Definition of Probability

Definition. Let $\mathcal{S}$ denote the set of all subsets of a sample space S.
A **probability** (or **probability model**) is a real function defined on $\mathcal{S}$ that satisfies the following three conditions (Axioms of Probability):

A1: For any event $A \subseteq S$, we have $0 \leq P(A) \leq 1$

A2: $P(S) = 1$ and $P(\emptyset) = 0$ ($\emptyset$ denotes the empty set)

A3: Additivity: if $A_1, A_2, \ldots$ is a sequence of disjoint (or mutually exclusive) events[1], then

$$P(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i).$$

[1] that is: $A_i \cap A_j = 0$ for any i,j such that $i \neq j$

Figure: A.N. Kolmogorov, 1963. Developed the Axioms of Probability (1933)

How do we construct a function P on $\mathcal{S}$ that satisfies the Axioms 1-3?
If S is finite or countable then a common approach is based on the following two steps:

S1: assign to each elementary event $\{a\}$ from $\mathcal{S}$ a number $P(a) \in [0,1]$ such that $P(\{a\}),\ a \in S$, satisfy

$$\sum_{\text{all } a \in S} P(\{a\}) = 1.$$

S2: assign to an event A from $\mathcal{S}$ a number $P(A)$ according to the rule

$$\boxed{P(A) := \sum_{\text{all } a \in A} P(\{a\}). \qquad (1)}$$

Comments:

- One can easily verify that the probability defined through steps S1-S2 above indeed satisfies the axioms of probability A1-A3 (see the notes).
- Note that if we want to construct a valid probability model that satisfies axiom A3 then in step S2 we have no choice but to use formula (1).
- Step S1 depends on the particular situation we want to model. In Chapter 5 we will look at several "standard models".
  We already know one such model, namely **discrete uniform distribution**, where for $N = |S|$ we take

$$P(\{a_1\}) = \cdots = P(\{a_N\}) = \frac{1}{N}.$$

Clicker Question(s).

Working with events

From the definition of probability we will be able to derive general[2] probability rules that will allow us to work with unions, intersections and compliments of given events.

Why do we need these? Because unions, intersections and compliments arise naturally in many applications.

By using these rules, we can express probabilities of more complex events in terms of simpler ones.

[2] general as they are valid as long as the Axioms A1-A3 are satisfied

Unions

- By definition, for two subsets A and B of S we have

$$A \cup B := \{s \in S : \ s \text{ belongs either to } A \textbf{ or} \text{ to } B\}.$$

Note: we will use "or" inclusively to also permit both.

- Using the "language of events":

event $A \cup B$ occurs iff A occurs **or** B occurs

or

event $A \cup B$ occurs iff at least one of A, B occurs

- Similarly

event $A \cup B \cup C$ occurs iff at least one of A, B, or C occurs

Intersections

- By definition, for two subsets A and B of S we have

$$A \cap B := \{s \in S : \; s \text{ belongs to both } A \textbf{ and } \text{to } B\}.$$

- Similarly

$$A \cap B \cap C := \{s \in S : \; s \text{ belongs to } A \text{ and to } B \text{ and to } C\}.$$

- To simplify the notation, we will use

$$AB \quad \text{for} \quad A \cap B$$

or

$$ABC \quad \text{for} \quad A \cap B \cap C.$$

- Using the "language of events":

$$\text{event } A \cap B \text{ occurs} \quad \text{iff} \quad A \text{ occurs } \textbf{and } B \text{ occurs}$$

or

$$\text{event } A \cap B \text{ occurs} \quad \text{iff} \quad \text{both } A \text{ and } B \text{ occur}$$

- Similarly,

$$\text{event } A \cap B \cap C \text{ occurs} \quad \text{iff} \quad \text{all } A,\ B, \text{ and } C \text{ occur}$$

Complement of an event

- By definition, the complement of an event A, denoted by $\bar{A}$ (or A^c), is defined as

$$\bar{A} := \{s \in S : \ s \text{ does not belong to } A\}.$$

In addition,

$$\bar{S} = \emptyset.$$

- Using the "language of events":

event $\bar{A}$ occurs iff A does not occur.

Note that by definition:

- the event S always occurs
- the event $\emptyset$ never occurs
- for any $A \subseteq S$, we have: $\bar{\bar{A}} = A$

De Morgan's Laws

- For two events:

$$\overline{A \cup B} = \bar{A} \cap \bar{B}$$

$$\overline{A \cap B} = \bar{A} \cup \bar{B}$$

- More generally:

$$\overline{\bigcup_{i=1}^{\infty} A_i} = \bigcap_{i=1}^{\infty} \bar{A}_i$$

$$\overline{\bigcap_{i=1}^{\infty} A_i} = \bigcup_{i=1}^{\infty} \bar{A}_i.$$

Examples: for the events A,B, and C from $\mathcal{S}$, find representations of the following events:

- A occurs but not B
- A occurs but not B nor C
- A and C occur but not B
- A or B occurs but not C

Clicker Question(s).

Exercise: for the events A, B, and C, find representations of the following events:

(i) exactly two of the three events occur

(ii) at least two of the three events occur[3]

[3] when we say "at least two" we mean two or more

Probability Rules

From the definition of probability, we can derive the following rules:

(R1) If $A_1, A_2, ...A_k$ is a finite sequence of mutually exclusive events, then

$$P(\cup_{i=1}^{k} A_i) = \sum_{i=1}^{k} P(A_i).$$

Proof: use Axioms 2 and 3.

(R2) For any event $A \in \mathcal{S}$, we have

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

Proof: use rule $(R1)$

(R3) If $A \subseteq B$, then $P(A) \leq P(B)$.

Proof: use rule $(R1)$ and Axiom 1.

(R4) For A and B such that $B \subseteq A$ we define

$$A \setminus B := A \cap \bar{B}.$$

Then

$$P(A \setminus B) = P(A) - P(B).$$

Proof: apply rule $(R1)$ to $A \setminus B$ and B

Rules for Unions

- For any two events A and B we have
$$P(A \cup B) = P(A) + P(B) - P(A \cap B).$$

- For any three events A, B, and C we have
$$\begin{aligned} P(A \cup B \cup C) &= P(A) + P(B) + P(C) - P(AB) - P(BC) \\ &\quad - P(AC) + P(ABC). \end{aligned}$$

- For any n events $A_1, A_2, \ldots, A_n$ we have the following *inclusion-exclusions* rule
$$\begin{aligned} P(A_1 \cup A_2 \cup \cdots \cup A_n) &= \sum_i P(A_i) - \sum_{i<j} P(A_iA_j) + \sum_{i<j<k} P(A_iA_jA_k) \\ &\quad - \sum_{i<j<k<l} P(A_iA_jA_kA_l) + \cdots . \end{aligned}$$

Clicker Question(s).

Dependent and Independent Events

Definition. Two events A and B are said to be **independent** if and only if

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

Comments:

- Events that are not independent are called **dependent.**
- This definition works two ways:
  - "$\Rightarrow$" if we know, or can assume, that A and B are independent then we can use the definition as a rule of probability to calculate $P(A \cap B)$
  - "$\Leftarrow$" if we know $P(A)$, $P(B)$, and $P(A \cap B)$, then we can determine whether or not A and B are independent by checking if

$$P(A \cap B) = P(A)P(B) \text{ holds.}$$

- The definition of independence tries to capture the usual common-sense interpretation of independence:

  A and B are independent iff

  knowing that A has occurred does not change the chances *of* B occurring, and vice versa.

- For example: when we roll two dice then normally we assume that the two rolled numbers are independent of each other.

- We will be able to provide an additional interpretation of the definition of independence once we introduce the concept of “conditional probability”.

Definition. The events $A_1, A_2, ...A_n$ are said to be (mutually) independent iff

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

for all possible subsets $\{i_1, i_2, \ldots, i_k\}$ of distinct subscripts selected from $\{1, 2, \ldots, n\}$.

For example, when $n = 3$ we need to check

$$\begin{aligned} P(A_1 \cap A_2) &= P(A_1)P(A_2) \\ P(A_2 \cap A_3) &= P(A_2)P(A_3) \\ P(A_1 \cap A_3) &= P(A_1)P(A_3) \\ P(A_1 \cap A_2 \cap A_3) &= P(A_1)P(A_2)P(A_3). \end{aligned}$$

Example. Consider rolling two fair six sided dice, and let

$$\begin{aligned} A &= \{\text{the sum is } 10\} \\ B &= \{\text{the first die is a } 6\} \\ C &= \{\text{the sum is } 7\}. \end{aligned}$$

1. Are A and B independent?
2. Are B and C independent?
3. Are A and C independent?

A common misconception is that if A and B are mutually exclusive, then A and B are independent.

Proposition. If A and B are independent and mutually exclusive, then either $P(A) = 0$ or $P(B) = 0$.

This result implies that if the events A and B are mutually exclusive and such that $P(A) > 0$ and $P(B) > 0$, then they must be dependent.

Useful fact

Proposition. If A and B are independent, then

$$\bar{A} \text{ and } B \text{ are independent.}$$

Clicker Question(s).

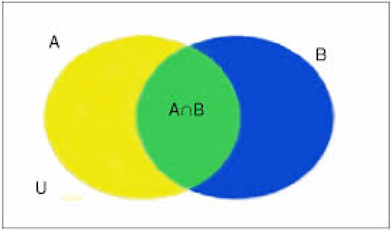
Conditional Probability

Definition. For two events A and B, the conditional probability of A *given* B is defined as

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

provided $P(B) > 0$.

Intuition: knowing that B has occurred, this event becomes our new sample space and hence now we measure chances of A occurring with respect to B.



Example. Consider rolling two fair six sided dice. Let $A = \{\text{the sum is 10}\}$, $B = \{\text{the first die is a 6}\}$ $C = \{\text{the sum is 7}\}$.

1. Compute $P(A|B)$
2. Compute $P(B|A)$
3. Compute $P(A|C)$
4. Compute $P(C|B)$

- **Equivalent definition of independence.**

  Two events A and B are independent if and only if either of the following statements is true

$$P(A|B) = P(A) \quad \text{or} \quad P(B|A) = P(B).$$

  Intuition: if knowing B does not change the chances of A occurring then the two events must be independent.

  Similarly: if knowing A does not change the chances of B occurring then the two events must be independent.

Rules for conditional probability

Working with a conditional probability is relatively simple, since it behaves in the same way as the usual probability!

In particular, for any fixed B such that $P(B) > 0$ and any event A we have:

1. $0 \leq P(A|B) \leq 1$
2. $P(S|B) = 1 = P(B|B)$
3. $P(\bar{A}|B) = 1 - P(A|B)$
4. If A_1 and A_2 are disjoint then

$$P(A_1 \cup A_2|B) = P(A_1|B) + P(A_2|B).$$

Example. Consider rearranging the letters in the word RACECAR at random to form a word.

1. What is the probability that the random word ends with an "R" given that the word starts with the three letter sequence "ACE".
2. Is the event that the word starts with "ACE" independent of the event that it ends with an "R"?

Product Rules

Definition. For events A and B,

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

provided $P(A) > 0$, $P(B) > 0$.

These equations are known as the **product rules**. They follow directly from the definition of the conditional probability.

Similarly

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

and so on.

Example. A box contains 12 red balls and 7 blue balls. Suppose that a ball is drawn at random, then, without replacement, a second ball is drawn at random.

What is the probability that both balls are red?

Clicker Question(s).

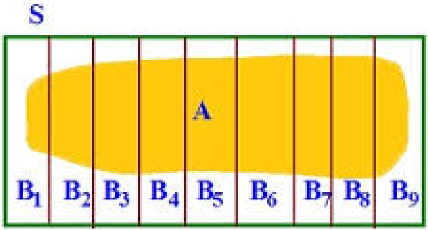
Law of Total Probability

Example. A box contains 12 red balls and 7 blue balls. Suppose that a ball is drawn at random, then, without replacement, a second ball is drawn at random.

1. What is the probability that both balls are red?
2. What is the probability that the second ball is red?

Definition. Sets $B_1, B_2, ..., B_k$ are said to **partition** the sample space S if $B_i \cap B_j = \emptyset$ for all $i \neq j$, and

$$\cup_{j=1}^{k} B_j = S.$$



***Theorem* (Law of total probability)**

Suppose that $B_1, B_2, ..., B_k$ is a partition of S such that $P(B_i) > 0$ for each i. Then, for any event A,

$$\begin{aligned} P(A) &= P(A|B_1)P(B_1) + P(A|B_2)P(B_2) + \cdots + P(A|B_k)P(B_k) \\ &= \sum_{i=1}^{k} P(A|B_i)P(B_i). \end{aligned}$$

Example. The simplest partition is of the from

$$S = B \cup \bar{B},$$

with $0 < P(B) < 1$, in which case we get

$$P(A) = P(A|B)P(B) + P(A|\bar{B})P(\bar{B}).$$

Example. A box contains 12 red balls and 7 blue balls. Suppose that a ball is drawn at random, then, without replacement, a second ball is drawn at random.

What is the probability that the second ball is red?

Exercise. We have two boxes with balls: in the first box there are 3 white balls and 5 red balls, while in the second box there 5 white balls and 3 red balls.

Suppose that a ball is randomly selected from the first box and put in the second box. Then a ball is randomly selected from the second box.

What is the probability that the ball selected from the second box is red?

Bayes Theorem

Theorem (Bayes Theorem). Suppose that $B_1, B_2, ..., B_k$ partition S and $P(B_i) > 0,\ i = 1, \ldots, k$. Then for any event A,

$$P(B_i|A) = \frac{P(A|B_i)P(B_i)}{\sum_{j=1}^{k} P(A|B_j)P(B_j)}, \quad i = 1, \ldots, k.$$

For the partition B and $\bar{B}$:

$$P(B|A) = \frac{P(A|B)P(B)}{P(A|B)P(B) + P(A|\bar{B})P(\bar{B})}.$$

Figure: Reverend Thomas Bayes: 1701-1761

Example. In an insurance portfolio, 10% of the policy holders are in Class A1 (high risk) and 90% are in Class A2 (medium-low risk).

The probability there is a claim on a Class A1 policy in a given year is 0.15, while similar probability for Class A2 is 0.03.

Find the probability that if a claim is made, it is made on a Class A1 policy.

Example. A rare disease occurs in 0.3% of a population. A test has been developed for the disease. If the test is performed on an individual who has the disease, the probability of a positive test result (i.e. the test says that the individual has the disease) is 0.96. If the test is performed on an individual who does not have the disease, the probability of a positive test result is 0.05. Suppose the test is performed on a random individual in the population.

(i) What is the probability that the test result will be positive?

(ii) Suppose the test comes back positive. What is the probability that the individual actually has the disease?

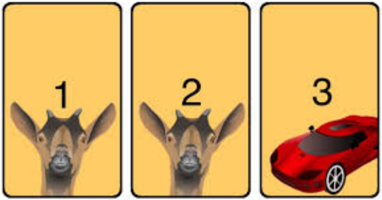
(iii) If the test comes back negative, what is the probability that the individual actually has the disease?

Exercise: Suppose a particular surveillance system has a 99% chance of correctly identifying a "threatening person" and a 99.99% chance of correctly identifying someone who is not a "threatening person".

If there are 1000 "threatening people" in a population of 300 million, and one of these 300 million is randomly selected, scrutinized by the system, and identified as a threat, what is the probability that the selected individual actually is a treat?

Clicker Question(s).

Exercise: The Monty Hall Problem, popularized by the show “Lets Make a Deal” (Problem 25 in Chapter 4):



You have been chosen as finalist on a tv show, and the host shows you three doors. Behind one door is a sports car, and behind the other two are goats. After you choose one door (but do not open it), the host, who knows what is behind each of the three doors, opens one (never the one you chose or the one with the car) and then says: “You are allowed to switch the door you chose if you find that advantageous”. Should you switch?

Useful Series and Sums

Some basic mathematical results:

- Geometric Series Formula

$$\sum_{i=0}^{\infty} x^i \equiv 1 + x + x^2 + \cdots = \frac{1}{1-x}, \quad \textit{if} \quad |x| < 1.$$

Partial Geometric Series

$$\sum_{i=0}^{k} x^i = \frac{1 - x^{k+1}}{1-x} \text{ for } x \neq 1.$$

- Binomial Theorem:
  - n is a positive integer:

$$(1+x)^n = \sum_{i=0}^{n} \binom{n}{i} x^i.$$

  - n is a real number:

$$(1+x)^n = \sum_{i=0}^{\infty} \binom{n}{i} x^i \quad \text{if } |x| < 1,$$

where

$$\binom{n}{i} = \frac{n^{(i)}}{i!}.$$

- Multinomial Theorem: for a positive integer n:

$$(t_1 + t_2 + \cdots + t_k)^n = \sum \frac{n!}{x_1!x_2!\cdots x_k!} t_1^{x_1} t_2^{x_2} \cdots t_k^{x_k}$$

  where the summation is over all non-negative integers $x_1, x_2, \ldots, x_k$ such that

$$x_1 + x_2 + \cdots x_k = n.$$

- Hypergeometric Identity

$$\sum_{i=0}^{\infty} \binom{a}{i}\binom{b}{n-i} = \binom{a+b}{n}.$$

  There will be only a finite number of terms if a and b are positive integers.

- Taylor Series for the Exponential Function:

$$e^x = \sum_{i=0}^{\infty} \frac{x^i}{i!}$$

- We will also use the following limit:

$$\lim_{n\to\infty} (1 + \frac{x}{n})^n = e^x \quad \text{for all real } x.$$