Section 1.5 Conditional Probabilities

Contents

1. The Definition of Conditional Probability
2. The Multiplication Rule
3. Total Probability Formula
4. Bayes' Theorem

Let us consider a easier problem.

Suppose that we flip two fair coins and that each of the 4 possible outcomes is equally likely to occur and hence has the probability 1/4. Suppose that we observe that a head appears on the flip of one coin. Then given this information, what is the probability that (H, H) occurs? To calculate this probability we reason as follows:

The Definition of Conditional Probability

- Given that a head appears, it follows that there can be three outcomes $(H, H), (H, T)$ and (T, H).
- Since each of these outcomes originally had the same probability of occurring, they should still have equal probabilities.

In fact, this probability can be calculated as follows.

Let $A = \{(H, H)\}$ and $B = \{$ a head appears on the flip of one coin$\}$.

$$P(A|B) = \frac{1}{3} = \frac{\#\{(H,H)\}}{\#\{(H,H),(H,T),(T,H)\}}$$

$$= \frac{\#AB}{\#B} = \frac{\frac{\#AB}{\#S}}{\frac{\#B}{\#S}} = \frac{P(AB)}{P(B)}.$$

The Definition of Conditional Probability

Definition

Given two events A and B with $P(B) > 0$, the ***conditional probability of*** *A* ***given*** *B is defined as the quotient of the joint probability of A and B, and the probability of B:*

$$P(A|B) = \frac{P(AB)}{P(B)} \tag{1}$$

Theorem

If $P(B) > 0$, then the conditional probability $P(A|B)$ is also a probability, that is,

(i) *for every event A,* $P(A|B) \geqslant 0$;

(ii) $P(\Omega|B) = 1$; ☆ what is a probability

The Definition of Conditional Probability

Theorem

(iii) *for every infinite sequence of countable disjoint events* $A_1, A_2, \cdots$,

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

Proof. (i) Since $P(B) \geqslant P(AB)$ and $P(B) > 0$, we get $P(A|B) \geqslant 0$ by equation (1).

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

(iii) $P\Big(\bigcup\limits_{i=1}^{\infty} A_i \Big| B\Big) = \Big(P\big(\bigcup\limits_{i=1}^{\infty} A_i B\big)\Big) \Big/ P(B) = \sum\limits_{i=1}^{\infty} \frac{P(A_i B)}{P(B)} = \sum\limits_{i=1}^{\infty} P(A_i|B)$. □

The Definition of Conditional Probability

Since the conditional probability is a probability, all properties of probabilities hold for conditional probabilities.

Property

If $P(B) > 0$, *then*
(1) $P(\emptyset|B) = 0$.
(2) *For every finite sequence of countable disjoint events* $A_1, A_2, \cdots, A_n$,

$$P\Big(\bigcup_{i=1}^{n} A_i \Big| B\Big) = \sum_{i=1}^{n} P(A_i|B).$$

(3) $P(\overline{A}|B) = 1 - P(A|B)$.
(4) *If* $A \subset C$, *then* $P(C - A|B) = P(C|B) - P(A|B)$ *and* $P(A|B) \leqslant P(C|B)$.
(5) $P(A \cup C|B) = P(A|B) + P(C|B) - P(AC|B)$.

The Definition of Conditional Probability

Example

A machine produces parts that are either good (90%), slightly defective (2%), or obviously defective (8%). Produced parts get passed through an automatic inspection machine, which is able to detect any part that is obviously defective and discard it. What is the quality of the parts that make it through the inspection machine and get shipped?

Solution. Let A (resp., B, C) be the event that a randomly chosen shipped part is good (resp., slightly defective, obviously defective).

We are told that $P(A) = 0.90, P(B) = 0.02$, and $P(C) = 0.08$.

The Definition of Conditional Probability

Solution. Let A (resp., B, C) be the event that a randomly chosen shipped part is good (resp., slightly defective, obviously defective).

We are told that $P(A) = 0.90, P(B) = 0.02$, and $P(C) = 0.08$.

We want to compute the probability that a part is good given that it passed the inspection machine (i.e., it is not obviously defective), which is

$$P(A|\overline{C}) = \frac{P(A\overline{C})}{P(\overline{C})} = \frac{P(A)}{1 - P(C)} = \frac{0.90}{1 - 0.08} = 90/92 = 0.978.$$

$\square$

The Definition of Conditional Probability

Example

Ten fair dice are rolled at one time. What is the conditional probability of the event at least two land on 1 given the event at least one of the dice lands on 1.

Solution. Let A be the event that at least one of the dice lands on 1, B be the event that at least two land on 1, and C be the event that exactly one of the dice lands on 1.
We see that $B \subset A, C \subset A$ and $B = A - C$. By using the definition of classical probability,

$$P(A) = 1 - P(\overline{A}) = 1 - \frac{5^{10}}{6^{10}}, \qquad P(C) = \frac{10 \times 5^9}{6^{10}}.$$

The Definition of Conditional Probability

So the required probability is

$$
\begin{aligned}
P(B|A) =& \frac{P(AB)}{P(A)} = \frac{P(B)}{P(A)} = \frac{P(A) - P(C)}{P(A)} \\
=& \frac{1 - \dfrac{5^{10}}{6^{10}} - \dfrac{10 \times 5^9}{6^{10}}}{1 - \dfrac{5^{10}}{6^{10}}} \approx 0.615.
\end{aligned}
$$

□

The Definition of Conditional Probability

Example

In describing the survival rate and life expectancy in a certain population, let A_N denote the event of a new-born to reach the age of N years. We are given that

$$P(A_{50}) = 0.913, \qquad P(A_{55}) = 0.881, \qquad P(A_{65}) = 0.746.$$

(a) What is the probability of a 50 years old man to reach the age of 55, i.e., what is $P(A_{55}|A_{50})$?
(b) If the probability that a man who just turned 65 will die within 5 years is 0.16, what is the probability for a man to survive till his 70th birthday, i.e., what is $P(A_{70})$?

The Definition of Conditional Probability

Solution. (a) Obviously, $A_{55} \cap A_{50} = A_{55}$. We have by definition,

$$P(A_{55}|A_{50}) = P(A_{55} \cap A_{50})/P(A_{50}) = P(A_{55})/P(A_{50}) \approx 0.965.$$

(b) Similarly, $P(A_{70}|A_{65}) = P(A_{70})/P(A_{65})$. So $P(A_{70}) = P(A_{65}) \cdot P(A_{70}|A_{65})$. We get

$$P(A_{70}|A_{65}) = 1 - 0.16 = 0.84.$$

Therefore,
$P(A_{70}) = P(A_{65}) \cdot P(A_{70}|A_{65}) = 0.746 \cdot 0.84 \approx 0.627.$ $\square$

Contents

1. The Definition of Conditional Probability
2. The Multiplication Rule
3. Total Probability Formula
4. Bayes' Theorem

The Multiplication Rule

The additivity of probability in Section 1.4 helped us to calculate the probabilities of the events that at least one of the events occurs.

How to deal with the probability of the events that many events occur at one time?

The multiplication rule (also known as the “Law of Multiplication”).

In a sense, the multiplication rule could be regarded as a perfect mathematical reflection of the Chinese phrase “proceed in an orderly way and step by step”.

The Multiplication Rule

Theorem

Assume that $P(B) > 0$. *Then*

$$P(AB) = P(B) \cdot P(A|B). \tag{2}$$

Or if $P(A) > 0$, *then*

$$P(AB) = P(A) \cdot P(B|A). \tag{3}$$

Theorem

Suppose that $A_1, A_2, \cdots, A_n$ *are events satisfying* $P(A_1 A_2 \cdots A_{n-1}) > 0$. *Then*

$$\begin{aligned} & P(A_1 A_2 \cdots A_n) \\ =& P(A_1) P(A_2|A_1) P(A_3|A_1 A_2) \cdots P(A_n|A_1 A_2 \cdots A_{n-1}). \end{aligned}$$

The Multiplication Rule

Example

(**Polya urn mode**) Suppose that an urn contains r red balls and b blue balls ($r \geqslant 2, b \geqslant 2$). Suppose that one ball is drawn randomly from the urn and its color is observed; it is then replaced in the urn, and c additional balls of the same color is added to the urn, and the selection process is repeated four times. We shall determine the probability of obtaining the sequence of outcomes red, blue, red, blue.

Solution. Let R_j denote the event that a red ball is obtained on the jth draw and B_j the event that a blue ball is obtained on the jth draw ($j = 1, 2, 3, 4$). Then

$$\begin{aligned} P(R_1 B_2 R_3 B_4) &= P(R_1)P(B_2|R_1)P(R_3|R_1B_2)P(B_4|R_1B_2R_3) \\ &= \frac{r}{r+b} \cdot \frac{b}{r+b+c} \cdot \frac{r+c}{r+b+2c} \cdot \frac{b+c}{r+b+3c}. \end{aligned}$$

$\square$

The Multiplication Rule

Theorem

Suppose that $A_1, A_2, \cdots, A_n, B$ *are events such that* $P(A_1A_2\cdots A_{n-1}|B) > 0$. *Then*

$$\begin{aligned}P(A_1A_2\cdots A_n|B) =&P(A_1|B)P(A_2|A_1B)P(A_3|A_1A_2B)\cdots\\&P(A_n|A_1A_2\cdots A_{n-1}B).\end{aligned}$$

Theoretically speaking, the multiplication rule holds no matter which one of these events happens first. But in actual applications, we usually use the multiplication rule in the situation that the events occur successively.

Contents

1. The Definition of Conditional Probability
2. The Multiplication Rule
3. Total Probability Formula
4. Bayes' Theorem

Total Probability Formula

Assume that A is the event caused by $B_1, B_2, \cdots, B_n$. The "Law of Total Probability" allows us to compute the probability of an event A by conditioning on cases, according to a partition $B_1, B_2, \cdots, B_n$ of the sample space.

Definition

*Let Ω denote the sample space of some experiment. n events $B_1, B_2, \cdots, B_n$ are said to form a **partition** of Ω if these events satisfy:*

(i) $B_1, B_2, \cdots, B_n$ *are disjoint and*

(ii) $\bigcup_{i=1}^{n} B_i = \Omega$.

Total Probability Formula

One way to partition Ω is to break it into sets B and $\overline{B}$, for any event B. Thus for event A, we have

$$\begin{aligned} P(A) &= P(A\Omega) = P(A \cap (B \cup \overline{B})) = P(AB \cup A\overline{B}) \\ &= P(AB) + P(A\overline{B}) = P(B) \cdot P(A|B) + P(\overline{B}) \cdot P(A|\overline{B}) \end{aligned}$$

It is the simplest form of the law of total probability. More generally,

Theorem

Suppose that the events $B_1, B_2, \cdots, B_n$ form a partition of the sample space Ω and $P(B_i) > 0$ for $i = 1, 2, \cdots, n$. Then, for every event A in Ω,

$$P(A) = \sum_{i=1}^{n} P(B_i)P(A|B_i). \tag{4}$$

Total Probability Formula

Example

Two cards from an ordinary deck of 52 cards are missing. What is the probability that a random card drawn from this deck is a spade?

Solution. Let A be the event that the randomly drawn card is a spade. Let B_i be the event that i spades are missing from the 50-card (defective) deck, for $i = 0, 1, 2$.
By conditioning on how many spades are missing from the original (good) deck, we get

$$P(A) = P(B_0)P(A|B_0) + P(B_1)P(A|B_1) + P(B_2)P(A|B_2)$$
$$= \frac{13}{50}\frac{\binom{13}{0}\binom{39}{2}}{\binom{52}{2}} + \frac{12}{50}\frac{\binom{13}{1}\binom{39}{1}}{\binom{52}{2}} + \frac{11}{50}\frac{\binom{13}{2}\binom{39}{0}}{\binom{52}{2}} \approx \frac{1}{4}.$$

$\square$

Total Probability Formula

How to ask a sensitive question?

- What proportion of people use illegal drugs?
- How many students ever cheated on an exam?
- ...

Surveying people directly and asking these types of sensitive questions are not likely to get honest responses and useful data.

Using probabilistic methods, statisticians have developed interesting ways, total probability methods, to ask sensitive questions that protect confidentiality.

Total Probability Formula

Example

Respondents are given a coin and told to flip it in private, not letting anyone see the outcome. If it lands heads, they answer the sensitive question of interest (e.g. "Have you ever taken illegal drugs?"). If tails, they answer an innocuous question such as "Were you born in the first half of the year?" The respondent reports a yes or no, but does not say which question they actually answered. From a sample of such yes-no responses, how can statisticians estimate the parameter of interest, such as the proportion of people who have ever taken illegal drugs?

Total Probability Formula

Solution. Let Y and N denote responses of yes and no, respectively. Let A denote the sensitive question and B the innocuous question. The unknown parameter that surveyors want to estimate is $p = P(Y|A)$, the probability that someone answers yes given that they were asked the sensitive question. We assume that the innocuous question is (i) easy to answer and (ii) has a known probability of yes and no, in this case 0.5 each.

Consider the unconditional probability $P(Y)$. By the law of total probability,

$$\begin{aligned} P(Y) &= P(Y|A)P(A) + P(Y|B)P(B) \\ &= p \cdot \frac{1}{2} + \frac{1}{2} \cdot \frac{1}{2} \\ &= \frac{p}{2} + \frac{1}{4}. \end{aligned}$$

Solution. When this survey is given to n people, the finial data will consist of n yes's and no's. The proportion of yes's is a simulated estimate of the unknown $P(Y)$. And thus

$$\frac{p}{2} + \frac{1}{4} = P(Y) \approx \frac{\text{Numbers of yes's in the sample}}{n}.$$

Solving for p gives

$$p \approx 2\left(\frac{\text{Numbers of yes's in the sample}}{n} - \frac{1}{4}\right),$$

which is the final estimate of the parameter of interest. □

Contents

1. The Definition of Conditional Probability
2. The Multiplication Rule
3. Total Probability Formula
4. Bayes' Theorem

Bayes' Theorem

Suppose that someone told you they had a nice conversation with someone on the train. Not knowing anything else about this conversation, the probability that they were speaking to a woman is 50%. Now suppose that they also told you that this person had long hair. It is now more likely they were speaking to a woman, since women are more likely to have long hair than men. The following result, which is known as *Bayes' theorem*, can be used to calculate the probability that the person is a woman.

Bayes' Theorem

Theorem

(***Bayes' Theorem***) *Let the events $B_1, B_2, \cdots, B_n$ form a partition of the sample space Ω such that $P(B_i) > 0$ for $i = 1, 2, \cdots, n$, and let A be an event such that $P(A) > 0$. Then, for $i = 1, 2, \cdots, n$,*

$$P(B_i|A) = \frac{P(B_i)P(A|B_i)}{\sum_{j=1}^{n} P(B_j)P(A|B_j)} \tag{5}$$

Proof. By the definition of conditional probability, $P(B_i|A) = \dfrac{P(B_iA)}{P(A)}$. By the multiplication rule for conditional probabilities, equation (2), $P(B_iA) = P(B_i)P(A|B_i)$. Thus $P(A) = \sum_{j=1}^{n} P(B_j)P(A|B_j)$. The equation 5 holds. □

全概率公式

Bayes' Theorem

Example

A new test has been devised for detecting a particular type of cancer. If the test is applied to a person who has this type of cancer, the probability that the person will have a positive reaction is 0.95 and the probability that the person will have a negative reaction is 0.05. If the test is applied to a person who does not have this type of cancer, the probability that the person will have a positive reaction is 0.05 and the probability that the person will have a negative reaction is 0.95. Suppose that in the general population, one person out of every 100,000 people has this type of cancer. If a person selected at random has a positive reaction to the test, what is the probability that he has this type of cancer?

Bayes' Theorem

Solution. Let A denote the event that a person has a positive reaction to the test, B_1 denote the event that he has this type of cancer and B_2 denote the event that he does not have this type of cancer. Now we have known that $P(A|B_1) = 0.95, P(\overline{A}|B_1) = 0.05, P(A|B_2) = 0.05, P(\overline{A}|B_2) = 0.95$ and $P(B_1) = 1/100000$. It now follows from Bayes' theorem that

$$\begin{aligned} P(B_1|A) &= \frac{P(B_1)P(A|B_1)}{P(B_1)P(A|B_1) + P(B_2)P(A|B_2)} \\ &= \frac{(1/100000) \cdot (0.95)}{(1/100000) \cdot (0.95) + (99999/100000) \cdot (0.05)} \approx 0.00019. \end{aligned}$$

$\square$

Bayes' Theorem

It seems impossible! The result is contrary to our common sense. So in practical applications, doctors generally increased the probability $P(B_1)$ through some other medical examinations in advance to increase the accuracy of the test. For example, if $P(B_1) = 0.2$, then $P(B_1|A) \approx 0.83$. This example illustrates not only the use of Bayes' theorem, but also the importance of taking into account all of the information available in a problem.

Example

Three different machines M_1, M_2 and M_3 were used for producing a large batch of similar manufactured items. Suppose that 20 percent of the items were produced by machine M_1, 30 percent of the items were produced by machine M_2, and 50 percent of the items were produced by machine M_3. Suppose further that 1 percent of the items produced by machine M_1 are defective, that 2 percent of the items produced by machine M_2 are defective, and that 3 percent of the items produced by machine M_3 are defective. Finally, suppose that one item is selected at random from the entire batch, and it is found to be defective. We shall determine the probability that this item was produced by machine $M_i (i = 1, 2, 3)$.

Bayes' Theorem

Solution. Let B_i be the event that the selected item was produced by machine $M_i (i = 1, 2, 3)$, and let A be the event that the selected item is defective. We must evaluate the conditinal probability $P(B_2|A)$. The probability $P(B_i)(i = 1, 2, 3)$ is as follows:

$$P(B_1) = 0.2, \qquad P(B_2) = 0.3, \qquad P(B_3) = 0.5.$$

Furthermore, the probability $P(A|B_i)$ that an item produced by machine M_i will be defective is:
$P(A|B_1) = 0.01, P(A|B_2) = 0.02, P(A|B_3) = 0.03$. It now follows from Bayes' theorem that

$$\begin{aligned} P(B_1|A) &= \frac{P(B_1)P(A|B_1)}{P(B_1)P(A|B_1) + P(B_2)P(A|B_2) + P(B_3)P(A|B_3)} \\ &= \frac{(0.2)(0.1)}{(0.2)(0.1) + (0.3)(0.02) + (0.5)(0.03)} = 0.087. \end{aligned}$$

By the similar way, we obtain
$P(B_2|A) = 0.261, P(B_3|A) = 0.652.$ $\square$

Bayes' Theorem

Remark

(i) *A probability like* $P(B_i)$ *is called the* ***prior probability*** *that the selected item will have been produced by machine* M_i*, because* $P(B_i)$ *is the probability of this event before the item is selected and before it is known whether the selected item is defective or nondefective.*

A probability like $P(B_i|A)$ *is called the* ***posterior probability*** *that the selected item was produced by machine* M_i*, because it is the probability of this event after it is known that the selected item is defective.*

Bayes' Theorem

Remark

(ii) *From above example, we can observe that*

$$P(B_1) = 0.2 > P(B_1|A) = 0.087,$$

$$P(B_3) = 0.5 < P(B_3|A) = 0.652.$$

That means there is no stationary numerical relationship between prior probability and posterior probability.

Application of Conditional Probability

Ask Marilyn

Figure: The Monty Hall problem

What would you do?
The critical factor in this problem is that the host knows what is behind the doors in advance.

Summary

- **Conditional Probability**
  $P(B|A) = P(AB)/P(A)$.
- **The multiplicational rule**
  $P(AB) = P(A)P(B|A)$. $= P(A)P(B)$
- **Total probability formula** independent
  $P(A) = \sum_{i=1}^{n} P(B_i)P(A|B_i)$.
- **Bayes' Theorem**

$$P(B_i|A) = \frac{P(B_i)P(A|B_i)}{\sum_{j=1}^{n} P(B_j)P(A|B_j)}$$

where the events $B_1, B_2, \cdots, B_n$ form a partition of the sample space Ω such that $P(B_i) > 0$ for $i = 1, 2, \cdots, n$.

Exercises

1.If A and B are two events such that $P(A) = 0.5, P(B) = 0.6$ and $P(B|\overline{A}) = 0.4$, then $P(B|A) =$?

Thank you for your patience !