

Discrete Probability Distributions

Chapter 6

Learning Objectives

- LO6-1 Identify the characteristics of a probability distribution
- LO6-2 Distinguish between discrete and continuous random variables
- LO6-3 Compute the mean, variance, and standard deviation of a discrete probability distribution
- LO6-4 Explain the assumptions of the binomial distribution and apply it to calculate probabilities
- LO6-5 Explain the assumptions of the hypergeometric distribution and apply it to calculate probabilities
- LO6-6 Explain the assumptions of the Poisson distribution and apply it to calculate probabilities

What is a Probability Distribution?

PROBABILITY DISTRIBUTION A listing of all the outcomes of an experiment and the probability associated with each outcome.

CHARACTERISTICS OF A PROBABILITY DISTRIBUTION

1. The probability of a particular outcome is between 0 and 1 inclusive.
2. The outcomes are mutually exclusive.
3. The list of outcomes is exhaustive. So the sum of the probabilities of the outcomes is equal to 1.

- ▶ Example: A drug manufacturer may claim a treatment will cause weight loss for 80% of the population. This claim could be tested by a consumer protection agency using a sample and statistical inference.

Probability Distribution Example

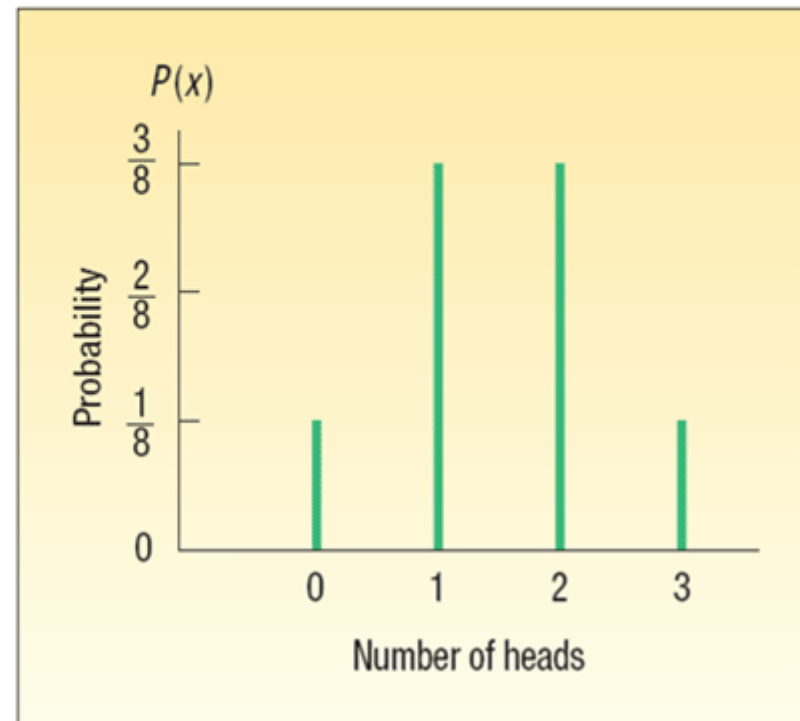
- ▶ Suppose we are interested in the number of heads showing face up with 3 tosses of a coin
- ▶ The possible outcomes are 0 heads, 1 head, 2 heads, and 3 heads

Possible Result	Coin Toss			Number of Heads
	First	Second	Third	
1	T	T	T	0
2	T	T	H	1
3	T	H	T	1
4	T	H	H	2
5	H	T	T	1
6	H	T	H	2
7	H	H	T	2
8	H	H	H	3

Probability Distribution Table

- Probability distribution table and chart for the events of zero, one, two, and three heads

Number of Heads, x	Probability of Outcome, $P(x)$
0	$\frac{1}{8} = .125$
1	$\frac{3}{8} = .375$
2	$\frac{3}{8} = .375$
3	$\frac{1}{8} = .125$
Total	$\frac{8}{8} = 1.000$



Random Variables

- ▶ In any experiment of chance, the outcomes occur randomly, and so are called random variables

RANDOM VARIABLE A quantity resulting from an experiment that, by chance, can assume different values.

- ▶ Examples
- ▶ The number of employees absent from the day shift on Monday, the number might be 0, 1, 2, 3, ... The number absent is the random variable
- ▶ The grade level (Freshman, Sophomore, Junior, or Senior) of the members of the St. James High School Varsity girls' basketball team. Grade level is the random variable

Two Types of Random Variables

- ▶ One type of random variable is the discrete random variable
- ▶ Discrete variables are usually the result of counting

DISCRETE RANDOM VARIABLE A random variable that can assume only certain clearly separated values.

- ▶ Examples
- ▶ Tossing a coin three times and counting the number of heads
- ▶ A bank counting the number of credit cards carried by a group of customers

Discrete Random Variable

- ▶ For example, the Bank of the Carolinas counts the number of credit cards carried by a group of customers
- ▶ The number of cards carried is the discrete random variable

Number of Credit Cards	Relative Frequency		
0		.03	
1		.10	
2		.18	
3		.21	
4 or more		.48	
Total		1.00	

Continuous Random Variables

- ▶ Continuous random variables can assume an infinite number of values within a given range
- ▶ Continuous variables are usually the result of measuring
- ▶ Examples
 - ▶ The time between flights between Atlanta and LA are 4.67 hours, 5.13 hours, and so on
 - ▶ The annual snowfall in Minneapolis, MN measured in inches

Mean and Variance of a Probability Distribution

- ▶ The mean is a typical value used to represent the central location of the data
- ▶ The mean is also referred to as the expected value

MEAN OF A PROBABILITY DISTRIBUTION

$$\mu = \sum [xP(x)]$$

[6-1]

- ▶ The amount of spread (or variation) in the data is described by the variance

VARIANCE OF A PROBABILITY DISTRIBUTION

$$\sigma^2 = \sum [(x - \mu)^2 P(x)]$$

[6-2]

- ▶ The standard deviation of the probability distribution is the positive square root of the variance

Probability Distribution Mean Example

John Ragsdale sells new cars for Pelican Ford. John usually sells the most cars on Saturday. He has developed a probability distribution for the number of cars he expects to sell on Saturday.

1. What type of distribution is this?
2. How many cars does John expect to sell on a typical Saturday?
3. What is the variance?

Number of Cars Sold, x	Probability, $P(x)$
0	.1
1	.2
2	.3
3	.3
4	.1
Total	1.0

Number of Cars Sold, x	Probability, $P(x)$	$x \cdot P(x)$
0	.1	0.0
1	.2	0.2
2	.3	0.6
3	.3	0.9
4	.1	0.4
Total	1.0	$\mu = 2.1$

Probability Distribution Variance Example

- ▶ The computational steps for variance
- ▶ Subtract the mean from each value of x and square
- ▶ Multiply each squared difference by its probability
- ▶ Sum the resulting products to arrive at the variance

Number of Cars Sold, x	Probability, $P(x)$	$(x - \mu)$	$(x - \mu)^2$	$(x - \mu)^2 P(x)$
0	.10	$0 - 2.1$	4.41	0.441
1	.20	$1 - 2.1$	1.21	0.242
2	.30	$2 - 2.1$	0.01	0.003
3	.30	$3 - 2.1$	0.81	0.243
4	.10	$4 - 2.1$	3.61	0.361
				$\sigma^2 = 1.290$

$$\mu = \sum XP(X) = 2(.5) + 8(.3) + 10(.2) = 5.4$$

Exercise (1 / 2)

Exercise 2: Compute the mean and variance of the following discrete probability distribution.

x	$P(x)$
2	.5
8	.3
10	.2

Exercise 4: Which of these variables are discrete and which are continuous random variables?

- ▶ a. The number of new accounts established by a salesperson in a year.
- ▶ b. The time between customer arrivals to a bank ATM.
- ▶ c. The number of customers in “Chez Choucric” barber shop
- ▶ d. The amount of fuel in your car’s gas tank.
- ▶ e. The number of minorities on a jury.
- ▶ f. The outside temperature today.

$$\mu = \sum XP(X) = 2(.5) + 8(.3) + 10(.2) = 5.4$$

Exercise (2/2)

Exercise 5: The information below is the number of daily emergency service calls made by the volunteer ambulance service for the last 50 days. To explain, there were 22 days on which there were 2 emergency calls, and 9 days on which there were 3 emergency calls.

Number of Calls	Frequency
0	8
1	10
2	22
3	9
4	1
Total	50

Number of Calls	Probability
0	0.16
1	0.20
2	0.44
3	0.18
4	0.02

- Convert this information on the number of calls to a probability distribution.
- Is this an example of a discrete or continuous probability distribution?
- What is the mean number of emergency calls per day?
- What is the standard deviation of the number of calls made daily?

Binomial Distribution

- ▶ There are four requirements of a binomial probability distribution
 1. There are only two possible outcomes and the outcomes are mutually exclusive, as either a success or a failure
 2. The number of trials is fixed and known
 3. The probability of a success is the same for each trial
 4. Each trial is independent of any other trial
- ▶ Example
- ▶ A young family has two children, both boys. The probability of the third birth being a boy is still .50. The gender of the third child is independent of the gender of the other two.

Binomial Probability Experiment

- ▶ Use the number of trials, n , and the probability of a success, π , to compute binomial probability

BINOMIAL PROBABILITY EXPERIMENT

1. An outcome on each trial of an experiment is classified into one of two mutually exclusive categories — a success or a failure.
2. The random variable is the number of successes in a fixed number of trials.
3. The probability of success is the same for each trial.
4. The trials are independent, meaning that the outcome of one trial does not affect the outcome of any other trial.

- ▶ Note: Do not confuse the symbol π , with the mathematical constant 3.1416

How is a Binomial Probability Computed?

BINOMIAL PROBABILITY FORMULA

$$P(x) = {}_nC_x \pi^x (1 - \pi)^{n-x} \quad [6-3]$$

The number of combinations **C** is:

number = n

Number_chosen = k

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

where:

C denotes a combination.

n is the number of trials.

x is the random variable defined as the number of successes.

π is the probability of a success on each trial.

$$P_{k,n} = \frac{n!}{(n-k)!}$$

There are five flights daily from Pittsburgh via US Airways into the Bradford Regional Airport in Bradford, Pennsylvania. Suppose the probability that any flight arrives late is .20.

What is the probability that none of the flights are late today?

$$\begin{aligned} P(x) &= {}_nC_r (\pi)^r (1 - \pi)^{n-r} \\ P(0) &= {}_5C_0 (.20)^0 (1 - .20)^{5-0} \\ &= (1)(1)(.3277) = .3277 \end{aligned}$$

Excel =COMBIN(5,1)

What is the probability that exactly one of the flights is late today?

$$\begin{aligned} P(x) &= {}_nC_r (\pi)^r (1 - \pi)^{n-r} \\ P(1) &= {}_5C_1 (.20)^1 (1 - .20)^{5-1} \\ &= (1)(1)(.4096) = .4096 \end{aligned}$$

Binomial Probability Distribution

There are five flights daily from Pittsburgh via US Airways into the Bradford Regional Airport in Bradford, Pennsylvania. Suppose the probability that any flight arrives late is .20. What is the probability that none of the flights are late today? What is the probability that exactly 1 of the flights is late today?

Number of Late Flights, x	$P(x)$	$xP(x)$	$x - \mu$	$(x - \mu)^2$	$(x - \mu)^2P(x)$
0	0.3277	0.0000	-1	1	0.3277
1	0.4096	0.4096	0	0	0
2	0.2048	0.4096	1	1	0.2048
3	0.0512	0.1536	2	4	0.2048
4	0.0064	0.0256	3	9	0.0576
5	0.0003	0.0015	4	16	0.0048
		$\mu = 1.0000$	$\sigma^2 = 0.7997$		

Shortcut Formulas

MEAN OF A BINOMIAL DISTRIBUTION

$$\mu = n\pi$$

[6-4]

VARIANCE OF A BINOMIAL DISTRIBUTION

$$\sigma^2 = n\pi(1 - \pi)$$

[6-5]

- ▶ Using the preceding example of flights into Bradford Airport; $n=5$ and $\pi = .20$ and the shortcut formulas

$$\mu = n\pi$$

$$\mu = (5)(.20) = 1.00$$

$$\sigma^2 = n\pi(1 - \pi)$$

$$\sigma^2 = (5)(.20)(1 - .20) = .80$$

Binomial Probability Tables

► Tables are already constructed for use as well

In the Southwest, 5% of all cell phone calls are dropped. What is the probability that out of six randomly selected calls, none was dropped? Exactly one? Exactly two? Exactly three? Exactly four? Exactly five? Exactly six out of six? See the table below for the answers.

TABLE 6-2 Binomial Probabilities for $n = 6$ and Selected Values of π

$n = 6$ Probability											
$x \backslash \pi$	.05	.1	.2	.3	.4	.5	.6	.7	.8	.9	.95
0	.735	.531	.262	.118	.047	.016	.004	.001	.000	.000	.000
1	.232	.354	.393	.303	.187	.094	.037	.010	.002	.000	.000
2	.031	.098	.246	.324	.311	.234	.138	.060	.015	.001	.000
3	.002	.015	.082	.185	.276	.313	.276	.185	.082	.015	.002
4	.000	.001	.015	.060	.138	.234	.311	.324	.246	.098	.031
5	.000	.000	.002	.010	.037	.094	.187	.303	.393	.531	.232
6	.000	.000	.000	.001	.004	.016	.047	.118	.262	.531	.735

Cumulative Binomial Probability Distributions

A study by the Illinois Department of Transportation concluded that 76.2% of front seat occupants wore seat belts. That is, both occupants of the front seat were using their seat belts. Suppose we decide to compare that information with current usage. We select a sample of 12 vehicles.

1. What is the probability that the front seat occupants in exactly 7 of the 12 vehicles are wearing seat belts?

$$P(x) = {}_nC_r(\pi)^r(1 - \pi)^{n - r}$$

$$P(x=7) = {}_{12}C_7(.762)^7(1 - .762)^{12 - 7} = 792(.149171)(.000764) = .0902$$

2. What is the probability that the front seat occupants in at least 7 of the 12 selected vehicles are wearing seat belts?

$$\begin{aligned} P(x \geq 7) &= P(x=7) + P(x=8) + P(x=9) + P(x=10) + P(x=11) + P(x=12) \\ &= .0902 + .1805 + .2569 + .2467 + .1436 + .0383 = .9562 \end{aligned}$$

The likelihood that the occupants of exactly 7 out of 12 sampled vehicles will be wearing their seat belts is about 9% and the probability of selecting 12 cars and finding that the occupants of 7 or more vehicles were wearing seat belts is .9562.

Exercise (1 / 2) - Binomial

Exercise 9: In a binomial situation, $n = 4$ and $\Phi = 0.25$. Determine the probabilities of the following events using the binomial formula.

a. $x = 2$

b. $x = 3$

Exercise 12: Assume a binomial distribution where $n = 5$ and $\Phi = 0.30$.

a. Refer to Appendix B.9, and list the probabilities for values of x from 0 to 5.

b. Determine the mean and standard deviation of the distribution from the general definitions given in formulas (6-1) and (6-2).

$n = 5$											
Probability											
x	0.05	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90	0.95
0	0.774	0.590	0.328	0.168	0.078	0.031	0.010	0.002	0.000	0.000	0.000
1	0.204	0.328	0.410	0.360	0.259	0.156	0.077	0.028	0.006	0.000	0.000
2	0.021	0.073	0.205	0.309	0.346	0.313	0.230	0.132	0.051	0.008	0.001
3	0.001	0.008	0.051	0.132	0.230	0.313	0.346	0.309	0.205	0.073	0.021
4	0.000	0.000	0.006	0.028	0.077	0.156	0.259	0.360	0.410	0.328	0.204
5	0.000	0.000	0.000	0.002	0.010	0.031	0.078	0.168	0.328	0.590	0.774

Exercise (2/2) - Binomial

Exercise 14: The United States Postal Service reports 95 percent of first class mail within the same city is delivered within two days of the time of mailing. Six letters are randomly sent to different locations.

- a. What is the probability that all six arrive within two days?
- b. What is the probability that exactly five arrive within two days?

In Excel → =BINOMDIST(5,6,0.95,FALSE)

- c. Find the mean number of letters that will arrive within two days.
- d. Compute the variance and standard deviation of the number that will arrive within two days.

Hypergeometric Distribution

- ▶ When sampling from relatively small populations without replacement, use the hypergeometric distribution

HYPERGEOMETRIC PROBABILITY EXPERIMENT

1. An outcome on each trial of an experiment is classified into one of two mutually exclusive categories — a success or a failure.
2. The random variable is the number of successes in a fixed number of trials.
3. The trials are not independent.
4. We assume that we sample from a finite population without replacement and $n/N > 0.05$. So, the probability of a success changes for each trial.

Hypergeometric Formula

HYPERGEOMETRIC DISTRIBUTION

$$P(x) = \frac{(_SC_x)(_N-S C_{n-x})}{_NC_n} \quad [6-6]$$

where:

N is the size of the population.

S is the number of successes in the population.

x is the number of successes in the sample. It may be 0, 1, 2, 3, ...

n is the size of the sample or the number of trials.

C is the symbol for a combination.

PlayTime Toys Inc. employs 50 people in the Assembly Dept. Forty of the employees belong to a union and 10 do not. Five employees are selected at random to form a committee. What is the probability that four of the five belong to a union.

$$\begin{aligned} P(4) &= \frac{(_{40}C_4)(_{{50-40}}C_{{5-4}})}{{50}C_5} \\ &= \frac{(91,390)(10)}{2,118,760} = .431 \end{aligned}$$

This the probability of selecting a union employee changes from one trial to the next

Hypergeometric Probabilities

Union Members	Probability		
0		.000	
1		.004	
2		.044	
3		.210	
4		.431	
5		.311	
		1.000	

$$P(3) = \frac{{}_{10}C_3({}_5C_1)}{{}_{15}C_4} = \frac{120(5)}{1365} = 0.4396$$

Exercise (1 / 2): Hypergeometric

- **Exercise 26:** A population consists of 15 items, 10 of which are acceptable. In a sample of 4 items, what is the probability that exactly 3 are acceptable? Assume the samples are drawn without replacement.

HYPERGEOMETRIC DISTRIBUTION

$$P(x) = \frac{{}_S C_x ({}_{N-S} C_{n-x})}{{}_N C_n}$$

[6-6]

- N = size of the population = 15
- S = Nbre of successes in the population = 10
- X = Number of successes in the sample = 3
- n = Size of the sample or the number of trials = 4
- C = symbol for a combination

$$P(3) = \frac{{}_{10}C_3({}_5C_1)}{{}_{15}C_4} = \frac{120(5)}{1365} = 0.4396$$



$$P(3) = \frac{({}_{10}C_3)({}_5C_1)}{{}_{15}C_4} = \frac{120(5)}{1365} = 0.4396$$

Exercise (2/2): Hypergeometric

- ▶ **Exercise 28:** The Computer Systems Department has 8 faculty, 6 of whom are tenured. Dr. Vonder, the chairman, wants to establish a committee of 3 department faculty members to review the curriculum. If she selects the committee at random:
 - ▶ a. What is the probability all members of the committee are tenured?
 - ▶ b. What is the probability that at least one member is not tenured? (Hint: For this question, use the complement rule.)

HYPERGEOMETRIC DISTRIBUTION

$$P(x) = \frac{({}_S C_x)({}_{N-S} C_{n-x})}{{}_N C_n} \quad [6-6]$$

- ▶ N = size of the population = 8
- ▶ S = Nbre of successes in the population = 6
- ▶ X = Number of successes in the sample = 3
- ▶ n = Size of the sample or the number of trials = 3

Poisson Probability Distribution

- ▶ This describes the number of times some event occurs during a specified interval
- ▶ The interval can be time, distance, area, or volume
- ▶ Two assumptions
 - ▶ The probability is proportional to the length of the interval
 - ▶ The intervals are independent
- ▶ The Poisson has many applications like describing
 - ▶ The distribution of errors in data entry
 - ▶ The number of accidents on I-75 during a three-month period

Poisson Distribution

POISSON PROBABILITY EXPERIMENT

1. The random variable is the number of times some event occurs during a defined interval.
2. The probability of the event is proportional to the size of the interval.
3. The intervals do not overlap and are independent.

POISSON DISTRIBUTION

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

[6-7]

where:

μ (mu) is the mean number of occurrences (successes) in a particular interval.

e is the constant 2.71828 (base of the Napierian logarithmic system).

x is the number of occurrences (successes).

$P(x)$ is the probability for a specified value of x .

MEAN OF A POISSON DISTRIBUTION

$$\mu = n\pi$$

[6-8]

Poisson Distribution Example

Budget Airlines is a seasonal airline that operates flights from South Carolina, to various cities in the northeast. Recently Budget has been concerned about the number of lost bags. Ann from the Analytics Department was asked to study the issue. She randomly selected a sample of 500 flights and found that a total of 20 bags were lost on the sampled flights.

The mean number of bags lost, μ , is found by $20/500 = .04$

The probability that no bags are lost is found using formula 6-7.

$$P(0) = \frac{\mu^x e^{-\mu}}{x!} = \frac{.04^0 e^{-0.04}}{0!} = .9608$$

Then calculate the probability that one or more bags is lost.

$$P(x \geq 1) = 1 - P(0) = 1 - \frac{\mu^x e^{-\mu}}{x!} = 1 - \frac{.04^0 e^{-0.04}}{0!} = 1 - .9608 = .0392$$

Poisson Probability Distribution Tables

NewYork-LA Trucking company finds the mean number of breakdowns on the New York to Los Angeles route is 0.30. From the table, we can locate the probability of no breakdowns on a particular run. Find the column 0.3, then read down that column to the row labeled 0; the value is .7408. The probability of 1 breakdown is .2222

μ									
x	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0	0.9048	0.8187	0.7408	0.6703	0.6065	0.5488	0.4966	0.4493	0.4066
1	0.0905	0.1637	0.2222	0.2681	0.3033	0.3293	0.3476	0.3595	0.3659
2	0.0045	0.0164	0.0333	0.0536	0.0758	0.0988	0.1217	0.1438	0.1647
3	0.0002	0.0011	0.0033	0.0072	0.0126	0.0198	0.0284	0.0383	0.0494
4	0.0000	0.0001	0.0003	0.0007	0.0016	0.0030	0.0050	0.0077	0.0111
5	0.0000	0.0000	0.0000	0.0001	0.0002	0.0004	0.0007	0.0012	0.0020
6	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0002	0.0003
7	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Exercise (1 / 2)

Exercise 32: In a Poisson distribution $\mu = 4$

- a. What is the probability that $x = 2$?
- b. What is the probability that $x \leq 2$?
- c. What is the probability that $x > 2$?

$$P(x \leq 2) = p(0) + p(1) + p(2)$$
$$P(x > 2) = 1 - [p(0) + p(1) + p(2)]$$

Exercise 33: Ms. Bergen is a loan officer at Coast Bank and Trust. From her years of experience, she estimates that the probability is .025 that an applicant will not be able to repay his or her installment loan. Last month she made 40 loans.

- a. What is the probability that 3 loans will be defaulted?
- b. What is the probability that at least 3 loans will be defaulted?

$$P(x \geq 3) = 1 - [p(0) + p(1) + p(2)]$$

Exercise (2/2)

Exercise 37: What is the difference between a random variable and a probability distribution?

A random variable is a quantitative or qualitative outcome, which results from a chance experiment. A probability distribution also includes the likelihood of each possible outcome

Exercise 43: In a recent survey, 35 percent indicated chocolate was their favorite flavor of ice cream. Suppose we select a sample of ten people and ask them to name their favorite flavor of ice cream.

- a. How many of those in the sample would you expect to name chocolate?
- b. What is the probability exactly four of those in the sample name chocolate?
- c. What is the probability four or more name chocolate?