Detection, Estimation, and Modulation Theory, Part I:
Detection, Estimation, and Linear Modulation Theory. Harry L. Van Trees
Copyright © 2001 John Wiley & Sons, Inc.
ISBNs: 0-471-09517-6 (Paperback); 0-471-22108-2 (Electronic)

Detection, Estimation, and Modulation Theory

Digital Global Systems
Exhibit 1012

Detection, Estimation, and Modulation Theory

Part I. Detection, Estimation, and Linear Modulation Theory

HARRY L. VAN TREES
George Mason University

A Wiley-Interscience Publication
JOHN WILEY & SONS, INC.
New York • Chichester • Weinheim • Brisbane • Singapore • Toronto

This text is printed on acid-free paper. ♾

Copyright © 2001 by John Wiley & Sons, Inc. All rights reserved.

Published simultaneously in Canada.

No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, scanning or otherwise, except as permitted under Section 107 or 108 of the 1976 United States Copyright Act, without either the prior written permission of the Publisher, or authorization through payment of the appropriate per-copy fee to the Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923, (978) 750-8400, fax (978) 750-4744. Requests to the Publisher for permission should be addressed to the Permissions Department, John Wiley & Sons, Inc., 605 Third Avenue, New York, NY 10158-0012, (212) 850-6011, fax (212) 850-6008, E-Mail: PERMREQ @ WILEY.COM.

For ordering and customer service, call 1-800-CALL-WILEY.

Library of Congress Cataloging in Publication Data is available.

ISBN 0-471-22108-2
This title is also available in print as **ISBN 0-471-09517-6**.

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

To Diane

and Stephen, Mark, Kathleen, Patricia,
Eileen, Harry, and Julia

and the next generation—
Brittany, Erin, Thomas, Elizabeth, Emily,
Dillon, Bryan, Julia, Robert, Margaret,
Peter, Emma, Sarah, Harry, Rebecca,
and Molly

Preface for Paperback Edition

In 1968, Part I of *Detection, Estimation, and Modulation Theory* [VT68] was published. It turned out to be a reasonably successful book that has been widely used by several generations of engineers. There were thirty printings, but the last printing was in 1996. Volumes II and III ([VT71a], [VT71b]) were published in 1971 and focused on specific application areas such as analog modulation, Gaussian signals and noise, and the radar–sonar problem. Volume II had a short life span due to the shift from analog modulation to digital modulation. Volume III is still widely used as a reference and as a supplementary text. In a moment of youthful optimism, I indicated in the the Preface to Volume III and in Chapter III-14 that a short monograph on optimum array processing would be published in 1971. The bibliography lists it as a reference, *Optimum Array Processing,* Wiley, 1971, which has been subsequently cited by several authors. After a 30-year delay, *Optimum Array Processing,* Part IV of *Detection, Estimation, and Modulation Theory* will be published this year.

A few comments on my career may help explain the long delay. In 1972, MIT loaned me to the Defense Communication Agency in Washington, D.C. where I spent three years as the Chief Scientist and the Associate Director of Technology. At the end of the tour, I decided, for personal reasons, to stay in the Washington, D.C. area. I spent three years as an Assistant Vice-President at COMSAT where my group did the advanced planning for the INTELSAT satellites. In 1978, I became the Chief Scientist of the United States Air Force. In 1979, Dr. Gerald Dinneen, the former Director of Lincoln Laboratories, was serving as Assistant Secretary of Defense for C3I. He asked me to become his Principal Deputy and I spent two years in that position. In 1981, I joined M/A-COM Linkabit. Linkabit is the company that Irwin Jacobs and Andrew Viterbi had started in 1969 and sold to M/A-COM in 1979. I started an Eastern operation which grew to about 200 people in three years. After Irwin and Andy left M/A-COM and started Qualcomm, I was responsible for the government operations in San Diego as well as Washington, D.C. In 1988, M/A-COM sold the division. At that point I decided to return to the academic world.

I joined George Mason University in September of 1988. One of my priorities was to finish the book on optimum array processing. However, I found that I needed to build up a research center in order to attract young research-oriented faculty and

doctoral students. The process took about six years. The Center for Excellence in Command, Control, Communications, and Intelligence has been very successful and has generated over $300 million in research funding during its existence. During this growth period, I spent some time on array processing but a concentrated effort was not possible. In 1995, I started a serious effort to write the Array Processing book.

Throughout the *Optimum Array Processing* text there are references to Parts I and III of *Detection, Estimation, and Modulation Theory*. The referenced material is available in several other books, but I am most familiar with my own work. Wiley agreed to publish Part I and III in paperback so the material will be readily available. In addition to providing background for Part IV, Part I is still useful as a text for a graduate course in Detection and Estimation Theory. Part III is suitable for a second level graduate course dealing with more specialized topics.

In the 30-year period, there has been a dramatic change in the signal processing area. Advances in computational capability have allowed the implementation of complex algorithms that were only of theoretical interest in the past. In many applications, algorithms can be implemented that reach the theoretical bounds.

The advances in computational capability have also changed how the material is taught. In Parts I and III, there is an emphasis on compact analytical solutions to problems. In Part IV, there is a much greater emphasis on efficient iterative solutions and simulations. All of the material in parts I and III is still relevant. The books use continuous time processes but the transition to discrete time processes is straightforward. Integrals that were difficult to do analytically can be done easily in Matlab®. The various detection and estimation algorithms can be simulated and their performance compared to the theoretical bounds. We still use most of the problems in the text but supplement them with problems that require Matlab® solutions.

We hope that a new generation of students and readers find these reprinted editions to be useful.

HARRY L. VAN TREES

Fairfax, Virginia
June 2001

Preface

The area of detection and estimation theory that we shall study in this book represents a combination of the classical techniques of statistical inference and the random process characterization of communication, radar, sonar, and other modern data processing systems. The two major areas of statistical inference are decision theory and estimation theory. In the first case we observe an output that has a random character and decide which of two possible causes produced it. This type of problem was studied in the middle of the eighteenth century by Thomas Bayes [1]. In the estimation theory case the output is related to the value of some parameter of interest, and we try to estimate the value of this parameter. Work in this area was published by Legendre [2] and Gauss [3] in the early nineteenth century. Significant contributions to the classical theory that we use as background were developed by Fisher [4] and Neyman and Pearson [5] more than 30 years ago. In 1941 and 1942 Kolmogoroff [6] and Wiener [7] applied statistical techniques to the solution of the optimum linear filtering problem. Since that time the application of statistical techniques to the synthesis and analysis of all types of systems has grown rapidly. The application of these techniques and the resulting implications are the subject of this book.

This book and the subsequent volume, Detection, Estimation, and Modulation Theory, Part II, are based on notes prepared for a course entitled "Detection, Estimation, and Modulation Theory," which is taught as a second-level graduate course at M.I.T. My original interest in the material grew out of my research activities in the area of analog modulation theory. A preliminary version of the material that deals with modulation theory was used as a text for a summer course presented at M.I.T. in 1964. It turned out that our viewpoint on modulation theory could best be understood by an audience with a clear understanding of modern detection and estimation theory. At that time there was no suitable text available to cover the material of interest and emphasize the points that I felt were

important, so I started writing notes. It was clear that in order to present the material to graduate students in a reasonable amount of time it would be necessary to develop a unified presentation of the three topics: detection, estimation, and modulation theory, and exploit the fundamental ideas that connected them. As the development proceeded, it grew in size until the material that was originally intended to be background for modulation theory occupies the entire contents of this book. The original material on modulation theory starts at the beginning of the second book. Collectively, the two books provide a unified coverage of the three topics and their application to many important physical problems.

For the last three years I have presented successively revised versions of the material in my course. The audience consists typically of 40 to 50 students who have completed a graduate course in random processes which covered most of the material in Davenport and Root [8]. In general, they have a good understanding of random process theory and a fair amount of practice with the routine manipulation required to solve problems. In addition, many of them are interested in doing research in this general area or closely related areas. This interest provides a great deal of motivation which I exploit by requiring them to develop many of the important ideas as problems. It is for this audience that the book is primarily intended. The appendix contains a detailed outline of the course.

On the other hand, many practicing engineers deal with systems that have been or should have been designed and analyzed with the techniques developed in this book. I have attempted to make the book useful to them. An earlier version was used successfully as a text for an in-plant course for graduate engineers.

From the standpoint of specific background little advanced material is required. A knowledge of elementary probability theory and second moment characterization of random processes is assumed. Some familiarity with matrix theory and linear algebra is helpful but certainly not necessary. The level of mathematical rigor is low, although in most sections the results could be rigorously proved by simply being more careful in our derivations. We have adopted this approach in order not to obscure the important ideas with a lot of detail and to make the material readable for the kind of engineering audience that will find it useful. Fortunately, in almost all cases we can verify that our answers are intuitively logical. It is worthwhile to observe that this ability to check our answers intuitively would be necessary even if our derivations were rigorous, because our ultimate objective is to obtain an answer that corresponds to some physical system of interest. It is easy to find physical problems in which a plausible mathematical model and correct mathematics lead to an unrealistic answer for the original problem.

We have several idiosyncrasies that it might be appropriate to mention. In general, we look at a problem in a fair amount of detail. Many times we look at the same problem in several different ways in order to gain a better understanding of the meaning of the result. Teaching students a number of ways of doing things helps them to be more flexible in their approach to new problems. A second feature is the necessity for the reader to solve problems to understand the material fully. Throughout the course and the book we emphasize the development of an ability to work problems. At the end of each chapter are problems that range from routine manipulations to significant extensions of the material in the text. In many cases they are equivalent to journal articles currently being published. Only by working a fair number of them is it possible to appreciate the significance and generality of the results. Solutions for an individual problem will be supplied on request, and a book containing solutions to about one third of the problems is available to faculty members teaching the course. We are continually generating new problems in conjunction with the course and will send them to anyone who is using the book as a course text. A third issue is the abundance of block diagrams, outlines, and pictures. The diagrams are included because most engineers (including myself) are more at home with these items than with the corresponding equations.

One problem always encountered is the amount of notation needed to cover the large range of subjects. We have tried to choose the notation in a logical manner and to make it mnemonic. All the notation is summarized in the glossary at the end of the book. We have tried to make our list of references as complete as possible and to acknowledge any ideas due to other people.

A number of people have contributed in many ways and it is a pleasure to acknowledge them. Professors W. B. Davenport and W. M. Siebert have provided continual encouragement and technical comments on the various chapters. Professors Estil Hoversten and Donald Snyder of the M.I.T. faculty and Lewis Collins, Arthur Baggeroer, and Michael Austin, three of my doctoral students, have carefully read and criticized the various chapters. Their suggestions have improved the manuscript appreciably. In addition, Baggeroer and Collins contributed a number of the problems in the various chapters and Baggeroer did the programming necessary for many of the graphical results. Lt. David Wright read and criticized Chapter 2. L. A. Frasco and H. D. Goldfein, two of my teaching assistants, worked all of the problems in the book. Dr. Howard Yudkin of Lincoln Laboratory read the entire manuscript and offered a number of important criticisms. In addition, various graduate students taking the course have made suggestions which have been incorporated. Most of the final draft was typed by Miss Aina Sils. Her patience with the innumerable changes is

sincerely appreciated. Several other secretaries, including Mrs. Jarmila Hrbek, Mrs. Joan Bauer, and Miss Camille Tortorici, typed sections of the various drafts.

As pointed out earlier, the books are an outgrowth of my research interests. This research is a continuing effort, and I shall be glad to send our current work to people working in this area on a regular reciprocal basis. My early work in modulation theory was supported by Lincoln Laboratory as a summer employee and consultant in groups directed by Dr. Herbert Sherman and Dr. Barney Reiffen. My research at M.I.T. was partly supported by the Joint Services and the National Aeronautics and Space Administration under the auspices of the Research Laboratory of Electronics. This support is gratefully acknowledged.

Harry L. Van Trees

Cambridge, Massachusetts
October, 1967.

REFERENCES

[1] Thomas Bayes, "An Essay Towards Solving a Problem in the Doctrine of Chances," *Phil. Trans*, **53**, 370–418 (1764).
[2] A. M. Legendre, *Nouvelles Méthodes pour La Détermination ces Orbites des Comètes*, Paris, 1806.
[3] K. F. Gauss, *Theory of Motion of the Heavenly Bodies Moving About the Sun in Conic Sections*, reprinted by Dover, New York, 1963.
[4] R. A. Fisher, "Theory of Statistical Estimation," *Proc. Cambridge Philos. Soc.*, **22**, 700 (1925).
[5] J. Neyman and E. S. Pearson, "On the Problem of the Most Efficient Tests of Statistical Hypotheses," *Phil. Trans. Roy. Soc. London*, A **231**, 289, (1933).
[6] A. Kolmogoroff, "Interpolation and Extrapolation von Stationären Zufälligen Folgen," *Bull. Acad. Sci. USSR, Ser. Math.* **5**, 1941.
[7] N. Wiener, *Extrapolation, Interpolation, and Smoothing of Stationary Time Series*, Tech. Press of M.I.T. and Wiley, New York, 1949 (originally published as a classified report in 1942).
[8] W. B. Davenport and W. L. Root, *Random Signals and Noise*, McGraw-Hill, New York, 1958.

Contents

1 Introduction 1

1.1 Topical Outline 1
1.2 Possible Approaches 12
1.3 Organization 15

2 Classical Detection and Estimation Theory 19

2.1 Introduction 19
2.2 Simple Binary Hypothesis Tests 23
Decision Criteria. Performance: Receiver Operating Characteristic.
2.3 M Hypotheses 46
2.4 Estimation Theory 52
Random Parameters: Bayes Estimation. Real (Nonrandom) Parameter Estimation. Multiple Parameter Estimation. Summary of Estimation Theory.
2.5 Composite Hypotheses 86
2.6 The General Gaussian Problem 96
Equal Covariance Matrices. Equal Mean Vectors. Summary.
2.7 Performance Bounds and Approximations 116
2.8 Summary 133
2.9 Problems 133

References 163

3 *Representations of Random Processes* 166

3.1 Introduction 166

3.2 Deterministic Functions: Orthogonal Representations 169

3.3 Random Process Characterization 174

Random Processes: Conventional Characterizations. Series Representation of Sample Functions of Random Processes. Gaussian Processes.

3.4 Homogeneous Integral Equations and Eigenfunctions 186

Rational Spectra. Bandlimited Spectra. Nonstationary Processes. White Noise Processes. The Optimum Linear Filter. Properties of Eigenfunctions and Eigenvalues.

3.5 Periodic Processes 209

3.6 Infinite Time Interval: Spectral Decomposition 212

Spectral Decomposition. An Application of Spectral Decomposition: MAP Estimation of a Gaussian Process.

3.7 Vector Random Processes 220

3.8 Summary 224

3.9 Problems 226

References 237

4 *Detection of Signals–Estimation of Signal Parameters* 239

4.1 Introduction 239

Models. Format.

4.2 Detection and Estimation in White Gaussian Noise 246

Detection of Signals in Additive White Gaussian Noise. Linear Estimation. Nonlinear Estimation. Summary: Known Signals in White Gaussian Noise.

4.3 Detection and Estimation in Nonwhite Gaussian Noise 287

"Whitening" Approach. A Direct Derivation Using the Karhunen-Loeve Expansion. A Direct Derivation with a Sufficient Statistic. Detection Performance. Estimation. Solution Techniques for Integral Equations. Sensitivity. Known Linear Channels.

4.4 Signals with Unwanted Parameters: The Composite Hypothesis Problem 333

Random Phase Angles. Random Amplitude and Phase.

4.5 Multiple Channels 366

Formulation. Application.

4.6 Multiple Parameter Estimation 370

Additive White Gaussian Noise Channel. Extensions.

4.7 Summary and Omissions 374

Summary. Topics Omitted.

4.8 Problems 377

References 418

5 *Estimation of Continuous Waveforms* 423

5.1 Introduction 423

5.2 Derivation of Estimator Equations 426

No-Memory Modulation Systems. Modulation Systems with Memory.

5.3 A Lower Bound on the Mean-Square Estimation Error 437

5.4 Multidimensional Waveform Estimation 446

Examples of Multidimensional Problems. Problem Formulation. Derivation of Estimator Equations. Lower Bound on the Error Matrix. Colored Noise Estimation.

5.5 Nonrandom Waveform Estimation 456

5.6 Summary 459

5.7 Problems 460

References 465

6 *Linear Estimation* 467

6.1 Properties of Optimum Processors 468

6.2 Realizable Linear Filters: Stationary Processes, Infinite Past: Wiener Filters 481

Solution of Wiener-Hopf Equation. Errors in Optimum Systems. Unrealizable Filters. Closed-Form Error Expressions. Optimum Feedback Systems. Comments.

6.3 Kalman-Bucy Filters 515

Differential Equation Representation of Linear Systems and Random Process Generation. Derivation of Estimator Equations. Applications. Generalizations.

6.4 Linear Modulation: Communications Context 575

DSB-AM: Realizable Demodulation. DSB-AM: Demodulation with Delay. Amplitude Modulation: Generalized Carriers. Amplitude Modulation: Single-Sideband Suppressed-Carrier.

6.5 The Fundamental Role of the Optimum Linear Filter 584

6.6 Comments 585

6.7 Problems 586

References 619

7 *Discussion* 623

7.1 Summary 623

7.2 Preview of Part II 625

7.3 Unexplored Issues 627

References 629

Appendix: A Typical Course Outline 635

Glossary 671

Author Index 683

Subject Index 687

Detection, Estimation, and Modulation Theory, Part I:
Detection, Estimation, and Linear Modulation Theory. Harry L. Van Trees
Copyright © 2001 John Wiley & Sons, Inc.
ISBNs: 0-471-09517-6 (Paperback); 0-471-22108-2 (Electronic)

Subject Index

Absolute error cost function, 54
Active sonar systems, 627
Adaptive and learning systems, 628
Additive Gaussian noise, 247, 250, 378, 387
Alarm, probability of false alarm, 31
AM, conventional DSB, with a residual carrier component, 424
Ambiguity, 627
Ambiguity function, radar, 627
AM-DSB, demodulation with delay, 578
 realizable demodulation, 576
AM/FM, 448
Amplitude, random, 349, 401
Amplitude estimation, random phase channel, 399
Amplitude modulation, double sideband, suppressed carrier, 424
 single sideband, suppressed carrier, 576, 581
Analog computer realization, 517, 519, 522–525, 534, 599
Analog message, transmission of an, 423, 626
Analog modulation system, 423
Angle modulation, pre-emphasized, 424
Angle modulation systems, optimum, 626
Angular prolate spheroidal functions, 193
Apertures, continuous receiving, 627
A posteriori estimate, maximum, 63
Applications, state variables, 546
Approach, continuous Gaussian processes, sampling, 231
 derivation of estimator equations using a state variable approach, 538
Approach, Markov process-differential equation, 629
 nonstructured, 12
 structured, 12
 whitening, 290
Approximate error expressions, 39, 116, 124, 125, 264
Approximation error, mean-square, 170
Arrays, 449, 464
Arrival time estimation, 276
ASK, incoherent channel, 399
 known channel, 379
Assumption, Gaussian, 471
Astronomy, radio, 9
Asymptotic behavior of incoherent M-ary system, 400
Asymptotically efficient, 71, 445
Asymptotically efficient estimates, 276
Asymptotically Gaussian, 71
Asymptotic properties, 71
 of eigenfunctions and eigenvalues, 205
Augmented eigenfunctions, 181
Augmented state-vector, 568

Bandlimited spectra, 192
Bandpass process, estimation of the center frequency of, 626
 representation, 227
Bandwidth, constraint, 282
 equivalent rectangular, 491
 noise bandwidth, definition, 226
Barankin bound, 71, 147, 286
Bayes, criterion, M hypotheses, 46
 two hypotheses, 24

Bayes, estimation, problems, 141
estimation of random parameters, 54
point estimator, 477
risk, 24
test, M hypotheses, 139
tests, 24
Beam patterns, 627
Bessel function of the first kind, modified, 338, 340
Bhattacharyya, bound, 71, 148, 284, 386
distance, 127
Bias, known, 64
unknown, 64
vector, 76
Biased estimates, Cramér-Rao bound, 146
Binary, communication, partially coherent, 345
detection, simple binary, 247
FSK, 379
hypothesis tests, 23, 134
nonorthogonal signals, random phase channel, error probability, 399
orthogonal signals, N Rayleigh channels, 415
orthogonal signals, square-law receiver, Rayleigh channel, 402
Binomial distribution, 145
Bi-orthogonal signals, 384
Biphase modulator, 240
Bit error probability, 384
Block diagram, MAP estimate, 431, 432
Boltzmann's constant, 240
Bound, Barankin, 71, 147, 286
Bhattacharyya, 71, 148, 284, 386
Chernoff, 121
Cramér-Rao, 66, 72, 79, 84, 275
ellipse, 81
$\mathrm{erfc}_*(X)$, 39, 138
error, optimal diversity, Rayleigh channel, 415
estimation errors, multiple nonrandom variables, 79
estimation errors, multiple random parameters, 84
intercept, 82
lower bound on, error matrix, vector waveform estimation, 453
mean-square estimation error, waveform estimation, 437
minimum mean-square estimate, random parameter, 72
Bound, matrix, 372
performance bounds, 116, 162
perfect measurement, 88
P_F, P_M, $\Pr(\epsilon)$, 122, 123
Brownian motion, 194
Butterworth spectra, 191, 502, 548–555

Canonical, feedback realization of the optimum filter, 510
realization, Number one (state variables), 522
realization, Number two (state variables), 524
realization, Number three (state variables), 525
receiver, linear modulation, rational spectrum, 556
Capacity, channel, 267
Carrier synchronization, 626
Cauchy distribution, 146
Channel, capacity, 267
kernels, 333
known linear, 331
measurement, 352, 358
measurement receivers, 405
multiple (vector), 366, 408, 537
partially coherent, 397, 413
randomly time-varying, 626
random phase, 397, 411
Rayleigh, 349-359, 414, 415
Rician, 360–364, 402, 416
Characteristic, receiver operating, ROC, 36
Characteristic function of Gaussian process, 185
Characteristic function of Gaussian vector, 96
Characterization, complete, 174
conventional, random process, 174
frequency-domain, 167
partial, 175
random process, 174, 226
second moment, 176, 226
single time, 176
time-domain, 167
Chernoff bound, 121
Chi-square density, definition, 109
Classical, estimation theory, 52
optimum diversity, 111
parameter estimation, 52
Classical theory, summary, 133
Closed form error expressions, colored noise, 505

Closed form error expressions, linear operations, 505
problems, 593
white noise, 498–505
Coded digital communication systems, 627
Coefficient, correlation (of deterministic signals), 173
Coefficients, experimental generation of, 170
Coherent channels, M-orthogonal signals, partially coherent channels, 397
N partially coherent channels, 413
on-off signaling, partially coherent channels, 397
Colored noise, closed form error expressions, 505
correlator realization, 293
detection performance, 301
estimation in the presence of, 307, 562
estimation in the presence of colored noise only, 572, 611
estimation of colored noise, 454
M-ary signals, 389
problems, 387
receiver derivation using the Karhunen-Loeve expansion, 297
receiver derivation using "whitening," 290
receiver derivation with a sufficient statistic, 299
sensitivity in the presence of, 326
singularity, 303
whitening realization, 293
Combiner, maximal ratio, 369
Comments, Wiener filtering, 511
Communication, methods employed to reduce intersymbol interference in digital, 160
partially coherent binary, 345
scatter, 627
Communication systems, analog, 9, 423
digital, 1, 160, 239, 627
Commutation, of efficiency, 84
maximum a posteriori interval estimation and linear filtering, 437
minimum mean-square estimation, 75
Complement, error function, 37
Complete characterization, 174
Complete orthonormal (CON), 171
Complex envelopes, 399, 626
Composite hypotheses, classical, 86
problems, classical, 151
Composite hypotheses, problems, signals, 394
signals, 333
Concavity, ROC, 44
Concentration, ellipses, 79, 81
ellipsoids, 79
Conditional mean, 56
Conjugate prior density, 142
Consistent estimates, 71
Constraint, bandwidth, 282
threshold, 281
Construction of $Q_n(t, u)$ and $g(t)$, 294
Continuous, Gaussian processes, sampling approach to, 231
messages, MAP equations, 431
receiving apertures, 627
waveform, estimation of, 11, 423, 459
Conventional, characterizations of random processes, 174
DSB-AM with a residual carrier component, 424
limiter-discriminators, 626
pulsed radar, 241
Convergence, mean-square, 179
uniform, 181
Convex cost function, 60, 144, 477, 478
Convolutional encoder, 618
Coordinate system transformations, 102
Correlated Gaussian random variables, quadratic form of, 396
Correlated signals, M equally, 267
Correlation, between $\mathbf{u}(t)$ and $\mathbf{w}(t)$, 570
coefficient, 173
error matrix, 84
operation, 172
receiver, 249
-stationary, 177
Correlator, estimator-correlator realization, 626
realization, colored noise, 293
-squarer receiver, 353
Cosine sufficient statistic, 337
Cost function, 24
absolute error, 54
nondecreasing, 61
square-error, 54
strictly convex, 478
symmetric and convex, 60, 144, 477
uniform, 54
Covariance, function, properties, 176
-stationary, 177
Cramér-Rao inequality, 74

Cramér-Rao inequality, biased estimates, 146
extension to waveform estimation, 437
multiple parameters, 79, 84
random parameters, 72
unbiased, nonrandom parameters, 66
waveform observations, 275
Criterion, Bayes, 24, 46
decision, 23
maximum signal to noise ratio, 13
Neyman-Pearson, 24, 33
Paley-Wiener, 512

$\mathbf{d}^2$, 69, 99
Data information matrix, 84
Decision, criterion, 21, 23
dimensionality of decision space, 34
one-dimensional decision space, 250
rules, randomized decision rules, 43
Definite, nonnegative (definition), 177
positive (definition), 177
Definition of, chi-square density, 109
d^2, 99
functions, error, 37
incomplete gamma, 110
moment-generating, 118
general Gaussian problem, 97
Gaussian processes, 183
Gaussian random vector, 96
$\mathbf{H}$ matrix, 108
information kernel, 441
inverse kernel, 294
$\mathbf{J}$ matrix, 80
jointly Gaussian random variables, 96
linear modulation, 427, 467
$\mu(s)$, 118
noise bandwidth, 226
nonlinear modulation, 427
periodic process, 209
probability of error, 37
state of the system, 517
sufficient statistic, 35
Delay, DSB-AM, demodulation with delay, 578
effect of delay on the mean-square error, 494
filtering with delay (state variables), 567
sensitivity to, 393
tapped delay line, 161
Demodulation, DSB-AM, realizable with delay, 578
Demodulation, DSB-AM realizable, 576
optimum nonlinear, 626
Density(ies), conjugate prior, 142
joint probability, Gaussian processes, 185
Markov processes, 228
Rician envelope and phase, 413
probability density, Cauchy, 146
chi-square, 109
phase angle (family), 338
Rayleigh, 349
Rician, 413
reproducing, 142
tilted, 119
Derivation, colored noise receiver, known signals, Karhunen-Loève approach, 297
sufficient statistic approach, 299
whitening approach, 290
Kalman-Bucy filters, 538
MAP equations for continuous waveform estimation, 426
optimum linear filter, 198, 472
receiver, signals with unwanted parameters, 334
white noise receiver, known signals, 247
Derivative, of signal with respect to message, 427
partial derivative matrix, $\nabla_{\mathbf{x}}$, 76, 150
Design, optimum signal, 302
Detection, classical, 19
colored noise, 287
general binary, 254
hierarchy, 5
M-ary, 257
models for signal detection, 239
in multiple channels, 366
performance, 36, 249
in presence of interfering target, 324
probability of, 31
of random processes, 585, 626
sequential, 627
simple binary, 247
Deterministic function, orthogonal representations, 169
Differential equation, first-order vector, 520
Markov process-differential equation approach, 629
representation of linear systems, 516
time-varying, 527, 531, 603
Differentiation of a quadratic form, 150
Digital communication systems, 1, 160, 239, 627

Dimensionality, decision space, 34
Dimension of the signal set, 380
Discrete, Kalman filter, 159
optimum linear filter, 157
time processes, 629
Discriminators, conventional limiter-, 626
Distance, between mean vectors, 100
Bhattacharyya, 127
Distortion, rate-, 626
Distortionless filters, 459, 598, 599
Distribution, binomial, 145
Poisson, 29
Diversity, classical, 111
frequency, 449
optimum, 414, 415, 416
polarization, 449
Rayleigh channels, 414
Rician channels, 416
space, 449
systems, 564
Doppler shift, 7
Double-sideband AM, 9, 424
Doubly spread targets, 627
Dummy hypothesis technique, 51
Dynamic system, state variables of a linear, 517, 534

Effective noise temperature, 240
Efficient, asymptotically, 71, 445
commutation of efficiency, 84
estimates, conditions for efficient estimates, 275, 439
definition of an efficient estimate, 66
Eigenfunctions, 180
augmented, 181
scalar with matrix eigenvalues, 223
vector, scalar eigenvalues, 221
Eigenfunctions and eigenvalues, asymptotic properties of, 205
properties of, 204
solution for optimum linear filter in terms of, 203
Eigenvalues, 180
F matrices, 526
maximum and minimum properties, 208
monotonicity property, 204
significant, 193
Eigenvectors, matrices, 104
Ellipse, bound, 81
concentration, 79, 81
Ellipsoids, concentration, 79
Ensemble, 174
Envelopes, complex, 626
Error, absolute error cost function, 54
bounds, optimal diversity, Rayleigh channel, 415
closed-form expressions, colored noise, 505
linear operations, 505
problems, 493
white noise, 498, 505
correlation matrix, 84
function, bounds on, 39
complement, 37
definition, 37
interval estimation error, 437
matrix, lower bound on, 453
mean-square error, approximation, 170
Butterworth family, 502
effect of delay on, 494
Gaussian family, 505
irreducible, 494
representation, 170
unrealizable, 496
mean-square error bounds, multiple nonrandom parameters, 79
multiple random parameters, 84
nonrandom parameters, 66
random parameters, 72
measures of, 76
minimum $\Pr(\epsilon)$ criterion, 30
probability of error, 37, 257, 397, 399
Estimates, estimation, 52, 141
amplitude, random phase channel, 400
arrival time (also PPM), 276
asymptotically efficient, 276
Bayes, 52, 141
biased, 146
colored noise, 307, 454, 562, 572, 611
consistent, 71
efficient, 66, 68
frequency, random phase channel, 400
general Gaussian (classical), 156
linear, 156
nonlinear, 156
linear, 271, 308
maximum a posteriori, 63, 426
parameter, 57
waveform interval, 430
waveform point, 470
maximum likelihood, 65

Estimates, maximum likelihood, parameter, 65
waveforms, 456, 465
minimum mean-square, 437
multidimensional waveform, 446
multiple parameter, 74, 150
sequential, 144, 158, 618, 627
state variable approach, 515
summary, of classical theory, 85
of continuous waveform theory, 459
unbiased, 64
Estimator-correlator realization, 626
-correlator receiver, 354
-subtractor filter, 295

Factoring of higher order moments of Gaussian process, 228
Factorization, spectrum, 488
Fading channel, Rayleigh, 352
False alarm probability, 31
Fast-fluctuating point targets, 627
Feedback systems, optimum feedback filter, canonic realization, 510
optimum feedback systems, 508
problems, 595
receiver-to-transmitter feedback systems, 629
Filters, distortionless, 459, 598, 599
-envelope detector receiver, 341
Kalman-Bucy, 515, 599
matched, 226, 249
matrix, 480
optimum, 198, 488, 546
postloop, 510, 511
-squarer receiver, 353
time-varying, 198
transversal, 161
whitening, approach, 290
realizable, 483, 586, 618
reversibility, 289
Wiener, 481, 588
Fisher's information matrix, 80
FM, 424
Fredholm equations, first kind, 315, 316
homogeneous, 186
rational kernels, 315, 316, 320
second kind, 315, 320
separable kernels, 316, 322
vector, 221, 368
Frequency, diversity, 449
domain characterization, 167
Frequency, estimation, random phase channel, 400
modulation, 424
FSK, 379
Function-variation method, 268

Gaussian, assumption, 471
asymptotically, 71
general Gaussian problem, classical, 96
detection, problems, 154
nonlinear estimation, problems, 156
processes, 182
definition, 183
factoring of higher moments of, 228
multiple, 185
problems on, 228
properties of, 182
sampling approach, 231
white noise, 197
variable(s), characteristic function, 96
definition, 96
probability density of jointly Gaussian variables, 97
quadratic form of correlated Gaussian variables, 98, 396
Generalized, likelihood ratio test, 92, 366
Q-function, 411
Generation of coefficients, 170
Generation of random processes, 518
Geometric interpretation, sufficient statistic, 35
Global optimality, 383
Gram-Schmidt procedure, 181, 258, 380

Hilbert transform, 591
Homogeneous integral equations, 186
Hypotheses, composite, 88, 151
dummy, 51
tests, general binary, 254
M-ary, 46, 257
simple binary, 23, 134

Impulse response matrix, 532
Incoherent reception, ASK, 399
asymptotic behavior, M-ary signaling, 400
definition, 343
N channels, on-off signaling, 411
orthogonal signals, 413
Inequality, Cramér-Rao, nonrandom parameters, 66, 275
Schwarz, 67

Information, kernel, 441, 444, 454
mutual, 585
matrix, data, 84
Fisher's, 80
Integral equations, Fredholm, 315
homogeneous, 180, 186
problems, 233, 389
properties of homogeneous, 180
rational kernels, 315
solution, 315
summary, 325
Integrated Fourier transforms, 215, 224, 236
Integrating over unwanted parameters, 87
Interference, intersymbol, 160
non-Gaussian, 377
other targets, 323
Internal phase structure, 343
Intersymbol interference, 160
Interval estimation, 430
Inverse, kernel, definition, 294
matrix kernel, 368, 408
Ionospheric, link, 349
point-to-point scatter system, 240
Irreducible errors, 494

Kalman-Bucy filters, 515
Kalman filter, discrete, 159
Karhunen-Loève expansion, 182
Kernels, channel, 333
information, 441, 444, 454
of integral equations, 180
inverse, 294, 368, 408
rational, 315
separable, 316
Kineplex, 406

Lagrange multipliers, 33
"Largest-of" receiver, 258
Learning systems, 160
Likelihood, equation, 65
function, 65
maximum-likelihood estimate, 65, 456, 465
ratio, 26
ratio test, generalized, 92, 366
ordinary, 26
l.i.m., 179
Limiter-discriminator, 626
Limit in the mean, 179
Linear, arrays, 464
channels, 393
dynamic system, 517
estimation, 308, 467
Linear filters, before transmission, 569
optimum, 198, 488, 546
time-varying, 198
Linear modulation, communications, context, 575, 612
definition, 427, 467
Linear operations on random processes, 176
Loop filter, optimum, 509

MAP equations, continuous messages, 431
Marcum's Q-function, 344
Markov process, 175
-differential equation approach, 629
probability density, 228
M-ary, 257, 380–386, 397, 399–405, 415–416
Matched filter, 226, 249, 341
Matrix(ices), bound, 372, 453
covariance, definition, 97
equal covariance matrix problem, 98
eigenvalues of, 104
eigenvectors of, 104
error matrix, vector waveform estimation, 453
F-, 520
impulse response, 532
information, 438, 454
inverse kernel, 368, 408
partial derivative, 150
Riccati equation, 543
state-transition, 530, 531
Maximal ratio combiner, 369, 565
Maximization of signal-to-noise ratio, 13, 226
Maximum, a posteriori, estimate, 57, 63
interval estimation, waveforms, 430
probability computer, 50
probability test, 50
likelihood estimation, 65
signal-to-noise ratio criterion, 13
Mean-square, approximation error, 170
convergence, 178
error, closed form expressions for, 498
commutation of, 75

Mean-square, error, effect of delay on, 494
unrealizable, 496
Measurement, channel, 352
perfect measurement, bound, 88
Pr(ϵ) in Rayleigh channel, 358
Measures of error, 76
Mechanism, probabilistic transition, 20
Mercer's theorem, 181
M hypotheses, classical, 46
general Gaussian, 154
Minimax, operating point, 45
tests, 33
Minimum-distance receiver, 257
Miss probability, 31
Model, observation, 534
Models for signal detection, 239
Modified Bessel function of the first kind, 338
Modulation, amplitude modulation, DSB, 9, 424, 576, 578
SSB, 581
double-sideband AM, 9
frequency modulation, 424
index, 445
linear, 467, 575, 612
multilevel systems, 446
nonlinear, 427
phase modulation, 424
pulse amplitude (PAM), 6
pulse frequency (PFM), 7, 278
Modulator, biphase, 240
Moment(s), of Gaussian process, 228
generating function, definition, 118
method of sample moments, 151
Most powerful (UMP) tests, 89
$\mu(s)$, definition, 118
Multidimensional waveform estimation, 446, 462
Multilevel modulation systems, 446
Multiple, channels, 366, 408, 446
input systems, 528
output systems, 528
parameter estimation, 74, 150, 370, 417
processes, 185, 446, 627
Multiplex transmission systems, 627
Multivariable systems and processes, 627
Mutual information, 585

Narrow-band signals, 626
Neyman-Pearson, criterion, 24, 33
tests, 33
Noise, bandwidth, 226
temperature, 240
white, 196
Non-Gaussian interference, 377, 629
Nonlinear, demodulators, optimum, 626
estimation, in colored Gaussian noise, 308
general Gaussian (classical), 157
in white Gaussian noise, 273
modulation, 427
systems, 584
Nonparametric techniques, 628
Nonrandom, multiple nonrandom variables, bounds on estimation errors, 79
parameters
Cramér-Rao inequality, 66
estimation (problems), 145
waveform estimation, 456, 465, 598, 599
Nonrational spectra, 511
Nonsingular linear transformation, state vector, 526
Nonstationary processes, 194, 527
Nonstructured approach, 12, 15

Observation space, 20
On-off signaling, additive noise channel, 379
N incoherent channels, 411
N Rayleigh channels, 414
partially coherent channel, 397
Operating, characteristic, receiver (ROC), 36
point, minimax, 45
Orthogonal, representations, 169
signals, *M* orthogonal–known channel, 261
N incoherent channels, 413
N Rayleigh channels, 415
N Rician channels, 416
one of *M*, 403
one Rayleigh channel, 356–359, 401
one Rician channel, 361–364, 402
single incoherent, 397, 400
Orthonormal, complete (CON), 171
Orthonormal functions, 168

Paley-Wiener criterion, 512
PAM, 244
Parameter-variation method, 269
Parseval's theorem, 171

Partial, characterization, 175
derivative matrix operator, 76, 150
fraction expansion, 492, 524
Passive sonar, 626
Pattern recognition, 160, 628
Patterns, beam, 627
Perfect measurement, bound, 90
in Rayleigh channel, 359
Performance bounds and approximations, 116
Periodic processes, 209, 235
P_F, P_M, $\Pr(\epsilon)$, approximations, 124, 125
bound, 39, 122, 123
PFM, 278
Phase-lock loop, 626
Phase modulation, 424
Physical realizations, 629
Pilot tone, 583
PM improvement, 443
PM/PM, 448, 463
Point, estimate, 470
estimation error, 199
estimator, Bayes, 477
target, fast-fluctuating, 627
target, slow-fluctuating, 627
Point-to-point ionospheric scatter system, 240
Poisson, distribution, 29, 41
random process, 136
Pole-splitting techniques, 556, 607
Positive definite, 177, 179
Postloop filter, optimum, 511
Power density spectrum, 178
Power function, 89
Probability, computer, maximum a posteriori, 50
density, Gaussian process, joint, 185
Markov process, 228
Rayleigh, 351
Rician envelope and phase, 413
detection probability, 31
of error [$\Pr(\epsilon)$], binary nonorthogonal signal, random phase, 399
bit, 384
definition, 37
minimum $\Pr(\epsilon)$ receiver, 37
orthogonal signals, uniform phase, 397
false alarm, 31
miss, 31
test, a maximum a posteriori, 50

Q-function, 344, 395, 411
$Q_n(t, u)$ and $g(t)$, 294
Quadratic form, 150, 396

Radar, ambiguity function, 627
conventional pulse, 244, 344, 412, 414
mapping, 627
Radial prolate spheroidal function, 193
Radio astronomy, 9, 626
Random, modulation matrix, 583
process, *see* Process
Randomized decision rules, 43, 137
Randomly time-varying channel, 626
Rate-distortion, 626
Rational kernels, 315
Rational spectra, 187, 485
Rayleigh channels, definition of, 352
M-orthogonal signals, 401
N channels, binary orthogonal signals, 415
N channels, on-off signaling, 414
$\Pr(\epsilon)$, orthogonal signals, 357
$\Pr(\epsilon)$, perfect measurement in, 359
ROC, 356
Realizable demodulation, 576
Realizable, linear filter, 481, 515
part operator, 488
whitening filter, 388, 586, 618
Realization(s), analog computer, 517, 519, 522–525, 534, 599
correlator, colored noise receiver, 293
estimator-correlator, 626
feedback, 508, 595
matched filter-envelope detector, 341
physical, 629
state variables, 522
whitening, colored noise receiver, 293
Receiver(s), channel measurement, 358
known signals in colored noise, 293
known signals in white noise
correlation, 249, 256
"largest-of," 258
matched filter, 249
minimum distance, 257
minimum probability of error, 30
operating characteristic (ROC), 36
signals with random parameters in noise, channel estimator-correlator, 354
correlator-squarer, 353
filter-squarer, 353
sub-optimum, 379

Rectangular bandwidth, equivalent, 491
Representation, bandpass process, 227
differential equation, 516
errors, 170
infinite time interval, 212
orthogonal series, 169
sampling, 227
vector, 173
Reproducing densities, 142
Residual carrier component, 424, 467
Residues, 526
Resolution, 324, 627
Reversibility, 289, 387
Riccati equation, 543
Rician channel, 360–364, 402, 413, 416
Risk, Bayes, 24
ROC, 36, 44, 250

Sample space, 174
Sampling, approach to continuous Gaussian processes, 231
representation, 227
Scatter communication, 626
Schwarz inequality, 66
Second-moment characterizations, 176, 226
Seismic systems, 627
Sensitivity, 267, 326–331, 391, 512, 573
Sequence of digits, 264, 627
Sequential detection and estimation schemes, 627
Sequential estimation, 144, 158, 618
Signals, bi-orthogonal, 384
equally correlated, 267
known, 3, 7, 9, 246, 287
M-ary, 257, 380
optimum design, 302, 393
orthogonal, 260–267, 356–359, 361–364, 397, 400, 402, 403, 413, 415, 416
random, 4
random parameters, 333, 394
Simplex, 267, 380, 383
Simplex signals, 267, 380, 383
Single-sideband, suppressed carrier, amplitude modulation, 581
Single-time characterizations, 176
Singularity, 303, 391
Slow-fading, 352
Slowly fluctuating point target, 627
Sonar, 3, 626
Space, decision, 250
Space, observation, 20
parameter, 53
sample, 174
Space-time system, 449
Spectra, bandlimited, 192
Butterworth, 191
nonrational, 511
rational, 187, 485
Spectral decomposition, applications, 218
Spectrum factorization, 488
Specular component, 360
Spheroidal function, 193
Spread targets, 627
Square-error cost function, 54
Square-law receiver, 340, 402
State of the system, definition, 517
State transition matrix, 530, 531
State variables, 517
State vector, 520
Structured, approach, 12
processes, 175
Sufficient statistics, cosine, 361
definition, 35
estimation, 59
Gaussian problem, 100, 104
geometric interpretation, 35
sine, 361

Tapped delay line, 161
Target, doubly spread, 627
fluctuating, 414, 627
interfering, 323
nonfluctuating, 412
point, 627
Temperature, effective noise, 240
Tests, Bayes, 24, 139
binary, 23, 134
generalized likelihood ratio, 92, 366
hypothesis, 23, 134
infinitely sensitive, 331
likelihood ratio, 26
maximum a posteriori probability, 50
M hypothesis, 46, 139
minimax, 33
Neyman-Pearson, 33
UMP, 89, 366
unstable, 331
Threshold constraint, 281
Threshold of tests, 26
Tilted densities, 119

Time-domain characterization, 166
Time-varying differential equations, 527, 603
Time-varying linear filters, 198, 199
Transformations, coordinate system, 102
  no-memory, 426
  state vector, nonsingular linear, 526
Transition matrix, state, 530, 531
Translation of signal sets, 380
Transversal filter, 161

Unbiased estimates, 64
Uniform cost function, 54
Uniformly most powerful (UMP) tests, 89, 366
Uniqueness of optimum linear filters, 476
Unknown bias, 64
Unrealizable filters, 496
Unstable tests, 331
Unwanted parameters, 87, 333, 394

Vector, bias, 76
  channels, 537
  differential equations, 515, 520
  eigenfunctions, 221
Vector, Gaussian, 96
  integrated transform, 224
  mean, 100, 107
  orthogonal expansion, 221
  random processes, 230, 236
  representation of signals, 172
  state, 520
  Wiener-Hopf equation, 480, 539

Waveforms, continuous waveform estimation, 11, 423
  multidimensional, 446
  nonrandom, 456, 465
Whitening, approach, 290
  property, 483
  realization, colored noise, 293
  reversible whitening filter, 290, 387, 388, 586, 618
White noise process, 196
Wiener, filter, 481, 588
  process, 194
Wiener-Hopf equation, 482

Detection, Estimation, and Modulation Theory, Part I:
Detection, Estimation, and Linear Modulation Theory. Harry L. Van Trees
Copyright © 2001 John Wiley & Sons, Inc.
ISBNs: 0-471-09517-6 (Paperback); 0-471-22108-2 (Electronic)

2

Classical Detection and Estimation Theory

2.1 INTRODUCTION

In this chapter we develop in detail the basic ideas of classical detection and estimation theory. The first step is to define the various terms.

The basic components of a simple decision-theory problem are shown in Fig. 2.1. The first is a *source* that generates an output. In the simplest case this output is one of two choices. We refer to them as hypotheses and label them H_0 and H_1 in the two-choice case. More generally, the output might be one of M hypotheses, which we label $H_0, H_1, \ldots, H_{M-1}$. Some typical source mechanisms are the following:

1. A digital communication system transmits information by sending ones and zeros. When "one" is sent, we call it H_1, and when "zero" is sent, we call it H_0.
2. In a radar system we look at a particular range and azimuth and try

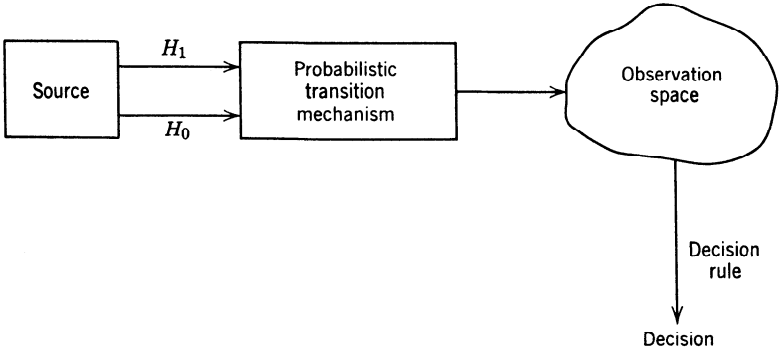

Fig. 2.1 Components of a decision theory problem.

because the details involved in constructing the likelihood ratios are the same as those in the binary case. Typical examples are given in the problems. Several important points should be emphasized.

1. The minimum dimension of the decision space is no more than $M - 1$. The boundaries of the decision regions are hyperplanes in the $(\Lambda_1, \ldots, \Lambda_{M-1})$ plane.
2. The optimum test is straightforward to find. We shall find however, when we consider specific examples that the error probabilities are frequently difficult to compute.
3. A particular test of importance is the minimum total probability of error test. Here we compute the a posteriori probability of each hypothesis $\Pr(H_i|\mathbf{R})$ and choose the largest.

These points will be appreciated more fully as we proceed through various applications.

These two sections complete our discussion of simple hypothesis tests. A case of importance that we have not yet discussed is the one in which several source outputs are combined to give a single hypothesis. To study this detection problem, we shall need some ideas from estimation theory. Therefore we defer the composite hypothesis testing problem until Section 2.5 and study the estimation problem next.

2.4 ESTIMATION THEORY

In the last two sections we have considered a problem in which one of several hypotheses occurred. As the result of a particular hypothesis, a vector random variable $\mathbf{r}$ was observed. Based on our observation, we shall try to choose the true hypothesis.

In this section we discuss the problem of *parameter estimation*. Before formulating the general problem, let us consider a simple example.

Example 1. We want to measure a voltage a at a single time instant. From physical considerations, we know that the voltage is between $-V$ and $+V$ volts. The measurement is corrupted by noise which may be modeled as an independent additive zero-mean Gaussian random variable n. The observed variable is r. Thus

$$r = a + n. \tag{116}$$

The probability density governing the observation process is $p_{r|a}(R|A)$. In this case

$$p_{r|a}(R|A) = p_n(R - A) = \frac{1}{\sqrt{2\pi}\,\sigma_n} \exp\left(-\frac{(R - A)^2}{2\sigma_n^2}\right). \tag{117}$$

The problem is to observe r and estimate a.

This example illustrates the basic features of the estimation problem.

A model of the general estimation problem is shown in Fig. 2.17. The model has the following four components:

Parameter Space. The output of the source is a parameter (or variable). We view this output as a point in a parameter space. For the single-parameter case, which we shall study first, this will correspond to segments of the line $-\infty < A < \infty$. In the example considered above the segment is $(-V, V)$.

Probabilistic Mapping from Parameter Space to Observation Space. This is the probability law that governs the effect of a on the observation.

Observation Space. In the classical problem this is a finite-dimensional space. We denote a point in it by the vector $\mathbf{R}$.

Estimation Rule. After observing $\mathbf{R}$, we shall want to estimate the value of a. We denote this estimate as $\hat{a}(\mathbf{R})$. This mapping of the observation space into an estimate is called the estimation rule. The purpose of this section is to investigate various estimation rules and their implementations.

The second and third components are familiar from the detection problem. The new features are the parameter space and the estimation rule. When we try to describe the parameter space, we find that two cases arise. In the first, the parameter is a random variable whose behavior is governed by a probability density. In the second, the parameter is an unknown quantity but not a random variable. These two cases are analogous to the

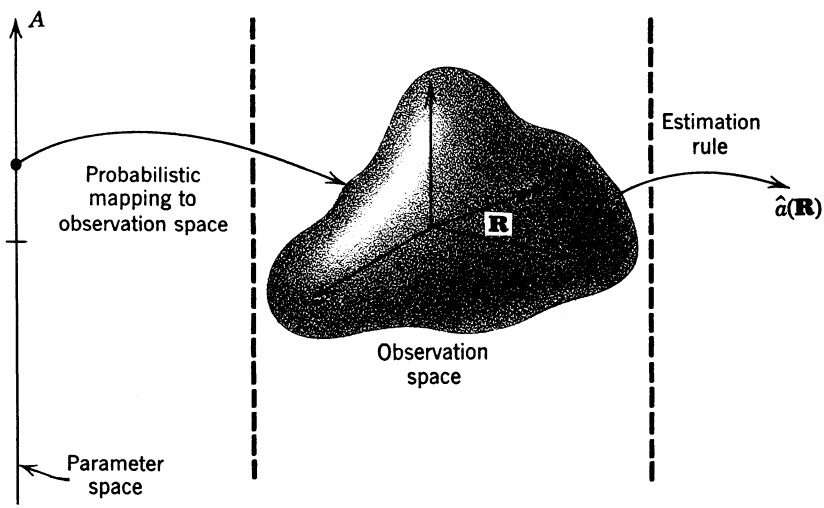

Fig. 2.17 Estimation model.

source models we encountered in the hypothesis-testing problem. To correspond with each of these models of the parameter space, we shall develop suitable estimation rules. We start with the random parameter case.

2.4.1 Random Parameters: Bayes Estimation

In the Bayes detection problem we saw that the two quantities we had to specify were the set of costs C_{ij} and the a priori probabilities P_i. The cost matrix assigned a cost to each possible course of action. Because there were M hypotheses and M possible decisions, there were M^2 costs. In the estimation problem a and $\hat{a}(\mathbf{R})$ are continuous variables. Thus we must assign a cost to all pairs $[a, \hat{a}(\mathbf{R})]$ over the range of interest. This is a function of two variables which we denote as $C(a, \hat{a})$. In many cases of interest it is realistic to assume that the cost depends only on the error of the estimate. We define this error as

$$a_\epsilon(\mathbf{R}) \triangleq \hat{a}(\mathbf{R}) - a. \tag{118}$$

The cost function $C(a_\epsilon)$ is a function of a single variable. Some typical cost functions are shown in Fig. 2.18. In Fig. 2.18*a* the cost function is simply the square of the error:

$$C(a_\epsilon) = a_\epsilon^2. \tag{119}$$

This cost is commonly referred to as the squared error cost function. We see that it accentuates the effects of large errors. In Fig. 2.18*b* the cost function is the absolute value of the error:

$$C(a_\epsilon) = |a_\epsilon|. \tag{120}$$

In Fig. 2.18*c* we assign zero cost to all errors less than $\pm\Delta/2$. In other words, an error less than $\Delta/2$ in magnitude is as good as no error. If $a_\epsilon > \Delta/2$, we assign a uniform value:

$$\begin{aligned} C(a_\epsilon) &= 0, \qquad |a_\epsilon| \le \frac{\Delta}{2}, \\ &= 1, \qquad |a_\epsilon| > \frac{\Delta}{2}. \end{aligned} \tag{121}$$

In a given problem we choose a cost function to accomplish two objectives. First, we should like the cost function to measure user satisfaction adequately. Frequently it is difficult to assign an analytic measure to what basically may be a subjective quality.

Our goal is to find an estimate that minimizes the expected value of the cost. Thus our second objective in choosing a cost function is to assign one that results in a tractable problem. In practice, cost functions are usually some compromise between these two objectives. Fortunately, in many

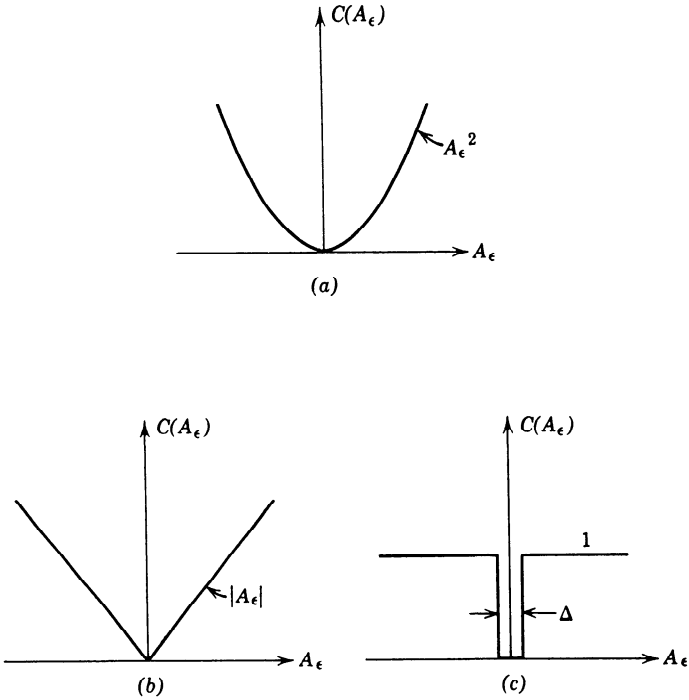

Fig. 2.18 Typical cost functions: (*a*) mean-square error; (*b*) absolute error; (*c*) uniform cost function.

problems of interest the same estimate will be optimum for a large class of cost functions.

Corresponding to the a priori probabilities in the detection problem, we have an a priori probability density $p_a(A)$ in the random parameter estimation problem. In all of our discussions we assume that $p_a(A)$ is known. If $p_a(A)$ is not known, a procedure analogous to the minimax test may be used.

Once we have specified the cost function and the a priori probability, we may write an expression for the risk:

$$\mathcal{R} \triangleq E\{C[a, \hat{a}(\mathbf{R})]\} = \int_{-\infty}^{\infty} dA \int_{-\infty}^{\infty} C[A, \hat{a}(\mathbf{R})] p_{a,\mathbf{r}}(A, \mathbf{R})\, d\mathbf{R}. \quad (122)$$

The expectation is over the random variable a and the observed variables $\mathbf{r}$. For costs that are functions of one variable only (122) becomes

$$\mathcal{R} = \int_{-\infty}^{\infty} dA \int_{-\infty}^{\infty} C[A - \hat{a}(\mathbf{R})] p_{a,\mathbf{r}}(A, \mathbf{R})\, d\mathbf{R}. \quad (123)$$

The Bayes estimate is the estimate that minimizes the risk. It is straightforward to find the Bayes estimates for the cost functions in Fig. 2.18. For the cost function in Fig. 2.18*a*, the risk corresponds to mean-square error. We denote the risk for the mean-square error criterion as $\mathcal{R}_{\text{ms}}$. Substituting (119) into (123), we have

$$\mathcal{R}_{\text{ms}} = \int_{-\infty}^{\infty} dA \int_{-\infty}^{\infty} d\mathbf{R}[A - \hat{a}(\mathbf{R})]^2 p_{a,\mathbf{r}}(A, \mathbf{R}). \tag{124}$$

The joint density can be rewritten as

$$p_{a,\mathbf{r}}(A, \mathbf{R}) = p_{\mathbf{r}}(\mathbf{R}) p_{a|\mathbf{r}}(A|\mathbf{R}). \tag{125}$$

Using (125) in (124), we have

$$\mathcal{R}_{\text{ms}} = \int_{-\infty}^{\infty} d\mathbf{R}\, p_{\mathbf{r}}(\mathbf{R}) \int_{-\infty}^{\infty} dA[A - \hat{a}(\mathbf{R})]^2 p_{a|\mathbf{r}}(A|\mathbf{R}). \tag{126}$$

Now the inner integral and $p_{\mathbf{r}}(\mathbf{R})$ are non-negative. Therefore we can minimize $\mathcal{R}_{\text{ms}}$ by minimizing the inner integral. We denote this estimate $\hat{a}_{\text{ms}}(\mathbf{R})$. To find it we differentiate the inner integral with respect to $\hat{a}(\mathbf{R})$ and set the result equal to zero:

$$\begin{aligned} &\frac{d}{d\hat{a}} \int_{-\infty}^{\infty} dA[A - \hat{a}(\mathbf{R})]^2 p_{a|\mathbf{r}}(A|\mathbf{R}) \\ &\qquad = -2 \int_{-\infty}^{\infty} A p_{a|\mathbf{r}}(A|\mathbf{R})\, dA + 2\hat{a}(\mathbf{R}) \int_{-\infty}^{\infty} p_{a|\mathbf{r}}(A|\mathbf{R})\, dA. \end{aligned} \tag{127}$$

Setting the result equal to zero and observing that the second integral equals 1, we have

$$\hat{a}_{\text{ms}}(\mathbf{R}) = \int_{-\infty}^{\infty} dA\, A p_{a|\mathbf{r}}(A|\mathbf{R}). \tag{128}$$

This is a unique minimum, for the second derivative equals two. The term on the right side of (128) is familiar as the mean of the a posteriori density (or the conditional mean).

Looking at (126), we see that if $\hat{a}(\mathbf{R})$ is the conditional mean the inner integral is just the a posteriori variance (or the conditional variance). Therefore the minimum value of $\mathcal{R}_{\text{ms}}$ is just the average of the conditional variance over all observations $\mathbf{R}$.

To find the Bayes estimate for the absolute value criterion in Fig. 2.18*b* we write

$$\mathcal{R}_{\text{abs}} = \int_{-\infty}^{\infty} d\mathbf{R}\, p_{\mathbf{r}}(\mathbf{R}) \int_{-\infty}^{\infty} dA[|A - \hat{a}(\mathbf{R})|] p_{a|\mathbf{r}}(A|\mathbf{R}). \tag{129}$$

To minimize the inner integral we write

$$I(\mathbf{R}) = \int_{-\infty}^{\hat{a}(\mathbf{R})} dA[\hat{a}(\mathbf{R}) - A]\, p_{a|\mathbf{r}}(A|\mathbf{R}) + \int_{\hat{a}(\mathbf{R})}^{\infty} dA[A - \hat{a}(\mathbf{R})]\, p_{a|\mathbf{r}}(A|\mathbf{R}). \tag{130}$$

Differentiating with respect to $\hat{a}(\mathbf{R})$ and setting the result equal to zero, we have

$$\int_{-\infty}^{\hat{a}_{\mathrm{abs}}(\mathbf{R})} dA\, p_{a|\mathbf{r}}(A|\mathbf{R}) = \int_{\hat{a}_{\mathrm{abs}}(\mathbf{R})}^{\infty} dA\, p_{a|\mathbf{r}}(A|\mathbf{R}). \tag{131}$$

This is just the definition of the median of the a posteriori density.

The third criterion is the uniform cost function in Fig. 2.18*c*. The risk expression follows easily:

$$\mathcal{R}_{\mathrm{unf}} = \int_{-\infty}^{\infty} d\mathbf{R}\, p_{\mathbf{r}}(\mathbf{R})\left[1 - \int_{\hat{a}_{\mathrm{unf}}(\mathbf{R})-\Delta/2}^{\hat{a}_{\mathrm{unf}}(\mathbf{R})+\Delta/2} p_{a|\mathbf{r}}(A|\mathbf{R})\, dA\right]. \tag{132}$$

To minimize this equation we maximize the inner integral. Of particular interest to us is the case in which Δ is an arbitrarily small but nonzero number. A typical a posteriori density is shown in Fig. 2.19. We see that for small Δ the best choice for $\hat{a}(\mathbf{R})$ is the value of A at which the a posteriori density has its maximum. We denote the estimate for this special case as $\hat{a}_{\mathrm{map}}(\mathbf{R})$, the *maximum a posteriori* estimate. In the sequel we use $\hat{a}_{\mathrm{map}}(\mathbf{R})$ without further reference to the uniform cost function.

To find $\hat{a}_{\mathrm{map}}$ we must have the location of the maximum of $p_{a|\mathbf{r}}(A|\mathbf{R})$. Because the logarithm is a monotone function, we can find the location of the maximum of $\ln p_{a|\mathbf{r}}(A|\mathbf{R})$ equally well. As we saw in the detection problem, this is frequently more convenient.

If the maximum is interior to the allowable range of A and $\ln p_{a|\mathbf{r}}(A|\mathbf{R})$ has a continuous first derivative then a necessary, but not sufficient, condition for a maximum can be obtained by differentiating $\ln p_{a|\mathbf{r}}(A|\mathbf{R})$ with respect to A and setting the result equal to zero:

$$\left.\frac{\partial \ln p_{a|\mathbf{r}}(A|\mathbf{R})}{\partial A}\right|_{A=\hat{a}(\mathbf{R})} = 0. \tag{133}$$

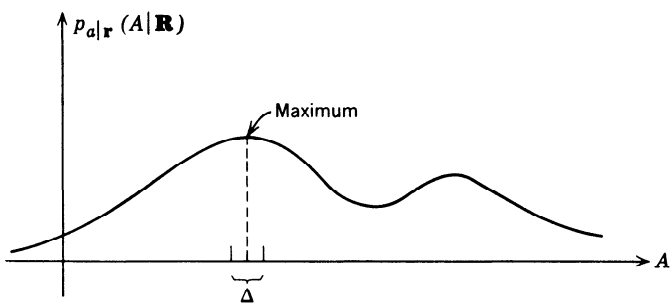

Fig. 2.19 An a posteriori density.

We refer to (133) as the MAP equation. In each case we must check to see if the solution is the absolute maximum.

We may rewrite the expression for $p_{a|\mathbf{r}}(A|\mathbf{R})$ to separate the role of the observed vector $\mathbf{R}$ and the a priori knowledge:

$$p_{a|\mathbf{r}}(A|\mathbf{R}) = \frac{p_{\mathbf{r}|a}(\mathbf{R}|A)p_a(A)}{p_{\mathbf{r}}(\mathbf{R})}. \tag{134}$$

Taking logarithms,

$$\ln p_{a|\mathbf{r}}(A|\mathbf{R}) = \ln p_{\mathbf{r}|a}(\mathbf{R}|A) + \ln p_a(A) - \ln p_{\mathbf{r}}(\mathbf{R}). \tag{135}$$

For MAP estimation we are interested only in finding the value of A where the left-hand side is maximum. Because the last term on the right-hand side is not a function of A, we can consider just the function

$$l(A) \triangleq \ln p_{\mathbf{r}|a}(\mathbf{R}|A) + \ln p_a(A). \tag{136}$$

The first term gives the probabilistic dependence of $\mathbf{R}$ on A and the second describes a priori knowledge.

The MAP equation can be written as

$$\left.\frac{\partial l(A)}{\partial A}\right|_{A=\hat{a}(\mathbf{R})} = \left.\frac{\partial \ln p_{\mathbf{r}|a}(\mathbf{R}|A)}{\partial A}\right|_{A=\hat{a}(\mathbf{R})} + \left.\frac{\partial \ln p_a(A)}{\partial A}\right|_{A=\hat{a}(\mathbf{R})} = 0. \tag{137}$$

Our discussion in the remainder of the book emphasizes minimum mean-square error and maximum a posteriori estimates.

To study the implications of these two estimation procedures we consider several examples.

Example 2. Let

$$r_i = a + n_i, \qquad i = 1, 2, \ldots, N. \tag{138}$$

We assume that a is Gaussian, $N(0, \sigma_a)$, and that the n_i are each independent Gaussian variables $N(0, \sigma_n)$. Then

$$\begin{aligned} p_{\mathbf{r}|a}(\mathbf{R}|A) &= \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi}\,\sigma_n} \exp\left(-\frac{(R_i - A)^2}{2\sigma_n^2}\right), \\ p_a(A) &= \frac{1}{\sqrt{2\pi}\,\sigma_a} \exp\left(-\frac{A^2}{2\sigma_a^2}\right). \end{aligned} \tag{139}$$

To find $\hat{a}_{\mathrm{ms}}(\mathbf{R})$ we need to know $p_{a|\mathbf{r}}(A|\mathbf{R})$. One approach is to find $p_{\mathbf{r}}(\mathbf{R})$ and substitute it into (134), but this procedure is algebraically tedious. It is easier to observe that $p_{a|\mathbf{r}}(A|\mathbf{R})$ is a probability density with respect to a for any $\mathbf{R}$. Thus $p_{\mathbf{r}}(\mathbf{R})$ just contributes to the constant needed to make

$$\int_{-\infty}^{\infty} p_{a|\mathbf{r}}(A|\mathbf{R})\, dA = 1. \tag{140}$$

(In other words, $p_{\mathbf{r}}(\mathbf{R})$ is simply a normalization constant.) Thus

$$p_{a|\mathbf{r}}(A|\mathbf{R}) = \left[\frac{\left(\prod_{i=1}^{N} \frac{1}{\sqrt{2\pi}\,\sigma_n}\right)\frac{1}{\sqrt{2\pi}\,\sigma_a}}{p_{\mathbf{r}}(\mathbf{R})}\right] \exp\left\{-\frac{1}{2}\left[\frac{\sum_{i=1}^{N}(R_i - A)^2}{\sigma_n^2} + \frac{A^2}{\sigma_a^2}\right]\right\}. \tag{141}$$

Rearranging the exponent, completing the square, and absorbing terms depending only on R_i^2 into the constant, we have

$$p_{a|\mathbf{r}}(A|\mathbf{R}) = k(\mathbf{R}) \exp\left\{-\frac{1}{2\sigma_p^2}\left[A - \frac{\sigma_a^2}{\sigma_a^2 + \sigma_n^2/N}\left(\frac{1}{N}\sum_{i=1}^{N} R_i\right)\right]^2\right\}, \tag{142}$$

where

$$\sigma_p^2 \triangleq \left(\frac{1}{\sigma_a^2} + \frac{N}{\sigma_n^2}\right)^{-1} = \frac{\sigma_a^2\sigma_n^2}{N\sigma_a^2 + \sigma_n^2} \tag{143}$$

is the a posteriori variance.

We see that $p_{a|\mathbf{r}}(A|\mathbf{R})$ is just a Gaussian density. The estimate $\hat{a}_{\mathrm{ms}}(\mathbf{R})$ is just the conditional mean

$$\hat{a}_{\mathrm{ms}}(\mathbf{R}) = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_n^2/N}\left(\frac{1}{N}\sum_{i=1}^{N} R_i\right). \tag{144}$$

Because the a posteriori variance is not a function of $\mathbf{R}$, the mean-square risk equals the a posteriori variance (see (126)).

Two observations are useful:

1. The R_i enter into the a posteriori density only through their sum. Thus

$$l(\mathbf{R}) = \sum_{i=1}^{N} R_i \tag{145}$$

is a *sufficient statistic*. This idea of a sufficient statistic is identical to that in the detection problem.

2. The estimation rule uses the information available in an intuitively logical manner. If $\sigma_a^2 \ll \sigma_n^2/N$, the a priori knowledge is much better than the observed data and the estimate is very close to the a priori mean. (In this case, the a priori mean is zero.) On the other hand, if $\sigma_a^2 \gg \sigma_n^2/N$, the a priori knowledge is of little value and the estimate uses primarily the received data. In the limit $\hat{a}_{\mathrm{ms}}$ is just the arithmetic average of the R_i.

$$\lim_{\frac{\sigma_n^2}{N\sigma_a^2} \to 0} \hat{a}_{\mathrm{ms}}(\mathbf{R}) = \frac{1}{N}\sum_{i=1}^{N} R_i. \tag{146}$$

The MAP estimate for this case follows easily. Looking at (142), we see that because the density is Gaussian the maximum value of $p_{a|\mathbf{r}}(A|\mathbf{R})$ occurs at the conditional mean. Thus

$$\hat{a}_{\mathrm{map}}(\mathbf{R}) = \hat{a}_{ms}(\mathbf{R}). \tag{147}$$

Because the conditional median of a Gaussian density occurs at the conditional mean, we also have

$$\hat{a}_{\mathrm{abs}}(\mathbf{R}) = \hat{a}_{ms}(\mathbf{R}). \tag{148}$$

Thus we see that for this particular example all three cost functions in Fig. 2.18 lead to the same estimate. This invariance to the choice of a cost function is obviously a useful feature because of the subjective judgments that are frequently involved in choosing $C(a_\epsilon)$. Some conditions under which this invariance holds are developed in the next two properties.†

† These properties are due to Sherman [20]. Our derivation is similar to that given by Viterbi [36].

Property 1. We assume that the cost function $C(a_\epsilon)$ is a symmetric, convex-upward function and that the a posteriori density $p_{a|\mathbf{r}}(A|\mathbf{R})$ is symmetric about its conditional mean; that is,

$$C(a_\epsilon) = C(-a_\epsilon) \qquad \text{(symmetry)}, \quad (149)$$

$$C(bx_1 + (1-b)x_2) \leq bC(x_1) + (1-b)\,C(x_2) \qquad \text{(convexity)} \quad (150)$$

for any b inside the range (0, 1) and for all x_1 and x_2. Equation 150 simply says that all chords lie above or on the cost function.

This condition is shown in Fig. 2.20*a*. If the inequality is strict whenever $x_1 \neq x_2$, we say the cost function is strictly convex (upward). Defining

$$z \triangleq a - \hat{a}_{\text{ms}} = a - E[a|\mathbf{R}] \quad (151)$$

the symmetry of the a posteriori density implies

$$p_{z|\mathbf{r}}(Z|\mathbf{R}) = p_{z|\mathbf{r}}(-Z|\mathbf{R}). \quad (152)$$

The estimate $\hat{a}$ that minimizes any cost function in this class is identical to $\hat{a}_{\text{ms}}$ (which is the conditional mean).

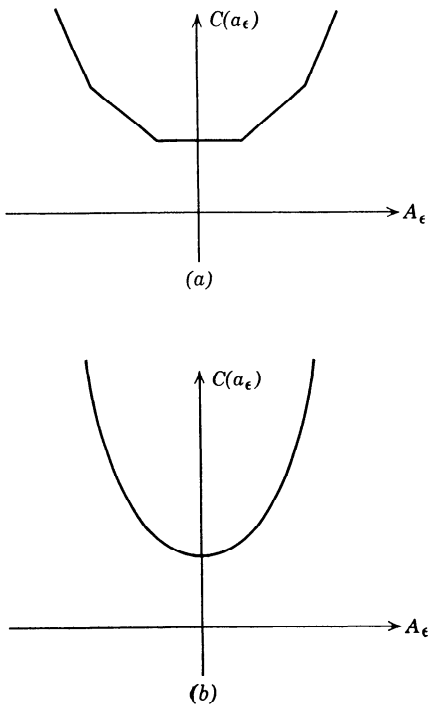

Fig. 2.20 Symmetric convex cost functions: (*a*) convex; (*b*) strictly convex.

Proof. As before we can minimize the conditional risk [see (126)]. Define

$$\mathcal{R}_B(\hat{a}|\mathbf{R}) \triangleq E_a[C(\hat{a} - a)|\mathbf{R}] = E_a[C(a - \hat{a})|\mathbf{R}], \tag{153}$$

where the second equality follows from (149). We now write four equivalent expressions for $\mathcal{R}_B(\hat{a}|\mathbf{R})$:

$$\mathcal{R}_B(\hat{a}|\mathbf{R}) = \int_{-\infty}^{\infty} C(\hat{a} - \hat{a}_{\text{ms}} - Z)p_{z|\mathbf{r}}(Z|\mathbf{R})\, dZ \tag{154}$$

[Use (151) in (153)]

$$= \int_{-\infty}^{\infty} C(\hat{a} - \hat{a}_{\text{ms}} + Z)p_{z|\mathbf{r}}(Z|\mathbf{R})\, dZ \tag{155}$$

[(152) implies this equality]

$$= \int_{-\infty}^{\infty} C(\hat{a}_{\text{ms}} - \hat{a} - Z)p_{z|\mathbf{r}}(Z|\mathbf{R})\, dZ \tag{156}$$

[(149) implies this equality]

$$= \int_{-\infty}^{\infty} C(\hat{a}_{\text{ms}} - \hat{a} + Z)p_{z|\mathbf{r}}(Z|\mathbf{R})\, dZ \tag{157}$$

[(152) implies this equality].

We now use the convexity condition (150) with the terms in (155) and (157):

$$\begin{aligned}\mathcal{R}_B(\hat{a}|\mathbf{R}) &= \tfrac{1}{2}E(\{C[Z + (\hat{a}_{\text{ms}} - \hat{a})] + C[Z - (\hat{a}_{\text{ms}} - \hat{a})]\}|\mathbf{R}) \\ &\geq E\{C[\tfrac{1}{2}(Z + (\hat{a}_{\text{ms}} - \hat{a})) + \tfrac{1}{2}(Z - (\hat{a}_{\text{ms}} - \hat{a}))]|\mathbf{R}\} \\ &= E[C(Z)|\mathbf{R}].\end{aligned} \tag{158}$$

Equality will be achieved in (158) if $\hat{a}_{\text{ms}} = \hat{a}$. This completes the proof. If $C(a_\epsilon)$ is strictly convex, we will have the additional result that the minimizing estimate $\hat{a}$ is unique and equals $\hat{a}_{\text{ms}}$.

To include cost functions like the uniform cost functions which are not convex we need a second property.

Property 2. We assume that the cost function is a symmetric, nondecreasing function and that the a posteriori density $p_{a|\mathbf{r}}(A|\mathbf{R})$ is a symmetric (about the conditional mean), unimodal function that satisfies the condition

$$\lim_{x \to \infty} C(x)p_{a|\mathbf{r}}(x|\mathbf{R}) = 0.$$

The estimate $\hat{a}$ that minimizes any cost function in this class is identical to $\hat{a}_{\text{ms}}$. The proof of this property is similar to the above proof [36].

The significance of these two properties should not be underemphasized. Throughout the book we consider only minimum mean-square and maximum a posteriori probability estimators. Properties 1 and 2 ensure that whenever the a posteriori densities satisfy the assumptions given above the estimates that we obtain will be optimum for a large class of cost functions. Clearly, if the a posteriori density is Gaussian, it will satisfy the above assumptions.

We now consider two examples of a different type.

Example 3. The variable a appears in the signal in a nonlinear manner. We denote this dependence by $s(A)$. Each observation r_i consists of $s(A)$ plus a Gaussian random variable n_i, $N(0, \sigma_n)$. The n_i are statistically independent of each other and a. Thus

$$r_i = s(A) + n_i. \tag{159}$$

Therefore

$$p_{a|\mathbf{r}}(A|\mathbf{R}) = k(\mathbf{R}) \exp\left(-\frac{1}{2}\left\{\frac{\sum_{i=1}^{N} [R_i - s(A)]^2}{\sigma_n^2} + \frac{A^2}{\sigma_a^2}\right\}\right). \tag{160}$$

This expression cannot be further simplified without specifying $s(A)$ explicitly.

The MAP equation is obtained by substituting (160) into (137)

$$\hat{a}_{\mathrm{map}}(\mathbf{R}) = \frac{\sigma_a^2}{\sigma_n^2} \sum_{i=1}^{N} [R_i - s(A)] \frac{\partial s(A)}{\partial A}\bigg|_{A = \hat{a}_{\mathrm{map}}(\mathbf{R})}. \tag{161}$$

To solve this explicitly we must specify $s(A)$. We shall find that an analytic solution is generally not possible when $s(A)$ is a nonlinear function of A.

Another type of problem that frequently arises is the estimation of a parameter in a probability density.

Example 4. The number of events in an experiment obey a Poisson law with mean value a. Thus

$$\Pr(n \text{ events} \mid a = A) = \frac{A^n}{n!} \exp(-A), \qquad n = 0, 1, \ldots. \tag{162}$$

We want to observe the number of events and estimate the parameter a of the Poisson law. We shall assume that a is a random variable with an exponential density

$$p_a(A) = \begin{cases} \lambda \exp(-\lambda A), & A > 0, \\ 0, & \text{elsewhere.} \end{cases} \tag{163}$$

The a posteriori density of a is

$$p_{a|n}(A|N) = \frac{\Pr(n = N \mid a = A) p_a(A)}{\Pr(n = N)}. \tag{164}$$

Substituting (162) and (163) into (164), we have

$$p_{a|n}(A|N) = k(N)[A^N \exp(-A(1 + \lambda))], \qquad A \geq 0, \tag{165}$$

where

$$k(N) = \frac{(1 + \lambda)^{N+1}}{N!} \tag{166}$$

in order for the density to integrate to 1. (As already pointed out, the constant is unimportant for MAP estimation but is needed if we find the MS estimate by integrating over the conditional density.)

The mean-square estimate is the conditional mean:

$$\hat{a}_{\mathrm{ms}}(N) = \frac{(1+\lambda)^{N+1}}{N!}\int_0^\infty A^{N+1}\exp[-A(1+\lambda)]\,dA$$

$$= \frac{(1+\lambda)^{N+1}}{(1+\lambda)^{N+2}}(N+1) = \left(\frac{1}{\lambda+1}\right)(N+1). \tag{167}$$

To find $\hat{a}_{\mathrm{map}}$ we take the logarithm of (165)

$$\ln p_{a|n}(A|N) = N\ln A - A(1+\lambda) + \ln k(N). \tag{168}$$

By differentiating with respect to A, setting the result equal to zero, and solving, we obtain

$$\hat{a}_{\mathrm{map}}(N) = \frac{N}{1+\lambda}. \tag{169}$$

Observe that $\hat{a}_{\mathrm{map}}$ is not equal to $\hat{a}_{\mathrm{ms}}$.

Other examples are developed in the problems. The principal results of this section are the following:

1. The minimum mean-square error estimate (MMSE) is always the mean of the a posteriori density (the conditional mean).
2. The maximum a posteriori estimate (MAP) is the value of A at which the a posteriori density has its maximum.
3. For a large class of cost functions the optimum estimate is the conditional mean whenever the a posteriori density is a unimodal function which is symmetric about the conditional mean.

These results are the basis of most of our estimation work. As we study more complicated problems, the only difficulty we shall encounter is the actual evaluation of the conditional mean or maximum. In many cases of interest the MAP and MMSE estimates will turn out to be equal.

We now turn to the second class of estimation problems described in the introduction.

2.4.2 Real (Nonrandom) Parameter Estimation†

In many cases it is unrealistic to treat the unknown parameter as a random variable. The problem formulation on pp. 52–53 is still appropriate. Now, however, the parameter is assumed to be nonrandom, and we want to design an estimation procedure that is good in some sense.

† The beginnings of classical estimation theory can be attributed to Fisher [5, 6, 7, 8]. Many discussions of the basic ideas are now available (e.g., Cramer [9]), Wilks [10], or Kendall and Stuart [11]).

A logical first approach is to try to modify the Bayes procedure in the last section to eliminate the average over $p_a(A)$. As an example, consider a mean-square error criterion,

$$\mathcal{R}(A) \triangleq \int_{-\infty}^{\infty} [\hat{a}(\mathbf{R}) - A]^2 \, p_{\mathbf{r}|a}(\mathbf{R}|A) \, d\mathbf{R}, \tag{170}$$

where the expectation is only over $\mathbf{R}$, for it is the only random variable in the model. Minimizing $\mathcal{R}(A)$, we obtain

$$\hat{a}_{\mathrm{ms}}(\mathbf{R}) = A. \tag{171}$$

The answer is correct, but not of any value, for A is the unknown quantity that we are trying to find. Thus we see that this direct approach is not fruitful. A more useful method in the nonrandom parameter case is to examine other possible measures of quality of estimation procedures and then to see whether we can find estimates that are good in terms of these measures.

The first measure of quality to be considered is the expectation of the estimate

$$E[\hat{a}(\mathbf{R})] \triangleq \int_{-\infty}^{+\infty} \hat{a}(\mathbf{R}) \, p_{\mathbf{r}|a}(\mathbf{R}|A) \, d\mathbf{R}. \tag{172}$$

The possible values of the expectation can be grouped into three classes

1. If $E[\hat{a}(\mathbf{R})] = A$, for all values of A, we say that the estimate is *unbiased.* This statement means that the average value of the estimates equals the quantity we are trying to estimate.
2. If $E[\hat{a}(\mathbf{R})] = A + B$, where B is not a function of A, we say that the estimate has a *known bias.* We can always obtain an unbiased estimate by subtracting B from $\hat{a}(\mathbf{R})$.
3. If $E[\hat{a}(\mathbf{R})] = A + B(A)$, we say that the estimate has an *unknown bias.* Because the bias depends on the unknown parameter, we cannot simply subtract it out.

Clearly, even an unbiased estimate may give a bad result on a particular trial. A simple example is shown in Fig. 2.21. The probability density of the estimate is centered around A, but the variance of this density is large enough that big errors are probable.

A second measure of quality is the variance of estimation error:

$$\mathrm{Var}\,[\hat{a}(\mathbf{R}) - A] = E\{[\hat{a}(\mathbf{R}) - A]^2\} - B^2(A). \tag{173}$$

This provides a measure of the spread of the error. In general, we shall try to find unbiased estimates with small variances. There is no straightforward minimization procedure that will lead us to the minimum variance unbiased estimate. Therefore we are forced to try an estimation procedure to see how well it works.

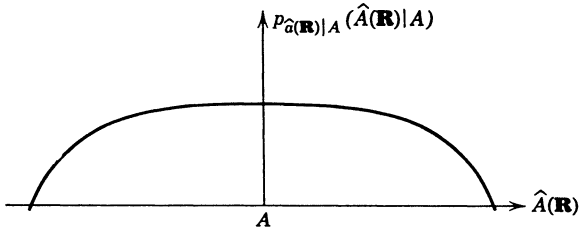

Fig. 2.21 Probability density for an estimate.

Maximum Likelihood Estimation. There are several ways to motivate the estimation procedure that we shall use. Consider the simple estimation problem outlined in Example 1. Recall that

$$r = A + n, \tag{174}$$

$$p_{r|a}(R|A) = (\sqrt{2\pi}\,\sigma_n)^{-1} \exp\left[-\frac{1}{2\sigma_n^2}(R - A)^2\right]. \tag{175}$$

We choose as our estimate the value of A that most likely caused a given value of R to occur. In this simple additive case we see that this is the same as choosing the most probable value of the noise ($N = 0$) and subtracting it from R. We denote the value obtained by using this procedure as a maximum likelihood estimate.

$$\hat{a}_{ml}(R) = R. \tag{176}$$

In the general case we denote the function $p_{\mathbf{r}|a}(\mathbf{R}|A)$, viewed as a function of A, as the *likelihood function.* Frequently we work with the logarithm, $\ln p_{\mathbf{r}|a}(\mathbf{R}|A)$, and denote it as the *log likelihood function.* The maximum likelihood estimate $\hat{a}_{ml}(\mathbf{R})$ is that value of A at which the likelihood function is a maximum. If the maximum is interior to the range of A, and $\ln p_{\mathbf{r}|a}(\mathbf{R}|A)$ has a continuous first derivative, then a necessary condition on $\hat{a}_{ml}(\mathbf{R})$ is obtained by differentiating $\ln p_{\mathbf{r}|a}(\mathbf{R}|A)$ with respect to A and setting the result equal to zero:

$$\left.\frac{\partial \ln p_{\mathbf{r}|a}(\mathbf{R}|A)}{\partial A}\right|_{A=\hat{a}_{ml}(\mathbf{R})} = 0. \tag{177}$$

This equation is called the *likelihood equation.* Comparing (137) and (177), we see that the ML estimate corresponds mathematically to the limiting case of a MAP estimate in which the a priori knowledge approaches zero.

In order to see how effective the ML procedure is we can compute the bias and the variance. Frequently this is difficult to do. Rather than approach the problem directly, we shall first derive a lower bound on the variance on *any* unbiased estimate. Then we shall see how the variance of $\hat{a}_{ml}(\mathbf{R})$ compares with this lower bound.

Cramér-Rao Inequality: Nonrandom Parameters. We now want to consider the variance of *any* estimate $\hat{a}(\mathbf{R})$ of the real variable A. We shall prove the following statement.

Theorem. (a) If $\hat{a}(\mathbf{R})$ is *any* unbiased estimate of A, then

$$\operatorname{Var}[\hat{a}(\mathbf{R}) - A] \geq \left(E\left\{\left[\frac{\partial \ln p_{\mathbf{r}|a}(\mathbf{R}|A)}{\partial A}\right]^2\right\}\right)^{-1} \tag{178}$$

or, equivalently,

(b)

$$\operatorname{Var}[\hat{a}(\mathbf{R}) - A] \geq \left\{-E\left[\frac{\partial^2 \ln p_{\mathbf{r}|a}(\mathbf{R}|A)}{\partial A^2}\right]\right\}^{-1}, \tag{179}$$

where the following conditions are assumed to be satisfied:

(c)

$$\frac{\partial p_{\mathbf{r}|a}(\mathbf{R}|A)}{\partial A} \quad \text{and} \quad \frac{\partial^2 p_{\mathbf{r}|a}(\mathbf{R}|A)}{\partial A^2}$$

exist and are absolutely integrable.

The inequalities were first stated by Fisher [6] and proved by Dugué [31]. They were also derived by Cramér [9] and Rao [12] and are usually referred to as the Cramér-Rao bound. Any estimate that satisfies the bound with an equality is called an *efficient* estimate.

The proof is a simple application of the Schwarz inequality. Because $\hat{a}(\mathbf{R})$ is unbiased,

$$E[\hat{a}(\mathbf{R}) - A] \triangleq \int_{-\infty}^{\infty} p_{\mathbf{r}|a}(\mathbf{R}|A)[\hat{a}(\mathbf{R}) - A]\, d\mathbf{R} = 0. \tag{180}$$

Differentiating both sides with respect to A, we have

$$\frac{d}{dA}\int_{-\infty}^{\infty} p_{\mathbf{r}|a}(\mathbf{R}|A)[\hat{a}(\mathbf{R}) - A]\, d\mathbf{R} = \int_{-\infty}^{\infty} \frac{\partial}{\partial A}\{p_{\mathbf{r}|a}(\mathbf{R}|A)[\hat{a}(\mathbf{R}) - A]\}\, d\mathbf{R} = 0, \tag{181}$$

where condition (c) allows us to bring the differentiation inside the integral. Then

$$-\int_{-\infty}^{\infty} p_{\mathbf{r}|a}(\mathbf{R}|A)\, d\mathbf{R} + \int_{-\infty}^{\infty} \frac{\partial p_{\mathbf{r}|a}(\mathbf{R}|A)}{\partial A}[\hat{a}(\mathbf{R}) - A]\, d\mathbf{R} = 0. \tag{182}$$

The first integral is just $+1$. Now observe that

$$\frac{\partial p_{\mathbf{r}|a}(\mathbf{R}|A)}{\partial A} = \frac{\partial \ln p_{\mathbf{r}|a}(\mathbf{R}|A)}{\partial A} p_{\mathbf{r}|a}(\mathbf{R}|A). \tag{183}$$

Substituting (183) into (182), we have

$$\int_{-\infty}^{\infty} \frac{\partial \ln p_{\mathbf{r}|a}(\mathbf{R}|A)}{\partial A} p_{\mathbf{r}|a}(\mathbf{R}|A)[\hat{a}(\mathbf{R}) - A]\, d\mathbf{R} = 1. \tag{184}$$

Rewriting, we have

$$\int_{-\infty}^{\infty} \left[\frac{\partial \ln p_{\mathbf{r}|a}(\mathbf{R}|A)}{\partial A} \sqrt{p_{\mathbf{r}|a}(\mathbf{R}|A)}\right]\left[\sqrt{p_{\mathbf{r}|a}(\mathbf{R}|A)}\,[\hat{a}(\mathbf{R}) - A]\right] d\mathbf{R} = 1, \tag{185}$$

and, using the Schwarz inequality, we have

$$\left\{\int_{-\infty}^{\infty} \left[\frac{\partial \ln p_{\mathbf{r}|a}(\mathbf{R}|A)}{\partial A}\right]^2 p_{\mathbf{r}|a}(\mathbf{R}|A)\, d\mathbf{R}\right\} \times \left\{\int_{-\infty}^{\infty} [\hat{a}(\mathbf{R}) - A]^2 p_{\mathbf{r}|a}(\mathbf{R}|A)\, d\mathbf{R}\right\} \geq 1, \tag{186}$$

where we recall from the derivation of the Schwarz inequality that equality holds if and only if

$$\frac{\partial \ln p_{\mathbf{r}|a}(\mathbf{R}|A)}{\partial A} = [\hat{a}(\mathbf{R}) - A]\, k(A), \tag{187}$$

for all $\mathbf{R}$ and A. We see that the two terms of the left side of (186) are the expectations in statement (*a*) of (178). Thus,

$$E\{[\hat{a}(\mathbf{R}) - A]^2\} \geq \left\{E\left[\frac{\partial \ln p_{\mathbf{r}|a}(\mathbf{R}|A)}{\partial A}\right]^2\right\}^{-1}. \tag{188}$$

To prove statement (b) we observe

$$\int_{-\infty}^{\infty} p_{\mathbf{r}|a}(\mathbf{R}|A)\, d\mathbf{R} = 1. \tag{189}$$

Differentiating with respect to A, we have

$$\int_{-\infty}^{\infty} \frac{\partial p_{\mathbf{r}|a}(\mathbf{R}|A)}{\partial A}\, d\mathbf{R} = \int_{-\infty}^{\infty} \frac{\partial \ln p_{\mathbf{r}|a}(\mathbf{R}|A)}{\partial A} p_{\mathbf{r}|a}(\mathbf{R}|A)\, d\mathbf{R} = 0. \tag{190}$$

Differentiating again with respect to A and applying (183), we obtain

$$\int_{-\infty}^{\infty} \frac{\partial^2 \ln p_{\mathbf{r}|a}(\mathbf{R}|A)}{\partial A^2} p_{\mathbf{r}|a}(\mathbf{R}|A)\, d\mathbf{R} + \int_{-\infty}^{\infty} \left(\frac{\partial \ln p_{\mathbf{r}|a}(\mathbf{R}|A)}{\partial A}\right)^2 p_{\mathbf{r}|a}(\mathbf{R}|A)\, d\mathbf{R} = 0 \tag{191}$$

or

$$E\left[\frac{\partial^2 \ln p_{\mathbf{r}|a}(\mathbf{R}|A)}{\partial A^2}\right] = -E\left[\frac{\partial \ln p_{\mathbf{r}|a}(\mathbf{R}|A)}{\partial A}\right]^2, \tag{192}$$

which together with (188) gives condition (b).