E1 222 Stochastic Models and Applications

P.S. Sastry
sastry@iisc.ac.in

Reference Material

- V.K. Rohatgi and A.K.Md.E. Saleh, An Introduction to probability and Statistics, Wiley, 2nd edition, 2018
- S.Ross, ‘Introduction to Probability Models’, Elsevier, 12th edition, 2019.
- P G Hoel, S Port and C Stone, Introduction to Probability Theory, 1971.
- P G Hoel, S Port and C Stone, Introduction to Stochastic Processes, 1971.
- Scott Sheffield, Probability and Random Variables, Massachusetts Institute of Technology, MIT OpenCourseWare: https://ocw.mit.edu/courses/mathematics/18-600-probability-and-random-variables-fall-2019/ License: Creative Commons BY-NC-SA.

Course Prerequisites / Background needed

- Calculus
  - continuity, differentiability, derivatives, functions of several variables, partial derivatives, integration, multiple integrals or integration over $\Re^n$, convergence of sequences and series, Taylor series
- Matrix theory
  - vector spaces, linear independence, linear transformations, matrices, rank, determinant, eigen values and eigen vectors
- In addition, knowledge of basic probability is assumed. I assume all students are familiar with the following:

  > *Random experiment, sample space, events, conditional probability, independent events, simple combinatorial probability computations*

  But we would review the basic probability in the first two classes.

Course grading

- Mid-Term Tests and Assignments: 70%
  Final Exam: 30%
- Two mid-term tests for 20 marks each. We will have 2-3 assignments for 10 marks. (Tentative)
- Please remember this is essentially a Maths course

Probability Theory

- Probability Theory – branch of mathematics that deals with modeling and analysis of random phenomena.
- Random (or Chance) Phenomena
  – individually not predictable but have a lot of regularity at a population level
- Recommender systems, opinion polls, sample surveys $\cdots$
  – useful because at a population level customer behaviour can be predicted.
- Statistics is the branch of Maths that deals with making inferences from data and Probability theory is needed for that.
- Example random phenomena: Tossing a coin, rolling a dice etc – familiar to you all
- It is also useful in many engineering systems, e.g., for analyzing behaviour under noise.

- In many engineering problems one needs to deal with random inputs where probability models are useful
  - Analysis of dynamical systems subjected to noise
  - System estimation
  - Policies for decision making under uncertainty
  - Pattern Recognition, prediction from data

    $\vdots$
- We may use probability models for analysing algorithms. (e.g., average case complexity of algorithms)
- We may deliberately introduce randomness in an algorithm

  (e.g., ALOHA protocol, Primality testing)

  $\vdots$

This is only a 'sample' of possible application scenarios!

Review of basic probability

We assume all of you are familiar with the terms:

random experiment, outcomes of random experiment, sample space, events etc.

We use the following Notation:

- Sample space – Ω
  Elements of Ω are the outcomes of the random experiment
  We write $\Omega = \{\omega_1, \omega_2, \cdots\}$ when it is countable
- An event is, by definition, a subset of Ω
- Set of all possible events – $\mathcal{F} \subset 2^{\Omega}$ (power set of Ω)
  Each event is a subset of Ω
  For now, we take $\mathcal{F} = 2^{\Omega}$ (power set of Ω)

Probability axioms

Probability (or probability measure) is a function that assigns a number to each event and satisfies some properties.

Formally, $P : \mathcal{F} \rightarrow \Re$ satisfying

A1 Non-negativity: $P(A) \geq 0, \forall A \in \mathcal{F}$

A2 Normalization: $P(\Omega) = 1$

A3 σ-additivity: If $A_1, A_2, \cdots \in \mathcal{F}$ satisfy $A_i \cap A_j = \phi, \forall i \neq j$ then

$$P(\cup_{i=1}^{n} A_i) = \sum_{i=1}^{n} P(A_i), \forall n; \quad \text{and} \quad P(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$$

Events satisfying $A_i \cap A_j = \phi, \forall i \neq j$ are said to be **mutually exclusive**

$(\Omega, \mathcal{F}, P)$ is called the **Probability Space**

Probability axioms

$P : \mathcal{F} \rightarrow \Re, \;\; \mathcal{F} \subset 2^{\Omega}$ (Events are subsets of Ω)

A1 $P(A) \geq 0, \; \forall A \in \mathcal{F}$

A2 $P(\Omega) = 1$

A3 If $A_i \cap A_j = \phi, \forall i \neq j$ then $P(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$

- For these axioms to make sense, we are assuming

$$\text{(i). } \Omega \in \mathcal{F} \;\; \text{ and } \;\; \text{(ii). } A_1, A_2, \cdots \in \mathcal{F} \Rightarrow (\cup_i A_i) \in \mathcal{F}$$

  When $\mathcal{F} = 2^{\Omega}$ this is true.
- Note that axiom $A3$ is about countable unions.
- As defined, the co-domain of the function P is $\Re$. However, the axioms imply that it takes values in $[0, 1]$.

Simple consequences of the axioms

- Notation:

  A^c is complement of A.

  $C = A + B$ implies $A \cap B = \phi$ and $C = A \cup B$.

- Let $A \subset B$ be events. Then $B = A + (B - A)$.

  $$P(B) = P\left(A + (B - A)\right) = P(A) + P(B - A) \geq P(A)$$

  Thus, $\quad A \subset B \;\Rightarrow\; P(A) \leq P(B)$

  This also shows $P(B - A) = P(B) - P(A)$ when $A \subset B$.

- If A is an event, then $A \subset \Omega$ and hence $P(A) \leq P(\Omega) = 1$.

- We can show $P(A^c) = 1 - P(A)$ as

  $$1 = P(\Omega) = P(A + A^c) = P(A) + P(A^c)$$

- There are many such properties (I assume familiar to you) that can be derived from the axioms.
- Here are a few important ones. (Proof is left to you as an exercise!)

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

$$\begin{aligned} P(U_{i=1}^{n} A_i) &= \sum_i P(A_i) - \sum_i \sum_{j>i} P(A_i \cap A_j) \\ &+ \sum_i \sum_{j>i} \sum_{k>j} P(A_i \cap A_j \cap A_k) - \cdots + (-1)^{n+1} P(\cap_i A_i) \end{aligned}$$

Known as inclusion-exclusion formula

Case of finite Ω – Example

- Let $\Omega = \{\omega_1, \cdots, \omega_n\}$, $\mathcal{F} = 2^\Omega$, and P is specified through 'equally likely' assumption.
- That is, $P(\{\omega_i\}) = \frac{1}{n}$. (Note the notation)
- Suppose $A = \{\omega_1, \omega_2, \omega_3\}$. Then

$$P(A) = P(\{\omega_1\} \cup \{\omega_2\} \cup \{\omega_3\}) = \sum_{i=1}^{3} P(\{\omega_i\}) = \frac{3}{n} = \frac{|A|}{|\Omega|}$$

- We can easily see this to be true for any event, A.
- This is the usual familiar formula: number of favourable outcomes by total number of outcomes.
- Thus, 'equally likely' is one way of specifying the probability function (in case of finite Ω).
- An obvious point worth remembering: specifying P for singleton events fixes it for all other events.

Case of Countably infinite Ω

- Let $\Omega = \{\omega_1, \omega_2, \cdots\}$.
- Once again, any $A \subset \Omega$ can be written as mutually exclusive union of singleton sets.
- Let $q_i, i = 1, 2, \cdots$ be numbers such that $q_i \geq 0$ and $\sum_i q_i = 1$.
- We can now set $P(\{\omega_i\}) = q_i, i = 1, 2, \cdots$.
  (Assumptions on q_i needed to satisfy $P(A) \geq 0$ and $P(\Omega) = 1$).
- This fixes P for all events: $P(A) = \sum_{\omega \in A} P(\{\omega\})$
- This is how we normally define a probability measure on countably infinite Ω.
- This can be done for finite Ω too.

Example: countably infinite Ω

- Let $\Omega = \{0, 1, 2, \cdots\}$
- Let $q_i = (1-p)^i p$ for some p, $0 < p < 1$.
- Easy to see: $q_i \geq 0$ and $\sum_{i=0}^{\infty} q_i = 1$.
- We can assign $P(\{k\}) = (1-p)^k p, k = 0, 1, \cdots$
- Consider a random experiment of tossing a biased coin repeatedly till we get a head. We take the outcome of the experiment to be the number of tails we had before the first head.
- A (reasonable) probability assignment is:

$$P(\{k\}) = (1-p)^k p, k = 0, 1, \cdots$$

  where p is the probability of head and $0 < p < 1$.
  (We assume you understand the idea of 'independent' tosses here).