.

Estimations

Exercice $\boxed{1}$ Soit p un réel appartenant à l'intervalle $]0,1[$. Soient $(X_n)_{n\in\mathbb{N}^*}$ une suite de variables aléatoires définies sur un même espace probabilisé, indépendantes et qui suivent toutes la loi de Bernoulli de paramètre p.

On pose, pour tout n de $\mathbb{N}^*$, $S_n = \displaystyle\sum_{i=1}^{n} X_i$.

Soit ε un réel strictement positif. Justifier que, pour tout n de $\mathbb{N}^*$, on a :

$$\mathbb{P}\left[\left|\frac{S_n}{n} - p\right| \geq \varepsilon\right] \leq \frac{p(1-p)}{n\varepsilon^2}.$$

Exercice $\boxed{2}$ On considère une suite de lancers indépendants d'une même pièce équilibrée. Pour tout n de $\mathbb{N}^*$, on note X_n la variable aléatoire égale à 1 si l'événement «on obtient pile» est réalisé au n-ième lancer et à 0 sinon.

Pour tout n de $\mathbb{N}^*$, on pose $\overline{X_n} = \dfrac{X_1 + X_2 + \ldots + X_n}{n}$.

1. Que représente la suite $\left(\overline{X_n}\right)_{n\in\mathbb{N}^*}$?
2. Justifier que, pour tout réel strictement positif ε, on a :

$$\lim_{n\to+\infty} \mathbb{P}\left[\left|\overline{X_n} - \frac{1}{2}\right| < \varepsilon\right] = 1.$$

3. Le résultat précédent est-il conforme à l'intuition ?

Exercice $\boxed{3}$ Dans tout l'exercice, X est une variable aléatoire suivant la loi de Poisson de paramètre $\lambda > 0$.

1. *Une première inégalité*
   (a) Montrer que l'on a $\mathbb{P}\left[|X-\lambda| \geq \lambda\right] \leq \dfrac{1}{\lambda}$.
   (b) En déduire $\mathbb{P}\left[X - \lambda \geq \lambda\right] \leq \dfrac{1}{\lambda}$ puis $\mathbb{P}\left[X \geq 2\lambda\right] \leq \dfrac{1}{\lambda}$.
2. *Une seconde inégalité*
   Soit Z une variable aléatoire (discrète ou à densité), d'espérance nulle et de variance v strictement positive.
   (a) Montrer que, pour tout couple (a,x) de $]0,+\infty[\times\mathbb{R}_+$, on a :
   $$\mathbb{P}[Z \geq a] \leq \mathbb{P}\left[(Z+x)^2 \geq (a+x)^2\right].$$
   (b) Justifier l'égalité $\mathbb{E}\left[Z^2\right] = v$, puis, en appliquant l'inégalité de Markov à la variable aléatoire $(Z+x)^2$, montrer que, pour tout couple (a,x) de $]0,+\infty[\times\mathbb{R}_+$, on a : $\mathbb{P}[Z \geq a] \leq \frac{v+x^2}{(a+x)^2}$.
   (c) Justifier que, pour tout réel a strictement positif, on a :
   $$\mathbb{P}[Z \geq a] \leq \frac{v+1}{(a+1)^2} \quad \text{et} \quad \mathbb{P}[Z \geq a] \leq \frac{v+a^2}{4a^2}.$$
   (d) Montrer que ensuite que, pour tout réel a strictement positif, on obtient, à l'aide de 2.a) : $\mathbb{P}[Z \geq a] \leq \frac{v}{v+a^2}$.
   Indication : étudier la fonction $x \longmapsto \dfrac{v+x^2}{(a+x)^2}$ sur $]0,+\infty[$.
   (e) En déduire $\mathbb{P}\left[X \geq 2\lambda\right] \leq \dfrac{1}{\lambda+1}$.
   Indication : appliquer le résultat de la question précédente à la variable aléatoire centrée associée à X.

Exercice $\boxed{4}$ Soit $\chi = (X_1 \ ; \ \dots \ ; \ X_n)$ un n-échantillon de variables aléatoires indépendantes et identiquement distribuées associé à n lancés d'une pièce truquée dont la probabilité d'obtenir la face désirée est $p \in]0; 1[$.

On note F_n la fréquence d'apparition de cette face au cours de ces n lancers (à réaliser).

1. Donner la loi des variables X_k $(k \leq n)$ et en rappeler l'espérance et la variance.
2. Ecrire F_n en fonction de $X_1, \dots, X_n$ et en déduire une expression de $\mathbb{E}[F_n]$ en fonction de n et de p.
3. Démontrer que $\mathbb{V}[F_n] \leq \dfrac{1}{4n}$.
4. Déterminer, dans l'inégalité de Bienaymé-Tchebychev appliquée correctement à F_n, une valeur de n la plus petite possible permettant d'assurer que *la fréquence d'apparition de la face attendue lors de n lancers de cette pièce soit p au centième près* survienne dans 95% des réalisations de série n lancers effectués.
5. Reprendre la démarche précédente mais avec 99% des cas.

Exercice $\boxed{5}$ On considère un n-échantillon $(X_1 \ ; \ \dots \ ; \ X_n)$ d'une loi uniforme continue sur $[0; a]$ et on pose $S_n = \max(X_k)_{k \leq n}$.

On cherche à estimer le paramètre a.

1. Démontrer que S_n est bien un estimateur de b et que son biais b_a est $\dfrac{-b}{n+1}$.
2. Est-il asymptotiquement sans biais, *i.e.*, a-t-on $\lim\limits_{n \to +\infty} \mathbb{E}_a[S_n - a] = 0$?
3. On définit le *risque quadratique* r_a de l'estimateur S_n par (sous couvert d'existence) :

$$r_a(S_n) = \mathbb{V}[S_n] + {b_a}^2$$

   Etablir que S_n admet bien un risque quadratique et qu'il vaut $\dfrac{2a^2}{(n+1)(n+2)}$

4. Déterminer $\lim\limits_{n \to +\infty} r_a(S_n)$.

Exercice $\boxed{6}$ Soit X une variable aléatoire qui suit la loi de Bernoulli de paramètre p ($p \in]0, 1[$). Soient n un entier strictement positif et $(X_1, \dots, X_n)$ un n-échantillon de X. On définit la variable aléatoire $\overline{X_n}$ par $\overline{X_n} = \dfrac{1}{n} \displaystyle\sum_{i=1}^{n} X_i$.

1. Justifier que $\overline{X_n}$ admet une espérance et une variance et les déterminer.
2. Montrer que, pour tout réel strictement positif ε, on a :

$$\mathbb{P}\left(\left|\overline{X_n} - p\right| \leq \varepsilon\right) \geq 1 - \frac{p(1-p)}{n\varepsilon^2}$$

3. En déduire :

$$\mathbb{P}\left[\left|\overline{X_n} - p\right| \leq \frac{\sqrt{p(1-p)}}{\sqrt{n \times 0,05}}\right] \geq 0,95$$

   puis $\mathbb{P}\left[\overline{X_n} - \dfrac{\sqrt{p(1-p)}}{\sqrt{n \times 0,05}} \leq p \leq \overline{X_n} + \dfrac{\sqrt{p(1-p)}}{\sqrt{n \times 0,05}}\right] \geq 0,95.$

4. En étudiant la fonction $x \longmapsto x(1-x)$, justifier que, pour tout p de $]0, 1[$, on a : $p(1-p) \leq \dfrac{1}{4}$.
5. En déduire :

$$\mathbb{P}\left[\overline{X_n} - \frac{1}{2\sqrt{n \times 0,05}} \leq p \leq \overline{X_n} + \frac{1}{2\sqrt{n \times 0,05}}\right] \geq 0,95.$$

6. Que peut-on dire de l'intervalle $\left[\overline{X_n} - \dfrac{1}{2\sqrt{n \times 0,05}}, \overline{X_n} + \dfrac{1}{2\sqrt{n \times 0,05}}\right]$ vis-à-vis de p ?
7. *Application* :

(a) 1 000 électeurs choisis au hasard ont été intérrogés avant une élection. 520 déclarent être favorables au candidat A.
Déterminer une estimation de l'intervalle de confiance à 95% de la proportion p d'électeurs favorables au candidat A dans la population.

(b) N électeurs choisis au hasard ont été intérrogés avant une élection. On suppose que la proportion de candidats favorables au candidat A obtenue est 0,52 et que l'intervalle de confiance à 95% de la proportion p d'électeurs favorables au candidat A dans la population (déterminé avec la même méthode que dans la question précédente) ne contient aucune valeur inférieure ou égale à 0,5.
Quelle est la valeur minimale de N ?

Exercice $\boxed{7}$ **D'après ESCP 2022**

On considère n variables aléatoires $X_1, \ \dots, \ X_n$ indépendantes et de même loi.
On suppose que cette loi permet d'écrire $\mathbb{E}[X_i] = a$ et $\mathbb{E}[{X_i}^2] = \frac{7}{6}a^2$ pour $i \leq n$ où a est un paramètre à estimer dans $]0;1]$.

On pose $\overline{X}_n = \dfrac{1}{n}\displaystyle\sum_{k=1}^{n} X_k$

1. Démontrer que les X_i ont bien une variance et la donner.
2. Déterminer l'espérance et la variance de la variable aléatoire $\overline{X}_n$.
3. En déduire que $\overline{X}_n$ est un estimateur sans biais de a.
4. Quel est le risque quadratique de $\overline{X}_n$?
   On donne le risque quadratique $r_a(T)$ d'un estimateur T : $r_a(T) = \mathbb{V}_a[T] + b_a(T)^2$
5. Soit $\varepsilon > 0$ fixé.
   (a) Ecrire l'inégalité de Bienaymé-Tchebychev pour $\overline{X}_n$ puis établir l'inégalité :
   $$\mathbb{P}\left[\left|\overline{X}_n - a\right| \geq \varepsilon\right] \leq \frac{1}{6n\varepsilon^2}$$
   (b) En déduire l'inégalité :
   $$\mathbb{P}\left[\overline{X}_n - \varepsilon \leq a \leq \overline{X}_n + \varepsilon\right] \geq 1 - \frac{1}{6n\varepsilon^2}$$
   (c) On a réalisé 1000 simulations $X_1, X_2, \dots, X_{1000}$ de cette loi par un programme informatique.
   En prenant $\varepsilon = \frac{1}{\sqrt{600}}$, donner l'intervalle de confiance pour a correspondant à cette situation (en fonction de $\overline{X}_{1000}$)
   (d) Quel est le niveau de confiance de ce dernier intervalle ?

Exercice $\boxed{8}$ **Un problème d'estimations à densité**

Soit $\theta > 0$ réel. On considère la fonction f définie par :
$$f(t) = e^{\theta - t}\mathbb{1}_{[\theta;+\infty[}(t)$$

1. Etude sommaire de f
   (a) Vérifier que f ainsi définie est bien une densité de probabilité.
   (b) Décrire la fonction de répartition F associée à f.
   (c) Déterminer l'espérance et la variance associées à la loi caractérisée par F.

   Soient alors T une variable aléatoire de densité f ainsi que n un entier supérieur ou égal à 2 et $(T_1, \dots, T_n)$ un n-échantillon de T.
2. On définit à présent une variable aléatoire par $Y_n = \dfrac{1}{n}\displaystyle\sum_{k=1}^{n} T_k$
   (a) Vérifier que Y_n admet une espérance et une variance et les déterminer.
   (b) Etablir alors que $\hat{Y}_n = Y_n - 1$ est un estimateur sans biais de θ
   (c) Quel est le risque quadratique de $\hat{Y}_n$?