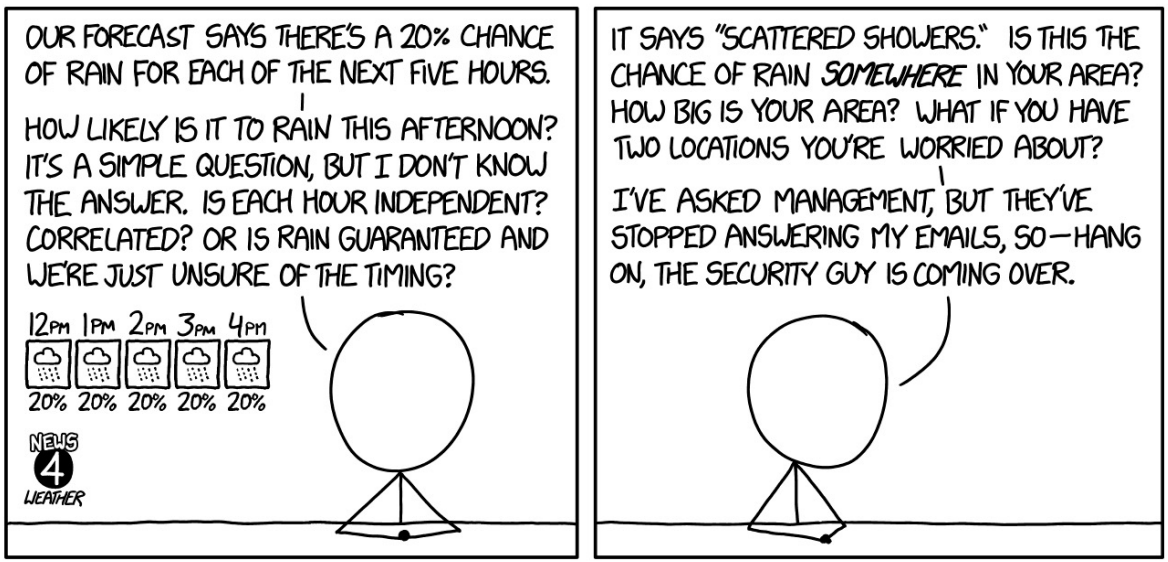
Stat 88: Probability and Mathematical Statistics in Data Science



https://imgs.xkcd.com/comics/meteorologist.png

Lecture 1: 1/18/2022

Course introduction and the basics, 1.1, 1.2

Shobhana M. Stoyanov

Agenda

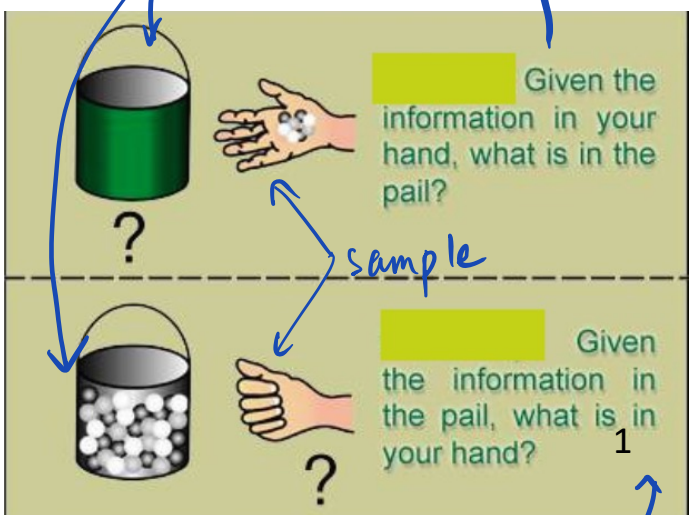
- Course resources:
  - Course site: http://stat88.org
  - Announcements and discussions: Piazza
  - Assignments and grades: Gradescope

- Write your questions on the google doc (we will have one for each lecture): https://tinyurl.com/2p9d58t6

- The Basics:
  - Section 1.1: Probabilities as Proportions
  - Section 1.2: Exact Calculation or Bound

Probability vs Statistics

- Discuss which is probability and which is statistics:



population → Statistics

sample → population
"inference."

coin P(H): toss it 100 times & try to figure out P(H)
or see the colors of marbles in my hand & guess at the composition of marbles in the bucket.

1

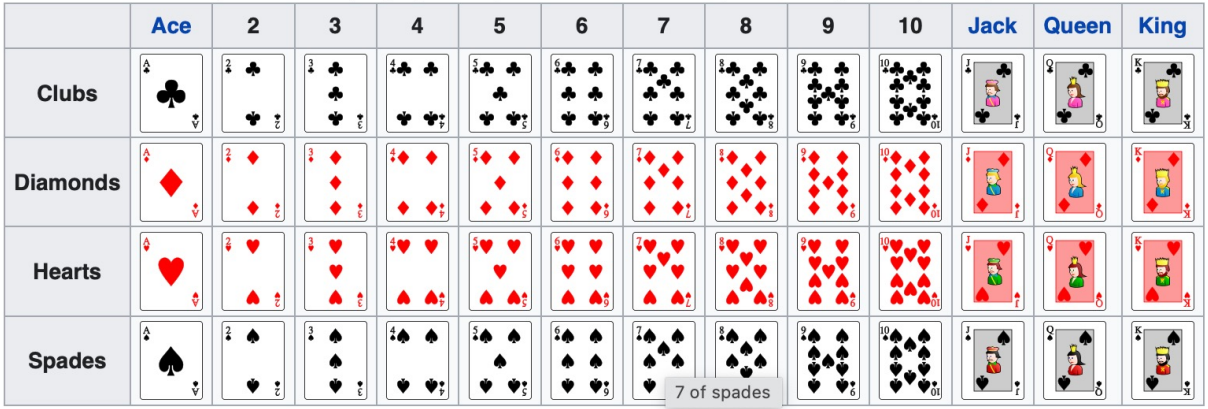
probability.

we know the population
take a "random" sample & compute the "chance" of specific items in the sample.

2

Cards

Example set of 52 playing cards; 13 of each suit clubs, diamonds, hearts, and spades



If you have a well-shuffled deck of cards, and deal 1 card from the top, what is the chance of it being the queen of hearts? What is the chance that it is a queen (any suit)? What assumptions are you making?

$\frac{1}{52}$

Assumption: all cards are equally likely to end up on top.

If you deal 2 cards, what is the chance that at least *one* of them is a queen? How do these relate to populations and samples?

Exercise reminder me to talk about this first thing on Thursday.

Deck: population Hand: sample.

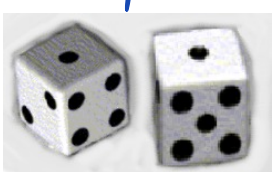
Section 1.1: Probabilities as proportions

- We can think about probability as a numerical measure of uncertainty, and we will define some basic principles for computing these numbers.
- These basic computational principles have been known for a long time, and in fact, gamblers thought about these ideas a lot. Then mathematicians investigated the principles.

De Méré's Paradox

- Famous problem: will the probability of **at least one six** in **four** throws of a die be equal to prob of **at least a double six** in 24 throws of a pair of dice.

Exercise how would you even begin to compute these probabilities?

- Note: single = die, plural = dice:  vs 

Origins of probability: de Méré's paradox

Questions that arose from gambling with dice.

Antoine Gombaud,
Chevalier de Méré

Blaise Pascal

Pierre de Fermat

The dice players
Georges de La Tour
(17th century)

Terminology

Example : Action: Toss a coin
$\{H, T\} = \Omega$

- Suppose we have an action that results in exactly one of several possible *outcomes* or results, and chance or randomness is involved - that is, each time we perform the action, the outcome will be different, and we don't know exactly which outcome will occur.
- Such an action is called an *experiment* or a *random experiment*.
- Examples: toss a coin, roll a die, take a random sample of people and see how many agree with Australian government's decision to deport Novak Djokovic.

Action : Roll a six-sided die
outcomes : $\{1, 2, 3, 4, 5, 6\}$
Ω
Sample space.

Djokovic admits attending interview with journalist while Covid positive

COVID-19 | NEWS | POLITICS | FOOTBALL | CELEBS | TV | MONEY

Pressure rises on Alex Hawke as 83% of Australians call for Novak Djokovic to be deported

Immigration minister Alex Hawke is under pressure to deport Novak Djokovic from Australia. According to a NewsCorp survey, 83% of people want to see the Serbian leave the country.

By Liam Llewellyn, Sports Trends Writer
11:07, 13 Jan 2022

10 COMMENTS

An overwhelming 83% of Australians want Novak Djokovic to be deported - piling pressure on immigration minister Alex Hawke.

Over 60,000 people responded to the survey conducted by NewsCorp, which indicates the feeling among members of the public Down Under.

Terminology

- Suppose we have an action that results in exactly one of several possible ***outcomes*** or results, and chance or randomness is involved - that is, each time we perform the action, the outcome will be different, and we don't know exactly which outcome will occur.
- Such an action is called an ***experiment*** or a ***random experiment***.

- A collection of all possible outcomes of an action is called a ***sample space*** or an ***outcome space*** . Usually denoted by Ω (sometimes also by S).
- An ***event*** is a collection of outcomes, so a subset of Ω.

denote events by A, B, C etc.

$A \subseteq \Omega$

Computing probabilities

Chance of top card $Q\heartsuit = \frac{1}{52}$

- If you have a well-shuffled deck of cards, and deal 1 card from the top, what is the chance of it being the queen of hearts? What is the chance that it is a queen (any suit)?

chance top card is a $Q = \frac{4}{52}$

- How did you do this? What were your assumptions?

all the cards equally likely to be in each of the positions.

$$\frac{\#(\text{queens})}{\#(\text{cards in deck})} = \frac{4}{52}$$

- Say we roll a die. What is Ω?

six-sided

$\Omega = \{1, 2, 3, 4, 5, 6\}$

- What is the chance that the die shows a multiple of 3? What were your assumptions?

Chance (probability) die shows multiple of 3. $= \frac{\#(\{3,6\})}{\#(\Omega)}$

$= \frac{2}{6} = \frac{1}{3}$

Chance of rolling a prime # $= \frac{\#(\{2,3,5\})}{\#(\Omega)}$

Exercise ① Write out Ω of action is rolling a PAIR of dice $\Omega = \{(1,1),(1,2)$ ... $\}$
② Write out Ω of action is tossing 3 coins.

Chance of a particular outcome

- We usually think of the chance of a particular outcome (roll a 6, coin lands heads etc) as the number of ways to get that outcome divided by the total possible number of outcomes.

$$\frac{\#\ of\ particular\ outcomes\ of\ interest}{total\ \#\ of\ outcomes\ possible}$$

- So if A is an event (subset of Ω), then $P(A) = \frac{\#(A)}{\#(\Omega)}, A \subseteq \Omega$

Cards

P(event) denotes the probability of that event.
↑ collection of outcomes

- If you have a well-shuffled deck of cards, and deal 1 card from the top, what is the chance of it being the queen of hearts? What is the chance that it is a queen (any suit)?

chance of $Q\heartsuit = P(Q\heartsuit) = \frac{1}{52}$

$P(Q) = 4/52$

- If you deal 2 cards, what is the chance that at least *one* of them is a queen?

$P(\overbrace{\text{at least 1 out of 2 cards dealt is a } Q}^{A}) = \frac{\#(A)}{\#(\Omega)} = \frac{\#\{QQ, Q\text{not}Q, \text{not}Q, Q\}}{52 \cdot 51}$

$52 \cdot 51$ = total # of outcomes possible

$= ??$