A NEW TEST FOR NORMALITY
AND INDEPENDENCE

Wiktor Ejsmont

Abstract. In this paper, a new test of normality and independence is proposed. This test is designed through a multivariate empirical characteristic by considering a result form [Ejsmont 2016].

1. Introduction and preliminaries

The original motivation for this paper comes from a desire to understand the results about the characterization of normal distribution which were shown in [Ejsmont 2016], where the author provides characterizations of the normal distribution by using a certain invariance of the noncentral chi-square distribution. Our construction of a new test of normality is based on the following characterization of normal distribution in terms of independent random vectors; see [Ejsmont 2016].

Theorem 1. *Let* $(X_1, ..., X_m, Y)$ *and* $(X_{m+1}, ..., X_n, Z)$ *be independent random vectors with all moments, where* X_i *are nondegenerate, and let statistic* $\sum_{i=1}^{n} a_i X_i + Y + Z$ *have a distribution which depends only on* $\sum_{i=1}^{n} a_i^2$, *where* $a_i \in \mathbb{R}$ *and* $1 \leq m < n$. *Then* X_i *are independent and have the same normal distribution with zero means and* $cov(X_i, Y) = cov(X_i, Z) = 0$ *for* $i \in \{1, \ldots, n\}$.

In this paper we denote a probability measure of random variable X. If X is a random vector defined on a probability space (Ω, Σ, P), then the expected

Wiktor Ejsmont
   Wrocław University of Economics
   email: wiktor.ejsmont@ue.wroc.pl
   ORCID: 0000-0002-4516-9374

value of X, denoted by $E(X)$ is defined as the Lebesgue integral

$$E(X) = \int_{\Omega} X(\omega)\, \mathbb{P}(d\omega).$$

A characteristic function is simply the Fourier transform, in probabilistic language. The characteristic function of a probability measure μ on $\mathbb{R}^n$ is the function $\varphi{:}\mathbb{R}^n{\rightarrow}\mathbb{C}$

$$\varphi_X(t_1,...,t_n) = E[\exp(itX)] = \int_{R^n} X \exp(itX) d\mu(x).$$

When we speak of the characteristic function φ_X of a random vector X, we have the characteristic function φ_{μ_X} of its distribution μ_X in mind. Apparently, it is not accidental that the characteristic function encodes the most important information about the associated random variables. A random vector $X = (X_1, ..., X_k)$ is said to have the multivariate normal distribution if there is a *n*-vector ξ and a symmetric, positive semidefinite $n \times n$ matrix Σ, such that the characteristic function of X is

$$\varphi_X(t_1,...,t_n) = \exp\left(i(t_1,...,t_n)^T \xi - \tfrac{1}{2}(t_1,...,t_n)^T \sum (t_1,...,t_n)\right)$$

2. Test statistic

In this section we will propose a new test statistic for independence and normality, which is a multidimensional analog of [Epps, Lawrence 1983]. We state this test formally without empirical verification of its critical value and power. The distance between the empirical and theoretical characteristic function provides a new approach to the problem of testing the joint independence and normality of random vectors. Our approach is based on the following reasoning (the author believes that the proposition of the test below is unknown). Given an observed random matrix

$$\left(X_{i,j}\right)_{i,j=1}^n = \begin{pmatrix} X_{1,1} & X_{1,2} & \cdots & X_{1,n} \\ X_{2,1} & X_{2,2} & \cdots & X_{2,n} \\ \ddots & \ddots & \ddots & \ddots \\ X_{m,1} & X_{m,2} & \cdots & X_{m,n} \end{pmatrix}, \tag{1}$$

for which we normalize its columns (because the test should be invariant under changes in the location and scale of data):

$$\left(\widetilde{X}_{i,j}\right)_{i,j=1}^{m,n} = \begin{pmatrix} X_{1,1} - \overline{X}_1 / S_1 & X_{1,2} - \overline{X}_2 / S_2 & \cdots & X_{1,n} - \overline{X}_n / S_n \\ X_{2,1} - \overline{X}_1 / S_1 & X_{2,2} - \overline{X}_2 / S_2 & \cdots & X_{2,n} - \overline{X}_n / S_n \\ \ddots & \ddots & \ddots & \ddots \\ X_{m,1} - \overline{X}_1 / S_1 & X_{m,2} - \overline{X}_2 / S_2 & \cdots & X_{m,n} - \overline{X}_n / S_n \end{pmatrix}, \tag{2}$$

where $\overline{X}_i = \sum_{j=1}^{m} X_{j,i} / m$ and $S_i = \sqrt{\sum_{j=1}^{m} \left(X_{j,i} - \overline{X}_i\right)^2}$, we propose a hypothesis in a form that can be applied both to independence and normality: H_0: all columns of $\left(\widetilde{X}_{j,i}\right)_{j,i=1}^{m,n}$ are independent and have a normal distributions vs. H_1: H_0 is not true.

The test statistic proposed by [Epps, Lawrence 1983] is based on the weighted integral, i.e. the squared modulus of the difference between the characteristic functions of the sample and of the theoretical normal distribution. Now we use a similar construction, namely the squared modulus of the difference between the characteristic functions of the sample and the function form the proof of Corollary 3.3 from [Ejsmont 2016] i.e.

$$E \exp\left(i \sum_{i=1}^{n} a_i X_i \right) = \exp\left(\sum_{i=1}^{n} \frac{a_i^2}{2} \right),$$

for all $a_i \in \mathbb{R}$. We adopt this test because Theorem 1 states that the distribution is normal if the corresponding statistic has a constant distribution on the sphere. Building a test based on the above formula is some approximation of this condition.

We define our proposed test of the composite hypothesis of normality and independence as the integral

$$\int_{R^n} \left| \phi_{n,m}(t) - \phi(t) \right|^2 dG(t),$$

where

- $t = (t_1, \ldots, t_n) \in \mathbb{R}$,

- $\phi_{n,m}(t)$ is the empirical multivariate characteristic function of the data $\left(\widetilde{X}_{j,i}\right)_{j,i=1}^{m,n}$ which is defined by

$$\phi_{n,m}(t) = \sum_{k=1}^{m} \exp\left(i \sum_{p=1}^{n} t_p \widetilde{X}_{k,p} \right) / m,$$

- $\phi(t) = \exp\left(-\sum_{j=1}^{n} t_i^2 / 2\right)$ is the multivariate theoretical characteristic function,

- $dG(t)$ is the weight $dG(t) = \exp\left(-\beta\sum_{j=1}^{n} t_i^2 / 2\right)$ with $\beta > 0$. We apply this function because it is a multidimensional analog of weight suggested by [Epps, Lawrence 1983]. The amount of information provided by β is too great in the real applications and it is too hard to find the appropriate β, so it should be selected through an empirical experiment as in [Epps, Lawrence 1983].

By $\tilde{X}_j$ we denote the j-th row of the matrix $\left(\tilde{X}_{j,i}\right)_{j,i=1}^{n,n}$ i.e. $\tilde{X}_j = \left(\tilde{X}_{j,1}, ..., \tilde{X}_{j,n}\right)^T$.

Proposition. *A statistic T has the form*

$$T = \frac{(2\pi / \beta)^{n/2}}{m^2} \sum_{k,r=1}^{m} \exp\left(-\left\|\tilde{X}_k - \tilde{X}_r\right\|^2 / (2\beta)\right)$$
$$-2\frac{(2\pi / \beta + 1)^{n/2}}{m} \sum_{k=1}^{m} \exp\left(-\left\|\tilde{X}_p\right\|^2 / (2\beta + 2)\right) + \left(\frac{2\pi}{\beta + 2}\right)^{n/2},$$

where $\|\cdot\|$ is the Euclidean norm on $\mathbb{R}^m$.

Proof. In the proof we use the well-known identity

$$\int_R \exp\left(-t^2 \beta / 2 + xit\right)dt = \exp\left(-x^2 / (2\beta)\right)\frac{\sqrt{2\pi}}{\sqrt{\beta}}.$$

By direct calculation we obtain a formula for T

$$T = \int_{R^n} \left|\phi_{n,m}(t) - \phi(t)\right|^2 dG(t) = \int_{R^n} \left(\phi_n(t) - \phi(t)\right)\left(\overline{\phi_n(t)} - \phi(t)\right)dG(t)$$

$$= \sum_{k,r=1}^{m} \int_{R^n} \exp\left(i\sum_{p=1}^{n} t_p\left(\tilde{X}_{k,p} - \tilde{X}_{r,p}\right) - \beta\sum_{p=1}^{n} t_p^2 / 2\right)dt / m^2$$

$$-\sum_{k=1}^{m} \int_{R^n} \exp\left(i\sum_{p=1}^{n} t_p\tilde{X}_{k,p} - (1+\beta)\sum_{p=1}^{n} t_p^2 / 2\right)dt / m$$

$$-\sum_{k=1}^{m} \int_{R^n} \exp\left(i\sum_{p=1}^{n} t_p\tilde{X}_{k,p} - (1+\beta)\sum_{p=1}^{n} t_p^2 / 2\right)dt / m$$

$$+\int_{R^n} \exp\left(-(2+\beta)\sum_{p=1}^{n} t_p^2 / 2\right)dt$$

$$= \sum\nolimits_{k,r=1}^{m} \prod\nolimits_{p=1}^{n} \left(\exp\left(- (\tilde{X}_{k,p} - \tilde{X}_{k,p})^2 (2\beta) \frac{\sqrt{2\pi}}{\sqrt{\beta}} \right) \middle/ m^2 \right.$$

$$-2\sum\nolimits_{k=1}^{m} \prod\nolimits_{p=1}^{n} \left(\exp\left(- \tilde{X}_{k,p}^2 \middle/ (2\beta+2) \frac{\sqrt{2\pi}}{\sqrt{\beta}} \right) \middle/ m + \prod\nolimits_{p=1}^{n} \frac{\sqrt{2\pi}}{\sqrt{\beta+2}} \right)$$

$$= \frac{\left(2\pi / \beta\right)^{n/2}}{m^2} \sum\nolimits_{k,r=1}^{m} \exp\left(\sum\nolimits_{p=1}^{n} - (\tilde{X}_{k,p} - \tilde{X}_{k,p})^2 / (2\beta) \right)$$

$$-2\frac{\left(2\pi / (\beta+1)\right)^{n/2}}{m} \sum\nolimits_{k=1}^{m} \left(\sum\nolimits_{p=1}^{n} - \tilde{X}_{k,p}^2 / (2\beta+2) \right) + \left(\frac{2\pi}{\beta+2} \right)^{n/2}.$$

$\square$

Bibliography

Ejsmont W. (2016). *A characterization of the normal distribution by the independence of a pair of random vectors*. Statistics and Probability Letters. No. 114. Pp. 1-5.

Epps T.W., Lawrence B. (1983). *A test for normality based on the empirical characteristic function*. Biometrika. No. 70(3). Pp. 723-726.