Feuille 4 : vers les statistiques inférentielles

Exercice 1 (Pile ou face)
On joue à pile ou face avec une pièce de monnaie dont on cherche à savoir si elle est équilibrée ou non. On note $p \in [0,1]$ la probabilité (inconnue) de tomber sur "pile". On suppose que les résultats des différents lancers sont indépendants.

1. *On joue n fois de suite, quelle est la loi du nombre de "pile" obtenus ? Décrire le modèle statistique sous-jacent, i.e. le triplet espace d'observation, tribu, famille de probabilités. Dans ce modèle, comment peut-on estimer le paramètre p ?*
2. *Soit T le nombre d'essais nécessaires avant de tomber sur "pile". Quelle est la loi de T ? Décrire à nouveau le modèle statistique relatif à l'expérience, i.e. le nouvel espace d'observation, la tribu, la famille de probabilités. Dans ce nouveau modèle, comment peut-on estimer p ?*
3. *Chacun des modèles ci-dessus est-il identifiable ?*

Exercice 2 (Fiabilité)
Un circuit électrique est composé de deux types de diodes A et B montées en série. Les durées de vie des diodes sont supposées indépendantes, de lois exponentielles de paramètres inconnus $\lambda_A > 0$ et $\lambda_B > 0$.

1. *Quelle est la loi suivie par la durée de vie du circuit en série ?*
2. *On observe les durées de vie de n circuits indépendants de ce type. Quel est le modèle statistique associé à cette expérience ?*
3. *Dans ce modèle, comment estimer le paramètre $\lambda_A + \lambda_B$?*
4. *Quelle expérience faudrait-il faire pour estimer λ_A et λ_B séparément ?*

Exercice 3 (Identifiabilité)
Décrire l'expérience statistique et étudier l'identifiabilité des modèles lorsque que l'on observe n réalisations indépendantes (X_i) des lois suivantes.

1. *les (X_i) suivent la loi uniforme dans l'intervalle $[-\alpha^2, +\alpha^2]$, avec $\alpha \in \mathbb{R}^*$.*
2. *les (X_i) suivent la loi uniforme dans $\{0, 1, ..., N\}$, $N \in \mathbb{N}^*$.*
3. *les (X_i) suivent la loi normale $\mathcal{N}(\nu, \sigma^2)$, $(\nu, \sigma^2) \in \mathbb{R} \times \mathbb{R}_+^*$.*

Exercice 4 (Statistiques)
On dispose de réalisations de v.a. i.i.d. $X_1, \ldots, X_n$. Parmi les variables aléatoires suivantes, lesquelles sont des statistiques ?

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i, \quad \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2, \quad S := n\hat{\sigma}^2/\sigma^2, \quad T = \sqrt{n-1}\left(\bar{X} - m\right)/\hat{\sigma}.$$

Exercice 5 (Échantillon gaussien)
On observe $Y_1, Y_2, ..., Y_n$ n v.a. i.i.d. de loi $\mathcal{N}(\nu, \sigma^2)$. On suppose que la variance σ^2 est connue mais que la moyenne ν est inconnue.

1. *Décrire le modèle statistique sous-jacent sous la forme d'un modèle linéaire.*
2. *Proposer un estimateur de ν et étudier ses propriétés asymptotiques (limite en probabilité, dans $\mathbb{L}^2$,etc.)*

Exercice 6 (Modèle linéraire)
Soient $\epsilon_1, \epsilon_2, ..., \epsilon_4$ *des v.a. i.i.d. de loi* $\mathcal{N}(0,1)$ *et soient* $\nu = (\nu_1, \ldots, \nu_4) \in \mathbb{R}^4$ *et* $\sigma \in \mathbb{R}_+^*$ *des paramètres. On suppose qu'on observe*

$$Y_1 = \nu_1 + \sigma\epsilon_1, \quad Y_2 = \nu_2 + \frac{\sigma}{\sqrt{3}}\epsilon_2, \quad Y_3 = \nu_3 + \sigma\epsilon_3, \quad Y_4 = \nu_4 + \frac{\sigma}{\sqrt{2}}\epsilon_4,$$

Écrire la représentation linéaire du modèle, autrement dit écrire le vecteur Y *sous la forme* $Y = \nu + \gamma\epsilon$ *où* $\epsilon \sim \mathcal{N}(0, Id_4)$ *et* γ *est une matrice appropriée. Quelle est la loi du vecteur* Y *?*

Exercice 7 (Moyenne mouvante)
On considère deux variables aléatoires réelles indépendantes X *et* R *définies sur l'espace probabilisé* $(\Omega, \mathcal{A}, \mathbb{P})$*, telles que* R *est une variable aléatoire de Rademacher, i.e.* $\mathbb{P}(R = 1) = \mathbb{P}(R = -1) = 1/2$*, et* X *suit la loi normale* $\mathcal{N}(0,1)$*. Pour* $\theta \in \mathbb{R}$*, on note* $Y_\theta = \theta R + X$*.*

1. *Déterminer la loi de la variable aléatoire* Y_θ*. Esquisser sa densité.*
2. *Soit une expérience dans laquelle on observe* n *réalisations indépendantes de la loi de* Y_θ*, avec* θ *inconnu. Préciser le modèle statistique sous-jacent. Est-il identifiable ?*
3. *On suppose ici que* $\theta \in \mathbb{R}^+$*. Comment peut-on estimer le paramètre* θ *?*

Exercice 8 (Estimation d'un minimum)
Soient $X_1, X_2, ..., X_n$ n *v.a. i.i.d. de loi* $\mathbb{P}_\theta(dx) = e^{\theta - x}\mathbb{1}_{\{[\theta,\infty[\}}(x)dx$, $\theta \in \mathbb{R}_+^*$.

1. *Décrire le modèle statistique associé aux observations. Est-il identifiable ?*
2. *Calculer* $\mathbb{E}_\theta(X_1)$ *et en déduire un estimateur de* θ *que l'on notera* $\hat{\theta}_n$*.*
3. *Étudier le comportement de* $\mathbb{E}_\theta((\hat{\theta}_n - \theta)^2)$ *en fonction de* n*.*
4. *Trouver un meilleur estimateur de* θ*, en quoi est-il meilleur ?*

Exercice 9 (Simulation)
On suppose que l'on sait simuler une variable U *de loi uniforme sur* $[0,1]$*. Le but de l'exercice est de simuler une variable de loi quelconque sur* $\mathbb{R}$*.*

1. *Expliciter une méthode pour simuler une loi discrète à support fini.*
2. *Soit* X *est variable aléatoire réelle de fonction de répartition* F*. On définit l'inverse généralisée de* F *par* $F^{-1}(x) := \inf\{t \in \mathbb{R}, F(t) > x\}$*. Quelle est la loi de* $F^{-1}(U)$ *?*
3. *Si* X *est une variable de fonction de répartition* F *continue, quelle est la loi de* $F(X)$ *?*