TD 3 : Estimation

Exercice 1. *Soit* $(X_1, \dots, X_n)$ *un échantillon de taille* n *d'une loi de moyenne* $\mu \in \mathbb{R}$ *et de variance* $\sigma^2 \geq 0$.

1. *Donner une condition nécessaire et suffisante sur les constantes réelles* $a_1, \dots, a_n$ *pour que :* $\sum_{i=1}^n a_i X_i$ *soit un estimateur sans biais de* μ.
2. *Parmi tous les estimateurs de* μ *de la forme* $\sum_{i=1}^n a_i X_i$, *quel est celui de variance minimale ? Quel est le biais de cet estimateur ?*
3. *Parmi tous les estimateurs de* μ *de la forme* $\sum_{i=1}^n a_i X_i$, *quel est celui dont l'erreur quadratique moyenne est minimale ?*
4. *Parmi les estimateurs sans biais de* μ *de la forme* $\sum_{i=1}^n a_i X_i$, *quel est celui de variance minimale ?*

Exercice 2. *Soit* $(X_1, \dots, X_n)$ *un échantillon de taille* n *d'une loi de Poisson de paramètre* $\lambda > 0$ *inconnu. On rappelle qu'une variable aléatoire réelle* X *ayant une loi de Poisson de paramètre* λ *vérifie :*

$$\forall x \in \mathbb{N} \quad \mathbb{P}(X = x) = \frac{\lambda^x \exp(-\lambda)}{x!}.$$

1. *Calculer l'estimateur de maximum de vraisemblance* $\widehat{\lambda}^{\mathrm{mv}}$ *de* λ.
2. *Calculer le biais et l'erreur quadratique moyenne de* $\widehat{\lambda}^{\mathrm{mv}}$.
3. *Quelles sont les propriétés asymptotiques de l'estimateur* $\widehat{\lambda}^{\mathrm{mv}}$.
4. *En utilisant la normalité asymptotique de l'estimateur* $\widehat{\lambda}^{\mathrm{mv}}$ *et la méthode-delta, établir la convergence en loi suivante*
$$\lim_{n \to +\infty} \mathbb{P}\left(\frac{\sqrt{\widehat{\lambda}^{\mathrm{mv}}} - \sqrt{\lambda}}{0.5/\sqrt{n}} \leq t \right) = \int_{-\infty}^{t} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx$$
5. *Le code R suivant permet de générer un échantillon expérimental de taille mille de la variable* $Z = (\sqrt{\widehat{\lambda}^{\mathrm{mv}}} - \sqrt{10})/(0.5/\sqrt{50})$ *et compare la loi de cette variable à une loi* $\mathcal{N}(0,1)$.

```
nb=1000
n=50
lambda=10
Z=numeric(nb)
for(i in (1:nb))
{
  X=rpois(n,lambda)
  Z[i]=(sqrt(mean(X))-sqrt(lambda))/(0.5/sqrt(n))
}
plot(density(Z))
curve(dnorm(x),add=TRUE,col="red")
Zord=sort(Z)
Y=(1:nb)/nb
plot(Zord,Y)
curve(pnorm(x),add=TRUE,col="red")
```

Que se passe-t-il lorsque l'on modifie dans ce code les valeurs de n, nb *et lambda.*

On a relevé le nombre d'accidents aux passages à niveau avec un train de 2006 *à* 2021 *au Danemark :*

Année	*2006*	*07*	*08*	*09*	*10*	*11*	*12*	*13*	*14*	*15*	*16*	*17*	*18*	*19*	*20*	*21*
Collisions	*8*	*5*	*4*	*2*	*9*	*2*	*5*	*5*	*5*	*2*	*2*	*2*	*3*	*1*	*0*	*2*

On modélise le nombre d'accidents au cours d'une année comme une variable aléatoire de loi de Poisson de paramètre inconnu $\lambda > 0$.

6. *En utilisant la question 1, déterminer une estimation ponctuelle de* λ.
7. *En utilisant le code R précédent et les données de cet exercice, semble-t-il raisonnable d'approcher la loi de* $Z = (\sqrt{\widehat{\lambda}^{\mathrm{mv}}} - \sqrt{\lambda})/(0.5/\sqrt{n})$ *par une loi normale* $\mathcal{N}(0,1)$ *?*
8. *Déterminer une estimation par intervalle pour* λ *de niveau de confiance asymptotique* $0,95$.

Exercice 3. *Soit* $(X_1,\ldots,X_n)$ *un échantillon de taille* n *d'une loi de exponentielle de paramètre* $\lambda > 0$ *inconnu. On rappelle que la densité d'une loi exponentielle de paramètre* λ *est :*

$$f_\lambda(x) = \lambda \exp(-\lambda x)\, \mathbf{1}_{x\geq 0}.$$

1. *Calculer l'estimateur de maximum de vraisemblance* $\widehat{\lambda}^{\mathrm{mv}}$ *de* λ.
2. *On pose* $\theta = 1/\lambda$. *Déduire de la question précédente un estimateur de* $\widehat{\theta}$ *de* θ.
3. *Calculer le biais et l'erreur quadratique moyenne de* $\widehat{\theta}$.
4. *Quelles sont les propriétés asymptotiques de l'estimateur* $\widehat{\theta}$.
5. *En utilisant la normalité asymptotique de l'estimateur* $\widehat{\theta}$ *et la méthode-delta, établir la convergence en loi suivante*
$$\lim_{n\to+\infty} \mathbb{P}\left(\frac{\ln(\widehat{\theta}) - \ln(\theta)}{1/\sqrt{n}} \leq t\right) = \int_{-\infty}^{t} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx$$
6. *En vous aidant du code R de l'exercice 2, illustrer la convergence en loi donnée à la question précédente.*

On a observé les durées de vie (en heure) de 30 *composants électroniques. Les résultats suivants ont été obtenus :*

0.1	7.4	1.0	7.9	2.1	1.8	17.9	9.3	6.5	3.3	5.6	7.7	0.1	24.3	8.1
10.0	11.9	1.6	2.7	0.5	5.8	42.5	5.1	2.0	0.2	15.0	3.5	6.4	0.6	3.3

On modélise la durée de vie par une variable aléatoire de loi de exponentielle de paramètre inconnu $\lambda > 0$.

7. *En utilisant la question 2, déterminer une estimation ponctuelle de* θ. *On pourra utiliser le code R suivant :*
```
heure=c(0.1, 7.4, 1.0, 7.9, 2.1, 1.8, 17.9, 9.3, 6.5, 3.3, 5.6, 7.7, 0.1, 24.3, 8.1,
10.0, 11.9, 1.6, 2.7, 0.5, 5.8, 42.5, 5.1, 2.0, 0.2, 15.0, 3.5, 6.4, 0.6, 3.3)
```
8. *En utilisant le code R vous avez implémenté et les données de cet exercice, semble-t-il raisonnable d'approcher la loi de* $Z = (\ln(\widehat{\theta}) - \ln(\theta))/(1/\sqrt{n})$ *par une loi normale* $\mathcal{N}(0,1)$ *?*
9. *Déterminer une estimation par intervalle pour* θ *de niveau de confiance asymptotique* $0,95$.
10. *Calculer la fonction de répartition d'une la loi exponentielle de paramètre* λ. *En utilisant la question 9, donner un encadrement de la probabilité que la durée de vie d'un composant soit inférieure à* $20h$, $30h$ *et* $40h$.

Exercice 4. *Soit* $(X_1,\ldots,X_n)$ *un échantillon de taille* n *d'une loi dont la densité est :*

$$f_\theta(x) = \frac{x}{\theta^2} \exp\left(-\frac{x}{\theta}\right) \mathbf{1}_{x\geq 0},$$

où $\theta > 0$ *est un paramètre inconnu.*

1. *Calculs préliminaires : On pose pour* $n \geq 0$ *:*
$$J_n(\theta) = \int_0^\infty x^n exp\left(-\frac{x}{\theta}\right) dx.$$
   (a) *Montrer, à l'aide d'une intégration par parties, que* $J_{n+1}(\theta) = (n+1)\theta J_n(\theta)$. *En déduire que pour* $n \geq 0$, $J_n(\theta) = \theta^{n+1} n!$.
   (b) *En déduire que* f_θ *définit bien une densité. Soit* X *est une variable aléatoire réelle dont la densité est* f_θ *montrer que* $\mathbb{E}(X) = 2\theta$ *et* $\mathrm{var}(X) = 2\theta^2$.
2. *Calculer l'estimateur du maximum de vraisemblance* $\widehat{\theta}$ *de* θ.
3. *Calculer le biais et l'erreur quadratique moyenne de* $\widehat{\theta}$.
4. *Quelles sont les propriétés asymptotiques de l'estimateur* $\widehat{\theta}$ *?*

Exercice 5 (Ecrits CAPES, 2020)**.** *On rappelle qu'une variable aléatoire X sur un espace probabilisé $(\Omega, \mathcal{P}(\Omega), P)$ fini suit une loi de Bernoulli de paramètre p si elle prend comme valeurs 1 et 0 avec les probabilités $P(X = 1) = p$ et $P(X = 0) = 1 - p$. Dans toute cette section, $X_1, X_2, \ldots, X_n$ désignent n variables aléatoires indépendantes de même loi de Bernoulli de paramètre p. On pose*

$$S_n = \sum_{i=1}^{n} X_i \text{ et } F_n = \frac{1}{n} S_n.$$

1. *Calculer l'espérance et la variance de X_1.*
2. *Étude des variables aléatoires S_n et F_n*
   a) *Calculer l'espérance et la variance de la variable S_n.*
   b) *Que représente la variable S_n ? Rappeler sa loi de probabilité.*
   c) *Calculer l'espérance et la variance de la variable F_n.*
   d) *Que représente la variable F_n ? Déterminer sa loi de probabilité.*
3. **Inégalité de Markov** *Soit Y une variable aléatoire positive définie sur Ω. On note $E(Y)$ son espérance. En décomposant $Y(\Omega) = Y_1 \cup Y_2$, avec $Y_1 = \{y \in Y(\Omega), y \geq a\}, Y_2 = \{y \in Y(\Omega), y < a\}$, démontrer que, pour tout nombre réel a strictement positif,*

$$P(Y \geq a) \leq \frac{E(Y)}{a}.$$

4. **Inégalité de Bienaymé-Tchebychev** *Soit X une variable aléatoire définie Ω. On note $E(X)$ son espérance et $V(X)$ sa variance. Démontrer que, pour tout nombre réel a strictement positif,*

$$P(|X - E(X)| \geq a) \leq \frac{V(X)}{a^2}.$$

5. *On reprend les notations de l'introduction de la partie E (i.e. de cet exercice).*
   a) *Démontrer que, pour tout réel ϵ strictement positif,*

$$P(|F_n - p| \geq \epsilon) \leq \frac{1}{4n\epsilon^2}$$

   b) *Expliquer comment, lorsque p est inconnu, cette inégalité permet d'en fournir une estimation. Comment s'appelle le théorème sous-jacent ?*

Application *Un problème historique dû au Chevalier de Méré est rapporté dans la correspondance entre Pascal et Fermat. Grand joueur, le chevalier de Méré s'intéressait aux jeux de hasard sur lesquels il misait de l'argent. À l'issue de nombreuses parties, il avait constaté avoir plus d'une chance sur deux d'obtenir au moins une fois un six en lançant quatre fois un dé à six faces et moins d'une chance sur deux d'obtenir au moins un double-six en lançant 24 fois deux dés. Ce résultat lui semblait en contradiction avec l'égalité des rapports $\frac{4}{6} = \frac{24}{36}$ et du nombre de lancers au nombre de faces.*

1. *Calculer la probabilité d'obtenir au moins un six à l'issue de 4 lancers d'un dé.*
2. *Calculer la probabilité d'obtenir au moins un double-six à l'issue de 24 lancers de deux dés.*
3. *(Ne pas traiter cette question) A-t-on plus ou moins d'une chance sur deux d'obtenir au moins un six en lançant quatre fois un dé à six faces ?*
4. *(Ne pas traiter cette question) A-t-on plus ou moins d'une chance sur deux d'obtenir au moins un double-six en lançant vingt-quatre fois deux dés à six faces ?*
5. *(Ne pas traiter cette question) Le texte ci-dessous reproduit l'extrait d'une lettre adressée par Fermat à Pascal en 1654. « Monsieur, Je n'ai pas eu le temps de vous envoyer la démonstration d'une difficulté qui étonnait fort M. de Méré. Il me disait donc qu'il avait trouvé fausseté dans les nombres par cette raison : si on entreprend de faire un six avec un dé, il y a avantage de l'entreprendre en 4, comme de 671 à 625. Si on entreprend de faire un double six avec deux dés, il y a désavantage de l'entreprendre en 24. Et néanmoins 24 est à 36 (qui est le nombre des faces de deux dés) comme 4 à 6 (qui est le nombre des faces d'un dé). Voilà quel était son grand scandale qui lui faisait dire hautement que les propositions n'étaient pas constantes et que l'arithmétique se démentait : mais vous en verrez bien aisément la raison par les principes où vous êtes. » Expliquer comment ce texte historique pourrait être utilisé en classe pour illustrer les réponses aux questions 3. et 4. Quelle est l'erreur de raisonnement commise par le Chevalier de Méré ?*

6. *En appliquant l'inégalité de Bienaymé-Tchebychev, déterminer au bout de combien de répétitions d'un lancer de quatre dés, la fréquence d'apparition d'au moins un six est supérieure ou égale à* $\frac{1}{2}$ *avec une probabilité supérieure ou égale à* $0,95$.
7. *En appliquant l'inégalité de Bienaymé-Tchebychev, déterminer au bout de combien de répétitions d'un lancer de vingt-quatre dés, la fréquence d'apparition d'au moins un double-six est inférieure ou égale à* $\frac{1}{2}$ *avec une probabilité supérieure ou égale à* $0,95$.
8. *(Ne pas traiter cette question) Quels commentaires vous inspirent ces résultats ?*