*Journal of the Korean Statistical Society (2005), **34**: 2, pp 109–123*

MOMENTS OF VARIOGRAM ESTIMATOR FOR A GENERALIZED SKEW t DISTRIBUTION

HYOUNG-MOON KIM[1]

ABSTRACT

Variogram estimation is an important step of spatial statistics since it determines the kriging weights. Matheron's variogram estimator can be written as a quadratic form of the observed data. In this paper, we extend a skew t distribution to a generalized skew t distribution and moments of the variogram estimator for a generalized skew t distribution are derived in closed forms. After calculating the correlation structure of the variogram estimator, variogram fitting by generalized least squares is discussed.

AMS 2000 subject classifications. Primary 11E25.
Keywords. Multivariate generalized skew t distribution, quadratic form, skewness, kurtosis, generalized least squares.

1. INTRODUCTION

Variogram estimation is an important step of spatial statistics since it determines the kriging weights. Matheron's classical variogram estimator of an intrinsic stationary spatial process, $\{Y(\mathbf{x}) : \mathbf{x} \in D \subset R^d, d \geq 1\}$, is as follows (Cressie, 1993):

$$2\hat{\gamma}(\mathbf{h}) = (1/N_{\mathbf{h}}) \sum_{N(\mathbf{h})} (Y(\mathbf{x}_i) - Y(\mathbf{x}_j))^2, \tag{1.1}$$
$$\mathbf{h} \in R^d, \;\; N(\mathbf{h}) = \{(\mathbf{x}_i, \mathbf{x}_j) : \mathbf{x}_i - \mathbf{x}_j = \mathbf{h}\},$$

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

$$2\hat{\gamma}(\mathbf{h}) = \mathbf{y}'\frac{1}{N_{\mathbf{h}}}A(\mathbf{h})\mathbf{y}, \;\; \mathbf{y} = (Y(\mathbf{x}_1), \cdots, Y(\mathbf{x}_n))' \tag{1.2}$$

Received October 2004; accepted April 2005.
[1]Department of Applied Statistics, Konkuk University, Seoul 143–701, Korea (e–mail : hmkim@konkuk.ac.kr)

where the spatial design matrix, $A(\mathbf{h})/N_{\mathbf{h}}$, is given by Genton(1998a) and Gorsich *et al.* (2002).

Understanding the statistical properties of the variogram estimator (1.2) is important. Because the same observation is used for different lags, variogram estimates at different spatial lags are correlated. As a consequence, variogram fitting by ordinary least squares is not satisfactory. Genton (1998a, 2000) uses generalized least squares for fitting a valid parametric model to variogram estimates. For data with the Gaussian distribution, the mean and variance (Cressie, 1993; Schott, 1997) of $2\hat{\gamma}(\mathbf{h})$, as well as its correlation structure (Genton, 1998a; Schott, 1997), are easily computed. These results have been extended to the elliptical distribution (Li, 1987) using characteristic function approach. Genton(2000) derived the correlation structure of Matheron's classical variogram estimator under the elliptical distribution with Muirhead's kurtosis, $\kappa = 0$. So he obviously excludes the case of a generalized t distribution where $\kappa = 2/(\nu - 4)$ (Muirhead, 1982). For a skew normal distribution, Genton et al.(2001) evaluated moments and correlation structure of Matheron's classical variogram estimator. We will extend it to a generalized skew t distribution.

To motivate our approach, more explanation will be given to the generalized least squares method. Genton (1998a, 2000) proposes to use the method of generalized least squares with an explicit formula for the covariance structure. It finds the estimator $\hat{\boldsymbol{\theta}}$ minimizing following equation:

$$F(\boldsymbol{\theta}) = (2\hat{\gamma} - 2\gamma(\boldsymbol{\theta}))'\Sigma^{-1}(2\hat{\gamma} - 2\gamma(\boldsymbol{\theta})), \tag{1.3}$$

where $2\hat{\gamma} = (2\hat{\gamma}(\mathbf{h}_1), \ldots, 2\hat{\gamma}(\mathbf{h}_d))' \in R^d$ is the random vector with covariance matrix $Var(2\hat{\gamma}) = \Sigma$, with $\mathbf{h}_i = i\mathbf{h}/||\mathbf{h}||$, $i = 1, \ldots, d$, and $2\gamma(\theta) = (2\gamma(\mathbf{h}_1; \theta), \ldots, 2\gamma(\mathbf{h}_d; \theta))' \in R^d$ is the vector of a valid parametric variogram. He suggests to use the matrix Σ defined by

$$\Sigma_{ij} = Corr(2\hat{\gamma}(\mathbf{h}_i), 2\hat{\gamma}(\mathbf{h}_j))\gamma(\mathbf{h}_i; \boldsymbol{\theta})\gamma(\mathbf{h}_j; \boldsymbol{\theta}))/\sqrt{N_i N_j}, \tag{1.4}$$

where N_i is the number of differences at lag $\mathbf{h}_i$, and the correlation $Corr(2\hat{\gamma}(\mathbf{h}_i), 2\hat{\gamma}(\mathbf{h}_j))$ is computed with an explicit formula in the multivariate independent Gaussian case. Therefore, the correlation $Corr(2\hat{\gamma}(\mathbf{h}_i), 2\hat{\gamma}(\mathbf{h}_j))$ in equation (1.4) can be improved by using Theorem 2 (will appear) for one dimension and Theorem 3 (will appear) for d dimension and estimating ν and λ, which will be discussed at next section, from the data using maximum likelihood estimation (Azzalini and Capitanio, 2003).

We first develop a generalized skew t distribution based on a skew t distribution (Azzalini and Capitanio, 2003). The moments of Matheron's variogram estimators for a generalized skew t distribution do not depend on a skewness parameter $\boldsymbol{\alpha}$ introduced in the next section. The second section present the closed forms of these moments of the observations in R^1. We also extend those to data in R^d.

2. GENERALIZED SKEW t DISTRIBUTION

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

$$\mathbf{y} = \boldsymbol{\mu} + V^{-1/2}\mathbf{z}, \tag{2.1}$$

where $\mathbf{z}$ has a skew normal distribution, $SN_n(\mathbf{0}, \Omega, \boldsymbol{\alpha})$, and $V \sim \chi^2_\nu/\nu$, independent of $\mathbf{z}$. An equivalent interpretation of $\mathbf{y}$ is to regard it as a scale mixture of skew normal variates, with mixing factor $V^{-1/2}$. Here $\Omega = w\bar{\Omega}w = (w_{rs})$ is a full-rank $n \times n$ covariance matrix, where

$$w = diag(w_1, \cdots, w_n) = diag(w_{11}, \cdots, w_{nn})^{1/2}, \tag{2.2}$$

$\bar{\Omega} = w^{-1}\Omega w^{-1}$ is the associated correlation matrix, and $\boldsymbol{\mu}, \boldsymbol{\alpha} \in R^n$. A skew normal distribution developed by Azzalini and Dalla Valle (1996), Azzalini and Capitanio (1999) is defined as follows:

$$f_{\mathbf{z}}(\mathbf{z}) = 2\phi_n(\mathbf{z}; \boldsymbol{\mu}, \Omega)\Phi(\boldsymbol{\alpha}' w^{-1}(\mathbf{z} - \boldsymbol{\mu})), \ \ \mathbf{z} \in R^n, \tag{2.3}$$

where $\phi_n(\mathbf{z}; \boldsymbol{\mu}, \Omega)$ is the n-dimensional normal pdf with mean $\boldsymbol{\mu}$ and covariance matrix Ω, $\Phi(\cdot)$ is the $N(0,1)$ cdf and $\boldsymbol{\alpha}$ is a n- dimensional vector. When $\mathbf{Z}$ has the pdf (2.3), we write $\mathbf{Z} \sim SN_n(\boldsymbol{\mu}, \Omega, \boldsymbol{\alpha})$. When $\boldsymbol{\alpha} = \mathbf{0}$, (2.3) reduces to $N_n(\boldsymbol{\mu}, \Omega)$ pdf; hence the parameter $\boldsymbol{\alpha}$ is referred to as a skewness or a shape parameter.

We extend a skew t distribution to a generalized skew t distribution similar to Arellano-Valle and Bolfarine (1995, an extension of a t distribution to a generalized t distribution). It can be simply done changing a mixing distribution of V as follows:

$$\mathbf{y} = \boldsymbol{\mu} + V^{1/2}\mathbf{z}, \tag{2.4}$$

where $\mathbf{z}$ has the same skew normal distribution in (2.1), and $V \sim IG(\nu/2, \lambda/2)$, independent of $\mathbf{z}$. Here $IG(\alpha, \beta)$ denotes an inverse gamma distribution with the density function

$$f_V(v) = \frac{\beta^\alpha}{\Gamma(\alpha)} \frac{e^{-\beta/v}}{v^{\alpha+1}}, \text{where} \quad v > 0, \alpha > 0, \text{and} \quad \beta > 0. \tag{2.5}$$

Therefore the pdf of $\mathbf{y}$ can be obtained by Appendix Lemma 1 (Azzalini and Capitanio, 2003) and some simple algebra as follows:

$$f_{\mathbf{y}}(\mathbf{y}) = 2\,Gt_n(\mathbf{y}; \lambda, \nu)\, T_1 \left(\boldsymbol{\alpha}' w^{-1}(\mathbf{y} - \boldsymbol{\mu}) \left(\frac{\nu + n}{Q_{\mathbf{y}} + \lambda} \right)^{1/2} ; \nu + n \right), \tag{2.6}$$

where

$$Q_{\mathbf{y}} = (\mathbf{y} - \boldsymbol{\mu})' \Omega^{-1} (\mathbf{y} - \boldsymbol{\mu}). \text{ Here} \tag{2.7}$$

$$\begin{aligned} Gt_n(\mathbf{y}; \lambda, \nu) &= |\Omega|^{-1/2} g_n(Q_{\mathbf{y}}; \lambda, \nu) \\ &= \frac{\Gamma((\nu+n)/2)}{|\Omega|^{1/2} (\pi\lambda)^{n/2} \Gamma(\nu/2)} (1 + Q_{\mathbf{y}}/\lambda)^{-(\nu+n)/2} \end{aligned} \tag{2.8}$$

is the pdf of a $n-$dimensional generalized t variate with two parameters ν and λ. The notation $T_1(x; \nu + n)$ denotes the scalar t cdf with $\nu + n$ degrees of freedom. The notation $\mathbf{y} \sim GSt_n(\boldsymbol{\mu}, \Omega, \boldsymbol{\alpha}, \lambda, \nu)$ is used to indicate that $\mathbf{y}$ has the pdf (2.6). Similar to the skew normal distribution, the pdf (2.6) reduces to the one of the multivariate generalized t distribution $t_n(\boldsymbol{\mu}, \Omega, \lambda, \nu)$ when $\boldsymbol{\alpha} = \mathbf{0}$. Furthermore it reduces to the one of the multivariate t distribution $t_n(\boldsymbol{\mu}, \Omega, \nu)$ when $\boldsymbol{\alpha} = \mathbf{0}$ and $\nu = \lambda$.

We might use characteristic function of a generalized skew t distribution after developing it. However the characteristic function of the generalized t distribution (see Fang et al. for the characteristic function of the multivariate t distribution, 1989) is still complicated, we use the simple stochastic relationship between a skew normal distribution and a generalized skew t distribution instead of using the characteristic function approach.

Moments of Matheron's variogram estimator under a skew t distribution up to matrix *trace* notation are derived by Kim and Mallick (2003) where Ω was a correlation matrix. However it is straightforward to prove that all moment formulas are exactly the same when Ω is changed from a correlation matrix to a covariance matrix (personal communication with Adelchi Azzalini). Furthermore an extension to a generalized t distribution results are as follows:

THEOREM 1. *If* $\mathbf{y} \sim GSt_n(\boldsymbol{\mu}, \Omega, \boldsymbol{\alpha}, \lambda, \nu)$*, where* $\boldsymbol{\mu}_{\mathbf{y}} = \mu_y \mathbf{1}_n$*, then the sample variogram estimator (1.2) with* $A = A_i = (1/N_{\mathbf{h}_i})A(\mathbf{h}_i)$*,* $i = 1, 2$*, satisfies:*

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

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

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

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

where $tr(\cdot)$ denotes the trace of a matrix. The proof follows from Theorem 4, 5, and 6 of Appendix and the fact that $A\boldsymbol{\mu}_{\mathbf{y}} = \mathbf{0}$. Intermediate calculation results for obtaining Theorem 1 are obtained and explained in Appendix. When $\boldsymbol{\alpha} = \mathbf{0}$, $\nu \to \infty$, and $\lambda \to \infty$, the formulas of Theorem 1 reduce to those obtained in the multivariate normal case (Cressie, 1993; Schott, 1997; Genton, 1998a). So $\lambda/(\nu-2)$ in (a) will be 1. Similarly $2\lambda^2/((\nu-2)(\nu-4))$ in (b) and (c) will be 2 and $2\lambda^2/((\nu-2)^2(\nu-4))$ in (b) and (c) will be 0. Furthermore $1/(\nu-2)$ in (d) will be 0. These results for the multivariate normal case can also be appplied to Theorems 2 and 3. When $\lambda = \nu$, the formulas of Theorem 1 reduce to those obtained in a skew t case (Kim and Mallick, 2003). The moments of Theorem 1 do not depend on $\boldsymbol{\alpha}$, so the statistical properties of variogram estimates do not depend on a skewness parameter in a generalized skew t distribution. Thus, they are exactly same as those of a generalized t distribution. Furthermore the correlation structure, (d), does not depend on a parameter λ. However these moments depend on a measure of multivariate kurtosis, $\beta_{2,n}$, introduced by Mardia (1970, 1974):

$$\begin{aligned}\beta_{2,n} &= E\{(\mathbf{y} - E(\mathbf{y}))'Var(\mathbf{y})^{-1}(\mathbf{y} - E(\mathbf{y}))\}^2 \\ &= \frac{\nu-2}{\nu-4}\left[\frac{3(\boldsymbol{\delta}'\Omega^{-1}\boldsymbol{\delta})^2}{\left(\frac{\lambda}{c^{*2}(\nu-2)} - \boldsymbol{\delta}'\Omega^{-1}\boldsymbol{\delta}\right)^2} + \frac{(2n+4)\boldsymbol{\delta}'\Omega^{-1}\boldsymbol{\delta}}{\frac{\lambda}{c^{*2}(\nu-2)} - \boldsymbol{\delta}'\Omega^{-1}\boldsymbol{\delta}} + n^2 + 2n\right], \quad (2.9)\end{aligned}$$

where $c^* = [(\lambda/\pi)^{1/2}\Gamma((\nu-1)/2)]/\Gamma(\nu/2)$ and $\boldsymbol{\delta} = \Omega\boldsymbol{\alpha}/((1+\boldsymbol{\alpha}'\Omega\boldsymbol{\alpha})^{1/2})$. This measure is calculated using the results of Theorem 5, and 6 (Appendix), and $(A + b\mathbf{u}\mathbf{v}')^{-1} = A^{-1} - (b/(1 + b\mathbf{v}'A^{-1}\mathbf{u}))A^{-1}\mathbf{u}\mathbf{v}'A^{-1}$, where $\mathbf{u}$ and $\mathbf{v}'$ are column and row vectors, respectively, b is a constant (Henderson and Searle, 1981).

3. Moments of Variogram Estimation in R^1

The spatial design matrix of Matheron's classical variogram estimator, $(1/(n-h))A(h)$ of size $n \times n$, is given as follows in one dimensional case (Genton, 1998a) when data are regularly spaced:

$$\frac{1}{n-h}A(h) = \frac{1}{n-h}\begin{pmatrix} I_{n-h} & -I_{n-h} \\ -I_{n-h} & I_{n-h} \end{pmatrix}. \tag{3.1}$$

It is built by superposing identity matrices I_{n-h}, of size $(n-h)\times(n-h)$. If $h < n/2$, the word, superposition, means that two elements located at the same place are added. Note that if data are irregularly spaced, "tolerance" regions around h are often used (Cressie, 1993). There are three possible matrices depending on h. For example, the spatial design matrix of $n = 4$ after removing $1/(n-h)$ are:

$$A(1) = \begin{pmatrix} 1 & -1 & 0 & 0 \\ -1 & 2 & -1 & 0 \\ 0 & -1 & 2 & -1 \\ 0 & 0 & -1 & 1 \end{pmatrix}, \quad A(2) = \begin{pmatrix} 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 \\ -1 & 0 & 1 & 0 \\ 0 & -1 & 0 & 1 \end{pmatrix},$$

$$A(3) = \begin{pmatrix} 1 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 1 \end{pmatrix}. \tag{3.2}$$

We will first derive the explicit formulas for Theorem 1 with an additional assumption $\Omega = \sigma^2 I_n$. Since the spatial design matrix is symmetric and tri-ridged, we can apply Appendix Lemma 2 (Genton, 1998a) to prove following theorem.

THEOREM 2. *If $\mathbf{y} \sim GSt_n(\boldsymbol{\mu}, \Omega, \boldsymbol{\alpha}, \lambda, \nu)$, where $\boldsymbol{\mu}_{\mathbf{y}} = \mu_y \mathbf{1}_n$ and $\Omega = \sigma^2 I_n$, then the sample variogram estimator (1.2) with $A = A_i = (1/N_{h_i})A(h_i) = (1/(n-h))A(h_i)$, $i = 1, 2$, satisfy:*

(a) $E(\mathbf{y}'A\mathbf{y}) = \frac{2\lambda\sigma^2}{\nu-2}$,

(b) $$Var(\mathbf{y}'A\mathbf{y}) = \begin{cases} \dfrac{4\lambda^2(3n-4h)\sigma^4}{(\nu-2)(\nu-4)(n-h)^2} + \dfrac{8\lambda^2\sigma^4}{(\nu-2)^2(\nu-4)} & \text{if } h < \dfrac{n}{2} \\ \dfrac{8\lambda^2\sigma^4}{(\nu-2)(\nu-4)(n-h)} + \dfrac{8\lambda^2\sigma^4}{(\nu-2)^2(\nu-4)} & \text{otherwise,} \end{cases}$$

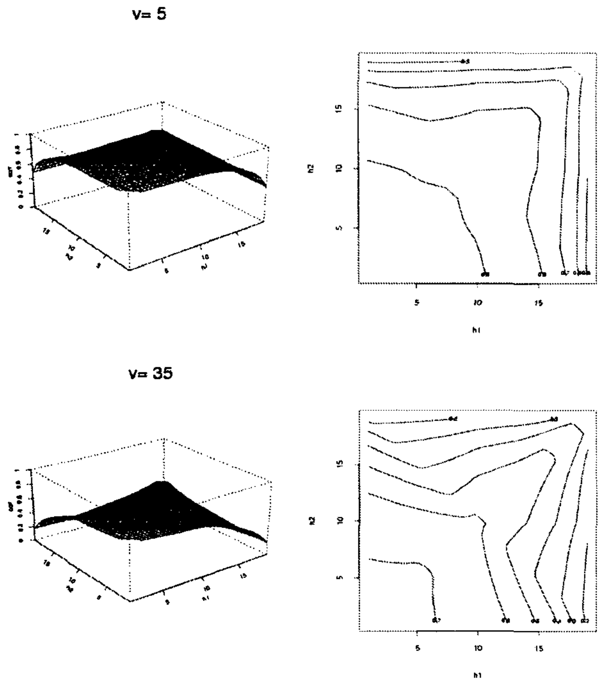

FIGURE 3.1 *These plots show the dependence of correlation on the lags h_1 and h_2 with $n = 20$ and different values of ν*

and for $h_1 < h_2$

(c) $Cov(\mathbf{y}'A_1\mathbf{y}, \mathbf{y}'A_2\mathbf{y})$

$$= \begin{cases} \dfrac{4\lambda^2(2n - h_1 - 2h_2)\sigma^4}{(\nu-2)(\nu-4)(n-h_1)(n-h_2)} + \dfrac{8\lambda^2\sigma^4}{(\nu-2)^2(\nu-4)} & \text{if } h_1 + h_2 < n \\ \dfrac{4\lambda^2\sigma^4}{(\nu-2)(\nu-4)(n-h_1)} + \dfrac{8\lambda^2\sigma^4}{(\nu-2)^2(\nu-4)} & \text{otherwise} \end{cases},$$

(d) $Corr(\mathbf{y}'A_1\mathbf{y}, \mathbf{y}'A_2\mathbf{y})$

$$= \begin{cases} \dfrac{(\nu-2)(2n-h_1-2h_2)+2(n-h_1)(n-h_2)}{\sqrt{(\nu-2)(3n-4h_1)+2(n-h_1)^2}\sqrt{(\nu-2)(3n-4h_2)+2(n-h_2)^2}} & \text{if } h_2 < \dfrac{n}{2} \\ \dfrac{(\nu-2)(2n-h_1-2h_2)+2(n-h_1)(n-h_2)}{\sqrt{(\nu-2)(3n-4h_1)+2(n-h_1)^2}\sqrt{2(\nu-2)(n-h_2)+2(n-h_2)^2}} & \text{if } h_2 \geq \dfrac{n}{2} \text{ and } h_1+h_2 < n \\ \dfrac{\sqrt{n-h_2}\{(\nu-2)+2(n-h_1)\}}{\sqrt{(\nu-2)(3n-4h_1)+2(n-h_1)^2}\sqrt{2(\nu-2)+2(n-h_2)}} & \text{if } h_1 < \frac{n}{2} \text{ and } h_1+h_2 \geq n \\ \dfrac{1}{2}\dfrac{\nu-2+2(n-h_1)}{\sqrt{\nu-2+n-h_1}\sqrt{\nu-2+n-h_2}}\sqrt{\dfrac{n-h_2}{n-h_1}} & \text{if } h_1 \geq \dfrac{n}{2}. \end{cases}$$

The perspective plots and corresponding contour plots of the correlation structure of the sample variogram estimator are plotted in Figure 3.1, where $n = 20$, different values of ν, and the formula (d) of Theorem 2 is used.

These results can be extended to the case where the covariance matrix Ω belongs to the particular family S of matrices:

$$S = \{\Omega | \Omega = \varphi I_n + \mathbf{1}_n \mathbf{a}' + \mathbf{a} \mathbf{1}_n'\}, \tag{3.3}$$

where $\varphi \in R$ and $\mathbf{a} = (a_1, \cdots, a_n)' \in R^n$ are defined in such a way that Ω is positive definite. We can choose $\varphi > 0$ such that

$$\varphi > \left(n\Sigma_{i=1}^n a_i^2\right)^{1/2} - \Sigma_{i=1}^n a_i, \tag{3.4}$$

to guarantee positive definiteness of Ω. The family S contains the uncorrelated case ($\varphi = \sigma^2, \mathbf{a} = \mathbf{0}$) and the equicorrelation case ($\varphi = 1 - \rho$, $\mathbf{a} = (\rho/2)\mathbf{1}_n$) as special cases. Because $A(h)\mathbf{1}_n = \mathbf{0}\ \forall h$ and $\mathbf{1}_n' A(h) = \mathbf{0}\ \forall h$, the results are exactly same as those of Theorem 2 except σ^2 is changed to φ.

4. SIMULATION

By the recommendation of one of the referees, we add a simulation part to improve reader's understanding. First, for simplicity, we generate 100 samples of $n = 100$ from a generalized skew t distribution with $\boldsymbol{\mu} = \mathbf{0}$, $\boldsymbol{\alpha} = \mathbf{0}$, $\lambda = \nu = 10$ and an exponential variogram

$$\gamma(h, c) = 2 - exp(-h/c) \tag{4.1}$$

when $h \neq 0$ and $c = 1, 5, 15$. Another reason we are using this simple form is that the correlation structure of the sample variogram estimator does not depend on the skewness parameter $\boldsymbol{\alpha}$. Generation of samples is straightforward using a stochastic representation, (2.4). Parameter c at (4.1) characterizes shape and sometimes range of the variogram.

Results of the simulations are shown in Table 4.1. For each situation, the mean of a parameter, with associated standard deviation, is computed over the 100 simulations. On each sample, the variogram is estimated by Matheron's classical estimator, denoted by M, and estimation of a parameter c, is done by generalized least squares(GLSE), (i) using correlation structure of a normal distribution(Genton, 1998a), (ii) using correlation structure of a generalized skew t distribution, (d) of Theorem 2.

$GLSE(i)$ and $GLSE(ii)$ denote that we use correlation structure of a normal distribution and a generalized skew t distribution, respectively. We write $\bar{c}$

TABLE 4.1 *Results of Simulations*

Variogram	*Parameter*	*M and GLSE(i)*	*M and GLSE(ii)*
exp(1)	$\bar{c}$	6.25	2.79
	$\hat{\sigma c}$	3.76	0.82
exp(5)	$\bar{c}$	11.70	8.18
	$\hat{\sigma c}$	1.54	0.95
exp(15)	$\bar{c}$	27.29	23.05
	$\hat{\sigma c}$	5.72	2.87

as the mean of a parameter c and $\hat{\sigma c}$ as corresponding standard deviation. It is straightforward to check that superiority of GLSE when we use appropriate correlation structure of the variogram estimator. More precise estimation can be done by using the highly robust estimator Q_{N_h} (Genton, 1998b) with GLSE(Genton, 1998a).

5. GENERALIZATION TO DATA IN R^d

We now extend Theorem 2 to data in R^d. Genton (1998a) extended the spatial design matrix to data in R^d with $d \geq 2$ using Kronecker products. Suppose spatial data are located on a hypercube with each edge holding n points (a total of n^d points), then the spatial design matrix is

$$\frac{1}{N_{\mathbf{h}_i}} A(\mathbf{h}) = \Sigma_{i=1}^{d} \left(\otimes_{k=1}^{i-1} I_n\right) \otimes \left(\frac{1}{N_{\mathbf{h}_i}} A_i(\mathbf{h})\right) \left(\otimes_{k=i+1}^{d} I_n\right), \tag{5.1}$$

where $\frac{1}{N_{\mathbf{h}_i}} A_i(\mathbf{h})$ is given by (3.1) and $\otimes$ is the Kronecker operator (Schott, 1997). Using the facts that $(A \otimes B)(C \otimes D) = (AC) \otimes (BD)$, and $tr(A \otimes B) = tr(A)tr(B)$, Theorem 3 of Genton (1998a) and above Theorem 2, we can prove the following Theorem. Basically Theorem 3 of Genton(1998a) shows that variance and covariance of Matheron's classical variogram estimator are linear combinations of one dimensional variances and covariances. Nevertheless, this is not the case for the correlation.

THEOREM 3. *If* $\mathbf{y} \sim GSt_{n^d}(\boldsymbol{\mu}, \Omega, \boldsymbol{\alpha}, \lambda, \nu)$, *where* $\boldsymbol{\mu}_{\mathbf{y}} = \mu_y \mathbf{1}_{n^d}$ *and* $\Omega = \sigma^2 I_{n^d}$, *then the sample variogram estimator (1.2) with* $A = A_i = (1/N_{\mathbf{h}_i}) A(\mathbf{h}_i)$ *and* $N_{\mathbf{h}_i} = dn^{d-1}(n - h_i)$, $i = 1, 2$, *satisfy:*

(a) $E(\mathbf{y}' A \mathbf{y}) = \frac{2\lambda\sigma^2}{\nu - 2}$,

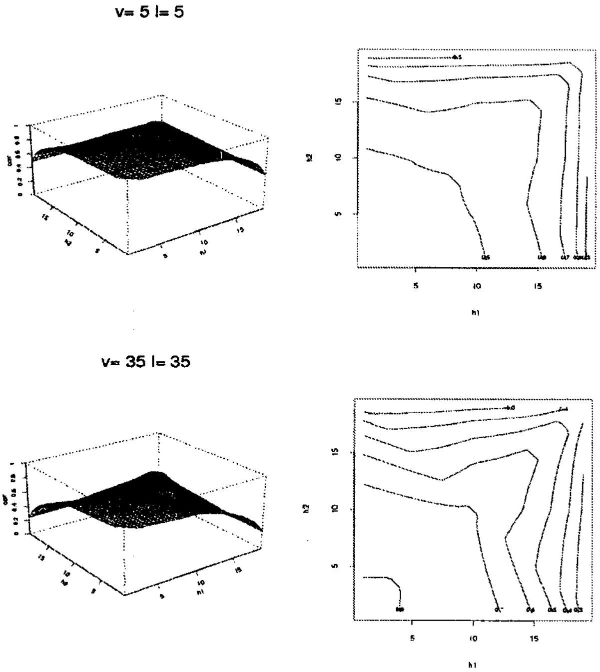

FIGURE 5.1 *These plots show the dependence of correlation on the lags h_1 and h_2 with $n = 20$ and different values of ν and λ for dimension 2. l denote λ.*

(b) $Var(\mathbf{y}'A\mathbf{y})$

$$= \begin{cases} \frac{4\sigma^4}{N_h}\left[\frac{\lambda^2(3n-4h)}{(\nu-2)(\nu-4)(n-h)} + \frac{2\lambda^2(n-h)}{(\nu-2)^2(\nu-4)} + \frac{2(d-1)(n-h)}{n}\right] & \text{if } h \text{ ¡ } \frac{n}{2} \\ \frac{8\sigma^4}{N_h}\left[\frac{\lambda^2}{(\nu-2)(\nu-4)} + \frac{\lambda^2(n-h)}{(\nu-2)^2(\nu-4)} + \frac{(d-1)(n-h)}{n}\right] & \text{otherwise} \end{cases},$$

and for $h_1 < h_2$

(c) $Cov(\mathbf{y}'A_1\mathbf{y}, \mathbf{y}'A_2\mathbf{y})$

$$= \begin{cases} \frac{4dn^{d-2}\sigma^4}{N_{h_1}N_{h_2}}\left[\frac{n\lambda^2(2n-h_1-2h_2)}{(\nu-2)(\nu-4)} + \frac{2n\lambda^2(n-h_1)(n-h_2)}{(\nu-2)^2(\nu-4)} + 2(d-1)(n-h_1)(n-h_2)\right] & \text{if } h_1+h_2<n \\ \frac{4dn^{d-2}\sigma^4}{N_{h_1}N_{h_2}}\left[\frac{n\lambda^2(n-h_2)}{(\nu-2)(\nu-4)} + \frac{2n\lambda^2(n-h_1)(n-h_2)}{(\nu-2)^2(\nu-4)} + 2(d-1)(n-h_1)(n-h_2)\right] & \text{otherwise} \end{cases}$$

(d) $Corr(\mathbf{y}'A_1\mathbf{y}, \mathbf{y}'A_2\mathbf{y})$

$$= \begin{cases} \frac{c_1}{\sqrt{v_1}\sqrt{v_2}} & if \;\; h_2 < \frac{n}{2} \\ \frac{c_1}{\sqrt{v_1}\sqrt{w_2}} & if \;\; h_2 \geq \frac{n}{2} \;\; and \;\; h_1 + h_2 < n \\ \frac{c_2}{\sqrt{v_1}\sqrt{w_2}} & if \;\; h_1 < \frac{n}{2} \;\; and \;\; h_1 + h_2 \geq n \\ \frac{c_2}{\sqrt{w_1}\sqrt{w_2}} & if \;\; h_1 \geq \frac{n}{2} \end{cases},$$

where $c_1 = n\lambda^2(\nu-2)(2n-h_1-2h_2)+2n\lambda^2(n-h_1)(n-h_2)+2(d-1)(\nu-2)^2(\nu-4)(n-h_1)(n-h_2)$, $c_2 = n\lambda^2(\nu-2)(n-h_2)+2n\lambda^2(n-h_1)(n-h_2)+2(d-1)(\nu-2)^2(\nu-4)(n-h_1)(n-h_2)$, $v_i = n\lambda^2(\nu-2)(3n-4h_i)+2n\lambda^2(n-h_i)^2+2(d-1)(\nu-2)^2(\nu-4)(n-h_i)^2$, and$w_i = 2(n-h_i)[n\lambda^2(\nu-2)+n\lambda^2(n-h_i)+(d-1)(\nu-2)^2 (\nu-4)(n-h_i)]$, $i = 1, 2$.

Note that correlation structure, (d), does depend on a parameter λ even though it does not depend on λ for one dimensional case (see Theorem 1 (d)). The perspective plots and corresponding contour plots of the correlation structure of the sample variogram estimator with 2 dimension are plotted in Figures 5.1, where $n = 20$, different values of ν, λ and the formula (d) of Theorem 3 is used. Comparing this plot with Figure 3.1, we can see the effect of dimension.

6. CONCLUTIONS

Moments of a variogram estimator for a generalized skew t distribution are calculated. Closed forms of those moments are derived. Surprisingly the correlation structure of a variogram estimator at different spatial lags for a generalized skew t distribution allowing both skewness and kurtosis is exactly same as that of a generalized t distribution only allowing kurtosis. So there is no effect of a skewness parameter when we apply generalized least squares suggested by Genton (1998a) under the assumption that data follows a generalized skew t distribution. One practical benefit regarding real data analysis from our approach is that we still have the explicit formula for the correlation structure of Matheron's classical variogram estimator under a generalized t distribution which is not included in Genton (2000). Since he only considered the elliptical distribution with Muirhead's kurtosis, κ, 0 (Muirhead, 1982). Muirhead's kurtosis of a generalized t distribution is $\kappa = 2/(\nu-4)$. Therefore we can improve variogram fitting using the correlation structure of Theorem 2 for one dimension and Theorem 3 for d dimension under the assumption that data follows a generalized t distribution.

APPENDIX

We first state two lemmas and derive the first four moments of a multivariate skew t distribution. These moments are used to calculate the first two moments of its quadratic forms.

LEMMA 1. *If $V \sim Gamma(\psi, \eta)$, then for any $a, b \in R$*

$$E_V\left\{\Phi(a\sqrt{V}+b)\right\} = P\left\{T \le a\sqrt{\psi/\eta}\right\},$$

where T denotes a non-central t variate with 2ψ degrees of freedom and non-centrality parameter $-b$.

LEMMA 2. *Suppose U and V are two symmetric, tri-ridged matrices, of size $n \times n$, given by*

$$U = \begin{pmatrix} a_1 & b_1 & & \\ & & \ddots & \\ b_1 & \ddots & & b_p \\ & \ddots & & \\ & b_p & & a_n \end{pmatrix}$$

and

$$V = \begin{pmatrix} c_1 & d_1 & & \\ & & \ddots & \\ d_1 & \ddots & & d_q \\ & \ddots & & \\ & d_q & & c_n \end{pmatrix}$$

where $1 \le p, q \le n$. We denote $\mathbf{a} = (a_1, \cdots, a_n)'$, $\mathbf{b} = (b_1, \cdots, b_p)'$, $\mathbf{c} = (c_1, \cdots, c_n)'$ and $\mathbf{d} = (d_1, \cdots, d_q)'$ as the ridge values. Then

$$tr(UV) = \begin{cases} \mathbf{a}\cdot\mathbf{c}, & p \ne q \\ \mathbf{a}\cdot\mathbf{c} + 2\mathbf{b}\cdot\mathbf{d}, & p = q \end{cases}$$

where $\cdot$ is the usual scalar product.

Azzalini and Capitanio (2003) developed a multivariate skew t distribution and gave the mean vector and variance matrix for the case when $\boldsymbol{\mu} = \mathbf{0}$. Genton et al. (2001) provided the moments of a skew normal distribution for the case when $\boldsymbol{\mu} = \mathbf{0}$ as follows:

THEOREM 4. *If* $\mathbf{z} \sim SN_n(\mathbf{0}, \Omega, \boldsymbol{\alpha})$, *then the first four moments of* $\mathbf{z}$ *are*

(a) $M_1 = E(\mathbf{z}) = \sqrt{\frac{2}{\pi}}\boldsymbol{\delta}$, *where* $\boldsymbol{\delta} = \frac{\Omega\boldsymbol{\alpha}}{(1+\boldsymbol{\alpha}'\Omega\boldsymbol{\alpha})^{1/2}}$,

(b) $M_2 = E(\mathbf{z} \otimes \mathbf{z}') = \Omega$,

(c) $M_3 = E(\mathbf{z} \otimes \mathbf{z}' \otimes \mathbf{z}) = \sqrt{\frac{2}{\pi}}[\boldsymbol{\delta} \otimes \Omega + vec(\Omega)\boldsymbol{\delta}' + (I_n \otimes \boldsymbol{\delta})\Omega - (I_n \otimes \boldsymbol{\delta})(\boldsymbol{\delta} \otimes \boldsymbol{\delta}')]$,

(d) $M_4 = E(\mathbf{z} \otimes \mathbf{z}' \otimes \mathbf{z} \otimes \mathbf{z}') = (I_{n^2} + K_{nn})(\Omega \otimes \Omega) + vec(\Omega)vec(\Omega)'$.

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). Also it is well-known that

$$E\left(V^{k/2}\right) = \frac{(\lambda/2)^{k/2}\Gamma((\nu-k)/2)}{\Gamma(\nu/2)}, \quad \text{where} \quad V \sim IG(\nu/2, \lambda/2). \tag{6.1}$$

Theorem 5 for the case when $\boldsymbol{\mu} \neq \mathbf{0}$ follows using the identities (2.4), (6.1), and some well-known properties of Kronecker and vec operators (Schott, 1997).

THEOREM 5. *If* $\mathbf{y} \sim GSt_n(\boldsymbol{\mu}, \Omega, \boldsymbol{\alpha}, \lambda, \nu)$, *then the first four moments of* $\mathbf{y}$ *are*

(a) $M_1 = \boldsymbol{\mu} + c^*\boldsymbol{\delta}$, where $c^* = \dfrac{(\lambda/\pi)^{1/2}\Gamma((\nu-1)/2)}{\Gamma(\nu/2)}$,

(b) $M_2 = \dfrac{\lambda}{\nu-2}\Omega + \boldsymbol{\mu}\boldsymbol{\mu}' + c^*(\boldsymbol{\mu}\boldsymbol{\delta}' + \boldsymbol{\delta}\boldsymbol{\mu}')$,

(c) $M_3 = \dfrac{\lambda}{\nu-2}[\Omega \otimes \boldsymbol{\mu} + \boldsymbol{\mu} \otimes \Omega + vec(\Omega) \otimes \boldsymbol{\mu}'] + \boldsymbol{\mu} \otimes \boldsymbol{\mu}' \otimes \boldsymbol{\mu} + \dfrac{c^*\lambda}{\nu-3}[\boldsymbol{\delta} \otimes \Omega + vec(\Omega)\boldsymbol{\delta}' + (I_n \otimes \boldsymbol{\delta})\Omega - \boldsymbol{\delta} \otimes \boldsymbol{\delta}' \otimes \boldsymbol{\delta}] + c^*[\boldsymbol{\delta} \otimes \boldsymbol{\mu}' \otimes \boldsymbol{\mu} + \boldsymbol{\mu} \otimes \boldsymbol{\delta}' \otimes \boldsymbol{\mu} + \boldsymbol{\mu} \otimes \boldsymbol{\mu}' \otimes \boldsymbol{\delta}]$,

(d) $M_4 = \dfrac{\lambda}{\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} + \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}' + \dfrac{\lambda^2}{(\nu-2)(\nu-4)}[(I_{n^2} + K_{nn})(\Omega \otimes \Omega) + vec(\Omega)vec(\Omega)'] + 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}' + \boldsymbol{\mu} \otimes \boldsymbol{\mu}' \otimes \boldsymbol{\mu} \otimes \boldsymbol{\delta}'] + \dfrac{c^*\lambda}{\nu-3}[\boldsymbol{\delta} \otimes \Omega \otimes \boldsymbol{\mu}' + vec(\Omega) \otimes \boldsymbol{\delta}' \otimes \boldsymbol{\mu}' + ((I_n \otimes \boldsymbol{\delta})\Omega) \otimes \boldsymbol{\mu}' + \boldsymbol{\delta}' \otimes \Omega \otimes \boldsymbol{\mu} + \boldsymbol{\delta} \otimes vec(\Omega)' \otimes \boldsymbol{\mu} + (\Omega(I_n \otimes \boldsymbol{\delta}')) \otimes \boldsymbol{\mu} + \boldsymbol{\mu}' \otimes \boldsymbol{\delta} \otimes \Omega + \boldsymbol{\mu}' \otimes (vec(\Omega)\boldsymbol{\delta}') + \boldsymbol{\mu}' \otimes ((I_n \otimes \boldsymbol{\delta})\Omega) + \boldsymbol{\mu} \otimes \boldsymbol{\delta}' \otimes \Omega + \boldsymbol{\mu} \otimes \boldsymbol{\delta} \otimes vec(\Omega)' + \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} - \boldsymbol{\mu}' \otimes \boldsymbol{\delta} \otimes \boldsymbol{\delta}' \otimes \boldsymbol{\delta} - \boldsymbol{\mu} \otimes \boldsymbol{\delta}' \otimes \boldsymbol{\delta} \otimes \boldsymbol{\delta}']$.

Using the first four moments of the random vector $\mathbf{y}$ given in Theorem 5, we can calculate the first two moments of its quadratic form.

THEOREM 6. *If* $\mathbf{y} \sim GSt_n(\boldsymbol{\mu}, \Omega, \boldsymbol{\alpha}, \lambda, \nu)$ *and* A, B *are two symmetric* $n \times n$ *matrices, then*

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

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

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

where $tr(\cdot)$ denotes the trace of a matrix. The proof rests on Theorem 5 and the following relations (Schott, 1997):

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

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

ACKNOWLEDGMENTS

The author thanks Adelchi Azzalini for valuable comments. This work was supported by the faculty research fund of Konkuk University in 2004.

REFERENCES

ARELLANO-VALLE, R.B. AND BOLFARINE, H. (1995). "On some characterizations of the t−distribution", *Statistics & Probability Letters*, **25**, 79–85.

AZZALINI, A. AND CAPITANIO (1999)."Statistical applications of the multivariate skew normal distribution", *Journal of the Royal Statistical Society Series B*, **61**, 579–602.

AZZALINI, A. AND CAPITANIO (2003). "Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t distribution", *Journal of the Royal Statistical Society Series B*, **65**, 367–389.

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.)*, Wiley, New York.

FANG, K.-T., KOTZ, S., AND NG, K.W. (1989). *Symmetric Multivariate and Related Distributions*, Chapman and Hall, New York.

GENTON, M.G. (1998a). "Variogram fitting by generalized least squares using an explicit formula for the covariance structure", *Mathematical Geology*, **30**, 323–345.

GENTON, M.G. (1998b). "Highly robust Variogram estimation", *Mathematical Geology*, **30**, 213–221.

GENTON, M.G. (2000). "The correlation structure of Matheron's classical variogram estimator under elliptically contoured distributions", *Mathematical Geology*, **32**, 127–137.

GENTON, M.G., HE, L. AND LIU, X. (2001). "Moments of skew-normal random vectors and their quadratic forms", *Statistics & Probability Letters*, **51**, 319–325.

GORSICH, D.J., GENTON, M.G., AND STRANG, G. (2002). "Eigenstructures of Spatial Design Matrices", *Journal of Multivariate Analysis*, **80**, 138–165.

HENDERSON, H.V. AND SEARLE, S.R. (1981). "On deriving the inverse of a sum of matrices", *SIAM Review*, **33**, 53–60.

KIM, H.-M., MALLICK, B.K. (2003). "Moments of random vectors with skew t distribution and their quadratic forms", *Statistics & Probability Letters*, **63**, 417-423.

LI, G. (1987). "Moments of a random vector and its quadratic forms", *Journal of Statistical Applied Probability*, **2**, 219–229.

MARDIA, K.V. (1970). "Measures of multivariate skewness and kurtosis with applications", *Biometrika*, **57**, 519–530.

MARDIA, K.V. (1974). "Applications of some measures of multivariate skewness and kurtosis in testing normality and robustness studies", *Sankhya Ser. B*, **36**, 115–128.

MUIRHEAD, R.J. (1982). *Aspects of Multivariate Statistical Theory*, Wiley, New York.

SCHOTT, JAMES R. (1997). *Matrix Analysis for Statistics*, Wiley, New York.