PC 9 : Intervalles de confiance

Exercice 1. (Intervalle de confiance et Méthode δ) Soit la variable aléatoire

$$Y = \mathbb{1}_{\{\theta > \xi\}},$$

où $\theta \in \mathbb{R}$ et ξ est une variable aléatoire de loi $\mathcal{N}(0,1)$. On dispose d'un échantillon $Y_1, \ldots, Y_n$ des réalisations i.i.d. de Y.

1. Soit Φ la fonction de répartition de la loi $\mathcal{N}(0,1)$. Montrer que $\hat{\theta}_n = \Phi^{-1}(\frac{1}{n}\sum_{i=1}^n Y_i)$ est l'estimateur de maximum de vraisemblance de θ. Cet estimateur est-il convergeant ?
2. Déterminer la limite en loi de
$$\sqrt{n}\left(\hat{\theta}_n - \theta\right).$$
3. Soit $0 < \alpha < 1$. Proposer un intervalle de confiance asymptotique de niveau $1-\alpha$ sous la forme
$$\left[\hat{\theta}_n \pm f_n(Y_1, \ldots, Y_n)\right],$$
où $f_n : \mathbb{R}^n \to \mathbb{R}$ est une fonction ne dépendant pas de θ.

Exercice 2. (Normalité asymptotique $\Rightarrow$ Estimateur convergeant) Soit $\hat{\theta}_n$ un estimateur asymptotiquement normal de θ pour un certain modèle statistique : il existe une fonction $g(\theta)$ telle que

$$\sqrt{n}(\hat{\theta}_n - \theta) \overset{\text{(loi)}}{\to} \mathcal{N}(0, g(\theta)).$$

Démontrer que $\hat{\theta}_n \to \theta$ en probabilité.

Exercice 3. (Modèle de Poisson) Soit $(X_1, \ldots, X_n)$ un échantillon i.i.d. de la loi de Poisson $\mathcal{P}(\lambda)$ de paramètre $\lambda > 0$ inconnu. On rappelle que

$$\mathbb{E}[X_1] = \mathrm{Var}(X_1) = \lambda.$$

Notons $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$ et $\hat{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \bar{X}_n)^2$.

1. Justifier que $\bar{X}_n$ est un estimateur sans biais de λ, qu'il est consistant et asymptotiquement normal.
2. On souhaite comparer $\bar{X}_n$ avec l'autre estimateur $\hat{\sigma}_n^2$. Montrer que $\hat{\sigma}_n^2$ est asymptotiquement normal. (On utilisera, sans le démontrer, que $\mathbb{E}_\lambda[(X_1 - \lambda)^4] = \lambda + 3\lambda^2$).
3. Est-ce que $\hat{\sigma}_n^2$ converge en probabilité vers λ ?
4. Trouver une fonction g telle que
$$\sqrt{n}\left(g(\bar{X}_n) - g(\lambda)\right) \xrightarrow{\mathcal{L}} \mathcal{N}(0,1).$$
5. Quel estimateur de λ est à privilégier : $\bar{X}_n$, $\hat{\sigma}_n^2$, ou $g(\bar{X}_n)$?

Exercice 4. (Réduction de variance dans une méthode de Monte Carlo). Soit g une fonction mesurable et bornée ($0 \leq g \leq 1$). On souhaite calculer de façon approchée $m = \int_0^1 g(x)dx$.
Soient X et Y des variables i.i.d. de loi uniforme sur $[0,1]$. On pose

$$U = \mathbf{1}_{Y \leq g(X)}, \qquad V = g(X) \quad \text{et} \quad W = \frac{g(X) + g(1-X)}{2}.$$

1. Calculer l'espérance et la variance de U, V et W. Démontrer que pour toute fonction g à valeurs dans $[0,1]$ on a
$$\mathbb{E}[W^2] \leq \mathbb{E}[V^2] \leq \mathbb{E}[U^2].$$
2. Expliquer comment en déduire trois méthodes de type Monte-Carlo pour estimer m.
3. À partir de maintenant on suppose que g est monotone. Démontrer que $\mathbb{E}(g(X)g(1-X)) \leq m^2$.
(Indication : on pourra d'abord montrer que $(g(x) - g(y))(g(1-x) - g(1-y)) \leq 0$ pour tous x, y.)
4. On considère une suite $(X_i)_{i\geq 1}$ de variables i.i.d. de loi uniforme sur $[0,1]$. On considère les estimateurs
$$A_n = \frac{1}{2n}\sum_{i=1}^{2n} g(X_i), \qquad B_n = \frac{1}{2n}\sum_{i=1}^{n}(g(X_i) + g(1-X_i)),$$
lequel choisiriez-vous pour approcher m ?

Exercice 5. (Intervalles de confiance et Bienaymé-Tchebychev) 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. Utiliser l'inégalité de Bienaymé-Tchebychev pour construire un intervalle de confiance pour p au niveau 95%.
2. On désire que la valeur estimée $\widehat{p}$ diffère de la proportion inconnue exacte p de moins de 0.005 avec une probabilité d'au moins 95%. Quel sera l'effectif d'un tel échantillon ?
3. Selon cette approche, quel est le niveau de confiance du même intervalle qu'à la question précédente avec un effectif de $n = 400$? Quelle conclusion peut-on en tirer ?