

Chapter 7

Random Variables and Discrete Probability Distributions

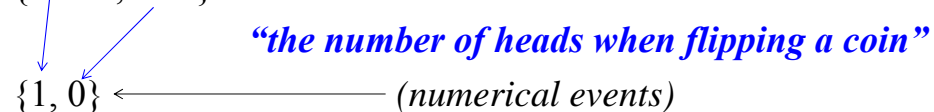
7.1

Random Variables...

A ***random variable*** is a function or rule that assigns a number to each outcome of an experiment.

Alternatively, the ***value*** of a random variable is a numerical event.

Instead of talking about the coin flipping event as {heads, tails} think of it as

 ***“the number of heads when flipping a coin”***
{1, 0} (numerical events)

7.2

Two Types of Random Variables...

Discrete Random Variable

- one that takes on a *countable* number of values
- E.g. values on the roll of dice: 2, 3, 4, ..., 12

Continuous Random Variable

- one whose values are *not discrete*, not countable
- E.g. time (30.1 minutes? 30.10000001 minutes?)

Analogy:

Integers are Discrete, while Real Numbers are Continuous

7.3

Probability Distributions...

A *probability distribution* is a table, formula, or graph that describes the values of a **random variable** and the probability associated with these values.

Since we're describing a **random variable** (which can be discrete or continuous) we have two types of probability distributions:

- Discrete Probability Distribution, (this chapter) and
- Continuous Probability Distribution (Chapter 8)

7.4

Probability Notation...

An upper-case letter will represent the *name* of the random variable, usually **X**.

Its lower-case counterpart will represent the *value* of the random variable.

The probability that the random variable **X** will equal x is:

$$P(\mathbf{X} = x)$$

or more simply

$$P(x)$$

7.5

Discrete Probability Distributions...

The probabilities of the values of a *discrete random variable* may be derived by means of probability tools such as tree diagrams or by applying one of the definitions of probability, so long as these two conditions apply:

$$1. \ 0 \leq P(x) \leq 1 \text{ for all } x$$

$$2. \ \sum_{\text{all } x_i} P(x) = 1$$

7.6

Example 7.1

The *Statistical Abstract of the United States* is published annually. It contains a wide variety of information based on the census as well as other sources. The objective is to provide information about a variety of different aspects of the lives of the country's residents. One of the questions asked households to report the number of color televisions in the household. The following table summarizes the data. Develop the probability distribution of the random variable defined as the number of color televisions per household.

7.7

Example 7.1

Number of Color Televisions	Number of Households (1,000s)
0	1,218
1	32,379
2	37,961
3	19,387
4	7,714
5	2,842
Total	101,501

7.8

Example 7.1

Probability distributions can be estimated from relative frequencies.

# of Televisions	# of Households	x	$P(x)$
0	1,218	0	0.012
1	32,379	1	0.319
2	37,961	2	0.374
3	19,387	3	0.191
4	7,714	4	0.076
5	2,842	5	0.028
	<u>101,501</u>		<u>1.000</u>

$$1,218 \div 101,501 = 0.012$$

e.g. $P(X=4) = P(4) = 0.076 = 7.6\%$

7.9

Example 7.1

E.g. what is the probability there is at least one television but no more than three in any given household?

# of Televisions	# of Households	x	$P(x)$
0	1,218	0	0.012
1	32,379	1	0.319
2	37,961	2	0.374
3	19,387	3	0.191
4	7,714	4	0.076
5	2,842	5	0.028
	<u>101,501</u>		<u>1.000</u>

“at least one television but no more than three”

$$P(1 \leq X \leq 3) = P(1) + P(2) + P(3) = .319 + .374 + .191 = .884$$

7.10

Population/Probability Distribution...

The discrete probability distribution represents a ***population***

Example 7.1 the population of number of TVs per household

Example 7.2 the population of sales call outcomes

Since we have ***populations***, we can describe them by computing various ***parameters***.

E.g. the population mean and population variance.

7.13

Population Mean (Expected Value)

The population mean is the ***weighted average*** of all of its values. The weights are the probabilities.

This parameter is also called the expected value of X and is represented by E(X).

$$E(X) = \mu = \sum_{all\ x} xP(x)$$

7.14

Population Variance...

The population variance is calculated similarly. It is the weighted average of the *squared deviations* from the mean.

$$V(X) = \sigma^2 = \sum_{\text{all } x} (x - \mu)^2 P(x)$$

As before, there is a “short-cut” formulation...

$$V(X) = \sigma^2 = \sum_{\text{all } x} x^2 P(x) - \mu^2$$

The standard deviation is the same as before: $\sigma = \sqrt{\sigma^2}$

7.15

Example 7.3...

Find the **mean**, variance, and standard deviation for the population of the number of color televisions per household... (from Example 7.1)

# of Televisions	# of Households	x	P(x)
0	1,218	0	0.012
1	32,379	1	0.319
2	37,961	2	0.374
3	19,387	3	0.191
4	7,714	4	0.076
5	2,842	5	0.028
<u>101,501</u>			<u>1.000</u>

$$\begin{aligned} E(X) = \mu &= \sum_{\text{all } x} xP(x) = 0 \cdot P(0) + 1 \cdot P(1) + \dots + 5 \cdot P(5) \\ &= 0(.012) + 1(.319) + 2(.374) + 3(.191) + 4(.076) + 5(.028) \\ &= \mathbf{2.084} \end{aligned}$$

7.16

Example 7.3...

Find the mean, **variance**, and standard deviation for the population of the number of color televisions per household... (from Example 7.1)

# of Televisions	# of Households	x	P(x)
0	1,218	0	0.012
1	32,379	1	0.319
2	37,961	2	0.374
3	19,387	3	0.191
4	7,714	4	0.076
5	2,842	5	0.028
	101,501		1.000

$$V(X) = \sigma^2 = \sum_{\text{all } x} (x - \mu)^2 P(x)$$

$$= (0 - 2.084)^2(.012) + (1 - 2.084)^2(.319) + \dots + (5 - 2.084)^2(.028)$$

$$= \mathbf{1.107}$$

7.17

Example 7.3...

Find the mean, variance, and **standard deviation** for the population of the number of color televisions per household... (from Example 7.1)

# of Televisions	# of Households	x	P(x)
0	1,218	0	0.012
1	32,379	1	0.319
2	37,961	2	0.374
3	19,387	3	0.191
4	7,714	4	0.076
5	2,842	5	0.028
	101,501		1.000

$$\sigma = \sqrt{\sigma^2} = \sqrt{1.107} = \mathbf{1.052}$$

7.18

Laws of Expected Value...

$$E(c) = c$$

The expected value of a constant (c) is just the value of the constant.

$$E(X + c) = E(X) + c$$

$$E(cX) = cE(X)$$

We can “pull” a constant out of the expected value expression (either as part of a sum with a random variable X or as a coefficient of random variable X).

7.19

Example 7.4...

Monthly sales have a mean of \$25,000 and a standard deviation of \$4,000. Profits are calculated by multiplying sales by 30% and subtracting fixed costs of \$6,000.

Find the **mean** monthly profit.

1) Describe the problem statement in algebraic terms:

sales have a mean of \$25,000 → $E(\text{Sales}) = 25,000$

profits are calculated by... → $\text{Profit} = .30(\text{Sales}) - 6,000$

7.20

Example 7.4...

Monthly sales have a mean of \$25,000 and a standard deviation of \$4,000. Profits are calculated by multiplying sales by 30% and subtracting fixed costs of \$6,000.

*Find the **mean** monthly profit.*

$$\begin{aligned} E(\text{Profit}) &= E[.30(\text{Sales}) - 6,000] \\ &= E[.30(\text{Sales})] - 6,000 && \text{[by rule \#2]} \\ &= .30E(\text{Sales}) - 6,000 && \text{[by rule \#3]} \\ &= .30(25,000) - 6,000 = 1,500 \end{aligned}$$

Thus, the mean monthly profit is **\$1,500**

7.21

Laws of Variance...

$$V(c) = 0$$

The variance of a constant (c) is zero.

$$V(X + c) = V(X)$$

The variance of a random variable and a constant is just the variance of the random variable (per 1 above).

$$V(cX) = c^2V(X)$$

The variance of a random variable and a constant coefficient is the coefficient squared times the variance of the random variable.

7.22

Example 7.4...

Monthly sales have a mean of \$25,000 and a standard deviation of \$4,000. Profits are calculated by multiplying sales by 30% and subtracting fixed costs of \$6,000.

Find the **standard deviation** of monthly profits.

1) Describe the problem statement in algebraic terms:

sales have a standard deviation of \$4,000

$$\rightarrow V(\text{Sales}) = 4,000^2 = 16,000,000$$

(remember the relationship between standard deviation and variance $\sigma = \sqrt{\sigma^2}$)

profits are calculated by... $\rightarrow \text{Profit} = .30(\text{Sales}) - 6,000$

7.23

Example 7.4...

Monthly sales have a mean of \$25,000 and a standard deviation of \$4,000. Profits are calculated by multiplying sales by 30% and subtracting fixed costs of \$6,000.

Find the **standard deviation** of monthly profits.

2) The **variance** of profit is = $V(\text{Profit})$

$$=V[.30(\text{Sales}) - 6,000]$$

$$=V[.30(\text{Sales})] \quad [\text{by rule \#2}]$$

$$=(.30)^2 V(\text{Sales}) \quad [\text{by rule \#3}]$$

$$=(.30)^2 (16,000,000) = 1,440,000$$

Again, **standard deviation** is the square root of **variance**,
so standard deviation of Profit = $(1,440,000)^{1/2} = \mathbf{\$1,200}$

7.24

Example 7.4 (summary)

Monthly sales have a mean of \$25,000 and a standard deviation of \$4,000. Profits are calculated by multiplying sales by 30% and subtracting fixed costs of \$6,000.

*Find the **mean** and **standard deviation** of monthly profits.*

The **mean** monthly profit is **\$1,500**

The **standard deviation** of monthly profit is **\$1,200**

7.25

Bivariate Distributions...

Up to now, we have looked at *univariate distributions*, i.e. probability distributions in **one** variable.

As you might guess, *bivariate distributions* are probabilities of combinations of **two** variables.

Bivariate probability distributions are also called *joint probability*. A joint probability distribution of X and Y is a table or formula that lists the joint probabilities for all *pairs* of values x and y, and is denoted P(x,y).

$$P(x,y) = P(X=x \text{ and } Y=y)$$

7.26

Discrete Bivariate Distribution...

As you might expect, the requirements for a bivariate distribution are similar to a univariate distribution, with only minor changes to the notation:

$$1. 0 \leq P(x,y) \leq 1$$

$$2. \sum_{all\ x} \sum_{all\ y} P(x,y) = 1$$

for all pairs (x,y).

7.27

Example 7.5...

Xavier and Yvette are real estate agents. Let X denote the number of houses that Xavier will sell in a month and let Y denote the number of houses Yvette will sell in a month. An analysis of their past monthly performances has the following joint probabilities (bivariate probability distribution).

		x			
		0	1	2	
y	0	0.12	0.42	0.06	0.6
	1	0.21	0.06	0.03	0.3
	2	0.07	0.02	0.01	0.1
		0.4	0.5	0.1	1.00

7.28

Marginal Probabilities...

As before, we can calculate the *marginal probabilities* by summing across rows and down columns to determine the probabilities of X and Y individually:

		x			
		0	1	2	
y	0	0.12	0.42	0.06	0.6
	1	0.21	0.06	0.03	0.3
	2	0.07	0.02	0.01	0.1
		0.4	0.5	0.1	1.00

x	P(x)
0	0.4
1	0.5
2	0.1

y	P(y)
0	0.6
1	0.3
2	0.1

E.g the probability that Xavier sells 1 house = $P(X=1) = 0.50$

7.29

Describing the Bivariate Distribution...

We can describe the mean, variance, and standard deviation of *each variable* in a bivariate distribution by working with the *marginal probabilities*...

x	P(x)
0	0.4
1	0.5
2	0.1

$E(X) =$	0.7
$V(X) =$	0.41
$\sigma_X =$	0.64

y	P(y)
0	0.6
1	0.3
2	0.1

$E(Y) =$	0.5
$V(Y) =$	0.45
$\sigma_Y =$	0.67

same formulae as for univariate distributions...

7.30

Covariance...

The *covariance* of two discrete variables is defined as:

$$COV(X,Y) = \sum_{all\ x} \sum_{all\ y} (x - \mu_x)(y - \mu_y)P(x,y)$$

or alternatively using this shortcut method:

$$COV(X,Y) = \sum_{all\ x} \sum_{all\ y} xyP(x,y) - \mu_x\mu_y$$

7.31

Coefficient of Correlation...

The coefficient of correlation is calculated in the same way as described earlier...

$$\rho = \frac{COV(X,Y)}{\sigma_x\sigma_y}$$

7.32

Example 7.6...

Compute the *covariance* and the *coefficient of correlation* between the numbers of houses sold by Xavier and Yvette.

$$COV(X,Y) = \sum_{all\ x} \sum_{all\ y} (x - \mu_x)(y - \mu_y)P(x,y)$$

$$COV(X,Y) = (0 - .7)(0 - .5)(.12) + (1 - .7)(0 - .5)(.42) + \dots \\ \dots + (2 - .7)(2 - .5)(.01) = \mathbf{-0.15}$$

$$\rho = \frac{COV(X,Y)}{\sigma_x \sigma_y} = -0.15 \div [(.64)(.67)] = \mathbf{-.35}$$

There is a weak, negative relationship between the two variables.

7.33

Example 7.6...

<u>X</u>	<u>Y</u>	<u>Probability</u>	<u>X - μ(x)</u>	<u>Y - μ(y)</u>	<u>[X - μ(x)][Y - μ(y)]</u>
0	0	0.12	-0.7	-0.5	0.042
0	1	0.21	-0.7	0.5	-0.074
0	2	0.07	-0.7	1.5	-0.074
1	0	0.42	0.3	-0.5	-0.063
1	1	0.06	0.3	0.5	0.009
1	2	0.02	0.3	1.5	0.009
2	0	0.06	1.3	-0.5	-0.039
2	1	0.03	1.3	0.5	0.020
<u>2</u>	<u>2</u>	0.01	1.3	1.5	<u>0.020</u>
0.7	0.5				-0.150

7.34

Example 7.6...

$$\rho = \frac{COV(X,Y)}{\sigma_x \sigma_y} = -0.150 \div [(.64)(.67)] = \textcolor{red}{-.35}$$

There is a weak, negative relationship between the two variables.

7.35

Sum of Two Variables...

The bivariate distribution allows us to develop the probability distribution of **any** combination of the two variables, of particular interest is the **sum** of two variables.

If we consider our example of Xavier and Yvette selling houses, we can create a probability distribution...

$x + y$	0	1	2	3	4
$P(x+y)$	0.12	0.63	0.19	0.05	0.01

...to answer questions like “what is the probability that two houses are sold”?

$$P(X+Y=2) = P(0,2) + P(1,1) + P(2,0) = .07 + .06 + .06 = .19$$

7.36

Sum of Two Variables...

Likewise, we can compute the expected value, variance, and standard deviation of $X+Y$ in the usual way...

$$E(X + Y) = 0(.12) + 1(.63) + 2(.19) + 3(.05) + 4(.01) = 1.2$$

$$V(X + Y) = (0 - 1.2)^2(.12) + \dots + (4 - 1.2)^2(.01) = .56$$

$$\sigma_{x+y} = \sqrt{\text{Var}(X + Y)} = \sqrt{.56} = .75$$

7.37

Laws...

We can derive laws of expected value and variance for the sum of two variables as follows...

$$E(X + Y) = E(X) + E(Y)$$

$$V(X + Y) = V(X) + V(Y) + 2\text{COV}(X, Y)$$

If X and Y are independent, $\text{COV}(X, Y) = 0$ and thus

$$V(X + Y) = V(X) + V(Y)$$

7.38

Laws

$$E(X + Y) = E(X) + E(Y) = .7 + .5 = 1.2$$

$$V(X + Y) = V(X) + V(Y) + 2\text{COV}(X, Y)$$

$$= .41 + .45 + 2(-.15) = .56$$

7.39

Portfolio Diversification and Asset Allocation

Consider an investor who forms a portfolio, consisting of only two stocks, by investing \$4,000 in one stock and \$6,000 in a second stock. Suppose that the results after 1 year are:

One-Year Results

Stock	Initial Investment	Value of Investment After One Year	Rate of Return on Investment
1	\$4,000	\$5,000	$R_1 = .25$ (25%)
2	\$6,000	\$5,400	$R_2 = -.10$ (-10%)
Total	\$10,000	\$10,400	$R_p = .04$ (4%)

OR

$$R_p = w_1R_1 + w_2R_2 = (.4)(.25) + (.6)(-.10) = .04$$

7.40

Portfolio Diversification and Asset Allocation

Mean and Variance of a Portfolio of Two Stocks

$$E(R_p) = w_1 E(R_1) + w_2 E(R_2)$$

$$\begin{aligned} V(R_p) &= w_1^2 V(R_1) + w_2^2 V(R_2) + 2w_1w_2 \text{COV}(R_1, R_2) \\ &= w_1^2 \sigma_1^2 + w_2^2 \sigma_2^2 + 2w_1w_2 \rho \sigma_1 \sigma_2 \end{aligned}$$

where w_1 and w_2 are the proportions or weights of investments 1 and 2, $E(R_1)$ and $E(R_2)$ are their expected values, σ_1 and σ_2 are their standard deviations, and ρ is the coefficient of correlation

7.41

Example 7.8

An investor has decided to form a portfolio by putting 25% of his money into McDonald's stock and 75% into Cisco Systems stock. The investor assumes that the expected returns will be 8% and 15%, respectively, and that the standard deviations will be 12% and 22%, respectively.

a Find the expected return on the portfolio.

b Compute the standard deviation of the returns on the portfolio assuming that

- (i) the two stocks' returns are perfectly positively correlated
- (ii) the coefficient of correlation is .5
- (iii) the two stocks' returns are uncorrelated

7.42

Example 7.8 Solution

a The expected values of the two stocks are

$$E(R_1) = .08 \text{ and } E(R_2) = .15$$

The weights are $w_1 = .25$ and $w_2 = .75$.

Thus,

$$\begin{aligned} E(R_p) &= w_1 E(R_1) + w_2 E(R_2) \\ &= .25(.08) + .75(.15) \\ &= .1325 \end{aligned}$$

7.43

Example 7.8 Solution

The standard deviations are $\sigma_1 = .12$ and $\sigma_2 = .22$. Thus,

$$\begin{aligned} V(R_p) &= w_1^2 \sigma_1^2 + w_2^2 \sigma_2^2 + 2w_1 w_2 \rho \sigma_1 \sigma_2 \\ &= (.25^2)(.12^2) + (.75^2)(.22^2) + 2(.25)(.75)\rho (.12)(.22) \\ &= .0281 + .0099 \rho \end{aligned}$$

When $\rho = 1$

$$V(R_p) = .0281 + .0099(1) = .0380$$

When $\rho = .5$

$$V(R_p) = .0281 + .0099(.5) = .0331$$

When $\rho = 0$

$$V(R_p) = .0281 + .0099(0) = .0281$$

7.44

Portfolio Diversification in Practice

The formulas introduced in this section require that we know the expected values, variances, and covariance (or coefficient of correlation) of the investments we're interested in. The question arises, How do we determine these parameters? (Incidentally, this question is rarely addressed in finance textbooks!) The most common procedure is to estimate the parameters from historical data, using sample statistics.

7.45

Portfolios with More Than Two Stocks

We can extend the formulas that describe the mean and variance of the returns of a portfolio of two stocks to a portfolio of any number of stocks.

Mean and Variance of a Portfolio of k Stocks

$$E(R_p) = \sum_{i=1}^k w_i E(R_i)$$

$$V(R_p) = \sum_{i=1}^k w_i^2 \sigma_i^2 + 2 \sum_{i=1}^k \sum_{j=i+1}^k w_i w_j \text{COV}(R_i, R_j)$$

Where R_i is the return of the i th stock, w_i is the proportion of the portfolio invested in stock i , and k is the number of stocks in the portfolio.

7.46

Portfolios with More Than Two Stocks

When k is greater than 2 the calculations can be tedious and time-consuming. For example, when $k = 3$, we need to know the values of the three weights, three expected values, three variances, and three covariances. When $k = 4$, there are four expected values, four variances and six covariances. [The number of covariances required in general is $k(k-1)/2$.] To assist you we have created an Excel worksheet to perform the computations when $k = 2, 3$, or 4. (For larger values of k , see the reference at the end of the chapter.) To demonstrate we'll return to the problem described in this chapter's introduction.

7.47

Binomial Distribution...

The ***binomial distribution*** is the probability distribution that results from doing a “***binomial experiment***”. Binomial experiments have the following properties:

Fixed number of trials, represented as ***n***.

Each trial has two possible outcomes, a “success” and a “failure”.

$P(\text{success})=p$ (and thus: $P(\text{failure})=1-p$), for all trials.

The trials are ***independent***, which means that the outcome of one trial does not affect the outcomes of any other trials.

7.48

Success and Failure...

...are just labels for a binomial experiment, there is no value judgment implied.

For example a coin flip will result in either heads or tails. If we define “heads” as success then necessarily “tails” is considered a failure (inasmuch as we attempting to have the coin lands heads up).

Other binomial experiment notions:

An election candidate wins or loses

An employee is male or female

7.49

Binomial Random Variable...

The random variable of a binomial experiment is defined as the number of successes in the n trials, and is called the ***binomial random variable***.

E.g. flip a fair coin 10 times...

- 1) Fixed number of trials ✓ $n=10$
- 2) Each trial has two possible outcomes ✓ {heads (success), tails (failure)}
- 3) $P(\text{success})=0.50$; $P(\text{failure})=1-0.50=0.50$ ✓
- 4) The trials are independent ✓ (i.e. the outcome of heads on the first flip will have no impact on subsequent coin flips).

Hence flipping a coin ten times is a binomial experiment since all conditions were met.

7.50

Binomial Random Variable...

The binomial random variable *counts* the number of successes in **n** trials of the binomial experiment. It can take on values from 0, 1, 2, ..., **n**. Thus, its a *discrete* random variable.

To calculate the probability associated with each value we use combintorics:

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

7.51

Example 7.9 Pat Statsdud...

Pat Statsdud is a (not good) student taking a statistics course. Pat's exam strategy is to rely on luck for the next quiz. The quiz consists of 10 multiple-choice questions. Each question has five possible answers, only one of which is correct. Pat plans to guess the answer to each question.

What is the probability that Pat gets no answers correct?

What is the probability that Pat gets two answers correct?

7.52

Pat Statsdud...

Pat Statsdud is a (not good) student taking a statistics course whose exam strategy is to rely on luck for the next quiz. The quiz consists of 10 multiple-choice questions. Each question has five possible answers, only one of which is correct. Pat plans to guess the answer to each question.

Algebraically then: $n=10$, and $P(\text{success}) = 1/5 = .20$

7.53

Pat Statsdud...

Is this a binomial experiment? Check the conditions:

- ✓ There is a fixed finite number of trials ($n=10$).
- ✓ An answer can be either correct or incorrect.

The probability of a correct answer ($P(\text{success})=.20$) does not change from question to question.

- ✓ Each answer is independent of the others.

7.54

Pat Statsdud...

n=10, and **P(success) = .20**

What is the probability that Pat gets **no answers** correct?

I.e. # success, x , = 0; hence we want to know $P(x=0)$

$$\begin{aligned}P(0) &= \frac{10!}{0!(10-0)!} (.2)^0 (1-.2)^{10-0} \\ &= 1(1)(.8)^{10} = .1074\end{aligned}$$

Pat has about an 11% chance of getting no answers correct using the guessing strategy.

7.55

Pat Statsdud...

n=10, and **P(success) = .20**

What is the probability that Pat gets **two answers** correct?

I.e. # success, x , = 2; hence we want to know $P(x=2)$

$$\begin{aligned}P(2) &= \frac{10!}{2!(10-2)!} (.2)^2 (1-.2)^{10-2} \\ &= 45(.04)(.1678) = .3020\end{aligned}$$

Pat has about a 30% chance of getting exactly two answers correct using the guessing strategy.

7.56

Cumulative Probability...

Thus far, we have been using the binomial probability distribution to find probabilities for individual values of x . To answer the question:

“Find the probability that Pat fails the quiz”

requires a *cumulative probability*, that is, $P(X \leq x)$

If a grade on the quiz is less than 50% (i.e. 5 questions out of 10), that's considered a failed quiz.

Thus, we want to know what is: $P(X \leq 4)$ to answer

7.57

Pat Statsdud...

$$P(X \leq 4) = P(0) + P(1) + P(2) + P(3) + P(4)$$

We already know $P(0) = .1074$ and $P(2) = .3020$. Using the binomial formula to calculate the others:

$$P(1) = .2684, P(3) = .2013, \text{ and } P(4) = .0881$$

$$\text{We have } P(X \leq 4) = .1074 + .2684 + \dots + .0881 = .9672$$

Thus, its about 97% probable that Pat will fail the test using the luck strategy and guessing at answers...

7.58

Binomial Table...

Calculating binomial probabilities by hand is tedious and error prone. There is an easier way. Refer to [Table 1](#) in Appendix B. For the Pat Statsdud example, **n=10**, so the first important step is to get the correct table!

n = 10		0.01	0.05	0.1	0.2	0.25	0.3	0.4	0.5	0.6	0.7	0.75	0.8	0.9	0.95	0.99
k																
0		0.9044	0.5987	0.3487	0.1074	0.0563	0.0282	0.0060	0.0010	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
1		0.9957	0.9139	0.7361	0.3758	0.2440	0.1493	0.0464	0.0107	0.0017	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000
2		0.9999	0.9885	0.9298	0.6778	0.5256	0.3828	0.1673	0.0547	0.0123	0.0016	0.0004	0.0001	0.0000	0.0000	0.0000
3		1.0000	0.9990	0.9872	0.8791	0.7759	0.6496	0.3823	0.1719	0.0548	0.0106	0.0035	0.0009	0.0000	0.0000	0.0000
4		1.0000	0.9999	0.9984	0.9672	0.9219	0.8497	0.6331	0.3770	0.1662	0.0473	0.0197	0.0064	0.0001	0.0000	0.0000
5		1.0000	1.0000	0.9999	0.9936	0.9803	0.9527	0.8338	0.6230	0.3669	0.1503	0.0781	0.0328	0.0016	0.0001	0.0000
6		1.0000	1.0000	1.0000	0.9991	0.9965	0.9894	0.9452	0.8281	0.6177	0.3504	0.2241	0.1209	0.0128	0.0010	0.0000
7		1.0000	1.0000	1.0000	0.9999	0.9996	0.9984	0.9877	0.9453	0.8327	0.6172	0.4744	0.3222	0.0702	0.0115	0.0001
8		1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9983	0.9893	0.9536	0.8507	0.7560	0.6242	0.2639	0.0861	0.0043
9		1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9990	0.9940	0.9718	0.9437	0.8926	0.6513	0.4013	0.0956

7.59

Binomial Table...

The probabilities listed in the tables are *cumulative*, i.e. $P(X \leq k)$ – k is the row index; the columns of the table are organized by $P(\text{success}) = p$

n = 10		0.01	0.05	0.1	0.2	0.25	0.3	0.4	0.5	0.6	0.7	0.75	0.8	0.9	0.95	0.99
k																
0		0.9044	0.5987	0.3487	0.1074	0.0563	0.0282	0.0060	0.0010	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
1		0.9957	0.9139	0.7361	0.3758	0.2440	0.1493	0.0464	0.0107	0.0017	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000
2		0.9999	0.9885	0.9298	0.6778	0.5256	0.3828	0.1673	0.0547	0.0123	0.0016	0.0004	0.0001	0.0000	0.0000	0.0000
3		1.0000	0.9990	0.9872	0.8791	0.7759	0.6496	0.3823	0.1719	0.0548	0.0106	0.0035	0.0009	0.0000	0.0000	0.0000
4		1.0000	0.9999	0.9984	0.9672	0.9219	0.8497	0.6331	0.3770	0.1662	0.0473	0.0197	0.0064	0.0001	0.0000	0.0000
5		1.0000	1.0000	0.9999	0.9936	0.9803	0.9527	0.8338	0.6230	0.3669	0.1503	0.0781	0.0328	0.0016	0.0001	0.0000
6		1.0000	1.0000	1.0000	0.9991	0.9965	0.9894	0.9452	0.8281	0.6177	0.3504	0.2241	0.1209	0.0128	0.0010	0.0000
7		1.0000	1.0000	1.0000	0.9999	0.9996	0.9984	0.9877	0.9453	0.8327	0.6172	0.4744	0.3222	0.0702	0.0115	0.0001
8		1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9983	0.9893	0.9536	0.8507	0.7560	0.6242	0.2639	0.0861	0.0043
9		1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9990	0.9940	0.9718	0.9437	0.8926	0.6513	0.4013	0.0956

7.60

Binomial Table...

“What is the probability that Pat gets *no answers* correct?”

i.e. what is $P(X = 0)$, given $P(\text{success}) = .20$ and $n=10$?

n = 10																	
k	0.01	0.05	0.1	0.2	0.25	0.3	0.4	0.5	0.6	0.7	0.75	0.8	0.9	0.95	0.99		
0	0.9044	0.5987	0.3487	0.1074	0.0563	0.0282	0.0060	0.0010	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000		
1	0.9957	0.9139	0.7361	0.3758	0.2440	0.1493	0.0464	0.0107	0.0017	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000		
2	0.9999	0.9885	0.9298	0.6778	0.5256	0.3828	0.1673	0.0547	0.0123	0.0016	0.0004	0.0001	0.0000	0.0000	0.0000		
3	1.0000	0.9990	0.9872	0.8791	0.7759	0.6496	0.3823	0.1719	0.0548	0.0106	0.0035	0.0009	0.0000	0.0000	0.0000		
4	1.0000	0.9999	0.9984	0.9672	0.9219	0.8497	0.6331	0.3770	0.1662	0.0473	0.0197	0.0064	0.0001	0.0000	0.0000		
5	1.0000	1.0000	0.9999	0.9936	0.9803	0.9527	0.8338	0.6230	0.3669	0.1503	0.0781	0.0328	0.0016	0.0001	0.0000		
6	1.0000	1.0000	1.0000	0.9991	0.9965	0.9894	0.9452	0.8281	0.6177	0.3504	0.2241	0.1209	0.0128	0.0010	0.0000		
7	1.0000	1.0000	1.0000	0.9999	0.9996	0.9984	0.9877	0.9453	0.8327	0.6172	0.4744	0.3222	0.0702	0.0115	0.0001		
8	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9983	0.9893	0.9536	0.8507	0.7560	0.6242	0.2639	0.0861	0.0043		
9	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9990	0.9940	0.9718	0.9437	0.8926	0.6513	0.4013	0.0956		

$$P(X = 0) = P(X \leq 0) = .1074$$

7.61

Binomial Table...

“What is the probability that Pat gets *two answers* correct?”

i.e. what is $P(X = 2)$, given $P(\text{success}) = .20$ and $n=10$?

n = 10																	
k	0.01	0.05	0.1	0.2	0.25	0.3	0.4	0.5	0.6	0.7	0.75	0.8	0.9	0.95	0.99		
0	0.9044	0.5987	0.3487	0.1074	0.0563	0.0282	0.0060	0.0010	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000		
1	0.9957	0.9139	0.7361	0.3758	0.2440	0.1493	0.0464	0.0107	0.0017	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000		
2	0.9999	0.9885	0.9298	0.6778	0.5256	0.3828	0.1673	0.0547	0.0123	0.0016	0.0004	0.0001	0.0000	0.0000	0.0000		
3	1.0000	0.9990	0.9872	0.8791	0.7759	0.6496	0.3823	0.1719	0.0548	0.0106	0.0035	0.0009	0.0000	0.0000	0.0000		
4	1.0000	0.9999	0.9984	0.9672	0.9219	0.8497	0.6331	0.3770	0.1662	0.0473	0.0197	0.0064	0.0001	0.0000	0.0000		
5	1.0000	1.0000	0.9999	0.9936	0.9803	0.9527	0.8338	0.6230	0.3669	0.1503	0.0781	0.0328	0.0016	0.0001	0.0000		
6	1.0000	1.0000	1.0000	0.9991	0.9965	0.9894	0.9452	0.8281	0.6177	0.3504	0.2241	0.1209	0.0128	0.0010	0.0000		
7	1.0000	1.0000	1.0000	0.9999	0.9996	0.9984	0.9877	0.9453	0.8327	0.6172	0.4744	0.3222	0.0702	0.0115	0.0001		
8	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9983	0.9893	0.9536	0.8507	0.7560	0.6242	0.2639	0.0861	0.0043		
9	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9990	0.9940	0.9718	0.9437	0.8926	0.6513	0.4013	0.0956		

$$P(X = 2) = P(X \leq 2) - P(X \leq 1) = .6778 - .3758 = .3020$$

remember, the table shows *cumulative* probabilities...

7.62

Binomial Distribution...

What is the probability that Pat fails the quiz”?

i.e. what is $P(X \leq 4)$, given $P(\text{success}) = .20$ and $n=10$?

n=10															
k	0.01	0.05	0.1	0.2	0.25	0.3	0.4	0.5	0.6	0.7	0.75	0.8	0.9	0.95	0.99
0	0.9044	0.5987	0.3487	0.1074	0.0563	0.0282	0.0060	0.0010	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
1	0.9957	0.9139	0.7361	0.3758	0.2440	0.1493	0.0464	0.0107	0.0017	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000
2	0.9999	0.9885	0.9298	0.6778	0.5256	0.3828	0.1673	0.0547	0.0123	0.0016	0.0004	0.0001	0.0000	0.0000	0.0000
3	1.0000	0.9990	0.9872	0.8791	0.7759	0.6496	0.3823	0.1719	0.0548	0.0106	0.0035	0.0009	0.0000	0.0000	0.0000
4	1.0000	0.9999	0.9984	0.9672	0.9219	0.8497	0.6331	0.3770	0.1662	0.0473	0.0197	0.0064	0.0001	0.0000	0.0000
5	1.0000	1.0000	0.9999	0.9936	0.9803	0.9527	0.8338	0.6230	0.3669	0.1503	0.0781	0.0328	0.0016	0.0001	0.0000
6	1.0000	1.0000	1.0000	0.9991	0.9965	0.9894	0.9452	0.8281	0.6177	0.3504	0.2241	0.1209	0.0128	0.0010	0.0000
7	1.0000	1.0000	1.0000	0.9999	0.9996	0.9984	0.9877	0.9453	0.8327	0.6172	0.4744	0.3222	0.0702	0.0115	0.0001
8	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9983	0.9893	0.9536	0.8507	0.7560	0.6242	0.2639	0.0861	0.0043
9	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9990	0.9940	0.9718	0.9437	0.8926	0.6513	0.4013	0.0956

$$P(X \leq 4) = .9672$$

7.63

Binomial Table...

The binomial table gives cumulative probabilities for $P(X \leq k)$, but as we've seen in the last example,

$$P(X = k) = P(X \leq k) - P(X \leq [k-1])$$

Likewise, for probabilities given as $P(X \geq k)$, we have:

$$P(X \geq k) = 1 - P(X \leq [k-1])$$

7.64

Binomial Distribution...

As you might expect, statisticians have developed general formulas for the mean, variance, and standard deviation of a binomial random variable. They are:

$$\mu = np$$

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

$$\sigma = \sqrt{np(1 - p)}$$

7.65

Poisson Distribution...

Named for Simeon Poisson, the ***Poisson distribution*** is a discrete probability distribution and refers to the number of events (a.k.a. successes) within a specific time period or region of space. For example:

The number of cars arriving at a service station in 1 hour. (The interval of time is 1 hour.)

The number of flaws in a bolt of cloth. (The specific region is a bolt of cloth.)

The number of accidents in 1 day on a particular stretch of highway. (The interval is defined by both time, 1 day, and space, the particular stretch of highway.)

7.66

The Poisson Experiment...

Like a binomial experiment, a ***Poisson experiment*** has four defining characteristic properties:

The number of successes that occur in any interval is independent of the number of successes that occur in any other interval.

The probability of a success in an interval is the same for all equal-size intervals

The probability of a success is proportional to the size of the interval.

The probability of more than one success in an interval approaches 0 as the interval becomes smaller.

7.67

Poisson Distribution...

The ***Poisson random variable*** is the number of successes that occur in a period of time or an interval of space in a Poisson experiment.

E.g. On average, 96 trucks arrive at a border crossing every hour.

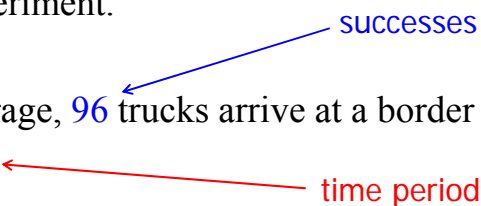


Diagram labels: successes (blue arrow pointing to 96), time period (red arrow pointing to every hour).

E.g. The number of typographic errors in a new textbook edition averages 1.5 per 100 pages.

successes (!)



Diagram label: successes (!) (blue arrow pointing to 1.5)

interval



Diagram label: interval (red arrow pointing to 100 pages)

7.68

Poisson Probability Distribution...

The probability that a Poisson random variable assumes a value of x is given by:

$$P(x) = \frac{e^{-\mu} \mu^x}{x!} \quad \text{for } x = 0, 1, 2, \dots$$

where μ is the mean number of successes in the interval
and e is the natural logarithm base.

FYI: $E(X) = V(X) = \mu$

7.69

Example 7.12...

A statistics instructor has observed that the number of typographical errors in new editions of textbooks varies considerably from book to book. After some analysis he concludes that the number of errors is Poisson distributed with a mean of 1.5 per 100 pages. The instructor randomly selects 100 pages of a new book. What is the probability that there are no typos?

7.70

Example 7.12...

A statistics instructor has observed that the number of typographical errors in new editions of textbooks varies considerably from book to book. After some analysis he concludes that the number of errors is Poisson distributed with a mean of 1.5 per 100 pages. The instructor randomly selects 100 pages of a new book. What is the probability that there are no typos? That is, what is $P(X=0)$ given that $\mu = 1.5$?

$$P(0) = \frac{e^{-\mu} \mu^x}{x!} = \frac{e^{-1.5} 1.5^0}{0!} = .2231$$

“There is about a 22% chance of finding zero errors”

7.71

Example 7.13...

Refer to Example 7.12. Suppose that the instructor has just received a copy of a new statistics book. He notices that there are 400 pages.

- a What is the probability that there are no typos?
- b What is the probability that there are five or fewer typos?

7.72

Poisson Distribution...

As mentioned on the Poisson experiment slide:

The probability of a success is proportional to the size of the interval

Thus, knowing an error rate of 1.5 typos per 100 pages, we can determine a mean value for a 400 page book as:

$$\mu = 1.5(4) = 6 \text{ typos / 400 pages.}$$

7.73

Example 7.13...

For a 400 page book, what is the probability that there are **no** typos?

$$P(X=0) = P(0) = \frac{e^{-\mu} \mu^x}{x!} = \frac{e^{-6} 6^0}{0!} = .002479$$

“there is a very small chance there are no typos”

7.74

Example 7.13...

For a 400 page book, what is the probability that there are **five or less** typos?

$$P(X \leq 5) = P(0) + P(1) + \dots + P(5) = \frac{e^{-6}6^0}{0!} + \frac{e^{-6}6^1}{1!} + \dots + \frac{e^{-6}6^5}{5!}$$

This is rather tedious to solve manually. A better alternative is to refer to [Table 2](#) in Appendix B...

... $k=5$, $\mu=6$, and $P(X \leq k) = .446$

“there is about a 45% chance there are 5 or less typos”