Creating new distributions by blunting cusps

Rose Baker
School of Business
University of Salford, UK

Keywords

cusp, derivatives, Laplace distribution, Pareto distribution

Creating new distributions by blunting cusps

Rose Baker
School of Business
University of Salford, UK

Abstract

A simple method is proposed for 'blunting' cusped distributions, i.e. removing the cusp. No additional parameters are required. The method is applied to the asymmetric Laplace distribution, to the van Dorp and Kotz double-sided power distribution, and to the double-sided asymmetric Pareto distribution, and some properties of the blunted distributions are derived. An example of fitting the blunted asymmetric Pareto distribution to data is given.

1. Introduction

Some probability distributions currently used in modelling have a cusp, a point where the derivative cannot be defined. In plainer English, a cusp is a sharp peak. The best-known example of this is the Laplace distribution. This generalizes to an asymmetric form, shown in Figure 1. Another example is the two-sided power distribution (Kotz and van Dorp, 2002) which is a bounded distribution that can be used as an alternative to the beta distribution. Yet another example is the Subottin (exponential power) distribution, a distribution that can be leptokurtic or platykurtic.

Distributions with cusps exemplify George Box's famous comment that 'essentially, all models are wrong, but some are useful' (Box and Draper, 1987). We surely do not believe that, if we had sufficient data, we would see a cusp becoming more and more clearly defined. Rather, we probably chose a distribution because it had the right tail or general behaviour, and the cusp was an unwanted but (just) ignorable artefact.

On the other hand, cusps can occasionally arise; this could happen if we looked at the difference of two exponentially-distributed random variables, which would follow the Laplace distribution. In practice, cusps seem to cause few problems for statistical inference (e.g. Hinkley and Revankar, 1977) and how undesirable they are is probably a matter of taste.

Preprint submitted to Elsevier *January 4, 2017*

Here, we describe and illustrate a method for blunting cusps (a term drawn from dentistry!) to produce distributions that behave more smoothly, for those who find cusps undesirable. This produces some flexible distributions, and an example of fitting to data is given.

Another method of cusp removal would be to add a random variable defined over $(-\epsilon, \epsilon)$, e.g. a uniform or other kernel distribution. This removes the cusp with minimal change to the distribution otherwise. This is the 'mollification' proposed by Friedrichs and investigated empirically by Rubio (web page given in bibliography). This method adds an extra parameter, and of course makes the pdf more complex as it is now a convolution.

Jones (2010) and Jones (2014) describe transformations of probability, and here again a distribution can be smoothed at the cost of introducing an additional parameter.

The blunting procedure proposed here was already used to remove cusps by Baker (2008) for a statistical curiosity, the Schlömilch-transformed exponential, with pdf $f(x) = c \exp(-|cx - b/cx|)$ for $x > 0$. The distribution is unimodal and has a cusp at the mode. It was modified into a smooth survival distribution with pdf $g(x) = (c/2)(1 + |cx - b/cx|) \exp(-|cx - b/cx|)$. Here we focus on more practical distributions.

Usually the parent distributions are 2-piece distributions, but not always, as the example just given shows. It should be pointed out that many 2-piece distributions are already differentiable at least once at the mode. Rubio and Steel (2015) discuss 2-piece distributions, focussing on skewness and kurtosis. They allow scale and shape parameters to vary in each half of the distribution, which requires 5 parameters (one location parameter, 2 scale and 2 shape parameters). These distributions are very useful in modelling skew and long-tailed data, and the present work allows more distributions to be given this treatment. We use asymmetric distributions and allow scale parameters to differ in each half, because this is often the reality. For simplicity, we do not always allow shape parameters to differ, although this is possible.

2. Methodology

2.1. Derivation of the pdf of the blunted distribution

Consider a pdf $f(x)$ with a cusp at m. We refer to cusped distributions as the 'parent' distribution, and denote pdf and distribution function as $f(x)$ and $F(x)$ respectively, using $g(x)$ and $G(x)$ for the 'daughter' or blunted distribution. It is helpful to decompose f into partial pdfs $f_+(x)$ and $f_-(x)$ defined on $x \geq m$ and $x < m$ respectively, so that $f_+(x) =$

$f(x)I(x \geq m), f_-(x) = f(x)I(x < m)$ where I is the indicator function, 1 if its argument is true, else zero. Then $f(x) = f_+(x) + f_-(x)$.

As an example of a cusped distribution, the asymmetric Laplace distribution is defined as

$$f(x) = \begin{cases} \lambda/(\kappa + \kappa^{-1}) \exp\{(\lambda/\kappa)(x - m)\} & \text{if } x < m \\ \lambda/(\kappa + \kappa^{-1}) \exp\{-(\lambda\kappa)(x - m)\} & \text{if } x \geq m. \end{cases}$$

where $m \in \mathcal{R}, \lambda > 0, \kappa > 0$.

We now construct a pdf on either side of m such that $g(x)$ will be cusp-free. Denote the first derivatives when $x \to m$ from above and below as $f'_+(m) = \lim_{\xi\to 0}(f(m+\xi) - f(m))/\xi$, $f'_-(m) = \lim_{\xi\to 0}(f(m) - f(m-\xi))/\xi$ respectively where $\xi > 0$. Define second derivatives $f''_+(m)$ and $f''_-(m)$ in the same way, and assume that first and second derivatives are finite. Then for $x \geq m$, on expanding in a Taylor series we have $f(x) = f(m) + f'_+(m)(x - m) + (1/2)f''_+(m)(x-m)^2 + \cdots$. We consider the pdf

$$g(x) \propto (1 - f'_+(m)(x-m)/f(m))f(x) = f(m) + \{(1/2)f'' - (f')^2/f\}(m)(x-m)^2 + O((x-m)^3). \tag{1}$$

On the other side of the cusp where $x < m$, we construct $g(x)$ similarly by multiplying by $1 - f'_-(m)(x - m)$. Thus we have that

$$g(x) = k\{1 - (\ln f)'_{\mathrm{sgn}(x)}(m)(x - m)\}f(x), \tag{2}$$

where k is a normalizing constant, and $\mathrm{sgn}(x) = 1$ $(+)$ if $x \geq 0$, else $\mathrm{sgn}(x) = -1$ $(-)$.

From (1) or (2) we see that $g(x)$ has a continuous first derivative, which is zero at m, so the cusp has been removed. To illustrate the method with the asymmetric Laplace distribution, we have $f'_-(m)/f(m) = \lambda/\kappa$ for $x < m$ and $f'_+(m)/f(m) = -\lambda\kappa$ for $x \geq m$. After renormalizing we obtain

$$g(x) = \begin{cases} (1/2)\lambda/(\kappa + \kappa^{-1})(1 - (\lambda/\kappa)(x - m)) \exp\{(\lambda/\kappa)(x - m)\} & \text{if } x < m \\ (1/2)\lambda/(\kappa + \kappa^{-1})(1 + (\lambda\kappa)(x - m)) \exp\{-(\lambda\kappa)(x - m)\} & \text{if } x \geq m. \end{cases}$$

This is shown in figure 1 along with the original distribution. The change in curvature at $x = m = 0$ can just be seen. In each half, $g(x)$ is an equal mixture of an exponential distribution with a gamma (Erlang) distribution. This is similar but not identical to the Lindley distribution, which has pdf $\propto (1 + x) \exp(-\delta x)$ with $\delta > 0$.

2.2. General properties of the blunted distributions

We now turn to the general properties of the blunted distributions, the pdf, moments, distribution function, quantiles and random number generation. For a unimodal cusped distribution $(\ln f)'_{\mathrm{sgn}(x)}(m)(x-m) < 0$ so $g(x) > 0$ always, but for an 'anticusp' where f has a minimum at m, $(\ln f)'_{\mathrm{sgn}(x)}(m)(x-m) > 0$ so $g(x)$ would become negative at large $|x|$. U-shaped distributions are bounded, but one needs to check that the constructed pdf g never becomes negative. Later, an example of the double-sided power distribution is given, where for some parameter values the pdf is indeed U-shaped.

Defining

$$\mu_+^{(n)} = \int_m^\infty (x-m)^n f(x)\,\mathrm{d}x,$$

$$\mu_-^{(n)} = \int_{-\infty}^m (x-m)^n f(x)\,\mathrm{d}x,$$

and writing $\mu_+^{(1)} = \mu_+$ and $\mu_-^{(1)} = \mu_-$, we obtain

$$g(x) = \frac{\{1-(\ln f)'_{\mathrm{sgn}(x)}(m)(x-m)\}f(x)}{1-(\ln f)'_+(m)\mu_+ - (\ln f)'_-(m)\mu_-}. \tag{3}$$

We see that $g(m) = kf(m)$ so that $g(x)$ is continuous at $x = m$, and that g has zero first derivative and so no cusp. If the parent distribution is symmetric about the mode, the second derivative is continuous, but otherwise the second derivative (curvature) will be discontinuous at $x = m$; this is only a minor problem and commonly occurs with 2-piece distributions. From (3) the quantities μ_+ and μ_- must exist for g to be defined. This means that the parent distribution must have a defined mean for the daughter distribution to have a valid pdf.

Note that we also assume finite derivatives at m. The derivative is infinite at zero for the Subottin distribution with its parameter $\alpha < 1$ (Subottin, 1923, Johnson *et al* 1995), so this method cannot be used for it. The Laplace distribution is a special case of the Subottin with $\alpha = 1$, where the derivative is finite and so this method does apply to it.

The moments about m are

$$\mathrm{E}(X-m)^n = \frac{\mu_+^{(n)} + \mu_-^{(n)} - (\ln f)'_+(m)\mu_+^{(n+1)} - (\ln f)'_-(m)\mu_-^{(n+1)}}{1-(\ln f)'_+(m)\mu_+ - (\ln f)'_-(m)\mu_-}. \tag{4}$$

so that the nth moment of g exists if the $n+1$th moment of f does. In particular

$$\mathrm{E}(X) = m + \frac{\mu_+ + \mu_- - (\ln f)'_+(m)\mu_+^{(2)} - (\ln f)'_-(m)\mu_-^{(2)}}{1 - (\ln f)'_+(m)\mu_+ - (\ln f)'_-(m)\mu_-}.$$

The central moments (i.e. about $\mathrm{E}(X)$) can be found from (4).

For distributions with infinite support, in the tail $g(x) \simeq -k(\ln f)'_{\mathrm{sgn}(x)}(m)xf(x)$. The tail behaviour will be similar to the parent distribution if this is not fat-tailed, for example has an exponential tail. If the parent distribution is fat-tailed (behaves like a power of x in the tail), for example the double Pareto distribution, the daughter distribution will have the same kind of tail behaviour as a distribution from the parent's family, with the 'tail index' reduced by 1. Hence tail behaviour can usually be made similar to the parent distribution by changing parameter values, despite the extra factor of $|x|$.

The distribution function is obtained by integrating (3). In particular

$$G(m) = \frac{F(m) - (\ln f)'_-(m)\mu_-}{1 - (\ln f)'_+(m)\mu_+ - (\ln f)'_-(m)\mu_-}.$$

Since for the examples given the distribution functions can always be easily computed, quantiles can be found by Newton-Raphson iteration. Given a trial value x_n the next iteration is

$$x_{n+1} = x_n - (G(x_n) - P)/g(x_n),$$

where P is the required quantile. The value of $G(m)$ at the erstwhile cusp is known, so it is known in which half of the distribution the quantile lies. One has to ensure that the iteration does not wander into the wrong half of the distribution. This can be achieved by checking whether x_{n+1} would be in the other half of the distribution, and, if so, setting x_{n+1} to $m + 1/2(x_n - m)$ for example.

We now give an interpretation of (3) that is relevant to random-number generation. Use the partial pdfs $f_+(x)$ and $f_-(x)$ and write $h_+(x)$ for the pdf of the length-biased distribution based on f for $x \geq m$ and similarly for $x < m$. We have

$$h_+(x) = (x-m)f_+(x)/\mu_+, h_-(x) = (x-m)f_-(x)/\mu_-,$$

so that $\int_m^\infty h_+(x)\,\mathrm{d}x = 1$, and $\int_{-\infty}^m h_-(x)\,\mathrm{d}x = 1$, Then the pdf of X is

$$g(x) = \begin{cases} k\{f_-(x) - (\ln f)'_-(m)\mu_- h_-(x)\} & \text{if } x < m \\ k\{f_+(x) - (\ln f)'_+(m)\mu_+ h_+(x)\} & \text{if } x \geq m \end{cases}$$

so that the daughter pdf is a positive mixture of the parent and length-weighted parent distributions. If random numbers can be generated from the parent and the length-weighted parent, then one computes $G(m)$ and with probability $G(m)$ takes $X < m$, else $X \geq m$. Then one generates the random number from the parent distribution with probability $F(m)/(F(m) - (\ln f)'_{-}(m)\mu_{-})$ if $X < m$, else from the length-biased distribution. If $X \geq m$ one generates from the parent distribution with probability $(1-F(m))/(1-F(m) - (\ln f)'_{+}(m)\mu_{+})$. In practice the length-biased pdf h may not correspond exactly to a standard distribution for which random number generators are available.

Other methods are the inversion method based on $G(x)$ or the rejection method. The rejection method can be used by generating candidate random numbers from any suitable distribution, but the parent cusped distribution cannot be used unless the support of the distribution is bounded, because $g(x)/f(x) \propto |x|$ in the tails. In this case, the analysis is simple only when the cusp is an 'anticusp', i.e. a minimum rather than a maximum. In this case $(\ln f)'_{\mathrm{sgn}(x)}(m)(x-m) > 0$, so $g(x)/k \leq f(x)$. Retaining random numbers from $f(x)$ with probability $g(x)/kf(x)$ then gives a method with efficiency $1/k = f(m)/g(m)$.

3. Application to the major cusped distributions

We now briefly describe the properties of the blunted distributions for the several 'parent' cusped distributions that we know of. The many detailed algorithms have been computer-checked, by generating random numbers by all the methods given in the text, and checking that the resulting moments and quantiles agree with those derived analytically, for large sample sizes and a range of parameter values.

3.1. Asymmetric Laplace distribution

The pdf has already been given. The distribution function is

$$G(x) = \begin{cases} \kappa^2(1+\kappa^2)^{-1}(1-(\lambda/2\kappa)x)\exp\{(\lambda/\kappa)(x-m)\} & \text{if } x < m \\ 1-(1+\kappa^2)^{-1}(1+(\lambda\kappa)x/2)\exp\{-(\lambda\kappa)(x-m)\} & \text{if } x \geq m \end{cases}$$

where $\kappa > 0, \lambda > 0$. Random numbers could be generated as follows, where U, U_1, U_2 are uniformly distributed random variables:

1. Choose $X \geq m$ with probability $1/(1+\kappa^2)$, else $X < m$.
2. If $X \geq m$, with probability $1/2$ let X be exponential (Erlang (1)) or Erlang (2).

3. If exponential, form $X = m - \ln(U)/\lambda\kappa$, else $X = m - \ln(U_1 U_2)/\lambda\kappa$.
4. similarly if $X < m$, form $X = m + \kappa \ln(U)/\lambda$, else $X = m + \kappa \ln(U_1 U_2)/\lambda$.

Here we have exploited the facts the blunted distribution is an equal mixture of exponential and Erlang (2) distributions, and that the Erlang (2) random variable is the sum of two exponential r.v.s. This method requires on average 3.5 uniform random numbers.

The moments all exist and can be simply derived. In general,

$$\mathrm{E}((X-m)^n) = \frac{(n+2)n!\lambda^{-n}(\kappa^{-(n+1)} - (-\kappa)^{n+1})}{2(\kappa+\kappa^{-1})},$$

from which

$$\mathrm{E}(X) = m + (3/2)\frac{(1-\kappa^2)}{\lambda\kappa},$$

$$\mathrm{var}(X) = \lambda^{-2}\frac{7\kappa^3 + 9\kappa + 9\kappa^{-1} + 7\kappa^{-3}}{4(\kappa+\kappa^{-1})}.$$

If we increase λ for the daughter distribution so that the variances of the original and blunted distributions are the same, figure 1 becomes figure 2, giving a better impression of the change in the original distribution.

3.2. Two-sided power distribution

The two-sided power distribution (Van Dorp and Kotz, 2002) defined on $[0,1]$ is one of the alternatives to the beta distribution. It is in fact slightly more tractable mathematically than the beta distribution, because the distribution function can be derived analytically, and random numbers are easier to generate. The pdf is

$$f(x) = \begin{cases} b(x/a)^{b-1} & \text{if } x < a \\ b(\{(1-x)/(1-a)\}^{b-1} & \text{if } x \geq a \end{cases}$$

where $0 < a < 1$ and $b > 0$. For $b < 1$ the pdf is U-shaped. We obtain

$$g(x) = \begin{cases} (1/2)(b+1)\{b(x/a)^{b-1} - (b-1)(x/a)^b\} & \text{if } x < a \\ (1/2)(b+1)\{b\{(1-x)/(1-a)\}^{b-1} - (b-1)\{(1-x)/(1-a)\}^b\} & \text{if } x \geq a \end{cases}$$

Figure 3 shows the original and blunted distributions.

The distribution function $G(x)$ is:

$$G(x) = \begin{cases} (a/2)\{(b+1)(x/a)^b - (b-1)(x/a)^{b+1}\} & \text{if } x < a \\ 1 - (1/2)(1-a)\{(b+1)\{(1-x)/(1-a)\}^b - (b-1)\{(1-x)/(1-a)\}^{b+1}\} & \text{if } x \geq a. \end{cases}$$

For $b > 1$ each part of the pdf can be rewritten as a mixture of (scaled) beta pdfs with parameters (α, β) given by $(b, 1)$ and $(b, 2)$ with weights $(b+1)/2b$ and $(b-1)/2b$ respectively. Thus for $b > 1$ random numbers can be generated in this way by the mixture method. Many platforms provide beta random variables, e.g. R and the NAG library.

One could also generate U and then find X using Newton-Raphson iteration, since G is easily computed. However, for low b it is quicker to use a rejection method, generating random numbers X from the cusped distribution and keeping (not rejecting) them with probability $g(x)/f(x) \times \frac{2\min(b,1)}{b+1}$. Kotz and van Dorp (2004) show how to compute random numbers X from their distribution.

The rejection method gives probability P of acceptance:

$$P = \begin{cases} \min(b,1)\{1 - \frac{b-1}{b}\frac{X}{a}\} & \text{if } U < a \\ \min(b,1)\{1 - \frac{b-1}{b}\frac{1-X}{1-a}\} & \text{if } U \geq a. \end{cases}$$

This method has efficiency $2\min(b,1)/(b+1)$.

The moments are:

$$\mathrm{E}(X) = \frac{(b-1)a + 3/2}{b+2},$$

$$\mathrm{var}(X) = \frac{4b^2 + 7b + 1 - 4(2b+3)(b-1)a(1-a)}{2(b+1)(b+2)^2(b+3)}.$$

3.3. Double-sided asymmetric Pareto distribution

This is the doubled type II Pareto distribution which generalizes the asymmetric Laplace distribution, and is shown in figure 4. It has 4 parameters, scale parameters $a, b > 0$, mode $m \in \mathcal{R}$ and tail power parameter $\beta > 0$. The pdf of the asymmetric double-sided Pareto distribution is

$$f(x) = \begin{cases} \frac{\beta}{a+b}(1 + |x-m|/a)^{-\beta-1} & \text{if } x < m \\ \frac{\beta}{a+b}(1 + |x-m|/b)^{-\beta-1} & \text{if } x \geq m. \end{cases}$$

The blunted distribution has pdf

$$g(x) = \begin{cases} \frac{\beta-1}{2(a+b)}\{(\beta+1)(1 + |x-m|/a)^{-\beta} - \beta(1 + |x-m|/a)^{-\beta-1}\} & \text{if } x < m \\ \frac{\beta-1}{2(a+b)}\{(\beta+1)(1 + |x-m|/b)^{-\beta} - \beta(1 + |x-m|/b)^{-\beta-1}\} & \text{if } x \geq m. \end{cases}$$

The distribution function is

$$G(x) = \begin{cases} \frac{a}{2(a+b)}\{(\beta+1)(1 + |x-m|/a)^{1-\beta} - (\beta-1)(1 + |x-m|/a)^{-\beta}\} & \text{if } x < m \\ 1 - \frac{b}{2(a+b)}\{(\beta+1)(1 + |x-m|/b)^{1-\beta} - (\beta-1)(1 + |x-m|/b)^{-\beta}\} & \text{if } x \geq m. \end{cases}$$

The distribution is a negative mixture of the asymmetric Pareto distribution with parameter $\beta - 1$ with weight $(\beta + 1)/2$, and the asymmetric Pareto distribution with parameter β, weight $(1 - \beta)/2$. The parent distribution requires $\beta > 0$, with the mean defined if $\beta > 1$ and the variance if $\beta > 2$; the blunted distribution requires $\beta > 1$ for the pdf to exist, $\beta > 2$ for the mean to exist, and $\beta > 3$ for the variance to exist.

Smoothing away the peak does not reduce the ability to model long tails, because in the tail the behaviour is now $\propto x^{-\beta}$ for $\beta > 1$, whereas before it was $\propto x^{-\beta-1}$ for $\beta > 0$.

Random numbers can be generated using the rejection method using a lighter-tailed distribution from the parent family. We generate random numbers from the double-sided Pareto distribution of pdf $h(x)$ with β replaced by $\beta - 1$. The random numbers from h are

$$X = \begin{cases} m + a\{1 - ((a+b)U/a)^{1/(1-\beta)}\} & \text{if } U < a/(a+b) \\ m + b\{(a+b)(1-U)/b)^{1/(1-\beta)} - 1\} & \text{if } U \geq a/(a+b). \end{cases}$$

The ratio $g(x)/h(x)$ is then $g(x)/h(x) = (1/2)\{(\beta+1) - \beta/(1+|x-m|/a)\}$ for $x < m$, from which we see that accepting random numbers from pdf h with probability $1 - \frac{\beta}{(\beta+1)(1+|x-m|/a)}$ for $x < m$ (and with b replacing a for $x \geq m$) gives random numbers from pdf g with efficiency $2/(\beta + 1)$, i.e. we need on average $(\beta + 1)/2$ random numbers from h.

For large β this method is inefficient. The Newton-Raphson iteration for the inversion method was faster if $\beta \geq 5$. Iteration to an accuracy of 10^{-8} takes just over 5 steps on average.

A third method to generate random numbers is to recognise that the two parts of the pdf $g(x)$ can be written as mixtures of beta prime (type II beta distribution) random variables. The algebra for this is simple but laborious. A random number Z from the beta prime distribution $\text{Bep}(\gamma, \delta)$ is most easily obtained from a $\text{beta}(\gamma, \delta)$ r.v. Y as $Z = Y/(1 - Y)$. Then the procedure is:

1. Choose $X < m$ with probability $a/(a + b)$, else $X > m$;
2. with probability $(\beta - 1)/2\beta$ generate Z from $\text{Bep}(1, \beta)$, else generate Z from $\text{Bep}(2, \beta - 1)$;
3. if $X < m$ form $X = m - aZ$, else form $X = m + bZ$.

Moments are readily calculated, and are

$$\text{E}(X) = \frac{3(b-a)}{2(\beta-2)}$$

for $\beta > 2$,

$$\mathrm{var}(X) = \frac{(\beta+1)\{(\beta-2)(a^2+b^2)+(a-b)^2\}}{2(\beta-2)^2(\beta-3)} - \frac{(\beta-1)(a^2+b^2)-(a-b)^2}{2(\beta-1)(\beta-2)},$$

for $\beta > 3$. It is possible to allow the power β to be different in each half of the distribution, giving the type of 5-parameter distribution described by Rubio and Steel (2015). Take the power as β for $x < m$ and γ for $x \geq m$. We then have as parent distribution:

$$f(x) = \begin{cases} \frac{\beta\gamma}{\gamma a+\beta b}(1+|x-m|/a)^{-\beta-1} & \text{if } x < m \\ \frac{\beta\gamma}{\gamma a+\beta b}(1+|x-m|/b)^{-\gamma-1} & \text{if } x \geq m. \end{cases}$$

The blunted pdf $g(x)$ is:

$$g(x) = \begin{cases} \frac{(\beta-1)(\gamma-1)}{(\gamma-1)a+(\beta-1)b}\{(\beta+1)(1+|x-m|/a)^{-\beta} - \beta(1+|x-m|/a)^{-\beta-1} & \text{if } x < m \\ \frac{(\beta-1)(\gamma-1)}{(\gamma-1)a+(\beta-1)b}\{(\gamma+1)(1+|x-m|/b)^{-\gamma} - \gamma(1+|x-m|/b)^{-\gamma-1} & \text{if } x \geq m. \end{cases}$$

4. Examples

To illustrate the use of these blunted distributions, the 2-sided Pareto distribution and the corresponding blunted distribution were fitted to daily (logarithmic) returns on the Nikkei and FTSE stock market indices. The AST distribution (e.g. Zhu and Galbraith, 2010, 2011) was used as a benchmark. This is the 4-parameter distribution in table 1, with different probability masses in the two halves, but with the same tail power. The 3-parameter distribution (the symmetric t-distribution, with equal scale and power parameters) and the full 5-parameter distribution of Rubio and Steel (2015) were also fitted.

From table 1 it is clear that for both the Nikkei and the FTSE returns, blunting the 2-sided Pareto distribution improved the fit, increasing the log-likelihood by 38 for the Nikkei and 44 for the FTSE-100. The blunted Pareto distribution fitted better than the AST benchmark distribution for the Nikkei but slightly worse for the FTSE-100. The point is not really that this model fits better for a particular dataset; we merely wish to demonstrate that it gives comparable fits to an established model.

Figure 5 shows the AST and blunted Pareto fits for the Nikkei data.

5. Conclusions

A simple method has been proposed for ‘blunting’ or de-cusping cusped distributions. Some of the new blunted distributions have been derived and their properties described. They could be called after their parent distribution, e.g. the blunted Laplace distribution. These models require no extra parameters, the downside being their greater complexity as mixtures of distributions, which entails more complicated expressions for the pdf, distribution function and moments.

Moments can be found analytically if they can for the parent distribution, and the distribution function can be found analytically if the parent distribution function can be integrated analytically. Random numbers can always be generated using the mixture method.

The surgery needed to remove the cusp has, as a referee noted, 'global repercussions' for a distribution, because it changes the tail behaviour. However, it is argued that by changing parameter values, one can often regain a distribution with tails not very different from the original.

The new distributions could be useful models, either because they are broadly similar to a useful cusped distribution, but without the cusp, or because they are simply flexible models in their own right. One of these blunted distributions in particular, the blunted 2-sided Pareto distribution, has been shown to give comparable fits to financial returns data to other models.

Further work along these lines is possible, either in developing new blunted distributions, or in gaining experience in fitting the ones described here to data. Random number generation is an area where very efficient methods can often be devised given ingenuity, and further work here would be useful if these distributions become widely used.

Acknowledgements

The author would like to thank two anonymous referees, whose comments have greatly improved the content and layout of the paper.

[1] Baker, R. (2008). Probabilistic Applications of the Schlömilch Transformation, Communications in Statistics, **37** (14), 2162-2176.

[2] Box, G. E. P., Draper, N. R. (1987), Empirical Model-Building and Response Surfaces, John Wiley and Sons, New York.

[3] Hinkley D.V. and Revankar N.S. (1977), Estimation of the Pareto law from underreported data: A further analysis. *Journal of Econometrics*, **5** (1):111.

[4] Johnson, N. L., Kotz, S. and Balakrishnan, N. (1995). Continuous Univariate Distributions, Wiley, New York.

[5] Jones M. C. (2010). Distributions generated by transformation of scale using an extended Cauchy-Schlömilch transformation, *Sankhyā*, **72** (2), 359-375.

[6] Jones, M. C. (2014). Generating distributions by transformation of scale. *Statistica Sinica*, **24** (2),749-771.

[7] Kotz, S. and van Dorp, J. R. (2004). Beyond Beta: Other Continuous Families of Distributions with Bounded Support and Applications, World Scientific, Singapore.

[8] Rubio, F. J. `https://sites.google.com/site/fjavierrubio67/resources/mollifiedtwopiecedistributions`

[9] Rubio, F. J. and Steel M. F. J. (2015). Bayesian modelling of skewness and kurtosis with two-piece scale and shape distributions, Electronic Journal of Statistics **9**, 1884-1912.

[10] Subbotin, M.T. (1923). On the law of frequency of error. Matemat. Sbornik 31, 296-301.

[11] Van Dorp, J R. and Kotz, S. (2002). The standard two-sided power distribution and its properties: with applications to financial engineering, American Statistician, **56**, 90-99.

[12] Zhu, D. and Galbraith, J. W. (2010). A generalized asymmetric Student-t distribution with application to financial econometrics, *Journal of Econometrics*, **157**, 297-305.

[13] Zhu, D. and Galbraith, J. W. (2011). Modeling and forecasting expected shortfall with the generalized asymmetric Student-t and asymmetric power distributions, *Journal of Empirical Finance*, **18**, 765-778.

Figures and tables

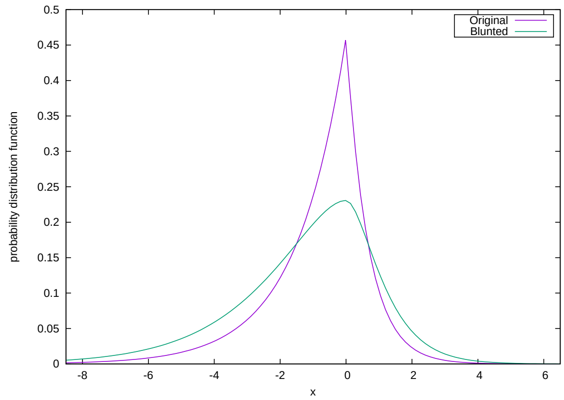

Figure 1: The asymmetric Laplace distribution ($\kappa = 1.5, m = 0, \lambda = 1$) and its 'blunted' analogue.

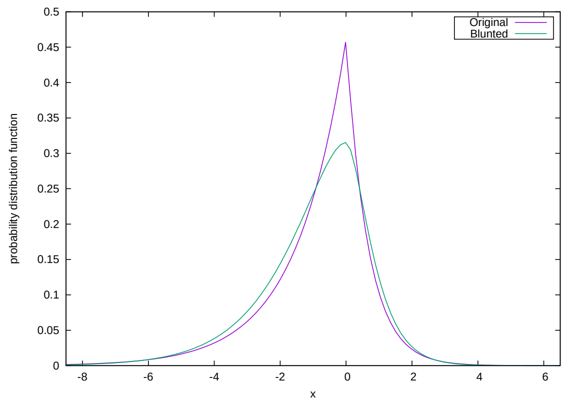

Figure 2: The asymmetric Laplace distribution ($\kappa = 1.5, m = 0, \lambda = 1$) and its 'blunted' analogue with λ increased to 1.367 to equalize variances.

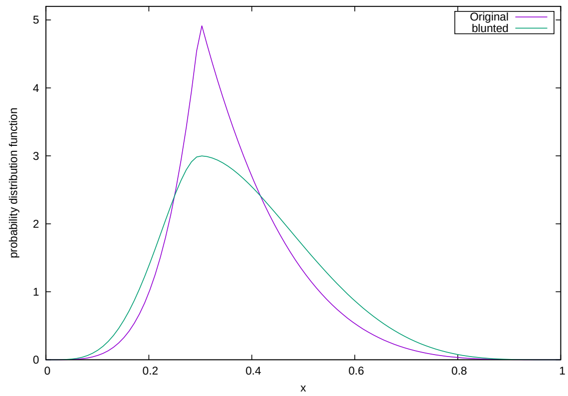

Figure 3: The two-sided power-law distribution ($a = 0.3, b = 5$) and its 'blunted' analogue.

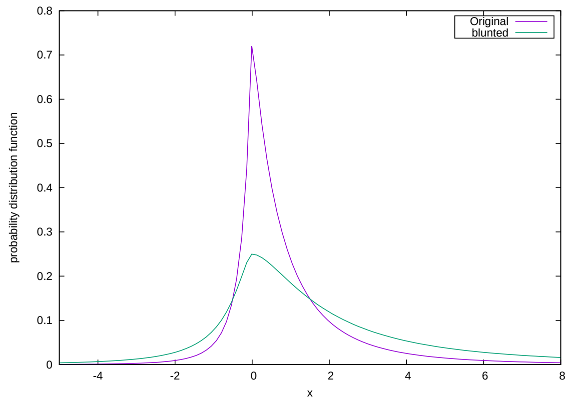

Figure 4: The two-sided asymmetric Pareto distribution ($a = 1, b = 3, \beta = 3, m = 0$) and its 'blunted' analogue.

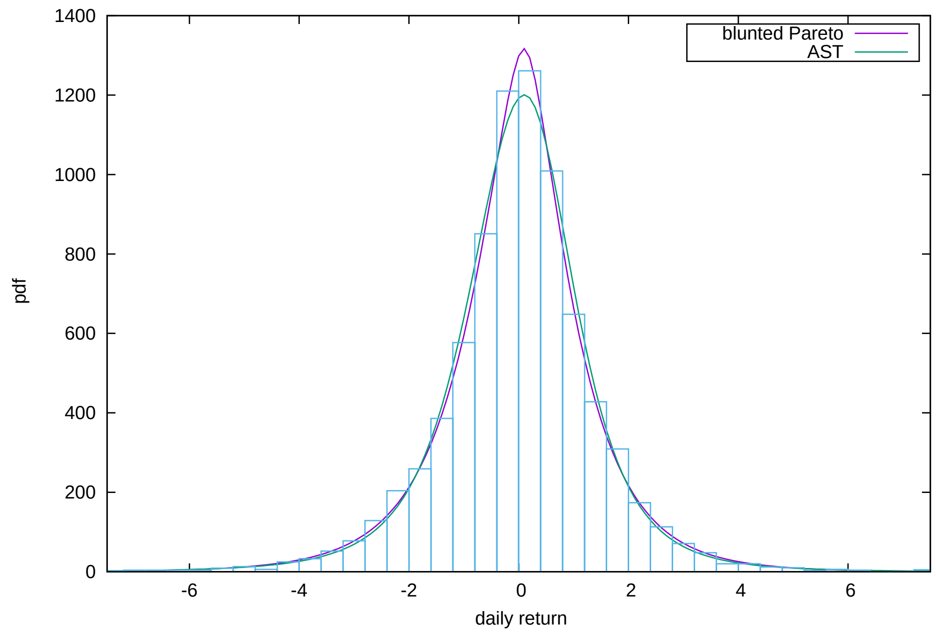

Figure 5: Daily returns on the Nikkei index with fitted AST and blunted 2-sided Pareto curves.

Dataset	Model	$-\ell$	AIC
Nikkei	AST(3)	13700.8	27407.7
Nikkei	AST(4)	13694.9	27397.9
Nikkei	AST(5)	13694.9	27399.8
Nikkei	Pareto	13707.8	27423.7
Nikkei	Blunted Pareto	13679.5	27366.9
FTSE	AST(3)	11951.3	23908.5
FTSE	AST(4)	11944.7	23897.5
FTSE	AST(5)	11944.4	23898.9
FTSE	Pareto	11999.2	24006.5
FTSE	Blunted Pareto	11955.7	23919.4

Table 1: Minus log-likelihood and AIC (Akaike Information Criterion) for fits of the AST model, the 2-sided Pareto and the blunted Pareto distributions to Nikkei and FTSE-100 data, to 5/7/2016. The AST(4) model has different probability mass in each tail, and the AST(5) model also has differing tail powers.