STAT 515 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
(Swiss)

Bernoulli trial

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

$S = \{ \text{success}, \ \text{failure} \}$

We often let p denote the probability of a "success".

$P(\text{success}) = p \in [0,1]$

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

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

$\mathcal{X} = \{0, 1\}$

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

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

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

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

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

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

$$\mathbb{E}X = 0(1-p) + 1(p) = p$$

$$\begin{aligned}
\operatorname{Var} X &= (1-p)(0-p)^2 + p(1-p)^2 \\
&= (1-p)p^2 + p(1-p)^2 \\
&= (1-p)\left(p^2 + p(1-p)\right) \\
&= (1-p)p \\
&= p(1-p).
\end{aligned}$$

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

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.

S = "make it"

① $S = \{$

- $\binom{4}{4} = 1$: SSSS — $(0.7)^4$
- $\binom{4}{3} = 4$: $\boxed{\text{SSSF, SSFS, SFSS, FSSS}}$ — each is $(.3)(.7)^3$
- $\binom{4}{2}$: SSFF, SFSF, FSSF, FFSS, FSFS, SFFS — each: $(.3)^2(.7)^2$
- $\binom{4}{1}$: $\boxed{\text{FFFS, FFSF, FSFF, SFFF}}$ — each $(.3)^3(.7)$
- $\binom{4}{0}$: FFFF — $(.3)^4$

$\}$

② $\underline{2}\cdot\underline{2}\cdot\underline{2}\cdot\underline{2} = 2^4 = 16$

③

x	0	1	2	3	4
$P(X=x)$	$(0.3)^4$	$4(0.3)^3(0.7)$	$6(0.3)^2(0.7)^2$	$4(.3)(.7)^3$	$(0.7)^4$

$n=4,\ p=0.7$

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

$$P(X=0)=\binom{4}{0}(0.7)^0(1-0.7)^{4-0}=(0.3)^4$$

$$P(X=1)=\binom{4}{1}(0.7)^1(1-0.7)^{4-1}=4(0.7)(0.3)^3$$

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

$$\vdots$$

$$\frac{4!}{2!(4-2)!}=\frac{4\cdot3\cdot2\cdot1}{2\cdot1\cdot2\cdot1}=6$$

$$X\sim\text{Binomial}(n=4,\ p=0.7)$$

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

n

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

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

Use R functions `dbinom()` and `pbinom()` to compute probabilities for X:

On average you make 7 free throws

$\mathbb{E}X = 10 \cdot (0.7) = 7.$

$$P(X = x) = \texttt{dbinom(x,n,p)}$$
$$P(X \leq x) = \texttt{pbinom(x,n,p)}$$

dbinom(3, 10, 0.7)

$X \sim \text{Binomial}(n = 10, \; p = 0.7)$

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

1. Compute $P(X = 3)$. $= \binom{10}{3}(.7)^3(1-.7)^{10-3} = 0.0090.$
2. Give the probability that you make at least one free throw.
3. Find $P(X > 6)$.
4. Find $P(X \leq 6)$.

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

(2) $P(X \geq 1) = 1 - P(X=0)$

$$= 1 - \binom{10}{0}(.7)^0(1-.7)^{10-0}$$

$$= 1 - (.3)^{10}$$

(4) $\underline{P(X \leq 6)} = P(X=0) + P(X=1) + \dots + \underline{P(X=6)}$

$$= \binom{10}{0}(.7)^0(1-.7)^{10-0} + \binom{10}{1}(.7)^1(1-.7)^{10-1}$$

$$+ \dots + \binom{10}{6}(.7)^6(1-.7)^{10-6}$$

$$= \sum_{x=0}^{6} \binom{10}{x}(.7)^x(1-.7)^{10-x}$$

$= \text{pbinom}(6, 10, 0.7)$

$= 0.3504$

(3) $P(X > 6) = 1 - P(X \leq 6)$

$= 1 - \text{pbinom}(6, 10, 0.7).$

Don't have an R function for this.

x	x_1	x_2	$\cdots$	x
$p(X=x)$	p_1	p_2	$\cdots$	

Binomial mean and variance

n Bernoulli trials

p success prob.

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

- $\mathbb{E}X = np. = \sum_{x=0}^{n} x \binom{n}{x} p^x (1-p)^{n-x}$
- $\text{Var}\, X = np(1-p). = \sum_{x=0}^{n} (x - np)^2 \binom{n}{x} p^x (1-p)^{n-x}$

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 $X = \#$ frees throws made in 4 attempts.
2. If you shoot 1000 free throws, how many do you "expect" to make?

① Draw 5.

$$P(\text{Draw 2 who vape}) = \frac{\binom{10}{2}\binom{90}{3}}{\binom{100}{5}} = 0.0702.$$

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

1. Give $P(X = 2)$ if we sample without replacement.
2. Give $P(X = 2)$ if we sample with replacement. $X \sim \text{Binomial}\left(n = 5,\ p = \frac{10}{100}\right)$

$$P(X=2) = \binom{5}{2}(.1)^2(1-.1)^{5-2} = \text{dbinom}(2, 5, .1) = 0.0729$$

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

$\underbrace{\text{pop of 100,}}_{N}$ $\underbrace{\text{10 vape,}}_{M}$ $\underbrace{\text{draw 5.}}_{K}$

$$P(X=2) = \frac{\binom{10}{2}\binom{100-10}{5-2}}{\binom{100}{5}}$$

$$P(X=x) = \frac{\binom{M}{x}\binom{N-M}{K-x}}{\binom{N}{K}}$$

non-red marbles

non-red in hand

N total marbles

M red marbles

Grab K marbles

X = # red marbles in hand.

$$P(X=x) = \frac{\binom{M}{x}\binom{N-M}{K-x}}{\binom{N}{K}}$$

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

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

Use R functions `dhyper()` and `phyper()` to compute probabilities for X:

\# red

\# non-red

$$P(X = x) = \texttt{dhyper(x,m,n,k)}$$
$$P(X \leq x) = \texttt{phyper(x,m,n,k)},$$

where `m` is our M, `n` is our $N - M$, and `k` is our K.

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

hypergeometric

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.

Binomial

1 $X \sim \text{Hypergeometric}(N = 100,\ M = 10,\ K = 5)$

$$\text{dhyper}(x, 10, \overset{N-M}{\overset{\downarrow}{90}}, 5)$$

$X \sim$ Hypergeometric $(N=100,\ M=10,\ K=5)$

x	0	1	2	3	4	5
$P(X=x)$	0.5834	0.3394	0.0702			
$P(X\le x)$	0.5834	.9231	.9934			

$P(X \le 0)$

cumulative probs

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

dbinom$(x,\ 5,\ 0.1)$.

x	0	1	2	3	4	5
$P(X=x)$	0.5905	0.328	0.0729	...		

Hypergeometric mean and variance

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

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

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 $x = 0, 1, 2, 3, 4$. $N = 52$ $M = 13$ $K = 4$
2. Give $\mathbb{E}X$.

①

x	0	1	2	3	4
$P(X=x)$	.304	.439	.213	0.041	0.003

(In R make x <- 0:4)

$$P(X=x) = \frac{\left(\begin{array}{c}\text{\# ways}\\ \text{to choose } x \text{ ♠'s}\end{array}\right)\left(\begin{array}{c}\text{\# ways to choose}\\ 4-x \text{ non ♠'s}\end{array}\right)}{\binom{52}{4}}$$

$$= \frac{\binom{13}{x}\binom{52-13}{4-x}}{\binom{52}{4}}$$

② $\mathbb{E}X = K\frac{M}{N} = 4 \cdot \frac{13}{52} = 1$

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

$X \sim \text{Hypergeometric}(N=100,\ M=10,\ K=5)$

x	0	1	2	3	4	5
$P(X=x)$	0.5834	0.3394	0.0702			

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

Hypergeom.

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.

Binomial.

① $X \sim \text{Hypergeometric}(N = 10{,}000,\ M = 1000,\ K = 5)$

$$P(X = x) = \frac{\binom{1000}{x}\binom{10000-1000}{5-x}}{\binom{10\,000}{5}}$$

x	0	1	2	3	4	5
$P(X=x)$	.590	.328	0.073	0.008	0.000	0.000

SAME to 3 decimal places.

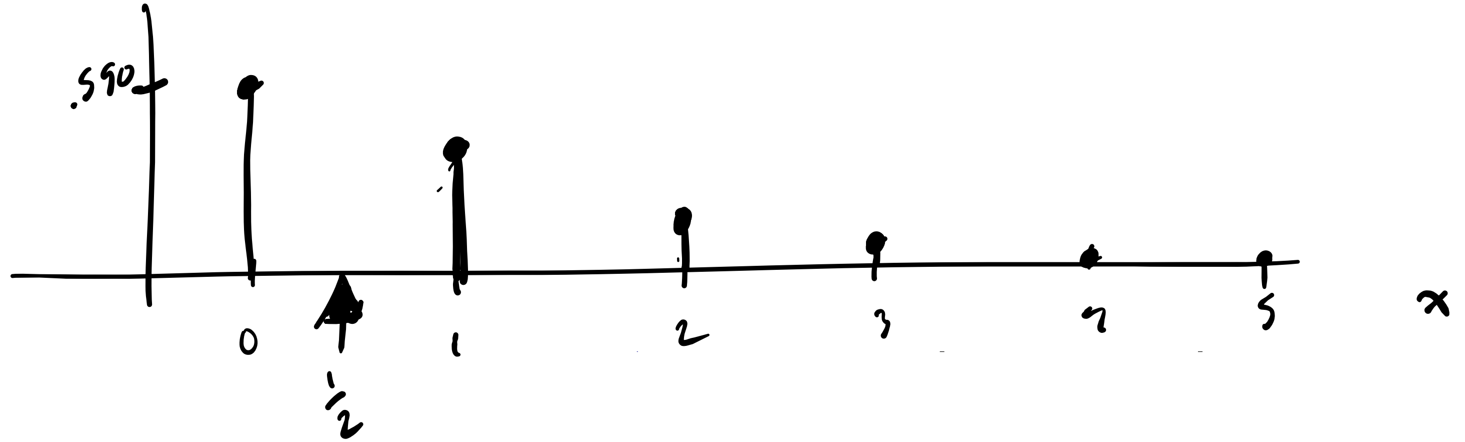
② $X \sim \text{Binomial}\left(n = 5,\ p = \frac{1}{10}\right)$

$$P(X = x) = \binom{5}{x}\left(\frac{1}{10}\right)^x\left(1 - \frac{1}{10}\right)^{5-x}$$

x	0	1	2	3	4	5
$P(X=x)$	.590	.328	0.073	0.008	0.000	0.000

$$\mathbb{E}X = np = 5 \cdot \frac{1}{10} = \frac{1}{2}$$

0



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.

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

Probability mass function

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

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

"not in"

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

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