5

The multivariate skew-normal distribution

5.1 Introduction

5.1.1 Definition and basic properties

A quite natural and simple extension of the skew-normal density (2.1) to the d-dimensional case, still of type (1.2), is given by

$$\varphi_d(x;\bar\Omega,\alpha) = 2\,\varphi_d(x;\bar\Omega)\,\Phi(\alpha^\top x), \qquad x\in\mathbb{R}^d, \tag{5.1}$$

where $\bar\Omega$ is a positive-definite $d\times d$ correlation matrix, $\varphi_d(x;\Sigma)$ denotes the density function of a $\mathrm{N}_d(0,\Sigma)$ variate and α is the d-dimensional vector parameter.

There are many other types of multivariate skew-normal distribution we might consider, some of which will indeed be examined later in this book. As already said, (5.1) represents what arguably is the simplest option involving a modulation factor of Gaussian type operating on a multivariate normal base density.

We shall refer to a variable Z with density (5.1) as a 'normalized' *multivariate skew-normal* variate. For applied work, we need to introduce location and scale parameters via the transformation

$$Y = \xi + \omega\, Z, \tag{5.2}$$

where $\xi\in\mathbb{R}^d$ and $\omega = \mathrm{diag}(\omega_1,\ldots,\omega_d) > 0$, leading to the general form of multivariate SN variables. It is immediate that the density function of Y at $x\in\mathbb{R}^d$ is

$$\det(\omega)^{-1}\,\varphi_d(z;\bar\Omega,\alpha) \;=\; 2\,\varphi_d(x-\xi;\Omega)\,\Phi(\alpha^\top\omega^{-1}(x-\xi)), \tag{5.3}$$

where $z = \omega^{-1}(x-\xi)$ and $\Omega = \omega\bar\Omega\omega$. We write $Y\sim \mathrm{SN}_d(\xi,\Omega,\alpha)$ and the parameter components will be called location, scale matrix and slant, respectively. When this notation is used, we shall be implicitly assuming that $\Omega > 0$. Note that ω can be written as

$$\omega = (\Omega\odot I_d)^{1/2}$$

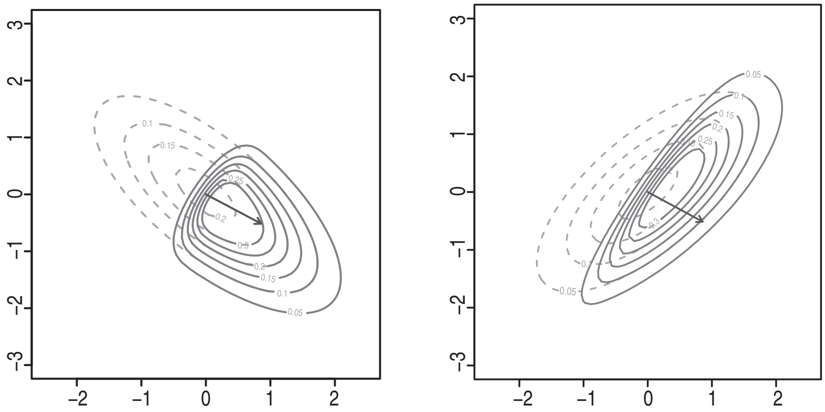


Figure 5.1 Contour plot of two bivariate skew-normal density functions when $\xi = (0, 0)$, $\alpha = (5, -3)^\top$, $\Omega_{11} = 1$, $\Omega_{22} = 1$ and $\Omega_{12} = -0.7$ in the left-side panel, and $\Omega_{12} = 0.7$ in the right-side panel. In each panel the dashed grey line represents the contour plot of the corresponding modulated bivariate normal distribution, and the arrow represents the vector α divided by its Euclidean norm.

where $\odot$ denotes the entry-wise or Hadamard product. We shall use this type of notation repeatedly in the following.

The shape of the multivariate SN density depends on the combined effect of Ω and α. For the bivariate case, a graphical illustration of the interplay between these components is provided in Figure 5.1, which shows the contour plots of two densities having the same parameter set except Ω_{12}. Here the corresponding base densities f_0, displayed by dashed grey lines, are formed by the reflection of each other with respect to the vertical axis, but the modulation effect produced by the same α leads to quite different densities.

Many properties of the univariate skew-normal distribution extend directly to the multivariate case. These are the simplest ones:

$$\varphi_d(x; \Omega, 0) = \varphi_d(x; \Omega), \quad \text{for all } x, \tag{5.4}$$

$$\varphi_d(0; \Omega, \alpha) = \varphi_d(0; \Omega)\,, \tag{5.5}$$

$$-Z \sim \mathrm{SN}_d(0, \bar\Omega, -\alpha)\,, \tag{5.6}$$

$$(Y-\xi)^\top\Omega^{-1}(Y-\xi) = Z^\top\,\bar\Omega^{-1}Z \sim \chi^2_d\,, \quad \text{for all } \alpha, \tag{5.7}$$

where Z has distribution (5.1) and Y is given by (5.2).

Proposition 5.1 *The* $\mathrm{SN}_d(\xi, \Omega, \alpha)$ *density (5.3) is log-concave, i.e., its logarithm is a concave function of x, for any choice of the parameters.*

The proof is a simple extension of Proposition 2.6 for the univariate case; see Problem 5.1. From this we conclude that the regions delimited by contour lines of the density are convex sets, and, of course, the mode is unique.

Before entering more technical aspects, some remarks on the choice of parameterization are appropriate. For algebraic simplicity, one might think of replacing $\omega^{-1}\alpha$ in the final factor of (5.3) by a single term η, say, and view the distribution as a function of (ξ, Ω, η). While use of η does simplify several expressions, and we shall make use of it at places, its adoption for parameterizing the family is questionable, since η reflects both the shape and the scale of the distribution. This choice would be similar to expressing the linear dependence between two variables via their covariance instead of their correlation.

Another notation in use replaces ω^{-1} in (5.3) by $\Omega^{-1/2}$. In this case the problem is that there are many possible options for the square root of Ω, leading to actually different densities, and there is no decisive reason for choosing one specific alternative.

5.1.2 Moment generating function

The following lemma is an immediate extension of Lemma 2.2; the proof follows by simply noticing that $h^\top U \sim \mathrm{N}(0, h^\top \Sigma h)$ if $U \sim \mathrm{N}_d(0, \Sigma)$. The subsequent statement illustrates the technique of 'completing the square' for a skew-normal type of integrand.

Lemma 5.2 *If* $U \sim \mathrm{N}_d(0, \Sigma)$ *then*

$$\mathbb{E}\{\Phi(h^\top U + k)\} = \Phi\left(\frac{k}{\sqrt{1 + h^\top \Sigma h}}\right), \qquad h \in \mathbb{R}^d,\ k \in \mathbb{R}. \tag{5.8}$$

Lemma 5.3 *If A is a symmetric positive definite d × d matrix, a and c are d-vectors and* c_0 *is a scalar, then*

$$\begin{aligned} I &= \int_{\mathbb{R}^d} \frac{1}{(2\pi)^{d/2}\ \det(A)^{1/2}} \exp\left\{-\tfrac{1}{2}(x^\top A^{-1} x - 2a^\top x)\right\} \Phi(c_0 + c^\top x)\,\mathrm{d}x \\ &= \exp\left(\tfrac{1}{2}a^\top A a\right) \Phi\left(\frac{c_0 + c^\top A a}{\sqrt{1 + c^\top A c}}\right). \end{aligned} \tag{5.9}$$

Proof In the integrand of I rewrite $x^\top A^{-1}\, x - 2a^\top x$ as $(x-\mu)^\top A^{-1}(x-\mu) - \mu^\top A^{-1}\mu$ where $\mu = Aa$, so that

$$I = \exp(\tfrac{1}{2}a^\top A a) \int_{\mathbb{R}^d} \varphi(y; A)\, \Phi\{c_0 + c^\top (y + \mu)\}\,\mathrm{d}y$$

https://doi.org/10.1017/CBO9781139248891.006 Published online by Cambridge University Press

after a change of variable. Use of Lemma 5.2 gives (5.9). QED

To compute the moment generating function $M(t)$ of $Y \sim \mathrm{SN}_d(\xi, \Omega, \alpha)$, write $Y = \xi + \omega Z$, where $Z \sim \mathrm{SN}_d(0, \bar{\Omega}, \alpha)$. Then, using Lemma 5.3, we obtain

$$\begin{aligned} M(t) &= \exp(t^\top \xi) \int_{\mathbb{R}^d} 2 \exp(t^\top \omega z)\, \varphi_d(z; \bar{\Omega})\, \Phi(\alpha^\top z)\, \mathrm{d}z \\ &= 2 \exp(t^\top \xi + \tfrac{1}{2} t^\top \Omega t)\, \Phi(\delta^\top \omega\, t), \qquad t \in \mathbb{R}^d, \end{aligned} \tag{5.10}$$

where

$$\delta = \left(1 + \alpha^\top \bar{\Omega} \alpha\right)^{-1/2} \bar{\Omega} \alpha\,. \tag{5.11}$$

For later use, we write down the inverse relationship:

$$\alpha = \left(1 - \delta^\top \bar{\Omega}^{-1} \delta\right)^{-1/2} \bar{\Omega}^{-1} \delta. \tag{5.12}$$

A simple corollary obtained using the above expression of $M(t)$ is the next statement, which is the multivariate extension of Proposition 2.3.

Proposition 5.4 *If $Y_1 \sim \mathrm{SN}_d(\xi, \Omega, \alpha)$ and $Y_2 \sim \mathrm{N}_d(\mu, \Sigma)$ are independent variables, then*

$$X = Y_1 + Y_2 \sim \mathrm{SN}_d(\xi + \mu, \Omega_X, \tilde{\alpha}),$$

where

$$\Omega_X = \Omega + \Sigma, \qquad \tilde{\alpha} = \left(1 + \eta^\top \Omega_X^{-1} \eta\right)^{-1/2} \omega_X \Omega_X^{-1} \Omega \eta,$$

having set $\eta = \omega^{-1} \alpha$ and $\omega_X = (\Omega_X \odot I_d)^{1/2}$.

Similarly to the univariate case, it can be shown that, when both summands Y_1 and Y_2 are 'proper' independent SN variates, that is with non-null slant, their sum is not SN. The proof is, in essence, the same as Proposition 5.5, given later. The only formal difference is in the leading sign of quadratic forms appearing in the $\exp(\cdot)$ terms, but this does not affect the argument.

5.1.3 Stochastic representations

Conditioning and selective sampling

Specification of (1.9)–(1.11) to the present context says that, if $X_0 \sim \mathrm{N}_d(0, \bar{\Omega})$ and $T \sim N(0, 1)$ are independent variables, then both

$$Z' = (X_0 | T > \alpha^\top X_0), \qquad Z = \begin{cases} X_0 & \text{if } T > \alpha^\top X_0, \\ -X_0 & \text{otherwise} \end{cases} \tag{5.13}$$

https://doi.org/10.1017/CBO9781139248891.006 Published online by Cambridge University Press

have distribution $\mathrm{SN}_d(0, \bar{\Omega}, \alpha)$.

This scheme can be rephrased in an equivalent form which allows an interesting interpretation. Define

$$X_1 = \left(1 + \alpha^\top \bar{\Omega} \alpha\right)^{-1/2} (\alpha^\top X_0 - T)$$

such that

$$X = \begin{pmatrix} X_0 \\ X_1 \end{pmatrix} \sim \mathrm{N}_{d+1}\left(0, \Omega^*\right), \qquad \Omega^* = \begin{pmatrix} \bar{\Omega} & \delta \\ \delta^\top & 1 \end{pmatrix}, \tag{5.14}$$

where δ is given by (5.11) and Ω^* is a full-rank correlation matrix. Then the variables in (5.13) can be written as

$$Z' = (X_0 | X_1 > 0), \qquad Z = \begin{cases} X_0 & \text{if } X_1 > 0, \\ -X_0 & \text{otherwise.} \end{cases} \tag{5.15}$$

For random number generation the second form is preferable, as discussed in Complement 1.1. However, the first form of (5.15) is interesting for its qualitative interpretation, since it indicates a link between the skew-normal distribution and a censoring mechanism, fairly common in an applied context, especially in the social sciences, where a variable X_0 is observed only when a correlated variable, X_1, which is usually unobserved, fulfils a certain condition. This situation is commonly referred to as selective sampling.

Another use of the first form of (5.15) allows us to express the distribution function of a multivariate SN variable, but we defer this point to the slightly more general case of § 5.3.3.

Additive representation

Consider a $(d+1)$-dimensional normal variable U which is partitioned into components U_0 and U_1 of dimensions d and 1, respectively, such that the joint distribution is

$$U = \begin{pmatrix} U_0 \\ U_1 \end{pmatrix} \sim \mathrm{N}_{d+1}\left(0, \begin{pmatrix} \bar{\Psi} & 0 \\ 0 & 1 \end{pmatrix}\right), \tag{5.16}$$

where $\bar{\Psi}$ is a full-rank correlation matrix. Given a vector $\delta = (\delta_1, \ldots, \delta_d)^\top$ with all elements in $(-1, 1)$, define similarly to (2.14)

$$Z_j = \left(1 - \delta_j^2\right)^{1/2} U_{0j} + \delta_j |U_1|, \tag{5.17}$$

for $j = 1, \ldots, d$. If $Z_1, \ldots, Z_d$ are arranged in a d-vector Z and we set

$$D_\delta = \left(I_d - \mathrm{diag}(\delta)^2\right)^{1/2}, \tag{5.18}$$

https://doi.org/10.1017/CBO9781139248891.006 Published online by Cambridge University Press

we can write more compactly

$$Z = D_\delta\, U_0 + \delta\, |U_1|\,. \tag{5.19}$$

Some algebraic work says that Z has a d-dimensional skew-normal distribution with parameters $(\bar\Omega, \alpha)$ related to δ and $\bar\Psi$ as follows:

$$\lambda = D_\delta^{-1}\, \delta\,, \tag{5.20}$$

$$\bar\Omega = D_\delta\, (\bar\Psi + \lambda\, \lambda^\top)\, D_\delta, \tag{5.21}$$

$$\alpha = \left(1 + \lambda^\top \bar\Psi^{-1} \lambda\right)^{-1/2} D_\delta^{-1} \bar\Psi^{-1} \lambda\,; \tag{5.22}$$

see Problem 5.2. In the scalar case, $\bar\Omega$ and $\bar\Psi$ reduce to 1 and $\lambda = \alpha$.

A direct link between the ingredients of the additive representation and those in (5.14) can be established by the standard orthogonalization scheme

$$U_1 = X_1, \quad U_0' = X_0 - \mathbb{E}\{X_0|X_1\} = X_0 - \delta\, X_1 \sim \mathrm{N}_d(0,\, \bar\Omega - \delta\, \delta^\top) \tag{5.23}$$

which, after transformation $U_0 = D_\delta^{-1} U_0'$ to have unit variances, leads to (5.16). Inversion of these relationships shows how to obtain X from U.

Minima and maxima

To introduce a stochastic representation in the form of minima and maxima, which generalizes the analogous one for the scalar case presented in Chapter 2, we make use of the variables and other elements introduced in the previous paragraph. Note that Z_j in (5.17) is algebraically equivalent to

$$Z_j = \operatorname{sgn}(\delta_j)\, \tfrac{1}{2}\, |V_j - W_j| + \tfrac{1}{2}\, (V_j + W_j)\,, \qquad j = 1, \ldots, d,$$

where

$$V_j = (1 - \delta_j^2)^{1/2}\, U_{0j} + \delta_j\, U_1\,, \qquad W_j = (1 - \delta_j^2)^{1/2}\, U_{0j} - \delta_j\, U_1\,.$$

The joint distribution of $V = (V_1, \ldots, V_d)$ and $W = (W_1, \ldots, W_d)$ is singular Gaussian, specifically

$$\begin{pmatrix} V \\ W \end{pmatrix} \sim \mathrm{N}_{2d}\left(0, \begin{pmatrix} D_\delta \bar\Psi D_\delta + \delta\delta^\top & D_\delta \bar\Psi D_\delta - \delta\delta^\top \\ D_\delta \bar\Psi D_\delta - \delta\delta^\top & D_\delta \bar\Psi D_\delta + \delta\delta^\top \end{pmatrix}\right), \tag{5.24}$$

where $\delta = (\delta_1, \ldots, \delta_d)^\top$ and D_δ is as in (5.18). The variables

$$V - W \sim \mathrm{N}_d(0, 4\, \delta\delta^\top), \qquad V + W \sim \mathrm{N}_d(0, 4\, D_\delta \bar\Psi D_\delta)$$

are independent, and the equality $V - W = 2U_1\delta$ confirms that $V - W$ has singular distribution. Recalling that

$$\max\{a, b\} = \tfrac{1}{2}\, |a - b| + \tfrac{1}{2}\, (a + b)\,, \quad \min\{a, b\} = -\tfrac{1}{2}\, |a - b| + \tfrac{1}{2}\, (a + b)$$

and writing

$$Z_j = \begin{cases} \max\{V_j, W_j\} & \text{if } \delta_j \geq 0, \\ \min\{V_j, W_j\} & \text{otherwise,} \end{cases} \tag{5.25}$$

it is visible that $Z = (Z_1, \ldots, Z_d)^\top \sim \mathrm{SN}_d$ with parameters (5.21)–(5.22).

5.1.4 Marginal distributions and another parameterization

Closure of the SN family with respect to marginalization follows from (5.10). More specifically, suppose that $Y \sim \mathrm{SN}_d(\xi, \Omega, \alpha)$ is partitioned as $Y^\top = (Y_1^\top, Y_2^\top)$ where the two components have dimension h and $d - h$, respectively, and denote by

$$\xi = \begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix}, \quad \Omega = \begin{pmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{pmatrix}, \quad \alpha = \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix}, \quad \delta = \begin{pmatrix} \delta_1 \\ \delta_2 \end{pmatrix} \tag{5.26}$$

the corresponding partitions of ξ, Ω, α and δ. Evaluation of (5.10) at $t = (s^\top, 0)^\top$ gives the moment generating function of Y_1 as

$$M_{Y_1}(s) = 2 \exp\left(s^\top \xi_1 + \tfrac{1}{2} s^\top \Omega_{11} s\right) \Phi(\delta_1^\top \omega_{11}\, s), \qquad s \in \mathbb{R}^h,$$

showing that Y_1 is of skew-normal type with location ξ_1 and scale matrix Ω_{11}. To find the slant parameter, $\alpha_{1(2)}$ say, we use (5.12) with δ replaced by δ_1, the first h components of (5.11). After some algebra, we arrive at

$$\alpha_{1(2)} = \left(1 + \alpha_2^\top \bar{\Omega}_{22\cdot 1} \alpha_2\right)^{-1/2} \left(\alpha_1 + \bar{\Omega}_{11}^{-1} \bar{\Omega}_{12} \alpha_2\right) \tag{5.27}$$

where

$$\bar{\Omega}_{11}^{-1} = (\bar{\Omega}_{11})^{-1}, \qquad \bar{\Omega}_{22\cdot 1} = \bar{\Omega}_{22} - \bar{\Omega}_{21} \bar{\Omega}_{11}^{-1} \bar{\Omega}_{12} \tag{5.28}$$

on partitioning $\bar{\Omega}$ similarly to Ω. To conclude, marginally

$$Y_1 \sim \mathrm{SN}_h(\xi_1, \Omega_{11}, \alpha_{1(2)}) \,. \tag{5.29}$$

Some remarks on the interpretation of the parameters $(\bar{\Omega}, \alpha)$ are now appropriate. From (5.17) it is apparent that the entries of the vector δ of the joint distribution coincide with the δ parameters of the marginal distributions. The same fact is visible also from the above expression of $M_{Y_1}(t)$, on taking $h = 1$. On the contrary, the jth entry of α does not individually provide information on the jth marginal of the joint distribution. In fact, from α_j we cannot even infer the sign of the corresponding component δ_j, that is, whether the jth marginal is positively or negatively asymmetric. However, a meaning can be attached to a null value of α_j, as we shall see in § 5.3.2 and § 5.3.5.

If one wants a parameterization where the parameter components have an interpretation as an individual slant parameter, this is possible on the basis of $(\bar{\Psi}, \lambda)$, recalling (5.20). We have seen that each choice of $\bar{\Psi}$ and λ in (5.16)–(5.19) corresponds to a distribution of type (5.1). The converse also holds: for each choice of $(\bar{\Omega}, \alpha)$ there is a corresponding choice of $(\bar{\Psi}, \delta)$ or equivalently of $(\bar{\Psi}, \lambda)$, in (5.16)–(5.19), leading to the same distribution; see Problem 5.3. Hence $(\bar{\Omega}, \alpha)$ and $(\bar{\Psi}, \lambda)$ are equivalent parameterizations for the same set of distributions. In both cases, the two components are variation independent, that is, they can be selected independently of each other. As an example of the contrary, $\bar{\Omega}$ and δ are not variation independent.

For the full class (5.3), write

$$\Omega = \psi(\bar{\Psi} + \lambda\lambda^\top)\psi = \Psi + \psi\lambda\lambda^\top\psi$$

where $\psi = \omega D_\delta = D_\delta\omega$ now represents the scale factor and $\Psi = \psi\bar{\Psi}\psi$; here D_δ is as in (5.18). Hence (5.3) can be equivalently expressed via the (ξ, Ψ, λ) parameter set as

$$2\,\varphi_d(x-\xi; \Psi + \psi\lambda\lambda^\top\psi)\,\Phi\left(\frac{1}{\sqrt{1+\lambda^\top\bar{\Psi}^{-1}\lambda}}\,\lambda^\top\bar{\Psi}^{-1}\psi^{-1}(x-\xi)\right). \qquad (5.30)$$

As already stated, the parameterization (ξ, Ψ, λ) has the advantage that the components of λ are interpretable individually. The reason why the parameterization (ξ, Ω, α) has been given a primary role is that it allows a simpler treatment in other respects. A basic fact is that (5.3) constitutes a simpler expression than (5.30). However, the reasons in favour of (Ω, α) are not indisputable, and one may legitimately prefer to use (Ψ, λ).

The skew-normal family is not closed under conditioning. A slight extension of the SN family which enjoys this property will be discussed in §5.3.

5.1.5 Cumulants and related quantities

From (5.10), the cumulant generating function of $Y \sim \mathrm{SN}_d(\xi, \Omega, \alpha)$ is

$$K(t) = \log M(t) = \xi^\top t + \tfrac{1}{2}t^\top\Omega\,t + \zeta_0(\delta^\top\omega\,t), \qquad t \in \mathbb{R}^d,$$

where $\zeta_0(x)$ is defined by (2.18). Taking into account (2.19), the first two derivatives of $K(t)$ are

$$\frac{\mathrm{d}}{\mathrm{d}t}K(t) = \xi + \Omega t + \zeta_1(\delta^\top\omega\,t)\,\omega\,\delta,$$

$$\frac{\mathrm{d}^2}{\mathrm{d}t\;\mathrm{d}t^\top}K(t) = \Omega + \zeta_2(\delta^\top\omega\,t)\,\omega\,\delta\,\delta^\top\omega,$$

https://doi.org/10.1017/CBO9781139248891.006 Published online by Cambridge University Press

and their values at $t = 0$ give

$$\mu = \mathbb{E}\{Y\} = \xi + \omega\, b\, \delta = \xi + \omega\, \mu_Z, \tag{5.31}$$

$$\Sigma = \mathrm{var}\{Y\} = \Omega - \omega\, \mu_Z\, \mu_Z^\top\, \omega = \omega\, \Sigma_Z\, \omega \tag{5.32}$$

where, analogously to the univariate case in § 2.1.2, we have set

$$\mu_Z = b\,\delta = \mathbb{E}\{Z\}\,, \qquad \Sigma_Z = \bar\Omega - \mu_Z\, \mu_Z^\top = \mathrm{var}\{Z\}$$

for $Z \sim \mathrm{SN}_d(0, \bar\Omega, \alpha)$. If $\xi = 0$, a quick way to obtain that $\mathbb{E}\{Y\, Y^\top\} = \Omega$ is by simply recalling the modulation invariance property.

The rth-order derivative of $K(t)$, for $r > 2$, takes the form

$$\frac{\mathrm{d}^r}{\mathrm{d}t_i\ \mathrm{d}t_j\ \cdots\ \mathrm{d}t_h} K(t) = \zeta_r(\delta^\top \omega\, t)\, \omega_i\, \omega_j \cdots \omega_h\, \delta_i\, \delta_j \cdots \delta_h, \tag{5.33}$$

where the expression of $\zeta_r(x)$ up to $r = 4$ is given by (2.20).

Evaluation at $t = 0$ of the above derivatives allows us to obtain an explicit expression of the coefficients of multivariate skewness and kurtosis introduced by Mardia (1970, 1974). Specifically, evaluation of (5.33) at $t = 0$ for $r = 3$ and insertion in (1.1) of Mardia (1974) lead to

$$\begin{aligned} \gamma^M_{1,d} = \beta^M_{1,d} &= \zeta_3(0)^2 \sum_{vst} \sum_{v's't'} \delta_v \delta_s \delta_t \delta_{v'} \delta_{s'} \delta_{t'} \sigma_Z^{vv'} \sigma_Z^{ss'} \sigma_Z^{tt'} \\ &= \left(\frac{4-\pi}{2}\right)^2 \left(\mu_Z^\top \Sigma_Z^{-1} \mu_Z\right)^3 \end{aligned} \tag{5.34}$$

where $\Sigma_Z^{-1} = (\sigma_Z^{st})$, and similarly when $r = 4$ we obtain

$$\begin{aligned} \gamma^M_{2,d} = \beta^M_{2,d} - d(d+2) &= \zeta_4(0) \sum_{rstu} \delta_v \delta_s \delta_t \delta_u \sigma_Z^{vs} \sigma_Z^{tu} \\ &= 2(\pi - 3) \left(\mu_Z^\top \Sigma_Z^{-1} \mu_Z\right)^2 \end{aligned} \tag{5.35}$$

from expressions (1.2) and (2.9) of Mardia (1974). The two measures, $\gamma^M_{1,d}$ and $\gamma^M_{2,d}$, depend on α and $\bar\Omega$ through the quadratic form $\mu_Z^\top \Sigma_Z^{-1} \mu_Z$, which in turn can be rewritten as

$$\mu_Z^\top \Sigma_Z^{-1} \mu_Z = \frac{(2/\pi)\, \alpha_*^2}{1 + (1 - 2/\pi)\, \alpha_*^2}, \tag{5.36}$$

where

$$\alpha_* = (\alpha^\top \bar\Omega\, \alpha)^{1/2} \in [0, \infty) \tag{5.37}$$

can then be seen as the regulating quantity. Therefore, as for Mardia's measures, the scalar quantity α_* encapsulates comprehensively the departure from normality. Equivalently, (5.36) and other expressions which will appear later can be written as functions of

$$\delta_* = (\delta^\top \bar{\Omega}^{-1} \delta)^{1/2} \in [0, 1), \tag{5.38}$$

where as usual δ is given by (5.11). These quantities are connected via

$$\delta_*^2 = \frac{\alpha_*^2}{1 + \alpha_*^2}, \qquad \alpha_*^2 = \frac{\delta_*^2}{1 - \delta_*^2}\,.$$

Some algebraic manipulation gives further insight about α_*. In (5.36) write α_* as a function $\delta(\alpha_*)$ according to (2.6) on p. 26. We can then rewrite (5.36) as $\mu_{\alpha_*}^2/\sigma_{\alpha_*}^2$, where the two components are functions of $\delta(\alpha_*)$ given by (2.26). Finally, we arrive at

$$\gamma_{1,d}^M = \left(\frac{4-\pi}{2}\right)^2 \left(\frac{\mu_{\alpha_*}^2}{\sigma_{\alpha_*}^2}\right)^3, \qquad \gamma_{2,d}^M = 2(\pi - 3)\left(\frac{\mu_{\alpha_*}^2}{\sigma_{\alpha_*}^2}\right)^2, \tag{5.39}$$

that is, $\gamma_{1,d}^M$ and $\gamma_{2,d}^M$ correspond to the square of γ_1 and to the γ_2 coefficient, respectively, for the distribution $\mathrm{SN}(0, 1, \alpha_*)$. These expressions arise from mere algebraic rewriting of (5.34) and (5.35), but they are notionally associated with a distribution $\mathrm{SN}(0, 1, \alpha_*)$. This idea will take a more precise shape in § 5.1.8.

An implication of (5.39) is that $\gamma_{1,d}^M$ and $\gamma_{2,d}^M$ range from 0 to $(\gamma_1^{\max})^2$ and to $\gamma_2^{\max}$, respectively, where $\gamma_1^{\max}$ and $\gamma_2^{\max}$ are given by (2.31).

5.1.6 Linear, affine and quadratic forms

From the moment generating function (5.10), it is visible that the family of multivariate skew-normal distributions is closed under affine transformations. More specifically, if $Y \sim \mathrm{SN}_d(\xi, \Omega, \alpha)$, A is a full-rank $d \times h$ matrix, with $h \le d$, and $c \in \mathbb{R}^h$, then some algebraic work shows that

$$X = c + A^\top Y \sim \mathrm{SN}_h(\xi_X, \Omega_X, \alpha_X) \tag{5.40}$$

where

$$\xi_X = c + A^\top Y\,, \tag{5.41}$$

$$\Omega_X = A^\top \Omega A\,, \tag{5.42}$$

$$\alpha_X = \left(1 - \delta^\top \omega A \Omega_X^{-1} A^\top \omega \delta\right)^{-1/2} \omega_X\, \Omega_X^{-1} A^\top \omega \delta \tag{5.43}$$

having set $\omega_X = (\Omega_X \odot I_h)^{1/2}$ and, as usual, δ is given by (5.11). When $h = 1$, so that A reduces to a vector, a say, (5.43) simplifies to

$$\alpha_X = \left(a^\top \Omega a - (a^\top \omega\delta)^2\right)^{-1/2} a^\top \omega\delta\,. \tag{5.44}$$

To examine the question of independence among components of an SN variable, we need the following preliminary result, which is also of independent interest.

Proposition 5.5 *For any choice of $a_1, a_2 \in \mathbb{R}$, $\mu_1, b_1 \in \mathbb{R}^p$, $\mu_2, b_2 \in \mathbb{R}^q$ such that $b_1 \neq 0$ and $b_2 \neq 0$ and symmetric positive-definite matrices Σ_1, Σ_2, there exist no $a, c \in \mathbb{R}$, $b, \mu \in \mathbb{R}^{p+q}$ and matrix Σ such that*

$$\begin{aligned}\varphi_p(x_1-\mu_1;\Sigma_1)\,\Phi(a_1+b_1^\top x_1)\,\varphi_q(x_2-\mu_2;\Sigma_2)\,\Phi(a_2+b_2^\top x_2)\\ = c\,\varphi_{p+q}(x-\mu;\Sigma)\,\Phi(a+b^\top\, x)\end{aligned} \tag{5.45}$$

for all $x_1 \in \mathbb{R}^p$, $x_2 \in \mathbb{R}^q$, $x = (x_1^\top, x_2^\top)^\top$.

Proof Select one non-zero component of b_1 and one of b_2, which exist. Set x_1 and x_2 to have value x_0 in these components and 0 otherwise. For these x_1 and x_2, (5.45) is a function of x_0 only and it is of the form (2.9), for which we know that equality cannot hold for all x_0. QED

In the special case with $a_1 = a_2 = 0$, (5.45) corresponds, up to a multiplicative constant, to the product of two multivariate SN densities, both with non-null slant parameter. The implication is that this product cannot be expressed as some other multivariate SN density. By repeated application of this fact we can state the following: if we partition $Y \sim \mathrm{SN}_d(0, \Omega, \alpha)$ in h blocks, so that $Y^\top = (Y_1^\top, \ldots, Y_h^\top)$, then joint independence of these h components requires that the parameters have a structure of the following form, in an obvious notation:

$$\Omega = \mathrm{diag}(\Omega_{11}, \ldots, \Omega_{hh}), \qquad \alpha = (0, \ldots, \alpha_j, \ldots, 0)^\top \tag{5.46}$$

so that the joint density (5.1) can be factorized into a product with separate variables.

This conclusion highlights an important aspect of the skew-normal distribution: independence among a set of components can hold only if at most one of them is marginally skew-normal. A direct implication of this fact is that two asymmetric marginal components of a multivariate skew-normal variate cannot be independent. Another implication is that the joint distribution of a set of independent skew-normal variables with non-zero slant (univariate or multivariate) cannot be multivariate SN.

https://doi.org/10.1017/CBO9781139248891.006 Published online by Cambridge University Press

As a further generalization, we now want to extend the above fact to a linear transformation $X = A^\top Y$, for a non-singular square matrix A.

Proposition 5.6 *Given $Y \sim \mathrm{SN}_d(0, \Omega, \alpha)$, consider the linear transform*

$$X = A^\top Y = \begin{pmatrix} X_1 \\ \vdots \\ X_h \end{pmatrix} = \begin{pmatrix} A_1^\top \\ \vdots \\ A_h^\top \end{pmatrix} Y \tag{5.47}$$

where A is a $d \times d$ non-singular matrix and $(A_1, \dots, A_h) = A$. Then $X_1, \dots, X_h$ are mutually independent variables if and only if the following conditions hold simultaneously:

(a) $A_i^\top \Omega A_j = 0$ for $i \neq j$,
(b) $A_i^\top \Omega \omega^{-1} \alpha \neq 0$ for at most one i.

Proof When condition (a) holds, use of (5.12), (5.42) and (5.43) yields

$$\Omega_X = \mathrm{diag}(A_1^\top \Omega A_1, \dots, A_h^\top \Omega A_h),$$
$$\alpha_X = \omega_X (A^\top \Omega A)^{-1} A^\top \Omega \omega^{-1} \alpha = \omega_X \begin{pmatrix} (A_1^\top \Omega A_1)^{-1} A_1^\top \Omega \omega^{-1} \alpha \\ \vdots \\ (A_h^\top \Omega A_h)^{-1} A_h^\top \Omega \omega^{-1} \alpha \end{pmatrix}.$$

From the last expression, it follows that, if condition (b) is fulfilled too, only one among the h blocks of α_X is non-zero. Hence the joint density of X can be factorized in an obvious manner and sufficiency is proved.

To prove necessity, note first that, if independence among $X_1, \dots, X_h$ holds, then the joint density of X equals the product of the h marginal densities. Taking into account Proposition 5.5, equality can occur if only one block of α_X is not zero and Ω_X is block diagonal. QED

Corollary 5.7 *Given $Y \sim \mathrm{SN}_d(0, \Omega, \alpha)$, consider the partition $\{s_1, \dots, s_h\}$ of $\{1, \dots, d\}$, and let $(Y_{s_1}^\top, \dots, Y_{s_h}^\top)$ denote the corresponding block partition of Y. Then $Y_{s_1}, \dots, Y_{s_h}$ are mutually independent variables if and only if the following conditions hold simultaneously:*

(a) $\Omega_{s_i s_j} = 0$ for $i \neq j$,
(b) $\alpha_{s_i} \neq 0$ for at most one i,

where $\Omega_{s_i s_j}$ is the block portion of Ω formed by rows s_i and columns s_j.

The next result states that another classical property of the multivariate normal distribution holds for the SN case as well.

Proposition 5.8 *If $Y \sim \mathrm{SN}_d(\xi, \Omega, \alpha)$, its univariate components are pairwise independent if and only if they are mutually independent.*

https://doi.org/10.1017/CBO9781139248891.006 Published online by Cambridge University Press

Proof Necessity is trivial. To prove sufficiency, note firstly that closure under marginalization ensures that the joint distribution of any pair of marginal components Y_i and Y_j, say, is of type $\mathrm{SN}_2(\xi', \Omega', \alpha')$, where the off-diagonal element of the matrix Ω' is Ω_{ij}. Also, from Proposition 5.6, Y_i and Y_j are independent if $\Omega_{ij} = 0$ and at least one between Y_i and Y_j is Gaussian, implying that the matrix Ω is diagonal and at least $d-1$ univariate marginal components of Y are Gaussian, that is, $d-1$ entries of δ defined in (5.11) should be zero. Mutual independence follows by noticing that the structure of the parameters Ω and α under pairwise independence guarantees that conditions (a) and (b) in Proposition 5.6 are fulfilled for $h = d$. QED

The skew-normal distribution with 0 location shares with the normal family the distributional properties of the associated quadratic forms, because of the modulation invariance property of Proposition 1.4. More specifically, the connection is as follows.

Corollary 5.9 *If $Y \sim \mathrm{SN}_d(0, \Omega, \alpha)$ and A is a $d \times d$ symmetric matrix, then*

$$Y^\top A Y \overset{\mathrm{d}}{=} X^\top A X \tag{5.48}$$

where $X \sim \mathrm{N}_d(0, \Omega)$.

This simple annotation immediately makes available the vast set of existing results for quadratic forms of multinormal variables. One statement of this type is property (5.7), among many others. The implications of modulation invariance are, however, not limited to a single quadratic form like in (5.48) by considering a q-valued even function $t(\cdot)$ in Proposition 1.4. For instance, the next result represents a form of Fisher–Cochran theorem.

Corollary 5.10 *If $Y \sim \mathrm{SN}_d(0, I_d, \alpha)$ and $A_1, \ldots, A_n$ are symmetric positive semi-definite matrices with rank $r_1, \ldots, r_n$ such that $A_1 + \cdots + A_n = I_d$, then a necessary and sufficient condition that $Y^\top A_j Y \sim \chi^2_{r_j}$ and are independent is that $r_1 + \cdots + r_n = d$.*

5.1.7 A characterization result

A classical result of normal distribution theory is that, if all linear combinations $h^\top Z$ of a multivariate random variable Z have univariate normal distribution, then Z is multinormal. The next proposition states a matching fact for the normalized skew-normal distribution.

https://doi.org/10.1017/CBO9781139248891.006 Published online by Cambridge University Press

Proposition 5.11 *Consider a d-dimensional random variable Z such that $R = \mathbb{E}\{Z Z^\top\}$ is a finite and positive-definite correlation matrix. If, for any $h \in \mathbb{R}^d$ such that $h^\top R h = 1$, there exists a value α_h such that $h^\top Z \sim \mathrm{SN}(0, 1, \alpha_h)$, then $Z \sim \mathrm{SN}_d(0, R, \alpha)$ for some $\alpha \in \mathbb{R}^d$ and R is a correlation matrix.*

Proof Denote $T = h^\top Z \sim \mathrm{SN}(0, 1, \alpha_h)$ which has moment generating function $M_T(t) = 2\, e^{t^2/2}\Phi(\delta_h t)$, where δ_h is related to α_h as in (2.6). First, note that $b\,\delta_h = \mathbb{E}\{T\} = \mathbb{E}\{h^\top Z\} = h^\top \mu$, where $\mu = \mathbb{E}\{Z\}$ and $b = \sqrt{2/\pi}$. Therefore, choosing $h_0 = w^{-1} R^{-1}\mu$ where $w^2 = \mu^\top R^{-1}\mu$, so that $h_0^\top R h_0 = 1$, we obtain $b\delta_{h_0} = h_0^\top \mu = w$, which implies $b^2 > \mu^\top R^{-1}\mu$. Then, the vector

$$\alpha = (b^2 - w^2)^{-1/2} R^{-1}\mu = \left(b^2 - \mu^\top R^{-1}\mu\right)^{-1/2} R^{-1}\mu$$

exists and, after some simple algebra, it turns out to fulfil this equality:

$$(1 + \alpha^\top R \alpha)^{-1/2}\, \alpha^\top R\, h = \delta_h\,.$$

Hence, taking into account that $h^\top R\, h = 1$, we can write

$$M_T(t) = 2\, \exp\left(\tfrac{1}{2} t^2 h^\top R\, h\right)\, \Phi\left((1 + \alpha^\top R\alpha)^{-1/2}\, \alpha^\top R\, h\, t\right)\,.$$

For any $u \in \mathbb{R}^d$, write it as $u = t\,h$ where t is a real and $h \in \mathbb{R}^d$ such that $h^\top R\, h = 1$. The moment generating function of Z at u is $M_Z(u) = \mathbb{E}\{\exp(t\, h^\top Z)\}$, which equals the above expression of $M_T(t)$ with $t\,h$ replaced by u. Comparing this with (5.10) evaluated at u, where δ is given by (5.11), we conclude that the moment generating function of Z is that of $\mathrm{SN}_d(0, R, \alpha)$, where R is a positive-definite correlation matrix. QED

This characterization result could be used to develop the skew-normal distribution theory taking this property as the one which *defines* the probability distribution, following a similar route to that taken for the normal distribution, as recalled at the beginning of this section; see Rao (1973, Section 8a.1) and Mardia *et al*. (1979, Section 3.1.2).

5.1.8 Canonical form

We focus now on a specific type of linear transformation of a multivariate skew-normal variable, having special relevance for theoretical developments but to some extent also for practical reasons.

Proposition 5.12 *Given a variable $Y \sim \mathrm{SN}_d(\xi, \Omega, \alpha)$, there exists an affine transformation $Z^* = A_*(Y - \xi)$ such that $Z^* \sim \mathrm{SN}_d(0, I_d, \alpha_{Z^*})$, where $\alpha_{Z^*} = (\alpha_*, 0, \ldots, 0)^\top$ and α_* is defined by (5.37).*

https://doi.org/10.1017/CBO9781139248891.006 Published online by Cambridge University Press

Proof Recall that in § 5.1 we have introduced the SN distribution assuming $\Omega > 0$ and the factorization $\Omega = \omega\bar{\Omega}\omega$ introduced right after (5.3); also let $\bar{\Omega} = C^\top C$ for some non-singular matrix C. If $\alpha \neq 0$, one can find an orthogonal matrix P with the first column proportional to $C\alpha$, while for $\alpha = 0$ we set $P = I_d$. Finally, define $A_* = (C^{-1}P)^\top \omega^{-1}$. It can be checked with the aid of formulae (5.41)–(5.43) for affine transformations that $Z^* = A_*(Y-\xi)$ has the stated distribution. QED

The variable Z^*, which we shall sometimes refer to as a 'canonical variate', comprises d independent components. The joint density is given by the product of $d-1$ standard normal densities and at most one non-Gaussian component $\mathrm{SN}(0, 1, \alpha_*)$; that is, the density of Z^* is

$$f_*(x) = 2 \prod_{i=1}^{d} \varphi(x_i)\, \Phi(\alpha_* x_1)\,, \qquad x = (x_1, \ldots, x_d)^\top \in \mathbb{R}^d.$$

In § 5.1.5, α_* has emerged as the summary quantity which regulates the Mardia coefficients of multivariate skewness and kurtosis $\gamma^M_{1,d}$ and $\gamma^M_{2,d}$. Among the set of $\mathrm{SN}(\xi, \Omega, \alpha)$ distributions sharing the same value of α_*, the canonical form can be regarded as the most 'pure' representative of this set, since all departure from normality is concentrated in a single component, independent from the others. Consequently, quantities which are invariant with respect to affine transformations can be computed more easily for the canonical form, and they hold for all distributions with the same value of α_*.

Therefore, expressions (5.39) of Mardia's measures could be derived as an instance of this scheme. A similar argument can be applied to compute the measures of multivariate skewness and kurtosis introduced by Malkovich and Afifi (1973). Since these measures are also invariant over affine transformations of the variable, we can reduce the problem to the canonical form, hence to the single univariate component possibly non-Gaussian. The implication is that we arrive again at expressions (5.39), equivalent to (5.34)–(5.35).

Inspection of the proof of Proposition 5.12 shows that, when $d > 2$, there exist several possible choices of A_*, hence many variables Z^*, all with the same distribution $f_*(x)$. However, this lack of uniqueness is not a problem. To draw an analogy, the canonical form plays a role loosely similar to the transformation which orthogonalizes the components of a multivariate normal variable, and also in that case the transformation is not unique.

Although Proposition 5.12 ensures that it is possible to obtain a canonical

https://doi.org/10.1017/CBO9781139248891.006 Published online by Cambridge University Press

form, and we have remarked that in general there are many possible ways to do so, it is not obvious how to achieve the canonical form in practice. The next result explains this.

Proposition 5.13 *For $Y \sim \mathrm{SN}_d(\xi, \Omega, \alpha)$ define $M = \Omega^{-1/2}\Sigma\Omega^{-1/2}$, where $\Sigma = var\{Y\}$ and $\Omega^{1/2}$ is the unique positive definite symmetric square root of Ω. Let $Q\Lambda Q^\top$ denote a spectral decomposition of M, where without loss of generality we assume that the diagonal elements of Λ are arranged in increasing order, and $H = \Omega^{-1/2}Q$. Then*

$$Z^* = H^\top(Y - \xi)$$

has canonical form.

Proof From the assumptions made, it follows that $H^{-1} = Q^\top\Omega^{1/2}$ and $\Sigma = (H^\top)^{-1}\Lambda H^{-1}$. In addition, use of (5.41) and (5.42) lends $\xi_{Z^*} = 0$ and $\Omega_{Z^*} = H^\top\Omega H = I_d$. In an obvious notation, therefore, we can write

$$\Sigma_{Z^*} = H^\top\Omega H - b^2\delta_{Z^*}\delta_{Z^*}^\top = I_d - b^2\delta_{Z^*}\delta_{Z^*}^\top,$$

where $b^2 = 2/\pi$ and $\delta_{Z^*} = H^\top\omega\delta$ on recalling (5.32). Since we can also write

$$\Sigma_{Z^*} = H^\top\Sigma H = H^\top(H^\top)^{-1}\Lambda H^{-1}H = \Lambda,$$

it follows that vector δ_{Z^*} can have at most one non-zero component, in the first position. This value will be $(\delta^\top\omega HH^\top\omega\delta)^{1/2} = (\delta^\top\bar\Omega^{-1}\delta)^{1/2} = \delta_*$, where the final equality follows from (5.38). Finally, from (5.12) and (2.15), we obtain

$$\alpha_{Z^*} = (1 - \delta_*)^{-1/2}(\delta_*, 0, \dots, 0)^\top = (\alpha_*, 0, \dots, 0)^\top. \qquad \text{QED}$$

So far we have employed the canonical form only to show simplified ways of computing multivariate coefficients of skewness and kurtosis. The next result, instead, seems difficult to prove without this notion. Recall that Proposition 5.1 implies that the multivariate SN density has a unique mode, like in the univariate case.

Proposition 5.14 *The unique mode of the distribution $\mathrm{SN}_d(\xi, \Omega, \alpha)$ is*

$$M_0 = \xi + \frac{m_0^*}{\alpha_*}\omega\bar\Omega\alpha = \xi + \frac{m_0^*}{\delta_*}\omega\delta, \tag{5.49}$$

where δ and δ_* are given by (5.11) and by (5.38), respectively, and m_0^* is the mode of the univariate $\mathrm{SN}(0, 1, \alpha_*)$ distribution.

Proof Given a variable $Y \sim \mathrm{SN}_d(\xi, \Omega, \alpha)$, consider first the mode of the corresponding canonical variable $Z^* \sim \mathrm{SN}_d(0, I_d, \alpha_{Z^*})$. We find this mode

https://doi.org/10.1017/CBO9781139248891.006 Published online by Cambridge University Press

by equating to zero the gradient of the density function, that is by solving the following equations with respect to $z_1, \ldots, z_d$:

$$z_1\,\Phi(\alpha_* z_1) - \varphi_1(\alpha_* z_1)\,\alpha_* = 0, \qquad z_j\,\Phi(\alpha_* z_1) = 0 \quad \text{for } j = 2, \ldots, d\,.$$

The last $d-1$ equations are fulfilled when $z_j = 0$, whilst the unique root of the first one corresponds to the mode, m_0^* say, of the $\mathrm{SN}(0, 1, \alpha_*)$ distribution. Therefore, the mode of Z^* is $M_0^* = (m_0^*, 0, \ldots, 0)^\top = (m_0^*/\alpha_*)\,\alpha_{Z^*}$. From Proposition 5.13, write $Y = \xi + \omega C^\top P Z^*$ and $\alpha_Z^* = P^\top C\alpha$. Since the mode is equivariant with respect to affine transformations, the mode of Y is

$$M_0 = \xi + \frac{m_0^*}{\alpha_*}\omega C^\top P P^\top C\alpha = \xi + \frac{m_0^*}{\alpha_*}\omega\bar{\Omega}\alpha = \xi + \frac{m_0^*}{\delta_*}\omega\delta,$$

where the last equality follows taking into account (5.11) and (5.38). QED

Equation (5.49) says that the mode lies on the direction of the vector $\omega\delta$ starting from location ξ. Recall from (5.31) that this is the same direction where the mean μ of this distribution is located. In other words, ξ, μ and M_0 are aligned points. Therefore, $\omega\delta$ is the direction where departure from Gaussianity displays more prominently its effect, and the intensity of this departure is summarized by α_*, or equivalently by δ_*. These conclusions are illustrated graphically in Figure 5.2, which refers to the case with

$$\xi = \begin{pmatrix} 3 \\ 5 \end{pmatrix}, \qquad \Omega = \begin{pmatrix} 2 & 2 \\ 2 & 4 \end{pmatrix}, \qquad \alpha = \begin{pmatrix} -5 \\ 2 \end{pmatrix}; \tag{5.50}$$

the labels of the contour lines will be explained in Complement 5.2.

Besides the theoretical value of (5.49), there is also a practical one. Finding the mode of $\mathrm{SN}_d(\xi, \Omega, \alpha)$ requires a numerical maximization procedure, and in principle this search should take place in the d-dimensional Euclidean space, but by means of (5.49) we can restrict the search to a one-dimensional set, from ξ along the direction $\omega\delta$.

5.1.9 Bibliographic notes

Azzalini and Dalla Valle (1996) have introduced the multivariate version of the skew-normal distribution via the additive construction (5.19). Therefore, the parameterization adopted initially was $(\bar{\Psi}, \lambda)$, and the density function so obtained was written as a function of $\bar{\Omega}$ and α. However, at that stage the latter quantities did not yet appear to form a parameter set. They

https://doi.org/10.1017/CBO9781139248891.006 Published online by Cambridge University Press

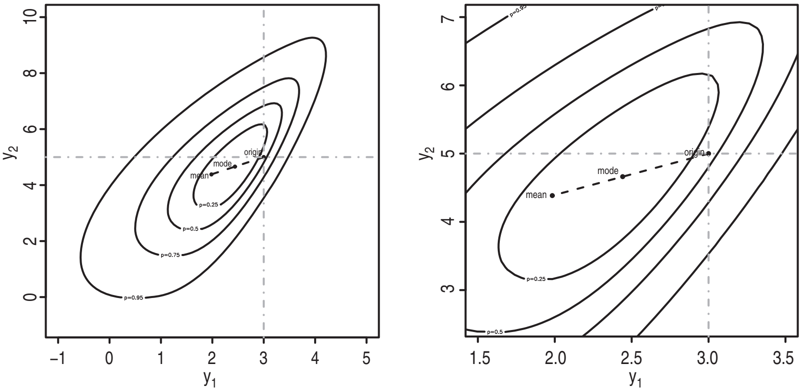


Figure 5.2 Contour lines plot of the bivariate skew-normal density whose parameters are given in (5.50) with the mean value and the mode superimposed, and their line of alignment. The right-hand plot enlarges the central portion of the left-hand plot.

have also shown that the same family of distributions can be generated by the conditioning mechanism (5.13), and have obtained some other results, notably the chi-square property (5.7), the moment generating function and the distribution function.

Azzalini and Capitanio (1999) have shown that the set of normalized SN distributions could equally be parameterized by $(\bar{\Omega}, \alpha)$. From this basic fact, much additional work has been developed, which represents the core part of the exposition in the preceding pages. One of their results was the canonical form, which has been explored further by Capitanio (2012); her paper includes Propositions 5.13 and 5.14, and other results to be recalled later. Before the general property of modulation invariance was discovered, various specific instances were obtained; see for instance Loperfido (2001). Loperfido (2010) verifies by a direct computation the coincidence of the multivariate indices of Mardia with those of Malkovich and Afifi; in addition, he shows a direct correspondence between the canonical form and the principal components, in the special case that α is an eigenvector of Ω. Representation (5.25) via minima and maxima is an extension of a result by Loperfido (2008). Proposition 5.8 seems to be new. Balakrishnan and Scarpa (2012) have examined a range of other multivariate measures of skewness for SN variates, some of vectorial type.

Following Azzalini and Capitanio (1999), most of the subsequent literature has adopted the (Ω, α) parameterization, or some variant of it, typically

https://doi.org/10.1017/CBO9781139248891.006 Published online by Cambridge University Press

(Ω, η) but often still denoted (Ω, α). Moreover in some papers, especially in the earlier ones, Ω denotes what here is $\bar{\Omega}$. So the reader should pay attention to which quantities are really intended. Adcock and Shutes (1999) adopt instead a parameterization of type (Ψ, λ), very similar to (5.30), which the author find preferable from the point of view of interpretability for financial applications. Their work is actually based on the extended distribution of § 5.3.

Genton *et al*. (2001) provide expression for moments up to the fourth order of SN variables and for lower-order moments of their quadratic forms. However, since what the authors denote Ω is a scale-free matrix, an adjustment is required to use these expressions in the general case: Ω must be interpreted as including the scale factor ω, that is, with the same meaning as in this book and δ must be replaced by $\omega\delta$ throughout.

The characterization result in Proposition 5.11 has been presented by Gupta and Huang (2002); the proof given here differs in two steps of the argument. Their paper includes other facts on linear and quadratic forms of skew-normal variates.

Javier and Gupta (2009) study the mutual information criterion for a multivariate SN distribution. Since its expression involves a quantity of type $\mathbb{E}\{\zeta_0(\alpha^\top Z)\}$, no explicit expression can be obtained, only reduced to a univariate integral, which is then expanded into an infinite series. Substantial additional work in this context, focusing on Shannon entropy and Kullback–Leibler divergence, has been carried out by Contreras-Reyes and Arellano-Valle (2012). Follow-up work by Arellano Valle *et al*. (2013) deals with similar issues for the broader class of skew-elliptical distributions, which are presented in Chapter 6.

Additional results on the multivariate SN distribution are recalled in the complements and in the set of problems at the end of the chapter.

5.2 Statistical aspects

5.2.1 Log-likelihood function and parameter estimation

Consider directly a regression setting where the ith component $y_i \in \mathbb{R}^d$ of $y = (y_1, \ldots, y_n)^\top$ is sampled from $Y_i \sim \mathrm{SN}_d(\xi_i, \Omega, \alpha)$, with independence among the Y_i's. Assume that the location parameter ξ_i is related to a set of p explanatory variables x_i via

$$\xi_i^\top = x_i^\top \beta, \qquad i = 1, \ldots, n, \tag{5.51}$$

https://doi.org/10.1017/CBO9781139248891.006 Published online by Cambridge University Press

for some $p \times d$ matrix β of unknown parameters, where the covariates vector x_i has a 1 in the first position. We arrange the vectors $x_1, \ldots, x_n$ in a $n \times p$ matrix X (with $n > p$), which we assume to have rank p.

We commonly say that the DP is formed by (β, Ω, α), but duplicated elements must be removed; hence the more appropriate expression is

$$\theta^{\text{DP}} = \begin{pmatrix} \text{vec}(\beta) \\ \text{vech}(\Omega) \\ \alpha \end{pmatrix}, \tag{5.52}$$

where vec(·) is the operator which stacks the columns of a matrix and vech(·) stacks the lower triangle, inclusive of the diagonal, of a symmetric matrix. From (5.3), the log-likelihood function is

$$\ell = c - \tfrac{1}{2} n \log \det(\Omega) - \tfrac{1}{2} n \operatorname{tr}(\Omega^{-1} S_\beta) + 1_n^\top \zeta_0(R_\beta\, \omega^{-1} \alpha) \tag{5.53}$$

where 1_n is the n-vector of all 1's,

$$c = -\tfrac{1}{2} n\, d\, \log(2\pi)\,, \qquad R_\beta = y - X\beta\,, \qquad S_\beta = n^{-1} R_\beta^\top R_\beta\,,$$

ζ_0 is defined by (2.18). The notation $\zeta_0(x)$ when x is a vector must be interpreted as component-wise evaluation, similarly to (3.15); in the following, we shall employ the same convention also for other functions.

Maximization of this log-likelihood must be pursued numerically, over a parameter space of dimension $pd + d(d+3)/2$, either by direct search of the function or using an EM-type algorithm. Here we describe a technique which works by direct optimization of the log-likelihood, combining analytical and numerical maximization.

First of all, notice that, for the purpose of this maximization, it is convenient to reparametrize temporarily the problem by replacing the component α of (5.52) with $\eta = \omega^{-1}\alpha$, since η enters only the final term of (5.53). Expression (5.53) without the last summand is the same as a Gaussian log-likelihood, and Ω does not enter the final term in the $(\text{vec}(\beta), \text{vech}(\Omega), \eta)$ parameterization. Using a well-known fact for Gaussian likelihoods, we can say immediately that, for any given β, maximization with respect to Ω is achieved at S_β. Plugging this expression into (5.53) lends the profile log-likelihood

$$\ell_*(\beta, \eta) = c - \tfrac{1}{2} n \log \det(S_\beta) - \tfrac{1}{2} n\, d + 1_n^\top \zeta_0(R_\beta\, \eta), \tag{5.54}$$

whose maximization must now be performed numerically with respect to $d(p+1)$ parameter components. This process can be speeded up

https://doi.org/10.1017/CBO9781139248891.006 Published online by Cambridge University Press

considerably if the partial derivatives

$$\frac{\partial \ell_*}{\partial \beta} = X^\top R_\beta S_\beta^{-1} - X^\top \zeta_1(R_\beta\,\eta)\,\eta^\top, \qquad \frac{\partial \ell_*}{\partial \eta} = R_\beta^\top\,\zeta_1(R_\beta\,\eta)$$

are supplied to a quasi-Newton algorithm. Once we have obtained the values $\hat\beta$ and $\hat\eta$ which maximize (5.54), the MLE of Ω is $\hat\Omega = S_{\hat\beta}$. From here we obtain $\hat\omega$, in an obvious notation, and the MLE of α as $\hat\alpha = \hat\omega\,\hat\eta$, recalling the equivariance property of MLE.

A form of penalized log-likelihood is possible, similarly to (3.30) with α^2 in (3.35) replaced by α_*^2. In this case, however, an equivalent of the profile log-likelihood (5.54) is not available.

After estimates of the parameters have been obtained, model adequacy can be examined graphically by comparing the fitted distributions with the data scatter, although in the multivariate case this must be reduced to a set of bivariate projections, or possibly trivariate projections when dynamic graphics can be employed.

Another device, aimed at an overall evaluation of the model fitting, is the perfect analogue of a diagnostic tool commonly in use for multivariate normal distributions (Healy, 1968), based on the empirical analogues of the Mahalanobis-type distances

$$d_i = (y_i - \hat\xi_i)^\top \hat\Omega^{-1} (y_i - \hat\xi_i), \qquad i = 1, \ldots, n, \tag{5.55}$$

whose approximate reference distribution is χ_d^2, recalling (5.7). Here $\hat\xi_i^\top = x_i^\top \hat\beta$, the estimated location parameter for the ith observation, becomes a constant value $\hat\xi$ in the case of a simple sample. From these d_i's, we construct QQ-plots and PP-plots similar to those employed in the univariate case.

For a simple illustration of the above graphical devices, we make use of some variables of the Grignolino wine data. Specifically, we introduce the following multivariate response linear regression model:

$$(\text{tartaric\_acid},\ \text{malic\_acid}) = \beta_0 + \beta_1\,(\text{fixed\_acidity}) + \varepsilon,$$

where $\varepsilon \sim \mathrm{SN}_2(0, \Omega, \alpha)$ and β_0, β_1 are vectors in $\mathbb{R}^2$.

After estimating β_0, β_1, Ω and α by maximum likelihood, the residuals of the fitted model have been plotted in Figure 5.3 with superimposed contour lines of the fitted error distribution. Each of these curves surrounds an area of approximate probability indicated by the respective curve label, using the method to be described in Complement 5.2. The visual impression is that the fitted distribution matches adequately the scatter of the residuals. It is true that there are four points out of 71 which fall outside the curve

https://doi.org/10.1017/CBO9781139248891.006 Published online by Cambridge University Press

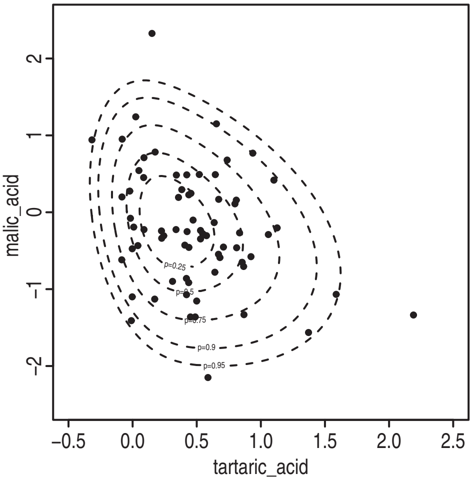


Figure 5.3 Grignolino wine data: empirical distribution of residuals and fitted parametric model after the linear component due to (fixed_acidity) has been removed from the joint distribution of (tartaric_acid, malic_acid).

labelled $p = 0.95$, somewhat more than expected, but two of them are just off the boundary. There is then some indication of a more elongated ‘tail’ than the normal one, but only in a mild form.

The overall impression of an essentially adequate data fit is supported also by the QQ-plot based on the distances (5.55), displayed in the left panel of Figure 5.4; there is only one point markedly distant from the ideal identity line. The right panel of the same figure displays the corresponding PP-plot, and compares it with the similar construct under normality assumption and least-squares (LS) fit. There is a quite clear indication of an improvement provided by the SN fit, whose points are visibly closer to the identity lines than the LS points.

Bibliographic notes

The above-described technique for maximization of the log-likelihood and the Healy-type graphical diagnostics have been put forward by Azzalini and Capitanio (1999, Section 6.1). Using these expressions of the partial derivatives, Azzalini and Genton (2008) deduce that the profile log-likelihood always has a stationary point at the point where β equals the least-squares estimate and $\eta = 0 = \alpha$; they also extend the result to a broader setting where the G_0 distribution of the modulation factor is not necessarily

https://doi.org/10.1017/CBO9781139248891.006 Published online by Cambridge University Press

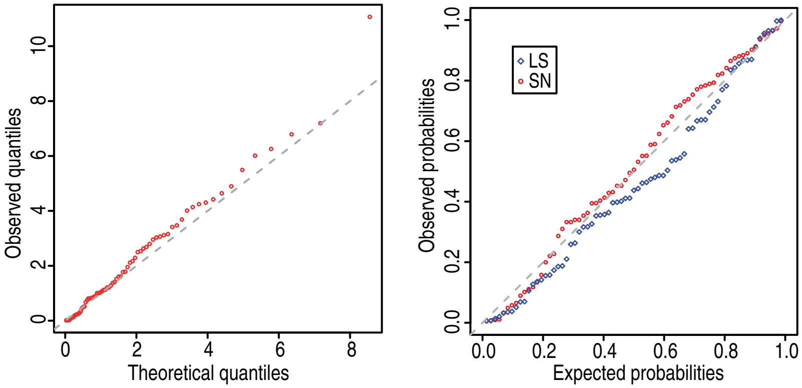


Figure 5.4 Grignolino wine data: QQ-plot (left) and PP-plot (right) of the same fitting as Figure 5.3; in the right plot the points corresponding to the least-squares fit are also displayed.

Gaussian. Instead of employing graphical diagnostics, the distributional assumption may be examined using a formal test procedure using the method proposed by Meintanis and Hlávka (2010) based on the empirical moment generating function; another option is to use the general procedure of Jiménez-Gamero *et al*. (2009) based the empirical characteristic function.

5.2.2 Fisher information matrix

Computation of the information matrix associated with the log-likelihood (5.53) is technically intricate. We only summarize the main facts, referring the reader to Arellano-Valle and Azzalini (2008) for a full treatment.

For mathematical convenience we consider again the parameterization, now denoted θ^{SP}, which replaces α in (5.52) with $\eta = \omega^{-1}\alpha$. Define the $d^2 \times [d(d+1)/2]$ duplication matrix D_d such that $\text{vec}(M) = D_d\,\text{vech}(M)$ for a symmetric matrix M, and let $Z_2 = -\text{diag}(\zeta_2(R_\beta\,\eta)) > 0$. Then it can be shown that

$$-\frac{\partial^2 \ell(\theta^{\text{SP}})}{\partial\theta^{\text{SP}}\,\partial(\theta^{\text{SP}})^\top} = \begin{pmatrix} \Omega^{-1}\otimes(X^\top X)+(\eta\eta^\top)\otimes(X^\top Z_2 X) & \cdot & \cdot \\ D_d^\top[\Omega^{-1}\otimes(\Omega^{-1}R_\beta^\top X)] & \frac{1}{2}n\,D_d^\top(\Omega^{-1}\otimes V)D_d & \cdot \\ I_d\otimes u^\top-\eta\otimes U^\top & 0 & R_\beta^\top Z_2 R_\beta \end{pmatrix} \tag{5.56}$$

where $u = X^\top\zeta_1(R_\beta\,\eta)$, $U = X^\top Z_2\,R_\beta$, $V = \Omega^{-1}(2S_\beta - \Omega)\,\Omega^{-1}$ and the upper triangle must be filled symmetrically. Evaluation of this matrix at the MLE, $\hat\theta^{\text{SP}}$, gives the observed information matrix, $\mathcal{J}(\hat\theta^{\text{SP}})$.

https://doi.org/10.1017/CBO9781139248891.006 Published online by Cambridge University Press

To compute the expected value of (5.56), consider $U \sim \mathrm{N}(0, \bar{\alpha}^2)$, where $\bar{\alpha}^2 = \alpha_*^2/(1+2\alpha_*^2)$, and define

$$a_0 = \mathbb{E}\left\{\Phi(U)^{-1}\right\}, \qquad a_1 = \mathbb{E}\left\{\Phi(U)^{-1} U \mu_c\right\},$$
$$A_2 = \mathbb{E}\left\{\Phi(U)^{-1}\left(U^2 \mu_c \mu_c^\top + \Omega_c\right)\right\},$$

where $\mu_c = \alpha_*^{-2}\Omega\eta = (\eta^\top\Omega\eta)^{-2}\Omega\eta$ and $\Omega_c = \Omega - \alpha_*^{-2}\Omega\eta\eta^\top\Omega$. Evaluation of these coefficients requires numerical integration, but only in the limited form of three 1-dimensional integrals, irrespective of d. The expected Fisher information matrix for θ^{SP} is then

$$\mathcal{I}(\theta^{\mathrm{SP}}) = \begin{pmatrix} (\Omega^{-1} + c_2\, a_0\, \eta\eta^\top) \otimes (X^\top X) & \cdot & \cdot \\ c_1 D_d^\top[\Omega^{-1} \otimes (\eta 1_n^\top X)] & \frac{1}{2} n\, D_d^\top(\Omega^{-1} \otimes \Omega^{-1}) D_d & \cdot \\ A_1^\top \otimes (1_n^\top X) & 0 & n\, c_2\, A_2 \end{pmatrix}, \tag{5.57}$$

where

$$c_k = (1 + k\,\eta^\top\Omega\eta)^{-1/2}\, 2\,(b/2)^k, \qquad A_1 = c_1(I_d + \eta\,\eta^\top\Omega)^{-1} - c_2\eta a_1^\top\,.$$

Conversion of either type of information matrix, $\mathcal{J}(\theta^{\mathrm{SP}})$ or $\mathcal{I}(\theta^{\mathrm{SP}})$, to its counterpart for θ^{DP} requires the Jacobian matrix of the partial derivatives of θ^{SP} with respect to θ^{DP}, that is,

$$D_{\theta^{\mathrm{DP}}}(\theta^{\mathrm{SP}}) = \begin{pmatrix} I_{pd} & 0 & 0 \\ 0 & I_{d(d+1)/2} & 0 \\ 0 & D_{32} & \omega^{-1} \end{pmatrix}$$

where

$$D_{32} = -\tfrac{1}{2}\sum_{i=1}^{d} (\Omega_{ii})^{-3/2} (\alpha^\top E_{ii} \otimes E_{ii}) D_d\,,$$

having denoted by E_{ii} the $d \times d$ matrix having 1 in the (i,i)th entry and 0 otherwise. Then the expected information matrix for θ^{DP} is

$$\mathcal{I}^{\mathrm{DP}}(\theta^{\mathrm{DP}}) = D_{\theta^{\mathrm{DP}}}(\theta)^\top \mathcal{I}(\theta^{\mathrm{SP}})\, D_{\theta^{\mathrm{DP}}}(\theta^{\mathrm{SP}}) \tag{5.58}$$

and a similar expression holds for $\mathcal{J}(\theta^{\mathrm{DP}})$.

5.2.3 The centred parameterization

We examine the multivariate extension of the centred parameterization discussed in § 3.1.4 for the univariate case. For simplicity of exposition, we refer to the case $p = 1$, hence with a constant location parameter ξ. This does not represent a restriction, since (3.23) indicates that only the first regression component changes moving between DP and CP.

https://doi.org/10.1017/CBO9781139248891.006 Published online by Cambridge University Press

A direct extension of the CP notion to the multivariate case is represented by (μ, Σ, γ_1), where the first two components are given by (5.31) and (5.32), respectively, and γ_1 is the d-vector of marginal coefficients of skewness obtained by component-wise application of (2.28) to each component of (5.11), via (2.26). Formally, only the non-replicated entries of Σ, namely $\mathrm{vech}(\Sigma)$, enter the parameter vector.

The above description says how to obtain CP from DP. As DP spans its feasible range, which is only restricted by the condition $\Omega > 0$, CP spans a corresponding set. An important difference is that, while the DP components are variation-independent, the same is not true for the CP components, if $d > 1$. However, if a certain parameter combination (μ, Σ, γ_1) belongs to the feasible CP parameter set, then there is a unique inverse point in the DP space. To see this, from the jth component of γ_1 obtain $\mu_{z,j}$ inverting (2.28) and from here δ_j, for $j = 1, \dots, d$. This gives a d-vector δ and a diagonal matrix σ_z whose jth non-zero entry is obtained from δ_j using the second expression of (2.26). After forming the diagonal matrix σ with the square-root of the diagonal entries of Σ, the first two DP components are given by

$$\xi = \mu + \sigma\sigma_z^{-1}\mu_z, \qquad \omega = \sigma\sigma_z^{-1}, \qquad \Omega = \Sigma + \omega\mu_z\mu_z^\top\omega$$

and α is as in (5.12). Therefore the CP, (μ, Σ, γ_1), represents a legitimate parameterization of the multivariate SN family.

One of the arguments in support of the CP in the univariate case was the simpler interpretation of mean, standard deviation and coefficient of skewness, compared to the corresponding component of the DP. This aspect holds *a fortiori* in the multivariate context, where the values taken on by the components of α are not easily interpretable. As a numerical illustration of this point, consider two sets of parameters, $(\Omega, \alpha^{(1)})$ and $(\Omega, \alpha^{(2)})$, where

$$\Omega = \begin{pmatrix} 2 & 1 & 3 \\ 1 & 2 & 4 \\ 3 & 4 & 9 \end{pmatrix}, \qquad \alpha^{(1)} = \begin{pmatrix} 5 \\ -3 \\ 4 \end{pmatrix}, \qquad \alpha^{(2)} = \begin{pmatrix} 5 \\ -3 \\ -4 \end{pmatrix}$$

whose corresponding coefficients of marginal skewness, rounded to two decimal digits, are

$$\gamma_1^{(1)} = (0.85,\ 0.04,\ 0.16)^\top, \qquad \gamma_1^{(2)} = (0.00,\ -0.21,\ -0.07)^\top,$$

respectively. Visibly, consideration of an individual component of α does not provide information on the corresponding component of γ_1, in fact not even on its sign, while in the univariate case there at least exists a monotonic relationship between α and γ_1.

https://doi.org/10.1017/CBO9781139248891.006 Published online by Cambridge University Press

For inferential purposes, ML estimates of the CP are simply obtained by transformation to the CP space of the ML estimates of the DP, by the equivariance property. The Fisher CP information matrix, $\mathcal{I}(\theta^{\text{CP}})$, is obtained from (5.57) by a transformation similar to (5.58) where the Jacobian matrix is now constituted by the partial derivatives of θ^{SP} with respect to θ^{CP}, which is also given by Arellano-Valle and Azzalini (2008). For mathematical convenience, an intermediate parameterization between θ^{SP} and θ^{CP} is introduced; consequently, this Jacobian matrix is expressed as the product of two such matrices.

Arellano-Valle and Azzalini (2008) have further considered the asymptotic behaviour of the resulting information matrix in the limiting case where $\gamma_1 \to 0$, or equivalently $\alpha \to 0$. While a limiting form of $\mathcal{I}(\theta^{\text{CP}})$ has been stated in the quoted paper, subsequent analysis has raised doubts on the correctness of this result, specifically on the diagonal block pertaining to γ_1, when $d > 1$. Further investigation on this issue is therefore required. If $d = 1$ the asymptotic expression is in agreement with the results of Chapter 3.

The previous passage prevents, at least currently, making use of the multivariate CP for inferential purposes in a neighbourhood of $\gamma_1 = 0$. Still, we feel like considering the usage of the CP in situations separate from $\gamma_1 = 0$, because the problematic aspects at one point do not prevent their use over the remaining parameter space, taking into account considerations on interpretability of the parameters discussed earlier.

5.3 Multivariate extended skew-normal distribution

5.3.1 *Definition and basic properties*

A d-dimensional version of the extended skew-normal distribution examined in §2.2 is given by

$$\varphi_d(x; \bar\Omega, \alpha, \tau) = \varphi_d(x; \bar\Omega)\, \frac{\Phi\{\alpha_0 + \alpha^\top x\}}{\Phi(\tau)}, \qquad x \in \mathbb{R}^d, \tag{5.59}$$

where $\tau \in \mathbb{R}$,

$$\alpha_0 = \tau(1 + \alpha^\top \bar\Omega \alpha)^{1/2} \tag{5.60}$$

and the other terms are as in (5.1). Using Lemma 5.2, it is straightforward to confirm that (5.59) integrates to 1. Similarly to the univariate case, τ effectively vanishes when $\alpha = 0$. A slightly different parameterization in use regards α_0 as a parameter component in place of τ, while here we shall use α_0 only as a short-hand notation for (5.60).

https://doi.org/10.1017/CBO9781139248891.006 Published online by Cambridge University Press

If Z has density (5.59) and $Y = \xi + \omega Z$ as in (5.2), the density of Y at $x \in \mathbb{R}^d$ is

$$\varphi_d(x - \xi; \Omega)\, \Phi\left\{\alpha_0 + \alpha^\top \omega^{-1}(x - \xi)\right\}\, \Phi(\tau)^{-1} \tag{5.61}$$

with the same notation of (5.3). In this case, we write $Y \sim \mathrm{SN}_d(\xi, \Omega, \alpha, \tau)$, where again the presence of the fourth parameter component indicates that the distribution is ‘extended’.

Using Lemma 5.3, the moment generating function of the distribution $Y \sim \mathrm{SN}_d(\xi, \Omega, \alpha, \tau)$ is readily seen to be

$$M(t) = \exp(t^\top \xi + \tfrac{1}{2} t^\top \Omega t)\, \Phi(\tau + \delta^\top \omega t)\, \Phi(\tau)^{-1}, \qquad t \in \mathbb{R}^d, \tag{5.62}$$

where δ is as in (5.11).

From $M(t)$, which matches closely (5.10) of the SN case, we can derive the distribution for marginal block components and for affine transformations of Y. Specifically, if Y is partitioned as $Y^\top = (Y_1^\top, Y_2^\top)$ where the two blocks have size h and $d{-}h$, as in § 5.1.4, then marginally

$$Y_1 \sim \mathrm{SN}_h(\xi_1, \Omega_{11}, \alpha_{1(2)}, \tau), \tag{5.63}$$

where the first three parameter components are the same as the SN case given by (5.29). For an affine transformation $X = c + A^\top Y$, where A is a full-rank $d \times h$ matrix ($h \le d$) and $c \in \mathbb{R}^h$, we have

$$X \sim \mathrm{SN}_h(\xi_X, \Omega_X, \alpha_X, \tau),$$

where the first three parameter components are given by (5.41)–(5.43).

Similarly to its univariate counterpart, density (5.59) does not satisfy the conditions for the property of modulation invariance (1.12). Hence the results of § 5.1.6 on quadratic forms of SN variates do not carry on here.

A mathematically appealing aspect of this distribution is first suggested by the observation that, if $X \sim \mathrm{SN}_d(\xi, \Omega, \alpha)$, then the conditional density of X given that a subset of its components takes on a certain value is of type (5.61); see Problem 5.12. This property is a simplified version of the closure property of the next paragraph.

5.3.2 Conditional distribution and conditional independence

An important property of the family (5.61) is its closure with respect to conditioning on the values taken on by some components. To see this, partition $Y \sim \mathrm{SN}_d(\xi, \Omega, \alpha, \tau)$ as $Y = (Y_1^\top, Y_2^\top)^\top$, where Y_1 has dimension h, and examine the conditional distribution of Y_2 given that $Y_1 = y_1$. Recall that,

https://doi.org/10.1017/CBO9781139248891.006 Published online by Cambridge University Press

if Y was a $\mathrm{N}_d(\xi, \Omega)$ variable, the parameters of the conditional distribution would be

$$\xi_{2\cdot1} = \xi_2 + \Omega_{21}\Omega_{11}^{-1}(y_1 - \xi_1), \qquad \Omega_{22\cdot1} = \Omega_{22} - \Omega_{21}\Omega_{11}^{-1}\Omega_{12} \tag{5.64}$$

and these quantities emerge again when we take the ratio of the normal densities involved by $(Y_2|Y_1 = y_1)$. Then, using (5.63), the conditional density of Y_2 given $Y_1 = y_1$ is

$$\varphi_{d-h}(y_2 - \xi_{2\cdot1}; \Omega_{22\cdot1}) \, \frac{\Phi\left\{\alpha_0' + \alpha_2^\top \omega_2^{-1}(y_2 - \xi_{2\cdot1})\right\}}{\Phi(\tau_{2\cdot1})}, \qquad y_2 \in \mathbb{R}^{d-h}, \tag{5.65}$$

where

$$\begin{aligned} \tau_{2\cdot1} &= \tau\left(1 + \alpha_{1(2)}^\top \bar{\Omega}_{11}\, \alpha_{1(2)}\right)^{1/2} + \alpha_{1(2)}^\top\, \omega_1^{-1}(y_1 - \xi_1)\,, \\ \alpha_0' &= \tau_{2\cdot1}\,(1 + \alpha_{2\cdot1}^\top\, \bar{\Omega}_{22\cdot1}^{-1}\, \alpha_{2\cdot1})^{1/2}\,, \\ \alpha_{2\cdot1} &= \omega_{22\cdot1}\, \omega_2^{-1}\, \alpha_2\,, \\ \omega_{22\cdot1} &= (\Omega_{22\cdot1} \odot I_{d-h})^{1/2} \end{aligned} \tag{5.66}$$

and we have used the notation in (5.27) and (5.28) on p. 130. To conclude, write

$$(Y_2|Y_1 = y_1) \sim \mathrm{SN}_{d-h}(\xi_{2\cdot1}, \Omega_{22\cdot1}, \alpha_{2\cdot1}, \tau_{2\cdot1}) \tag{5.67}$$

which states the property of closure with respect to conditioning.

The above expression of $\alpha_{2\cdot1}$ provides the key to interpret the presence of null components of α. Since $\alpha_{2\cdot1} = 0$ if and only if $\alpha_2 = 0$, then $\alpha_2 = 0$ means that $(Y_2|Y_1 = y_1)$ is Gaussian. Consequently, when the rth component of α is null, the conditional distribution of Y_r given all other components is Gaussian. These facts hold both in the ESN and in the SN case, since (5.67) holds also when $\tau = 0$, with a simplification in $\tau_{2\cdot1}$.

This type of argument can be carried on to examine conditional independence among components of the distribution of $(Y_2|Y_1 = y_1)$. Specifically, bearing in mind the relationship between $\alpha_{2\cdot1}$ and α_2 as given in (5.66) and that $\Omega_{22\cdot1} = (\Omega^{-1})_{22}$, conditions for conditional independence can be stated directly as conditions on α and Ω^{-1}. This fact is exploited to obtain the next result.

Proposition 5.15 *Consider any three-block partition $Y = (Y_1^\top, Y_{2a}^\top, Y_{2b}^\top)^\top$ of $Y \sim \mathrm{SN}_d(\xi, \Omega, \alpha, \tau)$. Then Y_{2a} and Y_{2b} are conditionally independent given Y_1 if and only if the following conditions hold simultaneously:*

(a) $(\Omega^{-1})_{ab} = 0$,
(b) at least one of α_a and α_b is the null vector,

where α_a and α_b denote the subsets of α associated with Y_{2a} and Y_{2b}, respectively, and $(\Omega^{-1})_{ab}$ is the corresponding block portion of Ω^{-1}.

Proof Since the value of τ does not affect the conditional independence among the components of $Y_2 = (Y_{2a}^\top, Y_{2b}^\top)^\top$, we can argue as if $\tau = 0$. Then the statement can be proved recalling that independence requires that the parameters of the conditional distribution must have the structure as in (5.46). In the present case, that structure holds for $h = 2$, the pertaining scale matrix is $\Omega_{22\cdot 1}$, that is, the scale matrix of the conditional distribution given Y_1, and the slant parameter $\alpha_{2\cdot 1}$ is computed from (5.66). QED

The property of closure under conditioning and the last proposition form the basis for developing graphical models of ESN variables. Some results in this direction will be presented in § 5.3.5.

5.3.3 Stochastic representations and distribution function

Some stochastic representations of the multivariate SN distribution extend naturally to the ESN case; others do not, or at least no such extension is known at the time of writing.

A stochastic representation via a conditioning mechanism is as follows. Starting from (X_0, X_1) distributed as in (5.14), a standard computation says that, for any $\tau \in \mathbb{R}$,

$$Z = (X_0 | X_1 + \tau > 0) \sim \mathrm{SN}_d(0, \bar{\Omega}, \alpha(\delta), \tau) \tag{5.68}$$

where $\alpha(\delta)$ is given by (5.12), similarly to the first expression in (5.15).

Representation (5.68) indicates how to compute the distribution function of Z. By a computation similar to (2.48), write

$$\begin{aligned}\mathbb{P}\{Z \le z\} &= \mathbb{P}\{X_0 \le z | X_1 + \tau > 0\}\\ &= \mathbb{P}\{(X_0 \le z) \cap (-X_1 < \tau)\} / \mathbb{P}\{-X_1 < \tau\}\\ &= \Phi_{d+1}((z^\top, \tau)^\top; \tilde{\Omega})/\Phi(\tau),\end{aligned} \tag{5.69}$$

where $\tilde{\Omega}$ is a matrix similar to Ω^* in (5.14) with δ replaced by $-\delta$. The general case $\mathrm{SN}_d(\xi, \Omega, \alpha, \tau)$ is handled as usual by reduction to a normalized variable Z. Therefore, the distribution function of a d-dimensional ESN, and then also of an SN, variable is computed by evaluating a suitable $(d+1)$-dimensional normal distribution function.

To introduce a form of additive representation of an ESN variate, start from the independent variables $U_0 \sim \mathrm{N}_d(0, \bar{\Psi})$, where $\bar{\Psi}$ is a full-rank correlation matrix, and $U_{1,-\tau}$ which is a $\mathrm{N}(0, 1)$ variable truncated below $-\tau$

https://doi.org/10.1017/CBO9781139248891.006 Published online by Cambridge University Press

for some $\tau \in \mathbb{R}$. Then a direct extension of (2.43), using the notation of (5.18)–(5.19), is

$$Z = D_\delta\, U_0 + \delta\, U_{1,-\tau} \tag{5.70}$$

such that $Z \sim \mathrm{SN}_d(0, \bar\Omega, \alpha, \tau)$ where $\bar\Omega$ and α are related to $\bar\Psi$ and δ as in (5.20)–(5.22); see Problem 5.13.

For the reasons discussed in § 2.2.2 for the univariate case, representation (5.70) is more convenient than (5.68) for random number generation.

5.3.4 Cumulants and related quantities

From (5.62) the cumulant generating function of $Y \sim \mathrm{SN}_d(\xi, \Omega, \alpha, \tau)$ is

$$K(t) = \log M(t) = \xi^\top t + \tfrac{1}{2} t^\top \Omega t + \zeta_0(\tau + \delta^\top \omega t) - \zeta_0(\tau), \qquad t \in \mathbb{R}^d,$$

where $\zeta_0(x)$ is defined by (2.18) along with its successive derivatives $\zeta_r(x)$. Evaluation at $t = 0$ of the first two derivatives of $K(t)$ leads to

$$\mathbb{E}\{Y\} = \xi + \zeta_1(\tau)\, \omega\, \delta = \xi + \omega\, \mu_z\,, \tag{5.71}$$

$$\mathrm{var}\{Y\} = \Omega + \zeta_2(\tau)\, \omega\, \delta\, \delta^\top \omega = \omega\, \Sigma_z\, \omega, \tag{5.72}$$

where

$$\mu_z = \mathbb{E}\{Z\} = \zeta_1(\tau)\, \delta, \qquad \Sigma_z = \mathrm{var}\{Z\} = \bar\Omega + \zeta_2(\tau)\, \delta\, \delta^\top$$

refer to $Z \sim \mathrm{SN}_d(0, \bar\Omega, \alpha, \tau)$. Higher-order derivatives of $K(t)$ are

$$\frac{\mathrm{d}^r}{\mathrm{d}t_i\ \mathrm{d}t_j\ \cdots\ \mathrm{d}t_h} K(t) = \zeta_r(\tau + \delta^\top \omega\, t)\, \omega_i\, \omega_j \cdots \omega_h\, \delta_i\, \delta_j \cdots \delta_h\,. \tag{5.73}$$

Proceeding similarly to § 5.1.5, we obtain that the Mardia coefficients of multivariate skewness and kurtosis are

$$\gamma^M_{1,d} = \left(\frac{\zeta_3(\tau)}{\zeta_1(\tau)^3}\right)^2 \left(\mu_z^\top \Sigma_z^{-1} \mu_z\right)^3 = \zeta_3(\tau)^2 \left(\frac{\delta_*^2}{1 + \zeta_2(\tau)\, \delta_*^2}\right)^3, \tag{5.74}$$

$$\gamma^M_{2,d} = \frac{\zeta_4(\tau)}{\zeta_1(\tau)^4} \left(\mu_z^\top \Sigma_z^{-1} \mu_z\right)^2 = \zeta_4(\tau) \left(\frac{\delta_*^2}{1 + \zeta_2(\tau)\, \delta_*^2}\right)^2, \tag{5.75}$$

where δ_* is as in (5.38). The two final expressions match those in (2.46) and (2.47) evaluated at δ_*, except that the Mardia coefficient of skewness when $d = 1$ corresponds to the square of the univariate coefficient. Therefore the range of $(\gamma^M_{1,d}, \gamma^M_{2,d})$ is the same as pictured in Figure 2.5 provided the γ_1-axis is square transformed.

https://doi.org/10.1017/CBO9781139248891.006 Published online by Cambridge University Press

5.3.5 Conditional independence graphs

The aim of this section is to present some introductory notions on *graphical models* for ESN variables, specifically in the form of conditional independence graphs. For background material on graphical models, we refer the reader to the monographs of Cox and Wermuth (1996) and Lauritzen (1996).

A graphical model is constituted by a graph, denoted $\mathscr{G} = (V, E)$, where the set V of the vertices or nodes is formed by the components of a multivariate random variable $Y = (Y_1, \ldots, Y_d)^\top$ and the set E of edges connecting elements of V is chosen to represent the dependence structure induced by the distribution of Y.

A conditional independence graph is a construction with the additional requirements that (a) the graph is *undirected*, which means that an edge is a set of two unordered elements of V, so that we do not make distinction among (i, j), (j, i) and $\{i, j\}$, and (b) the nodes i and j are not connected if Y_i and Y_j are conditionally independent given all other components of Y, for $i \neq j$. Formally we write $\{i, j\} \notin E$ if $Y_i \perp\!\!\!\perp Y_j |$(all other variables), where the symbol $\perp\!\!\!\perp$ denotes independence.

We now explore the above concepts when Y has a multivariate ESN distribution. The focus is on this family because closure with respect to conditioning plays a fundamental role here. From Proposition 5.15 it is immediate to state the following result.

Corollary 5.16 (Pairwise conditional independence) *If* $Y = (Y_1, \ldots, Y_d)^\top \sim \mathrm{SN}_d(\xi, \Omega, \alpha, \tau)$*, then*

$$Y_i \perp\!\!\!\perp Y_j | \textit{(all other variables)}$$

if and only if the following conditions hold simultaneously:

(a) $\Omega^{ij} = 0$*,*
(b) $\alpha_i \alpha_j = 0$,

where Ω^{ij} *denotes the* (i, j)*th entry of* Ω^{-1}.

This statement lends the operational rule to specify the conditional independence graph associated with Y:

$$(i, j) \in E \quad \Longleftrightarrow \quad \{\Omega^{ij} \neq 0 \quad \text{or} \quad \alpha_i \alpha_j \neq 0\}\,. \tag{5.76}$$

When $\alpha = 0$, we recover the classical rule for the Gaussian case based solely on the elements of the concentration matrix, that is, the inverse of the variance matrix.

https://doi.org/10.1017/CBO9781139248891.006 Published online by Cambridge University Press

So far, the graph built via (5.76) reflects the conditional independence for a pair of variables, but we are interested in establishing all conditional independence statements implied by this structure. This extension is possible thanks to the *global Markov property*, which applies to continuous random variables with density positive everywhere on the support, such as the ESN family; see the monographs cited earlier for a detailed discussion of these aspects. In essence, the global Markov property can be described as follows: if A, B and C are disjoint subsets of vertices and C separates A from B, then conditional independence $Y_A \perp\!\!\!\perp Y_B | Y_C$ holds for the corresponding set of variables, Y_A, Y_B, Y_C. Recall that C *separates* A from B if there is no sequence of edges connecting a node in A with a node in B without going through some node in C.

Clearly, for a given pair (Ω, α), the corresponding conditional independence graph is uniquely specified by (5.76). The converse is not true: a given graph is compatible with several patterns of (Ω, α). For instance a *complete* graph, where an edge exists between any pair of distinct vertices, can be obtained both from the pair $(I_d, a\, 1_d)$ where $a \neq 0$ and from the pair (Ω, α) where Ω^{-1} has no zero entries and α is arbitrary.

Stochastic representation (5.68) of Y indicates how this variable is related to a suitable $(d+1)$-dimensional normal variable X, as specified in (5.14). The next proposition indicates how the respective conditional independence graphs are related.

Proposition 5.17 *Given the conditional independence graph $\mathscr{G}_X$ of X with distribution (5.14), the conditional independence graph $\mathscr{G}_Z$ of Z, defined in (5.68), is uniquely identified and can be obtained by adding those edges needed to make the boundary of the vertex associated with X_1 complete and by deleting this vertex and the corresponding edges.*

Proof By making use of (5.12), the concentration matrix of $X = (X_0^\top, X_1)^\top$ can be written as

$$(\Omega^*)^{-1} = \begin{pmatrix} A & -\alpha\, c \\ -\alpha^\top c & c^2 \end{pmatrix}, \tag{5.77}$$

where $A = \bar\Omega^{-1} + \alpha\alpha^\top$ and $c = (1 - \delta_*^2)^{-1/2} > 0$ with δ_* defined by (5.38). If $A_{ij} \neq 0$, so that the edge (i, j) exists in $\mathscr{G}_X$, then this edge will also exist in $\mathscr{G}_Z$, from (5.76). If $A_{ij} = 0$ and $\alpha_i\alpha_j \neq 0$, then $\bar\Omega^{ij} \neq 0$. Hence we must add an edge (i, j) if vertex X_1 is connected to both i and j. QED

For simplicity of notation, Proposition 5.17 has been stated for the case of a normalized variable Z with zero location and unit scale factors, but it holds for the general case as well.

https://doi.org/10.1017/CBO9781139248891.006 Published online by Cambridge University Press

We now examine the conditions for separation when Y has an ESN distribution. In this case the possible presence of nodes with marginal Gaussian distribution introduces constraints on the structure of conditional dependence, so that some patterns are inhibited. Moreover, the existence of Gaussian nodes may provide an indication of the presence of 0 elements in α. To distinguish the two types of nodes, we mark the nodes having Gaussian marginal distribution with ‘G’, and the others with ‘SN’, dropping the ‘E’ of ESN for mere simplicity of notation. Correspondingly, V is partitioned into two disjoint sets, V_G and V_SN. The boundary set of vertex i formed by all vertices which share an edge with i is denoted by $\mathrm{bd}(i)$.

Proposition 5.18 *Consider the three-block partition $Y^\top = (Y_A^\top, Y_B^\top, Y_C^\top)$ where A, B and C are disjoint subsets of indices and $Y \sim \mathrm{SN}_d(\xi, \Omega, \alpha, \tau)$. If C separates A from B, one among the three following conditions must hold:*

(a) $A \cup C \subseteq V_\mathrm{G}$,
(b) $B \cup C \subseteq V_\mathrm{G}$,
(c) $C \nsubseteq V_\mathrm{G}$.

Proof Recall Corollary 5.7 which clearly holds also for ESN distributions. Since Y obeys the global Markov property, the fact that C separates A from B corresponds to the independence relationship $Y_A \perp\!\!\!\perp Y_B | Y_C$. Then Corollary 5.7 implies that $\bar\Omega^{AB} = 0$ and at least one of α_A and α_B is the null vector. Therefore, from (5.12), at least one of the two following equalities must hold:

$$\alpha_A = k\,(\bar\Omega^{AA}\delta_A + \bar\Omega^{AC}\delta_C) = 0, \qquad \alpha_B = k\,(\bar\Omega^{BB}\delta_B + \bar\Omega^{BC}\delta_C) = 0,$$

in an obvious notation, for some $k > 0$. Conditions (a) and (b) then follow because both $\bar\Omega^{AA} > 0$ and $\bar\Omega^{BB} > 0$. If both (a) and (b) fail, separation can only occur under (c). QED

Corollary 5.19 *Let (A, B, C) be a partition of V such that $A \cup C \subseteq V_\mathrm{G}$. If C separates A from B, then $\alpha_A = 0$.*

Proposition 5.20 *If $i \in V_\mathrm{G}$ and $\mathrm{bd}(i) \cap V_\mathrm{SN} = \{h\}$ [i.e., $\mathrm{bd}(i)$ has only one vertex in V_SN], then $\alpha_i \neq 0$.*

Proof Let h be the unique non-Gaussian vertex in $\mathrm{bd}(i)$. Then from (5.12) we have $\alpha_i \propto \Omega^{ih}\delta_h$. Since $\delta_h \neq 0$, it follows that $\alpha_i = 0$ if and only if $\Omega^{ih} = 0$, implying $(i, h) \notin E$. QED

https://doi.org/10.1017/CBO9781139248891.006 Published online by Cambridge University Press

Corollary 5.21 *If $i, j \subseteq V_{\mathrm{G}}$ and both* $\mathrm{bd}(i)$ *and* $\mathrm{bd}(j)$ *have exactly one vertex in V_{SN}, then $(i, j) \in E$.*

Proof Immediate from Propositions 5.18 and 5.20. QED

Operationally, these statements allow us to define two rules for checking the admissibility of a marked graph, with vertices labelled G or SN: (a) in any three-set partition of a marked graph, a subset of G vertices cannot separate two subsets each containing some SN vertices; (b) in a marked graph, there cannot exist two not connected G vertices having in their boundary sets exactly one SN vertex. From here, in some cases, we can identify which are the non-zero components of α.

The importance of identifying, for a given graph, which are the null elements of α and of Ω^{-1} lies in the possibility of using this information in the estimation stage. We have in mind the case where a marked conditional independence graph, associated with a certain applied problem, has been specified on the grounds of subject-matter considerations. For all pairs of vertices $\{i, j\}$ where an edge is missing, we know that $\Omega^{ij} = 0$ and at least one of α_i and α_j is zero. The use of this information in conjunction with the results established above can lead to a quite specific identification of the parameter structure; this process is exemplified in the next paragraph. The possibility of transferring the structure of the graph into constraints on the null elements of the parameter estimates can improve appreciably the estimation problem, avoiding the scan of a large set of compatible parameter patterns, and reducing variability of the estimates.

For an illustration, consider the graph in Figure 5.5, where the nature of vertex 5 is not yet specified. If we set $5 \in V_{\mathrm{G}}$, from Corollary 5.21 we conclude that the graph is not admissible since the G nodes 2 and 5 would have on their boundary a single vertex belonging to V_{SN}, but they are not connected to each other. If we set $5 \in V_{\mathrm{SN}}$, the graph becomes admissible. Since

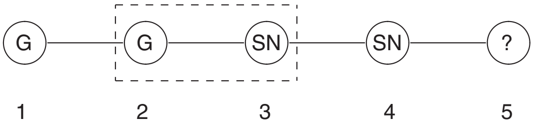


Figure 5.5 An example of a marked graph, where the labels G and SN denote Gaussian and extended skew-normal nodes, respectively, and the nature of the node marked '?' is discussed in the text. The dashed box indicates the nodes with possibly non-null α's in the joint 5-dimensional distribution.

https://doi.org/10.1017/CBO9781139248891.006 Published online by Cambridge University Press

$2 \in V_G$ separates $1 \in V_G$ from $\{3, 4, 5\} \subseteq V_{SN}$, condition (a) of Proposition 5.18 holds. The fact $Y_1 \perp\!\!\!\perp Y_{\{3,4,5\}}|Y_2$ is compatible both with $\alpha_1 = 0$ and with $\alpha_{\{3,4,5\}} = (0, 0, 0)^\top$. Corollary 5.19 indicates that we must have $\alpha_1 = 0$, and Proposition 5.20 implies that $\alpha_2 \neq 0$. Finally, the facts $Y_2 \perp\!\!\!\perp Y_5|Y_{\{1,3,4\}}$ and $Y_2 \perp\!\!\!\perp Y_4|Y_{\{1,3,5\}}$ lead us to say that the non-zero components of α can only be α_2 and α_3. Of these, we have established that α_2 is non-zero, while α_3 may be 0 or not.

5.3.6 Bibliographical notes

The multivariate ESN distribution has been studied by Adcock and Shutes (1999), Arnold and Beaver (2000a, Section 4) and Capitanio *et al.* (2003, Section 2 and Appendix). The first of these was motivated by application problems to quantitative finance, the main facts of which we shall recall in the next subsection. The other two papers present expressions for basic properties, such as the marginal and the conditional distributions, the moment generating function and lower-order moments, with inessential differences in the parameterization. Capitanio *et al.* (2003) also give expressions for the distribution function, the general expression of the cumulants and Mardia's coefficients.

The main target of Capitanio *et al.* (2003) is the development of a formulation for graphical models, of which § 5.3.5 represents an excerpt. Among the aspects not summarized here, this paper provides results for a parameter-based factorization of the likelihood function, which can simplify substantially complex estimation problems. Work on related graphical models has been presented by Stanghellini and Wermuth (2005). Capitanio and Pacillo (2008) propose a Wald-type test for the inclusion/exclusion of a single edge, and Pacillo (2012) explores the issue further.

5.3.7 Some applications

In quantitative finance, much work is developed under the assumption of multivariate normality, for convenience reasons. While it is generally agreed that normality is unrealistic, use of alternatives is often hampered by the lack of mathematical tractability. Adcock and Shutes (1999) have shown that various operations can be transferred quite naturally from the classical context of multivariate normal distribution to the ESN. They work with a parameterization which essentially is as in (5.30), with the introduction of an additional parameter, which leads to the ESN distribution. They obtain the moment generating function, lower-order moments and other basic properties. These results are employed to reconsider some classical

https://doi.org/10.1017/CBO9781139248891.006 Published online by Cambridge University Press

optimality problems in finance within this broader context. As a specific instance, denote by R a vector of d asset returns, and examine the problem of optimal allocation of weights w among these assets, under the expected utility function

$$\psi(w) = 1 - \mathbb{E}\{\exp(-w^\top R/\theta)\},$$

where $\theta > 0$ is a parameter which expresses the risk appetite of the investor. If we assume that R has joint ESN distribution, then maximization of $\psi(w)$ corresponds to minimization of the moment generating function of type (5.62) evaluated at $t = -w/\theta$, more conveniently so after logarithmic transformation. The problem allows a simple treatment even in the presence of linear inequality constraints. The authors also deal with analogues of efficient frontier and market model. See also Adcock (2004) for closely related work and some empirical illustrations.

Carmichael and Coën (2013) formulate a model for asset pricing where the log-returns are jointly multivariate skew-normal and the stochastic discount factor is a polynomial transform of a reference component of them. The ensuring construction is sufficiently tractable for the authors to obtain analytic expressions for various quantities of interest and this ‘sheds a new light on financial puzzles as the equity premium puzzle, the riskfree rate puzzle and could also be promising to deal with other well known financial anomalies’ (Section 4).

Similarly to finance, in various other application areas the assumption of multivariate normality is often made for convenience and the SN or ESN distribution can be adopted as a more realistic and still tractable model. A case in point is represented by the work of Vernic (2006) in the context of insurance problems. For the evaluation of risk exposure, she considers the ‘conditional tail expectation’ (TCE), regarded as preferable to the more common indicator represented by value at risk. The TCE is defined for a random variable X as

$$\mathrm{TCE}_X(x_q) = \mathbb{E}\left\{X | X > x_q\right\}, \qquad x_q \in \mathbb{R},$$

which is much the same concept of mean residual life used in other areas.

For an ESN variable $Z \sim \mathrm{SN}(0, 1, \alpha, \tau)$, the TCE function can easily be computed via integration by parts lending, in the notation of §2.2,

$$\begin{aligned}\mathrm{TCE}_Z(z_q) &= \frac{1}{1 - \Phi(z_q; \alpha, \tau)} \int_{z_q}^{\infty} z\, \varphi(z; \alpha, \tau)\, \mathrm{d}z \\ &= \frac{1}{1 - \Phi(z_q; \alpha, \tau)} \left[\varphi(z_q; \alpha, \tau) + \delta\zeta_1(\tau)\, \Phi\left(-\sqrt{1+\alpha^2}(z_q + \delta\tau)\right)\right]\end{aligned}$$

and the more general case $Y \sim \mathrm{SN}(\xi, \omega^2, \alpha, \tau)$ is handled by the simple connection $\mathrm{TCE}_Y(y_q) = \xi + \omega\,\mathrm{TCE}_Z(z_q)$, where $z_q = \omega^{-1}(y_q - \xi)$.

For the purpose of optimal capital allocation in the presence of random losses $Y = (Y_1, \ldots, Y_d)^\top$, it is of interest to compute $\mathbb{E}\left\{Y_i | S > s_q\right\}$ where S is the total loss $S = \sum_i Y_i$ or more generally a linear combination of type $S = w^\top Y$. Under a multivariate ESN assumption for Y, Vernic (2006) shows how this computation can be performed in explicit form, leading to the TCE formula for capital allocation. The author also considers another allocation formula with respect to an alternative optimality criterion.

The multivariate SN distribution has been employed in a range of other application areas. Early usage of the bivariate SN distribution for data fitting includes the works of Chu *et al.* (2001) as a model for random effects in the analysis of some pharmacokinetics data and of Van Oost *et al.* (2003) in a study of soil redistribution by tillage. Many more applications have followed, however, often quite elaborate. Since they generally feature other modelling aspects or they intersect with the use of related distributions, these other contributions will be recalled at various places later on, many of them in §8.2 but also elsewhere.

5.4 Complements

Complement 5.1 (Canonical form and scatter matrices) The construction of the canonical form $Z^* = H^\top(Y - \xi)$ of $Y \sim \mathrm{SN}_d(\xi, \Omega, \alpha)$ in Proposition 5.13 involves implicitly the simultaneous diagonalization of Ω and $\Sigma = \mathrm{var}\{Y\}$ to obtain matrix H. To see this, consider the equations

$$\Sigma\, h_j = \rho_j\, \Omega\, h_j, \qquad j = 1, \ldots, d, \tag{5.78}$$

where $h_j \in \mathbb{R}^d$ and $\rho_j \in \mathbb{R}$. The solution of the jth equation is obtained when ρ_j and h_j are an eigenvalue and the corresponding eigenvector of $\Omega^{-1}\Sigma$. Since this matrix is similar to matrix M appearing in Proposition 5.13, it easily follows that h_j constitutes the jth column of H.

This reading of the canonical form establishes a bridge with the results of Tyler *et al.* (2009) based on the simultaneous diagonalization of two scatter matrices. Recall that, given a d-dimensional random variable X, a matrix-valued functional $V(X)$ is a scatter matrix if it is positive definite, symmetric and satisfies the property $V(b+A^\top X) = A^\top V(X)\, A$ for any vector $b \in \mathbb{R}^d$ and any non-singular $d \times d$ matrix A. The authors show how, from the diagonalization of two scatter matrices, information about the properties of a model can be established, as the vectors h_j identify important

https://doi.org/10.1017/CBO9781139248891.006 Published online by Cambridge University Press

directions for inspecting data and they form an invariant coordinate system which, in the authors' words, 'can be viewed as a projection pursuit without the pursuit effort'. Also the ρ_j's provide information about the model; for instance, for an elliptical distribution they are all equal to each other. In our case, Ω and Σ represent two such scatter matrices.

Complement 5.2 (Regions of given probability) For a skew-normal variable Z with specified parameter values, we examine the problem of finding the region $R_{\rm SN} \subset \mathbb{R}^d$ of smallest geometrical size such that $\mathbb{P}\{Z \in R_{\rm SN}\} = p$, for any given value $p \in (0, 1)$. First of all, notice that the problem is location and scale equivariant, so that it can be reduced to the case $Z \sim {\rm SN}_d(0, \bar\Omega, \alpha)$ where $\bar\Omega$ is a correlation matrix. Secondly, it is immediate to state that the solution must be of type

$$R_{\rm SN} = \{x : \varphi_d(x; \bar\Omega, \alpha) \geq f_0\},$$

where f_0 is a suitable value which ensures that $\mathbb{P}\{Z \in R_{\rm SN}\} = p$. The question is how to find f_0. Log-concavity of the SN density implies that $R_{\rm SN}$ is a convex set.

The analogous problem for a normal variable $X \sim {\rm N}_d(0, \Sigma)$ has a neat solution represented by the ellipsoid

$$\begin{aligned} R_{\rm N} &= \{x : x^\top \Sigma^{-1} x \leq c_p\} \\ &= \{x : 2 \log \varphi_d(x; \Sigma) \geq -c_p - d \log 2\pi - d \log \det(\Sigma)\}, \end{aligned}$$

where c_p is the pth quantile of χ^2_d, on recalling that $X^\top \Sigma^{-1} X \sim \chi^2_d$. The region $R_{\rm N}$, with Σ replaced by $\bar\Omega$, provides a region of probability p also for Z, since the χ^2_d distribution is preserved, but in the SN case it does not represent the region of minimum geometrical size.

An exact expression of f_0 does not seem feasible, and an approximation must be considered. What follows summarizes the proposal of Azzalini (2001). As a first formulation, rewrite $R_{\rm N}$ replacing the normal density with $\varphi_d(x; \bar\Omega, \alpha)$, that is, consider the set

$$\tilde R_{\rm SN} = \{x : 2 \log \varphi_d(x; \bar\Omega, \alpha) \geq -c_p - d \log(2\pi) - \log \det(\bar\Omega)\} \qquad (5.79)$$

and let $\tilde p = \mathbb{P}\left\{Z \in \tilde R_{\rm SN}\right\}$.

To ease exposition, in the following we focus on the case $d = 2$, so that in (5.79) we have $c_p = -2 \log(1 - p)$ and $\det(\bar\Omega) = 1 - \bar\omega^2_{12}$. Evaluation of $\tilde p$ can be performed via simulation methods, for any given choice of the parameter set. For a range of values from $p = 0.01$ to $p = 0.99$, say, the corresponding values $\tilde p$ can be estimated by the relative frequencies of $\tilde R_{\rm SN}$

https://doi.org/10.1017/CBO9781139248891.006 Published online by Cambridge University Press

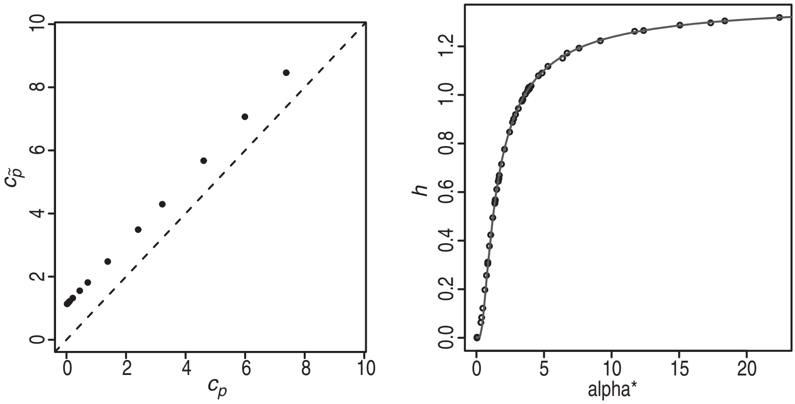


Figure 5.6 Construction of regions with given probability of SN distribution. Left plot: $(c_p, c_{\tilde{p}})$ for a set of p values when the distribution is bivariate skew-normal with parameters given in the text. Right plot: points (α_*, h) for a set of parameter combinations and interpolating curve.

in a set of sampled values. The left plot of Figure 5.6 refers to a simulation of 10^6 values sampled with $\bar{\omega}_{12} = -0.5$ and $\alpha = (2, 6)^\top$; after converting the $\tilde{p}$'s to the corresponding χ^2_2 quantiles, $c_{\tilde{p}}$'s say, these have been plotted versus c_p. While the plotted points do not lie on the ideal identity line, they are almost perfectly aligned along a line essentially parallel to the identity line, with only a very slight upturn when c_p is close to 0. Hence with good approximation we can write $c_{\tilde{p}} = c_p + h$ for some fixed h.

The pattern described above for that specific parameter set has been observed almost identically in a range of other cases, with different parameters. Invariably, the plotted points were aligned along the line $c_{\tilde{p}} = c_p + h$, where h varied with $\bar{\Omega}$ and α. Another interesting empirical indication is that h depends on the parameters only via α_* defined by (5.37). This is visible in the right panel of Figure 5.6, which plots a set of values of h, for several parameter combinations, versus α_*; the interpolating line will be described below. Therefore, a revised version of the approximate set is

$$\hat{R}_{\text{SN}} = \{x : 2 \log \varphi_d(x; \bar{\Omega}, \alpha) \geq -c_p + h - d \log(2\pi) - \log \det(\bar{\Omega})\} \quad (5.80)$$

where h is a suitable function of α_*.

Some more numerical work shows that a good approximation to h is provided by $h = 2 \log\{1 + \exp(-k_2/\alpha_*)\}$ where $k_2 = 1.544$, and this corresponds to the solid line in the right plot of Figure 5.6. This curve visibly interpolates the points satisfactorily, with only a little discrepancy near the origin.

As a check of the validity of the revised formulation, we evaluate $\hat{p} = \mathbb{P}\left\{Z \in \hat{R}_{\text{SN}}\right\}$ with the same method described for $\tilde{p}$. The numerical outcome for the earlier case with $\bar{\omega}_{12} = -0.5$ and $\alpha = (2, 6)^\top$ is summarized in the following table:

p	0.01	0.05	0.3	0.5	0.8	0.95	0.99
$\hat{p}$	0.043	0.077	0.306	0.500	0.797	0.949	0.990

There is a satisfactory agreement between p and $\hat{p}$ for moderate and large p, which are the cases of main practical interest. A similar agreement between p and $\hat{p}$ has been observed with other parameter combinations.

The contour lines in Figures 5.2 and 5.3 have been chosen using this method, followed by suitable location and scale transformations. Hence, for instance, the region delimited by the curve labelled $p = 0.9$ has this probability, up to the described approximation, and has minimal area among the regions with probability 0.9.

The same procedure works also for other values of d, provided k_2 in the above expression of h is replaced by k_d, where $k_1 = 1.854$, $k_3 = 1.498$, $k_4 = 1.396$.

Complement 5.3 (Extension of Stein's lemma) If $X \sim \text{N}(\mu, \sigma^2)$ and h is a differentiable function such that $\mathbb{E}\{|h'(X)|\} < \infty$, Stein's lemma states that

$$\text{cov}\{X, h(X)\} = \sigma^2\, \mathbb{E}\{h'(X)\} \,. \tag{5.81}$$

An extension to multivariate normal variables exists.

Adcock (2007) presents an extension of this result to the case of a multivariate ESN variable, which he developed as a tool for tackling an optimization problem in finance. His result was formulated in a parameterization essentially like (5.30) but, for homogeneity with the rest of our exposition, we recast the result for the parameterization $\text{SN}_d(\xi, \Omega, \alpha, \tau)$. Another difference is that the proof below makes use of the canonical form, which simplifies the logic of the argument. The transformation $Y^* = H^\top(Y - \xi)$ has been defined in Proposition 5.13 in connection with an SN distribution, hence with $\tau = 0$, but the same transformation works here, leading to $Y^* \sim \text{SN}_d(0, I_d, \alpha_{Z^*}, \tau)$; see Problem 5.14.

https://doi.org/10.1017/CBO9781139248891.006 Published online by Cambridge University Press

Lemma 5.22 *Let $Y \sim \mathrm{SN}_d(\xi, \Omega, \alpha, \tau)$ and denote by $h(x)$ a real-valued function on $\mathbb{R}^d$ such that $h'_i(x) = \partial h(x)/\partial x_i$ is continuous and $\mathbb{E}\{|h'_i(Y)|\}$ is finite, for $i = 1, \ldots, d$. Then*

$$cov\{Y, h(Y)\} = \Omega\, \mathbb{E}\{\nabla h(Y)\} + (\mathbb{E}\{Y\} - \xi) \Big(\mathbb{E}\{h(W)\} - \mathbb{E}\{h(Y)\}\Big), \qquad (5.82)$$

where $\nabla h(Y) = \Big(h'_1(Y), \ldots, h'_d(Y)\Big)^\top$, $W \sim \mathrm{N}_d(\xi - \tau\omega\delta, \Omega - \omega\delta\delta^\top\omega)$, $\mathbb{E}\{Y\} = \xi + \zeta_1(\tau)\, \omega\, \delta$ as in (5.71); here as usual δ is given by (5.11) and ζ_1 by (2.20).

Proof Consider the canonical form $Z^* = H^\top(Y - \xi) \sim \mathrm{SN}_d(0, I_d, \alpha_{Z^*}, \tau)$ where H is defined in Proposition 5.13. The variables $Z^*_1, \ldots, Z^*_d$ are mutually independent and the last $d-1$ components have N(0, 1) distribution. Therefore, by the original Stein's lemma (5.81), we have

$$\mathrm{cov}\{Z^*_i, h(Z^*)\} = \mathbb{E}\{h'_i(Z^*)\} \qquad (i = 2, \ldots, d)\,,$$

first arguing conditionally on the remaining components and then, by independence, unconditionally. For $Z^*_1 \sim \mathrm{SN}(0, 1, \alpha_*, \tau)$, we have

$$\mathbb{E}\{Z^*_1\, h(Z^*)\} = \int_{\mathbb{R}^{d-1}} \prod_{j=2}^{d} \varphi(z_i) \left[\int_{\mathbb{R}} h(z)\, z_1\, \varphi(z_1; \alpha_*, \tau)\, \mathrm{d}z_1\right] \mathrm{d}z_2 \cdots \mathrm{d}z_d$$

where $\varphi(z_1; \alpha_*, \tau)$ is given by (2.39). Expansion of the inner integral by parts lends

$$\int_{\mathbb{R}} \frac{\partial}{\partial z_1} h(z) \varphi(z_1; \alpha_*, \tau)\, \mathrm{d}z_1 + \frac{\alpha_* \varphi(\tau)}{\Phi(\tau)} \int_{\mathbb{R}} h(z)\, \varphi\Big((z_1 + \tau\delta(\alpha_*)) \sqrt{1 + \alpha_*^2}\Big)\, \mathrm{d}z_1,$$

where $\delta(\cdot)$ is given by (2.6). When this expression is inserted back in the d-dimensional integral, we get

$$\mathbb{E}\{Z^*_1 h(Z^*)\} = \mathbb{E}\{h'_1(Z^*)\} + \delta(\alpha_*)\, \zeta_1(\tau) \int_{\mathbb{R}^d} h(u)\, \varphi_d(u - \mu_U; \Omega_U)\, \mathrm{d}u,$$

where $\mu_U = (-\tau\delta(\alpha_*), 0, \ldots, 0)^\top$ and $\Omega_U = \mathrm{diag}(1 - \delta(\alpha_*)^2, 1, \ldots, 1)$. On recalling that $\delta(\alpha_*) = \delta_*$ where δ_* is defined by (5.38), write

$$\mathbb{E}\{Z^*_1\, h(Z^*)\} = \mathbb{E}\{h'_1(Z^*)\} + \zeta_1(\tau)\delta_* \mathbb{E}\{h(U)\}\,,$$

where $U \sim \mathrm{N}_d(\mu_U, \Omega_U)$, and from (5.71) we obtain

$$\mathrm{cov}\{Z^*, h(Z^*)\} = \mathbb{E}\{\nabla h(Z^*)\} + \mathbb{E}\{Z^*\} \Big(\mathbb{E}\{h(U)\} - \mathbb{E}\{h(Z^*)\}\Big).$$

Since $Z^* = H^\top(Y-\xi)$, then $H^\top\Omega H = I_d$, and so also $(H^\top)^{-1}H^{-1} = \Omega$. Moreover, since $\Omega_U = I_d - \zeta_1(\tau)^{-2}\mathbb{E}\{Z^*\}\,\mathbb{E}\{Z^*\}^\top$, we obtain $(H^\top)^{-1}\Omega_U H^{-1} = \Omega - \omega\delta\delta^\top\omega$, bearing in mind (5.71). From these facts and

$$\operatorname{cov}\{Y,h(Y)\} = \operatorname{cov}\{Y-\xi,h(Y)\} = (H^\top)^{-1}\operatorname{cov}\left\{Z^*, h(\xi+(H^\top)^{-1}Z^*)\right\},$$

we arrive at (5.82). QED

Problems

5.1 Prove Proposition 5.1.

5.2 Confirm that the distribution of $Z=(Z_1,\ldots,Z_d)$ whose components are defined by (5.19) is $\mathrm{SN}_d(0,\bar\Omega,\alpha)$, where $\bar\Omega$ and α are given by (5.21) and (5.22) (Azzalini and Dalla Valle, 1996).

5.3 Show that, for any choice of $\bar\Omega$ and α, there is a choice of $\bar\Psi$ and δ in (5.16)–(5.19) leading to the distribution $\mathrm{SN}_d(0,\bar\Omega,\alpha)$. From here show how the parameterization (ξ,Ψ,λ) of (5.30) can be mapped to (ξ,Ω,α) of (5.3), and conversely (Azzalini and Capitanio, 1999, Appendix of the full version).

5.4 In § 5.1.3 it is stated that $\bar\Omega$ and δ are not variation independent; hence not all choices $(\bar\Omega,\delta)$ are admissible. Show that a necessary and sufficient condition for their admissibility is $\bar\Omega - \delta\delta^\top > 0$.

5.5 Check (5.27). Also, show that in case $h=1$, the expression reduces to

$$\alpha_{1(2)} = \left(1+\alpha_{2*}^2-u^2\right)^{-1/2}\,(\alpha_1+u)$$

where $u=\bar\Omega_{12}\alpha_2$ and $\alpha_{2*}^2 = \alpha_2^\top\bar\Omega_{22}\alpha_2$. Finally, show that $\alpha_{1(2)}=\lambda_1$, the first component of vector λ in (5.20).

5.6 Confirm that the parameters of (5.40) are as given by (5.41)–(5.44).

5.7 Show that α_X in (5.43) can be written as

$$\left(1+\alpha^\top\omega^{-1}\left(\Omega-\Omega A\Omega_X^{-1}A^\top\Omega\right)\omega^{-1}\alpha\right)^{-1/2}\;\omega_X\,\Omega_X^{-1}A^\top\Omega\omega^{-1}\alpha\,.$$

5.8 Confirm the statement at the end of § 5.1.2 that the sum of two independent multivariate SN variables, both with non-zero slant, is not of SN type.

5.9 Consider the variable (U,Z), where $U\in\mathbb{R}$ and $Z=(Z_1,\ldots,Z_d)^\top\in\mathbb{R}^d$, with joint density

$$\begin{aligned} f(u,z) \;&=\; \frac{(1+\alpha_*^2)^{1/2}}{(2\pi)^{(d+1)/2}\,\det(\bar\Omega)^{1/2}}\,\exp\Big\{-\tfrac{1}{2}\Big(u^2\,(1+\alpha_*^2) \\ &\qquad -2\,(1+\alpha_*^2)^{1/2}\,|u|\,\alpha^\top z + z^\top\bar\Omega^{-1}z + (\alpha^\top z)^2\Big)\Big\} \end{aligned}$$

where $\bar{\Omega} > 0$ is a correlation matrix and $\alpha_*^2 = \alpha^\top \bar{\Omega} \alpha$. Show the following: (a) marginally, $U \sim \mathrm{N}(0,1)$ and $Z \sim \mathrm{SN}_d(0, \bar{\Omega}, \alpha)$; (b) Z and U are independent if and only if $\alpha = 0$; (c) $\mathrm{cov}\{U, Z_i\} = 0$ for $i = 1, \ldots, d$.

5.10 Consider a bivariate SN distribution with location $\xi = 0$, $\mathrm{vech}(\Omega) = (1, r, 1)^\top$ and $\alpha = a(-1,1)^\top$ where $r \in (-1,1)$ and $a \in \mathbb{R}$. Show that, if $a \to \infty$, then $\delta \to \frac{1}{2}\sqrt{1-r}\,(-1,1)^\top$. If further $r \to 1$, then $\delta \to 0$ and correspondingly $\gamma_1 \to 0$ for each marginal. Examine the numerical values of γ_1 and the contour lines plot of the density in the case $r = 0.9$ and $a = 100$. Comment on the qualitative implications.

5.11 For a $(d{+}1)$-dimensional normal distribution as in (5.14), consider the conditional distribution of X_0 under two-sided constraint of X_1, that is $Z = (X_0 | a < X_1 < b)$ where a and b are arbitrary, provided $a < b$. Obtain the density function and the lower-order moments of Z, specifically the marginal coefficients of skewness and kurtosis (Kim, 2008).

5.12 Suppose that $X \sim \mathrm{SN}_d(\xi, \Omega, \alpha)$ is partitioned as $X = (X_1^\top, X_2^\top)^\top$, where X_1 has dimension h. Then show, without using (5.67), that the distribution of X_2 conditionally on $X_1 = x_1$ is $\mathrm{SN}_{d-h}(\xi_{2\cdot 1}, \Omega_{22.1}, \alpha_{2\cdot 1}, \tau_c)$, of which the first three parameter components are as in (5.67) and $\tau_c = \alpha_{1(2)}^\top \omega_1^{-1}(x_1 - \xi_1)$, where $\alpha_{1(2)}$ is given by (5.27).

5.13 Show that the additive representation (5.70) of a multivariate ESN distribution is equivalent to representation (5.68).

5.14 Extend the idea of the canonical form of § 5.1.8 to the ESN case. Specifically, if $Y \sim \mathrm{SN}_d(\xi, \Omega, \alpha, \tau)$, show that there exist a matrix H such that $Z^* = H^\top(Y - \xi) \sim \mathrm{SN}_d(0, I_d, \alpha_{Z^*}, \tau)$, so that the distribution of Z^* can be factorized as the product of $d-1$ standard normal densities and that of $\mathrm{SN}(0, 1, \alpha_*, \tau)$, where α_{Z^*} and α_* are as for the SN case. Use this result to derive the final expressions in (5.74) and (5.75).

5.15 If $X = (X_1, \ldots, X_d)^\top \sim \mathrm{N}_d(0, \Omega)$ where $\Omega > 0$, Šidák (1967) has shown that the inequality

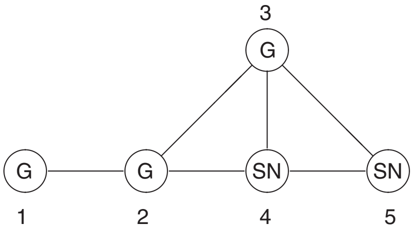
$$\mathbb{P}\{|X_1| \le c_1, \ldots, |X_d| \le c_d\} \ge \textstyle\prod_{i=1}^d \mathbb{P}\{|X_i| \le c_i\}$$

holds for any positive numbers $c_1, \ldots, c_d$. Prove that the same inequality holds when $X \sim \mathrm{SN}_d(0, \Omega, \alpha)$, and that for any $p \in [0,1]$ the choice of a sequence $c_1, \ldots, c_d$ such that $\mathbb{P}\{|X_1| \le c_1, \ldots, |X_d| \le c_d\} = 1 - p$ does not depend on the parameter α.

5.16 Show that the set of distributions $\mathrm{SN}_5(\xi, \Omega, \alpha, \tau)$ compatible with the marked conditional independence graph depicted below must satisfy the condition $\{\alpha_1 = \alpha_5 = 0\} \cap \{\alpha_2 \neq 0\}$. Also, show that changing the

https://doi.org/10.1017/CBO9781139248891.006 Published online by Cambridge University Press

graph to $1 \in V_{\mathrm{SN}}$ would make it incompatible with the above ESN assumption.



5.17 Consider the density

$$f(x) = 2\,\varphi_2(x; \bar{\Omega}, \alpha)\,\Phi\{\lambda\,(x_1^2 - x_2^2)\}, \qquad x = (x_1, x_2) \in \mathbb{R}^2,$$

where $\bar{\Omega}$ is a correlation matrix, $\alpha \in \mathbb{R}^2$ and $\lambda \in \mathbb{R}$, that is, a density similar to the one in (1.30) but with the bivariate normal density replaced by a bivariate SN. Confirm that $f(x)$ is a proper density and show that, if X has density $f(x)$ with $\alpha = a\,(1,1)^\top$ for some $a \in \mathbb{R}$, then $X^\top \bar{\Omega}^{-1} X \sim \chi_2^2$.

https://doi.org/10.1017/CBO9781139248891.006 Published online by Cambridge University Press