

Frontiers in Probability and the Statistical Sciences

Dimitris N. Politis

Model-Free Prediction and Regression

A Transformation-Based Approach to
Inference

 Springer

Frontiers in Probability and the Statistical Sciences

Editor-in Chief:

Somnath Datta
Department of Bioinformatics & Biostatistics
University of Louisville
Louisville, Kentucky, USA

Series Editors:

Frederi G. Viens
Department of Mathematics & Department of Statistics
Purdue University
West Lafayette, Indiana, USA

Dimitris N. Politis
Department of Mathematics
University of California, San Diego
La Jolla, California, USA

Hannu Oja
Department of Mathematics and Statistics
University of Turku
Turku, Finland

Michael Daniels
Section of Integrative Biology
Division of Statistics & Scientific Computation
University of Texas
Austin, Texas, USA

More information about this series at <http://www.springer.com/series/11957>

Dimitris N. Politis

Model-Free Prediction and Regression

A Transformation-Based Approach
to Inference

 Springer

Dimitris N. Politis
Department of Mathematics
University of California, San Diego
La Jolla, CA, USA

Frontiers in Probability and the Statistical Sciences
ISBN 978-3-319-21346-0 ISBN 978-3-319-21347-7 (eBook)
DOI 10.1007/978-3-319-21347-7

Library of Congress Control Number: 2015948372

Springer Cham Heidelberg New York Dordrecht London
© The Author 2015

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made.

Printed on acid-free paper

Springer International Publishing AG Switzerland is part of Springer Science+Business Media (www.springer.com)

*Für die zwei Violinen und die Viola des
Model-Freien Quartetts*

Preface

Prediction has been one of the earliest forms of statistical inference. The emphasis on parametric estimation and testing seems to only have occurred about 100 years ago; see Geisser (1993) for a historical overview. Indeed, parametric models served as a cornerstone for the foundation of Statistical Science in the beginning of the twentieth century by R.A. Fisher, K. Pearson, J. Neyman, E.S. Pearson, W.S. Gosset (also known as “Student”), etc.; their seminal developments resulted into a complete theory of statistics that could be practically implemented using the technology of the time, i.e., pen and paper (and slide-rule!).

While some models are inescapable, e.g., modeling a polling dataset as a sequence of independent Bernoulli random variables, others appear contrived, often invoked for the sole reason to make the mathematics work. As a prime example, the ubiquitous—and typically unjustified—assumption of Gaussian data permeates statistics textbooks to the day. Model criticism and diagnostics were developed as a practical way out; see Box (1976) for an account of the model-building process by one of the pioneers of applied statistics.

With the advent of widely accessible powerful computing in the late 1970s, computer-intensive methods such as resampling and cross-validation created a revolution in modern statistics. Using computers, statisticians became able to analyze big datasets for the first time, paving the way towards the “big data” era of the twenty-first century. But perhaps more important was the realization that the way we do the analysis could/should be changed as well, as practitioners were gradually freed from the limitations of parametric models. For instance, the great success of Efron’s (1979) bootstrap was in providing a complete theory for statistical inference under a nonparametric setting much like Maximum Likelihood Estimation had done half a century earlier under the restrictive parametric setup.

Nevertheless, there is a further step one may take, i.e., going beyond even nonparametric models, and this is the subject of the monograph at hand. To explain this, let us momentarily focus on regression, i.e., data that are pairs: $(Y_1, X_1), (Y_2, X_2), \dots, (Y_n, X_n)$, where Y_i is the measured response associated with a regressor value of X_i . There are several ways to model such a dataset; three main ones are listed below.

They all pertain to the standard, homoscedastic additive model:

$$Y_i = \mu(X_i) + \varepsilon_i \quad (1)$$

where the random variables ε_i are assumed to be independent, identically distributed (i.i.d.) from a distribution $F(\cdot)$ with mean zero.

- **Parametric model:** Both $\mu(\cdot)$ and $F(\cdot)$ belong to parametric families of functions, e.g., $\mu(x) = \beta_0 + \beta_1 x$ and $F(\cdot)$ is $N(0, \sigma^2)$.
- **Semiparametric model:** $\mu(\cdot)$ belongs to a parametric family, whereas $F(\cdot)$ does not; instead, it may be assumed that $F(\cdot)$ belongs to a smoothness class, etc.
- **Nonparametric model:** Neither $\mu(\cdot)$ nor $F(\cdot)$ can be assumed to belong to parametric families of functions.

Despite the nonparametric aspect of it, even the last option constitutes a model, and is thus rather restrictive. To see why, note that Eq. (1) with i.i.d. errors is not satisfied in many cases of interest even after allowing for heteroscedasticity of the errors. For example, consider the model $Y_i = G(X_i, \varepsilon_i)$, where the ε_i are i.i.d., and $G(\cdot, \cdot)$ is a nonlinear/non-additive function of two variables. It is for this reason, i.e., to render the data amenable to an additive model such as (1), that a multitude of transformations in regression have been proposed and studied over the years, e.g., Box-Cox, ACE, AVAS, etc.; see Linton et al. (1997) for a review.

Nevertheless, it is possible to shun Eq. (1) altogether and still conduct inference about a quantity of interest such as the conditional expectation function $E(Y|X = x)$. In contrast to nonparametric model (1), the following model-free assumption can be made:

- **Model-free regression:**

- **Random design.** The pairs $(Y_1, X_1), (Y_2, X_2), \dots, (Y_n, X_n)$ are i.i.d.
- **Deterministic design.** The variables X_1, \dots, X_n are deterministic, and the random variables Y_1, \dots, Y_n are independent with common conditional distribution, i.e., $P\{Y_j \leq y | X_j = x\} = D_x(y)$ not depending on j .

Inference for features, i.e., functionals, of the common conditional distribution $D_x(\cdot)$ is still possible under some regularity conditions, e.g., smoothness. Arguably, the most important such feature is the conditional mean $E(Y|X = x)$ that can be denoted $\mu(x)$. While $\mu(x)$ is crucial in the model (1) as the function explaining Y on the basis of $X = x$, it has a key function in model-free prediction as well: $\mu(x_f)$ is the mean squared error (MSE) optimal predictor of a future response Y_f associated with a regressor value x_f .

As will be shown in the sequel, it is possible to accomplish the goal of point and interval prediction of Y_f under the above model-free setup; this is achieved via the **Model-free Prediction Principle** described in Part I of the book. In so doing, the solution to interesting estimation problems is obtained as a by-product, e.g., inference on features of $D_x(\cdot)$; the prime example again is $\mu(x)$. Hence, a Model-free approach to frequentist statistical inference is possible, including prediction and confidence intervals.

In nonparametric statistics, it is common to try to develop some asymptotic theory for new methods developed. In addition to offering justification for the accuracy of these methods, asymptotics often provide insights on practical implementation, e.g., on the optimal choice of smoothing bandwidth, etc. All of the methods discussed/employed in the proposed Model-free approach to inference will be based on estimators that have favorable large-sample properties—such as consistency—under regularity conditions. Furthermore, asymptotic information on bandwidth rates, MSE decay rates, etc. will be given whenever available in the form of Facts or Claims together with suggestions on their proof and/or references. However, formal theorems and proofs were deemed beyond the scope of this monograph in order to better focus on the methodology, as well as keep the book’s length (and time of completion) under control. Perhaps more importantly, note that it is still unclear how to properly judge the quality of prediction intervals in an asymptotic setting; some preliminary ideas on this issue are given in Sects. 3.6.2 and 7.2.3, and the Rejoinder of Politis (2013).

Interestingly, the emphasis on prediction seems to be coming back full-circle in the twenty-first century with the recent boom in machine learning and data mining; see, e.g., the highly influential book on statistical learning by Hastie et al. (2009), and the recent monograph on predictive modeling by Kuhn and Johnson (2013). The Model-free prediction methods presented here are of a very different nature but share some similarities, e.g., in employing cross-validation and sample re-use for fine-tuning and optimization, and may thus complement well the popular model-based approaches to prediction and classification. Furthermore, ideas from statistical learning and model selection could eventually be incorporated in the Model-free framework as well, e.g., selecting a subset of regressors; this is the subject of ongoing work. Notably, the methods presented in this monograph are very computer-intensive; relevant R functions and software are given at: <http://www.math.ucsd.edu/~politis/DPsoftware.html>.

I would like to thank my colleagues in the Departments of Mathematics and Economics of UCSD for their support, and my Ph.D. students for bearing with some of the material. I have benefited immensely from suggestions and discussions with colleagues from all over the world; a very partial list includes: Ian Abramson, Ery Arias-Castro, Brendan Beare, Patrice Bertail, Ricardo Cao, Anirban DasGupta, Richard Davis, Brad Efron, Peter Hall, Xuming He, Nancy Heckman, Göran Kauermann, Claudia Klüppelberg, Piotr Kokoszka, Jens-Peter Kreiss, Michele La Rocca, Jacek Leskow, Tim McMurry, George Michailidis, Stathis Paparoditis, Mohsen Pourahmadi, Jeff Racine, Joe Romano, Dimitrios Thomakos, Florin Vaida, Slava Vasiliev, Philippe Vieu, and Michael Wolf. Further acknowledgements are given at the end of several chapters.

In closing, I would like to thank the Division of Mathematical Sciences of the National Science Foundation for their continuing support with multiple grants, the most recent ones being DMS-10-07513 and DMS 13-08319, and the John Simon Guggenheim Memorial Foundation for a 2011–2012 fellowship that helped me get started on this monograph. I would also like to thank Marc Strauss and Hannah

Bracken of Springer for a wonderful collaboration, and Somnath Datta and the Editorial Board of the Frontiers Series for hosting this project.

The impetus for putting together this monograph was to show how very different statistical problems can be approached afresh in a Model-free setting. Due to time and space limitations, I could only explicitly address a handful of areas of practical implementation, e.g., regression, autoregression, Markov processes, etc. It is my sincere hope that the monograph will incite the interest of readers to take another look at their favorite problem—either theoretical or applied—in this new light; the insights gained may be well worth it.

San Diego, CA, USA
Spring 2015

Dimitris N. Politis

Contents

Part I The Model-Free Prediction Principle

1	Prediction: Some Heuristic Notions	3
1.1	To Explain or to Predict?	3
1.2	Model-Based Prediction	6
1.3	Model-Free Prediction	9
2	The Model-Free Prediction Principle	13
2.1	Introduction	13
2.2	Model-Free Approach to Prediction	14
2.2.1	Motivation: The i.i.d. Case	14
2.2.2	The Model-Free Prediction Principle	14
2.3	Tools for Identifying a Transformation Towards i.i.d.–Ness	17
2.3.1	Model-Free Prediction as an Optimization Problem	17
2.3.2	Transformation into Gaussianity as a Stepping Stone	17
2.3.3	Existence of a Transformation Towards i.i.d.–Ness	19
2.3.4	A Simple Check of the Model-Free Prediction Principle	20
2.3.5	Model-Free Model-Fitting in Practice	21
2.4	Model-Free Predictive Distributions	22
2.4.1	Prediction Intervals and Asymptotic Validity	22
2.4.2	Predictive Roots and Model-Free Bootstrap	23
2.4.3	Limit Model-Free Resampling Algorithm	26
2.4.4	Prediction of Discrete Variables	28

Part II Independent Data: Regression

3	Model-Based Prediction in Regression	33
3.1	Model-Based Regression	33
3.2	Model-Based Prediction in Regression	36
3.3	A First Application of the Model-Free Prediction Principle	37
3.4	Model-Free/Model-Based Prediction	38

3.5	Model-Free/Model-Based Prediction Intervals	40
3.6	Pertinent Prediction Intervals	43
3.6.1	The i.i.d. Case	43
3.6.2	Asymptotic Pertinence of Bootstrap Prediction Intervals	45
3.7	Application to Linear Regression	47
3.7.1	Better Prediction Intervals in Linear Regression	47
3.7.2	Simulation: Prediction Intervals in Linear Regression	49
3.7.3	Model-Free vs. Least Squares: A Reconciliation	50
	Appendix 1: The Solution of Eq. (3.9)	53
	Appendix 2: L_1 vs. L_2 Cross-Validation	54
4	Model-Free Prediction in Regression	57
4.1	Introduction	57
4.2	Constructing the Transformation Towards i.i.d.–Ness	58
4.3	Model-Free Optimal Predictors	63
4.3.1	Model-Free and Limit Model-Free Optimal Predictors	63
4.3.2	Asymptotic Equivalence of Point Predictors	64
4.3.3	Cross-Validation for Model-Free Prediction	67
4.4	Model-Free Bootstrap	68
4.5	Predictive Model-Free Bootstrap	70
4.6	Model-Free Diagnostics	71
4.7	Simulations	72
4.7.1	When a Nonparametric Regression Model Is True	72
4.7.2	When a Nonparametric Regression Model Is Not True	76
	Acknowledgements	78
	Appendix 1: High-Dimensional and/or Functional Regressors	78
5	Model-Free vs. Model-Based Confidence Intervals	81
5.1	Introduction	81
5.2	Model-Based Confidence Intervals in Regression	82
5.3	Model-Free Confidence Intervals Without an Additive Model	85
5.4	Simulations	89
5.4.1	When a Nonparametric Regression Model Is True	89
5.4.2	When a Nonparametric Regression Model Is Not True	92
	Acknowledgements	93
Part III Dependent Data: Time Series		
6	Linear Time Series and Optimal Linear Prediction	97
6.1	Introduction	97
6.2	Optimal Linear Prediction	99
6.3	Linear Prediction Using the Complete Process History	100
6.3.1	Autocovariance Matrix Estimation	101
6.3.2	Data-Based Choice of the Banding Parameter l	102
6.4	Correcting a Matrix Towards Positive Definiteness	103

- 6.4.1 Eigenvalue Thresholding 103
- 6.4.2 Shrinkage of Problematic Eigenvalues 104
- 6.4.3 Shrinkage Towards White Noise 105
- 6.4.4 Shrinkage Towards a Second Order Estimate 106
- 6.5 Estimating the Length n Vector $\gamma(n)$ 106
- 6.6 Linear Prediction Based on the Model-Free Prediction Principle ... 107
 - 6.6.1 A First Idea: The Discrete Fourier Transform 107
 - 6.6.2 Whitening and the Model-Free Linear Predictor 108
 - 6.6.3 From Point Predictors to Prediction Intervals 111
- Acknowledgements 112

- 7 Model-Based Prediction in Autoregression 113**
 - 7.1 Introduction 113
 - 7.2 Prediction Intervals in AR Models: Laying the Foundation 114
 - 7.2.1 Forward and Backward Bootstrap for Prediction 114
 - 7.2.2 Prediction Intervals for Autoregressive Processes 116
 - 7.2.3 Pertinent Prediction Intervals in Model-Based Autoregression 117
 - 7.3 Bootstrap Prediction Intervals for Linear Autoregressions 119
 - 7.3.1 Forward Bootstrap with Fitted Residuals 120
 - 7.3.2 Forward Bootstrap with Predictive Residuals 121
 - 7.3.3 Forward Bootstrap Based on Studentized Roots 122
 - 7.3.4 Backward Bootstrap 123
 - 7.3.5 Generalized Bootstrap Prediction Intervals 126
 - 7.4 Alternative Approaches to Bootstrap Prediction Intervals for Linear Autoregressions 127
 - 7.5 Simulations: Linear AR Models 128
 - 7.5.1 Unconditional Coverage Level 128
 - 7.5.2 Conditional Coverage Level 132
 - 7.6 Bootstrap Prediction Intervals for Nonparametric Autoregression .. 134
 - 7.6.1 Nonparametric Autoregression with i.i.d Errors 134
 - 7.6.2 Nonparametric Autoregression with Heteroscedastic Errors 137
- Acknowledgements 139

- 8 Model-Free Inference for Markov Processes 141**
 - 8.1 Introduction 141
 - 8.2 Prediction and Bootstrap for Markov Processes 142
 - 8.2.1 Notation and Definitions 142
 - 8.2.2 Forward vs. Backward Bootstrap for Prediction Intervals .. 144
 - 8.3 Bootstrap Based on Estimates of Transition Density 145
 - 8.4 The Local Bootstrap for Markov Processes 148
 - 8.5 Hybrid Backward Markov Bootstrap for Nonparametric Autoregression 151
 - 8.6 Prediction Intervals for Markov Processes Based on the Model-Free Prediction Principle 153

8.6.1	Theoretical Transformation	154
8.6.2	Estimating the Transformation from Data	155
8.7	Finite-Sample Performance of Model-Free Prediction Intervals	159
8.8	Model-Free Confidence Intervals in Markov Processes	163
8.8.1	Finite-Sample Performance of Confidence Intervals	167
8.9	Discrete-Valued Markov Processes	170
8.9.1	Transition Densities and Local Bootstrap	171
8.9.2	Model-Free Bootstrap	174
	Acknowledgements	176
9	Predictive Inference for Locally Stationary Time Series	177
9.1	Introduction	177
9.2	Model-Based Inference	179
9.2.1	Theoretical Optimal Point Prediction	179
9.2.2	Trend Estimation and Practical Prediction	180
9.2.3	Model-Based Predictors and Prediction Intervals	183
9.3	Model-Free Inference	185
9.3.1	Constructing the Theoretical Transformation	186
9.3.2	Kernel Estimation of the “Uniformizing” Transformation	187
9.3.3	Local Linear Estimation of the “Uniformizing” Transformation	188
9.3.4	Estimation of the Whitening Transformation	189
9.3.5	Model-Free Point Predictors and Prediction Intervals	190
9.3.6	Special Case: Strictly Stationary Data	193
9.3.7	Local Stationarity in a Higher-Dimensional Marginal	194
	Acknowledgements	195
Part IV Case Study: Model-Free Volatility Prediction for Financial Time Series		
10	Model-Free vs. Model-Based Volatility Prediction	199
10.1	Introduction	199
10.2	Three Illustrative Datasets	202
10.3	Normalization and Variance-Stabilization	205
10.3.1	Definition of the NoVaS Transformation	205
10.3.2	Choosing the Parameters of NoVaS	206
10.3.3	Simple NoVaS Algorithm	207
10.3.4	Exponential NoVaS Algorithm	213
10.4	Model-Based Volatility Prediction	215
10.4.1	Some Basic Notions: L_1 vs. L_2	215
10.4.2	Do Financial Returns Have a Finite Fourth Moment?	219
10.5	Model-Free Volatility Prediction	221
10.5.1	Transformation Towards i.i.d.-Ness	222
10.5.2	Volatility Prediction Using NoVaS	224
10.5.3	Optimizing NoVaS for Volatility Prediction	226