Corrigé : Feuille de travaux dirigés 1

Solution Exercice 1

1. On observe n réplications $X_1, \ldots, X_n$ de la v.a. à valeurs dans $\Omega = \mathbb{R}$

$$X = \mu + \delta + \epsilon,$$

où $\delta = 0.1$, μ est le paramètre d'intérêt, ϵ représente l'erreur de mesure, de variance supposée connue $\sigma > 0$. L'espace des observations est $\mathcal{X} = \mathbb{R}$. Un échantillon de taille n est un vecteur de $\mathcal{X}^n$, le modèle pour une observation X ($n = 1$) est

$$\mathcal{P} = \{\mathrm{P} : \mathbb{E}_{\mathrm{P}}(X) = \delta + \mu \ , \mathbb{V}\mathrm{ar}_{\mathrm{P}}(X) = \sigma^2, \mu \in \mathbb{R}\},$$

avec σ^2 et δ connus, μ inconnu. Les X_n sont i.i.d., comme les bruits de mesure associés $\epsilon_1, \ldots, \epsilon_n$. Le modèle pour un échantillon i.i.d.de taille n est

$$\mathcal{P}_n = \{\mathrm{P}^{\otimes n} : \mathrm{P} \in \mathcal{P}\},$$

voir le poly pour un rappel sur les lois produits.

2. Le modèle statistique est non paramétrique dans la mesure où les paramètres (inconnus et connus) ne caractérisent pas la distribution de l'observation X : l'ensemble de toutes les lois de probabilité de variance donnée ne peut pas être paramétré par un ouvert d'un espace de dimension finie. Par exemple à paramètre $\mu > 0$ fixé, loi des observations est de moyenne $\mu + 0.1$ et de variance σ^2, comme le sont par exemple les lois $\mathcal{N}(\mu + 0.1;\ \sigma^2)$ et $\Gamma((\mu + 0.1)^2/\sigma^2, \sigma/(\mu + 0.1))$.
3. Lorsque le biais δ est connu, le paramètre μ est identifiable. En effet, deux lois identiques ont même moyenne, donc si P_1 et P_2 sont deux lois de paramètres respectifs μ_1 et μ_2 avec $\mu_1 \neq \mu_2$, alors on a $\mathbb{E}_{\mathrm{P}_1}(X) = \mu_1 + \delta \neq \mu_2 + \delta = \mathbb{E}_{\mathrm{P}_2}(X)$, donc $\mathrm{P}_1 \neq \mathrm{P}_2$.
4. Si le biais est inconnu, on considère alors un couple de paramètres (μ, δ). Le modèle devient
$$\mathcal{P} = \{\mathrm{P} : \mathbb{E}_{\mathrm{P}}(X) = \delta + \mu \ , \mathbb{V}\mathrm{ar}_{\mathrm{P}}(X) = \sigma^2, \mu \in \mathbb{R}, \delta \in \mathbb{R}.\},$$
Le couple n'est alors pas identifiable, toutes les valeurs appartenant à la droite $\mu + \delta = c$, pour une constante c donnée définissent la même loi de probabilité. Par contre, si le biais est connu mais pas σ^2, on vérifie que le vecteur de paramètre (μ, σ^2) est identifiable en utilisant le fait que deux lois égales ont même espérance et même variance.

Solution Exercice 2

1. Le modèle statistique relatif à l'observation X, à valeurs dans $\mathcal{X} = \{0, 1\}$, s'écrit
$$(\{0, 1\}, \mathcal{P}(\{0, 1\}), Ber(q),\ q \in (0, 1))\,.$$
L'espace des actions est $\mathcal{A} = \{a_0,\ a_1\}$. Il y a $(\#\mathcal{A})^{\#\mathcal{X}}$ règles de décisions $\delta : \mathcal{X} \to \mathcal{A}$ possibles.

2. On définit $\delta_1(x) \equiv a_0$, $\delta_2(x) \equiv a_1$, $\delta_3(x) = a_0\mathbb{I}\{x=0\} + a_1\mathbb{I}\{x=1\}$ et $\delta_4(x) = a_0\mathbb{I}\{x=1\} + a_1\mathbb{I}\{x=0\}$. En utilisant la matrice de coût $(C(\theta_i, a_i))_{i=0,\,1}$n vérifie que

$$\begin{aligned}
R(\theta_0, \delta_1) &= C(\theta_0, a_0) = 100 \text{ et } R(\theta_1, \delta_1) = C(\theta_1, a_0) = 100, \\
R(\theta_0, \delta_2) &= C(\theta_0, a_1) = 200 \text{ et } R(\theta_1, \delta_2) = C(\theta_1, a_1) = 0, \\
R(\theta_0, \delta_3) &= C(\theta_0, a_0)(1-q) + C(\theta_0, a_1)q = 120 \text{ et} \\
&\quad R(\theta_1, \delta_3) = C(\theta_1, a_0)(1-p) + C(\theta_1, a_1)p = 20, \\
R(\theta_0, \delta_4) &= C(\theta_0, a_0)q + C(\theta_0, a_1)(1-q) = 180 \text{ et} \\
&\quad R(\theta_1, \delta_4) = C(\theta_1, a_0)p + C(\theta_1, a_1)(1-p) = 80.
\end{aligned}$$

3. Le risque maximum est 100 pour la règle δ_1 (obtenu que le paramètre vaille θ_0 ou θ_1), 200 pour δ_2 (lorsque $\theta = \theta_0$), 120 pour δ_3 (lorsque $\theta = \theta_0$) et 180 pour δ_4 (lorsque $\theta = \theta_0$). La règle minimisant le risque maximum est donc δ_1, consistant à ne jamais forer, quelque soit la valeur observée pour X.

4. Si l'on dispose d'une information *a priori* (*i.e.* avant l'observation de X) sur la probabilité d'occurence des valeurs du paramètre θ, il est naturel de ne considérer le risque associé à une valeur que pondéré par la probabilité de se trouver dans l'état décrit par cette valeur. Ici, $\mathbb{P}\{\theta = \theta_0\} = \mathbb{P}\{\theta = \theta_1\} = 1/2$. Un critère naturel de risque pour une règle δ devient alors :

$$\rho(\delta) = R(\theta_0, \delta)/2 + R(\theta_1, \delta)/2.$$

Le risque Bayesien des règles envisageables est donné par : $\rho(\delta_1) = 100 = \rho(\delta_2)$, $\rho(\delta_3) = 70$ et $\rho(\delta_4) = 130$. La règle de moindre risque de Bayes est alors δ_3.

Solution Exercice 3

1. L'observation peut s'écrire de façon vectorielle $Y = \theta_1 A + V$ où $A = (x_1, \ldots, x_n)$ est un vecteur déterministe et $V = (V_1, \ldots, V_n)$ est un vecteur Gaussien (ses composantes sont des Gaussiennes i.i.d. $\mathcal{N}(0, \sigma^2)$) centré de matrice de covariance $\sigma^2 I_n$, I_n désignant la matrice $n \times n$ unité. La vraisemblance du modèle statistique (dominé par la mesure de Lebesgue) s'écrit donc :

$$p_\theta(Y) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^{n} (Y_i - x_i\theta_1)^2 \right).$$

2. On maximise la log-vraisemblance. Les equations de score s'écrivent :

$$\begin{aligned}
0 &= \frac{\partial}{\partial \sigma^2} \log p_\theta(Y) = -\frac{n}{2}\frac{1}{\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^{n} (Y_i - \theta_1 x_i)^2, \\
0 &= \frac{\partial}{\partial \theta_1} \log p_\theta(Y) = \frac{1}{\sigma^2} \sum_{i=1}^{n} x_i (Y_i - \theta_1 x_i).
\end{aligned}$$

La solution (on vérifiera qu'il s'agit d'un maximum global en calculant la hessienne) est donc :

$$
\begin{aligned}
\widehat{\theta}_1 &= \frac{\sum_{i=1}^n Y_i x_i}{\sum_{i=1}^n x_i^2}, \\
\widehat{\sigma}^2 &= \frac{1}{n}\sum_{i=1}^n (Y_i - \widehat{\theta}_1 x_i)^2.
\end{aligned}
$$