.

Estimations

Exercice $\boxed{1}$ Fréquence attendue

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 $\boxed{2}$ Moyenne et Variance empiriques : reprise des exemples du cours Soit X une variable aléatoire suivant une loi d'espérance $\mu \in \mathbb{R}$ et de variance $\sigma^2 > 0$. On désigne par $(X_1, \dots, X_n)$ un n-échantillon de X.

1. Justifier que $\overline{X}_n = \dfrac{1}{n}\displaystyle\sum_{k=1}^{n} X_k$ est un estimateur sans biais de μ
2. Dans cette question, on suppose à présent μ connu.
   (a) Etablir que $V_n = \dfrac{1}{n}\displaystyle\sum_{k=1}^{n}(X_k - \mu)^2$ est un estimateur sans biais de σ^2
   (b) On pose $V_\sigma = \dfrac{1}{n}\displaystyle\sum_{k=1}^{n}(X_k - \overline{X}_n)^2$. Démontrer que :
   $$V_\sigma = \frac{1}{n}\sum_{k=1}^{n} X_k^2 \quad - \overline{X}_n^2 \; = \; \frac{1}{n}\sum_{k=1}^{n}(X_k - \mu)^2 \quad - (\overline{X}_n - \mu)^2$$
   et en déduire que V_σ est un estimateur biaisé de σ^2 et déterminer la limite de ce biais lorsque n tend vers $+\infty$.

Exercice $\boxed{3}$ Estimateur naïf biaisé

On considère une variable aléatoire X de loi uniforme continue sur $[a;b]$ et on pose $\theta = (a;b) \in \mathbb{R}^2$, en supposant $a < b$. On désignera par $\chi = (X_1; \dots; X_n)$ un n-échantillon de X et par $\Theta \subset \mathbb{R}^2$ l'ensemble des valeurs possibles de θ.

1. Décrire Θ et en indiquer la nature topologique.
2. Justifier que $T_n = \max(\chi) - \min(\chi)$ est un estimateur biaisé de $g(\theta) = b - a$.
3. Démontrer que T_n est asymptotiquement sans biais.

Exercice $\boxed{4}$ On considère un n-échantillon $(X_1 \; ; \; \dots \; ; \; X_n)$ d'une loi uniforme continue sur $[a;b]$ et on pose $S_n = \max(X_k)_{k \leq n}$. On cherche à estimer le paramètre b (et on suppose a connu).

1. Démontrer que S_n est bien un estimateur de b et que son biais est $\dfrac{a-b}{n+1}$.
   Est-il asymptotiquement sans biais ?
2. Vérifier que le risque quadratique de S_n vaut $\dfrac{2(b-a)^2}{(n+1)(n+2)}$ après en avoir justifié l'existence.

Exercice $\boxed{5}$ *D'après Ens D2 - paris Saclay 2014*

Soit $(X_i)_{i\in\mathbb{N}^*}$ une suite de variables aléatoires mutuellement indépendantes et identiquement distribuées suivant une loi de Bernoulli de paramètre p. Pour tout $i \geq 1$ on pose $Y_i = X_i + X_{i+1}$

1. Donner la loi de Y_i pour $i \in \mathbb{N}^*$
2. Pour $n \geq 1$ on pose $T_n = \dfrac{Y_1 + \cdots + Y_n}{n}$. Calculer l'espérance mathématique et la variance de T_n.
3. Montrer que la suite $(T_n)_{n>0}$ converge en probabilités vers la variable aléatoire constante $X = 2p$.

Exercice $\boxed{6}$ Soit $\theta > 0$. On définit une famille $(p_k)_{k\in\mathbb{N}}$ de réels par :

$$\forall k \in \mathbb{N} \quad p_k = \frac{1}{1+\theta}\left(\frac{\theta}{1+\theta}\right)^k$$

1. Vérifier que la famille $(p_n)_{n\in\mathbb{N}}$ permet de définir une loi de probabilité (discrète).
2. Soit alors X une variable aléaoire telle que $\forall k \in \mathbb{N}\ \ \mathbb{P}[X = k] = p_k$. On notera $(X_1, \ldots, X_n)$ un n-échantillon de X.
   (a) Quelle est la loi de $Y = X + 1$?
   (b) Démontrer que $T_n = \frac{1}{n}\displaystyle\sum_{i=1}^{n} X_i$ est un estimateur sans biais de θ.

Exercice $\boxed{7}$ **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 $(T_1, \ldots, T_n)$ 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 $\hat{Y}_n$?
3. On définit cette fois une nouvelle variable aléatoire $Z_n = \min\{T_1, \ldots,\ T_n\}$ et on note G_n la fonction de répartition de Z_n.
   (a) Justifier que : $\forall x \in \mathbb{R} \quad 1 - G_n(x) = (1 - F(x))^n$
   (b) En déduire une expression de G_n puis établir que Z_n est à densité. Exprimer alors une densité g_n de Z_n.
   (c) Justifier que Z_n admet des moments d'ordre 1 et 2 puis calculer $\mathbb{E}[Z_n]$ ainsi que $\mathbb{V}[Z_n]$.
   (d) Démontrer que $\hat{Z}_n = Z_n - 1$ est un (autre) estimateur sans biais de θ puis déterminer le risque quadratique de $\hat{Z}_n$.
4. Comparer les risques quadratiques de $\hat{Y}_n$ et $\hat{Z}_n$.

Exercice $\boxed{8}$ **autour d'une loi de Poisson**

Soit X une variable aléatoire suivant une loi de Poisson dont le paramètre λ est à estimer. On notera $(X_1\ ;\ \ldots\ ;\ X_n)$ un n-échantillon associé.

1. Démontrer que $\overline{X}_n$ est un estimateur sans biais de λ. Quel $\Theta \subset \mathbb{R}$ choisir pour λ ?

2. Peut-on dire que ${S_n}^2$, variance empirique, est un (autre) estimateur sans biais de λ ? Asymptotiquement sans biais ?
3. On pose $T_n = \exp(-x\overline{X}_n)$ pour $x \in \mathbb{R}$ donné. On souhaite à présent estimer $g(\lambda) = e^{-\lambda x}$.
   (a) Justifier que, si X et Y suivent une même loi de Poisson de paramètre λ et sont indépendantes, alors $X + Y \hookrightarrow \mathcal{P}(2\lambda)$.
   (b) En déduire que $Y_n = n\overline{X}_n \hookrightarrow \mathcal{P}(n\lambda)$, pour tout $n \geq 1$.
   (c) Etablir que, pour $x \in \mathbb{R}$, $n \geq 1$ entier naturel, et $k \in \mathbb{N}$ on a :
   $$e^{-\frac{xk}{n}}\mathbb{P}_\lambda[Y_n = k] = \frac{\left(n\lambda e^{-x/n}\right)^k}{k!}e^{-\lambda n}$$
   (d) En déduire que l'estimateur T_n admet une espérance $\mathbb{E}_\lambda[T_n]$ et qu'elle vaut $\exp\left(\lambda n(e^{-x/n} - 1)\right)$.
   (e) L'estimateur T_n est-il biaisé ? Asymptotiquement biaisé ? Convergent ?

Exercice $\boxed{9}$ **autour d'une loi exponentielle** Soit X une variable aléatoire suivant une loi exponentielle dont le paramètre $\lambda > 0$ est à estimer. On notera $(X_1\ ;\ \ldots\ ;\ X_n)$ un n-échantillon associé.
On pose $T_n = n \cdot \min(X_1 \ldots X_n)$.

1. Démontrer que T_n est un estimateur sans biais de $\frac{1}{\lambda}$.
2. Etablir que T_n n'est en revanche pas convergent.

Exercice $\boxed{10}$ *D'après Ens D2 - paris Saclay 2015*
On observe un échantillon $X_1, \ldots, X_n$ indépendant et identiquement distribué de loi uniforme sur $[0; 2a]$ où a est un réel strictement positif, de densité :
$$f_a(x) = \frac{1}{2a} \cdot \mathbb{1}_{[0;2a]}(x)$$
Dans ce problème, on cherche à étudier deux estimateurs de a.

1. Soit $\hat{a} = \overline{X}_n = \dfrac{1}{n}\displaystyle\sum_{k=1}^{n} X_k$ moyenne des observations.
   (a) Calculer l'espérance d'une variable aléatoire de densité f_a
   (b) Montrer que $\hat{a}$ est un estimateur sans biais de a.
   (c) Calculer la variance de $\hat{a}$.
2. Soit $M = \max\{X_1, \ldots, X_n\}$
   (a) Déterminer la fonction de répartition F_a d'une variable aléatoire de densité f_a
   (b) Pour tout $t \geq 0$, calculer $\mathbb{P}[M \leq t]$
   (c) En déduire une densité de M.
   (d) Pour c réel donné, calculer l'espérance de cM.
   (e) En déduire un estimateur sans biais de a, que l'on notera $\tilde{a}$.
   (f) Calculer la variance de $\tilde{a}$.
3. Conclure : quel est le meilleur estimateur ?