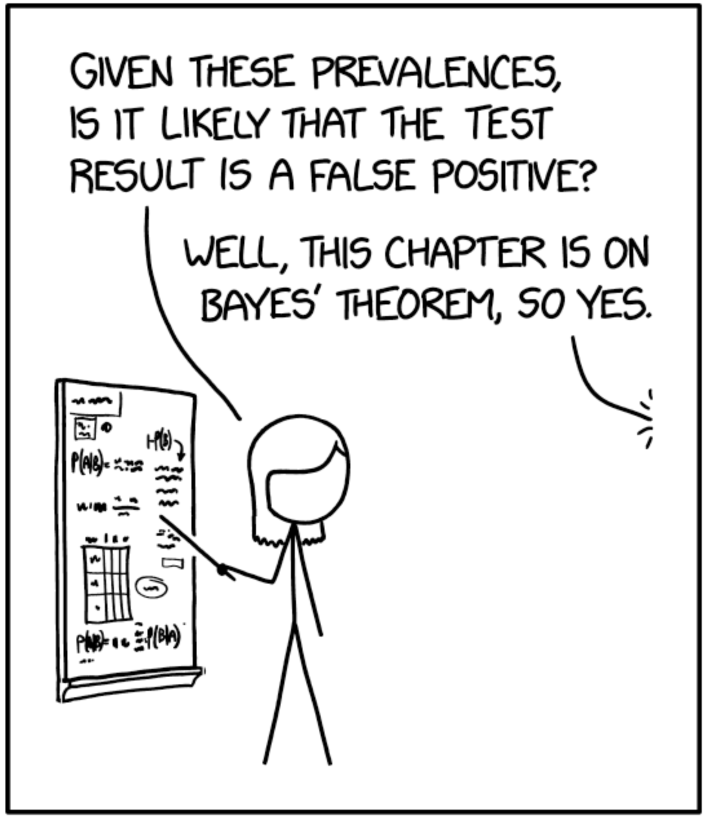
Stat 88: Probability & Math. Stat in Data Science



SOMETIMES, IF YOU UNDERSTAND BAYES' THEOREM WELL ENOUGH, YOU DON'T NEED IT.

https://xkcd.com/2545

Lecture 9: 2/6/2024

Random variables & their distributions, and a special distribution

3.1, 3.2, 3.3

Shobhana Stoyanov

Agenda & warm-up

- Random variables and their distributions
- The binomial distribution

- Warm up
    1. Deal 5 cards from a standard deck of 52. What is the chance that you have exactly 2 aces in your hand?

$\underline{52} \cdot \underline{51} \cdot \underline{50} \cdot \underline{49} \cdot \underline{48}$

$$\boxed{= \frac{52!}{47!\,5!}}$$

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

$$P\left(\text{exactly 2 Aces in hand of 5 cards}\right) = \frac{\binom{4}{2} \cdot \binom{48}{3}}{\binom{52}{5}}$$

note that this is not necessarily a "pair" as in Poker

    2. Roll a fair six-sided die 5 times. What is the chance of rolling exactly 2 aces (one spot)?

$$P(\text{exactly 2 aces in 5 rolls}) = \boxed{\binom{5}{2}} \left(\frac{1}{6}\right)^2 \cdot \left(\frac{5}{6}\right)^3$$

$\underline{NA}$ $\underline{A}$ $\underline{A}$ $\underline{NA}$ $\underline{NA}$

not not not

$\frac{1}{6} \cdot \frac{1}{6} \cdot \frac{5}{6} \cdot \frac{5}{6} \cdot \frac{5}{6}$

Section 3.1: Vocabulary

- 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) of outcomes ***Success***, and ***Failure***

every trial is independent

- Might be with replacement (like a coin toss) or without replacement (drawing a sample of voters from a city and checking number of registered voters) trials are DEPENDENT (like dealing cards)

- Read about Paul the octopus and Mani the parakeet and their soccer predictions
- Note that Paul made 8 correct 2010 WC predictions. What is the chance of 8 correct if picking completely at random? (like tossing a coin and getting all heads) Like tossing a coin 8 times

& getting all H $\text{Prob} = \left(\frac{1}{2}\right)^8 = \frac{1}{256}$

Back to counting outcomes of tosses

- Toss a coin 8 times, how many possible outcomes? $256 = 2^8$

- What is the chance of ***all*** heads? $P(\text{all } H) = \frac{1}{256}$

- If each of the students in this class present today flip a coin 8 times, what is the chance that ***at least 1 person*** gets all heads?

Assume 100 students

$P(\text{any one student getting all } H) = \frac{1}{256}$

Exercise for chocolate:

Prob of among 100 students at least 1 getting all H?

3.2 Random Variables

- A real number – we don't know exactly *what* value it will take, but we know the possible values.

success = Heads

- The number of heads when a coin is tossed 3 times could be 0, 1, 2, or 3.
- The sum of spots when a pair of dice is rolled could be 2, 3, 4, 5, …, 12.

- These are both examples of *random variables*.
- *Variable* because the number takes different values
- *Random variable* because the outcomes are not certain.

Ω outcome space $\{HHH, HHT, HTH, \ldots, TTT\}$ $\xrightarrow{X(TTT)}$ $\mathbb{R}$

$X(TTT) = 0$

$X(HTH) = 2$

X = # of H in 3 tosses of a fair coin

Random variables

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

- Using random variables helps to write events more clearly and concisely.

  - we can do arithmetic on outcomes

- It is a way to *map* the ~~function~~ OUTCOME space Ω to real numbers
- For example: Let X represent the number of heads in 3 tosses.

- We can write down the ***distribution*** of X, which consists of its possible values and their probabilities.

- The function describing the distribution is called the *probability mass function($f(x)$)*
- Note that the probabilities must add up to 1.
- We can visualize it using a *probability histogram*.

Random variables, distribution table & histogram (exercise from Friday)

- For example: Let X represent the **number of heads in 3 tosses**.
- We can write down the ***distribution*** of X, which consists of the possible values of X and the probabilities of X taking these values & make a histogram:

Outcome	X(outcome)	probability $\omega \in \Omega$
HHH	3	1/8
HHT	2	1/8
HTH	2	1/8
THH	2	1/8
HTT	1	1/8
THT	1	1/8
TTH	1	1/8
TTT	0	1/8

$P(X=3) = 1/8$

$P(X=2) = 3/8$

$P(X=1) = 3/8$

$P(X=0) = 1/8$

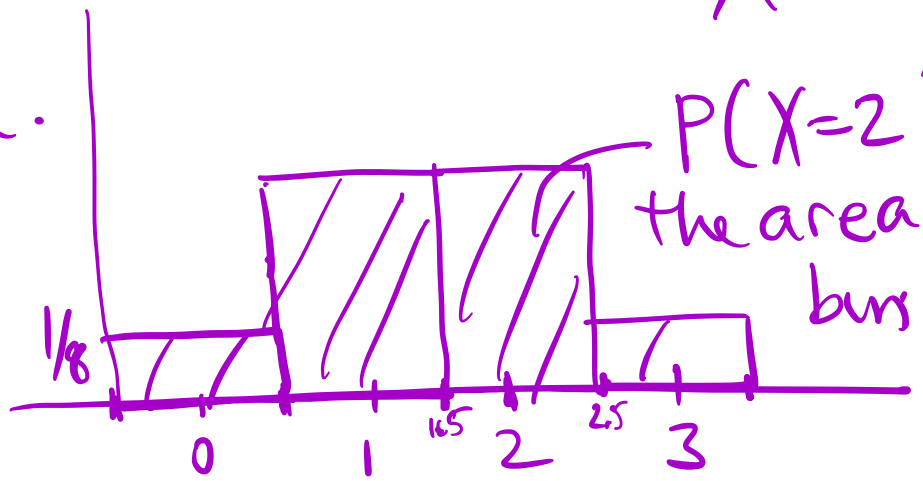
- The function describing the distribution is called the *probability mass function* $f(x)$, where $\boldsymbol{f(x) = P(X = x)}$

For this particular X

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

$$\boxed{f(x) = P(X=x)}$$

describes how the total "mass" (or prob. of 1 is distributed among the possible values that X can take.



P(X=2) is the area of this bar

Die 2

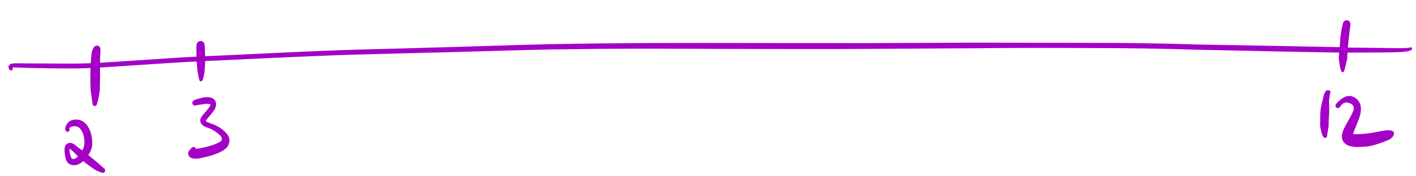
Die 1	1	2	3	4	5	6
1	11	12	13	14		
2						
3						
4						
5						
6						

Another example

- Let X be the **sum of spots** when a pair of dice is rolled.
- Write down the probability distribution table of X :

x	2	3	4	5	6	7	8	9	10	11	12
$f(x)$	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36

- Probability histogram: Exercise.



Random Variables

- Note that even if two random variables have the same distribution, they are not necessarily equal. For example, let X be the number of heads in 2 tosses of a fair coin, and Y be the number of tails.
- That is, we can talk about the *particular* values being equal and *distributions* being equal – and these are not the same thing.

X :

heads in 3 tosses

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

Y =

# of tails

y	0	1	2	3
$P(Y=y)$	1/8	3/8	3/8	1/8

$\boxed{X = 3 - Y}$ $\quad X \neq Y$

3.3 The Binomial distribution

generalisation of warm up problem

- 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
  - Each toss is an independent trial
  - A success is a head.
  - P(success) = 0.5

- 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.