

*NSF-CBMS Regional Conference Series
in Probability and Statistics
Volume 5*

**MIXTURE MODELS:
THEORY, GEOMETRY
AND APPLICATIONS**

Bruce G. Lindsay
Pennsylvania State University

Institute of Mathematical Statistics
Hayward, California

American Statistical Association
Alexandria, Virginia

Conference Board of the Mathematical Sciences

*Regional Conference Series
in Probability and Statistics*

Supported by the
National Science Foundation

The production of the *NSF-CBMS Regional Conference Series in Probability and Statistics* is managed by the Institute of Mathematical Statistics: John Collins, IMS Managing Editor/Statistics; Patrick Kelly, IMS Production Editor; Miriam Gasko Donoho, IMS Treasurer; and Barbara J. Lindeman, IMS Business Manager.

Library of Congress Catalog Card Number: 94-75429

International Standard Book Number: 0-940600-32-3

Copyright © 1995 Institute of Mathematical Statistics

All rights reserved

Printed in the United States of America

Acknowledgments

I have many persons to thank for their part in the writing of this monograph. First and foremost, John Grego was responsible for the successful proposal that led to the lecture series. He also assumed the onerous task of being the local host. Without him, and his strong supporting cast from the University of South Carolina, neither the lectures nor the notes would have come into being. Of course, I am also indebted to the NSF and CBMS for their sponsorship of the conference and the lecture notes.

Much of the work reported in this set of notes is based on collaborative efforts of myself with others. I owe much to the many people who I have worked with over the years on these problems. I must thank especially Kathryn Roeder and Dankmar Böhning, as a considerable bit of what I report here is due to them. While engaged in this research, I have been provided research support by the National Science Foundation and the Humboldt Foundation of Germany.

In addition, my current graduate students have all contributed in a substantial way to the final product you have in your hands. In the last frantic days of finishing, they pulled together with me to help construct the many figures, tables, and the bibliography. For this, I thank Matilde Sanchez, Liwen Xi and Ramani Pilla for their efforts, and add special thanks to Yanling Zuo for the bibliographical work and Yong Lin for the many figures found in Chapter 4.

I am also indebted to those friends and family members who contributed, both by positive support and by accepting my extended working hours. Finally, the book is dedicated to my parents, George Speers Lindsay (1907–1994) and Geneva Elizabeth Lindsay (1907–), whose formative role in my life becomes more and more apparent to me. The best that I have, I got from them. An additional dedication is to Clifford Clogg (1949–1995), whose premature death was a great personal loss to me and a tragedy for our profession.

Contents

CHAPTER 1: The Wide Scope	1
1.1. The finite mixture problem	2
1.1.1. A simple example	3
1.1.2. More complicated applications	5
1.2. The latent (or mixing) distribution	6
1.2.1. The discrete latent distribution	6
1.2.2. The continuous latent distribution	8
1.3. Many more variations and names	9
1.3.1. Known component densities	9
1.3.2. Linear inverse problems	10
1.3.3. Random effects models	11
1.3.4. Repeated measures models	12
1.3.5. Latent class and latent trait models	12
1.3.6. Missing covariates and data	13
1.3.7. Random coefficient regression models	13
1.3.8. Empirical and hierarchical Bayes	13
1.3.9. Nuisance parameter models	14
1.3.10. Measurement error models	14
1.3.11. Deconvolution problems	15
1.3.12. Robustness and contamination models	16
1.3.13. Overdispersion and heterogeneity	16
1.3.14. Hidden mixture structures	17
1.3.15. Clustering: A second kind	17
1.4. Be aware of limitations	17
1.4.1. Robustness characteristics	18
1.4.2. Extracting signal from noise	18
1.5. The likelihoods	20
1.5.1. The multinomial likelihood	21
1.5.2. Partly classified data	21
1.6. The mixture NPMLE theorem	22

1.6.1.	The fundamental theorem	22
1.7.	Related nonparametric problems	24
1.7.1.	The MLE of an unknown distribution	25
1.7.2.	Accurate and error-prone measurements	25
1.7.3.	Monotone density problems	26
1.7.4.	Censoring problems	26
1.8.	Similar statistical problems	27
CHAPTER 2: Structural Features		28
2.1.	Descriptive features	28
2.1.1.	Some simple moment results	28
2.1.2.	Shape and modality	29
2.1.3.	Overdispersion and sign changes	30
2.1.4.	Log convexity of ratios	32
2.1.5.	Moments and sign changes	33
2.1.6.	Dispersion models	34
2.2.	Diagnostics for exponential families	35
2.2.1.	Empirical ratio plots	35
2.2.2.	Gradient function plots	35
2.2.3.	Comparing gradient and ratio plots	38
2.3.	Geometry of multinomial mixtures	38
2.3.1.	Known component densities	39
2.3.2.	Basic convex geometry	40
2.3.3.	Identifiability of weight parameters	41
2.3.4.	Carathéodory's theorem	42
2.4.	Exponential family geometry	43
2.4.1.	Identifiable functions	43
2.4.2.	Identifiability of weights, m fixed	44
2.4.3.	Full identifiability of m components	46
2.4.4.	Hyperplanes and convex sets	47
2.4.5.	Identifiability of weights and supports	48
2.4.6.	Related problems	50
2.5.	Moment representations	51
2.6.	Certain nested mixture models	53
2.7.	Concluding remark	55
CHAPTER 3: Parametric Models		56
3.1.	Discrete versus continuous	56
3.1.1.	Continuous models: The conjugate family	57
3.2.	Discrete latent distribution	59
3.2.1.	Known component distributions	59
3.2.2.	Unknown component parameters	60
3.3.	Properties of the m -component MLE	60
3.4.	EM algorithm	61
3.4.1.	A description of the EM	61

3.4.2.	The EM for finite mixtures	62
3.4.3.	Algorithmic theory	63
3.5.	Multimodality and starting values	65
CHAPTER 4: Testing for Latent Structure		68
4.1.	Dispersion score tests	69
4.1.1.	The dispersion score	69
4.1.2.	Neyman and Scott's $C(\alpha)$ test	70
4.1.3.	Dispersion test optimality	72
4.1.4.	Auxiliary parameters	73
4.2.	LRT for number of components	74
4.2.1.	The testing problem	74
4.2.2.	Historical perspective	75
4.2.3.	Initial observations	76
4.3.	Asymptotic multinomial geometry	77
4.3.1.	The dagger simplex	77
4.3.2.	Maximum likelihood and projections	79
4.3.3.	Type I likelihood ratio testing	81
4.4.	The type II likelihood ratio problem	82
4.4.1.	Parameter constraints	82
4.4.2.	Convex cones	82
4.4.3.	The z -coordinate system	83
4.4.4.	Projections onto convex cones	85
4.4.5.	The dual basis	86
4.4.6.	Sector decomposition and projection	87
4.4.7.	The type II LRT	88
4.4.8.	Applications	90
4.5.	Asymptotic mixture geometry	91
4.5.1.	Directional score functions	91
4.5.2.	The gradient scores	92
4.5.3.	Other directional scores	93
4.5.4.	Simple binomial examples	94
4.5.5.	The nonparametric LRT	95
4.5.6.	A nonconvex score cone	96
4.6.	The LRT on nonconvex cones	97
4.6.1.	Projections onto nonconvex cones	97
4.6.2.	Measuring distances	99
4.6.3.	Tubes and distributions	101
4.6.4.	Approximations for tubes	103
4.6.5.	The arc length problem	104
4.6.6.	Final comments	106
CHAPTER 5: Nonparametric Maximum Likelihood		108
5.1.	The optimization framework	108
5.1.1.	Reformulating the problem	108

5.1.2.	The feasible region	109
5.1.3.	The objective function	111
5.2.	Basic theorems	112
5.2.1.	Existence and support size	112
5.2.2.	Closed and bounded?	113
5.2.3.	Gradient characterization	115
5.2.4.	Properties of the support set	116
5.3.	Further implications of the theorems	117
5.3.1.	Duality theorem	117
5.3.2.	Gradient bounds on the likelihood	118
5.3.3.	Link to m -component methods	119
5.3.4.	Moment and support point properties	119
5.4.	Applications	120
5.4.1.	A binomial mixture	121
5.4.2.	Empirical CDF	122
5.4.3.	Known component distributions	122
5.4.4.	The multinomial case	123
5.5.	Uniqueness and support size results	123
5.5.1.	The strategy	123
5.5.2.	A geometric approach to Task 1	124
5.5.3.	A gradient function representation	125
CHAPTER 6: Computation: The NPMLE		127
6.1.	The convergence issue	127
6.2.	Using the EM	128
6.3.	Gradient-based algorithms	128
6.3.1.	Design algorithms	129
6.3.2.	Keeping track of the support points	129
6.3.3.	Vertex direction and exchange methods	129
6.3.4.	Intrasimplex direction method	130
6.3.5.	Monotonicity	130
6.3.6.	Using the dual problem	131
6.4.	Ideal stopping rules	131
6.4.1.	The ideal rule	131
6.4.2.	A gradient-based rule	132
6.4.3.	Combining grid and gradient	133
6.4.4.	Bounding the second order score	134
6.4.5.	A conservative method	134
6.4.6.	Remarks	135
CHAPTER 7: Extending the Method		136
7.1.	Problems with ratio structure	136
7.1.1.	Example: Size bias	136
7.1.2.	NPMLE with ratio structure	137
7.1.3.	Example: Size bias	137

7.1.4.	Example: Weibull competing risks	138
7.1.5.	Mixed hazards NPMLE	138
7.2.	NPMLE with constraints on Q	139
7.2.1.	Profile likelihood	139
7.2.2.	Linear constraints	140
7.2.3.	Examples with linear constraints	140
7.2.4.	The constrained NPMLE	141
7.2.5.	A simple algorithm	142
7.3.	Smooth estimates of Q	143
7.3.1.	Roughening by smoothing	143
7.3.2.	Deconvolution	144
7.3.3.	Series expansion	144
7.3.4.	A likelihood method	144
CHAPTER 8: The Semiparametric MLE		146
8.1.	An equivalence theorem	146
8.2.	Exponential response models	148
8.2.1.	Example: Rasch model	148
8.2.2.	Type I conditional models	149
8.2.3.	The two-item example	149
8.2.4.	Efficiency theorem	150
8.2.5.	Equivalence theorem for mixture MLE	150
8.3.	Errors-in-variables and case-control studies	151
8.3.1.	The joint sampling model	151
8.3.2.	The retrospective model	152
8.3.3.	Prentice and Pyke's equivalency	152
8.3.4.	The measurement error extension	153
8.3.5.	The extended equivalency result	153
8.4.	A mixture index of fit	154
8.4.1.	The problem	154
8.4.2.	The concept	154
8.4.3.	Application to the multinomial	155
8.4.4.	Maximum likelihood estimation	156
8.4.5.	Inference on the lack-of-fit index	156
Bibliography		159