Ivan Wang

Probability and Applied Statistics

December 14, 2023

# Normal, Beta, and Gamma Distribution Paper

## What is a distribution in statistics? *[1]*

A **distribution** in statistics usually refers to how the values of a random variable are distributed across a range of values. A function's distribution shows how often certain values occur. It is important to note that a" distribution of an event consists not only of the input values that can be observed, but is made up of all possible values" [1]. Any distribution will have an associated graph which represents that distribution, which includes all possible values. This means that the total area under the curve is always 1, or 100%.

Distributions are everywhere, and are used to model real world phenomena, like predicting outcomes, financial modeling, or voting patterns. The ability to apply statistical distributions is a powerful tool across all industries which allows us to understand the world around us.

## Normal Distribution *[2, 3]*

The **normal distribution**, also known as the **bell curve**, is the most widely known of the distributions. The shape is perfectly symmetrical, with its mean in the middle. This means that the mean, median, and mode of the dataset would all be the same or very close to each other. The highest point (most frequent value) of the bell-shaped curve represents the mean/median/mode of the data. Half of the values are below the mean and half are above the mean. The shape of a normal distribution is determined by two parameters: the mean (μ) and the standard deviation (σ). The mean is the central tendency, and the standard deviation reflects the dispersion or spread of the distribution. This is shown in its formula:

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

Due to a normal distribution's unique shape, we can apply the **empirical rule** which states that around 68% of values are within 1 standard deviation from the mean, 95% of values are within 2 standard deviations, and 99.7% of values are within 3. The empirical rule is a quick and easy way to understand your data and check for outliers in a dataset.

The normal distribution is known to be so useful since many phenomena are roughly normally distributed. Some examples include height, SAT Scores, and blood pressure readings. Often, methods created using normal distribution theory work well, even in non-normal distributions [2].

However, the normal distribution has its limitations. The model is not applicable to skewed distributions or data with extreme outliers. Also, over-relying on the normal distribution can lead to incorrect conclusions, especially in datasets that do not follow this pattern.

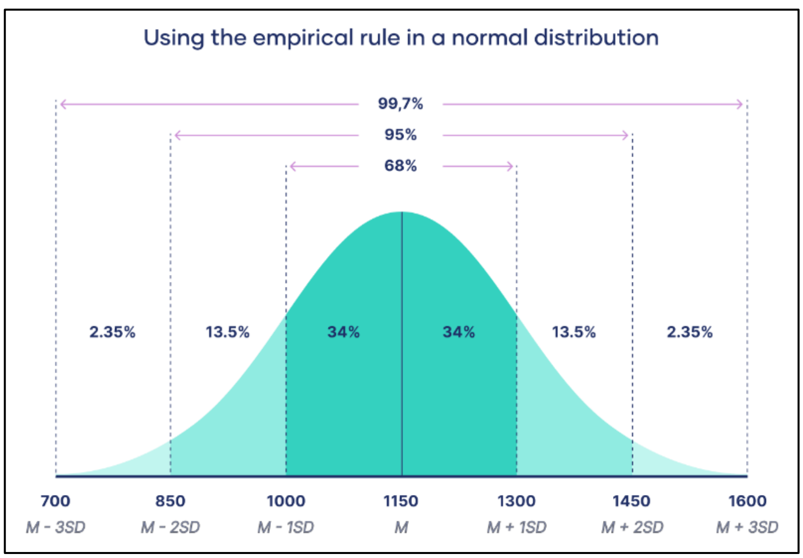



*Figure 1. A diagram of the empirical rule in a normal distribution from Scribbr [3].*

## **Probability Density Functions** *[4]*

To understand the following distributions, we must also understand **probability density functions (PDF).** Unlike with discrete variables where we can calculate the probability of exact values, PDFs show the probability of a variable falling between a certain interval. It is important to note that the probability of the variable being any single value is zero since there are infinite possibilities on a continuous interval.

## **Beta Distribution** *[5, 6, 7, 8]*

The **beta distribution** is for modeling events that fall in a specific range, usually between 0 and 1. This makes it ideal for representing anything between 0 and 1: probabilities, proportions, rates, etc. The distribution is defined by two parameters, alpha (α) and beta (β), which determine the distribution's shape. Alpha and beta represent the number of "successes" and "failures" of a particular event respectively. A larger alpha value means that the distribution is skewed toward 1 and a larger beta value would skew the distribution toward 0.

The shape of the distribution is ultimately defined by how both alpha and beta interact with each other. For example, if they are equal, the distribution will be symmetric. If one is greater than the other, the distribution will be skewed toward 0 or 1.

The PDF of the Beta Distribution:

$$f(x|\alpha,\beta) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)}$$

The formula for the Beta Distribution:

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

Unlike the normal distribution, the beta distribution can be symmetric, skewed, U-shaped, or J-shaped, depending on the specific use case. This versatility makes it useful in many settings and in modeling real world phenomena. In an example by David Robinson on Stats Stack Exchange, he explains beta distribution as a "probabilistic distribution of *probabilities*: the case where we don't know what a probability is in advance, but we have some reasonable guesses". He uses baseball batting averages as an example. A beta distribution would tell us the reasonable range of possible batting averages, which is what he means by a "probabilistic distribution of probabilities" [6].

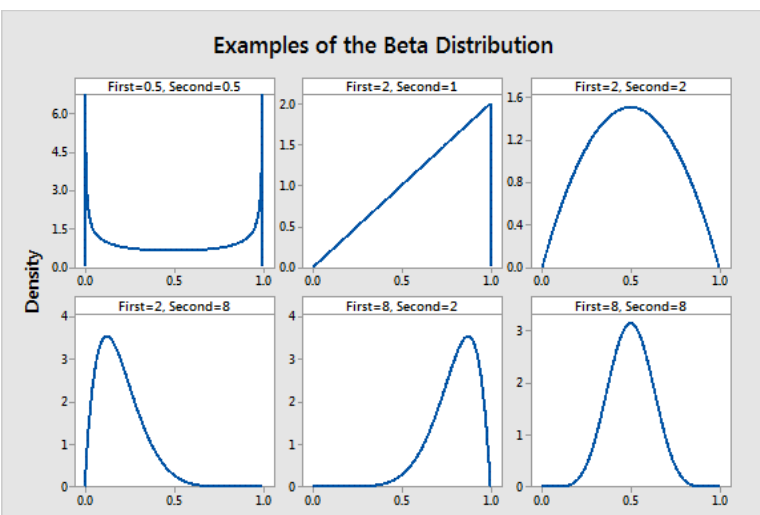



*Figure 2. Examples of Beta Distributions by Statistics By Jim [8].*

## **Gamma Distribution** *[9, 10]*

The Gamma distribution is another important continuous probability distribution used in statistical modeling. Unlike the Normal and Beta distributions, the Gamma distribution is used primarily for modeling waiting times or the time until an event occurs. It is particularly useful in fields such as queuing theory, reliability engineering, and weather forecasting.

The Gamma distribution is characterized by two parameters: the shape parameter (k) and the scale parameter (θ). These parameters determine the shape and scale of the distribution, respectively. The shape parameter k is often referred to as the order of the distribution. When k is an integer, the Gamma distribution represents the sum of k exponentially distributed random variables, each with mean θ.

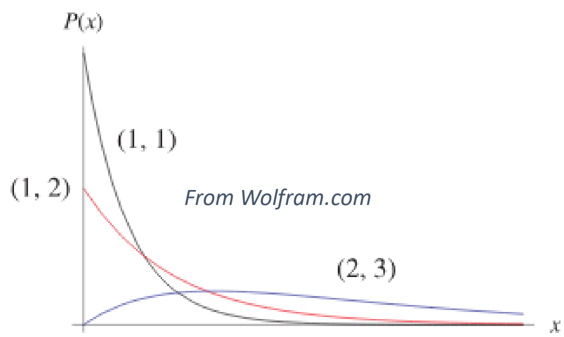
The PDF of the Gamma Distribution:

$$f(x|k,\theta) = \frac{1}{\theta^k \Gamma(k)} x^{k-1} e^{-\frac{x}{\theta}}$$

The CDF of the Gamma Distribution:

$$F(x|k,\theta) = \frac{1}{\Gamma(k)} \gamma\left(k, \frac{x}{\theta}\right)$$

The asymmetry of the beta distribution sets it apart from other distributions like the normal and beta distributions. The Gamma distribution is skewed to the right, meaning it has a long tail on the right side. Because of this characteristic, it can be used to model quantities like time that cannot be negative.




There are many uses for Gamma distributions. For example, in finance, it can model the time until a financial goal is reached. It can forecast the total amount of precipitation that will build up in a reservoir over time in meteorology. In health care, it can simulate the amount of time it will take until a specific event, like the failure of a medical device.

## **Conclusion**

Each of the three distributions—the normal, beta, and gamma—have special qualities and uses in statistical modeling. The Normal distribution, with its symmetric bell shape, is widely used for occurrences that are evenly distributed around a mean. To model proportions and probabilities, the Beta distribution, which is defined in the interval 0 to 1, is perfect. Because of its right-skewed shape, the Gamma distribution is especially helpful in modeling waiting times and the elapsed time until events. For those who work with statistics, comprehending these distributions and their uses is essential since they offer effective tools for examining and predicting real-world occurrences.

1. https://365datascience.com/tutorials/statistics-tutorials/distribution-in-statistics/
2. https://www3.nd.edu/~rwilliam/stats1/x21.pdf
3. https://www.scribbr.com/statistics/normal-distribution/#:~:text=In%20a%20normal%20distribution%2C%20data%20are%20symmetrically%20distributed%20with%20no,same%20in%20a%20normal%20distribution.
4. https://online.stat.psu.edu/stat414/lesson/14/14.1
5. https://web.stanford.edu/class/archive/cs/cs109/cs109.1166/pdfs/22%20Beta.pdf
6. https://byjus.com/maths/beta-distribution/
7. https://stats.stackexchange.com/questions/47771/what-is-the-intuition-behind-beta-distribution
8. https://statisticsbyjim.com/probability/beta-distribution/
9. https://mathworld.wolfram.com/GammaDistribution.html