Calculating the Conditional Probability of Events

Dmitri Nesteruk
QUANTITATIVE ANALYST

@dnesteruk http://activemesa.com

Goal:

Understand the concept of *conditional probability*.

Overview

What is Conditional Probability?

Independence of Events

Laws of Conditional Probability

Partitions and Total Probability

Bayes Theorem

Gambler’s Ruin Problem

Conditional vs Unconditional Probability

If want to do outdoor sports, you care about the probability of rain

Compare

- Probability it will rain today, given no additional information
- Probability it will rain today given it's been raining for an entire week

First value is the *unconditional* probability

Second value is the *conditional* probability, which depends on other observations

Conditional Probability

The conditional probability of event A given that event B has occurred, is

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

assuming $P(B) > 0$.

Conditional Probability Example

The unconditional probability that you roll a 3 on a balanced die roll is $1/6$

But what if you observed an odd number? This reduces the number of options to $\{1,3,5\}$, so the conditional probability is now $1/3$

Using our definition of conditional probability, the odds of rolling a 3 are:

$$P(3|\text{odd}) = \frac{P(3 \cap \text{odd})}{P(\text{odd})}$$

Since $P(3 \cap \text{odd}) = P(3) = 1/6$ and $P(\text{odd}) = 1/2$, we have

$$P(3|\text{odd}) = \frac{1/6}{1/2} = \frac{1}{3}$$

as expected

Independence of Events

Two events A and B are *independent* if <u>any</u> one of the following holds:

- $P(A|B) = P(A)$
- $P(B|A) = P(B)$
- $P(A \cap B) = P(A)P(B)$

If none of these hold, the events are *dependent*

Calculating Independence of Events

Consider these events generated by a single die roll:

- A: odd number rolled
- B: even number rolled
- $C = \{1,2\}$

Are A and B independent?

- We know that $P(A) = 1/2$, $P(B) = 1/2$
- Since $\text{A} \cap B = \emptyset$, $P(A|B) = 0$
- $P(A|B) \neq P(A) \Rightarrow$ events are dependent

Are A and C independent?

- $P(A|C) = \frac{1}{2} = P(A)$
- Events are independent

Calculating Independence of Events

Three brands of wine X, Y and Z are being ranked by judges

- A – wine X is ranked higher than Y
- B – wine X is ranked best
- C – wine X is ranked second best
- D – wine X is ranked worst

$E_1: XYZ$	$E_2: XZY$	$E_3: YXZ$
$E_4: YZX$	$E_5: ZXY$	$E_6: ZYX$

Possible outcomes for observations:

- $A = \{E_1, E_2, E_5\}$
- $B = \{E_1, E_2\}$
- $C = \{E_3, E_5\}$
- $D = \{E_4, E_6\}$

Calculate probabilities:

- $P(A) = \frac{3}{6} = \frac{1}{2}$
- $P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{1/3}{1/3} = 1$
- $P(A|C) = \frac{P(A \cap C)}{P(C)} = \frac{P(\{E_5\})}{1/3} = \frac{1/6}{1/3} = \frac{1}{2}$
- $P(A|D) = \frac{P(A \cap D)}{P(D)} = \frac{0}{P(D)} = 0$

A and C are independent

A and B are dependent
A and D are dependent

Multiplicative Law of Probability

From the formula

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

The probability of intersection of two events A and B is

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

If these events are independent, then

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

Multiplicative Law Example

Dependent events

- A – odd number
- B – number < 4

$$P(A) = \frac{1}{2};\ P(B) = \frac{1}{2}$$

$$P(A|B) = \frac{2}{3}; P(B|A) = \frac{2}{3}$$

$$P(A \cap B) = P(A)P(B|A) = \frac{1}{2} \cdot \frac{2}{3} = \frac{1}{3}$$

Additive Law of Probability

Probability of a union of events (we've met this before)

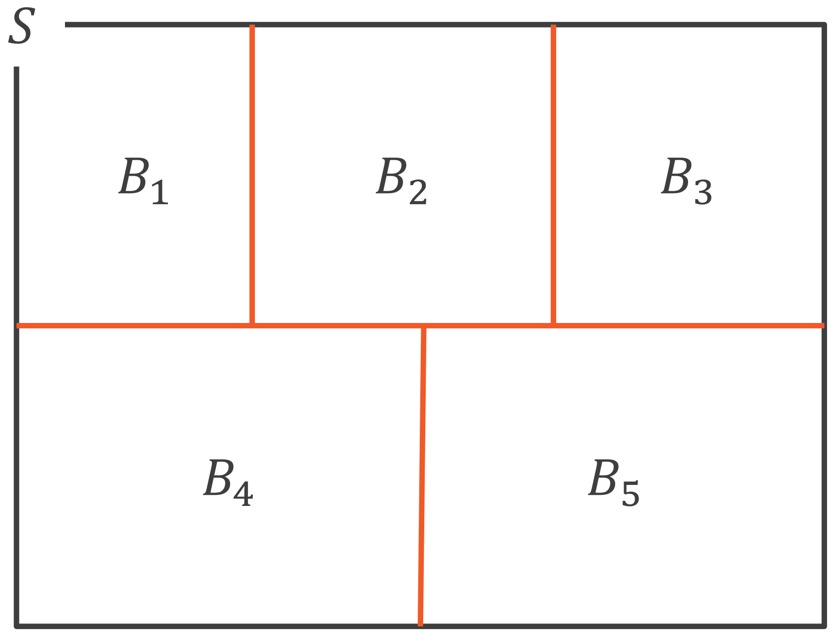
$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

If A and B are mutually exclusive,
$P(A \cap B) = 0$ and

$$P(A \cup B) = P(A) + P(B)$$

Both multiplicative and additive laws can be extended to calculate probabilities of >2 events

Partitions

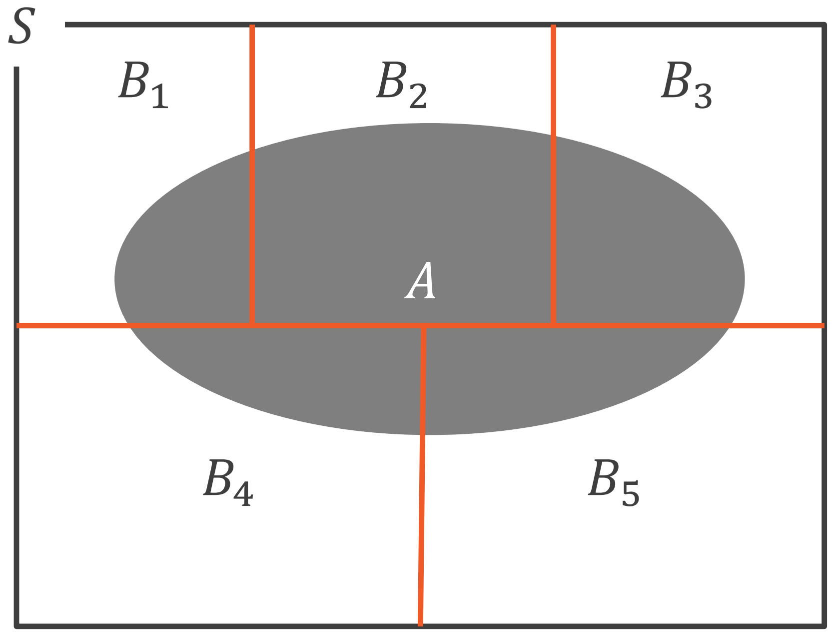


If S is the sample space of an experiment, given k disjoint events $B_1, \ldots, B_k$ s.t.

$$\bigcup_{i=1}^{k} B_i = S$$

We say that these events form a *partition* of S.

Law of Total Probability



Suppose $B_1, \ldots, B_k$ **are partitions over space** S **and** $P(B_j) > 0$ **for** $j = 1, \ldots, k$.

Then, for every event A **in** S,

$$P(A) = \sum_{j=1}^{k} P(B_j)P(A|B_j)$$

Conditional version:

$$P(A|C) = \sum_{j=1}^{k} P(B_j|C)P(A|B_j \cap C)$$

Total Probability Example

A player plays a game where their final score is somewhere from 1 to 50 (equally likely).

Player's score after first game is X.

Player continues to play until they obtain another score Y such that $Y \geq X$.

What is $P(Y = 50)$?

Let $B_i \stackrel{\text{def}}{=} P(X = i)$. Conditional on B_i, the value is likely to be any number $i, i+1, \dots, 50$. Each of these $(51 - i)$ values for Y are equally likely, so

$P(A|B_i) = P(Y = 50|B_i) = \frac{1}{51-i}$

$P(B_i) = 1/50$ for all i

$$P(A) = \sum_{i=1}^{50} \frac{1}{50} \cdot \frac{1}{51-i} = \frac{1}{50}\left(1 + \frac{1}{2} + \cdots + \frac{1}{50}\right) = 0.09$$

Bayes' Theorem Example

You can test a person for presence of absence of a particular disease

The test is 90% reliable

- If a person has the disease

$$P(\text{test positive}) = 0.9$$

- If a person does not have the disease

$$P(\text{test positive}) = 0.1$$

Chances of having this disease are 1 in 10,000

You test positive ☹

What's the probability you actually have this disease?

Bayes' Theorem

Let events $B_1, \ldots, B_k$ form a partition over S s.t. $P(B_j) > 0$ for $j = 1, \ldots, k$ and let A be an event s.t. $P(A) > 0$. Then, for $i = 1, \ldots, k$

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

Bayes' Theorem has a conditional version:

$$P(B_i|A \cap C) = \frac{P(B_i|C)P(A|B_i \cap C)}{\sum_{j=1}^{k} P(B_j|C)\,P(A|B_j \cap C)}$$

Bayes' Theorem Example

Probability that we have the disease is not 0.9 because it discounts the probability of 0.0001 that you had the disease before taking the test

Let B_1 denote that you have the disease and B_2 that you do not

B_1 and B_2 form a partition

Let A denote the event that the response to the test is positive

$$P(B_1|A) = \frac{P(A|B_1)P(B_1)}{P(A|B_1)P(B_1)+P(A|B_2)P(B_2)} = \frac{0.9 \cdot 0.0001}{0.9 \cdot 0.0001 + 0.1 \cdot 0.9999} = 0.0009$$

So the conditional probability that you have the disease is only 1 in 1000

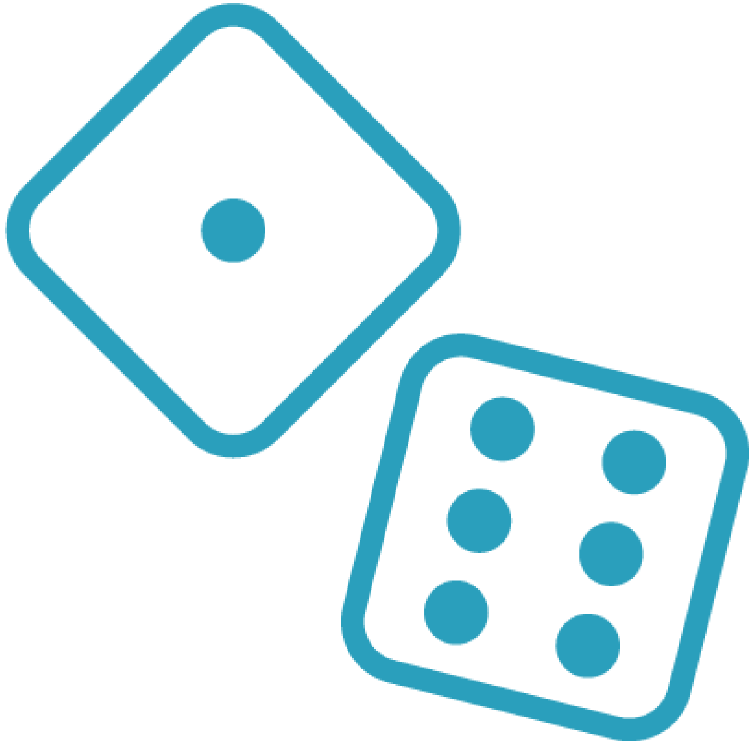
Gambler's Ruin Problem

Two gamblers A and B are playing a game against each other

On each play, probability that gambler A wins \$1 from gambler B is p (and for gambler B it's $1 - p$)

Initially, gambler A has i dollars and gambler B has $k - i$ dollars (total fortune $= k$)

Gamblers play repeatedly and independently until one of them reaches \$0



Consider the game from the perspective of player A

Initial balance of i dollars

$p > 1/2$ is favourable, $p < 1/2$ unfavourable

Game ends when

- Player A has k dollars (player B has \$0); or
- Player A has \$0 (player B has \$$k$)

Problem: determine the probability that the fortune of gambler A will reach k before it reaches \$0

"Gambler's Ruin" because one of the gamblers must go bankrupt

Gambler's Ruin Solution

Let a_i be the probability that gambler A reaches balance k before getting ruined given initial fortune of i

For each $j = 0, \dots, k$ each time we observe a set of plays that lead to A's fortune being j, the conditional probability that A wins is a_j

If gambler A's fortune reaches 0 then A is ruined, hence $a_0 = 0$

Let A_1 denote the event that A wins $1 on first play (and B_1 for gambler B)

Let W denote the event that the fortune of gambler A reaches k before reaching zero

$$\begin{aligned} P(W) &= P(A_1)P(W|A_1) + P(B_1)P(W|B_1) \\ &= pP(W|A_1) + (1-p)P(W|B_1) \end{aligned}$$

Gambler's Ruin Solution

Initial fortune of gambler A is i dollars ($i = 1, \ldots, k-1$), $P(W) = a_i$

If gambler A wins \$1 on first play, his fortune becomes $i+1$ and the conditional probability of $P(W|A_1) = a_{i+1}$

If A loses one dollar on the first play, his fortune becomes $i-1$ and $P(W|B_1) = a_{i-1}$

Thus

$$a_i = pa_{i+1} + (1-p)a_{i-1}$$

We know that $a_0 = 0$ and $a_k = 1$, so let $i = 1, \ldots, k-1$:

$$\begin{aligned} a_1 &= pa_2 \\ a_2 &= pa_2 + (1-p)a_1 \\ &\vdots \\ a_{k-2} &= pa_{k-1} + (1-p)a_{k-3} \\ a_{k-1} &= p + (1-p)a_{k-2} \end{aligned}$$

Rewrite left side of each term a_i as $pa_i + (1-p)a_i$ and rearrange

$$\begin{aligned} a_1 &= pa_2 \\ pa_1 + (1-p)a_1 &= pa_2 \\ (1-p)a_1 &= pa_2 - pa_1 \\ a_2 - a_1 &= \frac{1-p}{p}a_1 \end{aligned}$$

Gambler's Ruin Solution

Rewriting all relations in this form:

$$a_2 - a_1 = \frac{1-p}{p} a_1$$

$$a_3 - a_2 = \frac{1-p}{p}(a_2 - a_1) = \left(\frac{1-p}{p}\right)^2 a_1$$

$$\vdots$$

$$a_{k-1} - a_{k-2} = \frac{1-p}{p}(a_{k-2} - a_{k-3}) = \left(\frac{1-p}{p}\right)^{k-2} a_1$$

$$1 - a_{k-1} = \frac{1-p}{p}(a_{k-1} - a_{k-2}) = \left(\frac{1-p}{p}\right)^{k-1}$$

Taking the sum of left and right sides, we get

$$1 - a_1 = a_1 \sum_{i=1}^{k-1} \left(\frac{1-p}{p}\right)^i$$

Gambler's Ruin: Fair Game

Given $p = \frac{1}{2}$ and

$$1 - a_1 = a_1 \sum_{i=1}^{k-1} \left(\frac{1-p}{p}\right)^i$$

We end up with

$$1 - a_1 = a_1 \cdot (k-1)$$
$$a_1 = \frac{1}{k}$$

Example: suppose A has \$98 and B has \$2 ($i = 98, k = 100$). In a fair game, $a_{98} = \frac{98}{100} = 0.98$

Gambler's Ruin: Unfair Game

Suppose $p \neq \frac{1}{2}$. We can rewrite

$$1 - a_1 = a_1 \sum_{i=1}^{k-1} \left(\frac{1-p}{p}\right)^i$$

In the form

$$1 - a_1 = a_1 \frac{\left(\frac{1-p}{p}\right)^k - \left(\frac{1-p}{p}\right)}{\left(\frac{1-p}{p}\right) - 1}$$

$$a_1 = \frac{\left(\frac{1-p}{p}\right) - 1}{\left(\frac{1-p}{p}\right)^k - 1}$$

And similarly for values of a_i for $i = 2, \dots, k-1$

Gambler's Ruin: Unfair Game

Final solution for an unfair game:

$$a_i = \frac{\left(\frac{1-p}{p}\right)^i - 1}{\left(\frac{1-p}{p}\right)^k - 1} \text{ for } i = 1, \dots, k-1$$

Example: $p = 0.4$, **A has \$99, B has \$1**

$$\frac{1-p}{p} = \frac{3}{2}, i = 99, k = 100$$

$$a_i = \frac{\left(\frac{3}{2}\right)^{99} - 1}{\left(\frac{3}{2}\right)^{100} - 1} \approx \frac{1}{3/2} = \frac{2}{3}$$

Note this value is different from $1 - p$

Summary

Conditional probability of an event is the probability of an event *provided* some other event has been observed

Events can be tested for independence; independent events do not influence each other's probabilities

Bayes' theorem describes the probability of an event based on prior knowledge of conditions related to that event

Gambler's Ruin problem is a model for predicting the eventual outcome of a series of repeated bets