Probabilités 7 – Estimation

Exercice 1 – On veut estimer les paramètres a et b de la loi uniforme sur $[a,b]$ à l'aide d'un échantillon $(X_1,\dots,X_n)$ de variables aléatoires indépendantes suivant cette loi.

1. (a) Soit S_n la variable à densité définie par $S_n = \sup(X_1,\dots,X_n)$.
   Déterminer la fonction de répartition puis une densité de S_n et son espérance.
   (b) Montrer que la variance de S_n est égale à $\frac{n(b-a)^2}{(n+2)(n+1)^2}$
   (on pourra faire intervenir la variable U_n de densité $x \mapsto nx^{n-1}$ sur $[0,1]$ et nulle ailleurs, ainsi que la variable $a+(b-a)U_n$)
   (c) S_n est-il un estimateur sans biais de b ? asymptotiquement sans biais ? Convergent ?
2. On pose $I_n = \inf(X_1,\dots,X_n)$. Déterminer l'espérance de I_n et sa limite lorsque n tend vers $+\infty$. On admettra que $V(I_n) = V(S_n)$.
3. Exprimer a et b en fonction de $E(I_n)$ et $E(S_n)$, en déduire des estimateurs sans biais de a et b.

Exercice 2 –

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 $M_n = \frac{1}{n}\sum_{k=1}^{n} T_k$ 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 $V_n = \frac{1}{n}\sum_{k=1}^{n}(T_k - M_n)^2$. 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 3 – Soit $(\Omega, \mathcal{T}, P)$ un espace probabilisé. On dit qu'une variable aléatoire X définie sur Ω suit la loi de Pascal de paramètres n et p si $X(\Omega) = [\![n, +\infty[\![$, et si pour tout $k \geqslant n$, $P(X = k) = \binom{k-1}{n-1} p^n q^{k-n}$, où $q = 1-p$.
Soient $T_1,\dots,T_n$ des variables aléatoires indépendantes suivant la même loi géométrique de paramètre p. Pour tout $k \in [\![1,n]\!]$, on pose $S_k = T_1 + \cdots + T_k$.

1. Montrer par récurrence sur k que S_k suit une loi de Pascal de paramètres k et p.
2. Montrer que : $\forall n \in \mathbb{N}$, $\displaystyle\sum_{s=n}^{+\infty} \binom{s}{n} q^{s-n} = \frac{1}{(1-q)^{n+1}}$.
3. On suppose p inconnu, montrer que $P_n = \frac{n-1}{S_n - 1}$ est un estimateur sans biais de p.
4. Calculer, pour $n \geqslant 3$, l'espérance de la variable $\frac{(n-1)^2}{(S_n-1)(S_n-2)}$, et en déduire que P_n est un estimateur convergent.

Exercice 4 – 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 5 – 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 $Y_n = \ln(X_1) + \cdots + \ln(X_n)$.

   (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 6 – (Confidentialité, d'après EDHEC 2000)
Un sondage consiste à proposer l'affirmation A à certaines personnes d'une population donnée. Le sujet abordé étant délicat, le stratagème suivant est mis en place afin de mettre en confiance les personnes sondées pour qu'elles ne mentent pas. L'enquêteur dispose d'un paquet de 20 cartes, numérotées de 1 à 20, qu'il remet à la personne sondée. Celle-ci tire une carte au hasard, et ne la montre pas à l'enquêteur. La règle est alors la suivante :
- Si la carte porte le numéro 1, la personne sondée répond « vrai » si elle est d'accord avec l'affirmation A, et « faux » sinon.
- Si la carte porte un autre numéro, la personne sondée répond « faux » si elle est d'accord avec l'affirmation A et « vrai » sinon.

Le but de l'enquête est d'évaluer la proportion p de gens de la population qui sont réellement d'accord avec l'affirmation A.

1. On interroge une personne selon ce procédé, et l'on considère l'événement suivant, noté V : la personne répond « vrai ». On note $\theta = P(V)$.
   Exprimer θ en fonction de p, puis déduire p en fonction de θ.
2. On considère un échantillon aléatoire, de taille n, extrait de la population considérée, et l'on note S_n le nombre de réponses « vrai » obtenues. On suppose n assez grand pour pouvoir considérer que cet échantillonnage est assimilable à un tirage avec remise.
   (a) Donner la loi de S_n, ainsi que son espérance et sa variance.
   (b) Montrer que $\dfrac{S_n}{n}$ est un estimateur sans biais et convergent de θ.
3. Dans cette question, on suppose que l'on a réalisé un échantillon de 100 personnes, et l'on constate que 23 personnes ont répondu « vrai ».
   (a) Donner une estimation ponctuelle de θ et de p.
   (b) Donner un intervalle de confiance à 95% de θ, puis de p.

Exercice 7 – (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 8 – (Vraisemblance) Soient θ une variable aléatoire réelle discrète dépendant d'un paramètre θ, et $(X_1, \dots, X_n)$ un échantillon iid de variables suivant la même loi que X_θ. Pour toute réalisation $(x_1, \dots, x_n)$ de cet échantillon, la probabilité $P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n)$ dépend de $(x_1, \dots, x_n)$, et aussi de θ. On la note :

$$L(x_1, \dots, x_n, \theta) = P_\theta(X_1 = x_1, \dots, X_n = x_n).$$

La fonction L est la fonction de vraisemblance (Likelihood).

1. (a) Que représente $L(x_1, \dots, x_n, \theta)$?
   (b) Exprimer $L(x_1, \dots, x_n, \theta)$ sous forme d'un produit.
2. Exprimer la fonction de vraisemblance des lois suivantes :
   (a) La loi de Bernoulli de paramètre p.
   (b) Pour k donné, loi binomiale $\mathcal{B}(k, p)$ pour le paramètre p.
   (c) Loi géométrique de paramètre p.
3. Lorsqu'on estime le paramètre θ à partir d'une réalisation $(x_1, \dots, x_n)$ de l'échantillon $(X_1, \dots, X_n)$, il est naturel d'essayer de déterminer la valeur de θ qui rend maximale la probabilité d'observer la réalisation $(x_1, \dots, x_n)$.
   Montrer que rechercher le ou les éventuels maximums L pour un $(x_1, \dots, x_n)$ donné équivaut à rechercher la ou les éventuels maximums de $\ln L$.
4. Le maximum $\hat{\theta}$ éventuellement déterminé ci-dessus dépend de la réalisation $(x_1, \dots, x_n)$. On appelle estimateur du maximum de vraisemblance de θ l'estimateur obtenu en remplaçant $(x_1, \dots, x_n)$ par $(X_1, \dots, X_n)$ dans l'expression de ce maximum.
   (a) Montrer que l'estimateur du maximum de vraisemblance du paramètre p d'une variable de Bernoulli est la fréquence empirique.
   (b) Déteminer l'estimateur du maximum de vraisemblance du paramètre p d'une variable binomiale $\mathcal{B}(k, p)$ dont le paramètre k est connu.
   (c) Déterminer l'estimateur du maximum de vraisemblance du paramètre p d'une loi géométrique.

Exercice 9 – (Oral HEC 2009)
Toutes les variables aléatoires de cet exercice sont définies sur un espace probabilisé $(\Omega, \mathcal{A}, P)$.

1. Question de cours : Rappeler la définition de la convergence en probabilité. Démontrer qu'une suite $(Y_n)_{n\in N^*}$ de variables aléatoires définies sur le même espace probabilisé, suivant des lois de Poisson de paramètres θ_n , tels que $\lim\limits_{n\to+\infty} \theta_n = 0$, converge en probabilité vers 0.
2. On note F la fonction définie sur $\mathbb{R}$ par :

$$F(x) = \begin{cases} 0 & \text{si } x \leqslant -1 \\ \frac{1}{2}(x+1)^2 & \text{si } -1 < x \leqslant 0 \\ 1 - \frac{1}{2}(x-1)^2 & \text{si } 0 < x \leqslant 1 \\ 1 & \text{si } x > 1 \end{cases}$$

   (a) Démontrer que F est de classe $\mathcal{C}^1$ sur $\mathbb{R}$.
   (b) On note f la fonction dérivée de F et, pour tout nombre réel θ, f_θ la fonction définie sur $\mathbb{R}$ par :

$$\forall x \in \mathbb{R}, f_\theta(x) = f(x - \theta).$$

   Vérifier que f_θ est une densité de probabilité.
   (c) Soit X une variable aléatoire définie sur un espace probabilisé $(\Omega, \mathcal{A}, P)$, admettant f_θ pour densité. Montrer que X possède une espérance et une variance et les calculer.
3. Soit $(X_n)_{n\in\mathbb{N}^*}$ une suite de variables aléatoires, définies sur $(\Omega, \mathcal{A}, P)$, indépendantes, de même loi, admettant pour densité f_θ.
   Pour tout entier $n \geqslant 1$, on note $U_n = \inf(X_1, X_2, \dots, X_n)$ et $V_n = \sup(X_1, X_2, \dots, X_n)$.
   (a) Exprimer, à l'aide de F, θ et n, les fonctions de répartition de U_n et V_n.
   (b) Justifier, pour tout $\varepsilon \in]0, 1[$, l'inégalité :

$$P([-1 + \theta \leqslant U_n < -1 + \theta + 2\varepsilon] \cap [1 + \theta - 2\varepsilon < Vn \leqslant 1 + \theta]) \leqslant P\left(\left[\left|\frac{U_n + V_n}{2} - \theta\right| < \varepsilon\right]\right).$$

(c) En déduire que $\dfrac{U_n + V_n}{2}$ est un estimateur convergent de θ.

(d) Est-il sans biais ?

Exercice 10 –

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 $\lim\limits_{n \to +\infty} E(X_1 + \cdots + X_n) = \dfrac{1}{c-1}$.
   (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 11 – (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 $+\infty$.