

Prof. Dr. Timo Schmid - Institut für Statistik und Ökonometrie,
Freie Universität Berlin, Garystr. 21, 14195 Berlin

To whom it may concern

**Inst. f. Statistik u. Ökonometrie
FB Wirtschaftswissenschaft**

Prof. Dr. Timo Schmid
Garystraße 21
14195 Berlin

Telefon +49 30 838-51084

Fax +49 30 838-456629

E-Mail timo.schmid@fu-berlin.de

Bearb.-Zeichen

Bearbeiter/in

10.01.2014

Seminar: Rechnerintensive Statistik mit R und JMP

Das Seminar „Monte-Carlo Simulationsmethoden in R“ wird von Prof. Dr. Rendtel und Prof. Dr. Schmid angeboten. Im Rahmen des Seminars erhalten Studierende aus dem Masterstudiengang „Statistik“ oder dem „Master of Science in Economics“ die Möglichkeit, ihr Wissen auf dem Gebiet rechnerintensiver Methoden zu erweitern. Ein Teil der Vorträge steht auch für Teilnehmer aus dem Master Wirtschaftsinformatik offen. Das Seminar ist als semesterbegleitende Einheit ausgelegt und beinhaltet eine Posterpräsentation sowie die Anfertigung einer schriftlichen Ausarbeitung.

In diesem Semester beschäftigt sich das Seminar für Angewandte Statistik mit rechnerintensiven Methoden. Im ersten Teil wird anhand dreier Lehrbücher und unter Verwendung von R mit der Erzeugung von Zufallszahlen und Ihrer Anwendung bei der Integration, der Optimierung und bei der Schätzung von posteriori Verteilungen. Hierzu werden der Metropolis-Hastings Algorithmus, der Gibbs-Sampler und der Markov-Chain-Monte-Carlo Algorithmus verwendet.

Seit Jahreswechsel steht Studierenden und Mitarbeitern der FU das neue Softwarepaket JMP zur Verfügung, das einige neuere Verfahren aus dem Bereich des Data Mining anbietet. Dieser Teil des Seminars steht auch Teilnehmern aus dem Masterprogramm Wirtschaftsinformatik offen. Das klassische Verfahren der Entscheidungsbäume wird auf ganze Wälder ausgedehnt. Die Konstruktion der Wälder folgt dabei unterschiedlichen Heuristiken. Bei Regressionsanalysen mit großen Datensätzen ist manchmal die Anzahl der erklärenden Merkmale sehr groß. Es kommt dabei vor, dass die Anzahl der Beobachtungen übertroffen wird, so dass klassische Schätzer nicht existieren. Oder die erklärenden Merkmale sind so stark miteinander korreliert, dass massive Schätzprobleme auftreten. Beim Lasso-Ansatz wird neben dem üblichen kleinste Quadrate Kriterium ein Strafterm aus dem Absolutbetrag (L_1 Norm) des geschätzten Parametervektors minimiert. Netze fügen einen weiteren Strafterm auf Basis der Euklidischen Länge des Parametervektors (L_2 Norm) hinzu.

Veranstalter:

Die Professur „Statistik“ von Ulrich Rendtel gehört dem Institut für Statistik und Ökonometrie am Fachbereich Wirtschaftswissenschaft an. Die Forschungsschwerpunkte der Professur liegen in den Bereichen Panel-Erhebungen, Nonresponse, Panel-Attrition, Messfehler, Imputation, Gewichtung und Kalibration.

Die Professur „Angewandte Statistik“ von Timo Schmid gehört ebenfalls dem Institut für Statistik und Ökonometrie am Fachbereich Wirtschaftswissenschaft an. Die Schwerpunkte in der Forschung liegen in den Bereichen Small Area Estimation, Robuste und räumliche Verfahren, Varianzschätzmethoden und Untersuchung von Unternehmensdaten.

Termine:

03.02.2014, 16.00 – 18.00 Uhr	Unverbindliche Informationsveranstaltung zum Seminar (Raum 315)
17.02.2014	Frist für die Anmeldung zum Seminar
25.02.2014	Vergabe der Themen
25.03.2014	Frist für den Rücktritt vom Seminar
10.07.2014	Präsentationstermin
27.08.2014	Frist für die Abgabe der Seminararbeit (ca. 20 Seiten)

Themenliste:

1. Erzeugung von Zufallszahlen
2. Monte-Carlo Integration
3. Konvergenzgeschwindigkeit und -kontrolle
4. Monte-Carlo Optimierung
5. Metropolis-Hastings Algorithmen
6. Gibbs-Sampler
7. Anwendung von Gibbs-Samplern zur Berechnung der Posterior-Verteilung
8. Diagnostik bei MCMC Algorithmen
9. Ein Ausflug in die Bootstrap Wälder: Entscheidungsbäume, Random Forests, Bagging und Boosting.
10. Mit Lasso und Netzen: Regressionsanalyse mit hochdimensionalen Daten.

Als Grundlage für die meisten Themen werden folgende Bücher verwendet:

- C. Robert and G. Casella (2009), *Introducing Monte Carlo Methods with R*, Springer, New York.
- E. Suess and B. Trumbo (2010), *Introduction to Probability Simulation and Gibbs Sampling with R*, Springer, New York.
- G. Fishman (2006), *A First Course in Monte Carlo*, Duxbury, London.
- T. Hastie, R. Tibshirani and J. Friedman (2009): *The Elements of Statistical Learning*, Second Edition. Springer, Berlin.
- P. Bühlmann and S. v.d. Geer (2011): *Statistics for High-Dimensional Data*, Springer Berlin

- L. Breiman (1996): Bagging Predictors, Maschine Learning, 24,123-140.
- L. Breiman (2001): Random Forests, Maschine Learning, 45, 5-32.

Teilnehmerkreis:

An dem Seminar können Studierende, die im „Master of Science in Statistics“, dem „Master of Science in Economics“ oder dem Master Wirtschaftsinformatik eingeschrieben sind, teilnehmen. Die Teilnehmerzahl ist auf 10 Studierende begrenzt.

Teilnahmevoraussetzungen:

Studierende, die an dem Seminar teilnehmen, sollten im SoSe 2014 mindestens im 2. Fachsemester (Master) sein und Erfahrungen in den Bereichen deskriptive Statistik und multiple lineare Regression mitbringen. Darüber hinaus ist der sichere Umgang mit R für die Themen 1 bis 8 von Vorteil. Des Weiteren ist die Bereitschaft sich vertiefend mit einem über die Grundlagen hinausgehenden Bereich der Statistik zu beschäftigen eine entscheidende Voraussetzung.

Ablauf des Seminars:

Für Interessenten wird am 03.02.2014 eine Informationsveranstaltung angeboten, in deren Rahmen das Thema des Seminars vorgestellt wird und die Gelegenheit geboten wird, Fragen zum Seminar zu stellen. Der Besuch der Veranstaltung ist freiwillig und unverbindlich. Interessenten schicken bis zum 17.02.2014 einen Informationsbogen mit möglichem Thema per E-Mail an timo.schmid@fu-berlin.de (für die Anmeldung wird ein PDF-Formular zur Verfügung gestellt). Nach einer Prüfung der Unterlagen werden die zugelassenen Teilnehmer bis zum 25.02.2014 über ihr Thema und Betreuer informiert. Sollten sich mehr als 10 Personen bewerben, erfolgt die Auswahl anhand der Noten sowie der nachgewiesenen Kenntnisse.

Am 10.07.2014 stellt jeder Teilnehmer sein Thema in Form eines Posters in einer 20-minütigen Präsentation vor, der jeweils eine ca. 10-minütige Diskussion folgt. Bis zum 27.08.2014 reichen die Teilnehmer eine schriftliche Ausarbeitung (ca. 20 Seiten) zu dem betreuten Thema ein.

Benotung:

Die Abschlussnote bestimmt sich zu 80% aus der Bewertung des Posters und dessen schriftlicher Ausarbeitung und zu 20% aus der Bewertung der Präsentation/Diskussion. Für die Teilnahme am Seminar werden 5 ECTS vergeben.

Kontakt für Fragen im Vorfeld:

Ansprechpartner: Prof. Dr. Timo Schmid
Sprechzeiten: Freitag, 13-14 Uhr, nach Vereinbarung
E-Mail: timo.schmid@fu-berlin.de