



Psychoco – International Workshop on Psychometric Computing

February 9–10, 2012

Universität Innsbruck, Austria

Thursday, February 9

13:00	<i>Welcome</i>	
13:15	Rune HB Christensen	Psychometric and Statistical Models in the R Packages <i>sensR</i> and <i>ordinal</i>
14:00	Florian Wickelmaier	Parameter Estimation in Probabilistic Knowledge Structures with the <i>pks</i> Package
14:30	Edgar C Merkle	Generalized Measurement Invariance Tests for Factor Analysis

Coffee break

15:30	Ingrid Koller	Quasi-Exact Tests for the Rasch Model in <i>eRm</i>
16:00	Kathrin Gruber	Generating Matrices with Ordinal Responses and Fixed Margins
16:30	David Firth	Bias Reduction in the Estimation of Rasch Models
17:00	Hannah Frick	Mixtures of Rasch Models with R Package <i>psychomix</i>

Coffee break

18:00	Darja Tutschkow	A Bayesian Test for the Hot Hand Phenomenon
18:30	Cristina Bostan	What Do We Know about Health Based on Capacity vs. Performance?

Dinner (Stiftskeller)

Friday, February 10

09:00	Yves Rosseel	<i>lavaan</i> and the History of Structural Equation Modeling
09:30	Tobias Koch	A Simulation Study on the Performance of a Longitudinal Multilevel CTC($M - 1$) Model Using <i>MplusAutomation</i>
10:00	Manuel JA Eugster	A Unified Syntax for Structural Equation Modeling

Coffee break

11:00	Ingo Nader	Investigating Dimensionality by Means of Mokken Analysis and Confirmatory Factor Analysis
11:30	Rainer Alexandrowicz	Comparing Missing Values Handling Algorithms in the Context of the Rasch Model
12:00	Karoline Sachse	<i>eat</i> : An R Package for Automation of Data Preparation and IRT Modeling

Lunch break

14:00	Carolin Strobl	Detecting Differential Item and Differential Step Functioning with Partial Credit Trees
14:30	Georg Hosoya	Maximum Entropy Models: A Framework for IRT Modeling?
15:00	Ivailo Partchev	Estimating Hierarchical Structure GLMM Item Response Models in R

Dinner (Cammerlander)

Thursday, February 9, 13:00–15:00

Psychometric and Statistical Models in the R Packages *sensR* and *ordinal*

Rune HB Christensen (Technical University of Denmark)

The development of the R packages *sensR* and *ordinal* were motivated by problems in modeling sensory and signal detection theory (SDT) discrimination tasks. On a basic level *sensR* provides the means for standard difference and similarity testing, d' estimation and sample size estimation in sensory discrimination protocols such as the triangle and m-AFC tasks. On a more advanced level *sensR* facilitates modeling of psychometric and sensory discrimination experiments with generalized linear models (GLMs) where special purpose link functions derived from the psychometric functions for the different discrimination protocols directly relate success probability to the underlying Thurstonian δ aka d' (d-prime).

The *ordinal* package was similarly motivated by the need for fitting so-called unequal variance models to rating experiments (known as the A-not A with sureness task in sensometrics), but have since developed into a much more generally applicable statistical R package. The *ordinal* package now estimates a wide range of variations of cumulative link models (CLMs) including the famous proportional odds model; it estimates location-scale models, allows for partial and non-proportional odds, has `anova()`, `drop1()/add1()`, `profile()`, `confint()`, `predict()` and other convenience methods. A rather unique feature is that all these variants of CLMs can also be fitted with random effects, that is, as cumulative link mixed models (CLMMs). As such *ordinal* extends `glmer()` from *lme4* for GLMMs to ordinal responses.

This talk will focus not so much on the application of the *sensR* and *ordinal* packages, but rather on implementation considerations; the interesting problems we have met during development and the solutions we have attained. One focus point in the development of *ordinal* has been the implementation of the wide range of models included in the class of CLMs. It turns out that many seemingly unrelated variants of CLMs can be embedded in a common framework and a single efficient estimation routine copes with them all. Another focus point has been to give the user the opportunity to assess the accuracy of fitted models. This is partly motivated by unreliable results from other R implementations of CLMs of which I will show a few examples.

References:

Christensen RHB (2011). *ordinal: Regression Models for Ordinal Data*. R package version 2011.09-14. <http://CRAN.R-project.org/package=ordinal>.

Christensen RHB, Brockhoff PB (2011). *sensR: An R Package for Thurstonian Modelling of Discrete Sensory Data*. R package version 1.2.13. <http://CRAN.R-project.org/package=sensR>.

Parameter Estimation in Probabilistic Knowledge Structures with the *pks* Package

Florian Wickelmaier (Eberhard-Karls-Universität Tübingen), *Jürgen Heller* (Eberhard-Karls-Universität Tübingen)

Practical applications of the theory of knowledge structures often rely on a probabilistic version known as the basic local independence model (Doignon & Falmagne, 1999). This presentation outlines various procedures for estimating its parameters, including maximum likelihood (ML) via the EM algorithm, the computationally efficient minimum discrepancy (MD) estimation as well as MDML, a hybrid method combining the two approaches. Implementation of these parameter estimation procedures in the R package *pks* is discussed.

References:

Doignon JP, Falmagne JC (1999). *Knowledge Spaces*. Springer-Verlag, Berlin.

Generalized Measurement Invariance Tests for Factor Analysis

Edgar C Merkle (University of Missouri), *Achim Zeileis* (Universität Innsbruck)

The issue of measurement invariance commonly arises in factor-analytic contexts, with methods for assessment including likelihood ratio tests, Lagrange multiplier tests, and Wald tests. These tests all require advance definition of the number of groups, group membership, and offending model parameters. In this talk, we construct tests of measurement invariance using individuals' scores (i.e., casewise gradients of the likelihood function) from the estimated factor analysis model. These tests can be viewed as generalizations of the Lagrange multiplier test, and they are especially useful for: (1) isolating specific parameters affected by measurement invariance violations and (2) identifying subgroups of individuals that violated measurement invariance based on a continuous auxiliary variable. The tests are presented and illustrated in detail,

along with simulations examining the tests' abilities in controlled conditions. To carry out the tests, we use lavaan for model estimation, strucchange for testing, and some general methods for obtaining scores from lavaan objects.

Thursday, February 9, 15:30–17:30

Quasi-Exact Tests for the Rasch Model in eRm

Ingrid Koller (University of Vienna), *Reinhold Hatzinger* (Vienna University of Economics and Business)

Quasi-exact tests for the dichotomous Rasch model, introduced by Pononcy (2001), allow for the examination of the well-known properties of the model, even with small samples. The distributions of the test statistics are approximated by a Monte Carlo simulation algorithm, which generates sample pseudo-random matrices with fixed marginals from the sample space of all matrices with the same marginals as in an observed data matrix. Recently, some test statistics using the Markov chain Monte Carlo (MCMC) approach of Verhelst (2008) have been implemented in the R package eRm (Hatzinger, Mair, & Maier, 2011). In this talk, we present some test statistics for various model deviations (e.g., response dependence, trait dependence, subgroup invariance, unidimensionality, sufficiency of the raw scores, etc.) and discuss the relation to parametric tests for the Rasch model. Additionally, we present some results on the empirical power for some of the test statistics. The results suggest the application of some of the quasi-exact tests to evaluate Rasch model assumptions, particularly in the case of small samples.

Generating Matrices with Ordinal Responses and Fixed Margins

Kathrin Gruber (Vienna University of Economics and Business), *Reinhold Hatzinger* (Vienna University of Economics and Business)

Many statistical and mathematical problems, like creating a probabilistic basis for, or removing the effect of nuisance parameters on a test involve sampling matrices with given row and column sums. Also when the asymptotic distribution of a test statistic is unclear or unknown it is useful to build nonparametric tests by approximating the null distribution (the distribution if the model is valid) of the statistical model. In this talk a generalization of Verhelst (2008) for simulating binary data matrices with given row and column sums will be presented. The aim is to sample discrete data matrices with fixed marginals to build exact nonparametric tests for applications in the Rasch model context.

The item totals ($\sum_i t_{ij}$) and person scores ($\sum_j t_{ij}$) are minimal sufficient statistics for the item difficulties. A subsequent feature is the separation of parameters, that means the conditional probability of the data given the person scores only depends on the item difficulty parameters β_j (and same holds vice versa for the ability parameters). Rasch proposed that his model should be tested by conditioning on person scores as well as item totals. That means a resulting test would be specifically objective. As for the class of Rasch models the parameter estimates only depend on the marginal totals of the data matrix the following should also be possible for the Partial credit model (Masters, 1982) where responses to items are scored with successive integers.

If one can simulate from the uniform (or nearly uniform) distribution then the total number of counts of the tables, as well as an approximate of the distribution of the test statistic can be estimated. This is useful because in small samples parameter estimates may be inaccurate or the model tests based on an asymptotic distributions may be misleading. Since the probability for deviation from a null hypothesis (for e.g., the Rasch model is valid) can be calculated exactly by sampling matrices with the same marginals, this method is even adequate for small samples. The challenge is to derive a computational fast and efficient algorithm which can be generalized to n categorical responses. We try to model the responses by using a Markov-chain and defining a transition matrix and some special permutation rules (Verhelst, 2008).

In this talk some attempts for solving the problem of generating data matrices with ordinal responses and fixed margins are presented and will be compared in computational and theoretical performance.

Bias Reduction in the Estimation of Rasch Models

David Firth (University of Warwick), *Ioannis Kosmidis* (University College London), *Heather Turner* (University of Warwick)

Rasch models are popular models for analysing binomial outcomes in a subject-item arrangement. As with all models for binomial responses, the possibility of boundary (usually infinite) maximum likelihood estimates can cause havoc with asymptotic inference (e.g., Wald-type confidence intervals).

To address this problem we apply the bias reduction method of Firth (1993), which has previously been shown to eliminate the possibility of boundary estimates in the case of logistic regression.

Using the results of Kosmidis and Firth (2009, 2010), simple procedures are developed for obtaining the reduced-bias estimates in non-linear (2PL) Rasch models via the introduction of appropriate pseudo-responses. These procedures rely merely on the availability of software for maximum likelihood estimation (for example, the *gnm* R package) and are a direct extension of the procedure that was suggested in Firth (1992) for logistic regressions.

We demonstrate that the desirable properties of the reduced-bias estimates in logistic regression also apply to Rasch-type models and explore the impact of this approach on inference.

The reduced-bias estimator is used for the analysis of data on twenty roll calls in the US House of Representatives, 2001 (ADA, 2006).

References:

Americans for Democratic Action, ADA (2002). "2001 Voting Record: Shattered Promise of Liberal Progress." *ADA Today* **57**(1), 1–17.

Firth D (1992). "Bias Reduction, the Jeffreys Prior and GLIM," *Advances in GLIM and Statistical Modelling: Proceedings of the GLIM 92 Conference*. Munich, Springer-Verlag, 1992, 91–100.

Firth D (1993). "Bias Reduction of Maximum Likelihood Estimates." *Biometrika*, **80**(1), 27–38.

Kosmidis I, Firth D (2009). "Bias Reduction in Exponential Family Nonlinear Models." *Biometrika*, **96**(4), 793–804.

Kosmidis I, Firth D (2010). "A Generic Algorithm for Reducing Bias in Parametric Estimation." *Electronic Journal of Statistics*, **4**, 1097–1112.

Turner H, Firth D (2011). *Generalized Nonlinear Models in R: An Overview of the gnm Package*. R package version 1.0-1. <http://CRAN.R-project.org/package=gnm>.

Mixtures of Rasch Models with R Package psychomix

Hannah Frick (Universität Innsbruck), *Carolin Strobl* (University of Zurich), *Friedrich Leisch* (Universität für Bodenkultur Wien), *Achim Zeileis* (Universität Innsbruck)

Measurement invariance is an important assumption in the Rasch model and mixture models constitute a flexible way of checking for a violation of this assumption by detecting unobserved heterogeneity in item response data. Here, a general class of Rasch mixture models is established and implemented in R, using conditional maximum likelihood estimation of the item parameters (given the raw scores) along with flexible specification of two model building blocks: (1) Mixture weights for the unobserved classes can be treated as model parameters or based on covariates in a concomitant variable model. (2) The distribution of raw score probabilities can be parametrized in several ways, including a saturated model or a specification through mean and variance. Additionally, we suggest a conditional-type Rasch mixture model which is free of a specific score model as this is usually not of particular interest when checking for measurement invariance. The various model flavors are implemented in the R package *psychomix*, leveraging the general infrastructure for fitting mixture models in the *flexmix* package. The usefulness and relative advantages of the different score distributions – as well as those of concomitant variables, if available – are explored for artificial data scenarios as well as empirical data from a study of verbally aggressive behavior.

Thursday, February 9, 18:00–19:00

A Bayesian Test for the Hot Hand Phenomenon

Darja Tutschkow (University of Tuebingen), *Conor Dolan* (University of Amsterdam), *Gilles Dutilh* (University of Amsterdam), *Ruud Wetzels* (University of Amsterdam), *Sophie van der Sluis* (UV University Amsterdam), *Eric-Jan Wagenmakers* (University of Amsterdam)

In many sports it may appear that performance is streaky, as players can alternate runs of good performance with runs of poor performance. The idea that a player can be either in a hot state (i.e., perform well) or in a cold state (i.e., perform poorly) is known as the "hot hand phenomenon". Here we propose a Bayesian test to quantify the statistical evidence for and against the hot hand phenomenon. Specifically, we used the Bayes factor to compare a three-parameter two-state Bernoulli Hidden Markov Model (HMM) to a baseline model that assumes constant performance. The HMM has two parameters that represent the probability of a "hit" in each state and a third parameter that represents the probability of staying in a state. The advantage

of using an HMM in the context of the hot hand phenomenon (as opposed to commonly used statistics such as the length of the longest run of successes) is that the HMM structurally corresponds to the definition of a streaky player. The advantage of using the Bayes factor is that it naturally accounts for differences in model flexibility. Performance of the new test in comparison to commonly used statistics is explored in simulation studies and real data examples.

What Do We Know about Health Based on Capacity vs. Performance?

Cristina Bostan (Swiss Paraplegic Research), *Cornelia Oberhauser* (Ludwig-Maximilians-University), *Alarcos Cieza* (Swiss Paraplegic Research, Ludwig-Maximilians-University)

Background: Capacity and performance are terms used in the International Classification of Functioning, Disability and Health (ICF) which refer to different aspects of functioning. According to the definition of the ICF, capacity is the level of functioning in health and health-related domains measured as the internal capacity of a person and performance is the level of functioning of a person in the current environment. In practice it is important to determine the magnitude of the differences between capacity and performance, and the influence of possible contextual factors, in particular environmental factors, which could explain these differences.

Objective: The objective of this study is to determine the differences between capacity and performance by means of an innovative statistical approach based on which the influence of environmental factors in those differences can be analyzed.

Method: This is a psychometric study on a sample of 297 patients with low back pain, osteoarthritis, osteoporosis, rheumatoid arthritis, and chronic widespread pain. The following steps were performed: 1) Rasch analyses were used to create two interval measurement scales of functioning, one for capacity and the other for performance; 2) Item Response Theory (IRT) Calibration was used to calibrate both scales of step 1 on a common metric scale ranging from 0 (low level) to 100 (high level) and to make possible the calculation of the differences between the patients' capacity and performance. ICF categories of the body-functions ICF component, which are common for both scales, capacity and performance, were used as anchor items. 3) Group Lasso method was used for identifying the environmental factors that best explain the differences between the patients' capacity and performance.

Results: 23 capacity ICF categories and 26 performance ICF categories fitted the Rasch model and two different scales could be built based on them. After IRT calibration, the differences between performance and capacity ranged from 13.21 to 42.68 on the scale from 0 to 100. The environmental facilitators that best explain that differences are: e410 - Individual attitudes of immediate family members; e540, e570 - Transportation and social security services, systems and policies. The environmental barrier explaining that differences is the lack of support from Health Professionals (e355).

Conclusion: This study presents a list of relevant social and physical environmental barriers or facilitators that may influence the performance of actions and tasks in daily life.

Friday, February 10, 09:00–10:30

lavaan and the History of Structural Equation Modeling

Yves Rosseel (Ghent University)

The lavaan package can be used for structural equation modeling (SEM) with continuous data. Unlike many R packages, which are often created to implement new cutting-edge ideas or state-of-the-art statistical algorithms, the lavaan package is in many areas still trying to catch up with commercial software. In a sense, lavaan tries to capture a bit of the history in the SEM literature by implementing not just one, but several traditions that largely correspond with the three (commercial) programs LISREL, EQS and Mplus. All fitting functions in lavaan have a `mimic` argument which can be set to "LISREL", "EQS" or "Mplus" respectively. Depending on the `mimic` option, lavaan tries to mimic many of the computational details that are used by the corresponding original program.

During the presentation, I will argue that by studying the history of SEM at the computational level, one can learn many things that are easily overlooked if one only studies the theory.

A Simulation Study on the Performance of a Longitudinal Multilevel CTC($M - 1$) Model Using MplusAutomation

Tobias Koch (Freie Universität Berlin), *Martin Schultze* (Freie Universität Berlin) *Christian Geiser* (Utah State University)

Measurement designs can be complex. For example, different methods can be used to assess different traits on different occasions of measurement. In addition, some types of methods (interchangeable methods) may also imply a hierarchical data structure. According to Eid et al. (2008) interchangeable methods are methods that are randomly drawn out of set of equivalent methods (e.g., student ratings). In contrast, structurally different methods are methods that are not randomly drawn out of a common set of equivalent methods, but are fixed beforehand (e.g., teacher, student, parent ratings). Koch et al. (2011) defined a longitudinal multitrait-multimethod structural equation model for the combination for interchangeable and structurally different methods. In order to investigate the performance of this model a complex simulation study with 144 conditions à 500 replications was carried out. The simulations were done in Mplus using R (packages: MplusAutomation and OpenMx) to automate runs and extract a comprehensive overview of detailed results. Overall, results indicate that larger numbers of interchangeable methods, increases in sample size and more measurement occasions stabilize estimates and decrease estimation problems.

A Unified Syntax for Structural Equation Modeling

Armin Monecke (Ludwig-Maximilians-Universität München), *Manuel JA Eugster* (Ludwig-Maximilians-Universität München)

The specification of structural equation models is not easy – and consequently each R software package provides its own specification dialect. This can be frustrating for the researcher as no automatic translation is available.

We present and discuss first steps towards an unified domain specific language for the specification of structural equation models within R. It uncouples the model specification (equal for all packages) from the model representation (specific for each package). The language is based on the well-established R formula objects and is easily extensible. We present its usage for the sem, lavaan, and semPLS packages.

Friday, February 10, 11:00–12:30

Investigating Dimensionality by Means of Mokken Analysis and Confirmatory Factor Analysis

Ingo W Nader (University of Vienna), *Ulrich S Tran* (University of Vienna), *Patricia Baranyai* (University of Vienna), *Martin Voracek* (University of Vienna)

Dimensionality of psychometric scales is usually investigated using exploratory factor analysis (EFA), and contradictory results are often resolved by applying confirmatory factor analysis (CFA). With CFA, different and competing models can be compared in terms of their fit to the data, using a number of well-established fit indices. Mokken analysis, a nonparametric method within the framework of item response theory (IRT), offers an alternative approach for exploring dimensionality. Specifically, it allows for an automated selection of items that measure the same latent construct. Within this automated selection algorithm, the criterion that is used to judge whether an item belongs to a scale or not (i.e., the criterion of unidimensionality) can be varied, and this approach can be used to gain information about the dimensionality that cannot be gained as easily when using factor analysis. To allow comparisons with prior research, the fit of models resulting from Mokken analysis can be investigated by means of CFA.

This study investigated the dimensionality of a questionnaire on attitudes towards suicide in a German-speaking sample of $N = 571$ participants. The questionnaire measures, amongst other attitudes, “acceptability of suicide”, attitudes towards “communicating about suicide”, and views of “suicide as a solution”. Two prior studies (Eskin, 2004; Eskin, Voracek, Stieger, & Altinyazar, 2011) have found different factor structures for this questionnaire, with only one of these finding the factor “suicide as a solution”.

Mokken analysis revealed that the factor of seeing “suicide as a solution” is a facet of the factor “acceptability of suicide”, as this factor emerged only for very strict criteria of unidimensionality. The fit of different and competing models was additionally compared with CFA.

Thus, our results suggest that the different factor structures in prior research were caused by a different response behavior regarding “suicide as a solution” items in the two samples. This finding was facilitated by the use of Mokken analysis, whereas the additional use of CFA allowed for comparison of competing

models. Hence, conjointly using Mokken analysis and CFA can benefit investigations on the dimensionality of scales, combining the advantages of nonparametric item response theory with a well-established method to assess the fit of different (competing) models.

Comparing Missing Values Handling Algorithms in the Context of the Rasch Model

Rainer Alexandrowicz (University of Klagenfurt)

The Rasch family of models applies the conditional maximum likelihood estimation (CML) method, which, among others, allows for taking structurally missing values into consideration in an adequate way. Basically, a matrix is incorporated, indicating which item-respondent confrontation has actually happened and which not. By means of this matrix, the respective elements of the elementary symmetric function are chosen appropriately (Molenaar, 1995). Under weak conditions, all item parameters can be estimated on a common scale when items have not been presented to some participants in a prespecified manner (e.g., the use of testlets).

However, aside of taking structurally missing values into account, the method is also being applied to missing values appearing during data acquisition. Such an ad-hoc method is prone to bias as the cause of the missingness in the sense of Rubin (1976) is not being considered (cf. Schafer, 1997; Enders, 2010). Numerous methods of handling the different cases have been developed so far, however, they are rarely used when applying Rasch models.

In a simulation study the effects of several missing data handling mechanisms are compared to the CML-method with respect to their impact upon parameter estimates and model fit measures, taking the taxonomy of Rubin into account.

References:

Enders CK (2010). *Applied Missing Data Analysis*. Guilford, NY.

Molenaar IW (1995). "Estimation of Item Parameters." In: Fischer GH, Molenaar IW (Eds.) *Rasch Models. Foundations, Recent Developments, and Applications*. (pp. 39–51). Springer-Verlag, NY.

Rubin DB (1976). "Inference and missing data." *Biometrika*, **63**, 581–592.

Schafer JL (1997). *Analysis of Incomplete Multivariate Data*. Chapman & Hall/CRC, Boca Raton.

eat: An R Package for Automation of Data Preparation and IRT Modeling

Karoline Sachse (Humboldt-University), *Martin Hecht* (Humboldt-University), *Sebastian Weirich* (Humboldt-University), *Nicole Haag* (Humboldt-University), *Malte Jansen* (Humboldt-University), *Sebastian Wurster* (Humboldt-University), *Christiane Penk* (Humboldt-University), *Anna Lenski* (Humboldt-University), *Thilo Siegle* (Humboldt-University)

Large-scale assessments require extensive preparation of complex data, which is time-consuming and tedious. Also, syntax specification in standard software for modeling data using item response theory such as ConQuest can be error-prone. When preparation rules as well as model specifications are well-defined, the processes are automatable.

In this talk we present the package eat, which is designed to simplify data preparation processes and IRT modeling with the software ConQuest within the R programming environment. The package provides routines for automation of data pre-processing as well as an interface for ConQuest to specify and run several IRT models at once. Input and data are checked for consistency, the model syntax is generated automatically, and the output files are read into a result structure for further processing. This structure contains all model parameters, fit statistics and additional information. Parallel processing with several types of models is achieved through creation and execution of batch files. Furthermore, analyses can be logged automatically. Currently, uni- and multidimensional one parameter logistic models with and without regression, DIF, weighting and/or grouping variables, partial credit models as well as rating scale models are implemented. All methods are illustrated by means of examples.

Friday, February 10, 14:00–16:00

Detecting Differential Item and Differential Step Functioning with Partial Credit Trees

Basil Abou El-Kompoz (Ludwig-Maximilians-Universität München), *Achim Zeileis* (Universität Innsbruck), *Carolin Strobl* (University of Zurich)

The psychotree package offers a statistical toolbox for detecting parameter differences in psychometric models, including differential item functioning (DIF) in the Rasch model, that can lead to an unfair advantage or

disadvantage for certain groups of subjects in educational and psychological testing. The DIF detection method implemented in the psychotree package is based on a flexible model-based recursive partitioning framework employing generalized M-fuctuation tests for detecting differences in the item parameters between different groups of subjects. The main advantage of this approach is that it allows to detect groups of subjects exhibiting DIF, that are not pre-specified, but are detected automatically from combinations of covariates. This technique is now being extended to the detection of differential item and differential step functioning (DSF) in the partial credit model (PCM). The talk presents work in progress and outlines our current attempts at providing an illustration of the resulting parameter estimates that is intuitively interpretable for applied researchers.

Maximum Entropy Models: A Framework for IRT Modeling?

Georg Hosoya (Freie Universität Berlin)

Maximum entropy models, also known as maximum entropy classifiers, log-linear models, multinomial logit models or Markov random fields feature many structural similarities to the framework of IRT modeling as proposed by Rasch (1961).

In the theoretical part of the talk, these formal similarities are highlighted by means of the partial credit model (Masters, 1980) and the conditional Rasch model for dichotomous data.

As an example for the application of the maximum entropy modeling framework to research questions in the domain of IRT, a new Rasch model for the assessment of intraindividual variability is presented, its formal properties are discussed and results of the application to real world ambulatory assessment data are shown.

Estimating Hierarchical Structure GLMM Item Response Models in R

Ivailo Partchev (Cito, KU Leuven), *Paul De Boeck* (KU Leuven, UV Amsterdam)

We examine a class of item response models with two defining features: they all (i) have a tree representation, and (ii) are members of the family of generalized linear mixed models (GLMM). IRT models fitting in this framework include, among others, continuation ratio models, multinomial processing tree models based on binary trees, longitudinal models with serial dependency of latent variables, or hierarchical models of the bi-factor type. Because the models are based on trees, they are denoted as IRTree models. The tree representation can be used both for the measurement part and for the structural part of the models. The GLMM nature of the models implies that they can all be estimated with the `lmer()` function of the lme4 package in R. Some rather elaborate preparation of the data is usually necessary, and we present a package, `irtrees`, to facilitate these operations.