Random Variables & Probability Distributions

Wanwanut Boongsood

Model, Physical system, and Deviation

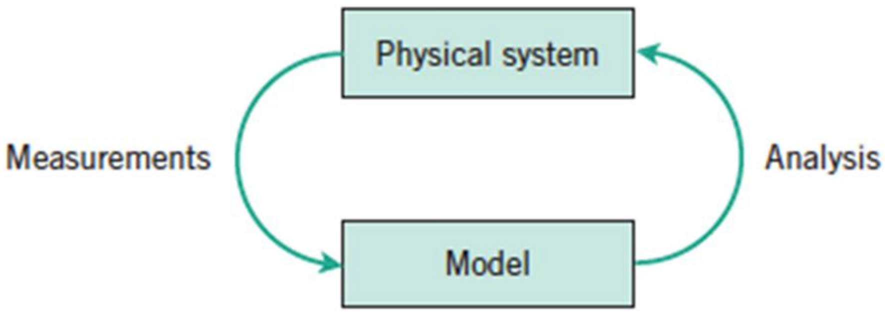


Figure 3-1 Continuous iteration between model and physical system.

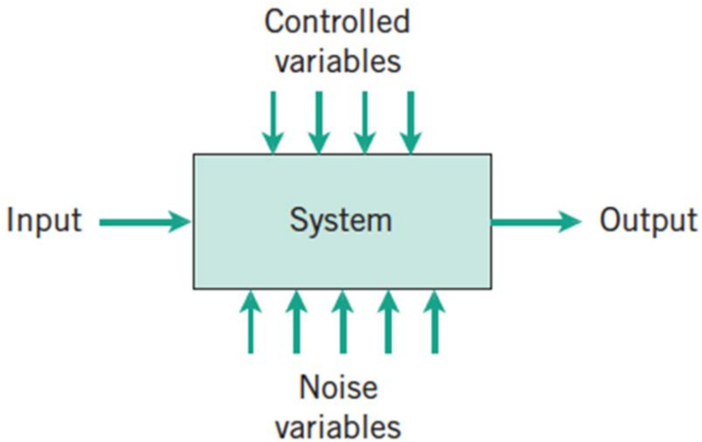


Figure 3-2 Noise variables affect the transformation of inputs to outputs.

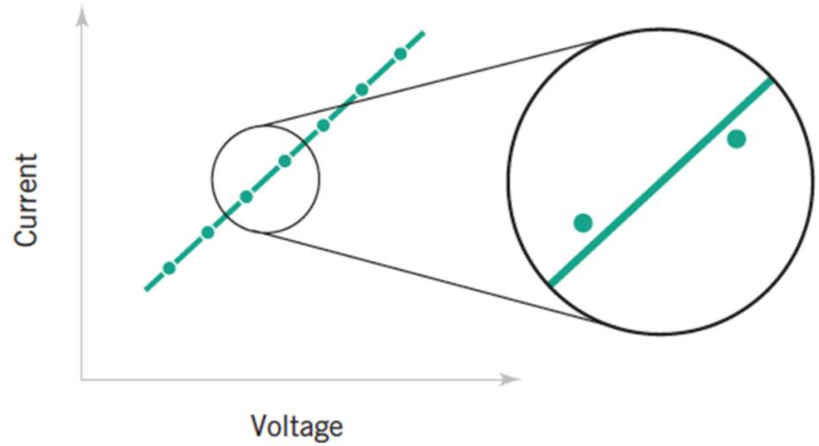


Figure 3-3 A closer examination of the system identifies deviations from the model.

OUTLINE

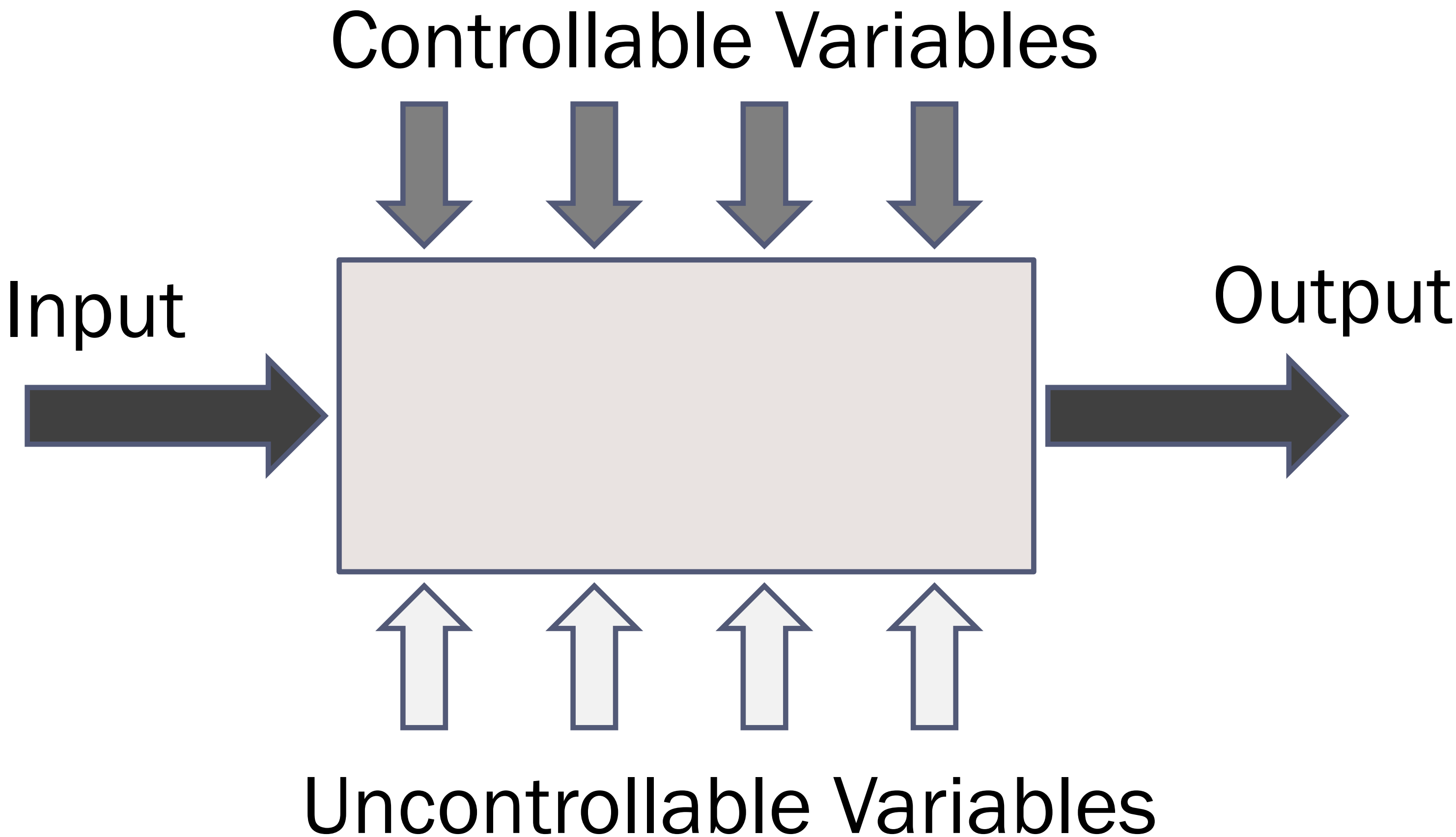
- Types of Random Variables

- The Important Distributions

OUTLINE

- **Types of Random Variables**
- The Important Distributions

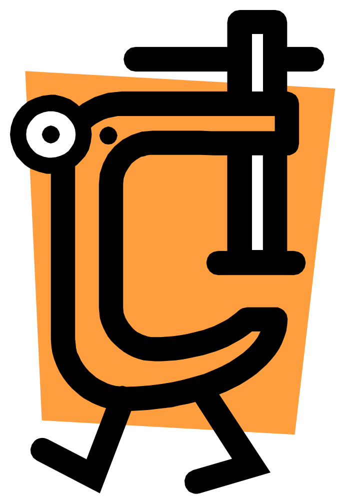
Random Variables

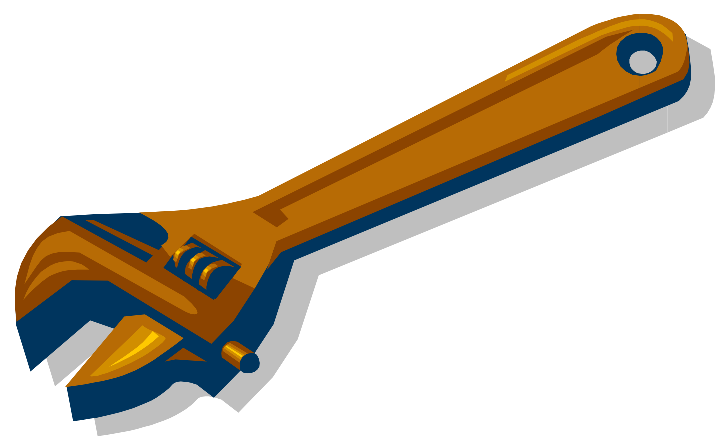


Random Variable: A Numerical variable whose measured value can change from one replicate of the experiment to another

Random Variables

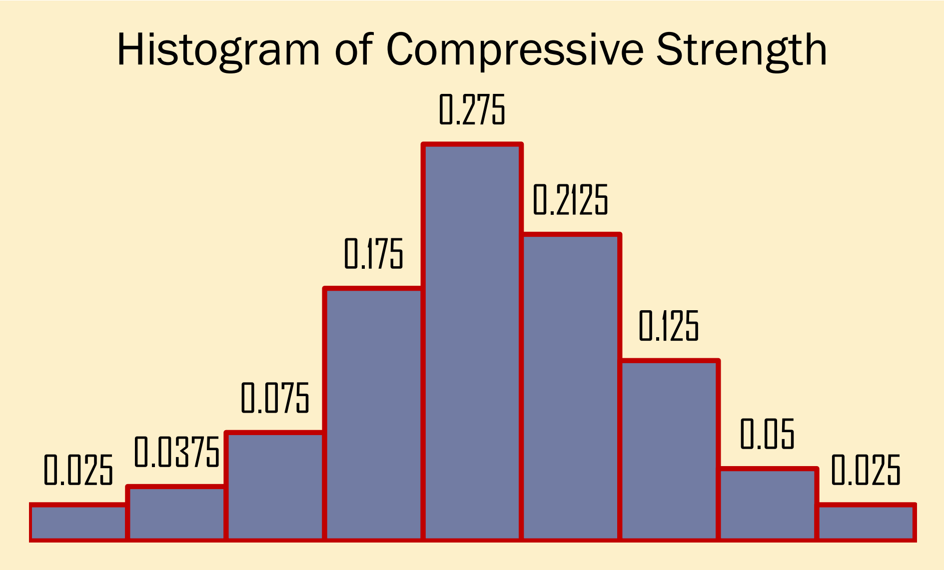
- Discrete random variables
- Continuous random variables

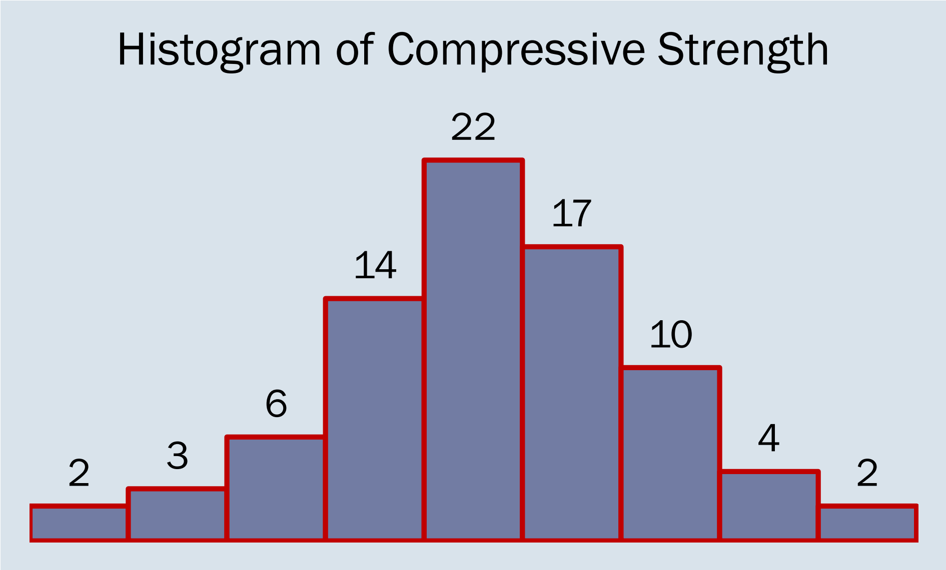




Probability

- The chance of “x”
- A degree of belief
- A relative frequency between “event frequency” to the “outcome frequency”





Probability

- A **probability** is usually expressed in terms of a **random variable**.

- For the part length example, *X* denotes the part length and the probability statement can be written in either of the following forms

$$P(X \in [10.8, 11.2]) = 0.25 \qquad \text{or} \qquad P(10.8 \leq X \leq 11.2) = 0.25$$

- Both equations state that the probability that the random variable *X* assumes a value in [10.8, 11.2] is 0.25.

Probability

Probability Properties

1. $P(X \in R) = 1$, where R is the set of real numbers.
2. $0 \leq P(X \in E) \leq 1$ for any set E. (3-1)
3. If $E_1, E_2, \ldots, E_k$ are mutually exclusive sets,

   $P(X \in E_1 \cup E_2 \cup \ldots \cup E_k) = P(X \in E_1) + \cdots + P(X \in E_k)$.

Continuous Random Variables

- Probability Density Function
- Cumulative Distribution Function
- Mean and Variance

Discrete Random Variables

- Probability Mass Function
- Cumulative Distribution Function
- Mean and Variance

Continuous Random Variables

- Probability Density Function
- Cumulative Distribution Function
- Mean and Variance

Discrete Random Variables

- Probability Mass Function
- Cumulative Distribution Function
- Mean and Variance

Probability Density Function

- The **probability distribution** or simply **distribution** of a random variable *X* is a description of the set of the probabilities associated with the possible values for *X*.

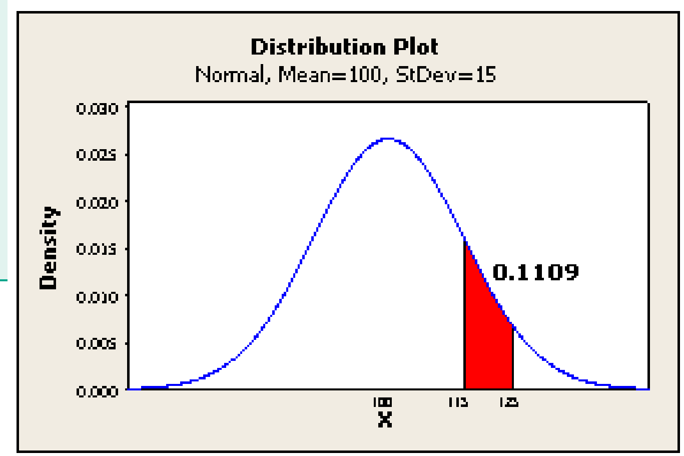
The **probability density function** (or pdf) $f(x)$ of a continuous random variable X is used to determine probabilities as follows:

$$P(a < X < b) = \int_a^b f(x)\,dx \qquad (3\text{-}2)$$

The properties of the pdf are

(1) $f(x) \geq 0$

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



Probability Density Function

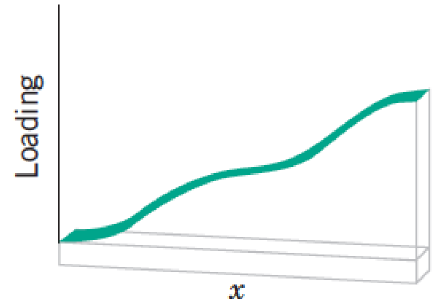


Figure 3-5 Density function of a loading on a long, thin beam.

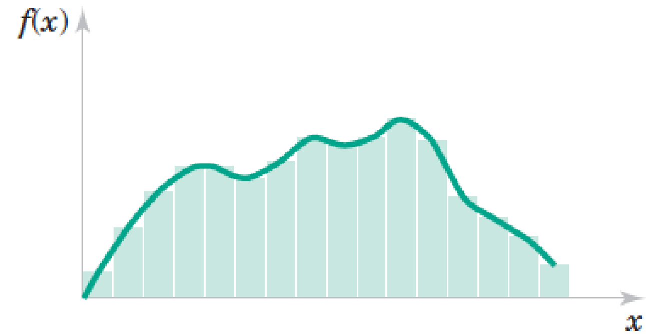


Figure 3-7 A histogram approximates a probability density function. The area of each bar equals the relative frequency of the interval. The area under $f(x)$ over any interval equals the probability of the interval.

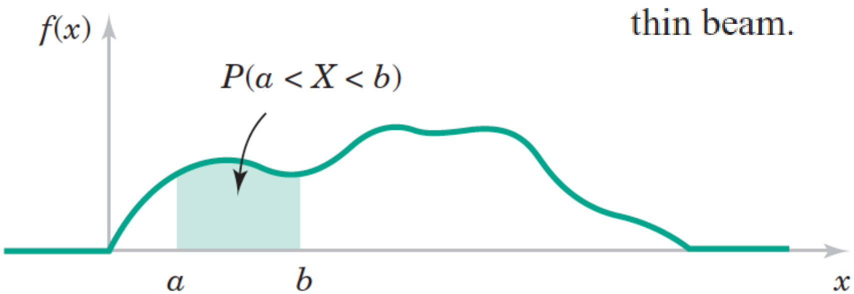


Figure 3-6 Probability determined from the area under $f(x)$.

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

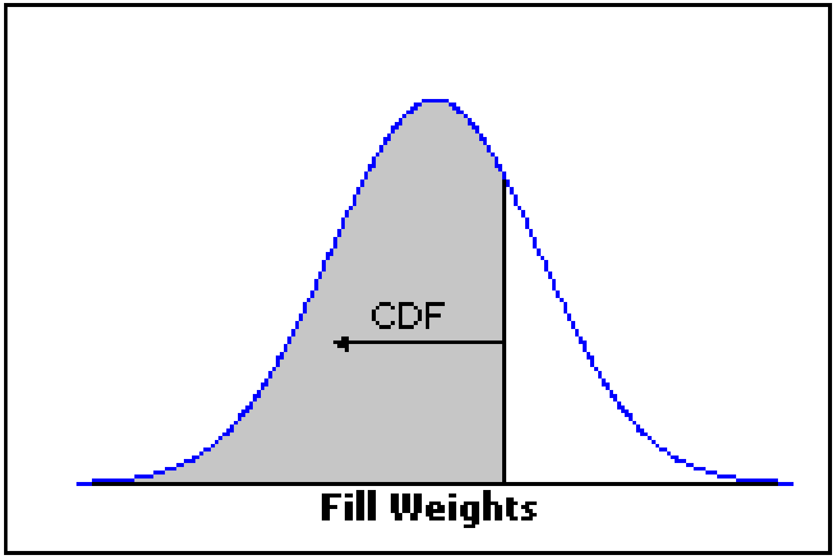
$$P(x_1 \leq X \leq x_2) = P(x_1 < X \leq x_2) = P(x_1 \leq X < x_2) = P(x_1 < X < x_2)$$

Cumulative Distribution Function

The **cumulative distribution function** (or cdf) of a continuous random variable X with probability density function $f(x)$ is

$$F(x) = P(X \le x) = \int_{-\infty}^{x} f(u)\,du$$

for $-\infty < x < \infty$.



Mean & Variance

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

$$\mu = E(X) = \int_{-\infty}^{\infty} x f(x)\, dx \tag{3-3}$$

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

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

The **standard deviation** of X is σ.

Continuous Random Variables

- Probability Density Function
- Cumulative Distribution Function
- Mean and Variance

Discrete Random Variables

- Probability Mass Function
- Cumulative Distribution Function
- Mean and Variance

Probability Mass Function

- For a discrete random variable, the distribution is often specified by just a list of the possible values along with the probability of each.

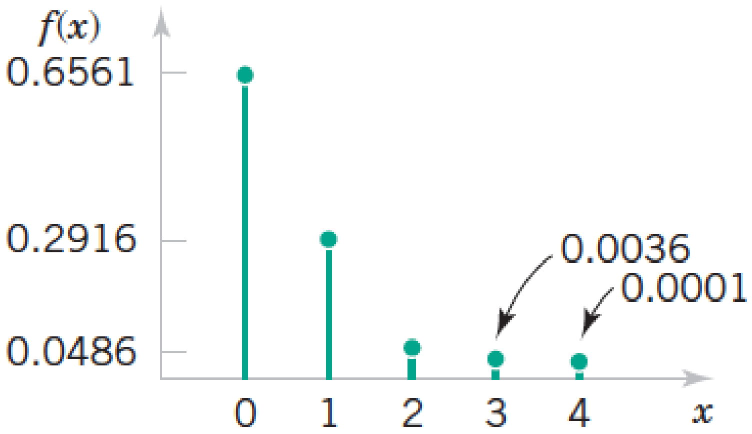


Figure 3-29 Probability distribution for X in Example 3-21.

For a discrete random variable X with possible values $x_1, x_2, \ldots, x_n$, the **probability mass function** (or pmf) is

$$f(x_i) = P(X = x_i) \tag{3-13}$$

Cumulative Distribution Function

The **cumulative distribution function** of a discrete random variable X is

$$F(x) = P(X \leq x) = \sum_{x_i \leq x} f(x_i)$$

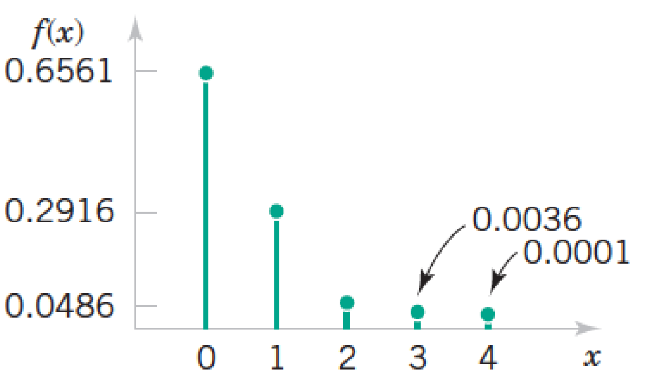


Figure 3-29 Probability distribution for X in Example 3-21.

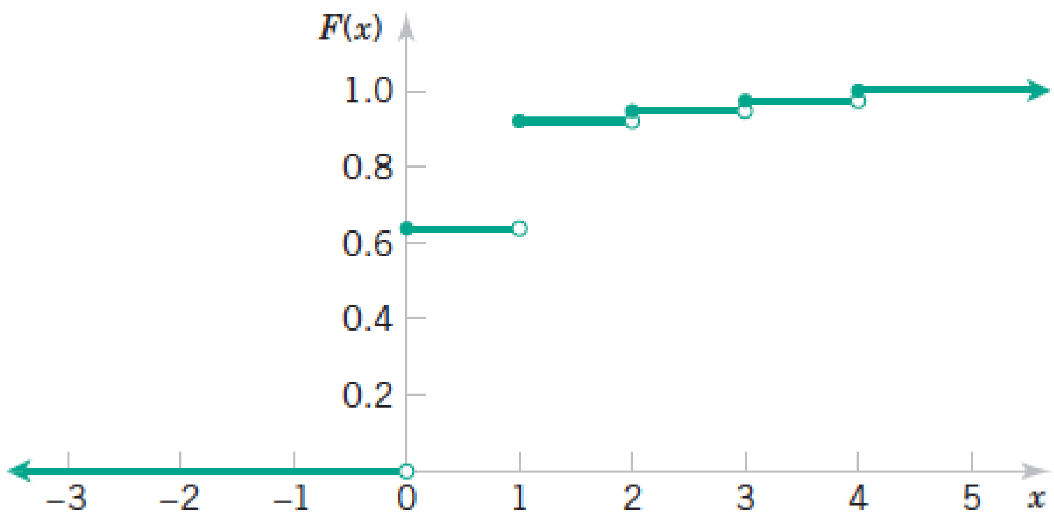


Figure 3-31 Cumulative distribution function for x in Example 3-22.

Mean & Variance

Let the possible values of the random variable X be denoted as $x_1, x_2, \ldots, x_n$. The pmf of X is $f(x)$, so $f(x_i) = P(X = x_i)$.

The **mean** or **expected value** of the discrete random variable X, denoted as μ or $E(X)$, is

$$\mu = E(X) = \sum_{i=1}^{n} x_i f(x_i) \tag{3-14}$$

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

$$\sigma^2 = V(X) = E(X - \mu)^2 = \sum_{i=1}^{n} (x_i - \mu)^2 f(x_i) = \sum_{i=1}^{n} x_i^2 f(x_i) - \mu^2$$

The **standard deviation** of X is σ.

OUTLINE

- Types of Random Variables
- **The Important Distributions**

Important Distributions

Continuous probability distribution

- Normal Distribution and t-Distribution
- Gamma and chi-square Distribution

Discrete Random Variables

- Binomial Distribution
- Poisson and Exponential Distribution (Continuous Probability Distribution)

Normal Approximation to the Binomial and Poisson Distributions

Normal Distribution

- Also referred to as a Gaussian Distribution

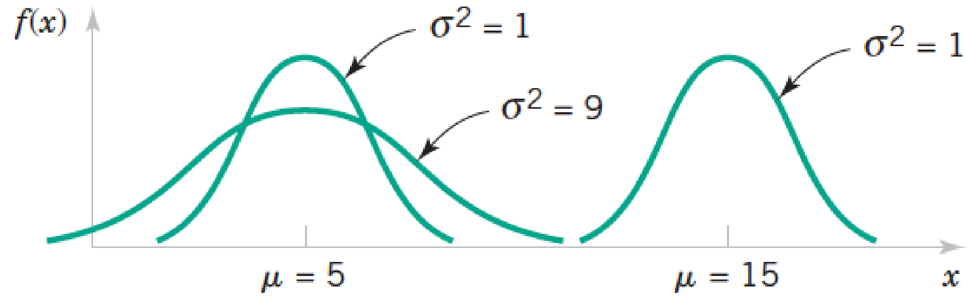


Figure 3-11 Normal probability density functions for selected values of the parameters μ and σ^2.

A random variable X with probability density function

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-(x-\mu)^2}{2\sigma^2}} \quad \text{for } -\infty < x < \infty \tag{3-4}$$

has a **normal distribution** (and is called a **normal random variable**) with parameters μ and σ, where $-\infty < \mu < \infty$, and $\sigma > 0$. Also,

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

The mean and variance of the normal distribution are derived at the end of this section.

Normal Distribution

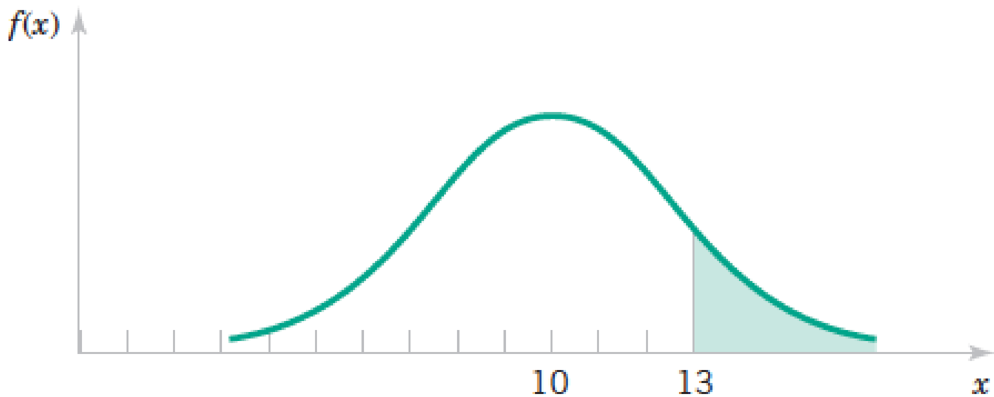
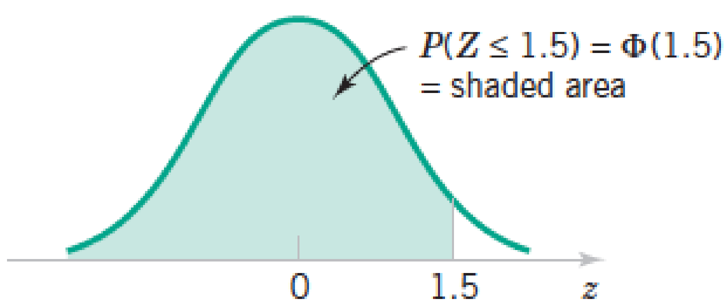


Figure 3-12 Probability that $X > 13$ for a normal random variable with $\mu = 10$ and $\sigma^2 = 4$ in Example 3-7.

$$Z = \frac{X - \mu}{\sigma}$$



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 3-14 Standard normal probability density function.

Normal Distribution: The 68-95-99.7 Rule

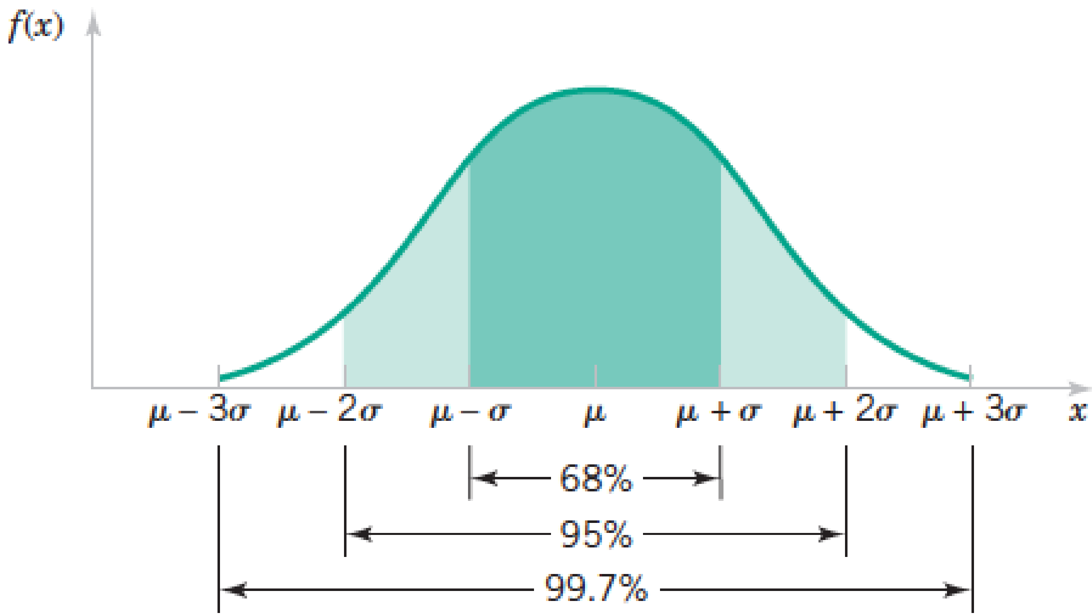


Figure 3-13 Probabilities associated with a normal distribution.

$$Z = \frac{X - \mu}{\sigma}$$

A normal random variable with $\mu = 0$ and $\sigma^2 = 1$ is called a **standard normal** random variable. A standard normal random variable is denoted as Z.

Normal Distribution: How can you tell?

- The normal probability plot

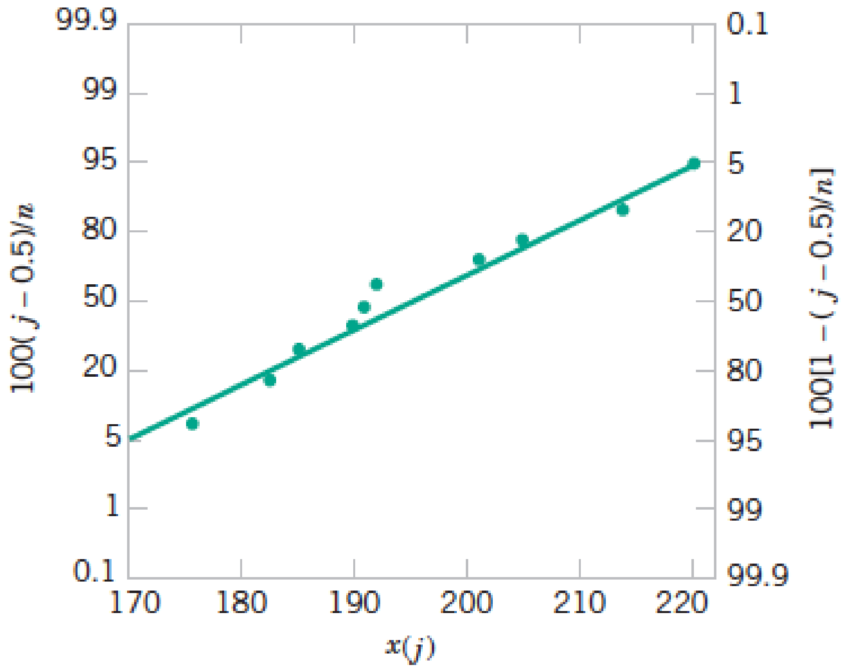
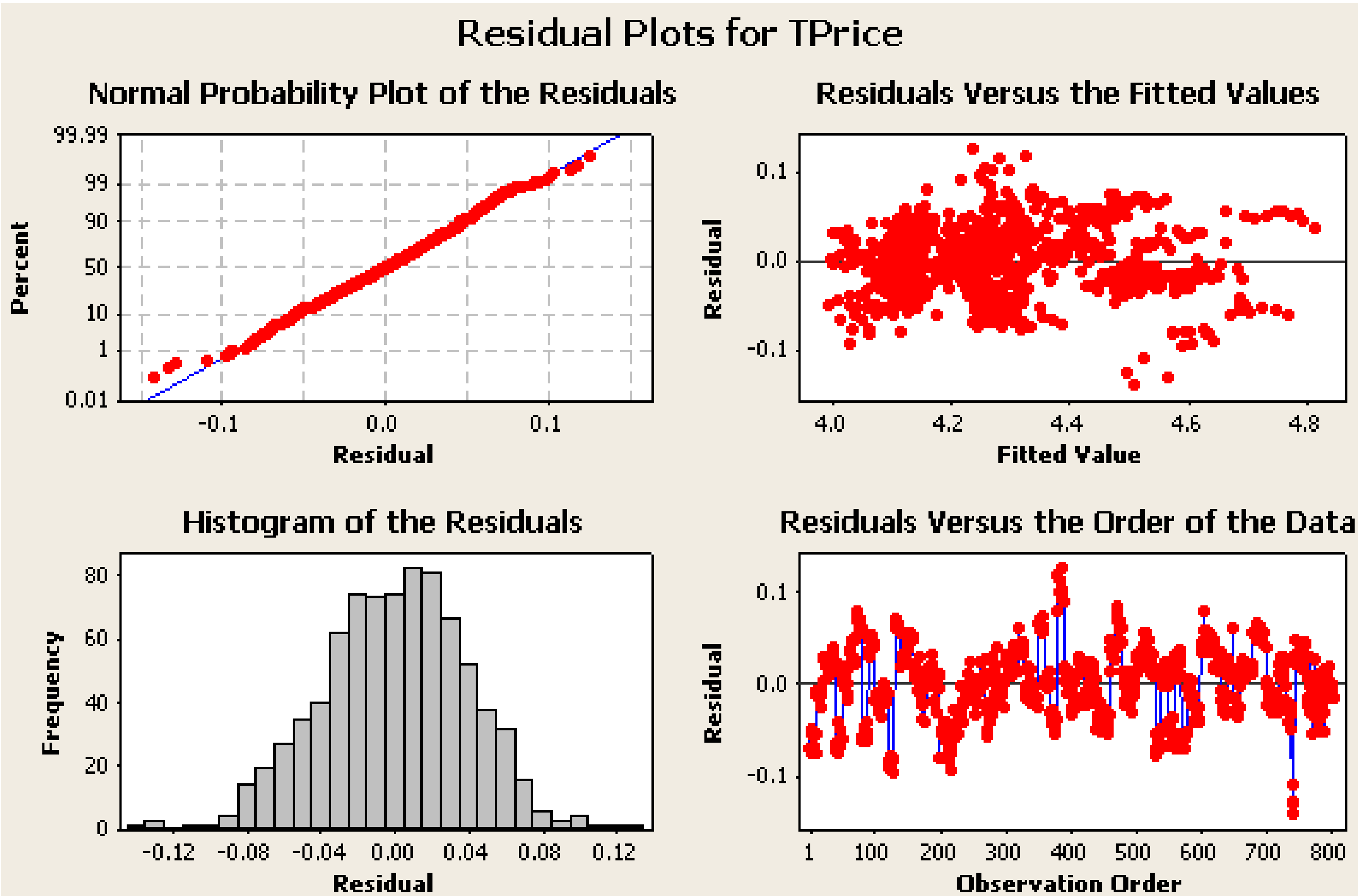


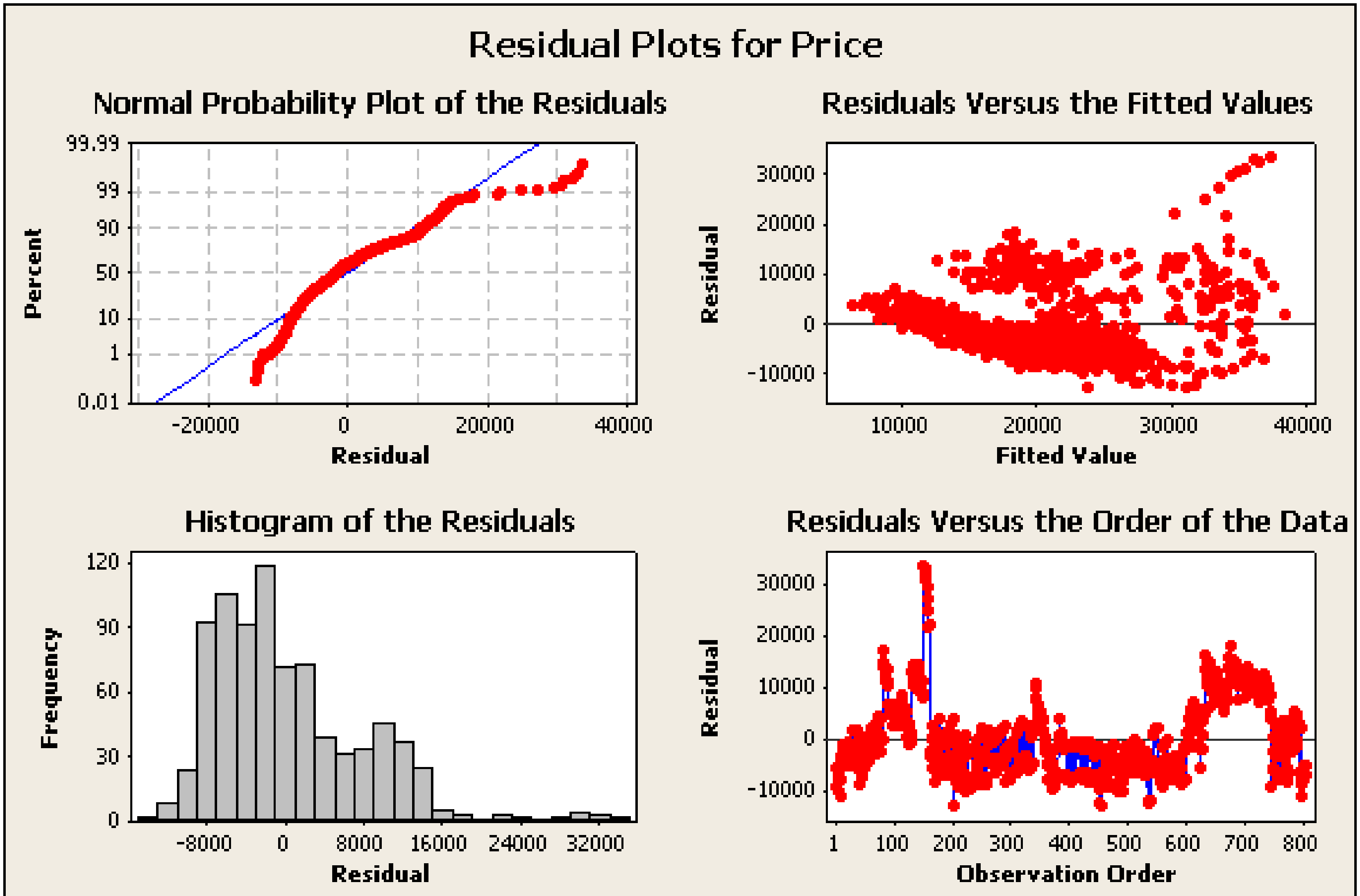
Figure 3-24 Normal probability plot for the battery life.

j	$x_{(j)}$	$(j-0.5)/10$
1	176	0.05
2	183	0.15
3	185	0.25
4	190	0.35
5	191	0.45
6	192	0.55
7	201	0.65
8	205	0.75
9	214	0.85
10	220	0.95

Normal Distribution: How can you tell?



Normal Distribution: How can you tell?



t-Distribution

- When σ is unknown
- Small sample size
- Degree of freedom (k) = n-1
- Significant level = α
- $t_{\alpha, k}$

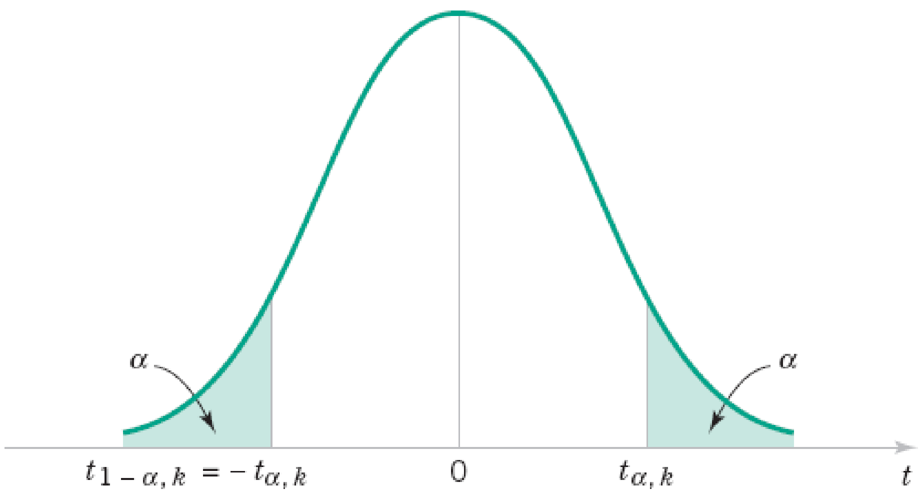


Figure 4-16 Percentage points of the t distribution.

Let $X_1, X_2, \ldots, X_n$ be a random sample for a normal distribution with unknown mean μ and unknown variance σ^2. The quantity

$$T = \frac{\overline{X} - \mu}{S/\sqrt{n}}$$

has a t distribution with $n - 1$ degrees of freedom.

Gamma Distribution

- Very useful for modeling a variety of random experiments.

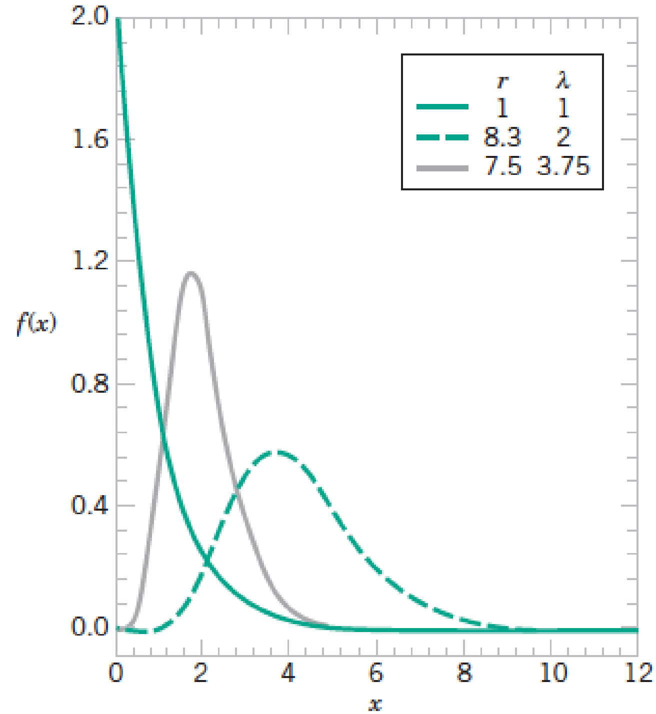


Figure 3-21 Gamma probability density functions for selected values of λ and r.

The random variable X with probability density function

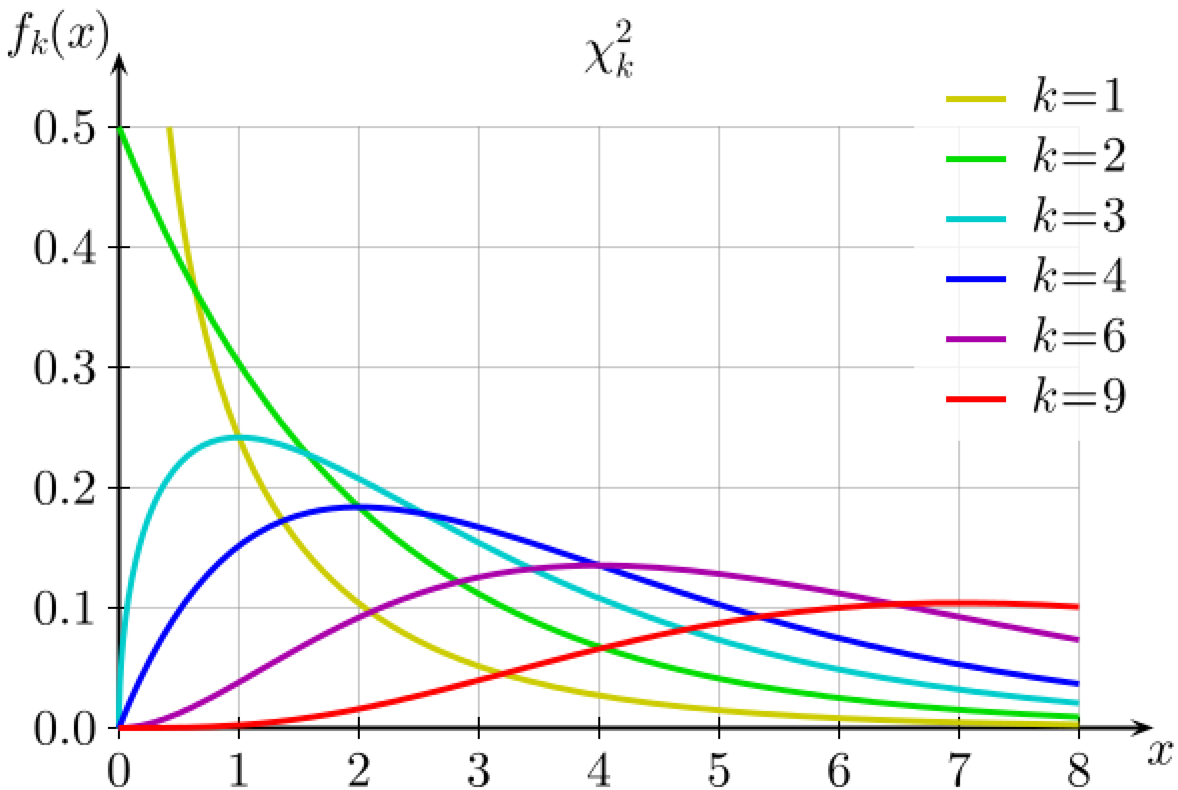
$$f(x) = \frac{\lambda^r x^{r-1} e^{-\lambda x}}{\Gamma(r)}, \quad \text{for } x > 0 \tag{3-9}$$

is a **gamma random variable** with parameters $\lambda > 0$ and $r > 0$. The mean and variance are

$$\mu = E(X) = r/\lambda \quad \text{and} \quad \sigma^2 = V(X) = r/\lambda^2 \tag{3-10}$$

Gamma Distribution

- The **chi-squared distribution** is a special case in which $\lambda=1/2$ and $r=1/2, 1, 3/2, 2,$ or … and used extensively in interval estimation and tests of hypotheses.



Binomial Distribution

- A trial with only **two possible outcomes** is used so frequently as a building block of a random experiment that it is called a **Bernoulli trial.**

- It is usually assumed that the trials that constitute the random experiment are **independent.** This implies that the outcome from one trial has no effect on the outcome to be obtained from any other trial.

- Furthermore, it is often reasonable to assume that the **probability of a success on each trial is constant.**

Binomial Distribution

A Bernoulli Trial

A random experiment consisting of n repeated trials such that

1. the trials are independent,
2. each trial results in only two possible outcomes, labeled as *success* and *failure*, and
3. the probability of a success on each trial, denoted as p, remains constant

is called a *binomial experiment.*

The random variable X that equals the number of trials that result in a *success* has a **binomial distribution** with parameters p and n where $0 \leq p \leq 1$ and $n = \{1, 2, 3, \ldots\}$. The pmf of X is

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x}, \quad x = 0, 1, \ldots, n \qquad (3\text{-}15)$$

Binomial Distribution

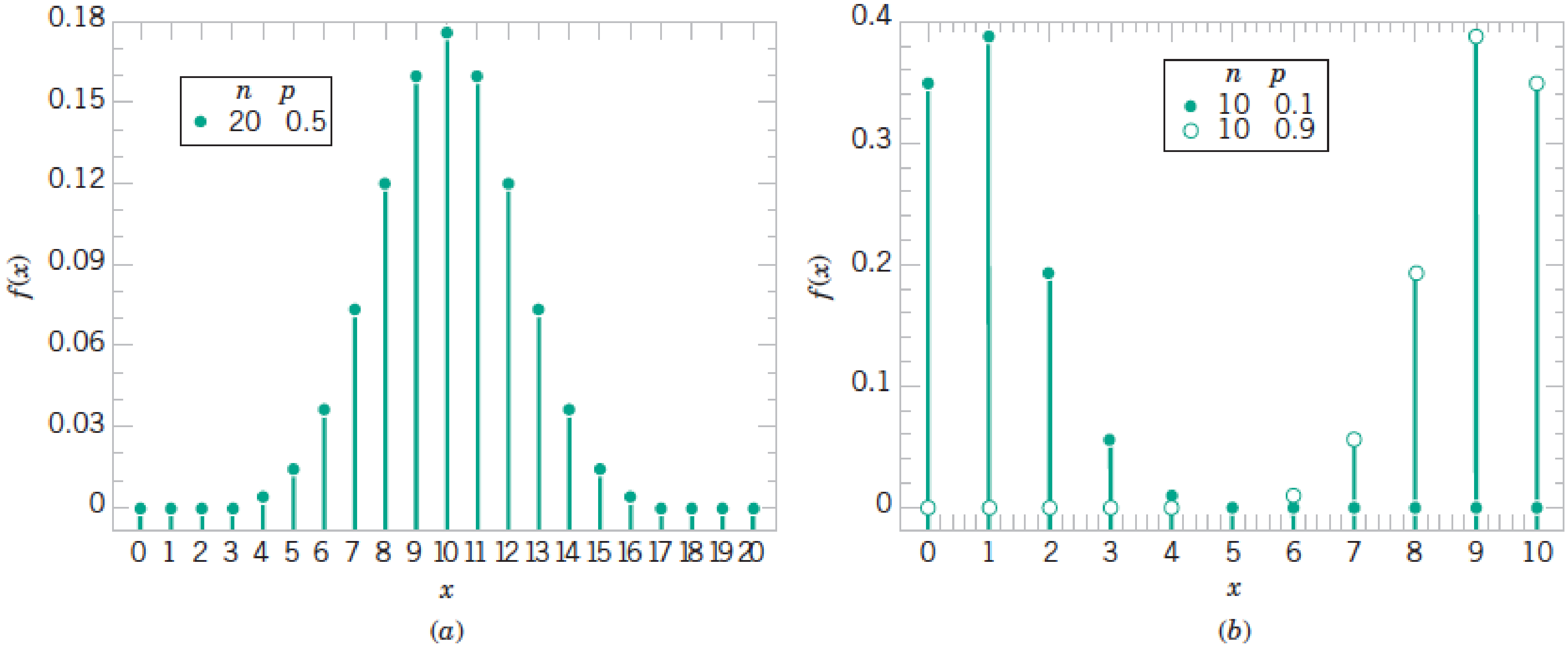


Figure 3-32 Binomial distribution for selected values of n and p.

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

$$\mu = E(X) = np \quad \text{and} \quad \sigma^2 = V(X) = np(1 - p)$$

Poisson Distribution

The random variable X that equals the number of events in a **Poison process** is a Poison random variable with parameter λ>0, and the probability mass function of X is

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

The mean and variance of X are

$$E(x) = \lambda \quad \text{and} \quad V(x) = \lambda$$

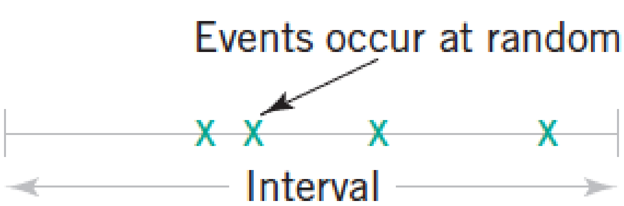


Figure 3-33 In a Poission process, events occur at random in an interval.

Poisson Distribution

In general, consider an interval T of real numbers partitioned into subintervals of small length Δt and assume that as Δt tends to zero,

(1) the probability of more than one event in a subinterval tends to zero,

(2) the probability of one event in a subinterval tends to $\lambda \Delta t / T$,

(3) the event in each subinterval is independent of other subintervals.

A random experiment with these properties is called a **Poisson process.**

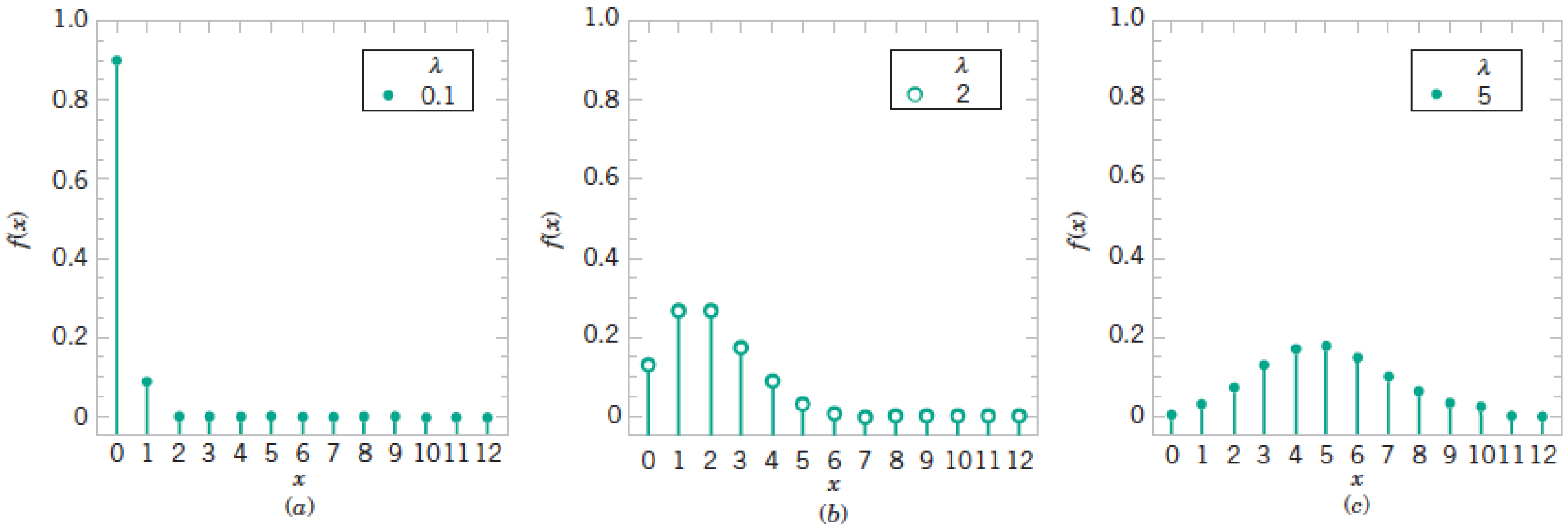


Figure 3-34 Poisson distribution for selected values of the parameter λ.

Poisson Distribution

A Poisson model can be used to approximate a Binomial model when the probability of a success, p, is very small and the number of trials, n, is very large,

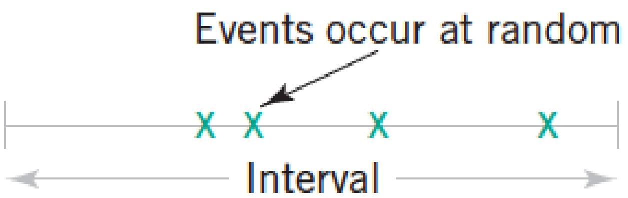
Poisson Process

The events that occur randomly in an interval

The number of events is a discrete random variable that is often modeled by a Poisson distribution

The length of the interval between events is often modeled by a **exponential distribution**

Figure 3-33 In a Poission process, events occur at random in an interval.



Exponential Distribution

The random variable *X* that equals the distance between the successive events of a Poisson process with mean λ>0 has an exponential distribution with parameter λ. The pdf of *X* is

$$f(x) = \lambda e^{-\lambda x}, \qquad \text{for } 0 \le x < \infty$$

The mean and variance of *X* are

$$E(X) = \frac{1}{\lambda} \quad \text{and} \quad V(X) = \frac{1}{\lambda^2}$$

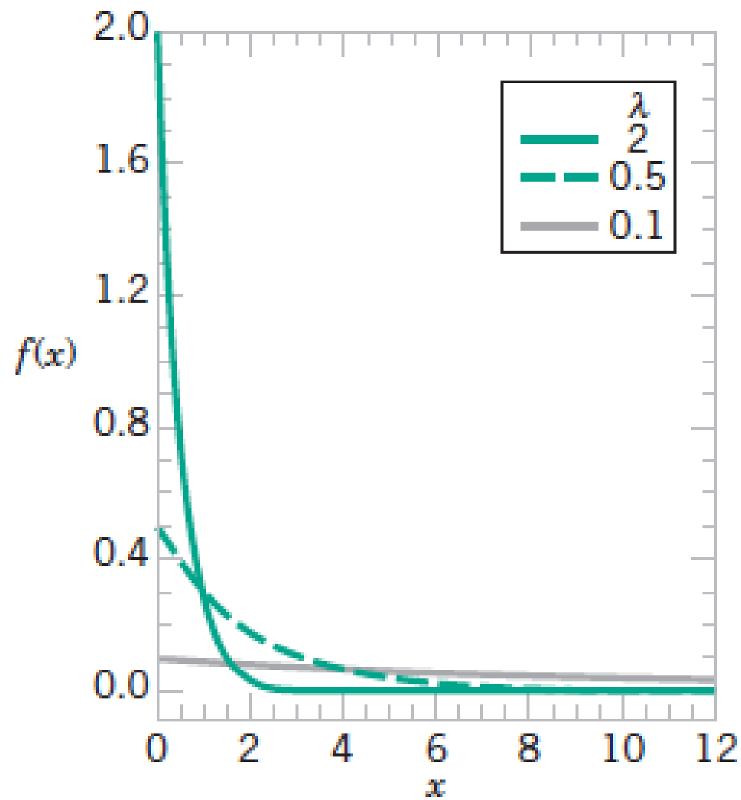


Figure 3-35 Probability density function of an exponential random variable for selected values of λ.

Normal Approximation to Binomial Distribution

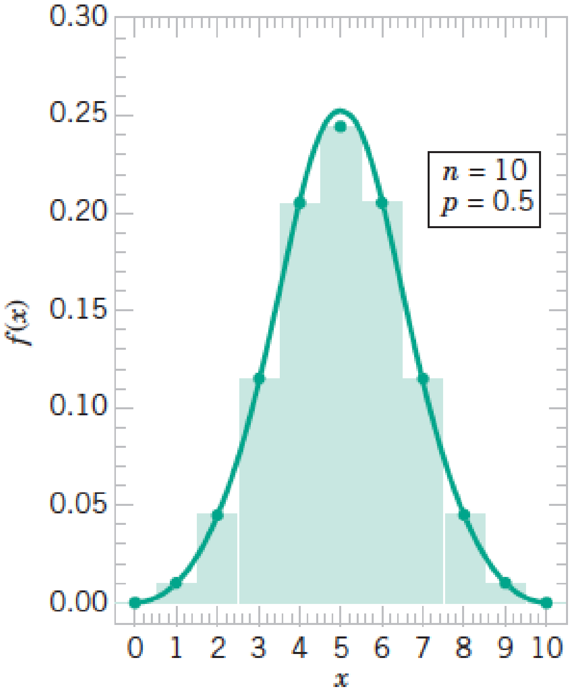


Figure 3-37 Normal approximation to the binomial distribution.

Normal model can approximate the Binomial – a much easier method

Normal model works pretty well if we expect to see at least 10 successes and 10 failures

$np \geq 10$ and $nq \geq 10$

If X is a binomial random variable,

$$Z = \frac{X - np}{\sqrt{np(1-p)}} \tag{3-21}$$

is approximately a standard normal random variable. Consequently, probabilities computed from Z can be used to approximate probabilities for X.

Why 10?

We require:	$\mu - 3\sigma > 0$
Or, in other words:	$\mu > 3\sigma$
For a Binomial that's:	$np > 3\sqrt{npq}$
Squaring yields:	$n^2p^2 > 9npq$
Now simplify:	$np > 9q$
Since $q \leq 1$, we can require:	$mp > 9$

Normal Approximation to Poisson Distribution

For large n when np is small, consider using a Poisson model instead.

The approximation is good for

$$\lambda > 5$$

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

$$Z = \frac{X - \lambda}{\sqrt{\lambda}} \qquad (3\text{-}22)$$

is approximately a standard normal random variable.