REVUE DE STATISTIQUE APPLIQUÉE

R. SNEYERS

Sur les tests de normalité

Revue de statistique appliquée, tome 22, n° 2 (1974), p. 29-36
<http://www.numdam.org/item?id=RSA_1974__22_2_29_0>

SUR LES TESTS DE NORMALITÉ (1)

R. SNEYERS
Institut Royal Météorologique de Belgique

1. INTRODUCTION

Dans un article récent [1], E. Morice a passé en revue un ensemble de tests proposés dans la littérature statistique pour vérifier l'hypothèse de la normalité de la répartition d'une série d'observations. Bien que l'auteur se soit défendu d'apporter un éventail complet de ces tests, on peut cependant considérer que ceux qui ont été réunis témoignent des diverses tendances qui ont présidé à leur élaboration.

En outre, en rappelant les résultats obtenus par les expériences Monte Carlo effectuées pour faire ressortir les puissances respectives de ces tests dans le cas d'alternatives bien déterminées, la question se trouvait implicitement soulevée de savoir pourquoi tel test donne en général un meilleur résultat que tel autre.

Dans ce qui suit nous avons replacé les tests envisagés dans le cadre général de l'adéquation d'une loi normale donnée à une série d'observations vue sous l'angle du principe des moindres carrés généralisés, et nous avons montré dans quelle mesure les statistiques des tests cités formaient des cas particuliers de la statistique du cas général.

Il est apparu de cette manière que la moindre puissance de certains tests par rapport à d'autres pouvait directement être rapportée à un moindre degré d'exhaustivité ou à une moindre efficacité (au sens de Fisher) des statistiques correspondantes.

2. DE L'ADEQUATION DE LA LOI NORMALE

Comme les moments de la loi normale existent pour n'importe quel ordre, on peut admettre que cette loi appartient à une famille de lois de probabilité à une infinité de paramètres fonctionnellement indépendants et que, dans cette famille, toute loi normale est elle-même définie par des valeurs particulières de ces paramètres.

----

(1) Article remis le 28/2/73.

Dans ces conditions, pour vérifier l'adéquation de la loi normale à une série de n observations indépendantes, on pourra procéder à l'estimation de n parmi l'infinité de paramètres, estimations que l'on peut supposer corrigées pour le biais, et procéder à un test de compatibilité des estimations obtenues avec les valeurs données de ces n paramètres sous l'hypothèse nulle d'une loi normale déterminée. En théorie, le choix des n paramètres importe peu puisque n observations conduisent tout au plus à n estimations fonctionnellement indépendantes.

Si $||\nu_{ij}||$ désigne alors la matrice des covariances des estimations des paramètres sous l'hypothèse nulle et si β_i, $\hat{\beta}_i$, $i = 1, 2, \ldots, n$ désignent respectivement les valeurs données de ces paramètres et leurs estimations, la compatibilité des estimations avec les valeurs données ou encore, l'adéquation de la loi normale considérée pourra, en .application du principe des moindres carrés généralisé, être véfifiée au moyen de la statistique :

$$X_n^2 = \Sigma (\hat{\beta}_i - \beta_i)\, \nu_{ij}^{(-1)} (\hat{\beta}_j - \beta_j),\ i, j = 1, 2, \ldots, n, \qquad (1)$$

où $||\nu_{ij}^{(-1)}||$ est la matrice inverse de la matrice $||\nu_{ij}||$.

Sous certaines conditions, cette statistique sera approximativement distribuée suivant une loi de χ^2 à n degrés de liberté et, de toutes manières, les grandes valeurs seront significatives d'un désaccord avec l'hypothèse nulle.

Si la valeur de la statistique X_n^2 n'est pas significative au niveau adopté l'hypothèse nulle d'une répartition suivant la loi normale de paramètres β_i, $i = 1, 2, \ldots, n$ pourra être acceptée. Dans le cas contraire, il conviendra de distinguer entre l'éventualité où le désaccrod peut être levé simplement par un meilleur choix des paramètres de position et d'échelle et celle où un tel choix ne suffit pas. Pour éviter cette ambiguïté il est clair qu'on peut calculer la statistique (1) en tenant compte de l'estimation préalable des paramètres de position et d'échelle, sachant que cette estimation entraîne une modification de la loi de répartition de la statistique correspondant à une réduction à $(n - 2)$ du nombre de degrés de liberté pour une loi de χ^2.

Signalons encore, qu'en cas de rejet, la méthode esquissée peut être conduite de manière à déterminer l'alternative qui s'accorde le mieux avec la série d'observations, par exemple, en modifiant tout d'abord le paramètre β_{i_1} pour lequel le rapport $(\hat{\beta}_{i_1} - \beta_{i_1})^2 / \nu_{i_1 i_1}$ est le plus élevé.

Cela étant, les directions dans lesquelles les tests de normalité ont été développés peuvent être caractérisées comme étant les cas particuliers du test précédent où les paramètres β_i sont soit les moments centrés réduits d'ordre supérieur, soit les espérances mathématiques des statistiques d'ordre, ou encore celles des valeurs pour ces statistiques de la fonction de répartition de la loi normale.

3. LES TESTS DE NORMALITE BASES SUR LES MOMENTS CENTRES REDUITS

Appliquée aux moments centrés réduits, la méthode qui vient d'être décrite conduit à une statistique de la forme :

$$X^2_{n-2} = \Sigma \hat{\beta}_i \nu^{(-1)}_{ij} \hat{\beta}_j \; ; i, j = 3, 4, \ldots, n \tag{1}$$

pour laquelle on peut supposer que les $\hat{\beta}_i$ sont des transformées normales de moyennes nulles des estimations de ces moments et où la matrice $||\nu_{ij}||$ est la matrice des convariances de ces transformées.

D'autre part, en raison de l'indépendance de chaque moment centré empirique d'ordre impair avec tous les moments centrés empiriques d'ordre pair, on peut prévoir que la corrélation entre les variables $\hat{\beta}_i$ relatives aux moments d'ordre impair et celles liées aux moments d'ordre pair sera négligeable malgré la corrélation introduite par l'emploi d'une estimation de l'écart-type dans le calcul des moments centrés réduits. On en déduit que la statistique X_{n-2} se décompose en la somme de deux statistiques pratiquement indépendantes $X^2_{1,n-2}$ et $X^2_{2,n-2}$, l'une dépendant des moments d'ordre impair et l'autre, des moments d'ordre pair.

Il s'ensuit que le test de normalité peut s'effectuer en procédant séparément à un test de symétrie des observations au moyen de la statistique $X^2_{1,n-2}$ et à un test de l'étalement de ces observations au moyen de la statistique $X^2_{2,n-2}$ en déterminant ensuite la probabilité associée au test global, par exemple, au moyen du test de Fisher (cf. [2] p.99). En d'autres termes, si α_1 est la probabilité pour laquelle la valeur de la statistique $X^2_{1,n-2}$ est significative et α_2, la probabilité correspondante pour $X^2_{2,n-2}$, la probabilité associée au test global se détermine au moyen de la statistique (en logarithmes népériens) :

$$X^2 = -2(\log \alpha_1 + \log \alpha_2) \tag{2}$$

qui possède une répartition de χ^2 à quatre degrés de liberté lorsque les probabilités α_1 et α_2 sont indépendantes.

Une version simplifiée de ce test peut s'obtenir en combinant de la manière indiquée les tests en $\sqrt{b_1}$ et b_2 de K. Pearson.

Rappelons à ce sujet que l'on a :

$$\sqrt{b_1} = m_3/(m_2)^{3/2} \quad \text{et} \quad b_2 = m_4/(m_2)^2 \tag{3}$$

avec

$$m_j = \Sigma (x_i - \bar{x})^j/n \quad , \quad j = 2, 3, 4,$$

où $\bar{x} = \Sigma\, x_i/n$ et où x_i, $i = 1, 2, \ldots, n$ sont les n observations, et que les probabilités associées aux valeurs de $\sqrt{b_1}$ et de b_2 peuvent se déterminer grâce aux tables publiées par E.S. Pearson et H.O. Hartley [3] ainsi qu'aux travaux de D'Agostino [4] et de D'Agostino et Tietjen [5].

L'emploi d'une forme simplifiée du test signifie qu'on néglige apparemment la part de l'information contenue dans les moments centrés empiriques d'ordre supérieur à 4. Toutefois, en raison de la corrélation qui lie le moment empirique d'ordre 3 aux moments empiriques d'ordre impair supérieur à 3 et celle qui lie le moment empirique d'ordre 4 aux moments empiriques d'ordre pair supérieur à 4, ce n'est qu'une fraction de cette part qui, en réalité, se trouve négligée.

4. LES TESTS DE NORMALITE BASES SUR LES STATISTIQUES D'ORDRE.

Si l'on considère les statistiques d'ordre déduites de la série ordonnée des observations, le test de normalité revient à vérifier au moyen de la statistique 2. (1), convenablement définie, la compatibilité de ces statistiques d'ordre avec leurs espérances mathématiques.

Si x_m, $m = 1, 2, \ldots, n$ désigne cette fois ces statistiques d'ordre et si μ et σ sont la moyenne et l'écart-type de la loi normale considérée, les variables $u_m = (x_m - \mu)/\sigma$ seront les statistiques d'ordre de la loi normale réduite. Si de plus, $\bar{u}_m$ sont des espérances mathématiques de ces variables et si $||\nu_{m m'}||$ désigne la matrice des covariances de ces statistiques, la statistique 2. (1) devient alors :

$$\begin{aligned} X^2 &= \Sigma (u_m - \bar{u}_m)\, \nu^{(-1)}_{mm'} (u_{m'} - \bar{u}_{m'}) \\ &= \Sigma (x_m - \mu - \sigma\, \bar{u}_m)\, \nu^{(-1)}_{m m'} (x_{m'} - \mu - \sigma\, \bar{u}_{m'})/\sigma^2 \qquad (1) \end{aligned}$$

où $||\nu^{(-1)}_{m m'}||$ désigne la matrice inverse dc la matrice $||\nu_{m m'}||$ et pour laquelle ici aussi les grandes valeurs seront significatives d'un désaccord entre la série d'observations et l'hypothèse nulle. Par ailleurs, il va de soi que le remplacement de μ et de σ par des estimations de la moyenne et de l'écart-type aura des conséquences analogues à celles de la même opération sur la statistique 2. (1).

Dans la pratique, un tel test de normalité ne semble pas avoir été développé jusqu'à présent, sans doute en raison du fait que la matrice $||\nu_{m m'}||$ n'est connue que pour les faibles valeurs de n et surtout par suite de la complexité des calculs qu'entraîne l'inversion de cette matrice.

Par contre, plusieurs tests de normalité peuvent être considérés comme une forme simplifiée du test précédent puisque les statistiques correspondantes sont de la forme :

$$X = \Sigma\, a_m\, x_m / s$$

c'est-à-dire, le rapport à l'écart-type empirique s d'une fonction linéaire des statistiques d'ordre.

Pour cette dernière statistique, on notera qu'en raison de la symétrie de la loi normale, elle doit pouvoir se mettre sous la forme :

$$X = \Sigma\, a_m\, (x_{n+1-m} - x_m)/s \; ; \; m = 1, 2, \ldots, (n-1)/2 \text{ ou } n/2 \qquad (3)$$

d'où il résulte qu'elle sera sensible à la fois dans le cas d'alternatives sans symétrie ou dans lesquelles l'étalement n'est pas conforme à celui de la loi normale. Une telle statistique doit donc, en principe et dans une certaine mesure, pouvoir remplir un rôle analogue à celui de la statistique 3. (1).

Pour le choix des valeurs les meilleures à attribuer aux constantes a_m, on retiendra que lorsque celles-ci sont fixées, l'espérance mathématique du numérateur de la statistique (3) est, à un facteur bien déterminé près, égal à la vraie valeur de l'écart-type σ. En outre, il est évident que le meilleur choix sera celui qui rend var X minimal. Il s'ensuit qu'on s'en rapprochera en plaçant au numérateur de la statistique (3) l'estimateur linéaire de σ à moindre variance.

Si l'on fait alors $s^2 = \Sigma (x_m - \bar{x})^2/n$, avec $\bar{x} = \Sigma\, x_m/n$, $W = X^2$ devient la statistique du test de Shapiro et Wilk [6] dont les petites valeurs sont significatives d'une répartition non normale et qui grâce aux tables de fractiles existantes peut être utilisée pour des échantillons d'effectifs $n = 3\ (1)\ 50$.

Pour $n > 50$, le test a été adapté par D'Agostino [7] en utilisant au numérateur de la statistique (3) l'estimateur linéaire de Downton [8], ce qui, à un facteur près, revient à poser :

$$a_m = (n + 1)/2 - m \tag{4}$$

La statistique a, en outre, été transformée de manière à avoir une moyenne nulle et une variance unité. Ce prolongement est particulièrement opportun puisque l'estimateur linéaire de Downton est asymptotiquement un estimateur linéaire de σ à moindre variance.

A ces deux statistiques, il convient d'ajouter la statistique W' de Shapiro et Francia [9] construite avec un estimateur linéaire de σ dérivé à partir d'un principe de moindres carrés approché et qui semble avoir des propriétés analogues.

Par contre, la statistique du test de Geary et celle du test de David [3] se déduisent de la statistique (3) à l'aide de nouvelles simplifications.

En faisant $a_m = 1/n$, $m = 1, 2, \ldots$, la statistique (3) devient, pour n pair, lasstatistique du test de Geary, tandis que pour n impair, la statistique de ce test prend, avec $a_m = 1/n$, la forme :

$$X + |x_{m_o} - \bar{x}|/ns \tag{5}$$

où $m_o = (n-1)/2 + 1$.

Enfin, avec $a_1 = 1$, $a_m = 0$, $m = 2, 3, \ldots$ et $s^2 = \Sigma (x_m - \bar{x})^2/(n-1)$; la statistique (3) se confond avec celle du test de David soit :

$$X = (x_n - x_1)/s \tag{6}$$

5. TESTS DE NORMALITE APPARENTES AU TEST DE KOLMOGOROV.

Si l'on considère les valeurs prises par la fonction de répartition $F(x)$ pour les diverses statistiques d'ordre, on obtient un nouveau groupe de n variables $F_m = F(x_m)$ pour lesquelles on a :

$$E(F_m) = m/(n+1) \tag{1}$$

et

$$\text{cov}\,(F_m, F_{m'}) = [E(F_m)]\,[1 - E(F_{m'})]/(n+2) \tag{2}$$

Il s'ensuit que si $F(x)$ est donnée et si on pose :

$$\nu_{mm'} = \text{cov}\,(F_m, F_{m'}) \tag{3}$$

pour vérifier l'adéquation de la fonction $F(x)$ à la série d'observations, on considèrera la statistique :

$$X^2 = \Sigma\,[F_m - m/(n+1)](\nu_{mm}^{-1}),\ [F_{m'} - m'/(n+1)] \tag{4}$$

où $||\nu_{mm'}^{(-1)}||$ désigne la matrice inverse de la matrice $||\nu_{mm'}||$ et dont les grandes valeurs seront à nouveau significatives d'une discordance entre la fonction de répartition $F(x)$ et la série d'observations.

Si, de plus, on adopte pour $F(x)$ la fonction de répartition $F(x)$ obtenue en ajustant une loi normale à la série d'observations, on obtiendra la statistique :

$$X^2 = \Sigma\,[\hat{F}_m - m/(n+1)](\hat{\nu}_{mm}^{-1}),\ [\hat{F}_{m'} - m'/(n+1)] \tag{5}$$

dont la répartition subira par rapport à celle de la statistique (4) une modification analogue à celle de la réduction à $(n-2)$ degrés de liberté dans le cas d'une variable X^2.

Comme sous l'hypothèse nulle la valeur prise par la statistique (4) permet d'associer une probabilité à l'ensemble de variables $d_m = F_m - m/(n+1)$, tout comme la statistique (5) le fait pour l'ensemble de n variables

$$\hat{d}_m = \hat{F}_m - m/(n+1),$$

tout procédé qui consiste à choisir une ou deux parmi ces n variables pour en faire la statistique d'un nouveau test peut être considéré comme une forme simplifiée du test optimal.

Dans cette optique, le test de Kolmogorov (cf. [10] p. 68) basé sur la statistique).

$$\Delta = \text{Max}\ [(F_m - (m-1)/n) \text{ et } (m/n - F_m)] \tag{6}$$

constitue une forme simplifiée du test optimal basé sur la statistique (4).

De même, les tests de normalité de Lilliefors [11] et de Kuiper [12] pour lesquelles les statistiques sont respectivement :

$$D = \text{Max } |\hat{F}_m - m/n| \tag{7}$$

et

$$V = \text{Max } (m/n - \hat{F}_m) - \text{Min } (m/n - \hat{F}_m) \tag{8}$$

appellent la même remarque relativement à la statistique (5).

Il convient de noter que ces tests présentent en outre le défaut de réunir, pour faire le choix, dans un même échantillon des variables dont les variances sont inégales, ce qui entraîne un biais dans les conclusions de ces tests.

Quoiqu'il en soit, on voit immédiatement qu'on se trouve en présence de statistiques simplifiées à un degré semblable à celui de la statistique 4. (6) de David.

6. DU CHOIX D'UN TEST DE NORMALITE.

En résumé, il est apparu que toutes les statistiques des tests de normalité considérés sont, à des degrés divers, des formes simplifiées de la statistique optimale de leur catégorie. Aucune ne peut donc être considérée comme définitivement la meilleure et on peut concevoir que certains progrès pourraient encore être accomplis dans ce domaine.

Il s'ensuit que, pour l'instant, lors de l'examen de la normalité de la répard'une série d'observations, le choix doit se porter sur les statistiques dont la forme est la plus proche de la statistique optimale. A ce titre, il convient de retenir le test basé sur la combinaison des deux statistiques de K.Pearson ainsi que ceux qui utilisent les statistiques de Shapiro-Wilk-D'Agostino-Francia.

Une telle conclusion se trouve d'ailleurs confirmée par les résultats des expériences Monte Carlo effectuées par Shapiro, Wilk et Chen [13] pour mettre en évidence la puissance des divers tests à l'égard d'un ensemble d'alternatives.

Pour l'illustrer, nous avons appliqué chaque test à une série de 45 observations de la température de l'air à Bruxelles-Uccle en janvier à l'altitude de 220 mb, apparemment réparties suivant une loi doublement exponentielle ($1^{ère}$ asymptote de Fisher-Tippett). Les niveaux α pour lesquels les valeurs obtenues par ces statistiques sont significatives ont été les suivants :

K. Pearson	Shapiro-Wilk	D'Agostino	Geary	David	Lilliefors	Kuiper
$< 0{,}001$	$< 0{,}01$	$< 0{,}02$	0,10	0,10	0,10	0,10

Il est clair que dans le cas d'alternatives particulières, il peut arriver que les statistiques les moins élaborées soient plus puissantes que les autres. C'est notamment le cas pour la statistique de David lorsque l'hypothèse alternative est la loi uniforme, loi pour laquelle le numérateur de la statistique est exhaustif (cf. [14] p. 28). Mais ceci ne saurait être un argument pour le cas général puisque si l'alternative est précisée, il existe une statistique optimale bien déterminée fournie par la méthode de Neyman-Pearson, pour séparer l'hypothèse nulle de l'hypothèse alternative.

En conclusion, la démarche qui dans la pratique paraît la plus raisonnable consiste dans l'emploi du test conjoint sur les deux statistiques de K. Pearson et du test sur les statistiques de Shapiro-Wilk-D'Agostino qui sont les plus sévères, et, en cas de rejet, dans un choix de l'hypothèse alternative s'appuyant sur les valeurs données par les statistiques de K. Pearson.

BIBLIOGRAPHIE

[1] MORICE E. – Tests de normalité d'une distribution observée, *Revue de Statistiques Appliquée,* 1972, vol. XX, n°2. p.5-35.

[2] FISHER R.A. – Statistical Methods for Research Workers, Oliver and Boyd, London, 1958, 13th Ed.

[3] PEARSON E.S. and HARTLEY H.O. – Biometrika Tables for Statisticians, Vol. 1, Cambridge University Press. 1966.

[4] D'AGOSTINO R.B. – Transformation to normality of the null distribution of g_1, *Biometrika,* 1970, 57, p. 679-681.

[5] D'AGOSTINO R.B. and TIETJEN G.L. – Simulation probability points of b_2 for small samples, *Biometrika,* 1971, 58, p. 669-672.

[6] SHAPIRO S.S. and WILK M.B. – An analysis of variance test for normality (complete samples), *Biometrika,* 1965, 52, p. 591-611.

[7] D'AGOSTINO R.B. – An omnibus test for normality for moderate and large size samples, *Biometrika,* 1971, 58, p.341-348.

[8] DOWNTON F. – Linear estimates with polynomial coefficients, *Biometrika,* 1966, 58, p.129-141.

[9] SHAPIRO S.S. and FRANCIA R.S. – An approximate analysis of variance test for normality, *Journ. Am. Stat. Assoc.,* 1972, 67, 337, p.215-216.

[10] VAN DER WAERDEN B.L. *Statistique Mathématique,* 1967, Dunod, Paris.

[11] LILLIEFORS H.W. – On the Kolmogorov-Smirnov test for normality with mean and variance unknown, *Journ. Am. Stat. Ass.,* 1967, p. 393-402.

[12] LOUTER S.A. and KOERTS J. – On the Kuiper test for normality with mean and variance unknown, *Stat. Neerl.* 1970, p. 83-87.

[13] SHAPIRO S.S., WILK M.B. and CHEN H.J. – A comparative study of various test for normality, *Journ. Am. Stat. Ass.,* 1968, 63, p, 1343-1372.

[14] KENDALL M.G. and STUART A. The Advanced Theory of Statistics, Vol. 2, 1967, Griffin, London.