Paul Johnson
2003-12-03 About the Negative Binomial Model

0.1 Recall Poisson.

The Poisson is a handy starting place because it has this nice feature that its central parameter, λ, is equal to its expected value and its variance. So we can create a modeling strategy in which suppose that each case has a different central parameter and we allow some predictor variables to "go into" determining the value of that parameter. That means each case has a customized probability model.

Often, the impact of inputs X_i depends on $X_i b$ in a simple way, like

$$\lambda_i = X_i b$$

or

$$\lambda_i = e^{X_i b}$$

In the case that y_i is Poisson(λ_i), it happens to be that the expected value of y equals λ_i.

$$\mu_i = E(y_i|X_i, b) = \lambda_i$$

The expected value of y_i–the mean for that case–is called μ_i ("mu"). Because I don't want us to get tangled up in notation, sometimes I just call it "input" and don't worry too much how it gets calculated.

$$E(y_i) = \mu_i = input_i$$

In the Poisson model, we usually have an exponential form:

$$input_i = \mu_i = exp(X_i b) = e^{X_i b} = e^{b_0 + b_1 X1_i}$$

the Poisson probability process takes that input and generates $y_i$and it has expected value $input_i$and variance $input_i$.

0.2 You want some Randomness in your Randomness?

It is not enough to say y_i follows a Poisson distribution. If you say that, it means all observations with the same input have same probability distribution. And we expect that's not true sometimes. We need some random variability over-and-above the Poisson variability. Sometimes this additional randomness is called "frailty".

Suppose we want to have some unpredictability, so that after $X_i b$ is given for several cases, then the probability distribution of outcomes can differ between observations. For example, we can say the Poisson process has an expected value of the original thing I called "input" times another random variable:

$$new\,input_i = input_i * \delta_i$$

Note that if

$$\delta_i = 1$$

then this thing just degenerates back to the original Poisson model.

Here's another way to think about it. Suppose u_i is a random variable. Suppose we make a "new" input.

$$new\,input_i = \mu_i = exp(X_i b + u_i)$$

note, that is the same as

$$= exp(X_i b) exp(u_i)$$

If you write

$$\delta_i = exp(u_i)$$

then it reduces to

$$= exp(X_i b) * \delta_i = input * \delta_i$$

If we choose δ_i, or equivalently, u_i carefully, then we can make an interesting model that will actually work.

Here's one other way to think of it:

$$new\,input = \mu_i = exp(X_i b + log(\delta))$$

What would be nice?

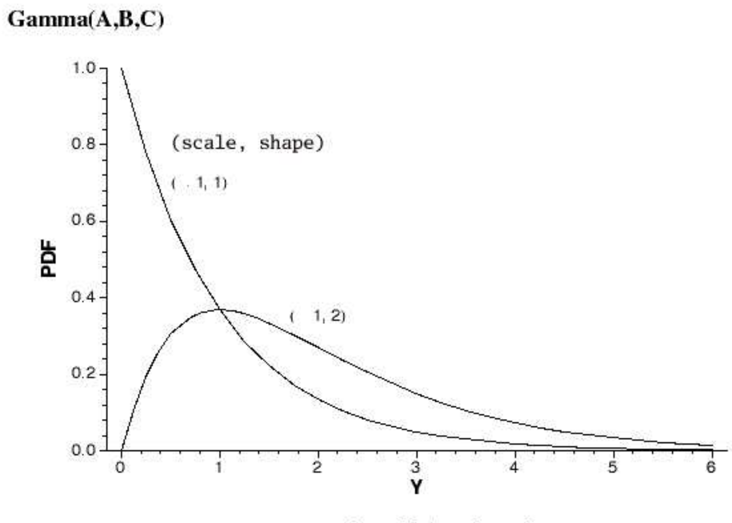
$$E(u_i) = 0$$

And we want something that has variance, of course. And that would imply

$$E(\delta_i) = E(exp(u_i)) = 1$$

0.3 What about Gamma distribution?

Gamma is for a continuous variable on $[0, \infty]$. It can look like a "ski slope" or it can look single-peaked.



It has 2 parameters, shape and scale. In some books, the scale is specified as $1/rate$. We don't care. We are going to try to throw that away by setting scale=1, so rate=1. No big!

If v_i is Gamma distributed, the probability density function is:

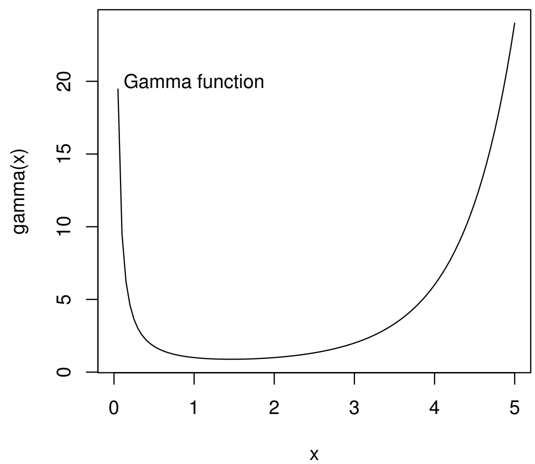
$$f(v_i) = \frac{1}{scale^{shape}\Gamma(shape)} v_i^{(shape-1)} e^{-(\frac{v_i}{scale})}$$

0.4 What is that Gamma function?

The function $\Gamma(shape)$ is the Gamma function (which is a complicated math thing I've never looked into very much). It is $\Gamma(s) = \int_0^\infty t^{s-1}e^{-t}dt$ if $s > 0$. If you pick s as an integer, $\Gamma(s)$ is very easy to calculate:

$$\Gamma(s) = (s-1)!\ s = 1, 2, ...$$

So, the value of $\Gamma(1) = 1$. And $\Gamma(2) = 1$. And $\Gamma(20)$ is some impossibly huge number. I never really worry about it, but in case you do, here's a graph of it:



0.5 Now, back to the Gamma probability distribution.

The Gamma probability distribution has these interesting properties:

$E(v_i) = shape * scale$

$Var(v_i) = shape * scale^2$

If you read a different book, you might find it rearanged like so:

$$f(v_i) = \frac{1}{scale * \Gamma(shape)} \left[\frac{v_i}{scale}\right]^{(shape-1)} e^{-(\frac{v_i}{scale})}$$

That's really just the same distribution.

The first simplification we use is to fix scale = 1. If you do that, you can adjust the shape up and down. I prepared a figure showing the histograms of some random samples from various Gamma distributions, just to convey the "feel" of the distribution.

0.6 Negative Binomial Model.

If you "assume away" the scale parameter, then the probability density formula for the Gamma(scale=1, shape) simplifies to:

$$f(v_i) = \frac{1}{\Gamma(shape)} v_i^{(shape-1)} e^{-v_i} \quad v_i > 0$$

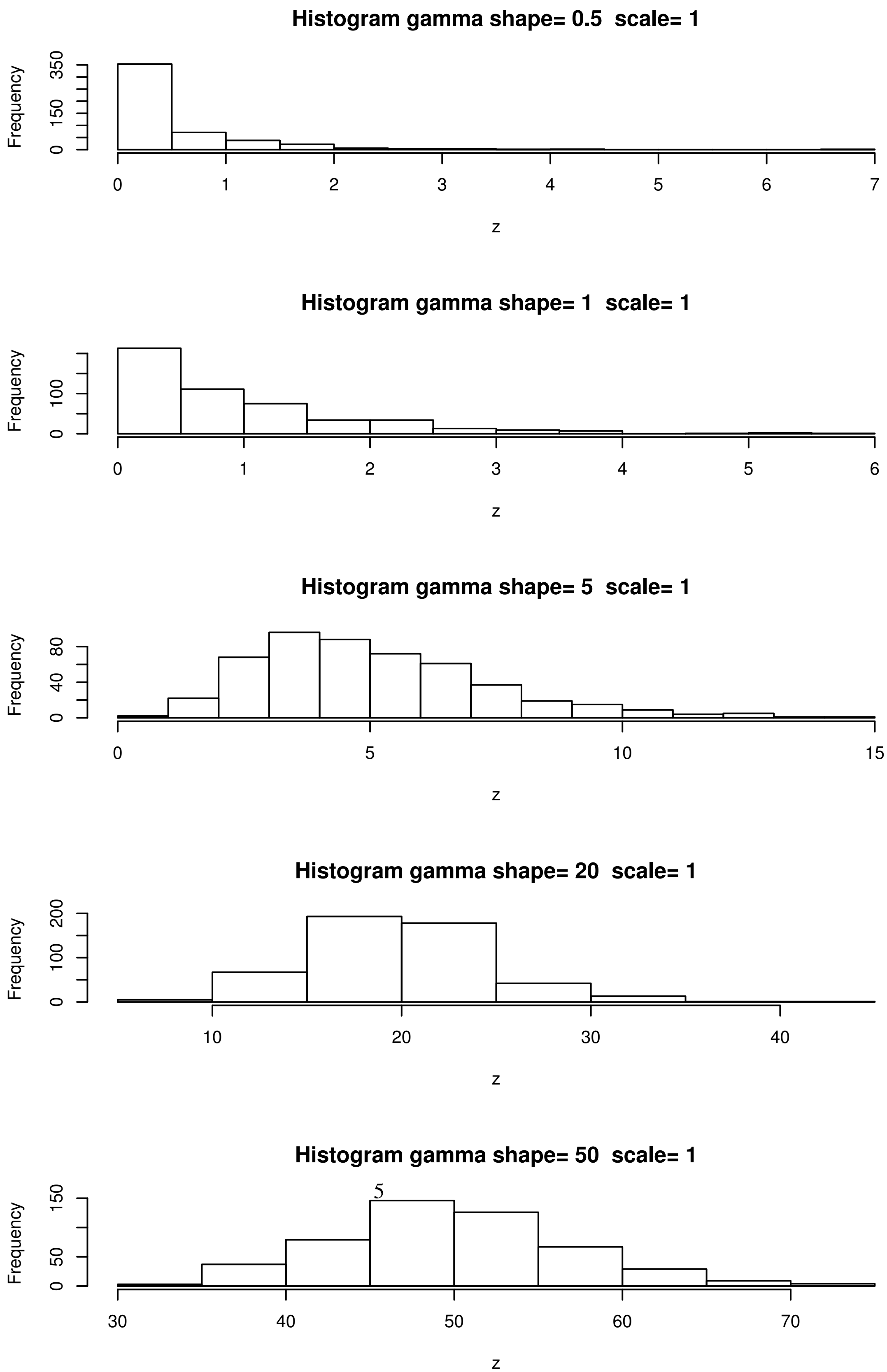
If shape=1, then this is an exponential distribution (because $\Gamma(1) = 1$). But, of course, the shape parameter can vary.

This has the same properties that the Poisson has (I just realized):

$E(v_i) = shape$

and

Figure 1: Gamma probability distribution



$Var(v_i) = shape$.

We made some progress there because it is easy to see what we have to do to make the random variable δ_i will have an expected value of 1.

Divide by shape!

So now the random variable that is being multiplied by e^{Xb} is $\delta_i = \frac{v_i}{shape}$. If v_i is Gamma(1,shape), and the "new randomness" is $\frac{v_i}{shape}$, that means that:

$E(\delta_i) = E(\frac{v_i}{shape}) = \frac{1}{shape}E(v_i) = \frac{shape}{shape} = 1$

and also

$V(\delta_i) = V(\frac{v_i}{shape}) = \frac{1}{shape^2}V(v_i) = \frac{shape}{shape^2} = \frac{1}{shape}$

If you go back and forth between books, you get a headache because no two book seem to treat these with exactly the same notation. But I'm pretty sure I've written it down correctly.

I wondered what a variable such $v_i/shape$ would look like, so I made a figure. See Figure 2.

So the "new input" for the model is:

$$exp(X_ib) * \delta_i = exp(X_ib + log(\delta_i)$$

$$= exp(X_ib + log(\frac{v_i}{shape}) = exp(X_ib + \frac{1}{shape}log(v_i))$$

And, if you are like me, and you wonder what the histogram of $log(\frac{v_i}{shape})$would look like, I'm proud to say I have that too in Figure 3. Please note that, as the math above indicates, the center of the distribution of $log(\frac{v_i}{shape})$ is centered on 0, and that as the shape increases, the variance of the distribution shrinks dramatically.

0.7 Estimating

I believe, if you wanted to, you could use a general purpose program for the estimation of "generalized linear mixed models" to estimate the b as well as the shape parameter. In doing so, you'd treat b as a "fixed" effect and then there is the additive random parameter that has the Gamma distribution. You'd want to estimate the shape parameter in order to find out how much uncertainty there is in the model.

However, there is a more elegant approach that works for this special case.

Here's the way we are going to think of this. There is a two-stage process. The error v_i (or δ_i) is chosen first. Then, knowing that, a draw from the Poisson distribution with mean $input_i * \delta_i$ is taken. Since, for each observation, a different δ_i is drawn, and then that is put together with the input before the y_i is drawn, it is possible for 2 cases with the exact same input e^{X_ib}could have y_i's drawn from different Poisson distributions.

Tools from probability theory will lead to an exact probability model for this.

If

$$Y \,|\, \delta_i \sim Poisson(input_i * \delta_i)$$

then

Figure 2: $v_i/shape$ where $v \sim Gamma(1, shape)$

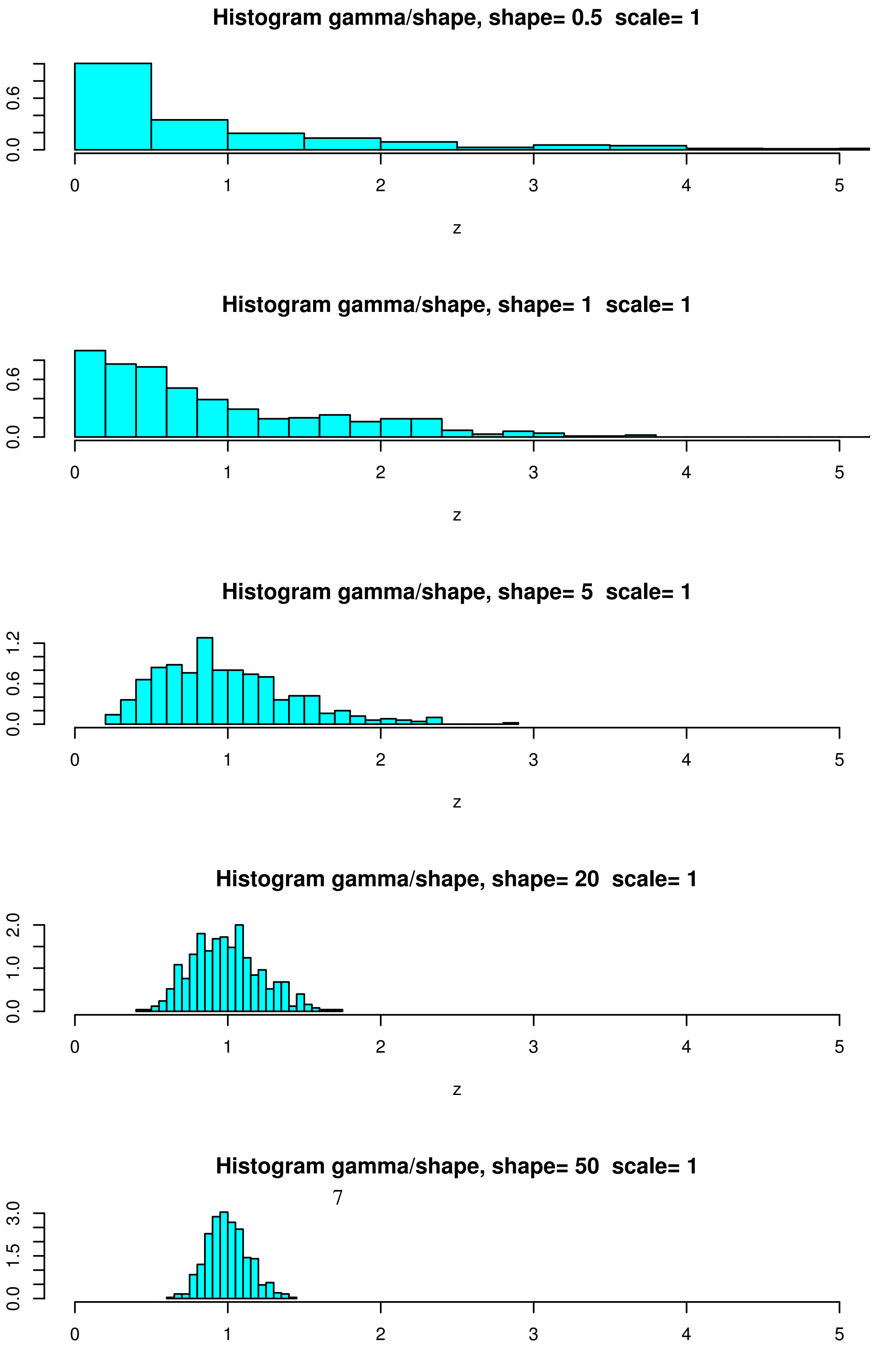
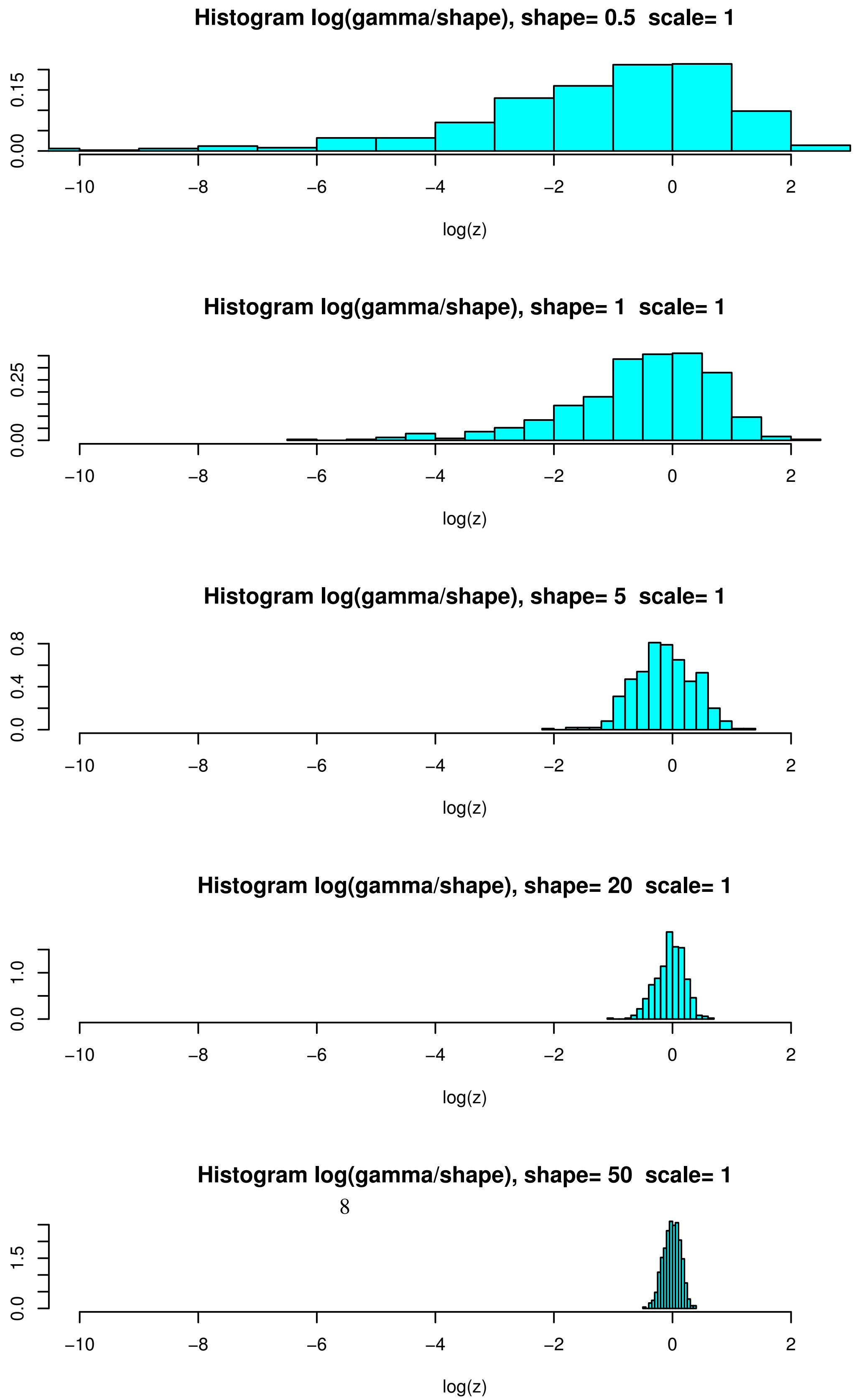


Figure 3: $log(v_i/shape)$ where $v \sim Gamma(1, shape)$



$$f_y(y|shape, input) = \frac{\Gamma(shape+y)}{\Gamma(shape)y!} \bullet \frac{input^y shape^{shape}}{(input+shape)^{shape+y}}$$

(Venables and Ripley, 4th ed, p. 206)

$$E(y_i) = input$$

$$Var(y_i) = input + input^2/shape$$

Note that if $shape = \infty$, then the variance of y_i is just $input_i$, meaning the original Poisson model is back! But for other values of the shape parameter, the variance of y_i is greater than in the Poisson model.

0.8 About that "shape" parameter

There are two ways to attack the problem. First, we might "guess" at a value for the "shape" parameter, and then estimate with the usual procedures for generalized linear models. If you want to try that, R has the procedure "glm" in which one can specify shape.

If we are interested in estimating *shape* as a part of the process, and that makes it interesting. R provides a procedure "glm.nb" which will do maximum likelihood to estimate the b's and the shape parameter. (In Venables & Ripley, p. 207, the "shape" parameter is called θ.