Chapter 14: random variables p394

- A **random variable (r. v.)** is a variable whose value is a numerical outcome of a random phenomenon.
-

Consider the experiment of tossing a coin.

Define a random variable as follows

X = 1 if a H comes up

= 0 if a T comes up.

- This is an example of a Bernoulli r.v.

Probability function of X

x	P(X = x)
0	q
1	p

p +q = 1

Probability distributions

Each value of a random variable is an event, so each value has probability. List of values and probabilities called *probability model*.

Tossing 3 coins:

# heads	0	1	2	3
Prob.	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

Combining values of random variable:

3 coins:

# heads	0	1	2	3
Prob.	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

- How likely are we to get two or more heads?
  - *add up probs: 3/8+1/8=4/8=1/2*

- How likely to get at least one head?
  - *P(no heads)=1/8, so P(at least one)=1-1/8=7/8*
    - *or: P(1 or 2 or 3)=3/8+3/8+1/8=7/8*

The mean of a random variable p394

Here's a random variable, called X:

Value of X	2	3	4	5
Probability	0.1	0.2	0.4	0.3

- Mean *not* (2+3+4+5)/4=3.5 because 4 and 5 more likely than 2 or 3.
- Have to account for more likely values when adding up:
  - times by probability:
  - 2(0.1)+3(0.2)+4(0.4)+5(0.3)=0.2+0.6 +1.6+1.5=3.9.
  - (Weighted average, weights sum to 1.)
- Median is value of X where summed-up probabilities first pass 0.5: 3 too small (total 0.1+0.2=0.3), 4 is right (0.1+0.2+0.4=0.7), so median 4.
- Mean a little smaller than median: left-skewed.

▪ The **variance** of a r. v. P397

- The variance of a r. v. is an average of the squared deviations $(X-\mu_X)^2$

- Variance of a discrete r. v. is

$$Var(X)=\sum(x-\mu)^2 P(x)$$

- The standard deviation of a r. v. is the positive square root of its variance.
- Examples

Continuous random variables

- So far: our random variables *discrete:* set of possible values, like 1,2,3,... , probability for each.
- Recall normal distribution: *any* decimal value possible, can't talk about probability of any *one* value, just eg. “less than 10”, “between 10 and 15”, “greater than 15”.
- Normal random variable example of *continuous.*
- Finding mean and SD of continuous random variable involves *calculus :-(*
- but if we are given mean/SD, work as above.

Probability Models p 405

The Binomial Model

Example:

A biased coin (P(H) = p = 0.6)) is tossed 5 times. Let X be the number of H's. Fine P(X = 2).

This X is a binomial r. v.

The binomial setting

1. There are a fixed number *n* of observations.
2. The *n* observations are independent.
3. Each observation falls into one of just two categories (successes and failures)
4. The probability of a success (call it *p*) is the same for each observation.

- Probability function of the binomial dist.

If X has a *B*(*n*, *p*),

$$P(X=x)={}^{n}C_{x}p^{x}(1-p)^{n-x} \text{ for } n=0,1,\ldots,n$$

Binomial table

The link to Statistical Tables on course website includes table of *binomial distribution probabilities.* In here, find chance of exactly k successes in n trials with success prob p.

Ex.

The probability that a certain machine will produce a defective item is 1/5. If a random sample of 6 items is taken from the output of this machine, what is the probability that there will be 5 or more defectives in the sample?

Ex There are 20 multiple-choice questions on an exam, each having responses a, b, c, d and e. Each question is worth 5 points. And only one response per question is correct. Suppose that a student guesses the answer to question and her guesses from question to question are independent. It the student needs at least 40 points to pass the test. What is the probability that the student will pass the test?

Ans. X~B(20, 0.2). P(X>=8) = 0.0322, adding the entries 8 through 20 in the appropriate of the binomial table

What is the expected (mean) score for this student. (later)

Ans. 20 x 0.2 = 4 and expected score =5 x 4= 20

Suppose n=8 and p=0.7. What is the probability of

- exactly 7 successes?
- 7 or more successes?

Idea: count failures instead of successes.

P(success)=0.7 means P(failure)=1-0.7=0.3

7 successes = 8-7=1 failure.

so look up n=8, p=0.3, k=1 prob=0.1977 which is answer we want.

7 or successes = 7, 8 successes

P(failure)=1-0.7=0.3

7, 8 successes = 1, 0 failures

prob we want is 0.1977+0.0576=0.2553.

- Mean and Variance of a binomial r. v. p

If *X* has a *Bin*(*n*, *p*)

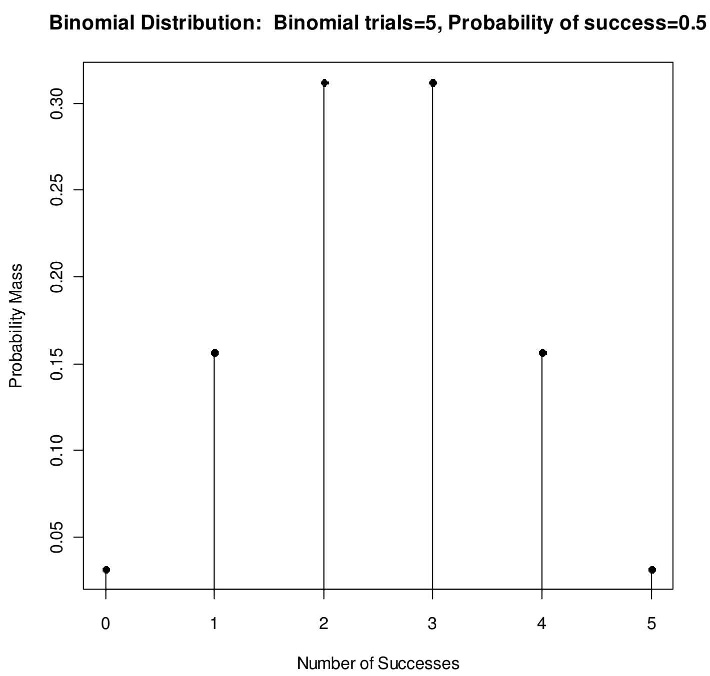
$mean = np$ and $SD = \sqrt{np(1-p)}$

Example

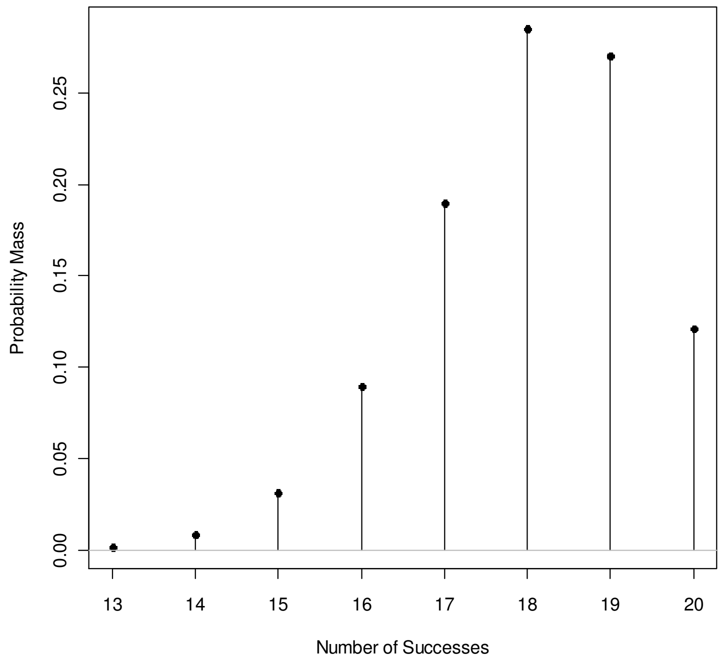
Use StatCrunch to explore the shapes of these binomial distributions:

(a) $n = 5, p = 0.5$
(b) $n = 20, p = 0.9$
(c) $n = 30, p = 0.2$
(d) $n = 500, p = 0.4$

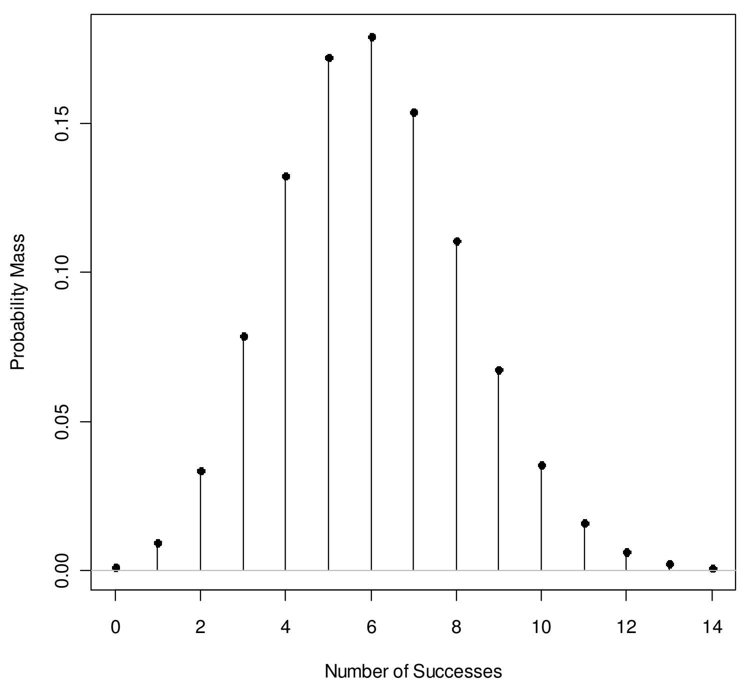
Which distribution does the last one resemble?

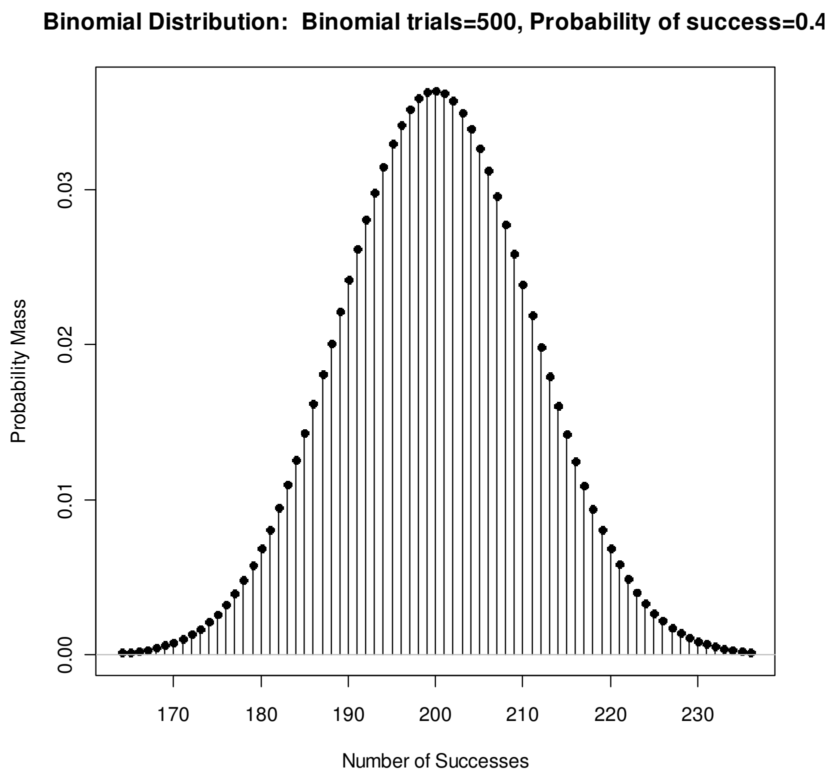


Binomial Distribution: Binomial trials=20, Probability of success=0.9



Binomial Distribution: Binomial trials=30, Probability of success=0.2




Binomial Distribution: Binomial trials=500, Probability of success=0.4
Probability Mass
0.00
0.01
0.02
0.03
170
180
190
200
210
220
230
Number of Successes

How does the shape depend on p?

p<0.5, skewed right;

p>0.5, skewed left;

p=0.5, symmetric

What happens to the shape as n increases?

- *shape becomes normal*

What does this suggest to do if n is too large for the tables?

If n too large for tables, try **normal approximation to binomial.**

Compute mean and SD of binomial, then pretend binomial actually normal:

- **Normal approximation for counts and proportions** p415

Draw a SRS of size *n* from a large population having population *p* of success. Let X be the count of success in the sample and $\hat{p} = X/n$ the sample proportion of successes. When n is large, the sampling distributions of these statistics are approximately normal:

X is approx. $N(np, \sqrt{np(1-p)})$

Works if n large and p not too far from 0.5

As a rule of thumb, we will use this approximation for values of n and p that satisfy $np \geq 10$ and $n(1-p) \geq 10$.

can relax this a bit if p close to 0.5.

According to government data, 21% of American children under the age of six live in households with incomes less than the official poverty level. A study of learning in early childhood chooses an SRS of 300 children.

(a) What is the mean number of children in the sample who come from poverty-level households?

What is the standard deviation of this number?

(b) Use the normal approximation to calculate the probability that at least 80 of the children in the sample live in poverty. Be sure to check that you can safely use the approximation.

(a) $\mu = (300)(0.21) = 63$, $\sigma = \sqrt{(300)(0.21)(0.79)} = 7.0548$. (b) $np = 63$ and $n(1-p) = 237$ are both more than 10, so we may approximate using the normal distribution: $P(X \geq 80) = P(Z \geq 2.41) = 0.0080$, or with the continuity correction: $P(X \geq 79.5) = P(Z \geq 2.34) = 0.0096$.