MA 6.101
Probability and Statistics

Tejas Bodas

Assistant Professor, IIIT Hyderabad

Logistics

- Feel free to contact me anytime at tejas.bodas@iiit.ac.in.
- Office @ A5304.
- TA list: Around 12 TAs, you will meet them during tutorials
- Lectures: Wednesday and Saturday 10:00 to 11:25
- Tutorial on Wed 11:40 to 1:00.
- Phones in pocket, laptops in bag!

Resources

- Wont be following any one particular book.
- Lecture slides will have material from variety of sources.
- Some popular books
  1. Introduction to probability by Bertsekas and Tsisiklis (Athena Scientific)
  2. Intro. to Probability and Statistics for Engineers and Scientists by Sheldon Ross (Elsevier)
  3. A first course in probability by Sheldon Ross (Prentice Hall)
- Some urls
  1. https://www.statlect.com/
  2. https://www.randomservices.org/
  3. https://www.probabilitycourse.com/

Evaluation scheme

- Quiz 1 : 10%
- Assignment 1: 10%
- Midsem exam: 25%.
- Quiz 2: 10%
- Assignment 2: 10%
- Endsem 35 %.

Course Outline

- Module 1 (4 Lectures)
  Motivation & Probability basics
- Module 2 (6 Lectures)
  All about random variables!
- Module 3 (4 Lectures)
  Random processes
- Module 4 (10 lectures)
  Probability inequalities and Statistics

Where is probability & statistics useful?

- Machine learning
- Reinforcement learning
- Insurance and Finance
- Inventory control & (dynamic) pricing
- Analysis of Computer systems (routing, scheduling)
- Forecasting (weather, demand)
- Biostatistics (clinical trials)

How is it used in all these fields?

- Draw inferences from the underlying data
- Theoretical tool to establish performance guarantees in such systems.
- Establish proof of concept for robustness of algorithms under randomness.

Laws

- There are laws of physics
- and there are laws of chemistry

This course is about laws of uncertainty/randomness.

Philosophy of Probability

Randomness in one way is defined as our inability to have a deterministic say about an outcome of some event.

For example in a coin toss, if you could mathematically capture all possible influences on the coin exactly (force of tossing, metal density, temperature, wind speed etc) you can predict the outcome of each coin toss experiment with certainty.

Probability theory is all about finding regularity and patterns is seemingly random experiments (experiments lacking a deterministic understanding of it) and expressing them mathematically.

Random experiments and Sample space

- Random experiment : Experiment involving randomness
  - Coin toss
  - Roll a dice
  - Pick a number at random from $[0, 1]$.
- Sample space Ω: set of all possible outcomes of the random experiment. It could be a finite or infinite set.
  - $\Omega_c = \{H, T\}$
  - $\Omega_d = \{1, 2, \ldots, 6\}$
  - $\Omega_u = [0, 1]$
  - $\Omega_{2c} = \{H, T\} \times \{H, T\} = \{(H, H), (H, T), (T, H), (T, T)\}$
- Sample space Ω should satisfy the following.
  - *Mutually exclusive outcomes*: When the random experiment is performed, only one of the things in Ω can happen.
  - *Mutually exhaustive outcomes*: Atleast one of the things in Ω can happen.

Outcomes and Events

- Element $\omega \in \Omega$ is called a **sample point** or possible outcome.
- We use the notation $\bar{\omega}$ to denote a **realized outcome**.
- A subset $A \subseteq \Omega$ is called an **event**.
- Examples of events
  - Events in the coin experiment: $C_1 = \{T\}$.
  - Events in the dice experiment: $D_1 = 6, D_2 = \{1, 3, 5\}$
  - Events in $U[0, 1]$ experiment: $U_1 = \{0.5\}$, $U_2 = [.25, .75]$.
- In this course, we are interested in probability of events.
- Probability of event A is denoted by $\mathbb{P}(\bar{\omega} \in A)$ or simply $\mathbb{P}(A)$.
- It may not be possible to measure/assign probability for every subset A (more later).
- Any guesses for $\mathbb{P}(C_1), \mathbb{P}(D_1), \mathbb{P}(D_2), \mathbb{P}(U_1)$ and $\mathbb{P}(U_2)$?

Probability theory

$\{$Random experiment, Sample space, Events$\}$ are the key ingredients in probability theory.

In probability theory, we are interested in **measuring** the probability of subsets of Ω (events).

Probability measure $\mathbb{P}$ is a **set function**, i.e. it acts on sets and measures the probability of such sets.

Set theory 101

Visualizing operations on events using Venn diagram!

- Compliments: A^c
- $\emptyset$ denotes empty set.$\emptyset \subseteq A$ for all A.
- Union: $A \cup B$
- Intersections: $A \cap B$
- Difference: $A \backslash B$
- Symmetric difference:
- Mutually exclusive or disjoint events A and B:
- Identity laws,Compliment laws,Associative, Commutative & Distributive laws, De'Morgans law.

Set theory 101–Cardinality & Countability

- Cardinality of A is denoted by $|A|$.
- Inclusion-exclusion principle $|A \cup B| = |A| + |B| - |A \cap B|$.
- Inclusion-exclusion principle for n sets ?
- Countable sets: Set A is said to be countable if it is either finite or has 1-1 correspondence with natural numbers $\mathbb{N}$.
- Uncountable sets: These are sets which are not countable.

Set theory 101 – Monotone sequence of sets

- Increasing sequence $A_1 \subseteq A_2 \subseteq A_3 \ldots$

- Decreasing sequence $A_1 \supseteq A_2 \supseteq A_3 \ldots$

- Examples from $U[0, 1]$:
  - $I_n = [0, 1 - \frac{1}{n}]$

  - $D_n = [0, \frac{1}{n}]$

Set theory 101 – Cartesian product of sets

- Cartesian product of sets A and B is denoted by $A \times B$.

- $A \times B$ is itself a set whose members are sets of the form (a, b) where $a \in A$ and $b \in B$.

- Suppose $A = \{1, 3, 5\}$ and $B = \{2, 4, 6\}$ what is $A \times B$?

- What is $[0, 1] \times [0, 1]$? unit square!

Set theory 101 – Powersets

Powerset of A is denoted by $\mathcal{P}(A)$ is a set whose members are all possible subsets of A. ($\mathcal{P}$ and $\mathbb{P}$ are different!)

- What is $\mathcal{P}(\Omega_c)$?
- What is $\mathcal{P}(\Omega_d)$?
- What is $\mathcal{P}(\Omega_u)$?
- What is the cardinality of $\mathcal{P}(\Omega_c), \mathcal{P}(\Omega_d), \mathcal{P}(\Omega_u)$?
- For discrete sets Ω, often the power set is denoted by 2^{Ω}.

functions and set functions

- What are functions? Functions are rules or maps that map elements from a domain $\mathcal{D}$ to elements in the range $\mathcal{R}$.
- $f : \mathcal{D} \to \mathcal{R}$.
- Example: $f : \mathbb{R} \to \mathbb{R}$ defined by $f(x) = x$.
- Read more on injection, surjection, bijection!
- What are set functions? these are functions that act on sets and hence domain $\mathcal{D}$ is a collection of sets.
- Example: length of closed segments on the real line.
- $l : \mathcal{D} \to \mathbb{R}_+$ where $\mathcal{D} = \{[a, b] : a \leq b, a, b \in \mathbb{R}\}$ and where $l([c, d]) = d - c$.

Back to $\mathbb{P}$

- Why this detour to set theory?
- Recall that Probability measure $\mathbb{P}$ acts on sets and measures the probability of such sets.
- In set theory 101 we looked at operations on sets A and B that gave new sets like $A \cup B,\ A \backslash B,\ A \times B, \mathcal{P}(A)$.
- So given $\mathbb{P}(A)$ and $\mathbb{P}(B)$, can we deduce $\mathbb{P}(A \cup B)$ or $\mathbb{P}(A/B)$?
- We want to understand how the probability measure $\mathbb{P}$ acts on sets such as $A \cup B,\ A \backslash B,\ A \times B$.

$\mathbb{P}$ axioms

Probability measure $\mathbb{P}$ is a **set function**.

Axiom 1: $\mathbb{P}(\emptyset) = 0, \mathbb{P}(\Omega) = 1$
Axiom 2: For a set $A \subseteq \Omega$ we have $0 \leq \mathbb{P}(A) \leq 1$.
Axiom 3: For a disjoint collection of events $A_1, A_2, \ldots$ (where $A_i \subseteq \Omega$)

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

- What is in general the domain of $\mathbb{P}$? Ω?
- $\mathcal{P}(\Omega)$? Recall $\mathcal{P}(\Omega) = \{A : A \subseteq \Omega\}$. Seems like a great choice!

Towards a formal definition of $\mathbb{P}$

Probability measure $\mathbb{P}$ can be defined as a set-function $\mathbb{P} : \mathcal{P}(\Omega) \to [0, 1]$ that satisfies the following 3 axioms.
Axiom 1: $\mathbb{P}(\emptyset) = 0, \mathbb{P}(\Omega) = 1$
Axiom 2: For a set $A \subseteq \Omega$ we have $0 \leq \mathbb{P}(A) \leq 1$.
Axiom 3: For a disjoint collection of events $A_1, A_2, \ldots$ (where $A_i \subseteq \Omega$)

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

- Is there a perceivable problem with this definition?
- The following counter-example will construct a set-function $\mathbb{P}$ for which you cannot assign valid probabilities to every subsets in Ω without violating these axioms.

Counter-example

- Random exp: Pick a number uniformly from the real line.
- $\Omega = \mathbb{R}$ and hence $\mathbb{P}(\mathbb{R}) = 1$.
- What is $\mathcal{P}(\mathbb{R})$? Collection of subsets of $\mathbb{R}$.
- $\mathbb{Z}, \mathbb{Q}, \mathbb{R}_+$, finite sets like $\{\pi\}, 0, \{1, 2, 4.5, 6, 10\}$ are all included in $\mathcal{P}(\mathbb{R})$. What else ?
- Sets of the form $[a, b], (a, b], [a, b), (a, b)$ for any $a \leq b$ are also part of the collection $\mathcal{P}(\mathbb{R})$.what else?
- Sets of the form $A_1 \cup A_2 \cup A_3 \cup \dots$ or $A_1 \cap A_2 \cap A_3 \cap \dots$ where A_i could be any of the sets described above.
- $\mathcal{P}(\mathbb{R})$ is unimaginably complex!

Counter-example

- Random exp: Pick a number uniformly from the real line.
- $\Omega = \mathbb{R}$ and hence $\mathbb{P}(\mathbb{R}) = 1$.
- We have $\mathbb{P} : \mathcal{P}(\mathbb{R}) \to [0,1]$.
- $\mathbb{P}(A)$ denotes the probability that the randomly picked number belongs to set A and has the property that sets of equal 'size' have equal probability.
- We know that $\mathbb{R} = \bigcup_{n=-\infty}^{\infty} [n, n+1)$ where $[n, n+1) \in \mathcal{P}(\mathbb{R})$.
- What is $\mathbb{P}[n, n+1)$?
- If we define $\mathbb{P}[n, n+1) = x$ for all $n \in \mathbb{Z}$ then $\mathbb{P}(\mathbb{R}) = \infty$!
- If we define $\mathbb{P}[n, n+1) = 0$ for all $n \in \mathbb{Z}$ then $\mathbb{P}(\mathbb{R}) = 0$!

Counter-example

- What is the takeaway from the counterexample?
- Not all set-functions (or measures) can be calibrated to measure every possible subset of your sample space.
- This is like you weighing scale at home, that is not able to weigh a piece of paper!
- There are much more complicated complications like this (see Vitali set for an example of a non-measurable set).
- What is the way out?
- Restrict your domain to only measurable sets.
- Possible domain for the counter example?
- $\mathcal{F} = \{\Phi, \Omega, \mathbb{R}_{-}, \mathbb{R}_{+}\}$. There can be many other domains one can define!

Towards sigma-algebra

- Restrict your domain of $\mathbb{P}$ to only measurable sets.
- $\mathcal{F} = \{\emptyset, \Omega, \mathbb{R}_-, \mathbb{R}_+\}$.
- The domain $\mathcal{F}$ has some nice and obvious properties.
- For example, if $B \in \mathcal{F}$, then $B^c \in \mathcal{F}$. Also, $\emptyset$ and Ω in $\mathcal{F}$.
- A domain with such nice properties is called as a *sigma-algebra*.

sigma-algebra as domain for $\mathbb{P}$

- Event space or *sigma-algebra* $\mathcal{F}$ is a collection of measurable sets that satisfy
  $\bullet\ \emptyset \in \mathcal{F} \quad \bullet\ A \in \mathcal{F} \implies A^c \in \mathcal{F}$
  $\bullet A_1, A_2, \ldots A_n, \ldots \in \mathcal{F} \implies \cup_{n=1}^{\infty} A_n \in \mathcal{F}$

- The $\sigma-$algebra is said to be closed under formation of compliments and countable unions.

- Is it also closed under the formation of countable intersections?

- When Ω is countable and finite, is $\mathcal{P}(\Omega)$ a sigma-algebra? Yes.

When Ω is countable and finite, we will consider power-set $\mathcal{P}(\Omega)$ as the domain.

Formal definition of Probability measure $\mathbb{P}$

Definition

A probability measure $\mathbb{P}$ on the *measurable space* $(\Omega, \mathcal{F})$ is a function $\mathbb{P} : \mathcal{F} \to [0, 1]$ s.t.

1. $\mathbb{P}(\emptyset) = 0, \quad \mathbb{P}(\Omega) = 1$
2. For a disjoint collection of event sets $A_1, A_2, \ldots$ from $\mathcal{F}$ we have
$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$$
(countable additivity)

- The trio $(\Omega, \mathcal{F}, \mathbb{P})$ is called as a probability space.
- Recall that when $|\Omega| < \infty$, we consider $\mathcal{F} = 2^{\Omega}$.

Formal definition of Probability measure $\mathbb{P}$

Definition

A probability measure $\mathbb{P}$ on the *measurable space* $(\Omega, \mathcal{F})$ is a function $\mathbb{P} : \mathcal{F} \to [0, 1]$ s.t.

1. $\mathbb{P}(\emptyset) = 0, \quad \mathbb{P}(\Omega) = 1$
2. For a disjoint collection of event sets $A_1, A_2, \ldots$ from $\mathcal{F}$ we have
$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$$
(countable additivity)

- The trio $(\Omega, \mathcal{F}, \mathbb{P})$ is called as a probability space.
- Identify the probability space in the coin and dice experiment.

Probability space for $U[0,1]$

- $\Omega = [0,1]$.
- Suppose $\mathcal{F} = \{\Phi, [0,1], [0,.5), [.5,1]\}$. Is there a problem in using this as a sigma-algebra?
- We cannot measure probability of sets like $[.25,.75]$ although we know $P([.25,.75]) = .5$.
- So lets include $[.25,.75]$ in $\mathcal{F}$.
- Now we have $\mathcal{F}^+ = \{\Phi, [0,1], [0,.5), [.5,1], [.25,.75]\}$. Is $\mathcal{F}^+$ a sigma-algebra? No.
- Can you make it a sigma-algebra by adding missing pieces ?

Probability space for $U[0,1]$

- $\mathcal{F}^+ = \{\Phi, [0,1], [0,.5), [.5,1], [.25,.75]\}$
- Can you make it a sigma-algebra by adding missing pieces ?
- Recall that sigma-algebras are closed under compliments, union and intersection.
- Intersection and union of $[.25,.75]$ with sets in $\mathcal{F}^+$ gives the collection $\{[.25,.5), [.5,.75], [.25,1], [0,0.75]\}$.
- Adding compliments, the collection enlarges by $\{[.25,.5), [.5,.75], [.25,1], [0,0.75], [0,.25) \cup [.5,1], [0,.5) \cup (.75,1], [0,.25), (0.75,1]\}$.
- Lets call it $\mathcal{F}^{++} =$ $\{\Phi, [0,1], [0,.5), [.5,1], [.25,.75], [.25,.5), [.5,.75], [.25,1], [0,0.75], [0,.25) \cup [.5,1], [0,.5) \cup (.75,1], [0,.25), (0.75,1]\}$

Probability space for $U[0,1]$

- $\mathcal{F}^{++} =$
  $\{\Phi, [0,1], [0,.5), [.5,1], [.25,.75], [.25,.5), [.5,.75], [.25,1],$
  $[0,0.75], [0,.25) \cup [.5,1], [0,.5) \cup (.75,1], [0,.25), (0.75,1]\}$
- Notice different type of sets with different brackets $[), (], ()$ that appear.
- But $\mathcal{F}^{++}$ is still not a sigma-algebra as each red set will demand a furthermore sets to be added.
- This operation we attempted is called generating a sigma-algebra!.
- Continuing on these lines, the resulting sigma algebra is called a borel-sigma algebra $\mathcal{B}[0,1]$.

Borel sigma-algebra $\mathcal{B}[0,1]$

- Borel $\sigma-$*algebra* $\mathcal{B}[0,1]$: When $\Omega = [0,1]$ the $\mathcal{B}[0,1]$ is the $\sigma-$*algebra* generated by closed sets of the form $[a,b]$ where $a \leq b$ and $a, b \in [0,1]$.

- Does this set contain sets of the form (a,b) or $[a,b)$ or $(a,b]$?

- $\boxed{(a,b) = \cup_{n=1}^{\infty}[a+\frac{1}{n}, b-\frac{1}{n}].}\boxed{(a,b] = \cap_{n=1}^{\infty}(a, b+\frac{1}{n})}$

> Borel $\sigma-$*algebra* $\mathcal{B}[0,1]$: $\mathcal{B}[0,1]$ is the $\sigma-$*algebra* generated by sets of the form $[a,b]$ or (a,b) or $(a,b]$ or even $[a,b)$ where $a \leq b$ and $a, b \in [0,1]$.

Borel sigma-algebra $\mathcal{B}(\mathbb{R})$

- Borel sigma-algebra $\mathcal{B}(\mathbb{R})$:
  If $\Omega = \mathbb{R}$, then $\mathcal{B}(\mathbb{R})$ is the sigma-algebra generated by open sets of the form (a, b) where $a \leq b$ and $a, b \in \mathbb{R}$.

- $\mathcal{B}(\mathbb{R})$ contains intervals of the form

$$
\begin{gathered}
[a, b] \\
[a, b) \\
(a, \infty) \\
[a, \infty) \\
(-\infty, b] \\
(-\infty, b) \\
\{a\}
\end{gathered}
$$

- How would you define $\mathcal{B}(\mathbb{R}^2)$?