Probability space \( (\Omega, \mathcal{F}, P) \):

- Sample space \( \Omega \)
- Set of events \( \mathcal{F} \)
- Probability measure \( P \):

\[ P : \mathcal{F} \to \mathbb{R} \]

Satisfies the (Kolmogorov) axioms of probabilities:

1. \( P(A) \geq 0 \) for all \( A \in \mathcal{F} \)
2. \( P(\Omega) = 1 \)
3. Countable additivity: If \( A_1, A_2, \ldots \in \mathcal{F} \) are disjoint (i.e., \( A_i \cap A_j = \emptyset \) for all \( i \neq j \)),
   \[ P\left( \bigcup_{i=1}^{\infty} A_i \right) = \sum_{i=1}^{\infty} P(A_i) \]

Basic laws of probability:

1. \( P(A^c) = 1 - P(A) \)
2. If \( A \subseteq B \), then \( P(A) \leq P(B) \)
3. \( P(A \cup B) = P(A) + P(B) - P(A \cap B) \)
4. \( P(A \cup B) \leq P(A) + P(B) \)
(4) Union of events bound:

\[ P(\bigcup_{i=1}^{\infty} A_i) \leq \sum_{i=1}^{\infty} P(A_i) \]

(5) Law of total probability: Let \( A_1, A_2, \ldots \) partition \( \Omega \) (i.e., \( A_1, A_2, \ldots \) be disjoint and \( \bigcup_{i=1}^{\infty} A_i = \Omega \)). Then

\[ P(B) = \sum_{i=1}^{\infty} P(A_i \cap B) \]

\[ \Omega \]

\[ A_1 \]

\[ A_2 \]

\[ A_3 \]

\[ A_4 \]

\[ A_5 \]

\[ A_6 \]

\[ A_7 \]

\[ B \]

\[ \Omega \]

Examples

(1) Flip a coin: \( \Omega = \{H, T\} \), \( \mathcal{F} = 2^\Omega = \{\emptyset, \{H\}, \{T\}, \{H, T\}\} \)

\[ P(\emptyset) = 0, \quad P(\{H\}) = p, \quad P(\{T\}) = 1-p, \quad P(\Omega) = 1 \]

The coin has "bias" \( p \)

(2) Flip a coin twice: \( \Omega = \{H, T\} \times \{H, T\} \)

\[ \mathcal{F} = 2^\Omega \]

\[ P(\emptyset) = 0, \quad P(\{(H, H)\}) = p^2, \quad P(\{(H, T)\}) = p(1-p) = P(\{(T, H)\}) \]

\[ P(\{T,T\}) = (1-p)^2, \quad P(\{(H, H), (H, T)\}) = p, \quad \ldots, \quad P(\Omega) = 1 \]
(3) Flip a coin $n$ times: 
\[ \Omega = \{H,T\}^n, \quad 2^n, \]
\[ P(\emptyset) = 0, \quad \ldots, \quad P(\Omega) = 1 \]

(4) Flip a coin until the first head:
\[ \Omega = \{H, TH, TTH, \ldots, \}, \]
\[ 2^n, \quad P \]

(5) Roll a fair die: 
\[ \Omega = \{1, 2, 3, 4, 5, 6\}, \quad 2^6 \]
\[ P(\emptyset) = 0, \quad P(\{1\}) = \ldots = P(\{6\}) = \frac{1}{6}, \]
\[ P(\{1, 2\}) = P(\{1, 2\}) = \ldots = P(\{4, 5, 6\}) = \frac{1}{3}, \]
\[ P(\{1, 2, 3\}) = \ldots = P(\{4, 5, 6\}) = \frac{1}{2}, \]
\[ P(\Omega) = 1 \]

When $\Omega$ is discrete (finite or countably infinite) and $\mathcal{F} = 2^\Omega$,
\[ P(A) = \sum_{\omega \in A} P(\{\omega\}) \]

In other words, probabilities of singletons determine $P$

(6) Pick a random number between 0 and 1
\[ \Omega = [0, 1] \]
\[ \mathcal{F} = \text{Borel field} = \text{"smallest} \ \sigma\text{-algebra that contains open intervals in } [0, 1] \text{"} \]
\[ P((a, b)) = b - a \text{ for all } a, b \text{ such that } 0 \leq a < b \leq 1 \]

When $\Omega$ is continuous (e.g., an interval or $\mathbb{R}$) and $\mathcal{F}$ is Borel, $P((a, b)), \forall a < b$, determines $P(A)$ for all $A \in \mathcal{F}$
For instance, \( P((a,b]) = \lim_{c \to b}^{c > b} P((a,c]) \cap_{c \text{ rational}} \)

A similar conclusion can be made with \( P((a,b]) \neq a < b \)

For instance, \( P((a,b)) = \lim_{c \to b}^{c < b} P((a,c)) \cap_{c \text{ rational}} \)

So far a probability measure is similar to any other measure (such as length, area, volume, weight, etc.).

- **Conditional probability**

Let \( B \) be an event with \( P(B) \neq 0 \). Then

\[
P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A,B)}{P(B)} \quad \text{notation}
\]

reads "conditional probability of \( A \) given \( B \)"

**Note:** \( P(\cdot | B) \) is a probability measure over \( \Omega \):

1. \( P(A|B) \geq 0 \), \( \forall A \)
2. \( P(\Omega|B) = \frac{P(B)}{P(B)} = 1 \)
3. \( P(\cup A_i|B) = \frac{P(U(A_i \cap B))}{P(B)} = \frac{P(U(A_i \cap B))}{P(B)} = \sum_{i=1}^{\infty} \frac{P(A_i \cap B)}{P(B)} = \sum P(A_i|B) \)
\textbf{Chain rule}

\[ P(A, B) = P(A) P(B|A) = P(B) P(A|B) \]

- True even when \( P(A) \) or \( P(B) = 0 \)

- Can be generalized to \( n \) events; e.g.,

\[ P(A_1, A_2, A_3) = P(A_1) P(A_2|A_1) P(A_3|A_1A_2) \]

- If \( P(B) \neq 0 \), then

\[ P(A|B) = \frac{P(B|A)}{P(B)} \cdot P(A) \]

\[ \text{posterior prob of } A \text{ (given } B) \]

\[ \text{prior prob of } A \]

\textbf{Bayes rule}

Let \( A_1, A_2, \ldots, A_n \) be nonzero probability events that partition \( \Omega \). Suppose we wish to compute

\[ P(A_j|B) \]

based on \( P(A_i) \) and \( P(B|A_i) \), \( i = 1, 2, \ldots, n \)

We know that

\[ P(A_j|B) = \frac{P(B|A_j) P(A_j)}{P(B)} \]

and

\[ P(B) = \sum P(A_i) P(B|A_i) \]
Hence,

\[ p(A_j | B) = \frac{p(B | A_j) \cdot p(A_j)}{\sum p(B | A_i) \cdot p(A_i)} \]  
(Bayes rule)

**Examples**

(1) **Binary communication channel**

\[ p(0) = 0.2 \quad p(1) = 0.8 \]

\[ p(0|0) = 0.9 \quad p(0|1) = 0.1 \quad p(1|0) = 0.9 \quad p(1|1) = 0.1 \]

\[ \Omega = \{ (0,0), (0,1), (1,0), (1,1) \}^3 \]

Let \( A = \{ 0 \text{ is sent } \} \) and \( B = \{ 0 \text{ is received} \} \)

Find \( p(A), p(B|A), p(B), p(A|B) \).

(2) **Finite state machine**

\[ \Omega = \{ (\text{initial state}, \text{next state}) \}^2 = \{ \xi, \beta, \alpha \} \]

\[ p(\alpha) = 0.5 \]

Let \( A = \{ \text{initial state is } \alpha \} \) and \( B = \{ \text{next state is } \alpha \} \)

Find \( p(A), p(B|A), p(B), p(A|B) \).