CC 511: Chapter 1

Probability Theory

Jae-kwang Kim

Outline

1. Sample spaces and events
2. Probability
3. Conditional Probability
4. Independence
5. Bayes' theorem

1.1 Sample space and Events

- Experiment: any process or procedure for which more than one outcome is possible
- Sample Space
  The sample space S of an experiment is a set consisting of all of the possible experimental outcomes.
- Event: subset of S.

Example 1

If we toss a coin twice, then

- Sample space

$$S = \{HH, HT, TH, TT\}$$

- The event that the first toss is head is

$$A = \{HH, HT\}$$

첫번째 던진 동전의 면 의 sample space : S` = {H,T} = { {HH,HT} , {TH,TT} }
H T

Set Operations

Let A and B be the subset of S:

- A^c: the complement of A

$$A^c = \{a \in S; a \notin A\}$$

- $A \cup B$: the union of A and B

$$A \cup B = \{a \in S; a \in A \text{ or } a \in B\}$$

- $A \cap B$: the intersection of A and B

$$A \cap B = \{a \in S; a \in A \text{ and } a \in B\}$$

- $A - B$: set difference

$$A - B = \{a \in S; a \in A \text{ and } a \notin B\}$$

Definition

Let $A_1, \cdots, A_n$ be n subsets of S.

- $A_1, \cdots, A_n$ are *disjoint* if $A_i \cap A_j = \phi$ whenever $i \neq j$.
- $A_1, \cdots, A_n$ is called a *partition* of S if they are disjoint and $\cup_{i=1}^{n} A_i = S$.

1.2 Probability

Motivation

1. Probability is a set function in the sense that it is a mapping from A to $[0, 1]$. We wish to assign a real number (between zero and one) to every event A in a reasonable way
2. Reasonable in the following senses

$$P(\phi) = 0$$
$$A \subset B \implies P(A) \leq P(B)$$
$$P(A^c) = 1 - P(A)$$
$$A \cap B = \phi \implies P(A \cup B) = P(A) + P(B)$$

Definition

Definition

A function $\mathbb{P}$ that assigns a real number $\mathbb{P}(A)$ to each event A in the sample space S is a *probability distribution* if it satisfies the following three axioms:

1. $\mathbb{P}(A) \geq 0$ for every A
2. $\mathbb{P}(S) = 1$
3. If $A_1, A_2, \cdots, A_n$ are disjoint then

$$\mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} \mathbb{P}(A_i)$$ n : countable (not infinity)

Remark

- Two interpretations
  - Frequency interpretation: $\mathbb{P}(A)$ is the long run proportion of times that A is true in repetitions.
  - Degree-of-belief interpretation: $\mathbb{P}(A)$ measures an observer's strength of belief that A is true.
- The differing interpretations lead to two schools of inference in statistics: the frequentist and the Bayesian schools
- In either interpretation, we require that Axioms 1 to 3 holds.

Lemma

Lemma

For any events A and B,

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B).$$

Proof:

P(A) + P(B) = P((A-B) U (A∩B)) + P(B) = P(A-B) + P(A∩B) + P(B) = P(AUB) + P (A∩B)

1.3 Conditional Probability

Definition

The conditional probability of event A conditional on event B is

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

for $\mathbb{P}(B) > 0$. It measures the probability that event A occurs when it is known that event B occurs.

- If $A \cap B = \phi$ then $\mathbb{P}(A \mid B) = 0$.
- If $B \subset A$, then $\mathbb{P}(A \mid B) = 1$.
- If $B = S$, then $\mathbb{P}(A \mid B) = \mathbb{P}(A)$.

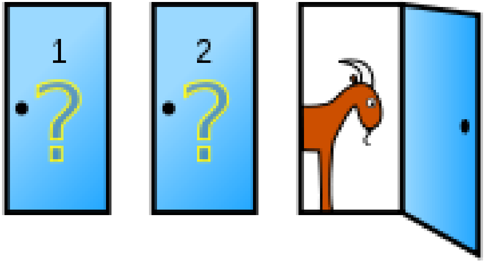
Example 2

Suppose we toss one fair, six-sided die. The sample space $S = \{1, 2, 3, 4, 5, 6\}$. Let $A =$ face is 2 or 3 and $B =$ face is even $(2, 4, 6)$. Calculate $\mathbb{P}(A|B)$.

Monty Hall Problem

The Monty Hall problem is a brain teaser loosely based on the American television game show "Let's Make a Deal" and named after its original host, Monty Hall. Here's the rule:

- There are 3 doors, behind which are two goats and a car.
- You pick a door (call it door 1). You're hoping for the car of course.
- Monty Hall, the game show host, examines the other doors (2 & 3) and always opens one of them with a goat (Both doors might have goats; he'll randomly pick one to open)



Monty Hall Problem (Cont'd)

Here's the game:
Now, you have an option to change your choice. Do you stick with door 1 (original guess) or switch to the other unopened door? Does it matter?

Example 3

	Right-handed	Left-handed
Males	43	9
Females	44	4

Let's denote the events M = the subject is male, F = the subject is female, R = the subject is right-handed, L = the subject is left-handed. Compute the following probabilities:

- $\mathbb{P}(M), \mathbb{P}(F), \mathbb{P}(R), \mathbb{P}(L)$
- $\mathbb{P}(M \cup R), \mathbb{P}(F \cup L), \mathbb{P}(M \cap F), \mathbb{P}(M \cap R), \mathbb{P}(F \cap L)$
- $\mathbb{P}(M^c)$
- $\mathbb{P}(R|M), \mathbb{P}(F|L), \mathbb{P}(L|F)$

Back to Monty Hall Problem

Define

D1: The Event of Monty Hall opening door 1.

D2: The Event of Monty Hall opening door 2.

D3: The Event of Monty Hall opening door 3.

C1: The Event of finding the car behind door 1.

C2: The Event of finding the car behind door 2.

C3: The Event of finding the car behind door 3.

사용자가 1번을 찍은 상태에서

1. Find $\underbrace{\mathbb{P}(D3 \mid C1)}_{=1/2}$, $\underbrace{\mathbb{P}(D3 \mid C3)}_{=0}$, and $\underbrace{\mathbb{P}(D3 \mid C2)}_{=1}$.
2. Show that $\mathbb{P}(C1 \mid D3) = 1/3$ and $\mathbb{P}(C2 \mid D3) = 2/3$.

1.4 Independent and Mutually Exclusive Events

Definition

Two events are independent if any of the following are true:

independent : 계산량 줄이기

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

In general $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B \mid A)$ holds. Under independence, we have $\mathbb{P}(B \mid A) = \mathbb{P}(B)$, which implies $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$.

Mutually Exclusive Events

Definition

A and B are mutually exclusive events if they cannot occur at the same time. This means that A and B do not share any outcomes and $\mathbb{P}(A \cap B) = 0$.

Suppose the sample space $S = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$. Let $A = \{1, 2, 3, 4, 5\}$, $B = \{4, 5, 6, 7, 8\}$, and $C = \{7, 9\}$. $A \cap B = \{4, 5\}$. $\mathbb{P}(A \cap B) = \frac{2}{10}$ and is not equal to zero. Therefore, A and B are not mutually exclusive. A and C do not have any numbers in common so $\mathbb{P}(A \cap C) = 0$. Therefore, A and C are mutually exclusive.

Example 5

In a bag, there are six red marbles and four green marbles. The red marbles are marked with the numbers 1, 2, 3, 4, 5, and 6. The green marbles are marked with the numbers 1, 2, 3, and 4.

- $R =$ a red marble
- $G =$ a green marble
- $O =$ an odd-numbered marble
- The sample space is $S = \{R1, R2, R3, R4, R5, R6, G1, G2, G3, G4\}$.

S has ten outcomes. If we pick one marble at random from the bag, what is $\mathbb{P}(G \cap O)$?

Example 6

Let event C = taking an English class. Let event D = taking a speech class. Suppose $\mathbb{P}(C) = 0.75$, $\mathbb{P}(D) = 0.3$, $\mathbb{P}(C|D) = 0.75$ and $\mathbb{P}(C \cap D) = 0.225$.

- Are C and D independent? yes
- Are C and D mutually exclusive? no
- What is $\mathbb{P}(D|C)$? 0.3
  1. we already know C and D are independent !
  2. P(D|C) = P(C ∩D) / P(C) = (P(D) * P(C|D)) / P(C)

Example 7

A box has two balls, one white and one red. We select one ball, put it back in the box, and select a second ball (sampling with replacement). Let T be the event of getting the white ball twice, F the event of picking the white ball first, S the event of picking the white ball in the second drawing.

- Compute $\mathbb{P}(T)$.
- Compute $\mathbb{P}(T|F)$.
- Are T and F independent?.
- Are F and S mutually exclusive?
- Are F and S independent?

1.5 Bayes Theorem

Lemma

Let $A_1, \cdots, A_k$ be a partition of S. Then, for any event B,

$$\mathbb{P}(B) = \sum_{i=1}^{k} \mathbb{P}(B \mid A_i)\mathbb{P}(A_i).$$

S = A1 U A2 U .... U Ak,
B = B ∩S = (B ∩A1) U (B ∩A2) ..... U (B ∩Ak)

Bayes' theorem

Theorem

Let $A_1, \cdots, A_k$ be a partition of S such that $\mathbb{P}(A_i) > 0$ for each i. If $\mathbb{P}(B) > 0$ then, for each $i = 1, \cdots, k$,

$$\mathbb{P}(A_i \mid B) = \frac{\mathbb{P}(B \mid A_i)\mathbb{P}(A_i)}{\sum_{j=1}^{k} \mathbb{P}(B \mid A_j)\mathbb{P}(A_j)}.$$

We call $\mathbb{P}(A_i)$ the *prior* probability of A and $\mathbb{P}(A_i \mid B)$ the *posterior probability* of A.

A1 ... Ak = 발생할 수 있는 전체 경우의 수들 , B = 우리가 특정하게 관심있는 이벤트
A1이 발생했을 때 B가 발생할 확률은 ? P(B|A1)

Example: Breast cancer test

Here is some background information about breast cancer:

- 1% of women have breast cancer (and therefore 99% do not).
- 80% of mammograms detect breast cancer when it is there (and therefore 20% miss it).
- 10% of mammograms detect breast cancer when it is not there (and therefore 90% correctly return a negative result).

Example: Breast cancer test (Cont'd)

The probabilities are as follows:

	Cancer (1%)	No Cancer (99%)
Test Pos	True Pos: 1% * 80%	False Pos: 99% *10%
Test Neg	False Neg: 1% * 20%	True Neg: 99% * 90%

What is the chance we really have cancer if we get a positive result?

Define T=test outcome, $C = 1$ if cancer and $C = 0$ otherwise. We wish to compute

$$\begin{aligned} & \mathbb{P}(C = 1 \mid T = +) \\ = & \frac{\mathbb{P}(T = + \mid C = 1)\mathbb{P}(C = 1)}{\mathbb{P}(T = + \mid C = 1)\mathbb{P}(C = 1) + \mathbb{P}(T = + \mid C = 0)\mathbb{P}(C = 0)} \end{aligned}$$

My knowledge of cancer status is updated by incorporating the test outcome. That is, the prior probability $\mathbb{P}(C = 1)$ is updated by the posterior probability $\mathbb{P}(C = 1 \mid T = +)$, which incorporates observed data.

Example: Breast cancer test (Cont'd)

We may use a frequency table instead of probability.

	Cancer	No Cancer
Test Pos	N_{11}	N_{12}
Test Neg	N_{21}	N_{22}
	100	9,900

Thus, the answer is

$$\mathbb{P}(C = 1 \mid T = +) = \frac{N_{11}}{N_{11} + N_{12}} = \frac{(\quad)}{(\quad) + (\quad)} = 0.0748$$

Thus, even if you get a positive test result, the chance of real cancer is only 7.48%.

Example: Breast cancer test (Cont'd)

Another interpretation:
Posterior probability can be understood in terms of odds:

$$\frac{\mathbb{P}(C=1 \mid T=+)}{\mathbb{P}(C=0 \mid T=+)} = \frac{\mathbb{P}(T=+ \mid C=1)}{\mathbb{P}(T=+ \mid C=0)} \times \frac{\mathbb{P}(C=1)}{\mathbb{P}(C=0)}.$$

That is

$$\text{New Odds} = \text{Evidence Adjustment} \times \text{Initial Odds}.$$

In our cancer example, the evidence adjustment is (   ). But, since the initial odd is low (   ), the new odds is still low.

Some further thoughts:
Suppose that you got a positive test result. Now, you want to take another independent test. What is the chance we really have cancer if we get a positive result in the second test?

Example: Breast cancer test (Cont'd)

May use

$$
\begin{aligned}
& \mathbb{P}(C = 1 \mid T_1, T_2) \\
= \quad & \frac{\mathbb{P}(T_2 \mid C = 1)\mathbb{P}(C = 1 \mid T_1)}{\mathbb{P}(T_2 \mid C = 1)\mathbb{P}(C = 1 \mid T_1) + \mathbb{P}(T_2 \mid C = 0)\mathbb{P}(C = 0 \mid T_1)}
\end{aligned}
$$

Thus, the cancer probability is updated as follows:

$$\mathbb{P}(C = 1) \Rightarrow \mathbb{P}(C = 1 \mid T_1) \Rightarrow \mathbb{P}(C = 1 \mid T_1, T_2).$$

Summary

- **Tests are not the event.** We have a cancer test, separate from the event of actually having cancer.
- **Tests are flawed.** Tests detect things that don't exist (false positive), and miss things that do exist (false negative).
- **Tests give us test probabilities, not the real probabilities.** People often consider the test results directly, without considering the errors in the tests.
- **False positives skew results.** Suppose you are searching for something really rare (1 in a million). Even with a good test, it's likely that a positive result is really a false positive on somebody in the 999,999.
- **Even science is a test.** At a philosophical level, scientific experiments can be considered "potentially flawed tests" and need to be treated accordingly. There is a test for a phenomenon, and there is the event of the phenomenon itself. Our tests and measuring equipment have some inherent rate of error.

Summary (Cont'd)

Bayes' theorem converts the results from your test into the real probability of the event. For example, you can:

- Correct for measurement errors. If you know the real probabilities and the chance of a false positive and false negative, you can correct for measurement errors.
- Relate the actual probability to the measured test probability. Bayes' theorem lets you relate $\mathbb{P}(A|X)$, the chance that an event A happened given the evidence X, and $\mathbb{P}(X|A)$, the chance the evidence X happened given that event A occurred. Given mammogram test results and known error rates, you can predict the actual chance of having cancer.