

# Contents

<i>Preface</i>	vii
1. Introduction	1
1.1 Kernels and their properties . . . . .	1
1.2 Use of MATLAB toolbox . . . . .	9
1.3 Complements . . . . .	11
2. Univariate kernel density estimation	15
2.1 Basic definition . . . . .	16
2.2 Statistical properties of the estimate . . . . .	19
2.3 Choosing the shape of the kernel . . . . .	23
2.4 Choosing the bandwidth . . . . .	24
2.4.1 Reference rule . . . . .	25
2.4.2 Maximal smoothing principle . . . . .	26
2.4.3 Cross-validation methods . . . . .	28
2.4.4 Plug-in method . . . . .	30
2.4.5 Iterative method . . . . .	31
2.5 Density derivative estimation . . . . .	33
2.5.1 Choosing the bandwidth . . . . .	36
2.6 Automatic procedure for simultaneous choice of the kernel, the bandwidth and the kernel order . . . . .	38
2.7 Boundary effects . . . . .	39
2.7.1 Generalized reflection method . . . . .	41
2.8 Simulations . . . . .	42
2.9 Application to real data . . . . .	42
2.9.1 Buffalo snowfall data . . . . .	42

2.9.2	Concentration of cholesterol . . . . .	43
2.10	Use of MATLAB toolbox . . . . .	45
2.10.1	Running the program . . . . .	45
2.10.2	Main figure . . . . .	46
2.10.3	Setting the parameters . . . . .	48
2.10.4	Eye-control method . . . . .	48
2.10.5	The final estimation . . . . .	50
2.11	Complements . . . . .	50
3.	Kernel estimation of a distribution function . . . . .	55
3.1	Basic definition . . . . .	55
3.2	Statistical properties of the estimate . . . . .	57
3.3	Choosing the bandwidth . . . . .	60
3.3.1	Cross-validation methods . . . . .	60
3.3.2	Maximal smoothing principle . . . . .	61
3.3.3	Plug-in methods . . . . .	61
3.3.4	Iterative method . . . . .	63
3.4	Boundary effects . . . . .	65
3.4.1	Generalized reflection method . . . . .	65
3.5	Application to data . . . . .	66
3.6	Simulations . . . . .	66
3.7	Application to real data . . . . .	66
3.7.1	Trout PCB data . . . . .	66
3.8	Use of MATLAB toolbox . . . . .	67
3.8.1	Running the program . . . . .	67
3.8.2	Main figure . . . . .	69
3.8.3	Setting the parameters . . . . .	69
3.8.4	Eye-control method . . . . .	70
3.8.5	The final estimation . . . . .	72
3.9	Complements . . . . .	73
4.	Kernel estimation and reliability assessment . . . . .	77
4.1	Basic Definition . . . . .	78
4.2	Estimation of ROC curves . . . . .	84
4.2.1	Binormal model . . . . .	84
4.2.2	Nonparametric estimates . . . . .	85
4.3	Summary indices based on the ROC curve . . . . .	88
4.3.1	Area under the ROC curve . . . . .	88

4.3.2	Maximum improvement of sensitivity over chance diagonal (MIS) . . . . .	92
4.4	Other indices of reliability assessment . . . . .	93
4.4.1	Cumulative Lift . . . . .	93
4.4.2	Lift Ratio . . . . .	95
4.4.3	Integrated Relative Lift . . . . .	96
4.4.4	Information Value . . . . .	97
4.4.5	KR index . . . . .	99
4.5	Application to real data . . . . .	101
4.5.1	Head trauma data . . . . .	101
4.5.2	Pancreatic cancer data . . . . .	103
4.5.3	Consumer loans data . . . . .	104
4.6	Use of MATLAB toolbox . . . . .	106
4.6.1	Running the program . . . . .	106
4.6.2	Start menu . . . . .	106
4.6.3	Simulation menu . . . . .	107
4.6.4	The final estimation . . . . .	109
5.	Kernel estimation of a hazard function . . . . .	111
5.1	Basic definition . . . . .	112
5.2	Statistical properties of the estimate . . . . .	114
5.3	Choosing the bandwidth . . . . .	115
5.3.1	Cross-validation method . . . . .	116
5.3.2	Maximum likelihood method . . . . .	116
5.3.3	Iterative method . . . . .	117
5.3.4	Acceptable bandwidths . . . . .	118
5.3.5	Points of the most rapid change . . . . .	120
5.4	Description of algorithm . . . . .	120
5.5	Application to real data . . . . .	121
5.5.1	Breast carcinoma data . . . . .	121
5.5.2	Cervix carcinoma data . . . . .	123
5.5.3	Chronic lymphocytic leukaemia . . . . .	124
5.5.4	Bone marrow transplant . . . . .	126
5.6	Use of MATLAB toolbox . . . . .	128
5.6.1	Running the program . . . . .	129
5.6.2	Main figure . . . . .	129
5.6.3	Setting the parameters . . . . .	130
5.6.4	Eye-control method . . . . .	131
5.6.5	The final estimation . . . . .	132



5.7	Complements . . . . .	134
6.	Kernel estimation of a regression function . . . . .	137
6.1	Basic definition . . . . .	138
6.2	Statistical properties of the estimate . . . . .	142
6.3	Choosing the bandwidth . . . . .	145
6.3.1	Mallows' method . . . . .	147
6.3.2	Cross-validation method . . . . .	148
6.3.3	Penalizing functions . . . . .	149
6.3.4	Methods based on Fourier transformation . . . . .	151
6.4	Estimation of the derivative of the regression function . . . . .	157
6.4.1	Choosing the bandwidth . . . . .	159
6.5	Automatic procedure for simultaneous choice of the kernel, the bandwidth and the kernel order . . . . .	159
6.6	Boundary effects . . . . .	160
6.7	Simulations . . . . .	161
6.7.1	Simulation 1 . . . . .	162
6.7.2	Simulation 2 . . . . .	163
6.8	Application to real data . . . . .	165
6.9	Use of MATLAB toolbox . . . . .	168
6.9.1	Running the program . . . . .	168
6.9.2	Main figure . . . . .	169
6.9.3	Setting the parameters . . . . .	170
6.9.4	Estimation of optimal parameters . . . . .	171
6.9.5	Eye-control method . . . . .	172
6.9.6	Comparing of methods for bandwidth selection . . . . .	173
6.9.7	The final estimation . . . . .	174
6.10	Complements . . . . .	175
7.	Multivariate kernel density estimation . . . . .	179
7.1	Basic definition . . . . .	179
7.2	Statistical properties of the estimate . . . . .	183
7.2.1	Exact MISE calculations . . . . .	184
7.3	Bandwidth matrix selection . . . . .	184
7.3.1	Cross-validation method . . . . .	186
7.3.2	Plug-in methods . . . . .	186
7.3.3	Reference rule . . . . .	187
7.3.4	Maximal smoothing principle . . . . .	187



7.3.5	Iterative method . . . . .	187
7.4	A special case for bandwidth selection . . . . .	190
7.4.1	M1 method . . . . .	191
7.4.2	M2 method . . . . .	192
7.5	Simulations . . . . .	194
7.6	Application to real data . . . . .	194
7.6.1	Old Faithful data . . . . .	196
7.6.2	UNICEF data . . . . .	196
7.6.3	Swiss bank notes . . . . .	198
7.6.4	Iris flower . . . . .	198
7.7	Use of MATLAB toolbox . . . . .	200
7.7.1	Running the program . . . . .	200
7.7.2	Main figure . . . . .	202
7.7.3	Setting the parameters . . . . .	203
7.7.4	The final estimation . . . . .	205
7.8	Complements . . . . .	205
<i>Notation</i>		207
<i>Bibliography</i>		213
<i>Index</i>		225