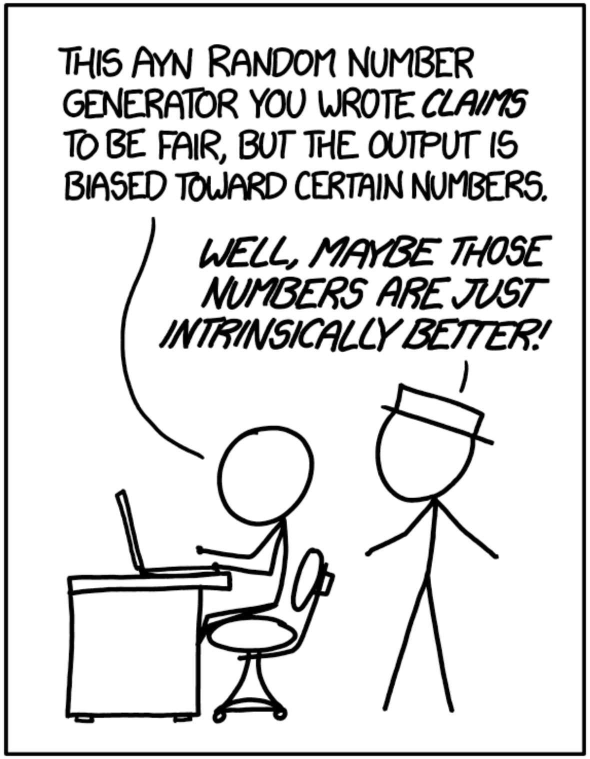
Stat 88: Probability & Math. Stat in Data Science



https://xkcd.com/1277

Lecture 8: 2/2/2024

Random variables & their distributions, and a special distribution

3.1, 3.2, 3.3

Shobhana Stoyanov

Agenda

- Counting permutations and combinations
- Random variables and their distributions
- The binomial distribution

Counting permutations

- Recall the product rule of counting, where we counted number of outcomes when we had a sequence of k actions, each with n_i outcomes, so the total number of outcomes is $n_1 \times n_2 \times \ldots n_k$
- # of ways to rearrange n things, taking them 1 at a time is $n!$
- If we have only $k \leq n$ spots to fill, then $n \cdot (n-1) \cdot \ldots \cdot \left(n-(k-1)\right)$
- # of perm. of n things taken k at a time.

- Count the number of sequences of 3 letters taken from the English alphabet without replacement. _ · _ · _

Counting combinations

- Suppose we don't care about the sequence but just *which* letters were chosen (so *abc* = *bca* = *cab* etc.) Then all of these combinations count as 1 selection. We need to take the number we got above and divide by the number of arrangements of 3 letters = _ · _ · _
- If we don't care about order, then we are counting subsets, and this number is denoted by $\binom{n}{k}$ (read as "n choose k") which we get by dividing: $n \cdot (n-1) \cdot \ldots \cdot \big(n-(k-1)\big)$ by $k!$, so $\displaystyle\binom{n}{k} = \frac{n!}{(n-k)!\,k!}$
- Note: $\binom{n}{n} = 1$, $\binom{n}{0} = 1$
-

Examples

Let's consider poker, in which each player is dealt 5 cards. How many hands of 5 cards are possible from a standard deck? Recall that a standard deck has 52 cards, consisting of 4 suits (♥, ♦, ♣, ♠) of 13 cards each (**2, 3, …,10, J, Q, K, A**)

- Chance of a pair in poker=
- Chance of two pairs =
- Chance of "full house in poker" =

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***

- 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)

- 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)

Back to counting outcomes of tosses

- Toss a coin 8 times, how many possible outcomes?

- What is the chance of ***all*** heads?

- 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?

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
- *Random variable* because the outcomes are not certain.

Random variables

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

- It is a way to ***map*** the function 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

- 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

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

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*)											

- Probability histogram:

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.

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

- 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.

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.

- $P(RBBB) =$
- How many such sequences are there?
- What is the probability of all such sequences (with1 R, 3B)?

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

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

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

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

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)

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.