Travaux pratiques 3 : p-valeur

Patrick Tardivel, Université de Bourgogne

Ci-dessous quelques commandes R utiles pour ces travaux pratiques :

- La commande : rnorm(n, mean =. . . , sd =. . .) permet de générer un échantillon expérimental de taille n d'une loi normale de moyenne : "mean" et d'écart-type : "sd".
- La commande : pt(q, df =. . .) permet d'évaluer la fonction de répartition d'une loi de Student à df degrés de liberté.
- La commande : pchisq(q, df =. . .) permet d'évaluer la fonction de répartition d'une loi de khi-deux à df degrés de liberté.
- Soit $x = (x_1, \ldots, x_n)$ un échantillon expérimental de taille n d'une loi dont la densité est f. On peut estimer f via la densité empirique. Pour représenter la densité empirique avec R on utilise la commande : plot(density(x)).
- Soit $x = (x_1, \ldots, x_n)$ un échantillon expérimental de taille n d'une loi dont la fonction de répartition est F. On peut estimer F via la fonction de répartition empirique donnée par

$$F_n(t) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}(x_i \leq t).$$

  Pour représenter une telle fonction avec R on effectue la démarche suivante :

  - On ordonne par valeur croissante les composantes x via la commande sort ; on obtient $x_{(1)} \leq x_{(2)} \leq \cdots \leq x_{(n)}$. On pose $xord = (x_{(1)}, x_{(2)}, \ldots, x_{(n)})$.
  - On construit le vecteur $y = (1/n, 2/n, \ldots, 1)$ via la commande : $y = (1 : n)/n$.
  - On construit le nuage de points plot($xord, y$).

Exercice 1 *Coder sur R une fonction qui à un échantillon expérimental $x = (x_1, \ldots, x_n)$ retourne le graphe de la fonction de répartition empirique.*

Exercice 2 (Test d'ajustement d'une moyenne) *On considère $X = (X_1, \ldots, X_{100})$ un échantillon de loi $\mathcal{N}(\mu, \sigma^2)$. On pose $T(X)$ la statistique*

$$T(X) = \frac{\overline{X}}{\sqrt{S^2_{corr}}/10} \text{ où } \overline{X} = \frac{1}{100}\sum_{i=1}^{100} X_i \text{ et } S^2_{corr} = \frac{1}{99}\sum_{i=1}^{100}(X_i - \overline{X})^2.$$

La p-valeur pour le test statistique $\mathcal{H}^0 : \mu = 0$ vs $\mathcal{H}^0 : \mu \neq 0$ est $p^0(X) = 2(1 - F_0(|T(X)|))$, où F_0 est la fonction de répartition d'une loi de Student à 99 degrés de liberté. De même, la p-valeur pour le test statistique $\mathcal{H}^0 : \mu \leq 0$ vs $\mathcal{H}^0 : \mu > 0$ est $p^{\leq 0}(X) = 1 - F_0(T(X))$.

1. *Via des simulations, représenter la densité empirique ainsi que la fonction de répartition empirique des variables aléatoires réelles $p^0(X)$ et $p^{\leq 0}(X)$ lorsque $\sigma = 1$ et que $\mu \in \{-0,5; -0,2; 0; 0,2; 0,5\}$.*

Exercice 3 (Loi de Benford discrète) *Des étudiantes en licence de mathématiques travaillent sur la loi de Benford. Pour leur projet, ces étudiantes cherchent à montrer que le premier chiffre du prix d'un article choisi aléatoirement peut être modélisé par une loi de Benford. Plus précisément, une variable aléatoire réelle X suit une loi de Benford discrète à valeurs dans $\{1, \ldots, 9\}$ si*

$$\forall k \in \{1, \ldots, 9\}\ Pr(X = k) = \log_{10}(k + 1) - \log_{10}(k) = \log_{10}(1 + 1/k).$$

Pour un échantillon expérimental de 790 prix, la Figure 1 donne les fréquences de chacun des chiffres observés en comparaison avec les probabilités de la loi de Benford.

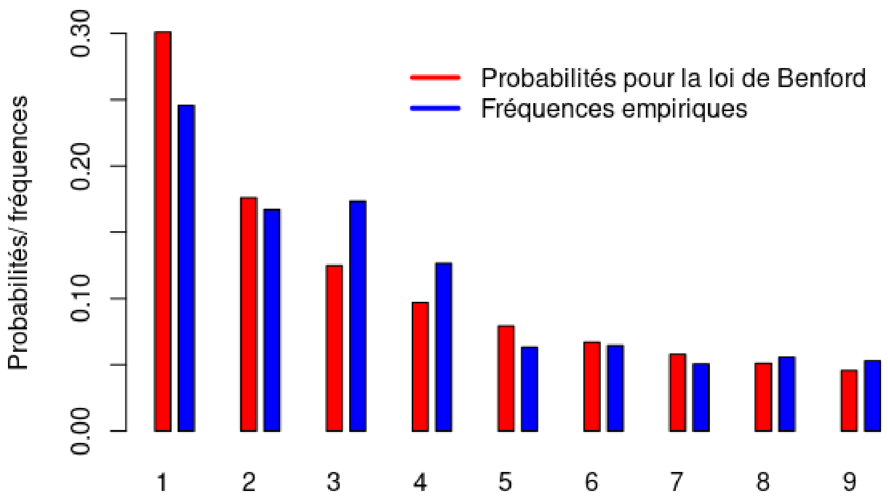

Figure 1: Les bâtons rouges représentent les probabilités pour la loi de Benford discrète. Les bâtons bleus représentent les fréquences d'apparition de chacun des chiffres de 1 à 9.

Tester l'hypothèse nulle : la variable aléatoire réelle qui modélise le premier chiffre du prix d'un article choisi aléatoirement suit une loi de Benford discrète contre l'hypothèse alternative : la loi de cette variable aléatoire ne suit pas une loi de Benford discrète. Vous donnerez la p-valeur, préciserez sa loi (asymptotique) sous l'hypothèse nulle et vous calculerez sa valeur expérimentale à partir des données $\widehat{\pi}_1^{\text{exp}} = \frac{194}{790}, \widehat{\pi}_2^{\text{exp}} = \frac{132}{790}, \widehat{\pi}_3^{\text{exp}} = \frac{137}{790}, \widehat{\pi}_4^{\text{exp}} = \frac{100}{790}, \widehat{\pi}_5^{\text{exp}} = \frac{50}{790}, \widehat{\pi}_6^{\text{exp}} = \frac{51}{790}, \widehat{\pi}_7^{\text{exp}} = \frac{40}{790}, \widehat{\pi}_8^{\text{exp}} = \frac{44}{790}$ *et* $\widehat{\pi}_9^{\text{exp}} = \frac{42}{790}$.

Pour effectuer ce test, vous pouvez utiliser le résultat suivant. Soit $(X_1, \ldots, X_n)$ *un échantillon de loi de Benford discrète. Pour* $k \in \{1, \ldots, 9\}$ *on pose* $\widehat{\pi}_k := \frac{1}{n}\sum_{i=1}^n \mathbf{1}(X_i = k)$ *le proportion aléatoire d'observations du chiffre* k *et on définit* Y_n *par :*

$$Y_n = n \sum_{k=1}^{9} \frac{(\widehat{\pi}_k - \log_{10}(1 + 1/k))^2}{\log_{10}(1 + 1/k)}.$$

Alors on a la convergence en loi suivante

$$Y_n \xrightarrow[n\to+\infty]{\text{Loi}} \chi^2(8).$$