Probability Bootcamp

2. Probability

References

1. Chapter 1, Mathematical Statistics and Data Analysis (3rd edition), John A. Rice
2. Chapter 2, Discrete Probability, Hugh Gordon

Introduction

- The idea of probability and randomness is quite old - many day to day tasks revolve around randomness and chance (eg weather)
- The rigorous axiomatization of probabilistic intuition only occurred relatively recently.
    - By developing a set of rigorous axioms (rules) and proofs, we are able to extend this intuition from simple random events to more complex events
- Some fields that rely on rigorous probability include finance, genetics, physics, economics, etc.
- This portion of the bootcamp introduces some of these axioms, proofs and common tools used in probability

Agenda

1. Sample Spaces
2. Probability Measures
3. Counting Methods
4. Conditional Probability
5. Independence

Sample Spaces

Sample Spaces

- Probability theory deals with situations where outcomes occur randomly
  - Most literature (and this bootcamp) refers to such situations as experiments
- The sample space refers to the set of all possible outcomes in an experiment
  - It is usually denoted by Ω
- Examples

Experiment	Ω
Flipping a coin	{H, T}
Flipping 2 coins	{HH, HT, TH, TT}
Number of heads when flipping N coins	{0, 1, …, N}
Earth population in 2500	{0, 1, …}
Number of seconds until the next earthquake	$\{t \mid t \geq 0\}$

Events

- We are often interested in subsets of Ω, which are called events in probability
- Example. Consider the experiment where we flip 3 coins
  - Ω = {HHH, HHT, HTH, HTT, THH, THT, TTH, TTT}
  - The following are some of the events of this experiment

Event	Description	Set
Ω	Sample Space	{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT}
A	First coin lands on Head	{HHH, HHT, HTH, HTT}
B	At least one Tail	{HHT, HTH, HTT, THH, THT, TTH, TTT}
C	Exactly one Head and One Tail	{} (empty set, also written as $\varnothing$}

Events

- All set operations (union, intersection, complement, etc) apply the same way here

Event	Description	Set
Ω	Sample Space	{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT}
A	First coin lands on Head	{HHH, HHT, HTH, HTT}
B	At least one Tail	{HHT, HTH, HTT, THH, THT, TTH, TTT}
C	Exactly one Head and One Tail	{} (empty set, also written as ∅}

- For example

Event	Set
$A \cap B$	{HHT, HTH, HTT}
$C \cup \{TTT\}$	{TTT}
B^C	{HHH}

Probability Measures

Probability Measure

- Given a sample space Ω, a probability measure P is a function that maps any event $A \subseteq \Omega$ to a value between [0, 1]. In addition, it must satisfy the following

	Condition	Why?
1	$P(\Omega) = 1$	The sample space represents the set of all possible outcomes. One of the outcomes must happen at the end of the experiment
2	For any $A \subseteq \Omega$, $P(A) \geq 0$	A probability must be non-negative
3	For any $A_1, A_2 \subseteq \Omega$ such that A_1, A_2 are disjoint (i.e. $A_1 \cap A_2 = \varnothing$), the following must hold: $P(A_1 \cup A_2)$ $= P(A_1) + P(A_2)$	If two outcomes have nothing in common, then they shouldn't affect each other. Hence, the probability of either one happening should be the sum of both probabilities

Probability Measures

- Intuitively, a probability measure is a function that maps events to a common scale from 0 to 1. This value allows us to quantitatively state and compare the likelihood of all possible outcomes.
- Example of a probability measure - Counting measure
  - Given a sample space $\mathbf{\Omega}$, an event A and the counting measure P,
  - $|A \cap \mathbf{\Omega}|$ is the number of outcomes in A that appears in the sample space
  - $|\mathbf{\Omega}|$ is the size of the sample space
  - Counting measure assumes all outcomes in the sample space are equally likely

$$P(A) = \frac{|A \cap \Omega|}{|\Omega|}$$

- Consider tossing two fair coins
  - $\mathbf{\Omega}$ = {HH, HT, TH, TT}
  - The counting measure can be used here as all outcomes in $\mathbf{\Omega}$ are equally likely
  - Define event A = {TH, HT, TT} (throwing at least 1 tail)
  - Using counting measure P, P(A) = 3/4

Probability Measures

- As an exercise, you can verify that the counting measure fulfills all the properties in the previous slide.
- Many other examples of measures. Basically, any probability distribution function (Normal, Poisson, Exponential, Binomial, etc) are all probability measures
- Since probability measures often require knowledge of the size of the event set, this leads to the next section - counting methods

Counting Methods

Counting Methods

- In the preceding examples, it was easy to count the number of outcomes in various events. For example,
  - When tossing 3 coins, count the number of outcomes where head occurred at least once
  - When rolling 2 dices, count the number of outcomes where both dices rolled an even number
- For more complicated experiments, we will require a more systematic method of computing such outcomes

Multiplication Principle

- If two experiments have m and n possible outcomes respectively, then there are m*n total possible outcomes for the 2 experiments
  - Example - If I toss a coin and roll a dice, there are 2*6 possible outcomes
- Extended multiplication principle - if k experiments have $n_1, n_2, \ldots, n_k$ possible outcomes respectively, then they collectively have $n_1 * n_2 * \ldots * n_k$ outcomes
  - Intuitively, same reasoning as the multiplication principle with 2 outcomes
  - Can formally prove this by induction (self exercise)
- Example - suppose you know that a password consists of
  - 8 alphanumeric lowercase characters,
  - First and last characters are numbers (digits)
  - Third character is an alphabet
  - There are $10 * (26 + 10) * 26 * (26 + 10)^4 * 10$ possible combinations

Permutations

- A permutation on a set is an ordered arrangement on a set
- A k-permutation on a set is an ordered arrangement of a subset of size k
- There are two types of k-permutations
  - Sampling with replacement
  - Sampling without replacement
- We can think of the two types as taking numbered balls from an urn. Suppose there are n balls in the urn and r balls are drawn
  - In sampling with replacement, after you select a ball, you return it back to the urn
    - It is possible to select multiple balls with the same number
    - There are n^r possible selections
  - In sampling without replacement, after you select a ball, it is discarded
    - All selections are unique
    - There are n * (n-1) * … * (n-r+1) possible selections
  - For both cases, the number of possible selections can be derived using the (extended) multiplication principle

Permutations

- Examples
    - How many different ways can 5 children be lined up?
        - 5!
    - How many different ways can 5 children be lined up in a queue of length 3?
        - 5 * 4 * 3
    - How many ways can 100 people be selected to win 4 distinct prizes?
        - 100 * 99 * 98 * 97

Combinations

- When dealing with permutations, ordering matters.
  - For example, when permuting a selection of 2 items from {1,2,3}, (1,2) and (2,1) are treated as 2 separate permutations
- When dealing with combinations, ordering does not matter
  - Example - The following combinations of 2 items can be drawn from {1,2,3} - (1,2), (1,3), (2,3)
- Combinations are more useful when only the objects matter, but not which order they appear
  - Example: In a lottery where you pick 6 numbers without replacement from 58 numbers, it does not matter which order you pick the 6 numbers
- There are $\binom{n}{r}$ ways of selecting r items without replacement from a set of n items when ordering does not matter, where

$$\binom{n}{r} = \frac{n!}{r!\,(n-r)!}$$

Intuition behind $\binom{n}{r} = \dfrac{n!}{r!\,(n-r)!}$

- Note that $\dfrac{n!}{(n-r)!}$ is actually the number of ways to permute r items from a set of n items
    - $\dfrac{n!}{(n-r)!} = \dfrac{n \cdot (n-1) \cdot \ldots \cdot (n-r+1) \cdot (n-r)!}{(n-r)!} = n \cdot (n-1) \cdot \ldots \cdot (n-r+1)$

- Given a set of r items, there are r! ways of permuting this set
- Note the following relationship
    - Number of ways to permute r items from a set of n items = $\dfrac{n!}{(n-r)!} = r! \cdot \binom{n}{r}$

Combinations

- Examples
  - In a lottery, a player chooses 6 numbers from a total of 58 numbers
    - There are $\binom{58}{6} = 40,475,358$ possible combinations
  - When performing quality control, only a fraction of the output of a manufacturing process is randomly sampled and examined for defects. Suppose in a batch of n items, r items are randomly chosen for selection. Also suppose that k defective items are manufactured. What is the probability the sample contains at least one defective item?
    - It is easier to compute the probability the sample does not contain any defective item. There are
      - $\binom{n}{r}$ Ways of choosing r items from the batch of n items
      - $\binom{n-k}{r}$ Ways of choosing r items from the batch of non-defective items
    - The probability of not choosing any defective items when sampling is therefore

$$\frac{No.\ of\, ways\, of\, not\ sampling\, any\, defectives}{No.\ of\, ways\, of\ sampling} = \frac{\binom{n-k}{r}}{\binom{n}{r}}$$

Conditional Probability

Conditional Probability

- Our judgement about the likelihood of an outcome may change when we have more information
  - For example, the likelihood of someone having Covid-19 is much higher if he has a positive ART test, as compared to someone who hasn’t taken the test
- Often, data comes broken down in groups. Example below, Covid-19 deaths by age group and vaccination status

Age 18–49			50–64			65+		
	Vaccine			Vaccine			Vaccine	
Death	No	Yes	Death	No	Yes	Death	No	Yes
Yes	155	7	Yes	290	23	Yes	561	158
No	2,666	1,523	No	1,755	2,447	No	1,668	7,132

Aggregate Tables

- Death by age group

	Death		Total	Death rate (no of deaths / total)
Age	Yes	No		
18-49	162	4,189	4,351	3.7%
50-64	313	4,202	4,515	6.9%
65+	719	8,800	9,519	7.6%
Total	1,194	17,191	18,385	6.5%

- If we do not know the patient’s age and vaccination status, we can only estimate his probability of death to be 6.5%
- Given that the patient is in the 18-49 age group, his risk of death is much lower at 3.7%

Conditional Probabilities (computation)

- The following data is from a clinical trial[1].
  - The digitalis medicine is sometimes administered to patients with heart failure, but some patients may develop toxicity towards digitalis.
  - A test to detect digitalis toxicity was developed, its accuracy is given below

$T+$ = high blood concentration (positive test)
$T-$ = low blood concentration (negative test)
$D+$ = toxicity (disease present)
$D-$ = no toxicity (disease absent)

	$D+$	$D-$	*Total*
$T+$	25	14	39
$T-$	18	78	96
Total	43	92	135

- Assuming this pattern holds true in larger groups, the probability of digitalis toxicity given positive/negative test results are the following

	$D+$	$D-$	*Total*
$T+$	.185	.104	.289
$T-$	.133	.578	.711
Total	.318	.682	1.000

1. Beller, George A., et al. "Digitalis intoxication: a prospective clinical study with serum level correlations." *New England Journal of Medicine* 284.18 (1971): 989-997.

Conditional Probabilities (computation)

	$D+$	$D-$	*Total*
$T+$	.185	.104	.289
$T-$	.133	.578	.711
Total	.318	.682	1.000

- Denote the probability that a patient shows toxicity given positive test by P(D+|T+)
  - This is known as the conditional probability of D+ given T+.

	$D+$	$D-$	*Total*
$T+$	25	14	39
$T-$	18	78	96
Total	43	92	135

- From the frequency table,
  - P(D+|T+) = 25/39 = 0.64
- From the probability table,
  - P(D+|T+) = 0.185/0.289 = 0.64
- Note that both are equivalent, since all terms in the probability table are just the relative frequency divided by a common denominator
- To formalize this intuition,

$$P(D+ \mid T+) = \frac{P(D+ \ \cap T+)}{P(T+)}$$

Independence

Independence

- Intuitively, two events A and B are independent if knowing A does not tell us anything about B, and vice versa
  - Let A = The colour of Mary’s dress and B = John’s score in IT5001
  - Knowing that Mary wore a red dress doesn’t tell us anything about John’s score, and vice versa. Hence, A and B are independent (written as A $\perp$ B)
- From the intuitive definition, we know the following conditional probability
  - $P(A) = P(A \mid B) := \frac{P(A \cap B)}{P(B)}$
- Therefore, if A and B are independent, we have $P(A \cap B) = P(A)P(B)$

Examples

- Tossing a coin and rolling a dice are independent. Let A = coin comes up heads and B = dice rolls a number greater than 3.
  - $P(A) = \frac{1}{2}$, $P(B) = \frac{1}{2}$
  - $P(A \cap B) = \frac{1}{2} * \frac{1}{2} = \frac{1}{4}$