AID-521 Mathematics for Data Science

Module: Probability | Lecture: 1

BASIC CONCEPTS OF PROBABILITY

random events, probability calculation

Basic Definitions

A random event is one of the outcomes of an underlying process, whose outcome at a particular time is not known prior to that time.

- → On Monday, that the Sun will rise in the east is a deterministic event.
- → On Monday, that the stock index will fall below 100 points is a random event.

The underlying process is called experiment.

Basic Definitions

A trial of a process is the (deliberate/natural) performance of the process once, after which its outcome can be observed.

- → A medical test is conducted on a patient to infer the effectiveness of a drug.
- → The duration of rainfall in Delhi can be shorter or longer.

Multiple trials of a given process may or may not be performed.

Basic Definitions

The sample space S of a trial is the set of all possible outcomes of the trial.

- $\rightarrow$ In a coin toss, the sample space is composed of heads and tails.
- $\rightarrow$ The sample space of the price of a new car contains all non-negative real numbers, i.e., $[0, \infty)$.

So, a random event A is a subset of S. Can you see this?

Defining Probability

Intuitive Definition

The probability of an outcome (event) is the proportion of times the outcome (event) would occur in a long run of repeated experiments.

- $\rightarrow$ Requirement of identical conditions for repeated experiments.
- $\rightarrow$ Unbiased vs. Biased coin: $P(H) = \lim_{n \to \infty} \frac{\#H}{n}$?

Defining Probability

Axiomatic Definition*

Given a sample space S composed of events, probability P is a function that satisfies the following conditions.

- $\rightarrow$ $P : S \rightarrow [0, \infty)$, i.e., $P(A \subset S) \geq 0$ for all events in S
- $\rightarrow$ $P(S) = 1$
- $\rightarrow$ $P(A \cup B) = P(A) + P(B)$, if events A and B are mutually exclusive (i.e., $A \cap B = \phi$)

*Does not say how to calculate/assign probabilty to an event. If a function P satisfies the above conditions, it is a probability function.

Properties of Probability

- $\rightarrow$ $P(A^c) = 1 - P(A)$
- $\rightarrow$ If $A \subset B$, then $P(A) \leq P(B)$
- $\rightarrow$ $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

for any two events A and B in S

Properties of Probability

Try yourself

In a large university, the freshman profile for one year's fall admission says that 40% of the students were in the top 10% of their high school class, and that 65% are white, of whom 25% were in the top 10% of their high school class. What is the probability that a freshman student selected randomly from this class either was in the top 10% of his or her high school class or is white?

Properties of Probability

Try yourself – solution

- $\rightarrow$ Let A be the event that a person chosen at random was in the top 10% of his or her high school class, and let B be the event that the student is white.
- $\rightarrow$ We have $P(A) = 0.4,\ P(B) = 0.65,\ P(A \cap B) = 0.25$
- $\rightarrow$ Required Event[1] is $A \cup B$
- $\rightarrow$ Required Probability $=\ 0.4 + 0.65 - 0.25$

[1] student chosen is white or was in the top 10% of his or her high school class

Counting: Ordered Objects

From n objects, we are sampling m items:

with replacement

total no. of ways
$= n$ ways $\times$ n ways $\times \cdots$ for m times
$= n^m$

without replacement

total no. of ways
$= n$ ways $\times$ $(n-1)$ ways $\times \cdots \times$ $(n-m+1)$ ways
$= \frac{n!}{(n-m)!} := {}^nP_m$

Counting: - Un-ordered* Objects

*A sample "ABC" is the same as "BAC" – i.e., they are not two different samples.

From n objects, we are sampling m items:

without replacement

total no. of ways
= n ways $\times$ n ways $\times \cdots$ for m times, divided by $m \times \cdots \times 1$
= $\frac{n!}{(n-m)!} \times \frac{1}{m!} \; := \; {}^{n}C_{m}$ or $\binom{n}{m}$

with replacement

total no. of ways
= equivalent to "choosing" m samples from $n + m - 1$ objects without replacement
= ${}^{n+m-1}C_{m}$

Counting: Objects in Classes

No. of ways to group n objects in m classes

with n_i objects in the i-th class ($i = 1, 2, ..., m$ and $\sum_{i=1}^{m} n_i = n$)

$$= \binom{n}{n_1 \cdot n_2 \cdots n_m}$$

Hint: how many samples are you choosing from the n objects?

$$= \frac{n!}{n_1! \cdot n_2! \cdots n_m!}$$

Counting & Probability

In the previous slides, we saw how to count cases.
But, where is probability?

Suppose there are N possible outcomes of an experiment, and

let n_A be the number of outcomes in an event A.

Then the probability of event A is

$$P(A) = \frac{n_A}{N}.$$

Try yourself

The admissions committee of a department at a U.S. university is selecting students.

Suppose that the admission committee decides to randomly choose seven graduate students from a pool of 30 applicants, of whom 20 are foreign and 10 are U.S. applicants.

What is the probability that a chosen seven will have four foreign students and three U.S. students?

Try yourself – solution

No. of ways of selecting 7 applicants out of 30 = $\binom{30}{7}$

No. of ways of selecting 4 foreign and 3 U.S. students = $\binom{20}{4} \times \binom{10}{3}$

Hence, required probability

$$= \frac{\binom{20}{4} \times \binom{10}{3}}{\binom{30}{7}} = \cdots = 0.286$$

IF THIS LECTURE WAS HEAVY FOR YOU...

It is strongly recommended to go through Ramachandran and Tsokos, and to do the solved examples yourself.