Chapter 1: Probability

August 30, 2009

1 Introduction

The mathematical theory of probability has been applied to a wide variety of phenomena; the following are some examples.

- Probability theory has been used in genetics as a model for mutations and ensuing natural variability, and plays a central role in bioinformatics.
- Probability theory is used to study complex systems and improve their reliability, such as in modern commercial or military aircraft.
- Probability theory is a cornerstone of the theory of finance.
- $\cdots\cdots$

2 Sample Spaces

2.1 Sample space

The term "experiment" is used in probability theory to describe virtually every process whose outcome is not known in advance with certainty. Here are some examples.

(1) In an experiment in which a coin is to be tossed 10 times, the experimenter might want to determine the probability that at least four heads will be obtained.

(2) In an experiment in which the air temperature at a certain location is to be observed every day at noon for 90 successive days, a person might want to determine the probability that the average temperature during this period will be less than some specified value.

(3) In evaluating an industrial research and development project at a certain time, a person might want to determine the probability that the project will result in the successful development of a new product within a specified number of months.

The interesting feature of an experiment is that each of its possible outcomes can be specified before the experiment is performed. The set of all possible outcomes is the **sample space** corresponding to an experiment. Each possible outcome is called a point, or an element, of the sample space. The following are some examples.

Ex A Driving to work, a commuter passes through a sequence of three intersections with traffic lights. At each light, she either stops, s, or continuous, c. The sample space is

$$\Omega = \{ccc, ccs, css, csc, sss, ssc, scc, scs\}.$$

Ex B The number of jobs in a print queue of a mainframe computer may be modeled as random. The sample space can be

$$\Omega = \{0, 1, 2, 3, \ldots\}.$$

2.2 Event

An **event** A is a certain subset of possible outcomes in the space Ω. We say the event A occurs if the outcome of the experiment is in the set A.

In Example A, the event that the commuter stops at the first light is the subset of Ω denoted by

$$A = \{sss, ssc, scc, scs\}.$$

In Example B, the event that there are fewer than five jobs in the print queue can be denoted by

$$A = \{0, 1, 2, 3, 4\}.$$

2.3 Set theory

The algebra of set theory carries over directly into probability theory. It is said that an event A is contained in another event B if every outcome that belongs to the subset defining the event A also belongs to the subset defining the event B. The relation is denoted by $A \subset B$. Mathematically, we have

$$A \subset B \Leftrightarrow x \in A \Rightarrow x \in B \qquad \text{(containment)}.$$

If two events are so related that $A \subset B$ and $B \subset A$, then $A = B$, which means that A and B must contain exactly the same sample points. Mathematically, we have

$$A = B \Leftrightarrow A \subset B \quad \text{and} \quad B \subset A. \qquad \text{(equality)}.$$

A subset of S that contains no outcomes is called the empty set, or null set, and it is denoted by the symbol $\emptyset$. For each event, it is true that $\subset A \subset S$.

Given any two events (or sets) A and B, we have the following elementary set operations:

Union: The union of A and B, written $A \cup B$, is the set of elements that belong to either A or B or both:

$$A \cup B = \{x : x \in A \text{ or } x \in B\}.$$

Intersection: The intersection of A and B, written $A \cap B$, is the set of elements that belong to both A and B:

$$A \cap B = \{x : x \in A \text{ and } x \in B\}.$$

Complements: The complement of A, written A^c, is the set of all elements that are not in A:

$$A^c = \{x : x \notin A\}.$$

Disjoint events: It is said that two events A and B are disjoint, or mutually exclusive, if A and B have no outcomes in common. It follows that A and B are disjoint if and only if $A \cap B = \emptyset$. The events $A_1, \ldots, A_n$ or the events $A_1, A_2, \ldots$ are disjoint if for every $i \neq j$, we have that A_i and A_j are disjoint, that is, $A_i \cap A_j = \emptyset$ for all $i \neq j$.

Ex A Consider the experiment of selecting a card at random from a standard deck and noting its suit: clubs (C), diamond (D), hearts (H), or spades (S). The sample space is

$$S = \{C, D, H, S\},$$

and some possible events are

$$A = \{C, D\}, \quad \text{and} \quad B = \{D, H, S\}.$$

From these events we can form

$$A \cup B = \{C, D, H, S\}, \quad A \cap B = \{D\}, \quad \text{and} \quad A^c = \{H, S\}.$$

Furthermore, notice that $A \cup B = S$ and $(A \cup B)^c = \emptyset$.

3 Probability Measures

3.1 Axioms and Basic Theorems

A probability measure on Ω is a function P from subsets of Ω to the real numbers that satisfies the following axioms:

Axiom 1 $P(\Omega) = 1$. If an event is certain to occur, then the probability of that event is 1.

Axiom 2 For every event A, $P(A) \geq 0$. In words, the probability of every event must be nonnegative.

Axiom 3 If A_1 and A_2 are disjoint, then

$$P(A_1 \cup A_2) = P(A_1) + P(A_2).$$

More generally, if $A_1, A_2, \ldots, A_m, \ldots$ are mutually disjoint, then

$$P\big(\bigcup_{i=1}^{\infty} A_i\big) = \sum_{i=1}^{\infty} P(A_i).$$

The mathematical definition of probability can now be given as follows: A *probability distribution*, or simply a *probability*, on a sample space Ω is a specification of numbers $P(A)$ that satisfy Axioms 1, 2, and 3.

Theorem 3.1 $P(\emptyset) = 0$.

Proof: Since $\emptyset \cup \emptyset = \emptyset$. Therefore, it follows from Axiom 3 that

$$P(\emptyset) = 2P(\emptyset).$$

The only real number with this property is 0. □

.

Theorem 3.2 *For every finite sequence of n disjoint events $A_1, \ldots, A_n$,*

$$P\big(\sum_{i=1}^{n} A_i\big) = \sum_{i=1}^{n} P(A_i).$$

Proof: Let $A_i = \emptyset$ for $i > n$. Therefore,

$$P\big(\sum_{i=1}^{n} A_i\big) = P\big(\sum_{i=1}^{\infty} A_i\big) = \sum_{i=1}^{n} P(A_i) + 0 = \sum_{i=1}^{n} P(A_i).$$

The proof is completed. □

Theorem 3.3 *For every event A, $P(A^c) = 1 - P(A)$.*

PROOF: Since $A \cup A^c = \Omega$, then $P(A) + P(A^c) = P(\Omega) = 1$ and $P(A^c) = 1 - P(A)$. $\square$

Theorem 3.4 *If $A \subset B$, then $P(A) \leq P(B)$.*

PROOF: Since B can be expressed as the union of two disjoint sets: $B = A \cup (B \cap A^c)$, the third axiom implies that $P(B) = P(A) + P(B \cap A^c) \geq P(A)$. $\square$

Theorem 3.5 *For any event A, $0 \leq P(A) \leq 1$.*

PROOF: For any event A, we have $\emptyset \subset A \subset \Omega$. Therefore, $0 \leq P(A) \leq 1$. $\square$

Theorem 3.6 *For every two events A and B,*

$$P(A \cup B) = P(A) + P(B) - P(AB).$$

PROOF: This is easy to see from the Venn diagram. Define $C = A \cap B^c$, $D = A \cap B$ and $E = A^c \cap B$. We then have, from the third axiom,

$$P(A \cup B) = P(C) + P(D) + P(E).$$

Also, $P(A) = P(C) + P(D)$ and $P(B) = P(D) + P(E)$. Putting these results together, we see that

$$P(A) + P(B) = P(C) + P(E) + 2P(D) = P(A \cup B) + P(D),$$

or

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

$\square$

Ex A Suppose that a fair coin is thrown twice. Let A denote the event of heads on the first toss, and let B denote the event of heads on the second toss. The sample space is

$$\Omega = \{hh, ht, th, tt\}.$$

We assume that each outcome in Ω is equally likely and has probability $1/4$. Then

$$P(C) = P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.5 + 0.5 - 0.25 = 0.75.$$

4 Computing Probabilities: Counting Methods

In an experiment of which the sample space contains only a finite number of points $s_1, \ldots, s_n$, a probability distribution on Ω is specified by assigning a probability p_i to each point $s_i \in S$. In order to satisfy the axioms for a probability distribution, the numbers $p_1, \ldots, p_n$ must satisfy the following two conditions:

$$p_i \geq 0, \qquad \text{for } i = 1, \ldots, n$$

and

$$\sum_{i=1}^{n} p_i = 1.$$

The probability of each even A can then be found by adding the probabilities p_i of all outcomes s_i that belong to A.

If all the outcomes are equally likely, then

$$P(A) = \frac{\text{number of ways } A \text{ can occur}}{\text{total number pf outcomes}}.$$

Ex A Suppose that three fair coins are tossed simultaneously. We shall determine the probability of obtaining exactly two heads.

The sample space is $\{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}$. It is natural to assume that the probability of each of these outcomes is 1/8. Therefore, the probability of obtaining exactly two heads is 3/8.

4.1 Multiplication rule

In many experiments, the number of outcomes in Ω is so large that a complete listing of these outcomes is too expensive, too slow, or too likely to be incorrect to be useful. In such an experiment, it is convenient to have a method of determining the total number of outcomes in the space Ω and in various events in Ω without compiling a list of all these outcomes.

Multiplication Principle If one experiment has m outcomes and another experiment has n outcomes, then there are mn possible outcomes for the two experiments.

The sample space will be composed of the following mn outcomes.

$$(x_1, y_1)(x_1, y_2) \dots (x_1, y_n)$$
$$(x_2, y_1)(x_2, y_2) \dots (x_2, y_n)$$
$$\vdots$$
$$(x_m, y_1)(x_m, y_2) \dots (x_m, y_m).$$

Ex A Playing cards have 13 face values and 4 suits. There are thus $4 \times 13 = 52$ face-value/suit combinations.

Ex B A class has 12 boys and 18 girls. The teacher selects 1 boy and 1 girl to do an experiment. She can do this in any of $12 \times 18 = 216$ different ways.

Ex C Suppose that there are three different routes from city A to city B and five different routes from city B to city C. Then the number of different routes from A to C that pass through B is $3 \times 5 = 15$.

Extended Multiplication Principle If there are p experiments and the first has n_1 possible outcomes, the second n_2, ..., and the pth n_p possible outcomes, then there are a total of $n_1 \times n_2 \times \cdots \times n_p$ possible outcomes for the p experiments.

4.2 Permutations

A permutation is an ordered arrangement of objects. Suppose that from the set $C = \{c_1, c_2, \dots, c_n\}$ we choose r elements and list them in order. The answer depends on whether we are allowed to duplicate items in the list. If no duplication is allowed, we are sampling without replacement. If duplication is allowed, we are sampling with replacement.

Proposition A For a set of size n and a sample of size r, there are n^r different ordered samples with replacement and $n(n-1)(n-2)\cdots(n-r+1)$ different ordered samples without replacement.

Corollary A The number of orderings of n elements is $n(n-1)\cdots 1 = n!$. Note that $0! = 1$.

Ex A (Arranging books) Suppose that six different books are to be arranged on a shelf. The number of possible permutations of the books is $6! = 720$.

Ex B (Ordered sampling without replacement) Suppose that a club consists of 25 members, and that a manager and a secretary are to be chosen from the membership. The number of all possible choices is $25 \times 24 = 600$.

Ex C A box contains n different balls. Suppose that sampling is done with replacement. So that there are n^r sample.

Ex D (The birthday problem) The problem is to determine the probability p that at least two people in a group of k people ($2 \le k \le 365$) will have the same birthday.

Under appropriate assumptions (independence and uniformity), the probability that all k persons will have different birthdays is $365 \times 364 \times (365 - k + 1)/365^k$. The probability that at least two of the people will have the same birthday is

$$p = 1 - 365 \times 364 \times (365 - k + 1)/365^k = 1 - \frac{365!}{(365-k)!365^k}.$$

k	p	k	p
5	0.027	25	0.569
10	0.117	30	0.706
15	0.253	40	0.891
20	0.411	50	0.970
22	0.476	60	0.994
23	0.507	100	0.999997

Ex E How many people must you ask to have a 50:50 chance of finding someone who shares your birthday? Let A be the event that someone's birthday is the same as yours. Then

$$P(A) = 1 - P(A^c) = 1 - \frac{364^k}{365^k}.$$

For the probability to be 0.5, k should be 253.

4.3 Combinations

Suppose that there is a set of n distinct elements from which it is desired to choose a subset containing k elements ($1 \le k \le n$). The number of different subsets that can be chosen is

$$\binom{n}{k} = \frac{n!}{(n-k)!k!}.$$

A subset is called a combination. For example, if the set contains four elements a, b, c and d and if each subset is to consist of two of these elements, then the following six different combinations can be obtained:

$$(a, b), (a, c), (a, d), (b, c), (b, d), (c, d).$$

When combinations are considered, the subsets (a, b) and (b, a) are identical and only one of these subsets is counted.

The number $\binom{n}{k}$ is called a binomial coefficient because it appears in the binomial theorem.

Proposition B For all numbers x and y and each positive integer n,

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}.$$

Note that

$$\binom{n}{0} = \binom{n}{n} = 1, \quad \binom{n}{k} = \binom{n}{n-k}, \quad \text{and} 2^n = \sum_{k=0}^{n} \binom{n}{k}.$$

Ex A (Unordered sampling without replacement) Suppose that a class contains 15 boys and 30 girls, and that 10 students are to be selected for a special assignment. We shall determine the probability that exactly 3 boys will be selected.

The desired probability is

$$p = \frac{\binom{15}{3}\binom{30}{7}}{\binom{45}{10}} = 0.2904.$$

Ex B The so-called capture/recapture method is sometimes used to estimate the size of a wildlife population. Suppose that 10 animals are captured. tagged, and released. On a latter occasion, 20 animals are captured, and it is found that 4 of them are tagged. How large is the population.

We assume that there are n animals in the population. The probability that 4 is recaptured is

$$P(n) = \frac{\binom{10}{4}\binom{n-10}{16}}{\binom{n}{20}}.$$

The population size can then be estimated by the value of n that makes the observed outcome most probable, i.e., maximizing the probability $P(n)$. The result is 50.

Proposition C The number of ways that n objects can be grouped into r classes with n_i in the ith class, $i = 1, \ldots, r$, and $\sum_{i=1}^{r} n_i = n$ is

$$\binom{n}{n_1, n_2, \cdots, n_r} = \frac{n!}{n_1! n_2! \cdots n_r!}.$$

The numbers $\binom{n}{n_1,n_2,\cdots,n_r}$ are called multinomial coefficients. They occur in the expansion

$$(x_1 + x_2 + \cdots + x_r)^n = \sum \binom{n}{n_1, n_2, \cdots, n_r} x_1^{n_1} x_2^{n_2} \cdots x_r^{n_r},$$

where the sum is over all nonnegative integers $n_1, n_2, \ldots, n_r$ such that $n_1 + n_2 + \cdots + n_r = n$.

Proof: There are $\binom{n}{n_1}$ ways to choose the objects for the first class. Having done that, there are $\binom{n-n_1}{n_2}$ ways of choosing the objects for the second class. Continuing in this manner, there are

$$\frac{n!}{n_1!(n-n_1)!} \frac{(n-n_1)!}{(n-n_1-n_2)!n_2!} \cdots \frac{(n-n_1-n_2-\cdots-n_{r-1})!}{0!n_r!}$$

choices in all. After cancellation, this yields the desired result. □

Ex A A committee of seven members is to be divided into three subcommittees of size three, two and two. This can be done in

$$\binom{7}{3,2,2} = \frac{7!}{3!2!2!} = 210.$$

Ex B In how many ways can the set of nucleotides $\{A, A, G, G, G, G, C, C, C\}$ be arranged in a sequence of nine letters?

$$\binom{9}{2,4,3} = \frac{9!}{2!4!3!} = 1260.$$

Ex C In how many ways can $n = 2m$ people be paired and assigned to m courts for the first round of a tennis tournament?

$$\binom{2m}{2,2,\cdots,2} = \frac{(2m)!}{2^m}.$$

Ex D **Playing cards** A deck of 52 cards contains 13 hearts. Suppose that the cards are shuffled and distributed among four players A,B, C and D so that each player receives 13 cards. The probability that A will have 6 hearts, B will have four hearts, C will have two hearts, and D will have one hearts is

$$p = \frac{\binom{13}{6,4,2,1}\binom{39}{7,9,11,12}}{\binom{52}{13,13,13,13}} = 0.00196.$$

5 Conditional Probability

A major use of probability in statistical inference is the updating of probabilities when certain events are observed. The updated probability of event A after we learned that event B has occurred is the conditional probability of A given B. The notation for this conditional probability is $P(A|B)$.

If we know that the event B has occurred, then we know that the outcome of the experiment is one of those included in B. Hence, to evaluate the probability that A will occur, we must consider the set of those outcomes in B that also result in the occurrence of A. These considerations lead to the following definition:

Definition 5.1 *Let A and B be two events such that $P(B) > 0$, then the conditional probability of A given B is defined to be*

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

The conditional probability $P(A|B)$ is not defined if $P(B) = 0$.

Rolling Dice Suppose that two dices were rolled and it was observed that the sum T of the two numbers was odd. We shall determine the probability that T was less than 8.

Let A denote the event that $T < 8$ and let B denote the event that T is odd, then AB is the event that T is 3,5, or 7. The sample space $S = \{(i, j) : i = 1, \ldots, 6, j = 1, \ldots, 6\}$, and each of the sample point is equally likely. Thus, we have

$$P(AB) = \frac{2}{36} + \frac{4}{36} + \frac{12}{36} = 1/3.$$

and

$$P(B) = \frac{2}{36} + \frac{4}{36} + \frac{12}{36} + \frac{4}{36} + \frac{2}{36} = 18/36 = 1/2.$$

The conditional probability is

$$P(A|B) = \frac{P(AB)}{P(B)} = 2/3.$$

Multiplication law Let A and B be events and assume $P(B) \neq 0$. Then

$$P(A \cap B) = P(A|B)P(B).$$

Ex A An urn contains three red balls and one blue ball. Two balls are selected without replacement. What is the probability that they are both red?

Let R_1 and R_2 denote the events that a red ball is drawn on the first trial and on the second trial, respectively. From the multiplication law,

$$P(R_1 \cap R_2) = P(R_1)P(R_2|R_1) = \frac{3}{4}\frac{2}{3} = \frac{1}{2}.$$

Ex B Suppose that if it is cloudy (B), the probability that it is raining (A) is 0.3, and that the probability that it is cloudy is $P(B) = 0.2$. The probability that it is cloudy and raining is

$$P(A \cap B) = P(A|B)P(B) = 0.3 \times 0.2 = 0.06.$$

Law of total probability Let $B_1, B_2, \ldots, B_n$ be such that $\cup_{i=1}^n B_i = \Omega$ and $B_i \cap B_j = \emptyset$ for $i \neq j$, with $P(B_i) > 0$ for all i. Then, for any event A,

$$P(A) = \sum_{i=1}^{n} P(A|B_i)P(B_i).$$

Proof:

$$\begin{aligned} P(A) &= P(A \cap \Omega) = P\left(A \cap (\cup_{i=1}^n B_i)\right) \\ &= P\left(\cup_{i=1}^n (A \cap B_i)\right) \\ &= \sum_{i=1}^{n} P(A \cap B_i) \quad \text{(the events } A \cap B_i \text{ are disjoint)} \\ &= \sum_{i=1}^{n} P(A|B_i)P(B_i), \end{aligned}$$

which concludes the proof. □

Ex C Referring to Example A, what is the probability that a red ball is selected on the second draw?

$$P(R_2) = P(R_2|R_1)P(R_1) + P(R_2|B_1)P(B_1) = \frac{2}{3} \times \frac{3}{4} + 1 \times \frac{1}{4} = \frac{3}{4}.$$

Ex D Suppose that occupations are grouped into upper (U), middle (M), and lower (L) levels. U_1 will denote the event that a father's occupation is upper-level; U_2 will denote the event that a son's occupation is upper-level, etc. Glass and Hall (1954) complied the following statistics on occupational mobility in England and Wales:

	U_2	M_2	L_2
U_1	0.45	0.48	0.07
M_1	0.05	0.70	0.25
L_1	0.01	0.50	0.49

Suppose that of the father's generation, 10% are in U, 40% in M, and 50% in L. The probability that a son in the next generation is in U is

$$P(U_2) = P(U_2|U_1)P(U_1) + P(U_2|M_1)P(M_1) + P(U_2|L_1)P(L_1) = 0.07.$$

The conditional probability

$$P(U_1|U_2) = \frac{P(U_1 \cap U_2)}{P(U_2)} = \frac{0.45 \times 0.1}{0.07} = 0.64.$$

Bayes' Rule Let A and $N_1, \ldots, B_n$ be events where the B_i are disjoint, $\cup_{i=1}^n B_i = \Omega$, and $P(B_i) > 0$ for all i. Then

$$P(B_i|A) = \frac{P(B_i)P(A|B_i)}{\sum_{j=1}^n P(B_j)P(A|B_j)}.$$

Ex F Polygraph test (lie-detector test) are often routinely administered to employees or prospective employees in sensitive positions. Let + denote the event that the polygraph reading is reading, indicating that the subject is lying; let T denote the event that the subject is telling the truth; and let L denote the event that the subject is lying. According to studies of polygraph reliability $P(+|L) = 0.88$, $P(-|L) = 0.12$, $P(-|T) = 0.86$ and $P(+|T) = 0.14$. Suppose that for a particular question the vast majority of subjects have no reason to lie so that $P(T) = 0.99$ and $P(L) = 0.01$. A subject produces a positive response on the polygraph. The probability that she is in fact telling truth is

$$\begin{aligned} P(T|+) &= \frac{P(+|T)P(T)}{P(+|T)P(T) + P(+|L)P(L)} \\ &= \frac{(0.14)(0.99)}{(0.14)(0.99) + (0.88)(0.01)} = 0.94 \end{aligned}$$

Ex G **Identifying the course of a defective item.** Three different machines M_1, M_2 and M_3 were used for producing a large batch of similar manufactured items. Suppose that 20, 30, and 50 percents of the items were produced by M_1, M_2 and M_3, respectively. Suppose further that 1 percent of the items produced by M_1 are defective, that 2 percents of the items produced by M_2 are defective, and that 3 percents of the items produced by M_3 are defective. Finally, suppose that one item is selected at random from the entire batch, and it is found to be defective.

Let B_i be the event that the selected item was produced by M_i, $i = 1, 2, 3$, and let A be the event that the selected item is defective. The probability that the selected item was produced by M_2 is

$$\begin{aligned} P(B_2|A) &= \frac{P(B_2)P(A|B_2)}{\sum_{i=1}^3 P(B_i)P(A|B_i)} \\ &= \frac{(0.3)(0.02)}{(0.2)(0.01) + (0.3)(0.02) + (0.5)(0.03)} = 0.26. \end{aligned}$$

6 Independence

Definition 6.1 *Two events are independent if* $P(AB) = P(A)P(B)$.

Suppose that $P(A) > 0$ and $P(B) > 0$, then it follows from the definitions of independence and conditional probability that A and B are independent if and only if $P(A|B) = P(A)$ and $P(B|A) = P(B)$, that is, knowing that one had occurred gave us no information about whether the other had occurred.

Ex A A card is selected randomly from a deck. Let A denote the event that it is an ace and D the event that it is a diamond. Checking the probabilities: $P(A) = 4/52 = 1/13$, $P(D) = 1/4$, $P(A \cap D) = 1/52$. Therefore, A and D are independent.

Ex B **Machine operations**. Suppose that two machines 1 and 2 in a factory are operated independently of each other. Let A be the event that machine 1 will become inoperative during a given 8-hour period; let B be the event that machine 2 will become inoperative during the same period; and suppose that $P(A) = 1/3$ and $P(B) = 1/4$.

The probability that both machines will become inoperative during the period is

$$P(AB) = P(A)P(B) = 1/12.$$

The probability that at least one of the machines will become inoperative during the period is

$$P(A \cup B) = P(A) + P(B) - P(AB) = \frac{1}{3} + \frac{1}{4} - \frac{1}{12} = \frac{1}{2}.$$

Things become more complicated when we consider more than two events.

Ex C **Pairwise independence**. Consider an experiment in which the sample space S contains four outcomes $\{s_1, s_2, s_3, s_4\}$, and suppose that the probability of each outcome is $1/4$. Let the three events A, B and C be defined as follows:

$$A = \{s_1, s_2\}, \quad B = \{s_1, s_3\}, \quad C = \{s_1, s_4\}.$$

Then $AB = AC = BC = ABC = \{s_1\}$. Hence,

$$P(A) = P(B) = P(C) = 1/2,$$

and

$$P(AB) = P(AC) = P(BC) = P(ABC) = 1/4.$$

These results can be summarized by saying that the events A, B, and C are pairwise independent, but all three events are not independent.

To encompass situations such as that in Example C, we define a collection of events, A_1, A_2, ..., A_n, to be mutually independent if for any subcollection, $A_1, \ldots, A_m$,

$$P(A_{i_1} \cap \cdots \cap A_{i_m}) = P(A_{i_1}) \cdots P(A_{i_m}).$$

Ex D Suppose that a machine produces a defective item with probability p $(0 < p < 1)$ and produces a nondefective item with probability $q = 1 - p$. Suppose further that six items produced by the machine are selected at random and inspected, and that the outcomes for these six items are independent. The probability that at least one of the six items will be defective is

$$p = 1 - P(D_1 D_2 D_3 D_4 D_5 D_6) = 1 - q^6,$$

where D_i denotes the event that the ith item will be defective.

The probability that exactly two of six items are defective is

$$p = \binom{6}{2} p^2 q^4,$$

where the binomial coefficient is the total number of distinct arrangement of two defective items and four nondefective items.