Stat 88: Probability & Math. Stat in Data Science

MY HOBBY: PLAYING BASKETBALL AGAINST SPACE

https://xkcd.com/2328

Lecture 10: 2/7/2024

The binomial and hypergeometric distributions

3.3, 3.4

Shobhana Stoyanov

Agenda & warm-up

- Warm up

- The binomial distribution

- The hypergeometric distribution

Warm up

- A quiz has 3 multiple choice questions. Each question has 2 possible answers, one of which is correct. A student answers all the questions by guessing at random. Let X be the number of questions the student gets right, and Y the number that the student gets wrong. What is the distribution of the student's score on the exam, if each correct answer is worth 1 point? Note that this value is X.

X is exactly like flipping a coin 3 times
& $X = \#H$

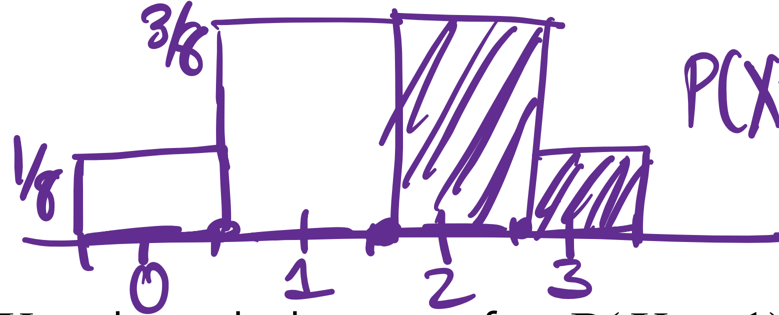
Dsn of X

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

$Y = 3 - X$

- Write down an expression for Y in terms of X, and the distribution of Y. Do X and Y have the same distribution?

Yes



$P(X>1) = \frac{3}{8} + \frac{1}{8} = \frac{4}{8} = \frac{1}{2}$

Probability histograms

- Draw the probability histogram for X and mark the area for $P(X > 1)$. What is the value of this area?

Let X_k = student's score on question k, $k = 1, 2, 3$

$$X_1 = \begin{cases} 0, & \text{with prob } 0.5 \\ 1, & \text{w/p } 0.5 \end{cases}$$

$$X = X_1 + X_2 + X_3$$

X_1, X_2, X_3 are called "BERNOULLI" random variables

$X_1 \sim \text{Bernoulli}(\frac{1}{2})$

"Distributed as"

Let R be $\sim \text{Bernoulli}(0.7)$

$f(x)$ for R

$$f(x) = P(R = x)$$

$$R = \begin{cases} 0 & \text{w.p. } 0.3 \\ 1 & \text{w.p. } 0.7 \end{cases}$$

Recall:

- 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 $P(X = x)$.

3.3 The Binomial distribution

- Many situations can be modeled using the following set up:
  - We have a ***fixed*** number of ***independent*** trials, each of which has ***two*** possible outcomes. “success”(S) and “failure”(F)
  - The probability of success stays ***constant*** from trial to trial.

- Example: toss a coin 10 times, count the number of heads w/ 10 tosses
  - Each toss is an independent trial
  - A success is a head.
  - P(success) = 0.5
    in a single trial

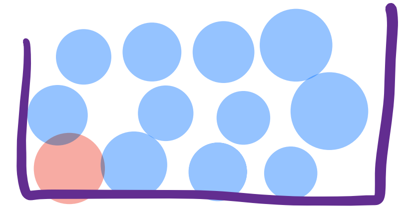
$$X = X_1 + X_2 + X_3 + \dots + X_{10}$$

$$X_k \sim \text{Bernoulli}\left(\tfrac{1}{2}\right)$$

- Need to specify number of trials (***n***), and P(success) (***p***)
  - Example: number of people who accept credit card offer from bank
  - Number of aces in 10 rolls of a die.

$$\left(\tfrac{1}{6}\right)\left(\tfrac{1}{6}\right)\left(\tfrac{1}{6}\right)\left(\tfrac{5}{6}\right) \dots = \left(\tfrac{1}{6}\right)^3 \left(\tfrac{5}{6}\right)^7 \cdot \binom{10}{3}$$

Binomial distribution: Example



- Consider a box with **one** red ball and **eleven** blue ones.

- One draw is made. What is the probability that the ball is red?
  - $n = 1, p = 1/12$
  - $P(R) = 1/12$

S E E E

RBBB
BRBB

- Now 4 draws are made, ***with replacement***. What is the probability that *exactly* 1 draw is red (out of the 4)?
  - Notice that this is like a tossing a coin 4 times, with P(head) = 1/12.

- $P(RBBB) = \left(\frac{1}{12}\right)^1 \left(\frac{11}{12}\right)^3$
- How many such sequences are there? $\binom{4}{1}$
- What is the probability of all such sequences (with1 R, 3B)?

P(1S & 3F) = P(1R & 3B) = P(RBBB) + P(BRBB)

$= \binom{4}{1}\left(\frac{1}{12}\right)^1\left(\frac{11}{12}\right)^3 \quad + P(BBRB) + P(BBBR)$

Binomial distribution: Example

- What if we want to compute the probability of **2** red balls in 4 draws? We need the number of sequences of R and B that have 2 R and 2 B.
- $P(RRBB) =$
- There are 6 such sequences (how?), so if we let X = # of red balls in 4 draws with replacement, we have that

$$\underline{P(X = 2)} = \underline{\binom{n}{k} \times p^2 \times (1-p)^2}$$

$$\binom{4}{2} \cdot \left(\frac{1}{12}\right)^2 \left(\frac{11}{12}\right)^2$$

where $p = P(red)$

- We say that X has the **Binomial distribution with parameters n and p**, and write it as $X \sim Bin(n, p)$ if X takes values $0, 1, \ldots, n$ and

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

k successes in n trials

Characteristics of the binomial distribution

- There are n trials, where n is FIXED beforehand.
- The chance (p) of a success stays the SAME from trial to trial
- Each trial results in either success (S) or failure (F)
- The trials are INDEPENDENT of each other.
- $X \sim Bin(n, p)$, possible values of X: $0, 1, 2, \ldots, n$
- Use python to compute numerical values of probabilities (read section in text, in 3.3)

~~Ex~~

1 R & 5 B balls   Success: R

draw 27 times with replacement.

What is the prob we see exactly 4 R balls

$$\underbrace{\underline{S}\ \_\ \underline{S}\ \_\ \_\ \_\ \_\ \underline{S}\ \_\ \_\ \underline{S}}_{27 \text{ spots}}$$

$P(S) = \frac{1}{6}$

$P(F) = \frac{5}{6}$ Let X = ~~#~~ of S in 27 draws

$X = 0, 1, 2, \ldots, 27$

$$P(X=4) = \binom{27}{4}\left(\frac{1}{6}\right)^4\left(\frac{5}{6}\right)^{27-4}$$

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

Sampling binary outcomes without replacement

- Deck of cards, deal 5, chance of 2 aces in hand? What about chance of 3 hearts in a hand of 5?

$$\frac{\binom{13}{3}\binom{52-13}{2}}{\binom{52}{5}}$$

- 25 balls, 10 red, 15 blue, pick 5 w/o repl. Chance of 2 red balls?

$$\frac{\binom{10}{2}\binom{15}{3}}{\binom{25}{5}}$$

Box with 27 success
33 failure $\boxed{S}$ $\boxed{F}$
Draw 10 times
w/o repl

$$P(6 \text{ succ}) = \frac{\binom{27}{6}\binom{33}{4}}{\binom{60}{10}}$$

Hypergeometric Random Variables

- Two kinds of tickets in box, but draws are ***without*** replacement (as opposed to the binomial setting, where the draws are independent).

- What information will we need? Total # of draws
Total # of tickets
" " " Successes ("Good")

- In this setting of drawing tickets without replacement, let X be the sample sum of tickets drawn from a box with tickets marked 0 and 1. Say that X has the ***hypergeometric*** distribution with parameters __

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