PROBABILITY THEORY

FINITE SET S OF ELEMENTARY EVENTS

$\{ (1,b), (2,b), (3,w), (4,w) \}$

PROBABLITY DISTRIBUTION

$P : S \to [0,1]$

- $P(s_i) \geq 0$
- $\sum_i P(s_i) = 1$

- EVENT A IS ANY SUBSET OF S

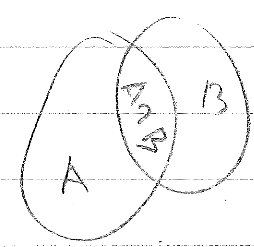
- $P(A) = \sum_{s_i \in A} P(s_i)$

SUM OVER ELEMENTARY EVENTS IN A

- AXIOMS:
  - $P(S) = 1$
  - $P(A \dot{\cup} B) = P(A) + P(B)$

↑ DISJOINT UNION

- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$



$$P(A|B) = \frac{P(A \cap B)}{P(B)} \quad \text{for } P(B) > 0$$

CONDITIONAL PROB. OF
EVENT A GIVEN EVENT B

URN

(1, b)
(2, b)
(3, w)
(4, w)

SAMPLE SPACE S

A ball is selected from an urn containing two black balls, numbered 1 and 2, and two white balls, numbered 3 and 4. The number and color of the ball is noted, so the sample space is $\{(1, b), (2, b), (3, w), (4, w)\}$. Assuming that the four outcomes are equally likely, find $P[A \mid B]$ and $P[A \mid C]$, where A, B, and C are the following events:

$A = \{(1, b), (2, b)\}$, "black ball selected,"

$B = \{(2, b), (4, w)\}$, "even-numbered ball selected," and

$C = \{(3, w), (4, w)\}$, "number of ball is greater than 2."

$$P(A \cap B) = P((2, b)) = .25$$

$$P(A \cap C) = P(\emptyset) = 0$$

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{.25}{.5} = .5 = P(A)$$

$$P(A|C) = \frac{P(A \cap C)}{P(C)} = \frac{0}{.5} = 0 \neq P(A)$$

In the first case, knowledge of B did not alter the probability of A. In the second case, knowledge of C implied that A had not occurred. ■■

If we multiply both sides of the definition of $P[A \mid B]$ by $P[B]$ we obtain

$$P[A \cap B] = P[A \mid B]P[B]. \tag{2.25a}$$

Similarly we also have that

$$P[A \cap B] = P[B \mid A]P[A]. \tag{2.25b}$$

INDEPENDENCE OF EVENTS

If knowledge of the occurrence of an event B does not alter the probability of some other event A, then it would be natural to say that event A is independent of B. In terms of probabilities this situation occurs when

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

The above equation has the problem that the right-hand side is not defined when $P[B] = 0$.

We will define two events A and B to be **independent** if

$$P[A \cap B] = P[A]P[B]. \tag{2.28}$$

Equation (2.28) then implies both

$$P[A \mid B] = P[A] \tag{2.29a}$$

and

$$P[B \mid A] = P[B] \tag{2.29b}$$

Note also that Eq. (2.29a) implies Eq. (2.28) when $P[B] \neq 0$ and Eq. (2.29b) implies Eq. (2.28) when $P[A] \neq 0$.

7.

S (1, b) (2, b) (3, w) (4, w)

$$A = \{(1, b), (2, b)\}, \quad \text{"black ball selected";}$$
$$B = \{(2, b), (4, w)\}, \quad \text{"even-numbered ball selected"; and}$$
$$C = \{(3, w), (4, w)\}, \quad \text{"number of ball is greater than 2."}$$

Are events A and B independent? Are events A and C independent?

First, consider events A and B. The probabilities required by Eq. (2.28)

$$P[A] = P[B] = \frac{1}{2}, \qquad \text{and}$$

$$P[A \cap B] = P[\{(2, b)\}] = \frac{1}{4}.$$

Thus

$$P[A \cap B] = \frac{1}{4} = P[A]P[B],$$

and the events A and B are independent. Equation (2.29b) gives more insight into the meaning of independence:

$$P[A \mid B] = \frac{P[A \cap B]}{P[B]} = \frac{P[\{(2, b)\}]}{P[\{(2, b), (4, w)\}]} = \frac{1/4}{1/2} = \frac{1}{2}$$

$$P[A] = \frac{P[A]}{P[S]} = \frac{P[\{(1, b), (2, b)\}]}{P[\{(1, b), (2, b), (3, w), (4, w)\}]} = \frac{1/2}{1}.$$

These two equations imply that $P[A] = P[A \mid B]$ because the proportion of outcomes in S that lead to the occurrence of A is equal to the proportion of outcomes in B that lead to A. Thus knowledge of the occurrence of B does not alter the probability of the occurrence of A.

Events A and C are not independent since $P[A \cap C] = P[\varnothing] = 0$ so

$$P[A \mid C] = 0 \neq P[A] = .5.$$

In fact, A and C are mutually exclusive since $A \cap C = \varnothing$, so the occurrence of C implies that A has definitely not occurred. ■■

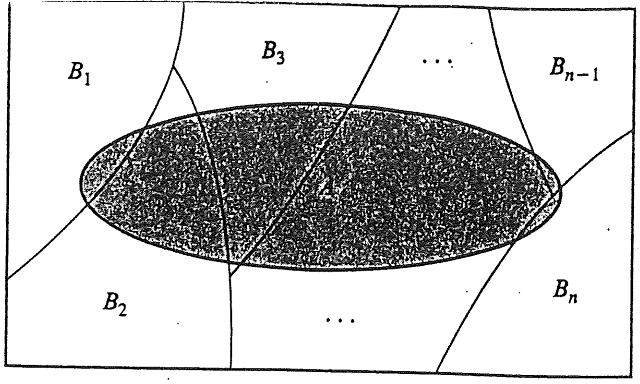


FIGURE 2.14 A partition of S into n disjoint sets.

Let $B_1, B_2, \ldots, B_n$ be mutually exclusive events whose union equals the sample space S as shown in Fig. 2.14. We refer to these sets as a **partition** of S. Any event A can be represented as the union of mutually exclusive events in the following way:

$$A = A \cap S = A \cap (B_1 \cup B_2 \cup \cdots \cup B_n)$$
$$= (A \cap B_1) \cup (A \cap B_2) \cup \cdots \cup (A \cap B_n).$$

See Fig. 2.14. By Corollary 4, the probability of A is

$$P[A] = P[A \cap B_1] + P[A \cap B_2] + \cdots + P[A \cap B_n].$$

By applying Eq. (2.25a) to each of the terms on the right-hand side, we obtain the **theorem on total probability**:

$$P[A] = P[A \mid B_1]P[B_1] + P[A \mid B_2]P[B_2] + \cdots + P[A \mid B_n]P[B_n].$$

KNOWLEDGE OF $P(A/B_i)$
AND $P(B_i)$
LETS US COMPUTE $P(A)$

10

Bayes' Rule

Let $B_1, B_2, \ldots, B_n$ be a partition of a sample space S. Suppose that event A occurs, what is the probability of event B_j? By the definition of conditional probability we have

$$P[B_j \mid A] = \frac{P[A \cap B_j]}{P[A]} = \frac{P[A \mid B_j]P[B_j]}{\sum_{k=1}^{n} P[A \mid B_k]P[B_k]}, \tag{2.27}$$

where we used the theorem on total probability to replace $P[A]$. Equation (2.27) is called **Bayes' rule.**

$P(B_j)$ PRIOR PROBILITIES

EXPERIMENT PERFORMED AND A OCCURRED

$P(B_j \mid A)$ POSTERIOR PROBABILITIES GIVEN ADDITIONAL INFORMATION

BAYES

- N EXPERTS / MODELS E_i

- IN EACH TRIAL t WE OBSERVE LABEL / DATUM y_t

ASSUMPTION :

- ONE EXPERT E_i GENERATED $(y_1, y_2, \ldots, y_T) = \bar{y}$
- PRIOR PROBABILITY OF EXPERT E_i IS $P(E_i)$

$y_t \in Y$ FINITE

PROBABILITY OF DATA $\bar{y}$ GIVEN E_i GENERATED IT :

$P(\bar{y} \mid E_i)$ DATA LIKELIHOODS

IMPORTANT SPECIAL CASE :

$y_1, y_2, \ldots, y_T$ ARE GENERATED INDEPENDENTLY AT RANDOM

THUS $P(y_1, \ldots, y_T \mid E_i) = \prod_{t=1}^{T} P(y_t \mid E_i)$

GENERAL CASE

$$P(y_1, \ldots, y_T \mid E_i) = \prod_{t=1}^{T} P(y_t \mid E_i, y_1, \ldots, y_{t-1})$$

FOR EXAMPLE: EXPERTS ARE COINS $Y = \{0, 1\}$

	E_1	E_2	E_3	E_4
$P(1 \mid E_i)$	.1	.2	.8	.9
$P(E_i)$	.2	.4	.3	.1

$\bar{y}_3 = (1, 1, 0)$

$$P(E_i \mid \bar{y}_3) = \frac{P(\bar{y}_3 \mid E_i)\, P(E_i)}{P(\bar{y}_3)}$$

POSTERIOR

$$= \frac{P(1 \mid E_i)^2 \, (1 - P(1 \mid E_i))\, P(E_i)}{P(\bar{y}_3)}$$

	E_1	E_2	E_3	E_4
$P(E_i \mid \bar{y}_3) \sim$	$.1^2 .9 .2$	$.2^2 .8 .4$	$.8^2 .2 .3$	$.9^2 .1 .1$
$\sim$	18	128	384	81

FOR 1-HEAVY SEQUENCES

POSTERIOR WILL BECOME $\approx \underset{i}{\arg\max}\ P(1 \mid E_i)$

PROVIDED THAT ALL $P(E_i) > 0$