Contrôle : Statistique pour les big data

Patrick Tardivel,
Université de Bourgogne

1er avril 2022

Exercice 1 (Question de cours) *On considère le problème de test suivant*

$$\begin{cases} X : (\Omega, \mathcal{F}, Pr) \to (\mathbb{X}, \mathcal{X}) \\ \mathbb{P}^X \in \mathcal{P}, \text{ où } \mathbb{P}^X \text{ est la loi } X \text{ et } \mathcal{P} \text{ est une famille de lois de probabilité sur } (\mathbb{X}, \mathcal{X}) \\ \mathcal{H}^0 : \mathbb{P}^X \in \mathcal{P}^0 \text{ où } \mathcal{P}^0 \text{ est un sous-ensemble de } \mathcal{P} \end{cases}.$$

Soit $(R_\alpha)_{\alpha\in[0,1]}$ *une famille emboîtée de régions de rejet telle que* $R_0 = \emptyset$, $R_1 = \mathbb{X}$, $R_\alpha \subset R_{\alpha'}$ *lorsque* $\alpha \leq \alpha'$ *et contrôlant le risque de première espèce au niveau* α, *c'est-à-dire que*

$$\forall \alpha \in [0,1], \sup\{\mathbb{P}(R_\alpha) : \mathbb{P} \in \mathcal{P}^0\} \leq \alpha.$$

On pose $p(X) := \inf\{\alpha \in]0,1] : X \in R_\alpha\}$. *Montrer que sous l'hypothèse nulle* $\mathcal{H}^0$ *la variable aléatoire réelle* $p(X)$ *satisfait*

$$\forall u \in [0,1], Pr(p(X) \leq u) \leq u.$$

Exercice 2

1. *Rappeler la procédure de Benjamini-Hochberg. Que peut-on dire du contrôle du FDR lorsque les p-valeurs sont indépendantes ? Même question lorsqu'aucune hypothèse d'indépendance n'est faîte sur les p-valeurs.*
2. *Soit* $p_1(X), \ldots, p_5(X)$ *des p-valeurs indépendantes associées aux hypothèses nulles* $\mathcal{H}^0_1, \ldots, \mathcal{H}^0_5$. *Les valeurs expérimentales des p-valeurs sont :* $p_1(X^{exp}) = 0,035, p_2(X^{exp}) = 0,012, p_3(X^{exp}) = 0,080, p_4(X^{exp}) = 0,016$ *et* $p_5(X^{exp}) = 0,052$. *Pour un contrôle du FDR au niveau* $0,05$, *donner les hypothèses nulles rejetées par la procédure de Benjamini-Hochberg. Vous justifierez votre réponse.*

Exercice 3

1. *Rappeler les procédures de Bonferroni et de Holm.*
2. *Soit* $p_1(X), \ldots, p_5(X)$ *des p-valeurs associées aux hypothèses nulles* $\mathcal{H}^0_1, \ldots, \mathcal{H}^0_5$. *Les valeurs expérimentales des p-valeurs sont :* $p_1(X^{exp}) = 0,027, p_2(X^{exp}) = 0,015, p_3(X^{exp}) = 0,008, p_4(X^{exp}) = 0,014$ *et* $p_5(X^{exp}) = 0,062$. *Lorsque le FWER est contrôlé au niveau* $0,05$, *donner les hypothèses nulles rejetées par la procédure de Bonferroni. Même question pour la procédure de Holm. Vous justifierez vos réponses.*

Exercice 4

1. *Soit* $X_1, \ldots, X_n$ *un échantillon de loi normale* $\mathcal{N}(\mu, \sigma^2)$ *où* $\mu \in \mathbb{R}$ *et* σ^2 *sont inconnues. On souhaite tester l'hypothèse nulle* $\mathcal{H}^0 : \mu \leq 0$ *contre l'hypothèse alternative* $\mathcal{H}^1 : \mu > 0$. *On pose* $\overline{X} := \frac{1}{n}\sum_{i=1}^n X_i$, $S^2_{corr}(X) := \frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X})^2$ *et* $T := \overline{X}/\sqrt{S^2_{corr}(X)/n}$. *Avec un risque de première espèce* $\alpha \in]0,1[$ *on rejette* $\mathcal{H}_0$ *lorsque* $T > F^{-1}(1-\alpha)$ *où* F *est la fonction de répartition d'une loi de Student à* $n-1$ *degrés de liberté.*

(a) *Déterminer la p-valeur de cette procédure de test.*

(b) *Pour quelle valeur de* $\mu \in \mathbb{R}$ *la p-valeur suit une loi uniforme sur* $[0,1]$ *? Vous justifierez votre réponse.*

2. *Soit* $X_1, \ldots, X_{n_1}$ *un échantillon de loi normale* $\mathcal{N}(\mu_X, \sigma_X^2)$ *indépendant de l'échantillon* $Y_1, \ldots, Y_{n_2}$ *de loi normale* $\mathcal{N}(\mu_Y, \sigma_Y^2)$ *où* $\mu_X \in \mathbb{R}, \mu_Y \in \mathbb{R}, \sigma_X^2 > 0$ *et* $\sigma_Y^2 > 0$ *sont inconnues. On souhaite tester l'hypothèse nulle* $\mathcal{H}^0 : \sigma_X^2 = \sigma_Y^2$ *contre l'hypothèse alternative* $\mathcal{H}^1 : \sigma_X^2 > \sigma_Y^2$*. On pose* $\overline{X} = \frac{1}{n_1}\sum_{i=1}^{n_1} X_i$, $\overline{Y} = \frac{1}{n_2}\sum_{i=1}^{n_2} Y_i$ $S_{corr}^2(X) := \frac{1}{n-1}\sum_{i=1}^{n_1}(X_i - \overline{X})^2$ *et* $S_{corr}^2(Y) := \frac{1}{n_2-1}\sum_{i=1}^{n_2}(Y_i - \overline{Y})^2$ *et* $S := S_{corr}^2(X)/S_{corr}^2(Y)$*. Avec un risque de première espèce* $\alpha \in]0,1[$ *on rejette* $\mathcal{H}_0$ *lorsque* $S > F^{-1}(1-\alpha)$ *où* F *est la fonction de répartition d'une loi de Fisher à* $n_1 - 1, n_2 - 1$ *degrés de liberté.*

   (a) *Déterminer la p-valeur de cette procédure de test.*

   (b) *Montrer que sous l'hypothèse nulle* $\mathcal{H}^0$*, lorsque* $\sigma_X^2 = \sigma_Y^2$*, la p-valeur suit une loi uniforme sur* $[0,1]$.