Feuille d'exercices

Convergences et approximations ; statistiques inférentielles

Exercice 1

Edhec 2006

Si Φ est la fonction de répartition de la loi normale centrée réduite, on donne les valeurs approchées suivantes : $\Phi\,(1{,}96) \cong 0{,}975$ *et* $\Phi\,(0{,}90) \cong 0{,}816$.

1\. a. Montrer que pour tout entier $k \geq 3$, l'intégrale $A_k = \int_1^{+\infty} \frac{\ln t}{t^k}\, dt$ est convergente.

b. À l'aide d'une intégration par parties, montrer que $A_k = \frac{1}{(k-1)^2}$.

2\. a. Montrer que la fonction f définie sur $\mathbb{R}$ par : $f(t) = \frac{16 \ln t}{t^5}$ si $t > 1$, $f(t) = 0$ si $t \leq 1$, définit une densité de variable aléatoire X.

b. Montrer que la variable aléatoire X admet une espérance, et la calculer.

c. Montrer que la variable aléatoire X admet une variance, et la calculer.

3\. On considère une suite (X_n) de variables aléatoires indépendantes suivant la loi définie par la densité f définie au 2° a. Pour tout entier $n \geq 1$, on définit la variable aléatoire Y_n par $Y_n = \frac{X_1 + \ldots + X_n}{n}$.

a. Quelle est la limite de la suite de variables aléatoires (Y_n) quand n tend vers $+\infty$?

b. Calculer une valeur approchée de la probabilité de l'évènement $\left(\left|Y_{68} - \frac{16}{9}\right| \leq 0{,}1\right)$.

c. Évaluer un entier n_0 tel que pour tout entier $n \geq n_0$ on ait $P\left(\left|Y_n - \frac{16}{9}\right| \leq 0{,}1\right) \geq 0{,}95$.

Exercice 2

Edhec 2012

On considère une variable aléatoire de Bernoulli X de paramètre p ($0 < p < 1$) inconnu et un n-échantillon $(X_1, \ldots, X_n)$ de la loi de X. On pose $S_n = \sum_{i=1}^{n} X_i$ et $F_n = \frac{S_n}{n}$.

On souhaite estimer p en donnant un intervalle de confiance.

1\) a) Donner la loi de S_n ainsi que son espérance et sa variance.

b) Montrer que F_n est un estimateur sans biais de p. Donner son risque quadratique.

2\) On note Φ la fonction de répartition de la loi normale centrée réduite et on note t le réel tel que $\Phi(t) = 0{,}975$.

a) Justifier que, pour n assez grand, on peut approcher la loi de $\dfrac{S_n - np}{\sqrt{np(1-p)}}$ par la loi normale centrée réduite.

b) En utilisant cette approximation, montrer que : $P(-t \leq \dfrac{S_n - np}{\sqrt{np(1-p)}} \leq t) = 2\Phi(t) - 1$.

c) En déduire : $P(F_n - \dfrac{t\sqrt{p(1-p)}}{\sqrt{n}} \leq p \leq F_n + \dfrac{t\sqrt{p(1-p)}}{\sqrt{n}}) = 0{,}95$.

3) a) Montrer que : $p(1-p) \leq \dfrac{1}{4}$.

b) On admet que $t = 1{,}96$. Établir que : $P\left(F_n - \dfrac{0{,}98}{\sqrt{n}} \leq p \leq F_n + \dfrac{0{,}98}{\sqrt{n}} \right) \geq 0{,}95$.

4) Lors d'élections, deux sondages sont effectués, chacun par un institut, et indépendamment l'un de l'autre.
Le premier sondage est effectué auprès de 625 personnes : 325 d'entre elles déclarent vouloir voter pour le président sortant.
Le deuxième sondage est effectué auprès de 400 autres personnes : 180 d'entre elles déclarent vouloir voter pour le président sortant.
On note p la proportion théorique d'électeurs souhaitant reconduire le président sortant.

a) Utiliser la première partie pour déterminer un intervalle de confiance pour p au risque 5% en utilisant le premier sondage.

b) Déterminer de même un intervalle de confiance pour p au risque 5% en utilisant le deuxième sondage.

Exercice 3

Edhec 2013

Dans cet exercice, n désigne un entier naturel supérieur ou égal à 2.
On considère une variable aléatoire réelle X dont une densité de probabilité est donnée par :

$f(x) = \begin{cases} e^{-(x-\theta)} \text{ si } x \geq \theta \\ 0 \text{ si } x < \theta \end{cases}$, θ désignant un réel strictement positif inconnu.

Soit $(X_1, \ldots, X_n)$ un n-échantillon de la loi de X. On pose :

$$Y_n = \frac{1}{n}\sum_{i=1}^{n} X_i \quad \text{et } Z_n = \operatorname{Inf}(X_1, X_2, \ldots, X_n).$$

1) a) Déterminer la fonction de répartition F de X puis reconnaître la loi de $X - \theta$. En déduire l'espérance et la variance de X.

b) Montrer que $T_{1,n} = Y_n - 1$ est un estimateur sans biais de θ.

2) a) Déterminer la fonction de répartition de Z_n puis reconnaître la loi de $Z_n - \theta$.

b) Montrer que $T_{2,n} = Z_n - \dfrac{1}{n}$ est un estimateur sans biais de θ.

3) Calculer les variances de $T_{1,n}$ et $T_{2,n}$. Que peut-on en conclure ?

4) On veut construire un estimateur T_n sans biais et convergent de θ, comme combinaison linéaire de $T_{1,n}$ et $T_{2,n}$. On souhaite, de plus, que la variance de T_n soit la plus petite possible.
On note ρ le coefficient de corrélation linéaire de $T_{1,n}$ et $T_{2,n}$, supposé constant, et on pose :

$$T_n = a_{1,n}T_{1,n} + a_{2,n}T_{2,n}$$

a) Donner la valeur de $a_{1,n} + a_{2,n}$.

b) Écrire la variance de T_n comme une fonction f de $a_{1,n}$, puis déterminer $a_{1,n}$ et $a_{2,n}$ (en fonction de n et de ρ) afin que cette fonction soit minimale.

c) Vérifier que la variance de T_n est inférieure ou égale à celles de $T_{1,n}$ et $T_{2,n}$.

Exercice 4

Edhec 2014

Dans cet exercice, n et N désignent deux entiers naturels supérieurs ou égaux à 2.

1) a) Montrer que, pour tout entier naturel k de $[\![0, N-1]\!]$, on a :

$$\frac{1}{N}\left(\frac{k}{N}\right)^n \leq \int_{\frac{k}{N}}^{\frac{k+1}{N}} t^n dt \leq \frac{1}{N}\left(\frac{k+1}{N}\right)^n$$

b) En déduire que :

$$\frac{1}{N}\sum_{k=0}^{N-1}\left(\frac{k}{N}\right)^n \leq \frac{1}{n+1}$$

2) Soit Y une variable aléatoire à valeurs dans $[\![1, N]\!]$.

a) Justifier que, pour tout k de $[\![1, N]\!]$, on a : $P(Y=k) = P(Y>k-1) - P(Y>k)$.
En déduire que :

$$E(Y) = \sum_{k=0}^{N-1} P(Y>k)$$

b) Montrer de même que :

$$E\big(Y(Y-1)\big) = 2\sum_{k=0}^{N-1} k\,P(Y>k)$$

Une urne contient des boules numérotées de 1 à N. On effectue n tirages au hasard d'une boule avec remise à chaque fois de la boule tirée et on note X_i le numéro de la boule obtenue au $i^{\text{ème}}$ tirage. On admet que les variables $X_1, \ldots, X_n$ sont mutuellement indépendantes.
On suppose N inconnu et on cherche à l'estimer grâce à deux estimateurs différents

3) On considère la variable aléatoire M_n, égale à la moyenne empirique de l'échantillon $(X_1,...,X_n)$. On a donc : $M_n = \frac{1}{n}\sum_{i=1}^{n} X_i$.

a) Calculer l'espérance de M_n et en déduire un estimateur sans biais M_n' de N.

b) Calculer le risque quadratique de M_n'. Quelle est sa limite lorsque n tend vers $+\infty$?

4) On considère maintenant la variable aléatoire Z_n définie par : $Z_n = \text{Max}(X_1,...,X_n)$.

a) On note F_n la fonction de répartition de Z_n.

Donner la valeur de $F_n(k)$ pour tout k de $Z_n(\Omega)$.

b) Utiliser les questions précédentes pour en déduire que : $E(Z_n) \geq N - \frac{N}{n+1}$. Établir alors que Z_n est un estimateur asymptotiquement sans biais de N.

c) Déterminer $E(Z_n(Z_n-1))$ et vérifier que : $\lim_{n\to+\infty} E(Z_n(Z_n-1)) = N(N-1)$.

En déduire que le risque quadratique de Z_n tend vers 0 lorsque n tend vers $+\infty$.

Exercice 5

Soit $a \in \mathbb{R}_+^*$, X une VAR définie sur un probabilisé $(\Omega, \mathcal{T}, \mathbb{P})$, suivant la loi uniforme $\mathcal{U}_{[0,a]}$, et $(X_1, ..., X_n)$ un n – échantillon *i.i.d.* de la loi de X sur $(\Omega, \mathcal{T}, \mathbb{P})$.

On cherche à construire un « bon » estimateur de a. On note $\overline{X_n}$ la moyenne empirique.

Q1_ Soit $T_n = 2\overline{X_n}$. Montrer que T_n est sans biais et déterminer son risque quadratique.

Q2_ Soit $S_n = \max(X_1, ..., X_n)$. Déterminer les fonctions de répartition de X et de S_n.

En déduire une densité de S_n, puis son biais et son risque quadratique.

Q3_ Soit $U_n = \frac{n+1}{n} S_n$. Déterminer son biais et son risque quadratique.

Q4_ Quel est le meilleur de ces trois estimateurs de a pour de grandes valeurs de n ?

Exercice 6

Soit $a \in]0,1[$, et X une VAR définie sur $(\Omega, \mathcal{T}, \mathbb{P})$, à valeurs dans $[0,1]$.

On suppose que la loi conditionnelle de X sachant l'événement $[X \leq a]$ est la loi uniforme sur $[0, a]$, et que sa loi conditionnelle sachant $[X > a]$ est la loi uniforme sur $[a, 1]$.

On suppose de plus que $\mathbb{P}\left([X \leq a]\right) = \dfrac{1}{2}$.

Q1_ Montrer que X est une VAR à densité.

Déterminer une densité de X, son espérance et sa variance.

On cherche maintenant à estimer le paramètre inconnu a.

Q2_ Soit $\left(X_1, ..., X_n\right)$ un n – échantillon *i.i.d.* de la loi de X. On pose :

$$M_n = \frac{1}{n} \sum_{k=1}^{n} X_k .$$

a_ Déterminer l'espérance $\mathbb{E}\left(M_n\right)$, et la variance $\mathbb{V}\left(M_n\right)$.

b_ En déduire un estimateur sans biais T_n de a.

c_ La suite d'estimateurs $\left(T_n\right)_{n \in \mathbb{N}^*}$ est-elle consistante ?

Q3_ Donner une majoration de $\mathbb{V}\left(T_n\right)$ quand $a \in \left]0, 1\right[$.

Pour n et α donnés, préciser, en utilisant l'inégalité de Bienaymé – Cebycev, les valeurs de ε pour lesquelles $\left[T_n - \varepsilon, T_n + \varepsilon\right]$ est un intervalle de confiance I_1 pour a au seuil de risque α.

Q4_ Quelle est la limite en loi de la suite $\left(\dfrac{T_n - a}{\sqrt{\mathbb{V}\left(T_n\right)}}\right)_{n \in \mathbb{N}^*}$?

On suppose n suffisamment grand pour identifier la loi de $\dfrac{T_n - a}{\sqrt{\mathbb{V}\left(T_n\right)}}$ à cette loi limite.

Q5_ En déduire un intervalle de confiance I_2 pour a au niveau de risque $5\ \%$.

Q6_ Au seuil de risque $5\ \%$, comparer les longueurs des intervalles de confiance I_1 et I_2.

Exercice 7

A l'issue d'un scrutin de type uninominal permettant à plusieurs centaines de milliers d'électeurs de départager deux candidats A et B d'importance comparable, on se propose, avant le dépouillement, de procéder à une estimation de la proportion p des voix obtenues par le candidat A. A cet effet, on répète n fois ($n \in \mathbb{N}^*$) l'expérience suivante : on retire au hasard un bulletin des urnes ; on note s'il est ou non en faveur de A et on le remet dans les urnes. On note X_n la VAR indiquant le nombre de suffrages favorables à A parmi les n bulletins dépouillés. Le quotient $Y_n = \dfrac{X_n}{n}$ indique donc la proportion des suffrages favorables à A parmi ces n bulletins. On pose enfin : $u_n = \mathbb{P}\left(\left|Y_n - p\right| > 0{,}01\right)$.

Soit $\varepsilon \in \left]0,1\right[$. On veut déterminer le nombre n de bulletins qu'il suffit de dépouiller pour avoir $u_n \leq \varepsilon$, *i.e.* pour connaître p à $0{,}01$ près avec un risque d'erreur inférieur à ε.

Q1_ Déterminer la loi de X_n. Calculer les espérances et les variances de X_n et Y_n.

Q2_ Montrer que : $\forall\ n \in \mathbb{N}^*,\ \mathbb{V}\left(X_n\right) \leq \dfrac{n}{4}$.

Q3_ Soit $\alpha \in \mathbb{R}_+^*$. En appliquant à Y_n l'inégalité de Bienaymé – Cebycev, donner une majoration de $\mathbb{P}\left(\left|Y_n - p\right| > \alpha\right)$. En déduire que $u_n \leq \dfrac{2500}{n}$.

Q4_ Comment suffit-il de choisir n pour que $u_n \leq \varepsilon$? Examiner le cas où $\varepsilon = 0{,}1$.

Q5_ En approchant X_n par une loi normale *(on justifiera l'approximation)*, montrer que $u_n = 2\left(1 - \Phi\left(\dfrac{0{,}01\sqrt{n}}{\sqrt{p\left(1-p\right)}}\right)\right)$, puis que $u_n \leq 2\left(1 - \Phi\left(0{,}02\sqrt{n}\right)\right)$.

Q6_ Justifier l'existence de la réciproque Φ^{-1} de Φ. Préciser son domaine de définition.

Déduire de ce qui précède que $u_n \leq \varepsilon$ dès que $n \geq 2500\left(\Phi^{-1}\left(1 - \dfrac{\varepsilon}{2}\right)\right)^2$.

Examiner le cas où $\varepsilon = 0{,}1$.

Q7_ Comparer la qualité de l'approximation fournie par l'inégalité de Bienaymé – Cebycev à celle fournie par l'approximation en loi (*i.e.* par le théorème central limite).

Exercice 8

Soit $\left(X_n\right)_{n \in \mathbb{N}^*}$ une suite de VAR définies sur un même univers probabilisé $\left(\Omega, \mathcal{T}, \mathbb{P}\right)$.

On suppose les $\left(X_n\right)_{n \in \mathbb{N}^*}$ mutuellement indépendantes, à valeurs dans $\mathbb{N} \cup \{-1\}$, et de loi donnée par : $\forall\ n \in \mathbb{N}^*,\ \forall\ k \in X_n(\Omega),\ \mathbb{P}\left(X_n = k\right) = \dfrac{1}{\mathrm{e}\,(k+1)!}$.

Q1_ Déterminer, pour tout entier $n \in \mathbb{N}^*$, la loi de $S_n = \sum_{k=1}^{n} X_k$.

hint *Pour* $k \in \mathbb{N}^*$*, on introduira la VAR* $Y_k = X_k + 1$.

Q2_ A l'aide de la formule de Taylor avec reste intégrale, établir que :

$$\forall\ n \in \mathbb{N}^*,\ \mathbb{P}\left(S_n \le 0\right) = 1 - \frac{1}{n!}\int_0^n t^n \mathrm{e}^{-t}\, dt .$$

Q3_ A l'aide du théorème central limite, montrer que : $\lim\limits_{n \to +\infty} \dfrac{1}{n!}\displaystyle\int_0^n t^n \mathrm{e}^{-t}\, dt = \dfrac{1}{2}$.

Exercice 9

Q1_ Montrer que $f : t \mapsto \dfrac{1}{a\, t^{1+\frac{1}{a}}}\mathbf{1}_{[1,+\infty[}(t)$ est une densité de probabilité ($a > 0$).

Soit X une VAR à densité définie sur un espace probabilisé $\left(\Omega, \mathcal{T}, \mathbb{P}\right)$, de densité f.

Q2_ Donner une CNS sur a pour que X admette une espérance. Calculer alors $\mathbb{E}(X)$.

Q3_ On suppose que X admet une espérance. On pose : $\forall\ n \in \mathbb{N}^*,\ \overline{X_n} = \dfrac{1}{n}\sum_{k=1}^{n} X_k$,

où $\left(X_1, \ldots, X_n\right)$ est n – échantillon *i.i.d.* de la loi de X.

Quelle fonction de a est-elle estimée sans biais par $\overline{X}$?

Q4_ Déterminer la loi de probabilité de la VAR Y définie par : $Y = \dfrac{1}{a}\ln(X)$.

Q5_ Si $\left(x_1, \ldots, x_n\right)$ est une réalisation de l'échantillon, on appelle ***vraisemblance*** de

cette réalisation la fonction $\Lambda : a \mapsto L\left(x_1, \ldots, x_n, a\right) = \prod_{i=1}^{n} f\left(x_i\right)$.

a_ Calculer en fonction de $x_1, \ldots, x_n$ la valeur $t\left(x_1, \ldots, x_n\right)$ de a qui maximise Λ.

L'estimateur $a = t\left(X_1, \ldots, X_n\right)$ *est dit* ***estimateur du maximum de vraisemblance****.*

b_ Montrer que a suit une loi Gamma. En déduire son espérance et sa variance.

c_ Calculer la dérivée seconde de $\ln \Lambda$, et donner sa valeur pour $a = t\left(x_1, \ldots, x_n\right)$.

d_ Montrer que a est un estimateur sans biais et convergent de a.