Chapter 4.6-4.10 Binomial Distributions

Jim Albert and Monika Hu

Chapter 4 Discrete Distributions

Coin-Tossing Distributions

- Consider the following random experiment. One takes a quarter and flip it ten times, recording the number of heads.
- There are four special characteristics of this simple coin-tossing experiment.

1. One is doing the same thing ten times. We will call an individual coin flip a trial.
2. On each trial, there are two possible outcomes, heads or tails.
3. In addition, the probability of flipping heads on any trial is $1/2$.
4. The results of different trials are independent.

Coin Tossing Random Variable

- One is interested in the random variable X, the number of heads one gets.
- In particular, one is interested in the probability of getting five heads, or $Prob(X = 5)$.
- This Binomial probability model applies to many different random phenomena in the real world.

Binomial probabilities

- Let's return to our experiment where a quarter is flipped ten times, recording X, the number of heads.
- One is interested in $Prob(X = 5)$.
- One possible outcome is

Trial	1	2	3	4	5	6	7	8	9	10
Result	H	H	T	T	H	T	T	H	H	T

- Another possible outcome is $TTHHTHTHHH$. The sample space consists of all possible ordered listings of ten letters, where each letter is either an H or a T.

Binomial probabilities

- Next consider computing the probability of a single outcome of ten flips such as the $HHTTHHTHHT$. The probability of this outcome is written as

P("H on toss 1" AND "H on toss 2" AND ... AND "T on toss 10").

- Using independence, this probability is written as the product

$P(H \text{ on toss } 1) \times \, P(H \text{ on toss } 2) \times \ldots \times \, P(T \text{ on toss } 10)$.

- Since the probability of heads (or tails) on a given trial is $1/2$, one has

$$P(HHTTHHTTHT) = \frac{1}{2} \times \frac{1}{2} \times \ldots \times \frac{1}{2} = \left(\frac{1}{2}\right)^{10}.$$

Original Question

- What is the probability that one gets exactly five heads?
- There are many ways to get five heads such as

$HHHHHTTTTT$ or $HHHHTTTTTH$ or $HHHTTTTTHH$

- If one observes exactly 5 heads, then one must choose five numbers from the possible trial numbers 1, 2, ..., 10 to place the five H's. There are $\binom{10}{5}$ ways of choosing these trial numbers. Each outcome has probability $\left(\frac{1}{2}\right)^{10}$, so that

$$Prob(X = 5) = \binom{10}{5}\left(\frac{1}{2}\right)^{10} = 0.246.$$

Binomial experiments

Many random experiments share the same basic properties as coin tossing. The following describes a Binomial experiment:

1. One repeats the same basic task or trial many times – let the number of trials be denoted by n.
2. On each trial, there are two possible outcomes, which are called "success" or "failure". One could call the two outcomes "black" and "white", or "0" or "1", but they are usually called success and failure.
3. The probability of a success, denoted by p, is the same for each trial.
4. The results of outcomes from different trials are independent.

Example: A sample survey.

The Gallup organization is interested in estimating the proportion of adults in the United States who use the popular auction website eBay. They take a random sample of 100 adults and 45 say that they use eBay.

1. The results of this survey can be considered to be a sequence of 100 trials where one trial is asking a particular adult if he or she uses eBay.
2. There are two possible responses to the survey question – either the adult says “yes” or “no”
3. Suppose the proportion of all adults that use eBay is p. Then the probability that the adult says “yes” will be p.
4. The responses of different adults to the question can be regarded as independent events.

Example: A baseball hitter's performance during a game

Suppose you are going to a baseball game and your favorite player comes to bat five times during the game. This particular player is a pretty good hitter and his batting average is about 0.300. You are interested in the number of hits he will get in the game.

1. The player will come to bat five times – these five at-bats can be considered the five trials of the experiment.
2. At each at-bat, there are two outcomes of interest – either the player gets a hit or he doesn't get a hit.
3. The probability that he will get a hit in a single at-bat can be assumed to be $p = 0.300$.
4. It is reasonable to assume that the results of the different at-bats are independent.

Example: Sampling without replacement.

-Suppose a committee of four will be chosen at random from five women and five men. You are interested in the number of women that will be in the committee. Is this a Binomial experiment?

1. If one thinks of selecting this committee one person at a time, then one can think this experiment as four trials .
2. On each trial, there are two possible outcomes – either one selects a woman or a man.
3. Is the probability of choosing a woman the same for each trial? Once this first person has been chosen, the probability of choosing a woman is not $5/10$ – it will be either $4/9$ or $5/9$ depending on the outcome of the first trial. So the probability of a "success" is not the same for all trials, so this violates the third property of a Binomial experiment.

Binomial computations

- A Binomial experiment is defined by two numbers

$n =$ the number of trials, and

$p =$ probability of a "success" on a single trial.

- If one recognizes an experiment as being Binomial, then all one needs to know is n and p to determine probabilities for the number of successes X.

- One can show that the probability of x successes in a Binomial experiment is given by

$$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}, \; k = x..., n.$$

Example: A baseball hitter's performance during a game

- Remember our baseball player with a true batting average of 0.300 is coming to bat five times during a game. What is the probability that he gets exactly two hits?
- This is a Binomial experiment where the number of trials is $n = 5$ and the probability of success on a single trial is $p = 0.3$. Using the formula, the probability of exactly $X = 2$ two hits is

$$P(X = 2) = \binom{5}{2}(0.3)^2(1 - 0.4)^{5-2} = 0.3087.$$

Another Probability

- What is the probability that the player gets at least one hit?
- Construct the collection of Binomial probabilities for $n = 5$ trials and probability of success $p = 0.3$.

x	$P(X = x)$
0	0.168
1	0.360
2	0.309
3	0.132
4	0.029
5	0.002

Then we can compute our probability:

Binomial Calculations in R

- By use of the `dbinom()` function in R, one can tabulate binomial probabilities for the baseball example.
- A data frame is constructed with the values of hits `x`, and the function `dbinom()` with arguments `size` and `prob` to compute probabilities:

```
require(dplyr)
data.frame(x = 0:5) %>%
  mutate(Probability = dbinom(x, size = 5, prob = .3))
```

```
##   x Probability
## 1 0     0.16807
## 2 1     0.36015
## 3 2     0.30870
## 4 3     0.13230
## 5 4     0.02835
## 6 5     0.00243
```

Binomial Calculations in R

- The function `pbinom()` will compute cumulative probabilities of the form $P(X \leq x)$. For example, to find the probability that number of hits X is 2 or less, $P(X \leq 2)$:

```r
pbinom(2, size = 5, prob = .3)
```

```
## [1] 0.83692
```

- One computes the probability $P(X \geq 2)$ by finding the cumulative probability $P(X \leq 1)$, and subtracting the result from 1:

```r
1 - pbinom(1, size = 5, prob = .3)
```

```
## [1] 0.47178
```

Mean and variance of a Binomial

There are simple formula for the mean and variance for a Binomial random variable. First let X_1 denote the result of the first Binomial trial where

$$X_1 = \begin{cases} 1 & \text{if we observe a success} \\ 0 & \text{if we observe a failure} \end{cases}$$

One can show that the mean and variance of X_1 are given by

$$E(X_1) = p, \quad Var(X_1) = p(1-p).$$

Mean and variance of a Binomial

- If $X_1, ..., X_n$ represent the results of the n Binomial trials, then the Binomial random variable X can be written as

$$X = X_1 + ... + X_n.$$

The mean and variance of X are given by

$$E(X) = E(X_1)+...+E(X_n), \quad Var(X) = Var(X_1)+...+Var(X_n).$$

- Using this result, we obtain

$$E(X) = p + ... + p = np,$$

and

$$Var(X) = p(1-p) + ... + p(1-p) = np(1-p).$$

Example

- Recall the first example where X denoted the number of heads when a fair coin is flipped 10 times.
- Here the number of trials and probability of success are given by $n = 10$ and $p = 0.5$.
- The expected number of heads would be

$$E(X) = 10(0.5) = 5$$

- The variance of the number of heads would be

$$V(X) = 10(0.5)(1 - 0.5) = 2.5.$$