

Kapitel 0

Einführung und Vorbemerkungen

Thorsten Dickhaus

Universität Bremen
Institut für Statistik

Mathematik 3: Stochastik
Universität Bremen, Fachbereich 03, SoSe 2025

Übersicht

- 1 Inhaltliche Einführung
- 2 Ablauf und Formalitäten
- 3 Zusätzliche Literatur

Übersicht

1 Inhaltliche Einführung

2 Ablauf und Formalitäten

3 Zusätzliche Literatur

Was ist Stochastik?

Stochastik = Wahrscheinlichkeitstheorie + Statistik

Abkürzungen:

- W'theorie: Wahrscheinlichkeitstheorie
- W'keit: Wahrscheinlichkeit
- stoch.: stochastisch
- stat.: statistisch

Elemente der Wahrscheinlichkeitstheorie

- Mathematische Theorie von Zufall
- Wahrscheinlichkeitsräume, -maße, -verteilungen
- Zufallsvariablen
- Erwartungswert, Varianz, Momente
- Kovarianz und Korrelation
- Grenzwertbegriffe und -sätze für Zufallsvariablen
- und vieles mehr...

Elemente der Statistik

- Beschreibende Statistik
(Stabdiagramme, Histogramme, stat. Kennzahlen, ...)
- Schließende Statistik (Inferenzstatistik)
- Schätzen
- Testen
- Klassifikation, Regression (überwachtes stat. Lernen)
- Clusteranalyse, Strukturerkennung in Datensätzen
(unüberwachtes statistisches Lernen)
- Entscheidungstheorie
- und vieles mehr...

Stochastische Anwendungen in der Informatik

- Randomisierte Algorithmen, Monte Carlo-Verfahren
- Medizinische Informatik: Zufallsstichproben ziehen, Imputation fehlender Werte, etc.
- Bioinformatik: Informationsextraktion aus (phylo-)genetischen Daten
- Künstliche Intelligenz, maschinelles Lernen
- Brain-Computer Interfacing (BCI)

2426–2431 *Nucleic Acids Research*, 2012, Vol. 40, No. 6
doi:10.1093/nar/gkr1073

Published online 29 November 2011

The allele distribution in next-generation sequencing data sets is accurately described as the result of a stochastic branching process

Verena Heinrich¹, Jens Stange², Thorsten Dickhaus², Peter Imkeller², Ulrike Krüger¹,
Sebastian Bauer¹, Stefan Mundlos¹, Peter N. Robinson¹, Jochen Hecht³ and
Peter M. Krawitz^{1,*}

¹Institute for Medical and Human Genetics, Charité Universitätsmedizin Berlin, Augustenburger Platz 1,
13353 Berlin, ²Department of Mathematics, Humboldt-University Berlin, Unter den Linden 6, 10099 Berlin and
³Berlin-Brandenburg Center for Regenerative Therapies (BCRT), Charité Universitätsmedizin Berlin,
Augustenburger Platz 1, 13353 Berlin, Germany

Received July 20, 2011; Revised October 19, 2011; Accepted October 28, 2011

*Statistical Applications in Genetics
and Molecular Biology*

Volume 11, Issue 4

2012

Article 12

How to analyze many contingency tables
simultaneously in genetic association studies

Thorsten Dickhaus, *Humboldt-University, Berlin*

Klaus Straßburger, *German Diabetes Center, Düsseldorf*

Daniel Schunk, *Johannes Gutenberg-Universität Mainz and
University of Zurich*

Carlos Morcillo-Suarez, *Universitat Pompeu Fabra,
Barcelona*

Thomas Illig, *Helmholtz Zentrum München*

Arcadi Navarro, *ICREA and Universitat Pompeu Fabra,
Barcelona*

Bioinformatics, 31(22), 2015, 3577–3583

doi: 10.1093/bioinformatics/btv457

Advance Access Publication Date: 6 August 2015

Original Paper



Genome analysis

Strategies to improve the performance of rare variant association studies by optimizing the selection of controls

Na Zhu¹, Verena Heinrich¹, Thorsten Dickhaus², Jochen Hecht³,
Peter N. Robinson¹, Stefan Mundlos^{1,4}, Tom Kamphans⁵ and
Peter M. Krawitz^{1,4,*}

¹Institute of Medical Genetics and Human Genetics, Charité Universitätsmedizin Berlin, 13353 Berlin, Germany,

²Institute for Statistics, University of Bremen, 28344 Bremen, Germany, ³Berlin-Brandenburg Center for Regenerative Therapies (BCRT), 13353 Berlin, Germany, ⁴Max Planck Institute for Molecular Genetics, 14195 Berlin, Germany and ⁵GeneTalk, 13189 Berlin, Germany

*To whom correspondence should be addressed.

Associate Editor: John Hancock

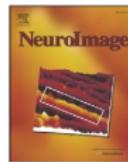
Received on May 25, 2015; revised on July 30, 2015; accepted on July 30, 2015



Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/ynim



Neurophysiological predictor of SMR-based BCI performance

Benjamin Blankertz^{a,b}, Claudia Sannelli^a, Sebastian Halder^c, Eva M. Hammer^c, Andrea Kübler^{c,d}, Klaus-Robert Müller^a, Gabriel Curio^e, Thorsten Dickhaus^{a,*}

^a Machine Learning Laboratory, Berlin Institute of Technology, Germany

^b Intelligent Data Analysis Group, Fraunhofer FIRST, Berlin, Germany

^c Institute of Medical Psychology and Behavioral Neurobiology, University of Tübingen, Germany

^d Department of Psychology I, University of Würzburg, Germany

^e Department of Neurology, Campus Benjamin Franklin, Charité University Medicine Berlin, Germany

ARTICLE INFO

Article history:

Received 25 January 2010

Revised 8 March 2010

Accepted 9 March 2010

Available online 17 March 2010

ABSTRACT

Brain-computer interfaces (BCIs) allow a user to control a computer application by brain activity as measured, e.g., by electroencephalography (EEG). After about 30 years of BCI research, the success of control that is achieved by means of a BCI system still greatly varies between subjects. For about 20% of potential users the obtained accuracy does not reach the level criterion, meaning that BCI control is not accurate enough to control an application. The determination of factors that may serve to predict BCI performance,

Anwendungen der Informatik in der Stochastik

- Kombinatorische Algorithmen
- Informationstheorie, Entropie
- Algorithmische Datenanalyse
- Effiziente Programmierung statistischer Verfahren
- Data Engineering, Data Warehousing
- Datensicherheit, Datenschutz, etc.

SCIENTIFIC REPORTS

OPEN

Combining Multiple Hypothesis Testing with Machine Learning Increases the Statistical Power of Genome-wide Association Studies

Received: 31 May 2016

Accepted: 06 October 2016

Published: 28 November 2016

Bettina Mieth^{1,*}, Marius Kloft^{2,*}, Juan Antonio Rodriguez^{3,†}, Sören Sonnenburg⁴, Robin Vobruba¹, Carlos Morcillo-Suárez³, Xavier Farre³, Urko M. Marigorta⁵, Ernst Fehr^{6,†}, Thorsten Dickhaus^{7,†}, Gilles Blanchard^{8,†}, Daniel Schunk^{9,†}, Arcadi Navarro^{3,10,11,†} & Klaus-Robert Müller^{1,12,†}

The standard approach to the analysis of genome-wide association studies (GWAS) is based on testing each position in the genome individually for statistical significance of its association with the phenotype under investigation. To improve the analysis of GWAS, we propose a combination of machine learning and statistical testing that takes correlation structures within the set of SNPs under investigation in a mathematically well-controlled manner into account. The novel two-step algorithm, COMBI, first trains a support vector machine to determine a subset of candidate SNPs and then performs hypothesis tests for these SNPs together with an adequate threshold correction. Applying COMBI to data from a WTCCC study (2007) and measuring performance as replication by independent GWAS published within the

Übersicht

- 1** Inhaltliche Einführung
- 2** Ablauf und Formalitäten
- 3** Zusätzliche Literatur

Mathematik 3 <i>Mathematics 3</i>		Modulnummer: IBGT-M3	
Bachelor Pflicht/Wahl <input checked="" type="checkbox"/> Wahl <input type="checkbox"/> Basis <input type="checkbox"/> Ergänzung <input type="checkbox"/> Sonderfall <input type="checkbox"/>		Zugeordnet zu Masterprofil Sicherheit und Qualität (SQ) <input type="checkbox"/> KI, Kognition, Robotik (KIKR) <input type="checkbox"/> Digitale Medien und Interaktion (DMI) <input type="checkbox"/>	
Modulbereich: Mathematik und Theoretische Informatik Modulteilbereich: (keine Angabe)			
Anzahl der SWS	V UE K S Prak. Proj. Σ 3 1 0 0 0 0 4	Kreditpunkte: 6	Turnus angeboten in jedem SoSe
Formale Voraussetzungen: -			
Inhaltliche Voraussetzungen: Inhalte von Mathematik 1			
Vorgesehenes Semester: 4. Semester			
Sprache: Deutsch			

Ziele:

- Fähig sein, mathematische Notation zu verstehen und zu verwenden.
- Im Stande sein, über mathematische Gegenstände und Sachverhalte zu kommunizieren.
- Logisches Denken und Abstraktionsfähigkeit trainiert haben.
- Mit den für die Informatik wichtigen Grundlagen der Wahrscheinlichkeitstheorie und Statistik vertraut sein, die elementaren Resultate aus diesen Gebieten kennen und sie anwenden können.
- Beweise verstehen, nachvollziehen und selbstständig durchführen können.

Inhalte: .

1 Zufall und Wahrscheinlichkeit:

- Bayesscher und frequentistischer Wahrscheinlichkeitsbegriff
- Bedingte Wahrscheinlichkeiten Stochastische Unabhängigkeit

2 Diskrete Verteilungen:

- Laplace-Verteilung
- Poisson-Verteilung
- Binomialverteilung

3 Stetige Verteilungen:

- Normal- und Standardnormalverteilung 1 Student-t-Verteilung
- Chi-Quadrat-Verteilung

4 Parameter:

- Erwartungswert, Standardabweichung, Varianz
- Ungleichungen (Chebychev, Chernov, Markov)

5 Deskriptive Statistik:

- Skalentypen
- Modus, Median, Mittelwert, empirische Varianz

6 Schließende Statistik:

- Punkt- und Intervallabschätzung des Erwartungswerts
- Varianzschätzer
- Hypothesentests

Unterlagen (Skripte, Literatur, Programme usw.):

- L. Fahrmeir, C. Heumann, R Künstler, I. Pigeot, G. Tutz: Statistik - Der Weg zur Datenanalyse, Springer, 2016.
- Hans-Otto Georgii: Stochastik. Einführung in die Wahrscheinlichkeitstheorie und Statistik, De Gruyter, 2015
- Klenke: Wahrscheinlichkeitstheorie, Springer, 2013

Form der Prüfung:

KP, PL1: xx%, PL2: xx%, Portfolio, Klausur

Arbeitsaufwand	Präsenz	56 h
	Übungsbetrieb/Prüfungsvorbereitung	124 h
	Summe	180 h

Lehrende:
SG Mathematik: N.N.

Verantwortlich:
Prof. Dr. C. Lutz

Durchführungshinweise

- Seit SoSe 2023 geändertes Format:
2V + 2Ü statt 3V + 1Ü
- Vorlesung ist montags 8h-10h in HS 1010
(Keksdose, Kleiner Hörsaal).
- Es gibt vier Übungsgruppen:
 - Donnerstags 12h-14h in MZH 1380/1400 (Senatssaal)
 - Donnerstags 12h-14h in MZH 1090
 - Donnerstags 14h-16h in MZH 1090
 - Freitags 12h-14h in NW2 C0290 (Hörsaal 1)

Durchführungshinweise

- Seit SoSe 2023 geändertes Format:
2V + 2Ü statt 3V + 1Ü
- Vorlesung ist montags 8h-10h in HS 1010
(Keksdose, Kleiner Hörsaal).
- Es gibt vier Übungsgruppen:
 - Donnerstags 12h-14h in MZH 1380/1400 (Senatssaal)
 - Donnerstags 12h-14h in MZH 1090
 - Donnerstags 14h-16h in MZH 1090
 - Freitags 12h-14h in NW2 C0290 (Hörsaal 1)

Durchführungshinweise

- Seit SoSe 2023 geändertes Format:
2V + 2Ü statt 3V + 1Ü
- Vorlesung ist montags 8h-10h in HS 1010
(Keksdose, Kleiner Hörsaal).
- Es gibt vier Übungsgruppen:
 - Donnerstags 12h-14h in MZH 1380/1400 (Senatssaal)
 - Donnerstags 12h-14h in MZH 1090
 - Donnerstags 14h-16h in MZH 1090
 - Freitags 12h-14h in NW2 C0290 (Hörsaal 1)

Fachliche (mathematische) Voraussetzungen (I)

■ Mengenlehre

(Menge, Teilmenge, Komplement, Durchschnitt, Vereinigung, Mengensysteme, Potenzmenge, ...)

■ Sicherer Umgang mit Termen und (Un-)Gleichungen

(Binomische Formeln, Betragsstriche auflösen, Potenzgesetze, Funktionalgleichungen von Logarithmus- und Exponentialfunktionen, Gleichungen und Ungleichungen lösen können, ...)

Fachliche (mathematische) Voraussetzungen (II)

- Folgen und Reihen, Konvergenzuntersuchung
- Analysis in einer reellen Veränderlichen
 - (Erste und zweite Ableitung, Extrema, Wendepunkte, Integralrechnung, partielle Integration, Integration durch Substitution, ...)
- Lineare Algebra und Matrizenrechnung
 - (Matrix-Vektor-Multiplikation, Matrizenmultiplikation, Invertierbarkeit, Determinante, ...)

Prüfungsmodalitäten

- Studienleistungen (25% der Modulnote):
Mindestens 50% der insgesamt zu erzielenden Punkte aus den Übungsaufgaben müssen erreicht werden.
Es soll nach Möglichkeit in festen Dreierteams abgegeben werden.
- Prüfungsleistungen (75% der Modulnote):
Die Modulabschlussprüfung wird als schriftliche Prüfung (Klausur) durchgeführt.

Organisation des Übungsbetriebs

- In jeder Lehrveranstaltungswoche ab der zweiten wird montagnachmittags ein Übungsblatt via Stud.IP zur Verfügung gestellt (Blatt 1 also am 14.04.2025).
- Dieses Übungsblatt wird in den Übungsstunden der gleichen Woche vorbesprochen.
- Die Bearbeitungszeit endet an dem Dienstag zwei Wochen nach Ausgabe des Blattes (für Blatt 1 also am 29.04.2025 um 23:59h).
- Eine weitere Woche später erhalten Sie Ihre korrigierten Zettel zurück und es werden in der Übungsstunde Musterlösungen besprochen.

Organisation des Übungsbetriebs

- In jeder Lehrveranstaltungswoche ab der zweiten wird montagnachmittags ein Übungsblatt via Stud.IP zur Verfügung gestellt (Blatt 1 also am 14.04.2025).
- Dieses Übungsblatt wird in den Übungsstunden der gleichen Woche **vorbesprochen**.
- Die Bearbeitungszeit endet an dem Dienstag **zwei Wochen nach Ausgabe des Blattes** (für Blatt 1 also am 29.04.2025 um 23:59h).
- Eine weitere Woche später erhalten Sie Ihre korrigierten Zettel zurück und es werden in der Übungsstunde Musterlösungen besprochen.

Organisation des Übungsbetriebs

- In jeder Lehrveranstaltungswoche ab der zweiten wird montagnachmittags ein Übungsblatt via Stud.IP zur Verfügung gestellt (Blatt 1 also am 14.04.2025).
- Dieses Übungsblatt wird in den Übungsstunden der gleichen Woche **vorbesprochen**.
- Die Bearbeitungszeit endet an dem Dienstag **zwei Wochen nach Ausgabe des Blattes** (für Blatt 1 also am 29.04.2025 um 23:59h).
- Eine weitere Woche später erhalten Sie Ihre korrigierten Zettel zurück und es werden in der Übungsstunde Musterlösungen besprochen.

Organisation des Übungsbetriebs

- In jeder Lehrveranstaltungswoche ab der zweiten wird montagnachmittags ein Übungsblatt via Stud.IP zur Verfügung gestellt (Blatt 1 also am 14.04.2025).
- Dieses Übungsblatt wird in den Übungsstunden der gleichen Woche **vorbesprochen**.
- Die Bearbeitungszeit endet an dem Dienstag **zwei Wochen nach Ausgabe des Blattes** (für Blatt 1 also am 29.04.2025 um 23:59h).
- Eine weitere Woche später erhalten Sie Ihre korrigierten Zettel zurück und es werden in der Übungsstunde Musterlösungen besprochen.

Übersicht

- 1 Inhaltliche Einführung**
- 2 Ablauf und Formalitäten**
- 3 Zusätzliche Literatur**



