Lecture 9: Conditional Probability and Independence

Lecturer: Dr. Krishna Jagannathan *Scribe: Vishakh Hegde*

9.1 Conditional Probability

Definition 9.1 *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Let $B \in \mathcal{F}$ such that $\mathbb{P}(B) > 0$. Then the* conditional probability of A given B *is defined as,*

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

Caution: We cannot condition on sets of zero probability measure. For example, if $\Omega = [0,1]$ endowed with the Borel σ-algebra and a uniform probability measure, we cannot condition on the set of rationals.

Theorem 9.2 *Let $B \in \mathcal{F}$ and $\mathbb{P}(B) > 0$. Then, $\mathbb{P}(\cdot\ |B) : \mathcal{F} \longmapsto [0,1]$ is a probability measure on $(\Omega, \mathcal{F})$.*

Proof: We need to show that the three properties of probability measure holds true, namely:

- $\mathbb{P}(\Omega|B) = 1$.
- $\mathbb{P}(\phi|B) = 0$.
- Countable additivity property.

We have,

$$\mathbb{P}(\Omega|B) = \frac{\mathbb{P}(\Omega \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(B)}{\mathbb{P}(B)} = 1.$$

$$\mathbb{P}(\phi|B) = \frac{\mathbb{P}(\phi \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(\phi)}{\mathbb{P}(B)} = 0.$$

We are now left with proving countable additivity property. Let $A_1, A_2, \ldots$ be disjoint. We need to show that,

$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i | B\right) = \sum_{i=1}^{\infty} \mathbb{P}(A_i|B).$$

Consider,

$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i | B\right) = \frac{\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i \cap B\right)}{\mathbb{P}(B)} = \frac{\mathbb{P}\left(\bigcup_{i=1}^{\infty} (A_i \cap B)\right)}{\mathbb{P}(B)}.$$

Since A_i are disjoint, $A_i \cap B$ are also disjoint. Therefore we can write the following:

$$\frac{\mathbb{P}\left(\bigcup_{i=1}^{\infty} (A_i \cap B)\right)}{\mathbb{P}(B)} = \sum_{i=1}^{\infty} \frac{\mathbb{P}(A_i \cap B)}{\mathbb{P}(B)} = \sum_{i=1}^{\infty} \mathbb{P}(A_i|B).$$

■

9.1.1 Properties of Conditional Probability

1. *The Law of Total Probability:* Let $A \in \mathcal{F}$ and let $\{B_i,\ i = 1, 2, \ldots\}$ be events that partition Ω (by partition we mean $\bigcup_{i\in\mathbb{N}} B_i = \Omega$ and $B_i \cap B_j = \phi$, $\forall i \neq j$), with $\mathbb{P}(B_i) > 0$, $\forall i$. Then,

$$\mathbb{P}(A) = \sum_{i=1}^{\infty} \mathbb{P}(A|B_i)\,\mathbb{P}(B_i).$$

Proof: We know that $\{B_i,\ i = 1, 2, \ldots\}$ partitions Ω. Hence $\{A \cap B_i,\ i = 1, 2, \ldots\}$ partitions A. Therefore, by the countable additivity property, we have

$$\mathbb{P}(A) = \mathbb{P}\left(\bigcup_{i=1}^{\infty} (A \cap B_i)\right) = \sum_{i=1}^{\infty} \mathbb{P}(A \cap B_i).$$

and $\mathbb{P}(A \cap B_i) = \mathbb{P}(A|B_i)\,\mathbb{P}(B_i)$, $\forall i$. Therefore,

$$\sum_{i=1}^{\infty} \mathbb{P}(A \cap B_i) = \sum_{i=1}^{\infty} \mathbb{P}(A|B_i)\,\mathbb{P}(B_i).$$

■

Note: In particular, if B is such that $0 < \mathbb{P}(B) < 1$, then,

$$\mathbb{P}(A) = \mathbb{P}(A|B)\,\mathbb{P}(B) + \mathbb{P}(A|B^c)\,\mathbb{P}(B^c).$$

2. *Bayes' Rule:* Let $A \in \mathcal{F}$, with $\mathbb{P}(A) > 0$ and let $\{B_i,\ i = 1, 2, \ldots\}$ be a partition of Ω such that $\mathbb{P}(B_i) > 0\ \forall i$. Then, we have,

$$\mathbb{P}(B_i|A) = \frac{\mathbb{P}(A|B_i)\,\mathbb{P}(B_i)}{\sum_{j=1}^{\infty} \mathbb{P}(A|B_j)\,\mathbb{P}(B_j)}.$$

Proof:

$$\mathbb{P}(B_i|A) = \frac{\mathbb{P}(A \cap B_i)}{\mathbb{P}(A)} = \frac{\mathbb{P}(B_i)\,\mathbb{P}(A|B_i)}{\mathbb{P}(A)} = \frac{\mathbb{P}(A|B_i)\,\mathbb{P}(B_i)}{\sum_{j=1}^{\infty} \mathbb{P}(A|B_j)\,\mathbb{P}(B_j)}.$$

■

3. For any sequence of events $\{A_i\}$, we have the following relation:

$$\mathbb{P}\left(\bigcap_{i=1}^{\infty} A_i\right) = \mathbb{P}(A_1) \prod_{i=2}^{\infty} \mathbb{P}(A_i|A_1 \cap A_2 \cap \ldots \cap A_{i-1}).$$

as long as all the conditional probabilities are well defined.
Proof: We know that the following holds for finite set of events:

$$\mathbb{P}\left(\bigcap_{i=1}^{n} A_i\right) = \mathbb{P}(A_1) \prod_{i=2}^{n} \mathbb{P}(A_i|A_1 \cap A_2 \cap \ldots \cap A_{i-1}).$$

Now taking limits, we have:

$$\lim_{n\to\infty} \mathbb{P}\left(\bigcap_{i=1}^{n} A_i\right) = \lim_{n\to\infty} \mathbb{P}(A_1) \prod_{i=2}^{n} \mathbb{P}(A_i|A_1 \cap A_2 \cap \ldots \cap A_{i-1}).$$

Now using continuity of probability, we get the required relation,

$$\mathbb{P}\left(\bigcap_{i=1}^{\infty} A_i\right) = \mathbb{P}(A_1) \prod_{i=2}^{\infty} \mathbb{P}(A_i|A_1 \cap A_2 \cap \ldots \cap A_{i-1}).$$

■

9.2 Independence

Definition 9.3 *Let* $(\Omega, \mathcal{F}, \mathbb{P})$ *be a probability space. Two events* A *and* B *are said to be* independent *(under the probability measure* $\mathbb{P}$*) if* $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$*.*

Note: If $\mathbb{P}(B) > 0$ and, A and B are independent, then we have,

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

Example: Can disjoint sets be independent at all? Let $A, B \in \mathcal{F}$ be two disjoint sets. Therefore, we have $\mathbb{P}(A \cap B) = \mathbb{P}(\phi)$. This means that $\mathbb{P}(A \cap B) = 0$. For independence, we need to have $\mathbb{P}(A)\mathbb{P}(B) = \mathbb{P}(A \cap B) = 0$. This can happen when $\mathbb{P}(A) = 0$ or $\mathbb{P}(B) = 0$. Therefore, in general, two disjoint events are independent if and only if at least one of them has zero probability.

Definition 9.4 $A_1, A_2, \ldots, A_n$ *are independent if for all non-empty* $I_0 \subseteq \{1, 2, \ldots, n\}$*, we have,*

$$\mathbb{P}\left(\bigcap_{i \in I_0} A_i\right) = \prod_{i \in I_0} \mathbb{P}(A_i).$$

Next, we define independence of an arbitrary collection of events.

Definition 9.5 $\{A_i, i \in I\}$ *are said to be independent if for every non-empty finite subset* I_0 *of* I*, we have*

$$\mathbb{P}\left(\bigcap_{i \in I_0} A_i\right) = \prod_{i \in I_0} \mathbb{P}(A_i).$$

9.2.1 Independence of σ-algebras

Definition 9.6 *Let* $\mathcal{F}_1$ *and* $\mathcal{F}_2$ *be two sub-*σ*-algebras of* $\mathcal{F}$*. We say that* $\mathcal{F}_1$ *and* $\mathcal{F}_2$ *are independent* σ*-algebras if for all* $A_1 \in \mathcal{F}_1$ *and* $A_2 \in \mathcal{F}_2$*,* A_1 *and* A_2 *are independent events.*

Example: A simple example we can construct is the following: Let $A, B \in \mathcal{F}$, then $\mathcal{F}_1 = \{\phi, \Omega, A, A^c\}$ and $\mathcal{F}_2 = \{\phi, \Omega, B, B^c\}$ are independent iff A and B are independent.

We now define independence on a collection of sub-σ algebras.

Definition 9.7 *Let* $\{\mathcal{F}_i, i \in I\}$ *(where* I *is an index set) be a collection of sub* σ *algebras of* $\mathcal{F}$*. Then,* $\{\mathcal{F}_i, i \in I\}$ *are said to be independent if for every choice of* $A_i \in \mathcal{F}_i$*, we have* $\{A_i, i \in I\}$ *are independent.*

Example (from [Lecture 2, MIT OCW]): Consider the infinite coin toss model discussed previously.

- Let A_i be the event that the i^{th} coin toss resulted in heads (say). If $i \neq j$, the events A_i and A_j are independent.
- The following infinite family of events are independent: $\{A_i | i \in \mathbb{N}\}$. This example captures the intuitive idea of independent coin tosses.

- Let $\mathcal{F}_1$ (respectively, $\mathcal{F}_2$) be the collection of all events whose occurrence can be decided by looking at the results of the coin toss at odd times (respectively, at even times) n. More formally, let H_i be the event that the i^{th} toss resulted in heads. Let $\mathcal{C} = \{H_i \mid i \text{ is odd}\}$ and let $\mathcal{F}_1 = \sigma(\mathcal{C})$, so that $\mathcal{F}_1$ is the smallest σ-algebra that contains all the events H_i, for odd i. We define $\mathcal{F}_2$ similarly, using even times instead of odd times. Then, the two σ-algebras $\mathcal{F}_1$ and $\mathcal{F}_2$ turn out to be independent. Intuitively, this implies that any event whose occurrence is determined completely by the outcomes of the tosses at odd times, is independent of any event whose occurrence is determined completely by the outcomes of the tosses at even times.
- Let $\mathcal{F}_n$ be the collection of all events whose occurrence can be decided by looking at the coin tosses $2n$ and $2n+1$. We know that $\mathcal{F}_n$ is a σ-algebra with finitely many events $\forall n \in \mathbb{N}$. It turns out that $\{\mathcal{F}_n, n \in \mathbb{N}\}$ are independent.

9.3 Exercises

1. (a) Let C, $C \in \mathcal{F}$, where $\mathcal{F}$ is a sigma algebra on Ω. Show that $\mathcal{F}_1 = \{\phi, \Omega, C, C^c\}$ and $\mathcal{F}_2 = \{\phi, \Omega, D, D^c\}$ are independent iff C and D are independent.

   (b) Let $\Omega = \{1, 2, 3, ...p\}$ where p is a prime, $\mathcal{F}$ be the collection of all subsets of Ω, and $\mathbb{P}(A) = \frac{|A|}{p}$ (where $|A|$ denotes cardinality of A) for all $A \in \mathcal{F}$. Show that, if A and B are independent events, then at least one of A and B is either ϕ or Ω.

2. In a box, there are four red balls, six red cubes, six blue balls and an unknown number of blue cubes. When an object from the box is selected at random, the shape and colour of the object are independent. Determine the number of blue cubes.

3. A man is known to speak the truth 3 out of 4 times. He throws a die and reports that it is a six. Find the probability that it is actually a six.

4. *[Exercise: Q29, Bertsekas & Tsitsiklis]* Let A and B be events such that $P(A|B) > P(A)$. Show that $P(B|A) > P(B)$ and $P(A|B^c) < P(A)$.

5. *[MIT OCW Assignment problem]* A coin is tossed independently n times. The probability of heads at each toss is p. At each time k $(k = 2, 3, ..., n)$ we get a reward at time $k+1$ if k^{th} toss was a head and the previous toss was a tail. Let A_k be the event that a reward is obtained at time k.

   a) Are events A_k and A_{k+1} independent?

   b) Are events A_k and A_{k+2} independent?

6. *[Assignment problem, University of Cambridge]* A drawer contains two coins. One is an unbiased coin, which when tossed, is equally likely to turn up heads or tails. The other is a biased coin, which will turn up heads with probability p and tails with probability $1-p$. One coin is selected (uniformly) at random from the drawer. Two experiments are performed:

   a) The selected coin is tossed n times. Given that the coin turns up heads k times and tails $n-k$ times, what is the probability that the coin is biased?

   b) The selected coin is toss repeatedly until it turns up heads k times. Given that the coin is tossed n times in total, what is the probability that the coin is biased?

7. *[MIT OCW Assignment problem]* Fred is giving out samples of dog food. He makes calls door to door, but he leaves a sample (one can) only on those calls for which the door is answered and a dog is in residence. On any call the probability of the door being answered is 3/4, and the probability that any household has a dog is 2/3. Assume that the events "door answered" and "a dog lives here" are independent and also that the outcomes of all calls are independent.

a) Determine the probability that Fred gives away his first sample on his third call.

b) Given that he has given away exactly four samples on his first eight calls, determine the conditional probability that Fred will give away his fifth sample on his eleventh call.

c) Determine the probability that he gives away his second sample on his fifth call.

d) Given that he did not give away his second sample on his second call, determine the conditional probability that he will leave his second sample on his fifth call.

e) We will say that Fred needs a new supply immediately after the call on which he gives away his last can. If he starts out with two cans, determine the probability that he completes at least five calls before he needs a new supply.

8. *[MIT OCW Assignment problem]* Let A, B, A_1, A_2, ... be events. Suppose that for each k, we have $A_k \subseteq A_{k+1}$, and that A_k is independent of B, $\forall k \geq 1$. If $A = \cup_{k\in\mathbb{N}} A_k$, then show that B is independent of A.

9. *[Assignment problem University of Cambridge]* Consider pairwise disjoint events B_1, B_2, B_3 and C, with $P(B_1) = P(B_2) = P(B_3) = p$ and $P(C) = q$, where $3p + q \leq 1$. Suppose $p = -q + \sqrt{q}$, then prove that the events $B_1 \cup C$, $B_1 \cup C$ and $B_1 \cup C$ are pairwise independent. Also, prove or disprove that there exist $p > 0$ and $q > 0$ such that these three events are independent.

References

[1] MIT OCW - 6.436J / 15.085J Fundamentals of Probability, Fall 2008, Lecture 2.