概率论与数理统计

Probability and Statistics

孔令臣 王立春 编著

$$P(A)=\lim_{n\to\infty}\frac{n(A)}{n}$$

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

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

北京交通大学出版社
http://www.bjtup.com.cn

概率论与数理统计

Probability and Statistics

孔令臣　王立春　编著

· 北京 ·

内 容 简 介

本书共 8 章，主要内容包括随机事件和概率、随机变量及其分布、随机变量的数字特征、大数定律和中心极限定理、数理统计的基本知识、参数估计、假设检验、方差分析。

本书可作为普通高等院校理工、经济管理类各专业的教材，也可供对概率论与数理统计感兴趣的读者参考。

图书在版编目（CIP）数据

概率论与数理统计 = Probability and Statistics/孔令臣，王立春编著．—北京：北京交通大学出版社，2015.3

ISBN 978-7-5121-2024-2

Ⅰ.① 概…　Ⅱ.① 孔…　② 王…　Ⅲ.① 概率论-高等学校-教材　② 数理统计-高等学校-教材　Ⅳ.① O21

中国版本图书馆 CIP 数据核字（2014）第 183905 号

责任编辑：黎　丹

出版发行：北京交通大学出版社　　电话：010-51686414

北京市海淀区高梁桥斜街 44 号　　邮编：100044

印 刷 者：北京艺堂印刷有限公司

经　　销：全国新华书店

开　　本：170×235　印张：10.25　字数：197 千字

版　　次：2015 年 3 月第 1 版　2015 年 3 月第 1 次印刷

书　　号：ISBN 978-7-5121-2024-2/O·139

印　　数：1～500 册　定价：32.00 元

Preface

Since 2010, we have admitted an international class, where the students are trained in English for all the courses. For the course, Probability and Statistics, we have tried to find an appropriate textbook written in English to use for the international class. For several years teaching, we have used three different textbooks. Since in our university, all the students need to master the similar knowledge and skills in the course of Probability and Statistics. However, to the best of our knowledge, there is no textbook written in English which has the similar contents as the one used in other Chinese classes. When we give the lectures, the students always want to own one textbook instead of several references. This motivates us to compile this textbook in English, which covers all the important knowledge and techniques as the Chinese students need to master. This textbook is based on our teaching experimences both in English and Chinese.

This book has eight Chapters, which provide an elementary introduction to probability and statistics with applications. Topics include: data description via graph and numerical measures, interpretations of probability, axioms of probability, and the use of counting methods for solving probability problems, conditional probability, Bayesian theorem, independence, random variables and distributions, expected values, the binomial, Poisson, normal and other distributions, the law of large numbers, and the central limit theorem. We will also cover applications of the above to the theory of statistical inference, including point estimation theory, confidence intervals, and hypothesis tests.

This book emerges from a very nice collaboration between the authors and their students. We acknowledge many people who have helped to improve the quality of the book. We would like to express our gratitude to our families for providing a different, interesting, supportive and beautiful environment. The work was also supported in part by the National Basic Research Program of China (2010CB732501), and the National Natural Science Foundation of China (11171018, 11371051).

Comments and suggestions for further improving the publication are welcome.

Dr. Lingchen Kong
Dr. Lichun Wang
Beijing Jiaotong University

Contents

Chapter 1 Probability

§1.1 Introduction

The term probability refers to the study of randomness and uncertainty. In any situation in which one of a number of possible outcomes may occur, the discipline of probability provides methods for quantifying the chances,or likelihood, associated with the various outcomes.The language of probability is constantly used in an informal manner in both written and spoken contexts. In this chapter, we first introduce some elementary probability concepts, indicate how probability can be interpreted, and show how the rules of probability can be applied to compute the probability of many interesting events.Then, we will introduce one of the most important concepts in probability theory, that is, conditional probability.

The study of probability as a branch of mathematics goes back over 300 years, when it had its genesis in connection with questions involving games of chance. Many books are devoted exclusively to probability, but our objective here is to cover only that part of the subject that has the most direct bearing on problems of statistical inference.

§1.2 Sample Space and Events

An experiment is any activity or process whose outcome is uncertain. Although the word experiment generally suggests a planned or carefully controlled laboratory testing situation, we use it here in a much wider sense. Thus experiments may be of interest including tossing a coin once or several times, selecting a card or cards from a deck, or measuring the compressive strengths of different steel beams. So, let's give the sample space of an experiment.

Definition 1.2.1 Consider an experiment whose outcome is not predictable with certainty. However, although the outcome of the experiment will not be known in advance, let us suppose that the set of all possible outcomes is known. This set of all possible outcomes of an experiment is known as the sample space of the experiment and is denoted by S or Ω.

The following are some examples.

Example 1.2.1 If the outcome of an experiment consists in the determination of the sex of a newborn baby, then

$$S = \{\text{girl}, \text{boy}\}.$$

■

Example 1.2.2 If the outcome of an experiment is the order of finish in a race among the 7 horses having post position 1, 2, 3, 4, 5, 6 and 7, then

$$S = \{\text{permutations of}(1,2,3,4,5,6,7)\}.$$

■

The outcome (2,3,1,6,5,4,7) means, for instance, that the number 2 horse comes in first, next the number 3 horse, then the number 1 horse, and so on.

Example 1.2.3 If the experiment consists of flipping two coins,then the sample space consists of the following four points:

$$S=\{(H,H),(H,T),(T,H),(T,T)\}.$$

H means head,while T means tail.

■

The outcome will be (H,H) if both coins are heads, (H,T) if the first coin is head and the second is tail, (T,H) if the first is tail and the second is head, and (T,T) if both coins are tails.

Example 1.2.4 If the experiment consists of tossing two dice, then the sample space consists of 36 points

$$S=\{(i,j)|i,j=1,2,3,4,5,6\}.$$

where the outcome (i, j) is said to occur if i appears on the first die and j on the other die.

■

Definition 1.2.2 An event A is a subset of a sample space S. In other words,an event is a set consisting of possible outcomes of the experiments. If the outcome of the experiment is contained in A, then we say that A has occurred.Events will be denoted by capital letters A, B, C, $\cdots$

The following are examples of some events from the examples above.

From example 1.2.1, we have A={girl}, B={boy}.

From example 1.2.2, we have A={all outcomes in S starting with 3}, B={all outcomes in S starting with 7}.

From example 1.2.3, we have $A = \{(H,H),(H,T)\}, B = \{(H,H),(T,T)\}$.

From example 1.2.4, we have $A=\{(1,6),(2,5)\}$, $B=\{(2,6),(4,5)\}$.

Definition 1.2.3 The union of the two events A and B, denoted by the symbol $A \cup B$, is the event containing all the elements that belong to A or B or both in A and B. That is, the event $A \cup B$ will occur if either A or B occurs. For example:

From example 1.2.1, we have $A\cup B$={girl,boy}=S.

From example 1.2.2, we have $A\cup B$={all outcomes in S starting with 3 or 7}.

From example 1.2.3, we have $A\cup B=A=\{(H,H),(H,T),(T,T)\}$.

From example 1.2.4, we have $A\cup B=A=\{(1,6),(2,5),(2,6),(4,5)\}$.

Definition 1.2.4 The intersection of two events A and B, denoted by the symbol $A\cap B$(sometimes written as AB), is the event containing all elements that are common to A and B.

From example 1.2.3, we have $AB=\{(H,H)\}$.

Example 1.2.5 Let M={a,e,i,o,u} and N={r,s,t}, then it follows that $M\cap N=\varnothing$. That is M and N have no elements in common and, therefore,cannot both occur simultaneously.

■

Definition 1.2.5 Two events A and B are mutually exclusive, or disjoint if $A\cap B=\varnothing$, that is, if A and B have no elements in common.

Definition 1.2.6 We define unions and intersections of more than two events in a similar manner. If E_1, E_2, $\cdots$ are events, then the union of these events, denoted by $\bigcup\limits_{n=1}^{+\infty} E_n$, is defined to be that event which consists of all outcomes that are in E_n for at least one value of $n=1,2,\cdots$. Similarly, the intersection of the events E_n, denoted by $\bigcap\limits_{n=1}^{+\infty} E_n$, is defined to be the event consisting of those outcomes which are in all of the events E_n, $n=1,2,\cdots$.

Definition 1.2.7 The complement of an event A with respect to S is the subset of all elements of S that are not in A. We denote the complement of A by the symbol $\overline{A}$ or A^{c}. For example:

Example 1.2.6 In example 1.2.4, if event $A=\{(1,6),(2,5),(3,4),(4,3),(5,2),(6,1)\}$, then A^{c} will occur when the sum of the dice does not equal 7. We also have $S^{\mathrm{c}}=\varnothing$.

■

Example 1.2.7 Consider the sample space

$$S=\{\text{book, catalyst, cigarette, precipitate, engineer,rivet}\}.$$

Let A = {catalyst, rivet, book, cigarette},then

$$A^{\mathrm{c}}=\{\text{precipitate, engineer}\}.$$

■

Definition 1.2.8 For any two events E and F, if all the outcomes in E are also in F, then we say that E is contained in F, or say E is a subset of F, and write $E\subset F$(or equivalently,

$F \supset E$, which we sometimes say as F is a superset of E). If $E \subset F$ and $F \subset E$, we say that E and F are equal and write $F = E$.

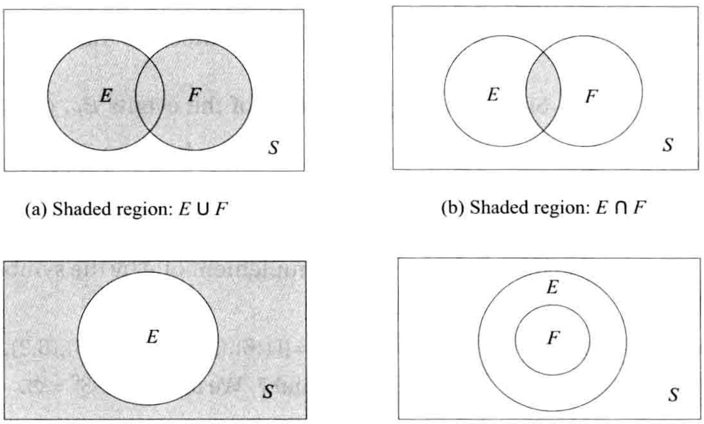
A graphical representation that is useful for illustrating logical relations among events is the Venn diagram. The sample space S is represented as consisting of all the outcomes in a large rectangle, and the events $E, F, G, \cdots$ are represented as consisting of all outcomes in given circles within the rectangle. Events of interest can then be indicated by shading appropriate regions of the diagram. For instance, in the four Venn diagram shown in Fig. 1.2.1, the shaded areas represent, respectively, the event $E \cup F$, $E \cap F$, E^c, $F \subset E$.

The operations of forming unions, intersections, and complements of events obey certain rules similar to the rules of algebra. We list a few of these rules:

Commutative laws: $E \cup F = F \cup E$, $\quad E \cap F = F \cap E$

Associative laws: $(E \cup F) \cup G = E \cup (F \cup G)$, $\quad (E \cap F) \cap G = E \cap (F \cap G)$

Distributive laws: $(E \cup F) G = E \cap G \cup F \cap G$, $\quad (E \cap F) \cup G = (E \cup G) \cap (F \cup G)$



(a) Shaded region: $E \cup F$ (b) Shaded region: $E \cap F$

(c) Shaded region: E^c (d) $F \subset E$

Fig. 1.2.1

These relations are verified by showing that any outcome that is contained in the event on the left side of the equality sign is also contained in the event on the right side, and vice versa. One way of showing this is by means of Venn diagram. Readers can prove those laws as exercises.

§1.3 Axioms of Probability

One way of defining the probability of an event is in terms of its relative frequency. Such a definition usually goes as follows. We suppose that an experiment, whose sample space is S, is repeatedly performed under the exactly the same conditions. For each event A of the sample space S,we define $n(A)$ to be the number of times in the first n repetitions of the experiment that the event A occurs. Then $P(A)$, the probability of the event A, is defined as

$$P(A) = \lim_{n\to\infty} \frac{n(A)}{n}$$

That is, $P(A)$ is defined as the (limiting) proportion of times that A occurs. It is thus the limiting frequency of A.

To insure that the probability assignment will be consistent with our intuitive notions of probability. Consider an experiment whose sample space is S. For each event A of the sample space S, we assume that a number $P(A)$ is defined and satisfies the following definition.

Definition 1.3.1 The probability of an event A is the sum of the weights of all sample points in A.Therefore,

Axioms 1 : $0 \leqslant \mathrm{P(A)} \leqslant 1$,

Axioms 2 : $P(S) = 1$,

Axioms 3: Furthermore,if $A_1, A_2, \cdots$ is a sequence of mutually exclusive events, then

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

We refer to $P(A)$ as the probability of the event A.

We can also deduce that $P(\varnothing) = 0$.

Example 1.3.1 If a die is rolled and we suppose that all six sides are equally likely to appear, then we would have

$$P(1) = P(2) = P(3) = P(4) = P(5) = P(6) = \frac{1}{6}.$$

From the Definition 1.3.1, it would thus follow that the probability of rolling an even number would equal

$$P(2,4,6) = P(2) + P(4) + P(6) = \frac{1}{2}.$$

■

Example 1.3.2 A coin is tossed twice. What is the probability that at least one head occurs?

Solution The sample space for this experiment is

$$S = \{HH, HT, TH, TT\}.$$

If the coin is balanced, each of this outcomes would be equally likely to occur. Therefore, we assign a probability of w to each sample point. Then $4w=1$,or $w=\frac{1}{4}$. If A represents the event of at least one head occurring,then

$$A=\{HH, HT, TH\} \qquad \text{and} \qquad P(A)=\frac{1}{4}+\frac{1}{4}+\frac{1}{4}=\frac{3}{4}.$$

■

Example 1.3.3 A die is loaded in such a way that an even number is twice as likely to occur as an odd number. If E is the event that a number less than 4 occurs on a single toss of the die, find $P(E)$.

Solution The sample space is $S=\{1,2,3,4,5,6\}$. We assign a probability of w to each odd number and a probability of $2w$ to each even number. Since the sum of the probability must be 1, we have $9w=1$ or $w=\frac{1}{9}$. Hence probability of $\frac{1}{9}$ and $\frac{2}{9}$ are assigned to each odd and even number, respectively. Therefore,

$$E=\{1,2,3\} \qquad \text{and} \qquad P(E)=\frac{1}{9}+\frac{2}{9}+\frac{1}{9}=\frac{4}{9}.$$

■

Example 1.3.4 In example 1.3.3 let A be the event that an even number turns up and let B be the event that a number divisible by 3 occurs. Find $P(A\cup B)$ and $P(A\cap B)$.

Solution For the events $A=\{2,4,6\}$ and $B=\{3,6\}$ we have

$$A\cup B=\{2,3,4,6\} \qquad \text{and} \qquad A\cap B=\{6\}.$$

By assigning a probability of $\frac{1}{9}$ to each odd number and $\frac{2}{9}$ to each even number, we have

$$P(A\cup B)=\frac{2}{9}+\frac{1}{9}+\frac{2}{9}+\frac{2}{9}=\frac{7}{9} \qquad \text{and} \qquad P(A\cap B)=\frac{2}{9}.$$

■

§1.4 Some Simple Propositions

In this section, we prove some simple propositions regarding probabilities.

Theorem 1.4.1 If $A\subset B$, then $P(A)\leqslant P(B)$.

Proof Since $A\subset B$, we can express B as

$$B=A\cup A^{c}B.$$

Hence,because A and A^cB are mutually exclusive, we obtain, from Axiom 3 ,

$$P(B)=P(A)+P(A^{c}B),$$

which proves the result, since $P(A^{c}B)\geqslant 0$.

Theorem 1.4.2 If A and B are two events,then

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

Proof From Axiom 3, we obtain

$$P(A \cup B) = P(A \cup A^c B) = P(A) + P(A^c B).$$

Furthermore, since $B = A \cap B \cup A^c \cap B$, we again obtain from Axiom 3,

$$P(B) = P(AB) + P(A^c B).$$

Thereby completing the proof.

Corollary 1 If A and B are mutually exclusive, then

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

Corollary 1 is an immediate result of Theorem 1.4.1. If A and B are mutually exclusive, $A \cap B = \varnothing$ and then $P(A \cap B) = P(\varnothing) = 0$.

Corollary 2 If $A_1, A_2, \cdots, A_n$ are mutually exclusive, then

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

A collection of events $A_1, A_2, \cdots, A_n$ of a sample space S is called a partition of S if $A_1, A_2, \cdots, A_n$ are mutually exclusive and $A_1 \cup A_2 \cup \cdots \cup A_n = S$. Thus we have

Corollary 3 If $A_1, A_2, \cdots, A_n$ is a partition of a sample space S, then

$$P(A_1 \cup A_2 \cup \cdots \cup A_n) = P(A_1) + P(A_2) + \cdots + P(A_n) = P(S) = 1.$$

As one might expect,Theorem 1.4.2 extends in an analogous fashion.

Theorem 1.4.3 For three events A, B and C,

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

Example 1.4.1 In a certain residential suburb, 60% of all households get Internet service from the local cable company, 80% get television service from that company,and 50% get both services from that company.If a household is randomly selected , what is the probability that it gets at least one of these two services from the company? And What is the probability that it gets exactly one of those services from the company?

Solution With A={gets Internet service} and B={gets TV service}, the given information implies that $P(A)=0.6$, $P(B)=0.8$, $P(A \cap B)=0.5$. The foregoing proposition now yields P(subscribe to at least one of the two services)$=P(A \cup B)= P(A)+P(B)-P(A \cap B) =0.6+0.8-0.5= 0.9$. The event that a household subscribes only to TV service can be written as $A^c \cap B$ [(not Internet)and TV]. Now Fig. 1.4.1 implies that

$$0.9 = P(A \cup B) = P(A) + P(A^c \cap B) = 0.6 + P(A^c \cap B),$$

from which $P(A^c \cap B) = 0.3$. Similarly,

$$P(A \cap B^c) = P(A \cup B) - P(B) = 0.1.$$

This is all illustrated in Fig. 1.4.1, from which we see that

$$P(\text{exactly one}) = P(A^c \cap B) + P(A \cap B^c) = 0.1 + 0.3 = 0.4.$$

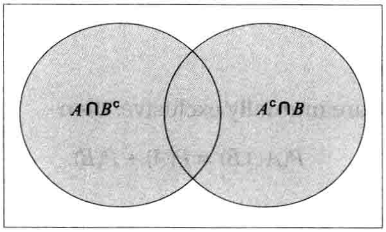


Fig. 1.4.1

■

Theorem 1.4.4 If A and A^c are complementary events, then

$$P(A) + P(A^c) = 1.$$

Proof Since $A \cup A^c = S$ and the sets A and A^c are disjoint, then

$$1 = P(S) = P(A \cup A^c) = P(A) + P(A^c).$$

Example 1.4.2 If the probability that an automobile mechanic will service 3, 4, 5, 6, 7, or 8 or more cars on any given workday is 0.12, 0.19, 0.28, 0.24, 0.10, and 0.07, respectively. What is the probability that he will service at least 5 cars on his next day at work?

Solution Let E be the event that at least 5 cars are serviced. Now, let $P(E) = 1 - P(E^c)$, where E^c is the event that fewer than 5 cars are serviced. Since

$$P(E^c) = 0.12 + 0.19 = 0.31,$$

it follows from Theorem 1.4.4 that

$$P(E) = 1 - 0.31 = 0.69.$$

■

§1.5 Sample Spaces Having Equally Likely Outcomes

In many experiments, it is natural to assume that all outcomes in the sample space are equally likely to occur. That is, consider an experiment whose sample space S is a finite set, say, $S=\{1,2,\cdots,N\}$, then it is natural to assume that

$$P(\{1\}) = P(\{2\}) = \cdots = P(\{N\}),$$

which implies, from Axiom 2 and 3, that

$$P(\{i\}) = \frac{1}{N}, \quad i = 1,2,\cdots,N.$$

From this equation, it follows from Axiom 3 that, for any event E

$$P(E) = \frac{\text{number of outcomes in } E}{\text{number of outcomes in } S}.$$

where S is the sample space.

Example 1.5.1 If two dice are rolled, what is the probability that the sum of the upturned faces will equal 7?

Solution We shall solve this problem under the assumption that all of the 36 possible outcomes are equally likely. Since there are 6 outcomes,namely, (1,6), (2,5), (3,4), (4,3), (5,2), and (6,1), that result in the sum of the dice being equal to 7,the desired probability is

$$\frac{6}{36} = \frac{1}{6}$$

■

Example 1.5.2 You have six unread mysteries on your bookshelf and six unread science fiction books. The first three of each type are hardcover, and the last three are paperback. Consider randomly selecting one of the six mysteries and then randomly selecting one of the six science fiction books to take on a post finals vacation to Acapulco (after all, you need something to read on the beach). Number the mysteries $1,2,\cdots,6$,and do the same for the science fiction books. Then each outcome is a pair of numbers such as (4,1), and there are $N = 36$ possible outcomes.With random selection as described, the 36 outcomes are equally likely. Nine of these outcomes are such that both selected books are paperbacks: $(4,4), (4,5), \cdots,(6,6)$. So the probability of the event A that both selected books are paperbacks is

$$P(A) = \frac{N(A)}{N} = \frac{9}{36} = 0.25.$$

■

Example 1.5.3 A committee of 5 is to be selected from a group of 6 men and 9 women.If the selection is made randomly, what is the probability that the committee consists of 3 men and 2 women?

Solution Because each of the $\binom{15}{5}$ possible committees is equally likely to be selected,the desired probability is

$$\frac{\binom{6}{3}\binom{9}{2}}{\binom{15}{5}}=\frac{240}{1001}.$$

■

§1.6 Conditional Probabilities

Suppose that we toss 2 dice, and suppose that each of the 36 possible outcomes is equally likely to occur and hence has the probability $\frac{1}{36}$. Suppose further that we observe that the first die is a 3. Then, given this information, what is the probability that the sum of the 2 dice is equals 8? To calculate this probability, we reason as follows. Given that the initial die is a 3, there can be at most 6 possible outcomes of our experiment, namely, (3,1), (3,2), (3,3), (3,4), (3,5), and (3,6). Since each of these outcomes originally had the same probability of occurring, the outcomes should still have equal probabilities. That is, given that the first die is a 3, the (conditional) probability of each of the outcomes (3,1), (3,2), (3,3), (3,4), (3,5) ,and (3,6) is $\frac{1}{6}$, whereas the (conditional) probability of the other 30 points in the sample space is 0. Hence, the desired probability will be $\frac{1}{6}$.

If we let A and B denote, respectively, the event that the sum of the die is 8 and the event that the first die is a 3, then the probability just obtained is called conditional probability that A occurs given that B occurred and is denoted by

$$P(A|B).$$

A general formula for $P(A|B)$ that is valid for all events A and B is derived in the same manner. If the event B occurs, then, in order for A to occur, it is necessary that the actual occurrence be a point both in A and in B; that is, it must be in AB. Now, since we know that B has occurred, it follows that B becomes our new, or reduced, sample space; hence, the probability the event $A\cap B$ occurs will equal the probability of AB relative to the probability of B. That is, we have the following definition.

Definition 1.6.1 The conditional probability of B, given A, denoted by $P(B|A)$ is defined by

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

As an additional illustration, suppose that our sample space S is the population of adults in a small town who have completed the requirement for a college degree, we shall categorize them according to sex and employment status in Table 1.1.

Table 1.1

	Employed	Unemployed	Total
Male	460	40	500
Female	140	260	400
Total	600	300	900

One of these individuals is to be selected at random for a tour throughout the country to publicize the advantages of establishing new industries in the town. We shall be concerned with the following events:

M: a man is chosen.

E: the one chosen is employed.

Using the reduced sample space E, we find that

$$P(M \mid E) = \frac{460}{600} = \frac{23}{30}.$$

Let $n(A)$ denote the number of elements in any set A. Using the notation, we can write

$$P(M|E) = \frac{n(E \cap M)}{n(E)} = \frac{n(E \cap M)/n(S)}{n(E)/n(S)} = \frac{P(E \cap M)}{P(E)},$$

where $P(M \cap E)$ and $P(E)$ are found from the original sample space S. To verify this result, note that

$$P(E) = \frac{600}{900} = \frac{2}{3} \quad \text{and} \quad P(M \cap E) = \frac{460}{900} = \frac{23}{45}.$$

Hence, $P(M|E) = \dfrac{23/45}{2/3} = \dfrac{23}{30}$, same as before.

Example 1.6.1 The conditional probability that a regularly scheduled flight departs on time is $P(D) = 0.83$; the probability that it arrives on time is $P(A) = 0.82$; and the probability that it departs and arrives on time is $P(D \cap A) = 0.78$. Find the probability that a plane

(1) arrives on time given that it departed on time, and

(2) departed on time given that it has arrived on time.

Solution (1) The probability that a plane arrives on time given that it departed on time is

$$P(A|D) = P(D \cap A)/P(D) = 0.78/0.83 = 0.94.$$

(2) The probability that a plane departed on time given that it has arrived on time is

$$P(D|A) = P(D \cap A)/P(A) = 0.78/0.82 = 0.95.$$

■