MA 6.101
Probability and Statistics

Tejas Bodas

Assistant Professor, IIIT Hyderabad

Logistics

- Second half of the course by Prof. Pawan Kumar.
- Feel free to contact me anytime at tejas.bodas@iiit.ac.in.
- Office @ A5304.
- TA list: Around 9 TAs, you will meet them during tutorials
- Lectures: Monday and Thursday 10:00 to 11:25
- Tutorial on Friday 14:00 to 15:30.
- 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. Intro. to Probability and Statistics for Engineers and Scientists by Sheldon Ross (Elsevier)
  2. Probability and Statistics by DeGroot and Schervish (Addison-Wesley)
  3. Introduction to probability by Bertsekas and Tsisiklis (Athena Scientific)
  4. 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 : 15%.
- Midsem exam: 30%.
- Quiz 2: 15%
- Endsem 40 %.

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
- Forecasting (weather, demand)
- Analysis of Computer systems (routing, scheduling)
- 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 $A_1, A_2, \dots$ (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 $A_1 \cup A_2 \cup A_3 \cup \ldots$ or $A_1 \cap A_2 \cap A_3 \cap \ldots$ 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.
- New domain for the counter example?
- $\mathcal{F} = \{\Phi, \Omega, \mathbb{R}_-, \mathbb{R}_+\}$. This is all that this $\mathbb{P}$ can measure!

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
  $\bullet\ \emptyset \in \mathcal{F} \quad \bullet\ A \in \mathcal{F} \implies A^c \in \mathcal{F}$
  $\bullet A_1, A_2, \ldots A_n, \ldots \in \Omega \implies \cup_{n=1}^{\infty} A_n \in \Omega$

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

More on sigma algebra

- Intersection of $\sigma-$*algebra*'s is a $\sigma-$*algebra*.

- For any set $S \subseteq \Omega$, the smallest $\sigma-$*algebra* containing S is $\{\emptyset, \Omega, S, S^c\}$

- For a collection of sets $\mathcal{C} \subset 2^{\Omega}$, the $\sigma-$*algebra* generated by $\mathcal{C}$, denoted by $\sigma(\mathcal{C})$ is the smallest $\sigma-$*algebra* containing $\mathcal{C}$. Here $\sigma(\mathcal{C}) = \{\cap \mathcal{F} : \mathcal{C} \subseteq \mathcal{F}\}$.

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 $a \le b$ 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 $a \le b$ 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 $\mathcal{B}(\mathbb{R}^2)$?

Consequences of the Probability Axioms

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.
- Identified the probability space in the coin, dice and experiment.

Consequences of the Probability Axioms

- $P(A^c) = 1 - P(A)$
- $P(A \cup B) = p(A) + P(B) - P(A \cap B)$.
- If $A \subseteq B$, prove that $P(A) \leq P(B)$. ($A \subseteq B$ has the interpretation that Event A implies event B)
- $P(\cup_{i=1}^{\infty} B_i) \leq \sum_{i=1}^{\infty} P(B_i)$ (Boole's/Bonferroni's inequality). HW
- What is $P(A \cup B \cup C)$?
- State and prove the inclusion-exclusion principle for $P(\cup_{i=1}^{n} A_i)$

Impossible event v/s Zero prob. event

- In $U[0,1]$ what is $P(\omega = 0.5)$? $= 0$.
- Intuitive reasoning for this is that a point has zero length!
- If $P(\omega \in [a,b]) = b - a$ then $P([.5,.5]) = P(\{.5\}) = 0$.
- This is a zero probability event. In fact, every outcome of this experiment is a zero probability event.
- This implies that events of zero probability can happen and they are not impossible events.
- $P(\emptyset) = 0$, then is $\emptyset$ also possible ?No!
- What is $P(\omega \in [0,.25] \cap [.75,1])$?
- $P([0,.25] \cap [.75,1]) = P(\emptyset) = 0$ This event will never happen.

Impossible event v/s Zero prob. event

- Note that in the $U[0,1]$ experiment, $\Omega = \bigcup_{\omega \in \Omega} \{\omega\}$
- $P(\Omega) = P(\bigcup_{\omega \in \Omega} \{\omega\}) = \sum_{\omega \in \Omega} P(\{\omega\}) = 0.$
- What is the problem above ?
- Ω is an uncountable set and the probability set-function only has a countable additive property.
- $\bigcup_{\omega \in \Omega} \{\omega\}$ is an uncountable disjoint union!

Recap

- Set theory 101
- $\mathbb{P}$ is a set-function.
- All sets in $\mathcal{P}(\Omega)$ need not be measurable.
- We restrict domain of $\mathbb{P}$ to sigma-algebra of measurable sets.
- $(\Omega, \mathcal{F}, \mathbb{P})$ is known as probability space.
- Formal definition of probability measure with its axioms.
- We looked at $\mathcal{B}([0,1])$ and $\mathcal{B}(\mathbb{R})$.

This class ...

- Continuity of set-function $\mathbb{P}$.
- Conditional Probability.
- Law of Total Probability.
- Independence and Correlation between sets.
- Conditional independence.

Limits and Continuity

- How do we define limit of a sequence $\{a_1, a_2, \ldots, \}$?
- Notation: $\lim_{n\to\infty} a_n = L$.
- How do you define limit of a function at a point c?
- Notation: $\lim_{x\to c} f(x) = L$
- How do you define continuity of a function $f(x)$ at c ?
- When do you say a function is continuous ?
- (ϵ, δ)-definition of limits and continuity?

Limits and Continuity

Definition in terms of limits of sequences.

For a continuous function $f(\cdot)$, as $x \to c$, we have $f(x) \to f(c)$

For a continuous set-function S, as $A_n \to A$, we have $S(A_n) \to S(A)$

- Recall that $\mathbb{P}$ is a set-function. Is it continuous?
- We will see the proof shortly.

Sequence of sets

- Given $(\Omega, \mathcal{F})$, If $A_1 \subset A_2 \dots$ is an increasing sequence of events defined on $\mathcal{F}$ and $\cup_{n=1}^{\infty} A_n = A \in \mathcal{F}$, then we say thats the sequence of sets A_n are increasing to A ($A_n \uparrow A$).

- Similarly when $A_1 \supset A_2 \dots$ is a decreasing sequence of events and $\cap_{n=1}^{\infty} A_n = A$, then we have $A_n \downarrow A$.

- Alternative notation: For an increasing sequence of sets A_n we often write $\lim\limits_{n\to\infty} A_n$ for $\cup_{n=1}^{\infty} A_n$ and for a decreasing sequence of sets A_n that $\lim\limits_{n\to\infty} A_n = \cap_{n=1}^{\infty} A_n$.

Continuity of set-function $\mathbb{P}$

Lemma

For sequence of events of the type $A_n \uparrow A$ or $A_n \downarrow A$, we have $\lim_{n\to\infty} \mathbb{P}(A_n) = \mathbb{P}(A)$.

Proof

- Consider increasing sequence first. Similar arguments follow for decreasing seq.
- Define $F_n = A_n - A_{n-1}$
- $\cup_{n=1}^{\infty} A_n = \cup_{n=1}^{\infty} F_n$.
- $\mathbb{P}(A) = \mathbb{P}(\cup_{n=1}^{\infty} A_n) = \mathbb{P}(\cup_{n=1}^{\infty} F_n)$
- But $\mathbb{P}(\cup_{n=1}^{\infty} F_n) = \lim_{n\to\infty} \sum_{i=1}^{n} P(F_i) = \lim_{n\to\infty} \mathbb{P}(A_n)$. $\square$

Equivalently if $An \to \emptyset$, then $\mathbb{P}(A_n) \to 0$.

Conditional probability

- Given/If dice rolls odd, what is the probability that the outcome is 1?
- Given/If $\bar{\omega} \in [0, 0.5]$ what is the probability that $\bar{\omega} \in [0, 0.25]$?
- The conditional probability of event B given event A is defined as $\mathbb{P}(B/A) := \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)}$ when $\mathbb{P}(A) > 0$.

When conditioned on event $A \in \mathcal{F}$, you focus on a random experiment with probability space $(\Omega_{/A}, \mathcal{F}_{/A}, \mathbb{P}_{/A})$ where

- $\Omega_{/A} = A$,
- $\mathcal{F}_{/A} = \{C \cap A : C \in \mathcal{F}\}$ and
- $\mathbb{P}_{/A}(B) := \mathbb{P}(B/A) = \frac{\mathbb{P}(B \cap A)}{\mathbb{P}(A)}$.

Conditional probability

- Show that $P(A/B)P(B) = P(B/A)P(A)$.
- Bayes rule: $P(B/A) = \frac{P(A/B)P(B)}{P(A)}$.
- What is $P(A/(B \cap C))$?.This is also denoted as $P(A/BC)$
- Prove the chain rule
  $P(A \cap B \cap C) = P(A)P(B/A)P(C/(AB))$.

HW: Prove the chain rule for conditional probability given by

$P(A_1 \cap A_2 \dots A_n) = P(A_1)P(A_2/A_1)P(A_3/A_1A_2)\dots P(A_n/A_{n-1}\dots A_1)$.

Conditional probability – Examples

- Suppose you draw 4 cards from a deck at random without replacement. What is the probability that (in order) these cards are 9 of club, 8 of diamond, king of spade and king of club?

- What if you do the above with replacement?

- Consider a finite sample space Ω where each outcome is equally likely. Then what is $P(B/A)$?

- $P(B/A) = \frac{|A \cap B|}{|A|}$.

Law of total probability

- $A = (A \cap B) \cup (A \cap B^c)$. What is $P(A)$?
- $P(A) = P(A \cap B) + P(A \cap B^c)$.
- This is same as $P(A) = P(A/B)P(B) + P(A/B^c)P(B^c)$.
- This formula is useful when $P(A)$ is not given or is difficult to find but $P(B)$ or $P(A/B)$ is readily available.

Let $B_1, B_2, \ldots B_n$ be the partition of the sample space Ω. Then for any event A we have

$$P(A) = \sum_{i=1}^{n} P(A \cap B_i) = \sum_{i=1}^{n} P(A/B_i)P(B_i).$$

Example

1. If an item is defective, a robot can spot it with 98% accuracy.
2. If an item is not defective, a robot will declare it so with 99% accuracy.
3. A total of 0.1% items are defective.
4. If the robot says that the item you drew at random is defective, what is the probability that the robot is correct?

- $P(\text{defective}/\text{robot says defective}) = \frac{P(\text{robot says defective}/\text{defective})P(\text{defective})}{P(\text{robot says defective})}$

- What is $P(\text{robot says defective})$?

Recap

- Continuity of set-function $\mathbb{P}$.
- Conditional Probability.
- Law of Total Probability.

This class

- Independence and Correlation between sets.
- Conditional independence.
- Principles of counting.

Bayes rule revisited

Let $B_1, B_2, \ldots B_n$ be the partition of the sample space Ω. Then for any event A with $P(A) > 0$ we have

$$P(B_j/A) = \frac{P(A/B_j)P(B_j)}{\sum_{i=1}^{n} P(A/B_i)P(B_i)}.$$

Independence

- Consider the experiment of tossing a coin and a dice simultaneously.
- Identify its underlying probability space.
- What is $\mathbb{P}(\{H, 6\})$?
- What is $\mathbb{P}(\{T, \text{ odd }\}) = \mathbb{P}(\{\cup_{i=1,3,5}\{T, i\}\})$?
- In both cases above we have $\mathbb{P}(A \cap B) = P(A)P(B)$.
- This implies that $\mathbb{P}(A/B) = \mathbb{P}(A)$.

> - Two events A, B are independent iff $P(A/B) = P(A)$ and $P(B/A) = P(B)$.
> - Two events A, B are independent iff $P(A \cap B) = P(A)P(B)$.

Independence

- Two events A, B are independent if and only if $P(A \cap B) = P(A)P(B)$.

- If A and B are independent, then so are A^c and B^c.
- What about A and B^c? Are they independent?
- If $A_1, A_2, \ldots, A_n$ are independent, then prove that

$$P(\cup_{i=1}^n A_i) = 1 - \prod_{i=1}^n [1 - P(A_i)]$$

Mutual and Pairwise Independence

- A collection of events $\{A_i, i \in I\}$ are said to be **mutually independent** if the $P(\cap_{j\in J} A_j) = \prod_{j\in J} P(A_j)$ for any subset J of I.
- A collection of events $\{A_i, i \in I\}$ are said to be **pairwise independent** if any pair of events from the collection are independent.
- Mutual independence implies pairwise independence but not the other way around.
- HW: Find an example where pairwise independence does not imply mutual independence.

Independence - Example

- Pick a number randomly from the set $\{1, \ldots, 10\}$.
- Event A says that the number is less than 7.
- Event B says that the number is less than 8.
- Event C says that the number is even.
- Are the events mutually independent?
- Which pair of event is independent?

Correlation between events

- Two events A, B are independent iff $P(A \cap B) = P(A)P(B)$.
- Two events A and B are positively correlated iff $P(A/B) > P(A)$.
- Two events A and B are negatively correlated iff $P(A/B) < P(A)$.
- A and B have the same correlation as A^c and B^c.
- A and B have the opposite correlation as A and B^c.

Mutually exclusive and Independence

- Two events A and B are mutually exclusive if occurrence of one implies that the other event cannot occur. Are they independent?

- If A and B are mutually exclusive, then they are not independent (and vice versa). This can be seen as follows.

A and B are Mutually Exclusive

- $P(A \cap B) = 0$
- $P(A/B) = 0$
- $P(A/B^c) = \frac{P(A \cap B^c)}{P(B^c)} = \frac{P(A)}{P(B^c)}$

A and B are Independent

- $P(A \cap B) = P(A)P(B)$
- $P(A/B) = P(A)$
- $P(A/B^c) = P(A)$

- If $A \subseteq B$, then two events are neither mutually exclusive nor independent.

Zero probability events and Independence

Zero probability events are always independent!

- Let E be a zero probability event, i.e. $P(E) = 0$.
- Then for any set F, we want to show that $P(E \cap F) = 0$.
- Note that $E \cap F \subseteq E$.
- This implies that $P(E \cap F) \leq P(E)$.

Conditional independence

- Recall : $P(A/B) = \frac{P(AB)}{P(B)}$.
- Also recall : $P(A/BC) = \frac{P(AB/C)}{P(B/C)}$
- This implies $P(AB/C) = P(A/BC)P(B/C)$.

> Two events A and B are said to be conditionally independent of event C ($P(C) > 0$) if $P((AB)/C) = P(A/C).P(B/C)$

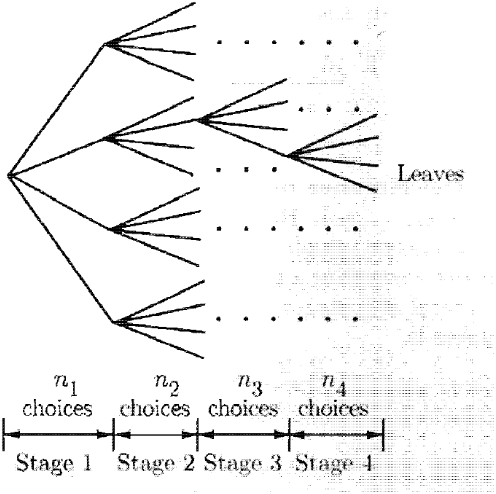
- As a consequence $P(A/BC) = P(A/C)$

HW: Verify if events A and B are conditionally independent of event C (in the experiment of picking number randomly in $\{1, .., 10\}$)

Conditional independence – example

- There are two coins, one fair and other fake (both heads). The experiment is to choose a coin uniformly and toss twice.
- Event A: First coin toss results in H. What is its probability? P(A) $= 3/4$.
- Event B: Second coin toss results in H. What is its probability? P(B) $= 3/4$.
- Event C: Coin 1 is chosen.
- What is $P(A/C)$ and $P(B/C)$? $1/2$
- What is $P((A \cap B)/C)$? $1/4$ Hence A and B are conditionally independent given C.
- Are A and B independent? HW

First principle of counting



Principle of counting

- Given n objects, in how many ways can you arrange them? n!
- Given n objects, how many distinct pairs can you form? ${}^{n}C_2 = \binom{n}{2} = \frac{n!}{n-2!2!}$.
- In general, given n objects, we can make ${}^{n}C_k = \binom{n}{k} = \frac{n!}{n-k!k!}$ distinct combination of k objects.
- Note that in each combination or group of k objects, the ordering within each group is immaterial. What if we also want to count this?
- ${}^{n}P_k = {}^{n}C_k \times k!$

Experiments with Sampling

- Sampling: Sampling from a set means choosing an element from the set.
- Sampling uniformly at random: All items in the set have equal probability of being chosen.
- Sampling can be with replacement or without replacement.
- Sampling can be ordered or unordered.
- In ordered sampling, $(a, b, c) \neq (c, b, a)$.
- This leaves us with 4 combinations.
  1. Ordered sampling with replacement
  2. Ordered sampling without replacement
  3. Unordered sampling with replacement
  4. Unordered sampling without replacement

Ordered sampling with replacement

- Suppose you want to sample k out of n objects with replacement and where the ordering of the k objects matters.
- Because we sample with replacement, repetition is allowed.
- How many ways can you choose k objects out of n this way?
- a) nk? b) $\binom{n}{k}$ c) k^n d) n^k ?
- There are k positions and n choices for every position.
- Total n^k.

Ordered sampling without replacement

- Suppose you want to sample k out of n objects now without replacement and where the ordering of the k objects matters.
- Because we sample without replacement, repetition is not allowed.
- How many ways can you choose k objects out of n this way?
- a) nk? b) $\binom{n}{k}$ c) k^n d) none ?
- There are k positions and $n - i + 1$ choices for every i^{th} position.
- Total $n \times (n-1) \times \ldots (n-k+1) = \frac{n!}{(n-k)!} = {}^nP_k$.

Unordered sampling without replacement

- Here you want to sample k out of n objects without replacement and the ordering of the k objects does not matters.
- Because we sample without replacement, repetition is not allowed.
- How many ways can you choose k objects out of n this way?
- a) nk? b) $\binom{n}{k}$ c) k^n d) none ?
- Essentially we want to count distinct k sized subsets from n objects without caring for ordering.
- ${}^nC_k$.

Unordered sampling with replacement

- Here you want to sample k out of n objects with replacement and the ordering of the k objects does not matters.
- Because we sample with replacement, repetition is allowed.
- How many ways can you choose k objects out of n this way?
- In any such sampling, any object i can appear atmost k times.
- Let x_i denote the number of times object i is chosen in k samples.
- Then any sampling satisfies $\displaystyle\sum_{i=1}^{n} x_i = k$
- How many solutions to the above equation tells you how many ways you can do the above sampling.
- $\binom{n+k-1}{k}$ Think(HW).