Chapter 6

The binomial distribution

Review of Last Week

Random variable: a numerical value obtained as or from the outcomes of a random experiment.

- discrete
- continuous

Describing discrete random variables

- probability mass function

$$f(x) = P(X = x)$$

- cumulative distribution function

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

Binomial experiment

A **binomial** experiment arises when a random process can be conceptualized as a sequence of smaller random processes called trials and

(a) the number of trials (usually denoted by n) is specified in advance,

(b) there are two outcomes (traditionally called success S and failure F) for each trial,

(c) the probability of a success (frequently denoted p or π) is the same in each trial, and

(d) each trial is independent of the other trials.

Sample space for a binomial experiment

The sample space of a binomial experiment consists of all the sequences of *Sucess* and *Failure* that could result.

Outcomes and probabilities for a binomial experiment with n=3 trials

1st trial	2nd trial	3rd trial	probability	Number of successes
Success	Success	Success	π^3	3
Success	Success	Failure	$\pi^2(1-\pi)$	2
Success	Failure	Success	$\pi^2(1-\pi)$	2
Success	Failure	Failure	$\pi(1-\pi)^2$	1
Failure	Failure	Failure	$(1-\pi)^3$	0
Failure	Failure	Success	$\pi(1-\pi)^2$	1
Failure	Success	Failure	$\pi(1-\pi)^2$	1
Failure	Success	Success	$\pi^2(1-\pi)$	2

The *number of successes* that occur in the *n* trials is called a **binomial random variable**.

Binomial Random Variable

Suppose that n independent trials, each of which results in a success with probability π and in a failure with probability $1 - \pi$, are to be performed. If X represents the number of successes that occur in the n trials, then X is said to be a binomial random variable with parameters (n, π).

We write $X \sim Binom(n, \pi)$. When $n = 1$, X is commonly referred to as a **Bernoulli** random variable.

Example 6.1

Consider the following random variables. Which ones are binomial?

a. X is the number of black marbles in a sample of 2 chosen randomly (with replacement) from two black (B_1, B_2) and one red (R).

Example 6.1

Consider the following random variables. Which ones are binomial?

b. X is the number of black marbles in a sample of 2 chosen randomly (without replacement) from two black (B_1, B_2) and one red (R).

Example 6.1

Consider the following random variables. Which ones are binomial?

c. The number of "good" rolls when a die is rolled four times. We call a roll "good" if the number rolled is greater than the roll number. So the first roll is good if it is a 2 or higher. The fourth roll is good if it is a 5 or a 6. Let X count the number of good rolls.

Example 6.1

Consider the following random variables. Which ones are binomial?

d. Independent trials consisting of the flipping of a coin are performed until a head is obtained. Let X denote the number of flips.

PMF of a Binomial

Theorem 6.1 Let $X \sim Binom(n, \pi)$. Then the PMF of X is given by

$$f(x) = P(X = x) = \binom{n}{x} \pi^x (1 - \pi)^{n-x}, \quad x = 0, 1, 2, \ldots, n$$

PMF of a Binomial

The sample space corresponding to the binomial experiment consists of all possible sequences of success and failure that that result from n independent trials:

$$\underbrace{S}_{\text{trial 1}} \times \underbrace{F}_{\text{trial 2}} \times \underbrace{F}_{\text{trial 3}} \cdots \times \underbrace{S}_{\text{trial n}}.$$

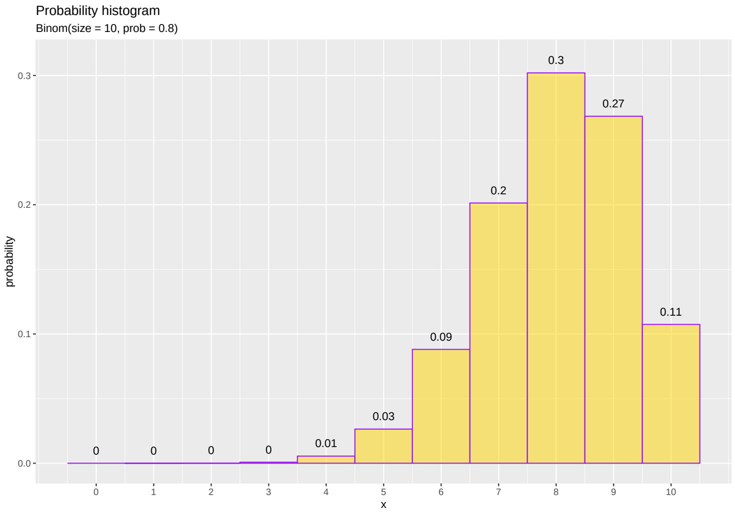
The number of successes x can equal any integer from 0 to n. By independence of the trials, any outcome with x successes will have probability $\pi^x(1-\pi)^{n-x}$ and the number of such outcomes is $\binom{n}{x}$ since we must select x of the n trials to be successful. So

$$\begin{aligned} P(X=x) &= \left(\begin{array}{c}\text{number of ways}\\ \text{to select } x\\ \text{of the } n \text{ trials}\end{array}\right) \cdot \left(\begin{array}{c}\text{probability of any}\\ \text{particular sequence of}\\ x \text{ successes and}\\ (n-x) \text{ failures}\end{array}\right), \\ &= \binom{n}{x}\pi^x(1-\pi)^{n-x}, x = 0, 1, 2, \ldots, n \end{aligned}$$

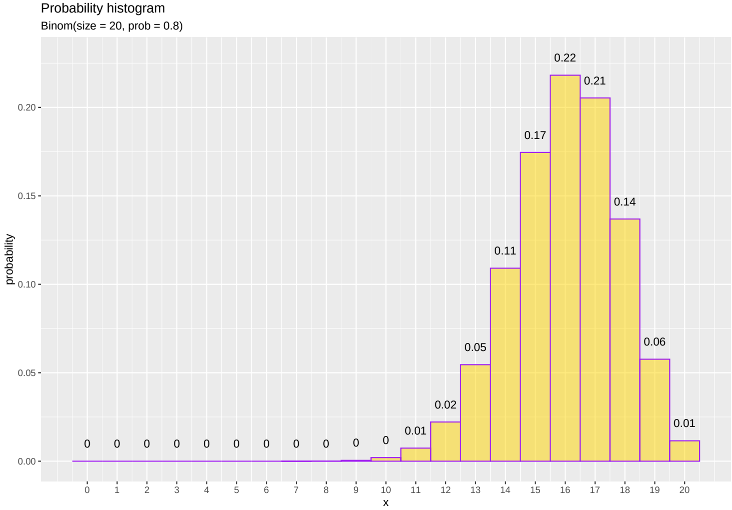
Example 6.2

Free throw Freddy is a 80% shooter which means the probability he makes a shot is 0.8. Which is more likely, that he makes at least 9 out of 10 shots or at least 18 out of 20 shots? What assumption are you making?

Probability Histogram of $X \sim Binom(10, 0.8)$



Probability Histogram of $X \sim Binom(20, 0.8)$



Binomial calculations in R

```
dbinom(x=9, size=10, prob=0.8)         #P(X=x) (the PMF)
```

```
## [1] 0.268
```

```
pbinom(q=8, size=10, prob=0.8)         #P(X<=q) (the CDF)
```

```
## [1] 0.624
```

```
pbinom(q=17, size=20, prob=0.8, lower.tail=F)      #P(X > q)
```

```
## [1] 0.206
```

```
#smallest x such that P(X <= x) >= p
qbinom(p = 0.9, size = 20, prob = 0.8)
```

```
## [1] 18
```

```
# 10 random draws of X.
set.seed(9898)            #set random number seed for reproducibility
rbinom(n = 10, size = 20, prob = 0.8)
```

```
##  [1] 19 16 15 15 14 15 18 17 16 17
```

Return to example 6.2 and write the R function for calculating the probabilities in R.

Probability Histogram code

```r
bball <- tibble( #enhanced data frame
   x = 0:20,      #allows me to define f in terms of x
   f = dbinom(x, size = 20, prob = 0.8)
)

# make probability histogram
ggplot(data = bball,
       mapping = aes(x = x, y = f) ) +
  geom_col(width = 1, color = "purple", fill = "gold", alpha = 0.5) +
  geom_text( mapping = aes( label = round(f, 2), y = f + 0.01) )+
  scale_x_continuous(breaks = 0:20)+
  labs(x = "x",
       y = "probability",
       title = "Probability histogram",
       subtitle = "Binom(size = 20, prob = 0.8)")
```