Section 4: Normal, Beta, and Gamma Distributions

Daniel Woods

Professor Byron Hoy

Continuous random variable probability distributions can take on certain shapes depending on the probability of the variable being measured. One can determine which class of distribution that a particular distribution belongs to by checking its probability density function (pdf) and comparing it to the formulas and preconditions that exist for each class. One can also look at the shape of the distribution and determine, with a good degree of accuracy, which class it belongs to. Probability distributions that can be attributed to one of these classes have the benefit of being able to solve for certain things more easily.

NORMAL PROBABILITY DISTRIBUTION

The Normal probability distribution is the most common shape that is most often thought of as “the bell curve”. This probability distribution has most of its probability occurring right at the mean. If a distribution is a normal distribution, the density function of Y is:

$$f(y) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(y-\mu)^2/(2a^2)} \qquad -\infty < y < \infty$$

Unfortunately, with a normal distribution there is no closed form method of finding the integral of an equation of that form. This necessitates the use of a table or computer assistance to find the area (probability) of a certain range of Y (continuous probability variable). To find areas under a normal distribution curve, we utilize the concept of z-scores. Z-scores are calculated as follows:

$$Z = \frac{Y-\mu}{\sigma}$$

This equation will give you the area of probability in units of standard deviations. For example, if we would like to find the probability that $40 < Y < 60$ on a distribution where the mean is 70 and the standard deviation is 8, then we would first calculate the z scores as follows:

$$z = \frac{40-70}{8} = -3.75 \quad and \quad z = \frac{60-70}{8} = -1.25$$

This means that our Y values are 6 and 2 standard deviations below the mean respectively. If we consult a z score table, we will find that the areas correspond to those z scores, (that is the area between them and Y=0) are .00009 corresponding to -3.75 and .10565 corresponding to -1.25. Because these areas are corresponding to the left, we must calculate the area by subtracting .00009 from .10565 which equals .10556, or a 10.6 percent change of Y being between 40 and 60 on a standard curve.

Examples of data sets that measure things like height, IQ, and other things where there is a strong average population.

GAMMA PROBABILITY DISTRIBUTION

A gamma probability distribution is said to be skewed to the right. This is seen much less commonly than a normal probability distribution. Most of the area of the curve is located on the left side of the curve, and closer to 0. A gamma distribution function will follow the following form:

$$\begin{cases} f(y) = \dfrac{y^{\alpha-1}e^{-y/\beta}}{\beta^{\alpha}\Gamma(a)}, 0 \le y < \infty \\ 0, elsewhere \end{cases} \quad where \quad \Gamma(a) = \int_0^{\infty} y^{\alpha-1}e^{-y}\,dy$$

The value of $\Gamma(n)$ can be simplified to $\Gamma(n) = (n-1)!$, if *n* is an integer. $\Gamma(\alpha)$ is called the gamma function. α is the variable that significantly changes the shape of the curve of the gamma function, and it is known as the shape parameter. β is the parameter that is multiplied by α and is referred to as the scale parameter because it produces a curve of the same shape, but on a different scale.

When α is not an integer and 0 < c < d < ∞, then it is not possible to calculate the area under the curve by evaluating the integral in conventional ways, except when $\alpha = 1$. You must use a computer or a chart, just as in a normal distribution. The expression of area of a gamma distribution between the points c and d is given by the expression:

$$\int_c^d \frac{y^{\alpha-1}e^{-y/\beta}}{\beta^{\alpha}\Gamma(a)}\,dy$$

While the area of a gamma function is hard to calculate, the mean and variance are easier to calculate:

$$\mu = \alpha\beta \quad and \quad \sigma^2 = \alpha\beta^2$$

A random variable with a chi-square distribution is called a chi-square random variable if the random variable Y is a gamma-distributed random variable where $\alpha = \frac{v}{2}\ and\ \beta = 2$ where v is a positive integer. V is referred to as degrees of freedom. The mean and variance of a chi-square random variable can be calculated as follows:

$$\mu = v\ and\ \sigma^2 = 2v$$

Chi-square distributions are preferable to standard gamma distributions because it is much easier to find tables for probability of chi-square distributions than regular gamma distributions. If the α of a gamma distribution can be represented as $\alpha = n/2$, then β can be represented by $\beta = 2Y/\beta$ and is a chi-square distribution.

If a gamma density function has parameter $\alpha = 1$, then that function is said to be an exponential density distribution. A function can be classified as an exponential distribution if $\beta > 0$ and the equation matches the following:

$$f(y) = \begin{cases} \dfrac{1}{\beta}e^{-y/\beta}, 0 \le y < \infty \\ 0, elsewhere \end{cases}$$

BETA PROBABILITY DISTRIBUTION

A beta density distribution can take on a wide variety of shapes and cannot easily be determined by looking at the curve. These distributions exist only if the density function is:

$$f(y) = \begin{cases} \dfrac{y^{\alpha-1}(1-y)^{\beta-1}}{B(\alpha,\beta)}, 0 \le y \le 1 \\ 0, elsewhere \end{cases}$$

Because y is defined as being between or equal to 0 and 1, you may define a new variable y* as:
$y* = (y-c)/(d-c)$ if $c \le y \le d$.

The cumulative distribution function of a beta random variable is known as the incomplete beta function. Values of the cumulative distribution function of a beta distribution can be found on a table of $\alpha\ and\ \beta$ values. Using a computer is the most efficient way to get probabilities from a beta distribution