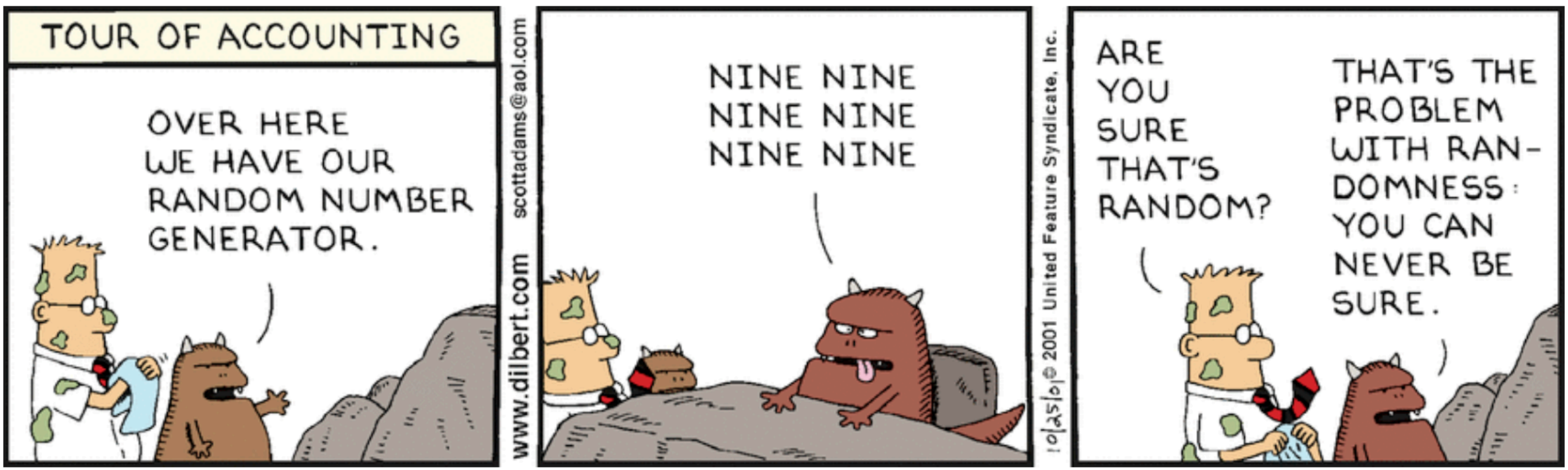
Stat 88: Probability & Math. Stat in Data Science



Lecture 11: 2/9/2024

More examples of binomial and hypergeometric, and cdf

3.5, 4.1

Shobhana Stoyanov

Recap

- When we have two kinds of tickets in a box and we draw tickets at random from this box, each draw is called a ***trial***
- We call the two kinds (binary) outcomes ***Success***, and ***Failure***
- Might be with replacement (like a coin toss) or without replacement (taking a simple random sample of Cal students and checking the number of people who are planning on going to cheer on the rugby team against BYU on Saturday.)
- Random variables (usually denoted by $X, \; Y$ etc) are numbers that ***map*** the outcome space Ω to real numbers, so they inherit a probability distribution.
- The probability distribution of a random variable X, is a description of the values taken by X, and the probabilities that X takes these values.
- The ***probability mass function*** of $X,$ denoted by $f(x)$, is a function that gives, for each value x taken by X, its chance $P(X = x)$.

Recap

- Can think of these problems as ***classifying*** and ***counting***.
- n independent trials each of which can result in one of two outcomes.
- We call these outcomes **S**uccess or **F**ailure, and can represent the random experiment by drawing n tickets **with** replacement from a box with tickets marked 0 or 1, where the proportion of tickets marked 1 is equal to the probability of a success in a trial (p)

$X_k = \begin{cases} 0 & \text{wp } 1-p \\ 1 & \text{wp } p \end{cases}$

- Each draw can be represented by a ***Bernoulli*** random variable, $X_k \sim$ ***Bernoulli***(p)
- If X is the number of successes in n trials, then X is the *sum of draws* from such a box as described above.

F F S F S . . . . . S ← k

- We say that $\boldsymbol{X \sim Bin(n,\ p)}$ and $\boldsymbol{P(X = k) = \binom{n}{k} \times p^k \times (1-p)^{n-k},\ k = 0, 1, \ldots n}$

with repl

- We might also draw ***without*** replacement, in which case, we say that X has the *hypergeometric*$(\boldsymbol{N,\ G,\ n})$ distribution, and

Deal 9 cards from std deck

$$P(\text{exactly } 4 \heartsuit) = \frac{\binom{13}{4}\binom{39}{5}}{\binom{52}{9}}$$

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

size of hand

- ***N*** is the number of tickets in the box, ***G*** is the number of successes possible, ***n*** is the number of draws (fixed beforehand).

Classify the following as binomial, hypergeometric, or neither

1. Number of heads in 12 tosses of a fair coin. $Bin(n = 12, p = \frac{1}{2})$
2. Number of tosses until we see two heads. Neither
3. Number of queens in a five card hand $HG(N = 52, n = 5, G = 4)$
4. Number of Democrats in a simple random sample of 500 adult voters drawn from the SF Bay Area. $HG(N = \text{SF Bay area pop}^n, n = 500, G = \#\text{ of Dems})$

Go to pollev.com/shobhana to answer

Example

- A large supermarket chain in Florida occasionally selects employees to receive management training. A group of women there claimed that female employees were passed over for this training in favor of their male colleagues. The company denied this claim. (A similar complaint of gender bias was made about promotions and pay for the 1.6 million women who work or who have worked for Wal-Mart. The Supreme Court heard the case in 2011 and ruled in favor of Wal-Mart.)
- Suppose that the large employee pool of the Florida chain (more than a 1000 people) that can be tapped for management training is half male and half female. Since this program began, none of the 10 employees chosen have been female. What would be the probability of 0 out of 10 selections being female, if there truly was no gender bias?

We have to make assumptions to make computations

- Method 1: pretend we are sampling with replacement, use Binomial dsn.

(1) No gender bias P(male) = P(Female)

(2) Employee was selected at random with replacement

Let X = # of females selected in 10 trials | $P(X=0) = \binom{10}{0}\left(\frac{1}{2}\right)^0\left(\frac{1}{2}\right)^{10} = \frac{1}{2^{10}} = \frac{1}{1024}$

$X \sim Bin(10, \frac{1}{2})$ $< 0.1\%$

Are we really sampling with replacement?

Probably not, sampling w/o repl. so $P(M) = P(F)$

$$X \sim HG(N=1000, G=500, n=10)$$

$$P(X=0) = \frac{\binom{500}{0}\binom{500}{10}}{\binom{1000}{10}}$$

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.
- What is the random variable?

We will begin with this on Monday.