STATISTICS FOR ECONOMISTS: A BEGINNING

John E. Floyd
University of Toronto

July 2, 2010

PREFACE

The pages that follow contain the material presented in my introductory quantitative methods in economics class at the University of Toronto. They are designed to be used along with any reasonable statistics textbook. The most recent textbook for the course was James T. McClave, P. George Benson and Terry Sincich, *Statistics for Business and Economics*, Eighth Edition, Prentice Hall, 2001. The material draws upon earlier editions of that book as well as upon John Neter, William Wasserman and G. A. Whitmore, *Applied Statistics*, Fourth Edition, Allyn and Bacon, 1993, which was used previously and is now out of print. It is also consistent with Gerald Keller and Brian Warrack, *Statistics for Management and Economics*, Fifth Edition, Duxbury, 2000, which is the textbook used recently on the St. George Campus of the University of Toronto. The problems at the ends of the chapters are questions from mid-term and final exams at both the St. George and Mississauga campuses of the University of Toronto. They were set by Gordon Anderson, Lee Bailey, Greg Jump, Victor Yu and others including myself.

This manuscript should be useful for economics and business students enrolled in basic courses in statistics and, as well, for people who have studied statistics some time ago and need a review of what they are supposed to have learned. Indeed, one could learn statistics from scratch using this material alone, although those trying to do so may find the presentation somewhat compact, requiring slow and careful reading and thought as one goes along.

I would like to thank the above mentioned colleagues and, in addition, Adonis Yatchew, for helpful discussions over the years, and John Maheu for helping me clarify a number of points. I would especially like to thank Gordon Anderson, who I have bothered so frequently with questions that he deserves the status of mentor.

After the original version of this manuscript was completed, I received some detailed comments on Chapter 8 from Peter Westfall of Texas Tech University, enabling me to correct a number of errors. Such comments are much appreciated.

J. E. Floyd
July 2, 2010

Chapter 3

Some Common Probability Distributions

3.1 Random Variables

Most of the basic outcomes we have considered thus far have been non-numerical characteristics. A coin comes up either heads or tails; a delivery is on the same day with the correct order, the next day with the incorrect order, etc. We now explicitly consider random trials or experiments that relate to a quantitative characteristic, with a numerical value associated with each outcome. For example, patients admitted to a hospital for, say, X days where $X = 1, 2, 3, 4, \ldots$. Canada's GNP this year will be a specific number on the scale of numbers ranging upwards from zero. When the outcomes of an experiment are particular values on a natural numerical scale we refer to these values as a *random variable.* More specifically, a random variable is a variable whose numerical value is determined by the outcome of a random trial or experiment where a unique numerical value is assigned to each sample point.

Random variables may be *discrete* as in the length of hospital stay in days or *continuous* as in the case of next month's consumer price index or tomorrow's Dow Jones Industrial Average, the calculated values of which, though rounded to discrete units for reporting, fall along a continuum. The essential distinction between discrete and continuous random variables is that the sample points can be enumerated (or listed in quantitative order) in the case of a discrete random variable—for example, we can list the number of potential days of a hospital stay.[1] In the case of continuous random variables it

[1] Hospital stays could also be treated as a continuous variable if measured in fractions

is not possible to list the sample points in quantitative order—next month's consumer price index, for example, could be 120.38947 or 120.38948 or it could take any one of an infinity of values between 120.38947 and 120.38948. The number of sample points for a continuous random variable is always infinite. For a discrete random variable the number of sample points may or may not be infinite, but even an infinity of sample points could be listed or enumerated in quantitative order although it would take an infinite length of time to list them all. In the case of a continuous random variable any sample points we might put in a list cannot possibly be next to each other—between any two points we might choose there will be an infinity of additional points.

3.2 Probability Distributions of Random Variables

The *probability distribution* for a discrete random variable X associates with each of the distinct outcomes x_i, $(i = 1, 2, 3, \ldots, k)$ a probability $P(X = x_i)$. It is also called the *probability mass function* or the *probability function.* The probability distribution for the hospital stay example is shown in the top panel of Figure 3.1. The *cumulative probability distribution* or *cumulative probability function* for a discrete random variable X provides the probability that X will be at or below any given value—that is, $P(X \leq x_i)$ for all x_i. This is shown in the bottom panel of Figure 3.1. Note that X takes discrete values in both panels so that the lengths of the bars in the top panel give the probabilities that it will take the discrete values associated with those bars. In the bottom panel the length of each bar equals the sum of the lengths of all the bars in the top panel associated with values of X equal to or less than the value of X for that bar.

A continuous random variable assumes values on a continuum. Since there are an infinity of values between any two points on a continuum it is not meaningful to associate a probability value with a point on that continuum. Instead, we associate probability values with intervals on the continuum. The *probability density function* of a continuous random variable X is a mathematical function for which the area under the curve corresponding to any interval is equal to the probability that X will take on a value in that interval. The probability density function is denoted by $f(x)$, which gives the probability density at x. An example is given in the top panel of Figure 3.2 with the shaded area being the probability that X will take a value between 6 and 7. Note that $f(x)$ is always positive.

of hours or days. They are normally measure discretely in days, however, with patients being in hospital 'for the day' if not released during a given period in the morning.

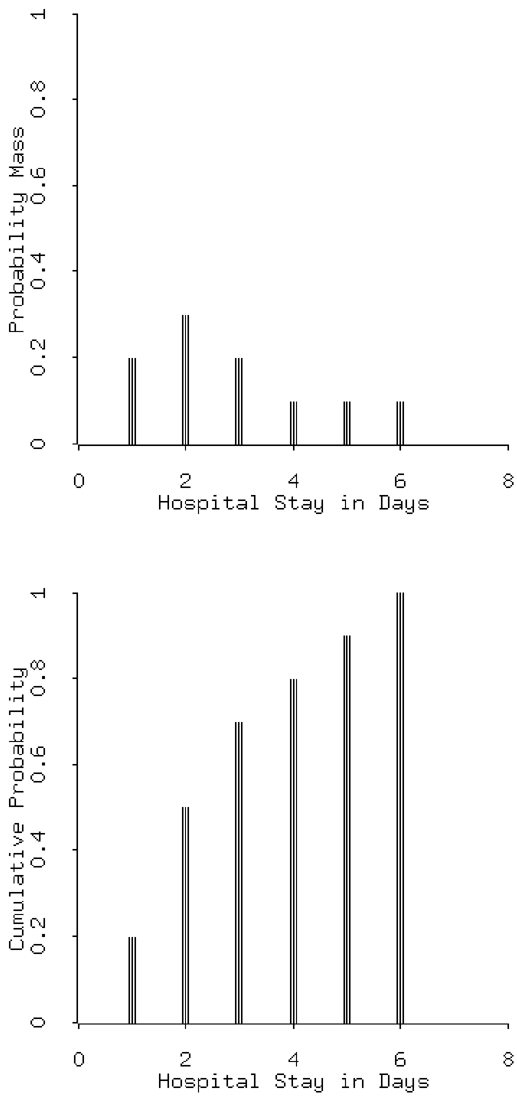


Figure 3.1: Probability mass function (top) and cumulative probability function (bottom) for the discrete random variable 'number of days of hospitalization'.

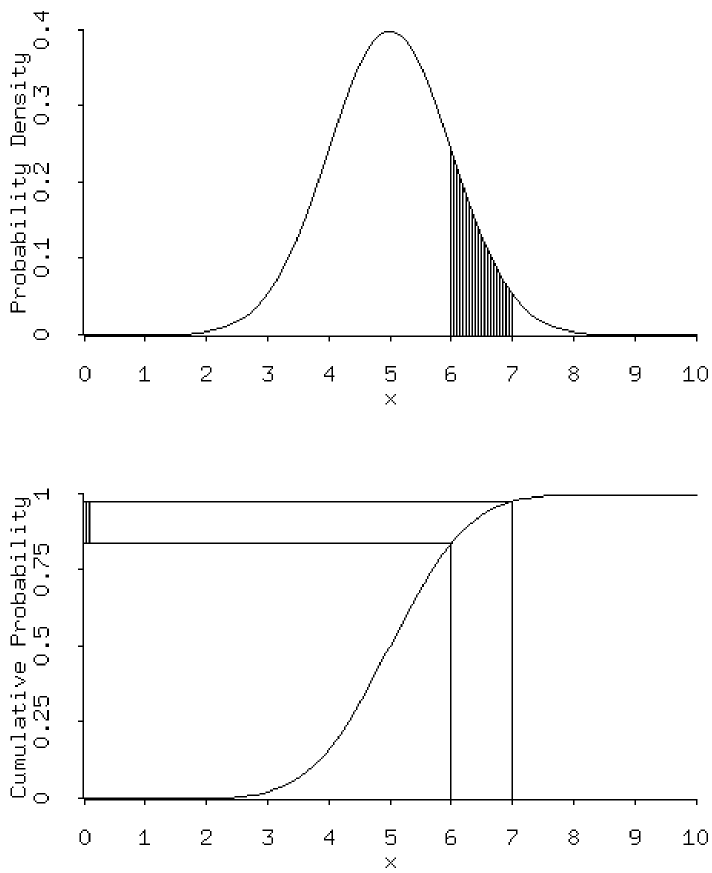


Figure 3.2: Probability density and cumulative probability functions for a continuouts random variable. The shaded area in the top panel equals the distance between the two vertical lines in the bottom panel.

The *cumulative probability function* of a continuous random variable X is denoted by $F(x)$ and is defined

$$F(x) = P(X \leq x) \tag{3.1}$$

where $-\infty \leq x \leq +\infty$. The cumulative probability function $F(x)$ gives the probability that the outcome of X in a random trial will be less than or equal to any specified value x. Thus, $F(x)$ corresponds to the area under the probability density function to the left of x. This is shown in the bottom panel of Figure 3.2. In that panel, the distance between the two horizontal lines associated with the cumulative probabilities at $X \leq 6$ and $X \leq 7$ is equal to the shaded area in the top panel, and the distance of the lower of those two horizontal lines from the horizontal axis is equal to the area under the curve in the top panel to the left of $X = 6$. In mathematical terms we can express the probability function as

$$P(a \leq x \leq b) \quad = \quad \int_a^b f(x)\,dx \tag{3.2}$$

and the cumulative probability function as

$$P(X \leq x) \quad = \quad F(x) \quad = \quad \int_{-\infty}^{x} f(u)\,du \tag{3.3}$$

where u represents the variable of integration.

3.3 Expected Value and Variance

The mean value of a random variable in many trials is also known as its expected value. The *expected value of a discrete random variable* X is denoted by $E\{X\}$ and defined

$$E\{X\} \quad = \quad \sum_{i=1}^{k} x_i P(x_i) \tag{3.4}$$

where $P(x_i) = P(X = x_i)$. Since the process of obtaining the expected value involves the calculation denoted by $E\{\}$ above, $E\{\,\}$ is called the *expectation operator.*

Suppose that the probability distribution in the hospital stay example in Figure 3.1 above is

x:	1	2	3	4	5	6
$P(x)$:	.2	.3	.2	.1	.1	.1

The expected value of X is

$$\begin{aligned} E\{X\} &= (1)(.2) + (2)(.3) + (3)(.2) + (4)(.1) + (5)(.1) + (6)(.1) \\ &= .2 + .6 + .6 + .4 + .5 + .6 \quad = \quad 2.9. \end{aligned}$$

Note that this result is the same as would result from taking the mean in the fashion outlined in Chapter 1. Let the probabilities be frequencies where the total hospital visits is, say, 100. Then the total number of person-days spent in the hospital is

$$\begin{aligned}&(1)(20) + (2)(30) + (3)(20) + (4)(10) + (5)(10) + (6)(10)\\ = \;& 20 + 60 + 60 + 40 + 50 + 60 \;=\; 290\end{aligned}$$

and the common mean is $290/100 = 2.9$. $E\{X\}$ is simply a weighted average of the possible outcomes with the probability values as weights. For this reason it is called the mean of the probability distribution of X. Note that the mean or expected value is a number that does not correspond to any particular outcome.

The *variance* of a discrete random variable X is denoted by $\sigma^2\{X\}$ and defined as

$$\sigma^2\{X\} \;=\; \sum_{i=1}^{k}(x_i - E\{X\})^2 P(x_i) \tag{3.5}$$

where $\sigma^2\{\,\}$ is called the *variance operator*. The calculation of the variance of the length of hospital stay can be organized in the table below:

x:	1	2	3	4	5	6
$P(x)$:	.20	.30	.20	.10	.10	.10
$x - E\{X\}$:	-1.90	-.90	.10	1.10	2.10	3.10
$(x - E\{X\})^2$:	3.61	.81	.01	1.21	4.41	9.61

from which

$$\begin{aligned}\sigma^2\{X\} &= (3.61)(.2) + (.81)(.3) + (.01)(.2) + (1.21)(.1) + (4.41)(.1)\\ &\quad +(9.61)(.1)\\ &= .722 + .243 + .002 + .121 + .441 + .961 \;=\; 2.49.\end{aligned}$$

The variance is a weighted average of the squared deviations of the outcomes of X from their expected value where the weights are the respective probabilities of occurrence. Thus $\sigma^2\{X\}$ measures the extent to which the outcomes of X depart from their expected value in the same way that the variance of the quantitative variables in the data sets examined in Chapter 1 measured the variability of the values about their mean. There is an important distinction, however, between what we are doing here and what we did in Chapter 1. In Chapter 1 we took an observed variable X and measured its observed variance. Here we are taking a *random variable* X and exploring the nature of its probability distribution.

Consider a random variable V for which $v_i = (x_i - E\{X\})^2$ in (3.5). Since each v_i has a corresponding x_i associated with it,

$$P(v_i) = P((x_i - E\{X\})^2) = P(x_i),$$

and (3.5) yields

$$\begin{aligned}\sigma^2\{X\} &= \sum_{i=1}^{k} v_i P(v_i) \\ &= E\{V\} \;=\; E\{(x_i - E\{X\})^2\}. \qquad (3.6)\end{aligned}$$

The variance is simply the expectation of, or expected value of, the squared deviations of the values from their mean. The *standard deviation*, denoted by σ, is defined as the square root of the variance.

The discrete random variable X can be *standardised* or put in *standardised form* by applying the relationship

$$Z_i = \frac{X_i - E\{X\}}{\sigma\{X\}} \qquad (3.7)$$

where the discrete random variable Z is the standardised form of the variable X. The variable Z is simply the variable X expressed in numbers of standard deviations from its mean. In the hospital stay example above the standardised values of the numbers of days of hospitalization are calculated as follows:

x:	1	2	3	4	5	6
$P(x)$:	.2	.3	.2	.1	.1	.1
$x - E\{X\}$:	-1.9	-.9	.1	1.1	2.1	3.1
$(x - E\{X\})^2$:	3.61	.81	.01	1.21	4.41	9.61
$(x - E\{X\})/\sigma\{X\}$:	-1.20	-.56	.06	.70	1.32	1.96

where $\sigma = \sqrt{2.49} = 1.58$.

The *expected value of a continuous random variable* is defined as

$$E\{X\} = \int_{-\infty}^{\infty} x f(x)\, dx. \qquad (3.8)$$

This is not as different from the definition of the expected value of a discrete random variable in (3.4) as it might appear. The integral performs the same role for a continuous variable as the summation does for a discrete one. Equation (3.8) sums from minus infinity to plus infinity the variable x with each little increment of x, given by dx, weighted by the probability $f(x)$ that the outcome of x will fall within that increment.[2] Similarly, the *variance of a continuous random variable* is defined as

$$\begin{aligned} \sigma^2\{X\} &= E\{(x - E\{X\})^2\} \\ &= \int_{-\infty}^{\infty} (x - E\{X\})^2 f(x)\, dx. \end{aligned} \tag{3.9}$$

In this equation the integral is taken over the probability weighted increments to $(x - E\{X\}^2)$ as compared to (3.8) where the integration is over the probability weighted increments to x.

Continuous random variables can be standardised in the same fashion as discrete random variables. The standardised form of the continuous random variable X is thus

$$Z = \frac{X - E\{X\}}{\sigma\{X\}}. \tag{3.10}$$

3.4 Covariance and Correlation

We noted in Chapter 1 that covariance and correlation are measures of the association between two variables. The variables in that case were simply quantitative data. Here we turn to an analysis of the covariance and correlation of two *random variables* as properties of their joint probability distribution. The *covariation* of the outcomes x_i and y_j of the discrete random variables X and Y is defined as

$$(x_i - E\{X\})(y_j - E\{Y\}).$$

The *covariance* of two random variables is the expected value of their covariation (i.e., their mean covariation after repeated trials). For two discrete random variables X and Y we thus have

$$\begin{aligned} \sigma\{X, Y\} &= E\{(x_i - E\{X\})(y_j - E\{Y\})\} \\ &= \sum_i \sum_j (x_i - E\{X\})(y_j - E\{Y\}) P(x_i, y_j) \end{aligned} \tag{3.11}$$

where $P(x_i, y_j)$ denotes $P(X = x_i \; \cap \; Y = y_j)$. We call $\sigma\{\,,\,\}$ the *covariance operator*. Consider the following example:

[2]Notice that the definition of probability requires that

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

	Y		
X	5	10	
2	.1	.4	.5
3	.3	.2	.5
	.4	.6	1.0

The two discrete random variables X and Y each take two values, 2 and 3 and 5 and 10 respectively. The four numbers in the enclosed square give the *joint probability distribution* of X and Y —that is, the probabilities

$$P(X = x_i \ \cap \ Y = y_j).$$

The numbers along the right and bottom margins are the marginal probabilities, which sum in each case to unity. On the basis of the earlier discussion it follows that

$$E\{X\} = (2)(.5) + (3)(.5) = 2.5$$

$$E\{Y\} = (5)(.4) + (10)(.6) = 8.0$$

$$\sigma^2\{X\} = (-.5^2)(.5) + (.5^2)(.5) = .25$$

and

$$\sigma^2\{Y\} = (-3^2)(.4) + (2^2)(.6) = 6.0$$

which renders $\sigma\{X\} = \sqrt{0.25} = .5$ and $\sigma\{Y\} = \sqrt{8} = 2.83$. The calculation of the covariance can be organized using the following table:

	$(X = 2 \cap Y = 5)$	$(X = 2 \cap Y = 10)$	$(X = 3 \cap Y = 5)$	$(X = 3 \cap Y = 10)$
$P(x_i, y_j)$	.1	.4	.3	.2
$(x_i - E\{X\})$	- .5	-.5	.5	.5
$(y_j - E\{Y\})$	- 3	2	- 3	2
$(x_i - E\{X\})(y_j - E\{Y\})$	1.5	-1	-1.5	1
$(x_i - E\{X\})(y_j - E\{Y\})P(x_i, y_j)$	.15	-.4	-.45	.2

The sum of the numbers in the bottom row gives

$$\sigma\{X, Y\} = \sum_i \sum_j (x_i - E\{X\})(y_j - E\{Y\})P(x_i, y_j) = -.5.$$

The *coefficient of correlation* of two random variables X and Y, denoted by $\rho\{X,Y\}$ is defined as

$$\rho\{X,Y\} = \frac{\sigma\{X,Y\}}{\sigma\{X\}\sigma\{Y\}}. \tag{3.12}$$

In the example above

$$\rho\{X,Y\} = -.5/((.5)(2.83)) = -.5/1.415 = -.35$$

which signifies a negative relationship between the two random variables. It is easy to show that the coefficient of correlation between X and Y is equivalent to the covariance between the standardised forms of those variables because the covariance of the standardised forms is the same as the covariance of the unstandardised variables and the standard deviations of the standardised forms are both unity. Thus, when the variables are standardised both the covariance and the correlation coefficient are unit free.

The *covariance of continuous random variables* X and Y is written

$$\begin{aligned}\sigma\{X,Y\} &= E\{(x - E\{X\})(y - E\{Y\})\} \\ &= \int\int (x - E\{X\})(y - E\{Y\}) f(x,y)\, dx\, dy \qquad (3.13)\end{aligned}$$

where $f(x,y)$ is the *joint probability density function* of X and Y. The shape of a typical joint probability density function is portrayed graphically in Figure 3.3 (both variables are in standardised form). The *coefficient of correlation between continuous random variables* is defined by equation (3.12) with the numerator being (3.13) and the denominator the product of the standard deviations of X and Y obtained by taking the square roots of successive applications of (3.9).

When two variables are statistically independent both the covariance and correlation between them is zero. The opposite, however, does not follow. Zero covariance and correlation do not necessarily imply statistical independence because there may be a non-linear statistical relationship between two variables. An example is shown in Figure 3.4. The covariance and correlation between the two variables is zero, but they are obviously systematically related.

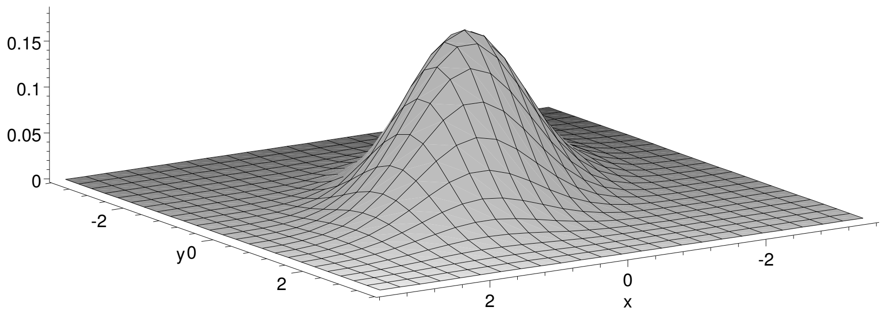


Figure 3.3: The joint probability density function of two continuous standardized random variables.

3.5 Linear Functions of Random Variables

Consider a linear function of the random variable X,

$$W \quad = \quad a + b\,X. \tag{3.14}$$

A number of relationships hold. First,

$$E\{W\} \quad = \quad E\{a + b\,X\} \quad = \quad E\{a\} + b\,E\{X\}, \tag{3.15}$$

which implies that

$$E\{a\} \quad = \quad a \tag{3.16}$$

and

$$E\{b\,X\} \quad = \quad b\,E\{X\}. \tag{3.17}$$

We can pass the expectation operator through a linear equation with the result that $E\{W\}$ is the same function of $E\{X\}$ as W is of X. Second,

$$\sigma^2\{W\} \quad = \quad \sigma^2\{a + b\,X\} \quad = \quad b^2\,\sigma^2\{X\} \tag{3.18}$$

which implies

$$\sigma^2\{a + X\} \quad = \quad \sigma^2\{X\}, \tag{3.19}$$

and

$$\sigma^2\{b\,X\} \quad = \quad b^2\,\sigma^2\{X\}. \tag{3.20}$$

This leads to the further result that

$$\sigma\{a + b\,X\} \quad = \quad |b|\,\sigma\{X\}. \tag{3.21}$$

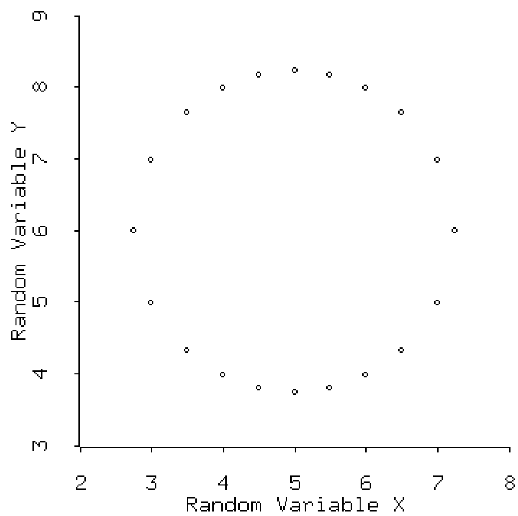


Figure 3.4: An example of two uncorrelated random variables that are not statistically independent.

3.6 Sums and Differences of Random Variables

If Z is the sum of two random variables X and Y, then the following two conditions hold:

$$E\{Z\} = E\{X+Y\} = E\{X\}+E\{Y\} \tag{3.22}$$

and

$$\sigma^2\{Z\} = \sigma^2\{X\}+\sigma^2\{Y\}+2\,\sigma\{X,Y\}. \tag{3.23}$$

When Z is the difference between X and Y, these become

$$E\{Z\} = E\{X-Y\} = E\{X\}-E\{Y\} \tag{3.24}$$

and

$$\sigma^2\{Z\} = \sigma^2\{X\}+\sigma^2\{Y\}-2\,\sigma\{X,Y\}. \tag{3.25}$$

To prove (3.23) and (3.25) we expand $\sigma^2\{Z\}$ using the definition of variance and the rules above:

$$\begin{aligned}
\sigma^2\{Z\} &= E\{(Z - E\{Z\})^2\} \\
&= E\{(X + Y - E\{X + Y\})^2\} \\
&= E\{((X - E\{X\}) + (Y - E\{Y\}))^2\} \\
&= E\{((X - E\{X\})^2 + 2\,(X - E\{X\})(Y - E\{Y\}) \\
&\qquad + (Y - E\{Y\})^2)\} \\
&= E\{(X - E\{X\})^2\} + 2\,E\{(X - E\{X\})(Y - E\{Y\})\} \\
&\qquad + E\{(Y - E\{Y\})^2\} \\
&= \sigma^2\{X\} + 2\,\sigma\{X, Y\} + \sigma^2\{Y\}. \qquad (3.26)
\end{aligned}$$

In the case where $Z = X - Y$ the sign of the covariance term changes but the variance of both terms remains positive because squaring a negative number yields a positive number.

When X and Y are statistically independent (and thus uncorrelated), $\sigma\{X, Y\} = 0$ and (3.23) and (3.25) both become

$$\sigma^2\{Z\} = \sigma^2\{X\} + \sigma^2\{Y\}.$$

More generally, if T is the sum of S *independent* random variables,

$$T = X_1 + X_2 + X_3 + \cdots + X_S,$$

where the X_i can take positive or negative values, then

$$E\{T\} = \sum_s^S E\{X_i\} \qquad (3.27)$$

and

$$\sigma^2\{T\} = \sum_s^S \sigma^2\{X_i\}. \qquad (3.28)$$

In concluding this section we can use the rules above to prove that the mean of a standardised variable is zero and its variance and standard deviation are unity. Let Z be the standardised value of X, that is

$$Z = \frac{X - E\{X\}}{\sigma\{X\}}.$$

Then

$$\begin{aligned} E\{Z\} &= E\left\{\frac{X - E\{X\}}{\sigma\{X\}}\right\} \\ &= \frac{1}{\sigma\{X\}} E\{X - E\{X\}\} \\ &= \frac{1}{\sigma\{X\}}\left(E\{X\} - E\{X\}\right) = 0 \end{aligned}$$

and

$$\begin{aligned} \sigma^2\{Z\} &= E\left\{\left(\frac{X - E\{X\}}{\sigma\{X\}} - 0\right)^2\right\} \\ &= E\left\{\left(\frac{X - E\{X\}}{\sigma\{X\}}\right)^2\right\} \\ &= \frac{1}{\sigma^2\{X\}} E\left\{(X - E\{X\})^2\right\} \\ &= \frac{\sigma^2\{X\}}{\sigma^2\{X\}} = 1. \end{aligned}$$

It immediately follows that $\sigma\{Z\}$ also is unity.

Finally, the correlation coefficient between two standardised random variables U and V will equal

$$\rho\{U,V\} = \frac{\sigma\{U,V\}}{\sigma\{U\}\,\sigma\{V\}} = \sigma\{U,V\}$$

since $\sigma\{U\}$ and $\sigma\{V\}$ are both unity.

3.7 Binomial Probability Distributions

We can think of many examples of random trials or experiments in which there are two basic outcomes of a qualitative nature—the coin comes up either heads or tails, the part coming off the assembly line is either defective or not defective, it either rains today or it doesn't, and so forth. These experiments are called *Bernoulli random trials.* To quantify these outcomes we arbitrarily assign one outcome the value 0 and the other the value 1. This random variable, $X_i = \{0, 1\}$ is called a *Bernoulli random variable.*

Usually we are interested in a whole sequence of random trials. In the process of checking the effectiveness of a process of manufacturing computer monitors, for example, we can let $X_i = 1$ if the ith monitor off the line is defective and $X_i = 0$ if the ith monitor is not defective. The X_i, $(i = 1, 2, 3, \ldots, n)$ can then be viewed as a sequence of Bernoulli random variables. Such a sequence is called a *Bernoulli process*. Let $X_1, X_2, X_3, \ldots, X_n$ be a sequence of random variables associated with a Bernoulli process. The process is said to be *independent* if the X_i are statistically independent and *stationary* if every $X_i = \{0\,, 1\}$ has the same probability distribution. The first of these conditions means that whether or not, say, the 5th monitor off the assembly line is defective will have nothing to do with whether the 6th, 7th, 100th, 200th, or any other monitor is defective. The second condition means that the probability of, say, the 10th monitor off the line being defective is exactly the same as the probability that any other monitor will be defective—the X_i are *identically distributed*. The random variables in the sequence are thus *independently and identically distributed*.

In a sequence of Bernoulli random trials we are typically interested in the number of trials that have the outcome 1. The sum $X_1 + X_2 + X_3 + \ldots + X_{300}$ would give the number of defective monitors in a sample of 300 off the line. The sum of n independent and identically distributed Bernoulli random variables, denoted by X,

$$X = X_1 + X_2 + X_3 + \ldots + X_n,$$

is called a *binomial random variable*. It can take $n + 1$ values ranging from zero (when $X_i = 0$ for all i) to n (when $X_i = 1$ for all i). This random variable is distributed according to the *binomial probability distribution*.

The *binomial probability function*, which gives the probabilities that X will take values $(0, \ldots, n)$, is

$$P(x) \;=\; \begin{pmatrix} n \\ x \end{pmatrix} p^x (1-p)^{n-x} \tag{3.29}$$

where $P(x) = P(X = x)$, $x = 0, 1, 2, \ldots, n$, and $0 \leq p \leq 1$. The parameter p is the probability that $X_i = 1$. It is the same for all i because the Bernoulli random variables X_i are identically distributed. The term

$$\begin{pmatrix} n \\ x \end{pmatrix}$$

represents a *binomial coefficient* which is defined as

$$\begin{pmatrix} n \\ x \end{pmatrix} \;=\; \frac{n!}{x!(n-x)!}$$

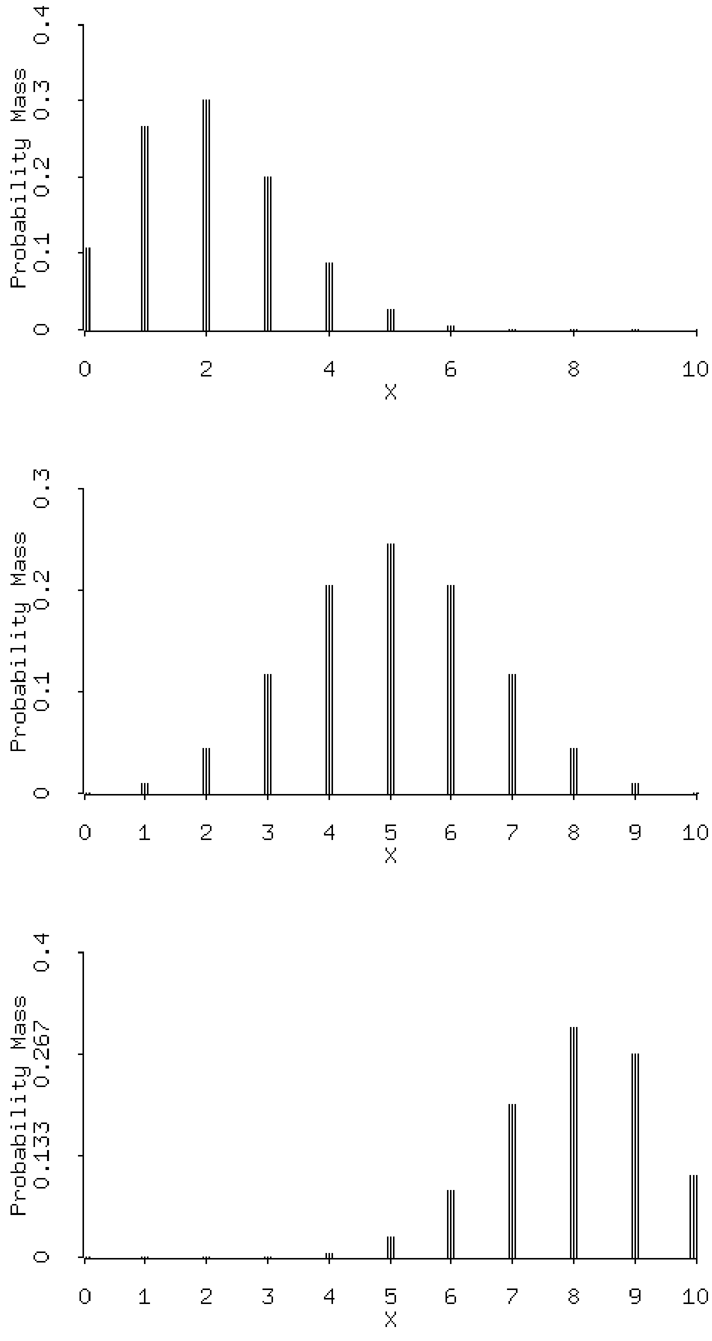


Figure 3.5: Binomial probability distributions with $n = 10$ and $p = .2$ (top), $p = .5$ (middle) and $p = .8$ (bottom.

where $a! = (a)(a-1)(a-2)(a-3)\ldots(1)$ and $0! = 1$.

The binomial probability distribution is a discrete probability distribution since X can only take the discrete values $0, 1, \ldots, n$. The parameters in the binomial probability distribution are p and n. Accordingly, there is a whole family of such distributions, one for each (p, n) combination. Figure 3.5 plots three examples—the distribution is skewed right if $p < .5$, skewed left if $p > .5$ and symmetrical if $p = .5$. The *mean of the binomial distribution* is

$$E\{X\} = np \tag{3.30}$$

and the variance is

$$\sigma^2\{X\} = np\,(1-p). \tag{3.31}$$

If we have two independent binomial random variables V and W with common probability parameter p and based on n_v and n_w trials, the sum $V+W$ is a binomial random variable with parameters p and $n = n_v + n_w$.

To more fully understand the workings of the binomial distribution consider the following problem. Four gauges are tested for accuracy. This involves four Bernoulli random trials $X_i = \{0, 1\}$ where 0 signifies that the gauge is accurate and 1 signifies that it is inaccurate. Whether or not any one of the four gauges is inaccurate has nothing to do with the accuracy of the remaining three so the X_i are statistically independent. The probability that each gauge is inaccurate is assumed to be .25. We thus have a binomial random variable X with $n = 4$ and $p = .25$. The sample space of X is

$$S = \{0, 1, 2, 3, 4\}.$$

Taking into account the fact that $n! = (4)(3)(2)(1) = 24$, the probability distribution can be calculated by applying equation (3.29) as follows:

x	$n!/(x!(n-x)!)$	p^x	$(1-p)^{n-x}$	$P(x)$
0	$24/(0!4!) = 1$	$.25^0 = 1.0000$	$.75^4 = .3164$	.3164
1	$24/(1!3!) = 4$	$.25^1 = .2500$	$.75^3 = .4219$	.4219
2	$24/(2!2!) = 6$	$.25^2 = .0625$	$.75^2 = .5625$	.2109
3	$24/(3!3!) = 4$	$.25^3 = .0156$	$.75^1 = .7500$	.0469
4	$24/(4!1!) = 1$	$.25^4 = .0039$	$.75^0 = 1.0000$	.0039
				1.0000

This probability distribution can be derived in a longer but more informative way by looking at the elementary events in the sample space and building up the probabilities from them. The four gauges are tested one after the other. There are 16 basic outcomes or sequences with probabilities attached to each sequence. The sequences are shown in Table 3.1. To see how the probabilities are attached to each sequence, consider sequence S_{12}. It consists of four outcomes of four independent and identically distributed Bernoulli random trials—0,1,0,0. The probability that 0 will occur on any trial is .75 and the probability that 1 will occur is .25. The probability of the four outcomes in the sequence observed is the product of the four probabilities. That is, the probability that first a 0 and then a 1 will occur is the probability of getting a 0 on the first trial times the probability of getting a 1 on the next trial. To obtain the probability that a sequence of 0,1,0 will occur we multiply the previously obtained figure by the probability of getting a zero. Then to get the probability of the sequence 0,1,0,0 we again multiply the previous figure by the probability of getting a zero. Thus the probability of the sequence S_{12} is

$$(.75)(.25)(.75)(.75) = (.25)^1(.75)^3 = .4219$$

which, it turns out, is the same as the probability of obtaining sequences S_8, S_{14} and S_{15}. Clearly, all sequences involving three zeros and a single one have the same probability regardless of the order in which the zeros and the one occur.

A frequency distribution of these sequences is presented in Table 3.2. There is one occurrence of no ones and four zeros, four occurrences of one and three zeros, six occurrences of two ones and two zeros, four occurrences of three ones and one zero, and one occurrence of four ones and no zeros. Thus, to find the probability that two ones and two zeros will occur we want the probability that any of the six sequences having that collection of ones and zeros will occur. That will be the probability of the union of the six elementary events, which will be the sum of the probabilities of the six sequences. Since all six sequences have the same probability of occurring the probability of two ones and two zeros is six times the probability associated with a single sequence containing two ones and two zeros.

Notice something else. Expand the expression $(x+y)^4$.

$$\begin{aligned}(x+y)^4 &= (x+y)(x+y)(x+y)(x+y)\\ &= (x+y)(x+y)(x^2+2\,x\,y+y^2)\\ &= (x+y)(x^3+3\,x^2\,y+3\,x\,y^2+y^3)\\ &= x^4+4\,x^3\,y+6\,x^2\,y^2+4\,x\,y^3+y^4.\end{aligned}$$

Table 3.1: Sequence of Outcomes in an Accuracy Test of Four Guages

X_1	X_2	X_3	X_4	Sequence	$X = \sum x_i$	Probability
1	1	1	1	S_1	4	$(.25)^4(.75)^0$
			0	S_2	3	$(.25)^3(.75)^1$
		0	1	S_3	3	$(.25)^3(.75)^1$
			0	S_4	2	$(.25)^2(.75)^2$
	0	1	1	S_5	3	$(.25)^3(.75)^1$
			0	S_6	2	$(.25)^2(.75)^2$
		0	1	S_7	2	$(.25)^2(.75)^2$
			0	S_8	1	$(.25)^1(.75)^3$
0	1	1	1	S_9	3	$(.25)^3(.75)^1$
			0	S_{10}	2	$(.25)^2(.75)^2$
		0	1	S_{11}	2	$(.25)^2(.75)^2$
			0	S_{12}	1	$(.25)^1(.75)^3$
	0	1	1	S_{13}	2	$(.25)^2(.75)^2$
			0	S_{14}	1	$(.25)^1(.75)^3$
		0	1	S_{15}	1	$(.25)^1(.75)^3$
			0	S_{16}	0	$(.25)^0(.75)^4$

Table 3.2: Frequency Distribution of Sequences in Table 3.1

x	Frequency	Probability of Sequence	$P(x)$
0	1	$(.25)^0(.75)^4$	$\times 1 = .3164$
1	4	$(.25)^1(.75)^3$	$\times 4 = .4219$
2	6	$(.25)^2(.75)^2$	$\times 6 = .2109$
3	4	$(.25)^3(.75)^1$	$\times 4 = .0469$
4	1	$(.25)^4(.75)^0$	$\times 1 = .0039$

It turns out that the coefficients of the four terms are exactly the frequencies of the occurrences in the frequency distribution above and the $x\,y$ terms become the sequence probabilities in the table when x is replaced by the probability of a one and y is replaced the probability of a zero and $n = 4$. The above expansion of $(x + y)^4$ is called the *binomial expansion*, whence the term binomial distribution. The easiest way to calculate the binomial coefficients for the simplest cases (where n is small) is through the use of *Pascal's Triangle.*

Pascal's Triangle

$$
\begin{array}{ccccccccccc}
 & & & & & 1 & & & & & \\
 & & & & 1 & & 1 & & & & \\
 & & & 1 & & 2 & & 1 & & & \\
 & & 1 & & 3 & & 3 & & 1 & & \\
 & 1 & & 4 & & 6 & & 4 & & 1 & \\
1 & & 5 & & 10 & & 10 & & 5 & & 1
\end{array}
$$

etc.. . . . . .

Additional rows can be added to the base by noting that each number that is not unity is the sum of the two numbers above it. The relevant binomial coefficients appear in the row with $n + 1$ entries.

Fortunately, all these complicated calculations need not be made every time we want to find a binomial probability. Probability tables have been calculated for all necessary values of n and p and appear at the end of every statistics textbook.[3]

3.8 Poisson Probability Distributions

The Poisson probability distribution applies to many random phenomena occurring during a period of time—for example, the number of inaccurate gauges coming off an assembly line in a day or week. It also applies to spatial phenomena such as, for example, the number of typographical errors on a page.

A *Poisson random variable* is a discrete variable that can take on any integer value from zero to infinity. The value gives the number of occurrences of the circumstance of interest during a particular period of time or within a particular spatial area. A unit probability mass is assigned to this sample space. Our concern is then with the probability that there will be, for example, zero, one, two, three, etc., calls to the police during a particular time period on a typical day, or that in typing this manuscript I will make zero, one, two, etc. errors on a particular page.

The *Poisson probability function* is

$$P(x) \quad = \quad \frac{\lambda^x \, e^{-\lambda}}{x!} \tag{3.32}$$

where

$$P(x) = P(X = x)$$

with

$$x = 0, 1, 2, 3, 4, \ldots, \infty$$

and $0 < \lambda < \infty$. The parameter $e = 2.71828$ is a constant equal to the base of natural logarithms.[4] Note that, whereas the binomial distribution had two parameters, n and p, the Poisson distribution has only one parameter, λ, which is the average number of calls over the period.

---

[3] These probabilities can also be calculated, and the various distributions plotted, using XlispStat and other statistical software.

[4] The number e is equal to

$$e = \lim_{n \to \infty} \left(1 + \frac{1}{n}\right)^n.$$

Consider an example. Suppose that the number of calls to the 911 emergency number between 8:00 and 8:30 PM on Fridays is a Poisson random variable X with $\lambda = 3.5$. We can calculate a portion of the probability distribution as follows:

x	$P(X = x)$					$P(X \leq x)$
0	$[3.5^0 e^{-3.5}]/0!$	=	$[(1)(.03019738)]/1$	=	.0302	.0302
1	$[3.5^1 e^{-3.5}]/1!$	=	$[(3.5)(.03019738)]/1$	=	.1057	.1359
2	$[3.5^2 e^{-3.5}]/2!$	=	$[(12/250)(.03019738)]/2$	=	.1850	.3208
3	$[3.5^3 e^{-3.5}]/3!$	=	$[(42.875)(.03019738)]/6$	=	.2158	.5366
4	$[3.5^4 e^{-3.5}]/4!$	=	$[(150.0625)(.03019738)]/24$	=	.1888	.7254
5	$[3.5^5 e^{-3.5}]/5!$	=	$[(525.2188)(.03019738)]/120$	=	.1322	.8576
6	$[3.5^6 e^{-3.5}]/6!$	=	$[(1838.266)(.03019738)]/720$	=	.0771	.9347
7	$[3.5^7 e^{-3.5}]/7!$	=	$[(6433.903)(.03019738)]/5040$	=	.0385	.9732

The figures in the right-most column are the cumulative probabilities. The probably of receiving 3 calls is slightly over .2 and the probability of receiving 3 or less calls is just under .54. Note that over 97 percent of the probability mass is already accounted for by $x \leq 7$ even though x ranges to infinity.

As in the case of the binomial distribution, it is unnecessary to calculate these probabilities by hand—Poisson tables can be found at the back of any textbook in statistics.[5] The mean and variance of a Poisson probability distribution are

$$E\{X\} = \lambda$$

and

$$\sigma^2\{X\} = \lambda.$$

Plots of Poisson distributions are shown in Figure 3.6. The top panel shows a Poisson distribution with $\lambda = .5$, the middle panel shows one with $\lambda = 3$

[5]And, as in the case of other distributions, probabilities can be calculated using statistical software such as XlispStat.

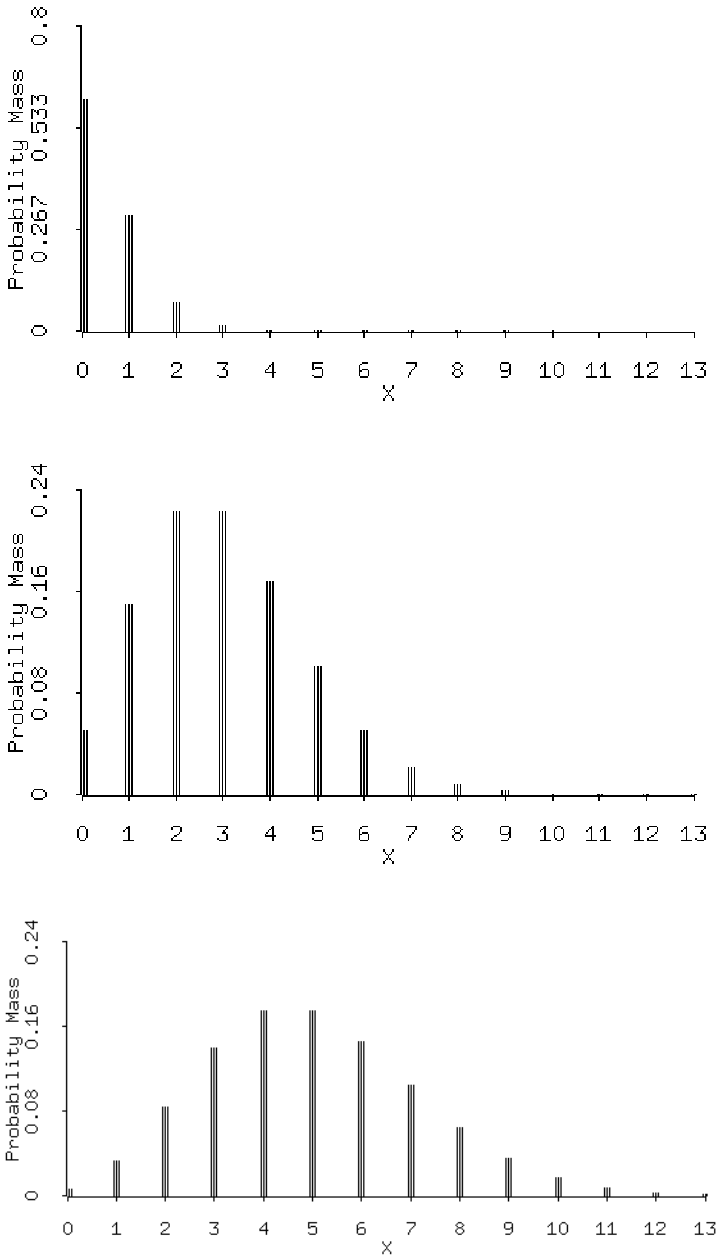


Figure 3.6: Poisson probability distributions with $\lambda = .5$ (top), $\lambda = 3$ (middle) and $\lambda = 5$ (bottom).

and the distribution plotted in the bottom panel has $\lambda = 5$. All Poisson probability distributions are skewed to the right although they become more symmetrical as λ gets larger.

Just as binomial distributions result from a Bernoulli process, Poisson distributions are the result of a *Poisson process*. A Poisson process is any process that generates occurrences randomly over a time or space continuum according to the following rules:

- The numbers of occurrences in non-overlapping time (space) intervals are statistically independent.
- The number of occurrences in a time (space) interval has the same probability distribution for all time (space) intervals.
- The probability of two or more occurrences in any interval $(t + \Delta t)$ is negligibly small relative to the probability of one occurrence in the interval.

When these postulates hold, the number of occurrences in a unit time (space) interval follows a Poisson probability distribution with parameter λ.

If V and W are two independent Poisson random variables with parameters λ_v and λ_w, respectively, the sum $V + W$ is a Poisson random variable with $\lambda = \lambda_v + \lambda_w$.

3.9 Uniform Probability Distributions

Uniform probability distributions result when the probability of all occurrences in the sample space are the same. These probability distributions may be either discrete or continuous.

Consider a computer random number generator that cranks out random numbers between 0 and 9. By construction of the computer program, the probability that any one of the 10 numbers will be turned up is 1/10 or 0.1. The probability distribution for this process is therefore

x:	0	1	2	3	4	5	6	7	8	9
$P(x)$:	.1	.1	.1	.1	.1	.1	.1	.1	.1	.1

This random variable is called a *discrete uniform random variable* and its probability distribution is a *discrete uniform probability distribution*. The discrete probability function is

$$P(x) = \frac{1}{s} \tag{3.33}$$

where

$$P(x) = P(X = x),$$
$$x = a, a+1, a+2, \ldots, a+(s-1).$$

The parameters a and s are integers with $s > 0$. Parameter a denotes the smallest outcome and parameter s denotes the number of distinct outcomes. In the above example, $a = 0$ and $s = 10$.

The mean and variance of a discrete uniform probability distribution are

$$E\{X\} = a + \frac{s-1}{2}$$

and

$$\sigma^2 = \frac{s^2-1}{12}.$$

In the example above, $E\{X\} = 0 + 9/2 = 4.5$ and $\sigma^2 = 99/12 = 8.25$. A graph of a discrete probability distribution is shown in the top panel of Figure 3.7.

The *continuous uniform* or *rectangular* probability distribution is the continuous analog to the discrete uniform probability distribution. A *continuous uniform random variable* has uniform probability density over an interval. The *continuous uniform probability density function* is

$$f(x) = \frac{1}{b-a} \tag{3.34}$$

where the interval is $a \leq x \leq b$. Its mean and variance are

$$E\{X\} = \frac{b+a}{2}$$

and

$$\sigma^2\{X\} = \frac{(b-a)^2}{12}$$

and the *cumulative probability function* is

$$F(x) = P(X \leq x) = \frac{x-a}{b-a}. \tag{3.35}$$

Suppose, for example, that a team preparing a bid on an excavation project assesses that the lowest competitive bid is a continuous uniform random variable X with $a =\$250.000$ and $b =\$300.000$. With X measured in units of one thousand, the density function will be

$$f(x) = 1/50 = .02$$

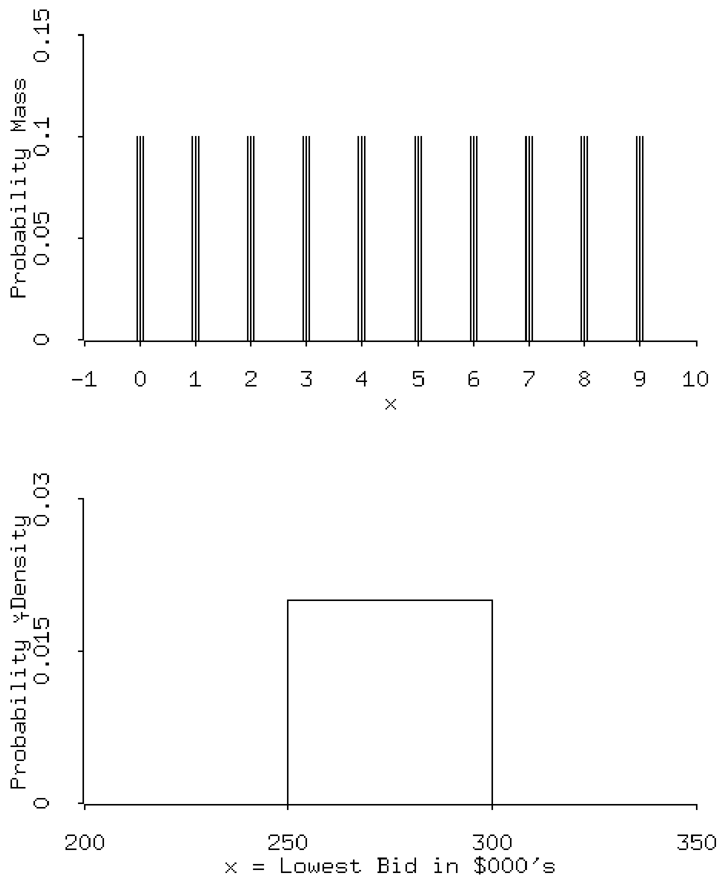


Figure 3.7: Discrete uniform probability distribution (top) and continuous uniform probability distribution (bottom).

where $250 \leq x \leq 300$. The graph of this distribution is shown in the bottom panel of Figure 3.7. The mean is 275 thousand and the variance is $50^2/12 = 2500/12 = 208.33$. The cumulative probability is the area to the left of x in the bottom panel of Figure 3.7. It is easy to eyeball the mean and the various percentiles of the distribution from the graph. The mean (and median) is the value of x that divides the rectangle in half, the lower quartile is the left-most quarter of the rectangle, and so forth. Keep in mind that, X being a continuous random variable, the probability that $X = x$ is zero.

3.10 Normal Probability Distributions

The family of normal probability distributions is the most important of all for our purposes. It is an excellent model for a wide variety of phenomena—for example, the heights of 10 year olds, the temperature in New York City at 12:01 on January 1, the IQs of individuals in standard IQ tests, etc. The *normal random variable* is a continuous one that may take any value from $-\infty$ to $+\infty$. Even though the normal random variable is not bounded, its probability distribution yields an excellent approximation to many phenomena.

The normal probability density function is

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(1/2)[(x-\mu)/\sigma]^2} \tag{3.36}$$

where $-\infty \leq x \leq +\infty$, $-\infty \leq \mu \leq +\infty$, $\sigma > 0$, $\pi = 3.14159$ and $e = 2.71828$.

The mean and variance of a normal probability distribution are

$$E\{X\} = \mu$$

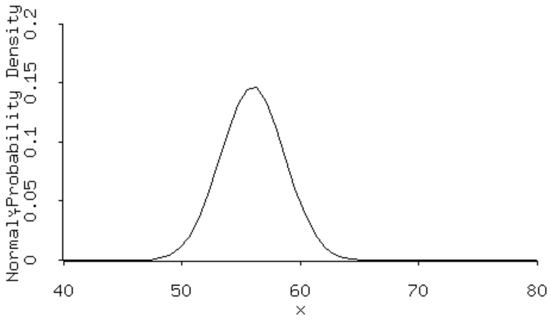
and

$$\sigma^2\{X\} = \sigma^2.$$

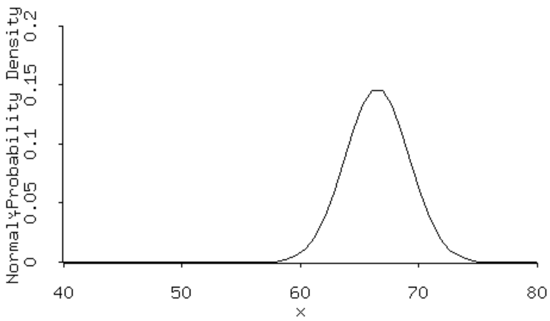
The distribution's two parameters, μ and σ, are its mean and standard deviation. Each parameter pair (μ, σ) corresponds to a different member of the family of normal probability distributions. Every normal distribution is bell shaped and symmetrical, each is centred at the value of μ and spread out according to the value of σ. Normal distributions are often referred to using the compact notation $N(\mu, \sigma^2)$. Three different members of the family of normal distributions are shown in Figure 3.8. In the top panel $\mu = 56$ and $\sigma = 2.7$ $[N(56, 7.29)]$ and in the middle panel $\mu = 66.5$ and $\sigma = 2.7$ $[N(66.5, 7.29)]$. In the bottom panel $\mu = 66.5$ and $\sigma = 4.1$ $[N(66.5, 16.81)]$.

The *standardised normal distribution* is the most important member of the family of normal probability distributions—the one with $\mu = 0$ and $\sigma = 1$. The normal random variable distributed according to the standard normal distribution is called the *standard normal variable* and is denoted by Z. It is expressed as

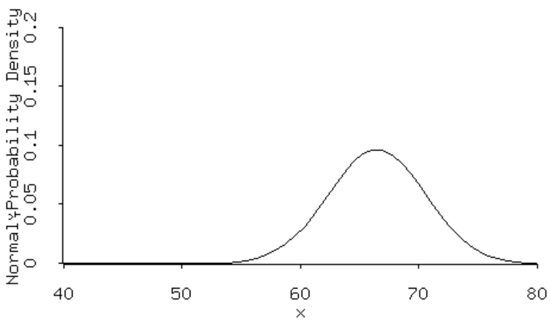
$$Z = \frac{X - \mu}{\sigma} \tag{3.37}$$



$\mu = 56\ ,\ \sigma = 2.7$



$\mu = 66.5\ ,\ \sigma = 2.7$



$\mu = 66.5\ ,\ \sigma = 4.1$

Figure 3.8: Three different members of the family of normal probability distributions.

A basic feature of normal distributions is that any linear function of a normal random variable is also a normal random variable. Thus

$$Z \quad = \quad -\frac{\mu}{\sigma} + \frac{1}{\sigma}\, X \tag{3.38}$$

and

$$X \quad = \quad \mu + \sigma\, Z \tag{3.39}$$

Figure 3.9 plots a normally distributed random variable in both its regular and standard form. It can be shown using (3.38) and (3.39) that $X = 67.715$ (i.e, 67.715 on the X scale) is equivalent to $Z = .45$ (i.e., .45 on the Z scale). This means that 67.715 is .45 standard deviations away from μ, which is 66.5. The probability that $X \geq 67.715$ is found by finding the corresponding value on the Z scale using (3.38) and looking up the relevant area to the left of that value in the table of standard normal values that can be found in the back of any textbook in statistics. To find the value of X corresponding to any cumulative probability value, we find the corresponding value of Z in the table of standard normal values and then convert that value of Z into X using (3.39). All calculations involving normal distributions, regardless of the values of μ and σ can thus be made using a single table of standard normal values.

If V and W are two independent normal random variables with means μ_v and μ_w and variances σ_v^2 and σ_w^2 respectively, the sum $V+W$ is a normal random variable with mean $\mu = \mu_v + \mu_w$ and variance $\sigma^2 = \sigma_v^2 + \sigma_w^2$. This extends, of course, to the sum of more than two random variables.

It is often useful to use the normal distribution as an approximation to the binomial distribution when the binomial sample space is large. This is appropriate when both np and $n(1-p)$ are greater than 5. To make a normal approximation we calculate the standard variate

$$Z = \frac{X - \mu}{\sigma} = \frac{X - np}{\sqrt{n\,p\,(1-p)}}. \tag{3.40}$$

We can then look up a value of Z so obtained in the normal distribution table. Alternatively, if we are given a probability of X being, say, less than a particular value we can find the value of Z from the table consistent with that probability and then use (3.39) to find the corresponding value of X.

For example, suppose we were to flip a coin 1000 times and want to know the probability of getting more than 525 heads. That is, we want to find the probability that $X \geq 525$. It turns out that

$$n\,p = n\;(1-p) = 500 > 5$$

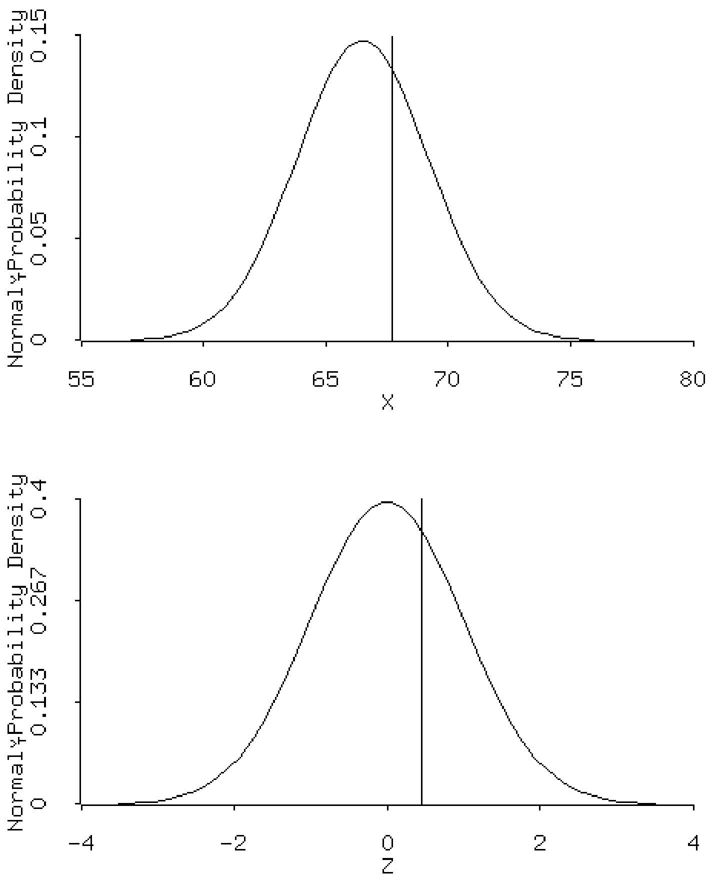


Figure 3.9: A normal distribution (top) and its standardized form (bottom). As marked by the vertical lines, 67.715 on the X scale in the top panel corresponds to .45 on the Z scale in the bottom panel.

so a normal approximation is appropriate. From (3.40) we have

$$Z = (525 - 500)/\sqrt{(1000)(.5)(.5)} = 25/\sqrt{250} = 1.58.$$

It can be seen from the probability tables for the normal distribution that

$$P(Z \leq 1.58) = .9429$$

which implies that

$$P(X \geq 525) = P(Z \geq 1.58) = 1 - .9429 = .0571.$$

There is almost a 6% chance of getting more than 525 heads.

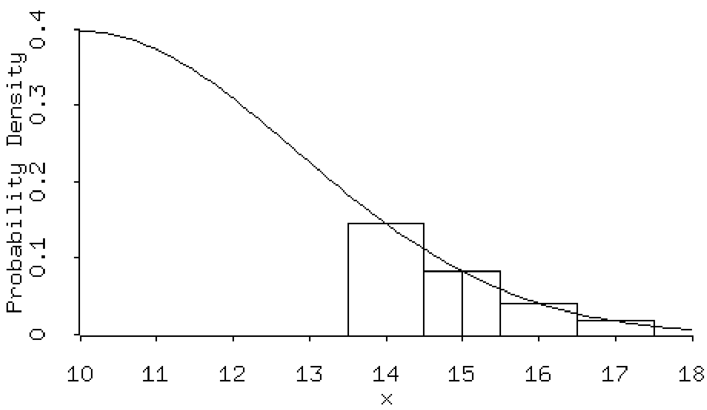


Figure 3.10: A normal approximation to a binomial distribution requiring correction for continuity.

Consider a second example of using a normal distribution to approximate a binomial one. Suppose the probability that a machine is in an unproductive state is $p = .2$. Let X denote the number of times the machine is in an unproductive state when it is observed at 50 random moments in time. It is permissible to use a normal approximation here because $(.2)(50) = 10$ and $(1-.2)(50) = 40$ and both these numbers exceed 5. The mean of distribution of X is $n\,p = 10$ and the standard deviation is

$$\sigma\{X\} = \sqrt{n\,p\,(1-p)} = \sqrt{(50)(.2)(.8)} = 2.83.$$

Now suppose we want to obtain the probability that $X = 15$. Since $n = 50$, X can be located at only 51 of the infinitely many points along the continuous line from 0 to 50. The probability that $X = 15$ on the continuum is zero. Since the underlying distribution is discrete, the probability that $X = 15$ is the area of the vertical strip under the probability density function between $X = 14.5$ and $X = 15.5$. This can be seen in Figure 3.10. So the probability that $X = 15$ becomes

$$\begin{aligned} P(X \leq 15.5) - P(X \leq 14.5) &= P(Z \leq (15.5-10)/2.83) \\ &\qquad - \; P(Z \leq (14.5-10)/2.83) \\ &= P(Z \leq 1.94) - P(Z \leq 1.59) \\ &= .9738 - .9441 \;=\; .0297. \end{aligned}$$

Similarly, if we want to calculate the probability that $X > 15$ we must calculate

$$P(X \geq 15.5) = (P(Z \geq 1.59) = 1 - P(Z \leq 1.59) = 1 - .9739 = .0262.$$

We base the calculation on $X \geq 15.5$ rather than $X \geq 15$ to correct for the fact that we are using a continuous distribution to approximate a discrete one. This is called a *correction for continuity.* It can be seen from Figure 3.10 that if we were to base our calculation on $X \geq 15$ the number obtained would be too large.

3.11 Exponential Probability Distributions

The Poisson probability distribution applies to the number of occurrences in a time interval. The *exponential* probability distribution applies to the amount of time between occurrences. For this reason it is often called the *waiting-time distribution.* It is a continuous distribution because time is measured along a continuum. An *exponential random variable* X is the time between occurrences of a random event. The probability density function is

$$f(x) = \lambda e^{-\lambda x}, \qquad (x > 0). \tag{3.41}$$

It turns out that the probability that $X \geq x$ is

$$P(X \geq x) = e^{-\lambda x}. \tag{3.42}$$

The mean and variance of an exponential distribution are

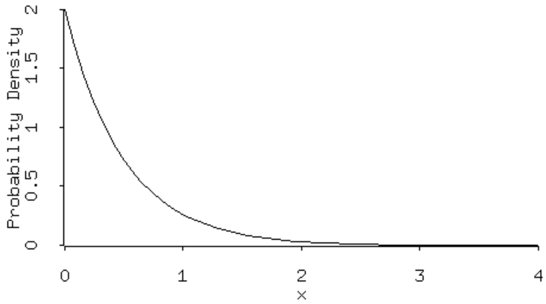
$$E\{X\} = \frac{1}{\lambda}$$

and

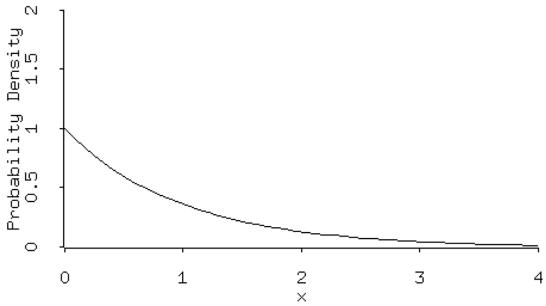
$$\sigma^2\{X\} = \frac{1}{\lambda^2}.$$

The shape of the exponential distribution is governed by the single parameter λ. As indicated in the plots of some exponential distributions in Figure 3.11, the exponential probability density function declines as x increases from zero, with the decline being sharper the greater the value of λ. The probability density function intersects the y-axis at λ.

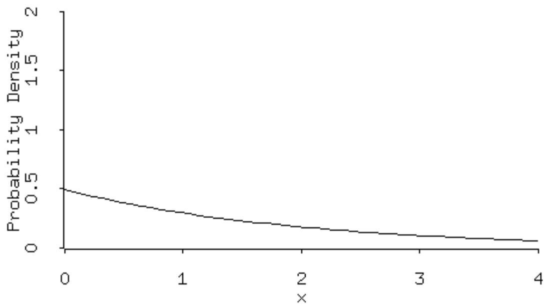
The area to the right of any value of x—that is, $P(X \geq x)$—can be looked up in the exponential distribution table at the back of any statistics textbook.



$\lambda = 2$



$\lambda = 1$



$\lambda = 0.5$

Figure 3.11: Three different members of the family of exponential probability distributions.

Consider an example. Suppose that the manufacturer of an electronic component has good reason to believe that its length of life in years follows an exponential probability distribution with $\lambda = 16$. He is considering giving a guarantee on the component and wants to know what fraction of the components he will have to replace if he makes the guarantee a five-year one. The mean time until the component breaks will be $1/\lambda = 1/16 = 6.25$ years. To find the fraction of components that will have to be replaced within 5 years we need $P(X \leq 5)$—that is, the area under the distribution to the left of $x = 5$. That area is equal to $(1 - P(X \geq 5))$ which can be found by using either equation (3.42) or the exponential distribution table. The value obtained is .550671. This means that about 55% of the components will have to be replaced within five years.

There is a close relationship between the exponential and Poisson distributions. If occurrences are generated by a Poisson process with parameter λ then the number of occurrences in equal non-overlapping units are independent random variables having a Poisson distribution with parameter λ and the durations between successive occurrences are independent random variables having the exponential distribution with parameter λ.

3.12 Exercises

1. The random variable X has a normal probability distribution with $\mu = 12$ and $\sigma = 16$. Estimate the following probabilities:

a) $P(X \leq 14.4)$

b) $P(7.2 \leq X \leq 12.8)$ (.35)

c) $P((X - \mu) \leq 5.6)$

d) $P(X \geq 8.0)$

2. The number of coating blemishes in 10-square-meter rolls of customized wallpaper is a Poisson random variable X_1 with $\lambda_1 = 0.3$. The number of printing blemishes in these 10-square-meter rolls of customized wallpaper is a Poisson random variable X_2 with $\lambda_2 = 0.1$. Assume that X_1 and X_2 are independent and let $T = X_1 + X_2$.

a) According to what distribution is the random variable T distributed?

b) What is the most probable total number of blemishes in a roll? (0)

c) If rolls with a total of two or more blemishes are scrapped, what is the probability that a roll will be scrapped? (.062)

d) What are the mean and standard deviation of the probability distribution of T?

3. There are three surviving members of the Jones family: John, Sarah, and Beatrice. All live in different locations. The probability that each of these three family members will have a stay of some length in the hospital next year is 0.2.

a) What is the probability that none of the three of them will have a hospital stay next year? (.512)

b) What is the probability that all of them will have a hospital stay next year?

c) What is the probability that two members of the family will spend time in hospital next year? (.096)

d) What is the probability that either John or Sarah, but not both, will spend time in the hospital next year?

Hint: Portray the sample space as a tree.

4. Based on years of accumulated evidence, the distribution of hits per team per nine-innings in Major League Baseball has been found to be approximately normal with mean 8.72 and standard deviation 1.10. What percentage of 9-inning Major League Baseball games will result in fewer than 5 hits?

5. The *Statistical Abstract of the United States, 1995* reports that that 24% of households are composed of one person. If 1,000 randomly selected homes are to participate in a Nielson survey to determine television ratings, find the approximate probability that no more than 240 of these homes are one-person households.

6. Suppose the f-word is heard in the main hall in a Toronto high school every 3 minutes on average. Find the probability that as many as 5 minutes could elapse without us having to listen to that profanity. (.188)

7. A manufacturer produces electric toasters and can openers. Weekly sales of these two items are random variables which are shown to have positive covariance. Therefore, higher sales volumes of toasters:

a) are less likely to occur than smaller sales volumes of toasters.

b) tend to be associated with higher sales volumes of can openers.

c) tend to be associated with smaller sales volumes of can openers.

d) are unrelated to sales of can openers.

8. Which of the following could be quantified as a Bernoulli random variable?

a) number of persons in a hospital ward with terminal diagnoses.

b) weights of deliveries at a supermarket.

c) square foot areas of houses being built in a suburban tract development.

d) whether or not employees wear glasses.

e) none of the above.

9. Fifteen percent of the patients seen in a pediatric clinic have a respiratory complaint. In a Bernoulli process of 10 patients, what is the probability that at least three have a respiratory complaint?

a) .1298

b) .1798

c) .1960

d) .9453

e) none of the above.

10. Two random variables X and Y have the following properties: $\mu_x = 10$, $\sigma_x = 4$, $\mu_y = 8$, $\sigma_y = 5$, $\sigma_{x,y} = -12$.

a) Find the expected value and variance of $(3X - 4Y)$.

b) Find the expected value of X^2. (Hint: work from the definition of the variance of X.)

c) Find the correlation between X and $(X + Y)$.

d) Find the covariance between the standardised values of X and Y.

11. John Daly is among the best putters on the PGA golf tour. He sinks 10 percent of all puts that are of length 20 feet or more. In a typical round of golf, John will face puts of 20 feet or longer 9 times. What is the probability that John will sink 2 or more of these 9 puts? What is the probability that he will sink 2 or more, given that he sinks one? Hint: If we know he is going to sink at least one then the only remaining possibilities are that he will sink only that one or two or more. What fraction of the remaining probability weight (excluding the now impossible event that he sinks zero) falls on the event 'two or more'.

12. Let X and Y be two random variables. Derive formulae for $E\{X+Y\}$, $E\{X-Y\}$, $\sigma^2\{X+Y\}$, and $\sigma^2\{X-Y\}$. Under what conditions does $\sigma^2\{X+Y\} = \sigma^2\{X-Y\}$?

13. According to the Internal Revenue Service (IRS), the chances of your tax return being audited are about 6 in 1000 if your income is less than \$50,000, 10 in 1000 if your income is between \$50,000 and \$99,999, and 49 in 1000 if your income is \$100,000 or more (*Statistical Abstract of the United States: 1995.*

a) What is the probability that a taxpayer with income less than \$50,000 will be audited by the IRS? With income between \$50,000 and \$99,999? With income of \$100,000 or more?

b) If we randomly pick five taxpayers with incomes under \$50,000, what is the probability that one of them will be audited? That more than one will be audited? Hint: What are n and p here?

14. Let $X_i = 1$ with probability p and 0 with probability $1-p$ where X_i is an independent sequence. For

$$X = \sum_{i=1}^{n} X_i$$

show that

$$E\{X\} = np$$

and

$$\sigma^2\{X\} = np(1-p).$$

15. The number of goals scored during a game by the Toronto Maple Leafs is a normally distributed random variable X with $\mu_x = 3$ and $\sigma_x = 1.2$. The number of goals given up during a game when Curtis Joseph is the goaltender for the Maple Leafs is a normally distributed random variable Y with $\mu_y = 2.85$ and $\sigma_y = 0.9$. Assume that X and Y are independent.

a) What is the probability that the Maple Leafs will win a game in which Curtis Joseph is the goaltender? (The probability of a game ending in a tie is zero here.)

b) What is the probability that the Maple Leafs will lose a game by 2 or more goals when Curtis Joseph is the goaltender?

c) Let T denote the total number of goals scored by both the Maple Leafs and their opponent during a game in which Curtis Joseph is the Leafs' goaltender. What is the expected value and variance of T?

d) Given your answer to a) and assuming that the outcomes of consecutive games are independent, what is the expected number of wins for the Maple Leafs over 50 games in which Curtis Joseph is the goaltender? Hint: What kind of process is occurring here?

16. The elapsed time (in minutes) between the arrival of west-bound trains at the St. George subway station is an exponential random variable with a value of $\lambda = .2$.

a) What are the expected value and variance of X?

b) What is the probability that 10 or more minutes will elapse between consecutive west-bound trains?

c) What is the probability that 10 or more minutes will elapse between trains, given that at least 8 minutes have already passed since the previous train arrived? Hint: What proportion of the probability weight that remains, given that a waiting time of less than 8 minutes is no longer possible, lies in the interval 8 minutes to 10 minutes?

17. The number of houses sold each month by a top real estate agent is a Poisson random variable X with $\lambda = 4$.

a) What are the expected value and standard deviation of X?

b) What is the probability that the agent will sell more than 6 houses in a given month?

c) Given that the agent sells at least 2 houses in a month, what is the probability that she will sell 5 or more?

18. In the National Hockey League (NHL), games that are tied at the end of three periods are sent to "sudden death" overtime. In overtime, the team to score the first goal wins. An analysis of NHL overtime games played between 1970 and 1993 showed that the length of time elapsed before the winning goal is scored has an exponential distribution with mean 9.15 minutes (*Chance*, Winter 1995).

a) For a randomly selected overtime NHL game, find the probability that the winning goal is scored in three minutes or less.

b) In the NHL, each period (including overtime) lasts 20 minutes. If neither team scores a goal in one period of overtime, the game is considered a tie. What is the probability of an NHL game ending in a tie?

19. A taxi service based at an airport can be characterized as a transportation system with one source terminal and a fleet of vehicles. Each vehicle takes passengers from the terminal to different destinations and then returns after some random trip time to the terminal and makes another trip. To improve the vehicle-dispatching decisions involved in such a system, a study was conducted and published in the *European Journal of Operational Research* (Vol. 21, 1985). In modelling the system, the authors assumed travel times of successive trips to be independent exponential random variables with $\lambda = .05$.

a) What is the mean trip time for the taxi service?

b) What is the probability that a particular trip will take more than 30 minutes?

c) Two taxis have just been dispatched. What is the probability that both will be gone more than 30 minutes? That at least one of the taxis will return within 30 minutes?

20. The probability that an airplane engine will fail is denoted by π. Failures of engines on multi-engine planes are independent events. A two engine plane will crash only if both of its engines fail. A four engine plane can remain airborne with two or more engines in operation. If $\pi = 0$ or $\pi = 1$, a traveller will clearly be indifferent between planes with two or four engines. What are the values of π that make a two engine plane safer than a four engine plane? Hint: Set the sample space up in tree form.