Dans ce sujet, par convention un vecteur x est un vecteur colonne. On notera $x^\top$ la transposée du vecteur (ou de la matrice) x. Pour $x \in \mathbb{R}^n$, $\|x\|$ désigne la norme euclidienne : $\|x\|^2 := x^\top x$.
$\mathbb{N}^*$ désigne l'ensemble des entiers naturels strictement positifs ; $\mathbb{R}_+$ est l'ensemble des réels positifs ou nuls et $\mathbb{R}_+^*$ est l'ensemble des réels strictement positifs.
Pour $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$, on pose $x_{n:n} := \max(x_1, \ldots, x_n)$.
Une loi Gamma de paramètres (p, λ) (définition IV.3-12 du polycopié) a pour espérance p/λ et pour variance p/λ^2.

Exercice 1. Soient $n \in \mathbb{N}$, $n \geq 3$ et $0 \leq t_1 < \cdots < t_n$ des instants connus. Nous définissons les vecteurs $\mathbf{1}$, $\mathbf{x}$ et la matrice $\mathbf{X}$ de taille $n \times 2$ par

$$\mathbf{1} := \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \in \mathbb{R}^n, \qquad \mathbf{x} := \begin{bmatrix} t_1 \\ \vdots \\ t_n \end{bmatrix} \in \mathbb{R}^n, \qquad \mathbf{X} := [\mathbf{1}, \mathbf{x}] \in \mathbb{R}^{n \times 2} .$$

Les hypothèses sur les composantes de $\mathbf{x}$ entraînent que $\mathbf{X}$ est de rang 2 et par suite, $\mathbf{X}^\top \mathbf{X}$ est inversible. Soit $\mathbf{Y} \in \mathbb{R}^n$ une observation du modèle statistique

$$(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \{\mathrm{N}(\mathbf{X}\boldsymbol{\beta}, \sigma^2 \mathrm{I}_n), (\boldsymbol{\beta}, \sigma^2) \in \Theta := \mathbb{R}^2 \times \mathbb{R}_+^*\}) , \qquad \mathbf{Y} = \begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} .$$

Pour $\theta = (\boldsymbol{\beta}, \sigma^2) \in \Theta$, nous notons $\mathrm{L}_n(\theta, \mathbf{Y})$ la vraisemblance de l'observation $\mathbf{Y}$, donnée par

$$\mathrm{L}_n(\theta, \mathbf{Y}) = (2\pi\sigma^2)^{-n/2} \exp(-\|\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}\|^2 / 2\sigma^2) .$$

Soit $\Theta_0 := \{(\beta_1, \beta_2, \sigma^2) \in \Theta, \beta_2 = 0\}$.

1. Montrer [en justifiant soigneusement le résultat] que l'estimateur du maximum de vraisemblance $(\hat{\boldsymbol{\beta}}, \widehat{\sigma}^2)$ de $(\boldsymbol{\beta}, \sigma^2)$ est donné par
$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{Y} , \qquad \hat{\sigma}^2 = n^{-1} \|\mathbf{Y} - \Pi \mathbf{Y}\|^2 , \tag{1}$$
où $\Pi := \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top$ est le projecteur orthogonal sur l'espace vectoriel engendré par les colonnes de $\mathbf{X}$.
2. En déduire que
$$\sup_{\theta \in \Theta} \mathrm{L}_n(\theta, \mathbf{Y}) = (2\pi\widehat{\sigma}^2)^{-n/2} \mathrm{e}^{-n/2} .$$
3. Montrer que $\sup_{\theta \in \Theta_0} \mathrm{L}_n(\theta, \mathbf{Y}) = \left(2\pi\tilde{\sigma}^2\right)^{-n/2} \mathrm{e}^{-n/2}$ où $\tilde{\sigma}^2 := n^{-1} \|\mathbf{Y} - \Pi_0 \mathbf{Y}\|^2$ avec $\Pi_0 = n^{-1} \mathbf{1}\mathbf{1}^\top$.
4. Montrer que
$$\tilde{\sigma}^2 = \left(\widehat{\sigma}^2 + n^{-1} \mathbf{Y}^\top \{\Pi - \Pi_0\} \mathbf{Y}\right)$$
où $\hat{\sigma}^2$ est l'estimateur défini par (1).

Nous considérons le test

$$H_0 : \ \theta \in \Theta_0, \quad \text{contre} \quad H_1 : \ \theta \notin \Theta_0. \tag{2}$$

Nous étudions le test du Rapport de Vraisemblance Généralisée (RVG) dont la statistique est donnée par

$$\Lambda(\mathbf{Y}) := \frac{\sup_{\theta\in\Theta_0} \mathrm{L}_n(\theta,\mathbf{Y})}{\sup_{\theta\in\Theta} \mathrm{L}_n(\theta,\mathbf{Y})} \ .$$

Pour $\alpha \in]0,1[$, nous considérons le test de fonction critique

$$\phi_\alpha(\mathbf{Y}) := \mathbb{1}_{[0,c_\alpha]}(\Lambda(\mathbf{Y})) \ ,$$

où $c_\alpha \geq 0$ vérifie : $\sup_{\theta\in\Theta_0} \mathbb{P}_\theta\left(\phi_\alpha(\mathbf{Y})=1\right) \leq \alpha$.

5. Montrer qu'il existe $k_\alpha \geq 0$ tel que le test RVG ϕ_α est équivalent au test de fonction critique $\psi_\alpha(\mathbf{Y}) := \mathbb{1}_{[k_\alpha,\infty[}(\mathrm{F}(\mathbf{Y}))$ où
$$\mathrm{F}(\mathbf{Y}) := (n-2)\frac{\mathbf{Y}^\top\{\Pi - n^{-1}\mathbf{1}\mathbf{1}^\top\}\mathbf{Y}}{n\hat\sigma^2} \ .$$
6. Montrer que, pour tout $\theta \in \Theta_0$, sous $\mathbb{P}_\theta$, la statistique $\mathrm{F}(\mathbf{Y})$ est distribuée suivant une loi de Fisher (voir polycopié, Section IV-3.6) dont on précisera les paramètres.
7. En déduire, pour $\alpha \in]0,1[$, un test de niveau α pour le problème (2).

Exercice 2. Soit $(X_1,\ldots,X_n)$ un n-échantillon du modèle statistique

$$(\mathbb{R}_+, \mathcal{B}(\mathbb{R}_+), \{p_\theta \cdot \mathrm{Leb} \,:\, \theta=(a,\beta)\in\Theta := \mathbb{R}_+^* \times \mathbb{R}_+^*\}) \ ,$$

où, pour tout $x \in \mathbb{R}_+$ et $\theta = (a,\beta) \in \Theta$, la densité de probabilité p_θ par rapport à la mesure de Lebesque Leb sur $\mathbb{R}_+$, est donnée par

$$p_\theta(x) := \frac{a}{\beta^a} x^{a-1}\mathbb{1}_{[0,\beta]}(x).$$

1. Montrer [en justifiant soigneusement le résultat] que l'estimateur du maximum de vraisemblance de (a,β) est donné par
$$\hat\beta_n := X_{n:n} \quad \text{et} \quad \hat a_n := \left\{\frac{1}{n}\sum_{i=1}^n \log(\hat\beta_n/X_i)\right\}^{-1} \ .$$
2. Montrer que pour tout $\theta \in \Theta$,
$$\mathbb{P}_{n,\theta}(\hat\beta_n \leq x) = \begin{cases} 0 & \text{si} \quad x \leq 0 \\ (x/\beta)^{an} & \text{si} \quad 0 \leq x \leq \beta \\ 1 & \text{si} \quad x \geq \beta \ . \end{cases}$$
3. En déduire que pour tout $\theta = (a,\beta)\in\Theta$, $\sqrt{n}\log(\hat\beta_n/\beta) \overset{\mathbb{P}_{n,\theta}-\mathrm{prob}}{\longrightarrow} 0$.
4. Montrer que pour $\theta=(a,\beta)\in\Theta$, sous p_θ.Leb, la loi de la variable $Y_1 := \log(\beta/X_1)$ est Gamma$(1,a)$ (voir Définition IV-3.12 dans le polycopié).
5. Montrer que pour tout $\theta = (a,\beta)\in\Theta$, $n^{-1/2}\sum_{i=1}^n\{\log(\beta/X_i) - 1/a\}$ converge en loi vers une limite que l'on identifiera.
6. En déduire que la suite d'estimateurs $\{\hat a_n,\ n\in\mathbb{N}\}$ est asymptotiquement normale (on précisera la variance asymptotique).

7. Pour $\alpha \in]0,1[$, déterminer un intervalle de confiance de a de probabilité de couverture asymptotique $1-\alpha$.
8. Montrer que la suite de fonctions $G_n(X_1,\ldots,X_n;a,\beta) := na(1-\hat{\beta}_n/\beta)$ est asymptotiquement pivotale. En déduire que la suite de fonctions $\tilde{G}_n(X_1,\ldots,X_n;\beta) := n\hat{a}_n(1-\hat{\beta}_n/\beta)$ est asymptotiquement pivotale.
9. Pour $\alpha \in]0,1[$, déterminer un intervalle de confiance de β de probabilité de couverture asymptotique $1-\alpha$.

Exercice 3. Nous considérons un problème de classification binaire avec abstention : la règle d'apprentissage a une possibilité de s'abstenir de classifier.
Nous utilisons les notations du polycopié, Chapitre III-1 : nous supposons que l'observation (X,Y) (où X est le vecteur de attributs et Y est l'étiquette) est un vecteur aléatoire défini sur un espace de probabilité $(\Omega,\mathcal{F},\mathbb{P})$ à valeurs dans $\mathbb{R}^d \times \{0,1\}$; et nous notons η la fonction de régression :

$$x \in \mathbb{R}^d \mapsto \eta(x) := \mathbb{P}\left(Y=1 \,|\, X=x\right) \ .$$

Nous définissons deux fonctions : une règle de classification $g : \mathbb{R}^d \to \{0,1\}$ et une règle d'abstention $r : \mathbb{R}^d \to \{0,1\}$. Lorsque $r(x)=0$, la règle d'apprentissage se prononce et décide $g(x)$; lorsque $r(x)=1$, la règle d'apprentissage s'abstient. Nous considérons la perte ℓ_ρ définie par

$$\ell_\rho(g(x),r(x),y) := \mathbb{1}_{\{g(x)\neq y\}}\mathbb{1}_{\{0\}}(r(x)) + \rho\,\mathbb{1}_{\{1\}}(r(x))$$

où $\rho \in [0,+\infty[$. Cette perte exprime que lorsque la règle d'apprentissage s'abstient, elle subit une perte ρ et lorsqu'elle se prononce, elle subit une perte de type $0-1$.
Dans la suite, les fonctions g et r sont toujours supposées mesurables ; le risque associé à la perte ℓ_ρ est noté

$$R_\rho(g,r) := \mathbb{E}[\ell_\rho(g(X),r(X),Y)] \ .$$

1. Montrer que pour toute paire de fonctions (g,r), nous avons

$$R_\rho(g,r) = \mathbb{E}[\mathbb{1}_{\{0\}}(r(X))\{\eta(X)\mathbb{1}_{\{0\}}(g(X)) + (1-\eta(X))\mathbb{1}_{\{1\}}(g(X))\}] + \rho\,\mathbb{P}(r(X)=1) \ . \tag{3}$$

Nous posons $h(x) := \min(\eta(x), 1-\eta(x))$.

2. Donner l'expression d'une règle de classification g_* telle que pour toute paire de fonctions (g,r) et tout $\rho \geq 0$,

$$R_\rho(g,r) \geq R_\rho(g_*,r) \ , \qquad \text{avec } R_\rho(g_*,r) = \mathbb{E}[\mathbb{1}_{\{0\}}(r(X))\ h(X)] + \rho\ \mathbb{P}(r(X)=1) \ .$$

3. Montrer que pour tout $\rho \geq 0$ et pour toute règle d'abstention r, on a

$$R_\rho(g_*,r) = \rho + \mathbb{E}[\mathbb{1}_{\{0\}}(r(X))\{h(X)-\rho\}]$$

4. Comment choisir la règle d'abstention r pour $\rho = 0$?
5. Comment choisir la règle d'abstention r pour $\rho \geq 1/2$?
6. Lorsque $\rho \in]0,1/2[$, quelle est la règle d'abstention optimale r_* ?

Exercice 4. Soit $Z = (X_1, \ldots, X_n)$ un n-échantillon du modèle $(\mathbb{N}, \mathcal{P}(\mathbb{N}), (\mathbb{P}_\theta)_{\theta\in\Theta})$, où $\Theta :=$ $\mathbb{N}^*$ et, pour tout $\theta \in \Theta$, $\mathbb{P}_\theta$ est la loi de densité

$$p_\theta(x) := \theta^{-1}\,\mathbb{1}_{\{1,\ldots,\theta\}}(x) = \begin{cases} \theta^{-1} & \text{si } x \in \{1,\ldots,\theta\} \\ 0 & \text{sinon} \end{cases} \quad , \quad x \in \mathbb{N}\,,$$

par rapport à la mesure de comptage sur $\mathbb{N}$.

Soit $\theta_0 \in \Theta$ et $\alpha \in \,]0,1[$, tels que $\theta_0\alpha^{1/n} \in \mathbb{N}^*$.

1. Montrer que pour toute statistique S définie sur $\mathbb{N}^n$ et pour tout $\theta > \theta_0$,

$$\mathbb{E}_\theta[S(Z)\mathbb{1}_{\{X_{n:n}\le\theta_0\}}] = \mathbb{E}_{\theta_0}[S(Z)](\theta_0/\theta)^n\,.$$

On considère le test

$$H_0:\ \theta \le \theta_0, \quad \text{contre} \quad H_1:\ \theta > \theta_0. \tag{4}$$

Nous allons montrer que le test randomisé de fonction critique

$$\phi_{\mathrm{m}}(Z) := \begin{cases} 1 & \text{si } X_{n:n} > \theta_0 \\ \alpha & \text{si } X_{n:n} \le \theta_0 \end{cases}$$

est U. P. P.(α) pour (4).

2. Montrer que le test randomisé de fonction critique ϕ_{m} est de niveau α pour (4).
3. Montrer que pour tout test randomisé de niveau α de fonction critique $\phi : Z \mapsto \phi(Z) \in$ $[0,1]$, et pour tout $\theta > \theta_0$,

$$\mathbb{E}_\theta[\phi_{\mathrm{m}}(Z)] - \mathbb{E}_\theta[\phi(Z)] \ge \mathbb{E}_\theta\left[(\alpha - \phi(Z))\mathbb{1}_{\{X_{n:n}\le\theta_0\}}\right] \ge 0\,.$$

Conclure.

On considère maintenant

$$H_0:\ \theta = \theta_0, \quad \text{contre} \quad H_1:\ \theta \ne \theta_0. \tag{5}$$

Nous considérons le test bilatéral de fonction critique

$$\phi_{\mathrm{b}}(Z) := \begin{cases} 1 & \text{si } X_{n:n} > \theta_0 \text{ ou si } X_{n:n} \le \theta_0\alpha^{1/n} \\ 0 & \text{autrement} \end{cases}$$

Nous allons montrer que ϕ_{b} est U. P. P.(α) pour (5). Dans la suite, ϕ est la fonction critique d'un test randomisé de niveau α de (5).

4. Montrer que le test randomisé de fonction critique ϕ_{b} est de niveau α pour le test (5).
5. Montrer que pour tout $\theta > \theta_0$, nous avons

$$\mathbb{E}_\theta[\phi_{\mathrm{b}}(Z)] - \mathbb{E}_\theta[\phi(Z)] \ge \mathbb{E}_\theta\left[(\phi_{\mathrm{b}}(Z) - \phi(Z))\,\mathbb{1}_{\{X_{n:n}\le\theta_0\}}\right] \ge 0\,.$$

6. Montrer que pour tout $\theta < \theta_0$, nous avons

$$\mathbb{E}_\theta[\phi_{\mathrm{b}}(Z)] - \mathbb{E}_\theta[\phi(Z)] = \mathbb{P}_\theta\left(X_{n:n} \le \theta_0\alpha^{1/n}\right) - \mathbb{E}_\theta[\phi(Z)]\,.$$

7. En déduire que pour $\theta < \theta_0$, nous avons $\mathbb{E}_\theta[\phi_{\mathrm{b}}(Z)] - \mathbb{E}_\theta[\phi(Z)] \ge 0$; on distinguera les cas $\theta \le \theta_0\alpha^{1/n}$ et $\theta_0\alpha^{1/n} \le \theta < \theta_0$.
   Conclure.