ELSEVIER

Statistics & Probability Letters 51 (2001) 319–325

STATISTICS & PROBABILITY LETTERS

www.elsevier.nl/locate/stapro

Moments of skew-normal random vectors and their quadratic forms

Marc G. Genton *, Li He, Xiangwei Liu

Department of Mathematics, Massachusetts Institute of Technology, Cambridge, MA 02139-4307, USA

Received April 2000; received in revised form July 2000

Abstract

In this paper, we derive the moments of random vectors with multivariate skew-normal distribution and their quadratic forms. Applications to time series and spatial statistics are discussed. In particular, it is shown that the moments of the sample autocovariance function and of the sample variogram estimator do not depend on the skewness vector.

Keywords: Autocovariance function; Multivariate skew-normal distribution; Quadratic form; Variogram

1. Introduction

The skew-normal distribution is a family of distributions including the normal one, but with an extra parameter to regulate skewness. It allows for a continuous variation from normality to non-normality, which is useful in many situations. Azzalini (1985, 1986) introduced the univariate skew-normal distribution and studied the properties of this class of density functions. The class of the multivariate skew-normal distributions represents a mathematically tractable extension of the multivariate normal distribution with the addition of a vector parameter to regulate skewness. The probabilistic properties of the multivariate skew-normal distributions were discussed by Azzalini and Dalla Valle (1996), whereas Azzalini and Capitanio (1999) emphasized statistical applications.

An n-dimensional random vector $\boldsymbol{z}$ is said to have a multivariate skew-normal distribution, denoted by $\mathrm{SN}_n(\boldsymbol{\mu}, \Omega, \boldsymbol{\alpha})$, if it is continuous with density function

$$2\phi_n(\boldsymbol{z}; \boldsymbol{\mu}, \Omega)\Phi(\boldsymbol{\alpha}^{\mathrm{T}}(\boldsymbol{z} - \boldsymbol{\mu})), \quad \boldsymbol{z} \in \mathbb{R}^n, \tag{1}$$

* Correspondence address: Department of Statistics, 209-D, North Carolina State University, Box 8203, Raleigh, NC 27695-8203, USA. Tel.: +1 919 515 1936; fax: +1 919 515 1169.

E-mail addresses: genton@stat.ncsu.edu (M.G. Genton), lihe@math.mit.edu (L. He), xwliu@math.mit.edu (X. Liu).

0167-7152/01/$ - see front matter
PII: S0167-7152(00)00164-4

where $\phi_n(\boldsymbol{z};\boldsymbol{\mu},\Omega)$ is the n-dimensional normal density with mean $\boldsymbol{\mu}$ and correlation matrix Ω, $\Phi(\cdot)$ is the standard normal $\mathrm{N}(0,1)$ distribution function and $\boldsymbol{\alpha}$ is an n-dimensional vector. When $\boldsymbol{\alpha}=\mathbf{0}$, density (1) reduces to the one of the multivariate normal distribution $N_n(\boldsymbol{\mu},\Omega)$. Thus, the parameter $\boldsymbol{\alpha}$ is referred to as a "shape parameter". Note that if $\boldsymbol{\alpha}^{\mathrm{T}}$ is replaced by $\boldsymbol{\alpha}^{\mathrm{T}}\omega^{-1}$ in (1), where $\omega=\mathrm{diag}(\omega_1,\dots,\omega_n)^{\mathrm{T}}$, then a general covariance matrix Ω is allowed (see Azzalini and Capitanio, 1999, p. 584). Azzalini and Capitanio (1999) showed that if

$$\begin{pmatrix} X_0 \\ \boldsymbol{x} \end{pmatrix} \sim \mathrm{N}_{n+1}(\mathbf{0},\Omega^*), \quad \Omega^* = \begin{pmatrix} 1 & \boldsymbol{\delta}^{\mathrm{T}} \\ \boldsymbol{\delta} & \Omega \end{pmatrix},$$

where X_0 is a scalar component, Ω^* is a correlation matrix and

$$\boldsymbol{\delta} = \frac{\Omega\boldsymbol{\alpha}}{(1+\boldsymbol{\alpha}^{\mathrm{T}}\Omega\boldsymbol{\alpha})^{1/2}},$$

then

$$\boldsymbol{z} = \begin{cases} \boldsymbol{x} & \text{if } X_0>0, \\ -\boldsymbol{x} & \text{otherwise,} \end{cases}$$

has a skew-normal distribution $\mathrm{SN}_n(\mathbf{0},\Omega,\boldsymbol{\alpha})$ where

$$\boldsymbol{\alpha} = \frac{\Omega^{-1}\boldsymbol{\delta}}{(1-\boldsymbol{\delta}^{\mathrm{T}}\Omega^{-1}\boldsymbol{\delta})^{1/2}}.$$

This provides an easy way to generate skew-normal random samples in applications. It also shows that $\boldsymbol{\delta}$ (and thus also $\boldsymbol{\alpha}$) regulates the skewness of the distribution.

The paper is organized as follows. In Section 2, we compute the first four moments of a random vector $\boldsymbol{z}$ with multivariate skew-normal distribution $\mathrm{SN}_n(\boldsymbol{\mu},\Omega,\boldsymbol{\alpha})$ and the first two moments of its quadratic form

$$\boldsymbol{z}^{\mathrm{T}}A\boldsymbol{z}, \tag{2}$$

for a symmetric matrix A. In Section 3, we discuss applications of our results to the estimation of the autocovariance function in time series and the estimation of the variogram in spatial statistics. In particular, we show that the moments of these estimators do not depend on the skewness vector $\boldsymbol{\delta}$ (nor on $\boldsymbol{\alpha}$).

2. Moments

Azzalini and Dalla Valle (1996) studied the multivariate skew-normal distribution and gave the moment generating function $M(\boldsymbol{t})$ for the case when $\boldsymbol{\mu}=\mathbf{0}$:

$$\begin{aligned} M(\boldsymbol{t}) &= 2\int_{\mathbb{R}^n} \exp(\boldsymbol{t}^{\mathrm{T}}\boldsymbol{z})\phi_n(\boldsymbol{z};\mathbf{0},\Omega)\Phi(\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{z})\,\mathrm{d}\boldsymbol{z} \\ &= 2\exp\{\tfrac{1}{2}(\boldsymbol{t}^{\mathrm{T}}\Omega\boldsymbol{t})\}\,\Phi(\boldsymbol{\delta}^{\mathrm{T}}\boldsymbol{t}). \end{aligned} \tag{3}$$

From here, we can calculate the partial derivatives of $M(\boldsymbol{t})$ which are directly related to the moments of the skew-normal distribution with $\boldsymbol{\mu}=\mathbf{0}$:

$$\frac{\partial M(\boldsymbol{t})}{\partial \boldsymbol{t}} = 2\exp\{\tfrac{1}{2}\boldsymbol{t}^{\mathrm{T}}\Omega\boldsymbol{t}\}[\Omega\boldsymbol{t}\Phi(\boldsymbol{\delta}^{\mathrm{T}}\boldsymbol{t}) + \boldsymbol{\delta}\phi(\boldsymbol{\delta}^{\mathrm{T}}\boldsymbol{t})], \tag{4}$$

$$\begin{aligned} \frac{\partial^2 M(\boldsymbol{t})}{\partial \boldsymbol{t}\partial \boldsymbol{t}^{\mathrm{T}}} &= 2\exp\{\tfrac{1}{2}\boldsymbol{t}^{\mathrm{T}}\Omega\boldsymbol{t}\}\{\Phi(\boldsymbol{\delta}^{\mathrm{T}}\boldsymbol{t})[\Omega + (\Omega\boldsymbol{t})\otimes(\Omega\boldsymbol{t})^{\mathrm{T}}] \\ &\quad + \phi(\boldsymbol{\delta}^{\mathrm{T}}\boldsymbol{t})[(\Omega\boldsymbol{t})\otimes\boldsymbol{\delta}^{\mathrm{T}} + \boldsymbol{\delta}\otimes(\Omega\boldsymbol{t})^{\mathrm{T}} - \boldsymbol{\delta}\otimes\boldsymbol{\delta}^{\mathrm{T}}(\boldsymbol{\delta}^{\mathrm{T}}\boldsymbol{t})]\}, \end{aligned} \tag{5}$$

$$\frac{\partial^3 M(\boldsymbol{t})}{\partial \boldsymbol{t}\partial \boldsymbol{t}^{\mathrm{T}}\partial \boldsymbol{t}} = 2\exp\{\tfrac{1}{2}(\boldsymbol{t}^{\mathrm{T}}\Omega\boldsymbol{t})\}\Phi(\boldsymbol{\delta}^{\mathrm{T}}\boldsymbol{t})\{(\Omega\boldsymbol{t})\otimes\Omega + \mathrm{vec}(\Omega)(\Omega\boldsymbol{t})^{\mathrm{T}} + (I_n\otimes(\Omega\boldsymbol{t}))[\Omega + (\Omega\boldsymbol{t})(\Omega\boldsymbol{t})^{\mathrm{T}}]\}$$
$$+ 2\exp\{\tfrac{1}{2}(\boldsymbol{t}^{\mathrm{T}}\Omega\boldsymbol{t})\}\phi(\boldsymbol{\delta}^{\mathrm{T}}\boldsymbol{t})\{\boldsymbol{\delta}\otimes\Omega + \mathrm{vec}(\Omega)\boldsymbol{\delta}^{\mathrm{T}} + (I_n\otimes(\Omega\boldsymbol{t}))[\boldsymbol{\delta}\otimes(\Omega\boldsymbol{t})^{\mathrm{T}} + (\Omega\boldsymbol{t})\otimes\boldsymbol{\delta}^{\mathrm{T}}$$
$$-(\boldsymbol{\delta}\otimes\boldsymbol{\delta}^{\mathrm{T}})(\boldsymbol{\delta}^{\mathrm{T}}\boldsymbol{t})] + (I_n\otimes\boldsymbol{\delta})[\Omega + (\Omega\boldsymbol{t})\otimes(\Omega\boldsymbol{t})^{\mathrm{T}} - (\boldsymbol{\delta}\otimes(\Omega\boldsymbol{t})^{\mathrm{T}})(\boldsymbol{\delta}^{\mathrm{T}}\boldsymbol{t}) - \boldsymbol{\delta}\otimes\boldsymbol{\delta}^{\mathrm{T}}$$
$$-((\Omega\boldsymbol{t})\otimes\boldsymbol{\delta}^{\mathrm{T}})(\boldsymbol{\delta}^{\mathrm{T}}\boldsymbol{t}) + (\boldsymbol{\delta}\otimes\boldsymbol{\delta}^{\mathrm{T}})(\boldsymbol{\delta}^{\mathrm{T}}\boldsymbol{t})^2]\}, \qquad (6)$$

where $\phi(\cdot)$ is the standard normal density, I_n is the identity matrix of size $n\times n$, and $\otimes$ and vec are the Kronecker product and vectorizing operator, respectively (e.g. Fang and Zhang, 1990). See the appendix for details about calculations of the partial derivatives. Letting $\boldsymbol{t}=\boldsymbol{0}$ in (4)–(6), we obtain the first three moments of the skew-normal distribution:

$$M_1 = \sqrt{\frac{2}{\pi}}\boldsymbol{\delta}, \quad M_2 = \Omega,$$

$$M_3 = \sqrt{\frac{2}{\pi}}[\boldsymbol{\delta}\otimes\Omega + \mathrm{vec}(\Omega)\boldsymbol{\delta}^{\mathrm{T}} + (I_n\otimes\boldsymbol{\delta})\Omega - (I_n\otimes\boldsymbol{\delta})(\boldsymbol{\delta}\otimes\boldsymbol{\delta}^{\mathrm{T}})].$$

To find the fourth moment, since we only need the value of $M^{(4)}(\boldsymbol{t})$ at $\boldsymbol{t}=\boldsymbol{0}$, we do not need to compute the complete expression of

$$M^{(4)}(\boldsymbol{t}) = \frac{\partial M(\boldsymbol{t})}{\partial\boldsymbol{t}\partial\boldsymbol{t}^{\mathrm{T}}\partial\boldsymbol{t}\partial\boldsymbol{t}^{\mathrm{T}}} = \frac{\partial M^{(3)}(\boldsymbol{t})}{\partial\boldsymbol{t}^{\mathrm{T}}}. \qquad (7)$$

Instead, we can simply single out all the terms in (7) that do not contain the factor $\boldsymbol{t}$ or $\boldsymbol{t}^{\mathrm{T}}$. By doing that, we find

$$M^{(4)}(\boldsymbol{0}) = 2\exp\{\tfrac{1}{2}(\boldsymbol{t}^{\mathrm{T}}\Omega\boldsymbol{t})\}\Phi(\boldsymbol{\delta}^{\mathrm{T}}\boldsymbol{t})\{\Omega\otimes\Omega + \mathrm{vec}(\Omega)\mathrm{vec}(\Omega^{\mathrm{T}}) + U_{n,n}(\Omega\otimes\Omega)\}|_{\boldsymbol{t}=\boldsymbol{0}}$$
$$= (I_{n^2} + U_{n,n})(\Omega\otimes\Omega) + \mathrm{vec}(\Omega)\mathrm{vec}(\Omega^{\mathrm{T}}). \qquad (8)$$

Here, $U_{n,n}$ is the permutation matrix associated with an $n\times n$ matrix (its size is $n^2\times n^2$). See Graham (1981, p. 32–36) for details about permutation matrices. The above moments are for the case when $\boldsymbol{\mu}=\boldsymbol{0}$, and it is straightforward to derive the case when $\boldsymbol{\mu}\neq\boldsymbol{0}$.

Proposition 1. *Let $\boldsymbol{z}$ be a random vector with a skew-normal distribution* $\mathrm{SN}_n(\boldsymbol{\mu},\Omega,\boldsymbol{\alpha})$. *Then the four first moments of $\boldsymbol{z}$ are*

(a) $M_1 = \boldsymbol{\mu}_z = \boldsymbol{\mu} + \sqrt{\frac{2}{\pi}}\boldsymbol{\delta},$

(b) $M_2 = \Omega + \boldsymbol{\mu}\boldsymbol{\mu}^{\mathrm{T}} + \sqrt{\frac{2}{\pi}}(\boldsymbol{\mu}\boldsymbol{\delta}^{\mathrm{T}} + \boldsymbol{\delta}\boldsymbol{\mu}^{\mathrm{T}}),$

(c) $M_3 = \Omega\otimes\boldsymbol{\mu} + \boldsymbol{\mu}\otimes\Omega + \mathrm{vec}(\Omega)\otimes\boldsymbol{\mu}^{\mathrm{T}} + \boldsymbol{\mu}\otimes\boldsymbol{\mu}^{\mathrm{T}}\otimes\boldsymbol{\mu} + \sqrt{\frac{2}{\pi}}[\boldsymbol{\delta}\otimes\Omega + \mathrm{vec}(\Omega)\boldsymbol{\delta}^{\mathrm{T}}$
$+ (I_n\otimes\boldsymbol{\delta})\Omega - \boldsymbol{\delta}\otimes\boldsymbol{\delta}^{\mathrm{T}}\otimes\boldsymbol{\delta} + \boldsymbol{\delta}\otimes\boldsymbol{\mu}^{\mathrm{T}}\otimes\boldsymbol{\mu} + \boldsymbol{\mu}\otimes\boldsymbol{\delta}^{\mathrm{T}}\otimes\boldsymbol{\mu} + \boldsymbol{\mu}\otimes\boldsymbol{\mu}^{\mathrm{T}}\otimes\boldsymbol{\delta}],$

(d) $M_4 = \Omega\otimes\boldsymbol{\mu}\otimes\boldsymbol{\mu}^{\mathrm{T}} + \boldsymbol{\mu}\otimes\Omega\otimes\boldsymbol{\mu}^{\mathrm{T}} + \mathrm{vec}(\Omega)\otimes\boldsymbol{\mu}^{\mathrm{T}}\otimes\boldsymbol{\mu}^{\mathrm{T}} + \boldsymbol{\mu}\otimes\boldsymbol{\mu}^{\mathrm{T}}\otimes\boldsymbol{\mu}\otimes\boldsymbol{\mu}^{\mathrm{T}} + \Omega\otimes\Omega$
$+ \mathrm{vec}(\Omega)\mathrm{vec}(\Omega)^{\mathrm{T}} + U_{n,n}(\Omega\otimes\Omega) + \boldsymbol{\mu}^{\mathrm{T}}\otimes\Omega\otimes\boldsymbol{\mu} + \boldsymbol{\mu}\otimes\boldsymbol{\mu}\otimes\mathrm{vec}(\Omega)^{\mathrm{T}} + \boldsymbol{\mu}\otimes\boldsymbol{\mu}^{\mathrm{T}}\otimes\Omega$
$+ \sqrt{\frac{2}{\pi}}[\boldsymbol{\delta}\otimes\Omega\otimes\boldsymbol{\mu}^{\mathrm{T}} + \mathrm{vec}(\Omega)\otimes\boldsymbol{\delta}^{\mathrm{T}}\otimes\boldsymbol{\mu}^{\mathrm{T}} + ((I_n\otimes\boldsymbol{\delta})\Omega)\otimes\boldsymbol{\mu}^{\mathrm{T}} + \boldsymbol{\delta}\otimes\boldsymbol{\mu}^{\mathrm{T}}\otimes\boldsymbol{\mu}\otimes\boldsymbol{\mu}^{\mathrm{T}}$
$+ \boldsymbol{\mu}\otimes\boldsymbol{\delta}^{\mathrm{T}}\otimes\boldsymbol{\mu}\otimes\boldsymbol{\mu}^{\mathrm{T}} + \boldsymbol{\mu}\otimes\boldsymbol{\mu}^{\mathrm{T}}\otimes\boldsymbol{\delta}\otimes\boldsymbol{\mu}^{\mathrm{T}} + \boldsymbol{\delta}^{\mathrm{T}}\otimes\Omega\otimes\boldsymbol{\mu} + \boldsymbol{\delta}\otimes\mathrm{vec}(\Omega)^{\mathrm{T}}\otimes\boldsymbol{\mu}$

$$+(\Omega(I_n\otimes\boldsymbol{\delta}^{\mathrm T}))\otimes\boldsymbol{\mu}+\boldsymbol{\mu}^{\mathrm T}\otimes\boldsymbol{\delta}\otimes\Omega+\boldsymbol{\mu}^{\mathrm T}\otimes(\mathrm{vec}(\Omega)\boldsymbol{\delta}^{\mathrm T})+\boldsymbol{\mu}^{\mathrm T}\otimes((I_n\otimes\boldsymbol{\delta})\Omega)$$
$$+\boldsymbol{\mu}\otimes\boldsymbol{\delta}^{\mathrm T}\otimes\Omega+\boldsymbol{\mu}\otimes\boldsymbol{\delta}\otimes\mathrm{vec}(\Omega)^{\mathrm T}+\boldsymbol{\mu}\otimes(\Omega(I_n\otimes\boldsymbol{\delta}^{\mathrm T}))+\boldsymbol{\mu}\otimes\boldsymbol{\mu}^{\mathrm T}\otimes\boldsymbol{\delta}\otimes\boldsymbol{\delta}^{\mathrm T}$$
$$-\boldsymbol{\delta}\otimes\boldsymbol{\delta}^{\mathrm T}\otimes\boldsymbol{\delta}\otimes\boldsymbol{\mu}^{\mathrm T}-\boldsymbol{\delta}^{\mathrm T}\otimes\boldsymbol{\delta}\otimes\boldsymbol{\delta}^{\mathrm T}\otimes\boldsymbol{\mu}-\boldsymbol{\mu}^{\mathrm T}\otimes\boldsymbol{\delta}\otimes\boldsymbol{\delta}^{\mathrm T}\otimes\boldsymbol{\delta}-\boldsymbol{\mu}\otimes\boldsymbol{\delta}^{\mathrm T}\otimes\boldsymbol{\delta}\otimes\boldsymbol{\delta}^{\mathrm T}].$$

Proof. Let $E(g(\boldsymbol{z}))$ and $E_0(g(\boldsymbol{z}))$ denote the expectations of $g(\boldsymbol{z})$ when the distribution of $\boldsymbol{z}$ is $\mathrm{SN}_n(\boldsymbol{\mu},\Omega,\boldsymbol{\alpha})$ and $\mathrm{SN}_n(\mathbf{0},\Omega,\boldsymbol{\alpha})$ respectively. The two expectation functions are related by the relationship $E(g(\boldsymbol{z}))=E_0(g(\boldsymbol{z}+\boldsymbol{\mu}))$. The proposition then follows directly from the moments of the $\boldsymbol{\mu}=\mathbf{0}$ case. □

With the first four moments of the random vector $\boldsymbol{z}$ given in Proposition 1, we can compute the first two moments of its quadratic form.

Proposition 2. *Let $\boldsymbol{z}$ be a random vector with a multivariate skew-normal distribution* $\mathrm{SN}_n(\boldsymbol{\mu},\Omega,\boldsymbol{\alpha})$. *Let A,B be two symmetric $n\times n$ matrices. Then*

(a) $E(\boldsymbol{z}^{\mathrm T}A\boldsymbol{z})=\mathrm{tr}[A\Omega]+\boldsymbol{\mu}^{\mathrm T}A\boldsymbol{\mu}+2\sqrt{\frac{2}{\pi}}\boldsymbol{\mu}^{\mathrm T}A\boldsymbol{\delta}$,

(b) $\mathrm{Var}(\boldsymbol{z}^{\mathrm T}A\boldsymbol{z})=2\,\mathrm{tr}\,[(A\Omega)^2]+4\boldsymbol{\mu}^{\mathrm T}(A\Omega A)(\boldsymbol{\mu}+2\sqrt{\frac{2}{\pi}}\boldsymbol{\delta})-2\sqrt{\frac{2}{\pi}}\left[2(\boldsymbol{\delta}^{\mathrm T}A\boldsymbol{\delta})(\boldsymbol{\mu}^{\mathrm T}A\boldsymbol{\delta})+2\sqrt{\frac{2}{\pi}}(\boldsymbol{\mu}^{\mathrm T}A\boldsymbol{\delta})^2\right]$,

(c) $\mathrm{Cov}(\boldsymbol{z}^{\mathrm T}A\boldsymbol{z},\boldsymbol{z}^{\mathrm T}B\boldsymbol{z})=2\,\mathrm{tr}[A\Omega B\Omega]+2\boldsymbol{\mu}^{\mathrm T}(A\Omega B+B\Omega A)\left(\boldsymbol{\mu}+2\sqrt{\frac{2}{\pi}}\boldsymbol{\delta}\right)$
$-2\sqrt{\frac{2}{\pi}}\left[(\boldsymbol{\delta}^{\mathrm T}A\boldsymbol{\delta})(\boldsymbol{\mu}^{\mathrm T}B\boldsymbol{\delta})+(\boldsymbol{\mu}^{\mathrm T}A\boldsymbol{\delta})(\boldsymbol{\delta}^{\mathrm T}B\boldsymbol{\delta})+2\sqrt{\frac{2}{\pi}}(\boldsymbol{\mu}^{\mathrm T}A\boldsymbol{\delta})(\boldsymbol{\mu}^{\mathrm T}B\boldsymbol{\delta})\right]$.

where $\mathrm{tr}[\,\cdot\,]$ *denotes the trace of a matrix.*

Proof. These results are derived from Proposition 1 and the following relations (e.g. Li, 1987):

$$\mathrm{E}(\boldsymbol{z}^{\mathrm T}A\boldsymbol{z})=\mathrm{tr}(AM_2),$$
$$\mathrm{E}((\boldsymbol{z}^{\mathrm T}A\boldsymbol{z})(\boldsymbol{z}^{\mathrm T}B\boldsymbol{z}))=\mathrm{tr}((A\otimes B)M_4).\quad □$$

When $\boldsymbol{\delta}=\mathbf{0}$ (or equivalently $\boldsymbol{\alpha}=\mathbf{0}$), formulas of Proposition 2 reduce to those obtained in the multivariate normal case (e.g. Muirhead, 1982, p. 47).

3. Applications to time series and spatial statistics

Quadratic forms of random vectors appear in various applications, and a first example comes from the time-series context. Consider a second-order stationary time-series $\{Z_t\colon t\in\mathbb{Z}\}$ and let $v(h)=\mathrm{Cov}(X_{t+h},X_t)$, $\forall\, t,h\in\mathbb{Z}$, be the autocovariance function of Z_t at lag h (e.g. Brockwell and Davis, 1991). The classical estimator for the autocovariance function, based on the method-of-moments, is

$$\hat v(h)=\frac{1}{n}\sum_{i=1}^{n-h}(Z_{i+h}-\bar Z)(Z_i-\bar Z),\quad 0\leqslant h\leqslant n-1, \tag{9}$$

where $\bar{Z} = (1/n)\sum_{i=1}^{n} Z_i$. The simple form of this estimator allows us to write (9) as a quadratic form. In fact, if $\boldsymbol{z} = (Z_1, \ldots, Z_n)^{\mathrm{T}}$ is the data vector and $D(h)$ is the time design matrix of the data at lag h, then estimator (9) has a quadratic form as in (2) with

$$A = \frac{1}{n} M D(h) M, \tag{10}$$

where $M = I_n - (1/n)\mathbf{1}_n\mathbf{1}_n^{\mathrm{T}}$ is a symmetric matrix satisfying $M^2 = M$, and $\mathbf{1}_n = (1, \ldots, 1)^{\mathrm{T}} \in \mathbb{R}^n$. The time design matrix $D(h)$, of size $n \times n$ is symmetric and defined by $D(h) = \frac{1}{2}(P(h) + P(h)^{\mathrm{T}})$, $0 \leqslant h \leqslant n-1$, where $P(h)$ is an $n \times n$ matrix with ones on the hth upper diagonal and zero elsewhere, $1 \leqslant h \leqslant n-1$ and $P(0) = I_n$. The size of the upper or lower diagonal of ones is $n - h$. Note that the expectation vector of $\boldsymbol{z}$ is constant, $\boldsymbol{\mu}_z = \mu_z \mathbf{1}_n$, due to stationarity, and thus the matrix A defined by (10) satisfies

$$A\boldsymbol{\mu}_z = \mathbf{0}. \tag{11}$$

As a second example, consider the estimation of the variogram in spatial statistics (e.g. Cressie, 1993). Let $\{Z(\boldsymbol{x}):\ \boldsymbol{x} \in D \subset \mathbb{R}^d\}$, $d \geqslant 1$, be a spatial stochastic process, intrinsically stationary. Matheron's (1962) classical variogram estimator, based on the method-of-moments, is

$$2\hat{\gamma}(\boldsymbol{h}) = \frac{1}{N_{\boldsymbol{h}}} \sum_{N(\boldsymbol{h})} (Z(\boldsymbol{x}_i) - Z(\boldsymbol{x}_j))^2, \quad \boldsymbol{h} \in \mathbb{R}^d, \tag{12}$$

where $N(\boldsymbol{h}) = \{(\boldsymbol{x}_i, \boldsymbol{x}_j):\ \boldsymbol{x}_i - \boldsymbol{x}_j = \boldsymbol{h}\}$ and $N_{\boldsymbol{h}}$ is the cardinality of $N(\boldsymbol{h})$. Again, the simple formulation of this estimator allows (12) to be written as a quadratic form. In fact, if $\boldsymbol{z} = (Z(\boldsymbol{x}_1), \ldots, Z(\boldsymbol{x}_n))^{\mathrm{T}}$ is the data vector and $A(\boldsymbol{h})$ is the spatial design matrix of the data at lag $\boldsymbol{h}$, then estimator (12) has a quadratic form as in (2) with

$$A = \frac{1}{N_{\boldsymbol{h}}} A(\boldsymbol{h}). \tag{13}$$

The spatial design matrix $A(\boldsymbol{h})$ is a symmetric matrix of size $n \times n$, derived from (12). For regularly spaced data in $\mathbb{R}^1$, $A(h)$ has three possible forms depending on h ($h < n/2$, $h = n/2$, and $h > n/2$). For data on a regularly spaced multidimensional grid in $\mathbb{R}^d$, $d > 1$, the spatial design matrix $A(\boldsymbol{h})$ is based on the spatial design matrix for unidimensional data, and described by Kronecker products of matrices (Genton, 1998; Gorsich et al., 2001). Here again, the matrix A defined by (13) satisfies (11). We now have the following result.

Proposition 3. *Let $\boldsymbol{z}$ be a random vector with a multivariate skew-normal distribution* $\mathrm{SN}_n(\boldsymbol{\mu}, \Omega, \boldsymbol{\alpha})$, *with $\boldsymbol{\mu}_z = \mu_z \mathbf{1}_n$. Then, the sample autocovariance function* (9) *with $A = A_i = (1/n)MD(h_i)M$ and the sample variogram estimator* (12) *with $A = A_i = (1/N_{\boldsymbol{h}_i})A(\boldsymbol{h}_i)$, $i = 1, 2$, satisfy*:
(a) $E(\boldsymbol{z}^{\mathrm{T}} A \boldsymbol{z}) = \mathrm{tr}[A\Omega]$,
(b) $\mathrm{Var}(\boldsymbol{z}^{\mathrm{T}} A \boldsymbol{z}) = 2\,\mathrm{tr}[A\Omega A\Omega]$,
(c) $\mathrm{Cov}(\boldsymbol{z}^{\mathrm{T}} A_1 \boldsymbol{z}, \boldsymbol{z}^{\mathrm{T}} A_2 \boldsymbol{z}) = 2\,\mathrm{tr}[A_1\Omega A_2\Omega]$,
(d) $\mathrm{Corr}(\boldsymbol{z}^{\mathrm{T}} A_1 \boldsymbol{z}, \boldsymbol{z}^{\mathrm{T}} A_2 \boldsymbol{z}) = \dfrac{\mathrm{tr}[A_1\Omega A_2\Omega]}{\sqrt{\mathrm{tr}[A_1\Omega A_1\Omega]\mathrm{tr}[A_2\Omega A_2\Omega]}}$.

Proof. These results are automatic by-products from Proposition 2 and property (11) that $A\boldsymbol{\mu}_z = \mathbf{0}$. □

Note that although the mean $\boldsymbol{\mu}_z$ and the covariance matrix $\Sigma_z = \Omega - (2/\pi)\boldsymbol{\delta}\boldsymbol{\delta}^{\mathrm{T}}$ depend on the skewness vector $\boldsymbol{\delta}$, the moments described in Proposition 3 do not depend on $\boldsymbol{\delta}$ or $\boldsymbol{\alpha}$. As a consequence, the statistical properties of the autocovariance or variogram estimates do not depend on the skewness vector describing the multivariate distribution of the data. This is important, for example when fitting a valid parametric model to variogram estimates by generalized least squares (Genton, 1998).

Assume that the matrix Ω belongs to the particular family of covariance matrices $\mathscr{S} = \{\Omega \mid \Omega = \tau I_n + \mathbf{1}_n \boldsymbol{a}^{\mathrm{T}} + \boldsymbol{a}\mathbf{1}_n^{\mathrm{T}}\}$, where $\tau \in \mathbb{R}$ and $\boldsymbol{a} = (a_1, \ldots, a_n)^{\mathrm{T}} \in \mathbb{R}^n$ are defined in such a way that Ω is positive definite. Then, explicit formulas for the quantities (a)–(d) of Proposition 3 have been computed, see Genton (1998) for the variogram case and Genton (1999) for the autocovariance case. In particular, they do not depend on the vector $\boldsymbol{a}$.

Similar results as those in Propositions 1 and 2 can be found in Li (1987) for random vectors with elliptically contoured distributions, i.e. allowing for kurtosis, but not for skewness. Applications to variogram and autocovariance estimates have been discussed by Genton (1999, 2000), respectively. A further extension would be to consider moments of quadratic forms from multivariate skew-elliptical random vectors, thus allowing for both skewness and kurtosis. Research is currently conducted towards this direction.

Acknowledgements

The authors thank Adelchi Azzalini and an anonymous referee for very valuable comments.

Appendix

Calculation of the partial derivatives of $M(\boldsymbol{t})$

The procedure we used to obtain the partial derivatives of $M(\boldsymbol{t})$ is lengthy but not difficult. The following facts were used in the calculations:

(1) $$\frac{\partial \boldsymbol{t}^{\mathrm{T}} A \boldsymbol{t}}{\partial \boldsymbol{t}} = 2A\boldsymbol{t}, \quad \frac{\partial \boldsymbol{a}^{\mathrm{T}} \boldsymbol{t}}{\partial \boldsymbol{t}} = \boldsymbol{a}, \quad \frac{\partial A\boldsymbol{t}}{\partial \boldsymbol{t}^{\mathrm{T}}} = A \quad \text{and} \quad \frac{\partial A\boldsymbol{t}}{\partial \boldsymbol{t}} = \mathrm{vec}(A),$$

for constant $n \times n$ matrix A and n-vector $\boldsymbol{a}$

(2) $$\frac{\partial (XY)}{\partial Z} = \frac{\partial X}{\partial Z}(I_q \otimes Y) + (I_p \otimes X)\frac{\partial Y}{\partial Z}$$

X, Y and Z are matrices of size $m \times n$, $n \times v$ and $p \times q$, respectively.

(3) $$\frac{\partial (X \otimes Y)}{\partial Z} = \frac{\partial X}{\partial Z} \otimes Y + (I_p \otimes U_1)\left(\frac{\partial Y}{\partial Z} \otimes X\right)(I_q \otimes U_2)$$

X, Y and Z are matrices of size $m \times n$, $u \times v$ and $p \times q$, respectively. U_1 is the permutation matrix associated with a $m \times u$ matrix and U_2 is the permutation matrix associated with a $n \times v$ matrix.

For proof of those formulas, see Graham (1981, Chapter 6). The derivatives we computed follow directly from those facts. Note that we also double-checked the correctness of our results via simulations using the method mentioned in the introduction.

References

Azzalini, A., 1985. A class of distributions which includes the normal ones. Scand. J. Statist. 12, 171–178.
Azzalini, A., 1986. Further results on a class of distributions which includes the normal ones. Statistica 46, 199–208.
Azzalini, A., Capitanio, A., 1999. Statistical applications of the multivariate skew normal distribution. J. Roy. Statist. Soc. B 61, 579–602.
Azzalini, A., Dalla Valle, A., 1996. The multivariate skew-normal distribution. Biometrika 83, 715–726.
Brockwell, P.J., Davis, R.A., 1991. Time series: Theory and Methods. Springer, New York.
Cressie, N., 1993. Statistics for Spatial Data. Wiley, New York.
Fang, K.-T., Zhang, Y.-T., 1990. Generalized Multivariate Analysis. Springer, Berlin.

Genton, M.G., 1998. Variogram fitting by generalized least squares using an explicit formula for the covariance structure. Math. Geol. 30, 323–345.

Genton, M.G., 1999. The correlation structure of the sample autocovariance function for a particular class of time series with elliptically contoured distribution. Statist. Probab. Lett. 41, 131–137.

Genton, M.G., 2000. The correlation structure of Matheron's classical variogram estimator under elliptically contoured distributions. Math. Geol. 32, 127–137.

Gorsich, D.J., Genton, M.G., Strang, G., 2001. Eigenstructures of spatial design matrices. Journal of Multivariate Analysis.

Graham, A., 1981. Kronecker Products and Matrix Calculus With Applications. Ellis Horwood Limited, Chichester.

Li, G., 1987. Moments of a random vector and its quadratic forms. J. Statist. Appl. Probab. 2, 219–229.

Matheron, G., 1962. Traité de Géostatistique Appliquée, Tome I. Mémoires du Bureau de Recherches Géologiques et Minières, no. 14, Editions Technip, Paris.

Muirhead, R.J., 1982. Aspects of Multivariate Statistical Theory. Wiley, New York.