

Chapter Four

Commonly Used Probability Distributions

4.1 Introduction

As we have seen in Charter 3, a probability density function (*pdf*) or a cumulative distribution function (*cdf*) can completely describe a random variable. Because of the physical phenomena or data patterns, different variables may follow different probability distributions. In this chapter, we will introduce several commonly used distributions in engineering analysis and design. We will also briefly discuss how to use Matlab Statistical Toolbox to calculate a *cdf* and *pdf*.

4.2 Normal Distribution

Normal distribution is the most commonly used distribution. The mean \mathbf{m}_x and standard deviation \mathbf{s}_x are the two parameters to determine a normal distribution. The notation of a normally distributed random variable X is $X \sim N(\mathbf{m}_x, \mathbf{s}_x)$.

The *pdf* of a normal distribution is expressed as

$$f_x(x) = \frac{1}{\sqrt{2\pi}\mathbf{s}_x} \exp\left[-\frac{1}{2}\left(\frac{x - \mathbf{m}_x}{\mathbf{s}_x}\right)^2\right], \quad -\infty < x < +\infty \quad (4.1)$$

The *cdf* is given by

$$F_x(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}\mathbf{s}_x} \exp\left[-\frac{1}{2}\left(\frac{x - \mathbf{m}_x}{\mathbf{s}_x}\right)^2\right] dx, \quad -\infty < x < +\infty \quad (4.2)$$

The *pdf* and *cdf* of a random variable $X \sim N(10,1)$ are plotted in Fig. 4.1 and 4.2 respectively. It is seen that a normal distribution is symmetric about its mean.

A random variable $X \sim N(\mathbf{m}_x, \mathbf{s}_x)$ can be transformed into a standard normal variable Z with the following equation.

$$Z = \frac{X - \mathbf{m}_x}{\mathbf{s}_x} \quad (4.3)$$

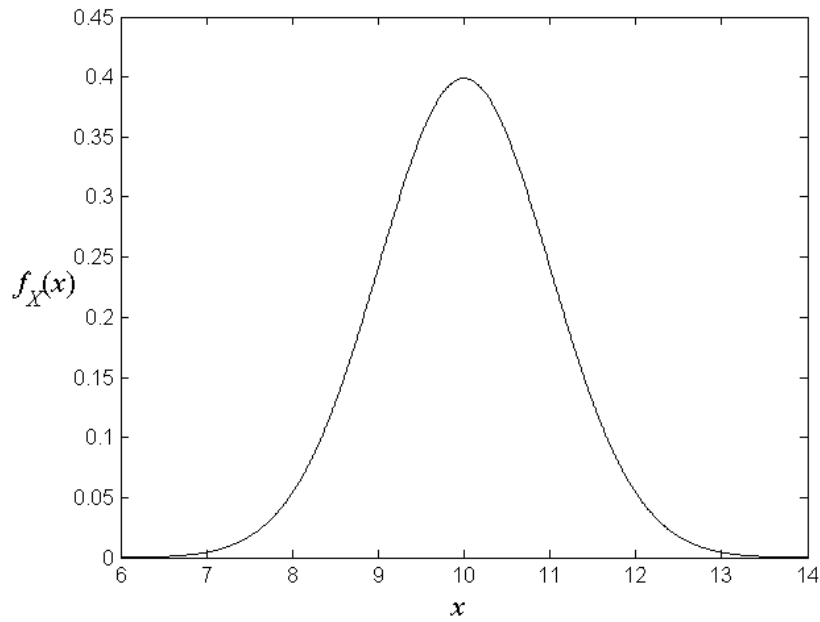


Figure 3.1 pdf of Random Variable $X \sim N(10,1)$

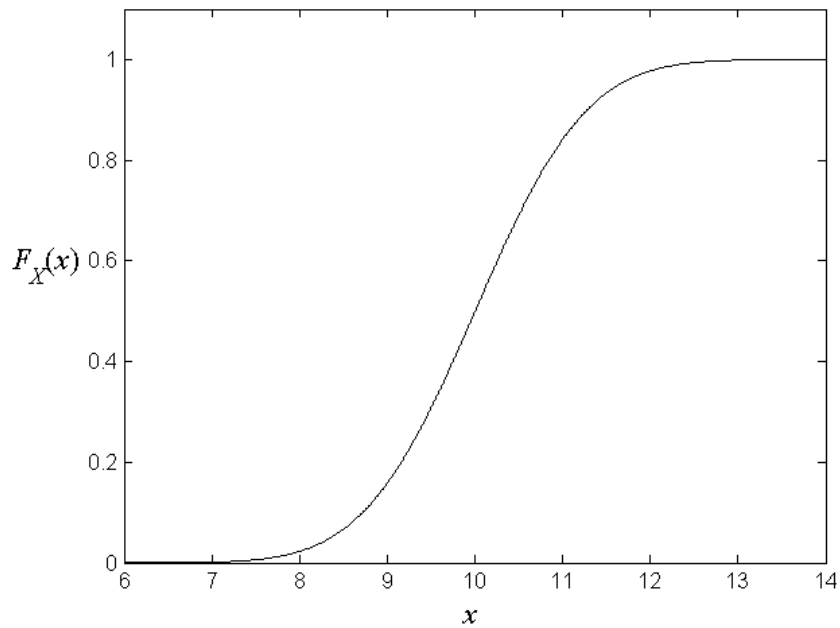


Figure 3.2 cdf of Random Variable $X \sim N(10,1)$

The standard normal distribution has a mean of 0 and standard deviation of 1 and is denoted as $Z \sim N(0,1)$. Its pdf $f(z)$ and cdf $\Phi(z)$ are given below.

$$f(z) = \frac{1}{\sqrt{2p}} \exp\left(-\frac{1}{2}z^2\right), \quad -\infty < z < +\infty \quad (4.4)$$

$$\Phi(z) = \int_{-\infty}^z \frac{1}{\sqrt{2p}} \exp\left(-\frac{1}{2}z^2\right) dz, \quad -\infty < z < +\infty \quad (4.5)$$

The *cdf* of a general normal distribution can be calculated from that of the standard normal distribution as

$$F_X(x) = P(X \leq x) = P(\mathbf{m}_X + Z\mathbf{s}_X \leq x) = P\left(Z \leq \frac{x - \mathbf{m}_X}{\mathbf{s}_X}\right) = \Phi\left(\frac{x - \mathbf{m}_X}{\mathbf{s}_X}\right)$$

Therefore, Φ is often used to calculate the *cdf* of a general normal distribution by the following equation

$$F_X(x) = \Phi\left(\frac{x - \mathbf{m}_X}{\mathbf{s}_X}\right) \quad (4.6)$$

The features of a normal distribution are listed below.

- The distribution is symmetric around its mean.
- Because of the symmetry, the median X_M is equal to the mean \mathbf{m}_X .
- Normal distributions are denser in the center and less dense in the tails.
- The distribution is defined by two parameters, the mean \mathbf{m}_X and the standard deviation \mathbf{s}_X .
- The range of the distribution extends from $-\infty$ to $+\infty$.
- $\mathbf{m}_X \pm 1\mathbf{s}_X$ contains 68.27% of items.
- $\mathbf{m}_X \pm 2\mathbf{s}_X$ contains 95.45% of items.
- $\mathbf{m}_X \pm 3\mathbf{s}_X$ contains 99.73% of items.
- 50% of all items fall between $\mathbf{m}_X \pm 0.674\mathbf{s}_X$.
- 95% of all items fall between $\mathbf{m}_X \pm 1.960\mathbf{s}_X$.
- 99% of all items fall between $\mathbf{m}_X \pm 2.576\mathbf{s}_X$.

The following example demonstrates that $\mathbf{m}_X \pm 3\mathbf{s}_X$ contains 99.73% of items.

$$\begin{aligned} P(\mathbf{m}_X - 3\mathbf{s}_X \leq X \leq \mathbf{m}_X + 3\mathbf{s}_X) &= \Phi\left[\frac{(\mathbf{m}_X + 3\mathbf{s}_X) - \mathbf{m}_X}{\mathbf{s}_X}\right] - \Phi\left[\frac{(\mathbf{m}_X - 3\mathbf{s}_X) - \mathbf{m}_X}{\mathbf{s}_X}\right] \\ &= \Phi(3) - \Phi(-3) = 0.9973 \end{aligned}$$

Many random variables follow a normal distribution. Examples include

- Dimensions of a product
- Measurement errors
- The intensity of laser light
- Test scores
- The time until you have your next car accident

Another important concept associated with a normal distribution is the z -score, which is an important concept in Design for Six Sigma (DFSS). DFSS is one of the probabilistic engineering design methods and will be discussed later in this book. If a normally distributed variable X is transformed into a standard normal variable Z by Eq. 4.3, the value z of Z is called a z -score. The equation is rewritten as below

$$z = \frac{x - \mathbf{m}_x}{\mathbf{s}_x} \quad (4.7)$$

Therefore, a z -score is a measure of the distance in standard deviations from the mean. It can be served as an indicator of quality level. Fig. 4.3 shows the cdf values for difference z -scores, ranging from -3 to $+3$.

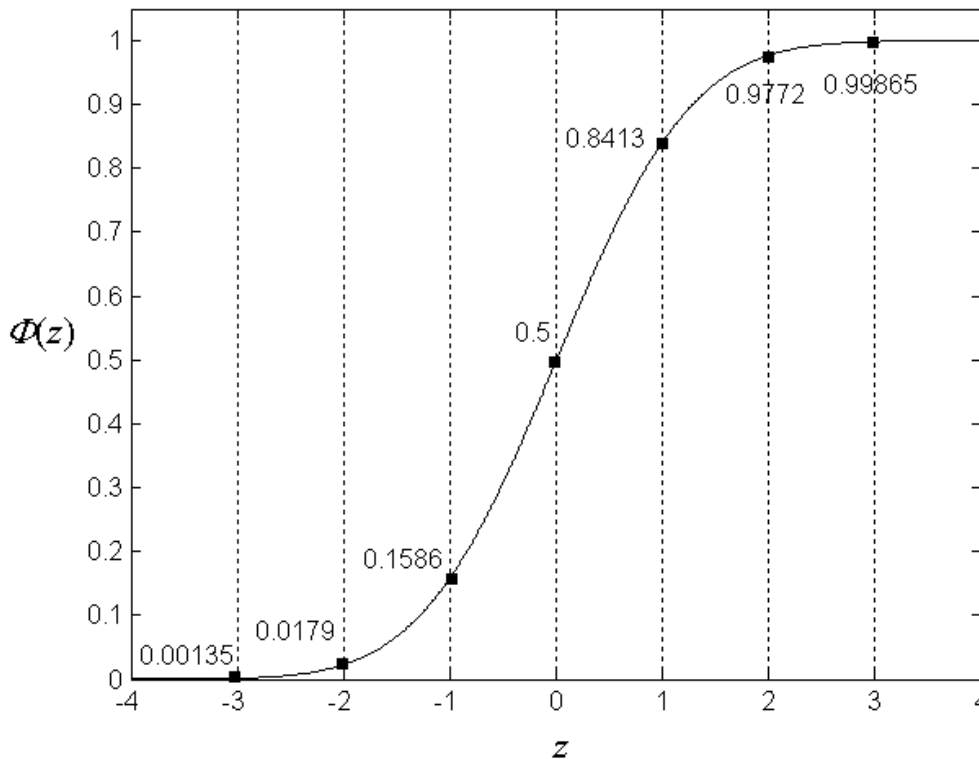
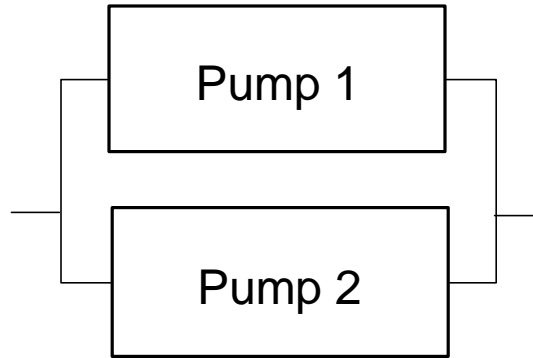


Figure 4.3 z-Score

Example 4.1

From filed data in an oil field, the time to failure of a pump, X , is known normally distributed. The mean and standard deviation of the time to failure are estimated from historical data as $m_x = 3200$ hr and $s_x = 600$ hr, respectively. (1) What is the probability that a pump will fail after it has worked for 2000 hours? (2) If two pumps work together as a system as shown in Fig. 4.4, what is probability that the system will fail after it has worked for 2000 hours?

**Figure 4.4 A Pumping System**

(1) The probability of failure of the pump is the probability that the time to failure X is less than 2000 hr; therefore

$$p_f = P(X < 2000) = F_x(1200) = \Phi\left(\frac{2000 - 3200}{600}\right) = \Phi(-2) = 0.0228$$

The reliability of the pump is then given by

$$R = 1 - p_f = 1 - 0.0228 = 0.9772$$

(2) The two pumps form a parallel system, from Eq. 2.3, the reliability of the system is

$$R_s = 1 - (1 - R)(1 - R) = 1 - (1 - 0.9772)(1 - 0.9772) = 0.9995$$

The probability of failure is then given by

$$p_f = 1 - R = 1 - 0.9995 = 0.0005$$

4.3 Uniform Distribution

If a random variable is uniformly distributed in a range $[a, b]$, it follows a uniform distribution. The *pdf* and *cdf* of a uniform distribution is given by

$$f_x(x) = \begin{cases} \frac{1}{b-a} & a \leq x \leq b \\ 0 & \text{otherwise} \end{cases} \quad (4.8)$$

and

$$F_x(x) = \begin{cases} 0 & x < a \\ \frac{x-a}{b-a} & a \leq x \leq b \\ 1 & x > b \end{cases} \quad (4.9)$$

respectively.

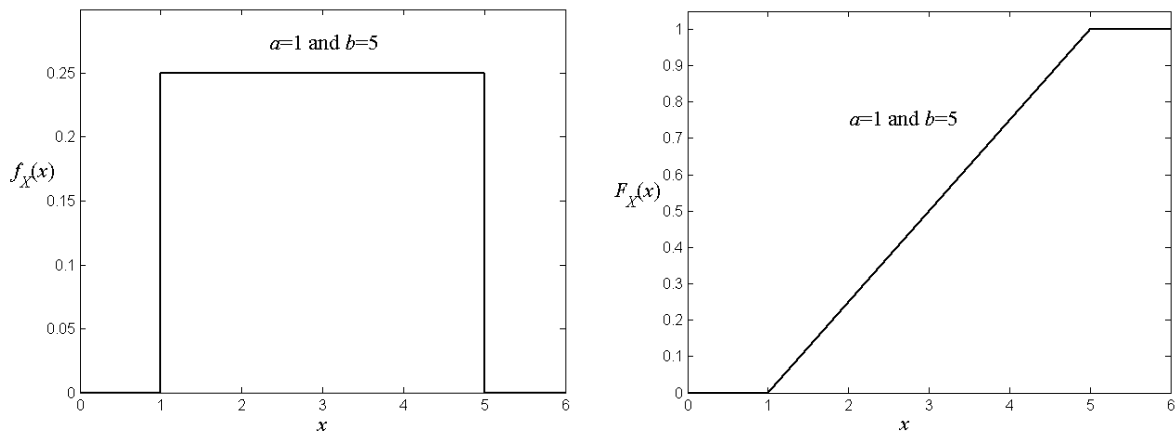


Figure 4.5 pdf and cdf of a Uniform Distribution

An example of uniform distribution is depicted in Fig. 4.5. The *pdf* is constant and the *cdf* is a straight line with a positive slope in the range of the random variable.

The mean of a uniform distribution is given by

$$m_x = \frac{a+b}{2} \quad (4.10)$$

and the variance is given by

$$s_x^2 = \frac{(b-a)^2}{12} \quad (4.11)$$

4.4 Lognormal Distribution

Random variable X follows a lognormal distribution if $Y = \ln(X)$ is normally distributed, where $\ln(X)$ is the natural logarithm to the base e . The *pdf* of X is given by

$$f_X(x) = \frac{1}{\sqrt{2p} s_Y x} \exp \left[-\frac{1}{2} \left(\frac{\ln x - m_Y}{s_Y} \right)^2 \right], \quad 0 \leq x < \infty \quad (4.12)$$

where m_Y and s_Y are the mean and standard deviation of $Y = \ln(X)$, respectively. The *pdf* of a lognormal distribution is shown in Fig. 4.6.

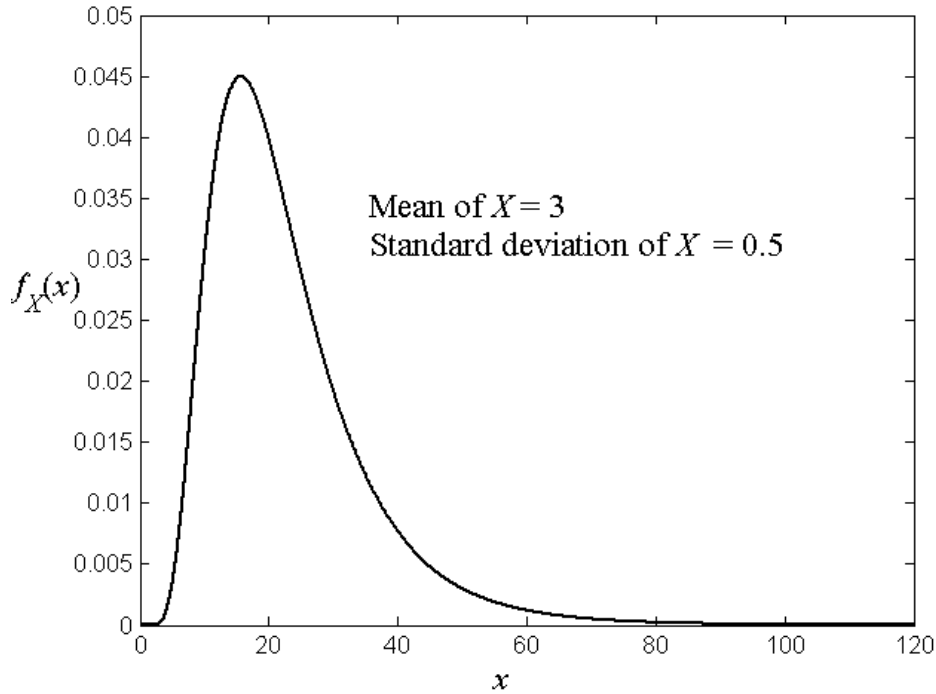


Figure 4.6 The pdf of A Lognormal Distribution

The *cdf* of a lognormal variable is given by

$$F_X(x) = \frac{1}{\sqrt{2p}} \int_{-\infty}^{\left(\frac{\ln b - m_Y}{s_Y}\right)} \exp\left(-\frac{1}{2}z^2\right) dz = \Phi\left(\frac{\ln b - m_Y}{s_Y}\right). \quad (4.13)$$

The relationships between means and standard deviations of X and Y are given by

$$\mathbf{s}_Y^2 = \ln \left[\left(\frac{\mathbf{s}_X}{\mathbf{m}_X} \right)^2 + 1 \right] \quad (4.15)$$

and

$$\mathbf{m}_Y = \ln \mathbf{m}_X - \frac{1}{2} \mathbf{s}_Y^2 \quad (4.16)$$

Examples of random variables that follow lognormal distribution include

- Fatigue life of materials
- Strength of some materials

4.5 Exponential Distribution

The *cdf* of an exponential distribution is given by

$$F_X(x) = \begin{cases} 1 - e^{-lx} & x \geq 0 \\ 0 & x < 0 \end{cases} \quad (4.17)$$

The *pdf* of the exponential distribution is given by

$$f_X(x) = \begin{cases} l e^{-lx} & x \geq 0 \\ 0 & x < 0 \end{cases} \quad (4.18)$$

The mean and standard deviation of the exponential distribution are

$$\mathbf{m}_X = 1/l \quad (4.19)$$

and

$$\mathbf{s}_X = 1/l^2 \quad (4.20)$$

respectively.

The *pdf* and *cdf* of an exponential distribution are depicted in Figs. 4.7 and 4.8, respectively.

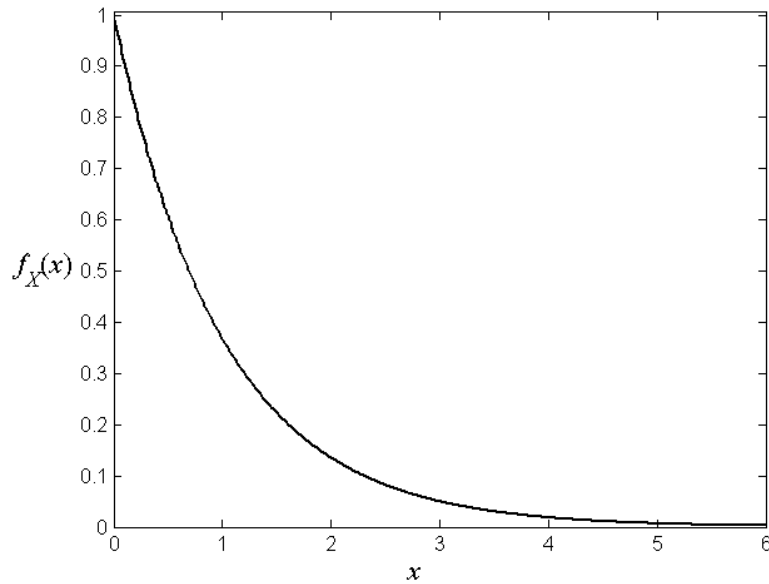


Figure 4.7 The *pdf* of an Exponential Distribution

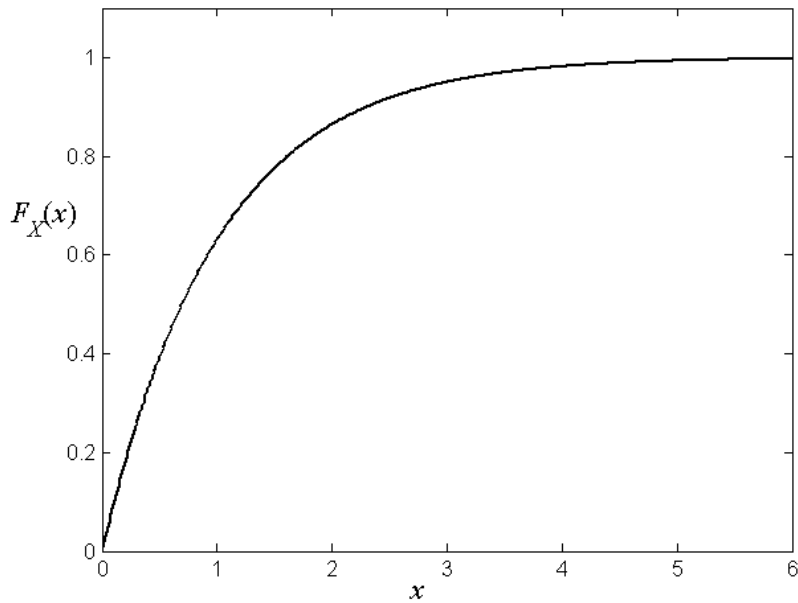


Figure 4.8 The *cdf* of an Exponential Distribution

Examples of variables that are exponentially distributed are:

- The life of electronic components
- The repair time of certain products
- The time until you have your next car accident
- The time until you get your next phone call

Note: More distributions will be added to this chapter.

Appendix MATLAB functions

Cumulative Distribution Functions (cdf)

normcdf – Normal cumulative distribution function

$P = \text{normcdf}(X, \text{MU}, \text{SIGMA})$ computes the normal cdf with mean MU and standard deviation SIGMA at the values in X .

logncdf – Lognormal cumulative distribution function

$P = \text{logncdf}(X, \text{MU}, \text{SIGMA})$ computes the lognormal cdf with mean MU and standard deviation SIGMA at the values in X .

expcdf – Exponential cumulative distribution function

$P = \text{expcdf}(X, \text{MU})$ returns the exponential cumulative distribution function with parameter MU at the values in X .

unifcdf – Uniform (continuous) cumulative distribution function

$P = \text{UNIFCDF}(X, \text{A}, \text{B})$ returns the cdf for the uniform distribution on the interval $[\text{A}, \text{B}]$ at the values in X .

Probability Density Functions

normpdf – Normal probability density function

$Y = \text{NORMPDF}(X, \text{MU}, \text{SIGMA})$ Returns the normal pdf with mean, MU , and standard deviation, SIGMA , at the values in X .

lognpdf – Lognormal probability density function

$Y = \text{LOGNPDF}(X, \text{MU}, \text{SIGMA})$ Returns the lognormal pdf at the values in X . The mean and standard deviation of $\log(Y)$ are MU and SIGMA .

exppdf – Exponential probability density function

$Y = \text{EXPPDF}(X, \text{MU})$ returns the exponential probability density function with parameter MU at the values in X .

unifpdf – Uniform (continuous) probability density function

$Y = \text{UNIFPDF}(X, \text{A}, \text{B})$ returns the continuous uniform pdf on the interval $[\text{A}, \text{B}]$ at the values in X . By default $\text{A} = 0$ and $\text{B} = 1$.

Inverse Cumulative Distribution Functions

norminv – Inverse of the normal cumulative distribution function

$X = \text{NORMINV}(P, \text{MU}, \text{SIGMA})$ finds the inverse of the normal cdf with mean, MU, and standard deviation, SIGMA. P is the probability.

logninv – Inverse of the lognormal cumulative distribution function

$X = \text{LOGNINV}(P, \text{MU}, \text{SIGMA})$ finds the inverse of the lognormal cdf with mean, MU, and standard deviation, SIGMA. P is the probability.

expinv – Inverse of the exponential cumulative distribution function

$X = \text{EXPINV}(P, \text{MU})$ returns the inverse of the exponential cumulative distribution function, with parameter MU, at the values in P.

unifinv – Inverse of uniform (continuous) distribution function

$X = \text{UNIFINV}(P, \text{A}, \text{B})$ returns the inverse of the uniform (continuous) distribution function on the interval [A,B], at the values in P. By default A = 0 and B = 1.