CS70 Discrete Probability II

Kelvin Lee

kelvinlee@berkeley.edu

November 1, 2020

Overview

1. Conditional Probability
2. Bayes' Rule
3. Law of Total Probability
4. Independence
5. Union Bound

Conditional Probability

Conditional Probability

- Given that event B happened, we are interested in finding the probability of event A, which is $P(A|B)$.

Conditional Probability

- Given that event B happened, we are interested in finding the probability of event A, which is $P(A|B)$.
- Since B is guaranteed to happen, we restrict our focus to the sample points in B instead of Ω.

Conditional Probability

- Given that event B happened, we are interested in finding the probability of event A, which is $P(A|B)$.
- Since B is guaranteed to happen, we restrict our focus to the sample points in B instead of Ω.
- Then we know that

Conditional Probability

- Given that event B happened, we are interested in finding the probability of event A, which is $P(A|B)$.
- Since B is guaranteed to happen, we restrict our focus to the sample points in B instead of Ω.
- Then we know that

$$P(B) = \sum_{\omega \in B} P(\omega),$$

Conditional Probability

- Given that event B happened, we are interested in finding the probability of event A, which is $P(A|B)$.
- Since B is guaranteed to happen, we restrict our focus to the sample points in B instead of Ω.
- Then we know that

$$P(B) = \sum_{\omega \in B} P(\omega),$$

which is less than 1 in general.

Conditional Probability

- Given that event B happened, we are interested in finding the probability of event A, which is $P(A|B)$.
- Since B is guaranteed to happen, we restrict our focus to the sample points in B instead of Ω.
- Then we know that

$$P(B) = \sum_{\omega \in B} P(\omega),$$

  which is less than 1 in general.
- To get the correct **normalization**, we scale the probability of each ω by $\frac{1}{P(B)}$ and so

Conditional Probability

- Given that event B happened, we are interested in finding the probability of event A, which is $P(A|B)$.
- Since B is guaranteed to happen, we restrict our focus to the sample points in B instead of Ω.
- Then we know that

$$P(B) = \sum_{\omega \in B} P(\omega),$$

which is less than 1 in general.

- To get the correct **normalization**, we scale the probability of each ω by $\frac{1}{P(B)}$ and so

$$P(\omega|B) = \frac{P(\omega)}{P(B)}.$$

Conditional Probability

Conditional Probability

- Now we are ready to compute $P(A|B)$:

Conditional Probability

- Now we are ready to compute $P(A|B)$:

$$P(A \mid B) = \sum_{\omega \in A \cap B} P(\omega \mid B) = \sum_{\omega \in A \cap B} \frac{P(\omega)}{P(B)} = \frac{P(A \cap B)}{P(B)}.$$

Conditional Probability

- Now we are ready to compute $P(A|B)$:

$$P(A \mid B) = \sum_{\omega \in A \cap B} P(\omega \mid B) = \sum_{\omega \in A \cap B} \frac{P(\omega)}{P(B)} = \frac{P(A \cap B)}{P(B)}.$$

- For events $A, B \subseteq \Omega$ in the same probability space such that $P(B) > 0$, the **conditional probability** of A given B is

Conditional Probability

- Now we are ready to compute $P(A|B)$:

$$P(A \mid B) = \sum_{\omega \in A \cap B} P(\omega \mid B) = \sum_{\omega \in A \cap B} \frac{P(\omega)}{P(B)} = \frac{P(A \cap B)}{P(B)}.$$

- For events $A, B \subseteq \Omega$ in the same probability space such that $P(B) > 0$, the **conditional probability** of A given B is

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}.$$

Conditional Probability

- Now we are ready to compute $P(A|B)$:

$$P(A \mid B) = \sum_{\omega \in A \cap B} P(\omega \mid B) = \sum_{\omega \in A \cap B} \frac{P(\omega)}{P(B)} = \frac{P(A \cap B)}{P(B)}.$$

- For events $A, B \subseteq \Omega$ in the same probability space such that $P(B) > 0$, the **conditional probability** of A given B is

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}.$$

- **Chain Rule.**

Conditional Probability

- Now we are ready to compute $P(A|B)$:

$$P(A \mid B) = \sum_{\omega \in A \cap B} P(\omega \mid B) = \sum_{\omega \in A \cap B} \frac{P(\omega)}{P(B)} = \frac{P(A \cap B)}{P(B)}.$$

- For events $A, B \subseteq \Omega$ in the same probability space such that $P(B) > 0$, the **conditional probability** of A given B is

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}.$$

- **Chain Rule.**

$$P(A \cap B) = P(A \mid B) \cdot P(B).$$

Bayes' Rule

Bayes' Rule

How do we compute $P(A \mid B)$ given $P(B \mid A)$?

Bayes' Rule

How do we compute $P(A \mid B)$ given $P(B \mid A)$?

- Using the definition and chain rule, we have

Bayes' Rule

How do we compute $P(A \mid B)$ given $P(B \mid A)$?

- Using the definition and chain rule, we have

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \frac{P(B \mid A) \cdot P(A)}{P(B)},$$

Bayes' Rule

How do we compute $P(A \mid B)$ given $P(B \mid A)$?

- Using the definition and chain rule, we have

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \frac{P(B \mid A) \cdot P(A)}{P(B)},$$

which is the **Bayes' Rule**.

Law of Total Probability

Law of Total Probability

Usually we are not given what $P(B)$ is, so we need to figure out $P(B)$ based on our current information.

Law of Total Probability

Usually we are not given what $P(B)$ is, so we need to figure out $P(B)$ based on our current information.

- Observe that

Law of Total Probability

Usually we are not given what $P(B)$ is, so we need to figure out $P(B)$ based on our current information.

- Observe that

$$P(B) = P(A \cap B) + P(\overline{A} \cap B),$$

Law of Total Probability

Usually we are not given what $P(B)$ is, so we need to figure out $P(B)$ based on our current information.

- Observe that

$$P(B) = P(A \cap B) + P(\overline{A} \cap B),$$

which is equivalent to

Law of Total Probability

Usually we are not given what $P(B)$ is, so we need to figure out $P(B)$ based on our current information.

- Observe that

$$P(B) = P(A \cap B) + P(\overline{A} \cap B),$$

  which is equivalent to

$$P(B) = P(B \mid A) \cdot P(A) + P(B \mid \overline{A})P(\overline{A}).$$

Law of Total Probability

Usually we are not given what $P(B)$ is, so we need to figure out $P(B)$ based on our current information.

- Observe that

$$P(B) = P(A \cap B) + P(\overline{A} \cap B),$$

which is equivalent to

$$P(B) = P(B \mid A) \cdot P(A) + P(B \mid \overline{A})P(\overline{A}).$$

This is exactly the **Law of Total Probability**, which is a very important law in probability theory.

Bayes' Rule

Bayes' Rule

Now putting back together what we have currently, we have

Bayes' Rule

Now putting back together what we have currently, we have

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

Bayes' Rule

Now putting back together what we have currently, we have

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

Bayes' Rule

Now putting back together what we have currently, we have

$$\begin{aligned} P(A|B) &= \frac{P(A \cap B)}{P(B)} \\ &= \frac{P(B \mid A) \cdot P(A)}{P(B)} \\ &= \frac{P(B \mid A) \cdot P(A)}{P(B \mid A) \cdot P(A) + P(B \mid \overline{A}) \cdot P(\overline{A})} \end{aligned}$$

Bayes' Rule

Now putting back together what we have currently, we have

$$\begin{aligned}
P(A|B) &= \frac{P(A \cap B)}{P(B)} \\
&= \frac{P(B \mid A) \cdot P(A)}{P(B)} \\
&= \frac{P(B \mid A) \cdot P(A)}{P(B \mid A) \cdot P(A) + P(B \mid \overline{A}) \cdot P(\overline{A})} \\
&= \frac{P(B \mid A) \cdot P(A)}{P(B \mid A) \cdot P(A) + P(B \mid \overline{A}) \cdot (1 - P(A))}.
\end{aligned}$$

Law of Total Probability

Law of Total Probability

Let's now consider **Bayes' Rule** and the **Law of Total Probability** in a more general context. First, let's define a partition of an event as follows:

Law of Total Probability

Let's now consider **Bayes' Rule** and the **Law of Total Probability** in a more general context. First, let's define a partition of an event as follows:

- **(Partition of an event).** We say that an event A is partitioned into n events $A_1, \ldots, A_n$ if

Law of Total Probability

Let's now consider **Bayes' Rule** and the **Law of Total Probability** in a more general context. First, let's define a partition of an event as follows:

- **(Partition of an event).** We say that an event A is partitioned into n events $A_1, \ldots, A_n$ if
  - $A = A_1 \cup A_2 \cup \cdots \cup A_n$,

Law of Total Probability

Let's now consider **Bayes' Rule** and the **Law of Total Probability** in a more general context. First, let's define a partition of an event as follows:

- **(Partition of an event).** We say that an event A is partitioned into n events $A_1, \ldots, A_n$ if
  - $A = A_1 \cup A_2 \cup \cdots \cup A_n$,
  - $A_i \cap A_j = \varnothing$ for all $i \neq j$ (i.e., $A_1, \ldots, A_n$ are mutually exclusive).

Law of Total Probability

Let's now consider **Bayes' Rule** and the **Law of Total Probability** in a more general context. First, let's define a partition of an event as follows:

- **(Partition of an event).** We say that an event A is partitioned into n events $A_1, \dots, A_n$ if
  - $A = A_1 \cup A_2 \cup \cdots \cup A_n$,
  - $A_i \cap A_j = \varnothing$ for all $i \neq j$ (i.e., $A_1, \dots, A_n$ are mutually exclusive).

  In simpler terms, each outcome in A belongs to exactly one of the events $A_1, \dots, A_n$.

Law of Total Probability

Let's now consider **Bayes' Rule** and the **Law of Total Probability** in a more general context. First, let's define a partition of an event as follows:

- **(Partition of an event).** We say that an event A is partitioned into n events $A_1, \ldots, A_n$ if
  - $A = A_1 \cup A_2 \cup \cdots \cup A_n$,
  - $A_i \cap A_j = \varnothing$ for all $i \neq j$ (i.e., $A_1, \ldots, A_n$ are mutually exclusive).

  In simpler terms, each outcome in A belongs to exactly one of the events $A_1, \ldots, A_n$.
- Now, let $A_1, \ldots, A_n$ be a **partition** of the sample space Ω. Then, the **Law of Total Probability** for any event B is

Law of Total Probability

Let's now consider **Bayes' Rule** and the **Law of Total Probability** in a more general context. First, let's define a partition of an event as follows:

- **(Partition of an event).** We say that an event A is partitioned into n events $A_1, \ldots, A_n$ if
  - $A = A_1 \cup A_2 \cup \cdots \cup A_n$,
  - $A_i \cap A_j = \varnothing$ for all $i \neq j$ (i.e., $A_1, \ldots, A_n$ are mutually exclusive).

  In simpler terms, each outcome in A belongs to exactly one of the events $A_1, \ldots, A_n$.
- Now, let $A_1, \ldots, A_n$ be a **partition** of the sample space Ω. Then, the **Law of Total Probability** for any event B is

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

Bayes' Rule

Bayes' Rule

Similarly, with the general version of **Law of Total Probability**, the general version of **Bayes' Rule**, assuming $P(B) \neq 0$, is given by

Bayes' Rule

Similarly, with the general version of **Law of Total Probability**, the general version of **Bayes' Rule**, assuming $P(B) \neq 0$, is given by

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

Independence

Independence

Independence.

Independence

Independence.

Event A and B are **independent** if and only if

Independence

Independence.

Event A and B are **independent** if and only if

$$P(A \cap B) = P(A) \cdot P(B) \text{ or } P(A \mid B) = P(A).$$

Independence

Independence.

Event A and B are **independent** if and only if

$$P(A \cap B) = P(A) \cdot P(B) \text{ or } P(A \mid B) = P(A).$$

The probability of A is not affected by whether or not B occurs.

Independence

Independence.

Event A and B are **independent** if and only if

$$P(A \cap B) = P(A) \cdot P(B) \text{ or } P(A \mid B) = P(A).$$

The probability of A is not affected by whether or not B occurs.

Mutual Independence.

Independence.

Event A and B are **independent** if and only if

$$P(A \cap B) = P(A) \cdot P(B) \text{ or } P(A \mid B) = P(A).$$

The probability of A is not affected by whether or not B occurs.

Mutual Independence.

Events $A_1, \ldots, A_n$ are said to be **mutually independent** if for every subset $I \subseteq \{1, \ldots, n\}$ with size $|I| \geq 2$,

Independence.

Event A and B are **independent** if and only if

$$P(A \cap B) = P(A) \cdot P(B) \text{ or } P(A \mid B) = P(A).$$

The probability of A is not affected by whether or not B occurs.

Mutual Independence.

Events $A_1, \ldots, A_n$ are said to be **mutually independent** if for every subset $I \subseteq \{1, \ldots, n\}$ with size $|I| \geq 2$,

$$P\left(\bigcap_{i \in I} A_i\right) = \prod_{i \in I} P(A_i).$$

Independence

Independence.

Event A and B are **independent** if and only if

$$P(A \cap B) = P(A) \cdot P(B) \text{ or } P(A \mid B) = P(A).$$

The probability of A is not affected by whether or not B occurs.

Mutual Independence.

Events $A_1, \ldots, A_n$ are said to be **mutually independent** if for every subset $I \subseteq \{1, \ldots, n\}$ with size $|I| \geq 2$,

$$P\left(\bigcap_{i \in I} A_i\right) = \prod_{i \in I} P(A_i).$$

Pairwise independence does not imply mutual independence!

Chain Rule

Chain Rule

Finding the probability of the intersection of mutually independent events is easy, but what if they are not mutually independent?

Chain Rule

Finding the probability of the intersection of mutually independent events is easy, but what if they are not mutually independent?

- We use the **chain rule** again but here's the generalized version.

Chain Rule

Finding the probability of the intersection of mutually independent events is easy, but what if they are not mutually independent?

- We use the **chain rule** again but here's the generalized version.

Chain Rule.

Chain Rule

Finding the probability of the intersection of mutually independent events is easy, but what if they are not mutually independent?

- We use the **chain rule** again but here's the generalized version.

Chain Rule. For any events $A_1, \ldots, A_n$,

Chain Rule

Finding the probability of the intersection of mutually independent events is easy, but what if they are not mutually independent?

- We use the **chain rule** again but here's the generalized version.

Chain Rule. For any events $A_1, \ldots, A_n$,

$$P\left(\bigcap_{i=1}^{n} A_i\right) = P(A_1) \cdot P(A_2 \mid A_1) \cdot P(A_3 \mid A_1 \cap A_2) \cdot \cdots \cdot P\left(A_n \middle| \bigcap_{i=1}^{n-1} A_i\right).$$

Union of Events

Union of Events

Sometimes we are also interested in finding the probability of the union of several events.

Union of Events

Sometimes we are also interested in finding the probability of the union of several events.

- If the events are **mutually exclusive**, then

Union of Events

Sometimes we are also interested in finding the probability of the union of several events.

- If the events are **mutually exclusive**, then

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i).$$

Union of Events

Sometimes we are also interested in finding the probability of the union of several events.

- If the events are **mutually exclusive**, then

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i).$$

- Otherwise, we use the **Inclusion-Exclusion Principle**:

Union of Events

Sometimes we are also interested in finding the probability of the union of several events.

- If the events are **mutually exclusive**, then

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i).$$

- Otherwise, we use the **Inclusion-Exclusion Principle**: let $A_1, \ldots, A_n$ be events in some probability space, where $n \geq 2$. Then, we have

Union of Events

Sometimes we are also interested in finding the probability of the union of several events.

- If the events are **mutually exclusive**, then

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i).$$

- Otherwise, we use the **Inclusion-Exclusion Principle**: let $A_1, \ldots, A_n$ be events in some probability space, where $n \geq 2$. Then, we have

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{k=1}^{n} (-1)^{k-1} \sum_{S \subseteq \{1,\ldots,n\}:|S|=k} P\left(\bigcap_{i \in S} A_i\right)$$

Union of Events

Sometimes we are also interested in finding the probability of the union of several events.

- If the events are **mutually exclusive**, then

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i).$$

- Otherwise, we use the **Inclusion-Exclusion Principle**: let $A_1, \ldots, A_n$ be events in some probability space, where $n \geq 2$. Then, we have

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{k=1}^{n} (-1)^{k-1} \sum_{S \subseteq \{1,\ldots,n\} : |S| = k} P\left(\bigcap_{i \in S} A_i\right)$$

$$= \sum_{i=1}^{n} P(A_i) - \sum_{i<j} P(A_i \cap A_j) + \ldots + (-1)^{n-1} P(A_1 \cap A_2 \cap \cdots \cap A_n).$$

Union Bound

Union Bound

Sometimes we are also interested in the upper bound of the probability of the union of several events.

Union Bound

Sometimes we are also interested in the upper bound of the probability of the union of several events.

- We have the **Union Bound**:

Union Bound

Sometimes we are also interested in the upper bound of the probability of the union of several events.

- We have the **Union Bound**: let $A_1, \ldots, A_n$ be events in some probability space. Then

Union Bound

Sometimes we are also interested in the upper bound of the probability of the union of several events.

- We have the **Union Bound**: let $A_1, \ldots, A_n$ be events in some probability space. Then

$$P\left(\bigcup_{i=1}^{n} A_i\right) \leq \sum_{i=1}^{n} P(A_i).$$

Union Bound

Sometimes we are also interested in the upper bound of the probability of the union of several events.

- We have the **Union Bound**: let $A_1, \ldots, A_n$ be events in some probability space. Then

$$P\left(\bigcup_{i=1}^{n} A_i\right) \leq \sum_{i=1}^{n} P(A_i).$$

Summing up all $P(A_i)$ only overestimate the probability of the union (equality holds when they are disjoint).

Problem Time!