Communications in Statistics—Theory and Methods, 35: 921–935, 2006

ISSN: 0361-0926 print/1532-415X online
DOI: 10.1080/03610920500501445

Quality of Life Analysis

Goodness of Fit Test and Latent Distribution Estimation in the Mixed Rasch Model

G. MARTYNOV[1] AND M. MESBAH[2]

[1]Institute for Information Transmission Problems, Russian Academy of Science, Moscow, Russia
[2]Department of Statistics, University Pierre et Marie Curie, Paris, France

The Cramér-von Mises test methodology is applied to build a goodness-of fit test for the mixed Rasch model. The Mixed Rasch Model is a probability model of a multivariate discrete random variable driven by an unknown latent continuous variable. The problem of estimation of the unknown fixed difficulty parameters and the latent density function is also considered. The theoretical results are illustrated through simulations and an application to real Quality of Life data.

Keywords Cramer-von Mises test; Latent distribution; Nonparametric estimation; Rasch model.

Mathematics Subject Classification 62P15; 62G10.

1. Introduction

The Rasch model was first developed by the Danish mathematician Rasch (1960). It is also sometimes called the one parameter logistic model because of its formulation

$$P(X_{vj} = x_{vj} \mid \theta_v) = \frac{e^{x_{vj}(\theta_v - \beta_j)}}{1 + e^{\theta_v - \beta_j}},$$

where v and j are respective subject and item parameter, and β_j and θ_v are scalars, $-\infty < \beta_j < \infty$, $-\infty < \theta_v < \infty$. The item parameter X_{vj} take two possible values, 1 (success) and 0 (failure). The probability of success, when viewed as a function of θ, is usually called the item characteristic curve. It is $1/2$, when $\theta_v = \beta_j$, and it is decreasing for fixed θ_v. So, a large values of β_j corresponds to a low probability of a correct response. The parameter β_j is often called the difficulty parameter of item j and θ_v, the ability or person parameter. The central portions of these curves

Received March 12, 2004; Accepted October 19, 2005
Address correspondence to M. Mesbah, Department of Statistics, University Pierre et Marie Curie, 175 rue de Chevaleret, 73013 Paris, France; E-mail: mesbah@ccr.jussieu.fr

are nearly straight lines. Another interesting property is that the different curves corresponding to different items are not crossing.

The Rasch model distinguishes itself from other more general models for items (so-called Item Response Theory Models (IRT), with item characteristic curve of any more general form) by the following property: an individual's total score (i.e., the number of correct answers) is a sufficient statistic for the ability parameter. This is a characteristic property for the Rasch model, and it is the only IRT model with this property. Hence, when we use the total score as the measure produced by the instrument, and when responses are dichotomous, we must assume that the Rasch model is true.

Another interesting measurement property of the Rasch model is its specific objectivity. This property states that the comparison of any two persons v and w does not involve the instrument parameter. Again, this is true only for the Rasch model and not for any other link function (Fischer and Molenaar, 1995).

The use and interpretation of the word "difficulty" and "ability" for the two parameters of the model is coming from the context of educational sciences, where the Rasch model was first developed. Applications to more general fields, mainly in social sciences, and more recently in medicine (psychiatry, quality of life) are increasing (Mesbah et al., 2002).

In Sec. 2, we will consider the mixed Rasch model, which is a special case of the original Rasch model with an ability parameter assumed random. It is proposed the representation of the model by the probabilities of the answers combination.

In Sec. 3, various representations of the item probability function are derived. These representations are used in Sec. 4 to develop a statistical test, analogous to the Cramér-von Mises test of goodness-of-fit, and in Sec. 5 to produce an estimate of the difficulty parameters and the a priori distribution of the unobserved density function of the ability parameter.

A short introduction to the Cramér-von Mises test theory is given in Sec. 3. Adaptation of this theory to the discrete Rasch model is done in Sec. 4. A new method for estimation of difficulty parameters β and the latent density function is described in Sec. 5. In Sec. 6, a numerical example is presented.

There is a growing literature on Rasch model. An extended discussion could be found in Fischer and Molenaar (1995). The importance of application of goodness-of-fit testing in the context of validation of questionnaire of quality of life is explained in Hamon et al. (2002).

2. Continuous Approximations for the Discrete Probabilities

The mixed Rasch model is a special case of the original Rasch model with an ability parameter assumed random. We will consider the random vector $X = (X_1, X_2, \dots, X_k)'$ with k components, containing the answers of an individual to the questions in a questionnaire, where each component can have only two values, 0 or 1. From now, we will consider the ability parameter θ_v as random and we will note its value z. It varies from one individual to another one. The answers of the individuals to the questionnaire $X = (X_1, X_2, \dots, X_k)'$ depends only on their z value which means that they are independent conditionally to this z value, which is unobserved.

In the Rasch model, the probabilities

$$p(x_1, x_2, \dots, x_k \mid z) = P(X_1 = x_1, X_2 = x_2, \dots, X_k = x_k \mid z)$$

are the conditional probabilities of the form

$$p(x_1, x_2, \ldots, x_k \mid z) = \prod_{j=1}^{k} \frac{e^{x_j(z-\beta_j)}}{1 + e^{(z-\beta_j)}},$$

where $\beta = (\beta_1, \beta_2, \ldots, \beta_k)'$ is the parameters vector. Then the probabilities

$$p(x_1, x_2, \ldots, x_k) = P(X_1 = x_1, X_2 = x_2, \ldots, X_k = x_k),$$

can be represented by the mixture

$$p(x_1, x_2, \ldots, x_k) = \int_{-\infty}^{\infty} \prod_{j=1}^{k} \frac{e^{x_j(z-\beta_j)}}{1 + e^{(z-\beta_j)}} g(z)dz,$$

where $g(z)$ is a density function, describing the r.v. z. The values of the vector $(X_1, X_2, \ldots, X_k)'$ belong to the set of 2^k binary numbers $B_0, B_1, \ldots, B_{2^k-1}$, where

$$B_0 = [00\ldots00], \quad B_1 = [00\ldots01],$$
$$B_2 = [00\ldots10], \ldots, B_i = [x_1x_2, \ldots, x_k], \ldots, B_{2^k-1} = [11\ldots11],$$

where for B_i the notation $[x_1x_2\ldots x_k]$ is the binary representation of the index i, $i \equiv N(x_1, x_2, \ldots, x_k)$, i.e.,

$$i = N(x_1, x_2, \ldots, x_k) = x_k + x_{k-1}2 + x_{k-2}2^2 + \cdots + x_1 2^{k-1}, \quad i = 0, \ldots, 2^k - 1.$$

We will introduce now the probabilities $p_i = p(x_1, x_2, \ldots, x_k), i = 0, \ldots, 2^k - 1$, defined as

$$p_i = P(X_1 = x_1, X_2 = x_2, \ldots, X_k = x_k).$$

Analogously, we can introduce the conditional probabilities $p_{i,z} = p(x_1, x_2, \ldots, x_k \mid z)$

$$p_{i,z} = P(X_1 = x_1, X_2 = x_2, \ldots, X_k = x_k \mid z).$$

If k is large, we need to develop an effective method for computation of each $p(x_1, x_2, \ldots, x_k)$. We will show below, how to reduce the number of integrals needed to get all $p(x_1, x_2, \ldots, x_k)$ from 2^k to $k+1$.

We can see that the probabilities p_i can be represented as follows:

$$\begin{aligned} p(x_1, x_2, \ldots, x_k) &= \int_{-\infty}^{\infty} \prod_{j=1}^{k} \frac{e^{x_j(z-\beta_j)}}{1 + e^{(z-\beta_j)}} g(z)dz \\ &= e^{-\sum_{j=1}^{k} x_j\beta_j} \int_{-\infty}^{\infty} \left\{ \prod_{j=1}^{k} \frac{1}{1 + e^{(z-\beta_j)}} \right\} e^{\sum_{j=1}^{k} x_j z} g(z)dz \\ &= e^{-\sum_{j=1}^{k} x_j\beta_j} \int_{-\infty}^{\infty} \frac{1}{\prod_{j=1}^{k}(1 + \alpha_j e^z)} e^{Mz} g(z)dz, \end{aligned}$$

where $M = \sum_{j=1}^{k} x_j$ and $\alpha_j = e^{-\beta_j}$. M takes only $k+1$ values $0, 1, \ldots, k$. Hence, the probabilities p_i can be written using the $k+1$ following integrals

$$S(M, \beta) = \int_{-\infty}^{\infty} \frac{1}{\prod_{j=1}^{k}(1+\alpha_j e^z)} e^{Mz} g(z) dz, \quad M = 0, 1, \ldots, k, \tag{1}$$

depending only from values M and β. The computation of all probabilities $p(x_1, x_2, \ldots, x_k)$ for given β can be provided now effectively with the formula

$$p_i = p(x_1, x_2, \ldots, x_k) = e^{-\sum_{i=1}^{k} x_i \beta_i} S(M, \beta), \quad i = 0, 1, \ldots, 2^k - 1. \tag{2}$$

3. Some Results from the Theory of the Cramér-Von Mises Test

Now, we will present some definitions and properties of the Cramér-von Mises test theory. Let $X^n = \{X_1, X_2, \ldots, X_n\}$ be a sample from the distribution function $F(x)$, $x \in R^1$. Let us consider the simple hypothesis

$$H_0 : F(x) = G(x),$$

where $G(x)$ is a continuous distribution function, and the complex hypothesis

$$H_0' : F(x) \in \mathcal{G} = \{G(x, \theta), \theta = (\theta_1, \theta_2, \ldots, \theta_k)^\top \in \Theta \subset R_k\},$$

where θ is an unknown parameter. For example, the family $\mathcal{G}$ can be the normal distribution family

$$\mathcal{N} = \left\{\Phi\left(\frac{x-m}{\sigma}\right), -\infty < m < \infty, \sigma > 0\right\},$$

where

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{y^2}{2}} dy$$

with unknown parameters m and σ.

The Cramér-von Mises test for simple hypotheses is based on the statistics

$$\omega_n^2 = n \int_{-\infty}^{\infty} (F_n(x) - G(x))^2 dG(x), \tag{3}$$

where

$$F_n(x) = \frac{1}{n} \sum_{i=1}^{n} \delta_x(X_i), \quad \delta_x(X_i) = \begin{cases} 0, & x \le X_i, \\ 1, & x > X_i. \end{cases}$$

After the transformation $T_i = G(X_i)$, $i = 1, 2, \ldots, n$, we get the standard form of the statistics:

$$\omega_n^2 = n \int_0^1 (F_n^*(t) - t)^2 \, dt = \int_0^1 \xi_n^2 \, dt,$$

where $F_n^*(t)$ is the empirical distribution function based on $T_i, i = 1, 2, \dots, n,$ and $\xi_n(t) = n^{1/2}(F_n^*(t) - t)$, is the unweighted empirical process.

Limit distribution of ω_n^2 under H_0 is the distribution of the functional

$$\omega^2 = \int_0^1 \xi^2(t)dt$$

of the Gauss process $\xi(t)$ with $E\xi(t) = 0$ and covariance function

$$K_0(t, \tau) = \min(t, \tau) - t\tau, \quad t, \tau \in (0, 1),$$

is called Brownion bridge $\xi(t)$.

The hypothesis H_0 is accepted, if $\omega_n^2 < t_{n,\alpha}$ and rejected, if $\omega_n^2 \geq t_{n,\alpha}$, where $P(\omega_n^2 \geq t_{n,\alpha}) = \alpha$. Approximately, it can be chosen $t_{n,\alpha} \approx t_\alpha$, and $P(\omega^2 \geq t_\alpha) \approx \alpha$.

In the case of parametrical hypotheses the Cramér-von Mises statistics is

$$\omega_n^2 = n \int_{-\infty}^{\infty} (F_n(x) - G(x, \theta_n))^2 \, dG(x, \theta_n), \tag{4}$$

where θ_n is an estimator of θ (see Kac et al., 1955).

Let θ_n be the maximum likelihood estimator of θ. Under some regularity conditions (Neuhaus, 1976) and under H_0', the limit distribution of the statistics ω_n^2 is the same as the distribution of the functional

$$\omega^2 = \int_0^1 \xi^2(t, \theta_0)dt$$

of the Gauss process $\xi(t, \theta_0)$ with $E\xi(t, \theta_0) = 0$ and with covariance function

$$K(t, \tau, \theta_0) = E(\xi(t, \theta_0)\xi(\tau, \theta_0)) = (K_0(t, \tau) - q^\top(t, \theta_0)I^{-1}(\theta_0)q(\tau, \theta_0)),$$

where θ_0 is an unknown value of the parameter θ,

$$q^\top(t, \theta) = \left(\frac{\partial G(x, \theta)}{\partial \theta_1}, \dots, \frac{\partial G(x, \theta)}{\partial \theta_k} \right) \Bigg|_{t=G(x,\theta)},$$

$I(\theta)$ is the Fisher information matrix,

$$I(\theta) = \left(E \frac{\partial}{\partial \theta_i} \log g(X, \theta) \frac{\partial}{\partial \theta_j} \log g(X, \theta) \right)_{1 \leq i,j \leq k}, \quad g(x, \theta) = \frac{\partial G(x, \theta)}{\partial x}.$$

A more known simple formula for the Cramér-von Mises statistics in the case of simple hypothesis (3) as well as complex hypothesis (4) is:

$$\omega_n^2 = \frac{1}{12n} + \sum_{i=1}^n \left(t_{(i)} - \frac{i - 1/2}{n} \right)^2,$$

where $t_{(i)} = G(X_{(i)})$ for (3) and $t_{(i)} = G(X_{(i)}, \theta_n)$ for (4).

In many regular case we can avoid the dependence of the limit distribution function from unknown parameters. One of this methods was proposed by Khmaladze (1981). This method use the transformed empirical process

$$w_n(x) = \sqrt{n}\Big\{F_n(x) - \int_{\infty}^{\infty} \Big[q^\top(y,\theta_n)\Big(\int_y^{\infty} q(s,\theta_n)q^\top(s,\theta_n)dG(s,\theta_n)\Big)^{-1} \times \int_y^{\infty} q(z,\theta_n)dF_n(z)\Big]dG(y,\theta_n)\Big\},$$

where θ_n is the maximum likelihood estimation of the parameter θ. The process $w_n(x)$ converges weakly in $L^2[0,1]$ to Wiener process, and the limit distribution of the test, based on the statistic

$$\omega_n^2 = \int_0^1 w_n^2(t)dt,$$

tends to the distribution of functional

$$\omega^2 = \int_0^1 w^2(t)dt,$$

where $w(t)$ is the Wiener process.

More materials on Cramér-von Mises theory can be find in Martynov (1992) and Deheuvels and Martynov (1996).

4. Cramér-von Mises Test for Rasch Model

In this section, we propose to build a goodness Cramér-von Mises test of fit for the mixed Rasch model. The mixed Rasch model allows to define the probabilities

$$p_i = p(x_1, x_2, \dots, x_k) = \int_{-\infty}^{\infty} \prod_{j=1}^{k} \frac{e^{x_j(z-\beta_j)}}{1+e^{(z-\beta_j)}} g(z)dz,$$

where $i = N(x_1, x_2, \dots, x_k)$ is a decimal representation of the binary number $[x_1x_2 \dots x_k]$. Statistical inference about the Rasch model consist on parameters estimation of $(\beta_j' s)$, goodness-of-fit tests for the probabilities p_i, estimation of the density function $g(t)$, or its parameters. As an example will show it in Sec. 7, the probabilities p_i, $i = 0, 1, \dots, 2^k - 1$, do not define completely the density function $g(z)$.

Let us present now the following Lemma giving a necessary condition to ensure that, under Rasch model, and when k is large, the limit of all probabilities is zero. Such a condition is preliminary to valid application of Cramér-von Mises theory, which is devoted to continuous distributions.

We note corresponding probabilities for each value k as $p_i^{(k)}$,

$$p_i^{(k)} = \int_{-\infty}^{\infty} \prod_{j=1}^{k} \frac{e^{x_{k,j}(z-\beta_{k,j})}}{1+e^{(z-\beta_{k,j})}} g(z)dz, \tag{5}$$

where the parameters vector $(\beta_{k,1}, \beta_{k,2}, \dots, \beta_{k,k})'$ and answers vector $(x_{k,1}, x_{k,2}, \dots, x_{k,k})'$ depend of k.

Lemma 4.1. *Let for some* $B > 0$

$$\sup_{k>0} k \max_{1\le j\le k} |\beta_{k,j}| \le B, \tag{6}$$

then

$$\lim_{k\to\infty} \max_{0\le i\le 2^k-1} p_i^{(k)} \to 0.$$

Proof. We have from (5)

$$\begin{aligned} p_i^{(k)} &= \exp\left\{-\sum_{j=1}^{k} x_j\beta_{k,j}\right\}\int_{-\infty}^{\infty}\frac{e^{Mz}g(z)dz}{\prod_{j=1}^{k}(1+e^{(z-\beta_{k,j})})} \le e^B\int_{-\infty}^{\infty}\frac{e^{Mz}g(z)dz}{\prod_{j=1}^{k}(1+e^{(z-\beta_{k,j})})} \\ &= e^B\int_{-\infty}^{0}\frac{e^{Mz}g(z)dz}{\prod_{j=1}^{k}(1+e^{(z-\beta_{k,j})})} + e^B\int_{0}^{\infty}\frac{e^{Mz}g(z)dz}{\prod_{j=1}^{k}(1+e^{(z-\beta_{k,j})})}. \end{aligned}$$

Here, $M = \sum_{j=1}^{k} x_{k,j} \le k$ and we have use the inequality

$$\exp\left\{-\sum_{j=1}^{k} x_j\beta_{k,j}\right\} \le \sup_{k\ge 1}\sum_{j=1}^{k}|\beta_{k,j}| \le B. \tag{7}$$

Furthermore, because $-\beta_{k,j} \ge -B/k \ge -B$,

$$\begin{aligned} p_i^{(k)} &\le e^B\int_{-\infty}^{0}\frac{g(z)dz}{(1+e^{-B}e^{z})^k} + e^B\int_{0}^{\infty}\frac{e^{kz}g(z)dz}{(1+e^{-B/k}e^{z})^k} \\ &\le e^B\int_{-\infty}^{0}\frac{g(z)dz}{(1+e^{-B}e^{z})^k} + e^{2B}\int_{0}^{\infty}\frac{g(z)dz}{(1+e^{B/k}e^{-z})^k} \\ &\le e^B\int_{-\infty}^{0}\frac{g(z)dz}{(1+e^{-B}e^{z})^k} + e^{2B}\int_{0}^{\infty}\frac{g(z)dz}{(1+e^{-z})^k}. \end{aligned}$$

Last sum tends to 0 as k tends to infinity. Hence, the probabilities $p_i^{(k)}$ tends to 0 uniformly.

This lemma support the proposition that if β_j, $j = 1, \ldots, k$, are relatively small and k is large, the probabilities p_i, $i = 1, \ldots, 2^k - 1$, can be relatively small.

But if, for example, all β_j tend to infinity, probability $p_{2^k-1} = p(1, 1, \ldots, 1)$ tends to 1 and all another probabilities tend to zero.

The statistical procedures considered below are based on the n observations of the binary vector X:

$$\begin{aligned} X^{(1)} &= (X_{11}, X_{12}, \ldots, X_{1k})', \\ X^{(2)} &= (X_{21}, X_{22}, \ldots, X_{2k})', \\ &\cdots \\ &\cdots \\ X^{(n)} &= (X_{n1}, X_{12}, \ldots, X_{nk})'. \end{aligned}$$

Each of this observations X^i is binary representation of some integer m_i, $m_i = 0, \dots, 2^k - 1$. We introduce the function N such that $m_i = N(X^{(i)})$. Hence, any initial binary vector sample can be replaced by a sequence of numbers $m_1 m_2, \dots, m_n$.

Because the probabilities p_i are ordered, we can now easily define the distribution function and the empirical distribution function. The "distribution function" $F(t)$ can be defined on the interval [0,1], as the distribution function, taking in the points $j/2^k$ the values

$$F\left(\frac{j}{2^k}\right) = \sum_{i=0}^{j} p_{i-1}, \quad j = 0, \dots, 2^k - 1,$$

where $p_{-1} = 0$. Let $F(t) = 0$, when $t < 0$, and $F(t) = 1$, when $t > 1$.

In the Rasch model, the probabilities p_i depend of the difficulty parameters $\beta = (\beta_1, \dots, \beta_k)'$ and of the possible parameters of the density g, $\theta = (\theta_1, \dots, \theta_m)'$ in the following form

$$p_{i,\beta,\theta} = \int_{-\infty}^{\infty} \prod_{j=1}^{k} \frac{e^{x_j(z-\beta_j)}}{1 + e^{(z-\beta_j)}} g(z, \theta_1, \dots, \theta_m) dz.$$

Hence, the introduced distribution function depend on the $k + m$ parameters in β and θ, $F(t) = F(t, \beta, \theta)$.

The n observations

$$X^{(1)} = (X_{11}, X_{12}, \dots, X_{1k})',$$
$$X^{(2)} = (X_{21}, X_{22}, \dots, X_{2k})', \dots, X^{(n)} = (X_{n1}, X_{12}, \dots, X_{nk})',$$

correspond to n values $\frac{N(X^{(i)})}{2^k}$, inside the interval [0,1].

Then the empirical distribution function can be represented now as

$$F_n(x) = \frac{1}{n} \sum_{i=1}^{n} I\left\{x < \frac{N(X^{(i)})}{2^k}\right\}, \quad x \in [0, 1],$$

where $I\{A\}$ indicate the indicator function of a set A.

The theory of the Cramér-von Mises-type test for testing complex hypothesis when the parameters are unknown needs to have exact formulas describing dependency between the distribution function and unknown parameters. So, it is not possible to apply "straightforward" the theory, developed for the Cramér-von Mises theory with the introduced above formula for distribution function. Nevertheless, we can use the Cramér-von Mises statistics as a "measure of discrepancy" between the estimated probabilities to the Rasch model theoretical probabilities. The Cramér-von Mises statistics can be represented now as

$$\omega_n^2 = n \int_0^1 (F_n(x) - F(x, \hat{\beta}, \hat{\theta}))^2 dF(x, \hat{\beta}, \hat{\theta}). \tag{8}$$

The distribution function $F(x, \hat{\beta}, \hat{\theta})$ was defined above. It is not uniform in the general case. The statistics ω_n^2 can be rewritten in the standard form

$$\omega_n^2 = n \int_0^1 (F_n^*(t) - t)^2 dt,$$

where $F_n^*(t)$ is constructed on the base of the transformed sample $T_i = F(\frac{N(X^{(i)}}{2^k}, \hat{\beta}, \hat{\theta})$. Hence

$$F_n^*(t) = \frac{1}{n}\sum_{i=1}^{n} I\left\{t < F\left(\frac{N(X^i)}{2^k}, \hat{\beta}, \hat{\theta}\right)\right\}, \quad t \in [0, 1]. \tag{9}$$

Let us, now, present our method by giving formally the algorithm allowing to apply Cramér-von Mises theory to Rasch model:

- Compute $\hat{\beta}$, estimation of β;
- Compute as above, the observed value $\hat{\omega}_n^2$ of the Cramér-von Mises statistics ω_n^2, using the previous estimation of β;
- Generate a large number N of samples each with size n using the same previous estimation of β, as a known parameter;
- Compute the estimate $\hat{\beta}_m$ of $\hat{\beta}$ for each sample, $1 \leq m \leq N$;
- For each generated sample above, compute the value $\hat{\omega}_n^2(m)$, $1 \leq m \leq N$, of the statistics ω_n^2;
- As an estimation of the p-value for the test based on ω_n^2, it is straightforward to take the proportion of observed (on generated samples) of $\hat{\omega}_n^2(m)$, $1 \leq m \leq N$, with values less than $\hat{\omega}_n^2$.

Readers will find a good review on goodness-of-fit tests for the Rasch model, in Glas and Verhelst (1995). Another good reference about χ^2-type test is proposed in Andersen (1973). Of course, there was no similar reference with use of Cramér-von Mises theory to Rasch modeling.

5. Estimation of Parameters β and the Latent Density Function $g(z)$

We will now consider the problem of estimating a density function $g(z)$ of the latent variable z in the model

$$p(x_1, x_2, \ldots, x_k) = \int_{-\infty}^{\infty} \prod_{j=1}^{k} \frac{e^{x_j(z-\beta_j)}}{1 + e^{(z-\beta_j)}} g(z)dz.$$

We have found in Sec. 2 that the probability p_i can be represented as (see (2))

$$p_i = p(x_1, x_2, \ldots, x_k) = e^{-\sum_{i=1}^{k} x_i\beta_i} S(M, \beta), \quad i = 0, \ 1, \ldots, 2^k - 1. \tag{10}$$

where

$$S(M, \beta) = \int_{-\infty}^{\infty} \frac{1}{\prod_{j=1}^{k}(1 + e^{(z-\beta_j)})} e^{Mz} g(z)dz, \tag{11}$$

depend only from values M, $M = 0, 1, \ldots, k$, and β.

Let $\hat{p}_i, i = 0, \ldots, 2^k - 1$, are the usual estimations of the probabilities p_i, $i = 0, \ldots, 2^k - 1$, obtained with the sample $X^{(1)}, X^{(2)}, \ldots, X^{(n)}$:

$$\hat{p}_i = \hat{p}(x_1, x_2, \ldots, x_k) = \frac{\text{number of } X^{(i)} \text{ with } X_{i1} = x_1, X_{i2} = x_2, \ldots, X_{i1} = x_k}{n},$$

where n is the total number of observations and i is the decimal representation of the binary answer $[x_1x_2 \ldots x_k]$. The estimates $\hat{p}_i$ tend to p_i when $n \to \infty$.

Changing p_i on his estimates $\hat{p}_i$ we have from (10) the follow 2^k equations

$$\hat{p}_i = p(x_1, x_2, \ldots, x_k) = e^{-\sum_{i=1}^{k} x_i \beta_i} S(M, \beta), \quad i = 0,\ 1, \ldots, 2^k - 1 \tag{12}$$

for $2k + 1$ variables $\beta_1, \ldots, \beta_k$ and $S(0, \beta), S(1, \beta), \ldots, S(k, \beta)$.

Let us now define new notations for the probabilities $\hat{p}_i$:

$$\hat{p}_0 = \hat{p}_{0,C_k^0} = \hat{p}_{0,1},$$
$$\hat{p}_1 = \hat{p}_{1,1},\ \hat{p}_2 = \hat{p}_{1,2}, \ldots, \hat{p}_{C_k^1} = \hat{p}_{1,C_k^1},$$
$$\hat{p}_{C_k^1+1} = \hat{p}_{2,1}, \hat{p}_{C_k^1+2} = \hat{p}_{2,2}, \ldots, \hat{p}_{C_k^1+C_k^2} = \hat{p}_{2,C_k^2},$$
$$\ldots$$
$$\ldots$$
$$\hat{p}_{C_k^{k-2}+\cdots+C_k^1+1} = \hat{p}_{k-1,1}, \ldots, \hat{p}_{C_k^{k-2}+\cdots+C_k^1+C_k^{k-1}} = \hat{p}_{k-1,C_k^{k-1}},$$
$$\hat{p}_{2^k-1} = \hat{p}_{k,1}.$$

Analogous notations is valid also for p_i.

Let $J_M = \ln S(M, \beta)$ $i = 0, \ldots, k$, and $\hat{b}_{i,j} = \ln \hat{p}_{i,j}$. We can transform the equations system (12) for each i and j with $\hat{p}_{i,j} \neq 0$ the follow linear equations system:

$$J_0 = \hat{b}_{0,1}, \quad \text{for } M = 0,$$
$$\beta_i + J_1 = \hat{b}_{1,i}, 1 \le i \le k,\ \text{for } M = 1,$$
$$\beta_{i_1} + \beta_{i_2} + J_2 = \hat{b}_{2,i}, \quad 1 \le i_1 < i_2 \le C_k^2,\ \text{for } M = 2,$$
$$\beta_{i_1} + \beta_{i_2} + \beta_{i_3} + J_3 = \hat{b}_{3,i},\ 1 \le i_1 < i_2 < i_3 \le C_k^3,\ \text{for } M = 3,$$
$$\ldots$$
$$\ldots$$
$$\beta_{i_1} + \cdots + \beta_{i_{k-1}} + J_{k-1} = \hat{b}_{k-1,i},\ 1 \le i_1 < \cdots < i_{k-1} \le C_k^{k-1},\ \text{for } M = k-1,$$
$$\beta_1 + \cdots + \beta_k + J_k = \hat{b}_{k,1},\ \text{for } M = k.$$

This is a classical system of linear equations with $2k+1$ unknown $(\beta_1, \ldots, \beta_k, J_0, \ldots, J_k)$ and $2^k - 1$ equations. We can write it, in a matrix form, first as:

$$AV = b,$$

where V is the $2k+1$ vector of unknowns, b the vector with components $b_{i,j} = \ln p_{i,j}$ and A is a singular matrix of $(2^k - 1)$ rows and $(2k+1)$ columns. Then, including estimation of the $p_{i,j}$ and using generalized linear models theory, the general solution can be obtained as:

$$\widehat{V} = ({}^tA\widehat{W}^{-1}A)^{-}({}^tA)\widehat{W}^{-1}\hat{b},$$

where the superscribe – indicate a generalized inverse of a matrix (Rao, 1973). $\hat{b}$ is the vector with components $\hat{b}_{i,j} = \ln \hat{p}_{i,j}$, and $\widehat{W}$ the estimated asymptotic variance matrix of $\hat{b}$, easily obtained from multinomial natural distribution on cells i, using delta-method (see Agresti, 1990).

One can easily show that the matrix A is of rank $2k$, so we must add an identifiability linear constraint of rank 1 to get estimable set of parameters (Rao, 1973). Setting the value of a specific β_{j_0} or the sum over all β_j to zero are the most classically used, with easy and interesting interpretation, for physical meaning of the parameter values obtained. Alternatively, a linear identifiability constraint on the J_M's parameters, can also, theoretically, be done, but will give new parameters with complicated interpretation. As usual, for similar problems, sampling zero's can occurs and can be managed, by grouping cells, or putting the observed number of individuals in cells as low as possible.

We have also got estimates of the parameters β non depending of the latent density function $g(z)$. The information on the density function is included in the estimated values of $S(M, \beta)$.

Using analogous methods with the estimated values of β, we can start from the following equation integrals

$$S(M, \beta) = \int_{-\infty}^{\infty} \frac{e^{M,z} g(z) dz}{\prod_j (1 + e^{(z-\beta_j)})}, \tag{13}$$

where M takes the $r + 1$ values from 0 to r. The $S(M, \beta)$ was estimated in the previous section, through the J_M's.

This an inverse problem. The $k + 1$ integral transformations in (13), can be rewritten as

$$S(M, \beta) = \int_{-\infty}^{\infty} K(z) g(z) dz. \tag{14}$$

We can directly derive a discrete version of that "integral equation" by choosing a finish set of z_i values, and try to evaluate g in those points. More precisely, the previous $k + 1$ integral transformations can be written:

$$S_i = \sum_1^l K_i(z_i) g(z_i)$$

where without confusion, we noted S and g the corresponding two vectors, of respective dimension $k + 1$ and l, and K, the corresponding matrix, with $(k + 1)$ rows and l column, where l is the suitably chosen number of z_i values.

We can write:

$$S = Kg.$$

Then, as, previously, applying weighted least square theory of linear models, we can get the general form of estimator of g below

$$\hat{g} = ({}^t\widehat{K}\widehat{\Omega}^{-1}\widehat{K})^{-}({}^t\widehat{K})\widehat{\Omega}^{-1}\widehat{S},$$

Table 1
Parameter estimates for β and σ^2 and their standard errors

	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_5$	$\hat{\beta}_6$	$\hat{\beta}_7$	$\hat{\beta}_8$	$\hat{\beta}_9$	$\hat{\sigma^2}$
Estimate	−.702	−0.270	.096	2.050	−1.269	0.838	−1.418	−1.376	−.601	0.7
SE MLE	0.111	0.105	0.104	0.153	0.124	0.112	0.125	0.125	0.109	

where the superscribe – indicate a generalized inverse of a matrix. $\widehat{\Omega}^{-1}$ is the estimated asymptotic variance of $\widehat{S}$, easily obtained, using the delta-method (see Agresti, 1990) from the previous generalized linear equations allowing to estimate β's and J's parameters. We must, of course, constraint the solution to be positive.

Our method is different from classical methods, like those that one can found in the book of Baker (1992), Glas and Ellis (1993), Andersen (1972), or Feddag et al. (2003), which concern only some specific parametrization of g (Gaussian distribution).

6. Application of the Cramér-Von Mises Test to Real Data and Simulations

In this section, we illustrate the application of the Cramér-von Mises methodology presented in Secs. 3 and 4, to the analysis of a real set of data from quality of life field. The sample is composed by answers of the 470 depressive patients to 9 questions, containing in the French version of the Sickness Impact Profile (SIP) questionnaire (see Feddag et al., 2003 for a detailed description). The Rasch model depends in this example on the 9-dimension parameter vector $\beta = (\beta_1, \ldots, \beta_9)$. As in Feddag et al. (2003), the density function $g(z)$ was chosen as the density function of the normal distribution with zero mean and unknown variance σ^2. This natural

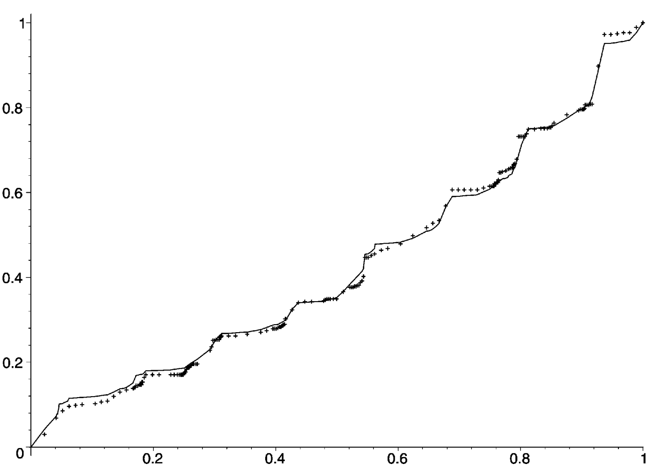

Figure 1. Non transformed empirical and theoretical distributions.

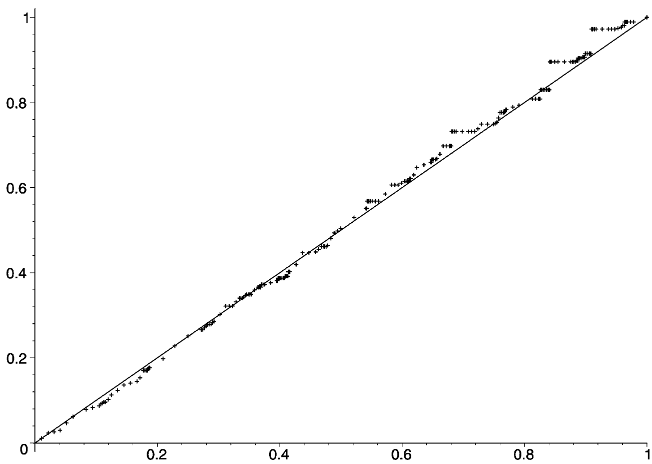

Figure 2. Transformed empirical and theoretical distributions.

choice for the parametrization of g is the most commonly maid in IRT. Other new and appealing choices, mainly coming when handling the problem with a Bayesian point of view can be done. All computations done in the following section were done using Mapple 8.0 software.

In Table 1, we can see the estimates obtained when using maximum likelihood (ML) method as given in Feddag et al. (2003).

In Fig. 1, we can see the plots of the empirical distribution of the data (dashed curve, $F_n(t)$ and the theoretical distribution (continuous curve, $F(t, \hat{\beta}, \hat{\theta})$, see (8)). For that data, the maximum value of $\hat{p}_i$, $i = 1, \ldots 512$, is equal to 0.0305.

Analogously in Fig. 2, we can see the plots of the empirical distribution of the transformed data (dashed curve, $F_n^*(t)$) and the theoretical distribution (continuous curve, $F(t) = t$)), see Deheuvels and Martynov (1996).

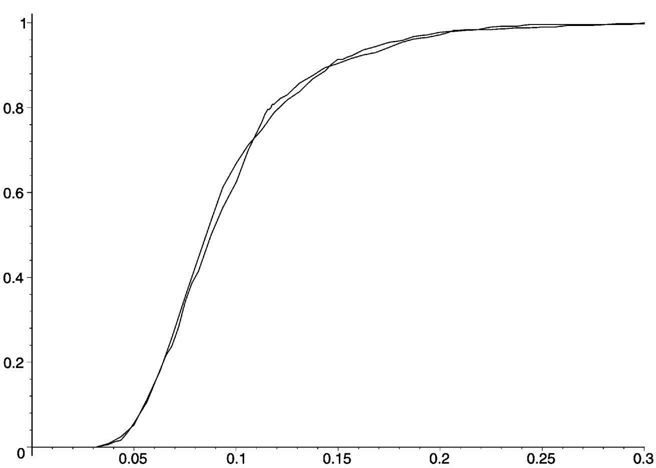

Figure 3. Comparison of the empirical distributions of ω_{470}^2, when $\beta = \hat{\beta}$ is ML estimate, and when $\beta = \beta^{(1)}$.

The computed value of the statistics ω^2_{470} is 0.137, obtained with the parameter β considered as known. $\beta = \hat{\beta}$ was the value taken, where $\hat{\beta}$ is the ML estimate given in Table 1. Then, we generated 500 samples with 470 observations (answers on 9 question) to estimate the p-value corresponding to the previous observed value of ω^2_{470}. Based on this simulations, we found a estimated p-value o 0.864. So, we cannot reject the nulle hypothesis, that using that data, the underlying model is a Rasch model.

Because the distribution of the statistics depends on unknown parameters, we did the same computing and simulations for another parameter chosen inside the confident intervals $\beta^{(1)} = (-0.8, -0.2, 0.15, 1.9, -1.4, 0, 9, -1.2, -1.5, -0.8)$. The corresponding estimated p-value is 0.876. The comparison of the two empirical distribution functions is represented in Fig. 3. It follows from these simulations, that variability of the distribution in the neighborhood of the estimated parameters is very low.

7. Conclusions

In this article:

(i) a Cramér-von Mises-type statistics is built after a continuous representation of the data was derived;
(ii) a set of like linear models estimating equations allowing to estimate jointly the difficulty parameters and some useful functionals of the latent density are established;
(iii) a set of like linear models estimating equations allowing to estimate latent density values on some specific chosen, obtained through an inversion of the integral transformation that specify the probability model, are established;
(iv) application to a real quality of life data set, with adapted simulations illustrate the interest of the Cramér-von Mises-type test given in the beginning.

Usually, Cramér-von Mises-type statistics needs continuous data. Our main idea was to first derive a continuous representation of the data (ordering the data values), then derive the usual Cramér-von Mises-type test. We have chosen specific, but natural, ordering of the probabilities $p_i = p(x_1, \ldots, x_k)$. If this order is chosen in another manner, the value of the statistics ω^2 changes, but our method is applicable, because after estimation of the β's, we need to find by simulation the distribution of the statistic, corresponding to the estimated set of β's so using, the same ordering of p_i.

Our simulation shows that the change of the parameters in the neighborhood of the true parameters values region lead to negligible change of the significance level.

Estimation of the unknown latent density function is based on the fact that the probability of the observed vectors are obtained by integral transformation of that latent density.

References

Agresti, A. (1990). *Categorical Data Analysis*. New-York: John Wiley, Inc.

Andersen, E. B. (1972). Asymptotic properties of conditional maximum likelihood estimators. *J. Roy. Statist. Soc. B* 32:283–301.

Andersen, E. B. (1973). A goodness-of-fit test for the Rasch model. *Psychometrika* 38:123–140.

Baker, F. B. (1992). *Item Response Theory. Parameter Estimation Theory*. New York, Basel, Hong Kong: Marcel Dekker Inc.

Deheuvels, P., Martynov, G. V. (1996). Cramér-von Mises-type tests with applications to tests of independence for multivariate extreme-value distributions. *Commun. Statist. Theor. Meth.* 25:871–908.

Feddag, M. L., Grama, I., Mesbah, M. (2003). Generalized estimating equations for mixed logistic models. *Commun. Statist. Theor. Meth.* 32:851–874.

Fischer, G. H., Molenaar, I. W., eds. (1995). *Rasch Models. Foundations, Recent Developments, and Applications*. New-York: Springer-Verlag.

Glas, C. A. W., Ellis, J. (1993). *Rasch Scaling Program: User's Manual Guide*. The Netherlands: iec ProGAMMA, Groningen.

Glas, C. A. W., Verhelst, N. D. (1995). Testing the Rasch model. In: Fischer, G., Molenaar, I., eds. *Rasch Models Foundations, Recent Developments and Applications*. New-York: Springer-Verlag.

Hamon, A., Dupuy, J. F., Mesbah, M. (2002). Validation of model assumptions in quality of life measurements. In: Huber, C., Balakhrishnan, N., Nikulin, M., Mesbah, M., eds. *Goodness of Fit and Validation of Models*. Boston: Birkhauser.

Kac, M., Kiefer, J., Wolfowitz, J. (1955). On test of normality and other tests of goodness-of-fit based on distance methods. *Ann. Math. Statist.* 26:189–211.

Khmaladze, E. V. (1981). A martingale approach in the theory of parametric goodness-of-fit tests. *Theor. Prob. Appl.* 26:N 2

Martynov, G. V. (1992). Statistical tests based on empirical processes and related questions. *J. Soviet. Math.* 61:2195–2271.

Mesbah, M., Cole, B. F., Lee, M. L. T., eds. (2002). *Statistical Design, Measurements and Analysis of Health Related Quality of Life*. Amsterdam: Kluwer.

Neuhaus, G. (1976). Asymptotic power properties of the Cramér-von Mises statistic under contiguous alternatives. *J. Multivariate Anal.* 6:95–110.

Rao, C. R. (1973). *Linear Statistical Inference and Its Applications.* 2nd ed. London: John Wiley, Inc.

Rasch, G. (1960). *Probabilistic Models for Some Intelligence and Attainment Test*. Copengagen: Danisch Institute for Education Research.