

Beta Distribution

Paul Johnson <pauljohn@ku.edu> and Matt Beverlin <mbeverlin@ku.edu>

March 11, 2006

1 Description

How likely is it that the Communist Party will win the next elections in Russia?

In my view, the probability is 0.33. You may think the chance is 0.22. Your brother thinks the chances are 0.6 and his sister-in-law thinks the chances are 0.1.

People disagree and we would like a way to summarize the chances that a person will say 0.4 or 0.5 or whatnot.

The Beta density function is a very versatile way to represent outcomes like proportions or probabilities. It is defined on the continuum between 0 and 1. There are two parameters which work together to determine if the distribution has a mode in the interior of the unit interval and whether it is symmetrical.

The Beta can be used to describe not only the variety observed across people, but it can also describe your subjective degree of belief (in a Bayesian sense). If you are not entirely sure that the probability is 0.22, but rather you think that is the most likely value but that there is some chance that the value is higher or lower, then maybe your personal beliefs can be described as a Beta distribution.

2 Mathematical Definition

The standard *Beta* distribution gives the probability density of a value x on the interval $(0,1)$:

$$Beta(\alpha, \beta) : prob(x|\alpha, \beta) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha, \beta)} \quad (1)$$

where B is the beta function

$$B(\alpha, \beta) = \int_0^1 t^{\alpha-1}(1-t)^{\beta-1} dt$$

2.1 Don't let all of those betas confuse you.

It is disappointingly confusing, but the word "beta" is used for 3 completely different meanings.

1. $Beta(\alpha, \beta)$ "Beta" is the name of the probability distribution

2. $B(\alpha, \beta)$ “Beta” is the name of a function that appears in the denominator of the density function
3. β “Beta” is the name of the second parameter in the density function

2.2 About the Beta function B

The Beta function B in the denominator plays the role of a “normalizing constant” which assures that the total area under the density curve equals 1.

The Beta function is equal to a ratio of Gamma functions:

$$B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha + \beta)}$$

Keeping in mind that for integers, $\Gamma(k) = (k - 1)!$, one can do some checking and get an idea of what the shape might be.

A 3 dimensional graph of the Beta function can be found in Figure 1.

3 Moments of the Beta

The expected value of a variable that is Beta distributed is:

$$E(x) = \mu = \frac{\alpha}{\alpha + \beta} \tag{2}$$

and the variance is

$$Variance(x) = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)} \tag{3}$$

People who are familiar with the Generalized Linear Model will notice that

$$V(\mu) = \frac{\beta}{(\alpha + \beta)(\alpha + \beta + 1)} \cdot \mu$$

is a variance function, $V(\mu)$, which indicates the dependence of the observed variance on the mean. For a fixed pair of parameters (α, β) , the variance is proportional to μ . A graph illustrating the Variance function is presented in Figure 2.

The third and fourth moments are:

$$Skewness(x) = \frac{2(\beta - \alpha)\sqrt{1 + \alpha + \beta}}{\sqrt{\alpha + \beta}(2 + \alpha + \beta)} \tag{4}$$

$$Kurtosis(x) = \frac{6[\alpha^3 + \alpha^2(1 - 2\beta) + \beta^2(1 + \beta) - 2\alpha\beta(2 + \beta)]}{\alpha\beta(\alpha + \beta + 2)(\alpha + \beta + 3)} \tag{5}$$

Figure 1: The Beta Function

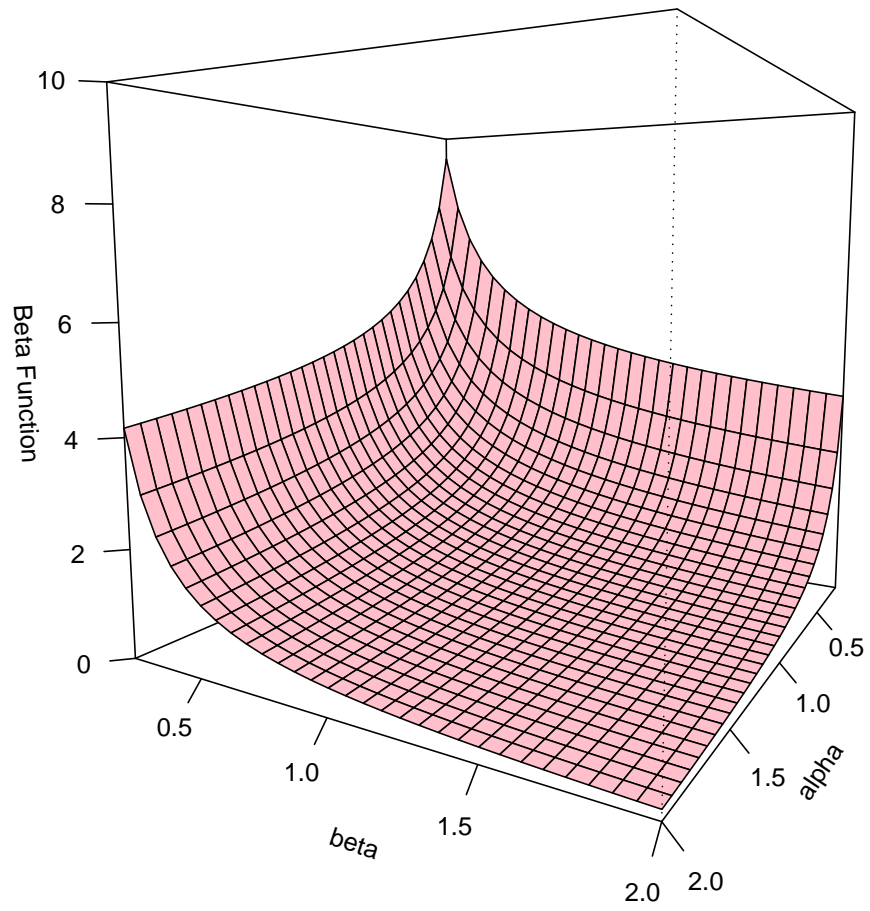
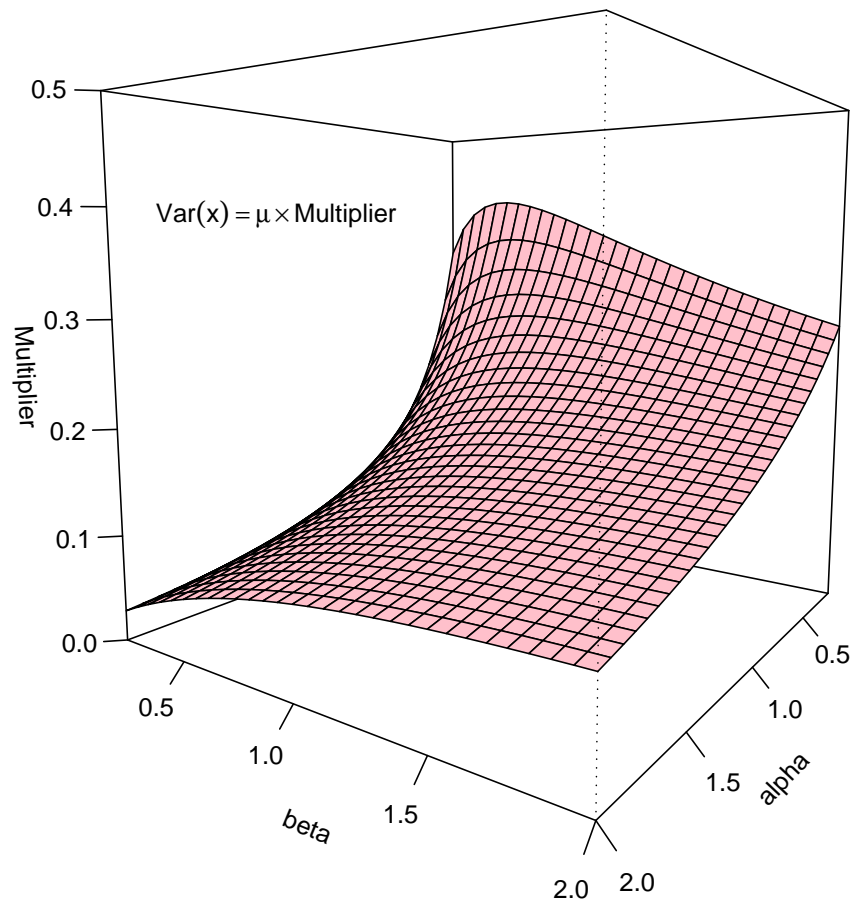


Figure 2: The Ratio of Variance to the Mean



3.1 The Mode

If $\alpha > 1$ and $\beta > 1$, the peak of the density is in the interior of $[0,1]$ and mode of the *Beta* distribution is

$$mode = \gamma = \frac{\alpha - 1}{\alpha + \beta - 2} \tag{6}$$

If α or $\beta < 1$, the mode may be at an edge.

As we will illustrate below, if $\alpha = \beta = 1$, then the Beta is identical to a Uniform distribution.

4 Illustration

One advantage of the Beta distribution is that it can take on many different shapes. If one believed that all scores were equally likely, then one could set the parameters $\alpha = 1$ and $\beta = 1$, as illustrated in Figure 3, this gives a “flat” probability density function.

In models of elections, one may need a distribution of ideal points to resemble a single-peaked distribution on the interval $[0,1]$. The Beta can be very useful in this kind of exercise. Consider Figure 4.

At one point, it fascinated me that the mode did not equal the mean and that the variance ends up characterizing the “slack” between those two things. Various densities in Figure 5 might be entertaining. In these examples, the Beta parameters are chosen to keep the mode constant at 0.3. Note how the mean and variance change across the illustrations.

Perhaps, as they say, the versatility of the Beta is a strength, but there is some reason to be cautious. If the parameters wander into the $(0,1)$ range, then some pretty surprising shapes can appear. Consider Figure 6.

5 About the connection between the mean, the mode, and the variance

In the pictures displaying the Beta density, one’s eye is drawn to the peak of the frequency distribution, which is the mode. We can set the Beta’s parameters in order to generate a distribution with a desired mode. Let the mode be represented by γ .

Here’s a simple starting point: Suppose the mode is .50. That is the same as the mean (its symmetric), and the mode formula (6) implies:

$$.50 = \frac{\alpha - 1}{\alpha + \beta - 2} \tag{7}$$

and

$$\begin{aligned} .50\alpha + .50\beta - 1 &= \alpha - 1 \\ .5\alpha &= .5\beta \\ \alpha &= \beta \end{aligned} \tag{8}$$

Figure 3: Beta(1,1) is the Uniform distribution

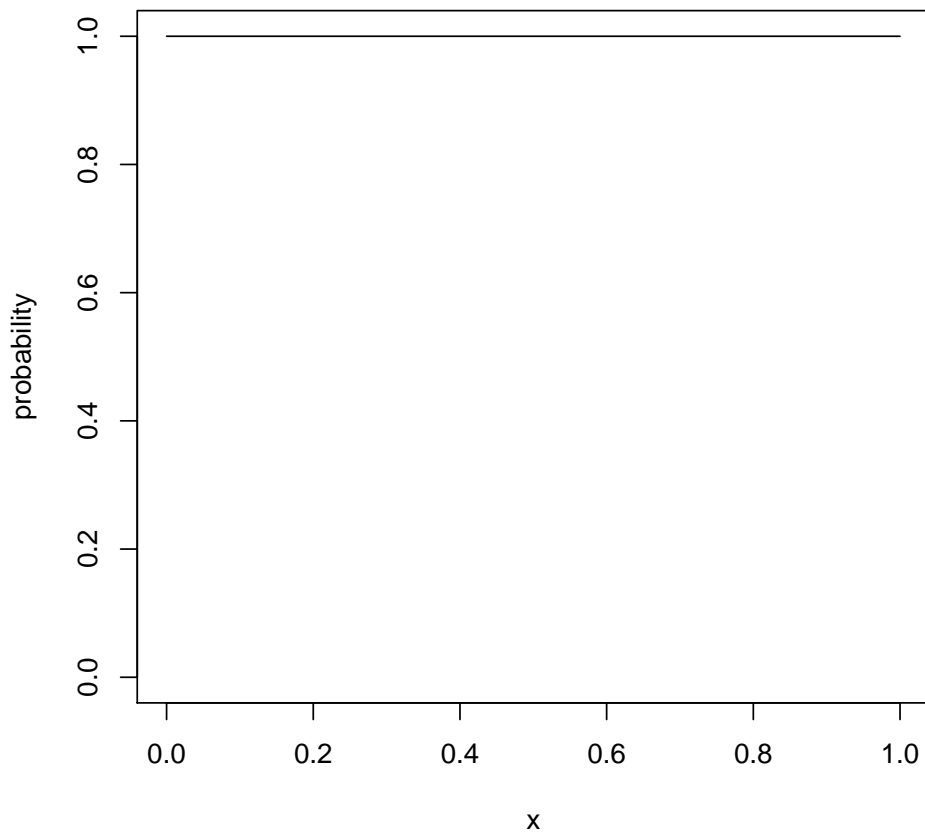


Figure 4: Some Pleasant Single-Peaked Betas

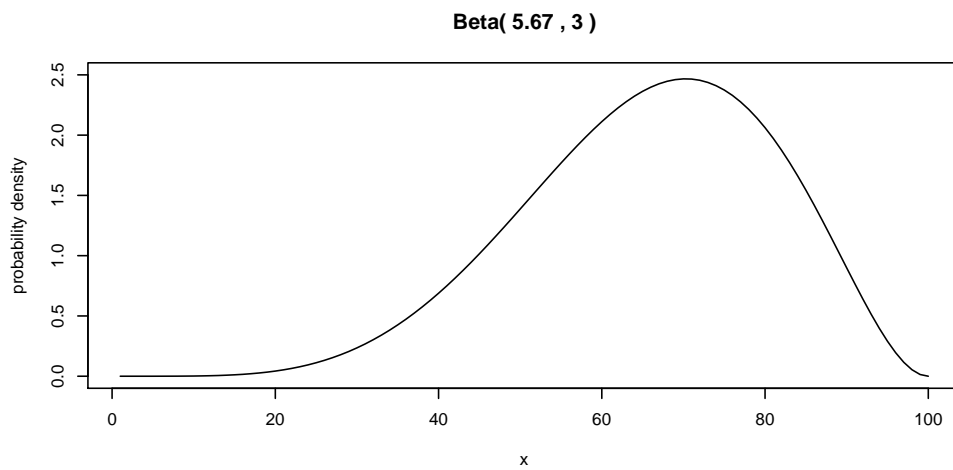
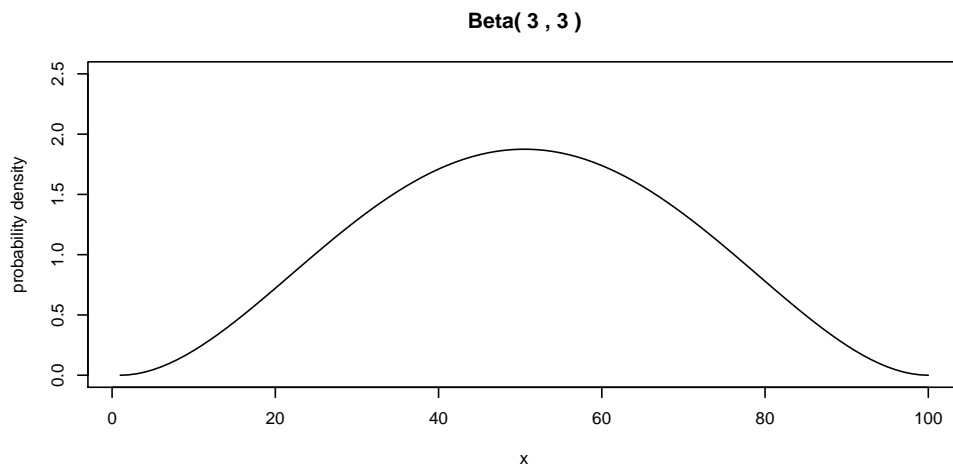
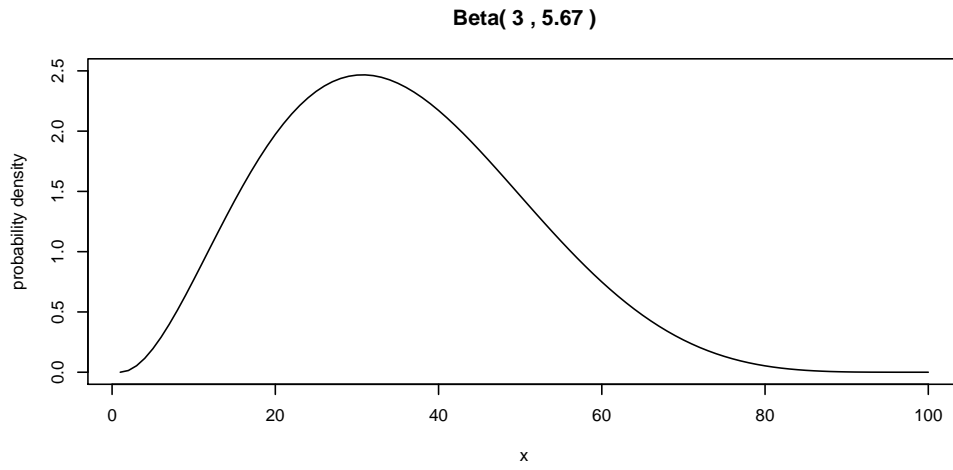


Figure 5: Beta Distributions with Mode=0.3

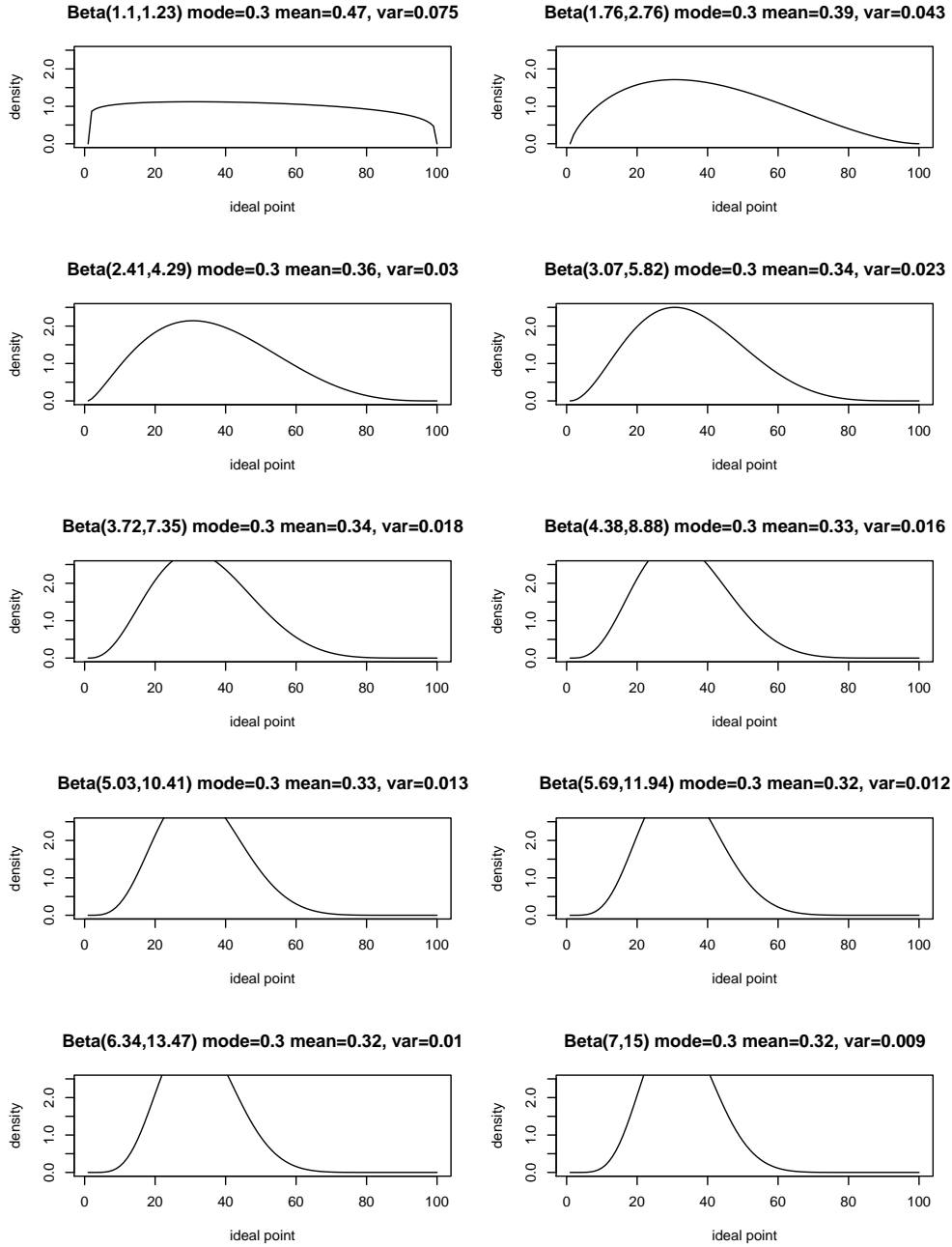
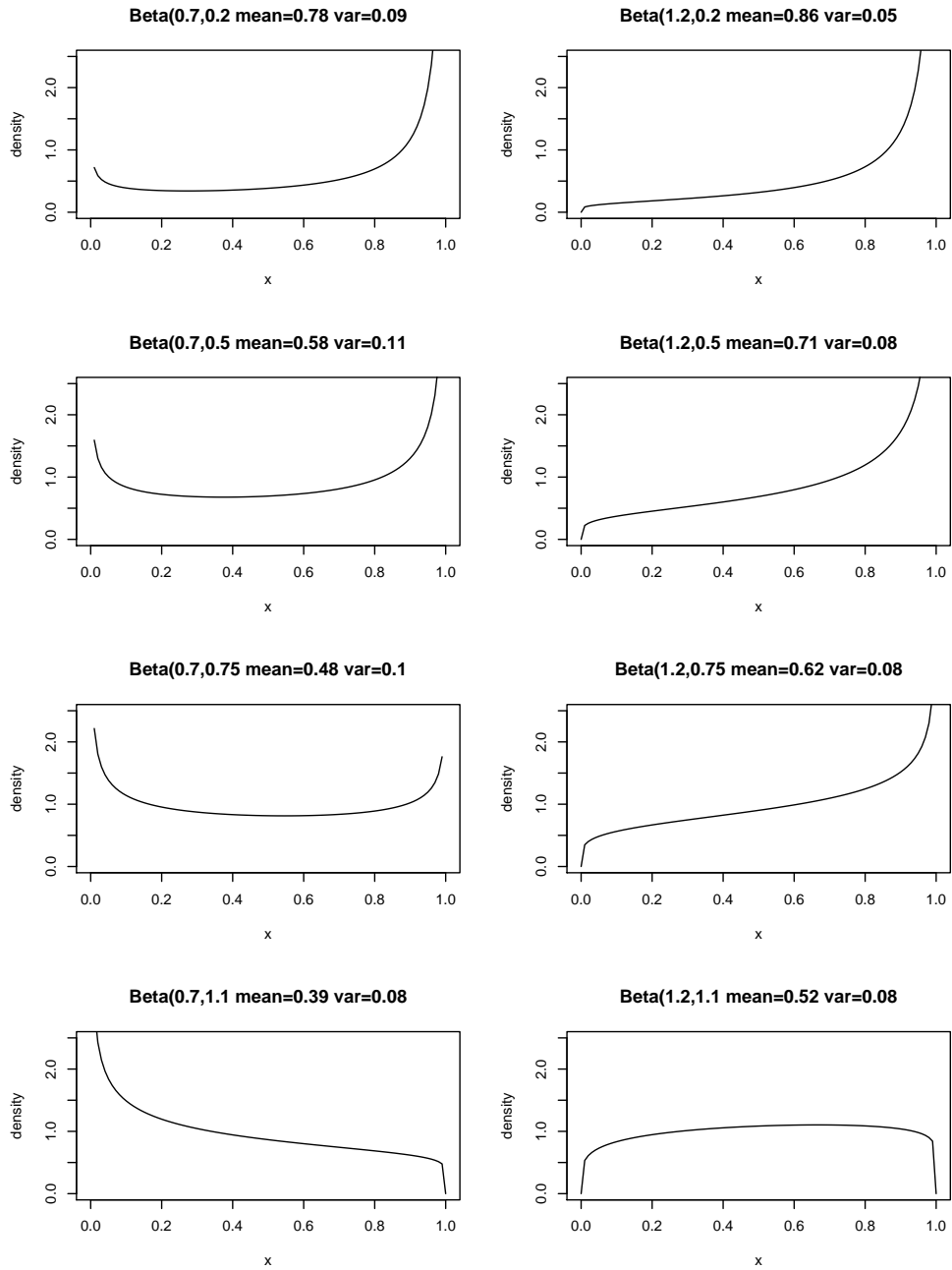


Figure 6: Some Unpleasant Betas



If one wants the mode to be in the middle, one can choose any value for α , as long as one chooses the same value for β . (Whew! What a relief. This exactly matched my intuition.)

If the mode is in the center, we know α and β are equal, but we don't know their values. The selection, it turns out, depends on how much diversity there is. If one wants a distribution to have points "tightly bunched" around the mode, then one should choose a large value for α , say 10.0,

$$\text{variance of Beta}(10, 10) = 0.01190 \quad (9)$$

In contrast, if $\alpha = 1.5$, the variance is much greater:

$$\text{variance of Beta}(1.5, 1.5) = 0.0625 \quad (10)$$

Seen in this light, the parameter α is a "homogeneity indicator." As α gets bigger, the distribution collapses around the mode.

Although this particular calculation works only for a mode in the center, it does outline the process that we can use to assign α and β for all other values of the mode.

Suppose the mode is .4. From equation 6

$$.40 = \frac{\alpha - 1}{\alpha + \beta - 2}$$

$$.40\alpha + .40\beta - 0.8 = \alpha - 1$$

$$.4\beta = .6\alpha - 0.2$$

$$\beta = \frac{3}{2}\alpha - \frac{1}{2} \quad (11)$$

or

$$.60\alpha = .2 + .40\beta$$

$$\alpha = \frac{1}{3} + \frac{2}{3}\beta \quad (12)$$

It is quite possible to calculate one parameter as a function of another, after specifying the mode, even if the mode is off center.

Generally speaking, for any value of the mode, $\gamma \in (0, 1)$ (keeping in mind the original stipulation that $\alpha, \beta > 1$):

$$\gamma = \frac{\alpha - 1}{\alpha + \beta - 2} \quad (13)$$

$$\gamma\alpha + \gamma\beta - 2\gamma = \alpha - 1 \quad (14)$$

$$(1 - \gamma)\alpha = \gamma\beta - 2\gamma + 1 \quad (15)$$

$$\alpha = \frac{\gamma\beta - 2\gamma + 1}{(1 - \gamma)} = \frac{\beta - 2 + \frac{1}{\gamma}}{(\frac{1}{\gamma} - 1)} = \frac{\gamma}{1 - \gamma}\beta - \frac{2\gamma + 1}{1 - \gamma} \quad (16)$$

So α is a linear function of β .

And

$$\gamma\beta = \alpha - 1 - \gamma\alpha + 2\gamma \quad (17)$$

$$\beta = \frac{\alpha - \gamma\alpha + 2\gamma - 1}{\gamma} = \frac{\alpha - \gamma(\alpha - 2) - 1}{\gamma} = \frac{(1 - \gamma)}{\gamma}\alpha - \frac{1 - 2\gamma}{\gamma} \quad (18)$$

This indicates that if we begin with the mode, and then take as given either α or β , we can calculate the missing parameter (β or α , as the case may be). As a result, instead of thinking of the Beta's shape as determined by parameters α and β , sometimes it is easier to think of it in terms of the mode (most likely value) and the homogeneity.