Probabilité 7 – Estimation

Exercice 1 –

1. Soit T une variable aléatoire qui suit la loi de Poisson de paramètre λ. Pour $i \in \mathbb{N}^*$, on note m_i le moment d'ordre i de cette variable. Déterminer m_i pour $i = 1, 2, 3$.
2. Soit $T_1, T_2, \dots, T_n$ des variables aléatoires indépendantes de même loi que T ; on cherche à estimer le paramètre inconnu λ.
   (a) Montrer que $M_n = \frac{1}{n} \sum\limits_{k=1}^{n} T_k$ est un estimateur sans biais de λ.
   (b) Préciser $E(T_k^2)$, $E(M_n^2)$ et $E(T_k M_n)$, pour $k \in [\![1, n]\!]$.
   (c) On pose $V_n = \frac{1}{n} \sum\limits_{k=1}^{n} (T_k - M_n)^2$. Est-ce que V_n est un estimteur sans biais de λ ? Proposer un estimateur W_n sans biais de λ obtenu à l'aide de V_n.
3. On admet que la variance de W_n est égale à $\frac{n}{(n-1)^2}\lambda(1+2\lambda)$. Quel est, entre M_n et W_n, le meilleur estimateur ?

Exercice 2 – On admet que la mesure d'une grandeur physique, dont la valeur exacte est m, suit une loi normale $\mathcal{N}\left(m, \frac{m^2}{100}\right)$, d'espérance m et d'écart-type $\frac{m}{10}$.
On effectue une série de n mesures indépendantes et l'on note Y_n la moyenne des résultats obtenus.

1. Montrer que Y_n est une estimateur sans biais et convergent de m.
2. Combien faut-il effectuer de mesures pour que l'erreur relative commise sur m soit inférieure à 1%, avec une probabilité supérieure à 0.9 ?

Exercice 3 – Soit α un réel strictement positif. On note f la fonction définie sur $\mathbb{R}$ par $f(x) = \frac{\alpha}{x^{\alpha+1}}$ si $x \geqslant 1$, et 0 sinon.

1. Soit X la variable aléatoire qui admet f pour densité. Quelle est la loi de la variable $\ln(X)$?
2. On considère un échantillon $(X_1, \dots, X_n)$ de variables aléatoires indépendantes suivant la même loi que X, et l'on pose $Y_n = \ln(X_1) + \dots + \ln(X_n)$.
   (a) Quelle est la loi de Y_n ?
   (b) Calculer l'espérance de Y_n.
3. (a) Déduire de ce qui précède un estimateur sans biais de $\frac{1}{\alpha}$.
   (b) Quelle est la variance de cet estimteur ? Conclure.

Exercice 4 – (Estimation par capture-recapture, oral ESCP)
On cherche à évaluer le nombre N de poissons dans un étang. On prélève dans l'étang un échantillon de m poissons, que l'on marque et que l'on remet dans l'étang.
On propose deux méthodes différentes pour estimer N.

Méthode 1

Soit n un entier non nul inférieur ou égal à m. On prélève des poissons dans l'étang au hasard et avec remise, et l'on note X_n la variable aléatoire égale au nombre de poissons qu'il a été nécessaire (et suffisant) de pêcher pour obtenir n poissons marqués.
Pour tout entier i de $[\![2, n]\!]$, on pose $D_i = X_i - X_{i-1}$. On pose de plus $D_1 = X_1$, et on suppose que les D_i sont des variables aléatoires indépendantes.

1. (a) Pour tout $i \in [\![2, n]\!]$, déterminer la loi de D_i, son espérance et sa variance.

(b) En déduire l'espérance et la variance de X_n.

(c) On pose $A_n = \frac{m}{n} X_n$. Montrer que A_n est un estimateur sans biais de N.

2. (a) Pour n assez grand, par quelle loi peut-on approcher la loi de la variable $\overline{X_n} = \frac{X_n}{n}$?

(b) On a marqué 200 poissons, puis effectué 450 prélèvements pour obtenir 50 poissons marqués.

On note σ l'écart-type de A_n. On a pu prouver par ailleurs que $\sigma \leqslant 100$. Déterminer un intervalle de confiance de niveau 0.9 pour N.

Méthode 2

On prélève successivement et avec remise n poissons. Soit Y_n le nombre de poissons marqués ainsi recueillis.

3. (a) Montrer que $\frac{1}{nm} Y_n$ est un estimateur sans biais de $\frac{1}{N}$.

(b) Pour quelle raison ne peut-on pas prendre $\frac{nm}{Y_n}$ comme estimateur de N ?

4. (a) On pose $B_n = \frac{m(n+1)}{Y_n+1}$. Calculer l'espérance de B_n.

(b) B_n est-il un estimateur sans biais de N ? Est-il asymptotiquement sans biais ? Est-il convergent ?

Exercice 5 – Soit θ un réel positif non nul. On considère un échantillon iid $(X_1, \ldots, X_n)$ de loi parente la loi uniforme sur $[0, \theta]$. Soit X une variable aléatoire uniforme sur $[0, \theta]$.

1. (a) Calculer $E(X)$. En déduire un estimateur sans biais S_n de θ.

(b) Cet estimateur est-il convergent ?

2. (a) Calculer $E(\ln X)$. En déduire un estimateur sans biais T_n de $\ln \theta - 1$, puis un estimateur S'_n de θ.

(b) Prouver que S'_n est un estimateur convergent de θ.

(c) Peut-on conclure que S'_n est un estimateur sans biais de θ ?

3. (a) Soit $S''_n = \max(X_1, \ldots X_n)$ un autre estimateur de θ. Montrer que cet estimateur est biaisé, calculer son biais et son risque quadratique.

(b) L'estimateur S''_n est-il convergent ?

(c) Déduire de S''_n un estimateur S'''_n sans biais de θ. Dire pourquoi il est convergent.

Exercice 6 –

Soit $(X_1, \ldots, X_n)$ un échantillon indépendant identiquement distribué de loi parente $\mathcal{E}(c)$, le réel c étant un paramètre strictement positif inconnu.

1. Soit $S_n = X_1 + \cdots + X_n$. Déterminer la loi de S_n, et donner son espérance et sa variance.
2. Soit $n \geqslant 2$, et $T_n = \dfrac{n}{S_n}$.

(a) Déterminer, à l'aide du théorème de transfert, l'existence et la valeur de $E(T_n)$.

(b) Montrer que T_n est un estimateur asymptotiquement sans biais de c.

(c) Justifier que $U_n = \dfrac{n-1}{n} T_n$ est un estimateur sans biais de c.

3. On suppose $n \geqslant 3$.

(a) Déterminer $E(T_n^2)$. En déduire que $V(T_n) = c^2 \cdot \dfrac{n^2}{(n-1)^2(n-2)}$.

(b) Déterminer $V(U_n)$.

4. (a) Montrer que U_n est un estimateur convergent de c.

(b) Montrer que T_n est un estimateur convergent de c.

5. Déterminer les risques quadratiques de T_n et U_n. Lequel de ces deux estimateurs de c est le meilleur ?
6. (a) Soit S_n^* la variable aléatoire centrée réduite associée à S_n. Exprimer S_n^* en fonction de S_n, n et c. Que peut-on dire de la loi de S_n^* lorsque n est grand ?

   (b) Déterminer un intervalle de confiance de c au taux de confiance 0.97, dont les bornes seront exprimées à l'aide de l'estimateur T_n et de l'entier n. (On donne $\Phi(2.17) = 0.985$)

   Application numérique : pour $n = 10000$, on a trouvé $T_n(\omega) = 2$. Donner l'estimation d'un intervalle de confiance de c au taux de confiance 0.97.

7. On suppose maintenant que c est toujours inconnu, mais vérifie $c > 1$. On suppose que pour tout $i \in [\![1, n]\!]$, X_i suit une loi exponentielle de paramètre c^i.
   (a) Montrer que $\displaystyle\lim_{n\to+\infty} E(X_1 + \cdots + X_n) = \frac{1}{c-1}$.
   (b) Peut-on dire que $X_1 + \cdots + X_n$ est un estimateur asymptotiquement sans biais de $\dfrac{1}{c-1}$?
   (c) Étudier la convergence en loi de la suite $(X_n)_{n\in\mathbb{N}^*}$.

Exercice 7 – (ESCP 2009) Soit θ un réel strictement positif. Soit X une variable aléatoire qui suit une loi uniforme sur $]0, \theta[$. Soit $(X_n)_{n\in\mathbb{N}^*}$ une suite de variables aléatoires indépendantes de même loi que X. Pour tout entier $n \geqslant 1$, on pose :

$$Y_n = \sup(X_1, \ldots, X_n), \qquad Z_n = \inf(X_1, \ldots, X_n),$$

$$T_n = \frac{2}{n}(X_1 + \cdots + X_n), \qquad T'_n = \frac{n+1}{n} Y_n \qquad \text{et} \qquad T''_n = Y_n + Z_n.$$

1. (a) Déterminer une densité, la fonction de répartition, l'espérance et la variance de la variable aléatoire X.
   (b) Montrer que $(T_n)_{n\in\mathbb{N}^*}$ est une suite convergente d'estimateurs sans biais de θ.
2. (a) Montrer que Y_n est une variable à densité. Déterminer son espérance et sa variance.
   (b) Montrer que $(T'_n)_{n\in\mathbb{N}^*}$ est une suite convergente d'estimateurs sans biais de θ, et comparer $V(T_n)$ et $V(T'_n)$.
3. (a) Montrer que Z_n est une variable aléatoire à densité. Déterminer son espérance et sa variance.
   (b) Retrouver l'égalité $V(Y_n) = V(Z_n)$ sans calcul.
   (c) Montrer que $V(T''_n) \leqslant 4V(Y_n)$.
   (d) Montrer que $(T''_n)_{n\in\mathbb{N}^*}$ est une suite convergente d'estimateurs sans biais de θ.
   (e) Comparer $V(T''_n)$ et $V(T_n)$ lorsque n tend vers $+\infty$.

Exercice 8 – (QC ESCP et HEC 2009) Soit $n \geqslant 1$, et $(X_1, \ldots, X_n)$ un échantillon identiquement distribué indépendant de loi de Poisson de paramètre inconnu $\lambda > 0$.
On pose $\overline{X_n} = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} X_i$ et $T_n = \sqrt{n} \cdot \frac{\overline{X_n} - \lambda}{\sqrt{\lambda}}$. À l'aide de T_n, déterminer, pour n grand, un intervalle de confiance de λ au risque α donné.

Exercice 9 – (Oral ESCP 2011)
Soit X une variable aléatoire définie sur un espace probabilisé $(\Omega, \mathcal{A}, P)$ admettant une variance et dont l'espérance $E(X) = \lambda$ est un paramètre réel inconnu.
Pour n entier $\geqslant 1$, soit $(X_1, X_2, \ldots, X_n)$ un n-échantillon i.i.d. de la loi de X. On pose $\overline{X}_n = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} X_i$.
On note S_λ l'ensemble des statistiques $U_n = g_n(X_1, X_2, \ldots, X_n)$, où g_n est une fonction de $\mathbb{R}^n$ dans $\mathbb{R}$, qui sont des estimateurs sans biais de λ et qui admettent une variance notée V.
On admet la propriété $\mathcal{P}$ suivante :

« Pour tout U_n de S_λ, on a $\mathrm{cov}(\overline{X}_n, U_n - \overline{X}_n) = 0$. »

On dit qu'un élément Z_n de S_λ est un *estimateur optimal* dans S_λ si pour tout U_n de S_λ, on a : $V(Z_n) \leqslant V(U_n)$.

1. Montrer que $\overline{X}_n$ est un estimateur optimal dans S_λ.
2. Soit Z_n un estimateur optimal dans S_λ. Pour α réel et U_n de S_λ, on pose :

$$W_n(\alpha) = \alpha U_n + (1 - \alpha) Z_n.$$

   (a) Montrer que pour tout α réel, $W_n(\alpha)$ est élément de S_λ.

   (b) Calculer $V(W_n(\alpha))$. En déduire que $\mathrm{cov}(Z_n, U_n - Z_n) = 0$.

   (c) Montrer que $Z_n = \overline{X}_n$ presque sûrement.

3. On suppose que X suit la loi de Poisson de paramètre $\lambda > 0$, et on admet que la propriété $(\mathcal{P})$ est vérifiée.

   Pour tout $n \geqslant 2$, on pose : $T_n = \dfrac{1}{n-1}\displaystyle\sum_{i=1}^{n}(X_i - \overline{X}_n)^2$.

   (a) Montrer que T_n est un estimateur sans biais de λ.

   (b) On admet sans démonstration l'existence de $V(T_n)$.

   Montrer que $\mathrm{cov}(T_n, \overline{X}_n) = \frac{\lambda}{n}$.

Exercice 10 – (ESCP 2007) Toutes les variables aléatoires de cet exercice sont définies sur le même espace probabilité $(\Omega, \mathcal{A}, P)$

Soit X une variable aléatoire qui suit la loi normale centrée, d'écart-type σ, le paramètre réel inconnu σ étant strictement positif.

1. Montrer que la variable $T = \frac{X^2}{2\sigma^2}$ suit la loi $\gamma\left(\frac{1}{2}\right)$.
2. Pour tout entier n de $\mathbb{N}^*$, on considère un n-échantillon $(X_1, \cdots, X_n)$ indépendant, identiquement distribué, de la loi de X.

   (a) Donner une densité de la variable aléatoire $S_n = \frac{1}{2\sigma^2}\sum\limits_{i=1}^{n} X_i^2$.

   (b) Soit Y_n définie par $Y_n = \frac{1}{n}\sum\limits_{i=1}^{n} X_i^2$. Montrer que Y_n est un estimateur sans biais de σ^2.

   (c) En justifiant son existence, calculer l'espérance $E(\sqrt{Y_n})$ en fonction de n et de σ.

   En déduire un estimateur $\widehat{\sigma_n}$ sans biais du paramètre σ.

3. (a) En justifiant son existence, calculer la variance $V(\widehat{\sigma_n})$ en fonction de n et de σ.

   (b) On admet que, pour tout réel $x > 0$, on a $\Gamma(x+n) \underset{n\to+\infty}{\sim} n^x(n-1)!$.

   Montrer que $\widehat{s_n}$ est un estimateur convergent de σ.