A Constructive Representation of Univariate Skewed Distributions

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

Abstract

We introduce a general perspective on the introduction of skewness into symmetric distributions. Through 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. The representation is used to introduce novel classes of skewed distributions, where we induce certain prespecified characteristics through particular choices of the skewing mechanism. Finally, we use a Bayesian linear regression framework to compare the new classes with some existing distributions in the context of two empirical examples.

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 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 construct novel classes. Finally, we use these classes in Bayesian regression modelling and briefly compare them with previously proposed classes of skewed distributions.

The most common approach to the creation of skewed distributions, and the one we follow 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 negative orthants (Fernández and Steel, 1998) and order statistics (Jones, 2004).

A 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 preserve a subset of these properties, specific to the method considered.

Here we propose a general perspective on a wide set of classes of 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 classes of skewed distributions can be compared in a common framework. The decomposition also brings inferential advantages, particularly in the elicitation of prior distributions. In addition, new skewing mechanisms can be designed so as to generate distributions with predefined 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. Any skewed version of the same

2

symmetric distribution can then be modelled directly by choosing a particular P.

In addition, we introduce new skewing mechanisms that generate distributions with a pre-specified set of properties. For example, unimodality is imposed and both the location of the mode and moment existence are left completely unaffected. Tails can either be constrained to be exactly equal or proportional. In addition, skewness does not depend on the particular choice of the symmetric distribution F.

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 for 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, new classes of skewed distributions are proposed, designed to possess a number of prespecified characteristics. In Section 4, we very briefly review some existing classes of skewed distributions, one of which can naturally be described through our representation. Section 5 describes the Bayesian regression setup. The various models are then used for the analysis of two applications in Section 6. The final section groups some concluding remarks.

3

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 $y \in (a,b)$, $-\infty \le a < b \le \infty$ is trivial. Throughout the paper, we assume that all density functions exist and are continuous. Proofs are relatively straightforward and are omitted for the sake of brevity.

The 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 $Y = F^{-1}(X)$. 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, *e.g.* by 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 clearly separated, provided the form of p is not linked to F. The latter condition is not satisfied by some existing classes, such as the ones mentioned later in Subsections 4.1 and

4.2. Also, the focus of this paper is different from that of Genton and Loperfido (2005) who define the class of generalized skew-elliptical distributions, with density $2f(y)\pi(y)$, motivated by invariance of even functions, such as quadratic forms, to the skewing function π (given certain restrictions on π). Their class can easily be represented through (1), at the cost of making the form of p depend on F, by choosing $p(x) = 2\pi[F^{-1}(x)]$.

The pdf $s(y|F, P)$ is a weighted version of $f(y)$, with weight function given by $p[F(y)]$. By specifying the skewing mechanism we fix the properties of S. 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 considering p_2, we can conclude that, irrespective of F, S will be right skewed, the left tail of the distribution will be shrunk and the right tail of the distribution will be proportional to the one of the original distribution F.

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 in Definition 1. 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 $1/2$.

Theorem 1(i) is rather intuitive, with $S = F$ corresponding to the case when P does not
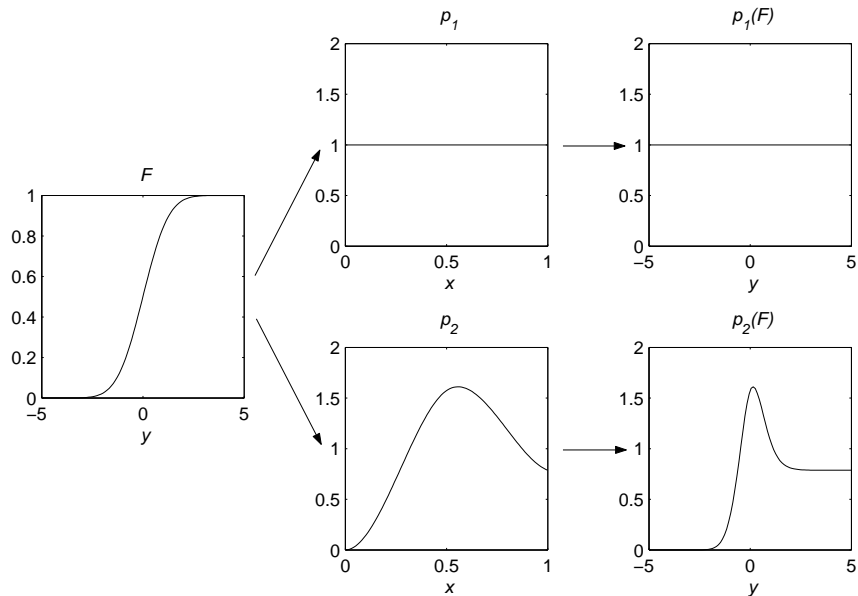
5

Figure 1: Effect of two distinct skewing mechanisms.

modify mass allocation. Theorem $1(ii)$ characterises the situation where P is never a *true* skewing mechanism, in the sense that it never introduces skewness. Thus, interesting classes of skewed distributions will typically avoid non-Uniform symmetric possibilities for P. 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 of no interest for our purpose.

Unimodality of S is often relevant in modelling real phenomena.

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, and is given by $AG = 1 - 2P(1/2)$.

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.

Theorem 3. The condition that, when x tends to zero and one, the limits of $p(x)$ are finite and non-zero implies that the moment existence of S is equal to that of F.

Intuitively, when limits of p at the extremes of $(0, 1)$ are both zero, moments may be *added* to the distribution. Similarly, when one of these limits is ∞, moments may be *removed*.

3 Constructed Skewing Mechanisms

In this section, we specifically design skewing mechanisms to generate classes of distributions with particular characteristics. Of course, one could argue about which set of characteristics is desirable, so this is partly to illustrate how the representation in (1) can be used to propose new classes of skewed distributions. However, we do feel that the particular characteristics chosen are appealing, and the resulting classes are useful for practical modelling of skewed distributions. A more restrictive special case will be analysed in Subsection 3.3, while the general class will be discussed in Subsection 3.4.

3.1 General Characteristics

There are a number of characteristics that we wish to impose:

(*i*) The skewing mechanism P does not depend on F. This really defines the focus of this paper, where we are primarily interested in distributions that allow for a meaningful separation of the skewing mechanism and the underlying symmetric distribution. It assigns distinct and clear interpretations to P and F, and also has advantages for inference and prior elicitation.

(*ii*) The only P that retrieves a symmetric S is the Uniform distribution. As we want to skew symmetric distributions, it is natural to impose (*ii*), as already suggested in Section 2.

7

The distribution F will then uniquely define a class of distributions S (for varying P) as being the only symmetric member of that class. By Theorem 1, this is imposed if the only symmetric P around $1/2$ in the class of skewing mechanisms is the Uniform distribution.

(iii) The unique mode of S equals the mode of F. This lends more interpretability to the mode, as it is unaffected by the degree of skewness, and aids modelling since the location parameter or regression function can be interpreted as the mode. It also removes the necessity of finding the mode of the distribution and simplifies the calculation of AG. From Theorem 2, we know that a choice of P with a unique mode at $1/2$ will ensure this.

(iv) Moment existence of S is the same as that of F. Thus, skewing F should not affect its tail behaviour, so that the latter can be modelled separately. Theorem 3 tells us that any $p(x)$ with finite and nonzero limits will impose this constraint.

(v) (a) Any AG skewness in $(-1, 1)$ can be achieved and (b) the AG measure is independent of F. Part (a) ensures that the class is practically useful even for extremely skewed data and (b) implies that the amount of skewness is entirely determined by the skewing mechanism.

A generic skewing mechanism $p(x)$ that always preserves unimodality and moment existence must be like the one in Figure 2.
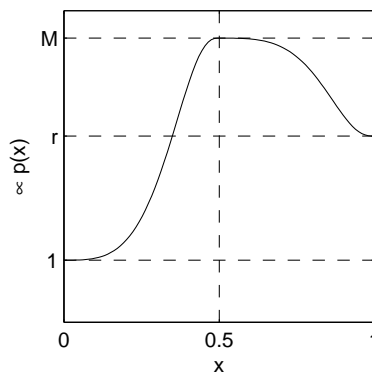

Figure 2: Skewing mechanism which preserves unimodality and moment existence.

8

Shapes of $p(x)$ as in Figure 2 can conveniently be generated by a (transition) function $g(x)$ from $[0, 1]$ to $\Re$ through

$$p(x) \propto \left[1 + \frac{M-1}{g(1/2)} \, g(x)\right] I_{(0,1/2]}(x) + \left[r + \frac{M-r}{g(1/2)} \, g(x)\right] I_{(1/2,1]}(x), \qquad (2)$$

where $g(x)$ has mode at $x = 1/2$, is decreasing in $|x - 1/2|$ and is zero at $x = 0$ and $x = 1$.

Therefore, for constructing such skewing mechanisms we have three things to choose: the function $g(x)$, the tail ratio $r > 0$ and the modal value $M \geq \max\{1, r\}$. In order to ensure that (ii) holds, we will take $M = 1$ whenever $g(x)$ is symmetric around $1/2$ and $r = 1$. Skewness is already introduced through the tails of the distribution if we choose $r \neq 1$, and we will now add a second source of skewness (around the mode) through the specification of $g(x)$.

3.2   Choosing the function g

We begin by selecting a continuous function $h(z|0)$, $z \in [0, 1/2]$, such that $h(0|0) = 0$ and $h(z|0) = 2 - h(1/2 - z|0)$. Such a function is chosen to be increasing in z and divides the rectangle $R = [0, 1/2] \times [0, 2]$ (with area one) into two equal parts (see the left plot of Figure 3). We then define a one-to-one increasing transformation v of $[0, 1/2]$ onto itself, parameterized by δ, and apply h to $v(z|\delta)$. Various choices of v are possible, but a particularly convenient one is a scaled version of $e^{\delta z} - 1$, leading to

$$h(z|\delta) = h\left[\frac{e^{\delta z} - 1}{2\left(e^{\delta/2} - 1\right)}\,\middle|\, 0\right], \qquad (3)$$

for some real $\delta \neq 0$. Since

$$\lim_{\delta \to 0} \frac{e^{\delta z} - 1}{2\left(e^{\delta/2} - 1\right)} = z, \quad \text{we have that} \quad \lim_{\delta \to 0} h(z|\delta) = h(z|0). \qquad (4)$$

The exponential transformation $v(z|\delta)$ modifies the manner in which R is divided. If $\delta \neq 0$, as illustrated by the centre plot in Figure 3, R is no longer divided into regions of equal size. If $\delta > 0$ ($\delta < 0$), the region represented with the darker shading will have larger (smaller) area
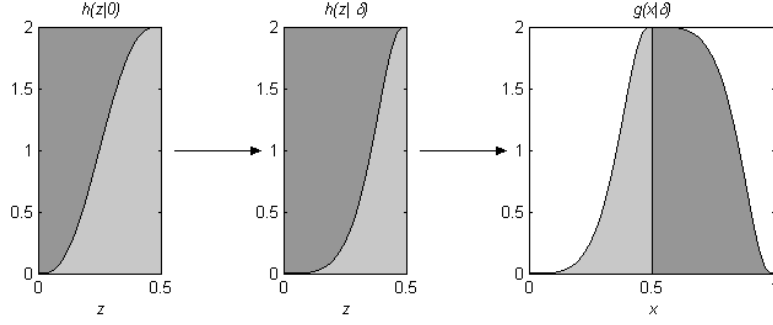
9

Figure 3: Illustration of the definition of $g(x|\delta)$.

than the region with the lighter shading. By varying δ, the region represented with the darker

shading can have any area in $(0,1)$. We finally define g as

$$g(x|\delta) = h(x|\delta)I_{[0,1/2]}(x) + \left[2 - h\left(x - \frac{1}{2}|\delta\right)\right]I_{(1/2,1]}(x), \tag{5}$$

which, with our choice of v, implies that $g(x|\delta) = g(1 - x| - \delta)$. As illustrated in Figure 3, g

has the properties mentioned in the previous subsection and integrates to one over its domain.

Also, the area to either side of the mode can take any value in $(0,1)$. If we define

$$A_\delta = \int_0^{1/2} g(x|\delta)dx = 1 - \int_{1/2}^1 g(x|\delta)dx,$$

the right-hand side of (2) integrates to $[A_\delta(r - 1) + M + 1]/2$. We can also immediately derive

that the implied amount of skewness is

$$AG = \frac{A_\delta(r + 1) + M(1 - 2A_\delta) - 1}{A_\delta(r - 1) + M + 1}. \tag{6}$$

With such a choice of g, p as defined in (2) 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 ith derivatives $h^{(i)}(0|0) =$

$h^{(i)}(1/2|0) = 0$, $i = 1, \ldots, d$, 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 $h(z|0)$. The restrictions on $h(z|0)$ mentioned at

the start of this subsection are easily imposed by choosing for $h(z|0)$ twice the cdf of a symmetric

10

distribution on $[0, 1/2]$. Taking the cdf of a Beta distribution with parameters $(d+1, d+1)$ leads to a convenient form for h, which is given by a polynomial of degree $2d+1$ satisfying $h^{(i)}(0|0) = h^{(i)}(1/2|0) = 0$, $i = 1, \ldots, d$. For various values of d, this gives:

$$d = 0 \quad : \quad h(z|0) = 4z$$

$$d = 1 \quad : \quad h(z|0) = -32z^3 + 24z^2$$

$$d = 2 \quad : \quad h(z|0) = 384z^5 - 480z^4 + 160z^3, \quad \text{etc.}$$

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

Now that we have fully specified g, we choose values for r and M in the next subsections.

3.3   Equal Tail Behaviour

First we consider a relatively simple case, where instead of only preserving moments as in characteristic (iv), we wish to impose the stronger constraint:

$(iv^\star)$ Both tails of S are not merely proportional, but equal. This means that skewness is introduced only around the mode of the distribution.

In this case we need to choose $r = 1$ in (2), which implies that (6) simplifies to

$$AG = \frac{(M-1)(1 - 2A_\delta)}{M+1}. \tag{7}$$

and, thus, AG is bounded between $1 - 2A_\delta$ and zero.

The definition of the skewing mechanism is now completed by choosing a value for M. Due to the one-to-one correspondence in (7) between AG and M, given $\delta \neq 0$ (and g), we can, equivalently, choose a particular amount of skewness, say, $AG^*(\delta)$. Expressing M as a function of $AG^*(\delta)$, the skewing mechanism has density given by

$$p(x|\delta) = 1 + \frac{AG^*(\delta)}{1 - 2A_\delta} \left[g(x|\delta) - 1 \right].$$

11

From (7), we deduce that $|AG^*(\delta)| < |1 - 2A_\delta|$. Straightforward analysis shows that for fixed δ, the value of $|1 - 2A_\delta|$ is minimized when $h(z|0) = 4z$. Using this linear h we have the following bound which will be valid (in the sense that there always exists a value $M > 1$ that leads to this $AG^*(\delta)$) for any d:

$$|AG^*(\delta)| < \left| 1 - 4 \int_0^{1/2} \frac{e^{\delta x} - 1}{e^{\delta/2} - 1} dx \right| = \left| 1 - \frac{4}{\delta} + \frac{2}{e^{\delta/2} - 1} \right|. \qquad (8)$$

In order to always satisfy (8), we choose

$$AG^*(\delta) = \text{sign}(\delta) \left(1 - \frac{4}{\delta} + \frac{2}{e^{\delta/2} - 1} \right)^2,$$

where $\text{sign}(\cdot)$ denotes the usual sign function.

This choice ensures that symmetry only occurs for $\delta = 0$, for which the skewing mechanism is chosen to be a Uniform distribution, thus imposing (ii). It also satisfies (v) and is an odd and monotonic function of the skewness parameter $\delta \in \Re$. As a consequence, $p(x|\delta) = p(1 - x| - \delta)$.

3.4 Proportional Tail Behaviour

This, more general, case allows for the tails to be proportional and thus, imposes only (iv) for the tail behaviour.

We now have two possible sources of skewness: the central part of the distribution (parameterised by δ) and the tails (parameterised by r). We feel it is sensible to impose that both sources lead to skewness in the same direction and, thus, we consider the following possible combinations of δ and r: $\{(\delta, r) : \delta \in \Re, \ r > 0, \ \text{and} \ \delta \log(r) \geq 0\}$.

From (6), we can show that, if we fix A_δ and allow r to vary, the space of achievable skewness is given by the shaded area in Figure 4(a).

Once δ and r are fixed, we must still choose M. Fortunately, for any $(\delta, r) \neq (0, 1)$, we still have a one-to-one relationship between AG and M through (6). Thus, as before, we choose M so as to obtain a particular amount of skewness, $AG^*(\delta, r)$. For fixed δ and r, achievable AG is
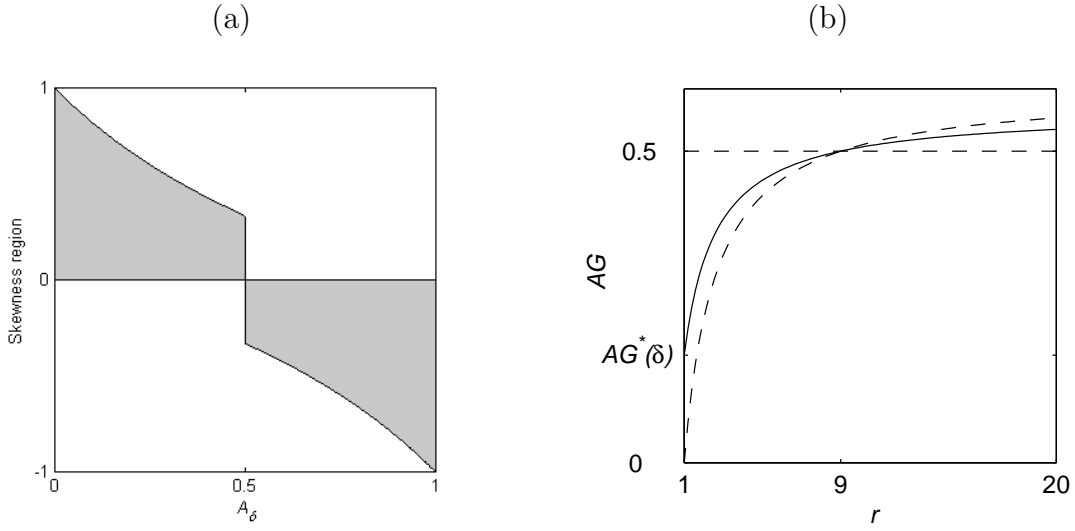
12

Figure 4: (a) Region of achievable skewness when r is allowed to vary. (b) Skewness bands for fixed δ. The dashed lines represent the bounds of achievable skewness and the solid line denotes $AG^*(\delta, r)$.

bounded in between $1 - 2A_\delta$ and a boundary denoted by $b(\delta, r)$, given by

$$\frac{A_\delta(1-r)+r-1}{A_\delta(r-1)+r+1} \text{ if } \delta \geq 0, \quad \frac{A_\delta(1-r)}{A_\delta(1-r)-2} \text{ if } \delta < 0. \tag{9}$$

For $r = [(1-A_\delta)/A_\delta]^2$ both boundaries have the same value, $1 - 2A_\delta$.

We select $AG^*(\delta, r)$ such that it always falls inside these bounds of achievable skewness and such that $AG^*(\delta, 1)$ equals $AG^*(\delta)$, as defined for the equal tails case in the previous subsection. Such an $AG^*(\delta, r)$ is given by

$$AG^*(\delta, r) = AG^*(\delta) + k_\delta b(\delta, r), \tag{10}$$

where k_δ is chosen such that

$$AG^*\left[\delta, \left(\frac{1-A_\delta}{A_\delta}\right)^2\right] = 1 - 2A_\delta.$$

The dashed lines in Figure 4(b) show the boundaries of AG, for δ (and g) such that $A_\delta = 0.25$ (thus, $\delta > 0$): the horizontal line denotes $1 - 2A_\delta$ and the curved dashed line is $b(\delta, r)$ in (9) as a function of $r > 1$ (in keeping with our choice of $\delta > 0$). The solid line represents $AG^*(\delta, r)$ in (10). For $\delta < 0$ and $r < 1$ an analogous graph can be drawn.

13

The resulting skewing mechanism follows from (2) by expressing M as a function of $AG^*(\delta, r)$ (through inverting (6)) and choosing $g(x|\delta)$ in (5), and satisfies $p(x|\delta, r) = p(1 - x| - \delta, 1/r)$.

4   Other Classes

We now briefly mention three common methods of generating univariate skewed distributions from symmetric ones. Only the third fits naturally into the framework introduced in Section 2. For the others, that can only be done by violating characteristic (i) of Subsection 3.1.

4.1   Hidden Truncation

Skewed distributions generated by hidden truncation ideas are probably the most common and most intensively studied. The skew-Normal distribution in Azzalini (1985) constitutes the first formulation of such a distribution specifically for skewness modelling. The most common versions of skewed distributions generated by hidden truncation have densities of the form

$$s(y) = 2f(y)G(\alpha y), \tag{11}$$

where G denotes the cdf of a symmetric distribution on $\Re$ and $\alpha \in \Re$. When G equals F (as assumed in the applications below) the distributions are usually denoted as skew-F.

4.2   Inverse Scale Factors

Another method for introducing skewness into a unimodal distribution F symmetric around the origin was introduced in Fernández and Steel (1998). The basic idea is 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

$$s(y|\gamma) = \frac{2}{\gamma + \frac{1}{\gamma}} f\left[y\gamma^{-\text{sign}(y)} \right]. \tag{12}$$

14

4.3   Order Statistics

The class of skewed distributions based on order statistics 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 p of this class is given by a Beta distribution parameterised by $\boldsymbol{\psi} = (\psi_1, \psi_2) \in \Re_+^2$. The class of distributions generated by this framework is also denoted by the generalised class (Amoroso, 1925). Ferreira and Steel (2004) suggested the use of a restricted parameterisation $\boldsymbol{\psi} = (\phi, 1/\phi)$, $\phi \in \Re_+$. With such a parameterisation, p is monotonic, and for the three choices of F used in the application, it always leads to unimodal S. In addition, this imposes characteristic (ii) and greatly aids in prior elicitation.

5   Regression Modelling

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

$$y_i = \boldsymbol{x}_i' \boldsymbol{\beta} + \sigma \epsilon_i, \tag{13}$$

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 or 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 adopt an Inverted-Gamma prior on σ^2 with parameters a and b. The priors on the parameters of the distributions of ϵ_i are chosen through the prior matching ideas suggested in Ferreira and Steel (2004). In particular, we fix a prior distribution on the AG measure and derive equivalent priors on the skewness parameters. For the classes generated as in Sections 3 and 4.2 the equivalence

15

is immediate and does not vary with the distribution that is being skewed. For the methods in Sections 4.1 and 4.3, only an approximate equivalence is possible.

6   Examples

In both applications, we use a polynomial h with $d = 2$. This leads to skewed densities s with two continuous derivatives provided f is smooth enough, and we feel that is sufficient. If smoother densities are required, a higher d can readily be used. We set $v = 100$ and $a = b = 0.1$, leading to 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 and is relatively flat for other values.

We compare skew versions of three distributions with different tail behaviour: the Normal, the Logistic and the t_2 distribution. Instead of the t_2, we could also use the more flexible t_ν distribution and conduct inference on the degrees of freedom parameter ν with an appropriate prior. For comparison, we also provide results for the original, symmetric, distributions.

For the distributions generated by order statistics we use the parameterisation suggested in Ferreira and Steel (2004). For this class we do not present results for the skewed Normal, as there is no one-to-one correspondence between ϕ and $AG(\phi)$ in that case, ruling out prior matching.

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 analysed in Jones and Faddy (2003). As in the latter paper, we use a location-scale model.

Table 1 presents the natural logarithm of the Bayes factors for all models with respect to the symmetric Normal model. Larger values indicate more support for the model (*i.e.* a better fit), and, assuming all models equally probable *a priori*, ratios of posterior probabilities (posterior

16

odds) between two models are given by the exponential of the difference in log Bayes factors. Thus, the hidden truncation model based on the t_2 does best. The evidence in favour of skewness is clear: the symmetric models perform worst throughout. The constructed mechanism with equal tails (Const-Eq.) was outperformed by the other skewed distributions for the distributions based on the Logistic and t_2. Allowing the tails to vary pays dividends: the constructed version with different tails (Const-Dif.) does a lot better for these cases, and also improves on the equal tails version for Normal F. Non-Normal tails are favoured for all distributions.

Table 1: Glass fibre data: 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.	Const-Eq.	Const-Dif.
Normal	0	3.15	2.66	-	3.39	3.83
Logistic	1.64	4.80	4.41	4.38	3.41	4.94
t_2	2.65	5.48	4.90	4.45	3.43	4.88

In Figure 5(a) we present estimates of the posterior predictive densities for the different skewed t_2 models, overlaid on a normalised histogram of the data. To avoid cluttering, the result for the order statistics model is not presented, but it is very close to that with inverse scale factors. The estimates are quite similar and show that left skewed distributions are more appropriate. A similar conclusion was reached by Jones and Faddy (2003). Insight into the tail behaviour is provided by Figure 5(b) where the logarithm of the posterior predictive densities is plotted. Clearly, the tail estimates for the constructed model with equal tails 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 5(a) the tails indeed appear to be different in the data. The constructed model with different (proportional) tails is much closer to the other models.
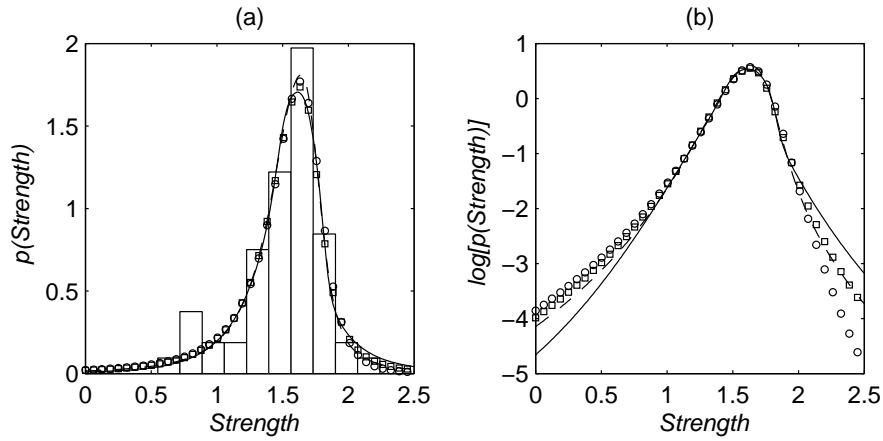
17

Figure 5: Glass fibre data: Posterior predictive densities (a) and log-densities (b) for the Hidden Truncation (circles), Inverse Scale Factors (squares), Constructed with Equal Tails (solid) and Constructed with Different Tails (dashed) 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 described in Hall (1993).

The data consist of 300 observations. We model 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. There is strong support for non-Normal tails, especially in the symmetric case. There is also evidence in favour of skewness, especially in the Normal case. For each choice of F it is a constructed model that performs best: the one with different tails is the best model for the Logistic case and this combination also has the highest overall marginal likelihood. The constructed model with equal tails actually beats the version with different tails in the t_2-case. The latter fact illustrates

18

the built-in parsimony preference of Bayes factors. 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 8.87, which corresponds to posterior odds of over 7100 to one. The Logistic distribution is favoured for all skewing mechanisms except the constructed with equal tails which favours the t_2 distribution.

Table 2: Market value data: 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.	Const-Eq.	Const-Dif.
Normal	0	14.70	11.32	-	19.70	20.19
Logistic	20.78	25.37	26.24	26.40	23.70	28.44
t_2	19.26	24.52	23.41	21.71	25.93	24.22

7 Conclusion

In this article, we provide a new perspective on the introduction of skewness into symmetric distributions through the use of a skewing mechanism which does not depend on the underlying symmetric distribution. Through inverse integral probability transformations we 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. In addition, the representation proposed here enables the definition of new classes, which can have certain prespecified characteristics. Whereas all continuous distributions can be put in the framework of our decomposition, most existing classes of skewed distributions imply that the form of the skewing mechanism depends on the underlying symmetric distribution. Thus, our classes are complementary to the main alternatives in the literature.

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. The insights derived from the constructive representation should make it easier to choose particular classes of skewed models for a given application, and help us to understand why certain classes work better than others. As we illustrated in Section 3, it also provides a powerful framework for the development of new classes of skewed distributions, with certain prespecified 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.

Matlab code is available from the authors upon request.

References

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

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, *Am. Statist.* **49**: 34–38.

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

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. (2004). Model comparison of coordinate-free multivariate skewed distributions with an application to stochastic frontiers, Statistics Research Report 423, University of Warwick.

Genton, M. G. and Loperfido, N. (2005). Generalized skew-elliptical distributions and their quadratic forms, *Ann. I. Stat. Math.* forthcoming.

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 (with discussion), *Test* **13**: 1–43.

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.

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.