

Kenneth Lange

Applied Probability



Springer

Kenneth Lange
Department of Biomathematics
UCLA School of Medicine
Los Angeles, CA 90095-1766
USA
klange@ucla.edu

Editorial Board

George Casella
Department of Statistics
University of Florida
Gainesville, FL 32611-8545
USA

Stephen Fienberg
Department of Statistics
Carnegie Mellon University
Pittsburgh, PA 15213-3890
USA

Ingram Olkin
Department of Statistics
Stanford University
Stanford, CA 94305
USA

Library of Congress Cataloging-in-Publication Data
Lange, Kenneth.

Applied probability / Kenneth Lange.

p. cm. — (Springer texts in statistics)

Includes bibliographical references and index.

ISBN 0-387-00425-4 (alk. paper)

1. Probabilities. 2. Stochastic processes. I. Title. II. Series.

QA273.L2684 2003

519.2—dc21

2003042436

ISBN 0-387-00425-4

Printed on acid-free paper.

© 2003 Springer-Verlag New York, Inc.

All rights reserved. This work may not be translated or copied in whole or in part without the written permission of the publisher (Springer-Verlag New York, Inc., 175 Fifth Avenue, New York, NY 10010, USA), except for brief excerpts in connection with reviews or scholarly analysis. Use in connection with any form of information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed is forbidden. The use in this publication of trade names, trademarks, service marks, and similar terms, even if they are not identified as such, is not to be taken as an expression of opinion as to whether or not they are subject to proprietary rights.

Printed in the United States of America.

9 8 7 6 5 4 3 2 1

SPIN 10908198

Typesetting: Pages created by the author using a Springer T_EX macro package.

www.springer-ny.com

Springer-Verlag New York Berlin Heidelberg
A member of BertelsmannSpringer Science+Business Media GmbH

Contents

Preface to the Second Edition	vii
Preface to the First Edition	ix
0.1 References	xi
1 Basic Principles of Population Genetics	1
1.1 Introduction	1
1.2 Genetics Background	1
1.3 Hardy-Weinberg Equilibrium	4
1.4 Linkage Equilibrium	8
1.5 Selection	9
1.6 Balance Between Mutation and Selection	12
1.7 Problems	13
1.8 References	19
2 Counting Methods and the EM Algorithm	21
2.1 Introduction	21
2.2 Gene Counting	21
2.3 Description of the EM Algorithm	23
2.4 Ascent Property of the EM Algorithm	24
2.5 Allele Frequency Estimation by the EM Algorithm	26
2.6 Classical Segregation Analysis by the EM Algorithm	27
2.7 Binding Domain Identification	31
2.8 Problems	32
2.9 References	37
3 Newton's Method and Scoring	39
3.1 Introduction	39
3.2 Newton's Method	39
3.3 Scoring	40
3.4 Application to the Design of Linkage Experiments	43
3.5 Quasi-Newton Methods	45
3.6 The Dirichlet Distribution	47
3.7 Empirical Bayes Estimation of Allele Frequencies	48
3.8 Empirical Bayes Estimation of Haplotype Frequencies	51
3.9 Problems	52
3.10 References	57

4	Hypothesis Testing and Categorical Data	59
4.1	Introduction	59
4.2	Hypotheses About Genotype Frequencies	59
4.3	Other Multinomial Problems in Genetics	62
4.4	The Z_{\max} Test	63
4.5	The W_d Statistic	65
4.6	Exact Tests of Independence	67
4.7	Case-Control Association Tests	69
4.8	The Transmission/Disequilibrium Test	70
4.9	Problems	73
4.10	References	76
5	Genetic Identity Coefficients	81
5.1	Introduction	81
5.2	Kinship and Inbreeding Coefficients	81
5.3	Condensed Identity Coefficients	84
5.4	Generalized Kinship Coefficients	86
5.5	From Kinship to Identity Coefficients	86
5.6	Calculation of Generalized Kinship Coefficients	88
5.7	Problems	91
5.8	References	94
6	Applications of Identity Coefficients	97
6.1	Introduction	97
6.2	Genotype Prediction	97
6.3	Covariances for a Quantitative Trait	99
6.4	Risk Ratios and Genetic Model Discrimination	102
6.5	An Affecteds-Only Method of Linkage Analysis	106
6.6	Problems	109
6.7	References	113
7	Computation of Mendelian Likelihoods	115
7.1	Introduction	115
7.2	Mendelian Models	115
7.3	Genotype Elimination and Allele Consolidation	118
7.4	Array Transformations and Iterated Sums	120
7.5	Array Factoring	122
7.6	Examples of Pedigree Analysis	124
7.7	Problems	133
7.8	References	137
8	The Polygenic Model	141
8.1	Introduction	141
8.2	Maximum Likelihood Estimation by Scoring	142
8.3	Application to Gc Measured Genotype Data	146

8.4	Multivariate Traits	147
8.5	Left and Right-Hand Finger Ridge Counts	149
8.6	QTL Mapping	150
8.7	Factor Analysis	151
8.8	A QTL Example	152
8.9	The Hypergeometric Polygenic Model	154
8.10	Application to Risk Prediction	157
8.11	Problems	158
8.12	References	165
9	Descent Graph Methods	169
9.1	Introduction	169
9.2	Review of Discrete-Time Markov Chains	170
9.3	The Hastings-Metropolis Algorithm and Simulated Annealing	173
9.4	Descent States and Descent Graphs	175
9.5	Descent Trees and the Founder Tree Graph	177
9.6	The Descent Graph Markov Chain	181
9.7	Computing Location Scores	184
9.8	Finding a Legal Descent Graph	185
9.9	Haplotyping	186
9.10	Application to Episodic Ataxia	187
9.11	The Lander-Green-Kruglyak Algorithm	188
9.12	Genotyping Errors	191
9.13	Marker Sharing Statistics	192
9.14	Problems	195
9.15	References	199
10	Molecular Phylogeny	203
10.1	Introduction	203
10.2	Evolutionary Trees	203
10.3	Maximum Parsimony	205
10.4	Review of Continuous-Time Markov Chains	209
10.5	A Nucleotide Substitution Model	211
10.6	Maximum Likelihood Reconstruction	214
10.7	Origin of the Eukaryotes	215
10.8	Codon Models	218
10.9	Variation in the Rate of Evolution	219
10.10	Illustration of the Codon and Rate Models	221
10.11	Problems	223
10.12	References	228
11	Radiation Hybrid Mapping	231
11.1	Introduction	231
11.2	Models for Radiation Hybrids	232
11.3	Minimum Obligate Breaks Criterion	233

11.4	Maximum Likelihood Methods	236
11.5	Application to Haploid Data	238
11.6	Polyploid Radiation Hybrids	239
11.7	Maximum Likelihood Under Polyploidy	240
11.8	Obligate Breaks Under Polyploidy	244
11.9	Bayesian Methods	245
11.10	Application to Diploid Data	248
11.11	Problems	250
11.12	References	253
12	Models of Recombination	257
12.1	Introduction	257
12.2	Mather's Formula and Its Generalization	258
12.3	Count-Location Model	260
12.4	Stationary Renewal Models	261
12.5	Poisson-Skip Model	264
12.6	Chiasma Interference	270
12.7	Application to <i>Drosophila</i> Data	273
12.8	Problems	274
12.9	References	278
13	Sequence Analysis	281
13.1	Introduction	281
13.2	Pattern Matching	281
13.3	Alphabets, Strings, and Alignments	283
13.4	Minimum Distance Alignment	285
13.5	Parallel Processing and Memory Reduction	289
13.6	Maximum Similarity Alignment	290
13.7	Local Similarity Alignment	291
13.8	Multiple Sequence Comparisons	292
13.9	References	296
14	Poisson Approximation	299
14.1	Introduction	299
14.2	The Law of Rare Events	300
14.3	Poisson Approximation to the W_d Statistic	300
14.4	Construction of Somatic Cell Hybrid Panels	301
14.5	Biggest Marker Gap	304
14.6	Randomness of Restriction Sites	306
14.7	DNA Sequence Matching	308
14.8	Problems	311
14.9	References	315
15	Diffusion Processes	317
15.1	Introduction	317

15.2	Review of Diffusion Processes	317
15.3	Wright-Fisher Model	321
15.4	First Passage Time Problems	322
15.5	Process Moments	325
15.6	Equilibrium Distribution	326
15.7	Numerical Methods for Diffusion Processes	328
15.8	Numerical Methods for the Wright-Fisher Process	332
15.9	Specific Example for a Recessive Disease	333
15.10	Problems	336
15.11	References	338
Appendix A: Molecular Genetics in Brief		341
A.1	Genes and Chromosomes	341
A.2	From Gene to Protein	343
A.3	Manipulating DNA	345
A.4	Mapping Strategies	346
A.5	References	348
Appendix B: The Normal Distribution		351
B.1	Univariate Normal Random Variables	351
B.2	Multivariate Normal Random Vectors	352
B.3	References	354
Index		355