1. Fundamental of Probability

1.1 What's Probability

Examples

1. Coin toss
    - "H" - head
    - "T" - tail
2. Roll a dice
    - every number in the set: $\{1, 2, 3, 4, 5, 6\}$
3. Tomorrow weather
    - {sunny, rainy, cloudy,...}
4. Randomly pick a number in $[0, 1]$

Although things are random, they are not haphazard/arbitrary. There are "patterns"

Example 1

If we repeat tossing a coin, then the fraction of times that we get a "H" goes to $\frac{1}{2}$ as the number of toss goes to infinity.

$$\frac{\text{\# of "H"}}{\text{total \# of toss}} = \frac{1}{2}$$

This number $\frac{1}{2}$ reflects how "likely" a "H" will appear in one toss (Even if the experiment is not repeated)

1.2 Probability Models

The *Sample space* Ω is the set consisting of al the possible outcomes of a random experiment.

Examples

1. $\{H, T\}$
2. $\{1, 2, 3, 4, 5, 6\}$
3. $\{sunny, rainy, cloudy, ...\}$
4. $[0, 1]$

An *event* $E \in \Omega$ is a subset of Ω

for which we can talk about "likelihood of happening"; for example

- in **2**:
  - {getting an even number} = {2, 4, 6}
- in **4**:
  - $\{\text{the point is between 0 and } 1/3\} = [0, 1/3]$ is an event
  - $\{\text{the point is rational}\} = Q \cap [0, 1]$

We say an even E "happens", if the result of the experiment turns out to belong to E (a subset of Ω)

A probability P is a set function (a mapping from events to real numbers)

$$\begin{aligned} P :&\xi \rightarrow R \\ &E \rightarrow P(E) \end{aligned}$$

which satisfies the following 3 properties:

1. $\forall E \in \xi, 0 \leq P(E) \leq 1$
2. $P(\Omega) = 1$
3. For
   - countably many disjoint events $E_1, E_2, ...$, we have $P(U_{i=1}^{\infty} E_i) = \sum_{i=1}^{\infty} P(E_i)$
   - countable: $\exists$ 1-1 mapping to natural numbers $1, 2, 3, ...$

Intuitively, one can think the probability of an event as the "likelihood/chance" for the event happens. If we repeat the experiment for a large number of events, the probability is the fraction of time that the event happens

$$P(E) = \lim_{n \rightarrow \infty} \frac{\text{\# of times the E happens in n ttrials}}{n}$$

Example 2 (cont'd)

$$\begin{gathered} P(\{1\}) = P(\{2\}) = \ldots = P(\{6\}) = \frac{1}{6} \\ E = \{\text{even number}\} = \{2, 4, 6\} \\ \Rightarrow P(E) = P(\{2\} \cup P(\{4\})) \cup P(\{6\}) = \frac{1}{2} \end{gathered}$$

Properties of probability:

1. $P(E) + P(E^c) = 1$
2. $P(\emptyset) = 0$
3. $E_1 \subseteq E_2 \Rightarrow P(E_1) \leq P(E_2)$
4. $P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 \cap E_2)$
   - $P(E_1 \cap E_2)$: E_1 and E_2 happen

Remark: why do we need the notion of event?

If the sample space Ω is **discrete**, then everything can has at most countable elements be built from the "atoms"

$$\Omega = \{w_1, w_2, \ldots\}$$
$$P(w_1) = P_i$$
$$P_i \in [0,1], \sum_{i=1}^{\infty} P_i = 1$$

Then for any event $E = \{w_1, i \in I\}$, $P(E) = \sum_{i \in I} P_i$

However, if the sample space Ω is continuous; e.g, $[0,1]$ in example 4, then such a construction can not be done for any $x \in [0,1]$ we get $P(\{x\} = 0$ (x: the point is exactly x)

We can not get $P([0,1/3])$ by adding $P(\{x\})$ for $x \leq 1/3$.

This is why we need the notion of event; and we define P as a set function from ξ to R rather than a function from Ω to R

To summarize: A **Probability Space** consists of a triplet (Ω, ξ, P):

- Ω: sample space,
- ξ: collection of events
- P: probability

1.3 Conditional Probability

If we know some information, the probability of an event can be updated

Let E, F be two events $P(F) > 0$

The conditional probability of E, given F is

$$P(E \mid F) = \frac{P(E \cap F)}{P(F)}$$

Again, think probability as the long-run frequency:

$$P(E \cap F) = \lim_{n \to \infty} \frac{\text{\# of times E and F happen in n trails}}{n}$$
$$P(F) = \lim_{n \to \infty} \frac{\text{\# of times F happen in n trails}}{n}$$
$$\Rightarrow \frac{P(E \cap F)}{P(F)} = \lim_{n \to \infty} \frac{\text{\# of times E and F happen}}{\text{\# of times F happen}}$$

By definition

$$P(E \cap F) = P(E \mid F) \cdot P(F)$$