International Statistical Review (2015), 83, 2, 198–202 doi:10.1111/insr.12072

Adelchi Azzalini[1] and Marc G. Genton[2]

[1] *Department of Statistical Sciences, University of Padua, Padua, Italy*
E-mail: azzalini@stat.unipd.it

[2] *CEMSE Division, King Abdullah University of Science and Technology, Thuwal 23955-6900, Saudi Arabia*
E-mail: marc.genton@kaust.edu.sa

1 Some General Preliminaries

First of all, we would like to congratulate the author for building a comprehensive framework that highlights the interconnections between co-existing and apparently unrelated formulations of distributions with shape parameters. This course of action helps people to grasp in detail the nature and the role of the numerous proposals, often a source of confusion for the non-specialist. Some authors seem to move in the opposite direction.

Of the many remarks, we especially concur with the final sentence of Section 7, that is, 'The ongoing challenge is to extract from the overwhelming plethora of possibilities those relatively few with the best and most appropriate properties that are of real potential value in practical applications'. This consideration is sometimes neglected in the flourishing industry for production of new distributions.

Because a number of our comments made to an earlier version of the paper have already been incorporated in the current version, we shift the aim towards slightly more personal views and, to some extent, also widen the scope beyond discussion of Professor Jones's paper.

2 Multivariate Distributions with Shape Parameters

The focus of the paper is explicitly stated to be on continuous univariate distributions. In a multivariate setting, univariate marginals must be combined, typically using some existing general mechanism. The chief tool of this type is of course the copula mechanism, which recurs at various places in the paper, but other ways of combining components are mentioned in the final paragraph of Section 5.1, especially in connection with cases where variables of mixed type are observed.

Our standpoint on the way of tackling multivariate observations is somewhat divergent from the one of Professor Jones for the reasons discussed later. We first introduce a general view and then elaborate on this in a set of more specific points.

The course taken in selecting an appropriate family of distributions for a given applied problem must not be regarded as an isolated step but must be embedded into the more comprehensive and complex problem of model building. We avoid entering a discussion of this even wider theme and refer instead to the magisterial guidelines compiled by Cox (1997), which are reproduced in Figure 1. A key concept in this scheme is the role of subject-matter theory or knowledge and its interplay with the data analysis process. We only add that the selection of

Desiderata for a probabilistic model can be formulated in various ways. One such is the following, not all points applying in any particular context:

1. the model should establish a link with underlying substantive knowledge or theory;
2. the model should allow comparisons with previous related studies of the topic;
3. the model should be consistent with or suggest a possible process that might have generated the data;
4. parameters defining primary features of the system should have individually clear subject matter interpretations;
5. the error structure should be represented sufficiently realistically that meaningful measures of precision are obtained for the primary comparisons;
6. the fit to data should be adequate.

Figure 1. *A passage from Cox (1997) on desiderata for a probabilistic model.*

a working family of distributions, as a part of the model-building strategy, should be framed within the same logic. The goodness of fit to the data is definitely not the only criterion and, presumably, it is not by chance that this aspect is the last one in the previously quoted list. On the contrary, it is quite commonly the case that a certain proposal in the literature is claimed to be preferable to others on the sole consideration of better data fitting in some selected examples. We do not deny that, in certain cases, the empirical data fitting may be the only point of interest, but this does not provide the basis for a general norm.

The general view just described expands into a number of more specific implications, which follow next. An accompanying note is that they do not claim to be of universal validity, never mind prescriptive, given the multitude of different forms that practical problems can take.

Mixed-type variables When the problem under consideration involves variables of different nature, that is, of discrete, unbounded and bounded continuous, ordered and unordered categorical type or at least more than one of these types, the existing techniques for generating distributions cannot be employed, as they operate on variables of homogeneous type. The classes of distributions produced by these mechanisms are typically of continuous type, but some recent work has dealt with discrete distributions; specific contributions are those of Genest & Nešlehová (2007) in the copula context and of Azzalini & Regoli (2014) for the Family 1 formulation.

It is hoped that these recent directions will expand further and possibly allow for mixed-type distributions. For the time being, when working with mixed-type variables, one has to resort on methods to combine these heterogeneous variables, such as a specification via conditional distributions. The following points apply then to groups of homogeneous variables, especially the continuous ones.

Continuous 'one-lump' multivariate distributions However, the focus of discussion here is on cases where the variables are more homogeneous; in fact, the emphasis in the paper under discussion is almost entirely on continuous variables. In these situations, the 'one-lump' multivariate distribution, as denoted in Section 5.1 of the paper, is a useful concept. It is particularly so in those non-rare cases where the variables are of homogeneous nature and with closely connected role; we are referring to cases such as a set of morphological measurements on an animal or a plant (Fisher's Iris data are the obvious example), the returns of a set of financial assets and wind speed measured on a group of weather stations. The more the variables under consideration are of homogeneous nature, the more the formulation of a 'one-lump' distribution is conceptually appealing and the link with underlying theory or knowledge is facilitated, provided of course a suitable choice of the distribution is made.

Flexibility versus interpretability A copula-like approach is operationally very flexible and can provide good and sometimes excellent empirical fit to the data, but the joint distribution has an 'artificial flavour'. Indeed, the copula can be chosen independently of the marginals, and this extreme level of flexibility also has the drawback of losing the aspect of interpretation and possible link to the data generation mechanism. For example, there are situations where the shape of the univariate/multivariate distribution is a consequence of one such mechanism. The formulation of Family 1 provides various such readings, thanks to its variety of stochastic representations; of these, one holds in complete generality, some others apply to important subclasses of the construction. An instance of this type of connection arises from the conditioning mechanism intimately linked to this construction, which has a natural connection with non-random sampling in the form of the Heckman selection model and its extension studied by Marchenko & Genton (2012) and references therein. Another exemplification is provided by the additive representation of certain distributions of Family 1, which has a direct interpretation in the context of stochastic frontier models in the econometric literature, again leading to extensions such as those studied by Domínguez-Molina *et al.* (2007). Of course, one can test the structure of the copula and identify classes of relevant copulas (e.g. Li & Genton, 2013), but this still does not help in selecting an interpretable copula. How does a copula-type construction pass requirements (i)–(iv) in Section 1 of Professor Jones's paper (as well as others mentioned here) when referred to the multivariate distribution?

Role of formal properties A major factor for the key role of the normal distribution in multivariate analysis is its mathematical tractability and a rich set of formal properties, such as closure under marginalisation, affine transformations and conditioning, and distributional properties of quadratic forms. All these features have allowed for important advances in many branches of multivariate analysis.

Still, over the years, it has emerged that the normal distribution is, in a range of cases, not fully adequate for the purpose of data fitting. This problem is becoming more and more evident as modern facilities allow to collect larger data sets so that the limitation of the normal distribution in this respect is accentuated. This situation has given impulse, especially in the last 15 years or so, to the exploration of other directions in handling the stochastic component of a model, alternative to normality.

This process is appropriate, but its motivations do not imply that mathematical tractability and convenient formal properties have become irrelevant aspects, shifting all interest on accuracy in data fitting. We can search for distributions that are more flexible than the normal and still are mathematically convenient. Clearly, the term 'normal' here can be replaced by some other names of classical distributions.

Formulations that illustrate the extension of normal-based constructions to more flexible ones, as envisaged earlier, are provided by the work of Adcock & Shutes (1999) and Carmichael & Coën (2013) exploiting properties of closure of the skew-normal family with respect to affine transformations that are at the basis of the original formulations in finance theory linked to the normal distribution. More examples of this sort exist; for instance, the work of Adcock (2010) provides an extension of similar nature with respect to the basic Student's t construction.

Therefore, one can ask whether Table 1 still holds in the multivariate setting and whether it needs to be extended to include properties such as closure under marginalisation and conditioning, linear transformations, and moments and global measures of multivariate skewness and kurtosis. We understand the latter properties for the 'Family 1' type of construction, possibly in some of its variant forms, but we are less clear about other families.

3 Miscellaneous Points

Here are a set of scattered remarks with no particular order or connections.

(1) Explicit expression of the mode has been much emphasised; see especially remark 'Rows 1 and 2' in Section 4. Undoubtedly, if present, this is a welcome feature, but little is lost when it is not present. The mode is not a parameter of major relevance in most statistical investigations. A related point of greater relevance is uniqueness of the mode. This requirement is especially meaningful if we are fitting a finite mixture of distributions where each component represents a sub-population. However, fitting finite mixtures can be computationally problematic, and this fact may motivate other directions. Also, if an interpretation via sub-populations is not envisaged, then unimodality is not an issue. For instance, in the work of Mazzuco & Scarpa (2015), age-specific fertility rates are fitted using a multimodal density multiplied by a positive parameter, and here, local maxima are not associated to subpopulations.

(2) Flexibility is a key concept with many facets, some of which have already been discussed. One aspect not always perceived is the potential harm caused by uncontrolled search for maximal flexibility and generality of the proposed families of distributions. In essence, the risk is a form of overfitting. In extreme cases, this has led to unidentifiable parametric families; some less extreme cases manifest themselves as instability of inference and flat likelihoods, at least for certain combinations of parameter values. In the stream of literature more familiar to us, relatively mild examples of this situation are the extended skew-normal and extended skew-*t* distributions, but the problem can arise in any approach. Many other and even more flexible distributions have been put forward, but the majority of them have never been tried 'on the road', besides an initial, merely illustrative, 'test drive'. A better understanding of this issue would be useful.

(3) As for the question discussed in paragraph 'Row 8' of Section 4, reasons for not regarding $\hat{\lambda} = \pm\infty$ as a satisfactory estimate, at least not on a regular basis, have been given in the references provided at the end of quoted paragraph, so we do not reproduce them here.

(4) There is an extension of the Family 1 arising when the truncation point of the underlying latent variable is not equal to its mean value, so that the normalizing constant is not 2. Much initial work in this logic has been summarised by Arnold & Beaver (2002), but this type of formulation has been subsequently much extended: Arellano-Valle & Azzalini (2006) deal with multiple latent constraints in the normal and the elliptical case, and these directions have been developed substantially by Arellano-Valle & Genton (2010a). Semiparametric modelling in this general setting was investigated by Ma *et al.* (2013), whereas Genton *et al.* (2012) investigated the special case of odd modulating function w when the normalizing constant is 2. The invariance property of Family 1 can be extended to some settings related to unified skew-elliptical distributions (Arellano-Valle & Genton, 2010b). A brief account of these ideas and additional references are provided in Sections 1.3.1 and 7.1 of Azzalini & Capitanio (2014).

(5) An interesting member of Family 2 is the class of Lambert W random variables (Goerg, 2011). It has a link with Tukey's h distribution; see in particular Section 8 of Goerg (2011).

Finally, we would like to underline that, even if on some aspects we have stated a diverging viewpoint from the one of Professor Jones, this does not affect our appreciation for this valuable work as we expressed at the beginning.

References

Adcock, C.J. (2010). Asset pricing and portfolio selection based on the multivariate extended skew-Student-*t* distribution. *Ann. Oper. Res.*, **176**, 221–234.

Adcock, C.J. & Shutes, K. (1999). Portfolio selection based on the multivariate-skew normal distribution. In *Financial Modelling*, Ed. A.M.J. Skulimowski, pp. 167–177. Kraków: Progress and Business Publishers. Available in 2001.

Arellano-Valle, R.B. & Azzalini, A. (2006). On the unification of families of skew-normal distributions. *Scand. J. Statist.*, **33**, 561–574.

Arellano-Valle, R.B. & Genton, M.G. (2010a). Multivariate unified skew-elliptical distributions. *Chil. J. Statist.*, **1**, 17–33.

Arellano-Valle, R.B. & Genton, M.G. (2010b). An invariance property of quadratic forms in random vectors with a selection distribution, with application to sample variogram and covariogram estimators. *Ann. Inst. Stat. Math.*, **62**, 363–381.

Azzalini, A. & Regoli, G. (2014). Modulation of symmetry for discrete variables and some extensions. *Stat*, **3**, 56–67.

Carmichael, B. & Coën, A. (2013). Asset pricing with skewed-normal return. *Finance Res. Lett.*, **10**, 50–57.

Cox, D.R. (1997). The current position of statistics: a personal view. *Int. Statist. Rev.*, **65**, 261–290.

Domínguez-Molina, J.A., González-Farías, G., Ramos-Quiroga, R. & Gupta, A.K. (2007). A matrix variate closed skew-normal distribution with applications to stochastic frontier analysis. *Commun. Statist. – Theory and Methods*, **36**, 1671–1703.

Genest, C. & Nešlehová, J. (2007). A primer on copulas for count data. *ASTIN Bull.*, **37**, 475–515.

Genton, M.G., Kim, M. & Ma, Y. (2012). Semiparametric location estimation under non-random sampling. *Stat*, **1**, 1–11.

Goerg, G.M. (2011). Lambert W random variables – a new family of generalized skewed distributions with applications to risk estimation. *Ann. Appl. Statist.*, **5**, 2197–2230.

Li, B. & Genton, M.G. (2013). Nonparametric identification of copula structures. *J. Amer. Statist. Assoc.*, **108**, 666–675.

Mazzuco, S. & Scarpa, B. (2015). Fitting age-specific fertility rates by a flexible generalized skew-normal probability density function. *J. R. Statist. Soc. Ser. A*. DOI: 10.1111/rssa.12053.

[*Received June 2014, accepted June 2014*]

International Statistical Review (2015), 83, 2, 202–207 doi:10.1111/insr.12074

Christophe Ley and Davy Paindaveine

Université libre de Bruxelles, ECARES and Département de Mathématique, Campus Plaine, CP210, Boulevard du Triomphe, B-1050, Brussels, Belgium
E-mail: chrisley@ulb.ac.be, dpaindav@ulb.ac.be

First of all, we would like to congratulate the author for an extremely interesting and insightful paper. In times where the need for flexible models has become obvious in view of the non-normal nature of most real data sets (be it in finance, biosciences, engineering, climatology, etc.) and where, consequently, more and more 'new' distributions are popping up in the literature, a survey paper reviewing the most important approaches is certainly highly welcomed. For researchers and—perhaps even more importantly—for practitioners,