PROBABILITY THEORY

1. Basics

Probability theory deals with the study of random phenomena, which under repeated experiments yield different outcomes that have certain underlying patterns about them. The notion of an experiment assumes a set of repeatable conditions that allow any number of identical repetitions. When an experiment is performed under these conditions, certain elementary events ξ_i occur in different but *completely uncertain* ways. We can assign nonnegative number $P(\xi_i)$, as the probability of the event ξ_i in various ways:

Laplace's Classical Definition: The Probability of an event A is defined a-priori without actual experimentation as

$$P(A) = \frac{\text{Number of outcomes favorable to } A}{\text{Total number of possible outcomes}}\,, \qquad (1\text{-}1)$$

provided all these outcomes are *equally likely*.

Consider a box with n white and m red balls. In this case, there are two elementary outcomes: white ball or red ball. Probability of "selecting a white ball" $= \dfrac{n}{n+m}$.

We can use above classical definition to determine the probability that a given number is divisible by a prime p.

If p is a prime number, then every p^{th} number (starting with p) is divisible by p. Thus among p consecutive integers there is one favorable outcome, and hence

$$P\{a \text{ given number is divisible by a prime } p\} = \frac{1}{p} \qquad (1\text{-}2)$$

Relative Frequency Definition: The probability of an event A is defined as

$$P(A) = \lim_{n \to \infty} \frac{n_A}{n} \qquad (1\text{-}3)$$

where n_A is the number of occurrences of A and n is the total number of trials.

We can use the relative frequency definition to derive (1-2) as well. To do this we argue that among the integers $1, 2, 3, \cdots, n,$ the numbers $p, 2p, \cdots$ are divisible by p.

Thus there are n/p such numbers between 1 and n. Hence

$$P\{a \text{ given number } N \text{ is divisible by a prime } p\}$$

$$= \lim_{n\to\infty} \frac{n/p}{n} = \frac{1}{p}. \qquad \text{(1-4)}$$

In a similar manner, it follows that

$$P\{\, p^2 \text{ divides any given number } N\} = \frac{1}{p^2} \qquad \text{(1-5)}$$

and

$$P\{\, pq \text{ divides any given number } N\} = \frac{1}{pq}. \qquad \text{(1-6)}$$

The axiomatic approach to probability, due to Kolmogorov, developed through a set of axioms (below) is generally recognized as superior to the above definitions, (1-1) and (1-3), as it provides a solid foundation for complicated applications.

The totality of all ξ_i, *known a priori*, constitutes a set Ω, the set of all experimental outcomes.

$$\Omega = \{ \xi_1, \xi_2, \cdots, \xi_k, \cdots \} \qquad (1\text{-}7)$$

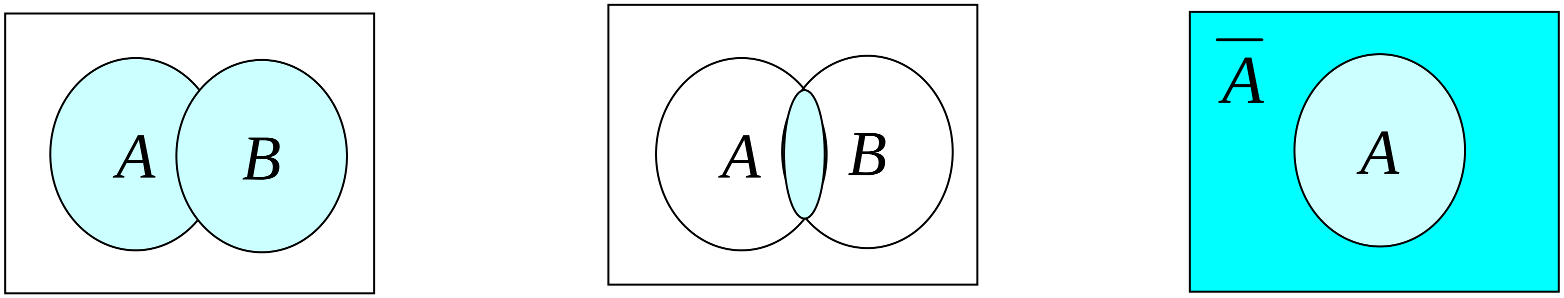
Ω has subsets $A, B, C, \cdots$. Recall that if A is a subset of Ω, then $\xi \in A$ implies $\xi \in \Omega$. From A and B, we can generate other related subsets $A \cup B$, $A \cap B$, $\overline{A}$, $\overline{B}$, etc.

$$A \cup B = \{ \xi \mid \xi \in A \text{ or } \xi \in B \}$$

$$A \cap B = \{ \xi \mid \xi \in A \text{ and } \xi \in B \}$$

and

$$\overline{A} = \{ \xi \mid \xi \notin A \} \qquad (1\text{-}8)$$

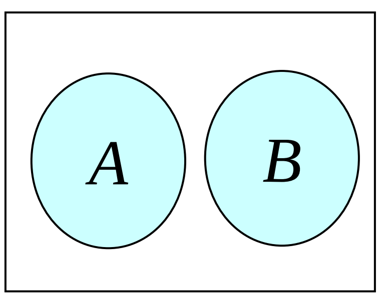


$A \cup B$ $\qquad$ $A \cap B$ $\qquad$ $\overline{A}$

Fig.1.1

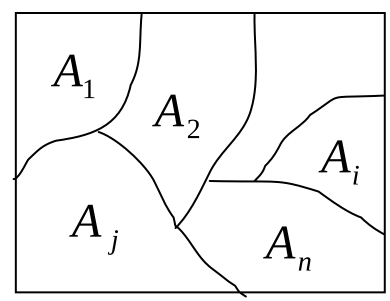
- If $A \cap B = \phi$, the empty set, then A and B are said to be mutually exclusive (M.E).
- A partition of Ω is a collection of mutually exclusive subsets of Ω such that their union is Ω.

$$A_i \cap A_j = \phi, \quad \text{and} \quad \bigcup_{i=1} A_i = \Omega. \tag{1-9}$$



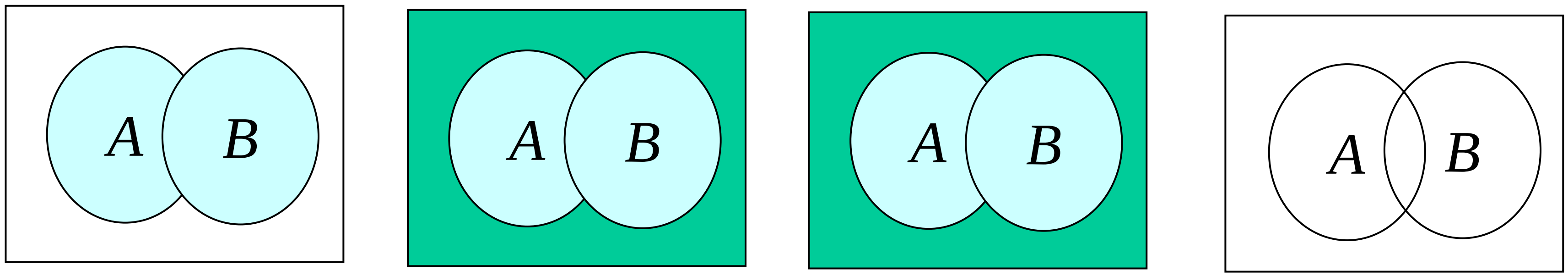
$A \cap B = \phi$

Fig. 1.2



De-Morgan's Laws:

$$\overline{A\cup B}=\overline{A}\cap\overline{B}; \qquad \overline{A\cap B}=\overline{A}\cup\overline{B} \qquad (1\text{-}10)$$



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

Fig.1.3

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

Example 1.1: Consider the experiment where two coins are simultaneously tossed. The various elementary events are

$$\xi_1 = (H,H),\ \ \xi_2 = (H,T),\ \ \xi_3 = (T,H),\ \ \xi_4 = (T,T)$$

and

$$\Omega = \{\xi_1, \xi_2, \xi_3, \xi_4\}.$$

The subset $A = \{\xi_1, \xi_2, \xi_3\}$ is the same as "Head has occurred at least once" and qualifies as an event.

Suppose two subsets A and B are both events, then consider

"Does an outcome belong to A or $B = A \cup B$"

"Does an outcome belong to A and $B = A \cap B$"

"Does an outcome fall outside A"?

Thus the sets $A \cup B$, $A \cap B$, $\overline{A}$, $\overline{B}$, etc., also qualify as events. We shall formalize this using the notion of a Field.

- **Field**: A collection of subsets of a nonempty set Ω forms a field F if

$$\begin{aligned} &\text{(i)} \quad \Omega \in F \\ &\text{(ii)} \quad \text{If } A \in F, \quad \text{then } \overline{A} \in F \\ &\text{(iii)} \quad \text{If } A \in F \quad \text{and} \quad B \in F, \quad \text{then } A \cup B \in F. \end{aligned} \tag{1-11}$$

Using (i) - (iii), it is easy to show that $A \cap B$, $\overline{A} \cap B$, etc., also belong to F. For example, from (ii) we have $\overline{A} \in F$, $\overline{B} \in F$, and using (iii) this gives $\overline{A} \cup \overline{B} \in F$; applying (ii) again we get $\overline{\overline{A} \cup \overline{B}} = A \cap B \in F$, where we have used De Morgan's theorem in (1-10).

Thus if $A \in F$, $B \in F$, then

$$F = \left\{ \Omega, A, B, \overline{A}, \overline{B}, A \cup B, A \cap B, \overline{A} \cup B, \cdots \right\}. \quad (1\text{-}12)$$

From here on wards, we shall reserve the term 'event' only for members of *F*.

Assuming that the probability $p_i = P(\xi_i)$ of elementary outcomes ξ_i of Ω are apriori defined, how does one assign probabilities to more 'complicated' events such as *A*, *B*, *AB*, etc., above?

The three axioms of probability defined below can be used to achieve that goal.

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 as axioms of probability.

(i) $P(A) \geq 0$ (Probability is a nonnegative number)

(ii) $P(\Omega) = 1$ (Probability of the whole set is unity) (1-13)

(iii) If $A \cap B = \phi$, then $P(A \cup B) = P(A) + P(B)$.

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

The following conclusions follow from these axioms:

a. Since $A \cup \overline{A} = \Omega$, we have using (ii)

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

But $A \cap \overline{A} \in \phi$, and using (iii),

$$\mathrm{P}(A \cup \overline{A}) = P(A) + \mathrm{P}(\overline{A}) = 1 \quad \text{or} \quad \mathrm{P}(\overline{A}) = 1 - P(A). \qquad \text{(1-14)}$$

b. Similarly, for any *A*, $A \cap \{\phi\} = \{\phi\}.$

Hence it follows that $P\big(A \cup \{\phi\}\big) = P(A) + P(\phi)$.

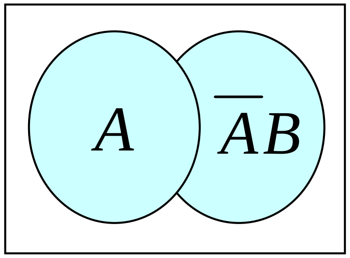
But $A \cup \{\phi\} = A$, and thus $P\{\phi\} = 0$. (1-15)

c. 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 Fig.1.4 we have

$$A \cup B = A \cup \overline{A}B, \qquad (1\text{-}16)$$



$A \cup B$

Fig.1.4

where A and $\overline{A}B$ are clearly M.E. events. Thus using axiom (1-13-iii)

$$P(A \cup B) = P(A \cup \overline{A}B) = P(A) + P(\overline{A}B). \qquad (1\text{-}17)$$

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

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

$$= (B \cap A) \cup (B \cap \overline{A}) = BA \cup B\overline{A} \qquad (1\text{-}18)$$

Thus

$$P(B) = P(BA) + P(B\overline{A}), \qquad (1\text{-}19)$$

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

From (1-19),

$$P(\overline{A}B) = P(B) - P(AB) \qquad \text{(1-20)}$$

and using (1-20) in (1-17)

$$P(A \cup B) = P(A) + P(B) - P(AB). \qquad \text{(1-21)}$$

- Question: Suppose every member of a denumerably infinite collection A_i of pair wise disjoint sets is an event, then what can we say about their union

$$A = \bigcup_{i=1}^{\infty} A_i \ ? \qquad \text{(1-22)}$$

i.e., suppose all $A_i \in F$, what about A? Does it belong to F? (1-23)

Further, if A also belongs to F, what about $P(A)$? (1-24)

The above questions involving infinite sets can only be settled using our intuitive experience from plausible experiments. For example, in a coin tossing experiment, where the same coin is tossed indefinitely, define

$$A = \text{“head eventually appears”}. \qquad (1\text{-}25)$$

Is A an event? Our intuitive experience surely tells us that A is an event. Let

$$\begin{aligned} A_n &= \{\text{head appears for the 1st time on the } n\text{th toss}\} \\ &= \{\underbrace{t, t, t, \cdots, t}_{n-1}, h\} \end{aligned} \qquad (1\text{-}26)$$

Clearly $A_i \cap A_j = \phi.$ Moreover the above A is

$$A = A_1 \cup A_2 \cup A_3 \cup \cdots \cup A_i \cup \cdots. \qquad (1\text{-}27)$$

We cannot use probability axiom (1-13-iii) to compute $P(A)$, since the axiom only deals with two (or a finite number) of M.E. events.

To settle both questions above (1-23)-(1-24), extension of these notions must be done, based on our intuition, as new axioms.

- **σ-Field (Definition)**:

A field F is a σ-field if in addition to the three conditions in (1-11), we have the following:

For every sequence $A_i, i = 1 \to \infty$, of pair wise disjoint events belonging to F, their union also belongs to F, i.e.,

$$A = \bigcup_{i=1}^{\infty} A_i \in F. \qquad (1\text{-}28)$$

In view of (1-28), we can add yet another axiom to the set of probability axioms in (1-13).

(iv) If A_i are pair wise mutually exclusive, then

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n). \tag{1-29}$$

Returning back to the coin tossing experiment, from experience we know that if we keep tossing a coin, eventually, a head must show up, i.e.,

$$P(A) = 1. \tag{1-30}$$

But $A = \bigcup_{n=1}^{\infty} A_n$, and using the fourth probability axiom in (1-29),

$$P(A) = P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n). \tag{1-31}$$

From (1-26), for a fair coin since only one in 2^n outcomes is in favor of A_n , we have

$$P(A_n) = \frac{1}{2^n} \quad \text{and} \quad \sum_{n=1}^{\infty} P(A_n) = \sum_{n=1}^{\infty} \frac{1}{2^n} = 1, \qquad \text{(1-32)}$$

which agrees with (1-30), thus justifying the

'reasonableness' of the fourth axiom in (1-29).

In summary, the triplet (Ω, F, P) composed of a nonempty set Ω of elementary events, a σ-field F of subsets of Ω, and a probability measure P on the sets in F subject the four axioms ((1-13) and (1-29)) form a probability model.

The probability of more complicated events must follow from this framework by deduction.

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

$$P(A) \approx \frac{N_A}{N}, \quad P(B) \approx \frac{N_B}{N}, \quad P(AB) \approx \frac{N_{AB}}{N}. \tag{1-33}$$

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

$$\frac{N_{AB}}{N_B} = \frac{N_{AB}/N}{N_B/N} = \frac{P(AB)}{P(B)} \tag{1-34}$$

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

$P(A|B) =$ Probability of “the event A given that B has occurred”.

We define

$$P(A \mid B) = \frac{P(AB)}{P(B)}, \tag{1-35}$$

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

We have

(i) $$P(A\,|\,B) = \frac{P(AB) \geq 0}{P(B) > 0} \geq 0, \tag{1-36}$$

(ii) $$P(\Omega\,|\,B) = \frac{P(\Omega B)}{P(B)} = \frac{P(B)}{P(B)} = 1, \text{ since } \Omega B = B. \tag{1-37}$$

(iii) Suppose $A \cap C = 0$. Then

$$P(A \cup C\,|\,B) = \frac{P((A \cup C) \cap B)}{P(B)} = \frac{P(AB \cup CB)}{P(B)}. \tag{1-38}$$

But $AB \cap AC = \phi$, hence $P(AB \cup CB) = P(AB) + P(CB).$

$$P(A \cup C\,|\,B) = \frac{P(AB)}{P(B)} + \frac{P(CB)}{P(B)} = P(A\,|\,B) + P(C\,|\,B), \tag{1-39}$$

satisfying all probability axioms in (1-13). Thus (1-35) defines a legitimate probability measure.

Properties of Conditional Probability:

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

$$P(A \mid B) = \frac{P(AB)}{P(B)} = \frac{P(B)}{P(B)} = 1 \qquad (1\text{-}40)$$

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

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

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

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

$$P(A \mid B) = \frac{P(AB)}{P(B)} = \frac{P(A)}{P(B)} > P(A). \qquad (1\text{-}41)$$

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

c. We can use the conditional probability to express the probability of a complicated event in terms of "simpler" related events.

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

$$\bigcup_{i=1}^{n} A_i = \Omega . \tag{1-42}$$

Thus

$$B = B(A_1 \cup A_2 \cup \cdots \cup A_n) = BA_1 \cup BA_2 \cup \cdots \cup BA_n . \tag{1-43}$$

But $A_i \cap A_j = \phi \Rightarrow BA_i \cap BA_j = \phi$, so that from (1-43)

$$P(B) = \sum_{i=1}^{n} P(BA_i) = \sum_{i=1}^{n} P(B \mid A_i) P(A_i). \qquad (1\text{-}44)$$

With the notion of conditional probability, next we introduce the notion of “independence” of events.

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

$$P(AB) = P(A) \cdot P(B). \qquad (1\text{-}45)$$

Notice that the above definition is a probabilistic statement, *not* a set theoretic notion such as mutually exclusiveness.

Suppose A and B are independent, then

$$P(A \mid B) = \frac{P(AB)}{P(B)} = \frac{P(A)P(B)}{P(B)} = P(A). \qquad (1\text{-}46)$$

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. An example will clarify the situation:

Example 1.2: 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"

$B_2 =$ "second ball removed is black"

We need $P(W_1 \cap B_2) = ?$ We have $W_1 \cap B_2 = W_1 B_2 = B_2 W_1$.
Using the conditional probability rule,

$$P(W_1 B_2) = P(B_2 W_1) = P(B_2 \mid W_1) P(W_1). \qquad \text{(1-47)}$$

But

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

and

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

and hence

$$P(W_1 B_2) = \frac{3}{5} \cdot \frac{4}{9} = \frac{12}{45} \approx 0.27.$$

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 (see (1-42)-(1-44))

$$P(B_2) = P(B_2 \mid W_1)P(W_1) + P(B_2 \mid B_1)P(B_1)$$

$$= \frac{4}{5+4} \cdot \frac{3}{5} + \frac{3}{6+3} \cdot \frac{4}{10} = \frac{4}{9} \cdot \frac{3}{5} + \frac{1}{3} \cdot \frac{2}{5} = \frac{4+2}{15} = \frac{2}{5},$$

and

$$P(B_2)P(W_1) = \frac{2}{5} \cdot \frac{3}{5} \neq P(B_2 W_1) = \frac{12}{45}.$$

As expected, the events W_1 and B_2 are *dependent*.

From (1-35),

$$P(AB) = P(A \mid B) P(B). \tag{1-48}$$

Similarly, from (1-35)

$$P(B \mid A) = \frac{P(BA)}{P(A)} = \frac{P(AB)}{P(A)},$$

or

$$P(AB) = P(B \mid A) P(A). \tag{1-49}$$

From (1-48)-(1-49), we get

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

or

$$P(A \mid B) = \frac{P(B \mid A)}{P(B)} \cdot P(A) \tag{1-50}$$

Equation (1-50) 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 in (1-50) take 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.

A more general version of Bayes' theorem involves partition of Ω. From (1-50)

$$P(A_i \mid B) = \frac{P(B \mid A_i)P(A_i)}{P(B)} = \frac{P(B \mid A_i)P(A_i)}{\sum_{i=1}^{n} P(B \mid A_i)P(A_i)}, \qquad (1\text{-}51)$$

where we have made use of (1-44). In (1-51), $A_i,\ i = 1 \to n$, represent a set of mutually exclusive events with associated a-priori probabilities $P(A_i),\ i = 1 \to n.$ With the new information "*B* has occurred", the information about A_i can be updated by the *n* conditional probabilities $P(B \mid A_i),\ i = 1 \to n,$ using (1-47).

Example 1.3: Two boxes B_1 and B_2 contain 100 and 200 light bulbs respectively. The first box (B_1) 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 B_1 has 85 good and 15 defective bulbs. Similarly box B_2 has 195 good and 5 defective bulbs. Let D = “Defective bulb is picked out”.

Then

$$P(D \mid B_1) = \frac{15}{100} = 0.15, \quad P(D \mid B_2) = \frac{5}{200} = 0.025 .$$

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

$$P(B_1) = P(B_2) = \frac{1}{2}.$$

Thus B_1 and B_2 form a partition as in (1-43), and using (1-44) we obtain

$$P(D) = P(D \mid B_1)P(B_1) + P(D \mid B_2)P(B_2)$$
$$= 0.15 \times \frac{1}{2} + 0.025 \times \frac{1}{2} = 0.0875 \ .$$

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? $P(B_1 \mid D) = ?$

$$P(B_1 \mid D) = \frac{P(D \mid B_1)P(B_1)}{P(D)} = \frac{0.15 \times 1/2}{0.0875} = 0.8571 \; . \qquad (1\text{-}52)$$

Notice that initially $P(B_1) = 0.5;$ 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 (1-52), $P(B_1 \mid D) = 0.857 > 0.5,$ and indeed it is more likely at this point that we must have chosen box 1 in favor of box 2. (Recall box1 has six times more defective bulbs compared to box2).