



## UvA-DARE (Digital Academic Repository)

### Asymptotic results in nonparametric Bayesian function estimation

Kirichenko, A.

**Publication date**

2017

**Document Version**

Other version

**License**

Other

[Link to publication](#)

**Citation for published version (APA):**

Kirichenko, A. (2017). *Asymptotic results in nonparametric Bayesian function estimation*. [Thesis, fully internal, Universiteit van Amsterdam].

**General rights**

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

**Disclaimer/Complaints regulations**

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

# Asymptotic results in nonparametric Bayesian function estimation

Alisa Kirichenko

Bayesian nonparametric methods are widely used in practical applications. They have numerous attractive features such as their philosophical appeal, conceptual simplicity, and ability to easily incorporate prior knowledge about the model. However, putting a prior on a large function space might result in erroneous estimates or suboptimal performance. Therefore, it is essential to study Bayesian procedures to gain insight about which priors to use and how their tuning affects the performance of these procedures. One way to do it is to take an asymptotic approach and to analyse Bayesian methods from the frequentist point of view by assuming that there exists a true underlying function and studying how fast a particular Bayesian procedure captures the truth as the number of observations goes to infinity.

In this thesis we consider function estimation problems in two different statistical settings. First, we discuss regression and binary classification problems on large graphs, where the goal is to estimate a smooth function defined on the vertices of a graph. In the second setting we aim to estimate the intensity of an inhomogeneous Poisson process from a realised point pattern. For both problems we develop adaptive Bayesian procedures and study their asymptotic behaviour from the frequentist perspective. In particular, we derive contraction rates for our procedures and show that they are optimal in a minimax sense.

Asymptotic results in nonparametric Bayesian function estimation

Alisa Kirichenko

*Asymptotic results in nonparametric  
Bayesian function estimation*

PROEFSCHRIFT

ter verkrijging van de graad van doctor  
aan de Universiteit van Amsterdam  
op gezag van de Rector Magnificus, prof.dr.ir. K.L.J. Maex  
ten overstaan van een door het College voor Promoties ingestelde commissie,  
in het openbaar te verdedigen in de Agnietenkapel  
op 5 oktober 2017, te 12:00 uur

door

Alisa Kirichenko

geboren te Sint-Petersburg, Rusland

Dit proefschrift is goedgekeurd door de promotoren en de samenstelling van de promotiecommissie is als volgt:

|                |                          |                                 |
|----------------|--------------------------|---------------------------------|
| Promotor:      | prof.dr. J.H. van Zanten | Universiteit van Amsterdam      |
| Copromotor:    | dr. B.J.K. Kleijn        | Universiteit van Amsterdam      |
| Overige leden: | dr. A.J. van Es          | Universiteit van Amsterdam      |
|                | prof.dr. S. Ghoshal      | North Carolina State University |
|                | prof.dr. M.C.M. de Gunst | Vrije Universiteit Amsterdam    |
|                | prof.dr. M.R.H. Mandjes  | Universiteit van Amsterdam      |
|                | prof.dr. R. Nunez Queija | Universiteit van Amsterdam      |

Faculteit der Natuurwetenschappen, Wiskunde en Informatica

Research supported by the Netherlands Organisation for Scientific Research (NWO).

# Contents

|  |           |
|--|-----------|
| <b>Introduction</b>  | <b>5</b>  |
| <b>1 Bayesian nonparametric statistics</b>                             | <b>11</b> |
| 1.1 Statistical inference . . . . .                                    | 11        |
| 1.2 Nonparametric models . . . . .                                     | 12        |
| 1.3 Frequentist asymptotics . . . . .                                  | 13        |
| 1.4 Bayesian approach . . . . .  | 15        |
| 1.5 Bayesian asymptotics . . . . .                                     | 16        |
| 1.6 Adaptation . . . . .   | 19        |
| <b>2 The problem of function estimation on large graphs</b>            | <b>21</b> |
| 2.1 Introduction . . . . .   | 21        |
| 2.2 Asymptotic geometry assumption on graphs . . . . .                 | 23        |
| 2.3 Smoothness assumption . . . . .                                    | 29        |
| 2.4 Minimax rates for function estimation problems on graphs . . . . . | 30        |
| 2.5 Proofs . . . . .   | 33        |
| <b>3 Bayesian function estimation on large graphs</b>                  | <b>43</b> |
| 3.1 Introduction . . . . .   | 43        |
| 3.2 General results on prior concentration . . . . .                   | 45        |
| 3.3 Function estimation on graphs . . . . .                            | 46        |
| 3.4 Function estimation on graphs with missing observations . . . . .  | 48        |
| 3.5 Proofs . . . . .   | 55        |
| <b>4 Poisson intensity estimation</b>                                  | <b>71</b> |
| 4.1 Introduction . . . . .   | 71        |
| 4.2 The SGCP model . . . . .   | 73        |
| 4.3 Main result . . . . .  | 75        |
| 4.4 Proofs . . . . .   | 76        |
| <b>References</b>  | <b>81</b> |
| <b>Summary</b>   | <b>87</b> |

|                             |           |
|-----------------------------|-----------|
| <b>Samenvatting</b>         | <b>89</b> |
| <b>Author contributions</b> | <b>91</b> |
| <b>Acknowledgments</b>      | <b>93</b> |

# Introduction

This thesis is dedicated to the scientific field of mathematical statistics. In order to understand what kind of problems we are dealing with consider the following example. Figure 1 shows a component of the social network graph from the Framingham Heart Study conducted in 2000. The vertices of the graph represent people and the edges show ties among siblings, spouses, and friends. Vertex size is made proportional to people's body mass index (BMI), and the vertex is coloured yellow if the BMI of the corresponding person is above 30 (which indicates obesity) and green otherwise. Christakis and Fowler [2007] show that the chance of a person becoming obese increases if his graph-neighbours (siblings/spouses or friends) became obese. In real life one wouldn't have the value of BMI for every person, hence some of the vertices are going to be blank. Using statistical analysis we can try to fill in that data by predicting a person's BMI based on the geometry of the social network and the information about other people in the graph.

The problem above is an example of function estimation problem, where the function of interest is the function on the vertices of the graph describing the BMI index. Of course, there are plenty of other scenarios in which the target function can be defined on different domains and take values in different spaces. In order to be able to make predictions a statistician chooses a model that incorporates his knowledge and assumptions about the underlying function. A fundamental problem in statistics is to develop models based on a sample of observations so that further analyses can be carried out. If the set of functions in the model is not very large (can be indexed by a subset of  $\mathbb{R}^d$ ) such a model is called parametric. If the collection of functions considered is infinite dimensional the model is nonparametric. When there is enough information about the underlying function, parametric models are favourable. However, such models lack flexibility and are sometimes not robust enough in the sense that even slight contamination of the data by observations not following the particular parametric model might lead to incorrect conclusions. Nonparametric models are more flexible in that sense, and more robust against misspecification.

The study of nonparametric models started in the mid twentieth century. According to Prakasa Rao [1983], the first paper in the area of nonparametric function estimation is Rosenblatt [1956]. Since then nonparametric models have greatly evolved and gained tremendous popularity. A good introduction to the topic is the