統計學

Fall 2025

政治大學統計系余清祥

2025年9月30日

第五章：離散機率分配

http://csyue.nccu.edu.tw

Chapter Contents

CENGAGE

Introduction

In this chapter, we extend the study of probability by introducing the concepts of random variables and probability distributions.

Random variables and probability distributions are models for populations of data.

The values of random variables represent the values of the data, and the probability distribution provides either the probability of each data value or a rule for computing the probability of each data value or a set of data values

The focus of this chapter is on probability distributions for discrete data, that is, discrete probability distributions.

We will introduce two types of discrete probability distributions.

- A table with one column for the values of the random variable and a second column for the associated probabilities.
- Three types of mathematical functions that represent a probability distribution: the binomial, Poisson, and hypergeometric distribution.

5.1 Examples of Discrete Random Variables

A **random variable** is a numerical description of the outcome of an experiment.

A **discrete random variable** may assume either a finite number of values or an infinite sequence of values.

Random Experiment	Random Variable (x)	Possible Values for the Random Variable
Flip a coin	Face of coin showing	1 if heads; 0 if tails
Roll a die	Number of dots showing on top of die	1, 2, 3, 4, 5, 6
Contact five customers	Number of customers who place an order	0, 1, 2, 3, 4, 5
Operate a health care clinic for one day	Number of patients who arrive	0, 1, 2, 3, …
Offer a customer the choice of two products	Product chosen by customer	0 if none; 1 if choose product A; 2 if choose product B

5.1 Examples of Continuous Random Variables

A **continuous random variable** may assume any numerical value in an interval or collection of intervals (*see notes.)

Random Experiment	Random Variable (x)	Possible Values for the Random Variable
Customer visits a web page	Time customer spends on web page in minutes	$x \geq 0$
Fill a soft drink can (max capacity = 12.1 ounces)	Number of ounces	$0 \leq x \leq 12.1$
Test a new chemical process	Temperature when the desired reaction takes place (min temperature = 150°F; max temperature = 212°F)	$150 \leq x \leq 212$
Invest $10,000 in the stock market	Value of investment after one year	$x \geq 0$

CENGAGE

5.2 Discrete Probability Distributions

The **probability distribution** for a random variable describes how probabilities are distributed over the values of the random variable.

We can describe a discrete probability distribution with a table, graph, or formula.

The Two Types of Discrete Probability Distributions

1. Uses the rules of assigning probabilities to experimental outcomes to determine probabilities for each value of the random variable.
2. Uses a special mathematical formula to compute the probabilities for each value of the random variable.

Probability Function

The **probability function**, denoted by *f*(*x*), defines the probability distribution by providing the probability for each value of the random variable.

The required conditions for a discrete probability function are:

$$\boldsymbol{f(x) \geq 0} \quad \text{and} \quad \boldsymbol{\sum f(x) = 1}$$

5.2 Empirical Discrete Distribution

There are three methods for assigning probabilities to random variables: classical method, subjective method, and relative frequency method.

The use of the relative frequency method to develop discrete probability distributions leads to an **empirical discrete distribution**.

Example: Number of automobiles sold during a day at DiCarlo Motors.

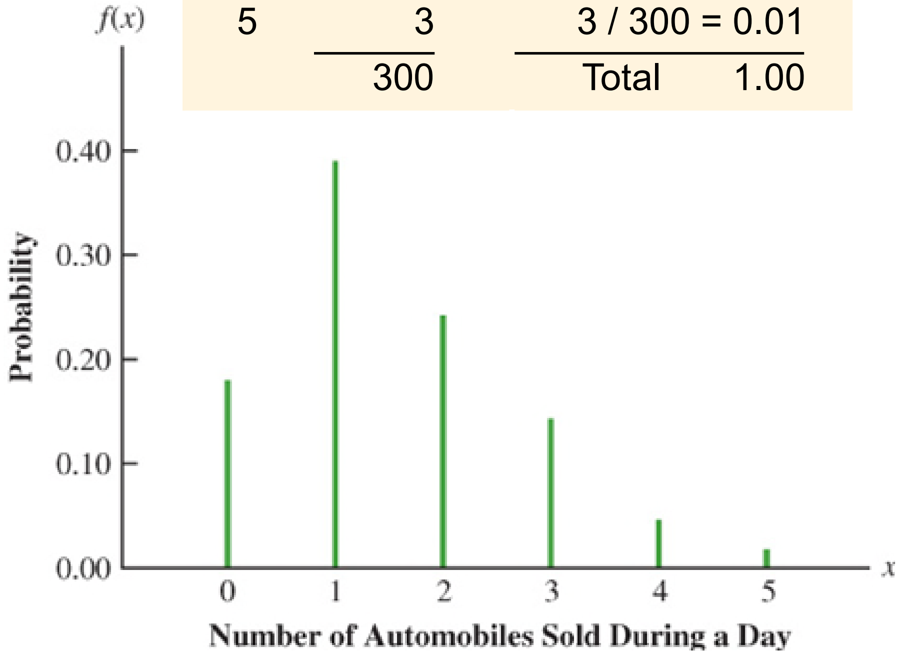
The table to the right shows the number of automobiles sold at DiCarlo over the past 300 days.

We define the random variable of interest as:

x = the number of automobiles sold during a day.

We can use the relative frequencies to represent graphically the empirical discrete distribution for x.

x	# of days	$f(x)$
0	54	54 / 300 = 0.18
1	117	117 / 300 = 0.39
2	72	72 / 300 = 0.24
3	42	42 / 300 = 0.14
4	12	12 / 300 = 0.04
5	3	3 / 300 = 0.01
	300	Total 1.00



5.2 Probability Distribution Given by a Formula

In addition to tables and graphs, a formula that gives the probability function, $f(x)$, for every value of *x* is often used to describe the probability distributions.

Typical discrete probability distributions specified by formulas are the discrete-uniform, binomial, Poisson, and hypergeometric distributions.

The **discrete uniform probability distribution** is the simplest example of a discrete probability distribution given by a formula.

The discrete uniform probability function is

$$f(x) = \frac{1}{n}$$

Where: *n* = the number of values the random variable may assume

An example of a discrete uniform probability distribution is provided by the experiment of rolling a die, in which $n = 6$ equally likely outcomes, the random variable x is represented by the number of dots on the upward face of the die, and $f(x) = 1/n = 1/6$.

CENGAGE

5.3 Expected Value

The **expected value**, or mean, of a random variable is a measure of its central location.

The expected value of a discrete random variable is calculated as a weighted average of the values the random variable may assume. The weights are the probabilities.

$$\boldsymbol{E(x) = \sum xf(x)}$$

The expected value for the DiCarlo Motors example is

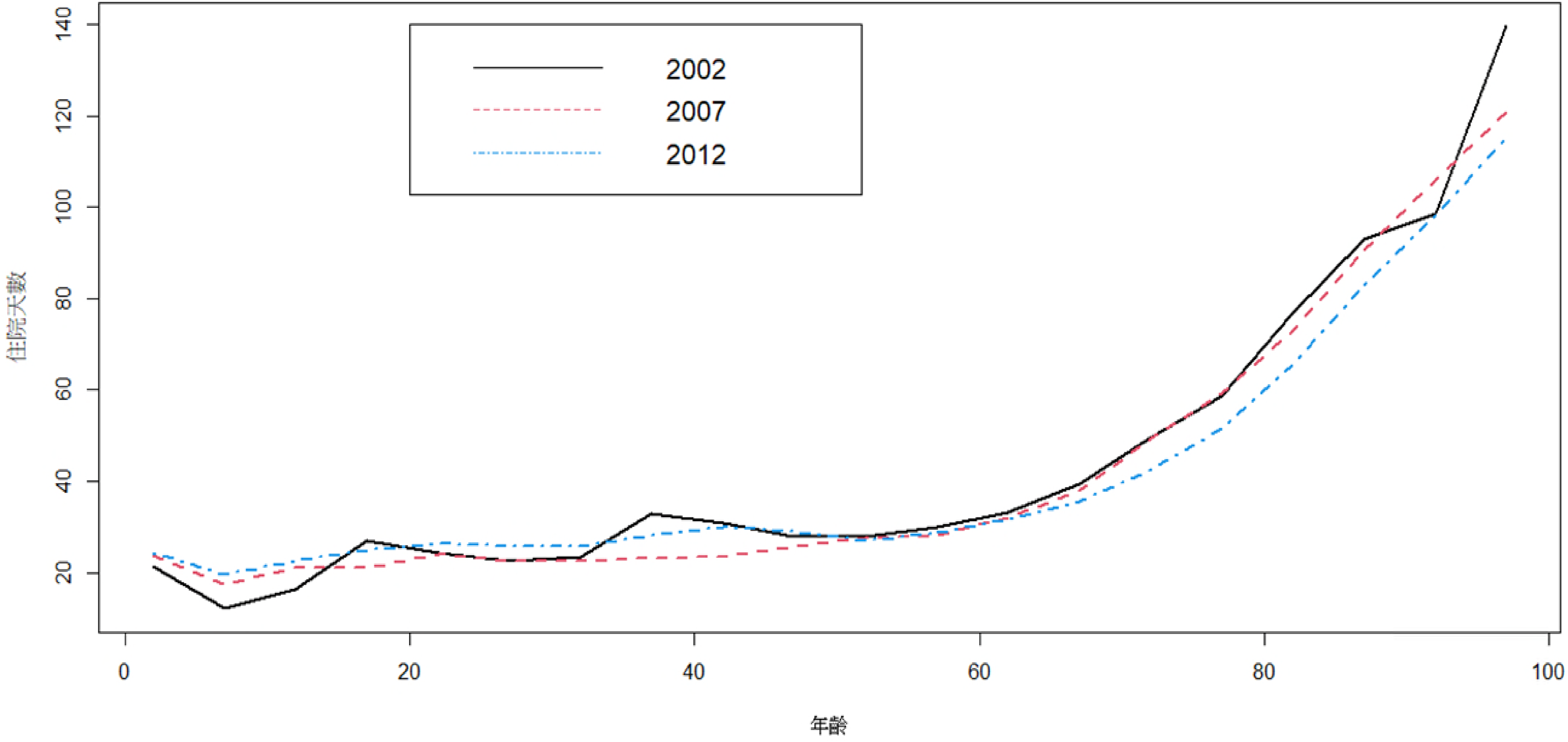
$$E(x) = \sum xf(x) = 1.50 \text{ automobiles (*see notes)}$$

Although sales from 0 to 5 automobiles are possible on any one day, over time DiCarlo can anticipate selling an average of 1.50 automobiles per day, or 30(1.50) = 45 automobiles per month.

x	***f*(*x*)**	***xf*(*x*)**
0	0.18	0(0.18) = 0.00
1	0.39	1(0.39) = 0.39
2	0.24	2(0.24) = 0.48
3	0.14	3(0.14) = 0.42
4	0.04	4(0.04) = 0.16
5	0.01	5(0.01) = 0.05
		1.50

CENGAGE

臺灣女性各年齡每年平均住院天數



5.3 Variance and Standard Deviation

The variance summarizes the variability in the values of a random variable (*see notes.)

The variance of a discrete random variable is calculated as

$$\boldsymbol{Var(x) = \sigma^2 = \sum (x - \mu)^2 f(x)}$$

For the DiCarlo Motors example, variance and standard deviation are calculated as

$$Var(x) = \sum (x - \mu)^2 f(x) = 1.25$$

$$\boldsymbol{\sigma} = \sqrt{\boldsymbol{Var(x)}} = \sqrt{1.25} = 1.118 \text{ automobiles}$$

Because the standard deviation is measured in the same units as the random variable, it is often preferred in describing the variability of a random variable.

x	$x - \mu$	$(x - \mu)^2$	$f(x)$	$(x - \mu)^2 f(x)$
0	0 − 1.5 = −1.50	2.25	0.18	2.25(0.18) = 0.4050
1	1 − 1.5 = −0.50	0.25	0.39	0.25(0.39) = 0.0975
2	2 − 1.5 = 0.50	0.25	0.24	0.25(0.24) = 0.0600
3	3 − 1.5 = 1.50	2.25	0.14	2.25(0.14) = 0.3150
4	4 − 1.5 = 2.50	6.25	0.04	6.25(0.04) = 0.2500
5	5 − 1.5 = 3.50	12.25	0.01	12.25(0.01) = 0.1225
				1.2500

CENGAGE

5.4 Bivariate Probability Distributions

A **bivariate probability distribution** is a probability distribution involving two random variables.

The outcomes of a **bivariate experiment** consist of the values of two random variables.

Examples of bivariate experiments:

- Rolling a pair of dice and recording the number for each die.
- Recording the percentage gain for a stock fund and a bond fund over a period of time.

When dealing with bivariate probability distributions, we are often interested in the relationship between the two random variables.

In this section, we introduce bivariate distributions and show how the covariance and correlation coefficient can be used as a measure of linear association between the random variables.

We shall also see how bivariate probability distributions can be used to construct and analyze financial portfolios.

CENGAGE

5.4 A Bivariate Experiment for DiCarlo Motors

Suppose we consider the bivariate experiment of observing a day of operations at DiCarlo Motors and recording the number of cars sold daily at the DiCarlo Motors automobile dealerships in Saratoga and Geneva, New York.

Let us define, $x =$ the number of cars sold at the Geneva dealership, and $y =$ the number of cars sold at the Saratoga dealership.

The table shows the number of cars sold (frequencies) at each of the dealerships over a 300-day period.

	Saratoga Dealership						
Geneva Dealership	**0**	**1**	**2**	**3**	**4**	**5**	**Total**
0	21	30	24	9	2	0	86
1	21	36	33	18	2	1	111
2	9	42	9	12	3	2	77
3	3	9	6	3	5	0	26
Total	54	117	72	42	12	3	300

CENGAGE

5.4 A Bivariate Empirical Discrete Probability Distribution for DiCarlo Motors

If we divide all the frequencies by the 300 observations, we develop a **bivariate empirical discrete probability distribution** for automobile sales at the two DiCarlo dealerships.

The probabilities in the lower and right margins provide the marginal distributions for the DiCarlo Motors Saratoga and Geneva dealerships, respectively.

The probabilities in the table provide the bivariate probability distribution for sales at both dealerships.

	Saratoga Dealership						
Geneva Dealership	**0**	**1**	**2**	**3**	**4**	**5**	**Total**
0	0.0700	0.1000	0.0800	0.0300	0.0067	0.0000	0.2867
1	0.0700	0.1200	0.1100	0.0600	0.0067	0.0033	0.3700
2	0.0300	0.1400	0.0300	0.0400	0.0100	0.0067	0.2567
3	0.0100	0.0300	0.0200	0.0100	0.0167	0.0000	0.0867
Total	0.18	0.39	0.24	0.14	0.04	0.01	1.0000

CENGAGE

5.4 Probability Distribution for Total Daily Sales at DiCarlo Motors

Let us define the total sales at DiCarlo Motors as: $s = x + y$.

From the bivariate probability distribution, we see that: $f(s{=}0) = 0.0700$, $f(s{=}1) = 0.0700 + 0.1000 = 0.1700$, $f(s{=}2) = 0.0300 + 0.1200 + 0.0800 = 0.2300$, and so on.

The table to the right shows the complete probability distribution, along with the computation of the expected value and variance of total daily sales at DiCarlo Motors:

$$E(s) = 2.6433$$

$$Var(s) = 2.3895$$

s	***f*(*s*)**	**s *f*(*s*)**	**s – *E*(*s*)**	**[s – *E*(*s*)]²**	**[s – *E*(*s*)]² *f*(*s*)**
0	0.0700	0.0000	-2.6433	6.9872	0.4891
1	0.1700	0.1700	-1.6433	2.7005	0.4591
2	0.2300	0.4600	-0.6433	0.4139	0.0952
3	0.2900	0.8700	0.3567	0.1272	0.0369
4	0.1267	0.5067	1.3567	1.8405	0.2331
5	0.0667	0.3333	2.3567	5.5539	0.3703
6	0.0233	0.1400	3.3567	11.2672	0.2629
7	0.0233	0.1633	4.3567	18.9805	0.4429
8	0.0000	0.0000	5.3567	28.6939	0.0000
		***E*(*s*) = 2.6433**			***Var*(*s*) = 2.3895**

CENGAGE

5.4 Expected Value and Variance of Daily Sales at the Geneva Dealership

With bivariate probability distributions, we often want to better understand the relationship existing between the two random variables by calculating the covariance and/or correlation coefficient as measures of association between two random variables.

To calculate the covariance, we need $Var(x)$, $Var(y)$, and $Var(s = x + y)$.

We have already computed the last two. Calculations of daily sales at the Geneva dealership follow:

$$E(x) = 1.1435$$

$$Var(x) = 0.8696$$

$$\sigma(x) = \sqrt{0.8696} = 0.9325$$

x	*f(x)*	*x f(x)*	*x* − *E*(*x*)	[*x* − *E*(*x*)]2	[*x* − *E*(*x*)]² *f*(*x*)
0	0.2867	0.0000	-1.1435	1.3076	0.3748
1	0.3700	0.3700	-0.1435	0.0206	0.0076
2	0.2567	0.5134	0.8565	0.7336	0.1883
3	0.0867	0.2601	1.8565	3.4466	0.2988
		***E*(*x*) = 1.1435**			***Var*(*x*) = 0.8696**

CENGAGE

5.4 Covariance and Correlation at DiCarlo Motors

The formula to compute the **covariance** between the random variables x and y is

$$\boldsymbol{\sigma_{xy}} = [\boldsymbol{Var(x+y)} - \boldsymbol{Var(x)} - \boldsymbol{Var(y)}]/\boldsymbol{2} = (2.3895 - 0.8696 - 1.25)/2 = 0.1350$$

The covariance of 0.1350 indicates that daily sales at DiCarlo's two dealerships have a positive relationship, but it is not suitable to get a sense of the strength of such relationship.

To assess the strength of a relationship, we use the **correlation coefficient** between *x* and *y*

$$\boldsymbol{\rho_{xy}} = \frac{\boldsymbol{\sigma_{xy}}}{\boldsymbol{\sigma_x \sigma_y}} = \frac{0.1350}{(0.9325)(1.118)} = 0.1295$$

A correlation coefficient of .1295 indicates that there is a weak positive relationship between the random variables representing daily sales at the two DiCarlo dealerships.

If the correlation coefficient had equaled zero, we would have concluded that daily sales at the two dealerships were independent.

5.4 Financial Applications

A financial advisor is considering a financial portfolio for the coming year and for each of the four economic scenarios has developed a probability distribution showing:

x = the percent return for investing in a large-cap stock fund the percent return

y = the percent return for investing in a long-term government bond fund

The bivariate probability distribution for x and y is simply a list with a separate row for each experimental outcome (economic scenario.)

Economic Scenario	Probability $f(x, y)$	Large-Cap Stock Fund (x)	Long-Term Government Bond Fund (y)
Recession	0.10	−40	30
Weak Growth	0.25	5	5
Stable Growth	0.50	15	4
Strong Growth	0.15	30	2

CENGAGE

5.4 Return and Risk of the Individual Funds

We can compute the expected percent return for investing in the stock fund, *E*(*x*), and the expected percent return for investing in the bond fund, *E*(*y*).

$$E(x) = 0.10(-40) + 0.25(5) + 0.50(15) + 0.15(30) = 9.25\%$$

$$E(y) = 0.10(30) + 0.25(5) + 0.50(4) + 0.15(2) = 6.55\%$$

Financial analysts recommend that investors also consider the standard deviation associated with an investment as a measure of risk.

First, we need to calculate the individual variances.

$$Var(x) = 0.1(-40 - 9.25)^2 + 0.25(5 - 9.25)^2 + 0.50(15 - 9.25)^2 + 0.15(30 - 9.25)^2 = 328.1875$$

$$Var(y) = 0.1(30 - 6.55)^2 + 0.25(5 - 6.55)^2 + 0.50(4 - 6.55)^2 + 0.15(2 - 6.55)^2 = 61.9475$$

Thus, the individual standard deviation for *x* and *y* are:

$$\sigma_x = \sqrt{328.1875} = 18.1159\% \quad \text{and} \quad \sigma_x = \sqrt{61.9475} = 7.8707\%$$

5.4 A Financial Portfolio of a Stock Fund and a Bond Fund

The computations of expected value and standard deviation for each of the two funds reveal that the large-cup stock fund *x* offers a more attractive return:

$$E(x) = 9.25\% > E(y) = 6.55\%$$

Conversely, investing in the long-term bond fund *y* is less risky:

$$\sigma_x = 18.12\% > \sigma_y = 7.87\%$$

The choice of investment may depend on our attitude toward risk and return:

- An aggressive investor may choose the stock fund because of the higher return.
- A conservative investor may choose the bond fund because of the lower risk.

Another option is investing in a portfolio consisting of a linear combination of the stock fund and the bond fund.

5.4 Properties of the Linear Combination of *x* and *y*

Let us define a random variable return, r, as a linear combination of *x* and *y*: $r = ax + by$

Where *a* and *b* are coefficients such that $a, b > 0$ and $a + b = 1$.

To build a financial portfolio made of 50% stock fund and 50% bond fund, we calculate $E(r)$ as

$$\boldsymbol{E(r) = E(ax + by) = aE(x) + bE(y)} = 0.5(9.25) + 0.5(6.55) = 7.9\%$$

Thus, a \$100 investment would return \$7.90.

The variance of a linear combination of two random variables x and y can be calculated as

$$\boldsymbol{Var(r) = Var(ax + by) = a^2 Var(x) + b^2 Var(y) + 2ab\sigma_{xy}}$$

Where σ_{xy} is the covariance of x and y.

Given $\sigma_{xy} = -135.3375$, variance and standard deviation of the financial portfolio are

$$Var(r) = 0.5^2(328.1875) + 0.5^2(61.9475) + 2(0.5)(0.5)(-135.3375) = 29.865$$

$$\sigma_r = \sqrt{29.865} = 5.465\%$$

CENGAGE

5.4 Risk and Return for the Three Financial Alternatives

The expected return of the financial portfolio consisting of investing 50% in the stock fund and 50% in the bond fund is halfway between that of the stock fund alone and the bond fund alone, but at a considerably lower risk, as shown by the smaller standard deviation.

Other linear combinations are possible. For example, a portfolio consisting of 25% stock fund and 75% bond fund reveals a comparable return to the 50%-50% investment, but at an even lower risk:

$$E(r) = 7.225\%$$

$$Var(r) = 4.6056$$

$$\sigma_r = 2.146\%$$

Investment Alternative	Expected Return (%)	Variance of Return	Standard Deviation of Return (%)
100% in Stock Fund	9.25	328.1875	18.1159
100% in Bond Fund	6.55	61.9475	7.8707
Portfolio (50% in stock fund, 50% in bond fund)	7.90	29.8650	5.4650

CENGAGE

5.5 The Binomial Experiment

A **binomial experiment** exhibits the following four properties.

1. The experiment consists of a sequence of n identical trials.
2. Two outcomes, *success* and *failure*, are possible in each trial.
3. The probability of a success, denoted by p, and the probability of a failure, denoted by $1 - p$, do not change from trial to trial (the *stationarity assumption.* *See notes.)
4. The trials are independent.

If properties 2-4 are present, we say the trials are generated by a Bernoulli process.

In a binomial experiment, our interest is in the number x of successes occurring in the n trials.

Because the number of successes is finite, x is a discrete random variable.

The probability distribution associated with this random variable is called the **binomial probability distribution**, and it is described by the *binomial probability function*.

5.5 Martin Clothing Store Problem

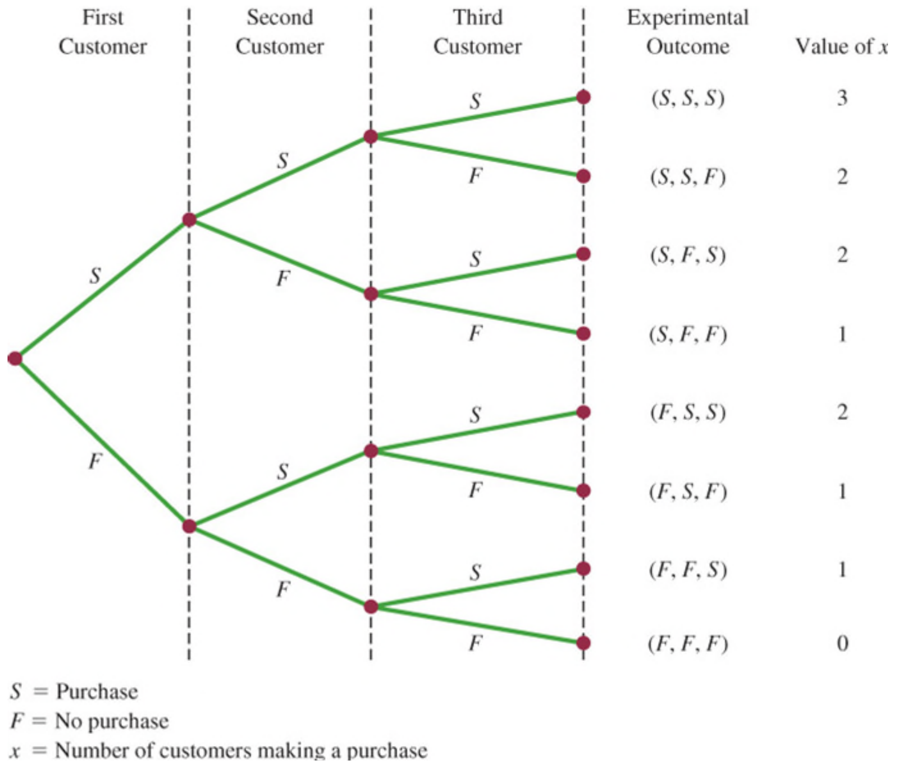
On the basis of past experience, the store manager at the Martin Clothing store estimates the probability that any one customer will make a purchase to be 0.30.

What is the probability that two of the next three customers will make a purchase?

The tree diagram to the right shows that the experiment has eight possible outcomes.

The requirements for a binomial experiment are met:

1. Three identical trials, one for each customer entering the store
1. Two outcomes: S = purchase, F = no purchase
2. Purchase: p = 0.30; no purchase: 1 – p = 0.70
3. Each purchase decision is independent



5.5 The Binomial Probability Function

The **binomial probability function** determines the probability of observing x successes in n trials in a binomial experiment.

$$\boldsymbol{f(x) = \binom{n}{x} p^{x}(1-p)^{n-x}}$$

Where:

x = the number of successes

p = the probability of a success in one trial

n = the number of trials

$f(x)$ = the probability of *x* successes in *n* trials

$\binom{n}{x} = \dfrac{n!}{x!\,(n-x)!}$ (the number of experimental outcomes providing *x* successes in *n* trials)

Table 5 of Appendix B gives the probability of x successes in n trials for a binomial experiment with varying values of the probability of success, *p* (*see notes.)

CENGAGE

5.5 The Binomial Probability Calculation for the Martin Clothing Store Problem

The binomial probability function is the product of two factors:

1. The number of experimental outcomes providing exactly x successes in n trials:

$$\binom{n}{x} = \frac{n!}{x!\,(n-x)!}$$

2. The probability of a particular sequence of trial outcomes with x successes in n trials:

$$p^x(1-p)^{n-x}$$

Thus, the probability that two $(x = 2)$ of the next three $(n = 3)$ customers will make a purchase at the Martin Clothing Store is

$$f(x) = \binom{3}{2}0.30^2(1-0.30)^{3-2} = \left(\frac{3!}{2!\,1!}\right)0.30^2(0.70)^1 = 3(0.063) = 0.189$$

5.5 The Binomial Probability Distribution for the Martin Clothing Store Problem

If we repeat the calculation of the previous slide for the probability that the number of customers making a purchase is either 0, 1, 2 or 3, we obtain the probability distribution.

The probability distribution may also be displayed with a graph, in which the height of a bar is proportional to the probability for a given value of x.

x	$f(x)$
0	$\frac{3!}{0!\,3!}(0.30)^0(0.70)^3 = 0.343$
1	$\frac{3!}{1!\,2!}(0.30)^1(0.70)^2 = 0.441$
2	$\frac{3!}{2!\,1!}(0.30)^2(0.70)^1 = 0.189$
3	$\frac{3!}{3!\,0!}(0.30)^3(0.70)^0 = 0.027$
	1.000

5.5 Properties of the Binomial Distribution

In the previous section, we provided formulas for computing the expected value and variance of a discrete random variable.

If the random variable has a binomial distribution with a known number of n trials and a known p probability of success, the general formulas for the expected value and variance are

$$\boldsymbol{E(x) = \mu = np}$$

$$\boldsymbol{Var(x) = \sigma^2 = np(1-p)}$$

For the Martin Clothing store problem, we have:

$E(x) = np = 3(0.30) = 0.9$ customers

$\sigma = \sqrt{\sigma^2} = \sqrt{np(1-p)} = \sqrt{3(0.30)(0.70)} = \sqrt{0.63} = 0.79$ customers

If the next month the Martin Clothing store forecasts 1,000 customers, we can expect

$(1000)(0.3) = 300$ customers making a purchase

with a standard deviation of $\sqrt{(1000)(0.3)(0.7)} = 14.5$ customers

5.6 Poisson Probability Distribution

A **Poisson Probability Distribution** describes *x*, the number of occurrences over a specified interval of time or space, if the following two properties of a **Poisson experiment** are respected.

1. The probability of an occurrence is the same for any two intervals of equal length.
2. The occurrence or nonoccurrence in any interval is independent of the occurrence or nonoccurrence in any other interval.

The **Poisson probability function** is described by the following equation

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

Where:

x = a discrete random variable describing the number of occurrences over an interval.

$f(x)$ = the probability of x occurrences in an interval

μ = expected value or mean number of occurrences in an interval

Also, in a Poisson probability distribution, $\boldsymbol{\sigma}^2 = \boldsymbol{\mu}$ (the variance is equal to the mean.)

CENGAGE

5.6 An Application of the Poisson Probability Distribution

An analysis of historical data shows that patients arrive at the emergency room of a large hospital on weekend mornings at the average rate $\mu = 10$ over a 15-minute interval.

What is the probability of exactly $x = 5$ arrivals in 15 minutes on a weekend morning?

In a Poisson experiment, we have

$$f(5) = \frac{\mu^x e^{-\mu}}{x!} = \frac{10^5 e^{-10}}{5!} = 0.0378$$

What if we want to know the probability of exactly $x = 1$ arrival over a 3-minute interval?

We observe that, on average, over a 3-minute interval, we have $\mu = (10/15)(3) = 2$. Thus, the probability of one arrival over a 3-minute period is

$$f(1) = \frac{\mu^x e^{-\mu}}{x!} = \frac{2^1 e^{-2}}{1!} = 0.2707$$

The probability of observing 5 arrivals over a 15-minute period is not the same as observing 1 arrival over a 3-minute period.

CENGAGE

5.6 Properties of the Poisson Probability Distribution

Because there is no stated upper limit for the number of occurrences, the Poisson probability function $f(x)$ is applicable for values $x = 0,1,2, \ldots$ without limit.

In practical applications, x will eventually become large enough so that $f(x)$ becomes approximately zero and the probability of any larger values of x negligible.

A Poisson probability distribution has the variance equal to the mean:

$$\boldsymbol{\mu = \sigma^2}$$

Probabilities can be calculated using a Poisson probability table (Table 7, Appendix B of the textbook).

For example, to obtain the probability of $x = 5$ arrivals over a 15-minute period when $\mu = 10$ using the table, follow these two steps:

1. Scroll down the leftmost column until you find $x = 5$
2. Move along the row for $x = 5$ until you reach the column corresponding to $\mu = 10$.

The answer is 0.0378. As shown before, if using $x = 1$ and $\mu = 2$, we get 0.2707 (*see notes.)

5.7 Hypergeometric Probability Distribution

The **hypergeometric probability distribution** is closely related to the binomial distribution, but it differs in two key respects:

- The trials are not independent.
- The probability of success changes from trial to trial.

The **hypergeometric probability function** is described by the following equation:

$$\boldsymbol{f(x)} = \frac{\binom{\boldsymbol{r}}{\boldsymbol{x}}\binom{\boldsymbol{N-r}}{\boldsymbol{n-x}}}{\binom{\boldsymbol{N}}{\boldsymbol{n}}} \quad \text{for} \quad 0 \le x \le r$$

Where:

x = the number of successes

n = the number of trials

$f(x)$ = the probability of x successes in n trials

N = the number of elements in the population

r = the number of elements in the population labeled success

CENGAGE

5.7 Hypergeometric Probability Function

The hypergeometric probability function, introduced in the previous slide, is used to compute the probability that in a random selection of n elements, selected without replacement, we obtain x elements labeled *success* and $n - x$ elements labeled *failure*.

For this outcome to occur, we must obtain x successes from the r successes in the population and $n - x$ failures from the $N - r$ failures.

The hypergeometric probability function, $f(x)$, the probability of obtaining x successes in n trials, shown in the previous slide, is made of the following three components:

$\binom{r}{x}$ = the number of ways that x successes can be selected from r successes in the population

$\binom{N-r}{n-x}$ = the number of ways that $n - x$ failures can be selected from a total of $N - r$ failures

$\binom{N}{n}$ = the number of ways n elements can be selected from a population of size N

CENGAGE

5.7 An Application of the Hypergeometric Distribution

If a box of $N = 12$ electric fuses contains exactly $r = 5$ defective fuses, what is the probability that a quality inspector will find exactly $x = 1$ defective fuse out of $n = 3$ selected?

$$f(x) = \frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}} = \frac{\binom{5}{1}\binom{12-5}{3-1}}{\binom{12}{3}} = \frac{\binom{5}{1}\binom{7}{2}}{\binom{12}{3}} = \frac{\left(\frac{5!}{4!}\right)\left(\frac{7!}{2!\,5!}\right)}{\left(\frac{12!}{3!\,9!}\right)} = \frac{(5)(21)}{(220)} = 0.4773$$

If we want instead the probability of finding at least one defective fuse ($x \geq 1$), it is easier to find the probability that there are no defective fuses ($x = 0$), and then take the complement.

$$f(0) = \frac{\binom{5}{0}\binom{7}{3}}{\binom{12}{3}} = \frac{\left(\frac{5!}{0!\,5!}\right)\left(\frac{7!}{3!\,4!}\right)}{\left(\frac{12!}{3!\,9!}\right)} = 0.1591$$

Thus, the probability of finding at least one defective fuse is

$$f(x \geq 1) = 1 - f(0) = 1 - 0.1591 = 0.8409$$

CENGAGE

5.7 Properties of the Hypergeometric Distribution

The expected value (mean) of a hyper-geometric distribution is calculated as

$$\boldsymbol{E(x) = \mu = n\left(\frac{r}{N}\right)}$$

In the electric fuses example, we have $n = 3$, $N = 12$, and $r = 5$. Thus

$$\mu = n\left(\frac{r}{N}\right) = 3\left(\frac{5}{12}\right) = 1.25 \text{ fuses}$$

The variance of a hypergeometric distribution can be calculated as

$$\boldsymbol{Var(x) = \sigma^2 = n\left(\frac{r}{n}\right)\left(1 - \frac{r}{n}\right)\left(\frac{N-n}{N-1}\right)}$$

In the electric fuses example, we can calculate the standard deviation as

$$\sigma = \sqrt{\sigma^2} = \sqrt{3\left(\frac{5}{12}\right)\left(1 - \frac{5}{12}\right)\left(\frac{12-3}{12-1}\right)} = \sqrt{0.60} = 0.77 \text{ fuses}$$

CENGAGE

Summary

- A random variable provides a numerical description of the outcome of an experiment.
- For any discrete random variable, the probability distribution defined by a probability function provides the probability associated with each value of the random variable.
- We introduced two types of discrete probability distributions:
  - A table consisting of the values of the random variable and associated probabilities.
  - A mathematical function that provides the probabilities for the random variable.
- We described a probability distribution with the expected value, a measure of central location, and the variance and standard deviation as measures of variability.
- We showed how to compute the covariance and correlation coefficient as measures of a bivariate relationship, and how bivariate distributions involving market returns on financial assets can be used to create financial portfolios.
- Finally, we introduced three discrete probability distributions that are described by a mathematical function: the binomial, Poisson, and hypergeometric distributions.