Exercices du chapitre 14

Estimations ponctuelle et par intervalle de confiance

●●○○ Exercice 1 - Estimation de l'ensemble image d'une loi uniforme

On considère une variable aléatoire X suivant la loi uniforme sur $[a;b]$, où $a,b\in\mathbb{R}$ avec $a<b$. On considère également une suite de variables aléatoires $(X_k)_{k\in\mathbb{N}^*}$, indépendantes, de même loi que X.
Pour tout $n\in\mathbb{N}^*$, on pose $Y_n=\min(X_1,\dots,X_n)$ et $Z_n=\max(X_1,\dots,X_n)$.

1. Justifier que Y_n est un estimateur de a et Z_n un estimateur de b.
2. 2.a. Démontrer que Y_n est une variable aléatoire à densité et en déterminer une densité.
   2.b. Calculer $\mathbb{E}(Y_n)$ ainsi que $\lim\limits_{n\to+\infty}\mathbb{E}(Y_n)$.
   2.c. Calculer $\mathbb{V}(Y_n)$.
   2.d. Etudier la convergence en loi de $(Y_n)_{n\in\mathbb{N}^*}$.
3. 3.a. Démontrer que Z_n est une variable aléatoire à densité et en déterminer une densité.
   3.b. Calculer $\mathbb{E}(Z_n)$ ainsi que $\lim\limits_{n\to+\infty}\mathbb{E}(Z_n)$.
   3.c. Calculer $\mathbb{V}(Z_n)$.
   3.d. Etudier la convergence en loi de $(Z_n)_{n\in\mathbb{N}^*}$.

●●○○ Exercice 2 - Maximum de vraisemblance

1. Déterminer l'estimateur du maximum de vraisemblance pour le paramètre λ d'une loi de Poisson.
2. On considère une variable aléatoire X suivant la loi uniforme sur $[0;\theta]$. Déterminer l'estimateur du maximum de vraisemblance pour θ.

●●○○ Exercice 3 - Intervalles de confiance pour l'espérance d'une $\mathcal{N}(m,1)$

On considère une variable aléatoire X suivant une loi $\mathcal{N}(m,1)$, où m est inconnu. On dispose d'un n-échantillon $(X_1,\dots,X_n)$ de X.

1. Donner la loi de $\overline{X_n}$ et celle de $\overline{X_n}^*=\sqrt{n}(\overline{X_n}-m)$.
2. Soit $\alpha\in]0;1[$. Justifier l'existence d'un unique réel t_α tel que $\Phi(t_\alpha)=1-\dfrac{\alpha}{2}$, où Φ désigne la fonction de répartition d'une variable aléatoire suivant la loi $\mathcal{N}(0;1)$.
3. En déduire que $\left[\overline{X_n}-t_\alpha\dfrac{1}{\sqrt{n}},\overline{X_n}+t_\alpha\dfrac{1}{\sqrt{n}}\right]$ est un intervalle de confiance de m au niveau de confiance $1-\alpha$.
4. A l'aide de l'inégalité de Bienaymé-Tchebychev, démontrer que $\left[\overline{X_n}-\dfrac{1}{\sqrt{n\alpha}},\overline{X_n}+\dfrac{1}{\sqrt{n\alpha}}\right]$ est un intervalle de confiance de m au niveau de confiance $1-\alpha$.
5. **Comparaison des deux intervalles.** On considère $n=100$ et $\alpha=0,05$. A l'aide d'une table de la loi $\mathcal{N}(0;1)$, déterminer une valeur approchée de t_α.
   Comparer alors l'amplitude des deux intervalles de confiance obtenus dans les questions précédentes.
6. Écrire un programme `Python` qui simule 1000 réalisations d'un échantillon de taille 100, avec le même paramètre m, et qui compte le nombre de réalisations fournissant un intervalle de confiance contenant la valeur m choisie.

●●○○ Exercice 4 - Prise de décision

Une entreprise souhaite acquérir une machine qui fabrique un certain type d'objets et qui, en fonctionnement normal, produit une proportion p $(0<p<1)$ d'objets défectueux. Le directeur veut connaître la valeur de p. Pour cela, il teste la machine et prélève un échantillon de n objets qu'il analyse, avec $n\geq 1$.
Pour tout $i\in[\![1,n]\!]$, soit X_i la variable aléatoire de Bernoulli définie par

$$X_i=\begin{cases}1 & \text{si le } i\text{-ème objet prélevé est défectueux}\\ 0 & \text{sinon}\end{cases}$$

On suppose que dans les conditions de prélèvement, les variables aléatoires $X_1,\dots,X_n$ sont indépendantes.

On pose $S_n=\displaystyle\sum_{k=1}^{n}X_k$.

1. 1.a. Montrer que $F_n=\dfrac{S_n}{n}$ est un estimateur de p d'espérance p.
   1.b. Calculer la variance v_n de F_n. Déterminer $\lim\limits_{n\to+\infty}v_n$.
2. Soit α un réel de $]0,1[$. On souhaite déterminer dans cette question un intervalle de confiance du paramètre p inconnu, au niveau de confiance $1-\alpha$, à partir de l'échantillon $(X_1,\dots,X_n)$.
   2.a. Quelle est la limite en loi de la suite $\left(\sqrt{n}\dfrac{F_n-p}{\sqrt{p(1-p)}}\right)_{n\in\mathbb{N}^*}$?

2.b. Justifier l'existence d'un unique réel t_α tel que $\Phi(t_\alpha) = 1 - \dfrac{\alpha}{2}$, où Φ désigne la fonction de répartition d'une variable aléatoire suivant la loi normale centrée réduite.
Montrer qu'un intervalle de confiance de p au niveau $1-\alpha$ est donné par $\left[F_n - \dfrac{t_\alpha}{2\sqrt{n}}, F_n + \dfrac{t_\alpha}{2\sqrt{n}}\right]$.

2.c. On suppose dans cette question qu'en fonctionnement normal la machine produit une proportion $p = 0,05$ d'objets défectueux. Le directeur analyse 10000 objets et compte 600 objets défectueux sur cet échantillon. Décide-t-il d'acheter la machine, avec un risque égal à $0,05$?
On donne $\Phi(2) \simeq 0,975$.

●●○○ Exercice 5

Soit $(X_k)_{k\in\mathbb{N}^*}$ une suite de variables aléatoires indépendantes suivant la loi $\mathcal{U}([0,\theta])$, où θ est inconnu. On pose, pour tout $n \in \mathbb{N}^*$, $M_n = \max(X_1, \dots, X_n)$.

1. 1.a. Démontrer que pour $\theta > 0$ fixé, il existe deux réels x_1 et x_2, que l'on exprimera en fonction de θ, tels que
$$F_{M_n}(x_1) = 0,025 \quad ; \quad F_{M_n}(x_2) = 0,975$$
   1.b. En déduire un intervalle de confiance de θ au niveau de confiance $0,95$.
2. 2.a. Démontrer que
$$n\left(1 - \frac{M_n}{\theta}\right) \xrightarrow[n\to+\infty]{\mathscr{L}} X$$
   où X suit la loi exponentielle $\mathscr{E}(1)$.
   2.b. Déterminer un intervalle de confiance asymptotique de θ au niveau de risque α de la forme $[M_n, U_n]$, où U_n est un estimateur de θ.

●●○○ Exercice 6 - EDHEC 2020 E

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

1. Démontrer : $\forall x \in \mathbb{R}$, $F_X(-x) = 1 - F_X(x)$.
2. On pose $Y = |X|$ et on admet que Y est une variable aléatoire.
   2.a. Montrer que la fonction de répartition de Y est la fonction, notée F_Y, définie par :
$$F_Y : x \longmapsto \begin{cases} 2F_X(x) - 1 & \text{si } x \geq 0 \\ 0 & \text{si } x < 0 \end{cases}$$
   2.b. En déduire que Y est une variable à densité et donner une densité f_Y de Y.
   2.c. Montrer que Y possède une espérance et que l'on a : $\mathbb{E}(Y) = \sigma\sqrt{\dfrac{2}{\pi}}$.
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 $(Y_1, Y_2, \dots, Y_n)$ composé de variables aléatoires, mutuellement indépendantes, et ayant toutes la même loi que Y.
   On note S_n la variable aléatoire définie par : $S_n = \dfrac{1}{n}\displaystyle\sum_{k=1}^{n} Y_i$.
   3.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.
   3.b. Rappeler la valeur du moment d'ordre 2 de X, puis déterminer $\mathbb{E}(Y^2)$, $\mathbb{V}(Y)$ et $\mathbb{V}(S_n)$.
   3.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.
4. Écrire une fonction **Python** prenant en argument d'entrée un réel strictement positif σ et renvoyant une réalisation des variables aléatoires S_n et T_n.

●●○○ Exercice 7 - EDHEC 2016 S

Soit X une variable aléatoire suivant la loi normale de paramètres m et σ^2 (avec $\sigma > 0$). On suppose que l'on ne connaît pas les paramètres $\theta_1 = m$ et $\theta_2 = \sigma^2$ et on souhaite les estimer par une méthode appelée méthode du maximum de vraisemblance.
Pour ce faire, on considère un n-échantillon $(X_1, \dots, X_n)$ de la loi de X, avec $n \geq 2$.
On appelle vraisemblance du couple (θ_1, θ_2), la fonction notée L définie par :
$$L(\theta_1, \theta_2) = \prod_{i=1}^{n} \varphi_{\theta_1,\theta_2}(x_i), \text{ où } x_1, \dots, x_n \text{ sont des nombres réels donnés}$$

1. Donner l'expression de $L(\theta_1, \theta_2)$, puis celle de $\ln\big(L(\theta_1, \theta_2)\big)$ en fonction de θ_1, θ_2 et $x_1, \dots, x_n$.
2. 2.a. Justifier que la fonction $f : (\theta_1, \theta_2) \mapsto \ln\big(L(\theta_1, \theta_2)\big)$, définie sur l'ouvert $U = \mathbb{R} \times \mathbb{R}_*^+$, est de classe $\mathscr{C}^2$ sur U.
   2.b. Montrer que f admet un seul point critique $A = \left(\widehat{\theta}_1, \widehat{\theta}_2\right)$ sur U tel que :
$$\widehat{\theta}_1 = \frac{1}{n}\sum_{i=1}^{n} x_i \text{ et } \widehat{\theta}_2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \widehat{\theta}_1^{\,2}$$
   2.c. Déterminer les valeurs des dérivées partielles d'ordre 2 de f en A.
   On vérifiera en particulier que : $\partial^2_{2,2}(f)\left(\widehat{\theta}_1, \widehat{\theta}_2\right) = \dfrac{-n}{2\widehat{\theta}_2^{\,2}}$.
   2.d. En déduire que f admet un maximum local en $(\widehat{\theta}_1, \widehat{\theta}_2)$.
   2.e. Expliquer pourquoi la fonction L admet aussi un maximum local en $(\widehat{\theta}_1, \widehat{\theta}_2)$.