Foundation of Data Science and Analytics

Continuous RV and Probability Distributions

Arun K. Timalsina

Material Adaptation : Applied Statistics and Probability for Engineers, by Montgomery and Runger. John Wiley & Sons,

Continuous Random Variables

The dimensional length of a manufactured part is subject to small variations in measurement due to vibrations, temperature fluctuations, operator differences, calibration, cutting tool wear, bearing wear, and raw material changes.

This length X would be a continuous random variable that would occur in an interval (finite or infinite) of real numbers.

The number of possible values of X, in that interval, is uncountably infinite and limited only by the precision of the measurement instrument.

Continuous Density Functions

Density functions, in contrast to mass functions, distribute probability continuously along an interval.

The loading on the beam between points a & b is the integral of the function between points a & b.

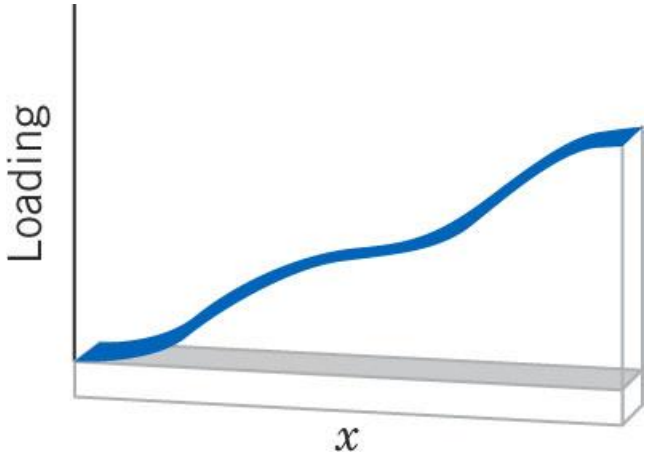


Figure 4-1 Density function as a loading on a long, thin beam. Most of the load occurs at the larger values of *x*.

A probability density function $f(x)$ describes the probability distribution of a continuous random variable. It is analogous to the beam loading.

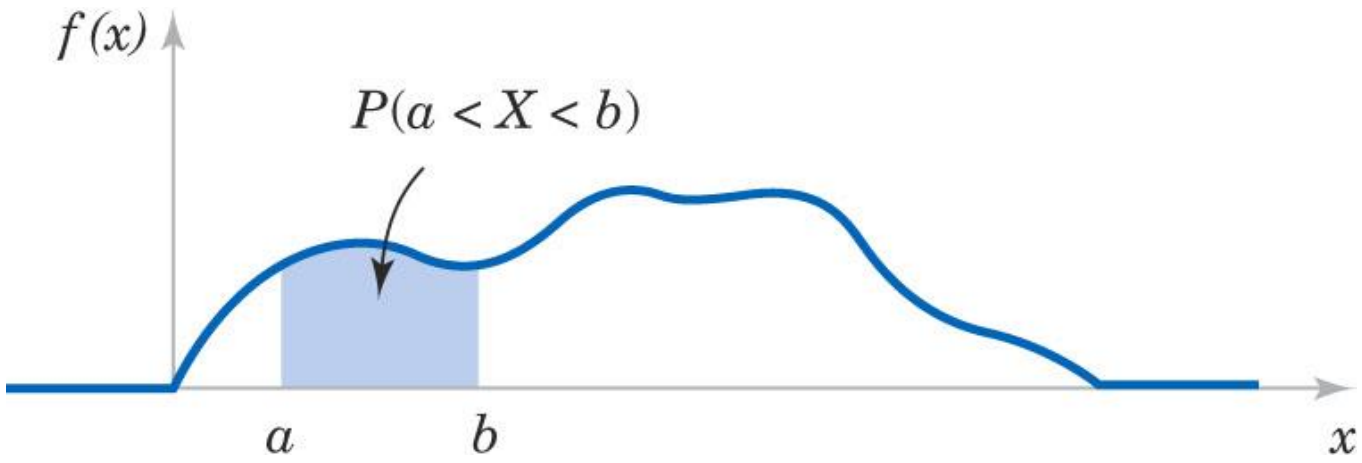


Figure 4-2 Probability is determined from the area under $f(x)$ from a to b.

Probability Density Function

For a continuous random variable X,

a probability density function is a function such that

(1) $f(x) \geq 0$ means that the function is always non-negative.

(2) $\int_{-\infty}^{\infty} f(x)dx = 1$

(3) $P(a \leq X \leq b) = \int_{a}^{b} f(x)dx =$ area under $f(x)dx$ from a to b

(4) $f(x) = 0$ means there is no area exactly at x.

Histograms

A histogram is graphical display of data showing a series of adjacent rectangles. Each rectangle has a base which represents an interval of data values. The height of the rectangle creates an area which represents the relative frequency associated with the values included in the base.

A continuous probability distribution $f(x)$ is a model approximating a histogram. A bar has the same area of the integral of those limits.

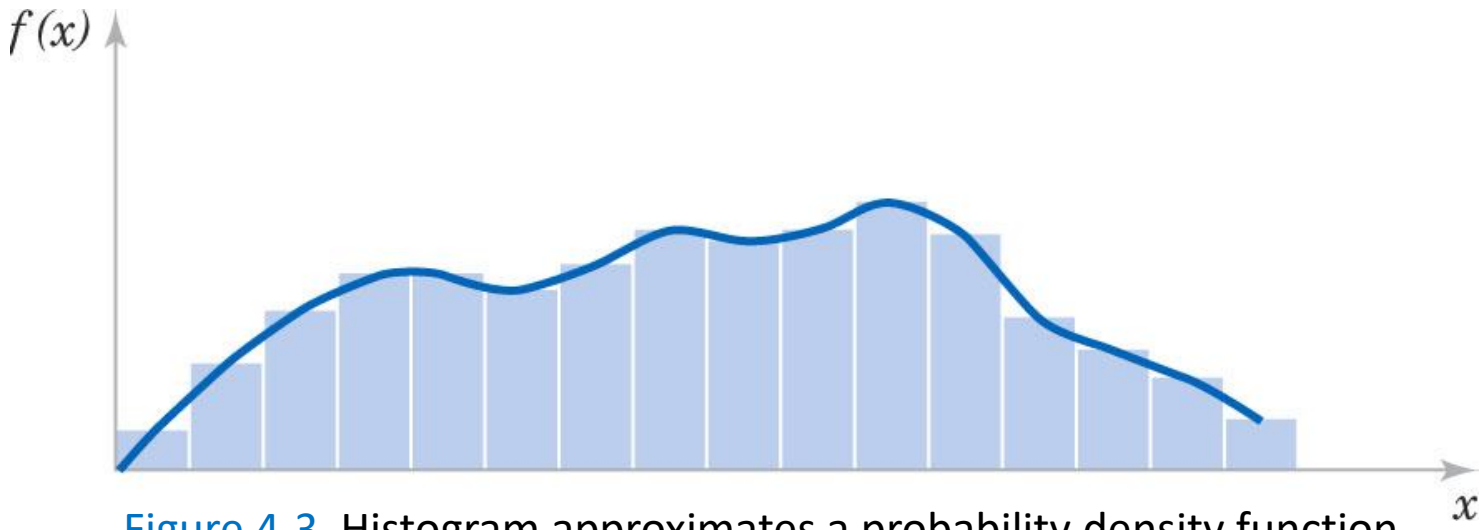


Figure 4-3 Histogram approximates a probability density function.

Area of a Point

If X is a continuous random variable, for any x_1 and x_2,

$$P(x_1 \le X \le x_2) = P(x_1 < X \le x_2) = P(x_1 \le X < x_2) = P(x_1 < X < x_2) \quad (4\text{-}2)$$

which implies that $P(X = x) = 0$.

From another perspective:

As x_1 approaches x_2, the area or probability becomes smaller and smaller.

As x_1 becomes x_2, the area or probability becomes zero.

Example 4-1: Electric Current

Let the continuous random variable X denote the current measured in a thin copper wire in milliamperes (mA). Assume that the range of X is $0 \le x \le 20$ and $f(x) = 0.05$. What is the probability that a current is less than 10mA?

Answer:

$$P(X < 10) = \int_0^{10} 0.5\,dx = 0.5$$

Another example,

$$P(5 < X < 20) = \int_5^{20} 0.5\,dx = 0.75$$

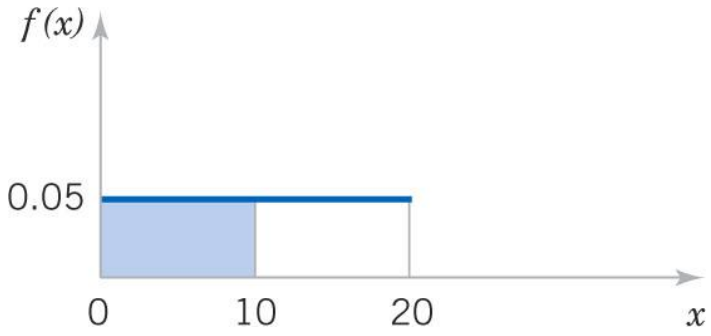


Figure 4-4 P(X < 10) illustrated.

Example 4-2: Hole Diameter

Let the continuous random variable X denote the diameter of a hole drilled in a sheet metal component. The target diameter is 12.5 mm. Random disturbances to the process result in larger diameters. Historical data shows that the distribution of X can be modeled by $f(x) = 20e^{-20(x-12.5)}$, $x \geq 12.5$ mm. If a part with a diameter larger than 12.60 mm is scrapped, what proportion of parts is scrapped?

Answer:

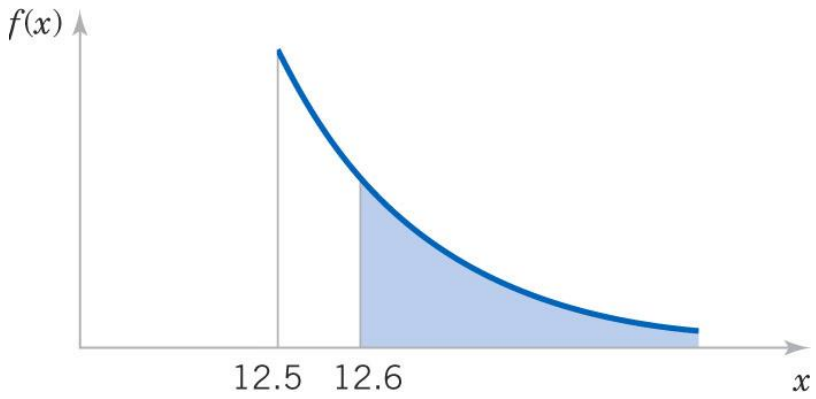


Figure 4-5 $P(X > 12.60) = \int_{12.6}^{\infty} 20e^{-20(x-12.5)}dx = 0.135$

Cumulative Distribution Functions

The cumulative distribution function of a continuous random variable X is,

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(u)\,du \quad \text{for} \quad -\infty < x < \infty \qquad (4\text{-}3)$$

Example 4-3: Electric Current

For the copper wire current measurement in Exercise 4-1, the cumulative distribution function (CDF) consists of three expressions to cover the entire real number line.

$$F(x) = \begin{cases} 0 & x < 0 \\ 0.05x & 0 \le x \le 20 \\ 1 & 20 < x \end{cases}$$

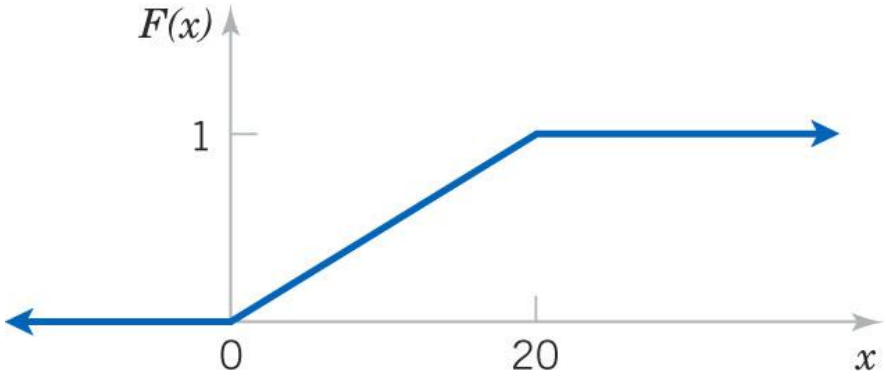


Figure 4-6 This graph shows the CDF as a continuous function.

Example 4-4: Hole Diameter

For the drilling operation in Example 4-2, F(x) consists of two expressions. This shows the proper notation.

$$F(x) = 0 \qquad \text{for } x < 12.5$$

$$F(x) = \int_{12.5}^{x} 20e^{-20(u-12.5)}du$$

$$= 1 - e^{-20(x-12.5)} \quad \text{for } x \geq 12.5$$

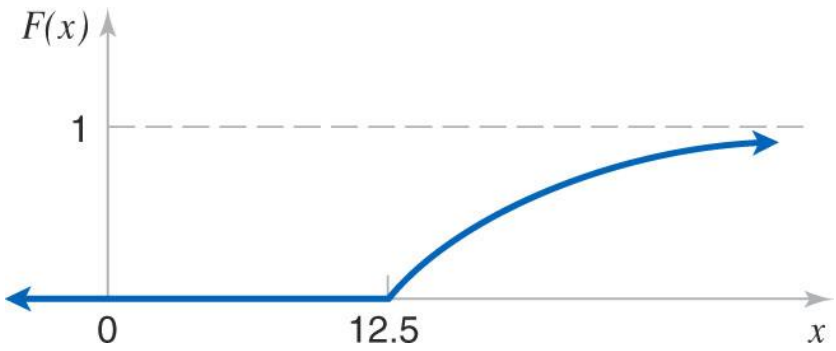


Figure 4-7 This graph shows F(x) as a continuous function.

Density vs. Cumulative Functions

- The probability density function (PDF) is the derivative of the cumulative distribution function (CDF).
- The cumulative distribution function (CDF) is the integral of the probability density function (PDF).

$$\text{Given } F(x),\ f(x) = \frac{dF(x)}{dx} \text{ as long as the derivative exists.}$$

Exercise 4-5: Reaction Time

- The time until a chemical reaction is complete (in milliseconds, ms) is approximated by this CDF:

$$F(x) = \begin{cases} 0 & \text{for } x < 0 \\ 1 - e^{-0.01x} & \text{for } 0 \le \text{x} \end{cases}$$

- What is the PDF?

$$f(x) = \frac{dF(x)}{dx} = \frac{d}{dx}\begin{cases} 0 \\ 1 - e^{-0.01x} \end{cases} = \begin{cases} 0 & \text{for } x < 0 \\ 0.01e^{-0.01x} & \text{for } 0 \le \text{x} \end{cases}$$

- What proportion of reactions is complete within 200 ms?

$$P(X < 200) = F(200) = 1 - e^{-2} = 0.8647$$

Mean & Variance

Suppose X is a continuous random variable with probability density function $f(x)$. The mean or expected value of X, denoted as μ or $E(X)$, is

$$\mu = E(X) = \int_{-\infty}^{\infty} x f(x)\, dx \qquad (4\text{-}4)$$

The variance of X, denoted as $V(X)$ or σ^2, is

$$\sigma^2 = V(X) \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\, dx = \int_{-\infty}^{\infty} x^2 f(x)\, dx - \mu^2$$

The standard deviation of X is $\sigma = \sqrt{\sigma^2}$.

Example 4-6: Electric Current

For the copper wire current measurement in Exercise 4-1, the PDF is $f(x) = 0.05$ for $0 \le x \le 20$. Find the mean and variance.

$$E(X) = \int_0^{20} x \cdot f(x)\,dx = \left.\frac{0.05x^2}{2}\right|_0^{20} = 10$$

$$V(X) = \int_0^{20} (x-10)^2 \, f(x)\,dx = \left.\frac{0.05(x-10)^3}{3}\right|_0^{20} = 33.33$$

Mean of a Function of a Random Variable

If X is a continuous random variable

with a probability density function $f(x)$,

$$E\left[h(x)\right] = \int_{-\infty}^{\infty} h(x) f(x)\, dx \qquad (4\text{-}5)$$

Example 4-7: In Example 4-1, *X* is the current measured in mA. What is the expected value of the squared current?

$$E\left[h(x)\right] = E\left[X^2\right] = \int_{0}^{20} x^2 f(x)\, dx$$

$$= \int_{0}^{20} 0.05x^2 dx = \frac{0.05x^3}{3}\Bigg|_{0}^{20} = 133.33 \text{ mA}^2$$

Example 4-8: Hole Diameter

For the drilling operation in Example 4-2, find the mean and variance of X using integration by parts. Recall that $f(x) = 20e^{-20(x-12.5)}dx$ for $x \geq 12.5$.

$$E(X) = \int_{12.5}^{\infty} xf(x)dx = \int_{12.5}^{\infty} x20e^{-20(x-12.5)}dx$$

$$= -xe^{-20(x-12.5)} - \frac{e^{-20(x-12.5)}}{20}\Bigg|_{12.5}^{\infty} = 12.5 + 0.05 = 12.55 \text{ mm}$$

$$V(X) = \int_{12.5}^{\infty} (x-12.55)^2 f(x)dx = 0.0025 \text{ mm}^2 \text{ and } \sigma = 0.05 \text{ mm}$$

Continuous Uniform Distribution

- This is the simplest continuous distribution and analogous to its discrete counterpart.
- A continuous random variable X with probability density function

$$f(x) = 1 / (b-a) \text{ for } a \le x \le b \qquad (4\text{-}6)$$

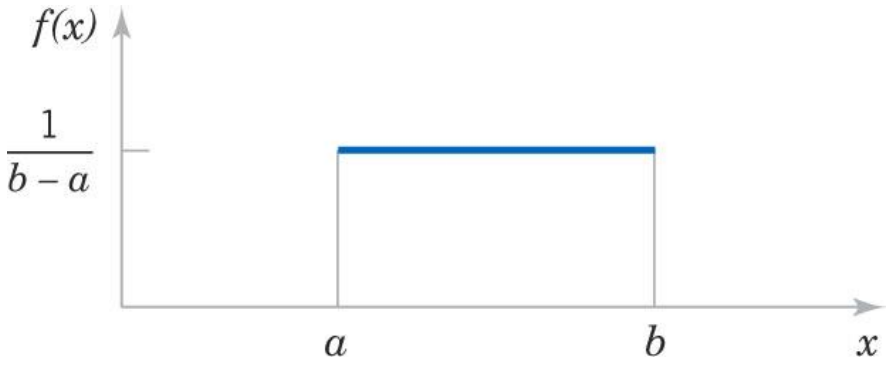


Figure 4-8 Continuous uniform PDF

Mean & Variance

- Mean & variance are:

$$\mu = E(X) = \frac{(a+b)}{2} \quad \text{and} \quad \sigma^2 = V(X) = \frac{(b-a)^2}{12} \qquad (4\text{-}7)$$

- Derivations are shown in the text. Be reminded that $b^2 - a^2 = (b + a)(b - a)$

Example 4-9: Uniform Current

Let the continuous random variable *X* denote the current measured in a thin copper wire in mA. Recall that the PDF is F(x) = 0.05 for 0 ≤ x ≤ 20.

What is the probability that the current measurement is between 5 & 10 mA?

$$P(5 < x < 10) = \int_{5}^{10} 0.05 dx = 5(0.05) = 0.25$$

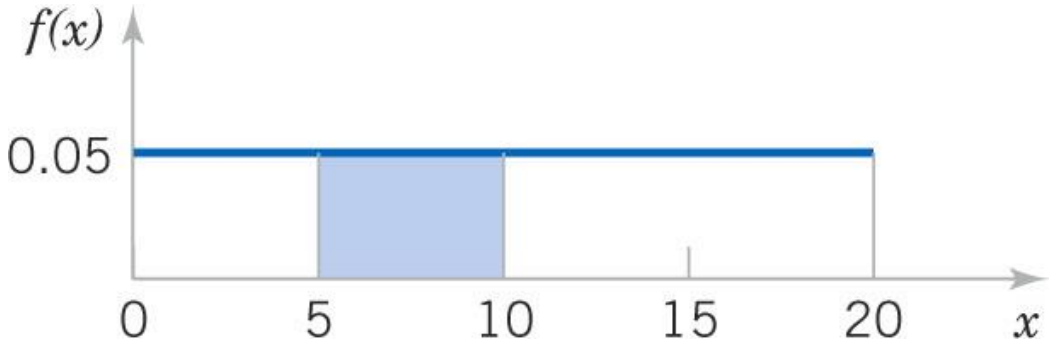


Figure 4-9

Continuous Uniform CDF

$$F(x) = \int_a^x \frac{1}{(b-a)} du = \frac{x-a}{b-a}$$

The CDF is completely described as

$$F(x) = \begin{cases} 0 & x < a \\ (x-a)/(b-a) & a \le x < b \\ 1 & b \le x \end{cases}$$

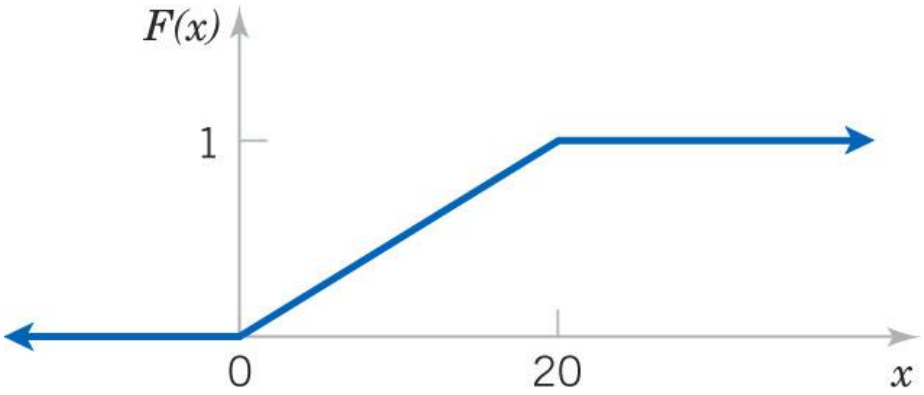


Figure 4-6 (again) Graph of the Cumulative Uniform CDF

Normal Distribution

- The most widely used distribution is the normal distribution, also known as the Gaussian distribution.
- Random variation of many physical measurements are normally distributed.
- The location and spread of the normal are independently determined by mean (μ) and standard deviation (σ).

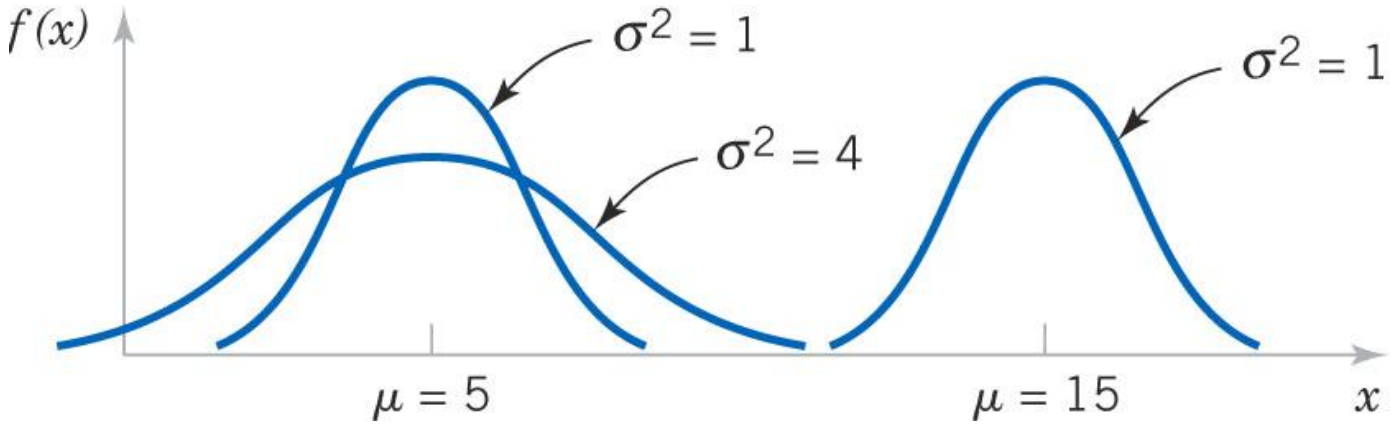


Figure 4-10 Normal probability density functions

Normal Probability Density Function

A random variable X with probability density function

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-(x-\mu)^2}{2\sigma^2}} \qquad -\infty < x < \infty \qquad (4\text{-}8)$$

is a normal random variable with parameters μ,

where $-\infty < \mu < \infty$, and $\sigma > 0$. Also,

$$E(X) = \mu \quad \text{and} \quad V(X) = \sigma^2 \qquad (4\text{-}9)$$

and the notation $N(\mu, \sigma^2)$ is used to denote the distribution.

Note that $f(X)$ cannot be intergrated analytically,
so $F(X)$ is expressed through numerical integration
with Excel or Minitab, and written as Appendix A, Table III.

Example 4-10: Normal Application

Assume that the current measurements in a strip of wire follows a normal distribution with a mean of 10 mA & a variance of 4 mA^2. Let *X* denote the current in mA.

What is the probability that a measurement exceeds 13 mA?

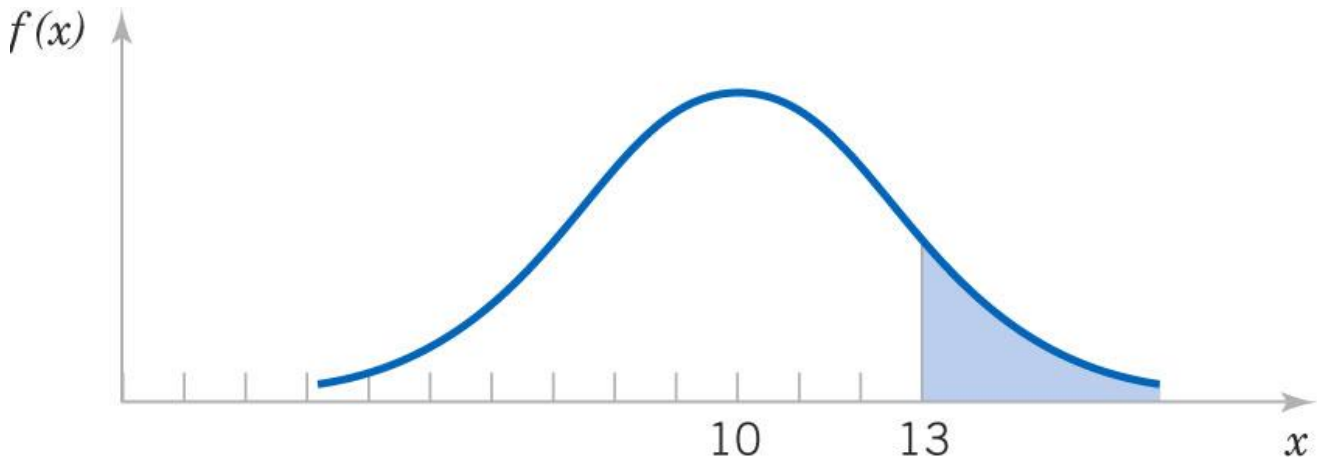


Figure 4-11 Graphical probability that $X > 13$ for a normal random variable with $\mu = 10$ and $\sigma^2 = 4$.

Empirical Rule

$$P(\mu - \sigma < X < \mu + \sigma) = 0.6827$$

$$P(\mu - 2\sigma < X < \mu + 2\sigma) = 0.9545$$

$$P(\mu - 3\sigma < X < \mu + 3\sigma) = 0.9973$$

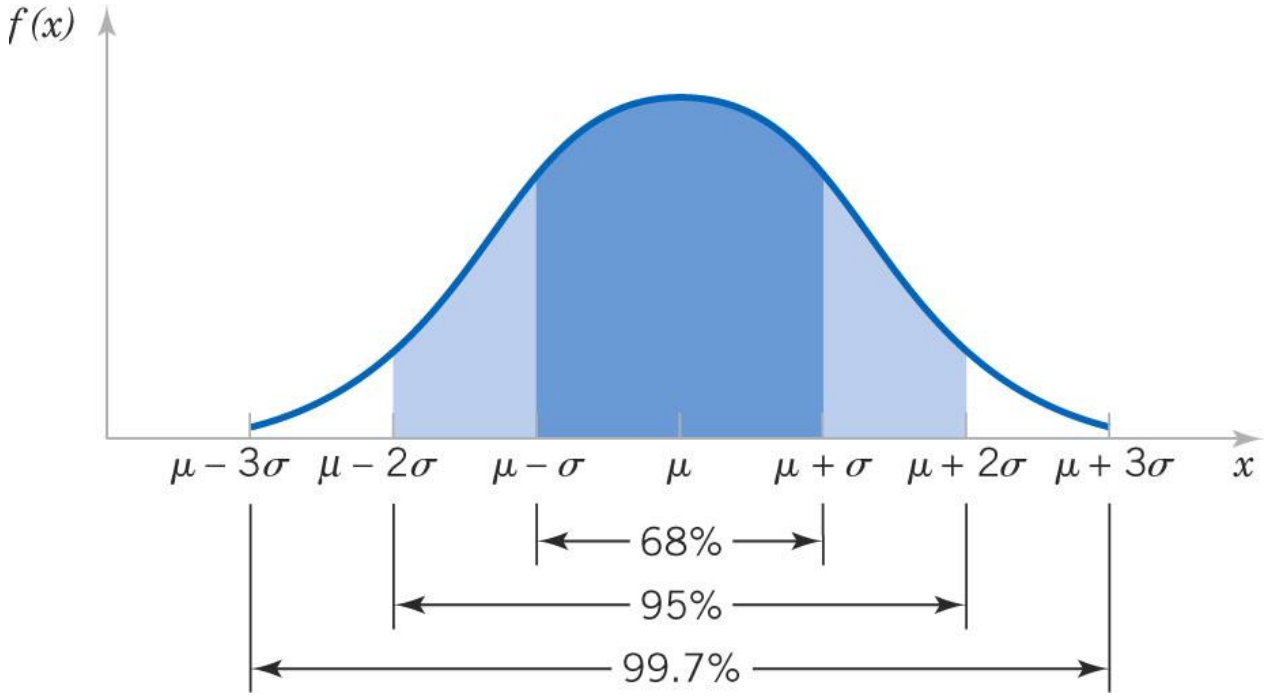


Figure 4-12 Probabilities associated with a normal distribution – well worth remembering to quickly estimate probabilities.

Standard Normal Distribution

A normal random variable with

$$\mu = 0 \text{ and } \sigma^2 = 1$$

Is called a standard normal random variable and is denoted as Z. The cumulative distribution function of a standard normal random variable is denoted as:

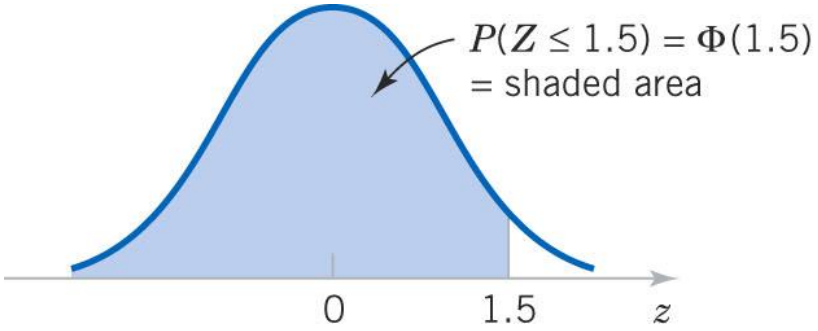
$$\Phi(z) = P(Z \leq z) = F(z)$$

Values are found in Appendix Table III and by using Excel and Minitab.

Example 4-11: Standard Normal Distribution

Assume Z is a standard normal random variable.

Find $P(Z \leq 1.50)$. Answer: 0.93319



z	0.00	0.01	0.02	0.03
0	0.50000	0.50399	0.50398	0.51197
⋮		⋮		
1.5	0.93319	0.93448	0.93574	0.93699

Figure 4-13 Standard normal PDF

Find $P(Z \leq 1.53)$. Answer: 0.93699

Find $P(Z \leq 0.02)$. Answer: 0.50398

Example 4-12: Standard Normal Exercises

1. $P(Z > 1.26) = 0.1038$
2. $P(Z < -0.86) = 0.195$
3. $P(Z > -1.37) = 0.915$
4. $P(-1.25 < 0.37) = 0.5387$
5. $P(Z \leq -4.6) \approx 0$
6. Find z for $P(Z \leq z) = 0.05$, z = -1.65
7. Find z for $(-z < Z < z) = 0.99$, z = 2.58

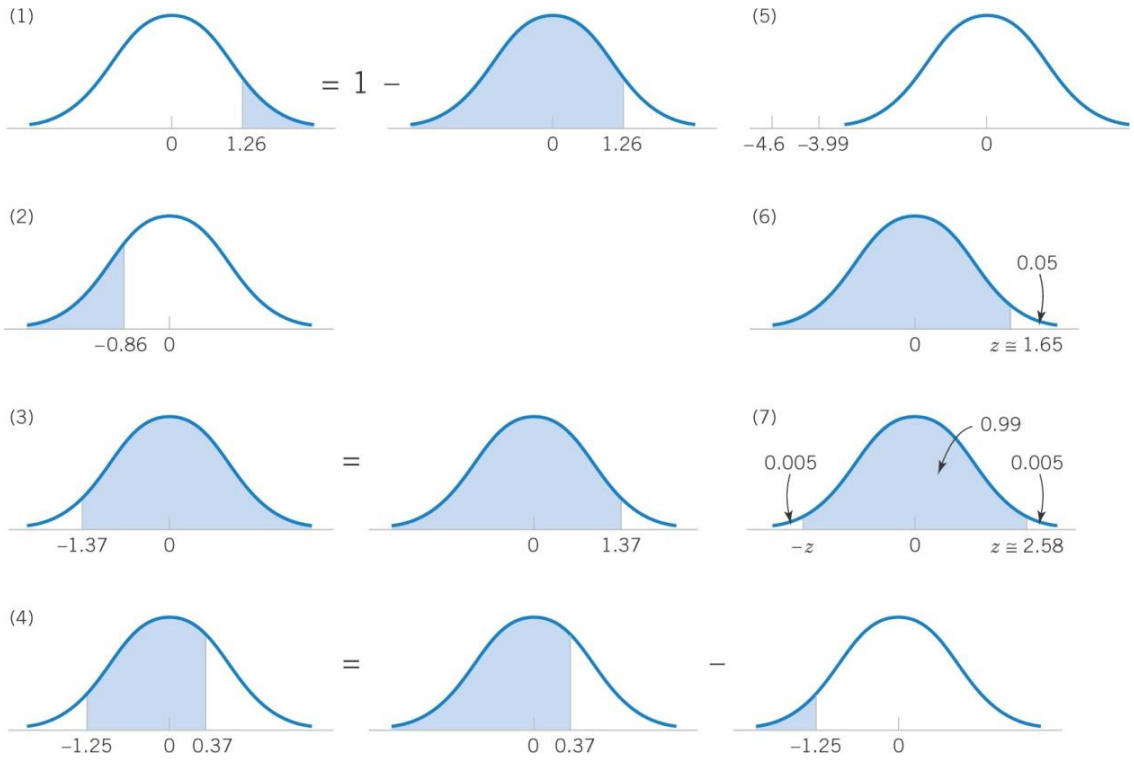


Figure 4-14 Graphical displays for standard normal distributions.

Standardizing

Suppose X is a normal random variable with mean μ and variance σ^2.

Then, $P(X \le x) = P\left(\frac{X-\mu}{\sigma} \le \frac{x-\mu}{\sigma}\right) = P(Z \le z)$ (4-11)

where Z is a standard normal random variable, and

$z = \frac{(x-\mu)}{\sigma}$ is the z-value obtainedby standardizing X.

The probability is obtained by using Appendix Table III

with $z = \frac{(x-\mu)}{\sigma}$.

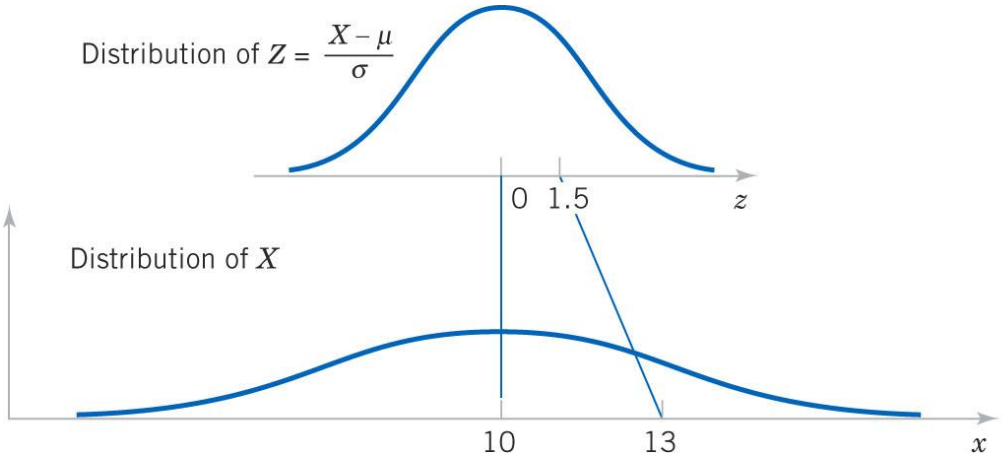
Example 4-14: Normally Distributed Current-1

From a previous example with μ = 10 and σ = 2 mA, what is the probability that the current measurement is between 9 and 11 mA?

$$P(9 < X < 11) = P\left(\frac{9-10}{2} < \frac{x-10}{2} < \frac{11-10}{2}\right)$$
$$= P(-0.5 < z < 0.5)$$
$$= P(z < 0.5) - P(z < -0.5)$$
$$= 0.69146 - 0.30854 = 0.38292$$

Answer:

Using Excel	
0.38292	= NORMDIST(11,10,2,TRUE) - NORMDIST(9,10,2,TRUE)



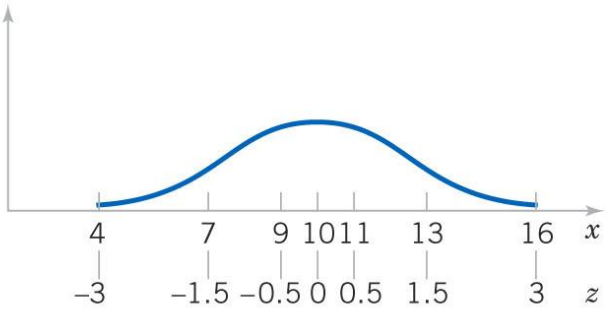


Figure 4-15 Standardizing a normal random variable.

Example 4-14: Normally Distributed Current-2

Determine the value for which the probability that a current measurement is below this value is 0.98.

Answer:

$$P(X < x) = P\left(\frac{X - 10}{2} < \frac{x - 10}{2}\right)$$

$$= P\left(Z < \frac{x - 10}{2}\right) = 0.98$$

$z = 2.05$ is the closest value.

$$z = 2(2.05) + 10 = 14.1 \text{ mA}.$$

Using Excel	
14.107	= NORMINV(0.98,10,2)

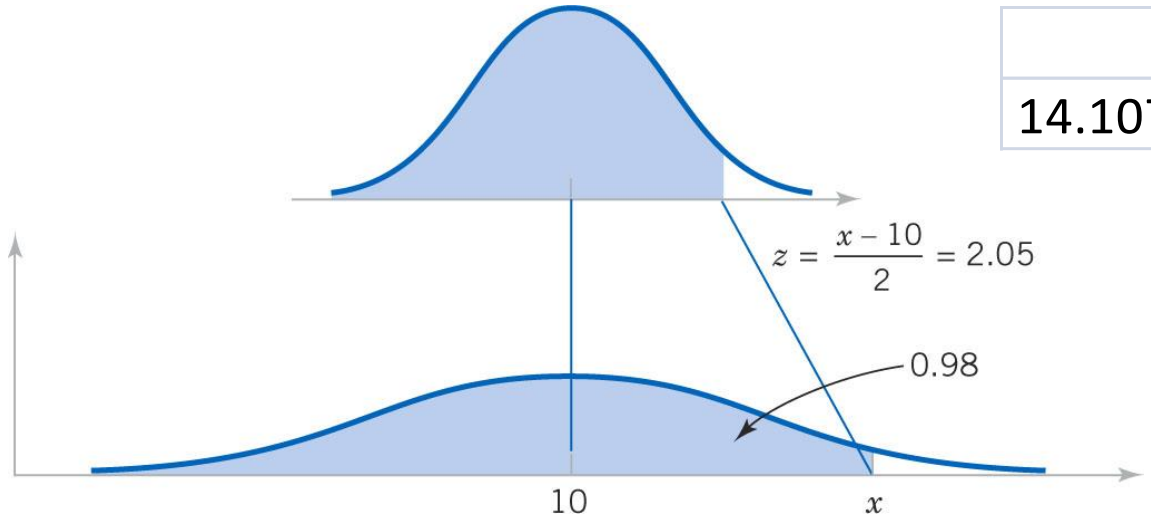


Figure 4-16 Determining the value of x to meet a specified probability.

Example 4-15: Signal Detection-1

Assume that in the detection of a digital signal, the background noise follows a normal distribution with μ = 0 volt and σ = 0.45 volt. The system assumes a signal 1 has been transmitted when the voltage exceeds 0.9. What is the probability of detecting a digital 1 when none was sent? Let the random variable *N* denote the voltage of noise.

$$P(N > 0.9) = P\left(\frac{N-0}{0.45} > \frac{0.9}{0.45}\right) = P(Z > 2)$$

$$= 1 - 0.97725 = 0.02275$$

Using Excel	
0.02275	= 1 - NORMDIST(0.9,0,0.45,TRUE)

This probability can be described as the probability of a false detection.

Example 4-15: Signal Detection-2

Determine the symmetric bounds about 0 that include 99% of all noise readings. We need to find x such that $P(-x < N < x) = 0.99$.

$$P(-x < N < x) = P\left(\frac{-x}{0.45} < \frac{N}{0.45} < \frac{x}{0.45}\right)$$

$$= P\left(\frac{-x}{0.45} < Z < \frac{x}{0.45}\right) = P(-2.58 < Z < 2.58)$$

$$x = 2.58(0.45) + 0 = 1.16$$

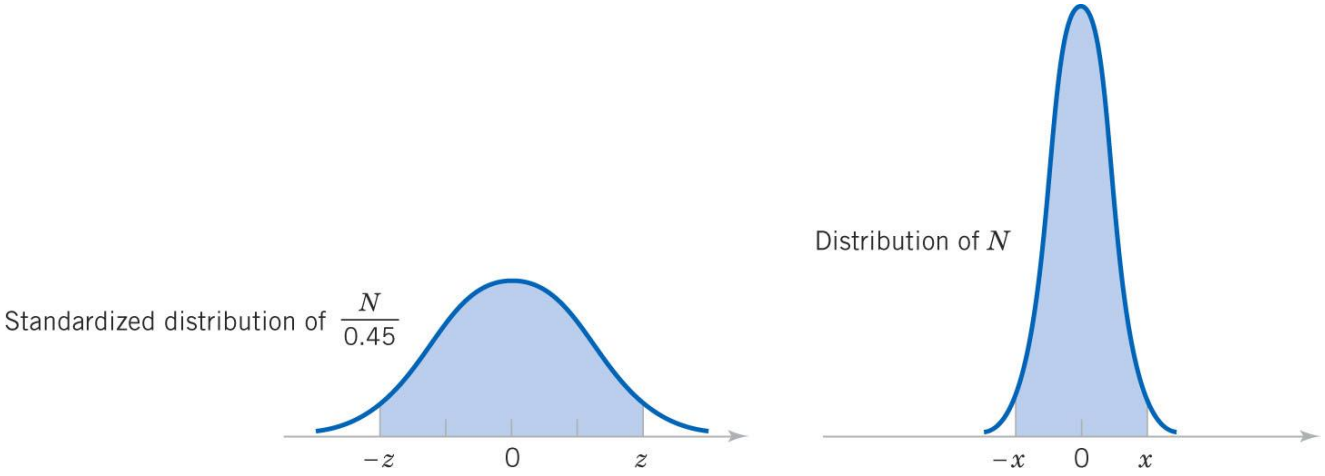


Figure 4-17 Determining the value of x to meet a specified probability.

Example 4-15: Signal Detection-3

Suppose that when a digital 1 signal is transmitted, the mean of the noise distribution shifts to 1.8 volts. What is the probability that a digital 1 is not detected? Let *S* denote the voltage when a digital 1 is transmitted.

$$P(S < 0.9) = P\left(\frac{S - 1.8}{0.45} < \frac{0.9 - 1.8}{0.45}\right)$$

$$= P(Z < -2) = 0.02275$$

Using Excel	
0.02275	= NORMDIST(0.9, 1.8, 0.45, TRUE)

This probability can be interpreted as the probability of a missed signal.

Example 4-16: Shaft Diameter-1

The diameter of the shaft is normally distributed with μ = 0.2508 inch and σ = 0.0005 inch. The specifications on the shaft are 0.2500 ± 0.0015 inch. What proportion of shafts conform to the specifications? Let *X* denote the shaft diameter in inches.

Answer:

$$P(0.2485 < X < 0.2515)$$

$$= P\left(\frac{0.2485 - 0.2508}{0.0005} < Z < \frac{0.2515 - 0.2508}{0.0005}\right)$$

$$= P(-4.6 < Z < 1.4)$$

$$= P(Z < 1.4) - P(Z < -4.6)$$

$$= 0.91924 - 0.0000 = 0.91924$$

Using Excel	
0.91924	= NORMDIST(0.2515, 0.2508, 0.0005, TRUE) - NORMDIST(0.2485, 0.2508, 0.0005, TRUE)

Example 4-16: Shaft Diameter-2

Most of the nonconforming shafts are too large, because the process mean is near the upper specification limit. If the process is centered so that the process mean is equal to the target value, what proportion of the shafts will now conform?

Answer:

$$P(0.2485 < X < 0.2515)$$

$$= P\left(\frac{0.2485 - 0.2500}{0.0005} < Z < \frac{0.2515 - 0.2500}{0.0005}\right)$$

$$= P(-3 < Z < 3)$$

$$= P(Z < 3) - P(Z < -3)$$

$$= 0.99865 - 0.00135 = 0.99730$$

Using Excel	
0.99730	= NORMDIST(0.2515, 0.25, 0.0005, TRUE) - NORMDIST(0.2485, 0.25, 0.0005, TRUE)

By centering the process, the yield increased from 91.924% to 99.730%, an increase of 7.806%

Normal Approximations

- The binomial and Poisson distributions become more bell-shaped and symmetric as their means increase.
- For manual calculations, the normal approximation is practical – exact probabilities of the binomial and Poisson, with large means, require technology (Minitab, Excel).
- The normal is a good approximation for the:
  - Binomial if $np > 5$ and $n(1-p) > 5$.
  - Poisson if $\lambda > 5$.

Normal Approximation to the Binomial

Suppose we have a binomial distribution with $n = 10$ and $p = 0.5$. Its mean and standard deviation are 5.0 and 1.58 respectively.

Draw the normal distribution over the binomial distribution.

The areas of the normal approximate the areas of the bars of the binomial with a continuity correction.

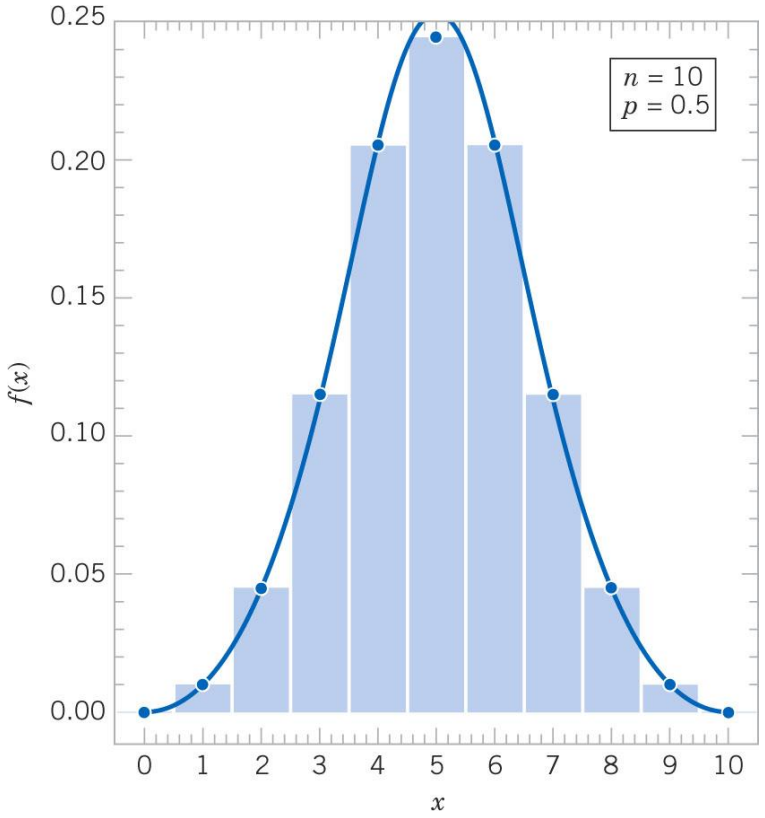


Figure 4-19 Overlaying the normal distribution upon a binomial with matched parameters.

Example 4-17:

In a digital comm channel, assume that the number of bits received in error can be modeled by a binomial random variable. The probability that a bit is received in error is 10^{-5}. If 16 million bits are transmitted, what is the probability that 150 or fewer errors occur? Let *X* denote the number of errors.

Answer:

$$P(X \leq 150) = \sum_{x=0}^{150} C_x^{16000000} \left(10^{-5}\right)^x \left(1-10^{-5}\right)^{16000000-x}$$

Using Excel	
0.2280	= BINOMDIST(150,16000000,0.00001,TRUE)

Can only be evaluated with technology. Manually, we must use the normal approximation to the binomial.

Normal Approximation Method

If X is a binomial random variable with parameters n and p,

$$Z = \frac{X - np}{\sqrt{np(1-p)}} \qquad (4\text{-}12)$$

is approximately a standard normal random variable. To approximate a binomial probability with a normal distribution, a continuity correction is applied as follows:

$$P(X \le x) = P(X \le x + 0.5) = P\left(Z \le \frac{x + 0.5 - np}{\sqrt{np(1-p)}}\right)$$

and

$$P(X \ge x) = P(X \le x - 0.5) = P\left(Z \le \frac{x - 0.5 - np}{\sqrt{np(1-p)}}\right)$$

The approximation is good for $np > 5$ and $n(1-p) > 5$. Refer to Figure 4-19 to see the rationale for adding and subtracting the 0.5 continuity correction.

Example 4-18: Applying the Approximation

The digital comm problem in the previous example is solved using the normal approximation to the binomial as follows:

$$P(X \le 150) = P(X \le 150.5)$$

$$= P\left(\frac{X - 160}{\sqrt{160\left(1 - 10^{-5}\right)}} \le \frac{150.5 - 160}{\sqrt{160\left(1 - 10^{-5}\right)}} \right)$$

$$= P\left(Z \le \frac{-9.5}{12.6491} \right) = P(-0.75104) = 0.2263$$

Using Excel	
0.2263	= NORMDIST(150.5, 160, SQRT(160*(1-0.00001)), TRUE)
-0.7%	= (0.2263-0.228)/0.228 = percent error in the approximation

Example 4-19: Normal Approximation-1

Again consider the transmission of bits. To judge how well the normal approximation works, assume n = 50 bits are transmitted and the probability of an error is p = 0.1. The exact and approximated probabilities are:

$$P(X \le 2) = C_0^{50} 0.9^{50} + C_1^{50} 0.1\left(0.9^{49}\right) + C_2^{50} 0.1^2 \left(0.9^{48}\right) = 0.112$$

$$P(X \le 2) = P\left(\frac{X - 5}{\sqrt{50(0.1)(0.9)}} < \frac{2.5 - 5}{\sqrt{50(0.1)(0.9)}} \right)$$

$$= P(Z < -1.18) = 0.119$$

Using Excel	
0.1117	= BINOMDIST(2,50,0.1,TRUE)
0.1193	= NORMDIST(2.5, 5, SQRT(5*0.9), TRUE)
6.8%	= (0.1193 - 0.1117) / 0.1117 = percent error

Example 4-19: Normal Approximation-2

$$P(X > 8) = P(X \geq 9) \approx P(X > 8.5)$$

$$= P\left(Z > \frac{8.5-5}{2.12}\right) = P(Z > 1.65) = 0.05$$

$$P(X = 5) \approx P(4.5 < X < 5.5)$$

$$= P\left(\frac{4.5-5}{2.12} < Z < \frac{5.5-5}{2.12}\right)$$

$$= P(-0.24 < Z < 0.24)$$

$$= P(Z < 0.24) - P(Z < 0.24) = 0.19$$

Using Excel	
0.1849	= BINOMDIST(5,50,0.1,FALSE)
0.1863	= NORMDIST(5.5, 5, SQRT(5*0.9), TRUE) - NORMDIST(4.5, 5, SQRT(5*0.9), TRUE)
0.8%	= (0.1863 - 0.1849) / 0.1849 = percent error

Reason for the Approximation Limits

The np > 5 and n(1-p) > 5 approximation rule is needed to keep the tails of the normal distribution from getting out-of-bounds.

As the binomial mean approaches the endpoints of the range of x, the standard deviation must be small enough to prevent overrun.

Figure 4-20 shows the asymmetric shape of the binomial when the approximation rule is not met.

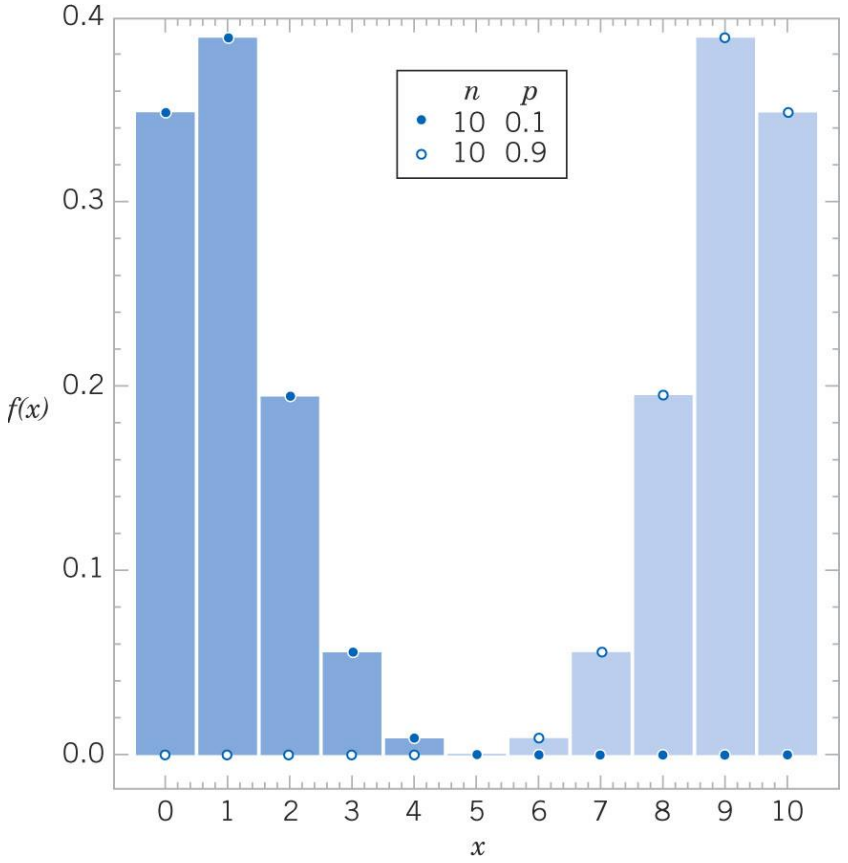


Figure 4-20 Binomial distribution is not symmetric as *p* gets near 0 or 1.

Normal Approximation to Hypergeometric

Recall that the hypergeometric distribution is similar to the binomial such that $p = K / N$ and when sample sizes are small relative to population size.

Thus the normal can be used to approximate the hypergeometric distribution also.

hypergeometric distribution	≈	binomial distribution	≈	normal distribution
	$n / N < 0.1$		$np < 5$	
			$n(1-p) < 5$	

Figure 4-21 Conditions for approximatine hypergeometric and binomial with normal probabilities

Normal Approximation to the Poisson

If X is a Poisson random variable with $E(X) = \lambda$ and $V(X) = \lambda$,

$$Z = \frac{X - \lambda}{\sqrt{\lambda}} \qquad (4\text{-}13)$$

is approximately a standard normal random variable. The same continuity correction used for the binomial distribution can also be applied. The approximation is good for

$$\lambda \geq 5$$

Example 4-20: Normal Approximation to Poisson

Assume that the number of asbestos particles in a square meter of dust on a surface follows a Poisson distribution with a mean of 100. If a square meter of dust is analyzed, what is the probability that 950 or fewer particles are found?

$$P(X \le 950) = \sum_{x=0}^{950} \frac{e^{-1000}1000^{x}}{x!} \quad \text{... too hard manually!}$$

$$\approx P(X < 950.5) = P\left(Z < \frac{950.5 - 1000}{\sqrt{1000}}\right)$$

$$= P(Z < -1.57) = 0.058$$

Using Excel	
0.0578	= POISSON(950,1000,TRUE)
0.0588	= NORMDIST(950.5, 1000, SQRT(1000), TRUE)
1.6%	= (0.0588 - 0.0578) / 0.0578 = percent error

Exponential Distribution

- The Poisson distribution defined a random variable as the number of flaws along a length of wire (flaws per mm).
- The exponential distribution defines a random variable as the interval between flaws (mm's between flaws – the inverse).

Let X denote the number of flaws in x mm of wire.

If the mean number of flaws is λ per mm,

N has a Poisson distribution with mean λx.

$$P(X > x) = \mathrm{P}(N = 0) = \frac{e^{-\lambda x}(\lambda x)^0}{0!} = e^{-\lambda x}$$

$$F(x) = P(X \le x) = 1 - e^{-\lambda x}, \quad x \ge 0, \text{ the CDF.}$$

Now differentiating:

$$f(x) = \lambda e^{-\lambda x}, \quad x \ge 0, \quad \text{the PDF.}$$

Exponential Distribution Definition

The random variable X that equals the distance between successive events of a Poisson process with mean number of events $\lambda > 0$ per unit interval is an exponential random variable with parameter λ. The probability density function of X is:

$$f(x) = \lambda e^{-\lambda x} \quad \text{for} \quad 0 \leq x < \infty \qquad (4\text{-}14)$$

Exponential Distribution Graphs

The y-intercept of the exponential probability density function is λ.

The random variable is non-negative and extends to infinity.

$F(x) = 1 - e^{-\lambda x}$ is well-worth committing to memory – it is used often.

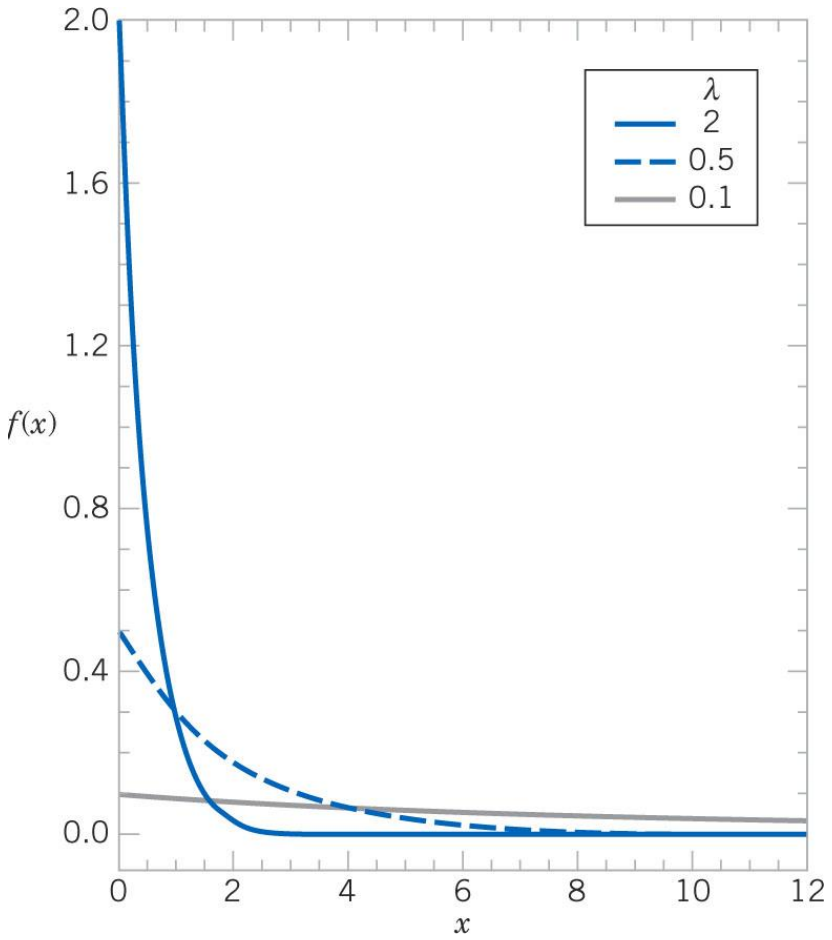


Figure 4-22 PDF of exponential random variables of selected values of λ.

Exponential Mean & Variance

If the random variable X has an exponential distribution with parameter λ,

$$\mu = E(X) = \frac{1}{\lambda} \quad \text{and} \quad \sigma^2 = V(X) = \frac{1}{\lambda^2} \qquad (4\text{-}15)$$

Note that, for the:

- Poisson distribution, the mean and variance are the same.
- Exponential distribution, the mean and standard deviation are the same.

Example 4-21: Computer Usage-1

In a large corporate computer network, user log-ons to the system can be modeled as a Poisson process with a mean of 25 log-ons per hour. What is the probability that there are no log-ons in the next 6 minutes (0.1 hour)? Let X denote the time in hours from the start of the interval until the first log-on.

$$P(X > 0.1) = \int_{0.1}^{\infty} 25e^{25x}dx = e^{-25(0.1)}$$

$$= 1 - F(0.1) = 0.082$$

Using Excel	
0.0821	= 1 - EXPONDIST(0.1,25,TRUE)

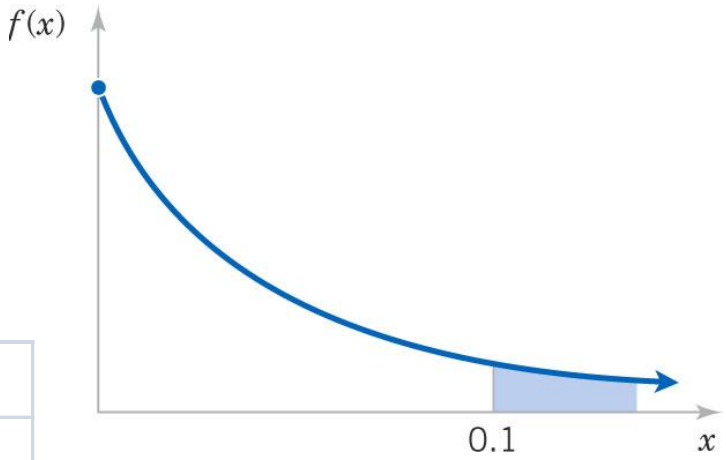


Figure 4-23 Desired probability.

Example 4-21: Computer Usage-2

Continuing, what is the probability that the time until the next log-on is between 2 and 3 minutes (0.033 & 0.05 hours)?

$$P(0.033 < X < 0.05) = \int_{0.033}^{0.05} 25e^{-25x}$$

$$= -e^{=25x}\Big|_{0.033}^{0.05} = 0.152$$

$$= F(0.05) - F(0.033) = 0.152$$

Using Excel	
0.148	= EXPONDIST(3/60, 25, TRUE) - EXPONDIST(2/60, 25, TRUE)
	(difference due to round-off error)

Example 4-21: Computer Usage-3

- Continuing, what is the interval of time such that the probability that no log-on occurs during the interval is 0.90?

$$P\left(X > x\right) = e^{-25x} = 0.90, \quad -25x = \ln\left(0.90\right)$$

$$x = \frac{-0.10536}{-25} = 0.00421 \text{ hour} = 0.253 \text{ minute}$$

- What is the mean and standard deviation of the time until the next log-in?

$$\mu = \frac{1}{\lambda} = \frac{1}{25} = 0.04 \text{ hour} = 2.4 \text{ minutes}$$

$$\sigma = \frac{1}{\lambda} = \frac{1}{25} = 0.04 \text{ hour} = 2.4 \text{ minutes}$$

Characteristic of a Poisson Process

- The starting point for observing the system does not matter.
- The probability of no log-in in the next 6 minutes [P(X > 0.1 hour) = 0.082], regardless of whether:
  - A log-in has just occurred or
  - A log-in has not occurred for the last hour.
- A system may have different means:
  - High usage period , e.g., $\lambda = 250$ per hour
  - Low usage period, e.g., $\lambda = 25$ per hour

Example 4-22: Lack of Memory Property

- Let *X* denote the time between detections of a particle with a Geiger counter. Assume X has an exponential distribution with E(X) = 1.4 minutes. What is the probability that a particle is detected in the next 30 seconds?

$$P(X < 0.5) = F(0.5) = 1 - e^{-0.5/1.4} = 0.30$$

Using Excel	
0.300	= EXPONDIST(0.5, 1/1.4, TRUE)

- No particle has been detected in the last 3 minutes. Will the probability increase since it is "due"?

$$P(X < 3.5 \mid X > 3) = \frac{P(3 < X < 3.5)}{P(X > 3)} = \frac{F(3.5) - F(3)}{1 - F(3)} = \frac{0.035}{0.117} = 0.30$$

  - No, the probability that a particle will be detected depends only on the interval of time, not its detection history.

Lack of Memory Property

- Areas $A+B+C+D=1$
- $A = P(X < t_2)$
- $A+B+C = P(X<t_1+t_2)$
- $C = P(X<t_1+t_2 \cap X>t_1)$
- $C+D = P(X>t_1)$
- $C/(C+D) = P(X<t_1+t_2 \mid X>t_1)$
- $A = C/(C+D)$

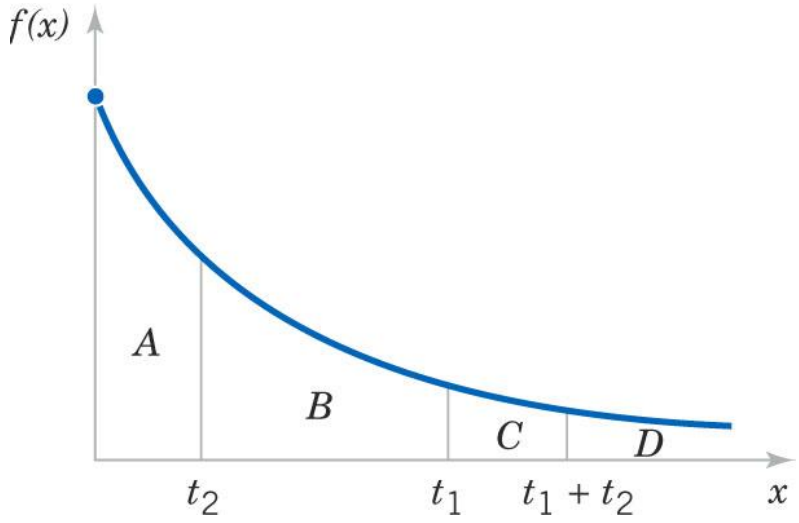


Figure 4-24 Lack of memory property of an exponential distribution.

Exponential Application in Reliability

- The reliability of electronic components is often modeled by the exponential distribution. A chip might have mean time to failure of 40,000 operating hours.
- The memoryless property implies that the component does not wear out – the probability of failure in the next hour is constant, regardless of the component age.
- The reliability of mechanical components do have a memory – the probability of failure in the next hour increases as the component ages. The Weibull distribution is used to model this situation.

Erlang & Gamma Distributions

- The Erlang distribution is a generalization of the exponential distribution.
- The exponential models the interval to the 1^{st} event, while the Erlang models the interval to the r^{th} event, i.e., a sum of exponentials.
- If r is not required to be an integer, then the distribution is called gamma.
- The exponential, as well as its Erlang and gamma generalizations, is based on the Poisson process.

Example 4-23: Processor Failure

The failures of CPUs of large computer systems are often modeled as a Poisson process. Assume that units that fail are repaired immediately and the mean number of failures per hour is 0.0001. Let *X* denote the time until 4 failures occur. What is the probability that *X* exceed 40,000 hours?

Let the random variable *N* denote the number of failures in 40,000 hours. The time until 4 failures occur exceeds 40,000 hours *iff* the number of failures in 40,000 hours is ≤ 3.

$$P(X > 40{,}000) = P(N \le 3)$$

$$E(N) = 40{,}000(0.0001) = 4 \text{ failure in } 40{,}000 \text{ hours}$$

$$P(N \le 3) = \sum_{k=0}^{3} \frac{e^{-4}4^k}{k!} = 0.433$$

Using Excel	
0.433	= POISSON(3, 4, TRUE)

Erlang Distribution

Generalizing from the prior exercise:

$$P(X > x) = \sum_{k=0}^{r-1} \frac{e^{\lambda x}(\lambda x)^k}{k!} = 1 - F(x)$$

Now differentiating $F(x)$:

$$f(x) = \frac{\lambda^r x^{r-1} e^{-\lambda x}}{(r-1)!} \quad \text{for } x > 0 \text{ and } r = 1, 2, \ldots$$

Gamma Function

The gamma function is the generalization of the factorial function for r > 0, not just non-negative integers.

$$\Gamma(r) = \int_0^\infty x^{r-1} e^{-x} dx, \quad \text{for } r > 0 \qquad (4\text{-}17)$$

Properties of the gamma function

$$\Gamma(r) = (r-1)\Gamma(r-1) \quad \text{recursive property}$$

$$\Gamma(r) = (r-1)! \quad \text{factorial function}$$

$$\Gamma(1) = 0! = 1$$

$$\Gamma(1/2) = \pi^{1/2} = 1.77 \quad \text{useful if manual}$$

Gamma Distribution

The random variable *X* with a probability density function:

$$f(x) = \frac{\lambda^r x^{r-1} e^{-\lambda x}}{\Gamma(r)}, \text{ for } x > 0 \qquad (4\text{-}18)$$

has a gamma random distribution with parameters $\lambda > 0$ and $r > 0$. If r is an positive integer, then *X* has an Erlang distribution.

Mean & Variance of the Gamma

- If X is a gamma random variable with parameters λ and r,

$$\mu = E(X) = r / \lambda \quad \text{and} \quad \sigma^2 = V(X) = r / \lambda^2 \qquad (4\text{-}19)$$

- r and λ work together to describe the shape of the gamma distribution.

Gamma Distribution Graphs

- The λ and r parameters are often called the "shape" and "scale", but may take on different meanings.
- Different parameter combinations change the distribution.
- The distribution becomes symmetric as *r* (and μ) increases.

Name	Text	Excel	Minitab
Scale	λ	β = 1 / λ	1 / λ
Shape	*r*	α	*r*

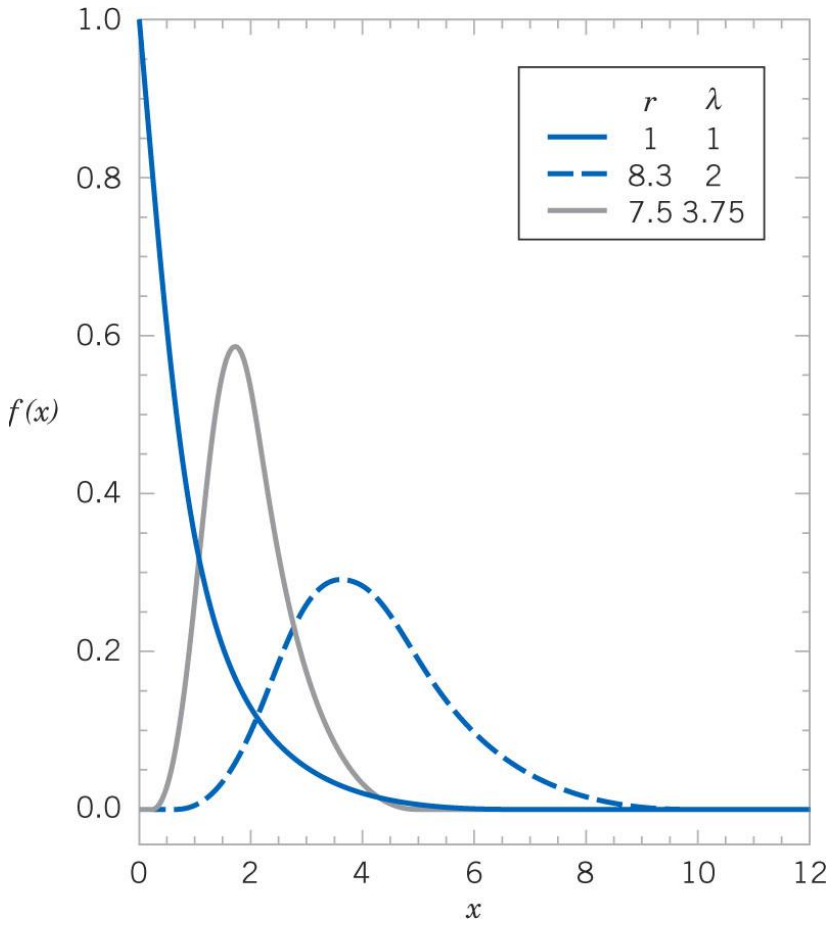


Figure 4-25 Gamma probability density functions for selected values of λ and *r*.

Example 4-24: Gamma Application-1

The time to prepare a micro-array slide for high-output genomics is a Poisson process with a mean of 2 hours per slide. What is the probability that 10 slides require more than 25 hours?

Let *X* denote the time to prepare 10 slides. Because of the assumption of a Poisson process, *X* has a gamma distribution with λ = ½, *r* = 10, and the requested probability is P(X > 25).

Using the Poisson distribution, let the random variable *N* denote the number of slides made in 10 hours. The time until 10 slides are made exceeds 25 hours *iff* the number of slides made in 25 hours is ≤ 9.

$$P(X > 25) = P(N \le 9)$$

$$E(N) = 25(1/2) = 12.5 \text{ slides in 25 hours}$$

$$P(N \le 9) = \sum_{k=0}^{9} \frac{e^{-12.5}(12.5)^k}{k!} = 0.2014$$

Using Excel	
0.2014	= POISSON(9, 12.5, TRUE)

Using the gamma distribution, the same result is obtained.

$$P(X > 25) = 1 - \int_0^{25} \frac{0.5^{10} x^9 e^{-0.5x}}{\Gamma(10)} dx$$

Using Excel	
0.2014	= 1 - GAMMADIST(25,10,2,TRUE)

Example 4-24: Gamma Application-2

What is the mean and standard deviation of the time to prepare 10 slides?

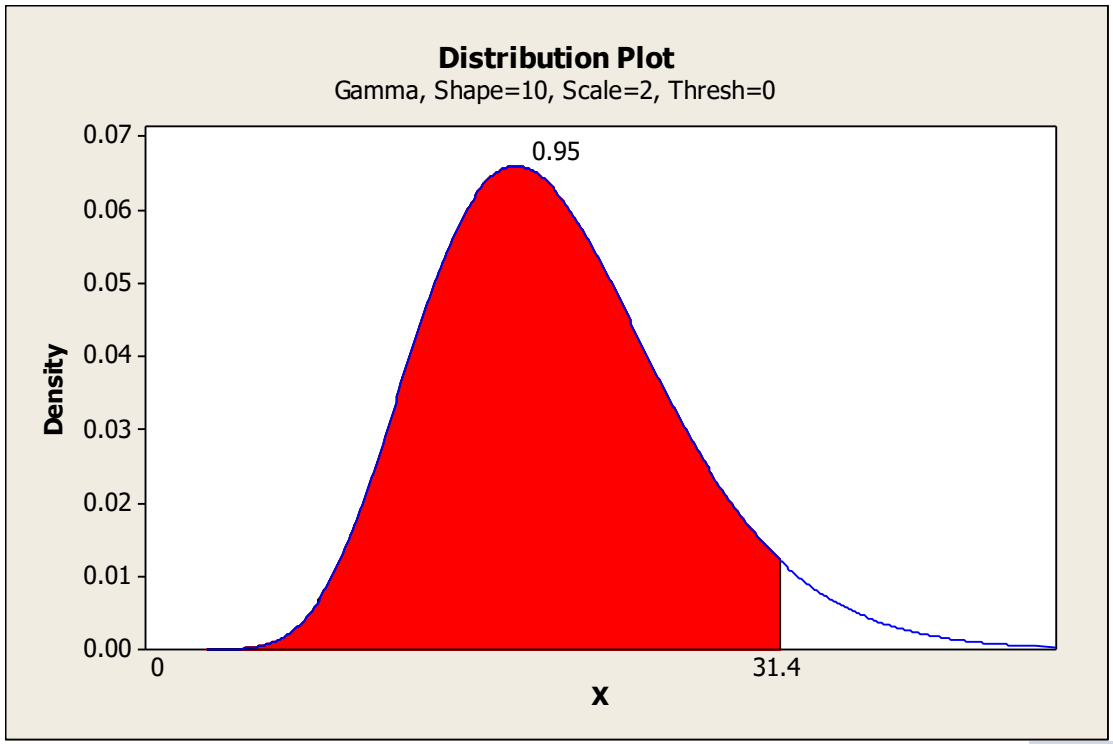
$$E(X) = \frac{r}{\lambda} = \frac{10}{0.5} = 20 \text{ hours}$$

$$\mathrm{V}(X) = \frac{r}{\lambda^2} = \frac{10}{0.25} = 40 \text{ hours}^2$$

$$SD(X) = \frac{\sqrt{10}}{\lambda} = \sqrt{40} = 6.32 \text{ hours}$$

Example 4-24: Gamma Application-3

The slides will be completed by what length of time with 95% probability? That is: $P(X \leq x) = 0.95$



Minitab: Graph > Probability Distribution Plot > View Probability

Using Excel	
31.41	= GAMMAINV(0.95, 10, 2)

Chi-Squared Distribution

- The chi-squared distribution is a special case of the gamma distribution with
  - $\lambda = 1/2$
  - $r = \nu/2$ where ν (nu) = 1, 2, 3, …
  - ν is called the "degrees of freedom".
- The chi-squared distribution is used in interval estimation and hypothesis tests as discussed in Chapter 7.

Weibull Distribution

- The Weibull distribution is often used to model the time until failure for physical systems in which failures:
  - Increase over time (bearings)
  - Decrease over time (some semiconductors)
  - Remain constant over time (subject to external shock)
- Parameters provide flexibility to reflect an item's failure experience or expectation.

Weibull PDF

The random variable X with probability density function

$$f(x) = \frac{\beta}{\delta^{\beta}} x^{\beta-1} e^{-(x/\delta)^{\beta}} \quad \text{for } x > 0 \qquad (4\text{-}20)$$

is a Weibull random variable with

scale parameter $\delta > 0$ and shape parameter $\beta > 0$.

The cumulative density function is:

$$F(x) = 1 - e^{-(x/\delta)^{\beta}} \qquad (4\text{-}21)$$

$$\mu = E(X) = \delta \cdot \Gamma\left(1 + \frac{1}{\beta}\right)$$

$$\sigma^2 = V(X) = \delta^2 \left[\Gamma\left(1 + \frac{2}{\beta}\right)\right] - \delta^2 \left[\Gamma\left(1 + \frac{1}{\beta}\right)\right]^2 \qquad (4\text{-}21a)$$

Weibull Distribution Graphs

Added slide

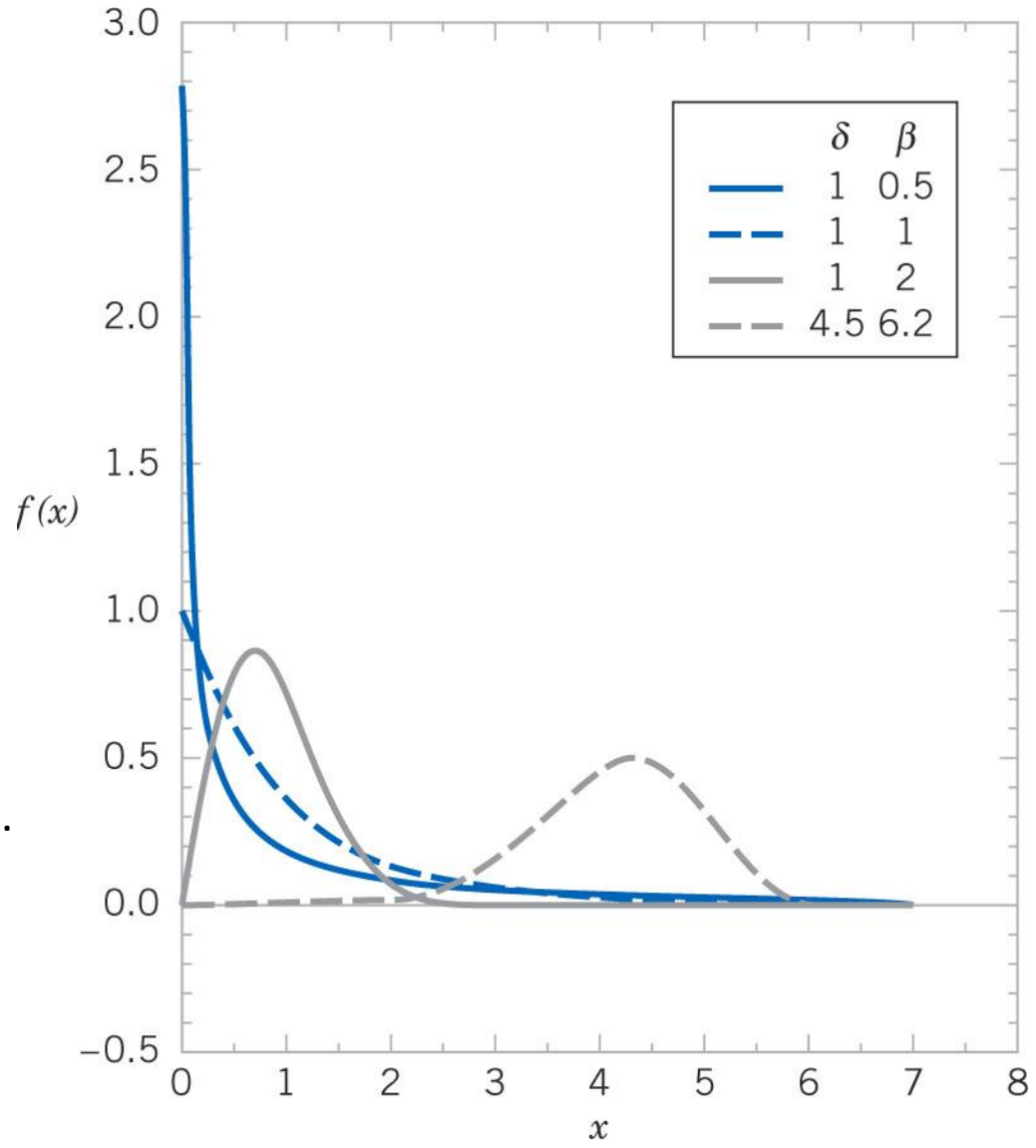


Figure 4-26 Weibull probability density function for selected values of δ and β.

Example 4-25: Bearing Wear

- The time to failure (in hours) of a bearing in a mechanical shaft is modeled as a Weibull random variable with β = ½ and δ = 5,000 hours.
- What is the mean time until failure?

$$E(X) = 5000 \cdot \Gamma(1+1/2) = 5000 \cdot \Gamma(1.5)$$
$$= 5000 \cdot 0.5\sqrt{\pi} = 4,431.1 \text{ hours}$$

Using Excel	
4,431.1	= 5000 * EXP(GAMMALN(1.5))

- What is the probability that a bearing will last at least 6,000 hours? (error in text solution)

$$P(X > 6,000) = 1 - F(6,000) = e^{-\left(\frac{6000}{5000}\right)^{0.5}}$$
$$= e^{-1.0954} = 0.334$$

Using Excel	
0.334	= 1 - WEIBULL(6000, 1/2, 5000, TRUE)

Lognormal Distribution

- Let *W* denote a normal random variable with mean of θ and variance of ω^2, i.e., $E(W) = \theta$ and $V(W) = \omega^2$
- As a change of variable, let $X = e^W = \exp(W)$ and $W = \ln(X)$
- Now X is a lognormal random variable.

$$F(x) = P[X \le x] = P[\exp(W) \le x] = P[W \le \ln(x)]$$

$$= P\left[Z \le \frac{\ln(x) - \theta}{\omega}\right] = \Phi\left[\frac{\ln(x) - \theta}{\omega}\right] = \quad \text{for} \quad x > 0$$

$$= 0 \quad \text{for} \quad x \le 0$$

$$f(x) = \frac{1}{x\omega\sqrt{2\pi}} e^{-\left[\frac{\ln(x)-\theta}{2\omega}\right]^2} \qquad \text{for } 0 < x < \infty$$

$$E(X) = e^{\theta+\omega^2/2} \qquad \text{and} \qquad V(X) = e^{2\theta+\omega^2}\left(e^{\omega^2} - 1\right) \qquad (4\text{-}22)$$

Lognormal Graphs

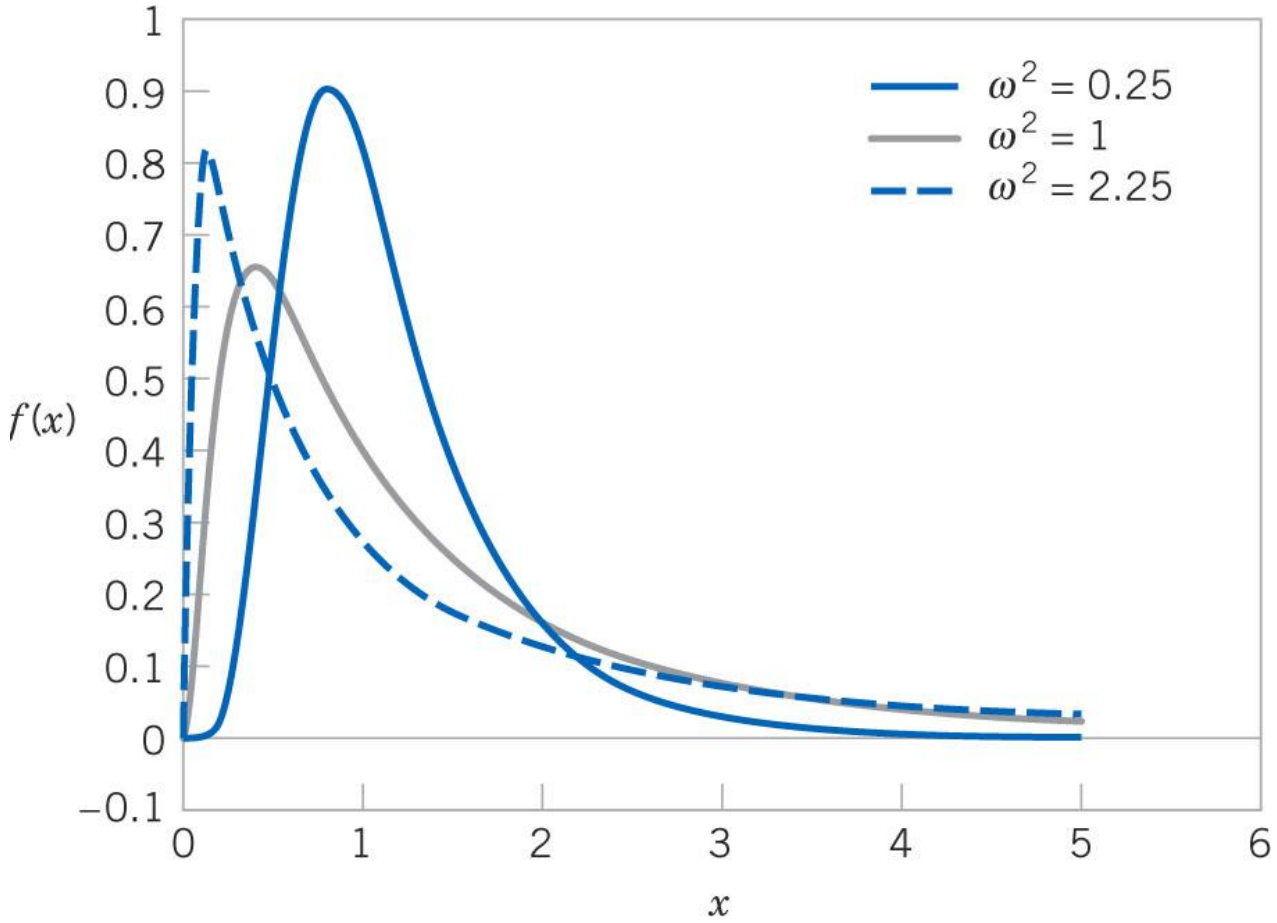


Figure 4-27 Lognormal probability density functions with θ = 0 for selected values of ω^2.

Example 4-27: Semiconductor Laser-1

The lifetime of a semiconductor laser has a lognormal distribution with θ = 10 and ω = 1.5 hours.

- What is the probability that the lifetime exceeds 10,000 hours?

$$
\begin{aligned}
P(X > 10,000) &= 1 - P\left[\exp(W) \le 10,000\right] \\
&= 1 - P\left[W \le \ln(10,000)\right] \\
&= 1 - \Phi\left(\frac{\ln(10,000) - 10}{1.5}\right) \\
&= 1 - \Phi(-0.5264) = 0.701
\end{aligned}
$$

1 - NORMDIST(LN(10000), 10, 1.5, TRUE) =	0.701

Example 4-27: Semiconductor Laser-2

- What lifetime is exceeded by 99% of lasers?

$$P(X > x) = P(\exp(W) > x) = P(W > \ln(x))$$

$$= 1 - \Phi\left(\frac{\ln(x) - 10}{1.5}\right) = 0.99$$

$$= 1 - \Phi(z) = 0.99 \text{ therefore } z = -2.33$$

$$\frac{\ln(x) - 10}{1.5} = -2.33 \text{ and } x = \exp(6.505) = 668.48 \text{ hours}$$

-2.3263	= NORMSINV(0.99)
6.5105	= -2.3263 * 1.5 + 10 = ln(x)
672.15	= EXP(6.5105)
(difference due to round-off error)	

- What is the mean and variance of the lifetime?

$$E(X) = e^{\theta + \varpi^2/2} = e^{10 + 1.5^2/2}$$

$$= \exp(11.125) = 67,846.29$$

$$V(X) = e^{2\theta + \omega^2}\left(e^{\omega^2} - 1\right) = e^{2 \cdot 10 + 1.5^2}\left(e^{1.5^2} - 1\right)$$

$$= \exp(22.25) \cdot \left[\exp(2.25) - 1\right] = 39,070,059,886.6$$

$$SD(X) = 197,661.5$$

Beta Distribution

A continuous distribution that is flexible, but bounded over the [0, 1] interval is useful for probability models. Examples are:

- Proportion of solar radiation absorbed by a material.
- Proportion of the max time to complete a task.

The random variable X with probability density function

$$f(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\cdot\Gamma(\beta)}\, x^{\alpha-1}(1-x)^{\beta-1} \quad \text{for} \quad 0 \le x \le 1$$

is a beta random variable with parameters $\alpha > 0$ and $\beta > 0$.

Beta Shapes Are Flexible

Distribution shape guidelines:

1. If $\alpha = \beta$, symmetrical about $x = 0.5$.
2. If $\alpha = \beta = 1$, uniform.
3. If $\alpha = \beta < 1$, symmetric & U- shaped.
4. If $\alpha = \beta > 1$, symmetric & mound-shaped.
5. If $\alpha \neq \beta$, skewed.

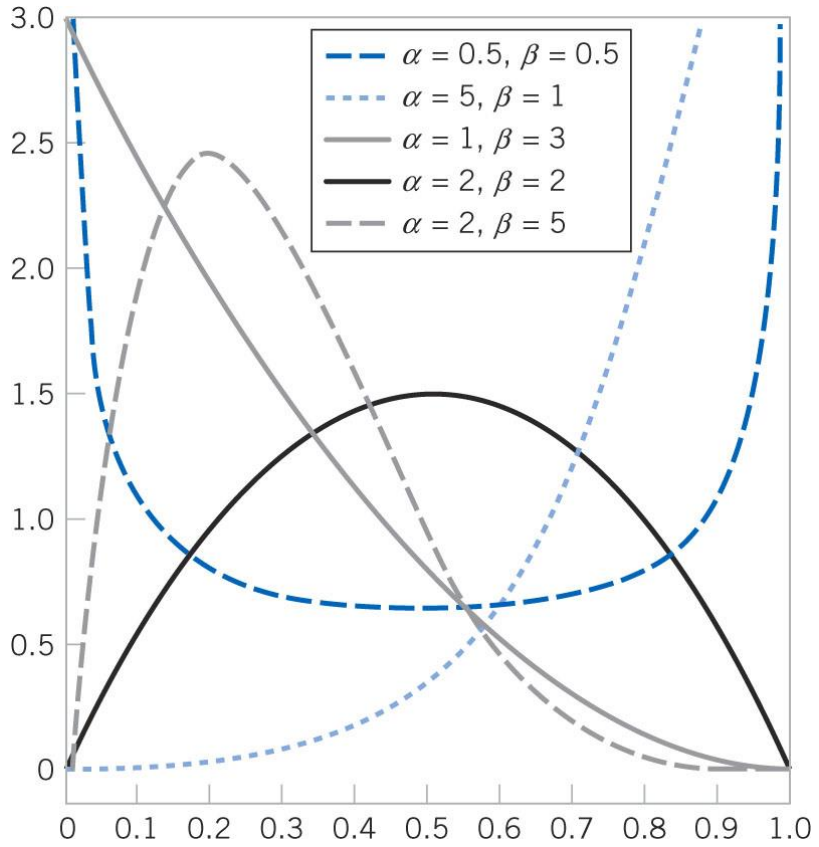


Figure 4-28 Beta probability density functions for selected values of the parameters α and β.

Example 4-27: Beta Computation-1

Consider the completion time of a large commercial real estate development. The proportion of the maximum allowed time to complete a task is a beta random variable with $\alpha = 2.5$ and $\beta = 1$. What is the probability that the proportion of the max time exceeds 0.7? Let *X* denote that proportion.

$$P(X > 0.7) = \int_{0.7}^{1} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\cdot\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}dx$$

$$= \frac{\Gamma(3.5)}{\Gamma(2.5)\cdot\Gamma(1)} \int_{0.7}^{1} x^{1.5}dx$$

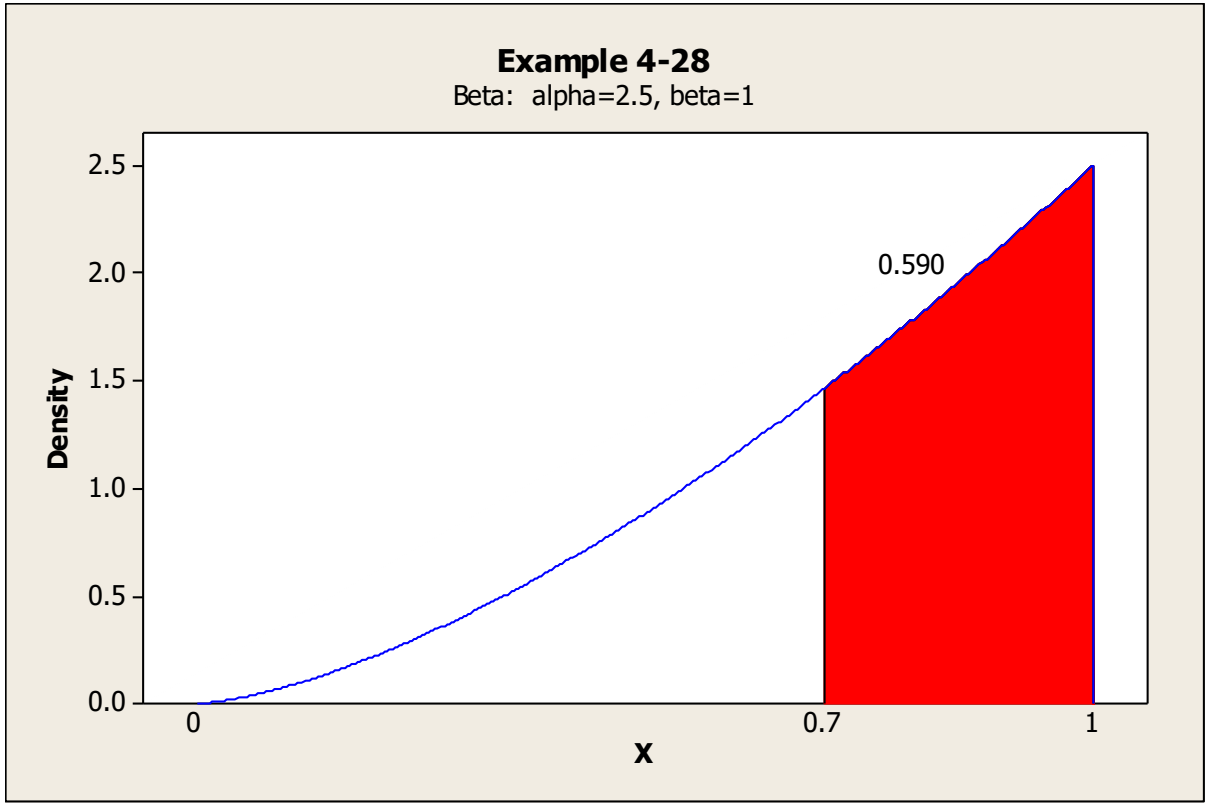
$$= \frac{2.5(1.5)(0.5)\sqrt{\pi}}{1.5(0.5)\sqrt{\pi}\cdot 1} \frac{x^{2.5}}{2.5}\Big|_{0.7}^{1}$$

$$= 1-(0.7)^{2.5} = 0.59$$

Using Excel	
0.590	= 1 - BETADIST(0.7,2.5,1,0,1)

Example 4-27: Beta Computation-2

This Minitab graph illustrates the prior calculation. **FIX**



Mean & Variance of the Beta Distribution

If *X* has a beta distribution with parameters α and β,

$$\mu = E(X) = \frac{\alpha}{\alpha + \beta}$$

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

Example 4-28: In the prior example, α = 2.5 and β = 1. What are the mean and variance of this distribution?

$$\mu = \frac{2.5}{2.5+1} = \frac{2.5}{3.5} = 0.71$$

$$\sigma^2 = \frac{2.5(1)}{(2.5+1)^2(2.5+1+1)} = \frac{2.5}{3.5^2(4.5)} = 0.045$$

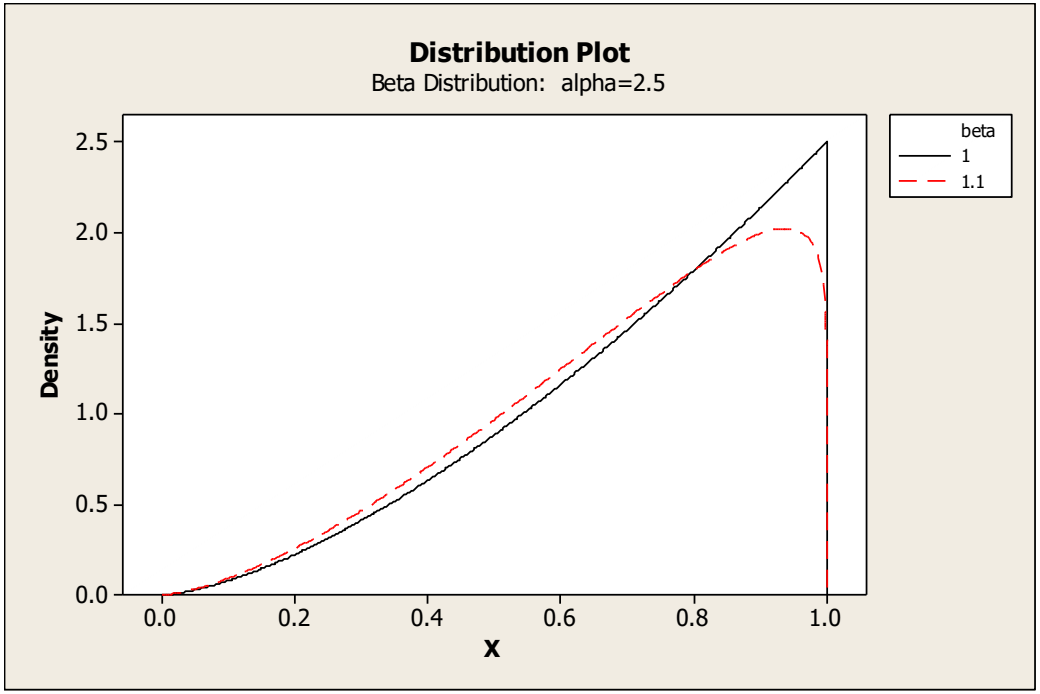
Mode of the Beta Distribution

If α >1 and β > 1, then the beta distribution is mound-shaped and has an interior peak, called the mode of the distribution. Otherwise, the mode occurs at an endpoint.

General formula:

$$\text{Mode} = \frac{\alpha - 1}{\alpha + \beta - 2}$$

for $\alpha > 0$ and $\beta > 0$.



case	alpha	beta	mode	
Example 4-28	2.25	1	1.00	= (2.5-1) / (2.5+1.0-2)
Alternate	2.25	1.1	0.94	= (2.5-1) / (2.5+1.1-2)

Extended Range for the Beta Distribution

The beta random variable X is defined for the [0, 1] interval. That interval can be changed to [a, b]. Then the random variable W is defined as a linear function of X:

$$W = a + (b - a)X$$

With mean and variance:

$$E(W) = a + (b - a)E(X)$$

$$V(W) = (b-a)^2 V(X)$$

Important Terms & Concepts of Chapter 4

Beta distribution
Chi-squared distribution
Continuity correction
Continuous uniform distribution
Cumulative probability distribution for a continuous random variable
Erlang distribution
Exponential distribution
Gamma distribution
Lack of memory property of a continuous random variable
Lognormal distribution
Mean for a continuous random variable
Mean of a function of a continuous random variable
Normal approximation to binomial & Poisson probabilities
Normal distribution
Probability density function
Probability distribution of a continuous random variable
Standard deviation of a continuous random variable
Standardizing
Standard normal distribution
Variance of a continuous random variable
Weibull distribution