EE5110: PROBABILITY FOUNDATIONS FOR ELECTRICAL ENGINEERS

JULY - NOV 2024

1 Conditional Probability

1.1 References

- See Chapter 1: Sample Space and Probability, Bertsekas and Tsitsiklis, for a review of conditional probability.
- See Chapter 3: Conditional Probability and Independence, Sheldon M Ross, for a review of conditional probability.

1.2 Solved Problems

1. Consider a probability space $(\Omega, \mathcal{F}, \mathsf{P})$. Show that the conditional probability $\mathsf{P}(\cdot|A)$, for $A \in \mathcal{F}$ such that $\mathsf{P}(A) > 0$ inspires a valid probability space.

   We need to show that $(\Omega, \mathcal{F}, \mathsf{P}(\cdot|A))$ is a valid probability space, where $A \in \mathcal{F}$, $\mathsf{P}(A) > 0$ and $\mathsf{P}(B|A)$ for $B \in \mathcal{F}$ is defined as $\mathsf{P}(B|A) = \frac{\mathsf{P}(AB)}{\mathsf{P}(A)}$.

   Clearly, Ω is a valid sample space, and $\mathcal{F}$ is a valid event space. In fact, $(\Omega, \mathcal{F})$ remains a valid measurable space independent of the probability measure assigned.

   Hence, we need only to prove that $\mathsf{P}(\cdot|A)$ is a valid probability measure on the measurable space $(\Omega, \mathcal{F})$. (We note that $\mathsf{P}(\cdot|A)$ is well defined when $\mathsf{P}(A) > 0$).

   To show that $\mathsf{P}(\cdot|A)$ is a valid probability measure, we need to validate that the conditional probability satisfies the axioms of probability.

   (a) Non-negativity: $\mathsf{P}(B|A) = \frac{\mathsf{P}(AB)}{\mathsf{P}(A)}$. As $\mathsf{P}(AB) \geq 0$ and $\mathsf{P}(A) > 0$, we have $\mathsf{P}(B|A) \geq 0$.

   (b) Normalized:
$$\mathsf{P}(\Omega|A) = \frac{\mathsf{P}(A\Omega)}{\mathsf{P}(A)} = \frac{\mathsf{P}(A)}{\mathsf{P}(A)} = 1$$

   Hence, the conditional probability measure is normalized.

   (c) Countably additive: Let $B_1, B_2, \cdots$ be a disjoint collection of events in $\mathcal{F}$.
$$\mathsf{P}\left(\bigcup_{i=1}^{\infty} B_i|A\right) = \frac{1}{\mathsf{P}(A)}\mathsf{P}\left(\bigcup_{i=1}^{\infty} AB_i\right)$$

But $AB_1, AB_2, \cdots$ are disjoint as well. Hence, using countable additivity of the probability measure, we have,

$$\begin{aligned}
\mathsf{P}\left(\bigcup_{i=1}^{\infty} B_i | A\right) &= \frac{1}{\mathsf{P}(A)}\mathsf{P}\left(\bigcup_{i=1}^{\infty} AB_i\right) \\
&= \frac{1}{\mathsf{P}(A)}\sum_{i=1}^{\infty}\mathsf{P}(AB_i) \\
&= \sum_{i=1}^{\infty}\frac{\mathsf{P}(AB_i)}{\mathsf{P}(A)} \\
&= \sum_{i=1}^{\infty}\mathsf{P}(B_i|A)
\end{aligned}$$

which is the desired result. □

2. Let $A_1, A_2, \cdots$ be events in $\mathcal{F}$ such that $\mathsf{P}\left(\bigcap_{i=1}^{n} A_i\right) > 0$ for any n. Then, show that

$$\mathsf{P}\left(\bigcap_{i=1}^{\infty} A_i\right) = \prod_{i=1}^{\infty}\mathsf{P}(A_i|A_1 \cap \cdots \cap A_{i-1})$$

Let $A_1, A_2, \cdots$ be events in $\mathcal{F}$. From the continuity theorem, we have,

$$\mathsf{P}\left(\bigcap_{i=1}^{\infty} A_i\right) = \lim_{n\to\infty}\mathsf{P}\left(\bigcap_{i=1}^{n} A_i\right)$$

Using the multiplicative rule of the conditional probabilities for any finite n, we have,

$$\begin{aligned}
\mathsf{P}\left(\bigcap_{i=1}^{\infty} A_i\right) &= \lim_{n\to\infty}\mathsf{P}\left(\bigcap_{i=1}^{n} A_i\right) \\
&= \lim_{n\to\infty}\prod_{i=1}^{n}\mathsf{P}(A_i|A_1\cdots A_{i-1}) \\
&= \prod_{i=1}^{\infty}\mathsf{P}(A_i|A_1\cdots A_{i-1})
\end{aligned}$$

where the last equality identifies the limit of the monotone (bounded) infinite sequence $\{\prod_{i=1}^{n}\mathsf{P}(A_i|A_1\cdots A_{i-1})\}$. □

3. Let $A_1, A_2, \cdots$ be a partition of Ω such that $\mathsf{P}(A_i) > 0$ for all i. Then, show that, for any $B \in \mathcal{F}$

$$\mathsf{P}(B) = \sum_{i=1}^{\infty}\mathsf{P}(B|A_i)\mathsf{P}(A_i)$$

The result follows easily from finite additivity.

$$
\begin{aligned}
\mathsf{P}(B) = \mathsf{P}\left(B \cap \bigcup_{i=1}^{\infty} A_i\right) &= \mathsf{P}\left(\bigcup_{i=1}^{\infty} BA_i\right) \\
&= \sum_{i=1}^{\infty} \mathsf{P}(BA_i) \\
&= \sum_{i=1}^{\infty} \mathsf{P}(A_i)\mathsf{P}(B|A_i)
\end{aligned}
$$

□

4. (Ross) A student is taking a one-hour-time-limit makeup examination. Suppose the probability that the student will finish the exam in less than x hours is $\frac{x}{2}$ for all $0 \leq x \leq 1$. Then, given that the student is still working after 3/4 hour, what is the conditional probability that the full hour is used?

   Let us assume that an appropriate probability space $(\Omega, \mathcal{F}, \mathsf{P})$ exists.

   Let A_x denote the event that the student will finish the exam in less than x hours. For $0 \leq x \leq 1$, we know that $\mathsf{P}(A_x) = \frac{x}{2}$.

   Let B denote the event that the student is still working after 0.75 hour. Note that $B = A_{0.75}^c$. We need to compute the probability of the event that the full hour is used for the exam given the knowledge of occurrence of the event B, i.e., $\mathsf{P}(A_1^c|B) = 1 - \mathsf{P}(A_1|B)$.

$$
\mathsf{P}(A_1|B) = \frac{\mathsf{P}(BA_1)}{\mathsf{P}(B)}
$$

   We note that $BA_1 = A_1 - A_{0.75}$ and $\mathsf{P}(BA_1) = \mathsf{P}(A_1) - \mathsf{P}(A_{0.75})$ (since $A_{0.75} \subset A_1$). From the description, we have $\mathsf{P}(A_1) = 0.5$ and $\mathsf{P}(A_{0.75}) = \frac{3}{8}$. Hence,

$$
\mathsf{P}(BA_1) = \frac{1}{2} - \frac{3}{8} = \frac{1}{8}
$$

   Also, $\mathsf{P}(B) = 1 - \mathsf{P}(A_{0.75}) = \frac{5}{8}$. Now,

$$
\mathsf{P}(A_1|B) = \frac{\mathsf{P}(BA_1)}{\mathsf{P}(B)} = \frac{\frac{1}{8}}{\frac{5}{8}} = \frac{1}{5}
$$

   Then, $1 - \mathsf{P}(A_1|B) = \frac{4}{5}$. □

1.3 Practice Problems

1. (Bertsekas and Tsitsiklis) We are given three coins: one has heads in both faces, the second has tails in both faces, and the third has a head in one face and a tail in the other. We choose a coin at random, toss it, and the result is heads. What is the probability that the opposite face is tails?

2. (Bertsekas and Tsitsiklis) A batch of one hundred items is inspected by testing four randomly selected items. If one of the four is defective, the batch is rejected. What is the probability that the batch is accepted if it contains five defectives?

3. State and prove the conditional version of the total probability theorem.

4. (Grimmett and Stirzaker) Assume that $\mathsf{P}(A)\mathsf{P}(B) > 0$. Show that if $\mathsf{P}(A|B) > \mathsf{P}(A)$, then $\mathsf{P}(B|A) > \mathsf{P}(B)$.

5. (Grimmett and Stirzaker) There are n urns of which the rth contains $r-1$ red balls and $n-r$ magenta balls. You pick an urn at random and remove two balls at random without replacement. Find the probability that:

   (a) the second ball is magenta;

   (b) the second ball is magenta, given that the first is magenta.

6. (Ross) Show that if $\mathsf{P}(A) > 0$, then $\mathsf{P}(AB|A) \geq \mathsf{P}(AB|A \cup B)$.

1.4 Associated Problems

1. Read about Monty Hall Problem (see Wikipedia for a detailed discussion). Can you write a tree representation for the Monty Hall problem identifying the probabilities along the path.

2. Read about Bertrand's paradox (see Wikipedia for a detailed discussion).

2 Independence

2.1 References

- See Chapter 1: Sample Space and Probability, Bertsekas and Tsitsiklis, for a review of independence.

- See Chapter 3: Conditional Probability and Independence, Sheldon M Ross, for a review of independence.

2.2 Solved Problems

1. Let A and B be events in an event space. Show that the two definitions of independence are equivalent if $\mathsf{P}(A) > 0$.

   Let $\mathsf{P}(AB) = \mathsf{P}(A)\mathsf{P}(B)$. Given $\mathsf{P}(A) > 0$. Hence, we can define the conditional probability $\mathsf{P}(B|A)$ as

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

   Let us assume that $\mathsf{P}(A) > 0$ and $\mathsf{P}(B|A) = \mathsf{P}(B)$. Then,

   $$\mathsf{P}(AB) = \mathsf{P}(B|A)\mathsf{P}(A) = \mathsf{P}(B)\mathsf{P}(A)$$

   Thus, the two definitions are equivalent if $\mathsf{P}(A) > 0$. $\square$

2. Let A and B be events. If A is independent of B, then A is independent of B^c.

   Given A is independent of B, i.e., $\mathsf{P}(AB) = \mathsf{P}(A)\mathsf{P}(B)$.

   Using the partition B, B^c, we can write $\mathsf{P}(A)$ as follows.

   $$\mathsf{P}(A) = \mathsf{P}(AB) + \mathsf{P}(AB^c) = \mathsf{P}(A)\mathsf{P}(B) + \mathsf{P}(AB^c)$$

   Rearranging the terms, we get,

   $$\mathsf{P}(AB^c) = \mathsf{P}(A) - \mathsf{P}(A)\mathsf{P}(B) = \mathsf{P}(A)(1 - \mathsf{P}(B)) = \mathsf{P}(A)\mathsf{P}(B^c)$$

   $\square$

3. Let A, B and C be events in an event space. Suppose that $\mathsf{P}(C) > 0$. Show that the two definitions of conditional independence are equivalent if $\mathsf{P}(A|C) > 0$.

   Let $\mathsf{P}(AB|C) = \mathsf{P}(A|C)\mathsf{P}(B|C)$. Given $\mathsf{P}(A|C) > 0$. We will now show that $\mathsf{P}(B|AC) = \mathsf{P}(B|C)$. $\mathsf{P}(B|AC)$ is well defined as $\mathsf{P}(AC) = \mathsf{P}(C)\mathsf{P}(A|C) > 0$.

   $$\mathsf{P}(B|AC) = \frac{\mathsf{P}(BAC)}{\mathsf{P}(AC)} = \frac{\mathsf{P}(C)\mathsf{P}(AB|C)}{\mathsf{P}(C)\mathsf{P}(A|C)}$$

   Substituting $\mathsf{P}(AB|C) = \mathsf{P}(A|C)\mathsf{P}(B|C)$, we get,

   $$\mathsf{P}(B|AC) = \frac{\mathsf{P}(C)\mathsf{P}(AB|C)}{\mathsf{P}(C)\mathsf{P}(A|C)} = \frac{\mathsf{P}(C)\mathsf{P}(A|C)\mathsf{P}(B|C)}{\mathsf{P}(C)\mathsf{P}(A|C)} = \mathsf{P}(B|C)$$

   Now, let us assume that $\mathsf{P}(A|C) > 0$ and $\mathsf{P}(B|AC) = \mathsf{P}(B|C)$. Then,

   $$\mathsf{P}(AB|C) = \frac{\mathsf{P}(ABC)}{\mathsf{P}(C)} = \frac{\mathsf{P}(C)\mathsf{P}(A|C)\mathsf{P}(B|AC)}{\mathsf{P}(C)} = \mathsf{P}(A|C)\mathsf{P}(B|AC)$$

We can define $\mathsf{P}(B|AC)$ as $\mathsf{P}(AC) = \mathsf{P}(C)\mathsf{P}(A|C) > 0$. Now, using $\mathsf{P}(B|AC) = \mathsf{P}(B|C)$, we get,

$$\mathsf{P}(AB|C) = \mathsf{P}(A|C)\mathsf{P}(B|AC) = \mathsf{P}(A|C)\mathsf{P}(B|C)$$

$\square$

4. Show that if $A_1, A_2, \cdots, A_n$ are independent events, then

$$\mathsf{P}\left(\bigcup_{i=1}^{n} A_i\right) = 1 - \prod_{i=1}^{n}(1 - \mathsf{P}(A_i))$$

Let $A_1, \cdots, A_n$ be mutually independent events. Then, $A_1^c, \cdots, A_n^c$ are mutually independent as well. Now,

$$\mathsf{P}(A_1^c A_2^c \cdots A_n^c) = \prod_{i=1}^{n} \mathsf{P}(A_i^c)$$

and,

$$1 - \mathsf{P}(A_1^c A_2^c \cdots A_n^c) = 1 - \prod_{i=1}^{n} \mathsf{P}(A_i^c) = 1 - \prod_{i=1}^{n}(1 - \mathsf{P}(A_i))$$

Now, $1 - \mathsf{P}(A_1^c A_2^c \cdots A_n^c) = \mathsf{P}((A_1^c \cdots A_n^c)^c)$ and from DeMorgan's law we know that $(A_1^c \cdots A_n^c)^c = \bigcup_{i=1}^{n} A_i$. Hence,

$$\mathsf{P}\left(\bigcup_{i=1}^{n} A_i\right) = 1 - \prod_{i=1}^{n}(1 - \mathsf{P}(A_i))$$

$\square$

2.3 Practice Problems

1. Consider 3 (mutually) independent tosses of a fair coin (i.e., the occurrence of Head (or Tail) in the three tosses are mutually independent). Describe the sample space and the probabilities on the elementary outcomes using a tree representation for the experiment.

2. Consider 2 (mutually) independent tosses of a fair coin. How many independent events A_i can you describe such that $\mathsf{P}(A_i) = 0.5$ for all i?

3. (Bertsekas and Tsitsiklis) A computer network connects two nodes A and B through intermediate nodes as shown in Figure 3. For every pair of directly connected nodes, say i and j, there is a probability that the link between them is up, p_{ij}. Assume that the link failures are independent of each other. What is the probability that there is a path connecting A and B in which all the links are up?

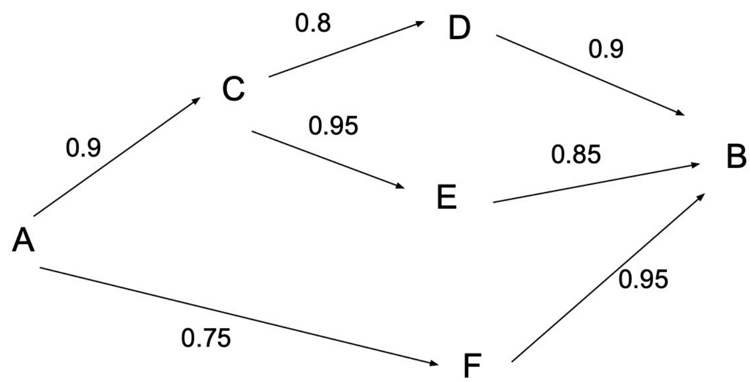

Figure 1: A computer network.

4. (Bertsekas and Tsitsiklis) Communication through a noisy channel: A source transmits a string of symbols through a noisy channel. Each symbol is 0 or 1 with probability $1-p$ and p, respectively. The transmitted symbols 0 and 1 are correctly received over the channel with probability $1-\epsilon_0$ and $1-\epsilon_1$, respectively. Assume that the errors in different symbol transmissions are independent.

   (a) What is the probability that the kth symbol is received correctly?

   (b) What is the probability that the string of symbols 1101 is received correctly?

   (c) Suppose you transmit every symbol three times. Then, what is the probability that the kth symbol is received correctly if the receiver uses the majority decision to decode the symbol?

5. (Bertsekas and Tsitsiklis) How do you use a biased coin to make an unbiased (fair) decision?

2.4 Associated Problems

1. (Bertsekas and Tsitsiklis) Gambler's Ruin: A gambler makes a sequence of independent bets. In each bet, he wins 1 dollar with probability p and loses 1 dollar with probability $1-p$. Initially, the gambler has k dollars, and plays until he either accumulates n dollars or has no money left. What is the probability that the gambler will end up with n dollars?

2. The Borel-Cantelli Lemma: Let $A_1, A_2, \cdots$ be an infinite collection of events.

(a) Assume that $\sum_{i=1}^{\infty} \mathsf{P}(A_i) < \infty$. Show that the event an infinite number of events $\{A_i\}$ occur with probability zero, i.e., $\mathsf{P}\left(\bigcap_{i=1}^{\infty}\bigcup_{j=i}^{\infty} A_j\right) = 0$.

(b) Assume that the events are independent and that $\sum_{i=1}^{\infty} \mathsf{P}(A_i) = \infty$. Show that an infinite number of events $\{A_i\}$ occur with probability one, i.e., $\mathsf{P}\left(\bigcap_{i=1}^{\infty}\bigcup_{j=i}^{\infty} A_j\right) = 1$.

3. Consider the uniform probability measure on $[0, 1]$. How many (mutually) independent events A_i can you describe such that $\mathsf{P}(A_i) = 0.5$ for all i?