Exercices - Chapitre 12 - Estimation et intervalles de confiance

Exercice 1
Soit X_1, ..., X_n des variables aléatoires i.i.d. telles que :

$$P(X_1=-1)=(1-p)^2,\quad P(X_1=0)=2p(1-p),\quad P(X_1=1)=p^2$$

On cherche à estimer $p\in]0,1[$. Montrer que

$$Z_n=\sum_{k=1}^{n}\frac{1+X_k}{2n}$$

est un estimateur sans biais et convergent de p.

Exercice 2
Soient λ un réel strictement positif et n un entier naturel supérieur ou égal à 3.

On considère un n-échantillon $(X_1,\cdots,X_n)$ indépendant identiquement distribué de la loi exponentielle de paramètre λ que l'on cherche à estimer.

On note $S_n=X_1+\cdots+X_n$ et $Y_n=\frac{1}{S_n}$

1. Déterminer un estimateur sans biais et convergent de $\frac{1}{\lambda}$.
2. Déterminer la loi de $T_n=\lambda S_n$.
3. Justifier que S_n est une variable à densité puis déterminer une densité de S_n.
4. Montrer que Y_n admet une espérance et une variance et les calculer.
5. En déduire un estimateur sans biais du paramètre λ.
6. Cet estimateur est-il convergent?

Exercice 3
Recherche d'un intervalle de confiance

On désigne par α un réel donné de $]0,1[$.
Soit $(X_n)_{n\in\mathbb{N}}$ une suite de variables aléatoires indépendantes, suivant toutes la loi normale de paramètres m et 1.

On note, pour tout entier naturel non nul, $\overline{X}_n=\dfrac{1}{n}\sum_{k=1}^{n}X_k$.

1. On note $Y_n=\sqrt{n}\left(\overline{X}_n-m\right)$. Quelle est la loi de Y_n ?
2. (a) Justifier l'existence d'un réel t_α strictement positif tel que $\Phi(t_\alpha)=1-\dfrac{\alpha}{2}$.
   (b) Montrer que $P\left(-t_\alpha\leq Y_n\leq t_\alpha\right)=1-\alpha$.
   (c) En déduire un intervalle de confiance de m au niveau de confiance $1-\alpha$.

Exercice 4
Soit $(X_1,\cdots,X_n)$ un échantillon i.i.d. issu de la variable aléatoire X qui admet une espérance μ et une variance σ^2.
On note $T_n=\frac{1}{n}\sum\limits_{i=1}^{n}(X_i-\overline{X_n})^2$. et $Z_n=\frac{1}{n-1}\sum\limits_{i=1}^{n}(X_i-\overline{X_n})^2$.

1. Montrer que T_n admet une espérance et $E(T_n)=\dfrac{n-1}{n}\sigma^2$.
2. En déduire que $\lim_{n\to+\infty}E(T_n)=\sigma^2$.
   On dit alors que T_n est un estimateur asymptotiquement sans biais de σ^2 - cette notion est HP
3. Montrer que Z_n est un estimateur sans biais de σ^2.

Exercice 5
Soit θ un paramètre réel strictement positif.

Soit $(\Omega,\mathcal{A},P)$ un espace probabilisé et X une variable aléatoire ayant pour densité la fonction f définie par :

$$\forall t\in\mathbb{R}\quad f(t)=\frac{2t}{\theta^2}.\mathbb{1}_{[0,\theta]}(t)$$

On considère $(X_i)_{i\in\mathbb{N}^*}$ une suite de variables aléatoires i.i.d. de même loi que X.

Pour tout $n\in\mathbb{N}^*$, on note $\overline{X_n}=\frac{1}{n}\sum\limits_{i=1}^{n}X_i$.

1. (a) Vérifier que f est une densité de probabilité. On notera X une variable aléatoire à densité de densité f.
   (b) Déterminer la fonction de répartition de la variable X.
   (c) Montrer que X admet une espérance et une variance et les calculer.
2. (a) Calculer $E(\overline{X_n})$ et $V(\overline{X_n})$.
   (b) Construire à partir de $\overline{X}_n$ un estimateur sans biais convergent du paramètre θ. On le notera T_n.
3. Appliquer le Théorème Central Limite à la variable $\overline{X_n}$.
   En déduire un intervalle de confiance asymptotique de θ, au niveau de confiance 95%.
   Si on note Φ la fonction de répartition d'une variable aléatoire qui suit une loi $\mathcal{N}(0,1)$, on donne, pour $t=1,96$:
   $2\Phi(t)-1=0,95$
4. Soit G_n la variable aléatoire définie par : $G_n=\sup(X_1,\ldots,X_n)$.
   (a) Déterminer une densité de la variable G_n.
   (b) Calculer l'espérance et la variance de G_n, puis montrer que $\lim_{n\to+\infty}E(G_n)=\theta$.
   (c) Construire, à l'aide de G_n un estimateur sans biais et convergent de θ noté W_n.
5. <u>Simulation</u> On suppose dans cette question que $\theta=1$ pour simplifier un peu.
   (a) A l'aide de la méthode d'inversion, écrire une ligne de code permettant de simuler la variable X.
   (b) Ecrire un programme qui simule $N=1000$ fois les deux estimateurs T_n et W_n, avec $n=100$. Puis tracer les deux histogrammes des valeurs prises par ces estimateurs (on pourra s'inspirer de l'exercice du cours !).
   (c) Quel semble être le meilleur de ces deux estimateurs ?
6. Démontrer par le calcul la conjecture précédente.

Exercice 6
On dispose d'un n-échantillon $(X_1,X_2,\cdots,X_n)$ de variables aléatoires indépendantes suivant la loi de Poisson de paramètre λ, avec λ inconnu.
On cherche à estimer la probabilité $e^{-q\lambda}$ de n'observer que des 0 au cours de q expériences consécutives.
On pose $T_n=e^{-q\overline{X_n}}$.

1. T_n est-il un estimateur sans biais de $e^{-q\lambda}$?
2. Montrer que T_n est un estimateur convergent de $e^{-q\lambda}$.

Exercice 7
Estimation par capture-recapture

On cherche à évaluer le nombre N de poissons présents dans un étang.
Pour cela, on prélève un échantillon de m poissons, que l'on marque puis que l'on rejette dans l'étang.
Soit $n \in \mathbb{N}^*$, où $n \leq m$. On prélève successivement des poissons au hasard et avec remise dans l'étang. On note X_n la v.a.r. égale au nombre de poissons qu'il a été nécessaire de pêcher avant d'obtenir n poissons marqués. On note $D_i = X_i - X_{i-1}$ où $i > 1$ et $D_1 = X_1$. On admet que les v.a.r. D_i sont indépendantes. On note $Z_k = m.D_k$.

1. (a) Quelle est la loi de D_i, où $i \geq 1$? *à justifier rapidement !!!*
   Donner son espérance, sa variance.
   (b) En déduire l'espérance et la variance de la v.a.r. X_n.
2. Montrer que $\overline{Z_n} = \frac{1}{n}(\sum_{k=1}^{n} Z_k)$ où $n \in \mathbb{N}^*$ est un estimateur sans biais convergent de N.
3. (a) Justifier que la suite $\left(\frac{\sqrt{N(N-m)}}{\sqrt{\overline{Z_n}(\overline{Z_n}-m)}}\right)$ converge en probabilité vers 1.
   (b) On admet le lemme de Slutsky : si (X_n) et (Y_n) sont deux suites de variables aléatoires telles que $(X_n)_{n\in\mathbb{N}}$ converge **en loi** vers X, et $(Y_n)_{n\in\mathbb{N}}$ converge **en probabilité** vers une variable aléatoire certaine égale à c, alors la suite $(X_nY_n)_{n\in\mathbb{N}}$ converge en loi vers la variable aléatoire cX.

   Prouver alors que la suite $\left(\sqrt{n} \times \frac{\overline{Z_n}-N}{\sqrt{\overline{Z_n}(\overline{Z_n}-m)}}\right)$ converge en loi vers une variable V suivant la loi $\mathcal{N}(0,1)$.
4. Rappel : si $\alpha \in]0,1[$, il existe un unique réel $t_\alpha > 0$ tel que $\Phi(t_\alpha) = 1 - \frac{\alpha}{2}$.
   Prouver que
   $$\left[\overline{Z_n} - t_\alpha\sqrt{\frac{\overline{Z_n}(\overline{Z_n}-m)}{n}};\quad \overline{Z_n} + t_\alpha\sqrt{\frac{\overline{Z_n}(\overline{Z_n}-m)}{n}}\right]$$
   est un intervalle de confiance asymptotique de N au risque α.

Exercice 8
On suppose que la probabilité qu'un individu contagieux transmette son virus à un individu sain est inconnue et on le note $p \in]0,1[$.

On cherche donc à estimer cette probabilité.
On considère la variable de Bernoulli Y qui vaut 1 si l'individu contagieux transmet son virus et 0 sinon.
Pour tout entier naturel m supérieur ou égal à 1, on considère un m-échantillon $(Y_1, Y_2, \ldots, Y_m)$ de variables aléatoires i.i.d. de même loi que Y.
On note $\overline{Y_m} = \dfrac{1}{m}\sum_{k=1}^{m} Y_k$.

1. Montrer que $\overline{Y_m}$ est un estimateur sans biais et convergent de p.
2. Montrer que $p(1-p) \in]0, \frac{1}{4}]$.
3. A l'aide de l'inégalité de Bienaymé-Tchebychev, montrer que $\left[\overline{Y_m} - \sqrt{\dfrac{5}{m}}, \overline{Y_m} + \sqrt{\dfrac{5}{m}}\right]$ est un intervalle de confiance de p au niveau de confiance 0.95

Exercice 9
Loi de Pareto et estimateur du maximum de vraisemblance (classique)

1. Pour $a > 0$, montrer que la fonction :
   $$f_a : t \mapsto \begin{cases} 0 & \text{si } t \leq 1 \\ \dfrac{a}{t^{a+1}} & \text{si } t > 1 \end{cases}$$
   est une densité de probabilité (**loi de Pareto**).
2. Soit $a > 0$ et X une variable de densité f_a. A quelle condition X admet-elle une espérance ?
   Si X admet une espérance, calculez $E(X)$.

   Dans la suite de cet exercice, on considère un n-échantillon $(X_1, \cdots, X_n)$ i.i.d. de loi à densité f_a, où $a > 0$ est un paramètre **inconnu**.

3. On définit la **fonction de vraisemblance** :
   $$L : (x_1, \cdots, x_n, a) \in \mathbb{R}^n \times \mathbb{R}_+^* \mapsto L(x_1, \cdots, x_n, a) = \prod_{k=1}^{n} f_a(x_k)$$
   (a) Expliciter, pour $(x_1, \cdots, x_n, a) \in \mathbb{R}^n \times \mathbb{R}_+^*$, la valeur de $L(x_1, \cdots, x_n, a)$.
   (b) Pour $(x_1, \cdots, x_n) \in]1, +\infty[^n$ donné, étudier les variations de la fonction $h : a \mapsto L(x_1, \cdots, x_n, a)$ et montrer qu'elle atteint un maximum en un point $\hat{a}$.

   Puisque $\hat{a}$ dépend de x_1, ..., x_n, il existe une fonction $\varphi :]1; +\infty[^n \to \mathbb{R}$, telle que $\hat{a} = \varphi(x_1, \cdots, x_n)$.
   On considère alors la variable aléatoire $T_n = \varphi(X_1, \cdots, X_n)$ appelée **estimateur du maximum de vraisemblance** de a.
4. (a) Pour $k \in [[1, n]]$, montrer que la variable aléatoire $Y_k = \ln(X_k)$ suit une loi exponentielle dont vous préciserez le paramètre.
   (b) En déduire la loi de la variable $a.(Y_1 + \cdots + Y_n)$, puis une densité de $S_n = Y_1 + \cdots + Y_n$.
   (c) Calculer $E(T_n)$ et $V(T_n)$ (on pourra réutiliser les résultats de l'exercice 2 !!!).
   (d) En déduire que T_n est un estimateur convergent de a.

Exercice 10
Oral ESCP
Soit n un entier naturel supérieur ou égal à 2.

On dispose de n variables aléatoires indépendantes, notées $X_1, X_2, \ldots, X_n$ suivant la même loi de Poisson de paramètre θ inconnu ($\theta > 0$) et définies sur le même espace probabilisé $(\Omega, \mathscr{A}, P)$.
On souhaite estimer $\exp(-\theta)$ (probabilité d'obtenir 0).
On définit, pour tout i élément de $[[1, n]]$, la variable aléatoire Y_i par :
$$Y_i(\omega) = \begin{cases} 1 & \text{si } X_i(\omega) = 0 \\ 0 & \text{si } X_i(\omega) \neq 0 \end{cases}$$
On note $\bar{Y_n} = \dfrac{1}{n}\sum_{i=1}^{n} Y_i$.

1. (a) Donner la loi de Y_i pour tout i élément de $[[1, n]]$.
   (b) En déduire que $\bar{Y_n}$ est un estimateur sans biais et convergent de $\exp(-\theta)$.
2. Pour tout $k \in [[1, n]]$ on note $S_k = \sum_{i=1}^{k} X_i$.
   (a) Rappeler la loi de S_k pour tout k élément de $[[1, n]]$.
   (b) On note, pour tout entier naturel j, $\varphi(j) = P_{(S_n=j)}(X_1 = 0)$
      i. Montrer que pour tout entier naturel j, on a $\varphi(j) = \left(1 - \dfrac{1}{n}\right)^j$
      On définit l'estimateur $W_n = \varphi(S_n) = \left(1 - \dfrac{1}{n}\right)^{S_n}$
      ii. Montrer que $\varphi(S_n)$ admet une espérance et que $\varphi(S_n)$ est un estimateur sans biais de $\exp(-\theta)$.
      iii. Calculer $V(W_n)$, puis montrer que W_n est un estimateur convergent de $e^{-\theta}$.
3. On souhaite ici comparer les performances des estimateurs $\bar{Y_n}$ et $\varphi(S_n)$
   (a) Justifier que la fonction exponentielle est convexe puis montrer que :
   $$\forall n \in \mathbb{N}^*, \quad \exp\left(\frac{\theta}{n}\right) \leq \frac{\exp(\theta)}{n} + \frac{n-1}{n}$$
   (b) Montrer l'inégalité $V(\varphi(S_n)) \leq V(\bar{Y_n})$.
   (c) Conclure