Feuille 6 de travaux dirigés

Dans toute la suite on considère le cadre de test multiple général avec $H \in \{0,1\}^m$ et $(p_j, 1 \leq j \leq m) \sim P \in \mathcal{P}_H$ tel que

$$\text{pour tout } j \in \{1,\dots,m\} \text{ avec } H_j = 0, \text{ pour tout } t \in [0,1],\ \mathbb{P}(p_j \leq t) \leq t.$$

On note $\mathcal{H}_0(H) = \{1 \leq j \leq m \;:\; H_j = 0\}$. Le paramètre du modèle est noté $\theta = (H, P)$ et l'espace des paramètres est Θ.

Exercice 1. [FWER plus stricte que FDR]
Soit une procedure de test multiple d'ensemble de rejet R.

1. Montrer $\#(\mathcal{H}_0(H) \cap R) \leq \#R \times \mathbf{1}\{\#(\mathcal{H}_0(H) \cap R) \geq 1\}$.
2. En déduire $\mathrm{FDR}_\theta(R) \leq \mathrm{FWER}_\theta(R)$.
3. Montrer que la procédure de Bonferroni contrôle le FDR (sans hypothèse supplémentaire).
4. Montrer que la procédure BH fait toujours plus de rejets que la procédure de Bonferroni. Quelle procédure doit-on utiliser ?

Exercice 2. [Inegalité de Simes] En utilisant un résultat du cours, montrer que pour une famille de variables aléatoires $(U_j, 1 \leq j \leq m)$ identiquement distribuées et uniformes sur $[0,1]$, qui est de plus supposée mutuellement indépendante, on a

$$\mathbb{P}(\exists j \in \{1,\dots,m\} \;:\; U_{(j)} \leq \alpha j/m) = \alpha,$$

où l'on a noté $U_{(1)} \leq \cdots \leq U_{(m)}$.

Exercice 3. [*Fonction de répartition des alternatives en gaussien unilatéral*]

Soit $F_1(t) = \overline{\Phi}(\overline{\Phi}^{-1}(t) - \delta)$ pour $\delta > 0$. Montrer que $t \mapsto F_1(t)/t$ est strictement décroissante sur $]0,1[$ et tend vers l'infini en 0.

Exercice 4. [*Modèle Dirac-Uniforme*]

On considère le cas où les p-values sont indépendantes avec $p_j \sim U(0,1)$ lorsque $H_j = 0$ et $p_j = 0$ lorsque $H_j = 1$ (modèle Dirac-uniforme).

1. Calculer $\widehat{G}(t) = m^{-1}\sum_{j=1}^m \mathbf{1}\{p_j \leq t\}$ en fonction de $\widehat{G}_0(t) = m_0^{-1}\sum_{j=1}^m (1-H_i)\mathbf{1}\{p_j \leq t\}$ et $\pi_0 = m_0/m$. Représenter graphiquement $\widehat{G}$. Où peut-on lire le nombre de rejets $\widehat{\ell}$ de la procédure BH ?

2. On considère une démarche asymptotique en m en supposant que $m_0/m \to \pi_{0,\infty} \in (0,1)$. En utilisant un résultat du cours, déterminer l'équation que satisfait la limite $t^\star$ de $\widehat{t} = \alpha\widehat{\ell}/m$. En déduire $t^\star$ en fonction de $\pi_{0,\infty}$. Determiner les limites de $t^\star$ lorsque $\pi_{0,\infty}$ s'approche de 0 ou de 1. Interpréter le résultat.

Exercice 5. [*Propriété PRDS*]

On observe $X \sim \mathcal{N}(\Delta, \Gamma)$ pour $\Delta \in \mathbb{R}^m$ et pour Γ une matrice $(m \times m)$ avec $\Gamma_{j,j} = 1$ et $\Gamma_{j,j'} \geq 0$ pour tout j, j'. On se propose de montrer que la famille de p-values définie par $p_j = \overline{\Phi}(X_j)$, $1 \leq j \leq m$, est PRDS sur $\{1, \dots, m\}$, c'est-à-dire, pour tout ensemble $D \subset [0,1]^m$ croissant (mesurable) et pour tout $j \in \{1, \dots, m\}$, la fonction

$$u \mapsto \mathbb{P}(p \in D \mid p_j \leq u)$$

est croissante. On rappelle que la distribution de X_{-j} conditionnellement à $X_j = x$ est une loi normale de moyenne $\Delta_{-j} + \Gamma_{-j,j}(x - \Delta_j)$ et de matrice de covariance $\Gamma_{-j,-j} - \Gamma_{-j,j}\Gamma_{j,-j}/\Gamma_{j,j}$, où l'on a noté $X_{-j} = (X_{j'})_{j' \neq j}$; $\Delta_{-j} = (\Delta_{j'})_{j' \neq j}$; $\Gamma_{-j,-j} = (\Gamma_{j',j''})_{j' \neq j, j'' \neq j}$; $\Gamma_{-j,j} = (\Gamma_{j',j})_{j' \neq j}$; $\Gamma_{-j,j} = (\Gamma_{j,j'})_{j' \neq j}$. On fixe $D \subset [0,1]^m$ croissant (mesurable) et $j \in \{1, \dots, m\}$.

1. Pour tout $x \in \mathbb{R}$, déterminer $D_x \subset \mathbb{R}^{m-1}$, un ensemble décroissant, tel que

$$\mathbb{P}(p \in D \mid p_j = \overline{\Phi}(x)) = \mathbb{P}(X_{-j} \in D_x \mid X_j = x).$$

2. Pour tout $x, x' \in \mathbb{R}$ avec $x \leq x'$ montrer que

$$\mathbb{P}(X_{-j} \in D_{x'} \mid X_j = x') \leq \mathbb{P}(X_{-j} \in D_x \mid X_j = x).$$

   En déduire que $u \mapsto \mathbb{P}(p \in D \mid p_j = u)$ est croissante.

3. Montrer le résultat annoncé en écrivant

$$\mathbb{P}(p \in D \mid p_j \leq u) = \mathbb{E}[\mathbb{P}(p \in D \mid p_j) \mid p_j \leq u].$$

Exercice 6. [*Violation de la propriété PRDS*]

On observe $X \sim \mathcal{N}(\Delta, \Gamma)$ pour $\Delta \in \mathbb{R}^m$ et pour Γ une matrice $(m \times m)$ avec $\Gamma_{j,j} = 1$ et $\Gamma_{j,j'} \geq 0$ pour tout j, j'. On considère les p-values bilatérales $p_j = 2\,\overline{\Phi}(|X_j|)$, $1 \leq j \leq m$. Montrer que la propriété PRDS n'est plus vérifiée.

[Indication : on pourra considérer le cas de dépendance maximale]

Exercice 7. [*Procédure de Benjamini-Yekutieli (BY)*]

Soit $\widehat{\ell}$ le nombre de rejets de la procédure de BY. Donner une condition nécessaire et suffisante de type $\widehat{\ell} \leq c_m$ pour que le seuil de BY soit plus conservatif que celui de la procédure de Bonferroni. Que devient cette condition lorsque $m = 1000$? Que pensez vous de la procédure de BY lorsque peu de rejets sont attendus (données sparse ou avec faible signal) ?