TD 17

Estimation

Estimation ponctuelle

Exercice 1 (★)

Soit $a \in \mathbb{R}_+^*$. On considère la fonction f définie sur $\mathbb{R}$ par :

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

1. Montrer que f est une densité de probabilité.

On considère un moteur qui fonctionne sans interruption, en émettant du gaz carbonique dans l'atmosphère. Un capteur mesure en permanence le taux de gaz carbonique émis par le 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 variable de valeurs $E(T) = \dfrac{3a}{2}$ et $V(T) = \dfrac{3a^2}{4}$.
3. On met en route $n \in \mathbb{N}^*$ moteurs de modèle identique au précédent et indépendants.

   On note $T_1, \ldots, T_n$ les temps respectifs pendant lesquels ces moteurs ont un taux de gaz carbonique réglementaire ($T_1, \ldots, T_n$ suivent donc la même loi que T et sont indépendantes).

   (a) Justifier que la variable $Z_n = \dfrac{2}{3n} \displaystyle\sum_{k=1}^{n} T_k$ est un estimateur de a.

   (b) Calculer l'espérance et la variance de Z_n.

   (c) Montrer que, pour tout $\varepsilon > 0$, $\displaystyle\lim_{n \to +\infty} P(|Z_n - a| \geq \varepsilon) = 0$.
   Comment interpréter ce résultat ?

Exercice 2 (★★)

On considère la fonction f définie par :

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

On cherche à déterminer l'inconnu $R > 0$.

1. (a) Montrer que f est une densité de probabilité.

   (b) On considère X est une variable aléatoire de densité f.
   Déterminer la fonction de répartition F de X.

   (c) Calculer l'espérance et la variance de X.

Dans la suite, on considère un n-échantillon $(X_1, X_2, \ldots, X_n)$ de la loi de X.

2. On note $T_n = \dfrac{3}{2n} \displaystyle\sum_{i=1}^{n} X_i$.

   (a) Justifier que T_n est un estimateur de R.

(b) Calculer l'espérance et la variance de T_n.

3. On note $M_n = \max(X_1, X_2, \ldots, X_n)$.

(a) Montrer que pour tout x réel, $P(M_n \leq x) = (F(x))^n$.

(b) En déduire que M_n est une variable aléatoire à densité et déterminer une densité g_n de M_n.

(c) Montrer que $E(M_n) = \dfrac{2n}{2n+1}R$ et $V(M_n) = \dfrac{n}{(n+1)(2n+1)^2}R^2$.

(d) Donner un estimateur W_n de R, proportionnel à M_n, c'est-à-dire de la forme $\lambda_n M_n$ où λ_n est un réel non aléatoire, tel que $E(W_n) = R$.

(e) Comparer la variance de W_n et celle de T_n. Que peut-on en déduire ?

Exercice 3 (★★)

1. On considère un n-échantillon $(X_1, \ldots, X_n)$ d'une loi de Poisson de paramètre $\theta > 0$ inconnu que l'on cherche à estimer, ainsi que $(x_1, \ldots, x_n)$ un n-uplet de $\mathbb{N}^n$ fixé.

(a) On considère la fonction de vraisemblance L_n définie sur $\mathbb{R}_+^*$ par :

$$L_n(\theta) = \prod_{i=1}^{n} P_\theta(X_i = x_i).$$

Exprimer L_n en fonction de $s_n = \displaystyle\sum_{i=1}^{n} x_i$ et $p_n = \displaystyle\prod_{i=1}^{n} (x_i!)$.

(b) Calculer $L_n'(\theta)$ pour tout $\theta > 0$ et vérifier que L_n est maximale en $\theta^* = \dfrac{s_n}{n}$.

(c) En déduire l'expression de l'estimateur de maximum de vraisemblance pour la loi de Poisson.

2. En utilisant la même méthode que précédemment, montrer que l'estimateur du maximum de vraisemblance d'une loi géométrique est donné pour un n-échantillon $(X_1, \ldots, X_n)$ par la formule :

$$\hat{\theta}_n = \frac{n}{X_1 + \ldots + X_n}.$$

Exercice 4 (★★)

Dans cet exercice, θ désigne un réel strictement positif et n un entier naturel supérieur ou égal à 2.

1. Pour tout $k \in \mathbb{N}$, on pose : $u_k = \dfrac{1}{1+\theta}\left(\dfrac{\theta}{1+\theta}\right)^k$.

   Montrer que la suite $(u_k)_{k\in\mathbb{N}}$ définit une loi de probabilité.

2. On considère une variable aléatoire X prenant ses valeurs dans $\mathbb{N}$ et dont la loi est donnée par : $\forall k \in \mathbb{N},\ P(X = k) = u_k$.

   On pose $Y = X + 1$. Reconnaitre la loi de Y, puis en déduire l'espérance et la variance de X.

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 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 $x_1, x_2, \ldots, x_k$ 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 $\widehat{\theta_n}$ et que l'on exprimera en fonction de S_n. Que représente $\widehat{\theta_n}$ pour la fonction L ?

4. On appelle estimateur du maximum de vraisemblance pour θ la variable $T_n = \frac{1}{n}\sum_{i=1}^{n} X_i$.

   (a) Justifier que T_n est un estimateur de θ.

   (b) Calculer l'espérance et la variance de T_n.

   (c) Montrer que, pour tout $\varepsilon > 0$, $\lim_{n\to+\infty} P(|T_n - \theta| \geq \varepsilon) = 0$.
   Comment interpréter ce résultat ?

Exercice 5 (★★)

Soit $a \in \mathbb{R}_+^*$.

1. Montrer que, pour tout $n \in \mathbb{N}$, l'intégrale $I_n = \int_0^{+\infty} x^n e^{-\frac{x^2}{2a^2}}\,dx$ est convergente.

2. (a) Rappeler une densité d'une variable aléatoire qui suit la loi normale d'espérance nulle et de variance a^2.
   En déduire que $I_0 = a\sqrt{\frac{\pi}{2}}$.

   (b) Calculer la dérivée de l'application $\varphi : \mathbb{R} \to \mathbb{R}$ définie, pour tout $x \in \mathbb{R}$, par : $\varphi(x) = e^{-\frac{x^2}{2a^2}}$.
   En déduire que $I_1 = a^2$.

3. (a) Montrer que, pour tout entier $n \geq 2$ et pour tout $t \in [0, +\infty[$:

$$\int_0^t x^n e^{-\frac{x^2}{2a^2}}\,dx = -a^2 t^{n-1} e^{-\frac{t^2}{2a^2}} + (n-1)a^2 \int_0^t x^{n-2} e^{-\frac{x^2}{2a^2}}\,dx.$$

   (b) En déduire, pour tout entier $n \geq 2$, que : $I_n = (n-1)a^2 I_{n-2}$.

   (c) Calculer I_2 et I_3.

4. On considère l'application $g_a : \mathbb{R} \to \mathbb{R}$ définie, pour tout $x \in \mathbb{R}$, par :

$$g_a(x) = \begin{cases} 0 & \text{si } x \leq 0, \\ \dfrac{x}{a^2} e^{-\frac{x^2}{2a^2}} & \text{si } x > 0. \end{cases}$$

   Montrer que g_a est une densité.

5. On considère une variable aléatoire X admettant g_a comme densité.
   Déterminer la fonction de répartition de la variable aléatoire X.

6. Montrer que la variable aléatoire X admet une espérance $E(X)$ et que $E(X) = a\sqrt{\frac{\pi}{2}}$.

7. Montrer que la variable aléatoire X admet une variance $V(X)$ et calculer $V(X)$.

8. (a) On considère une variable aléatoire U suivant la loi uniforme sur l'intervalle $]0,1]$.
Montrer que la variable aléatoire $Z = a\sqrt{-2\ln(U)}$ suit la même loi que X.
   (b) En déduire un programme en langage `Python` simulant la variable aléatoire X, le réel a strictement positif étant entré par l'utilisateur.

9. On considère $n \geq 2$ variables aléatoires indépendantes $X_1, X_2, \ldots, X_n$ suivant toutes la même loi que la variable aléatoire X.

   On considère également la variable aléatoire $A_n = \dfrac{\sqrt{2}}{n\sqrt{\pi}}(X_1 + X_2 + \ldots + X_n)$.

   (a) Justifier que la variable aléatoire A_n est un estimateur de a.
   (b) Déterminer l'espérance et la variance de l'estimateur A_n.

10. On définit la variable aléatoire $M_n = \min(X_1, X_2, \ldots, X_n)$.
   (a) Calculer la fonction de répartition de M_n.
   (b) Montrer que M_n est une variable à densité, admettant g_b comme densité avec $b = \dfrac{a}{\sqrt{n}}$.
   (c) Montrer que la variable aléatoire M_n admet une espérance $E(M_n)$ et une variance $V(M_n)$. Calculer $E(M_n)$ et $V(M_n)$.
   (d) En déduire un estimateur B_n de la forme $\lambda_n M_n$, avec $\lambda_n \in \mathbb{R}$, tel que $E(B_n) = a$.
   (e) Déterminer la variance de l'estimateur B_n.
   (f) Comparer les estimateurs A_n et B_n.

Intervalles de confiance

Exercice 6 (★★)

Soit un dé cubique truqué dont la probabilité d'obtenir la face 1 est p à déterminer. On lance n fois le dé et on note X_i la variable aléatoire valant 1 si on obtient la face 1 au i-ième lancer et 0 sinon. On décide d'estimer le paramètre p par la fréquence F_n d'apparition de la face 1 au cours des n lancers.

1. Donner l'expression de F_n en fonction des variables X_i.

2. Montrer que F_n a une espérance et une variance et les calculer.

3. Montrer que pour tous $p \in [0,1]$, $p(1-p) \leq \dfrac{1}{4}$.

4. Énoncer l'inégalité de Bienaymé-Tchebychev pour la variable F_n puis en déduire que
$$P(F_n - \varepsilon \leq p \leq F_n + \varepsilon) \geq 1 - \frac{1}{4n\varepsilon^2}.$$

5. Donner une valeur de ε telle que l'intervalle $[F_n - \varepsilon; F_n + \varepsilon]$ est un intervalle de confiance de p au niveau de confiance $1 - \alpha$.

6. (a) On effectue 1000 lancers.
Donner alors un intervalle de confiance de p au niveau de confiance de 90%.
   (b) Déterminer le nombre minimal n de lancers à effectuer pour que $[F_n - 0,01; F_n + 0,01]$ soit un intervalle de confiance de p au niveau de confiance de 99%.

Exercice 7 (★★)

10 000 000 électeurs sont inscrits sur les listes électorales d'un pays. Une élection oppose deux candidats A et B. On cherche à estimer la proportion p d'électeurs du pays favorables au candidat A.
On considère un échantillon quelconque de n électeurs choisis au hasard (avec remise). Pour tout $i \in [\![1;n]\!]$, on note X_i la variable aléatoire égale à 1 si le i-ième électeur interrogé vote pour le candidat A, et à 0 sinon.
On note N_n le nombre d'électeurs favorables à A dans l'échantillon et F_n la proportion d'électeurs favorables à A dans l'échantillon et F_n^* la variable aléatoire centrée réduite associée à F_n. On choisit F_n comme estimateur de p.

1. Exprimer F_n en fonction de N_n puis en fonction des variables $X_1, X_2, \dots, X_n$.
2. Déterminer un réel t tel que :
$$\lim_{n\to+\infty} P(-t \leq F_n^* \leq t) = 0,99.$$
3. Montrer que :
$$\lim_{n\to+\infty} P\left(p - t\frac{\sqrt{p(1-p)}}{\sqrt{n}} \leq F_n \leq p + t\frac{\sqrt{p(1-p)}}{\sqrt{n}}\right) = 0,99.$$
4. Montrer que, pour tout $p \in [0;1]$, $0 \leq p(1-p) \leq \frac{1}{4}$.
5. En déduire que :
$$\lim_{n\to+\infty} P\left(p - t\frac{1}{2\sqrt{n}} \leq F_n \leq p + t\frac{1}{2\sqrt{n}}\right) \geq 0,99.$$
6. En déduire un intervalle de confiance asymptotique de p et donner le niveau de confiance associé.
7. Dans un échantillon observé de 100 électeurs, 64 sont déclarées favorables à A.

   Donner l'intervalle de confiance de p au niveau de confiance asymptotique $0,99$.

   A est-il certain d'être élu ?

Exercice 8 (★★)

Soient $X_1, \dots, X_n$ des variables aléatoires i.i.d. suivant la loi normale d'espérance θ inconnue et de variance σ^2 connue. On pose $\overline{X_n} = \frac{1}{n}\sum_{i=1}^{n} X_i$.

1. Quelle est la loi de $\overline{X_n}$?
2. Soit $\alpha \in]0,1[$ et t_α l'unique réel tel que $\Phi(t_\alpha) = 1 - \frac{\alpha}{2}$. Montrer que
$$\left[\overline{X_n} - \frac{t_\alpha \sigma}{\sqrt{n}}, \overline{X_n} + \frac{t_\alpha \sigma}{\sqrt{n}}\right]$$
   est un intervalle de confiance de θ au niveau de confiance $1 - \alpha$.

Exercice 9 (★★)

Soient $X_1, \dots, X_n$ un échantillon de la loi exponentielle de paramètre $\lambda > 0$ inconnu.
On pose alors $M_n = \min(X_1, \dots, X_n)$ et on considère $\alpha \in]0,1[$.

1. Montrer que M_n suit une loi exponentielle dont on précisera le paramètre.
2. Déterminer deux réels a_n et b_n tels que $P(n\lambda M_n \leq a_n) = \dfrac{\alpha}{2} = P(n\lambda M_n \geq b_n)$.
3. En déduire que $\left[\dfrac{nM_n}{b_n}, \dfrac{nM_n}{a_n}\right]$ est un intervalle de confiance de $\dfrac{1}{\lambda}$ au niveau de confiance $1-\alpha$.

Exercice 10 (★★)

Soit $a \in]0,1[$ et f_a la fonction définie sur $\mathbb{R}$ par :

$$f_a(x) = \begin{cases} \dfrac{1}{2a} & \text{si } x \in]0,a[, \\ \dfrac{1}{2(1-a)} & \text{si } x \in [a,1], \\ 0 & \text{sinon.} \end{cases}$$

1. Montrer que f_a est une densité.
2. On considère une variable aléatoire X admettant f_a comme densité.
   (a) Montrer que X admet une espérance et que $E(X) = \dfrac{2a+1}{4}$.
   (b) Montrer que X admet une variance et que $V(X) = \dfrac{4a^2-4a+5}{48}$.

On cherche à estimer le paramètre inconnu a. On considère un n-échantillon $(X_1, X_2, \dots, X_n)$ de variables aléatoires indépendantes de même loi que X.

3. On pose $S_n = \dfrac{1}{n}\displaystyle\sum_{k=1}^{n} X_k$.
   (a) Déterminer l'espérance et la variance de S_n.
   (b) En déduire un estimateur T_n de a tel que $E(T_n) = a$.
   (c) Donner une majoration de la variance de T_n quand $a \in]0,1[$.
   (d) Pour n et α donnés, préciser en utilisant l'inégalité de Bienaymé-Tchebychev, les valeurs de ε pour lesquelles $[T_n - \varepsilon, T_n + \varepsilon]$ est un intervalle de confiance au risque α pour a.
4. (a) Quelle est la limite en loi de la suite $\left(\dfrac{T_n - a}{\sqrt{V(T_n)}}\right)_{n\in\mathbb{N}^*}$?
   (b) On suppose que n est suffisamment grand pour identifier la loi de $\dfrac{T_n - a}{\sqrt{V(T_n)}}$ à cette loi limite.
   En déduire un intervalle de confiance pour a à un risque inférieur à 5%.
   On donne $\Phi^{-1}(0.975) = 1.96$.
   (c) Au niveau de risque 5%, comparer les longueurs des deux intervalles de confiance trouvés aux questions 3.(d) et 4.(b).
   On donne $\sqrt{5} \simeq 2.23$.
