TD n$^{\text{o}}$ 10 : Bases de l'estimation paramétrique

Exercice 1. On désire estimer la proportion p inconnue de brochets dans un grand lac. Pour ce faire, on pêche des poissons jusqu'à ce qu'on obtienne n brochets. Soit X la *var* égale au nombre de poissons qu'il a fallu pêcher. On admet que la proportion p ne varie pas au cours de la pêche. La loi de X est donnée par

$$\mathbb{P}(X = k) = \binom{k-1}{n-1} p^n (1-p)^{k-n}, \qquad k \in \mathbb{N} - \{0, \ldots, n-1\}.$$

1. Montrer que, pour tout $k \in \mathbb{N} - \{0, \ldots, n-1\}$, $\dfrac{n-1}{k-1} \times \dbinom{k-1}{n-1} = \dbinom{k-2}{n-2}$.

2. Montrer que $\widehat{p} = \dfrac{n-1}{X-1}$ est un estimateur sans biais de p *(on rappelle la formule du binôme négatif : on a* $\sum_{k=m}^{\infty} \binom{k}{m} x^{k-m} = \frac{1}{(1-x)^{m+1}}$, $x \in]-1, 1[$, $m \in \mathbb{N}$*).*

Exercice 2. Soit X la *var* égale au nombre de pannes que subit un certain type d'appareil électroménager. On suppose que X suit la loi de Poisson $\mathcal{P}(\lambda)$:

$$\mathbb{P}(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}, \qquad k \in \mathbb{N},$$

où $\lambda > 0$ est un paramètre inconnu. Soient $n \in \mathbb{N}^*$ et $(X_1, \ldots, X_n)$ un n-échantillon de X. On pose

$$S_n = \sum_{i=1}^{n} X_i, \qquad \overline{X}_n = \frac{S_n}{n} = \frac{1}{n} \sum_{i=1}^{n} X_i.$$

1. Déterminer la loi de S_n. Montrer que $\overline{X}_n$ est un estimateur sans biais de λ. Étudier sa convergence.

2. Dorénavant, on cherche à estimer la probabilité qu'il n'y ait aucune panne. Cette probabilité est notée θ.

   (a) Exprimer θ en fonction de λ.

   (b) On pose $T_n = e^{-\overline{X}_n}$. Calculer $\mathbb{E}(T_n)$ et montrer que T_n n'est pas un estimateur sans biais de θ. Est-il asymptotiquement sans biais ?

   (c) On pose $\widehat{\theta}_n = \left(1 - \frac{1}{n}\right)^{S_n}$. Montrer que $\widehat{\theta}_n$ est un estimateur sans biais de θ.

   (d) Calculer la variance de $\widehat{\theta}_n$. Est-ce que $\widehat{\theta}_n$ est un estimateur convergent ?

3. *Application.* Un expérimentateur a relevé le nombre de pannes que subissent 10000 appareils de ce type. Les mesures obtenues, notées $x_1, \ldots, x_{10000}$, donnent $\sum_{i=1}^{10000} x_i = 20000$. Donner une estimation ponctuelle de la probabilité qu'il n'y ait aucune panne.

Exercice 3. Soient $a > 0$, X une *var* suivant la loi uniforme $\mathcal{U}([0,a])$, *i.e.* de densité :

$$f(x) = \begin{cases} \dfrac{1}{a} & \text{si } x \in [0,a], \\ 0 & \text{sinon.} \end{cases}$$

$n \in \mathbb{N}^*$ et $(X_1, \ldots, X_n)$ un n-échantillon de X. Ici, a est un paramètre inconnu que l'on souhaite estimer à l'aide de $(X_1, \ldots, X_n)$. On pose

$$U_n = \frac{2}{n}\sum_{i=1}^{n} X_i, \qquad V_n = \sup(X_1, \ldots, X_n), \qquad W_n = \left(\frac{n+1}{n}\right)\sup(X_1, \ldots, X_n).$$

1. Montrer que U_n est un estimateur sans biais de a. Déterminer son risque quadratique.
2. Déterminer la fonction de répartition de X, puis celle de V_n. En déduire une densité de V_n. Calculer le biais de V_n, et son risque quadratique.
3. Déterminer le biais et le risque quadratique de W_n.
4. Entre U_n, V_n et W_n, quel est le meilleur estimateur de a lorsque n est supposé être grand ?

Exercice 4. Soient $a \in]0,1[$, X une *var* de densité :

$$f(x) = \begin{cases} \dfrac{1}{2a} & \text{si } x \in [0,a], \\ \dfrac{1}{2(1-a)} & \text{si } x \in]a,1], \\ 0 & \text{sinon.} \end{cases}$$

$n \in \mathbb{N}^*$ et $(X_1, \ldots, X_n)$ un n-échantillon de X. Ici, a est un paramètre inconnu que l'on souhaite estimer à l'aide de $(X_1, \ldots, X_n)$. On pose

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

1. Calculer $\mathbb{E}(X)$. Montrer que
$$\mathbb{V}(X) = \frac{4 + (2a-1)^2}{48}.$$
2. Calculer $\mathbb{E}\left(\overline{X}_n\right)$ et $\mathbb{V}\left(\overline{X}_n\right)$. En déduire un estimateur T_n sans biais de a et étudier sa convergence.
3. En considérant une majoration de $\mathbb{V}(T_n)$ indépendante de a, construire un intervalle de confiance bilatéral I_1 pour a au niveau 95% *(on pourra utiliser l'inégalité de Bienaymé-Tchebychev)*.
4. Étudier la convergence en loi de $\left(\dfrac{T_n - a}{\sigma(T_n)}\right)_{n \in \mathbb{N}^*}$. On suppose que n est suffisamment grand pour approcher la loi de $\dfrac{T_n - a}{\sigma(T_n)}$ à cette loi limite. En utilisant cette approximation et une majoration convenable de $\mathbb{V}(T_n)$, construire un intervalle de confiance bilatéral I_2 pour a au niveau 95%.
5. Comparer les intervalles I_1 et I_2.