Special discrete distributions

– Math 161a, Spring 2019, San Jose State University

Prof. Guangliang Chen

February 28, 2019

Outline

Our treatment plan for each distribution

- Examples
- Definition (including pmf)
- Expected value
- Variance
- Other useful properties (if any)

Bernoulli distribution

Consider the following experiments:

Ex 0.1 (Toss a fair coin). $X = 1$ (heads) and 0 (tails).

Ex 0.2 (Randomly select a ball from an urn that has 10 red and 20 green balls). Let $Y = 1$ (if the selected ball is red) and 0 (otherwise).

Ex 0.3 (Randomly select an individual from a population 40% of which have certain characteristic). Let $Z = 1$ (if the selected individual has the characteristic) and 0 (otherwise).

These experiments all share the following traits:

- There is only **one trial**;
- It has only **two outcomes**, "success" or "failure";
- The **probability of getting a success is some number** p;
- X **is a indicator variable**: $X = 1$ (success) or 0 (failure)

We say that such a random variable has a **Bernoulli distribution with parameter** p, and denote it as $X \sim \text{Bernoulli}(p)$.

Such an experiment is called a **Bernoulli trial**.

Ex 0.4 (Toss a fair coin). $X = 1$ (heads) and 0 (tails).

Answer: $X \sim$ Bernoulli($\frac{1}{2}$)

Ex 0.5 (Randomly select a ball from an urn that has 10 red and 20 green balls). Let $Y = 1$ (if the selected ball is red) and 0 (otherwise).

Answer: $Y \sim$ Bernoulli($\frac{1}{3}$)

Ex 0.6 (Randomly select an individual from a population 40% of which have certain characteristic). Let $Z = 1$ (if the selected individual has the characteristic) and 0 (otherwise).

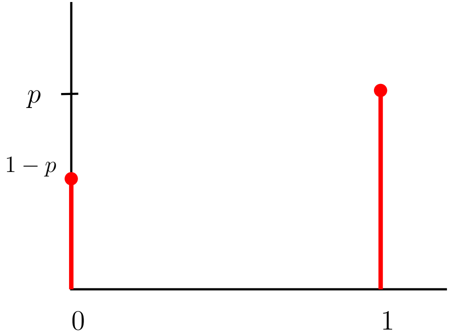
Answer: $Z \sim$ Bernoulli(0.4)

Clearly, if a discrete random variable X follows a Bernoulli distribution with parameter p, then its pmf has the following form (and vice versa):

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

(and $f_X(x) = 0$ for all other x)

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



Theorem 0.1. *Let $X \sim$ Bernoulli(p). Then*

$$E(X) = p$$
$$Var(X) = p(1-p)$$

Proof. We have already obtained them in class. $\square$

Binomial

The following experiments are identical in nature:

- (Toss a fair coin 10 times) $X = \#$heads
- (Answer 10 multiple-choiced questions by random guessing) $X=$ $\#$correctly answered questions
- (Select with replacement 10 balls at random from an urn containing 30 red and 20 blue balls) $X=\#$ red balls obtained

We make the following abstraction:

- There are n **repeated trials** in the experiment
- **Each trial has only two outcomes**, "success" and "failure"
- The **probability** p **of getting successes is fixed**
- The n Bernoulli trials are **independent**
- X denotes the **total number of successes**

In short, X counts the total number of successes in n independent Bernoulli trials with fixed probability of success p.

$$\underset{1}{\underline{\qquad}}\quad\underset{2}{\underline{\qquad}}\quad\underline{\qquad}\quad\underline{\qquad}\quad\underline{\qquad}\quad\underline{\qquad}\quad\underset{n}{\underline{\qquad}}$$

In the above scenario, we say that X follows a **binomial distribution with parameters** n, p, and write $X \sim B(n, p)$

Example.

- (Toss a fair coin 10 times) $X = \#$heads $\quad B(10, \frac{1}{2})$
- (Answer 10 multiple-choiced questions by random guessing) $X=$ $\#$correctly answered questions $\quad B(10, \frac{1}{4})$
- (Draw with replacement 10 balls from an urn containing 30 red and 20 blue balls at random) $X=\#$ red balls obtained $\quad B(10, 0.6)$

Question. In the last example, if balls are drawn without replacement, is X still a binomial random variable? Why?

Theorem 0.2. *The pmf of $X \sim B(n, p)$ is*

$$f_X(x) = \binom{n}{x} p^x (1-p)^{n-x}, \qquad x = 0, 1, \ldots n.$$

	H	H			H	
1	2					n

How to understand this result:

- $\binom{n}{x}$: # ways of having x successes in n trials
- p^x: probability of having exactly x successes
- $(1-p)^{n-x}$: probability of having exactly $n-x$ failures

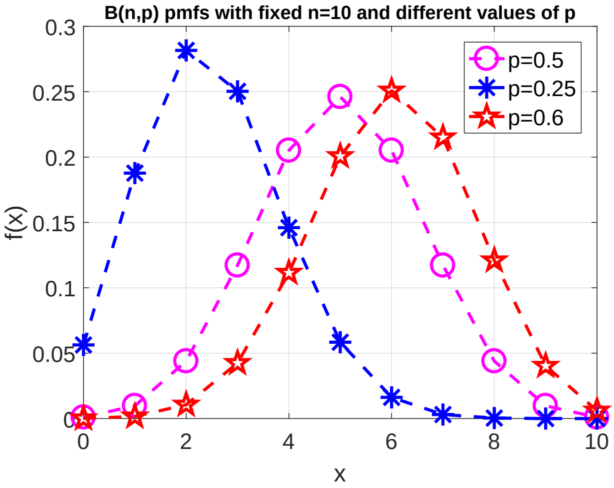
Ex 0.7. What is the probability of getting 0 heads in 10 independent flips of a fair coin? Exactly 1 head? At least two heads?

Ex 0.8 (Answer 10 multiple-choiced questions by random guessing)**.** Let $X=$ #correctly answered questions. Find $P(X = x)$ for $x = 0, 2, 9$.

Binomial probabilities (continued)

Entry is $P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$

		p								
n	k	.10	.15	.20	.25	.30	.35	.40	.45	.50
9	0	.3874	.2316	.1342	.0751	.0404	.0207	.0101	.0046	.0020
	1	.3874	.3679	.3020	.2253	.1556	.1004	.0605	.0339	.0176
	2	.1722	.2597	.3020	.3003	.2668	.2162	.1612	.1110	.0703
	3	.0446	.1069	.1762	.2336	.2668	.2716	.2508	.2119	.1641
	4	.0074	.0283	.0661	.1168	.1715	.2194	.2508	.2600	.2461
	5	.0008	.0050	.0165	.0389	.0735	.1181	.1672	.2128	.2461
	6	.0001	.0006	.0028	.0087	.0210	.0424	.0743	.1160	.1641
	7			.0003	.0012	.0039	.0098	.0212	.0407	.0703
	8				.0001	.0004	.0013	.0035	.0083	.0176
	9						.0001	.0003	.0008	.0020
10	0	.3487	.1969	.1074	.0563	.0282	.0135	.0060	.0025	.0010
	1	.3874	.3474	.2684	.1877	.1211	.0725	.0403	.0207	.0098
	2	.1937	.2759	.3020	.2816	.2335	.1757	.1209	.0763	.0439
	3	.0574	.1298	.2013	.2503	.2668	.2522	.2150	.1665	.1172
	4	.0112	.0401	.0881	.1460	.2001	.2377	.2508	.2384	.2051
	5	.0015	.0085	.0264	.0584	.1029	.1536	.2007	.2340	.2461
	6	.0001	.0012	.0055	.0162	.0368	.0689	.1115	.1596	.2051
	7		.0001	.0008	.0031	.0090	.0212	.0425	.0746	.1172
	8			.0001	.0004	.0014	.0043	.0106	.0229	.0439
	9					.0001	.0005	.0016	.0042	.0098
	10							.0001	.0003	.0010


B(n,p) pmfs with fixed n=10 and different values of p
p=0.5
p=0.25
p=0.6
f(x)
x
0.3
0.25
0.2
0.15
0.1
0.05
0
0
2
4
6
8
10

Theorem 0.3. *Let* $X \sim B(n, p)$. *Then*

$$E(X) = np,$$
$$Var(X) = np(1-p)$$

Proof. We have already proved this result in class. □

Hypergeometric

Ex 0.9. Draw 10 balls at random from an urn containing 30 red and 20 blue balls, and let $X = \#$ red balls.

We obtained that $X \sim B(10, 0.6)$ **if the experiment is performed with replacement**.

We also mentioned that X **is not binomial if it is without replacement.**

In fact, in the latter case, X has the following pmf:

$$f_X(x) = \frac{\binom{30}{x}\binom{20}{10-x}}{\binom{50}{10}}, \quad x = 0, 1, \ldots, 10$$

More generally, consider the following example:

Ex 0.10. Draw, without replacement, n balls at random from an urn containing r red and $N-r$ blue balls. Let $X = \#$ red balls. Then the pmf of X is

$$f_X(x) = \frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}}, \quad x = 0, 1, \ldots, n \qquad (x \le r,\ n-x \le N-r)$$

How to understand this result:

- $\binom{r}{x}$: #ways of choosing x red balls out of r
- $\binom{N-r}{n-x}$: #ways of choosing $n-x$ blue balls out of $N-r$
- $\binom{N}{n}$: #ways of choosing n balls out of N in total (ignoring color)

Def 0.1 (HyperGeom(N, r, n))**.** We say that the above random variable X follows a *hypergeometric* distribution, and write $X \sim$ HyperGeom(N, r, n).

Ex 0.11. In the previous example, $X \sim$ HyperGeom$(50, 30, 10)$.

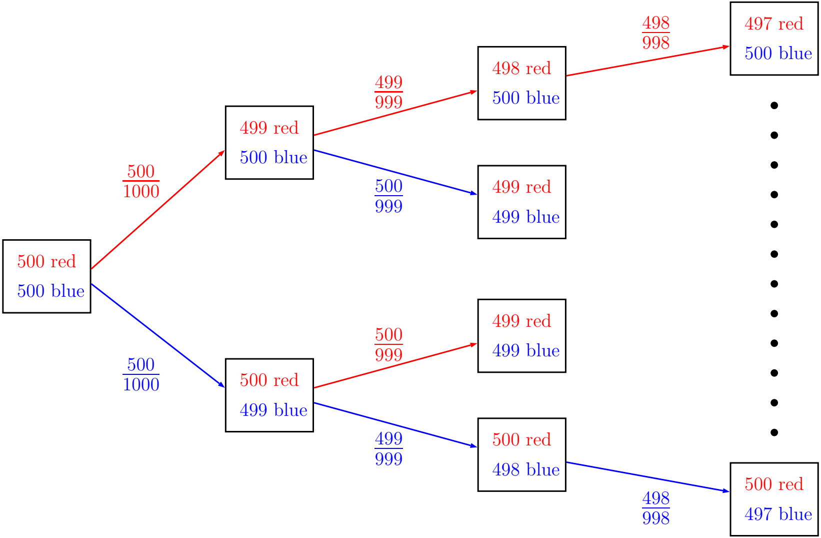
Ex 0.12. Draw without replacement n voters at random from the whole pool of N that are registered, r of which support certain presidential candidate. Let $X =$ #selected supporters of the candidate. Then $X \sim$ HyperGeom(N, r, n).

The hypergeometric pmf has a complicated formula, but in certain case, it can be well approximated by the binomial pmf.

Theorem 0.4. *When N, r are both large (relative to n), then*

$$HyperGeom(N, r, n) \approx B(n, p = \frac{r}{N}).$$

Remark. In the last example, both r and N are large. The theorem implies that, if $r/N = 0.4$ and $n = 500$, then the number of voters (among the 500 selected) that support the candidate is approximately binomial: $X \overset{\text{approx}}{\sim} B(500, 0.4)$.


500 red
500 blue
$\frac{500}{1000}$
$\frac{500}{1000}$
499 red
500 blue
500 red
499 blue
$\frac{499}{999}$
$\frac{500}{999}$
$\frac{500}{999}$
$\frac{499}{999}$
498 red
500 blue
499 red
499 blue
499 red
499 blue
500 red
498 blue
$\frac{498}{998}$
$\frac{498}{998}$
497 red
500 blue
500 red
497 blue

Ex 0.13. Select 5 balls at random from an urn containing 300 red and 200 blue balls, and let $X=$ $\#$selected red balls. Find both the exact probability and its binomial approximation of $P(X = 3)$.

Exact answer (by using hypergeometric): 0.3473, *binomial approx.*: 0.3456

Theorem 0.5. *Let* $X \sim$ *HyperGeom*(N, r, n) *and* $p = \frac{r}{N}$. *Then*

$$E(X) = \frac{nr}{N} = np$$

$$Var(X) = np(1-p)\left(\frac{N-n}{N-1}\right)$$