[Probabilistic Systems Analysis and Applied Probability](#) (2010)

[Probabilistic Systems Analysis and Applied Probability](#) (2013)

[Introduction to Probability (Supplemental Course)](#)

Broad Course Objectives

- develop the art of describing uncertainty in terms of probabilistic models
- develop the skill of probabilistic reasoning

Specific Learning Objectives

- describe the generic structure of probabilistic models and their basic properties

Additional Resources

The Bayesian New Statistics: Hypothesis testing, estimation,meta-analysis, and power analysis from a Bayesian perspective [paper](#)

Combinatorial principles [wiki](#)

Notes on [Reading](#)

1.1 Sets

Amusing anecdote on tension of general use of the word probability

define probability in terms of frequency of occurrence, as a percentage of sucesses in a moderately large number of similar situations

- this may be appropriate in some situations with uncertainty, but many where its not
  - e.g. a scholar who says the iliad and odyssey were composed by the same person with a probability of 90%, this doesnt convey information based on frequencies, because the subject is a one-time event (*difference*), rather this probability is an expression of the scholar's subjective belief
  - many argue subjective beliefs aren't interesting for math and science POVs
    - however people are often making decisions in the presence of uncertainty, and a systematic way of making use of their beliefs is a prereq for successful, or at least consistent decision making
    - the choices and actions of rational people can reveal a lot about the inner held subjective probabilities (*I see another layer of why I want to do this, we're walking into a realm of new possibilities, and dont want to resort to status quo tactics, but also dont want to be wrong/unviable/nonexistent/unreal*)

probabilistic models (in the scope of this material) assign probabilities to collections (sets) of possible outcomes

- that is why set notation & operations are foundational to probability theory and models

sets

- a collection of objects which are the elements of the set,
- If S is a set and x is an element of S,
  - we write $x \in S$
- If x is not an element of S, we write
  - $x \notin S$
- A set can have no elements, the empty set, $\emptyset$
- if S contains a finite number of elements, we write it in braces
  - $S = \{x_1, x_2, \ldots, x_n\}$.
  - eg set of possible outcomes of a die role is $\{1, 2, 3, 4, 5, 6\}$, the set of possible outcomes of a coin toss is $\{H, T\}$ where H is 'heads' and T is 'tails'
- if S contains infinitely many elements which can be enumerated in a list (ie there are as many elements are there are positive integers), we write
  - $S = \{x_1, x_2, \ldots\}$
  - and we say S is countably infinite
    - e.g. the set of even integers can be written as $\{0, 2, -2, 4, -4, \ldots\}$ and is countably infinite
- we can consider the set of all x that have a certain property P, and denote it by
  - $\{x | x \text{ satisfies } P\}$
    - symbol "|" is to be read "such that"
  - eg the set of even integers can be written as $\{k | k/2 \text{ is integer}\}$
  - e.g. the set of all scalars x in the interval $[0, 1]$ can be written as $\{x | 0 \leq x \leq 1\}$.
    - elements x take a continuous range of values, and cannot be written down in a list (theres a proof for this)
    - such as set is said to be **uncountable**
- if every element of a set S is also an element of a set T, we say that S is a subset of T, and we write $S \subset T$ or $T \supset S$
  - if $S \subset T$ & $T \subset S$, the two sets are equal $S = T$
- the universal set, denoted by Ω, which contains all objects that could conceivably be of interest in a particular context
  - when context is specified in terms of set Ω, we only consider sets S that are subsets of Ω, $S \subset \Omega$

Set Operations

- The **complement** of a set S with respect to the universe Ω is the set $\{x \in \Omega \mid x \notin S\}$
  - of all elements of Ω that do not belong to S and is denoted by S^c
  - note: $\Omega^c = \emptyset$.
- The **union** of two sets S and T is the set of all elements that belong to S or T (or both), and is denoted by $S \cup T$
  - $S \cup T = \{x \mid x \in S \text{ or } x \in T\}$
- The **intersection** of two sets S and T is the set of all elements that belong to both S and T, and is denoted by $S \cap T$
  - $S \cap T = \{x \mid x \in S \text{ and } x \in T\}$
- When considering the union of several, even infinitiely many sets
  - e.g. if for every positive integer n, we are given a set S_n

- $\bigcup_{n=1}^{\infty} S_n = S_1 \cup S_2 \cup \cdots = \{x \mid x \in S_n \text{ for some n}\}$ and,
- $\bigcap_{n=1}^{\infty} S_n = S_1 \cap S_2 \cap \cdots = \{x \mid x \in S_n \text{ for all n}\}$

- Two sets are said to be **disjoint** if their intersection is empty, or more generally, several sets ar said to be disjoint if no two of them have a common element
  - a collection of sets is said to be a *partition* of a set S if the sets in the collection are disjoint and their union is S
- if x and y are two objects, we use (x, y) to denote the **ordered pair** of x and y,
- the set of scalars (real numbers) is denoted as $\Re$, the set of pairs (two dimensional plane) is denoted as $\Re^2$or triplets (three dimensional plane) is denoted as $\Re^3$
- Sets and operations visualized:

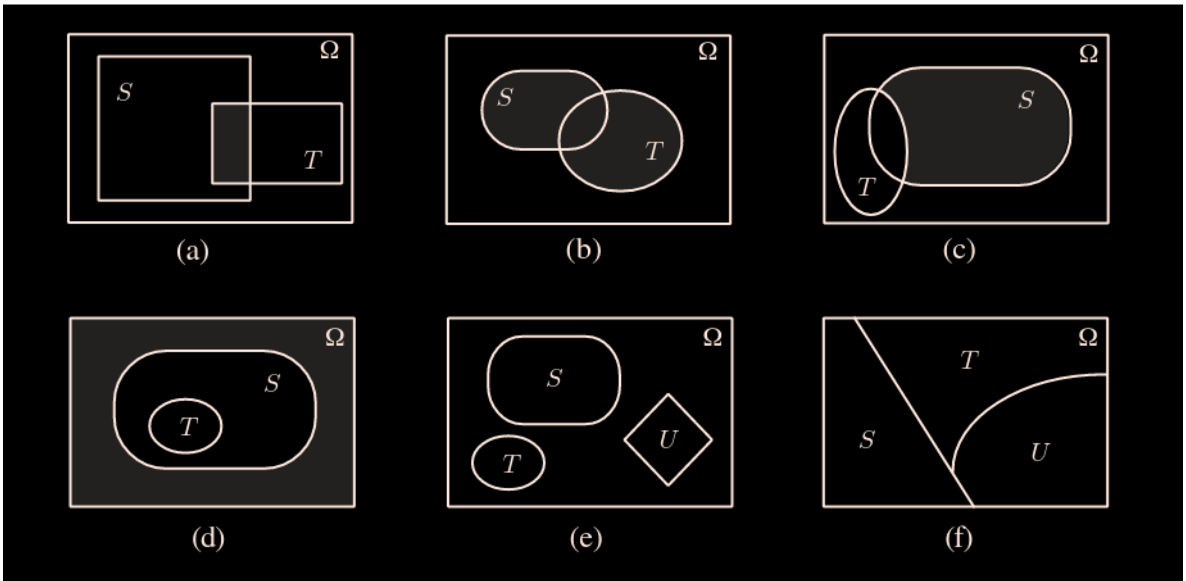


Figure 1.1: Examples of Venn diagrams. (a) The shaded region is $S \cap T$. (b) The shaded region is $S \cup T$. (c) The shaded region is $S \cap T^c$. (d) Here, $T \subset S$. The shaded region is the complement of S. (e) The sets S, T, and U are disjoint. (f) The sets S, T, and U form a partition of the set Ω.

The algebra of Sets

Some examples:

$S \cup T = T \cup S$	$S \cup (T \cup U) = (S \cup T) \cup U$
$S \cap (T \cup U) = (S \cap T) \cup (S \cap U)$	$S \cup (T \cap U) = (S \cup T) \cap (S \cup U)$
$(S^c)^c = S$	$S \cap S^c = \emptyset$
$S \cup \Omega = \Omega$	$S \cap \Omega = S$

de Morgan's laws

Two particularly useful properties are given by de Morgan's laws:

The first law	The second law
$(S \cup T)^c = S^c \cap T^c$	$(S \cap T)^c = S^c \cup T^c$
Alternative notation: $\overline{A \cup B} = \bar{A} \cap \bar{B}$	Alternative notation: $\overline{A \cap B} = \bar{A} \cup \bar{B}$
"The complement of the union is equal to the intersection of the complements"	"The complement of the intersection is equal to the union of the complements"

The Set Notation Proof for N Sets:

The first law	The second law (the converse of first law)
Set notation (for n number of sets): $\left(\bigcup_n S_n\right)^c = \bigcap_n S_n^c$	Set notation: $\left(\bigcap_n S_n\right)^c = \bigcup_n S_n^c$
Suppose: $x \in (\cup*{n}S*{n})^c	$x \in (\cap_n S_n)^c$
Then, $x \notin \cup_n S_n$	$\implies x \notin \cap_n S_n$
which implies for every n, we have $x \notin S_n$	$\implies$ for every n, $x \notin S_n$
thus, x belongs to the complement of every S_n	$\implies$ x belongs to the complement of every S_n
and, $x_n \in \cap_n S_n^c$	$\implies x_n \in \cup_n S_n^c$
$\therefore (\cup_n S_n)^c \subset \cap_n S_n^c$	$\therefore (\cap_n S_n)^c \subset \cup_n S_n^c$

Note: this proof is for n Sets, it's helpful to think of just 2 sets as a baseline, though the proof is the same, the proof of 2 sets explained here

These laws can be stated in terms of logical convention as

First Law	Second Law
$\neg(p \wedge q) = \neg p \vee \neg q$	$\neg(p \vee q) = \neg p \wedge \neg q$
"Neither p nor q is equal to not p and not q"	"Not (p and q) is equal to Not p or Not q"

where $\neg$ means not, $\wedge$ means and, $\vee$ means or

Note: it's helpful to recreate de Morgan's laws visually, with venn diagrams

1.2 Probabilistic Models

A **probabilistic model** is a mathematical description of an uncertain situation (in accordance to a fundamental framework, explained in this section)

The two main elements of a probabilistic model:

- the **sample space** Ω (Omega), which is the set of all possible outcomes of an experiment
- the **probability law**, which assigns to a set of A possible outcomes (also called an **event**) a nonnegative number $P(A)$ (Called the *probability of A) that encodes our *knowledge or belief about the collective 'likelihood'* of the elements of A.

- **The probability law** must satisfy certain properties, as we'll see

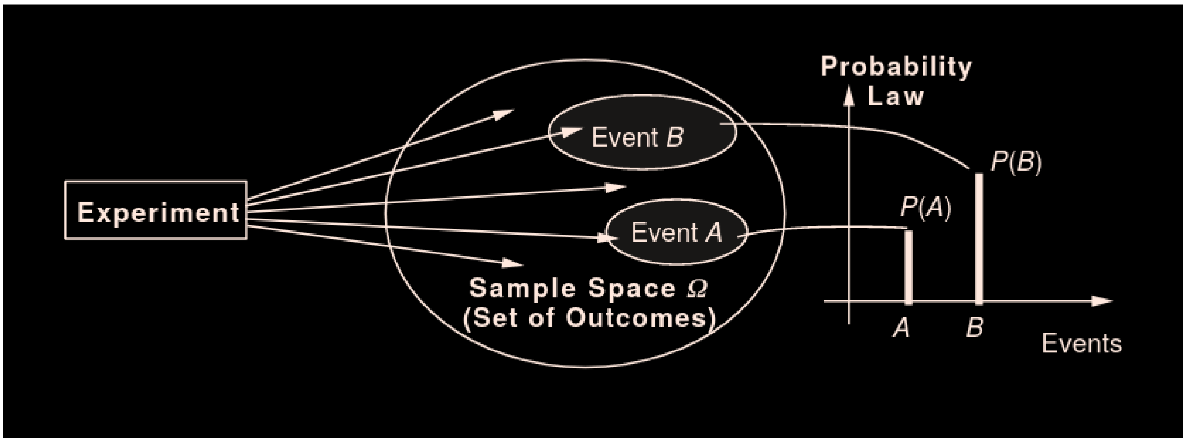


Figure 1.2: The main ingredients of a probabilistic model.

Sample Spaces and Events

- Every probabilistic model involves an underlying process, called the **experiment**, that produces exactly ONE out of several possible **outcomes**
    - no restrictions on what constitutes an experiment
        - a single coin toss, three tosses, an infinite sequence of tosses
        - however in this formulation of a probabilistic model, theres only a single experiment, rather than 3 experiments
- The set of all possible outcomes is called the **sample space** of the experiment, and is denoted by Ω
    - sample space of an experiment may consist of a finite or infinite number of possible outcomes
        - finite spaces are conceptually and mathematically simpler
        - sample spaces with an infinite number of elements are quite common
            - e.g. throwing a dart on a square target and viewing the point of impact as the outcome
- A subset of the sample space, that is, a collection of possible outcomes, is called an **event**
    - in practice, any collection of possible outcomes, including the entire sample space Ω and its complement, the empty set $\emptyset$, may qualify as an event
        - there are theoretical examples in uncountably infinite sample spaces, where this is not the case, but this theoretical point can ignored in practice
- choosing an appropriate sample space
    - regardless of their number, different elements of the sample space should be distinct and **mutually exclusive**
        - the experiment must have have a unique outcome
        - e.g. the sample space of a die cannot contain "1 or 3" as a possible outcome and "1 or 4" as another outcome, when the roll is a 1 the outcome of the experiment would not be unique
    - a given physical situation can be modeled in several different ways, depending onthe kind of questions we're interested in
    - generally, the sample space must be **collectively exhaustive**
        - i.e. no matter what happens in the experiment, we always obtain an outcome that has been included in the sample space

- the sample space should have enough detail to distinguish between all outcomes of interest to the modeler, while avoiding irrelevant details

- example: a game where you receive $1 each time a head comes up for 10 tosses
  - only the total number of heads in a ten toss sequence matters
  - the sample space consists of ten possible outcomes 1,...,10
- example 2: a game where you receive $1 for *every* coin toss, up to and including the first time a head comes up, then the amount you get doubles each time a head comes up after
  - in this game the order of heads and tails is also important
  - a more detailed sample space is necessary, consisting of every possible ten-long sequence of heads and tails

Sequential Models

many experiments are sequential in character

- e.g. tossing a coin 3 times, observing stock prices for 5 subsequent days, receiving a code with 8 digits
- useful to use a **tree-based sequential description**

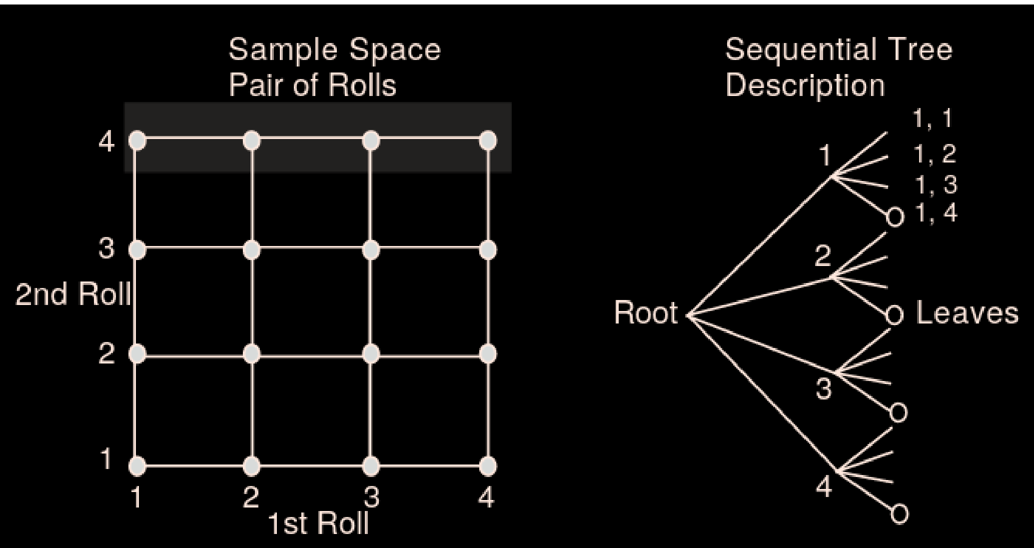


Figure 1.3: Two equivalent descriptions of the sample space of an experiment involving two rolls of a 4-sided die. The possible outcomes are all the ordered pairs of the form (i, j), where i is the result of the first roll, and j is the result of the second. These outcomes can be arranged in a 2-dimensional grid as in the figure on the left, or they can be described by the tree on the right, which reflects the sequential character of the experiment. Here, each possible outcome corresponds to a leaf of the tree and is associated with the unique path from the root to that leaf. The shaded area on the left is the event $\{(1,4),\ (2,4),\ (3,4),\ (4,4)\}$ that the result of the second roll is 4. That same event can be described as a set of leaves, as shown on the right. Note also that every node of the tree can be identified with an event, namely, the set of all leaves downstream from that node. For example, the node labeled by a 1 can be identified with the event $\{(1,1),(1,2),(1,3),(1,4)\}$ that the result of the first roll is 1.

Probability Laws

to complete the probabilistic model, we must introduce

- this specifies the 'likelihood' of any outcome, or any set of possible outcomes (an event)
- the probability law assigns to every event A a number $P(A)$, called the probability of A, satisfying the following axioms:

Probability Axioms

Name	Description
Nonnegativity	$P(A) \geq 0$ for every event A
Additivity:	If A and B are two **disjoint** events, then the probability of their union satisfies $P(A \cup B) = P(A) + P(B)$. Furthermore, if the sample space has an infinite number of elements and $A_1, A_2, \ldots$ is a sequence of disjoint events, then the probability of their union satisfies $P(A_1 \cup A_2 \cup \ldots) = P(A_1) + P(A_2) + \ldots$
Normalization	The probability of the entire sample space Ω is equal to 1, that is $P(\Omega) = 1$

Visualize a probability law as a unit of mass 'spread' over a sample space. $P(A)$ is the total mass assigned collectively to the elements of A. The additivity axiom can be seen inutitively; the total mass in a sequence of disjoint events is the sum of their individual masses.

The relative frequency interpretation: $P(A) = 2/3$ represents the belief that event A wil materialize about two thirds out of a large number of repetitions of the experiment. This inrepretation can sometimes be helpful, but not always appropriate & will be revisited in the study of limit theorems.

Additional Properties of a Probability Law

Not included in the axioms but can be derived

Axiom	Consequenc
Nonnegativity: $P(A) \geq 0$	$P(A \leq 1)$
Normalization:$P(\Omega) = 1$	$P(\emptyset) = 0$
Additivity: $P(A \cup B) = P(A) + P(B)$ (for disjoint events)	$P(A) + P(A^c) = 1$ $P(A \cup B \cup C) = P(A) + P(B) + P(C)$ and similarly for k disjoint events (see Discrete Probability Law)

video explanation

Example 1: the normalization and additivity axioms imply:

$$1 = P(\Omega) = P(\Omega \cup \emptyset) = P(\Omega) + P(\emptyset) = 1 + P(\emptyset)$$

This shows the probability of the empty event is 0:

$$P(\emptyset) = 0$$

Example 2: consider 3 disjoint events $A_1, A_2, \text{ and } A_3$. We can use the additivity axiom for two disjoint events repeatedly, to obtain:

$$\begin{aligned} P(A_1 \cup A_2 \cup A_3) &= P(A_1 \cup (A_2 \cup A_3)) \\ &= P(A_1) + P(A_2 \cup A_3) \\ &= P(A_1) + P(A_2) + P(A_3) \end{aligned}$$

The probability of the union of finitely many disjoint events is always equal to the sum of the probabilities of these events

Discrete Models

Examples of constructing a probability law, starting with some common sense assumptions about a model

Example 1:

Coin Tosses: Consider an experiment involving a single coin toss. Two possible outcomes: heads (H) and tails (T), the sample space is $\Omega = \{H, T\}$ and the events are

$$\{H, T\}, \{H\}, \{T\}, \emptyset.$$

If the coin is fair, i.e., if we believe that heads and tails are 'equally likely', we should assign equal probabilities to the two possible outcomes, specifying: $P(\{H\}) = P(\{T\}) = 0.5$. The additivity axiom implies

$$P(\{H, T\}) = P(\{H\}) + P(\{T\}) = 1,$$

which is consistent with the normalization axiom.

The probability law is given by

$$P(\{H, T\}) = 1, \qquad P(\{H\}) = 0.5, \qquad P(\{T\}) = 0.5, \qquad P(\emptyset) = 0,$$

and satisfies all three axioms.

Example 2:

Three Coin Tosses: The outcome will now be a 3-long string of heads or tails. The sample space is

$$\Omega = \{HHH,\ HHT,\ HTH,\ HTT,\ THH,\ THT,\ TTH,\ TTT\}.$$

We assume each possible outcome has the same probability of 1/8.

Let's construct a probability law that satisfies the three axioms.

For example, let's consider the event:

$$A = \{\text{exactly 2 heads occur}\} = \{HHT,\ HTH,\ THH\}.$$

Using additivity, the probability of A is the sum of the probabilities of its elements:

$$\begin{aligned} P(\{HHT,\ HTH,\ THH\}) &= P(\{HHT\}) + P(\{HTH\}) + P(\{THH\}) \\ &= \frac{1}{8} + \frac{1}{8} + \frac{1}{8} \\ &= \frac{3}{8}. \end{aligned}$$

Similarly, the probability of any event is equal to 1/8 times the number of possible outcomes contained in the event. This defines a probability law that satisfies the three axioms.

By using the additivity axiom and by generalizing the reasoning in the preceding example, we reach the following conclusion:

Discrete Probability Law

If the sample space with a finite number of possible outcomes, then the probability law is specified by the probabilities of the events that consist of a single element. In particular, the probability of any event $\{s_1, s_2, \ldots s_n\}$ is the sum of the probabilities of its elements:

$$P(\{s_1, s_2, \ldots s_n\}) = P(\{s_1\}) + P(\{s_2\}) + \cdots + P(\{s_n\}).$$

Discrete Uniform Probability Law

In the special case where the probabilities $P(\{s_1\}), \ldots, P(\{s_n\})$ are all the same (by necessity, equal to $1/n$, in view of the normalization axiom), we obtain the Discrete *Uniform* Probability Law:

If the sample space consists of n possible outcomes which are equally likely (i.e., all single-element events have the same probabilitiy), then the probability of any event A is given by

$$P(A) = \frac{\text{Number of Elements of } A}{n}.$$

Some examples of sample spaces and probability laws.

Example 1:

Dice: Consider the experiment of rolling a pair of 4-sided dice. we assume the dice are fair, and we interpret this assumption to mean that each of the sixteen possible outcomes [ordered pairs (i, j), with $i, j = 1, 2, 3, 4$], has the sample probability of $1/16$. To calculate the probability of an event, we must count the number of elements of event and divide by 16 (the total number of possible outcomes).

Here are some event probabilities calculated in this way:

note: it may be helpful to draw out all 16 possible outcomes to illustrate these calculations

$$\begin{aligned}
P(\{\text{the sum of the rolls is even}\}) &= 8/16 = 1/2 \\
P(\{\text{the sum of the rolls is odd}\}) &= 8/16 = 1/2 \\
P(\{\text{the first roll is equal to the second}\}) &= 4/16 = 1/4 \\
P(\{\text{the first roll is larger than the second}\}) &= 6/16 = 3/8 \\
P(\{\text{at least one roll is equal to 4}\}) &= 7/16.
\end{aligned}$$

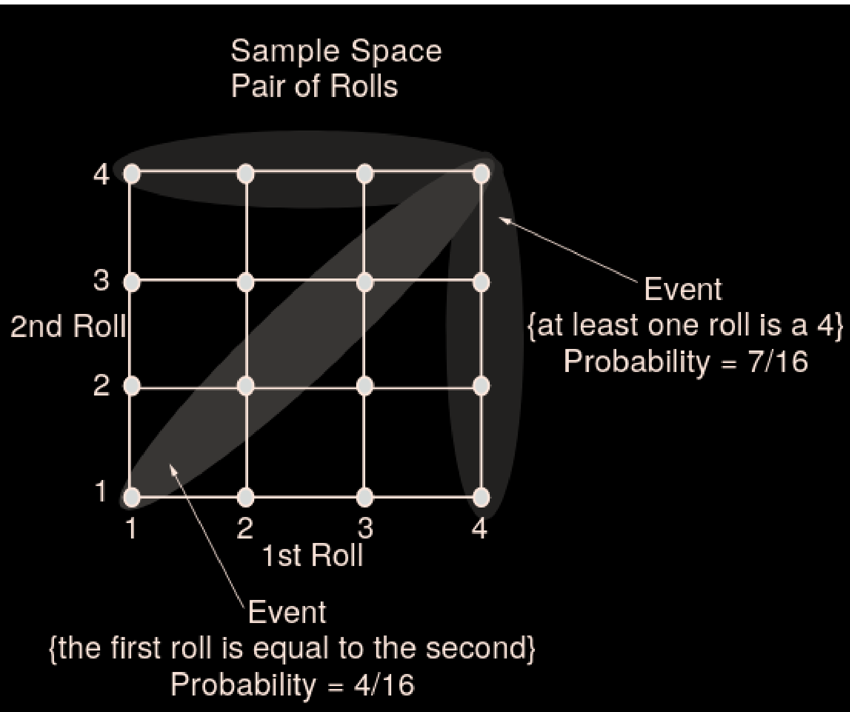


Figure 1.4: Various events in the experiment of rolling a pair of 4-sided dice, and their probabilities, calculated according to the discrete uniform law.

Continuous Models

Probabilistic models with continuous sample spaces differ from their discrete as the probabilities of single-element events may not be sufficient to characterize the probability law

Example 1:

A wheel of fortune whos possible outcomes of a spin (the experiment) are the numbers in the interval $\Omega = [0, 1]$

A fair wheel, means all outcomes equally likely

What is the probability of th event consisting of a single element?

It can't be positive, bc, using the additivity axiom, it would follow that events with a sufficiently large number of elements would have probability larger than 1

Therefore, the probability of any event that consists of a single element must be 0

It would make sense to assign probability $b - a$ to any subinterval of $[a, b]$ of $[0, 1]$ and to caclualte hte probabilty of a more complicated set by evaluating its 'length'[^1]

Example 2:

Romeo and Juliet have a date at a set time, and will arrive late, between 0 and 1 hour, with all pairs of delays equally likely, the first to arrive will wait for 15 minutes and will leave if the other has not arrived. What is the probability they will meet?

The sample space, the quare $\Omega = [0, 1] \times [0, 2]$, the elements are the possible pairs of delays for both people

"equally likely" pairs of delays can be interpreted as letting the probabily of a subset of Ω be equal to its area

- this probability law satisfies the 3 probability axioms (nonnegativity, additivity, normalization)

the event that Romeo and Juliet will meet is the shaded region, with a calculated probabilty of $7/16$

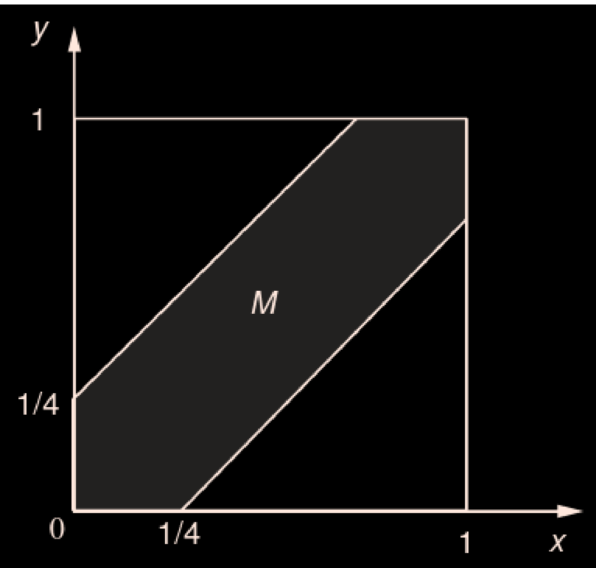


Figure 1.5: The event M that Romeo and Juliet will arrive within 15 minutes of each other (cf. Example 1.5) is

$$M = \Big\{(x,y) \;\Big|\; |x-y| \le 1/4,\ 0 \le x \le 1,\ 0 \le y \le 1\Big\},$$

and is shaded in the figure. The area of M is 1 minus the area of the two unshaded triangles, or $1-(3/4)\cdot(3/4) = 7/16$. Thus, the probability of meeting is $7/16$.

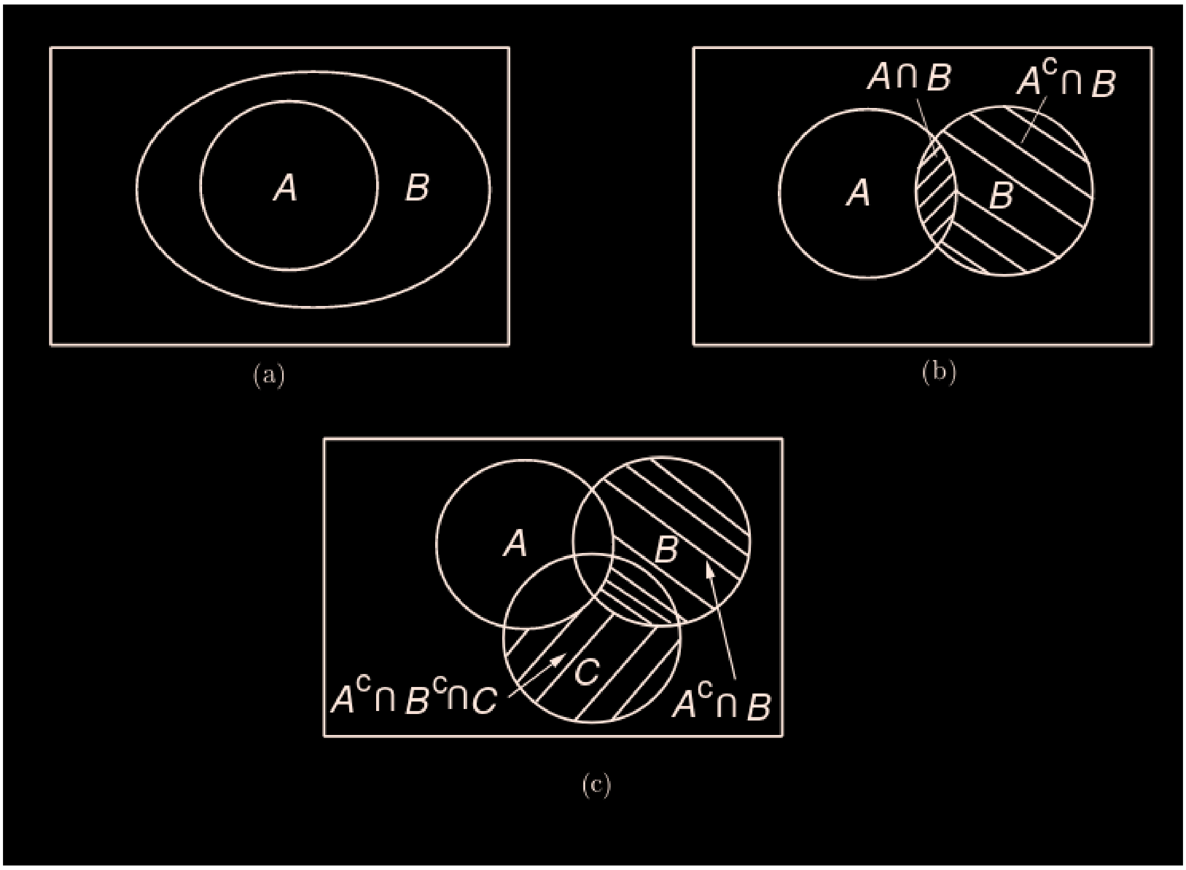
Properties of Probability Laws

Laws Deduced from the axioms, some include:

Consider a probability law, let $A,\ B, \text{and } C$ be events.

- (Property a) If $A \subset B$, then $P(A) \le P(B)$.
- (Property b) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ (Inclusion-Exclusion Principle)
- (Property c) $P(A \cup B) \le P(A) + P(B)$ (when would it be less than?)
- (Property d) $P(A \cup B \cup C) = P(A) + P(A^c \cap B) + P(A^c \cap B^c \cap C)$. (Is it assumed continuous events are not disjoint?)

Note: It's helpful to visualize these with venn diagrams for events $A,\ B, \text{and } C$.



- (diagram a) if $A \subset B$, then B is the union of the two disjoint events A and $A^c \cap B$
  - therefore, by the additivity axiom:
    $$P(B) = P(A) + P(A^c \cap B) \geq P(A),$$
    where the inequality follows from the nonnegativity axiom, and verifies (property a)
- (diagram b)

we can express the(Property c) can be applied repeatedly to obtain the inequality

$$P(A_1 \cup A_2 \cup \cdots \cup A_n) \leq \sum_{i=1}^{n} P(A_i).$$

In more detail, applying property (c) to the sets A_1 and $A_2 \cup \cdots \cup A_n$, to obtain

$$P(A_1 \cup A_2 \cup \cdots \cup A_n) \leq P(A_1) + P(A_2 \cup \cdots \cup A_n),$$

continue similarly, and finally add.

Models and Reality