

# Statistical Modeling with Spline Functions Methodology and Theory

Mark H. Hansen  
University of California at Los Angeles

Jianhua Z. Huang  
University of Pennsylvania

Charles Kooperberg  
Fred Hutchinson Cancer Research Center

Charles J. Stone  
University of California at Berkeley

Young K. Truong  
University of North Carolina at Chapel Hill

Copyright ©2006

by M. H. Hansen, J. Z. Huang, C. Kooperberg, C. J. Stone, and Y. K. Truong

January 5, 2006



# Contents

<b>1</b>	<b>Introduction</b>	<b>11</b>
1.1	Overview . . . . .	12
1.1.1	Why do we end up using splines? . . . . .	13
1.1.2	Broad outline of methods used . . . . .	16
1.1.3	Broad outline of theory . . . . .	19
1.1.4	Chapter by chapter overview . . . . .	21
1.2	Background . . . . .	22
1.2.1	Other smoothing methods . . . . .	22
1.2.2	Some history . . . . .	22
1.3	Software . . . . .	22
<b>2</b>	<b>Preliminaries</b>	<b>25</b>
2.1	What is a Spline? . . . . .	26
2.1.1	Polynomials . . . . .	27
2.1.2	Piecewise polynomials . . . . .	27
2.1.3	Splines . . . . .	28
2.1.4	B-splines: Definition . . . . .	30
2.1.5	Important properties of B-splines . . . . .	34
2.2	Function Approximation . . . . .	37
2.2.1	Properties of Polynomial Approximation . . . . .	37
2.2.2	Why Splines? . . . . .	38
2.2.3	Why B-splines? . . . . .	39
2.2.4	Distance from a function to a spline space . . . . .	39
2.3	Tensor products of splines . . . . .	41

4 Contents

2.3.1	Tensor products of linear spaces . . . . .	41
2.3.2	Tensor products of B-splines . . . . .	41
2.3.3	Approximation properties . . . . .	42
2.4	Concavity . . . . .	43
2.4.1	One-dimensional case . . . . .	43
2.4.2	Multi-dimensional case . . . . .	45
2.4.3	Checking concavity . . . . .	45
2.4.4	Existence and the uniqueness of the maximum . . . . .	46
2.5	Optimization . . . . .	48
2.5.1	Preview of the methods . . . . .	49
2.5.2	Gradient Methods – steepest ascent . . . . .	51
2.5.3	Newton–Raphson Method . . . . .	52
2.5.4	Quasi-Newton Method . . . . .	54
2.5.5	Conjugate Directions . . . . .	58
2.5.6	One-Dimensional Optimization — step length search . . . . .	61
2.5.7	Step-halving . . . . .	62
2.5.8	How to terminate an iteration . . . . .	63
2.6	B-splines with repeated knots . . . . .	63
2.6.1	Polynomial interpolation . . . . .	63
2.6.2	Divided difference – efficient way to evaluate the coefficients . . . . .	64
2.6.3	Divided differences with repetition . . . . .	64
2.6.4	Properties of the divided difference . . . . .	65
2.6.5	Computing the divided differences recursively . . . . .	67
2.6.6	B-Splines with repeated knots . . . . .	68
2.6.7	Basis: Curry–Schoenberg Theorem . . . . .	70
2.6.8	Examples . . . . .	72
2.6.9	Interpolation Errors via divided difference . . . . .	72
2.7	Continuity of divided differences . . . . .	74
2.8	Partial derivatives of B-splines with respect to knot locations . . . . .	75
<b>3</b>	<b>Linear Models</b> . . . . .	<b>101</b>
3.1	Examples . . . . .	102
3.1.1	Smoothing and extrinsic catastrophists . . . . .	102
3.1.2	Global warming and tree migration . . . . .	109
3.2	Regression modeling and approximation spaces . . . . .	114
3.2.1	Linear spaces and ordinary least squares . . . . .	115
3.2.2	The bias-variance tradeoff . . . . .	117
3.2.3	How smooth? Some simple model selection criteria . . . . .	120
3.3	Curve estimation . . . . .	124
3.3.1	From polynomials to splines . . . . .	124
3.3.2	Model error for fixed-knot splines . . . . .	132
3.3.3	Adaptive knot placement . . . . .	134
3.4	Multivariate models . . . . .	141
3.4.1	From multivariate polynomials to splines . . . . .	141

3.4.2	Model error and functional ANOVA . . . . .	146
3.4.3	Adaptation for multivariate splines . . . . .	148
3.5	A survey of multivariate spline methods . . . . .	154
3.6	Properties of spline estimates . . . . .	154
3.6.1	Knot spacing . . . . .	154
3.6.2	Boundary conditions . . . . .	158
3.6.3	Degrees of freedom associated with knot placement . . . . .	160
3.7	Representation and computation . . . . .	163
3.7.1	Selecting a basis . . . . .	163
3.7.2	Implementing stepwise addition . . . . .	166
3.7.3	Connection to smoothing splines . . . . .	169
3.8	A second look at the examples . . . . .	171
3.8.1	Assessing uncertainty in curve fitting . . . . .	171
3.8.2	Test set prediction error . . . . .	171
3.8.3	Multivariate responses . . . . .	173
3.9	Conclusion . . . . .	200
<b>4</b>	<b>Generalized Linear Models</b>	<b>203</b>
4.1	Applications . . . . .	204
4.1.1	Health effects of particulate matter . . . . .	204
4.1.2	Obesity and urban sprawl . . . . .	214
4.2	GLMs and approximation spaces . . . . .	220
4.2.1	Conditional Likelihood for a GLM . . . . .	220
4.2.2	Canonical linear regression and approximation spaces . . . . .	223
4.2.3	Link functions . . . . .	225
4.3	Estimation and adaptation . . . . .	226
4.3.1	Quadratic approximations . . . . .	226
4.3.2	Application to GLMs . . . . .	229
4.4	A general methodology . . . . .	232
<b>5</b>	<b>Polychotomous Regression and Multiple Classification</b>	<b>233</b>
5.1	An example . . . . .	233
5.1.1	The vowel data . . . . .	233
5.1.2	Background . . . . .	235
5.1.3	A Polyclass model for the vowel data . . . . .	237
5.2	The Polyclass methodology . . . . .	238
5.2.1	The Polyclass model . . . . .	238
5.2.2	Fitting Polyclass models . . . . .	241
5.2.3	Model selection . . . . .	242
5.3	Further analysis of the vowel data . . . . .	243
5.4	Applying Polyclass to large data sets . . . . .	246
5.4.1	The fruit data . . . . .	246
5.4.2	Analysis of cpu-time required for large data sets . . . . .	251
5.4.3	PolyMARS: A least squares approximation of the ad- dition process . . . . .	254

6 Contents

5.4.4	Fitting Polyclass models with large data sets and many basis functions . . . . .	257
5.4.5	Further analysis of the fruit data . . . . .	264
5.5	Technical details of the Polyclass algorithm . . . . .	271
5.5.1	Maximum number of basis functions . . . . .	271
5.5.2	Optimizing the location of a new knot. . . . .	273
5.6	Notes . . . . .	275
<b>6</b>	<b>Density Estimation</b>	<b>277</b>
6.1	An example . . . . .	277
6.1.1	The income data . . . . .	277
6.1.2	Background . . . . .	278
6.1.3	Log spline density estimation . . . . .	280
6.2	The Log spline methodology . . . . .	282
6.2.1	The Log spline model . . . . .	282
6.2.2	Basis functions . . . . .	285
6.2.3	Fitting Log spline models . . . . .	287
6.2.4	Knot selection . . . . .	287
6.3	How much to smooth: more examples . . . . .	288
6.4	Free knot splines and inference . . . . .	293
6.4.1	Free knot splines . . . . .	293
6.4.2	The bootstrap . . . . .	294
6.4.3	A comparison . . . . .	294
6.5	Censoring and truncation . . . . .	296
6.5.1	The Fyn diabetes data . . . . .	296
6.5.2	Implications for Log spline . . . . .	297
6.5.3	The Fyn diabetes data analyzed . . . . .	300
6.6	Multivariate density estimation . . . . .	301
6.7	Technical details . . . . .	303
6.7.1	Initial knot placement . . . . .	303
6.7.2	Stepwise addition for Log spline . . . . .	304
6.7.3	Numerical integration . . . . .	305
6.7.4	Constrained optimization . . . . .	305
6.8	Notes . . . . .	306
<b>7</b>	<b>Survival Analysis</b>	<b>309</b>
7.1	An example . . . . .	309
7.1.1	The bone marrow transplant data . . . . .	309
7.1.2	Background . . . . .	313
7.1.3	Linear models for the conditional log-hazard function . . . . .	314
7.1.4	A Hare model for the bone marrow transplant data . . . . .	316
7.2	The Hare methodology . . . . .	318
7.2.1	The Hare model . . . . .	318
7.2.2	Allowable spaces . . . . .	318
7.2.3	Model selection . . . . .	319

7.2.4	Fitting Hare models . . . . .	321
7.2.5	Inference . . . . .	322
7.3	Further analysis of the bone marrow transplant data . . . . .	323
7.3.1	Does a simpler model fit the data? . . . . .	323
7.3.2	Partially linear Hare models . . . . .	327
7.3.3	Proportional hazards regression . . . . .	329
7.4	Extensions . . . . .	333
7.4.1	The Colorado Plateau uranium miners data . . . . .	333
7.4.2	Time-dependent covariates . . . . .	334
7.4.3	Left truncation . . . . .	335
7.4.4	Analysis of the Colorado Plateau uranium miners data	336
7.4.5	Interval censored data . . . . .	339
7.4.6	Heft . . . . .	340
7.4.7	Severe censoring and the penalty parameter . . . . .	342
7.5	Technical details . . . . .	344
7.5.1	Numerical integration for Heft . . . . .	344
7.6	Notes . . . . .	345
<b>8</b>	<b>Estimation of the Spectral Distribution</b>	<b>349</b>
8.1	An example . . . . .	349
8.1.1	The network data . . . . .	349
8.1.2	Background . . . . .	350
8.1.3	Mixed Spectra . . . . .	353
8.1.4	An Lspec model for the network data . . . . .	355
8.2	The Lspec methodology . . . . .	356
8.2.1	The Lspec model . . . . .	356
8.2.2	Model selection for Lspec models . . . . .	358
8.3	Further analysis of the network data . . . . .	359
8.4	Extensions . . . . .	362
8.5	Notes . . . . .	362
<b>9</b>	<b>Multivariate Splines</b>	<b>365</b>
9.1	Preliminaries . . . . .	365
9.2	An application . . . . .	370
9.3	The methodology . . . . .	372
9.3.1	Bivariate spline spaces . . . . .	372
9.3.2	Maximum likelihood estimation. . . . .	378
9.3.3	A stepwise algorithm . . . . .	379
9.3.4	Stepwise addition . . . . .	382
9.3.5	Stepwise deletion . . . . .	386
9.4	The example revisited . . . . .	388
9.5	Simulation results . . . . .	390
9.6	Extensions . . . . .	393
<b>10</b>	<b>Alternate Optimization Methods</b>	<b>399</b>

10.1	Normal linear regression revisited . . . . .	400
10.1.1	Greedy methods and the $^{87}\delta\text{Sr}$ data . . . . .	402
10.1.2	Results from an exhaustive search . . . . .	405
10.2	Bayesian Formulations . . . . .	408
10.2.1	A single linear space . . . . .	409
10.2.2	Many spaces . . . . .	413
10.2.3	Computation . . . . .	417
10.2.4	Connection with model selection criteria . . . . .	420
10.2.5	Theoretical justification . . . . .	420
10.3	Normal linear regression . . . . .	420
10.3.1	Prior specification . . . . .	420
10.3.2	Computation . . . . .	424
10.4	Extended linear models . . . . .	425
10.4.1	Logspline density estimation . . . . .	425
10.4.2	Triogram regression . . . . .	425
10.4.3	ELM Prior specification . . . . .	425
10.4.4	Computation . . . . .	431
10.4.5	Logspline density estimation . . . . .	432
10.4.6	Triogram regression . . . . .	439
10.5	Other optimization methods . . . . .	447
10.5.1	Simulated annealing . . . . .	448
10.5.2	Genetic algorithms . . . . .	448
10.5.3	Gradient descent machines . . . . .	448
10.6	Combining models . . . . .	448
<b>11</b>	<b>Rates of Convergence in Extended Linear Modeling</b>	<b>449</b>
11.1	Theoretical Framework and Basic Results . . . . .	450
11.1.1	Extended Linear Models . . . . .	450
11.1.2	Consistency and Rates of Convergence . . . . .	454
11.1.3	ANOVA modeling . . . . .	458
11.2	The main result on rates of convergence . . . . .	463
11.2.1	Approximation Error. . . . .	464
11.2.2	Estimation Error. . . . .	465
11.3	Functional ANOVA . . . . .	467
11.3.1	Functional ANOVA Decompositions . . . . .	467
11.3.2	Construction of Model Space and Estimation Spaces using Functional ANOVA . . . . .	470
11.3.3	Rates of Convergence of ANOVA Components . . . . .	473
11.4	Verification of Technical Conditions . . . . .	476
11.4.1	Preliminaries . . . . .	476
11.4.2	Theoretical and Empirical Inner Products . . . . .	478
11.4.3	Generalized Regression . . . . .	481
11.4.4	Density Estimation . . . . .	486
11.4.5	Hazard Regression . . . . .	487
11.5	Notes . . . . .	490



<b>12</b>	<b>Extended Linear Modeling with Free Knot Splines</b>	<b>493</b>
12.1	Main Results . . . . .	494
12.1.1	Statement of Main Results . . . . .	494
12.1.2	Uniformity in Rates of Convergence . . . . .	496
12.1.3	Adaptive Parameter Selection . . . . .	497
12.2	Free Knot Splines and Their Tensor Products . . . . .	500
12.3	Verification of Technical Conditions . . . . .	502
12.3.1	Preliminary Lemmas . . . . .	503
12.3.2	Density Estimation . . . . .	505
12.3.3	Generalized Regression . . . . .	506
12.4	Proofs of Lemmas in Section 12.2 . . . . .	513

