
TEST DU χ^2 ET DE KOLMOGOROV

Exercice 1 - Test du χ^2 .

Le tableau suivant donne un ensemble d'observations de variables aléatoires indépendantes et de même loi :

4.13	1.41	-1.16	-0.75	1.96	2.46	0.197	0.24	0.42	2.00
2.08	1.48	1.73	0.82	0.33	-0.76	0.42	4.60	-2.83	0.197
2.59	0.54	4.06	-0.69	4.99	0.67	2.45	5.61	2.13	1.76
5.03	0.85	1.29	0.17	-0.38	2.76	-1.03	1.87	4.48	0.73

Est-il raisonnable de penser que ces observations sont issues d'une population dont la loi de probabilité est $\mathcal{N}(1, 4)$? (on effectuera un test du χ^2 avec 4 classes et un risque $\alpha = 0.01$).

Correction: Comme précisé en cours, les quatre classes doivent être équiprobables. En utilisant la symétrie de la densité de probabilité normale, les quatre classes sont de la forme $C_1 =]-\infty, x_1[$, $C_2 =]x_1, x_2 = 1[$, $C_3 =]1, x_3[$ et $C_4 =]x_3, +\infty[$ avec

$$\int_{-\infty}^{x_1} p_{1,4}(x)dx = \int_{x_1}^{x_2} p_{1,4}(x)dx = \int_{x_2}^{x_3} p_{1,4}(x)dx = \int_{x_3}^{+\infty} p_{1,4}(x)dx = \frac{1}{4}$$

où $p_{1,4}$ est la densité de probabilité de la loi normale $\mathcal{N}(1, 4)$. En effectuant le changement de variables $y = \frac{x-1}{2}$, on obtient

$$\int_{-\infty}^{\frac{x_1-1}{2}} p_{0,1}(u)du = \frac{1}{4} \Leftrightarrow F\left(\frac{x_1-1}{2}\right) = \frac{1}{4}$$

où F est la fonction de répartition de la loi normale $\mathcal{N}(0, 1)$. On en déduit

$$x_1 = 2F^{-1}\left(\frac{1}{4}\right) + 1 \approx -0.34.$$

Par symétrie de la densité normale, on a aussi $x_3 - 1 = 1 - x_1$ soit $x_3 = 2 - x_1 \approx 2.34$. La statistique du test du χ^2 est

$$\phi = \sum_{k=1}^K \frac{(n_k - np_k)^2}{np_k}$$

avec

- n le nombre d'observations (ici $n = 40$)
- K le nombre de classes choisi (ici $K = 4$)
- n_k le nombre d'observations de la classe C_k (ici, on obtient $n_1 = 7, n_2 = 12, n_3 = 10$ et $n_4 = 11$)
- p_k la probabilité d'appartenir à la classe C_k (ici, $p_k = 1/4$ puisque les classes sont équiprobables).

Un calcul élémentaire conduit à

$$\phi = \frac{9}{10} + \frac{4}{10} + \frac{0}{10} + \frac{1}{10} = \frac{14}{10} = 1.4.$$

Comme il y a $K = 4$ classes, ϕ suit une loi du χ^2 à $K - 1 = 3$ degrés de liberté sous l'hypothèse H_0 . On a donc

$$\alpha = 0.01 = P[\text{Rejeter } H_0 | H_0 \text{ vraie}] = P[\phi > S_\alpha | \phi \sim \chi_3^2] = 1 - F_{\chi_3^2}(S_\alpha)$$

où $F_{\chi^2_3}$ est la fonction de répartition d'une loi du χ^2_3 . En conséquence

$$F_{\chi^2_3}(S_\alpha) = 1 - \alpha = 0.99 \Leftrightarrow S_\alpha = F^{-1}(0.99) \approx 11.345.$$

Comme $\phi = 1.4 < S_\alpha$, on accepte l'hypothèse H_0 avec le risque $\alpha = 0.01$.

Exercice 2 - Test de Kolmogorov.

Une série d'observations de v.a. indépendantes et de même loi a donné les résultats suivants

0.68	0.48	0.39	0.25	0.685	0.80	0.87	0.10	0.49	0.88
0.0078	0.67	0.063	0.73	0.896	0.32	0.79	0.40	0.996	0.53

Effectuer un test de Kolmogorov et décider si on peut considérer que ces observations sont issues d'une population dont la densité de probabilité est uniforme sur l'intervalle $[0, 1]$?

Correction: la statistique du test de Kolmogorov est

$$D_n = \sup_{x \in \mathbb{R}} |\hat{F}(x) - F_0(x)| > S_\alpha$$

où F_0 est la fonction de répartition de la loi testée (ici $F_0(x_i) = x_i$) et \hat{F} est la fonction de répartition des données. Comme expliqué en cours, si on a pris soin de ranger les observations par ordre croissant, on obtient

$$D_n = \max_{i \in \{1, \dots, n\}} \max\{E_i^+, E_i^-\}$$

avec $E_i^+ = \left| \frac{i}{n} - F_0(x_i) \right|$ et $E_i^- = \left| \frac{i-1}{n} - F_0(x_i) \right|$.

$F_0(x_i) = x_i$	0.0078	0.063	0.10	0.25	0.32	0.39	0.40	0.48	0.49	0.53
E_i^+	0.0422	0.037	0.05	0.05	0.07	0.09	0.05	0.08	0.04	0.03
E_i^-	0.0078	0.013	0	0.10	0.12	0.14	0.10	0.13	0.09	0.08
$\max(E_i^+, E_i^-)$	0.0422	0.037	0.05	0.1	0.12	0.14	0.1	0.13	0.09	0.08

$F_0(x_i) = x_i$	0.67	0.68	0.69	0.73	0.79	0.80	0.87	0.88	0.90	0.996
E_i^+	0.12	0.08	0.04	0.03	0.04	0	0.02	0.02	0.05	0.004
E_i^-	0.17	0.13	0.09	0.08	0.09	0.05	0.07	0.03	0	0.046
$\max(E_i^+, E_i^-)$	0.17	0.13	0.09	0.08	0.09	0.05	0.07	0.03	0.05	0.046

On a donc

$$D_n = \sup_{x \in \mathbb{R}} |\hat{F}(x) - F_0(x)| = 0.17.$$

Si on choisit $\alpha = 0.01$, on a $S_\alpha = 0.352$. Comme $D_n < 0.352$, on accepte l'hypothèse H_0 avec le risque $\alpha = 0.01$.