Test (2008) 17: 370–380
DOI 10.1007/s11749-006-0037-3

ORIGINAL PAPER

Modeling maxima of longitudinal contralateral observations

Nicola Loperfido

Received: 26 November 2005 / Accepted: 22 July 2006 /
Published online: 10 March 2007

Abstract The paper gives the joint distribution of maxima of contralateral observations taken from the same individual at several occasions, when the data are normal with given constraints on the parameters. Different constraints lead to different well-known generalizations of the normal distribution. As an immediate consequence, evaluation of several features of these maxima is greatly simplified. The results of the paper are also useful in modeling other phenomena, such as measures of physical fitness based on the best of two trials. Further applications deal with statistical inference, for example, testing whether an identified treatment is best.

Keywords Canonical fundamental skew-normal distribution · Exchangeability · Extended skew-normal distribution · Multivariate skew-normal distribution · Nonlinear prediction

Mathematics Subject Classification (2000) Primary 62E15 · 62H10 · Secondary 62P10

1 Introduction

Sometimes physicians (i.e., ophthalmologists, otologists, pneumologists) take observations from both sides of the same body (i.e., from each eye, each ear, each lung) at different occasions. Data are then arranged in two vectors of lateral observations, that is, observations from the same side. Data collected at the same time from different sides form a vector of contralateral observations. Let $y_L = (Y_{1L}, \dots, Y_{pL})^T$ and $y_R = (Y_{1R}, \dots, Y_{pR})^T$ be the vectors of lateral observations from the left and

N. Loperfido (✉)
Istituto di Scienze Economiche, Facoltá di Economia, Universitá degli Studi di Urbino "Carlo Bo", Via Saffi 42, 61029 Urbino (PU), Italy
e-mail: nicola@econ.uniurb.it

right side of the same individual measured at p occasions, respectively. The corresponding vectors of contralateral observations are $(Y_{1L}, Y_{1R})^T, \ldots, (Y_{pL}, Y_{pR})^T$. Hence, the vectors of contralateral observations are bivariate longitudinal data. Moreover, let $x = (X_1, \ldots, X_k)^T$ denote a set of covariates. The vector of maxima of contralateral observations is $\{\max(Y_{1L}, Y_{1R}), \ldots, \max(Y_{pL}, Y_{pR})\}^T$. In a more compact notation, we write $\max(y_L, y_R)$ to denote the vector of maxima taken componentwise. The vector $\min(y_L, y_R)$ is defined in a similar way. The paper only deals with $\max(y_L, y_R)$, since corresponding results for $\min(y_L, y_R)$ easily follow from the identity $\min(y_L, y_R) = -\max(-y_L, -y_R)$. The joint distribution of x, y_L, and y_R is commonly assumed to be multivariate normal, whose properties a reviewed in detail by Tong (1990), with constraints on the parameters to model bilateral symmetries in the human body: $E(y_L) = E(y_R)$, $\mathrm{var}(y_L) = \mathrm{var}(y_R)$, $\mathrm{cov}(x, y_R) = \mathrm{cov}(x, y_L)$. For example, Olkin and Viana (1995) modeled age X, vision from left eye Y_L, and vision from right eye Y_R as follows:

$$\begin{pmatrix} X \\ Y_L \\ Y_R \end{pmatrix} \sim N_3 \left\{ \begin{pmatrix} \mu_0 \\ \mu_1 \\ \mu_1 \end{pmatrix}, \begin{pmatrix} \sigma^2 & \gamma\sigma\tau & \gamma\sigma\tau \\ \gamma\sigma\tau & \tau^2 & \rho\tau^2 \\ \gamma\sigma\tau & \rho\tau^2 & \tau^2 \end{pmatrix} \right\}, \tag{1.1}$$

where $2\gamma^2 < 1 + \rho$ and $\rho^2 < 1$. The model implies irrelevance of left–right labeling: $P(X \le a, Y_L \le b, Y_R \le c) = P(X \le a, Y_L \le c, Y_R \le b)$ for any real values a, b, c. Olkin and Viana (1995) only consider first and second moments of the vector $\{X, \max(Y_L, Y_R), \min(Y_L, Y_R)\}^T$. This paper gives the exact distributions of $\max(Y_L, Y_R)$, $\{X, \max(Y_L, Y_R)\}^T$, and $X|\max(Y_L, Y_R) = m$. It also shows that they belong to well-known families (which will be described in Sect. 2), possessing very useful analytical properties.

The same authors, in a subsequent paper (Viana and Olkin 2000), assume the normality, irrelevance of left–right labeling at each time point, and lateral-contralateral homogeneity (correlations between data observed at different times only depend on times themselves). The corresponding distribution is

$$\begin{pmatrix} y_L \\ y_R \end{pmatrix} \sim N_{2p} \left\{ \begin{pmatrix} \mu \\ \mu \end{pmatrix}, \begin{pmatrix} \Omega & \Omega - \Lambda \\ \Omega - \Lambda & \Omega \end{pmatrix} \right\}, \tag{1.2}$$

where $\Omega - \Lambda/2$ is a symmetric positive definite $p \times p$ matrix, and Λ is the diagonal matrix whose diagonal elements are the nonnegative scalars $\lambda_1, \ldots, \lambda_p$. Note that, for $p = 1$, the joint distribution of Y_L and Y_R in (1.1) is the same as that of y_L and y_R in (1.2), with $\mu = \mu_1$, $\Omega = \tau^2$, and $\Lambda = \tau^2(1-\rho)$. Similarly to model (1.1), model (1.2) implies that $P(y_L \le h, y_R \le k) = P(y_L \le k, y_R \le k)$ for any pair of p-dimensional real vectors h, k. Equivalently, we can say that the pairs Y_L ,Y_R and y_L, y_R are exchangeable in models (1.1) and (1.2), respectively. More formally, n q-dimensional random vectors $x_1, \ldots, x_n$ are exchangeable if $P(x_1 \le a_1, \ldots, x_n \le a_n) = P(x_{\pi(1)} \le a_1, \ldots, x_{\pi(n)} \le a_n)$, where $\pi(1), \ldots, \pi(n)$ is a permutation of the first n positive integers, and $a_1, \ldots, a_n$ are n q-dimensional real vectors. Aldous (1985) gives a detailed review of the role of exchangeability in probability and statistics.

Interest in maxima of contralateral observations arises in several situations. In the first place, they may be relevant in their own right. Parving and Christiansen (1990) use best ear's hearing as a measure of deafness in mentally retarded adults. In the second place, they may be useful predictors. Frenkel and Shin (1986) examine the predictive value of worst vision in elderly patients with glaucoma. In the third place, they may be the only available data. As an example, consider a survey on visual acuities in the adult population of the USA, when limited grants prevent hiring professional ophthalmologists. A possible solution is looking at driving licenses, since criterions for an unrestricted driving license are based on visual acuity from the best eye, in most states of the USA (Fishman et al. 1993).

This paper focuses on modeling maxima of contralateral observations, but maxima of two normally distributed random vectors also appear elsewhere, as shown by the following miscellaneous examples. Roberts (1966) considers an application in epidemiology, related to the study of twins. Tomohiro et al. (2003) compare high and low intensity aerobic exercises using the best of two trials for several strength tests. Gupta and Gupta (2001) study the monotonicity of the hazard rate for parallel (series) systems, whose failure times are the maximum (minimum) of the failure times of their components, when the components themselves are normal.

Further applications of the maximum of two normally distributed random variables can be found in hypothesis testing, where it appears as a test statistic. Cox and Wermuth (1991) advocate its use whenever there are two qualitatively different types of departures from the null hypothesis. The same statistic appears in connection to the problem of testing whether an identified treatment is best (Laska and Meisner 1989). Another case for this statistic is the evaluation of p-values when two statistical tests are applied to the same data set, but only the most convenient result is reported, as in (Jiang 1997) and (Loperfido 2002). Azzalini and Cox (1984) introduce a new test for the analysis of variance which can be expressed in terms of maxima of two normally distributed random variables. All these examples can be generalized to maxima of two random vectors.

The paper gives the exact distribution of $\max(z, w)$ when z and w are exchangeable normal vectors with given constraints on their covariance. All these distributions (also known as skew-normal distributions) are well-known extensions of the normal one. Hence, results in the paper connect order statistics from longitudinal observations with skew-normal distributions, with mutual benefit. The former benefit from the many analytical properties of skew-normal distributions, including simple formulas for expectations, variances, skewness, and correlations. The latter benefit from the many applications related to order statistics, which contribute to fulfill the need for more illustrations of practical utility of skew-normal models pointed out by Azzalini (2002) and Arnold and Beaver (2002).

The paper is organized as follows: Sect. 2 describes some skew-normal distributions. Section 3 contains main results. Section 4 applies them to visual assessment data. Section 5 contains all proofs.

2 Skew-normal distributions

The maximum of two random variables whose joint distribution is bivariate normal has been considered by Gupta and Pillai (1965), Roberts (1966), Nagarajah (1982), Kella (1986), Cain (1994), Cain and Pan (1995), Olkin and Viana (1995), Loperfido (2002). Its distribution and its main features have been evaluated in the general case as well as under several equality constraints on means and variances. In particular, the density function of the maximum of two standardized random variables whose joint distribution is bivariate normal with correlation ρ is $2\phi(z)\Phi(\psi z)$, where $\psi = \sqrt{(1-\rho)/(1+\rho)}$, $\Phi(z)$, and $\phi(z)$ are the distribution function and density function of a standard normal variable, respectively. Azzalini and Capitanio (2003) and Dalla Valle (2004) address as a research problem the generalization of this result to the multivariate case. This paper shows that the answer is related to two generalizations of the normal distribution: the multivariate skew-normal distribution (Azzalini and Dalla Valle 1996) and the canonical fundamental skew-normal distribution (Arellano-Valle and Genton 2005). The paper also shows that the predictive distribution based on the maximum of two random variables, under normality and exchangeability assumptions, is the extended skew-normal introduced by Arnold and Beaver (2000). Azzalini (2005) gives a good review of the previous literature on this class of distributions.

Azzalini and Dalla Valle (1996), in their seminal paper, introduced the multivariate skew-normal distribution (SN, hereafter) with the density function

$$f(z;\xi,\Omega,\eta) = 2\phi_p(z;\xi,\Omega)\Phi\left\{\eta^T(z-\xi)\right\},$$

where $z,\xi,\eta \in \mathbb{R}^p$, and $\phi_p(z;\xi,\Omega)$ is the density function of $N_p(\xi,\Omega)$. The vector ξ, the matrix Ω, and the vector η are commonly referred to as the location vector, the scale matrix, and the shape vector of the p-variate skew-normal distribution $SN_p(\xi,\Omega,\eta)$. When the shape parameter equals the null vector, the above density is a normal one.

Arellano-Valle and Genton (2005) introduced the canonical fundamental skew-normal distribution (CFSN, hereafter) with the density function

$$f(z;\xi,\Omega,\Delta) = 2^m\phi_p(z;\xi,\Omega)\Phi_m\left\{\Delta^T\Omega^{-1/2}(z-\xi); I_m - \Delta^T\Delta\right\},$$

where $z;\xi \in \mathbb{R}^p$, $\Omega \in \mathbb{R}^p \times \mathbb{R}^p$, $\Delta \in \mathbb{R}^p \times \mathbb{R}^m$, and $\Phi_m(z; I_m - \Delta^T\Delta)$ is the distribution function of $N_p(0_p, I_m - \Delta^T\Delta)$. We write $CFSN_p(\xi,\Omega,\Delta)$ to denote the corresponding distribution and refer to the vector ξ, the matrix Ω, and the matrix Δ as the location vector, the scale matrix, and the shape matrix, respectively. The SN distribution is a CFSN distribution when the shape matrix has only one column. Expectation and variance of a CFSN random vector have a simple analytical form.

Arnold and Beaver (2000) introduced the extended skew-normal distribution (ESN, hereafter), another generalization of the normal distribution with the density function

$$f(z;\xi,\Omega,\delta,\tau) = \frac{\phi_p(z;\xi,\Omega)}{\Phi(\tau)}\Phi\left\{\frac{\tau + \delta^T\Omega^{-1}(z-\xi)}{\sqrt{1-\delta^T\Omega^{-1}\delta}}\right\},$$

where $\xi \in \mathbb{R}^p$, $\tau \in \mathbb{R}$, Ω is a symmetric positive definite $p \times p$ matrix, and δ is a p-dimensional vector such that $\delta^T\Omega^{-1}\delta < 1$. We write $ESN_p(\xi,\Omega,\delta,\tau)$ to denote the corresponding distribution and refer to the vector ξ, the matrix Ω, the vector δ, and the scalar τ as the location vector, the scale matrix, the shape vector, and the truncation parameter, respectively. The SN distribution is an ESN distribution with the truncation parameter equal to zero. Arnold and Beaver (2004) give a good review of the previous literature related to this distribution.

The above parameterizations suffer from some drawbacks and alternative parameterizations are discussed by Azzalini (1985) and Capitanio et al. (2003). However, they are quite common in the literature and allow simple representations of moments and cumulants. For example, the expectation and variance of $ESN_p(\xi, \Omega, \delta, \tau)$ are

$$E(z) = \xi + \delta \frac{\phi(\tau)}{\Phi(\tau)}, \qquad \operatorname{var}(z) = \Omega - \delta\delta^T \frac{\phi(\tau)}{\Phi(\tau)} \left\{ \tau + \frac{\phi(\tau)}{\Phi(\tau)} \right\}.$$

The ratio $\phi(\tau)/\Phi(\tau)$ is known in the literature with several names. Econometricians usually call it the Mill ratio. It is a positive, decreasing, and nonlinear function of τ, and

$$\lim_{\tau \to +\infty} \frac{\phi(\tau)}{\Phi(\tau)} = 0, \qquad \lim_{\tau \to -\infty} \frac{-\phi(\tau)}{\tau \Phi(\tau)} = 1.$$

3 Main results

We denote by ξ, Ω, and ψ a p-dimensional vector, a symmetric positive definite $p \times p$ matrix, and the parameter $\sqrt{(1-\rho)/(1+\rho)}$, respectively. Theorem 3.1 shows that maxima of contralateral observations have a SN distribution, when the joint distribution of lateral observations belongs to a class of models which includes (1.1) as a special case.

Theorem 3.1 *Let β be a nonnegative vector such that $\beta^T \Omega^{-1} \beta < 2$ and*

$$\begin{pmatrix} z \\ w \end{pmatrix} \sim N_{2p} \left\{ \begin{pmatrix} \xi \\ \xi \end{pmatrix}, \begin{pmatrix} \Omega & \Omega - \beta\beta^T \\ \Omega - \beta\beta^T & \Omega \end{pmatrix} \right\}.$$

Then $\max(z, w) \sim SN_p\{\xi, \Omega, \Omega^{-1}\beta/(2 - \beta^T \Omega^{-1} \beta)^{1/2}\}$.

The following theorem shows that maxima of contralateral observations have a CFSN distribution, when the joint distribution of lateral observations is (1.2). We write $A^{1/2}$ to denote the square root of a symmetric positive definite matrix A: $A = A^{1/2} A^{1/2}$.

Theorem 3.2 *Let y_L and y_R be two random vectors with joint distribution* (1.2). *Then* $\max(y_L, y_R) \sim CFSN\{\xi, \Omega, \Omega^{-1/2}(\Lambda/2)^{1/2}\}$.

Theorems 3.1 and 3.2 model $\operatorname{cov}(y_L, y_R)$ in two different ways. The former assumes that $\operatorname{cov}(y_L, y_R)$ equals $\operatorname{var}(y_L) = \operatorname{var}(y_R)$ minus a symmetric matrix of rank less or equal to one with nonnegative elements. The latter assumes that $\operatorname{cov}(y_L, y_R)$ equals $\operatorname{var}(y_L) = \operatorname{var}(y_R)$ minus a diagonal matrix (of rank possibly greater than one) with nonnegative diagonal elements. Assumptions coincide when $p = 1$, leading to $\max(Y_L, Y_R) \sim SN_1(\mu_1, \tau^2, \psi/\tau)$ with $\xi = \mu = \mu_1$, $\Omega = \tau^2$, and $\beta = \Lambda = \tau(1-\rho)^{1/2}$. This result was independently found by Roberts (1966) and Loperfido (2002), so that Theorems 3.1 and 3.2 generalize it to the multivariate case in two different ways.

Theorems 3.1 and 3.2 lead to the joint distribution of $\max(y_L, y_R)$ and a vector of covariates x. It suffices to let $\tilde{y}_L^T = (x^T, y_L^T)$ and $\tilde{y}_R^T = (x^T, y_R^T)$. If assumptions in Theorems 3.1 or 3.2 hold for y_L and y_R, they also hold for $\tilde{y}_L$ and $\tilde{y}_R$. As an example, consider the joint distribution of X and $\max(Y_L, Y_R)$ under model (1.1):

Corollary 3.1 *Let the joint distribution of the random variables X, Y_L, and Y_R be* (1.1). *Then the joint distribution of X and* $\max(Y_L, Y_R)$ *is*

$$SN_2\left\{\begin{pmatrix}\mu_0\\ \mu_1\end{pmatrix}, \begin{pmatrix}\sigma^2 & \gamma\sigma\tau\\ \gamma\sigma\tau & \tau^2\end{pmatrix}, \sqrt{\frac{(1-\rho)/(1-\gamma^2)}{1-2\gamma^2+\rho}}\begin{pmatrix}-\tau\gamma/\sigma\\ 1/\tau\end{pmatrix}\right\}.$$

The result easily follows from Theorem 3.1 by letting $z = (X, Y_L)^T$, $w = (X, Y_R)^T$, $\xi = E(z)$, $\Omega = \operatorname{var}(z)$, and $\beta = \tau\sqrt{1-\rho}(0, 1)^T$.

Theorems 3.1 and 3.2 also lead to an interesting property of bivariate normal orthant probabilities.

Corollary 3.2 *Let $L(h, k, \rho) = P(Z_1 > h, Z_2 > k)$, where Z_1 and Z_2 are two standardized variables whose joint distribution is normal with correlation $-1 < \rho < 1$. Then $L(a, a, \rho) = 2L\{a, 0, \sqrt{(1-\rho)/2}\}$.*

From Theorem 3.1 we know that $M = \max(Z_1, Z_2) \sim SN_1(0, 1, \psi)$, where $\psi = \sqrt{(1-\rho)/(1+\rho)}$. Azzalini and Dalla Valle (1996) show that $P(M > a) = P(Y_1 > a|Y_2 > 0)$, where Y_1 and Y_2 are two standardized variables whose joint distribution is normal with correlation $\sqrt{\psi/(1+\psi^2)} = \sqrt{(1-\rho)/2}$. Since $P(Y_2 > 0) = 0.5$, we have

$$P(M > a) = 2P(Y_1 > a, Y_2 > 0) = 2L\big\{a, 0, \sqrt{(1-\rho)/2}\big\}.$$

In order to complete the proof, it suffices to recall that, by definition, $M = \max(Z_1, Z_2)$, which implies $P(M > a) = L(a, a, \rho)$. It is worth noticing that the proof is based on probabilistic arguments only, which are quite different from direct evaluation of integrals and series expansions, the most common approaches for dealing with bivariate normal orthant probabilities (Kotz et al. 2000). Corollary 3.2 can be used for approximating bivariate normal orthant probabilities. Cox and Wermuth (1991) propose a simple approximation of $L(h, k, \rho)$, which is quite accurate but when $h = k = a \neq 0$. By Corollary 3.2, their approximation can be used even when $h = k = a \neq 0$ if applied to $2L\{a, 0, \sqrt{(1-\rho)/2}\}$.

Olkin and Viana (1995) obtain the correlations and covariances between X and $\max(Y_L, Y_R)$ using model (1.1). They did not deal with the predictive distribution of X based on $\max(Y_L, Y_R)$, which is given in the following theorem.

Theorem 3.3 *Let* (1.1) *be the joint distribution of X, Y_L, and Y_R. Then the distribution of $X|$*$\max(Y_L, Y_R) = m$ *is*

$$ESN_1\left\{\mu_0 + \frac{m-\mu_1}{\tau}\gamma\sigma, \sigma^2\big(1-\gamma^2\big), \sigma\frac{\rho-\gamma}{\sqrt{1-\rho^2}}, \frac{m-\mu_1}{\tau}\psi\right\}.$$

Practical implications are twofold. In the first place, it provides the nonlinear predictor $E\{X|\max(Y_L, Y_R) = m\}$ of X based on $\max(Y_L, Y_R)$, which is better than the linear one considered by Olkin and Viana (1995) under quadratic loss. In the second place, it allows analytical evaluation of various features of $X|\max(Y_L, Y_R) = m$ (henceforth $X|m$, for brevity), such as the expectation and variance, using results in Capitanio et al. (2003):

$$E(X|m) = \mu_0 + \gamma\sigma m^* + \sigma\frac{\rho-\gamma}{\sqrt{1-\rho^2}}M\big(m^*\psi\big),$$

$$\mathrm{var}(X|m) = \sigma^2\big(1-\gamma^2\big) - \sigma^2\frac{(\rho-\gamma)^2}{1-\rho^2}M\big(m^*\psi\big)\big\{m^*\psi + M\big(m^*\psi\big)\big\}$$

where $M(m^*) = \phi(m^*)/\Phi(m^*)$ is the Mill ratio and $m^* = (m-\mu_1)/\tau$. The conditional expectation $E(X|m)$ is a nonlinear function of m, since the Mill ratio is a nonlinear function, too. The eteroschedasticity follows from $\mathrm{var}(X|m)$ being a nonconstant function of m.

Future research in this area might consider the joint distribution of the covariates and linear combinations of order statistics from bilateral observations. This issue has been addressed by Gupta and Pillai (1965), Nagarajah (1982), and Viana (1998). The latter also points out its relevance in visual sciences.

4 Visual assessment data

This section applies the results of the previous section to Best's vitelliform macular dystrophy (BVMD) and intraocular pressure (IOP). Fishman et al. (1993) evaluated 43 patients with BVMD for age (X), vision of left eye (Y_L), and vision of right eye (Y_R). Olkin and Viana (1995) applied model (1.1) to these data and computed maximum-likelihood estimates (simply mle, hereafter): $\hat{\mu}_0 = 28.833$, $\hat{\mu}_1 = 0.424$, $\hat{\sigma} = 19.182$, $\hat{\tau} = 0.386$, $\hat{\rho} = 0.496$, $\hat{\gamma} = 0.581$. They also estimated first and second moments of X, $\min(Y_L, Y_R)$, and $\max(Y_L, Y_R)$, together with best linear predictions of one variable based on another variable or on both remaining variables.

We shall first consider the marginal distribution of $\max(Y_L, Y_R)$. We already know from Theorem 3.1 that it has a SN distribution whose parameters are known functions of μ_1, τ, and ρ. Hence, the mle of the distribution of $\max(Y_L, Y_R)$ is $SN_1(0.424, 0.1490, 1.5037)$. Results in Azzalini (1985) lead to mle of skewness and kurtosis of $\max(Y_L, Y_R)$, which are 0.0359 and 0.0103, respectively, hinting that the distribution of $\max(Y_L, Y_R)$ is very similar to a normal one. It seems then reasonable to approximate $SN_1(0.424, 0.1490, 1.5037)$ with $SN_1(0.424, 0.1490, 0)$, that is, $N(0.424, 0.1490)$, which physicians are much more comfortable to work with. However, this approximation might not be appropriate. In the first place, the graphs of the density functions of $SN_1(0.424, 0.1490, 1.5037)$ and $N(0.424, 0.1490)$ are markedly different. More precisely, the right (left) tail of the former is much heavier (lighter) than the right (left) tail of the latter. The approximation can be also evaluated using the Kolmogorov distance between $SN_1(0.424, 0.1490, 1.5037)$ and $N(0.424, 0.1490)$ (i.e., the supremum, taken over all possible real values, of the absolute value of the difference between the corresponding distribution functions), which equals 0.1674

(Azzalini 1985). It can be shown that this value equals the difference between the cumulative distribution functions, evaluated in 0.424, of $N(0.424, 0.1490)$, that is, 0.5, and $SN_1(0.424, 0.1490, 1.5037)$, that is, 0.3326. This result has some practical implications, since physicians typically asses the health status of patients using quantiles: the lower the quantile corresponding to the patient, the lower his health status. Hence, a physician using the approximation $N(0.424, 0.1490)$ would not be too concerned about a patient with BVMD whose visual acuity score were about 0.424, which corresponds to the median of the approximating distribution. However, use of $SN_1(0.424, 0.1490, 1.5037)$ implies that this score corresponds to a much lower quantile, hence suggesting the possible presence of a health problem.

We shall now consider the conditional distributions of $\max(Y_L, Y_R)$ given the observed value of X. Model (1.1) implies that

$$\begin{pmatrix} Y_L \\ Y_R \end{pmatrix} \Bigg| X = x \sim N_2 \left\{ \begin{pmatrix} 1 \\ 1 \end{pmatrix} \left(\mu_1 - \gamma\tau \frac{x - \mu_0}{\sigma} \right), \tau^2 \begin{pmatrix} 1-\gamma^2 & \rho-\gamma^2 \\ \rho-\gamma^2 & 1-\gamma^2 \end{pmatrix} \right\}.$$

Hence, Y_L and Y_R are exchangeable normal random variables conditionally on X, too. It follows that $\max(Y_L, Y_R)|X = x$ has a SN distribution, whose estimate is $SN_1(0.0869 + 0.0117x, 0.0987, 2.2943)$. Estimated expectation, variance, skewness, and kurtosis of $\max(Y_L, Y_R)|X = x$ are $0.2313 + 0.0117x$, 0.0772, 0.0769, and 0.0286, respectively.

We shall now consider the conditional distribution of $X|m$, which is ESN by Theorem 3.3. The corresponding mle is $ESN_1(16.588 + 28.872m, 243.74, -1.8810, 1.5037m - 0.6376)$. A straightforward application of results in Capitanio et al. (2003) implies that expectation, variance, skewness, and kurtosis depend on the observed value of $\max(Y_L, Y_R)$ and that $E(X|m)$ is a nonlinear function of m.

We shall now consider intraocular pressure, with emphasis on the association between pretreatment and posttreatment maximal intraocular pressure. Sonty et al. (1996) observed 15 subjects with glaucomatous eyes. For each subject, the intraocular pressure was measured from each eye before and after topical beta blocker therapy. Let the pretreatment and posttreatment intraocular pressure from left (right) eye be $Y_{L1}(Y_{R1})$ and $Y_{L2}(Y_{R2})$, respectively. Viana and Olkin (2000) modeled these data using (1.2). The corresponding estimates of the corresponding scale and shape parameters are

$$\begin{pmatrix} 12.667 & 9.8237 \\ 9.8237 & 17.190 \end{pmatrix}, \qquad \begin{pmatrix} 0.6661 & -0.3434 \\ -0.2320 & 0.8271 \end{pmatrix}$$

Mle of $\mathrm{var}\{\max(Y_{R1}, Y_{L1})\}$, $\mathrm{var}\{\max(Y_{R2}, Y_{L2})\}$, $\mathrm{cov}\{\max(Y_{R1}, Y_{L1}), \max(Y_{R2}, Y_{L2})\}$ are 9.8674, 11.0591, 9.8237, respectively (Arellano-Valle and Genton 2005). The correlation between $\max(Y_{R1}, Y_{L1})$ and $\max(Y_{R2}, Y_{L2})$ is then 0.9404. The correlation between Y_{L1} (Y_{R1}) and $\max(Y_{R2}; Y_{L2})$ equals 0.6657, as it follows from results in Olkin and Viana (1995). Hence, the maximal pretreatment IOP is a better predictor of the max posttreatment IOP than the pretreatment IOP from any eye (or a weighted average of IOP from both eyes).

5 Proofs

Proof of Theorem 3.1 Without loss of generality we can assume that $\xi = 0_p$. Let the random variable U and random vector x be such that $z = x + \delta U$ and $w = x - \delta U$. Hence,

$$\begin{pmatrix} U \\ x \end{pmatrix} \sim N_{p+1}\left\{\begin{pmatrix} 0 \\ 0_p \end{pmatrix}, \begin{pmatrix} 1 & 0_p^T \\ 0_p & \Omega - \delta\delta^T \end{pmatrix}\right\}$$

where 0_p is the p-dimensional null vector, and $\delta = \beta/2^{1/2}$. Consider now the identity $\max(Z_i, W_i) = (Z_i + W_i)/2 + |Z_i - W_i|/2$. From the definitions of U and x and nonnegativeness of β we obtain $\max(Z_i, W_i) = X_i + \delta_i|U|$. The density of U is symmetric at the origin. It follows that $|U| \sim U|U > 0$. Hence, $\max(z, w) \sim z|U > 0$. The joint distribution of U and z is

$$\begin{pmatrix} U \\ z \end{pmatrix} \sim N_{p+1}\left\{\begin{pmatrix} 0 \\ 0_p \end{pmatrix}, \begin{pmatrix} 1 & \delta^T \\ \delta & \Omega \end{pmatrix}\right\}.$$

Azzalini and Dalla Valle (1996) show that the distribution of $z|U > 0$ is $SN_p\{0_p, \Omega, \Omega^{-1}\delta/\sqrt{(1 - \delta^T\Omega^{-1}\delta)}\}$. As a consequence, $\max(z, w) \sim SN_p\{0_p, \Omega, \Omega^{-1}\delta/\sqrt{(1 - \delta^T\Omega^{-1}\delta)}\}$. Recall now the definition of δ and complete the proof. □

Proof of Theorem 3.2 Without loss of generality we can assume that $\xi = 0_p$ and let $\Delta = \Omega^{-1/2}(\Lambda/2)^{1/2}$, $z = (4\Omega - 2\Lambda)^{-1/2}(y_L + y_R)$, and $w = (2\Lambda)^{-1/2}(y_L - y_R)$. In matrix notation, we have

$$\begin{pmatrix} z \\ w \end{pmatrix} = \begin{pmatrix} 4\Omega - 2\Lambda & O_p \\ O_p & 2\Lambda \end{pmatrix}^{-1/2} \begin{pmatrix} I_p & I_p \\ I_p & -I_p \end{pmatrix} \begin{pmatrix} y_L \\ y_R \end{pmatrix},$$

where O_p is the null $p \times p$ matrix. Ordinary properties of the multivariate normal distribution imply that z and w are two independent and identically distributed random vectors with distribution $N_p(0_p, I_p)$. Their definitions also imply that $y_L = (\Omega - \Lambda/2)^{1/2}z + (\Lambda/2)^{1/2}w$. Simple matrix algebra leads to $y_L = \Omega^{1/2}\{(I_p - \Omega^{-1/2}\Lambda\Omega^{-1/2}/2)^{1/2}z + \Omega^{-1/2}(\Lambda/2)^{1/2}w\}$. Recall now the definition of $\Delta : y_L = \Omega^{1/2}\{(I_p - \Delta\Delta^T)^{1/2}z + \Delta w\}$. Let $|v|$ be the vector whose the ith component is the absolute value of the ith component of the vector v for $i = 1, \ldots, p$. Ordinary properties of maxima imply that $\max(y_L, y_R) = (y_L + y_R)/2 + |y_L - y_R|/2$. Recall now the definitions of z, w, and Δ and write $\max(y_L, y_R) = \Omega^{1/2}\{(I_p - \Delta\Delta^T)^{1/2}z + \Delta|w|\}$. We already know that z and w are two independent and identically distributed random vectors with distribution $N_p(0_p, I_p)$. Hence the last identity implies that $\max(y_L, y_R) \sim CFSN_p(0_p, \Omega, \Delta)$ by Proposition 2.2 in Arellano-Valle and Genton (2005). □

Proof of Theorem 3.3 Let $M = \max(Y_L, Y_R)$. The density function of $X|M = m$ can be represented as follows:

$$f(x|m) = f(x|m, Y_L < Y_R)P(Y_L < Y_R|m) + f(x|m, Y_L > Y_R)P(Y_L > Y_R|m).$$

The random vectors $(X, Y_L, Y_R)^T$ and $(X, Y_R, Y_L)^T$ are identically distributed, so that $P(Y_L < Y_R|m) = P(Y_L > Y_R|m) = 0.5$ and $f(x|m, Y_L < Y_R) = f(x|m, Y_L > Y_R)$. By ordinary properties of maxima, these equations lead to $f(x|m) = f(x|m, Y_L < Y_R)$ and $f(x|m) = f(x|Y_L < Y_R = m)$. Consider now the conditional distribution of $-Y_L$ and z given $Y_R = m$:

$$N_2\left[\left\{\begin{array}{c}-\mu_1 - \rho(m-\mu_1)\\ \mu_0 + \gamma(m-\mu_1)/\tau\end{array}\right\}, \left\{\begin{array}{cc}\tau^2(1-\rho^2) & \tau\sigma(\rho-\gamma)\\ \tau\sigma(\rho-\gamma) & \sigma^2(1-\gamma^2)\end{array}\right\}\right].$$

Let the random variables U and W be jointly distributed as $-Y_L$ and X conditionally on $Y_R = m$. Arnold and Beaver (2000) show that the distribution of $w|U > -m$ is

$$ESN_1\left\{\mu_0 + \frac{m-\mu_1}{\tau}\gamma\sigma, \sigma^2\left(1-\gamma^2\right), \frac{\sigma(\rho-\gamma)}{(1-\rho^2)^{1/2}}, \frac{m-\mu_1}{\tau}\psi\right\},$$

where $\psi = \sqrt{(1-\rho)/(1+\rho)}$. We already know that the conditional distribution of $W|U > -m$ equals the conditional distribution of $X|m$. □

Acknowledgements The author wish to thank Prof. Azzalini for reading an earlier version of this manuscript and two anonymous referees for helpful comments.

References

Aldous DJ (1985) Exchangeability and related topics. In: Hennequin PL (ed) Ecole d'été de probabilités de Saint-Flour, XIII. Springer series in statistics. Springer, Berlin

Arellano-Valle RB, Genton MG (2005) On fundamental skew distributions. J Multivar Anal 96:93–116

Arnold BC, Beaver RJ (2000) Hidden truncation models. Sankhya Ser A 62:22–35

Arnold BC, Beaver RJ (2002) Skewed multivariate models related to hidden truncation and/or selective reporting (with discussion). Test 11:7–54

Arnold BC, Beaver RJ (2004) Elliptical models subject to hidden truncation or selective sampling. In: Genton MG (ed) Skew-elliptical distributions and their applications: a journey beyond normality. Chapman & Hall/CRC, Boca Raton, pp 101–112

Azzalini A (1985) A class of distributions which includes the normal ones. Scand J Stat 12:171–178

Azzalini A (2002) Discussion of Arnold and Beaver's paper "Skewed multivariate models related to hidden truncation and/or selective reporting". Test 11:36

Azzalini A (2005) The skew-normal distribution and related multivariate families (with discussion). Scand J Stat 32:159–200

Azzalini A, Capitanio A (2003) Distributions generated by perturbation of symmetry with emphasis on a multivariate skew-t distribution. J Roy Stat Soc Ser B 65:367–389

Azzalini A, Cox DR (1984) Two new tests associated with analysis of variance. J Roy Stat Soc Ser B 46:335–343

Azzalini A, Dalla Valle A (1996) The multivariate skew-normal distribution. Biometrika 83:715–726

Cain M (1994) The moment-generating function of the minimum of bivariate normal random variables. Am Stat 48:124–125

Cain M, Pan M (1995) Moments of the minimum of bivariate normal random variables. Math Sci 20:119–122

Capitanio A, Azzalini A, Stanghellini E (2003) Graphical models for skew-normal variates. Scand J Stat 30:129–144

Cox DR, Wermuth N (1991) A simple approximation for bivariate and trivariate normal integrals. Int Stat Rev 59:263–269

Dalla Valle A (2004) The skew-normal distribution. In: Genton MG (ed) Skew-elliptical distributions and their applications: a journey beyond normality. Chapman & Hall/CRC, Boca Raton, pp 3–24

Fishman GA, Baca W, Alexander KR, Derlacky DJ, Glenn AM, Viana MAG (1993) Visual acuity in patients with best vitelliform macular distrophy. Ophthalmology 100:1665–1670

Frenkel R, Shin D (1986) Prevention and management of delayed suprachoroidal hemorrhage after filtration surgery. Arch Ophthalmol 104:1459–1463

Gupta PL, Gupta RC (2001) Failure rate of the minimum and maximum of a multivariate normal distribution. Metrika 53:39–49

Gupta SS, Pillai S (1965) On linear functions of ordered correlated normal random variables. Biometrika 52:367–379

Jiang J (1997) Sharp upper and lower bounds for asymptotic levels of some statistical tests. Stat Probab Lett 35:395–400

Kella O (1986) On the distribution of maximum of bivariate normal random variables with general means and variances. Commun Stat Theory Methods 15:3265–3276

Kotz S, Balakrishnan N, Johnson NL (2000) Continuous multivariate distributions, vol 1. Wiley, New York

Laska EM, Meisner MJ (1989) Testing whether an identified treatment is best. Biometrics 45:1139–1151

Loperfido N (2002) Statistical implications of selectively reported inferential results. Stat Probab Lett 56:13–22

Nagarajah HN (1982) A note on linear functions of ordered correlated normal random variables. Biometrika 69:284–285

Olkin I, Viana M (1995) Correlation analysis of extreme observations from a multivariate normal distribution. J Am Stat Assoc 90:1373–1379

Parving A, Christiansen B (1990) Hearing of the mentally retarded living at home. Ugeskr Læ 152:3161–3164

Roberts C (1966) A correlation model useful in the study of twins. J Am Stat Assoc 61:1184–1190

Sonty SP, Sonty S, Viana MAG (1996) The additive ocular hypotensive effect of topical 2% dorzolamide on glaucomatous eyes on topical beta blocker therapy. Investig Ophthalmol Vis Sci 37:1100

Tomohiro O, Yoshio N, Kiyoji T (2003) Effects of exercise intensity on physical fitness and risk factors for coronary heart disease. Obes Res 11:1131–1139

Tong YL (1990) The multivariate normal distribution. Springer series in statistics. Springer, Berlin

Viana M (1998) Linear combinations of ordered symmetric observations with application to visual acuity. In: Balakrishnan N, Rao CR (eds) Order statistics: applications. Handbook of statistics, vol 17. North Holland, Amsterdam, pp 513–524. Chap 19

Viana M, Olkin I (2000) Symmetrically dependent models arising in visual assessment data. Biometrics 56:1188–1191