TD 3 : Quantités empiriques et méthode des moments

Exercice 1 (Propriétés des moyenne et médiane empiriques) Soit $X_1, \dots, X_n$ un échantillon de fonction de répartition F avec $\mathbb{E}(X_1^2) < \infty$.

1. (a) Montrer que $\mathbb{E}(X_1) = \underset{t\in\mathbb{R}}{\operatorname{argmin}}\ \mathbb{E}[(X_1 - t)^2]$.
   (b) Déterminer $\underset{t\in\mathbb{R}}{\operatorname{argmin}} \sum_{i=1}^n (X_i - t)^2/n$.
2. On note $x_{1/2} = F^{-1}(1/2)$ la médiane de la loi de X_1.
   (a) Montrer que si F est continue sur $\mathbb{R}$ et strictement croissante sur un voisinage de $x_{1/2}$, alors $x_{1/2} = \underset{t\in\mathbb{R}}{\operatorname{argmin}}\ \mathbb{E}[|X_1 - t|]$.
   Indication : pour toute v.a. positive Y, $\mathbb{E}(Y) = \int_0^{+\infty} \mathbb{P}(Y > s)\, \mathrm{d}s$.
   (b) Déterminer, selon la parité de n, $\underset{t\in\mathbb{R}}{\operatorname{argmin}} \sum_{i=1}^n |X_i - t|/n$.
   Indication : on pourra étudier la fonction $t \in \mathbb{R} \mapsto \sum_{i=1}^n |X_i - t|/n$ sur les intervalles définis par les statistiques d'ordre $X_{(1)} \le \dots \le X_{(n)}$.

Exercice 2 (Estimation empirique dans un modèle de translation) Soient X une variable aléatoire symétrique (i.e. X et $-X$ ont même loi), de variance σ^2, de fonction de répartition F, et $X_1, \dots, X_n$ un échantillon de loi $F_\theta(x) = F(x - \theta)$, c'est-à-dire que X_1 a même loi que $X + \theta$. On s'intéresse à l'estimation de $\theta \in \mathbb{R}$.

1. Calculer $\mathbb{E}(X)$ et en déduire $\mathbb{E}_\theta(X_1)$.
2. Montrer que si F est continue et strictement croissante sur un voisinage de 0, alors sa médiane vaut 0. En déduire que $\theta = x_{1/2} = F_\theta^{-1}(1/2)$.
3. Dans chacun des cas suivants, comparer les variances des lois limites des estimateurs de θ construits à partir de la moyenne et de la médiane empirique :
   (a) F est la fonction de répartition de la loi normale $\mathcal{N}(0, \sigma^2)$;
   (b) F admet pour densité $f(x) = \frac{3}{4}(1 - x^2)$ sur $[-1, 1]$;
   (c) F admet pour densité $\frac{1}{2}e^{-|x|}$ (loi de Laplace).

Exercice 3 (Loi uniforme) Soit $X_1, \dots, X_n$ un échantillon de loi $\mathcal{U}([0, \theta])$, on veut estimer $\theta > 0$.

1. Déterminer un estimateur $\widehat{\theta}_1$ de θ par la méthode des moments.

On considère les estimateurs suivants :

$$\widehat{\theta}_2 = 2F_n^{-1}(1/2) \quad \text{et} \quad \widehat{\theta}_3 = \max_{1\le i\le n} X_i,$$

où F_n^{-1} est l'inverse généralisée de la fonction de répartition empirique.

2. Expliquer les idées conduisant à proposer les estimateurs $\widehat{\theta}_2$ et $\widehat{\theta}_3$.
3. Déterminer les lois limites de $\widehat{\theta}_1, \widehat{\theta}_2, \widehat{\theta}_3$. Que pouvez-vous dire des propriétés de ces estimateurs ?
4. Comparer les performances des trois estimateurs.
5. Donner un intervalle de confiance non asymptotique de niveau $1-\alpha$ pour θ.

Exercice 4 (Intervalles de confiance pour la médiane) Soient $X_1, \ldots, X_n$ des variables aléatoires i.i.d. de même fonction de répartition F inconnue, continue sur $\mathbb{R}$. Soit $x_{1/2}$ la médiane de F et $x_{1/2}(n)$ la médiane empirique. Le but est d'estimer $x_{1/2}$.

1. On suppose ici que F est dérivable en $x_{1/2}$, de dérivée strictement positive. Quelle est la loi limite de $\sqrt{n}(x_{1/2}(n) - x_{1/2})$? Ce résultat permet-il d'obtenir un intervalle de confiance asymptotique pour la médiane $x_{1/2}$?
2. Appliquer le théorème central limite à la variable $F_n(x_{1/2})$.
3. En déduire un intervalle de confiance de niveau asymptotique $1-\alpha$ pour $x_{1/2}$.
4. Pour $n = 10^4$, donner un intervalle de confiance asymptotique à 95% pour la médiane, en fonction des statistiques d'ordre de l'échantillon.

Exercice 5 (Décembre 2016)
Soit $\theta > 0$ un paramètre inconnu. On considère la densité f_θ par rapport à la mesure de Lebesgue sur $\mathbb{R}$ définie par

$$f_\theta(x) = \frac{1}{2\theta}(\mathbf{1}_{[0,\theta]}(x) + \mathbf{1}_{[2\theta,3\theta]}(x)).$$

Dans toute la suite, on suppose disposer de n observations i.i.d. $X_1, \ldots, X_n$ de densité f_θ.

1. Calculer $\mathbb{E}[X_1]$ et $\mathrm{Var}(X_1)$. Déduire de la méthode des moments un estimateur $\widehat{\theta}_{M,n}$ de θ. Préciser la loi limite (après une renormalisation convenable) de $\widehat{\theta}_{M,n} - \theta$.
2. Représenter f_θ. Calculer la fonction de répartition F_θ de X_1 et la représenter.
3. Que vaut le premier quartile $q = x_{1/4} = F_\theta^{-1}(1/4)$? En déduire un estimateur $\widehat{\theta}_{q,n}$ consistant pour θ et préciser la loi limite (après une renormalisation convenable) de $\widehat{\theta}_{q,n} - \theta$.
4. Que vaut le troisième quartile $Q = x_{3/4} = F_\theta^{-1}(3/4)$? En déduire un estimateur $\widehat{\theta}_{Q,n}$ consistant pour θ et préciser la loi limite (après une renormalisation convenable) de $\widehat{\theta}_{Q,n} - \theta$.
5. Parmi $\widehat{\theta}_{M,n}$, $\widehat{\theta}_{q,n}$, $\widehat{\theta}_{Q,n}$, quel estimateur choisiriez-vous asymptotiquement ?
6. Soit $X_{(n)} = \max_{1 \leq i \leq n} X_i$. Pour $0 < x < \theta$, que vaut la probabilité $\mathbb{P}(X_{(n)} \leq 3\theta - x)$? En déduire un estimateur $\widehat{\theta}_{S,n}$ consistant pour θ.
7. Montrer que $\mathbb{P}(\widehat{\theta}_{S,n} \leq \theta - t/n)$ converge, pour tout $t \geq 0$, vers une limite non nulle que l'on calculera. Préciser la loi limite de $n(\theta - \widehat{\theta}_{S,n})$.
8. Si vous avez un grand nombre d'observations, lequel des estimateurs précédents choisir ?
9. Que vaut la médiane $m = x_{1/2} = F_\theta^{-1}(1/2)$ de la loi de X_1 ? Que dire de la médiane empirique comme estimateur de θ ?
10. ($\star$) On veut trouver un intervalle de confiance bilatère non asymptotique pour θ au niveau $(1-\alpha)$. Proposer une méthode.

Exercice 6 ($\star$) (Loi Gamma) Soit X de loi $\gamma(q, \theta)$, où $q > 0$ et $\theta > 0$, et $X_1, \ldots, X_n$ un échantillon de même loi que X.

1. On suppose $q > 0$ connu et $\theta > 0$ inconnu.
   (a) Proposer un estimateur de θ.
   (b) Construire un intervalle de confiance de niveau asymptotique $1-\alpha$ pour θ. Expliciter l'intervalle de niveau 95% lorsque $q = 2$ et $n = 10$.
2. On suppose $\theta > 0$ connu et $q > 0$ inconnu.
   (a) Proposer un estimateur de q.
   (b) Construire un intervalle de confiance de niveau asymptotique $1 - \alpha$ pour q.
3. On suppose $\theta > 0$ et $q > 0$ inconnus. Proposer des estimateurs empiriques de $\theta > 0$ et $q > 0$. Sont-ils consistants ?

Exercice 7 ($\star$) **(Exponentielle translatée)** Soient X une v.a. 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 de même loi que X.

1. Estimer θ par la méthode des moments.
2. Quelle est la loi limite de l'estimateur $\widehat{\theta}$ ainsi obtenu ? Etudier également sa convergence en moyenne quadratique.
3. Cet estimateur est-il meilleur que $X_{(1)} = \min_{1 \le i \le n} X_i$? Rappel : cet estimateur a été étudié au TD 2.