theta \neq 1$.

Exercice 5 (Loi Bêta dilatée) On considère un échantillon i.i.d. $(X_i)_{1\leq i\leq n}$ avec X_1 de densité
$$f_\theta(x) = \frac{2}{\theta^2} x \mathbb{1}_{[0,\theta]}(x),$$
où $\theta > 0$ est un paramètre inconnu.

1. On pose $X_{(n)} = \max_{1\leq i\leq n} X_i$.
   (a) Donner la densité de $X_{(n)}$.
   (b) Donner les deux premiers moments de $X_{(n)}$. En déduire sa variance.
   (c) Montrer que $X_{(n)}$ converge en probabilité vers θ.
   (d) $X_{(n)}$ est-il un estimateur fortement consistant de θ ?
2. Étudier la convergence de $\overline{X}_n$. En déduire un estimateur consistant de θ.
3. Qui de $X_{(n)}$ ou $\frac{3}{2}\overline{X}_n$ choisiriez-vous pour estimer θ ?
4. Construire un intervalle de confiance pour θ de niveau $1 - \alpha$.
5. Construire un test de niveau α de l'hypothèse nulle $\theta = 1$ contre l'alternative $\theta \neq 1$. On observe $x_{(20)} = 0.85$, quelle est la p-valeur ? Rejette-t-on H_0 au niveau 5% ?

Exercice 6 ($\star$)(Exponentielle translatée) Soit X une v.a.r. admettant pour densité
$$f_\theta(x) = \frac{1}{\theta} \exp(-(x - \theta)/\theta)\mathbb{1}_{[\theta,+\infty[}(x),$$
où $\theta > 0$ est inconnu, et $(X_1, \ldots, X_n)$ un échantillon i.i.d. de même loi que X.

1. Montrer que $X/\theta - 1$ suit une loi exponentielle $\mathcal{E}(1)$.
2. Déterminer la fonction de répartition de $X_{(1)} = \min_{1 \le i \le n} X_i$.
3. Etudier la convergence de $X_{(1)}$ vers θ en moyenne quadratique.
4. Construire un intervalle de confiance pour θ de niveau $1 - \alpha$.

Exercice 7 ($\star$)(Défaillances et loi de Poisson) Un statisticien observe chaque jour, pendant n jours, le nombre d'ampoules défaillantes à la sortie d'une chaîne de fabrication. Il veut estimer la probabilité de n'avoir aucune ampoule défaillante. Il note X_i le nombre d'ampoules défaillantes à la sortie de la chaîne le i^{e} jour.

1. Dans un premier temps, il compte le nombre N_n de jours où aucune défaillance n'a été observée. Il propose d'estimer $\mathbb{P}(X_1 = 0)$ par $\widehat{p}_1 = \dfrac{N_n}{n}$.
   (a) Montrer, en supposant juste les X_i i.i.d., que N_n/n est un estimateur sans biais de $\mathbb{P}(X_1 = 0)$. Calculer son risque quadratique et donner sa loi limite.
   (b) Donner des intervalles de confiance pour $\mathbb{P}(X_1 = 0)$ de niveau $1 - \alpha$.
2. Le statisticien suppose de plus que $X_1 \sim \mathcal{P}(\lambda)$. Il propose d'estimer $\mathbb{P}(X_1 = 0) = \exp(-\lambda)$ par $\widehat{p}_2 = e^{-\overline{X}_n}$.
   (a) Préliminaire : soient $Y_1 \sim \mathcal{P}(\lambda_1)$ et $Y_2 \sim \mathcal{P}(\lambda_2)$, deux v.a. indépendantes. Etudier la loi de $Y_1 + Y_2$.
   (b) Expliquer sa démarche. Montrer que $\widehat{p}_2$ est biaisé. Calculer sa variance et son biais. Déterminer des équivalents asymptotiques des quantités précédentes.
   (c) Lequel de $\widehat{p}_1$ et $\widehat{p}_2$ choisiriez-vous pour estimer $e^{-\lambda}$?