university of groningen

University Medical Center Groningen

University of Groningen

Bayesian model determination in complex systems

Mohammadi, Abdolreza

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2015

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):
Mohammadi, A. (2015). *Bayesian model determination in complex systems*. [Thesis fully internal (DIV), University of Groningen]. University of Groningen.

Copyright
Other than for strictly personal use, 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), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment.

Take-down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): http://www.rug.nl/research/portal. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Download date: 24-01-2025

Summary

In this thesis, we address several problems related to modeling complex systems. The difficulty of modeling complex systems lies partly in their topology and how they form rather complex networks. From this perspective, our interest in networks (graphs) is part of a broader current of research on complex systems.

Graphical models provide powerful tools to model and make statistical inference regarding complex relationships among variables. In this context, Gaussian graphical models are commonly used, since inference in such models is often tractable. In Chapter 2, we introduce a novel Bayesian framework for Gaussian graphical model determination. We carry out the posterior inference by using an efficient sampling scheme which is a trans-dimensional MCMC approach based on birth-death process. In particular, we construct an efficient search algorithm which explores the graph space to detect the underlying graph with high accuracy. We cover the theory and computational details of the proposed method. We then apply the method to large-scale real applications from mammary gene expression studies to show its empirical usefulness.

The method that we propose in chapter 2 is limited only to the data that follows the Gaussianity assumption. In Chapter 3, we propose a Bayesian approach for graphical model determination based on a Gaussian copula approach that can deal with continuous, discrete, or mixed data. We embed a graph selection procedure inside a semi-parametric Gaussian copula. We implement our approach to discovering potential risk factors related to Dupuytren disease.

In chapter 4, we introduce an R package BDgraph which efficiently performs the Bayesian approaches that proposed in chapters 2 and 3. The core of the BDgraph package efficiently implemented in C++ to maximize computational speed.

The most promising statistical model that can be used for network modelling is the class of Exponential Random Graph Models (ERGMs). However, they are restricted to the models that regarded the network as given. In chapter 5, we develop a novel Bayesian statistical framework which combines the class of ERGMs with graphical models capable of modelling non-observed networks. Our proposed method greatly extends the scope of

the ERGMs to more applied research areas.

In chapter 6, we introduce a Bayesian framework in an M/G/1 queuing system with an optional second service. We estimate system parameters, predictive densities and some performance measures related to this queuing system such as stationary system size and waiting time.

Samenvatting

In dit proefschrift bekijken we een aantal problemen met betrekking tot het modelleren van complexe systemen. De moeilijkheid in het modelleren van complexe systemen ligt deels in hun topologie en deels in hoe ze tamelijk complexe netwerken vormen. Onze interesse in netwerken (grafen) is onderdeel van een bredere beweging in de richting van onderzoek naar complexe systemen.

Na de introductie, in hoofdstuk 2, introduceren we een nieuw Bayesiaans kader voor Gaussische grafische modelbepaling, namelijk een trans-dimensionale Markov Chain Monte Carlo (MCMC) aanpak op basis van een continue-tijd geboorte-dood-proces. We behandelen de theorie en computationele details van de voorgestelde methode. Vervolgens passen we de methode toe op borstklier genexpressie studies om zijn empirische nut tonen.

De werkwijze die we in hoofdstuk 2 voorstellen is beperkt tot de gegevens die de Gaussiaanse aanname volgt. In hoofdstuk 3 introduceren we een Bayesiaanse aanpak gebaseerd op de Gaussiaanse copula benadering die tegelijkertijd met binaire, gewone of continue variabelen kan werken. We verwerken een grafische selectieprocedure binnen een semi-parametrische Gaussische copula. Wij passen een posterieure inferentie toe door gebruik te maken van een efficiënt bemonsteringsschema: een trans-dimensionale MCMC aanpak op basis van een geboorte-dood-proces. Wij implementeren onze aanpak met als doel het ontdekken van potentiële risicofactoren bij de ziekte van Dupuytren.

In hoofdstuk 4, introduceren we een op R gebaseerd softwarepakket, BDgraph, dat efficiënt de Bayesiaanse benaderingen die in de hoofdstukken 2 en 3 zijn geïntroduceerd uitvoert. De kern van het BDgraph pakket is efficiënt geïmplementeerd in C++ om de rekensnelheid te bevorderen.

De meest veelbelovende statistische modellen die kunnen worden gebruikt voor netwerk-modellering zijn de exponentiële Random Graph Models (ERGMs). Ze zijn echter beperkt tot de modellen die het netwerk als gegeven beschouwen. In hoofdstuk 5 ontwikkelen we een nieuw Bayesiaans statistisch kader, welke de klasse van ERGMs combineert met grafische modellen die in staat zijn niet-waargenomen netwerken te modelleren. Onze voorgestelde methode vergroot het bereik van de ERGMs met meer toegepaste onderzoeks-

gebieden.

In hoofdstuk 6, stellen we een Bayesiaans kader voor in een M/G/1 wachtrijsysteem met een optionele tweede dienstverlening (service). Wij schatten parameters van het systeem, voorspellen dichtheden en een aantal prestatie-indicatoren met betrekking tot dit wachtrijsysteem zoals stationaire systeemgrootte en de wachttijd.

Acknowledgements

First of all I would like to acknowledge Prof. Ernst Wit, who gave me the opportunity to pursue my PhD project in his group. His continuous, trust support and guidance have been invaluable to the success of my PhD project. I hope we will have more fruitful collaborations in the future. I truly benefit from his ideas and his enthusiasm.

I wish to thank Prof. Nial Friel, Prof. Nicolai Petkov and Prof. Goeran Kauermann for serving as the reading committee.

Chapter 3 of this thesis is richer as a result of collaborations with Prof. Edwin van den Heuvel and Dr Fentaw Abegaz and Chapter 6 with Dr Mohammadreza Salehi-Rad. I would like to thanks them with their helpful collaborations.

I owe my loving thanks to my Pariya for her supports during my PhD study. Without her encouragement and understanding, it would have been impossible for me to finish my work. I owe my gratitude to my family for their unconditional supports. I am also grateful toward my family-in-low, specially my mother-in-low and my father-in-low, for their encouragements and their supports during my PhD study.

I would like to thank to my friends, Javier, Nynke, Ivan, Antonio, Vladimir and Lotsi who make my stay in Groningen so full of good memories. I am very lucky to meet Dick and Anneke whom I never forget their kindness and their continuous supports.

My special thanks to Dr. Frank Brokken whom I learned a lot in his great C++ course. We had very informative talks about C++ programming. I acknowledge Sourabh for his help in building BDgraph package and Erik for translating the summary and abstract into Dutch.

Abdolreza Mohammadi
Groningen
April, 2015