EE8103 Random Processes

Chap 1 : Experiments, Models, and Probabilities

Introduction

- Real word exhibits randomness
  - Today's temperature;
  - Flip a coin, head or tail?
  - At a bus station, how long do you wait for the arrival of a bus?

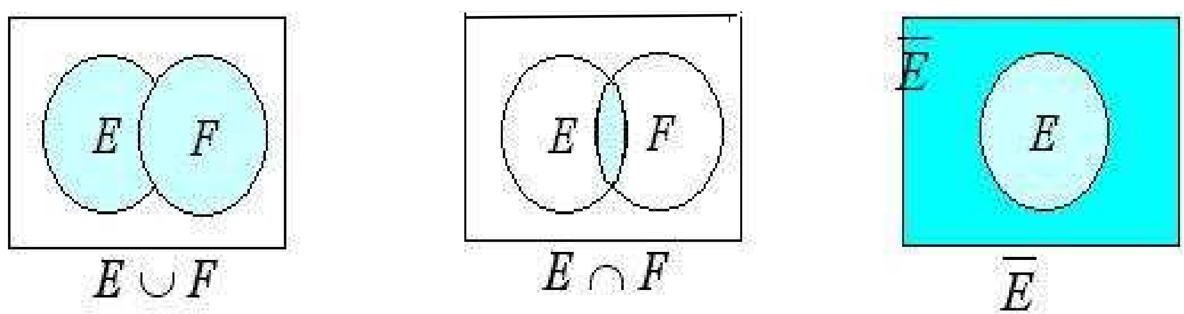
- We create models to analyze since real experiment are generally too complicated, for example, waiting time depends on the following factors:
  - The time of a day (is it rush hour?);
  - The speed of each car that passed by while you waited;
  - The weight, horsepower, and gear ratio of the bus;
  - The psychological profile and work schedule of drivers;
  - The status of all road construction within 100 miles of the bus stop.
- It would be apparent that it would be too difficult to analyze the effects of all the factors

on the likelihood that you will wait less than 5 minutes for a bus. Therefore, it is necessary to study and create a model to capture the critical part of the actual physical experiment.

- **Probability theory** deals with the study of random phenomena, which under repeated experiments yield different outcomes that have certain *underlying patterns* about them.

Review of Set Operation

- Universal set Ω: include all of the elements
- Set operations: for $E \subset \Omega$ and $F \subset \Omega$
  - Union: $E \cup F = \{s \in \Omega : s \in E \text{ or } s \in F\}$;
  - Intersection: $E \cap F = \{s \in \Omega : s \in E \text{ and } s \in F\}$;
  - Complement: $E^c = \bar{E} = \{s \in \Omega : s \notin E\}$;
  - Empty set: $\Phi = \Omega^c = \{\}$.
- Only complement needs the knowledge of Ω.



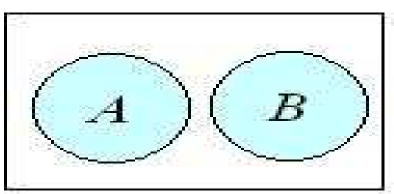
Several Definitions

- **Disjoint**: if $A \cap B = \phi$, the empty set, then A and B are said to be *mutually exclusive (M.E)*, or *disjoint*.
- **Exhaustive**: the collection of sets has

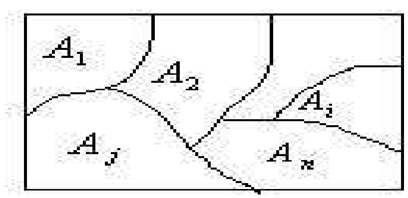
$$\sum_{j=1}^{\infty} A_j = \Omega$$

- **A partition** of Ω is a collection of mutually exclusive subsets of Ω such that their union is Ω (Partition is a stronger condition than Exhaustive.):

$$A_i \cap A_j = \phi \quad \text{and} \quad \cup_{i=1}^{n} A_i = \Omega$$

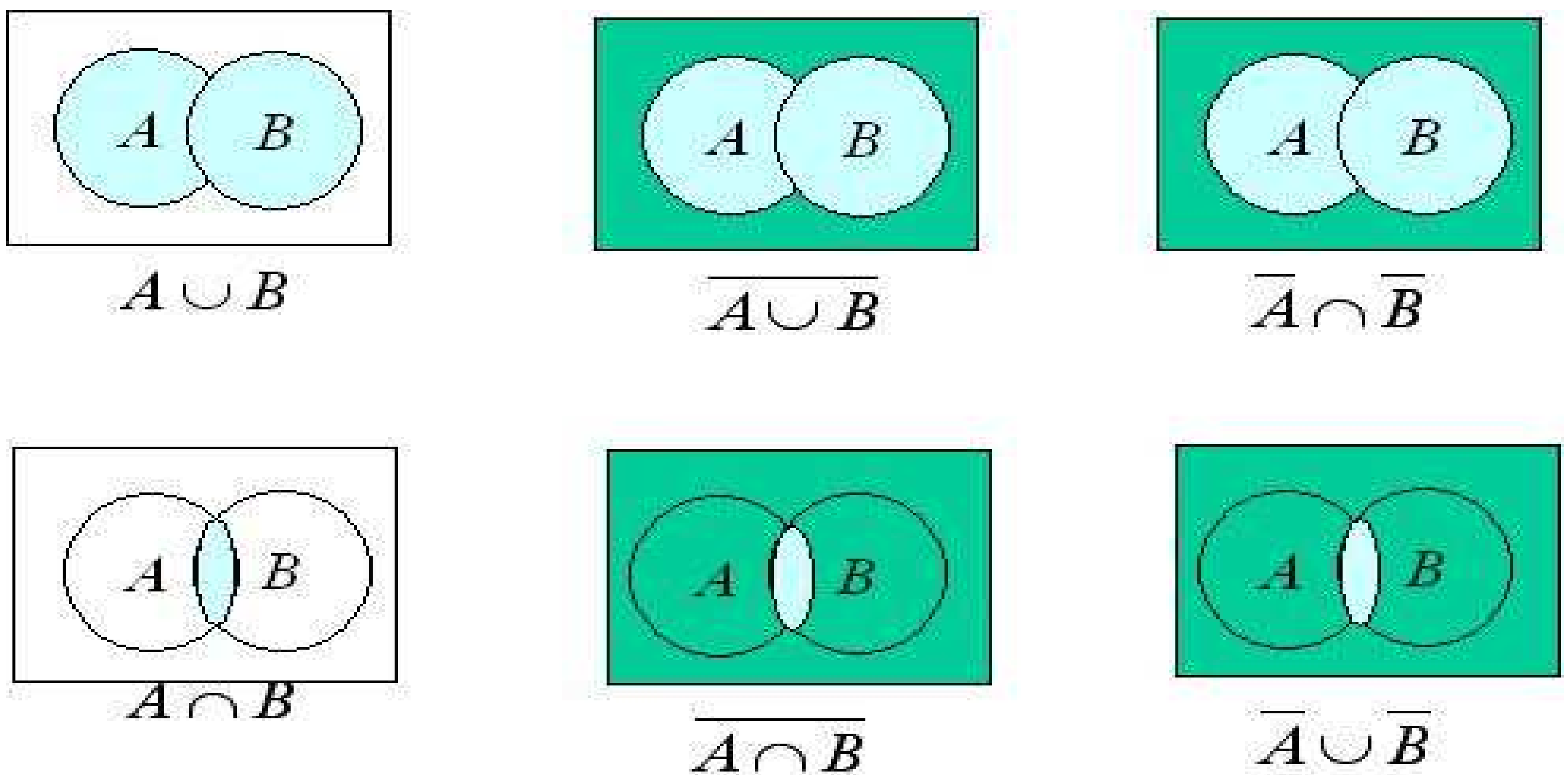


$A \cap B = \phi$



De-Morgan's Law

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



Sample Space, Events and Probabilities

- **Outcome**: an outcome of an experiment is any possible observations of that experiment.
- **Sample space**: is the *finest-grain*, mutually exclusive, collectively exhaustive set of all possible outcomes.
- **Event**: is a set of outcomes of an experiment.
- **Event Space**: is a collectively exhaustive, mutually exclusive set of events.

Sample Space and Event Space

- Sample space: contains all the details of an experiment. It is a set of all outcomes, each outcome $s \in S$. Some example:
  - coin toss: $S = \{H, T\}$
  - two coin toss: $S = \{HH, HT, TH, TT\}$
  - roll pair of dice: $S = \{(1, 1), \cdots, (6, 6)\}$
  - component life time: $S = \{t \in [0, \infty)\}$ *e.g.*lifespan of a light bulb
  - noise: $S = \{n(t); t : \text{real}\}$

- Event Space: is a set of events.

- **Example 1**: coin toss 4 times:

  The *sample space* consists of 16 four-letter words, with each letter either h (head) or t (tail).

  Let B_i=outcomes with i heads for $i = 0, 1, 2, 3, 4$. Each B_i is an *event* containing one or more outcomes, say, $B_1 = \{ttth, ttht, thtt, httt\}$ contains four outcomes. The set $B = \{B_0, B_1, B_2, B_3, B_4\}$ is an event space. It is not a sample space because it lacks the finest-grain property.

- **Example 2**: Toss two dice , there are 36 elements in the sample space. If we define the event as the sum of two dice,

$$\text{Event space:} \quad \Omega = \{B_2, B_3, \cdots, B_{12}\}$$

there are 11 events.

Probability Defined on Events

Often it is meaningful to talk about at least some of the subsets of S as events, for which we must have mechanism to compute their probabilities.

Example 3: Consider the experiment where two coins are simultaneously tossed. The sample space is $S = \{\xi_1, \xi_2, \xi_3, \xi_4\}$ where

$$\xi_1 = [H, H], \quad \xi_2 = [H, T], \quad \xi_3 = [T, H], \quad \xi_4 = [T, T]$$

If we define

$$A = \{\xi_1, \xi_2, \xi_3\}$$

The event of A is the same as "Head has occurred at least once" and qualifies as an event.

Probability measure: each event has a probability, $P(E)$

Definitions, Axioms and Theorems

- Definitions: establish the logic of probability theory
- Axioms: are facts that we have to accept without proof.
- Theorems are consequences that follow logically from definitions and axioms. Each theorem has a proof that refers to definitions, axioms, and other theorems.
- There are only three axioms.

Axioms of Probability

For any event A, we assign a number $P(A)$, called the probability of the event A. This number satisfies the following three conditions that act the axioms of probability.

1. probability is a nonnegative number

$$P(A) \geq 0 \tag{1}$$

2. probability of the whole set is unity

$$P(\Omega) = 1 \tag{2}$$

3. For any countable collection $A_1, A_2, \cdots$ of mutually exclusive events

$$P(A_1 \cup A_2 \cup \cdots) = P(A_1) + P(A_2) + \cdots \tag{3}$$

Note that (3) states that if A and B are mutually exclusive (M.E.) events, the probability of their union is the sum of their probabilities.

We will build our entire probability theory on these axioms.

Some Results Derived from the Axioms

The following conclusions follow from these axioms:

- Since $A \cup \overline{A} = \Omega$, using (2), we have

$$P(A \cup \overline{A}) = P(\Omega) = 1$$

  But $A \cap \overline{A} = \phi$, and using (3),

$$P(A \cup \overline{A}) = P(A) + P(\overline{A}) = 1 \quad \text{or} \quad P(\overline{A}) = 1 - P(A)$$

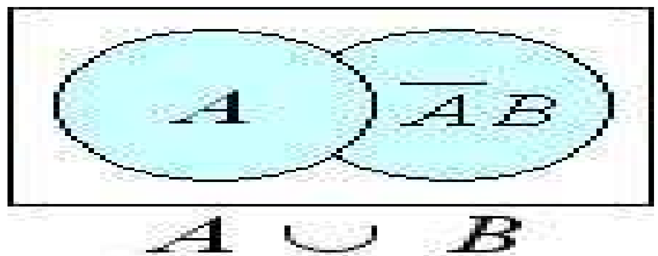
- Similarly, for any A, $A \cap \{\phi\} = \{\phi\}$. hence it follows that $P(A \cup \{\phi\}) = P(A) + P(\phi)$. But $A \cup \{\phi\} = A$ and thus

$$\boxed{P\{\phi\} = 0}$$

- Suppose A and B are not mutually exclusive (M.E.)? How does one compute $P(A \cup B)$?

  To compute the above probability, we should re-express $(A \cup B)$ in terms of M.E. sets so that we can make use of the probability axioms. From figure below,

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



where A and $\overline{A}B$ are clearly M.E. events. Thus using axiom (3)

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

To compute $P(\overline{A}B)$, we can express B as

$$B = B \cap \Omega = B \cap (A \cup \overline{A}) = (B \cap A) \cup (B \cap \overline{A}) = BA \cup B\overline{A}$$

Thus

$$P(B) = P(BA) + P(B\overline{A})$$

since $BA = AB$ and $B\overline{A} = \overline{A}B$ are M.E. events, we have

$$P(\overline{A}B) = P(B) - P(AB)$$

Therefore

$$\boxed{P(A \cup B) = P(A) + P(B) - P(AB)}$$

- Coin toss revisited:

$$\xi_1 = [H, H], \quad \xi_2 = [H, T], \quad \xi_3 = [T, H], \quad \xi_4 = [T, T]$$

Let $A = \{\xi_1, \xi_2\}$: the event that the first coin falls heads

Let $B = \{\xi_1, \xi_3\}$: the event that the second coin falls head

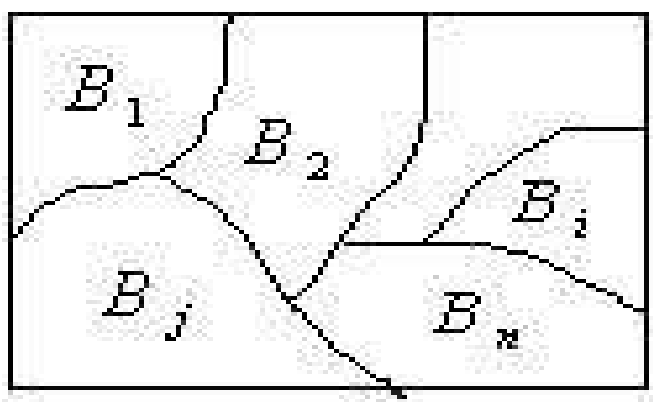
$$P(A \cup B) = P(A) + P(B) - P(AB) = \frac{1}{2} + \frac{1}{2} - \frac{1}{4} = \frac{3}{4}$$

where $A \cup B$ denotes the event that at least one head appeared.

Theorem

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

$$A = C_1 \cup C_2 \cup \cdots; \qquad P(A) = \sum P(C_i)$$



Example 4: Coin toss 4 times, let A equal the set of outcomes with less than three heads, as

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

Let $\{B_0, B_1, \cdots, B_4\}$ denote the event space in which B_i={outcomes with i heads }.

Let $C_i = A \cap B_i (i = 0, 1, 2, 3, 4)$, the above theorem states that

$$\begin{aligned} A &= C_0 \cup C_1 \cup C_2 \cup C_3 \cup C_4 \\ &= (A \cap B_0) \cup (A \cap B_1) \cup (A \cap B_2) \cup (A \cap B_3) \cup (A \cap B_4) \end{aligned}$$

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 $i = 3, 4$, $A \cap B_i = \phi$, 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 the events for "zero head", "one head", and "two heads".

Example 5: A company has a model of telephone usage. It classifies all calls as L (long), B (brief). It also observes whether calls carry voice(V), fax (F), or data(D). The sample space has six outcomes $S = \{LV, BV, LD, BD, LF, BF\}$. The probability can be represented in the table as

	V	F	D
L	0.3	0.15	0.12
B	0.2	0.15	0.08

Note that $\{V, F, D\}$ is an event space corresponding to $\{B_1, B_2, B_3\}$ in the previous theorem (and L is equivalent as the event A). Thus, we can apply the theorem to find

$$P(L) = P(LV) + P(LD) + P(LF) = 0.57$$

Conditional Probability and Independence

In N independent trials, suppose N_A, N_B, N_{AB} denote the number of times events A, B and AB occur respectively. According to the frequency interpretation of probability, for large N

$$P(A) = \frac{N_A}{N} \quad P(B) = \frac{N_B}{N} \quad P(AB) = \frac{N_{AB}}{N}$$

Among the N_A occurrences of A, only N_{AB} of them are also found among the N_B occurrences of B. Thus the ratio

$$\frac{N_{AB}}{N_B} = \frac{N_{AB}/N}{N_B/N} = \frac{P(AB)}{P(B)}$$

is a measure of the event A given that B has already occurred. We denote this conditional probability by

$$P(A|B) = \text{Probability of the event } A \text{ given that } B \text{ has occurred.}$$

We define

$$\boxed{P(A|B) = \frac{P(AB)}{P(B)}} \tag{4}$$

provided $P(B) \neq 0$. As we show below, the above definition satisfies all probability axioms discussed earlier. We have

1. Non-negative
$$P(A|B) = \frac{P(AB) \geq 0}{P(B) > 0} \geq 0$$

2.
$$P(\Omega|B) = \frac{P(\Omega B)}{P(B)} = \frac{P(B)}{P(B)} = 1 \quad \text{since } \Omega B = B$$

3. Suppose $A \cap C = \phi$, then
$$P(A \cup C|B) = \frac{P((A \cup C) \cap B)}{P(B)} = \frac{P(AB \cup CB)}{P(B)}$$

   But $AB \cap CB = \phi$, hence $P(AB \cup CB) = P(AB) + P(CB)$,
$$P(A \cup C|B) = \frac{P(AB)}{P(B)} + \frac{P(CB)}{P(B)} = P(A|B) + P(C|B)$$

satisfying all probability axioms. Thus $P(A|B)$ defines a legitimate probability measure.

Properties of Conditional Probability

1. If $B \subset A$, $AB = B$, and

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

since if $B \subset A$, then occurrence of B implies automatic occurrence of the event A. As an example, let

$$A = \{\text{outcome is even}\}, \quad B = \{\text{outcome is 2}\}$$

in a dice tossing experiment. Then $B \subset A$ and $P(A|B) = 1$.

2. If $A \subset B$, $AB = A$, and

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

In a dice experiment, $A = \{\text{outcome is 2}\}$, $B = \{\text{outcome is even}\}$, so that $A \subset B$. The statement that B has occurred (outcome is even) makes the probability for "outcome is 2" greater than that without that information.

3. We can use the conditional probability to express the probability of a complicated event in terms of simpler related events - **Law of Total Probability**.

Let $A_1, A_2, \cdots, A_n$ are pair wise disjoint and their union is Ω. Thus $A_i \cap A_j = \phi$, and

$$\cup_{i=1}^{n} A_i = \Omega$$

thus

$$B = B\Omega = B(A_1 \cup A_2 \cup \cdots \cup A_n) = BA_1 \cup BA_2 \cup \cdots \cup BA_n$$

But $A_i \cap A_j = \phi \rightarrow BA_i \cap BA_j = \phi$, so that

$$\boxed{P(B) = \sum_{i=1}^{n} P(BA_i) = \sum_{i=1}^{n} P(B|A_i)P(A_i)} \tag{5}$$

Above equation is referred as the "law of total probability". Next we introduce the notion of "independence" of events.

Independence: A and B are said to be independent events, if

$$\boxed{P(AB) = P(A)P(B)}$$

Notice that the above definition is a probabilistic statement, NOT a set theoretic notion such as mutually exclusiveness, (*independent and disjoint are not synonyms*).

More on Independence

- Disjoint events have no common outcomes and therefore $P(AB) = 0$. Independent does not mean (cannot be) disjoint, except $P(A) = 0$ or $P(B) = 0$. If $P(A) > 0$, $P(B) > 0$, and A, B independent implies $P(AB) > 0$, thus the event AB cannot be the null set.
- Disjoint leads to probability sum, while independence leads to probability multiplication.
- Suppose A and B are independent, then

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

  Thus if A and B are independent, the event that B has occurred does not shed any more light into the event A. It makes no difference to A whether B has occurred or not.

Example 6: A box contains 6 white and 4 black balls. Remove two balls at random without replacement. What is the probability that the first one is white and the second one is black?

Let W_1 = "first ball removed is white" and B_2 = "second ball removed is black". We need to find $P(W_1 \cap B_2) =?$

We have $W_1 \cap B_2 = W_1 B_2 = B_2 W_1$. Using the conditional probability rule,

$$P(W_1 B_2) = P(B_2 W_1) = P(B_2|W_1)P(W_1)$$

But

$$P(W_1) = \frac{6}{6+4} = \frac{6}{10} = \frac{3}{5}$$

and

$$P(B_2|W_1) = \frac{4}{5+4} = \frac{4}{9}$$

and hence

$$P(W_1 B_2) = \frac{3}{5}\frac{4}{9} = \frac{20}{81} = 0.27$$

Are the events W_1 and B_2 independent? Our common sense says No. To verify this we need to compute $P(B_2)$. Of course the fate of the second ball very much depends on that of the first ball. The first ball has two options: W_1 = "first ball is white" or
B_1 = "first ball is black". Note that $W_1 \cap B_1 = \phi$ and $W_1 \cup B_1 = \Omega$. Hence W_1 together with B_1 form a partition. Thus

$$\begin{aligned} P(B_2) &= P(B_2|W_1)P(W_1) + P(B_2|B_1)P(B_1) \\ &= \frac{4}{5+4} \cdot \frac{3}{5} + \frac{3}{6+3} \cdot \frac{4}{10} = \frac{4}{9} \cdot \frac{3}{5} + \frac{1}{3} \cdot \frac{5}{5} = \frac{4+2}{15} = \frac{2}{5} \end{aligned}$$

and

$$P(B_2)P(W_1) = \frac{2}{5} \cdot \frac{3}{5} \neq P(B_2W_1) = \frac{20}{81}$$

As expected, the events W_1 and B_2 are dependent.

Bayes' Theorem

Since

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

similarly,

$$P(B|A) = \frac{P(BA)}{P(A)} = \frac{P(AB)}{P(A)} \rightarrow P(AB) = P(B|A)P(A)$$

We get

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

or

$$\boxed{P(A|B) = \tfrac{P(B|A)}{P(B)} \cdot P(A)} \tag{6}$$

The above equation is known as **Bayes'theorem**.

Although simple enough, Bayes theorem has an interesting interpretation: $P(A)$ represents the a-priori probability of the event A. Suppose B has occurred, and assume that A and B are not independent. How can this new information be used to update our knowledge about A? Bayes rule takes into account the new information ("B has occurred") and gives out the a-posteriori probability of A given B.

We can also view the event B as new knowledge obtained from a fresh experiment. We know something about A as $P(A)$. The new information is available in terms of B. The new information should be used to improve our knowledge/understanding of A. Bayes theorem gives the exact mechanism for incorporating such new information.

Bayes' Theorem

A more general version of Bayes theorem involves partition of Ω as

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{P(B)} = \frac{P(B|A_i)P(A_i)}{\sum_{j=1}^{n} P(B|A_j)P(A_j)} \tag{7}$$

In above equation, $A_i, i = [1, n]$ represent a set of mutually exclusive events with associated a-priori probabilities $P(A_i), i = [1, n]$. With the new information "B has occurred", the information about A_i can be updated by the n conditional probabilities $P(B|A_j), j = [1, n]$.

Example 7: Two boxes $B1$ and $B2$ contain 100 and 200 light bulbs respectively. The first box $(B1)$ has 15 defective bulbs and the second 5. Suppose a box is selected at random and one bulb is picked out.

(a) What is the probability that it is defective?

Solution: Note that box $B1$ has 85 good and 15 defective bulbs. Similarly box $B2$ has 195 good and 5 defective bulbs. Let $D =$ "Defective bulb is picked out". Then,

$$P(D|B1) = \frac{15}{100} = 0.15, \quad P(D|B2) = \frac{5}{200} = 0.025$$

Since a box is selected at random, they are equally likely.

$$P(B1) = P(B2) = 1/2$$

Thus $B1$ and $B2$ form a partition, and using Law of Total Probability, we obtain

$$P(D) = P(D|B1)P(B1) + P(D|B2)P(B2) = 0.15 \cdot \frac{1}{2} + 0.025 \cdot \frac{1}{2} = 0.0875$$

Thus, there is about 9% probability that a bulb picked at random is defective.

(b) Suppose we test the bulb and it is found to be defective. What is the probability that it came from box 1? $P(B1|D) =?$

$$P(B1|D) = \frac{P(D|B1)P(B1)}{P(D)} = \frac{0.15 \cdot 0.5}{0.0875} = 0.8571 \quad (8)$$

Notice that initially $P(B1) = 0.5$; then we picked out a box at random and tested a bulb that turned out to be defective. Can this information shed some light about the fact that we might have picked up box 1?

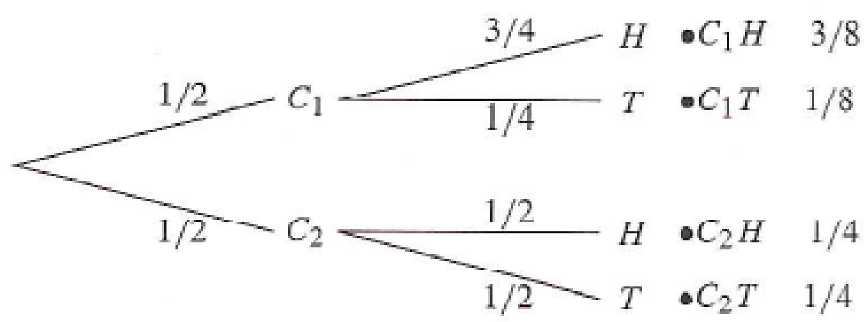
From (8), $P(B1|D) = 0.875 > 0.5$, and indeed it is more likely at this point that we must have chosen box 1 in favor of box 2. (Recall that the defective rate in Box 1 is 6 times of that in Box 2).

Example: (textbook Example 1.27)

Suppose you have two coins, one biased, one fair, but you don't know which coin is which. Coin 1 is biased. It comes up heads with probability $3/4$, while coin 2 will flip heads with probability $1/2$. Suppose you pick a coin at random and flip it. Let C_i denote the event that coin i is picked. Let H and T denote the possible outcomes of the flip. Given that the outcome of the flip is a head, what is $P[C_1|H]$, the probability that you picked the biased coin? Given that the outcome is a tail, what is the probability $P[C_1|T]$ that you picked the biased coin?

Solution:

First, we construct the sample tree.



To find the conditional probabilities, we see

$$P[C_1|H] = \frac{P[C_1H]}{P[H]} = \frac{P[C_1H]}{P[C_1H]+P[C_2H]} = \frac{3/8}{3/8+1/4} = \frac{3}{5}. \qquad (1.52)$$

Similarly,

$$P[C_1|T] = \frac{P[C_1T]}{P[T]} = \frac{P[C_1T]}{P[C_1T]+P[C_2T]} = \frac{1/8}{1/8+1/4} = \frac{1}{3}. \qquad (1.53)$$

As we would expect, we are more likely to have chosen coin 1 when the first flip is heads, but we are more likely to have chosen coin 2 when the first flip is tails.