E. MORICE

Tests de normalité basés sur l'emploi des statistiques d'ordre

Revue de statistique appliquée, tome 23, nº 3 (1975), p. 75-79
<http://www.numdam.org/item?id=RSA_1975__23_3_75_0>

TESTS DE NORMALITÉ BASÉS SUR L'EMPLOI DES STATISTIQUES D'ORDRE (1)

E. MORICE

Depuis l'étude publiée par Kolmogorov dans le Bulletin de l'Académie des Sciences (U.R.S.S., 1933), de nombreux tests de l'hypothèse de normalité ont été proposés et tabulés.

On peut les ranger en deux catégories :

a) Tests basés sur les écarts entre les fréquences cumulées estimées à partir de l'échantillon et les valeurs correspondantes de la fonction de répartition de la loi normale (définie a priori ou estimée à partir de l'échantillon).

b) Tests basés sur le rapport d'une estimation linéaire de la variance (ou de l'écart-type), calculés à partir de l'échantillon, aux estimations définies à partir de $\Sigma (x_i - \overline{x})^2$.

A – TESTS DE KOLMOGOROV ET TESTS DERIVES

Etant donné un échantillon de n observations ordonnées :

$$x_1 \leqslant x_2 \leqslant \ldots \leqslant x_n ,$$

soit z_i la valeur pour $X = x_i$ de la fonction de répartition de la loi normale considérée, avec :

$$z_i = P(x_i) \text{ pour une loi spécifiée a priori}$$

ou $$z_i = F\left(\frac{x_i - \overline{x}}{s}\right) \text{ pour une loi estimée à partir de l'échantillon,}$$

avec $$s^2 = \sum_i (x_i - \overline{x})^2 /(n - 1).$$

Les principaux tests proposés utilisent les statistiques suivantes :

1/ Test de Kolmogorov-Smirnov

$$D = \text{Max}\left\{\max\left[\frac{i}{n} - z_i\right], \max\left[z_i - \frac{i-1}{n}\right]\right\}$$

---

(1) Cette note apporte un complèment à l'étude publiée par le même auteur : "Tests de normalité d'une distribution observée". RSA – n : 2. 1972

2/ Test de Cramer-von Mises

$$W^2 = \sum_{i=1}^{n} \left[z_i - \frac{2i-1}{2n} \right]^2 + \frac{1}{12n}$$

3/ Test de Kuiper

$$V = \text{Max} \left[\frac{i}{n} - z_i \right] + \text{Max} \left[z_i - \frac{i-1}{n} \right]$$

4/ Test de Watson

$$U^2 = W^2 - n(\bar{z} - 0{,}5)^2$$

où $\bar{z}$ est la moyenne des valeurs z_i.

5/ Test de Anderson-Darling

$$A = \left\{ - \left[\sum_{i=1}^{n} (2i-1) \left\{ \log z_i + \log (1 - z_{n+1-i}) \right\} \right] / n \right\} - n$$

Les tables relatives a ces tests, pour z_i, défini a priori ou estimé, ont été présentées sous diverses formes, mais elles sont toutes résumées dans les tables de Stephens, utilisables pour toutes valeurs de n, publiées dans Biometrika Tables (Tome II p. 359), où l'on trouvera aussi les références relatives aux travaux originaux.

6/ Test de Durbin – Biometrika 1961 p. 41-55

On peut aussi ranger dans cette catégorie le test utilisant la statistique

$$D = \underset{i}{\text{Max}} \left[\frac{i}{n} - \sum_{j=1}^{i} (n+2-j)(C_j - C_{j-1}) \right], i = 1 \ldots n$$

avec $$0 \leqslant C_0 \leqslant C_1 \leqslant \ldots \leqslant C_n$$

$\ldots C_j^x \ldots$ étant les différences ordonnées de la forme $C_i = z_i - z_{i-1}$

$$C_0 = 0, \quad C_1 = z_1, \quad C_2 = z_2 - z_1 \ \ldots \ C_{n+1} = 1 - z_n$$

B – TESTS BASES SUR DES ESTIMATIONS DE VARIANCE OU D'ECART-TYPE

Dans ce qui suit, pour l'échantillon ordonné

$$x_1 \leqslant x_2 \leqslant \ldots \leqslant x_j \leqslant \ldots x_n$$

posons $$S^2 = \sum_{j=1}^{n} (x_j - \bar{x})^2, \ j = 1, 2, \ldots, n$$

Les statistiques utilisées sont les suivantes :

1/ Test de David – Biometrika 1954, p. 482-93 pour $n < 1\,000$

Compte tenu de la laison entre l'étendue de l'échantillon et l'écart-type de la distribution, il utilise la statistique

$$U = \sqrt{n-1}\,(x_n - x_1) / \left[\sum_1^n (x_i - \bar{x})^2\right]^{1/2}$$

2/ Test de Shapiro-Wilk – Biometrika 1965 p. 591-611, tables pour

$$n \leqslant 50 \qquad W = b^2/S^2$$

$$b = \sum_i a_i (x_{n+1-i} - x_i)$$

pour $$i = 1, 2, \ldots \frac{n}{2} \text{ ou } \frac{n-1}{2}$$

suivant que n est pair ou impair.

Pour $n \leqslant 50$, les tables de S.W. donnent les valeurs des coefficients a_i en fonction de i et n, ainsi que les valeurs critiques de W.

3/ Test de Shapiro-Francia – Journal of the American Statistical Association, 1972, n° 337 p. 214-17. Tables pour n = 50 à 100.

Il utilise la statistique

$$W' = b'^2/S^2$$

de même forme que celle de S.W. avec

$$b' = \sum_i a'_i (x_{n+1-i} - x_i)$$

$i = 1, 2 \ldots \frac{n}{2}$ ou $\frac{n-1}{2}$ selon que n est pair ou impair, mais les coefficients a'_i qui n'ont pas été tabulés doivent dans chaque cas être calculés à partir des formules

$$a'_i = \zeta_{in} / \left[2 \sum_i \zeta_{in}^2\right]^{1/2}$$

où ζ_{in} désigne l'espérance mathématique de la $i^{\text{ème}}$ valeur d'un échantillon de n valeurs de la variable normale réduite ordonnée (Biometrika Tables Tome II).

Les tables de S.F. prolongent celles de S.W. jusqu'à n = 99, en ce qui concerne les valeurs critiques de W', mais leur emploi – en plus du calcul des a'_i – semble de faible valeur pratique en raison des très faibles écarts entre les valeurs $W'_{1-\alpha}$ lorsque le risque α varie. Ainsi par exemple pour n = 50, $W'_{1-\alpha}$ varie de 0,994 pour $\alpha = 0{,}01$ à 0,987 pour $\alpha = 0{,}20$.

4/ Test de d'Agostino – Biometrika, 1971, n° 2 p. 341-348. Tables pour n = 50 à 1 000.

Il utilise la statistique Y définie par la série de calculs ci-après :

$$T = \sum_{i=1}^{n} \left[i - \frac{1}{2}(n+1) \right] x_i \ , \ i = 1 \ldots n$$

$$D = T/n^{3/2} \left[\sum_i (x_i - \bar{x})^2 \right]^{1/2}$$

$$Y = [D - 0,28209479] / \frac{0,02998598}{\sqrt{n}}$$

Pour $n = 50(10) \ldots 100(50) \ldots 1\,000$, la table de d'A. donne divers fractiles de la distribution de Y, permettant un test bilatéral aux risques $\alpha = 0,01 - 0,02 - 0,05 - 0,10$.

La table est d'un emploi facile, les valeurs critiques étant nettement distinctes lorsque α varie.

Ainsi pour

$$n = 50, \ Y_{0,005} = -3,949 \quad \text{et} \quad Y_{0,995} = 1,192$$

alors que $\quad Y_{0,025} = -2,757 \quad$ et $\quad Y_{0,975} = 1,038$.

REMARQUES GENERALES

I – La puissance de ces tests, généralement croissante avec l'effectif n de l'échantillon, dépend évidemment de l'hypothèse alternative opposée à l'hypothèse normale que l'on veut tester.

Cette hypothèse alternative étant le plus souvent inconnue, il résulte des études comparatives faites par simulation, pour diverses valeurs de n (10 – 15 – 20 – 30 – 50), par Shapiro-Wilk-Chen (Journal of the American Statistical Association 1966, p. 1343-1372), d'Agostino (Biometrika 1971, n° 2 p. 341) et Shapiro-Francia (Jal Amer. Stat. Ass. 1972 p. 215-216) :

– que les tests W, W' et Y sont généralement plus puissants que les tests du type Kolmogorov et dérivés, en particulier pour les petits échantillons ($n < 20$).

– que, dans le cas d'une hypothèse alternative uniforme le test U est légèrement plus puissant que le test W,

– que pour $n < 50$, le test W est généralement aussi puissant ou plus puissant que d'autres tests classiques de K. Pearson basés soit sur les moments (tests $\sqrt{b_1}$ et b_2), soit sur la comparaison des fréquences des classes (tests χ^2) (Biometrika Tables Tome II et Biometrika 1973 n° 3, p. 610-625).

II – Quel que soit le test utilisé, il lui correspond un risque α fixé a priori – en général assez faible : 1 % – 5 %, de rejeter l'hypothèse de normalité alors qu'elle est vraie et de se priver ainsi de l'emploi des propriétés de lois dérivées de la loi normale.

Mais il lui correspond aussi un risque inconnu, qui peut être important, d'accepter comme normale une distribution qui ne l'est pas.

Alors, une question se pose, importante bien que généralement négligée par le praticien : dans quelle mesure la manière dont la distribution considérée à partir d'un échantillon, s'écarte de la normalité, assure-t-elle cependant la validité pratique de l'emploi des tests t , χ^2 , F utilisés dans les problèmes de comparaison de moyennes et de variances ?

Une telle question ne peut pas avoir de réponse certaine, faute de pouvoir en toute rigueur mesurer "l'écart" entre la distribution d'où provient l'échantillon et une distribution normale.

Cependant si on admet en première approximation que la forme d'une distribution est suffisamment caractérisée par ses quatre premiers moments et, plus particulièrement, par les deux paramètres de Pearson.

$$\pm\sqrt{\beta_1} = \frac{\mu_3}{\mu_2^{3/2}} \quad , \quad \beta_2 = \frac{\mu_4}{\mu_2^2} \, ,$$

respectivement égaux à 0 et 3 dans le cas d'une distribution normale, il est évidemment possible – par une méthode de simulation, utilisant un très grand nombre d'échantillons d'effectifs 5, 10, 20, ... provenant de populations correspondant à diverses valeurs de β_1 , β_2, de voir comment se comportent les tests t , χ^2 , F , comparés à leurs résultats théoriques en cas de normalité.

Une telle étude a été entreprise et est poursuivie par le Professeur E. Pearson et son équipe. Pour des échantillons aussi petits que n = 10 elle a montré en particulier la robustesse du test t de comparaison de deux moyennes aussi bien pour $\sqrt{\beta_1} = 0$ et $\beta_2 = 3$ (distribution normale), que pour $\sqrt{\beta_1} = 0{,}6$ et β_2 variant de 2 à 4,4 (distributions dissymétriques assez largement étalées vers les petites ou les grandes valeurs de la variable). Le test t relatif à une seule moyenne est peu sensible aux variations de β_2 lorsque $\sqrt{\beta_1} = 0$, mais pour $\sqrt{b_1} = 0{,}6$, les écarts entre les risques théoriques $\alpha = 0{,}01$ ou 0,05 et les risques réels associés aux tests, sont assez importants et plus particulièrement dans les tests unilatéraux.

Par contre les tests relatifs à une variance ou à la comparaison de deux variances, sont aussi bien pour $\sqrt{\beta_1} = 0$ que pour $\sqrt{\beta_1} = 0{,}6$, très sensibles aux variations de β_2.

Ces résultats, obtenus par simulation, correspondent à des valeurs de $\sqrt{\beta_1}$ et β_2, fixées a priori, alors que dans chaque cas particulier on ne connaît que leurs estimations $\sqrt{b_1}$ et b_2 à partir des moments empiriques

$$m_q = \frac{\Sigma\,(x - \bar{x})^q}{n}$$

Une étude des fluctuations de $\sqrt{b_1}$ et b_2, autour des valeurs fixées $\sqrt{\beta_1}$ et β_2, devrait sans doute permettre de donner une portée plus réaliste aux résultats ci-dessus.

Une telle étude implique la réalisation et l'analyse d'une masse considérable de calculs, dont le résultat, lorsqu'il sera publié, permettra sans doute de donner aux praticiens des règles d'emploi valide des tests classiques.