Quadratic forms of skew t distributions with an application to Spatial and Time series analysis

Hyoung-Moon Kim[1]

Abstract

Moments of skew t random vectors and their quadratic forms are derived. It is shown that the moments of the sample autocovariance function and of the sample variogram estimator do not depend on a skewness parameter.

Keywords: Multivariate skew t distribution; Quadratic form; skewness

1 Introduction

The skew t distribution was recently developed by Azzalini and Capitanio (2003), which includes the t distribution as a special case. It is related to the skew normal distribution by the following equation:

$$\mathbf{y} = \boldsymbol{\mu} + V^{-1/2}\mathbf{z}, \tag{1}$$

where $\mathbf{z}$ has the skew normal distribution, $SN_n(\mathbf{0}, \Omega, \boldsymbol{\alpha})$, and $V \sim \chi^2_\nu/\nu$, independent of $\mathbf{z}$. The skew normal distribution developed by Azzalini and Dalla Valle (1996), Azzalini and Capitanio (1999) is defined as follows:

$$f_{\mathbf{z}}(\mathbf{z}) = 2\phi_n(\mathbf{z}; \boldsymbol{\mu}, \Omega)\Phi(\boldsymbol{\alpha}'(\mathbf{z} - \boldsymbol{\mu})), \;\; \mathbf{z} \in R^n, \tag{2}$$

where $\phi_n(\mathbf{z}; \boldsymbol{\mu}, \Omega)$ is the n-dimensional normal pdf with mean $\boldsymbol{\mu}$ and correlation matrix Ω, $\Phi(\cdot)$ is the $N(0,1)$ cdf and $\boldsymbol{\alpha}$ is a n-dimensional vector. When $\mathbf{Z}$ has the pdf (2), we write $\mathbf{Z} \sim SN_n(\boldsymbol{\mu}, \Omega, \boldsymbol{\alpha})$. When $\boldsymbol{\alpha} = \mathbf{0}$, (2) reduces to $N_n(\boldsymbol{\mu}, \Omega)$ pdf; hence the parameter $\boldsymbol{\alpha}$ is referred to as a 'skewness (shape) parameter'.

To avoid the complexity of the characteristic function of the skew multivariate t distribution, we exploit the relationship in (1). In Section 2, the first four moments of a multivariate skew t distribution and the first two moments of its quadratic form, $\mathbf{z}'A\mathbf{z}$, are obtained for a symmetric matrix A. In Section 3, applications to the sample autocovariance function and the sample variogram estimator are discussed. In particular, it is shown that the moments of these estimators do not depend on the skewness parameter $\boldsymbol{\alpha}$.

[1]Hyoung-Moon Kim is Assistant professor, Department of Applied Statistics, Konkuk University, Seoul 143-701, Republic of Korea (E-mail: hmkim@konkuk.ac.kr).

2 Moments

Azzalini and Capitanio (2003) developed the multivariate skew t distribution and gave the mean vector and variance matrix for the case when $\boldsymbol{\mu} = \mathbf{0}$. Those moments are extended up to the fourth moment with $\boldsymbol{\mu} \neq \mathbf{0}$. To avoid the complexity of the characteristic function of the skew multivariate t distribution, we will use the simple identity, (1), related to the skew normal distribution. Genton et al. (2001) provided the moments of the skew normal distribution for the case when $\boldsymbol{\mu} = \mathbf{0}$.

Theorem 1 for the case when $\boldsymbol{\mu} \neq \mathbf{0}$ follows immediately using the identity (1), the results of Genton et al. (2001), and the fact that $(A_1A_2) \otimes (B_1B_2) = (A_1 \otimes B_1)(A_2 \otimes B_2)$.

Theorem 1. If $\mathbf{y} \sim St_n(\boldsymbol{\mu}, \Omega, \boldsymbol{\alpha}, \nu)$, then the first four moments of $\mathbf{y}$ are

$$
\begin{aligned}
\text{(a)}\quad M_1 &= \boldsymbol{\mu} + c\boldsymbol{\delta} \quad \text{, where} \quad c = \frac{(\nu/\pi)^{1/2}\Gamma((\nu-1)/2)}{\Gamma(\nu/2)}, \\
\text{(b)}\quad M_2 &= \frac{\nu}{\nu-2}\Omega + \boldsymbol{\mu}\boldsymbol{\mu}' + c(\boldsymbol{\mu}\boldsymbol{\delta}' + \boldsymbol{\delta}\boldsymbol{\mu}'), \\
\text{(c)}\quad M_3 &= \frac{\nu}{\nu-2}[\Omega \otimes \boldsymbol{\mu} + \boldsymbol{\mu} \otimes \Omega + vec(\Omega) \otimes \boldsymbol{\mu}'] + \boldsymbol{\mu} \otimes \boldsymbol{\mu}' \otimes \boldsymbol{\mu} + \frac{c\nu}{\nu-3}[\boldsymbol{\delta} \otimes \Omega + \\
&\quad vec(\Omega)\boldsymbol{\delta}' + (I_n \otimes \boldsymbol{\delta})\Omega - \boldsymbol{\delta} \otimes \boldsymbol{\delta}' \otimes \boldsymbol{\delta}] + c[\boldsymbol{\delta} \otimes \boldsymbol{\mu}' \otimes \boldsymbol{\mu} + \boldsymbol{\mu} \otimes \boldsymbol{\delta}' \otimes \boldsymbol{\mu} + \\
&\quad \boldsymbol{\mu} \otimes \boldsymbol{\mu}' \otimes \boldsymbol{\delta}], \\
\text{(d)}\quad M_4 &= \frac{\nu}{\nu-2}[\Omega \otimes \boldsymbol{\mu} \otimes \boldsymbol{\mu}' + \boldsymbol{\mu} \otimes \Omega \otimes \boldsymbol{\mu}' + vec(\Omega) \otimes \boldsymbol{\mu}' \otimes \boldsymbol{\mu}' + \boldsymbol{\mu}' \otimes \Omega \otimes \boldsymbol{\mu} + \\
&\quad \boldsymbol{\mu} \otimes \boldsymbol{\mu} \otimes vec(\Omega)' + \boldsymbol{\mu} \otimes \boldsymbol{\mu}' \otimes \Omega] + \boldsymbol{\mu} \otimes \boldsymbol{\mu}' \otimes \boldsymbol{\mu} \otimes \boldsymbol{\mu}' + \\
&\quad \frac{\nu^2}{(\nu-2)(\nu-4)}[(I_{n^2} + K_{nn})(\Omega \otimes \Omega) + vec(\Omega)vec(\Omega)'] + \\
&\quad c[\boldsymbol{\delta} \otimes \boldsymbol{\mu}' \otimes \boldsymbol{\mu} \otimes \boldsymbol{\mu}' + \boldsymbol{\mu} \otimes \boldsymbol{\delta}' \otimes \boldsymbol{\mu} \otimes \boldsymbol{\mu}' + \boldsymbol{\mu} \otimes \boldsymbol{\mu}' \otimes \boldsymbol{\delta} \otimes \boldsymbol{\mu}' \\
&\quad + \boldsymbol{\mu} \otimes \boldsymbol{\mu}' \otimes \boldsymbol{\mu} \otimes \boldsymbol{\delta}'] + \frac{c\nu}{\nu-3}[\boldsymbol{\delta} \otimes \Omega \otimes \boldsymbol{\mu}' + vec(\Omega) \otimes \boldsymbol{\delta}' \otimes \boldsymbol{\mu}' + \\
&\quad ((I_n \otimes \boldsymbol{\delta})\Omega) \otimes \boldsymbol{\mu}' + \boldsymbol{\delta}' \otimes \Omega \otimes \boldsymbol{\mu} + \boldsymbol{\delta} \otimes vec(\Omega)' \otimes \boldsymbol{\mu} + \\
&\quad (\Omega(I_n \otimes \boldsymbol{\delta}')) \otimes \boldsymbol{\mu} + \boldsymbol{\mu}' \otimes \boldsymbol{\delta} \otimes \Omega + \boldsymbol{\mu}' \otimes (vec(\Omega)\boldsymbol{\delta}') + \\
&\quad \boldsymbol{\mu}' \otimes ((I_n \otimes \boldsymbol{\delta})\Omega) + \boldsymbol{\mu} \otimes \boldsymbol{\delta}' \otimes \Omega + \boldsymbol{\mu} \otimes \boldsymbol{\delta} \otimes vec(\Omega)' + \\
&\quad \boldsymbol{\mu} \otimes (\Omega(I_n \otimes \boldsymbol{\delta}')) - \boldsymbol{\delta} \otimes \boldsymbol{\delta}' \otimes \boldsymbol{\delta} \otimes \boldsymbol{\mu}' - \boldsymbol{\delta}' \otimes \boldsymbol{\delta} \otimes \boldsymbol{\delta}' \otimes \boldsymbol{\mu} - \\
&\quad \boldsymbol{\mu}' \otimes \boldsymbol{\delta} \otimes \boldsymbol{\delta}' \otimes \boldsymbol{\delta} - \boldsymbol{\mu} \otimes \boldsymbol{\delta}' \otimes \boldsymbol{\delta} \otimes \boldsymbol{\delta}'].
\end{aligned}
$$

Here, K_{nn} is the commutation matrix associated with an $n \times n$ matrix and its size is actually $n^2 \times n^2$, and $\otimes$ and vec are the Kronecker operator and the vec operator, respectively (Schott, 1997). Using the first four moments of the random vector $\mathbf{y}$ given in Theorem 1, we can calculate the first two moments of its quadratic form.

Theorem 2. If $\mathbf{y} \sim St_n(\boldsymbol{\mu}, \Omega, \boldsymbol{\alpha}, \nu)$ and A, B are two symmetric $n \times n$ matrices, then

(a) $E(\mathbf{y}'A\mathbf{y}) = \frac{\nu}{\nu-2}tr(A\Omega) + \boldsymbol{\mu}'A\boldsymbol{\mu} + 2c\boldsymbol{\mu}'A\boldsymbol{\delta},$

(b) $Var(\mathbf{y}'A\mathbf{y}) = \frac{2\nu^2}{(\nu-2)(\nu-4)}tr((A\Omega)^2) + \frac{2\nu^2}{(\nu-2)^2(\nu-4)}(tr(A\Omega))^2$
$+\frac{4\nu}{\nu-2}\boldsymbol{\mu}'(A\Omega A)\left(\boldsymbol{\mu} + \frac{2c(\nu-2)}{\nu-3}\boldsymbol{\delta}\right) + \frac{4c\nu}{(\nu-2)(\nu-3)}\boldsymbol{\mu}'A\boldsymbol{\delta}tr(A\Omega)$
$-\frac{2c\nu}{\nu-3}[2\boldsymbol{\mu}'A\boldsymbol{\delta}\boldsymbol{\delta}'A\boldsymbol{\delta} + \frac{2c(\nu-3)}{\nu}(\boldsymbol{\mu}'A\boldsymbol{\delta})^2],$

(c) $Cov(\mathbf{y}'A\mathbf{y}, \mathbf{y}'B\mathbf{y}) = \frac{2\nu^2}{(\nu-2)(\nu-4)}tr(A\Omega B\Omega) +$
$\frac{2\nu^2}{(\nu-2)^2(\nu-4)}tr(A\Omega)tr(B\Omega) + \frac{2\nu}{\nu-2}\boldsymbol{\mu}'(A\Omega B + B\Omega A)\left(\boldsymbol{\mu} + \frac{2c(\nu-2)}{\nu-3}\boldsymbol{\delta}\right) +$
$\frac{2c\nu}{(\nu-2)(\nu-3)}[\boldsymbol{\mu}'B\boldsymbol{\delta}tr(A\Omega) + \boldsymbol{\mu}'A\boldsymbol{\delta}tr(B\Omega)] - \frac{2c\nu}{\nu-3}[\boldsymbol{\delta}'A\boldsymbol{\delta}\boldsymbol{\mu}'B\boldsymbol{\delta} +$
$\boldsymbol{\mu}'A\boldsymbol{\delta}\boldsymbol{\delta}'B\boldsymbol{\delta} + \frac{2c(\nu-3)}{\nu}\boldsymbol{\mu}'A\boldsymbol{\delta}\boldsymbol{\mu}'B\boldsymbol{\delta}],$

where $tr(\cdot)$ denotes the trace of a matrix. The proof rests on Theorem 1 and the following relations (Schott, 1997):

$$\begin{aligned} E(\mathbf{y}'A\mathbf{y}) &= tr(AM_2),\ Cov(\mathbf{y}'A\mathbf{y}, \mathbf{y}'B\mathbf{y}) = tr((A\otimes B)M_4) - tr(AM_2)tr(BM_2), \\ tr(AB) &= tr(BA),\ tr(A\otimes B) = tr(A)tr(B), \end{aligned}$$

and $tr(K_{nn}(A\otimes B)) = tr(AB) = (vec(A'))'vec(B)$.

3 Time series and spatial statistics

Genton et al. (2001) apply quadratic forms of random vectors with skew normal distribution to a second-order stationary time series and a spatial intrinsically stationary stochastic process. They show that the mean and the covariance matrix of the sample autocovariance function and the sample variogram estimator do not depend on the skewness parameter under the assumption that the location parameter, $\boldsymbol{\mu}$, is proportional to the unity vector, i.e. $\boldsymbol{\mu} = \mu\mathbf{1}$. For our case Theorem 3 reveals that these moments do not depend on a skewness parameter $\boldsymbol{\alpha}$ when the underlying distribution is changed to a multivariate skew t distribution.

We first note that the classical autocovariance function estimator (Wei, 1994),

$$\hat{\gamma}(h) = (1/n)\sum_{i=1}^{n-h}(Y_{i+h} - \bar{Y})(Y_i - \bar{Y}),\ 0 \le h \le n-1,\ \bar{Y} = (1/n)\Sigma_{i=1}^{n}Y_i, \quad (3)$$

of a second-order stationary time-series, $\{Y_t : t \in Y\}$, can be written as a quadratic form

$$\hat{\gamma}(h) = \mathbf{y}'\frac{1}{n}MD(h)M\mathbf{y},\ \mathbf{y} = (Y_1, \cdots, Y_n)', \quad (4)$$

where $M = I_n - (1/n)\mathbf{1}_n\mathbf{1}_n'$, $D(h) = (1/2)(P(h) + P(h)')$, $0 \le h \le n-1$, and $P(h)$ is an $n \times n$ matrix with ones on the hth upper diagonal and zero otherwise, $1 \le h \le n-1$ with $P(0) = I_n$. For the second example, Matheron's classical variogram estimator of an intrinsic stationary spatial process, $\{Y(\mathbf{x}) : \mathbf{x} \in D \subset R^d, d \ge 1\}$, is as follows (Cressie, 1993):

$$\begin{aligned} 2\hat{\gamma}(\mathbf{h}) &= (1/N_{\mathbf{h}}) \sum_{N(\mathbf{h})} (Y(\mathbf{x}_i) - Y(\mathbf{x}_j))^2, \quad \mathbf{h} \in R^d, \\ N(\mathbf{h}) &= \{(\mathbf{x}_i, \mathbf{x}_j) : \mathbf{x}_i - \mathbf{x}_j = \mathbf{h}\}, \end{aligned} \tag{5}$$

where $N_{\mathbf{h}}$ is the cardinality of $N(\mathbf{h})$. This estimator also can be expressible as a quadratic form

$$2\hat{\gamma}(\mathbf{h}) = \mathbf{y}'\frac{1}{N_{\mathbf{h}}}A(\mathbf{h})\mathbf{y}, \quad \mathbf{y} = (Y(\mathbf{x}_1), \cdots, Y(\mathbf{x}_n))', \tag{6}$$

where the spatial design matrix, $A(\mathbf{h})$, is given by Genton(1998).

Theorem 3. If $\mathbf{y} \sim St_n(\boldsymbol{\mu}, \Omega, \boldsymbol{\alpha}, \nu)$, where $\boldsymbol{\mu}_{\mathbf{y}} = \mu_y \mathbf{1}_n$, then the sample autocovariance function (4) with $A = A_i = (1/n)MD(h_i)M$ and the sample variogram estimator (6) with $A = A_i = (1/N_{\mathbf{h}_i})A(\mathbf{h}_i)$, $i = 1, 2$, satisfy:

(a) $E(\mathbf{y}'A\mathbf{y}) = \dfrac{\nu}{\nu-2}tr(A\Omega)$,

(b) $Var(\mathbf{y}'A\mathbf{y}) = \dfrac{2\nu^2}{(\nu-2)(\nu-4)}tr((A\Omega)^2) + \dfrac{2\nu^2}{(\nu-2)^2(\nu-4)}(tr(A\Omega))^2$,

(c) $Cov(\mathbf{y}'A_1\mathbf{y}, \mathbf{y}'A_2\mathbf{y}) =$
$\dfrac{2\nu^2}{(\nu-2)(\nu-4)}tr(A_1\Omega A_2\Omega) + \dfrac{2\nu^2}{(\nu-2)^2(\nu-4)}tr(A_1\Omega)tr(A_2\Omega)$,

(d) $Corr(\mathbf{y}'A_1\mathbf{y}, \mathbf{y}'A_2\mathbf{y}) =$
$\dfrac{tr(A_1\Omega A_2\Omega) + \frac{1}{\nu-2}tr(A_1\Omega)tr(A_2\Omega)}{\sqrt{tr((A_1\Omega)^2) + \frac{1}{\nu-2}(tr(A_1\Omega))^2}\sqrt{tr((A_2\Omega)^2) + \frac{1}{\nu-2}(tr(A_2\Omega))^2}}$.

The proof follows from Theorem 2 and the fact that $A\boldsymbol{\mu}_{\mathbf{y}} = \mathbf{0}$. The moments of Theorem 3 do not depend on $\boldsymbol{\delta}$ or $\boldsymbol{\alpha}$, even though the mean $\boldsymbol{\mu}_{\mathbf{y}}$ and the covariance matrix $Var(\mathbf{y}) = \Sigma_{\mathbf{y}} = (\nu/(\nu-2))\Omega - c^2\boldsymbol{\delta}\boldsymbol{\delta}'$ depend on $\boldsymbol{\delta}$ as well as a skewness parameter $\boldsymbol{\alpha}$. So the statistical properties of the autocovariance or the variogram estimates do not depend on the skewness parameter of the skew t distribution. An important issue arises to fit a valid parametric model to variogram estimates using generalized least squares (Genton, 1998). Because the same observation is used for different lags, variogram estimates of different spatial lags are correlated. As a consequence, variogram fitting by ordinary least squares is not satisfactory.

References

Azzalini, A. and Capitanio, A., 1999. Statistical applications of the multivariate skew normal distribution. J. Roy. Statist. Soc. Ser. B 61, 579-602.

Azzalini, A. and Capitanio, A., 2002. Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t distribution. to appear in J. Roy. Statist. Soc. Ser. B.

Azzalini, A. and Dalla Valle, A., 1996. The multivariate skew-normal distribution. Biometrika 83, 715-726.

Cressie, N., 1993. Statistics for Spatial Data(rev. ed.). New York:Wiley.

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., He, L. and Liu, X., 2001. Moments of skew-normal random vectors and their quadratic forms. Statist. Probab. Lett. 51, 319-325.

Schott, James R., 1997. Matrix Analysis for Statistics. New York:Wiley.

Wei, W.W., 1994. Time Series Analysis. New York:Addison-Wesley.