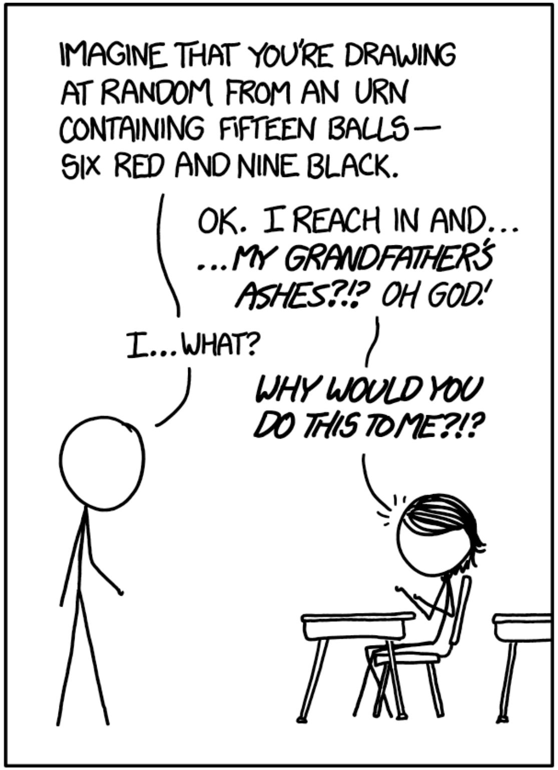
Stat 88: Probability & Math. Stat in Data Science



https://xkcd.com/1374/

Lecture 7: 2/3/2021

Random variables & their distributions + 2 special distributions

Sections 3.1, 3.2, 3.3, 3.4

Agenda

- 3.1, 3.2, 3.3, 3.4
- Success and failure
- Random variables
- The binomial distribution
- The hypergeometric distribution
- (binomial approximations next time)

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) outcomes ***Success***, and ***Failure***

- Might be with replacement (like a coin toss) or without replacement (drawing voters from a city and checking number of Bernie supporters)

- Read about Paul and Mani in section 3.1
- Note that Paul made 8 correct predictions. What is the chance of 8 winners if picking completely at random?

= chance of all H if tossing a fair coin 8 times. $= \frac{1}{256}$

Exercise: What is the chance of (numerical answer) at least 1 person out of 200 getting all H if tossing 8 coins. >50%

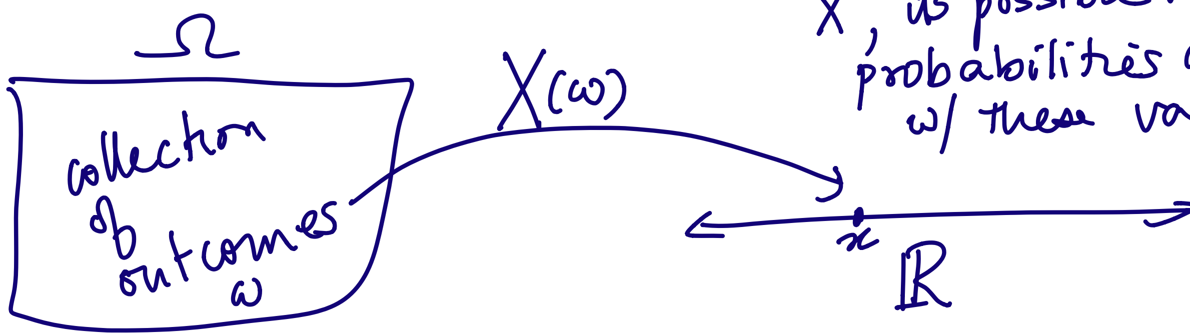
3.2 Random Variables

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

- 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 and *Random* because the outcomes are not certain.
- Formally: map the function space Ω to real numbers

X, its possible values, probabilities associated w/ these values

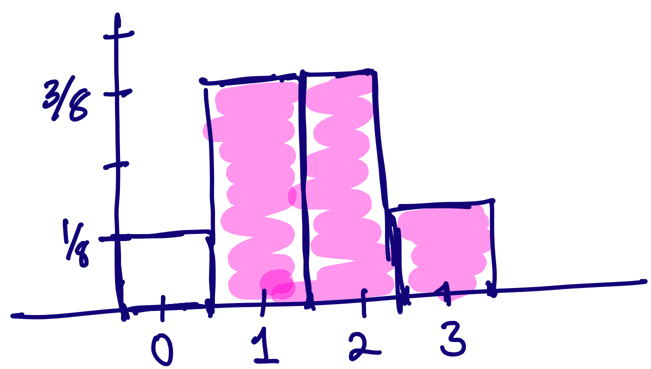


Random variables, prob. dsn table & histogram

- 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:

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

outcome: ω	$x = X(\omega)$	$P(X = x)$
HHH	3	1/8
HHT	2	1/8
HTH	2	1/8
THH	2	1/8
TTT	0	1/8
TTH	1	1/8
THT	1	1/8
HTT	1	1/8



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

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

$$P(X > 0) = f(1) + f(2) + f(3) = \frac{3+3+1}{8}$$

$= 1 - P(X \le 0) = 1 - P(X=0) = 1 - \frac{1}{8} = \frac{7}{8}$

Another example

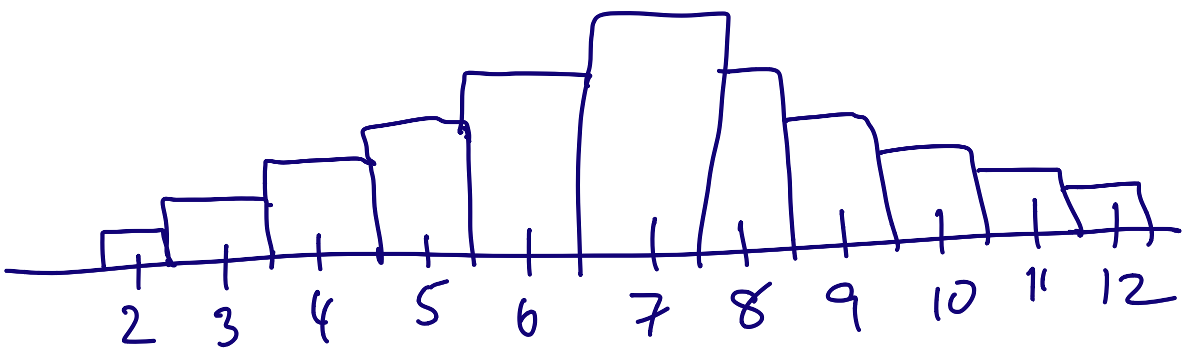
equally likely

36 outcomes when we roll a pair of dice.

- 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:



Random Variables and Equality

- Note that even if two random variables have the same distribution, they are not necessarily equal. That is, we can talk about the particular values being equal and distributions being equal.

- Let X= # of Heads in 3 tosses of a fair coin and Y = # of tails in 3 tosses of a fair coin.

X :

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

Y =

y	0	1	2	3
$f(y)$	$1/8$	$3/8$	$3/8$	$1/8$

Two r.v.'s can have exactly the same dsn, but not equal values. $X = 3 - Y$

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

Can model as binomial

- 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.

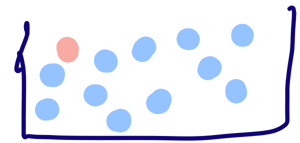
$n = 10$, $p = P(S) = P(\boxed{\cdot}) = \frac{1}{6}$

$\overset{1-p}{\boxed{0}}$ $\overset{p}{\boxed{1}}$ draws / replacement n times

$X =$ sum of draws, $P(\boxed{1}) = p$

sum of draws = # of 1's . $P(\boxed{0}) = 1-p$

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
- 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.

  → biased coin

- P(RBBB) = $\frac{1}{12}\cdot\frac{11}{12}\cdot\frac{11}{12}\cdot\frac{11}{12} = \left(\frac{1}{12}\right)^1\left(\frac{11}{12}\right)^3$ ← chance of each seq. w/ 1 R, 3 B
- How many such sequences are there? 4
- What is the probability of all such sequences (1 R, 3B)?

$4\cdot\left(\frac{1}{12}\right)\left(\frac{11}{12}\right)^3$ ← sum of all the probs.

Binomial distribution: Example

$n = \#$ of trials.

- 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) = $\left(\frac{1}{12}\right)\left(\frac{1}{12}\right)\left(\frac{11}{12}\right)\left(\frac{11}{12}\right)$

$\#(\text{sequences w/ } 2R \,\&\, 2B) = \binom{4}{2}$

- There are 6 such sequences (how?), so if we let X = # of red balls in 4 draws with replacement, we have that

$p = \frac{1}{12}$

Success = Red ball

$$P(X = 2) = \binom{4}{2} \times p^2 \times (1 - p)^2$$

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

where p = P(red) = 1/12

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

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

P(a single sequence with K succ. & (n-K) failures)

$= p^k (1-p)^{n-k}$

X ~ Bin(n, p)

Characteristics of the binomial distribution

- There are n trials, where n is FIXED beforehand.
- The chance 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 ~ Bin(n,p), possible values of X: 0, 1, 2, …, n
- Use python to compute probabilities

Identifying binomial random variables

Which of the following are binomial random variables?

- Number of heads in 12 tosses of a fair coin.
- Number of tosses until we see two heads.
- Number of queens in a five card hand
- Number of Democrats in a simple random sample of 500 adult voters drawn from the SF Bay Area.