

Tests for Normality: Comparison of Powers

Test di Normalità: Confronto delle Potenze

Daniele Coin¹ and Roberto Corradetti¹

Dipartimento di Statistica e Matematica Applicata, Università di Torino
e-mail: coin@econ.unito.it

Abstract In questo lavoro viene stimata per simulazione la potenza di alcuni test di normalità rispetto ad una vasta gamma di distribuzioni alternative. L'obiettivo è identificare la statistica con la miglior capacità di identificazione della non normalità. La scelta delle procedure da includere nell'analisi è stata guidata da criteri di grande diffusione, recente presentazione e risultati emersi in precedenti lavori. In seguito a questa analisi il test di Shapiro-Wilk è risultato complessivamente il più potente, è emersa inoltre una sostanziale inefficacia di test molto diffusi come lo Jarque-Bera.

Keywords: test for normality, power, alternative distribution

1. Introduction

An important issue in Statistics is to determine whether the distribution from which a sample is drawn is normal. This short paper presents simulation study results on of various test for normality power with composite null hypothesis, which means μ and σ unknown. The aim is to determine the overall most powerful test of normality. Interesting reviews of the main tests are contained in the book of D'Agostino and Stephens (1986) and Thode Jr. (2002). Some comparisons were already done in two old papers by Shapiro and J. (1968) and by Pearson and Bowman (1977), due to the fact that some tests were presented more recently and that automatic computation is evolved, the necessity to repeat this analysis emerges.

Seven tests were used for this comparison, they were chosen for modernity or wide diffusion, they are summarized in section 2. These tests' empirical power, with $\alpha = 0.05$, was calculated simulating 10000 samples for many alternative distributions, catalogued in section 3. The main results are provided in section 4.

2. The tests used

We considered only test for normality with composite null hypothesis, formally given a sample $\mathbf{X}_n = (X_1, \dots, X_n)$ drawn from X , we test $H_0 : \mathbf{X} \sim N(\mu, \sigma)$ (with any μ and any σ). The goodness of fit statistic tests for normality can be catalogued in four categories.

The first category is composed by statistics based on different distances between theoretical distribution function ($\Phi_X(x)$), given by the null hypothesis, and the empirical distribution function ($F_n(x)$) of the sample ($X_{(i)}$ will denote the i ordered value of the sample $\mathbf{X}_n$ through out the paper). We considered the statistic proposed by Kolmogorov (1933) for its wide diffusion, it is defined as

$$K = \sup_{\mathbf{X}} |\Phi_X(x) - F_n(x)| \quad (1)$$

Many authors (see D'Agostino and Stephens (1986) pag. 370-374) suggest the use of the statistics proposed by Anderson and Darling (1954), defined as

$$A^2 = -n - \sum_{i=1}^n (2i-1) \frac{\ln P_i - P_{n+1-i}}{n} \quad (2)$$

where $P_i = \Phi_Z(Z_{(i)}) = \int_{-\infty}^{Z_{(i)}} \frac{e^{-\frac{t^2}{2}}}{\sqrt{2\pi}}$ and $Z_{(i)} = \frac{X_{(i)} - \bar{X}}{S}$ is the standardized i ordered value of the sample $\mathbf{X}_n$, $\bar{X}$ is the sample mean and S is the sample standard deviation.

The second category contains statistics based on skewness and Kurtosis. We define the standardized third and fourth moments of the sample $\mathbf{X}$ as $\sqrt{b_1} = \frac{m_3}{m_2^{3/2}}$ and $b_2 = \frac{m_4}{m_2^2}$

where $m_r = \frac{\sum_{i=1}^n (X_i - \bar{X})^r}{n}$. In this family we used the test proposed by D'Agostino and Pearson (1973) for its wide diffusion, it is defined as

$$K^2 = X^2(\sqrt{b_1}) + X^2(b_2) \quad (3)$$

where $X(\sqrt{b_1})$ and $X(b_2)$ are standardized normal equivalent deviates. Another test used is the one proposed by Bowman and Shenton (1975) (it is known as Jarque-Bera too), defined as

$$B = \frac{n}{6} \left(b_1 + \frac{(b_2 - 3)^2}{4} \right) \quad (4)$$

Under the null hypothesis both (3) and (4) are asymptotically distributed as a chi square variable with two degrees of freedom.

The third category contains a class of tests, generalizing the classical Pearson's χ^2 , based on the deviation of a density estimator from the true density. While traditional tests as Pearson's one are usually not recommended for testing the composite hypothesis of normality due to their inferior power properties compared to other tests (for details see D'Agostino and Stephens (1986) and Thode Jr. (2002)), it is interesting to consider the special case of a statistic derived from the projection density estimators see Neyman (1937). Suppose that $X \sim f \in L^2$ that is a Hilbert separable space. For a fixed positive integer k let $E_k = \{e_0, e_1, \dots, e_k\}$ be an orthonormal system with $e_0 \equiv 1$ in L^2 . We define

$$FTF = n \|f_n - 1\|^2 \quad (5)$$

where $\|\cdot\|$ denotes L^2 norm and f_n is the density estimator of f on E_k . For details and computations in the special case of composite hypotheses see Rayner and Best (1989) and Rayner and Best (1990).

In the category containing statistics based on regression we considered the test proposed by Shapiro and Wilk (1965) defined as

$$W = \frac{[\sum_{i=1}^n c_i X_{(i)}]^2}{nm_2} \quad (6)$$

where $\mathbf{c}' = \frac{\boldsymbol{\alpha}' \mathbf{V}^{-1}}{\sqrt{\boldsymbol{\alpha}' \mathbf{V}^{-1} \mathbf{V}^{-1} \boldsymbol{\alpha}}}$ and $\boldsymbol{\alpha}_n = (\alpha_1, \dots, \alpha_n)$ denote the vector of n expected values of standard normal order statistics, and $\mathbf{V} = \{v_{ij}\}$ the corresponding $n \times n$ covariance matrix. We considered too the test proposed by D'Agostino (1971), defined as

$$D = \frac{\sum_{i=1}^n \left(i + \frac{n+1}{2}\right) X_{(i)}}{n^2 \sqrt{m_2}} \quad (7)$$

3. Alternative Distributions Studied

Our aim is to study the behaviour of the test summarized above in different situations. Thus we included in our study different alternative distributions, particularly we considered: *symmetric distributions*, the Beta distribution with shape parameters $a = 1, b = 1$ (d1), $a = 2, b = 2$ (d2), $a = 5, b = 5$ (d3), the Student's t distribution with 1 (d4), 2 (d5), 4 (d6) and 10 (d7) degrees of freedom; *asymmetric distributions*, the Beta distribution with shape parameters $a = 2, b = 1$ (d8), $a = 3, b = 2$ (d9), $a = 3, b = 1$ (d10), $a = 1, b = 3$ (d11) and $a = 5, b = 2$ (d12), the Log-normal distribution with $\ln(\mu) = 0, \ln(\sigma) = 0.1$ (d13), $\ln(\mu) = 0, \ln(\sigma) = 1$ (d14) and $\ln(\mu) = 0, \ln(\sigma) = 5$ (d15), the Chi square distribution with 1 (d16), 2 (d17), 4 (d18) and 10 (d19) degrees of freedom, the Weibull distribution with scale parameter equal to 1 and shape parameter equal to 0.5 (d20), 1 (d21), 2 (d22) and 4 (d23); *fifty percent mixed bimodal distributions*, two Beta distributions $a = 1, b = 3$ and $a = 3, b = 1$ (d24), two Normal distributions with $\mu = 0, \sigma = 1$ and $\mu = 5, \sigma = 1$ (d25)

The empirical power of these tests with $\alpha = 0.05$, was calculated simulating 10000 samples for any alternative distributions. Originally we considered samples of different sizes from 50 to 1000, but due to the obvious fact the power of all the tests increase with large samples, we focus only on the case of 50 units samples reported in Table 1.

	K	A^2	K^2	B	FTF	W	D
d(1)	0.2455	0.5725	0.7945	0.0000	0.4420	0.7435	0.5490
d(2)	0.4480	0.7085	0.3395	0.1050	0.6860	0.8380	0.0615
d(3)	0.0780	0.1290	0.2280	0.0010	0.1130	0.1475	0.2780
d(4)	0.1300	0.1900	0.1345	0.0080	0.1750	0.2080	0.1150
d(5)	0.6430	0.8925	0.5730	0.4645	0.8900	0.9520	0.2295
d(6)	0.6520	0.8870	0.5505	0.4520	0.8860	0.9590	0.2175
d(7)	0.2485	0.3845	0.2395	0.1495	0.3870	0.4915	0.0775
d(8)	0.0465	0.0485	0.0450	0.0020	0.0590	0.0375	0.0760
d(9)	0.0770	0.1005	0.1205	0.0885	0.0800	0.1185	0.0785
d(10)	0.9945	1.0000	0.9980	0.9965	1.0000	1.0000	0.9855
d(11)	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
d(12)	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
d(13)	0.9625	0.9975	0.9670	0.9560	0.9980	0.9985	0.8970
d(14)	0.7195	0.8945	0.7960	0.7570	0.8720	0.9540	0.5830
d(15)	0.3065	0.4590	0.4575	0.4050	0.4200	0.5770	0.2650
d(16)	0.9930	0.9975	0.9930	0.9940	0.9960	0.9970	0.9985
d(17)	0.7930	0.8670	0.8655	0.8730	0.7660	0.8720	0.9130
d(18)	0.2910	0.4080	0.5080	0.4995	0.2800	0.4725	0.5290
d(19)	0.0980	0.1295	0.2055	0.1895	0.0770	0.1625	0.1840
d(20)	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
d(21)	0.9580	0.9970	0.9660	0.9580	0.9970	1.0000	0.8905
d(22)	0.2075	0.2920	0.2870	0.2180	0.2690	0.4105	0.1250
d(23)	0.0490	0.0485	0.0320	0.0115	0.0420	0.0415	0.0505
d(24)	0.5990	0.9260	0.9885	0.0035	0.7960	0.9780	0.4345
d(25)	0.9870	0.9995	0.9995	0.0245	0.9990	0.9985	0.0125

Table 1: Power of the Normality Tests for different alternative Hypotheses, 50 units samples.

4. Results

The analysis of table 1 lead us to the following general findings.

All the considered tests behave well against asymmetric alternative distributions, in this case there are not considerable differences in the estimated power. Considering symmetric alternative distributions there are huge dissimilarities. The statistics defined in (4) and in (7) are very insensitive, while (2), (5) and particularly (6) behave well.

Considering bimodal alternative distributions the tests based on (7) and (4) are again very insensitive while (2), (5) and (6) again behave well; in this case even the (1) has a good behaviour.

It is possible to assert the test based on (6) provides a generally superior omnibus measure of non normality, the second one is the (2). The statistic defined at (5) behaves well but it is not the best, considering its very high time of computation, we do not suggest to use it.

References

- Anderson T.W. and Darling D.A. (1954) A test of goodness of fit, *Journal of the American Statistical Association*, 49, 141–160.
- Bowman K.O. and Shenton B.R. (1975) Omnibus test contours for departures from normality based on $\sqrt{b_1}$ and b_2 , *Biometrika*, 62, 243–250.
- D’Agostino R. and Stephens M. (1986) *Goodness-of-Fit Techniques*, Marcel Dekker, New York.
- D’Agostino R.B. (1971) An omnibus test of normality for moderate and large sample size, *Biometrika*, 58, 341–348.
- D’Agostino R.B. and Pearson E.S. (1973) Testing for departures from normality, *Biometrika*, 60, 613–622.
- Kolmogorov A. (1933) Sulla determinazione empirica di una legge di distribuzione, *Giornale dell’Istituto Italiano Attuari*, 4, 83–91.
- Neyman J. (1937) Smooth test for goodness of fit, *Scandinavian Aktuar*, 20, 119–128.
- Pearson E. S. D’Agostino R.B. and Bowman K.O. (1977) Tests for departure from normality, comparison of powers, *Biometrika*, 64, 231–246.
- Rayner J.C.W. and Best D.J. (1989) *Smooth Tests of Goodness of Fit*, Oxford University Press.
- Rayner J.C.W. and Best D.J. (1990) Smooth tests of goodness of fit: an overview, *International Statistical Review*, 58, 9–17.
- Shapiro S.S. and Wilk M.B. (1965) An analysis of variance test for normality (complete samples), *Biometrika*, 52, 591–611.
- Shapiro S. S. W.M.B. and J. C.H. (1968) A comparative study on various test for normality, *Journal of the American Statistical Association*, 63, 1343–1372.
- Thode Jr. H. (2002) *Testing for Normality*, Marcel Dekker, New York.