

CRC

**DECISION ANALYSIS
IN MEDICINE:
METHODS AND
APPLICATIONS**

Edward A. Patrick

CRC

PRESS

Decision Analysis in Medicine: Methods and Applications

Author

Edward A. Patrick

Professor of Electrical Engineering

Purdue University

West Lafayette, Indiana

and

Physician in Charge of Clinical Computing

The Jewish Hospital

Cincinnati, Ohio

and

Staff Physician

St. Elizabeth Hospital Medical Center

Lafayette, Indiana

and

Staff Physician

~~Dea~~ ~~ness~~ ~~Hospital~~

Cincinnati, Ohio



CRC PRESS, INC.
Boca Raton, Florida 33431

Library of Congress Cataloging in Publication Data

Patrick, Edward A. 1942-
Decision analysis in medicine.

Bibliography: p.

Includes index.

1. Medicine — Decision making. 2. Statistical
decision. I. Title.

R723.P37 616'.001'82 77-19085
ISBN 0-8493-5255-X

This book represents information obtained from authentic and highly regarded sources. Reprinted material is quoted with permission, and sources are indicated. A wide variety of references are listed. Every reasonable effort has been made to give reliable data and information, but the author and the publisher cannot assume responsibility for the validity of all materials or for the consequences of their use.

All rights reserved. This book, or any parts thereof, may not be reproduced in any form without written consent from the publisher.

© 1979 by CRC Press, Inc.

International Standard Book Number 0-8493-5255-X

Library of Congress Card Number 77-19085
Printed in the United States

PREFACE

New methods exist to model and evaluate the diagnosis treatment process together with the outcome. Also, new methods exist for simultaneous processing of signs, symptoms, and laboratory tests which can make available more accurate decisions for physicians to use in patient management. These methods for both the evaluation and improved decision making have foundations in statistical pattern recognition, a discipline under development since about 1960. During the last few years the phrase medical decision analysis has increasingly been applied to the study of these methods.

Basic to the methods is estimation of the multidimensional probability distribution of the joint collection of signs, symptoms, and laboratory tests (features). There is increased information in the multidimensional representation of features over that in the representation of one or two features. Whereas a physician is limited to the latter low dimensional processing, a digital computer can implement the multidimensional representation to achieve more accurate decisions. Similarly, evaluation of outcome is more accurate when a multidimensional representation is used.

Whereas more accurate decisions and evaluations are now possible with the new methods, it must be recognized that the physician (as a manager), rather than a digital computer, must analyze the decisions and evaluations presented to him. Put differently, the physician has a substantial field of knowledge about his patient and the cultural and technological assets or liabilities available to manage the patient's health. On the other hand, the physician should understand that the methods described can offer him more accurate decisions and evaluations on the basis of which he should manage the patient's health.

On the premise that the physician rather than a machine will manage the patient, the physician must be provided with well-defined decisions within the context of his sphere of operation. Ideally, those clinical problems should be identified which have a fixed structure and frequently recur in clinical practice. For example, a decision from a 12 lead electrocardiogram is in this category, as is the decision from a CAT scan. Other examples include the differential diagnosis of chest pain, acute abdominal pain, pancreas disease, renal disease, and cardiac arrhythmia. An example is evaluating outcomes for various treatments of breast cancer. Other examples include evaluating the expected outcome of a brain scan, the expected outcome of surgical vs. medical treatment of coronary artery disease, or the expected outcome of surgical treatment vs. no treatment of pancreas cancer.

The first part of this book deals with the methods while the second part deals with applications. One objective of this book is presentation of methods so that in the future applications can be precisely described and evaluated. The author thus has converted many of the applications presented to the basic methodology. This makes possible a precise description of the application, the assumptions made, and possible extensions.

The book begins with an introduction to probability theory and then proceeds to a presentation of statistical decision rules. At this point classes, features, complex classes, subclasses, complex features, a priori class probabilities, a posteriori probabilities, differential diagnosis, and class conditional probability density functions are defined and discussed. This is all for a well-defined, structured problem in differential diagnosis or decision making.

Next a model for expected outcome loss is presented whereby outcomes for various treatments and decisions can be evaluated. Many special cases of expected outcome loss are developed, and later in the applications part of the book the special cases of expected outcome loss are developed.

A distinction then is made between consulting and diagnosis. Various forms of consulting are discussed including the rather simple concept of a protocol. Then after

defining subsystems, activation rules, sequential feature extraction, and new subsystems, a model for automated computer assisted diagnosis and consulting is presented.

Because the needs for assisted diagnosis and consulting of a busy emergency room physician, an internist, a surgeon a dermatologist, or a clinical pathologist are different, a single system for automated computer assisted diagnosis and consulting is not feasible. Furthermore, because of the differences in cultural and operating spheres imposed on physicians, automated computer-assisted diagnosis and consulting is doomed if the attempt is to make it too general. Thus methods for accurate decisions and/or evaluations of well-structured, repeating problems have the greatest likelihood of clinical usefulness.

Applications presented include to treat or not to treat, expected outcome of treating pancreas cancer, coronary artery disease, and breast cancer with surgery. Other applications include the computer-assisted diagnosis of renal disease, abdominal diseases, thyroid diseases, congenital heart disease, acute chest pain, hematological disease, some disease in ophthalmology, chest trauma, and liver diseases. The predicted outcome of a radionuclear brain scan also is presented.

The advent of the fast microcomputer and the theory presented in this book made possible the development of CONSULT I® in 1979. CONSULT I® implements basic consulting, inverse retrieval, and automatic decision making for well-structured, recurring problems. It was appropriate to develop the inexpensive CONSULT I® for the field of emergency medicine and ambulatory care.

Furthermore, CONSULT II® was developed with a videodisc player interfaced to CONSULT I®. The CONSULT II® system with the videodisc player adds color pictures or motion picture segments to basic consulting. Important uses include dermatology, radiology, electrocardiograms, hematology, and other areas of medicine.

Edward A. Patrick, M.D., Ph.D.
January 1979

THE AUTHOR

Dr. Edward A. Patrick received a Ph.D. in Electrical Engineering from Purdue University and the M.D. degree from Indiana University School of Medicine. His graduate medical education includes completion of a residency combined in medicine and surgery. Previously, he received the B.S. and M.S. degrees both in Electrical Engineering from Massachusetts Institute of Technology.

Dr. Patrick is President of the Systems, Man, and Cybernetics Society of IEEE, author of *Fundamentals of Pattern Recognition*, Prentice Hall, 1972, "Medical Science and Tomorrow's American" in *Tomorrow's American*, Oxford Press, 1976, and a contributor to *Frontiers of Pattern Recognition*. He is author of over 100 technical papers dealing with pattern recognition, medical diagnosis, computer techniques, communications, and heart-lung systems.

In January 1966, Dr. Patrick became an Assistant Professor of Electrical Engineering at Purdue University. In 1969, he had risen to Associate Professor of Electrical Engineering at Purdue and later an Associate Professor in the School of Medicine, Indiana University. In 1974, he was promoted to Full Professor at these institutions, a position currently held. In 1975, Dr. Patrick joined the staff of Jewish Hospital, University of Cincinnati, as a physician and Physician in Charge of Clinical Computing. At the same time, along with former astronaut Neil A. Armstrong, Henry J. Heimlich, M.D., and George Rieveschl, Jr., Sc.D., Ph.D., founded the Institute of Engineering and Medicine.

Dr. Patrick has been a consultant to numerous corporations, universities, and government agencies including the National Academy of Sciences, National Science Foundation, Dupont Corporation, Texas Instruments, Magnavox Corporation, Regenstrief Institute, Department of the Navy and Air Force, and the United States President's Office.

Dr. Patrick is an associate editor of *Computers in Biology and Medicine*. He served as consultant since 1975 for the establishment of a computerized monitoring and information system for the 24-bed coronary care unit of the Jewish Hospital. He served on the Emergency Medical Services Committee of the National Academy of Sciences which drafted the current protocol for diagnosis and treatment of foreign body airway obstruction and cardiac arrest. This protocol is taught to all physicians and other emergency medical members of the health care team.

At the 1978 Scientific Meeting of the American Medical Association, Dr. Patrick and his colleagues demonstrated the first videodisc player controlled by a microcomputer for medicine. With the American Medical Association, Dr. Patrick offered the first course to physicians for credit using the microcomputer-controlled videodisc player (CONSULT II®). This course teaches physicians' techniques for Positive End Expiratory Pressure and Cardiopulmonary Resuscitation.

Dr. Patrick is recognized as a discoverer of the solution to "learning without a teacher" (unsupervised estimation) in 1965 to 1967. This is a theoretical solution to the fundamental number of diseases in a particular problem.

In 1978, he established Patrick Consult Inc. in order to bring together research-orientated physicians, computer scientists, and medical researchers to further the development of CONSULT I® and CONSULT II®. A panel of respected medical consultants has agreed to review CONSULT I® and CONSULT II® programs.

TABLE OF CONTENTS

Chapter 1	
Introduction	1
Chapter 2	
Introduction to Probability and Statistics	3
Basic Notation	3
Sets	3
Set Operations	3
Events	3
Mutually Exclusive Sets	4
Probability Spaces	4
Conditional Probability	6
Total Probability	6
Bayes' Theorem	7
Independent Events	8
Random Variable, Distribution Function, and Probability Density Function	8
Multiple Random Variables (Multidimensional)	12
Correlation; Covariances	12
Vectors, Spaces, and Matrix Notation	14
Introduction	14
Definition	14
Chapter 3	
Decision Rules and Estimation	21
Introduction	21
Minimizing Decision Loss and Probability of Error	22
Gaussian Decision Rule	24
Third K-Nearest Neighbor Decision Rule	26
Adaptive Sample Set Construction Discussion	27
Estimating Class-Conditional Probability Density Functions	28
Introduction	28
Gaussian	28
Binomial	30
Multinomial	31
Chapter 4	
Diagnosis, Consulting, and Evaluation	33
Spaces, Classes, and Features	33
Systems and Subsystems	34
Consulting and Diagnosis	34
Joint vs. Conditional Probability Density Functions	38
Reality Behind Bayes' Theorem	38
Expected Decision Cost	39
Decision Making	39
Utility and Loss	41
Dimensionality Reduction	43
Independent Features with Discrete Values	52
Discrete vs. Continuous Feature Values	53
Dependence Tree Approximation — Discrete Feature Values	54
Any Differential Diagnosis as a Two-Class Problem	55
Feature Evaluation	57

Chapter 5	
Evaluating Diagnosis, Treatment, and Outcome	65
Expected Outcome Loss for Fixed Patient State	65
Introduction	65
Treatments and Outcomes	65
Clinical Example — Foreign Body Airway Obstruction vs. Heart Attack	68
Diagnosis, Treatment, and Outcome at Different Patient States	73
Introduction	73
Subclasses, Complex Class, Significant Feature Vector	74
States	75
The Patient's Path	76
Outcome	77
A Posteriori Outcome Probability	77
Generalized Expected Outcome Loss	79
Loss and Utility Vectors	80
Sequential Decision Making	80
Historical Perspective	82
 Chapter 6	
Automated Consulting and Diagnosis Systems	87
Introduction	87
Perspective	87
Patrick-Shen-Stelmack System	88
Subsystem Activation Rules	90
Basic Consulting	93
A Working Compatible Consulting System	96
Logic Consulting	109
System Design Considerations	114
Overview	114
CASNET System	120
Reasoning	121
 Chapter 7	
Applications	125
To Treat or Not to Treat	125
Example of Expected Outcome Loss — Pancreas Cancer	126
Coronary Artery Surgery — Expected Outcome Utility	129
Breast Cancer Screening, Diagnosis, Treatment, and Outcome	132
Patrick's Formulation	132
Eddy's Model for Evaluating Breast Cancer Screening	134
Ophthalmology — Special Case	137
Diabetes Management	138
Heart Attacks — Early Diagnosis Given Chest Pain	140
Chest Pain — Pneumonia vs. Heart Disease	154
Congenital Heart Disease	161
A Computer Operating System — The Help System	162
Survival after Multiple Trauma	166
Lower G.I. Tract Disorders	167
Acid-Base Disorders	169
Bleich's System	176
HEME — Hematologic Diseases	178
Present Illness History Taking — Pauker, Gorry, Kassirer, and Schwartz System	182

Surgery — Critical Care	183
CARE	183
Psychiatric Diseases	188
Acute Viral Hepatitis vs. Chronic Active Hepatitis	191
Acute Renal Failure	194
Thyroid Diseases	196
Drug Interactions	205
Genetics	205
Pedigree Analysis — Familial Diseases	205
Protocol for Diagnosis and Treatment of Foreign Body Airway Obstruction	209
Introduction	209
Decision Analysis Approach	209
Treatments	212
Utility of Brain Scanning	220
Mortality Loss and Dollar Cost — Differential Diagnosis of Chest Pain	222
Introduction	222
Upper and Lower Bounds	224
Emergency Medicine	224
Introduction	224
Medical Emergencies	225
General Automated Consulting and Diagnosis	228
Consulting in Neurology	230
Introduction	230
Subsystems in the Human Nervous System	231
Generic Measurements	232
Other Subsystems	233
Diagnosing Acute Pulmonary Embolism	236
Introduction	236
Subclasses of acute pulmonary embolism	241
Subclasses of no pulmonary embolism	244
Malabsorption Syndromes	245
Introduction	245
Classes and Measurements	245
Activation Rules	245
Hyperlipidemia	245
Introduction	245
Classes and Measurements	246
Differential Diagnosis of Thyroid Disease with Minimal Testing	246
Introduction	246
Medical Decision Analysis	249
Discussion	250
Hypercalcemia-Differential Diagnosis	250
Introduction	250
Classes and Features	251
Data Base Structure	251
Treatments	252
Conclusion	252
Fever and Unknown Origin	252
Introduction	252
Protocol Design	253
Discussion	253
Rheumatic Fever	254
Introduction	254

Acute Renal Failure	258
Introduction	258
Recognizing Stroke	263
Introduction	263
Recognizing Acute Pancreatitis	266
Introduction	266
Determining Classes, Subsystems, and Structured Recurring Problems.....	268
Differential Diagnosis of Jaundice (Increases Serum Bilirubin).....	269
Introduction	269
The Acute Abdomen	271
Introduction	271
Cerebrovascular Subsystem	278
Introduction	278
Shock Subsystem	281
Introduction	281
Expanded Model for Differential Diagnosis of Chest Pain.....	282
Introduction	282
Tendon Reflex Subsystem.....	285
Introduction	285
Hypersplenism Subsystem.....	287
Introduction	287
Collagen Disease Subsystem	287
 Appendix A.	
Model-Oriented Literature Review	293
Approaches.....	293
Characteristics Sought in Application Papers	293
Default Properties of Applications	294
Summary of Applications.....	294
Conclusion	309
 Appendix B.	
Historial Perspective	315
Approaches.....	315
Decision-making Involves a Decision Rule	316
Subsystems and Their Interconnection	317
Loss Functions	318
Data Structures	318
Index	325

Chapter 1

INTRODUCTION

This book is intended for those having a serious interest in the use of computers, statistics, and statistical pattern recognition to improve diagnosis, treatment, and management in patient care. The book is self-contained, including chapters entitled Introduction to Probability and Statistics and Decision Rules and Estimation. A student in engineering, mathematics, statistics, computer science, economics, and other fields of science should be able to comprehend the methods presented. Doctors with similar interests can make important contributions.

Experience has shown that with extra instruction on the material in the early chapters, the methods can be comprehended by many residents and practicing doctors. On the other hand, they are in a better position to appreciate the applications. Ideally, a team approach involving one or more research physicians and practitioners of this science can make the most progress.

Early books, such as *Fundamentals of Pattern Recognition* by Patrick,¹ are at such an advanced level as to be difficult to integrate by one not doing research in that area. Yet, the methods developed in statistical pattern recognition have often been overlooked by those developing solutions to problems in computer-assisted diagnosis. Other early books have been more concerned with medical examples involving credible but elementary methods from statistical pattern recognition. This is not meant as a general criticism because some of the problems have been solved, resulting in working systems.

A book by Lusted² discusses elementary properties of the a posteriori probability of a disease in a differential diagnosis. This work does not deal with the fundamental problems of estimating multidimensional probability density functions of diseases, ways to introduce a priori medical knowledge (such as through complex features), and other methods of dimensionality reduction. He does discuss the decision tree approach to decision making, which often is an approximation to the optimum approach. Lusted presents applications to Cushing's syndrome and congenital heart disease. He provides a reference list for heart diseases, ECG processing, Cushing's syndrome, thyroid disease, gastrointestinal disorders, bone tumors, neurology, hematology, screening, diabetes, urology, dermatology, lung cancer, and surgery.

A book edited by Jacquez³ deals extensively with a well-known but simple concept called the ROC (receiver operating characteristic), which is just the probability of a true positive as a function (plotted against) of the probability of a false positive. This is of interest to the radiologist in understanding how he sets his decision boundary or, more precisely, what losses he assigns to different kinds of errors. The general formulation, however, are the decision rules presented in this book.

Jacquez also discusses the decision tree, the a posteriori probability calculated under many assumptions, and some aspects of decision analysis. In that regard, a model is presented which involves diagnosis, treatment, and outcomes. Applications discussed are aching and weakness in the lower extremities, psychiatric disorders, hematologic diseases, chromosome analysis, serum immunoglobulins, and ECG analysis.

This book presents the methodology of statistical pattern recognition and decision analysis in the early chapters. Later, when over 25 applications are presented, the

reader can precisely identify the assumptions made by the authors of the applications presented. I must confess that it has always bothered me to hear someone proclaim that a particular application program for computer-assisted medical diagnosis “does not work” or that I as a physician should perform better. This usually is a statement made out of context. Has the program been designed to function to resolve a differential diagnosis when the critic is thinking about monitoring the patient through multiple states with time, while the program is written as a differential diagnosis at a fixed state? Often the author of an application program does not precisely define classes, features, subclasses, complex classes, or class-conditional probability density functions. Then, other researchers are not able to modify the model by updating it or adding to it.

It is hoped that this book will provide one more step toward the goal of using statistical pattern recognition or decision analysis and computers to improve health care delivery.

REFERENCES

1. Patrick, E. A., *Fundamentals of Pattern Recognition*, Prentice-Hall, Englewood Cliffs, N.J., 1972.
2. Lusted, L. B., *Introduction to Medical Decision Making*, Charles C Thomas, Springfield, Ill., 1968.
3. Jacquez, J. A., *Computer Diagnosis and Diagnostic Methods*, Charles C Thomas, Springfield, Ill., 1972.

Chapter 2

INTRODUCTION TO PROBABILITY AND STATISTICS

BASIC NOTATION

This chapter presents an introduction to probability and statistics which will be useful in later chapters. More depth is provided in the References.¹⁻⁶

Sets

A set is a collection of objects denoted $\xi_1, \xi_2, \dots, \xi_n$, and the set denoted

$$A = [\xi_1, \xi_2, \dots, \xi_n] \quad (2.1)$$

where $\{ \}$ means the "collection of" or "set of". Equation 2.1 is also written

$$A = [\xi_i]_{i=1}^n \quad (2.2)$$

where the objects in the set are accounted for by the index $i = 1, 2, \dots, n$. A particular element ξ_i in the set A can be denoted $\xi_i \in A$ (ξ_i is an element of A) or, if ξ_i is not in the set, denoted $\xi_i \notin A$ (ξ_i is not an element of A).

Examples: $\{1, 2, 3, 4\}$ = a set of real numbers; $\{\text{head, tail}\}$ = possible results upon flipping an unbiased coin; $\{\text{heart attack, food choking, fainting}\}$ = a set of causes of unconsciousness.

Set Operations

Subset: $B \subset A$ – B is a subset of A if every element of B also is an element of A .

Equality: $B = A$ if each element of B is an element of A ; also, each element of A is an element of B .

Sums: $A + B$ indicates a set whose elements are all the elements of A , of B , or of both.

Product: AB indicates a set of all elements that are common to set A and B .

Complement of a Set: The complement of A is a set \bar{A} containing all elements (or events) of S not in A .

Difference: The difference $A - B$ is the set consisting of the elements of A that are not in B . It follows that $A - B = A\bar{B}$.

Laws: The Associative Law of Addition is expressed as $(A + B) + C = A + (B + C)$; the Commutative Law of Addition by $A + B = B + A$; The Associative Law of Product by $A(B + C) = (AB) + (AC)$; and the Commutative Law of Product by $(AB)C = A(BC) = ABC$.

Events

The set of all possible events in an experiment or model is denoted S . Given one set of elements $\{\xi_i\}_{i=1}^n$, an event may be a single ξ_i . Given two sets of elements $\{\xi_i\}_{i=1}^n, \{\eta_j\}$

$r_{j=1}$, an event may be $\xi_1 + \eta_1$. Another event is $\xi_2 + \eta_2$. The totality of all events considered part of the experiment must be defined and is denoted S.

Example: A coin is tossed twice, and thus there are two sets of elements

$$[\xi_i]_{i=1}^2 = [\text{heads (h), tails (t)}]; \text{ first tossing} \tag{2.3}$$

$$[\xi_j]_{j=1}^2 = [\text{heads (h), tails (t)}]; \text{ second tossing} \tag{2.4}$$

The set S (set of all possible events) can be, $S = \{hh, ht, th, tt\}$ and the entries are events. Subsets of S can be formed:

- $A = \{\text{heads at the first tossing}\} = \{hh, ht\}$
- $B = \{\text{only one head came up}\} = \{ht, th\}$
- $C = \{\text{heads came up at least once}\} = \{hh, ht, th\}$

Mutually Exclusive Sets

Two sets A and B are called mutually exclusive or disjoint if they have no common elements; i.e., $AB = \phi$, where ϕ denotes the empty set, i.e., the set containing no objects.

M sets A_1, A_2, \dots, A_M are mutually exclusive if $A_i A_j = \phi$ for all combinations of i and j where $i \neq j$. This often is written

$$A_i A_j = \phi \forall i, j, i \neq j \tag{2.5}$$

where $\forall =$ for all, \neq means does not equal.

PROBABILITY SPACES

Experiment: The set S contains all possible events, and therefore, S is the certain event. The impossible event is denoted O. Redefine for convenience the members of S to be $\mu_1, \mu_2, \dots, \mu_M$; then

$$S = [\mu_i]_{i=1}^M \tag{2.6}$$

The single performance of a well-defined experiment produces a single outcome, μ_i .

Probability: In the experiment, there exists a number $P(\mu_i)$ called the probability of the event μ_i . For convenience, μ_i can be represented generally by A.

Axioms of Probability: A model called the axiomatic model of probability theory is characterized by the following three axioms, called the Axioms of Probability:

$$P(A) \geq 0 \tag{2.7}$$

$$P(S) = 1 \tag{2.8}$$

$$\text{If } AB = \phi, \text{ then } P(A + B) = P(A) + P(B) \tag{2.9}$$

where A and B are any sets of events in S. How are probabilities in this model determined? The basic method is through application of symmetry. For example, if

the experiment is tossing an unbiased coin, $S = \{h, t\}$; then by symmetry $P(h) = \frac{1}{2}$, $P(t) = \frac{1}{2}$. From Equations 2.8 and 2.9

$$P(h+t) = P(h) + P(t) = \frac{1}{2} + \frac{1}{2} = 1 \quad (2.10)$$

because h and t are mutually exclusive events.

A corollary to the three Axioms of Probability is that, in general (if $AB \neq \phi$),

$$P(A+B) = P(A) + P(B) - P(AB) \quad (2.11)$$

Relative Frequency: In the real world, how does one estimate the probabilities of events in S ? Symmetry, useful for determining probabilities of events when tossing dice or playing cards, may not apply.

Example: Suppose a patient presents in the emergency room with chest pain, and one of these three possible events has occurred:

- μ_1 — Myocardial infarction (MI)
- μ_2 — Coronary insufficiency (CI)
- μ_3 — Chest pain, noncardiac cause (CP)

Thus, $S = \{MI, CI, CP\}$. We don't know if an assumption such as $P(MI) = P(CI) = P(CP)$ is valid. A solution is another approach called relative frequency definition of probability, otherwise known as estimation. We identify the μ_i (there is only one) for each of n patients presenting with chest pain. Three estimators then are defined:

$$\hat{P}(\mu_1) = \hat{P}(MI) = \frac{n_1}{n} \quad (2.12)$$

$$\hat{P}(\mu_2) = \hat{P}(CI) = \frac{n_2}{n} \quad (2.13)$$

$$\hat{P}(\mu_3) = \hat{P}(CP) = \frac{n_3}{n} \quad (2.14)$$

where n_1 patients have MI, n_2 patients have CI, and n_3 patient have CP and $n = n_1 + n_2 + n_3$. The experiment has been repeated n times, an example of the relative frequency approach to probability.

If an experiment is repeated n times and the event A occurs n_a times, then

$$\hat{P}(A) = \frac{n_a}{n} \quad (2.15)$$

Additivity of Probabilities of Disjoint Events: if $A_i A_j = \phi$, $i \neq j$, $\forall i, j$, then

$$\begin{aligned} P(A_1 + A_2 + \dots + A_M) &= P(A_1) + P(A_2) + \dots + P(A_M) \\ &= \sum_{i=1}^M P(A_i) \end{aligned} \quad (2.16)$$

Construction of the Probability Space: As always, $S = \{\mu_1, \mu_2, \dots, \mu_M\}$. Define

$$P[\mu_i] = P_i \quad (2.17)$$

elementary probabilities. Let A be an arbitrary subset of the elements in S , say the k th possible subset and denote its elements $\mu_{k1}, \mu_{k2}, \dots, \mu_{kr}$. These are called the elementary events:

$$A = [\mu_{k_1}, \mu_{k_2}, \dots, \mu_{k_r}] \quad (2.18)$$

Then, from Equations 2.16 and 2.17,

$$P(A) = P_{k_1} + P_{k_2} + \dots + P_{k_r} \quad (2.19)$$

Thus, the probability of any event A (Equations 2.18 and 2.19) can be determined in terms of the elementary probabilities, i.e., the probabilities of the elementary events.

CONDITIONAL PROBABILITY

Definition: Suppose that a set S of elementary events has been defined as a probability space. Let X be any subset of S and ω another subset of S . The probability of X given ω is denoted $P(X|\omega)$; thus,

$$P(X|\omega) = \frac{P(X, \omega)}{P(\omega)} \quad (2.20)$$

is also called the conditional probability of X . The ratio on the right of Equation 2.20 may be viewed as the probability of the part of X included in ω divided by the probability of ω .

Relative Frequency Interpretation: If the experiment is performed n times and X occurs n_x times; ω occurs n_ω times; and (X, ω) occurs $n_{x\omega}$ times; then

$$\hat{P}(X) = n_x/n \quad (2.21)$$

$$\hat{P}(\omega) = n_\omega/n \quad (2.22)$$

$$\hat{P}(X, \omega) = n_{x\omega}/n \quad (2.23)$$

and thus, as expected from Equation 2.20

$$\hat{P}(X|\omega) = \frac{(n_{x\omega}/n)}{(n_\omega/n)} = \frac{n_{x\omega}}{n_\omega} \quad (2.24)$$

TOTAL PROBABILITY

Given M mutually exclusive subsets $\omega_1, \omega_2, \dots, \omega_M$ which are also exhaustive subsets ($S = \omega_1 + \omega_2 + \dots + \omega_M$), then with X as any subset of S (an arbitrary event)

$$X = (X, \omega_1) + (X, \omega_2) + \dots + (X, \omega_M) \quad (2.25)$$

Furthermore, because $X\omega_1, X\omega_2, \dots, X\omega_M$ are mutually exclusive*

* $X\omega_1$ and (X, ω_1) are different notations for the product on intersection of two sets.

$$P(X) = P(X, \omega_1) + P(X, \omega_2) + \dots + P(X, \omega_M) \quad (2.26)$$

which is called the total probability. Then, using Equation 2.20

$$P(X) = \sum_{i=1}^M P(X|\omega_i) P_i \quad (2.27)$$

$$P_1 + P_2 + \dots + P_M = 1 \quad (2.28)$$

where $P_i = P(\omega_i)$. Equations 2.27 and 2.28 illustrate also a mixture probability where the P_i are called mixing parameters. An example of Equations 2.27 and 2.28 arises in medicine when there is a differential diagnosis consisting of M diseases, and X is an observed subset of signs, symptoms, or laboratory tests for the patient. This concept of M diseases and an observed subset X is basic to decision analysis in medicine and occurs in many places in this book.

BAYES' THEOREM

Introduction: Let X and ω be as defined above under Conditional Probability. Using the fact that $P(X, \omega) = P(\omega, X)$ and applying Equation 2.20 to each side of the above equality, we obtain

$$P(X|\omega) P(\omega) = P(\omega|X) P(X) \quad (2.29)$$

or

$$P(\omega|X) = \frac{P(X|\omega) P(\omega)}{P(X)} \quad (2.30)$$

which is Bayes' Theorem. Recall that ω is a subset of S . Equation 2.30, Bayes' Theorem, gives the probability of the subset ω , given the subset X .

To illustrate how Bayes' Theorem applies to medical diagnosis, let ω be one of the diseases in a differential diagnosis. We see above in Total Probability that (X, ω) is a subset of X , where X is a set of signs, symptoms, and laboratory tests. That is, (X, ω) is a subset of the signs, symptoms, and laboratory tests which can occur for patients with disease ω . Returning to Equation 2.30, the term $P(\omega)$ is the probability of the subset ω (without knowing the subset X). Given the subset X , a new conditional probability $p(\omega|X)$ is defined and calculated according to Equation 2.30. $P(\omega)$ is called the a priori probability. Given the subset X , a new probability, $P(\omega|X)$ called the a posteriori probability, is calculated. The probability $p(X|\omega)$ is called a class-conditional probability.

Bayes' Theorem for a Mixture: In general, there are M mutually exclusive and exhaustive subsets $\omega_1, \omega_2, \dots, \omega_M$ (as under Total Probability). Thus, Equations 2.29 and 2.30 can be applied for any subset ω_i :

$$P(\omega_i|X) = \frac{P(X|\omega_i) P(\omega_i)}{P(X)} \quad (2.31)$$

Substituting the mixture (or equations for total probability) Equations 2.27 and 2.28 for $P(X)$ into Equation 2.31 gives