Chapter 4

Discrete Random Variables and Probability Distributions

4.1 Random Variables

A quantity resulting from an experiment that, by chance, can assume different values. A random variable is a variable that takes on numerical values determined by the outcome of a random experiment.

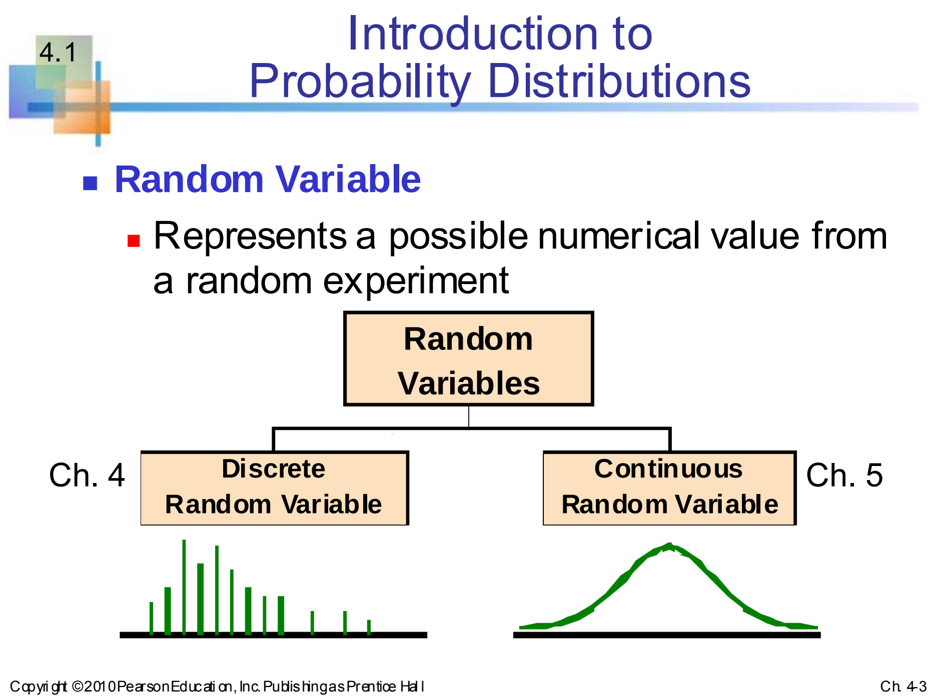


Figure 4.1:

- There are two types of random variables:

1. **Discrete Random Variables**

A random variable is discrete if it can take on no more than a *countable* number of values.

eg: X = number of heads in two flips of a coin.

2. **Continuous Random Variables**

A random variable is continuous if it can take on *any* value in an interval.

eg: X = time required to run 100 metres.

Notation

- Capital letters will denote random variables.
- Lower case letters denote the outcome of a random variable.
- $P(X = x)$ represents the probability of the random variable X having the outcome x.

4.2 Probability Distributions for Discrete Random Variables

- We can characterize the behavior of a discrete random variable X by attaching *probabilities* to each possible value, x, that X can take on. The probability distribution function, P(x), of a discrete random variable X expresses the probability that X takes the value x, as a function of x. That is P(x)=P(X=x), for all values of x.

4.3 A Probability Distribution (PDF)

- For a discrete random variable X is a table, graph, or formula that shows all possible values that X can assume along with associated probabilities.

- It is a complete (probability) description of the random variable X.

Notes:

- $0 \leq P(X = x) \leq 1$.

- $\sum_x P(X = x) = 1$.

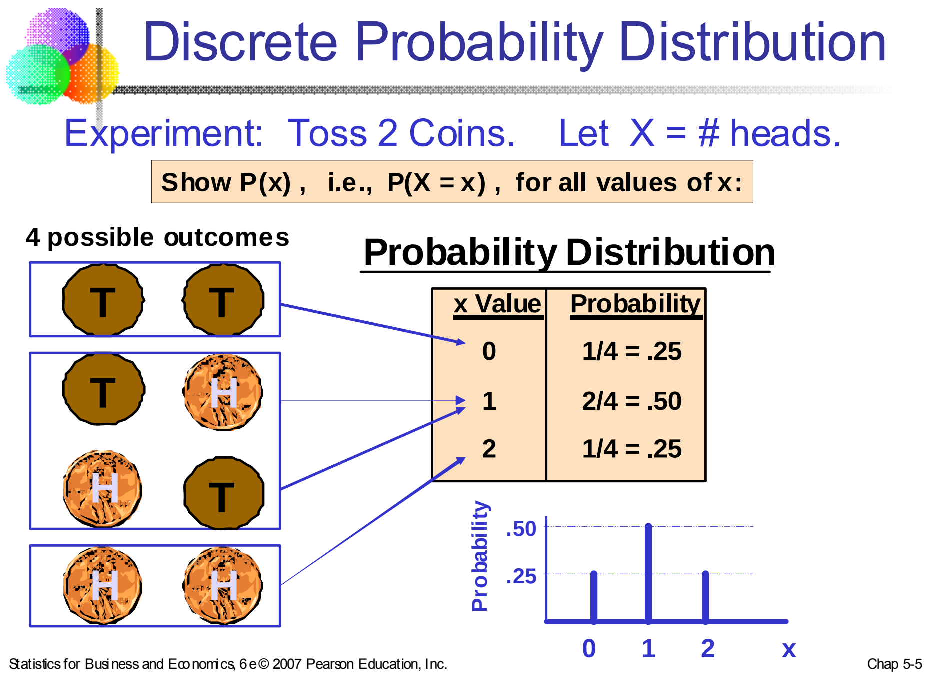

Discrete Probability Distribution
Experiment: Toss 2 Coins. Let X = # heads.
Show P(x) , i.e., P(X = x) , for all values of x:
4 possible outcomes
T T
T H
H T
H H
Probability Distribution
x Value | Probability
0 | 1/4 = .25
1 | 2/4 = .50
2 | 1/4 = .25
Probability
.50
.25
0 1 2 x
Statistics for Business and Economics, 6e © 2007 Pearson Education, Inc.
Chap 5-5

Figure 4.2:

4.4 Cumulative Distribution Function

- Let X be a random variable, then the cumulative distribution $F(x_0)$, is the function:

$$F(x_0) = P(X \leq x_0)$$

ie. $F(x)$ is the probability that the random variable X takes on a value *less than or equal* to x_0.

- Let X be a discrete random variable which can take on the values $x_1, x_2, \ldots, x_n$, and $x_1 < x_2 < \cdots < x_n$. Then (for $r \leq n$):

- $F(x_r) = P(X \leq x_r) = \sum_{i=1}^{r} P(X = x_i) \quad \forall r \leq n.$

- $0 \leq F(x_r) \leq 1 \quad \forall r.$

- If $r \leq s$ then $F(x_r) \leq F(x_s)$.

- $F(x_1) = P(X = x_1)$.

- $F(x_n) = 1$.

4.5 Descriptive Measures For Discrete Random Variables

4.5.1 Expected Value of a Discrete Random Variable

The expected value or mean of a discrete random variable X, denoted $E[X]$ or μ, is

$$E[X] = \mu = \sum_x xP(X = x)$$

- it is a weighted average of all possible values of X, the weights being the associated probabilities . $P(X = x)$

- The expected value is a measure of central tendency in that the probability distribution of X will be centered around (or balanced at) μ.

-

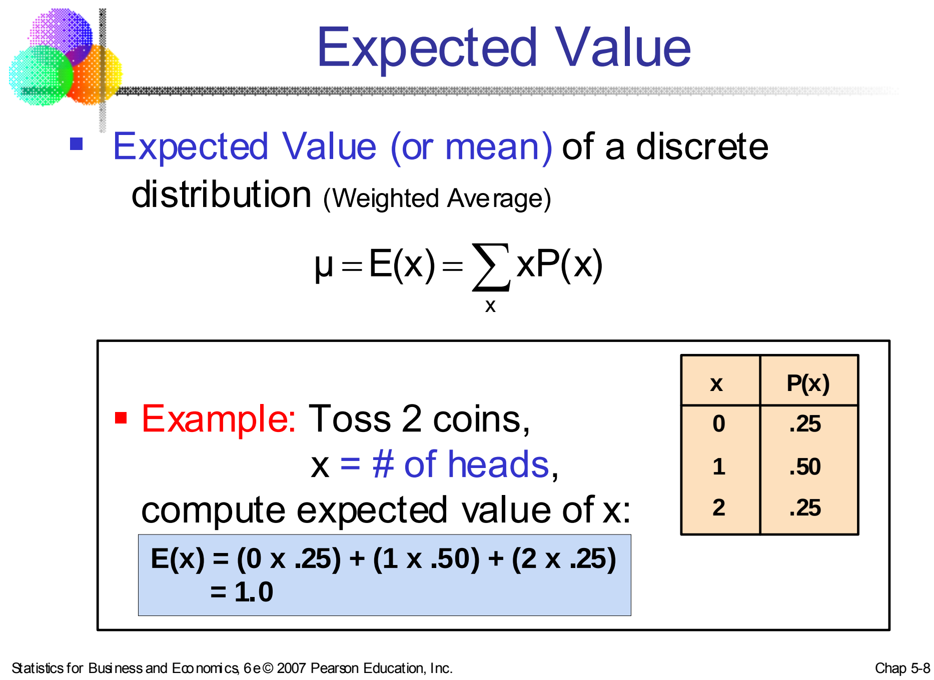


Figure 4.3:

- Note the expected value of X is not necessarily a value that X can assume. Consider tossing a fair coin once and let X equal the number of heads observed. $X = 0, 1$.

$$P(X = 0) = P(X = 1) = .5$$
$$E[X] = 0 \times 0.5 + 1 \times 0.5 = .5.$$

- *Expected Value as the balancing point of the distribution* [Transparency 5.2]

4.5.2 Variance and Standard Deviation

Variance measures the dispersion of X around it's expected value.

If $E[X] = \mu_x$, then the variance of X is:

$$V[X] = \sigma_X^2 = E[(X - \mu_X)^2] = \sum_x (x - \mu_X)^2 P(X = x)$$

Notes:

- $V[X] \geq 0$
- We can show

$$V[X] = E[(X - E[X])^2] = E[X^2] - \mu_x^2.$$

- $\sigma_X = \{V[X]\}^{1/2} \geq 0$ is the **standard deviation** of X.
- We often write $E[X] = \mu$,and $V[X] = \sigma^2$.

4.5.3 Example of Expected Value and Variance

Let $X =$ the number of heads in two straight tosses of a coin.

$X = x$	$P(X = x)$	$xP(X = x)$	$(x - \mu_x)^2 P(X = x)$	$x^2 P(X = x)$
0	.25	0	.25	0
1	.50	.50	0	.50
2	.25	.50	.25	1
		$\mu_x = 1$	$\sigma^2 = .5$	$E[X^2] = 1.5$

Note $E[X^2] - \mu^2 = 1.5 - 1 = .5 = \sigma^2$.

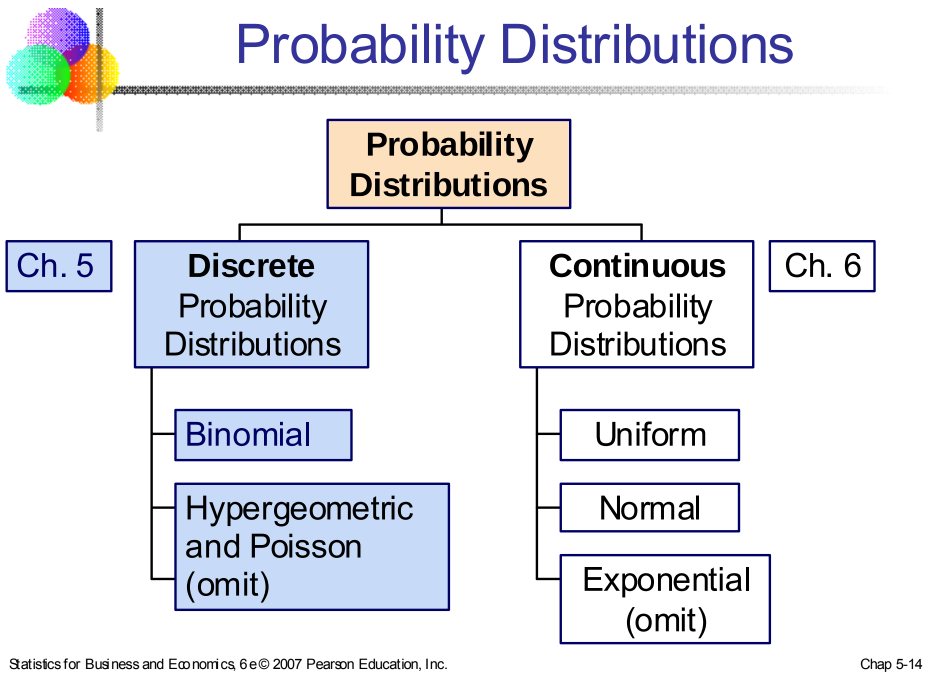


Figure 4.4:

- Chapter 4 discrete and Chapter 5 is continuous

4.6 Properties of Expectations

Let A and B be two constants and X be a discrete random variable.

4.6.1 Expectation of a Constant

- $$E[B] = \sum_x B\, P(X = x) = B \sum_x P(X = x) = B.$$

- $$E[E[X]] \;=\; E[X]$$

  since $E[X]$ **is a constant**.

- $$E[AX] = \sum_x A\, x\, P(X = x) = A \sum_x xP(X = x) = AE(X)$$

4.6.2 Expectation of a function of a random variable

Let X be a random variable with probability distribution $P(X = x)$.

Let

$$U = g(X).$$

then

$$E[U] = E[g(X)] = \sum_x g(x)P(X = x).$$

- Note setting $g(X) = (X - \mu)^2$ gives the formula for the variance

4.6.3 Expectation of a linear function of X

$$U = AX + B$$

$$E[U] = E[AX + B] = AE[X] + B$$

4.6.4 Expectation of a sum of random variables

$$E[X+Y] = E[X] + E[Y]$$

$$E[X-Y] = E[X] - E[Y]$$

4.6.5 Expectation of the Product of Independent Variables

If X and Y are independent,

$$E[XY] = E[X]E[Y]$$

4.7 Properties of Variance

4.7.1 Variance of a Constant

$$V[B] = E([B - E(B)]^2) = E([B - B]^2) = 0$$

- $$V[AX] = E[(AX - E[AX])^2] = E[(AX)^2 - 2AX\ E[AX] + (E[AX])^2]$$

  $$= A^2\ (E[X^2] - 2XE[X] + (E[X])^2) = A^2V[X]$$

4.7.2 Variance of a linear function of a random variable

Let

$$U = AX + B.$$

then

$$V[U] = V[AX + B] = A^2V[X].$$

4.7.3 Variance of Independent Random variables

If X and Y are **independent**

$$V[X+Y] = V[X] + V[Y],$$

since:

$$V[X+Y] = E([X+Y-E(X+Y)]^2)$$

$$= E([(X-E(X)]^2 - 2(X-E(X))(Y-E(Y)) + [Y-E(Y)]^2)$$

$$= E([X-E(X)]^2) + E([Y-E(Y)]^2)$$

(by independence the cross product terms are zero).

- Variance of the sum is the sum of the variances for independent random variables

$$V[X-Y] \;=\; V[X] \;+\; V[Y].$$

- Variance of the difference of independent random variables is the sum of the variances

- Questions: NCT 5.8-5.10

4.8 Examples of Discrete Probability Distributions

4.8.1 Binomial Distibution

In order to apply the binomial distribution **three conditions** must hold.

1. There are a fixed number of trials, n , of an experiment with only two possible outcomes for each trial: "success" or "failure". These trials are called Bernoulli trials.
2. The probability of success on any trial, p , is constant. **Note that in earlier versions of notes** $p = \pi$
3. The outcome of each trial is **independent** of every other trial.

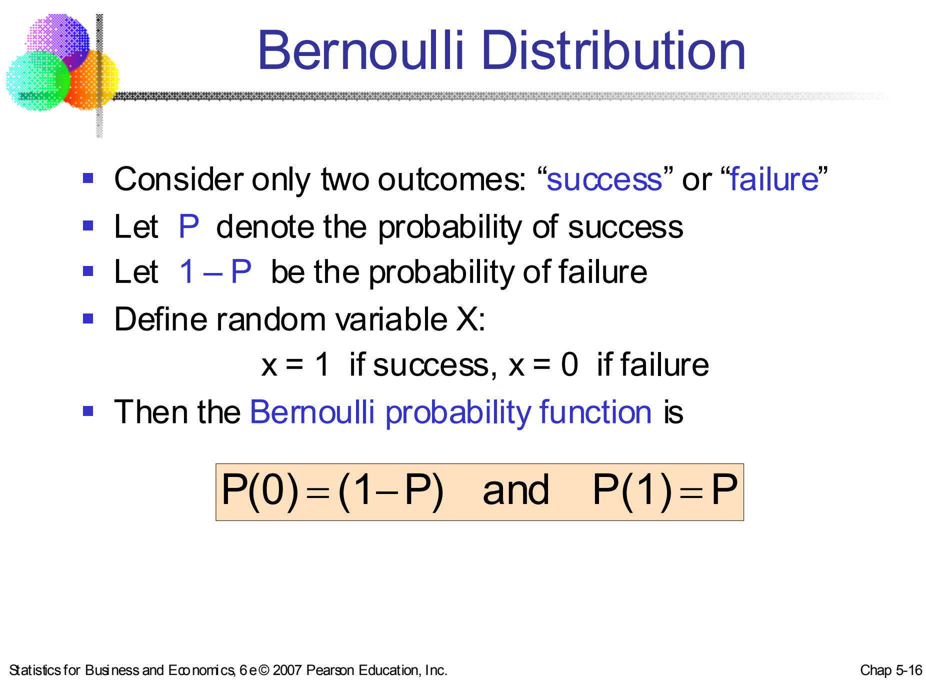

Bernoulli Distribution
Consider only two outcomes: "success" or "failure"
Let P denote the probability of success
Let 1 – P be the probability of failure
Define random variable X:
x = 1 if success, x = 0 if failure
Then the Bernoulli probability function is
P(0) = (1 − P) and P(1) = P
Statistics for Business and Economics, 6e © 2007 Pearson Education, Inc.
Chap 5-16

Figure 4.5:

Notation

- p= Probability of a success on a single trial
- $1 - p$ = Probability of a failure on a single trial
- n = Number of trials
- X = Number of successes (X is the binomial variable)

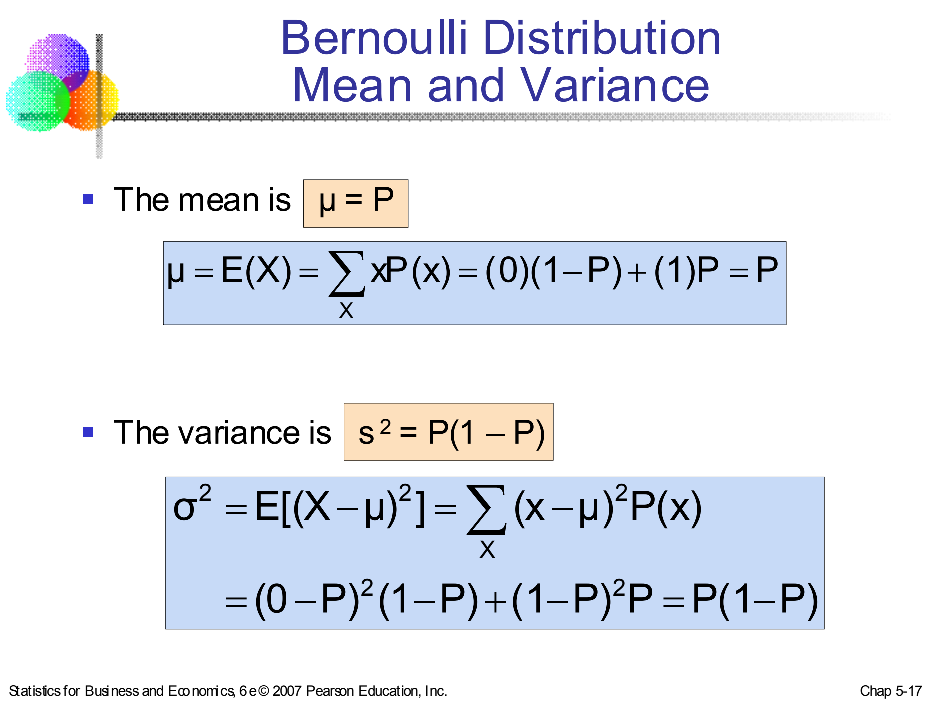


Figure 4.6:

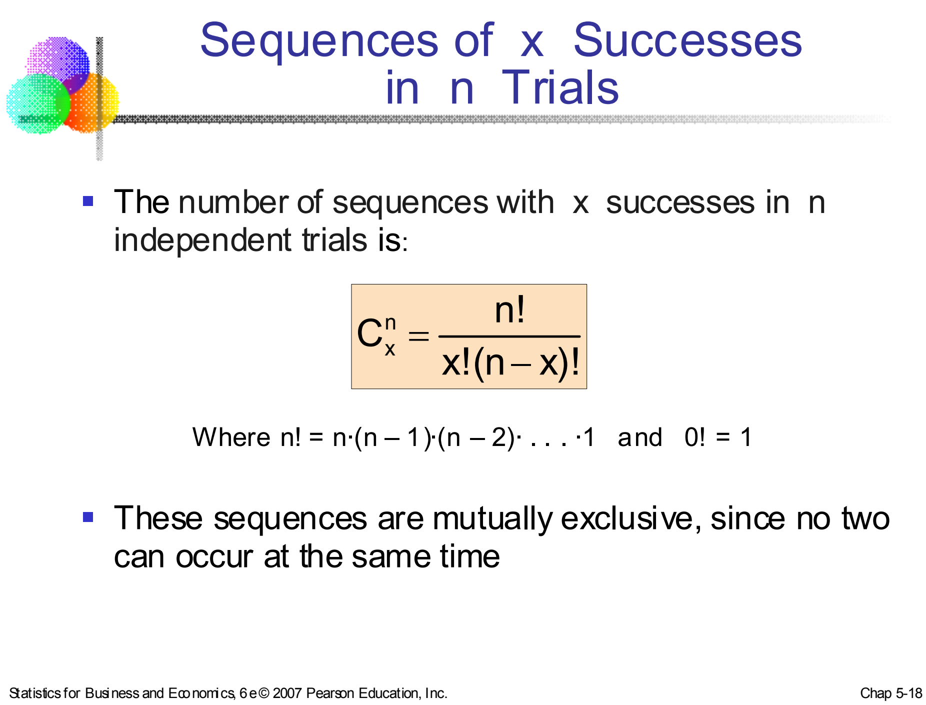

Sequences of x Successes in n Trials
The number of sequences with x successes in n independent trials is:
$$C_x^n = \frac{n!}{x!(n-x)!}$$
Where n! = n·(n – 1)·(n – 2)· . . . ·1 and 0! = 1
These sequences are mutually exclusive, since no two can occur at the same time
Statistics for Business and Economics, 6e © 2007 Pearson Education, Inc.
Chap 5-18

Figure 4.7:

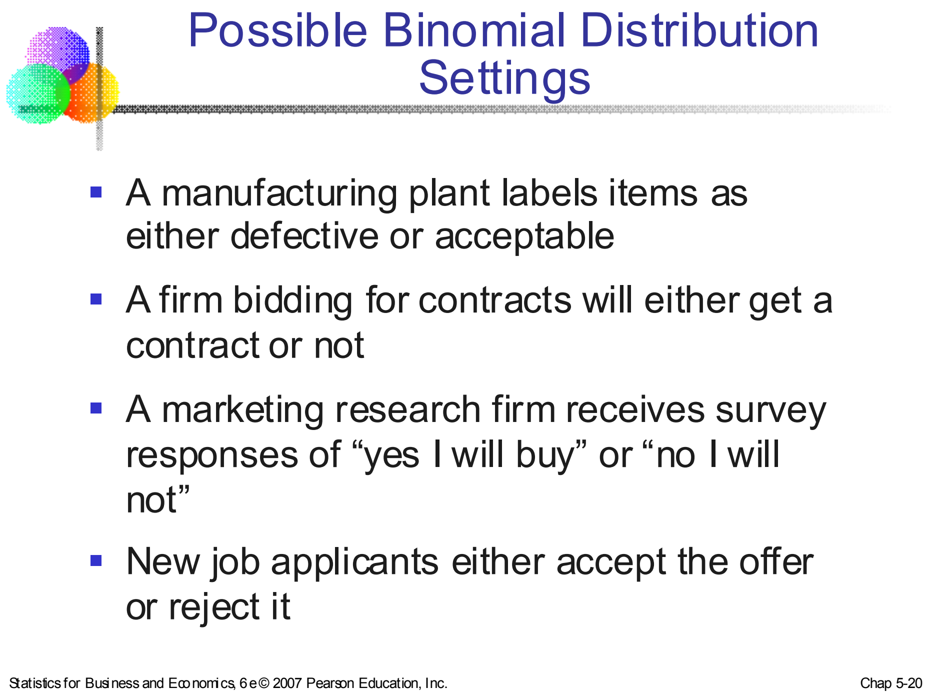

Possible Binomial Distribution Settings
A manufacturing plant labels items as either defective or acceptable
A firm bidding for contracts will either get a contract or not
A marketing research firm receives survey responses of “yes I will buy” or “no I will not”
New job applicants either accept the offer or reject it
Statistics for Business and Economics, 6e © 2007 Pearson Education, Inc.
Chap 5-20

Figure 4.8:

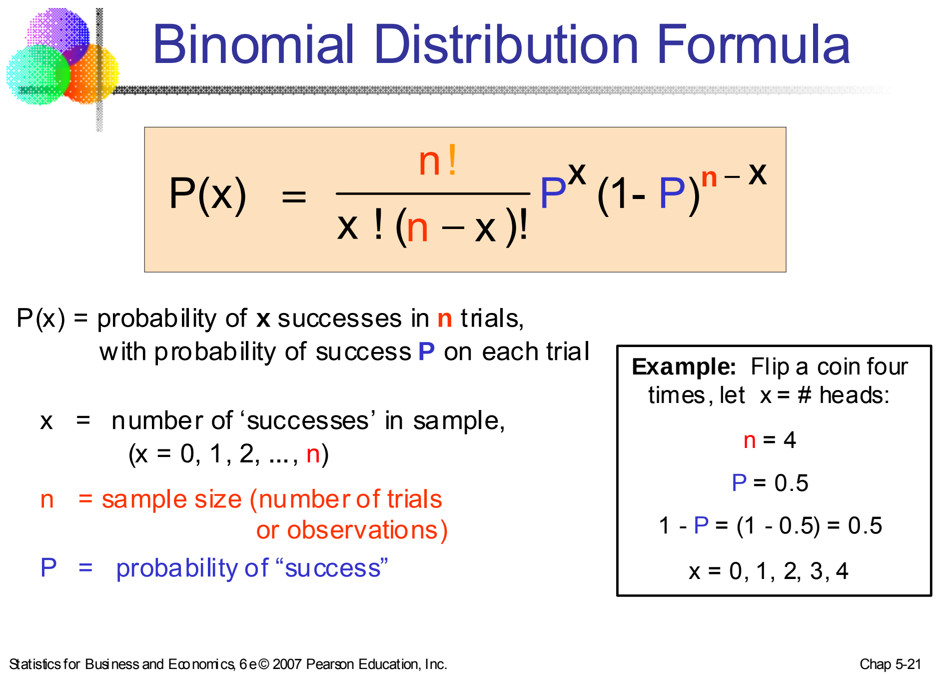

Binomial Distribution Formula
$P(x) = \frac{n!}{x!(n-x)!} P^x (1-P)^{n-x}$
P(x) = probability of x successes in n trials, with probability of success P on each trial
x = number of 'successes' in sample, (x = 0, 1, 2, ..., n)
n = sample size (number of trials or observations)
P = probability of "success"
Example: Flip a coin four times, let x = # heads:
n = 4
P = 0.5
1 - P = (1 - 0.5) = 0.5
x = 0, 1, 2, 3, 4
Statistics for Business and Economics, 6e © 2007 Pearson Education, Inc.
Chap 5-21

Figure 4.9:

The formula for binomial probability is:

$$P(X = x|n, p) = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x} \text{ for } x = 0, 1, 2, ..., n$$

Notes

- Usually we use the binomial distribution when sampling is done:

a. with replacement so that trials are independent and p is constant.
b. without replacement when the population is large relative to n (so that the change in p from trial to trial is not significant).

- $p^x(1-p)^{n-x}$ is the probability of *one* sequence (of n trials) containing x **successes** and $n-x$ **failures**.
- $\binom{n}{x} = C_x^n =$ is the number of sequences that contain x successes and $n-x$ failures.
- $E(X) = np$ (mean of a binomial variable)
- $V(X) = np(1-p)$ (variance of a binomial)
- Each different combination of n and p results in a different binomial distribution.

4.8.2 Example of a Binomial Calculation

- Toss a fair coin three times and let H be a success and T a failure.
- We have $n = 3, p = 0.5$ and $X =$ number of heads observed. Then:

$$P(\text{we observe only one } H) = P(X = 1) = \binom{3}{1}(0.5)^1(0.5)^2 = \frac{3}{8}$$

- $E(X) = np = 3 \times 0.5 = 1.5$
- $V(X) = np(1-p) = 3 \times (0.5) \times (0.5) = \frac{3}{4}$

4.8.3 Cumulative Binomial Probabilities

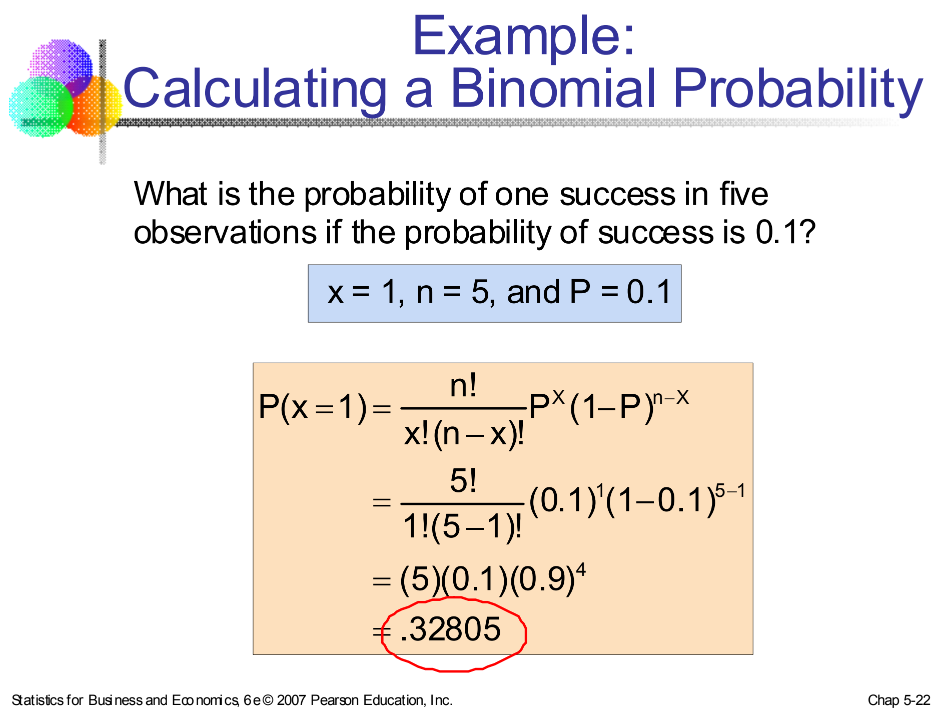


Figure 4.10:

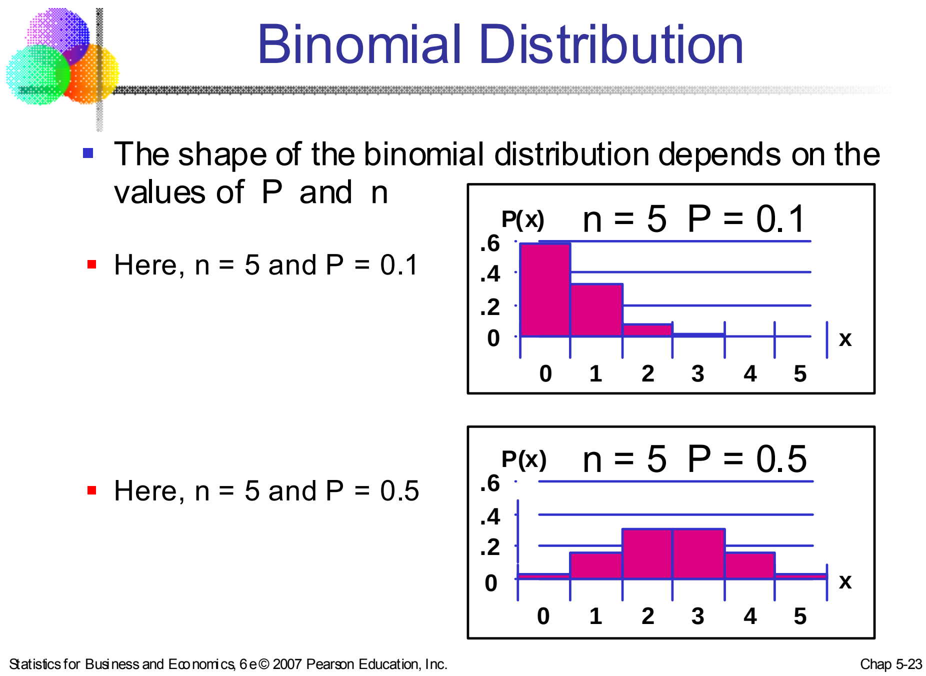

Binomial Distribution
The shape of the binomial distribution depends on the values of P and n
Here, n = 5 and P = 0.1
P(x)
n = 5 P = 0.1
.6
.4
.2
0
0 1 2 3 4 5
x
Here, n = 5 and P = 0.5
P(x)
n = 5 P = 0.5
.6
.4
.2
0
0 1 2 3 4 5
x
Statistics for Business and Economics, 6e © 2007 Pearson Education, Inc.
Chap 5-23

Figure 4.11:

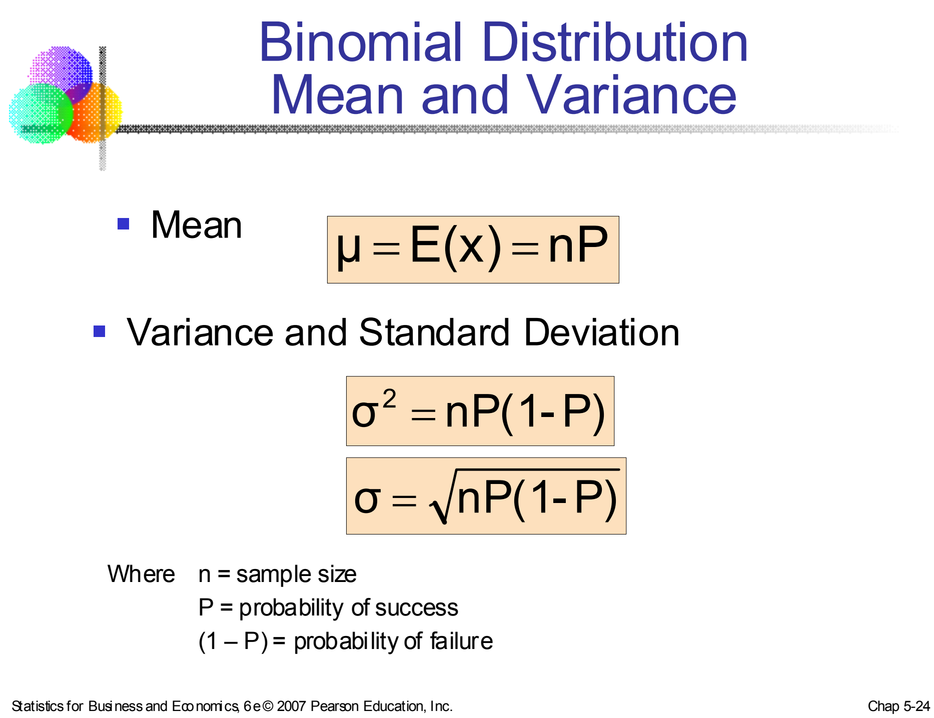


Figure 4.12:

The cumulative probability function for the binomial distribution is:

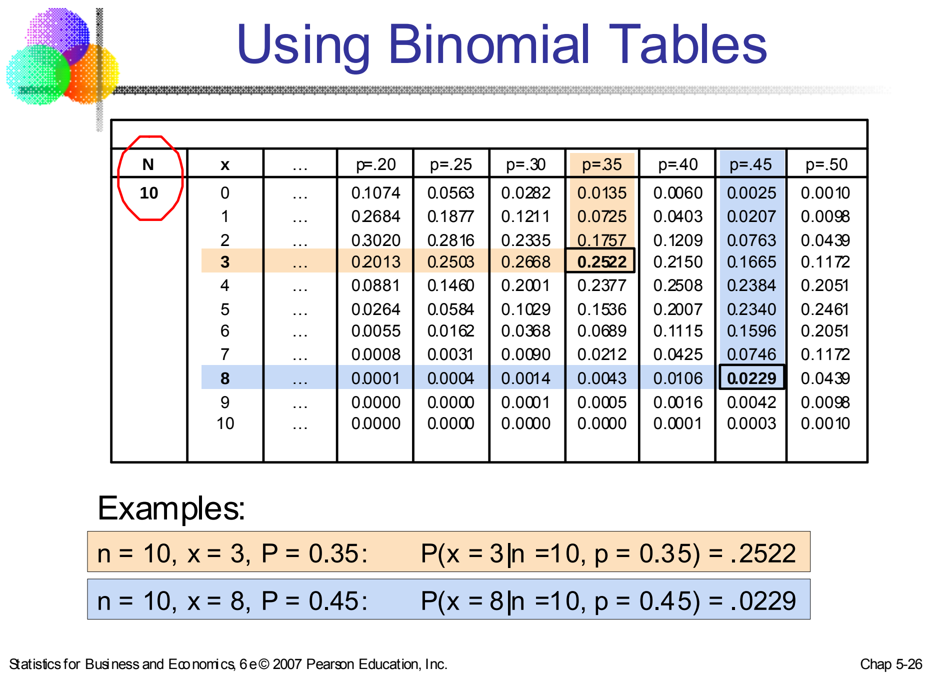
$$P(X \leq x|n,p) = \sum_{k=0}^{x} \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k}.$$

- For instance

$$P(X \leq 2|n = 3, p = .2) = \sum_{k=0}^{2} \frac{3!}{k!(3-k)!}.2^k (.8)^{n-k}.$$

- Calculating and summing individual binomial probabilities can take a great deal of work!

- Values of the binomial and cummulative function are computed in NCT Appendix **Table 2 and 3** for various values of n, p and x are from a table at the back of this chapter.



Using Binomial Tables

N	x	…	p=.20	p=.25	p=.30	p=.35	p=.40	p=.45	p=.50
10	0	…	0.1074	0.0563	0.0282	0.0135	0.0060	0.0025	0.0010
	1	…	0.2684	0.1877	0.1211	0.0725	0.0403	0.0207	0.0098
	2	…	0.3020	0.2816	0.2335	0.1757	0.1209	0.0763	0.0439
	3	…	0.2013	0.2503	0.2668	**0.2522**	0.2150	0.1665	0.1172
	4	…	0.0881	0.1460	0.2001	0.2377	0.2508	0.2384	0.2051
	5	…	0.0264	0.0584	0.1029	0.1536	0.2007	0.2340	0.2461
	6	…	0.0055	0.0162	0.0368	0.0689	0.1115	0.1596	0.2051
	7	…	0.0008	0.0031	0.0090	0.0212	0.0425	0.0746	0.1172
	8	…	0.0001	0.0004	0.0014	0.0043	0.0106	**0.0229**	0.0439
	9	…	0.0000	0.0000	0.0001	0.0005	0.0016	0.0042	0.0098
	10	…	0.0000	0.0000	0.0000	0.0000	0.0001	0.0003	0.0010

Examples:

n = 10, x = 3, P = 0.35: P(x = 3|n =10, p = 0.35) = .2522

n = 10, x = 8, P = 0.45: P(x = 8|n =10, p = 0.45) = .0229

Statistics for Business and Economics, 6e © 2007 Pearson Education, Inc. Chap 5-26

Figure 4.13:

- Table 3 does the cummulative binomial probabilities
- Note that marginal probabilities can be calculated from the Table as follows:

$$P(X = x) = P(X \leq x) - P(X \leq x - 1).$$

- Note as well:

$$P(X \geq x) = 1 - P(X \leq x - 1).$$

- Table 2 only lists a few values of n and p , so we can approximate using the nearest values.

4.8.4 Example of Cumulative Binomial Distribution

Managers for the State Department of Transportation know that 70 % of the cars arrive at a toll for a bridge have the correct change. If 20 cars pass through the toll in the next 5 minutes, what is the probabiltiy that between 10 and 15 cars, inclusive, have the correct change?

Answer

Let X be the number of people with the correct change. Clearly this is a binomial problem (verify that it satisfies the 3 conditions set out earlier). Clearly we have **independence** of trials since the fact that one driver has the correct change in no way influences the next driver. From the problem we are given $n = 20$ and $p = .70$. However our tables only have p up to .5 so we need to redefine the problem . Let Y be the number of people with wrong change so that we are looking for the interval between 5 and 10 inclusive with the wrong change and the appropriate $p = .3$

$$\begin{aligned} P(5 \leq Y \leq 10 \,|n = 20, p = .30) &= P(Y \leq 10) - P(Y \leq 4) \\ &= .983 - .238 \\ &= .745 \end{aligned}$$

4.9 Shape of the Binomial Distribution

1. As p approaches .5 the distribution becomes more symmetric.
a. If $p < .5$ the distribution is skewed to the right.
b. If $p > .5$ the distribution is skewed to the left.
2. As n increases the distribution becomes more bell-shaped.

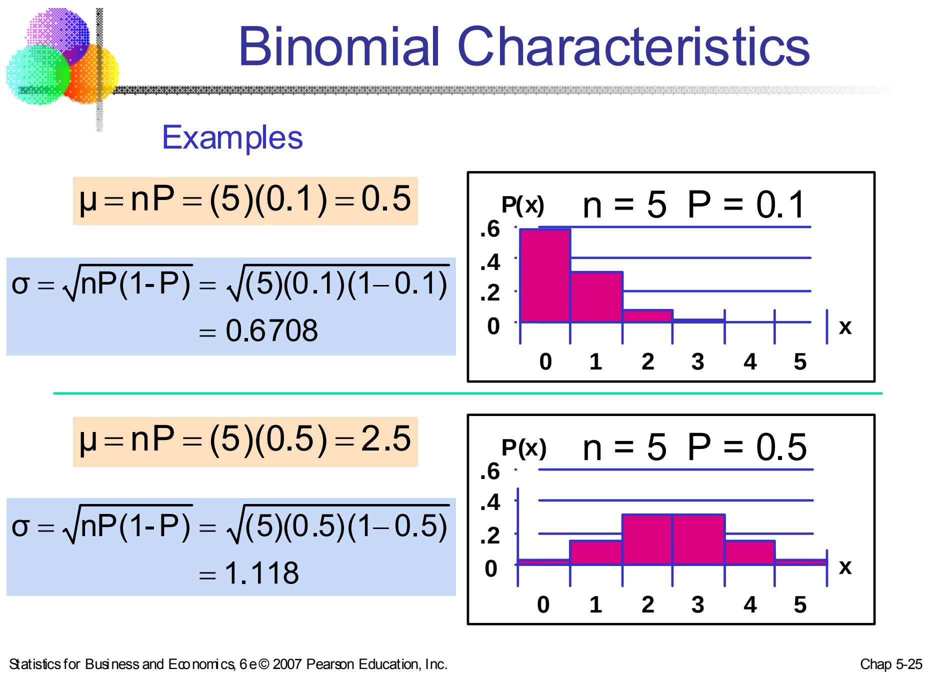


Figure 4.14:

4.10 The Binomial Fraction of Successes

We denote the fraction of succeses

$$f = \frac{X}{n}$$

.

Notes:

- As X takes one of the values $0, 1, 2, ..., n$; f takes the corresponding value $0, \frac{1}{n}, \frac{2}{n}, ..., 1$, and the $P(X = x) = P(f = \frac{x}{n})$, so the probability distribution of f is easily derived from the probability distribution of X.

- We can use the formulas developed earlier for finding the expectation and variance of a linear function of a random variable to find the expectation and variance for the fraction of successes.

- Recall if
-
$$U = AX + B$$
then
$$E[U] = AE[X] + B$$
and
$$V(U) = A^2 V[X].$$

- Applying this logic to the binomial:
$$E(f) = E[\frac{X}{n}] = \frac{1}{n}E[X] = \frac{1}{n}np = p$$

$$V(f) = V[\frac{X}{n}] = \frac{1}{n^2}V(X) = \frac{1}{n^2}np(1-p) = \frac{p(1-p)}{n}$$

4.11 Jointly Distributed Discrete Random Variables

- In this section we study the distribution of 2 discrete wandom variables
- Similar to Chapter 4 where we considered the marginal, conditional and joint distribution of two events A and$B$
- We also will show how two variables are **linearly** related

4.11.1 Joint and Marginal Probabilities

- Let X and Y be two discrete random variables, we can denote the **joint probability** distribution as:
$$P(x, y) = P(X = x \cap Y = y)$$
- To obtain the marginal distribution of X, we need to sum over all possible values of $Y = y$
$$P(x) = \sum_y P(x, y) = \sum_y P(X = x \cap Y = y)$$
- To get the marginal distribution of Y, we need to sum over all possible values of $X = x$
$$P(y) = \sum_x P(x, y) = \sum_x P(X = x \cap Y = y)$$

- Since these are probabilities we have the following:

$$
\begin{aligned}
0 &\leq P(x,y) \leq 1 \\
\sum_x \sum_y P(x,y) &= 1 \\
P(x) &\geq 0 \;, P(y) \geq 0 \\
\sum_x P(x) &= \sum_y P(y) = 1
\end{aligned}
$$

4.11.2 Conditional Probabilities and Independence

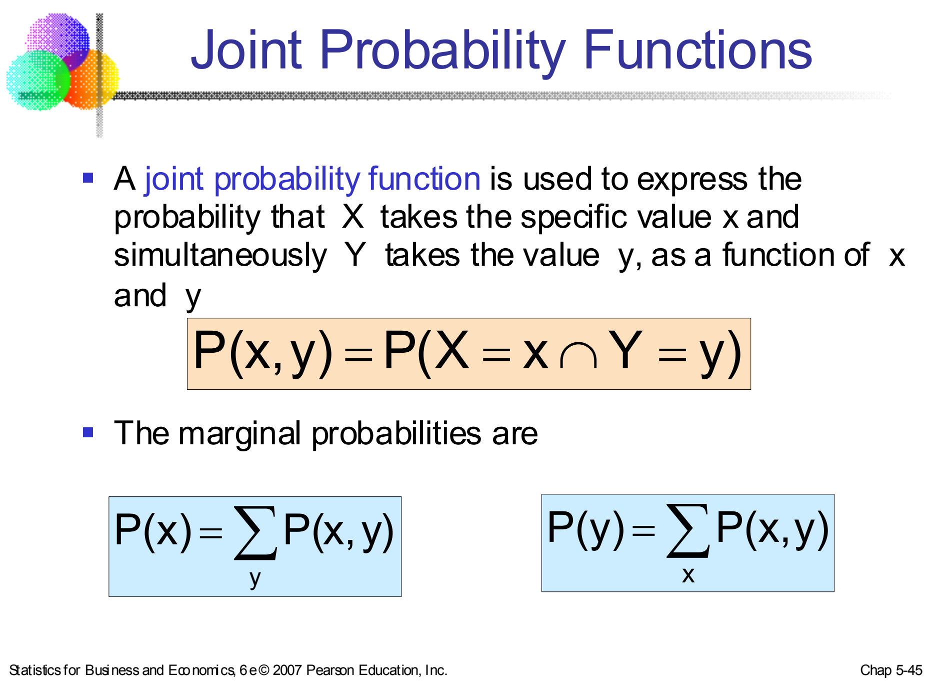


Figure 4.15:

- Again, as in Chapter 4, we can use the conditional probability formual to define:

$$P(x \mid y) = \frac{P(x,y)}{P(y)}$$

and

$$P(y \mid x) = \frac{P(x,y)}{P(x)}$$

- Independence implies that

$$\begin{aligned} P(x,y) &= P(x) \times P(y) \\ P(x &\mid y) = P(x) \\ P(y &\mid x) = P(y) \end{aligned}$$

4.11.3 Expected Value : Function of Jointly Distributed Ran-

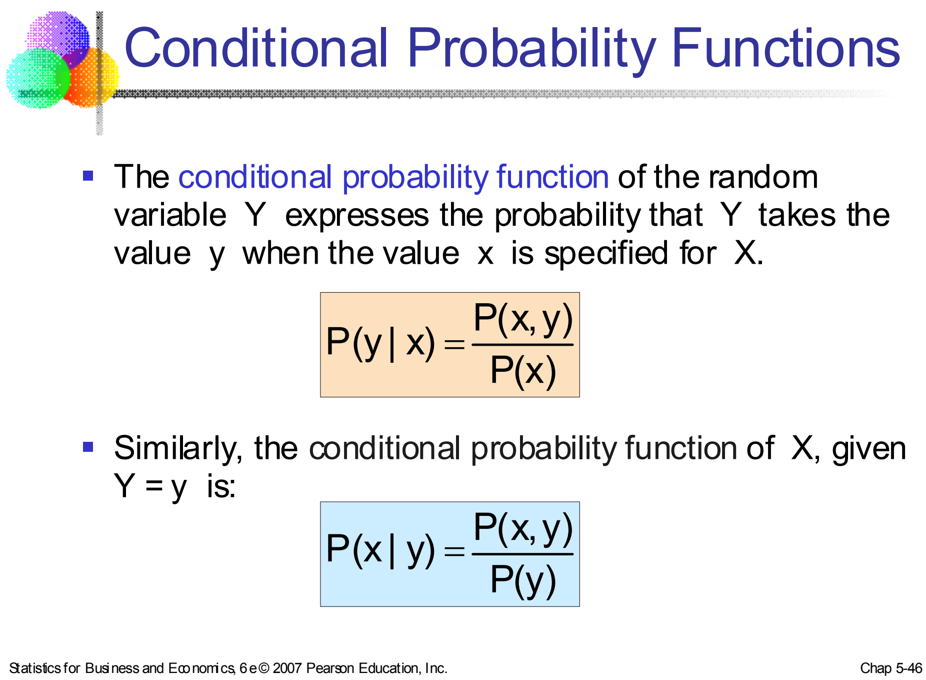


Figure 4.16:

dom Variables

- Let X and Y be two discrete random variables with joint probaility density functions $P(x, y)$. The expectation of any function $g(X, Y)$ is defined as:

$$E[g(X,Y)] = \sum_x \sum_y g(x,y) \times P(x,y)$$

- **Example from Text:** Suppose Charlotte Kind has 2 stocks, A and B. Assume that there are only 4 possible returns for each of these stocks

$$0\%, 5\%, 10\%, 15\%$$

with joint probability

	Y returns				$P(x)$
X returns	0%	5%	10%	15%	
0%	.0625	.0625	.0625	.0625	.25
5%	.0625	.0625	.0625	.0625	.25
10%	.0625	.0625	.0625	.0625	.25
15%	.0625	.0625	.0625	.0625	.25
$P(y)$	.25	.25	.25	.25	1.0

- Clearly these returns for A and B are **independent** (why?)

- Suppose that each stock costs a dollar and we have 1 and A and 2 B what is the **expected net return** of the portfollio

$$g(XX, YY) = XX + 2YY - 3$$

then (notice that we first need to express the returns as **gross returns**: ie. 0% return is agross return of 1.0 which is $xx = (1 + x) = 1.0$, 5% is a gross return of $xx = 1.05$ and so on:)

$$\begin{aligned}
E[g(X,Y)] &= \sum_{xx} \sum_{yy} (XX + 2YY - 3) \times P(xx, yy) \\
&= \sum_{xx} \sum_{yy} (XX + 2YY) P(xx) \times P(yy) - 3 \quad \text{because of independence} \\
&= [(1 \times 1.0 + 2 \times 1.0) \times .25 \times .25] + [(1 \times 1.0 + 2 \times 1.05) \times .25 \times .25 + \cdots \\
&\quad +(1 \times 1.15 + 2 \times 1.15) \times .25 \times .25 - 3 \\
&= (E[XX] + 2E[YY]) = 3 \\
&= (1.075 + 2.15) - 3 = .22
\end{aligned}$$

on a \$3 investment you are expected to earn 22 cents

4.12 Covariance

- Covariance tells us how variables move together relative to their means

- Does one variable tend to be high when another is low? (a **negative** covariance)

- Or do variable move together so both are high (relative to their means) together (a **positive** covarinace)

- Or is there no association relative to their means (a **zero** covariance)

- Definition of a covariance.. Let X and Y be 2 discrete random variables with population means of μ_X and μ_Y repsectively.

- The expected value of the product of $(X - \mu_X) \times (Y - \mu_Y)$ is the covariance:

$$\begin{aligned} Cov(X,Y) &= E\left[(X - \mu_X) \times (Y - \mu_Y)\right] = \sum_x \sum_y (X - \mu_X) \times (Y - \mu_Y) P(x,y) \\ &= E[XY] - \mu_X \mu_Y = \sum_x \sum_y xyP(x,y) - \mu_X \mu_Y \end{aligned}$$

Notice that the above expression is in units of x times y.

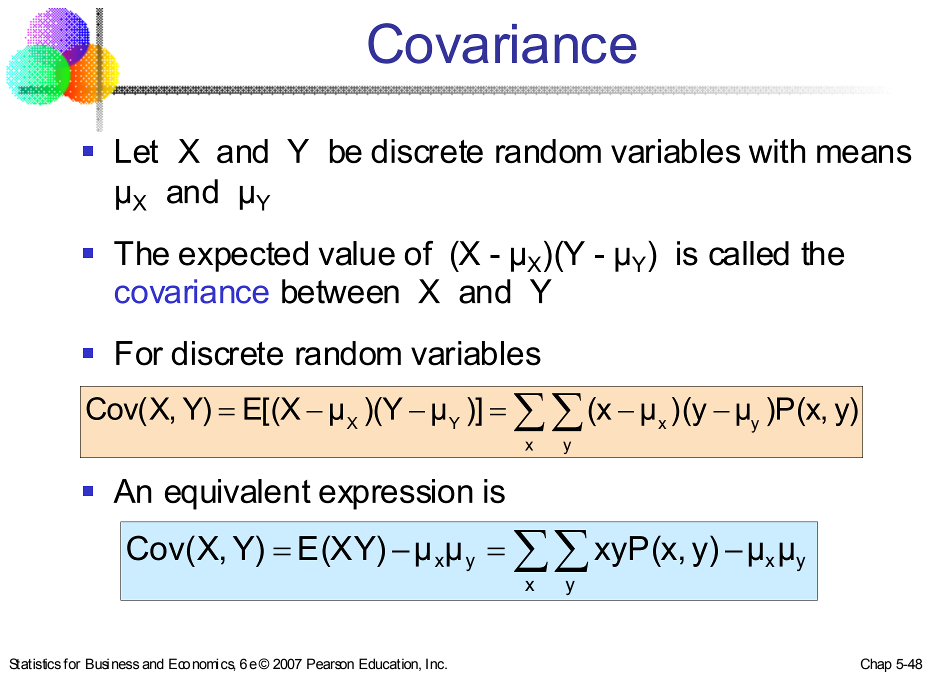

Covariance
Let X and Y be discrete random variables with means μ_X and μ_Y
The expected value of $(X - \mu_X)(Y - \mu_Y)$ is called the covariance between X and Y
For discrete random variables
$\mathrm{Cov}(X,Y) = E[(X-\mu_X)(Y-\mu_Y)] = \sum_x \sum_y (x-\mu_x)(y-\mu_y)P(x,y)$
An equivalent expression is
$\mathrm{Cov}(X,Y) = E(XY) - \mu_x\mu_y = \sum_x \sum_y xyP(x,y) - \mu_x\mu_y$
Statistics for Business and Economics, 6e © 2007 Pearson Education, Inc.
Chap 5-48

Figure 4.17:

4.13 Correlation

- We can define a measure, called the correlation coeffcient which is unit free and bound in the interval of $(-1, 1)$

$$\rho_{XY} = Corr(X, Y) = \frac{Cov(x, y)}{\sigma_X \sigma_y}$$

  this is a measure of **linear association**

- $$-1 \leq \rho_{XY} \leq 1$$

- To see that we measure linear assocation, imagine that linear relation exists for the variables

$$Y = a + bX$$

- We know the following from earlier work

$$\begin{aligned} \mu_Y &= a + b\mu_x \\ \sigma_y &= \mid b \mid \sigma_X \end{aligned}$$

- Substituting this in for

$$\begin{aligned} Cov(X, Y) &= E\left[(X - \mu_X) \times (Y - \mu_Y)\right] \\ &= E[X - \mu_X) \times (a + bX - (a + b\mu_x)] \\ &= E\left[[(X - \mu_X)[b(X - \mu_X)\right] \\ & b\sigma_X^2 \end{aligned}$$

- Now we use this and substitute into the expression for ρ_{XY}

$$\begin{aligned} \rho_{XY} &= \frac{b\sigma_X^2}{\sigma_X \mid b \mid \sigma_X} \\ &= sign(b) \times 1 \end{aligned}$$

  that is the correlation is ± 1, depending on the sign of b

4.14 Independence, Covariance and Correlation

- If the discrete random variables X and Y are independent,

$$\begin{aligned} Cov(X, Y) &= \sum_x \sum_y (X - \mu_X) \times (Y - \mu_Y) P(x, y) \\ &= \sum_x \sum_y (X - \mu_X) \times (Y - \mu_Y) P(x) \times P(y) \\ &= \sum_x (X - \mu_X) P(x) \times \sum_y (Y - \mu_Y P(y) \\ &= 0 \times 0 \\ & 0 \end{aligned}$$

so that

$$\rho_{XY} = Corr(X,Y) = \frac{Cov(x,y)}{\sigma_X \sigma_y} = \frac{0}{\sigma_X \sigma_y} = 0$$

- Note that **zero covariance does not imply independence**.
- Covariance measures linear association and independence is about any association (nonlinear as well)