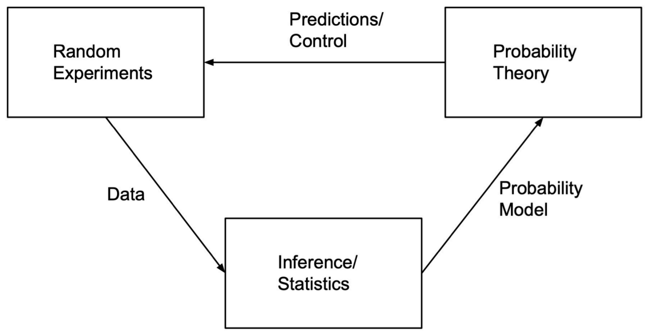
EE5110: PROBABILITY FOUNDATIONS FOR ELECTRICAL ENGINEERS

JULY - NOV 2024

1 Introduction to Probability

1. Definitions of probability
    - classical probability
    - relative frequency
    - subjective belief
    - axiomatic probability
2. In all versions, probability is non-negative, normalized and additive!
3. Andrey Kolmogorov formalized the field and provided a rigorous mathematical foundation with the axiomatic framework[1].
    - the axiomatic framework is based on measure theory
    - there are no paradoxes with the framework!
4. A brief history
    - 18th century: Jacob Bernoulli proposes law of large numbers; De Moivre proves the first form of central limit theorem; contributions by Daniel Bernoulli, Leibnitz, Bayes, Lagrange
    - 19th century: contributions by Laplace, Poisson, Chebyshev, Markov, Lyapunov; Legendre and Gauss apply probability theory to astronomical predictions; probability is viewed as a natural science;
    - 20th century: contributions by Borel, Lebesgue; Kolmogorov introduces the axiomatic framework for probability theory; probability is viewed as a branch of mathematics;
    - Probability theory is applied in physics for statistical mechanics and quantum mechanics, in biology for genetic studies and epidemiology, in finance for risk assessment and stock market analysis, in engineering for reliability testing, signal processing, and control systems, and in computer science for algorithms and artificial intelligence. It is crucial in fields that require modeling uncertainty and predicting outcomes.
5. Probability theory is a mathematical framework to study experiments whose outcomes are not deterministic.

[1] Foundations of the Theory of Probability, 1956



1.1 Axioms of Probability

1. A random experiment
   - the outcome of the experiment is not deterministic
2. Study begins with the description of a probability space $(\Omega, \mathcal{F}, \mathsf{P})$
   - sample space Ω: set of all outcomes
   - event space $\mathcal{F}$: a sigma-algebra of subsets of Ω
   - probability measure P: a non-negative, normalized and countably additive measure
3. Sample space, Ω, is the set of possible outcomes of the experiment
   - elements of the sample space are called sample points, ω
   - outcome of a random experiment is mapped to one and only one $\omega \in \Omega$
   - sample space must be non-empty, exhaustive, distinguishable and at the right granularity
   - sample space can be finite, countable or uncountable
4. Event space, $\mathcal{F}$, is a sigma-algebra of subsets of Ω
   - $\mathcal{F}$ is non-empty, closed under complement and countable union, i.e., $\Omega \in \mathcal{F}$, if $A \in \mathcal{F}$ then $A^c \in \mathcal{F}$, and if $\{A_i\}_{i=1}^{\infty}$ are in $\mathcal{F}$, then, $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$.
   - elements in $\mathcal{F}$ (subsets of Ω) correspond to events of interest
   - event $A \in \mathcal{F}$ is said to occur if $\omega \in A$ occurs
5. Probability measure $\mathsf{P} : \mathcal{F} \to \mathcal{R}$
   - P is non-negative, i.e., $\mathsf{P}(A) \geq 0$ for all $A \in \mathcal{F}$

- $\mathsf{P}(\Omega) = 1$ (measure is normalized)
- (countable additivity) Suppose that $\{A_i\}_{i=1}^{\infty}$ are mutually exclusive events in $\mathcal{F}$, then,

$$\mathsf{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathsf{P}(A_i)$$

- $\mathsf{P}(A)$ denotes the probability of the occurence of event A, for $A \in \mathcal{F}$
- $\Omega \in \mathcal{F}$ is called the certain event and $\phi \in \mathcal{F}$ is called the impossible event!

6. A discrete probability space for a single coin-toss experiment:
   - $\Omega = \{H, T\}$ (a finite sample space)
   - $\mathcal{F} = \{\Omega, \{H\}, \{T\}, \phi\}$ (power set of Ω)
   - $\mathsf{P}(\cdot) = \{1, p, 1-p, 0\}$ for $0 \leq p \leq 1$ is a valid probability measure (Verify!)
   - Note that we denote probability of occurrence of the outcome Head as $\mathsf{P}(\{\text{H}\})$ instead of $\mathsf{P}(\text{H})$[2].
7. Can you describe a probability space for an uncountable sample space?

1.2 Properties

1. $\Omega \in \mathcal{F}$ and $\phi \in \mathcal{F}$

Proof: $\mathcal{F}$ is non-empty. Hence, there exists an $A \in \mathcal{F}$.
Given $A \in \mathcal{F}$, then $A^c \in \mathcal{F}$ as well (as $\mathcal{F}$ is closed under complement).
Let $A_1 = A, A_2 = A^c, A_3 = A^c, A_4 = A^c, \cdots$ be a countable collection of events in $\mathcal{F}$. Then, their countable union $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$. But, $\bigcup_{i=1}^{\infty} A_i = A_1 \cup A_2 = A \cup A^c = \Omega$. Hence, $\Omega \in \mathcal{F}$.
Given $\Omega \in \mathcal{F}$, $\Omega^c = \phi \in \mathcal{F}$ as $\mathcal{F}$ is closed under complement. □

2. If $\{A_i\}_{i=1}^{\infty}$ are in $\mathcal{F}$, then $\bigcap_{i=1}^{\infty} A_i \in \mathcal{F}$.

Proof: Given $\{A_i\}_{i=1}^{\infty}$ are in $\mathcal{F}$, then $\{A_i^c\}_{i=1}^{\infty}$ are in $\mathcal{F}$, as $\mathcal{F}$ is closed under complement.
Given $\{A_i^c\}_{i=1}^{\infty}$ are in $\mathcal{F}$, then $\bigcup_{i=1}^{\infty} A_i^c \in \mathcal{F}$ (as $\mathcal{F}$ is closed under countable union).
Given $\bigcup_{i=1}^{\infty} A_i^c \in \mathcal{F}$, then $\left(\bigcup_{i=1}^{\infty} A_i^c\right)^c \in \mathcal{F}$ (as $\mathcal{F}$ is closed under complement). But, $\left(\bigcup_{i=1}^{\infty} A_i^c\right)^c = \bigcap_{i=1}^{\infty} A_i$. Hence, $\bigcap_{i=1}^{\infty} A_i \in \mathcal{F}$. □

3. $\mathsf{P}(\phi) = 0$.

[2] $\mathsf{P}(\{\text{Head}\})$ is the recommended usage (since $\mathsf{P} : \mathcal{F} \to \mathcal{R}$) but we will also refer it as $\mathsf{P}(\text{Head})$ sometimes (out of laziness).

Proof: $\phi = \Omega^c \in \mathcal{F}$. Hence, $\mathsf{P}(\phi)$ is well defined.

Let $A_1 = \Omega, A_2 = \phi, A_3 = \phi, A_4 = \phi, \cdots$ be events in $\mathcal{F}$. Clearly, $\{A_i\}$ are mutually exclusive. Then, from countable additivity axiom, we have,

$$\mathsf{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathsf{P}(A_i) = \mathsf{P}(A_1) + \sum_{i=2}^{\infty} \mathsf{P}(A_2) = \mathsf{P}(\Omega) + \sum_{i=2}^{\infty} \mathsf{P}(\phi)$$

We know that $\bigcup_{i=1}^{\infty} A_i = A_1 \cup A_2 = \Omega$. Hence, $\mathsf{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \mathsf{P}(\Omega) = 1$. Substituting in the above expression, we get,

$$1 = \mathsf{P}(\Omega) + \sum_{i=2}^{\infty} \mathsf{P}(\phi)$$

or,

$$0 = \sum_{i=2}^{\infty} \mathsf{P}(\phi)$$

This implies that $\mathsf{P}(\phi) = 0$. $\square$

4. (Finite Additivity) If $A_1, A_2, \cdots, A_n$ are disjoint events in $\mathcal{F}$, then

$$\mathsf{P}\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} \mathsf{P}(A_i)$$

(Countable additivity implies finite additivity, but not the other way around.)

Proof: Let $A_1, A_2, \cdots, A_n$ be disjoint events in $\mathcal{F}$.

Define $A_{n+1} = \phi, A_{n+2} = \phi, A_{n+3} = \phi, \cdots$. Then, $\{A_i\}_{i=1}^{\infty}$ are disjoint events in $\mathcal{F}$.

From countable additivity axiom, we have,

$$\mathsf{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathsf{P}(A_i) = \sum_{i=1}^{n} \mathsf{P}(A_i) + \sum_{i=n+1}^{\infty} \mathsf{P}(\phi) = \sum_{i=1}^{n} \mathsf{P}(A_i)$$

(as $\mathsf{P}(\phi) = 0$.)

But, we know that $\bigcup_{i=1}^{\infty} A_i = \bigcup_{i=1}^{n} A_i$. Hence, substituting in the above expression, we have,

$$\mathsf{P}\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} \mathsf{P}(A_i)$$

In particular, if A_1, A_2 are disjoint events in $\mathcal{F}$, then $\mathsf{P}(A_1 \cup A_2) = \mathsf{P}(A_1) + \mathsf{P}(A_2)$. $\square$

5. For $A \in \mathcal{F}$, $\mathsf{P}(A^c) = 1 - \mathsf{P}(A)$.

Proof: Given $A \in \mathcal{F}$, then $A^c \in \mathcal{F}$. And, $\mathsf{P}(A^c)$ is well-defined.
Note that A and A^c are disjoint events in $\mathcal{F}$. Hence, from finite additivity, we have,

$$\mathsf{P}(A \cup A^c) = \mathsf{P}(A) + \mathsf{P}(A^c)$$

But, $A \cup A^c = \Omega$ and $\mathsf{P}(\Omega) = 1$. Hence,

$$1 = \mathsf{P}(A) + \mathsf{P}(A^c)$$

or,

$$1 - \mathsf{P}(A) = \mathsf{P}(A^c)$$

□

6. (Monotonicity) If $A, B \in \mathcal{F}$ and $A \subset B$, then $\mathsf{P}(A) \leq \mathsf{P}(B)$. In particular, $\mathsf{P}(A) \leq 1$ for all $A \in \mathcal{F}$.

Proof: Let $A \subset B$. Then, $B = A \cup (A^c B)$.
Clearly, A and $A^c B$ are disjoint events in $\mathcal{F}$. Hence, from finite additivity, we have,

$$\mathsf{P}(B) = \mathsf{P}(A) + \mathsf{P}(A^c B) \geq \mathsf{P}(A)$$

where the last inequality follows from the fact that $\mathsf{P}(A^c B) \geq 0$. □

7. (Inclusion-Exclusion Principle) If A_1 and A_2 are events in $\mathcal{F}$, then

$$\mathsf{P}(A_1 \cup A_2) = \mathsf{P}(A_1) + \mathsf{P}(A_2) - \mathsf{P}(A_1 \cap A_2)$$

Proof: Let A_1 and A_2 be events in $\mathcal{F}$. Then, $A_1 \cup A_2$ is in $\mathcal{F}$ and its probability is well-defined.
Define $A_1 \cup A_2 = A_1 \cup (A_1^c A_2)$. Then, A_1 and $A_1^c A_2$ are disjoint events in $\mathcal{F}$. Hence, from finite additivity, we have,

$$\mathsf{P}(A_1 \cup A_2) = \mathsf{P}(A_1) + \mathsf{P}(A_1^c A_2)$$

We can also define $A_2 = A_2 \cap \Omega = A_2 \cap (A_1 \cup A_1^c) = (A_2 A_1) \cup (A_2 A_1^c)$. Clearly, $A_2 A_1$ and $A_2 A_1^c$ are disjoint events in $\mathcal{F}$. Hence, from finite additivity, we have,

$$\mathsf{P}(A_2) = \mathsf{P}(A_2 A_1) + \mathsf{P}(A_2 A_1^c)$$

Substituting for $\mathsf{P}(A_1^c A_2)$ in the former expression, we have,

$$\mathsf{P}(A_1 \cup A_2) = \mathsf{P}(A_1) + \mathsf{P}(A_2) - \mathsf{P}(A_2 A_1)$$

□

8. (Inclusion-Exclusion Principle) If $A_1, A_2, \cdots, A_n$ are events in $\mathcal{F}$, then

$$\mathsf{P}\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} \mathsf{P}(A_i) - \sum_{1 \le i < j \le n} \mathsf{P}(A_i A_j) + \cdots + (-1)^n \mathsf{P}(A_1 A_2 \cdots A_n)$$

(Hint: Use mathematical induction.)

9. (Continuity) Let $A_1, A_2, \cdots$ be events in $\mathcal{F}$. Then,

$$\mathsf{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \lim_{n\to\infty} \mathsf{P}\left(\bigcup_{i=1}^{n} A_i\right)$$

If $A_1 \subset A_2 \subset A_3 \cdots$ be events in $\mathcal{F}$, then,

$$\mathsf{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \lim_{n\to\infty} \mathsf{P}\left(A_n\right)$$

10. (Continuity) Let $A_1, A_2, \cdots$ be events in $\mathcal{F}$. Then,

$$\mathsf{P}\left(\bigcap_{i=1}^{\infty} A_i\right) = \lim_{n\to\infty} \mathsf{P}\left(\bigcap_{i=1}^{n} A_i\right)$$

If $A_1 \supset A_2 \supset A_3 \cdots$ be events in $\mathcal{F}$, then,

$$\mathsf{P}\left(\bigcap_{i=1}^{\infty} A_i\right) = \lim_{n\to\infty} \mathsf{P}\left(A_n\right)$$

11. (Union Bound) If $A_1, A_2, \cdots, A_n$ are events in $\mathcal{F}$, then

$$\mathsf{P}\left(\bigcup_{i=1}^{n} A_i\right) \le \sum_{i=1}^{n} \mathsf{P}(A_i)$$

(Also, see Bonferroni's inequalities.)

1.3 Uniform Probability Measure on $[0, 1]$

1. Let us consider the following valid probability space

   - $\Omega = \mathcal{R}$
   - $\mathcal{F} = \mathcal{B}(\mathcal{R})$
   - $\mathsf{P}([a, b]) = \int_{[a,b]\cap[0,1]} 1 \text{ dx} = \int_a^b 1_{[0,1]}(x) \text{ dx}$, for $-\infty < a < b < \infty$

2. The sample space can be $[0, 1]$ or all of $\mathcal{R}$.

3. Borel sigma-algebra of $\mathcal{R}$, $\mathcal{B}(\mathcal{R})$

   - smallest sigma-algebra containing open intervals $\{(a, b) : -\infty < a < b < \infty\}$
   - $\mathbb{B}(\mathbb{R})$ contains $\{a\}, [a, b], (a, b], [a, b), (-\infty, b), (-\infty, b], (a, \infty), [a, \infty), \mathbb{Q}, \mathbb{P}, \mathbb{Z}, \mathbb{R}$
   - $\mathbb{B}(\mathbb{R})$ is not the power-set of $\mathbb{R}$
   - $\mathbb{B}(\mathbb{R})$ is not the sigma-algebra generated by singleton sets

4. If $A \in \mathbb{B}(\mathbb{R})$, then $A \cap \mathbb{B}(\mathbb{R})$ is also a sigma-algebra.

   - for example, $[0, 1] \cap \mathbb{B}(\mathbb{R}) = \mathbb{B}([0, 1])$ is also a sigma-algebra

5. (Uniform probability measure) $\mathsf{P}([a, b])$ for $0 \leq a < b \leq 1$ can be computed as

$$\mathsf{P}([a, b]) = \int_a^b 1 \text{ dx} = (b - a)$$

   You may consider using Riemann integration for the exercise above.

6. Compute $\mathsf{P}([a, b] \cup [c, d])$ for $0 \leq a < b < c < d \leq 1$.

$$\begin{aligned} \mathsf{P}([a, b] \cup [c, d]) &= \int_0^1 1_{[a,b]\cup[c,d]}(x) \text{ dx} \\ &= \int_a^b 1 \text{ dx} + \int_c^d 1 \text{ dx} \\ &= (b - a) + (c - d) \end{aligned}$$

7. Compute $\mathsf{P}(\{a\})$ for $0 < a < 1$.

$$\begin{aligned} \mathsf{P}(\{a\}) = \mathsf{P}\left(\bigcap_{n=1}^{\infty} \left(a - \frac{1}{2n}, a + \frac{1}{2n}\right)\right) &= \lim_{n\to\infty} \mathsf{P}\left(\left(a - \frac{1}{2n}, a + \frac{1}{2n}\right)\right) \\ &= \lim_{n\to\infty} \left(\left(a + \frac{1}{2n}\right) - \left(a - \frac{1}{2n}\right)\right) \\ &= \lim_{n\to\infty} \frac{1}{n} = 0 \end{aligned}$$

8. Compute $\mathsf{P}(\mathbb{Q} \cap [0, 1])$.

$$\begin{aligned} \mathsf{P}(\mathbb{Q} \cap [0, 1]) &= \sum_{q\in\mathbb{Q}\cap[0,1]} \mathsf{P}(\{q\}) \\ &= \sum_{q\in\mathbb{Q}\cap[0,1]} 0 = 0 \end{aligned}$$

   Could you compute the integral $\int_{\mathbb{Q}\cap[0,1]} 1 \text{ dx} = \int_0^1 1_{\mathbb{Q}}(x) \text{ dx}$ using Riemann integration?

9. Uniform probability measure cannot be meaningfully defined for all subsets of $[0, 1]$.

   - notion of length measure cannot be extended for all subsets of $\mathbb{R}$
   - Borel sigma-algebra is a convenient and useful sigma-algebra of $\mathbb{R}$
   - the definitions and computations of probability measure for the continuous sample space and the Borel sigma-algebra are inspired by Lebesgue and Lebesgue-Stieltjes integration.

1.4 Examples of Probability Measures

For discrete distributions, $\mathsf{P}(A) = \sum_{i \in A} \mathsf{P}(\{i\})$ and for continuous distributions, $\mathsf{P}(A) = \int_A f(x)\ \mathrm{dx}$.

(a) **Bernoulli** $(p : 0 \le p \le 1)$: $(\{0, 1\}, 2^{\Omega}, \mathsf{P})$ where

$$\mathsf{P}(\{0\}) = 1 - p, \text{ and } \mathsf{P}(\{1\}) = p$$

(b) **Uniform** $(N : N \in \mathbb{N})$: $(\{1, 2, \cdots, N\}, 2^{\Omega}, \mathsf{P})$ where

$$\mathsf{P}(\{i\}) = \frac{1}{N} \text{ for all } \ i = 1, 2, \cdots, N$$

(c) **Geometric** $(p : 0 < p \le 1)$: $(\mathbb{N}, 2^{\Omega}, \mathsf{P})$ where

$$\mathsf{P}(\{i\}) = p\ (1 - p)^{i-1} \text{ for all } \ i \in \mathbb{N}$$

Show that $\mathsf{P}(\mathrm{Odd}) = \sum_{i=1}^{\infty} p(1-p)^{2(i-1)} = \frac{1}{2-p}$.

(d) **Uniform** $(a, b : a < b \in \mathbb{R}) : (\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathsf{P})$ where

$$\mathsf{P}([x, y]) = \int_{[x,y] \cap [a,b]} \frac{1}{b - a}\ \mathrm{du}$$

(e) **Exponential** $(\lambda : \lambda > 0)$: $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathsf{P})$ where

$$\mathsf{P}([x, y]) = \int_{[x,y] \cap [0,\infty)} \lambda e^{-\lambda u}\ \mathrm{du}$$

(f) **Gaussian** $(\mu, \sigma^2 : \mu, \sigma \in \mathbb{R})$: $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathsf{P})$ where

$$\mathsf{P}([x, y]) = \int_x^y \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(u-\mu)^2}{2\sigma^2}}\ \mathrm{du}$$