Some recent developments in the theory of distributions and their applications

Alcuni recenti sviluppi nella teoria delle distribuzioni e loro applicazioni

Adelchi Azzalini
Dipartimento di Scienze Statistiche, Università di Padova
e-mail: azzalini@stat.unipd.it

Riassunto: Viene presentata un'introduzione ad un particolare filone di letteratura che si colloca nell'ambito della teoria delle distribuzioni di probabilità per variabili casuali continue. Tale filone si caratterizza per il meccanismo costruttivo che, presa come 'base' una densità simmetrica, la modifica introducendovi ampi margini di flessibilità. Seppure tale modifica abbia una struttura matematica decisamente semplice, essa può agire in modo radicale sull'andamemento della funzione risultante, eppure allo stesso tempo essa è tale da mantenere valide talune delle proprietà formali della distribuzione 'base'. Dopo una breve rassegna dei principali aspetti di natura probabilistica, ci si concentra sugli aspetti applicativi, e si evidenziano le connessioni con altri filoni della letteratura.

Keywords: adaptive designs, compositional data, distribution theory, financial markets, kurtosis, parametric class, robustness, skewness, selective sampling, skew-normal distribution, skew-elliptical distribution, stochastic frontier analysis.

1. An ever-green well flourishing in Italy

The study of properties of probability distributions, generally in the sense of properties of suitable classes of distributions, has ever been a persistent theme of statistics and of applied probability. The work of Karl Pearson, especially in connection with his well-known system of frequency distributions, is often regarded as the starting point for this area of the literature. During the century which has elapsed since Pearson's construction, much work has been devoted to this area, although not always with the same momentum. With its ups and downs, this theme has always maintained some level of attention from the statistical community.

The main aims in the development of parametric families of distributions are: (i) wide flexibility of the distribution shape as its parameters vary, in order to accommodate different sort of data patterns; (ii) mathematical tractability and possibly the existence of a set of useful formal properties; (iii) existence of some mechanisms which generate random variables of the given type, to allow 'physically' motivated data modelling, as well as for random numbers generation.

Simultaneous and extensive fulfilment of the above requisites is not so easy and this fact largely explains why so many proposals have been put forward over the years. It is especially in the realm of multivariate analysis where meeting the above criteria has faced more problems. Hence, for the analysis of continuous multivariate data, the multivariate normal distribution still remains the basic ingredient, in most cases.

The present contribution provides an account of a relatively recent strand of the literature in the theory of distributions for continuous random variables. As explained later in more detail, the approach to be presented operates by starting from a symmetric, possibly multivariate, density and modifying it to increase its degree of flexibility. The set of perturbating functions must satisfy a simple condition to ensure that the resulting function is still a proper probability density. As a final product, families of remarkable flexibility can be constructed in a quite a simple manner. Since the identity function is included among the possible modifications, the original 'basis' function is still a member of the family. In particular, we can construct extensions of the normal class having the classical normal distribution as a proper member of the new family, not as a limiting case, which is what happens with many other parametric classes.

One point to be underlined is the fundamental role played by Italian statisticians in the development of this stream of literature, a claim widely documented by several important contributions mentioned later on. In addition, seminal ideas of the above sort have appeared in a very early piece of work due to an Italian statistician, Fernando de Helguero, although this fact has not been recognised until very recently. It is then a special pleasure to have been offered the chance of presenting this account at a meeting of the Italian Statistical Society.

Various review papers with similar target have been presented in the last few years; see specifically those of Arnold & Beaver (2002), Kotz & Vicari (2005), Azzalini (2005, 2006). Another relevant reference, whose role is markedly different, is the book edited by Genton (2004), containing a set of contributions of theoretical and of applied nature, many of which reflect work in progress at the time of writing. To avoid extensive overlap with these other presentations, the present one is specially targeted to the connections of existing material with certain topics in applied fields. This choice inevitably translates into a less exhaustive coverage of the probabilistic aspects and lack of discussion of the inferential aspects; for the missing items, see the review of Azzalini (2005) which in a sense can be regarded as a natural complement to the present account.

2. Skew-normal density on the real line

It is convenient from the expository viewpoint to start from the a very simple instance of the general context, since this case already contains in a simple form several of the concepts to be discussed at a later stage in greater generality. In addition, this option allows us to follow at large the time sequence of the literature.

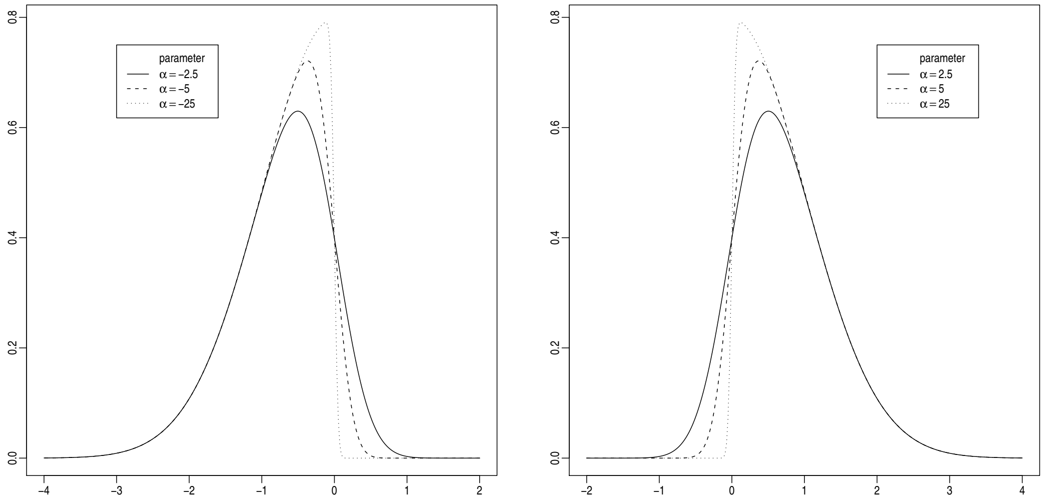
In probability theory, especially in many instances of parametric families of distributions, the normal or Gaussian family represents a suitable *limiting* case of a large variety of formulations. In applied statistical work, the normal family appears instead to represent a sort of *central* case, in the sense that the actual behaviour of observed data almost inevitably exhibit some form of departure from normality, in one or another of many possible directions, and the normal density represents the 'ideal shape' where these various departures find their point of equilibrium so to speak.

This remark motivates the construction of a parametric class of distribution functions for continuous random variables where the classical normal density is the 'central' case of a set of densities, as a suitable parameter moves along its range. A class of density functions accomplishing the above requirement in a simple form is given by

$$f_{\mathrm{SN}}(x) = 2\,\phi(x)\,\Phi(\alpha\,x)\,, \qquad (-\infty < x < \infty), \tag{1}$$

where $\phi(\cdot)$ and $\Phi(\cdot)$ denote the $N(0,1)$ density function and distribution function, respectively, and α is a parameter taking values in $(-\infty, \infty)$. It is obvious that setting $\alpha = 0$ reproduces the $N(0,1)$ family. It is less obvious that the '2' term in (1) provides the appropriate normalising constant for any choice of α; this fact can be deduced from a general proposition given later on. As α varies along the real lines, different shapes of the density are obtained, having positive skewness when $\alpha > 0$ and negative skewness when $\alpha < 0$; this fact gives a first explanation the term skew-normal (SN) for density (1). Figure 1 illustrates the behaviour of the SN density for some choices of α.

Figure 1: *SN density for a few values of the shape parameter (negative value of α on the left-hand side panel, positive values on the right-hand side).*



The adoption of the term 'skew-normal' for the family (1) is supported not only by the inclusion of the $N(0,1)$ density as a special case, but also by several analogies. Among these, a particularly noteworthy fact is that, if a random variable Z has density (1), then $Z^2 \sim \chi_1^2$, irrespectively of the value of α.

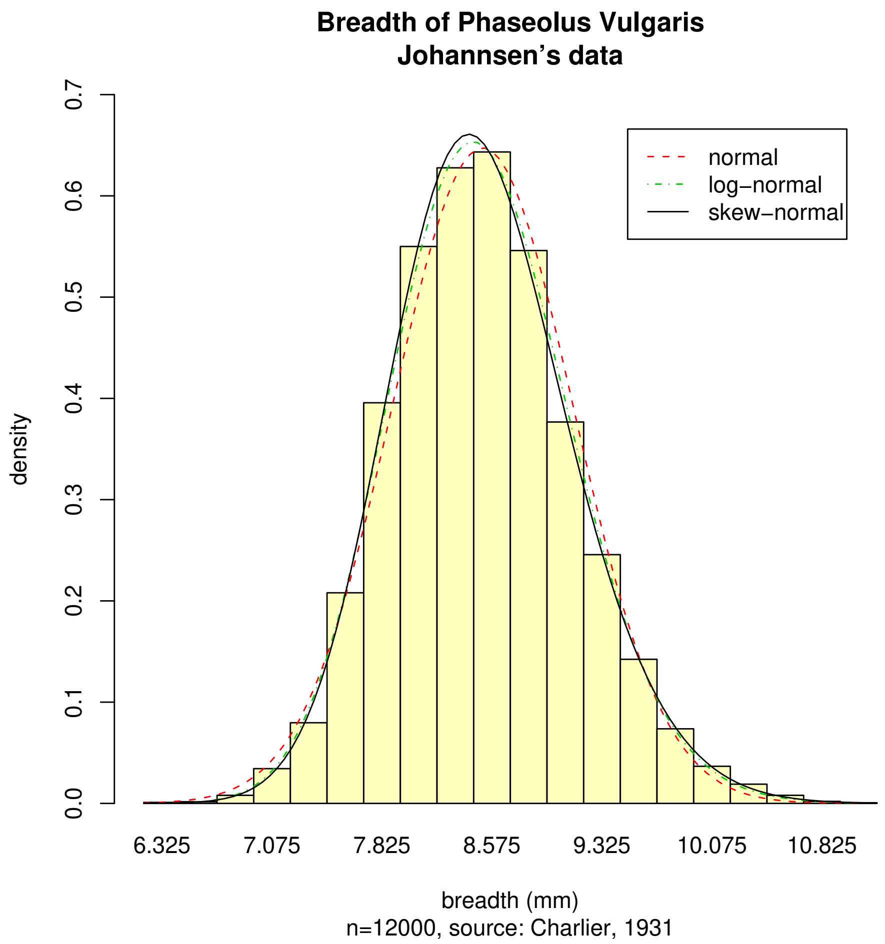
Obviously in practical statistical work, one introduces in addition a location and a scale parameter to the SN density. Specifically, if Z is a random variable with density function (1), then the transformation $Y = \xi + \omega Z$ (with $\omega > 0$) generates a three-parameter class of distributions. The adoption of the symbols ξ and ω in place of the more common μ and σ is made to mark the fact that they do no represent the mean and the standard deviation of the random variable, since

$$\mathbb{E}(Y) = \xi + \sqrt{2/\pi}\omega\delta, \qquad \mathrm{var}(Y) = \omega^2\{1 - (2/\pi)\delta^2\} \tag{2}$$

where $\delta = \alpha/\sqrt{1+\alpha^2}$.

To illustrate numerically the effect of the introduction of the shape parameter α, consider the data collected by Johannsen and presented by Charlier (1931, p. 73). They provide the frequency distribution of the breadth of $n = 12000$ common beans (*Phaseolus Vulgaris*) classified in 20 cells having width 0.25 mm each; the histogram of data is displayed in Figure 2. While the visual impression of the histogram suggests at first that the data could be considered normally distributed, a closer inspection gives different indication. The sample index of skewness is indeed small, since $\hat{\gamma}_1 = 0.288$, but the large

Figure 2: *Johannsen's data: breadth of* Phaseolus Vulgaris *and some fitted distributions.*



sample size transforms this value into a strong indication of departure from symmetry, since $\hat{\gamma}_1\sqrt{n/6} = 12.9$.

Indeed, the large sample size makes quite challenging to fit these data adequately. Figure 2 also displays, superimposed to the histogram, the curve obtained by fitting a normal, a log-normal and a skew-normal distribution; the estimation method was maximum likelihood for grouped data. Table 1 provides summary values of Pearson's X^2 statistic for a few parametric families which have been fitted to the data. The top three lines of the table are reproduced from Cramér (1946, p. 440) who has also considered the same data, except that he has grouped some of the cells to improve the χ^2 approximation to the X^2 distribution; hence his outcomes refer to 16 cells. The main conclusion to be extracted from Table 1 is that only the Edgeworth expansion to the second order term and the SN family provide an adequate numerical fit, but the former is based on one extra parameter. Considering in addition that the truncated Edgeworth expansion lacks any natural interpretation or a link to a generation mechanism, then the overall indication is slanted in favour of the SN family, for this simple example.

The above discussion has touched the issue of the forms of genesis of random variables with distribution (1). One of the attractive features of the SN class is that it can be originated in various different ways, a fact which will turn out to be relevant in connection

Table 1: *Johannsen's data: value of X^2 statistics and its observed significance level for some fitted distributions.*

family	X^2	d. f.	p-value
Normal	196.5	13	< 0.001
Edgeworth, first order	34.3	12	< 0.001
Edgeworth, second order	14.9	11	0.19
Log-normal	35.5	17	0.005
Skew-normal, 16 bins	14.2	12	0.29
Skew-normal, 20 bins	21.2	16	0.17

with data modelling, besides providing methods for random numbers generation.

Genesis via conditioning If (U_0, U_1) is a bivariate random variable having jointly normal distribution with standardised marginals and $\mathrm{corr}(U_0, U_1) = \delta$, then the distribution of $(U_1 | U_0 > 0)$ is as in (1), where $\alpha = \delta/\sqrt{1-\delta^2}$.

Genesis via a sum If V_0, V_1 are independent $N(0, 1)$ variates and $\delta \in (-1, 1)$, then

$$Z = \delta|V_0| + \sqrt{1-\delta^2}V_1 \tag{3}$$

has still density (1) with parameter $\alpha = \delta/\sqrt{1-\delta^2}$.

Genesis via maxima/minima If (U_0, U_1) is as indicated above, then

$$Z_1 = \min(U_0, U_1), \qquad Z_2 = \max(U_0, U_1)$$

have distribution of type (1) with parameters $\mp\sqrt{(1-\delta)/(1+\delta)}$.

The genesis via conditioning can be considered in a more general form associated to the conditional distribution $(U_1|U_0 + \tau > 0)$ for some real parameter τ, leading to the so-called 'extended skew-normal' density

$$f_{\mathrm{ESN}}(x) = \phi(x)\,\frac{\Phi(\tau\sqrt{1+\alpha^2} + \alpha\, x)}{\Phi(\tau)} \tag{4}$$

which reduces to (1) when $\tau = 0$. This extended form shares some but not all of the properties of the basic skew-normal density; in particular, the above-mentioned χ^2 property does not hold.

As for the historical development of the above results, and of the many others not mentioned here, appearances of (1) have occurred repeatedly in the literature as the result of the manipulation of some normal variates by one or another of the mechanisms described above, to tackle some specific applied problem. Two early references of this type are Birnbaum (1950) and Roberts (1966). Consideration of (1) as a distribution of independent interest, for the reasons and with the role described at the beginning of this section, is more recent, and it seems to start with Azzalini (1985, 1986). Further work on the properties of the class of skew-normal densities and on the associated inferential problems has been developed by a several authors, including Salvan (1986), Liseo (1990), Arnold et al. (1993), Chiogna (1998, 2005), Azzalini & Capitanio (1999), Pewsey (2000, 2003), Loperfido (2002), Monti (2003), Liseo & Loperfido (2006), Sartori (2006).

3. Distributions generated by perturbation of symmetry

The discussion of the previous section has focused on a very special, although important, case of a formulation which can be extended in many directions. The following lemma provides a useful tool for the construction of an extremely versatile family starting from a 'basis' symmetric density and applying a whole range of possible forms of perturbation.

Lemma 1 *If f_0 is a d-dimensional probability density function such that $f_0(x) = f_0(-x)$ for $x \in \mathbb{R}^d$, G is a one-dimensional differentiable distribution function such that G' is a density symmetric about 0, and w is real-valued function such that $w(-x) = -w(x)$ for all $x \in \mathbb{R}^d$, then*

$$f(x) = 2\, f_0(x)\, G\{w(x)\}\,, \qquad z \in \mathbb{R}^d, \tag{5}$$

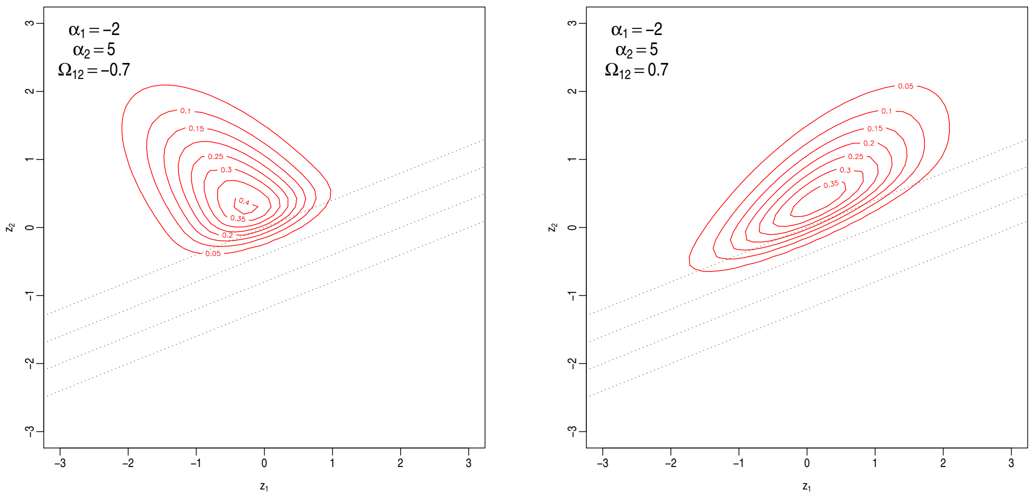
is a density function on $\mathbb{R}^d$.

As a simple yet important illustration of use of this result, consider the case where f_0 is $\phi_d(x;\Omega)$, the density function of a $N_d(0,\Omega)$ variable, and $w(x) = \alpha^\top \omega^{-1} x$ for some $\alpha \in \mathbb{R}^d$ and ω equal to the diagonal matrix formed by the standard deviations of Ω. Then the above lemma ensures that

$$f_{\mathrm{SN},d}(x) = 2\, \phi_d(x;\Omega)\, \Phi(\alpha^\top \omega^{-1}\, x) \tag{6}$$

is a proper density function, which is called multivariate skew-normal density. Clearly, if $d = 1$ and $\Omega = 1$, we return to (1). Figure 3 displays the shape of the bivariate SN distributions for two choices of the parameter set; in both cases $\Omega_{11} = \Omega_{22} = 1$, hence ω is the identity matrix. To illustrate how the weighting factor $\Phi(\alpha_1\, x_1 + \alpha_2 x_2)$ operates to down-weight the original normal density in (6), some lines where $\alpha_1\, x_1 + \alpha_2 x_2$ is constant are also shown.

Figure 3: *Contour levels of the bivariate SN density for two choices of the set of parameters. The dotted lines correspond to $\alpha_1\, x_1 + \alpha_2 x_2 = c$ for a few values of c.*



The next statement provides insight on the underlying stochastic mechanism associated to distributions of type (5); the subsequent one indicates that the χ^2 property of the SN distribution is a very special case of a far more general fact.

Stochastic representation Under the conditions of Lemma 1, if $X \sim G'$ and $Y \sim f_0$ are independent variables, then

$$Z = \begin{cases} Y & \text{if } X < w(Y), \\ -Y & \text{otherwise,} \end{cases} \tag{7}$$

has distribution (5).

Perturbation invariance Under the conditions of Lemma 1, if $Y \sim f_0$ and $Z \sim f$, then

$$t(Y) \stackrel{d}{=} t(Z) \tag{8}$$

for any real-valued function such that $t(x) = t(-x)$ for all $x \in \mathbb{R}^d$, irrespectively of the choice of G and w.

In the last few years, much work has been developed directly or indirectly linked to form (5), with special attention to the case when f_0 is of elliptical type, that is a density of the form

$$f_0(x) = \text{constant} \times \tilde{f}(x^\top \Omega^{-1} x)$$

which takes on constant value on the ellipsoids where $x^\top \Omega^{-1} x = c$, for any positive constant c and some fixed matrix $\Omega > 0$. By a suitable choice of the so-called generator $\tilde{f}$, we can obtain densities whose tail behaviour can be regulated by some parameter. Within this context, special attention has been given to the case where $w(x)$ is a linear function of x, leading to the construction of so-called skew-elliptical distributions.

By combining an elliptical density with the perturbation mechanism $G\{w(x)\}$, one provides the means for regulating simultaneously skewness and kurtosis of a distribution. This process is particularly attractive since one operates starting from a specific 'basis' function f_0, often the normal one; hence one can work by successive generalisation of a familiar model, without need of jumping to an entirely different formulation if skewness and kurtosis are an issue. Within the skew-elliptical family, the multivariate skew-t distribution appears to be a particularly appealing instance, given that the multivariate t distribution has already been used by various author for flexible modelling of long-tailed distributions.

This formulation effectively provides an approach to robust statistical inference, alternative to the traditional one based on M-estimators and alike. In case the error distribution is not symmetric, this approach offers the advantage of an explicit specification of the quantities being estimated, namely the parameters of the distribution, while M-estimators converge to the implicit solution of an equation.

The contour levels of a skew-elliptical density are qualitatively similar to those in Figure 3, except that the amount of spacing between the contour levels is regulated by $\tilde{f}$ instead of the bivariate normal density.

Clearly, one is not restricted to linear functions in the choice of $w(x)$. If this function is highly non-linear in x, then the corresponding contour levels may delimit non-convex regions, or even non-connected sets, i.e. the density function can be multimodal. This direction of work is potentially very fruitful, but little explored so far.

The above discussion summarises work which has been largely developed by Azzalini & Dalla Valle (1996), Azzalini & Capitanio (1999, 2003), Branco & Dey (2001), Loperfido (2001), Gupta & Huang (2002), Gupta (2003), Liseo & Loperfido (2003), DiCiccio & Monti (2004), Gupta et al. (2004), Genton & Loperfido (2005); see also various contributions to the book edited by Genton (2004).

Another important tool for building families of distributions is the following very general expression whose importance has been stressed by Arellano-Valle et al. (2002).

Lemma 2 *Suppose the (U_0, U_1) is a $(m + d)$-dimensional random vector whose U_1 component has marginal density f_{U_1}, then*

$$f(x) = f_{U_1}(x)\,\frac{\Pr\{U_0 > 0|U_1 = x\}}{\Pr\{U_0 > 0\}}, \qquad (x \in \mathbb{R}^d), \tag{9}$$

where the notation $U_0 > 0$ means that the inequality must hold for each of the m components, is a density function on $\mathbb{R}^d$.

This formula is of complete generality and it does not involve assumptions of symmetry or others, but it requires to compute two integrals which are feasible only in some favourable cases. A simple instance of this sort is offered by (4) and, via (9), that formulation can be extended to the case where (U_0, U_1) is a multivariate normal vector with $m \geq 1$ and $d \geq 1$; see Arellano-Valle & Azzalini (2006) for a discussion which encompasses earlier formulations and extends it to elliptical components.

This section cannot be closed without a mention of the pioneering work of Fernardo de Helguero, a young Italian scientist that in 1908 presented a paper which contained very important seminal ideas. To fully appreciate the value of his innovative formulation, one must bear in mind that in those years the Pearson's system of distributions and the Edgeworth expansion were considered the key approaches when one had to fit a distribution to data. In this historical context, de Helguero starts off arguing as follows.

> *Il compito della statistica nelle sue varie applicazioni alle scienze economiche e biologiche non consiste solo nel determinare la legge di dipendenza dei diversi valori ed esprimerla con pochi numeri, ma anche nel fornire un aiuto allo studioso che vuole cercare le cause della variazione e le loro modificazioni.*
>
> *(. . .) Invece le curve teoriche studiate dal* PEARSON *e dall'*EDGEWORTH *per la perequazione delle statistiche abnormali di materiale omogeneo (. . .) nulla ci fanno sapere sulla legge di dipendenza, quasi nulla sulle relazioni colla curva normale che pure deve essere considerata come fondamentale.*
>
> *Io penso che miglior aiuto per lo studioso potrebbero essere delle equazioni che supponessero una perturbazione della variabilità normale per opera di cause estranee.*

The formulation could not be more innovative and lucid: our target is not simply to search for a numerical agreement between the observed data and the fitted distribution, but we must also attempt to model the departure from normality (which is regarded as the underlying mechanism) aiming at understanding which mechanism has perturbated the original normal distribution and produced the observed departure from normality. Of this broad program de Helguero examines two possible directions of work, one leading to consideration of the mixture of two normal populations, and another one which he called 'Curves perturbed by selection'. This second instance is of our concern here, because it is directly linked to Lemma 1 above and the corresponding stochastic representation

(7). Specifically, de Helguero assumed that data are generated from an underlying normal distribution but, because of a selection process, a datum x is actually observed with a probability depending on x itself, through a function which he assumes to be a linear function of x. If this formulation is phrased in terms of (5), then f_0 is the $N(0,1)$ density, G is the distribution with uniform density over $(-1,1)$ and $w(x) = \alpha x$, arguing on the basis of a construction which is essentially the top branch of (7). Not knowing Lemma 1, de Helguero obtained an approximate expression for the normalising constant; he also derived a recurrence relationship for moments which allowed him to fit his new distribution to some data on wages of workers.

The basis of a major change in statistical thinking had been formulated, and the initial part of the path had been developed. Unfortunately, Fernando de Helguero died a few month later in the earthquake of Messina, 28th December 1908, and his ideas passed unnoticed for a long time.

4. Applications and connections with other topics

The potential applications of these probabilistic tools are very numerous. Here we focus on the most popular ones and those having a direct link with the various forms of stochastic representations described above. An associated motivation of this section is to highlight the connections existing between applications area which apparently are totally unrelated.

4.1. Selective sampling

The concept of selective sampling has been mentioned repeatedly in the preceding discussion, and the connection with Heckman's (1976, 1979) model is transparent when one compares certain ingredients. In its simplest form, Heckman's formulation starts from the relationships, supposed to hold for a generic individual of the population,

$$Y_0 = X_0\beta_0 + U_0, \quad Y_1 = X_1\beta_1 + U_1,$$

where (U_0, U_1) is a bivariate normal variable, and β_0, β_1 are unknown parameters. The X's and Y's are observables, at least in principle but, because of a selection mechanism in the sampling process, we observe Y_1 only when $Y_0 > 0$. The construction is then the same of the genesis by conditioning leading to the extended SN distribution (4), as remarked by Copas & Li (1997).

Establishing this link between the two streams of literature allows to transfer theoretical results from the theory of distributions to the area of models for selective sampling. An instance of this type is the work of Grilli & Rampichini (2005), which involves a distribution of type (9) with components of multivariate normal type. One can envisage that additional flexibility can be brought in this sort of modelling by considering other types of components, especially those of elliptical form which allow regulation of the tail behaviour.

4.2. Stochastic frontier models

The theme of stochastic frontier models refers to a special formulation of regression models of type

$$\begin{aligned} Y &= x^\top\beta + W_1 - |W_0| && (10)\\ &= x^\top\beta + \omega(\sqrt{1-\delta^2}\,V_1 - \delta|V_0|) && (11) \end{aligned}$$

where the W_j's are two independent 0-mean random terms; in the simplest version of the model, they are taken to be normally distributed, with variance σ_j^2, say, for $j = 0, 1$. The motivation of this formulation lies in the analysis of productivity of units for which the vector x denotes the input ingredients of the production process, and Y the corresponding output. The key distinction between (10) and a standard regression model is the presence of the additional error term $|W_0|$ which represents the inefficiency of the production unit with respect to the output level of full efficiency, $x^\top\beta$. Clearly the ingredients of (10) can be re-arranged into the form (11) for suitably defined quantities, and the latter expression is of type (3).

Many variants and extensions of (11) have been examined in the econometric literature, for instance assuming that $|W_0|$ is exponentially distributed, to allow for a long-tail distribution of the inefficiency. Again, the results discussed in Section 3 can offer a fairly general framework for more general formulations. Specifically, the skew-elliptical families provide a very flexible formulation, with the advantage of enjoying a set of convenient formal properties. A first exploration in this direction has been pursued by Tancredi (2002).

4.3. Observation of the maximal component

In a number of situations, especially of the medical context, observations are taken in pairs, but one is especially interested in the maximum value (or the minimum, in other cases). For instance, in ophthalmology, the visual acuity of both eyes is commonly measured, but the maximum of these two values has a special relevance and it can be regarded as the single response value, for certain purposes.

Under the assumption of joint normality and equal marginal distribution of the two measurements, the distribution of the maximum value is of skew-normal type, if one recalls the third mechanism of genesis described in Section 2. In fact, Roberts (1966) has obtained what we now call the skew-normal distribution by consideration of the maximal measurement in the study of twins. Recently, Loperfido (2002, 2005) has re-examined and extended this result, by including consideration of covariate values and longitudinal observations.

4.4. Financial markets

In financial applications, presence of long tails in the observed distribution is almost ubiquitous, and data modelling strongly requires a corresponding formulation for the error term, involving for instance a stable or a Student's t distribution. In recent times, consideration of skewness has come into the picture, for a more accurate data modelling. Besides support from empirical observations, there are qualitative arguments which motivate this step, since financial markets react in opposite direction but with different amplitude to positive and to negative information coming for instance from other markets.

Adcock (2004) has shown how to extend a standard economic formulation from the normal assumption of the stochastic component to a more realistic skew-normal assumption, and still maintain the key properties of the economic formulation.

The Black-Litterman technique allows managers to construct portfolios that account for their views on a set of expected returns, under normality assumption. Meucci (2006) pick-ups the idea of regulating simultaneously skewness and kurtosis and extends the Black-Litterman framework to a much more general market distribution.

Cappuccio et al. (2004) and De Luca et al. (2005) have developed variants of classical models for financial time series to accommodate the presence of skewness via ingredients of the form described above. In particular, the multivariate skew-GARCH model proposed De Luca et al. (2005) links in a natural way economic considerations to properties of the skew-normal distribution.

4.5. Adaptive designs in clinical trials

Due to the steady increase in the astronomical cost of clinical trials conducted for drug development, a theme currently of interest in medical statistics goes under the heading of adaptive designs, which attempt to limit these costs. Within this context, one direction of work is the combination of the outcome from a phase II study with the one from a phase III study. There are two complications to handle here: one is that the phase III study is conducted conditionally on successful outcome of the phase II study; the other one is that the two studies often consider a different endpoint. The conditioning mechanism in action suggests that, under normality assumption of the variables, a skew-normal component of the resulting likelihood function can be envisaged; Azzalini & Bacchieri (2006) work out the detailed construction.

4.6. Compositional data

Compositional data appear in very different areas of applications, but it is in the geological context where they represent the regular situation. For the analysis of this kind of data, a standard device is to transform the $d+1$ original components belonging to the simplex to d components in $\mathbb{R}^d$ using the additive log-ratio transform, followed by analysis based on methods for normal data.

After the additive log-ratio transformation has taken place, normality assumption can be replaced by assumption of skew-normality on the transformed data, to improve adequacy in data fitting. This assumption on $\mathbb{R}^d$ induces back a distribution on the simplex which enjoys certain formal desirable properties, which are due to the properties of closure under marginalisation and affine transformation of the skew-normal distribution, inducing some corresponding properties on the simplex. For a discussion of these aspects, see Aitchison et al. (2003).

Acknowledgement I am grateful to Professor Donato Michele Cifarelli for kindly drawing my attention to the work of Fernando de Helguero.

References

Adcock, C. (2004). Capital asset pricing in UK stocks under the multivariate skew-normal distribution. In M. G. Genton (Ed.), *Skew-elliptical distributions and their applications: a journey beyond normality* chapter 11, (pp. 191–204). Chapman & Hall/CRC.

Aitchison, J., Mateu-Figueras, G., & Ng, K. W. (2003). Characterization of distributional forms for compositional data and associated distributional tests. *Math. Geol.*, 35(6), 667–680.

Arellano-Valle, R. B., del Pino, G., & San Martín, E. (2002). Definition and probabilistic properties of skew-distributions. *Statist. Probab. Lett.*, 58, 111–121.

Arellano-Valle, R. B. & Azzalini, A. (2006). On the unification of families of skew-normal distributions. *Scand. J. Statist.*, 33, in press.

Arnold, B. C., Beaver, R. J., Groeneveld, R. A., & Meeker, W. Q. (1993). The nontruncated marginal of a truncated bivariate normal distribution. *Psychometrika*, 58, 471–478.

Arnold, B. C. & Beaver, R. J. (2002). Skewed multivariate models related to hidden truncation and/or selective reporting (with discussion). *Test*, 11, 7–54.

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

Azzalini, A. (1986). Further results on a class of distributions which includes the normal ones. *Statistica*, XLVI, 199–208.

Azzalini, A. (2005). The skew-normal distribution and related multivariate families (with discussion). *Scand. J. Statist.*, 32, 159–188 (C/R 189–200).

Azzalini, A. (2006). Skew-normal family of distributions. In S. Kotz, N. Balakrishnan, C. B. Read, & B. Vidakovic (Eds.), *Encyclopedia of Statistical Sciences*, second edition, volume 12 (pp. 7780–7785). J. Wiley & Sons, New York.

Azzalini, A. & Bacchieri, A. (2006). A proposal in adaptive designs: combining phase II and phase III in drug developement. Work in progress.

Azzalini, A. & Capitanio, A. (1999). Statistical applications of the multivariate skew normal distributions. *J. R. Stat. Soc., ser. B*, 61, 579–602.

Azzalini, A. & Capitanio, A. (2003). Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t distribution. *J. R. Stat. Soc., ser. B*, 65, 367–389.

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

Birnbaum, Z. W. (1950). Effect of linear truncation on a multinormal population. *Ann. Math. Statist.*, 21, 272–279.

Branco, M. D. & Dey, D. K. (2001). A general class of multivariate skew-elliptical distributions. *J. Multivariate Anal.*, 79, 99–113.

Cappuccio, N., Lubian, D., & Raggi, D. (2004). MCMC Bayesian estimation of a skew-GED stochastic volatility model. *Studies in nonlinear dynamics and econometrics*, 8(2). `http://www.bepress.com/snde/vol8/iss2/art6`.

Charlier, C. V. L. (1931). *Vorlesungen über die Grundzüge der mathematischen Statistik*. C. W. K. Gleerups Vörlag, Lund.

Chiogna, M. (1998). Some results on the scalar skew-normal distribution. *J. Ital. Statist. Soc*, 7, 1–13.

Chiogna, M. (2005). A note on the asymptotic distribution of the maximum likelihood estimator for the scalar skew-normal distribution. *Stat. Meth. & Appl.*, 14, 331–341.

Copas, J. B. & Li, H. G. (1997). Inference for non-random samples (with discussion). *J. R. Statist. Soc.* B, 59, 55–95.

Cramér, H. (1946). *The mathematical methods of statistics*. Princeton University Press, Princeton.

de Helguero, F. (1909). Sulla rappresentazione analitica delle curve abnormali. In G. Castelnuovo (Ed.), *Atti del IV Congresso Internazionale dei Matematici (Roma, 6–11 Aprile 1908)*, volume III (sez. III-B). Roma: R. Accademia dei Lincei.

De Luca, G., Genton, M. G., & Loperfido, N. (2005). A multivariate skew-GARCH model. *Advances in Econometrics*, 20, 33–56.

DiCiccio, T. J. & Monti, A. C. (2004). Inferential aspects of the skew exponential power distribution. *J. Amer. Statist. Assoc.*, 99, 439–450.

Genton, M. G., Ed. (2004). *Skew-elliptical distributions and their applications: a journey beyond normality*. Chapman & Hall/CRC, Boca Raton.

Genton, M. G. & Loperfido, N. (2005). Generalized skew-elliptical distributions and their quadratic forms. *Ann. Inst. Statist. Math.*, 57, 389–401.

Grilli, L. & Rampichini, C. (2005). Sample selection in multilevel models. Presented ad the *5th International Amsterdam Conference on Multilevel Analysis*. Amsterdam, 21-22 March 2005.

Gupta, A. K. (2003). Multivariate skew t-distribution. *Statistics*, 37(4), 359–363.

Gupta, A. K. & Chen, J. T. (2004). A class of multivariate skew-normal models. *Ann. Inst. Statist. Math.*, 56, 305–315.

Gupta, A. K., González-Farías, G., & Domínguez-Molina, J. A. (2004). A multivariate skew normal distribution. *J. Multivariate Anal.*, 89(1), 181–190.

Heckman, J. J. (1976). The common structure of statistical models of truncation, sample selection and limited dependent variables, and a simple estimator for such models. *Ann. Econ. Socl. Measmnt.*, 5, 475–492.

Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica*, 47, 153–161.

Kotz, S. & Vicari, D. (2005). Survey of developments in the theory of continuous skewed distributions. *Metron*, LXIII, 225–261.

Liseo, B. (1990). La classe delle densità normali sghembe: aspetti inferenziali da un punto di vista bayesiano. *Statistica*, L, 59–70.

Liseo, B. & Loperfido, N. (2003). A Bayesian interpretation of the multivariate skew-normal distribution. *Statist. Probab. Lett.*, 61, 395–401.

Liseo, B. & Loperfido, N. (2006). A note on reference priors for the scalar skew-normal distribution. *J. Statist. Plann. Inference*, 136, 373–389.

Loperfido, N. (2001). Quadratic forms of skew-normal random vectors. *Statist. Probab. Lett.*, 54, 381–387.

Loperfido, N. (2002). Statistical implications of selectively reported inferential results. *Statist. Probab. Lett.*, 56, 13–22.

Loperfido, N. (2005). Modelling maxima of longitudinal contralateral observations. Submitted for publication.

Meucci, A. (2006). Beyond Black-Litterman: views on non-normal markets. *Risk Magazine*, 19, 87–92.

Monti, A. C. (2003). A note on the estimation of the skew normal and the skew exponential power distributions. *Metron* **XLI**, 205–219.

Pewsey, A. (2000). Problems of inference for Azzalini's skew-normal distribution. *Journal of Applied Statistics*, 27, 859–770.

Pewsey, A. (2003). The characteristic functions of the skew-normal and wrapped skew-normal distributions. In *27 Congreso Nacional de Estadistica e Investigación Operativa* (pp. 4383–4386). Lleida, España.

Roberts, C. (1966). A correlation model useful in the study of twins. *J. Amer. Statist. Assoc.*, 61, 1184–1190.

Salvan, A. (1986). Test localmente più potenti tra gli invarianti per la verifica dell'ipotesi di normalità. In *Atti della XXXIII Riunione Scientifica della Società Italiana di Statistica*, volume II (pp. 173–179). Bari: Cacucci.

Sartori, N. (2006). Bias prevention of maximum likelihood estimates for scalar skew normal and skew t distributions. *J. Statist. Plann. Inference*, in press.

Tancredi, A. (2002). Accounting for heavy tails in stochastic frontier models. Working paper. Dipartimento di Scienze Statistiche, Università di Padova, n. 2002.16.