

# STAT 515 fa 2023 Lec 05 slides

## Bernoulli trials, binomial and hypergeometric distributions

Karl B. Gregory

University of South Carolina

These slides are an instructional aid; their sole purpose is to display, during the lecture, definitions, plots, results, etc. which take too much time to write by hand on the blackboard. They are not intended to explain or expound on any material.



## Bernoulli trial

A *Bernoulli* trial is an experiment with the two outcomes “success” and “failure”.

We often let  $p$  denote the probability of a “success”.

### Examples:

- 1 Flip a coin and call “heads” a “success”. If the coin is fair,  $p = 1/2$ .
- 2 Shoot a free throw and call making it a “success”. What is your  $p$ ??

Consider a rv  $X$  that encodes the outcome of a Bernoulli trial such that

$$X = \begin{cases} 1 & \text{if "success"} \\ 0 & \text{if "failure"} \end{cases}$$

## Bernoulli distribution

Let  $X$  be a rv with support  $\mathcal{X} = \{0, 1\}$  such that  $P(X = 1) = p$ . Then  $X$  has the *Bernoulli distribution* with success probability  $p$ .

We write  $X \sim \text{Bernoulli}(p)$ .

The probabilities  $P(X = x)$  are given by

$$P(X = x) = p^x(1 - p)^{1-x} \quad \text{for } x \in \{0, 1\}.$$

**Exercise:** Let  $X \sim \text{Bernoulli}(p)$ .

- 1 Find  $\mathbb{E}X$ .
- 2 Find  $\text{Var } X$ .

Consider the number of successes in a sequence of independent Bernoulli trials. . .

**Exercise:** Let  $X = \#$  heads in 4 coin tosses.

- 1 Give the sample space of the experiment of tossing a coin 4 times.
- 2 Tabulate the probability distribution of  $X$ .
- 3 Add cumulative probabilities to the table.

This leads to the binomial distribution.

## Binomial distribution

Let  $X = \#$  “successes” in  $n$  of indep. Bernoulli trials, each with success prob.  $p$ . Then  $X$  has the *Binomial distribution* based on  $n$  Bernoulli trials with success probability  $p$ .

We write  $X \sim \text{Binomial}(n, p)$ .

The probabilities  $P(X = x)$  for  $x \in \{0, 1, \dots, n\}$  are given by

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

**Exercise:** Let  $X = \#$  free throws you make in 4 attempts. Let  $p = 0.7$ .

- 1 Give the sample space of the experiment.
- 2 Assign a probability to each outcome in the sample space.
- 3 Tabulate the probability distribution of  $X$ .

**Exercise:** Let  $X = \#$  free throws you make in 10 attempts. Let  $p = 0.7$ .

- 1 Compute  $P(X = 3)$ .
- 2 Give the probability that you make at least one free throw.
- 3 Find  $P(X > 6)$ .
- 4 Find  $P(X \leq 6)$ .

Discuss R functions `dbinom()` and `pbinom()`.

## Binomial mean and variance

If  $X \sim \text{Binomial}(n, p)$ , then

- $\mathbb{E}X = np$ .
- $\text{Var} X = np(1 - p)$ .

Discuss how we would get these expressions.

**Exercise:** Suppose you make free throws with  $p = 0.7$  and attempts are indep.

- 1 Compute  $\mathbb{E}X$  when  $X = \#$  free throws made in 4 attempts.
- 2 If you shoot 1000 free throws, how many do you “expect” to make?

**Exercise:** Consider sampling 5 ppl from a population of 100 of whom 10 vape. Let  $X = \#$  in sample who vape.

- 1 Find  $P(X = 1)$  if we sample without replacement.
- 2 Find  $P(X = 1)$  if we sample with replacement.



## Hypergeometric distribution

Draw  $K \geq 0$  marbles from a bag of  $N \geq 0$  marbles, of which  $M \geq 0$  are red. If  $X = \#$  red marbles drawn, then  $X$  has the *Hypergeometric distribution*.

We write  $X \sim \text{Hypergeometric}(N, M, K)$ .

If  $X \sim \text{Hypergeometric}(N, M, K)$ , then we have

$$P(X = x) = \frac{\binom{M}{x} \binom{N-M}{K-x}}{\binom{N}{K}} \quad \text{for } x = \max\{K - (N - M), 0\}, \dots, \min\{M, K\}$$

**Exercise:** Consider sampling 5 ppl from a population of 100 of whom 10 vape. Let  $X = \#$  in sample who vape.

- 1 Tabulate  $P(X = x)$  for  $x = 0, 1, \dots, 5$  if we sample without replacement.
- 2 Tabulate  $P(X = x)$  for  $x = 0, 1, \dots, 5$  if we sample with replacement.

## Hypergeometric mean and variance

If  $X \sim \text{Hypergeometric}(N, M, K)$ , then

- $\mathbb{E}X = K \frac{M}{N}$ .
- $\text{Var } X = K \frac{M}{N} \left[ \frac{(N-K)(N-M)}{N(N-1)} \right]$ .

Discuss how we would get these expressions.

**Exercise:** Draw 4 cards from a 52-card deck and let  $X = \#\spadesuit$ s in hand.

- 1 Tabulate  $P(X = x)$  for  $x = 0, 1, 2, 3, 4$ .
- 2 Give  $\mathbb{E}X$ .

**Exercise:** Consider sampling 5 ppl from a population of 10,000 of whom 1,000 vape. Let  $X = \#$  in sample who vape.

- 1 Find  $P(X = 1)$  if we sample without replacement.
- 2 Find  $P(X = 1)$  if we sample with replacement.

Use R functions `dhyper()` and `phyper()`.

Note that with `dhyper(x,m,n,k)`, `m` is our  $M$ , `n` is our  $N - M$ , and `k` is our  $K$ .

Hypergeometric probs approach binomial probs as  $N \rightarrow \infty$  and  $M/N \rightarrow p$ .



If the pop. is large, sampling with/without replacement are practically the same!

Discuss treating samples from finite-but-large populations as independent draws.

## Probability mass function

The *probability mass function (pmf)* of a discrete rv  $X$  with support  $\mathcal{X}$  is the function given by

$$p(x) = P(X = x) \quad \text{for } x \in \mathcal{X}.$$

For  $x \notin \mathcal{X}$ ,  $p(x) = 0$ .

If  $X$  is an rv with pmf  $p$ , then we write  $X \sim p$ .