MITx SDS MicroMasters

Course Notes

6.431x: Probability - The Science of Uncertainty and Data

Author:
Yolanda Huitian Diao

February 26, 2023

Contents

List of Figures

List of Tables

List of Abbreviations

LAH	**L**ist **A**bbreviations **H**ere
WSF	**W**hat (it) **S**tands **F**or

List of Symbols

$\mathbb{R}$	Real numbers
$\mathbb{N}$	Natural numbers
Ω	Universal set
$\emptyset$	Empty set

Chapter 1

Probability models and axioms

1.1 Math Overview

1.1.1 Sets and De Morgan's Laws

Sets

- A collection of distinct element
- Can be finite or infinite

Unions and intersections

- $S \cup T : \quad x \in S \cup T \Leftrightarrow x \in S$ or $x \in T$
- $S \cap T : \quad x \in S \cap T \Leftrightarrow x \in S$ and $x \in T$
- $x \in \bigcup_n S_n \Leftrightarrow x \in S_n$, for some n
- $x \in \bigcap_n S_n \Leftrightarrow x \in S_n$, for all n

Set properties

- $S \cup T = T \cup S$
- $S \cap (T \cup U) = (S \cap T) \cup (S \cap U)$
- $(S^c)^c = S$
- $S \cup \Omega = \Omega$
- $S \cup (T \cup U) = (S \cup T) \cup U$
- $S \cup (T \cap U) = (S \cup T) \cap (S \cup U)$
- $S \cap S^c = \varnothing$
- $S \cap \Omega = S$

De Morgan's laws

- $(\bigcap_n S_n)^c = \bigcup_n S_n^c$
- $(\bigcup_n S_n)^c = \bigcap_n S_n^c$

Sequences and their limits

- **Definition of Sequence**

  - function$f : \mathbb{N} \to S, f(i) = a_i$

- **Convergence of Sequence**
  - $a_i \underset{i\to\infty}{\to} a, \lim\limits_{i\to\infty} a_i = a$
  - For any $\epsilon > 0$, there exists i_0, such that if $i \geq i_0$, then $|a_i - a| < \epsilon$

1.1.2 Sequences and their limits

- If $a_i \geq a_{i+1}$, for all i, then either:
  - the sequence "converges to ∞"
  - the sequence converges to some real number a
- If $|a_i - a| \leq b_i$, for all i, and $b_i \to 0$, then $a_i \to a$
- Properties of convergent sequences
  - If $a_i \to a$ and $b_i \to b$, then
    * $a_i + b_i \to a + b$
    * $a_i b_i \to ab$
  - If $a_i \to a$ and g is a continuous function, then
    * $g(a_i) \to g(a)$

1.1.3 Infinite series

Provided limit exists: $\sum\limits_{i=1}^{\infty} a_i = \lim\limits_{n\to\infty} \sum\limits_{i=1}^{n} a_i$

- If $a_i \geq 0$: limit exists
- If term a_i do not all have the same sign:
  - limit need not exist
  - limit may exist but be different if we sum in a different order
  - **Fact:** limit exists and independent of order of summation if $\sum\limits_{i=1}^{\infty} |a_i| < \infty$

1.1.4 Geometric series

$\sum\limits_{i=0}^{\infty} \alpha^i = 1 + \alpha + \alpha^2 + ... + = \frac{1}{1-\alpha}\ |\alpha| < 1$

1.1.5 Sums with multiple indices

$\sum\limits_{i\geq 1, j\geq 1} a_{ij}$

- If the sum converges, this double series will be well defined.
- If $\sum |a_{ij}| < \infty$, then order of summation does not matter.

1.1.6 Countable and uncountable sets

- Countable: can put in 1-1 correspondence with positive integers
  - positive integers
  - integers
  - pairs of positive integers
  - rational number q, with $0 < q < 1$
    * $1/2, 1/2, 2/3, 1/4, 2/4, 3/4, 1/5, 2/5...$
- Uncountable: not countable
  - the interval $[0, 1]$
  - the reals, the plane, ...
- The reals are uncountable
  - Cantor's diagonalization argument

1.2 Lecture 1: Probability models and axioms

1.2.1 Sample space

- Two steps:
  - Describe possible outcomes
  - Describe beliefs about likelihood of outcomes
- List (set) of possible outcomes, Ω
  - Mutually exclusive
  - Collectively exhaustive
  - At the "right" granularity
- Examples
  - Discrete / finite
    * Two rolls of a tetrahedral die
    * Sequential description (decision tree)
  - Continuous
    * (x, y) such that $0 \leq x, y \leq 1$

1.2.2 Probability laws

- **Event:** a subset of the sample space
  - Probability is assigned to events
- **Axioms:**
  - Nonnegativity: $P(A) \geq 0$
  - Normalization: $P(\Omega) = 1$
  - (Finite) additivity: (to be strengthened later)
    * If $A \cap B = \emptyset$, then $P(A \cup B) = P(A) + P(B)$

1.2.3 Some simple consequences of the axioms

- $P(A) \leq 1$
- $P(\emptyset) = 0$
- $P(A) + P(A^C) = 1$
- $P(A \cup B \cup C) = P(A) + P(B) + P(C)$ and similarly for k disjoint events $P(s_1, s_2, ...s_k) = P(s_1) + ... + P(s_k)$
- $A \cup A^C = \Omega$
- $A \cap A^C = \emptyset$

1.2.4 More consequences of the axioms

- If $A \subset B$, then $P(A) \leq P(B)$
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
- $P(A \cup B) \leq P(A) + P(Bs)$
- $P(A \cup B \cup C) = P(A) + P(A^C \cap B) + P(A^C \cap B^C \cap C)$

Examples

- **Discrete / finite example:** Two rolls of a tetrahedral die
  - $X = Firstroll, Y = Secondroll, Z = min(X, Y)$
  - $P(X = 1) = 4/16 = 1/4$, $P(Z = 2) = 5/16$

1.2.5 Discrete uniform law

- Assume Ω consists of n equally likely elements
- Assume A consists of k elements
- $P(A) = k \cdot \frac{1}{n}$

1.2.6 Uniform probability law

- $Probabiliy = Area$

1.2.7 Probability calculation steps

- Specify the sample space
- Specify a probability law
- Identify an event of interest
- Calculate ...

1.2.8 Countable additivity axiom

- If $A_1, A_2, A_3, ...$ is an infinite sequence of disjoint events,
  - Then $P(A_1 \cup A_2 \cup A_3 \cup ...) = P(A_1) + P(A_2) + P(A_3) + ...$

1.2.9 Interpretations of probabilities

- A narrow view: a branch of math
  - Axioms $\Rightarrow$ theorems
- Are probabilities frequencies?
  - P(coin toss yields heads) $= 1/2$
  - P(the president of ... will be reelected) $= 0.7$
- Probabilities are often interpreted as:
  - Description of beliefs
  - Betting preferences

1.2.10 The role of probability theory

- A framework for analyzing phenomena with uncertain outcomes
  - Rules for consistent reasoning
  - Used for predictions and decisions
- Diagram
  - **Real world** $\Rightarrow$ data $\Rightarrow$ **Inference/Statistics**
  - **Inference/Statistics** $\Rightarrow$ Models $\Rightarrow$ **Probability theory (Analysis)**
  - **Probability theory (Analysis)** $\Rightarrow$ Predictions / Decisions $\Rightarrow$ **Real world**

Chapter 2

Conditioning and independence

2.1 Lecture 2: Conditioning an Baye's rule

The idea of conditioning: Use new information to revise a model

2.1.1 Conditional Probability

The idea of conditioning: Use new information to revise a model

Definition of conditional probability

- $P(A|B) =$ "probability of A, given that B occurred"
- $P(A|B) = \frac{P(A \cap B)}{P(B)}$ defined only when $P(B) > 0$

Two rolls of a 4-sided die

- Let B be the event: $min(X, Y) = 2$. Let $M = max(X, Y)$
  - $P(M = 1|B) = 0$
  - $P(M = 3|B) = \frac{P(M=3andB)}{P(B} = \frac{2/16}{5/16} = 2/5$

Conditional probabilities hsare properties of ordinary probabilities

- $P(A|B) \geq 0$, assuming $P(B) > 0$
- $P(\Omega|B) = \frac{P(\Omega\ capB)}{P(B)} = 1$
- $P(B|B) = 1$
- If $A\ capC = \emptyset$, then $P(A \cup C|B) = P(A|B) + P(C|B)$

2.1.2 Three important tools: Multiplication rule; Total probability theorem; Baye's rule

- Multiplication rule
  - $P(A|B) = \frac{P(A \cap B)}{P(B)}$
  - $P(A \cap B) = P(B)P(A|B) = P(A)P(B|A)$
- Total probability theorem
  - Partition of sample space into $A_1, A_2, A_3, ...$
  - Have $P(A_i)$, for every i

- Have $P(B|A_i)$, for every$i$
- $P(B) = \sum_i P(A_i)P(B|A_i)$

Bayes' rule and inference

- Thomas Bayes, presbyterian minister (c. 1701 - 1761)
- "Bayes' theorem", published pothumously
- systematic approach for incorporatin new evidence
- **Bayesian inference**
  - initial beliefs $P(A_i)$ on possible causes of an observed event B
  - model of the world under each A_i: $P(B|A_i)$
    * $A_i \xrightarrow[\text{model}]{P(B|A_i)} B$
  - draw conclusions about causes
    * $B \xrightarrow[\text{inference}]{P(A_i|B)} A_i$

2.2 Lecture 3: Independence

2.2.1 Independence of two events

- **Intuitive "definition"**: $P(B|A) = P(B)$
  - occurrence of A provides no new information about B
- **Definition of independence**: $P(A \cap B) = P(A) \cdot P(B)$
  - Symmetric with respect to A and B
  - implies $P(A|B) = P(A)$
  - applies even if $P(A) = 0$
- If A and B are independent, then A and B^C are independent.

2.2.2 Conditional independence

- Conditional independence, given C, is defined as independence under the probability law $P(\cdot|C)$

Conditioning may affect independence

- Two unfair coins, A and B: $P(H|\text{coin A}) = 0.9, P(H|\text{coin B}) = 0.1$
- Choose either coin with equal probability
  - $P(toss11 = H) = 0.5$
  - $P(toss11 = H|\text{first 10 tosses are heads}) = 0.9$

2.2.3 Independence of a collection of events

- **Intuitive "definition":** Information on some of the events does not change probability related to the remaining events
- **Definition:** Events $A_1, A_2, ..., A_n$ are called **independent** if $(A_i \cap A_j \cap ... \cap A_m) = P(A_i)P(A_j)P(A_m)$ for any distinct indices $i, j, ...m$

2.2.4 Pairwise independence

- Two independent fair coin tosses
  - H_1: First toss is H
  - H_2: Second toss is H
  - C: the two tosses had the same result
- Independence between H_1, H_2 and C
  - $P(H_1 \cap C) = P(H_1 \cap H_2) = 1/4$, $P(H_1)P(C) = 1/2 \cdot 1/2 = 1/4$
    * H_1 and C: independent, H_2 and C: independent
  - $P(H_1 \cap H_2 \cap C) = P(HH) = 1/4$, $P(H_1)P(H_2)P(C) = 1/8$
    * H_1, H_2 and C: not independent
- **Conclusion:** H_1, H_2, C are pairwise independent, but not independent

2.2.5 Reliability

- Independent units:
  - p_1 and p_2 and p_3 (series)
    * $P(\text{system up}) = p_1p_2p_3$
  - p_1 or p_2 or p_3 (parallel)
    * $P(\text{system up}) = 1 - (1 - p_1)(1 - p_2)(1 - p_3)$

2.2.6 The king's sibling puzzle

- The king comes from a family of two children what is the probability that his sibling is female? (boy have precedence)
- Combinations: BB, BG, GB, GG
- $P(\text{sibling is female}|\text{king}) = 2/3$

Chapter 3

Counting

3.1 Lecture 4: Counting

3.1.1 Discrete uniform law

- Assume Ω consists of n equaly likely elements
- Assume A consists of k elements
- Then: $P(A) = \frac{\text{number of elements of } A}{\text{number of elements of } \Omega} = \frac{k}{n}$
- Applications
  - Permutations, combinations
  - Partitions
  - Number of subsets
  - Binomial probabilities

3.1.2 Basic counting principle

Example

- 4 shirts, 3 ties, 2 jackets: number of possible attires?
  - r selection stages
  - n_i choices at stage i
- Number of choices is: $n_1 \cdot n_2 \cdot n_3 ... \cdot n_r$

More examples

- Number of license plates with 2 letters followed by 3 digits:
  - $26 \cdot 26 \cdot 10 \cdot 10 \cdot 10$
- **Permutations:** Number of ways of ordering n elements:
  - $n \cdot (n-1) \cdot (n-2) ... \cdot 1 = n!$
- Number of subsets of 1, ..., n:
  - $2 \cdot 2 \cdot ... \cdot 2 = 2^n$

3.1.3 Combinations

- Definition: $\binom{n}{k}$: number of k-element subsets of a given n-element set $= \frac{n!}{k!(n-k)!}$
- Two ways of constructing an **ordered** sequence of k **distinct** items:
  - Choose the k items one at a time
  - Choose k items, then order them

3.1.4 Partitions

- $n \geq 1$ distinct items; $r \geq 1$ persons, given n_i items to person i
  - here $n_1, ..., n_r$ are given nonnegative integers
  - with $n_1 + ...n_r = n$
- Ordering n items: $n!$
  - Deal n_i to each persons i, and then order
- Number of partitions $= \frac{n!}{n_1!n_2!...n_r!}$ (multinomial effect)

Appendix A

Frequently Asked Questions

A.1 How do I change the colors of links?

The color of links can be changed to your liking using:

`\hypersetup{urlcolor=red}`, or

`\hypersetup{citecolor=green}`, or

`\hypersetup{allcolor=blue}`.

If you want to completely hide the links, you can use:

`\hypersetup{allcolors=.}`, or even better:

`\hypersetup{hidelinks}`.

If you want to have obvious links in the PDF but not the printed text, use:

`\hypersetup{colorlinks=false}`.