# TD 4 : Estimations et intervalles de confiances

# I Estimateurs

**Exercice 1.** Nous considérons dans chacun des cas ci-dessous un $n$-échantillon $X_1, \ldots, X_n$ de loi $p_\theta$. Déterminer à chaque fois $\hat{\theta}_n$ l'estimateur du maximum de vraisemblance de $\theta$, puis étudier ses propriétés (biais, consistance).

**a)** $p_\theta$ est la loi géométrique de paramètre $\theta \in ]0,1[$.

**b)** $p_\theta$ est la loi Normale $\mathcal{N}(\theta)$ avec $\theta = (m, \sigma^2) \in \mathbb{R} \times ]0, +\infty[$.

**c)** $p_\theta$ est la loi $\mathcal{E}(\lambda)$ translaté de $\theta > 0$, (c'est-à-dire de densité $\lambda e^{-\lambda(x-\theta)} \mathbb{1}_{[\theta,+\infty[}(x)$) où $\lambda > 0$ est un paramètre supposé connu. On montrera que $\hat{\theta}_n \sim \theta + \mathcal{E}(n\lambda)$. Quel autre estimateur « naturel » fournit la méthode des moments ?

**d)** $p_\theta$ est la loi de densité $f_\theta$ avec $\theta \in ]1/2, 1[$, où

$$\forall x \in \mathbb{R} \qquad f_\theta(x) = \frac{\theta}{1-\theta}\, x^{(2\theta-1)/(1-\theta)} \mathbb{1}_{]0,1[}(x)$$

**Exercice 2.** Soit $T$ une variable aléatoire réelle dont la densité est donnée par :

$$f(x) = \begin{cases} \frac{2x}{R^2} & \text{si } 0 \leq x \leq R, \\ 0 & \text{sinon} \end{cases}$$

Soient aussi $T_1, \ldots, T_n$ $n$ variables aléatoires indépendantes de même loi que $T$. On pose

$$X_n = \frac{1}{n}\sum_{i=1}^{n} T_i \qquad \text{et} \qquad Y_n = \sup_{1 \leq i \leq n} T_i$$

**1)** Calculer $\mathbb{E}[T]$ et $\mathbb{V}\mathrm{ar}(T)$.

**2)** Calculer $\mathbb{E}[X_n]$ et $\mathbb{V}\mathrm{ar}(X_n)$.

**3)** Calculer la densité $f_n$ de $Y_n$, puis son espérance et sa variance.

**4)** Construire à partir de $X_n$ et de $Y_n$ respectivement deux estimateurs $\hat{X}_n$ et $\hat{Y}_n$ sans biais de $R$. Ces deux estimateurs sont-ils consistants ?

# II Intervalles de confiance

**Exercice 3** (**Elimination de la dépendance en** $\theta$)**.** Nous avons vu en cours que, dans le modèle de Bernoulli (où $\theta \in ]0,1[$ est le paramètre recherché), si $\overline{X}_n$ désigne la moyenne empirique de l'échantillon alors

$$\sqrt{n}\frac{\overline{X}_n - \theta}{\sqrt{\overline{X}_n(1-\overline{X}_n)}} \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}(0,1)$$

(Nous avons utilisé le TCL, la convergence en probabilité de $\sqrt{\overline{X}_n(1-\overline{X}_n)}$ vers $\sqrt{\theta(1-\theta)}$ puis le lemme de Slutsky). Ceci nous a permis d'obtenir un intervalle de confiance de $\theta$. Nous allons voir maintenant une autre méthode :

**1)** En utilisant le TCL et la méthode Delta, trouver une fonction $g$ telle que

$$\sqrt{n}\left(g(\overline{X}_n) - g(\theta)\right) \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}(0,1).$$

*Indice : La fonction* Arcsin *est dérivable sur* $]-1,1[$ *et* $\mathrm{Arcsin}'(x) = (1-x^2)^{-1/2}$*, pour tout* $x \in ]-1,1[$.

**2)** En déduire un intervalle de confiance de niveau $1-\alpha$ pour $\theta$, que l'on exprimera en fonction des quantiles $q_r$ de $\mathcal{N}(0,1)$.

**Exercice 4.** Un standard téléphonique reçoit $N$ appels par heures où $N$ suit une loi de Poisson de paramètre $\lambda$. Une étude portant sur 80 heures a donné un nombre moyen de 21 appels par heures. Déterminer l'intervalle de confiance asymptotique à 95% du nombre moyen d'appels par heure $\lambda$. *Indication : si* $Y \sim \mathcal{N}(0,1)$ *et* $t = 1,960$*, alors* $\mathbb{P}(Y \leq t) \approx 0,975$.

**Exercice 5.** Une usine fabrique des câbles. On suppose que la charge maximale supportée par un câble, exprimée en tonnes, est une variable aléatoire réelle qui suit une loi normale $\mathcal{N}(\mu, 1/4)$. Une étude portant sur 50 câbles a donné une moyenne des charges maximales supportées égale à $12,2$ tonnes.

**1)** Déterminer l'intervalle de confiance à 99% de la charge maximale moyenne $\mu$ de tous les câbles fabriqués par l'usine. *Indication : si* $Y \sim \mathcal{N}(0,1)$ *et* $t = 2,575$*, alors* $\mathbb{P}(Y \leq t) \approx 0,995$.

**2)** Déterminer la taille minimale $n$ de l'échantillon étudié pour que la longueur de l'intervalle de confiance à 99% soit inférieure ou égale à $0,2$.

**Exercice 6. (Lien entre intervalles de confiance asymptotiques et non-asymptotiques)**
On admettra le théorème suivant :

***Borne de Berry-Essen :*** *Soient* $X_1, \ldots, X_n$ *des v.a. i.i.d. telles que* $\mathbb{E}[X_1] = 0$ *et* $\mathbb{E}[X_1^2] = \sigma^2$. *On suppose que* $\rho = \mathbb{E}[|X_1|^3] < \infty$. *Si on définit*

$$F_n(x) = \mathbb{P}\left(\frac{X_1 + \ldots + X_n}{\sigma\sqrt{n}} \leq x\right) \qquad \text{et} \qquad \Phi(x) = \mathbb{P}\big(\mathcal{N}(0,1) \leq x\big),$$

*alors pour tout* $x \in \mathbb{R}$

$$|F_n(x) - \Phi(x)| \leq \frac{3\rho}{\sigma^3\sqrt{n}}.$$

En déduire un intervalle de confiance non-asymptotique de niveau $1-\alpha$ dans le modèle de Bernoulli, à partir des intervalles obtenus grâce au théorème central limite et en supposant que $\theta_0 \in [\delta, 1-\delta]$ pour un $\delta \in ]0, 1/2[$ (de sorte que $\sigma^2 \geq \delta(1-\delta) > 0$).