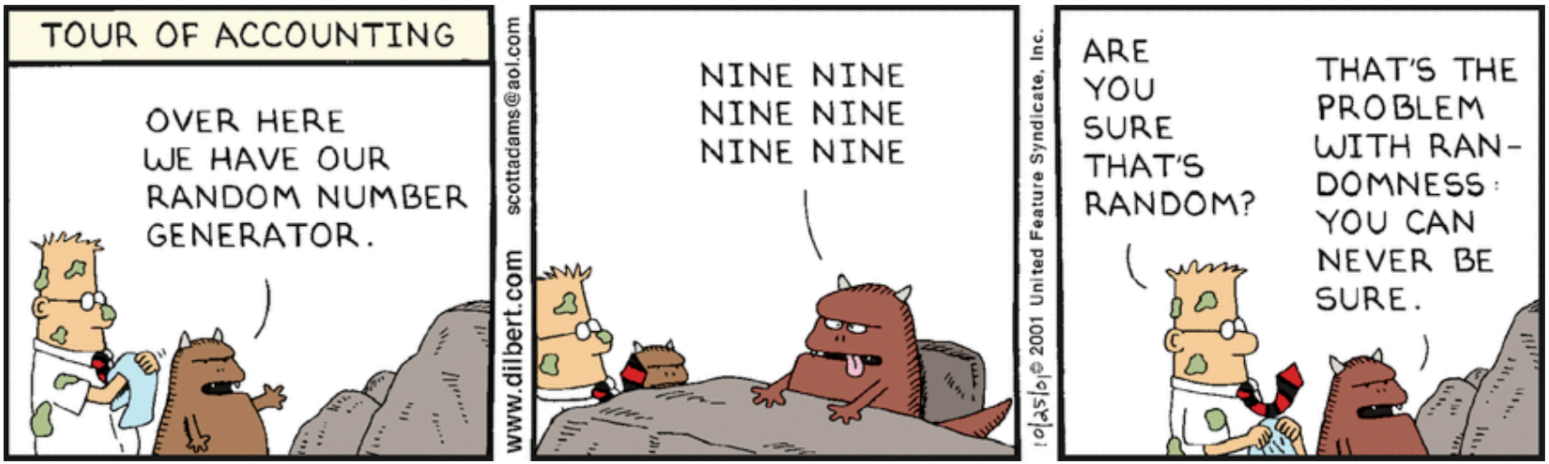
Stat 88: Probability & Mathematical Statistics in Data Science



Lecture 9: 2/8/2021

Examples, CDF, waiting times

Sections 3.5, 4.1,4.2

Agenda

- 3.5: Examples
- 4.1 The cumulative distribution function
- 4.2 Waiting times

Problem solving techniques

- See if problem can be broken into smaller problems
- See which distribution applies to the situation
- Identify the parameters
- Use the addition and multiplication rules carefully

An advisor at a university provides guidance to 10 students. Each student has to meet with her once a month during the school year which consists of nine months.

Each month the advisor schedules one day of meetings. Each student has to sign up for one meeting that day. Students have the choice of meeting her in the morning or in the afternoon.

Assume that every month each student, independently of other students and other months, chooses to meet in the afternoon with probability 0.75.

What is the chance that she has both morning and afternoon meetings in all of the months except one?

Advisors and their students

- Need to figure out a random variable. First fix one month, any month.
- Figure out the chance in that month, *all* the students choose the afternoon OR *all* the students choose the morning: this would mean that the meetings happen *only* in the morning OR *only* in the afternoon.
- We need the chance of the complement of this event.
- Where is the random variable?

Say in a particular month, figure out the chance of the advisor having meetings both in the AM and in the PM.

If she has meetings both in the morning & the afternoon this means that at least one student chose the morning AND at least one the afternoon

P(at least one student in AM $\underline{\underline{\&}}$ at least 1 in the afternoon)

$= 1 - P(\underbrace{\text{no one in the AM}}_{A}\ \underline{\text{OR}}\ \underbrace{\text{no one in the afternoon}}_{B})$

$= 1 - \left(P(A) + P(B) - \cancel{P(AB)}^{0}\right)$ b/c A & B are disjoint

$= 1 - P(A) - P(B) = \left.1 - (0.75)^{10} - (0.25)^{10}\right\} p$

Let X = # of months in which she has both morning & afternoon meetings

$X \sim Bin(9, p)$, We want $P(X=8)$

$$P(X=8) = \binom{9}{8} p^{8} (1-p)^{1}$$

Randomized Controlled Experiments

Two randomized controlled experiments are being run independently of each other. In each experiment, a simple random sample of **half** the participants will be assigned to the treatment group and the other half to control. Expt 1 has 100 participants of whom 20 are men. Expt 2 has 90 participants of whom 30 are men.

What is the chance that the treatment and control groups in Experiment 1 contain the same number of men?

T_1 = # of men in treatment group of expt 1

If treatment & ctrl gps have the same # of men, must be 10 in each.

$$P(T_1 = 10) = \frac{\binom{20}{10}\binom{80}{40}}{\binom{100}{50}}$$

$k=10$

$$T_1 \sim HG(\underset{N}{100}, \underset{G}{20}, \underset{n}{50})$$

Exercise 1: Put this into Python & check text answer.

Exercise ②: Pretend its binomial, $P(S) = \frac{2}{10}$ & see what $P(X=10)$ is. $n=50$

Problems, continued

What is the chance that the treatment groups in the two experiments have the **same** number of men?

- Notice this is a bit tricky. There are many disjoint cases (each of the treatment groups has 1 man, or 2 men or 3 men etc. What is the max?
- We will have to split the chance into the chance of each of the cases and add them.
-

T_1 : # of men in trt gp of Expt 1

T_2 : " " " " " 2

$$P(T_1 = T_2) = \sum_{k=0}^{20} P(T_1 = T_2 = k)$$

Did the treatment have an effect?

- RCE with 100 participants, 60 in Treatment, 40 in Control
- T: 50 recover, out of 60 (83%), C: 30 recover out of 40 (75%)

- Suppose treatment had no effect, and these 80 just happened to recover. What is the chance they would have recovered no matter what and 50 were assigned to the treatment group by chance?

Suppose 80 recover. no matter what

Q: What is the prob that 50 of 80 just happened to be assigned to the trt gp?

$N=100, \ G=80, \ n=60$ ← #(Trt)

$X \sim HG(100, 80, 60)$

$$P(X=50) = \frac{\binom{80}{50}\binom{20}{10}}{\binom{100}{60}}$$

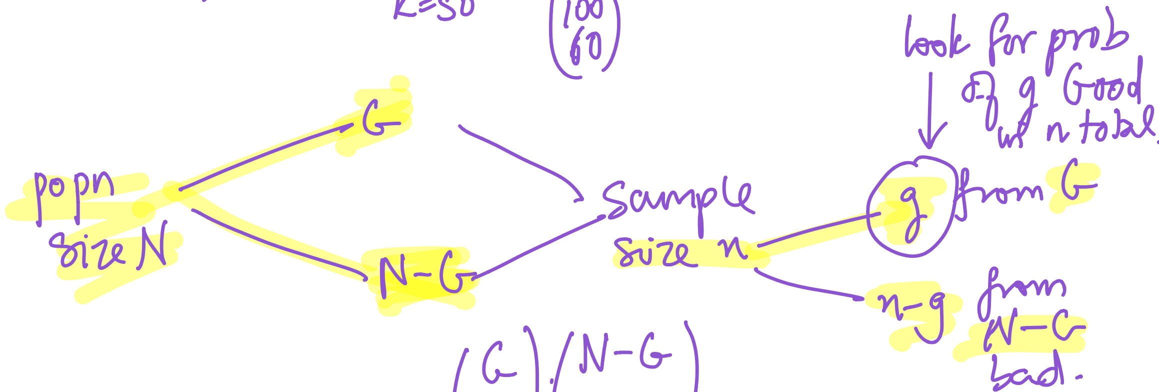
Google: Ronald Fisher & Lady Tasting Tea.

Book looks at $P(X \geq 50)$, $X \sim HG(100, 80, 60)$

This is b/c book sets up a hyp. test & we want P-value.

(Null: Trt has no effect
Alt: More people recover w/ trt)

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



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

Hypergeometric but don't know N

- A state has several million households, half of which have annual incomes over 50,000 dollars. In a simple random sample of 400 households taken from the state, what is the chance that more than 215 have incomes over 50,000 dollars?

How should we do this? n = 400, k = 215, G=N/2, N=???

Use the binomial dsn, $P(S) = \frac{1}{2}$

4.1: Back to random variables and their distributions

$f(x)$ is the prob mass function

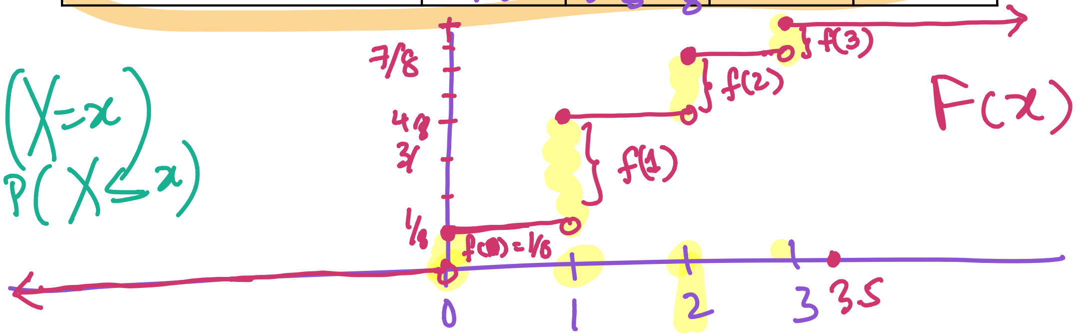
- X, f(x)=P(X=x)
- Consider X = number of H in 3 tosses, then X ~ Bin(3,1/2)
- We can also define a new function **F**, called the ***cumulative distribution function,*** that, for each real number x, tells us how much mass has been accumulated by the time X reaches x.
- $\boldsymbol{F(x) = P(X \leq x) = \sum_{k \leq x} \binom{3}{k} p^k (1-p)^{n-k}}$.

$$F(x) = P(X \leq x) = \sum_{k \leq x} f(k) = \sum_{k \leq x} P(X=k)$$

x	0	1	2	3
$f(x) = P(X = x)$	1/8	3/8	3/8	1/8
$F(x)$	1/8	$\frac{1}{8}+\frac{3}{8}=\frac{4}{8}$		

$f(x) = P(X=x)$

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



F(x) is defined FOR EVERY REAL # x

$F(-27) = 0$, $F(10^7) = 1$, $F(2) = 7/8$

$F(3.5) = 1$

F(x) ⟶ f(x)?

- How to recover the pmf from the cdf? Draw the graph of F(x):

F(x) : jump or step function
- Flat in b/w the jumps
- Right continuous fn.
- Piecewise defined

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

- What are the properties of F(x)? What is its domain? Range?

Domain $(-\infty, \infty)$

Range $[0, 1]$

$$f(x) = \begin{cases} 1/8, & x = 0 \\ 3/8, & x = 1 \\ 3/8, & x = 2 \\ 1/8, & x = 3 \end{cases}$$

Exercise 4.5.2

- A random variable W has the distribution shown in the table below. Sketch a graph of the cdf of W.

w	-2	-1	0	1	3
$P(W = w)$	0.1	0.3	0.25	0.2	0.15