Journal of Multivariate Analysis 123 (2014) 386–394

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

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

Note(s)

A note on the fourth cumulant of a finite mixture distribution

Nicola Loperfido

Dipartimento di Economia, Società e Politica, Università degli Studi di Urbino "Carlo Bo", via Saffi 42, 61029 Urbino, Italy

ARTICLE INFO

Article history:
Received 18 May 2012
Available online 27 September 2013

AMS subject classifications:
62H15
15A69

Keywords:
Cluster analysis
Covariance testing
Finite mixture model
Generalized skew-normal
Kurtosis
Tensor rank

ABSTRACT

The paper shows that the fourth cumulant of a finite mixture distribution might be decomposed into the mean of the components' fourth cumulants and the fourth cumulant of the components' means, when the mixture's components have the same second and third cumulants. Statistical applications include robustness properties of likelihood-based testing procedures and kurtosis-based projection methods. Practical relevance of theoretical results in the paper are illustrated with two well-known data sets.

1. Introduction

Let $x = (X_1, \ldots, X_d)^T$ be a d-dimensional random vector with mean $\mu = (\mu_1, \ldots, \mu_d)^T$, covariance matrix $\Sigma = \{\sigma_{ij}\}$ and finite fourth-order moments: $E(|X_i X_j X_h X_k|) < +\infty$, for $i, j, h, k = 1, \ldots, d$. The fourth cumulant $\mathcal{K}_4 = \{\kappa_{ijhk}\}$ of x is the d-dimensional, symmetric tensor of order 4 whose elements are the fourth-order derivatives of the cumulant generating function of x: $\kappa_{ijhk} = \log E[\exp(\iota t^T x)] / \partial t_i \partial t_j \partial t_h \partial t_k$, where $\iota = \sqrt{-1}$ and $t^T = (t_1, \ldots, t_d)$. An equivalent representation of κ_{ijhk} is

$$E[(X_i - \mu_i)(X_j - \mu_j)(X_h - \mu_h)(X_k - \mu_k)] - \sigma_{ij}\sigma_{hk} - \sigma_{ih}\sigma_{jk} - \sigma_{ik}\sigma_{jh}.$$

The elements κ_{ijhk} might be arranged into the $d^2 \times d^2$ block matrix $\kappa_4(x) = \{M_{pq}\}$, where $M_{pq} = \log E[\exp(\iota t^T x)] / \partial t_p \partial t_q \partial t \partial t^T$ for $p, q = 1, \ldots, d$. The matrix $\kappa_4(x)$ is the unfolded version of $\mathcal{K}_4$ (see, for example [41]) and can be represented as

$$E(y \otimes y^T \otimes y \otimes y^T) - (I_{d^2} + K_{d,d})(\Sigma \otimes \Sigma) - vec(\Sigma)\, vec^T(\Sigma),$$

where $y = x - \mu$, $\otimes$ denotes the Kronecker product, $vec(\Sigma)$ is the vectorization of Σ and $K_{d,d}$ is the $d^2 \times d^2$ commutation matrix [25]. With a slight abuse of notation, we shall refer to the matrix $\kappa_4(x)$ as to the fourth cumulant of x. Loperfido [23] examines some spectral properties of $\kappa_4(x)$.

In the general case, the number of distinct elements in $\kappa_4(x)$ increases very quickly with the dimension of x. If x is d- dimensional, $\kappa_4(x)$ might contain up to $d(d+1)(d+2)(d+3)/24$ distinct elements (see, for example, [16]). This suggests that statistical applications of the fourth order cumulant might greatly benefit from its parsimonious modelling, especially when they deal with multivariate kurtosis, as measured by functions of the fourth standardized cumulant. As a first example, Mardia [27], Malkovich and Afifi [26], Henze [13] investigate different measures of multivariate kurtosis for

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

0047-259X/$ – see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.jmva.2013.09.007

testing the hypothesis of multivariate normality. As a second example, kurtosis is used in Independent Component Analysis to recover the independent components themselves, when they are assumed to be leptokurtic (see, for example, [15]). As a third example, likelihood-based procedures for testing hypotheses on covariance matrices might be very sensitive to the kurtosis of the sampled distribution, when the latter is erroneously assumed to be multivariate normal [14,28,43,45].

Finite mixtures of multivariate distributions have often been used to achieve parsimonious modelling. Let $F_1, \ldots, F_g$ be d-dimensional cumulative distribution functions, and let $\pi_1, \ldots, \pi_g$ be nonnegative real numbers which add up to one. The weighted average $F = \pi_1 F_1 + \cdots \pi_g F_g$ is said to be a finite mixture distribution (or model), whose i-th component and i-th weight are F_i and π_i, respectively. Let μ_i and Ω_i be the mean and the variance of the i-th mixture's component F_i, respectively, for $i = 1, \ldots, g$. It is well-known that the mixture's mean μ, i.e. its first cumulant, is the mean of the components' means. It is also well-known that the mixture's covariance Σ, i.e. its second cumulant, is the mean of the components' covariances plus the covariance of the components' means. The above representations of the mean and the variance are appealing in that they are easily expressed both in words and in matrix notation. Higher-order cumulants of finite mixture distributions might be obtained via the law of total cumulance [5]. Unfortunately, it leads to results which are neither easily interpretable nor admit simple representations in the matrix form, thus limiting their use in statistical modelling.

In recent years, finite mixtures of elliptical distributions with proportional scatter matrices have been used to explore the statistical properties of kurtosis-based multivariate procedures. Projections which either maximize or minimize kurtosis have been used both in cluster analysis and in outlier detection [36–39]. In tensor terminology, they might be regarded as the best rank-one approximations to the fourth standardized cumulant. Tyler et al. [44] and Peña et al. [40] used a kurtosis matrix independently introduced by Cardoso [7] and Mori et al. [34] to uncover several features of multivariate data. The same matrix might be regarded as the sum of the d diagonal blocks of the fourth cumulant, which are $d \times d$ symmetric matrices [19]. Both approaches have good statistical properties when sampling from a finite mixture of elliptical distributions with proportional scatter matrices. At present time, however, no one investigated their robustness to violation of the underlying assumptions, not even in the special case of a mixture of two multivariate normal distributions with the same variance.

This paper addresses the above mentioned problems within the framework of finite mixtures whose components have identical second and third cumulants. They include several classes of well-known finite mixture models, most notably finite mixtures of normal distributions with equal covariance matrices. McLachlan and Peel [32] report many applications of such models and remark that often the component-covariances are restricted to being the same ([32], page 83). Their widespread use is partly due to the little inferential problems they pose, compared to other finite normal mixtures [33]. An additional advantage, from this paper's perspective, is their very parsimonious modelling of the fourth cumulant. However, the class of finite mixtures with equal second and third cumulants is much wider, since it also includes location mixtures. Skewness-based projection pursuit might be helpful in detecting clusters, when the sampled distribution is a location mixture of two multivariate, symmetric distributions [24]. Principal points of location mixtures of spherically symmetric distributions have nice theoretical properties [20,30,31]. Section 3 in this paper discusses location mixtures of multivariate skew-normal distributions.

We shall show that, when mixture's components have identical second and third cumulants, the fourth cumulant of the mixture equals the mean of the components' fourth cumulants, plus the fourth cumulant of the components' means. Statistical applications deal with robustness of multivariate statistical procedures. First we shall assess the robustness of MANOVA statistics when the data are drawn from a normal mixture with two homoscedastic components. Then we shall use mixtures with two skew-normal components to assess the robustness of the kurtosis-based procedures proposed by Peña and Prieto [36–39], Tyler et al. [44] and Peña et al. [40]. Other theorems in the paper, regarding fourth multivariate cumulants and moments, are instrumental in proving the above results as well as being interesting in their own right.

The rest of the paper is organized as follows. Section 2 contains the main results. Section 3 discusses some statistical applications. Section 4 illustrates their practical relevance with two well-known data sets. All proofs are deferred to the Appendix.

2. Main results

The following theorem represents the fourth moment $\mu_4(x-c)$ of the difference $x-c$, where x is a d-dimensional random vector and c is a real vector of the same dimension, as a function of the first four moments of x: $\mu = \mu_1 = E(x)$, $\mu_2 = E\left(xx^T\right)$, $\mu_3 = E\left(x \otimes x^T \otimes x\right)$ and $\mu_4 = E\left(x \otimes x^T \otimes x \otimes x^T\right)$.

Theorem 1. *Let μ_1, μ_2, μ_3, μ_4 be the first, second, third and fourth moment of the d-dimensional random vector x. Then the fourth moment of $x - c$, where c is a d-dimensional real vector, is*

$$\begin{aligned}
&\mu_4 - \mu_3^T \otimes c - \mu_3 \otimes c^T - c^T \otimes \mu_3 - c \otimes \mu_3^T + \mu_2 \otimes cc^T + vec(\mu_2) \otimes c^T \otimes c^T \\
&\quad + K_{d,d}\left(cc^T \otimes \mu_2\right) + K_{d,d}\left(\mu_2 \otimes cc^T\right) + c \otimes c \otimes vec^T(\mu_2) + cc^T \otimes \mu_2 \\
&\quad - \mu_1 c^T \otimes cc^T - c\mu_1^T \otimes cc^T - cc^T \otimes \mu_1 c^T - cc^T \otimes c\mu_1^T + cc^T \otimes cc^T.
\end{aligned}$$

As a direct consequence, the fourth central moment of a random vector might be represented via the first noncentral moments of the vector itself. More precisely, let μ_1, μ_2, μ_3, μ_4 be the first, second, third and fourth moment of the d-