Estimation ponctuelle

Exercice 1

On considère la variable aléatoire X dont la loi est donnée par

$$P(X=-1)=p, \quad P(X=0)=1-2p, \quad P(X=1)=p$$

pour un certain paramètre $p \in]0;\frac{1}{2}[$. On dispose d'un n-échantillon $(X_1,\dots,X_n)$ de X, et on cherche à déterminer le paramètre p.

1. Calculer $E(\bar{X}_n)$. En déduire que l'estimateur $\bar{X}_n$ est biaisé.
2. Peut-on trouver des réels a et b tels que aX_n+b soit un estimateur sans biais de p ?
3. On note $T_n=\dfrac{1}{2n}\displaystyle\sum_{i=1}^{n} X_i^2$.
   (a) Montrer que T_n est un estimateur sans biais de p.
   (b) Calculer la variance $V_\theta(T_n)$.
   (c) Montrer que l'estimateur est convergent.

Exercice 2

Soit a un réel strictement positif. On note f la fonction définie sur $\mathbb{R}$ par

$$f(x)=\begin{cases} 0, & \text{si } x<a \\ \dfrac{3a^3}{x^4}, & \text{si } x\geqslant a\end{cases}$$

1. Montrer que f est une densité de probabilité.
   Un capteur mesure en permanence le taux de gaz carbonique émis par un moteur. On suppose que le temps écoulé entre le démarrage du moteur et l'instant précis (en heures) où son taux de gaz carbonique devient non réglementaire est une variable aléatoire T de densité f.
2. Montrer que T admet une espérance et une variance de valeurs :
$$E(T)=\frac{3a}{2}, \quad V(t)=\frac{3a^2}{4}.$$
3. (a) Déterminer la fonction de répartition de T.
   (b) Calculer les probabilités $P(T>2a)$ et $P_{(T>2a)}(T>6a)$.
4. On met en route n moteurs de modèle identique au précédent, et indépendants. On note $T_1,T_2,\dots,T_n$ les temps respectifs pendant lesquels ces moteurs ont un taux de gaz carbonique réglementaire ($T_1,T_2,\dots,T_n$ suivent donc la même loi que T et sont indépendantes).
   (a) Montrer que la variable
$$Z_n=\frac{2}{3n}\sum_{k=1}^{n} T_k$$
   est un estimateur sans biais du paramètre a.
   (b) Calculer son risque quadratique $r(Z_n)$. En déduire que T_n est un estimateur convergent.

Exercice 3

. Soit (X_k) une suite de v.a.i.i.d. suivant une loi de Poisson $\mathcal{P}(\lambda)$, où $\lambda>0$. On pose, pour tout $n\in\mathbb{N}^*, S_n=X_1+X_2+\ldots+X_n$.

1. (a) Montrer que $S_2\hookrightarrow\mathcal{P}(2\lambda)$.
   (b) En déduire, par récurrence sur $n\in\mathbb{N}^*$, que $S_n\hookrightarrow\mathcal{P}(n\lambda)$.
2. Pour $n\in\mathbb{N}, n\geqslant 2$, on pose, $Y_n=\left(1-\dfrac{1}{n}\right)^{S_n}$.
   (a) Montrer que Y_n est une v.a. discrète à valeurs dans $]0;1]$.
   (b) Déterminer la loi de Y_n, son espérance et sa variance.
   (c) En déduire un estimateur sans biais de $e^{-\lambda}$.

Exercice 4

La sécurité routière fait une enquête sur le nombre d'accidents survenus par semaine sur un tronçon d'autoroute. Soit X la v.a. égale au nombre d'accidents en une semaine. On suppose que X suit une loi $\mathcal{P}(\lambda), \lambda>0$. On se propose d'estimer le paramètre $e^{-\lambda}=P(X=0)$. On note $(X_1,\dots,X_n)$ un n-échantillon de X.

1. Soit Y_n le nombre de fois où l'on a pas observé d'accident pendant la semaine, i.e.
$$Y_n=\text{Card}\left(\{i\in[\![1;n]\!]; X_i=0\}\right).$$
   (a) Montrer que Y_n/n est un estimateur non biaisé de $e^{-\lambda}$.
   (b) Déterminer son risque quadratique, noté ici $r(Y_n/n)$.
2. Vérifier que $\bar{X}_n$ est un estimateur sans biais de λ.

3. On pose $S_n = X_1 + \ldots + X_n$. Quelle est la loi de S_n ? À l'aide du théorème de transfert, déterminer l'espérance de $e^{-\bar{X}_n}$. En déduire que $e^{-\bar{X}_n}$ est un estimateur biaisé de $e^{-\lambda}$.

Exercice 5 — EDHEC 2014

Dans cet exercice, θ désigne un réel strictement positif et n un entier naturel supérieur ou égal à 2 . Pour tout k de $\mathbb{N}$, on pose

$$u_k = \frac{1}{1+\theta}\left(\frac{\theta}{1+\theta}\right)^k$$

1. Montrer que la suite (u_k) définit bien une loi de probabilité.

   On considère maintenant une variable aléatoire X prenant ses valeurs dans $\mathbb{N}$ et dont la loi est donnée par

$$\forall k \in \mathbb{N}, \quad P(X=k) = u_k$$

2. (a) On pose $Y = X + 1$. Reconnaître la loi de Y et en déduire l'espérance et la variance de X.

   (b) Compléter la fonction Python suivante pour qu'elle simule la loi d'une variable aléatoire X :

```
def simul_X(theta):
        Y=1
        while ...:
                Y=Y+1
        return ...
```

3. Dans cette question, on souhaite estimer le paramètre θ par la méthode du maximum de vraisemblance. Pour ce faire, on considère un échantillon $(X_1, X_2, \ldots, X_n)$ composé de variables aléatoires indépendantes ayant toutes la même loi que X et on introduit la fonction L, de $\mathbb{R}_+^*$ dans $\mathbb{R}$, définie par

$$\forall \theta \in \mathbb{R}_+^*, \quad L(\theta) = \prod_{k=1}^{n} P(X_k = x_k)$$

   où $x_1, x_2, \ldots, x_n$ désignent des entiers naturels éléments de $X(\Omega)$. L'objectif est de choisir la valeur de θ qui rend $L(\theta)$ maximale.

   (a) Écrire $\ln(L(\theta))$ en fonction de θ et de $S_n = \sum_{k=1}^{n} x_k$.

   (b) On considère la fonction φ, définie par

$$\forall \theta \in]0; +\infty[, \quad \varphi(\theta) = S_n \ln\theta - (S_n + n)\ln(1+\theta)$$

   Montrer que la fonction φ admet un maximum, atteint en un seul réel que l'on notera $\hat{\theta_n}$ et que l'on exprimera en fonction de S_n. Que représente $\hat{\theta_n}$ pour la fonction L ?

   On pose dorénavant

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

   La variable T_n est appelée estimateur du maximum de vraisemblance pour θ.

   (c) Vérifier que T_n est un estimateur sans biais de θ.

   (d) Calculer le risque quadratique $r_{T_n}(\theta)$ de T_n et vérifier que $\lim\limits_{n\to+\infty} r_{T_n}(\theta) = 0$.

Exercice 6 — EDHEC 2020

On considère une variable aléatoire X suivant la loi normale $\mathcal{N}(0, \sigma^2)$, où σ est strictement positif dont on note f_X une densité et F_X la fonction de répartition.

1. Montrer que : $\forall x \in \mathbb{R}, F_X(-x) = 1 - F_X(x)$.
2. On pose $Y = |X|$ et on admet que Y est une variable aléatoire.

   (a) Montrer que la fonction de répartition de Y est la fonction, notée F_Y, définie par :

$$F_Y(x) = \begin{cases} 2F_X(x) - 1, \text{ si } x \geqslant 0 \\ 0, \text{ si } x < 0 \end{cases}$$

   (b) En déduire que Y est une variable à densité et donner une densité f_Y de Y.

   (c) Montrer que Y possède une espérance et que l'on a $E(Y) = \sigma\sqrt{\frac{2}{\pi}}$.

3. On suppose, dans cette question seulement, que σ est inconnu et on se propose de l'estimer. Soit n un entier naturel supérieur ou égal à 1 . On considère un échantillon $(Y_1, Y_2, \ldots, Y_n)$ composé de variables aléatoires, mutuellement indépendantes, et ayant toutes la même loi que Y On note S_n la variable aléatoire définie par $S_n = \frac{1}{n}\sum_{k=1}^{n} Y_k$.

   (a) Montrer que S_n est un estimateur de σ, donner la valeur de son biais, puis proposer un estimateur sans biais de σ, que l'on notera T_n, construit de façon affine à partir de S_n.

   (b) Rappeler la valeur du moment d'ordre 2 de X, puis déterminer $E(Y^2), V(Y)$ et $V(S_n)$.

   (c) Déterminer le risque quadratique de T_n en tant qu'estimateur de σ. En déduire que T_n est un estimateur convergent de σ.

Exercice 7 — EDHEC 2012

On désigne par λ, un réel strictement positif et on considère la fonction f, définie sur $\mathbb{R}$, par

$$f(x) = \lambda|x|e^{-\lambda x^2}$$

1. (a) Montrer que f est paire.
   (b) Établir la convergence et calculer la valeur de l'intégrale $\displaystyle\int_0^{+\infty} f(x)\mathrm{d}x$.
   (c) Montrer que la fonction f peut être considérée comme densité d'une variable aléatoire X que l'on suppose, dans la suite, définie sur un certain espace probabilisé $(\Omega; \mathcal{A}; P)$.
2. (a) Justifier la convergence de l'intégrale $\displaystyle\int_0^{+\infty} xf(x)\mathrm{d}x$.
   (b) En déduire que la variable aléatoire X possède une espérance, notée $E(X)$, et donner sa valeur.
3. (a) Montrer, grâce à une IPP, la convergence et donner la valeur de l'intégrale $\displaystyle\int_0^{+\infty} x^2f(x)\mathrm{d}x$.
   (b) En déduire que la variable aléatoire X possède une variance, notée $V(X)$, et donner sa valeur.
4. On pose $Y = X^2$ et on admet que Y est une variable aléatoire à densité, elle aussi définie sur le même espace probabilisé.
   (a) Donner l'expression de la fonction de répartition F_Y de la variable aléatoire Y à l'aide de la fonction de répartition F_X de la variable aléatoire X.
   (b) Déterminer une densité f_Y de Y, puis vérifier que Y suit la loi exponentielle de paramètre λ.
   (c) Retrouver alors sans calcul la valeur de $V(X)$.
5. Soit U une variable aléatoire suivant la loi uniforme sur $[0; 1\ [$. On pose $W = -\dfrac{1}{\lambda}\ln(1-U)$ et on admet que W est une variable aléatoire.
   (a) Déterminer la fonction de répartition de W et en déduire la loi suivie par la variable aléatoire W.
   (b) Vérifier que la probabilité que X prenne des valeurs positives est égale à la probabilité que X prenne des valeurs négatives.

   On suppose, dans la suite, que le paramètre λ est inconnu et on souhaite l'estimer en utilisant la loi de Y.

   On désigne par n un entier naturel supérieur ou égal à 2 et on considère un échantillon $Y_1, Y_2, .., Y_n$ de la loi de Y, c'est à dire des variables $Y_1, Y_2, .., Y_n$ définies sur $(\Omega; \mathcal{A}; P)$, indépendantes et de même loi que Y.
6. On considère des réels $x_1, x_2, .., x_n$ strictement positifs, ainsi que la fonction L, à valeurs dans $\mathbb{R}$, définie sur $]0; +\infty[$ par

$$L(\lambda) = \prod_{k=1}^{n} f_Y(x_k), \quad \lambda > 0$$

   (a) Exprimer $L(\lambda)$, puis $\ln(L(\lambda))$ en fonction de $\lambda, x_1, \ldots, x_n$.
   (b) On considère la fonction φ, définie pour tout réel $\lambda > 0$ par

$$\varphi(\lambda) = n\ln(\lambda) - \lambda\sum_{=1}^{n} x_k$$

   Montrer que la fonction φ admet un maximum, atteint en un seul réel que l'on notera z et que l'on exprimera en fonction de $x_1, x_2, \ldots, x_n$. Que peut-on dire de z pour la fonction L ?
7. On pose dorénavant, toujours avec $n \geqslant 2$,

$$Z_n = \frac{n}{Y_1 + Y_2 + \ldots + Y_n}$$

   On admet que Z_n est une variable aléatoire définie, elle aussi, sur l'espace probabilisé $(\Omega; \mathcal{A}; P)$. La suite $(Z_n)_{n\geqslant 2}$ est appelée estimateur du maximum de vraisemblance pour λ.
   On admet que la variable aléatoire $\bar{Y}_n = Y_1 + Y_2 + \ldots + Y_n$ admet pour densité la fonction f_n définie par

$$f_n(t) = \begin{cases} 0, & \text{si } t < 0 \\ \dfrac{\lambda^n}{(n-1)!}t^{n-1}e^{-\lambda t}, & \text{si } t \geqslant 0 \end{cases}.$$

   (a) En remarquant que $\displaystyle\int_0^{+\infty} f_{n-1}(t)\mathrm{d}t = 1$, montrer que Z_n possède une espérance et que

$$E(Z_n) = \frac{n}{n-1}\lambda$$

   (b) Déterminer un estimateur non biaisé de λ, noté Z'_n, construit à partir de Z_n.

Estimation par intervalles de confiance

Exercice 8 — ESSEC II 2016

Soit X d'espérance m et de variance σ^2. On considère alors un n-échantillon $(X_1, \dots, X_n)$ de X et on pose $T_n = X_1 + X_2 + \dots + X_n$. On suppose que σ^2 est connue, mais pas m.

1. Déterminer $E(T_n)$ et $V(T_n)$.
2. Montrer que, pour tout $\varepsilon > 0$
$$P(|T_n - nm| > n\varepsilon) \leqslant \frac{\sigma^2}{n\varepsilon^2}$$
3. Quel est le risque de l'intervalle de confiance
$$\left[\frac{T_n}{n} - \varepsilon; \frac{T_n}{n} + \varepsilon\right]$$
pour m ?

Exercice 9

Soit $n \in \mathbb{N}^*$. On considère un échantillon $(X_1, \dots, X_n)$ de la loi normale $\mathcal{N}(m, m^2/25)$, avec $m > 0$, le paramètre m étant inconnu.

Soit $\alpha \in]0;1[$ et t_α le réel positif tel que $\Phi(t_\alpha) = 1 - \dfrac{\alpha}{2}$. On suppose que $n > \dfrac{t_\alpha^2}{25}$. On note

$$\overline{X_n} = \frac{X_1 + \dots + X_n}{n}, \quad \text{et} \quad Y_n = 5\sqrt{n}\frac{\overline{X_n} - m}{m}$$

1. Justifier que la variable aléatoire Y_n converge en loi vers une loi normale centrée réduite.
2. Justifier alors que pour n assez grand, on peut écrire
$$P(|Y_n| \leqslant t_\alpha) = 1 - \alpha$$
3. Montrer que l'intervalle
$$\left[5\sqrt{n}\frac{\overline{X_n}}{5\sqrt{n} + t_\alpha}; 5\sqrt{n}\frac{\overline{X_n}}{5\sqrt{n} - t_\alpha}\right]$$
est un intervalle de confiance de m au niveau de confiance $1 - \alpha$.
4. Pour $n = 100$, une réalisation de ce n-échantillon nous donne une moyenne empirique de 12 .
Déterminer une estimation d'un intervalle de confiance de m à 95%. On donne $\Phi(1,96) = 0,975$.

Exercice 10

Soient $n \in \mathbb{N}^*$ et $a > 0$. On considère une suite de v.a. (X_n) dont la fonction de répartition est donnée par

$$F_{X_n}(t) = \begin{cases} 0, \text{ si } x < 0 \\ 1 - \exp\left(-ax - \dfrac{x^2}{2n}\right), \text{ si } 0 \leqslant x < 2n \\ 1, \text{ si } x \geqslant 2n \end{cases}$$

1. Montrer que (X_n) converge en loi vers une v.a dont on précisera la loi.
2. Soient $Z \hookrightarrow \mathcal{E}(1)$ et $\alpha \in]0;1[$.
   (a) Déterminer deux réels c et d strictement positifs tels que
   $$P(c \leqslant Z \leqslant d) = 1 - \alpha, \quad \text{et} \quad P(Z \leqslant c) = \frac{\alpha}{2}$$
   (b) Quelle est la loi de aZ ?
   (c) Montrer que
   $$\lim_{n \to +\infty} P\left(a \in \left[\frac{c}{X_n}; \frac{d}{X_n}\right]\right) = 1 - \alpha$$
   (d) Que peut-on dire de l'intervalle
   $$\left[\frac{c}{X_n}; \frac{d}{X_n}\right]?$$

Exercice 11

Soit $p \in]0;1[$, on considère une variable aléatoire X qui suit la loi géométrique de paramètre p. On pose $q = 1 - p$. Pour tout $n \in \mathbb{N}^*$, on considère un n-échantillon $(X_1, \dots, X_n)$ de la loi de X. On pose

$$S_n = \sum_{k=1}^{n} X_k, \quad Y_n = \frac{n}{S_n} \quad \text{et} \quad \bar{X}_n = \frac{1}{Y_n} = \frac{1}{n}S_n$$

1. Montrer que $\bar{X}_n$ est un estimateur sans biais de $\dfrac{1}{p}$. Quel est son risque quadratique ?
2. Pour tout $n \in \mathbb{N}^*$, on pose $T_n = \sqrt{\dfrac{n}{p^2 q}}(Y_n - p)$. On admet que T_n converge en loi vers une variable aléatoire $T \hookrightarrow \mathcal{N}(0,1)$.

(a) Soit $\alpha \in]0;1[$ et a_α l'unique réel vérifiant $P([T > a_\alpha]) = \dfrac{\alpha}{2}$.
Montrer que $P(-a_\alpha \leqslant T \leqslant a_\alpha) = 1-\alpha$.

(b) Pour n assez grand on considère alors que $P(-a_\alpha \leqslant T_n \leqslant a_\alpha) = 1-\alpha$. En déduire que :

$$P\left(Y_n - a_\alpha p\sqrt{\frac{q}{n}} \leqslant p \leqslant Y_n + a_\alpha p\sqrt{\frac{q}{n}}\right) = 1-\alpha.$$

(c) Étudier la fonction $f : x \mapsto x\sqrt{1-x}$ sur l'intervalle $[0;1]$.

(d) Déduire des deux questions précédentes que pour n assez grand :

$$P\left(Y_n - \frac{2a_\alpha}{3\sqrt{3n}} \leqslant p \leqslant Y_n + \frac{2a_\alpha}{3\sqrt{3n}}\right) \geqslant 1-\alpha$$

(e) On suppose que $n = 900$, une réalisation de l'échantillon $(X_1, \ldots, X_{900})$ a donné la valeur 4 à $\bar{X}_{900}$.
Donner alors la réalisation y_{900} de la variable Y_{900}.
On se donne un niveau de risque $\alpha = 0,05$, le nombre $a_{0,05}$ vaut à peu près 2 .
Trouver une fourchette pour p avec un niveau de confiance d'au moins 0,95 . On donne $\dfrac{2}{45\sqrt{3}} \simeq 0,026$.

Exercice 12

Soit $n \in \mathbb{N}^*$. On considère un n-échantillon $(X_1, X_2, \ldots, X_n)$ de la loi exponentielle $\mathcal{E}(\lambda)$ de paramètre $\lambda > 0$. On pose $L_n = \min(X_1, \ldots, X_n)$ et $M_n = \max(X_1, \ldots, X_n)$.

1. Déterminer les fonctions de répartition des variables L_n puis M_n.
2. Quelle est la loi de la variable $Y_n = n\lambda L_n$?
3. On maintenant alors $Z_n = \lambda M_n - \ln(n)$.
   (a) Déterminer la fonction de répartition F_n de Z_n.
   (b) Pour $t \in \mathbb{R}$, déterminer la limite $F(t)$ de $F_n(t)$ lorsque $n \to +\infty$.
   (c) En déduire soigneusement que Z_n converge en loi vers une variable aléatoire à densité, que l'on notera Z. (On dit que Z suit la loi de Gumbel.)

   La durée de vie d'une ampoule est modélisée par une variable aléatoire $X \hookrightarrow \mathcal{E}(\lambda)$ où $\lambda > 0$ est inconnu. On cherche à estimer la durée de vide moyenne $\mu = E(X) = 1/\lambda$ et on dispose d'un échantillon de n ampoules (dont les durées de vie sont supposées indépendantes).
4. Dans cette question, on suppose que la seule information dont on dispose est la durée de vie de l'ampoule qui a grillé le plus tôt.
   (a) À l'aide de la Question (2), proposer un estimateur $\tilde{L}_n$ de μ, construit à partir de L_n, qui soit sans biais.
   (b) Quel est son risque quadratique ?
   (c) Montrer que, pour tout $\varepsilon > 0$

$$P\left(\left|\tilde{L}_n - \mu\right| > \varepsilon\right) = 1 - e^{-1+\lambda\varepsilon} + e^{-1-\lambda\varepsilon}$$

   (d) L'estimateur $\tilde{L}_n$ est-il convergent ?
   (e) Soit $\alpha \in]0\,;1[$. Montrer que si $Y \hookrightarrow \mathcal{E}(1)$, alors

$$P\left(Y < -\ln\left(1-\frac{\alpha}{2}\right)\right) = P\left(Y > -\ln\left(\frac{\alpha}{2}\right)\right) = \frac{\alpha}{2}$$

   (f) Montrer alors que l'intervalle

$$I_{\alpha,n} = \left[\frac{\tilde{L}_n}{-\ln(\alpha/2)}, \frac{\tilde{L}_n}{-\ln(1-\alpha/2)}\right]$$

   est un intervalle de confiance au seuil $1-\alpha$ pour μ.
   Remarque. On aurait pu utiliser la convergence en loi de Z_n pour obtenir un intervalle de confiance asymptotique pour μ (après avoir fait calculer les quantiles de la loi de Gumbel)

$$J_{\alpha,n} = \left[\frac{M_n}{\ln(n) - \ln(-\ln(1-\alpha/2))}; \frac{M_n}{\ln(n) - \ln(-\ln(\alpha/2))}\right].$$