STAT 88: Lecture 2

Contents

Last time Probability allows you to learn about a random sample from a known population

Probability: population $\rightarrow$ sample

Statistics: sample $\rightarrow$ population

Sec 1.1 Probability as Proportions

We call the set of all outcomes of an experiment Ω, the outcome space or the sample space. Let $A \subseteq \Omega$ be an event.

For equally likely outcomes, $P(A) = \frac{\#A}{\#\Omega}$.

Sec 1.2 Probability Bounds

When we don't know how much events overlap, we sometimes need to give upper and lower bounds for what a probability is.

Example: Let $A, B, C \subseteq \Omega$ with $P(A) = 0.1, P(B) = 0.05, P(C) = 0.01$. Then

$$P(A \cup B \cup C) \leq P(A) + P(B) + P(C) = 0.16 \text{ and}$$
$$P(A \cup B \cup C) \geq P(A) = 0.1.$$

Warm up: Based on historical averages,

- $P(A)$ = Chance that you catch bus to school = 70%.
- $P(B)$ = Chance that it rains = 50%.

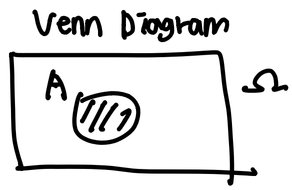
- $P(C)$ = Chance you make it to class on time = 10%.

What is the chance that at least one of these events occurs (making no assumptions). If it cannot be found exactly, find the best lower bound and upper bound that you can.

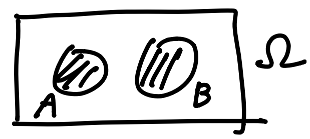
1.3. Fundamental Rules

In 1933, the Russian mathematician Andrey Kolmogorov established the axioms of the modern theory of probability. Intuitively, the axioms are generalizations of the natural properties of measures like length, area, or volume (think of Venn diagrams where $P(A)$ is the area of circle A).

Formally, probability is a function on events, $P : A \subseteq \Omega \mapsto [0, 1]$, satisfying the following 3 axioms:



1. $P(A) \geq 0$ for all $A \subseteq \Omega$.
2. $P(\Omega) = 1$.
3. **Addition Rule.** If A and B are mutually exclusive, i.e. $A \cap B = \emptyset$, then $P(A \cup B) = P(A) + P(B)$.

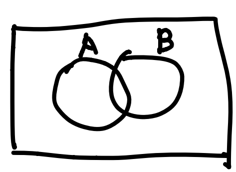


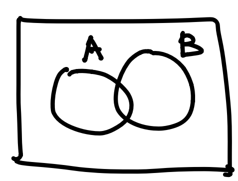
Consequence of three axioms:

1. Complement rule: $P(A^c) = 1 - P(A)$.
2. Difference rule: If $B \subseteq A$ then $P(A \backslash B) = P(A) - P(B)$.
3. Boole/Bonferroni's inequality: $P(\cup_{i=1}^{n} A_i) \leq \sum_{i=1}^{n} P(A_i)$ for all events A_i's.

Boole/Bonferroni's inequality is widely used in statistics and machine learning theory. In many cases, however, it gives rather a crude bound, e.g. if $P(A) = 0.8$ and $P(B) = 0.9$ then $P(A \cup B) \leq 1.7$ which is true but rather silly as we already know $P(A \cup B) \leq 1$.

De Morgan's laws: $(A \cap B)^c = A^c \cup B^c$.





Example:

- $P(A)$ = Chance that you catch bus to school = 70%.
- $P(B)$ = Chance that it rains = 50%.

Find the best lower bound for $P(A \cap B)$ you can without making any assumptions.

Example:

- $P(A)$ = Chance that you catch bus to school = 70%.
- $P(B)$ = Chance that it rains = 50%.
- $P(C)$ = Chance you make it to class on time = 10%.

Find the best lower bound for $P(A \cap B \cap C)$ you can without making any assumptions.