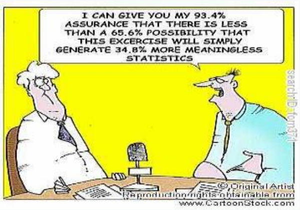
Chapter 7: Random Variables and Discrete Probability Distributions

Chapter 7: Random Variables and Discrete Probability Distributions



Introduction

In this chapter, we extend the concepts and techniques of probability by presenting random variables and probability distributions, which are essential in the development of statistical inference.

The concepts and techniques of probability introduced in this chapter will allow us to calculate the probability of observing a given experimental outcome before we move onto the more advanced concepts of bivariate and other discrete probability distributions.

2

Variables and Probability Distributions

Consider an experiment where we flip two balanced coins and observe the results:
- Heads on the first coin and heads on the second coin
- Heads on the first coin and tails on the second coin
- Tails on the first coin and heads on the second coin
- Tails on the first coin and tails on the second coin

We can list the events as: 2 heads 1 head, 1 tail 1 tail, 1 head 0 heads

The number of heads is called **random variable X**; X=0,1,2.

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

3

Discrete and Ransom Variables Examples

Discrete Random Variable – usually count data [Number of]

* one that takes on a *countable* number of values – this means you can sit down and list all possible outcomes without missing any, although it might take you an infinite amount of time.

X = values on the roll of two dice: X has to be either 2, 3, 4, …, or 12.

Y = number of customer at Starbucks during the day

Y has to be 0, 1, 2, 3, 4, 5, 6, 7, 8, ……………"real big number"

Continuous Random Variable – usually measurement data [time, weight, distance, etc]

* one that takes on an uncountable number of values – this means you can never list all possible outcomes even if you had an infinite amount of time.

X = time it takes you to walk home from class: X > 0, might be 5.1 minutes measured to the nearest tenth but in reality the actual time is 5.10000001…………………. minutes?)

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

- *P(X = x)* or simply *P(x)*.

4

Chapter 7: Random Variables and Discrete Probability Distributions

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.

These two conditions apply:

1. $0 \leq P(x) \leq 1$ for all x
2. $\sum_{all\ x} P(x) = 1$

Probability Distribution of Persons per Household

The following table, published in The Statistical Abstract of the United States, summarizes the number of persons living in a household.

Number of Persons	Millions of Households
1	35.2
2	43.5
3	19.5
4	16.2
5	7.3
6	2.8
7 or more	1.6
Total	126.1

Questions:

- Develop the probability distribution for the number of persons per household.
- What is the probability of a household with at least 4 persons?

Probability Distribution of Persons per Household

The probability of each value of *X*, the number of persons per household, is computed as the relative frequency.

Number of Persons	P(x)
1	35.2 / 126.1 = .279
2	43.5 / 126.1 = .345
3	19.5 / 126.1 = .155
4	16.2 / 126.1 = .128
5	7.3 / 126.1 = .058
6	2.8 / 126.1 = .022
7 or more	1.6 / 126.1 = .013
Total	126.1 / 126.1 = 1.000

Addition rule applies; the probability of a household with at least 4 persons:

$P(X \geq 4) = .128 + .058 + .022 + .013 = .221$

Describing the Population of the Number of Persons per Household

Find the mean, variance, and standard deviation for the population of the number of persons per household.

For this example, we will assume that the last category is exactly seven persons.

Number of Persons	P(x)
1	35.2 / 126.1 = .279
2	43.5 / 126.1 = .345
3	19.5 / 126.1 = .155
4	16.2 / 126.1 = .128
5	7.3 / 126.1 = .058
6	2.8 / 126.1 = .022
7 or more	1.6 / 126.1 = .013
Total	126.1 / 126.1 = 1.000

Chapter 7: Random Variables and Discrete Probability Distributions

Describing the Population/Probability Distribution

Because a discrete probability distribution represents a population, we can describe it by computing various parameters.

Population mean (expected value)

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

Population variance

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

Shortcut calculations for population variance:

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

Population standard deviation

$$\sigma = \sqrt{\sigma^2}$$

9

Describing the Population of the Number of Persons per Household

The mean of X is:

$$E(X) = \mu = \sum_{\text{all x}} xP(x) = 1P(1) + 2P(2) + \cdots + 7P(7)$$

$$= 1(.279) + 2(.345) + \cdots + 7(.013) = 2.46$$

The variance of X is:

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

$$= (1 - 2.46)^2(.279) + (2 - 2.46)^2(.345) + \cdots + (7 - 2.46)^2(.013) = 1.931$$

The standard deviation of X is:

$$\sigma = \sqrt{\sigma^2} = \sqrt{1.931} = 1.39$$

10

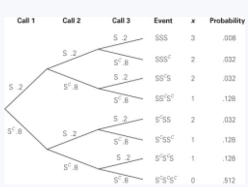
Probability Distribution of the Number of Sales

A mutual fund salesperson has arranged to call on **three people**. Based on past experience, the salesperson knows that there is a 20% chance of closing a sale on each call.

Determine the probability distribution of the number of sales the salesperson will make.

Let S denote closing a sale: $P(S) = .2$
Thus, S^C is not closing a sale. $P(S^C) = .8$

We can use the multiplication rule for independent events to develop a probability tree, and use it to calculate the **probability distribution** of making a number of X sales:



x	Events	P(X = x)
0	$S^CS^CS^C$	.512
1	SS^CS^C, S^CSS^C, S^CS^CS	.128 + .128 + .128 = .384
2	SSS^C, SS^CS, S^CSS	.032 + .032 + .032 = .096
3	SSS	.008

Where, for example:

for $X = 0$, $P(S^CS^CS^C) = P(S^C)\,P(S^C)\,P(S^C) = (.8)(.8)(.8) = .512$
for $X = 1$, $P(SS^CS^C) = P(S)\,P(S^C)\,P(S^C) = (.2)(.8)(.8) = .128$
for $X = 2$, $P(SSS^C) = P(S)\,P(S)\,P(S^C) = (.2)(.2)(.8) = .032$
for $X = 3$, $P(SSS) = P(S)\,P(S)\,P(S) = (.2)(.2)(.2) = .008$

11

The number of pizzas delivered to university students each month is a random variable with the following probability distribution.

x	0	1	2	3
$P(x)$	.1	.3	.4	.2

a. Find the probability that a student has received delivery of two or more pizzas this month.

b. Determine the mean and variance of the number of pizzas delivered to students each month.

12

The number of pizzas delivered to university students each month is a random variable with the following probability distribution.

x	0	1	2	3
P(x)	.1	.3	.4	.2

a. Find the probability that a student has received delivery of two or more pizzas this month.

b. Determine the mean and variance of the number of pizzas delivered to students each month.

Solutions:

a. $P(X \geq 2) = P(2) + P(3) = .4 + .2 = .6$

b. $\mu = E(X) = \sum xP(x) = 0(.1) + 1(.3) + 2(.4) + 3(.2) = 1.7$

$\sigma^2 = V(X) = \sum (x-\mu)^2 P(x) = (0-1.7)^2(.1) + (1-1.7)^2(.3) + (2-1.7)^2(.4) + (3-1.7)^2(.2) = .81.$

13

Grouped Data

- The weighted mean computation can be used to obtain approximations of the mean, variance, and standard deviation for the grouped data.
- To compute the weighted mean, we treat the midpoint of each class as though it were the mean of all items in the class.
- We compute a weighted mean of the class midpoints using the class frequencies as weights.
- Similarly, in computing the variance and standard deviation, the class frequencies are used as weights.

14

Mean for Grouped Data

- Sample Data

$$\bar{x} = \frac{\sum f_i M_i}{n}$$

- Population Data

$$\mu = \frac{\sum f_i M_i}{N}$$

where:

f_i = frequency of class i
M_i = midpoint of class i

15

Sample Mean for Grouped Data

- Example: Apartment Rents

Rent ($)	f_i	M_i	f_iM_i
420-439	8	429.5	3436.0
440-459	17	449.5	7641.5
460-479	12	469.5	5634.0
480-499	8	489.5	3916.0
500-519	7	509.5	3566.5
520-539	4	529.5	2118.0
540-559	2	549.5	1099.0
560-579	4	569.5	2278.0
580-599	2	589.5	1179.0
600-619	6	609.5	3657.0
Total	70		34525.0

$$\bar{x} = \frac{34{,}525}{70} = 493.21$$

This approximation differs by $2.41 from the actual sample mean of $490.80.

16

Variance for Grouped Data

- For sample data

$$s^2 = \frac{\sum f_i (M_i - \bar{x})^2}{n-1}$$

- For population data

$$\sigma^2 = \frac{\sum f_i (M_i - \mu)^2}{N}$$

17

Sample Variance for Grouped Data

- Example: Apartment Rents

Rent ($)	f_i	M_i	$M_i - x$	$(M_i - x)^2$	$f_i(M_i - x)^2$
420-439	8	429.5	-63.7	4058.96	32471.71
440-459	17	449.5	-43.7	1910.56	32479.59
460-479	12	469.5	-23.7	562.16	6745.97
480-499	8	489.5	-3.7	13.76	110.11
500-519	7	509.5	16.3	265.36	1857.55
520-539	4	529.5	36.3	1316.96	5267.86
540-559	2	549.5	56.3	3168.56	6337.13
560-579	4	569.5	76.3	5820.16	23280.66
580-599	2	589.5	96.3	9271.76	18543.53
600-619	6	609.5	116.3	13523.36	81140.18
Total	70				208234.29

continued →

18

Sample Variance for Grouped Data

- Example: Apartment Rents
  - Sample Variance

    $s^2 = 208{,}234.29/(70 - 1) = 3{,}017.89$

  - Sample Standard Deviation

    $s = \sqrt{3{,}017.89} = 54.94$

This approximation differs by only $.20 from the actual standard deviation of $54.74.

19

Finance Example of Expected Returns

- Expected returns are based on the probabilities of possible outcomes

$$E(R) = \sum_{i=1}^{n} p_i R_i$$

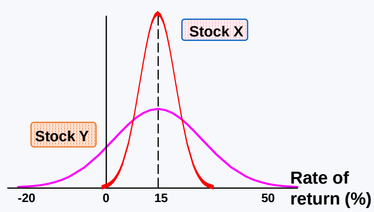
Where:

pi = the probability of state "i" occurring
Ri = the expected return on an asset in state i

20

Chapter 7: Random Variables and Discrete Probability Distributions

Probability distribution



- Which stock is riskier? Why?

21

Variance and Standard Deviation

- Variance and standard deviation measure the volatility of returns
- Variance = Weighted average of squared deviations
- Standard Deviation = Square root of variance

$$\boxed{\sigma^2 = \sum_{i=1}^{n} p_i (R_i - E(R))^2}$$

22

Example - Assume the Following Investment Alternatives

Economy	Prob.	T-Bill	Stock Al	Stock Bob	Stock Cuc	Market
Recession	0.10	8.0%	-22.0%	28.0%	10.0%	-13.0%
Below avg.	0.20	8.0	-2.0	14.7	-10.0	1.0
Average	0.40	8.0	20.0	0.0	7.0	15.0
Above avg.	0.20	8.0	35.0	-10.0	45.0	29.0
Boom	0.10	8.0	50.0	-20.0	30.0	43.0
	1.00					

23

Calculate the Expected Rate of Return on Each Alternative

$\hat{r}$ = expected rate of return.

$$\hat{r} = \sum_{i=1}^{n} r_i P_i.$$

$\hat{r}_{Al}$ = 0.10(-22%) + 0.20(-2%) + 0.40(20%) + 0.20(35%) + 0.10(50%) = 17.4%.

	$\hat{r}$
Market	15.0
Cuc	13.8
T-bill	8.0
Bob	1.7

Al has the highest rate of return. Does that make it best?

24

What is the Standard Deviation of Returns for Each Alternative?

$$\sigma = \text{Standard deviation}$$

$$\sigma = \sqrt{\text{Variance}} = \sqrt{\sigma^2}$$

$$= \sqrt{\sum_{i=1}^{n}\left(r_i - \hat{r}\right)^2 P_i}.$$

σ_{Al} = ((-22 - 17.4)^{2}0.10 + (-2 - 17.4)^{2}0.20 + (20 - 17.4)^{2}0.40 + (35 - 17.4)^{2}0.20 + (50 - 17.4)^{2}0.10)$^{1/2}$ = 20.0%.

$\sigma_{T-bills} = 0.0\%$
$\sigma_{Bob} = 13.4$
$\sigma_{Cuc} = 18.8$
$\sigma_{Market} = 15.3\%$

25

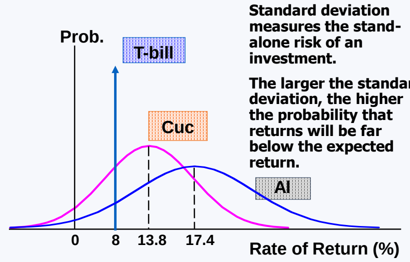
$$\sigma = \sqrt{\sum_{i=1}^{n}\left(r_i - \hat{r}\right)^2 P_i}.$$

Al:

σ = ((-22 - 17.4)^{2}0.10 + (-2 - 17.4)^{2}0.20 + (20 - 17.4)^{2}0.40 + (35 - 17.4)^{2}0.20 + (50 - 17.4)^{2}0.10)$^{1/2}$ = 20.0%.

$\sigma_{T\text{-bills}}$ = 0.0%.
σ_{Al} = 20.0%.
σ_{Bob} = 13.4%.
σ_{Cuc} = 18.8%.
σ_{Market} = 15.3%.

26



Standard deviation measures the stand-alone risk of an investment.

The larger the standard deviation, the higher the probability that returns will be far below the expected return.

An alternative measure of stand-alone risk is the Coefficient of Variation (CV). CV = Standard deviation/expected return

27

Expected Return versus Risk versus Coefficient of Variation

Security	Expected return	Risk, σ
Al	17.4%	20.0%
Market	15.0	15.3
Cuc	13.8	18.8
T-bills	8.0	0.0
Bob	1.7	13.4

$CV_{T\text{-bills}}$ = 0.0%/8.0% = 0.0.
CV_{Al} = 20.0%/17.4% = 1.1.
CV_{Bob} = 13.4%/1.7% = 7.9.
CV_{Cuc} = 18.8%/13.8% = 1.4.
CV_{Market} = 15.3%/15.0% = 1.0.

28

Chapter 7: Random Variables and Discrete Probability Distributions

Bivariate Distributions

A **bivariate distribution** provides joint probabilities of the combination of two variables. We refer to the probability distribution of one variable we saw previously as a *univariate* distribution.

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*). These two conditions apply to a discrete bivariate distribution:

1. $0 \le P(x,y) \le 1$ for all pair of values (x,y)
2. $\sum_{\text{all } x} \sum_{\text{all } y} P(x,y) = 1$

29

Bivariate Distribution of the Number of House Sales

Xavier and Yvette are real estate agents. Let *X* denote the number of houses that Xavier will sell in a month and *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	.12	.42	.06
	1	.21	.06	.03
	2	.07	.02	.01

As we did in Chapter 6, we can calculate the marginal probabilities by summing across rows or down columns.

The marginal probability distributions of X and Y are as follows:

x	*P(X=x)*
0	.12 + .21 + .07 = .4
1	.42 + .06 + .02 = .5
2	.06 + .03 + .01 = .1
y	*P(Y=y)*
0	.12 + .42 + .06 = .6
1	.21 + .06 + .03 = .3
2	.07 + .02 + .01 = .1

30

Example Describing the Bivariate Distribution of the Number of House Sales

As we did with the univariate distribution, we can describe the bivariate distribution in the example by computing the mean, variance, and standard deviation of **each variable** utilizing the respective **marginal probabilities**.

Marginal Distribution of *X*

$E(X) = \mu_x = \sum_{\text{all } x} xP(x) = 0(.4) + 1(.5) + 2(.1) = 0.7$

$V(X) = \sigma_x{}^2 = \sum_{all\, x} (x-\mu_x)^2 P(x) = (0-.7)^2(.4) + (1-.7)^2(.5) + (2-.7)^2(.1) = 0.41$

$\sigma_x = \sqrt{\sigma_x{}^2} = \sqrt{0.41} = 0.64$

Marginal Distribution of *Y*

$E(Y) = \mu_y = \sum_{\text{all } y} yP(y) = 0(.6) + 1(.3) + 2(.1) = 0.5$

$V(Y) = \sigma_y{}^2 = \sum_{all\, y} (y-\mu_y)^2 P(y) = (0-.5)^2(.6) + (1-.5)^2(.3) + (2-.5)^2(.1) = 0.45$

$\sigma_y = \sqrt{\sigma_y{}^2} = \sqrt{0.45} = 0.67$

31

7-2b Covariance and Coefficient of Correlation

The **covariance** and the **coefficient of correlation**, previously introduced in Chapter 4, describe the relationship between the two discrete variables of the bivariate distribution.

Covariance:

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

Shortcut calculation for population covariance:

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

Coefficient of correlation: $\rho = \frac{\sigma_{xy}}{\sigma_x \sigma_y}$

32

Chapter 7: Random Variables and Discrete Probability Distributions

Covariance and Coefficient of Correlation for the Number of House Sales

Calculation of the covariance using the shortcut method:

$$\mathrm{COV}(X,Y) = \sigma_{xy} = \sum_{\text{all x}}\sum_{\text{all } y} xyP(x,y) - \mu_x\mu_y$$
$$= (0)(0)(.12) + (1)(0)(.42) + +(2)(0)(.06) + (0)(1)(.21) + (1)(1)(.06)$$
$$+ +(2)(1)(.03) + (0)(2)(.07) + (1)(2)(.02) + +(2)(2)(.01) - (0.7)(0.5)$$
$$= -0.15$$

Given the previously calculated standard deviations for *X* and *Y*:

$$\rho = \frac{\sigma_{xy}}{\sigma_x\sigma_y} = \frac{-0.15}{(0.64)(0.67)} = -.35$$

There is a weak negative relationship between the two variables: the number of houses Xavier will sell in a month (*X*) and the number of houses Yvette will sell in a month (*Y*).

33

Sum of Two Variables

Of particular interest to us is **the sum of two variables**.

To develop the probability distribution of the sum of two variables, add together the joint probabilities corresponding to each total value of *X* + *Y*.

In the example, the possible values of *X* + *Y* are:

0, 1, 2, 3, and 4 houses.

Example: if we want to calculate the probability associated to the total sale of 2 houses, we write:

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

		X		
		0	1	2
Y	0	.12	.42	.06
	1	.21	.06	.03
	2	.07	.02	.01

34

Sum of Two Variables

We calculate the probabilities of the other values of *X* + *Y*, similarly, producing the following table:

x + y	Events	P(x + y)
0	P(0,0)	.12
1	P(0,1) + P(1,0)	.63
2	P(0,2) + P(1,1) + P(0,2)	.19
3	P(1,2) + P(2,1)	.05
4	P(2,2)	.01

We can now compute the parameters of *X* + *Y* as we learned to do for univariate distributions:

$$E(X+Y) = .63 + 2(.19) + 3(.05) + 4(.01) = 1.2$$
$$V(X+Y) = (0-1.2)^2(.12) + (1-1.2)^2(.63)$$
$$+(2-1.2)^2(.19) + (3-1.2)^2(.05) + (4-1.2)^2(.01) = 0.56$$

35

Laws of Expected Value and Variance of the Sum of Two Variables

1. $E(X+Y) = E(X) + E(Y)$
2. $V(X+Y) = V(X) + V(Y) + 2\,COV(X,Y)$

If *X* and *Y* are independent, $COV(X,Y) = 0$, thus:

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

Applying the formulas to the example, we obtain the same values shown for the distribution of the sum of two variables:

$$E(X+Y) = E(X) + E(Y) = 0.7 + 0.5 = 1.2$$
$$V(X+Y) = V(X) + V(Y) + 2\,COV(X,Y) = .41 + .45 + 2(-0.15) = 0.56$$
$$\sigma_{x+y} = \sqrt{V(X+Y)} = \sqrt{0.56} = .75$$

36

Chapter 7: Random Variables and Discrete Probability Distributions

Binomial Distribution

*We start defining different rules and formulas of probability distribution based on the **type & properties** of the variable that we are dealing with.*

The binomial distribution is the result of a *binomial experiment*, which has the following properties:

Binomial Experiment

1. The binomial experiment consists of a fixed number of n trials.
2. Each trial has two possible outcomes: a success and a failure.
3. The probability of success is p. The probability of failure is $1 - p$.
4. The trials are independent.

A **Bernoulli process** is a trial that satisfies properties 2-4.

The **binomial random variable** is the random variable of a binomial experiment. It is defined as the number of successes in the experiment's n trials.

37

Examples of Binomial Experiments

1. Flip a coin 10 times.
   - $n = 10$
   - Two outcomes: heads (success) and tails (failure).
   - If coin is fair: P(heads) = P(tails) = .5
   - Each coin toss is independent.
2. Draw five cards out of a shuffled deck.
   - $n = 5$
   - We label success any rank and/or suit of cards we seek, such as a five or clubs.
   - If success is a club, then P(club) = 13 / 52; if it is a rank of five, then P(five) = 1/13.
   - Each draw is independent only if we replace the drawn card in the deck and reshuffle each time.
3. A political survey of 1,500 voters about an upcoming election.
   - $n = 1{,}500$
   - If there are more than two candidates, the candidate of choice is a success, the other a failure.
   - The actual value of p is unknown. The task of the statistics practitioner is to estimate its value.
   - The trials are independent because the choice of one voter does not affect the choice of others.

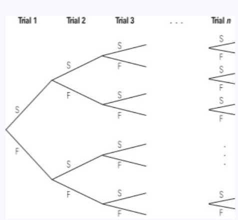
38

Binomial Random Variable

The binomial random variable is the number of successes X in the experiment's n trials. It can take on values 0, 1, 2, …, n. Thus, the random variable X is discrete.

We can draw the n trials of a binomial experiment with a probability tree. The stages represent the outcomes for each of the n trials. At each stage, there are two branches representing success (S) and failure (F).

To calculate the probability that there are X successes in n trials, we multiply together the probability of a success, p, or a failure, $1 - p$, for each stage of the sequence. And if there are X successes, there must be $n - X$ failures.



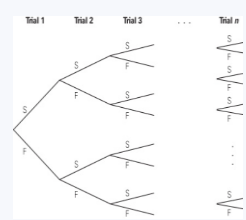
39

Binomial Random Variable

The binomial random variable is the number of successes X in the experiment's n trials. It can take on values 0, 1, 2, …, n. Thus, the random variable X is discrete.

We can draw the n trials of a binomial experiment with a probability tree. The stages represent the outcomes for each of the n trials. At each stage, there are two branches representing success (S) and failure (F).

To calculate the probability that there are X successes in n trials, we multiply together the probability of a success, p, or a failure, $1 - p$, for each stage of the sequence. And if there are X successes, there must be $n - X$ failures.



40

Chapter 7: Random Variables and Discrete Probability Distributions

Counting Rule for Combinations

This rule allows us to count the number of experimental outcomes when we select ***n*** objects from a (usually larger) set of ***N*** objects.

The number of *N* objects taken *n* at a time is

$$C_n^N = \binom{N}{n} = \frac{N!}{n!(N-n)!}$$

where

$$N! = N(N-1)(N-2)\ldots(2)(1)$$
$$n! = n(n-1)(n-2)\ldots(2)(1)$$

And by definition

$$0! = 1$$

41

7-4a Binomial Probability Distribution

We can write the probability for each sequence of branches that represents *x* successes with probability *p*, and *n* – *x* failures with probability 1 – *p* as:

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

The combinatorial formula yields the count of branch sequences that produces *x* successes and *n* – *x* failures:

$C_x^n = \frac{n!}{x!(n-x)!}$ Keep in mind that here n is total number of trials.

Combining the two components yields the **binomial probability distribution**:

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

Where n! = n(n-1)(n-2)…(2)(1).
For example, 3! = 3 x 2 x 1 = 6.
Incidentally, although it may not appear to be logical, 0! = 1.

42

Example Pat Statsdud and the Statistics Quiz

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.

a. What is the probability that Pat gets no answers correct?
b. What is the probability that Pat gets two answers correct

Solution:
This is a binomial experiment because:

1. *n* = 10
2. Two outcomes: correct and incorrect answer.
3. Probability of correct answer: *p* = 1/5 = .2.
4. Answers to questions are independent.

We can apply the binomial probability distribution to answer both questions:

a. *x* = 0: $P(0) = \frac{10!}{0!(10-0)!}.2^0.8^{(10-0)} = 1(1)(.8)^{10} =$ .1074

b. *x* = 2: $P(2) = \frac{10!}{2!(10-2)!}.2^2.8^{(10-2)} =$ $\frac{9 \cdot 10}{2}(.2)^2(.8)^8$
$= 45(.04)(.1678) = .3020$

43

Cumulative Probability

The probability that a random variable is less than or equal to a value *x* is called a cumulative probability, and it is represented as *P*(*X* ≤ *x*).

In the case of a discrete probability distribution, such as the binomial distribution, we can write:

$$P(X \le x) = \sum_{X=0}^{x} P(X = x)$$

44

Chapter 7: Random Variables and Discrete Probability Distributions

Example – Will Pat Fail the Quiz?

Find the probability that Pat fails the quiz. For the purpose of this exercise, a mark is considered a failure if it is less than 50%.

Solution:

Because there are 10 questions, 50% corresponds to a mark of 5. Because the marks must be integers, a mark of 4 or less is a failure.

$$P(X \leq 4) = P(0) + P(1) + P(2) + P(3) + P(4)$$
$$= .1074 + .2684 + .3020 + .2013 + .0881 = .9672$$

There is a 96.72% probability that Pat will fail the quiz by guessing the answer for each question.

45

Binomial Table

There exists a cumulative binomial probabilities table for selected values of *n* and *p*.
(posted under chapter 7 on Canvas) alternate table

We can use this table to answer the question on the previous slide, where we need *P(X ≤ 4)*. The table to the right shows the values of *P(X ≤ 4)* for *n = 10* and *p = .2*.

We can use the table and the complement rule to determine probabilities of the type P(*X = x*) and *P(X ≥ x)*.

For example, to find the probability that Pat answer exactly two questions:

$$P(2) = P(X \leq 2) - P(X \leq 1) = .6778 - .3758 = .3020$$
$$\boldsymbol{P(x) = P(X \leq x) - P(X \leq [x-1])}$$

And the probability that Pat passes the quiz is:

$$P(X \geq 5) = 1 - P(X \leq 4) = 1 - .9672 = .0328$$
$$\boldsymbol{P(X \geq x) = 1 - P(X \leq [x-1])}$$

Cumulative Binomial Probabilities with *n = 10* and *p = .2*

X	P(X ≤ x)
0	.1074
1	.3758
2	.6778
3	.8791
4	.9672
5	.9936
6	.9991
7	.9999
8	1.0000

46

Mean and Variance of a Binomial Distribution

Statisticians have developed general formulas for the mean, variance, and standard deviation of a binomial random variable:

$$\mu = np$$
$$\sigma^2 = np(1-p)$$
$$\sigma = \sqrt{np(1-p)}$$

Example: Suppose that a professor has a class full of students like Pat.

a. What is the mean mark? $\mu = np = 10(.2) = 2$

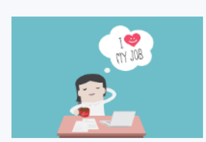
b. What is the standard deviation? $\sigma = \sqrt{np(1-p)} = \sqrt{10(.2).8)} = 1.26$

.

47

In a recent survey, the Pew Research Center asked graduates of private universities whether they were satisfied with their current job, and 72% said they were. Suppose you take a sample of four private university graduates and ask each whether they are satisfied with their current jobs.

a. What is the probability that all four say they are satisfied?
b. What is the probability that two say they are satisfied?
c. Determine the expected number of people who are satisfied.



48

In a recent survey, the Pew Research Center asked graduates of private universities whether they were satisfied with their current job, and 72% said they were. Suppose you take a sample of four private university graduates and ask each whether they are satisfied with their current jobs.

a. What is the probability that all four say they are satisfied?

b. What is the probability that two say they are satisfied?

c. Determine the expected number of people who are satisfied.

Solutions:

Excel with n = 4 and p = .72

a. $P(X = 4) = .2687$

b. $P(X = 2) = .2439$

c. $E(X) = np = 4(.72) = 2.88$

49

Poisson Distribution

A binomial random variable is the number of successes in a set number of trials, whereas a **Poisson random variable** is the number of successes in an interval of time or specific region of space.

Poisson Experiment

1. The number of successes that occur in any interval is independent of the number of successes that occur in any other interval.
2. The probability of a success in an interval is the same for all equal-size intervals.
3. The probability of a success in an interval is proportional to the size of the interval.
4. The probability of more than one success in an interval approaches 0 as the interval becomes smaller.

Examples

1. The number of cars arriving at a service station in 1 hour *(the interval of time is 1 hour)*
2. The number of flaws in a bolt of cloth *(the specific region is a bolt of clothes)*
3. 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)*

50

Poisson Probability Distribution

The probability that a Poisson random variable assumes a value of *x* in a specific interval is:

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

Where μ is the mean number of successes in the interval or region *(given)* and *e* is the base of the natural logarithm (approximately 2.71828).

The variance of a Poisson random variable is equal to its mean; that is, $\sigma^2 = \mu$.

51

Examples - Probability of the Number of Typographical Errors in Textbooks ...

A statistics instructor has observed that the number of typos in new editions of textbooks 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 typographical errors?

We want to determine the probability that a Poisson random variable with a mean of 1.5 is equal to 0.

Using the formula for Poisson probability distribution, with $x = 0$, and $\mu = 1.5$, we get:

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

The probability that in the 100 pages selected there are no errors is .2231.

52

Chapter 7: Random Variables and Discrete Probability Distributions

Examples - Probability of the Number of Typographical Errors in Textbooks …

The instructor now receives a copy of a new statistics book and notices that there are 400 pages. Calculate the probability that for a book of 400 pages there are **(a)** no typos, and **(b)** no more than five typos.

If there are 1.5 error per 100 pages, then there must be 4 x 1.5 = 6 errors per 400 pages of textbook, or $\mu = 6$.

a. Thus, the probability that there are no typos is:

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

b. And the probability that there five or less typos is:

$$P(X \leq 5) = P(0) + P(1) + P(2) + P(3) + P(4) + P(5)$$
$$= .002479 + .01487 + .04462 + .08924 + .1339 + .1606 = .4457$$

53

Poisson Table

Table 2 in Appendix B of the textbook provides cumulative Poisson probabilities for selected values of μ.

We can use this table to answer the question in the example, where we need $P(X \leq 5)$. The table to the right shows the values of $P(X \leq 5)$ for $\mu = 6$.

We can use the table and the complement rule to determine probabilities of the type $P(X = x)$ and $P(X \geq x)$.

For example, to find the probability that there are exactly ten typos:
$P(10)=P(X\leq10)-P(X\leq9)=.9574-.9161=.0413$

And the probability that there are more than five typos:
$P(X\geq6)=1-P(X\leq5)=1-.4457=.5543$

Cumulative Poisson Probabilities for $\mu = 6$

x	$P(X \leq x)$
0	.0025
1	.0174
2	.0620
3	.1512
4	.2851
5	.4457
6	.6063
7	.7440
8	.8472
9	.9161
10	.9574
11	.9799
12	.9912
13	.9964

54

In older cities across North America, infrastructure is deteriorating, including water lines that supply homes and businesses. A report to the Toronto city council stated that there are on average 30 water line breaks per 100 kilometers per year in the city of Toronto. Outside of Toronto, the average number of breaks is 15 per 100 kilometers per year.

a. Find the probability that in a stretch of 100 kilometers in Toronto there are 35 or more breaks next year.

b. Find the probability that there are 12 or fewer breaks in a stretch of 100 kilometers outside of Toronto next year.

Solutions:

a. Excel with $\mu = 30$: $P(X \geq 35) = 1 - P(X \leq 34) = 1 - .79731 = .20269$

b. Excel with $\mu = 15$: $P(X \leq 12) = .26761$

55

The number of trucks crossing at the Ambassador Bridge connecting Detroit, Michigan, and Windsor, Ontario, is Poisson distributed with a mean of 1.5 per minute.

a. What is the probability that in any 1-minute time span two or more trucks will cross the bridge?

b. What is the probability that fewer than four trucks will cross the bridge over the next 4 minutes?

Solutions:

a. Table 2 with $\mu = 1.5$: $P(X \geq 2) = 1 - P(X \leq 1) = 1 - .5578 = .4422$

b. Table 2 $\mu = 6$: $P(X < 4) = P(X \leq 3) = .1512$

56