Selecting the Family of Distributions

Example 2–2 Record High Temperatures

These data represent the record high temperatures in degrees Fahrenheit (°F) for each of the 50 states. Construct a grouped frequency distribution for the data using 7 classes.

112	100	127	120	134	118	105	110	109	112
110	118	117	116	118	122	114	114	105	109
107	112	114	115	118	117	118	122	106	110
116	108	110	121	113	120	119	111	104	111
120	113	120	117	105	110	118	112	114	114

Source: *The World Almanac and Book of Facts.*

Step 1 Determine the classes.

Find the highest value and lowest value: $H = 134$ and $L = 100$.

Find the range: $R = \text{highest value} - \text{lowest value} = H - L$, so

$R = 134 - 100 = 34$

Select the number of classes desired (usually between 5 and 20). In this case, 7 is arbitrarily chosen.

Find the class width by dividing the range by the number of classes.

$$\text{Width} = \frac{R}{\text{number of classes}} = \frac{34}{7} = 4.9$$

Round the answer up to the nearest whole number if there is a remainder: $4.9 \approx 5$. (Rounding *up* is different from rounding *off*. A number is rounded up if there is any decimal remainder when dividing. For example, $85 \div 6 = 14.167$ and is rounded up to 15. Also, $53 \div 4 = 13.25$ and is rounded up to 14. Also, after dividing, if there is no remainder, you will need to add an extra class to accommodate all the data.)

Select a starting point for the lowest class limit. This can be the smallest data value or any convenient number less than the smallest data value. In this case, 100 is used. Add the width to the lowest score taken as the starting point to get the lower limit of the next class. Keep adding until there are 7 classes, as shown, 100, 105, 110, etc.

Subtract one unit from the lower limit of the second class to get the upper limit of the first class. Then add the width to each upper limit to get all the upper limits.

$$105 - 1 = 104$$

The first class is 100–104, the second class is 105–109, etc.

Find the class boundaries by subtracting 0.5 from each lower class limit and adding 0.5 to each upper class limit:

99.5–104.5, 104.5–109.5, etc.

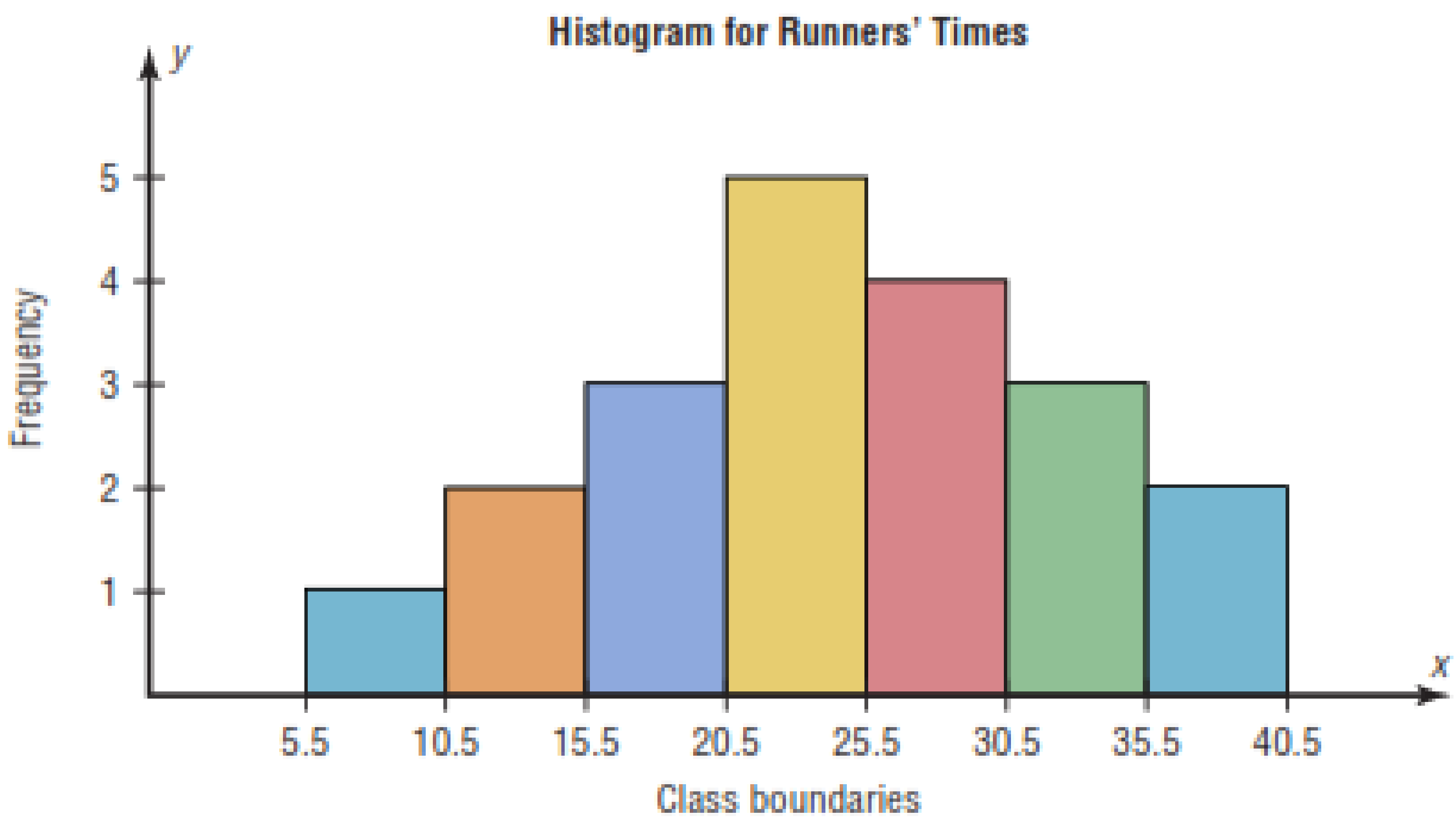
Step 2 Tally the data.

Step 3 Find the numerical frequencies from the tallies.

The completed frequency distribution is

Class limits	Class boundaries	Tally	Frequency
100–104	99.5–104.5	//	2
105–109	104.5–109.5	~~////~~ ///	8
110–114	109.5–114.5	~~////~~ ~~////~~ ~~////~~ ///	18
115–119	114.5–119.5	~~////~~ ~~////~~ ///	13
120–124	119.5–124.5	~~////~~ //	7
125–129	124.5–129.5	/	1
130–134	129.5–134.5	/	1
			$n = \Sigma f = 50$

The frequency distribution shows that the class 109.5–114.5 contains the largest number of temperatures (18) followed by the class 114.5–119.5 with 13 temperatures. Hence, most of the temperatures (31) fall between 109.5 and 119.5°F.


Histogram for Runners' Times
y
5
4
3
2
1
Frequency
x
5.5
10.5
15.5
20.5
25.5
30.5
35.5
40.5
Class boundaries

(a) Histogram

Table 1 Number of Arrivals in a 5-Minute Period

Arrivals per Period	*Frequency*	*Arrivals per Period*	*Frequency*
0	12	6	7
1	10	7	5
2	19	8	5
3	17	9	3
4	10	10	3
5	8	11	1

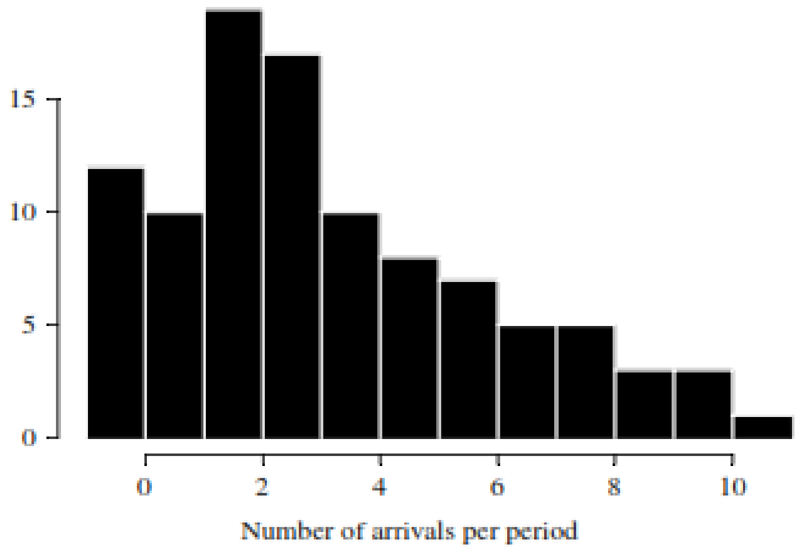
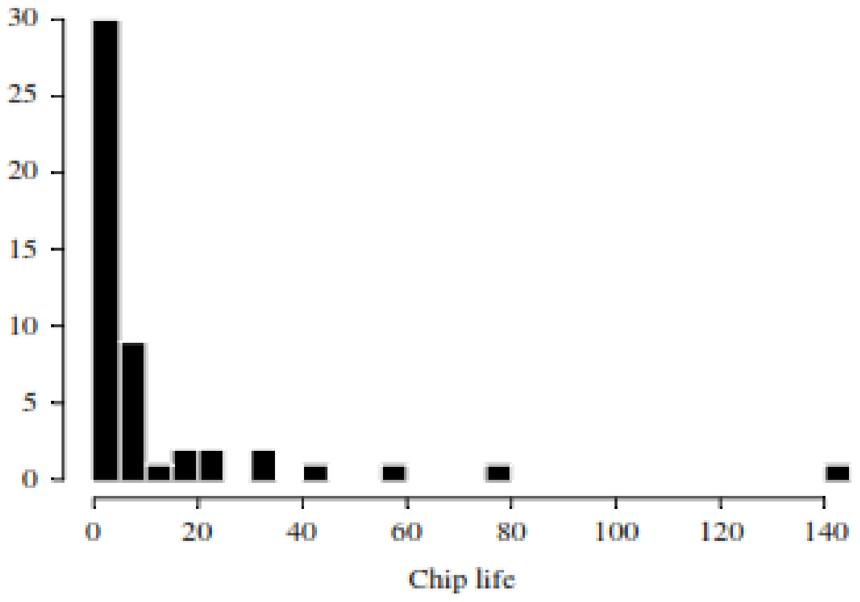


Figure 4 Histogram of number of arrivals per period.

Example 6: Continuous Data

Life tests were performed on a random sample of electronic components at 1.5 times the nominal voltage, and their lifetime (or time to failure), in days, was recorded:

79.919	3.081	0.062	1.961	5.845
3.027	6.505	0.021	0.013	0.123
6.769	59.899	1.192	34.760	5.009
18.387	0.141	43.565	24.420	0.433
144.695	2.663	17.967	0.091	9.003
0.941	0.878	3.371	2.157	7.579
0.624	5.380	3.148	7.078	23.960
0.590	1.928	0.300	0.002	0.543
7.004	31.764	1.005	1.147	0.219
3.217	14.382	1.008	2.336	4.562



6.2
USEFUL PROBABILITY DISTRIBUTIONS

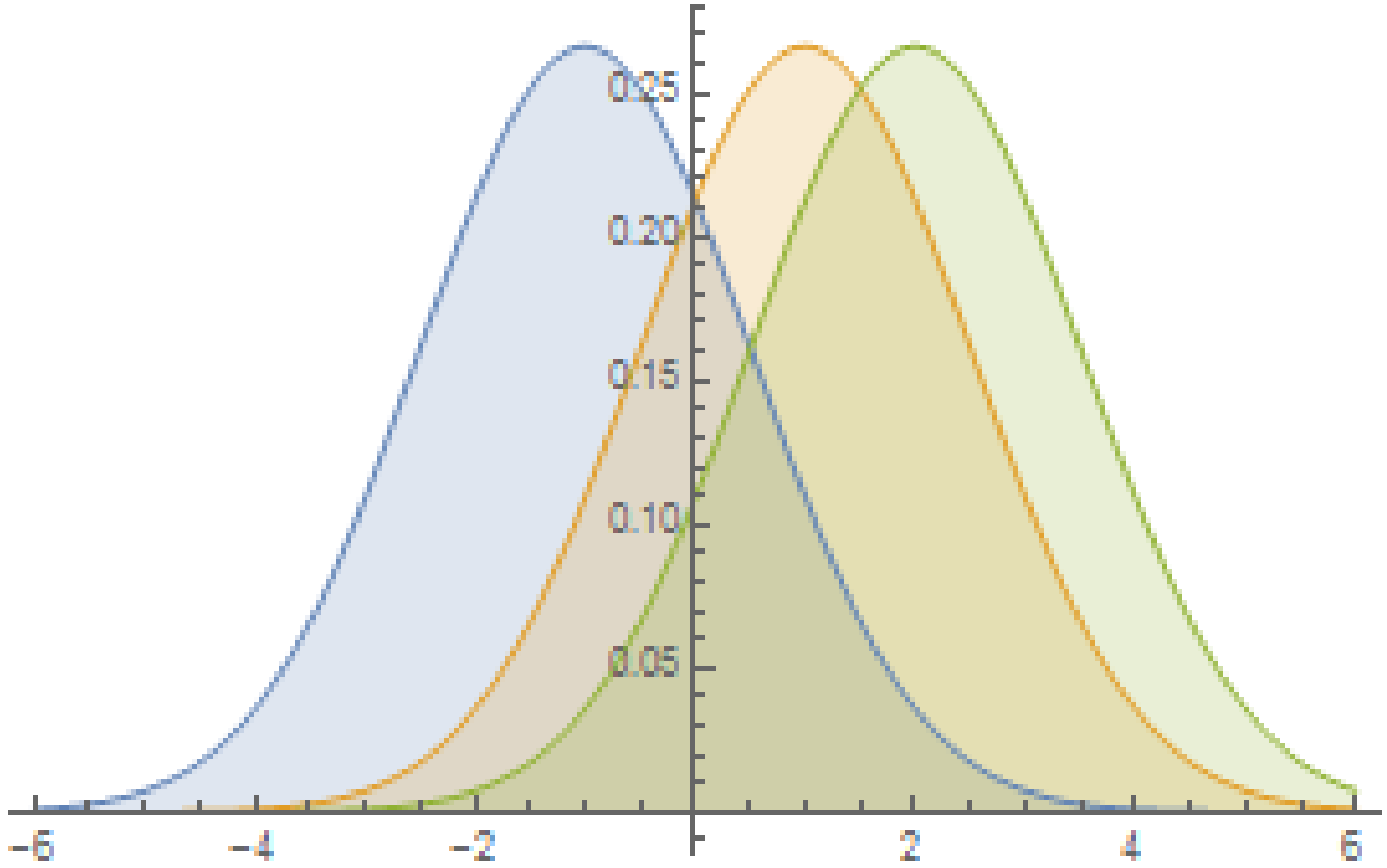
The purpose of this section is to discuss a variety of distributions that have been found to be useful in simulation modeling and to provide a unified listing of relevant properties of these distributions.

6.2.1 Parameterization of Continuous Distributions

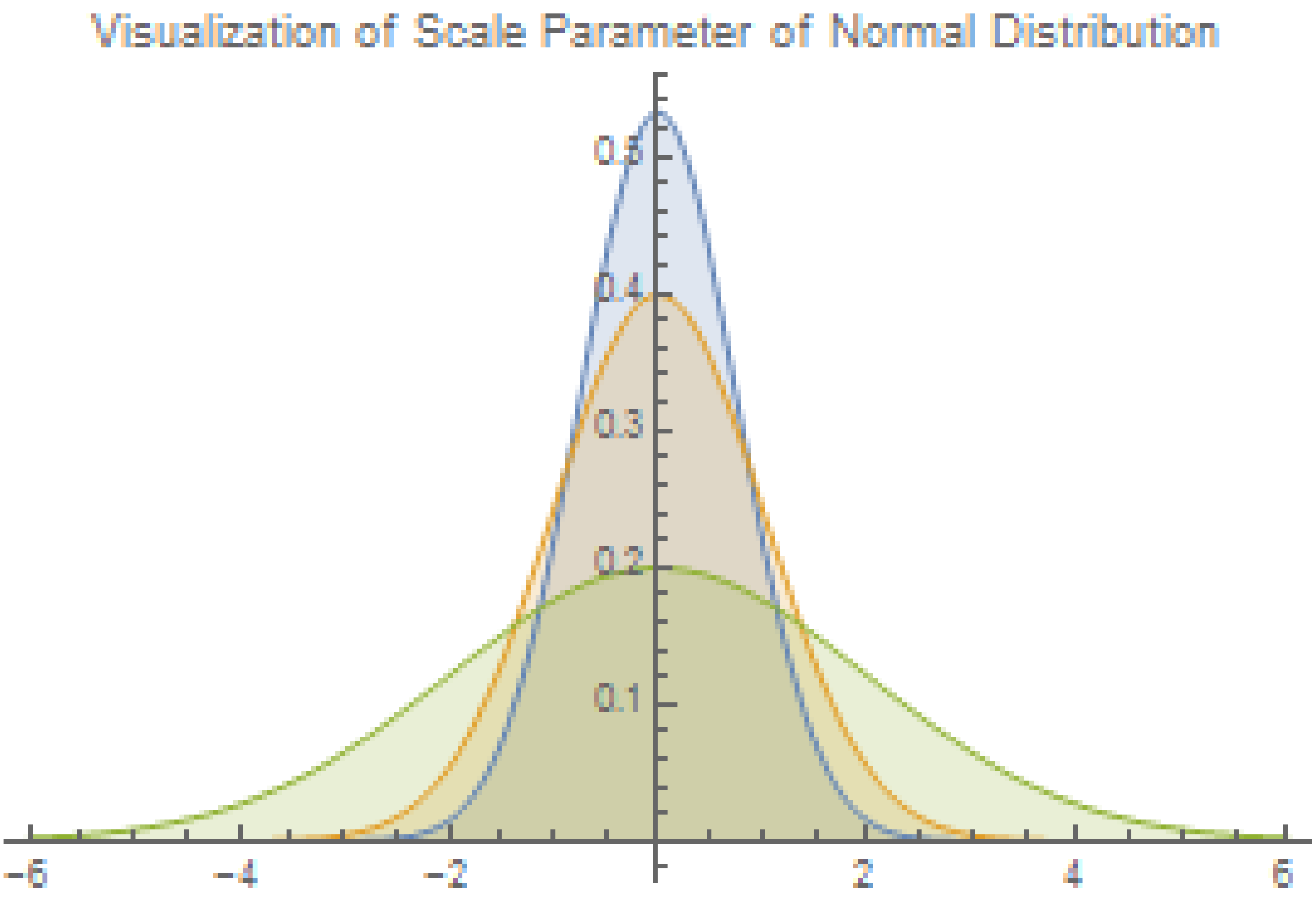
For a given family of continuous distributions, e.g., normal or gamma, there are usually several alternative ways to define, or *parameterize*, the probability density function. However, if the parameters are defined correctly, they can be classified, on the basis of their physical or geometric interpretation, as being one of three basic types: location, scale, or shape parameters.

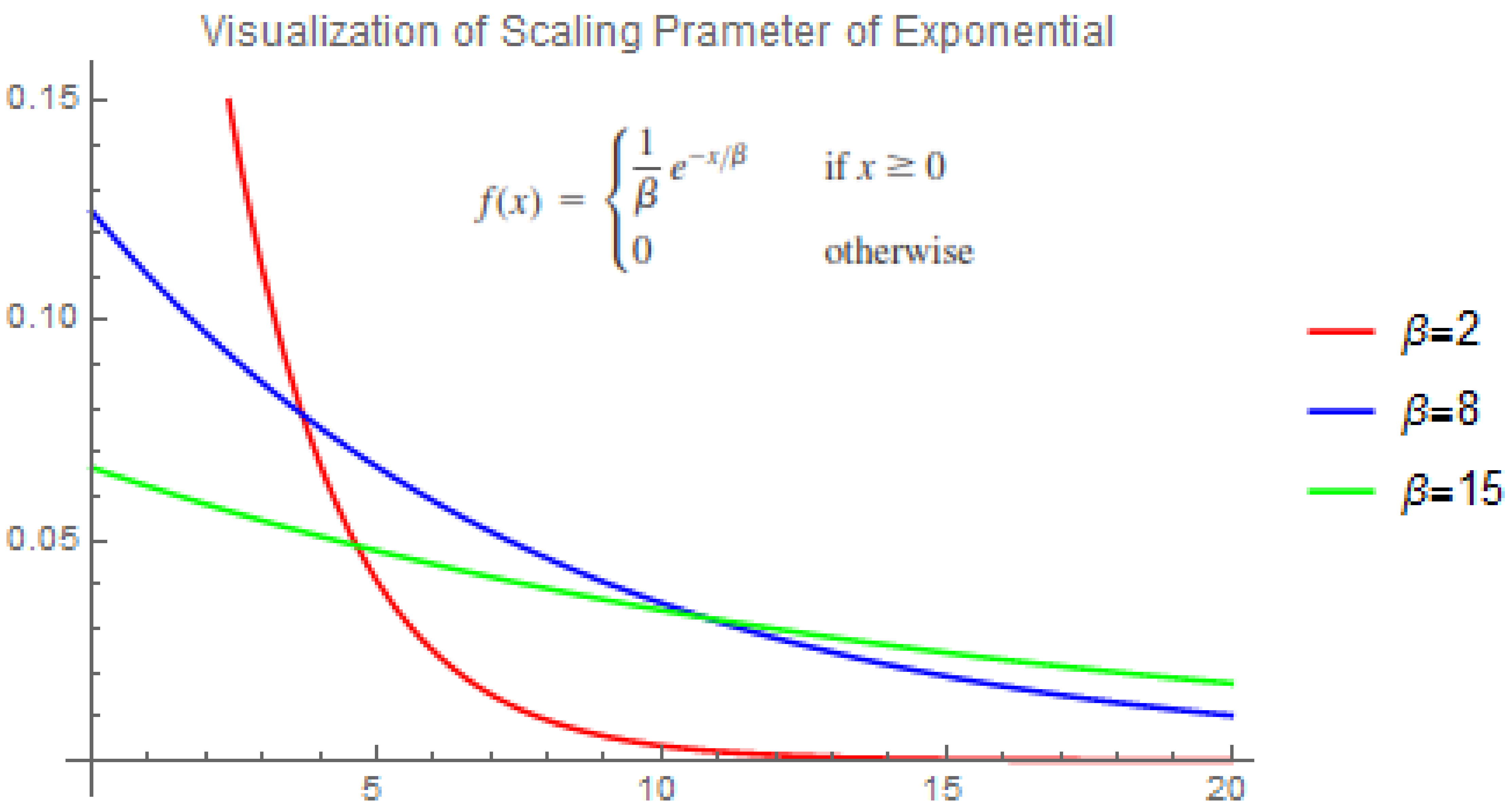
A location parameter specifies an abscissa (x axis) location point of a distribution's range of values; usually is the midpoint (e.g., the mean for a normal distribution)
or lower endpoint (see Sec. 6.8) of the distribution's range. (In the latter case, location parameters are sometimes called shift parameters.)
As it changes, the associated distribution merely shifts left or right without otherwise changing.

Visualization of Location Parameter of Normal Distribution



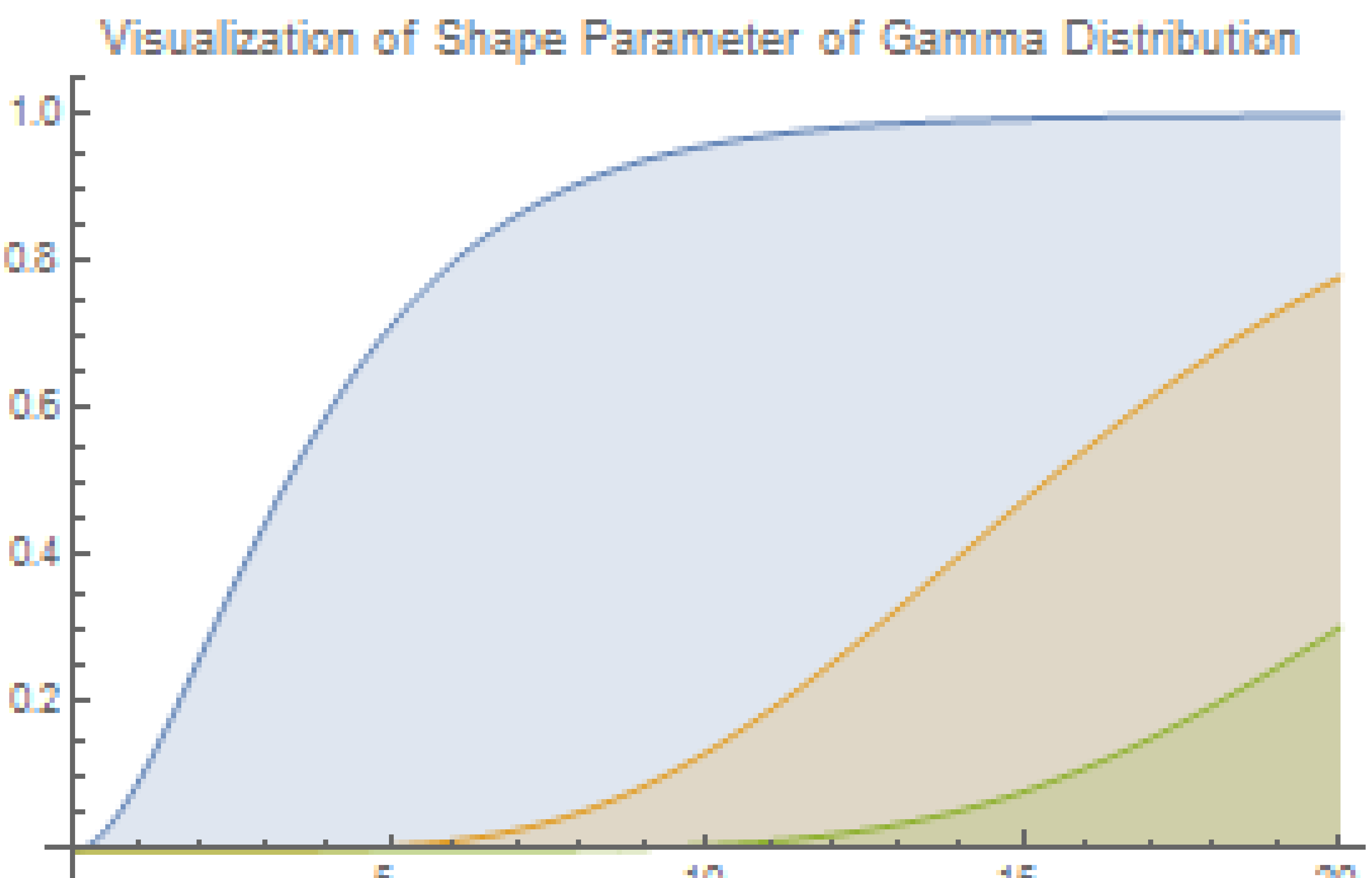
A scale parameter determines the scale (or unit) of measurement of the values in the range of the distribution. (The standard deviations is a scale parameter for the normal distribution.) A change in b compresses or expands the associated distribution without altering its basic form.


Visualization of Scale Parameter of Normal Distribution
0.5
0.4
0.3
0.2
0.1
−6
−4
−2
2
4
6



Visualization of Scaling Prameter of Exponential
0.15
0.10
0.05
5
10
15
20
$f(x) = \begin{cases} \frac{1}{\beta} e^{-x/\beta} & \text{if } x \geq 0 \\ 0 & \text{otherwise} \end{cases}$
β=2
β=8
β=15

A shape parameter a determines, distinct from location and scale, the basic form or shape of a distribution within the general family of distributions of interest. A change in a generally alters a distribution’s properties (e.g., skewness) more fundamentally than a change in location or scale. Some distributions (e.g., exponential and normal) do not have a shape parameter, others (e.g., beta) may have two.



Visualization of Shape Parameter of Gamma Distribution
1.0
0.8
0.6
0.4
0.2
5
10
15
20

6.2.2 Continuous Distributions

Uniform $\mathbf{U}(a, b)$

Continuous distributions

Uniform	**U(*a*, *b*)**
Possible applications	Used as a "first" model for a quantity that is felt to be randomly varying between a and b but about which little else is known. The U(0, 1) distribution is essential in generating random values from all other distributions (see Chaps. 7 and 8).
Density (See Fig. 6.5)	$f(x) = \begin{cases} \dfrac{1}{b-a} & \text{if } a \le x \le b \\ 0 & \text{otherwise} \end{cases}$
Distribution	$F(x) = \begin{cases} 0 & \text{if } x < a \\ \dfrac{x-a}{b-a} & \text{if } a \le x \le b \\ 1 & \text{if } b < x \end{cases}$
Parameters	a and b real numbers with $a < b$; a is a location parameter, $b - a$ is a scale parameter

Uniform	U(a, b)
Range	$[a, b]$
Mean	$\frac{a+b}{2}$
Variance	$\frac{(b-a)^2}{12}$
Mode	Does not uniquely exist
MLE	$\hat{a} = \min_{1 \le i \le n} X_i$, $\hat{b} = \max_{1 \le i \le n} X_i$
Comments	1. The U(0, 1) distribution is a special case of the beta distribution (when $\alpha_1 = \alpha_2 = 1$). 2. If $X \sim U(0, 1)$ and $[x, x + \Delta x]$ is a subinterval of $[0, 1]$ with $\Delta x \geq 0$, $$P(X \in [x, x+\Delta x]) = \int_x^{x+\Delta x} 1\,dy = (x + \Delta x) - x = \Delta x$$ which justifies the name "uniform."

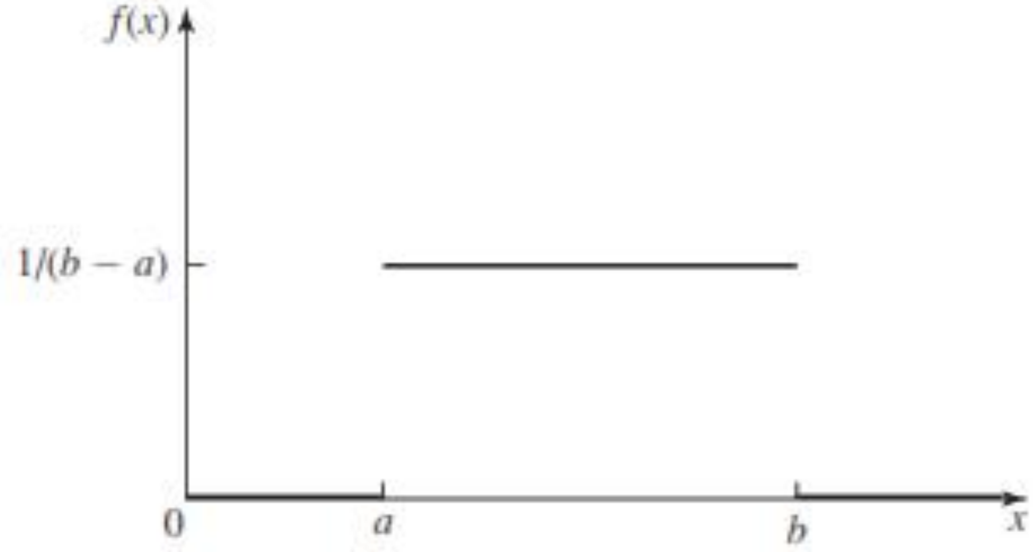


FIGURE 6.5
U(a, b) density function.

Gamma	**gamma(α, β)**

Gamma	gamma(α, β)
Possible applications	Time to complete some task, e.g., customer service or machine repair
Density (see Fig. 6.7)	$f(x) = \begin{cases} \dfrac{\beta^{-\alpha} x^{\alpha-1} e^{-x/\beta}}{\Gamma(\alpha)} & \text{if } x > 0 \\ 0 & \text{otherwise} \end{cases}$ where $\Gamma(\alpha)$ is the *gamma function*, defined by $\Gamma(z) = \int_0^\infty t^{z-1} e^{-t} dt$ for any real number $z > 0$. Some properties of the gamma function: $\Gamma(z + 1) = z\Gamma(z)$ for any $z > 0$, $\Gamma(k + 1) = k!$ for any nonnegative integer k, $\Gamma(k + \frac{1}{2}) = \sqrt{\pi} \cdot 1 \cdot 3 \cdot 5 \cdots (2k - 1)/2^k$ for any positive integer k, $\Gamma(1/2) = \sqrt{\pi}$

Gamma	**gamma(α, β)**
Distribution	If α is not an integer, there is no closed form. If α is a positive integer, then $$F(x) = \begin{cases} 1 - e^{-x/\beta} \sum_{j=0}^{\alpha-1} \frac{(x/\beta)^j}{j!} & \text{if } x > 0 \\ 0 & \text{otherwise} \end{cases}$$
Parameters	Shape parameter $\alpha > 0$, scale parameter $\beta > 0$
Range	$[0, \infty)$
Mean	$\alpha\beta$
Variance	$\alpha\beta^2$
Mode	$\beta(\alpha - 1)$ if $\alpha \geq 1$, 0 if $\alpha < 1$
MLE	The following two equations must be satisfied: $$\ln \hat{\beta} + \Psi(\hat{\alpha}) = \frac{\sum_{i=1}^{n} \ln X_i}{n}, \qquad \hat{\alpha}\hat{\beta} = \overline{X}(n)$$ which could be solved numerically. [$\Psi(\hat{\alpha}) = \Gamma'(\hat{\alpha})/\Gamma(\hat{\alpha})$ and is called the *digamma function*; Γ' denotes the derivative of Γ.] Alternatively, approximations to $\hat{\alpha}$ and $\hat{\beta}$ can be obtained by letting $T = [\ln \overline{X}(n) - \sum_{i=1}^{n} \ln X_i/n]^{-1}$, using Table 6.21 (see App. 6A) to obtain $\hat{\alpha}$ as a function of T, and letting $\hat{\beta} = \overline{X}(n)/\hat{\alpha}$. [See Choi and Wette (1969) for the derivation of this procedure and of Table 6.21.]

Gamma	gamma(α, β)
Comments	1. The expo(β) and gamma(1, β) distributions are the same. 2. For a positive integer m, the gamma(m, β) distribution is called the m-Erlang(β) distribution. 3. The chi-square distribution with k df is the same as the gamma($k/2$, 2) distribution. 4. If $X_1, X_2, \ldots, X_m$ are independent random variables with $X_i \sim$ gamma(α_i, β), then $X_1 + X_2 + \cdots + X_m \sim$ gamma($\alpha_1 + \alpha_2 + \cdots + \alpha_m$, β). 5. If X_1 and X_2 are independent random variables with $X_i \sim$ gamma(α_i, β), then $X_1/(X_1 + X_2) \sim$ beta(α_1, α_2). 6. $X \sim$ gamma(α, β) if and only if $Y = 1/X$ has a Pearson type V distribution with shape and scale parameters α and $1/\beta$, denoted PT5(α, $1/\beta$). 7. $\lim_{x \to 0} f(x) = \begin{cases} \infty & \text{if } \alpha < 1 \\ \frac{1}{\beta} & \text{if } \alpha = 1 \\ 0 & \text{if } \alpha > 1 \end{cases}$

and of Table 6.21.]

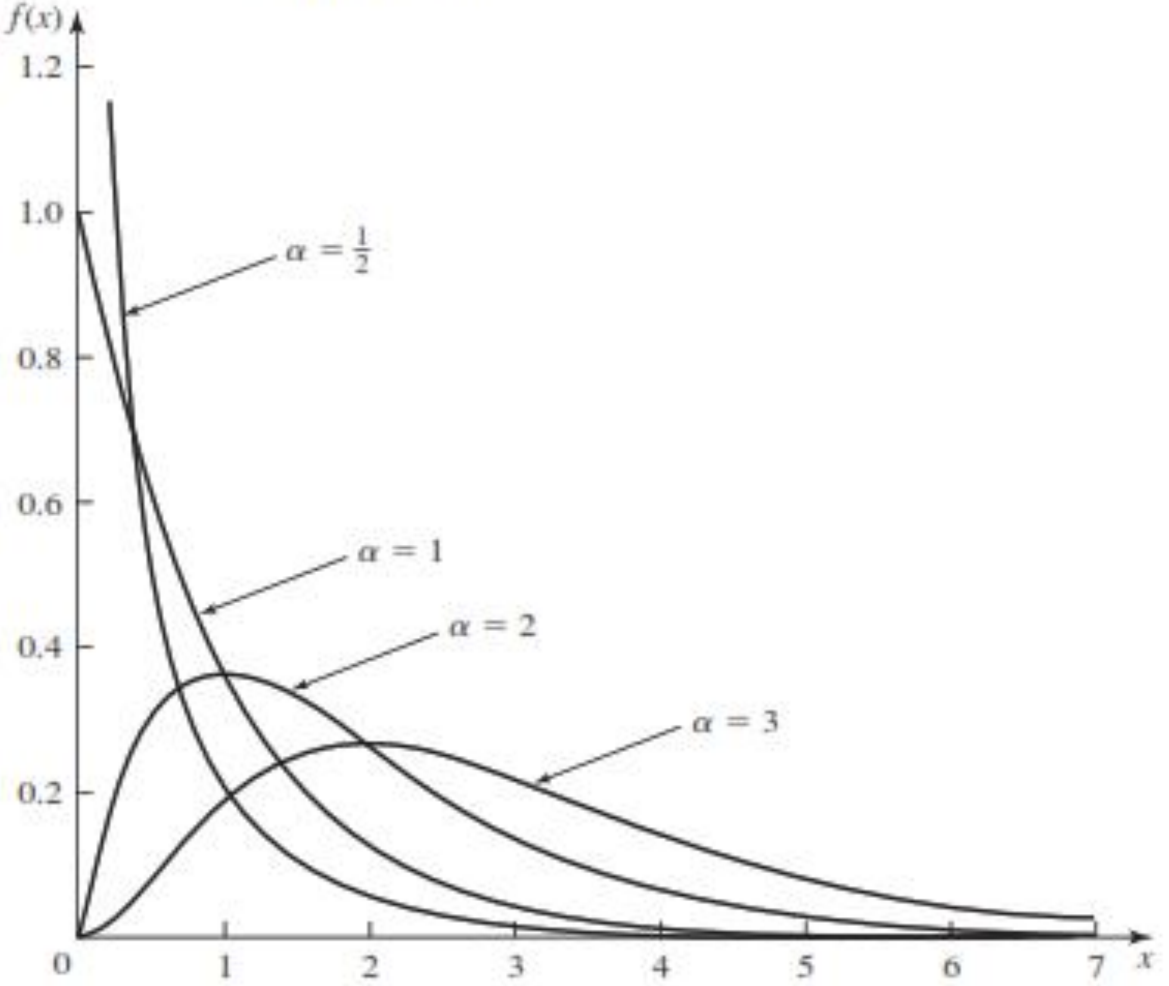


FIGURE 6.7
gamma(α, 1) density functions.

(continued)

Exponential	expo(β)
Possible applications	Interarrival times of "customers" to a system that occur at a constant rate, time to failure of a piece of equipment.
Density (see Fig. 6.6)	$f(x) = \begin{cases} \frac{1}{\beta} e^{-x/\beta} & \text{if } x \geq 0 \\ 0 & \text{otherwise} \end{cases}$
Distribution	$F(x) = \begin{cases} 1 - e^{-x/\beta} & \text{if } x \geq 0 \\ 0 & \text{otherwise} \end{cases}$
Parameter	Scale parameter $\beta > 0$
Range	$[0, \infty)$
Mean	β
Variance	β^2
Mode	0

Exponential	expo(β)
MLE	$\hat{\beta} = \bar{X}(n)$
Comments	1. The expo(β) distribution is a special case of both the gamma and Weibull distributions (for shape parameter $\alpha = 1$ and scale parameter β in both cases). 2. If $X_1, X_2, \ldots, X_m$ are independent expo(β) random variables, then $X_1 + X_2 + \cdots + X_m \sim$ gamma(m, β), also called the *m-Erlang(β) distribution*. 3. The exponential distribution is the only continuous distribution with the memoryless property (see Prob. 4.30).

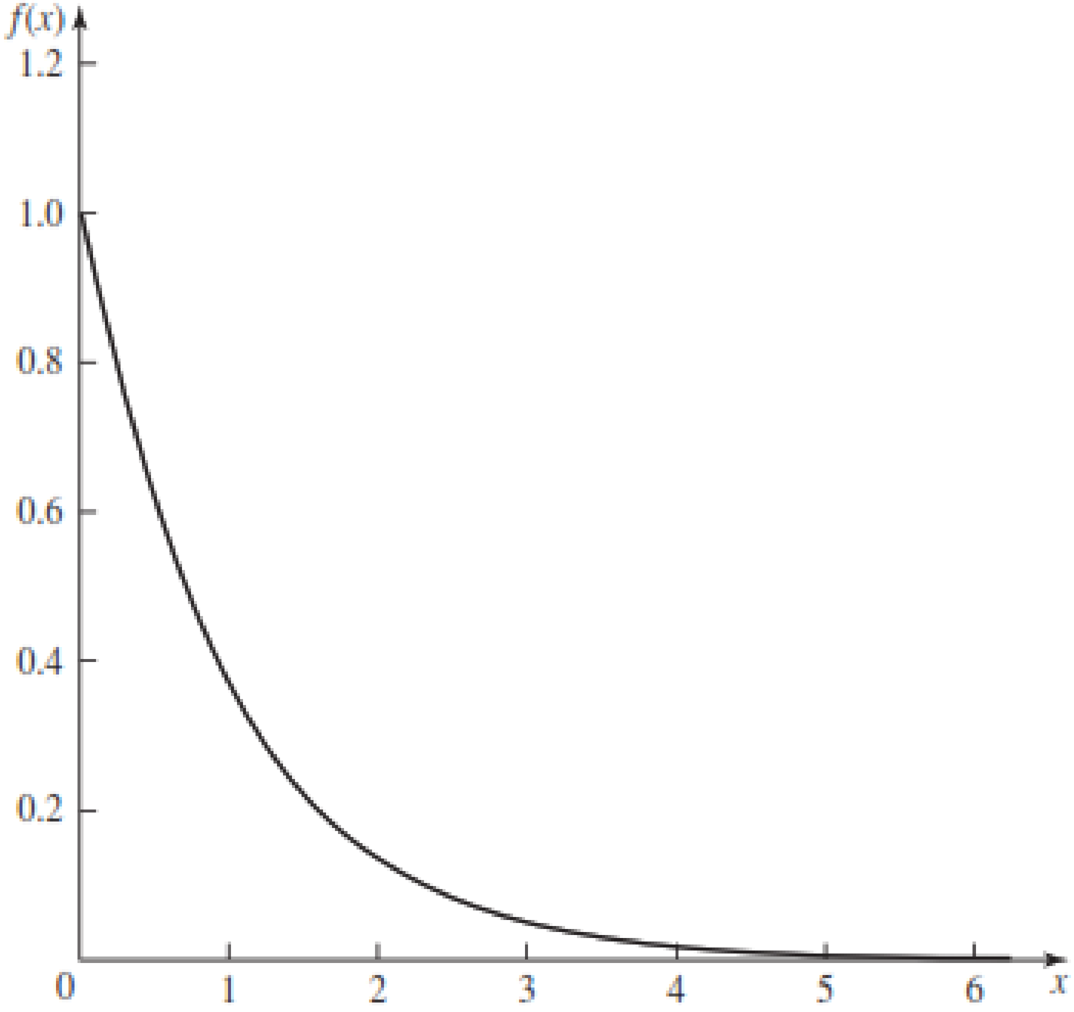


FIGURE 6.6
expo(1) density function.

Normal	$\mathbf{N}(\boldsymbol{\mu}, \boldsymbol{\sigma}^2)$
Possible applications	Errors of various types, e.g., in the impact point of a bomb; quantities that are the sum of a large number of other quantities (by virtue of central limit theorems)
Density (see Fig. 6.9)	$f(x) = \dfrac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/(2\sigma^2)}$ for all real numbers x
Distribution	No closed form
Parameters	Location parameter $\mu \in (-\infty, \infty)$, scale parameter $\sigma > 0$
Range	$(-\infty, \infty)$

Normal	$\mathbf{N}(\boldsymbol{\mu}, \boldsymbol{\sigma}^2)$
Mean	μ
Variance	σ^2
Mode	μ
MLE	$\hat{\mu} = \overline{X}(n), \qquad \hat{\sigma} = \left[\frac{n-1}{n} S^2(n)\right]^{1/2}$

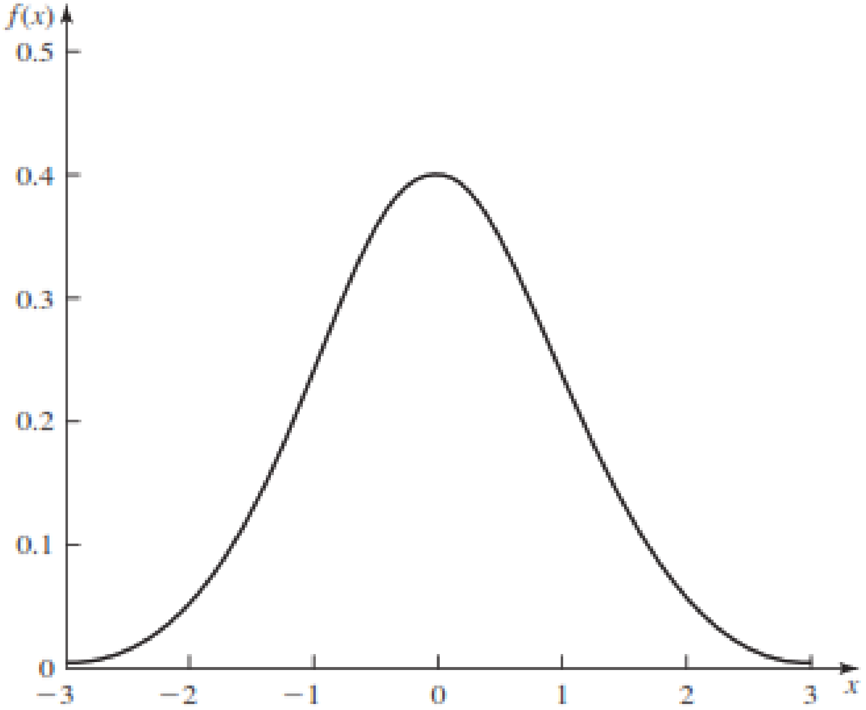


FIGURE 6.9
N(0, 1) density function.

(continued)

Poisson	**Poisson(λ)**
Possible applications	Number of events that occur in an interval of time when the events are occurring at a constant rate (see Sec. 6.12); number of items in a batch of random size; number of items demanded from an inventory
Mass (see Fig. 6.23)	$p(x) = \begin{cases} \dfrac{e^{-\lambda}\lambda^x}{x!} & \text{if } x \in \{0, 1, \ldots\} \\ 0 & \text{otherwise} \end{cases}$
Distribution	$F(x) = \begin{cases} 0 & \text{if } x < 0 \\ e^{-\lambda} \displaystyle\sum_{i=0}^{\lfloor x \rfloor} \dfrac{\lambda^i}{i!} & \text{if } x \geq 0 \end{cases}$
Parameter	$\lambda > 0$
Range	$\{0, 1, \ldots\}$
Mean	λ
Variance	λ

Poisson	Poisson(λ)
Mode	$\begin{cases} \lambda - 1 \text{ and } \lambda & \text{if } \lambda \text{ is an integer} \\ \lfloor \lambda \rfloor & \text{otherwise} \end{cases}$
MLE	$\hat{\lambda} = \overline{X}(n)$.
Comments	1. Let $Y_1, Y_2, \ldots$ be a sequence of nonnegative IID random variables, and let $X = \max\{i: \sum_{j=1}^{i} Y_j \le 1\}$. Then the distribution of the Y_i's is expo($1/\lambda$) if and only if $X \sim$ Poisson(λ). Also, if $X' = \max\{i: \sum_{j=1}^{i} Y_j \le \lambda\}$, then the Y_i's are expo(1) if and only if $X' \sim$ Poisson(λ) (see also Sec. 6.12). 2. If $X_1, X_2, \ldots, X_m$ are independent random variables and $X_i \sim$ Poisson(λ_i), then $X_1 + X_2 + \cdots + X_m \sim$ Poisson($\lambda_1 + \lambda_2 + \cdots + \lambda_m$).

order. Let $X_{(i)}$ denote the ith smallest of the X_j's, so that $X_{(1)} \le X_{(2)} \le \cdots \le X_{(n)}$. Then F is given by

$$F(x) = \begin{cases} 0 & \text{if } x < X_{(1)} \\ \dfrac{i-1}{n-1} + \dfrac{x - X_{(i)}}{(n-1)(X_{(i+1)} - X_{(i)})} & \text{if } X_{(i)} \le x < X_{(i+1)} \\ & \quad \text{for } i = 1, 2, \ldots, n-1 \\ 1 & \text{if } X_{(n)} \le x \end{cases}$$

Figure 6.24 gives an illustration for $n = 6$. Note that $F(x)$ rises most rapidly over those ranges of x in which the X_i's are most densely distributed, as desired.

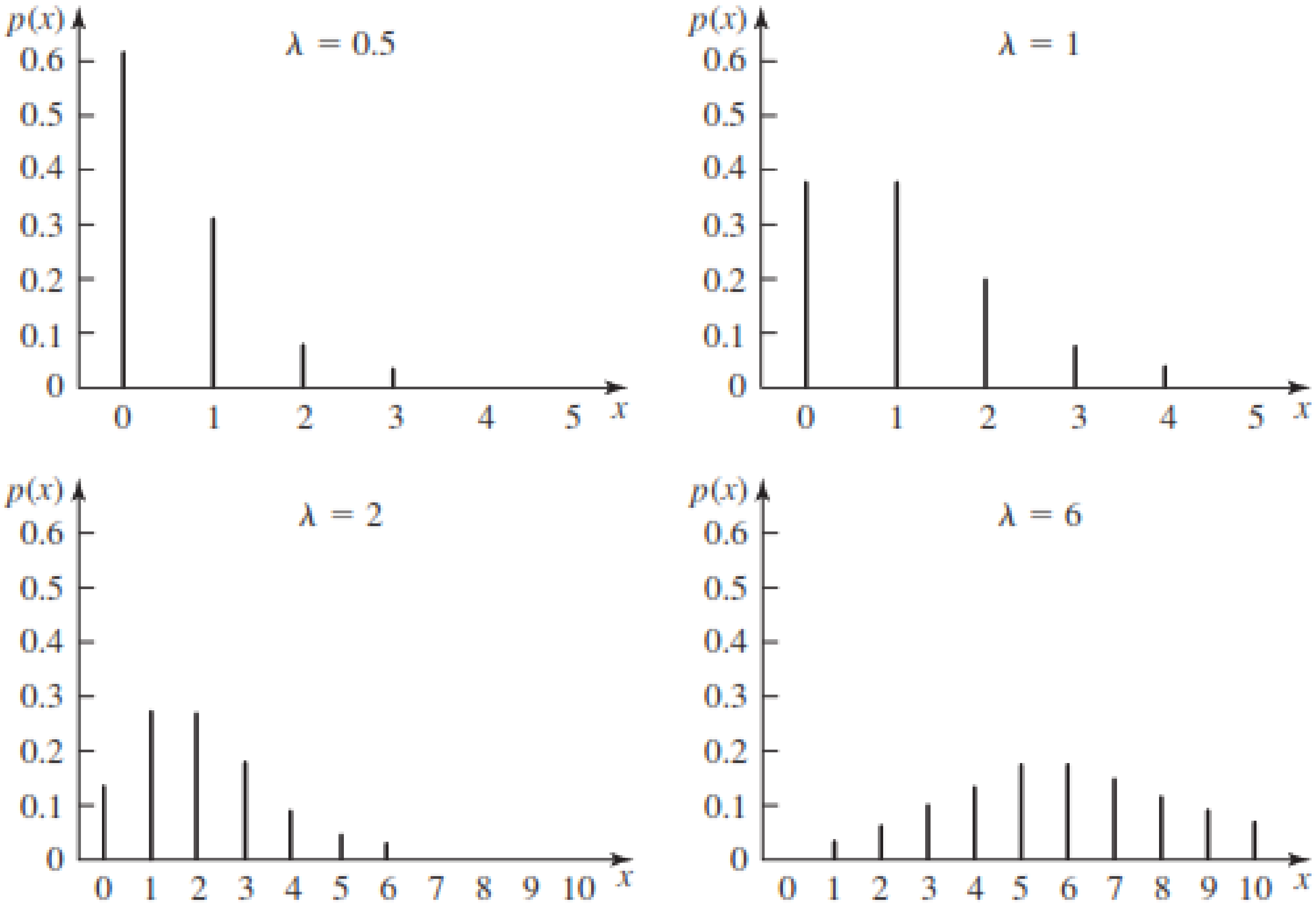


FIGURE 6.23
Poisson(λ) mass functions.

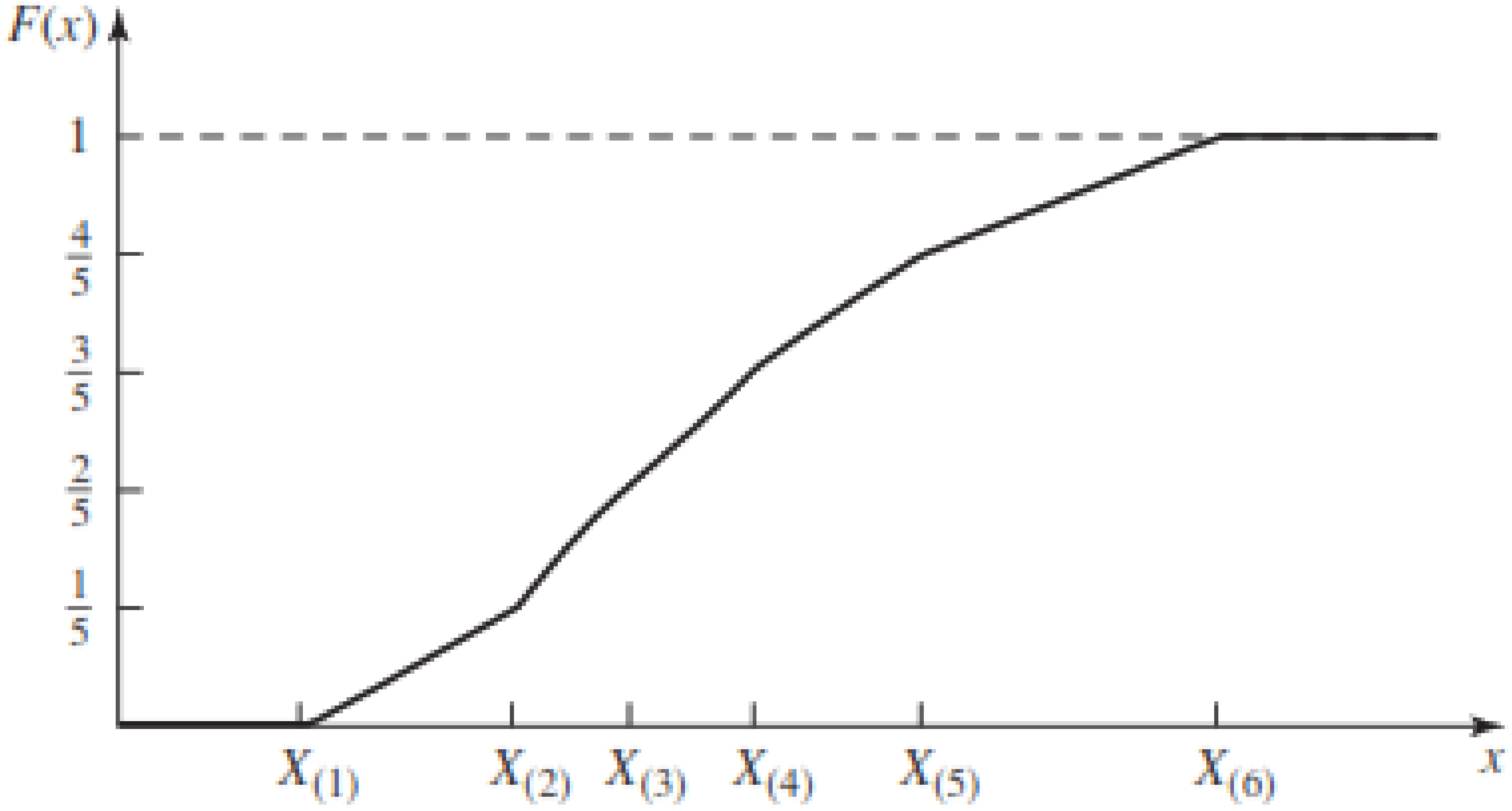


FIGURE 6.24
Continuous, piecewise-linear empirical distribution function from original data.

19. The Crosstowner was a bus that cut a diagonal path from northeast Atlanta to southwest Atlanta. The time required to complete the route was recorded by the bus operator. The bus runs from Monday through Friday. The times of the last fifty 8:00 A.M. runs, in minutes, are as follows:

92.3	92.8	106.8	108.9	106.6
115.2	94.8	106.4	110.0	90.9
104.6	72.0	86.0	102.4	99.8
87.5	111.4	105.9	90.7	99.2
97.8	88.3	97.5	97.4	93.7
99.7	122.7	100.2	106.5	105.5
80.7	107.9	103.2	116.4	101.7
84.8	101.9	99.1	102.2	102.5
111.7	101.5	95.1	92.8	88.5
74.4	98.9	111.9	96.5	95.9

How are these run times distributed? Develop and test a suitable model.

```
In[•]:= Histogram[{92.3, 115.2, 104.6, 87.5, 97.8, 99.7, 80.7, 84.8, 111.7, 74.4, 92.8, 94.8,
   72, 111.4, 88.3, 122.7, 107.9, 101.9, 101.5, 98.9, 106.8, 106.4, 86, 105.9, 97.5,
   100.2, 103.2, 99.1, 95.1, 111.9, 108.9, 110.0, 102.4, 90.7, 97.4, 106.5, 116.4,
   102.2, 92.8, 96.5, 106.6, 90.9, 99.8, 99.2, 93.7, 105.5, 101.7, 102.5, 88.5, 95.9},
  10]
```

Out[•]=