SOME PROPERTIES OF THE MULTIVARIATE SPLIT NORMAL DISTRIBUTION

ROLF LARSSON AND MATTIAS VILLANI

Abstract. The multivariate split normal distribution extends the usual multivariate normal distribution by a set of parameters which allows for skewness in the form of contraction/dilation along a subset of the principal axes. This note derives some properties for this distribution, including its moment generating function, multivariate skewness and multivariate kurtosis.

1. Introduction

The univariate split normal distribution, or the two-piece normal, extends the symmetric normal distribution with an additional parameter to model skewness. This distribution was originally introduced by Gibbons and Mylroie (1973), with most of its known properties derived by John (1982); see also Kimber (1985) for some additional results. Johnson, Kotz and Balakrishnan (1994) contains references to papers where the split normal distribution is used as a statistical model. The split normal distribution has also been recognized as a convenient vehicle for elicitation of subjective beliefs, see *e.g.* Blix and Sellin (1998) and Kadane, Chan and Wolfson (1996), which in turn have motivated extensions to the multivariate case; see *e.g.* the bivariate translation approach in Blix and Sellin (2000) and the discussion of Kadane *et al.* (1996) in Bauwens, Polasek and van Dijk (1996).

In an influential paper on Monte Carlo integration, Geweke (1989), apparently unaware of earlier work in this area, suggested a multivariate generalization of the split normal distribution to be used in the construction of an importance function. The density was only given up to a constant and no properties of this distribution were presented. In this note, some properties of the multivariate split normal distribution will be derived using a suitable reparametrization of Geweke's distribution.

The paper is outlined as follows. The next section gives a short review of the univariate split normal distribution. Section 3 describes the multivariate split normal distribution and the fourth section derives some of its properties. The proofs have been collected in an appendix.

Key words and phrases. Multivariate analysis; Elicitation of distributions; Multivariate skewness; Moment generating function: Multivariate kurtosis.

The second author gratefully acknowledges financial support from the Swedish Council of Research in Humanities and Social Sciences (HSFR), grant no. F0582/1999 and the Swedish Research Council (Vetenskapsrådet) grant no. 412-2002-1007.

2. The univariate split normal distribution

The following definition is a reparametrization of the univariate split normal distribution in John (1982).

Definition 1. $x \in R$ follows the univariate split normal distribution, $x \sim SN(\mu, \lambda^2, \tau^2)$, if it has density

$$f(x) = \begin{cases} c \cdot \exp\left[-\frac{1}{2\lambda^2}(x-\mu)^2\right] & \text{if } x \leq \mu \\ c \cdot \exp\left[-\frac{1}{2\tau^2\lambda^2}(x-\mu)^2\right] & \text{if } x > \mu, \end{cases}$$

where $c = \sqrt{2/\pi}\lambda^{-1}(1+\tau)^{-1}$.

The density of the $SN(\mu, \lambda^2, \tau^2)$-distribution is thus proportional to the density of the $N(\mu, \lambda^2)$-distribution to the left of the mode, μ, whereas to the right of the mode it is proportional to the density of the $N(\mu, \tau^2\lambda^2)$-distribution. For $\tau < 1$ the distribution is skewed to the left, for $\tau > 1$ it is skewed to the right and for $\tau = 1$ it reduces to the usual symmetric normal distribution.

John (1982) derived several properties of the univariate split normal distribution. The following result will be useful in the sequel.

Lemma 1. If $X \sim SN(\mu, \lambda^2, \tau^2)$, then

$$\begin{aligned} E(X) &= \mu + \sqrt{2/\pi}\lambda(\tau - 1) \\ Var(X) &= b\lambda^2 \end{aligned}$$

where $b = \frac{\pi-2}{\pi}(\tau-1)^2 + \tau$.

The next lemma gives the univariate skewness

$$\beta_1 = \frac{E[x - E(x)]^3}{[Var(x)]^{3/2}}$$

and univariate kurtosis

$$\beta_2 = \frac{E[x - E(x)]^4}{[Var(x)]^2}$$

of a $SN(\mu, \lambda^2, \tau^2)$ variable.

Lemma 2. If $x \sim SN(\mu, \lambda^2, \tau^2)$ then

$$\beta_1 = b^{-3/2}\sqrt{2/\pi}(\tau-1)[(4/\pi - 1)(\tau-1)^2 + \tau]$$

and

$$\beta_2 = b^{-2}q,$$

where $q = 3(1+\tau^5)/(1+\tau) - 4\pi^{-2}(1-\tau)^2\left[(3+\pi)(1+\tau^2) + 3(\pi-2)\tau\right]$.

The next lemma gives the moment generating function $\phi_x(t) = E(e^{tx})$ of a univariate split normal variable as derived by John (1982).

Lemma 3. If $x \sim SN(\mu, \lambda^2, \tau^2)$, then

$$\phi_x(t) = \frac{2\lambda\left\{\exp(-\lambda^2 t^2/2)\Phi(-\lambda t) + \tau\exp(-\lambda^2\tau^2 t^2/2)\Phi(-\lambda\tau t)\right\}}{\lambda(1+\tau)\exp(\mu t)}.$$

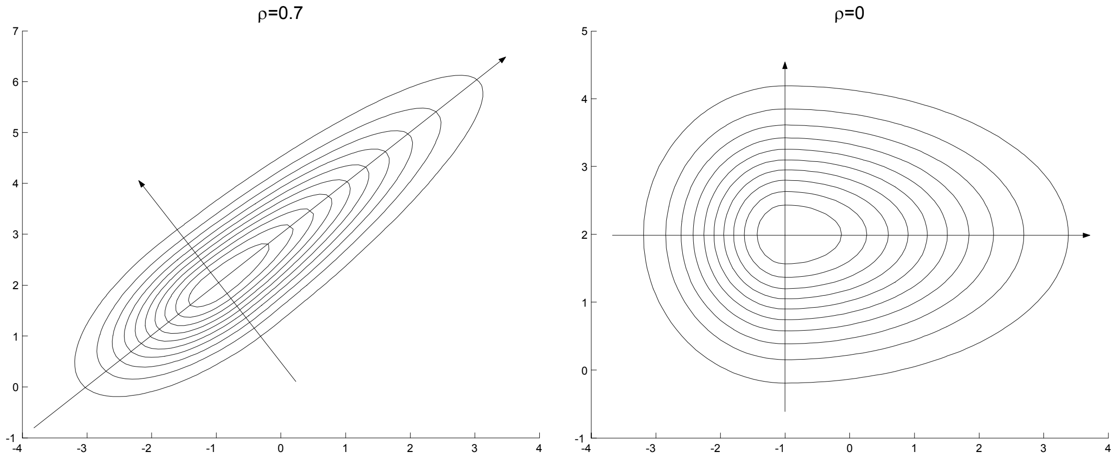

FIGURE 1. Contour plots of bivariate 1-split normal density functions. $\mu = (-1, 2)$, $\Sigma = (1, \rho; \rho, 1)$, $\mathcal{Q} = 1$ and $\tau = 2$.

3. The multivariate split normal distribution

The following definition is a natural generalization of the univariate split normal distribution in John (1982) to the multivariate setting and is a reparametrization of the multivariate split normal distribution in Geweke (1989).

Definition 2. A vector $x \in R^p$ follows the q-split normal distribution, $x \sim SN_p(\mu, \Sigma, \tau, \mathcal{Q})$, if its principal components are independently distributed as

$$v_i' x \sim \begin{cases} SN(v_i'\mu, \lambda_i^2, \tau_i^2) & \text{if } i \in \mathcal{Q} \\ N(v_i'\mu, \lambda_i^2) & \text{if } i \in \mathcal{Q}^c, \end{cases}$$

where $\mathcal{Q} \subseteq \{1, ..., p\}$ of size q, $\mathcal{Q}^c = \{1, 2, ..., p\} \backslash \mathcal{Q}$ is the complement of $\mathcal{Q}$, v_i is the eigenvector corresponding to the ith largest eigenvalue in the spectral decomposition of $\Sigma = V \Lambda V'$, $\Lambda = \text{diag}(\lambda_1^2, ..., \lambda_p^2)$ and $\tau = (\tau_i)_{i \in \mathcal{Q}}$ is q-dimensional vector of contraction/dilation parameters.

Definition 2 coheres with the idea that multivariate data are driven by a small number of underlying variables (here represented by the principal components) and it is these underlying variables which are allowed to have skewed distributions. Let us consider the case $\mathcal{Q} = \{r\}$ for illustration, i.e. when only the rth principal component has a skewed distribution. It is then easy to see that the density of x is

$$f(x) = \begin{cases} c \cdot \exp\left\{-\frac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu)\right\} & \text{if } \ v_r'(x-\mu) \leq 0 \\ c \cdot \exp\left\{-\frac{1}{2}(x-\mu)'\bar{\Sigma}^{-1}(x-\mu)\right\} & \text{if } \ v_r'(x-\mu) > 0, \end{cases}$$

where $\bar{\Sigma} = V \bar{\Lambda} V'$, $\bar{\Lambda} = diag(\lambda_1^2, ..., \tau_1^2 \lambda_r^2, ..., \lambda_p^2)$ and $c^{-1} = \frac{1}{2}(2\pi)^{p/2} |\Lambda|^{1/2} (1 + \tau_1)$. This should be compared to the univariate case in Definition 1. Figure 1 illustrates two possible shapes of the $SN_2(\mu, \Sigma, \tau, \mathcal{Q})$-distribution.

The general $SN_p(\mu, \Sigma, \tau, \mathcal{Q})$-distribution amounts to using different multivariate normal distributions, all with mode μ, over 2^q regions of R^p separated by the q hyperplanes $v_i'(x - \mu) = 0$, for $i \in \mathcal{Q}$. Other forms of the separating hyperplanes, or more general changes in covariance structure between the 2^q regions, produces ill-behaved densities with sharp ridges.

4. Some properties of the $SN_p(\mu, \Sigma, \tau, \mathcal{Q})$ distribution

In the following theorems, let $b_i = \frac{\pi-2}{\pi}(\tau_i - 1)^2 + \tau_i$ for $i \in \mathcal{Q}$. Our first result generalizes Lemma 1 to the multivariate setting.

Theorem 4. If $X \sim SN_p(\mu, \Sigma, \tau, \mathcal{Q})$, then

$$\begin{aligned} E(X) &= \mu + \sqrt{2/\pi} \sum_{\mathcal{Q}} \lambda_i(\tau_i - 1)v_i \\ Var(X) &= V\Lambda_{\mathcal{Q}}V' \end{aligned}$$

where $\Lambda_{\mathcal{Q}}$ is a diagonal matrix with ith element equal to λ_i^2 if $i \in \mathcal{Q}^c$ or $b_i\lambda_i^2$ if $i \in \mathcal{Q}$.

Let

$$M_{xz} = (x - m)'S^{-1}(z - m),$$

be the Mahalanobis distance between two p-dimensional independent identically distributed random vectors x and z, where m and S are the common mean and covariance matrix, respectively. Mardia (1970) used M_{xz} to define a widely used measure of *multivariate skewness*

$$\beta_{1,p} = E(M_{xz}^3).$$

Note that if $x \sim N_p(\mu, \Sigma)$, then $\beta_{1,p} = 0$. $\beta_{1,p}$ is related to the univariate skewness through the equality $\beta_{1,1} = \beta_1^2$. We have the following result.

Theorem 5. If $x \sim SN(\mu, \Sigma, \tau, \mathcal{Q})$ then

$$\beta_{1,p} = \sum_{\mathcal{Q}} b_i^{-3}(2/\pi)(\tau_i - 1)^2[(4/\pi - 1)(\tau_i - 1)^2 + \tau_i]^2.$$

The Mahalanobis distance may also be used to define *multivariate kurtosis* (Mardia, 1970)

$$\beta_{2,p} = E(M_{xx}^2).$$

If $x \sim N_p(\mu, \Sigma)$, then $\beta_{2,p} = p(p+2)$. Note also that $\beta_{2,1} = \beta_2$. The next theorem gives $\beta_{2,p}$ for the multivariate split normal distribution.

Theorem 6. If $x \sim SN(\mu, \Sigma, \tau, \mathcal{Q})$ then

$$\beta_{2,p} = p(p+2) + \sum_{\mathcal{Q}} b_i^{-2}q_i - 3q,$$

where $q_i = 3(1+\tau_i^5)/(1+\tau_i) - 4\pi^{-2}(1-\tau_i)^2\left[(3+\pi)(1+\tau_i^2) + 3(\pi-2)\tau_i\right]$.

The moment generating function $\phi_x(t) = E[\exp(t'x)]$ of a $SN_p(\mu, \Sigma, \tau, \mathcal{Q})$ variable is given in the next result.

Theorem 7. If $x \sim SN_p(\mu, \Sigma, \tau, \mathcal{Q})$, then

$$\begin{aligned} \phi_x(t) &= \left[\prod_{\mathcal{Q}} \frac{2\lambda_i\left\{\exp[-(\lambda_i v_i't)^2/2]\Phi(-\lambda_i v_i't) + \tau_i \exp[-(\lambda_i\tau_i v_i't)^2/2]\Phi(-\lambda_i\tau_i v_i't)\right\}}{\lambda_i(1+\tau_i)\exp(\mu_i v_i't)}\right] \\ &\quad \times \exp\left\{\sum_{\mathcal{Q}^c}[\mu_i v_i't - \frac{1}{2}(v_i't)^2\lambda_i^2]\right\}. \end{aligned}$$

References

[1] Bauwens, L., Polasek, W. and van Dijk, H. K. (1996). Editors' introduction, *Journal of Econometrics*, **75**, 1-5.

[2] Blix, M. and Sellin, P. (1998). Uncertainty bands for inflation forecasts, Sveriges Riksbank Working Paper Series No. **65**.

[3] Blix, M. and Sellin, P. (2000). A bivariate distribution for inflation and output forecasts, Sveriges Riksbank Working Paper Series No. **102**.

[4] Geweke, J. (1989). Bayesian inference in econometric models using Monte Carlo integration, *Econometrica*, **57** (6), 1317-1339.

[5] Gibbons, J. F. and Mylroie, S. (1973). Estimation of impurity profiles in ion-implanted amorphous targets using joined half-Gaussian distributions, *Applied Physics Letters*, **22**, 568-569.

[6] John, S. (1982). The two-piece normal family of distributions and its fitting, *Communications in Statistics - Theory and Methods*, **11** (8), 879-885.

[7] Johnson, N. L., Kotz, S. and Balakrishnan, N. (1994). *Continuous Univariate Distributions*, Vol. I.

[8] Kadane, J. B., Chan, N. H. and Wolfson, L. J. (1996). Priors for unit root models, *Journal of Econometrics*, **75**, 99-111.

[9] Kimber, A. C. (1985). Methods for the two-piece normal distribution, *Communications in Statistics - Theory and Methods*, **14** (1), 235-245.

[10] Mardia, K. V. (1970). Measures of multivariate skewness and kurtosis with applications, *Biometrika*, **57**, 519-530.

5. Appendix

5.1. **Proof of Lemma 2.**

From John (1982) we have

$$E[x-E(x)]^3 = \sqrt{2/\pi}\lambda(\tau-1)[(4/\pi-1)(\lambda(\tau-1))^2+\lambda^2\tau].$$

Make the transformation $y=(x-\mu)/\lambda$. It is easy to see that $y \sim SN(0,1,\tau^2)$ with $Var(y) = \lambda^{-2}Var(x) = b$. Since skewness is invariant to linear transformations we have

$$\beta_1(x) = \beta_1(y) = \frac{E[y-E(y)]^3}{[Var(y)]^{3/2}} = b^{-3/2}\sqrt{2/\pi}(\tau-1)[(4/\pi-1)(\tau-1)^2+\tau].$$

Similarly, since kurtosis is invariant to linear transformations

$$\beta_2(x) = \beta_2(y) = \frac{E[y-E(y)]^4}{[Var(y)]^2} = b^{-2}E[y-E(y)]^4,$$

where

$$E[y-E(y)]^4 = E(y^4) - 4E(y^3)E(y) + 6E(y^2)[E(y)]^2 - 3[E(y)]^4. \tag{5.1}$$

and

$$\begin{aligned} E(y) &= \sqrt{2/\pi}(\tau-1) \\ E(y^2) &= (1-\tau)^2+\tau \\ E(y^3) &= 2\sqrt{2/\pi}(\tau^4-1)/(1+\tau) \\ E(y^4) &= 3(1+\tau^5)/(1+\tau). \end{aligned}$$

Inserting these moments into (5.1) and simplifying yields

$$E[y-E(y)]^4 = 3(1+\tau^5)/(1+\tau) - 4\pi^{-2}(1-\tau)^2[(3+\pi)(1+\tau^2)+3(\pi-2)\tau)],$$

which proves the result.

5.2. **Proof of Theorem 4.** Since $x = Vy$, where y is the vector of principal components, we have

$$\begin{aligned} E(x) &= VE(y) = \sum_{\mathcal{Q}} v_i E(y_i) + \sum_{\mathcal{Q}^c} v_i E(y_i) = \sum_{\mathcal{Q}} v_i [v_i'\mu + \sqrt{2/\pi}\lambda_i(\tau_i - 1)] + \sum_{\mathcal{Q}^c} v_i v_i'\mu \\ &= \mu + \sqrt{2/\pi} \sum_{\mathcal{Q}} \lambda_i(\tau_i - 1) v_i, \end{aligned}$$

by Lemma 1.

The covariance matrix can be written

$$Var(x) = V \cdot Var(y) \cdot V' = \sum_{i=1}^{p} Var(y_i) v_i v_i' = \sum_{\mathcal{Q}} b_i \lambda_i^2 v_i v_i' + \sum_{\mathcal{Q}^c} \lambda_i^2 v_i v_i' = V \Lambda_{\mathcal{Q}} V',$$

again using Lemma 1.

5.3. **Proof of Theorem 5.** Since $x = Vy$, where y are the principal components of x,

$$\beta_{1,p}(x) = \beta_{1,p}(Vy) = \beta_{1,p}(y),$$

by the invariance of $\beta_{1,p}$ under linear transformations (Mardia, 1970). Let v and w be independent random vectors from the same distribution of y, $m = (m_1, ..., m_p)' = E(y)$ and $Var(y) = \Lambda_{\mathcal{Q}} = Diag(\sigma_1^2, ..., \sigma_p^2)$, where $\sigma_i^2 = \lambda_i^2$ if $i \in \mathcal{Q}^c$ and $\sigma_i^2 = b_i \lambda_i^2$ if $i \in \mathcal{Q}$. By definition, $\beta_{1,p}(y) = E(M_{vw}^3)$, where M_{vw} may be decomposed as

$$M_{vw} = (v - m)' \Lambda_{\mathcal{Q}}^{-1} (w - m) = \Sigma_{i=1}^{p} \sigma_i^{-2} (v_i - m_i)(w_i - m_i) = \Sigma_{i=1}^{p} M_{v_i w_i},$$

and therefore

$$M_{vw}^3 = \Sigma_{r_1 + ... + r_p = 3} \frac{3!}{r_1! \cdots r_p!} M_{v_1 w_1}^{r_1} \cdots M_{v_p w_p}^{r_p}.$$

Since $E(M_{v_i w_i}) = 0$ for $i = 1, ..., p$, by the independence of the elements of v and w, we have

$$E(M_{vw}^3) = \Sigma_{i=1}^{p} E(M_{v_i w_i}^3),$$

which proves that

$$\beta_{1,p}(y) = \Sigma_{i=1}^{p} \beta_{1,1}(y_i).$$

Since $\beta_{1,1} = \beta_1^2$ and $\beta_i = 0$ for $i \in \mathcal{Q}^c$, the result now follows from Lemma 2.

5.4. **Proof of Theorem 6.** Since $x = Vy$, where y are the principal components of x,

$$\beta_{2,p}(x) = \beta_{2,p}(Vy) = \beta_{2,p}(y),$$

by the invariance of $\beta_{2,p}$ under linear transformations (Mardia, 1970). Now, by the diagonality of $Var(y) = \Lambda_{\mathcal{Q}}$,

$$M_{yy}^2 = \Sigma_{i=1}^{p} M_{y_i y_i}^2 + 2 \sum_{i<j} M_{y_i y_i} M_{y_j y_j}.$$

Thus,

$$\begin{aligned} E(M_{yy}^2) &= \Sigma_{i=1}^{p} E(M_{y_i y_i}^2) + 2 \sum_{i<j} E(M_{y_i y_i}) E(M_{y_j y_j}) \\ &= \Sigma_{i=1}^{p} E(M_{y_i y_i}^2) + p(p-1). \end{aligned}$$

since $E(M_{y_i y_i}) = 1$ for $i = 1, ..., p$. Thus

$$\beta_{2,p}(y) = \Sigma_{i=1}^{p} \beta_{2,p}(y_i) + p(p-1) = \sum_{\mathcal{Q}} \beta_2(y_i) - 3q + p(p+2).$$

since $\beta_{2,p}(y_i) = \beta_2(y_i)$ and $\beta_2(y_i) = 3$ for $i \in \mathcal{Q}^c$. The result now follows from Lemma 2.

5.5. **Proof of Theorem 7.** Since $x = Vy$,

$$\begin{aligned} \phi_x(t) &= E[\exp(t'x)] = E[\exp(t'Vy)] = \prod_{\mathcal{Q}^c} E[\exp(t'v_i y_i)] \prod_{\mathcal{Q}} E[\exp(t'v_i y_i)] \\ &= \prod_{\mathcal{Q}^c} \phi_{y_i}(v_i' t) \prod_{\mathcal{Q}} \phi_{y_i}(v_i' t) = \exp\left\{ \sum_{\mathcal{Q}^c} [\mu_i v_i' t - \frac{1}{2}(v_i' t)^2 \lambda_i^2] \right\} \prod_{\mathcal{Q}} \phi_{y_i}(v_i' t), \end{aligned}$$

where, using Lemma 3,

$$\phi_{y_i}(v_i' t) = \frac{2\lambda_i \left\{ \exp[-(\lambda_i v_i' t)^2/2] \Phi(-\lambda_i v_i' t) + \tau_i \exp[-(\lambda_i \tau_i v_i' t)^2/2] \Phi(-\lambda_i \tau_i v_i' t) \right\}}{\lambda_i (1+\tau_i) \exp(\mu_i v_i' t)} \quad \text{for } i \in \mathcal{Q}.$$

Department of Information Science, Uppsala University, Box 513, SE-751 20 Uppsala, Sweden.
E-mail address: rolf.larsson@dis.uu.se

Department of Statistics, Stockholm University, SE-106 91 Stockholm, Sweden
E-mail address: mattias.villani@stat.su.se