Fitting Beta Densities to Loss Data

Daniel R. Corro

Fitting Beta Densities to Loss Data

Dan Corro
National Council on Compensation Insurance, Inc.
March, 2002

Abstract: *This short note details how to match the mean and variance of any loss distribution on a finite interval to a Beta density, scaled to that interval.*

Most loss distributions that actuaries use (c.f. [1], Appendix) are naturally defined on $(0,\infty)$. In this note we consider instead loss distributions defined on a finite interval *(0,L)* of positive width *L*>0. We require throughout that all loss distributions considered have a positive mean and finite mean and variance. So let $\mu > 0$ denote the mean and σ^2 the variance of any such distribution. Also, let $\gamma = \dfrac{\sigma}{\mu}$ be the coefficient of variation and X be the associated random variable of such a loss distribution. Then, since $X < L$, we have the following inequality that will come in handy later:

$$\mu(1+\gamma^2) = \frac{\mu^2+\sigma^2}{\mu} = \frac{E(X^2)}{E(X)} = \frac{\sum x^2 p(x)}{\sum xp(x)} < \frac{\sum Lxp(x)}{\sum xp(x)} = L.$$

The Beta density on (0,1) is among the most useful of this class of loss densities. Recall that the Beta distribution is a two-parameter, α, β, distribution that is usually defined in terms of its probability density function [PDF]:

$$f(\alpha,\beta;x) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{\mathrm{B}(\alpha,\beta)} = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1} \quad x \in (0,1), \alpha > 0, \beta > 0$$

where B and Γ denote the usual Beta and Gamma functions (c.f. [1], p. 48). For this Beta density, the mean and variance are:

$$\mu = \frac{\alpha}{\alpha+\beta} \quad \text{and} \quad \sigma^2 = \frac{\alpha\beta}{(\alpha+\beta+1)(\alpha+\beta)^2}.$$

Indeed, the reduction formula $\Gamma(x) = (x-1)\Gamma(x-1)$ leads directly to:

$$E(X^n) = \frac{\mathrm{B}(\alpha+n,\beta)}{\mathrm{B}(\alpha,\beta)} = \prod_{i=0}^{n-1} \frac{\alpha+i}{\alpha+\beta+i}.$$

Thus, the moments of the Beta density are easy to compute from the parameters.

It is also easy to verify that two ordered pairs of parameters α, β and α', β' have the same mean if and only if the points (α, β) and (α', β') lie on the same line through the origin, i.e., if and only if $\alpha' = \rho\alpha$ and $\beta' = \rho\beta$ for some fixed proportionality constant $\rho > 0$. It is then apparent that the pair α, β is uniquely determined by the mean and variance.

With these preliminaries out of the way, suppose we have bounded loss data with loss amounts in $(0,L)$ that we want to model or otherwise approximate using a continuous PDF on $(0,L)$. Assume we have determined the following statistics for the data:

$$\text{mean} = m > 0 \qquad \text{and} \qquad \text{variance} = s^2 > 0.$$

Let $c = \dfrac{s}{m}$ and consider what the data looks like scaled into the interval $(0,1)$. Evidently the scaled data has mean and standard deviation:

$$\hat{m} = \frac{m}{L} > 0 \qquad \hat{s} = c\hat{m} = \frac{s}{L}.$$

In particular, the above inequality implies [$\Rightarrow$] that:

$$\hat{m}(1+c^2) < 1 \Rightarrow 1 - \hat{m} - c^2\hat{m} > 0.$$

So we may define:

$$\alpha = \frac{1-\hat{m}-c^2\hat{m}}{c^2} > 0 \quad \text{and} \quad \beta = \left(\frac{1-\hat{m}}{\hat{m}}\right)\alpha > 0$$

as the parameters of a Beta density. Now observe that this Beta density has mean

$$\frac{\alpha}{\alpha+\beta} = \frac{\alpha}{\alpha + \left(\frac{1-\hat{m}}{\hat{m}}\right)\alpha} = \frac{1}{1+\left(\frac{1-\hat{m}}{\hat{m}}\right)} = \frac{\hat{m}}{\hat{m}+1-\hat{m}} = \hat{m}.$$

Observe next that:

$$\frac{\alpha}{\alpha+\beta}=\hat{m} \Rightarrow \frac{\alpha+\beta}{\alpha}=\frac{1}{\hat{m}} \Rightarrow \alpha+\beta=\frac{\alpha}{\hat{m}}.$$

And then we find, as no surprise, that the variance is:

$$\frac{\alpha\beta}{(\alpha+\beta+1)(\alpha+\beta)^2}=\frac{\alpha^2\left(\frac{1-\hat{m}}{\hat{m}}\right)}{\left(\frac{\alpha}{\hat{m}}+1\right)\left(\frac{\alpha}{\hat{m}}\right)^2}=\hat{m}^2\left(\frac{1-\hat{m}}{\alpha+\hat{m}}\right)$$

$$=\hat{m}^2\left(\frac{1-\hat{m}}{\frac{1-\hat{m}-c^2\hat{m}}{c^2}+\hat{m}}\right)=\hat{m}^2\left(\frac{1-\hat{m}}{\frac{1-\hat{m}}{c^2}}\right)=(c\hat{m})^2=\hat{s}^2.$$

It follows that these parameters define a Beta density whose mean and variance equal those of the rescaled loss data.

In terms of the original scale, the approximating Beta continuous PDF is:

$$g(\alpha,\beta;z)=\frac{z^{\alpha-1}(L-z)^{\beta-1}}{\mathrm{B}(\alpha,\beta)L^{\alpha+\beta+1}} \quad z\in(0,L).$$

It is worth emphasizing that this construction is quite sensitive to the choice of L. In general, unless some other applicable loss limitation prevails, it is usually best to select L at or near the maximum observed loss.

By a continuous density on the finite interval $(0,L)$, we mean a density that can be specified via a PDF, $f(x)$, that is defined and continuous on $(0,L)$. We may summarize what we have shown in two simple results:

Proposition 1: *The following condition is both necessary and sufficient for a pair μ,σ^2 of positive real numbers to be the mean and variance of a continuous density on* $(0,1)$:

$$\mu\left(1+\left(\frac{\sigma}{\mu}\right)^2\right)<1.$$

Proof: Necessity follows from the inequality shown earlier and sufficiency from our discussion of the Beta density, which also establishes:

Proposition 2: *Within the class of all continuous densities on* (0,1) *with a given mean* $= \mu$ *and variance* $= \sigma^2$, *there is exactly one Beta density that is uniquely determined by the parameters:*

$$\alpha = \frac{1-\mu-\gamma^2\mu}{\gamma^2} > 0$$

$$\beta = \left(\frac{1-\mu}{\mu}\right)\alpha > 0$$

where $\gamma = \dfrac{\sigma}{\mu}$.

The so-called "central moments" μ, σ^2 may not be the most convenient for this purpose. Let $\mu = \mu_1 = E(X)$ and $\mu_2 = E(X^2)$ be the first and second moments of a continuous density on (0,1), then:

$$X \le X^2 \Rightarrow \mu_1 = E(X) \le E(X^2) = \mu_2$$

$$E((X-E(X))^2) \ge 0 \Rightarrow E(X^2) - E(X)^2 \ge 0 \Rightarrow \mu_2 = E(X^2) \ge E(X)^2 = \mu_1^2 .$$

And, equivalent to the above, we obtain the corresponding Beta density parameters:

$$\alpha = \mu_1\left(\frac{\mu_1 - \mu_2}{\mu_2 - \mu_1^2}\right) > 0 \quad \text{and} \quad \beta = \frac{(\mu_1 - \mu_2)(1-\mu_1)}{\mu_2 - \mu_1^2} > 0.$$

References:

[1] *Hogg, Robert V., Klugman, Stuart A.,* Loss Distributions, *Wiley Series in Probability and Statistics, John Wiley & Sons, Inc., 1984.*