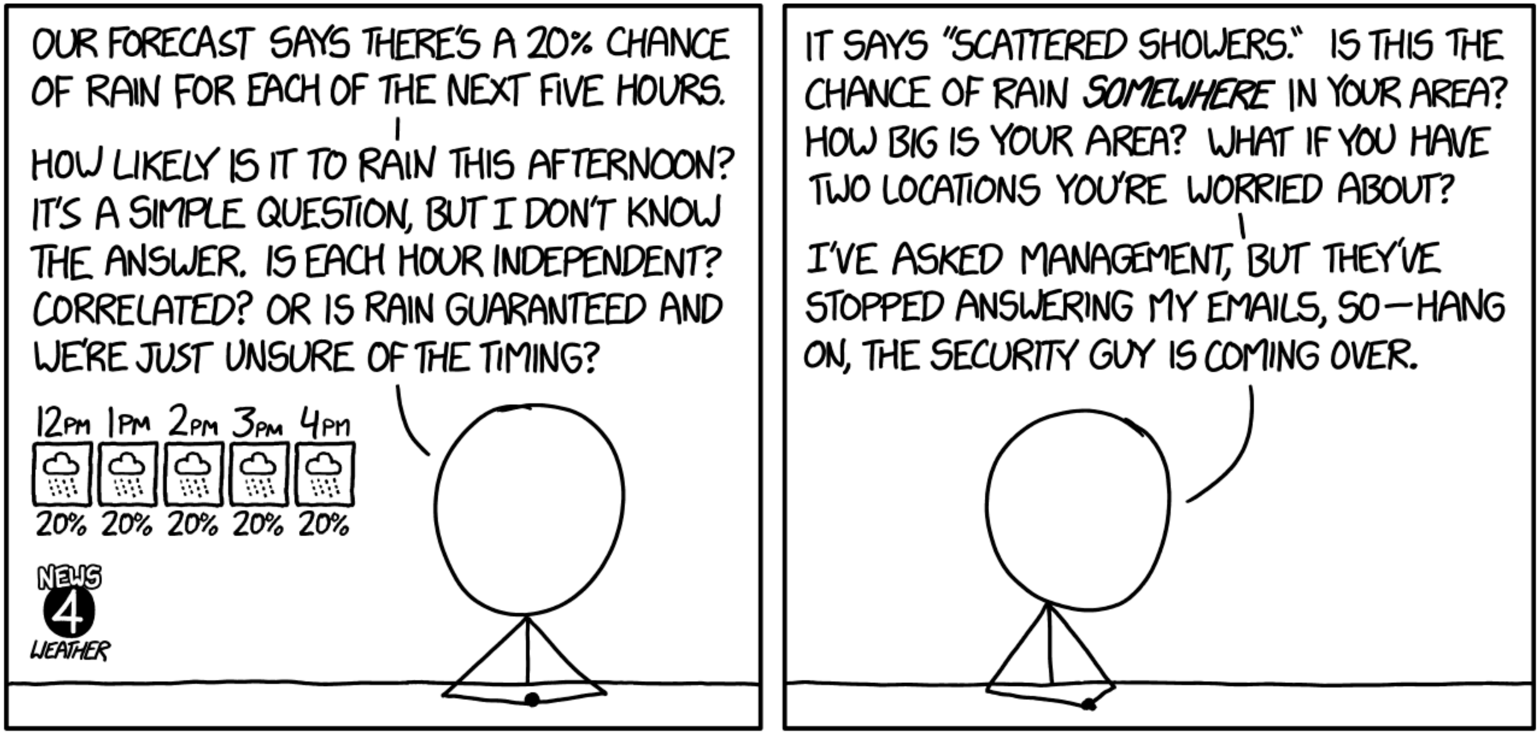
Stat 88: Probability and Mathematical Statistics in Data Science



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

Lecture 1: 1/17/2024

Course introduction and the basics

Shobhana Stoyanov

Agenda

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

- Put your questions about the course and today's lecture on the thread for Lecture 1

- Introduce yourself to two people sitting near you, tell them your name, where you were born, and what you would be famous for, if you were famous.

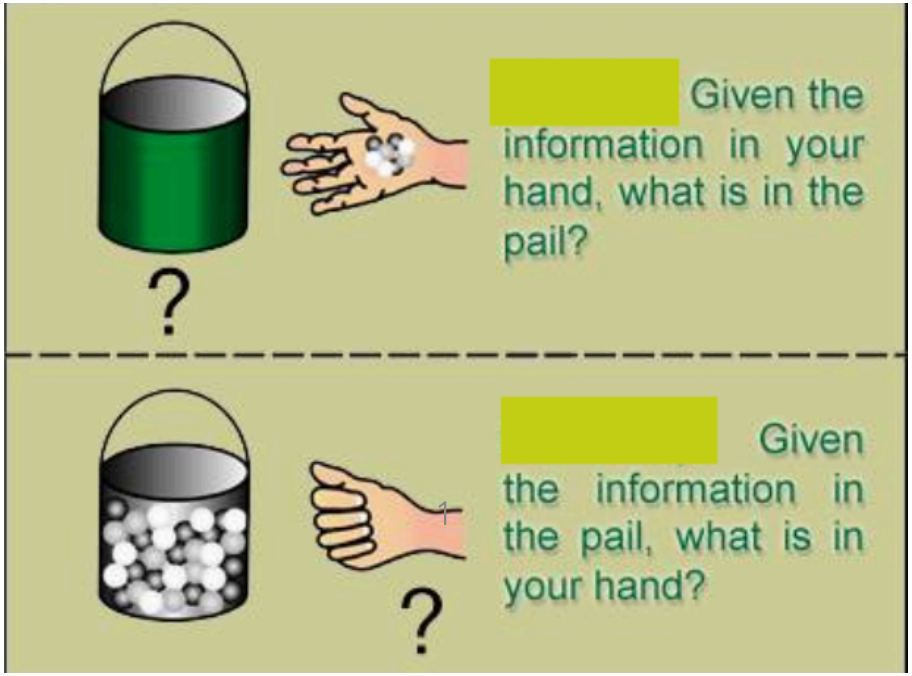
- The Basics:
  - terminology
  - assumptions
  - proportions
  - distribution

Probability vs Statistics

- Discuss which is probability and which is statistics:

A

B



Section 1.1.1: Basic vocabulary or terminology

- The act of shuffling a deck and then drawing a card has an element of chance - you won't always get the same card.
- Any activity that has chance associated with it is called an ***experiment*** or a random experiment if there is 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.

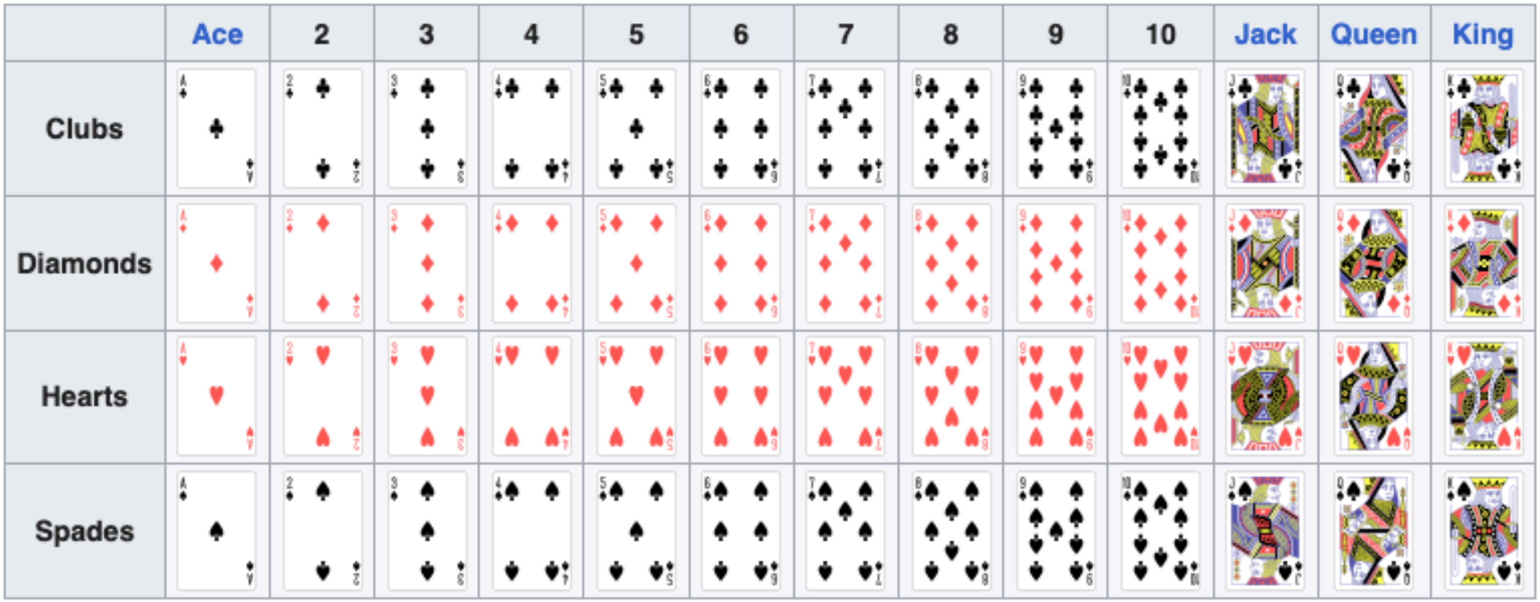
- Which of the following are experiments?
  - Roll a pair of dice
  - Read your textbook
  - Buy a raffle ticket
  - Draw 52 cards from a standard deck, without *replacement.*

Red \ Green	1	2	3	4	5	6
1	1,1	1,2	1,3	1,4	1,5	1,6
2	2,1	2,2	2,3			
3						
4						
5						
6						

- An ***event*** is a description of the result, and might include several outcomes. For example, rolling a die and having the sum of the rolls be 4.

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?

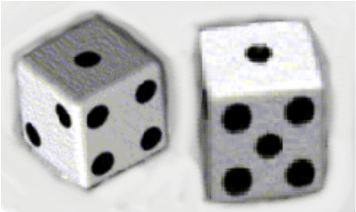
- If you deal 2 cards, what is the chance that at least *one* of them is a queen? How do these relate to populations and samples?

De Méré's Paradox

- 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.

- 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.

- Note: single = die, plural = dice:



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
(17^{th} century)

Terminology

- **Experiment**: 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.

- An ***event*** is a collection of outcomes.

- 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 always a subset of Ω. Suppose we call the event A, then we write this as $A \subset \Omega$

Computing probabilities: what do we often assume?

- 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)?

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

- Say we roll a die. What is Ω?

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

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 $P(A) = \dfrac{\#(A)}{\#(\Omega)},\ A \subseteq \Omega$

- If an experiment has a **finite** number of possible *equally likely* outcomes, then the probability of an event is the proportion of outcomes that are included in the event.