Probability

Jan-April 2014

Chennai Mathematical Institute

Professor Nandini Kannan

Chapter 2

Conditional Probability and Independence

So far, we have computed probabilities of events assuming no information is available about the experiment other than the sample space. In many examples, some partial information may be available. This information allows us to "update" the sample space, conditional on this information.

Conditional probability may be viewed in terms of an observer in possession of partial information. A probability space $(\mathcal{S}, \mathcal{B}, P)$ describes an experiment, governed by chance, which produces a result s distributed according to P; $P(A)$ = probability that the point $s \in A$.
Suppose we know that s lies in B. Using this partial information, the probability that s also lies in A represents the conditional probability of event A occurring given event B has occurred. We denote this probability by $P(A|B)$.
The fact that B has occurred restricts the sample space to B. The only elements of A that are of interest are those that are also elements of B, i.e. the elements of $A \cap B$.
This leads to the definition of the function $P(A|B)$:

<u>Definition:</u> Let A and B be two events defined on the sample space $\mathcal{S}$. Then

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \qquad \text{provided} \quad P(B) > 0.$$

For any B for which $P(B) > 0$, we can show that $P(.|B)$ is a probability function, i.e. it satisfies the three axioms.

Theorem 2.0.1. *$P(.|B)$ is a probability function for all $B \in \mathcal{B}$, such that $P(B) > 0$.*

<u>Proof:</u>

1. $P(A|B) = \frac{P(A \cap B)}{P(B)} \geq 0 \qquad \forall A \in \mathcal{F}.$

2. $P(\mathcal{S}|B) = \frac{P(\mathcal{S}\cap B)}{P(B)} = \frac{P(B)}{P(B)} = 1$

3. If A_i's are disjoint,

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

■

The original probability space $(\mathcal{S}, \mathcal{B}, P)$ is converted to $(B, \mathcal{B} \cap B, P(.|B))$.

Does $\mathcal{B} \cap B$ look familiar?

Example: Bowl I contains 1 white and 2 black chips, Bowl II contains 1 black and 2 white chips. A fair coin is tossed. If the coin turns up heads, a chip is drawn from Bowl I; otherwise a chip is drawn from Bowl II. The chip drawn was black. Find the probability that the coin turned up heads.

Solution: The sample space is given by

$$\mathcal{S} = \{HB_1^I, HB_2^I, HW_1^I, TB_1^{II}, TW_1^{II}, TW_2^{II}\}.$$

Let H denote the event that a head occurs, and B denote the event that the chip is black. We need $P(H|B)$.
What is $P(H \cap B)$? This is the event that a head occurs and a black chip was drawn.

$$H \cap B = \{HB_1^I, HB_2^I\}.$$

Therefore $P(H \cap B) = 2/6$. $P(B) = 3/6$. Therefore

$$P(H|B) = 2/3.$$

■

Example: An urn contains r red and b blue marbles. Two marbles are drawn at random from the urn without replacement. Find the probability that the first marble is red and the second is blue.

Solution: Let A be the event that the first marble drawn is red, and B the event that the second marble is blue. We need to find $P(A \cap B)$. We have

$$P(A) = \frac{r}{r+b}; \qquad P(B|A) = \frac{b}{r+b-1}.$$

We know that

$$P(B|A) = \frac{P(A \cap B)}{P(B)} \Rightarrow P(A \cap B) = \left(\frac{b}{r+b-1}\right)\left(\frac{r}{r+b}\right).$$

■

Example: Bowl I contains 6 red chips and 4 blue chips. Five of these 10 chips are selected at random and without replacement and put in Bowl II, which was originally empty. Given that this chip is blue, find the conditional probability that 2 red chips and 3 blue chips are transferred from bowl I to bowl II.

Solution: Let B be the event that a blue chip was drawn. Let A be the event that 2 red and 3 blue chips are transferred. We need $P(A|B)$.

■

Result: Multiplication Rule. Let A and B be two events defined on the sample space $\mathcal{S}$. Then

$$P(A \cap B) = P(A|B) \times P(B) = P(B|A) \times P(A).$$

This follows from the definition of conditional probability.

Example: Game of Craps. A player rolls 2 dice and observes the sum of the two faces.

- If the sum is 7 or 11, the player wins.
- If the sum is 2,3, or 12, the player loses.
- If sum is between 4 and 10 (inclusive), the dice are rolled again and again until the

sum is either 7 or the original value. If the 7 appears first, the player loses; if the initial outcome reoccurs before the 7, the player wins

Find the probability the player wins.

Solution: Let P_0 be the probability of winning on the 1st roll. Then

$$P_0 = P(7 \text{ or } 11) = P(7) + P(11) = \frac{8}{36}.$$

Suppose the sum is 4. Then

$$\begin{aligned} P(\text{Player wins}) &= P(4 \text{ appears before a } 7) \\ &= \frac{3}{36} + \left(\frac{27}{36}\right)\left(\frac{3}{36}\right) + \left(\frac{27}{36}\right)^2\left(\frac{3}{36}\right) + \ldots \\ &= \frac{3}{36}\left[1 + \frac{27}{36} + \left(\frac{27}{36}\right)^2 + \ldots\right] \\ &= \frac{3/36}{1-\frac{27}{36}} = \frac{1}{3}. \end{aligned}$$

Therefore

$$P(\text{Sum } = 4 \text{ and winning}) = \left(\frac{3}{36}\right)\left(\frac{1}{3}\right).$$

Now, suppose the sum is 5. Repeat the same process to show that

$$P(5 \text{ appears before a } 7) = \frac{4}{10}.$$

Therefore

$$P(\text{Sum } = 5 \text{ and winning}) = \left(\frac{4}{36}\right)\left(\frac{4}{10}\right).$$

$$P(\text{Sum } = 6 \text{ and winning}) = \left(\frac{5}{36}\right)\left(\frac{5}{11}\right).$$

$$P(\text{Sum } = 8 \text{ and winning}) = \left(\frac{5}{36}\right)\left(\frac{5}{11}\right).$$

$$P(\text{Sum } = 9 \text{ and winning}) = \left(\frac{4}{36}\right)\left(\frac{4}{10}\right).$$

$$P(\text{Sum } = 10 \text{ and winning}) = \left(\frac{3}{36}\right)\left(\frac{1}{3}\right).$$

The probability of winning is the sum of all these probabilities and is equal to **0.493**. ■

Example: The Monty Hall problem. The problem is named for its similarity to the Let's Make a Deal TV game show hosted by Monty Hall. The problem is stated as follows. Assume that a room is equipped with three doors. Behind two of the doors are goats, and behind the third is a new car. Monty Hall knows what is behind all three doors.

The contestant is asked to pick a door, and will win whatever is behind it. Before the door is opened, however, Monty Hall opens one of the other two doors, revealing a goat, and asks the contestant if she wishes to change her selection to the third door (i.e., the door which neither she picked nor he opened).
What is the probability that the contestant wins if (a) she sticks with her first choice; or (b) decides to switch?

Solution: Assume there is a winning door, and that the remaining doors A and B have goats behind them.
There are three options:

1. Contestant chooses door with the car behind it. She is shown either A or B. If she switches, she loses. If she stays with the original choice, she wins,

2. Contestant chooses A. Monty opens door B. If she switches she wins, otherwise she loses.

3. Contestant chooses B. Monty opens door A. If she switches she wins, otherwise she loses.

Each of the above three options has a 1/3 probability of occurring.
If the contestant switches, then

$$P(\text{winning}) = 0 \times \frac{1}{3} + 1 \times \frac{1}{3} + 1 \times \frac{1}{3} = \frac{2}{3}.$$

If the contestant does not switch, then

$$P(\text{winning}) = 1 \times \frac{1}{3} + 0 \times \frac{1}{3} + 0 \times \frac{1}{3} = \frac{1}{3}.$$

■

Variations of the problem exist: One famous version is called the Prisoner's Dilemma.

Theorem 2.0.2. Bayes' Theorem. *Let $A_1, A_2, \ldots$ be a partition of $\mathcal{S}$, and let B be any event.*

$$P[A_i|B] = \frac{P(B|A_i)P(A_i)}{\sum_{j=1}^{\infty} P(B|A_j)P(A_j)}. \tag{2.1}$$

Proof: We know that

$$P[A_i|B] = \frac{P(A_i \cap B)}{P(B)}.$$

Use the Multiplication Rule to rewrite the numerator. Using Theorem 1.5.10 (Law of Total Probability), we can simplify the denominator to obtain the result. ■

Each term in Bayes' theorem has a conventional name. The term $P(A)$ is called the **prior probability** of A. It is "prior" in the sense that it precedes any information about the event B. $P(A)$ is also called the **marginal probability** of A.

The term $P(A|B)$ is called the **posterior probability** of A, given B. It is "posterior" in the sense that it is derived from the specified value of B.

Example: An insurance company believes that individuals can be divided into two classes: (a) Accident Prone (A), and (b) Not accident prone (N).

Data shows that an A type person will have an accident at some time within a fixed year period with probability 0.4, whereas the probability decreases to 0.2 for a type N person. We know that 30% of the population is type A.

(i) Find the probability that a new policy holder will have an accident within a year.

(ii) If a new policy holder has an accident within a year, what is the probability they were accident prone?

Solution: Let E be the event that a policy holder has an accident within a year. We have

$$P[A] = 0.3, P[E|A] = 0.4, P[E|N] = 0.2.$$

Then

$$P[E] = P[E \cap A] + P[E \cap N] = P[E|A]P(A) + P[E|N|P(N) = (.4)(.3) + (.2)(.7) = .26.$$

$$P[A|N] = \frac{(.12)}{(.26)}.$$

■

Example: Imagine there is a screening test for liver cancer which exhibits the following properties: 80% of individuals who have liver cancer will have a positive test result. However, 9.6% of individuals who do not have liver cancer will show a false positive test result.. Given that the rate of liver cancer is 1%, what is the probability that an individual who has a positive test result actually has liver cancer?

Solution: Let E be the event that the test is positive for liver cancer. Let C be the event that the individual has cancer. We have

$$P[C] = 0.01, P[E|C] = 0.8, P[E|C^c] = 0.096.$$

Then

$$P[C|E] = \frac{0.8 \times .01}{0.8 \times .01 + .096 \times 0.99} = 0.0776.$$

■

Result: Extension of the Multiplication Rule. Let $A_1, A_2, \ldots, A_n$ be n events defined on the sample space $\mathcal{S}$. Then

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

2.0.1 Independence of Events

If $P(A|B) = P(A)$, then occurrence of B has no effect on A. We say that A and B are independent events.

Definition: Two events A and B are said to be statistically independent or stochastically independent if

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

Theorem 2.0.3. *If A and B are independent events, then*

- *A and B^c are independent.*
- *A^c and B are independent.*
- *A^c and B^c are independent.*

Proof: Part 1. From Theorem 1.5.4 we have

$$\begin{aligned}
&P(A) = P[A \cap B^c] + P(A \cap B) \\
&\Rightarrow P(A) = P[A \cap B^c] + P(A)P(B) \\
&\Rightarrow P[A \cap B^c] = P(A) - P(A)P(B) = P(A)[1 - P(B)] \\
&\Rightarrow P[A \cap B^c] = P(A)P(B^c).
\end{aligned}$$

From the definition, we have A and B^c are independent

■

How do we extend the definition to more than two events? For three events, we may require

$$P(A \cap B \cap C) = P(A)P(B)P(C).$$

Example: Consider 2 rolls of a die. Let A be the event that a double occurs, B be the event that the sum is 7, 8, 9, or 10, and C be the event that the sum is 2, 7, or 8. We have

$$P(A) = \frac{1}{6}; P(B) = \frac{18}{36}; P(C) = \frac{12}{36}.$$

We have

$$P(A \cap B \cap C) = P(\{4, 4\}) = \frac{1}{36} = P(A).P(B).P(C).$$

However,

$$P(B \cap C) = P(7 \text{ or } 8) = \frac{11}{36} \neq P(B).P(C).$$

So the events B and C are not pairwise independent.

■

Suppose we assume pairwise independence.

Example: Consider 2 rolls of a die. Let E be the event that an odd number occurs on roll 1, F be the event that an odd number occurs on roll 2, and G be the event that the sum is odd. We have

$$P(E) = \frac{1}{2}; P(F) = \frac{1}{2}; P(G) = \frac{1}{2}.$$

We have

$$P(E \cap F) = \frac{1}{4}; P(F \cap G) = \frac{1}{4}; P(E \cap G) = \frac{1}{4}.$$

The three events are pairwise independent. However,

$$P(E \cap F \cap G) = 0 \neq P(E).P(F).P(G).$$

■

Definition: The events $A_1, \ldots, A_n$ are said to be **mutually independent** if for any subcollection $A_{i_1}, \ldots, A_{i_k}$, we have

$$P\left(\bigcap_{j=1}^{k} A_{i_j}\right) = \prod_{i=1}^{k} P(A_{i_j}).$$