Introduction to Probability

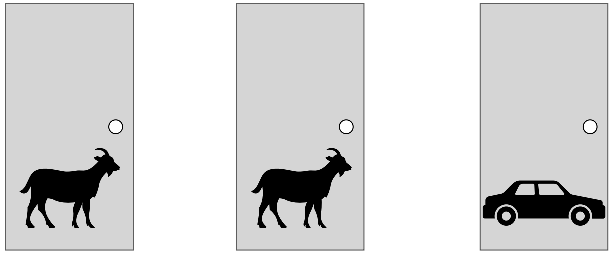
Lecture 7

Termeh Shafie

1

probabilities are rarely intuitive

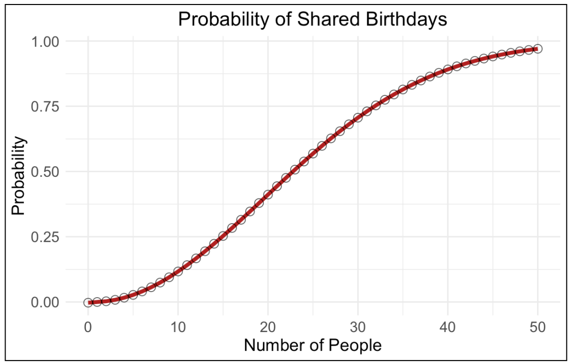
example 1: Monty Hall



2

probabilities are rarely intuitive

example 2: happy birthday you two!



3

what is probability?

Frequentist interpretation of the probability

- probability as the long-run fraction of time that it would happen if the random process occurs over and over again under the same conditions
- many interesting random phenomena cannot be repeated over and over again, e.g., weather

Bayesian interpretation of the probability

- probability as a subjective degree of belief: for the same event, two separate people could have different viewpoints and therefore assign different probabilities

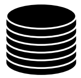
> "Probability is orderly opinion… inference from data is nothing other than the revision of such opinion in the light of relevant new information."
> – Thomas Bayes (1701-1761)

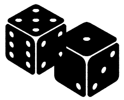
both interpretations agree on the probability rules introduced!

4

terminology for probability theory

- experiment
  process of observation or measurement
- outcome
  result obtained through an experiments
- sample space
  set of all possible outcomes of an experiment
  - finite
  - infinite
- events
  a subset of a sample space (what we are interested in)

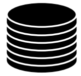




terminology for probability theory

example

Flip a coin 3 times and record the side facing up each times



Sample space Ω = {HHH, HHT, HTH, HTT, THH, THT, TTH, TTT}

Events:

{all heads} = {HHH}

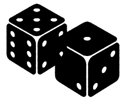
{get exactly one heads} = {HTT, THT, TTH}

{get at least two heads} = {HHT, HTH, THH, HHH}

exercise 1

Roll two dice, each with numbers 1-6.

Describe the event of rolling a total of 7 with the two dice.

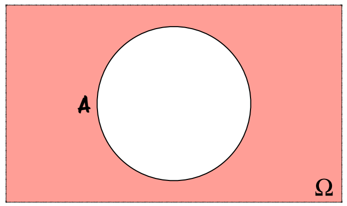


the algebra of events

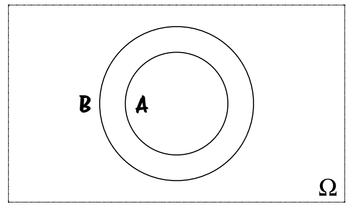
- Often we are interested in combinations of two or more events
- Events are sets (i.e. subsets of the sample space Ω) so we can do the usual set operations
- Assume sample space with two events A and B
  - complement $\overline{A}$ (also denoted A^c or A')
    all elements of S that are not in A
  - subset $A \subset B$
    all elements of A are also elements of B
  - union $A \cup B$
    all elements of Ω that are in A or B
  - intersection $A \cap B$
    all elements of Ω that are in A and B
- These operations can be represented graphically using Venn diagrams

venn diagrams

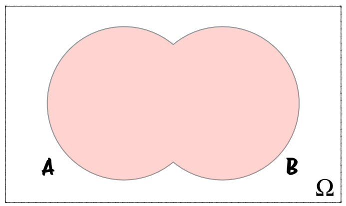
$\overline{A}$



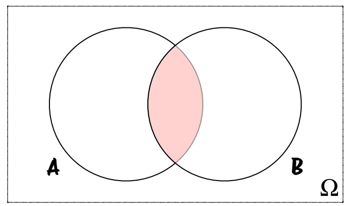
$A \subset B$



$A \cup B$



$A \cap B$



union and intersection: operator rules

Let E_1, E_2, E_3 denote events in Ω

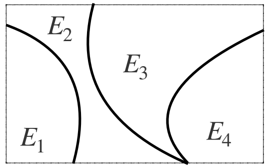
- Commutative
  $E_1 \cup E_2 = E_2 \cup E_1$
  $E_1 \cap E_2 = E_2 \cap E_1$

- Associative
  $(E_1 \cup E_2) \cup E_3 = E_1 \cup (E_2 \cup E_3)$
  $(E_1 \cap E_2) \cap E_3 = E_1 \cap (E_2 \cap E_3)$

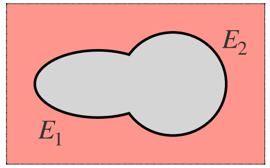
- Distributive
  $(E_1 \cup E_2) \cap E_3 = (E_1 \cap E_3) \cup (E_2 \cap E_3)$
  $(E_1 \cap E_2) \cup E_3 = (E_1 \cup E_3) \cap (E_2 \cup E_3)$

the complement

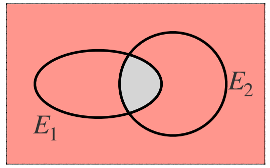
MECE = mutually exclusive and collectively exhaustive events



$\overline{E_1 \cup E_2} = \overline{E_1} \cap \overline{E_2}$



$\overline{E_1 \cap E_2} = \overline{E_1} \cup \overline{E_2}$



axioms of probability

1. The probability of an event is a nonnegative real number $P(A) \geq 0$ for any $A \subset \Omega$
2. $P(\Omega) = 1$ (also denoted $P(S) = 1$)
3. If $A_1, A_2, A_3, \ldots$ is a sequence of **mutually exclusive events** of Ω, then:

$$P(A_1 \cup A_2 \cup A_3 \cup \cdots = P(A_1) + P(A_2) + P(A_3) + \cdots$$

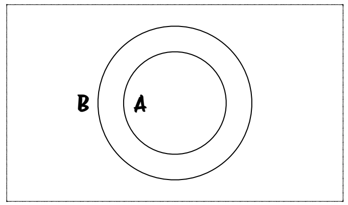
further properties

- $P(\varnothing) = 0$

Proof:
$1 = P(\Omega) + P(\Omega^c)$
$1 = 1 + P(\varnothing) \implies P(\varnothing) = 0$
Also evident from set theory:
$\Omega \cup \varnothing = \Omega \implies P(\Omega) + P(\varnothing) = P(\Omega) \implies P(\Omega) = 0$

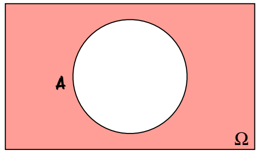
- if $A \subset B$ then $P(A) \leq P(B)$

Proof:
$B = A \cup (B \cap \overline{A})$
$\implies P(B) = P(A) + P(B \cap \overline{A}) \geq P(A)$



further properties

- $P(A \cup \overline{A}) = P(A) + P(\overline{A}) = 1 \implies P(\overline{A}) = 1 - P(A)$
  (this is also referred to as the complement rule coming up shortly…)



- $0 \leq P(A) \leq 1$ for any event A
  Directly follows from axiom (1) and (2).
  Also directly evident from set theory:
  $\varnothing \subset A \subset \Omega$ for any event $A \implies P(\varnothing) \leq P(A) \leq P(\Omega) \implies 0 \leq A \leq 1$

probability of an event

If A is an event in a discrete sample space A and $O_1, O_2, O_3, \ldots$ are the individual outcomes comprising A, then $P(A) = P(O_1) + P(O_2) + P(O_3) + \ldots$

example
We flip a fair coin twice. What's the probability of obtaining at least one head?

The sample space $S = \{HH, HT, TH, TT\}$

As the coin is fair, all outcomes are equally likely: $P(HH) = P(HT) = P(TH) = P(TT) = \frac{1}{4}$

The event of obtaining at least one head is $A = \{HH, HT, TH\}$

$$\implies P(A) = P(HH) + P(HT) + P(TH) = \frac{1}{4} + \frac{1}{4} + \frac{1}{4} = \frac{3}{4}$$

probability of an event

exercise 2
Roll two dice, each with numbers 1–6. Let X denote the first roll and Y the second roll.

(a) Find the probability $P(X = 1)$

(b) Let $Z = \min(X, Y)$ and find the probability $P(Z = 6)$

(c) Let $Z = \min(X, Y)$ and find the probability $P(Z = 3)$

6						
5						
4						
3						
2						
1						
	1	2	3	4	5	6

equally likely outcomes

If an experiment can result in N equally likely outcomes, and if n of these outcomes constitute an event A, then $P(A) = \frac{n}{N}$

This theorem is consistent with the frequency interpretation of probability theory:
probability of an event is the proportion of the time events of the same kind occur in the long run.

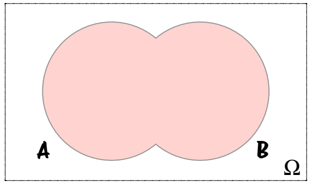
exercise 3
Assume all letters occur equally often in English. Then what's the probability of a three-letter word only consisting of vowels?

probability rules

The Addition Rule

If A and B are events in the sample space Ω, then

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



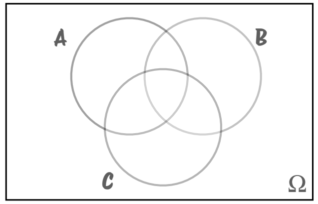
Proof:

Let $a = P(A \cap \overline{B}), b = P(A \cap B), c = P(B \cap \overline{A})$

$P(A \cup B) = a + b + c$

$P(A) + P(B) - P(A \cap B) = [a + b] + [b + c] - b = a + b + c$

probability rules



Inclusion-Exclusion Rule

For three events $A,\ B,\ C$

$$P(A \cup B \cup C) = P(A) + P(\overline{A} \cap B) + P(\overline{A} \cap \overline{B} \cap C)$$

$$\begin{aligned} P(A \cup B \cup C) = P(A) &+ P(B) + P(C) \\ &- P(A \cap B) - P(A \cap C) - P(B \cup C) \\ &+ P(A \cap B \cap C) \end{aligned}$$

For any number of finite events events $A_1, A_2, \ldots, A_n$,

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{k=1}^{n} (-1)^{k+1} \left(\sum_{1 \le i_1 < i_2 < \cdots < i_k \le n} P(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_k}) \right).$$

probability rules

Conditional Probability Rule

If A and B are events in the sample space Ω, then the conditional probability of A given B where $P(B) > 0$ is given by

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

$\Longrightarrow$ The Multiplication Rule

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

since $A \cap B = B \cap A$

$\Longrightarrow \quad P(A \cap B) = P(B) P(A \mid B)$

Independent Events and Their Complement

Two events A and B are independent if and only if

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

Two events A and B are independent then A and $\overline{B}$ are also independent.

probability rules

Complement Rule

If A be an event in the sample space Ω, then the probability of its complement is given by

$$P(\overline{A}) = 1 - P(A)$$

exercise 4

What is the probability of at least one head (H) in four tosses of a coin?

probability rules

Rule of Total Probability

If $\{A_1, A_2, \ldots, A_k\}$ are mutually exclusive and collectively exhaustive (a partition of Ω), and $P(A_i) \neq 0 \ \forall i$, then

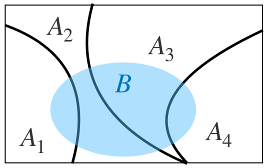
$$P(B) = \sum_{i=1}^{k} P(A_i)P(B \mid A_i)$$

Proof: $P(B) = P(B \cap (A_1 \cup A_2 \cup \cdots \cup A_k))$

$= P((B \cap A_1) \cup (B \cap A_2) \cup \cdots \cup (B \cap A_k))$ {events collectively exhaustive}

$= \sum_{i=1}^{k} P(B \cap A_i)$ {mutually exclusive}

$= \sum_{i=1}^{k} P(A_i)P(B \mid A_i)$ ■



so this rule only applies to MECE events!

probability rules

Bayes Rule

If the events $A_1, A_2, \ldots, A_k$ form a mutually exclusive and collectively exhaustive partition of the sample space Ω, and if $P(A_i) \neq 0$ for all i, then for any event B with $P(B) \neq 0$:

$$P(A_i \mid B) = \frac{P(A_i)P(B \mid A_i)}{P(A_1)P(B \mid A_1) + P(A_2)P(B \mid A_2) + \cdots + P(A_k)P(B \mid A_k)}$$

$$= \frac{P(A_i)P(B \mid A_i)}{\sum_{j=1}^{k} P(A_j \mid B)P(A_j)}$$ {rule of total probability}

$$= \frac{P(A_i)P(B \mid A_i)}{P(B)}$$

$P(A_i \mid B) = \frac{\text{prior of } A_i \times \text{likelihood of } B \text{ given } A_i}{\text{total probability of } B}$

This theorem is consistent with the Bayesian interpretation of probability theory

probability rules

exercise 5

In an experiment on human memory, participants have to memorize a set of words (B_1), numbers (B_2), and pictures (B_3). These occur in the experiment with the probabilities $P(B_1) = 0.5,\ P(B_2) = 0.4,\ P(B_3) = 0.1$.

Then participants have to recall the items (where A is the recall event). The results show that $P(A \mid B_1) = 0.4,\ P(A \mid B_2) = 0.2,\ P(A \mid B_3) = 0.1$.

(a) Compute $P(A)$, the probability of recalling an item.

(b) What is the probability that an item that is correctly recalled (A) is a picture (B_3)?

counting outcomes

A permutation of items is an arrangement of the items in a certain order, where each item can be used only once in the sequence: $n! = n(n-1)(n-2)\cdots(2)(1)$

A permutation of n items taken k at of time is the number of ways to select k items from n distinct items and arranging them in order:

$$P(n, k) = \frac{n!}{(n-k)!}$$

use when you're asked to arrange k things out of n

A combination of n items taken k at of time any selection of k items from n elements where order is not important:

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

use when you're asked to select k things out of n

counting outcomes

example

In a group of six men and four women, I select a committee of three at random. What is the probability that all three committee numbers are women?

The number of ways to select a three-women committee from four women: $\binom{4}{3} = \frac{3!}{3!(4-3)!} = 4$

The number of ways to select a three-person committee from the 10 people (the total number of outcomes):

$$\binom{10}{3} = \frac{10!}{10!(10-3)!} = 120$$

Probability of all female committee: $\frac{4}{120} = \frac{1}{30} = 0.03$

odds and log odds

odds are the ratio of something happening to something not happening

the odds of my team winning is 1 to 4: $\frac{1}{4} = = 0.25$

the odds of my team winning is 5 to 3: $\frac{5}{3} = = 1.7$

the probability is the ratio of something happening to everything that could happen

$$= \frac{1}{5} = 0.20$$

$$= \frac{5}{8} = 0.625$$

odds and log odds

the probability is the ratio of something happening to everything that could happen

the probability of winning: $= \frac{5}{8} = 0.625$

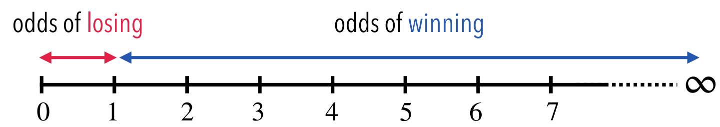
the probability of losing: $= \frac{3}{8} = 1 - \frac{5}{8} = 0.375$

the ratio of probability of winning to the probability of losing $= \frac{p}{(1-p)} = \frac{5/8}{3/8} = \frac{5}{3} =$

the ratio of probabilities are the same as the ratio of the raw counts resulting in the same odds

odds and log odds

why log odds?

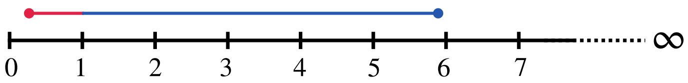


odds and log odds

why log odds?

example: odds against 1/6=0.17 but odds in favor 6/1=6

taking the log of the odds makes everything symmetrical

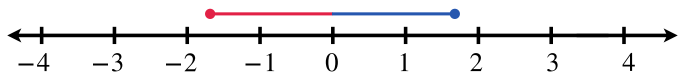


odds and log odds

why log odds?

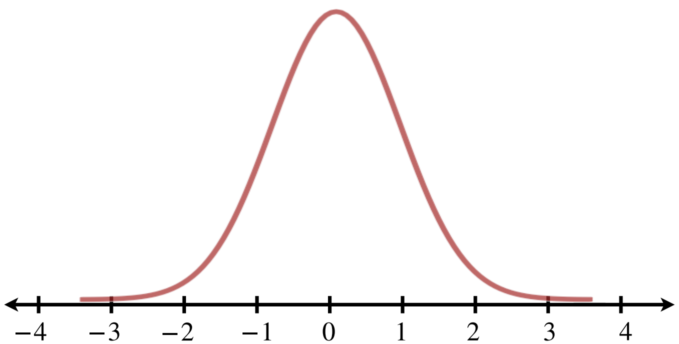
example: log odds against log(0.17)=−1.79 but log odds in favor log(6)=1.79

taking the log of the odds makes everything symmetrical



odds and log odds

why log odds?



odds and log odds and probability

Probability p	Odds p/(1-p)	Log Odds log[p/(1-p)]
0.1	0.1111	-2.1972
0.5	1	0
0.9	9	2.1972