A Constructive Representation of Univariate Skewed Distributions

José T.A.S. Ferreira and Mark F.J. Steel*

Department of Statistics
University of Warwick, UK

Abstract

We introduce a general perspective on the introduction of skewness into symmetric distributions. Making use of inverse probability integral transformations we provide a constructive representation of skewed distributions, where the skewing mechanism and the original symmetric distributions are specified separately. We study the effects of the skewing mechanism on *e.g.* modality, tail behaviour and the amount of skewness generated. In light of the constructive representation, we review a number of characteristics of three classes of skew distributions previously defined in the literature. The representation is also used to introduce two novel classes of skewed distributions. Finally, we incorporate the different classes of distributions into a Bayesian linear regression framework and analyse their differences and similarities.

Keywords: Arnold and Groeneveld skewness measure, Bayesian regression model, inverse probability integral transformation, modality, skewing mechanism, tail behaviour

1 Introduction

Recent years have seen a resurgent interest in the theory and application of distributions that can account for skewness. This article studies skewness in univariate data and its objective is threefold. First, we present a general constructive representation of univariate skewed distributions. We then use this representation to study classes of such distributions previously proposed in the literature and to construct novel classes. Finally, we use these classes in Bayesian regression modelling.

The most common approach to the creation of skewed distributions, and the one we are interested in here, is to introduce skewness into an originally symmetric distribution. This approach underlies the general classes of skewed distributions generated, for example, by hidden truncation models (see *e.g.* Azzalini, 1985 and Arnold and Beaver, 2002), inverse scale factors in the positive and the negative orthant (Fernández and Steel, 1998) and, more recently, order statistics (Jones, 2004).

The key advantage of skewing a symmetric distribution F is that in doing so it is possible to retain some of the properties of F, which are often well known. All methods mentioned in the previous paragraph keep a subset of these properties, with distinct models leading to distinct subsets.

Here we propose a unified perspective on skewed distributions. The idea is to separate the skewing mechanism from the symmetric distribution that serves as a starting point. This is appealing both methodologically and in the context of applications. By separating the two components, different

* *Address for correspondence*: Mark Steel, Department of Statistics, University of Warwick, Coventry, CV4 7AL, U.K. Tel.: +44-24-7652 3369; Fax: +44-24-7652 4532; Email M.F.Steel@stats.warwick.ac.uk.

classes of skewed distributions can be compared in a common framework. The decomposition also brings inferential advantages, particularly in the elicitation of prior distributions for these skewed distributions. In addition, new skewing mechanisms can be designed so as to generate distributions with pre-defined characteristics, tailored to suit specific requirements.

Separation of the components is achieved via inverse probability integral transformations. The probability density function (pdf) of the skewed distribution can be decomposed into one (fixed) factor which is the original symmetric pdf, and another defined by a probability distribution P in $(0, 1)$, which represents and models the asymmetry. One immediate benefit is that any skewed version of the same symmetric distribution can then be modelled directly by choosing a particular P.

Using this representation, we study the three classes of skewed distributions mentioned above. They can all be rather easily factorised according to our representation which facilitates an analysis of their differences. The representation also assists us in verifying some of the known properties of the classes. In addition, we introduce two new skewing mechanisms with rather distinct characteristics. The first was constructed in order to generate distributions with a pre-specified set of properties. For example, skewness is introduced around the mode and tail behaviour is left completely unaffected. The second type of new skewing mechanism is a very flexible one, given by Bernstein densities (Petrone and Wasserman, 2002) of arbitrary order.

We are particularly interested in the effect of different skewing mechanisms on tail behaviour. In order to illustrate this effect we analyse several skewed versions of three very common distributions: Normal, Logistic and Student-t_ν (henceforth t_ν), where ν denotes the degrees of freedom. For the skewed versions of these distributions we derive a number of results through the new representation.

Throughout, skewness is quantified according to the measure proposed in Arnold and Groeneveld (1995), defined as one minus twice the mass to the left of the mode. This measure, which takes values in $[-1, 1]$, is fairly intuitive for unimodal distributions, with negative (positive) values for left (right) skewed distributions and zero corresponding to symmetric distributions. It is particularly suitable for quantifying skewness of heavy tailed distributions because it does not require the existence of any moments, in constrast to most other measures.

Finally, we consider skewed regression models by using a linear regression structure, with errors distributed according to the different skewed distributions, and compare these regression models in the context of two applications. Formal model comparison is conducted through Bayes factors. Prior distributions for the skewness parameters are based on prior matching with a certain prior distribution on the skewness measure mentioned above.

Section 2 introduces the constructive representation of skewed distributions and a number of properties based on this representation. In Section 3, familiar classes of skewed distribution are decomposed and compared in terms of skewing mechanisms. Two new classes of skewed distributions are defined in Section 4. Section 5 defines the Bayesian regression setup. Analysis of the applications and comparison of the models is presented in Section 6. Section 7 groups some concluding remarks. All proofs are deferred to the Appendix, without explicit mention in the body of the text.

2 Representation of Univariate Skewness

In the sequel, S, F and P denote, respectively, a skewed distribution on the real line, a symmetric distribution on the real line and a distribution on $(0, 1)$, or their cumulative distribution functions (cdfs). Pdfs are denoted by the corresponding lower case letters. Further, let $x \in (0,1)$ and $y \in \Re$. Extension to the case where $y \in (a,b)$, $-\infty < a < b < \infty$ is trivial. Throughout the paper, we assume that all density functions exist and are continuous.

The unified representation of univariate skewed distributions that we introduce in this article is based on the inverse probability integral transformation. In particular, if the random variable X has distribution P on $(0,1)$, then we consider $Y = F^{-1}(X)$. In fact, since $\Pr(Y < y) = \Pr(X < F(y)) = P[F(y)]$ we can choose this to be any value by an appropriate choice of P. Thus, distributions defined in this way will cover the entire class of continuous distributions. As our interest is in characterising skewed versions of a symmetric distribution F, we will aim to restrict the classes we define in the sequel by *e.g.* only allowing one symmetric member, namely F itself (corresponding to uniform P). This means the classes of distributions can be indexed by F and can be thought of as grouping skewed versions of F. However, we will not impose any restrictions at this stage, and will first examine some general properties of the induced distribution of Y, denoted by S. In terms of densities, this leads to the following definition.

Definition 1. A distribution S is said to be a skewed version of the symmetric distribution F, generated by the skewing mechanism P, if its pdf is of the form

$$s(y|F,P) = f(y)p[F(y)]. \tag{1}$$

Given our interest in skewed versions of symmetric distributions, this representation is particularly useful in that the roles played by the symmetric distribution and by the skewing mechanism are obvious. The pdf $s(y|F,P)$ is a weighted version of $f(y)$, with weight function given by $p[F(y)]$. It is by specifying the skewing mechanism that the properties of S are fixed. In particular, P determines how characteristics of S will relate to those of F. This allows us to concentrate our attention on P.

Distribution P transforms the mass allocation of F into that of S. The plots in Figure 1 illustrate the effect of the skewing mechanism. The left-hand side plot shows the cdf of the original symmetric distribution. To this different skewing mechanisms, given by *e.g.* P_1 and P_2, can be applied (centre plots), leading to different weight functions (right-hand side plots). As P_1 is chosen to be Uniform, the weight function is constant, leading to $S = F$ (this is, of course, the usual application of the inverse probability integral transformation). The case of P_2 is more interesting. The asymmetry of p_2 implies an asymmetric weight function leading to a skewed S. Merely by analysing p_2, we can conclude that, irrespective of F, S will be right skewed, that the left tail of the distribution will be shrunk and that the right tail of the distribution will be proportional to the one of the original distribution.

A number of simple but powerful results can be obtained from the decomposition in Definition 1. The first one specifies the conditions on P that lead to symmetric S.

Theorem 1. Let, F, P and S be as above. Then,

(i) S is equal to F only in the case when P is the Uniform distribution on $(0,1)$.

(ii) Let P be fixed and vary F. Obtaining a symmetric S for any F is equivalent to symmetry of P around $1/2$.

The result in Theorem 1(i) is a rather intuitive one, with $S = F$ corresponding to the case when P does not modify mass allocation. Theorem 1(ii) characterises the situation where P is

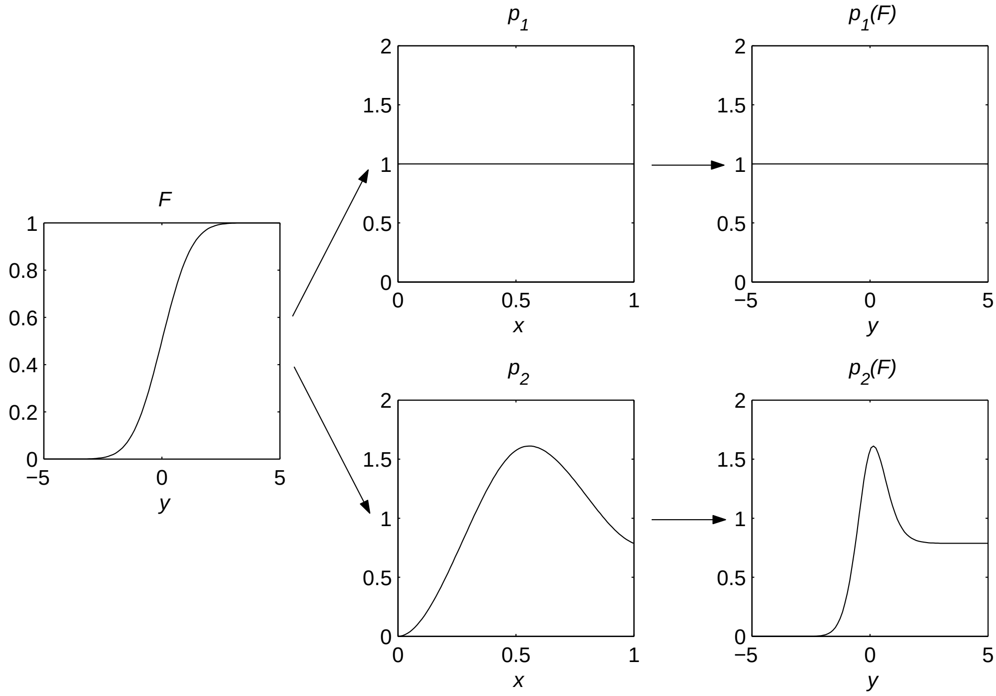

Figure 1: Effect of two distinct skewing mechanisms.

never a *true* skewing mechanism, in the sense that it can never introduce skewness into an originally symmetric distribution. Thus, interesting classes of skewed distributions will typically avoid non-Uniform symmetric possibilities for P. We should point out that for certain distributions F, S can be symmetric even if P is asymmetric. For instance, that is the case when S is a shifted version of F. These cases are no interest for our purposes.

The modality of S is of special relevance in modelling real phenomena. For instance, it is not uncommon to assume unimodality. Determining the modes of the distribution involves determining the stationary points of s, namely solving

$$s'(y|F,P) = 0 \Leftrightarrow -\frac{f'(y)}{f^2(y)} = \frac{p'[F(y)]}{p[F(y)]}, \tag{2}$$

which often is not possible analytically so that numerical methods have to be employed. However, for one important case we have a simple result.

Theorem 2. Let F be any symmetric and unimodal distribution, with mode at y_0. Further, let P be any unimodal distribution with mode at $1/2$. Then, the skewed distribution S is also unimodal and its mode is at y_0.

This property is useful in the definition of mode-preserving skewing mechanisms. When F and P are as in Theorem 2, calculation of the skewness of S through the Arnold and Groeneveld (1995) measure, denoted by AG is greatly simplified in that it never depends on F, being given by

$$AG = 1 - 2P(1/2). \tag{3}$$

We now turn our attention to the existence of moments of S, and in particular to how P affects the relation to the moment existence of F. To facilitate the discussion we introduce a definition on moment characteristics of a distribution.

Definition 2. Let G be the distribution of a random variable y in $\Re$. We define:

(i) Largest left moment of $G : M_l^G = \sup\{m \in \Re : \int_{-\infty}^0 y^m dG < \infty\}$.

(ii) Largest right moment of $G : M_r^G = \sup\{m \in \Re : \int_0^{\infty} y^m dG < \infty\}$.

(iii) Largest moment of $G : M^G = \max\{M_l^G, M_r^G\}$.

If distribution G is symmetric with continuous pdf, as is the case for F, then $M_l^G = M_r^G = M^G$. As examples, the Normal and Logistic distributions have $M^G = \infty$ while the heavier tailed t_ν has $M^G = \nu$.

Theorem 3. The condition that, when x tends to zero and one, the limits of $p(x)$ are finite and non-zero:

(i) is equivalent to $M_l^S = M_r^S = M^F$, if $0 < M^F < \infty$.

(ii) implies that $M_l^S = M_r^S = M^F$, if $M^F = \infty$.

Intuitively, when limits of p at the extremes of $(0, 1)$ are zero, moments are *added* to the distribution (with 0 corresponding to the left tail and 1 to the right tail). Similarly, when these limits are ∞, moments are *removed*.

Theorem 3 provides a strong result on the existence of moments for skewed versions of distribution with finite M^F. Theorem 3 is applicable both to the definition of moment preserving skewing mechanisms and to the evaluation of moment characteristics for already defined classes of skewed distributions. The result is weaker for transformations of F when $M^F = \infty$. This is due to the impossibility of adding to the moments of F and to the possibility of removing any finite number of moments without altering $M^S = \infty$.

Often, a complete characterisation of the moment existence of S will not be possible, but we can still ensure the existence of certain moments:

Theorem 4. If $p(x)$ is bounded, then $M_l^S, M_r^S \geq M^F$. If, in addition, the limit when x tends to zero (one) is positive then M_l^S (M_r^S) equals M^F.

The previous results deal with the case when p is bounded and, in the case of Theorem 3, when its limits as x tends to zero and one are finite. The following theorems deal with the case when P does not have either or both of these properties. For the sake of brevity, we only analyse the effect on M_l^S. Treatment for M_r^S is entirely analogous.

Theorem 5. If $\lim_{y \to -\infty} \frac{p[F(y)]}{|y|^b}$ is positive and finite for some $b \in \Re$, then $M_l^S = M^F - b$.

As will be seen in the sequel, the result in Theorem 5 is particularly interesting for determining the moments of some skew-t_ν alternatives. A weaker result, but one with a wider scope is presented in the following theorem.

Theorem 6. Let $b, K > 0$. Then if, for $F(y) < \epsilon < 1/2$,

(i) $p[F(y)] \leq K|y|^{-b}$ then $M_l^S \geq M^F + b$;

(ii) $p[F(y)] \geq K|y|^b$ then $M_l^S \leq M^F - b$.

Theorem 6(ii) provides a useful result especially when M^F is finite. When the latter is ∞, it is possible that $p[F(y)] > K|y|^b$, for any $b > 0$ and suitable K while S still has the same moment existence as F.

3 Familiar Classes

In this section we put three common methods of generating skewed distributions from symmetric ones into the framework introduced in Section 2. We do not conduct an exhaustive study of any of these classes of distributions as that is not the aim of this work. Readers interested in further details are referred to the references provided. In the sequel, we assume that F is a unimodal distribution.

3.1 Hidden Truncation

Skewed distributions generated by hidden truncation ideas are probably the most common and most intensively studied skewed distributions. Arnold and Beaver (2002) presents an overview, both for the univariate as for the multivariate cases. The skew-Normal distribution in Azzalini (1985) constitutes the first explicit formulation of such a distribution specifically for skewness modelling.

The most common versions of univariate skewed distributions generated by hidden truncation have densities that are of the form

$$s(y) = 2f(y)G(\alpha y) \tag{4}$$

where G denotes the cdf of a distribution on $\Re$ and $\alpha \in \Re$. Positive (negative) values of α generate right (left) skewed distributions while $\alpha = 0$ leads to $S = F$. When G equals F the distributions are usually called skew-F.

The case when F and G are not identical has also been studied (*i.a.* Nadarajah and Kotz, 2003). In addition, recent research has shown that it is possible to extend models such as the one in (4) even to the case when G is not a distribution (Arellano-Valle and Genton, 2003). Further, often the cdf G is applied not to αy but to other transformations of y, such as $\alpha_1 + \alpha_2 y$. All these cases can be accommodated in the constructive representation. Here we focus on the original, most common, case when G is a distribution and is applied to αy.

Viewed in terms of the representation in (1), the distribution with pdf given by (4) is a skewed version of distribution F, with skewing mechanism P with pdf

$$p(x|\alpha) = 2G[\alpha F^{-1}(x)],$$

where p is explicitly conditioned on α as different values for this parameter lead to distinct skewing mechanisms.

Any real α leads to a monotonic p, taking values in $(0, 2)$. Therefore, if $f'(y)/f^2(y)$ is decreasing then, from (2), S is necessary unimodal. This is the case of a number of distributions F, including the Normal, Logistic and t_ν. When $\alpha = 0$ we obtain $S = F$ with P the Uniform distribution on $(0, 1)$. We now consider the case when $\alpha > 0$, the case with negative α being similar. For $\alpha > 0$, p is an increasing function of x, with

$$\lim_{x \to 0^+} p(x|\alpha) = 0 \quad \text{and} \quad \lim_{x \to 1^-} p(x|\alpha) = 2.$$

As the limit of p when x tends to one is positive and finite, by Theorem 4, M_r^S will necessary be equal to M^F. As the limit of p when x tends to zero is zero, the only general comment that can be made about M_l^S is that it is not less than M^F. More precise results can only be determined if both F and G are specified. Let us now focus on skewed versions of the Normal, the Logistic and the t_ν distributions. Due to M^{Normal} and M^{Logistic} being equal to infinity, M_l^S for their skewed versions is unchanged (from Theorem 4 and the fact that one cannot increase ∞!). For the t_ν distribution, it is possible to show that

$$\lim_{y \to -\infty} \frac{p[F(y)]}{|y|^b}$$

is finite and non-zero only in the case when $b = -\nu$. Thus, from Theorem 5, we obtain $M_l^S = 2\nu$.

In general, it is not possible to calculate the skewness of distributions generated by hidden truncation analytically. This is the case for most measures of skewness, including AG.

3.2 Inverse Scale Factors

Another method for introducing skewness into a unimodal distribution F symmetric around the origin was introduced in Fernández and Steel (1998). Their basic idea was to introduce inverse scale factors in the positive and the negative half real lines. Let γ be a scalar in $(0, \infty)$. Then, S is defined by means of the pdf,

$$s(y|\gamma) = \frac{2}{\gamma + \frac{1}{\gamma}} f\left[y\gamma^{-\text{sign}(y)}\right] \tag{5}$$

where $\text{sign}(\cdot)$ is the usual sign function in $\Re$. If the skewness parameter γ is unity, we retrieve the original symmetric density. The mode of the density is unchanged, remaining at zero irrespective of the particular value of γ. Another relevant feature of this method is that $M_l^S = M_r^S = M^F$.

The decomposition of Definition 1 is obtained by defining p as

$$p(x|\gamma) = \frac{2}{\gamma + \frac{1}{\gamma}} \frac{f[\gamma^{\text{sign}(1/2 - x)} F^{-1}(x)]}{f[F^{-1}(x)]}.$$

Like for the class of skewed distributions described in Subsection 3.1, p is always monotonic and it is constant only if $\gamma = 1$. We know that this method always transfers moment characteristics from F to S. As an illustration of the use of the results given in Section 2, we analyse moment existence for the skewed t_ν. We select γ to be larger than unity, corresponding to right skewed distributions. The limit of $p(x|\gamma)$ when x tends to zero is proportional to $\gamma^{-(\nu+1)}$ and the limit when x tends to one is proportional to $\gamma^{\nu+1}$ for the t_ν version. Therefore, by Theorem 3, we immediately conclude that $M_l^S = M_r^S = M^{t_\nu} = \nu$.

Measuring the skewness of distributions generated by inverse scale factors is straightforward. The AG measure is a function of γ alone, given by (see Fernández and Steel, 1998)

$$AG(\gamma) = \frac{\gamma^2 - 1}{\gamma^2 + 1},$$

which is a strictly increasing function of γ and can take any value in $(-1, 1)$.

3.3 Order Statistics

The last of the familiar classes of skewed distributions that we analyse here is based on order statistics, and has been recently introduced in Jones (2004). As pointed out by the latter, a generalised version of distributions based on order statistics "...is also the result of applying the inverse probability integral transformation to the beta distribution". Indeed, the skewing mechanism of this class is given by the Beta distribution with pdf

$$p(x|\psi_1, \psi_2) = [B(\psi_1, \psi_2)]^{-1} x^{\psi_1 - 1}(1-x)^{\psi_2 - 1}, \tag{6}$$

where $B(\cdot, \cdot)$ is the Beta function and $\boldsymbol{\psi} = (\psi_1, \psi_2) \in \Re_+^2$ parameterises the distribution. Distributions generated by this setting are sometimes also said to belong to the generalised class (Amoroso, 1925).

In contrast to the previous cases, S can be bimodal for certain unimodal distributions F and values of $\boldsymbol{\psi}$.

Left tail behaviour of the distribution is controlled by ψ_1. If $\psi_1 > 1$, as $\lim_{x \to 0^+} p(x|\psi_1, \psi_2) = 0$, no left moments can be lost compared to the ones of F. Equivalently, if $\psi_1 < 1$, as $\lim_{x \to 0^+} p(x|\psi_1, \psi_2) = \infty$, no left moments can be gained. The effect of ψ_2 on the right tail is similar.

A number of distributions proposed in the literature are members of this class. The Beta-Normal distribution, generated when F is the Normal distribution, has been analysed in detail in Eugene (2001), with some findings given in Eugene et al. (2002). This distribution has properties that can make it less attractive for applications. As observed in Eugene (2001), the distribution can be bimodal but is necessarily unimodal if the parameters of the skewing mechanism are both larger than 0.241. Also, the amount of skewness is not a monotonic function of $\boldsymbol{\psi}$. Finally, the distribution is not flexible enough to be highly skewed. All these characteristics can be inferred from Figure 1. There we present a grey-scale plot of the AG skewness for Beta-Normal distributions with varying $\boldsymbol{\psi}$, presented on the log-scale, together with the relevant colour bar. Dark (bright) areas correspond to values of $\boldsymbol{\psi}$ leading to right (left) skewed distributions. The maximum absolute values of skewness obtained for $10^{-3} \leq \psi_1, \psi_2 \leq 10^3$ is roughly 0.58. The area of the plot with vertical hatching, corresponding to small values of both ψ_1 and ψ_2, denotes the parameter region leading to bimodality (and therefore not suitable to be classified with the AG measure). Further, we observe that the degree of skewness does not necessarily increase by increasing one element of $\boldsymbol{\psi}$ whilst decreasing the other.

When F is the Logistic distribution, applying the skewing mechanism in (6) generates the, so-called, log F distribution (Aroian, 1941). This distribution has a number of interesting results in analytic form, some of which we present here. The moment generating function is given by

$$M(t) = \frac{\Gamma(\psi_1 + t)\Gamma(\psi_2 - t)}{\Gamma(\psi_1)\Gamma(\psi_2)}. \tag{7}$$

From (7) we deduce that

$$\begin{aligned} E(y) &= \Psi(0, \psi_1) - \Psi(0, \psi_2) \\ V(y) &= \Psi(1, \psi_1) + \Psi(1, \psi_2), \end{aligned}$$

where $\Psi(n, z)$ is the polygamma function, with $n = 1, 2, \ldots$ and $\Psi(0, z)$ the digamma function. The distribution is always unimodal with mode at $y = \log\left(\frac{\psi_1}{\psi_2}\right)$. Finally, the AG measure is given by

$$AG = 1 - 2 I_{\frac{\psi_1}{\psi_1 + \psi_2}}(\psi_1, \psi_2), \tag{8}$$

with $I_x(a, b)$ the incomplete Beta function ratio. We can deduce from (8) that by varying $\boldsymbol{\psi}$ it is possible to have any AG skewness value in $(-1, 1)$. In addition, AG is increasing in ψ_1 and decreasing in ψ_2.

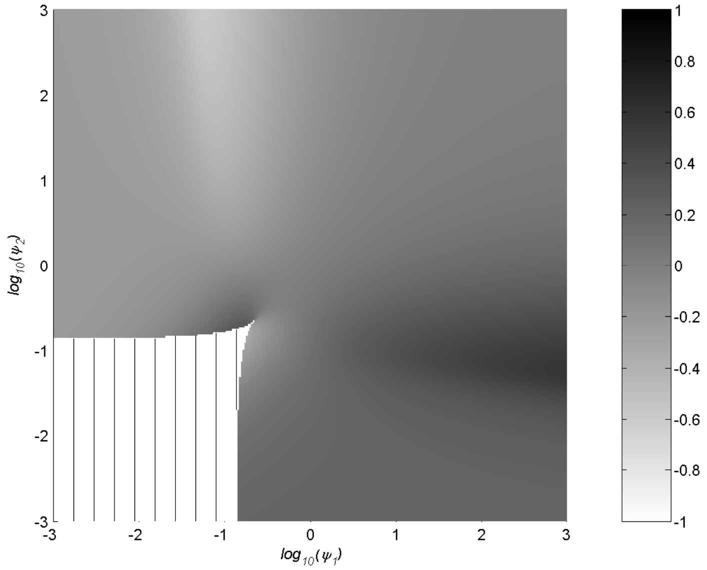

Figure 2: Grayscale plot of AG for Beta-Normal distributions with varying $\boldsymbol{\psi}$. Values in the lower left corner of the plot, with vertical hatching, correspond to bimodal Beta-Normal distributions.

The effects of skewing the t_ν distribution using the Beta distribution have been studied in Jones and Faddy (2003) and Jones (2004). Many results are provided in these references. Of special interest is the fact that tail behaviour is severely affected so that $M_l^S = \psi_1 M^{t_\nu} = \psi_1 \nu$ and $M_r^S = \psi_2 M^{t_\nu} = \psi_2 \nu$. This also follows from Theorem 5. This characteristic can be a disadvantage in the sense that modelling skewness and fat tails cannot be done independently. Unimodality of S is ensured only when ν is small. Some numerical experimentation revealed that for $\nu \leq 5.8$ S is unimodal irrespective of $\boldsymbol{\psi}$. If $\nu \geq 5.9$ then the distribution can be bimodal. In addition, when ν is large, interpretation of the effects of $\boldsymbol{\psi}$ on skewness is difficult, in the same sense as it is for the Beta-Normal distribution. Further, only for small ν can the full range of possible AG skewness be covered. As expected, the larger the value of ν, the more similar the skewed version of the t_ν distribution is to the Beta-Normal. Finally, the skew-t defined in Jones and Faddy (2003) is obtained by skewing a normalised version of the t_2 distribution, with the normalisation depending on $\boldsymbol{\psi}$. Nevertheless, the usual (un-normalised) version of the t_2 distribution can be employed without loosing any of the attractive analytical results.

The potential bimodality of S and the effect that the skewing mechanism has on M_l^S and M_r^S suggest that for this class of distributions it may be of interest to find a restricted parameterisation of (6). Ferreira and Steel (2004b) suggested the use of $\boldsymbol{\psi} = (\phi, 1/\phi)$, $\phi \in \Re_+$. With such a parameterisation, p is monotonic, and for the three choices of F used here, it always leads to unimodal S. In addition, this excludes the possibility of leading to symmetric distributions S different from F, which would otherwise be generated by choosing $\psi_1 = \psi_2 \neq 1$ (see Theorem 1(ii)).

3.4 Comparison of Skewing Mechanisms

In the previous subsections we have analysed three general methods for skewing symmetric distributions. We showed that they can all be interpreted in terms of the constructive representation introduced in Section 2. The methods always skew a symmetric distribution F, using skewing mechanism P. By fixing F and comparing the different Ps we can gain further insight in the properties of the various classes of asymmetric distributions.

Figure 2 presents p for each of the skewing mechanisms such that the obtained skewed distributions have three levels of AG skewness: 0.2 (left column), 0.5 (middle columns) and 0.8 (right column), for three different choices of F: Normal (top row), Logistic (middle row) and t_2 (bottom row). For the mechanism based on order statistics, the parameterisation suggested in Ferreira and Steel (2004b) was employed. Plots (b) and (c) do not contain the plot of p for the model based on order statistics because the relevant values of AG cannot be attained. The figure highlights a number of important issues. In all cases, the value of $p(x)$ when x approaches zero and unity are very different, potentially leading to substantially different tails of S. For the hidden truncation technique, the shape of the pdf implies that a considerable amount of skewness is introduced around the mode of the distribution. For the method derived from order statistics the shape of p reveals that skewness is introduced largely through the right tail of the distribution. Examining the plots by columns shows that for the different methods the skewing mechanism does not depend greatly on the particular distribution F at hand. This suggests that distributions P that do not depend on F may be a sensible choice. This idea is further explored in the next section.

4 New Classes

4.1 Construct

Here we use the constructive representation of Section 2 to define a novel class of skewed distributions with a set of pre-defined characteristics induced by an appropriate choice of skewing mechanism.

We would like to have the following properties:

(i) The skewing mechanism P does not depend on F.

(ii) The only P that retrieves a symmetric S is the Uniform distribution.

(iii) The unique mode of S equals the mode of F.

(iv) (a) skewness is introduced around the mode of the distribution with
(b) both tails of S behaving identically.

(v) Any AG skewness in $(-1, 1)$ can be achieved and the AG measure is independent of F.

(vi) The measure of skewness is an odd function of δ, the scalar parameter of P.

Defining a skewing mechanism that meets requirement (i) is appealing because after fixing such a P we can concentrate on selecting an adequate distribution F. As we want to skew symmetric distributions, it is natural to impose (ii), as already suggested in Section 2. By Theorem 1, this is imposed if the only symmetric P around 1/2 in the class of skewing mechanisms is the Uniform distribution. The distribution F will then uniquely define a class of distributions S (for varying P) as being the only symmetric member of that class. Property (iii) serves a dual purpose. Firstly, it removes the necessity of finding the mode of the distribution. Secondly, it will make the calculation of AG simpler (see (3)). With (iv) we want to impose that the asymmetry of S is located around the mode and not in the tails. In particular, we will impose that as y increases in absolute value, $s(y)$ is approximatively $Kf(y)$, with K a strictly positive and finite constant. Characteristic (v),

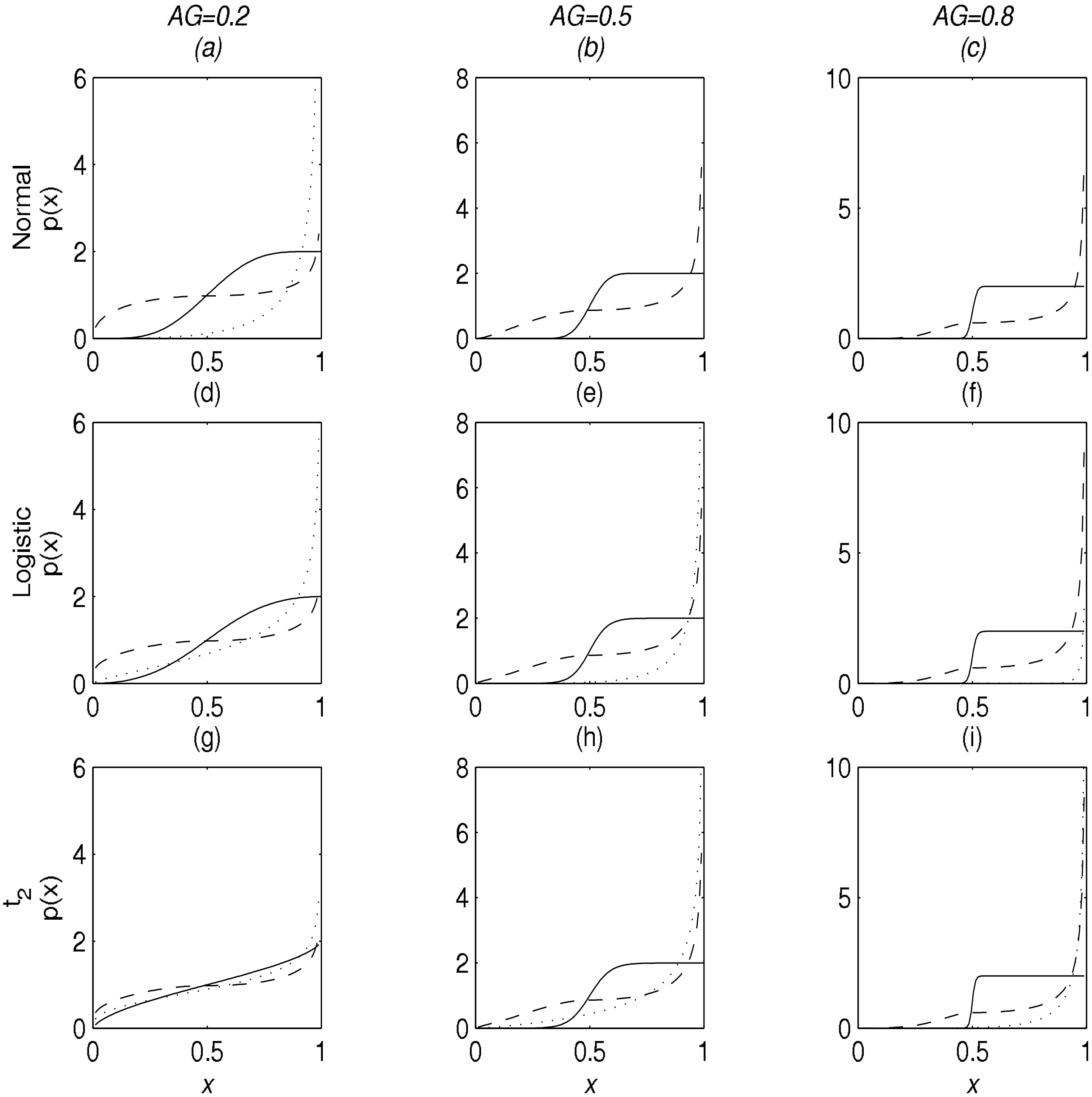

Figure 3: Densities p for each of the skewing mechanisms leading to three levels of AG skewness: 0.2 (left column), 0.5 (middle columns) and 0.8 (right column), for Normal (top row), Logistic (middle row) and t_2 (bottom row). In each plot the solid line stands for the hidden truncation model, the dashed line for the inverse scale factors model and the dotted line for the model based on order statistics.

especially when combined with (i), facilitates the interpretation of the model parameters and, thus, prior elicitation. Finally, (vi) imposes that by changing the sign of δ we obtain the same amount of skewness, but in the opposite direction.

We do not argue that such properties are ideal, or that such skewing mechanisms are appropriate for all applications. The objective here is twofold. Up to our knowledge, a skewing mechanism with the characteristics above has not been previously defined. In addition, we want to illustrate how the constructive representation can be used to impose certain characteristics on S.

By imposing (i), the construction of the skewing mechanism is somewhat simplified as we do not need to take F into account. We construct a skewing mechanism that meets the above requirements in two steps. First, we define a class of non-negative continuous functions $g(x|\delta)$, $x \in [0,1]$, $\delta \in \Re$, that has mode at $x = 1/2$, is decreasing in $|x - 1/2|$, is zero at both zero and one, and integrates to one. We continue by defining a continuous symmetric function $l(\delta)$ that is strictly increasing in the absolute value of δ, is zero for $\delta = 0$ and has $\lim_{\delta\to\pm\infty} l(\delta) = 1$. The overall skewing mechanism is then defined by

$$p(x|\delta) = 1 + l(\delta)[g(x|\delta) - 1]. \tag{9}$$

Even before defining the specific forms of g and l we can derive some properties of the mechanism generated by (9). If $\delta = 0$, p is the Uniform density. By imposing that g is symmetric only if $\delta = 0$, we guarantee (ii). The density p is bounded below by $1 - l(\delta)$, positive for finite δ, and as g is continuous with $g(0|\delta) = g(1|\delta) = 0$, p also has a positive upper bound, thus implying $M_l^S = M_r^S = M^F$ (Theorem 3). Also, as $p(0|\delta) = p(1|\delta) = 1 - l(\delta) > 0$, the tails of S will be proportional to the tails of F, with the same proportionality constant, thus meeting requirement (iv) (b). The mode of p is at 1/2. Therefore, by Theorem 2, the mode of s is the same as the mode of f, thus satisfying (iii). As the mode is know, the AG measure is given by (3), which here becomes

$$AG(\delta) = l(\delta)\left[1 - 2\int_0^{1/2} g(x|\delta)dx\right]. \tag{10}$$

Immediately, as $l(0) = 0$, $AG(0) = 0$. Finally, if for any $x \in [0,1]$, $g(x|\delta)$=$g(1-x|-\delta)$ then

$$AG(-\delta) = -AG(\delta), \tag{11}$$

and the class satisfies (vi).

Characteristics (iv) (a) and (v) are addressed in the definition of g and l.

4.1.1 Defining g

As is evident from (9) and (10), it is g that determines the asymmetry of p. Therefore, only if g can assign any mass to either side of 1/2 can $AG(\delta)$ take any value in $(-1,1)$. Here, we will define g such that for negative (positive) values of δ the function g, and consequently p, is left (right) skewed.

The definition of g is graphically depicted in Figure 4. We begin by selecting a continuous function $h(z)$, $z \in [0,1/2]$, such that $h(0) = 0$ and $h(x) = 2 - h(1/2 - x)$. Such a function is increasing in z and divides the rectangle $A = [0,1/2] \times [0,2]$ (with area one) into two parts with the same area. We then define

$$h^*(z|\delta) = h\left[\frac{e^{\delta z} - 1}{2\left(e^{\frac{\delta}{2}} - 1\right)}\right], \tag{12}$$

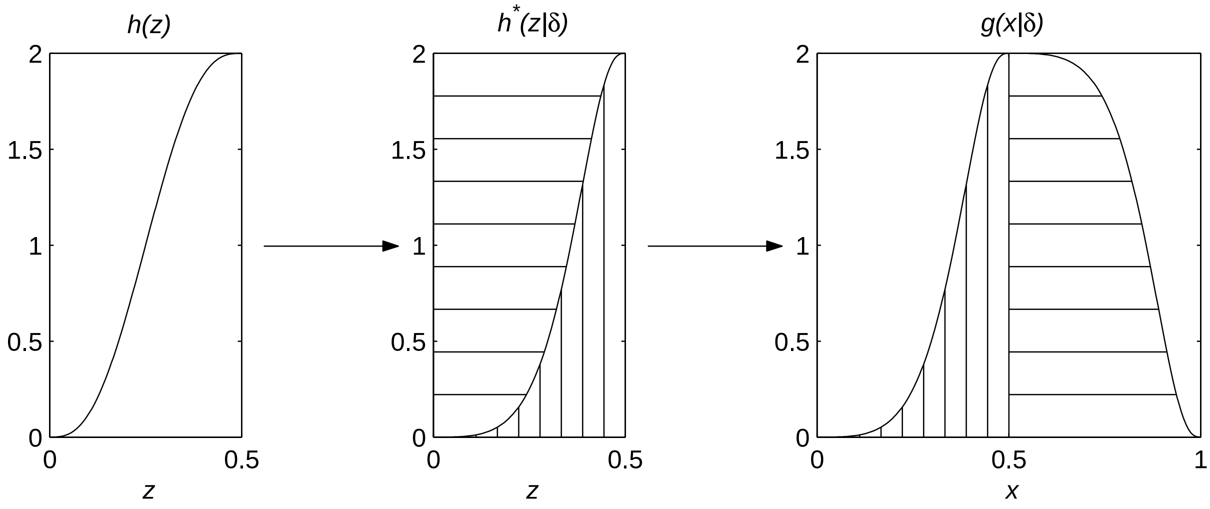

Figure 4: Illustration of the definition of $g(x|\delta)$.

for $\delta \neq 0$. In addition, we define $h^*(z|0) = h(z)$, since

$$\lim_{\delta \to 0} \frac{e^{\delta z} - 1}{2\left(e^{\frac{\delta}{2}} - 1\right)} = z. \tag{13}$$

The exponential transformation in the argument of h above modifies the manner in which A is divided. If $\delta \neq 0$, as seen from the middle plot in Figure 4, A is no longer divided into regions of equal size. If $\delta > 0$ ($\delta < 0$), the region represented with the horizontal lines will have larger (smaller) area than the region with vertical lines. By varying δ, the region represented with the horizontal lines can have any area in $(0, 1)$. We then define g as

$$g(x|\delta) = h^*(x|\delta) I_{[0,1/2]}(x) + \left[2 - h^*\left(x - \frac{1}{2} \mid \delta\right)\right] I_{(1/2,1]}(x), \tag{14}$$

which imposes (11), as $g(x|\delta) = g(1 - x| - \delta)$.

The right-hand side plot in Figure 4 gives the geometric interpretation of g. In $[0, 1/2]$, $g = h^*$, whilst in $(1/2, 1]$ g is $2 - h^*$ shifted $1/2$ to the right. Clearly, g integrates to one over its domain. The unique mode of the distribution is always at $1/2$. Also, the area to either side of the mode can take any value in $(0, 1)$. From the same plot we can observe that property (iv) is met. Skewness is concentrated clearly in the centre part of $[0, 1]$, with g, and therefore p, behaving identically in the extremes of the interval.

With such a choice of g, p as defined in (9) is a continuous function of x for any value of δ. Further, by controlling the smoothness of g, the smoothness of p can be controlled. If we impose that h has d continuous derivatives, *i.e.* h is $\mathcal{C}(d)$ and that the i^{th} derivatives $h^{(i)}(0) = h^{(i)}(1/2) = 0,\ i = 1, \ldots, d$, then p is $\mathcal{C}(d)$ in $(0, 1)$. This leads to s being $\mathcal{C}(\min\{d^*, d\})$, where d^* denoted the smoothness of f.

We now turn our attention to specific forms of h. The simplest form for h is given by polynomials of degree $2d + 1$ satisfying $h(0) = 0 = 2 - h(1/2)$, and $h^{(i)}(0) = h^{(i)}(1/2) = 0,\ i = 1, \ldots, d,\ d \in \mathbb{N}$.

Some particular choices of h are given below:

$$\begin{aligned} d=0 &\quad : \quad h(z)=4z \\ d=1 &\quad : \quad h(z)=-32z^3+24z^2 \\ d=2 &\quad : \quad h(z)=384z^5-480z^4+160z^3 \\ d=3 &\quad : \quad h(z)=-5120z^7+8960z^6-5376z^5+1120z^4. \end{aligned}$$

Choosing h of polynomial form has the advantage that it leads to the availability of $P(x|\delta)$ in analytical form. This is of special interest as it implies that the cdf S is easily calculated from F.

Polynomial choices of h cannot lead to $p(x|\delta)$ being in $\mathcal{C}(\infty)$. In order to have complete smoothness of $p(x|\delta)$, $h(x)$ must be a $\mathcal{C}(\infty)$ function in $[0,1/2]$, with $h^{(i)}(0)=h^{(i)}(1/2)=0,\ i=1,2,\ldots$. However, such functions can be less tractable than polynomials. Here we only present one particular example, defined by $h(0)=0=2-h(1/2)$ and by

$$h(z)=1-e^{e^2}\left\{\exp\left[-\left(e^{-(2z)^{-2}}-e^{-1}\right)^{-2}\right]-\exp\left[-\left(e^{-(1-2z)^{-2}}-e^{-1}\right)^{-2}\right]\right\}, \quad \text{if } 0<z<\frac{1}{2}. \tag{15}$$

Different choices of h are plotted in Figure 5. The functions can be compared according to the manner the function increases with z. For the polynomial of degree one, as $h^{(1)}=h'$ is constant, the transition is uniform across the entire domain. For the other functions in the figure the transition happens mainly around 1/4 and is faster as d increases. Function h as defined in (15) leads to the most abrupt transition.

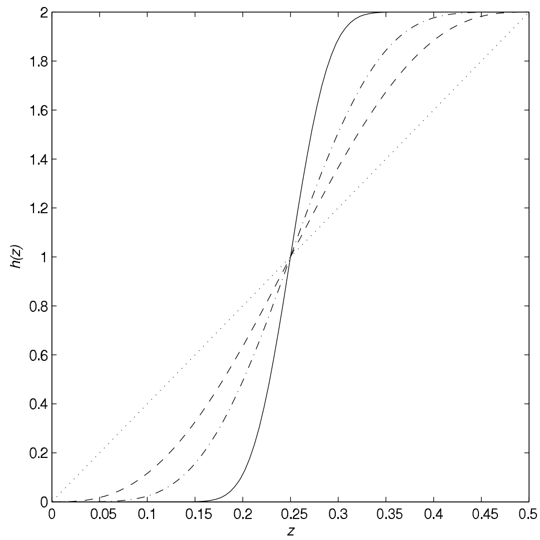

Figure 5: Plot of different functions h. The solid line is the function in (15). The remaining are the polynomial choices with d equal to zero (dotted), two (dashed) and five (dot-dashed).

4.1.2 Defining l

The function l is chosen so that characteristic (v) is met. From (10), as $0 \le l(\delta) \le 1$, the most extreme value of skewness possible for δ, in absolute terms, is given by

$$\left|1-2\int_0^{1/2} g(x|\delta)dx\right|, \tag{16}$$

which can be calculated analytically for polynomial h. Now suppose that we define some odd function $AG^*(\delta)$ that always satisfies

$$|AG^*(\delta)| \leq \left| 1 - 2\int_0^{1/2} g(x|\delta)dx \right|. \tag{17}$$

If we then define $l(\delta)$ as

$$l(\delta) = \frac{AG^*(\delta)}{1 - 2\int_0^{1/2} g(x|\delta)dx},$$

for $\delta \neq 0$, and $l(0) = \lim_{\delta \to 0} l(\delta) = 0$, we have that for any $\delta \in \Re$ the AG skewness measure of S will be $AG^*(\delta)$. In addition, if $AG^*(\delta)$ is a continuous function with

$$\lim_{\delta \to -\infty} AG^*(\delta) = -1 \quad \text{and} \quad \lim_{\delta \to \infty} AG^*(\delta) = 1, \tag{18}$$

then the complete range of skewness values is covered. By choosing AG^* to be an odd function, property (vi) is met.

The definition of the skewing mechanism is completed by specifying $AG^*(\delta)$ that satisfies (17)-(18) and is continuous. Straightforward analysis shows that for fixed δ, minimum skewness (in absolute value) is obtained when $h(z) = 4z$. As such, if $AG^*(\delta)$ satisfies (17) for $h(z) = 4z$ then this also holds for any other choice of g. For this last h we have that

$$|AG^*(\delta)| \leq \left| 1 - 4\int_0^{1/2} \frac{e^{\delta x} - 1}{e^{\delta/2} - 1} dx \right| = \left| 1 - \frac{4}{\delta} + \frac{2}{e^{\delta/2} - 1} \right|. \tag{19}$$

Choosing $AG^*(\delta)$ to be a bijective, odd function of δ with the properties in (18) and (19) imposes characteristics (v) and (vi) and provides an immediate interpretation of δ. After some manipulation, we found that

$$AG^*(\delta) = \left[\frac{2}{\pi} \arctan\left(\frac{2\delta}{5} \right) \right]^3 \tag{20}$$

is such a choice. Equation (20) is based on a powered cdf of a Cauchy distribution.

4.1.3 Resulting Skewing Mechanism

In the sequel, we always use a polynomial h with $d = 2$. Figure 6 exhibits a surface plot of $p(x|\delta)$ for varying values of x and δ. We can observe that for small $|\delta|$, $p(x|\delta)$ is very close to the Uniform density. For large values of $|\delta|$, the mass is shifted to the right (positive values) or to the left. Important to note is that, irrespective of the value of δ, the limits as x tend to zero or one are equal positive values, implying identical tail behaviour.

Figure 7 presents plots of the densities of skewed versions of the Normal (a), the Logistic (b) and the t_2 (c) distributions, with AG equal to 0.2, 0.5 and 0.8. As intended, the distributions are changed around the origin and not in the tails. Of course, the skewing mechanism for the different distributions is the same, by construction.

The skewing mechanism above was constructed so as to meet a set of pre-defined characteristics. In practice, the particular set of characteristics we wish to impose may well depend on the application at hand. For example, one characteristic that could be unsuitable for certain applications is the restriction that both tails behave identically (see our first example in Section 6).

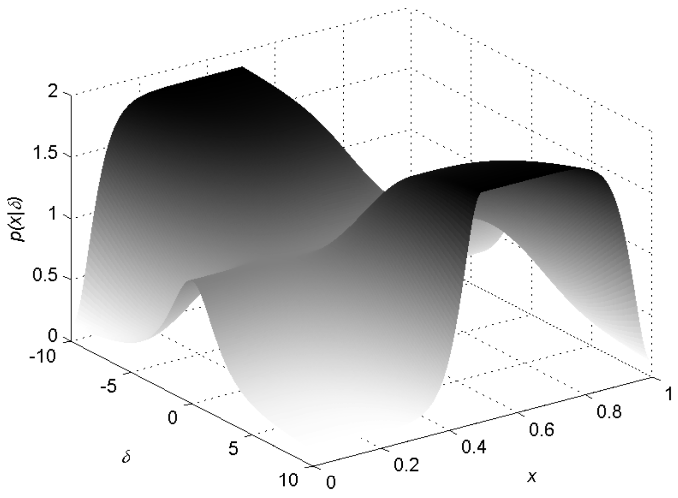

Figure 6: Surface plot of $p(x|\delta)$, as a function of x and δ when $g(x)$ is the 5-th degree polynomial.

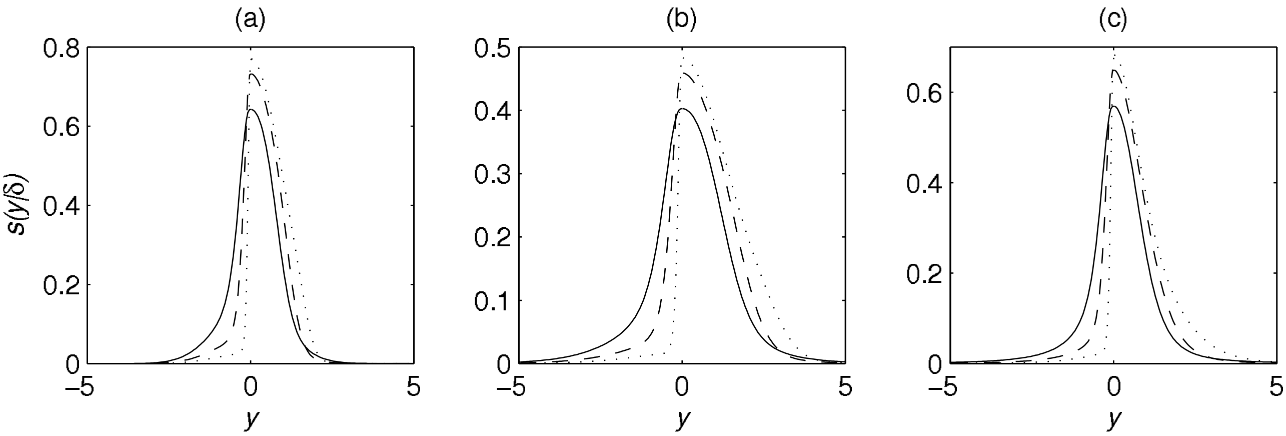

Figure 7: Densities s corresponding to skewed versions of the Normal (a), the Logistic (b) and the t_2 (c) distributions. In each case we plot the densities for distributions with AG equal to 0.2 (solid line), 0.5 (dashed line) and 0.8 (dotted line)

4.2 Bernstein Densities

Given the constructive representation of skewed distributions introduced in Section 2, a completely nonparametric treatment of the skewing mechanism may look appealing. As P can be any distribution in $(0, 1)$, the possibility to model it in an unrestricted fashion seems tempting. However, as already indicated in Section 2, this would effectively generate the entire class of continuous distributions and, thus, we would lose control over the properties of the resulting skewed distributions.

Here we try to reach a compromise between a totally flexible skewing mechanism and one for which some interesting results are still available. We make use of Bernstein densities (see *e.g.* Petrone and Wasserman, 2002) to model p. Given a positive integer m and a vector $\boldsymbol{\omega} = (\omega_1, \dots, \omega_m)$, a Bernstein density is given by

$$p(x|m, \boldsymbol{\omega}) = \sum_{j=1}^{m} \omega_j Be(x|j, m - j + 1),$$

where $Be(\cdot|a, b)$ denotes the pdf of the Beta distribution with parameters a and b, $\omega_j \geq 0,\ j = 1, \dots, m$ and $\sum_{j=1}^{m} \omega_j = 1$. Bernstein densities have some characteristics that make their use as flexible skewing mechanisms quite attractive. As the parameters of the Beta distributions in the mixture are never smaller than unity, $p(x|m,\boldsymbol{\omega})$ is bounded and therefore, by Theorem 4, $M_l^S, M_r^S \geq M^F$. For any choice of m, if $\omega_i = 1/m,\ i = 1,\ldots,m$, then P is Uniform which implies that we retain the original symmetric distribution F. Further, as long as there is a $j^* \in \{1,...,m^*\}$, with $m^* = m/2$ if m is even and $m^* = (m-1)/2$ otherwise, such that $\omega_{j^*} \neq \omega_{m-j^*+1}$, $p(x|\boldsymbol{\omega})$ is asymmetric.

Member of the class of distributions generated by this skewing mechanism are often multimodal. Multimodality becomes more common as m increases.

Figure 8 presents the effect of using Bernstein densities as skewing mechanisms. In (a) we present p for four different values of $\boldsymbol{\omega}$ and $m = 9$, illustrating the flexibility of the densities. Multimodality in p is the rule rather than the exception. Figure 8(b) presents plots for the corresponding skewed Normal densities. In one case, multimodality is transferred from p to s.

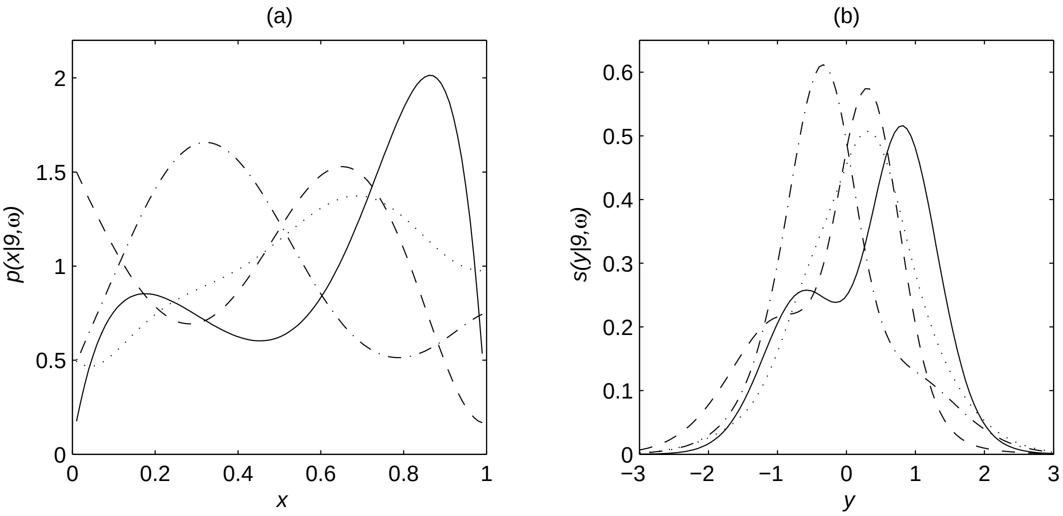

Figure 8: Plot (a) shows four different skewing mechanisms p based on Bernstein densities. In (b) the equivalent densities for the skewed versions of the Normal distribution are presented.

5 Regression Modelling

In the sequel, we assume that the observables $y_i \in \Re,\ i = 1,\ldots,n$ are generated from

$$y_i = \boldsymbol{x}_i'\boldsymbol{\beta} + \sigma\epsilon_i, \tag{21}$$

where $\boldsymbol{\beta} = (\beta_1,\ldots,\beta_k)' \in \Re^k$ is a vector of regression coefficients, $\boldsymbol{x}_i$ is a k-dimensional vector of explanatory variables, and $\sigma \in \Re_+$ is a scale parameter. We consider the cases where ϵ_i is i.i.d. with a distribution from one of the classes in Sections 3 and 4.

We adopt an independent prior structure for $\boldsymbol{\beta}, \sigma$ and the parameters of the distribution of ϵ_i. The priors on $\boldsymbol{\beta}$ and σ are chosen to be vague but proper. For $\boldsymbol{\beta}$, we choose a multivariate Normal distribution with independent components, each with mean zero and variance v. We define the prior on σ via an Inverted-Gamma prior on σ^2 with parameters a and b. Except for distributions based on Bernstein densities, the priors on the parameters of the distributions of ϵ_i are defined using the prior matching ideas suggested in Ferreira and Steel (2004c). There, we fix a prior distribution on the AG measure and derive equivalent priors on the skewness parameters of the skewed distributions. For the classes generated as in Sections 3.2 and 4.1 the equivalence is immediate and does not vary with the distribution that is being skewed. For the methods in Sections 3.1 and 3.3, only an approximate equivalence is possible. For the skewing mechanism based on Bernstein densities, we choose a Dirichlet prior on $\boldsymbol{\omega}$ with all parameters equal to c. Such a prior centres the skewing mechanism around the Uniform distribution, with c determining the prior variability of the mechanism. In the examples presented in the next section, we fixed the order m.

6 Examples

In the following two applications, we set $v = 100$ and $a = b = 0.1$. To these values correspond rather flat priors on $\boldsymbol{\beta}$ and $\ln(\sigma)$. On the AG measure of skewness we use a Beta prior distribution with both parameters equal to five, rescaled to the interval $(-1, 1)$. This prior puts little mass on extremely skewed distributions (more that 90% of the mass to one of the sides of the mode) and is relatively flat for other values. We also fixed c equal to one and selected $m = 9$. After some simulations, we found that this value for c leads to a prior that is not too restrictive on the skewing mechanism. If required, inference on m could easily be accomplished by using inference methods that allow the parameter space to vary (see *e.g.* Green, 1995 and Stephens, 2000).

We compare skew versions of three distributions: the Normal, the Logistic and the t_2 distribution. We also provide results for the original, symmetric, distributions. The reason for using the t_2 distribution is twofold. Firstly, this distribution underlies the skew-t as defined in Jones and Faddy (2003). Secondly, as opposed to the two other alternatives, it is a heavy tailed distribution. Alternatively, we could use a more flexible t_ν distribution and conduct inference on the degrees of freedom parameter ν with an appropriate prior, as *e.g.* in Fernández and Steel (1998).

For the class of distributions generated by order statistics we use the parameterisation suggested in Ferreira and Steel (2004b). However, for this class we do not present results for the skewed Normal distribution. This is due to the fact that there is no one-to-one correspondence between ϕ and $AG(\phi)$.

Inference is conducted using fairly standard Markov chain Monte Carlo methods. For the sake of brevity, we omit details of the samplers. These can be obtained from the authors.

Formal model comparison is done using Bayes factors, with marginal likelihoods estimated using the p_4 measure in Newton and Raftery (1994), setting their δ to 0.1.

6.1 Strength of Glass Fibres

The data for this example appear in Smith and Naylor (1987) and relate to the breaking strength of $n = 63$ glass fibres. In the context of univariate skewed distributions, this problem was recently analysed in Jones and Faddy (2003). As in the latter paper, we use a location-scale model for these data.

Table 1 presents the logarithm of the Bayes factor estimates for the different models with respect to the standard symmetric Normal model. Firstly, the evidence in favour of skewness is clear: the symmetric versions of the models perform worst throughout. In all cases, the alternatives generated using Bernstein densities seem to fit the data best. The Construct mechanism was outperformed by the other skewed distributions for the Logistic- and t_2-based distributions. Apart from the Bernstein density skewing mechanism, heavier tailed distributions are favoured. The greater flexibility of the Bernstein densities allows to adequately fit the tails to the data, even if the tails of F are too thin.

Table 1: Log of Bayes factors for the different models with respect to Normal model for the Strength of Glass Fibres data

	Symmetric	Hidden Trunc.	Inv. Scale	Order Stats.	Construct	Bernstein
Normal	0	3.15	2.66	-	2.99	6.93
Logistic	1.64	4.80	4.41	4.38	3.10	6.59
t_2	2.65	5.48	4.90	4.45	3.26	5.83

In Figure 9(a) we present estimates of the posterior predictive densities for the different skewed t_2 models, overlaid on a normalised histogram of the data. The estimates are quite similar and show that left skewed distributions are more appropriate. A similar conclusion was reached by Jones and Faddy (2003). Further insight is brought by Figure 9(b) where the logarithm of the estimated posterior predictive densities is plotted. Clearly, the tail estimates for the Construct model are different from those for the other models, potentially explaining the poorer performance of this method. For these data, it appears that the right-hand side tail is too "heavy", whilst the left-hand side one is too "light". From the histogram in Figure 9(a) the tails indeed appear to be different in the data. Thus, imposing that both tails are identical seems to be in conflict with the data, as evidenced by the Bayes factors in Table 1.

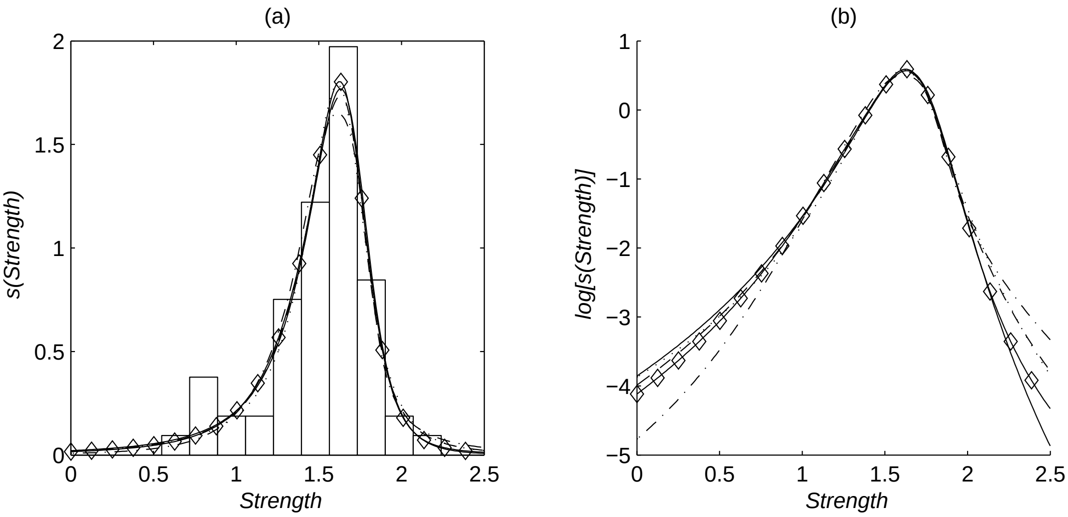

Figure 9: Posterior predictive densities (a) and log-densities (b) for the Hidden Truncation (solid), Inverse Scale Factors (dashed), Order Statistics (dotted), Construct (dot-dashed) and Bernstein densities (solid with diamonds) skewed t_2 distributions. A normalised histogram of the data is presented in plot (a).

6.2 Market Value Growth

Our second application studies the impact of Research and Development effort (R&D) and investment on the growth in market value between 1980 and 1990 of U.S. publicly traded manufacturing companies. The data derive from a panel of firms created by Bronwyn H. Hall (see Hall, 1993). These data have been analysed in a multivariate context in Ferreira and Steel (2004a), where more details can be found.

The dataset that we analyse consists of 300 observations. The target variable is the difference in the logarithm of the market value of firms, measured in millions of dollars, between 1990 and 1980. We include an intercept and as regressors we use R&D and investment, both measured as the ratio between quantity spent and total assets, and both standardised to have mean zero and unit variance.

The log Bayes factors with respect to the Normal model are presented in Table 2. Clearly, there is evidence in favour of skewness. As for the application in Section 6.1, the skewed models based on Bernstein densities again find most support in the data. The Construct-based alternatives perform quite well, being second to the Bernstein density ones for the skewed versions of the Normal and the t_2 distributions. The adequacy of the skewed versions of the Normal distribution varies extensively between the different skewing mechanisms. The logarithm of the Bayes factor between the best and the worst skewed Normal versions is 13.67. Also, this provides another illustration of the fact that the Bernstein density skewing mechanism can fit the data on the basis of almost any F. The Logistic distribution is favoured for all skewing mechanisms except the ones defined in Section 4, which favour the t_2 distribution. For all methods, Normal distributions behave worst.

Table 2: Log of Bayes factors for the different models with respect to Normal model for the Market Value Growth data

	Symmetric	Hidden Trunc.	Inv. Scale	Order Stats.	Construct	Bernstein
Normal	0	14.70	11.32	-	19.39	24.99
Logistic	20.78	25.37	26.24	26.40	25.18	26.86
t_2	19.26	24.52	23.41	21.71	26.69	28.15

We now focus on a graphical comparison of the skewing mechanisms. Figure 10 shows the mean posterior densities p for the different models and the three distributions. Except for the skewing mechanism based on Bernstein densities, p always has more mass on the left half-interval of $(0, 1)$, irrespective of the distribution that is being skewed. For the Bernstein densities models p always has more mass to the right of $1/2$. This seems counterintuitive but a closer look reveals an interesting, though perhaps undesirable, characteristic. Bernstein densities are so flexible they can adapt to the case where the main mass of S is moved. The flexible Bernstein skewing mechanism allows for a change in the location of the distribution (through the intercept) and adapts accordingly. It turns out to be beneficial to centre the distribution in a location that is different from the one favoured by the other models. From Figure 10, the Bernstein density skewing mechanism also suggests that the tail behaviour in both tails is quite similar, which partly explains the relative success of the Construct models for these data.

Finally, we can compare the different models in terms of the amount of skewness they induce. Figure 11 presents boxplots of the AG measure derived from the MCMC chain for the skewed versions of the Normal (a), Logistic (b) and t_2 (c) distributions. In all circumstances, negative values of AG are the norm, indicating strongly that the distribution of the regression residuals is left skewed. In all cases, the Construct method provided a smaller median value for AG and larger absolute values for extreme skewness, followed by the Inverse Scale Factors method.

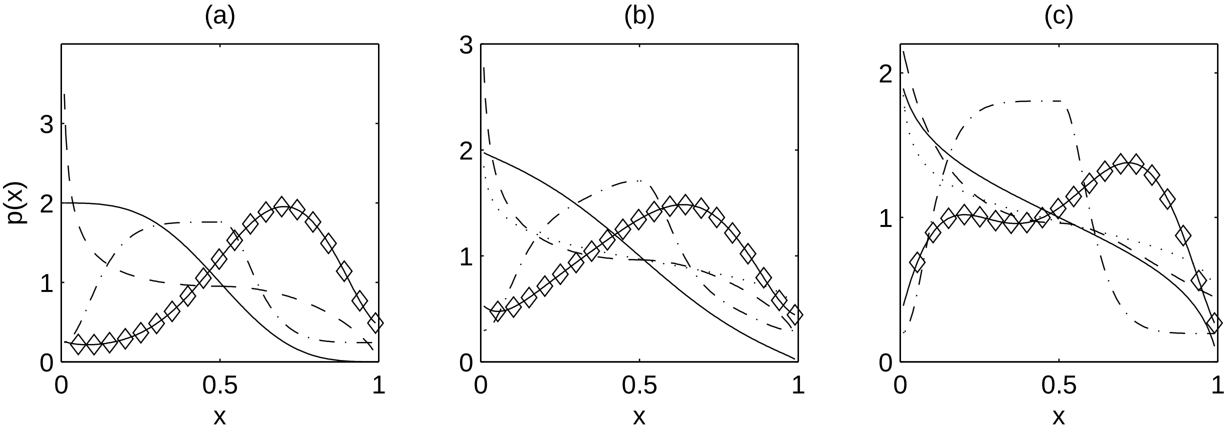

Figure 10: Plots of the mean posterior pdfs p for the skewed versions of Normal (a), Logistic (b) and t_2 (c) for Hidden Truncation (solid), Inverse Scale Factors (dashed), Order Statistics (dotted), Construct (dot-dashed) and Bernstein Densities (solid with diamonds).

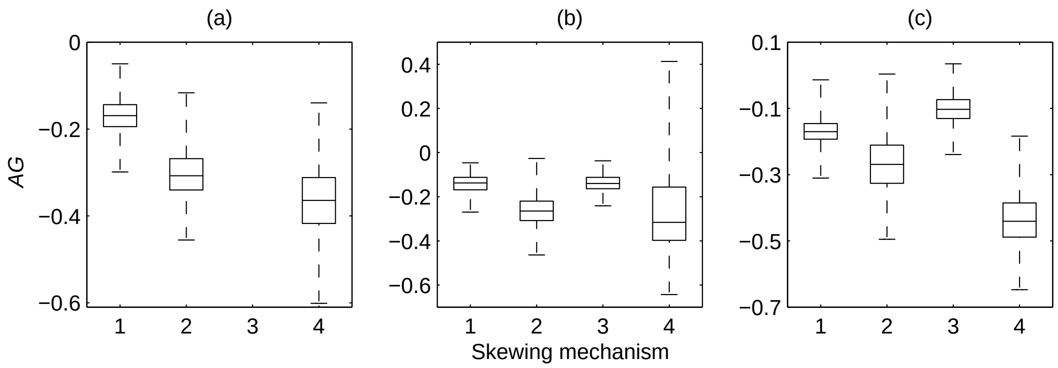

Figure 11: Boxplots of the AG measure for the skewed versions of Normal (a), Logistic (b) and t_2 (c) for Hidden Truncation: 1, Inverse Scale Factors: 2, Order Statistics: 3 and Construct: 4.

7 Conclusion

In this article we provide a unified view of the introduction of skewness into symmetric distributions. We make use of inverse integral probability transformations and separate out the skewing mechanism from the symmetric distribution. This decomposition immediately allows us to link properties of the skewing mechanism with key characteristics of the resulting skewed distributions, such as modality and moment existence. Using our decomposition to represent well-known classes of skewed distributions brings new insight in these previously defined classes and their comparison. In addition, the representation proposed here enables the definition of new classes with certain pre-specified characteristics.

An important contribution of the constructive representation is that it provides a common framework for comparing apparently unrelated skewed versions of a common symmetric distribution. As a consequence, we hope that the link between the behaviour of a real phenomenon and that induced by a particular class of skewed distributions can be more easily established. In other words, the insights derived from the constructive representation should make it easier to choose particular classes of skewed models for any given application, and help us to understand why certain classes work better than others. As we illustrated with the new classes developed in Section 4, it also provides a powerful framework for the development of new classes of skewed distributions, possibly with certain pre-set characteristics. Once we have imposed the set of properties that we want the skewed distributions to have, through the choice of skewing mechanism, we can then combine this with any symmetric distribution.

We believe that the definition of useful and realistic classes of skewing mechanisms is still partly an open question, and that it should be the focus of some debate. Of course, it depends on the particular data at hand. The two applications in Section 6 illustrate the importance of the question of how flexible a skewing mechanism should be. Even though we want to describe the underlying process in an appropriate manner, often we will also want to control a number of characteristics of the distribution classes we use for modelling. This will usually aid in the interpretation of the results and also in the out-of-sample predictive performance of the model. Very flexible skewing mechanisms like the Bernstein densities are not necessarily the ones we want to use in practice, but they can give valuable hints as to which properties a more restrictive class of models should have in order to fit a particular data set well. The trade-off between flexibility and control is as relevant here as in many other areas of statistics.

Matlab code written for conducting the inference described here is available from the authors upon request.

Acknowledgments: The work of the first author was supported by grant SFRH BD 1399 2000 from Fundação para a Ciência e Tecnologia, Ministério para a Ciência e Tecnologia, Portugal.

A Proofs and auxiliary results

Proof of Theorem 1. The proof of part (i) is immediate. In order to prove part (ii) we first prove that if P is symmetric around 1/2 then S is necessarily symmetric. We then show that for any asymmetric P it is possible to find F so that S is skewed.

For the first part, let F have mean and mode at y_0. If P is symmetric around 1/2, then

$$s(y_0 - y) = f(y_0 - y)p[F(y_0 - y)] = f(y_0 + y)p[1 - F(y_0 + y)] = f(y_0 + y)p[F(y_0 + y)] = s(y_0 + y),$$

with the penultimate equality following because p is symmetric around 1/2.

For the second part let P be asymmetric and let F be such that its density is zero outside $I = [-K, K]$ and $(2K)^{-1}$ otherwise. Then, for $y \in I$

$$s(y) = \frac{1}{2K} p\left(\frac{1}{2K}y + \frac{1}{2}\right),$$

and therefore, for any $K > 0$, such s is an asymmetric density. □

Proof of Theorem 2. We verify that the derivative of the pdf of S is zero at y_0 by realising that

$$s'(y_0|F, P) = f'(y_0)p[F(y_0)] + f^2(y_0)p'[F(y_0)] = 0\ p(1/2) + f^2(y_0)p'(1/2) = 0$$

As $F(y)$ is non-decreasing, by the unimodality of F and P, and by analysing the equation on the right-hand side equation in (2), in particular the sign of both sides, we conclude that y_0 is the only point where $s'(y_0|F, P)$ vanishes and that it corresponds to a maximum. □

Proof of Theorem 3. The existence of moment $r \in \Re$ of S is equivalent to

$$\int_{\Re} |y|^r dS(y|F,P) = \int_{\Re} |y|^r f(y) p[F(y)] dy < \infty.$$

Left (right) moments exist if the integral is finite when the integration domain is replaced by $\Re_-$ ($\Re_+$).

We first prove that if $0 < \lim_{x \to 0^+} p(x), \lim_{x \to 1^-} p(x) < \infty$, then $M_l^S = M_r^S = M^F$. If the limits are both positive and finite, and P is a continuous distribution, then it is possible to find $0 < K_1 \leq K_2 < \infty$ such that $K_1 \leq p(x) \leq K_2$ and therefore

$$K_1 \int_A |y|^r f(y) dy \leq \int_A |y|^r f(y) p(F(y)) dy \leq K_2 \int_A |y|^r f(y) dy,$$

for any set A. Now as $\int_A |y|^r f(y) dy$ does not depend on P, Theorem 3 (ii) is proved as is the implication part of Theorem 3 (i).

We now prove sufficiency of Theorem 3 (i). Without loss of generality we prove the result when $\lim_{x \to 0^+} p(x)$ is zero or infinity and analyse the effect on M_l^S. If $\lim_{x \to 0^+} p(x) = 0$ then it is possible to find $0 < a < M^F$ and $K > 0$ such that for $y \in \Re_-$

$$p[F(y)] \leq K|y|^{-a}$$

and therefore

$$\int_{\Re_-} |y|^r dS(y|F,P) \leq K \int_{\Re_-} |y|^{r-a} f(y) dy.$$

As the right-hand side in the above equation is finite for any $r < M^F + a$ we have $M_l^S > M^F$. If $\lim_{x \to 0^+} p(x) = \infty$ then it is possible to find $0 < a < M^F$ and $K > 0$ such that for $y \in \Re_-$

$$p[F(y)] \geq K|y|^a$$

and therefore

$$\int_{\Re_-} |y|^r dS(y|F,P) \geq K \int_{\Re_-} |y|^{r+a} f(y) dy.$$

As the integral on the right-hand side in the above equation does not exist for any $r > M^F - a$ we obtain $M_l^S < M^F$. The proof for M_r^S is completely analogous. □

Proof of Theorem 4. Follows immediately from the proof of Theorem 3. □

Proof of Theorem 5. If $\lim_{y \to -\infty} \frac{p[F(y)]}{|y|^b}$ is finite then, as P is a continuous distribution, it is possible to find $0 < K_1 \leq K_2 < \infty$ such that $K_1|y|^b \leq p[F(y)] \leq K_2|y|^b$ and as such,

$$K_1 \int_{\Re_-} |y|^{r+b} f(y) dy \leq \int_{\Re_-} |y|^r dS(y|F,P) \leq K_2 \int_{\Re_-} |y|^{r+b} f(y) dy.$$

Now, as the largest moment of F is M^F, the left and right-hand side of equation above are finite if and only if $r + b < M^F$ and consequently $M_l^S = M^F - b$. □

Proof of Theorem 6. The proof follows the same rationale as the proof of Theorem 5. □

References

Amoroso, L. (1925). Ricerche intorno alla curva dei redditi, *Annali de Mathematica* **2**: 132–159.

Arellano-Valle, R. B. and Genton, M. G. (2003). On fundamental skew distributions. Institute of Statistics Mimeo Series #2551, North Carolina State University.

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

Arnold, B. C. and Groeneveld, R. A. (1995). Measuring skewness with respect to the mode, *Amer. Statist.* **49**: 34–38.

Aroian, L. A. (1941). A study of R.A. Fisher's z distribution and the related F distribution, *Ann. Math. Stat.* **12**: 429–448.

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

Eugene, N. (2001). *A Generalized Normal Distribution: Properties, Estimation and Applications*, PhD thesis, Central Michigan University, Mount Pleasant, Michigan.

Eugene, N., Lee, C. and Famoye, F. (2002). Beta-Normal distribution and its applications, *Commun. Stat. A-Theor.* **31**: 497–512.

Fernández, C. and Steel, M. F. J. (1998). On Bayesian modeling of fat tails and skewness, *J. Amer. Statist. Assoc.* **93**: 359–371.

Ferreira, J. T. A. S. and Steel, M. F. J. (2004a). Bayesian multivariate skewed regression modelling with an application to firm size, in M. G. Genton (ed.), *Skew-Elliptical Distributions and Their Applications: A Journey Beyond Normality*, CRC Chapman & Hall. forthcoming.

Ferreira, J. T. A. S. and Steel, M. F. J. (2004b). Discussion of M. C. Jones, "Families of distributions arising from distributions of order statistics", *Test.* forthcoming.

Ferreira, J. T. A. S. and Steel, M. F. J. (2004c). Model comparison of coordinate-free multivariate skewed distributions, mimeo, University of Warwick.

Green, P. J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination, *Biometrika* **82**: 711–732.

Hall, B. H. (1993). The stock market's valuation of R&D investment during the 1980's, *Amer. Econ. Rev.* **83**: 259–263.

Jones, M. C. (2004). Families of distributions arising from distributions of order statistics, *Test.* forthcoming.

Jones, M. C. and Faddy, M. J. (2003). A skew extension of the t-distribution, with applications, *J. Roy. Statist. Soc. B* **65**: 159–174.

Nadarajah, S. and Kotz, S. (2003). Skewed distributions generated by the Normal kernel, *Statist. Prob. Lett.* **65**: 269–277.

Newton, M. A. and Raftery, A. E. (1994). Approximate Bayesian inference with the weighted likelihood bootstrap (with discussion), *J. Roy. Statist. Soc. B* **56**: 3–48.

Petrone, S. and Wasserman, L. (2002). Consistency of Bernstein polynomial posteriors, *J. Roy. Statist. Soc. B* **64**: 79–100.

Smith, R. L. and Naylor, J. C. (1987). A comparison of maximum likelihood and Bayesian estimators for the three-parameter Weibull distribution, *Appl. Statist.* **36**: 358–369.

Stephens, M. (2000). Bayesian analysis of mixture models with an unknown number of components - An alternative to reversible jump method, *Ann. Statist.* **28**: 40–74.