Journal of Multivariate Analysis 129 (2014) 186–192

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

Linear transformations to symmetry

Nicola Loperfido

Dipartimento di Economia, Società e Politica, Università degli Studi di Urbino "Carlo Bo", Via Saffi 42, 61029 Urbino (PU), Italy

ARTICLE INFO

Article history:
Received 16 August 2013
Available online 9 May 2014

AMS subject classifications:
primary 62E15
secondary 15A18

Keywords:
Finite mixture
Multivariate analysis of variance
Nonrandom sampling
Singular value decomposition
Symmetrization

ABSTRACT

We obtain random vectors with null third-order cumulants by projecting the data onto appropriate subspaces. Statistical applications include, but are not limited to, the robustification of Hotelling's T^2 test against nonnormality. Our approach only requires the existence of the third-order moments and leads to normal transformed variables when the parent distribution belongs to well-known classes of sample selection models.

1. Introduction

Let $\mu = (\mu_1, \ldots, \mu_d)^T$ be the mean of a d-dimensional random vector $x = (X_1, \ldots, X_d)^T$ satisfying $E\left(\left|X_i X_j X_k\right|\right) < +\infty$ for $i, j, k = 1, \ldots, d$. The third cumulant of x is the $d^2 \times d$ matrix $\kappa_3(x) = E\left\{(x-\mu) \otimes (x-\mu)^T \otimes (x-\mu)\right\}$, where "$\otimes$" denotes the Kronecker product (see, for example, [6]). In the following, when referring to a third cumulant, we implicitly assume the existence of all the third-order moments of the corresponding random vector. The third cumulant of x is a null matrix when x is symmetric about a real vector $c \in \mathbb{R}^d$, that is if $x - c$ and $c - x$ are identically distributed. However, the converse is not necessarily true, as shown by many univariate examples. Bearing this in mind, we shall refer to random vectors whose third cumulants are null matrices as to weakly symmetric vectors. Weak symmetry, or lack of it, plays a fundamental role in probability and statistics. As a first example, the asymptotic distributions of commonly used MANOVA statistics greatly simplify when the sampled distribution is weakly symmetric [14]. As a second example, multivariate sample means admitting valid Edgeworth expansions converge to normality at a quicker rate, when the observations are weakly symmetric [32]. Similar comments hold for the asymptotic distribution of maximum likelihood estimates [36]. As a third example, theoretical and empirical results [29,30,7,5] hint that sampling distribution of Hotelling's T^2 statistic is quite robust to nonnormality, when the sampled distribution is weakly symmetric. Moreover, theoretical results in [10] imply that the Kolmogorov distance between the sampling distribution of Hotelling's statistic and the chi-squared distribution with d degrees of freedom converges to zero at a faster rate when the sampled distribution is d-dimensional, weakly symmetric, centered at the origin and has finite moments of appropriate order.

Symmetry is usually pursued by means of power transformations, primarily the Box–Cox one. Statistical applications include skewness removal from Hotelling's T^2 statistic when testing hypotheses about a multivariate mean [8,34]. However, power transformations suffer from some serious drawbacks, as pointed out by Hubert and van der Veeken [18] and Lin and Lin [22], among others. In the first place, the transformed variables are neither affine invariant nor robust to outliers. In the second place, they might not be easily interpretable nor jointly normal.

E-mail addresses: nicola.loperfido@uniurb.it, nicola.loperfido@econ.uniurb.it.

http://dx.doi.org/10.1016/j.jmva.2014.04.018
0047-259X/

For the sake of completeness, we shall mention two symmetrization techniques different from power transformations. Hall [15] studied empirical transformations for removing most of the skewness of an asymmetric statistic using a monotone and an invertible cubic polynomial. Fujioka and Maesono [11] also propose a transformation for removing skewness from U-statistics. Both transformations are limited to univariate data.

The present paper deals with the above issues by means of appropriate linear transformations, with special emphasis on Hotelling's T^2 statistic and nonrandom sampling. The approach is nonparametric in nature, since it applies to any multivariate data with finite third-order moments. It also leads to multivariate normal transformed variables, under some additional assumptions. Both real and simulated data encourage its use in statistical practice.

The rest of the paper is organized as follows. Sections 2 and 3 describe the symmetrization methods for the bivariate case and the multivariate case, respectively. Section 4 applies the method described in Section 3 to nonrandom samples from multivariate normal distributions. Sections 5 and 6 assess the practical relevance of the theoretical results in the previous sections by means of simulation studies and numerical examples, respectively. Section 6.1 contains some concluding remarks and hints for future research. All proofs are deferred to the Appendix.

2. The bivariate case

This section investigates the simplest case of linear transformations to symmetry, which involves two random variables only. We shall motivate it with the following example. Let Z_1, Z_2, Z_3 be three independent, identically distributed gamma variables. Also, let $W_1 = Z_1 - Z_3$ and $W_2 = Z_2 - Z_3$. Then W_1 and W_2 are symmetric random variables but the third cumulant of $w = (W_1, W_2)^T$ is not a null matrix. As a direct consequence, no componentwise power transformation $\left(W_1^a, W_2^b\right)^T$, with $a, b \in \mathbb{R}$, has a third cumulant which is a null matrix. However, there are three linear functions of w which are symmetric: W_1, W_2 and $W_1 - W_2$.

A natural question to ask is whether any two random variables with finite third moments X_1 and X_2 might be linearly combined to form another random variable $a_1X_1 + a_2X_2$ whose third cumulant is zero. Surprisingly enough, the answer is in the affirmative, as it can be shown constructively. When the third cumulant of either variable is zero, the linear function might be taken as the variable itself. Hence, without loss of generality, we shall assume that the third cumulants of both X_1 and X_2 are different from zero. The third cumulant $\kappa_3(W)$ of the random variable $W = wX_1 + X_2$ is $E\left\{(wY_1 + Y_2)^3\right\} = w^3E\left(Y_1^3\right) + 3w^2E\left(Y_1^2Y_2\right) + 3wE\left(Y_1Y_2^2\right) + E\left(Y_2^3\right)$, where $Y_1 = X_1 - E(X_1)$ and $Y_2 = X_2 - E(X_2)$. The third cumulant of W is then a cubic polynomial in w: $\kappa_3(W) = aw^3 + bw^2 + cw + d$, where $a = E\left(Y_1^3\right)$, $b = 3E\left(Y_1^2Y_2\right)$, $c = 3E\left(Y_1Y_2^2\right)$, $d = E\left(Y_2^3\right)$. By elementary algebra the cubic equation $ax^3 + bx^2 + cx + d = 0$ has at least one real root, that is $s + t - v$, where $s = \sqrt[3]{r + u}$, $t = \sqrt[3]{r - u}$, $u = \sqrt{q^3 + r^2}$, $q = \left(3ac - b^2\right)/\left(9a^2\right)$, $r = \left(9abc - 27a^2d - 2b^3\right)/\left(54a^3\right)$, $v = E\left(Y_1^2Y_2\right)/E\left(Y_1^3\right)$.

Sample moments provide convenient choices for the variables X_1 and X_2 in many statistical applications. For example, let M_n and Q_n be the first and second sample moment of n random variables with finite sixth-order moments. Then there is a real value w such that the third cumulant of $wM_n + Q_n$ is zero. As a direct consequence, under very general conditions, it is possible to find an affine function of the first and second sample moment which converges to the standard normal distribution at a faster rate than the standardized sample mean itself.

The method naturally generalizes to any d-dimensional random vector $x = (X_1, \ldots, X_d)^T$, with $d > 2$ and finite third-order moments. Without loss of generality, we can assume that the variance of x is a positive definite matrix and that all components of x are standardized random variables whose third moments are different from zero. Also, let $\beta_1, \ldots, \beta_d$ be d-dimensional real vectors such that the ith component of β_i is zero, for $i = 1, \ldots, d$. We can then apply the above described method to the pairs $\left(X_1, \beta_1^Tx\right), \ldots, \left(X_d, \beta_d^Tx\right)$ to obtain the weakly symmetric random variables $Y_1 = \alpha_1X_1 + \beta_1^Tx, \ldots, Y_d = \alpha_dX_d + \beta_d^Tx$, where $\alpha_i \in \mathbb{R}_0$, for $i = 1, \ldots, d$. Judiciously chosen vectors $\beta_1, \ldots, \beta_d$ will yield a vector $y = (Y_1, \ldots, Y_d)^T$ whose variance is a positive definite matrix. As an example, consider the trivariate random vector $x = (X_1, X_2, X_3)^T$ and apply the method described to the pairs (X_1, X_2), (X_2, X_3) and (X_3, X_1) to obtain the weakly symmetric random variables $Y_1 = \alpha_1X_1 + X_2$, $Y_2 = \alpha_2X_2 + X_3$, $Y_3 = \alpha_3X_3 + X_1$. It follows that the variance of $(Y_1, Y_2, Y_3)^T$ is a positive definite matrix and that $\beta_1 = (0, 1, 0)^T$, $\beta_2 = (0, 0, 1)^T$, $\beta_3 = (1, 0, 0)^T$. However, the vector $y = (Y_1, \ldots, Y_d)^T$ is not in general weakly symmetric, despite the fact that all its components are. We shall deal with this problem in the next section, by making some assumptions about the rank of the third cumulant.

3. The multivariate case

This section deals with linear transformations to symmetry of several variables, motivated by the problem of testing the hypothesis that the mean $\mu = (\mu_1, \ldots, \mu_d)^T$ of a d-dimensional random vector $x = (X_1, \ldots, X_d)^T$ equals a known real vector $\mu_0 = (\mu_{01}, \ldots, \mu_{0d})^T$. In a nonparametric setting, the natural test statistic is Hotelling's T^2, whose distribution is approximately chi-squared when the sample size is large and the parent population is not too skewed. When this is not the case, the power method approaches the problem by looking for a transformation $y = \left(X_1^{\lambda_1}, \ldots, X_d^{\lambda_d}\right)^T$ which is symmetric, where $\lambda_i \in \mathbb{R}$ for $i = 1, \ldots, d$. Under the null hypothesis, the mean of y is in general different from $\left(\mu_{01}^{\lambda_1}, \ldots, \mu_{0d}^{\lambda_d}\right)^T$, and it