PSYCHOMETRIKA—VOL. 73, NO. 1, 145–151
MARCH 2008
DOI: 10.1007/S11336-007-9023-2

ON THE RELATIONSHIPS BETWEEN SUM SCORE BASED ESTIMATION AND JOINT MAXIMUM LIKELIHOOD ESTIMATION

GUIDO DEL PINO AND ERNESTO SAN MARTÍN

PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE

JORGE GONZÁLEZ AND PAUL DE BOECK

K.U. LEUVEN

This paper analyzes the sum score based (SSB) formulation of the Rasch model, where items and sum scores of persons are considered as factors in a logit model. After reviewing the evolution leading to the equality between their maximum likelihood estimates, the SSB model is then discussed from the point of view of pseudo-likelihood and of misspecified models. This is then employed to provide new insights into the origin of the known inconsistency of the difficulty parameter estimates in the Rasch model. The main results consist of exact relationships between the estimated standard errors for both models; and, for the ability parameters, an upper bound for the estimated standard errors of the Rasch model in terms of those for the SSB model, which are more easily available.

Key words: Rasch model, standard error, information matrix, pseudo-likelihood.

1. Introduction

The Rasch model has motivated a large field of psychometric research dealing not only with its multiple extensions (see, e.g. Fischer & Molenaar, 1995; Van der Linden & Hambleton, 1997; Boomsma, van Duijn, & Snijders, 2001), but also with its theoretical properties. This paper intends to be a contribution to the theoretical understanding of some estimation aspects of the Rasch model, whose standard specification is as follows:

$$Y_{ij} \sim \text{Bern}(p_{ij}), \quad i = 1, \ldots, n, \ j = 1, \ldots, k, \qquad \perp\!\!\!\perp_{i,j} Y_{ij}, \qquad \text{logit}(p_{ij}) = \theta_i - \beta_j, \tag{1}$$

where $\perp\!\!\!\perp_{i,j} Y_{ij}$ denotes the mutual independence of the Y_{ij}, and both the θ_i and the β_j are unknown parameters. The $n + k$ parameters of model (1) are identified after imposing a linear restriction on the β_j, like $\sum_{j=1}^{k} \beta_j = 0$. In educational tests, $Y_{ij} = 1$ if person i answers item j correctly, $\theta_i \in \mathbb{R}$ represents the ability of person i, and $\beta_j \in \mathbb{R}$ represents the difficulty of item j.

To make a distinction with the conditional (CMLE) and the marginal (MMLE) maximum likelihood estimates, the MLE for model (1), say $\widehat{\theta_i}$, $i = 1, \ldots, n$, and $\widehat{\beta_j}$, $j = 1, \ldots, k$, are called joint or unconditional MLE and will be denoted by JMLE. Andersen (1980) proved the inconsistency of the $\widehat{\beta_j}$ as $n \to \infty$ for $k = 2$, while Ghosh (1995) extended this result for general k, using a proof by contradiction. As a practical solution to this problem, Andersen (1980, p. 245) proposed the multiplicative bias correcting factor $(k - 1)/k$. There is empirical evidence that the bias-corrected JMLE is close to the CMLE or the MMLE for large k. Holland (1990)

The authors acknowledge partial financial support from the FONDECYT Project No. 1060722 from the Chilean Government, and the BIL05/03 grant to P. De Boeck, E. Lesaffre and G. Molenberghs (Flanders) for a collaboration with G. del Pino, E. San Martín, F. Quintana and J. Manzi (Chile).

Requests for reprints should be sent to Ernesto San Martín, Department of Statistics, Pontificia Universidad Católica de Chile, Casilla 306, Santiago 22, Chile. E-mail: esanmart@mat.puc.cl

© 2007 The Psychometric Society

proposes a theoretical formulation that gives some intuition on the closeness of both JMLE and CMLE to MMLE. Thus, for large-scale tests with many items, the biased corrected JMLE may be considered to be a reasonable alternative.

Using the + sign to indicate summation over all possible values of an argument, the sum score of examinee i is Y_{i+}. Letting $I_t = \{i \mid Y_{i+} = t\}$ be the set of all examinees with sum score t, it is easily shown that the JMLE $\widehat{\theta_i}$ are constant within I_t, a fact already pointed out by Rasch himself (1960, chapter VI). Based on this fact, Perline, Wright, and Wainer (1979, p. 239) stated that "a practical implementation of the model is that statistical estimates of abilities and item difficulties proceed *as if* everyone with the same sum score has exactly the same ability". This can be achieved by posing an artificial statistical model, in which the sum score and the item appear as categorical factors, for the sole purpose of obtaining point estimates. We refer to such a model as *sum score based* (SSB) and denote the parameter estimates by SSBE.

Though the original spirit of the SSBE was to obtain an approximation of the JMLE, it turns out that they are identical. Mellenbergh and Vijn (1981) used a log-linear formulation, in which items and person sum scores are used as factors. These authors reported that the estimates of the item parameters and of the score group parameters, computed using iterative proportional fitting, were very similar to the JMLE. An alternative SSB formulation, which will be used throughout this paper, is the additive logit model

$$Y_{ij} \sim \text{Bern}(p_{ij}), \quad i = 1, \ldots, n, \ j = 1, \ldots, k, \qquad \coprod_{i,j} Y_{ij}, \qquad \text{logit}(p_{ij}) = \gamma_{t(i)} - \beta_j, \quad (2)$$

where $\gamma_{t(i)}$ represents a proxy of the ability for any examinee i whose sum score is equal to t. The equivalent grouped form of this model (obtained by a sufficiency reduction) is

$$N_{tj} \sim \text{Bin}\big(n_t, p_j^t\big), \qquad \text{logit}\big(p_j^t\big) = \gamma_t - \beta_j, \quad t = 1, \ldots, k-1, \ j = 1, \ldots, k, \qquad \coprod_{t,j} N_{tj}, \quad (3)$$

where N_{tj} is the random variable indicating the number of persons with a sum score t who give a correct response to item j, n_t is the number of persons with sum score t and p_j^t is the probability of a person with sum score t to give a correct response to item j.

Verhelst and Molenaar (1988) analyzed model (2) and developed an iteratively reweighted least square estimation method starting from the proportions in (3) and showed that this method leads to a point estimate that is equivalent to the JMLE (see also Molenaar, 1995, Section 3). Blackwood and Bradley (1989) provided a formal proof that the likelihood equations for models (1) and (2) have the same solution, which we write as JMLE = SSBE. More recently, Haberman (2004) reported the same result in terms of (3) and provided compelling evidence of the computational savings achieved by fitting this model using standard software, particularly for large data sets. He also suggested that the JMLE can be used as a starting point for a Newton Raphson algorithm that computes the CMLE. It may also be mentioned that although MMLE is currently the most used, it is a matter of some controversy what the effect is of a misspecification of the ability distribution.

In this paper, Section 2 analyzes the equality JMLE = SSBE, interpreting SSBE as a pseudo-likelihood estimate (PLE). In this context, this equality constitutes a remarkable behaviour for a PLE. This is then used to provide insight into the source of the inconsistency of the difficulty parameters in the Rasch model. Section 3 shows that the equality of the point estimates can be extended to the asymptotic standard errors of the difficulty parameters obtained with the JMLE and with the SSBE. As far as the ability parameters are concerned, this section provides an exact formula and useful bounds that link the asymptotic standard errors of the JMLE to those of the SSBE (for the parameters related with the sum score factor in the SSB model). The theoretical results in Section 3 are new, and they may be of interest not only in practice, if one would use the SSB model, but also as additional properties satisfied by estimation methods for the Rasch model.

2. The SSB Formulation of the Rasch Model as a Pseudo-Likelihood

Since the statistical assumptions of the SSB model (2) are in contradiction with those of the original Rasch model, the SSBE can be considered as an instance of a pseudo-likelihood estimate (PLE) in the sense of Besag (1975), as the MLE for a misspecified model. The motivation for the PLE is that it is computationally easier to obtain than the MLE for the original model. In the case of the SSBE, the misspecification arises from ignoring the randomness involved in defining the groups I_t by the sum scores, and the resulting correlation between the Y_{ij} and the Y_{i+}; see Verhelst and Molenaar (1988, Section 6). In general, the PLE does not coincide with the MLE, except asymptotically, so that the equality SSBE = JMLE constitutes a remarkable exception. This equality also provides a new insight into the source of the bias and inconsistency of the JMLE for $\widehat{\beta}_j$. Unlike the Rasch model, the SSB model has no incidental parameters, i.e. the number of parameters does not change with the number of subjects. The bias and inconsistency of the JMLE can now be explained by the fact that the SSB model is misspecified.

Another view on the inconsistency and bias of the $\widehat{\beta}_j$ is obtained by assuming that the examinees fall into $k+1$ latent classes, that is, as many classes as there are sum scores (each with a constant ability ω_t, $t = 0, 1, \ldots, k$). For known class probabilities, these parameters can be consistently estimated from the frequencies of the sum scores. Assuming these class probabilities, $\widehat{\gamma}_t$ converges in probability to some value γ_t. Then the consistency of the JMLE $\widehat{\beta}_j$ can only be achieved if $\gamma_t = \omega_t$ for $t = 0, 1, \ldots, k$. But this is not the case, since the SSB model implicitly assumes that any examinee with sum score t falls in the class with ability parameter ω_t, thus ignoring the possibility of misclassification. The more extreme cases are $t = 0$ and $t = k$, where $\gamma_0 = -\infty$ and $\gamma_k = \infty$.

3. Information Matrices and Standard Errors

3.1. Main Results and Illustrations

The (asymptotic) standard errors of SSBE and JMLE are the square root diagonal elements of the inverses of the corresponding information matrices. Since (1) and (3) are generalized linear models with a canonical link, their information matrices coincide with the negative Hessian of the corresponding log-likelihoods (McCullagh & Nelder, 1989). When evaluated at the MLE we denote these matrices by $\mathcal{I}_{\text{JMLE}}$ and $\mathcal{I}_{\text{SSBE}}$, respectively. Under the general identification restriction $\sum_{j=1}^{k} c_j \beta_j = 0$, the following key equalities are straightforward:

$$
\begin{aligned}
&[\mathcal{I}_{\text{SSBE}}]_{tt} = n_t [\mathcal{I}_{\text{JMLE}}]_{ii}, \quad i \in I_t,\ 1 \le t < k; \qquad [\mathcal{I}_{\text{SSBE}}]_{tt'} = 0, \quad 1 \le t \ne t' < k,\\
&[\mathcal{I}_{\text{SSBE}}]_{jj} = [\mathcal{I}_{\text{JMLE}}]_{jj}, \quad 1 \le j < k; \qquad [\mathcal{I}_{\text{SSBE}}]_{jj'} = [\mathcal{I}_{\text{JMLE}}]_{jj'}, \quad 1 \le j \ne j' < k,\\
&[\mathcal{I}_{\text{SSBE}}]_{tj} = n_t [\mathcal{I}_{\text{JMLE}}]_{ij}, \quad i \in I_t,\ 1 \le t, j < k.
\end{aligned}
\tag{4}
$$

Relationships between the inverse information matrices lead to the following results on the estimated standard errors $s.e.(\widehat{\beta}_j)$ and $s.e.(\widehat{\theta}_i)$ with $i \in I_t$, and $s.e.(\widehat{\gamma}_t)$ with $t \in \mathcal{T}$:

$$
s.e.\big(\widehat{\beta}_j\big) \text{ are identical for the SSBE and the JMLE,} \tag{5}
$$

$$
\big(s.e.\big(\widehat{\theta}_i\big)\big)^2 = \big(s.e.\big(\widehat{\gamma}_t\big)\big)^2 + \frac{n_t - 1}{[\mathcal{I}_{\text{SSBE}}]_{tt}} = \big(s.e.\big(\widehat{\gamma}_t\big)\big)^2 + \frac{n_t - 1}{n_t v_{t+}}, \tag{6}
$$

$$
\frac{1}{\sqrt{v_{t+}}\, s.e.(\widehat{\gamma}_t)} \le \frac{s.e.(\widehat{\theta}_i)}{s.e.(\widehat{\gamma}_t)} \le \sqrt{n_t} \quad \text{for all } i \in I_t,\ t \in \mathcal{T}, \tag{7}
$$

with $v_{tj} = u''(\widehat{\eta}_{ij})$, where $\widehat{\eta}_{ij} = \widehat{\theta}_i - \widehat{\beta}_j = \widehat{\gamma}_t - \widehat{\beta}_j$ when $i \in I_t$ and $t \in \mathcal{T}$, and $u(\eta) = \log(1 + e^{\eta})$.

TABLE 1.
Data matrix used for the SSBE with $n = 10$ and $k = 5$.

t	j	n_t	n_{tj}	t	j	n_t	n_{tj}	t	j	n_t	n_{tj}	t	j	n_t	n_{tj}
1	1	1	0	2	1	5	4	3	1	2	1	4	1	2	2
1	2	1	1	2	2	5	1	3	2	2	2	4	2	2	2
1	3	1	0	2	3	5	4	3	3	2	2	4	3	2	2
1	4	1	0	2	4	5	0	3	4	2	1	4	4	2	2
1	5	1	0	2	5	5	1	3	5	2	0	4	5	2	0

Note that $s.e.(\widehat{\theta}_i) \geq s.e.(\widehat{\gamma}_t)$, with equality only attained when there is just one examinee with a sum score equal to t. The upper bound in (7) is a useful approximation to $s.e.(\widehat{\theta}_i)$, since it tends to be quite sharp for large-scale tests with many items. Moreover, (7) implies that $1/\sqrt{v_{t+}} \leq s.e.(\widehat{\theta}_i)$; here $1/\sqrt{v_{t+}}$ coincides with the estimated standard error when the item parameter estimates are taken as if they were the true values. The proof of these results actually comes from analytic equalities and inequalities involving the inverse information matrices for the Rasch and SSB models.

Before sketching a proof of (5), (6) and (7) we illustrate their use, as well as the structure of the inverse information matrices, with a small example. Assume there are $n = 10$ examinees and $k = 5$ items, with response patterns: $Y_1 = (0, 1, 0, 0, 0)$, $Y_2 = (1, 0, 1, 0, 0)$, $Y_3 = (0, 1, 1, 0, 0)$, $Y_4 = (1, 0, 0, 0, 1)$, $Y_5 = (1, 0, 1, 0, 0)$, $Y_6 = (1, 0, 1, 0, 0)$, $Y_7 = (0, 1, 1, 1, 0)$, $Y_8 = (1, 1, 1, 0, 0)$, $Y_9 = (1, 1, 1, 1, 0)$ and $Y_{10} = (1, 1, 1, 1, 0)$. To obtain the SSBE, the data are stored in the 20×4 array obtained by stacking the four blocks in Table 1 under each other.

With the identification restriction $\beta_5 = -\sum_{j=1}^{4} \beta_j$, S-PLUS produces the following matrix $[\mathcal{I}_{\text{SSBE}}]^{-1} \equiv \widehat{\text{Cov}}(\widehat{\gamma}_1, \ldots, \widehat{\gamma}_4, \widehat{\beta}_1, \ldots, \widehat{\beta}_4)$:

$$[\mathcal{I}_{\text{SSBE}}]^{-1} = \left(\begin{array}{cccc|cccc} 1.636 & 0.089 & -0.002 & -0.160 & 0.140 & 0.083 & 0.206 & -0.095 \\ 0.089 & 0.317 & 0.016 & -0.114 & 0.112 & 0.109 & 0.086 & -0.028 \\ -0.002 & 0.016 & 0.750 & -0.005 & 0.004 & 0.046 & -0.052 & 0.108 \\ -0.160 & -0.114 & -0.005 & 1.186 & -0.174 & -0.134 & -0.225 & 0.106 \\ \hline 0.140 & 0.112 & 0.004 & -0.174 & 0.557 & -0.029 & -0.048 & -0.129 \\ 0.083 & 0.109 & 0.046 & -0.134 & -0.029 & 0.511 & -0.054 & -0.108 \\ 0.206 & 0.086 & -0.052 & -0.225 & -0.048 & -0.054 & 0.668 & -0.168 \\ -0.095 & -0.028 & 0.108 & 0.106 & -0.129 & -0.108 & -0.168 & 0.590 \end{array}\right). \tag{8}$$

Using (4) we obtain $\mathcal{I}_{\text{JMLE}}$, with the following inverse:

$$\left(\begin{array}{ccccccc} \textit{1.636} & 0.089 & 0.089 & 0.089 & 0.089 & 0.089 & -0.002 \\ 0.089 & \textit{1.293} & \textit{0.073} & \textit{0.073} & \textit{0.073} & \textit{0.073} & 0.016 \\ 0.089 & \textit{0.073} & \textit{1.293} & \textit{0.073} & \textit{0.073} & \textit{0.073} & 0.016 \\ 0.089 & \textit{0.073} & \textit{0.073} & \textit{1.293} & \textit{0.073} & \textit{0.073} & 0.016 \\ 0.089 & \textit{0.073} & \textit{0.073} & \textit{0.073} & \textit{1.293} & \textit{0.073} & 0.016 \\ 0.089 & \textit{0.073} & \textit{0.073} & \textit{0.073} & \textit{0.073} & \textit{1.293} & 0.016 \\ -0.002 & 0.016 & 0.016 & 0.016 & 0.016 & 0.016 & \textit{1.470} \\ -0.002 & 0.016 & 0.016 & 0.016 & 0.016 & 0.016 & \textit{0.031} \\ -0.160 & -0.114 & -0.114 & -0.114 & -0.114 & -0.114 & -0.005 \\ -0.160 & -0.114 & -0.114 & -0.114 & -0.114 & -0.114 & -0.005 \\ \hline 0.140 & 0.112 & 0.112 & 0.112 & 0.112 & 0.112 & 0.004 \\ 0.083 & 0.109 & 0.109 & 0.109 & 0.109 & 0.109 & 0.046 \\ 0.206 & 0.086 & 0.086 & 0.086 & 0.086 & 0.086 & -0.052 \\ -0.095 & -0.028 & -0.028 & -0.028 & -0.028 & -0.028 & 0.108 \end{array}\right)$$

$$\left(\begin{array}{ccc|cccc}
-0.002 & -0.160 & -0.160 & 0.140 & 0.083 & 0.206 & -0.095 \\
0.016 & -0.114 & -0.114 & 0.112 & 0.109 & 0.086 & -0.028 \\
0.016 & -0.114 & -0.114 & 0.112 & 0.109 & 0.086 & -0.028 \\
0.016 & -0.114 & -0.114 & 0.112 & 0.109 & 0.086 & -0.028 \\
0.016 & -0.114 & -0.114 & 0.112 & 0.109 & 0.086 & -0.028 \\
0.016 & -0.114 & -0.114 & 0.112 & 0.109 & 0.086 & -0.028 \\
\mathit{0.031} & -0.005 & -0.005 & 0.004 & 0.046 & -0.052 & 0.108 \\
\mathit{1.470} & -0.005 & -0.005 & 0.004 & 0.046 & -0.052 & 0.108 \\
-0.005 & \mathit{2.175} & \mathit{0.197} & -0.174 & -0.134 & -0.225 & 0.106 \\
-0.005 & \mathit{0.197} & \mathit{2.175} & -0.174 & -0.134 & -0.225 & 0.106 \\
\hline
0.004 & -0.174 & -0.174 & 0.557 & -0.029 & -0.048 & -0.129 \\
0.046 & -0.134 & -0.134 & -0.029 & 0.511 & -0.054 & -0.108 \\
-0.052 & -0.225 & -0.225 & -0.048 & -0.054 & 0.668 & -0.168 \\
0.108 & 0.106 & 0.106 & -0.129 & -0.108 & -0.168 & 0.590
\end{array}\right)$$

which matches the results produced when directly computing the JMLE. The equality of the bottom right blocks of these matrices illustrates (5). $\mathcal{I}_{\text{JMLE}}^{-1}$ exhibits many equal values, which corresponds to the property of invariance under permutations within a group of persons with a common sum score. This induces a block decomposition of this matrix and it can be observed that the entries in most blocks are identical to some entry in the smaller matrix (8). The exception concerns the $n_t \times n_t$ diagonal blocks, shown in italics. That the diagonal elements are constant within each block just reflects the fact that the standard errors of $\widehat{\theta}_i$s depend only on the sum score. It is also seen that $s.e.(\widehat{\theta}_i) \geq s.e.(\widehat{\gamma}_t)$ with equality for $n_t = 1$. Applying (6), the standard errors of the JMLE can be recovered from those of the SSBE, as illustrated in Table 2, where the sixth column is the sum of the fourth and fifth columns. Finally, a comparison of the last two columns shows how good the upper bound (7) is.

The upper bound derived in (7) is much better for larger data sets. To illustrate this, a data set of $n = 100$ examinees and $k = 30$ items was simulated. The abilities were generated from a standard normal distribution, and the difficulties were chosen to be equally spaced between -2.5 and 2.5. Both the SSBE and the JMLE were separately and directly computed using the glm function of S-PLUS. Table 3 shows that the bound is actually quite tight.

3.2. *Sketch of the Proof of (5), (6) and (7)*

Let $n = \sum_{t \in \mathcal{T}} n_t$, $\boldsymbol{n} = (n_t : t \in \mathcal{T})$, and $m = k - 1$. Denote the matrices $\mathcal{I}_{\text{JMLE}}$, $\mathcal{I}_{\text{SSBE}}$ and their corresponding inverses as

$$\mathcal{I}_{\text{JMLE}} = \begin{bmatrix} D & A \\ A' & H \end{bmatrix}, \qquad \mathcal{I}_{\text{JMLE}}^{-1} = \begin{bmatrix} X & B \\ B' & W \end{bmatrix},$$

$$\mathcal{I}_{\text{SSBE}} = \begin{bmatrix} \overline{D} & \overline{A} \\ \overline{A'} & H \end{bmatrix}, \qquad \mathcal{I}_{\text{SSBE}}^{-1} = \begin{bmatrix} E & F \\ F' & S \end{bmatrix},$$

TABLE 2.
Computation of $(s.e.(\widehat{\theta}_i))^2$ from $(s.e.(\widehat{\gamma}_t))^2$ and illustration of the upper bound (7).

t	n_t	$[\mathcal{I}_{\text{SSBE}}]_{tt}$	$(s.e.(\widehat{\gamma}_t))^2$	$(n_t - 1)/[\mathcal{I}_{\text{SSBE}}]_{tt}$	$(s.e.(\widehat{\theta}_i))^2$	Quotient	$\sqrt{n_t}$
1	1	0.665	1.636	0.000	1.636	1.0000	1.0000
2	5	4.097	0.317	0.976	1.293	2.0196	2.2361
3	2	1.389	0.750	0.720	1.470	1.4000	1.4142
4	2	1.011	1.186	0.989	2.175	1.3542	1.4142

TABLE 3.
Numerical illustration of relationship (7).

t	n_t	Est. value	$s.e.(\widehat{\gamma_t})$	$s.e.(\widehat{\theta_i})$	Quotient	$\sqrt{n_t}$
2	1	−3.856	0.799	0.799	1.000	1.000
3	1	−3.319	0.688	0.688	1.000	1.000
6	4	−2.211	0.283	0.558	1.972	2.000
7	3	−1.914	0.313	0.537	1.719	1.732
8	3	−1.636	0.303	0.521	1.722	1.732
9	3	−1.373	0.295	0.508	1.723	1.732
10	6	−1.122	0.205	0.497	2.422	2.449
11	5	−0.880	0.220	0.489	2.218	2.236
12	7	−0.646	0.184	0.482	2.615	2.646
13	3	−0.417	0.276	0.477	1.726	1.732
14	4	−0.192	0.238	0.473	1.988	2.000
15	9	0.030	0.160	0.471	2.953	3.000
16	8	0.251	0.169	0.471	2.788	2.828
17	5	0.472	0.213	0.472	2.216	2.236
18	7	0.695	0.182	0.475	2.609	2.646
19	7	0.922	0.184	0.480	2.606	2.646
20	3	1.154	0.283	0.487	1.722	1.732
21	5	1.394	0.225	0.496	2.209	2.236
22	3	1.645	0.296	0.509	1.720	1.732
23	4	1.909	0.266	0.525	1.977	2.000
24	3	2.193	0.318	0.546	1.717	1.732
25	1	2.504	0.575	0.575	1.000	1.000
26	1	2.854	0.617	0.617	1.000	1.000
27	4	3.267	0.347	0.681	1.965	2.000

where D is $n \times n$, A is $n \times m$, H and S are $m \times m$, $\overline{D}$ is $T \times T$, and $\overline{A}$ is $T \times m$. Denote by $\mathcal{C}(\boldsymbol{n})$ the class of all matrices with the structure shown in the numerical example. The proof consists in the following steps:

- $\mathcal{I}_{\text{JMLE}} \in \mathcal{C}(\boldsymbol{n})$: This follows from (4).
- $C \in \mathcal{C}(\boldsymbol{n})$ implies that $C^{-1} \in \mathcal{C}(\boldsymbol{n})$: Performing the same permutation for those rows and columns of C corresponding to persons with raw score t leads to a matrix of the form $P_t C P_t{}'$. The condition $C \in \mathcal{C}(\boldsymbol{n})$ is equivalent to $P_t C P_t^{'} = C$ for all such P_t. Taking the inverse in both sides and using the fact that $P_t^{-1} = P_t^{'}$, it follows that $P_t C^{-1} P_t^{'} = C^{-1}$.
- $W = S$: This proves (5) and it follows from $W = (H - A' D^{-1} A)^{-1}$ and $S = (H - \overline{A}'(\overline{D})^{-1}\overline{A})^{-1}$.
- Let $\widetilde{B}$ be the $T \times m$ matrix with $\widetilde{B}_{jl}$ coinciding with an entry of the lth column of the t block of B. Then $F = \widetilde{B}$: This follows from $CC^{-1} = I_{(n+k-1)\times(n+k-1)}$ and $\overline{C}(\overline{C})^{-1} = I_{(T+k-1)\times(T+k-1)}$.
- Equality (6): This follows from combining the equations $DX + AB' = I_{n\times n}$ and $(\overline{D}E) + \overline{A}\ F'\overline{D}E + \overline{A}\ \tilde{B}' = I_{T\times T}$. This proves the relation between $s.e.(\widehat{\theta_i})$ and $s.e.(\widehat{\gamma_t})$ with $i \in I_t$.
- Inequality (7): This follows from (6) and the fact that for any positive definite matrix M, $1/M_{ii} \leq M^{ii}$, where M^{ii} denotes the ii-entry of its inverse. This proves the bounds for the ratio $s.e.(\widehat{\theta_i})/s.e.(\widehat{\gamma_t})$ with $i \in I_t$.

4. Concluding Remarks

This paper provides additional relationships between the JMLE in the Rasch model and sum square based estimates (SSBE). The SSBE are seen to be an example of pseudo-likelihood estimates, which satisfy the very special property that they coincide with the corresponding MLE in the original Rasch model. Since the SSB model is misspecified, this known equality allows us to attribute the well-known bias and inconsistency of the JMLE to this misspecification, rather than to the presence of incidental parameters. Concerning the estimated standard errors of the JMLE and the SSBE, they are shown to be equal for the difficulties, but not for the abilities. In this second case, an exact formula is provided, and it is supplemented by upper and lower bounds on their ratio. The sharpness of the upper bound is illustrated. The importance of these new relationships is primarily theoretical, but they may have some practical value as well. As suggested by Haberman (2004), the SSBE could at least be used as initial points for other estimation procedures, and it is useful to have an idea of precision to compare the new estimates with the SSBE.

References

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

Besag, J. (1975). Statistical analysis of non-lattice data. *The Statistician*, *24*, 179–195.

Blackwood, L.G., & Bradley, E.L. (1989). The equivalence of two methods of parameter estimation for the Rasch model. *Psychometrika*, *54*, 751–754.

Boomsma, A., van Duijn, M.A.J., & Snijders, T.A.B. (2001). *Essays on item response theory*. Lectures Notes in Statistics, Vol. 157. New York: Springer.

Fischer, G.H., & Molenaar, I.W. (1995). *Rasch models. Foundations, recent developments and applications*. New York: Springer.

Ghosh, M. (1995). Inconsistent maximum likelihood estimators for the Rasch model. *Statistical and Probability Letters*, *23*, 165–170.

Haberman, S.J. (2004). Joint and conditional maximum likelihood estimation for the Rasch model with binary responses. ETS Research Report RR-04-20, May 2004.

Holland, P. (1990). On the sampling theory foundations of item response theory models. *Psychometrika*, *55*, 577–601.

McCullagh, P., & Nelder, J.A. (1989). *Generalized linear models* (2nd ed.). New York: Chapman & Hall/CRC.

Mellenbergh, G.J., & Vijn, P. (1981). The Rasch model as a loglinear model. *Applied Psychological Measurement*, *5*, 369–376.

Molenaar, I.W. (1995). Estimation of item parameters. In: G.H. Fischer & I.W. Molenaar (Eds.), *Rasch models. Foundations, recent developments and applications* (Chap. 3). New York: Springer.

Perline, R., Wright, B.D., & Wainer, H. (1979). The Rasch model as additive conjoint measurement. *Applied Psychological Measurement*, *3*, 237–255.

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen: The Danish Institute for Educational Research. (Expanded edition, 1980, Chicago: The University of Chicago Press.)

Van der Linden, W., & Hambleton, R.K. (1997). *Handbook of modern item response theory*. New York: Springer.

Verhelst, N., & Molenaar, I. (1988). Logit-based parameter estimation in the Rasch model. *Statistica Neerlandica*, *42*, 273–295.

Manuscript received 14 MAR 2006
Final version received 12 FEB 2007
Published Online Date: 18 JUL 2007