T. D. n° 5
Estimation ponctuelle

Exercice 1. Une inégalité

Soit X une variable aléatoire admettant un moment d'ordre deux et suivant une loi d'espérance m et d'écart-type σ.

1. Montrer que pour toute statistique T nous avons :

$$(\mathbb{E}[T])^2 \leqslant \mathbb{E}\left[T^2\right].$$

Dans quel cas avons-nous l'égalité ?

2. Déterminer un estimateur T^2 sans biais de σ^2 ; dans quel cas T est-il sans biais pour σ ?

3. En déduire que généralement, la statistique :

$$\sqrt{\frac{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2}{n-1}}$$

n'est pas un estimateur sans biais de l'écart-type σ.

Exercice 2. Modèle de Poisson

Soit X une variable aléatoire suivant une loi de Poisson $\mathcal{P}(\lambda)$ de paramètre λ. Nous considérons un échantillon aléatoire de taille n de loi parente X.

1. Déterminer deux estimateurs $\widehat{\lambda}_1$ et $\widehat{\lambda}_2$ sans biais du paramètre λ fondés sue les caractéristiques empiriques d'un échantillon aléatoire de taille n de loi parente X.

2. Comparer ces deux estimateurs de λ.

Exercice 3. Loi uniforme

Soit X une variable aléatoire uniforme sur l'intervalle $[0, 2a]$. Nous considérons une suite de n variables aléatoires indépendantes et de même loi que X. Nous posons $\overline{X} = \frac{1}{n}(X_1 + \cdots + X_n)$ et $T = \operatorname{Max}(X_1, \ldots, X_n)$.

1. Déterminer l'espérance et la variance de la variable aléatoire X.

2. Montrer que $\overline{X}$ est un estimateur convergent de a.

3. Déterminer la densité de la variable T. Calculer l'espérance et la variance de T. En déduire un autre estimateur $T^\star$ sans biais de a.

4. Comparer les deux estimateurs de a.

Exercice 4. Quantiles et comparaison d'estimateurs

Définition. *Soit X une variable aléatoire réelle dont nous supposons la fonction de répartition F continue et strictement croissante. Pour tout $p \in]0,1[$ nous appelons quantile d'ordre p et nous notons q_p la racine (unique d'après les hypothèses) de l'équation en x : $F(x) = p$, soit $q_p = F^{-1}(p)$. En particulier si $p = 1/2$, $q_{1/2}$ est appelée médiane et sera notée $me(X)$.*

Définition. *Soit $(X_1, \ldots, X_n)$ un échantillon aléatoire de loi parente X et $(Y_1, \ldots, Y_n)$ l'échantillon ordonné associé. Si $p \in]0,1[$ est tel que $n \times p$ n'est pas entier nous appelons quantile empirique d'ordre p et nous notons $Q_p(n)$ la variable aléatoire $Y_{[n\times p]+1}$, où $[n \times p]$ désigne la partie entière de $[n \times p]$.*

Remarque. Si $n \times p$ est entier nous appellerons quantile d'ordre p toute variable aléatoire comprise entre $Y_{n\times p}$ et $Y_{n\times p+1}$.

1. Soit X une variable aléatoire réelle de fonction de répartition F et $(X_1, \ldots, X_n)$ un échantillon aléatoire de loi parente X. Soit E_x l'événement $\{X < x\}$, $x \in \mathbb{R}$. Que vaut $\mathbb{P}[E_x]$? Nous posons $R_n(x)$ le nombre de répétitions de E_x en n expériences indépendantes. Pour $x \in \mathbb{R}$ fixé, quelle est la loi de $R_n(x)$?

Définition. *En fonction de $x \in \mathbb{R}$, $F_n(x) = R_n(x)/n$, où $R_n(x)$ est défini à la question **1.**, définit une fonction aléatoire appelée fonction de répartition empirique associée à X basée sur l'échantillon $(X_1, \ldots, X_n)$.*

2. En utilisant la loi forte des grands nombres, montrez que, pour $x \in \mathbb{R}$ fixé, $F_n(x) \overset{p.s.}{\to} F(x)$.
3. Montrer que si X est une variable aléatoire réelle dont la fonction de répartition F est continue strictement croissante alors :

$$Q_p(n) \overset{p.s.}{\to} q_p, \quad n \to +\infty.$$

   Vous pourrez utiliser le théorème de Glivenko-Cantelli dont l'énoncé est le suivant.

Théorème. *Pour $\mathbb{P}-$presque tout ω, la suite des fonctions de répartition $F_n(\cdot, \omega)$ converge uniformément vers F, autrement dit, nous avons*

$$\mathbb{P} - p.s. \quad \lim_n \sup_{x\in\mathbb{R}} |F_n(x, \cdot) - F(x)| = 0.$$

4. Soit X une variable aléatoire admettant une densité continue $f > 0$, alors la variable aléatoire $T_p = \sqrt{n}\,(Q_p - q_p)$ possède la propriété suivante :

$$T_p \overset{\mathcal{L}}{\to} \mathcal{N}\left(0, \frac{\sqrt{p(1-p)}}{f(q_p)}\right).$$

5. Déduire de la question **4.** que si l'échantillon aléatoire $(X_1, \ldots, X_n)$ a pour loi parente X une loi normale centrée réduite alors $\sqrt{n}Q_{1/2}$ suit asymptotiquement une normale $\mathcal{N}(0, \sqrt{\pi/2})$. Montrez que $\overline{X}_n = \widehat{\mu}_n$ et $Q_{1/2}$ sont deux estimateurs convergents de la moyenne d'une loi normale $\mathcal{N}(\mu, \sigma)$. Lequel est-il le plus efficace ?
6. Qu'en est-il pour la loi logistique, dont la fonction de densité est définie pour $x \in \mathbb{R}$ par $f(x) = \exp(-x)/(1 + \exp(-x))^2$, et la loi de Laplace , dont la fonction de densité est définie pour $x \in \mathbb{R}$ par $f(x) = \exp(-|x|)/2$.

Exercice 5. Famille exponentielle 1

Soit X une variable aléatoire suivant une loi de Poisson $\mathcal{P}(\lambda)$ de paramètre λ.
Montrer que la loi de Poisson $\mathcal{P}(\lambda)$ de paramètre λ appartient à la famille exponentielle.

Exercice 6. Famille exponentielle 2

Soit X une variable aléatoire suivant une loi binomiale $\mathcal{B}(n, p)$ de paramètres n et p.
Montrer que la loi binomiale $\mathcal{B}(n, p)$ de paramètres n et p appartient à la famille exponentielle.

Exercice 7. Famille exponentielle 3

Soit X une variable aléatoire suivant une loi normale $\mathcal{N}(\mu, \sigma)$.

1. Nous supposons que μ est inconnu et que σ est connu. Montrer que la loi normale $\mathcal{N}(\mu, \sigma)$ appartient à la famille exponentielle.
2. Nous supposons que μ et σ sont inconnus. Montrer que la loi normale $\mathcal{N}(\mu, \sigma)$ appartient à la famille exponentielle.

Exercice 8. Famille exponentielle 4

Soit X une variable aléatoire suivant une loi $\Gamma(r, \lambda)$ de paramètres r et λ. Montrer que la loi $\Gamma(r, \lambda)$ de paramètres r et λ appartient à la famille exponentielle.

Exercice 9. Famille exponentielle et exhaustivité

Soit X une variable aléatoire suivant une loi uniforme $U_{[0,\theta]}$ sur $[0, \theta]$ avec $\theta > 0$.

1. La famille de cette loi est-elle exponentielle ?
2. Soit $(X_1, \ldots, X_n)$ un échantillon aléatoire de taille n de loi parente X. Déterminer une statistique exhaustive pour θ.

Exercice 10. Exhaustivité et loi exponentielle

Soient X une variable aléatoire suivant une loi exponentielle de paramètre λ et $(X_1, \ldots, X_n)$ un échantillon aléatoire de taille n de loi parente X.
Montrer que la statistique

$$S = \sum_{i=1}^{n} X_i$$

est une statistique exhaustive.

Exercice 11. Exhaustivité et loi normale

Soient X une variable aléatoire suivant une loi normale $\mathcal{N}(\mu, \sigma)$ de paramètres μ et σ et $(X_1, \ldots, X_n)$ un échantillon aléatoire de taille n de loi parente X.
Considérons les statistiques

$$U = \sum_{i=1}^{n} X_i^2 \quad \text{et} \quad \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

Montrer que le couple $T = (\overline{X}, U)$ est une statistique exhaustive pour le couple (μ, σ^2).

Exercice 12. Exhaustivité et loi de Poisson

Soit X une variable aléatoire suivant une loi de Poisson $\mathcal{P}(\lambda)$ de paramètre λ et $(X_1, \ldots, X_n)$ un échantillon aléatoire de taille n de loi parente X.
Montrer que la statistique

$$S = \sum_{i=1}^{n} X_i$$

est une statistique exhaustive :

1. En montrant que $\mathbb{P}(X_1, \ldots, X_n | S = s)$ est indépendante de λ.
2. En utilisant le critère de factorisation.
3. En utilisant l'exercice 5.

Exercice 13. Exhaustivité et conditionnement

Nous considérons une variable aléatoire X continue de densité de probabilité $f(x, \theta)$ et un échantillon aléatoire de taille n de loi parente X.
Soit $T_n(X_1, \ldots, X_n)$ une statistique exhaustive. Nous tirons un échantillon aléatoire de taille n de loi parente la loi de X conditionnée par l'événement $x \in I$, I partie mesurable de $\mathbb{R}$.
La statistique $T_n(X_1, \ldots, X_n)$ est-elle toujours une statistique exhaustive ?

Exercice 14. Minimisation du risque pour l'estimation de la variance d'une loi normale

Soient X une variable aléatoire suivant une loi normale $\mathcal{N}(\mu, \sigma)$ de paramètres μ et σ et $(X_1, \ldots, X_n)$ un échantillon aléatoire de taille n de loi parente X.
Considérons la statistique suivante :

$$S_{n,c}^2 = \frac{1}{n-1} \sum_{i=1}^{n} \left(X_i - \overline{X}\right)^2.$$

1. Calculer $\mathbb{E}[S_{n,c}^2]$ et $\mathrm{Var}[S_{n,c}^2]$.

2. Soit k un nombre réel positif. Nous introduisons les estimateurs $T(k)$ de σ^2 de la forme $T(k) = kS_{n,c}^2$.
Calculer $EQM(T(k)) = \mathbb{E}\left[\left(T(k) - \sigma^2\right)^2\right]$.

3. Déterminer la valeur $k^\star$ de k qui minimise $R(k)$. Calculer alors $R(k^\star)$. Que dire de l'estimateur $t(k^\star)$?