Effects of Parameters of Transformed Beta Distributions

Gary G Venter

Guy Carpenter Instrat

The transformed beta distribution was introduced to the insurance literature in Venter(1983) and independently to the economics literature in McDonald (1984). The parameterization discussed here was introduced by Rodney Kreps in order to make the parameters more independent of each other in the estimation process. The resulting parameters have somewhat separate roles in determining the shape of the distribution, and this note examines those effects.

Effects of Parameters of Transformed Beta Distributions

The transformed beta is considered parameterized so that $f(x) \propto (x/d)^{b-1}(1+(x/d)^c)^{-(a+b)/c}$. Each of the parameters will be considered in alphabetical order. In general terms, a determines the heaviness of the tail, b the shape of the distribution and the behavior near zero, c moves the middle around, and d is a scale parameter.

a

All positive moments $E(X^k)$ exist for $k < a$, but not otherwise. Thus a determines the heaviness of the tail. One way to measure tail heaviness is to look at the ratio of a high percentile to the median. For a large company, say with 50,000 expected claims, a pretty large claim would be one of the five largest – say the 1/10,000 probability claim.

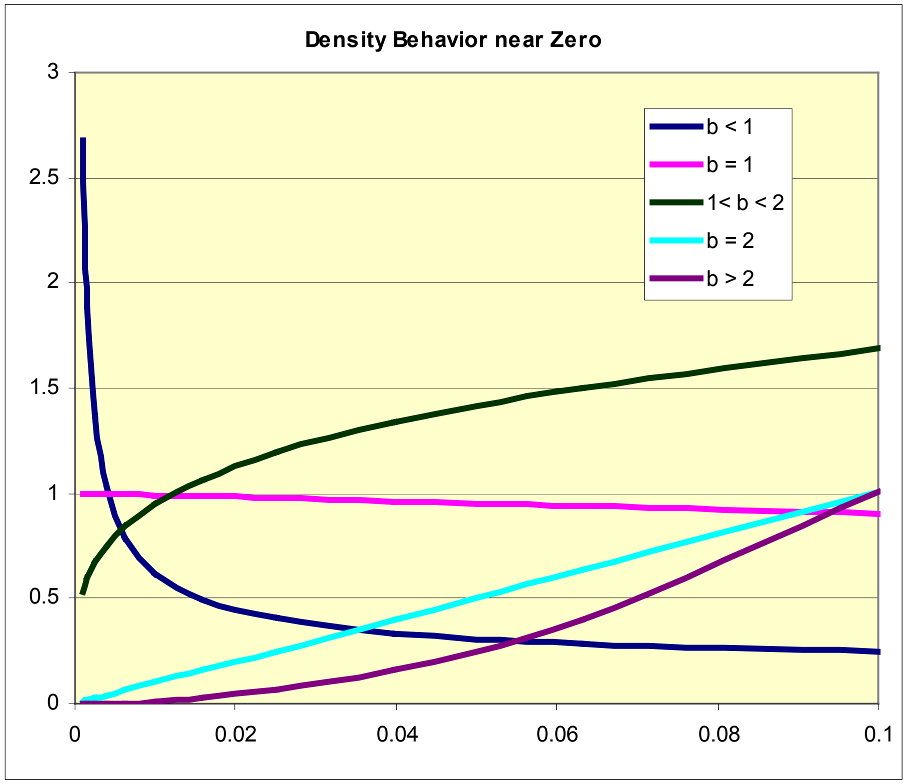
The ballasted Pareto $F(x) = 1 - (1+ x/d)^{-a}$ will be used to illustrate tail heaviness, as it is easy to deal with and has essentially the same tail heaviness as the general case. If B is a (big) number, the 1 - 1/Bth percentile is $d(B^{1/a} - 1)$. For the example with B=10,000, this is $d(10^{4/a} - 1)$. The ratio of this to the median (B=2) is thus $(10^{4/a} - 1)/(2^{1/a} - 1)$. This is very sensitive to a, especially for a between 1 and 2, where it often is. The ratio of pretty large claim to median in this case is 9254 at a=1.01 down to 788 for a=1.5. Thus the estimate of a could have a big impact on excess losses.

b

Negative moments $E((1/X)^k)$ exist for $k<b$ and not otherwise. This parameter governs the behavior of the distribution near 0. In that region, the density is close to constant$*(x/d)^{b-1}$so the derivative of the density is proportional to $(b-1)x^{b-2}$. This can be used to ascertain the shape of the density for smaller claims, which really determines the overall shape of the distribution.

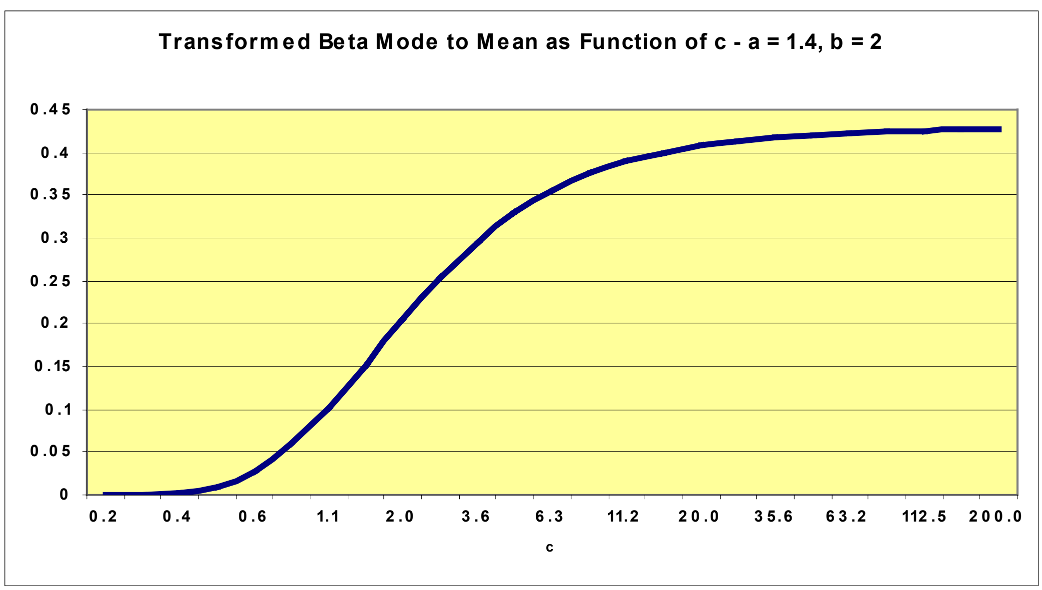
If b < 1, the slope of the density at zero is negative infinity, so the density is asymptotic to the vertical axis. For b=1, the other factor in the density becomes significant, and the slope is a negative number. The mode of the distribution is at zero in both of these cases. For 1 < b < 2, the slope at zero is positive infinity, so the density is rising and tangent to the vertical axis. For b=2, the slope is a positive number, and for b > 2, the slope is zero, so the density is tangent to the horizontal axis. For b > 1, then, there is a positive mode.

This graph show the behavior near zero and how that depends on b. Since the right tail is an inverse power curve, the behavior near zero determines the overall look of the distribution. The case b > 2 gives the usual shape of a density function people think of, which rises gradually then more steeply before falling off with the inverse power relationship. The transformed beta in this case looks like a heavy tailed lognormal. The case b = 1 is also seen a lot, for instance in the exponential and ballasted Pareto distributions.

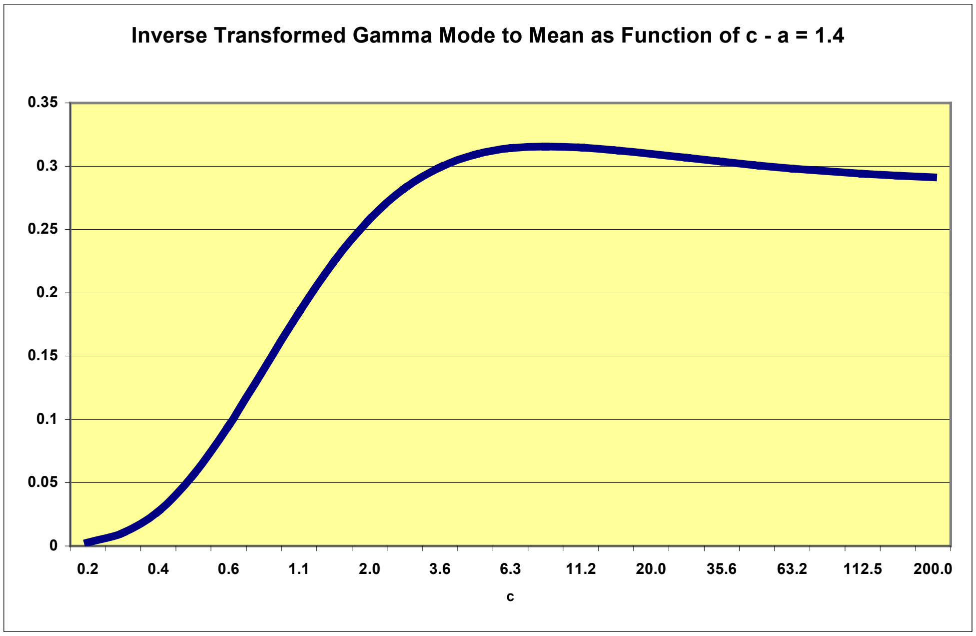


c

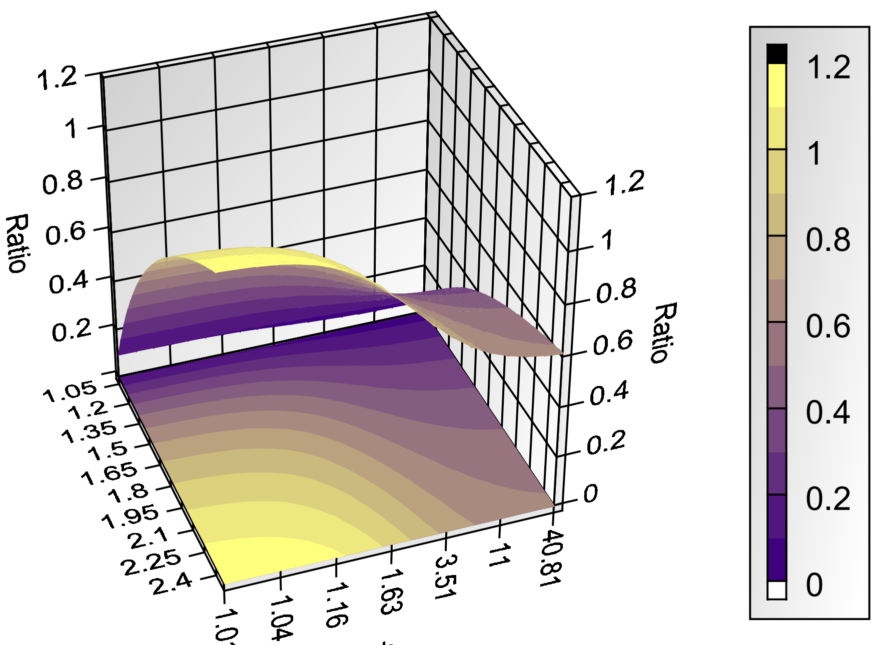
The c parameter introduces a power transform x –> x^c into the transformed beta. This tends to move the middle of the distribution around. A useful measure of where the middle is is the mode, as related to the mean. The ratio of mode to mean, when the mean exists (a > 1) and the mode is positive (b > 1), is for the most part an



increasing function of c for fixed a and b. The graph above shows the case a=1.4, b=2, and the graph below shows the same a for the inverse transformed gamma, which is the limit of the transformed beta as b goes to infinity. This is a high enough a to have some cases with a reasonably high mode, say 40% of the mean, but a is still small enough to be of potential use in US liability insurance. For b=2, the ratio is a strictly increasing function of c. For the limiting case, the ratio reaches a peak and then declines slightly after that. This will also be the behavior for large values of b.



Thus something near the highest value of the mode-to-mean ratio for a given a and b is provided by the limit of the transformed beta when c goes to infinity, which is in fact the split simple Pareto distribution. Its density f(x) is proportional to $(x/d)^{b-1}$ for x<d and to $(x/d)^{-a-1}$ for x>d. The density is continuous but not differentiable at d, which is the mode when b > 1. The mean is dab/[(b+1)(a-1)]. Thus the ratio of mode to mean is (a-1)(b+1)/ab = (1–1/a)(1+1/b). This is increasing in a and decreasing in b. The graph above shows this ratio for a from 1.05 to



2.5 and b-1 from 0.01 to 40 on a geometric scale.

For low values of a, the ratio cannot get very high, as the mean is increased by the heavy tail. The ratio for this distribution is close to the upper limit for the transformed beta with the same a and b, so for low values of a, the c parameter is not going to be able to have much effect on the mode for any transformed beta distribution.

The ratio declines for increasing b, but rather slowly. It is interesting that for this limiting distribution, the maximum mode-to-mean ratio is as b approaches 1, while for the transformed beta the mode is zero at b=1. The split simple Pareto at b=1 is the uniform Pareto, which is uniform up to d and Pareto after that. Thus its mode is undefined, or it could be considered to be the whole interval [0,b].

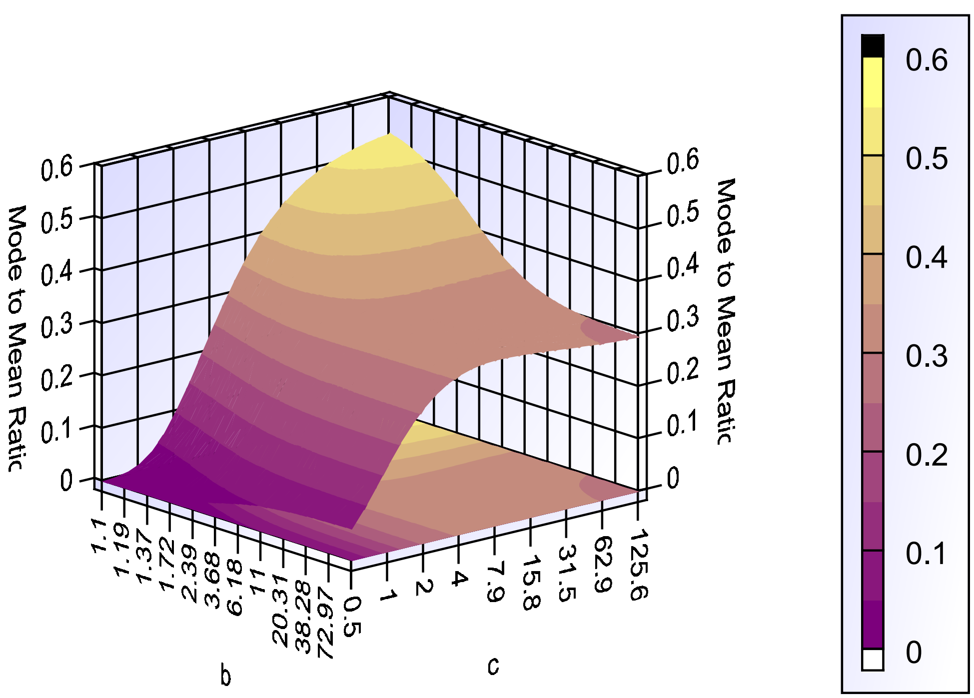
The split simple Pareto shows the maximum, and thus the range of mode-to-mean ratios for any a and b. How is this ratio affected by c? The transformed beta mean is:

$d\Gamma(b/c+1/c)\Gamma(a/c-1/c)/[\Gamma(a/c)\Gamma(b/c)]$, for $a > 1$, and the mode is:

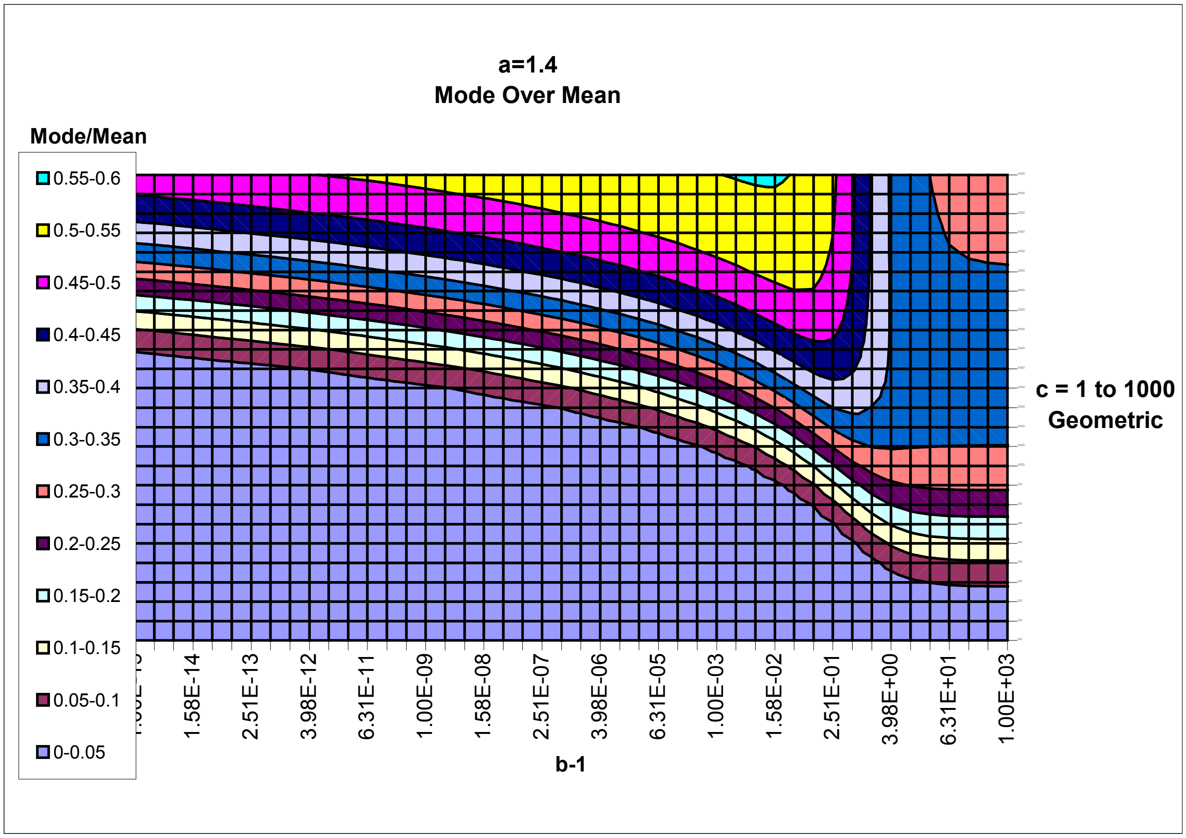
$d[(b-1)/(a+1)]^{1/c}$, for $b>1$

This makes the mode to mean ratio:

$[(b-1)/(a+1)]^{1/c}\Gamma(a/c)\Gamma(b/c) /[\Gamma(b/c+1/c)\Gamma(a/c-1/c)]$ for a, b >1. The graph below shows the ratio for a=1.4 and a range of b's. The high values of b can be seen to have a relationship to c similar to that for the inverse transformed gamma, with a decrease in the ratio for higher values of c. The



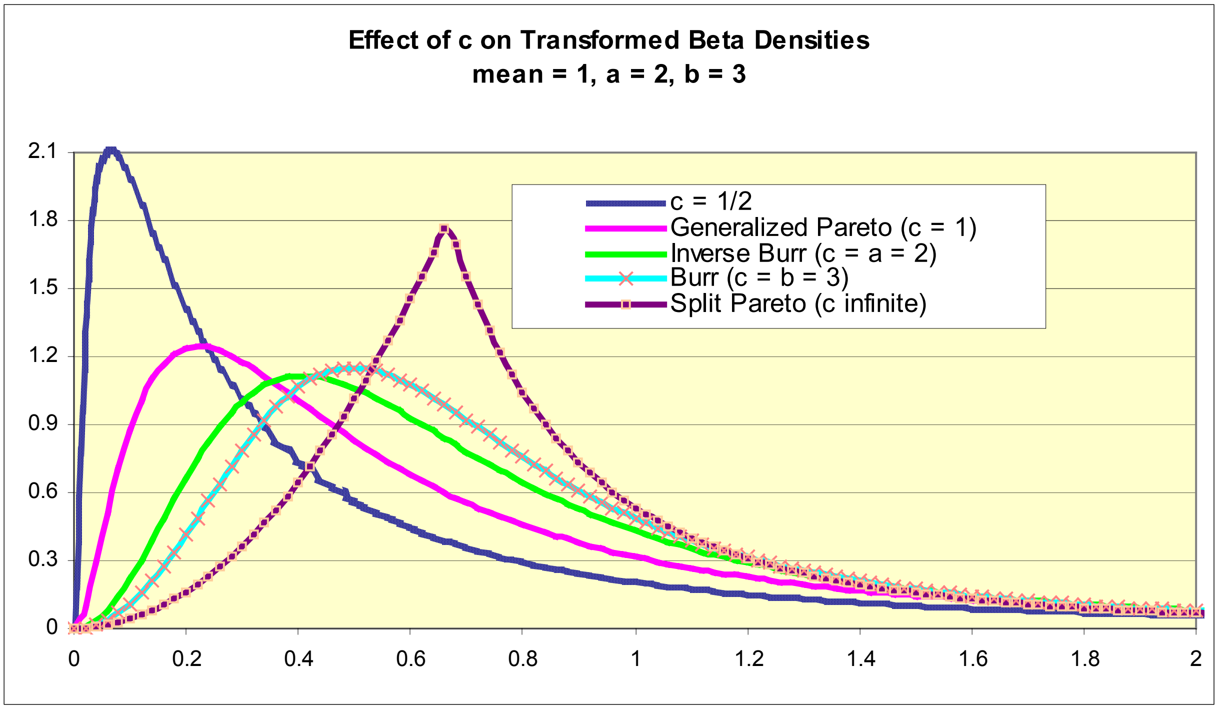
lower values of b have a strictly increasing function of c, like the case b=2 above. The contours of this surface are shown below for a wider range of values.

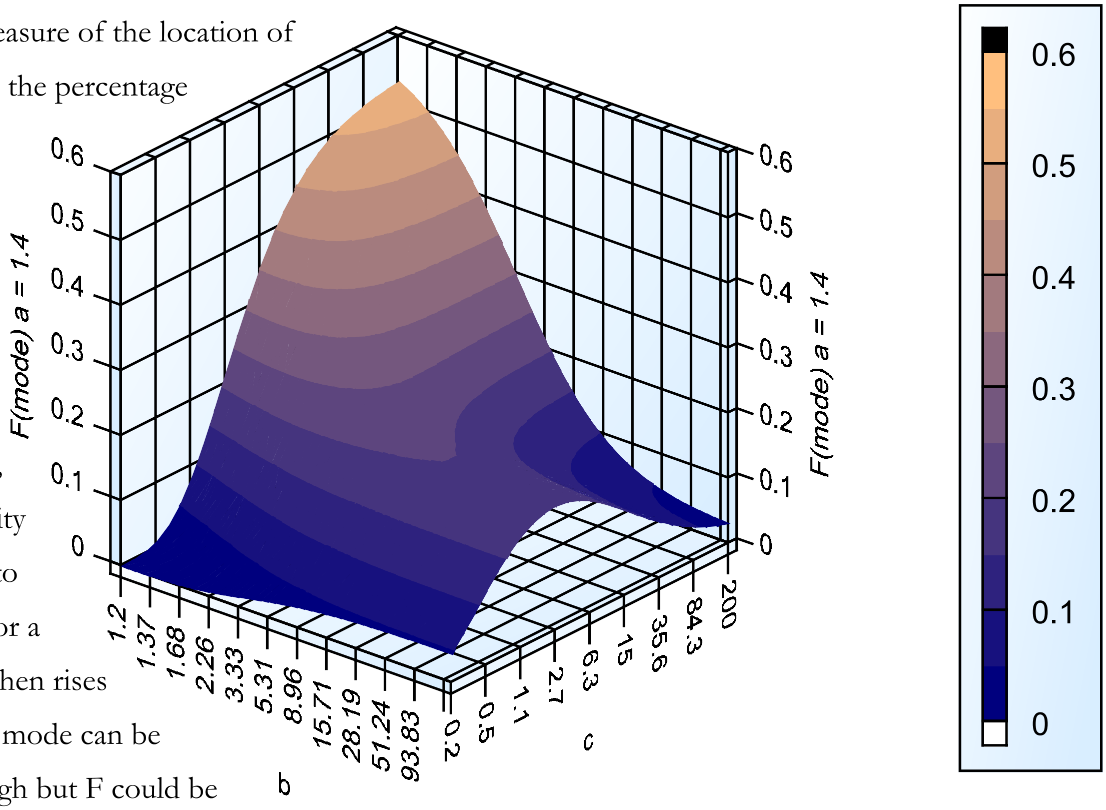


The vertical lines are the contours of the function for fixed values of b, like what was graphed for b=2 above. The horizontal lines are the contours of the function for fixed values of c. For the smaller c's the ratio is an increasing function of b, but very slowly increasing, so b has little impact. For larger c's the function increases then decreases. The maximum seems to hit fairly early, like around 1.01 to 1.25. This is somewhat surprising, in that for b=1, the mode is zero. Thus the mode increases rapidly for b just above 1, especially for higher values of c.

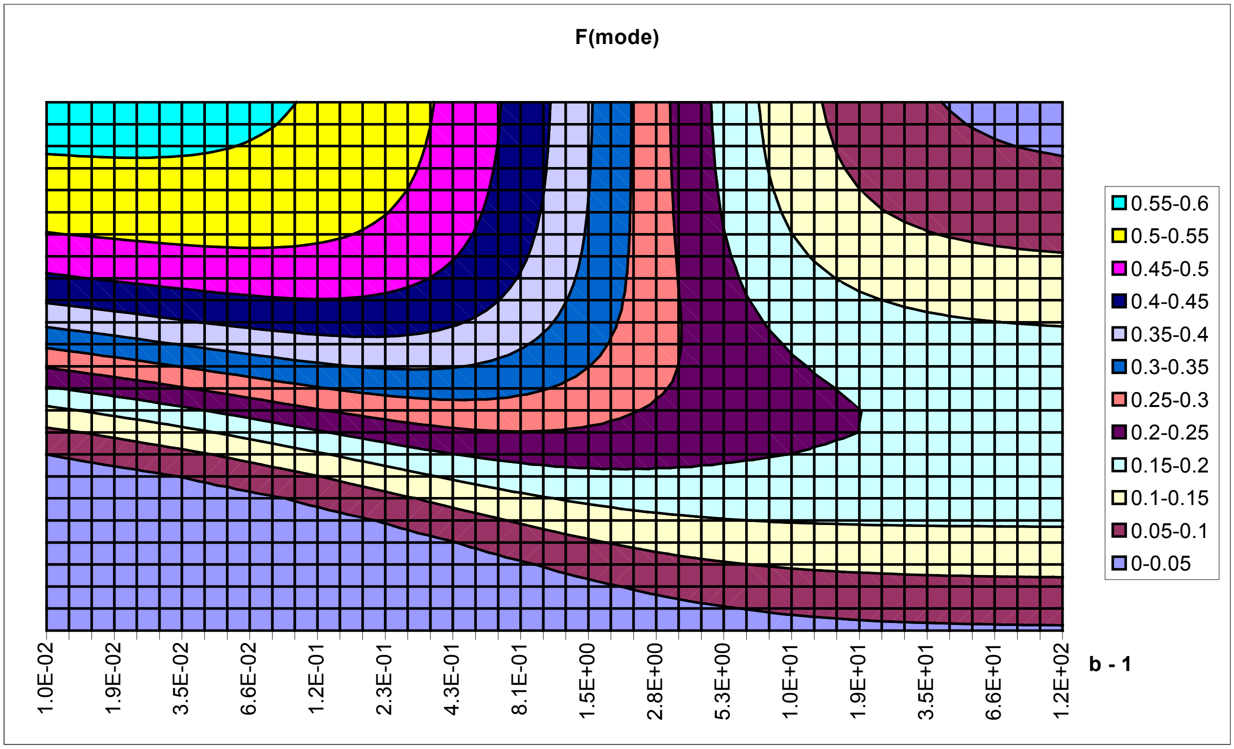
The graph above starts at $b–1= 10^{–15}$, which is the smallest value for which Excel can do this calculation. Even at this level, higher values of c give modes substantially above zero. The mode is above 1% of the mean for c as low as 9 for $b–1= 10^{–15}$.

To illustrate the effect of the c parameter, and thus the mode, on the density function, several cases are illustrated on the graph below. All the distributions have a=2, b=3, and mean=1.

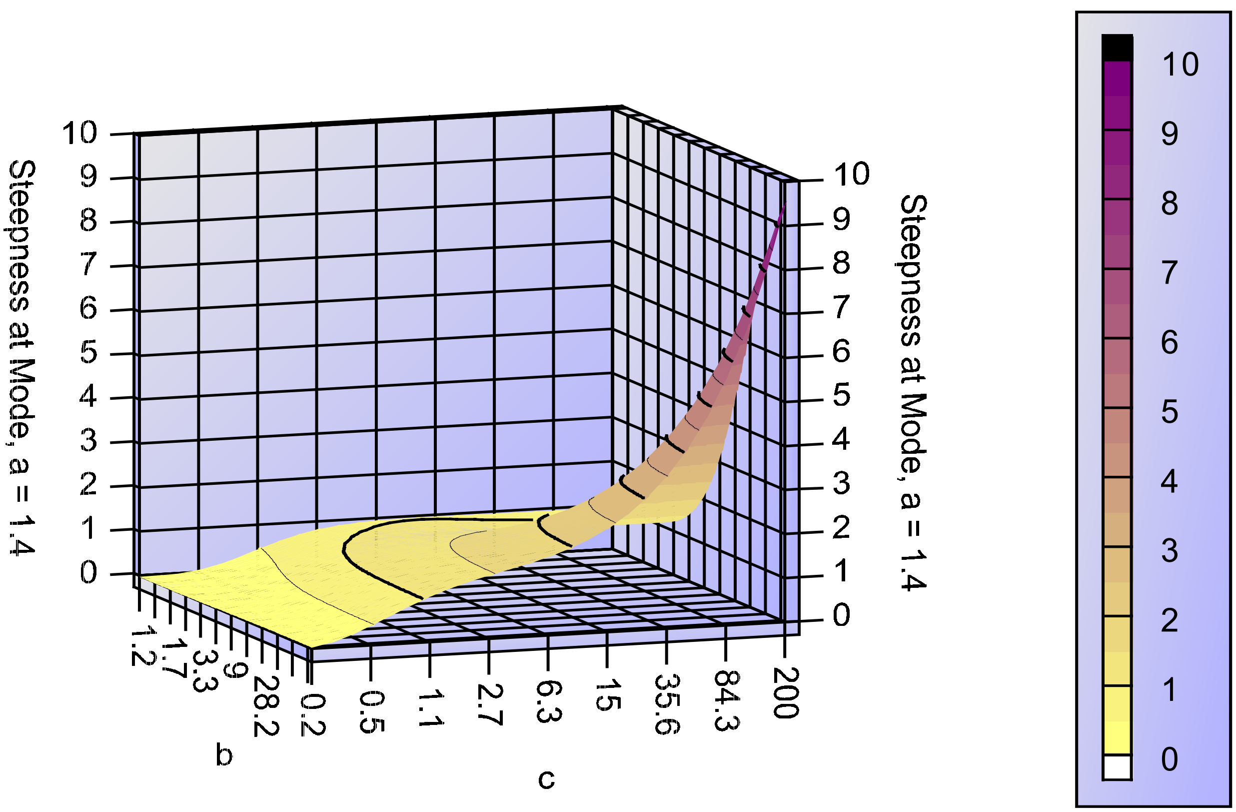




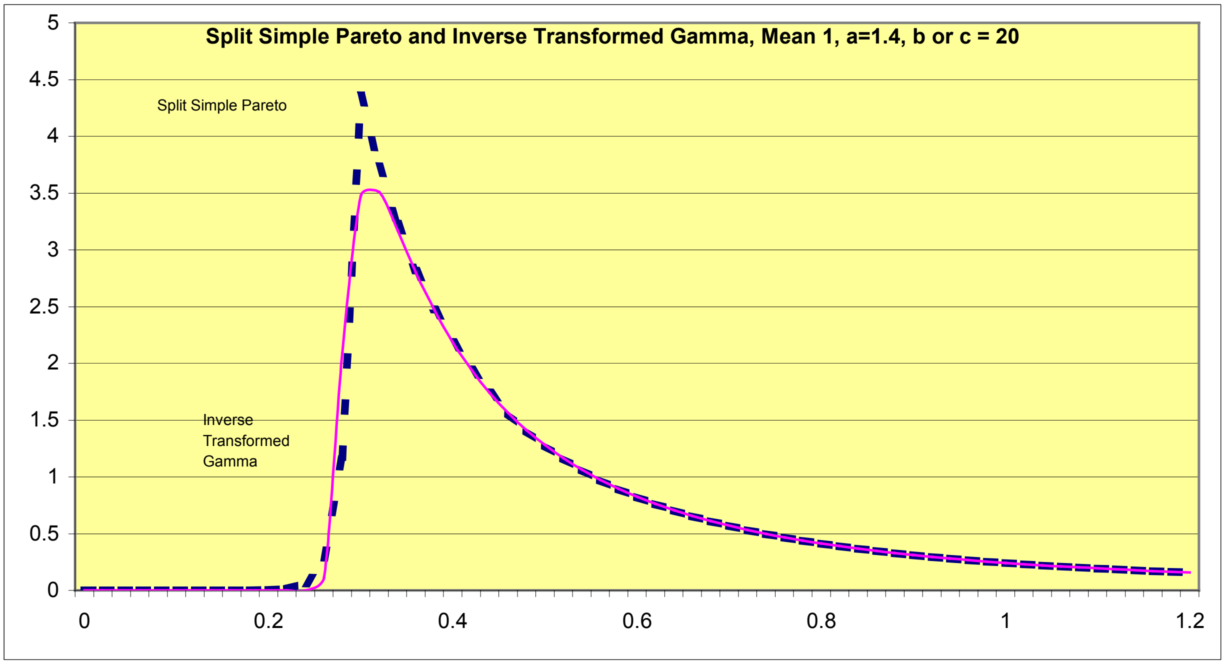
Another measure of the location of the mode is the percentage of the distribution that is below it, i.e., F(mode). If b is high, so the density stays close to the x-axis for a while, and then rises steeply, the mode can be relatively high but F could be low at that point. This is generally the case for high b and c. The graph above shows F(mode) for a wide range of b and c values for a=1.4. The contours are shown below.



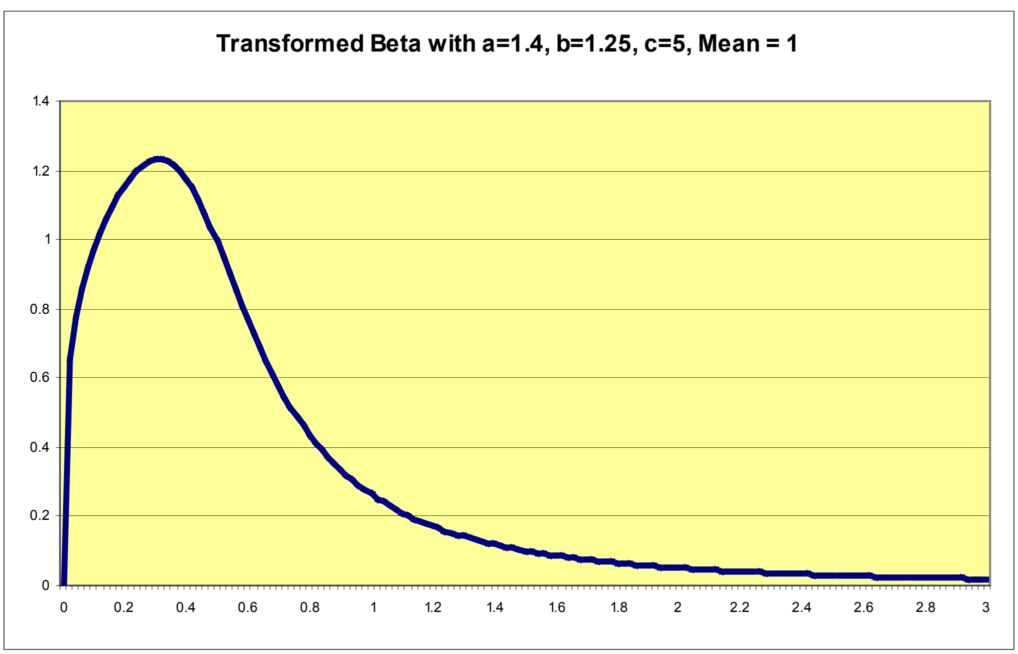
This has a very similar shape to the mode-to-mean ratio. Dividing that by this gives an indication of how steep is the distribution just below the mode. A graph of the steepness is below.



The steepness measure is between zero and two for most of the range of b and c values. It is only in the upper right, with high values of b and c, that the steepness gets very high. The limiting distribution where both b and c go to infinity is the simple Pareto: $F(x) = 1 - (d/x)^a$. This is both the limit of the inverse transformed gamma as c goes to infinity and the split simple Pareto as b goes to infinity.



Examples of both that goes towards the simple Pareto limit are graphed above. Both show a steep rise to the mode. For the split simple Pareto, the mode is at 0.30, and F(mode) = 0.065. This is closer to the limiting case of 0.3 and 0 than is the inverse transformed gamma, with mode of 0.31 and F(mode) = 0.11.

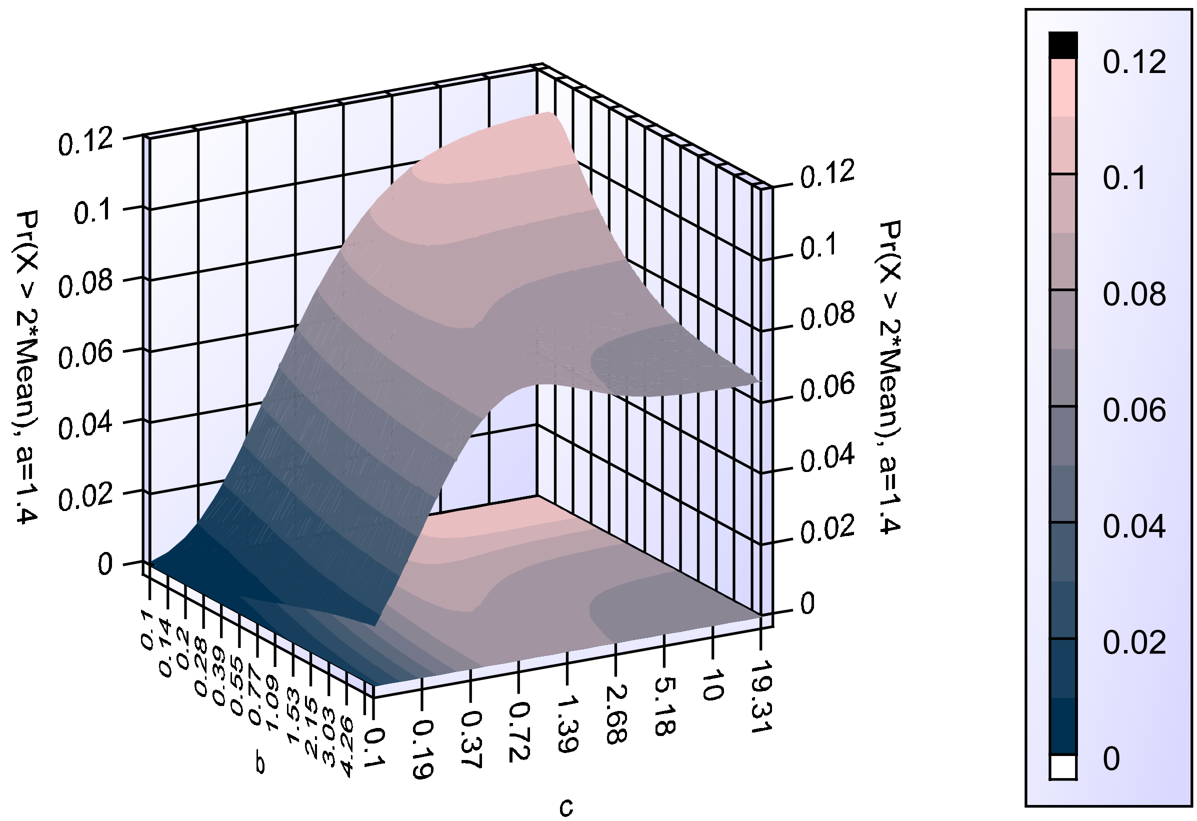


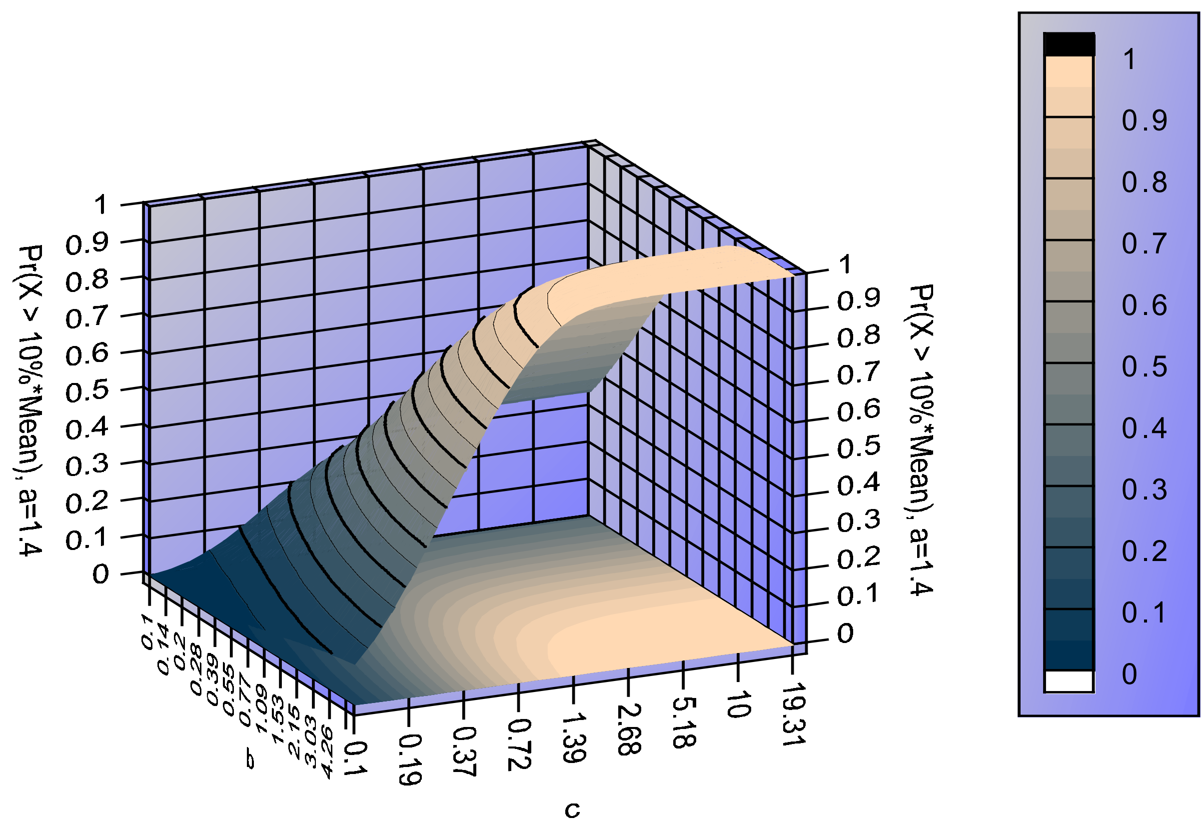
In contrast, a low b with a high c puts a lot more of the probability below the mode. The graph here shows the case b=1.25, c=5, which also has a mode of 0.31 but a higher F(mode) of 0.32. It is clear from the graph that a lot more of the distribution is below the mode in this case.

Another effect of low values of c can be to increase the tails, even though this might not show up in moments. An interesting example is the Weibull distribution, for which a is infinite so all positive moments exist, and b=c. Taking b=c=0.2 gives a fairly heavy-tailed distribution for which all positive moments exist. This has been traditionally used in the US workers compensation line. As an example, take d=100, with a mean of 12,000. In this case, the pretty large loss – 1-in-10,000 claim – is 6.6M, or 550 times the mean. This is heavier-tailed by this measure than most Pareto distributions. For instance, with a=1.4, this ratio is 287. The cv^2 for this Weibull distribution is 251, so the cv itself is almost 16. For contrast, a lognormal with the same mean and cv would have the 1-in-10,000 loss about 4,750,000. The Weibull has another strange feature, however. As b is so small, negative moments do not exist except for powers closer to zero than –0.2. This means that a lot of the distribution is packed in towards zero. In fact, about 33% of the claims are less than 1, and the median claim is 16. The comparable lognormal, which can be given in the limit of a and b both going to infinity, has only 0.2% of its claims less than 1, even though the mode is 3. The median claim is 756 for this distribution.

A small c can pump up the tail of the transformed beta as well. For instance, taking a=1.4, and b=c=0.2 gives a Burr distribution where the 1-in-10,000 claim is 910 times the mean, and over 50% of the claims are below $1/12{,}000^{th}$ of the mean. Keeping this value of c, but letting b get larger, can allow the pretty large claim to be a high multiple of the mean without so many small claims. For instance, taking b=1 (which gives the Pareto T), the pretty large loss is 795 times the mean, and only 7% of claims are below $1 when the mean is $12,000. Taking b up to 5, keeping a=1.4 and c=0.2, these numbers come down to 673 times and 0.1%, which is still very heavy tailed without pushing so many claims to unrealistically small sizes. Although this distribution has a positive mode, it is at 0.065% of the mean, or 7.8 for a mean of 12,000, so is close to zero.

To get an idea of how b and c influence the tail heaviness, the probability that a loss is greater than twice the mean is shown by b and c for a=1.4 below. The graph after that shows the probability of being greater than 10% of the mean.


Pr(X > 2*Mean), a=1.4
0.12
0.1
0.08
0.06
0.04
0.02
0
b
0.1 0.14 0.2 0.28 0.39 0.55 0.77 1.09 1.53 2.15 3.03 4.26
c
0.1 0.19 0.37 0.72 1.39 2.68 5.18 10 19.31


Pr(X > 10%*Mean), a=1.4
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
b
0.1 0.14 0.2 0.28 0.39 0.55 0.77 1.09 1.53 2.15 3.03 4.26
c
0.1 0.19 0.37 0.72 1.39 2.68 5.18 10 19.31

d

The parameter d is a scaling factor. It's effect is just like re-scaling the x-axis. For instance, to convert a distribution expressed in pounds to Canadian dollars, just multiply the scale parameter by 3 (typically). Then a probability for an amount expressed in Canadian dollars would be the same as for the equivalent amount expressed in pounds.

Where did b and c go?

Several two-parameter cases of the transformed beta have just the a and d parameters. To understand what they are doing, it is helpful to know how b and c were disposed of. Some examples:

Ballasted Pareto: b=c=1, so moments in (-1,a), mode zero. Closed form and invertible.
Loglogistic: a=b=c, so moments in (-a,a), and thus the mode is positive if the mean exists, but is probably pretty small with a low steepness. Closed form and invertible.
Inverse Weibull: b infinite, c=a, also closed form and invertible for simulation. Mode is always positive .
Inverse Gamma: b infinite, c=1, so mode positive but usually less than for inverse Weibull. Not closed form.
Simple Pareto: b=c=infinity, so positive mode, infinite steepness, F(mode) = 0. The opposite extreme from the ballasted Pareto for b and c. Invertible.
Uniform Pareto: c infinite, b=1. Mode ambiguous – whole range from 0 to d is uniform. Intermediate between ballasted and simple Paretos and mirror image of inverse gamma in parameters. Invertible.

References

James McDonald (james_mcdonald@byu.edu) Some Generalized Functions for the Size Distribution of Income. Econometrica, 1984, vol. 52, issue 3, pages 647-63
Gary G. Venter (gary.g.venter@guycarp.com) Transformed Beta and Gamma Distributions and Aggregate Losses. PCAS LXX, 1983 pages 156-93
Rodney Kreps (rodney.e.kreps@guycarp.com) Continuous Distributions.doc