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.

Jakob Bernoulli

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

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

x	0	1
$P(X=x)$	$1-p$	p

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.

X "follows" the Bernoulli...

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

$X \sim \boxed{\text{Name of dist.}}$ ("parameters")

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

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

$$P(X=0) = p^0(1-p)^{1-0} = 1-p$$

$$P(X=1) = p^1(1-p)^{1-1} = p$$

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

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

x	0	1
$P(X=x)$	$1-p$	p

① $\mathbb{E}X = 0(1-p) + 1(p) = p.$

$$\text{Var } X = (1-p)(0-p)^2 + p(1-p)^2$$

$$= (1-p)p^2 + p(1-p)^2$$

$$= (1-p)[\underbrace{p^2 + p(1-p)}_{p}]$$

$$= p(1-p)$$

$$\text{Var } X = p_1(x_1-\mu)^2 + p_2(x_2-\mu)^2 + p_3(x_3-\mu)^2 + \dots$$

$$\sigma = \sqrt{\text{Var } X} = \sqrt{p(1-p)}$$

successes in n indep Bernoulli trials has the Binomial distribution.

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

n

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.

$2 \cdot 2 \cdot 2 \cdot 2 = 16$

This leads to the binomial distribution.

$S = \{$ HHHH, HHHT, HHTH, HTHH, THHH, HHTT, HTHT, THHT, TTHH, THTH, HTTH, TTTH, TTHT, THTT, HTTT, TTTT $\}$

$\mathcal{X} = \{0, 1, 2, 3, 4\}$

x	0	1	2	3	4
$P(X=x)$	$\frac{1}{16}$	$\frac{4}{16}$	$\frac{6}{16}$	$\frac{4}{16}$	$\frac{1}{16}$

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.

parameters

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

The probabilities $P(X = x)$ for $x \in \{0, 1, \ldots, n\}$ 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.

H = make it

T = miss

$p = 0.7$

$$S = \left\{ \underbrace{HHHH}_{(.7)(.7)(.7)(.7) = (.7)^4},\ \underbrace{\begin{matrix} (.7)(.7)(.7)(.3) \\ HHHT \\ (.7)(.7)(.3)(.7) \\ HHTH \\ HTHH \\ THHH \end{matrix}}_{(.7)^3(.3)},\ \underbrace{\begin{matrix} HHTT \\ HTHT \\ THHT \\ TTHH \\ THTH \\ HTTH \end{matrix}}_{(.7)^2(.3)^2},\ \underbrace{\begin{matrix} TTTH \\ TTHT \\ THTT \\ HTTT \end{matrix}}_{(.7)(.3)^3},\ \underbrace{TTTT}_{(.3)^4} \right\}$$

x	0	1	2	3	4
$P(X=x)$	$1(.7)^0(.3)^4$	$4(.7)(.3)^3$	$6(.7)^2(.3)^2$	$4(.7)^3(.3)$	$1(.7)^4(.3)^0$
	$\binom{4}{0}(.7)^0(.3)^{4-0}$	$\binom{4}{1}(.7)^1(.3)^{4-1}$	$\binom{4}{2}(.7)^2(.3)^{4-2}$	$\binom{4}{3}(.7)^3(.3)^{4-3}$	$\binom{4}{4}(.7)^4(.3)^{4-4}$

$\binom{4}{1} = \binom{4}{3}$ why?

$$P(X=x) = \binom{4}{x}(.7)^x(\underbrace{1-.7}_{.3})^{4-x}$$

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

$$\binom{N}{0} = \frac{N!}{\underbrace{0!}_{=1}(N-0)!} = \frac{N!}{N!} = 1$$

$$\mathcal{X} = \{0, 1, 2, \dots, 10\}$$

Exercise: Let $X = \#$ free throws you make in $\underbrace{10}_{n=10}$ attempts. Let $p = 0.7$.

1. Compute $P(X = 3)$. $= \binom{10}{3}(.7)^3(1-.7)^{10-3} = \text{dbinom}(3, 10, .7) = 0.009$
2. Give the probability that you make at least one free throw.
3. Find $P(X > 6)$.
4. Find $P(X \le 6)$. $= 1 - P(X > 6)$

$\text{pbinom}(7, 10, .7) = P(X \le 7)$

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

② $P(X \geq 1) = 1 - \underbrace{P(X < 1)}_{P(X=0)} = 1 - P(X=0)$

$$= 1 - \underbrace{\binom{10}{0}}_{1}\underbrace{(.7)^0}_{1}\underbrace{(1-.7)^{10-0}}_{(.3)^{10}} = 1-(.3)^{10} = .99999$$

(3)

$$P(X>6) = P(X=7) + P(X=8) + P(X=9) + P(X=10)$$

$$= 1 - P(X \le 6) = 1 - \text{pbinom}(6, 10, .7) = 0.6496.$$

$$= \binom{10}{7}(.7)^{7}(1-.7)^{10-7} + \dots + \binom{10}{10}(.7)^{10}(1-.7)^{10-10}$$

$$= \sum_{x=7}^{10} \binom{10}{x}(.7)^{x}(1-.7)^{10-x}$$

$$= \sum_{m=7}^{10} \binom{10}{m}(.7)^{m}(1-.7)^{10-m}$$

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

$$\text{pbinom}(x, n, p) = P(X \le x)$$

$$= P(X=0) + \dots + P(X=x)$$

$$= \sum_{m=0}^{x} \binom{n}{m} p^{m}(1-p)^{n-m}$$

Binomial mean and variance

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

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

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 10 attempts. $\mathbb{E}X = 10 \cdot (.7) = 7.$
2. If you shoot 1000 free throws, how many do you "expect" to make?

$$\underbrace{\overbrace{V\,V\,V\,V\,V\,V\,V\,V\,V\,V}^{10}\,(\underbrace{N\,N\cdots N}_{90})}_{100}$$

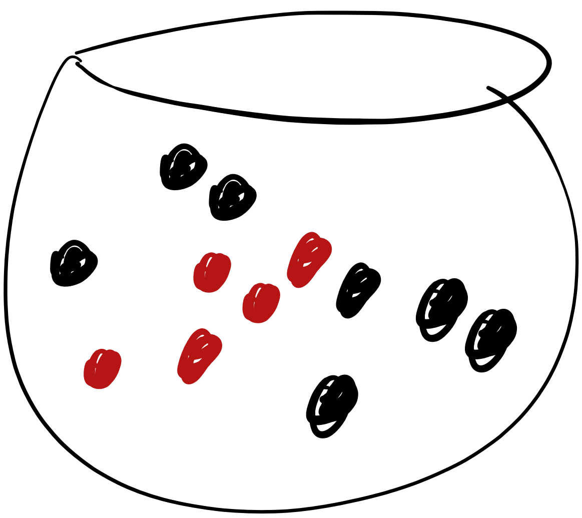
Exercise: Consider sampling 5 ppl from a population of 100 of whom 10 vape. Let $X = \#$ in sample who vape.

1. Find $P(X = 1)$ if we sample without replacement. ← We do this in surveys
2. Find $P(X = 1)$ if we sample with replacement.

① $P(X=1) = \dfrac{\binom{10}{1}\binom{90}{4}}{\binom{100}{5}} = 0.339$

② $X \sim \text{Binomial}(n=5,\ p=0.1)$

$P(X=1) = \binom{5}{1}(0.1)^1(1-0.1)^{5-1} = 5(.1)(.9)^4 = \text{dbinom}(1,5,0.1)$
$= .328$



N marbles

M red

draw K marbles from bag.

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

$\mathcal{X} = \{ \max\{K-(N-M), 0\}, \ldots, \min\{M, K\} \}$

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

$K = 5$ $N = 100$ $M = 10$

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

$\mathcal{X} = \{0, 1, 2, 3, 4, 5\}$

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

①

x	0	1	2	3	4	5
$P(X=x)$		$\frac{\binom{10}{1}\binom{90}{4}}{\binom{100}{5}}$				

$N - M$ = # non red marbles

dhyper(x, m, n, k)

M = # red marbles

K

round(dhyper(0:5, 10, 90, 5), 3)

digits

---

② Now $X \sim$ Binomial$(n = 5,\ p = 0.10)$

x	0	1	2	3	4	5
$p(X=x)$		$= .328$				

round(dbinom(0:5, 5, 0.10), 3)

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.

$N = 52$
$M = 13$
$K = 4$

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 $\mathbb{E}X$.

$$\mathbb{E}X = K\frac{M}{N} = 4 \cdot \frac{13}{52} = 1.$$

100 10

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. ← dhyper(1, 1000, 9000, 5)
2. Find $P(X = 1)$ if we sample with replacement. ← dbinom(1, 5, 0.10)

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

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

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 $X \sim p$.