STAT 515 Lec 05 slides

Bernoulli trials, binomial and hypergeometric distributions

Karl 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". $0 < p < 1$

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??

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

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

"Bernoulli" random variable

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

"is distributed as", "follows"

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

"Let the random variable X have the Bernoulli dist with success prob. p."

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

$x=0$: $\quad p^{0}(1-p)^{1-0} = 1-p$

$x=1$: $\quad p^{1}(1-p)^{1-1} = p$

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

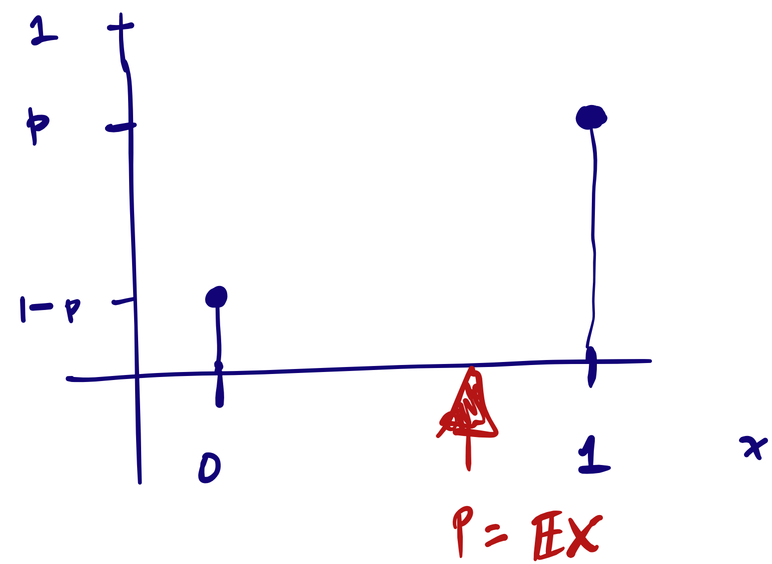
The probabilities $P(X = x)$ are given by

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

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

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

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



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

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

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

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.

$S = \{$ HHHH, HHHM, HHMH, HMHH, MHHH, HHMM, HMHM, HMMH, MHHM, MHMH, MMHH, MMMH, MMHM, MHMM, HMMM, MMMM $\}$

$(0.7)^4$

each one: $(0.7)^3(0.3)$

(0.3) ← $1-0.7$

$(0.7)^2(0.3)^2$

$(0.7)^1(0.3)^3$

$(0.3)^4$

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

same

$n=4$

$\binom{4}{0}(.7)^0(.3)^{4-0}$ $\quad \binom{4}{1}(.7)^1(.3)^{4-1}$ $\quad \binom{4}{2}(.7)^2(.3)^{4-2}$

$1 \qquad \frac{4!}{1!(4-1)!}=4 \qquad \frac{4!}{2!(4-2)!}=\frac{4!}{2!2!}=\frac{4\cdot3\cdot2\cdot1}{2\cdot1\cdot2\cdot1}=6$

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

$$\binom{n}{x}=\frac{n!}{x!(n-x)!} \qquad \binom{n}{0}=\frac{n!}{0!(n-0)!}=\frac{n!}{0!\,n!}=\frac{1}{0!}=1$$

$X =$

0	MMMM
1	MMMH MMHM MHMM HMMM
2	HHMM HMMH MHMH HMHM MHHM MMHH
3	HHHM HHMH HMHH MHHH
4	HHHH

n = # Bernoulli trials (independent)

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

$x \geq$

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

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

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

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

On your own.

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

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 $P(X \leq 6)$.
4. Find $P(X > 6)$.

$n = 10 \quad p = 0.7$

(1) $P(X=3) = \binom{10}{3}(0.7)^3(1-0.7)^{10-3} = 0.009$

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

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

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

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

$$\approx 1$$

(3) $P(X \leq 6) = P(X=0) + P(X=1) + \dots + P(X=6)$

$$= \sum_{x=0}^{6} P(X=x)$$

$$= \sum_{x=0}^{6} \binom{10}{x}(0.7)^x(1-0.7)^{10-x}$$

R

$$= 0.350$$

Show me this kind of expression on an exam.

$$P(X=x) = \binom{n}{x} p^x (1-p)^{n-x} \quad \text{for } x = 0, 1, \dots, n$$

$$\mathbb{E}X = \sum_{x=0}^{n} x \cdot \binom{n}{x} p^x (1-p)^{n-x} = \underset{\text{(steps omitted)}}{\dots} = np.$$

$$\text{Var } X = \sum_{x=0}^{n} (x-\mu)^2 \binom{n}{x} p^x (1-p)^{n-x} = \dots = np(1-p)$$

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.

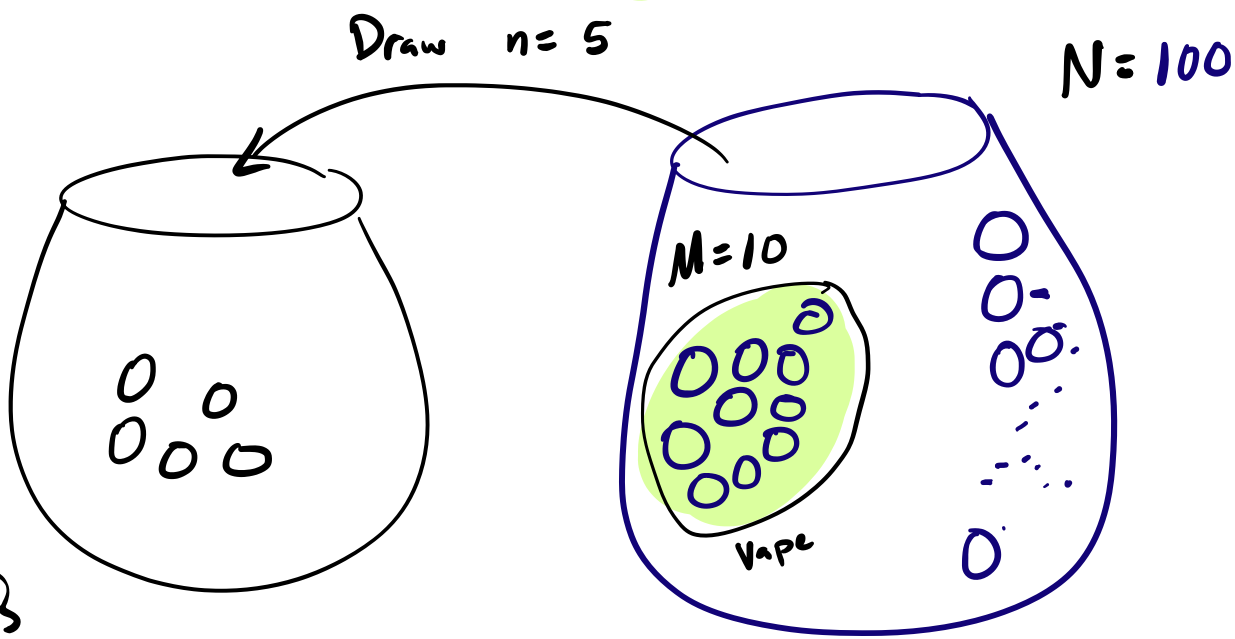
1. Compute $\mathbb{E}X$ when $X = \#$ frees throws made in 10 attempts.
2. If you shoot 1000 free throws, how many do you "expect" to make? 700

1 $\mathbb{E}X = np = 10 \cdot (0.7) = 7$

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

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

1



$X = \#$ ppl I draw who vape.

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

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

$$P(X=3) = \frac{\#\{\text{ways to draw 3 who vape}\}}{\#\{\text{ways to draw 5 from 100}\}}$$

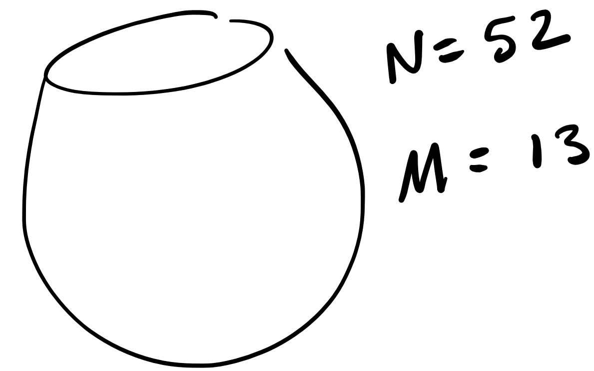
select 3 who vape

$$= \frac{\binom{10}{3}\binom{100-10}{5-3}}{\binom{100}{5}} = 0.0064$$

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

If we sample n from pop. of size N, M of whom vape, without replacement, then

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



Exercise: Draw 5 cards from a 52-card deck and let $X = \#\spadesuit$s in hand.

1. Tabulate $P(X = x)$ for $x = 0, 1, 2, 3, 4, 5$.
2. What if you replace each card in the deck and re-shuffle before drawing again?

(1) $$P(X=x) = \frac{\binom{13}{x}\binom{52-13}{5-x}}{\binom{52}{5}}$$

x	0	1	2	3	4	5
$P(X=x)$			see	R	code	

② $X = \#$ spades. $X \sim \text{Binomial}\left(n=5,\ p=\frac{13}{52}\right)$

$$P(X=x) = \binom{5}{x}\left(\frac{13}{52}\right)^{x}\left(1-\frac{13}{52}\right)^{5-x}$$

Hypergeometric distribution

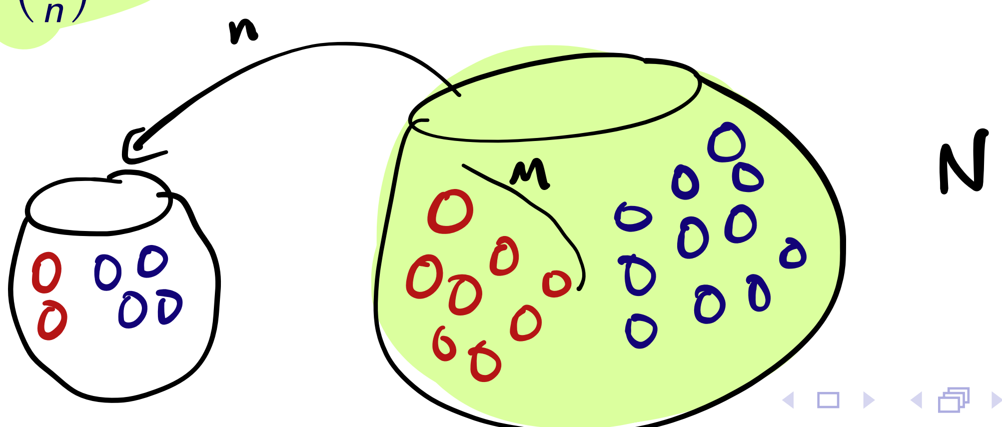
Draw $n \geq 0$ marbles from a bag of $N \geq 0$ marbles, of which $M \geq 0$ are red. If $X = \#$ red marbles drawn, then X has the *Hypergeometric distribution*.

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

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

$x = 0, 1, \ldots, n$

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



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

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

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

May be simpler to compute the probabilities as

$$P(X = x) = \frac{\texttt{choose(M,x)*choose(N-M,n-x)}}{\texttt{choose(N,n)}}.$$

Hypergeometric mean and variance

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

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

Discuss how we would get these expressions.

$n=5$

$N=520$

$M=130$

Exercise: Mix ten 52-card decks together. Then draw 5 cards from the combined deck and let $X = \#\spadesuit$s in hand.

1. Tabulate $P(X = x)$ for $x = 0, 1, 2, 3, 4, 5$.
2. Give $\mathbb{E}X$.
3. What if you replace each card in the deck and re-shuffle before drawing again?

① $$P(X=x) = \frac{\binom{130}{x}\binom{520-130}{n-x}}{\binom{520}{5}}$$

$$P(X \le 3) = P(X=0) + P(X=1) + P(X=2) + P(X=3)$$

$$= \sum_{x=0}^{3} P(X=x)$$

$$= \sum_{x=0}^{3} \frac{\binom{130}{x}\binom{520-130}{5-x}}{\binom{520}{5}}$$

$= 0.985$ (see R code).

(2)

$$P(X=x) = \binom{5}{x}\left(\frac{13}{52}\right)^{x}\left(1-\frac{13}{52}\right)^{5-x} = \text{dbinom}(x, 5, 1/4)$$

$$P(X \le 3) = \sum_{x=0}^{3} \binom{5}{x}\left(\frac{13}{52}\right)^{x}\left(1-\frac{13}{52}\right)^{5-x}$$

$= 0.984$ (see R code)

$= \text{pbinom}(3, 5, 1/4)$

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

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

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

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.

HW 3

Q3 (d)

$N = 100$

$n = 10$

X = # defectives in sample.

$P(\text{No defects}) = P(X=0) \geq 0.90$

$$P(X=0) = \frac{\binom{d}{0}\binom{100-d}{10-0}}{\binom{100}{10}} =$$

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 $X \sim p$.

Give Bernoulli, Binomial, and hypergeometric pmfs.