Introduction to Probability

Matt Brems | *Patrick Wales-Dinan*

GENERAL ASSEMBLY

Data Science Process

1. Define problem. → What is the problem I want to solve? You need to define that problem… & you may have to take a real-world problem and turn it into a data science problem.
2. $\boxed{\text{Gather data.}}$ → Collecting the information. We might get it from the internet, administer a survey, do an experiment to gather the data...
3. Explore data. → What does the information tell us? This is the data exploration process.
4. $\boxed{\text{Model with data.}}$ → We want to use the different models that we have: Linear Regression, Random Forest, Neural Net? We are trying to gain insights from our data.
5. Evaluate model. → Specifically is there a way to measure if the model is good or not
6. Answer problem.

Definitions

Experiment: A procedure that can be repeated an infinite number of times and has a well-defined set of outcomes.

Sample Space: The set of all possible outcomes of an experiment, usually denoted S

Event: Any collection of outcomes of an experiment.

Examples

Experiment: Flip a coin twice.

Sample Space:

S = { {H,H}, {H,T}, {T,T} }

Event:

A = flip as least one H
= {{H,H}, {H,T}}

Experiment: Roll one die.

Sample Space:

S = {1, 2, 3, 4, 5, 6}

Event:

B = Roll even = {2, 4, 6}
C = Roll 5 or 6 = {5, 6}
D = Roll prime = {2, 3, 5}

Definitions

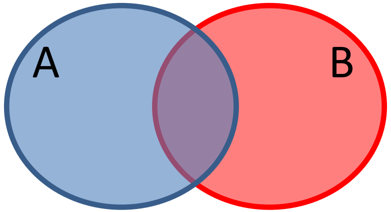
Set: An unordered collection of distinct objects.

- {Annie, *e*, basketball}

Element: An object that is a member of a set.

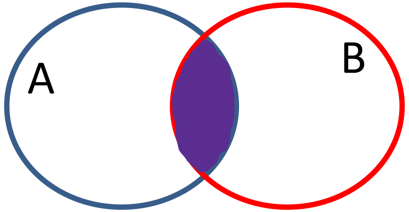
- Annie
- *e*
- basketball

Set Operations



Union: A ∪ B = the set of elements in set A **or** in set B

Intersection: A ∩ B = the set of elements in set A **and** set B



Example:

A = {2, 4, 6, 8}

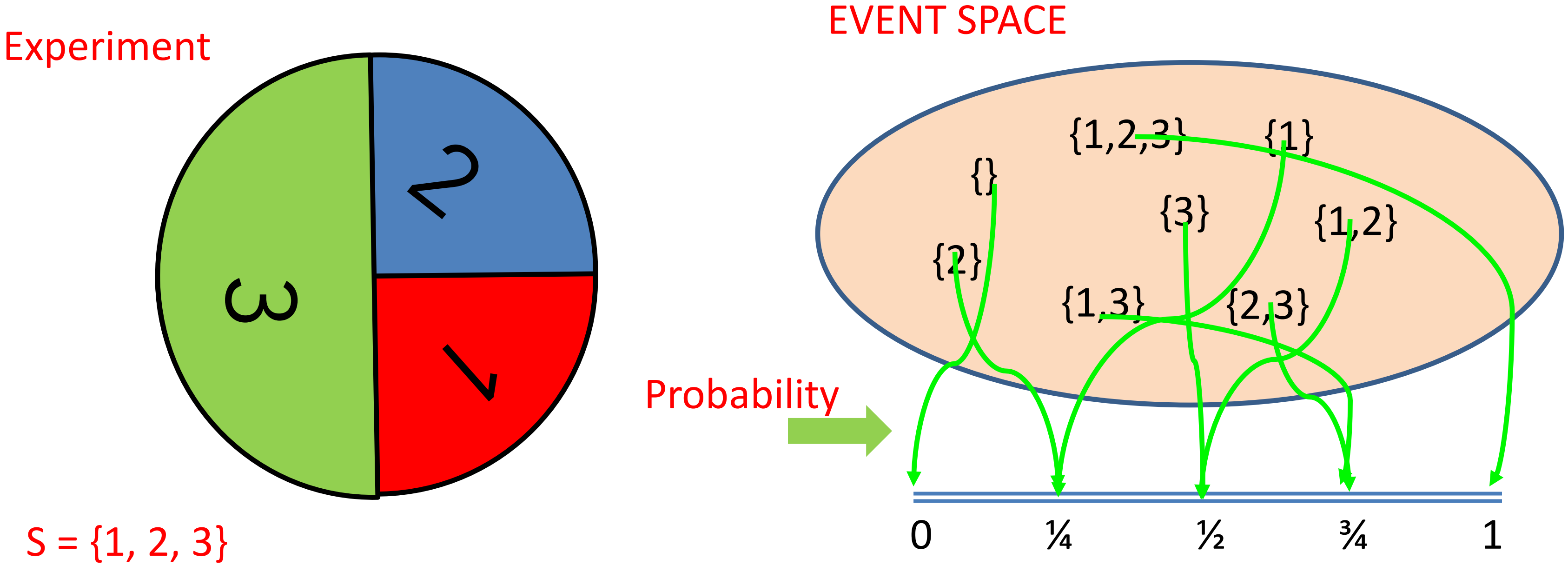
B = {2, 3, 5, 7}

A ∪ B = {2, 4, 6, 8} ∪ {2, 3, 5, 7} = {2, 3, 4, 5, 6, 7, 8}

A ∩ B = {2, 4, 6, 8} ∩ {2, 3, 5, 7} = {2}

Probability

We are often interested in the probability of some event(s) occurring.



We write $\boxed{P(A)}$ to mean the probability that event A occurs.

Probability - Practice

A: “a U.S. birth results in twin females”

B: “a U.S. birth results in identical twins”

C: “a U.S. birth results in twins”

In words, what does $P(A \cap C)$ mean?

The probability that a US birth results in twin females

In words, what does $P(A \cap B \cap C)$ mean?

The probability that a US birth results in identical twin females

Example adapted from *Statistical Inference*, 2nd Ed., Casella & Berger

Probability Rules

When trying to find the probability of a complex event, it's not straightforward.

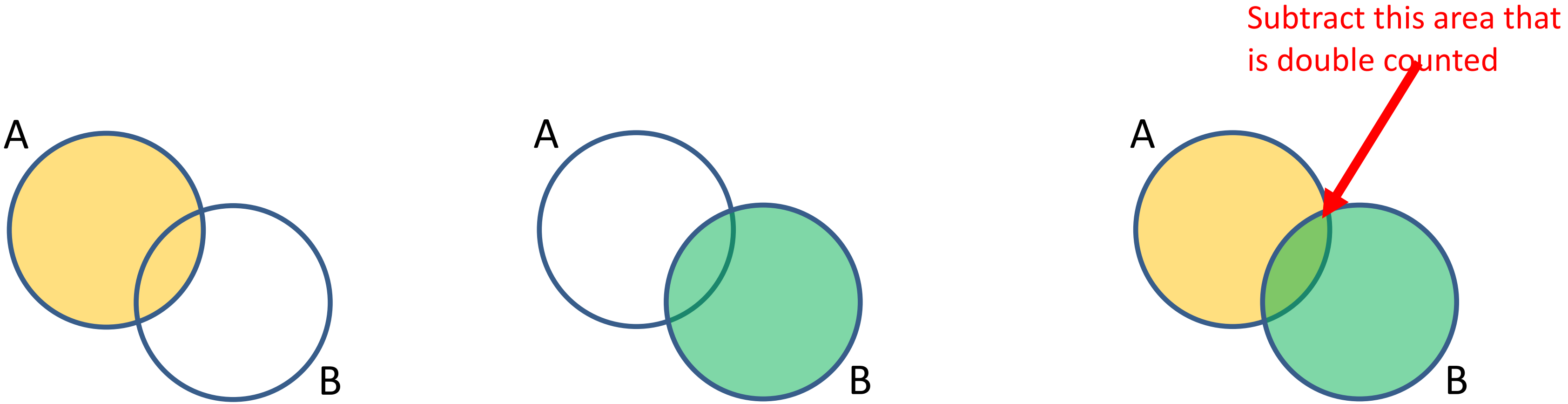
> There are 12 red and 12 black balls. If you draw one ball, then a second ball without replacing the first, what is the probability they are the same color?
> Suppose you roll three dice. What is the probability that the three dice are rolled in increasing order?
> You call 3 friends of yours in Seattle and ask each independently if it's raining. Each of your friends tells you the truth ⅔ of the time. All 3 friends tell you it is raining. Based on historical evidence, it rains ¼ of the time in Seattle. What is the probability that it's actually raining in Seattle right now?

There are three probability rules that come in handy.

Probability Rule 1: P(A ∪ B)

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Venn diagrams can help *illustrate* this, but Venn diagrams are not formal proofs!



Probability Rule 2: $P(A \mid B)$

Conditional Probability

$$P(A \mid B) = P(A \cap B) / P(B)$$

A | B means "A given B" or "A conditional on the fact that B happens."

A = roll a 2

B = roll even

A | B = roll a 2 given even

$$P(A \mid B) = P(A \cap B) / P(B)$$

$$= P(2 \cap \text{even}) / P(\text{even})$$

$$= P(2) / P(\text{even})$$

$$= (1/6) / (1/2) = 1/3$$

Probability Rule 3: P(A ∩ B)

P(A ∩ B) = P(A | B) * P(B)

We just took the last rule and multiplied both sides by P(B).

We can rearrange these as well: **P(B ∩ A) = P(B | A) * P(A)**

A = Lokma (Greek)

B = Pizzetta (Italian)

P(Lokma ∩ Pizzetta)

P(Pizzetta ∩ Lokma)

This isn't limited to two events: **P(A ∩ B ∩ C) = P(A | B ∩ C) * P(B | C) * P(C)**

Probability Practice

There are 12 red and 12 black balls. If you draw one ball, then a second ball without replacing the first, what is the probability that they are the same color?

P (draw 2 balls, same color) = P (draw 2 red or draw 2 black)

$= P(R \cap R \cup B \cap B)$

$P(R \cap R) = P(R|R)P(R)$

$P(B \cap B) = P(B|B)P(B)$

$(11/23) * (12/24) + (11/23) * (12/24)$

When by hand is tough...

Oftentimes, it's challenging to evaluate probabilities by hand.

But it's important to understand the ideas behind probability.

- For example: are two events independent of one another?
- Events A and B are **independent** if $\mathbf{P(A \mid B) = P(A)}$.

We often think of probability as how frequently an event occurs.

- We can use computer simulations to give us a good approximation of the true probability of some event.
- We can often use this to "check our work" or to tackle more challenging probability problems!

