Feuille de travaux dirigés 1

Exercice 1 (Modèle statistique, identifiabilité)**:**
Un instrument est utilisé pour effectuer n mesures d'une grandeur physique constante μ. L'instrument est biaisé d'une quantité positive δ connue ($\delta = 0.1$). On sait que les erreurs de mesures sont indépendantes et on connait leur variance $\sigma^2 > 0$.

1. Écrire formellement le problème statistique : identifier l'espace des observations, le modèle statistique en jeu (c'est-à-dire, décrire l'ensemble des lois possibles des observations).
2. Le modèle est-il paramétrique ?
3. Le paramètre d'intérêt μ est-il identifiable ?
4. Même question en supposant le biais δ inconnu. Qu'en est-il si, à l'inverse, on connait le biais δ mais pas la variance σ^2 ?

Exercice 2 (Risque, pire des cas, risque intégré)**:**
On reprend l'exemple de la prospection pétrolière donné en cours (fin du chapitre 1 du poly), dans un cadre simplifié : on considère qu'il y a deux actions possibles : forer (a_1) ou ne pas forer (a_0).

Les deux états possibles de la nature sont θ_1 (présence de pétrole) et θ_0 (absence de pétrole). La fonction de coût est donnée par le tableau suivant :

	a_0	a_1
θ_0	100	200
θ_1	100	0

TABLE 1 – Fonction de coût d'un forage pétrolier

Pour obtenir une information sur θ, on réalise un forage préliminaire, produisant une donnée $X \in \{0, 1\}$ de nature aléatoire, telle que $\mathbb{P}_{\theta_1}(X = 1) = p$, $\mathbb{P}_{\theta_1}(X = 0) = 1 - p$; $\mathbb{P}_{\theta_0}(X = 1) = q$, $\mathbb{P}_{\theta_0}(X = 0) = 1 - q$; avec $p = 1 - q = 0.8$.

1. Préciser le modèle statistique, l'espace des observations $\mathcal{X}$ et l'espace des actions. Combien y a-t-il de règles de décisions $\delta : \mathcal{X} \to \mathcal{A}$?
2. Calculer les points de risque $R(\theta_0, \delta), R(\theta_1, \delta)$ pour chaque règle de décision et représenter ces points dans le plan ($R(\theta_0, \delta)$ en abscisse, $R(\theta_1, \delta)$ en ordonnée).
3. Un exploitant souhaite adoper une stratégie lui assurant de perdre le moins d'argent (en moyenne) dans le pire des cas (le θ le moins favorable). Il cherche donc la règle δ^* minimisant le risque « dans le pire des cas »,

$$\text{Pire risque } (\delta) := \max_\theta R(\theta, \delta^*),$$

c'est-à-dire la stratégie dite *minimax*

$$\delta^* = \operatorname*{argmin}_\delta \max_\theta R(\theta, \delta).$$

Déterminer la règle minimax dans cet exemple.

4. *Pour aller plus loin :* on dispose de l'information suivante : la moitié des terrains candidats à l'exploitation contient effectivement du pétrole. La stratégie minimax vous paraît-elle raisonnable ? Proposer un critère $\rho(\delta)$ représentant le risque moyen (sur l'ensemble des θ possibles, pondérés selon l'information a priori dont on dispose). Quell est maintenant la fonction de décision $\delta' = \operatorname{argmin}_\delta \rho(\delta)$ qui minimise le critère $\rho(\delta)$?
   N.B : Ceci est un premier exemple illustratif de l'approche Bayésienne. On verra plus tard que ρ est appelé « risque intégré » et sa valeur minimale $\rho(\delta')$ est le « risque de Bayes ».

Exercice 3 (Maximum de vraisemblance pour le modèle linéaire simple)**:**
Pour tout $i = 1, \dots, n$, on considère

$$X_i = a_i \theta_1 + Z_i \tag{1}$$

où les Z_i sont des variables aléatoires indépendantes, de même loi $\mathcal{N}(0, \sigma^2)$. Les coefficients a_i sont des variables déterministes connues (et non toutes nulles). Les paramètres θ_1 et σ^2 sont inconnus. On observe $X = (X_1, \dots, X_n)^\top$. On paramètre le modèle pour X par $\theta = (\theta_1, \sigma^2) \in \mathbb{R} \times \mathbb{R}^+$.

1. Donner la densité p_θ de X par rapport à la mesure de Lebesgue en fonction de $a = (a_1, \dots, a_n)^\top$ et de σ^2.
2. Exprimer l'estimateur $\widehat{\theta}$ du maximum de vraisemblance pour le paramètre θ.
3. Montrer que $\widehat{\theta}_1$ (la première coordonnée de $\widehat{\theta}$) est non biaisé.