

The Gamma Distribution (and connections with Exponential, Geometric, Negative Binomial, Chi-Squared, and Normal Distributions).

Page 175 shows some pictures of Gamma distribution. To understand the parametrization, note that the density can be written in terms of constant times function of x/β . In other words, β is really a scale parameter, a bit like σ in the Normal distribution. The thing that really controls the shape of the Gamma is the parameter α . So if we set $\beta=1$ (so we can forget about it for a while), we have the "standard gamma" and we can see how the standard gamma depends on α . See Figure 4.26b on p 175.

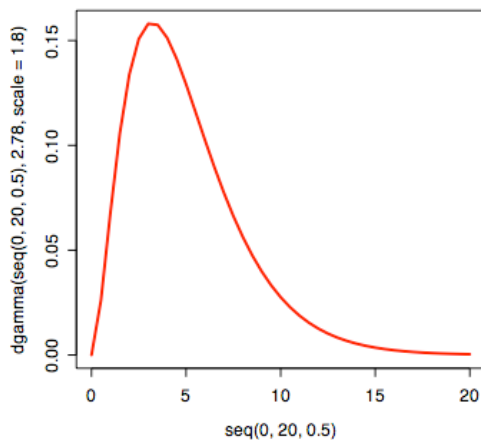
Note that when $\alpha=1$, the density looks like an exponential decay – in fact the standard gamma with $\alpha=1$ is the exponential distribution (defined on p 177 even though it is a special case of the gamma defined on p 175.)

The mean and the variance of the general gamma are $\alpha\beta$ and $\alpha\beta^2$, respectively. Of course the mean and SD are $\alpha\beta$ and $\alpha^{1/2}\beta$. Suppose you had some sample data for a measurement and the sample mean and SD were 5 and 3. Suppose you suspect the underlying population distribution is skewed and perhaps gamma, and you want to choose a reasonable α and β to see if it fits the data? One way to do this is to solve the following for α and β .

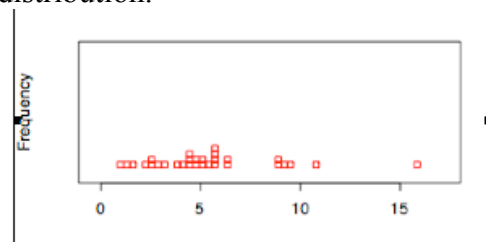
$$\begin{aligned}\alpha\beta &= 5 \\ \alpha^{1/2}\beta &= 3\end{aligned}$$

This produces $\alpha^{-1/2} = 3/5$ so $\alpha = 25/9 = 2.78$ and then $\beta = 5/2.78=1.80$

Here is the density function for the gamma RV with $\alpha = 2.78$ and $\beta = 1.80$



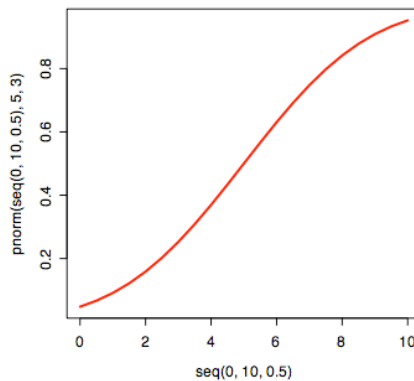
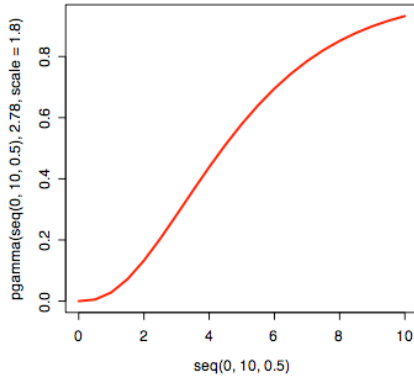
Here is a sample (n=30) from this same distribution.



Does it look like mean = 5 and SD = 3?

This is the cdf for the same density.

Note the shape – it is an S curve but the right skew turns up in the cdf as a gradual curve at the top of the cdf. Note the normal cdf – the second graph below.



The title of these notes was: " The Gamma Distribution (and connections with Exponential, Geometric, Negative Binomial, Chi-Squared, and Normal Distributions)." Now we outline these connections.

Recall that the Geometric distribution was a natural model for the number of Bernoulli trials that were failures before the first success occurred. Now to find the number of Bernoulli trials that were failures until the r th success occurred, we really have to wait for r Geometrics. That is, if Y is Negative Binomial with parameters r and p , and X is Geometric with parameter p , then

$$Y = \sum_{i=1}^r X_i \quad (*)$$

The sum of r Geometrics (with the same parameter p) is Negative Binomial.

Now recall that the **Geometric is the discrete analogue of the Exponential distribution**. It turns out that (*) is also true if the X_i are Exponential with parameter λ and Y is Gamma with parameters r and λ^{-1} . (The parametrization of the exponential on p 177 of your book defines λ in such a way that the mean is λ^{-1}). Look at the mean and SD of the Gamma. These can also be written
 mean = $r \lambda^{-1}$ and SD = $r^{1/2} \lambda^{-1}$
 in the parameterization that relates best to the exponential.

So we have a sort of box of links:

Geometric	Exponential
Neg Binomial	Gamma

The Central Limit Theorem (to come, p 239) says that averages (or sums) of IID (Independent, Identically Distributed) Random Variables is approximately Normal, and the approximation gets better and better as the things averaged gets large.

But as we have just shown Neg Binom and Gamma are IID Sums. As r gets large these two distributions become closer and closer to being Normal. In other words, **Negative Binomial and Gamma get close to Normal when r is large.**

Next we have a connection between the Gamma and Chi-Squared Distribution (p 179).

We already discussed that the Exponential was a special case of the Gamma. It is also true that the Chi-squared is a special case of the Gamma. The Chi-squared is usually described in terms of one parameter ν (sounds like gnu or new). It is called, for reasons we will understand shortly, the "degrees of freedom" of the Chi-squared distribution. The Gamma with $r=\nu/2$ and $\beta=2$ is exactly the chi-squared distribution on ν degrees of freedom. One way to think of the **Chi-squared (or χ^2) distribution is that it is the sum of squares of ν standard normal random variables** (this is not obvious – but just a helpful fact.) : That is,

$$\chi_{\nu}^2 = \sum_{i=1}^{\nu} z_i^2$$

where z_i are $N(0,1)$ and independent (so IID).

Even **the chi-squared distribution gets close to normal for large ν .**

We will skip Sections 4.5 and 4.6. On to Ch 5! This is where "inference" begins.

