Electronic Journal of Applied Statistical Analysis
EJASA, Electron. J. App. Stat. Anal.
http://siba-ese.unisalento.it/index.php/ejasa/index
e-ISSN: 2070-5948
DOI: 10.1285/i20705948v7n1p37

A modified minimum divergence estimator: some preliminary results for the Rasch model

By Lando and Bertoli-Barsotti

April 26, 2014

Electronic Journal of Applied Statistical Analysis
Vol. 7, Issue 1, 2014, 37-57
DOI: 10.1285/i20705948v7n1p37

A modified minimum divergence estimator: some preliminary results for the Rasch model

Tommaso Lando*[a] and Lucio Bertoli-Barsotti[b]

[a] *Department of Finance, VŠB-Technical University of Ostrava, Ostrava, Czech Republic*
[b] *Department of Management, Economics and Quantitative Methods, University of Bergamo, Italy*

April 26, 2014

Since its introduction, the joint maximum likelihood (JML) has been widely used as an estimation method for Rasch measurement models. As is well known, when the JML method is used, all item and person parameters are regarded as unknowns to be estimated. In this paper we focus on some drawbacks of the JML for the Rasch model: viz. i) the occasional non-existence of estimates, and ii) the bias of item parameter estimates. We propose a new estimation method which is based on the Minimum Divergence Estimation approach and consists in appropriately modifying the empirical distribution function. We provide empirical evidence that this method can solve the problem of the non-existence of the estimates and, at the same time, can reduce the bias of item parameter estimates compared to those obtained with both traditional JML estimation and the $(k-1)/k$ correction factor (where k is the number of items) commonly applied in JML software.

keywords: Rasch model, MLE, minimum divergence estimate, bias.

1 Introduction

The Rasch model (RM, Rasch 1960) is a well-known special case of the general Item Response Theory (IRT). In that context, the RM refers to the problem of measurement of a latent (unobservable) quantitative trait, say "ability", given a set of items used to measure that latent trait. In the RM, each item is also characterized by an item

*Corresponding author: lucio.bertoli-barsotti@unibg.it.

ISSN: 2070-5948
http://siba-ese.unisalento.it/index.php/ejasa/index

parameter, which may be interpreted as the "difficulty" of that item. Since its introduction, the RM – sometimes also referred to as the one-parameter logistic model – has been widely applied in many areas other than psychometrics alone. Indeed, there are some remarkable properties of the RM which are of large practical use for resolving the problem of the "measurement" of latent variables. First, the RM is usually considered able to measure on an interval scale level. Second, the RM fulfils the requirement of generalizability, in that it produces estimates (measurements) that do not depend on conditions of measurement. For example, estimates of a person's ability do not depend on the sample of persons being examined and the particular items used in the test. This important property is also known as specific objectivity. To estimate the item and person parameters, the likelihood may be jointly maximized. This method is usually called joint maximum likelihood estimation (JML) – and it coincides with the usual maximum likelihood estimation (MLE). Hence, in what follows, we will use these terms interchangeably.

The JML procedure suffers from a number of problems, the most important of which is the presence of infinitely many incidental parameters. Moreover, in particular,

i. (*existence problem*) for some particular datasets, finite estimates of parameters are not available;

ii. (*bias problem*) item parameter estimates are biased, especially for short tests, independently from the sample size.

These problems also hamper the normal approximation algorithm PROX (Linacre 2009; for a recent application of the use of PROX method see Lucadamo 2010). Problem i) is due to the fact that, for fixed sample size, there is always a positive probability of observing data lying on the boundary of the sample space. This is a well-known problem in the field of logistic regression modeling (see Agresti 2007), and more generally within the class of the exponential family of distributions. In classic logistic regression, this problem is known as *perfect discrimination* (Agresti 2007, p.153), or also as the problem of missing *overlap* (between successes and failures, (Albert and Anderson, 1984). Several techniques have been suggested to remedy this problem. For example, in regard to the classic logistic regression model Rousseeuw and Christmann (2003) proposed the Maximum Estimated Likelihood (MEL) method and proved that it yields finite estimates under certain assumptions. More specifically, the authors presented a *hidden logistic model*: the model assumes that the true responses are unobservable, so that the true empirical observations are substituted by so-called *pseudo-observations*, which are linked to the true ones by probabilistic assumptions in the form of probabilities of misclassification errors. Type I and type II error probabilities are given by $(1 - \delta_1)$ and δ_0, where δ_0 and δ_1 are unknown parameters. Although similar formulas (estimators) can be derived from Rousseeuw and Christmann's approach, problem formulation and applications are clearly quite different from ours. We shall return to this point later. We can also cite Kosmidis (2007) who proved that the modified-score functions approach to bias reduction proposed by Firth (1993) fulfils -under certain regularity assumptions- the same property of finiteness estimates. These important results still have to be studied in detail in the special case of the JML estimation of the RM (that is, in a context in which the dimension of the unknown parameter is not fixed, but grows with the sample size). We recall that, for the RM, the necessary and sufficient conditions for nonexistence of MLE are stated by Fischer (1981). They occur: a) in the presence of zero and/or perfect person totals; b) in the presence of zero and/or perfect items totals (that is, null-categories); c) for other special configurations of the dataset (so-called ill-conditioned datasets). The bias problem ii) is also sometimes referred to as a problem of "inconsistency" (Jansen et al., 1988). Indeed, an estimator is consistent if, as sample size approaches infinity, the estimated value approaches the true value. Andersen (1980) proved that when the MLE is used, there is no guarantee that increasing sample size (that is, the number of respondents) will yield better item parameter estimates. At present, the best-known solution for correcting the bias of the MLEs of the item parameters of the RM is the one suggested by Wright and Douglas (1977) (see also Wright 1977, Wright 1988): to be specific, Wright and Douglas argue that a correction factor of $(k\text{-}1)/k$, where k is the number of items, remove most of the bias. This approach will henceforth be denoted as corrected-JML (C-JML). To jointly solve the estimation problems at their source a new method is introduced in the present paper: its key principle consists in the minimum divergence (MD) estimation procedure (Ali and Silvey, 1965). As well known, the MD methods include, as special cases, the classic maximum likelihood method as well as minimum chi-squared methods. In particular, the proposed method consists in adjusting the empirical distribution of the observations. Since the sensitivity of this adjustment depends on an arbitrarily small positive number ε, the method is called *ε-adjustment*. In particular, this paper focuses on the adjusted version of the ML, which is simply based on the adjustment of the Kullback-Leibler divergence. The right choice of ε is somewhat troublesome, since it can depend on several elements including the size of the dataset; nonetheless it can be found empirically and approximately. Simulation studies show that, on setting ε conveniently, the method provides acceptable estimates, even for well-conditioned datasets. Moreover, we found that the ε-adjustment method may also be very useful for correcting the bias of the MLEs of the item parameters and it seems to work even better for this purpose than the C-JML which is known to be, at present, the best solution for the bias problem. For this reason, in section 4.6 we compare the estimates provided by these two alternative methods with a simulation study. In section 2 we present the estimation of the RM as a MD problem. Then, in section 3, we analyze the situations which typically produce estimation problems: that is, extreme scores, null categories and ill-conditioned datasets. In section 4 we propose the new method and we analyze it using an ill-conditioned dataset. Finally, with a simulation study, we compare the accuracy of the estimates yielded by the new method with those furnished by the C-JML method. To perform the simulations and other computations we used the Mathematica software package (Wolfram Research, 2003).

2 MD estimation of the RM

2.1 MD method: a brief overview

In a loose sense, a divergence measure quantifies the disparity, or "distance", between two statistical distributions. Divergence measures are defined in a somewhat general way, for example, in Ali and Silvey (1965). For our purpose (estimation), we need to define divergence measures between theoretical and empirical distributions. Consider a parametric family F, of discrete distributions, defined on a set S. We call divergence measure between $F_\xi \in$F, where ξ is a parameter, and the empirical distribution function of the data, F_m, defined on $S_m \subset S$, any function with the form A) or B), respectively given by:

$$\Psi\left(F_\xi, F_m\right) = \sum_{x\in S} \psi\left(\frac{f_m(x)}{f_\xi(x)}\right) f_\xi(x) \tag{1}$$

where ψ is a strictly convex function, or

$$\Phi\left(F_m, F_\xi\right) = \sum_{x\in S_m} \phi\left(\frac{f_\xi(x)}{f_m(x)}\right) f_m(x) \tag{2}$$

where ϕ is strictly convex and decreasing. f_ξ and f_m are respectively the probability function and the empirical probability function corresponding to F_ξ and F_m. Many well known divergence measures belong in this class. For example, from expression (1) we obtain the *Chi-square* divergence, for $\psi(x) = (x-1)^2$ and the *Hellinger* divergence (Simpson, 1987) for $\psi(x) = (\sqrt{x}-1)^2$; on the other hand, from expression (2) we obtain the *Kullback-Leibler* divergence for $\phi(x) = -\ln(x)$ (Kullback and Leibler, 1951). Divergences are used to measure the "distance" between statistical distributions. In particular take F_{ξ_1}, $F_{\xi_2}\in$F : if F_{ξ_1} is "more equal" to F_m than F_{ξ_2}we should have a smaller divergence between F_{ξ_1} and F_m rather than between F_{ξ_2} and F_m. Hence we can use any divergence measure to produce a MD, or minimum distance, estimation (Wolfowitz, 1957). Put simply, MD estimation consists in finding the distribution $F_\xi\in$Fwhich minimizes the divergence with respect to F_m.

2.2 The MD estimation approach in the RM

Estimation in the RM can then also be seen as a MD problem. In the RM, we consider a $n \times k$ dataset, with elements X_{vi}, $v = 1, ..., n$, $i = 1, ..., k$, where X_{vi} are independent Bernoulli random variables representing the answer of n persons to i items. The probability function depends on the parameters $\beta_i (i = 1,\ 2,\ \ldots,\ k)$, denoting the item difficulty, and θ_v $(v = 1,\ 2,\ \ldots,\ n)$, denoting the person ability. It is given by:

$$p\left(x_{vi}\right) = P\left(X_{vi} = x_{vi} \mid \theta_v, \beta_i\right) = \frac{exp\left\{x_{vi}(\theta_v - \beta_i)\right\}}{1 + exp\left\{(\theta_v - \beta_i)\right\}} \tag{3}$$

Note that in our situation we have to deal (for every cell (v, i) of the data matrix) with an empirical distribution function, say F_{mvi}, (let us denote with f_{mvi} the corresponding frequency function) which is necessarily degenerate at the point 1 or 0 (depending on the success x_{vi} =1 or failure x_{vi} =0). For clarity, observe that:

$$f_{mvi}\left(x\right) = \begin{cases} 1 & if \quad x = x_{iv} \\ 0 & if \quad x \neq x_{iv} \end{cases}, \tag{4}$$

where x can be 0 or 1. Hence, for the estimation of the $n+k$ unknown parameters in the RM (minus one, for an identifiability constraint), we consider $n \times k$ divergences (one for each cell), between F_{mvi} and a theoretical distribution function $F_{\theta\beta vi}$ with corresponding probability function $f_{\theta\beta vi}$. For clarity, observe that:

$$f_{\theta,\beta,vi}(x) = \begin{cases} p(x_{vi}) & if \quad x = x_{iv} \\ 1 - p(x_{vi}) & if \quad x \neq x_{iv} \end{cases}, \tag{5}$$

where x can be 0 or 1. In the general case, for any divergence measure of the form A or B, we want to find the minimum of:

$$\sum_{i,v} \Psi(F_{\theta\beta vi}, F_{mvi}) \text{ or } \sum_{i,v} \Phi(F_{mvi}, F_{\theta\beta vi}) \tag{6}$$

To clarify the interpretation of this formula, there follow some special and well known examples of MD problems applied to the RM.

1. *Kullback-Leibler divergence.* The problem of the minimization of the following function (formula (2)):

$$\sum_{i,v} -\ln p(x_{vi}) \ = -l \tag{7}$$

is equivalent to that of the maximization of the function l, that is the log-likelihood function.

2. *Chi-square divergence.* For the generic cell (v,i), the formula (1) yields:

$$\Psi(F_{\theta\beta vi}, F_{mvi}) = \frac{(p(x_{vi}) - 1)^2}{p(x_{vi})} + \frac{(1 - p(x_{vi}))^2}{1 - p(x_{vi})} = \frac{(1 - p(x_{vi}))^2}{(1 - p(x_{vi}))\, p(x_{vi})}. \tag{8}$$

Then, the joint minimization of the $n \times k$ Chi-square divergences leads to the problem of the minimization of the function:

$$\sum_{vi} \left[\frac{(1 - p(x_{vi}))^2}{(1 - p(x_{vi}))\, p(x_{vi})} \right] = \sum_{vi} \left[\exp\left\{ (\theta_v - \beta_i)(-1)^{x_{vi}} \right\} \right] \tag{9}$$

which provides the minimum Chi-square estimates of $\beta_i$and θ_v (see also Linacre 2004). (Incidentally, it may be worth noting that this approach is not related to the so-called MINCHI procedure; Molenaar 1995)

3 Estimation problems with the RM

As outlined above, in a number of cases, estimates of the parameters of the RM may not exist. We will briefly discuss these situations in what follows.

3.1 Extreme scores and null categories

We obtain an extreme score when an individual gives correct responses to each item (perfect score) or to none of them (zero score). Similarly, a null (or "unused") category occurs when every individual gives the same (correct or incorrect) response to an item. If the individual v obtains an extreme score, MD methods cannot find a finite estimate for θ_v, which will then tend to -8 (zero score) or +8 (perfect score). Similarly, if item i obtains an extreme score, the estimate of β_i will tend to -8, or +8. This estimation problem has a logical explanation related to the MD method: a perfect individual score (response pattern (1,1,1,. . . ,1)), for example, suggests finding an ability parameter θ such that the theoretical distribution is the closest one possible to the degenerate distribution at the point 1 (for any item). Obviously, a higher value of θ provides a higher probability of a correct response, so that the estimate diverges. This is not the case, for example, of the pattern (0,1,1,1. . . ,1), because a single zero score is enough to prevent the estimate of θ from diverging. The estimation problem can be solved by erasing the rows (or columns) with extreme scores, or by modifying the total scores, adding (in the case of zero scores) or subtracting (in the case of perfect scores) a fixed number between 0 and 1. For example, the value 0.3 is used, by default, by the Winsteps (Linacre, 2009) computer program.

3.2 Ill-conditioned datasets

As regards the JML estimation, it is well known that the absence of extreme scores and null categories is not a sufficient condition for the existence of the estimates. Indeed, persons with higher ability tend to give correct responses to the easier items, and persons with lower ability tend to give incorrect responses to the more difficult items: if this tendency becomes a "rule" in the dataset, another estimation problem may arise. In particular, a dataset is ill-conditioned (Fischer 1981, Bertoli-Barsotti 2005) if there exists a partition of the respondents into two non-empty subsets such that any individual in the first subset obtains, for any item, a score which is not less than the score of any other individual in the second subset. As a limit case, this situation may obviously be due to the presence of extreme scores, but this is not a necessary condition. In particular, we shall focus on the situation in which, after removing any extreme scores or null categories, the data matrix exhibits the particular structure represented in Figure 1.

$$\begin{bmatrix} 1\backslash 0 & 1 \\ 0 & 1\backslash 0 \end{bmatrix}$$

Figure 1: Non-extreme ill-conditioned matrix structure. See also Fischer (1981).

We will call this special case the "non-extreme" (n.e.) ill-conditioned dataset, which is simply understood to be an ill-conditioned dataset without extreme-scores or null categories. In this case we can identify two sets of individuals and two sets of items. The individuals in the first set (constituted by the persons with higher ability) will give correct responses to every item in the second set (constituted by the less difficult items). At the same time, the individuals in the second set (the persons with lower ability) will give incorrect responses to every item in the first set (the more difficult items). In this situation, no individual in the second set can obtain a score higher, for any item, than that obtained by any individual in the first set. Hence, there is an ability gap between the two sets that cannot be mathematically measured. For this reason all the estimates of θ for the first set of persons will tend to +8 and all the estimates of θ for the second set of persons will tend to -8. Analogously for the two sets of items, we cannot find any finite estimate. Besides, more generally, this estimation problem may also hamper the MD method. Indeed, the divergence is minimized by a theoretical distribution which is the closest to 1 in the second quadrant of the matrix, and to 0 in the third quadrant. Consequently, one should have $\theta\to\infty$ in the first set and $\theta\to-\infty$ in the second; $\beta\to\infty$ in the first set and $\beta\to-\infty$ in the second. Most software fails to recognize that MLEs of the parameters of the RM go to infinity when an n.e. ill-conditioned dataset is considered.

4 The ε-adjustment method

4.1 Stating the procedure

As explained in section 2, estimation problems may arise if the model gets "closer" to the data as the parameter values approach infinity. Our proposal for solving this problem is to "fuzzify" the empirical observations, so as to avoid the risk of a theoretical distribution perfectly matching the empirical one, for infinite parameter values. An intuitive explanation of the method is as follows. Suppose that each item is submitted 100 times to each individual and that, in each case, we obtain 99 correct and 1 incorrect responses, instead of 100% of correct responses; or the opposite case of 99 incorrect and 1 incorrect responses, instead of 100% of incorrect responses. In this situation the empirical distributions, in each cell of the matrix, are not degenerate distributions. In fact, in the case of success the distribution would be 0 with frequency 0.01 and 1 with frequency 0.99, in the case of failure it would be 0 with frequency 0.99 and 1 with frequency 0.01. If we apply the MD method to these modified empirical distributions, and use it to estimate a RM on an n.e. ill-conditioned dataset, we find values of θ and β such that the theoretical distributions are as close as possible to 0.99, in the second quadrant of the matrix (see Figure 1) and to 0.01 in the third quadrant. To reach those values (0.99 or 0.01) there is no need for the parameter estimates to diverge to $+\infty$ or $-\infty$. This can also be applied in the more general case of ill-conditioned datasets. With extreme scores or null categories, the same result holds: the existence of the estimates. The problem of non-existence is very common and involves not only the Rasch model. We now define a new general method to solve the problem of non-existence due to degenerate empirical distributions. Let F_ξ be a discrete distribution defined on a set S. Suppose that the empirical distribution of the data F_m is degenerate at the point $x_0 \in S$, thus we get:

$$f_m(x) = \begin{cases} 1 & if \quad x = x_0 \\ 0 & if \quad x \neq x_0 \end{cases} . \tag{10}$$

In many situation, the MD estimate for the parameter ξ does not exist. The problem can be solved by a "fuzzification" of the empirical distribution that leads to a new empirical distribution which is not degenerate. For a given positive and arbitrarily small number ε (<1), it is possible to "fuzzify" F_m by attributing weight $1-\varepsilon$ (rather than 1) to the point $x = x_0$. As for the other points in S, it is always possible to define an appropriate subdivision of ε into a number of addends which matches the points in S except from x_0, and then apply the MD method. This procedure requires to re-define the empirical distribution, however this assumption can be avoided with the following alternative solution. Suppose our only information is that the relative frequency (or empirical probability) of point x_0 is given by $1-\varepsilon$. More precisely:

$$\begin{cases} f_m^{\varepsilon}(x) = 1-\varepsilon & if \quad x = x_0 \\ \sum_{x \neq x_0} f_m^{\varepsilon}(x) = \varepsilon & if \quad x \neq x_0 \end{cases} . \tag{11}$$

As (11) do not completely identify (except for the case $\#S$=2) the new empirical distribution say F_m^{ε}, the MD method cannot be applied properly. Hence we propose to estimate the unknown parameter by minimizing an *improper* divergence measure between F_ξ and F_m^{ε}, defined as:

$$\Psi^*(F_\xi, F_m^{\varepsilon}) = \psi\left(\frac{1-\varepsilon}{f_\xi(x_0)}\right) f_\xi(x_0) + \psi\left(\frac{\varepsilon}{1-f_\xi(x_0)}\right)(1-f_\xi(x_0)), \tag{12}$$

where ψ is a strictly convex function, or

$$\Phi^*(F_m^{\varepsilon}, F_\xi) = \phi\left(\frac{f_\xi(x_0)}{1-\varepsilon}\right)(1-\varepsilon) + \phi\left(\frac{1-f_\xi(x_0)}{\varepsilon}\right)\varepsilon, \tag{13}$$

where ϕ is a strictly convex and decreasing function.
The improper divergence measures defined above simply attribute weight $1-\varepsilon$ to the "observed" and weight ε to the "unobserved". We call this fuzzification of the empirical distribution function *ε-adjustment*. It is worth noting that if $\#S$=2 (as in the case of the RM) the empirical distribution is completely identified by (11), which means that (11) and (12) define real divergence measures: in other words when $\#S$=2 $\Psi^*(F_\xi, F_m^{\varepsilon}) = \Psi(F_\xi, F_m^{\varepsilon})$ and $\Phi^*(F_m^{\varepsilon}, F_\xi) = \Phi(F_m^{\varepsilon}, F_\xi)$. The ε-adjustment method solves the estimation problems in the RM by approximating each empirical distribution F_{mvi} (see expression (4)) with its ε-adjusted version F_{mvi}^{ε}, (denote with f_{mvi}^{ε} its ε-adjusted empirical probability function) such that:

$$f_{mvi}^{\varepsilon}(x) = \begin{cases} 1-\varepsilon & if \quad x = x_{iv} \\ \varepsilon & if \quad x \neq x_{iv} \end{cases} , \tag{14}$$

where ε is an arbitrarily small positive number. As we will show below, our alternative approach to estimation in the RM is simply based on the MD method applied to the modified empirical distributions shown in expression (11). We can apply the ε-adjustment to any divergence measure. The general formula is:

$$\sum_{i,v} \Psi\left(F^{\varepsilon}_{mvi}, F_{\theta\beta iv}\right) = \sum_{i,v}\left[\psi\left(\frac{1-\varepsilon}{p\left(x_{vi}\right)}\right)p\left(x_{vi}\right) + \psi\left(\frac{\varepsilon}{1-p\left(x_{vi}\right)}\right)\left(1-p\left(x_{vi}\right)\right)\right] \quad (15)$$

for any divergence measure Ψ of the form A, expression (1), and

$$\sum_{i,v} \Phi\left(F^{\varepsilon}_{mvi}, F_{\theta\beta iv}\right) = \sum_{i,v}\left[\phi\left(\frac{p\left(x_{vi}\right)}{1-\varepsilon}\right)(1-\varepsilon) + \phi\left(\frac{1-p\left(x_{vi}\right)}{\varepsilon}\right)\varepsilon\right] \quad (16)$$

for any divergence measure Φ of the form B, expression (2).

Thus, for the special case of the Kullback-Leibler divergence, we obtain the ε-adjusted MLE by minimizing (with respect to the unknown parameters) the function:

$$-l^{\varepsilon} = \sum_{i,v}\left[-\left(1-\varepsilon\right)\ln p\left(x_{vi}\right) - \varepsilon\ln\left(1-p\left(x_{vi}\right)\right)\right]. \quad (17)$$

It is interesting to note that the the ε-adjusted MLE belongs to a more general class of estimators: that is those obtained by maximizing a modified version of the log-likelihood function, say the ε-adjusted log-likelihood function. Specifically:

$$l^{\varepsilon} = l + A_{\varepsilon}, \quad (18)$$

where the function $A_{\varepsilon} = \sum_{i,v} \varepsilon \ln\left[(1-p(x_{vi}))/p(x_{vi})\right]$ is allowed to depend on both the data and the parameters. To be noted is that the above mentioned Firth's formula (Firth, 1993) belongs to the same class of estimators, but clearly leads to a different modified log-likelihood function. Analogously, in regard to the Chi-square divergence, we obtain an ε-adjusted minimum Chi-square estimate by minimizing the function:

$$\sum_{i,v}\left[\frac{\left(p\left(x_{vi}\right)-(1-\varepsilon)\right)^2}{p\left(x_{vi}\right)} + \frac{\left(1-p\left(x_{vi}\right)-\varepsilon\right)^2}{1-p\left(x_{vi}\right)}\right] = \sum_{i,v}\left[\frac{\left(p\left(x_{vi}\right)-(1-\varepsilon)\right)^2}{\left(1-p\left(x_{vi}\right)\right)p\left(x_{vi}\right)}\right] \quad (19)$$

Simulations (only partially reported here) show that, on setting ε conveniently, these methods provide estimates not too far from traditional MLE and minimum Chi-square when traditional estimation is possible, and acceptable estimates when these traditional methods fail to give finite estimates. As regards the RM, Fischer (1981) and Bertoli-Barsotti (2005) show that, under the problematic conditions specified in 3.1 and 3.2, the ML method will always produce infinite estimates. In these same particular situations, empirical evidence shows that the ε-adjusted JML estimates do exist. Indeed, the ε-adjusted JML estimator always exists and is unique, as is proved in what follows.

4.2 Existence and uniqueness of the ε-adjusted JML estimate for the RM

Let $p(x_{vi})$ be defined as in (3). Moreover let, for short, $\theta_v - \beta_i = \sigma_{vi}$ and $D_{vi} = 1 + \exp(\sigma_{vi})$. The ε-adjusted log-likelihood function to be maximized thus becomes

$$l^{\varepsilon}(\sigma) = \sum_{vi} \left[(1-\varepsilon)(\sigma_{vi}x_{vi} - \ln D_{vi}) + \varepsilon \left(\ln\left(1 + \exp\sigma_{vi} - \exp(\sigma_{vi}x_{vi})\right) - \ln D_{vi} \right) \right] =$$

$$= -\sum_{vi} \ln D_{vi} + \sum_{vi} \left[(1-\varepsilon)\sigma_{vi}x_{vi} + \varepsilon \cdot \ln\left(1 + \exp\sigma_{vi} - \exp(\sigma_{vi}x_{vi})\right) \right] = \sum_{vi} \gamma(\sigma_{vi}, x_{vi}). \tag{20}$$

Interestingly enough, a similar formula can be obtained by applying the cited MEL method (Rousseeuw and Christmann, 2003) to the RM, that is:

$$l^{EST}(\sigma) = \sum_{vi} \left[\left((1-x_{vi})\delta_0 + x_{vi}\delta_1\right)(\sigma_{vi} - \ln D_{vi}) + \left(1 - (1-x_{vi})\delta_0 - x_{vi}\delta_1\right)(-\ln D_{vi}) \right]. \tag{21}$$

Nevertheless, it is important to stress that l^{ε} and l^{EST} are differently motivated. Indeed, there are several important differences between MEL and ε-adjustment methods:

1. The MEL approach is focused on binary regression models, while the ε-adjustment method can be applied to a more general class of models as well (see section 4.1).

2. The MEL technique is based on the maximum likelihood estimation, while the ε-adjustment method can be applied to a more general framework of divergence measures.

3. The aim of the MEL approach is to develop a model to explain the occurrence of misclassification errors (remember that for binary data a misclassification error is a transposition $0 \to 1$ or $1 \to 0$), as for example in the case of a medical diagnosis (Rousseeuw and Christmann 2003, p.316). Actually, Rousseeuw and Christmann assume that the true response is unobservable. Instead, in our context, we exclude the occurrence of misclassification, according to standard IRT modelling where it is assumed that the response variable is observed without errors.

4. The MEL approach is motivated by the problem of the bias induced by the presence of misclassifications (outliers), while our problem is, in a sense, the opposite, that is the bias induced by the absence of outliers. Indeed, our estimation problem with the RM originates from the probability of occurrence of degenerate samples when the parameter is "close" to the boundary of the parameter space.

5. The MEL approach introduces in the model two unknown parameters to be estimated, δ_0 and δ_1, related to type I and type II error probabilities. By definition, there are no mathematical constraints between these parameters (except that:

$0 < \delta_0 < \delta_1 < 1$), because they represent probabilities on two *different* probability spaces. On the contrary, $\{\varepsilon, 1-\varepsilon\}$ (where ε, instead, has not to be estimated) is a set of two positive constants that must to sum up to one, to define a proper cumulative distribution function.

All that said, if we substitute $\delta_0 = \varepsilon$ and $\delta_1 = 1-\varepsilon$ we may derive, in particular,

$$l^{EST}(\sigma) = -\sum_{vi} \ln D_{vi} + \sum_{vi} (x_{vi} + \varepsilon - 2\varepsilon x_{vi})\,\sigma_{vi} = \sum_{vi} \lambda(\sigma_{vi}, x_{vi}). \tag{22}$$

Hence, in particular, the functions l^{ε} and l^{EST} become mathematically (even if not conceptually) equal, because

$$\gamma(\sigma_{vi}, 0) = \lambda(\sigma_{vi}, 0) = -\ln D_{vi} + \varepsilon\sigma_{vi}, \tag{23}$$

$$\gamma(\sigma_{vi}, 1) = \lambda(\sigma_{vi}, 1) = -\ln D_{vi} + (1-\varepsilon)\,\sigma_{vi}. \tag{24}$$

Since 0 and 1 are the only possible values for x_{vi}, there is no loss of generality if we consider the problem of the maximization of $l^{EST}(\sigma)$ instead of that of $l^{\varepsilon}(\sigma)$. We thus immediately conclude that the ε-adjusted MLE always exists and is unique by virtue of what is proved in Rousseeuw and Christmann (2003, Property 1, p.320).

4.3 ε-adjusted JML estimate of an ill-conditioned dataset

An important matter of fact is that the method proposed can work while JML (and then the C-JMLE, as far as the item parameter estimates are concerned) fails, which means that, at present, this alternative represents a valuable option. This can be illustrated by the following example. Tables 4 and 5 in Appendix A present the ε-adjusted ML estimates for an n.e. ill-conditioned 30×10 matrix. The dataset was randomly generated by using 10 item difficulties fixed between -4 and 4, (-4,-3.5,-2.5,-1.5,-0.5.0.5,1.5,2.5,3.5,4), and two different sets of 15 ability parameters each, generated from a normal distribution, with standard deviation 1 and means -1.5 and +1.5, respectively. These somewhat strange parametric values are only due to the fact that it is not easy to randomly generate an n.e. ill-conditioned dataset (especially for medium/large test lenghts). The results reported in these tables show that the ε-adjusted ML method (in this case $\varepsilon = 0.04$) yields acceptable estimates, because their values $\hat{\theta}_v^{\varepsilon}$ and $\hat{\beta}_i^{\varepsilon}$ are not too far from the real ones, although the size of the dataset is quite small. In the last two columns of the tables we report - for exemplificative purposes only - the estimates produced by RUMM2020 (Andrich et al., 2003) and Winsteps (Linacre, 2009) computer programs. Because these estimates are quite distant from the true values, they cannot be considered as acceptable, as their values strongly depend on the number of iterations used by the programs. A larger number of iterations would make the estimates grow and grow to infinity: these softwares seem not to recognize the estimation problem and to treat this matrix as a non-problematic one.

4.4 On the possible influence of extreme scores

It is well known that extreme row scores have no influence on the MLE of item parameters. In fact a zero, or perfect, score does not furnish any information about the item difficulties (similarly, items with unused category have no influence on the MLE of the person parameters). However, we may conjecture that there is a little difference if the extreme scores are all zero or all perfect scores. Is it possible to obtain some information from the ratio between the number of zero scores and the number of perfect scores? The following example shows that the ε-adjusted MLE follows this line of reasoning. Consider the 10×6 dataset (randomly generated) shown in Table 1.

Table 1: Data matrix without extreme scores

1	0	0	0	1	0
1	1	1	1	0	1
1	0	0	0	0	0
1	0	1	0	0	0
1	1	0	0	1	0
1	1	0	0	0	0
1	1	0	0	0	0
1	0	0	0	1	0
1	1	1	0	0	0
0	1	1	1	1	0

If we add 6 rows with extreme scores to this matrix we have 7 different possibilities: 6 zero scores *(6z)*, 5 zero scores and 1 perfect score *(5z,1p)*, 4 zero and 2 perfect scores *(4z,2p)*, 3 zero and 3 perfect scores *(3z,3p)*, 2 zero and 4 perfect scores *(2z,4p)*, 1 zero and 5 perfect scores *(1z,5p)* and 6 perfect scores *(6p)*. Table 2 shows how the ε-adjusted MLE (ε=0.01) of the β parameters changes in these different situations. As can be seen, if the number of perfect scores increases (with respect to the number of zero scores), the difficulties of the first and the last items (β_1 and β_6) decrease. On the other hand, the difficulties of the remaining items $\beta_2,\ldots,\beta_5$ slightly decrease. In this example, the differences among the results are anyway quite small, and the ratio between extreme and non-extreme rows is 6/10: if we had a smaller ratio we could have much smaller differences.

Hence, the ε-adjusted MLE may somehow be influenced by extreme scores: this may be a good property because the changes in the estimates could correspond to a deeper analysis of the dataset with respect to the traditional MLE. However, this particular aspect should be elaborated further, with other examples and simulation studies.

Table 2: Dependence of extreme row scores on ε-adjusted MLEs $\hat{\beta}_i^{\varepsilon}$ of item parameters. Results for zero scores that vary from 6 to 0 and perfect scores that vary in a complementary way, from 0 to 6.

	(6z)	*(5z,1p)*	*(4z,2p)*	*(3z,3p)*	*(2z,4p)*	*(1z,5p)*	*(6p)*
$\hat{\beta}_1^{\varepsilon}$	-2.6930	-2.7136	-2.7353	-2.7581	-2.7822	-2.8076	-2.8344
$\hat{\beta}_2^{\varepsilon}$	-0.9378	-0.9292	-0.9205	-0.9117	-0.9027	-0.8935	-0.8841
$\hat{\beta}_3^{\varepsilon}$	0.0509	0.0616	0.0722	0.0829	0.0937	0.1045	0.1155
$\hat{\beta}_4^{\varepsilon}$	1.2897	1.2969	1.3040	1.3109	1.3177	1.3245	1.3313
$\hat{\beta}_5^{\varepsilon}$	0.0509	0.0616	0.0722	0.0829	0.0937	0.1045	0.1155
$\hat{\beta}_6^{\varepsilon}$	2.2393	2.2227	2.2074	2.1931	2.1799	2.1676	2.1563

4.5 Comparisons

The ε-adjusted ML estimates cannot be directly compared with the C-JML estimates in the special case of n.e. ill-conditioned datasets because in this case the JML fails to provide finite estimates. Moreover, at least in the case of extreme scores, most softwares use an ad-hoc correction method (as explained in section 3.1), to provide finite estimates. In the sequel we focus on the estimation of the item parameters, considering the individual abilities as nuisance parameters. We will compare ε-adjusted MLE with MLE and C-JMLE with a simulation study, focusing on the minimization of the estimation error for the item parameters. Since the C-JMLE method is well known and commonly used, it can be considered to be the "standard correction" method – commonly implemented in the JML software. Nevertheless, as pointed out in Jansen et al. (1988), this bias correction is not fully satisfactory, since the bias may depend on the skewness of the item difficulty distribution. Also, the C-JMLE method cannot remove the bias when the number of items k is particularly small (k<10). Moreover, as proved in Bertoli-Barsotti and Punzo (2012), the C-JMLE generally works better than Firth's correction for the RM. Hence, in our comparison analysis, we decided to consider only this "standard" method, as a benchmark.

4.6 Simulations

In our simulation study, different combinations of values for k and n were considered, that is: k=5, 10, 20, 30, and n=200, 300, 500, 1000. One hundred datasets were randomly generated for each combination of the values of k and n. For each dataset, the θ parameters were randomly generated from a normal distribution with mean zero and standard deviation equal to 0.5 and 0.7. We report our simulation results in Tables 6 and 7 in Appendix B. The values in each cell are the averages of 100 empirical mean squared error (MSE) obtained with the different estimation methods considered. In bold are the smallest values across the methods considered. In the first set of simulations, the β parameters were fixed and equally spaced in the range considered. In the second set of simulations, the β parameters were re-centered at zero, after being randomly generated from uniform distributions with different ranges, that is, U(-2,2), U(-3,3). Thus each dataset was generated by different values of parameters (θ and β), in order to consider a large class of different situations.

As said, the estimation methods considered were 1) MLE, 2) C-JMLE, and 3) ε-adjusted MLE. The latter was considered for different values of ε (at least the ones which seemed the most suitable – for this first analysis). As a general rule, the larger n is, the higher the probability of obtaining extreme column scores. But this is not a problem, because we are concerned only with the estimates of the β parameters, which are not influenced by extreme column scores. Besides, there is no risk of obtaining extreme row scores with the combination of parameters (θ and β) that we used, so the that MLE estimates of the β parameters always exist. We summarize the results of our two sets of simulations as follows.

1. β fixed and equally spaced.

   It appears that the ε-adjusted ML yields the best results for suitable choices of ε: the larger k is, the smaller ε should be. If k increases, the difference among the three methods is less evident: for high values of k, the correction methods may even become superfluous because the risk of having extreme scores or ill conditioned matrices is close to zero.

2. β randomly generated from uniform distribution.

   This second simulation study only confirms what we found in the first one. The sole difference is that the MSEs are a little larger (on average) than before. This is due to the fact that equally spaced parameters are easier to estimate than ones randomly generated from a uniform distribution.

5 Conclusion

Our preliminary empirical results show that the ε-adjusted MLE method provides acceptable estimates in both troublesome situations: extreme scores and n.e. ill-conditioned datasets. In the latter case, this may be the only viable estimation method, since the other well-known and widely-used methods (for the RM) are known to fail. Hopefully, this empirical and preliminary finding will be extended to cover more general cases in our future studies. Moreover, although it was not conceived for the purpose, the ε-adjusted MLE method appears also to be a good solution for correcting the bias of the MLE, as shown in the simulations, especially when the size of the dataset justifies the use of the method (smaller datasets have a higher risk of n.e. ill-conditioned or extreme scores). On the other hand, the "optimal" values for ε can only be found empirically. This is certainly a drawback, and further simulations could furnish us more accurate information. Nevertheless, the simulations performed thus far show that there is a relation between the "ideal" choice of ε and the number of items k. This can be summarized in Table 3, which provides a useful guidance for application of the ε-adjustment method. Although we cannot say that the choice of these values for ε is absolutely the best, we surely are certainly confident that the ε values considered are enough to improve not only the traditional ML estimates but also the "corrected" ones (C-JMLE). As a rule of thumb, we found that a good choice of ε could be ε=0.3/k. Finally, the question of the influence of extreme scores on the estimates (studied in Section 4.4) should be investigated further with simulations, given that it could be another good property for this method.

Table 3: "Ideal" choice of ε for different test lengths k

k	ε
5	0.05
10	0.03
20	0.02
30	0.01

Appendix A: An instance of an n.e. ill-conditioned dataset

Table 4: Comparisons between ε-adjusted MLEs $\hat{\theta}_v^{\varepsilon}$ ($\varepsilon = 0.04$) of person parameters and the corresponding estimates given by two estimation programs RUMM2020 (*Rumm*) and Winsteps (*Win*), with default settings (both these softwares have convergence problems)

										Total	θ_v *true*	$\hat{\theta}_v^{\varepsilon}$	*MLE*	*Rumm*	*Win*
1	1	0	0	0	0	0	0	0	0	2	-2.73	-2.79	*does not exist*	-5.63	-5.99
1	1	0	0	0	0	0	0	0	0	2	-2.47	-2.79	*does not exist*	-5.63	-5.99
0	1	1	0	0	0	0	0	0	0	2	-2.47	-2.79	*does not exist*	-5.63	-5.99
1	0	1	0	0	0	0	0	0	0	2	-2.19	-2.79	*does not exist*	-5.63	-5.99
1	1	0	0	0	0	0	0	0	0	2	-1.85	-2.79	*does not exist*	-5.63	-5.99
0	1	1	0	0	0	0	0	0	0	2	-1.36	-2.79	*does not exist*	-5.63	-5.99
1	1	1	0	0	0	0	0	0	0	3	-2.03	-1.90	*does not exist*	-2.39	-3.94
1	1	1	1	0	0	0	0	0	0	4	-1.72	-1.00	*does not exist*	-0.38	-1.74
1	1	1	1	1	0	0	0	0	0	5	-1.59	-0.07	*does not exist*	1.33	0.24
1	1	1	1	1	0	0	0	0	0	5	-1.27	-0.07	*does not exist*	1.33	0.24
1	1	1	1	1	0	0	0	0	0	5	-1.07	-0.07	*does not exist*	1.33	0.24
1	1	1	0	1	0	1	0	0	0	5	-0.78	-0.07	*does not exist*	1.33	0.24
1	1	1	1	1	0	0	0	0	0	5	-0.70	-0.07	*does not exist*	1.33	0.24
1	1	1	1	1	0	0	0	0	0	5	-0.13	-0.07	*does not exist*	1.33	0.24
1	1	1	1	1	0	0	0	0	0	5	0.43	-0.07	*does not exist*	1.33	0.24
1	1	1	1	1	0	0	0	0	0	5	0.50	-0.07	*does not exist*	1.33	0.24
1	1	1	1	1	1	0	0	0	0	6	1.29	0.87	*does not exist*	2.70	2.13
1	1	1	1	1	1	0	0	0	0	6	1.31	0.87	*does not exist*	2.70	2.13
1	1	1	1	1	0	1	0	0	0	6	1.48	0.87	*does not exist*	2.70	2.13
1	1	1	1	1	1	0	0	0	0	6	1.98	0.87	*does not exist*	2.70	2.13
1	1	1	0	1	1	1	0	0	0	6	2.23	0.87	*does not exist*	2.70	2.13
1	1	1	1	1	1	1	0	0	0	7	0.67	1.81	*does not exist*	3.90	3.80
1	1	1	1	1	1	1	0	0	0	7	0.76	1.81	*does not exist*	3.90	3.80
1	1	1	1	1	1	1	0	0	0	7	1.97	1.81	*does not exist*	3.90	3.80
1	1	1	1	1	1	1	0	0	0	7	2.39	1.81	*does not exist*	3.90	3.80
1	1	1	1	1	0	1	1	1	0	8	0.67	2.79	*does not exist*	5.32	5.65
1	1	1	1	1	1	0	1	0	1	8	1.05	2.79	*does not exist*	5.32	5.65
1	1	1	1	1	1	1	1	0	0	8	1.69	2.79	*does not exist*	5.32	5.65
1	1	1	1	1	1	1	1	0	0	8	2.63	2.79	*does not exist*	5.32	5.65
1	1	1	1	1	1	1	1	0	0	8	3.54	2.79	*does not exist*	5.32	5.65

Table 5: Comparisons between ε-adjusted MLEs of item parameters $\hat{\beta}_i^{\varepsilon}$ ($\varepsilon = 0.04$) and the corresponding estimates given by two estimation programs RUMM2020 (*Rumm*) and Winsteps (*Win*), with default settings (both these softwares have convergence problems).

Item	*Total*	β_i*true*	$\hat{\beta}_i^{\varepsilon}$	*MLE*	*C-JMLE*	*Rumm*	*Win*
1	28	-4	-3.26	*does not exist*	*does not exist*	6.56	7.22
2	29	-3.5	-3.75	*does not exist*	*does not exist*	6.57	7.22
3	27	-2.5	-2.85	*does not exist*	*does not exist*	4.49	4.77
4	21	-1.5	-1.04	*does not exist*	*does not exist*	2.73	2.04
5	22	-0.5	-1.30	*does not exist*	*does not exist*	2.81	2.40
6	12	0.5	0.89	*does not exist*	*does not exist*	-1.12	-2.05
7	11	1.5	1.09	*does not exist*	*does not exist*	0.42	-1.35
8	5	2.5	2.46	*does not exist*	*does not exist*	-6.98	-6.02
9	1	3.5	3.88	*does not exist*	*does not exist*	-8.08	-7.57
10	1	4	3.88	*does not exist*	*does not exist*	-7.39	-6.67

Appendix B: Simulation results

Table 6: Simulation comparing ε-adjusted MLE (for different values of ε), MLE, and C-JMLE for the item parameters. True β parameters fixed and equally spaced.

k	n	dist β	sd(θ)	MLE (β)	C-JMLE (β)	$\hat{\beta}^{0.05}$	$\hat{\beta}^{0.08}$
5	200	eq(-2,2)	0.5	0.3168	0.0527	**0.0269**	0.0480
5	200	eq(-3,3)	0.5	1.3291	0.3032	**0.0486**	0.2168
5	500	eq(-2,2)	0.5	0.2613	0.0272	**0.0124**	0.0349
5	500	eq(-3,3)	0.5	0.9500	0.1459	**0.0364**	0.2131
5	1000	eq(-2,2)	0.5	0.2349	0.0162	**0.0071**	0.0319
5	1000	eq(-3,3)	0.5	0.9360	0.1353	**0.0262**	0.2083

k	n	dist β	sd(θ)	MLE (β)	C-JMLE (β)	$\hat{\beta}^{0.01}$	$\hat{\beta}^{0.02}$	$\hat{\beta}^{0.03}$
10	200	eq(-2,2)	0.7	0.0700	0.0453	0.0466	0.0327	**0.0263**
10	200	eq(-3,3)	0.7	0.1766	0.1062	0.0829	0.0420	**0.0352**
10	500	eq(-2,2)	0.7	0.0467	0.0247	0.0268	0.0155	**0.0112**
10	500	eq(-3,3)	0.7	0.1268	0.0653	0.0504	0.0196	**0.0192**
10	1000	eq(-2,2)	0.7	0.0373	0.0169	0.0190	0.0091	**0.0059**
10	1000	eq(-3,3)	0.7	0.1033	0.0472	0.0364	**0.0112**	0.0144

k	n	dist β	sd(θ)	MLE(β)	C-JMLE(β)	$\hat{\beta}^{0.01}$	$\hat{\beta}^{0.02}$	$\hat{\beta}^{0.03}$
20	200	eq(-2,2)	0.7	0.0396	0.0301	0.0311	**0.0276**	0.0284
20	200	eq(-3,3)	0.7	0.0654	0.0436	0.0397	**0.0375**	0.0518
20	500	eq(-2,2)	0.7	0.0190	0.0122	0.0129	**0.0115**	0.0141
20	500	eq(-3,3)	0.7	0.0373	0.0189	**0.0176**	0.0191	0.0360
20	1000	eq(-2,2)	0.7	0.0121	0.0060	0.0067	**0.0059**	0.0089
20	1000	eq(-3,3)	0.7	0.0272	0.0097	**0.0092**	0.0119	0.0294

k	n	dist β	sd(θ)	MLE(β)	C-JMLE(β)	$\hat{\beta}^{0.01}$
30	300	eq(-2,2)	0.7	0.0233	0.0192	**0.0187**
30	300	eq(-3,3)	0.7	0.0422	0.0312	**0.0266**
30	500	eq(-2,2)	0.7	0.0158	0.0121	**0.0118**
30	500	eq(-3,3)	0.7	0.0256	0.0174	**0.0155**
30	1000	eq(-2,2)	0.7	0.0085	0.0057	**0.0056**
30	1000	eq(-3,3)	0.7	0.0157	0.0084	**0.0077**

Table 7: Simulation comparing ε-adjusted MLE (for different values of ε), MLE, and C-JMLE for the item parameters. True β parameters randomly generated by drawing from a uniform distribution.

k	n	dist β	sd(θ)	MLE(β)	C-JMLE(β)	$\hat{\beta}^{0.05}$	$\hat{\beta}^{0.08}$
5	200	U(-2,2)	0.5	0.1472	0.0304	**0.0250**	0.0257
5	200	U(-3,3)	0.5	0.8512	0.2229	**0.0319**	0.1001
5	500	U(-2,2)	0.5	0.1448	0.0172	**0.0107**	0.0160
5	500	U(-3,3)	0.5	0.4748	0.0761	**0.0138**	0.0660
5	1000	U(-2,2)	0.5	0.1144	0.0093	**0.0069**	0.0132
5	1000	U(-3,3)	0.5	0.4159	0.0550	**0.0109**	0.0680

k	n	dist β	sd(θ)	MLE(β)	C-JMLE(β)	$\hat{\beta}^{0.01}$	$\hat{\beta}^{0.02}$	$\hat{\beta}^{0.03}$
10	200	U(-2,2)	0.7	0.0538	0.0368	0.0388	0.0296	**0.0251**
10	200	U(-3,3)	0.7	0.1282	0.0791	0.0656	0.0377	**0.0316**
10	500	U(-2,2)	0.7	0.0364	0.0205	0.0224	0.0143	**0.0110**
10	500	U(-3,3)	0.7	0.0873	0.0451	0.0384	0.0169	**0.0139**
10	1000	U(-2,2)	0.7	0.0267	0.0124	0.0145	0.0075	**0.0049**
10	1000	U(-3,3)	0.7	0.0744	0.0349	0.0289	**0.0110**	0.0112

k	n	dist β	sd(θ)	MLE(β)	C-JMLE(β)	$\hat{\beta}^{0.01}$	$\hat{\beta}^{0.02}$	$\hat{\beta}^{0.03}$
20	200	U(-2,2)	0.7	0.0375	0.0284	0.0296	0.0259	**0.0259**
20	200	U(-3,3)	0.7	0.0652	0.0441	0.0407	**0.0365**	0.0464
20	500	U(-2,2)	0.7	0.0165	0.0112	0.0118	**0.0109**	0.0131
20	500	U(-3,3)	0.7	0.0334	0.0172	0.0164	**0.0163**	0.0286
20	1000	U(-2,2)	0.7	0.0108	0.0055	0.0062	**0.0055**	0.0080
20	1000	U(-3,3)	0.7	0.0224	0.0087	**0.0086**	0.0104	0.0237

k	n	dist β	sd(θ)	MLE(β)	C-JMLE(β)	$\hat{\beta}^{0.01}$
30	300	U(-2,2)	0.7	0.0233	0.0192	**0.0188**
30	300	U(-3,3)	0.7	0.0370	0.0285	**0.0251**
30	500	U(-2,2)	0.7	0.0140	0.0110	**0.0108**
30	500	U(-3,3)	0.7	0.0264	0.0179	**0.0156**
30	1000	U(-2,2)	0.7	0.0079	0.00556	**0.0054**
30	1000	U(-3,3)	0.7	0.0156	0.00896	**0.0080**

Aknowledgement

This paper has been elaborated in the framework of the project Opportunity for young researchers, reg. no. CZ.1.07/2.3.00/30.0016, supported by Operational Programme Education for Competitiveness and co-financed by the European Social Fund and the state budget of the Czech Republic.

References

Agresti, A. (2007). *An introduction to categorical data analysis.* Wiley, NY.

Albert, A. and Anderson, J. (1984). On the existence of maximum likelihood estimates in logistic regression models. *Biometrika*, 71:1–10.

Ali, S. M. and Silvey, S., D. (1965). A general class of coefficients of divergence of one distribution from another. *Journal of the Royal Statistical Society, Series B*, 28:131–142.

Andersen, E. (1980). *Discrete statistical models with social sciences applications.* North-Holland, Amsterdam.

Andrich, D., Lyne, A., Sheridan, B., and G., L. (2003). *RUMM2020 (computer program).* RUMM Laboratory, Perth, Australia.

Bertoli-Barsotti, L. (2005). On the existence and uniqueness of jml estimates for the partial credit model. *Psychometrika*, 70:517–531.

Bertoli-Barsotti, L. and Punzo, A. (2012). Comparison of two bias reduction techniques for the rasch model. *Electronic Journal of Applied Statistical Analysis*, 5:360–366.

Firth, D. (1993). Bias reduction of maximum likelihood estimates. *Biometrika*, 80:27–38.

Fischer, G., H. (1981). On the existence and uniqueness of maximum likelihood estimates in the rasch model. *Psychometrika*, 46:59–77.

Jansen, P., G. W., van den Wollenberg, A., L., and Wierda, F., W. (1988). Correcting unconditional parameter estimates in the rasch model for inconsistency. *Applied Psychological Measurement*, 12:297–306.

Kosmidis, I. (2007). *Bias Reduction in Exponential Family Nonlinear Models*. PhD thesis, University of Warwick, http://www.ucl.ac.uk/ ucakiko/.

Kullback, S. and Leibler, R., A. (1951). On information and sufficiency. *The Annals of Mathematical Statistics*, 22:79–86.

Linacre, J. M. (2004). Rasch model estimation: further topics. *Journal of Applied Measurement*, 5:95–110.

Linacre, J., M. (2009). *WINSTEPS Rasch measurement computer program.* Beaverton, Oregon: Winsteps.com.

Lucadamo, A. (2010). Rasch analysis and multilevel models for the evaluation of the customer satisfaction. *Electronic Journal Of Applied Statistical Analysis*, 3(1):44–51.

Molenaar, I. (1995). *Rasch Models: Foundations, Recent Developments, and Applications*, chapter Estimation of item parameters, in , pp. 39-51., pages 39–51. New York: Springer-Verlag.

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen: Danish Institute for Educational Research. Reprinted, 1993, Chicago: MESA Press.

Rousseeuw, P., J. and Christmann, A. (2003). Robustness against separation and outliers in logistic regression. *Computational Statistics & Data Analysis*, 43:315–332.

Simpson, D. (1987). Minimum hellinger estimation for the analysis of count data. *Journal of the American Statistical Association*, 82:802–807.

Wolfowitz, J. (1957). The minimum distance method. *The Annals of Mathematical Statistics*, 28:75–88.

Wolfram Research, I. (2003). *Mathematica, Version 5.0.* Champaign, IL.

Wright, B., D. (1977). Best procedures for sample-free item analysis. *Applied Psychological Measurement*, 1:281–295.

Wright, B., D. (1988). The efficacy of unconditional maximum likelihood bias correction: comment on jansen, van den wollenberg and wierda. *Applied Psychological Measurement*, 12:315–316.

Wright, B., D. and Douglas, G., A. (1977). Solving measurement problems with the rasch model. *Journal of Educational Measurement*, 14:97–116.