* Announcement:

① HW3 due Tue (9/15)

② Study group survey by 5pm Tue (9/15)

③ Quiz 2 on Thu (9/17) - Chapter 3

④ OH: Wed 2-4pm ⇝ Tue 5-7pm
(only this week)

STAT 88: Lecture 8

Contents

Section 3.5: Examples
Section 4.1: Cumulative Distribution Function (CDF)

Last time

<u>Sec 3.4</u> The Hypergeometric distribution:

Picture:

$N = 9 + 4 = 13$

$\underbrace{\text{H H H H H H H H H}}_{G=9}$ $\underbrace{\text{T T T T}}_{B=4}$

↓ draw n w/o replacement

$g=3,\ b=2 \qquad n = g+b = 5$

HHH TT

Chance of g good in sample:

$$P(X = g) = \frac{\binom{G}{g}\binom{N-G}{n-g}}{\binom{N}{n}},$$

or in this case

$$\frac{\frac{9!}{3!6!} \cdot \frac{4!}{2!2!}}{\frac{13!}{5!8!}} = \frac{\overset{G}{\binom{9}{3}}\overset{B}{\binom{4}{2}}}{\binom{13}{5}} = \binom{5}{3}\overset{H}{\frac{9}{13}}\,\overset{H}{\frac{8}{12}}\,\overset{H}{\frac{7}{11}}\,\overset{T}{\frac{4}{10}}\,\overset{T}{\frac{3}{9}}$$

$\binom{n}{k} = \frac{n!}{k!(n-k)!}$

Simplify ↗

Notice if draw w/ replacement this is $\binom{5}{3}\left(\frac{9}{13}\right)^3\left(\frac{4}{13}\right)^2$ "Binomial formula"

Warm up: Three cards are dealt from a standard 52 card deck. Find the chance:

(a) The first card is red and the second two black.

(b) Exactly one of the cards dealt is red.

(c) $\boxed{\text{At least}}$ one of the cards dealt is red.

(a) $P(\text{1st R}\underset{\text{and}}{,}\ \text{2nd B}\underset{\text{and}}{,}\ \text{3rd B}) = P(\text{1st R} \cap \text{2nd B} \cap \text{3rd B})$

$\binom{n}{k}$ = n choose k $= \dfrac{n!}{k!(n-k)!}$

Mult? — $= P(\text{1st R})\,P(\text{2nd B} \mid \text{1st R})\,P(\text{3rd B} \mid \text{1st R, 2nd B})$

$= \dfrac{26}{52} * \dfrac{26}{51} * \dfrac{25}{50} \quad := P$

(b) (i) X = # red cards out of 3.

$\sim HG(N=52,\ G=26,\ n=3)$

$$P(X=1) = \frac{\binom{26}{1}\binom{26}{2}}{\binom{52}{3}} = 0.382$$

(ii) $P(\text{one is R}) = P(RBB) + P(BRB) + P(BBR) = 3 * P$

3 different ways to arrange RBB $\rightarrow$ $= \underline{\binom{3}{1}} * P$

$= 0.382$

$$P(X=0) = \frac{\binom{26}{0}\binom{26}{3}}{\binom{52}{3}}$$

" $1 - P(X=1) - P(X=2) - P(X=3)$

(c) $P(\text{At least one is R}) \underset{\text{Complement}}{=} 1 - P(\text{all 3 B})$

$= 1 - P(\text{1st B, 2nd B, 3rd B})$

$= 1 - \dfrac{26}{52} * \dfrac{25}{51} * \dfrac{24}{50}$

≈ 0.882

3.5. Examples (continued)

Fisher Exact Test In a randomized controlled experiment with 100 participants, 60 participants are in the treatment group and 40 are in the control group. In the treatment group, 50 out of the 60 participants recover after the treatment. In the control group, 30 out of the 40 participants recover.

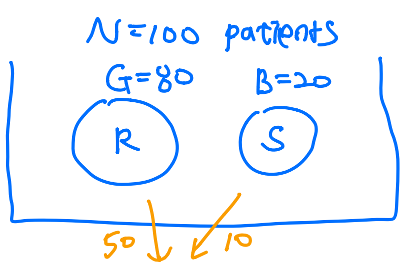
A total of 80 patients recovered out of 100.

Question. Suppose the treatment is not effective. What is the chance that 50 or more of the recovered patients are randomly assigned to the treatment group?

(if the answer is really small, then the treatment is probably effective)

Start with: What is the chance that 50 of the recovered patients are randomly assigned to the treatment group?

"If treatment group has no special effect and in fact acted like control group, then 80 patients who recovered would have recovered in whichever group they were assigned to!"



draw n=60 w/o replacement
↓
assign to treatment group.

X = # recovered patients in treatment group.

$$\begin{cases} N=100 \\ G=80 \\ n=60 \end{cases} \Rightarrow X \sim HG(100, 80, 60)$$

$$P(X \geq 50) = \sum_{g=50}^{60} \frac{\binom{80}{g} \cdot \binom{20}{60-g}}{\binom{100}{60}}$$

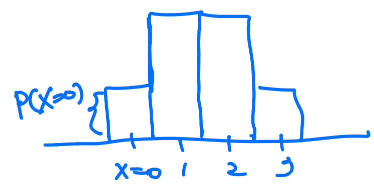
$$\approx \boxed{22\%}$$

4.1. Cumulative Distribution Function (CDF)

To specify a probability distribution, we have used a probability mass function (pmf):

Example: $X \sim \text{Binomial}(3, 1/2)$.

x	0	1	2	3
$P(X=x)$	$1/8$	$3/8$	$3/8$	$1/8$



You can also specify a probability distribution by giving the chance that the value of X is at most x, $F(x) = P(X \leq x)$. This is called the cummulative distribution function (CDF).

$0 \leq F(x) \leq 1$

Example: $X \sim \text{Binomial}(3, \frac{1}{2})$ $\quad n=3$, $p=\frac{1}{2}$

x	0	1	2	3
$P(X=x)$	$1/8$	$3/8$	$3/8$	$1/8$
$F(x)$	$1/8$	$4/8$	$7/8$	1

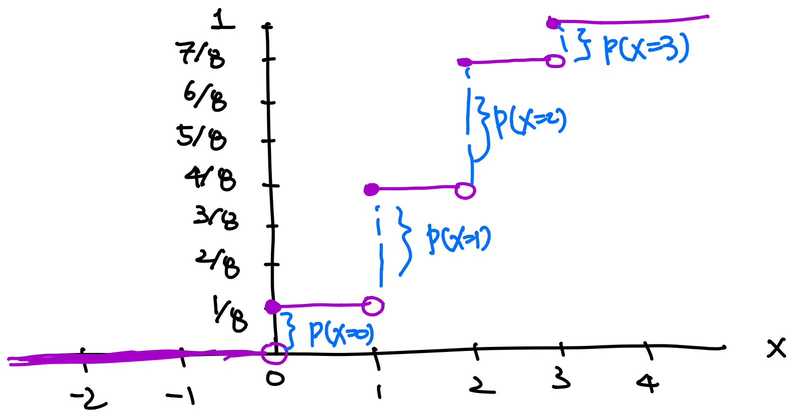
$F(x) = P(X \leq x)$

$F(0) = P(X \leq 0) = P(X=0)$

$F(1) = P(X \leq 1) = P(X=0) + P(X=1)$

$$F(x) = \begin{cases} 0, & x < 0 \\ 1/8, & 0 \leq x < 1 \\ 4/8, & 1 \leq x < 2 \\ 7/8, & 2 \leq x < 3 \\ 1, & x \geq 3 \end{cases}$$

Graph of the CDF We can define $F(x) = P(X \le x)$ on the entire x-axis even though it only "jumps" at $x = 0, 1, 2, 3$.



$$F(x) = \begin{cases} 0, & x < 0 \\ 1/8, & 0 \le x < 1 \\ 4/8, & 1 \le x < 2 \\ 7/8, & 2 \le x < 3 \\ 1, & x \ge 3 \end{cases}$$

$$F(1) = P(X \le 1) = P(X=0) + P(X=1)$$

Why does the CDF specify the distribution?

$$P(X = x) = P(X \le x) - P(X \le x - 1) = F(x) - F(x-1).$$

So knowing $F(x)$ tells us $P(X = x)$.

$\Rightarrow$ $F(x)$ and $P(X=x)$ gives the same info.

Why is CDF useful? Solutions to many problems can be expressed in terms of CDF and Python has built-in CDF function.