Statistique inférentielle - TD 1

Exercice 1.

Dans chacun des cas suivants, écrire le modèle statistique.

1. On observe $(X_1, ..., X_n)$ qui sont i.i.d. de loi binomiale de paramètres (m, p).
2. On observe $(X_i)_{i\in\mathbb{N}^*}$ i.i.d. de loi de Poisson de paramètre λ.
3. On observe $(X_i)_{i\in\mathbb{N}^*}$ de loi normale de paramètres (m, σ^2).
4. On observe $(X_1, ..., X_n)$ i.i.d. de loi exponentielle de paramètre λ.
5. On observe $(X_1, ..., X_n)$ tels que $X_i|X_{i-1} = x_{i-1}, ..., X_1 = x_1$ suit une loi $\mathcal{N}(x_{i-1}, 1)$ pour $i > 2$, et X_1 suit une loi $\mathcal{N}(0, 1)$.

Exercice 2. Notion de modèle statistique dominé.

Un modèle statistique $(\Omega, \mathcal{F}, (\mathbb{P}_\theta)_{\theta\in\Theta})$ est dit dominé par une mesure μ si et seulement si :

$$A \in \mathcal{F}\ t.q.\ \mu(A) = 0, \implies \forall\theta \in \Theta,\ \mathbb{P}_\theta(A) = 0.$$

Reprendre les modèles statistiques de l'exercice précédent, et montrer qu'ils sont tous dominés en proposant une mesure μ qui les domine.

Exercice 3. On considère X qui suit une loi de Bernoulli de paramètre p inconnu, où l'on suppose que $p \in \{1/4, 1/2\}$.

1. Ecrire le modèle statistique correspondant.
2. Lister tous les estimateurs de p possibles à valeurs dans $\Theta = \{1/4, 1/2\}$.
3. On considère la perte quadratique. Déterminer la fonction de risque $\theta \to E_\theta[(\hat{p} - \theta)^2]$ pour chaque estimateur $\hat{p}$ de p.
4. Déterminer l'estimateur minimax.
5. On considère une approche bayésienne. Quelles sont les distributions a priori que l'on peut mettre sur Θ ?
6. Exprimer le risque bayésien en fonction de l'a priori, et déterminer l'estimateur bayésien correspondant.

Exercice 4.

On considère la perte quadratique. On appelle **biais** d'un estimateur $\hat{\theta}$ la quantité

$$B(\hat{\theta}) = E_\theta[\hat{\theta}] - \theta.$$

1. Montrer qu'on peut décomposer

$$E_\theta[(\hat{\theta} - \theta)^2] = B(\hat{\theta})^2 + Var_\theta(\hat{\theta}).$$

2. On considère $(X_1, ..., X_n)$ i.i.d. de loi $\mathcal{N}(\theta, \sigma^2)$ où l'on cherche à estimer θ. On considère

$$\hat{\theta} = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

   Quelle est la fonction de risque correspondante ?

3. On considère $(X_1, ...X_n)$ i.i.d. de loi $\mathcal{E}(\lambda)$, et on cherche à estimer λ. Proposer un estimateur "raisonnable" de λ, et donner le signe de son biais.

4. Déterminer le comportement asymptotique de cet estimateur.
5. Pour estimer $\theta = 1/\lambda$, montrer que l'estimateur de la question 3 n'est pas admissible (indication : on pourra considérer la classe d'estimateurs $\bar{X}_k = \frac{1}{n+k}\sum_{i=1}^n X_i$ pour $k = \{-(n-1), ..., -1, 1, 2, ...\}$).

<u>Exercice 5.</u> Mélange Gamma-Poisson.

On considère un portefeuille d'assurance où le nombre de sinistres X_i de l'assuré i dans une année suit une loi de Poisson de paramètre θ_i. On suppose que les assurés sont hétérogènes. Les θ_i sont distincts et inconnus. On suppose qu'ils ont été tirés aléatoirement suivant une distribution π.

1. A votre avis, pourquoi le choix d'une loi de Poisson, et pourquoi pas une loi binomiale, pour modéliser le nombre de sinistres ?
2. Comment s'interprète la distribution π dans l'exemple considéré ?
3. On suppose que π est une loi Gamma de paramètres (r, λ), i.e. de densité par rapport à la mesure de Lebesgue

$$f_{r,\lambda}(t) = \frac{1}{\Gamma(r)}\lambda^r t^{r-1}\exp(-\lambda t)\mathbf{1}_{t\geq 0},$$

   où $r > 1$ et $\lambda > 0$. Soit X le nombre de sinistre dans une année d'un assuré pris au hasard. Montrer que sa distribution est celle d'une loi **binomiale négative** dont on précisera les paramètres.
   (Y suit une loi binomiale négative de paramètre (a, p) si, pour tout $n \in \mathbb{N}$

$$\mathbb{P}(Y = n) = \frac{\Gamma(a+n)}{\Gamma(a)n!}p^a(1-p)^n.$$

4. Déduire de la question précédente l'espérance d'une loi binomiale négative.
5. Proposer des estimateurs naturels de r et λ à partir d'un échantillon $(X_1, ..., X_n)$ i.i.d.