Handout 11: Normal, Gamma, Beta

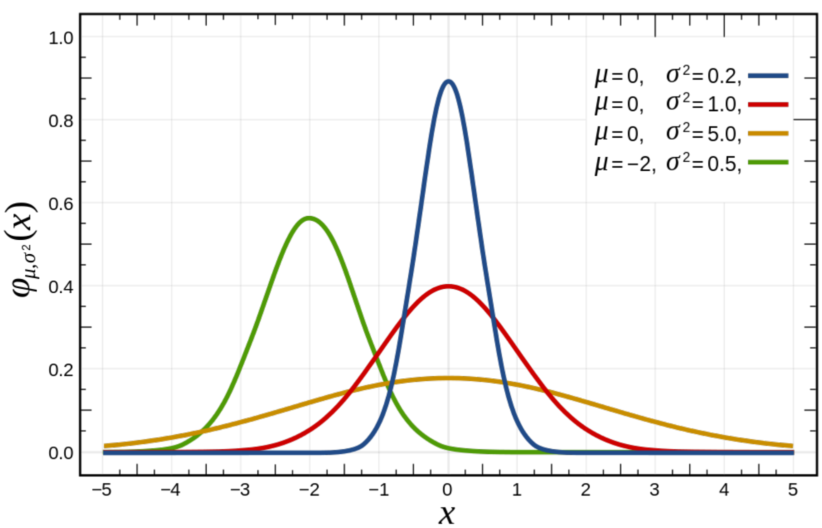
Today we will investigate three particularly well-known and useful continuous probability models.

The Normal Distribution

The most widely used continuous probability model is the so-called normal distribution. It is often called the "Gaussian distribution" in honor of the great German mathematician Karl Friedrich Gauss who introduced the "normal model" in a 1809 monograph as a plausible model for measurement error. Using this normal law as a generic model for errors in scientific experiments, Gauss formulated what is now known in the statistical literature as the "non-linear weighted least-squares" method. His work brought much attention to the model's utility in both theoretical and applied work. The normal distribution is entirely determined by two parameters, its mean p and its variance σ^2. We will therefore denote the normal distribution by $N(\mu, \sigma^2)$. The probability density is given by:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

The graph of which looks like:



The normal distribution occurs in a wide array of applications. Physical attributes like the heights and weights of members of certain subpopulations typically have distributions having bell-shaped densities. If we were to obtain a random sample of one hundred 20-year-old males attending a particular university, the histogram that we would construct for their heights or their weights (which graphically displays the relative frequency of occurrence of heights or weights in consecutive intervals of equal length) is very likely to look like an approximation to a bell-shaped curve. The measured characteristics of many natural populations have distributions that are approximately normal. As Gauss himself pointed out, the normal model often provides a good model for measurement or experimental errors. Another reason that the normal distribution arises in practice with some regularity is the fact that the distribution of numerical averages is often approximately normal, a consequence of the famous "Central Limit Theorem" which we will examine closely towards the end of the semester.

The distribution function of the normal model does not have a closed form expression for any value of σ and μ. To show that this is a valid probability distribution function we can use one of the most clever, in my opion, integration tricks.

Theorem 1 *We have:*

$$\int_{-\infty}^{\infty} e^{-x^2} dx = \sqrt{\pi}.$$

Proof. The first step of the trick is to square the integral, switch the dummy variable in one integral, and then collect all of the variable under a double integral sign:

$$\begin{aligned}\left(\int_{-\infty}^{\infty} e^{-x^2} dx\right)^2 &= \int_{-\infty}^{\infty} e^{-x^2} dx \cdot \int_{\infty}^{\infty} e^{-x^2} dx \\ &= \int_{-\infty}^{\infty} e^{-y^2} dy \cdot \int_{\infty}^{\infty} e^{-x^2} dx \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-(x^2+y^2)} dxdy\end{aligned}$$

Now, this is just an integral over all of $\mathbb{R}^2$. We can switch to polar coordinates ρ and θ, keeping in mind that $dxdy = \rho d\rho d\theta$. This is useful mathematically because otherwise we have no way of dealing with the squared term in the exponent:

$$\left(\int_{-\infty}^{\infty} e^{-x^2} dx\right)^2 = \int_{0}^{2\pi} \int_{0}^{\infty} e^{-(\rho^2)} \rho d\rho d\theta$$

Setting $r = \rho^2$ gives $dr/2 = \rho d\rho$:

$$\begin{aligned}\left(\int_{-\infty}^{\infty} e^{-x^2} dx\right)^2 &= \frac{1}{2} \int_{0}^{2\pi} \int_{0}^{\infty} e^{-r} \rho dr d\theta \\ &= \frac{2\pi}{2} \cdot -e^{-r}\Big|_{0}^{\infty} \\ &= \pi\end{aligned}$$

Taking the square root of both sides yields the result $\blacksquare$.

From this theorem it is an easy change of variables to show that the normal distribution is a proper probability distribution. In particular, it illustrates where the mysterious $\sqrt{\pi}$ comes from in the denominator of the normalizing constant.

Theorem 2 *If $X \sim N(0,1)$, the moment generating function $m_X(t)$ is equal to $e^{t^2/2}$.*

Proof. By definition:

$$\begin{aligned}
m_X(t) &= \int_{-\infty}^{\infty} e^{tx} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2+xt} dx \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-(z^2-2zt+t^2-t^2)} dx \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-(z-t)^2/2+t^2/2} dx \\
&= e^{t^2/2} \cdot \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-(z-t)^2/2} dx \\
&= e^{t^2/2}
\end{aligned}$$

The last step comes because the integral is the density of a $N(t,1)$ distributed random variable. The algebraic manipulations in the exponent comes from an application of completing the square ■.

One of the most important properties of the normal distribution is that it is closed under linear transformations. In particular, we have the following result:

Theorem 3 *If $X \sim N(\mu, \sigma^2)$, then $Y = aX + b$ we have $Y \sim N(a\mu + b, a^2\sigma^2)$.*

Theorem 4 *If $X \sim N(\mu, \sigma^2)$, the moment generating function $m_X(t)$ is equal to $exp\{\mu t + \frac{1}{2}\sigma^2 t^2\}$.*

Proof. Let $X = \mu + \sigma Z$ where $Z \sim N(0,1)$. We know that:

$$\begin{aligned}
m_X(t) &= e^{\mu t} m_Z(\sigma t) \\
&= e^{\mu t} \cdot e^{(\sigma t)^2/2} \\
&= e^{\mu t + t^2\sigma^2/2}
\end{aligned}$$

Which completes the result ■.

The Gamma Distribution

The so-called gamma integral, given by

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} \cdot e^{-x} dx$$

where α is a positive constant, is a rather famous object in the history of mathematics. The celebrity status of $\Gamma(\alpha)$ among integrals of real-valued functions is due to the fact that it arises in a wide variety of mathematical applications. The importance of being able to evaluate the integral became increasingly clear over time. However, one of the distinguishing features of the integral is that there is no closed-form expression for it, that is, no formula that gives the value of the integral for arbitrary values of α . Today, there are tables which allow you to obtain an excellent approximation to the value of Γ (α) for any value of α of interest.

While the value of $\Gamma(\alpha)$ cannot be obtained analytically for most values of α , it can be expressed in closed form when α is a positive integer. For example, $\Gamma(1)$ is simply equal to 1. For other positive integers α, the value of $\Gamma(\alpha)$ may be obtained from the following recursive relation satisfied by gamma integrals.

Theorem 5 *For any $\alpha > 1$, $\Gamma(\alpha) = (\alpha - 1) \cdot \Gamma(\alpha - 1)$. For a positive integer n, then, $\Gamma(n) = (n - 1)!$.*

Proof. We may prove the claim by applying integration by parts to $\Gamma(\alpha)$. We have that:

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-1} dx$$

Now, let $u = x^{\alpha\ -1}$ and $dv = e^{-x} dx$. It follows that $du = (\alpha - 1)x^{\alpha\ -2} dx$ and $v = -\ e^{-x}$. Since

$$\int_a^b u dv = uv|_a^b - \int_a^b v du.$$

we may write

$$\begin{aligned}\Gamma(\alpha) &= -x^{\alpha-1} e^{-x}|_0^\infty + \int_0^\infty (\alpha - 1) x^{\alpha-2} e^{-x} dx \\ &= (0 - 0) + (\alpha - 1)\Gamma(\alpha - 1) \\ &= (\alpha - 1)\Gamma(\alpha - 1)\end{aligned}$$

The second part of the statement follows by calculating $\Gamma(1)$:

$$\begin{aligned}\Gamma(1) &= \int_0^\infty x^{1-1} \cdot e^{-x} dx \\ &= \int_0^\infty e^{-x} dx \\ &= -e^{-x}\big|_0^\infty \\ &= -e^{-\infty} + e^0 \\ &= 1\end{aligned}$$

And appling the definition of the factorial function $\blacksquare$.

From this result, we then know that the following is a valid probability density function:

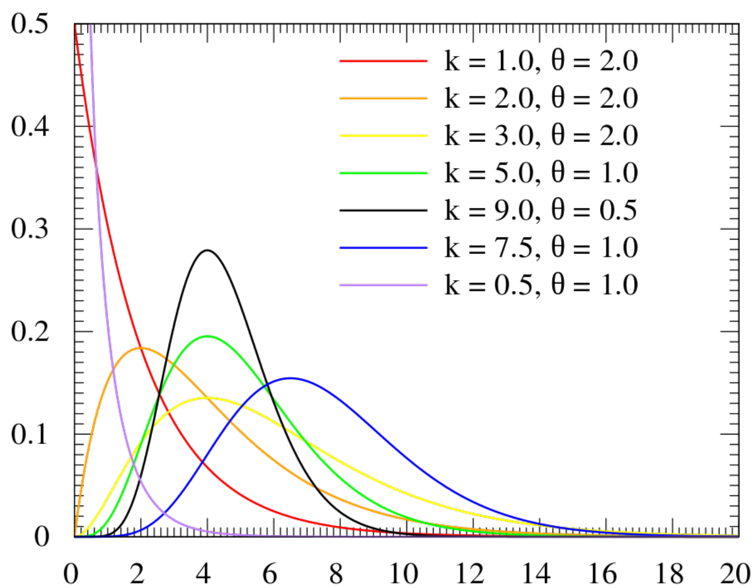
$$f(x) = \frac{1}{\Gamma(\alpha)} \cdot x^{\alpha-1} e^{-x}, \quad x > 0.$$

That is, it is a non-negative function on the interval $(0, \infty)$ that integrates to 1 over that interval. This is a simple version of the gamma model. It is a probability model indexed by one parameter, the "shape" parameter $\alpha > 0$.

The usual form of the gamma density includes a second parameter $\beta > 0$ called the "scale" parameter of the model. We will call the two-parameter model with density:

$$f(x) = \frac{1}{\Gamma(\alpha)\beta^\alpha} \cdot x^{\alpha-1} e^{-x/\beta}, \quad x > 0.$$

The *gamma model* and denote it by $Gamma(\alpha, \beta)$. Its distribution looks like this:



This model has found application in many and diverse studies, including lifetime modeling in reliability studies in engineering and in survival analysis in public health, the modeling of selected physical variables of interest in hydrology and meteorology and the modeling the waiting time until the occurrence of k events.

Theorem 6 *If $X \sim Gamma(\alpha, \beta)$, then $\mathbb{E}X = \alpha \cdot \beta$.*

Theorem 7 *If $X \sim Gamma(\alpha, \beta)$, then $V(X) = \alpha \cdot \beta^2$.*

Theorem 8 *If $X \sim Gamma(\alpha, \beta)$, then $m_X(t) = (1 - \beta t)^{-\alpha}$ for $t < 1/\beta$.*

We will prove these theorems all at once using the moment generating function.

Proof. By definition and basic arithmetic re-arranging:

$$\begin{aligned} m_X(t) &= \mathbb{E}e^{tX} \\ &= \int_0^\infty e^{tx} \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta} dx \\ &= \frac{1}{\Gamma(\alpha)\beta^\alpha} \int_0^\infty x^{\alpha-1} e^{-x(\beta^{-1}-t)} dx \\ &= \frac{1}{\Gamma(\alpha)\beta^\alpha} \Gamma(\alpha) \left(\frac{\beta}{1-\beta t}\right)^\alpha \cdot \int_0^\infty \frac{1}{\Gamma(\alpha)\left(\frac{\beta}{1-\beta t}\right)^\alpha} \cdot x^{\alpha-1} e^{-x/\left(\frac{\beta}{1-\beta t}\right)} dx \end{aligned}$$

The whole part inside of the integral is the density of a $\Gamma(\alpha, \beta/(1-\beta t))$ variable for $t < 1/\beta$, and therefore integrates to 1. This gives:

$$\begin{aligned} m_X(t) &= \frac{1}{\Gamma(\alpha)\beta^\alpha} \Gamma(\alpha) \left(\frac{\beta}{1-\beta t}\right)^\alpha \\ &= \left(\frac{1}{1-\beta t}\right)^\alpha \\ &= (1-\beta t)^{-\alpha} \end{aligned}$$

Taking derivatives we have:

$$\begin{aligned} \frac{\partial}{\partial t} m_X(t) &= \alpha\beta(1-\beta t)^{-\alpha-1} \\ \frac{\partial^2}{\partial^2 t} m_X(t) &= \alpha(1+\alpha)\beta^2(1-\beta t)^{-\alpha-2} \end{aligned}$$

Setting equal to $t = 0$, we get:

$$\begin{aligned} \mu_1' &= \alpha\beta \\ \mu_2' &= \alpha(1+\alpha)\beta^2 \end{aligned}$$

Which gives the mean and variance as:

$$\begin{aligned}
\mathbb{E}X &= \alpha\beta \\
\mathbb{E}(X-\mathbb{E}X)^2 &= \mathbb{E}(X)^2 - (\mathbb{E}X)^2 \\
&= \mu_2' - \mu_1^2 \\
&= \alpha(1+\alpha)\beta^2 - \alpha^2\beta^2 \\
&= \alpha\beta^2 + \alpha^2\beta^2 - \alpha^2\beta^2 \\
&= \alpha\beta^2
\end{aligned}$$

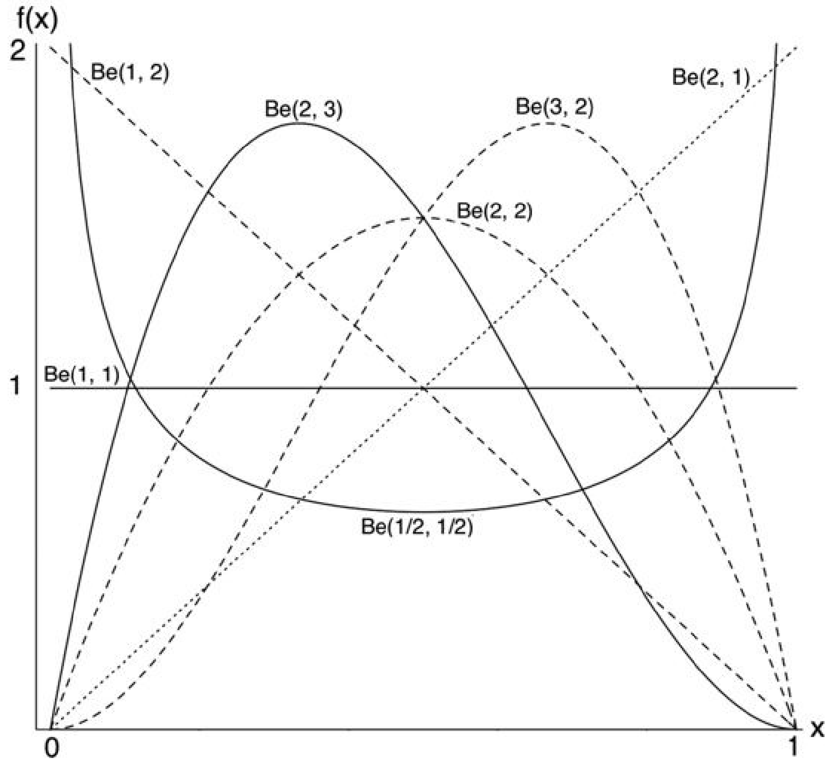
And this completes all of the results ■.

The Beta Distribution

We have seen the utility of a continuous distribution of a uniform set. A somewhat broader view of the U(0,1) distribution is as a model for a randomly chosen proportion. After all, a random proportion will take values in the interval (0, 1), and if the proportion drawn was thought to be equally likely to fall in any subinterval of (0,1) of a given length, the model would clearly be appropriate. But the uniform distribution is just one member of a large and flexible family of models for random proportions. The Beta family is indexed by a pair of positive parameters α and β . The model we will denote by $Beta(\alpha, \beta)$ has density function:

$$f(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}, \quad x \in [0,1].$$

The graph for several values of α and β looks like this:



We will state the following result without proof.

Theorem 9 *If* $X \sim Beta(\alpha, \beta)$ *then* $\mathbb{E}X = \frac{\alpha}{\alpha+\beta}$ *and:*

$$Var(X) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$$

You might wonder where you would encounter the need for modeling a random probability p. One place in which it arises quite naturally is in the area of Bayesian inference. We will see this application in the upcoming weeks.