The Minimax Distribution: A Beta-Type Distribution With Some Tractability Advantages

M.C. Jones

A two-parameter family of distributions on $(0, 1)$ is explored which has many similarities to the beta distribution and a number of advantages in terms of tractability (it also, of course, has some disadvantages). The minimax distribution, so-called because of its genesis in terms of uniform order statistics, particularly has straightforward distribution and quantile functions which do not depend on special functions (and hence afford very easy random variate generation). The distribution might, therefore, have a particular role when a quantile-based approach to statistical modelling is taken, and its tractability has appeal for pedagogical uses.

KEY WORDS: Beta distribution; Distribution theory; Order statistics.

1. INTRODUCTION

Despite the many alternatives and generalisations (Kotz and van Dorp, 2004, Nadarajah and Gupta, 2004), it remains fair to say that the beta distribution provides the premier family of continuous distributions on bounded support (which is taken to be $(0, 1)$). The beta distribution, Beta(a, b), has density

$$g(x) = \frac{1}{B(a,b)}\, x^{a-1}(1-x)^{b-1}, \qquad 0 < x < 1, \tag{1.1}$$

where its two shape parameters a and b are positive and $B(\cdot,\cdot)$ is the beta function. Beta densities are unimodal, uniantimodal, increasing, decreasing

M.C. Jones is Professor of Statistical Science, Department of Statistics, The Open University, Walton Hall, Milton Keynes, MK7 6AA, U.K. (E-mail: *m.c.jones@open.ac.uk*).

or constant depending on the values of a and b relative to 1 and have a host of other attractive properties (Johnson, Kotz and Balakrishnan, 1994, Chapter 25). The beta distribution is pretty tractable, but in some ways not fabulously so; in particular, its distribution function is an incomplete beta function ratio and its quantile function the inverse thereof.

In this paper, I look at an alternative two-parameter distribution on $(0, 1)$ which I call the minimax distribution, Minimax(α, β), where I have called its two positive shape parameters α and β. It has many of the same properties as the beta distribution but has some advantages in terms of tractability. Its density is

$$f(x) = f(x; \alpha, \beta) = \alpha\beta x^{\alpha-1}(1 - x^{\alpha})^{\beta-1}, \qquad 0 < x < 1. \tag{1.2}$$

This is not entirely new and alert readers might recognise it in some way, but it seems that this distribution has not been investigated systematically before nor has its relative interchangeability with the beta distribution been appreciated. For example, minimax densities are also unimodal, uniantimodal, increasing, decreasing or constant depending in the same way on the values of α and β. (Boundary behaviour and the main special cases are also common to beta and minimax distributions.) And yet the normalising constant in (1.2) is very simple and the corresponding distribution and quantile functions also need no special functions. The latter gives the minimax distribution an advantage if viewed from the quantile modelling perspective popular in some quarters (Parzen, 1979, Gilchrist, 2001). Some other properties are also more readily available, mathematically, than their counterparts for the beta distribution. Yet the beta distribution also has its particular advantages and I hesitate to claim whether, in the end, the tractability advantages of the minimax distribution will prove to be of immense practical significance; at the very least, the minimax distributions might find a pedagogical role.

The background and genesis of the minimax distribution are given in Section 2 along with its principal special cases. The basic properties of the minimax distribution are given in Section 3 whose easy reading reflects the tractability of the distribution. A deeper investigation of skewness and kurtosis properties of the minimax distribution is given in Section 4. Inference by maximum likelihood is investigated in Section 5 while, in Section 6, a number of further related distributions are briefly considered. It should be the case that similarities and differences between the beta and minimax distributions are made clear as the paper progresses but, in any case, they are summarised and discussed a little more in the closing Section 7.

Note that the linear transformation $\ell + (u - \ell)X$ moves a random variable X on $(0, 1)$ to any other bounded support (ℓ, u). So, provided ℓ and u don't depend on α or β and are known, there is no need to mention such an extension further.

2. GENESIS, FOREBEARS AND SPECIAL CASES

Temporarily, set $a = m, b = n + 1 - m$ in the beta distribution, where m and n are positive integers. Then, as is well known, the beta distribution is the distribution of the m'th order statistic from a random sample of size n from the uniform distribution (on $(0, 1)$). Now consider another simple construction involving uniform order statistics. Take a set of n independent random samples each of size m from the uniform distribution and collect their maxima; take X, say, to be the minimum of the set of maxima; then X has the minimax distribution — terminology thus explained — with parameters m and n. (Likewise, by symmetry, one minus the maximum of a set of n minima of independent uniform random samples of size m has the same minimax distribution.)

Now, the maxima themselves constitute a random sample of size n from the power function distribution which is the Beta$(m, 1)$ distribution (it has density $mx^{m-1}, 0 < x < 1$). Hence X is also the minimum of a random sample from the power function distribution.

As with the beta distribution, for greater generality the integer-valued parameters m and n may be replaced in the minimax distribution by real-valued, positive, parameters α and β.

It is also clear that both beta and minimax distributions are special cases of the three-parameter distribution with density

$$g(x) = \frac{p}{B(\gamma, \delta)}\, x^{\gamma p - 1}(1 - x^p)^{\delta - 1}, \qquad 0 < x < 1, \tag{2.1}$$

and $p > 0$. This is the generalised beta distribution of McDonald (1984). It is the distribution of the $1/p$'th power of a Beta(γ, δ) random variable or of the γ'th order statistic of a sample of size $\gamma + \delta - 1$ from the power function distribution Beta$(p, 1)$ (for γ, δ integer). (The order statistics version of the generalised beta density can be found, for example, in Example 2.2.2 of Arnold, Balakrishnan and Nagaraja, 1992.) Beta and minimax distributions are therefore the $(p, \gamma, \delta) = (1, a, b)$ and $(\alpha, 1, \beta)$ special cases, respectively, of (2.1). The similarities between beta and minimax distributions that will become clear through the rest of this article lead, however, to the conclusion that this generalised beta distribution is not very useful in practice.

The beta and minimax distributions share their main special cases. Beta(a, 1) and Minimax(α, 1) are both the power function distribution mentioned above and Beta(1, a) and Minimax(1, α) are the distribution of one minus that power function random variable. Beta(1, 1) and Minimax(1, 1) are both the uniform distribution.

A further special case of the minimax distribution has also appeared elsewhere. The Minimax(2, β) distribution is that of the "generating variate" $R = \sqrt{x_1^2 + x_2^2}$ when $\{x_1, x_2\}$ follow a bivariate Pearson Type II distribution (Fang, Kotz and Ng, 1990, Section 3.4.1). This has been used in an algorithm to generate univariate symmetric beta random variates by Ulrich (1984) and Devroye (1986, p.436).

3. BASIC PROPERTIES

The distribution function of the minimax distribution is

$$F(x) = 1 - (1 - x^\alpha)^\beta, \qquad 0 < x < 1. \tag{3.1}$$

This compares extremely favourably in terms of simplicity with the beta distribution's incomplete beta function ratio.

The distribution function is readily invertible to yield the quantile function

$$Q(y) = F^{-1}(y) = \{1 - (1 - y)^{1/\beta}\}^{1/\alpha}, \qquad 0 < y < 1. \tag{3.2}$$

As already mentioned, this facilitates ready quantile-based statistical modelling (Parzen, 1979, Gilchrist, 2001). Moreover, I know of no other two-parameter quantile family on (0,1) so simply defined and yet with such good behaviour (as follows, with an explicit simple density function to boot!). The popular generalised lambda family (Ramberg and Schmeiser, 1984, Gilchrist, 2001, Section 7.3) with quantile function $y^\gamma - (1 - y)^\delta$ has support (0,1) for $\gamma, \delta > 0$ but encompasses bimodality, repeated incarnations of the uniform distribution and complicated patterns of skewness and kurtosis (e.g. Karvanen and Nuutinen, 2007). A better competitor, almost as tractable as the minimax distribution and with some similar properties, is Tadikamalla and Johnson's (1982) L_B distribution.

Formula (3.2) also facilitates trivial random variate generation. If $U \sim U(0,1)$, then $X \sim f$ if

$$X = (1 - U^{1/\beta})^{1/\alpha}. \tag{3.3}$$

This compares extremely favourably with the sophisticated algorithms preferred to generate random variates from the beta distribution (Devroye, 1986, Section IX.4, Johnson et al., 1994, Section 25.2).

It can be shown that the minimax distribution has the same basic shape properties as the beta distribution, namely:

$$\alpha > 1, \beta > 1 \;\Rightarrow\; \text{unimodal}; \qquad \alpha < 1, \beta < 1 \;\Rightarrow\; \text{uniantimodal};$$

$$\alpha > 1, \beta \leq 1 \;\Rightarrow\; \text{increasing}; \qquad \alpha \leq 1, \beta > 1 \;\Rightarrow\; \text{decreasing};$$

$$\alpha = \beta = 1 \;\Rightarrow\; \text{constant}.$$

In the first two cases, the mode/antimode is at

$$x_0 = \left(\frac{\alpha - 1}{\alpha\beta - 1} \right)^{1/\alpha}. \tag{3.4}$$

Both beta and minimax densities are log-concave if and only if both their parameters are greater than or equal to 1.

The behaviour of the minimax density also matches that of the beta density at the boundaries of their support:

$$f(x) \sim x^{\alpha-1} \;\text{ as }\; x \to 0;$$

$$f(x) \sim (1-x)^{\beta-1} \;\text{ as }\; x \to 1.$$

Some illustrative examples of minimax densities are plotted in Figure 1. Note the similarity to analogous depictions of the beta family (e.g. Johnson et al., 1994, Figure 25.1b).

* * * Figure 1 about here * * *

In Jones (2002), I suggested swapping the roles of distribution and quantile functions for distributions on $(0,1)$ to produce "complementary distributions". There, I applied the idea to the beta distribution to produce the complementary beta distribution. Applying the same idea here, a pleasing symmetry is observed that means that nothing is really gained. The density of the complementary distribution is the quantile density function, $q(y) = Q'(y)$, of the minimax distribution, and

$$q(y) = \frac{1}{\alpha\beta}(1-y)^{(1/\beta)-1}\{1-(1-y)^{1/\beta}\}^{(1/\alpha)-1} = f(1-y; 1/\beta, 1/\alpha).$$

The moments of the minimax distribution are both immediate and obtainable from McDonald (1984) or Arnold et al. (1992, p.15):

$$E(X^r) = \beta \mathrm{B}\left(1+\frac{r}{\alpha}, \beta\right). \tag{3.5}$$

Like those of the beta distribution they exist for all $r > -\alpha$. I note the first appearance of a special function, although it is still only the (complete) beta function. (It might be interesting to note the moment formula in terms of the binomial coefficient with non-integer arguments: $E(X^r) = 1 \Big/ \binom{\beta+(r/\alpha)}{\beta}$.) In particular,

$$E(X) = \beta \mathrm{B}\left(1+\frac{1}{\alpha}, \beta\right);$$

$$\mathrm{Var}(X) = \beta \mathrm{B}\left(1+\frac{2}{\alpha}, \beta\right) - \left\{\beta \mathrm{B}\left(1+\frac{r}{\alpha}, \beta\right)\right\}^2.$$

The first L-moment is the mean. Expressions for higher L-moments (Hosking, 1990, Hosking and Wallis, 1997) are explicitly available — an improvement on the situation for the beta distribution. However, the general formula for the r'th L-moment is a sum of r terms involving several gamma functions; see Appendix A for details. Here, just the scale measure which is the second L-moment (half the Gini mean difference) is given:

$$\lambda_2 = \beta\left\{\mathrm{B}\left(1+\frac{1}{\alpha}, \beta\right) - 2\mathrm{B}\left(1+\frac{1}{\alpha}, 2\beta\right)\right\}.$$

Distributions of order statistics and their moments are also relatively tractable but not especially edifying (but this too is an improvement on the situation for the beta distribution). They are dealt with briefly in Appendix B.

4. SKEWNESS AND KURTOSIS

4.1 Skewness

A strong skewness ordering between distributions is the classical one due to van Zwet (1964). It is immediate that skewness to the right increases, in this sense, with decreasing α for fixed β. This is because the transformation from $X_1 \sim \text{Minimax}(\alpha_1, \beta)$ to a $\text{Minimax}(\alpha_2, \beta)$ random variate is of the form $X_1^{\alpha_1/\alpha_2}$ which is convex for $\alpha_1 > \alpha_2$. There seems to be no such simple property for changing β and fixed α.

Various scalar skewness measures can be plotted as functions of α and β. For example, the L-skewness (Hosking, 1990; formula from Appendix A),

$$\tau_3 = \frac{\lambda_3}{\lambda_2} = \frac{B\left(1+\frac{1}{\alpha},\beta\right) - 6B\left(1+\frac{1}{\alpha},2\beta\right) + 6B\left(1+\frac{1}{\alpha},3\beta\right)}{B\left(1+\frac{1}{\alpha},\beta\right) - 2B\left(1+\frac{1}{\alpha},2\beta\right)},$$

is shown in Figure 2. Note the increasing nature of the skewness as Figure 2 is traversed diagonally from bottom right to top left. Of course, this measure respects the van Zwet ordering, being a decreasing function of α for fixed β. Similar patterns arise for the classical third-moment measure and the quantile-based skewness measure, $\{Q(3/4) - 2Q(1/2) + Q(1/4)\}/\{Q(3/4) - Q(1/4)\}$ (Bowley, 1937); these are not shown. For $\alpha, \beta > 1$, one can also plot Arnold and Groeneveld's (1995) skewness measure $1 - 2F(x_0)$, with similar results again (not shown).

* * * Figure 2 about here * * *

The pattern in Figure 2 is broadly in line with similar pictures for the beta distribution, although the latter are symmetric about the line $a = b$. They both contrast with the more complicated patterns of skewness associated with the generalised lambda distribution (Karvanen and Nuutinen, 2007, Figure 7(a),(c) for τ_3).

4.2 Symmetry and Near-Symmetry

Perhaps the least attractive feature of the minimax distribution, at least at first thought, is that, unlike the beta distribution, it has no symmetric members other than the uniform distribution ($\alpha = \beta = 1$). It does, however, have a range of almost-symmetric members. Some of these are shown in Figure 3. They correspond to zero skewness in the sense of Arnold and Groeneveld (1995). I chose this measure (and used it for uniantimodal densities as well as unimodal ones) because it has a simple formula:

$$\gamma = 1 - 2F(x_0) = 2\left\{\frac{\alpha(\beta-1)}{\alpha\beta-1}\right\}^{\beta} - 1;$$

it is easy to see that $\gamma = 0$ whenever

$$\alpha = 1/\left\{\beta - (\beta-1)2^{1/\beta}\right\}.$$

The zero curves for other skewness measures follow very similar trajectories.

* * * Figure 3 about here * * *

The densities in Figure 3 are indeed reasonably symmetric and, provided α and β are not too big, resemble symmetric beta distributions. It is only when β becomes large that, while almost-symmetry is retained (as is a small variance), something happens which is different from — but not necessarily less desirable than — the beta distribution: the modal location, (3.4), shifts away from the centre. (Almost-symmetry about a point towards the right of the unit interval would necessitate $f(1-x)$.)

4.3 Kurtosis

The L-kurtosis (Hosking, 1990; formula from Appendix A), defined by $\tau_4 = \lambda_4/\lambda_2$, is

$$\tau_4 = \frac{B\left(1+\frac{1}{\alpha},\beta\right) - 12B\left(1+\frac{1}{\alpha},2\beta\right) + 30B\left(1+\frac{1}{\alpha},3\beta\right) - 20B\left(1+\frac{1}{\alpha},4\beta\right)}{B\left(1+\frac{1}{\alpha},\beta\right) - 2B\left(1+\frac{1}{\alpha},2\beta\right)};$$

it is plotted as a function of α and β in Figure 4. Notice how the L-kurtosis increases along the line of 'near-symmetry' as the parameters get bigger. It also increases as one goes away from near-symmetry when either parameter decreases. This behaviour is also observed for the classical fourth-moment and the third-difference-quantile-based kurtosis measures (not shown) as well as being, once more, in line with similar pictures for the beta distribution, except for the symmetry about the diagonal line there. Again, the generalised lambda distribution on $(0,1)$ displays a more complicated pattern of kurtosis (Karvanen and Nuutinen, 2007, Figure 7(b),(d) for τ_4).

* * * Figure 4 about here * * *

5. LIKELIHOOD INFERENCE

In this section, I consider maximum likelihood estimation for the minimax distribution. Let $X_1, ..., X_n$ be a random sample from the minimax distribution and let circumflexes denote maximum likelihood estimates of parameters. Differentiating the log likelihood with respect to β leads immediately to the relation

$$\widehat{\beta} = -n \Big/ \sum_{i=1}^{n} \log(1 - X_i^{\hat{\alpha}}) \, . \tag{5.1}$$

I can now concentrate on the equation to be satisfied by $\widehat{\alpha}$ arising from substituting for $\widehat{\beta}$ in the other score equation (Mäkeläinen, Schmidt and Styan, 1981, p.762). This is

$$S(\alpha) \equiv \left\{ \frac{1}{\alpha} + T_1(\alpha) + \frac{T_2(\alpha)}{T_3(\alpha)} \right\} = 0 \tag{5.2}$$

where

$$T_1(\alpha) = n^{-1} \sum_{i=1}^{n} \frac{\log X_i}{1 - Y_i}, \quad T_2(\alpha) = n^{-1} \sum_{i=1}^{n} \frac{Y_i \log X_i}{1 - Y_i}, \quad T_3(\alpha) = n^{-1} \sum_{i=1}^{n} \log(1 - Y_i)$$

and $Y_i = X_i^{\alpha}$, $i = 1, ..., n$, remembering that each Y_i is a function of α. It is shown in Appendix C that $\lim_{\alpha \to 0} S(\alpha) > 0$ and $\lim_{\alpha \to \infty} S(\alpha) < 0$ and so, given the continuity of $S(\alpha)$, there is at least one zero of (5.2) in $(0, \infty)$. However, although it is possible to obtain various further properties of $T_j(\alpha)$ and its derivatives, $j = 1, 2, 3$, I have been able to derive no further general properties of the shape of $S(\alpha)$. The safest approach, therefore, is to evaluate $S(\alpha)$ over a fine grid of values (on the $\log \alpha$ scale), plot it if desired, identify all zeroes of the function and then evaluate the log likelihood at those zeroes to identify its global maximum.

Properties of the maximum likelihood estimates thus obtained follow from the usual asymptotic likelihood theory. It is readily shown that the elements of the observed information matrix are:

$$\widehat{\mathcal{I}}_{\alpha\alpha} = \frac{n}{\alpha^2} + \frac{(\beta - 1)}{\alpha^2} \sum_{i=1}^{n} \frac{Y_i (\log Y_i)^2}{(1 - Y_i)^2}; \qquad \widehat{\mathcal{I}}_{\alpha\beta} = \frac{1}{\alpha} \sum_{i=1}^{n} \frac{Y_i \log Y_i}{(1 - Y_i)}; \qquad \widehat{\mathcal{I}}_{\beta\beta} = \frac{n}{\beta^2}.$$

And a little more manipulation taking advantage of the fact that $Y \sim \text{Beta}(1,\beta)$ gives the elements of the expected information matrix as:

$$n^{-1}\mathcal{I}_{\alpha\alpha} = \frac{\mathcal{A}}{\alpha^2}; \qquad n^{-1}\mathcal{I}_{\alpha\beta} = \frac{\mathcal{B}}{\alpha}; \qquad n^{-1}\mathcal{I}_{\beta\beta} = \frac{1}{\beta^2} \tag{5.3}$$

where

$$\mathcal{A} = \mathcal{A}(\beta) = 1 + \frac{\beta}{\beta-2}\left[\{\psi(\beta)-\psi(2)\}^2 - \{\psi'(\beta)-\psi'(2)\}\right]$$

and

$$\mathcal{B} = \mathcal{B}(\beta) = -\{\psi(\beta+1)-\psi(2)\}/(\beta-1) < 0.$$

Here, $\psi(z) = d\log\Gamma(z)/dz$ is the digamma function. I can verify that $\det(\mathcal{I}) = \mathcal{A} - \beta^2\mathcal{B}^2 > 0$ for $\beta \geq 2$ mathematically, but only have numerical evidence of its truth for $0 < \beta < 2$.

It follows from (5.3) that, asymptotically,

$$n^{-1}\text{Var}(\widehat{\alpha}) \simeq \frac{\alpha^2}{\mathcal{A}-\beta^2\mathcal{B}^2}; \quad n^{-1}\text{Var}(\widehat{\beta}) \simeq \frac{\beta^2\mathcal{A}}{\mathcal{A}-\beta^2\mathcal{B}^2}; \quad \text{Corr}(\widehat{\alpha},\widehat{\beta}) \simeq \frac{-\beta\mathcal{B}}{\sqrt{A}} > 0.$$

Notice that the standard deviation of $\widehat{\alpha}$ is proportional to α and (it seems, numerically) the constant of proportionality decreases with β; the standard deviation of $\widehat{\beta}$ is independent of α and increases with β. The correlation between the parameter estimates does not depend on α and increases with β from somewhere around $1/4$ for small β to 1 for large β; this behaviour is very similar to that of the correlation between maximum likelihood estimators of the parameters of the beta distribution traced along the line $a = b$.

6. RELATED DISTRIBUTIONS

6.1 Limiting Distributions

It is straightforward to see that if the minimax distribution is normalised by looking at the distribution of $Y = \beta^{1/\alpha}X$ (on $(0,\beta^{1/\alpha})$), then its density, $\alpha y^{\alpha-1}\{1-(y^\alpha/\beta)\}^{\beta-1}$, tends to $\alpha y^{\alpha-1}\exp(-y^\alpha)$ (on $(0,\infty)$), the density of the Weibull distribution, as $\beta \to \infty$. This is the interesting analogue of the gamma limit that arises in similar circumstances in the case of the beta distribution.

Similarly, the distribution of $Z = \alpha(1 - X)$ has limiting density

$$\beta e^{-z}(1 - e^{-z})^{\beta - 1} \tag{6.1}$$

(on $(0, \infty)$) as $\alpha \to \infty$. This is the distribution of minus the logarithm of the Beta$(1, \beta)$ power function distribution (since for finite α, Z is minus the Box-Cox transformation with power $1/\alpha$ of a Beta$(1, \beta)$ random variable). This distribution has recently become quite popular under the name "generalised exponential distribution" (Gupta and Kundu, 1999, 2007).

As both α and β become large (in any relationship to one another), the limit associated with the overall normalisation $\alpha(1 - \beta^{1/\alpha} X)$ is the extreme value density $e^{-x} \exp(-e^{-x})$.

The exact distribution of the overall normalised random variable used above is, in fact, the kappa distribution of Hosking (1994). Notice, however, that this observation does not invalidate the novelty of the remainder of the paper since the support of the kappa distribution depends on α and β while that of the minimax distribution does not. All of the limiting distributions correspond, of course, with those mentioned in Hosking (1994) or obtained from extreme value theory when $\alpha = m, \beta = n$.

6.2 Distributions Related by Transformation

Let us ignore the glib implication of the subsection title which is "all of them" and consider transformations from a limited class. First, let us briefly consider some of the most obvious transformations to positive half-line and whole real line supports. These might include the odds ratio $Y = X/(1-X)$ from $(0, 1)$ to $(0, \infty)$ and what I have argued (Jones, 2007) is a natural extension $Z = (Y - (1/Y))/2$ from $(0, \infty)$ to $(-\infty, \infty)$ which maintain the power tails of the minimax distribution. Alternatively, take logs ($Y = -\log(1-X)$, $Z = \log Y$) to decrease tailweights with each transformation. Or combine the two. In particular, the minimax odds distribution is quite an interesting heavy-tailed alternative to the F distribution. It has density

$$\alpha\beta \frac{y^{\alpha}}{(1 + y)^{\alpha\beta - 1}} \left\{(1 + y)^{\alpha} - y^{\alpha}\right\}^{\beta - 1}, \quad y > 0.$$

On the other hand, $Y = -\log(1-X)$ yields a scaled version of the exponentially-tailed distribution with density (6.1).

A natural way of generating families of distributions on some other support from a simple starting distribution with density h and distribution function H, say, is to apply the quantile function H^{-1} to a family of distributions on $(0, 1)$. See Jones (2004) for this idea applied to the beta distribution. For the minimax distributions, the resulting family has density

$$\alpha\beta h(x)H^{\alpha-1}(x)\{1 - H^{\alpha}(x)\}^{\beta-1}.$$

The transformations of the previous paragraph can, of course, be interpreted in this light, and are amongst the simplest transformations of this type. I particularly like to generate families of distributions on the whole real line from a symmetric H, α and β then becoming the shape parameters associated with the family.

7. CONCLUSIONS

To assist the reader in deciding whether the minimax distribution might be of use to him or her in terms of either research or teaching, I summarise the pros, cons and equivalences between the two below. (Some of the pros of the beta distribution have not been mentioned previously in this paper.)

The minimax and beta distributions have the following attributes in common:

- their general shapes (unimodal, uniantimodal, monotone or constant) and the dependence of those shapes on the values of their parameters;
- power function and uniform distributions as special cases;
- straightforward interpretations in terms of order statistics from the uniform distribution;
- explicit expressions for the mode/antimode (where appropriate);
- behaviour of densities as $x \to 0, 1$;
- good behaviour of skewness and kurtosis measures as functions of the parameters of the distribution;
- broadly similar maximum likelihood estimation;
- simple standard (if different) limiting distributions.

The minimax distribution has the following advantages over the beta distribution:

• a simple explicit formula for its distribution function not involving any special functions;

• ditto for the quantile function;

• as a consequence of the simplicity of the quantile function, a simple formula for random variate generation;

• explicit formulae for L-moments;

• simpler formulae for moments of order statistics.

The beta distribution has the following advantages over the minimax distribution:

• simpler formulae for moments and the moment generating function;

• a one-parameter subfamily of symmetric distributions;

• simpler moment estimation;

• more ways of generating the distribution via physical processes;

• in Bayesian analysis, conjugacy with a simple distribution, the binomial.

This paper has offered the minimax distribution as a viable alternative to the beta distribution that shares many of the latter's properties while being easier to handle in several ways. I cautiously commend it to the reader. However, the minimax distribution is certainly not superior to the beta distribution in every way! But it might be worth consideration from time to time by researchers who wish to utilise one or more of its simpler properties (e.g. quantile-based work and random variate generation) and it provides a useful new example for teaching of distributions.

APPENDIX A: L-MOMENTS

A general form for the L-moments for $r \geq 2$ is

$$\lambda_r = \frac{1}{r}\sum_{j=0}^{r-2}(-1)^j \binom{r-2}{j}\binom{r}{j+1} J(r-1-j, j+1)$$

where

$$J(i_1, i_2) = \int_0^1 F^{i_1}(x)(1-F)^{i_2}(x)du$$

(Hosking, 1989). In the case of the minimax distribution

$$\begin{aligned}
J(i_1,i_2) &= \int_0^1 \{1-(1-x^\alpha)^\beta\}^{i_1}(1-x^\alpha)^{\beta i_2}dx \\
&= \frac{1}{\alpha}\int_0^1 \{1-w^\beta\}^{i_1} w^{\beta i_2}(1-w)^{(1/\alpha)-1}dw \\
&= \frac{1}{\alpha}\sum_{k=0}^{i_1}\binom{i_1}{k}(-1)^k B\left(\beta(k+i_2)+1,\frac{1}{\alpha}\right) \\
&= \beta\sum_{k=0}^{i_1}\binom{i_1}{k}(-1)^k (k+i_2) B\left(\beta(k+i_2),1+\frac{1}{\alpha}\right).
\end{aligned}$$

Therefore,

$$\begin{aligned}
\lambda_r &= \frac{\beta}{r}\sum_{j=0}^{r-2}\sum_{k=0}^{r-j-1}(-1)^{j+k}\binom{r-2}{j}\binom{r}{j+1}\binom{r-1-j}{k}(k+j+1)B\left(\beta(k+j+1),1+\frac{1}{\alpha}\right). \\
&= \frac{\beta}{r}\sum_{\ell=1}^{r}(-1)^{\ell-1}\ell\left\{\sum_{j=0}^{\min(\ell-1,r-2)}\binom{r-2}{j}\binom{r}{j+1}\binom{r-1-j}{r-\ell}\right\}B\left(\beta\ell,1+\frac{1}{\alpha}\right) \\
&= \frac{\beta}{r}\sum_{\ell=1}^{r}(-1)^{\ell-1}\ell\binom{r}{\ell}\left\{\sum_{j=0}^{\min(\ell-1,r-2)}\binom{r-2}{j}\binom{\ell}{j+1}\right\}B\left(\beta\ell,1+\frac{1}{\alpha}\right) \\
&= \frac{\beta}{r}\sum_{\ell=1}^{r}(-1)^{\ell-1}\ell\binom{r}{\ell}\binom{r+\ell-2}{r-1}B\left(\beta\ell,1+\frac{1}{\alpha}\right).
\end{aligned}$$

The final equality follows from (3.20) and (3.29) of Gould (1972).

APPENDIX B: MOMENTS OF ORDER STATISTICS

The i'th order statistic, $X_{i:n}$, of a random sample of size n from the minimax distribution has density

$$\frac{\alpha\beta}{\mathrm{B}(i,n+1-i)}x^{\alpha-1}(1-x^\alpha)^{\beta(n+1-i)-1}\{1-(1-x^\alpha)^\beta\}^{i-1}, \quad 0<x<1.$$

It follows that

$$E(X_{i:n}^r) = \frac{\alpha\beta}{\mathrm{B}(i,n+1-i)}\int_0^1 x^{r+\alpha-1}\{1-(1-x^\alpha)^\beta\}^{i-1}(1-x^\alpha)^{\beta(n+1-i)-1}dx$$

$$
\begin{aligned}
&= \frac{\beta}{\mathrm{B}(i, n+1-i)} \int_0^1 \{1 - w^\beta\}^{i-1} w^{\beta(n+1-i)-1} (1-w)^{(r/\alpha)} dw \\
&= \frac{\beta}{\mathrm{B}(i, n+1-i)} \sum_{k=0}^{i-1} \binom{i-1}{k} (-1)^k B\left(\beta(k+n+1-i), 1 + \frac{r}{\alpha}\right) \\
&= \frac{\beta}{\mathrm{B}(i, n+1-i)} \sum_{\ell=n+1-i}^{n} \binom{i-1}{n-\ell} (-1)^{\ell-(n+1-i)} B\left(\beta\ell, 1 + \frac{r}{\alpha}\right).
\end{aligned}
$$

APPENDIX C: LIMITING BEHAVIOUR OF $S(\alpha)$

(I) $\alpha \to 0$. First, note that $1 - X_i^\alpha \simeq -\alpha \log X_i > 0$ and hence that $\log(1 - X_i^\alpha) \simeq \log \alpha < 0$. It follows that

$$
T_1(\alpha) \simeq -\frac{1}{\alpha} + \frac{1}{2n} \sum_{i=1}^n \log X_i, \quad T_2(\alpha) \simeq -\frac{1}{\alpha} \quad \text{and} \quad T_3(\alpha) \simeq \log \alpha
$$

so that

$$
S(\alpha) \simeq \frac{-1}{\alpha \log \alpha} \to \infty.
$$

(II) $\alpha \to \infty$. In this case,

$$
T_1(\alpha) \simeq n^{-1} \sum_{i=1}^n \log X_i, \quad T_2(\alpha) \simeq n^{-1} \sum_{i=1}^n X_i^\alpha \log X_i, \quad \text{and} \quad T_3(\alpha) \simeq -n^{-1} \sum_{i=1}^n X_i^\alpha.
$$

It follows that

$$
\begin{aligned}
S(\alpha) &\simeq -\frac{\sum_{i=1}^n (-\log X_i)}{n} + \frac{\sum_{i=1}^n X_i^\alpha (-\log X_i)}{\sum_{i=1}^n X_i^\alpha} \\
&\simeq -\frac{\sum_{i=1}^n (-\log X_i)}{n} + (-\log X_{n:n}) < 0
\end{aligned}
$$

because the mean of positive quantities is larger than their minimum.

REFERENCES

Arnold, B.C., Balakrishnan. N., and Nagaraja, H.N. (1992), *A First Course in Order Statistics*, New York: Wiley.

Arnold, B.C., and Groeneveld, R.A. (1995), "Measuring Skewness With Respect to the Mode," *The American Statistician*, 49, 34–38.

Bowley, A.L. (1937), *Elements of Statistics*, Sixth Edition, New York: Scribner.

Devroye, L. (1986), *Non-Uniform Random Variate Generation.* New York: Springer. Available at `http://cg.scs.carleton.ca/~luc/rnbook index.html`

Fang, K.T., Kotz, S., and Ng, K.W. (1990), *Symmetric Multivariate and Related Distributions*, London: Chapman and Hall.

Gilchrist, W.G. (2001), *Statistical Modelling With Quantile Functions*, Boca Raton, LA: Chapman & Hall/CRC.

Gould, H.W. (1972), *Combinatorial Identities.* Morgantown, VA: Morgantown Printing and Binding Co.

Gupta, R.D., and Kundu, D. (1999), "Generalized Exponential Distributions," *Australian and New Zealand Journal of Statistics*, 41, 173–188.

Gupta, R.D., and Kundu, D. (2007), "Generalized Exponential Distribution: Existing Results and Some Recent Developments," *Journal of Statistical Planning and Inference*, 137, 3537–3547.

Hosking, J.R.M. (1989), "The Theory of Probability Weighted Moments," Research report RC 12210 (Revised Version), IBM Research Division, Yorktown Heights, NY.

Hosking, J.R.M. (1990), "L-Moments: Analysis and Estimation of Distributions Using Linear Combinations of Order Statistics," *Journal of the Royal Statistical Society, Series B*, 52, 105–124.

Hosking, J.R.M. (1994), "The Four-Parameter Kappa Distribution," *IBM Journal of Research and Development*, 38, 251–258.

Hosking, J.R.M., and J.R. Wallis (1997), *Regional Frequency Analysis; an Approach Based on L-Moments*, Cambridge: Cambridge University Press.

Johnson, N.L., Kotz, S., and Balakrishnan, N. (1994), *Continuous Univariate Distributions*, Volume 2, Second Edition, New York: Wiley.

Jones, M.C. (2002), "The Complementary Beta Distribution," *Journal of Statistical Planning and Inference*, 104, 329–337.

Jones, M.C. (2004), "Families of Distributions Arising From Distributions of Order Statistics," (With Discussion) *Test*, 13, 1–43.

Jones, M.C. (2007), "Connecting Distributions With Power Tails on the Real Line, the Half Line and the Interval," *International Statistical Review*, 75, 58–69.

Karvanen, J., and Nuutinen, A. (2007), "Characterizing the Generalized Lambda Distribution by L-Moments," *Computational Statistics and Data Analysis*, to appear.

Kotz, S., and van Dorp, J.R. (2004), *Beyond Beta; Other Continuous Families of Distributions With Bounded Support and Applications*, New Jersey: World Scientific.

Mäkeläinen, T., Schmidt, K., and Styan, G.P.H. (1981), "On the Existence and Uniqueness of the Maximum Likelihood Estimate of a Vector-Valued Parameter in Fixed-Size Samples," *Annals of Statistics*, 9, 758–767.

McDonald, J.B. (1984), "Some Generalized Functions for the Size Distribution of Income," *Econometrica*, 52, 647–664.

Nadarajah, S., and Gupta, A.K. (2004), "Generalizations and Related Univariate Distributions," in *Handbook of Beta Distribution and Its Applications*, eds. A.K. Gupta, and S. Nadarajah, New York: Dekker, pp. 97–163.

Parzen, E. (1979), "Nonparametric Statistical Data Modelling," (With Comments), *Journal of the American Statistical Association*, 74, 105–131.

Ramberg, J.S., and Schmeiser, B.W. (1974), "An Approximate Method for Generating Asymmetric Random Variables," *Communications of the Association for Computing Machinery*, 17, 78–82

Tadikamalla, P.R., and Johnson, N.L. (1982), "Systems of Frequency Curves Generated by Transformations of Logistic Variables," *Biometrika*, 69, 461–465.

Ulrich, G. (1984), "Computer Generation of Distributions on the M-Sphere," *Applied Statistics*, 33, 158–163.

van Zwet, W.R. (1964), *Convex Transformations of Random Variables*, Amsterdam: Mathematisch Centrum.

Figure 1. Some minimax densities. $\alpha = 5, \beta = 2$: left-skewed unimodal density; $\alpha = 2, \beta = 2.5$: almost symmetric unimodal density; $\alpha = 1/2, \beta = 1/2$: uniantimodal density; $\alpha = 1, \beta = 3$: decreasing density.

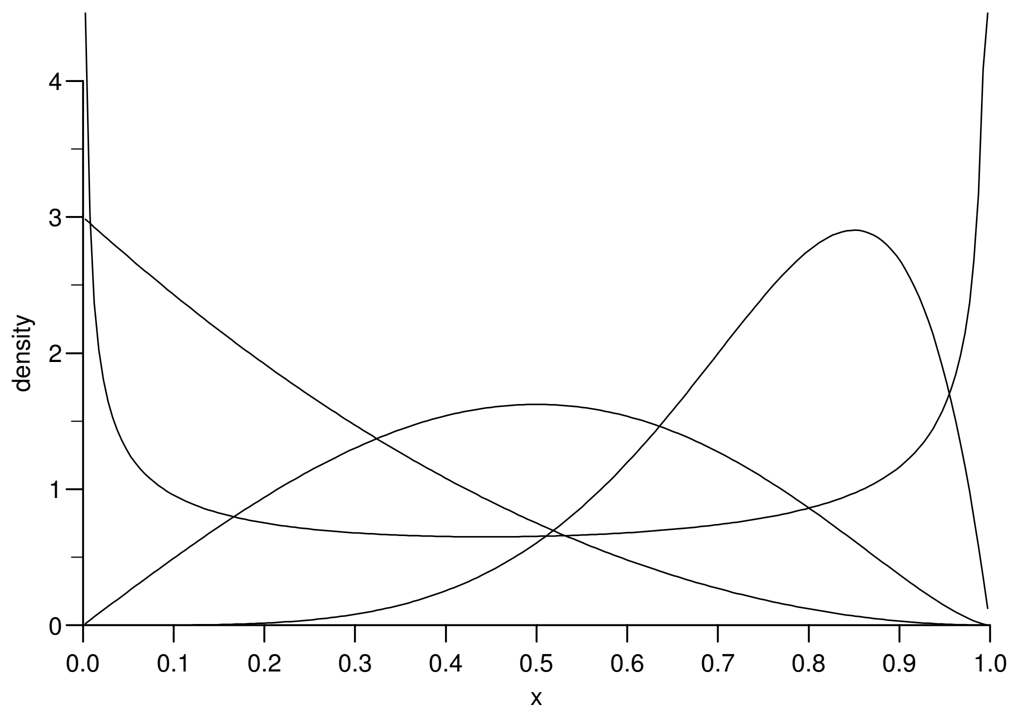

Figure 2. L-skewness for the minimax distribution, plotted as a function of $\log(\alpha)$ and $\log(\beta)$. The straight lines superimposed on the plot correspond to $\alpha = 1$ and $\beta = 1$.

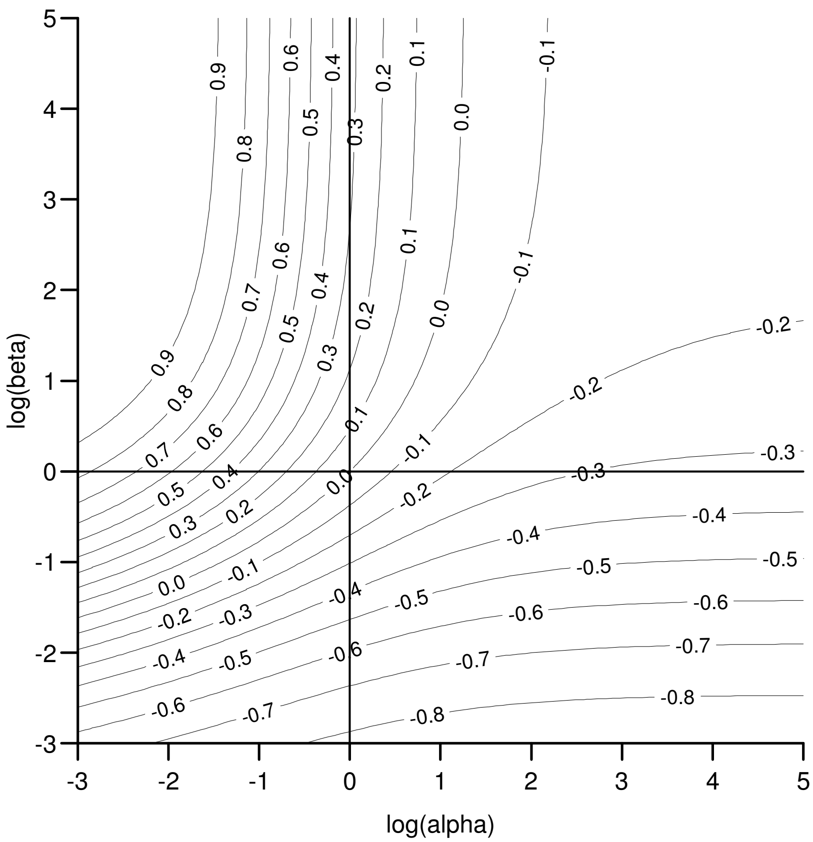

Figure 3. Some almost-symmetric minimax densities. In order of decreasing height of mode/antimode: (i) $\alpha = 3.211, \beta = 100$; (ii) $\alpha = 2.470, \beta = 5$; (iii) $\alpha = 1.707, \beta = 2$; (iv) $\alpha = \beta = 1$ (the symmetric uniform density); (v) $\alpha = 2/5, \beta = 1/2$; (vi) $\alpha = 0.082, \beta = 1/4$.

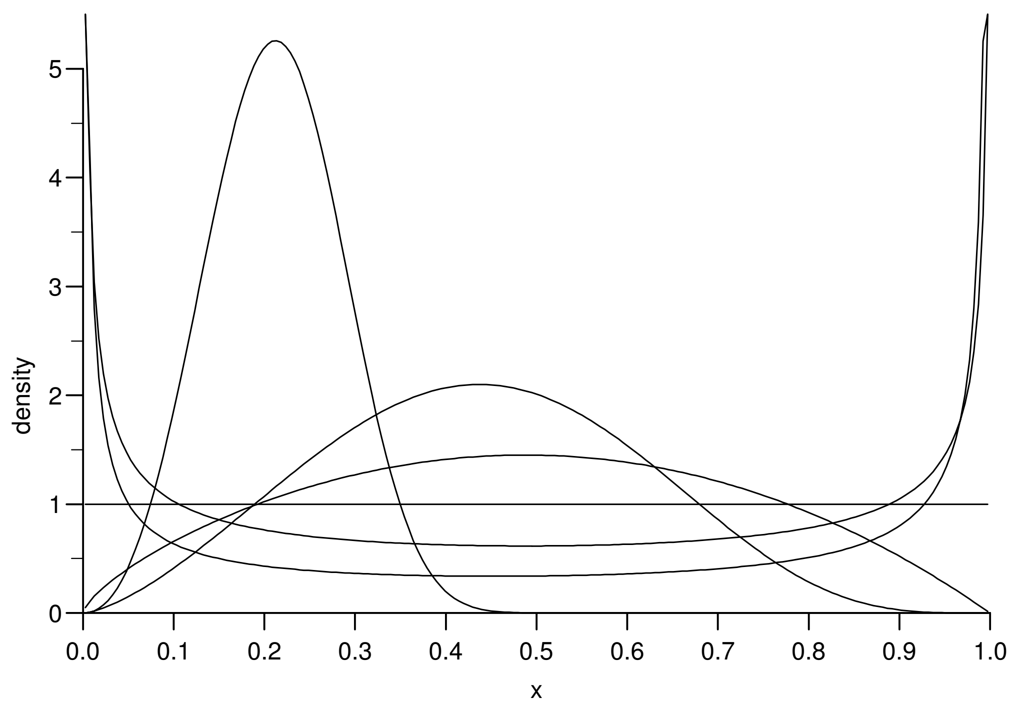

Figure 4. L-kurtosis for the minimax distribution, plotted as a function of log(α) and log(β). The straight lines superimposed on the plot correspond to $\alpha = 1$ and $\beta = 1$.