Chapter 6

Tuesday, October 17, 2023 4:03 PM

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 – the possible values are finite or countable
  ex: $x = 1/2, 1, 3/2$
- continuous → possible values are (a, b) (an interval)

Error in notation

→ $\text{range}(X) = \{0, 1, 2, \ldots, 12\} = [0, 12]$ ✗

$\{0, 1, 2, \ldots, \infty\} = [0, \infty)$ ✗

Describing discrete random variables

- probability mass function PMF

lower case → $f(x) = P(X = x)$ $\forall x$

lower case $\rightarrow f(x) = P(X = x)$ $\forall x$

- cumulative distribution function CDF

uppercase $\rightarrow F(x) = P(X \leq x)$ $\forall x$

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.

EX: Toss a coin 3 times (textbook ex of binomial)

- trial = coin toss

- $n = 3$ (# of trials)
- on every toss, 2 outcomes: head or tail
  ↳ success ↳ failure
- P(head) should be same for every trial
- whether or not we see a "head" on one coin toss does not affect the chances of "head" on another toss

Sample space for a binomial experiment

does not affect the chances of ... toss

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 (X)
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**.

$P(X=0) = (1-\pi)^3 = \binom{3}{0}\pi^0(1-\pi)^3$

$P(X=1) = 3\pi(1-\pi)^2 = \binom{3}{1}\pi^1(1-\pi)^2$

$P(X=2) = 3.\pi^2(1-\pi)^1 = \binom{3}{2}\pi^2(1-\pi)$

$P(X=3) = \pi^3 = \binom{3}{3}\pi^3(1-\pi)^0$

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).

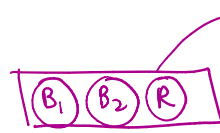
inputs

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

distributed as

Example 6.1

Consider the following random variables. Which ones are binomial?



draw 2
$X = \#$ black

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).

What is the trial? drawing one marble from box

1) Is number of trials fixed? Yes $n=2$ ✓

2) Are there 2 outcomes on each trial? Yes, black/red
↳ success

3) Is prob of success the same for every trial? ✓

1st trial, 2nd trial

$S = \{ (B_1, B_2), (B_1, R), (B_2, B_1), (B_2, R), (R, B_1), (R, B_2),$

3) Is prob of success the same for every trial? $\xrightarrow{1^{st}\text{ trial}}$ $\xrightarrow{2^{nd}\text{ trial}}$

$S = \{(B_1, B_2), (B_1, R), (B_2, B_1), (B_2, R), (R, B_1), (R, B_2), (B_1, B_1), (B_2, B_2), (R, R)\}$

all 9 outcomes equally likely.

P(Black marble on 1^{st} draw) $= \frac{6}{9} = \frac{2}{3}$

P(Black " " 2^{nd} draw) $= \frac{6}{9} = \frac{2}{3}$

yes, they are the same!

4) Are trials independent? ✓

What is P(Black on 2^{nd} draw | Black on 1^{st} draw) $= 4/6 = 2/3$

Reduced sample space: $(B_1, B_2), (B_1, R), (B_2, B_1), (B_2, R), (B_1, B_1), (B_2, B_2)$

$=$ P(Black on 2^{nd} draw)

$\therefore$ trials are independent!

Yes this is a binomial expt!

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).

$\boxed{B_1, B_2, R}$ 2 draws randomly $X = \#$ black

1) $n = 2$ ✓

2) 2 outcomes on each trial ✓

3) Is Prob of success the same for every trial? ✓

$S = \{(B_1, B_2), (B_1, R), (B_2, B_1), (B_2, R), (R, B_1), (R, B_2)\}$

all equally likely.

P(Black on 1^{st} draw) $= \frac{4}{6} = 2/3$

P(Black on 2^{nd} ") $= 4/6 = 2/3$

same

4) Are trials independent? ✗

P(black on 2^{nd} | black on 1^{st}) $= 2/4 \neq$ P(black on 2^{nd})

reduced sample space: $(B_1, B_2), (B_1, R), (B_2, B_1), (B_2, R)$

Not a binomial expt.

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.

What is a trial? Rolling a die once

- $n = 4$ ✓
- good / bad
  ↳ success
- P(good on Roll 1) = 5/6
  P(good on Roll 4) = 2/6] ✗

Not binomial Expt

Example 6.1

Consider the following random variables. Which ones are binomial?

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

What is a trial? Toss a coin once.

- # trials fixed? No.

Not binomial expt

PMF of a Binomial

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

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

ways you can choose the trials that are successes.

prob. of x successes and $(n-x)$ failures

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

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

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?

X = # baskets in 10 throws

we assume $X \sim \text{Binom}(n = 10, \pi = .8)$ which is reasonable provided the throws are independent.

want to calculate

$$\begin{aligned} P(X \geq 9) &= P(X=9) + P(X=10) \\ &= \binom{10}{9} .8^9 \, .2^1 + \binom{10}{10} .8^{10} \, .2^0 \\ &= .2684 + .1074 = \boxed{.3758} \end{aligned}$$

$$= .2684 + .1074 = \boxed{.3758}$$

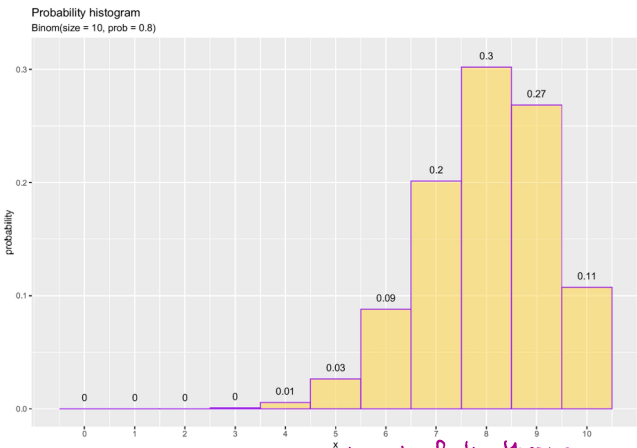
Y = # baskets in 20 throws

we assume $Y \sim \text{Binom}(n=20, \pi=.8)$

$$P(Y \geq 18) = P(Y=18) + P(Y=19) + P(Y=20)$$
$$= \binom{20}{18} \cdot .8^{18} \cdot .2^{2} + \binom{20}{19} \cdot .8^{19} \cdot .2^{1} + \binom{20}{20} \cdot .8^{20} \cdot .2^{0}$$
$$= .1369 + .0576 + .0115$$
$$= \boxed{.2060}$$

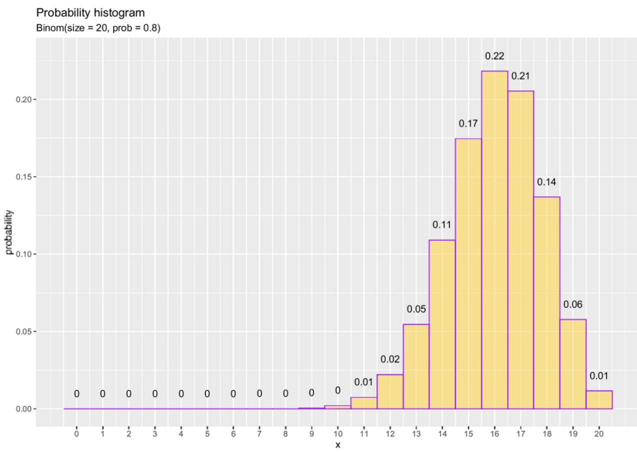
More likely to score above 90% when $n=10$

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



baskets made out of 10 throws

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



Binomial calculations in R

n = # trials

π

```
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
```

#trials

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

how many X's should I generate?

```
##  [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.

$X \sim Binom(n=10,\ \pi=.8)$

a. $P(X \geq 9) = 1 - P(X < 9) = 1 - P(X \leq 8)$

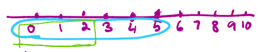
$= 1-$ pbinom(q=8, size=10, prob=.8)

b. $P(X \geq 9) = P(X > 8)$

= pbinom(q=8, size=10, prob=.8, lower.tail=F)

Ex: $P(2 < X \leq 5) = P(X \leq 5) - P(X \leq 2)$

= pbinom(q=5, size=10, prob=.8) − pbinom(q=2, size=10, prob=.8)



Probability Histogram code

```
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)")
```