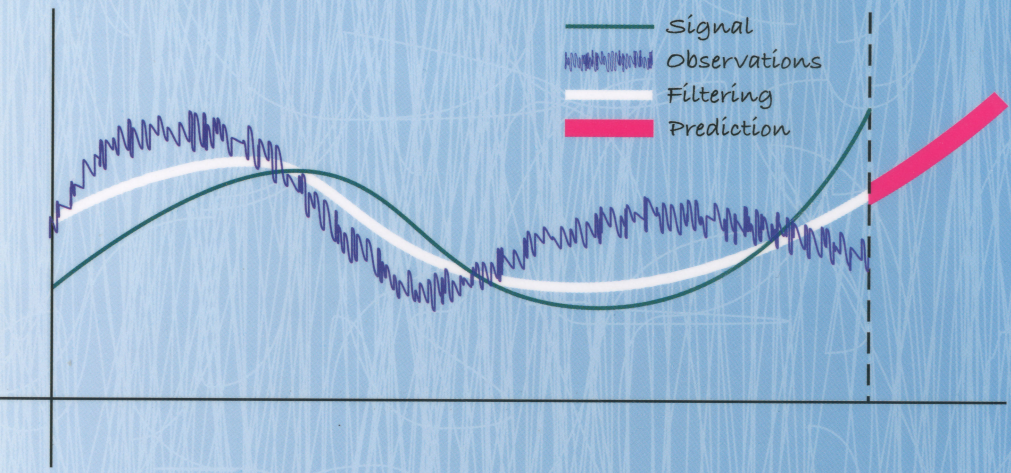
STUDENT MATHEMATICAL LIBRARY
Volume 38

Filtering and Prediction: A Primer

B. Fristedt
N. Jain
N. Krylov



AMS
American Mathematical Society

http://dx.doi.org/10.1090/stml/038

Filtering and Prediction: A Primer

This page intentionally left blank

STUDENT MATHEMATICAL LIBRARY
Volume 38

Filtering and Prediction: A Primer

B. Fristedt
N. Jain
N. Krylov

AMS
AMERICAN MATHEMATICAL SOCIETY
Providence, Rhode Island

Editorial Board

Gerald B. Folland
Robin Forman (Chair)
Brad G. Osgood
Michael Starbird

2000 *Mathematics Subject Classification.* Primary 60–01; Secondary 60G99.

Nicolai Krylov was supported in part by NSF grants DMS-0140405 and DMS-0653121.

For additional information and updates on this book, visit
www.ams.org/bookpages/stml-38

Library of Congress Cataloging-in-Publication Data

Filtering and prediction : a primer / B. Fristedt, N. Jain, N. Krylov.
p. cm. — (Student mathematical library, ISSN 1520-9121 ; v. 38)
Includes bibliographical references and index.
ISBN 978-0-8218-4333-8 (alk. paper)
1. Markov processes. 2. Filters (Mathematics) 3. Prediction theory. I. Fristedt, Bert, 1937– II. Jain, N. (Naresh), 1937– III. Krylov, N. V. (Nikolai Vladimirovich)

QA274.7.F53 2007
519.2'33—dc22 2007060783

Copying and reprinting. Individual readers of this publication, and nonprofit libraries acting for them, are permitted to make fair use of the material, such as to copy a chapter for use in teaching or research. Permission is granted to quote brief passages from this publication in reviews, provided the customary acknowledgment of the source is given.

Republication, systematic copying, or multiple reproduction of any material in this publication is permitted only under license from the American Mathematical Society. Requests for such permission should be addressed to the Acquisitions Department, American Mathematical Society, 201 Charles Street, Providence, Rhode Island 02904-2294, USA. Requests can also be made by e-mail to `reprint-permission@ams.org`.

© 2007 by the American Mathematical Society. All rights reserved.
The American Mathematical Society retains all rights
except those granted to the United States Government.
Printed in the United States of America.

⊗ The paper used in this book is acid-free and falls within the guidelines established to ensure permanence and durability.
Visit the AMS home page at `http://www.ams.org/`

10 9 8 7 6 5 4 3 2 1 12 11 10 09 08 07

Contents

Preface ix

Chapter 1. Preliminaries 1

§1. Series 2

1:1. Sums and series. 2

1:2. Some examples 7

1:3. The theorems of Fubini and Tonelli. 8

§2. Probability concepts 10

2:1. Random variables. 10

2:2. Expectations of real-valued random variables that are not necessarily discrete 14

2:3. Second moments, variance, and standard deviation 15

2:4. Independence 16

2:5. Some important distributions 18

§3. Conditioning in the discrete case 19

3:1. Generalities. 19

3:2. Applications in estimating 29

Chapter 2. Markov chains 33

§1. Random walks 33

§2. Discrete time and space Markov chains 42

§3. Further problems on Markov chains 57

Chapter 3. Filtering of discrete Markov chains 61

§1. Filtering 63

§2. Parameter estimation for discrete Markov chains 72

§3. Interpolation for discrete Markov chains 80

3:1. Interpolation equation 80

3:2. Conjugate equations 87

3:3. Other problems related to interpolation 89

§4. Prediction for discrete Markov chains 92

Chapter 4. Conditional expectations 95

§1. L_2-spaces 95

1:1. Case of finite Ω 95

1:2. General case 104

§2. Definition of conditional expectations 108

2:1. General case 108

2:2. The case of Gaussian random variables 112

§3. Conditional expectations and densities 118

Chapter 5. Filtering of continuous-space Markov chains 127

§1. Filtering of $\mathbb{R}^d$-valued Markov chains 127

§2. Discrete-time Kalman filter 132

2:1. One-dimensional case 133

2:2. Multidimensional case 137

§3. Linear filtering 147

Chapter 6. Wiener process and continuous time filtering 161

§1. Introduction 161

§2. Definition and simplest properties of the Wiener process 162

§3. Integration against the Wiener process 168

§4. Recalling power series and systems of ODEs 172

§5. Kalman filter in continuous time 179

Chapter 7. Stationary sequences 191

§1. Definition and simplest properties 191

§2. Spectral densities 197

§3. Filtering stationary sequences 204

§4. The Bochner-Khinchin theorem (optional) 207

§5. The law of large numbers (optional) 210

§6. Spectral representation of stationary sequences (optional) 213

Chapter 8. Prediction of stationary sequences 225

§1. Some properties of rational spectral densities 227

§2. Predicting one step ahead 233

§3. Predicting many steps ahead 240

§4. Dynamic predicting 243

Bibliography 247

Index 249

This page intentionally left blank

Preface

In many scientific activities, observations are made on moving objects and they are generally corrupted by random errors. The theory of filtering provides means of filtering out the "noise" in the observations and extracting from them the most precise information about the moving object, which itself may or may not be moving in a random fashion. Then the prediction step follows. On the basis of information about the past behavior of the object, one tries to predict its future path. Two simple examples of such moving objects are the motion of a rocket and the changing price of a commodity in the stock market.

The theory of filtering and prediction is highly developed and a huge literature exists that deals with numerous aspects of the subject. Even though the theory at the advanced level requires a basic knowledge of advanced analysis and probability and is highly technical, several texts have been written at the undergraduate level. At that level, the compromises that have been made by other authors seem to be somewhat unnecessary. We feel that a fairly rigorous treatment of the basic ideas can be given in such a way that the interest of a serious undergraduate student can be sustained.

The material in the current text has been offered by us a few times in the past as a one-semester course. Based on the experience, we designed the text basically to consist of two parts: the first three chapters and the five remaining ones.

The first three chapters deal with an introduction to the filtering theory of Markov chains with discrete time and space; interpolation and prediction questions are also introduced. In particular, we treat Hidden Markov Models. Even though one does not need any heavy machinery here, all the definitions are carefully made and the treatment is quite rigorous, yet in this "simple" setting the problems can be quite challenging. We believe there are enough exercises to illustrate the main ideas and also to challenge a strong student. The main purpose, though, is to develop the main ideas. Some of the exercises have an asterisk to indicate that the later text will use the contents of those exercises. Altogether in the book there are about 85 starred exercises and 166 additional ones.

The next five chapters form the second part of the book. This part requires a more mature attitude and more commitment than normally expected from average undergraduates. However, our experience shows that the priming that the student gets by going through the first three chapters does lead to this kind of maturity and commitment. In the Spring semester of 2006 we had ten students finish the course; three of them were graduate students from three different departments and the remaining seven were undergraduates mostly from the department of mathematics. One of the best students turned out to be a junior from our Lower Division.

Chapter 4 introduces the "abstract" notion of conditional expectation. The introduction of this notion in terms of L_2-theory seems most natural in the context of filtering and prediction. We have made a rigorous presentation of the ideas without making it too technical. All the "usual" elementary definitions of conditional probability and expectation are shown to be special cases of this general setting. This chapter lays the foundation for the development of the theory in continuous time and space in the next four chapters.

Chapter 5 deals with discrete time Markov chains whose state space is $\mathbb{R}^d$. The discrete Kalman filter for $d = 1$ and $d \geq 1$, as well as linear filtering, are discussed in this chapter. Special attention is paid to so-called Hidden Markov Models. We also derive general filtering equations for Markov chains whose transition probability functions have a density function (kernel) with respect to Lebesgue measure.

Even though Lebesgue measure or even Riemann integration are not introduced in this text, those unfamiliar with these notions may simply assume the kernel to be jointly continuous. Once integration of continuous functions is understood, the rest is easy to follow and is self-contained.

Chapter 6 deals with continuous time filtering in the important case of the Wiener process. The exercises are designed to help the student with the development of the ideas while minimizing technicalities without sacrificing rigor. Stochastic integral with respect to the Wiener process is defined only for smooth nonrandom summands for obvious reasons. Some of the exercises could be challenging for even the strongest students without helpful hints from the instructor.

Stationary sequences (wide sense) are introduced in Chapter 7. We decided to include an elementary case of the Bochner-Khinchin representation theorem for positive definite functions since it plays an important role in this context. In addition to filtering such sequences we have included some optional sections on the law of large numbers and the spectral representation of stationary sequences. These sections may be skipped if time becomes a factor.

Finally, Chapter 8 contains an introduction to the prediction of stationary sequences. We only deal with autoregressive and autoregressive-moving average sequences. Some ideas from the theory of complex analysis are included rather than referenced in order to make the text self contained as far as possible.

The reader who would like to pursue further study on related topics will certainly benefit from consulting references [1] through [7].

We are thankful to our students for their patience because the text was under development for a number of years, which necessitated changes and revisions. We hope that this is a version that will be of sufficient interest to other colleagues and we hope to get comments from them. We already got quite a few comments and deep observations from J. Baxter, who taught this course in Spring 2007 and whose contribution is greatly appreciated.

B. Fristedt, N. Jain, N. Krylov, May 2007.

This page intentionally left blank

This page intentionally left blank

Bibliography

[1] Bar-Shalom Y., Li X.R, and Kirubarajan T., "Estimation with applications to tracking and navigation", Wiley Inter-Science, 2001.

[2] Davis M.H.A., "Linear estimation and stochastic control", Chapman and Hall, London, 1977.

[3] Davis M.H.A. and Vinter R.B., "Stochastic modelling and control", Chapman and Hall, London, 1985.

[4] Gīhman, Ĭ.Ī. and Skorokhod A.V., "The theory of stochastic processes. I", *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*, **210**, Springer-Verlag, Berlin-New York, 1980.

[5] Liptser R.Sh. and Shiryayev A.N., "Statistics of random processes", Springer-Verlag, New York, 1977-1978.

[6] Rabiner L. and Juang B-H., "Fundamentals of speech recognition", PTR Prentice Hall, Englewood Cliffs, NJ, 1993.

[7] Yaglom A.M., "An introduction to the theory of stationary random functions", Prentice-Hall, Englewood Cliffs, NJ, 1962.

This page intentionally left blank

Index

(p_{ij}^k), 52
$:=$, 8
$=:$, 8
$A \times B$, 8
$C_{[k,n]}$, 43
$E(X-k)^2$, 15
$E(X \mid Y_{[1,k]} = y_{[1,k]})$, 109
$E(X \mid Y_{[1,k]})$, 108
$E(X \mid Z)$, 23
$Ef(X)$, 11
e^{tA}, 174
F_X, 14
g_n^a, 91
$H_1 H_2 \ldots$, 13
I_H, 12
$I_{X=a}$, 12
$k^{\pm}$, 3
L_a^{-1}, 201
$L_2(-\pi, \pi)$, 215
$L_2(Y_{(-\infty,n]})$, 194
L_a, 201
$P(A \mid Y_{[1,k]} = y_{[1,k]})$, 109
$P(A \mid Y_{[1,k]})$, 108
$P(G \mid H)$, 19
$P(H \mid Z)$, 22
$P(H_1, H_2, \ldots)$, 13
$P(X \mid H)$, 19
$p(c, z)$, 47
$p(x \mid y)$, 121
$p^{\star k}$, 51
$p_{n+1}(y \mid x)$, 44
Q_{XY}, 120
$q^{rs}(u, v)$, 63
$q^{x_1 x_2}(y_1, y_2)$, 130
Tf, 127
T_n^a, 90
$X \perp Y$, 98
$Z_{[k,n]}$, 43
$z_{[k,n]}$, 43
$\bar{Q}_{XX}$, 120
$\bar{U}_{|n}$, 139
$\bar{X}_{n+1\,,,n}$138
$\bar{X}_n$, 61
$\bar{\boldsymbol{\kappa}}_{nm}^a$, 82
$\bar{\Sigma}_{n+1}$, 138
$\bigcap_i H_i$, 12
$\bigcup_i H_i$, 12
$\mathbb{N}$, 2
$\boldsymbol{\pi}_{nm}^r$, 82
$\mathbb{Z}$, 2
$\mathbb{Z}^+$, 2
$\mathbb{Z}^+ \setminus \{0\}$, 2
$\mathbb{Z}^d$, 2, 33
$\boldsymbol{\zeta}_{nm}^{ar}$, 89
$\emptyset$, 13
$\eta_n^r(b_{[0,n]})$, 81
$\frac{c}{0}$, 21
$\hat{Y}_{n|k}$, 195
$\kappa_{nm}^a(b_{[n,m]})$, 83
l.i.m , 97
$\langle X, Y \rangle$, 97
$\langle x, y \rangle_{\mathbb{R}^n}$, 97
Ω, 12, 95
Π_Y, 102
$\Pi_{\mathcal{L}}$, 99
$\phi_n^a(b_{[0,n]})$, 63
$\pi_{nm}^{a,b}$, 92
π_{nm}^{ar}, 87
π_{nm}^a, 81
$\pi_{nm}^a(b_{[0,m]})$, 80
π_n^a, 64

$\pi_n^a(b_{[0,n]})$, 64
π^n, 44
$\pi_0(x \mid y)$, 130
$\pi_n(b_{[0,n]})$, 64
$\pi_n(x \mid y_{[0,n]})$, 130
$\prod$, 16
$\prod_i$, 12
ρ_n^a, 67
Σ_{XY}, 115
Span, 99
$\sigma(Y_{[0,t]})$, 180
$\sigma(Y_{[1,k]})$, 108
$\sum_{a \in A} f(a)$, 2
Var(X), 16
$\wedge$, 163
$\xi_{n+1}(b_{[0,n+1]})$, 64
$||A||$, 172
$||X||_{\mathbb{R}^n}$, 97
$||x||_{\mathbb{R}^n}$, 97
(a.s.), 23

almost surely, 23
amplitude response, 206

basis, 100
Bayesian setting, 73
Bernoulli
  distribution, 18
  type, 18
binomial distribution, 18

Cartesian product, 8
Cauchy sequence, 98
Cauchy's inequality, 98
closed subspace, 100
complete space, 97
conditional
  convergence of series, 8
  density, 121
  expectation, 108
  expectation of X given Z, 23
  probability distribution, 31
  probability of H given Z, 22
convergence
  absolute, 5
  in the mean-square sense, 97
  of series conditional, 8
countable
  additivity, 13
  set, 1
countably infinite set, 1
covariance, 114
covariance matrix, 115

density, 118
  posterior, 121
discrete random variable, 11
  distribution of, 11
  expectation of, 11
  expected value of, 11
  mean of, 11
discrete random variables
  independent, 16
discrete spectrum, 200, 209, 214
disjoint events, 13
distribution
  Bernoulli, 18
  binomial, 18
  conditional, 31
  equilibrium, 54
  function, 14
  geometric, 18
  initial, 34, 43
  invariant, 54
  of random variable, 11
  posterior, 31
  prior, 31
  spectral, 209
  step, 34
  unnormalized filtering, 67
double-norm, 172

effective phase space, 48
equilibrium distribution, 54
estimate
  best mean square, 15
  mean-square optimal, 29
  of X given Y, 29
events
  disjoint, 13
  independent, 16
expectation, 11
  conditional, 23, 108
  of discrete random variable, 11
  undefined, 11
expected value, 11

filter
  characterized by H, 205
  input, 205
  output, 205
filtering equations, 63
  unnormalized, 63
finite set, 1
Fourier coefficients, 216
frequency response, 206
Fubini's Theorem
  for series, 10
function
  k-step transition, 51
  n^{th} transition, 47
  characteristic, 112
  correlation, 191
  delta-, 192
  distribution, 14
  spectral distribution, 209
  transition, 47

fundamental solution, 88

Gaussian
  (m, Σ) random variable, 115
  process, 163
  random variables, 113
geometric
  distribution, 18
  series, 7

hidden Markov models, 66, 153

independent
  events, 16
  random variables, 16
indicator random variable, 12
inequality
  Cauchy's, 98
  triangle, 98
infinite series, 5
initial
  distribution, 34, 43
  probability vector, 52
invariant distribution, 54

jump of random walk, 33

Kalman's equations, 136

Law of Total Probability, 22
least mean squared error, 15
length, 97
linear
  filtering problem, 103
  regression, 103
  space, 215

Markov chain
  in discrete time, 42
  time-homogeneous, 47
Markov property, 43
matrix Riccati equation, 144, 182
mean, 11
  value, 115
mean-square
  convergence, 97
  optimal estimate, 29
minimax setting, 73
moment
  second, 15
moving averages, 193

negative part, 3
nonlinear filtering problem, 103
nonnegative definite matrix, 115
norm, 97
  of a matrix, 172
normal random variables, 113

Ornstein-Uhlenbeck process, 179
orthogonal
  projection, 99
  random variables, 98

parallelogram identity, 99
Parseval's identity, 216
periodogram, 202
positive part, 3
posterior
  density, 121
  distribution, 31
power spectrum, 200
prior distribution, 31
probability of heads, 18
probability vector at time n, 52
product
  Cartesian, 8
  scalar, 97, 105, 215
property of the correlation function, 204
pseudo inverse matrices, 144

random
  orthogonal measure, 221
  process, 163
  spectral measure, 213
  vector, 11
random variable
  discrete, 11
  distribution of, 11
  expectation of, 11
  expected value of, 11
  mean of, 11
random variables
  independent, 16
random walk, 33
  jump of, 33
  on $\mathbb{Z}^+ \cup \{\infty\}$, 58
  recurrent, 38
  simple, 34
  simple symmetric, 34
  starting state of, 34
  state of, 34
  step of, 33
  symmetric, 34
response
  amplitude, 206
  frequency, 206
Riccati matrix equation, 144, 182

scalar product, 97, 105, 215
second moment, 15
self-similarity, 164
sequence
  AR, 235
  ARMA, 235
  autoregressive, 235
  autoregressive-moving average, 235

Cauchy, 98
innovating, 145, 235
mean-square stationary, 191
Ornstein-Uhlenbeck, 129, 194
positive definite, 196
stationary, 129
white noise, 200
series, 5
conditional convergence of, 8
divergence of, 5
Fubini's Theorem for, 10
geometric, 7
infinite, 5
Tonelli's Theorem for, 8
set
countable, 1
finite, 1
uncountable, 2
setting
Bayesian, 73
minimax, 73
simple random walk, 34
simple symmetric random walk, 34
space
complete, 97
spectral
density, 200
measure, 209
standard deviation, 16
starting state, 34
state of random walk, 34
starting, 34
state space, 43
step
distribution, 34
of random walk, 33
stochastic integral, 168
subspace, 99
closed, 100
success probability, 18
sum
of series, 6
sure event, 12
symmetric
matrix, 115
random walk, 34

time-homogeneous
Markov chain, 47
Tonelli's Theorem
for series, 8
Total Probability, law of, 22
trajectory, 163
transition
kernel, 127
matrix, 52
operator, 127
transition function, 47
k-step, 51
triangle inequality, 98

uncountable set, 2
unnormalized filtering
density, 132
distribution, 67
equations, 63
probabilities, 67

variance, 16
vector
initial probability, 52
probability at time n, 52
random, 11

Weierstrass's theorem, 174
white noise, 200
Wiener process, 163

Filtering and prediction is about observing moving objects when the observations are corrupted by random errors. The main focus is then on filtering out the errors and extracting from the observations the most precise information about the object, which itself may or may not be moving in a somewhat random fashion. Next comes the prediction step where, using information about the past behavior of the object, one tries to predict its future path.

The first three chapters of the book deal with discrete probability spaces, random variables, conditioning, Markov chains, and filtering of discrete Markov chains. The next three chapters deal with the more sophisticated notions of conditioning in nondiscrete situations, filtering of continuous-space Markov chains, and of Wiener process. Filtering and prediction of stationary sequences is discussed in the last two chapters.

The authors believe that they have succeeded in presenting necessary ideas in an elementary manner without sacrificing the rigor too much. Such rigorous treatment is lacking at this level in the literature. In the past few years the material in the book was offered as a one-semester undergraduate/beginning graduate course at the University of Minnesota. Some of the many problems suggested in the text were used in homework assignments.

ISBN 978-0-8218-4333-8
9 780821 843338
STML/38

For additional information and updates on this book, visit
www.ams.org/bookpages/stml-38

AMS *on the* Web
www.ams.org