Chapter 1
Conditional Probability and Independence
Conditional Probability

Outline

Introduction

Toss a fair coin 3 times. Let winning be "at least two heads out of three"

HHH	HHT	HTH	HTT
THH	THT	TTH	TTT

If we now know that the first coin toss is heads, then only the top row is possible and we would like to say that the probability of winning is

$$\frac{\#(\text{outcome that result in a win and also have a heads on the first coin toss})}{\#(\text{outcomes with heads on the first coin toss})}$$

$$= \frac{\#\{\text{HHH, HHT, HTH}\}}{\#\{\text{HHH, HHT, HTH, HTT}\}} = \frac{3}{4}.$$

We can take this idea to create a formula in the case of equally likely outcomes for the statement the conditional probability of A given B.

$$P(A|B) = \text{the proportion of outcomes in } A \text{ that are also in } B = \frac{\#(A \cap B)}{\#(B)}.$$

Definition

We can turn this into a more general statement using only the probability, P, by dividing both the numerator and the denominator in this fraction by $\#(\Omega)$.

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

We thus take this version of the identity as the general definition of conditional probability for any pair of events A and B as long as the denominator $P(B) > 0$.

Exercise. Pick an event B so that $P(B) > 0$. Define, for every event A,

$$Q(A) = P(A|B).$$

Show that Q satisfies the three axioms of a probability. In words, *a conditional probability is a probability*.

Definition

Let's check the axioms:

1. For an event A,

$$Q(A) = P(A|B) = \frac{P(A \cap B)}{P(B)} \geq 0$$

2. For the sample space Ω,

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

3. For mutually exclusive events $\{A_j; j \geq 1\}$,

$$\begin{aligned} Q\left(\cup_{j=1}^{\infty} A_j\right) &= P\left(\cup_{j=1}^{\infty} A_j | B\right) = \frac{P((\cup_{j=1}^{\infty} A_j) \cap B)}{P(B)} = \frac{P(\cup_{j=1}^{\infty} (A_j \cap B))}{P(B)} \\ &= \frac{\sum_{j=1}^{\infty} P(A_j \cap B)}{P(B)} = \sum_{j=1}^{\infty} \frac{P(A_j \cap B)}{P(B)} = \sum_{j=1}^{\infty} Q(A_j) \end{aligned}$$

Introduction

Roll two dice. The event {first roll is 3} is indicated.

(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

Then $P\{\text{sum is 8}|\text{first die shows 3}\} = 1/6$, and $P\{\text{ sum is 8}|\text{first die shows 1}\} = 0$

Exercise. Roll two 4-sided dice. With the numbers 1 through 4 on each die, the value of the roll is the number on the side facing downward. Assume equally likely outcomes, find $P\{\text{sum is at least 5}\}$, $P\{\text{first die is 2}\}$ and $P\{\text{sum is at least 5}|\text{first die is 2}\}$.
Answers. $5/8$, $1/4$, and $1/2$.

Introduction

Exercise. If $A \subset B$ then $P(A|B) = P(A)/P(B)$. If $B \subset A$ then $P(A|B) = 1$.

If $A \subset B$ then

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

If $B \subset A$ then

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

Introduction

Example. Now suppose that we are to roll two dice until one of two numbers k_1 and k_2 is observed for the sum. Our sample space Ω is the set of all 36 outcomes on a pair of dice.

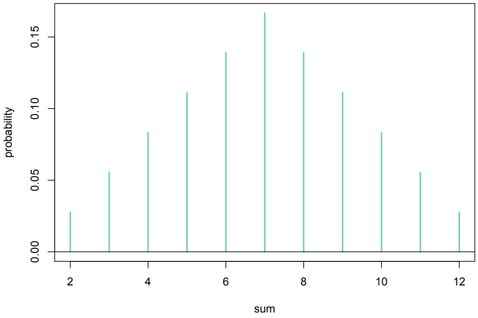

Figure: Probabilities for the sum $k = 2, \ldots, 12$ on two dice.

$$\begin{aligned} & P\{\text{sum is } k_1 | \text{sum is } k_1 \text{ or } k_2\} \\ = \; & \frac{P\{\text{sum is } k_1\}}{P\{\text{sum is } k_1 \text{ or } k_2\}} \\ = \; & \frac{(6 - |k_1 - 7|)/36}{(6 - |k_1 - 7| + 6 - |k_2 - 7|)/36} \\ = \; & \frac{6 - |k_1 - 7|}{12 - |k_1 - 7| - |k_2 - 7|} \end{aligned}$$

The Multiplication Principle

The defining formula for conditional probability can be rewritten to obtain the multiplication principle, $P(A \cap B) = P(A|B)P(B)$.

$$\begin{aligned} P\{\text{ace on first 2 cards}\} &= P\{\text{ace on 2nd card}|\text{ace on 1st card}\}P\{\text{ace on 1st card}\} \\ &= \frac{3}{51} \times \frac{4}{52} = \frac{1}{17} \times \frac{1}{13} \end{aligned}$$

We can continue this process to obtain a chain rule:

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

Thus, $P\{\text{ace on first 3 cards}\}$

$$\begin{aligned} &= P\{\text{ace on 3rd card}|\text{ace on 1st and 2nd card}\}P\{\text{ace on 2nd card}|\text{ace on 1st card}\} \\ &\quad \times P\{\text{ace on 1st card}\} = \frac{2}{50} \times \frac{3}{51} \times \frac{4}{52} = \frac{1}{25} \times \frac{1}{17} \times \frac{1}{13}. \end{aligned}$$

The Multiplication Principle

Exercise. For an urn with b brown balls and g green balls, find

- the probability of green, brown, green (in that order)
  $(g)_2(b)_1/(b+g)_3$
- the probability of green, green, brown (in that order)
  $(g)_2(b)_1/(b+g)_3$
- $P\{$exactly 2 out of 3 are green$\}$
  $3(g)_2(b)_1/(b+g)_3$,
- $P\{$exactly 2 out of 4 are green$\}$
  $6(g)_2(b)_2/(b+g)_4$

The Law of Total Probability

A partition of the sample space Ω is a finite collection of pairwise mutually exclusive events

$$\{C_1, C_2, \dots, C_n\}$$

whose union is Ω.

Thus, every outcome $\omega \in \Omega$ belongs to *exactly* one of the C_i. In particular, distinct members of the partition are mutually exclusive. ($C_i \cap C_j = \emptyset$, if $i \neq j$)

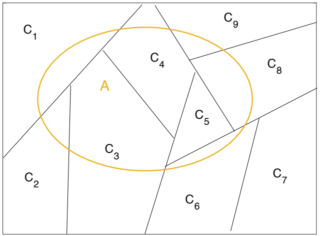

Figure: A partition of Ω into $n = 9$ events.

An event A can be written as the union

$$(A \cap C_1) \cup \cdots \cup (A \cap C_n)$$

of mutually exclusive events.

The Law of Total Probability

If we know, for each state, the fraction of the population whose ages are from 18 to 25, then we cannot just average these values to obtain this fraction over the whole country.This method fails because it give equal weight to California and Wyoming.

The law of total probability says that we should weigh these conditional probabilities by the probability of residence in a each state and then sum over the states.

Let $\{C_1, C_2, \ldots, C_n\}$ be a partition of Ω chosen so that $P(C_i) > 0$ for all i. Then, for any event A,

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

To check this identity:

$$\begin{aligned} P(A) &= P((A \cap C_1) \cup \cdots \cup (A \cap C_n)) = P(A \cap C_1) + \cdots + P(A \cap C_n) \\ &= P(A|C_1)P(C_1) + \cdots + P(A|C_n)P(C_n) \end{aligned}$$

NB. For the partition $\{C, C^c\}$, $P(A) = P(A|C)P(C) + P(A|C^c)P(C^c)$.

The Law of Total Probability

Example. To play craps, first role two dice.

- If the sum is 7 or 11, then the player wins immediately.
- If the sum is 2, 3, or 12, then the player loses immediately.
- If the sum is 4, 5, 6, 8, 9, or 10, and if the player rolls this number a second time before rolling a 7, then the player wins.

So,

$$P\{\text{winning immediately}\} = \frac{6}{36} + \frac{2}{36} = \frac{8}{36} = \frac{2}{9}.$$

The Law of Total Probability

For $k = 4, 5, 6, 8, 9$, or 10,

$$\begin{aligned} P\{\text{winning with } k\} &= P\{\text{rolling } k \text{ before } 7|\text{first roll is } k\} \cdot P\{\text{first roll is } k\} \\ &= \frac{6-|k-7|}{12-|k-7|} \cdot \frac{6-|k-7|}{36} = \frac{(6-|k-7|)^2}{36(12-|k-7|)} \end{aligned}$$

This yields the table

first roll	4	5	6	8	9	10
probability of winning	$\frac{9}{36\cdot 9}$	$\frac{16}{36\cdot 10}$	$\frac{25}{36\cdot 11}$	$\frac{25}{36\cdot 11}$	$\frac{16}{36\cdot 10}$	$\frac{9}{36\cdot 9}$

The Law of Total Probability

first roll	4	5	6	8	9	10
probability of winning	$\frac{9}{36\cdot 9}$	$\frac{16}{36\cdot 10}$	$\frac{25}{36\cdot 11}$	$\frac{25}{36\cdot 11}$	$\frac{16}{36\cdot 10}$	$\frac{9}{36\cdot 9}$

Thus,

$$\begin{aligned} P\{\text{winning after the first toss}\} &= \frac{2}{36}\left(1+\frac{8}{5}+\frac{25}{11}\right) \\ &= \frac{1}{18}\left(\frac{55+88+125}{55}\right) = \frac{268}{18\cdot 55} = \frac{134}{9\cdot 55} \end{aligned}$$

and

$$P\{\text{winning}\} = \frac{2}{9} + \frac{134}{9\cdot 55} = \frac{110+134}{9\cdot 55} = \frac{244}{495} = 0.493.$$

Independence

An event A is independent of B if the occurrence of B does not alter the probability of A:

$$P(A) = P(A|B).$$

Multiply this equation by $P(B)$ and use the multiplication rule to obtain

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

The formula $P(A)P(B) = P(A \cap B)$ is symmetric in A and B. Consequently, we say that A and B are independent.

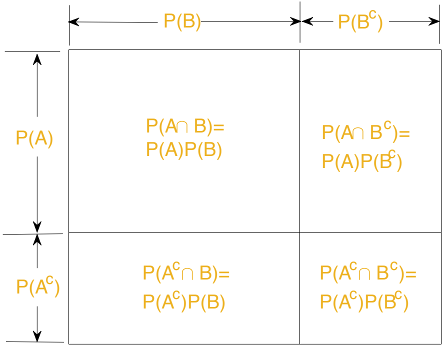

Figure: The Venn diagram for independent events.

Independence

Exercise. If A and B are independent, then show that A^c and B, A and B^c, A^c and B^c are also independent.

$$P(A) = P(A \cap B^c) + P(A \cap B) = P(A \cap B^c) + P(A)P(B)$$

$$P(A \cap B^c) = P(A) - P(A)P(B) = P(A)(1 - P(B)) = P(A)P(B^c)$$

For the second statement reverse the role of A and B.

For the third, because A and B are independent, then so is A^c and B as is the first event and the the complement of the second.

Independence

The events $A_1, \cdots, A_n$ are called independent if for any choice $A_{i_1}, A_{i_2}, \cdots, A_{i_k}$ taken from this collection of n events, then

$$P(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_k}) = P(A_{i_1})P(A_{i_2}) \cdots P(A_{i_k}).$$

A similar product formula holds if some of the events are replaced by the complement.

Exercise. Flip 10 biased coins. Their outcomes are independent with the i-th coin turning up heads with probability p_i. Find

$$P\{\text{first coin heads, third coin tails, seventh \& ninth coin heads}\}$$

$p_1(1 - p_3)p_7 p_9$,

$$P\{\text{fifth coin heads, seventh \& eighth coins agree}\}.$$

$p_5(p_7 p_8 + (1 - p_7)(1 - p_8))$

Association

If two traits are genetically linked, then the appearance of one increases the probability of the other. Let,

$$A_i = \{\text{individual has allele for trait } i\}, \quad 1 = 1, 2.$$

$$P(A_1|A_2) > P(A_1) \quad \text{implies} \quad P(A_2|A_1) > P(A_2).$$

Indeed, both are equivalent to

$$P(A_1 \cap A_2) > P(A_1)P(A_2)$$

In this case, we say that A_1 and A_2 are positively associated.

Reverse the inequalities for negatively associated events.

Association

Linkage disequilibrium

$$D_{A_1,A_2} = P(A_1)P(A_2) - P(A_1 \cap A_2).$$

measures the non-independent association of alleles at two loci on single chromosome. Thus if $D_{A,B} = 0$, the the two events are independent.

Exercise. Show that $D_{A_1,A_2^c} = -D_{A_1,A_2}$

Because A_1 is the disjoint union of $A_1 \cap A_2$ and $A_1 \cap A_2^c$, we have

$$P(A_1) = P(A_1 \cap A_2) + P(A_1 \cap A_2^c) \quad \text{or} \quad P(A_1 \cap A_2^c) = P(A_1) - P(A_1 \cap A_2).$$

Thus,

$$\begin{aligned} D_{A_1,A_2^c} &= P(A_1)P(A_2^c) - P(A_1 \cap A_2^c) \\ &= P(A_1)(1 - P(A_2)) - (P(A_1) - P(A_1 \cap A_2)) \\ &= -P(A_1)P(A_2) + P(A_1 \cap A_2) = -D_{A_1,A_2}. \end{aligned}$$