Feuille 4 de travaux dirigés

Exercice 1. [*p-value d'un test*]
On considère un problème de test simple général : soit $X_1, \dots, X_n$ i.i.d. de loi P_θ avec $\theta \in \Theta \subset \mathbb{R}$. On cherche à tester H_0 : "$\theta \in \Theta_0$" contre H_1 : "$\theta \notin \Theta_0$" pour un sous-ensemble $\Theta_0 \subset \Theta$ donné. On souhaite rejeter H_0 lorsque une statistique de test $S(X)$ est "grande". On considère la fonction de répartition de $S(X)$ donnée par

$$F_\theta(s) = \mathbb{P}_\theta(S(X) \leq s), \quad s \in \mathbb{R}, \quad \theta \in \Theta,$$

et on suppose qu'il existe une configuration $\theta_0 \in \Theta_0$ "la plus défavorable", c'est-à-dire telle que $\forall \theta \in \Theta_0$, $\forall s \in \mathbb{R}$, $F_\theta(s) \geq F_{\theta_0}(s)$. La fonction F_{θ_0} est supposée de plus continue et strictement croissante sur $\mathbb{R}$.

1. Montrer que le test rejetant si $S(X) > F_{\theta_0}^{-1}(1-\alpha)$ est de niveau α.
2. Calculer la p-value du test, c'est-à-dire $p(X) = \inf\{\alpha \in [0,1] : H_0 \text{ est rejetée au niveau } \alpha\}$.
3. Montrer que pour toute valeur observée $S(x)$ de la statistique $S(X)$,
$$p(x) = \sup_{\theta \in \Theta_0} \{\mathbb{P}_\theta(S(X) \geq S(x))\}.$$
4. Montrer la propriété suivante : $\forall \theta \in \Theta_0,\ \forall t \in [0,1],\ \mathbb{P}_\theta(p(X) \leq t) \leq t$.
5. Montrer que si Θ_0 est réduit à un singleton, i.e. $\Theta_0 = \{\theta_0\}$, alors $p(X) \sim U(0,1)$ sous H_0.

Exercice 2. [*Loi de l'infimum de m variables uniformes indépendantes*]
Soit p_j, $1 \leq j \leq m$, i.i.d. uniformes sur $[0,1]$. On considère la variable $p_{(1)} = \min_{1 \leq j \leq m}\{p_j\}$.

1. Calculer $F(t) = \mathbb{P}(p_{(1)} \leq t)$ pour tout $t \in [0,1]$.
2. Montrer que $p_{(1)}$ admet une densité f par rapport à la mesure de Lebesgue et calculer f puis $\mathbb{E}[p_{(1)}]$. Que se passe t'il lorsque m grandit ?
3. Trouver la loi limite de $r_m p_{(1)}$ puis de $r_m(p_{(1)} - \mathbb{E}[p_{(1)}])$ lorsque m tend vers l'infini, pour une suite $(r_m)_m$ bien choisie.

Exercice 3. [*Modèle de test multiple dans le cas gaussien unilatéral*]

1. Test simple : soit $X^{(1)}, \dots, X^{(n)}$ i.i.d. $\mathcal{N}(\mu, 1)$ pour un paramètre $\mu \in \mathbb{R}$ inconnu.
   (a) Construire un test de niveau α pour H_0 : "$\mu = 0$" contre H_1 : "$\mu > 0$". Calculer la p-value p du test.
   (b) Montrer la propriété suivante : si $\mu = 0$, $p \sim U(0,1)$.

(c) Donner la fonction de répartition de p en fonction de μ lorsque H_1 est vraie.

2. Test multiple : soit $X^{(1)}, \dots, X^{(n)}$ i.i.d. $\mathcal{N}_m(\Delta, \Gamma_m)$ pour un paramètre $\Delta \in (\mathbb{R}_+)^m$ inconnu et une matrice de covariance Γ_m $m \times m$ connue avec des coefficients diagonaux tous égaux à 1.
   (a) Pour chaque $j \in \{1, \dots, m\}$, construire une p-value p_j pour le test de $H_{0,j}$: "$\Delta_j = 0$" contre $H_{1,j}$: "$\Delta_j > 0$".
   (b) Exprimer ce cadre comme un cas particulier du modèle de test multiple général vu en cours en précisant ce que sont $H \in \{0,1\}^m$ et $\mathcal{P}_H$. Le modèle est-il identifiable ? Décrire le sous-modèle correspondant au cas $\Gamma_m = I_m$ et $\Delta_j = \delta$ lorsque $\Delta_j > 0$ (pour un $\delta > 0$ fixé) en précisant la fonction de répartition des p-values sous l'alternative. Que devient-elle lorsque n tends vers l'infini ?

Exercice 4. [*Modèle de test multiple dans le cas gaussien bilatéral*]
Refaire l'exercice précédent dans le cas de tests bilatéraux.

Exercice 5. [*Construction de p-values par permutation*]
On rappelle que si $(T_0, \dots, T_B)$ est échangeable si pour toute permutation σ de $\{0, 1, \dots, B\}$, $(T_{\sigma(0)}, \dots, T_{\sigma(B)}) \sim (T_0, \dots, T_B)$. Dans ce cas, on admettra que la variable

$$W = (B+1)^{-1} \left(1 + \sum_{b=1}^{B} \mathbf{1}\{T_b \geq T_0\}\right)$$

satisfait la propriété $\mathbb{P}(W \leq t) \leq t$ pour tout $t \in [0,1]$. On note ν_n la loi uniforme sur le groupe des permutations de $\{1, \dots, n\}$.

1. Test simple : soit $X^{(1)}, \dots, X^{(n_x)}$ i.i.d. $\sim P$ et $Y^{(1)}, \dots, Y^{(n_y)}$ i.i.d. $\sim Q$, où P et Q sont des lois sur $\mathbb{R}$. On cherche à tester H_0 : "$P = Q$" contre H_1 : "$P \neq Q$". On note $Z = (X^{(1)}, \dots, X^{(n_x)}, Y^{(1)}, \dots, Y^{(n_y)})$ et $n = n_x + n_y$.
   (a) Proposer une statistique de test $T(Z)$ pour tester H_0.
   (b) En notant $Z^\sigma = (Z^{(\sigma(1))}, \dots, Z^{(\sigma(n))})$, montrer que sous H_0, $Z^\sigma \sim Z$, pour toute permutation σ de $\{1, \dots, n\}$ (déterministe) et donc aussi pour toute permutation aléatoire $\sigma \sim \nu_n$ (indépendante de Z).
   (c) Soit $\sigma_1, \dots, \sigma_B$ i.i.d. $\sim \nu_n$ (indépendants de Z). On pose

$$p = (B+1)^{-1} \left(1 + \sum_{b=1}^{B} \mathbf{1}\{T(Z^{\sigma_b}) \geq T(Z)\}\right).$$

   Montrer que sous H_0, pour tout $t \in [0,1]$, $\mathbb{P}(p \leq t) \leq t$.

2. Test multiple : soit $X^{(1)}, \dots, X^{(n_x)}$ i.i.d. $\sim P$ et $Y^{(1)}, \dots, Y^{(n_y)}$ i.i.d. $\sim Q$, où P et Q sont des lois sur $\mathbb{R}^m$. On note $Z = (X^{(1)}, \dots, X^{(n_x)}, Y^{(1)}, \dots, Y^{(n_y)})$ et $n = n_x + n_y$. Pour chaque $j \in \{1, \dots, m\}$, on note $P_j = \mathcal{L}(X_j^{(1)})$ et $Q_j = \mathcal{L}(Y_j^{(1)})$.
   (a) Pour chaque $j \in \{1, \dots, m\}$, proposer une p-value pour tester $H_{0,j}$: "$P_j = Q_j$" contre $H_{1,j}$: "$P_j \neq Q_j$".
   (b) Vérifier qu'on retrouve le modèle de test multiple général vu en cours en précisant ce que sont $H \in \{0,1\}^m$ et $\mathcal{P}_H$.