Feuille de travaux dirigés 3 : risque quadratique

Exercice 1 (Loi de Poisson)**:**
Soient $X_1, \ldots, X_n$ n variables aléatoires indépendantes identiquement distribuées de loi de Poisson de paramètre $\lambda > 0$.

1. Proposer un estimateur sans biais de λ et calculer son risque quadratique.
   On souhaite à présent trouver un estimateur du paramètre $\theta = \mathbb{P}_\lambda(X_1 = 0)$.
2. Proposer un estimateur biaisé et un estimateur sans biais de θ.
3. Les estimateurs précédents sont ils efficaces ?

Exercice 2 (Modèle de Bernoulli : moyenne et variance)**:**
Soient $X_1, \ldots, X_n$, n variables aléatoires i.i.d. de loi $B(\theta)$ (Bernoulli de paramètre θ).

1. Montrez que
$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$
   est un estimateur de θ efficace.
2. Trouver un estimateur sans biais de la variance des X_i de la forme $\widehat{v} = \eta \overline{X}(1 - \overline{X})$.

Exercice 3 (modèle de Pareto)**:**
On reprend l'exemple du modèle de Pareto pour la modélisation du trafic internet, vu au TD 2. On rappelle que une variable aléatoire X_1 suit une loi de Pareto $\mathcal{P}ar(u, \theta)$, avec $u > 0$ et $\theta > 0$, si la fonction de répartition de X_1 est

$$F_\theta(x) = \begin{cases} 0 & \text{si } x \leq u \\ 1 - \left(\frac{u}{x}\right)^\theta & \text{si } x > u. \end{cases}.$$

Dans la suite in fixe $u > 0$ supposé connu. On pourra utiliser les résultats suivant :

1. **Loi Gamma** Une variable aléatoire Y suit une loi Gamma de paramètres α et λ ($\alpha > 0$ et $\lambda > 0$), notée $\mathcal{G}amma(\alpha, \lambda)$, si elle admet une densité par rapport à la mesure de Lebesgue donnée par
$$f^{\mathcal{G}}_{(\alpha,\lambda)}(y) = \mathbb{1}_{y>0} \frac{\lambda^\alpha}{\Gamma(\alpha)} y^{\alpha-1} e^{-\lambda y}.$$
On rappelle que pour $\alpha > 0$, $\Gamma(\alpha+1) = \alpha\Gamma(\alpha)$. Si $Y \sim \mathcal{G}amma(\alpha, \lambda)$, on a
$$\mathbb{E}_{\alpha,\lambda}(Y) = \frac{\alpha}{\lambda} \quad ; \quad \mathbb{V}\mathrm{ar}_{\alpha,\beta}(Y) = \frac{\alpha}{\lambda^2}.$$

2. **Loi Inverse Gamma** Si $Y \sim \mathcal{G}amma(\alpha, \lambda)$, alors $T := \frac{1}{Y}$ suit une loi dite 'inverse gamma' $\mathcal{IG}(\alpha, \lambda)$, de densité
$$f^{\mathcal{IG}}_{\alpha,\lambda}(t) = \mathbb{1}_{t>0} \frac{\lambda^\alpha}{\Gamma(\alpha)} \frac{1}{t^{\alpha+1}} e^{-\lambda/t},$$
et l'on a, lorsque $\alpha > 1$ (*resp.* $\alpha > 2$) :
$$\mathbb{E}_{\alpha,\lambda}(T) = \frac{\lambda}{\alpha-1} \quad ; \quad (\text{resp. } \mathbb{V}\mathrm{ar}_{\alpha,\lambda}(T) = \frac{\lambda^2}{(\alpha-1)^2(\alpha-2)}.)$$

3. **Somme d'exponentielles** Si $(Z_i)_{i=1}^n \overset{i.i.d.}{\sim} \mathcal{E}(\lambda)$, alors $Y := \sum_{i=1}^n Z_i$ suit une loi Gamma de paramètres (n, λ) :
$$\sum_{i=1}^n Z_i \sim \mathcal{G}amma(n, \lambda).$$

1. Rappeler l'expression de l'estimateur de maximum de vraisemblance $\widehat{\theta}_{MV}$ (TD2).

On se demande maintenant si cet estimateur est « bon », au sens du risque quadratique.

2. Montrer que $Y_i = \log(X_i/u)$ suit une loi exponentielle dont on précisera le paramètre.
3. Montrer que si une variable aléatoire $V \sim \mathcal{G}amma(\alpha, \lambda)$, alors $\lambda V \sim \mathcal{G}amma(\alpha, 1)$.
4. Montrer ensuite que $\widehat{\theta}_{MV}(X) \sim \mathcal{IG}(n, n\theta)$.
5. Quelle est le biais de $\widehat{\theta}_{MV}$? Quel est son risque quadratique ?
6. Montrer que l'estimateur 'corrigé'
$$\widehat{\theta}_c(X) = \frac{n-1}{\sum_{i=1}^n \log(X_i/u)}$$
est un estimateur sans biais de θ.
7. Calculer l'information de Fisher $I_n(\theta)$ pour le modèle $\mathcal{P}^{\otimes n}$ à n observations indépendantes.
8. Calculer la variance de $\widehat{\theta}_c(X)$.
9. Les estimateurs $\widehat{\theta}_{MV}$ et $\widehat{\theta}_c$ sont -ils efficaces ?
10. Si l'utilisateur est sensible au risque quadratique sur l'erreur d'estimation, quel estimateur doit-il préférer : $\widehat{\theta}_c$ ou $\widehat{\theta}_{MV}$?