



FAKULTÄT FÜR
MATHEMATIK

Forschungsbericht 2017

Institut für Mathematische Stochastik

INSTITUT FÜR MATHEMATISCHE STOCHASTIK

Universitätsplatz 2, 39106 Magdeburg
Tel. +49 (0)391 67 58651, Fax +49 (0)391 67 11172
imst@ovgu.de

1. Leitung

Prof. Dr. Alexandra Carpentier
Prof. Dr. Claudia Kirch - geschäftsführende Leiterin (bis 31.03.2017)
Prof. Dr. Rainer Schwabe - geschäftsführender Leiter (ab 01.04.2017)

apl. Prof. Dr. Waltraud Kahle
Dr. Heiko Großmann

2. HochschullehrerInnen

Prof. Dr. Alexandra Carpentier
Prof. Dr. Claudia Kirch
Prof. Dr. Cornelia Pokalyuk (bis 31.03.2017 Vertretungsprofessur)
Prof. Dr. Rainer Schwabe

apl. Prof. Dr. Berthold Heiligers (extern)
apl. Prof. Dr. Waltraud Kahle

Professoren im Ruhestand:

Prof. Dr. Otfried Beyer
Prof. Dr. Gerd Christoph
Prof. Dr. Norbert Gaffke

3. Forschungsprofil

Mathematische Stochastik (Mathematische Statistik und Maschinelles Lernen): Prof. Dr. Alexandra Carpentier

- High or Infinite-Dimensional Adaptive Inference
- Uncertainty Quantification and Adaptive Confidence Sets
- Composite-Composite Testing Theory
- Sequential Sampling, Bandit Theory
- Optimisation of Computational Resources
- Inverse Problems and Compressed Sensing
- Applications in Statistical Problems (like regression/non-parametric estimation/matrix completion/extreme value theory/anomaly detection, etc)

Mathematische Stochastik (Stochastische Prozesse): Prof. Dr. Gerd Christoph; apl. Prof. Dr. Waltraud Kahle

- Asymptotische Methoden in der Stochastik
- Edgeworth und Cornish-Fisher Entwicklungen
- Statistik in Abnutzungsprozessen mit unvollständiger Reparatur
- Optimale unvollständige Instandhaltung in Abnutzungsprozessen
- Optimale Instandhaltung in allgemeinen Ausfall-Reparatur-Prozessen bei diskreten Lebensdauerverteilungen

Mathematische Stochastik (Mathematische Statistik): Prof. Dr. Norbert Gaffke

- Statistische Regressionsmodelle
- Experimental Design: Theorie und Algorithmen
- Tests und Konfidenzschranken
- Statistische Modellierung interdisziplinär

Mathematische Stochastik (Mathematische Statistik): Prof. Dr. Claudia Kirch

- Zeitreihenanalyse
- Changepoint-Analyse
- Resampling-Verfahren für Zeitreihen
- Statistische Methoden im Frequenzbereich
- Sequentielle Methoden
- Funktionale/Hochdimensionale Daten
- Bayessche semiparametrische Verfahren zur Zeitreihenanalyse

Mathematische Stochastik (Stochastische Prozesse und ihre Anwendungen): Prof. Dr. Cornelia Pokalyuk

- Stochastische Prozesse der Populationsgenetik, insbesondere Darstellung und Analyse dieser mittels Graphen, z.B. durch den ancestralen Selektionsgraphen
- Untersuchung des Effekts von Selektion auf Genealogien
- Modellierung der Evolution des Zytomegalievirus

Mathematische Stochastik (Statistik und ihre Anwendungen): Prof. Dr. Rainer Schwabe; Dr. Heiko Großmann

- Planung und Auswertung statistischer Experimente
 - Conjoint-Analyse (Psychologie, Marktforschung)
 - Intelligenzforschung (Psychologie)
 - Populationspharmakokinetik (Arzneimittelforschung)
 - Adaptive und gruppensequenzielle Verfahren
 - Diagnostische Studien mit räumlicher Datenstruktur und zeitlicher Verlaufskontrolle (Perimetrie in der Augenheilkunde)
 - Klinische Dosisfindungsstudien
 - Statistik in industriellen Anwendungen
 - Multivariate Äquivalenz und Nichtunterlegenheit
 - Multizentrische Studien
 - Lineare, verallgemeinert lineare und nichtlineare gemischte Modelle

4. Serviceangebot

Beratung und Unterstützung bei allen statistischen Fragestellungen

Das Institut für Mathematische Stochastik bietet Beratung zur Planung und statistischen Auswertung von Experimenten an, insbesondere:

- zur Unterstützung von Abschlussarbeiten bei der Konzeption und Durchführung von Studien

- bei der Stichproben-/ Versuchsplanung, Datengewinnung und Sicherstellung der Datenqualität
- bei der Auswahl und Anwendung geeigneter Analysemethoden
- bei der Interpretation und Präsentation der Untersuchungsergebnisse

Dieses Angebot richtet sich an ...

- Studierende und Promovierende der Otto-von-Guericke-Universität Magdeburg (OVGU)
- Ausgenommen sind Personen, die mit dem Fachbereich Medizin assoziiert sind. (Das Universitätsklinikum bietet über das Institut für Biometrie und Medizinische Informatik Statistikberatungen an.)

<http://www.statistik.ovgu.de/Statistische+Beratung.html>

5. Kooperationen

- Dr. Celine Duval, Universite Paris Descartes, France
- Dr. Debarghya Ghoshdastidar, Universitaet Tubingen, Germany
- Dr. Etienne Roquain, Universite Paris VI, France
- Dr. Frenkel, Beer Sheva, Israel Sami Shamoon College of Engineering, Israel
- Dr. Fritjof Freise, TU Dortmund
- Dr. Maureen Cerc, INRIA Sophia Antipolis, France
- Dr. Michal Valko, INRIA Lille Nord Europe, France
- Dr. Nicolas Verzelen, INRA Montpellier, France
- Dr. Olga Klopp, ESSEC Business School, France
- Dr. Sylvain Delattre, Universite Paris VI, France
- Juliette Achdou, HEC and Telecom Paris, France
- Oleksandr Zadorozhnyi, Universitaet Potsdam, Germany
- Priv.-Doz. Dr. Ekkehard Glimm, Novartis Pharma AG, Basel
- Priv.-Doz. Dr. Norbert Benda, BfArM, Bonn
- Prof. Dr. Andreas Greven, Universität Erlangen-Nürnberg
- Prof. Dr. Anton Wakolbinger, Goethe-Universität Frankfurt
- Prof. Dr. Arlene K.H. Kim, Sungshin Women's University, Korea
- Prof. Dr. Bharath Sriperumbudur, Penn State University, USA
- Prof. Dr. Christian Paroissin, Universität Pau, Frankreich
- Prof. Dr. Gilles Blanchard, Universitaet Potsdam, Germany
- Prof. Dr. Haeran Cho, University of Bristol
- Prof. Dr. Heinz Holling, Westfälische Wilhelms-Universität Münster
- Prof. Dr. Hernando Ombao, University of California, Irvine
- Prof. Dr. Idris Eckley, Lancaster University
- Prof. Dr. John Aston, University of Cambridge
- Prof. Dr. Laura Gibson, University of Massachusetts Medical School, USA
- Prof. Dr. Luc Pronzato, Université de Nice, CNRS-13R
- Prof. Dr. Martin Wendler, Universität Greifswald
- Prof. Dr. Peter Pfaffelhuber, Albert-Ludwigs-Universität Freiburg
- Prof. Dr. Radoslav Harman, Comenius-Universität Bratislava
- Prof. Dr. Renate Meyer, University of Auckland, New Zealand
- Prof. Dr. Richard Nickl, University of Cambridge, UK
- Prof. Dr. Samory Kpotufe, Princeton University, USA
- Prof. Dr. Sophie Mercier, Universität Pau, Frankreich
- Prof. Dr. Timothy Kowalik, University of Massachusetts Medical School, USA
- Prof. Dr. Ulrike von Luxburg, Universitaet Tubingen, Germany
- Prof. V. Ulyanov, Moskauer Staatliche Universität

6. Forschungsprojekte

Projektleitung: Prof. Dr. Claudia Kirch
Kooperationen: Silke Weber, KIT
Förderer: Haushalt; 01.01.2017 - 31.12.2018

A novel change point approach for the detection of gas emission sources using remotely contained concentration data

We consider a multivariate epidemic mean change model with dependent errors where the mean changes in each dimension at a time point t_1 and returns back at a time point t_2 . These two change points can differ in each dimension but each change point has a functional connection depending on the dimension. To find t_1 and t_2 in each dimension we developed an asymptotic testing procedure. Therefore we use two different types of test statistics, the multivariate test statistic and the projection test statistic where we transform the multivariate data into univariate data.

For simulations we consider the situation that we search for a gas emission source in a big area. So we assume that we have data from an air plane which measures the gas concentration in the air.

Our testing procedure helps us to decide whether there exists a gas emission source in this area. If there is a source we want to estimate its coordinates as near as possible to the real location. Therefore we assume that outside of the gas plume the data have a constant mean and inside the plume the mean increases to a higher level. With the knowledge of the form of the gas plume and the gas concentration with the corresponding coordinates of the measurement points we can draw conclusions for the location of the gas emission source.

Additionally we use our method for real data to locate a landfill.

This is joint work with Idris Eckley and Claudia Kirch.

Projektleitung: Prof. Dr. Claudia Kirch
Projektbearbeitung: Kerstin Reckrühm, M.Sc.
Förderer: Haushalt; 01.04.2015 - 31.12.2018

Die Detektion multipler Strukturbrüche basierend auf dem MOSUM-Verfahren

Es existieren zwei grundlegende Verfahren zur Erkennung multipler Strukturbrüche in Zeitreihen im klassischen Modell der Erwartungswertänderung, die binäre Segmentierung und das MOSUM-Verfahren. Das Segmentierungsverfahren ist eine iterative Methode, die ausnutzt, dass Tests für Ein-Change-point-Alternativen weiterhin Macht im Fall von multiplen Änderungen besitzen. Die zweite Methode hingegen basiert auf Statistiken, die gleitende Summen verwenden. Ein Vorteil des MOSUM-Verfahrens besteht darin, dass das Gesamtsignifikanzniveau kontrolliert werden kann. Tests und statistische Eigenschaften von Change-point Schätzern, die auf derartige Statistiken gleitender Summen basieren, wurden von Kirch und Muhsal (2015+) im klassischen Erwartungswert-Modell detailliert untersucht. Diese Resultate sollen nun für verschiedene Change-point Situationen verallgemeinert werden. Durch die Verwendung von MOSUM-Statistiken basierend auf Schätzfunktionen können Modelle verschiedener Parameteränderungen in ein Erwartungswert-Modell der Schätzfunktion transformiert werden. Dazu muss lediglich der globale Schätzer ermittelt werden, was einen großen Vorteil in Bezug auf den Rechenaufwand darstellt. Wir konstruieren eine entsprechende Teststatistik und analysieren ihr asymptotisches Verhalten unter der Nullhypothese und Alternativen. Weiterhin werden die zugehörigen Change-point Schätzer hinsichtlich ihrer Konsistenzeigenschaften näher untersucht.

Das Hauptproblem des MOSUM-Verfahrens besteht darin, dass die Güte dieser Methode im Wesentlichen von der Wahl der Bandbreite G abhängt. Dies erweist sich insbesondere dann als sehr problematisch, wenn die Abstände zwischen den Change-points stark variieren. So eignen sich große Bandbreiten zur Detektion kleiner Änderungen und kleine Bandbreiten zur Erkennung großer Änderungen. Eine Lösungsmöglichkeit wurde kürzlich im Zusammenhang mit Änderungen in Punktprozessen von Messer et al. (2014) vorgeschlagen. Ein Multiskalenverfahren basierend auf MOSUM-Statistiken soll dementsprechend konstruiert und untersucht werden. Da es für dieses Verfahren bisher noch keinerlei theoretische Untersuchungen gibt, wollen wir hier zunächst bei dem einfachen Erwartungswert-Modell bleiben.

Projektleitung: Prof. Dr. Claudia Kirch
Projektbearbeitung: Alexander Meier, M.Sc.
Förderer: Deutsche Forschungsgemeinschaft (DFG); 01.01.2016 - 31.12.2018
Ein nichtparametrischer Ansatz zur Bayesschen Zeitreihenanalyse

Die Bayessche Zeitreihenanalyse hat in den letzten Jahren große Beachtung in der Fachliteratur gefunden. Dennoch existieren nur wenige nichtparametrische Ansätze, insbesondere für multivariate Zeitreihen. Die meisten der etablierten Methoden basieren auf der Whittleschen Likelihood, welche die zweite Ordnungsstruktur einer stationären Zeitreihe durch die Spektraldichtematrix einbindet.

Letztere wird oft auf Ebene der Cholesky-Zerlegung modelliert, um positive Definitheit zu garantieren. Jedoch sind wichtige asymptotische Gütekriterien wie Konsistenz und ggf. Kontraktionsraten unter solchen a priori Verteilungen nicht bekannt.

Eine andere Idee besteht darin, die Spektraldichtematrix direkt mittels zufälligen Maßen zu modellieren. Wir verwenden eine Mischung von Beta-Dichten als Modellierungsansatz, mit matrixwertigen Mischungsgewichten, die durch ein vollständig zufälliges matrixwertiges Maß induziert werden.

Zur Konstruktion des vollständig zufälliges matrixwertiges Maßes verwenden wir eine unendlich teilbare matrixwertige Verallgemeinerung der Gamma-Verteilung.

Mittels optimierter Markov-Ketten-Monte-Carlo-Verfahren kann die Bayessche a posteriori Verteilung numerisch approximiert werden.

Während dieses Verfahren gute Ergebnisse auf simulierten und echten Datensätzen liefert, ist es uns auch möglich, Konsistenz nachzuweisen und Kontraktionsraten herzuleiten.

Projektleitung: Prof. Dr. Claudia Kirch

Projektbearbeitung: Christina Stöhr, Dipl.-Wirt.-Math.

Förderer: Haushalt; 01.01.2017 - 31.12.2018

Erkennung von Änderungen in der Kovarianzstruktur funktionaler Zeitreihen mit Blick auf Daten der funktionellen Magnetresonanztomographie

Funktionelle Magnetresonanztomographie (fMRT) wird eingesetzt, um Interaktionen zwischen verschiedenen Gehirnregionen zu analysieren. Die entsprechenden Aufnahmen können bezüglich gewisser Aufgaben oder Reize, aber auch in einem Ruhezustand gemacht werden. Letzteres wird durchgeführt, um Gehirnaktivitäten unabhängig von externen Einflussfaktoren analysieren zu können. Der Patient erhält dazu die Anweisung, während der Aufnahme an nichts zu denken und nicht einzuschlafen. Dies kann jedoch oft nicht eingehalten werden, sodass anschließende Analysen, welche auf der Stationaritätsannahme beruhen, verfälscht werden. Dazu haben J.A.D. Aston und C.Kirch 2012 bereits ein entsprechendes Testverfahren eingeführt, um Mittelwertänderungen in fMRT-Daten erkennen und korrigieren zu können. Wir sind nun daran interessiert, auch Änderungen in der Kovarianzstruktur zu detektieren. fMRT-Daten können als funktionelle Zeitreihen modelliert werden, wobei wir annehmen, dass die einzelnen Beobachtungen der Zeitreihe, d.h. das jeweils aufgenommene Bild zu einem bestimmten Zeitpunkt, als Funktion dargestellt werden können. Durch die Anwendung bestimmter Techniken zur Dimensionsreduktion erhalten wir eine multivariate Zeitreihe. Unter gewissen Voraussetzungen impliziert eine Änderung in der Kovarianzstruktur der funktionellen Zeitreihe auch eine Änderung in der Kovarianzstruktur der multivariaten Zeitreihe, welche sich aus der Dimensionsreduktion ergibt. Wir führen ein asymptotisches Testverfahren ein, um solche Änderungen zu erkennen. Allerdings erfordert dies die Schätzung der Langzeitkovarianz, welche insbesondere für hohe Dimensionen im Vergleich zum Stichprobenumfang statistisch sehr instabil ist. Daher verwenden wir einen misspezifizierten Schätzer, welcher nur die Diagonalelemente, d.h. die Langzeitvarianzen schätzt. Die asymptotische Verteilung der Teststatistik hängt dann jedoch von unbekanntem Parametern ab, sodass Resampling-Verfahren eingesetzt werden müssen, um den kritischen Wert zu erhalten.

Projektleitung: Prof. Dr. Claudia Kirch

Kooperationen: Silke Weber, KIT

Förderer: Haushalt; 01.01.2016 - 31.12.2018

Modified sequential change point procedures based on estimating functions

A large class of sequential change point tests are based on estimating functions where estimation is computationally efficient as (possibly numeric) optimization is restricted to an initial estimation. This includes examples as diverse as mean changes, linear or non-linear autoregressive and binary models. While the standard cumulative-sum-detector (CUSUM) has recently been considered in this general setup, we consider several modifications that have faster detection rates

in particular if changes do occur late in the monitoring period. More precisely, we use three different types of detector statistics based on partial sums of a monitoring function, namely the modified moving-sum-statistic (mMOSUM), Page cumulative-sum-statistic (Page-CUSUM) and the standard moving-sum-statistic (MOSUM). The statistics only differ in the number of observations included in the partial sum. The mMOSUM uses a bandwidth parameter which multiplicatively scales the lower bound of the moving sum. The MOSUM uses a constant bandwidth parameter, while Page-CUSUM chooses the maximum over all possible lower bounds for the partial sums. So far, the first two schemes have only been studied in a linear model, the MOSUM only for a mean change. We develop the asymptotics under the null hypothesis and alternatives under mild regularity conditions for each test statistic, which include the existing theory but also many new examples. In a simulation study we compare all four types of test procedures in terms of their size, power and run length. Additionally we illustrate their behavior by applications to exchange rate data as well as the Boston homicide data.

Projektleitung: Prof. Dr. Claudia Kirch

Förderer: Fördergeber - Sonstige; 01.01.2016 - 31.12.2017

Multiple Strukturbruch-Verfahren basierend auf MOSUM-Statistiken

In diesem Projekt wollen wir Verfahren zur Schätzung multipler Changepoints basierend auf MOSUM-Statistiken auf verschiedene Arten von Strukturänderungen verallgemeinern sowie robuste Methoden für hochdimensionale Daten entwickeln. Es gibt dazu eine Vielzahl möglicher Ansätze, die wir nicht nur durch asymptotische Betrachtungen vergleichen wollen, sondern auch durch umfangreiche Simulationsstudien, wobei auch Resampling-Verfahren berücksichtigt werden sollen.

Das Hauptproblem bei obigem Vorgehen besteht darin, dass die Güte des Verfahrens im Wesentlichen von der Wahl der Bandbreite "G" abhängt. Grundsätzlich sollte diese so groß wie möglich gewählt werden, ohne dass sich in einem Fenster der Länge "2G" zwei Changepoints befinden, grob gesprochen also halb so groß wie der minimale Abstand zwischen zwei Changepoints. Dies ist offensichtlich problematisch, wenn die Abstände zwischen Changepoints stark variieren. Grundsätzlich können Changepoint-Tests große Änderungen bereits in kleinen Stichprobenumfängen zuverlässig erkennen (und liefern somit auch gute Schätzer für die Changepoints selbst), benötigen aber für kleine Änderungen längere Stichproben. Im Falle der MOSUM-Statistiken bedeutet dies, dass man mit einer großen Bandbreite kleine Änderungen erkennen kann, die aber dafür von langen Segmenten gesäumt sind. Umgekehrt kann man große Änderungen, die nur von kurzen Segmenten gesäumt sind, sehr gut mit kleinen Bandbreiten erkennen. Asymmetrische Bandbreiten sind nötig, um kleine Änderungen zu erkennen, die nur auf einer Seite von einem langen Segment umgeben sind.

Wir wollen verschiedene Methoden vergleichen, um die Information aus verschiedenen Bandbreiten (auch asymmetrischen) zu poolen.

Projektleitung: Prof. Dr. Claudia Kirch

Projektbearbeitung: Alexander Meier, M.Sc.

Kooperationen: Prof. Dr. Renate Meyer, University of Auckland, New Zealand

Förderer: Deutsche Forschungsgemeinschaft (DFG); 01.01.2016 - 31.12.2018

Neue semiparametrische Ansätze zur Bayesschen Zeitreihenanalyse

- Neue semiparametrische Ansätze zur Bayesschen Zeitreihenanalyse (mit R. Meyer, M. Edwards, A. Meier)
- In den letzten Jahren haben nichtparametrische Bayessche Verfahren stark an Aufmerksamkeit und Bedeutung gewonnen. Dennoch sind nur wenige Ansätze für die Zeitreihenanalyse entwickelt worden. Die Schwierigkeit besteht darin, dass Bayessche statistische Verfahren der vollständigen Spezifikation einer Likelihood-Funktion bedürfen, was einer nichtparametrischen Herangehensweise zunächst entgegen steht. Mehrere Autoren haben das Problem mit Hilfe der Whittle-Likelihood gelöst, einer Approximation der wahren Likelihood, die von der Spektraldichte als der wichtigsten nichtparametrischen Kenngröße von Zeitreihen abhängt. Selbst bei nicht-Gaußschen Zeitreihen, die nicht vollständig durch ihre erste und zweite Ordnungsstruktur gekennzeichnet

sind, erhält man so in vielen Situationen asymptotisch gültige statistische Verfahren jedoch häufig auf Kosten der Effizienz. Parametrische Modelle, auf der anderen Seite, sind im korrekt spezifizierten Fall wesentlich mächtiger, brechen jedoch bei starker Misspezifikation vollständig zusammen. Moderne nichtparametrische Bootstrap-Verfahren für Zeitreihen setzen sich mit den gleichen Schwierigkeiten auseinander und verwenden implizit ebenfalls Approximationen der wahren Likelihood-Funktion. In diesem Projekt werden wir Approximationen moderner Resampling-Verfahren für Zeitreihen für die Bayessche nichtparametrische Analyse verwenden. Hierzu kombinieren wir eine erste parametrische Modellierung der Zeitreihe im Zeitbereich mit einer Bayesschen Korrektur im Frequenzbereich, die auf nichtparametrischen Priors basiert. Dieser vollständig neue semiparametrische Ansatz zur Bayesschen Zeitreihenanalyse geht weit über bestehende Ansätze hinaus.

Projektleitung: Prof. Dr. Claudia Kirch

Projektbearbeitung: Christina Stöhr, Dipl.-Wirt.-Math.

Förderer: Haushalt; 01.04.2015 - 31.03.2018

Robuste Verfahren in der sequentiellen Change-Point Analyse basierend auf U-Statistiken

Change-Point Analyse dient dazu, Strukturbrüche in zeitlich geordneten Datensätzen zu erkennen. Hierzu gibt es die folgenden zwei Ansätze: Einerseits verwendet das a-posteriori (offline) Verfahren den vollständig beobachteten Datensatz zum Testen auf einen Change-Point. Andererseits gibt es das sequentielle (online) Verfahren, bei dem eine sogenannte Testperiode bekannt ist und darauf aufbauend neue Daten gesammelt werden. Nach jeder neuen Beobachtung wird erneut ein Test auf eine Strukturveränderung vorgenommen.

Die Entscheidung, ob eine Veränderung eingetreten ist oder nicht, wird mittels einer Teststatistik gefällt. Ist diese größer als ein kritischer Wert, entdeckt der a-posteriori Test eine Veränderung. Im Falle des sequentiellen Verfahrens wird statt eines kritischen Wertes eine kritische Kurve verwendet. Sobald die Teststatistik über der kritischen Kurve liegt, stoppt das Verfahren und entdeckt eine Veränderung.

Robuste Verfahren in der Change-Point Analyse sind von großer Bedeutung, um auch beim Auftreten stark abweichender Beobachtungen. Solche Beobachtungen können beispielsweise durch schiefe Verteilungen, dicke Flanken oder Ausreißer verursacht werden. Insbesondere in der online Change-Point Analyse sind robuste Verfahren bedeutsam, da beispielsweise bei der Überwachung von Patienten- oder Maschinendaten bei Erkennung eines Strukturbruches schnell eingegriffen werden muss. Ist das verwendete Verfahren nicht robust, können Ausreißer in den Daten leicht einen falschen Alarm auslösen.

In diesem Projekt konstruieren wir robuste sequentielle Verfahren basierend auf U-Statistiken. Dazu muss zunächst die entsprechende Teststatistik konstruiert werden. Für diese Teststatistik werden dann die asymptotischen Verteilungen unter der Null- und Alternativhypothese hergeleitet. Aus den asymptotischen Verteilungen können schließlich die kritische Kurve sowie die Güte des Tests abgeleitet werden. Das Verhalten des Tests bei endlichen Stichproben wird mittels Simulationen untersucht.

Sequentielle Verfahren weisen grundsätzlich eine gewisse zeitliche Verzögerung in der Erkennung eines Strukturbruches auf, da nach der Änderung zunächst genügend Beobachtungen gesammelt werden müssen, um statistische Signifikanz zu erhalten. Diese zeitliche Verzögerung ist von großer Bedeutung, da insbesondere beim Einsatz sequentieller Verfahren zur Überwachung nach einem Strukturbruch möglichst schnell eingegriffen werden muss. Dazu leiten wir die asymptotische Verteilung der zugehörigen Stoppzeit her.

Projektleitung: Prof. Dr. Rainer Schwabe

Projektbearbeitung: Parisa Parsamaram, MSc

Förderer: Fördergeber - Sonstige; 15.03.2017 - 15.09.2017

Empirisch und hierarchisch Bayes-optimale Designs für logistische Regressionsmodelle mit zufälligen Effekten

Binäre Beobachtungen werden statistisch oft mit Hilfe logistischer Regressionsmodelle analysiert. Um die Qualität dieser Analyse zu verbessern, ist das Design des zugrundeliegenden Experiments zu optimieren. Wegen der Nichtlinearität der Modellierung hängen die optimalen Designs von unbekanntem Parametern ab. Um dieses Problem zu umgehen, kann Bayessche a-priori-Information eingeführt werden, die ihrerseits hierarchisch modelliert und mit empirischer Vorinformation validiert werden kann. Zusätzlich ist beim Vorliegen von Messwiederholungen

interindividuelle Variabilität zu berücksichtigen. Alle diese Konzepte sollen in einem einheitlichen Ansatz zusammengefasst werden.

Projektleitung: Prof. Dr. Rainer Schwabe

Projektbearbeitung: Maryma Prus, Dr.

Kooperationen: Dr. Heiko Großmann, Otto-von-Guericke-Universität Magdeburg, Institut für Mathematische Stochastik; Dr. Norbert Benda, Bundesinstitut für Arzneimittel und Medizinprodukte; Prof. Luc Pronzato, Université de Nice, Sophia Antipolis; Prof. Norbert Gaffke, Otto-von-Guericke-Universität Magdeburg, Institut für Mathematische Stochastik; Prof. Radoslav Harman, Comenius-Universität, Bratislava

Förderer: Deutsche Forschungsgemeinschaft (DFG); 16.02.2017 - 15.02.2019

Generierung optimaler und effizienter Experimentaldesigns zur individualisierten Vorhersage in hierarchischen Modellen

Das Ziel des vorliegenden Projektes ist die Entwicklung analytischer Ansätze zur Gewinnung optimaler Designs für die Vorhersage in hierarchischen linearen Modellen sowie in verallgemeinerten linearen und nichtlinearen gemischten Modellen mit zufälligen Parametern. Derartige Modelle wurden ursprünglich in den Bio- und Agrarwissenschaften entwickelt und werden heutzutage in den unterschiedlichsten statistischen Anwendungsgebieten vielfältig eingesetzt.

Projektleitung: Prof. Dr. Rainer Schwabe

Projektbearbeitung: Frank Röttger, MSc

Kooperationen: JProf. Dr. Thomas Kahle, OVGU, FMA, IAG

Förderer: Deutsche Forschungsgemeinschaft (DFG); 01.04.2017 - 31.03.2020

Geometrie optimaler Designs für nichtlineare Modelle in der Statistik

Geometrische Beschreibungen optimaler Designbereiche sind in Zeiten zunehmender Komplexität statistischer Modelle von wachsendem Interesse. Das Ziel dieses Projektes besteht in der Suche von Optimalitätsbereichen von experimentellen Designs für derartige statistische Modelle, insbesondere für verallgemeinerte lineare Modelle mit Poisson- oder logistisch verteilten Zielvariablen. Diese Bereiche können durch Systeme von polynomialen Ungleichungen im Parameterraum beschrieben werden, was bedeutet, dass sie nichts anderes als semialgebraische Mengen sind. Somit können Methoden der algebraischen Geometrie benutzt werden, um die Eigenschaften dieser Optimalitätsbereiche zu studieren. Als Beispiel können im Paarvergleichsmodell nach Bradley-Terry, das ein statistisches Modell für den Vergleich verschiedener Alternativen auf der Basis logistischen Antwortverhaltens ist, die Optimalitätsbereiche für sogenannte saturierte Designs, d.h. Designs mit einer minimalen Anzahl von Trägerpunkten, bestimmt werden.

Projektleitung: Prof. Dr. Rainer Schwabe

Projektbearbeitung: Björn Holzhauer, Dipl.-Math.

Kooperationen: Priv.-Doz. Dr. Ekkehard Glimm, Novartis Pharma AG, Basel

Förderer: Fördergeber - Sonstige; 01.02.2015 - 20.10.2017

Metaanalyse von unerwünschten Ereignissen in klinischen Studien basierend auf Aggregatdaten

Zur Charakterisierung der Nebenwirkungen von medizinischen Behandlungen ist es von Interesse, die Evidenz aus mehreren klinischen Studien zu kombinieren und auch historische Daten über die Kontrollgruppe zu berücksichtigen, weil in jeder einzelnen Studie oft nur wenige unerwünschte medizinische Ereignisse auftreten. Wenn für alle Studien detaillierte Patientendaten verfügbar sind, werden üblicherweise Überlebenszeitmethoden angewandt, um das Auftreten von unerwünschten Ereignissen unter Berücksichtigung der Dauer des Beobachtungszeitraums zu analysieren. Dies ist besonders dann wichtig, falls die Beobachtungsdauer sich zwischen Behandlungsgruppen unterscheidet oder falls die Austauschbarkeit von Modellparametern zwischen Studien von verschiedener Länge angenommen wird. Traditionelle Überlebenszeitmethoden können allerdings nicht angewandt werden, wenn für einige Studien lediglich Aggregatdaten verfügbar sind. Dies stellt ein Problem in der Metaanalyse dar, da Metaanalysen meist auf Veröffentlichungen in medizinischen Fachzeitschriften beruhen, welche in aller Regel keine individuellen Patientendaten enthalten.

Ziel dieses Projektes ist es, statistische Methoden zu entwickeln und zu evaluieren, die Überlebenszeitmetaanalysen unter Berücksichtigung historischer Studien basierend auf Aggregatdaten ermöglichen.

Projektleitung: Prof. Dr. Rainer Schwabe

Projektbearbeitung: Helmi Shat, MSc

Förderer: Deutscher Akademischer Austauschdienst e.V. (DAAD); 01.10.2017 - 30.09.2020

Optimale Planung multi-variabler Accelerated-Degradation-Tests

Die rasante Entwicklung moderner Fertigungstechniken zusammen mit den Bedürfnissen der Verbraucher nach hochqualitativen Produkten dienen als Motivation für Industrieunternehmen, Produkte zu entwickeln und herzustellen, die ohne Ausfall über Jahre oder gar Jahrzehnte funktionieren können. Für derartig langlebige Produkte ist es jedoch eine nicht einfache Aufgabe, innerhalb kurzer verfügbarer Zeit Zuverlässigkeitsaussagen zu treffen, da nicht genügend Daten für eine akkurate Schätzung der Lebensdauer gewonnen werden können. Dementsprechend ist eine Lebensdauerprüfung unter Normalbedingungen nicht sinnvoll. Daher werden Ermüdungstests mit wiederholte Messungen ("repeated measures accelerated degradation tests") häufig in der produzierenden Industrie angewendet, um Lebensdauerverteilungen hochzuverlässiger Produkte zu bestimmen, die bei traditionellen oder beschleunigten Lebensdauererests nicht ausfallen würden. In diesen Experimenten werden Beobachtungen bei hohen Belastungsstufen (z.B. Temperatur, Stromspannung oder Druck) mit Hilfe eines physikalisch sinnvollen statistischen Modells extrapoliert, um Schätzungen der Lebensdauer für niedrigere Belastungen unter Normalbedingungen zu erhalten. Zusätzlich ist zu beachten, dass verschiedene Faktoren wie die Häufigkeit der Messungen, die Stichprobengrößen und die Dauer des Experiments Einfluss auf die Kosten und die Genauigkeit der Schätzung haben.

Im Rahmen dieses Projektes werden zuerst adäquate und relevante Computerexperimente identifiziert und robuste Methoden der Regressionsanalyse entwickelt. Danach werden Optimalitätskriterien für experimentelle Designs definiert, die auf der Qualität der ausgewählten robusten Methoden basieren, und Simulationsbasierte Designs werden entwickelt, um einen einheitlichen Zugang zur Generierung optimaler oder zumindest effizienter Designs für die robuste Analyse in Computerexperimenten zu erhalten.

Projektleitung: Prof. Dr. Rainer Schwabe

Projektbearbeitung: Maryna Prus, Dr.

Kooperationen: Dr. Norbert Benda, Bundesinstitut für Arzneimittel und Medizinprodukte; Dr. Tobias Mielke, Aptiv Solutions, Köln

Förderer: Fördergeber - Sonstige; 16.06.2016 - 15.02.2017

Optimales Design für die Vorhersage zufälliger individueller Effekte in Mehrgruppenmodellen

In der statistischen Datenanalyse werden Modelle mit zufälligen Parametern in verschiedenen Anwendungsbereichen, insbesondere in den Biowissenschaften und der individualisierten Medizinforschung, häufig verwendet. In diesen Modellen ist neben der Schätzung des Populationsparameters die Vorhersage individueller zufälliger Effekte von Interesse. Der individualisierte Ansatz ist vor allem für die Studien, in denen nur wenige Beobachtungen pro Individuum möglich sind, z.B. in der Onkologie, von großer Bedeutung. Ziel dieses Projektes ist es, optimale oder zumindest effiziente und anwendbare Versuchspläne (Designs) für die Vorhersage von individuellen zufälligen Effekten für solche Experimentalsituationen zu konstruieren und zu validieren, in denen in unterschiedlichen Gruppen nur Querschnittsdesigns angewendet werden können.

Projektleitung: Prof. Dr. Rainer Schwabe

Projektbearbeitung: Osama I. O. Idais, MSc

Förderer: Land (Sachsen-Anhalt); 01.10.2016 - 30.09.2019

Optimales Design für multivariate verallgemeinerte lineare Modelle mit stetigen Zielfunktionen

In vielen Anwendungssituationen, in denen Daten gesammelt werden, werden nicht nur eine einzelne, sondern mehrere Zielvariablen gleichzeitig beobachtet, die miteinander korreliert sein können. Derartige multivariate Beobachtungen werden oft mit einer multivariaten Normalverteilung modelliert. In einigen Situationen ist dies jedoch nicht angebracht, insbesondere wenn die beobachteten Merkmale nicht stetig sind. Für diese Situationen ist das Konzept der verallgemeinerten linearen Modelle entwickelt worden, die sich speziell bei binären Daten (z.B. logistische Regression) oder Zähldaten (z.B. Poisson-Regression) bewährt haben. Jedoch kann auch bei stetigen Merkmalen statt der Normalverteilungsannahme eine andere Verteilungsannahme angemessener sein, die sich über ein

verallgemeinertes lineares Modell mit nichtlinearer Linkfunktion beschreiben lässt. Ziel des Projektes ist es, für derartige Modelle asymptotische Eigenschaften unter verschiedenen Korrelationsstrukturen zu bestimmen und auf dieser Basis optimale Designs zu generieren, die zu einer Verbesserung der Datenanalyse führen.

Projektleitung: Prof. Dr. Rainer Schwabe

Projektbearbeitung: Fritjof Freise, Dr.

Förderer: Deutsche Forschungsgemeinschaft (DFG); 01.01.2016 - 31.12.2017

Optimales Design für online generierte adaptive Intelligenztestverfahren (III)

In diesem Projekt sollen adaptive Intelligenztests zur Messung der allgemeinen Intelligenz entwickelt werden. Die Items werden durch einen automatischen Itemgenerator regelbasiert und online generiert und adaptiv dargeboten. Selektiert werden die Items anhand der Parameterschätzungen für erweiterte linear-logistische Testmodelle. Die Parameterschätzungen erfolgen anhand optimaler Designs, so dass mit einem Minimum an darzubietenden Items ein Maximum an Präzision bei der Intelligenzmessung erzielt werden kann. Konkret sollen vier Arten regelgeleiteter Testverfahren zur Messung von allgemeiner Intelligenz konstruiert und hierfür die erforderlichen statistischen Grundlagen entwickelt werden.

In der ersten Phase wurden Items zur Verarbeitungskapazität regelbasiert entworfen und empirisch anhand D-optimaler Versuchspläne mittels linear-logistischer Testmodelle kalibriert. Dazu wurden optimale Versuchspläne für linear-logistische Testmodelle mit festen und zufälligen Faktoren entwickelt. Weiterhin entstand ein Programmsystem zur automatischen Generierung dieser Items, ihrer adaptiven Darbietung und Personenparameterschätzung.

In der zweiten Phase wurden die Arbeiten aus der ersten Phase fortgesetzt. Dazu wurden analog zu den in der ersten Phase entwickelten Items zur Verarbeitungskapazität regelbasierte Items zur Bearbeitungsgeschwindigkeit konstruiert, die sich für eine adaptive Testung dieser Intelligenzkomponente eignen. Da es sich hier um Speed-Tests handelt, war es erforderlich, anstelle des logistischen Rasch-Modells erweiterte Formen des Rasch Poisson Count-Modells als statistische Grundlage heranzuziehen. Für diese Modelle wurden optimale Versuchspläne zur Itemkalibrierung und adaptiven Testung entwickelt.

Ziel der dritten Phase ist es, in Fortsetzung und Ergänzung der Arbeit in den ersten beiden Phasen bei der Modellierung der Intelligenzkomponenten zeitliche, zumeist nichtlineare Trends in longitudinalen Studien zu berücksichtigen und hierfür optimale Designs zu entwickeln, die adaptiv eingesetzt werden können. Darüber hinaus werden Designs für Itempools unter Nebenbedingungen an die Anzahl der verwendeten Regeln bereitgestellt.

Projektleitung: Prof. Dr. Rainer Schwabe

Projektbearbeitung: Martin Radloff, Dipl.-Math.

Förderer: Haushalt; 01.01.2017 - 30.09.2018

Optimales Design für Sphärische Versuchsbereiche

Die Gültigkeit statischer Modelle ist oft auf einen lokalen Bereich der erklärenden Variablen beschränkt. Dieser wird in vielen Anwendungsbereichen als rechteckig angenommen, d.h. die erklärenden Variablen können unabhängig voneinander variieren. In manchen Situationen sind jedoch sphärische Bereiche sinnvoller, die durch einen beschränkten Euklidischen oder Mahalanobis-Abstand zu einem zentralen Punkt für die Versuchseinstellungen beschrieben werden können.

Ziel der Versuchsplanung ist es, optimale oder zumindest effiziente Einstellungen für die erklärenden Variablen zu bestimmen, um die Qualität der statistischen Analyse zu optimieren. Beim Vorliegen klassischer linearer Regressionsmodelle sind Charakterisierungen optimaler Designs für sphärische Versuchsbereiche mit Hilfe von Invarianzen und Symmetrien schon seit längerem bekannt. Fragestellung dieses Projekts ist es, für die in der statistischen Praxis zunehmend verwendeten verallgemeinerten linearen Modelle bzw. nichtlinearen Modelle optimale Designs auf derartigen sphärischen Versuchsbereichen zu bestimmen. Erste Ergebnisse für Poisson-verteilte Zähldaten zeigen deutliche Abweichungen der hierfür benötigten optimalen Designs von denjenigen für klassische lineare Modelle.

Projektleitung: Prof. Dr. Rainer Schwabe

Projektbearbeitung: Dennis Schmidt, Dipl.-Math.

Kooperationen: Dr. Heiko Großmann, Otto-von-Guericke-Universität Magdeburg, Institut für Mathematische Stochastik

Förderer: Fördergeber - Sonstige; 01.04.2013 - 30.06.2017

Optimales Design für statistische Modelle mit zensierten Daten

In vielen technischen und biologischen Bereichen spielt die statistische Analyse zensierter Daten eine zunehmende Rolle. Diese Zensierungen können deterministisch (feste Studiendauer, Nachweisgrenzen) oder zufallsgesteuert (zufällige Studiendauer, zufälliges Ausscheiden aus der Studie) sein. Die beobachteten, teilweise zensierten Größen können zusätzlich von weiteren Einflussfaktoren (Behandlungen und Kovariablen) abhängen, was beispielsweise über ein "proportional hazards"-Modell beschrieben werden kann.

Während die statistische Analyse derartiger Daten schon relativ weit entwickelt ist, gibt es relativ wenig Resultate zur effizienten Planung derartiger Studien oder Experimente. Ziel des vorliegenden Projekts ist es, für eine Reihe von relevanten Modellsituationen optimale oder zumindest effiziente Designs zu charakterisieren und analytisch zu bestimmen, um Anleitungen für eine möglichst effektive Ausnutzung der vorhandenen Ressourcen beim Vorliegen zensierter Daten bereit zu stellen.

Projektleitung: Prof. Dr. Rainer Schwabe

Projektbearbeitung: Marius Schmidt, Dipl.-Math.

Kooperationen: Dr. Hermann Kulmann, Bayer, Berlin; Dr. Thomas Schmelter, Bayer, Berlin; Dr. Tobias Mielke, Aptiv Solutions, Köln; Priv.-Doz. Dr. Steffen Uhlig, Quo Data, Dresden; Prof. Dr. Heinz Holling, Universität Münster, Institut für Psychologie IV

Förderer: Haushalt; 01.10.2013 - 30.09.2018

Optimales Design für verallgemeinerte lineare gemischte Modelle

Gemischte Modelle spielen zunehmend eine wichtige Rolle nicht nur in Biowissenschaften sondern auch bei wirtschafts- und sozialwissenschaftlichen Fragestellungen, um individuelle Effekte der verschiedenen Beobachtungseinheiten als Repräsentanten einer größeren Grundgesamtheit bei der statistischen Datenanalyse berücksichtigen und erfassen zu können. Mit verallgemeinerten linearen gemischten Modellen werden Zusammenhänge für binäre ("Erfolg - Misserfolg") und diskrete Zielgrößen ("Anzahlen") beschrieben, die nicht sinnvoll durch standardmäßige lineare gemischte Modelle für metrische Daten dargestellt werden können. Für die zufälligen Effekte können dann neben normalverteilten individuellen Einflüssen auch solche aus konjugierten Familien angenommen werden, die eine explizitere Analyse erlauben. Wie in allen statistischen Analysen hängt auch hier die Qualität der Ergebnisse wesentlich vom Beobachtungs- oder Experimentaldesign, d.h. der Wahl der Beobachtungseinheiten und Beobachtungszeitpunkte, ab. Ziel dieses Projektes ist es, optimale oder zumindest effiziente Designs für verallgemeinerte lineare gemischte Modelle zu entwickeln, die sowohl normalverteilte als auch Effekte aus konjugierten Verteilungen beinhalten können, und diese zu validieren.

Projektleitung: Prof. Dr. Rainer Schwabe

Projektbearbeitung: Iyad Aboud, MSc

Kooperationen: Prof. Dr. Heinz Holling, Universität Münster, Institut für Psychologie IV

Förderer: Fördergeber - Sonstige; 01.10.2015 - 30.09.2019

Optimales Design für Zähldaten

Neben klassischen Ansätzen stetiger, metrischer oder diskreter, binärer Daten (Messungen oder Ja / Nein -Antworten) spielt in statistischen Anwendungen die Analyse von Zähldaten eine zunehmende Rolle. Derartige Beobachtungen von Anzahlen treten zum Beispiel im Transportwesen, bei der Schadstoffmessung, in der Psychologie oder in medizinischen Anwendungen auf. Klassischerweise werden Anzahlen über Modelle mit Poisson-Verteilungen beschrieben, die Äquidispersion, d.h. Gleichheit von Mittelwert und Varianz, aufweisen. In der Praxis ist diese Annahme aber zu restriktiv, so dass alternativ häufig Modelle mit Überdispersion (Vorliegen einer größeren Varianz) und/oder exzessiven Nullen (zero inflation, verstärkte Beobachtung von Nullanzahlen) verwendet werden, um die Daten adäquat zu beschreiben. Während die Datenanalyse von Zähldaten relativ weit entwickelt und entsprechende Software verfügbar ist, gibt es nur wenige Resultate zur Planung von Experimenten mit Zähldaten. Diese beschränken sich auch bisher auf den klassischen Poisson-Ansatz oder auf ein spezielles Modell der Überdispersion mit negativ-binomialverteilten Daten, das sich als Mischmodell über eine konjugierte a-priori-Verteilung der Modellparameter ergibt. Dabei spielt die

Planung in Experimentalsituationen eine immens wichtige Rolle, da nur mit einer vernünftigen Auswahl der Experimentaleinstellungen die vorhandenen Ressourcen effektiv ausgenutzt sowie Kosten und Aufwand verringert werden können. Ziel dieses Projektes ist es, optimale Designs für Zähldaten zu generieren, die auch beim Vorliegen von Überdispersion und/oder exzessiven Nullen eine effiziente Auswertung der Beobachtungen ermöglichen.

Projektleitung: Prof. Dr. Rainer Schwabe

Projektbearbeitung: Eric Nyarko, MSc

Kooperationen: Dr. Heiko Großmann, Otto-von-Guericke-Universität Magdeburg, Institut für Mathematische Stochastik

Förderer: Deutscher Akademischer Austauschdienst e.V. (DAAD); 01.10.2016 - 30.09.2019

Optimales Design in der Discrete-Choice-Analyse bei geblockten Beobachtungen

Die Discrete-Choice-Analyse ist ein häufig angewendetes Verfahren der Marktforschung. Sie wird verwendet, um das Präferenzverhalten von Konsumenten zu untersuchen und den Nutzen zu ermitteln, den die verschiedenen Attribute eines Produktes besitzen. Die den Konsumenten dabei vorgelegten Auswahlfragen erfordern den Vergleich von Produktbeschreibungen, welche unter Verwendung eines experimentellen Designs zusammengestellt werden. Die Qualität der Ergebnisse eines solchen Experiments hängt folglich stark vom verwendeten Design ab. Bei der Modellierung der Daten und der Wahl des Designs wird häufig jedoch nicht berücksichtigt, dass den teilnehmenden Personen mehrere Fragen gestellt werden und die resultierenden Antworten daher korreliert sein können. Ziel des vorliegenden Projektes ist es, optimale und effiziente Designs unter Berücksichtigung von Blockeffekten zur Modellierung dieser potenziellen Abhängigkeiten zu entwickeln und zu validieren.

Projektleitung: Prof. Dr. Alexandra Carpentier

Förderer: Deutsche Forschungsgemeinschaft (DFG); 01.10.2017 - 14.06.2019

Active learning for matrix completion

Matrix completion is an essential problem in modern machine learning, as it is e.g. important for the calibration of the recommendation systems. We consider the problem of matrix completion in the setting where the learner can choose where to sample. In this setting, it can be of interest to target more specifically parts of the matrix where it is discovered that the complexity is high (higher local rank), where the knowledge is limited (few sampled points), or where the noise is high. This project plans to consider first the problem of active learning for matrix completion when the matrix can be subdivided into block submatrices of small ranks that are known, and then in the more general case where this cannot be done.

Projektleitung: Prof. Dr. Alexandra Carpentier

Förderer: Haushalt; 01.10.2017 - 31.12.2020

Adaptive rejection sampling using non-parametric estimators

Rejection Sampling is a basic Monte-Carlo method, used to sample from distributions admitting a probability density function, for which evaluations of the density are costly but still tractable. This technique is however rarely used as it stands, because of the high rejection rate it implies. Several methods have been explored to cope with this problem. Most of them either take advantage of strong assumptions on the form of the density, or do not present any performance guarantee. We aim at studying the minimax properties of this problem over the class of smooth density functions

Projektleitung: Prof. Dr. Alexandra Carpentier

Projektbearbeitung: Joseph Lam, M.Sc.

Förderer: Haushalt; 01.11.2017 - 31.10.2020

Adaptive two sample test in the density setting

We consider the problem of testing between two samples of (non necessarily uniform) density. While minimax signal detection in the case where the null hypothesis density is uniform is well understood, recent works in the case of multinomial distributions have highlighted the amelioration in the minimax rate that can come when considering non uniform null hypothesis density. We want to study this problem in the two sample testing case, which is significantly more complex, and extend it to smooth densities.

Projektleitung: Prof. Dr. Alexandra Carpentier

Förderer: Deutsche Forschungsgemeinschaft (DFG); 01.10.2017 - 14.06.2019

MuSyAD on Anomaly Detection

Anomaly detection is an interdisciplinary domain, borrowing elements from mathematics, computer science, and engineering. The main aim is to develop efficient techniques for detecting anomalous behaviour of systems. In the classical scenario a monitor receives data from a system and compares this data to a reference system with some single normal behaviour. Ideally no strong assumptions are made on the nature of anomalous behaviours, so the problem of anomaly detection is by essence a non parametric problem. Here I propose to study a more complex scenario, which will be referred to as multisystem anomaly detection. In this setting, reference systems can have a variety of normal behaviours, and moreover, there are many systems under the monitors surveillance, and the monitor must allocate its resources wisely among them. In this situation new theoretical and computational challenges arise. The overall objective of this proposal is to find efficient methods to solve the problem of multi-system anomaly detection. This aim will be reached by addressing the following sub-objectives. First, we will generalise the theoretical framework of anomaly detection to the broader setting of multi-system anomaly detection. Second, multi-system anomaly detection methods will be developed, by taking ideas from the non parametric testing field and applying them to the new framework. Third, we will study optimal monitoring strategies for cases where the multiple systems cannot be monitored simultaneously. Here, it is important that the monitor allocates its resources among the systems in a way that is as efficient as possible. To this end, sequential and adaptive sampling methods that target the anomaly detection problem will be designed. Since anomaly detection is a non parametric problem, elements in the theory of non parametric confidence sets will be used. Finally, the newly developed methods will be applied to practical problems: a methodological example in extreme value theory, an econometric application for speculative bubble detection and two applications in a Brain Computer Interface framework.

Projektleitung: Prof. Dr. Alexandra Carpentier

Förderer: Deutsche Forschungsgemeinschaft (DFG); 01.10.2017 - 14.06.2019

Online adaptation to the parameters of the problem in active learning

We consider the problem of active learning in the setting of classification and optimisation, which are baseline problems in applied mathematics and machine learning. The problem of adaptivity (to unknown distributional parameters) has remained open in many contexts (e.g. smooth decision boundary for classification, or optimisation of the cumulative regret). While some recent advances on this problems established adaptive rates in some contexts, adaptivity in most of the real world setting has so far remained elusive. In this project, we plan to investigate the problem of adapting to the unknown parameters of the problem (e.g. smoothness, margin assumptions, measure of the near optimal points, etc), and intend to develop minimax rates on the learning efficiency with respect to an oracle learner.

Projektleitung: Prof. Dr. Alexandra Carpentier

Förderer: Deutsche Forschungsgemeinschaft (DFG); 01.10.2017 - 30.12.2018

Projekt on Data Assimilation

This project is concerned with the problem of learning sequentially, adaptively and in partial information on an uncertain environment. In this setting, the learner collects sequentially and actively the data, which is not available before-hand in a batch form. The process is as follows: at each time t , the learner chooses an action and receives a data point, that depends on the performed action. The learner collects data in order to learn the system, but also to achieve a goal (characterized by an objective function) that depends on the application. In this project, we will aim at solving this problem under general objective functions, and dependency in the data collecting process exploring variations of the so-called bandit setting which corresponds to this problem with a specific objective function.

As a motivating example, consider the problem of sequential and active attention detection through an eye tracker. A human user is looking at a screen, and the objective of an automatized monitor (learner) is to identify through an eye tracker zones of this screen where the user is not paying sufficient attention. In order to do so, the monitor is allowed at each time t to flash a small zone a_t in the screen, e.g. light a pixel (action), and the eye tracker detects through the eye movement if the user has observed this flash. Ideally the monitor should focus on these difficult zones and flash more often there (i.e. choose more often specific actions corresponding to less identified zones). Therefore, sequential and adaptive learning methods are expected to improve the performances of the monitor.

Projektleitung: Prof. Dr. Alexandra Carpentier
Förderer: Haushalt; 01.10.2017 - 14.06.2019

Smoothness testing in the Sobolev sense

We want to develop a test to determine whether a function lying in a fixed L_2 -Sobolev-type ball of smoothness t , and generating a noisy signal, is in fact of a given smoothness s larger than t or not. While it is impossible to construct a uniformly consistent test for this problem on every function of smoothness t , it becomes possible if we remove a sufficiently large region of the set of functions of smoothness t . The functions that we remove are functions of smoothness strictly smaller than s , but that are very close to s -smooth functions. This problem has been considered in the case of specific Besov bodies where it is easier, and we plan to extend it to more usual Sobolev ellipsoids.

Projektleitung: Prof. Dr. Alexandra Carpentier
Förderer: Haushalt; 01.10.2017 - 30.12.2020

Testing the presence of jumps in Levy processes, and on conditions for the equivalence of jumps to a Gaussian noise

In this project we aim at first developing a test for detecting the presence of jumps in Levy processes, and at establishing a minimax testing framework for describing the cases where testing is, or is not possible. We then plan to use these results to output conditions under which small jumps can efficiently be modeled by a Brownian motion, i.e. in the case where they are indistinguishable from a Brownian motion

Projektleitung: Prof. Dr. Alexandra Carpentier
Förderer: Haushalt; 01.10.2017 - 31.12.2020

Testing the sparsity in the sparse linear model

Consider the Gaussian linear regression model. We study the twin problems of estimating the number of non-zero components of the parameter and testing whether this number is smaller than some value. For testing, we aim at establishing the minimax separation distances for this model and introduce a minimax adaptive test. Extensions to the case of unknown variance will also be discussed. Rewriting the estimation of the number of non-zero components of the parameter as a multiple testing problem, we aim at deriving a new way of assessing the optimality of a sparsity estimator.

Projektleitung: Prof. Dr. Alexandra Carpentier
Förderer: Haushalt; 01.10.2017 - 31.12.2020

Tight rates for estimating and testing the number of true null hypotheses in multiple testing

In this project we consider a multiple testing problem, and we aim at estimating the proportion of true null hypotheses. This is a complex problem as it amounts effectively to estimating the sparsity of a vector. We aim at considering two cases : one sided testing, and two sided testing, which correspond to two different usual settings in multiple testing.

Projektleitung: Prof. Dr. Alexandra Carpentier
Förderer: Deutsche Forschungsgemeinschaft (DFG); 01.10.2017 - 14.06.2019

Two sample tests in structured networks

The study of networks leads to a wide range of high dimensional inference problems. In most practical scenarios, one needs to draw inference from a small population of large networks. The present paper studies hypothesis testing of graphs in this high-dimensional regime. We consider the problem of testing between two populations of inhomogeneous random graphs defined on the same set of vertices. We are interested in tests based on estimates of the Frobenius and operator norms of the difference between the population adjacency matrices, in both the "large graph, small sample" and "small graph, large sample" regimes. We aim at deriving lower bounds on the minimax separation rate for the associated testing problems, and show that the constructed tests are near optimal. We plan to focus on the structured graph setting (e.g. sparse graphs, etc).

Projektleitung: Prof. Dr. Gerd Christoph
Förderer: Haushalt; 01.10.2017 - 31.12.2019

Edgeworth und Cornish-Fisher Entwicklungen

Für asymptotisch normalverteilte Statistiken werden Edgeworth und Cornish-Fisher Entwicklungen hergeleitet, die bessere Approximationen der unbekannt Quantile der zugrunde liegenden Statistik liefern können. In Anwendungen ist öfter der Stichprobenumfang ebenfalls vom Zufall abhängig. Für Stichproben mit zufälligen

Stichprobenumfang ändert sich oft das Grenzverteilung der untersuchten Statistik, anstelle der oft erwarteten Normalverteilung tritt z.B. die Student- oder Laplace-Verteilung. Untersuchungen wurden zum arithmetischen Mittel und zum Median bei Stichproben mit zufälligen Umfang durchgeführt.

Projektleitung: Prof. Dr. Norbert Gaffke
Projektbearbeitung: Prof. Dr. Norbert Gaffke
Kooperationen: Prof. Dr. Rainer Schwabe, OVGU, FMA-IMST
Förderer: Fördergeber - Sonstige; 01.10.2015 - 28.09.2020

Algorithmen zum optimalen Design für lineare Regressionsmodelle.

Im Rahmen der approximativen Design-Theorie für lineare Regressionsmodelle sollen optimale Designs algorithmisch berechnet werden (insbesondere D-optimale und I-optimale Designs). Ein universell einsetzbarer Algorithmus existiert nicht. Ob die {em konzeptuell} vorhandenen Algorithmen zur Anwendung kommen können, hängt von der Komplexität des Modells ab und erfordert ggf. weiteren theoretischen Input. Im Projekt sollen unsere Quasi-Newton Methoden (s. Gaffke, Graßhoff, Schwabe, 2014) auf zwei Modellklassen angewendet werden: Zum Einen Querschnitts-Designs bei longitudinalen Daten, z.B. im Kontext von "accelerated life testing"-Untersuchungen in der Qualitätskontrolle (vgl. Weaver and Meeker, 2014). Zum anderen der Fall eines {em endlichen} Versuchsbereichs, wobei auch Stratifizierungs- oder Kostenrestriktionen einbezogen werden. Hierfür sind in den letzten Jahren Algorithmen vom Silvey-Titterington-Torsney Typ wieder aufgegriffen worden (vgl. Harman, 2014). Diese wollen wir mit unseren Quasi-Newton Methoden kontrastieren.

Literatur:

Gaffke,N.; Graßhoff,U.; Schwabe,R.: Algorithms for approximate linear regression design applied to a first order model with heteroscedasticity.
Computational Statistics and Data Analysis 71 (2014),1113-1123.

Weaver,B.P.; Meeker, W.Q.: Methods for Planning Repeated Measures Accelerated Degradation Tests.
Applied Stochastic Models in Business and Industry 30 (2014), 658-671,

Harman,R.: Multiplicative methods for computing D-optimal stratified designs of experiments.
Journal of Statistical Planning and Inference 146 (2014), 82-94.

Projektleitung: Prof. Dr. Norbert Gaffke
Förderer: Fördergeber - Sonstige; 01.04.2015 - 31.12.2018

Die Verteilung einer nicht-parametrischen Teststatistik für den Erwartungswert.

Im nicht-parametrischen Modell mit n i.i.d. nicht negativen Zufallsvariablen wurde von Gaffke (2005) eine Teststatistik vorgeschlagen, die auch die Kontruktion einer unteren Konfidenzschranke für den Erwartungswert erlaubt Immer noch offen ist die Frage, ob das nominelle Konfidenzniveau eingehalten wird. Äquivalent ist die Frage, ob die Verteilung der Teststatistik stochastisch größer (oder gleich) der Standard-Rechteck-Verteilung ist, unter jeder zu Grunde liegenden Verteilung mit Erwartungswert gleich 1. Es besteht einige numerische Evidenz, dass die Antwort positiv ist. Bewiesen ist aber wenig: Nur der Fall $n = 2$ (und der triviale Fall $n = 1$), sowie die asymptotische Aussage, dass für $n \rightarrow \infty$ die Verteilung der Teststatistik gegen die Standard-Rechteck-Verteilung konvergiert. Das erste Ziel des Projektes ist es, den Fall $n = 3$ zu beantworten.

Literatur:

Gaffke, N. (2005): Three test statistics for a nonparametric one-sided hypothesis on the mean of a nonnegative random variable.
Mathematical Methods of Statistics 14, 451-467.

Projektleitung: Prof. Dr. Cornelia Pokalyuk

Förderer: Deutsche Forschungsgemeinschaft (DFG); 01.01.2016 - 31.12.2017

Balancing selection in viral populations

Die Rolle von Balancing Selection als evolutionäre Kraft ist bisher weitgehend ungeklärt. Das Zytomegalievirus zeichnet sich durch eine enorme genetische Variabilität aus. In diesem Projekt möchten wir herausfinden, ob und wie sich diese Variabilität (zumindestens teilweise) durch Balancing Selection erklären lässt. Die Basis dieser Analyse bilden stochastische Modelle, die Balancing Selection in einem Wirt-Virus-System beschreiben.

Projektleitung: Prof. Dr. Cornelia Pokalyuk

Förderer: Deutsche Forschungsgemeinschaft (DFG); 01.01.2016 - 31.12.2017

Selektive Sweeps in strukturierten Populationen

Wie kann man den Einfluss von Selektion auf die Evolution einer Population quantifizieren? Diese Frage ist in der Populationsgenetik von großem Interesse. Aufschluss darüber gibt die Analyse des Fixationsprozesses eines selektiv bevorzugten Allels. In diesem Projekt geht es darum diesen Fixationsprozess mittels eines genealogischen Ansatzes zu verstehen.

Projektleitung: apl. Prof. Dr. Waltraud Kahle

Projektbearbeitung: Waltraud Kahle

Förderer: Haushalt; 01.01.2016 - 31.12.2018

Unvollständige Instandhaltung in Abnutzungsprozessen

Es wird ein Wiener Prozess zur Modellierung des Abnutzungsverhaltens betrachtet. Ein Ausfall erfolgt, wenn der Abnutzungsprozess ein vorgegebenes Niveau erstmalig erreicht.

Zur Vermeidung von Ausfällen werden regelmäßig vorbeugende Instandhaltungsmaßnahmen durchgeführt, die das Abnutzungslevel verringern. Im Projekt werden folgende Fragestellungen betrachtet;

- Einführung der Konzepte des virtuellen Alters und des Reparaturgrades, die bei der Betrachtung von vorbeugenden Instandhaltungen in der Lebensdaueranalyse von Systemen verwendet werden,
 - Einfluß der Instandhaltung auf die Lebensdauerverteilung,
 - Definition von Kostenfunktionen der vorbeugenden Instandhaltung in Abhängigkeit vom Reparaturgrad,
 - Kostenoptimale Instandhaltung.
-

Projektleitung: Dr. Heiko Großmann

Projektbearbeitung: Bairu Zhang, MSc

Förderer: Fördergeber - Sonstige; 06.01.2014 - 06.01.2018

Funktionale Datenanalyse von Ganganalyse-Daten

Bestimmte neurologische Erkrankungen beeinträchtigen die Gehfähigkeit der betroffenen Individuen. In diesem Projekt werden Verfahren der funktionalen Datenanalyse entwickelt, um Daten zu analysieren, die mit Hilfe bildgebender Verfahren in einem Ganglabor bei Kindern und Jugendlichen erhoben werden. Im angewandten Teil des Projekts wird unter anderem untersucht, wie sich bestimmte medizinische Hilfsmittel (Orthesen) auf das Gehverhalten auswirken.

7. Eigene Kongresse, wissenschaftliche Tagungen und Exponate auf Messen

- Prof. Dr. Claudia Kirch: NeurIT Workshop, 18.10.2017 in der Lukasklause, Magdeburg
- Dr. Maryna Prus: Organisation der Sektion "Optimal Design in Mixed Models" auf der CEN-ISBS-Tagung, 28.08.-01.09.2017, Wien

8. Veröffentlichungen

Begutachtete Zeitschriftenaufsätze

Alonso Cabrera, Jesús; Schwabe, Rainer

D-optimal design for a model with interaction between a qualitative and a quantitative factor in the presence of random block effects

In: Journal of statistical theory and practice - London [u.a.]: Taylor & Francis, Bd. 11.2017, 4, S. 759-770

Benda, Norbert; Bürkner, Paul-Christian; Freise, Fritjof; Holling, Heinz; Schwabe, Rainer

Adaptive designs for quantal dose-response experiments with false answers

In: Journal of statistical theory and practice - London [u.a.]: Taylor & Francis, Bd. 11.2017, 3, S. 361-374

Häfner, Franziska; Kirch, Claudia

Moving fourier analysis for locally stationary processes with the bootstrap in view

In: Journal of time series analysis: a journal sponsored by the Bernoulli Society for Mathematical Statistics and Probability - Oxford [u.a.]: Wiley-Blackwell, Bd. 38.2017, 6, S. 895-922

[Imp.fact.: 0,975]

Maier, Oliver; Györfi, Benedikt; Wrede, Jürgen; Kasper, Roland

Design and validation of a multi-body model of a front suspension bicycle and a passive rider for braking dynamics investigations

In: Multibody system dynamics - Dordrecht [u.a.]: Springer Science + Business Media B.V, insges. 27 S., 2017

[Imp.fact.: 1,389]

Schmidt, Dennis; Schwabe, Rainer

Optimal design for multiple regression with information driven by the linear predictor

In: Statistica Sinica - Taipei: Statistica Sinica, Institute of Statistical Science, Academia Sinica, Bd. 27.2017, 3, S. 1371-1384

[Imp.fact.: 0,899]

Schmidt, Marius; Bück, Andreas; Tsotsas, Evangelos

Experimental investigation of the influence of drying conditions on process stability of continuous spray fluidized bed layering granulation with external product separation

In: Powder technology: an international journal on the science and technology of wet and dry particulate systems - Amsterdam [u.a.]: Elsevier Science, Bd. 320.2017, S. 474-482

[Imp.fact.: 2,942]

Zhang, Bairu; Twycross-Lewis, Richard; Großmann, Heiko; Morrissey, Dylan

Testing gait with ankle-foot orthoses in children with cerebral palsy by using functional mixed-effects analysis of variance

In: Scientific reports - [London]: Macmillan Publishers Limited, part of Springer Nature, Vol. 7.2017, Art. 11081, insgesamt 12 S.

[Imp.fact.: 4,259]

Nicht begutachtete Zeitschriftenaufsätze

Carpentier, Alexandra; Kim, Arlene

An iterative hard thresholding estimator for low rank matrix recovery with explicit limiting distribution

In: De.arxiv.org - [S.l.]: Arxiv.org, insges. 40 S., 2017

Christoph, Gerd; Monakhov, M. M.; Ulyanov, V. V.

Second order Chebyshev-Edgeworth and Cornish-Fisher expansions for distributions of statistics constructed from samples with random sizes

In: Magdeburg Universität, Fakultät für Mathematik, 2017, 30 Seiten, Illustrationen - (Preprint; Fakultät für Mathematik, Otto-von-Guericke-Universität Magdeburg; 2017,Nr.08)

[Literaturverzeichnis: Seite 28-30]

Ghoshdastidar, Debarghya; Gutzeit, Maurilio; Carpentier, Alexandra; Luxemburg, Ulrike

Two-sample tests for large random graphs using network statistics

In: De.arxiv.org - [S.l.]: Arxiv.org, insges. 24 S., 2017

Harman, Radoslav; Prus, Maryna

Computing optimal experimental designs with respect to a compound Bayes risk criterion

In: De.arxiv.org - [S.I.]: Arxiv.org, insges. 13 S., 2017

Locatelli, Andrea; Carpentier, Alexandra; Kpotufe, Samory

Adaptivity to noise parameters in nonparametric active learning

In: De.arxiv.org - [S.I.]: Arxiv.org, insges. 32 S., 2017

Begutachtete Buchbeiträge

Dietrich, Stephanie; Kahle, Waltraud

Optimal imperfect maintenance in a multi-state system

In: Recent Advances in Multi-state Systems Reliability: Theory and Applications - Cham: Springer International Publishing, S. 179-193, 2017

Dissertationen

Holzhauser, Björn; Schwabe, Rainer [AkademischeR BetreuerIn]

Meta-analysis of aggregate data on medical events. - Magdeburg, 2017, V, 152 Seiten, Illustrationen

[Literaturverzeichnis: Seite 111-124]

Schmidt, Dennis; Schwabe, Rainer [AkademischeR BetreuerIn]

Versuchsplanung für nichtlineare multiple Regressionsmodelle mit Anwendung auf zensierte Daten. - Magdeburg, 2017,

IV, 127 Seiten