Chapter 4. Continuous Probability Distributions

Jae Seok Jang

VIRTUAL REALITY LABORATORY

Objective

- Continuous probability distributions
  - The probability density function, the expectations and variances of the distributions
    - Depends on some parameter values
  - The probability value
    - Straightforward to calculate
    - Use of a software package

In This Chapter…

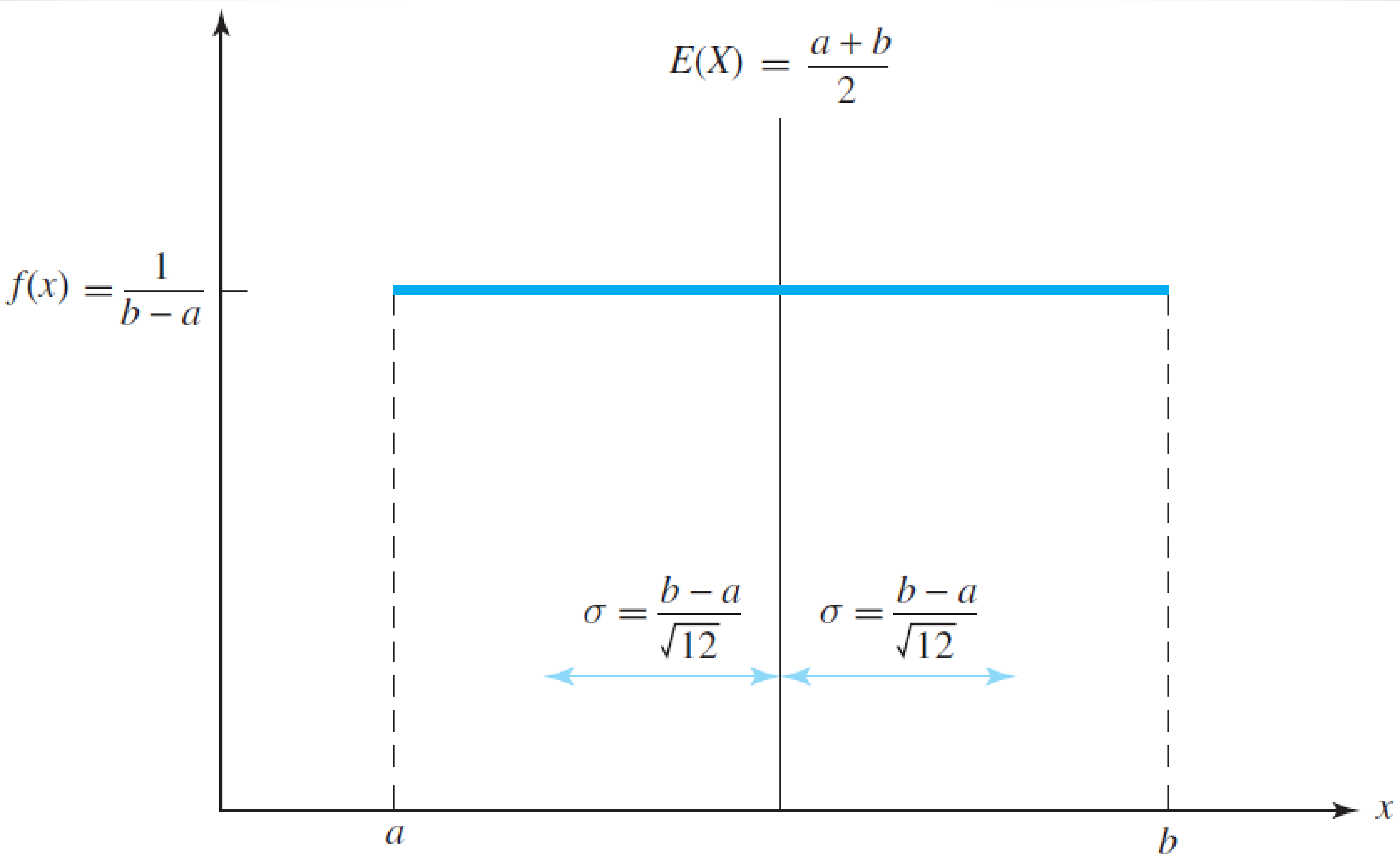
- The uniform distribution
- The exponential distribution
  - Modeling failure rates or waiting times
  - General form
    - The gamma distribution, and the Weibull distribution
  - To discuss Poisson process, simple stochastic process
- The beta distribution
  - Modeling proportions

- Statistical data analysis
  - Approximate to the normal distribution

Uniform Distribution: Definition

- A flat probability density function with bracket(a and b)
- Sometimes, uniform distribution between a and b
  - $X \sim U(a, b)$
  - $\int_a^b f(x)\, dx = 1$
    - The height of the probability functions, $\frac{1}{b-a}$
    - $f(x) = \frac{1}{b-a}$, for $a \le x \le b$, otherwise $f(x) = 0$
  - The cumulative distribution function
    - $F(x) = \int_{y=a}^{y=x} f(y)\, dy = \frac{x-a}{b-a}$, for $a \le x \le b$
  - The probability within a given interval of length δ, $\frac{\delta}{b-a}$

Uniform Distribution: Definition



Uniform Distribution

- The expectation and variance
  - $E(X) = \int_a^b \frac{x}{b-a} dx = \frac{1}{2(b-a)} [x^2]_a^b = \frac{b^2-a^2}{2(b-a)} = \frac{a+b}{2}$
    - Simply, for the symmetric distribution, $E(X) = \frac{a+b}{2}$
  - $E(X^2) = \int_a^b \frac{x^2}{b-a} dx = \frac{1}{3(b-a)} [x^3]_a^b = \frac{b^3-a^3}{3(b-a)} = \frac{(b-a)(b^2+ab+a^2)}{3(b-a)} = \frac{a^2+ab+b^2}{3}$
  - $Var(X) = E(X^2) - E(X)^2 = \frac{a^2+ab+b^2}{3} - \frac{(a+b)^2}{4} = \frac{(b-a)^2}{12}$
- pth quantile
  - $(1-p)a + pb$ →interporlation
  - The interquartile range
    - $(1-0.75)a + 0.75b - (1-0.25)a - 0.25b = \frac{b-a}{2}$
- A standard uniform distribution, $Y \sim U(0,1)$
  - $Y = \frac{X-a}{b-a}$

KNU VIRTUAL REALITY LABORATORY

The Uniform Distribution: Definition

The Uniform Distribution

A random variable X with a flat probability density function between two points a and b, so that

$$f(x) = \frac{1}{b-a}$$

for $a \leq x \leq b$ and $f(x) = 0$ elsewhere, is said to have a **uniform distribution**, which is written $X \sim U(a, b)$. The cumulative distribution function is

$$F(x) = \frac{x-a}{b-a}$$

$$E(X^2) = \int_a^b x^2 \frac{1}{b-a}\, dx = \frac{b^3 - a^3}{3(b-a)} = \frac{a^2 + ab + b^2}{3}$$

for $a \leq x \leq b$, and the expectation and variance are

$$E(X) = \frac{a+b}{2} \qquad \text{and} \qquad \mathrm{Var}(X) = \frac{(b-a)^2}{12}.$$

The Uniform Distribution: Example

- Pearl oyster farming

When pearl oysters are opened, pearls of various sizes are found. Suppose that each oyster contains a pearl with a diameter in mm that has a $U(0, 10)$ distribution. The expected pearl diameter is therefore 5 mm, with a variance of

$$\mathrm{Var}(X) = \frac{(10-0)^2}{12} = 8.33$$

and a standard deviation of $\sigma = \sqrt{8.33} = 2.89$ mm, as shown in Figure 4.2.

  - Commercial value
    - a diameter of at least 4mm
      - $P(X \geq 4) = 1 - F(4) = 1 - 0.4 = 0.6$

Suppose that a farmer retrieves ten oysters out of the water and that the random variable Y represents the number of them containing pearls of commercial value. If the oysters grow pearls independently of one another, Y has a binomial distribution with parameters $n = 10$ and $p = 0.6$, and the probability that at least 8 of the oysters contain pearls of commercial value is

  - $P(Y \geq 8) = ?$

The Exponential Distribution: Definition

The **exponential distribution** has a state space $x \geq 0$ and is often used to model failure or waiting times and interarrival times. It has a probability density function

$$f(x) = \lambda e^{-\lambda x}$$

for $x \geq 0$ and $f(x) = 0$ for $x < 0$, which depends upon a parameter $\lambda > 0$. The cumulative distribution function is

$$F(x) = \int_0^x \lambda e^{-\lambda y}\, dy = 1 - e^{-\lambda x}$$

for $x \geq 0$.

The Exponential Distribution: Definition

- Expectation and variance
  - $E(X) = \int_0^\infty xf(x)dx = \int_0^\infty x\lambda e^{-\lambda x}\, dx$
    - $\frac{dg(x)e^{f(x)}}{dx} = g'(x)e^{f(x)} + g(x)f'(x)e^{f(x)}$
      - $\int g(x)f'(x)e^{f(x)} = g(x)e^{f(x)} - \int g'(x)e^{f(x)}$
    - $-\int_0^\infty x(-\lambda e^{-\lambda x})dx = -\left(xe^{-\lambda x} + \frac{1}{\lambda}e^{-\lambda x}\right)_0^\infty = \frac{1}{\lambda}$
  - $E(X^2) = \int_0^\infty x^2(\lambda e^{-\lambda x})dx = \frac{2}{\lambda^2}$
  - $Var(X) = E(X^2) - E(X)^2 = \frac{1}{\lambda^2}$

The Exponential Distribution: Definition

The Exponential Distribution

An **exponential distribution** with parameter $\lambda > 0$ has a probability density function

$$f(x) = \lambda e^{-\lambda x}$$

for $x \geq 0$ and $f(x) = 0$ for $x < 0$, and a cumulative distribution function

$$F(x) = 1 - e^{-\lambda x}$$

for $x \geq 0$. It is useful for modeling failure times and waiting times. Its expectation and variance are

$$E(X) = \frac{1}{\lambda} \qquad \text{and} \qquad \mathrm{Var}(X) = \frac{1}{\lambda^2}$$

- pth quantile, $F(x) = 1 - e^{-\lambda x} = p$
  - $x = -\frac{\ln(1-p)}{\lambda}$
  - The median of the distribution, $-\frac{\ln(1-0.5)}{\lambda} = E(X)\ln 2$
  - The interval of interquartile
    - $E(X)\ln 3$

The Exponential Distribution: Definition

Figure 4.3 shows the probability density function of an exponential distribution with parameter $\lambda = 1$, and Figure 4.4 shows the probability density function of an exponential distribution with parameter $\lambda = 1/2$. The first distribution has a mean and standard deviation of 1, and the second distribution has a mean and standard deviation equal to 2. Notice that the shapes of the probability density functions are smooth exponential decays.

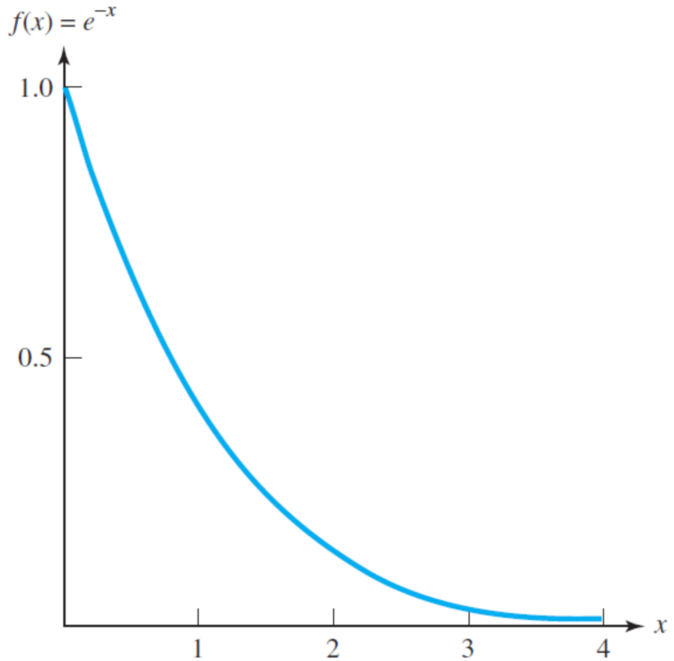


FIGURE 4.3

Probability density function of an exponential distribution with parameter $\lambda = 1$

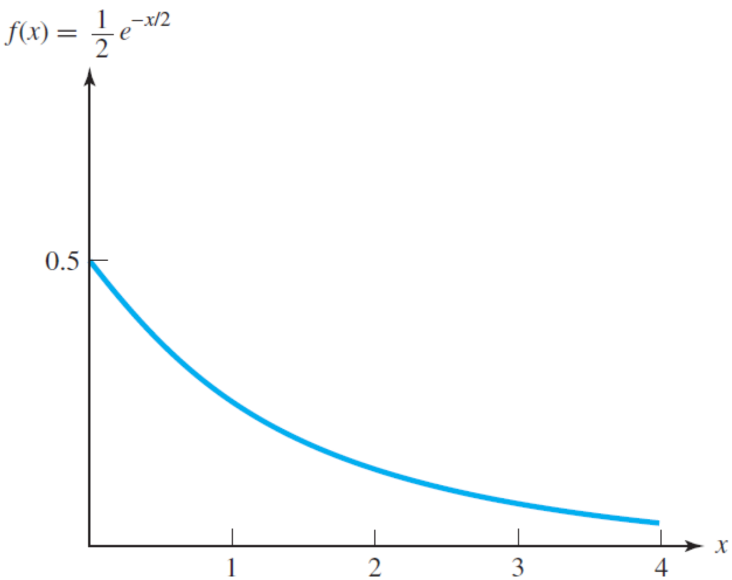


FIGURE 4.4

Probability density function of an exponential distribution with parameter $\lambda = 1/2$

KNU VIRTUAL REALITY LABORATORY

The Exponential Distribution: Memoryless Property

An important aspect of the exponential distribution is its **memoryless property**. This property states that if X has an exponential distribution with parameter λ, then conditional on $X \geq x_0$ for some fixed value x_0, the quantity $X - x_0$ also has an exponential distribution with parameter λ. In other words, if X measures the time until a certain event occurs and the event has not occurred by time x_0, the *additional* waiting time for the event to occur beyond x_0 has the same exponential distribution as X. The process seems to "forget" that a time x_0 has already elapsed and acts as though it is just starting afresh at time zero.

The memoryless property can be shown in the following manner. Notice that if X has an exponential distribution with parameter λ, then

$$P(X \geq x) = 1 - F(x) = e^{-\lambda x}$$

Then if the random variable Y represents the additional time beyond x_0 that elapses before the event occurs,

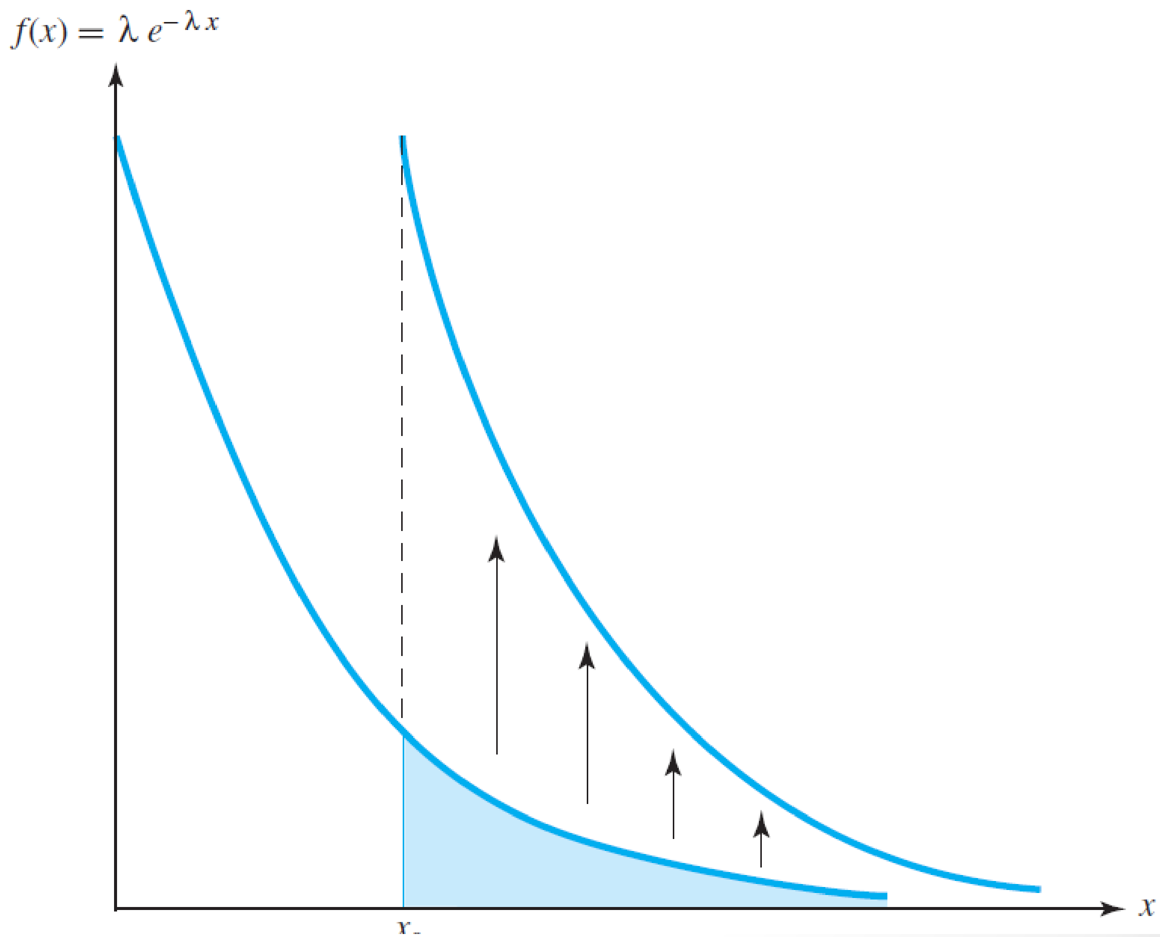
$$P(Y \geq y) = P(X \geq x_0 + y \mid X \geq x_0) = \frac{P(X \geq x_0 + y)}{P(X \geq x_0)} = \frac{e^{-\lambda(x_0+y)}}{e^{-\lambda x_0}} = e^{-\lambda y}$$

so that Y also has an exponential distribution with parameter λ. In graphical terms, the memoryless property follows from the fact that the section of the probability density function of an exponential distribution beyond a certain point x_0 is just a *scaled* version of the whole probability density function, as illustrated in Figure 4.5.

The Exponential Distribution: Memoryless Property

FIGURE 4.5

Illustration of the memoryless property of the exponential distribution. The part of the probability density function beyond x_0 is a **scaled** version of the whole probability density function



The Exponential Distribution: Memoryless Property

The implications of the memoryless property can be rather confusing when first encountered. Suppose that you are waiting at a bus stop and that the time in minutes until the arrival of the bus has an exponential distribution with $\lambda = 0.2$. The expected time that you will wait is consequently $1/\lambda = 5$ minutes. However, if after 1 minute the bus has not yet arrived, what is the expectation of the *additional* time that you must wait?

Unfortunately, it has not been reduced to 4 minutes but is still, as before, 5 minutes. This is because the additional waiting time until the bus arrives beyond the first minute during which you know the bus did not arrive still has an exponential distribution with $\lambda = 0.2$. In fact, as long as the bus has not arrived, no matter how long you have waited, you always have an expected additional waiting time of 5 minutes! This is true right up until the time you first spot the bus coming.

- Application for the exponential distribution
  - The amount of time until an earthquake occurs
  - The amount of time until a new war breaks
  - The amount of time until a telephone call you receive turns out to be a wrong number
    - Waiting time
  - Insurance
    - Prediction of lifetime or accidents

The Exponential Distribution: Memoryless Property

- Example) Suppose that a number of miles that a car run before its battery wears out is exponentially distributed with an average value of 10,000 miles. If a person desires to take a 5,000 mile trip, what is the probability that he will be able to complete his trip without to replace the battery?
  - Let X be a random variable including the remaining lifetime of the battery in thousand miles
    - $E(X) = \frac{1}{\lambda} = 10,\ \lambda = \frac{1}{10}$
    - $P(X > 5) = 1 - F(5) = e^{-5\lambda} = e^{-\frac{1}{2}}$

The Exponential Distribution

- Important property
  - The Poisson model
    - The number of events that occur in a block of time
    - Constant average rate v, the mean after a time t
      - $\lambda = vt$
    - Discrete distribution
      - P(# of events)
  - The Exponential model
    - The time elapsing between events
      - For the time to the first event
    - Continuous distribution
      - P(time till first event)

The Exponential Distribution

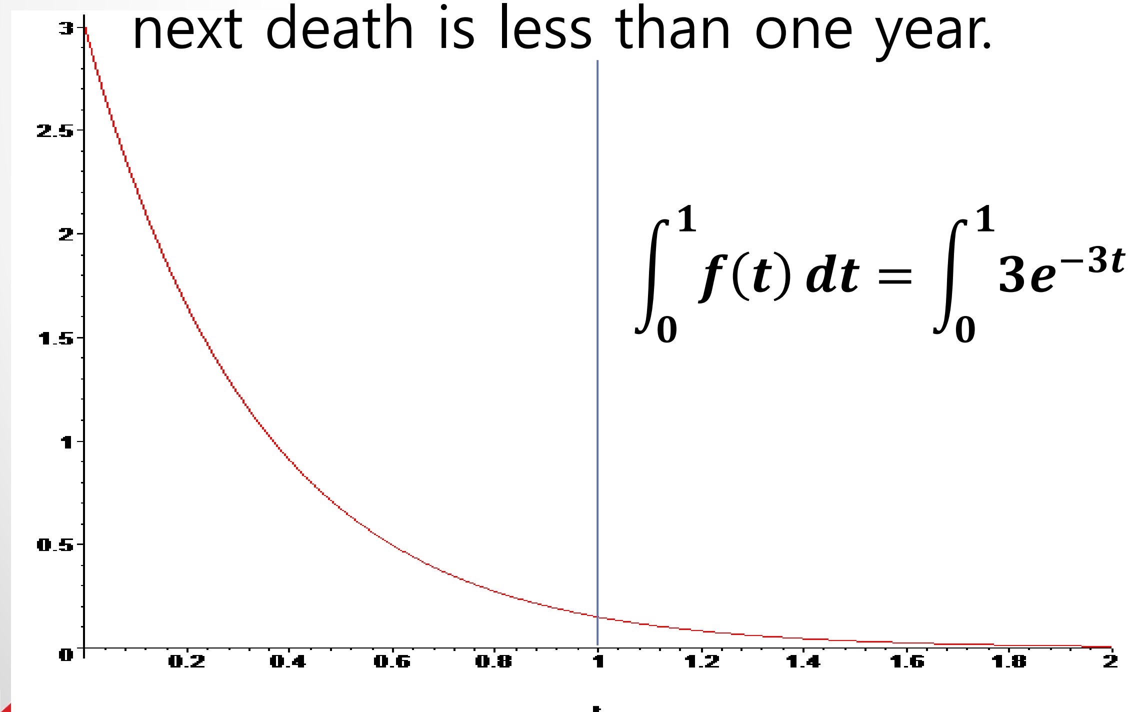
- Example) Which of the following is most likely to be well modelled by a Poisson distribution?
  - Number of trains arriving at Dong-daegu station every hour
  - Number of lottery winners each year that live in Daegu
  - Number of days between solar eclipses
  - Number of days until a component fails

The Exponential Distribution

- Example) Which of the following is most likely to be well modelled by a Poisson distribution?
  - Number of trains arriving at Dong-daegu station every hour
    - No, arrive regularly on a timetable not at random
  - Number of lottery winners each year that live in Daegu
    - Yes, number of random events in fixed interval
  - Number of days between solar eclipses
    - No, solar eclipses are not random events and this is a time between random events, not the number of some fixed interval
  - Number of days until a component fails
    - No, random events, but this is time until a random event, not the number of random events

The Exponential Distribution

- Example) On average lightening kills three people each year in the UK, $\lambda = 3$. So, the rate is $v = 3/year$. Assuming strikes occur randomly at any time during the year so v is constant, time from today until the next fatality has the probability density function $f(t) = ve^{-vt} = 3e^{-3t}$, where t is in years. Calculate the probability value for the time till the next death is less than one year.



$$\int_0^1 f(t)\,dt = \int_0^1 3e^{-3t}\,dt = -\left[e^{-3t}\right]_0^1 = 1 - e^{-3}$$

The Exponential Distribution

- Example) Reliability
  - The time till failure of an electronic component has an Exponential distribution and it is known that 10% of components have failed by 1000 hours.
    - What is the probability that a component is still working after 5000 hours?
    - Find the mean and standard deviation of the time till failure

The Exponential Distribution

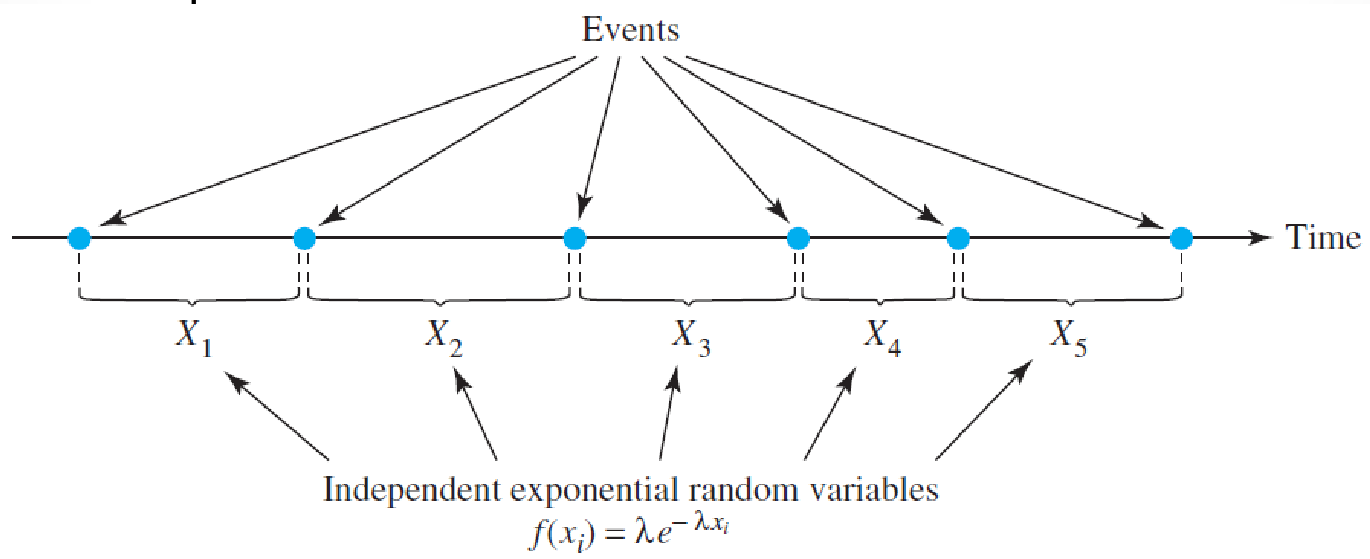
- Example) Reliability
  - The time till failure of an electronic component has an Exponential distribution and it is known that 10% of components have failed by 1000 hours.
    - What is the probability that a component is still working after 5000 hours?
      - Let X be a random variable which is the time for the first failure in thousand hour
      - Firstly, we have to calculate λ
        - $P(X \leq 1) = \int_0^1 \lambda e^{-\lambda x}\, dx = -\left[e^{-\lambda x}\right]_0^1 = 1 - e^{-\lambda} = 0.1$
          - $\lambda = \ln \frac{10}{9} \approx 1.054 \times 10^{-1}$
        - $P(X \geq 5) = \int_5^\infty \lambda e^{-\lambda x}\, dx = -\left[e^{-\lambda x}\right]_5^\infty = e^{-5\lambda} \approx 0.594$

The Exponential Distribution

- Example) Reliability
  - The time till failure of an electronic component has an Exponential distribution and it is known that 10% of components have failed by 1000 hours.
    - Find the mean and standard deviation of the time till failure
      - $E(X) = \frac{1}{\lambda} = \frac{1}{\ln 10 - \ln 9} \approx 10000/1054 \approx 9.4877$
      - $\sigma(X) = \sqrt{Var(X)} = \frac{1}{\lambda} = E(X)$

The Exponential Distribution: The Poisson Process

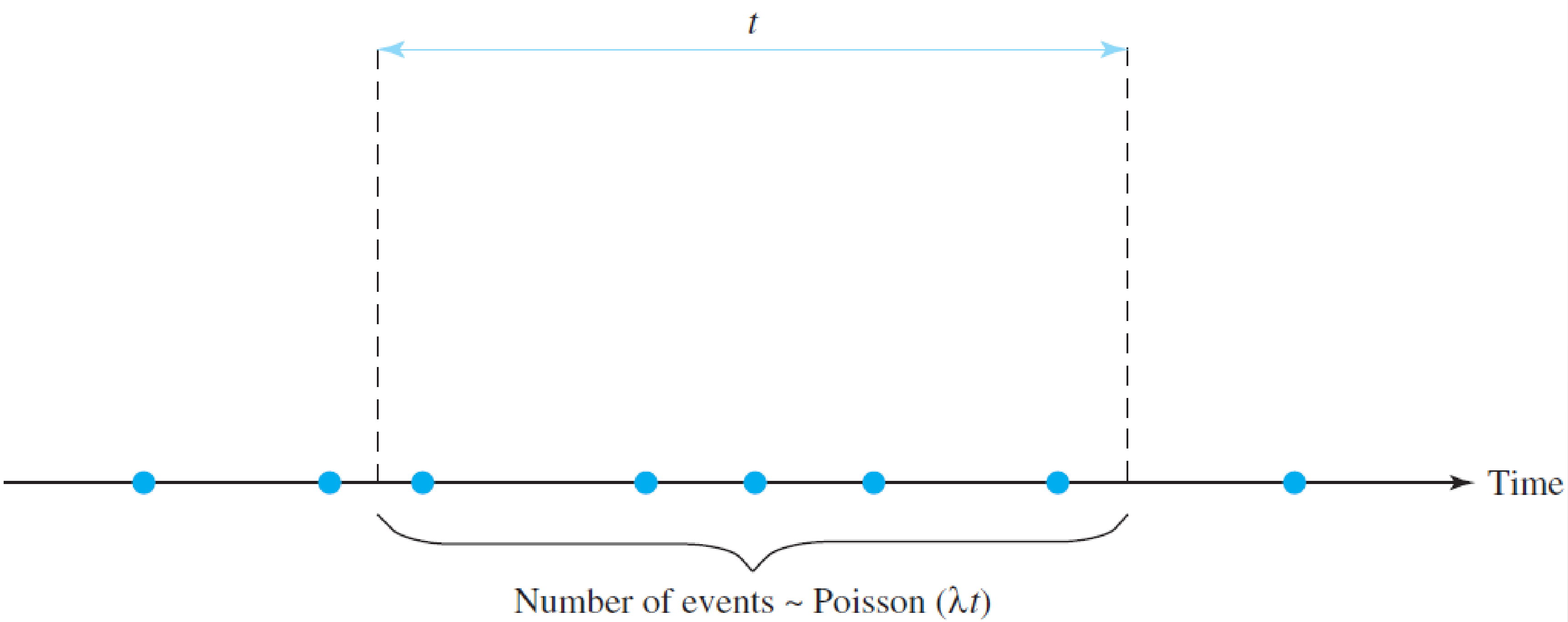
- A stochastic process
  - A series of random events
  - A sequence of events occurring over time with specified distributions of the time intervals between the occurrences of adjacent events
  - Poisson process
    - The arrival of calls at a switchboard
    - The addition of new elements to a queue
    - The positions of deformities within a substance



The Exponential Distribution: The Poisson Process

The expected waiting time between two events in a Poisson process is $1/\lambda$ because it is simply the expected value of an exponential distribution with parameter λ. Furthermore, the expected number of events occurring within a fixed time interval of length t is λt. Moreover, the number of events occurring within such a time interval has a Poisson distribution with mean λt. In other words, if the random variable X counts the number of events occurring within a fixed time interval of length t, then

$$X \sim P(\lambda t)$$



The Exponential Distribution: Example

- Shipwreck hunts

A team of underwater salvage experts sets sail to search the ocean floor for the wreckage of a ship that is thought to have sunk within a certain area. Their boat is equipped with underwater sonar with which they hope to detect unusual objects lying on the ocean floor.

The captain's experience is that in similar situations it has taken an average of 20 days to locate a wreck. Consequently, the captain surmises that the time in days taken to locate the wreck can be modeled by an exponential distribution with parameter

$$\lambda = \frac{1}{E(X)} = \frac{1}{20} = 0.05$$

The captain considers the memoryless property of the exponential distribution to be suitable since, with such vast areas of the ocean floor to be searched, the unfruitful searching of certain areas does not appreciably alter the chance of finding the wreck in the future.

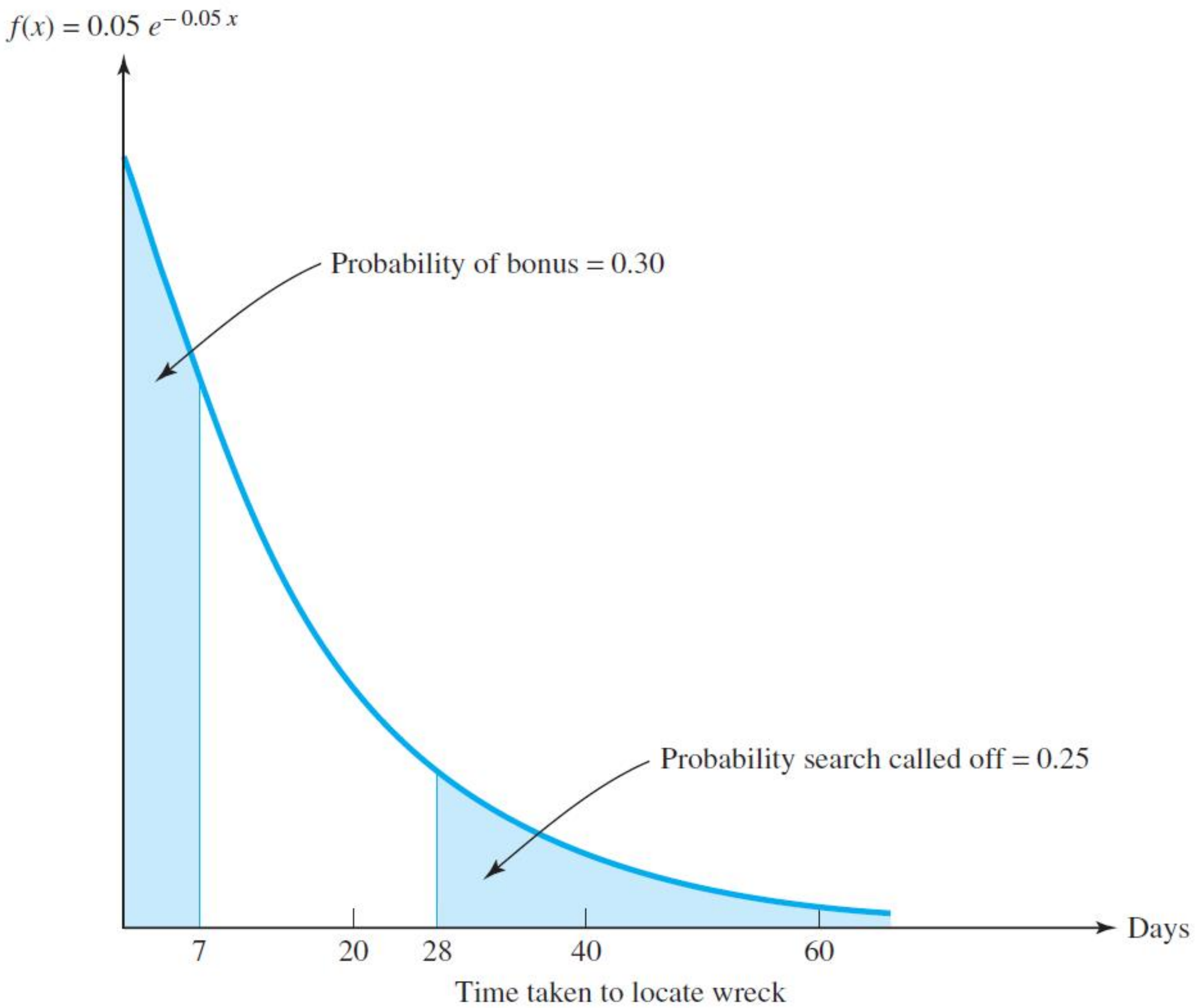
The captain's contractors have offered a sizeable bonus if it is possible to reduce searching costs by locating the wreck within the first week. The captain estimates the probability of this to be

$$P(X \leq 7) = F(7) = 1 - e^{-0.05\times 7} = 0.30$$

The Exponential Distribution: Example

FIGURE 4.8

Probability density function for shipwreck hunt



The Exponential Distribution: Example

- Streel girder fractures

An engineer examines the edges of steel girders for hairline fractures. The girders are 10 m long, and it is discovered that they have an average of 42 fractures each. If a girder has 42 fractures, then there are 43 "gaps" between fractures or between the ends of the girder and the adjacent fractures. The average length of these gaps is therefore $10/43 = 0.23$ m. The fractures appear to be randomly spaced on the girders, so the engineer proposes that the location of fractures on a particular girder can be modeled by a Poisson process with

$$\lambda = \frac{1}{0.23} = 4.3$$

According to this model, the length of a gap between any two adjacent fractures has an exponential distribution with $\lambda = 4.3$, as illustrated in Figure 4.9. In this case, the probability that a gap is less than 10 cm long is

$$P(X \le 0.10) = F(0.10) = 1 - e^{-4.3 \times 0.10} = 0.35$$

The probability that a gap is longer than 30 cm is

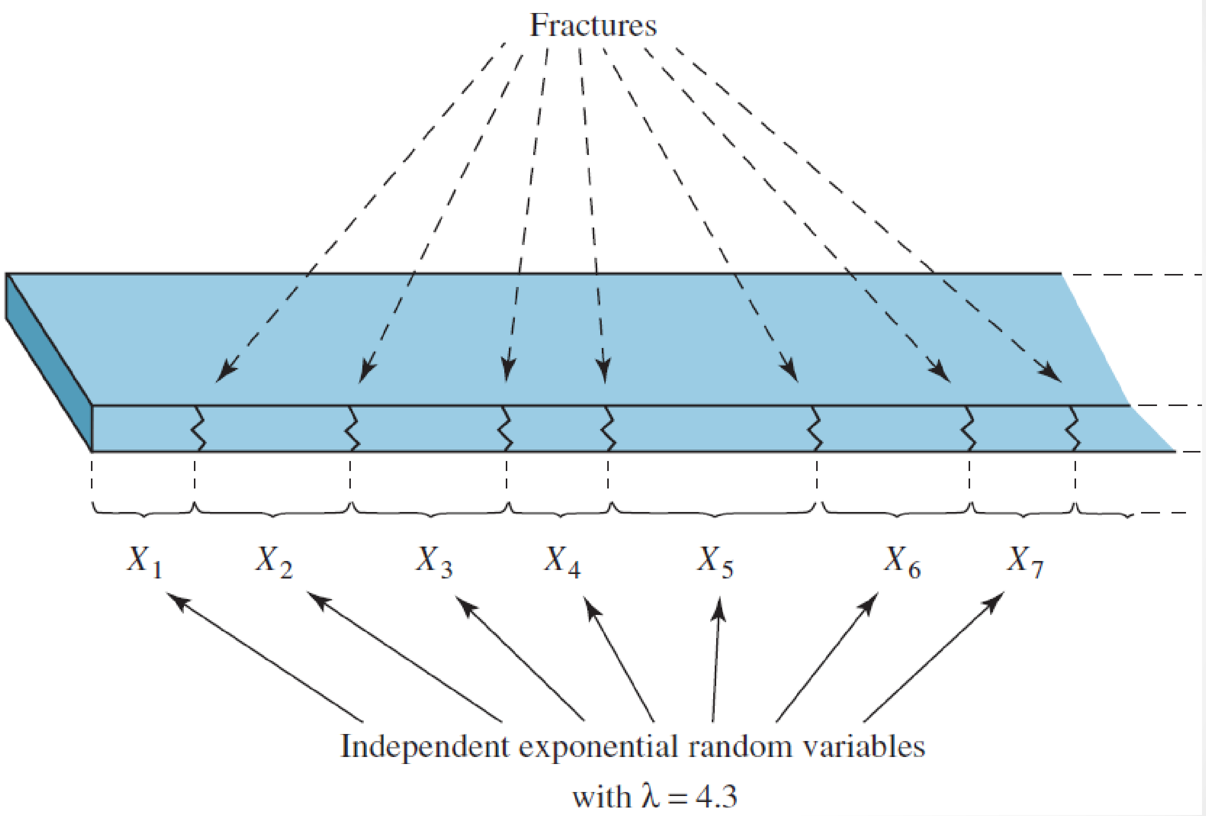
$$P(X \ge 0.30) = 1 - F(0.30) = e^{-4.3 \times 0.30} = 0.28$$

The Exponential Distribution: Example

- Streel girder fractures

FIGURE 4.9

Poisson process modeling fracture locations on a steel girder



The Exponential Distribution: Example

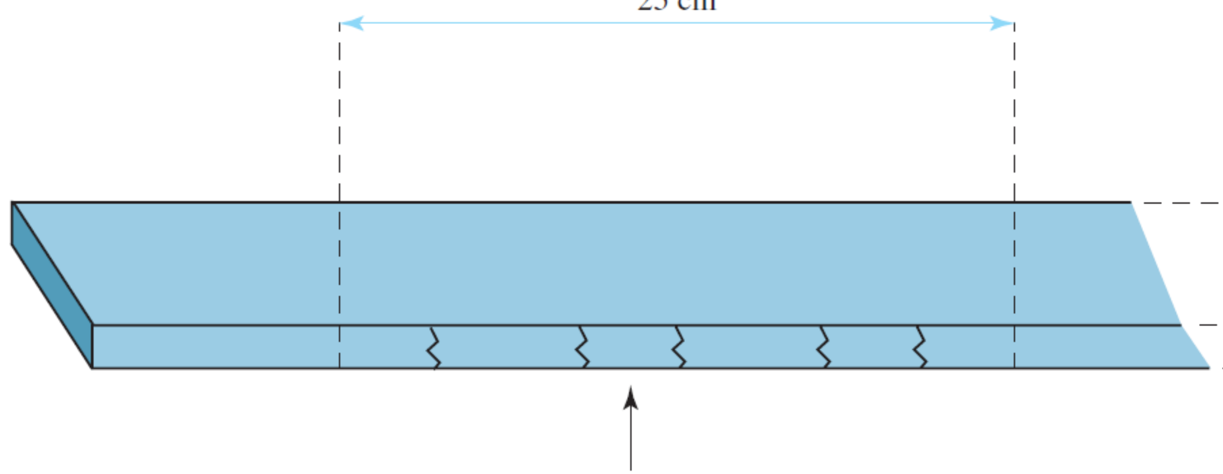


FIGURE 4.10

The number of fractures in a 25-cm segment of the steel girder has a Poisson distribution with mean 1.075

If a 25-cm segment of a girder is selected, the number of fractures it contains has a Poisson distribution with mean

$$\lambda \times 0.25 = 4.3 \times 0.25 = 1.075$$

as illustrated in Figure 4.10. The probability that the segment contains at least two fractures is therefore

$$\begin{aligned} P(X \geq 2) &= 1 - P(X = 0) - P(X = 1) \\ &= 1 - \frac{e^{-1.075} \times 1.075^0}{0!} - \frac{e^{-1.075} \times 1.075^1}{1!} \\ &= 1 - 0.341 - 0.367 = 0.292 \end{aligned}$$

The Exponential Distribution: Example

- Car body assembly line

The engineer in charge of the car panel manufacturing process pays particular attention to the arrival of metal sheets at the beginning of the panel construction lines. These metal sheets are brought one by one from other parts of the factory floor, where they have been cut into the required sizes. On average, about 96 metal sheets are delivered to the panel construction lines in 1 hour.

The engineer decides to model the arrival of the metal sheets with a Poisson process. The average waiting time between arrivals is $60/96 = 0.625$ minute, so a value of

$$\lambda = \frac{1}{0.625} = 1.6$$

is used. This model assumes that the waiting times between arrivals of metal sheets are independently distributed as exponential distributions with $\lambda = 1.6$, as shown in Figure 4.11. For example, the probability that there is a wait of more than 3 minutes between arrivals is

$$P(X \geq 3) = 1 - F(3) = e^{-1.6\times 3} = 0.008$$

The Exponential Distribution: Example

- Car body assembly line

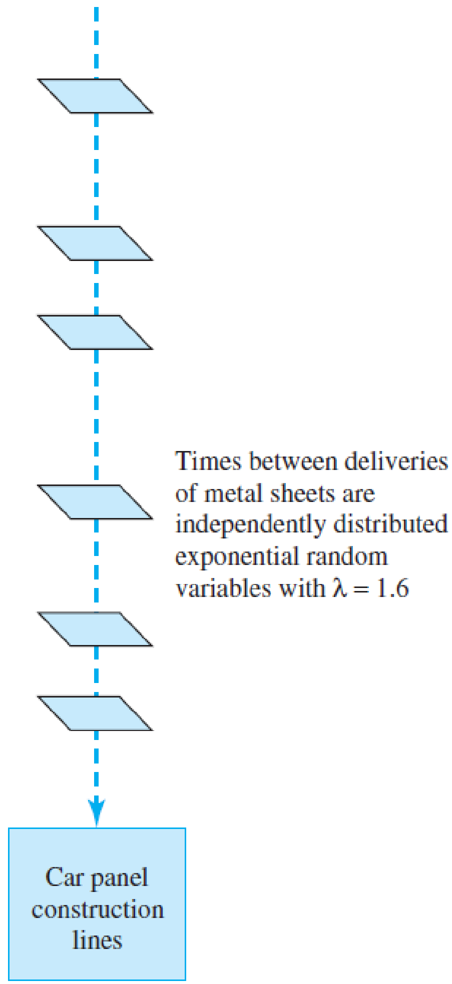


FIGURE 4.11

Poisson process modeling arrival times of metal sheets to car panel construction lines

$X \sim$ Poisson (24.0)		$X \sim$ Poisson (24.0)	
x	$P(x \leq x)$	x	$P(x \leq x)$
7	0.0000	27	0.7677
8	0.0002	28	0.8225
9	0.0004	29	0.8679
10	0.0011	30	0.9042
11	0.0025	31	0.9322
12	0.0054	32	0.9533
13	0.0107	33	0.9686
14	0.0198	34	0.9794
15	0.0344	35	0.9868
16	0.0563	36	0.9918
17	0.0871	37	0.9950
18	0.1283	38	0.9970
19	0.1803	39	0.9983
20	0.2426	40	0.9990
21	0.3139	41	0.9995
22	0.3917	42	0.9997
23	0.4728	43	0.9998
24	0.5540	44	0.9999
25	0.6319	45	1.0000
26	0.7038		

FIGURE 4.12

The cumulative distribution function of a Poisson random variable with mean 24.0

The Exponential Distribution: Example

- Car body assembly line

The number of metal sheets arriving at the panel construction lines during a specific 15-minute period has a Poisson distribution with mean

$$\lambda \times 15 = 1.6 \times 15 = 24.0$$

Figure 4.12 shows the cumulative distribution function of this Poisson distribution. The probability that no more than 16 sheets arrive during the 15-minute period, for example, is about 0.056. On the other hand, the engineer can be about 95% confident that no more than 32 metal sheets will arrive during the period under consideration.

The Gamma Distribution: Definition

The **gamma distribution** has many important applications in areas such as reliability theory, and it is also used in the analysis of a Poisson process. It has a state space $x \geq 0$ and a probability density function

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

for $x \geq 0$ and $f(x) = 0$ for $x < 0$, which depends upon two parameters $k > 0$ and $\lambda > 0$. The function $\Gamma(k)$ is known as the **gamma function**. It provides the correct scaling to ensure that the total area under the probability density function is equal to 1.

Notice that if $k = 1$, the gamma distribution simplifies to the exponential distribution with parameter λ. The expectation and variance of a gamma distribution are given in the following box.

- Gamma function $\Gamma(k)$
  - $\Gamma(k) = \int_0^\infty x^{k-1} e^{-x}\, dx$
  - Some special case, $\Gamma(1) = 1$ and $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$
  - In general, $\Gamma(k) = (k-1)\Gamma(\mathrm{k}-1)$
  - $\Gamma(k) = (k-1)!$ For $k > 1$ and n is a positive integer
  - Except for special cases, in general no closed-form expression

The Gamma Distribution: Definition

- Expectation and variance
  - $E(X) = \int_0^\infty \frac{x\lambda^k x^{k-1} e^{-\lambda x}}{\Gamma(k)} dx = \frac{\lambda^k}{\Gamma(k)} \int_0^\infty x^k e^{-\lambda x}\, dx,\ t = \lambda x$
    - $\frac{\lambda^k}{\Gamma(k)} \int_0^\infty x^k e^{-\lambda x}\, dx = \frac{\lambda^k}{\Gamma(k)} \int_0^\infty \left(\frac{t}{\lambda}\right)^k e^{-t} \frac{1}{\lambda} dt = \frac{\lambda^k}{\lambda^{k+1}\Gamma(k)} \int_0^\infty t^k e^{-t}\, dt$
    - $\frac{\lambda^k}{\lambda^{k+1}\Gamma(k)} \int_0^\infty t^k e^{-t}\, dt = \frac{\Gamma(k+1)}{\lambda\Gamma(k)} = \frac{k\Gamma(k)}{\lambda\Gamma(k)} = \frac{k}{\lambda}$
  - $E(X^2) = \int_0^\infty \frac{\lambda^k x^{k+1} e^{-\lambda x}}{\Gamma(k)} dx = \frac{\lambda^k}{\Gamma(k)} \int_0^\infty x^{k+1} e^{-\lambda x}\, dx,\ t = \lambda x$
    - $\frac{\lambda^k}{\Gamma(k)} \int_0^\infty x^{k+1} e^{-\lambda x}\, dx = \frac{\lambda^k}{\Gamma(k)} \int_0^\infty \left(\frac{t}{\lambda}\right)^{k+1} e^{-t} \frac{1}{\lambda} dt =$
      $\frac{1}{\lambda^2\Gamma(k)} \int_0^\infty t^{k+1} e^{-t}\, dt = \frac{\Gamma(k+2)}{\lambda^2\Gamma(k)} = \frac{k(k+1)}{\lambda^2}$
  - $Var(X) = E(X^2) - E(X)^2 = \frac{k^2+k}{\lambda^2} - \frac{k^2}{\lambda^2} = \frac{k}{\lambda^2}$

VIRTUAL REALITY LABORATORY

The Gamma Distribution: Definition

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

The Gamma Distribution

A **gamma distribution** with parameters $k > 0$ and $\lambda > 0$ has a probability density function

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

$$\begin{aligned} \boldsymbol{F(X)} &= \boldsymbol{P(X \le x) = 1 - P(X > x)} \\ &= \boldsymbol{1 - P(\#\ of\ events\ in\ [0, x] < \alpha)} \\ &= \boldsymbol{1 - P(Y \le \alpha - 1),\ Y \sim P(\lambda x)} \end{aligned}$$

for $x \ge 0$ and $f(x) = 0$ for $x < 0$, where $\Gamma(k)$ is the gamma function. It has an expectation and variance of

$$E(X) = \frac{k}{\lambda} \quad \text{and} \quad \mathrm{Var}(X) = \frac{k}{\lambda^2}$$

- Extension form of exponential distribution
  - Distribution for waiting time till n-times occurrences in Poisson process
  - Distribution for waiting time till n-th event occurrence in exponential distribution
  - Similar to negative distribution in binomial distribution

The Gamma Distribution: Definition

The parameter k is often referred to as the *shape* parameter of the gamma distribution, and λ is referred to as the *scale* parameter. Figure 4.13 shows the probability density functions of gamma distributions with $\lambda = 1$ and $k = 1, 3$, and 5. As the shape parameter increases, the peak of the density function is seen to move farther to the right. Figure 4.14 shows the probability density functions of gamma distributions with $\lambda = 1, 2$, and 3 and $k = 3$. This illustrates how the parameter λ "scales" the distribution function.

FIGURE 4.13

Probability density functions of gamma distributions

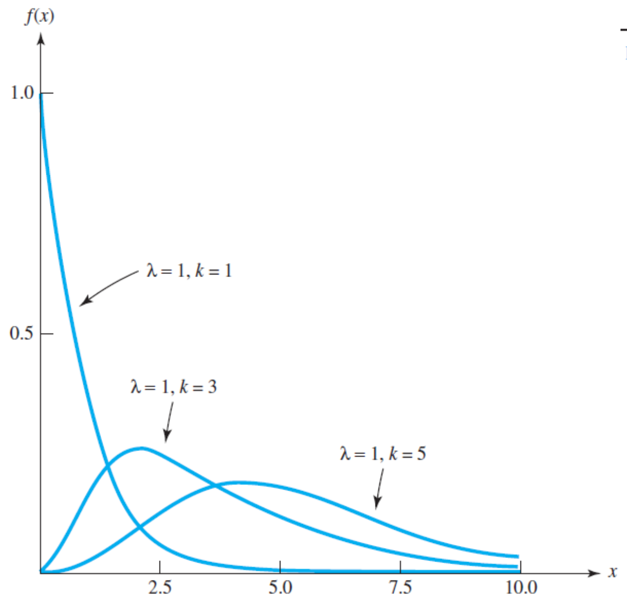
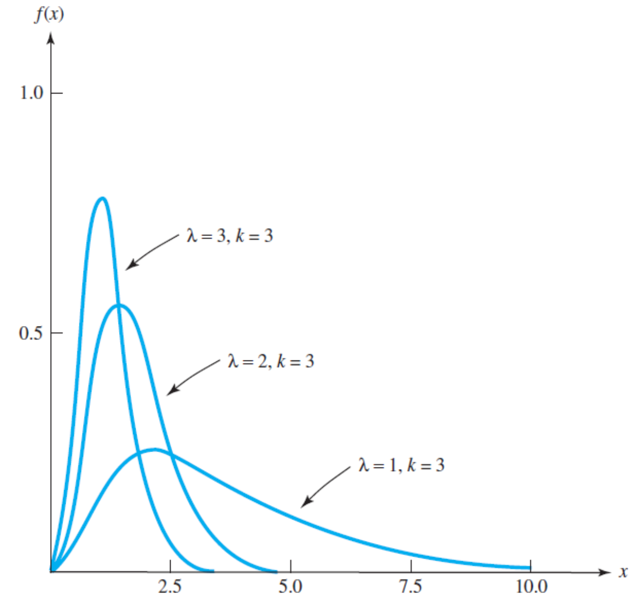


FIGURE 4.14

Probability density functions of gamma distributions



The Gamma Distribution: Properties

- If X_i, $i = 1, \ldots, n$ are independent gamma random variables with respect parameters (k_i, λ)
  - $\sum_{i=1}^{n} X_i$ is a gamma random variable with parameter $(\sum_{i=1}^{n} k_i, \lambda)$
- If X_i, $i = 1, \ldots, n$ are independent exponential random variables, each having rate λ
  - $\sum_{i=1}^{n} X_i$ is a gamma random variable with parameter (n, λ)

The Gamma Distribution:

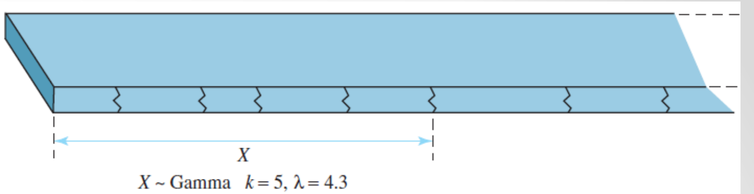


FIGURE 4.15
Distance to fifth fracture has a gamma distribution with parameters $k = 5$ and $\lambda = 4.3$

- Streel girder fractures

Suppose that the random variable X measures the length between one end of a girder and the *fifth* fracture along the girder, as shown in Figure 4.15. If the fracture locations are modeled by a Poisson process as discussed previously, X has a gamma distribution with parameters $k = 5$ and $\lambda = 4.3$. The expected distance to the fifth fracture is therefore

$$E(X) = \frac{k}{\lambda} = \frac{5}{4.3} = 1.16 \text{ m}$$

A software package can be used to show that the 0.05 quantile point of this distribution is $x = 0.458$ m, so that

$$F(0.458) = 0.05$$

Consequently, the engineer can be 95% sure that the fifth fracture is at least 46 cm away from the end of the girder. A software package can also be used to calculate the probability that the fifth fracture is within 1 m of the end of the girder, which is

$$F(1) = 0.4296$$

It is interesting to note that this latter probability can also be obtained using the Poisson distribution. The number of fractures within a 1-m section of the girder has a Poisson distribution with mean

$$\lambda \times 1 = 4.3$$

The probability that the fifth fracture is within 1 m of the end of the girder is the probability that there are at least five fractures within the first 1-m section, which is therefore

$$P(Y \geq 5) = 0.4296$$

where $Y \sim P(4.3)$.

The Gamma Distribution: Example

- Car body assembly line

Suppose that the engineer in charge of the car panel manufacturing process is interested in how long it will take for 20 metal sheets to be delivered to the panel construction lines. Under the Poisson process model, this time X has a gamma distribution with parameters $k = 20$ and $\lambda = 1.6$. The expected waiting time is consequently

$$E(X) = \frac{k}{\lambda} = \frac{20}{1.6} = 12.5 \text{ minutes}$$

The variance of the waiting time is

$$\text{Var}(X) = \frac{k}{\lambda^2} = \frac{20}{1.6^2} = 7.81$$

so that the standard deviation is $\sigma = \sqrt{7.81} = 2.80$ minutes, as illustrated in Figure 4.16.

A software package can be used to show that for this distribution,

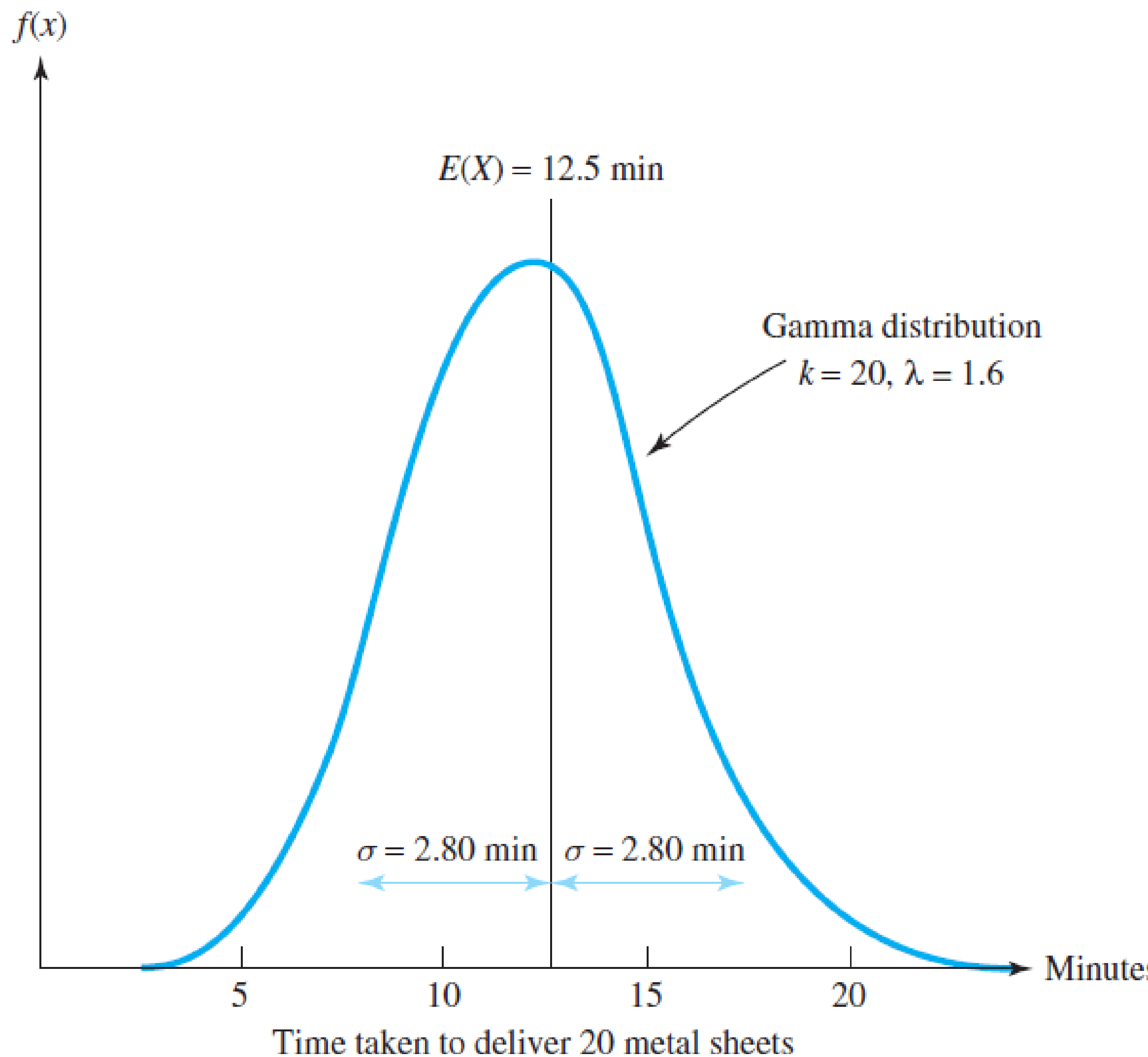
$$F(17.42) = 0.95 \qquad \text{and} \qquad F(15) = 0.8197$$

The engineer can therefore be 95% confident that 20 metal sheets will have arrived within 18 minutes, say. Furthermore, there is a probability of about 0.82 that they will all arrive within 15 minutes. This latter probability can also be obtained from the probabilities of a Poisson distribution with mean 24.0, which is shown in Figure 4.12. This Poisson distribution is the distribution of the number of sheets arriving during a 15-minute period, and $1 - 0.1803 = 0.8197$ is seen to be the probability that at least 20 sheets arrive during this time interval.

The Gamma Distribution: Example

FIGURE 4.16

Probability density function of the time taken to deliver 20 metal sheets to the car panel construction lines



The Weibull Distribution: Definition

The **Weibull distribution** is often used to model failure and waiting times. It has a state space $x \geq 0$ and a probability density function

$$f(x) = a\,\lambda^a\,x^{a-1}\,e^{-(\lambda x)^a}$$

for $x \geq 0$ and $f(x) = 0$ for $x < 0$, which depends upon two parameters $a > 0$ and $\lambda > 0$. Notice that taking $a = 1$ gives the exponential distribution as a special case.

The cumulative distribution function of a Weibull distribution is

$$F(x) = \int_0^x a\,\lambda^a\,y^{a-1}\,e^{-(\lambda y)^a}\,dy = 1 - e^{-(\lambda x)^a}$$

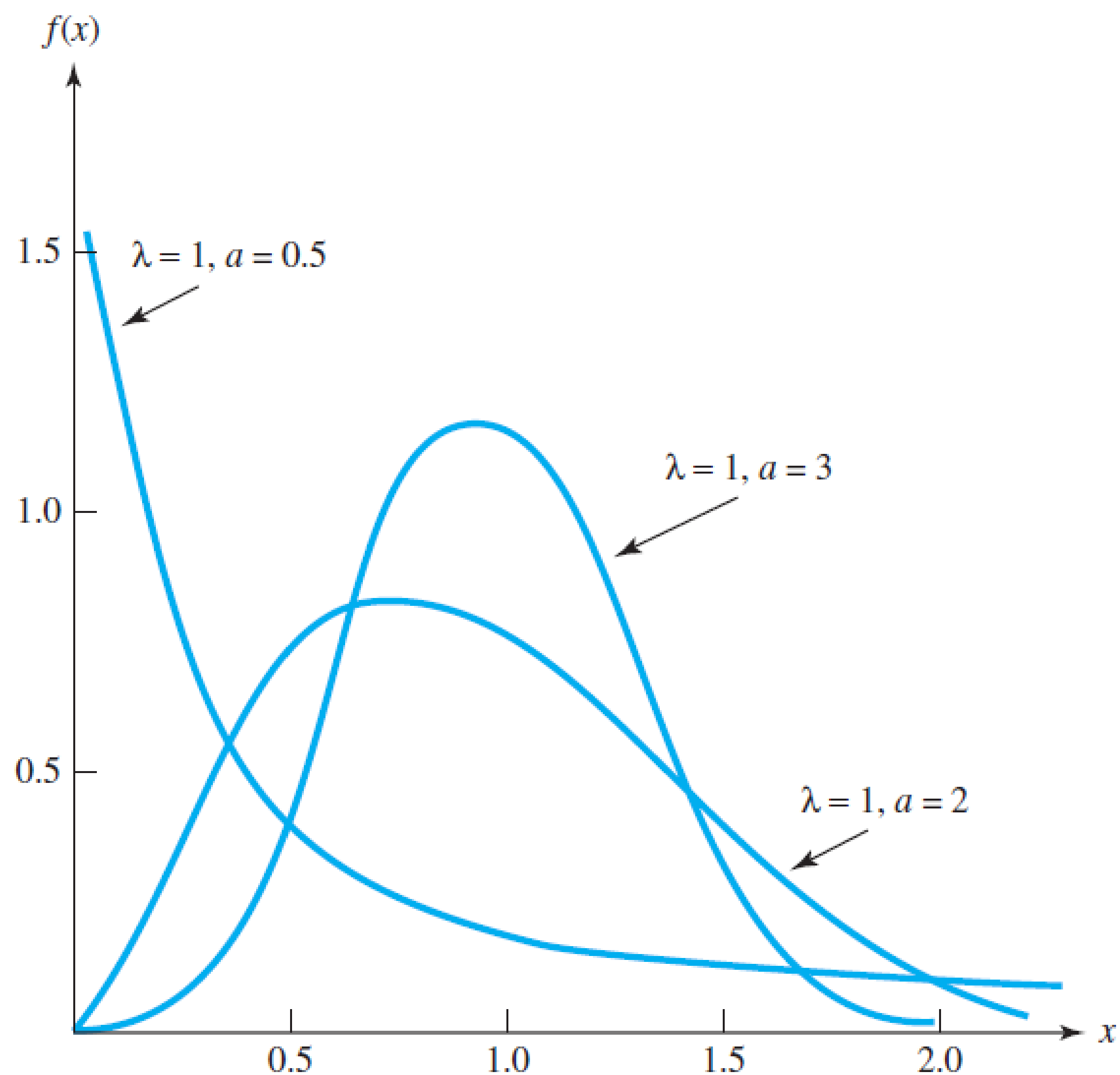
for $x \geq 0$. The expectation and variance of a Weibull distribution depend upon the gamma function and are given in the following box.

As with the gamma distribution, λ is called the *scale* parameter of the distribution, and a is called the *shape* parameter. A useful property of the Weibull distribution is that the probability density function can exhibit a wide variety of forms, depending on the choice of the two parameters. Figure 4.17 illustrates some probability density functions with $\lambda = 1$ and various values of the shape parameter a. Figure 4.18 illustrates some probability density functions with $a = 3$ and with various values of λ.

The Weibull Distribution: Definition

FIGURE 4.17

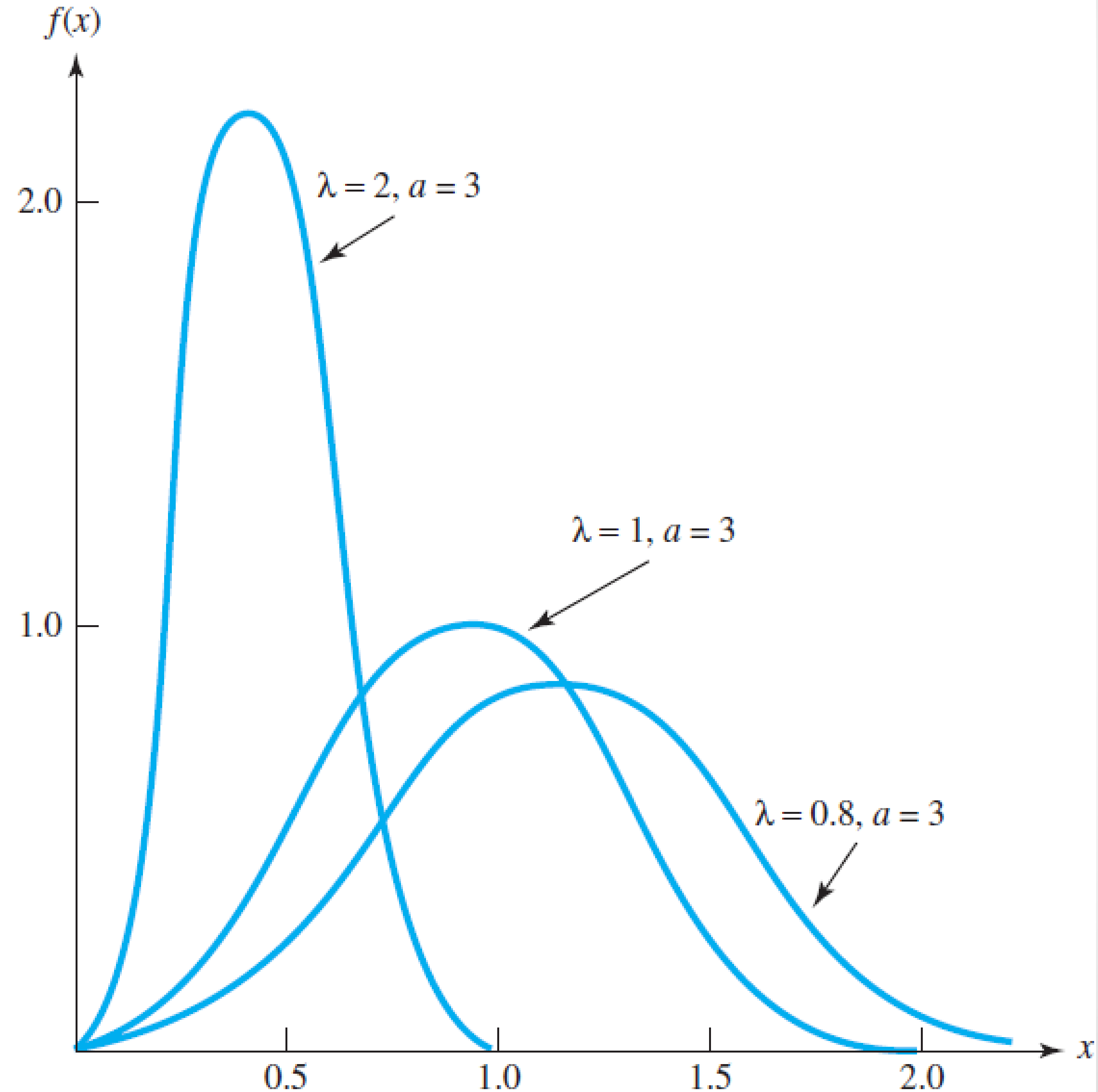
Probability density functions of the Weibull distribution



The Weibull Distribution: Definition

FIGURE 4.18

Probability density functions of the Weibull distribution



The Weibull Distribution: Example

- Bacteria lifetimes

Suppose that the random variable X measures the lifetime of a bacterium at a certain high temperature, and that it has a Weibull distribution with $a = 2$ and $\lambda = 0.1$. This distribution is illustrated in Figure 4.19.

The expected survival time of a bacterium is

$$E(X) = \frac{1}{0.1} \times \Gamma\left(1 + \frac{1}{2}\right) = 10 \times \frac{1}{2} \times \Gamma\left(\frac{1}{2}\right) = 10 \times \frac{1}{2} \times \sqrt{\pi} = 8.86 \text{ minutes}$$

The variance of the bacteria lifetimes is

$$\mathrm{Var}(X) = \frac{1}{0.1^2} \times \left\{\Gamma\left(1 + \frac{2}{2}\right) - \Gamma\left(1 + \frac{1}{2}\right)^2\right\}$$

$$= 100 \times \left\{1 - \left(\frac{\sqrt{\pi}}{2}\right)^2\right\} = 21.46$$

so that the standard deviation is $\sigma = \sqrt{21.46} = 4.63$ minutes.

- The probability that a bacterium dies within 5 minutes?
- The probability that a bacterium lives longer than 15 minutes?

The Weibull Distribution: Example

- Bacteria lifetimes

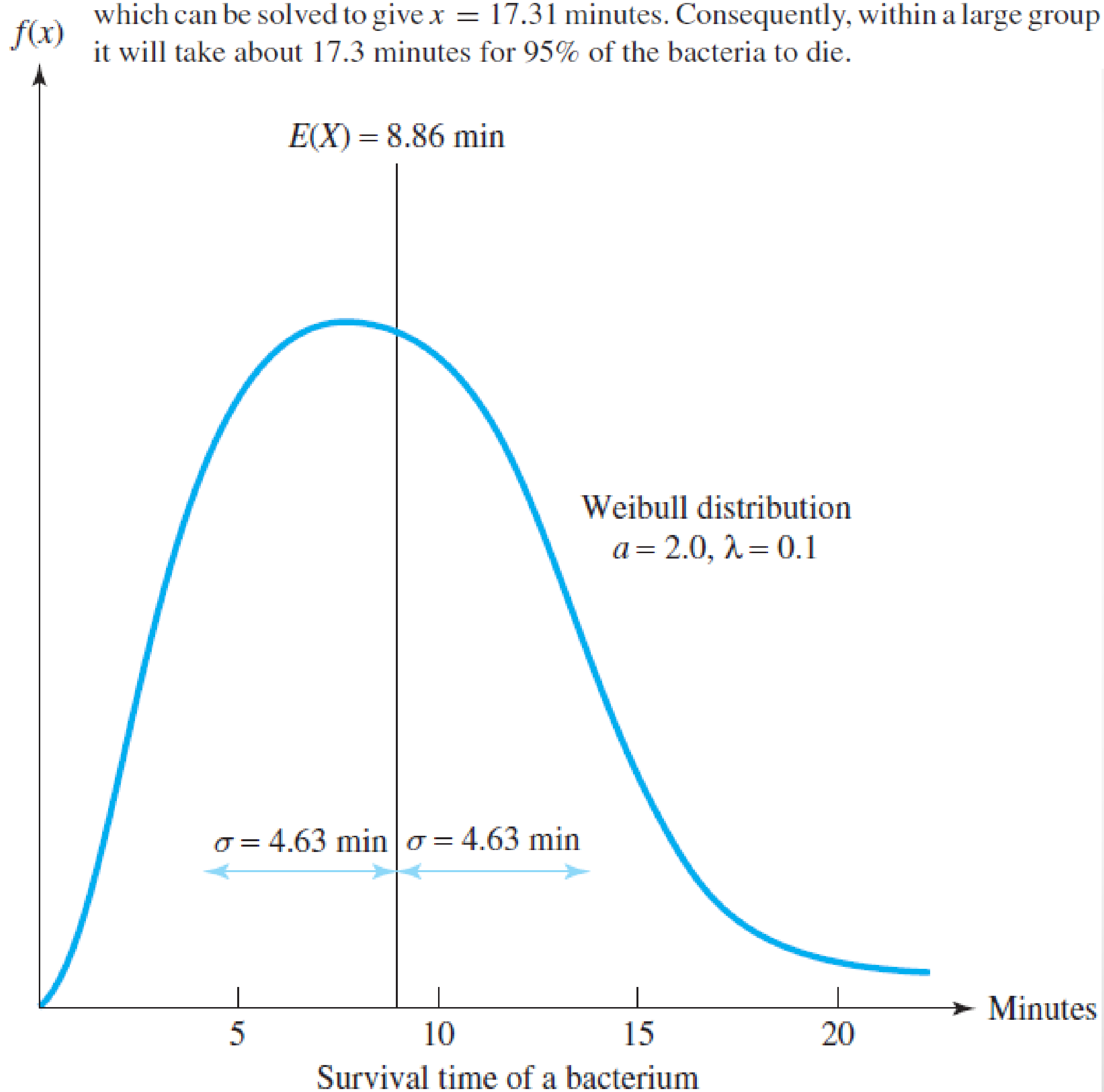
Notice that if $F(x) = 0.95$, then

$$0.95 = 1 - e^{-(0.1 \times x)^2}$$

which can be solved to give $x = 17.31$ minutes. Consequently, within a large group of bacteria, it will take about 17.3 minutes for 95% of the bacteria to die.

FIGURE 4.19

Distribution of bacteria survival times



The Weibull Distribution: Example

- Car break pad wear

A brake pad made from a new compound is tested in cars that are driven in city traffic. The random variable X, which measures the mileage in 1000-mile units that the cars can be driven before the brake pads wear out, has a Weibull distribution with parameters $a = 3.5$ and $\lambda = 0.12$. This distribution is shown in Figure 4.20.

The median car mileage is the value x satisfying

$$0.5 = F(x) = 1 - e^{-(0.12 \times x)^{3.5}}$$

which can be solved to give $x = 7.50$. Consequently, it should be expected that about half of the brake pads will last longer than 7500 miles. The probability that a set of brake pads last longer than 10,000 miles is

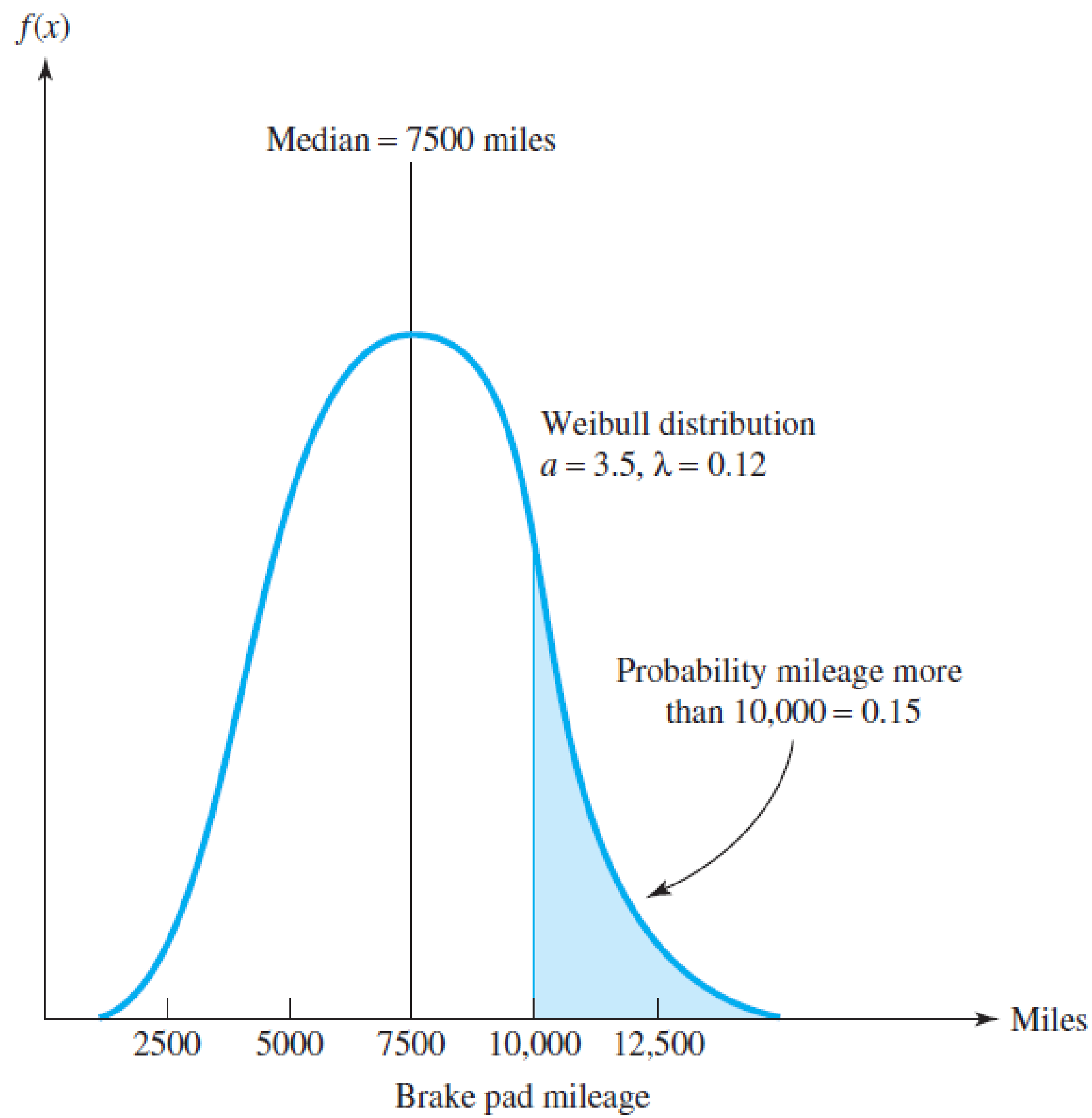
$$P(X \geq 10) = 1 - F(10) = e^{-(0.12 \times 10)^{3.5}} = 0.15.$$

The Weibull Distribution: Example

- Car break pad wear

FIGURE 4.20

Distribution of brake pad mileage



The Beta Distribution: Definition

The **beta distribution** has a state space $0 \leq x \leq 1$ and is often used to model proportions.

The Beta Distribution

A **beta distribution** with parameters $a > 0$ and $b > 0$ has a probability density function

$$f(x) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1} (1-x)^{b-1}$$

for $0 \leq x \leq 1$ and $f(x) = 0$ elsewhere. It is useful for modeling proportions. Its expectation and variance are

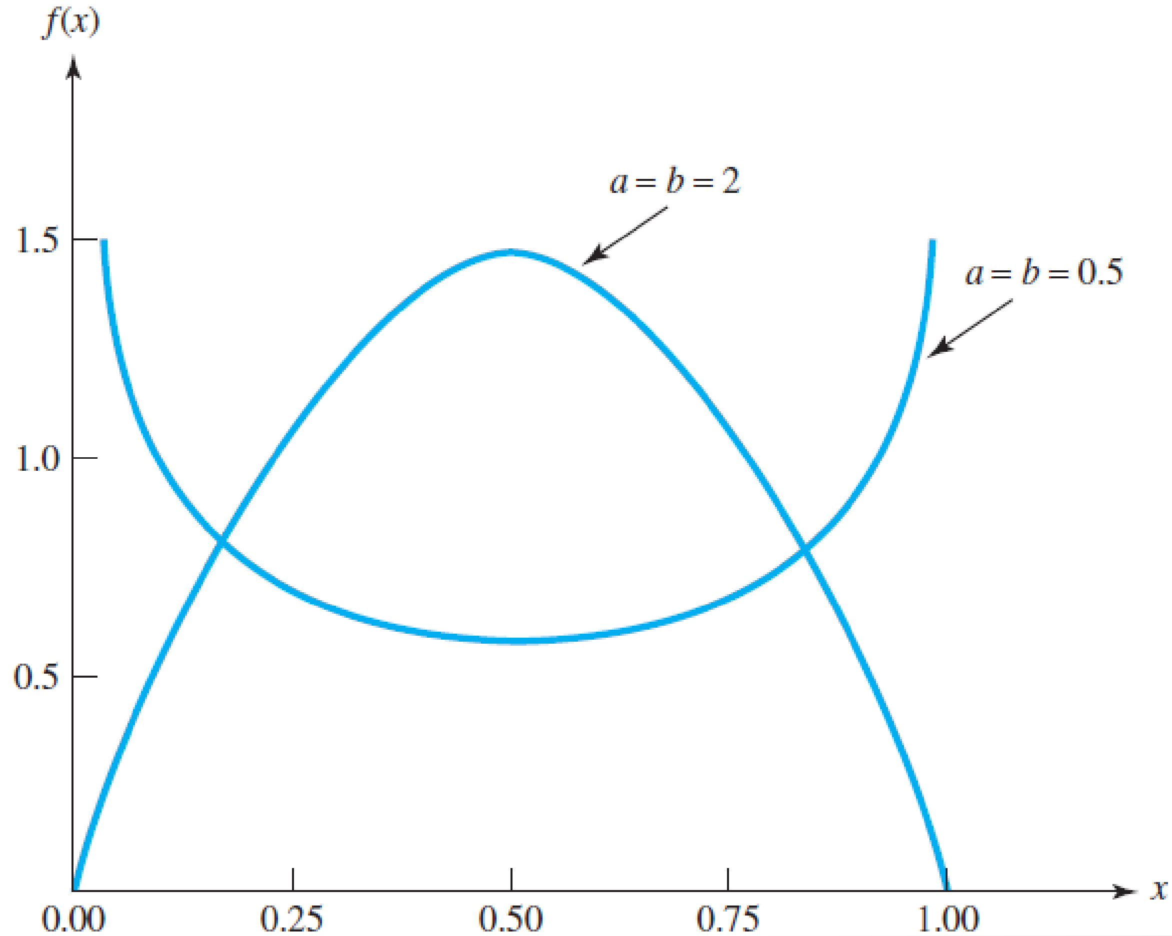
$$E(X) = \frac{a}{a+b} \qquad \text{and} \qquad \mathrm{Var}(X) = \frac{ab}{(a+b)^2(a+b+1)}$$

Figure 4.21 illustrates the probability density functions of beta distributions with $a = b = 0.5$ and $a = b = 2$. While their shapes are quite different, they are both symmetric about $x = 0.5$. In fact, all beta distributions with $a = b$ are symmetric. Figure 4.22 illustrates the probability density functions of beta distributions with $a = 0.5$, $b = 2$ and with $a = 4$, $b = 2$.

The Beta Distribution: Definition

FIGURE 4.21

Probability density functions of the beta distribution

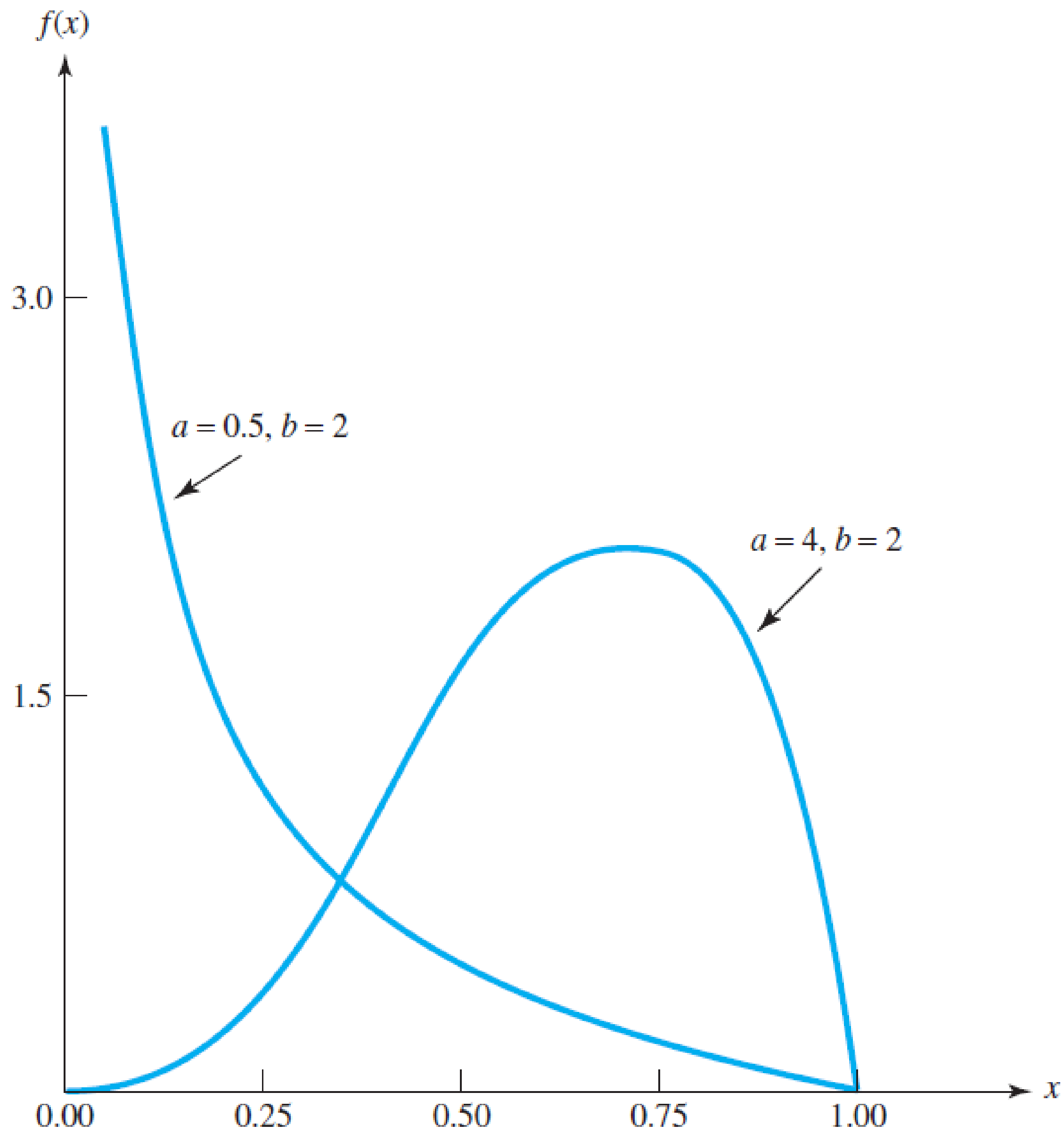


VIRTUAL REALITY LABORATORY

The Beta Distribution: Definition

FIGURE 4.22

Probability density functions of the beta distribution



The Beta Distribution: Example

- Stock prices

A Wall Street analyst has built a model for the performance of the stock market. In this model the *proportion* of listed stocks showing an increase in value on a particular day has a beta distribution with parameter values a and b, which depend upon various economic and political factors. On each day the analyst predicts suitable values of the parameters for modeling the subsequent day's stock prices. Suppose that on Monday the analyst predicts that parameter values $a = 5.5$ and $b = 4.2$ are suitable for the next day. What does this indicate about stock prices on Tuesday?

The distribution of the proportion of stocks increasing in value on Tuesday is shown in Figure 4.23. The expected proportion of stocks increasing in value on Tuesday is

$$E(X) = \frac{5.5}{5.5 + 4.2} = 0.57$$

The variance in the proportion is

$$\mathrm{Var}(X) = \frac{5.5 \times 4.2}{(5.5 + 4.2)^2 \times (5.5 + 4.2 + 1)} = 0.0229$$

so that the standard deviation is $\sigma = \sqrt{0.0229} = 0.15$. A software package can be used to calculate the probability that more than 75% of the stocks increase in value as

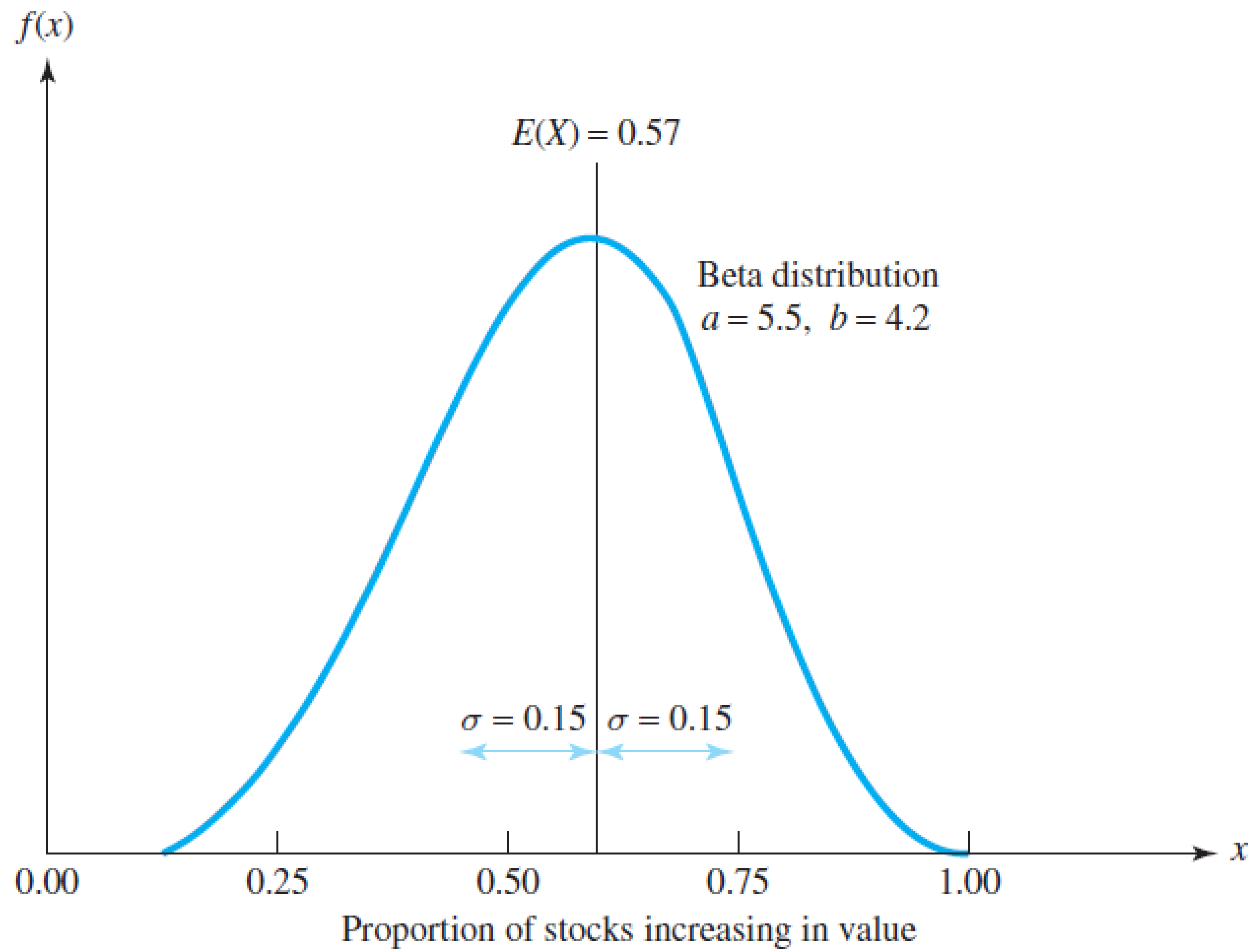
$$P(X \geq 0.75) = 1 - F(0.75) = 1 - 0.881 = 0.119$$

The Beta Distribution: Example

- Stock prices

FIGURE 4.23

Distribution of the proportion of stocks that increase in value



The Beta Distribution: Example

- Bee colonies

When a queen bee leaves a bee colony to start a new hive, a certain proportion of the worker bees take flight and follow her. An entomologist models the proportion X of the worker bees that leave with the queen using a beta distribution with parameters $a = 2.0$ and $b = 4.8$. This distribution is illustrated in Figure 4.24.

The expected proportion of bees leaving is

$$E(X) = \frac{2.0}{2.0 + 4.8} = 0.29$$

The variance in the proportion is

$$\text{Var}(X) = \frac{2.0 \times 4.8}{(2.0 + 4.8)^2 \times (2.0 + 4.8 + 1)} = 0.0266$$

so that the standard deviation is $\sigma = \sqrt{0.0266} = 0.16$. The probability that more than half of the bee colony leaves with the queen can be calculated from a software package to be

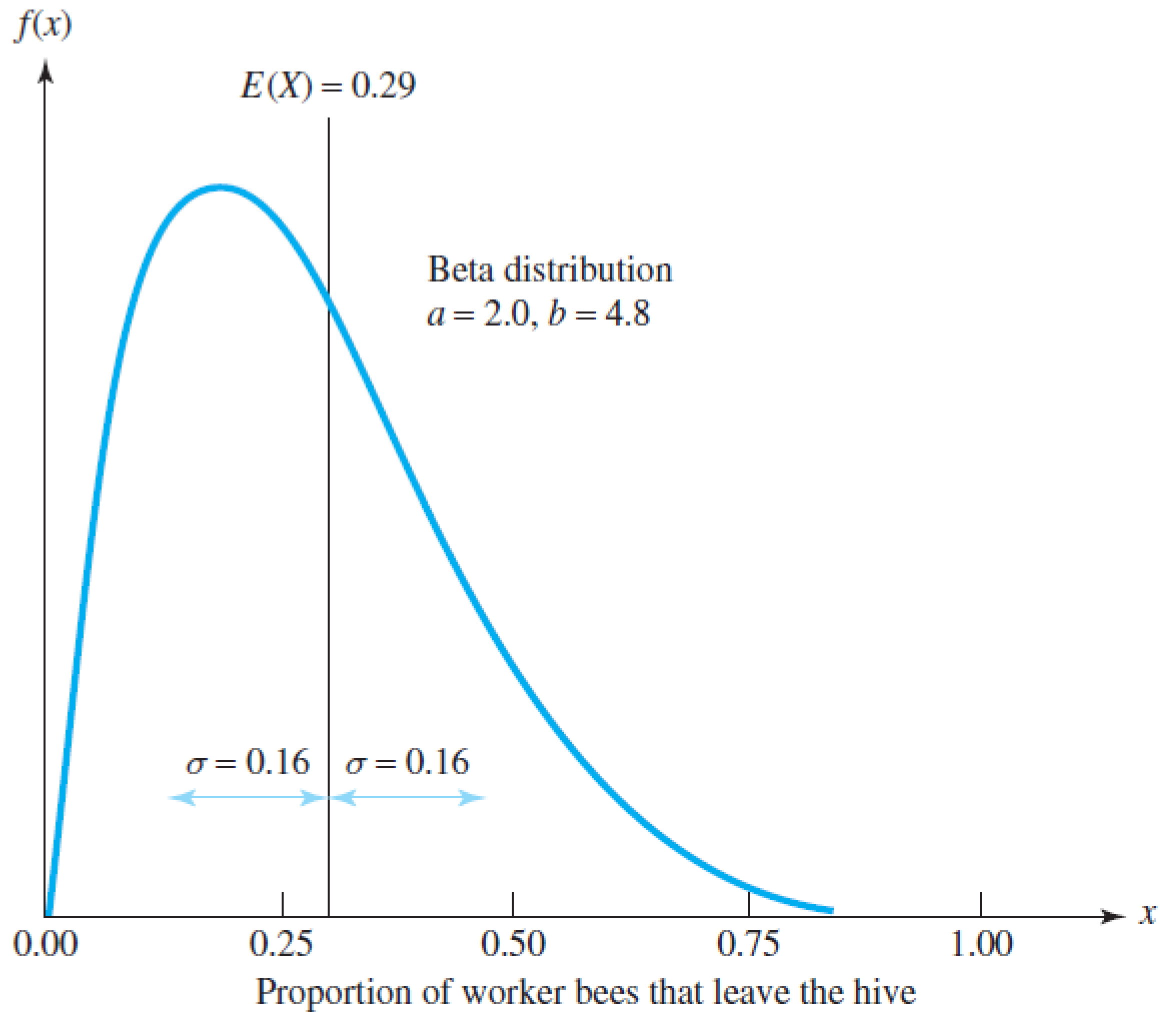
$$P(X \geq 0.5) = 1 - F(0.5) = 1 - 0.878 = 0.122$$

The Beta Distribution: Example

- Bee colonies

FIGURE 4.24

Distribution of the proportion of worker bees leaving with queen



Continuous Probability Distributions: Case Study

- Microelectronic solder joints

A Weibull distribution can be used to model the number of temperature cycles that an assembly can be subjected to before it fails. In this case, experience dictates that it is best to define the cumulative distribution function of the failure time distribution in terms of the logarithm of the number of cycles, so that

$$P(\text{assembly fails within } t \text{ cycles}) = P(X \le t) = 1 - e^{-(\lambda \ln(t))^a}$$

The values of the parameters a and λ will depend upon the specific design of the assembly. Suppose that if an epoxy of type I is used for the underfill, then $a = 25.31$ and $\lambda = 0.120$, whereas if an epoxy of type II is used, then $a = 27.42$ and $\lambda = 0.116$. The solution of

$$P(X \le t) = 1 - e^{-(0.120 \ln(t))^{25.31}} = 0.01$$

is $t = 1041$, and

$$P(X \le t) = 1 - e^{-(0.120 \ln(t))^{25.31}} = 0.5$$

is solved with $t = 3691$. Consequently, if epoxy of type I is used for the underfill, then 99% of the assemblies can survive 1041 temperature cycles, whereas half of them can survive 3691 temperature cycles. In addition, the solution of

$$P(X \le t) = 1 - e^{-(0.116 \ln(t))^{27.42}} = 0.01$$

is $t = 1464$ and

$$P(X \le t) = 1 - e^{-(0.116 \ln(t))^{27.42}} = 0.5$$

is solved with $t = 4945$, so that if epoxy of type II is used for the underfill, then 99% of the assemblies can survive 1464 temperature cycles, whereas half of them can survive 4945 temperature cycles. These calculations reveal that an underfill with epoxy of type II produces an assembly with better reliability.

Continuous Probability Distributions: Case Study

- Internet marketing

When a individual has logged on to the organisation's website, the length of the idle periods in minutes is distributed as a gamma distribution with $k = 1.1$ and $\lambda = 0.9$. Consequently, the idle periods have an expectation of $k/\lambda = 1.1/0.9 = 1.22$ minutes, and the standard deviation is $\sqrt{k}/\lambda = \sqrt{1.1}/0.9 = 1.17$ minutes.

Suppose that the individual is automatically logged out when the idle period reaches 5 minutes. What proportion of the idle periods result in the individual being automatically logged out?

This can be calculated as

$$P(\text{Gamma}(k = 1.1, \lambda = 0.9) \geq 5) = 1 - P(\text{Gamma}(k = 1.1, \lambda = 0.9) \leq 5)$$
$$= 1 - 0.986 = 0.014$$

so that the proportion is 1.4%.