Probability and Stochastic Processes

Chapter 1

Section 1.1

Set Theory

Theorem 1.1

De Morgan's law relates all three basic operations:

$$(A \cup B)^c = A^c \cap B^c.$$

Proof: Theorem 1.1

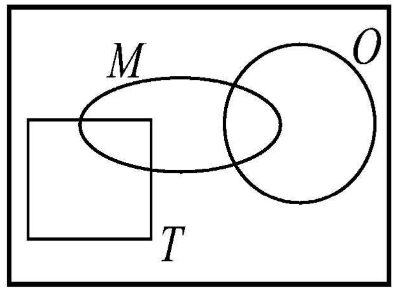
There are two parts to the proof:

- To show $(A \cup B)^c \subset A^c \cap B^c$, suppose $x \in (A \cup B)^c$. That implies $x \notin A \cup B$. Hence, $x \notin A$ and $x \notin B$, which together imply $x \in A^c$ and $x \in B^c$. That is, $x \in A^c \cap B^c$.

- To show $A^c \cap B^c \subset (A \cup B)^c$, suppose $x \in A^c \cap B^c$. In this case, $x \in A^c$ and $x \in B^c$. Equivalently, $x \notin A$ and $x \notin B$ so that $x \notin A \cup B$. Hence, $x \in (A \cup B)^c$.

Quiz 1.1

A pizza at Gerlanda's is either regular (R) or Tuscan (T). In addition, each slice may have mushrooms (M) or onions (O) as described by the Venn diagram at right. For the sets specified below, shade the corresponding region of the Venn diagram.



(1) R

(2) $M \cup O$

(3) $M \cap O$

(4) $R \cup M$

(5) $R \cap M$

(6) $T^c - M$

Section 1.2

Applying Set Theory to Probability

Example 1.1

An experiment consists of the following procedure, observation, and model:

- Procedure: Flip a coin and let it land on a table.
- Observation: Observe which side (head or tail) faces you after the coin lands.
- Model: Heads and tails are equally likely. The result of each flip is unrelated to the results of previous flips.

Example 1.2

Flip a coin three times. Observe the sequence of heads and tails.

Example 1.3

Flip a coin three times. Observe the number of heads.

Definition 1.1 Outcome

An outcome *of an experiment is any possible observation of that experiment.*

Definition 1.2 Sample Space

The sample space *of an experiment is the finest-grain, mutually exclusive, collectively exhaustive set of all possible outcomes.*

Mutually exclusive if and only if $A_i \cap A_j = \phi, \quad i \neq j.$

Collectively exhaustive if and only if $A_1 \cup A_2 \cup \cdots \cup A_n = S.$

Example 1.4

- The sample space in Example 1.1 is $S = \{h, t\}$ where h is the outcome "observe head," and t is the outcome "observe tail."

- The sample space in Example 1.2 is

$$S = \{hhh, hht, hth, htt, thh, tht, tth, ttt\}$$

- The sample space in Example 1.3 is $S = \{0, 1, 2, 3\}$.

Definition 1.3 Event

An event *is a set of outcomes of an experiment.*

Example 1.6

Suppose we roll a six-sided die and observe the number of dots on the side facing upwards. We can label these outcomes $i = 1, \ldots, 6$ where i denotes the outcome that i dots appear on the up face. The sample space is $S = \{1, 2, \ldots, 6\}$. Each subset of S is an event. Examples of events are

- The event $E_1 = \{$Roll 4 or higher$\} = \{4, 5, 6\}$.
- The event $E_2 = \{$The roll is even$\} = \{2, 4, 6\}$.
- $E_3 = \{$The roll is the square of an integer$\} = \{1, 4\}$.

Example 1.7

Wait for someone to make a phone call and observe the duration of the call in minutes. An outcome x is a nonnegative real number. The sample space is $S = \{x|x \geq 0\}$. The event "the phone call lasts longer than five minutes" is $\{x|x > 5\}$.

Example 1.8

A short-circuit tester has a red light to indicate that there is a short circuit and a green light to indicate that there is no short circuit. Consider an experiment consisting of a sequence of three tests. In each test the observation is the color of the light that is on at the end of a test. An outcome of the experiment is a sequence of red (r) and green (g) lights. We can denote each outcome by a three-letter word such as rgr, the outcome that the first and third lights were red but the second light was green. We denote the event that light n was red or green by R_n or G_n. The event $R_2 = \{grg, grr, rrg, rrr\}$. We can also denote an outcome as an intersection of events R_i and G_j. For example, the event $R_1 G_2 R_3$ is the set containing the single outcome $\{rgr\}$.

Definition 1.4 Event Space

An event space *is a collectively exhaustive, mutually exclusive set of events.*

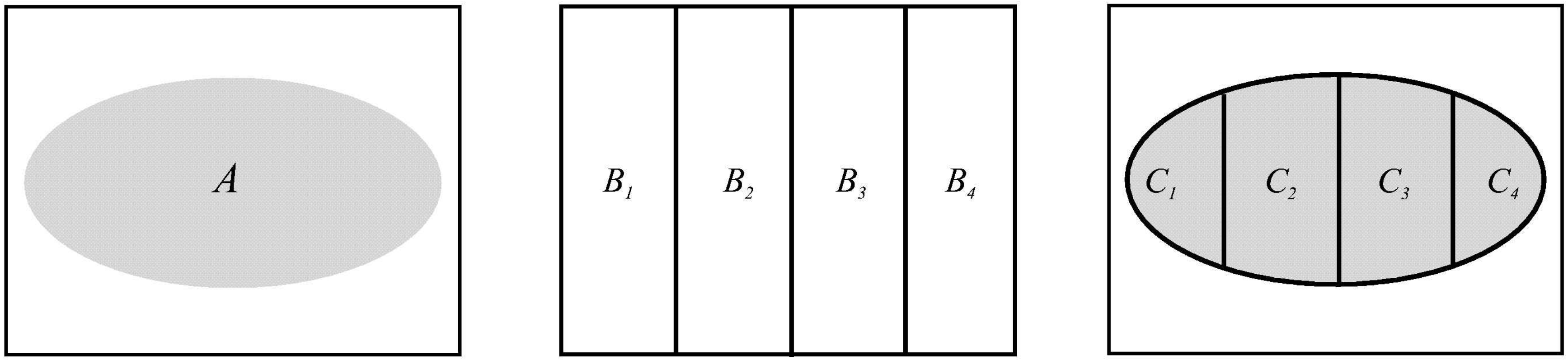
Example 1.9 Problem

Flip four coins, a penny, a nickel, a dime, and a quarter. Examine the coins in order (penny, then nickel, then dime, then quarter) and observe whether each coin shows a head (h) or a tail (t). What is the sample space? How many elements are in the sample space?

Example 1.10

Continuing Example 1.9, let $B_i = \{\text{outcomes with } i \text{ heads}\}$. Each B_i is an event containing one or more outcomes. For example, $B_1 = \{ttth, ttht, thtt, httt\}$ contains four outcomes. The set $B = \{B_0, B_1, B_2, B_3, B_4\}$ is an event space. Its members are mutually exclusive and collectively exhaustive. It is not a sample space because it lacks the finest-grain property. Learning that an experiment produces an event B_1 tells you that one coin came up heads, but it doesn't tell you which coin it was.

Figure 1.1



In this example of Theorem 1.2, the event space is $B = \{B_1, B_2, B_3, B_4\}$ and $C_i = A \cap B_i$ for $i = 1, \ldots, 4$. It should be apparent that $A = C_1 \cup C_2 \cup C_3 \cup C_4$.

Theorem 1.2

For an event space $B = \{B_1, B_2, \ldots\}$ and any event A in the sample space, let $C_i = A \cap B_i$. For $i \neq j$, the events C_i and C_j are mutually exclusive and

$$A = C_1 \cup C_2 \cup \cdots .$$

Example 1.11

In the coin-tossing experiment of Example 1.9, let A equal the set of outcomes with less than three heads:

$$A = \{tttt, httt, thtt, ttht, ttth, hhtt, htht, htth, tthh, thth, thht\}.$$

From Example 1.10, let $B_i = \{$outcomes with i heads$\}$. Since $\{B_0, \ldots, B_4\}$ is an event space, Theorem 1.2 states that

$$A = (A \cap B_0) \cup (A \cap B_1) \cup (A \cap B_2) \cup (A \cap B_3) \cup (A \cap B_4)$$

In this example, $B_i \subset A$, for $i = 0, 1, 2$. Therefore $A \cap B_i = B_i$ for $i = 0, 1, 2$. Also, for $i = 3$ and $i = 4$, $A \cap B_i = \phi$ so that $A = B_0 \cup B_1 \cup B_2$, a union of disjoint sets. In words, this example states that the event "less than three heads" is the union of events "zero heads," "one head," and "two heads."

Quiz 1.2

Monitor three consecutive phone calls going through a telephone switching office. Classify each one as a voice call (v) if someone is speaking, or a data call (d) if the call is carrying a modem or fax signal. Your observation is a sequence of three letters (each letter is either v or d). For example, two voice calls followed by one data call corresponds to vvd. Write the elements of the following sets:

(1) $A_1 = \{$first call is a voice call$\}$

(2) $B_1 = \{$first call is a data call$\}$

(3) $A_2 = \{$second call is a voice call$\}$

(4) $B_2 = \{$second call is a data call$\}$

(5) $A_3 = \{$all calls are the same$\}$

(6) $B_3 = \{$voice and data alternate$\}$

(7) $A_4 = \{$one or more voice calls$\}$

(8) $B_4 = \{$two or more data calls$\}$

For each pair of events A_1 and B_1, A_2 and B_2, and so on, identify whether the pair of events is either mutually exclusive or collectively exhaustive or both.

Section 1.3

Probability Axioms

Definition 1.5 Axioms of Probability

A probability measure $P[\cdot]$ is a function that maps events in the sample space to real numbers such that

Axiom 1 *For any event A, $P[A] \geq 0$.*

Axiom 2 $P[S] = 1$.

Axiom 3 *For any countable collection $A_1, A_2, \ldots$ of mutually exclusive events*

$$P[A_1 \cup A_2 \cup \cdots] = P[A_1] + P[A_2] + \cdots .$$

Section 1.4

Some Consequences of the Axioms

Theorem 1.7

The probability measure $P[\cdot]$ satisfies

(a) $P[\phi] = 0$.

(b) $P[A^c] = 1 - P[A]$.

(c) For any A and B (not necessarily disjoint),

$$P[A \cup B] = P[A] + P[B] - P[A \cap B].$$

(d) If $A \subset B$, then $P[A] \leq P[B]$.

Theorem 1.8

For any event A, and event space $\{B_1, B_2, \ldots, B_m\}$,

$$P[A] = \sum_{i=1}^{m} P[A \cap B_i].$$

Proof: Theorem 1.8

The proof follows directly from Theorem 1.2 and Theorem 1.4. In this case, the disjoint sets are $C_i = \{A \cap B_i\}$.

Theorem 1.2

For an event space $B = \{B_1, B_2, \ldots\}$ and any event A in the sample space, let $C_i = A \cap B_i$. For $i \neq j$, the events C_i and C_j are mutually exclusive and

$$A = C_1 \cup C_2 \cup \cdots .$$

Theorem 1.4

If $A = A_1 \cup A_2 \cup \cdots \cup A_m$ and $A_i \cap A_j = \phi$ for $i \neq j$, then

$$P[A] = \sum_{i=1}^{m} P[A_i].$$

Quiz 1.4

Monitor a phone call. Classify the call as a voice call (V) if someone is speaking, or a data call (D) if the call is carrying a modem or fax signal. Classify the call as long (L) if the call lasts for more than three minutes; otherwise classify the call as brief (B). Based on data collected by the telephone company, we use the following probability model: $P[V] = 0.7$, $P[L] = 0.6$, $P[VL] = 0.35$. Find the following probabilities:

(1) $P[DL]$

(2) $P[D \cup L]$

(3) $P[VB]$

(4) $P[V \cup L]$

(5) $P[V \cup D]$

(6) $P[LB]$

Section 1.5

Conditional Probability

Definition 1.6 Conditional Probability

The conditional probability of the event A *given the occurrence of the event* B *is*

$$P[A|B] = \frac{P[AB]}{P[B]}.$$

Example 1.15 (Why conditional probability?)

- Testing two circuits in the same silicon wafer
  - Accepted (a) or rejected (b)
  - $S=\{rr,ra,ar,aa\}$
  - Let B = Event that the first chip tested is rejected = $\{rr,ra\}$
    A = Event that the second circuit is a failure = $\{rr,ar\}$
- A priori probabilities
  - $P[rr]$=0.01, $P[ra]$=0.01, $P[ar]$=0.01, $P[aa]$=0.97
  - $P[A] = P[rr] + P[ar] = 0.02$, $P[B] = P[rr] + P[ra] = 0.02$
  - $P[AB] = P[\text{both rejected}] = P[rr] = 0.01$
- The conditional probability
  - $$P[A\,|\,B]=\frac{P[AB]}{P[B]}=0.01/0.02=0.5$$
  - A posteriori probability

Theorem 1.9

A conditional probability measure $P[A|B]$ has the following properties that correspond to the axioms of probability.

Axiom 1: $P[A|B] \geq 0$.

Axiom 2: $P[B|B] = 1$.

Axiom 3: If $A = A_1 \cup A_2 \cup \cdots$ with $A_i \cap A_j = \phi$ for $i \neq j$, then

$$P[A|B] = P[A_1|B] + P[A_2|B] + \cdots$$

Theorem 1.10 Law of Total Probability

For an event space $\{B_1, B_2, \ldots, B_m\}$ with $P[B_i] > 0$ for all i,

$$P[A] = \sum_{i=1}^{m} P[A|B_i]P[B_i].$$

Proof: Theorem 1.10

This follows from Theorem 1.8 and the identity $P[AB_i] = P[A|B_i]P[B_i]$, which is a direct consequence of the definition of conditional probability.

Theorem 1.8

For any event A, and event space $\{B_1, B_2, \ldots, B_m\}$,

$$P[A] = \sum_{i=1}^{m} P[A \cap B_i].$$

Example 1.19 Problem

A company has three machines B_1, B_2, and B_3 for making 1 kΩ resistors. It has been observed that 80% of resistors produced by B_1 are within 50 Ω of the nominal value. Machine B_2 produces 90% of resistors within 50 Ω of the nominal value. The percentage for machine B_3 is 60%. Each hour, machine B_1 produces 3000 resistors, B_2 produces 4000 resistors, and B_3 produces 3000 resistors. All of the resistors are mixed together at random in one bin and packed for shipment. What is the probability that the company ships a resistor that is within 50 Ω of the nominal value?

Example 1.19 Solution

Let $A = \{$resistor is within 50 Ω of the nominal value$\}$. Using the resistor accuracy information to formulate a probability model, we write

$$P[A|B_1] = 0.8, \quad P[A|B_2] = 0.9, \quad P[A|B_3] = 0.6$$

The production figures state that $3000 + 4000 + 3000 = 10{,}000$ resistors per hour are produced. The fraction from machine B_1 is $P[B_1] = 3000/10{,}000 = 0.3$. Similarly, $P[B_2] = 0.4$ and $P[B_3] = 0.3$. Now it is a simple matter to apply the law of total probability to find the accuracy probability for all resistors shipped by the company:

$$\begin{aligned} P[A] &= P[A|B_1]P[B_1] + P[A|B_2]P[B_2] + P[A|B_3]P[B_3] \\ &= (0.8)(0.3) + (0.9)(0.4) + (0.6)(0.3) = 0.78. \end{aligned}$$

For the whole factory, 78% of resistors are within 50 Ω of the nominal value.

Theorem 1.11 Bayes' theorem

$$P[B|A] = \frac{P[A|B]P[B]}{P[A]}.$$

Proof: Theorem 1.11

$$P[B|A] = \frac{P[AB]}{P[A]} = \frac{P[A|B]P[B]}{P[A]}.$$

Example 1.20 Problem

In Example 1.19 about a shipment of resistors from the factory, we learned that:

- The probability that a resistor is from machine B_3 is $P[B_3] = 0.3$.
- The probability that a resistor is *acceptable*, i.e., within 50 Ω of the nominal value, is $P[A] = 0.78$.
- Given that a resistor is from machine B_3, the conditional probability that it is acceptable is $P[A|B_3] = 0.6$.

What is the probability that an acceptable resistor comes from machine B_3?

Example 1.20 Solution

Now we are given the event A that a resistor is within 50 Ω of the nominal value, and we need to find $P[B_3|A]$. Using Bayes' theorem, we have

$$P[B_3|A] = \frac{P[A|B_3]P[B_3]}{P[A]}.$$

Since all of the quantities we need are given in the problem description, our answer is

$$P[B_3|A] = (0.6)(0.3)/(0.78) = 0.23.$$

Similarly we obtain $P[B_1|A] = 0.31$ and $P[B_2|A] = 0.46$. Of all resistors within 50 Ω of the nominal value, only 23% come from machine B_3 (even though this machine produces 30% of all resistors). Machine B_1 produces 31% of the resistors that meet the 50 Ω criterion and machine B_2 produces 46% of them.

Section 1.6

Independence

Definition 1.7 Two Independent Events

Events A *and* B *are* independent *if and only if*

$$P[AB] = P[A]P[B].$$

Example 1.21 Problem

Suppose that for the three lights of Example 1.8, each outcome (a sequence of three lights, each either red or green) is equally likely. Are the events R_2 that the second light was red and G_2 that the second light was green independent? Are the events R_1 and R_2 independent?

Example 1.21 Solution

Each element of the sample space

$$S = \{rrr, rrg, rgr, rgg, grr, grg, ggr, ggg\}$$

has probability $1/8$. Each of the events

$$R_2 = \{rrr, rrg, grr, grg\} \quad \text{and} \quad G_2 = \{rgr, rgg, ggr, ggg\}$$

contains four outcomes so $P[R_2] = P[G_2] = 4/8$. However, $R_2 \cap G_2 = \phi$ and $P[R_2 G_2] = 0$. That is, R_2 and G_2 must be disjoint because the second light cannot be both red and green. Since $P[R_2 G_2] \neq P[R_2]P[G_2]$, R_2 and G_2 are not independent. Learning whether or not the event G_2 (second light green) occurs drastically affects our knowledge of whether or not the event R_2 occurs. Each of the events $R_1 = \{rgg, rgr, rrg, rrr\}$ and $R_2 = \{rrg, rrr, grg, grr\}$ has four outcomes so $P[R_1] = P[R_2] = 4/8$. In this case, the intersection $R_1 \cap R_2 = \{rrg, rrr\}$ has probability $P[R_1 R_2] = 2/8$. Since $P[R_1 R_2] = P[R_1]P[R_2]$, events R_1 and R_2 are independent. Learning whether or not the event R_2 (second light red) occurs does not affect our knowledge of whether or not the event R_1 (first light red) occurs.

Definition 1.8 3 Independent Events

A_1, A_2, *and* A_3 *are* independent *if and only if*

(a) A_1 *and* A_2 *are independent,*

(b) A_2 *and* A_3 *are independent,*

(c) A_1 *and* A_3 *are independent,*

(d) $P[A_1 \cap A_2 \cap A_3] = P[A_1]P[A_2]P[A_3]$.

Example 1.23 Problem

In an experiment with equiprobable outcomes, the event space is $S = \{1, 2, 3, 4\}$. $P[s] = 1/4$ for all $s \in S$. Are the events $A_1 = \{1, 3, 4\}$, $A_2 = \{2, 3, 4\}$, and $A_3 = \phi$ independent?

Example 1.23 Solution

These three sets satisfy the final condition of Definition 1.8 because $A_1 \cap A_2 \cap A_3 = \phi$, and

$$P[A_1 \cap A_2 \cap A_3] = P[A_1]P[A_2]P[A_3] = 0.$$

However, A_1 and A_2 are not independent because, with all outcomes equiprobable,

$$P[A_1 \cap A_2] = P[\{3, 4\}] = 1/2 \neq P[A_1]P[A_2] = 3/4 \times 3/4.$$

Hence the three events are dependent.

More than Two Independent

Definition 1.9 Events

If $n \geq 3$, the sets $A_1, A_2, \ldots, A_n$ are independent if and only if

(a) every set of $n-1$ sets taken from $A_1, A_2, \ldots A_n$ is independent,

(b) $P[A_1 \cap A_2 \cap \cdots \cap A_n] = P[A_1]P[A_2]\cdots P[A_n]$.

Quiz 1.6

Monitor two consecutive phone calls going through a telephone switching office. Classify each one as a voice call (v) if someone is speaking, or a data call (d) if the call is carrying a modem or fax signal. Your observation is a sequence of two letters (either v or d). For example, two voice calls corresponds to vv. The two calls are independent and the probability that any one of them is a voice call is 0.8. Denote the identity of call i by C_i. If call i is a voice call, then $C_i = v$; otherwise, $C_i = d$. Count the number of voice calls in the two calls you have observed. N_V is the number of voice calls. Consider the three events $N_V = 0$, $N_V = 1$, $N_V = 2$. Determine whether the following pairs of events are independent:

(1) $\{N_V = 2\}$ and $\{N_V \geq 1\}$

(2) $\{N_V \geq 1\}$ and $\{C_1 = v\}$

(3) $\{C_2 = v\}$ and $\{C_1 = d\}$

(4) $\{C_2 = v\}$ and $\{N_V$ is even$\}$

Section 1.11

MATLAB

Example 1.45

```
>> X=rand(1,4)
X =
  0.0879   0.9626   0.6627   0.2023
>> X<0.5
ans =
     1      0      0      1
```

Since `rand(1,4)<0.5` compares four random numbers against 0.5, the result is a random sequence of zeros and ones that simulates a sequence of four flips of a fair coin. We associate the outcome 1 with $\{\text{head}\}$ and 0 with $\{\text{tail}\}$.

Example 1.46 Problem

Using MATLAB, perform 75 experiments. In each experiment, flip a coin 100 times and record the number of heads in a vector $\mathbf{Y}$ such that the ith element Y_i is the number of heads in subexperiment i.

Example 1.46 Solution

```
X=rand(75,100)<0.5;
Y=sum(X,2);
```

The MATLAB code for this task appears on the left. The 75×100 matrix $\mathbf{X}$ has i, jth element $X_{ij} = 0$ (tails) or $X_{ij} = 1$ (heads) to indicate the result of flip j of subexperiment i.

Since $\mathbf{Y}$ sums $\mathbf{X}$ across the second dimension, Y_i is the number of heads in the ith subexperiment. Each Y_i is between 0 and 100 and generally in the neighborhood of 50.

Example 1.47 Problem

Simulate the testing of 100 microprocessors as described in Example 1.43. Your output should be a 4×1 vector $\mathbf{X}$ such that X_i is the number of grade i microprocessors.

Example 1.47 Solution

```
%chiptest.m
G=ceil(4*rand(1,100));
T=1:4;
X=hist(G,T);
```

The first line generates a row vector `G` of random grades for 100 microprocessors. The possible test scores are in the vector `T`. Lastly, `[` hist]X=hist(G,T) returns a histogram vector `X` such that `X(j)` counts the number of elements `G(i)` that equal `T(j)`.

Note that "`help hist`" will show the variety of ways that the hist function can be called. Morever, `[` hist]X=hist(G,T) does more than just count the number of elements of `G` that equal each element of `T`. In particular, `hist(G,T)` creates bins centered around each `T(j)` and counts the number of elements of `G` that fall into each bin.

Quiz 1.11

The flip of a thick coin yields heads with probability 0.4, tails with probability 0.5, or lands on its edge with probability 0.1. Simulate 100 thick coin flips. Your output should be a 3×1 vector $\mathbf{X}$ such that X_1, X_2, and X_3 are the number of occurrences of heads, tails, and edge.