EECS 126 Notes

Japjot Singh

March 22, 2020

Course notes for EECS 126 taken in Spring 2020.

Contents

1 Tuesday, January 21 **2**

1.1 Fundamentals . . . 2

1.2 Axioms of Probability (Kolmogov) . . . 3

1.3 Fundamental facts about probability . . . 3

1.4 Discrete Probability . . . 3

1.5 Conditional Probability . . . 3

1.6 Product (Multiplication) Rule . . . 4

1.7 Total Probability . . . 4

1.8 Bayes' Theorem . . . 4

2 Thursday, January 23 **5**

2.1 Announcement . . . 5

2.2 Birthday Paradox . . . 5

2.3 Bayes Rule False Positive Problem . . . 5

2.4 Independence . . . 5

2.4.1 Conditional Independence . . . 6

2.4.2 Independence of a collection of events . . . 6

3 Markov Chains **7**

3.1 Discrete Time Markov Chains . . . 7

3.2 Recurrence and Transience . . . 7

1 Tuesday, January 21

Understand problem as an "experiment" and then solve it using tools in your skillset: combinatorics, calculus, common sense.

1.1 Fundamentals

Definition 1. Sample Space Ω of an experiment is the set of all oucomes of the experiment.

Example 1.1

Your experiment is 2 fair coins $\Omega = \{HH, HT, TH, TT\}$ these outcomes (base outcomes) are **mutually exclusive (ME) and collectively exhaustive (CE)**

Example 1.2

Toss a coin till the first "Heads" $\Omega = \{H, TH, TTH, \ldots\}; |\Omega| = \infty$

Example 1.3

Waiting at the bus-stop for next bus $\Omega = (0, T)$

Visual 1 - We have the experiment which produces outcomes, once you have the outcome space the next definition is the definition of events

Definition 2 (Events)**.** Allowable subsets of Ω (collections of outcomes)

Example 1.4

Get at least 1 Head in experiment 1, $\{HH, HT, TH\}$, $p = \frac{3}{4}$

Defining events carefully is the key to tackling many tough problems.

Example 1.5

Ex 2.2: Get an even number of tosses

Example 1.6

Ex 2.3: Waiting time ¡ 5 min

Definition 3 (Probability Space)**.** A **probability space** $(\Omega, \mathcal{F}, \mathcal{P})$ is a mathematical construct to model "experiments" and has 3 components:

1. Ω is the set of all possible outcomes
2. $\mathcal{F}$ set of all events (composition of outcomes), where each event is a set containing 0 or more base outcomes, $\emptyset$ **is a base outcome** where $\mathbb{P}(\emptyset) = 0$. $\mathcal{F}$ is intuitively a powerset (i.e. for the experiment in example 1.1 $\mathcal{F} = \{\emptyset, \{H, H\}, \{H, T\}, \ldots\}$).
3. $\mathcal{P}$ is the proability measure which assigns a number in $[0, 1]$ to each event in $\mathcal{F}$.

Base oucomes must be ME and CE that is when writing out Ω as a collection of all the base outcomes, they should be the most simplified components.

1.2 Axioms of Probability (Kolmogov)

What properties do we need the probability measure $\mathcal{P}$ to satisfy?

1. $\mathbb{P}(\emptyset) = 0$
2. $\mathbb{P}(\Omega) = 1$, really just a normalization
3. $\mathbb{P}(A_1 \cup A_2 \cup \ldots) = P(A_1) + P(A_2) + \ldots$ for disjoint (ME) events $A_1, A_2, \ldots$

for disjoint $\mathbb{P}(\bigcup_{i=1}^n A_i) = \sum_{i=1}^n \mathbb{P}(A_i)$

1.3 Fundamental facts about probability

1. $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$
2. $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$ **Vis2 Venn Diagram**
3. Union-bound $\mathbb{P}(A_1 \cup A_2 \cup \ldots \cup A_n) \leq \sum_{i=1}^n \mathbb{P}(A_i)$
4. Inclusion-Exclusion, a generalized version of number 2

Theorem 4 (Inclusion-Exclusion)
$\mathbb{P}(A_1 \cup A_2 \cup \ldots \cup A_n) = \sum_{k=1}^n \sum_{1 \leq i_1} \sum_{\leq i_2} \ldots \sum_{i_k \leq n} (-1)^{i+1}$

Proof. here □

1.4 Discrete Probability

$$\mathbb{P}(A) = \sum_{\omega \in A} \mathbb{P}(\omega)$$

In a uniform sample space, all the outcomes are equally likely so then

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

1.5 Conditional Probability

Similar to events, conditioning on the right event will bail you out of tricky problems.

Definition 5. $P(A|B) := \mathbb{P}(\text{Event A given that Event B has occured})$

Thus for any event A if $\mathbb{P}(B) \neq 0$,

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

Example 1.7

Consider 2 six-sided dice. Let A be the event that the first dice rolls is a 6. Let B be the event the sum of the two dice is 7. Then

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(\{6,1\})}{\mathbb{P}(\{6,1\},\{5,2\},\ldots,\{1,6\})}$$

Similarly $\mathbb{P}(A|\text{sum is } 11) = \frac{1}{2}$

When conditioning on B, B becomes to new Ω.

1.6 Product (Multiplication) Rule

1.7 Total Probability

1.8 Bayes' Theorem

2 Thursday, January 23

2.1 Announcement

Readings B&T ch1 and 2, HW 1 due next wednesday one minute before midnight

2.2 Birthday Paradox

Assuming a group of n individuals whose birth dates are distributed uniformly at random. Given $k = 365$ days in the year what is the probability that at least 2 people in the group share the same birthday. Our sample space is the consists of each possible set of assignments of birth dates to the n students in the class. Since there are 365 posisble days for each of the n students in the group $|\Omega| = k^n = 365^n$. Now we can define our event of interest, A, that at least 2 people have the same birthday. Since this is a hard event to work with we can look at the complement A^c the event that no two people share a birth date. We can reach the solution with a counting argument

$$\mathbb{P}(A^c) = \frac{|A^c|}{|\Omega|} = \frac{365 * 364 * \ldots * (365 - (n-1))}{365^n}$$

or with a probabilistic argument using the chain rule

$$\mathbb{P}(A^c) = 1(1 - \frac{1}{k})(1 - \frac{2}{k}) \cdots (1 - \frac{n-1}{k})$$

the latter expression can be approximated using Taylor Series which say $e^x \approx 1 + x$ for $|x| << 1$.

$$\mathbb{P}(A^c) \approx 1 \cdot e^{-\frac{1}{k}} \cdot e^{-\frac{2}{k}} \cdots e^{-\frac{n-1}{k}}$$

thus $\mathbb{P}(A) = 1 - \mathbb{P}(A^c) \approx 1 - e^{-\frac{n^2}{2k}}$

2.3 Bayes Rule False Positive Problem

Suppoes there is a new test for a rare disease.

- If a person has the disease, test positive with $p = 0.95$
- If person does not have disease, test negative with $p = 0.95$
- Random person has the diseaes with $p = 0.001$

Suppose a person tested positive, what is the probability that person has the disease. Let A be the event has disease and B be the event test positive then by applying Bayes Rule directly

$$\mathbb{P}(A|B) = \frac{(0.95)(0.001)}{(0.95)(0.001) + (0.999)(0.05)} = 0.1875$$

the factor heavily contributing to this number is the prior, how rare the diesase is in the first place.

2.4 Independence

Definition 6. Two events are independent if the occurence of one provides **no information** about the occurence of the other (i.e. $\mathbb{P}(A|B) = \mathbb{P}(A)$).

insert vis1 Indepedence can also be written as

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$$

Note: Disjoint events are **not** Independent. Events A and B are disjoint if and only if $\mathbb{P}(A \cap B) = 0 \implies \mathbb{P}(A) = 0 \vee \mathbb{P}(B) = 0$. Thus since Base outcomes of a random experiment are disjoint (ME) and have non-zero probabilities they **must be dependent**.

2.4.1 Conditional Independence

$$\mathbb{P}(A \cap B|C) = \mathbb{P}(A|C) \cdot \mathbb{P}(B|C)$$

Note that

- Dependent events can be conditionally independent
- Independent events can be conditionally dependent

Example 2.1

Consider 2 indistinguishable coins: one is two-tailed and the other is two-headed. You pick on of the 2 coins at random and flip it twice.

Let H_i be the event that the i^{th} flip is a Head ($i = 1, 2$). By itself $\mathbb{P}(H_1) = \mathbb{P}(H_2) = \frac{1}{2}$ and $\mathbb{P}(H_2|H_1) = 1 \neq \mathbb{P}(H_2) = 1/2$. Furthermore, $\mathbb{P}(H_1 \cap H_2|A) = \mathbb{P}(H_1|A)\mathbb{P}(H_2|A \cap H_1) = \mathbb{P}(H_1|A)\mathbb{P}(H_2|A)$ which by definition tells us that H_1, H_2 are conditionally independent given A.

2.4.2 Independence of a collection of events

For all possible subsets of your events $A_{1:n}$, each subset must be independent that is

$$\mathbb{P}(\bigcap_{i \in S} A_i) = \prod_{i \in S}(\mathbb{P}(A_i)), \forall S$$

where S is any subset of the collection of events. Pairwise independence **does not imply** Joint independence of 3 or more events.

3 Markov Chains

Interested in models where the effect of the past on the future is summarized by a state, which changes over time given probabilities.

3.1 Discrete Time Markov Chains

In **discrete-time Markov chains**, state changes at certain discrete time instants, indexed by an integer variable n. At each step n, the state of the chain is denoted X_n, and belongs to a **finite** set $\mathcal{S}$ of possible states, called the state space. WLOG let $S = \{1, \ldots, m\}$.The Markov Chain is described in terms of its transition probabilities p_{ij} where

$$p_{ij} = \mathbb{P}(X_{n+1} = j | X_n = i) \quad i, j \in \mathcal{S}$$

The key assumption underlying these chains is that the transition probabilities apply whenever i is visited, no matter what happened in the past, and no matter how i was reached, formally this is the **Markov property**, requiring that:

$$\mathbb{P}(X_{n+1} = j | X_n = i, X_{n-1} = i_{n-1}, \ldots, X_0 = i_0) = \mathbb{P}(X_{n+1} = j | X_n = i) = p_{ij}$$

Furthermore the transition probabilities p_{ij} must be nonnegative, and sum to one:

$$\sum_{j=1}^{m} p_{ij} = 1, \text{ for all } i$$

Another efficient way to encode the MC chain model is a transition probability matrix, a 2D array whose element at row i and column j is p_{ij}, the transition probability from i to j

$$P = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1m} \\ p_{21} & p_{22} & \cdots & p_{2m} \\ \vdots & \vdots & \vdots & \vdots \\ p_{m1} & p_{m2} & \cdots & p_{mm} \end{bmatrix}$$

3.2 Recurrence and Transience

For each $x \in \mathcal{X}$, the random variable T_x represents the first time that the chain visits state x, that is $T_x := \min\{n \in \mathbb{N} : X_n = x$. T_x is called the hitting time of state x. We further define T_x^+ which is the hitting time for state x execept T_x^+ cannot equal 0, to avoid the trivial condition in which the chain starts at x. Let $\rho_{x,y} := \mathbb{P}_x(T_y^+ < \infty)$, the probability that starting from state x we eventually reach state y and shorthand $\rho_x = \rho_{x,x}$.

Definition 7. A state x is **recurrent** if $\rho_x = 1$ and **transient** if $\rho_x < 1$.

Proposition 8. Let N_x be the total number of visits to state x. If x is recurrent then $N_x = \infty$ in probability almost surely. Thus $\mathbb{E}_x[N_x] = \mathbb{E}[N_x | X_0 = x] = \infty$. On the other hand, if x is transient then $\mathbb{E}_x[N_x] < \infty$, in fact,

$$\mathbb{E}_x[N_x] = \frac{\rho_x}{1 - \rho_x}$$