V. Méthodes de contraste (maximum de vraisemblance et moindres carrés)

V.1. Exercice. On considère l'expérience statistique

$$\mathcal{E} = (\{0,1\}, \mathcal{P}(\{0,1\}), (P_\theta)_{\theta \in \{\frac{1}{10}, \frac{8}{10}\}}),$$

$$\text{où} \quad P_\theta(dx) = \theta \delta_{\{0\}}(dx) + (1-\theta)\delta_{\{1\}}(dx).$$

(1) Montrer que le modèle est identifiable, dominé, calculer sa fonction de vraisemblance.
(2) En déduire l'estimateur du maximum de vraisemblance.

V.2. Exercice. Calculer l'estimateur du maximum de vraisemblance (s'il existe et s'il est bien défini) et étudier ses propriétés (vitesse de convergence, intervalle de confiance, loi limite, lien avec un estimateur par substitution) dans les cas suivants :

(1) On observe $(X_1, \ldots, X_n)$ i.i.d. de loi de Poisson de paramètre $\lambda > 0$.
(2) On observe $X \sim \mathcal{B}(n,p)$, où $p \in]0,1[$.
(3) On observe $(X_1, \ldots, X_n)$ i.i.d. de loi $\mathcal{N}(\mu, \sigma^2)$, où $\mu \in \mathbb{R}$ et $\sigma > 0$.
(4) On observe $(X_1, \ldots, X_n)$ i.i.d. de loi uniforme sur $[0, \theta]$, où $\theta \in \mathbb{R}_+^*$.

V.3. Exercice*. Soient $X_1, \ldots, X_n$ un n-échantillon de la loi uniforme sur $[0,\theta]$, avec $\theta > 0$.

(1) Décrire le modèle statistique associé.
(2) Proposer un estimateur $\hat{\theta}_n$ de θ obtenu par méthode des moments. Etudier sa loi limite.
(3) Soit T_n l'estimateur du maximum de vraisemblance de θ. Montrer que pour tout réel t

$$P_\theta^n\{n(T_n - \theta) \leq t\} \to e^{t/\theta} 1_{\{t \leq 0\}} + 1_{\{t > 0\}}$$

quand n tend vers l'infini. Interprétation.

V.4. Exercice. Soit $c > 0$ un paramètre fixé connu. On considère la loi de Weibull P_λ, $\lambda > 0$, de densité

$$\lambda c x^{c-1} e^{-\lambda x^c} 1_{]0,+\infty[}(x)$$

par rapport à la mesure de Lebesgue. On observe un n-échantillon de la loi P_λ.

(1) Calculer l'estimateur du maximum de vraisemblance $\hat{\lambda}_n$ de λ.
(2) Calculer son risque quadratique $\mathrm{E}_\lambda^n\{(\hat{\lambda}_n - \lambda)^2\}$ où P_λ^n désigne la loi du n-échantillon.

V.5. Exercice. On considère pour $\mu > 0$, $\alpha > 0$, la loi de probabilité $P_{\mu,\alpha}$ dont la fonction de répartition est continue sur $\mathbb{R}$ et vaut : $P_{\mu,\alpha}(X \leq x) = 1 - Cx^{-\alpha}$ si $x \geq \mu$ et 0 sinon.

(1) Calculer la constante C.
(2) Calculer l'estimateur du maximum de vraisemblance $(\hat{\mu}_n, \hat{\alpha}_n)$ de (μ, α).
(3) Quelle est la loi limite de $n(\hat{\mu}_n - \mu)$?
(4) Montrer que

$$\sqrt{n}\Big(\hat{\alpha}_n^{-1} - \frac{1}{n}\sum_{i=1}^n \log(X_i/\mu)\Big) \xrightarrow{P} 0$$

puis en déduire la loi limite de $\sqrt{n}(\hat{\alpha}_n - \alpha)$.

V.6. **Exercice*.** Soit $g : [0,1] \to]0,+\infty[$ une densité continue connue. Pour $n \geq 1$, on observe $(X_1^n, \ldots, X_n^n)$, où

$$X_i^n = \theta + \frac{1}{\sqrt{g(i/n)}} \varepsilon_i^n$$

et les ε_i^n sont i.i.d. de loi $\mathcal{N}(0, \sigma^2)$, avec $\sigma^2 > 0$ connue.

(1) Ecrire le modèle statistique correspondant. Montrer qu'il est identifiable et dominé.
(2) Calculer l'estimateur du maximum de vraisemblance $\hat{\theta}_n$ de θ .
(3) Calculer le risque quadratique de $\hat{\theta}_n$ et sa limite quand $n \to \infty$.

V.7. **Exercice.** On suppose que la durée de vie discrétisée T d'un appareil suit une loi géométrique de paramètre $\theta \in]0,1[$. On arrête l'observation d'un tel appareil à un instant r arbitraire et on pose

$$Y = T 1_{\{T \leq r\}} + (r+1) 1_{\{T \geq r+1\}}.$$

(1) Quelle est la loi de Y ?
(2) On observe un n-échantillon de la loi de Y. Décrire l'expérience associée, calculer sa vraisemblance. On notera M le nombre (aléatoire) de Y_i égaux à $r+1$.
(3) Déterminer l'estimateur du maximum de vraisemblance de θ.
(4) Etudier son comportement asymptotique.

V.8. **Exercice*.** Soient Y_1, Y_2, Y_3, Y_4 des variables aléatoires gaussiennes indépendantes telles que $\mathrm{E}\{Y_1\} = \mathrm{E}\{Y_2\} = \mu_1$, $\mathrm{E}\{Y_3\} = \mu_2$, $\mathrm{E}\{Y_4\} = \mu_3$, $\mathrm{Var}(Y_1) = \mathrm{Var}(Y_3) = \sigma^2$, $\mathrm{Var}(Y_2) = \sigma^2/3$ $\mathrm{Var}(Y_4) = \sigma^2/2$; μ_1, μ_2, μ_3 et σ^2 sont inconnus et $\sigma^2 > 0$.

(1) Ecrire le modèle linéaire correspondant.
(2) Déterminer les estimateurs des moindres carrés $\hat{\mu_1}, \hat{\mu_2}, \hat{\mu_3}$ et $\hat{\sigma^2}$. Quelle est la loi de $\hat{\mu_1}$?
(3) Déterminer un intervalle de confiance pour μ_1 de niveau α.
(4) On suppose maintenant que μ_2 et μ_3 sont égaux à μ_1.
   (a) Quel est le modèle linéaire associé?
   (b) Déterminer l'estimateur $\tilde{\mu}_1$ des moindres carrés de μ_1. Comparer les performances de $\hat{\mu}_1$ et de $\tilde{\mu}_1$.
(5) On utilise $\tilde{\mu}_1$ pour estimer μ_1 dans le premier modèle.
   (a) Calculer son erreur quadratique moyenne.
   (b) Montrer que si $|\mu_2 + 2\mu_3 - 3\mu_1|$ est inférieur à un seuil que l'on calculera, l'erreur quadratique moyenne de $\tilde{\mu}_1$ est inférieure à celle de $\hat{\mu}_1$.

V.9. **Exercice.** On considère le modèle de régression linéaire

$$Y_i = b_0 + b_1 x_i + \varepsilon_i, \quad i = 1, \ldots, n$$

où les ε_i sont des variables aléatoires indépendantes $\mathcal{N}(0, \sigma^2)$. b_0, b_1 et σ^2 sont inconnus.

(1) Quels sont les estimateurs des moindres carrés ordinaires $\hat{b}_0$, $\hat{b}_1$ et $\hat{\sigma}^2$ de ces paramètres ? Quelle est la loi du couple $((\hat{b}_0, \hat{b}_1), \hat{\sigma}^2)$?

(2) On dispose d'une observation y_0 sur une unité statistique pour laquelle la valeur de x_0 de la variable explicative est inconnue On suppose que y_0 est la réalisation d'une variable Y_0 s'écrivant

$$Y_0 = b_0 + b_1 x_0 + \eta,$$

où η est une variable aléatoire $\mathcal{N}(0, \sigma^2)$ indépendante du vecteur $(\varepsilon_1, \dots, \varepsilon_n)$. On cherche un intervalle de confiance pour x_0. On fera l'hypothèse supplémentaire que

$$|x_0 - \bar{x}| \leq 1 \quad \text{où } \bar{x} = n^{-1} \sum x_i.$$

(a) Quelle est la loi de $Y_0 - \hat{b}_0 - \hat{b}_1 x_0$?

(b) En utilisant l'estimateur $\hat{\sigma}$ de σ, déterminer un intervalle de confiance I_1 de niveau α pour x_0.

(3) On dispose maintenant de m observations $y_{01}, \dots, y_{0m}$ correspondant à la valeur x_0 inconnue; ce sont des observations de m variables aléatoires telles que

$$Y_{0j} = b_0 + b_1 x_0 + \eta_j,$$

où $(\eta_1, \dots, \eta_m)$ et $(\varepsilon_1, \dots, \varepsilon_n)$ sont indépendantes, et η_j suit la loi $\mathcal{N}(0, \sigma^2)$.

(a) Montrer que

$$\widetilde{\sigma}^2 = \frac{(n-2)\hat{\sigma}^2 + \sum_{j=1}^m (Y_{0j} - \bar{Y}_0)^2}{n+m-3},$$

où $\bar{Y}_0 = \frac{1}{m} \sum_{j=1}^m Y_{0j}$, est un autre estimateur sans biais de σ^2. Quelle est sa loi ?

(b) Quelle est la loi de $\bar{Y}_0 - \hat{b}_0 - \hat{b}_1 x_0$?

(c) A l'aide de $\widetilde{\sigma}^2$ et de $\bar{Y}_0$ donner un intervalle de confiance I_2 pour x_0 de niveau α.

(d) Aurait-on pu construire un intervalle de confiance I_3 pour x_0 à l'aide de $\hat{\sigma}^2$ et de $\bar{Y}_0$?

V.10. Exercice*. On considère le modèle suivant

$$Y = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix} \sim \mathcal{N}\left(X\beta, \sigma^2 S\right)$$

avec

$$X = \begin{pmatrix} x_{1,1} & \dots & x_{1,p} \\ \vdots & \ddots & \vdots \\ x_{n,1} & \dots & x_{n,p} \end{pmatrix}, \beta = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_p \end{pmatrix} \text{ et } S = \begin{pmatrix} f(x_{1,1}) & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & f(x_{n,1}) \end{pmatrix}.$$

On suppose que l'on observe Y, que la matrice X et la fonction f sont connues, et l'on souhaite estimer les paramètres inconnus β et σ^2. Un tel modèle est dit "modèle linéaire hétéroscédastique", par opposition au modèle linéaire défini dans le cours, avec $S = I_n$, appelé "modèle linéaire homoscédastique".

(1) Construire une matrice M telle que $Z = MY$ suive un modèle linéaire homoscédastique.

(2) En déduire un estimateur $\hat{\beta}$ de β en utilisant la procédure des moindres carrés sur Z.

(3) On peut aussi construire l'estimateur des moindres carrés directement sur Y,

$$\widetilde{\beta} = \arg\min_{b \in \mathbb{R}^p} \|Y - Xb\|^2.$$

Comparer $\mathrm{E}(\|\hat{\beta} - \beta\|^2)$ et $\mathrm{E}(\|\widetilde{\beta} - \beta\|^2)$, conclure.

V.11. Exercice. Soit $\mathcal{E} = (\mathbb{R}^n, (P_\theta)_{\theta \in \Theta})$ une modèle statistique où $\Theta = \{\theta_1, \theta_2, \ldots, \theta_d\}$ est une ensemble finie de cardinal d et pour chaque θ la probabilité P_θ admet une densité $L_\theta : \mathbb{R}^n \to \mathbb{R}_+$. On se donne une probabilité π sur Θ c'est-à-dire un d-uplet $(\pi(\theta_1), \ldots, \pi(\theta_d)) \in [0,1]^d$ tel que $\sum_{i=1}^d \pi(\theta_i) = 1$.

(1) Si $\hat{\theta}$ est un estimateur de θ, à valeurs dans Θ, montrer que

$$\sum_{i=1}^d \pi(\theta_i) P_{\theta_i}(\hat{\theta} = \theta_i) = \int_{\mathbb{R}^n} \pi(\hat{\theta}(z)) L_{\hat{\theta}(z)}(z) dz.$$

(2) En déduire un estimateur θ^* de θ qui rend minimal le risque de classification suivant

$$R(\hat{\theta}) = \sum_{i=1}^d \pi(\theta_i) P_{\theta_i}(\hat{\theta} \neq \theta_i).$$

(3) Montrer que

$$R(\theta^*) = 1 - \int_{\mathbb{R}^n} \max_{1 \le i \le d} \big(\pi(\theta_i) L_{\theta_i}(z)\big)\, dz \le 1 - \max_{1 \le i \le d} \pi(\theta_i) \le 1 - 1/d.$$

(4) Dans le cas où $\Theta = \{\theta_1, \theta_2\} \subset \mathbb{R}$, $\pi(\theta_1) = \pi(\theta_2)$ et P_θ est la loi d'un n-échantillon gaussien de la loi $\mathcal{N}(\theta, 1)$, montrer que

$$P_{\theta_1}(\theta^* \neq \theta_1) = P_{\theta_2}(\theta^* \neq \theta_2) = P(Z > \sqrt{n}|\theta_1 - \theta_2|/2)$$

où $Z \sim \mathcal{N}(0,1)$.

V.12. Exercice. Dans une urne contenant 1000 tickets, 20 sont marqués θ et 980 sont marqués 10 θ, où θ est un entier naturel inconnu.

(1) Donner l'estimateur du maximum de vraisemblance $\hat{\theta}$ de θ lorsque l'on tire un unique ticket de valeur X, et montrer que $\mathbb{P}(\hat{\theta} = \theta) \ge 0,98$.

(2) On renumérote les tickets marqués 10θ par $a_i\theta$, $1 \le i \le 980$ où les a_i sont des réels connus, deux-à-deux distincts, et compris dans l'intervalle $[10, 10.1]$. Donner le nouvel estimateur du maximum de vraisemblance $\widetilde{\theta}$ et montrer que $\mathbb{P}(\widetilde{\theta} < 10\theta) = 0,02$. Ce résultat vous semble-t-il paradoxal?