Lecture Notes For: The Foundamentals of Probability

Ali Fele Paranj
alifele@student.ubc.ca

January 19, 2024

This lecture note contains my notes while I was studying the foundamentals of the probability probability. The main content is from the course in term winter2, UBC, 2023. However, I have exanded the content and examples using the material of two books:

- Statistical Modeling and Computation by Kroese
- Probability - Random Variables and Stochastic Processes by Papoulis

1 Foundamentals

The main concept in the field of statistics and probability is the set theory. Basically all we deal with the sets. The whole theroy of statistics can be built on that. Let's discuss some foundamental concepts in statistics and then build the theory.

1.1 Random Experiment

To understand the meaning of random experiment, do not over think! The first thing that comes into our minds when we hear the word "random experiment" is its definition! In a nutshell, random experiment is an experiment that its outcome is unkown to us. Like:

- Tossing two coin
- Rolling a dice
- Measuring the number of possible ReadWrite operations on a piece of EEPROM chip

Do not overthink about that. Yes we can go further and discuss stuff like "we can compute the exact movement of dice or coin so it is not random but determenistic" and etc. Here I will not touch the philosophical topics that are very deep and do not necessarily converge the a unified point of view!

The random experiments can be modeled and despite the fact that a random experiment is random, we can deduce many useful information from modeling that. To model a random experiment, we use three important concepts: sample space, events, probability. In the following section, we will discuss each of them in detail.

1.2 Sample Space

Definition: Sample Space

Sample space Ω is simply a set that contains *all possible outcomes* of a random experiment/

For each of random experiments described above, we can define a sample space. For example:

- Ω of Tossing Tow Coins:
$$\Omega = \{HH, HT, TH, TT\}$$
- Ω of Rolling a Dice:
$$\Omega = \{1, 2, 3, 4, 5, 6\}$$
- Ω of Rolling Two Dices:

$$\Omega = \{(1,1), (1,2), \ldots, (1,6), \ldots, (6,6)\}$$

- Ω of Number of possible ReadWrite operations on a EEPROM chip:

$$\Omega = \mathbb{N}$$

1.3 Events

Definition: Events

Event E is a set of outcomes of a random experiment and is the subset of sample space Ω.

$$E \in \Omega$$

For example for any of the sample spaces specified above, we can define so many possible events. In fact any set that is a subset of the sample space is a valid event of that sample space. For example:

- Tossing Three Coins
  - There are at least on Heads:

    $$E = \{HHH, HHT, HTH, THH, HTT, THT, TTH\}$$

  - There are only two Tails:

    $$E = \{TTH, THT, HTT\}$$

- Rolling Two Dices
  - The sum of two dices is 4:

    $$E = \{(1, 3), (2, 2), (3, 1)\}$$

  - there are at least one prime number in the outcome:

    $$E = \{(1, 2), (1, 3), (1, 5), (2, 1), (3, 1), (5, 1), (2, 2), (2, 3), (2, 5), \dots, (5, 5)\}$$

Since we have define everything on the basics of set theory, then now we can correspond the everyday concepts to specific operations in the set theory.

Example: The Mapping Between Everyday Language and Sets in the Theory of Probability

- At least one of two events $A, B \in \Omega$ happens: $E = A \cup B$.
- Tow events $A, B \in \Omega$ occures at the same time: $E = A \cap B$.
- Event $A \in \Omega$ does not happen: $E = \overline{A} = \Omega - A$.

- The event A happens but B does not happen: $E = A - B$.

In probability and statistics, we are dealing with three important concepts: sample space Ω, event E, and probability P.

Definition: Disjoint events

If two events has no common elements (i.e. $A \cap B = \varnothing$) then we say that two events are *disjoint*. Basically, if two sets in the venn diagram has nothing is common they are considerent to be disjoint sets.
For example for the random experiment of tossing two coins, the events 1) both coins are heads: $A = \{HH\}$ and 2) both coins are tails: $B = \{TT\}$. Two events A, B are two disjoint events. **Two events being disjoing is NOT the same as being independent**. We will talk about independet events in future.

Note that since the events are basically sets, we can use theorems of set theory to solve the problems.

Theorem: De Morgan's Laws

If A, B are two sets then:

$$\overline{A \cap B} = \overline{A} \cup \overline{B}$$

$$\overline{A \cup B} = \overline{A} \cap \overline{B}$$

Proof. the proof is left as an excerise! □

1.4 Probability

The last foundamental ingeridient in modeling a random experiment, is to define a probability for each event. The probability should intuitively reflect how likely an event is probable to happen. This probability should satisfy some foundamental properties which are explained as follows.

Axiom: Axioms of probability (Kolmogorov axioms)

Suppose that $A, B \in \Omega$ is an event and $\mathbb{P}$ is a probability function. Then $\mathbb{P}$ should satisfy the following properites:

1. $0 \leq \mathbb{P}(A) \leq 1$
2. $\mathbb{P}(\Omega) = 1$
3. For the events $E_1, E_2, ..., E_n \in \Omega$ that are mutually exclusive (i.e. disjoint

events):

$$\mathbb{P}(\bigcup_i E_i) = \sum_i \mathbb{P}(E_i)$$

These axioms are called the foundamental axioms of probability and also the Kolmogorov axioms. We are free to define any kind of probabilty function that we want but it is important that 1) It should align with our common sense, 2) It should satisfy the Kolmogorov axioms.

Using the axioms above, we can observe and prove several interesting properties of the probability function. In the following box we have expressed some of them.

Theorem: Basic Properties of the Probability Function

Suppose that $\mathbb{P}$ is a probability function and $A, B \in \Omega$ are events of the sample space Ω. We can show that the probability function has the following properties:

1. $\mathbb{P}(\varnothing) = 0$
2. If $A \subset B$ then $\mathbb{P}(A) \leq \mathbb{P}(B)$.
3. $\mathbb{P}(\overline{A}) = 1 - \mathbb{P}(A)$.
4. $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$.

Proof. The properties can be proved using the basic set theory theorems.

1. Since $\varnothing$ is the complement of Ω, so these two sets are disjoint (i.e. $\varnothing \cap \Omega = \varnothing$). On the other hand from the set theory we know that $\varnothing \cup \Omega = \Omega$. So $\mathbb{P}(\varnothing \cup \Omega) = \mathbb{P}(\Omega)$. On the other hand, using the third axiom we can write: $\mathbb{P}(\varnothing \cup \Omega) = \mathbb{P}(\varnothing) + \mathbb{P}(\Omega)$. Comparing the two recent equations we can conclude that $\mathbb{P}(\varnothing) = 0$.

The proofs for 2,3,4 are left as a exercise. However, the solutions can be found in the book "Statistical Modeling and Computation by Kroese" chapter 1.

$\square$

Example: Defining a simple probability function

Let's define a probability function for the rolling n dice experiment that is both aligned with our common sense and also satisfy the Kolmogorov equations. Suppose that the Ω is the sample space and $E \in \Omega$ is an event. Then let's define:

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

in which the $|E|$ means the cardinality (number of elements) of the set E.

1.5 Conditional Probability and Independentness

1.6 Law of Total Probability

1.7 Baye's Rule

1.8 Some Solved Problems

Example: Hats problem

Question. In a party, three men through their hats on a table and then pick a hat in a random. What is the probability that non of men has got their own hats.

Solutions. We can solve this question in two ways. The first solution is kind of intuitive and very short. However, to get use to the mathematical structure of the probability theory, I have also forced my self to solve this problem with the usual notation of the conditional probability. This is a very important thing to practice. Because the real world problems are so massive that there are no ways to tackle them with the intuition. That is why I really need to get equipped with the standard machinery of the statistical stuff that works no matter what is the size of problem!

Solution 1. This problem can be though of as a permutation problem. In other words we have the numbers $1, 2, 3$ and we want to permute them in a way that non of them has its previous position. I will utilize the matrix like notation for the permutation here. The set of all of permutations of $1, 2, 3$ is:

$$\Omega = \{\begin{pmatrix}1 & 2 & 3\\1 & 2 & 3\end{pmatrix}, \begin{pmatrix}1 & 2 & 3\\1 & 3 & 2\end{pmatrix}, \begin{pmatrix}1 & 2 & 3\\2 & 3 & 1\end{pmatrix}, \begin{pmatrix}2 & 2 & 3\\2 & 3 & 2\end{pmatrix}, \begin{pmatrix}1 & 2 & 3\\3 & 2 & 1\end{pmatrix}, \begin{pmatrix}1 & 2 & 3\\3 & 1 & 2\end{pmatrix}\},$$

To elaborate more about this notation, as an instance, the permutation

$$\begin{pmatrix}1 & 2 & 3\\3 & 1 & 2\end{pmatrix}$$

means that the first man had picked up the third man's had. The second man had picked up the first man's had and the third man has picked up the second man's hat. So in fact the set Ω is in fact the sample space of our random experiment. So the probability that no one will pick his own hat is $P(E) = \frac{2}{6} = \frac{1}{3}$.

Solution 2. In this section I will use the inclusion-exclusion identity along with the concept of the conditional probability. First let's define the following events:

E_1: The first man has picked up his own hat.
E_2: The second man has picked up his own hat.
E_3: The third man has picked up his own hat.

Given this definitions, we can compute $P(E_i)$ for $i \in \{1,2,3\}$, and $P(E_i \cap E_j)$ for $i, j \in 1, 2, 3$ and $i \neq j$ and finally $P(E_1 \cap E_2 \cap E_3)$.
The first expression $P(E_i)$ is very easy to calculate. Since every man picks his own hat in a random way, then $P(E_i) = 1/3$. To calculate $P(E_i \cap E_j)$ we need to use the law of conditional probability:

$$P(E_i \cap E_j) = P(E_i|E_j)P(E_j).$$

We know that $P(E_j) = 1/3$. And for $P(E_i|E_j)$, it means that what is the chance that given one man has chosen his hat right, then other man would choose his hat right. Imagine the scenario. One man has chosen his hat right. Then there are two hats on the table and the other man wants to choose his hat. The chance that he gets his own hat is 1/2. So $P(E_i|E_j) = 1/6$. As the last step we need to calculate $P(E_1 \cap P_2 \cap E_3)$. Again using the conditional probability we can rewrite this as:

$$P(E_1 \cap E_2 \cap E_3) = P(E_1|E_2 \cap E_3)P(E_2 \cap E_3).$$

As calculated in the last step $P(E_1|E_2)$ is 1/6. And $P(E_1|E_2 \cap E_3)$ is 1. Since when both E_2 and E_3 has gotten their hats right, then there is only one hat remaining on the table. So the first person will get his hat right with 1 probability. So in conclusion $P(E_1 \cap E_2 \cap E_3) = 1/6$
Now it is time to use the inclusive-exclusive law:

$$\begin{aligned} P(E_1 \cup E_2 \cup E_3) &= P(E_1) + P(E_2) + P(E_3) - P(E_1 \cap E_2) \\ &- P(E_1 \cap E_3) - P(E_2 \cap E_3) + P(E_1 \cap E_2 \cap E_3) \\ &= 1/3 + 1/3 + 1/3 - 1/6 - 1/6 - 1/6 + 1/6 = 1 - 1/3 = 2/3 \end{aligned}$$

However note that this is not the final answer. $P(E_1 \cup E_2 \cup E_3)$ shows the probability that at least on of the men had picked up a right hat. So the answer to the original question will be $1 - P(E_1 \cup E_2 \cup E_3) = 1 - 2/3 = 1/3$.

Example: Tossing Coin

Choose one number in $\{1, 2, \ldots, 10\}$ randomly. Then toss that many dices. What is the probability that you get the sum equal to 3.

Example: Medical Test

Null Hypothesis (H_0): The person is *not* healthy (so if the results of test is positive it means the null hypothesis is accepted and the person is infected and is not healthy).
In a nutshell: positive test $\rightsquigarrow$ person is infected!
Question: If the test if positive what is the probability that person is infected.

Example: Monty Hall Problem

Suppose that I have chose the box #1 and the Monty Hall has opened the box #2. What is the chance that the prise is in box #1 and what is the probability that the prise is in box #2.