KM 13(1984)
pag 131-147



# COMPUTING THE SECOND-ORDER DERIVATIVES OF THE SYMMETRIC FUNCTIONS IN THE RASCH MODEL

Paul G.W. Jansen*

*Abstract*

When applying conditional maximum likelihood procedures to estimate the item parameters of the Rasch model, computation of the so-called basic symmetric functions is necessary. This paper demonstrates that the second-order derivatives of the symmetric functions, computation of which is required for various purposes, can be rewritten as a simple function of the basic symmetric functions, of the first-order derivatives of the basic symmetric functions, and of the item parameters. As a consequence, the rather tedious explicit computation of the second-order derivatives of the symmetric functions may be avoided in practical applications of the Rasch model.

Additionally, the investigation requires the computation of a number of formulas which are of relevance for theoretical studies of the Rasch model.

---

*Afdeling Research
Rijks Psychologische Dienst
P.O. Box 20013
2500 EA The Hague
The Netherlands
Tel: 070 - 514001 tst. 111

## 1. Introduction

### 1.1. *The Rasch model*

Suppose N subjects respond to K dichotomous items. The probability of subject v responding positively to item i may be represented by the parametric probability function

$$p(+|v,i) = \frac{\theta_v \varepsilon_i}{1 + \theta_v \varepsilon_i} , \tag{1}$$

which is known as the *Rasch model*. Since (1) is monotonically increasing in $\theta$, $\theta_v$ may be conceived as representing the 'ability' of subject v in the context of intelligence items e.g., and correspondingly $\varepsilon_i$ stands for the 'easiness' of item i. Since $\varepsilon_i$ may be simply defined as the inverse $1/\theta_w$ of the ability of a subject w for whom $p(+|w,i)=\frac{1}{2}$, item response model (1) assumes one latent trait underlying the observed responses, viz. the $\theta$-dimension. Mostly, the item parameters are subject to the norming restraint $\prod_{i=1}^{K} \varepsilon_i = 1$; we will assume this to be the case in the sequel.

For an extensive introduction to latent trait theory the reader is referred to Molenaar (1982), who presents an overview of various unidimensional latent trait models, including the Rasch model.

Denoting the response of subject v to item i by the binary indicator variable $a_{vi}$ ($a_{vi}=1,0$), the likelihood of a single response $a_{vi}$ in the Rasch model (1) is

$$p(a_{vi}|\theta_v,\varepsilon_i) = \frac{1}{1 + \exp(\xi_v - \sigma_i)} * \exp(-\sigma_i * a_{vi}) * \exp(\xi_v * a_{vi}), \tag{2}$$

where $\xi_v = \ln(\theta_v)$ and $\sigma_i = -\ln(\varepsilon_i)$. Formula (2) shows that $p(a_{vi}|\theta_v,\varepsilon_i)$ is a member of the one-parameter exponential family when the likelihood is considered as a function of either $\theta_v$ or $\varepsilon_i$ only. Therefore, $\sum_i a_{vi} =: a_{v0}$ is a sufficient statistic for $\theta_v$, and $\sum_v a_{vi} =: a_{0i}$ is a sufficient statistic for $\varepsilon_i$ (Mood, Graybill & Boes, 1974, 326).

The property of sufficiency makes it possible to use the procedure of *conditional maximum likelihood estimation* (CML) in the Rasch model. Essentially, in the CML the number of parameters is reduced by conditioning the likelihood of the data on the sufficient statistics for the parameters to be eliminated. For instance, when the aim is to estimate the item parameters $\varepsilon_i$, the unknown $\theta_v$ may be 'conditioned away' by replacing it by the observable $a_{v0}$. Since the statistic $a_{v0}$ is sufficient for the parameter $\theta_v$, no relevant information is lost in the CML approach to estimation. To estimate the $\underline{\varepsilon}=:(\varepsilon_1,\ldots,\varepsilon_i,\ldots,\varepsilon_K)$ the unconditional likelihood $L(\underline{a}_v|\theta_v,\underline{\varepsilon})$ of the data vector $\underline{a}_v=:(a_{v1},\ldots,a_{vi},\ldots,a_{vK})$ of a subject v and the conditional likelihood $L(\underline{a}_v|a_{v0},\underline{\varepsilon})$ of this data vector are equally useful.

Conditional estimation has both statistical and measurement-theoretical advantages. Statistical advantages are consistency of the estimators (Andersen, 1973b), and the existence of model tests with known and desirable properties (Andersen, 1973a; Van den Wollenberg, 1979, 1982a,b; Gustafsson, 1980b). The measurement-theoretical advantage of the Rasch model was labelled 'specific objectivity' by Rasch (1966, 1977); for a description the reader is referred to Molenaar (1982, 17-18) in the present journal.

In the remainder of this study the focus will be on the estimation of the item parameters $\varepsilon_i$, given the sufficient statistics $a_{v0}$ for the subject parameters $\theta_v$. To simplify notation, $a_{v0}$ will be written as $r_v$ in the sequel; $r_v$ represents the number of items responded to positively by subject v. In this case the estimation equation for the $\varepsilon_i$ is obtained by equating the sufficient statistics for $\varepsilon_i$ to their conditional expectations:

$$a_{0i} \mathrel{\hat{=}} E(A_{0i}|r_v,\underline{\varepsilon}) \qquad (i=1,\ldots,K), \tag{3}$$

which is a general property of exponential distributions (Fischer, 1974, 236; Andersen, 1980a,b).

Equation (3) can be rewritten as

$$a_{0i} \mathrel{\hat{=}} \sum_{v=1}^{N} E(A_{vi}|r_v,\underline{\varepsilon}) = \sum_{v=1}^{N} \pi_{r_v i} \qquad (i=1,\ldots,K). \tag{4}$$

Because of sufficiency, the conditional probability $\pi_{r_v i}$ only depends on the subject v through the raw score $r_v$. The probability is the same for all subjects having this marginal total. Therefore, the index v may be dropped and the conditional probability may be written as $\pi_{ri}$ in the sequel. The unconditional probability (i.e. (1)) will be written as $\pi_{vi}$ below. Thus

$$\pi_{r_v i} = \pi_{ri} =: \mathrm{Prob}(+|a_{v0}=r,i), \tag{5}$$

$$\pi_{vi} =: \mathrm{Prob}(+|v,i). \tag{6}$$

Although this notation may become ambiguous when numbers are substituted for letters, it seems to be standard in the literature on the Rasch model since Fischer (1974, 236, 238) (cf. e.g. Van den Wollenberg, 1979, 34, 73). Of course, in theoretical studies such as these, it is unambiguous.

1.2. *The symmetric functions*

The conditional probability $\pi_{ri}$ is defined as the ratio of $p(A_{vi}=1,r|\theta_v,\underline{\varepsilon})$ to $p(r|\theta_v,\underline{\varepsilon})$ as (5) shows. Performing this division, eventually the following expression is found (cf. Fischer, 1974, 221-232)

$$\pi_{ri} = \varepsilon_i \frac{\gamma_{r-1}^{(i)}}{\gamma_r}, \tag{7}$$

so that the estimation equations (4) become

$$a_{0i} = \sum_{r=1}^{K} \hat{\varepsilon}_i \frac{\hat{\gamma}_{r-1}^{(i)}}{\hat{\gamma}_r} \qquad (i=1,\ldots,K). \tag{8}$$

Eq. (7) formulates the regression of item score $A_{vi}$ on test score r, i.e. the item-test regression. The factor $\gamma_r$ in (7) is the *elementary symmetric function of order r* in the parameters $\varepsilon_1,\ldots,\varepsilon_K$. The function $\gamma_r$ is defined as

$$\gamma_r = \sum_{(r)} \prod_i \varepsilon_i^{a_{vi}}, \tag{9}$$

in which $\sum_{(r)}$ means that summation is over all $\binom{K}{r}$ possible ways of obtaining the raw score r in a K-item test. From definition (9)

$$\gamma_1 = \varepsilon_1^1\varepsilon_2^0\ldots\varepsilon_K^0 + \varepsilon_1^0\varepsilon_2^1\varepsilon_3^0\ldots\varepsilon_K^0 + \ldots + \varepsilon_1^0\ldots\varepsilon_{K-1}^0\varepsilon_K^1$$

$$= \varepsilon_1 + \varepsilon_2 + \ldots + \varepsilon_K \qquad (\tbinom{K}{1} \text{ terms}),$$

$$\gamma_2 = \varepsilon_1^1\varepsilon_2^1\varepsilon_3^0\ldots\varepsilon_K^0 + \varepsilon_1^1\varepsilon_2^0\varepsilon_3^1\varepsilon_4^0\ldots\varepsilon_K^0 + \ldots + \varepsilon_1^0\varepsilon_2^0\ldots\varepsilon_{K-1}^1\varepsilon_K^1$$

$$= \varepsilon_1\varepsilon_2 + \varepsilon_1\varepsilon_3 + \ldots + \varepsilon_{K-1}\varepsilon_K \qquad (\tbinom{K}{2} \text{ terms}),$$

$$\vdots$$

$$\gamma_K = \varepsilon_1\varepsilon_2\ldots\varepsilon_K \qquad (\tbinom{K}{K} \text{ terms}).$$

Finally, $\gamma_0$ is defined to be equal to 1.

The symmetric function of order r-1, $\gamma_{r-1}^{(i)}$ is defined analogously to the r-th order symmetric function (9), except that $\gamma_{r-1}^{(i)}$ does not contain the item parameter $\varepsilon_i$.

Estimation of the $\varepsilon_i$ occurs by solving (8) iteratively. The greatest problem in solving these equations lies in the computation of the symmetric functions in every iteration. Fischer (1974) presents some recursive formulas that lighten the burden of computing all the symmetric functions in every iteration:

$$\gamma_r = \varepsilon_i\gamma_{r-1}^{(i)} + \gamma_r^{(i)} , \tag{10}$$

$$r\gamma_r = \sum_{i=1}^{K} \varepsilon_i\gamma_{r-1}^{(i)} \quad . \tag{11}$$

From (10) it is easily obtained that

$$\frac{\partial}{\partial\varepsilon_i}\gamma_r = \frac{\partial}{\partial\varepsilon_i}(\varepsilon_i\gamma_{r-1}^{(i)} + \gamma_r^{(i)}) = \gamma_{r-1}^{(i)} , \tag{12}$$

which reveals that $\gamma_{r-1}^{(i)}$ is the first-order derivative of $\gamma_r$ with respect to $\varepsilon_i$. In this study, the focus will be on the second-order derivative of $\gamma_r$ with respect to $\varepsilon_i$ and $\varepsilon_j$.

Still, the computational problems are heavy. Recently, Gustafsson (1980a) devised an algorithm that seems to be both fast and accurate for rather large K. However, in addition to the computation of $\gamma_r$ and $\gamma_{r-1}^{(i)}$, application of the test statistics of Van den Wollenberg (1979, 1982a,b) and of Martin-Löf (described by Gustafsson, 1980b) requires the computation of the binary conditional probability $\pi_{rij}$, which represents the probability that a subject with raw score r will respond positively to both items i and j ($i \neq j$; this will be assumed throughout this study: i and j are different items). Since $\pi_{ri} = \text{Prob}(A_{vi}=1 \mid r) \geq \text{Prob}(A_{vi}=1 \mid a_{vj}=1, r)$ it follows that

$$\pi_{rij} \leq \pi_{ri}\pi_{rj}. \tag{13}$$

In the Rasch model, $\pi_{rij}$ is given by (cf. Fischer, 1974, 237)

$$\pi_{rij} = \varepsilon_i \varepsilon_j \frac{\gamma_{r-2}^{(i,j)}}{\gamma_r}. \tag{14}$$

The function $\gamma_{r-2}^{(i,j)}$ is defined analogously to $\gamma_{r-1}^{(i)}$. Applying (12) recursively, i.e. on $\gamma_{r-1}^{(i)}$, it follows that

$$\frac{\partial}{\partial \varepsilon_j} \gamma_{r-1}^{(i)} = \gamma_{r-2}^{(i,j)} \quad ,$$

which reveals $\gamma_{r-2}^{(i,j)}$ to be the second-order derivative of $\gamma_r$ to $\varepsilon_i$ and $\varepsilon_j$. As these second derivatives are contained in the Fisher -information matrix (Fischer, 1974, 235f), computation of $\gamma_{r-2}^{(i,j)}$ is also necessary to determine the Fisher -information matrix and subsequently the (co-)variances of the conditional estimators. In the same way, determination of $\gamma_{r-2}^{(i,j)}$ is necessary if the Hessian is to be computed, e.g. if Newton-Raphson algorithms are to be employed in solving (8).

Again, computing $\hat{\gamma}_{r-2}^{(i,j)}$ from estimates of the K item parameters is not easily done, as may be illustrated by the following quotations from Gustafsson (1980b):

"The greatest problem in solving the equations, which must

> be done iteratively, lies in efficiently and accurately computing the $\gamma_r$ and their first derivatives with respect to each of the items ($\gamma_{r-1}^{(i)}$) and sometimes also their second order derivatives with respect to the items two at a time ($\gamma_{r-2}^{(i,j)}$)"

(o.c., 211), and

> "When K is large, the test is quite tedious to compute since it requires the computation of K-1 matrix inversions as well as the second derivatives of the symmetric functions",

(o.c., 213).

In the next section, assuming $\varepsilon_i \neq \varepsilon_j$, $\pi_{rij}$ will be rewritten as a simple function of $\pi_{ri}$ and $\pi_{rj}$ using an analogue of the recursive relation (10). To my knowledge, this simplified formula, which may, in circumstances, save the explicit computation of the $\gamma_{r-2}^{(i,j)}$, is not generally known; Gustafsson (1982) has confirmed this. In section 3 a method is studied for computing $\pi_{rij}$ when $\varepsilon_i = \varepsilon_j$.

## 2. *A simple formula for the case* $\varepsilon_i \neq \varepsilon_j$

### 2.1. *Derivation*

Applying (10) to $\gamma_{r-1}^{(i)}$ and $\gamma_{r-1}^{(j)}$, it follows that

$$\gamma_{r-1}^{(i)} = \varepsilon_j \gamma_{r-2}^{(i,j)} + \gamma_{r-1}^{(i,j)}, \tag{15}$$

and

$$\gamma_{r-1}^{(j)} = \varepsilon_i \gamma_{r-2}^{(i,j)} + \gamma_{r-1}^{(i,j)}. \tag{16}$$

Subtracting (16) from (15):

$$(\varepsilon_j - \varepsilon_i)\gamma_{r-2}^{(i,j)} = \gamma_{r-1}^{(i)} - \gamma_{r-1}^{(j)}, \tag{17}$$

which if $\varepsilon_i \neq \varepsilon_j$ can be rewritten as

$$\gamma_{r-2}^{(i,j)} = \frac{\gamma_{r-1}^{(i)} - \gamma_{r-1}^{(j)}}{\varepsilon_j - \varepsilon_i} \quad (\varepsilon_i \neq \varepsilon_j). \tag{18}$$

Formula (18) shows that the second-order derivatives of the symmetric functions can be written as a simple function of the first-order derivatives and the respective parameters if $\varepsilon_i \neq \varepsilon_j$.

Substituting (18) for $\gamma_{r-2}^{(i,j)}$ in (14), it is found that

$$\pi_{rij} = \varepsilon_i \varepsilon_j \frac{\gamma_{r-1}^{(i)} - \gamma_{r-1}^{(j)}}{(\varepsilon_j - \varepsilon_i)\gamma_r}$$

$$= \frac{\varepsilon_j}{\varepsilon_j - \varepsilon_i} \varepsilon_i \frac{\gamma_{r-1}^{(i)}}{\gamma_r} - \frac{\varepsilon_i}{\varepsilon_j - \varepsilon_i} \varepsilon_j \frac{\gamma_{r-1}^{(j)}}{\gamma_r} \qquad \text{(apply (7))}$$

$$= \frac{\varepsilon_j \pi_{ri} - \varepsilon_i \pi_{rj}}{\varepsilon_j - \varepsilon_i} \qquad (\varepsilon_i \neq \varepsilon_j). \tag{19}$$

Equation (19) shows that the second-order probability $\pi_{rij}$ may be obtained as a simple function of the item parameters $\varepsilon_i$ and $\varepsilon_j$ and of the first order probabilities $\pi_{ri}$ and $\pi_{rj}$. As in the conditional approach

$$\pi_i = \frac{1}{N} \sum_{r=1}^{K} n_r \pi_{ri} \tag{20}$$

($n_r$: number of subjects in score group r), equation (19) may be generalized to the item probabilities $\pi_i$, $\pi_j$ and $\pi_{ij}$ as expected in the conditional approach:

$$\pi_{ij} = \frac{\varepsilon_j \pi_i - \varepsilon_i \pi_j}{\varepsilon_j - \varepsilon_i} . \tag{21}$$

Equations (19) and (20) also hold for the unconditional case, as is easily shown. Since in the latter approach $\pi_{vij} = \pi_{vi}\pi_{vj}$ because of the assumption of local stochastic independence (which is basic to latent trait models, cf. Molenaar, 1982, 7-8), it can be derived that

$$\pi_{vij} = \pi_{vi}\pi_{vj} \qquad \text{(apply (1))}$$

$$= \frac{\theta_v \varepsilon_i}{1+\theta_v \varepsilon_i} * \frac{\theta_v \varepsilon_j}{1+\theta_v \varepsilon_j}$$

$$= \frac{\theta_v \varepsilon_i \varepsilon_j}{(\varepsilon_j - \varepsilon_i)(1+\theta_v \varepsilon_i)} - \frac{\theta_v \varepsilon_i \varepsilon_j}{(\varepsilon_j - \varepsilon_i)(1+\theta_v \varepsilon_j)}$$

$$= \frac{\varepsilon_j \pi_{vi} - \varepsilon_i \pi_{vj}}{\varepsilon_j - \varepsilon_i} \quad (\varepsilon_i \neq \varepsilon_j). \tag{22}$$

However, since the CML estimation procedure has a number of important advantages as explained in section 1.1, concentration will almost invariably be on the formulas (18)-(21) in the sequel.

2.2. *Discussion*

As was indicated in section 1.2, formulas (18)-(21) could not be found in the literature on the Rasch model. A formula that resembles (21) is the following equivalent of Fischer's equation (13.3.6) (Fischer, 1974, 217)

$$\frac{\varepsilon_i}{\varepsilon_j} = \frac{\pi_i - \pi_{ij}}{\pi_j - \pi_{ij}} \tag{23}$$

which, in fact, motivated the present study. It seems that Fischer did not realize the implications of (23), which is equivalent to (21), with respect to a drastic simplification of the symmetric functions.

In the remainder of this section, it is discussed whether and how (19) and (21) may be used for actually computing $\pi_{rij}$.

Because of the conditional estimation equation (3), $\hat{\pi}_i$ will always be equal to the observed $a_{0i}/N =: p_i$ when estimating the item parameters of the Rasch model. To apply (21) therefore, $\pi_i$ and $\pi_j$ may be replaced by their respective observed counterparts and $\varepsilon_i$ and $\varepsilon_j$ by their estimated values.

Computation of (19) is a bit more tedious since $\pi_{ri}$ and $\pi_{rj}$ will have to be estimated from the parameters estimates $\hat{\varepsilon}_i$ (i=1,...,K) by means of (7). In general, $\hat{\pi}_{ri}$ will be unequal to the observed $p_{ri}$; in fact, the $Q_1$ test statistic of Van den Wollenberg (1979, 1982b) is based on the comparison of observed $p_{ri}$ to expected $\hat{\pi}_{ri}$.

However, in the case of the $Q_2$ test statistic of Van den Wollenberg (o.c.) computation becomes relatively simple. Since in this procedure the item parameters are reestimated for every subject group r (i.e. $\varepsilon_{ri}$ is estimated instead of $\varepsilon_i$), the analogue of (3) applies with $p_{ri}$ instead of $p_i$ and $\varepsilon_{ri}$ instead of $\varepsilon_i$. In such a case, $\hat{\pi}_{ri}$ is of necessity equal to $p_{ri}$, and (19) can be computed from the observed $p_{ri}$, $p_{rj}$ and estimated $\hat{\varepsilon}_{ri}$, $\hat{\varepsilon}_{rj}$. Thus, application of (19) is pre-eminently feasible in the computation of the $Q_2$ test statistic.

To illustrate the discussion, two examples of applying (19) are presented. Below, $n_r$ denotes the number of subjects in score group r and $n_{ri}$ the number of positive responses to item i in score group r; i=1,2,3 in the examples. Note that it does not matter whether the 3 items comply with the Rasch model, since (18)-(21) only function as computational 'shortcuts'. Furthermore, in the comparisons below the algorithm for explicitly computing $\gamma_{r-2}^{(i,j)}$ is assumed to be perfectly accurate (which might not always be realistic).

*First example*

Suppose r=5, $n_r$=23, $n_{r1}$=4, $n_{r2}$=20, $n_{r3}$=22. Then $p_{r1}$=4/23= .1739, $p_{r2}$= .8696, $p_{r3}$= .9565. Conditional maximum likelihood estimation in this score group yielded $\hat{\varepsilon}_{r1}$= .0342, $\hat{\varepsilon}_{r2}$= .3395, $\hat{\varepsilon}_{r3}$=1.0517. Using explicit computation of the symmetric functions, i.e. applying (14), the following expected probabilities were found: $\hat{\pi}_{r12}$= .0961, $\hat{\pi}_{r13}$= .1477, $\hat{\pi}_{r23}$= .8284. Application of (19) yields: $\hat{\pi}_{r12}$= .0960, $\hat{\pi}_{r13}$= .1476, $\hat{\pi}_{r23}$= .8282. Differences between the two kinds of estimates are negligible.

*Second example*

In this example $p_{r1} \approx p_{r2}$. Suppose r=2,3,4 i.e. the combination of the score groups 2,3 and 4. Furthermore suppose $n_r$=590, and $p_{r1}$= .0814, $p_{r2}$= .0915, $p_{r3}$= .1644. Estimation yielded $\hat{\varepsilon}_{r1}$= .5774, $\hat{\varepsilon}_{r2}$= .6546, $\hat{\varepsilon}_{r3}$=1.2473, $\hat{\pi}_{r12}$= .0053, $\hat{\pi}_{r13}$= .0098, $\hat{\pi}_{r23}$= .0110. Applying (19) it is found that $\hat{\pi}_{r12}$= .0059, $\hat{\pi}_{r13}$= .0099, $\hat{\pi}_{r23}$= .0110. Although for the item pair (1,2) the differences are somewhat larger than in the first example, they again appear at the fourth decimal place.

Thus, in these two examples $\pi_{rij}$ can be computed by means of (19). However, whether in all cases ($\varepsilon_i \neq \varepsilon_j$) $\pi_{rij}$ can be reliably computed by means of (19) is a question that only be answered both by empirical and theoretical error studies.

For instance, it may be expected that, as $\varepsilon_i$ and $\varepsilon_j$ are measured on an exponential scale, even very small differences between $\pi_{ri}$ and $\pi_{rj}$ will, in general, result in a difference $(\varepsilon_j - \varepsilon_i)$ that is large enough for the quotient (19) to be stably computable. On the other hand, suppose the computed values of $\pi_{ri}$ and $\pi_{rj}$ have additive errors of $\Delta_{ri}$ and $\Delta_{rj}$ respectively, and especially assume $\Delta_{ri}=-\Delta_{rj}$ (this situation was suggested by Charles Lewis). Substituting these in (19) gives

$$\frac{\varepsilon_j(\pi_{ri}+\Delta_{ri}) - \varepsilon_i(\pi_{rj}-\Delta_{ri})}{\varepsilon_j - \varepsilon_i} = \frac{\varepsilon_j\pi_{ri} - \varepsilon_i\pi_{rj}}{\varepsilon_j - \varepsilon_i} + \underbrace{\Delta_{ri}\left(\frac{\varepsilon_j+\varepsilon_i}{\varepsilon_j-\varepsilon_i}\right)}_{E} .$$

Even assuming the division can be carried out with unlimited precision, the error term E may be many times larger than $\Delta_{ri}$ when $\varepsilon_i$ and $\varepsilon_j$ are large relative to their difference. So in this case $\hat{\pi}_{rij}$ as computed by (19) from $\hat{\pi}_{ri}$ and $\hat{\pi}_{rj}$ may be radically different from $\pi_{rij}$.

It should be remarked, however, that it is unclear how the explicit 'traditional' computation of $\pi_{rij}$ performs in this specific situation. In general, a study of the computational merits of (19) should in all cases be a comparative study of (19) to the traditional, explicit computation procedure.

## 3. *The case* $\varepsilon_i=\varepsilon_j$

When $\varepsilon_i=\varepsilon_j$, the quotient (18) is not defined. This section is devoted to this case. In section 3.1 both for the unconditional and the conditional cases formulas are presented by means of which the binary probabilities $\pi_{vij}$ and $\pi_{rij}$ are rewritten as functions of the corresponding probabilities $\pi_{vi}$ or $\pi_{ri}$, and the item parameters. It appears however that these formulas are not suited for practical applications. Therefore, concentrating on the conditional case exclusively, in the next section an algorithm is presented by means of which $\pi_{rij}$ might be computed when $\varepsilon_i=\varepsilon_j$. Note that this extensive treatment of the case $\varepsilon_i=\varepsilon_j$ is of practical importance since in applications the estimates $\hat{\varepsilon}_i$ and $\hat{\varepsilon}_j$ are equal when $a_{0i}=a_{0j}$, $a_{0i}$ being sufficient for $\varepsilon_i$ in the Rasch model.

### 3.1. *Formulas*

#### 3.1.1. *The unconditional case*

When $\varepsilon_i=\varepsilon_j$, $\pi_{vij}$ is given by the limit

$$\lim_{\varepsilon_i\to\varepsilon_j} \pi_{vij} . \tag{24}$$

In the Rasch model, this limit is equal to (substituting (22) in (24))

$$\lim_{\varepsilon_i \to \varepsilon_j} \frac{\varepsilon_j \pi_{vi} - \varepsilon_i \pi_{vj}}{\varepsilon_j - \varepsilon_i} . \tag{25}$$

Applying l'Hopital's limit rule and assuming $\pi_{vj}$ to be independent of $\varepsilon_i$, i.e. assuming an unconditional approach, (25) can be rewritten as

$$\lim_{\varepsilon_i \to \varepsilon_j} \frac{\frac{\partial}{\partial \varepsilon_i}(\varepsilon_j \pi_{vi} - \varepsilon_i \pi_{vj})}{\frac{\partial}{\partial \varepsilon_i}(\varepsilon_j - \varepsilon_i)} = \pi_{vj} - \varepsilon_j \pi'_{vi}, \tag{26}$$

in which

$$\pi'_{vi} =: \frac{\partial}{\partial \varepsilon_i} \pi_{vi} . \tag{27}$$

Since in the unconditional case $\pi'_{vi} = \pi'_{vj}$ when $\varepsilon_i = \varepsilon_j$, (26) can be reformulated as

$$\pi_{vj} - \varepsilon_j \pi'_{vj} . \tag{28}$$

Formula (28) represents a formulation of the binary $\pi_{vij}$ into a function of first-order probabilities and of the item parameters. It is not of much practical importance, since $\pi_{vij} = \pi_{vi} \pi_{vj}$ in the unconditional approach. It is easily proved that for the total group of N subjects the analogue of (26) holds:

$$\lim_{\varepsilon_i \to \varepsilon_j} \pi_{ij} = \pi_j - \varepsilon_j \pi'_j . \tag{29}$$

Since $\pi_i$ as defined in (20) is a monotonically increasing function of $\varepsilon_i$, (29) implies that $\pi_{ij} < \pi_i$ $(\varepsilon_i = \varepsilon_j)$, as it of course should be. We shall not bother about the computation of (29) in practice, since it is only the conditional approach to estimation in the Rasch model that yields the advantages discussed in section 1.1. Still, (29) may be of use in theoretical studies; for an example we refer to Jansen (1983, ch.3).

### 3.1.2. *The conditional case*

Since in the conditional approach $\pi_{rj}$ depends on $\pi_i$ as (7) shows, it is not possible to transfer the unconditional result (28) to this situation. The derivative of $\pi_{rj}$ to $\varepsilon_i$ is equal to (using (7) and (12))

$$\frac{\partial}{\partial\varepsilon_i}\pi_{rj} = \frac{\partial}{\partial\varepsilon_i}\varepsilon_j\frac{\gamma^{(j)}_{r-1}}{\gamma_r}$$

$$= \varepsilon_j\frac{\gamma_r\gamma^{(i,j)}_{r-2} - \gamma^{(j)}_{r-1}\gamma^{(i)}_{r-1}}{\gamma^2_r}$$

$$= \varepsilon_j\frac{\gamma^{(i,j)}_{r-2}}{\gamma_r} - \varepsilon_j\frac{\gamma^{(i)}_{r-1}\gamma^{(j)}_{r-1}}{\gamma^2_r}$$

$$= \frac{\pi_{rij} - \pi_{ri}\pi_{rj}}{\varepsilon_i}, \tag{30}$$

which implies

$$\pi_{rij} = \pi_{ri}\pi_{rj} + \varepsilon_i\frac{\partial}{\partial\varepsilon_i}\pi_{rj}. \tag{31}$$

When $\varepsilon_i=\varepsilon_j$, (31) can be rewritten as

$$\pi_{rij} = \pi^2_{rj} + \varepsilon_j\frac{\partial}{\partial\varepsilon_i}\pi_{rj} \quad (\varepsilon_i=\varepsilon_j). \tag{32}$$

Note that (31) implies (using (13))

$$\frac{\partial}{\partial\varepsilon_i}\pi_{rj} < 0. \tag{33}$$

Inequality (33) implies that $\pi_{rj}$ is a monotonically decreasing function of $\varepsilon_i$, which agrees with intuition: The probability that a subject with score r (r fixed) will respond positively to item j will get smaller when some other item i in the test gets easier. For, the same raw score indicates a higher ability when i is difficult than when i is easy.

When $\frac{\partial}{\partial\varepsilon_i}\pi_{rj}$ is small over a whole range of values of $\varepsilon_i$, $\pi_{rj}$ is relatively flat when considered as a function of $\varepsilon_i$, and increases or decreases in $\varepsilon_i$ will not affect $\pi_{rj}$ very much. Thus, the actual range of values of the derivative of $\pi_{rj}$ to $\varepsilon_i$ over an acceptable range of values for $\varepsilon_i$, may be considered as indicating the dependence of $\pi_{rj}$ on item i. Locally, $\frac{\partial}{\partial \varepsilon_i}\pi_{rj}$ indicates to what extent $\pi_{rj}$ is affected by small shifts of $\varepsilon_i$. Thus, generally, the derivative of $\pi_{rj}$ with respect to $\varepsilon_i$ indicates the dependency of $\pi_{rj}$ on another item i.

Also, from (30)

$$\frac{\varepsilon_i}{\varepsilon_j} = \frac{\frac{\partial}{\partial \varepsilon_j}\pi_{ri}}{\frac{\partial}{\partial \varepsilon_i}\pi_{rj}}\,, \tag{34}$$

which gives some insight into the interpretation of the 'easiness'-scale on which the Rasch model parameters $\varepsilon_i$ are measured. As was seen above, $\frac{\partial}{\partial \varepsilon_i}\pi_{rj}$ indicates the dependency of $\pi_{rj}$ on the presence of another item i. Therefore, when $\varepsilon_i<\varepsilon_j$ in (34) item i affects $\pi_{rj}$ more than item j affects $\pi_{ri}$.

Eq. (32) is not of much use for the computation of $\pi_{rij}$ $(\varepsilon_i=\varepsilon_j)$ in practical applications. Therefore in the next section an algorithm will be studied for computing $\pi_{rij}$ $(\varepsilon_i=\varepsilon_j)$ which avoids the computation of the second-order derivatives of the symmetric functions and which seems practically feasible.

## 3.2. *An algorithm for the conditional case*

The algorithm to be discussed below presupposes that (19) works acceptably well when $\varepsilon_i \approx \varepsilon_j$. In such a case, it may be possible to compute an estimate $\pi^{*}_{rij}$ of $\pi_{rij}$ when $\varepsilon_i=\varepsilon_j$ by raising $\varepsilon_j$ somewhat and lowering $\varepsilon_i$ somewhat, and by subsequently applying (19). Below, an algorithm is discussed in which $\varepsilon_j$ is replaced by $\Delta\varepsilon_j$ and $\varepsilon_i$ by $\varepsilon_i/\Delta$ $(\Delta>1)$. It will be demonstrated that this procedure yields theoretically *exact* values of $\pi_{rij}$.

Suppose $\varepsilon_i$ is replaced by $\varepsilon^{*}_i =: \varepsilon_i/\Delta$ and $\varepsilon_j$ by $\varepsilon^{*}_j =: \Delta\varepsilon_j$ $(\Delta>1)$. Since

$$\varepsilon^{*}_i\varepsilon^{*}_j \prod_{\substack{h=1\\ h\neq i,j}}^{K} \varepsilon_h = \prod_{h=1}^{K} \varepsilon_h = 1,$$

the parameters $\varepsilon_h$ (h=1,...,K; h≠i,j) may remain unchanged: renorming is not necessary. Now the relation between $\pi^*_{rij}$ and $\pi_{rij}$ is given by (using (14))

$$\pi_{rij} = \varepsilon_i \varepsilon_j \frac{\gamma^{(i,j)}_{r-2}}{\gamma_r}$$

$$= \varepsilon^*_i \varepsilon^*_j \frac{\gamma^{(i,j)}_{r-2}}{\gamma^*_r} \frac{\gamma^*_r}{\gamma_r}$$

$$= \frac{\gamma^*_r}{\gamma_r} \pi^*_{rij}. \qquad (35)$$

All terms at the right side of (35) can be computed without using second-order derivatives of symmetric functions. The basic symmetric function $\gamma_r$ can be readily computed; in fact, its value will be known from the phase of estimating the original item parameters $\varepsilon_h$ by means of (8). The $\pi^*_{rij}$ follows from $\varepsilon^*_i$, $\varepsilon^*_j$, $\pi^*_{ri}$, and $\pi^*_{rj}$ by means of (19). This implies that $\pi^*_{ri}$ and $\pi^*_{rj}$ have to be computed, which is possible by application of (7). Using (7) to compute $\pi^*_{ri}$ and $\pi^*_{rj}$ from $\varepsilon^*_i$ and $\varepsilon^*_j$, and $\varepsilon_h$ (h=1,...,K; h≠i,j) implies that $\gamma^*_r$ will be computed. Thus, all terms at the right side of (35) can be computed from zero-order or first-order derivatives of the symmetric functions.

In the same way $\gamma^{(i,j)}_{r-2}$ can be computed (using (18))

$$\gamma^{(i,j)}_{r-2} = \gamma^{*(i,j)}_{r-2}$$

$$= \frac{\gamma^{*(i)}_{r-1} - \gamma^{*(j)}_{r-1}}{\varepsilon^*_j - \varepsilon^*_i}$$

$$= \frac{\gamma^{*(i)}_{r-1} - \gamma^{*(j)}_{r-1}}{\Delta\varepsilon_j - \varepsilon_i/\Delta}$$

$$= \frac{\Delta}{\Delta^2-1} \frac{\gamma^{*(i)}_{r-1} - \gamma^{*(j)}_{r-1}}{\varepsilon_j} . \qquad (36)$$

Of course, $\gamma^{*(i)}_{r-1}$ and $\gamma^{*(j)}_{r-1}$ can be readily computed.

Formulas (35) and (36) can be applied in situations in which $\varepsilon_i=\varepsilon_j$. Use of these formulas requires the computation of $\gamma^{*(i)}_{r-1}$, $\gamma^{*(j)}_{r-1}$, $\gamma^*_r$, and $\gamma_r$, i.e. of $\gamma_r$, $\gamma^*_r$, and of first-order derivatives of $\gamma^*_r$.

When $\Delta$ is chosen in an appropriate way, computation of the quotient $\gamma^*_r/\gamma_r$ in (35) may even not be necessary because the quotient approximates 1 to a satisfactory degree. In the 1983 version of the computer program *RADI* (Raaijmakers & Van den Wollenberg, 1979) the term $\gamma^*_r/\gamma_r$ is neglected in (35); it appeared that by setting $\Delta$ to 1.01 $\pi^*_{rij}$ was acceptably close to $\pi_{rij}$ for practical purposes (Van den Wollenberg, 1983).

Note however that the actual computation of $\gamma_r$ and $\gamma^*_r$ cannot be avoided: of $\gamma_r$ because of the necessity of solving the estimation equation (8), and of $\gamma^*_r$ because of the necessity of computing $\pi^*_{ri}$ and $\pi^*_{rj}$ (in order to compute $\pi^*_{rij}$ by means of (19)). That being the case, one may as well compute the quotient $\gamma^*_r/\gamma_r$ and consequently compute an estimate of $\pi_{rij}$ by means of (35) that is, theoretically, exact.

## 4. *Conclusion*

The derivations of (18) and (19) in section 2.1 and of (35) and (36) in section 3.2 demonstrate that, at least in theory, it may be possible to compute $\pi_{rij}$ according to the Rasch model, both for the cases $\varepsilon_i\neq\varepsilon_j$ and $\varepsilon_i=\varepsilon_j$, and at the same time avoid the explicit computation of the second-order derivatives of the symmetric functions. This is of importance since the computation of these second-order derivatives is tedious and time-consuming. Still, the only empirical corroboration of the practical use of (19) is Van den Wollenberg (1983). Hopefully, this study will stimulate other researchers, either to develop other formulas that might suggest computational simplifications of the symmetric functions, or to test the practical relevance of (19), e.g. by means of simulation studies.

Besides their possible practical use, the various formulas presented in this paper may be of relevance to theoretical studies of the Rasch model. For instance, as far as the present author knows formulas (18),(19),(21) and (30)-(36) have not been derived before in the published literature on conditional estimation in the Rasch model. Still, more work should, and may, be done in this domain.

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

Andersen, E.B. *Conditional inference and models for measuring*. Copenhagen: Mental-Hygiejnisk Forlag, 1973. (b)

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

Andersen, E.B. Latent structure analysis. *Invited paper to European Meeting of the Psychometric Society*, Groningen, June, 1980. (b)

Fischer, G.H. *Einführung in die Theorie psychologischers Tests*. Bern: Huber, 1974.

Gustafsson, J.-E. A solution of the conditional estimation problem for long tests in the Rasch model for dichotomous items. *Educational and Psychological Measurement*, 1980, *40*, 377-385. (a)

Gustafsson, J.-E. Testing and obtaining fit of data to the Rasch model. *British Journal of Mathematical and Statistical Psychology*, 1980, *33*, 205-233. (b)

Gustafsson, J.-E. *Personal communication*, 1982.

Jansen, P.G.W. *Rasch analysis of attitudinal data* (doct. diss.). The Hague: Rijks Psychologische Dienst, 1983.

Molenaar, I.W. Mensen die het beter meten. Een inleiding tot de latente trek modellen. *Kwantitatieve Methoden*, 1982, *5*, 3-29.

Mood, A.M., Graybill, F.A., & Boes, D.C. *Introduction to the theory of statistics* (3rd ed.). Tokyo: McGraw-Hill, 1974.

Raaijmakers, M.H., & Wollenberg, A.L., van den. RADI: Program for the dichotomous Rasch model. *Program Bulletin*. Nijmegen: The University of Nijmegen, 1979.

Rasch, G. An informal report on a theory of objectivity in comparisons. *Proceedings of the NUFFIC international summer session in science at "Het Oude Hof"*, The Hague, 14-28 July, 1966.

Rasch, G. On specific objectivity: An attempt at formalizing the request for generality and validity of scientific statements. *Danish Yearbook of Philosophy*, 1977, *14*, 58-94.

Wollenberg, A.L., van den. *The Rasch model and time limit tests* (doct. diss.). Nijmegen: Studentenpers, 1979.

Wollenberg, A.L., van den. On the applicability of the $Q_2$ test for the Rasch model. *Kwantitatieve Methoden*, 1982, *3*, 30-55. (a)

Wollenberg, A.L., van den. Two new test statistics for the Rasch model. *Psychometrika*, 1982, *47*, 123-140. (b)

Wollenberg, A.L., van den. *Personal communication*, 1983.

*Ontvangen: 5-7-1983*