TRAVAUX DIRIGÉS : ESTIMATION

Exercice 1. Soit X une variable aléatoire qui suit la loi $\mathcal{N}(m,\sigma^2)$, et soit $(X_1,...,X_n)$ un n-échantillon de la loi de X. On désigne par $\overline{X_n}$ la moyenne empirique de $(X_1,...,X_n)$, et l'on pose :

$$V_n = \frac{1}{n}\sum_{k=1}^{n}(X_k - \overline{X_n})^2.$$

(1) Déterminer la loi de la variable aléatoire $\overline{X_n}$.
(2) Montrer que $\overline{X_n}$ est un estimateur sans biais et convergent de m.
(3) Etablir la formule suivante : $E(V_n) = \frac{1}{n}\sum_{k=1}^{n} V(X_k - \overline{X_n})$.
(4) Montrer que $\lim\limits_{n\to+\infty} E(V_n) = \sigma^2$.
(5) En déduire un réel a_n tel que $T_n = a_n V_n$ soit un estimateur sans biais de σ^2.

Exercice 2. On se propose d'estimer le paramètre θ de la loi uniforme sur $[0,\theta]$ à l'aide d'un échantillon $(X_1,...,X_n)$ de variables aléatoires indépendantes et qui suivent cette loi. Pour ce faire, on pose :

$$\overline{X_n} = \frac{1}{n}\sum_{k=1}^{n} X_k \quad \text{et} \quad M_n = \sup\{X_1,...,X_n\}.$$

(1) Déterminer en fonction de $\overline{X_n}$ un estimateur sans biais T_n de θ.
(2) Justifier que la variable aléatoire M_n est à densité et en donner une densité.
(3) Déterminer en fonction de M_n un estimateur sans biais U_n de θ.
(4) Les estimateurs T_n et U_n de θ sont-ils convergents? Justifier.
(5) Qui, de T_n ou de U_n, est le meilleur estimateur de θ? Justifier.

Exercice 3. On dit qu'une variable aléatoire X définie sur un espace probabilisé $(\Omega,\mathcal{A},P)$ suit la loi de Pascal de paramètres $n\in\mathbb{N}^*$ et $p\in]0,1[$ si $X(\Omega) = [\![n,+\infty[\![$, et si pour tout $k\geq n$, on a :

$$P([X=k]) = \binom{k-1}{n-1}p^n q^{k-n} \quad \text{où} \quad q = 1-p.$$

Soient $T_1,...,T_n$ des variables aléatoires indépendantes suivant toutes la loi géométrique de paramètre p. Pour tout $k\in\{1,...,n\}$, on pose $S_k = T_1+...+T_k$. *Par la suite, on admet que, pour tout* $m\in\mathbb{N}$:

$$\sum_{i=m}^{+\infty}\binom{i}{m}q^{i-m} = \frac{1}{(1-q)^{m+1}}.$$

(1) Montrer par récurrence sur k que S_k suit la loi de Pascal de paramètres k,p.
(2) On suppose p inconnu. Montrer que $P_n = \dfrac{n-1}{S_n-1}$ est un estimateur sans biais de p.
(3) Calculer l'espérance de la variable aléatoire $X_n = \dfrac{(n-1)^2}{(S_n-1)(S_n-2)}$ pour tout $n\geq 3$.
(4) Montrer que $E(X_n)\geq E(P_n^2)$ et en déduire une majoration de $V(P_n)$ en fonction de p et n.
(5) En déduire que P_n est un estimateur convergent de p.

Exercice 4. (ESCP 2016) Soit X une variable aléatoire qui suit la loi de Poisson de paramètre inconnu $\theta>0$. Pour $n\geq 1$, on considère un n-échantillon $(X_1,...,X_n)$ de la loi de X et l'on pose :

$$M_n = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

(1) Montrer que, pour tout $c>0$, la variable aléatoire e^{-cM_n} admet une espérance et la calculer.
(2) Pour tout $n\geq 2$, on pose $c_n = n\ln\left(\dfrac{n}{n-1}\right)$. On choisit $T_n = e^{-c_n M_n}$ comme estimateur de $e^{-\theta}$.
   (a) Montrer que T_n est un estimateur sans biais de $e^{-\theta}$.
   (b) Calculer la variance de T_n.
   (c) En déduire que T_n est un estimateur convergent de $e^{-\theta}$.

Exercice 5. Soit m un réel > 0. On admet que la mesure d'une grandeur physique, dont la valeur exacte est égale à m, suit la loi normale d'espérance m et de variance $\frac{m^2}{100}$. On effectue une série de n mesures indépendantes notées $X_1, ..., X_n$, et on désigne par Y_n la moyenne des résultats obtenus. Enfin, on définit *l'erreur relative commise sur m* par :

$$Z_n = \frac{Y_n - m}{m}.$$

(1) Justifier que Y_n est un estimateur sans biais et convergent de m, puis donner la loi de Y_n.
(2) A l'aide de l'inégalité de Bienaymé-Tchebychev, déterminer combien de mesures il faut effectuer pour que l'erreur relative commise sur m soit inférieure à 1% avec une probabilité supérieure à $0,9$.
(3) Répondre à la question précédente à l'aide de la loi de Y_n (*note :* $\Phi^{-1}(0,95) \simeq 1,645$). Que constate-t-on?

Exercice 6. Afin d'étudier le pourcentage p de consommateurs satisfaits par un produit A, on réalise un sondage auprès de 100 consommateurs, au cours duquel 56 d'entre eux se sont déclarés satisfaits par A.

(1) A l'aide de l'inégalité de Bienaymé-Tchebychev, calculer un intervalle de confiance de p à 95%.
(2) Calculer un intervalle de confiance asymptotique de p à 95% (*note :* $\Phi^{-1}(0,975) \simeq 1,96$).

Exercice 7. (ESCP 2016) Soit $\theta \in \mathbb{R}_+^*$ et soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires indépendantes de même loi ayant pour densité la fonction f définie par :

$$f(t) = \begin{cases} \dfrac{2t}{\theta^2} & \text{si} \quad t \in [0, \theta] \\ 0 & \text{si} \quad t \notin [0, \theta] \end{cases}.$$

(1) Pour tout $n \in \mathbb{N}^*$, on pose : $\overline{X}_n = \dfrac{1}{n}\displaystyle\sum_{k=1}^{n} X_k$.
 (a) Calculer $E(X_1)$ et en déduire que, pour tout $n \in \mathbb{N}^*$, $T_n = \dfrac{3}{2}\overline{X}_n$ est un estimateur sans biais de θ.
 (b) Etudier la convergence en probabilités de $(T_n)_{n\geq 1}$.
(2) A l'aide du théorème limite central, déterminer un intervalle de confiance au niveau de confiance 95% pour θ basé sur l'estimateur $\overline{X}_n$ (on rappelle que, si Φ est la fonction de répartition de la loi $\mathcal{N}(0,1)$, alors $2\Phi(t) - 1 = 0,95$ pour $t = 1,96$).
(3) Pour tout $n \geq 1$, on pose : $M_n = \sup\{X_1, ..., X_n\}$.
 (a) Calculer la fonction de répartition G_n de M_n.
 (b) Soit $\delta > 0$. Etudier la convergence de la série $\sum_{n\geq 1} P(|M_n - \theta| > \delta)$.
 (c) Calculer $E(M_n)$ et étudier les propriétés de $M_n' = \dfrac{2n+1}{2n} M_n$ en tant qu'estimateur de θ.
 (d) Quel est, de T_n et de M_n', le meilleur estimateur de θ? Justifier.

Exercice 8. (ESCP 2017) Soit $\sigma \in \mathbb{R}_+^*$ et soit X une variable aléatoire qui suit la loi $\mathcal{N}(0, \sigma^2)$.

(1) Montrer que $U = \dfrac{X^2}{2\sigma^2}$ suit une loi γ dont on précisera le paramètre.

 Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires indépendantes de même loi que X.

(2) (a) Pour tout $n \in \mathbb{N}^*$, déterminer la loi de $S_n = \dfrac{1}{\sigma^2}\displaystyle\sum_{i=1}^{n} X_i^2$.
 (b) Pour tout $n \in \mathbb{N}^*$, montrer que $Y_n = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} X_i^2$ est un estimateur sans biais de σ^2.
(3) Pour tout $n \geq 1$, on pose : $T_n = \dfrac{\Gamma(n/2)}{\Gamma((n+1)/2)}\sqrt{\dfrac{nY_n}{2}}$.
 (a) Montrer que $\sqrt{Y_n}$ admet une espérance et la calculer en fonction de n et σ.
 (b) En déduire que T_n est un estimateur sans biais de σ.
 (c) Montrer que T_n admet une variance et la calculer en fonction de n et σ.
 (d) Montrer que T_n est un estimateur convergent de σ.
 (*On admetta que, pour tout $x > 0$, on a :* $\Gamma(n+x) \underset{n\to+\infty}{\sim} n^x (n-1)!$).

1. Exercices supplémentaires

Exercice 9. Soit T une variable aléatoire qui suit la loi de Poisson de paramètre λ, et soient $T_1, ..., T_n$ des variables aléatoires indépendantes de même loi que T. On pose :

$$M_n = \frac{1}{n}\sum_{k=1}^{n} T_k \quad \text{et} \quad V_n = \frac{1}{n}\sum_{k=1}^{n}(T_k - M_n)^2.$$

(1) Calculer les moments d'ordre $1, 2, 3$ de T.
(2) Montrer que M_n est un estimateur sans biais de λ.
(3) Calculer $E(T_k^2)$ $E(M_n^2)$ et $E(T_k M_n)$ pour tout $k \in \{1, ..., n\}$.
(4) Est-ce que V_n est un estimateur sans biais de λ? Justifier.
(5) Proposer un estimateur W_n sans biais de λ construit à partir de V_n.
(6) On admet que $V(W_n) = \frac{n\lambda(1+2\lambda)}{(n-1)^2}$. Quel est, entre M_n et W_n, le meilleur estimateur de λ?

Exercice 10. Soit α un paramètre inconnu réel > 0. On considère une variable aléatoire à densité X, dont une densité f est définie par $f(x) = \dfrac{\alpha}{x^{\alpha+1}}$ si $x \geq 1$ et $f(x) = 0$ sinon. De plus, on considère un n-échantillon $(X_1, ..., X_n)$ de la loi de X, et l'on pose $Y_n = \ln(X_1) + ... + \ln(X_n)$.

(1) Déterminer la loi de la variable aléatoire $\ln(X)$.
(2) Calculer l'espérance de Y_n, et en déduire un estimateur sans biais de $\frac{1}{\alpha}$.
(3) Quelle est la variance de cet estimateur? Conclusion?

Exercice 11. Soit a un réel > 0. Soit X une variable aléatoire telle que $Y = X - a$ suit la loi exponentielle de paramètre 1, et soit $(X_1, ..., X_n)$ un n-échantillon de la loi de X. On pose :

$$Z_n = \frac{1}{n}\sum_{k=1}^{n} X_k \quad \text{et} \quad T_n = \inf\{X_1, ..., X_n\}.$$

(1) Etablir que la variable aléatoire $W_n = T_n - a$ suit la loi exponentielle de paramètre n.
(2) Montrer que $Z'_n = Z_n - 1$ est un estimateur sans biais et convergent de a.
(3) Montrer que $T'_n = T_n - \dfrac{1}{n}$ est un estimateur sans biais et convergent de a.
(4) Quel est, entre les estimateurs Z'_n et T'_n de a, le meilleur des deux? Justifier.

Exercice 12. On considère un n-échantillon $(X_1, ..., X_n)$ de la loi de Poisson de paramètre $\lambda > 0$. On suppose que λ est inconnu et on cherche à l'estimer par un intervalle de confiance. On pose :

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \quad \text{et} \quad T_n = \sqrt{n}\frac{\overline{X}_n - \lambda}{\sqrt{\lambda}}.$$

(1) Donner une approximation de la loi de T_n.
(2) Soit $\alpha \in]0, 1[$ et posons $t_\alpha = \Phi^{-1}(1 - \frac{\alpha}{2})$. Résoudre l'inéquation $(\lambda - \overline{X}_n)^2 \leq \frac{\lambda}{n} t_\alpha^2$ d'inconnue λ.
(3) Déterminer en fonction de $\overline{X_n}$ un intervalle de confiance asymptotique de λ au niveau de confiance $1 - \alpha$.

Exercice 13. (ESCP 2009) Soit $\theta \in \mathbb{R}_+^*$, soit X une variable aléatoire qui suit la loi $\mathcal{U}(]0, \theta[)$ et soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires indépendantes de même loi que X. Pour tout $n \geq 1$, on pose :

$$Y_n = \sup(X_1, ..., X_n), \quad Z_n = \inf(X_1, ..., X_n), \quad T_n = \frac{2}{n}(X_1 + ... + X_n), \quad T'_n = \frac{n+1}{n} Y_n, \quad T''_n = Y_n + Z_n.$$

(1) (a) Déterminer une densité, la fonction de répartition, l'espérance et la variance de X.
    (b) Montrer que (T_n) est une suite convergente d'estimateurs sans biais de θ.
(2) (a) Montrer que Y_n est une variable à densité, puis déterminer son espérance et sa variance.
    (b) Montrer que (T'_n) est une suite convergente d'estimateurs sans biais de θ.
    (c) Comparer $V(T_n)$ et $V(T'_n)$. Quel est, de T_n et de T'_n, le meilleur estimateur de θ?
(3) (a) Montrer que Z_n est une variable à densité, puis déterminer son espérance et sa variance.
    (b) Retrouver l'égalité $V(Y_n) = V(Z_n)$ sans calcul.
    (c) Montrer que $V(T''_n) \leq 4V(Y_n)$.
    (d) Montrer que (T''_n) est une suite convergente d'estimateurs sans biais de θ.
    (e) Comparer $V(T''_n)$ et $V(T_n)$ lorsque n tend vers $+\infty$.
    (f) Quel est, de T_n et de T''_n, le meilleur estimateur de θ?

Exercice 14. (Méthode de capture-recapture - ESCP 2012) On cherche à évaluer le nombre N de lions d'Asie, espèce en voie de disparition, encore en vie dans la forêt de Gir. Pour cela, on capture d'abord en une seule fois m lions (avec $m \in \mathbb{N}^*$) que l'on tatoue avant de les relâcher dans la nature. On admet que, pendant toute la durée de l'étude, il n'y a ni décès ni naissance, puis on utilise l'une des deux méthodes suivantes.

(1) **Première méthode.**
On capture successivement, au hasard (donc avec équiprobabilité) et avec remise en liberté après observation du sujet, n lions. Soit Y_n le nombre de lions tatoués parmi eux.
 (a) Déterminer la loi de Y_n. En déduire que $\frac{1}{nm}Y_n$ est un estimateur sans biais et convergent de $\frac{1}{N}$.
 (b) Pourquoi ne peut-on pas prendre $\frac{nm}{Y_n}$ comme estimateur de N? Justifier.
 (c) On pose $B_n = \frac{m(n+1)}{Y_n+1}$. Calculer l'espérance de B_n et montrer que $\lim\limits_{n\to+\infty} B_n = N$.

(2) **Deuxième méthode.**
On se donne $n \in \mathbb{N}^*$. On capture également, un à un, des lions de Gir au hasard et avec remise en liberté après l'observation du sujet. Pour tout $i \in [\![1, n]\!]$, on désigne par X_i le nombre de lions qu'il a été juste nécessaire de capturer pour en obtenir i tatoués. On pose $D_1 = X_1$ et pour tout $i \in [\![2, n]\!]$, $D_i = X_i - X_{i-1}$. On admet que les variables aléatoires $D_1, ..., D_n$ sont indépendantes.
 (a) Pour tout $i \in [\![2, n]\!]$, que représente concrètement D_i?
 (b) Pour tout $i \in [\![2, n]\!]$, déterminer la loi de D_i, son espérance et sa variance. 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 et convergent de N et déterminer son risque quadratique.
 (d) Pour n assez grand, par quelle loi peut-on approcher la loi de $\widetilde{X}_n = \frac{1}{n}X_n$?
 (e) On a tatoué $m = 200$ lions, puis capturé 450 lions, pour obtenir $n = 50$ lions marqués. On désigne par σ l'écart-type de A_{50}, et on a pu prouver que $\sigma \leq 100$. Déterminer un intervalle de confiance pour N au niveau de confiance $0,9$ (on rappelle que $\Phi(1,64) \simeq 0,95$).

Exercice 15. (HEC 2018) Soit (X_n) une suite de variables aléatoires indépendantes et suivant toutes la loi uniforme sur $[-\theta, \theta]$, où θ est un paramètre inconnu. Pour tout $n \in \mathbb{N}^*$, on pose :

$$U_n = \inf\{X_1, ..., X_n\} \quad \text{et} \quad V_n = \sup\{X_1, ..., X_n\}.$$

(1) Montrer que V_n est un estimateur convergent de θ.
(2) Pour tout $n \in \mathbb{N}^*$, on pose : $T_n = \frac{1}{2}\sup\{X_i - X_j, (i,j) \in [\![1, n]\!]\}$.
 (a) Exprimer T_n en fonction de U_n et V_n.
 (b) En déduire que T_n est un estimateur convergent de θ.

Exercice 16. (ESCP 2021) Soit n un entier ≥ 2, soit $\theta \in \left[-\frac{1}{2}, \frac{1}{2}\right]$ et posons : $f_\theta(t) = \begin{cases} \dfrac{1-\theta}{2} & \text{si } t \in [-1, 0[\\ \dfrac{1+\theta}{2} & \text{si } t \in [0, 1] \\ 0 & \text{sinon} \end{cases}$.

Par la suite, on désigne par X une variable aléatoire admettant f_θ comme densité et par $(X_1, ..., X_n)$ un n-échantillon de la loi de X.

(1) On pose : $F_n = \dfrac{1}{n}\sum\limits_{k=1}^{n} X_k$.
 (a) Déterminer un réel c tel que $\widehat{F_n} = cF_n$ soit un estimateur sans biais de θ.
 (b) Montrer que $\widehat{F_n}$ est un estimateur convergent de θ.
(2) Pour tout $k \in [\![1, n]\!]$, on désigne par Y_n le nombre de variables parmi les variables X_k qui ont pris une valeur positive ou nulle.
 (a) Montrer que Y_n suit une loi binomiale dont on donnera les paramètres.
 (b) Montrer que $\widehat{\theta_n} = \dfrac{2}{n}Y_n - 1$ est un estimateur sans biais de θ.
 (c) Montrer que $\widehat{\theta_n}$ est un estimateur convergent de θ.
(3) (a) Déterminer un réel $\lambda > 0$ tel que : $\left|\sqrt{1 - \widehat{\theta_n}^2} - \sqrt{1-\theta^2}\right| \leq \lambda \left|\widehat{\theta_n} - \theta\right|$.
 (b) Montrer que $\sqrt{1 - \widehat{\theta_n}^2}$ est un estimateur convergent de $\sqrt{1-\theta^2}$.
 (c) Montrer que la suite $\left(\sqrt{n}\dfrac{\widehat{\theta_n} - \theta}{\sqrt{1-\theta^2}}\right)$ converge en loi vers une variable normale centrée réduite.