

2

Modeling Process Quality

CHAPTER OUTLINE

2-1 DESCRIBING VARIATION

2-1.1 The Stem-and-Leaf Plot

2-1.2 The Histogram

2-1.3 Numerical Summary of Data

2-1.4 The Box Plot

2-1.5 Probability Distributions

2-2 IMPORTANT DISCRETE DISTRIBUTIONS

2-2.1 The Hypergeometric Distribution

2-2.2 The Binomial Distribution

2-2.3 The Poisson Distribution

2-2.4 The Pascal and Related Distributions

2-3 IMPORTANT CONTINUOUS DISTRIBUTIONS

2-3.1 The Normal Distribution

2-3.2 The Lognormal Distribution

2-3.3 The Exponential Distribution

2-3.4 The Gamma Distribution

2-3.5 The Weibull Distribution

2-4 PROBABILITY PLOTS

2-4.1 Normal Probability Plots

2-4.2 Other Probability Plots

2-5 SOME USEFUL APPROXIMATIONS

2-5.1 The Binomial Approximation to the Hypergeometric

2-5.2 The Poisson Approximation to the Binomial

2-5.3 The Normal Approximation to the Binomial

2-5.4 Comments on Approximations

Supplemental Material for Chapter 2

S2-1 Independent Random Variables

S2-2 Development of the Poisson Distribution

S2-3 The Mean and Variance of the Normal Distribution

S2-4 More about the Lognormal Distribution

S2-5 More about the Gamma Distribution

S2-6 The Failure Rate for the Exponential Distribution

S2-7 The Failure Rate for the Weibull Distribution

The supplemental material is on the textbook website www.wiley.com/college/montgomery.

Learning Objectives

1. Construct and interpret visual data displays, including the stem-and-leaf plot, the histogram, and the box plot
2. Compute and interpret the sample mean, the sample variance, the sample standard deviation, and the sample range
3. Explain the concepts of a random variable and a probability distribution
4. Understand and interpret the mean, variance, and standard deviation of a probability distribution
5. Determine probabilities from probability distributions
6. Understand the assumptions for each of the discrete probability distributions presented
7. Understand the assumptions for each of the continuous probability distributions presented
8. Select an appropriate probability distribution for use in specific applications
9. Use probability plots
10. Use approximations for some hypergeometric and binomial distributions

2-1 DESCRIBING VARIATION

Stem-and-Leaf Display

Table 2-1 Days to Pay Employee Health Insurance Claims

Claim	Days	Claim	Days	Claim	Days	Claim	Days
1	48	11	35	21	37	31	16
2	41	12	34	22	43	32	22
3	35	13	36	23	17	33	33
4	36	14	42	24	26	34	30
5	37	15	43	25	28	35	24
6	26	16	36	26	27	36	23
7	36	17	56	27	45	37	22
8	46	18	32	28	33	38	30
9	35	19	46	29	22	39	31
10	47	20	30	30	27	40	17

Stem-and-Leaf Display: Days

Stem-and-leaf of Days

N = 40

Leaf Unit = 1.0

```

 3 1 677
 8 2 22234
13 2 66778
(8) 3 00012334
19 3 555666677
10 4 02224
 6 4 56678
 1 5
 1 5 6
```

Figure 2-1 Stem-and-leaf plot for the health insurance claim data.

- Easy to find percentiles of the data; see page 43

Plot of Data in Time Order

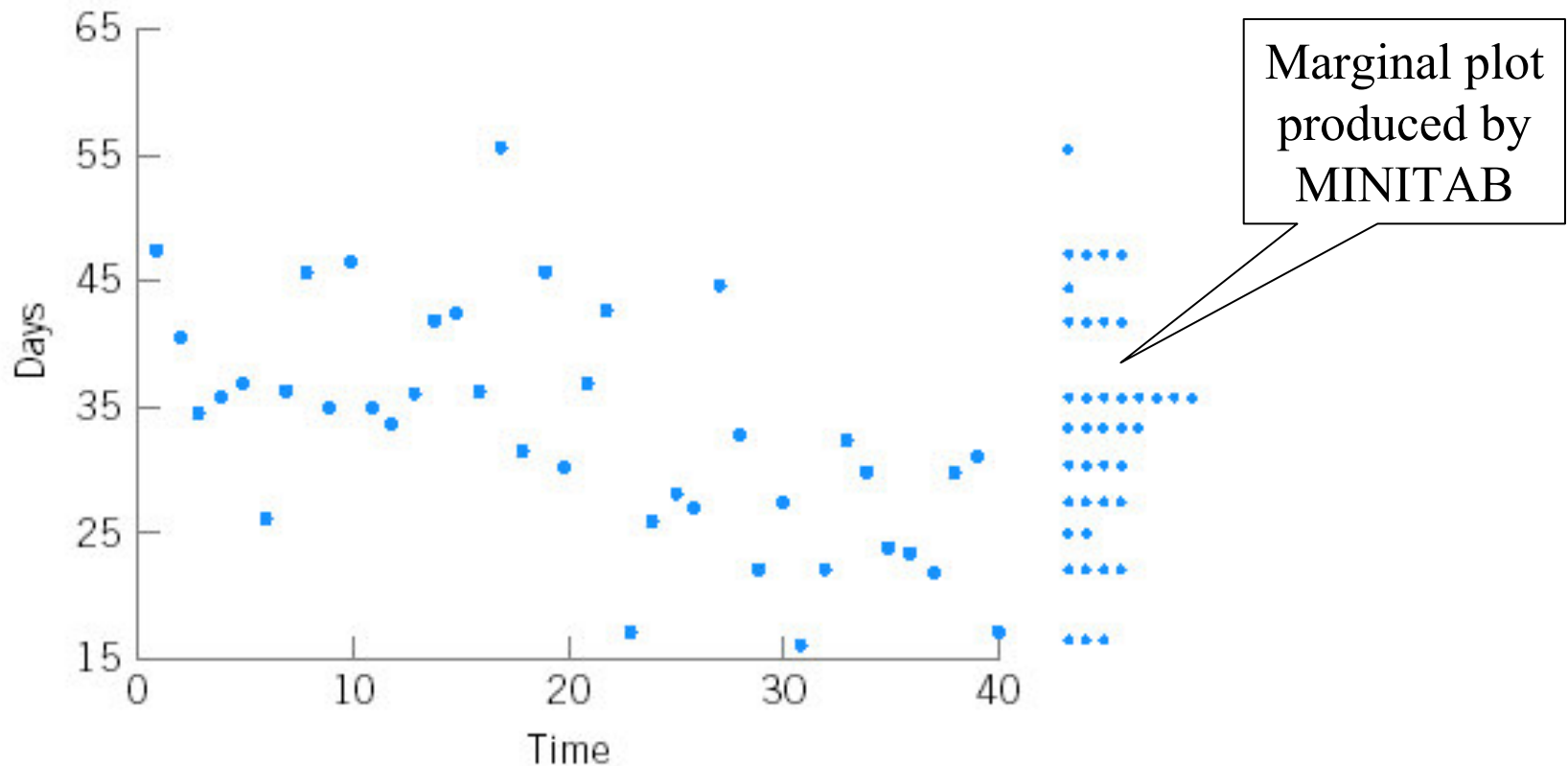


Figure 2-2 A time series plot of the health insurance data in Table 2-1.

- Also called a *run chart*

Histograms – Useful for large data sets

Table 2-2 Layer Thickness (Å) on Semiconductor Wafers

438	450	487	451	452	441	444	461	432	471
413	450	430	437	465	444	471	453	431	458
444	450	446	444	466	458	471	452	455	445
468	459	450	453	473	454	458	438	447	463
445	466	456	434	471	437	459	445	454	423
472	470	433	454	464	443	449	435	435	451
474	457	455	448	478	465	462	454	425	440
454	441	459	435	446	435	460	428	449	442
455	450	423	432	459	444	445	454	449	441
449	445	455	441	464	457	437	434	452	439

- Group values of the variable into bins, then count the number of observations that fall into each bin
- Plot frequency (or relative frequency) versus the values of the variable

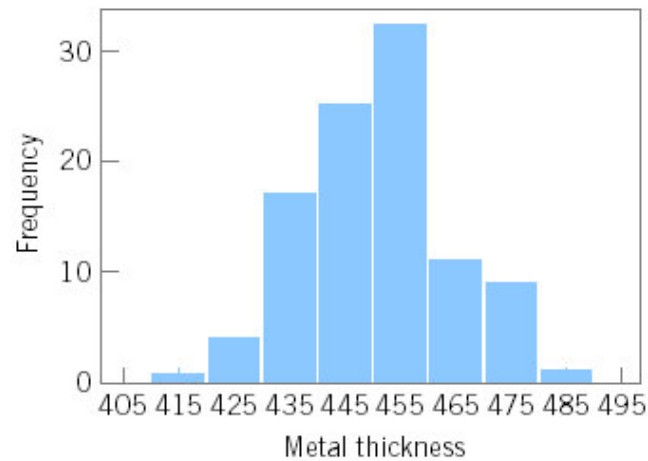


Figure 2-3 Minitab histogram for the metal layer thickness data in Table 2-2.

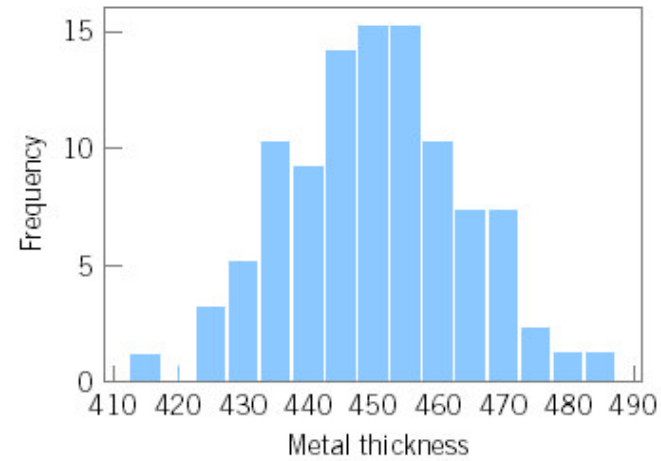


Figure 2-4 Minitab histogram with 15 bins for the metal layer thickness data.

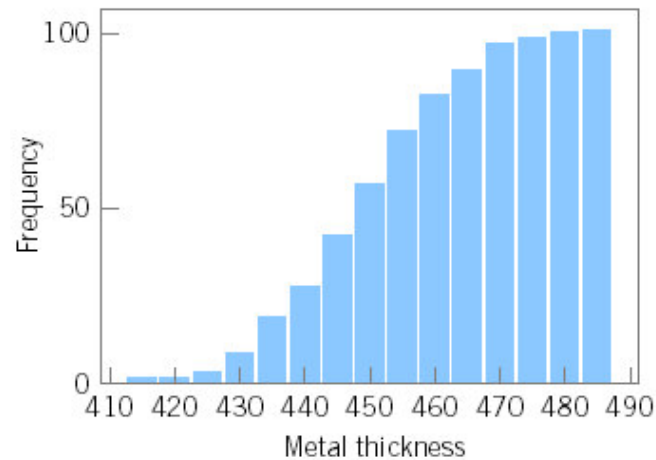


Figure 2-5 A cumulative frequency plot of the metal thickness data from Minitab.

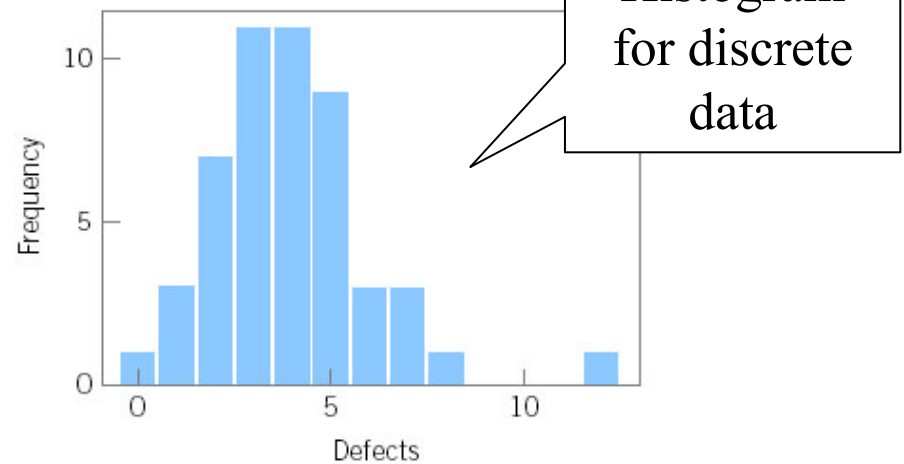


Figure 2-6 Histogram of the number of defects in painted automobile hoods (Table 2-3).

Table 2-3 Surface Finish Defects in Painted Automobile Hoods

6	1	5	7	8	6	0	2	4	2
5	2	4	4	1	4	1	7	2	3
4	3	3	3	6	3	2	3	4	5
5	2	3	4	4	4	2	3	5	7
5	4	5	5	4	5	3	3	3	12

Figure 2-6 is the histogram of the defects. Notice that the number of defects is a discrete variable. From either the histogram or the tabulated data we can determine

$$\text{Proportions of hoods with at least 3 defects} = \frac{39}{50} = 0.78$$

and

$$\text{Proportions of hoods with between 0 and 2 defects} = \frac{11}{50} = 0.22$$

These proportions are examples of relative frequencies.

Numerical Summary of Data

Suppose that $x_1, x_2, \dots, x_n$ are the observations in a sample. The most important measure of central tendency in the sample is the **sample average**,

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n} = \frac{\sum_{i=1}^n x_i}{n} \quad (2-1)$$

Note that the sample average $\bar{x}$ is simply the arithmetic mean of the n observations. The sample average for the metal thickness data in Table 2-2 is

$$\bar{x} = \frac{\sum_{i=1}^{100} x_i}{100} = \frac{45.001}{100} = 450.01 \text{ \AA}$$

Refer to Fig. 2-3 and note that the sample average is the point at which the histogram exactly “balances.” Thus, the sample average represents the center of mass of the sample data.

The variability in the sample data is measured by the **sample variance**,

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1} \quad (2-2)$$

Note that the sample variance is simply the sum of the squared deviations of each observation from the sample average $\bar{x}$, divided by the sample size minus one. If there is no variability in the sample, then each sample observation $x_i = \bar{x}$, and the sample variance $s^2 = 0$. Generally, the larger is the sample variance s^2 , the greater is the variability in the sample data.

The units of the sample variance s^2 are the square of the original units of the data. This is often inconvenient and awkward to interpret, and so we usually prefer to use the square root of s^2 , called the **sample standard deviation** s , as a measure of variability.

It follows that

$$s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}} \quad (2-3)$$

The primary advantage of the sample standard deviation is that it is expressed in the original units of measurement. For the metal thickness data, we find that

$$s^2 = 180.2928 \text{ Å}^2$$

and

$$s = 13.43 \text{ Å}$$

The Box Plot (or Box-and-Whisker Plot)

Table 2-4 Hole Diameters (in mm) in Wing Leading Edge Ribs

120.5	120.4	120.7
120.9	120.2	121.1
120.3	120.1	120.9
121.3	120.5	120.8

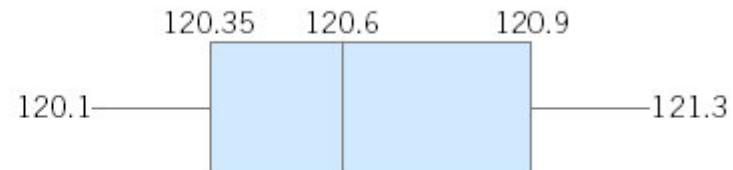


Figure 2-7 Box plot for the aircraft wing leading edge hole diameter data in Table 2-4.

Comparative Box Plots

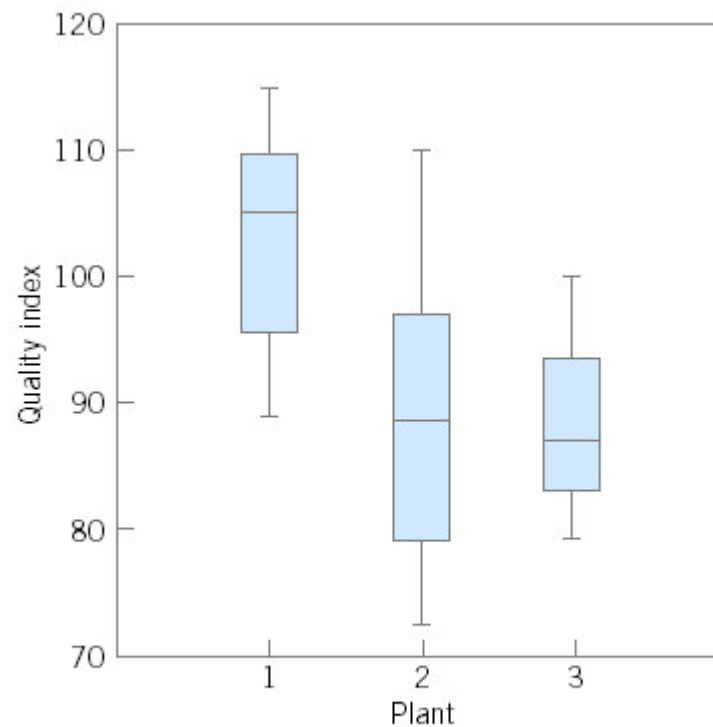


Figure 2-8 Comparative box plots of a quality index for products produced at three plants.

Probability Distributions

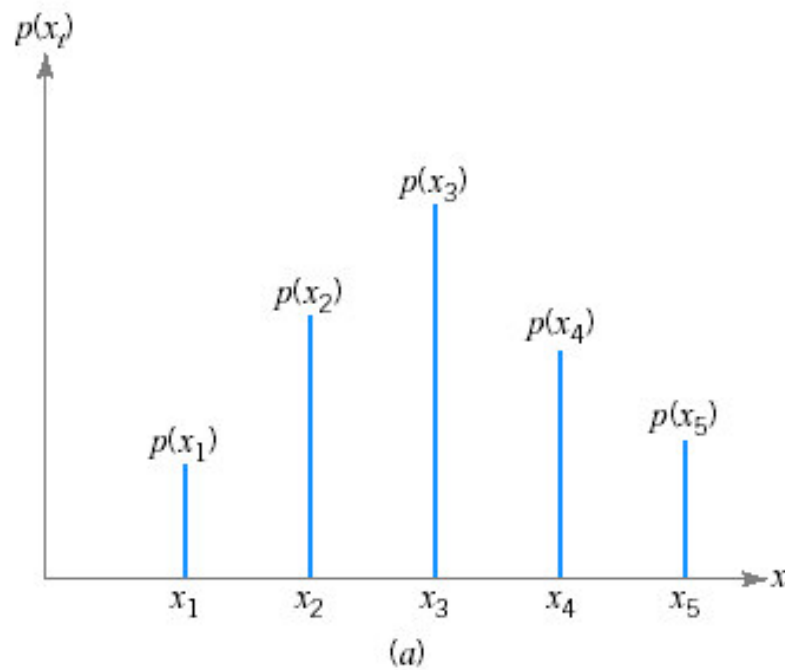
The histogram (or stem-and-leaf plot, or box plot) is used to describe *sample* data. A **sample** is a collection of measurements selected from some larger source or **population**. For example, the measurements on layer thickness in Table 2-2 are obtained from a sample of wafers selected from the manufacturing process. The population in this example is the collection of all layer thicknesses produced by that process. By using statistical methods, we may be able to analyze the sample layer thickness data and draw certain conclusions about the process that manufactures the wafers.

A **probability distribution** is a mathematical model that relates the value of the variable with the probability of occurrence of that value in the population. In other words, we might visualize layer thickness as a **random variable**, because it takes on different values in the population according to some random mechanism, and then the probability distribution of layer thickness describes the probability of occurrence of any value of layer thickness in the population. There are two types of probability distributions.

Definition

1. **Continuous distributions.** When the variable being measured is expressed on a continuous scale, its probability distribution is called a *continuous distribution*. The probability distribution of metal layer thickness is continuous.
2. **Discrete distributions.** When the parameter being measured can only take on certain values, such as the integers 0, 1, 2, . . . , the probability distribution is called a *discrete distribution*. For example, the distribution of the number of non-conformities or defects in printed circuit boards would be a discrete distribution.

Sometimes called a
probability mass function



Sometimes called a
probability density function

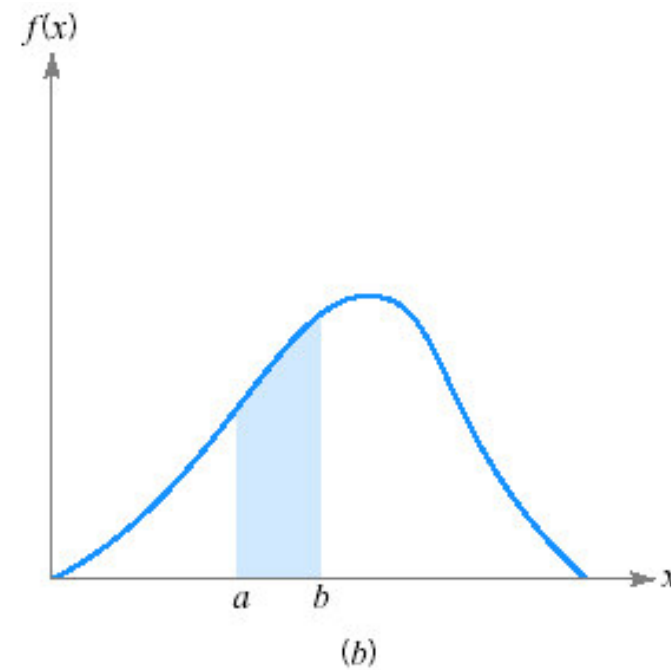


Figure 2-9 Probability distributions. (a) Discrete case. (b) Continuous case.

Will see many examples in the text

..... **EXAMPLE 2-5**

A Discrete Distribution

A manufacturing process produces thousands of semiconductor chips per day. On the average, 1% of these chips do not conform to specifications. Every hour, an inspector selects a random sample of 25 chips and classifies each chip in the sample as conforming or nonconforming. If we let x be the random variable representing the number of nonconforming chips in the sample, then the probability distribution of x is

$$p(x) = \binom{25}{x} (0.01)^x (0.99)^{25-x} \quad x = 0, 1, 2, \dots, 25$$

where $\binom{25}{x} = 25!/[x!(25-x)!]$. This is a *discrete* distribution, since the observed number of nonconformances is $x = 0, 1, 2, \dots, 25$, and is called the **binomial distribution**. We may calculate the probability of finding one or fewer nonconforming parts in the sample as

$$\begin{aligned} P(x \leq 1) &= P(x = 0) + P(x = 1) \\ &= p(0) + p(1) \\ &= \sum_{x=0}^1 \binom{25}{x} (0.01)^x (0.99)^{25-x} \\ &= \frac{25!}{0!25!} (0.99)^{25} (0.01)^0 + \frac{25!}{1!24!} (0.99)^{24} (0.01)^1 \\ &= 0.7778 + 0.1964 = 0.9742 \end{aligned}$$

.....

..... **EXAMPLE 2-6**

A Continuous Distribution

Suppose that x is a random variable that represents the actual contents in ounces of a 1-lb bag of coffee beans. The probability distribution of x is assumed to be

$$f(x) = \frac{1}{1.5} \quad 15.5 \leq x \leq 17.0$$

This is a *continuous* distribution, since the range of x is the interval $[15.5, 17.0]$. This distribution is called the **uniform distribution**, and it is shown graphically in Fig. 2-10. Note that the area under the function $f(x)$ corresponds to probability, so that the probability of a bag containing less than 16.0 oz is

$$P\{x \leq 16.0\} = \int_{15.5}^{16.0} f(x) dx = \int_{15.5}^{16.0} \frac{1}{1.5} dx = \left. \frac{x}{1.5} \right|_{15.5}^{16.0} = \frac{16.0 - 15.5}{1.5} = 0.3333$$

This follows intuitively from inspection of Fig. 2-9.

.....

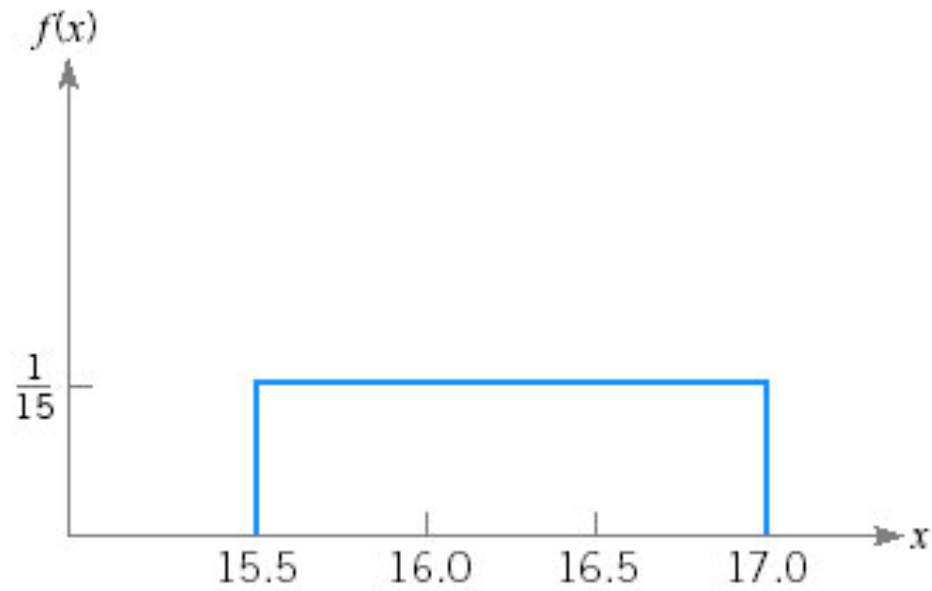


Figure 2-10 The uniform distribution for Example 2-6.

The **mean** μ of a probability distribution is a measure of the **central tendency** in the distribution, or its **location**. The mean is defined as

$$\mu = \begin{cases} \int_{-\infty}^{\infty} xf(x) dx, & x \text{ continuous} \\ \sum_{i=1}^{\infty} x_i p(x_i), & x \text{ discrete} \end{cases} \quad \begin{matrix} (2-5a) \\ (2-5b) \end{matrix}$$

For the case of a discrete random variable with exactly N equally likely values [that is, $p(x_i) = 1/N$], then equation 2-5b reduces to

$$\mu = \frac{\sum_{i=1}^N x_i}{N}$$

Note the similarity of this last expression to the sample average $\bar{x}$ defined in equation 2-1.

The mean is the point at which the distribution exactly “balances” (see Fig. 2-11). Thus, the mean is simply the center of mass of the probability distribution. Note from Fig. 2-11*b* that the mean is not necessarily the fiftieth percentile of the distribution (the **median**), and from Fig. 2-11*c* it is not necessarily the most likely value of the variable (which is called the **mode**). The mean simply determines the **location** of the distribution, as shown in Fig. 2-12.

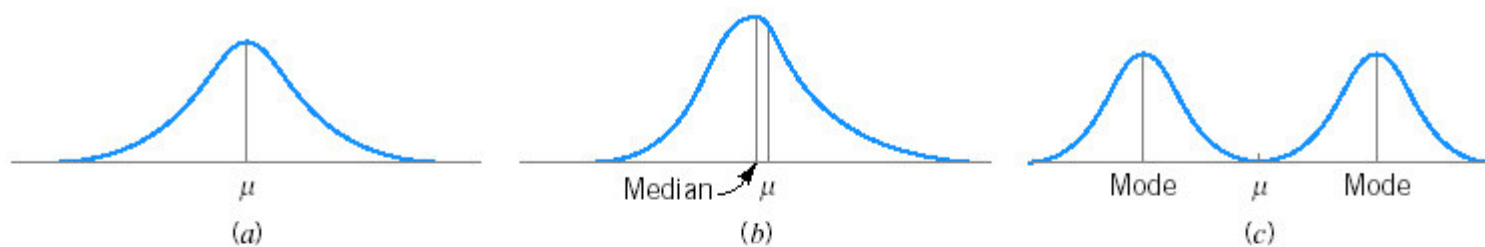


Figure 2-11 The mean of a distribution.

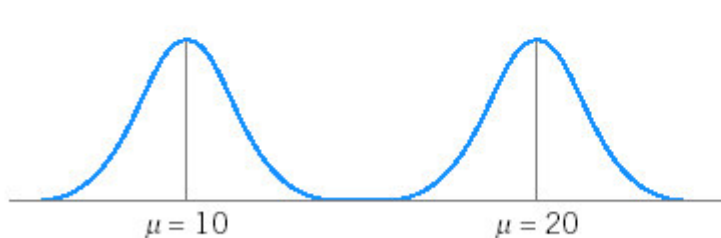


Figure 2-12 Two probability distributions with different means.

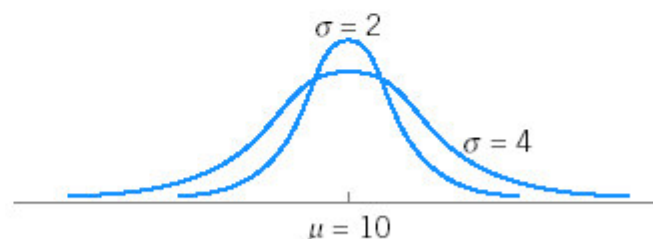


Figure 2-13 Two probability distributions with the same mean but different standard deviations.

The scatter, spread, or variability in a distribution is expressed by the **variance** σ^2 . The definition of the variance is

$$\sigma^2 = \begin{cases} \int_{-\infty}^{\infty} (x - \mu)^2 f(x) dx, & x \text{ continuous} & (2-6a) \\ \sum_{i=1}^{\infty} (x_i - \mu)^2 p(x_i), & x \text{ discrete} & (2-6b) \end{cases}$$

when the random variable is discrete with N equally likely values, then equation 2-6b becomes

$$\sigma^2 = \frac{\sum_{i=1}^N (x_i - \mu)^2}{N}$$

$$\sigma = \sqrt{\sigma^2} = \sqrt{\frac{\sum_{i=1}^N (x_i - \mu)^2}{N}} \quad (2-7)$$

The standard deviation is a measure of spread or scatter in the population expressed in the original units. Two distributions with the same mean but different standard deviations are shown in Fig. 2-13.

2-2 IMPORTANT DISCRETE DISTRIBUTIONS

The Hypergeometric Distribution

Definition

The **hypergeometric probability distribution** is

$$p(x) = \frac{\binom{D}{x} \binom{N-D}{n-x}}{\binom{N}{n}} \quad x = 0, 1, 2, \dots, \min(n, D) \quad (2-8)$$

The mean and variance of the distribution are

$$\mu = \frac{nD}{N} \quad (2-9)$$

and

$$\sigma^2 = \frac{nD}{N} \left(1 - \frac{D}{N}\right) \left(\frac{N-n}{N-1}\right) \quad (2-10)$$

The hypergeometric distribution is the appropriate probability model for selecting a random sample of n items without replacement from a lot of N items of which D are nonconforming or defective. By a random sample, we mean a sample that has been selected in such a way that all possible samples have an equal chance of being chosen. In these applications, x usually represents the number of nonconforming items found in the sample. For example, suppose that a lot contains 100 items, 5 of which do not conform to requirements. If 10 items are selected at random without replacement, then the probability of finding one or fewer nonconforming items in the sample is

$$\begin{aligned} P\{x \leq 1\} &= P\{x = 0\} + P\{x = 1\} \\ &= \frac{\binom{5}{0}\binom{95}{10}}{\binom{100}{10}} + \frac{\binom{5}{1}\binom{95}{9}}{\binom{100}{10}} = 0.923 \end{aligned}$$

In Chapter 14 we show how probability models such as this can be used to design acceptance-sampling procedures.

The Binomial Distribution

Basis is in **Bernoulli trials**

Definition

The **binomial distribution** with parameters $n \geq 0$ and $0 < p < 1$ is

$$p(x) = \binom{n}{x} p^x (1-p)^{n-x} \quad x = 0, 1, \dots, n \quad (2-11)$$

The mean and variance of the binomial distribution are

$$\mu = np \quad (2-12)$$

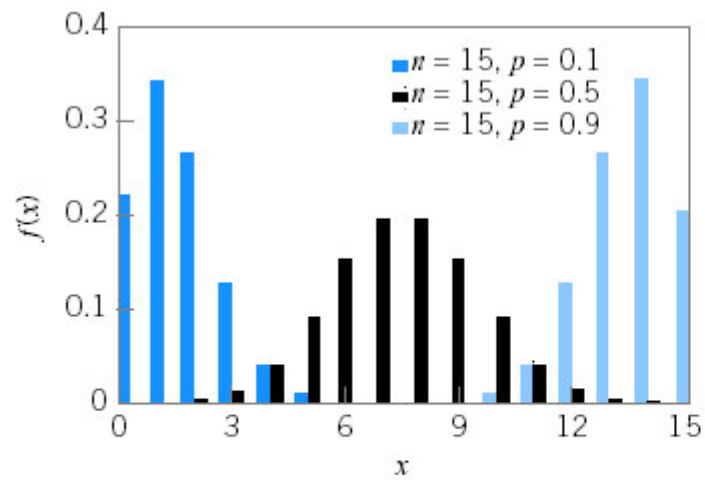
and

$$\sigma^2 = np(1-p) \quad (2-13)$$

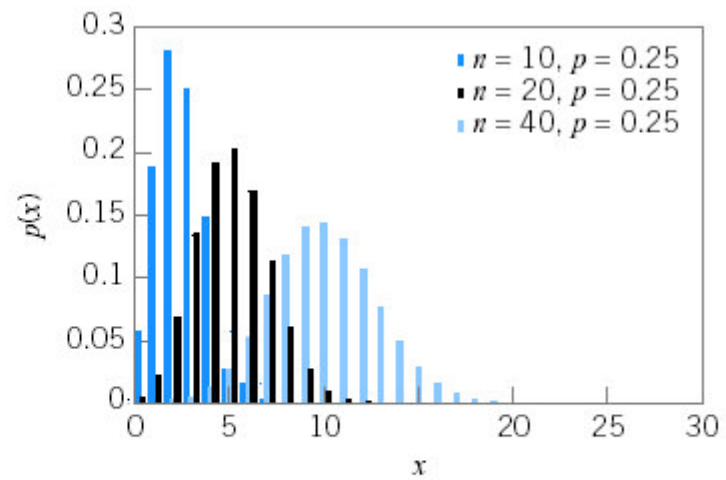
The random variable x is the number of successes out of n Bernoulli trials with constant probability of success p on each trial

The binomial distribution is used frequently in quality engineering. It is the appropriate probability model for sampling from an infinitely large population, where p represents the fraction of defective or nonconforming items in the population. In these applications, x usually represents the number of nonconforming items found in a random sample of n items. For example, if $p = 0.10$ and $n = 15$, then the probability of obtaining x nonconforming items is computed from equation 2-11 as

x	$p(x)$
0	0.2059
1	0.3432
2	0.2669
3	0.1285
4	0.0428
5	0.0105
6	0.0019
7	0.0003
8	0.0000
.	.
.	.
.	.
15	0.0000



(a) Binomial distributions for different values of p with $n = 15$.



(b) Binomial distributions for different values of n with $p = 0.25$.

Figure 2-14 Binomial distributions for selected values of n and p .

A random variable that arises frequently in statistical quality control is

$$\hat{p} = \frac{x}{n} \quad (2-14)$$

where x has a binomial distribution with parameters n and p . Often $\hat{p}$ is the ratio of the observed number of defective or nonconforming items in a sample (x) to the sample size (n) and this is usually called the **sample fraction defective** or **sample fraction nonconforming**. The “^” symbol is used to indicate that $\hat{p}$ is an estimate of the true, unknown value of the binomial parameter p . The probability distribution of $\hat{p}$ is obtained from the binomial, since

$$P\{\hat{p} \leq a\} = P\left\{\frac{x}{n} \leq a\right\} = P\{x \leq na\} = \sum_{x=0}^{[na]} \binom{n}{x} p^x (1-p)^{n-x}$$

where $[na]$ denotes the largest integer less than or equal to na . It is easy to show that the mean of $\hat{p}$ is p and that the variance of $\hat{p}$ is

$$\sigma_{\hat{p}}^2 = \frac{p(1-p)}{n}$$

The Poisson Distribution

Definition

The **Poisson distribution** is

$$p(x) = \frac{e^{-\lambda} \lambda^x}{x!} \quad x = 0, 1, \dots \quad (2-15)$$

where the parameter $\lambda > 0$. The **mean** and **variance** of the Poisson distribution are

$$\mu = \lambda \quad (2-16)$$

and

$$\sigma^2 = \lambda \quad (2-17)$$

Frequently used as a model for count data

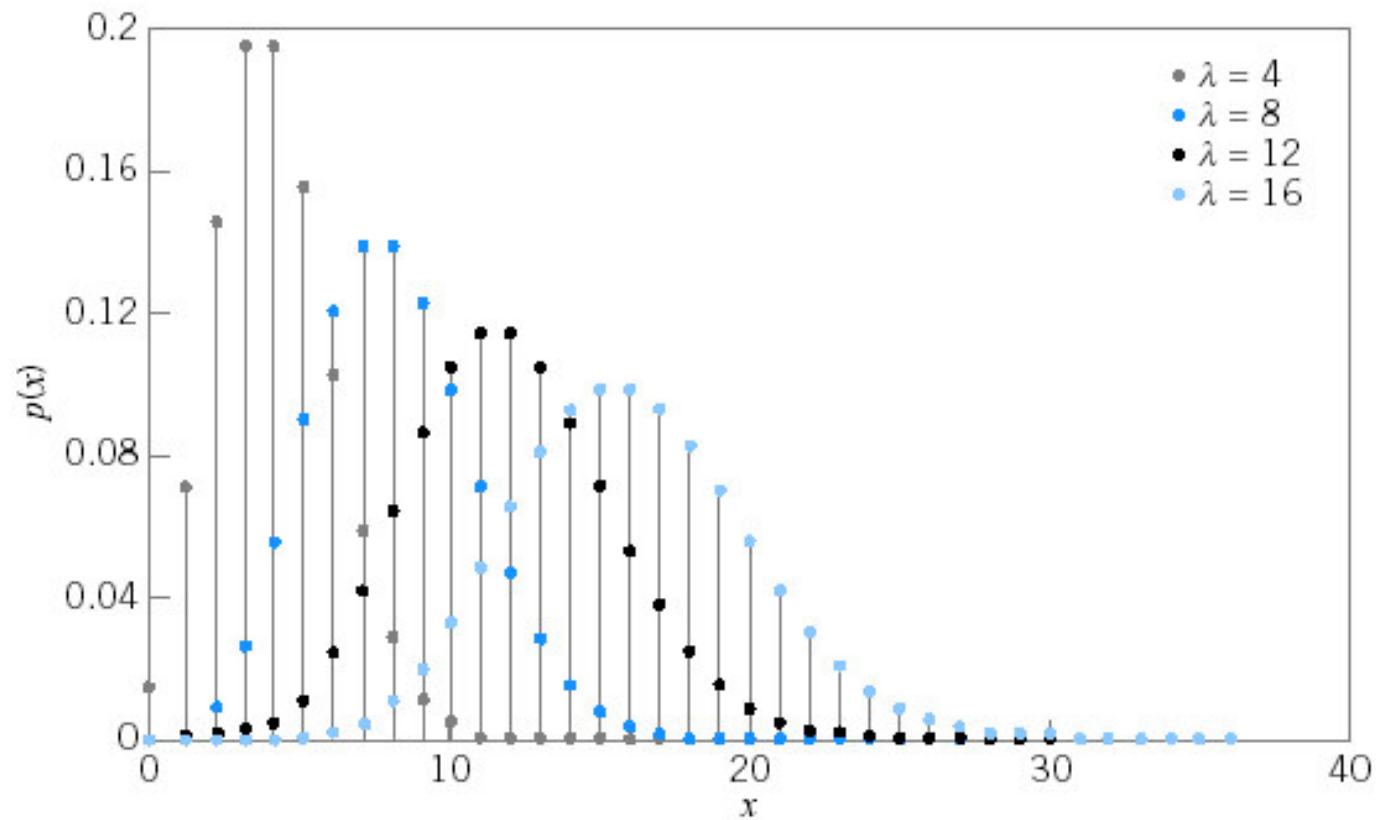


Figure 2-15 Poisson probability distributions for selected values of λ .

The Pascal Distribution

Definition

The **Pascal distribution** is

$$p(x) = \binom{x-1}{r-1} p^r (1-p)^{x-r} \quad x = r, r+1, r+2, \dots \quad (2-18)$$

where $r \geq 1$ is an integer. The **mean** and **variance** of the Pascal distribution are

$$\mu = \frac{r}{p} \quad (2-19)$$

and

$$\sigma^2 = \frac{r(1-p)}{p^2} \quad (2-20)$$

respectively.

The random variable x is the number of Bernoulli trials upon which the r th success occurs

- When $r = 1$ the Pascal distribution is known as the **geometric** distribution
- The geometric distribution has many useful applications in SQC

2-3 IMPORTANT CONTINUOUS DISTRIBUTIONS

The Normal Distribution

Definition

The **normal distribution** is

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad -\infty < x < \infty \quad (2-21)$$

The mean of the normal distribution is μ ($-\infty < \mu < \infty$) and the variance is $\sigma^2 > 0$.

The normal distribution is used so much that we frequently employ a special notation, $x \sim N(\mu, \sigma^2)$, to imply that x is normally distributed with mean μ and variance σ^2 . The visual appearance of the normal distribution is a symmetric, unimodal or **bell-shaped** curve and is shown in Fig. 2-16.

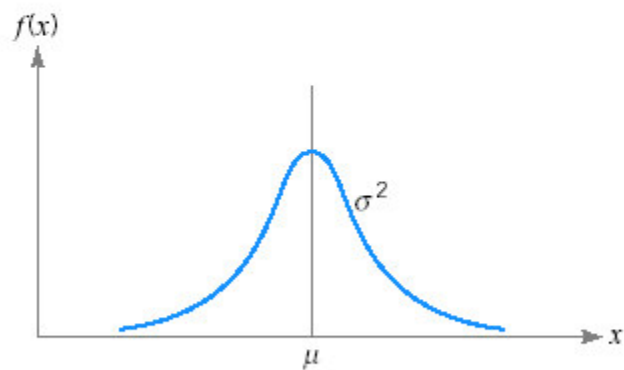


Figure 2-16 The normal distribution.

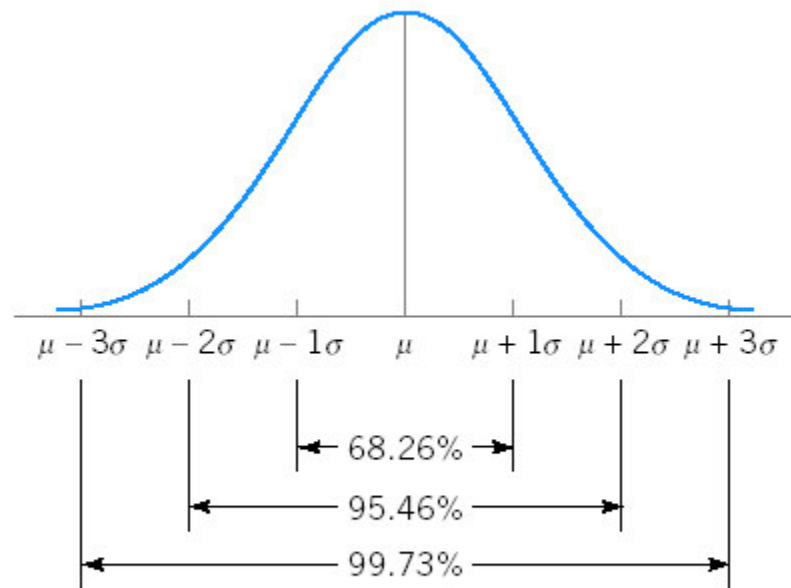


Figure 2-17 Areas under the normal distribution.

The cumulative normal distribution is defined as the probability that the normal random variable x is less than or equal to some value a , or

$$P\{x \leq a\} = F(a) = \int_{-\infty}^a \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx \quad (2-22)$$

This integral cannot be evaluated in closed form. However, by using the change of variable

$$z = \frac{x - \mu}{\sigma} \quad (2-23)$$

the evaluation can be made independent of μ and σ^2 . That is,

$$P\{x \leq a\} = P\left\{z \leq \frac{a - \mu}{\sigma}\right\} \equiv \Phi\left(\frac{a - \mu}{\sigma}\right) \quad (2-23)$$

where $\Phi(\cdot)$ is the cumulative distribution function of the **standard normal distribution** (mean = 0, standard deviation = 1). A table of the cumulative standard normal distribution is given in Appendix Table II. The transformation (2-23) is usually called **standardization**, because it converts a $N(\mu, \sigma^2)$ random variable into an $N(0, 1)$ random variable.

..... **EXAMPLE 2-7**

The tensile strength of paper used to make grocery bags is an important quality characteristic. It is known that the strength—say, x —is normally distributed with mean $\mu = 40 \text{ lb/in}^2$ and standard deviation $\sigma = 2 \text{ lb/in}^2$, denoted $x \sim N(40, 2^2)$. The purchaser of the bags requires them to have a strength of at least 35 lb/in^2 . The probability that a bag produced from this paper will meet or exceed this specification is $P\{x \geq 35\}$. Note that

$$P\{x \geq 35\} = 1 - P\{x \leq 35\}$$

To evaluate this probability from the standard normal tables, we standardize the point 35 and find

$$P\{x \leq 35\} = P\left\{z \leq \frac{35 - 40}{2}\right\} = P\{z \leq -2.5\} = \Phi(-2.5) = 0.0062$$

Consequently, the desired probability is

$$P\{x \geq 35\} = 1 - P\{x \leq 35\} = 1 - 0.0062 = 0.9938$$

Figure 2-18 shows the tabulated probability for both the $N(40, 2^2)$ distribution and the standard normal distribution. Note that the shaded area to the left of 35 lb/in^2 in Fig. 2-18 represents the fraction nonconforming or “fallout” produced by the bag manufacturing process.

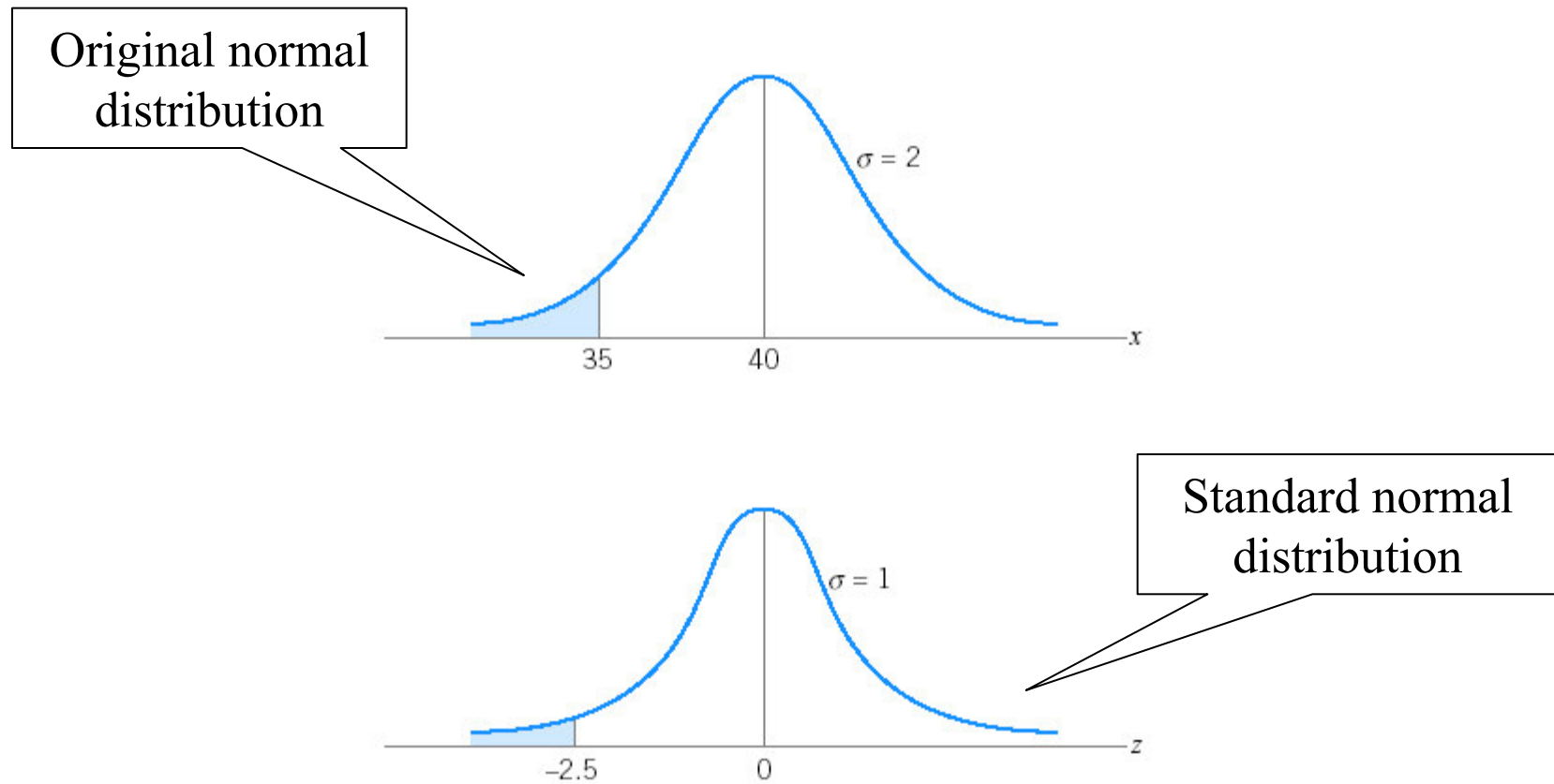


Figure 2-18 Calculation of $P\{x \leq 35\}$ in Example 2-7.

..... **EXAMPLE 2-9**

Sometimes instead of finding the probability associated with a particular value of a normal random variable, we find it necessary to do the opposite—find a particular value of a normal random variable that results in a given probability. For example, suppose that $x \sim N(10, 9)$. We wish to find the value of x —say, a —such that $P\{x > a\} = 0.05$. Thus

$$P\{x > a\} = P\left\{z > \frac{a-10}{3}\right\} = 0.05$$

or

$$P\left\{z \leq \frac{a-10}{3}\right\} = 0.95$$

From Appendix Table II, we have $P\{z \leq 1.645\} = 0.95$, so

$$\frac{a-10}{3} = 1.645$$

or

$$a = 10 + 3(1.645) = 14.935$$

.....

The normal distribution has many useful properties. One of these is relative to **linear combinations** of normally and independently distributed random variables. If $x_1, x_2, \dots, x_n$ are normally and independently distributed random variables with means $\mu_1, \mu_2, \dots, \mu_n$ and variances $\sigma_1^2, \sigma_2^2, \dots, \sigma_n^2$, respectively, then the distribution of

$$y = a_1x_1 + a_2x_2 + \dots + a_nx_n$$

is normal with mean

$$\mu_y = a_1\mu_1 + a_2\mu_2 + \dots + a_n\mu_n \quad (2-27)$$

and variance

$$\sigma_y^2 = a_1^2\sigma_1^2 + a_2^2\sigma_2^2 + \dots + a_n^2\sigma_n^2 \quad (2-28)$$

where $a_1, a_2, \dots, a_n$ are constants.

Definition: The Central Limit Theorem

If $x_1, x_2, \dots, x_n$ are independent random variables with mean μ_i and variance σ_i^2 , and if $y = x_1 + x_2 + \dots + x_n$, then the distribution of

$$\frac{y - \sum_{i=1}^n \mu_i}{\sqrt{\sum_{i=1}^n \sigma_i^2}}$$

approaches the $N(0, 1)$ distribution as n approaches infinity.

- Practical interpretation – the sum of independent random variables is approximately normally distributed regardless of the distribution of each individual random variable in the sum

The Lognormal Distribution

Definition

Let w have a normal distribution mean θ and variance ω^2 ; then $x = \exp(w)$ is a **lognormal random variable**, and the lognormal distribution is

$$f(x) = \frac{1}{x\omega\sqrt{2\pi}} \exp\left[-\frac{(\ln(x) - \theta)^2}{2\omega^2}\right] \quad 0 < x < \infty \quad (2-29)$$

The mean and variance of x are

$$\mu = e^{\theta + \omega^2/2} \quad \text{and} \quad \sigma^2 = e^{2\theta + \omega^2} (e^{\omega^2} - 1) \quad (2-30)$$

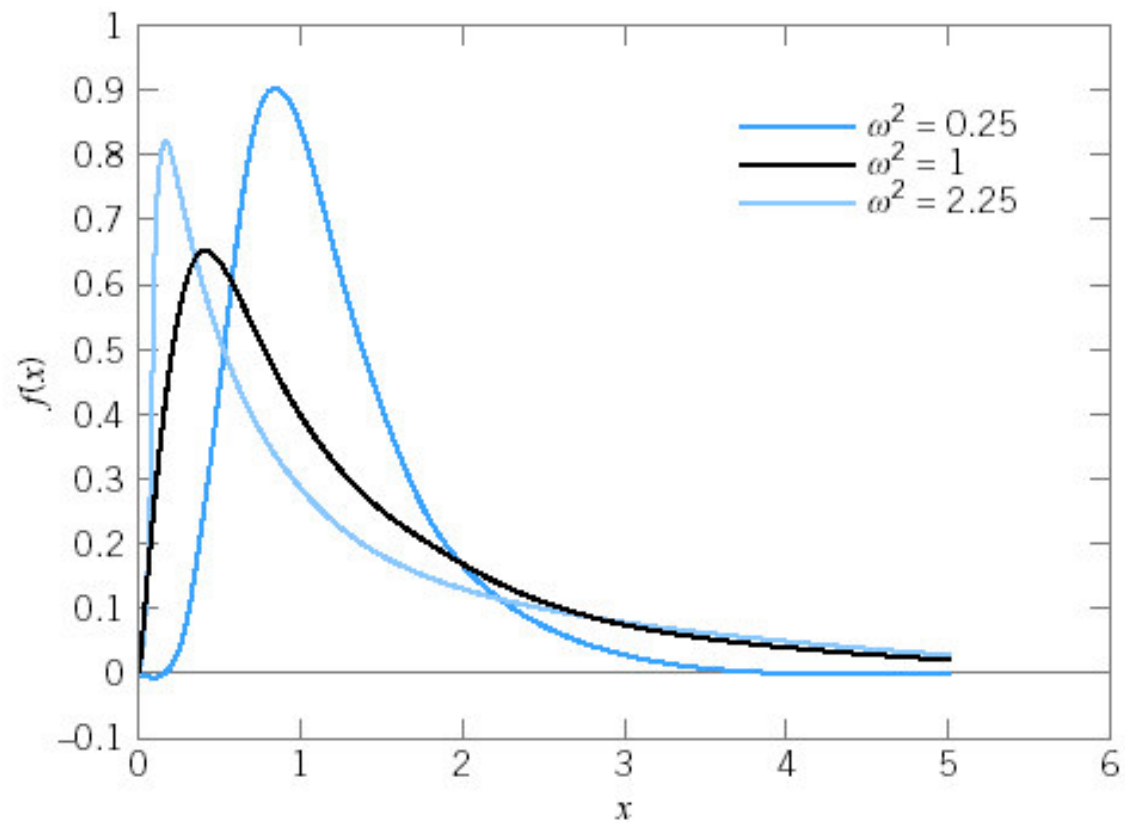


Figure 2-20 Lognormal probability density functions with $\theta = 0$ for selected values of ω^2 .

..... **EXAMPLE 2-10**

The lifetime of a medical laser used in ophthalmic surgery has a lognormal distribution with $\theta = 6$ and $\omega = 1.2$ hours. What is the probability the lifetime exceeds 500 hours?

From the cumulative distribution function for the lognormal random variable

$$\begin{aligned} P(x > 500) &= 1 - P[\exp(w) \leq 500] = 1 - P[w \leq \ln(500)] \\ &= \Phi\left(\frac{\ln(500) - 6}{1.2}\right) = 1 - \Phi(0.1788) = 1 - 0.5710 = 0.4290 \end{aligned}$$

What lifetime is exceeded by 99% of lasers? Now the question is to determine a such that $P(x > a) = 0.99$. Therefore,

$$P(x > a) = P[\exp(w) > a] = P[w > \ln(a)] = 1 - \Phi\left(\frac{\ln(a) - 6}{1.2}\right) = 0.99$$

From Appendix Table II, $1 - \Phi(a) = 0.99$ when $a = -2.33$. Therefore,

$$\frac{\ln(a) - 6}{1.2} = -2.33 \quad \text{and} \quad a = \exp(3.204) = 24.63 \text{ hours}$$

Determine the mean and standard deviation of lifetime. Now,

$$\mu = e^{\theta + \omega^2/2} = \exp(6 + 0.72) = 828.82 \text{ hours}$$

$$\sigma^2 = e^{2\theta + \omega^2} (e^{\omega^2} - 1) = \exp(12 + 1.44) [\exp(1.44) - 1] = 2,212,419.85$$

so the standard deviation of the lifetime is 1487.42 hours. Notice that the standard deviation of the lifetime is large relative to the mean.

The Exponential Distribution

Definition

The **exponential distribution** is

$$f(x) = \lambda e^{-\lambda x} \quad x \geq 0 \quad (2-31)$$

where $\lambda > 0$ is a constant. The **mean** and **variance** of the exponential distribution are

$$\mu = \frac{1}{\lambda} \quad (2-32)$$

and

$$\sigma^2 = \frac{1}{\lambda^2} \quad (2-33)$$

respectively.

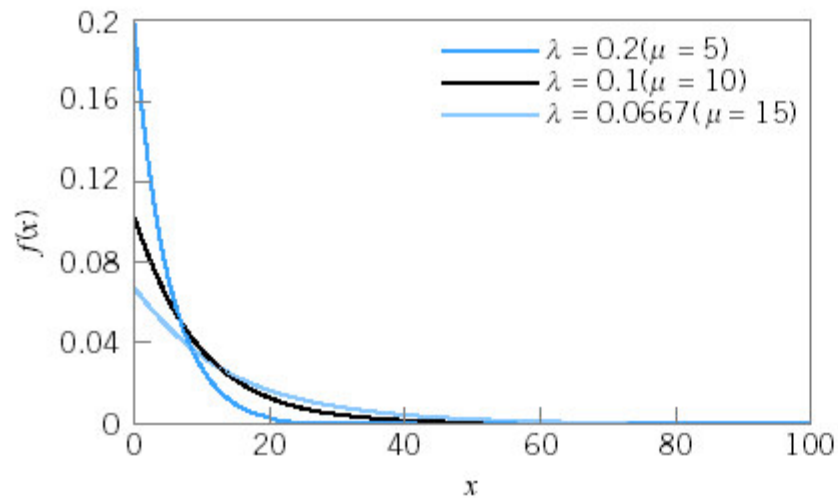


Figure 2-21 Exponential distributions for selected values of λ .

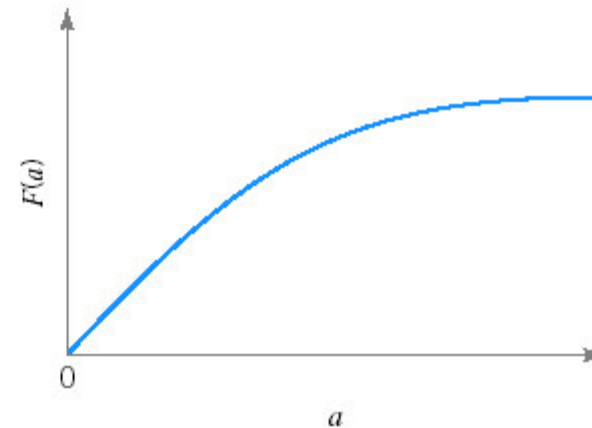


Figure 2-22 The cumulative exponential distribution function.

Relationship between the Poisson and exponential distributions

The Gamma Distribution

Definition

The **gamma distribution** is

$$f(x) = \frac{\lambda}{\Gamma(r)} (\lambda x)^{r-1} e^{-\lambda x} \quad x \geq 0 \quad (2-36)$$

with **shape parameter** $r > 0$ and **scale parameter** $\lambda > 0$. The **mean** and **variance** of the gamma distribution are

$$\mu = \frac{r}{\lambda} \quad (2-37)$$

and

$$\sigma^2 = \frac{r}{\lambda^2} \quad (2-38)$$

respectively.³

³ $\Gamma(r)$ in the denominator of equation 2-36 is the gamma function, defined as $\Gamma(r) = \int_0^{\infty} x^{r-1} e^{-x} dx, r > 0$. If r is a positive integer, then $\Gamma(r) = (r - 1)!$

- When r is an integer, the gamma distribution is the result of summing r independently and identically exponential random variables each with parameter λ

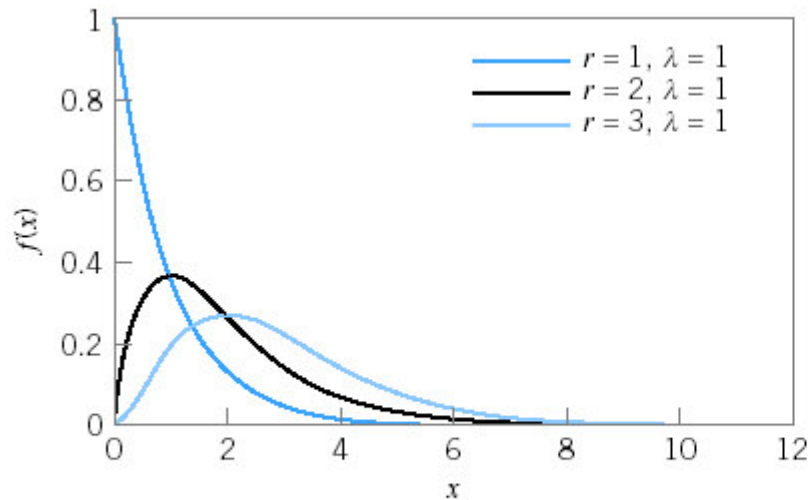


Figure 2-23 Gamma distributions for selected values of r and $\lambda = 1$.

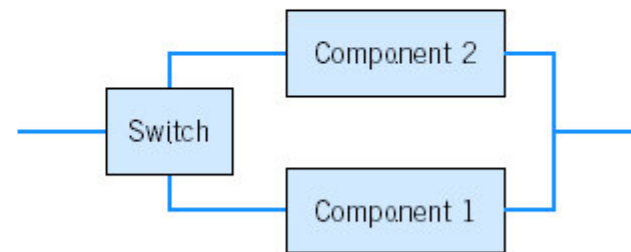


Figure 2-24 The standby redundant system for Example 2-11.

- The gamma distribution has many applications in reliability engineering; see Example 2-121, text page 71

The Weibull Distribution

Definition

The **Weibull distribution** is

$$f(x) = \frac{\beta}{\theta} \left(\frac{x}{\theta} \right)^{\beta-1} \exp \left[- \left(\frac{x}{\theta} \right)^{\beta} \right] \quad x \geq 0 \quad (2-41)$$

where $\theta > 0$ is the **scale parameter**, and $\beta > 0$ is the **shape parameter**. The **mean** and **variance** of the Weibull distribution are

$$\mu = \theta \Gamma \left(1 + \frac{1}{\beta} \right) \quad (2-42)$$

and

$$\sigma^2 = \theta^2 \left[\Gamma \left(1 + \frac{2}{\beta} \right) - \left\{ \Gamma \left(1 + \frac{1}{\beta} \right) \right\}^2 \right] \quad (2-43)$$

respectively.

- When $\beta = 1$, the Weibull distribution reduces to the exponential distribution

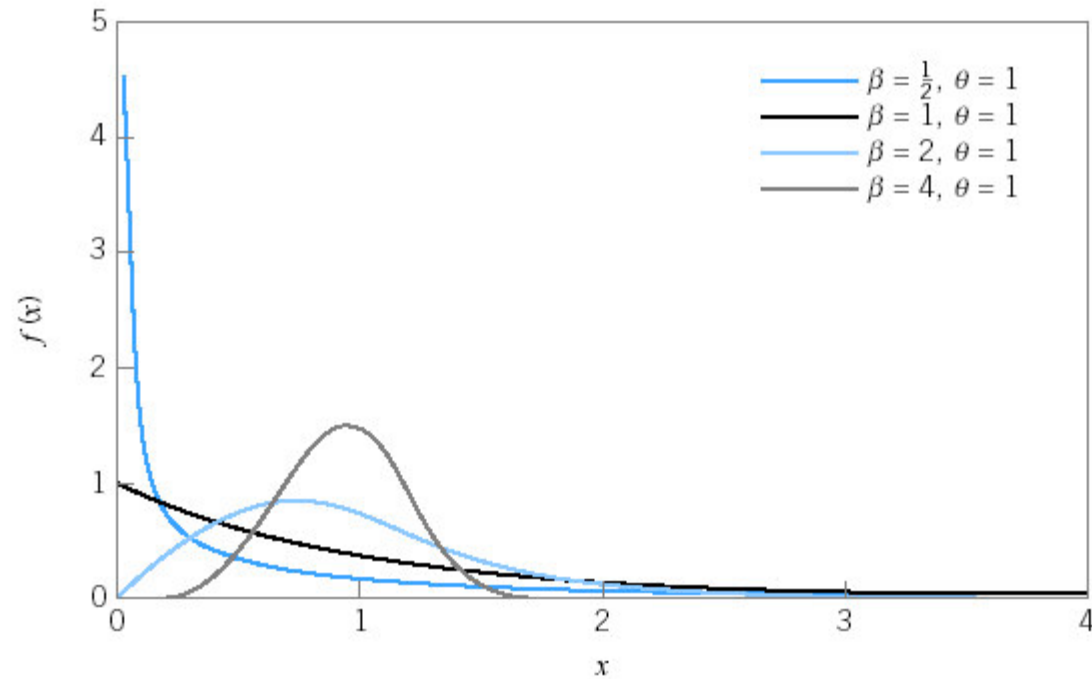


Figure 2-25 Weibull distributions for selected values of the shape parameter β and scale parameter $\theta = 1$.

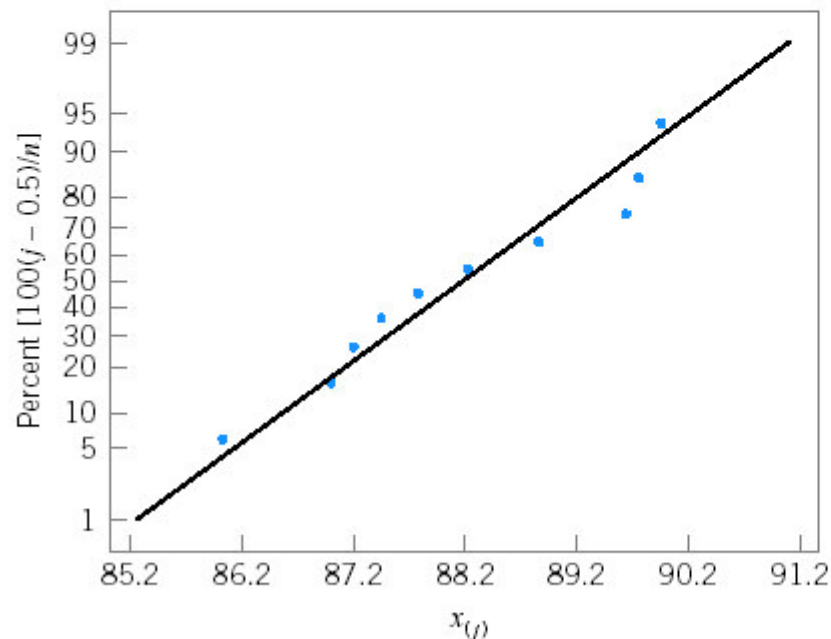
2-4 PROBABILITY PLOTS

- Determining if a sample of data might reasonably be assumed to come from a specific distribution
- Probability plots are available for various distributions
- Easy to construct with computer software (MINITAB)
- Subjective interpretation

Normal Probability Plot

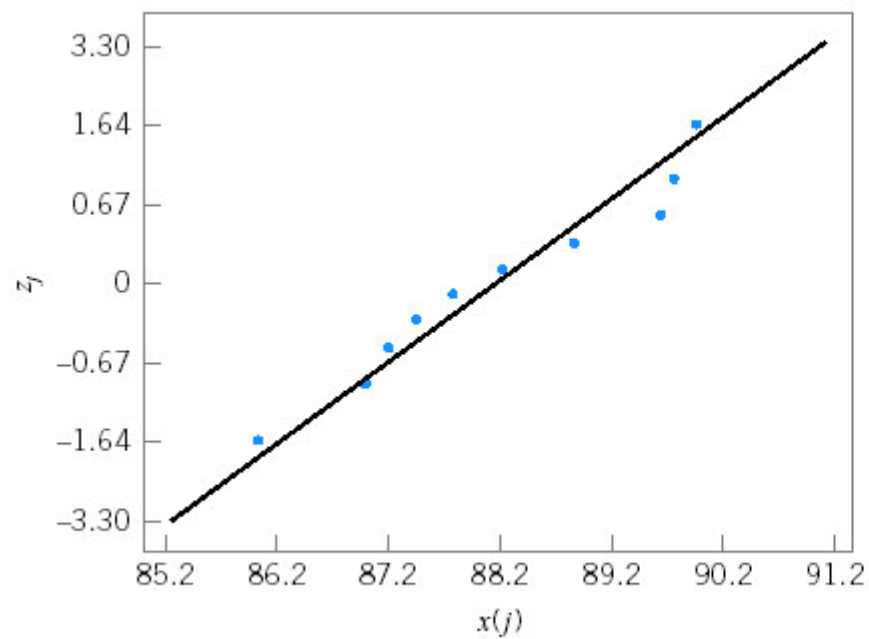
..... EXAMPLE 2-13

Observations on the road octane number of ten gasoline blends are as follows: 88.9, 87.0, 90.0, 88.2, 87.2, 87.4, 87.8, 89.7, 86.0, and 89.6. We hypothesize that octane number is adequately modeled by a normal distribution. To use probability plotting to investigate this hypothesis, first arrange the observations in ascending order and calculate their cumulative frequencies $(j - 0.5)/10$ as shown in the accompanying table.



j	$x_{(j)}$	$(j - 0.5)/10$
1	86.0	0.05
2	87.0	0.15
3	87.2	0.25
4	87.4	0.35
5	87.8	0.45
6	88.2	0.55
7	88.9	0.65
8	89.6	0.75
9	89.7	0.85
10	90.0	0.95

Figure 2-26 Normal probability plot of the road octane number data.



j	$x_{(j)}$	$(j - 0.5)/10$	z_j
1	86.0	0.05	-1.64
2	87.0	0.15	-1.04
3	87.2	0.25	-0.67
4	87.4	0.35	-0.39
5	87.8	0.45	-0.13
6	88.2	0.55	0.13
7	88.9	0.65	0.39
8	89.6	0.75	0.67
9	89.7	0.85	1.04
10	90.0	0.95	1.64

Figure 2-27 Normal probability plot of the road octane number data with standardized scores.

Other Probability Plots

- What is a reasonable choice as a probability model for these data?

Table 2-5 Aluminum Contamination (ppm)

30	30	60	63	70	79	87
90	101	102	115	118	119	119
120	125	140	145	172	182	
183	191	222	244	291	511	

From “The Lognormal Distribution for Modeling Quality Data When the Mean Is Near Zero,” *Journal of Quality Technology*, 1990, pp. 105–110.

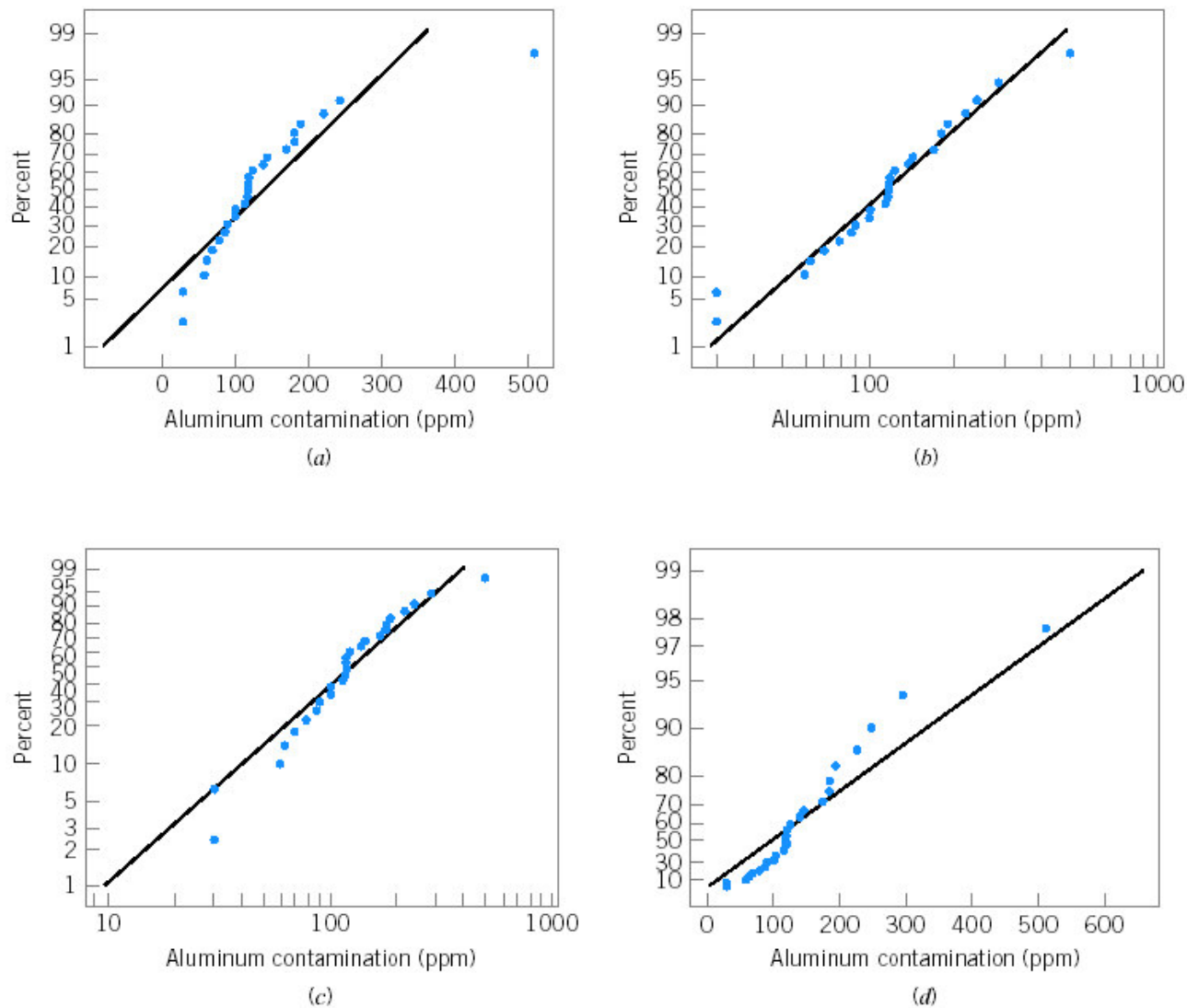


Figure 2-28 Probability plots of the aluminum contamination data in Table 2-5. (a) Normal. (b) Lognormal. (c) Weibull. (d) Exponential.

2-5 SOME USEFUL APPROXIMATIONS

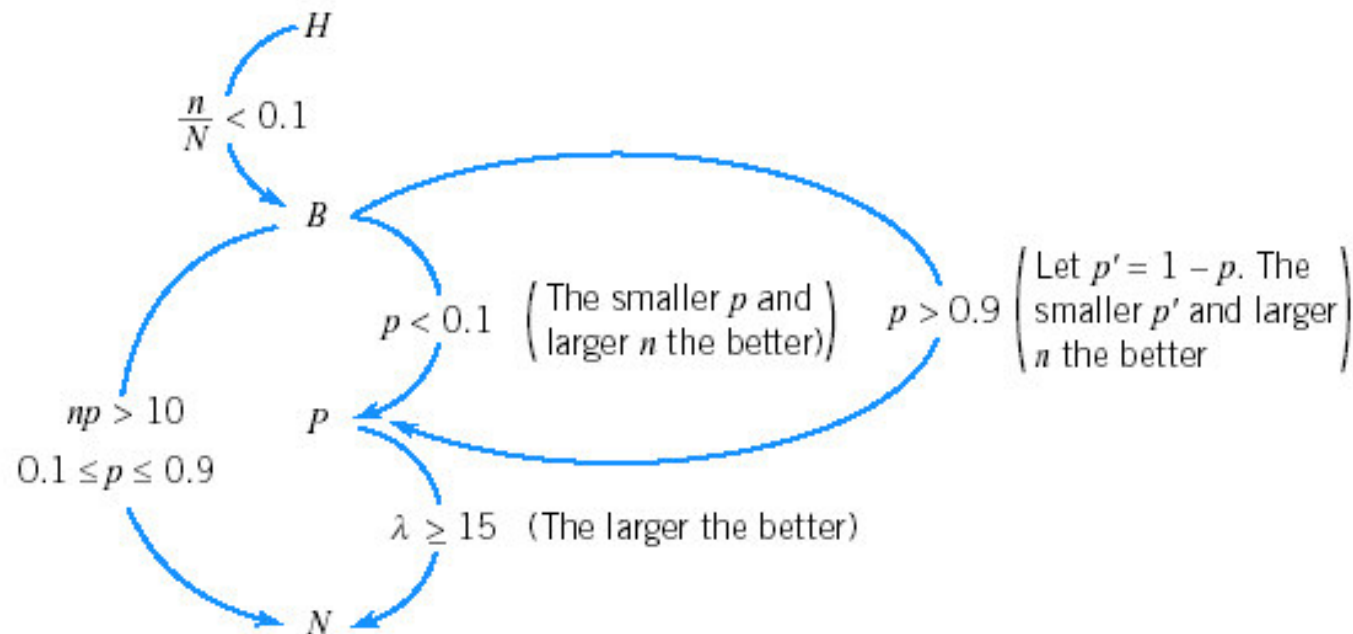


Figure 2-29 Approximations to probability distributions

IMPORTANT TERMS AND CONCEPTS

Approximations to probability distributions	Percentile
Binomial distribution	Poisson distribution
Box plot	Population
Continuous distribution	Probability distribution
Control limit theorem	Probability plotting
Descriptive statistics	Quartile
Discrete distribution	Random variable
Exponential distribution	Run chart
Gamma distribution	Sample
Geometric distribution	Sample average
Histogram	Sample standard deviation
Hypergeometric distribution	Sample variance
Interquartile range	Standard deviation of a distribution
Lognormal distribution	Standard normal distribution
Mean of a distribution	Statistics
Median	Stem-and-leaf display
Negative binomial distribution	Time series plot
Normal distribution	Uniform distribution
Normal probability plot	Variance of a distribution
Pascal distribution	Weibull distribution

Learning Objectives

1. Construct and interpret visual data displays, including the stem-and-leaf plot, the histogram, and the box plot
2. Compute and interpret the sample mean, the sample variance, the sample standard deviation, and the sample range
3. Explain the concepts of a random variable and a probability distribution
4. Understand and interpret the mean, variance, and standard deviation of a probability distribution
5. Determine probabilities from probability distributions
6. Understand the assumptions for each of the discrete probability distributions presented
7. Understand the assumptions for each of the continuous probability distributions presented
8. Select an appropriate probability distribution for use in specific applications
9. Use probability plots
10. Use approximations for some hypergeometric and binomial distributions