# Discussion of "The Skew-normal"

MARC G. GENTON

*Department of Statistics, Texas A&M University*

First, I would like to congratulate Professor Azzalini for an excellent overview paper on the topic of the skew-normal distribution and its various extensions. This is a fast-growing area of research that has been pioneered by Professor Azzalini's seminal article on 'A class of distributions which includes the normal ones' (Azzalini, 1985). Presently, both frequentist and Bayesian statisticians are actively involved in the study of these univariate and multivariate skewed distributions. A sample of current researchers have recently contributed to an edited book entitled 'Skew-elliptical distributions and their applications: a journey beyond normality' (Genton, 2004a). Professor Azzalini fully deserves the credit for generating such a diverse family of researchers!

My discussion consists of some additional links to existing results from the literature as well as suggestions for further research. My comments are centred around three main themes. The first one, in section 7, is devoted to the construction of skew-normal distributions, their properties, and some extensions. The second theme is concerned with flexibility and inference with skewed distributions in section 8. The third theme, in section 9, is motivated by applications.

## 7. Skew-normal and related distributions

I enjoyed the parallel presentation of the univariate and multivariate skew-normal distributions in sections 2 and 3, respectively. I would like to start with some comments about the form of the perturbation function. As mentioned in the article, the expression $G\{w(z)\}$ appearing in lemmas 1 and 3 can be replaced by an arbitrary skewing function $\pi : \mathbb{R}^d \to [0, 1]$ satisfying $\pi(-z) = 1 - \pi(z)$, see Wang *et al.* (2004a) and Genton & Loperfido (2005). Moreover, the stochastic representation and the perturbation invariance property still hold in this framework. Note that a stochastic representation equivalent to (2) but slightly more convenient to use is

$$Z = \begin{cases} Y & \text{if } U < \pi(Y), \\ -Y & \text{otherwise,} \end{cases} \tag{36}$$

where $Y \sim f_0$ and $U \sim U(0, 1)$, a uniform random variable. Indeed, this formulation avoids the generation of $X \sim G'$. It holds both in the univariate and multivariate setting.

When $d = 1$ and $\pi(z) = \Phi(\alpha z)$, lemma 2 is the key result in computing the moment generating function of the univariate skew-normal distribution. Denoting by $\Phi_m(z, \Sigma)$ the cumulative distribution function of an $N_m(0, \Sigma)$ variate at a point $z \in \mathbb{R}^m$, the result analogous to lemma 2 for $d > 1$ is the following.

**Lemma 2b**

*If* $u \sim N_d(0, I_d)$ *and* $a \in \mathbb{R}^m$, $B \in \mathbb{R}^{m\times d}$, *then*

$$\mathbb{E}\{\Phi_m(a + Bu, I_m)\} = \Phi_m(a, I_m + BB^{\mathrm{T}}). \tag{37}$$

The proof can be found in Arellano-Valle & Genton (2005). Lemma 2b with $m = 1$ allows for the computation of the moment generating function of the multivariate skew-normal distribution with $\pi(z) = \Phi(\alpha^{\mathrm{T}} z)$, yielding the explicit formula (17). However, it suggests also that more general families of multivariate skew-normal distributions can be constructed for $m > 1$, for example, such as those described by (27).

As noted by Wang *et al.* (2004b), the perturbation invariance property discussed in sections 2.1 and 3.1 implies an equivalence between chi-square and certain skew-normal distributions. Specifically, if $Z$ is a random variable with probability density function $f$, then $Z^2 \sim \chi_1^2$ if and only if there exists a skewing function $\pi$ such that $f(z) = 2\phi(z)\pi(z)$. Similarly in the multivariate case with the additional condition that $f_0$ is spherical, if $z \in \mathbb{R}^d$ is a random vector with probability density function $f$, then $z^{\mathrm{T}}z \sim \chi_d^2$ if and only if there exists a skewing function $\pi$ such that $f(z) = 2\phi_p(z)\pi(z)$. Note that the sphericity condition for $f_0$ is important because non-spherical $f_0$ yielding $z^{\mathrm{T}}z \sim \chi_d^2$ can be constructed. An important class of functions $t$ resulting in the perturbation invariance property (13) is given by quadratic forms $t(x) = x^{\mathrm{T}}Ax$, where $x \in \mathbb{R}^d$ and $A \in \mathbb{R}^{d \times d}$. Indeed, quadratic forms appear naturally in many different contexts involving multivariate data. For instance, the sample autocovariance estimator of a stationary time series and the sample variogram estimator of a stationary random field are quadratic forms in the data vector. The perturbation invariance implies that the joint distribution of these estimators at various lags does not depend on the skewing function $\pi$; see Genton *et al.* (2001) for the skew-normal case, Kim & Mallick (2003) for the skew-$t$, and Genton (2004b) for general skew-symmetric distributions.

The quantiles of the skew-normal distribution are worth investigating. Following the framework developed by Parzen (1979, p. 106), the quantile function of the skew-normal distribution defined by (3) and (f) is $Q(u; \alpha) = \Phi^{-1}(u; \alpha)$ and its score function is

$$J(u;\alpha) = \frac{-\phi'(Q(u;\alpha);\alpha)}{\phi(Q(u;\alpha);\alpha)} = Q(u;\alpha) - \alpha\frac{\phi(\alpha Q(u;\alpha))}{\Phi(\alpha Q(u;\alpha))} \approx (1+\alpha^2)Q(u;\alpha), \tag{38}$$

where we used the fact that $1 - \Phi(z) \approx \phi(z)/z$ for large $z$. Thus, the tail behaviour of the skew-normal distribution is the same as the normal distribution. Note in particular that $Q(u; 0) = J(u; 0) = \Phi^{-1}(u)$ for the standard normal distribution. In order to compare the quantiles of a skew-normal distribution with the quantiles of a normal distribution with same mean and variance, I plot in Fig. 8 the comparison distribution function

$$D(u;\alpha) = \Phi\left(\frac{\sqrt{2}\alpha}{\sqrt{\pi(1+\alpha^2)}} + \sqrt{1 - \frac{2\alpha^2}{\pi(1+\alpha^2)}}\,\Phi^{-1}(u);\alpha\right), \tag{39}$$

for various values of the skewness parameter $\alpha = 1, 2, 3, 5, 10$. The diagonal of the unit square is obtained for $\alpha = 0$ and the comparison distribution does not change significantly more for $\alpha > 10$. More generally, Chang & Genton (2004) have shown that the extreme value distribution of univariate skew-symmetric distributions based on the symmetric density $f_0$ is the same as the extreme value distribution of $f_0$ under mild conditions on the skewing function $\pi$. They provide various illustrative examples such as the skew-normal, skew-$t$, and skew-Cauchy distributions.

An important setting where skew-normal and related distributions appear is motivated by selective sampling; see the discussion at the beginning of section 2.3. I illustrate next that the selection criteria does not need to be restricted to a unilateral one. As a simple example, consider the height ($H$) and weight ($W$) of crews on a submarine, who are selected only if they satisfy some height requirements. For instance, suppose that candidates for working on the submarine must be at least 165 cm tall (to perform some specific tasks), but no more than 195 cm tall (because of the confinement due to the submarine's structure). A bivariate normal distribution might well be acceptable to model the joint distribution of height and weight in the general population. However, the height restriction will result in a skewed distribution for the weight of the selected individuals, the density of which has been derived by Arnold *et al.* (1993). The top row of Fig. 9 depicts this situation based on the simulated data from Fig. 2.

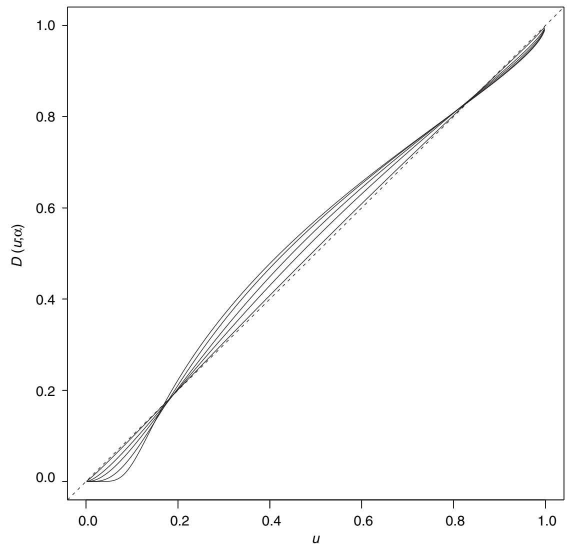


*Fig. 8.* Comparison distribution between a skew-normal distribution and a normal distribution with same mean and variance, for various values of the skewness parameter $\alpha = 1, 2, 3, 5, 10$.

The left-hand side plot shows the simulated data using different symbols for points inside the interval $165 < H < 195$. The right-hand side plot displays the histogram and the theoretical density of $W|165 < H < 195$, as well as the marginal density of $W$ from the untruncated bivariate normal distribution. Next, suppose that on another type of submarine, candidates must be either less than 120 cm tall (to perform some very specific tasks) or more than 180 cm tall (to have access to some special parts of the submarine). Based on a bivariate normal distribution of height and weight for the general population, the height restriction will result in a skewed distribution, possibly bimodal, for the weight of the selected individuals, the density of which has the form

$$\phi(w;\mu_2,\sigma_{22})\frac{1-\Phi\left((1-\rho^2)^{-1/2}\left[\frac{180-\mu_1}{\sqrt{\sigma_{11}}}-\rho\frac{w-\mu_2}{\sqrt{\sigma_{22}}}\right]\right)+\Phi\left((1-\rho^2)^{-1/2}\left[\frac{120-\mu_1}{\sqrt{\sigma_{11}}}-\rho\frac{w-\mu_2}{\sqrt{\sigma_{22}}}\right]\right)}{1-\Phi\left(\frac{180-\mu_1}{\sqrt{\sigma_{11}}}\right)+\Phi\left(\frac{120-\mu_1}{\sqrt{\sigma_{11}}}\right)}, \tag{40}$$

where $\rho = \sigma_{12}/\sqrt{\sigma_{11}\sigma_{22}}$ is the correlation in the untruncated population. Note that if $\rho = 0$, then the density (40) is normal (not skewed). The bottom row of Fig. 9 depicts this situation based on the simulated data from Fig. 2. The left-hand side plot shows the simulated data using different symbols for points outside the interval $120 < H < 180$. The right-hand side plot displays the histogram and the theoretical density of $W|(H < 120 \text{ or } H > 180)$, as well as the marginal density of $W$ from the untruncated bivariate normal distribution. Note the bimodality of the theoretical density (40). These remarks show that many extensions of the skew-normal distributions can be obtained by selective sampling based on various constraints for $H$, some yielding multimodal distributions or other interesting characteristics. With my co-workers, Professor Arellano-Valle and Professor Branco, we have now derived a unified theory for multivariate skewed distributions resulting from arbitrary sets defining the selection mechanism. These results will be published elsewhere in the near future.

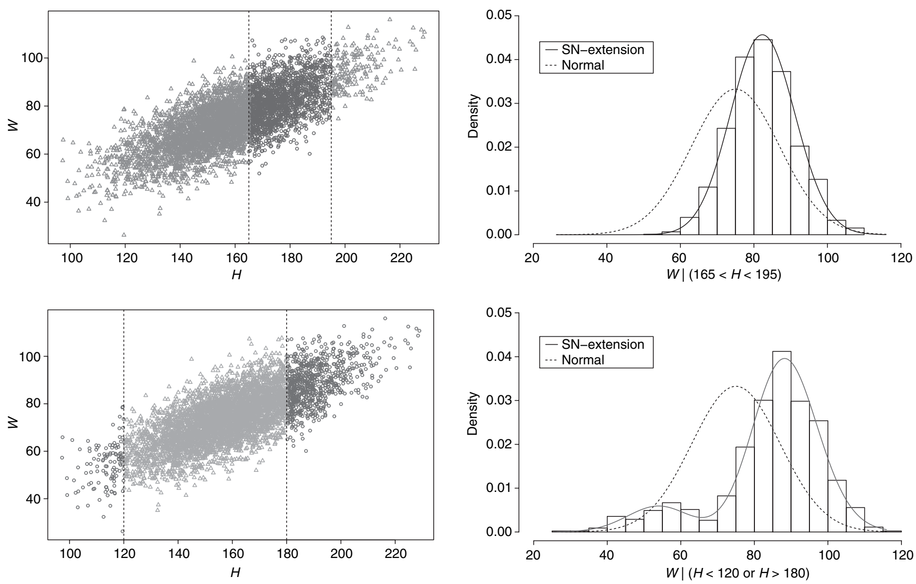


*Fig. 9.* The left-hand side plots show the simulated data using different symbols for points inside the interval $165 < H < 195$ (top row) and outside the interval $120 < H < 180$ (bottom row). The right-hand side plots display the histogram and the theoretical density of $W|165 < H < 195$ (top row) and $W|(H < 120$ or $H > 180)$ (bottom row), as well as the marginal density of $W$ from the untruncated bivariate normal distribution.

As a complement to the discussion about the skew-normal distribution (3) arising as the stationary distribution of a threshold autoregressive process, I would like to mention the example described in Tong (1990, p. 146). Briefly, the setting is defined by an abstract generalization of a threshold autoregressive process to $\mathbb{R}^d$ that yields a stationary distribution with density of the form

$$\left[\frac{2(1-\alpha^2)}{\pi}\right]^{d/2} \exp\left[-\frac{1}{2}(1-\alpha^2)z^{\mathrm{T}}z\right]\prod_{i=1}^{d}\Phi(-\alpha z_i), \tag{41}$$

where $z = (z_1, \ldots, z_d)^{\mathrm{T}}$ and $-1 < \alpha < 1$. For $d = 1$, the skew-normal-like density (41) reduces to a skew-normal density similar to (3), but for $d > 1$ it is not of the form (16). I speculate that the stationary distribution of threshold autoregressive processes indexed by $\mathbb{R}^d$ (e.g. a spatial process on a lattice for $d = 2$) may yield stationary distributions with densities of the form (16). More investigations are needed to validate this conjecture.

Finally, I discuss some ideas related to the skew-*t* distribution. As mentioned in sections 2.5 and 4.2, many applications require the possibility of modelling both the skewness and the thickness of the tails of a distribution. Skew-*t* distributions, and more generally skew-elliptical distributions, allow for such flexibility. An interesting alternative to the skew-*t* distribution is the skew-slash distribution proposed by Wang & Genton (2005). It is simply defined as the ratio of a skew-normal random vector and an independent uniform random variable $U$ on the interval (0, 1) raised to the power $1/q$, $q > 0$. Similar tail heaviness as the skew-*t* can be achieved by changing the parameter $q$. For example, Cauchy-like tails are obtained with $q = 1$. One advantage of the skew-slash over the skew-*t* distribution is that its moments are available explicitly due to the property that $\mathbb{E}(U^{-k/q}) = q/(q-k)$, $q > k$. Another attractive feature is that simulations from the skew-slash distribution are straightforward with software that permits simulations from the skew-normal distribution. A possible drawback is that the density of the skew-slash distribution does not have a closed form.

## 8. Flexibility and inference

### *8.1. Parametric*

Flexible skewed distributions based on the centrally symmetric density $f_0$ can be obtained by a suitable choice of the skewing function $\pi(z) = G\{w(z)\}$. For example, Ma & Genton (2004) proposed the use of odd polynomials in order to approximate the odd function $w$. The degree of the polynomial is linked in a complex fashion to the number of modes of the resulting skewed distribution. In the univariate case, for instance, it can be shown that $2\phi(z)\Phi(\alpha z + \beta z^3)$ can have at most two modes. More general statements are however difficult to obtain. In practice, the order of the polynomial is most conveniently chosen adaptively via model selection strategies such as likelihood ratio tests or AIC/BIC criteria.

Instead of odd polynomials, one can investigate other sets of functions that form a complete orthogonal system, for example such as Fourier or wavelet bases, to approximate the odd function $w$. I explore ideas from Fourier analysis (e.g. Tolstov, 1962) and consider the univariate Fourier sine series

$$w_M(z) = \sum_{k=1}^{M} b_k \sin(kz), \tag{42}$$

where $M$ is a positive integer and $b_k \in \mathbb{R}$. A well-known result is that an odd function $w$, periodic on the interval $[-\pi, \pi]$, admits a Fourier sine series, that is, $w(z) = \lim_{M\to\infty} w_M(z)$, where $b_k = 1/\pi \int_{-\pi}^{\pi} w(x)\sin(kx)\mathrm{d}x$. This suggests that $G\{w_M(z)\}$ provides flexibility in the resulting skewed distribution. Extensions to the multivariate setting are straightforward. For instance, consider the bidimensional case and the double Fourier series

$$w_{M,N}(z_1, z_2) = \sum_{k=0}^{M}\sum_{l=0}^{N}\Big[b_{k,l}\sin(kz_1)\cos(lz_2) + c_{k,l}\cos(kz_1)\sin(lz_2)\Big], \tag{43}$$

where $M$, $N$ are positive integers and $b_{k,l}$, $c_{k,l} \in \mathbb{R}$. If $w$ satisfies $w(-z_1, -z_2) = -w(z_1, z_2)$ and is periodic on the domain $[-\pi, \pi] \times [-\pi, \pi]$, then $w(z_1, z_2) = \lim_{M\to\infty, N\to\infty} w_{M,N}(z_1, z_2)$, where $b_{k,l} = 1/\pi^2 \int_{-\pi}^{\pi}\int_{-\pi}^{\pi} w(x, y)\sin(kx)\cos(ly)\mathrm{d}x\,\mathrm{d}y$ and $c_{k,l} = 1/\pi^2 \int_{-\pi}^{\pi}\int_{-\pi}^{\pi} w(x, y)\cos(kx)$ $\sin(ly)\mathrm{d}x\,\mathrm{d}y$. Note that in all the expressions above, a simple change of variables can be used to transform the interval of integration from $[-\pi, \pi]$ to $[-L, L]$, $L > 0$. For practical applications, $L$ can always be chosen so that the corresponding domain covers the range of the data. Figure 10 illustrates the flexibility achieved by $w_3(z)$ (top row) from (42), and $w_{1,1}(z_1, z_2)$ (bottom row) from (43), when skewing a standard normal distribution with density $f_0$.

### *8.2. Semiparametric*

Sometimes, the interest lies in the parameters, $\beta \in \mathbb{R}^p$ say, of the central model $f_0$ rather than the skewing function $\pi$. For example, if the data are obtained only for a selected portion of the population of interest, then the data can be modelled as a random sample from a distribution with density $2f_0(z; \beta)\pi(z)$; see the survey article by Bayarri & DeGroot (1992) on such selection models and references therein. Another point of view, as mentioned in the article at the end of section 4, is given by robustness considerations. In that case, the skewing function $\pi$ can be seen as a contaminating function, producing asymmetric outliers in the observed sample. In

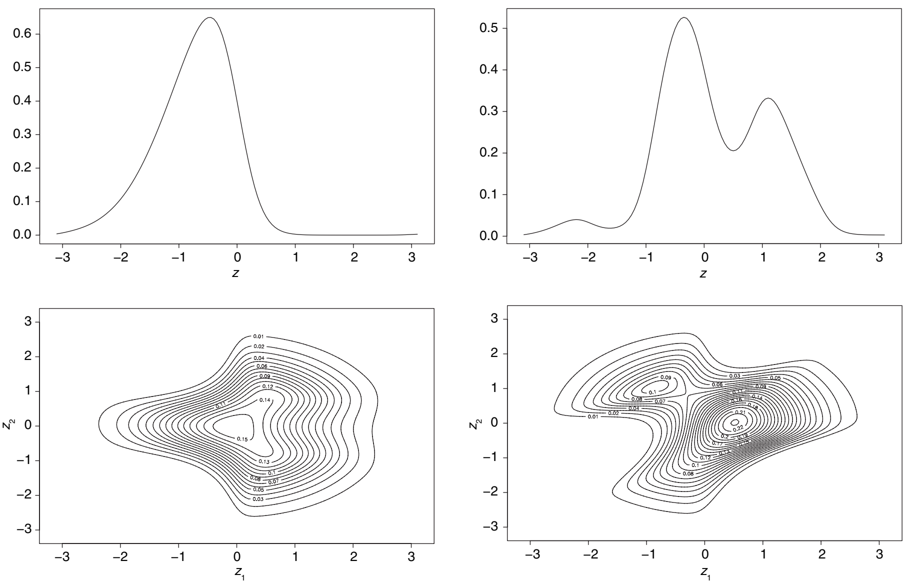

*Fig. 10.* Flexible densities resulting from a standard normal density $f_0$ and skewing functions based on Fourier series. Top row: $w_3(z) = b_1 \sin(z) + b_2 \sin(2z) + b_3 \sin(3z)$. Bottom row: $w_{1,1}(z_1, z_2) = b_{1,0} \sin(z_1) + c_{0,1} \sin(z_2) + b_{1,1} \sin(z_1)\cos(z_2) + c_{1,1} \cos(z_1)\sin(z_2)$.

both situations, the goal is to provide consistent and asymptotically normal estimators of $\beta$ in the presence of a nuisance skewing function $\pi$.

This view point naturally suggests that inference on $\beta$ should be carried out in the framework of semiparametric estimation, where $\beta$ is the finite dimensional parameter of interest and $\pi$ is an infinite dimensional nuisance parameter. Various semiparametric estimation methods can be implemented. Ma *et al.* (2005) took a geometric approach to construct regular asymptotically linear (RAL) estimators, and linked RAL estimators to influence functions. By identifying all the influence functions, all the RAL estimators can be found. Following Bickel *et al.* (1993), the space of all the mean zero functions is a Hilbert space $\mathcal{H}$, where the inner product is defined by covariance. Here, the mean and covariance are computed with respect to the true distribution. The space where influence functions belong is the space that is orthogonal to the nuisance tangent space $\Lambda$, which is the mean square closure of all linear combinations of nuisance score functions. As soon as $\mathcal{H}$ and $\Lambda$ are identified, the space $\Lambda^{\perp}$ orthogonal to the nuisance tangent space can be characterized and hence semiparametric RAL estimators can be constructed. In particular, among all the RAL estimators, the most efficient semiparametric estimator can be found.

In the class of skew-symmetric distributions, Ma *et al.* (2005) have identified $\Lambda^{\perp}$ and determined the class of all the semiparametric estimators of $\beta$. They also constructed an estimator that is locally efficient, in the sense that this estimator contains an infinite dimensional parameter $\eta$ that can be chosen arbitrarily in a large class. The locally efficient estimator is robust in that no matter what one chooses for $\eta$, the resulting estimator provides a $\sqrt{n}$-consistent estimator for $\beta$, where $n$ denotes the sample size; if one can find an optimal $\eta$, then the estimator is also semiparametric efficient. Ma *et al.* (2005) showed that the optimal choice for $\eta$ is in fact the true skewing function $\pi$. However, besides a possible approximation of $\pi$ using the flexible class of skewing functions proposed by Ma & Genton (2004) based on odd polynomials, they did not provide a method for estimating $\pi$. Hence, the estimator does not necessarily achieve the promised optimal efficiency.

### *8.3. Non-parametric*

Recently, new developments have suggested a way to achieve the optimal efficiency. These methods inevitably involve non-parametric estimation of the skewing function $\pi$. Because this is ongoing research, I only provide here a brief sketch of the main ideas.

Non-parametric estimation of an unknown density $f$ is a fundamental problem in statistics. Splines, local polynomials and kernel estimators are among the preferred methods. In the framework of skew-normal and related distributions, the task is complicated in several respects, including the fact that the central density model has a specific finite dimensional parameter $\beta$, and that the infinite dimensional parameter $\pi$ is subject to two constraints.

Using a local polynomial estimation in conjunction with either a profile likelihood or the locally efficient semiparametric estimator mentioned above, Ma & Hart (2005) have derived a non-parametric estimator that not only achieves the optimal semiparametric efficiency for $\beta$, but also results in a non-parametric density estimator that has the standard $O_p(h^2)$ bias and $O_p(1/(nh))$ variance. Other non-parametric estimators can also be used to replace the local polynomial estimator, although, at least for spline estimators, the asymptotics may be non-trivial to derive.

From a different point of view, Hjort & Glad (1995) have developed an approach designed to reduce bias in a broad non-parametric neighbourhood of a given parametric class of densities, $f(z, \theta)$ say. The basic idea is to start with a parametric estimate $f(z, \hat{\theta})$, and then multiply it with a non-parametric estimate of the correction function $r(z) = f(z)/f(z, \hat{\theta})$. In the context of skew-normal distributions, a natural candidate for the parametric start is the normal or skew-normal density. In the latter case, the non-parametric estimate of the correction factor provides further non-parametric adjustment to the skewing function $\Phi(\alpha z)$. This of course can be generalized to the various extensions of the skew-normal distribution. Further research and implementations are needed to pursue these ideas.

## 9. Applications

As a first application, I revisit the Australian athletes data analysed in section 3.6 by means of a linear model for LBM and BMI with a normal and a skew-normal assumption for the distribution of the errors. Additional flexibility can be obtained by using a flexible skew-normal distribution as suggested by Ma & Genton (2004). Due to the perturbation invariance for this class of distributions discussed in section 7, the quadratic form (29) is still distributed as $\chi^2_d$. The top panel in Fig. 11 depicts the bivariate residuals and the contour levels of the fitted error distribution (by maximum likelihood) under the assumption of a bivariate flexible skew-normal distribution with a skewing function based on an odd polynomial of degree 3:

$$\pi(z_1, z_2) = \Phi(\alpha_1 z_1 + \alpha_2 z_2 + \alpha_3 z_1^3 + \alpha_4 z_2^3 + \alpha_5 z_1^2 z_2 + \alpha_6 z_1 z_2^2), \quad (44)$$

where I have set $\alpha_5 = \alpha_6 = 0$ for parsimony of the model. The observed value of the likelihood ratio test for the nullity of the additional parameters $\alpha_3$ and $\alpha_4$ in the flexible skew-normal model compared with the skew-normal model is 0.8 which is not significant on the $\chi^2_2$ scale. The bottom panel in Fig. 11 depicts the case where all parameters in (44) are estimated. The observed value of the likelihood ratio test for the nullity of the additional parameters $\alpha_3, \ldots, \alpha_6$ in the flexible skew-normal model compared with the skew-normal model is 28.2, which is

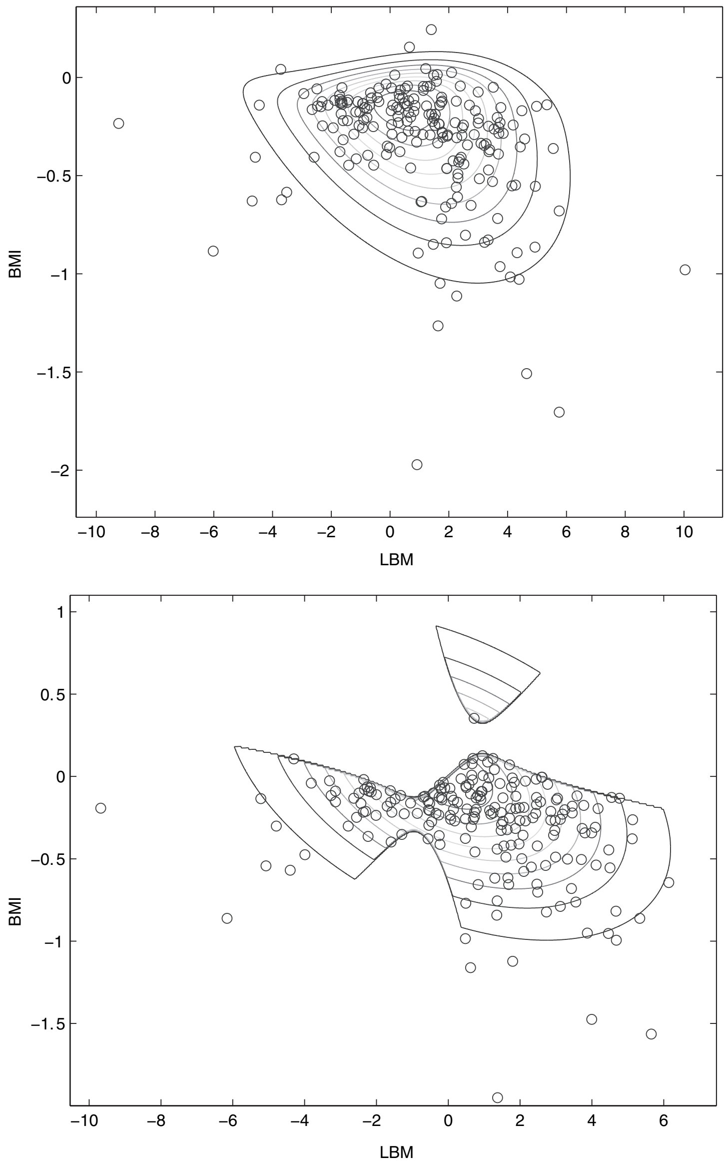


*Fig. 11.* AIS data: plot of residuals after fitting a linear model to the bivariate response variable (LBM, BMI) under the assumption of a bivariate flexible skew-normal distribution based on (44) for the error term: parsimonious model (top panel); full model (bottom panel).

significant on the $\chi^2_4$ scale. It seems that the inferential issues discussed in section 2.4 are still present in the multivariate case as can be seen in the bottom panel of Fig. 11, where some estimated coefficient in (44) are large and yield a 'half-normal-like' fit. The Healy's plots on the probability scale corresponding to Fig. 11 indicate that there is still room for improvement. Of course, one could also use a flexible skew-*t* distribution to model the distribution of the errors in a similar spirit as in section 4.2. However, it seems quite likely that the difference with a regular skew-*t* distribution will not be significant.

Another important application, as mentioned at the end of section 3.6, is concerned with continuous longitudinal responses in biomedical studies. The first attempt to use skew-normal and related distributions to relax the standard assumption of normality of the random effects in linear mixed models has been developed by Ma *et al.* (2004). Specifically, they modelled the random effects by means of flexible skew-normal and flexible skew-*t* distributions. Because the likelihood function does not have a closed form in this setting, they proposed inference based on the EM algorithm as well as inference via MCMC simulations in a Bayesian framework. In this approach, the choice of the degree of the polynomial involved in the skewing function was identified by means of model selection criteria. A simulation study indicated that a flexible model for the distribution of random effects in the linear mixed model results in more efficient estimators of the fixed effects and also more efficient estimators of the mean and the variance of the unobserved random effects. The special case where the distribution of the random effects is skew-normal permits a closed form expression for the likelihood function and has been studied by Arellano-Valle *et al.* (2005a). The use of skew-normal and related distributions in biomedical sciences has been further promoted by Sahu & Dey (2004) for the development of survival models with a skewed frailty, and by Chen (2004) for the construction of skewed link models for categorical response data.

Finally, I believe that skew-normal and related distributions can play an important role in applications arising from geosciences. As was mentioned at the end of section 3.6, Kim & Mallick (2004) used the skew-normal distribution to model spatial data. In the context of data assimilation, Naveau *et al.* (2004) developed a skewed Kalman filter for the analysis of climatic time series. Specifically, they studied the impact of strong but short-lived perturbations from large explosive volcanic eruptions on climate. The use of skew-normal distributions gave a more realistic representation of volcanic forcing than the normal distribution. Genton & Thompson (2003) used skew-elliptical time series to model sea levels and to evaluate the risk of coastal flooding in Charlottetown, Canada.

Besides the appeal of extending the classical multivariate normal theory, the current impetus for research in the field of skew-normal and related distributions is clearly driven by its potential for new applications. I expect to see many more novel applications in the near future!

## References

Bayarri, M. J. & DeGroot, M. (1992). A BAD view of weighted distributions and selection models. In *Bayesian Statistics 4* (eds. J. M. Bernardo, J. O. Berger, A. P. Dawid & A. F. M. Smith), 17–29. Oxford University Press, Oxford.

Bickel, P., Klaassen, C. A. J., Ritov, Y. & Wellner, J. A. (1993). *Efficient and adaptive inference in semiparametric models*. Johns Hopkins University Press, Baltimore, MD.

Chang, S.-M. & Genton, M. G. (2004). *Extreme values of skew-symmetric distributions*. Technical report, TAMU, College Station, TX.

Chen, M.-H. (2004). Skewed link models for categorical response data. In *Skew-elliptical distributions and their applications: a journey beyond normality* (ed. M. G. Genton), 131–151. Chapman & Hall/CRC, Boca Raton, FL.

Genton, M. G. & Thompson, K. R. (2003). Skew-elliptical time series with application to flooding risk. In *Time series analysis and applications to geophysical systems* (eds D. R. Brillinger, E. A. Robinson & F. P. Schoenberg), 169–186. IMA Volume in Mathematics and its Applications, Vol. 139. Springer, New York.

Hjort, N. L. & Glad, I. K. (1995). Nonparametric density estimation with a parametric start. *Ann. Statist.* **23**, 882–904.

Ma, Y., Genton, M. G. & Tsiatis, A. A. (2005). Locally efficient semiparametric estimators for generalized skew-elliptical distributions. *J. Am. Statist. Assoc.*, in press.

Ma, Y. & Hart, J. (2005). *Constrained local polynomial estimators for semiparametric skew-symmetric distributions*. Technical report, TAMU, College Station, TX.

Naveau, P., Genton, M. G. & Ammann, C. (2004). Time series analysis with a skewed Kalman filter. In *Skew-elliptical distributions and their applications: a journey beyond normality* (ed. M. G. Genton), 259–278. Chapman & Hall/CRC, Boca Raton, FL.

Parzen, E. (1979). Nonparametric statistical data modeling. *J. Am. Statist. Assoc.* **74**, 105–121.

Sahu, S. K. & Dey, D. K. (2004). On a Bayesian multivariate survival model with a skewed frailty. In *Skew-elliptical distributions and their applications: a journey beyond normality* (ed. M. G. Genton), 321–338. Chapman & Hall/CRC, Boca Raton, FL.

Tolstov, G. P. (1962). *Fourier Series*. Dover, New York.

Tong, H. (1990). *Nonlinear time series. A dynamical system approach*. Oxford University Press, New York.

Wang, J., Boyer, J. & Genton, M. G. (2004b). A note on an equivalence between chi-square and generalized skew-normal distributions. *Statist. Probab. Lett.* **66**, 395–398.

Wang, J. & Genton, M. G. (2005). The multivariate skew-slash distribution. *J. Statist. Plann. Inference*, in press.

*Received February 2005*

Marc G. Genton, Department of Statistics, Texas A&M University, College Station, TX 77843-3143, USA.
E-mail: genton@stat.tamu.edu