

In This Issue

1. Learn the basics of PDF, CDF, and ICDF
2. Learn how to read and interpret the CDF or S-Curve based on its four moments

“How do you interpret a probability distribution histogram or cumulative distribution chart (S-Curve)?”

Contact Us

Real Options Valuation, Inc.

4101F Dublin Blvd., Ste. 425,
Dublin, California 94568 U.S.A.

admin@realoptionsvaluation.com
www.realoptionsvaluation.com
www.rovusa.com

The Basics of PDF, CDF, and ICDF

This paper briefly explains the probability density function (PDF) for continuous distributions, which is also called the probability mass function (PMF) for discrete distributions (we use these terms interchangeably), where given some distribution and its parameters, we can determine the probability of occurrence given some outcome or random variable x . In addition, the cumulative distribution function (CDF) can also be computed, which is the sum of the PDF values up to this x value. Finally, the inverse cumulative distribution function (ICDF) is used to compute the value x given the cumulative probability of occurrence.

In mathematics and Monte Carlo risk simulation, a probability density function (PDF) represents a **continuous** probability distribution in terms of integrals. If a probability distribution has a density of $f(x)$, then intuitively the infinitesimal interval of $[x, x + dx]$ has a probability of $f(x) dx$. The PDF, therefore, can be seen as a smoothed version of a probability histogram; that is, by providing an empirically large sample of a continuous random variable repeatedly, the histogram using very narrow ranges will resemble the random variable's PDF.

The probability of the interval between $[a, b]$ is given by $\int_a^b f(x)dx$, which means that the total integral of the function f must be 1.0.

It is a common mistake to think of $f(a)$ as the probability of a . This is incorrect. In fact, $f(a)$ can sometimes be larger than 1—consider a uniform distribution between 0.0 and 0.5. The random variable x within this distribution will have $f(x)$ greater than 1. The probability, in reality, is the function $f(x)dx$ discussed previously, where dx is an infinitesimal amount.

The cumulative distribution function (CDF) is denoted as $F(x) = P(X \leq x)$ indicating the probability of X taking on a less than or equal value to x . Every CDF is monotonically increasing, is continuous from the right, and at the limits has the following properties:

$$\lim_{x \rightarrow -\infty} F(x) = 0 \text{ and } \lim_{x \rightarrow +\infty} F(x) = 1.$$

Further, the CDF is related to the PDF by $F(b) - F(a) = P(a \leq X \leq b) = \int_a^b f(x)dx$, where the

PDF function f is the derivative of the CDF function F . In probability theory, a probability mass function, or PMF, gives the probability that a **discrete** random variable is exactly equal to some value. The PMF differs from the PDF in that the values of the latter, defined only for continuous random variables, are not probabilities; rather, its integral over a set of possible values of the random variable is a probability. A random variable is discrete if its probability distribution is discrete and can be characterized by a PMF.

Therefore, X is a discrete random variable if $\sum_u P(X = u) = 1$ as u runs through all possible values of the random variable X .

Here are some tips to help decipher the characteristics of a distribution when looking at different PDF and CDF charts:

- For each distribution, a continuous distribution's PDF is shown as area charts (Figure 1) whereas discrete distributions' PMF are shown as bar charts (Figure 2).
- If the distribution can only take a single shape (e.g., normal distributions are always bell shaped, with the only difference being the central tendency measured by the mean and the spread measured by the standard deviation), then typically only one PDF area chart will be shown with an overlay PDF line chart (Figure 3) showing the effects of various parameters on the distribution.
- The CDF charts, or S-Curves, are shown as line charts (Figure 4), and sometimes as bar graphs.
- The central tendency of a distribution (e.g., the mean of a normal distribution) is its central location (Figure 3).
- Multiple area charts and line charts will be shown (e.g., beta distribution) if the distribution can take on multiple shapes (e.g., the beta distribution is a uniform distribution when $\alpha = \beta = 1$; a parabolic distribution when $\alpha = \beta = 2$; a triangular distribution when $\alpha = 1$ and $\beta = 2$, or vice versa; a positively skewed distribution when $\alpha = 2$ and $\beta = 5$, and so forth). In this case, you will see multiple area charts and line charts (Figure 5).
- The starting point of the distribution is sometimes its minimum parameter (e.g., parabolic, triangular, uniform, arcsine, etc.) or its location parameter (e.g., the beta distribution's starting location is 0, but a beta 4 distribution's starting point is the location parameter; Figure 5's first chart shows a beta 4 distribution with location = 10, its starting point on the x -axis).
- The ending point of the distribution is sometimes its maximum parameter (e.g., parabolic, triangular, uniform, arcsine, etc.) or its natural maximum multiplied by the factor parameter shifted by a location parameter (e.g., the original beta distribution has a minimum of 0 and maximum value of 1, but a beta 4 distribution with location = 10 and factor = 2 indicates that the shifted starting point is 10 and ending point is 11, and its width of 1 is multiplied by a factor of 2, which means that the beta 4 distribution now will have an ending value of 12, as shown in Figure 5's first chart).
- Interactions between parameters are sometimes evident. For example, in the beta 4 distribution, if the $\alpha = \beta$, the distribution is symmetrical, whereas it is more positively skewed the greater the difference between $\beta - \alpha$, and the more negatively skewed, the greater the difference between $\alpha - \beta$ (Figure 5's second chart).
- Sometimes a distribution's PDF is shaped by two or three parameters called *shape* and *scale*. For instance, the Laplace distribution has two input parameters, α location and β scale, where α indicates the central tendency of the distribution (like the mean in a normal distribution) and β indicates the spread from the mean (like the standard deviation in a normal distribution).
- The narrower the PDF (Figure 3's normal distribution with a mean of 10 and standard deviation of 2), the steeper the CDF S-Curve looks (Figure 4), and the smaller the width on the CDF curve.
- A 45-degree straight line CDF (an imaginary straight line connecting the starting and ending points of the CDF) indicates a uniform distribution; an S-Curve CDF with equal amounts above and below the 45-degree straight line indicates a symmetrical and somewhat like a bell- or mound-shaped curve; a CDF completely curved above the 45-degree line indicates a positively skewed distribution, while a CDF completely curved below the 45-degree line indicates a negatively skewed distribution, as shown in Figure 5's fourth chart.
- A CDF line that looks identical in shape but shifted to the right or left indicates the same distribution but shifted by some location, and a CDF line that starts from the same point but is pulled both to the left and right indicates a multiplicative effect on the distribution such as a factor multiplication, as shown in the last two graphs in Figure 5.
- An almost vertical CDF indicates a high kurtosis distribution with fat tails and where the center of the distribution is pulled up (e.g., see the Cauchy distribution) versus a relatively flat CDF, a very wide and perhaps flat-tailed distribution is indicated.
- Some discrete distributions can be approximated by a continuous distribution if its number of trials is sufficiently large and its probability of success and failure are fairly symmetrical (e.g., see the binomial and negative binomial distributions). For instance, with a small number of trials and a low probability of success, the binomial distribution is positively skewed, whereas it approaches a symmetrical normal distribution when the number of trials is high and the probability of success is around 0.50.
- Many distributions are both flexible and interchangeable—refer to the details of each distribution in the Risk Simulator user manual—e.g., binomial is Bernoulli repeated multiple times; arcsine and parabolic are special cases of beta; Pascal is a shifted negative binomial; binomial and Poisson approach normal at the limit; chi-square is the squared sum of multiple normal; Erlang is a special case of gamma; exponential is the inverse of the Poisson but on a continuous basis; F is the ratio of two chi-squares; gamma is related to the lognormal, exponential, Pascal, Erlang, Poisson, and chi-square distributions; Laplace comprises two exponential distributions in one; the log of a lognormal approaches normal; the sum of multiple discrete uniforms approach normal; Pearson V is the inverse of gamma; Pearson VI is the ratio of two gammas; PERT is a modified beta; a large degree of freedom T approaches normal; Rayleigh is a modified Weibull; and so forth.

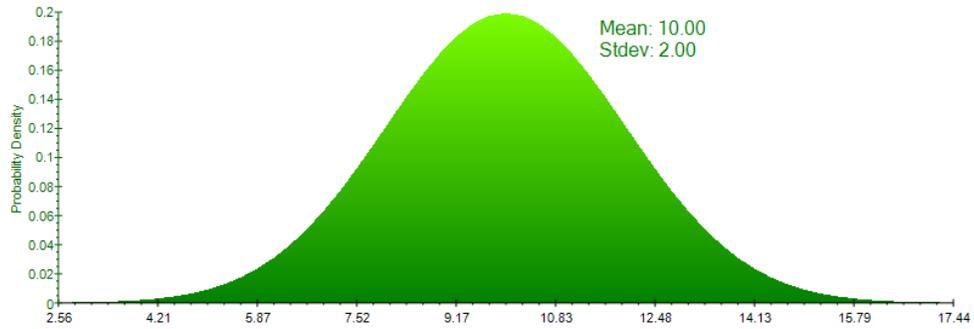


Figure 1. Continuous PDF (Area Chart)

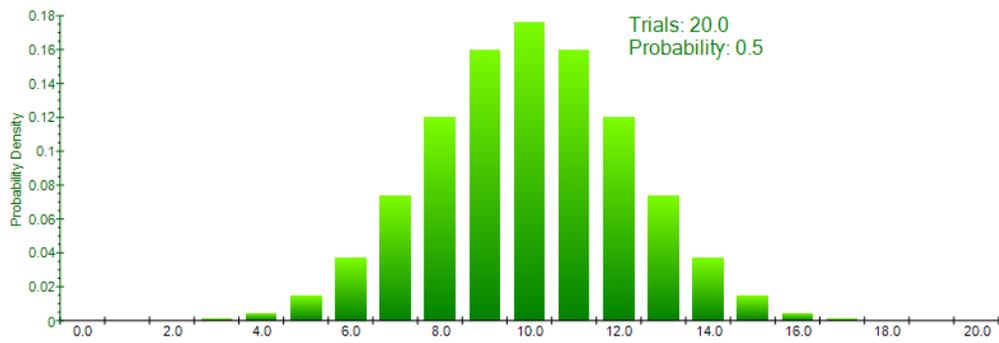


Figure 2. Discrete PMF (Bar Chart)

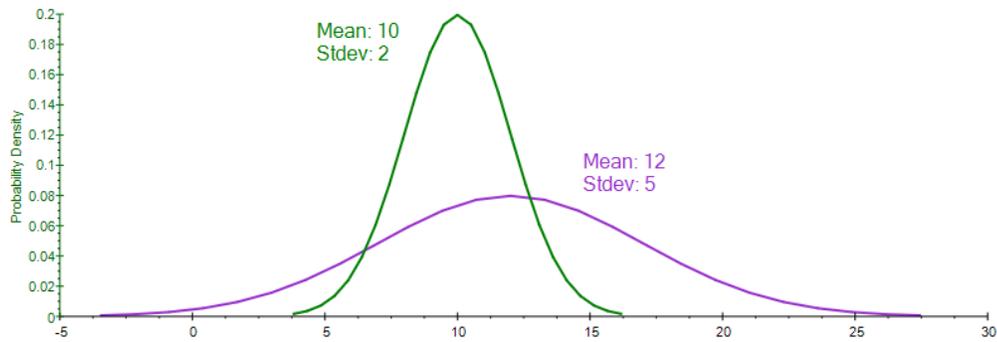


Figure 3. Multiple Continuous PDF Overlay Charts

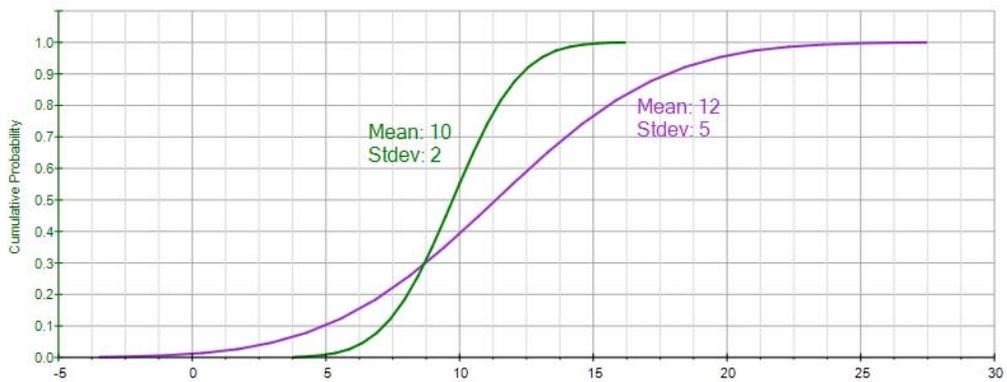


Figure 4. CDF Overlay Charts

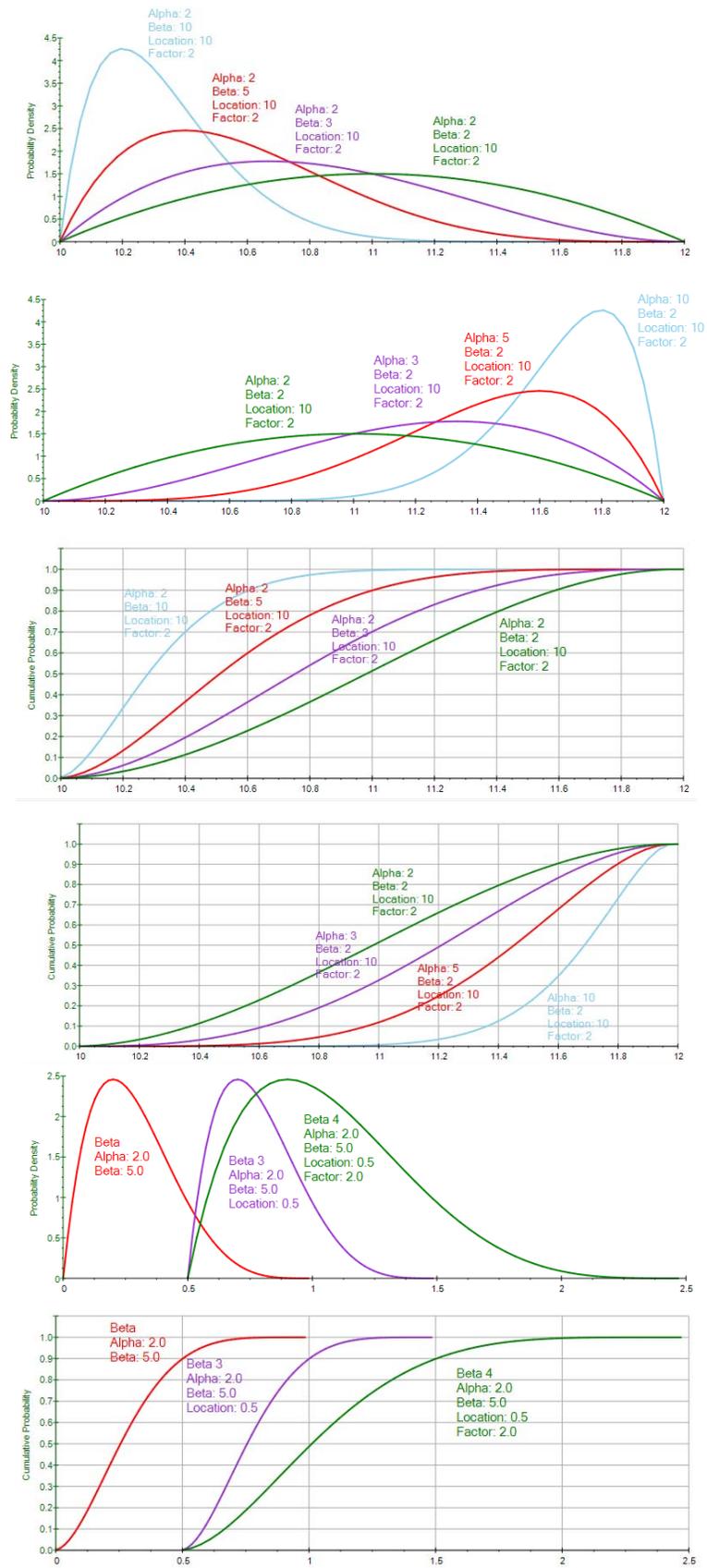


Figure 5. PDF and CDF Characteristics