Pseudo-likelihood Approach to Item Response Theory Models: Theory and Applications

Por
Eduardo Rodríguez Faúndez

Tesis sometida como requisito para optar al grado de

Doctor en Estadística

Pontificia Universidad Católica de Chile,

Santiago, Chile

Diciembre 2009

PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE
DEPARTAMENTO DE ESTADÍSTICA

Los que abajo firman certifican haber leído y recomiendan a la Facultad de Matemáticas la aceptación de la Tesis titulada "**Pseudo-likelihood Approach to Item Response Theory Models: Theory and Applications**" de **Eduardo Rodríguez** como requerimiento para optar al grado de **Doctor en Estadística**.

10 de Diciembre 2009.

Supervisor de Tesis: ______________________________

Ernesto San Martín

Comité Examinador: ______________________________

Manuel Galea

____

Guido del Pino

____

Jorge González

A mi madre, padre, hermanos, mi amor
Carla y mi hijo Mariano.

Agradecimientos

He escuchado de mucha gente que hace ciencia, que un doctorado es solo el *comienzo* de una vida llena de misterios, de magia y de mucho aprendizaje, y creo que tienen razón, pero vamos por parte, el hoy para mi es el *término* de un ciclo que empezó hace muchos años con una primera formación de matemático y ahora dando un pequeño giro derivando en la estadística. En este largo peregrinar de mi vida académica he encontrado mucho apoyo, muy buenos amigos y gente que me ha enriquecido en varios ambitos, creo que todas estas personas e instituciones que quiero mencionar lo haré en forma segmentada:

Mi familia, un pilar fundamental, mi padres Elsa y Sergio son los que me enseñaron y formaron desde pequeño a empezar un camino en línea recta y me dijeron que había que sobrepasar cualquier obstáculo, a ellos le debo todo. Mis hermanos Yaqueline y Girman, siempre recibí un apoyo incondicional de ustedes, los quiero mucho. A mi amor Carla le agradezco muchas cosas, apoyo, comprensión, la amo mucho, mi hijo Mariano lo más lindo del mundo fue aquel angelito que me dio el último impulso para poder terminar esta tesis y además Jesus quien me da mucha alegría y cariño.

Mis Amigos, Mario Tagle, Cafú y Bombero es el trío con el que pasé largas jornadas de estudios y camaradería acompañados siempre de buenos víveres. La Paty y el Cojo mis compañeros y amigos quienes pasamos grandes momentos de estudios y amistad. A Hugo inolvidable amigo y compañero, muchas gracias por su apoyo.

Facultad, la Sra Tina, creo que fui su regalón, le agradezco mucho su apoyo, Don Oscar, ¡grande en la seguridad!, el profesor Guido del Pino quien me ayudó mucho en los momentos de inflexión de mi carrera. A mi tutor o como le digo yo mi Maestro, muchas gracias, la profesora Pilar Iglesias que me dio mucho apoyo en los primeros años que son los más difíciles y además fue la propulsora para que estudiara el doctorado y las secretarias Rosy, Sole y Miriam, a todos ellos un abrazo grande.

Financiamiento, el aporte monetario de la Universidad hacia mi persona en cuanto a Beca de Matricula y de Residencia fueron primordiales para dejar a un lado la preocupación del financiamiento de mis estudios por lo que le agradezco muy en especial a la Beca VRAID.

Proyectos e Instituciones, le doy los agradecimientos a MIDE-UC quien aportó con el acceso a la pruebas SIES y sus datos en el Capitulo 2 y 3, y al proyecto "Estrategias de medición y análisis de datos educacionales usando modelos con variables aleatorias" Proyecto VRAID-2008 (Ernesto San Martín and David Preiss).

En estos cortos párrafos me quedo mucha gente que debo de agradecer, pero lo haré en forma personal.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

In 1948, Neyman and Scott (1948) introduced a set of statistical models parameterized by two types of parameters: the *incidental* parameters and the *structural* parameters. The incidental parameters increase with the sample size and, consequently, can be viewed as individual characteristics of the statistical units. The structural parameters are parameters upon which all individuals (of the sample) depend and, consequently, their dimension does not depends on the sample size. As an example, consider the first statistical model analyzed by Neyman and Scott (1948):

$$X_{ij} \sim \mathcal{N}(\mu_i, \sigma^2) \quad j = 1, \dots, m, \;\; i = 1, \dots, n,$$

where m is fixed and $\{X_{ij} : 1 \leq i \leq n,\ 1 \leq j \leq m\}$ are mutually independent. In this case, the statistical model (or likelihood function) is indexed by $(\mu_1, \dots, \mu_n, \sigma^2) \in \mathbb{R}^n \times \mathbb{R}_+$; here, the incidental parameters are the μ_i's, whereas the structural parameter is σ^2. Statistical models involving incidental parameters are, therefore, characterized by having a parameter space which increases with the sample size. For this class of models, Neyman and Scott (1948) showed that the maximum likelihood estimator (MLE) of the structural parameters is inconsistent due to the presence of the incidental parameters. In the statistical literature, this problem is known as the *incidental problem*; see Lancaster (2000).

A relevant example of the incidental problem is the Rasch model, which is widely used in educational measurement and psychometrics. The Rasch model assumes that the probability that a person answers correctly an item depends on two factors: a person-specific factor and an item-specific factor; the incidental parameters correspond to the person-specific factor; see Rasch (1960b). For two items, Andersen (1980) proved that the item-specific factor estimate is inconsistent. Thereafter, Ghosh (1995) extended the proof to an arbitrary number of items, using an argument by contradiction.

The incidental problem opened two approaches to its solution. The first consists in factorizing the likelihood function (which is indexed by structural and incidental parameters) into two factors in such way that one of the factors only depends on the structural parameters; for details, see Pfanzagl (1993). This factorization can be performed if there exists a sufficient statistic for the incidental parameters (when the structural parameters are fixed). In the case of the Rasch model, this is possible because the total score of a person is a sufficient statistic for the person-specific factor (when the item-specific factors are fixed); see (Chapter 6 Andersen, 1980). A second approach consists in considering the incidental parameters as realization of an iid process. Typically, this process is considered as a latent process generating person-specific latent factors. The statistical model, which refers to the observable variables only, can be obtained after integrating out the latent variables; the parameters indexing this statistical model include the item-specific factors as well as the parameters indexing the latent distribution. Kiefer and Wolfowitz (1956) and Pfanzagl (1970) proved that the MLE of these parameters are consistent provided they are identified. In general, this condition is difficult to verify because the statistical model typically doesn't have a closed form and, consequently, the identification problem reduces to verify the invertibility of an integral operator.

This doctoral dissertation deals with the incidental problem in the context of educational measurement, more precisely when a test includes items with partial credit. A popular statistical model to analyze this kind of data is the so-called Partial Credit Model (PCM); see Master (1982).

These models can be considered as an extension of the Rasch model. Since the framework of the Rasch model (with person-specific effects specified as unknown parameters) creates an example of the incidental problem, it seems natural that this is inherited by the PCM. The question is how to verify such a conjecture. This dissertation focus its attention on this question and not only attempts to answer it, but also tries to explain why the incidental problem is produced. In this dissertation it is claimed that the reason is that the MLE of the structural parameters is equivalent to the MLE of the parameters of a *misspecified* statistical model, which is called *SSB-model*, and will be defined later. This material is discussed in Chapter 2.

For the Rasch model, there exist bias correction factors which successfully help to "repair" that inconsistency; see Haberman (1977) and Andersen (1980). Taking advantage of the similarity between a Rasch model and a PCM, Master and Wright (1997) show, through simulations studies, that such correction factors work in the PCM context. In this dissertation, the results of this simulation study are confirmed; to do it, specific SAS codes were produced to compute the MLE of the PCM and its SSB-version. It should be mentioned that the codes are extremely efficient from a computational point of view. A natural question is, therefore, the following: in which context the MLE of the parameters of the SSB-version of the PCM (and their corrections) are useful? To answer this question, a simulation study is performed to compare the MLE of the item-specific parameters of a PCM model obtained after integrating out the person-specific factors, with the item-specific factors estimated under the SSB-version of the PCM. It is shown that, when the distribution generating the person-specific factors is correctly specified, both estimators are very similar. Furthermore, in the case where the distribution generating the person-specific factors is misspecified , it is shown that the estimators of the item-specific factors obtained with the SSB-version of the PCM are quite similar to the true item-specific factors. A theoretical explanation for these findings is also provided. This material is also developed in Chapter 2.

Taking into account that the item-specific factors can efficiently be estimated using the SSB-version of the PCM, this dissertation illustrates their use in specific application of educational measurement. Thus, in Chapter 2, it is studied the structure of the item parameters of a Chilean mathematics test with items partially scored. In Chapter 3, it is shown how a PCM can be reduced to a Linear Logistic Test Model (LLTM); see Fischer (1973). These models are used to evaluate if the items of a test are answered following specific sub-tasks or cognitive operations. To illustrate this aspect, the cognitive content of a Chilean test (the SEPA test) was evaluated using the SSB-version of the LLTM. A specific SAS-code was developed, which is very efficient from a computational point of view. A similar application was developed using a Rasch Poisson Count Model; the details are provided in Appendix B.

The SAS-codes are gathered in the Appendices. Appendix A contains some matrix theory needed in the theoretical development of Chapter 2.

Chapter 2

A Pseudo likelihood Aproach to the Partial Credit Model

2.1 Introduction

2.1.1 Partial Credit Model: Model Specification and Parameter Interpretation

The Partial Credit Model (PCM) is a unidimensional parametric item response theory (IRT) model for the analysis of responses recorded in two or more ordered categories; see Master (1982) and Master and Wright (1997). For each item these ordered categories correspond to ordered levels of performance on each item (partial success) and thereby awards partial credit for each answer. The usual motivation for partial credit scoring is the hope that it will lead to a more precise estimate of a person's ability than a simple pass/fail score. Suppose, for instance, that a person is asked to solve the following problem (taken from Master, 1982): $\sqrt{7.5/0.3 - 16}$. The correct answer is obtained after computing $\sqrt{9}$; it corresponds to the highest level of performance. This is getting after mastering two additional levels: the first one corresponds to solve 7.5/0.3, which is equal to 25; the second one, which supposes to master the first level, corresponds to solve the difference $25-16$, which is equal to 9. Thus, the item categories, along with the corresponding partial scores, are the following:

Failed	......	0
$7.5/0.3 = 25$	......	1
$25 - 16 = 9$	......	2
$\sqrt{9} = 3$	......	3

The PCM is specified as follows: consider the responses of n persons to a sequence of k items, $I_1, \ldots, I_i, \ldots, I_k$. Each person may respond to item I_i in $m+1$ ($m \geq 1$) *ordered* categories. The response of person v to item I_i will be represented by a selection vector $\boldsymbol{x}'_{vi} = (x_{vi0}, \ldots, x_{vih}, \ldots, x_{vim})$, where $\boldsymbol{x}_{vi}$ is an observation from the random variable $\boldsymbol{X}_{vi}$ defined as follows: $X_{vih} = 1$ if the correct answer is in category h, and $X_{vih} = 0$ otherwise. The model assumes that, for each item, the subject chooses one and only one of the $m+1$ categories. The probability function of the PCM is, therefore, given by

$$p_{vih} \doteq P(X_{vih} = 1) = \frac{\exp\left(h\theta_v - \sum_{g=0}^{h} \beta_{ig}\right)}{\sum_{z=0}^{m} \exp\left(z\theta_v - \sum_{g=0}^{z} \beta_{ig}\right)}, \quad h = 0, 1, \ldots, m, \quad (2.1.1)$$

where θ_v is a person parameter representing the ability of person v, whereas β_{ih} correspond to the difficulty of category h of the item I_i. The model is completed by assuming that the $\{X_{vih} : v = 1, \ldots, n;\ i = 1, \ldots, k;\ h = 0, \ldots, m\}$ are mutually independent.

Following San Martín et al. (2009), the parameters of interest should be distinguished from the identified parameters. In the case of a PCM, the identified parameters correspond to the probabilities $\{p_{vih} : v = 1, \ldots, n;\ i = 1, \ldots, k;\ h = 0, \ldots, m\}$ because each X_{vih} is distributed according to a Bernoulli distribution of parameter p_{vih} and the mapping $p_{vih} \longmapsto \text{Bern}\,(p_{vih})$ is injective. The parameters of interest are $\{\theta_v, \beta_{ih} : v = 1, \ldots, n;\ i = 1, \ldots, k;\ h = 0, \ldots, m\}$. To identify the parameters of interest, an injective relationship between them and the identified parametrization should be established. The identified quotients $p_{vih}/(p_{vih} + p_{vi,h-1})$ satisfy the following identity (see Master, 1982):

$$\frac{p_{vih}}{p_{vih} + p_{vi,h-1}} = \frac{\exp(\theta_v - \beta_{ih})}{1 + \exp(\theta_v - \beta_{ih})} \qquad h = 1, \ldots, m. \tag{2.1.2}$$

Consequently, β_{i0} (for $i = 1, \ldots, k$) should be fixed at 0; if not, we would have parameters that are not related to the identified parameters. Furthermore, for each v and i, the differences $\theta_v - \beta_{ih}$ for $h = 1, \ldots, m$ are identified; therefore, a linear restriction is needed to identify (θ_v, β_{ih}) for $h = 1, \ldots, m$. In this chapter, the following identification restriction is used:

$$\beta_{i0} = 0 \text{ for } i = 1, , \ldots, k; \quad \beta_{11} = 0. \tag{2.1.3}$$

Under this restriction, there remains $n + k(m-1) - 1$ parameters of interest.

The identification restrictions (2.1.3) also allows us to statistically interpret the parameters of interest; see San Martín et al. (2009). As a matter of fact, the restriction $\beta_{11} = 0$ implies that

$$\theta_v = \ln\left(\frac{p_{v11}}{p_{v10}}\right), \qquad v = 1, \ldots, n. \tag{2.1.4}$$

Thus, what it is commonly called "ability of person v" corresponds to the logarithm of the ratio between his/her probability to correctly answer the standard item 1 in category 1 and the probability to correctly answer the standard item 1 in category 0. This means that if the individual characteristic $\theta_v > 0$ (respect. < 0), then his/her probability to correctly answer item 1 in category 1 is greater (respect. lesser) than his/her probability to correctly answer item 1 in category 0.

Similarly, for each item i and each category $h \geq 2$, what it is typically called "difficulty of item i in category h" corresponds to

$$\beta_{ih} = \theta_v - \ln\left(\frac{p_{vih}}{p_{vi,h-1}}\right) = \ln\left(\frac{p_{v11}}{p_{v10}} \cdot \frac{p_{vi,h-1}}{p_{vih}}\right). \tag{2.1.5}$$

Therefore, $\beta_{ih} > \beta_{11} = 0$ has a precise meaning in terms of probability of correctly answer items in specific categories, namely

$$\frac{p_{vi,h-1}}{p_{vih}} > \frac{p_{v10}}{p_{v11}}, \quad h = 2, \ldots, m.$$

Two remarks should be added. First, the statistical interpretation of the parameters θ_v and β_{ih} always involves two categories of an item. Secondly, abilities and difficulties are in the same scale, namely the logarithm of probability ratios; see equations (2.1.4) and (2.1.5). Thus, the simultaneous representation ability-difficulty has an explicit statistical meaning: $\beta_{ih} > \theta_v$ means that $p_{vi,h-1} > p_{vih}$, that is, the probability that person v correctly answers item i in category $h-1$ is greater than his/her probability to correctly answer the same item in category h. In applications this simultaneous representation is crucial, particularly in criterion-referenced measurements; for details, see Berk (1970); Livinston and Zieky (1989); Cizek (2001).

2.1.2 Estimation procedures in the context of IRT models

IRT models are typically estimated using three different approaches: joint maximum likelihood (JML), conditional maximum likelihood (CML) and marginal maximum likelihood (MML). The first two approaches consider the θ_v's as *parameters*. The JML-approach is the general maximum likelihood estimation method (MLE) applied to the estimation of the parameters of interest. In the context of the Rasch model, it has a number of drawbacks (see Embretson and Reise, 2000, pp. 209-210), one of them being the inconsistency of the difficulty parameters estimates as the number of persons grows (for details, see Andersen, 1980; Ghosh, 1995); as was mentioned in the introduction this is an example of the famous *incidental parameter problem* (see Neyman and Scott, 1948; Lancaster, 2000). Nevertheless, bias correcting factors have been proposed (see Haberman, 1977; Andersen, 1980), providing empirical evidence that the corrected estimators (denoted as BC-JMLE) work well for a large number of items. The CML-approach uses the fact that the sum score is a sufficient statistic for the ability parameter θ_v (when the difficulty parameters are "fixed"). This leads to factorize the likelihood into two factors: one corresponds to the conditional distribution given the sum score parameterized by the difficulty parameters only; the other one corresponds to the marginal distribution of the sum score parameterized by both the person parameters and the difficulty parameters. The difficulty parameters are estimated maximizing the conditional likelihood given the total score. The abilities can be estimated in a second step from the marginal distribution of the sum score after substituting the difficulty parameters by their estimates. The MML considers the abilities as mutually independent random variables with a common distribution F^{ϕ}, typically normal. In this case, model (2.1.1) is viewed as a conditional distribution given θ_v, and the estimation is focused on the difficulty parameters and the parameter ϕ indexing F. These parameters are consistently estimated provided they are identified by the observations (see, e.g., Kiefer and Wolfowitz, 1956). The abilities are finally estimated using an Empirical Bayes procedure. For details and references, see De Boeck and Wilson (2004) and Baker and Kim (2004).

2.1.3 Estimation procedures for the PCM

These approaches have also been implemented for the PCM (Master and Wright, 1997; Baker and Kim, 2004). In particular, the JML procedure is still widely used in practice in computer programs such as QUEST (Adams and Khoo, 1993) and WINSTEPS (Linacre and Wright, 2000). In this context, Bertoli-Barsotti (2005) provides a necessary and sufficient condition for the existence and uniqueness of the JMLE of both the person parameters and the difficulty parameters. Taking advantage on the structural similarity between the Rasch model and the PCM, the related literature widely accepts (without a formal proof) that the inconsistency of the JMLE for the Rasch model is inherited by the PCM. Accordingly, correction factors similar to that used for the Rasch model have been proposed for the PCM. Master and Wright (1997) shows, suggest that such correction factors work well when compared with the CML-estimator.

2.1.4 Purpose of this chapter

The first concern of this chapter is to explain why the JMLE of the PCM does not work correctly. It is actually shown that the JMLE of the PCM is equivalent to the MLE of a misspecified model, which is called *SSB (sum score based)-model*. Broadly speaking, the SSB-model is obtained from the PCM after replacing the person parameter θ_v by a proxy of it based on the sum score of person v (that is, the observed score obtained by person v when answering all the items). The model obtained in this way actually is a conditional model given the sum score; the incorrectness of the SSB-model comes from the fact that it is *explicitly* treated as a marginal model rather than as a conditional model. Following the terminology introduced by Besag (1974, 1975), the SSB-model corresponds to a pseudo-likelihood; the MLE obtained from the SSB-model will be denoted as SSBE. It is proved that JMLE = SSBE; the inconsistency of the JMLE is accordingly explained by the fact that it is equal to the MLE of an *incorrect model*, namely the SSB-model. Exact relationships between the estimated standard errors for the PCM and its SSB-version are also obtained. Thus, a complete statistical description of a pseudo-likelihood (in our case, the SSB-model) in terms of the original statistical model (in our case, the PCM) is provided.

Due to the inconsistency problems of the JMLE, PCM are also estimated using a MML-approach. This requires to specify the distribution generating the person abilities F^{φ}, typically a normal distribution $\mathcal{N}(0, \varphi^2)$. The statistical model –obtained after integrating out the person abilities– is indexed by the difficulty parameters β_{ih} and the parameter φ; we call this marginal model *structural PCM*. The person abilities are typically estimated using an Empirical Bayes procedure. In this context, the simultaneous representation of abilities and difficulties mentioned in Section 2.1.1 is meaningless although the corresponding estimations belong to the sample space (i.e. they are functions of the observations), and can accordingly be compared. Taking into account the relevance in applications of such a representation, it should be asked under which conditions the abilities and difficulties estimated in the context of a structural PCM can be considered simultaneously represented. This is the second concern of this chapter, which is discussed in Section 2.5.

This chapter is organized as follows: the SSB-formulation of the PCM is developed in Section 2.2. The equality between the JMLE and the SSBE is proved in Section 2.3, whereas the relationships between the corresponding standard errors are established in Section 2.4. Section 2.5 composes SSB-estimates with MML-estimates and suggest a way in which BC-SSB-estimates are useful.

2.2 SSB-formulation of the Partial Credit Model

Let $\boldsymbol{\theta}' = (\theta_1, \ldots, \theta_n)$ and $\boldsymbol{\beta}' = (\beta_1', \beta_2', \ldots, \beta_k')$, where $\beta_i' = (\beta_{i0}, \beta_{i1}, \ldots, \beta_{im})$, with $i = 1, \ldots, k$. It can easily be verified that the log-likelihood of the PCM is given by

$$l_{\text{JMLE}}(\boldsymbol{\theta}, \boldsymbol{\beta}) = \sum_{v=1}^{n} \theta_v \sum_{h=0}^{m} h x_{v+h} - \sum_{i=1}^{k} \sum_{h=0}^{m} \beta_{ih} \sum_{w=h}^{m} x_{+iw} - G_{\text{JMLE}}(\boldsymbol{\theta}, \boldsymbol{\beta}), \tag{2.2.1}$$

where $G_{\text{JMLE}}(\boldsymbol{\theta}, \boldsymbol{\beta}) = \sum_{i=1}^{k} \sum_{v=1}^{n} \log \left[\sum_{z=0}^{m} \exp\left(z\theta_v - \sum_{g=0}^{z} \beta_{ig} \right) \right]$. It follows that the sufficient statistic for (θ_v, β_{ih}) is given by

$$\left(\sum_{h=0}^{m} hx_{v+h}, \sum_{w=h}^{m} x_{+iw} \right).$$

From the exponential family theory, it is well known that all persons having the same total score $X_{v++} = \sum_{h=0}^{m} hx_{v+h}$ have the same estimation for their abilities. Thus, a total score $t \in \{0, \dots, mk\}$ can be considered as a proxy of the abilities of persons having a total score $X_{v++} = t$. The SSB-formulation of the PCM is based on this fact. More precisely, the SSB-model corresponding to the PCM is obtained by replacing the θ_v in (2.1.1) by a $\gamma_t \in \mathbb{R}$ when $v \in I_t = \{v : X_{v++} = t\}$, with $t \in \{0, \dots, mk\}$. Thus,

$$p_{vih} \doteq P(X_{vih} = 1) = \frac{\exp\left(h\gamma_t - \sum_{g=0}^{h} \beta_{ig}\right)}{\sum_{z=0}^{m} \exp\left(z\gamma_t - \sum_{g=0}^{z} \beta_{ig}\right)}, \quad h = 0, 1, \dots, m, \text{ for all } v \in I_t, \tag{2.2.2}$$

where γ_t represents a proxy of the ability of a person v who obtained a sum score $X_{v++} = t$. It is still assumed that $\{X_{vih} : v = 1, \dots, n, \ i = 1, \dots, k, \ , h = 0, \dots, m\}$ are mutually independent. This model is called *SSB-model*.

Let N_{tih} be a random variable indicating the number of persons with a sum score equal to t who correctly answered the step h for the item i; let also n_t be the number of persons with a sum score equal to t. Using a sufficiency reduction, model (2.2.2) can equivalently be rewritten as

$$\boldsymbol{N_{ti}}' \sim \text{Mult}\,(n_t, (p^t_{i0}, \ldots, p^t_{im})), \quad t = 0, \ldots, mk,\ i = 1, \ldots, k, \quad \mathop{\perp\!\!\!\perp}_{t,i} \boldsymbol{N'_{ti}}, \tag{2.2.3}$$

where $\boldsymbol{N_{ti}}' = (N_{ti0}, \ldots, N_{tim})$, $\mathop{\perp\!\!\!\perp}_{t,i} \boldsymbol{N_{ti}}'$ stands for the mutual independence of $\{\boldsymbol{N_{ti}}'\ :\ t = 0, \ldots, mk,\ i = 1, \ldots, k\}$, and for $t = 1, \ldots, mk-1$, $i = 1, \ldots, k$ and $h = 0, \ldots, m$,

$$p^t_{ih} \doteq p_{vih} \quad \text{when } v \in I_t, \tag{2.2.4}$$

namely the probability that a person with sum score equal to t correctly answers the step h for the item i. It can be noticed that in (2.2.3) and (2.2.4) the total score $t = 0$ and $t = mk$ have been excluded in order to avoid infinite estimates. In the SSB-method, these cases are associated with an absence of randomness, since all the x_{vih} are equal to 0 for $X_{v++} = 0$ and to 1 for $X_{v++} = mk$, respectively; they thus provide no information about the difficulty parameters β_{ih}.

Similarly to the PCM, the identification restrictions of the SSB-model are given by (2.1.3). Accordingly, the SSB-model involves $t + km - 1$ free parameters. Since $t < n$, the number of parameters to be estimated using the SSB-model is smaller than the quantity of parameter of the original PCM. These parameters are estimated using the corresponding log-likelihood, which is given by

$$l_{\text{SSBE}}(\boldsymbol{\gamma}, \boldsymbol{\beta}) = \sum_{t=1}^{mk-1} \gamma_t \sum_{h=0}^{m} h n_{t+h} - \sum_{i=1}^{k} \sum_{h=0}^{m} \beta_{ih} \sum_{w=h}^{m} n_{+iw} - G_{\text{SSBE}}(\boldsymbol{\gamma}, \boldsymbol{\beta}), \tag{2.2.5}$$

where $\boldsymbol{\gamma}' = (\gamma_1, \ldots, \gamma_t, \ldots, \gamma_{mk-1})$ and

$$G_{\text{SSBE}}(\boldsymbol{\gamma}, \boldsymbol{\beta}) = \sum_{t=1}^{mk-1} \sum_{i=1}^{k} n_t \log \left[\sum_{z=0}^{m} \exp\left(z\gamma_t - \sum_{g=0}^{z} \beta_{ig} \right) \right].$$

The sufficient statistic for (γ_t, β_{ih}) is, therefore, given by

$$\left(\sum_{h=0}^{m} h n_{t+h}, \sum_{w=h}^{m} n_{+iw} \right).$$

Let us finish this section pointing out that the γ_t's are not properly parameters because they are data dependent through of t, whereas the θ_v's in the PCM are actually incidental parameters. More precisely, for each t γ_t depends on the random variable X_{v++}, *but in the SSB-model* γ_t *is not longer treated as a random variable, but as a fixed (but unknown) constant.* Not only this fact shows that the SSB-model is a misspecified one, but also that the misspecification arises from ignoring the randomness involved in defining the groups I_t by the sum scores, and the resulting correlation between X_{vih} and the X_{v++}. The SSB version of PCM can be viewed as a pseudo-likelihood in the sense of Besag (1974, 1975).

2.3 Equality of the JMLE and the SSBE

In this section we prove that JMLE = SSBE. Both models (2.1.1) and (2.2.3) belong to the exponential family; therefore, the corresponding likelihood equations have a unique solution. This uniqueness property means that we need only to verify that the JMLE satisfies the SSB-likelihood equations. The conclusion follows from the uniqueness of the MLE. Let us start by examining the JML-equations:

JML-equations: Since (2.1.1) belongs to the exponential family, the equations are grouped into two sets: (a.1) $E\left(\sum_{h=0}^{m} hX_{v+h}\right) = \sum_{h=0}^{m} hx_{v+h}$, $v = 1, \ldots, n$; and (b.1) $E\left(\sum_{w=h}^{m} X_{+iw}\right) = \sum_{w=h}^{m} x_{+iw}$, $i = 1, \ldots, k$, $h = 0, \ldots, m$. These equations can respectively be rewritten as

$$\sum_{h=0}^{m} h \sum_{i=1}^{k} p_{vih}(\theta_v, \boldsymbol{\beta}) = \sum_{h=0}^{m} hx_{v+h}, \quad v = 1, \ldots, n. \tag{2.3.1}$$

$$\sum_{w=h}^{m} \sum_{v=1}^{n} p_{viw}(\theta_v, \beta_{iw}) = \sum_{w=h}^{m} x_{+iw}, \quad i = 1, \ldots, k; \;\; h = 0, \ldots, m. \tag{2.3.2}$$

It follows that $\widehat{\theta}_v$ is the unique solution of $\sum_{h=0}^{m} h \sum_{i=1}^{k} p_{vih}(\eta, \widehat{\boldsymbol{\beta}}) = \sum_{h=0}^{m} hx_{v+h}$ for all $\eta \in \mathbb{R}$. Denoting by $\widetilde{\theta}_t = \widetilde{\theta}_t(\widehat{\boldsymbol{\beta}})$ the unique real number satisfying

$$\sum_{h=0}^{m} h \sum_{i=1}^{k} p_{vih}(\widetilde{\theta}_t(\widehat{\boldsymbol{\beta}}), \widehat{\boldsymbol{\beta}}) = t, \quad t \in \mathcal{T} = \{t | 0 < t < mk\}, \tag{2.3.3}$$

$\widehat{\theta}_v$ and $\widetilde{\theta}_t$ are linked through the following relation:

$$\widehat{\theta}_v = \widetilde{\theta}_t(\widehat{\boldsymbol{\beta}}) \qquad \text{if } v \in I_t = \left\{ v \,\middle|\, \sum_{h=0}^{m} hX_{v+h} = t \right\}. \tag{2.3.4}$$

Substituting the JMLE in (2.3.1), we obtain that

$$\sum_{w=h}^{m} \sum_{t \in \mathcal{T}} n_t p_{vih}(\widetilde{\theta}_t(\widehat{\boldsymbol{\beta}}), \widehat{\beta}_{ih}) = \sum_{w=h}^{m} x_{+iw} \quad i = 1, \ldots, k; \;\; h = 0, \ldots, m. \tag{2.3.5}$$

Since the likelihood equations have a unique solution, the JMLE for the PCM is fully characterized by equations (2.3.3), (2.3.4) and (2.3.5).

SSB-likelihood equations: Since (2.2.4) belongs to the exponential family, the likelihood equations are grouped into two sets: (a.2) $E\left(\sum_{h=0}^{m} hN_{t+h}\right) = \sum_{h=0}^{m} hn_{t+h}$, $t \in \mathcal{T} = \{t|0 < t < mk\}$; and (b.2) $E\left(\sum_{w=h}^{m} N_{+iw}\right) = \sum_{w=h}^{m} n_{+iw}$, $i = 1, \ldots, k$, $h = 0, \ldots, m$. These equations can respectively be rewritten as follows:

$$n_t \sum_{h=0}^{m} h \sum_{i=1}^{k} p_{ih}^{t}(\gamma_t, \boldsymbol{\beta}) = \sum_{h=0}^{m} hn_{t+h}, \quad t \in \mathcal{T}. \tag{2.3.6}$$

$$\sum_{w=h}^{m} \sum_{t \in \mathcal{T}} n_t p_{iw}^{t}(\gamma_t, \beta_{ih}) = \sum_{w=h}^{m} n_{+iw}, \quad i = 1, \ldots, k, h = 0, \ldots, m. \tag{2.3.7}$$

Equality between the SSBE and the JMLE: To verify that $\widetilde{\theta}_t(\widehat{\boldsymbol{\beta}})$ and $\widehat{\beta}_{ih}$ are solutions of (2.3.6) and (2.3.7) we only need to use two simple facts:

$$\text{(i)} \sum_{h=0}^{m} hn_{t+h} = tn_t; \qquad \text{(ii)} \sum_{w=h}^{m} n_{+ih} = \sum_{w=h}^{m} x_{+ih}.$$

The second fact is true by equations (2.3.5) and (2.3.7); the first one comes from the following equality:

$$\sum_{h=0}^{m} hn_{t+h} = \sum_{v \in I_t} \sum_{i=1}^{k} \sum_{h=0}^{m} hx_{vih} = \sum_{v \in I_t} \sum_{h=0}^{m} hx_{v+h} = \sum_{v \in I_t} t = tn_t$$

□

This ends the proof.

Remark 1. It is important to remark that, when computing the expectations in equations (a.2) and (b.2), it is explicitly ignored that the total score t is a random variable. It is rather assumed to be a constant; this fact precisely leads to consider the SSB-model as a pseudo-likelihood; furthermore, when fitting the SSB-model, the γ_t 's are treated as constant, and not as random variables.

The SSBE is an unusual instance of pseudo-likelihood estimate, since it agrees exactly with the MLE for the original PCM. This equality implies that the SSBE does not eliminate the inconsistency of the $\widehat{\beta}_{ih}$, even though this cannot now be attributed to the presence of incidental parameters. The source of the inconsistency is that the SSBE is actually derived from the misspecified model (2.2.2). Reversing the argument, the inconsistency of the $\widehat{\beta}_{ih}$ in the PCM can be explained by the fact that the JMLE coincides with the MLE of a misspecified model.

2.4 Relationships between the standard errors of the JMLE and the SSBE

In which sense the equality of the point estimates established in Section 2.3 can be extended to the corresponding (asymptotic) standard errors? This question is motivated by the fact that the Fisher Information Matrices of both the PCM and its SSB-version are related. It is expected, therefore, that their inverses are related and, consequently, the (asymptotic) standard errors too. In this section we establish exact relationships between them using a technique sketched in del Pino et al. (2008). We motivate the kind of relationships we want to establish through an example. A relevant aspect is that these inverse matrices are theoretically obtained without imposing additional hypotheses leading to a simplification of the structure of the Fisher Information Matrices (as done, for instance, by Baker and Kim, 2004). To derive the main results some mathematical machinery is developed, which is explained in Appendix A.

2.4.1 Relationships between the information matrices

The (asymptotic) standard errors of SSBE and JMLE are the square root diagonal elements of the inverses of the corresponding information matrices. Since (2.1.1) and (2.2.2) are generalized linear models, their information matrices coincide with the negative Hessian of the corresponding log-likelihoods (McCullagh and Nelder, 1989). When evaluated at the MLE we denote these matrices by $\mathcal{I}_{\text{JMLE}}$ and $\mathcal{I}_{\text{SSBE}}$, respectively. Under the identification restriction (2.1.3), the $\mathcal{I}_{\text{JMLE}}$ is given by

$$[\mathcal{I}_{\text{JMLE}}]_{vv} = \sum_{i=1}^{k} \frac{\sum_{z=1}^{m} z^2 \exp\left(z\widehat{\gamma}_t - \sum_{g=1}^{z} \widehat{\beta}_{ig}\right) + \sum_{w=1}^{m-1}\sum_{j=1}^{m-w} w^2 \exp\left((2j+w)\widehat{\gamma}_t - 2\sum_{q=1}^{j} \widehat{\beta}_{iq} - \sum_{r=j+1}^{w+j} \widehat{\beta}_{ir}\right)}{\left(\sum_{z=0}^{m} \exp\left(z\widehat{\gamma}_t - \sum_{g=1}^{z} \widehat{\beta}_{ig}\right)\right)^2}, \quad v \in I_t.$$

$$[\mathcal{I}_{\text{JMLE}}]_{vv'} = 0, \quad v \neq v'.$$

$$[\mathcal{I}_{\text{JMLE}}]_{ih,v} = \frac{\sum_{z=h}^{m} z \exp\left(z\widehat{\gamma}_t - \sum_{g=1}^{z} \widehat{\beta}_{ig}\right) + \sum_{j=1}^{h-1}\sum_{w=h-j}^{m-j} w^2 \exp\left((2j+w)\widehat{\gamma}_t - 2\sum_{q=1}^{j} \widehat{\beta}_{iq} - \sum_{r=j+1}^{w+j} \widehat{\beta}_{ir}\right)}{\left(\sum_{z=0}^{m} \exp\left(z\widehat{\gamma}_t - \sum_{g=1}^{z} \widehat{\beta}_{ig}\right)\right)^2},$$

$$i = 1,\ldots,k; h = 1,\ldots,m, \quad \text{excluding the pair}(i=1, h=1).$$

$$[\mathcal{I}_{\text{JMLE}}]_{ih,ih'} = \sum_{t\in\mathcal{T}} n_t \frac{\left(\sum_{w=h'}^{m} \exp\left(w\widehat{\gamma}_t - \sum_{j=1}^{w} \widehat{\beta}_{ij}\right)\right)\left(1 + \sum_{w=1}^{h-1} \exp\left(w\widehat{\gamma}_t - \sum_{j=1}^{w} \widehat{\beta}_{ij}\right)\right)}{\left(\sum_{z=0}^{m} \exp\left(z\widehat{\gamma}_t - \sum_{g=1}^{z} \widehat{\beta}_{ig}\right)\right)^2}, \quad i = 1,\ldots,k; h = 1,\ldots,m; h \leq h'.$$

$$[\mathcal{I}_{\text{JMLE}}]_{iz,i'z'} = 0, \quad i \neq i'. \tag{2.4.1}$$

Here, we defined $\sum_{w=1}^{0} D_w = 0$, where D_w is an algebraic expression in w. Performing similar computations for $\mathcal{I}_{\text{SSBE}}$ and comparing with (2.4.1) we obtain the following key relationships:

$$[\mathcal{I}_{\text{SSBE}}]_{tt} = n_t[\mathcal{I}_{JMLE}]_{vv}, \quad t = 1,\ldots,mk-1; v \in I_t.$$

$$[\mathcal{I}_{\text{SSBE}}]_{tt'} = 0 \quad, t \neq t', \quad t = 1,\ldots,mk-1.$$

$$[\mathcal{I}_{\text{SSBE}}]_{ih,t} = n_t[\mathcal{I}_{JMLE}]_{ih,v}, \quad i = 1,\ldots,k; h = 1,\ldots,m; t = 1,\ldots,mk-1; v \in I_t.$$

$$[\mathcal{I}_{\text{SSBE}}]_{ih,ih'} = [\mathcal{I}_{JMLE}]_{ih,ih'}, \quad i = 1,\ldots,k; h = 1,\ldots,m; h \leq h'.$$

$$[\mathcal{I}_{\text{SSBE}}]_{ih,i'h'} = 0, \quad i \neq i' \quad, i = 1,\ldots,k. \tag{2.4.2}$$

2.4.2 Relationships between the standard errors

Illustration

Let us start this section with the following illustration: consider $n = 10$ persons, $k = 5$ items and $m = 2$ categories. Consider the following pattern responses:

$$\begin{array}{llll} Y_1 = (0,2,1,1,0), & Y_2 = (1,0,1,1,1), & Y_3 = (2,2,1,1,1), & Y_4 = (2,2,2,0,2), \\ Y_5 = (1,2,1,2,1), & Y_6 = (1,0,1,0,0), & Y_7 = (0,1,1,0,0), & Y_8 = (1,0,0,2,0), \\ Y_9 = (1,1,1,0,1), & Y_{10} = (1,1,1,1,0). & & \end{array}$$

These patterns have not being simulated, but have been arbitrary selected; this does not represent any disadvantage since we are looking for *exact* relationships between the standard errors. As mentioned in Section 2.2, the SSBE can efficient be obtained if the data are represented as in Table 2.1. Using the identification restriction (2.1.3), the variance-covariance matrix $\mathcal{I}^{-1}_{\text{SSBE}}$ is automatically obtained through the PROC NLIN procedure detailed in Appendix C. Since there are five different total scores, the first five columns (and five rows) of $\mathcal{I}^{-1}_{\text{SSBE}}$ correspond to $\widehat{\gamma}_2, \widehat{\gamma}_3, \widehat{\gamma}_4, \widehat{\gamma}_7, \widehat{\gamma}_8$. Moreover, since we are considering $k = 5$ items with $h = 2$ steps, the second nine columns (and nine rows) correspond to $\widehat{\beta}_{12}, \widehat{\beta}_{21}, \ldots, \widehat{\beta}_{52}$. The corresponding Fisher Information Matrix evaluated at the MLE is the inverse of $\mathcal{I}^{-1}_{\text{SSBE}}$, namely $\mathcal{I}_{\text{SSBE}}$. Both matrices are shown on page 27.

Using the Fisher Information Matrix $\mathcal{I}_{\text{SSBE}}$, along with the information provided by the n_t, namely $n_2 = 2$, $n_3 = 1$, $n_4 = 4$, $n_7 = 2$ and $n_8 = 1$, equality (2.4.2) leads to obtain the Fisher Information Matrix $\mathcal{I}_{\text{JMLE}}$ corresponding to the JMLE; its inverse $\mathcal{I}^{-1}_{\text{JMLE}}$ corresponds to the variance-covariance matrix of the JMLE of the PCM. Both matrices are shown on page 28. These ma-

Table 2.1: Data matrix used for fitting the SSB-model $n = 10$, $k = 5$ and $h = 2$

t	i	n_t	n_{ti1}	n_{ti2}
2	1	2	1	0
2	2	2	1	0
2	3	2	2	0
2	4	2	0	0
2	5	2	0	0
3	1	1	1	0
3	2	1	0	0
3	3	1	0	0
3	4	1	0	1
3	5	1	0	0
4	1	4	3	0
4	2	4	2	1
4	3	4	4	0
4	4	4	3	0
4	5	4	2	0
7	1	2	1	1
7	2	2	0	2
7	3	2	2	0
7	4	2	1	1
7	5	2	2	0
8	1	1	0	1
8	2	1	0	1
8	3	1	0	1
8	4	1	0	0
8	5	1	0	1

trices agree exactly with that obtained after directly fitting the PCM with te SAS procedure PROC NLIN and inverting the estimated covariance matrix. The $n(2k-1)$ person-item combinations and the $n+(2k-1)$ parameters generate a $n(2k-1) \times (n+(2k-1))$ design matrix, which after applying the identifiability restriction is reduced by one column; this is the design matrix which was used for fitting the model. The first 10 diagonal elements of $\mathcal{I}^{-1}_{\text{JMLE}}$ correspond to the estimated variances of the $\widehat{\theta}_v$'s. There are only 5 different values; the number of repetitions for each of these values is determined by the n_t's, namely $n_2 = 2$, $n_3 = 1$, $n_4 = 4$, $n_7 = 2$ and $n_8 = 1$.

In this example, the following relationships between $\mathcal{I}^{-1}_{\text{SSBE}}$ and $\mathcal{I}^{-1}_{\text{JMLE}}$ can be observed:

1. The block corresponding to the correlations between $\widehat{\beta}_{ih}$ and $\widehat{\gamma}_t$ in $\mathcal{I}^{-1}_{\text{SSBE}}$ is equal to the corresponding block in $\mathcal{I}^{-1}_{\text{JMLE}}$.

2. The variance-covariance matrices of the $\widehat{\beta}_{ih}$'s are equal in both matrices. In particular, s.e.$(\widehat{\beta_{ih}})$ are identical for the SSBE and the JMLE.

3. When $n_t = 1$, the variance of $\widehat{\gamma}_t$ is equal to the variance of $\widehat{\theta}_v$ for $v \in I_t$. Thus $Var(\widehat{\gamma}_3) = Var(\widehat{\theta}_8) = 1.348$ and $Var(\widehat{\gamma}_8) = Var(\widehat{\theta}_4) = 2.277$.

4. When $n_t > 1$, the following inequalities are satisfied:

$$\frac{\text{s.e.}(\widehat{\theta}_v)}{\text{s.e.}(\widehat{\gamma}_2)} = \frac{1.447}{1.002} \leq \sqrt{n_2} = \sqrt{2}, \quad v \in I_2 = \{6, 7\};$$

$$\frac{\text{s.e.}(\widehat{\theta}_v)}{\text{s.e.}(\widehat{\gamma}_4)} = \frac{1.348}{0.914} \leq \sqrt{n_7} = \sqrt{4}, \quad v \in I_4 = \{1, 2, 9, 10\};$$

$$\frac{\text{s.e.}(\widehat{\theta}_v)}{\text{s.e.}(\widehat{\gamma}_7)} = \frac{1.945}{1.577} \leq \sqrt{n_4} = \sqrt{2}, \quad v \in I_7 = \{3, 5\}.$$

Main results

The relationships previously illustrated are *always* valid. As a matter of fact, given that $\mathcal{I}_{\text{JMLE}}$ and $\mathcal{I}_{\text{SSBE}}$ are related through equation (2.4.2), we can derive a relationship between the standard errors. In order to do so, Appendix A introduces a particular class of partitioned matrices, denoted as $\mathcal{C}(\boldsymbol{n})$, where $\boldsymbol{n}$ is a vector of integers which are used to define the sub-blocks of the corresponding matrix. It can easily be verified that $\mathcal{I}_{\text{JMLE}}$ belongs to this class with $T = m = k - 1$. Using Theorem A.5.1, the following results are obtained:

$$s.e.(\widehat{\beta}_{ih}) \text{ are identical for the SSBE and the JMLE.} \tag{2.4.3}$$

$$\left(s.e.(\widehat{\theta}_v)\right)^2 = (s.e.(\widehat{\gamma}_t))^2 + \frac{n_t - 1}{[\mathcal{I}_{\text{SSBE}}]_{tt}} = (s.e.(\widehat{\gamma}_t))^2 + \frac{n_t - 1}{n_t [\mathcal{I}_{\text{JMLE}}]_{vv}}. \tag{2.4.4}$$

$$\frac{1}{\sqrt{[\mathcal{I}_{\text{JMLE}}]_{vv}} \times s.e.(\widehat{\gamma}_t)} \leq \frac{s.e.(\widehat{\theta}_v)}{s.e.(\widehat{\gamma}_t)} \leq \sqrt{n_t} \quad \forall v \in I_t; \;\; t = 1, \ldots, mk - 1. \tag{2.4.5}$$

Table 2.2: Computation of $(s.e.(\widehat{\theta}_v))^2$ from $(s.e.(\widehat{\gamma}_t))^2$ with $v \in I_t$ using (2.4.4)

t	n_t	$[\mathcal{I}_{\text{SSBE}}]_{tt}$	$(s.e.(\widehat{\gamma}_t))^2$	$(n_t-1)/\left[\mathcal{I}_{\text{SSBE}}\right]_{tt}$	$(s.e.(\widehat{\theta}_v))^2$
2	2	2.247	1.002	0.445	1.447
3	1	1.444	1.348	0	1.348
4	4	6.440	0.914	0.466	1.380
7	2	2.713	1.577	0.369	1.945
8	1	1.150	2.277	0	2.277

From these relationships it can be concluded that, for $v \in I_t$, $s.e.(\widehat{\theta}_v) \geq s.e.(\widehat{\gamma}_t)$, with equality only attained when there is just one examinee with a sum score equal to t. The upper bound in (2.4.5) is a useful approximation to $s.e.(\widehat{\theta}_v)$, since it tends to be quite sharp for large-scale test with many items. Moreover, (2.4.5) implies that $1/\sqrt{[\mathcal{I}_{JMLE}]_{vv}} \leq s.e.(\widehat{\theta}_v)$; here $1/\sqrt{[\mathcal{I}_{JMLE}]_{vv}}$ coincides with the estimated standard error when then parameter estimated are taken as if they were the true values.

To end this subsection, let us illustrate the equality (2.4.4) with the same example above-mentionated. In Table 2.2 the values of 1.447, 1.348, 1.380, 1.945 and 2.277 for $\left(s.e.(\widehat{\theta}_v)\right)^2$ are identical to the five values on the diagonal of the upper left part of the variance-covariance $\mathcal{I}_{\text{JMLE}}^{-1}$ matrix shown on page 28.

$$
\mathcal{I}_{\mathrm{SSBE}}^{-1} = \left(\begin{array}{ccccc|ccccccccc}
1.002 & 0.594 & 0.620 & 0.654 & 0.658 & 0.738 & 0.800 & 0.699 & 0.680 & 0.717 & 0.625 & 0.651 & 0.639 & 0.646 \\
0.594 & 1.348 & 0.699 & 0.765 & 0.774 & 0.783 & 0.747 & 0.774 & 0.770 & 0.838 & 0.765 & 0.767 & 0.760 & 0.764 \\
0.620 & 0.699 & 0.914 & 0.870 & 0.887 & 0.780 & 0.740 & 0.818 & 0.838 & 0.955 & 0.890 & 0.878 & 0.888 & 0.886 \\
0.654 & 0.765 & 0.870 & 1.577 & 1.292 & 0.744 & 0.760 & 0.820 & 0.911 & 1.351 & 1.059 & 1.328 & 1.343 & 1.442 \\
0.658 & 0.774 & 0.887 & 1.292 & 2.277 & 0.752 & 0.762 & 0.812 & 0.896 & 1.429 & 1.064 & 1.512 & 1.413 & 1.649 \\
\hline
0.738 & 0.783 & 0.780 & 0.744 & 0.752 & 1.519 & 0.760 & 0.771 & 0.770 & 0.832 & 0.512 & 0.754 & 0.755 & 0.749 \\
0.800 & 0.747 & 0.740 & 0.760 & 0.762 & 0.760 & 1.996 & 0.754 & 0.752 & 0.828 & 0.746 & 0.720 & 0.754 & 0.758 \\
0.699 & 0.774 & 0.818 & 0.820 & 0.812 & 0.771 & 0.754 & 1.396 & 0.794 & 0.890 & 0.816 & 0.815 & 0.688 & 0.819 \\
0.680 & 0.770 & 0.838 & 0.911 & 0.896 & 0.770 & 0.752 & 0.794 & 1.413 & 0.955 & 0.855 & 0.887 & 0.892 & 0.789 \\
0.717 & 0.838 & 0.955 & 1.351 & 1.429 & 0.832 & 0.828 & 0.890 & 0.955 & 2.204 & 1.072 & 1.284 & 1.274 & 1.367 \\
0.625 & 0.765 & 0.890 & 1.059 & 1.064 & 0.512 & 0.746 & 0.816 & 0.855 & 1.072 & 1.728 & 1.018 & 1.022 & 1.059 \\
0.651 & 0.767 & 0.878 & 1.328 & 1.512 & 0.754 & 0.720 & 0.815 & 0.887 & 1.284 & 1.018 & 2.620 & 1.261 & 1.382 \\
0.639 & 0.760 & 0.888 & 1.343 & 1.413 & 0.755 & 0.754 & 0.688 & 0.892 & 1.274 & 1.022 & 1.261 & 2.248 & 1.354 \\
0.646 & 0.764 & 0.886 & 1.442 & 1.649 & 0.749 & 0.758 & 0.819 & 0.789 & 1.367 & 1.059 & 1.382 & 1.354 & 2.949
\end{array}\right)
$$

$$
\mathcal{I}_{\mathrm{SSBE}} = \left(\begin{array}{rrrrr|rrrrrrrrr}
2.247 & 0 & 0 & 0 & 0 & -0.484 & -0.393 & -0.359 & -0.256 & -0.043 & -0.146 & -0.018 & -0.027 & -0.007 \\
0 & 1.444 & 0 & 0 & 0 & -0.317 & -0.123 & -0.256 & -0.206 & -0.053 & -0.190 & -0.020 & -0.045 & -0.013 \\
0 & 0 & 6.440 & 0 & 0 & -1.128 & -0.296 & -1.107 & -1.029 & -0.390 & -1.175 & -0.148 & -0.401 & -0.132 \\
0 & 0 & 0 & 2.713 & 0 & -0.053 & -0.023 & -0.159 & -0.278 & -0.512 & -0.316 & -0.347 & -0.554 & -0.421 \\
0 & 0 & 0 & 0 & 1.150 & -0.006 & -0.005 & -0.027 & -0.064 & -0.229 & -0.081 & -0.238 & -0.225 & -0.265 \\
\hline
-0.484 & -0.317 & -1.128 & -0.053 & -0.006 & 1.501 & 0 & 0 & 0 & 0 & 0.488 & 0 & 0 & 0 \\
-0.393 & -0.123 & -0.296 & -0.023 & -0.005 & 0 & 0.817 & 0 & 0 & 0 & 0 & 0.023 & 0 & 0 \\
-0.359 & -0.256 & -1.107 & -0.159 & -0.027 & 0 & 0 & 1.687 & 0 & 0 & 0 & 0 & 0.221 & 0 \\
-0.256 & -0.206 & -1.029 & -0.278 & -0.064 & 0 & 0 & 0 & 1.697 & 0 & 0 & 0 & 0 & 0.136 \\
-0.043 & -0.053 & -0.390 & -0.512 & -0.229 & 0 & 0 & 0 & 0 & 1.120 & 0 & 0 & 0 & 0 \\
-0.146 & -0.190 & -1.175 & -0.316 & -0.081 & 0.488 & 0 & 0 & 0 & 0 & 1.419 & 0 & 0 & 0 \\
-0.018 & -0.020 & -0.148 & -0.347 & -0.238 & 0 & 0.023 & 0 & 0 & 0 & 0 & 0.749 & 0 & 0 \\
-0.027 & -0.045 & -0.401 & -0.554 & -0.225 & 0 & 0 & 0.221 & 0 & 0 & 0 & 0 & 1.031 & 0 \\
-0.007 & -0.013 & -0.132 & -0.421 & -0.265 & 0 & 0 & 0 & 0.136 & 0 & 0 & 0 & 0 & 0.701
\end{array}\right)
$$

$$
\mathcal{I}_{\mathsf{JMLE}}^{-1} = \left(\begin{array}{cc|c|cccc|cc|c|ccccccccc}
\mathbf{1.447} & 0.557 & 0.594 & 0.620 & 0.620 & 0.620 & 0.620 & 0.654 & 0.654 & 0.658 & 0.738 & 0.800 & 0.699 & 0.680 & 0.717 & 0.625 & 0.651 & 0.639 & 0.646 \\
0.557 & \mathbf{1.447} & 0.594 & 0.620 & 0.620 & 0.620 & 0.620 & 0.654 & 0.654 & 0.658 & 0.738 & 0.800 & 0.699 & 0.680 & 0.717 & 0.625 & 0.651 & 0.639 & 0.646 \\
\hline
0.594 & 0.594 & \mathbf{1.348} & 0.699 & 0.699 & 0.699 & 0.699 & 0.765 & 0.765 & 0.774 & 0.783 & 0.747 & 0.774 & 0.770 & 0.838 & 0.765 & 0.767 & 0.760 & 0.764 \\
\hline
0.620 & 0.620 & 0.699 & \mathbf{1.380} & 0.759 & 0.759 & 0.759 & 0.870 & 0.870 & 0.887 & 0.780 & 0.740 & 0.818 & 0.838 & 0.955 & 0.890 & 0.878 & 0.888 & 0.886 \\
0.620 & 0.620 & 0.699 & 0.759 & \mathbf{1.380} & 0.759 & 0.759 & 0.870 & 0.870 & 0.887 & 0.780 & 0.740 & 0.818 & 0.838 & 0.955 & 0.890 & 0.878 & 0.888 & 0.886 \\
0.620 & 0.620 & 0.699 & 0.759 & 0.759 & \mathbf{1.380} & 0.759 & 0.870 & 0.870 & 0.887 & 0.780 & 0.740 & 0.818 & 0.838 & 0.955 & 0.890 & 0.878 & 0.888 & 0.886 \\
0.620 & 0.620 & 0.699 & 0.759 & 0.759 & 0.759 & \mathbf{1.380} & 0.870 & 0.870 & 0.887 & 0.780 & 0.740 & 0.818 & 0.838 & 0.955 & 0.890 & 0.878 & 0.888 & 0.886 \\
\hline
0.654 & 0.654 & 0.765 & 0.870 & 0.870 & 0.870 & 0.870 & \mathbf{1.945} & 1.208 & 1.292 & 0.744 & 0.760 & 0.820 & 0.911 & 1.351 & 1.059 & 1.328 & 1.343 & 1.442 \\
0.654 & 0.654 & 0.765 & 0.870 & 0.870 & 0.870 & 0.870 & 1.208 & \mathbf{1.945} & 1.292 & 0.744 & 0.760 & 0.820 & 0.911 & 1.351 & 1.059 & 1.328 & 1.343 & 1.442 \\
\hline
0.658 & 0.658 & 0.774 & 0.887 & 0.887 & 0.887 & 0.887 & 1.292 & 1.292 & \mathbf{2.277} & 0.752 & 0.762 & 0.812 & 0.896 & 1.429 & 1.064 & 1.512 & 1.413 & 1.649 \\
\hline
0.738 & 0.738 & 0.783 & 0.780 & 0.780 & 0.780 & 0.780 & 0.744 & 0.744 & 0.752 & 1.519 & 0.760 & 0.771 & 0.770 & 0.832 & 0.512 & 0.754 & 0.755 & 0.749 \\
0.800 & 0.800 & 0.747 & 0.740 & 0.740 & 0.740 & 0.740 & 0.760 & 0.760 & 0.762 & 0.760 & 1.996 & 0.754 & 0.752 & 0.828 & 0.746 & 0.720 & 0.754 & 0.758 \\
0.699 & 0.699 & 0.774 & 0.818 & 0.818 & 0.818 & 0.818 & 0.820 & 0.820 & 0.812 & 0.771 & 0.754 & 1.396 & 0.794 & 0.890 & 0.816 & 0.815 & 0.688 & 0.819 \\
0.680 & 0.680 & 0.770 & 0.838 & 0.838 & 0.838 & 0.838 & 0.911 & 0.911 & 0.896 & 0.770 & 0.752 & 0.794 & 1.413 & 0.955 & 0.855 & 0.887 & 0.892 & 0.789 \\
0.717 & 0.717 & 0.838 & 0.955 & 0.955 & 0.955 & 0.955 & 1.351 & 1.351 & 1.429 & 0.832 & 0.828 & 0.890 & 0.955 & 2.204 & 1.072 & 1.284 & 1.274 & 1.367 \\
0.625 & 0.625 & 0.765 & 0.890 & 0.890 & 0.890 & 0.890 & 1.059 & 1.059 & 1.064 & 0.512 & 0.746 & 0.816 & 0.855 & 1.072 & 1.728 & 1.018 & 1.022 & 1.059 \\
0.651 & 0.651 & 0.767 & 0.878 & 0.878 & 0.878 & 0.878 & 1.328 & 1.328 & 1.512 & 0.754 & 0.720 & 0.815 & 0.887 & 1.284 & 1.018 & 2.620 & 1.261 & 1.382 \\
0.639 & 0.639 & 0.760 & 0.888 & 0.888 & 0.888 & 0.888 & 1.343 & 1.343 & 1.413 & 0.755 & 0.754 & 0.688 & 0.892 & 1.274 & 1.022 & 1.261 & 2.248 & 1.354 \\
0.646 & 0.646 & 0.764 & 0.886 & 0.886 & 0.886 & 0.886 & 1.442 & 1.442 & 1.649 & 0.749 & 0.758 & 0.819 & 0.789 & 1.367 & 1.059 & 1.382 & 1.354 & 2.949
\end{array}\right)
$$

$$
\mathcal{I}_{\mathsf{JMLE}} \left(\begin{array}{cc|c|cccc|cc|c|ccccccccc}
\mathbf{1.124} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -0.242 & -0.196 & -0.180 & -0.128 & -0.022 & -0.073 & -0.009 & -0.014 & -0.003 \\
0 & \mathbf{1.124} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -0.242 & -0.196 & -0.180 & -0.128 & -0.022 & -0.073 & -0.009 & -0.014 & -0.003 \\
\hline
0 & 0 & \mathbf{1.444} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -0.317 & -0.123 & -0.256 & -0.206 & -0.053 & -0.190 & -0.020 & -0.045 & -0.013 \\
\hline
0 & 0 & 0 & \mathbf{1.610} & 0 & 0 & 0 & 0 & 0 & 0 & -0.282 & -0.074 & -0.277 & -0.257 & -0.098 & -0.294 & -0.037 & -0.100 & -0.033 \\
0 & 0 & 0 & 0 & \mathbf{1.610} & 0 & 0 & 0 & 0 & 0 & -0.282 & -0.074 & -0.277 & -0.257 & -0.098 & -0.294 & -0.037 & -0.100 & -0.033 \\
0 & 0 & 0 & 0 & 0 & \mathbf{1.610} & 0 & 0 & 0 & 0 & -0.282 & -0.074 & -0.277 & -0.257 & -0.098 & -0.294 & -0.037 & -0.100 & -0.033 \\
0 & 0 & 0 & 0 & 0 & 0 & \mathbf{1.610} & 0 & 0 & 0 & -0.282 & -0.074 & -0.277 & -0.257 & -0.098 & -0.294 & -0.037 & -0.100 & -0.033 \\
\hline
0 & 0 & 0 & 0 & 0 & 0 & 0 & \mathbf{1.356} & 0 & 0 & -0.027 & -0.011 & -0.080 & -0.139 & -0.256 & -0.158 & -0.174 & -0.277 & -0.210 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \mathbf{1.356} & 0 & -0.027 & -0.011 & -0.080 & -0.139 & -0.256 & -0.158 & -0.174 & -0.277 & -0.210 \\
\hline
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \mathbf{1.150} & -0.006 & -0.005 & -0.027 & -0.064 & -0.229 & -0.081 & -0.238 & -0.225 & -0.265 \\
\hline
-0.242 & -0.242 & -0.317 & -0.282 & -0.282 & -0.282 & -0.282 & -0.027 & -0.027 & -0.006 & 1.501 & 0 & 0 & 0 & 0 & 0.488 & 0 & 0 & 0 \\
-0.196 & -0.196 & -0.123 & -0.074 & -0.074 & -0.074 & -0.074 & -0.011 & -0.011 & -0.005 & 0 & 0.817 & 0 & 0 & 0 & 0 & 0.023 & 0 & 0 \\
-0.180 & -0.180 & -0.256 & -0.277 & -0.277 & -0.277 & -0.277 & -0.080 & -0.080 & -0.027 & 0 & 0 & 1.687 & 0 & 0 & 0 & 0 & 0.221 & 0 \\
-0.128 & -0.128 & -0.206 & -0.257 & -0.257 & -0.257 & -0.257 & -0.139 & -0.139 & -0.064 & 0 & 0 & 0 & 1.697 & 0 & 0 & 0 & 0 & 0.136 \\
-0.022 & -0.022 & -0.053 & -0.098 & -0.098 & -0.098 & -0.098 & -0.256 & -0.256 & -0.229 & 0 & 0 & 0 & 0 & 1.120 & 0 & 0 & 0 & 0 \\
-0.073 & -0.073 & -0.190 & -0.294 & -0.294 & -0.294 & -0.294 & -0.158 & -0.158 & -0.081 & 0.488 & 0 & 0 & 0 & 0 & 1.419 & 0 & 0 & 0 \\
-0.009 & -0.009 & -0.020 & -0.037 & -0.037 & -0.037 & -0.037 & -0.174 & -0.174 & -0.238 & 0 & 0.023 & 0 & 0 & 0 & 0 & 0.749 & 0 & 0 \\
-0.014 & -0.014 & -0.045 & -0.100 & -0.100 & -0.100 & -0.100 & -0.277 & -0.277 & -0.225 & 0 & 0 & 0.221 & 0 & 0 & 0 & 0 & 1.031 & 0 \\
-0.003 & -0.003 & -0.013 & -0.033 & -0.033 & -0.033 & -0.033 & -0.210 & -0.210 & -0.265 & 0 & 0 & 0 & 0.136 & 0 & 0 & 0 & 0 & 0.701
\end{array}\right)
$$

2.4.3 Numerical Illustration of the Main Results

Let us illustrate the previous theoretical results using a Chilean data set called SIES (*Sistema de Ingreso a la eduación superior*). The data set corresponds to an experimental test applied in 2003. The test was couched in the context of a project leading to implement a new entrance university national test (for details, see Bravo and Manzi, 2004). As a part of this project, a polychotomous test in Mathematics was produced. The test composed of 30 items was applied to 1,090 examinees; each item has 5 alternatives: only one alternative corresponds to 2 points; 7 items contains two alternatives corresponding to 1 point; 23 items contains only one alternative scored with 1 points. Consequently, the items have two steps or categories. The reliability of the test was 0.724.

As it was mentioned in Section 2.2, examinees with total score equal to 0 or equal to the maximum were excluded, remaining 1,088 students. A random sample of size 50 was used. The estimates and the corresponding estimated standard errors of the difficulty parameters for each item and for each step are gathered in Table 2.3; they were computed using both a JML-approach and a SSB-approach. As expected, estimates and the estimated standard errors are equal.

The estimates for the ability parameters, along with their respective estimated standard errors, can be found in Table 2.4; they also were computed using both approaches. Several facts can be verified:

1. When $n_t = 1$, as it is the case for $t = 2$, $t = 23$ and $t = 32$, the estimated standard errors are equal.

Table 2.3: JMLE and SSBE for the difficulty parameters, along with the corresponding estimated standard errors

	JMLE				SSBE			
item i	$\widehat{\beta}_{i1}$	$s.e(\widehat{\beta}_{i1})$	$\widehat{\beta}_{i2}$	$s.e(\widehat{\beta}_{i2})$	$\widehat{\beta}_{i1}$	$s.e(\widehat{\beta}_{i1})$	$\widehat{\beta}_{i2}$	$s.e.(\widehat{\beta}_{i2})$
1	0	0	$-0,5619$	$0,8221$	0	0	$-0,5619$	$0,8221$
2	$-1,0689$	$0,4921$	$-0,1059$	$0,6109$	$-1,0689$	$0,4921$	$-0,1059$	$0,6109$
3	$0,0278$	$0,5929$	$-2,3078$	$0,6120$	$0,0278$	$0,5929$	$-2,3078$	$0,6120$
4	$-0,0178$	$0,5942$	$-2,3738$	$0,6092$	$-0,0178$	$0,5942$	$-2,3738$	$0,6092$
5	$-1,0076$	$0,5006$	$-0,6528$	$0,5691$	$-1,0076$	$0,5006$	$-0,6528$	$0,5691$
6	$-1,2177$	$0,5026$	$-0,8168$	$0,5471$	$-1,2177$	$0,5026$	$-0,8168$	$0,5471$
7	$-1,5700$	$0,5519$	$-1,9611$	$0,5162$	$-1,5700$	$0,5519$	$-1,9611$	$0,5162$
8	$-2,4255$	$0,5070$	$0,8630$	$0,6560$	$-2,4255$	$0,5070$	$0,8630$	$0,6560$
9	$1,2121$	$0,7117$	$-1,9922$	$0,7932$	$1,2121$	$0,7117$	$-1,9922$	$0,7932$
10	$0,5681$	$0,6129$	$-1,7677$	$0,6803$	$0,5681$	$0,6129$	$-1,7677$	$0,6803$
11	$1,2468$	$0,7113$	$-1,8245$	$0,8084$	$1,2468$	$0,7113$	$-1,8245$	$0,8084$
12	$-0,7356$	$0,5960$	$-2,7000$	$0,5724$	$-0,7356$	$0,5960$	$-2,7000$	$0,5724$
13	$-0,5593$	$0,5252$	$-1,3253$	$0,5693$	$-0,5593$	$0,5252$	$-1,3253$	$0,5693$
14	$0,8149$	$0,8344$	$-3,9363$	$0,8275$	$0,8149$	$0,8344$	$-3,9363$	$0,8275$
15	$-0,7016$	$0,5298$	$-1,5768$	$0,5538$	$-0,7016$	$0,5298$	$-1,5768$	$0,5538$
16	$-0,6423$	$0,5180$	$-1,1452$	$0,5695$	$-0,6423$	$0,5180$	$-1,1452$	$0,5695$
17	$-0,6083$	$0,5689$	$-2,3268$	$0,5645$	$-0,6083$	$0,5689$	$-2,3268$	$0,5645$
18	$1,1025$	$0,6492$	$0,3241$	$1,1860$	$1,1025$	$0,6492$	$0,3241$	$1,1860$
19	$-1,0125$	$0,5551$	$-2,1627$	$0,5399$	$-1,0125$	$0,5551$	$-2,1627$	$0,5399$
20	$-0,8679$	$0,4963$	$-0,2147$	$0,6168$	$-0,8679$	$0,4963$	$-0,2147$	$0,6168$
21	$-1,4417$	$0,4851$	$0,5449$	$0,6668$	$-1,4417$	$0,4851$	$0,5449$	$0,6668$
22	$1,5424$	$0,8211$	$-2,7872$	$0,8670$	$1,5424$	$0,8211$	$-2,7872$	$0,8670$
23	$-0,6993$	$0,5039$	$-0,5048$	$0,6045$	$-0,6993$	$0,5039$	$-0,5048$	$0,6045$
24	$-0,4595$	$0,5469$	$-1,8437$	$0,5681$	$-0,4595$	$0,5469$	$-1,8437$	$0,5681$
25	$1,8165$	$0,8191$	$-1,0654$	$1,0754$	$1,8165$	$0,8191$	$-1,0654$	$1,0754$
26	$1,1771$	$0,7120$	$-2,1387$	$0,7815$	$1,1771$	$0,7120$	$-2,1387$	$0,7815$
27	$0,5681$	$0,6129$	$-1,7677$	$0,6803$	$0,5681$	$0,6129$	$-1,7677$	$0,6803$
28	$0,9957$	$0,7144$	$-2,6822$	$0,7488$	$0,9957$	$0,7144$	$-2,6822$	$0,7488$
29	$1,3155$	$0,7108$	$-1,3904$	$0,8591$	$1,3155$	$0,7108$	$-1,3904$	$0,8591$
30	$0,3147$	$0,5634$	$-0,8028$	$0,7043$	$0,3147$	$0,5634$	$-0,8028$	$0,7043$

Table 2.4: JMLE and SSBE for the ability parameters, and the corresponding standard errors

t	n_t	$\widehat{\gamma}_t$	$s.e.(\widehat{\gamma}_t)$	$\widehat{\theta}_i$	$s.e.(\widehat{\theta}_v)$	$s.e.(\widehat{\theta}_v)/s.e.(\widehat{\gamma}_t)$	$\sqrt{n_t}$
2	1	$-3,4443$	$0,7709$	$-3,4443$	$0,7709$	$1,0000$	$1,0000$
6	1	$-2,4757$	$0,5373$	$-2,4757$	$0,5373$	$1,0000$	$1,0000$
9	1	$-2,1145$	$0,4934$	$-2,1145$	$0,4934$	$1,0000$	$1,0000$
10	2	$-2,0178$	$0,4342$	$-2,0178$	$0,4846$	$1,1161$	$1,4142$
11	2	$-1,9287$	$0,4303$	$-1,9287$	$0,4775$	$1,1097$	$1,4142$
12	1	$-1,8457$	$0,4717$	$-1,8457$	$0,4717$	$1,0000$	$1,0000$
13	1	$-1,7678$	$0,4668$	$-1,7678$	$0,4668$	$1,0000$	$1,0000$
14	3	$-1,6940$	$0,4076$	$-1,6940$	$0,4627$	$1,1352$	$1,7321$
15	5	$-1,6238$	$0,3948$	$-1,6238$	$0,4592$	$1,1631$	$2,2361$
16	3	$-1,5565$	$0,4051$	$-1,5565$	$0,4562$	$1,1261$	$1,7321$
17	4	$-1,4917$	$0,3976$	$-1,4917$	$0,4536$	$1,1408$	$2,0000$
18	2	$-1,4291$	$0,4159$	$-1,4291$	$0,4514$	$1,0854$	$1,4142$
19	3	$-1,3684$	$0,4026$	$-1,3684$	$0,4495$	$1,1165$	$1,7321$
20	4	$-1,3092$	$0,3960$	$-1,3092$	$0,4479$	$1,1311$	$2,0000$
21	5	$-1,2514$	$0,3920$	$-1,2514$	$0,4465$	$1,1390$	$2,2361$
22	1	$-1,1948$	$0,4453$	$-1,1948$	$0,4453$	$1,0000$	$1,0000$
23	3	$-1,1392$	$0,4008$	$-1,1392$	$0,4443$	$1,1085$	$1,7321$
28	1	$-0,8709$	$0,4418$	$-0,8709$	$0,4418$	$1,0000$	$1,0000$
31	1	$-0,7133$	$0,4420$	$-0,7133$	$0,4420$	$1,0000$	$1,0000$
32	1	$-0,6607$	$0,4424$	$-0,6607$	$0,4424$	$1,0000$	$1,0000$
33	2	$-0,6077$	$0,4119$	$-0,6077$	$0,4429$	$1,0753$	$1,4142$
37	1	$-0,3911$	$0,4465$	$-0,3911$	$0,4465$	$1,0000$	$1,0000$
38	1	$-0,3349$	$0,4479$	$-0,3349$	$0,4479$	$1,0000$	$1,0000$
39	1	$-0,2777$	$0,4495$	$-0,2777$	$0,4495$	$1,0000$	$1,0000$

2. When $n_t > 1$, then $s.e.(\widehat{\theta}_v) > s.e.(\widehat{\gamma}_t)$ when $v \in I_t$, as can be verified in Table 2.4.

3. The upper bound derived in (2.4.5) is sharp.

4. The mapping $t \longmapsto \widehat{\gamma}_t$ is increasing.

2.5 Simultaneous representation of item and person parameters when using a structural PCM

Typically, using a pseudo-likelihood and maximizing it may in some cases be a useful method for obtaining point estimates; examples are Besag (1974) in the context of lattice spatial data and, more recently, Rabe-Hasketh and Skrondal (2006) in the context of multilevel models for complex survey data. In the case considered in this chapter, the SSB-formulation of the PCM not only helps to explain the inconsistency of the JMLE, but also provides a computationally efficient way to obtain the SSB-estimators; its efficiency is practically independent on both the number of items and the number of persons. Moreover, it can be computed using a SAS code which is provided in Appendix C.

The concern of this section is take advantage of this computational efficiency of the SSB-method and to use it in some specific application. Concretely, the question which is explored in this section is the following: under which conditions the simultaneous representation of both person and items parameters is valid when these parameters are estimated using the MMLE procedure?. This question has sense in a standard-setting procedure, which is explained in Section 2.5.1. In the context of polytomous data, we propose a standard setting procedure, which produces an ability for the minimal competent student; it is denoted as $\widehat{\theta}_{MCS}$. The idea is to assume different true distributions generating the individual abilities; with them, different patterns are simulated and both person and items parameters are estimated using a SSB-procedure and a MMLE-procedure. The latter one is performed using NLMIXED, as tipically done is psychometrics (see Tuerlinckx et al. (2004)); this SAS procedure assumes that individual abilities are normally distributed. With these results, persons are classified following two different way:

1) comparing the $\widehat{\theta}_{SSBE}$ with $\widehat{\theta}_{MCS}$

2) comparing the $\widehat{\theta}_{MMLE}$ with $\widehat{\theta}_{MCS}$

The previous two classifications are compared and therefore the pertinence of a simultaneous representation of both the $\widehat{\theta}_{MMLE}$ and $\widehat{\beta}_{MMLE}$ can be evaluated. The simulation study is detailed in Section 2.5.2; the standard setting procedure for polythomous data is explained in Section 2.5.1; and a discussion of the results is gathered in section 2.5.3.

The $\widehat{\theta}_{MCS}$ is computed using a book-mark procedure (explained later); from a statistical point of view, $\widehat{\theta}_{MCS}$ depends on the difficulty parameters. Therefore, a second concern of this section is to explore which is the impact of the misspecification of the distribution of random effects (i.e. individual abilities) on the estimates of the difficulty parameters.

2.5.1 Standard setting and criterion-referenced measurement

The SSB-estimators could be useful when respondents are categorized in different levels of proficiency with respect to a particular criterion or set of criteria. The categories are most often dichotomous, like in classifying masters and non-masters, but they can also be polytomous, like in grading the performance on an exam. Most standard setting methods rely on a continuum view of mastery (Meskauskas, 1976): mastering a trait or educational objective is conceived as a gradual process. A cutoff point on the continuum indicates the point of sufficient proficiency to be classified as a master. Judgments of experts are used to determine the cutoff points on the observed test score scale. In an examinee-centered method, judges classify respondents into masters, non-masters and/or borderline cases. The cutoff score is set by determining the point on the observed test score scale that is most consistent with these classifications. In a test-centered method, like for example that of Angoff (1971), judges review the items in the test, by rating the probability of success that is expected from masters of the domain. The cutoff on the test score scale is set at the sum of the expected performances on the items of the test.

The continuum view of mastery implies that the acquisition of the underlying trait or ability measured by the criterion-referenced test consists in a progression along a continuum. Hence, when a criterion-referenced test is administered to a heterogeneous sample of respondents, one can expect that the respondents are ordered along one dimension. This assumption is implicit in the use of the observed test score as an indication of the level of proficiency on the measured criterion. However, it can also be inferred from the continuum view of mastery that the items of a criterion-referenced test can in principle be ordered along the same continuum as a function of their difficulty. This idea of ordering the items along the same continuum as the persons seems to be implicit in Glaser (1963)'s original chapter on criterion-referenced measurement: "the standard against which a student's performance is compared … is the behavior which defines each point along the achievement continuum" (Glaser, 1963, p. 519). By consequence, "along such a continuum of attainment, a student's score on a criterion-reference measure provides explicit information as to what the individual can or cannot do" Glaser (1963, p. 519-520).

The assumption of ordering both respondents and items on the same continuum naturally leads to unidimensional IRT models as the adequate formalization of the continuum view of mastery. In these models, persons and items are positioned on the same continuum on the basis of their ability and difficulty, respectively, so that their positions can be compared directly; for details, see Janssen et al. (2000). However, this procedure is valid in an IRT model where the abilities are viewed as unknown parameters. As a matter of fact, for the PCM, equality (2.1.2) implies that

$$\ln\left(\frac{p_{vih}}{p_{vi,h-1}}\right) = \theta_v - \beta_{ih} \quad v = 1, \ldots, n;\ i = 1, \ldots, k;\ h = 0, \ldots, m.$$

This means that the difference between a person parameter and an item parameter is in the logarithmic scale, a ratio between the probabilities p_{vih} and $p_{vi,h-1}$. Thus, if $\theta_v > \beta_{ih}$, then $p_{vih} > p_{vi.h-1}$ i.e. the person parameter θ_v is greater than the difficulty of category h of item i if and only if the probability that person v answers correctly item i at category h is greater than the probability to correctly answers the same item at category $h - 1$.

IRT models have been used for criterion-referenced measurement, as well as a part of a standard setting procedure. In general, given that a criterion-referenced measure for defining mastery in a certain domain conforms to a

Table 2.5: True item-parameters for the simulation study

Item	step 1	step 2	Item	step 1	step 2
1	0	-0,8846	16	-0,9026	-1,2425
2	-0,5737	-0,6807	17	0,5005	-2,6542
3	-0,8371	-1,739	18	0,5912	-1,115
4	0,3182	-2,2671	19	-0,6525	-1,4034
5	-0,1427	-1,4991	20	-0,842	-0,6935
6	-1,7522	-0,2252	21	-1,7916	-0,0251
7	-1,176	-2,1133	22	0,5752	-2,4872
8	-2,7895	0,4789	23	-0,1844	-0,959
9	0,1304	-1,2041	24	-0,4154	-1,6914
10	0,1971	-0,7849	25	1,6767	-0,6138
11	1,5895	-2,0216	26	2,1879	-3,4323
12	-0,2968	-2,7038	27	1,4663	-2,7424
13	-0,9942	-0,7677	28	0,5005	-2,6542
14	0,7367	-4,0345	29	0,6224	-0,8203
15	-1,1721	-1,4317	30	-0,0345	-1,0737

unidimensional IRT model, one could proceed in two ways to set a standard, just like in the existing methods using the observed test score scale. In an examinee-centered approach, one could determine the cutoff as a function of the performance of a selected group of respondents. In a test-centered approach, one could determine the cutoff as a function of the expected average performance of a master on the items.

2.5.2 Design of the Simulation Study

A simulation study consisting of the following four parts was performed:

(a) A test composed of $k = 30$ items, each of them with $h = 2$ categories, was considered. The "true" difficulties correspond to the estimates of the item parameters obtained with SSBE method with a sample of 50 students taken from the SIES test; see Table 2.5.

(b) $n = 200$ individual abilities were generated from specific distributions, which were different from a normal distribution $\mathcal{N}(\mu, \sigma^2)$.

(c) From (a) and (b), the corresponding patterns responses were generated using the PCM.

(d) Using the patterns responses obtained in (c) both items parameters and abilities were estimated by means of two estimation procedures: (i) NLMIXED (from SAS) with $N(0, \sigma^2)$ as the distribution of the abilities; (ii) SSB-method as implemented in SAS. As it is well known, the estimations of the abilities obtained with the NLMIXED are computed with an Empirical Bayes procedure.

It is relevant to make explicit the random-effects specification of the PCM. Let

$$\beta_i = (\beta_{i0}, \ldots, \beta_{im})' \quad \text{for} \quad i = 1, \ldots, k$$
$$X_v = (X_{v1}, \ldots, X_{vk})' \quad \text{for} \quad v = 1, \ldots, n$$

where $X_{vi} = (X_{vi0}, \ldots, X_{vim})'$. The hypotheses underlying the PCM can be written as follows:

(i) X_{vi} depends on (θ_v, β_i) according to equation (2.1.1).

(ii) For each person v, his/her responses $\{X_{vi}, \ldots, X_{vk}\}$ are mutually independent conditionally on $(\theta_v, \beta_1, \ldots, \beta_k)$. This is the so-called Axiom of Local Independence. It means that the process generating X_{vi} conditionally on

$(X_{v1}, \ldots, X_{v,i-1}, X_{v,i+1}, \ldots, X_{vk}, \theta_v, \beta_1, \ldots \beta_k)$ only depends on $(\theta_v, \beta_1, \ldots, \beta_k)$; that is, the responses of person v to the items only depends on his/her ability θ_v and on the difficulty parameters. In the classical literature, this was called the Hypothesis of the Common Cause; see Laplace (1820) and Reichenbach (1956). For details on the Axiom of Local Independence, see Lazarsfeld (1950).

(iii) $X_1, \ldots, X_n$ are mutually independent given $\theta_1, \ldots, \theta_v, \beta_1, \ldots, \beta_k$. This means that once the abilities and difficulties are known, the responses of the examinees are mutually independent.

(iv) $(\theta_i \mid \varphi) \overset{i.i.d}{\sim} F^{\varphi}$, where F is a known probability distribution parameterized with $\varphi \in \Phi$.

(v) $\varphi \perp\!\!\!\perp (\beta_1, \ldots, \beta_k)$.

In order to grasp the meaning of condition (v), two comments should be added:

1. From (i) and (ii), it follows that

$$X_v \perp\!\!\!\perp \varphi \mid \theta_v, \beta_1, \ldots, \beta_k \tag{2.5.1}$$

   Similarly, from (iv) it follows that

$$\theta_v \perp\!\!\!\perp \beta_1, \ldots, \beta_k \mid \varphi \tag{2.5.2}$$

   Condition (2.5.1) means that the parameters $(\beta_1, \ldots, \beta_k)$ are sufficient to describe the conditional process X_v given θ_v. Condition (2.5.2) means that φ is sufficient to describe the marginal process generating θ_v.

In this context , $\varphi \perp\!\!\!\perp \beta_1, \ldots, \beta_k$ defines a cut (see Barndorff-Nielsen (1978); Florens et al. (1990a), Chapter 3); this means that the conditional process and the marginal process are cutted in the sense that the parameters describing both the conditional and the marginal processes are not functionally related.

2. In condition (v) the prior distributions are left unspecified. The only relevant aspect here is the structure of the specification, and not the computation of posterior distribution. Therefore, this way of specifying the model is also valid from a sampling-theory framework. In fact, condition (v) could be replaced by a variation-free property, but in most of the cases, it is difficult to characterized it; for details, see Engle et al. (1983).

   The hypotheses above-mentioned imply that $X_1, \ldots, X_n$ are mutually independent with a common distribution given by

$$P[X_v = x_v \mid \beta_1, \ldots, \beta_k, \varphi] = \int_{\mathbb{R}} \prod_{i=i} P[X_v = x_v \mid \beta_i, \theta] F^{\varphi}(d\theta)] \quad (2.5.3)$$

   where $P[X_v = x_v \mid \beta_i, \theta]$ should be computed using (2.1.1). For a proof, see Mouchart and San Martin (2003).

   The SAS-procedure NLMIXED uses (2.5.3) to estimated the parameters of interest. In our case , we use F^{φ} as a $N(0, \sigma^2)$. To fit the model, some identification restriction is needed. In fact, it can be proved that $(\beta_1, \ldots, \beta_k, \sigma^2)$ is identified by X_1; it is enough to apply the arguments developed by San Martín and Rolin (2007) to equation (2.1.2) $(\theta_v \mid \sigma^2) \sim N(0, \sigma^2)$ in order to identify (β_{ih}, σ^2).

Using the above-mentioned simulation design, three studies were performed, each with a different distribution for the abilities or random effects (see step (b) above):

Normal distribution	$\mathcal{N}(-0.2, 1)$
Mixture of Normal Distributions with equal weights	$0.5\mathcal{N}(-2.7, 1.3) + 0.5\mathcal{N}(1.5, 1.2)$
Mixture of Normal Distributions with different weights	$0.7\mathcal{N}(-1.9, 1.8) + 0.3\mathcal{N}(2.5, 1.3)$.

The parameters of these distributions were chosen in such a way that the true-β values -see step (a) above- lies where the mass of the distribution is 0.9 ($\pm$). The individual abilities were generated as follows: For mixtures of normal distribution $pN(\mu_1, \sigma_1^2) + (1-p)N(\mu_2, \sigma_2^2)$, first a $u \sim U(0, 1)$ is generated. Secondly, if $u < p$, the θ is generated from a $N(\mu_1, \sigma_1^2)$; if not, θ is generated from a $N(\mu_2, \sigma_2^2)$.

For the mixture normal distributions, steps (c) and (d) were replicated $N = 10$ times, whereas for the normal distribution $N = 50$ replications were performed. The focus of the simulation studies is to compare estimations of the difficulties and the individual abilities obtained with the SSB-method (with bias-corrected factors) and the NLMIXED procedure. Consequently, we are interested in comparing these estimations for each replication, and not to compare Monte-Carlo estimators. Furthermore, each replication does not show the same total scores and, accordingly, if the abilities estimations were computed using the Monte-Carlo procedure, then each would be based on a different number of replications. Therefore, contingency tables comparing the classifications induced by the MMLE and SSBE were generated; thereafter an average contingency table was obtained. Let us finally mention that each NLMIXED procedure, including the Empirical Bayes computation, took 3 hours in a PC Intel(R) Xeon(TM) CPU 3.00 GHz 1.00 of RAM.

2.5.3 Discussion of the results I: difficulty parameters

From the simulation study, two types of findings need to be discussed: that related with the β-estimates , and that related with the θ-estimates. For the three different distributions of the random effects, the β-estimates obtained by both the MML method and the SSB method are practically invariant between them, and also very near to the true difficulties. Taking into account that the β-estimates are monotonic functions of the empirical difficulties, the correlation between $\widehat{\beta}_{SSBE}$ and $\widehat{\beta}_{MMLE}$ is almost 1. However , this is not enough to empirically prove an almost perfect agreement between $\widehat{\beta}_{SSBE}$ and $\widehat{\beta}_{MMLE}$.

A way to compare $\widehat{\beta}_{SSBE}$ and $\widehat{\beta}_{MMLE}$ is through the distances between difficulties, namely $\widehat{\beta}_{SSBE,ih} - \widehat{\beta}_{SSBE,i-1,h}$ and $\widehat{\beta}_{MMLE,ih} - \widehat{\beta}_{MMLE,i-1,h}$ (for example, see figure 2.1 and 2.2 for both methods). This is the right way to make a comparison because both difficulties and abilities are represented in an interval scale; the role of the identification restriction is to fix the 0 of the scale .

The invariance of the estimations of the difficulty parameters with respect to the misspecification of the distribution of the individual abilities is due to the sufficiency of total score with respect to φ. In fact, from the exponential family theory it follows that X_{v++} is sufficient for θ_v (when the β's are fixed).

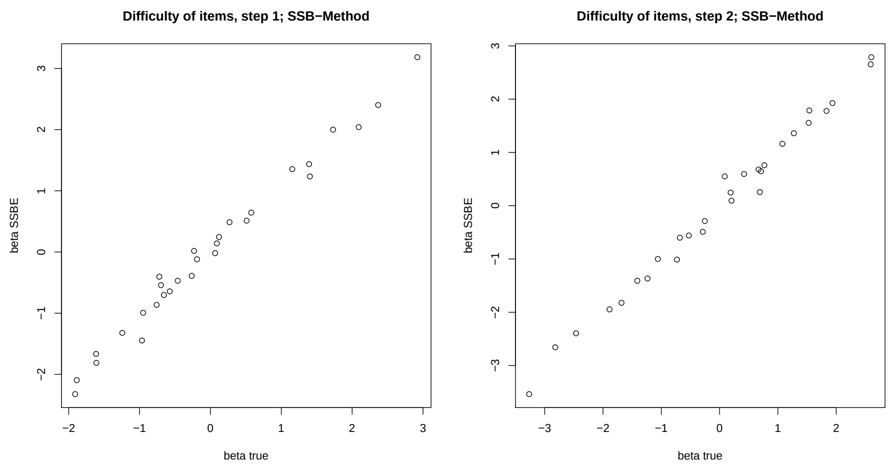

Figure 2.1: beta-true vs beta-SSBE to the Mixture of Normal Distributions with different weights, SSB-Method

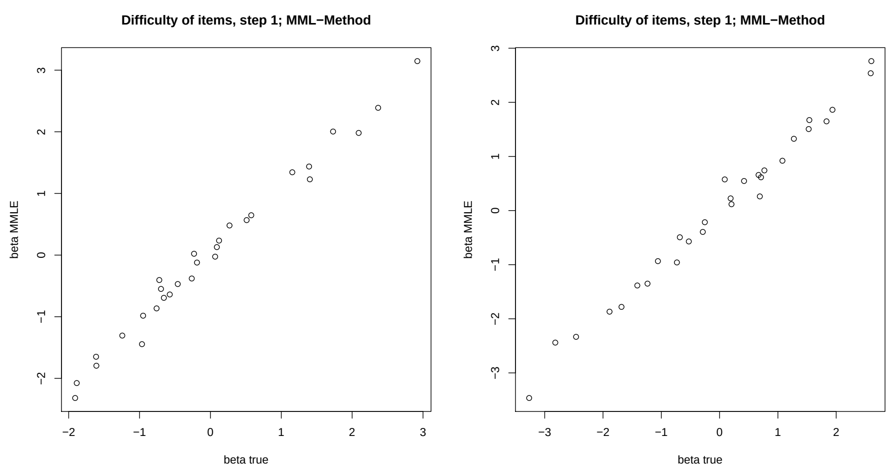

Figure 2.2: beta-true vs beta-MMLE to the Mixture of Normal Distributions with different weights, MML-Method

This can be written as

$$X_v \perp\!\!\!\perp \theta_v \mid X_{v++}, \beta_1, \ldots, \beta_k \tag{2.5.4}$$

On the other hand

$$X_v \perp\!\!\!\perp \varphi \mid \theta_v, \beta_1, \ldots, \beta_k; \tag{2.5.5}$$

but X_{v++} is a function of X_v. It follows that

$$(X_v, X_{v++}) \perp\!\!\!\perp \varphi \mid \theta_v, \beta_1, \ldots, \beta_k, \tag{2.5.6}$$

which in turn implies that

$$X_v \perp\!\!\!\perp \varphi \mid \theta_v, X_{v++}, \beta_1, \ldots, \beta_k, \tag{2.5.7}$$

2.5.4 and 2.5.7 are equivalent to

$$X_v \perp\!\!\!\perp \varphi, \theta_v \mid X_{v++}, \beta_1, \ldots, \beta_k, \tag{2.5.8}$$

which implies that

$$X_v \perp\!\!\!\perp \varphi \mid X_{v++}, \beta_1, \ldots, \beta_k. \tag{2.5.9}$$

That is X_{v++} is a sufficient statistic for φ given $\beta_1, \ldots, \beta_k$. Let us remark that classical sufficiency and Bayesian sufficiency are equivalent for all prior distribution, for details, see Florens et al. (1990b).

Therefore, the likelihood (for one person) obtained after integrating out the random effect θ_v can be factorized as

$$P[X_v = x_v \mid \varphi, \beta_1, \ldots, \beta_k] = P[X_v = x_v \mid X_{v++}, \beta_1, \ldots, \beta_k] \times$$
$$\times \int P[X_{v++} = t \mid \theta, \beta_1, \ldots, \beta_k] \cdot F^{\varphi}(d\theta)$$

Thus, the conditional likelihood $P[X_v = x_v \mid X_{v++,\beta_1,\ldots,\beta_k}]$ only provides information on the β's, and consequently the misspecification of F^{φ} does not affect their estimation.

Let us remark that, in the statistical literature, it is known (also through simulations) that the estimation of the fixed effects (in our case, the difficulty parameters) are robust with respect to the miss-specification of the distribution of the random effects(in our case, the individual difficulties); see Agresti et al. (2004), Heagerty and Kurland (2001), Verbeke and Lesaffre (1996).

Taking into account these considerations, the BC-SSBE of the difficulty parameters can in practice be used to describe the structure of the test. The computational efficiency of the BC-SSBE is an advantage for this kind of descriptions.

Let us illustrate this issue with the estimations of the difficulty parameters of the SIES-test reported at Table 2.3. It can be verified that 7 items satisfy the condition that $\widehat{\beta}_{i1} < \widehat{\beta}_{i2}$, that is, the first step of the item is more easy than the second step.

There are 23 items such that $\widehat{\beta}_{i1} > \widehat{\beta}_{i2}$, that is, the first step of the item is more difficult than the second step. Now when $\widehat{\beta}_{i1} > \widehat{\beta}_{i2}$, the Category Probability Curves (CPC) for the step 1, namely $P_{i1}(\cdot \mid \theta_v)$, is "flatter". which means that a good deal of ability is necessary to obtain a "high" probability of answering correctly step 1 of the item.

When $\widehat{\beta}_{i1} < \widehat{\beta}_{i2}$, the CPC for the step 1 is “stteper”, thus a “not too much ” ability is necessary to obtain a “high” probability of answering step 1 correctly.

Thus , it can be concluded that the Mathematics-SIES test is a difficult test because for 23 items step 1 more difficult than step 2.

2.5.4 Discussion of the results II: person parameters

The concern of this section is to compare classifications of persons using both their $\widehat{\theta}$-estimates and their θ-true abilities. To do it, we simulate a standard setting procedure. The basic ideas are the following:

1. A set of judges receive the items ordered by difficulty.
2. They are asked to put a mark on the item they consider a minimal competent student should answer.
3. It is assumed in the literature that the minimal competent student answer this item with a probability of 0.67 or 0.70 or 0.85. Thus, judgement of each judge can be transformed into the ability of a minimal competent student by solving the following equation:

$$\frac{\exp\left(2\theta_v - \sum_{g=0}^{2}\beta_{ig}\right)}{\sum_{z=0}^{2}\exp\left(z\theta_v - \sum_{g=0}^{z}\beta_{ig}\right)} \leq 0.7 \tag{2.5.10}$$

In this application we suppose that all the judges arrive at the same $\widehat{\theta}_{MCS}$. In practice, this value is obtained as the median of different $\widehat{\theta}_{MCS}$, or other tendency measure. Moreover, we compute the $\widehat{\theta}_{MCS}$ assuming that the probability (2.5.10) is at most equal to 0.70. Once the $\widehat{\theta}_{MCS}$ is obtained, we compute the classifications induced by $\widehat{\theta}_{SSBE}$ and $\widehat{\theta}_{MML}$ with the true classifications (based on the true abilities). If the estimated ability $\widehat{\theta}_{MC}$ is such that $\widehat{\theta}_{MC} \leq 0.7$, we say that the corresponding examinee FAILS the selection test; in other case, we say that he/she PASSES. Similarly, for the true-ability.

Tables 2.6 , 2.7 and 2.8 show the average of the simulations previously described. The reported results were computed using two types of true distributions: mixture of normal distributions and a normal distribution. These tables should be read as follows: in Table 2.6, consider the first row. 88.53% of the examinees who failed (taking into account their true scores) were classified as FAILED EXAMINEES taking into account their abilities estimated MML-procedure; the 11.47% corresponds, therefore, to the proportion of failed examinees (w.r.t. their true scores) who were classified as PASSED EXAMINEES when they are classified using their abilities estimated with the MML-procedure. Similarly, for the abilities estimated with the SSB-procedure. Tables 2.6 , 2.7 and 2.8 reported classifications of examinees with respect to two items: item 18 and item 3. The difficulty of these items approximatively corresponds to the solution of equation (2.5.10) for 0.67 and 0.70, respectively.

Tables 2.6 and 2.7 report the results when the true distribution generating the abilities is estimated under the /wrong) assumption that this distribution is normal. It can be seen that, for item 18, the classifications performed with the SSB-procedure are better than those performed with the MML-procedure. However, for item 3, the opposite holds true. More precisely, the

Table 2.6: Mixture of Normal Distributions with different weights

	Item 18			
	MML		SSBE	
True	Fail	Pass	Fail	Pass
Fail	88.53%	11.47%	98.64%	1.51%
Pass	0%	100%	7.75%	95.35%
	Item 3			
True	Fail	Pass	Fail	Pass
Fail	72.12%	27.88%	97.48%	3.14%
Pass	0%	100%	5.90%	95.28%

Table 2.7: Mixture of Normal Distributions with equal weights

	Item 18			
	MML		SSBE	
True	Fail	Pass	Fail	Pass
Fail	87.84%	12.16%	96.42%	3.97%
Pass	0%	100%	7.63%	95.35%
	Item 3			
True	Fail	Pass	Fail	Pass
Fail	86.92%	13.08%	97.48%	3.67%
Pass	0%	100%	4.96%	97.52%

SSB-procedure seems better than the MML-procedure when classifying failed examinees, whereas when classifying passed examinees the MML-procedure is better. Table 2.8 shows the results when the true distribution is normal, in which case the conclusions are similar.

These results show, therefore, that the estimators of the random effects when compared with the SSB-estimators are in general bad when compared with the true values. Non-parametric procedures seem to be unescapable, but this subject is outside the limits of this research.

Table 2.8: Normal Distribution

	Item 18			
	MML		SSBE	
True	Fail	Pass	Fail	Pass
Fail	75.49%	24.51%	87.34%	12.92%
Pass	1.94%	99.05%	7.25%	93.62%
	Item 3			
True	Fail	Pass	Fail	Pass
Fail	75.49%	24.51%	87.34%	12.92%
Pass	1.94%	99.05%	7.25%	93.62%

2.6 Concluding Remarks

This chapter deals with the use and limitations of a pseudo-likelihood estimation method which can be employed as an alternative for a common estimation method used for the JML formulation of the Partial Credit Model. The first result is that the alternative method, the sum score based estimation (SSBE), provides point estimates which are proven to be identical to those of the JMLE. The equality of the point estimates allows the JMLE to be interpreted as a pseudo-likelihood estimate, and this offers some insight in the features of the JMLE for the Rasch model.

The second result is that the standard errors for the JMLE and the SSBE are equal for the difficulties, but not for the abilities. The equality for the difficulties is rather surprising, since in pseudo-likelihood estimation the standard errors typically do not agree with those of the MLE for the true model. As far as the abilities are concerned, an exact formula relating the standard error of the JMLE to that of the SSBE is provided. This is supplemented by upper and lower bounds on the ratio between the two standard errors, one of which is quite sharp.

In order to obtain the results for the standard errors, a special class of patterned partitioned matrices has been defined and it has been shown how to obtain their inverses efficiently, something that may be useful beyond its application to this paper. Moreover, the Fisher Information Matrix evaluated at the JMLE can be exactly recovered from the estimated covariance matrix of the SSBE.

The relationships established in this paper are not only of theoretical interest, but they have also a practical value. They imply that standard software for the estimation of generalized linear models (GLM) can be used for the joint maximum likelihood estimation without a complicated set up to estimate an ability parameter for each person. The standard errors and information matrix for the JMLE estimates of the individual abilities can be obtained through rather simple equations starting from the SSBE results. In particular, the SSB-estimators helped us to compare the estimations of the difficulty parameters with that obtained when the abilities are interpreted as a latent variable and their distribution is misspecified. Through simulations, we prove that these estimators are quite similar. Using the jargon of generalized linear mixed models, we would say that the estimation of the fixed effects is robust with respect to the misspecification of the distribution of the random effects. We also offer a theoretical justification of this fact. A similar comparison was performed between the MLE of the abilities with the predictions of them using an Empirical Bayes procedure. The simulation results show that these estimators are not robust with respect to the misspecification of the distribution of the random effects.

Chapter 3

A Pseudo likelihood Aproach to the Linear Logistic Test Model

3.1 Introduction

The Linear Logistic Test Model (LLTM) can be considered a particular case of a Rasch model in the sense that the probability of answering correctly an item depends on both an individual ability and a difficulty parameter, but the difficulty is explained by a set of sub-task difficulties. As a matter of fact, when a Rasch model is considered, the difficulty parameters can be ordered in an interval scale and, therefore, it is possible to describe these difficulties. However, a Rasch model can not tell us why such an item is more difficult than such other one. The LLTM is intended to answer this question. Each item is associated to specific cognitive operations or subtasks which are necessary to master in order to correctly answer the item. These tasks or cognitive operations are actually an explanation of the difficulty of the item.

In this chapter, the LLTM is reduced to a SSB-model. As a corollary of the results obtained in Chapter 2, it is shown that the SSB-estimators of both the difficulty parameters and the item parameters are equivalent to the MLE of a misspecified model. To do it, it is enough to show how the LLTM can be reduced to a PCM. These issues are discussed in Section 3.2. In order to save a material produced before the developments of Chapter 2, Section 3.3. gathers the derivations dealing with the relationships between the standard errors of the SSB-estimators with that of the JML-estimators. The reader could skip this section without affecting his/her comprehension of this work.

In any case, the main contribution of this chapter deals with a practical use of the SSB-estimation. In fact, as discussed in the last chapter, the SSB-estimators of the difficulty parameters are practically equivalent to the estimators obtained through the marginal maximum likelihood; and both are practically equal to the true values. Furthermore, the SSB-method is computationally very efficient. Therefore, the SSB-method for the LLTM can be used to perform a post-hoc evaluation of the content of the items. More explicitly, suppose we applied a test and we get the results. Typically, we are interesting in analyzing the internal structure of the items. This leads to compute the reliability of the test, to analyze the properties of the items (empirical difficulties, item-test correlation and so on). A complementary analysis could be the post-hoc evaluation of their difficulty. In fact, the items-builders and/or external judges can be asked on specific cognitive operations or sub-tasks which are necessary to correctly answer the items. Thus, a matrix where the rows represent items and the columns represent sub-tasks is defined in such a way that an entry equal to 1 means that the respective items involves a specific sub-tasks.

The idea is to verify whether the cognitive-structure of the items can be empirically sustained. To do it, the Rasch-difficulties are estimated using the SSB-procedure. In parallel, the sub-tasks difficulties are also computed using the SSB-procedure developed in this chapter. Once the latters are obtained, LLTM-difficulties are getting and, therefore, can be compared with the Rasch-difficulties. If a good agreement is obtained, it can be concluded that the cognitive-structure of the items is empirically sustained.

This type of applications is useful when the content of an item need to be empirically evaluated. By this way, the accent moves from an analysis at the individual side to an analysis at the item side. At the individual side, it is typically studied whether the latent trait is well measured by the items; to do it, factor-analysis techniques and reliability are typically used. The analysis we are proposing in this chapter could be viewed as a complementary one. It must be confessed that the items should be constructed following the LLTM philosophy: first, to define a pool of cognitive operations or subtasks; and thereafter to map items to these subtasks. The empirical evaluation way we are suggesting in this chapter could be useful to evaluate the pertinence of the cognitive-operations pool. Unfortunately, in Chile (and probably in international assessment programs), the items are constructed taking into account the Bloom-taxonomy (1957) or variations of it, these focus their attention on the latent trait it is supposed to be measured by the test. But such a construction could be evaluated at the item level following the way above-proposed.

In this chapter we illustrate both ways, namely a test where the items were constructed after subtasks were defined; and a test where the items were evaluated using the LLTM philosophy after their construction. The first one is taken from Fischer (1973) and developed in Section 3.4.1; the second one was performed using the SEPA-test and two judges which did a codification of the SEPA-Mathematics test applied to 4-th and 5-th level in 2007; this material is developed in Section 3.4.2.

3.2 Link between PCM and LLTM

3.2.1 Model formulation

As discussed in Chapter 2, the probability function for the PCM is given by

$$p_{vih} \equiv P(X_{vih} = 1) = \frac{\exp\left(h\theta_v - \sum_{g=0}^{h} \beta_{ig}\right)}{\sum_{z=0}^{m} \exp\left(z\theta_v - \sum_{g=0}^{z} \beta_{ig}\right)}, \quad h = 0, 1, \ldots, m \tag{3.2.1}$$

where θ_v is a person parameter and β_{ih} correspond to the difficulty parameter of category h of item i.

If we consider $m = 1$ y $\beta_{ig} = \beta_{j1} = \sum_{l=1}^{p} w_{jl}\alpha_l + c$, we have, that;

$$p_{vj} \equiv P(X_{vj} = 1) = \frac{\exp\left(\theta_v - \beta_{j1}\right)}{\sum_{z=0}^{1} \exp\left(z\theta_v - \sum_{g=0}^{z} \beta_{i1}\right)} \tag{3.2.2}$$

By convention, $\beta_{i0} = 0$ y $\sum_{g=0}^{0} \beta_{i1} = 0$, then

$$p_{vj} = p_{vj} \equiv P(X_{vj} = 1) = \frac{\exp\left(\theta_v - \sum_{l=1}^{p} w_{jl}\alpha_l + c\right)}{1 + \exp\left(\theta_v - \sum_{l=1}^{p} w_{jl}\alpha_l + c\right)}$$

which corresponds to the probability to answers correctly an item j for the LLTM.

Here θ_v is the ability of the individual v, α_l, $l = 1, \ldots, p$, is the difficulty of the cognitive operation l, w_{jl} is the hypothetical frequencies with which each component l influences the solution of the each item j, and c is the usual additive normalization constant.

When

$$\beta_j = \sum_{l=1}^{p} w_{jl}\alpha_l + c, j = 1, \ldots, k \tag{3.2.3}$$

we have the RM.

3.2.2 Identification

It clear that the model is unidentified, so we need to impose a restriction, typically $\sum_{j=1}^{k} \beta_j = 0$. If we replace this condition in (3.2.3) we obtain

$$c = -\frac{1}{k} \sum_{j=1}^{k} \sum_{l=1}^{p} w_{jl}\alpha_l$$

and then the difficulty of the item j is given by

$$\beta_j = \sum_{l=1}^{p} \alpha_l \left(w_{jl} - \overline{w}_{\cdot l}\right) \tag{3.2.4}$$

where $\overline{w}_{\cdot l} = \frac{1}{k}\sum_{j=1}^{k} w_{jl}$.

However , it is difficult to interpret the difference $w_{jl} - \overline{w}_{\cdot l}$, particularly when $w_{jl} \in \{0, 1\}$. A way to provide a more interpretable parametrization and to get simplicity is to fix the difficulty of an item at 0, for instance $\beta_1 = 0$; it follows that

$$\beta_j = \sum_{l=1}^{p} \alpha_l \left(w_{jl} - w_{1l}\right) \tag{3.2.5}$$

Other possibility is to impose $c = 0$ and $\sum_j \beta_j = 0$. In such a case, the structure (in terms of tasks or cognitive operation) of, say, the last item need to be equal to $-\sum_j \sum_l \alpha_l w_{jl}$. This is clearly difficult to build. This explains why, in principle, $c \neq 0$.

3.2.3 The formulation of the SSB in the context of the LLTM

Let N_{tj} be the the random variable that indicates the number of persons with a sum score t who give a correct response to item j. The SSB model is given by:

$$\begin{aligned} &N_{tj} \quad \text{are mutually independent, with} \quad N_{tj} \sim Bin(n_t, p_j^t), \\ &logit(p_j^t) = \gamma_t - \sum_{l=1}^{p} \alpha_l \left(w_{jl} - \overline{w}_{\cdot l}\right), \quad t = 1, \ldots, k-1; \;\; j = 1, \ldots, k \end{aligned} \tag{3.2.6}$$

where n_t is the number of persons with sum score t, γ_t represents a proxy of the ability of an examinee i obtaining a sum score equal to t and p_j^t is the probability of a person with sum score t to give a correct response to item j.

3.2.4 JMLE and SSBE

As discussed in Chapter 2, the JMLE of the parameters of interest for the PCM are equivalent to the MLE of a misspecified model, namely the SSBE-formulation of the PCM. This model actually is a conditional model given the individual total score, but it is explicitly forgotten this aspect and the model is accordingly treated as a marginal one. This explains why the JMLE for the PCM is ill- pased. The same conclusion can be drawn for the LLTM; this conclusion directly follows from the fact that the LLTM can be reduced to the PCM

3.3 Standard errors for the SSBE and JMLE

Not only the JMLE is related to the SSBE for the LLTM, but also the corresponding standard errors. Such relationships straight forward follow both the relationship between the PCM and its SSB-formulation and that between PCM and LLTM. However, while this research was developed, the first relationship between an IRT model and a SSB-formulation was studied for the LLTM. Thereafter, we noticed that the treatment of the PCM was enough to explain why the JMLE for an IRT model with separated individual and item parameters (for the terminology, see Rasch (1960a) is ill- conditioned. In spite of that, we collect the corresponding relationships between the standard errors for the SSBE and JMLE in the context of the LLTM is this section. The reader, however, can skips it without his/her comprehension of the research.

3.3.1 Relationships between the information matrices

Let $u(\eta) = \log(1 + e^{\eta})$. The log-likelihood function for the JMLE can be written as

$$l_{JMLE}(\theta, \alpha) = \sum_{l=1}^{n} x_{i+}\theta_i - \sum_{j=1}^{k} x_{+j} \left(\sum_{l=1}^{p} (w_{jl} - \overline{w}_{\cdot l})\, \alpha_l \right) - G_{JMLE}(\theta, \alpha),$$

where $G_{JMLE}(\theta, \alpha) = \sum_{i=1}^{n} \sum_{j=1}^{k} u(\theta_i - \beta_j)$.

For the SSBE

$$l_{SSBE}(\gamma, \alpha) = \sum_{t=1}^{k-1} \sum_{j=1}^{k} n_{tj}(\gamma_t - \beta_j) - G_{SSBE}(\gamma, \alpha),$$

where $G_{SSBE}(\gamma, \alpha) = \sum_{t=1}^{k-1} n_t \sum_{j=1}^{k} u(\gamma_t - \beta_j)$.

The matrices $\mathcal{I}_{JMLE}$ and $\mathcal{I}_{SSBE}$ are the evaluation of the MLE and these matrices of information reduce to the negative Hessianas of G_{JMLE} and G_{SSBE} respectively.

Since point estimates of the $JMLE$ and $SSBE$ coincide, let us write $\widehat{\eta}_{ij} = \widehat{\theta}_i - \sum_{l=1}^{p} (w_{jl} - \overline{w}_{\cdot l})\, \widehat{\alpha}_l = \widehat{\gamma}_t - \sum_{l=1}^{p} (w_{jl} - \overline{w}_{\cdot l})\, \widehat{\alpha}_l$, with $i \in I_t = \{i | X_{i+} = t\}$,

$t \in \mathcal{T} = \{t|0 < t < k, I_t \neq \emptyset\}$.

For $i \in I_t$, denote by v_{tj} the variance of X_{ij} evaluated at the MLE. Then

$$u''(\widehat{\eta}_{ij}) = v_{tj} = \frac{e^{\widehat{\gamma}_t - \sum_{l=1}^{p} (w_{jl} - \overline{w}_{\cdot l})\, \widehat{\alpha}_l}}{\left(1 + e^{\widehat{\gamma}_t - \sum_{l=1}^{p} (w_{jl} - \overline{w}_{\cdot l}) \widehat{\alpha}_l}\right)^2}.$$

We assume $l \neq l'$ and $t \neq t'$ then

$$\begin{aligned}
&[\mathcal{I}_{JMLE}]_{ii} = v_{t+}, i \in I_t, \quad [\mathcal{I}_{JMLE}]_{ii'} = 0, i \neq i' \\
&[\mathcal{I}_{JMLE}]_{ll} = \sum_{j=1}^{k} (w_{jl} - \overline{w}_{\cdot l})^2 \sum_{t=1}^{k-1} n_t v_{tj}, \quad 1 \leq l \leq p \\
&[\mathcal{I}_{JMLE}]_{ll'} = \sum_{j=1}^{k} [(w_{jl} - \overline{w}_{\cdot l})\,(w_{jl'} - \overline{w}_{\cdot l'})] \sum_{t=1}^{k-1} n_t v_{tj}, \quad 1 \leq l \neq l' \leq p \\
&[\mathcal{I}_{JMLE}]_{il} = -\sum_{j=1}^{k} (w_{jl} - \overline{w}_{\cdot l})\, v_{tj}, \quad 1 \leq t, 1 \leq l \leq p \\
&[\mathcal{I}_{JMLE}]_{il} = [\mathcal{I}_{JMLE}]_{li}, \quad 1 \leq t, 1 \leq l \leq p.
\end{aligned} \tag{3.3.1}$$

Performing similar computations for $\mathcal{I}_{SSBE}$ yields

$$
\begin{aligned}
[\mathcal{I}_{SSBE}]_{tt} &= n_t v_{t+},\ 1 \le t < k, \quad [\mathcal{I}_{SSBE}]_{tt'} = 0,\ 1 \le t \ne t' < k \\
[\mathcal{I}_{SSBE}]_{ll} &= \sum_{j=1}^{k} \left(w_{jl} - \overline{w}_{\cdot l}\right)^2 \sum_{t=1}^{k-1} n_t v_{tj}, \quad 1 \le l \le p \\
[\mathcal{I}_{SSBE}]_{ll'} &= \sum_{j=1}^{k} [\left(w_{jl} - \overline{w}_{\cdot l}\right)\left(w_{jl'} - \overline{w}_{\cdot l'}\right)] \sum_{t=1}^{k-1} n_t v_{tj}, \quad 1 \le l \ne l' \le p \\
[\mathcal{I}_{SSBE}]_{tl} &= -n_t \sum_{j=1}^{k} \left(w_{jl} - \overline{w}_{\cdot l}\right) v_{tj}, \quad 1 \le t, 1 \le l \le p \\
[\mathcal{I}_{SSBE}]_{tl} &= [\mathcal{I}_{SSBE}]_{lt}, \quad 1 \le t, 1 \le l \le p.
\end{aligned} \tag{3.3.2}
$$

Comparing (5) with (6) we obtain the following key relationships:

$$
\begin{aligned}
&[\mathcal{I}_{SSBE}]_{tt} = n_t [\mathcal{I}_{JMLE}]_{tt},\ i \in I_t, 1 \le t < k, \quad [\mathcal{I}_{SSBE}]_{tt'} = 0,\ 1 \le t \ne t' < k \\
&[\mathcal{I}_{SSBE}]_{ll} = [\mathcal{I}_{JMLE}]_{ll}, \quad 1 \le l \le p \qquad [\mathcal{I}_{SSBE}]_{ll'} = [\mathcal{I}_{JMLE}]_{ll'} \quad 1 \le l \ne l' \le p \\
&[\mathcal{I}_{SSBE}]_{tl} = n_t [\mathcal{I}_{JMLE}]_{il},\ i \in I_t, 1 \le t, 1 \le l \le p
\end{aligned} \tag{3.3.3}
$$

Therefore $\mathcal{I}_{JMLE} \in \mathcal{C}(\mathbf{n})$. Using Theorem A.5.1 of Appendix A , it follows that

$$
\left(s.e.(\widehat{\theta}_i)\right)^2 = \left(s.e.(\widehat{\gamma}_t)\right)^2 + \frac{n_t - 1}{[\mathcal{I}_{SSBE}]_{tt}} \tag{3.3.4}
$$

and the bounds

$$
\frac{1}{\sqrt{v_{t+}}} \le s.e.(\widehat{\theta}_i) \le \sqrt{n_t} \cdot s.e.(\widehat{\gamma}_t). \tag{3.3.5}
$$

Table 3.1: Subtask, Estimation via JMLE and SSBE

No. task	α_{SSBE}	$s.e._{\alpha_{SSBE}}$	α_{JMLE}	$s.e._{\alpha_{JMLE}}$
1	$0,3314$	$0,1431$	$0,3314$	$0,1431$
2	$-0,0687$	$0,0730$	$-0,0687$	$0,0730$
3	$0,2712$	$0,0659$	$0,2712$	$0,0659$
4	$1,8747$	$0,0884$	$1,8747$	$0,0884$
5	$0,7841$	$0,0774$	$0,7841$	$0,0774$
6	$0,9113$	$0,0603$	$0,9113$	$0,0603$
7	$0,0138$	$0,0738$	$0,0138$	$0,0738$
8	$0,4009$	$0,0760$	$0,4009$	$0,0760$

3.4 Applications to real data sets

3.4.1 Ilustration 1

A test was applied to 287 student of secondary school in Austria (see Fischer (1973)). The test was composed of 29 items dealing with derivatives. To solve them, 8 tasks are involved: (1) differentiation of the polynomial, (2) product rule, (3) quotient rule, (4) compound functions, (5) $\sin(x)$, (6) $\cos(x)$, (7) $\exp(x)$ and (8) $\ln(x)$.

Example of items:

Item 1

$$x^3(x^2+1)^5 \text{ , Involves tasks 1, 2 y 4}$$

Item 2

$$\tfrac{x^2-3}{5x+4}\text{, Involves tasks 1 y 3}$$

Table 3.1 shows the results of the 8 tasks together with their respective standard errors; it can be observed that the estimates are identical for both the JMLE and the SSBE and the the same holds true for their standard errors.

Table 3.2 shows that $\widehat{\theta}_i = \widehat{\gamma}_t$ with $i \in I_t = \{i|Y_{i+} = t\}$,$t \in \mathcal{T} = \{t|0 < t < k, I_t \neq \emptyset\}$. In particular, the bound (3.3.5) is clearly illustrated. Note that for $t = 28$ and $t = 29$, the data set do not present patterns responses.

Table 3.3 shows the $\widehat{\beta}$ as estimated by the Rasch Model; $\widehat{\beta}^*$ is the reconstruction of the difficulties starting from the $\widehat{\alpha}$'s, where these are estimated with the SSB method; $\widehat{\beta}_p$ are the "true value" (see Fischer, 1973), $\widehat{\beta}_c$ is the bias correction proposed by Haberman (1977, pp.834-835) and the weights w_{jl} are those involved in each item.

The SSBE and JMLE of the α's are identical -as in the previous section-. The advantage of the SSBE is its computational efficiency. Let us provide same comments on the results. The more difficult task for students is the differentiation of composite function because $2,8\%$ of the $\widehat{\theta}_i$'s are greater than the difficulty of this task; whereas the more easy tasks for students in the product rule: 70.63% are greater than it. The more difficult item in this test is the question 29; it is composed of the three more difficult substask, namely the composite function the derivative of sin and that of cos.

The correlation between the difficulties estimated with the Rasch Model and with the LLTM is 0.99359; see also figure 3.1. Therefore, the explanation of the difficulties through the tasks is pretty well sustained empirically.

Table 3.2: Estimation ability via JMLE and SSBE

t	n_t	$\widehat{\gamma}_t$	$s.e._{\widehat{\gamma}_t}$	$\widehat{\theta}_i$	$s.e._{\widehat{\theta}_i}$	$s.e._{\widehat{\theta}_i}/s.e._{\widehat{\gamma}_t}$	$\sqrt{n_t}$
1	1	$-3,679$	1.0348	$-3,679$	1.0348	1,0000	1,0000
2	1	$-2,915$	0.7555	$-2,915$	0.7555	1,0000	1,0000
3	3	$-2,440$	0.3673	$-2,440$	0.6357	1,7307	1,7321
4	8	$-2,081$	0.2008	$-2,081$	0.5664	2,8207	2,8284
5	6	$-1,788$	0.2130	$-1,788$	0.5207	2,4446	2,4495
6	10	$-1,534$	0.1548	$-1,534$	0.4882	3,1537	3,1623
7	13	$-1,308$	0.1291	$-1,308$	0.4641	3,5949	3,6056
8	17	$-1,101$	0.1084	$-1,101$	0.4456	4,1107	4,1231
9	8	$-0,909$	0.1527	$-0,909$	0.4314	2,8251	2,8284
10	17	$-0,728$	0.1021	$-0,728$	0.4204	4,1175	4,1231
11	18	$-0,555$	0.0972	$-0,555$	0.4121	4,2397	4,2426
12	12	$-0,388$	0.1173	$-0,388$	0.4060	3,4612	3,4641
13	19	$-0,225$	0.0923	$-0,225$	0.4019	4,3543	4,3589
14	13	$-0,064$	0.1109	$-0,064$	0.3997	3,6041	3,6056
15	19	0,095	0.0916	0,095	0.3991	4,3570	4,3589
16	19	0,255	0.0919	0,255	0.4004	4,3569	4,3589
17	14	0,416	0.1079	0,416	0.4035	3,7396	3,7417
18	14	0,581	0.1093	0,581	0.4085	3,7374	3,7417
19	17	0,750	0.1011	0,750	0.4159	4,1137	4,1231
20	16	0,927	0.1068	0,927	0.4259	3,9878	4,0000
21	11	1,114	0.1327	1,114	0.4392	3,3097	3,3166
22	10	1,314	0.1448	1,314	0.4568	3,1547	3,1623
23	4	1,533	0.2404	1,533	0.4803	1,9979	2,0000
24	8	1,779	0.1816	1,779	0.5124	2,8216	2,8284
25	3	2,064	0.3224	2,064	0.5579	1,7305	1,7321
26	4	2,412	0.3140	2,412	0.6274	1,9981	2,0000
27	1	2,877	0.7481	2,877	0.7481	1,0000	1,0000
28	0	–	–	–	–	–	–
29	1	–	–	–	–	–	–

Table 3.3: Reconstruction of the difficulties

$item$	$\widehat{\beta}$	$\widehat{\beta}^*$	$\widehat{\beta}_p$	$\widehat{\beta}_c$	w_{jl}							
1	$-0,270$	$-0,262$	$-0,204$	$-0,253$	1	1	0	1	0	0	0	0
2	$-1,807$	$-1,796$	$-1,604$	$-1,735$	1	0	1	0	0	0	0	0
3	$-0,148$	$-0,193$	$-0,143$	$-0,186$	1	0	0	1	0	0	0	0
4	$0,077$	$0,078$	$0,146$	$0,076$	1	0	1	1	0	0	0	0
5	$-0,339$	$-0,193$	$-0,143$	$-0,186$	1	0	0	1	0	0	0	0
6	$-0,846$	$-0,772$	$-0,768$	$-0,746$	0	1	0	0	1	1	0	0
7	$-1,160$	$-1,012$	$-0,977$	$-0,977$	1	0	1	0	1	0	0	0
8	$-0,131$	$-0,101$	$-0,219$	$-0,098$	1	0	1	0	1	1	0	0
9	$1,136$	$0,990$	$0,905$	$0,955$	1	0	1	1	0	1	0	0
10	$0,719$	$0,718$	$0,615$	$0,694$	1	0	0	1	0	1	0	0
11	$1,012$	$0,862$	$0,773$	$0,833$	1	0	1	1	1	0	0	0
12	$-0,322$	$-0,179$	$-0,163$	$-0,173$	1	0	0	1	0	0	1	0
13	$-0,079$	$-0,179$	$-0,163$	$-0,173$	1	0	0	1	0	0	1	0
14	$-1,729$	$-1,783$	$-1,623$	$-1,721$	1	0	1	0	0	0	1	0
15	$0,757$	$0,664$	$0,535$	$0,641$	1	1	0	1	0	1	1	0
16	$0,007$	$0,092$	$0,127$	$0,089$	1	0	1	1	0	0	1	0
17	$0,181$	$0,208$	$0,245$	$0,201$	1	0	0	1	0	0	0	1
18	$0,571$	$0,479$	$0,534$	$0,463$	1	0	1	1	0	0	0	1
19	$0,268$	$0,208$	$0,245$	$0,201$	1	0	0	1	0	0	0	1
20	$-1,079$	$-1,156$	$-1,007$	$-1,116$	0	1	0	0	0	1	0	1
21	$0,834$	$1,051$	$0,943$	$1,014$	1	1	0	1	0	1	0	1
22	$0,776$	$0,718$	$0,615$	$0,694$	1	0	0	1	0	1	0	0
23	$-1,202$	$-1,352$	$-1,328$	$-1,306$	1	1	0	0	1	0	0	0
24	$0,645$	$0,591$	$0,483$	$0,571$	1	0	0	1	1	0	0	0
25	$-0,270$	$-0,262$	$-0,204$	$-0,253$	1	1	0	1	0	0	0	0
26	$-0,010$	$0,078$	$0,146$	$0,076$	1	0	1	1	0	0	0	0
27	$1,032$	$0,990$	$0,905$	$0,955$	1	0	1	1	0	1	0	0
28	$-0,270$	$-0,262$	$-0,204$	$-0,253$	1	1	0	1	0	0	0	0
29	$1,646$	$1,774$	$1,531$	$1,713$	1	0	1	1	1	1	0	0

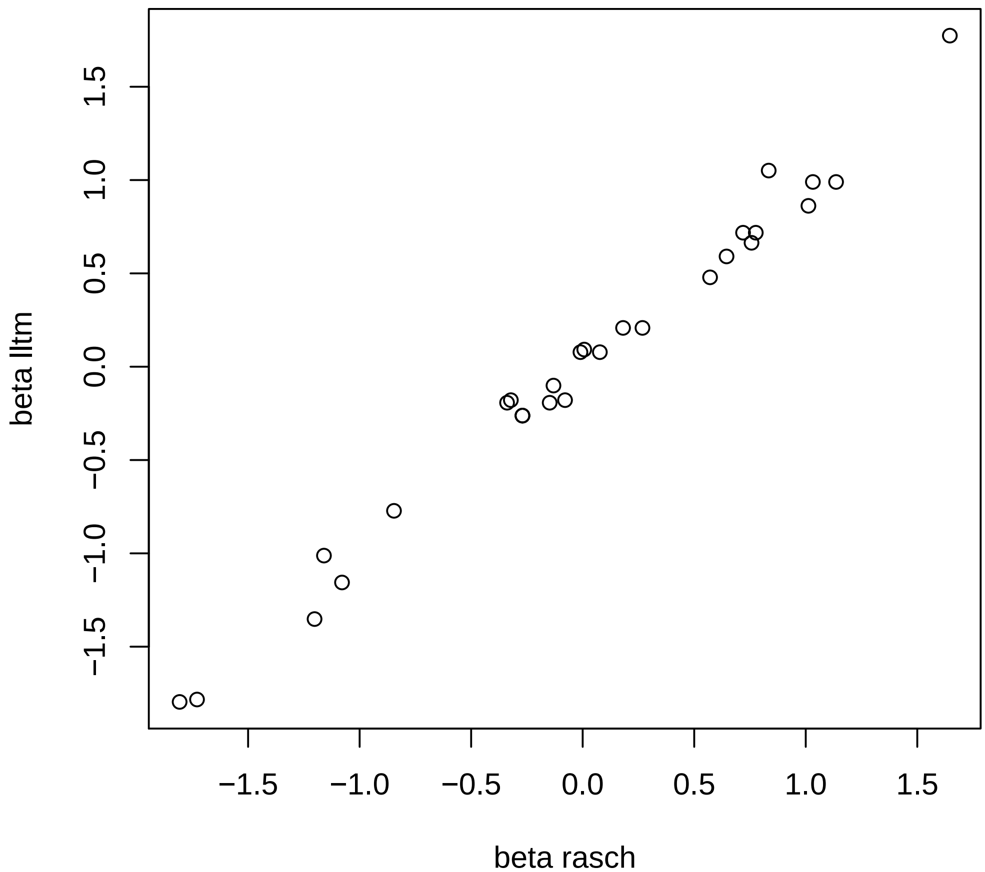

Figure 3.1: Difficulty of items, beta-RASCH vs beta-LLTM ; Ilustration 1

Table 3.4: Percentage of students that exceeds the subtask

Subtasks	Percentage
differentiation of the polynomial	35.66%
product rule	70.63%
quotient rule	35.66%
compount functions	2.80%
sin(x)	19.93%
cos(x)	19.93%
exp(x)	48.95%
ln(x)	35.66%

3.4.2 Ilustration 2

The SEPA test is an instrument generated by the Measurement Center MIDE UC in the context of a value-added project. The test is applied to the eight different levels of primary school.

In the context of the present research, we evaluated the content of a SEPA test using the LLTM-approach. More specifically we used the Mathematics SEPA-test applied to examinees of the 4-th level of primary school.

After defining a set of substasks , the items of the test were linked with them. This linkage procedure was performed through 2 judges; after in good agreement (measured using the Kappa-statistics), a Q-matrix was constructed the entries of this matrix are the w_{jl} depend in equation (3.2.3). This matrix satisfies the equality

$$Q\alpha + c\mathbb{1} = \beta$$

where $\beta = (\beta_1, \ldots, \beta_k)$ are the difficulty parameters, and $\alpha = (\alpha_1, \ldots, \alpha_l)'$ are the substask difficulties.

By identifiability, it is needed that $l \leq k$, that is the number of substasks should be at mats equal to the number of items. This means that $k = 40$ items are explained with only $l = 14$ substasks. In other words, the explanation consists in a dimensional reduction, as typically done in Factor Analysis.

14 subtasks/strategies (see summary figure 3.5) will be considered, which are potentially presentin the items. For this purpose they will be divided into 7 concepts:

1. Knowledge: It refers to the need of managing some concepts or procedures to be able to correctly respond to the item.

   Mathematical knowledge:

   a) Mathematical Conceptual Knowledge (CCM): It is necessary to manage mathematical concepts such as sequence, formation pattern, ratio, decomposition, predecessor, successor, etc. to be able to solve the problem.

   b) Information Management Conceptual Knowledge (CCMI): It is necessary to manage concepts related to information management such as interpretation of graphics or tables to solve the problem.

   c) Geometric Conceptual Knowledge (CCMG): It s necessary to manage concepts related to geometry such as area, perimeter, to be able to solve the problem.

   d) Procedural Mathematical Knowledge (CPM): It refers to the need of managing some mathematical procedures such as fraction operations (amplification, simplification, comparison) or area, perimeter to be able to solve the item.

Mathematical Non Conceptual Knowledge (CCNM): It is necessary to manage non mathematical concepts to be able to solve the item; for instance: century, liter, measures in general.

2. Translation: It refers to the need of translating certain concepts into mathematical terms to be able to solve the item.

    a) Translation from verbal to numeric (TVN): It is necessary to translate verbal information into mathematical terms or operations to be able to correctly solve the exercise.

    b) Translation of figural into numeric (TFN): It is necessary to translate certain figural information into numeric terms (graphic interpretation, etc.)

3. Working with space elements (TE): The item demands working with space elements such as geometric networks, element displacement, to be correctly solved.

4. To dismiss irrelevant numeric information (NR): In this case the problem offers numeric data which are irrelevant for the solution of the problem, which the subject must be able to dismiss to correctly solve the exercise.

5. To perform a mathematical operation: It is necessary when the solution of the item requires the application of some mathematical operation, addition, subtraction, division, or multiplication.

a) To perform a mathematical operation of addition or subtraction (OSR): The problem requires an addition or subtraction to be solved.

b) To perform a mathematical operation of multiplication or division (OMD): The problem requires a multiplication or division to be solved.

6. Type of problem: It refers to the type of exercise to be solved; fractions or integers.

   a) Fraction (FRC): The exercise presented implies working with fractions.

   b) Integer Numbers: The exercise presented implies working with integer numbers.

7. Way of solving the problem: The exercise offers more than one solution strategy (MES): The item can be solved in two or more different ways.

As an example, we will observe the item 18: At an ice-cream parlor, Carlos ordered three quarter liter of ice-cream. Which of the following statements is correct?

(a) Carlos ordered more than three liters of ice-cream.

(b) Carlos ordered less than a liter of ice-cream.

(c) Carlos ordered exactly a liter of ice-cream.

(d) Carlos ordered exactly three liters of ice-cream.

Table 3.5: Subtasks used in the SEPA test

NAME	CODE
Mathematics Conceptual Knowledge	CCMA
Conceptual Knowledge Information Management	CCMI
Geometric Conceptual Knowledge	CCMG
Mathematical Procedural Knowledge	CPM
Conceptual Knowledge nonmathematical	CCNM
Working with spatial elements	TE
Translation of verbal to numerical	TVN
Translation of figural to numerical	TFN
Discard irrelevant numerical information	NR
Execute mathematical operations of addition or subtraction	OSR
Execute mathematical operations of multiplication or division	OMD
Fractions	FRC
Integers	ENT
Exercise has more than one alternative solution	MES

In this item, judges determined that tasks involved are: Mathematical Conceptual Knowledge (CCMA) and fractions (FRC). Results from estimates were obtained via SSB, Table 3.6 shows these results: here it can be observed that both estimates are considered moderately difficult, $\widehat{\alpha}_{CCMA} = 0.6404$ and $\widehat{\alpha}_{FRC} = 0.5668$, the most difficult cognitive sub task for students is given by Geometric Conceptual Knowledge (CCMG) and the least difficult is *The exercise offers more than one solution alternative* (MES) with $\widehat{\alpha}_{CCMG} = 1.0915$ and $\widehat{\alpha}_{MES} = -0.4503$ respectively.

When reconstructing the difficulties of items through sub tasks estimated in LLTM, it was found that there was not the precision obtained when sub tasks are previously generated (see table 3.6) as compared to estimates from RASCH model. Correlation of difficulties of these two models is 0.78 (see Figure 3.2). It is worth mention that judges test several combinations of pres-

Table 3.6: Estimation subtasks and Percentage of teachers that exceeds the subtask

Subtasks	Estimation-SSBE	S.E	Percentage LLTM	Percentage RASCH
CCMI	0.8496	0.0876	18.89%	10.62%
CCMA	0.6404	0.0842	26.14%	13.32%
CCMG	1.0915	0.0842	13.32%	7.59%
CPM	0.7617	0.0776	18.89%	11.63%
CCNM	0.5997	0.091	26.14%	13.32%
TVN	-0.0394	0.0639	57.34%	30.02%
TF	0.7334	0.0535	22.26%	11.63%
TE	-0.1738	0.0976	64.76%	33.56%
NR	0.1076	0.0755	40.98%	26.14%
OSR	0.0511	0.0409	48.9%	26.14%
OMD	0.4798	0.0435	30.02%	14.67%
FRC	0.5668	0.0828	26.14%	14.67%
ENT	-0.0257	0.0719	57.34%	30.02%
MES	-0.4503	0.0946	79.42%	48.9%

ences of sub tasks, obtaining no important correlation. Figures 3.3 and 3.4 show a little bias towards the left on the part of LLTM in reference to model RASCH, which implies that items are considered less difficult for students in LLTM; this makes certain students appearing more skilful than they are when taking into account the estimates with model RASCH. Anyway, what is relevant is that the explanation proposed by judges related to tasks involved in each item is a good one, since the correlation between the descriptive estimates (that is those calculated with model RASCH) and explanatory estimates (that is those calculated with model LLTM) is equal to 0.78, though for this model family, the estimates of fix effects are rather good when compared with real values (in spite of an eventual bad specification of distribution generating individual skills), results obtained can be used to evaluate the con-

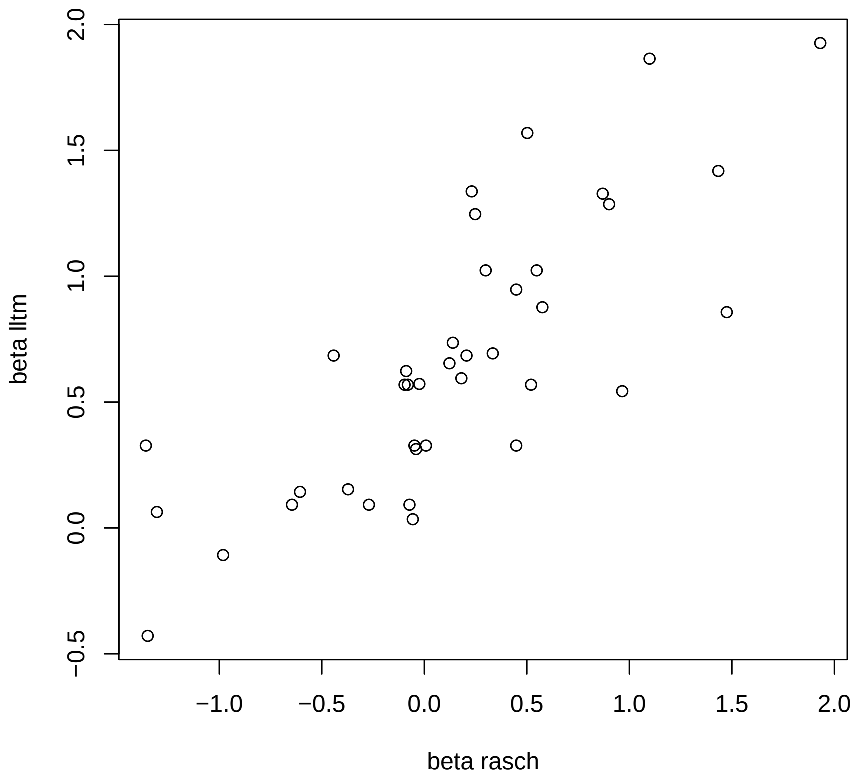

Figure 3.2: Difficulty of items, beta-RASCH vs beta-LLTM ; Ilustration 2

tents of a test. In this way, for instance, such certain items can be chosen whose descriptive and explanatory difficulty is very different and eventually correct that item or simply rule it out from the test. In other words, we have a procedure (SSBE procedure) which due to its Computational quickness, can be used on-line to evaluate cognitive processes underlying an item.

On estimates of skills, correlation between these two models is very good, reaching 0.99 (see Figure 3.5), but as it happens with difficulty estimates,

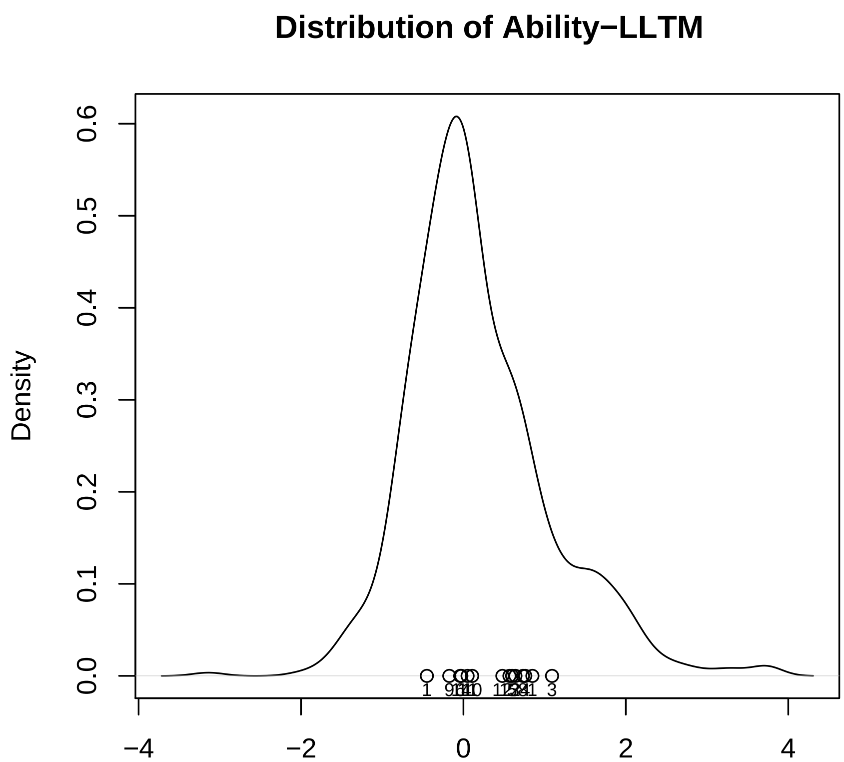

Figure 3.3: Distribution of abilities and difficulties. The circles represent the estimated difficulties-LLTM

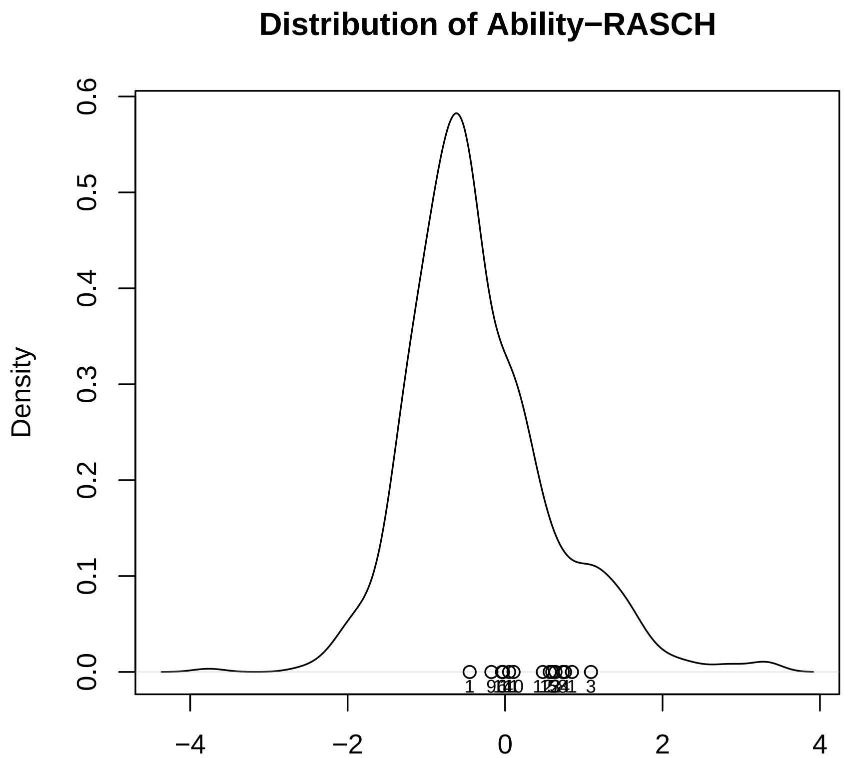

Figure 3.4: Distribution of abilities and difficulties. The circles represent the estimated difficulties-Rasch

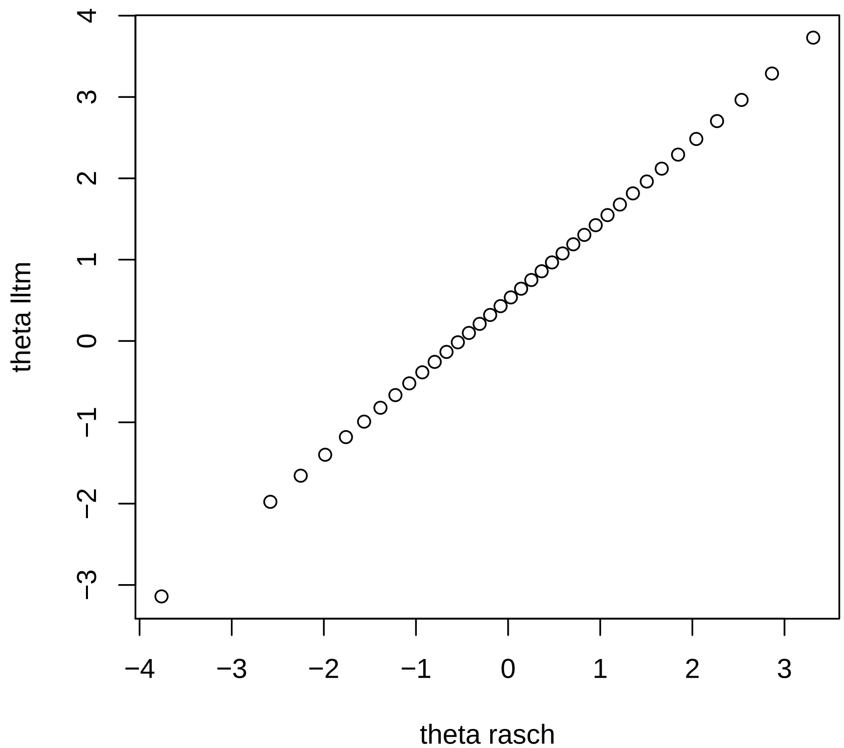

Figure 3.5: Individual Ability; theta-Rasch vs theta-LLTM

there is a little bias which makes students getting a certain mastery of cognitive operation without being a master. Table 3.6 clearly shows that students in LLTM are better classified at a higher percentage than in model RASCH.

3.4.3 Discussion

One of the advantages of SSB estimations is that the number of subjects submitted to the test is fairly irrelevant. This is due to the fact that the sufficient statistic of ability concentrates this information. This minimizes the computational cost in comparison to a traditional methods like the MML, where the cost of getting the estimates increases susbtantially with the number of subjects. Using the LLTM model instead of the RASCH model provides us with information about the underlying difficulties in the subtasks presented in the items. In this regard, Illustration 1 shows more precisely the convenience of previously shaping the subtasks, since, as can be seen in Illustration 2 some biases can result which would imply the bad classification of reaching or not certain knowledge of individuals.

Besides this practical use of the SSB-version of the LLTM model, let us emphasize that our results are valid for a large set of statistical models belonging to the exponential family. As a matter of fact, as it is shown in this chapter and in Appendix B, the PCM can be particularized into several specific models belonging to the exponential family. All these models involve incidental parameters and it is shown that their MLE is equivalent to the MLE of a pseudo-likelihood in Besag's sense. It is important to remark that these results are not in contradiction with the exponential-family theory dealing with the MLE. In fact, when incidental parameters are involved, the parameter space is a function of the sample size and the exponential-family theory of the MLE assumes that the parameter space is fixed for all sample sizes.

Appendix A

A class of partitioned matrices

A.1 Introduction

The main purpose of this appendix is to provide an explicit formula for the inverse of a class of partitioned matrices, whose form is motivated by the Fisher information matrices $\mathcal{I}_{\mathrm{JMLE}}$ and $\mathcal{I}_{\mathrm{SSBE}}$ of the PCM model and its SSB-formulation; for examples, see pages 27 and 28. The content of this Appendix is a detailed explanation of what was sketched by del Pino et al. (2008).

Let us introduce some notation to deal with these patterned matrices. Letting $S_k = \{1, \ldots, k\}$, an entry $C_{rr'}$ of a squared matrix of order k can be identified with the value of a function defined on $S_k \times S_k$. An integer array $\boldsymbol{n} = (n_1, \ldots, n_{T+1})$, with $\sum_{i=1}^{T+1} n_i = k$, induces a partition of S_k:

$$\left(\{1, \ldots, n_1\}, \{n_1 + 1, \ldots, n_1 + n_2\}, \ldots, \left\{\sum_{i=1}^{T} n_i + 1, \ldots, \sum_{i=1}^{T+1} n_i\right\}\right).$$

The increasing order of the integers which appear in each subset of the partition should be respected. In turn this partition induces a block pattern for C, as illustrated by the matrices $\mathcal{I}_{\text{JMLE}}$ and $\mathcal{I}^{-1}_{\text{JMLE}}$ at page 28. Finally, $r \in S_k$ is in one to one correspondence with the pair (t, s), where t labels one subset in the partition and s identifies its s-th element. In this way, we may write $C_{rr'} = c(t, s; t', s')$ for some function c with four integer arguments. Take for instance matrix $C = \mathcal{I}^{-1}_{\text{JMLE}}$ at page 28. There $\boldsymbol{n} = (2, 1, 4, 2, 1, 9)$ and $T = 5$. The element $C_{49} = 0.870$ corresponds to the third subset of the partition $(t = 2)$ and its first element $(s = 1)$, and to the fourth subset of the partition $(t' = 4)$ and its first element $(s' = 1)$. Therefore, we write $C_{49} = c(3, 2; 4, 2)$.

A.2 The class $\mathcal{C}(\boldsymbol{n})$

The blocks of the patterned matrix as those of pages 27 and 28 satisfy a large number of equality constraints. The behavior is different between $t \leq T$ and $t = T + 1$, and so we rewrite the local index s as j, and similarly s' as j'. We also write $n = \sum_{i=1}^{T} n_i$ and $m = n_{T+1}$. With this convention, the constraints essentially convey the idea that the local indices (s, s') do not affect the entry values of the matrix. There is however one exception, namely that for $t = t'$, it is relevant whether these local indices are equal or not.

An alternative way of expressing this idea is that the values of the function c may be written as values of functions with fewer arguments, and this is what is done in the following definition.

Definition A.2.1. *Let $t \neq t' \leq T$. A square matrix C of order $m + n$ is said to belong to the class $\mathcal{C}(\boldsymbol{n})$ (or just $C \in \mathcal{C}(\boldsymbol{n})$) if there exist functions*

d, e, u, v satisfying $c(t, s; t, s) = d(t)$ with $1 \leq s \leq n_t$; $c(t, s; t, s') = e(t, t)$ for $1 \leq s \neq s' \leq n_t$; $c(t, s; t', s') = e(t, t')$ with $1 \leq s \leq n_t$ and $1 \leq s' \leq n_{t'}$; $c(t, s; T+1, j') = u(t, j')$; and $c(T+1, j; t', s') = v(t, j)$. The values $c(T+1, j; T+1, j')$ are arbitrary.

$\mathcal{C}_S(\boldsymbol{n})$ is the subclass formed by all symmetric matrices in $\mathcal{C}(\boldsymbol{n})$, in which case $v = u$ and $e(t, t') = e(t', t)$. $\mathcal{C}_{S0}(\boldsymbol{n})$ is the subclass of $\mathcal{C}_S(\boldsymbol{n})$ determined by the constraint $e = 0$.

Theorem A.2.1. *If a nonsingular C belongs to the class $\mathcal{C}(\boldsymbol{n})$ then also its inverse does. The same holds for $\mathcal{C}_S(\boldsymbol{n})$.*

A.3 The bar-operation

Let us introduce the *bar-operation* which transforms a $(n+m) \times (n+m)$ matrix $C \in \mathcal{C}_{S0}(\boldsymbol{n})$ into a square matrix $\overline{C}$ of $(T+m) \times (T+m)$ according to

$$\begin{array}{rcll} \overline{C}_{t,t} & = & n_t d_t, & t \in S_T \\ \overline{C}_{t,t'} & = & 0, & t \neq t' \in S_T \\ \overline{C}_{t,T+j'} & = & n_t\, u(t, j'), & t \in S_T,\ j' \in S_m \\ \overline{C}_{T+j,T+j'} & = & c(T+1, j; T+1, j'), & j, j' \in S_m. \end{array} \tag{A.3.1}$$

Clearly $\overline{C}$ determines C uniquely. If C is written in the following partitioned form

$$C = \begin{bmatrix} D & A \\ A' & H \end{bmatrix}, \tag{A.3.2}$$

with $D_{n\times n}$, $H_{m\times m}$ and $A_{n\times m}$, then $\overline{C}$ is also written in partitioned form by attaching the bar sign to the submatrices, that is,

$$\overline{C}_{(T+m)\times(T+m)} = \begin{bmatrix} \overline{D} & \overline{A} \\ \overline{A}' & \overline{H} \end{bmatrix}. \tag{A.3.3}$$

By definition, $\overline{H} = H$. For instance, let $C = \mathcal{I}_{\text{JMLE}}$ as given at page 28, then $\overline{C}$ is given by

$$\overline{C} = \left(\begin{array}{rrrrr|rrrrrrrrr}
2.248 & 0 & 0 & 0 & 0 & -0.484 & -0.392 & -0.360 & -0.256 & -0.044 & -0.146 & -0.018 & -0.028 & -0.006 \\
0 & 1.444 & 0 & 0 & 0 & -0.317 & -0.123 & -0.256 & -0.206 & -0.053 & -0.190 & -0.020 & -0.045 & -0.013 \\
0 & 0 & 6.440 & 0 & 0 & -1.128 & -0.296 & -1.108 & -1.028 & -0.392 & -1.176 & -0.148 & -0.400 & -0.132 \\
0 & 0 & 0 & 2.712 & 0 & -0.054 & -0.022 & -0.160 & -0.278 & -0.512 & -0.316 & -0.348 & -0.554 & -0.420 \\
0 & 0 & 0 & 0 & 1.150 & -0.006 & -0.005 & -0.027 & -0.064 & -0.229 & -0.081 & -0.238 & -0.225 & -0.265 \\
\hline
-0.484 & -0.317 & -1.128 & -0.054 & -0.006 & 1.501 & 0 & 0 & 0 & 0 & 0.488 & 0 & 0 & 0 \\
-0.392 & -0.123 & -0.296 & -0.022 & -0.005 & 0 & 0.817 & 0 & 0 & 0 & 0 & 0.023 & 0 & 0 \\
-0.360 & -0.256 & -1.108 & -0.160 & -0.027 & 0 & 0 & 1.687 & 0 & 0 & 0 & 0 & 0.221 & 0 \\
-0.256 & -0.206 & -1.028 & -0.278 & -0.064 & 0 & 0 & 0 & 1.697 & 0 & 0 & 0 & 0 & 0.136 \\
-0.044 & -0.053 & -0.392 & -0.512 & -0.229 & 0 & 0 & 0 & 0 & 1.120 & 0 & 0 & 0 & 0 \\
-0.146 & -0.190 & -1.176 & -0.316 & -0.081 & 0.488 & 0 & 0 & 0 & 0 & 1.419 & 0 & 0 & 0 \\
-0.018 & -0.020 & -0.148 & -0.348 & -0.238 & 0 & 0.023 & 0 & 0 & 0 & 0 & 0.749 & 0 & 0 \\
-0.028 & -0.045 & -0.400 & -0.554 & -0.225 & 0 & 0 & 0.221 & 0 & 0 & 0 & 0 & 1.031 & 0 \\
-0.006 & -0.013 & -0.132 & -0.420 & -0.265 & 0 & 0 & 0 & 0.136 & 0 & 0 & 0 & 0 & 0.701
\end{array}\right).$$

The following proposition will be used in the proof of Theorem A.5.1:

Proposition A.3.1. *If $C \in \mathcal{C}_S(\boldsymbol{n})$ is positive definite, so is $\overline{C}$.*

A.4 The tilde-operation

Let $C \in \mathcal{C}_{\mathcal{S}}(\boldsymbol{n})$ be a matrix of orden $m+n$ partitioned as (A.3.2). The tilde-operation is applied to the block A. A matrix $\widetilde{A}$ of order $T \times m$ is obtained, which is defined as follows:

$$\widetilde{A}_{tj'} = u(t, j') \qquad t \in S_T, \;\; j' \in S_m,$$

where $u(t, j') = c(t, s; T+1, j')$ as given at Definition A.2.1. In order words, $\widetilde{A}_{tj'}$ coincide with the j'-th column of the t block of A. As an example, consider the matrix $C = \mathcal{I}_{\mathrm{JMLE}}^{-1}$ shown at page 28. In this case, $T = 5$ and $m = 9$. Therefore, $\widetilde{A}$ is given by

$$\widetilde{A} = \begin{pmatrix} 0.738 & 0.800 & 0.699 & 0.680 & 0.717 & 0.625 & 0.651 & 0.639 & 0.646 \\ 0.783 & 0.747 & 0.774 & 0.770 & 0.838 & 0.765 & 0.767 & 0.760 & 0.764 \\ 0.780 & 0.740 & 0.818 & 0.838 & 0.955 & 0.890 & 0.878 & 0.888 & 0.886 \\ 0.744 & 0.760 & 0.820 & 0.911 & 1.351 & 1.059 & 1.328 & 1.343 & 1.442 \\ 0.752 & 0.762 & 0.812 & 0.896 & 1.429 & 1.064 & 1.512 & 1.413 & 1.649 \end{pmatrix}.$$

A.5 Main results

The final objective is to obtain a convenient formula for C^{-1} when $C \in \mathcal{C}_{S0}(\boldsymbol{n})$. By Theorem A.2.1, $C^{-1} \in \mathcal{C}(\boldsymbol{n})$. Let us consider C partitioned as in (A.3.2) and $\overline{C}$ partitioned as in (A.3.3). Their corresponding inverses are partitioned as follows:

$$C^{-1} = \begin{bmatrix} X & B \\ B' & W \end{bmatrix}, \quad \overline{C}^{-1} = \begin{bmatrix} E & F \\ F' & S \end{bmatrix},$$

where X is of $n \times n$, B is of $n \times m$, H and S are of $m \times m$, E is of $T \times T$, and F is of $T \times m$.

Theorem A.5.1. *For the matrices C, C^{-1}, $\overline{C}$ and $\overline{C}^{-1}$, the following relationships are true:*

(i) $W = S$.

(ii) $F = \widetilde{B}$.

(iii) For any nonsingular matrix A, denote by A^{ij} the ij-entry of its inverse. For a given $r \leq n$, let t be determined by:

$$r = \sum_{i=1}^{t-1} n_i + s, \qquad \textit{with } 1 \leq s \leq n_t.$$

Then

$$C^{rr} = \overline{C}^{tt} + \frac{n_t - 1}{n_t d_t}, \qquad \textit{for } r \in S_n. \tag{A.5.1}$$

(iv) If $C \in \mathcal{C}_{\mathcal{S}_0}(\boldsymbol{n})$ is positive definite then

$$\frac{1}{d_t} \leq C^{rr} \leq n_t \cdot \overline{C}^{tt}, \qquad \textit{for } r \in S_n. \tag{A.5.2}$$

A.6 Proofs

Proof of Theorem A.2.1: The first n rows and columns of the matrix C form T natural groups of size $n_t, t \in S_T$. Performing the same permutation P_t for the rows and columns in the t-th group produces a matrix $P_t C P_t'$. But this operation is equivalent to performing a permutation of the elements of $\{\sum_{k=1}^{t-1} n_k + 1, \ldots, \sum_{k=1}^{t} n_k\}$, before evaluating the function $c(t, s; t', s')$. This shows that $C \in \mathcal{C}(\boldsymbol{n})$ if, and only if, $P_t C P_t' = C$ for all such permutations matrices. Since $P_t' = P_t^{-1}$ it follows that

$$P_t C^{-1} P_t' = \left[(P_t')^{-1} C P_t^{-1}\right]^{-1} = (P_t C P_t')^{-1} = C^{-1}.$$

For $\mathcal{C}_S(\boldsymbol{n})$ just use the fact that the inverse of a symmetric matrix is also symmetric.

□

Proof of Proposition A.3.1: $\boldsymbol{y}_{n+m\times 1}$ is a $\mathcal{C}(\boldsymbol{n})$-vector if their entries are constant within the T sets that partition S_n. This means that there exist a function f satisfying $y_r = f(t)$, for $r \leq n$. For $\boldsymbol{y}_{n+m\times 1}$ define $\overline{\boldsymbol{y}}_{T+m\times 1}$ with entries $\overline{y}_i = f(i), i \leq T$ and $\overline{y}_{T+j} = y_{n+j}, 1 \leq j \leq m$. Straightforward computations show that $\boldsymbol{y}'C\boldsymbol{y} = \overline{\boldsymbol{y}}\overline{C}\overline{\boldsymbol{y}}$. Since C is positive definite, $\boldsymbol{y}'C\boldsymbol{y} > 0$, for all $\boldsymbol{y} \neq \boldsymbol{0}$ and it follows that $\overline{\boldsymbol{y}}\overline{C}\overline{\boldsymbol{y}} > 0$ for all $\overline{\boldsymbol{y}} \neq \boldsymbol{0}$.

□

Proof of Theorem A.5.1: Denote by I_k the identity matrix of order k. Then the key relationships follow from $C\,C^{-1} = I_{(n+m)\times(n+m)}$ and $\overline{C}\,(\overline{C})^{-1} = I_{(T+m)\times(T+m)}$, where the multiplications are performed using the partitioned form.

• *Proof of (i)*: The standard formula for an inverse partitioned matrix yields $W = (H - A'D^{-1}A)^{-1}$ and $S = (H - \overline{A}'(\overline{D})^{-1}\overline{A})^{-1}$. The entries jj' of $A'D^{-1}A$ and $\overline{A}'\overline{D}^{-1}\overline{A}$ are

$$\sum_r \frac{1}{d_t}\widetilde{a}_{tj}\widetilde{a}_{tj'} = \sum_{t=1}^{T} \frac{1}{d_t} \sum_{s=1}^{n_t} \widetilde{a}_{tj}\widetilde{a}_{tj'} = \sum_{t=1}^{T} \frac{n_t}{d_t}\widetilde{a}_{tj}\widetilde{a}_{tj'}$$

and

$$\sum_{t=1}^{T} \frac{1}{n_t d_t}(n_t\widetilde{a}_{tj})(n_t\widetilde{a}_{tj'}) \quad = \quad \sum_{t=1}^{T} \frac{n_t}{d_t}\widetilde{a}_{tj}\widetilde{a}_{tj'},$$

respectively. The equality of these two expressions implies $A'D^{-1}A = \overline{A}'\overline{D}^{-1}\overline{A}$ and hence $W = S$.

• *Proof of (ii)*: From $CC^{-1} = I_{(n+m)\times(n+m)}$ it follows that $(DB + AW)_{ij} = D_{ii}B_{ij} + \sum_{s=1}^{m} A_{is}W_{sj} = d_t b_{tj} + \sum_{s=1}^{m} \widetilde{a}_{ts}W_{sj} = 0$. Similarly, $\overline{C}(\overline{C})^{-1} =$

$I_{(T+m)\times(T+m)}$ and $W = S$ yield the equivalent equation $n_t d_t F_{tj} + \sum_{s=1}^{m} n_t \widetilde{a}_{ts} W_{sj} = 0$. Therefore, $F_{tj} = b_{tj}$ for all t, j. Therefore, $F = \widetilde{B}$.

- *Proof of (iii)*: By definition, $(C^{-1})^{rr} = X_{rr} \equiv d_t^*$ and $[(\overline{C})^{-1}]^{tt} \equiv E_{tt}$. We must then show that $d_t^* = E_{tt} + \frac{n_t - 1}{\overline{D}_t}$. Now, $DX + AB' = I_{n\times n}$ implies $(DX)_{rr} + (AB')_{rr} = 1$ and so $d_t d_t^* + u(t,t) = 1$, where $u_t = \sum_s \widetilde{a}_{ts} \widetilde{b}_{ts}$. Similarly, $(\overline{D}E) + \overline{A}\ F' = \overline{D}E + \overline{A}\ \widetilde{B}' = I_{T\times T}$, yields $n_t d_t E_{tt} + n_t u(t,t) = 1, t \in S_T$. Eliminating u_t from these two equations, (A.5.1) is obtained.

- *Proof of (iv)*: Since $C \in \mathcal{C}(\boldsymbol{n})$ is positive definite, then by Proposition A.3.1 $\overline{C}$ is also positive definite. For any positive define M, the inequality $\frac{1}{M_{ii}} \leq M^{ii}$ holds. Applying it to $(M = C, i = r \leq n)$ and $(M = \overline{C}, t \leq T)$ yields $\frac{1}{d_t} \leq C^{rr}$ and $\frac{1}{n_t d_t} \leq \overline{C}^{tt}$ respectively. Multiplying the second inequality by $n_t - 1$, adding $\overline{C}^{tt}$ to both sides, and using (A.5.1) the upper bound in (A.5.2) follows.

$\square$

Appendix B

Rasch Poisson Count Model

B.1 Introduction

The Rasch Poisson Count model (RPCM) was developed by Rasch (1960a) (see also Lord and Novick (1968)) in the context of testing the reading ability of persons. The model assumes that the reading of each word of a text constitutes an independent Bernoulli trial with the probability of an error on that word being some small value that depends on the difficulty of the word. Denoting the probability of an error by examinee p on a fixed word in text i by γ_{pi}, Rasch (1960a) assumed that it can be expressed as the ratio $\gamma_{pi} = \delta_i/\xi_p$ of a test difficulty parameter $\delta_i \in \mathbb{R}_+$ and an ability parameter $\xi_p \in \mathbb{R}_+$. For computational convenience the reparametrization $\beta_i = -\ln(\delta_i) \in \mathbb{R}$ and $\theta_p = -\ln(\xi_p) \in \mathbb{R}$ is typically used; by so doing, it is avoided to work on bounded parameter spaces. Using the approximation of a Poisson distribution by a sum of independent distributed Bernoulli random variables (see Lord and Novick (1968), section 21.3) and assuming that the values of γ_{pi} are constant within a text, Rasch (1960a) concluded that the number of misreadings Y_{pi} of an examinee p when reading the text i is distributed according to a Poisson distribution of parameter

$$\lambda_{pi} = \frac{\gamma_i}{\xi_p} = \exp(\theta_p - \beta_j); p = 1, \ldots, n; i = 1, \ldots, K \tag{B.1.1}$$

The specification is completed by assuming that $\{Y_{pi} : 1 \leq i \leq K; 1 \leq p \leq g\}$ are mutually independent. The Rasch Poisson Count Model (RPCM) is, therefore, parameterized by $(\gamma_1, \ldots, \gamma_K; \xi_1, \ldots, \xi_n) \in \mathbb{R}_+^K \times \mathbb{R}_+^n$ or, equivalently by $(\beta_1, \ldots, \beta_K;\, \theta_1, \ldots, \theta_n) \in \mathbb{R}^K \times \mathbb{R}^n$. Nevertheless, as is well known, these parameters are not identified, so an identification restriction is needed; typically the most difficult text is taken as a point of reference by choosing a value of $\gamma_1 = 1$ (or, equivalently, $\beta_1 = 0$); (see Rasch (1960a)). Due to the fact that the individual abilities are considered as unknown parameters, we refer to this specification as the fixed-effects specification of the RPCM.

B.2 Link between PCM and RPCM

Rasch Poisson Count Model(RPCM) is represented as, Y_{ij} the number of errors the individual i makes when reading the text j, the model is given by:

$$Y_{ij} \sim Poisson(\lambda_{ij}) \;\; , i = 1, \ldots, n; j = 1, \ldots, k \;\; ; \lambda_{ij} = \exp(\theta_i - \beta_j) \tag{B.2.1}$$

where θ_i is the ability of individual i in the item with difficulty β_j. Its likelihood function is given by:

$$P(Y_{ij} = y) = \frac{\exp(y(\theta_i - \beta_j))}{y! \exp(\exp(\theta_i - \beta_j))}, \quad y = 0, 1, \ldots, \infty \tag{B.2.2}$$

Now this model (RPCM) can be expressed as PCM, if replacing the following terms in B.2.2 we have; $y! = \prod_{z=1}^{y} z = \exp\left(\sum_{z=1}^{y} log(z)\right)$ and we define $\beta_j = \beta_{j1}$ we have:

$$\frac{\exp(y(\theta_i - \beta_{j1}))}{y!} = \exp\left(y(\theta_i - \beta_{j1}) - \sum_{z=1}^{y} log(z)\right)$$

And if we consider $\beta_{j0} = 0$ and $\beta_{ij} = \beta_{j1} + \log j$ for $j > 0$ we have:

$$\begin{aligned}
\frac{\exp(y(\theta_i - \beta_{j1}))}{y! \exp(\exp(\theta_i - \beta_{j1}))} &= \frac{\exp\left(j\theta_i - \sum_{z=0}^{j} \beta_{jz}\right)}{\sum_{w=0}^{\infty} \frac{\exp(w(\theta_i - \beta_{j1}))}{k!}} \\
&= \frac{\exp\left(j\theta_i - \sum_{z=0}^{j} \beta_{jz}\right)}{\sum_{w=0}^{\infty} \exp\left(w\theta_i - \sum_{z=0}^{w} \beta_{jz}\right)}, \quad j = 0, 1, \ldots, \infty
\end{aligned}$$

Which gives the probability function of PCM. The log-likelihood function is given by:

$$l_{JMLE}(\boldsymbol{\theta}, \boldsymbol{\beta}) = \sum_{i=1}^{n} \theta_i y_{i+} - \sum_{j=1}^{k} \beta_j y_{+j} - \sum_{i=1}^{n} \sum_{j=1}^{k} \log(y_{ij}!) - \sum_{i=1}^{n} \sum_{j=1}^{k} \exp(\theta_i - \beta_j)$$

here sufficient statistics are $(\theta_i, \beta_j) = (y_{i+}, y_{+j}) = \left(\sum_{i=1}^{n} y_{ij}, \sum_{j=1}^{k} y_{ij}\right)$

B.3 SSB version of the Rasch Poisson Count Model

B.3.1 Model formulation

N_{tj} represents the number of individuals having t errors when reading the text j. The model is given by:

$$N_{tj} \sim Poisson(n_t \lambda_{tj}) \ , t = 1, 2, \ldots, \infty; j = 1, \ldots, k; \quad \lambda_{tj} = \exp(\gamma_t - \beta_j),$$

where n_t is the number of people with t score, γ_t represents the proxy of ability of an individual i obtaining a score equal to t and p_j^t is the probability of a person with t score giving a response y to item j

$$p_j^t = \frac{(n_t \exp(\gamma_t - \beta_j))^{n_{tj}}}{n_{tj}! \exp(n_t \exp(\gamma_t - \beta_j))},$$

The log-likelihood function is given by:

$$l_{SSBE}(\boldsymbol{\gamma}, \boldsymbol{\beta}) = \sum_{t=1}^{\infty} \gamma_t \sum_{j=1}^{k} n_{tj} - \sum_{j=1}^{k} \beta_j \sum_{t=1}^{\infty} n_{tj} + \sum_{t=1}^{\infty} \sum_{j=1}^{k} n_{tj} \log(n_t) - \sum_{t=1}^{\infty} \sum_{j=1}^{k} n_t \exp(\gamma_t - \beta_j)$$
$$- \sum_{t=1}^{\infty} \sum_{j=1}^{k} \log(n_{tj}!)$$

here sufficient statistics are $(\gamma_t, \beta_j) = (n_{t+}, n_{+j}) = \left(\sum_{j=1}^{k} n_{tj}, \sum_{t=1}^{\infty} n_{tj} \right)$

B.4 Applications to real data sets

B.4.1 Illustration, Evaluation of Teacher Talk

Consider data about evaluation of teacher talk of first-cycle teachers, first to fourth grades of Elementary School, who work on public schools in Chile, in Language subject. We will illustrate a test for 118 students and 14 items. The purpose of this study is quantifying the performance of teachers on the questions they make and the follow ups on student's interventions, to control the flux of the class or check information. This study about communicational performance of teachers will be considered both from the point of view of teacher's ability and the difficulty imposed by every type of question on the teacher and follow up; the component of difficulty of the task to be performed will be divided in three groups:

Elicitation, when the teacher intends to cause physical or psychic immediate responses on student or students.

Exhibition, when the teacher gives information to boys and girls looking for no immediate reaction from them.

Follow up, It is a tracking performed by the teacher to student's response to the questions made to a student, a group, or to the class as a whole.

The next picture shows all the 14 difficulties to be studied:

Elicitation	Demand Control Information Implement Elaborate Review Opinion
Exhibition	The teacher delivers information content The teacher delivers information about cognitive processes to perform The teacher delivers information about the specific tasks to perform
Follow up	Monosyllable/Neutral Repeat Evaluate Reformulate

Let's define as Y_{ij} the number of times that teacher i performs the task j, $i = 1, \dots, n; j = 1, \dots, k$, this random variable is distributed $Y_{ij} \big| \theta_i, \beta_j \sim Poisson(exp(\theta_i - \beta_j))$, where θ_i is the ability of teacher i and β_j is the difficulty of task j.

B.4.2 Results

In Table B.1 we can see that least difficult tasks to be performed by the teacher are two from Elicitation: Control ($\beta_2 = -0.9315$) and Implementation ($\beta_4 = -0.9928$ and a Follow Up which is repeated ($\beta_{12} = -0.9373$). The most difficult is cognitive processes ($\beta_9 = 1.7554$). This means that it is easier for the teacher to evaluate if students are following the class and organize some tasks, train students in language to make them use their verbal abilities, and respond to student contribution partially or completely repeating their answer. Here almost every teacher reaches this level of difficulty. The most difficult task for teacher was to give information about cognitive processes to be performed or being performed during the session: It includes

Table B.1: MMLE and SSBE for the difficulty parameters, along with the corresponding estimated standard errors

	Tarea	Item	$\widehat{\beta}_{ssbe}$	$s.e_{ssbe}$	$\widehat{\beta}_{mml}$	$s.e_{mml}$	%-ssbe	%-mml
Elicitation	Demand	1	0	-	0	-	93,22	93,22
	Control	2	-0,9315	0,059	-0,9315	0,059	98,31	100,000
	Information	3	-0,6438	0,061	-0,6438	0,061	94,92	99,150
	Implement	4	-0,9928	0,058	-0,9928	0,058	98,31	100,000
	Elaborate	5	0,5745	0,083	0,5745	0,083	84,75	85,590
	Review	6	0,9933	0,096	0,9933	0,096	64,41	64,410
	Opinion	7	0,1957	0,074	0,1957	0,074	91,53	93,220
Exposición	Content Information	8	0,1206	0,073	0,1206	0,072	93,22	93,220
	Information cognitive processes	9	1,7554	0,129	1,7554	0,129	7,63	4,230
	Information tasks to be performed	10	-0,7378	0,060	-0,7378	0,060	96,61	99,150
Follow ups	Monosyllable/Neutral	11	0,5702	0,083	0,5702	0,083	85,6	85,590
	Repeat	12	-0,9373	0,059	-0,9373	0,059	98,31	100,000
	Evaluate	13	-0,1844	0,067	-0,1844	0,067	93,22	94,920
	Reformulate	14	-0,6068	0,062	-0,6068	0,062	94,92	99,150

every type of expressions that do not relate to discipline contents, but talk about what will be done or is being made in the session in terms of cognitive processes or strategies involved and their difficulty level. In this task, a small percentage of teachers reach or master this difficulty.

Difficulties estimates $\widehat{\beta}'s$ are exactly the same for both methods. In relation to teacher abilities, although the correlation of both methods is very high (0.9903), in MML an estimation shrinking is produced at extremes (see Figure B.1) and studies of simulation in chapter refcPCM make us think that in the case of being forced to choose some type of methodology, SSBE presents two important advantages: firstly, it reaches a high percentage of real values; secondly the delay time of estimation results is low, no matter the number of individuals; that is, the computational cost is minimal.

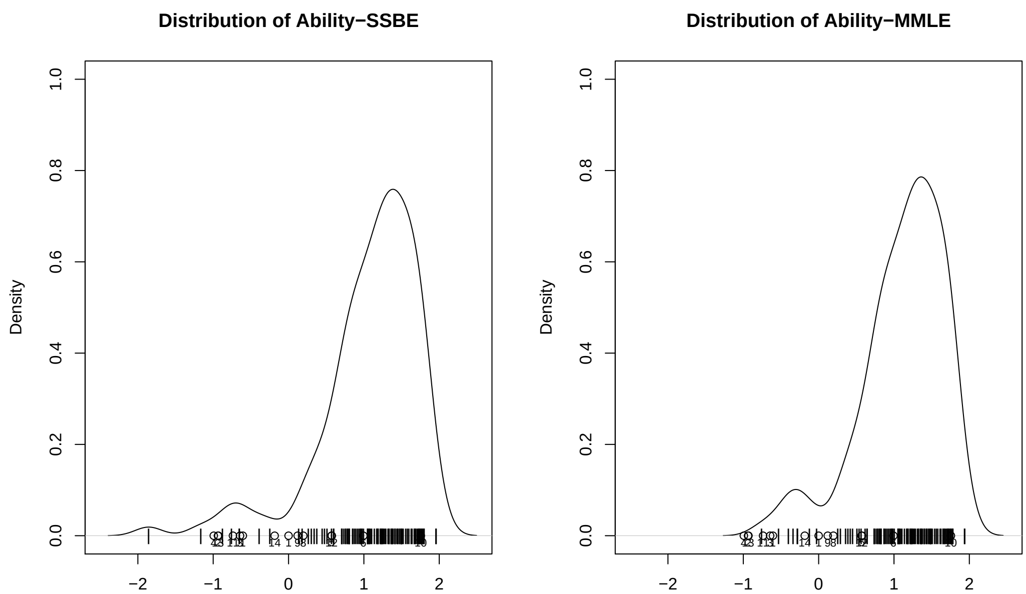

Figure B.1: Distribution Ability

Appendix C

SAS code for fitting the SSB-version

C.1 PCM

Let k be total number of items, n_{tih} indicating the number of persons with a sum score equal to t who answer correctly the step h for the item i and n_t the number of persons with a sum score equal t.

The original data set need to be reduced to a design matrix of the following form:

$$
\text{data file} = \left(\begin{array}{cccc|ccc|ccc}
col1 & col2 & \dots & coln & col(n+1) & \dots & col(n+k) & nti1 & nti2 & nt \\
1 & & & & & & & n111 & n112 & n1 \\
1 & & & & & I_k & & n121 & n122 & n1 \\
\vdots & & & & & & & & \vdots & \\
1 & & & & & & & n1k1 & n1k2 & n1 \\
\hline
 & 1 & & & & & & n211 & n212 & n2 \\
 & 1 & & & & I_k & & n221 & n222 & n2 \\
 & \vdots & & & & & & & \vdots & \\
 & 1 & & & & & & n2k1 & n2k2 & n2 \\
\hline
 & & \ddots & & & \vdots & & & \vdots & \\
\hline
 & & & 1 & & & & nt11 & nt12 & nt \\
 & & & 1 & & I_k & & nt21 & nt22 & nt \\
 & & & \vdots & & & & & \vdots & \\
 & & & 1 & & & & n2k1 & ntk2 & nt
\end{array}\right)
$$

The SAS-procedure needed to fit the SSB-version of PCM is the following:

```
proc nlin data=name_data  maxiter=1000 sigsq=1;
like = 0;
PARMS b1_2-b1_k=0 b2_1-b2_k=0 theta1-thetan=0;
theta=theta1*COL1+theta2*COL2+...+thetan*COLn;
b1_1=0;
beta1=b1_1*COL(n+1)+b1_2*COL(n+2)+...+b1_k*COL(n+k);
beta2=b2_1*COL(n+1)+b2_2*COL(n+2)+...+b2_k*COL(n+k);
exp1=exp(theta-beta1);
exp2=exp(2*theta-beta1-beta2);
denom=1+exp1+exp2;
p1=1/denom;
p2=exp1/denom;
p3=exp2/denom;
model like = sqrt(- 2 * ( nti1*log(p2)+nti2*log(p3)+(nt-nti1-nti2)*log(p1) ) );
run;
```

C.2 LLTM

Let k be total number of items, n_{tj} the random variable that indicates the number of persons with a sum score t , n_t the number of persons with a sum score equal t and L the number the subtaks.

The original data set need to be reduced to a design matrix of the following form:

$$\text{data file} = \left(\begin{array}{cccc|ccc|cc}
col1 & col2 & \dots & coln & col(n+1) & \dots & col(n+L) & ntj & nt \\
1 & & & & & & & n11 & n1 \\
1 & & & & & I_k & & n12 & n1 \\
\vdots & & & & & & & & \vdots \\
1 & & & & & & & n1k & n1 \\
\hline
 & 1 & & & & & & n21 & n2 \\
 & 1 & & & & I_k & & n22 & n2 \\
 & \vdots & & & & & & & \vdots \\
 & 1 & & & & & & n2k & n2 \\
\hline
 & & \ddots & & & \vdots & & & \vdots \\
\hline
 & & & 1 & & & & nt1 & nt \\
 & & & 1 & & I_k & & nt2 & nt \\
 & & & \vdots & & & & & \vdots \\
 & & & 1 & & & & n2k & nt
\end{array}\right)$$

The SAS-procedure needed to fit the SSB-version of PCM is the following:

```
proc nlin data=name_data maxiter=1000 sigsq=1;
like = 0;
PARMS  alpha1-alphaL=0 theta1-thetan=0;
theta=theta1*COL1+theta2*COL2+...+thetan*COLn;
alpha=alpha1*COL(n+1)+alpha2*COL(n+2)+...+alphaL*COL(n+L);
exp=exp(theta-alpha);
```

```
model like=sqrt(-2*(ntj*log(exp/(1+exp) )+(nt-ntj)*log(1/(1+exp))));
run;
```

C.3 RPCM

Let k be total number of text, n_{tj} represents the number of individuals who have t errors they make while reading the text j.

The original data set need to be reduced to a design matrix of the following form:

$$
\text{data file} = \left(\begin{array}{cccc|ccc|cc}
col1 & col2 & \dots & coln & col(n+1) & \dots & col(n+k) & ntj & nt \\
1 & & & & & & & n11 & n1 \\
1 & & & & & I_k & & n12 & n1 \\
\vdots & & & & & & & & \vdots \\
1 & & & & & & & n1k & n1 \\
\hline
 & 1 & & & & & & n21 & n2 \\
 & 1 & & & & I_k & & n22 & n2 \\
 & \vdots & & & & & & & \vdots \\
 & 1 & & & & & & n2k & n2 \\
\hline
 & & \ddots & & & \vdots & & & \vdots \\
\hline
 & & & 1 & & & & nt1 & nt \\
 & & & 1 & & I_k & & nt2 & nt \\
 & & & \vdots & & & & & \vdots \\
 & & & 1 & & & & n2k & nt
\end{array}\right)
$$

The SAS-procedure needed to fit the SSB-version of PCM is the following:

```
proc nlin data=name_data maxiter=1000 sigsq=1;
like = 0;
PARMS  beta2-betak=0 theta1-thetan=1;
```

```
theta=theta1*COL1+theta2*COL2+...+thetan*COLn;
beta1=0;
beta=beta1*COL(n+1)+beta2*COL(n+2)+...+betak*COL(n+k);;
exp=exp(theta-beta);
th=theta-beta;
model like=sqrt(-2*(ntj*log(nt)+ntj*th-log(fact(ntj))-nt*exp));
run;
```

Let us remark that, when we have a large number of examinees, it is necessary to implement a SAS-procedure to get the design matrix.

Bibliography

Adams, R. A. and S. T. Khoo (1993). *QUEST*. Australian Council for Educational Research.

Agresti, A., B. Caffo, and P. Ohman-Strickland (2004). Examples in which misspecification of a random effects distribution reduces efficiency. *Computational Statistics and Data Analysis 47*, 639–653.

Andersen, E. B. (1980). *Discrete Statistical Models with Social Sciences Applications*. North-Holland.

Angoff, W. H. (1971). Scale, norms, and equivalent scores. In R. L. Thorndike (Ed.), *Educational Measurement*, pp. 508–600. American Council of Education.

Baker, F. B. and S. H. Kim (2004). *Item response Theory. Parameter Estimation Techniques*. Marcel Dekker.

Barndorff-Nielsen, O. (1978). *Information and Exponential Families in Statistical Theory*. New York: Wiley.

Berk, R. H. (1970). Consistency a posteriori. *The Annals of Statistics 41*, 894–906.

Bertoli-Barsotti, L. (2005). On the Existence and Uniqueness of JML Estimates for the Partial Credit Model. *Psychometrika 70*, 517–531.

Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems (with Discussion). *Journal of the Royal Statistical Society, Series B 36*, 192–236.

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

Bravo, D. and J. Manzi (2004). Antecedentes y características del Proyecto SIES. Technical report, Proyecto FONDEF SIES. Available at *www.sectormatematica.cl/ppt/SIES.ppt*.

Cizek, G. J. (2001). *Setting performance standards: concepts, methods, and perspectives*. Lawrence Erlbaum.

De Boeck, P. and M. Wilson (2004). *Explanatory Item Response Models. A Generalized Linear and Nonlinear Approach*. Springer.

del Pino, G., E. San Martín, J. González, and P. De Boeck (2008). On the Relationships Between Sum Score Based Estimation and Joint Maximum Likelihood Estimation. *Psychometrika 73*, 145–151.

Embretson, S. E. and S. P. Reise (2000). *Item Response Theory for Psychologists*. Lawrence Elbaurm Associates.

Engle, R. E., D. F. Hendry, and J. F. Richard (1983). Exogeneity. *Econometrica 51*, 277–304.

Fischer, G. (1973). The linear logistic test model as an instrument in educational research. *Acta Psychologica 37*(6), 359–374.

Florens, J.-P., M. Mouchart, and J.-M. Rolin (1990a). *Elements of Bayesian Statistics*. Marcel Dekker.

Florens, J.-P., M. Mouchart, and J.-M. Rolin (1990b). *Elements of Bayesian Statistics*. Marcel Dekker.

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

Glaser, R. (1963). Instructional technology and the measurement of learning outcomes. *American Psychologist 18*, 519–521.

Haberman, S. J. (1977). Maximum Likelihood Estimates in Exponential Response Models. *The Annals of Statistics 5*, 815–841.

Heagerty, P. J. and B. F. Kurland (2001). Misspecified maximum likelihood estimates and generalized linear mixed models. *Biometrika 88*, 973–985.

Janssen, R., F. Tuerlinckx, and P. De Boeck (2000). A Hierarchical IRT Model for Criterior-Reference Measurement. *Journal of Educational and Behavioral Statistics 25*, 285–306.

Kiefer, J. and J. Wolfowitz (1956). Consistency of the Maximum Likelihood Estimators in the Presence of Infinitely Many Incidental Parameters. *The Annals of Mathematical Statistics 27*, 887–906.

Lancaster, T. (2000). The incidental parameter problem since 1948. *Journal of Econometrics 95*, 391–413.

Laplace, P. S. (1820). *Théorie Analytique des Probabilités.* Paris: Mme. Ve. Courcier, Imprimeur-Libraire pour les Mathématiques.

Lazarsfeld, P. F. (1950). The Logical and Mathematical Foundation of Llatent Structural Analysis. In S. S. Stout (Ed.), *Measurement and Prediction*, pp. 362–412. Science Editions.

Linacre, J. M. and B. D. Wright (2000). *WINSTEPS: Multiple-choice, Rating Scale and Partial Credit Rasch Analysis.* MESA Press.

Livinston, S. and M. Zieky (1989). A comparative study of standard-setting methods. *Applied Measurement in Education 2*, 121–141.

Lord, F. M. and M. R. Novick (1968). *Statistical Theories of Mental Test Scores.* Addison Wesley.

Master, G. N. (1982). A rasch model for partial credit scoring. *Psychometrika 47*, 149–174.

Master, G. N. and B. D. Wright (1997). The Partial Credit Model. In W. J. van der Linden and R. K. Hambleton (Eds.), *Handbook of Modern Item Response Theory*, pp. 101–121. Springer.

McCullagh, P. and J. A. Nelder (1989). *Generalized linear models.* Chapmann & Hall.

Meskauskas, J. A. (1976). Evaluation models for criterion-referenced testing: Views regarding mastery in standard setting. *Review of Educational Research 45*, 133–158.

Mouchart, M. and E. San Martin (2003). Specification and identification issues in models involving a latent hierarchical structure. *Journal of Statistical Planning and Inference 111*, 143–163.

Neyman, J. and E. Scott (1948). Consistent Estimates Based on Partially Consistent Observations. *Econometrica 16*, 1–32.

Pfanzagl, J. (1970). Consistent Estimation in the Presence of Incidental Parameters. *Metrika 15*, 141–148.

Pfanzagl, J. (1993). Incidental versus Random Nuisance Parameters. *The Annals of Statistics 21*, 1663–1691.

Rabe-Hasketh, S. and A. Skrondal (2006). Multilevel modelling of complex data survey. *Journal of the Royal Statistical Society, Series A 169*, 805–827.

Rasch, G. (1960a). On general laws and the meaning of measurement in psychology. In *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, pp. 321–333. University of California Press.

Rasch, G. (1960b). *Probabilistic Models for Some Intelligence and Attainment Tests*. The Danish Institute for Educational Research (Expanded Edition, 1980, Chicago: The University Chicago Press).

Reichenbach, H. (1956). *The direction of time*. University CaliforniaPress, Berkeley.

San Martín, E., J. González, and Tuerlinckx (2009). Identified Parameters, Parameters of Interest and Their Relationships. *Measurement: Interdisciplinary Research & Perspective 7:2*, 97 – 105.

San Martín, E. and J. M. Rolin (2007). Identifiability and Estimability of Parametric Rasch-Type Models. *Submitted*.

Tuerlinckx, F., F. Rijmen, W. Molenberghs, M. Meulders, and P. De Boeck (2004). Estimation and software. In P. De Boeck and M. Wilson (Eds.), *Explanatory Item Response Models: A Generalized Linear and Nonlinear Approach*. Springer.

Verbeke, G. and E. Lesaffre (1996). A linear mixed-effects model with heterogeneity in the random effects population. *Journal of the American Statistical Association 91*, 217–221.