A Multivariate Skew Normal Distribution

by

Arjun K. Gupta
Bowling Green State University, U.S.A.
gupta@bgnet.bgsu.edu

Graciela González-Farías
Centro de Investigación en Matemáticas, México
farias@cimat.mx

and

J. Armando Domínguez-Molina
Universidad de Guanajuato, México
jadguez@cimat.mx

Abstract: In this paper, we define a new class of multivariate skew-normal distribution. Its properties are studied. In particular we derive its density, moment generating function and the first two moments. The moment generating function of the corresponding quadratic form and its moments are also given.

1 Introduction

The univariate skew normal distribution was introduced by Azzalini (1985, 1986) and its multivariate version by Azzalini and Dalla Valle (1996) and Azzalini and Capitanio (1999). These classes of distributions include the normal and have some properties like the normal and yet are skew. They are useful in studying robustness. Whereas Azzalini and Dalla Valle (1996) and Azzalini and Capitanio (1999) obtain the multivariate distribution by conditioning on one random variable being positive; we condition on the same number of random variables being positive. Hence, by construction, in the univariate case the two families are the same. From related reference ones refers to Kelker (1970), Fang et al (1990), Gupta and Varga (1993), and Gupta and Nagar (2000).

We study in this work this new family of multivariate skew normal (MSN) distributions. This class of distributions includes the normal distribution and has some properties like the normal family and yet are skewed. In the next section we give the definition of this distribution, some of its properties and a method for simulating random vectors with this distribution. In Section 3 we compute the first two moments of the MSN distribution and the expectation of some quadratic forms. We discuss in detail the bivariate skew normal distribution in Section 4. Finally in Section 5 we discuss a general version of the MSN distribution.

2 The multivariate skew normal distribution

A continuous random vector $Y\ (p \times 1)$ is said to have a *multivariate skew normal distribution* (MSN) if its p.d.f. is given by

$$f_p\left(y; \mu, \Sigma, D\right) = \frac{1}{\Phi_p\left(0; I + D\Sigma D'\right)} \phi_p\left(y; \mu, \Sigma\right) \Phi_p\left[D\left(y - \mu\right)\right], \tag{2.1}$$

where $\mu \in \mathbb{R}^p,\ \Sigma > 0,\ D\,(p \times p)\,,\ \phi_p\,(\cdot;\mu,\Sigma)$ and $\Phi_p\,(\cdot;\Sigma)$ denote the p.d.f. and the c.d.f. of p-dimensional normal distribution with mean μ and covariance matrix $\Sigma > 0$. However we will denote $\phi_p\,(\cdot;0,I)$ by $\phi_p\,(\cdot)\,,$ and $\Phi_p\,(\cdot;0,\Sigma)$ by $\Phi_p\,(\cdot;\Sigma)\,.$ We will denote that a random vector is distributed according to MSN with parameters μ,Σ,D by writing $Y \sim SN_p\,(\mu,\Sigma,D)\,.$

In this section we first prove a lemma which is used in the sequel (see Azzalini, 1985, Zacks 1981, Marsaglia, 1967).

2.1 Some properties of $SN_p\,(\mu,\Sigma,D)$

I) If $D' = (0,...,0,\delta,0,...,0)\,,\ \delta = (\delta_1,...,\delta_p)'\,,\ 0\,(p \times 1)$ is a null vector, then

$$f_p\,(y;\mu,\Sigma,D) = Azzf\left(y;\mu,\Sigma,\delta'\,(y-\mu)\right),$$

where $Azzf$ denotes the skew normal density of Azzalini and Dalla Valle (1996).

II) If $D = [\mathrm{diag}\,(\delta_1,...,\delta_p)]\,\Sigma^{-1/2}$ then the density function (2.1) reduces to

$$f\,(y;\mu,\Sigma,D) = 2^p\phi_p\,(y;\mu,\Sigma)\,\Phi_p\left\{[\mathrm{diag}\,(\delta_1,...,\delta_p)]\,\Sigma^{-1/2}\,(y-\mu)\right\}.$$

III) The distribution function of $Y \sim SN_p\,(\mu,\Sigma,D)$ is

$$\begin{aligned}
F_p\,(y;\mu,\Sigma,D) &= \Pr\,(Y \le y) = \Pr\,(Y_1 \le y_1,...,Y_p \le y_p) \\
&= \frac{1}{\Phi_p\,(0;I+D\Sigma D')}\int_{-\infty}^{y_1}\cdots\int_{-\infty}^{y_p}\phi_p\,(t;\mu,\Sigma)\,\Phi_p\,[D\,(t-\mu)]\,dt \\
&= \frac{1}{\Phi_p\,(0;I+D\Sigma D')}\int_{-\infty}^{y_1}\cdots\int_{-\infty}^{y_p}\int_{-\infty}^{[D(s-\mu)]_1}\cdots\int_{-\infty}^{[D(s-\mu)]_p}\phi_p\,(t;\mu,\Sigma)\,\phi_p\,(s)\,dsdt \\
&= \frac{1}{\Phi_p\,(0;I+D\Sigma D')}\Pr\,(U_1 \le y_1,...,U_p \le y_p, V \le D\,(U-\mu))
\end{aligned}$$

where $D\,(t-\mu) = \left\{[D\,(t-\mu)]_1\,,...,[D\,(t-\mu)]_p\right\}'$, $U \sim N_p\,(\mu,\Sigma)$, $V \sim N_p\,(0,I)$ and U is independent of V. If $W = V - D\,(U-\mu)$ we have that

$$\begin{pmatrix} U \\ W \end{pmatrix} \sim N_{2p}\left[\begin{pmatrix} \mu \\ 0 \end{pmatrix},\begin{pmatrix} \Sigma & -\Sigma D' \\ -D\Sigma & I+D\Sigma D' \end{pmatrix}\right].$$

Thus, the distribution function of Y is

$$F_p\,(y;\mu,\Sigma,D) = \Phi_{2p}\left[\begin{pmatrix} y-\mu \\ 0 \end{pmatrix};\begin{pmatrix} \Sigma & -\Sigma D' \\ -D\Sigma & I+D\Sigma D' \end{pmatrix}\right].$$

Proposition 2.1. Let $Y \sim SN_p\,(\mu,\Sigma,D)\,.$ If $\Sigma = \mathrm{diag}\left(\sigma_1^2,...,\sigma_p^2\right),\ D = \mathrm{diag}\,(\delta_1,...,\delta_p)\,,$ then

$$\begin{aligned}
f_p\,(y;\mu,\Sigma,D) &= \Phi_p^{-1}\,(0;I+D\Sigma D')\,\phi_p\,(y;\mu,\Sigma)\,\Phi_p\,[D\,(y-\mu)] \\
&= \prod_{i=1}^{p}2\phi\left(y_i;\mu_i,\sigma_i^2\right)\Phi\,[\delta_i\,(y_i-\mu_i)] \\
&= \prod_{i=1}^{p}f_1\left(y_i;\mu_i,\sigma_i^2,\delta_i\right),
\end{aligned}$$

i.e., Y_i's are independently distributed as $SN_1\left(\mu_i, \sigma_i^2, \delta_i\right)$. This means that if we have a collection of n independent random variables with distribution skew normal then the joint distribution of the n random variables is multivariate skew normal. This property does not hold for the multivariate skew normal distribution defined by Azzalini and Dalla Valle (1996). Thus, $\Sigma = \sigma^2 I$ and $D = \delta I$ give us the i.i.d. case. Notice that Azzalini (1985, Section 4) proposed a generalization which is a particular case of Proposition 2.1.

2.1.1 The moment generating function

The following lemma is useful for evaluating some integrals that we will use in the rest of the paper.

Lemma 2.2. Let B be a constant $(p \times p)$ matrix, and $a \in \mathbb{R}^p$. If $V \sim N_p(\mu_1, \Sigma)$ then

$$E_V\left[\Phi_p\left(a + BV; \mu_2, \Omega\right)\right] = \Phi_p\left(a - \mu_2 + B\mu_1; \Omega + B\Sigma B'\right).$$

Proof.

$$\begin{aligned} E_V\left[\Phi_p\left(a + BV; \mu_2, \Omega\right)\right] &= E_V\left[\Pr\left(U \leq a + BV | V\right)\right] \\ &= \Pr\left(U \leq a + BV\right) \end{aligned}$$

where $U \sim N_p(\mu_2, \Omega)$. Then

$$E_V\left[\Phi_p\left(a + BV; \mu_2, \Omega\right)\right] = \Pr\left(U - BV \leq a\right)$$

Given that $U - BV \sim N_p(\mu_2 - B\mu_1, \Omega + B\Sigma B')$ we get

$$\Pr\left(U - BV \leq a\right) = \Phi_p\left(a; \mu_2 - B\mu_1, \Omega + B\Sigma B'\right).$$

∎

Proposition 2.3. If $Y \sim SN_p(\mu, \Sigma, D)$ then its m.g.f. is given by

$$M_Y(t) = \frac{\Phi_p\left(D\Sigma t; I + D\Sigma D'\right)}{\Phi_p\left(0; I + D\Sigma D'\right)} e^{\mu' t + \frac{1}{2} t' \Sigma t}, \quad t \in \mathbb{R}^p. \tag{2.2}$$

Proof. By definition of m.g.f. we have

$$\begin{aligned} M_Y(t) = Ee^{t'Y} &= \Phi_p^{-1}\left(0; I + D\Sigma D'\right) \int_{\mathbb{R}^p} e^{t'y} \phi_p\left(y; \mu, \Sigma\right) \Phi_p\left[D\left(y - \mu\right)\right] dy \\ &= \Phi_p^{-1}\left(0; I + D\Sigma D'\right) e^{t'\mu + \frac{1}{2} t' \Sigma t} \int_{\mathbb{R}^p} \phi_p\left(y; \mu + \Sigma t, \Sigma\right) \Phi_p\left[D\left(y - \mu\right)\right] dy \\ &= \frac{E_W\left[\Phi_p\left(DW - D\mu\right)\right]}{\Phi_p\left(0; I + D\Sigma D'\right)} e^{t'\mu + \frac{1}{2} t' \Sigma t}, \end{aligned}$$

where $W \sim N_p(\mu + \Sigma t, \Sigma)$. The result follows from Lemma 2.2. ∎

If $D = \operatorname{diag}(\delta_1, ..., \delta_p)$ and $\Sigma = \operatorname{diag}\left(\sigma_1^2, ..., \sigma_p^2\right)$ then

$$M_Y(t) = 2^p \prod_{i=1}^{p} \Phi\left(\frac{\delta_i \sigma_i^2 t_i}{\sqrt{1 + \delta_i^2 \sigma_i^2}}\right) e^{t_i \mu_i + \frac{1}{2} \sigma_i^2 t_i^2},$$

which is the independent case.

From Proposition 2.3 we give the distribution of a linear transform of Y.

Proposition 2.4. Let $Y \sim SN_p(\mu, \Sigma, D)$. Let $A(p \times p)$ a non-singular matrix and $b \in \mathbb{R}^p$ be constants, then $AY + b \sim SN_p\left(b + A\mu; A\Sigma A', DA^{-1}\right)$.

Proof. For $t \in \mathbb{R}^p$ the m.g.f. of $AY + b$ is given by

$$\begin{aligned} M_{AY+b}(t) &= E\left[e^{t'(AY+b)}\right] = e^{t'b} M_Y(A't) \\ &= \frac{\Phi_p(D\Sigma A't; I + D\Sigma D')}{\Phi_p(0; I + D\Sigma D')} e^{t'(A\mu+b)+\frac{1}{2}t'A\Sigma A't} \\ &= \frac{\Phi_p\left[\left[DA^{-1}\right][A\Sigma A']t; I + \left[DA^{-1}\right][A\Sigma A']\left(DA^{-1}\right)'\right]}{\Phi_p\left(0; I + [DA^{-1}][A\Sigma A'](DA^{-1})'\right)} e^{t'(A\mu+b)+\frac{1}{2}t'A\Sigma A't}. \end{aligned}$$

■

2.2 Construction of the MSN

In this section we give a derivation of the MSN distribution based on a partitioned-conditional method. This procedure is useful for simulating random vectors with this distribution.

Proposition 2.5. Let $X \sim N_p(\xi, I + D\Sigma D')$ and $Y \sim N_p(\mu, \Sigma)$ with $\text{Cov}(X, Y) = D\Sigma$. Then the distribution of the random vector $Y|(X > \xi)$ is $SN_p(\mu, \Sigma, D)$.

Proof. Write the joint distribution of X and Y as

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim N_{2p}\left[\begin{pmatrix} \xi \\ \mu \end{pmatrix}, \begin{pmatrix} I + D\Sigma D' & D\Sigma \\ \Sigma D' & \Sigma \end{pmatrix}\right],$$

then

$$X|Y \sim N_p(\xi + D(y - \mu), I).$$

By using the fact

$$f_{Y|X}(y|X > \xi) = \frac{f_Y(y)\Pr(X > \xi|y)}{\Pr(X > \xi)}$$

we observe that

$$\begin{aligned} f_{Y|X}(y|X > \xi) &= \frac{\phi_p(y; \mu, \Sigma)}{\Pr(\xi - X < 0)} \Pr(\xi - X < 0|y) \\ &= \frac{\phi_p(y; \mu, \Sigma)}{\Phi_p(0; I + D\Sigma D')} \Phi_p[D(y - \mu)]. \end{aligned}$$

■

3 Some expected values

3.1 First and second moment of the MSN distribution

Let Y be a random variable with distribution $SN_p(\mu, \Sigma, D)$. In order to compute the first and second moment for the MSN distribution we consider the derivatives of the m.g.f. given in equation (2.2). From the Appendix A we have that the mean of Y is

$$EY = \mu + \frac{G^p(0; D\Sigma, I + D\Sigma D')}{\Phi_p(0; I + D\Sigma D')},$$

where

$$G^p(0; D\Sigma, I + D\Sigma D') = \left.\frac{\partial}{\partial t}\Phi_p(D\Sigma t, I + D\Sigma D')\right|_{t=0}.$$

From Lemma C.1, we get that the i^{th} element of $G^p(0; D\Sigma, I + D\Sigma D')$ is

$$G_i^p(0; D\Sigma, I + D\Sigma D') = \frac{1}{\sqrt{2\pi}\,|I + D\Sigma D'|^{1/2}} \sum_{j=1}^{p} (D\Sigma)_{ji} \left|H_{(j)}^{-1}\right|^{1/2} \Phi_{p-1}\left[0; H_{(j)}^{-1}\right],$$

where $H_{(j)}$ is the matrix obtained by eliminating the j^{th} row and the j^{th} column from $(I + D\Sigma D')^{-1}$.

If $D = \text{diag}(\delta_1, ..., \delta_p)$ and $\Sigma = \text{diag}(\sigma_1^2, ..., \sigma_p^2)$ then

$$G_i^p(0; D\Sigma, I + D\Sigma D') = \frac{1}{2^{p-1}\sqrt{2\pi}} \frac{\delta_i \sigma_i^2}{1 + \delta_i^2 \sigma_i^2}, \tag{3.1}$$

which implies that

$$\begin{aligned} EY &= \mu + 2^p G^p(0; D\Sigma, I + D\Sigma D') \\ &= \mu + \sqrt{\frac{2}{\pi}} \left\{ \frac{\delta_1 \sigma_1^2}{1 + \delta_1^2 \sigma_1^2}, ..., \frac{\delta_p \sigma_p^2}{1 + \delta_p^2 \sigma_p^2} \right\}'. \end{aligned}$$

From Appendix B we get that the second moment of Y is

$$EYY' = \Sigma + \mu \frac{G^{p\prime}(0; D\Sigma, I + D\Sigma D')}{\Phi_p(0; I + D\Sigma D)} + \frac{G_{[2]}^p(0; D\Sigma, I + D\Sigma D')}{\Phi_p(0; I + D\Sigma D)} + \frac{G^p(0; D\Sigma, I + D\Sigma D')}{\Phi_p(0; I + D\Sigma D)}\mu' + \mu\mu', \tag{3.2}$$

where

$$G^{p\prime}(t; D\Sigma, I + D\Sigma D') = [G^p(t; D\Sigma, I + D\Sigma D')]'$$

$$G_{[2]}^p(0; D\Sigma, I + D\Sigma D') = \left.\frac{\partial}{\partial t} G^{p\prime}(t; D\Sigma, I + D\Sigma D')\right|_{t=0} = \left.\frac{\partial^2}{\partial t \partial t'}\Phi_p(D\Sigma t; I + D\Sigma D)\right|_{t=0},$$

The expression for $G^p_{[2]}(0; D\Sigma, I + D\Sigma D')$ is given in the Appendix by Lemma C.2. If $D = \text{diag}\left(\delta_1^2, ..., \delta_p^2\right)$ and $\Sigma = \text{diag}\left(\sigma_1^2, ..., \sigma_p^2\right)$ are diagonal then by equation (C.8) we get

$$G^p_{i,j}(0; D\Sigma, I + D\Sigma D') = \begin{cases} \dfrac{\delta_i \delta_j \sigma_i^2 \sigma_j^2}{2^{p-1}\pi\left(1+\delta_i^2\sigma_i^2\right)\left(1+\delta_j^2\sigma_j^2\right)}, & \text{if } i \neq j, \\ 0, & \text{if } \quad i = j. \end{cases} \tag{3.3}$$

In order to evaluating the variance of Y we need the following quantity

$$\begin{aligned}(EY)(EY)' = \mu\mu' &+ \frac{G^p(0; D\Sigma, I + D\Sigma D')}{\Phi_p(0; I + D\Sigma D')}\mu' + \mu\frac{G^{p\prime}(0; D\Sigma, I + D\Sigma D')}{\Phi_p(0; I + D\Sigma D)} \\ &+ \frac{G^p(0; D\Sigma, I + D\Sigma D')\, G^{p\prime}(0; D\Sigma, I + D\Sigma D')}{\Phi_p^2(0; I + D\Sigma D')},\end{aligned}$$

thus

$$\text{Var}(Y) = \Sigma - \frac{G^p(0; D\Sigma, I + D\Sigma D')\, G^{p\prime}(0; D\Sigma, I + D\Sigma D')}{\Phi_p^2(0; I + D\Sigma D')} + \frac{G^p_{[2]}(0; D\Sigma, I + D\Sigma D')}{\Phi_p(0; I + D\Sigma D)}.$$

If $D = \text{diag}(\delta_1, ..., \delta_p)$ and $\Sigma = \text{diag}\left(\sigma_1^2, ..., \sigma_p^2\right)$ are diagonal then from equations (3.1) and (3.3) and letting $\Delta = \text{diag}\left(\frac{\delta_1\delta_1}{1+\delta_1^2\sigma_1^2}, ..., \frac{\delta_p\delta_p}{1+\delta_p^2\sigma_p^2}\right)$ we get

$$\text{Var}(Y) = \Sigma - \frac{2}{\pi}\Delta^2. \tag{3.4}$$

3.2 Moment generating function of quadratic forms

In this section we consider the quadratic form $Y'AY$, where $Y \sim SN_p(0, \Sigma, D)$. By using Lemma 2.2 we can obtain the m.g.f. of $Y'AY$.

Proposition 3.1. Let $Y \sim SN_p(0, \Sigma, D)$ then
(i) The m.g.f. of $Y'AY$ is

$$M_{Y'AY}(t) = \frac{\Phi_p\left[0; I + D\Sigma(I - 2tA\Sigma)^{-1} D'\right]}{\Phi_p(0; I + D\Sigma D')} |I - 2tA\Sigma|^{-1/2}.$$

(ii) The joint m.g.f. of $Y'A_1Y$ and $Y'A_2Y$ is

$$M_{Y'A_1Y, Y'A_2Y}(t_1, t_2) = \frac{\Phi_p\left[0; I + D\Sigma\left(I - 2\left(t_1A_1 + t_2A_2\right)\Sigma\right)^{-1} D'\right]}{\Phi_p(0; I + D\Sigma D')} \left|I - 2\left(t_1A_1 + t_2A_2\right)\Sigma\right|^{-1/2}.$$

Proof. (i) We have

$$\begin{aligned}M_{Y'AY}(t) = Ee^{tY'AY} &= \Phi_p^{-1}(0; I + D\Sigma D) \int\limits_{\mathbb{R}^p} e^{ty'Ay}\phi_p(y; 0, \Sigma)\,\Phi_p(Dy)\,dy \\ &= \Phi_p^{-1}(0; I + D\Sigma D)\,|I - 2tA\Sigma|^{-1/2} \int_{\mathbb{R}^p} \phi_p\left[y; 0, \Sigma(I - 2tA\Sigma)^{-1}\right]\Phi_p(Dy)\,dy.\end{aligned}$$

Now by using Lemma 2.2 we get

$$M_{Y'AY}(t) = \frac{\Phi_p\left[0; I + D\Sigma(I - 2tA\Sigma)^{-1} D\right]}{\Phi_p(0; I + D\Sigma D)} |I - 2tA\Sigma|^{-1/2}.$$

(ii) By noting that $M_{Y'A_1Y, Y'A_2Y}(t_1, t_2) = M_{Y'(A_1t_1 + A_2t_2)Y}(1)$ and using (i) the proof is complete. ■

Corollary 3.2. If Σ and D are diagonal matrices and $Y \sim SN_p(0, \Sigma, D)$, then $Y'\Sigma^{-1}Y \sim \chi_p^2$.

From (3.2) and the identity $E(Y'AY) = \text{tr}[AE(Y'Y)]$ we can evaluate the expectation of $Y'AY$. For computing moments of higher orders is better to use the derivatives of $M_{Y'AY}(t)$ evaluated at zero. The arguments for derivative are similar as those given for calculating $E(Y'Y)$ in lemmas C.1 and C.2. The expressions of $E(Y'AY)$ and $\text{Var}(Y'AY)$ have a closed form and can be evaluated without major problem, however they are tedious and will not be included here.

4 The bivariate skew normal distribution

In this section we obtain a closed expression for the density of a random vector with distribution $SN_2(\mu, \Sigma, D)$.

When $D = \begin{pmatrix} \delta_1 & \delta_2 \\ 0 & 0 \end{pmatrix}$ or $D = \begin{pmatrix} 0 & 0 \\ \delta_1 & \delta_2 \end{pmatrix}$ the bivariate skew normal density reduces to

$$\begin{aligned} f_2(x, y; \mu, \Sigma, D) &= \Phi_2^{-1}\left[0; \begin{pmatrix} 1 + \delta_1^2\sigma_1^2 + 2\delta_1\delta_2\sigma_1\sigma_2\rho + \delta_2^2\sigma_2^2 & 0 \\ 0 & 1 \end{pmatrix}\right] \phi_2(x, y; \mu, \Sigma) \\ &\times \Phi_2\left[\begin{pmatrix} \delta_1(x - \mu_1) + \delta_2(y - \mu_2) \\ 0 \end{pmatrix}\right] \\ &= 2\phi_2(x, y; \mu, \Sigma)\, \Phi\left[\delta_1(x - \mu_1) + \delta_2(y - \mu_2)\right] \end{aligned}$$

which is the same as given by Azzalini and Dalla Valle (1996). This case will not be discussed in this paper.

We will consider the general case where $D = \begin{pmatrix} \delta_{11} & \delta_{12} \\ \delta_{21} & \delta_{22} \end{pmatrix}$. The next lemma helps us to evaluate the constant of the density (2.1).

Lemma 4.1. If $R = \begin{pmatrix} r_1^2 & r_{12} \\ r_{12} & r_2^2 \end{pmatrix}$ is a positive definite matrix, then $\Phi_2(0; R) = \frac{1}{2} - \frac{1}{2\pi} \arccos \frac{r_{12}}{r_1 r_2}$.

Note that if $r_{12} = 0$ we get $\Phi_2(0; R) = \frac{1}{4}$, because $\arccos(0) = \frac{1}{2}\pi$, this result coincides with the fact that when $R = \text{diag}\left(r_1^2, r_2^2\right)$ it follows that

$$\Phi_2(0; R) = \Phi\left(0, r_1^2\right)\Phi\left(0, r_2^2\right) = \left(\tfrac{1}{2}\right)\left(\tfrac{1}{2}\right) = \tfrac{1}{4}.$$

Corollary 4.2: If $\Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho \\ \sigma_1\sigma_2\rho & \sigma_2^2 \end{pmatrix}$ is a positive definite matrix, and $D = \begin{pmatrix} \delta_{11} & \delta_{12} \\ \delta_{21} & \delta_{22} \end{pmatrix}$ is an arbitrary matrix, then

$$\Phi_2(0; I + D\Sigma D') = \tfrac{1}{2} - \tfrac{1}{2\pi} \arccos(\rho_{D\Sigma}), \tag{4.1}$$

where

$$\rho_{D\Sigma} = \frac{\delta_{21}\delta_{11}\sigma_1^2 + \delta_{22}\delta_{12}\sigma_2^2 + (\delta_{12}\delta_{21} + \delta_{22}\delta_{11})\,\sigma_1\sigma_2\rho}{\sqrt{\left(1 + \delta_{11}^2\sigma_1^2 + 2\delta_{11}\delta_{12}\sigma_1\sigma_2\rho + \delta_{12}^2\sigma_2^2\right)\left(1 + \delta_{21}^2\sigma_1^2 + 2\delta_{21}\delta_{22}\sigma_1\sigma_2\rho + \delta_{22}^2\sigma_2^2\right)}}.$$

If $\rho = 0$

$$\rho_{D\Sigma} = \frac{\delta_{21}\delta_{11}\sigma_1^2 + \delta_{22}\delta_{12}\sigma_2^2}{\sqrt{\left(1 + \delta_{11}^2\sigma_1^2 + \delta_{12}^2\sigma_2^2\right)\left(1 + \delta_{21}^2\sigma_1^2 + \delta_{22}^2\sigma_2^2\right)}}$$

and if $D = \mathrm{diag}\,(\delta_1, \delta_2)$

$$\rho_{D\Sigma} = \frac{\delta_2\delta_1\sigma_1\sigma_2\rho}{\sqrt{\left(1 + \delta_1^2\sigma_1^2\right)\left(1 + \delta_2^2\sigma_2^2\right)}}, \tag{4.2}$$

and if $D = \mathrm{diag}\,(\delta_1, \delta_2)$ and $\delta_1 = 0$ and/or $\delta_2 = 0$ then $\rho_{D\Sigma} = 0$.
If $\delta_{21} = \delta_{11}$ and $\delta_{12} = \delta_{22}$, then

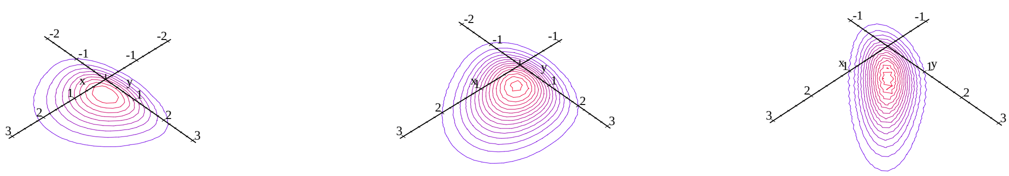
$$\rho_{D\Sigma} = \frac{\delta_{11}^2\sigma_1^2 + \delta_{22}^2\sigma_2^2 + 2\delta_{22}\delta_{11}\sigma_1\sigma_2\rho}{1 + \delta_{11}^2\sigma_1^2 + 2\delta_{11}\delta_{22}\sigma_1\sigma_2\rho + \delta_{22}^2\sigma_2^2}$$

Let $\mu = (\mu_1, \mu_2)'$ and $\Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho \\ \sigma_1\sigma_2\rho & \sigma_2^2 \end{pmatrix}$. From (2.1) and (4.1) we have that the density of the *bivariate skew normal distribution* (BSN) is given by

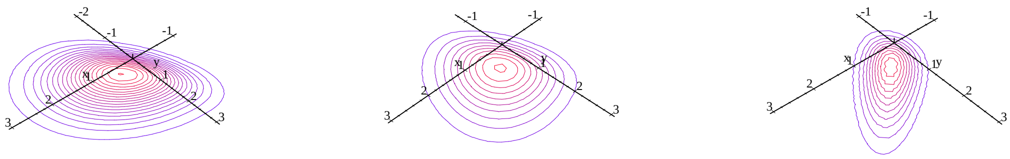
$$f_2\,(x, y; \mu, \Sigma, D) = \frac{\exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{(x-\mu_1)^2}{\sigma_1^2} - 2\frac{\rho(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2} + \frac{(y-\mu_2)^2}{\sigma_2^2}\right]\right\}}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}\left[\frac{1}{2} - \frac{1}{2\pi}\arccos\left(\rho_{D\Sigma}\right)\right]}$$
$$\times\,\Phi\left[\delta_{11}(x-\mu_1) + \delta_{12}(y-\mu_2)\right]\Phi\left[\delta_{21}(x-\mu_1) + \delta_{22}(y-\mu_2)\right].$$

With the explicit expression for the density of the BSN distribution we can draw some contours of this density. Note that these contours are not elliptical.

(i) For $\mu = \binom{0}{0}$, $\Sigma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$, $D = \begin{pmatrix} 3 & 0 \\ 0 & 1 \end{pmatrix}$ and $\rho = 0, \frac{1}{2}, \frac{9}{10}$, respectively:



(ii). For $\mu = \binom{0}{0}$, $\Sigma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$, $D = \begin{pmatrix} 3 & 1 \\ 1 & 1 \end{pmatrix}$ and $\rho = 0, \frac{1}{2}, \frac{9}{10}$, respectively:



4.1 The marginal and conditional density of the BSN

From Proposition 2.3 we have that the m.g.f. of (X, Y) is

$$M_{X,Y}(t_1, t_2) = \frac{\Phi_p(D\Sigma t; I + D\Sigma D')}{\Phi_p(0; I + D\Sigma D')} e^{\mu' t + \frac{1}{2} t' \Sigma t},$$

thus, the m.g.f. of X is given by

$$\begin{aligned} M_X(t_1) &= M_{X,Y}(t_1, 0) \\ &= \frac{\Phi_2\left[(D_1\Sigma_{11} + D_2\Sigma_{21}) t_1; I + D\Sigma D'\right]}{\Phi_2(0; I + D\Sigma D')} e^{\mu_1 t_1 + \frac{1}{2}\sigma_1^2 t_1^2}, \end{aligned} \tag{4.3}$$

where D was partitioned as $D = \begin{pmatrix} D_1(2 \times 1) & D_2(2 \times 1) \end{pmatrix}$. It is easy to verify that (4.3) corresponds to a random variable with distribution GMSN given in Section 5,

$$GSN_{1,2}\left(\mu_1, \Sigma_{11}, D_1 + D_2\Sigma_{21}\Sigma_{11}^{-1}, D_1\mu_1 + D_2\Sigma_{21}\Sigma_{11}^{-1}\mu_1, I + D_2\left(\Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}\right) D_2'\right).$$

Hence the marginal density of X is given by

$$\begin{aligned} f_X(x, \mu_1, \Sigma, D) \quad &= \quad \Phi_2^{-1}(0; I + D\Sigma D')\, \phi(x; \mu_1, \Sigma_{11}) \\ &\quad \times \Phi_2\left[\left(D_1 + D_2\Sigma_{21}\Sigma_{11}^{-1}\right)(x - \mu_1); I + D_2\left(\Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}\right) D_2'\right]. \end{aligned}$$

By using (4.1) and replacing $\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$ by $\Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho \\ \sigma_1\sigma_2\rho & \sigma_2^2 \end{pmatrix}$ and $D = (D_1, D_2)$ by $D = \begin{pmatrix} \delta_{11} & \delta_{12} \\ \delta_{21} & \delta_{22} \end{pmatrix}$ we get

$$\begin{aligned} f_X(x, \mu_1, \Sigma, D) = & \frac{\phi\left(x; \mu_1, \sigma_1^2\right)}{\frac{1}{2} - \frac{1}{2\pi}\arccos\left(\rho_{D\Sigma}\right)} \\ & \times \Phi_2\left\{\left[\begin{pmatrix} \delta_{11} \\ \delta_{21} \end{pmatrix} + \begin{pmatrix} \delta_{12} \\ \delta_{22} \end{pmatrix}\frac{\sigma_2}{\sigma_1}\rho\right](x - \mu_1); \begin{pmatrix} 1 + \delta_{12}^2\sigma_2^2\left(1 - \rho^2\right) & \delta_{12}\delta_{22}\sigma_2^2\left(1 - \rho^2\right) \\ \delta_{12}\delta_{22}\sigma_2^2\left(1 - \rho^2\right) & 1 + \delta_{22}^2\sigma_2^2\left(1 - \rho^2\right) \end{pmatrix}\right\}, \end{aligned}$$

and if $D = \text{diag}(\delta_1, \delta_2)$, then $\rho_{D\Sigma}$ is given by (4.2) and the marginal density of X reduces to

$$\begin{aligned} f_X(x, \mu_1, \Sigma, D) &= \frac{\phi(x; \mu_1, \sigma_1^2)}{\frac{1}{2} - \frac{1}{2\pi}\arccos(\rho_{D\Sigma})}\Phi\left\{\left[\begin{pmatrix}\delta_1 \\ 0\end{pmatrix} + \begin{pmatrix}0 \\ \delta_2\end{pmatrix}\frac{\sigma_2}{\sigma_1}\rho\right](x-\mu_1); \begin{pmatrix}1 & 0 \\ 0 & 1+\delta_2^2\sigma_2^2(1-\rho^2)\end{pmatrix}\right\} \\ &= \frac{\phi(x; \mu_1, \sigma_1^2)}{\frac{1}{2} - \frac{1}{2\pi}\arccos(\rho_{D\Sigma})}\Phi(\delta_1(x-\mu_1))\,\Phi\left(\frac{\delta_2\sigma_2\rho(x-\mu_1)}{\sigma_1\sqrt{1+\delta_2^2\sigma_2^2(1-\rho^2)}}\right). \end{aligned} \tag{4.4}$$

Now, the conditional density of Y given $X = x$ is

$$\begin{aligned} f_{Y|X}(y|x; \mu, \Sigma, D) &= \frac{\phi\left[y; \mu_2 + \Sigma_{21}\Sigma_{11}^{-1}(x-\mu_1), \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}\right]\Phi_2\left[D_2(y-\mu_2); -D_1(x-\mu_1)\right]}{\Phi_2\left[\left(D_1 + D_2\Sigma_{21}\Sigma_{11}^{-1}\right)(x-\mu_1)\right]} \\ &= \frac{\phi\left[y; \mu_2 + \frac{\sigma_2}{\sigma_1}\rho(x-\mu_1), \sigma_2^2\left(1-\rho^2\right)\right]}{\Phi_2\left\{\left[\begin{pmatrix}\delta_{11} \\ \delta_{21}\end{pmatrix} + \begin{pmatrix}\delta_{12} \\ \delta_{22}\end{pmatrix}\frac{\sigma_2}{\sigma_1}\rho\right](x-\mu_1); \begin{pmatrix}1+\delta_{12}^2\sigma_2^2\left(1-\rho^2\right) & \delta_{12}\delta_{22}\sigma_2^2\left(1-\rho^2\right) \\ \delta_{12}\delta_{22}\sigma_2^2(1-\rho^2) & 1+\delta_{22}^2\sigma_2^2(1-\rho^2)\end{pmatrix}\right\}} \\ &\quad \times \Phi\left[\delta_{12}(y-\mu_2) + \delta_{11}(x-\mu_1)\right]\Phi\left[\delta_{22}(y-\mu_2) + \delta_{21}(x-\mu_1)\right], \end{aligned}$$

which correspond to the distribution (5.1)

$$GSN_{1,2}\left(\mu_2 + \Sigma_{21}\Sigma_{11}^{-1}(x-\mu_1), \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}, D_2, D_2\mu_2 - D_1(x-\mu_1), I\right).$$

If $D = \text{diag}(\delta_1, \delta_2)$, $\rho_{D\Sigma}$ is given by (4.2) and the density of $Y|X$ reduces to

$$f_{Y|X}(y|x; \mu, \Sigma, D) = \frac{\phi\left[y; \mu_2 + \frac{\sigma_2}{\sigma_1}\rho(x-\mu_1), \sigma_2^2\left(1-\rho^2\right)\right]}{\Phi\left[\frac{\delta_2\sigma_2\rho}{\sigma_1\sqrt{1+\delta_2^2\sigma_2^2(1-\rho^2)}}(x-\mu_1)\right]}\Phi\left[\delta_2(y-\mu_2)\right].$$

The m.g.f. of the conditional distribution of Y given $X = x$ is

$$\begin{aligned} M_{Y|X}(t) &= \frac{\Phi_2\left[D_2\left(\Sigma_{21}\Sigma_{11}^{-1}(x-\mu_1-\Sigma_{12}t) + \Sigma_{22}t\right); -D_1(x-\mu_1), I + D_2\left(\Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}\right)D_2'\right]}{\Phi_2\left[D_2\Sigma_{21}\Sigma_{11}^{-1}(x-\mu_1); -D_1(x-\mu_1); I + D_2\left(\Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}\right)D_2'\right]} \\ &\quad \times e^{\mu_2 t + \frac{\sigma_2}{\sigma_1}\rho(x-\mu_1)t + \frac{1}{2}\sigma_2^2\left(1-\rho^2\right)t^2}. \end{aligned}$$

If $D = \text{diag}(\delta_1, \delta_2)$, then $\rho_{D\Sigma}$ is given by (4.2) and the m.g.f. of $Y|X$ reduces to

$$\begin{aligned} M_{Y|X}(t) &= \frac{\Phi_2\left[\begin{pmatrix}0 \\ \delta_2\end{pmatrix}\frac{\sigma_2}{\sigma_1}\rho(x-\mu_1-\sigma_1\sigma_2\rho t) + \begin{pmatrix}0 \\ \delta_2\end{pmatrix}\sigma_2^2 t; -\begin{pmatrix}\delta_1 \\ 0\end{pmatrix}(x-\mu_1), \begin{pmatrix}1 & 0 \\ 0 & 1+\delta_2^2\sigma_2^2(1-\rho^2)\end{pmatrix}\right]}{\Phi_2\left[\begin{pmatrix}0 \\ \delta_2\end{pmatrix}\frac{\sigma_2}{\sigma_1}\rho(x-\mu_1); -\begin{pmatrix}\delta_1 \\ 0\end{pmatrix}(x-\mu_1); \begin{pmatrix}1 & 0 \\ 0 & 1+\delta_2^2\sigma_2^2(1-\rho^2)\end{pmatrix}\right]} \\ &\quad \times e^{\mu_2 t + \frac{\sigma_2}{\sigma_1}\rho(x-\mu_1)t + \frac{1}{2}\sigma_2^2\left(1-\rho^2\right)t^2} \\ &= \frac{\Phi\left[\frac{\delta_2\frac{\sigma_2}{\sigma_1}\rho(x-\mu_1) + \delta_2\sigma_2^2 t\left(1-\rho^2\right)}{\sqrt{1+\delta_2^2\sigma_2^2(1-\rho^2)}}\right]}{\Phi\left(\frac{\delta_2\sigma_2\rho(x-\mu_1)}{\sigma_1\sqrt{1+\delta_2^2\sigma_2^2(1-\rho^2)}}\right)}e^{\mu_2 t + \frac{\sigma_2}{\sigma_1}\rho(x-\mu_1)t + \frac{1}{2}\sigma_2^2\left(1-\rho^2\right)t^2}. \end{aligned} \tag{4.5}$$

We get the conditional expectation of $Y|X$ by differentiating the m.g.f. (4.5), thus

$$\frac{d}{dt}M_{Y|X}(t) = \left[\mu_2 + \rho\frac{\sigma_2}{\sigma_1}(x-\mu_1) + \left(1-\rho^2\right)\sigma_2^2 t\right] \frac{e^{\mu_2 t + \frac{\sigma_2}{\sigma_1}\rho(x-\mu_1)t + \frac{1}{2}\sigma_2^2\left(1-\rho^2\right)t^2}}{\Phi\left[\frac{\rho\delta_2\sigma_2(x-\mu_1)}{\sigma_1\sqrt{1+\sigma_2^2\delta_2^2(1-\rho^2)}}\right]}$$
$$\times\,\Phi\left[\frac{\rho\delta_2\frac{\sigma_2}{\sigma_1}(x-\mu_1) + \left(1-\rho^2\right)\delta_2\sigma_2^2 t}{\sqrt{1+\sigma_2^2\delta_2^2\left(1-\rho^2\right)}}\right]$$
$$+\,\frac{e^{\mu_2 t + \frac{\sigma_2}{\sigma_1}\rho(x-\mu_1)t + \frac{1}{2}\sigma_2^2\left(1-\rho^2\right)t^2}}{\Phi\left[\frac{\rho\delta_2\sigma_2(x-\mu_1)}{\sigma_1\sqrt{1+\sigma_2^2\delta_2^2(1-\rho^2)}}\right]}\phi\left[\frac{\rho\delta_2\frac{\sigma_2}{\sigma_1}(x-\mu_1) + \left(1-\rho^2\right)\delta_2\sigma_2^2 t}{\sqrt{1+\sigma_2^2\delta_2^2\left(1-\rho^2\right)}}\right]\left[\frac{\left(1-\rho^2\right)\delta_2\sigma_2^2}{\sqrt{1+\sigma_2^2\delta_2^2\left(1-\rho^2\right)}}\right].$$

From the last expression we get that the conditional expectation of Y given $X = x$ is

$$E(Y|x) = \left.\frac{d}{dt}E\left(e^{tY}|X\right)\right|_{t=0}$$
$$= \mu_2 + \rho\frac{\sigma_2}{\sigma_1}(x-\mu_1) + \frac{\left(1-\rho^2\right)\delta_2\sigma_2^2}{\sqrt{1+\left(1-\rho^2\right)\delta_2^2\sigma_2^2}}\frac{\phi\left[\frac{\rho\delta_2\sigma_2(x-\mu_1)}{\sigma_1\sqrt{1+\delta_2^2\sigma_2^2(1-\rho^2)}}\right]}{\Phi\left[\frac{\rho\delta_2\sigma_2(x-\mu_1)}{\sigma_1\sqrt{1+\delta_2^2\sigma_2^2(1-\rho^2)}}\right]}.$$

If $\delta_2 = 0$ then $\rho_{D\Sigma} = 0$, and by using equation (4.4) we get that the marginal densities of X and Y are $SN_1\left(\mu_1, \sigma_1^2, \delta_1\right)$ and $SN_1\left(\mu_2, \sigma_2^2, 0\right) = N\left(\mu_2, \sigma_2^2\right)$, respectively. Thus, when $D = \operatorname{diag}(\delta_1, 0)$ the conditional expectation of Y, given $X = x$, is the usual regression line of Y on X. Note that in this case the BSN density of $(X, Y)'$ reduces to

$$f_2(x, y; \mu, \Sigma, D) = 2\phi_2(x, y; \mu, \Sigma)\,\Phi\left[\delta_1(x-\mu_1)\right],$$

which is the Azzalini-Dalla Valle density with $\alpha = (\delta_1, 0)$.

5 General MSN distribution

In order to have a closed family such that it contains its marginal and conditional densities it is necessary to define a general version of the MSN distribution. We define the *general multivariate skew normal distribution* (GMSN) as a distribution whose density is of the form

$$f_{p,q}(y; \mu, \Sigma, D, \nu, \Delta) = \Phi_q^{-1}\left(D\mu; \nu, \Delta + D\Sigma D'\right)\phi_p(y; \mu, \Sigma)\,\Phi_q(Dy; \nu, \Delta), \tag{5.1}$$

where $\mu \in \mathbb{R}^p$, $\nu \in \mathbb{R}^q$, and $\Sigma\,(p \times p)$ and $\Delta\,(q \times q)$ are two covariance matrices and $D\,(q \times p)$ is an arbitrary matrix. We say that a random variable W has distribution GMSN by writing $W \sim GSN_{p,q}(\mu, \Sigma, D, \nu, \Delta)$.

The fact that (5.1) is a density is readily verified with the help of Lemma 2.2.

The m.g.f.. of this density is given by:

Proposition 5.1. If $Y \sim SN_{p,q}(\mu, \Sigma, D, \nu, \Delta)$ its m.g.f. is given by

$$M_Y(t) = \frac{\Phi_q\left[D(\mu + \Sigma t); \nu, \Delta + D\Sigma D'\right]}{\Phi_q\left(D\mu; \nu; \Delta + D\Sigma D'\right)}e^{\mu' t + \frac{1}{2}t'\Sigma t}. \tag{5.2}$$

6 Concluding remarks

The skew normal distribution is very useful in applied statistics for modelling the skewness, see for example the works of Azzalini and Capitanio (1999) and Genton (2001).

For the bivariate case we saw that the contours of the BSN density are not elliptical as we showed in Section 4. When the contours of a bivariate density are not elliptical the correlation coefficient is not a good measure of association between variables. Bjerve and Doksum (1993) suggest to use a correlation curve which is a natural local measure of the strength of the association between Y and X near $Y = x$. We are working on the computation of the correlation curve for the BSN distribution and on the properties and applications of the general multivariate skew normal distribution see Domínguez-Molina *et al.* (2001).

Appendix

A Computations for the first moment of the MSN distribution

In order to compute the first and second moment for the MSN distribution we consider the derivatives of the m.g.f.. The first derivative of

$$M_Y(t) = Ee^{t'Y} = \frac{\Phi_p(D\Sigma t; I + D\Sigma D')}{\Phi_p(0; I + D\Sigma D')} e^{\mu' t + \frac{1}{2} t' \Sigma t},$$

is given by

$$\frac{\partial}{\partial t} M_Y(t) = \left[(\mu + \Sigma t) \frac{\Phi_p(D\Sigma t; I + D\Sigma D')}{\Phi_p(0; I + D\Sigma D')} + \frac{G^p(t; D\Sigma, I + D\Sigma D')}{\Phi_p(0; I + D\Sigma D')} \right] e^{\mu' t + \frac{1}{2} t' \Sigma t}. \tag{A.1}$$

Then the mean of Y is

$$EY = \left. \frac{\partial}{\partial t} M_Y(t) \right|_{t=0} = \mu + \frac{1}{\Phi_p(0; I + D\Sigma D')} \sum_{i=1}^{p} e_i G_i^p(0; D\Sigma, I + D\Sigma D'),$$

where $e_i\,(p \times 1)$ has one in the i^{th} position and zero elsewhere and

$$G^p(t; D\Sigma, I + D\Sigma D') = \frac{\partial}{\partial t} \Phi_p(D\Sigma t; I + D\Sigma D').$$

$$G^p(0; D\Sigma, I + D\Sigma D') = \left. \frac{\partial}{\partial t} \Phi_p(D\Sigma t; I + D\Sigma D') \right|_{t=0}$$

From Lemma C.2, we get

$$G_i^p(0; D\Sigma, I + D\Sigma D') = \frac{1}{\sqrt{2\pi} \left| I + D\Sigma D' \right|^{1/2}} \sum_{j=1}^{p} (D\Sigma)_{ji} \left| H_{(j)} \right|^{1/2} \Phi_{p-1} \left[0; \left(H_{(j)} \right)^{-1} \right] \tag{A.2}$$

where $H_{(j)}$ is the matrix obtained by eliminating the j^{th} row and the j^{th} column from $(I + D\Sigma D')^{-1}$.

B Computations for the second moment of the MSN distribution

From (A.1) we obtain

$$\frac{\partial^2}{\partial t \partial t'} E e^{t'Y} = \left[\Sigma \frac{\Phi_p\left(D\Sigma t; I + D\Sigma D\right)}{\Phi_p\left(0; I + D\Sigma D\right)} + (\mu + \Sigma t) \frac{G^{p\prime}\left(t; D\Sigma, I + D\Sigma D'\right)}{\Phi_p\left(0; I + D\Sigma D\right)} \right.$$
$$+ \frac{G^p_{[2]}\left(t; D\Sigma, I + D\Sigma D'\right)}{\Phi_p\left(0; I + D\Sigma D\right)} + (\mu + \Sigma t)(\mu + \Sigma t)' \frac{\Phi_p\left(D\Sigma t; I + D\Sigma D\right)}{\Phi_p\left(0; I + D\Sigma D\right)}$$
$$\left. + \frac{G^p\left(t; D\Sigma, I + D\Sigma D'\right)}{\Phi_p\left(0; I + D\Sigma D\right)} (\mu + \Sigma t)' \right] e^{\mu' t + \frac{1}{2} t' \Sigma t}$$

where

$$G^{p\prime}\left(t; D\Sigma, I + D\Sigma D'\right) = \left[G^p\left(t; D\Sigma, I + D\Sigma D'\right)\right]'$$

$$G^p_{[2]}\left(t; D\Sigma, I + D\Sigma D'\right) = \frac{\partial}{\partial t'} G^p\left(t; D\Sigma, I + D\Sigma D'\right) = \frac{\partial^2}{\partial t \partial t'} \Phi_p\left(D\Sigma t; I + D\Sigma D\right).$$

See Lemma C.2 for the full expression of $G^p_{[2]}\left(t; D\Sigma, I + D\Sigma D'\right)$. Thus

$$EYY' = \Sigma + \mu \frac{G^{p\prime}\left(0; D\Sigma, I + D\Sigma D'\right)}{\Phi_p\left(0; I + D\Sigma D\right)} + \frac{G^p_{[2]}\left(0; D\Sigma, I + D\Sigma D'\right)}{\Phi_p\left(0; I + D\Sigma D\right)} + \frac{G^p\left(0; D\Sigma, I + D\Sigma D'\right)}{\Phi_p\left(0; I + D\Sigma D\right)} \mu' + \mu\mu'.$$

C Derivatives of the multinormal integral

C.1 First derivative of the multinormal integral

Lemma C.1. Let $A\,(p \times p)$ be an arbitrary matrix and let Ω be a positive definite matrix. Then

$$G^p\left(0; A, \Omega\right) = \frac{\partial}{\partial t} \Phi_p\left(At; \Omega\right) = \sum_{i=1}^{p} e_i G^p_i\left(0; A, \Omega\right),$$

where $e_i\,(p \times 1)$ has one in the i^{th} position and zero elsewhere and

$$G^p_i\left(0; A, \Omega\right) = \left.\frac{\partial}{\partial t_i} \Phi_p\left(At; \Omega\right)\right|_{t=0} = \frac{1}{\sqrt{2\pi}\left|\Omega\right|^{1/2}} \sum_{j=1}^{p} a_{ji} \left|\Omega_{(j)}\right|^{1/2} \Phi_{p-1}\left[0; \left(\Omega^{-1}_{(j)}\right)^{-1}\right]$$

and where $\Omega^{-1}_{(i)}$ is the matrix constructed by eliminating the i^{th} row and the i^{th} column of Ω^{-1}. With the convention that $\Omega^{-1}_{(i)} = \left(\Omega^{-1}\right)_{(i)}$.

Proof. For $At = \left[(At)_1, ..., (At)_p\right]'$ write

$$\Phi_p\left(At; \Omega\right) = \int_{-\infty}^{(At)_1} \cdots \int_{-\infty}^{(At)_p} \phi_p\left(x_1, ..., x_p; \Omega\right) dx_1 \cdots dx_p.$$

Then the partial derivative with respect to t_i is

$$\frac{\partial}{\partial t_i}\Phi_p\left(At;\Omega\right)=\sum_{k=1}^{p}\int_{-\infty}^{(At)_1}\cdots\int_{-\infty}^{(At)_{k-1}}\int_{-\infty}^{(At)_{k+1}}\cdots\int_{-\infty}^{(At)_p}a_{ki}\phi_p\left[x_1,...,x_{k-1},(At)_k,x_{k+1},...,x_p;\Omega\right] \times dx_1\cdots dx_{k-1}dx_{k+1}\cdots dx_p, \quad \text{(C.1)}$$

whence

$$\begin{aligned}\left.\frac{\partial}{\partial t_i}\Phi_p\left(At;\Omega\right)\right|_{t=0} &= \sum_{k=1}^{p}\int_{-\infty}^{0}\cdots\int_{-\infty}^{0}a_{ki}\phi_p\left(x_1,...,x_{k-1},0,x_{k+1},...,x_p;\Omega\right)dx_1\cdots dx_{k-1}dx_{k+1}\cdots dx_p,\\ &= \frac{1}{\sqrt{2\pi}\left|\Omega\right|^{1/2}}\sum_{k=1}^{p}a_{ki}\left|\left(\Omega_{(k)}^{-1}\right)^{-1}\right|^{1/2}\int_{-\infty}^{0}\cdots\int_{-\infty}^{0}\phi_{p-1}\left[x_1,...,x_{k-1},x_{k+1},...,x_p;\left(\Omega_{(k)}^{-1}\right)^{-1}\right]\\ &\quad\times dx_1\cdots dx_{k-1}dx_{k+1}\cdots dx_p\\ &= \frac{1}{\sqrt{2\pi}\left|\Omega\right|^{1/2}}\sum_{k=1}^{p}a_{ki}\left|\left(\Omega_{(k)}^{-1}\right)^{-1}\right|^{1/2}\Phi_{p-1}\left[0;\left(\Omega_{(k)}^{-1}\right)^{-1}\right].\end{aligned}$$

■

If $\Omega=\mathrm{diag}\left(\omega_1^2,...,\omega_p^2\right)$ the expression for $G_i^p\left(0;A,\Omega\right)$ reduces to

$$G_i^p=\frac{1}{2^{p-1}\sqrt{2\pi}}\sum_{j=1}^{p}\frac{a_{ji}}{\omega_j}. \quad \text{(C.2)}$$

C.2 Second derivative of the multinormal integral

Lemma C.2. Let $A\left(p\times p\right)$ be an arbitrary matrix and let $\Omega\left(p\times p\right)$ be a positive definite matrix. Then

$$\begin{aligned}G_{[2]}^p\left(0;A,\Omega\right) &= \frac{\partial}{\partial t\partial t'}\Phi_p\left(At;\Omega\right)\\ &= \sum_{i=1}^{p}\sum_{j=1}^{p}G_{i,j}^p\left(0;A,\Omega\right)H_{ij},\end{aligned} \quad \text{(C.3)}$$

where the matrix $H_{ij}\left(m\times m\right)$ has unit element at the $\left(i,j\right)^{\text{th}}$ place and zero elsewhere. And

$$\begin{aligned}G_{i,j}^p\left(0;A,\Omega\right) &= \left.\frac{\partial}{\partial t_i\partial t_j}\Phi_p\left(At;\Omega\right)\right|_{t=0} \quad \text{(C.4)}\\ &= \frac{1}{2\pi\left|\Omega\right|^{1/2}}\sum_{k=1}^{p}\sum_{l\neq k}^{p}a_{li}a_{kj}\left|\Omega_{(k,l)}^{-1}\right|\Phi_{p-2}\left(0;\left[\Omega_{(k,l)}^{-1}\right]^{-1}\right)\\ &+\sum_{k=1}^{p}\sum_{\substack{i=1\\ i\neq k}}^{p}\int_{-\infty}^{0}\cdots\int_{-\infty}^{0}\int_{-\infty}^{0}\cdots\int_{-\infty}^{0}a_{ki}a_{kj}\Omega^{ki}x_i\\ &\times\phi_{p-1}\left[x_1,...,x_{k-1},x_{k+1},...,x_p;\left[\Omega_{(k)}^{-1}\right]^{-1}\right]dx_1\cdots dx_{k-1}dx_{k+1}dx_p.\end{aligned}$$

where Ω^{kk} is the element (k,k) of Ω^{-1} and $\Omega^{-1}_{(i,j)}$ is the matrix constructed by eliminating the i^{th} and j^{th} rows and the i^{th} and j^{th} columns of Ω^{-1}, with $\Omega^{-1}_{(i,j)} = \left(\Omega^{-1}\right)_{(i,j)}$.

Proof. From equation (C.1) we have

$$\begin{aligned}\frac{\partial}{\partial t_i \partial t_j}\Phi_p\left(At;\Omega\right) = &\sum_{k=1}^{p}\sum_{l\neq k}^{p}\int_{-\infty}^{(At)_1}\cdots\int_{-\infty}^{(At)_{k-1}}\int_{-\infty}^{(At)_{k+1}}\cdots\int_{-\infty}^{(At)_{l-1}}\int_{-\infty}^{(At)_{l+1}}\cdots\int_{-\infty}^{(At)_p}\\ &a_{li}a_{kj}\phi_p\left[x_1,...,x_{k-1},(At)_k,x_{k+1},...,x_{l-1},(At)_l,x_{l+1}x_p;\Omega\right]\\ &\times dx_1\cdots dx_{k-1}dx_{k+1}\cdots dx_{l-1}dx_{l+1}dx_p\\ &+\sum_{k=1}^{p}\int_{-\infty}^{(At)_1}\cdots\int_{-\infty}^{(At)_{k-1}}\int_{-\infty}^{(At)_{k+1}}\cdots\int_{-\infty}^{(At)_p}a_{ki}a_{kj}q_k\left(t,x,A,\Omega\right)\\ &\times\phi_p\left[x_1,...,x_{k-1},(At)_k,x_{k+1},...,x_p;\Omega\right]\\ &\times dx_1\cdots dx_{k-1}dx_{k+1}dx_p\end{aligned}$$

where

$$q_k\left(t,x,A,\Omega\right) = 2e_k'\Omega^{-1}x^* + 2\left(At\right)_k\Omega^{kk},$$

with $x^* = (x_1,...,x_{k-1},0,x_{k+1},...,x_p)'$ (see C.6 and C.7) and $e_k\ (p\times 1)$ has one in the k^{th} position and zero elsewhere.

Note that

$$q_k\left(0,x,A,\Omega\right) = e_k'\Omega^{-1}x^* = \sum_{\substack{i=1\\ i\neq k}}^{p}\Omega^{ki}x_i,$$

and evaluating $\frac{\partial}{\partial t_i\partial t_j}\Phi_p\left(At;\Omega\right)$ at $t=0$ and after some manipulations we get

$$\begin{aligned}\left.\frac{\partial}{\partial t_i\partial t_j}\Phi_p\left(At;\Omega\right)\right|_{t=0} = &\frac{1}{2\pi\left|\Omega\right|^{1/2}}\sum_{k=1}^{p}\sum_{l\neq k}^{p}a_{li}a_{kj}\left|\Omega^{-1}_{(kl)}\right|\Phi_{p-2}\left(0;\left[\Omega^{-1}_{(k,l)}\right]^{-1}\right)\\ &+\sum_{k=1}^{p}\sum_{\substack{i=1\\ i\neq k}}^{p}\int_{-\infty}^{0}\cdots\int_{-\infty}^{0}\int_{-\infty}^{0}\cdots\int_{-\infty}^{0}a_{ki}a_{kj}\Omega^{ki}x_i\\ &\times\phi_{p-1}\left[x_1,...,x_{k-1},x_{k+1},...,x_p;\left[\Omega^{-1}_{(k)}\right]^{-1}\right]dx_1\cdots dx_{k-1}dx_{k+1}dx_p.\end{aligned}\tag{C.5}$$

In order to obtain

$$q_k\left(t,x,A,\Omega\right) = a_{ki}^{-1}\frac{d}{dt_i}\left[\left(x_1,...,x_{k-1},(At)_k,x_{k+1},...,x_p\right)'\Omega^{-1}\left(x_1,...,x_{k-1},(At)_k,x_{k+1},...,x_p\right)\right],\tag{C.6}$$

first observe that,

$$\begin{aligned}&\left(x_1,...,x_{k-1},(At)_k,x_{k+1},...,x_p\right)'\Omega^{-1}\left(x_1,...,x_{k-1},(At)_k,x_{k+1},...,x_p\right)\\ &=\left[x^*+(At)_k e_k\right]'\Omega^{-1}\left[x^*+(At)_k e_k\right]\\ &=(x^*)'\Omega^{-1}x^*+2(At)_k e_k'\Omega^{-1}x^*+(At)_k^2 e_k'\Omega^{-1}e_k = (x^*)'\Omega^{-1}x^*+2(At)_k e_k'\Omega^{-1}x^*+(At)_k^2\Omega^{kk}\end{aligned}$$

where $x^* = (x_1, ..., x_{k-1}, 0, x_{k+1}, ..., x_p)'$. Given that

$$\frac{d}{dt_i}(At)_k = \frac{d}{dt_i}\sum_{l=1}^{p} a_{kl}t_l = a_{ki},$$

and we get

$$q_k(t, x, A, \Omega) = 2a_{ki}e_k'\Omega^{-1}x^* + 2(At)_k a_{ki}\Omega^{kk} \tag{C.7}$$

where Ω^{kk} is the element (k, k) of Ω^{-1}.

If $A\,\mathrm{diag}(a_1, ..., a_p) = \Omega = \mathrm{diag}(\omega_1^2, ..., \omega_p^2)$ the expression for $G_{i,j}^p(0; A, \Omega)$ reduces to

$$G_{i,j}^p(0; A, \Omega) = \begin{cases} \dfrac{a_i a_j}{2^{p-1}\pi\omega_i\omega_j}, & \text{if } i \neq j, \\ 0 & \text{if } \quad i = j. \end{cases} \tag{C.8}$$

References

[1] Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scand. J. Statist.* **12,** 171-178.

[2] Azzalini, A. (1986). Further results on a class of distributions which includes the normal ones. *Statistica,* **46**, 199-208.

[3] Azzalini, A. and Capitanio, A.(1999). Statistical applications of the multivariate skew normal distribution. *Journal of the Royal Statistical Society,* B, **61**, 579-602.

[4] Azzalini, A. and Dalla-Valle, A. (1996). The multivariate skew-normal distribution. *Biometrika*, **83**, 715-726.

[5] Bjerve, S. and Doksum, K. (1993). Correlation curves: Measures of association as function of covariate values. *Annals of Statistics,* **21,** No. 2, 890-902.

[6] Domínguez-Molina, J. A., González-Farías, G. and Gupta, A.K. (2001) A General Multivariate Skew Normal Distribution. Department of Mathematics and Statistics, Bowling Green State University, Technical Report No. 01-09.

[7] Fang, K.-T., Kotz, S. and Ng, K.-W. (1990). *Symmetric Multivariate and Related Distributions.* London: Chapman and Hall.

[8] 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.

[9] Graham, A. (1981). *Kronecker products and matrix calculus with applications.* Hellis Horwood Limited; Chichester.

[10] Gupta, A.K. and Nagar, D.K. (2000). *Matrix variate distributions.* Chapman And Hall/CRC, Boca Raton.

[11] Gupta, A.K. and Varga, T. (1993). *Elliptically Contoured Models in Statistics* (Mathematics and Its Applications, Vol 240). Kluwer Academic Publishers Group, Dordrecht.

[12] Kelker, D. (1970). Distribution theory of spherical distributions and location-scale parameter generalization. *Sankhȳa,* **32**, 419-430.

[13] Marsaglia, G. (1963). Expressing the normal distribution with covariance matrix $A+B$ in terms of one with covariance matrix A. *Biometrika* 50, 535-538.

[14] Zacks, S. (1981). *Parametric Statistical Inference.* Oxford, Pergamon.