

Pascal Massart

Concentration Inequalities and Model Selection

Ecole d'Eté de Probabilités
de Saint-Flour XXXIII - 2003

Editor: Jean Picard

 Springer

Contents

1	Introduction	1
1.1	Model Selection	1
1.1.1	Minimum Contrast Estimation	3
1.1.2	The Model Choice Paradigm	5
1.1.3	Model Selection via Penalization	7
1.2	Concentration Inequalities	10
1.2.1	The Gaussian Concentration Inequality	10
1.2.2	Suprema of Empirical Processes	11
1.2.3	The Entropy Method	12
2	Exponential and Information Inequalities	15
2.1	The Cramér-Chernoff Method	15
2.2	Sum of Independent Random Variables	21
2.2.1	Hoeffding's Inequality	21
2.2.2	Hoeffding's Inequality	23
2.2.3	Bernstein's Inequality	24
2.3	Basic Information Inequalities	27
2.3.1	Duality and Variational Formulas	27
2.3.2	Some Links Between the Moment Generating Function and Entropy	29
2.3.3	Fischer's Inequality	31
2.3.4	Rényi's Lemma	32
2.4	Entropy on Product Spaces	35
2.4.1	Marton's Coupling	37
2.4.2	Tensorization Inequality for Entropy	40
2.5	ϕ -Entropy	43
2.5.1	Necessary Condition for the Concavity of ϕ -Entropy	45
2.5.2	A Duality Formula for ϕ -Entropy	46
2.5.3	A Direct Proof of the Tensorization Inequality	49
2.5.4	Efron-Stein's Inequality	50

2	Gaussian Processes	53
2.1	Introduction and Basic Remarks	53
2.2	Concentration of the Gaussian Measure on $\mathbb{R}^N$	56
2.2.1	The Isoperimetric Nature of the Concentration Phenomenon	57
2.2.2	The Gaussian Isoperimetric Theorem	59
2.2.3	Gross' Logarithmic Sobolev Inequality	62
2.2.4	Application to Suprema of Gaussian Random Vectors	64
2.3	Comparison Theorems for Gaussian Random Vectors	66
2.3.1	Stein's Lemma	66
2.4	Metric Entropy and Gaussian Processes	70
2.4.1	Metric Entropy	70
2.4.2	The Chaining Argument	72
2.4.3	Continuity of Gaussian Processes	74
2.5	The Isocentral Process	77
2.5.1	Definition and First Properties	77
2.5.2	Continuity Sets with Examples	79
4	Gaussian Model Selection	83
4.1	Introduction	83
4.1.1	Examples of Gaussian Frameworks	83
4.1.2	Some Model Selection Problems	86
4.1.3	The Least Squares Procedure	87
4.2	Selecting Linear Models	88
4.2.1	A First Model Selection Theorem for Linear Models	89
4.2.2	Lower Bounds for the Penalty Term	94
4.2.3	Mixing Several Strategies	98
4.3	Adaptive Estimation in the Minimax Sense	101
4.3.1	Minimax Lower Bounds	102
4.3.2	Adaptive Properties of Penalized Estimators for Gaussian Sequences	115
4.3.3	Adaptation with Respect to Ellipsoids	116
4.3.4	Adaptation with Respect to Arbitrary ℓ_p -Bodies	117
4.3.5	A Special Strategy for Besov Bodies	122
4.4	A General Model Selection Theorem	125
4.4.1	Statement	125
4.4.2	Selecting Ellipsoids: A Link with Regularization	131
4.4.3	Selecting Nets Toward Adaptive Estimation for Arbitrary Compact Sets	139
4.5	Appendix: From Function Spaces to Sequence Spaces	144
5	Concentration Inequalities	147
5.1	Introduction	147
5.2	The Bounded Difference Inequality via Marton's Coupling	149
5.3	Concentration Inequalities via the Entropy Method	154

5.3.1	ϕ -Sobolev and Moment Inequalities	155
5.3.2	A Poissonian Inequality for Self-Bounding Functionals	157
5.3.3	ϕ -Sobolev Type Inequalities	162
5.3.4	From Efron-Stein to Exponential Inequalities	166
5.3.5	Moment Inequalities	172
6	Maximal Inequalities	183
6.1	Set-Indexed Empirical Processes	184
6.1.1	Random Vectors and Rademacher Processes	184
6.1.2	Vapnik-Chervomenskii Classes	186
6.1.3	L_1 -Entropy with Bracketing	190
6.2	Function-Indexed Empirical Processes	192
7	Density Estimation via Model Selection	201
7.1	Introduction and Notations	201
7.2	Penalized Least Squares Model Selection	202
7.2.1	The Nature of Penalized LSE	204
7.2.2	Model Selection for a Polynomial Collection of Models	211
7.2.3	Model Subset Selection Within a Localized Basis	219
7.3	Selecting the Best Histogram via Penalized Maximum Likelihood Estimation	225
7.3.1	Some Deeper Analysis of Chi-Square Statistics	228
7.3.2	A Model Selection Result	230
7.3.3	Choice of the Weights $\{x_m, m \in M\}$	236
7.3.4	Lower Bound for the Penalty Function	237
7.4	A General Model Selection Theorem for MLE	239
7.4.1	Local Entropy with Bracketing Conditions	239
7.4.2	Finite Dimensional Models	245
7.5	Adaptive Estimation in the Minimax Sense	251
7.5.1	Lower Bounds for the Minimax Risk	251
7.5.2	Adaptive Properties of Penalized LSE	263
7.5.3	Adaptive Properties of Penalized MLE	267
7.6	Appendix	273
7.6.1	Kullback-Leibler Information and Hellinger Distance	273
7.6.2	Moments of Log-Likelihood Ratios	276
7.6.3	An Exponential Bound for Log-Likelihood Ratios	277
8	Statistical Learning	279
8.1	Introduction	279
8.2	Model Selection in Statistical Learning	280
8.2.1	A Model Selection Theorem	281

8.3 A Refined Analysis for the Risk of an Empirical Risk Minimizer	287
8.3.1 The Main Theorem	288
8.3.2 Application to Bounded Regression	293
8.3.3 Application to Classification	296
8.4 A Refined Model Selection Theorem	301
8.4.1 Application to Bounded Regression	303
8.5 Advanced Model Selection Problems	307
8.5.1 Hold-Out as a Margin Adaptive Selection Procedure	308
8.5.2 Data-Driven Penalties	314
References	319
Index	325
List of Participants	331
List of Short Lectures	335