Journal of Statistical Software

June 2014, Volume 58, Issue 6. http://www.jstatsoft.org/

KernSmoothIRT: An R Package for Kernel Smoothing in Item Response Theory

Angelo Mazza
University of Catania

Antonio Punzo
University of Catania

Brian McGuire
Montana State University

Abstract

Item response theory (IRT) models are a class of statistical models used to describe the response behaviors of individuals to a set of items having a certain number of options. They are adopted by researchers in social science, particularly in the analysis of performance or attitudinal data, in psychology, education, medicine, marketing and other fields where the aim is to measure latent constructs. Most IRT analyses use parametric models that rely on assumptions that often are not satisfied. In such cases, a nonparametric approach might be preferable; nevertheless, there are not many software implementations allowing to use that.

To address this gap, this paper presents the R package **KernSmoothIRT**. It implements kernel smoothing for the estimation of option characteristic curves, and adds several plotting and analytical tools to evaluate the whole test/questionnaire, the items, and the subjects. In order to show the package's capabilities, two real datasets are used, one employing multiple-choice responses, and the other scaled responses.

Keywords: kernel smoothing, item response theory, principal component analysis, probability simplex.

1. Introduction

In psychometrics the analysis of the relation between latent continuous variables and observed dichotomous/polytomous variables is known as item response theory (IRT). Observed variables arise from items of one of the following formats: *multiple-choice* in which only one alternative is designed to be correct, *multiple-response* in which more than one answer may be keyed as correct, *rating scale* in which the phrasing of the response categories must reflect a scaling of the responses, *partial credit* in which a partial credit is given in accordance with an examinee's degree of attainment in solving a problem, and *nominal* in which there is neither a correct option nor an option ordering. Naturally, a set of items can be a mixture of these

item formats. Hereafter, for consistency's sake, the term "option" will be used as the unique term for several often used synonyms like: (response) category, alternative, answer, and so on; also the term "test" will be used to refer to a set of items comprising any psychometric test or questionnaire.

Our notation and framework can be summarized as follows. Consider the responses of an n-dimensional set $\mathcal{S} = \{S_1, \ldots, S_i, \ldots, S_n\}$ of subjects to a k-dimensional sequence $\mathcal{I} = \{I_1, \ldots, I_j, \ldots, I_k\}$ of items. Let $\mathcal{O}_j = \{O_{j1}, \ldots, O_{jl}, \ldots, O_{jm_j}\}$ be the m_j-dimensional set of options conceived for I_j, and let x_{jl} be the weight attributed to O_{jl}. The actual response of S_i to I_j can be so represented as a selection vector $\boldsymbol{y}_{ij} = (y_{ij1}, \ldots, y_{ijm_j})^\top$, where $\boldsymbol{y}_{ij}$ is an observation from the random variable $\boldsymbol{Y}_{ij}$ and $y_{ijl} = 1$ if the option O_{jl} is selected, and 0 otherwise. From now on it will be assumed that, for each item $I_j \in \mathcal{I}$, the subject selects one and only one of the m_j options in $\mathcal{O}_j$; omitted responses are permitted.

The central problem in IRT, with reference to a generic option O_{jl} of I_j, is the specification of a mathematical model describing the probability of selecting O_{jl} as a function of ϑ, the underlying latent trait which the test attempts to measure (the discussion is here restricted to models for items that measure one continuous latent variable, i.e., *unidimensional latent trait models*). According to Ramsay (1991), this function, or curve, will be referred to as *option characteristic curve* (OCC), and it will be denoted with

$$p_{jl}(\vartheta) = \mathsf{P}(\text{select } O_{jl} \,|\vartheta) = \mathsf{P}(Y_{jl} = 1 \,|\vartheta),$$

$j = 1, \ldots, k$ and $l = 1, \ldots, m_j$. OCCs are important for a number of different reasons. For example, in the analysis of multiple-choice items, which has typically relied on numerical statistics such as the proportion of subjects selecting each option and the point biserial correlation (quantifying item discrimination), it might be more informative to take into account all of the OCCs (Lei, Dunbar, and Kolen 2004). Moreover, the OCCs are the starting point for a wide range of IRT analyses (see, e.g., Baker and Kim 2004). Note that, the term "option characteristic curve" is not by any means universal. Among the different terms found in literature, there are *category characteristic curve*, *operating characteristic curve*, *category response function*, *item category response function*, *option response function* and more (see Ostini and Nering 2006, p. 10, and DeMars 2010, p. 23, for a survey of the different names used).

To estimate the OCCs, in analogy with the classic statistical modeling, at least two routes are possible. The first, and most common, is the *parametric* one (PIRT: parametric IRT), in which a parametric structure is assumed so that the estimation of an OCC is reduced to the estimation of a parameter vector, of dimension varying from model to model, for each item in $\mathcal{I}$ (see, e.g., Thissen and Steinberg 1986, Van der Linden and Hambleton 1997, Ostini and Nering 2006, and Nering and Ostini 2010, to have an idea of the existing PIRT models). This vector is usually considered to be of direct interest and its estimate is often used as a summary statistic of some item aspects such as difficulty and discrimination (see Lord 1980). The second route is the *nonparametric* one (NIRT: nonparametric IRT), in which estimation is made directly on $\boldsymbol{y}_{ij}$, $i = 1, \ldots, n$ and $j = 1, \ldots, k$, without assuming any mathematical form for the OCCs, in order to obtain more flexible estimates which, according to Van der Linden and Hambleton (1997, p. 348), can be assumed to be closer to the true OCCs than those provided by PIRT models. Accordingly, Ramsay (1997) argues that NIRT might become the reference approach unless there are substantive reasons for preferring a certain parametric model. Moreover, although nonparametric models are not characterized by parameters of

direct interest, they encourage the graphical display of results; Ramsay (1997, p. 384), by personal experience, confirms the communication advantage of an appropriate display over numerical summaries. These are only some of the motivations which justify the growth of NIRT research in recent years; other considerations can be found in Junker and Sijtsma (2001) who identify three broad motivations for the development and continued interest in NIRT.

This paper focuses on NIRT. Its origins – prior to interest in PIRT – are found in the scalogram analysis of Guttman (1947, 1950a,b). Nevertheless, the work by Mokken (1971) is recognized as the first important contribution to this paradigm; he not only gave a nonparametric representation of the item characteristic curves in the form of a basic set of formal properties they should satisfy, but also provided the statistical theory needed to check whether these properties would hold in empirical data. Among these properties, *monotonicity* with respect to ϑ was required. The R (R Core Team 2014) package **mokken** (Van der Ark 2007, 2012) provides tools to perform a Mokken scale analysis. Several other NIRT approaches have been proposed (see Van der Ark 2001). Among them, kernel smoothing (Ramsay 1991) is a promising option, due to conceptual simplicity as well as advantageous practical and theoretical properties. The computer software **TestGraf** (Ramsay 2000) performs kernel smoothing estimation of OCCs and related graphical analyses. In this paper we present the R package **KernSmoothIRT** (Mazza, Punzo, and McGuire 2014), available from the Comprehensive R Archive Network (CRAN) at `http://CRAN.R-project.org/package=KernSmoothIRT`, which offers most of the **TestGraf** features and adds some related functionalities. Note that, although R is well-provided with PIRT techniques (see de Leeuw and Mair 2007 and Wickelmaier, Strobl, and Zeileis 2012), it does not offer nonparametric analyses, of the type described above, in IRT. Nonparametric smoothing techniques of the kind found in **KernSmoothIRT** are commonly used and often cited exploratory statistical tools; as evidence, consider the number of times in which classical statistical studies use the functions `density()` and `ksmooth()`, both in the **stats** package, for kernel smoothing estimation of a density or regression function, respectively. Consistent with its exploratory nature, **KernSmoothIRT** can be used as a complementary tool to other IRT packages; for example a **mokken** package user may use it to evaluate monotonicity. OCCs smoothed by kernel techniques, due to their statistical properties (see Douglas 1997, 2001 and Douglas and Cohen 2001), have been also used in PIRT analysis as a benchmark to estimate the best OCCs in a pre-specified parametric family (Punzo 2009).

The paper is organized as follows. Section 2 retraces kernel smoothing estimation of the OCCs and Section 3 illustrates other useful IRT functions based on these estimates. The relevance of the package is shown, via two real data sets, in Section 4, and conclusions are finally given in Section 5.

2. Kernel smoothing of OCCs

Ramsay (1991, 1997) popularized nonparametric estimation of OCCs by proposing regression methods, based on kernel smoothing approaches, which are implemented in the **TestGraf** program (Ramsay 2000). The basic idea of kernel smoothing is to obtain a nonparametric estimate of the OCC by taking a (local) weighted average (see Altman 1992, Härdle 1990, and Simonoff 1996) of the form

$$\widehat{p}_{jl}(\vartheta) = \sum_{i=1}^{n} w_{ij}(\vartheta)\, y_{ijl}, \tag{1}$$

$j = 1, \ldots, k$ and $l = 1, \ldots, m_j$, where the weights $w_{ij}(\vartheta)$ are defined so as to be maximal when $\vartheta = \vartheta_i$ and to be smoothly non-increasing as $|\vartheta - \vartheta_i|$ increases, with ϑ_i being the value of ϑ for $S_i \in \mathcal{S}$. The need to keep $\widehat{p}_{jl}(\vartheta) \in [0,1]$, for each $\vartheta \in \mathbb{R}$, requires the additional constraints $w_{ij}(\vartheta) \geq 0$ and $\sum_{i=1}^{n} w_{ij}(\vartheta) = 1$; as a consequence, it is preferable to use Nadaraya-Watson weights (Nadaraya 1964 and Watson 1964) of the form

$$w_{ij}(\vartheta) = \frac{K\left(\dfrac{\vartheta - \vartheta_i}{h_j}\right)}{\displaystyle\sum_{r=1}^{n} K\left(\frac{\vartheta - \vartheta_r}{h_j}\right)}, \tag{2}$$

where $h_j > 0$ is the *smoothing parameter* (also known as *bandwidth*) controlling the amount of smoothness (in terms of bias-variance trade-off), while K is the *kernel function*, a nonnegative, continuous ($\widehat{p}_{jl}$ inherits the continuity from K) and usually symmetric function that is non-increasing as its argument moves further from zero.

Since the performance of (1) largely depends on the choice of h_j, rather than on the kernel function (see, e.g., Marron and Nolan 1988) a simple Gaussian kernel $K(u) = \exp\left(-u^2/2\right)$ is often preferred (this is the only setting available in **TestGraf**). Nevertheless, **KernSmoothIRT** allows for other common choices such as the uniform kernel $K(u) = \mathbb{I}_{[-1,1]}(u)$, and the quadratic kernel $K(u) = \left(1 - u^2\right)\mathbb{I}_{[-1,1]}(u)$, where $\mathbb{I}_A(u)$ represents the indicator function assuming value 1 on A and 0 otherwise. In addition to the functionalities implemented in **TestGraf**, **KernSmoothIRT** allows the bandwidth h_j to vary from item to item (as highlighted by subscript j). This is an important aspect, since different items may not require the same amount of smoothing to obtain smooth curves (Lei *et al.* 2004, p. 8).

2.1. Estimating abilities

Unlike the standard kernel regression methods, in (1) the dependent variable Y_{jl} is a binary variable and the independent one is the latent variable ϑ. Although ϑ cannot be directly observed, kernel smoothing can still be used, but each ϑ_i in (2) must be replaced with a reasonable estimate $\widehat{\vartheta}_i$ (Ramsay 1991) leading to

$$\widehat{p}_{jl}(\vartheta) = \sum_{i=1}^{n} \widehat{w}_{ij}(\vartheta)\, y_{ijl}, \tag{3}$$

where

$$\widehat{w}_{ij}(\vartheta) = \frac{K\left(\dfrac{\vartheta - \widehat{\vartheta}_i}{h_j}\right)}{\displaystyle\sum_{r=1}^{n} K\left(\frac{\vartheta - \widehat{\vartheta}_r}{h_j}\right)}.$$

The choice of the scale of $\widehat{\vartheta}_i$ is arbitrary, since in this context only rank order considerations make sense (Bartholomew 1983 and Ramsay 1991, p. 614). Therefore, as most IRT models do, the estimation process begins (Ramsay 1991, p. 615 and Ramsay 2000, pp. 25–26) with:

1. Computation of the transformed rank $r_i = \text{rank}(S_i) / (n+1)$, with $\text{rank}(S_i) \in \{1, \ldots, n\}$,

induced by some suitable statistic t_i, the total score

$$t_i = \sum_{j=1}^{k} \sum_{l=1}^{m_j} y_{ijl} x_{jl}$$

being the most obvious choice. **KernSmoothIRT** also allows, through the argument `RankFun` of the `ksIRT()` function, the use of common summary statistics available in R, such as `mean()` and `median()`, or of a custom user-defined function. Alternatively, the user may specify the rank of each subject explicitly through the argument `SubRank`, allowing subject ranks to come from another source than the test being studied.

2. Replacement of r_i by the quantile $\widehat{\vartheta}_i$ of some distribution function F. The estimated ability value for S_i then becomes $\widehat{\vartheta}_i = F^{-1}(r_i)$. In these terms, the denominator $n+1$ of r_i avoids an infinity value for the biggest $\widehat{\vartheta}_i$ when $\lim_{\vartheta \to +\infty} F(\vartheta) = 1^-$. Note that the choice of F is equivalent to the choice of the ϑ-metric. Historically, the standard Gaussian distribution $F = \Phi$ has been heavily used (see Bartholomew 1988). However, **KernSmoothIRT** allows the user specification of F through one of the classical continuous distributions available in R.

Since these preliminary ability estimates are rank-based, they are usually referred to as *ordinal ability estimates*. Note that even a substantial amount of error in the ranks has only a small impact on the estimated curve values. This can be demonstrated both by mathematical analysis and through simulated data (see Ramsay 1991, 2000 and Douglas 1997). Further theoretical results can be found in Douglas (2001) and Douglas and Cohen (2001). The latter also asserts that, if nonparametric estimated curves are meaningfully different from parametric ones, this parametric model – defined on the particular scale determined by F – is an incorrect model for the data. In order to make this comparison valid, it is fundamental that the same F is used for both nonparametric and parametric curves. Thus, in the choice of a parametric family, visual inspections of the estimated kernel curves can be useful (Punzo 2009).

2.2. Operational aspects

Operationally, the kernel OCC is evaluated on a finite grid, $\vartheta_1, \ldots, \vartheta_s, \ldots, \vartheta_q$, of q equally-spaced values spanning the range of the ordinal ability estimates, so that the distance between two consecutive points is δ. Thus, starting from the values of y_{ijl} and $\widehat{\vartheta}_i$, by grouping we can define the two sequences of q values

$$\widetilde{y}_{sjl} = \sum_{i=1}^{n} \mathbb{I}_{[\vartheta_s - \delta/2, \vartheta_s + \delta/2)} \left(\widehat{\vartheta}_i\right) y_{ijl} \quad \text{and} \quad v_s = \sum_{i=1}^{n} \mathbb{I}_{[\vartheta_s - \delta/2, \vartheta_s + \delta/2)} \left(\widehat{\vartheta}_i\right).$$

Up to a scale factor, the sequence $\widetilde{y}_{sjl}$ is a grouped version of y_{ijl}, while v_s is the corresponding number of subjects in that group. It follows that

$$\widehat{p}_{jl}(\vartheta) \approx \frac{\sum_{s=1}^{q} K\left(\frac{\vartheta - \vartheta_s}{h_j}\right) \widetilde{y}_{sjl}}{\sum_{s=1}^{q} K\left(\frac{\vartheta - \vartheta_s}{h_j}\right) v_s}, \qquad \vartheta \in \left\{\vartheta_1, \ldots, \vartheta_s, \ldots, \vartheta_q\right\}. \tag{4}$$

2.3. Cross-validation selection for the bandwidth

Two of the most frequently used methods of bandwidth selection are the plug-in method and the cross-validation (for a more complete treatment of these methods see, e.g., Härdle 1990).

The former approach, widely used in kernel density estimation, often leads to rules of thumb. Motivated by the need to have fast automatically generated kernel estimates, the function `ksIRT()` of **KernSmoothIRT** adopts, as default, the common rule of thumb of Silverman (1986, p. 45) for the Gaussian kernel density estimator. In our context, this is formulated as

$$h_j = h = 1.06\,\sigma_\vartheta\,n^{-1/5}, \tag{5}$$

where σ_ϑ – that in the original framework is a sample estimate – simply represents the standard deviation of ϑ, induced by F. Note that (5), with $\sigma_\vartheta = 1$, is the only approach considered in **TestGraf**.

The second approach, cross-validation, requires a considerably higher computational effort; nevertheless, it is simple to understand and widely applied in nonparametric kernel regression (see, e.g., Wong 1983, Rice 1984 and Mazza and Punzo 2011, 2013a,b, 2014). Its description, in our context, is as follows. Let $\boldsymbol{y}_j = \left(\boldsymbol{y}_{1j}, \ldots, \boldsymbol{y}_{ij}, \ldots, \boldsymbol{y}_{nj}\right)$ be the $m_j \times n$ selection matrix referred to I_j. Moreover, let

$$\widehat{\boldsymbol{p}}_j\left(\vartheta\right) = \left(\widehat{p}_{j1}\left(\vartheta\right), \ldots, \widehat{p}_{jm_j}\left(\vartheta\right)\right)^\top$$

be the m_j-dimensional vector of kernel-estimated probabilities, for I_j, at the evaluation point ϑ. The probability kernel estimate evaluated in ϑ, for I_j, can thus be written as

$$\widehat{\boldsymbol{p}}_j\left(\vartheta\right) = \sum_{i=1}^{n} \widehat{w}_{ij}\left(\vartheta\right)\boldsymbol{y}_{ij} = \boldsymbol{y}_j\widehat{\boldsymbol{w}}_j\left(\vartheta\right),$$

where $\widehat{\boldsymbol{w}}_j\left(\vartheta\right) = \left(\widehat{w}_{1j}\left(\vartheta\right), \ldots, \widehat{w}_{ij}\left(\vartheta\right), \ldots, \widehat{w}_{nj}\left(\vartheta\right)\right)^\top$ denotes the vector of weights.

In detail, cross-validation simultaneously fits and smooths the data contained in $\boldsymbol{y}_j$ by removing one "data point" $\boldsymbol{y}_{ij}$ at a time, estimating the value of $\boldsymbol{p}_j$ at the corresponding ordinal ability estimate $\widehat{\vartheta}_i$, and then comparing the estimate to the omitted, observed value. So the cross-validation statistic is

$$\mathrm{CV}\left(h_j\right) = \frac{1}{n}\sum_{i=1}^{n}\left(\boldsymbol{y}_{ij} - \widehat{\boldsymbol{p}}_j^{(-i)}\left(\widehat{\vartheta}_i\right)\right)^\top\left(\boldsymbol{y}_{ij} - \widehat{\boldsymbol{p}}_j^{(-i)}\left(\widehat{\vartheta}_i\right)\right),$$

where

$$\widehat{\boldsymbol{p}}_j^{(-i)}\left(\widehat{\vartheta}_i\right) = \frac{\displaystyle\sum_{\substack{r=1\\ r\neq i}}^{n} K\left(\frac{\widehat{\vartheta}_i - \widehat{\vartheta}_r}{h_j}\right)\boldsymbol{y}_{rj}}{\displaystyle\sum_{\substack{r=1\\ r\neq i}}^{n} K\left(\frac{\widehat{\vartheta}_i - \widehat{\vartheta}_r}{h_j}\right)}$$

is the estimated vector of probabilities at $\widehat{\vartheta}_i$ computed by removing the observed selection vector $\boldsymbol{y}_{ij}$, as denoted by the superscript in $\widehat{\boldsymbol{p}}_j^{(-i)}$. The value of h_j that minimizes $\mathrm{CV}\left(h_j\right)$ is

referred to as the cross-validation smoothing parameter, h_j^{CV}, and it is possible to find it by systematically searching across a suitable smoothing parameter region.

2.4. Approximate pointwise confidence intervals

In visual inspection and graphical interpretation of the estimated kernel curves, pointwise confidence intervals at the evaluation points provide relevant information, because they indicate the extent to which the kernel OCCs are well defined across the range of ϑ considered. Moreover, they are useful when nonparametric and parametric models are compared.

Since $\widehat{p}_{jl}(\vartheta)$ is a linear function of the data, as can be easily seen from (3), and as $Y_{ijl} \sim \mathrm{Ber}\left[p_{jl}\left(\widehat{\vartheta}_i\right)\right]$,

$$\begin{aligned}\mathsf{VAR}\left[\widehat{p}_{jl}(\vartheta)\right] &= \sum_{i=1}^{n}\left[\widehat{w}_i(\vartheta)\right]^2 \mathsf{VAR}\left(Y_{ijl}\right) \\ &= \sum_{i=1}^{n}\left[\widehat{w}_i(\vartheta)\right]^2 p_{jl}\left(\widehat{\vartheta}_i\right)\left[1-p_{jl}\left(\widehat{\vartheta}_i\right)\right].\end{aligned}$$

The above formula holds if independence of the Y_{ijl}s, with respect to the subjects, is assumed and possible error variation in the arguments, $\widehat{\vartheta}_i$, are ignored (Ramsay 1991). Substituting p_{jl} for $\widehat{p}_{jl}$ yields the $(1-\alpha)\,100\%$ approximate pointwise confidence intervals

$$\widehat{p}_{jl}(\vartheta) \mp z_{1-\frac{\alpha}{2}}\sqrt{\sum_{i=1}^{n}\left[\widehat{w}_i(\vartheta)\right]^2 \widehat{p}_{jl}\left(\widehat{\vartheta}_i\right)\left[1-\widehat{p}_{jl}\left(\widehat{\vartheta}_i\right)\right]}, \tag{6}$$

where $z_{1-\frac{\alpha}{2}}$ is such that $\Phi\left[z_{1-\frac{\alpha}{2}}\right] = 1-\frac{\alpha}{2}$. Other more complicated approaches to interval estimation for kernel-based nonparametric regression functions are described in Azzalini, Bowman, and Härdle (1989) and Härdle (1990, Section 4.2).

3. Functions related to the OCCs

Once the kernel estimates of the OCCs are obtained, several other quantities can be computed based on them. In what follows we will give a concise list of the most important ones.

3.1. Expected item score

In order to obtain a single function for each item in $\mathcal{I}$ it is possible to define the expected value of the score $X_j = \sum_{l=1}^{m_j} x_{jl}Y_{jl}$, conditional on a given value of ϑ (see, e.g., Chang and Mazzeo 1994), as follows

$$e_j(\vartheta) = \mathsf{E}\left(X_j \,|\vartheta\right) = \sum_{l=1}^{m_j} x_{jl}p_{jl}(\vartheta), \tag{7}$$

$j = 1,\ldots,k$, that takes values in $[x_{j\,\min}, x_{j\,\max}]$, where $x_{j\,\min} = \min\left\{x_{j1},\ldots,x_{jm_j}\right\}$ and $x_{j\,\max} = \max\left\{x_{j1},\ldots,x_{jm_j}\right\}$. The function $e_j(\vartheta)$ is commonly known as *expected item score*

(EIS) and can be viewed (Lord 1980) as a regression of the item score X_j onto the ϑ scale. Naturally, for dichotomous and multiple-choice IRT models, the EIS coincides with the OCC of the correct option.

Starting from (7), it is straightforward to define the kernel EIS estimate as follows

$$\widehat{e}_j(\vartheta) = \sum_{l=1}^{m_j} x_{jl}\widehat{p}_{jl}(\vartheta) = \sum_{l=1}^{m_j} x_{jl} \sum_{i=1}^{n} \widehat{w}_{ij}(\vartheta)\, y_{ijl} = \sum_{i=1}^{n} \widehat{w}_{ij}(\vartheta) \sum_{l=1}^{m_j} x_{jl} y_{ijl}.$$

For the EIS, in analogy with Section 2.4, the $(1-\alpha)\,100\%$ approximate pointwise confidence interval is given by

$$\widehat{e}_j(\vartheta) \mp z_{1-\frac{\alpha}{2}} \sqrt{\widehat{\mathsf{VAR}\left[\widehat{e}_j(\vartheta)\right]}}, \tag{8}$$

and, since $Y_{ijl}Y_{ijt} \equiv 0$ for $l \neq t$, one has

$$\mathsf{VAR}\left[\widehat{e}_j(\vartheta)\right] = \sum_{i=1}^{n} \left[\widehat{w}_{ij}(\vartheta)\right]^2 \mathsf{VAR}\left(\sum_{l=1}^{m_j} x_{jl} Y_{ijl}\right)$$

where

$$\begin{aligned}
\mathsf{VAR}\left(\sum_{l=1}^{m_j} x_{jl} Y_{ijl}\right) &= \sum_{l=1}^{m_j} x_{jl}^2 \mathsf{VAR}\left(Y_{ijl}\right) + \sum_{l=1}^{m_j} \sum_{t \neq l} x_{jl} x_{jt} \mathsf{COV}\left(Y_{ijl}, Y_{ijt}\right) \\
&= \sum_{l=1}^{m_j} x_{jl}^2 \mathsf{VAR}\left(Y_{ijl}\right) - \sum_{l=1}^{m_j} \sum_{t \neq l} x_{jl} x_{jt} \mathsf{E}\left(Y_{ijl}\right) \mathsf{E}\left(Y_{ijt}\right) \\
&= \sum_{l=1}^{m_j} x_{jl}^2 p_{jl}\left(\widehat{\vartheta}_i\right)\left[1 - p_{jl}\left(\widehat{\vartheta}_i\right)\right] - \sum_{l=1}^{m_j} \sum_{t \neq l} x_{jl} x_{jt} p_{jl}\left(\widehat{\vartheta}_i\right) p_{jt}\left(\widehat{\vartheta}_i\right).
\end{aligned}$$

Substituting p_{jl} with $\widehat{p}_{jl}$ in $\mathsf{VAR}\left[\widehat{e}_i(\vartheta)\right]$, one obtains $\widehat{\mathsf{VAR}\left[\widehat{e}_i(\vartheta)\right]}$, quantity that has to be inserted in (8).

Really, intervals in (6) and (8) are, respectively, intervals for $\mathsf{E}\left[\widehat{p}_{jl}(\vartheta)\right]$ and $\mathsf{E}\left[\widehat{e}_j(\vartheta)\right]$, rather than for $p_{jl}(\vartheta)$ and $e_j(\vartheta)$; thus, they share the bias present in $\widehat{p}_{jl}$ and $\widehat{e}_j$, respectively (for the OCC case, see Ramsay 1991, p. 619).

3.2. Expected test score

In analogy to Section 3.1, a single function for the whole test can be obtained as follows

$$e(\vartheta) = \sum_{j=1}^{k} e_j(\vartheta) = \sum_{j=1}^{k} \sum_{l=1}^{m_j} x_{jl} p_{jl}(\vartheta).$$

It is called *expected test score* (ETS). Its kernel smoothed counterpart can be specified as

$$\widehat{e}(\vartheta) = \sum_{j=1}^{k} \sum_{l=1}^{m_j} x_{jl} \widehat{p}_{jl}(\vartheta) \tag{9}$$

and may be preferred in substitution of ϑ, for people who are not used to IRT, as display variable on the x-axis to facilitate the interpretation of the OCCs, as well as of other output-plots of **KernSmoothIRT**. This possibility is considered through the argument `axistype = "scores"` of the `plot()` method. Note that, although it can happen that (9) fails to be completely increasing in ϑ, this event is rare and tends to affect the plots only at extreme trait levels.

3.3. Relative credibility curve

For a generic subject $S_i \in \mathcal{S}$, we can compute the relative likelihood

$$L_i(\vartheta) = M^{-1} \prod_{j=1}^{k} \prod_{l=1}^{m_j} [\widehat{p}_{jl}(\vartheta)]^{y_{ijl}} \tag{10}$$

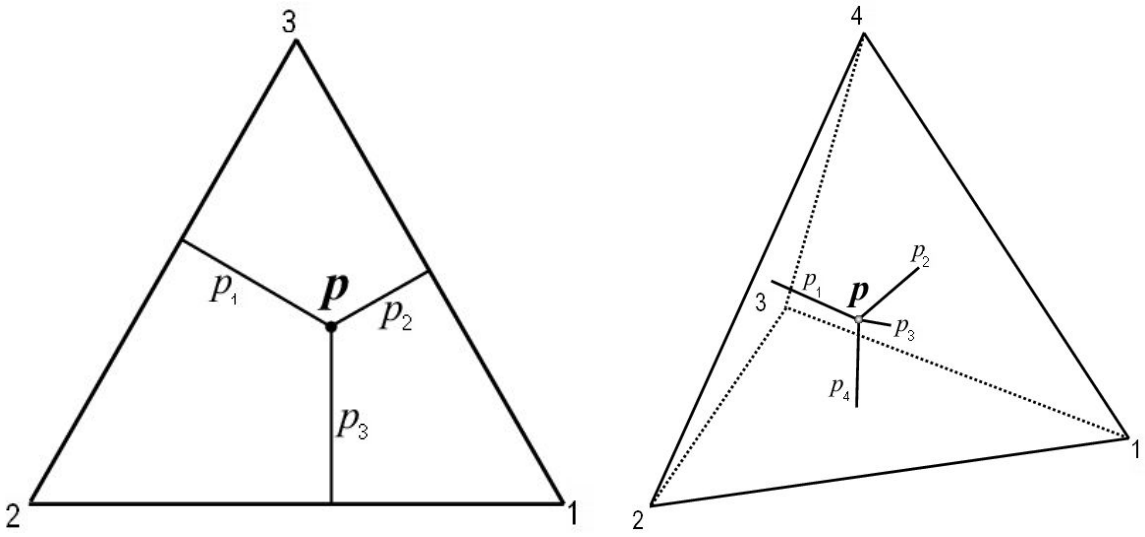
of the various values of ϑ given his pattern of responses and given the kernel-estimated OCCs. In (10), $M = \max_\vartheta \left\{ \prod_{j=1}^{k} \prod_{l=1}^{m_j} [\widehat{p}_{jl}(\vartheta)]^{y_{ijl}} \right\}$. The function in (10) is also known as *relative credibility curve* (RCC; see, e.g, Lindsey 1973). The ϑ-value, say $\widehat{\vartheta}^{\mathrm{ML}}$, such that $L_i(\vartheta) = 1$, is called the maximum likelihood (ML) estimate of the ability for S_i (see also Kutylowski 1997). Differently from simple summary statistics like the total score, $\widehat{\vartheta}^{\mathrm{ML}}$ considers, in addition to the whole pattern of responses, also the characteristics of the items as described by their OCCs; thus, it will tend to be a more accurate estimate of the ability.

Finally, as Kutylowski (1997) and Ramsay (2000) do, the obtained values of $\widehat{\vartheta}^{\mathrm{ML}}$ may be used as a basis for a second step of the kernel smoothing estimation of OCCs. This iterative process, consisting in cycling back the values of $\widehat{\vartheta}^{\mathrm{ML}}$ into estimation, can clearly be repeated any number of times with the hope that each step refines or improves the estimates of ϑ. However, as Ramsay (2000) states, for the vast majority of applications, no iterative refinement is really necessary, and the use of $\widehat{\vartheta}_i$ or $\widehat{\vartheta}_i^{\mathrm{ML}}$ for ranking examinees works fine. This is the reason why we have not considered the iterative process in the package.

3.4. Probability simplex

With reference to a generic item $I_j \in \mathcal{I}$, the vector of probabilities $\widehat{\boldsymbol{p}}_j(\vartheta)$ can be seen as a point in the probability simplex $\mathbb{S}^{m_j}$, defined as the $(m_j - 1)$-dimensional subset of the m_j-dimensional space containing vectors with nonnegative coordinates summing to one. As ϑ varies, because of the assumptions of smoothness and unidimenionality of the latent trait, $\widehat{\boldsymbol{p}}_j(\vartheta)$ moves along a curve; the item analysis problem is to locate the curve properly within the simplex. On the other hand, the estimation problem for S_i is the location of its position along this curve.

As illustrated in Aitchison (2003, pp. 5–9), a convenient way of displaying points in $\mathbb{S}^{m_j}$, when $m_j = 3$ or $m_j = 4$, is represented, respectively, by the *reference triangle* in Figure 1(a) – an equilateral triangle having unit altitude – and by the *regular tetrahedron*, of unit altitude, in Figure 1(b). Here, for any point $\boldsymbol{p}$, the lengths of the perpendiculars $p_1, \ldots, p_{m_j}$ from $\boldsymbol{p}$ to the sides opposite to the vertices $1, \ldots, m_j$ are all greater than, or equal to, zero and have a unitary sum. Since there is a unique point with these perpendicular values, there is a one-to-one correspondence between $\mathbb{S}^3$ and points in the reference triangle, and between $\mathbb{S}^4$ and points in the regular tetrahedron. Thus, we have a simple means for representing the vector



(a) Triangle (b) Tetrahedron

Figure 1: Convenient way of displaying a point in the probability simplex $\mathbb{S}^{m_j}$ when $m_j = 3$ (on the left) and $m_j = 4$ (on the right).

of probabilities $\widehat{\boldsymbol{p}}_j(\vartheta)$ when $m_j = 3$ and $m_j = 4$. Note that for items with more than four options there is no satisfactory way of obtaining a visual representation of the corresponding probability simplex; nevertheless, with **KernSmoothIRT** we can perform a partial analysis which focuses only on three or four of the options.

4. Package description and illustrative examples

The main function of the package is `ksIRT()`; it creates an S3 object of class '`ksIRT`', which provides a `plot()` method as well as a suite of functions that allow the user to analyze the subjects, the options, the items, and the overall test. What follows is an illustration of the main capabilities of the **KernSmoothIRT** package.

4.1. Kernel smoothing with the `ksIRT()` function

The `ksIRT()` function performs the kernel smoothing. It requires `responses`, a $(n \times k)$-matrix, with a row for each subject in $\mathcal{S}$ and a column for each item in $\mathcal{I}$, containing the selected option numbers. Alternatively, `responses` may be a data frame or a list object.

Arguments for setting the item format: `format`, `key`, `weights`

To use the basic weighting schemes associated with each item format, the following combination of arguments have to be applied.

- For *multiple-choice* items, use `format = 1` and provide in `key`, for each item, the option number that corresponds to the correct option. For *multiple-response* items, one way to score them is simply to count the correctly classified options; to do this, a preliminary conversion of every option into a separate true/false item is necessary.

Values	Description
`$itemcor`	Vector of item point polyserial correlations.
`$evalpoints`	Vector of evaluation points used in curve estimation.
`$subjscore`	Vector of observed subjects' overall scores.
`$subjtheta`	Vector of subjects' quantiles on the distribution specified in `thetadist`.
`$subjthetaML`	Vector of $\widehat{\vartheta}_i^{\mathrm{ML}}$, $i = 1, \ldots, n$.
`$subjscoreML`	Vector of subjects' ML scores $\widehat{e}\left(\widehat{\vartheta}_i^{\mathrm{ML}}\right)$, $i = 1, \ldots, n$.
`$subjscoresummary`	Vector of quantiles, of probability 0.05, 0.25, 0.50, 0.75, 0.95, for the observed overall scores.
`$subjthetasummary`	Vector as `subjscoresummary` but computed on `subjtheta`.
`$OCC`	Matrix of dimension $\left(\sum_{j=1}^{k} m_j\right) \times (3 + q)$. The first three columns specify the item, the option, and the corresponding weight x_{jl}. The additional columns contain the kernel smoothed OCCs at each evaluation point.
`$stderrs`	Matrix as `OCC` containing the standard errors of `OCC`.
`$bandwidth`	Vector of bandwidths h_j, $j = 1, \ldots, k$.
`$RCC`	List of n vectors containing the q values of $L_i(\vartheta)$, $i = 1, \ldots, n$.

Table 1: List of most the important components of class '`ksIRT`'.

- For *rating scale* and *partial credit* items, use `format = 2` and provide in `key` a vector with the number of options of every item. If all the items have the same number of options, then `key` may be a scalar.
- For *nominal* items, use `format = 3`; `key` is omitted. Note that to analyze items or options, subjects have to be ranked; this can only be done if the test also contains non-nominal items or if a prior ranking of subjects is provided with `SubRank`.

If the test consists of a mixture of different item formats, then `format` must be a numeric vector of length equal to the number of items. More complicated weighting schemes may be specified using `weights` in lieu of both `format` and `key` (see the help page for details).

Arguments for smoothing: `evalpoints`, `nevalpoints`, `thetadist`, `kernel`, `bandwidth`

The user can select the q evaluation points of Section 2.2, the ranking distribution F of Section 2.1, the type of kernel function K and the bandwidth(s). The number q of OCCs evaluation points may be specified in `nevalpoints`. By default they are 51 and their range is data dependent. Alternatively, the user may directly provide evaluation points using `evalpoints`. As to F, it is by default Φ; any other distribution, with its parameters values, may be provided in `thetadist`. The default kernel function is the Gaussian; uniform or quadratic kernels

Methods	Description
`plot()`	Allows for a variety of exploratory plots.
`itemcor()`	Returns a vector of item point polyserial correlations.
`subjscore()`	Returns a vector of subjects' observed overall scores.
`subjthetaML()`	Returns a vector of $\widehat{\vartheta}_i^{\mathrm{ML}}$, $i = 1, \ldots, n$.
`subjscoreML()`	Returns a vector of $\widehat{e}\left(\widehat{\vartheta}_i^{\mathrm{ML}}\right)$, $i = 1, \ldots, n$.
`subjOCC()`	Returns a list of k matrices, of dimension $(m_j \times n)$, $j = 1, \ldots, k$, containing $\mathsf{P}\left(Y_{jl} = 1 \,|S_i\right)$, $i = 1, \ldots, n$ and $l = 1, \ldots, m_j$. The argument `stype` governs the scale on which to evaluate each subject; among the possible alternatives there are the observed score t_i and the ML estimates $\widehat{\vartheta}_i^{\mathrm{ML}}$, $i = 1, \ldots, n$.
`subjEIS()`	Returns a $(k \times n)$ matrix of subjects' expected item scores.
`subjETS()`	Returns a vector of subjects' expected test scores.
`PCA()`	Returns a list of class '`prcomp`' as defined in the **stats** package.
`subjOCCDIF()`	Returns a list containing, for each group, the same object returned by `subjOCC()`.
`subjEISDIF()`	Returns a list containing, for each group, the same object returned by `subjEIS()`.
`subjETSDIF()`	Returns a list containing, for each group, the same object returned by `subjETS()`.

Table 2: Methods implemented for class '`ksIRT`'.

Option	Description
`"OCC"`	Plots the OCCs.
`"EIS"`	Plots and returns the EISs.
`"ETS"`	Plots and returns the ETSs.
`"RCC"`	Plots the RCCs.
`"triangle"/"tetrahedron"`	Displays a simplex plot with the highest 3 or 4 probability options.
`"PCA"`	Displays a PCA plot of the EISs.
`"OCCDIF"`	Plots OCCs for multiple groups.
`"EISDIF"`	Plots EISs for multiple groups.
`"ETSDIF"`	Plots ETSs for multiple groups.

Table 3: Main options for argument `plottype` of the `plot()` method.

may be selected with `kernel`. The global bandwidth is computed by default according to the rule of thumb in Equation 5. Otherwise, the user may either input a numerical vector of bandwidths for each item or opt for cross-validation estimation, as described in Section 2.3, by specifying `bandwidth = "CV"`.

Arguments to handle missing values: `miss`, `NAweight`

Several approaches are implemented for handling missing answers. The default, `miss = "option"`, treats them as further options, with weight specified in `NAweight`, that by default is 0. When OCCs are plotted, the new options will be added to the corresponding items. Other choices impute the missing values according to some discrete probability distributions taking values on $\{1, \ldots, m_j\}$, $j = 1, \ldots, k$; the uniform distribution is specified by `miss = "random.unif"`, while the multinomial distribution, with probabilities equal to the frequencies of the non-missing options of that item, is specified by `miss = "random.multinom"`. Finally, `miss = "omit"` deletes from the data all the subjects with at least one omitted answer.

The '`ksIRT`' class

The `ksIRT()` function returns an S3 object of class '`ksIRT`'; its main components, along with their brief descriptions, can be found in Table 1. Methods implemented for this class are illustrated in Table 2. The `plot()` method allows for a variety of exploratory plots, which are selected with the argument `plottype`; its main options are described in Table 3.

4.2. Psych 101

The first tutorial uses the Psych 101 dataset included in the **KernSmoothIRT** package. This dataset contains the responses of $n = 379$ students, in an introductory psychology course, to $k = 100$ multiple-choice items, each with $m_j = 4$ options as well as a key. These data were also analyzed in Ramsay and Abrahamowicz (1989) and in Ramsay (1991).

To begin the analysis, create a '`ksIRT`' object. This step performs the kernel smoothing and prepares the object for analysis using the many types of plots available.

```
R> data("Psych101", package = "KernSmoothIRT")
R> Psych1 <- ksIRT(responses = Psychresponses, key = Psychkey, format = 1)
R> Psych1

    Item Correlation
1      1  0.23092838
2      2  0.09951663
3      3  0.19214764
.      .       .
.      .       .
.      .       .
99    99  0.01578162
100  100  0.24602614
```

The command `data("Psych101", package = "KernSmoothIRT")` loads both `Psychresponses` and `Psychkey`. The function `ksIRT()` produces kernel smoothing estimates using, by default,

a Gaussian distribution F (`thetadist = list("norm", 0, 1)`), a Gaussian kernel function K (`kernel = "gaussian"`), and the rule of thumb (5) for the global bandwidth. The last command, `Psych1`, prints the point polyserial correlations, a traditional descriptive measure of item performance given by the correlation between each dichotomous/polythomous item and the total score (see Olsson, Drasgow, and Dorans 1982, for details). As documented in Table 2, these values can be also obtained via the command `itemcor(Psych1)`.

Once the '`ksIRT`' object `Psych1` is created, several plots become available for analyzing each item, each subject and the overall test. They are displayed through the `plot()` method, as described below.

Option characteristic curves

The code

```
R> plot(Psych1, plottype = "OCC", item = c(24, 25, 92, 96))
```

produces the OCCs for items 24, 25, 92, and 96 displayed in Figure 2.

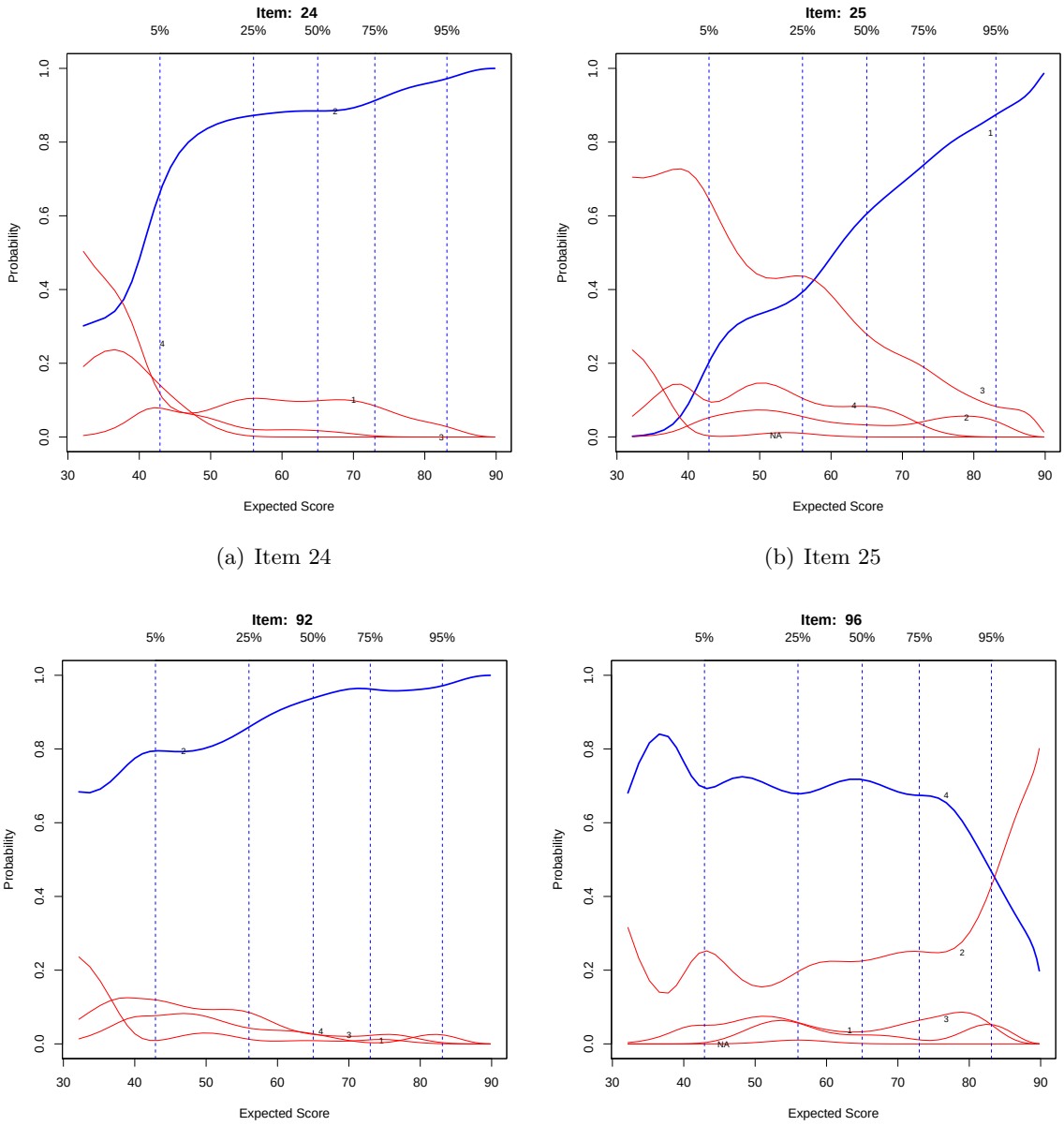
The correct options are displayed in blue and the incorrect options in red. The default specification `axistype = "scores"` uses the expected total score (9) as display variable on the x-axis. The vertical dashed lines indicate the scores (or quantiles if `axistype = "distribution"`) below which 5%, 25%, 50%, 75%, and 95% of subjects fall. Since the argument `miss` has not been specified, by default (`miss = "option"`) an additional OCC is plotted for items receiving nonresponses, as we can see from Figure 2(b) and Figure 2(d).

The OCC plots in Figure 2 show four very different items. Globally, apart from item 96 in Figure 2(d), the other items appear to be monotone enough. Item 96 is problematic for the Psych 101 instructor, as subjects with lower trait levels are more likely to select the correct option than higher trait level examinees. In fact, examinees with an expected total score of 90 are the least likely to select the correct option. Perhaps the question is misworded or it is measuring a different trait. On the contrary, items 24, 25, and 92, do a good job in differentiating between subjects with low and high trait levels. In particular item 24, in Figure 2(a), displays a high discriminating power for subjects with expected total scores near 40, and a low discriminating power for subjects with expected total scores greater than 50; above 50, subjects have roughly the same probability of selecting the correct option regardless of their expected total score. Item 25 in Figure 2(b) is also an effective one, since only the top students are able to recognize option 3 as incorrect; option 3 was selected by about 30.9% of the test takers, that is 72.7% of those who answered incorrectly. Note also that, for subjects with expected total scores below about 58, option 3 constitutes the most probable choice. Finally, item 92 in Figure 2(c), aside from being approximately monotone, is also easy, since a subject with expected total score of about 30 already has a 70% chance of selecting the correct option; only a few examinees are consequently attracted by the incorrect options 1, 3, and 4.

Expected item scores

Through the code

```
R> plot(Psych1, plottype = "EIS", item = c(24, 25, 92, 96))
```



(a) Item 24

(b) Item 25

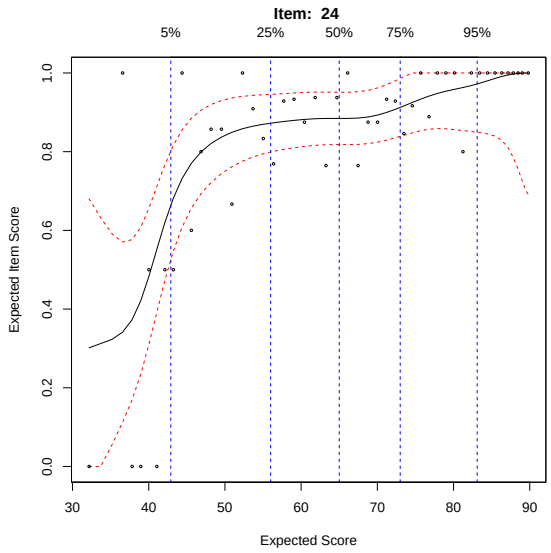
(c) Item 92

(d) Item 96

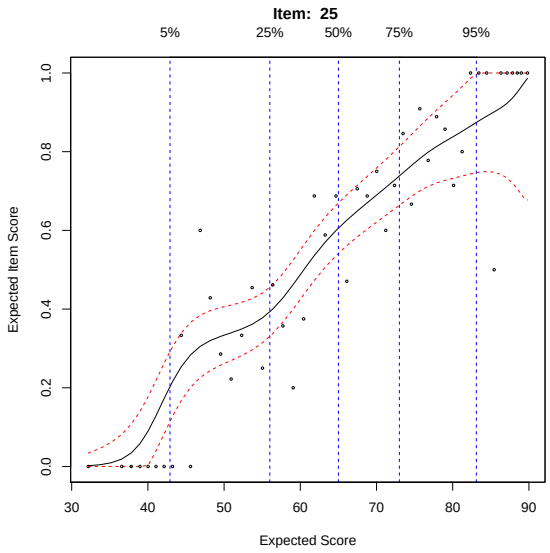
Figure 2: OCCs for items 24, 25, 92, and 96 of the introductory psychology exam.

we obtain, for the same set of items, the EISs displayed in Figure 3.

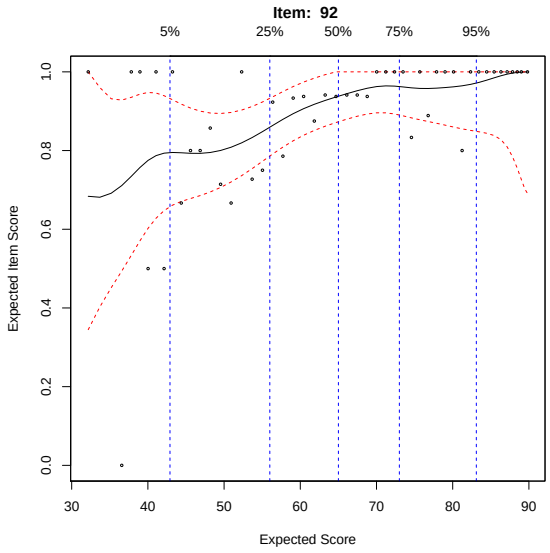
Due to the 0/1 weighting scheme, the EIS is the same as the OCC (shown in blue in Figure 2) for the correct option. EISs by default show the 95% approximate pointwise confidence intervals (dashed red lines) illustrated in Section 2.4. Via the argument `alpha`, these confidence intervals can be removed entirely (`alpha = FALSE`) or changed by specifying a different value. In this example relatively wide confidence intervals, for expected total scores at extremely high or low levels, are obtained. This is due to the fact that there are less data for estimating the curve in these regions and thus there is less precision in the estimates. Finally, the points on



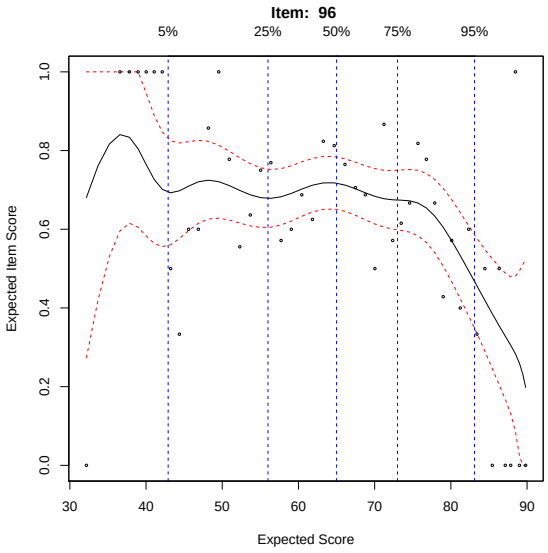
(a) Item 24



(b) Item 25



(c) Item 92



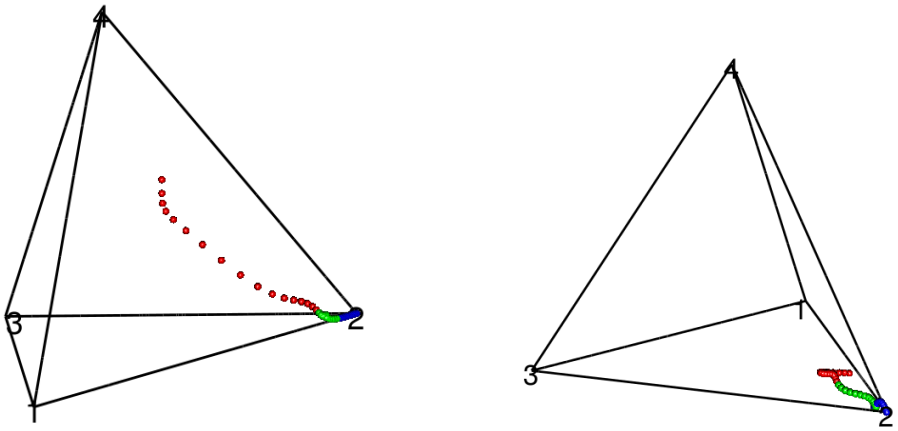
(d) Item 96

Figure 3: EISs, and corresponding 95% pointwise confidence intervals (dashed red lines), for items 24, 25, 92, and 96 of the introductory psychology exam. Grouped subject scores are displayed as points.

the EIS plots show the observed average score for the subjects grouped as in (4).

Probability simplex plots

To complement the OCCs, the package includes triangle and tetrahedron (simplex) plots that, as illustrated in Section 3.4, synthesize the OCCs. When these plots are used on



(a) Item 24 (b) Item 92

Figure 4: Probability tetrahedrons for two items of the introductory psychology exam. Low trait levels are plotted in red, medium in green, and high in blue.

items with more than 3 or 4 options (including the missing value category), only the options corresponding to the 3 or 4 highest probabilities will be shown; naturally, these probabilities are normalized in order to allow the simplex representation. This seldom loses any real information since experience tends to show that in a very wide range of situations people tend to eliminate all but a few options.

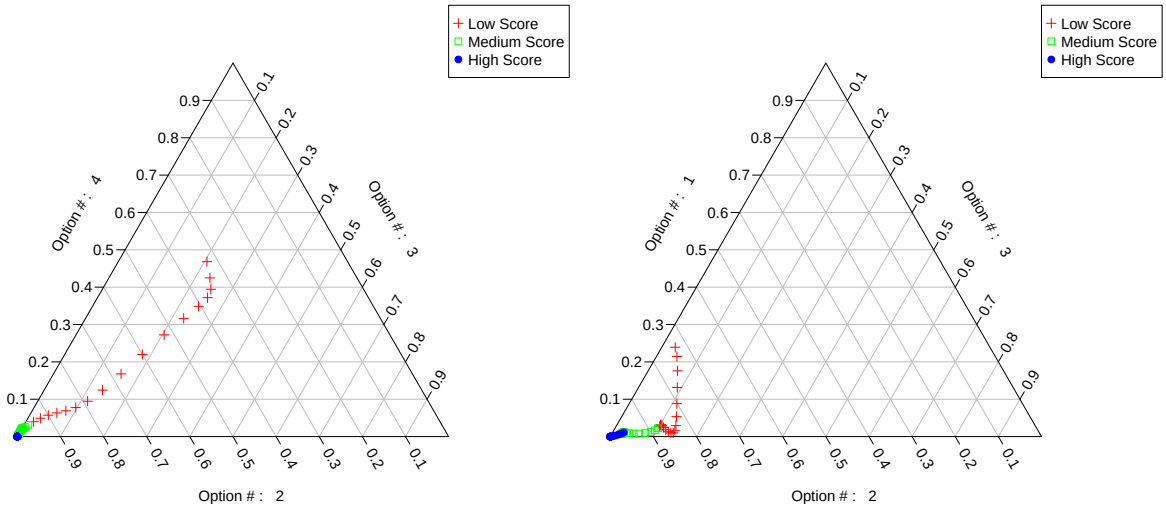
The tetrahedron is the natural choice for the items 24 and 92, characterized by four options and without "observed" missing responses; for these items the code

```
R> plot(Psych1, plottype = "tetrahedron", items = c(24, 92))
```

generates the tetrahedron plots displayed in Figure 4.

These plots may be manipulated with the mouse or keyboard. Inside the tetrahedron there is a curve constructed on the q (`nevalpoints`) evaluation points. In particular, low, medium, and high trait levels are identified by red, green, and blue points, respectively, where the levels are simply the values of `evalpoints` broken into three equal groups. Considering this ordering in the trait level, it is possible to make some considerations.

- A basic requirement of a reasonable item, of this format, is that the sequence of points terminates at or near the correct answer. In these terms, as can be noted in Figure 4(a) and Figure 4(b), items 24 and 92 satisfy this requirement since the sequence of points moves toward the correct option.
- The length of the curve is very important. The individuals with the lowest trait levels should be far from those with the highest. Item 24, in Figure 4(a), is a fairly good example. By contrast very easy items, such as item 92 in Figure 4(b), have very short curves concentrated close to the correct answer, with only the worst students showing a slight tendency to choose a wrong answer.
- The spacing of the points is related to the speed at which probabilities of choice change; compare the worst students of Figure 4(a) with those in Figure 4(b) and also the corresponding results in Figure 2(a) and Figure 2(c), respectively.



(a) Item 24 (b) Item 92

Figure 5: Probability triangles for two items of the introductory psychology exam.

For the same items, the code

```
R> plot(Psych1, plottype = "triangle", items = c(24, 92))
```

produces the triangle plots displayed in Figure 5. From Figure 5(a) we can see that in the set of the three most often chosen options, the second one has a much higher probability of being selected while the other two share almost the same probability, and so the sequence of points approximately lies on the bisector of the angle associated to the second option.

Principal component analysis

By performing a principal component analysis (PCA) of the EISs at each evaluation point, the **KernSmoothIRT** package provides a way to simultaneously compare items and to show the relationships among them. Since EISs may be defined on different ranges $[x_{j\,\min}, x_{j\,\max}]$, the transformation $\left(\widehat{e}_j\left(\vartheta\right) - x_{j\,\min}\right) / \left(x_{j\,\max} - x_{j\,\min}\right)$, $j = 1, \ldots, k$, is preliminarily applied. Furthermore, as stated in Section 2.1, in this paradigm only rank order considerations make sense, so the zero-centered ranks of $\widehat{e}_1\left(\vartheta_s\right), \ldots, \widehat{e}_k\left(\vartheta_s\right)$, for each $s = 1, \ldots, q$, are computed and the PCA is carried out on the resulting $(q \times k)$-matrix. In particular, the code

```
R> plot(Psych1, plottype = "PCA")
```

produces the graphical representation in Figure 6.

A first glance at this plot shows that:

- The first principal component, on the horizontal axis, represents item difficulty, since the most difficult items are placed on the right and the easiest ones on the left. The small plots on the left and on the right show the EISs for the two extreme items with

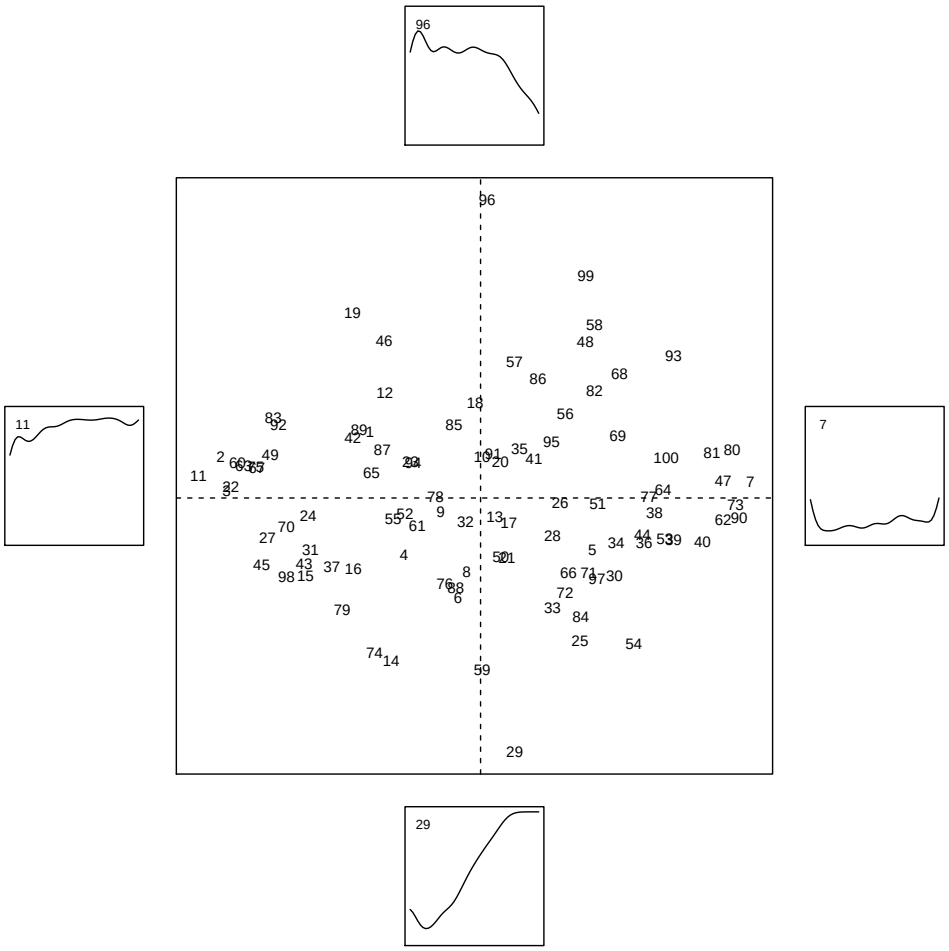
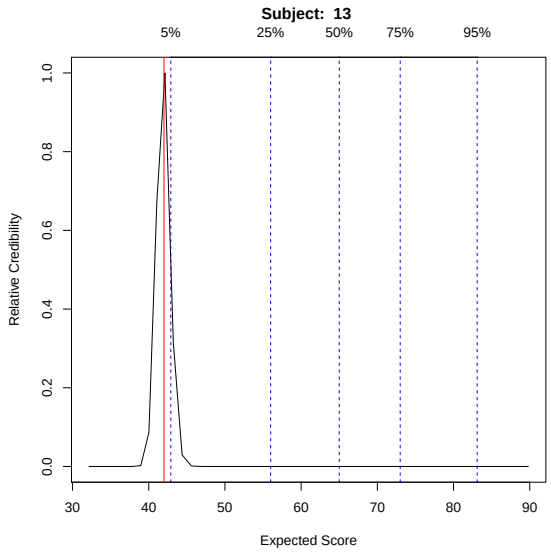


Figure 6: First two principal components for the introductory psychology exam. In the interior plot, numbers are the identifiers of the items. The vertical component represents discrimination, while the horizontal one difficulty. The small plots show the EISs for the most extreme items for each principal component.

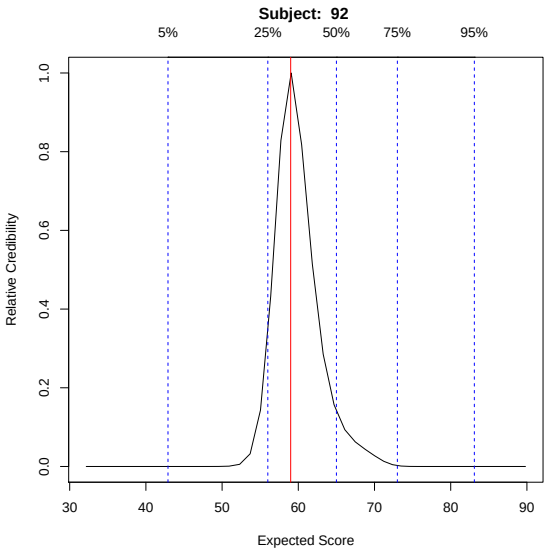
respect to this component and help the user in identifying the axis-direction with respect to difficulty (from low to high or from high to low). Here, I_7 shows high difficulty, as test takers of all ability levels receive a low score, while I_{11} is extremely easy.

- The second principal component, on the vertical axis, corresponds to item discrimination, since low items tend to have an high positive slope while high items tend to have an high negative slope. Also in this case, the small plots on the bottom and on the top show the EISs for the two extreme items with respect to this component and help the user in identifying the axis-direction with respect to discrimination (from low to high or vice versa). Here, while both I_{96} and I_{29} possess a very strong discrimination, I_{96} is clearly ill-posed, since it discriminates negatively.

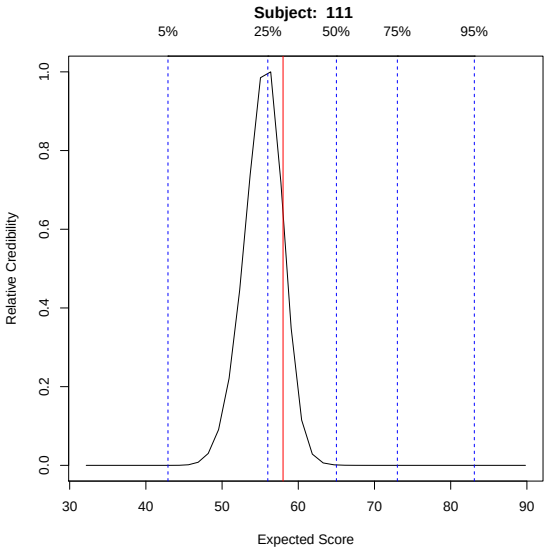
Concluding, the principal components plot tends to be a useful overall summary of the composition of the test. Figure 6 is fairly typical of most academic tests and it is also usual to have only two dominant principal components reflecting item difficulty and discrimination.



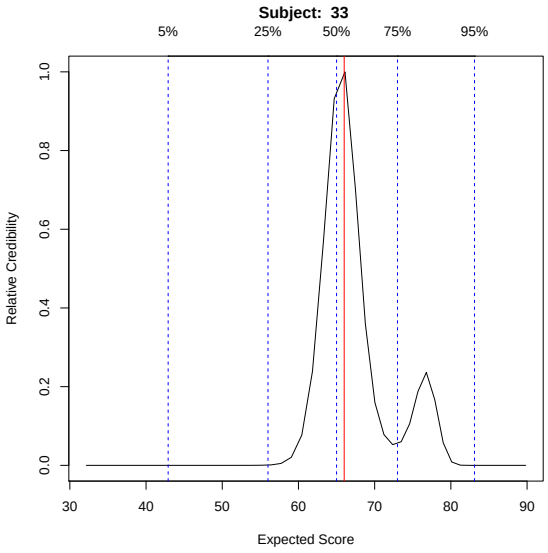
(a) Subject 13



(b) Subject 92



(c) Subject 111



(d) Subject 33

Figure 7: RCCs for some subjects. The vertical red line shows the actual score the subject received.

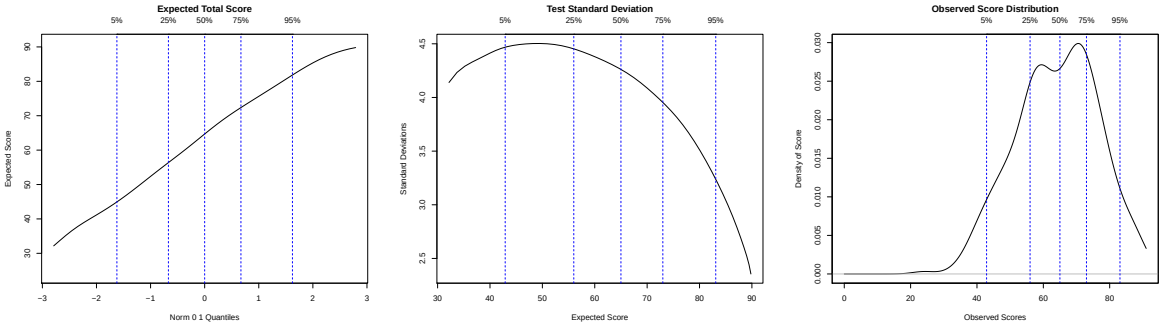
Relative credibility curves

The RCCs shown in Figure 7 are obtained by the command

```
R> plot(Psych1, plottype = "RCC", subjects = c(13, 92, 111, 33))
```

In each plot, the red line shows the subject's actual score t.

For both subjects considered in Figure 7(a) and Figure 7(b), there is a substantial agreement



(a) Expected Test Score (b) Standard Deviation (c) Density

Figure 8: Test summary plots for the introductory psychology exam.

between the maximum of the RCC, $\widehat{e}\left(\widehat{\vartheta}^{\mathrm{ML}}\right)$, and t. Nevertheless, there is a difference in terms of the precision of the ML-estimates; for S_{13} the RCC is indeed more spiky, denoting a higher precision. In Figure 7(c) there is a substantial difference between $\widehat{e}\left(\widehat{\vartheta}^{\mathrm{ML}}_{111}\right)$ and t_{111}. This indicates that the correct and incorrect answers of this subject are more consistent with a lower score than they are with the actual score received. Finally, in Figure 7(d), although there is a substantial agreement between $\widehat{e}\left(\widehat{\vartheta}^{\mathrm{ML}}_{33}\right)$ and t_{33}, a small but prominent bump is present in the right part of the plot. Although S_{33} is well represented by his total score, he passed some, albeit few, difficult items and this may imply that he is more able than t_{33} suggests.

The commands

```
R> subjscore(Psych1)

[1] 74 56 89 70 56 57 ...

R> subjscoreML(Psych1)

[1] 72.36589 59.06626 88.47615 67.47167 57.71787 55.03844 ...
```

allow us to evaluate the differences between the values of t_i and $\widehat{e}\left(\widehat{\vartheta}^{\mathrm{ML}}_i\right)$, $i = 1, \ldots, n$.

Test summary plots

The **KernSmoothIRT** package also contains many analytical tools for an overall assessment of the test. Figure 8 shows a few of these, obtained via

```
R> plot(Psych1, plottype = "expected")
R> plot(Psych1, plottype = "sd")
R> plot(Psych1, plottype = "density")
```

Figure 8(a) shows the ETS as a function of the quantiles of the standard normal distribution Φ; it is nearly linear for the Psych 101 dataset. Note that, in the nonparametric context, the

ETS may be non-monotone due to either ill-posed items or random variations. In the latter case, a slight increase of the bandwidth may be advisable.

The total score, for subjects having a particular value ϑ, is a random variable, in part because different examinees, or even the same examinee on different occasions, cannot be expected to make exactly the same choices. The standard deviation of these values, graphically represented in Figure 8(b), is therefore also a function of ϑ. Figure 8(b) indicates that the standard deviation reaches the maximum for examinees at around a total score of 50, where it is about 4.5 items out of 100. This translates into 95% confidence limits of about 41 and 59 for a subject getting 50 items correct.

Figure 8(c) shows a kernel density estimate of the distribution of the total score. Although such distribution is commonly assumed to be "bell-shaped", from this plot we can note that this assumption might not be justified for these data. In particular, a negative skewness can be noted which is a consequence of the test having relatively more easy items than hard ones. Moreover, bimodality is evident.

4.3. Voluntary HIV-1 counseling and testing efficacy study group

It is often useful to explore if, for a specific item on a test, its expected score differs when estimated on two or more different groups of subjects, commonly formed by gender or ethnicity. This is called differential item functioning (DIF) analysis in the psychometric literature. In particular, DIF occurs when subjects with the same ability but belonging to different groups have a different probability of choosing a certain option. DIF can properly be called *item bias* because the curves of an item should depend only on ϑ, and not directly on other person factors. Zumbo (2007) offers a recent review of various DIF detection methods and strategies.

The **KernSmoothIRT** package allows for a nonparametric graphical analysis of DIF, based on kernel smoothing methods. To illustrate this analysis, we use data coming from the Voluntary HIV-1 counseling and testing efficacy study, conducted in 1995–1997 by the Center for AIDS Prevention Studies at University of California, San Francisco (see The Voluntary HIV-1 Counseling and Testing Efficacy Study Group 2000a,b, for details). This study was concerned with the effectiveness of HIV counseling and testing in reducing risk behavior for the sexual transmission of HIV. To perform this study, $n = 4292$ persons were enrolled. The whole dataset – downloadable from `http://caps.ucsf.edu/research/datasets/`, which also contains other useful survey details – reported 1571 variables for each participant. As part of this study, respondents were surveyed about their attitude toward condom use via a bank of $k = 15$ items. Respondents were asked how much they agreed with each of the statements on a 4-point response scale, with 1 = "strongly disagree", 2 = "disagree more than I agree", 3 = "agree more than I disagree", 4 = "strongly agree". Since 10 individuals omitted all the 15 questions, they have been preliminarily removed from the data used. Moreover, given the ("negative") wording of the items I_2, I_3, I_5, I_7, I_8, I_{11}, and I_{14}, a respondent who strongly agreed with such statements was indicating a less favorable attitude toward condom use. In order to uniform the data, the score for these seven items was preliminarily reversed. The dataset so modified can be directly loaded from the **KernSmoothIRT** package with the code

```
R> data("HIV", package = "KernSmoothIRT")
R> HIV
```

```
     SITE GENDER AGE  1  2  3  4  5  6  7  8  9  10  11  12  13  14  15
1     Ken      F  17  4  1  1  4  1  2  4  4  4   4   3   4   1   2   4
2     Ken      F  17  4  2  4  4  2  3  1  4  3   3   2   3   4   1   4
3     Ken      F  18  4  4  4  4  4  1  4  4  4   1  NA   4   1  NA   4
.     .        .   .  .  .  .  .  .  .  .  .  .   .   .   .   .   .   .
.     .        .   .  .  .  .  .  .  .  .  .  .   .   .   .   .   .   .
.     .        .   .  .  .  .  .  .  .  .  .  .   .   .   .   .   .   .
4281  Tri      M  79  4  4  1  4  1 NA  4 NA  4  NA  NA  NA  NA   1   4
4282  Tri      M  80  4 NA  4  4  1  4 NA NA NA   1  NA   4   1   4  NA
```

```
R> attach(HIV)
```

As it can be easily seen, the above data frame contains the following person factors:

`SITE`	=	"site of the study" (`Ken` = Kenya, `Tan` = Tanzania, `Tri` = Trinidad)
`GENDER`	=	"subject's gender" (`M` = male, `F` = female)
`AGE`	=	"subject's age" (age at last birthday)

Each of these factors can potentially be used for a DIF analysis. These data have been also analyzed, through some well-known parametric models, by Bertoli-Barsotti, Muschitiello, and Punzo (2010) which also perform a DIF analysis. Part of this sub-questionnaire has been also considered by De Ayala (2003, 2009) with a Rasch analysis.

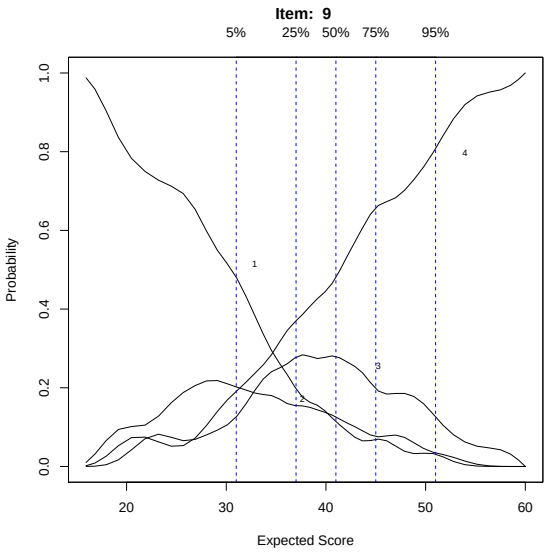
The code below

```
R> HIVres <- ksIRT(HIV[, -(1:3)], key = HIVkey, format = 2, miss = "omit")
R> plot(HIVres, plottype = "OCC", item = 9)
R> plot(HIVres, plottype = "EIS", item = 9)
R> plot(HIVres, plottype = "tetrahedron", item = 9)
```

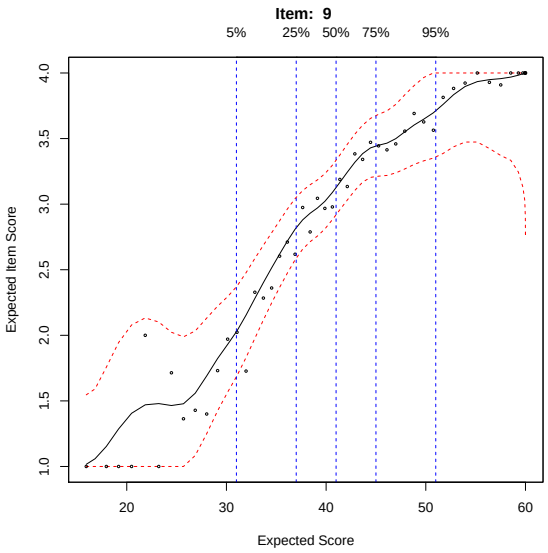
produces the plots, for I_9, displayed in Figure 9. The option `miss = "omit"` excludes from the nonparametric analysis all the subjects with at least one omitted answer, leading to a sample of 3473 respondents; the option `format = 2` specifies that the data contain rating scale items. Figure 9(a) displays the OCCs for the considered item. As expected, subjects with the smallest scores are choosing the first option while those with the highest ones are selecting the fourth option. Generally, as the total scores increase, respondents are approximately estimated to be more likely to choose an higher option and this reflects the typical behavior of a rating scale item. Figure 9(b) shows the EIS for I_9. Note how the expected item score climbs consistently as the total test score increases. Moreover, the EIS displays a fairly monotone behavior that covers the entire range $[1, 4]$. Finally, Figure 9(c) shows the tetrahedron for I_9. It corroborates the good behavior of I_9 already seen in Figure 9(a) and Figure 9(b). The sequence of points herein, as expected, starts from (the vertex) option 1 and smoothly tends to option 4, passing by option 2 and option 3.

The following example demonstrates DIF analysis using the person factor `GENDER`. To perform this analysis, a new '`ksIRT`' object must be created with the addition of the `groups` argument by which the different subgroups may be specified. In particular, the code

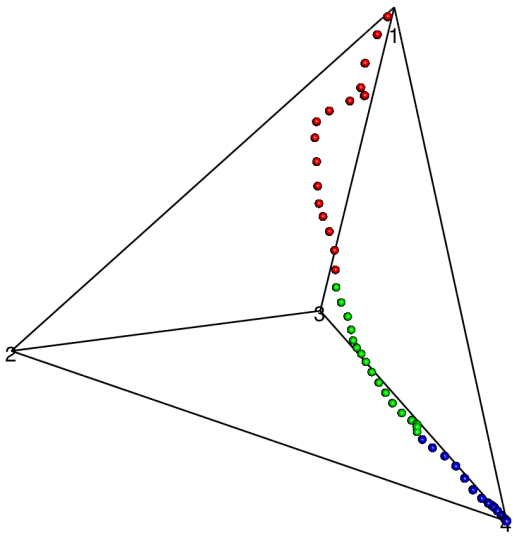
```
R> DIF1 <- ksIRT(res = HIV[, -(1:3)], key = HIVkey, format = 2,
+    groups = GENDER, miss = "omit")
```



(a) OCCs



(b) EIS



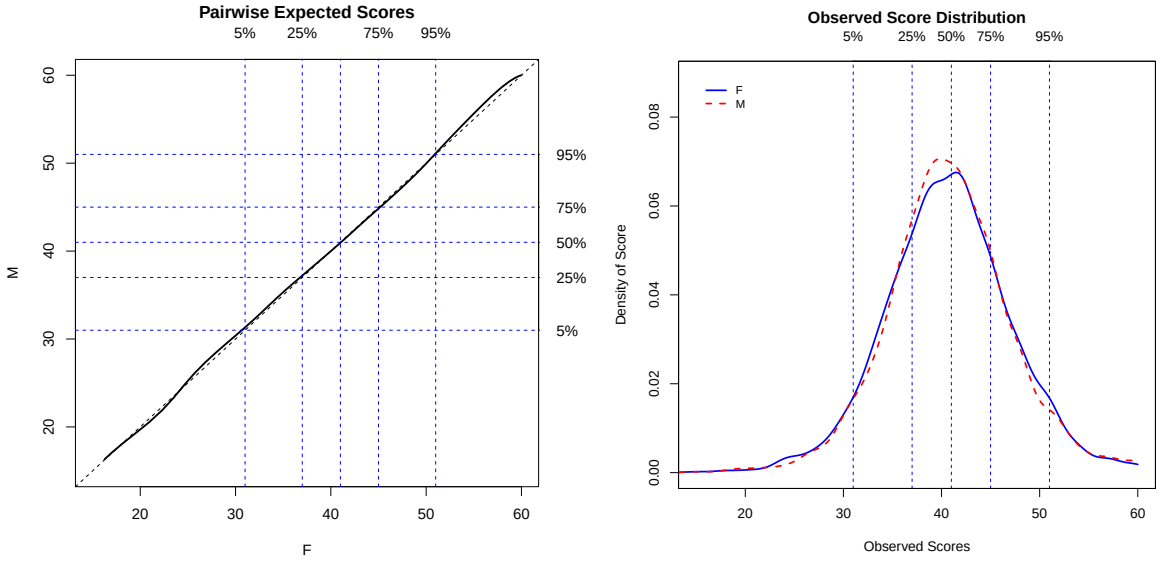
(c) Tetrahedron

Figure 9: Item 9 from the voluntary HIV-1 counseling and testing efficacy study group.

```
R> plot(DIF1, plottype = "expectedDIF", lwd = 2)
R> plot(DIF1, plottype = "densityDIF", lwd = 2)
```

produces the plots in Figure 10.

Figure 10(a) displays the QQ-plot between the distributions of the expected scores for males and females; if the performances of the two groups are about the same, the relationship will appear as a nearly diagonal line (a dotted diagonal line is plotted as a reference). Figure 10(b) shows the density functions for the two groups. Both plots confirm that there is a strong agreement in behavior for males and females with respect to the test.



(a) Expected scores

(b) Density

Figure 10: Behavior of males (`M`) and females (`F`) on the test. In the QQ-plot on the left, the dashed diagonal line indicates the reference situation of no difference in performance for the two groups; the horizontal and vertical dashed blue lines indicate the 5%, 25%, 50%, 75%, and 95% quantiles for the two groups.

After this preliminary phase, the DIF analysis proceeds by considering the item by item group comparisons. Figure 11, obtained via the command

```
R> plot(DIF1, plottype = "OCCDIF", cex = 0.5, item = 3)
```

displays the OCCs for the (rating scale) item I_3.

These plots allow the user to compare the two groups at the item level. Lack of DIF is evident by nearly overlapping OCCs for all the four options. DIF may also be evaluated in terms of the expected score of the groups, as displayed in Figure 12.

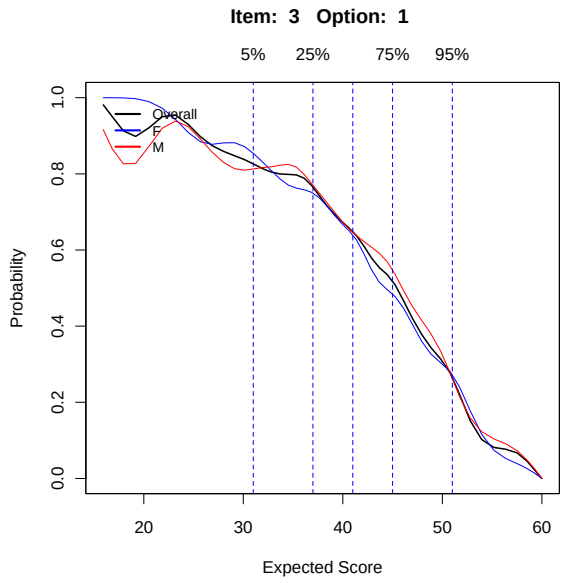
This plot is obtained with the code

```
R> plot(DIF1, plottype = "EISDIF", cex = 0.5, item = 3)
```

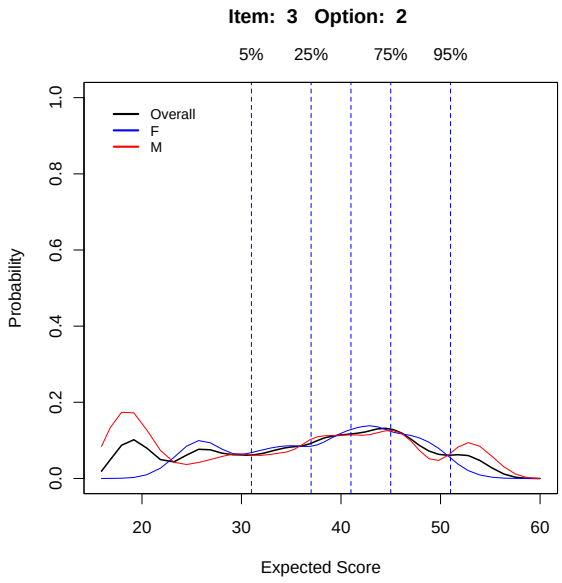
The different color points on the plot represent how individuals from the groups actually scored on the item. Although we focused the attention only on I_3, similar results are obtained for all of the other items in $\mathcal{I}$, and this confirms that `GENDER` is not a variable producing DIF in this study. This result is corroborated in Bertoli-Barsotti *et al.* (2010). Note that, for both OCCs and EISs, it is possible to add confidence intervals through the `alpha` argument.

The code

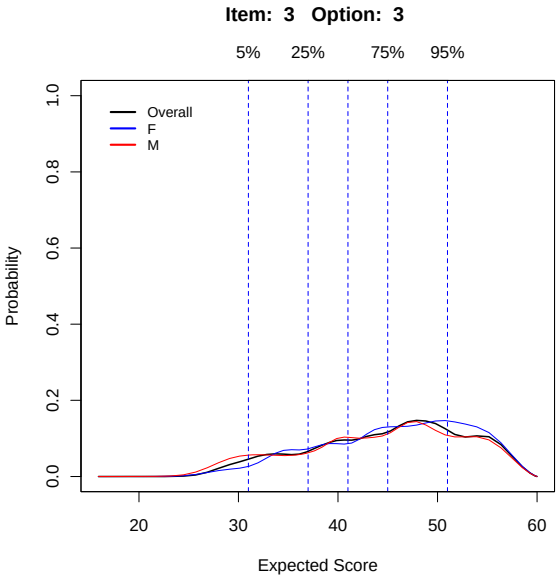
```
R> DIF2 <- ksIRT(res = HIV[, -(1:3)], key = HIVkey, format = 2,
+    groups = SITE, miss = "omit")
R> plot(DIF2, plottype = "expectedDIF", lwd = 2)
R> plot(DIF2, plottype = "densityDIF", lwd = 2)
```



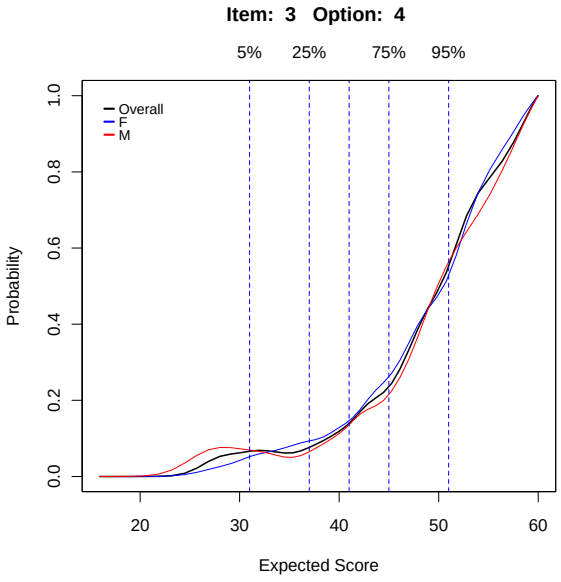
(a) Option 1



(b) Option 2



(c) Option 3



(d) Option 4

Figure 11: OCCs, for males and females, related to item 3 of the voluntary HIV-1 counseling and testing efficacy study group. The overall OCCs are superimposed.

produces separate plots for subjects with different `SITE` levels (Figure 13).

Among the 3473 subjects answering to all the 15 items, 984 come from Trinidad, 1143 from Kenya and 1346 from Tanzania. As highlighted by Bertoli-Barsotti *et al.* (2010), there are differences among these groups, and Figure 13 shows this. The three pairwise QQ-plots of the expected score distributions show that according to Figure 13(c) there is a slight dominance of people from Kenia over people from Trinidad (in the sense that people from Kenia have, in

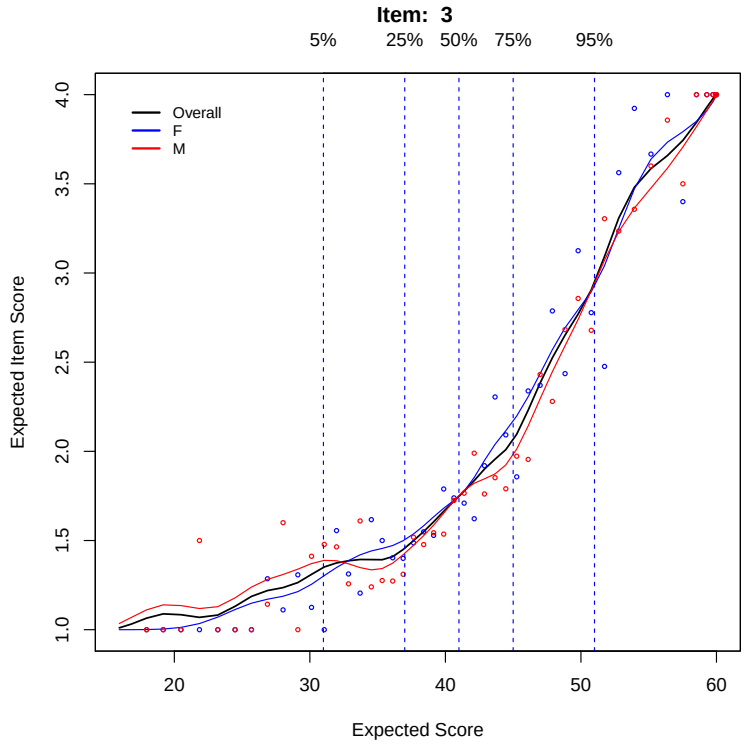


Figure 12: Overall EIS and EIS of males and females, for item 3.

distribution, a slightly more positive attitude toward condom use than people from Trinidad), and a large discrepancy between the performances of people from Tanzania and those of the other two groups, as shown in Figure 13(a) and Figure 13(b). The above dominance, and the peculiar behavior of people from Tanzania compared with the other countries, can be also noted by looking at the observed total score densities in Figure 13(d). Here, there is higher variability in the total score for people from Tanzania. But what about DIF? The command

```
R> plot(DIF2, plottype = "EISDIF", item = c(6, 11))
```

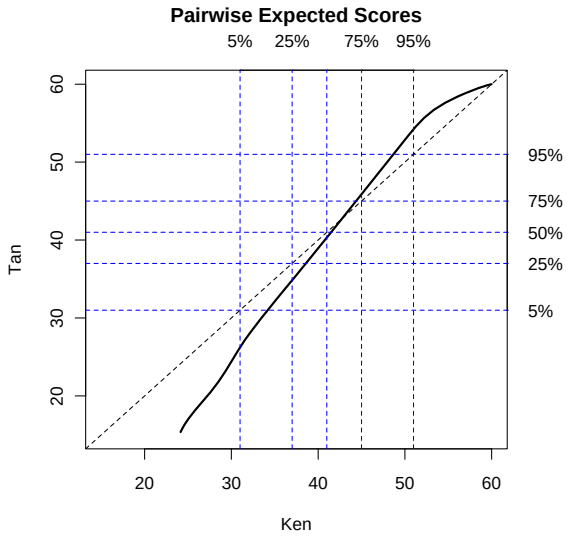
produces, for I_6 and I_{11}, the EISs in Figure 14.

In both the plots we have a graphical indication of the presence of DIF, and this confirms the results by Bertoli-Barsotti *et al.* (2010) that detect `SITE`-based DIF for these and other items in the test.

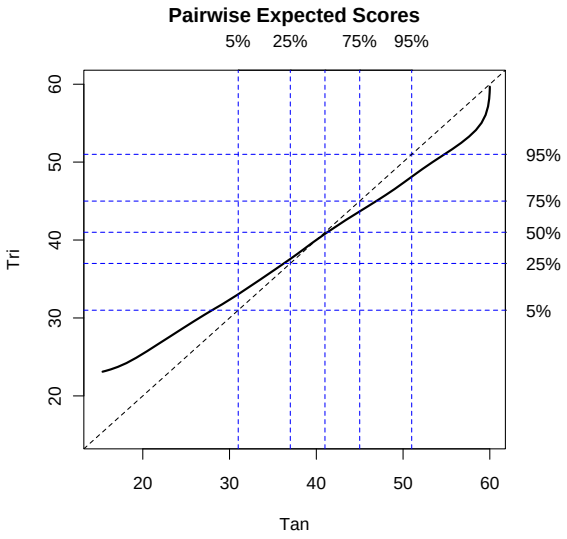
5. Conclusions

In this paper, package **KernSmoothIRT** for the R environment, which allows for kernel smoothing within the IRT context, has been introduced. Two applications have been discussed, along with some theoretical and practical issues.

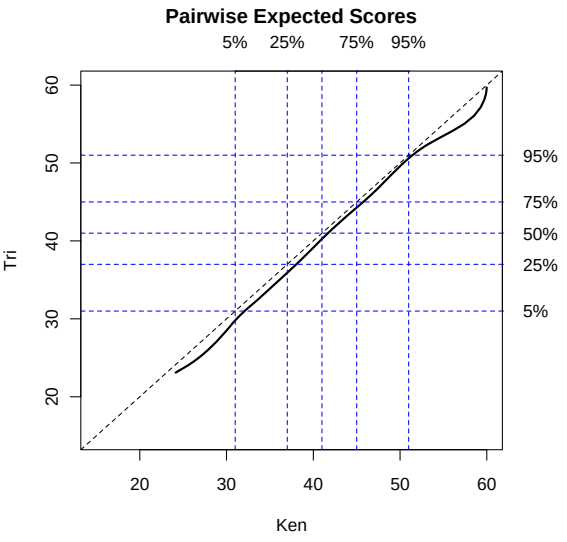
The advantages of nonparametric IRT modeling are well known. Ramsay (2000) recommends its application, at least as an exploratory tool, to guide users in their choice of an appropriate parametric model. Moreover, while currently most IRT analyses are conducted with



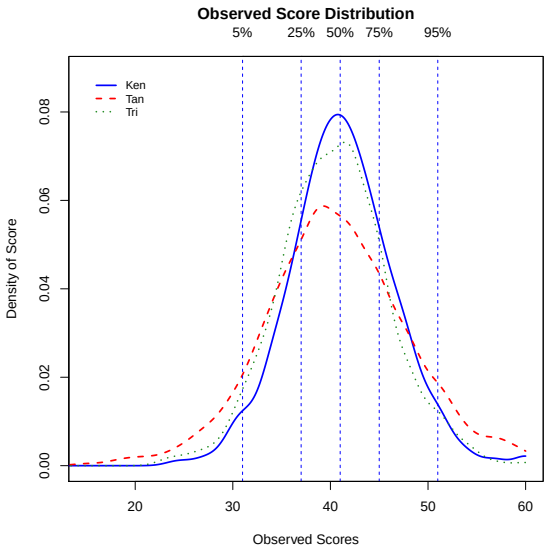
(a) QQ-plot (`Ken` vs. `Tan`)



(b) QQ-plot (`Tan` vs. `Tri`)



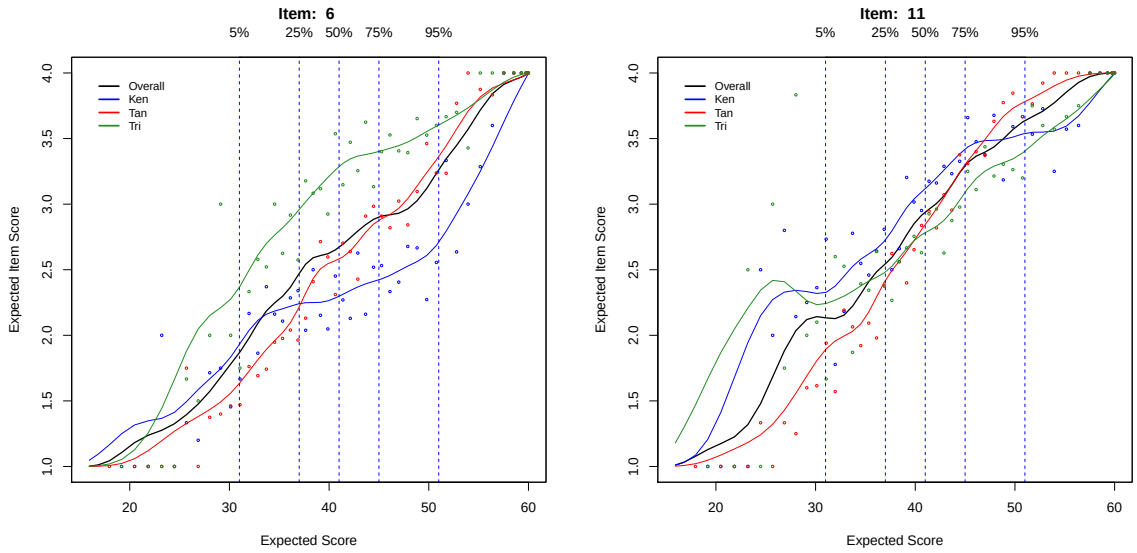
(c) QQ-plot (`Ken` vs. `Tri`)



(d) Total score density

Figure 13: Behavior of people from Kenya (`Ken`), Tanzania (`Tan`), and Trinidad (`Tri`), on the test. In all the pairwise QQ-plots, the dashed diagonal line indicates the reference situation of no difference in performance for the two groups; the horizontal and vertical dashed blue lines indicate the 5%, 25%, 50%, 75%, and 95% quantiles for the two groups.

parametric models, quite often the assumptions underling parametric IRT modeling are not preliminarily checked. One reason for this may be the lack, apart from **TestGraf**, of available software. **TestGraf** has set a milestone in this field as the first computer program to implement a kernel smoothing approach to IRT and has been the most prominent software for years. Compared to **TestGraf**, **KernSmoothIRT** has the major advantage of running within the R environment. Users do not have to export their results into another piece of software in



(a) Item 6 (b) Item 11

Figure 14: Overal EISs and EISs for people from Kenya (`Ken`), Tanzania (`Tan`), and Trinidad (`Tri`) separately, for items 6 and 11 of the voluntary HIV-1 counseling and testing efficacy study group.

order to perform non-standard data analysis, to produce customized plots or to perform parametric IRT, which is possible in R using any of several add-on packages (see also Mair 2014). Furthermore, **KernSmoothIRT** allows more flexibility in bandwidth and kernel selection, as well as in handling missing values.

We believe that **KernSmoothIRT** may prove useful to educators, psychologists, and other researchers developing questionnaires, enabling them to spot ill-posed questions and to formulate more plausible wrong options. Future works will consider extending the package by allowing for kernel smoothing estimation of test and item information functions. Although well-established in parametric IRT, information functions present serious statistical problems in the NIRT context, as underlined by Ramsay (2000, p. 66). Currently available nonparametric-based IRT programs, such as **TestGraf**, estimate test and item information functions based on parametric OCCs.

Acknowledgments

The Voluntary HIV-1 counseling and testing efficacy study was sponsored by UNAIDS/WHO, AIDSCAP/Family Health International, and the Center for AIDS Prevention Studies at the University of California, San Francisco.

References

Aitchison J (2003). *The Statistical Analysis of Compositional Data*. Blackburn Press, Cald-

well.

Altman NS (1992). "An Introduction to Kernel and Nearest-Neighbor Nonparametric Regression." *The American Statistician*, **46**(3), 175–185.

Azzalini A, Bowman AW, Härdle W (1989). "On the Use of Nonparametric Regression for Model Checking." *Biometrika*, **76**(1), 1–11.

Baker FB, Kim SH (2004). *Item Response Theory: Parameter Estimation Techniques*. Statistics: A Dekker Series of Textbooks and Monographs. Marcel Dekker, New York.

Bartholomew DJ (1983). "Latent Variable Models for Ordered Categorical Data." *Journal of Econometrics*, **22**(1–2), 229–243.

Bartholomew DJ (1988). "The Sensitivity of Latent Trait Analysis to Choice of Prior Distribution." *British Journal of Mathematical and Statistical Psychology*, **41**(1), 101–107.

Bertoli-Barsotti L, Muschitiello C, Punzo A (2010). "Item Analysis of a Selected Bank from the Voluntary HIV-1 Counseling and Testing Efficacy Study Group." *Technical Report 1*, Dipartimento di Matematica, Statistica, Informatica e Applicazioni (Lorenzo Mascheroni), Università degli Studi di Bergamo. URL http://aisberg.unibg.it/bitstream/10446/444/1/WPMateRi01(2010).pdf.

Chang HH, Mazzeo J (1994). "The Unique Correspondence of the Item Response Function and Item Category Response Functions in Polytomously Scored Item Response Models." *Psychometrika*, **59**(3), 391–404.

De Ayala RJ (2003). "The Effect of Missing Data on Estimating a Respondent's Location Using Ratings Data." *Journal of Applied Measurement*, **4**(1), 1–9.

De Ayala RJ (2009). *The Theory and Practice of Item Response Theory*. Methodology in the Social Sciences. Guilford Press, New York.

de Leeuw J, Mair P (2007). "An Introduction to the Special Volume on "Psychometrics in R"." *Journal of Statistical Software*, **20**(1), 1–5. URL http://www.jstatsoft.org/v20/i01.

DeMars C (2010). *Item Response Theory*. Understanding Statistics. Oxford University Press, New York.

Douglas JA (1997). "Joint Consistency of Nonparametric Item Characteristic Curve and Ability Estimation." *Psychometrika*, **62**(1), 7–28.

Douglas JA (2001). "Asymptotic Identifiability of Nonparametric Item Response Models." *Psychometrika*, **66**(4), 531–540.

Douglas JA, Cohen A (2001). "Nonparametric Item Response Function Estimation for Assessing Parametric Model Fit." *Applied Psychological Measurement*, **25**(3), 234–243.

Guttman L (1947). "The Cornell Technique for Scale and Intensity Analysis." *Educational and Psychological Measurement*, **7**(2), 247–279.

Guttman L (1950a). "Relation of Scalogram Analysis to other Techniques." In SA Stouffer, L Guttman, FA Suchman, PF Lazarsfeld, SA Star, JA Clausen (eds.), *Measurement and Prediction*, volume 4 of *Studies in Social Psychology in World War II*, pp. 172–212. Princeton University Press, Princeton.

Guttman LA (1950b). "The Basis for Scalogram Analysis." In SA Stouffer, L Guttman, FA Suchman, PF Lazarsfeld, SA Star, JA Clausen (eds.), *Measurement and Prediction*, volume 4 of *Studies in Social Psychology in World War II*, pp. 60–90. Princeton University Press, Princeton.

Härdle W (1990). *Applied Nonparametric Regression*, volume 19 of *Econometric Society Monographs*. Cambridge University Press, Cambridge.

Junker BW, Sijtsma K (2001). "Nonparametric Item Response Theory in Action: An Overview of the Special Issue." *Applied Psychological Measurement*, **25**(3), 211–220.

Kutylowski AJ (1997). "Nonparametric Latent Factor Analysis of Occupational Inventory Data." In J Rost, R Langeheine (eds.), *Applications of Latent Trait and Latent Class Models in the Social Sciences*, pp. 253–266. Vaxmann, New York.

Lei PW, Dunbar SB, Kolen MJ (2004). "A Comparison of Parametric and Non-Parametric Approaches to Item Analysis for Multiple-Choice Tests." *Educational and Psychological Measurement*, **64**(3), 1–23.

Lindsey JK (1973). *Inferences from Sociological Survey Data: A Unified Approach*, volume 3 of *Progress in Mathematical Social Sciences*. Elsevier Scientific Publishing Company, Amsterdam.

Lord FM (1980). *Application of Item Response Theory to Practical Testing Problems*. Lawrence Erlbaum Associates, Hillsdale.

Mair P (2014). "CRAN Task View: Psychometric Models and Methods." Version 2014-03-18, URL http://CRAN.R-project.org/view=Psychometrics.

Marron JS, Nolan D (1988). "Canonical Kernels for Density Estimation." *Statistics & Probability Letters*, **7**(3), 195–199.

Mazza A, Punzo A (2011). "Discrete Beta Kernel Graduation of Age-Specific Demographic Indicators." In S Ingrassia, R Rocci, M Vichi (eds.), *New Perspectives in Statistical Modeling and Data Analysis*, Studies in Classification, Data Analysis, and Knowledge Organization, pp. 127–134. Springer-Verlag.

Mazza A, Punzo A (2013a). "Graduation by Adaptive Discrete Beta Kernels." In A Giusti, G Ritter, M Vichi (eds.), *Classification and Data Mining*, Studies in Classification, Data Analysis, and Knowledge Organization, pp. 243–250. Springer-Verlag.

Mazza A, Punzo A (2013b). "Using the Variation Coefficient for Adaptive Discrete Beta Kernel Graduation." In P Giudici, S Ingrassia, M Vichi (eds.), *Statistical Models for Data Analysis*, Studies in Classification, Data Analysis, and Knowledge Organization, pp. 225–232. Springer-Verlag.

Mazza A, Punzo A (2014). "**DBKGrad**: An R Package for Mortality Rates Graduation by Discrete Beta Kernel Techniques." *Journal of Statistical Software, Code Snippets*, **57**(2), 1–18. URL http://www.jstatsoft.org/v57/c02/.

Mazza A, Punzo A, McGuire B (2014). ***KernSmoothIRT**: Nonparametric Item Response Theory*. R package version 6.1, URL http://CRAN.R-project.org/package=KernSmoothIRT.

Mokken RJ (1971). *A Theory and Procedure of Scale Analysis*, volume 1 of *Methods and Models in the Social Sciences*. The Gruyter, Berlin, Germany.

Nadaraya EA (1964). "On Estimating Regression." *Theory of Probability and Its Applications*, **9**(1), 141–142.

Nering ML, Ostini R (2010). *Handbook of Polytomous Item Response Theory Models*. Taylor & Francis, New York.

Olsson U, Drasgow F, Dorans NJ (1982). "The Polyserial Correlation Coefficient." *Psychometrika*, **47**(3), 337–347.

Ostini R, Nering ML (2006). *Polytomous Item Response Theory Models*, volume 144 of *Quantitative Applications in the Social Sciences*. Sage, London.

Punzo A (2009). "On Kernel Smoothing in Polytomous IRT: A New Minimum Distance Estimator." *Quaderni di Statistica*, **11**, 15–37.

Ramsay JO (1991). "Kernel Smoothing Approaches to Nonparametric Item Characteristic Curve Estimation." *Psychometrika*, **56**(4), 611–630.

Ramsay JO (1997). "A Functional Approach to Modeling Test Data." In WJ Van der Linden, RK Hambleton (eds.), *Handbook of Modern Item Response Theory*, pp. 381–394. Springer-Verlag, New York.

Ramsay JO (2000). ***TestGraf**: A Program for the Graphical Analysis of Multiple Choice Test and Questionnaire Data*. URL http://www.psych.mcgill.ca/faculty/ramsay/ramsay.html.

Ramsay JO, Abrahamowicz M (1989). "Binomial Regression with Monotone Splines: A Psychometric Application." *Journal of the American Statistical Association*, **84**(408), 906–915.

R Core Team (2014). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org.

Rice J (1984). "Bandwidth Choice for Nonparametric Regression." *The Annals of Statistics*, **12**(4), 1215–1230.

Silverman BW (1986). *Density Estimation for Statistics and Data Analysis*, volume 26 of *Monographs on Statistics & Applied Probability*. Chapman & Hall, London.

Simonoff JS (1996). *Smoothing Methods in Statistics*. Springer Series in Statistics. Springer-Verlag, New York.

The Voluntary HIV-1 Counseling and Testing Efficacy Study Group (2000a). "Efficacy of Voluntary HIV-1 Counseling and Testing in Individual and Couples in Kenya, Tanzania, and Trinidad: A Randomised Trial." *Lancet*, **356**(9224), 103–112.

The Voluntary HIV-1 Counseling and Testing Efficacy Study Group (2000b). "The Voluntary HIV-1 Counseling and Testing Efficacy Study: Design and Methods." *AIDS and Behavior*, **4**(1), 5–14.

Thissen D, Steinberg L (1986). "A Taxonomy of Item Response Models." *Psychometrika*, **51**(4), 567–577.

Van der Ark LA (2001). "Relationships and Properties of Polytomous Item Response Models." *Applied Psychological Measurement*, **25**(3), 273–282.

Van der Ark LA (2007). "Mokken Scale Analysis in R." *Journal of Statistical Software*, **20**(11), 1–19. URL http://www.jstatsoft.org/v20/i11/.

Van der Ark LA (2012). "New Developments in Mokken Scale Analysis in R." *Journal of Statistical Software*, **48**(5), 1–27. URL http://www.jstatsoft.org/v48/i05/.

Van der Linden WJ, Hambleton RK (1997). *Handbook of Modern Item Response Theory*. Springer-Verlag, New York.

Watson GS (1964). "Smooth Regression Analysis." *Sankhya A*, **26**(4), 359–372.

Wickelmaier F, Strobl C, Zeileis A (2012). "Psychoco: Psychometric Computing in R." *Journal of Statistical Software*, **48**(1), 1–5. URL http://www.jstatsoft.org/v48/i01/.

Wong WH (1983). "On the Consistency of Cross-Validation in Kernel Nonparametric Regression." *The Annals of Statistics*, **11**(4), 1136–1141.

Zumbo BD (2007). "Three Generations of Differential Item Functioning (DIF) Analyses: Considering Where it Has Been, Where it is Now, and Where it is Going." *Language Assessment Quarterly*, **4**(2), 223–233.

Affiliation:

Angelo Mazza, Antonio Punzo
Department of Economics and Business
University of Catania
Corso Italia, 55, 95129 Catania, Italy
E-mail: a.mazza@unict.it, antonio.punzo@unict.it
URL: http://www.economia.unict.it/a.mazza/,
http://www.economia.unict.it/punzo/

Brian McGuire
Department of Statistics
Montana State University
Bozeman, Montana, United States of America
E-mail: mcguirebc@gmail.com

Journal of Statistical Software http://www.jstatsoft.org/
published by the American Statistical Association http://www.amstat.org/
Volume 58, Issue 6 *Submitted:* 2012-05-08
June 2014 *Accepted:* 2014-02-10