Chapter 5
Network Psychometrics in Educational Practice

Maximum Likelihood Estimation of the Curie-Weiss Model

M. Marsman, C. C. Tanis, T. M. Bechger and L. J. Waldorp

Abstract In network psychometrics undirected graphical models—such as the Ising model from statistical physics—are used to characterize the manifest probability distribution of psychometric data. In practice, we often find that it is extremely difficult to apply graphical models as the Ising model to educational data because (i) the model's likelihood is impossible to compute for the big data that we typically observe in educational measurement, and (ii) the model cannot handle the partially observed data that stem from incomplete test designs. In this chapter, we therefore propose to use a simplified Ising model that is known as the Curie-Weiss model. Unlike the more general Ising model, the Curie-Weiss model is computationally tractable, which makes it suitable for applications in educational measurement. The objective of this chapter is to study the statistical properties of the Curie-Weiss model and discuss its estimation with complete or incomplete data. We demonstrate that our procedures work using a simulated example, and illustrate the analysis of fit of the Curie-Weiss model using real data from the 2012 Cito Eindtoets.

5.1 Introduction

The item response theory (IRT) model is ubiquitous in the analysis of pupil responses to the questions in educational tests. In the model, a latent variable is posited to represent the ability of a pupil that is measured by the test, and then characterizes the probability that the pupil responds correctly or incorrectly to the test questions as a function of his or her ability. But these abilities are never directly observed. What we can estimate directly, however, are the proportion of pupils in a given population

M. Marsman (✉) · C. C. Tanis · L. J. Waldorp
Psychological Methods, University of Amsterdam, Nieuwe Achtergracht 129B,
PO Box 15906, 1001 NK Amsterdam, The Netherlands
e-mail: m.marsman@uva.nl

T. M. Bechger
Cito, Arnhem, The Netherlands

that obtain particular configurations of correct and incorrect responses to the test questions (Cressie and Holland 1983). Modeling these *manifest probabilities* has been the focus of several areas in the psychometric literature, such as that related to the Dutch Identity (Holland 1990; Hessen 2012; Ip 2002), log-multiplicative association models (Anderson and Vermunt 2000; Anderson and Yu 2007), and marginal models (Bergsma 1997; Bergsma and Rudas 2002). A recent addition to the psychometric literature on modeling manifest probability distributions is the network psychometric approach (van der Maas et al. 2006; Borsboom 2008; Epskamp 2017), which utilizes undirected graphical models to characterize the manifest probabilities and posits observables (e.g., responses to test questions) on a graphical or network structure.

In network psychometrics, the Ising (1925; Lenz 1920) model is a commonly used graphical model for binary random variables X, which may be used to encode the correct ($X = 1$) and incorrect ($X = 0$) responses to the questions in the test. The model is characterized by the following probability distribution over the configurations of k binary variables $\mathbf{X}$,

$$p\left(\mathbf{X} = \mathbf{x}\right) = \frac{\exp\left(\mathbf{x}^{\mathsf{T}}\boldsymbol{\mu} + \mathbf{x}^{\mathsf{T}}\boldsymbol{\Sigma}\mathbf{x}\right)}{\sum_{\mathbf{x}}\exp\left(\mathbf{x}^{\mathsf{T}}\boldsymbol{\mu} + \mathbf{x}^{\mathsf{T}}\boldsymbol{\Sigma}\mathbf{x}\right)}, \tag{5.1}$$

where the sum in the denominator ranges across all possible configurations $\mathbf{x}$ of $\mathbf{X}$, $\boldsymbol{\Sigma} = [\sigma_{ij}]$ is the symmetric $k \times k$ *connectivity matrix* that encodes the strength of interactions between the response variables—i.e., the network structure—and $\boldsymbol{\mu}$ is the $k \times 1$ vector that encodes influences from outside the network. In an educational testing context, the σ_{ij} may be taken to represent the latent processes that are shared by questions i and j, and μ_i may be taken to represent factors that are attributed to a specific question—e.g., its difficulty. This interpretation of the Ising model comes from recent work that characterizes it as the marginal distribution of a multidimensional IRT (MIRT) model (Marsman et al. 2015; Epskamp et al. 2018),

$$p\left(\mathbf{X} = \mathbf{x}\right) = \int_{\mathbb{R}^k} \prod_{i=1}^{k} p\left(X_i = x_i \mid \boldsymbol{\theta}\right) f\left(\boldsymbol{\theta}\right) \, \mathrm{d}\boldsymbol{\theta},$$

where $-\mu_i$ was shown to be equal to the MIRT model's difficulty parameter, and the square root of the r-th term in the eigenvalue decomposition of the connectivity matrix,

$$\boldsymbol{\Sigma} = \sum_{r=1}^{k} \lambda_r \mathbf{q}\mathbf{q}^{\mathsf{T}},$$

was shown to be equal to the loadings of the vector of responses on the r-th ability dimension. Viewed in this way, the two-parameter logistic (2PL) model corresponds to an Ising model with a rank one connectivity matrix.

The use of graphical models such as the Ising model in large-scale educational applications remains problematic, however. One major issue with the Ising model,

for instance, is that the model is computationally intractable except for very small or highly constrained problems. This intractability resides in the model's normalizing constant, which requires the evaluation of 2^k distinct terms. With as little as 20 questions there already are over one million terms that need to be evaluated, and educational tests often consist of much more than 20 questions. This is particularly problematic for estimating the model's parameters, since closed form expressions are unavailable, and iterative procedures are needed to estimate them. The model's normalizing constant then needs to be evaluated several times in each of the iterations. Another important issue is that the Ising model consists of too many unknown parameters. With only $k = 20$ questions on the test there already are $\frac{1}{2}\left(k^2 + k\right) = 210$ free parameters, and with $k = 50$ test questions there are over one thousand free parameters. When the number of questions in the test increases, both the number of terms that are evaluated in the normalizing constant and the number of free parameters quickly grow. This makes the Ising model impractical for the large applications that are often encountered in educational measurement. Finally, the Ising model is not closed under marginalization (Marsman et al. 2017); that is,

$$\underbrace{p\left(\mathbf{X}^{(i)} = \mathbf{x}^{(i)}\right)}_{\text{not Ising model}} = \sum_{x_i} \underbrace{p\left(\mathbf{X} = \mathbf{x}\right)}_{\text{Ising model}},$$

where $\mathbf{x}^{(i)}$ is the vector $\mathbf{x}$ without element i. As a result, it is complicated to handle data that is collected with the incomplete test designs that are commonly used in educational testing.

As a way to work around these problems we will use a simplified Ising model that is known as the Curie-Weiss model (Kac 1968), in which the connectivity matrix is the scalar

$$\mathbf{\Sigma} = \sigma\, \mathbf{1}_k,$$

where $\mathbf{1}_k$ denotes a $k \times k$ matrix of ones, and there is a constant interaction $\sigma > 0$ among variables in the network. Even though the Curie-Weiss model may seem to be an oversimplification of the full Ising model, it is an incredibly rich model. For example, it provides an analytic expression for the marginal Rasch model (Bock and Aitken 1981), and is a special case of the extended Rasch model (Tjur 1982; Cressie and Holland 1983), two well-known psychometric models for the manifest probability distribution. But it also remains of interest in theoretical physics (e.g., Kochmański et al. 2013), and it is strongly related to the mean-field approximation that is often used to study theoretical properties of Ising models in the context of, for instance, magnetism. Moreover, it offers a pragmatic approximation to the full Ising model, as the Ising network can be factored into cliques—fully connected subgraphs—and the distribution of variables in such cliques can be closely approximated with a Curie-Weiss model. What matters here is that, unlike the full Ising model, the Curie-Weiss model is computationally tractable, which makes it suitable for applications in educational measurement.

The objective of the current chapter is to study the statistical properties of the Curie-Weiss model and discuss its estimation with complete or incomplete data. First, we introduce the model and analyze its statistical properties. In particular, we examine what happens when we either condition on or marginalize a part of the Curie-Weiss network. This also provides us with the opportunity to discuss its relation to the two Rasch-type models, and how its properties relate to these two models. Hereafter we show how to estimate the Curie-Weiss model and its asymptotic standard errors in both the complete data case and the incomplete data case, and then illustrate our procedures using simulated data. Data from a large-scale educational testing application—the 2012 Cito Eindtoets—is used to illustrate model fit procedures. We end the chapter with a discussion.

5.2 The Curie-Weiss Model

In a now famous series of lectures (Chrétien et al. 1968), Marc Kac lectured about the mathematical mechanisms of phase transitions (Kac 1968; Kac and Thompson 1966), and in particular the phase transitions that are observed in the study of magnetism. The Ising model is often used in this context to study dynamic properties such as the shift from a non-magnetic to a magnetic state when the material is cooled. But since it is very difficult to produce analytic results with the Ising model, Kac proposed to start with a simpler model based on the theories of magnetism of Pierre Curie and Pierre-Ernest Weiss. His version of the Curie-Weiss model is characterized by the probability distribution over the configurations of k random variables Y that take values in $\{-1, +1\}$,

$$p\left(\mathbf{Y} = \mathbf{y}\right) = \frac{\exp\left(\frac{J}{k} \sum_i \sum_j y_i y_j\right)}{\sum_{\mathbf{y}} \exp\left(\frac{J}{k} \sum_i \sum_j y_i y_j\right)},$$

where the interaction strength J/k depends on the size of the network and is constant throughout the network. In this context, the variables Y refer to the magnetic moments of electrons, which may point up ($Y = +1$) or point down ($Y = -1$). Since every variable in the Curie-Weiss network is related to all other variables, it is sometimes referred to as the *fully-connected Ising network* (Gould and Tobochnik 2010). For a detailed analysis of its dynamical properties we refer the interested reader to recent work by Kochmański et al. (2013).

Our interest in this chapter will focus on the statistical properties of the Curie-Weiss model, which inspired us to work with the following version of it

$$p\left(\mathbf{X} = \mathbf{x}\right) = \frac{\exp\left(\sum_i x_i \mu_i + \sigma \sum_i \sum_j x_i x_j\right)}{\sum_{\mathbf{x}} \exp\left(\sum_i x_i \mu_i + \sigma \sum_i \sum_j x_i x_j\right)}, \tag{5.2}$$

which differs in three important ways from the original version that was introduced by Kac. Firstly, our version of the Curie-Weiss model is used to model the distribution of $\{0, 1\}$-variables X instead of $\{-1, +1\}$-variables Y. This formulation suits the typical $\{0, 1\}$ coding that is used to score responses to educational test items. Secondly, we have introduced the *external fields* $\boldsymbol{\mu}$ (c.f. Kochmański et al. 2013), which are used here to model differences in item difficulty. Moreover, since our interest is focused on its statistical instead of its dynamical properties—e.g., phase transitions—we will not investigate networks that increase in size. Therefore we have simplified the association strength from J/k to σ so that it does not explicitly depend on the network's size.

5.2.1 Some Statistical Properties of the Curie-Weiss Model

Before we continue with the problem of estimating the model's parameters $\boldsymbol{\mu}$ and σ, we will first review the model and its conditioning and marginalization properties. From our formulation of the Curie-Weiss model in Eq. (5.2) we readily find the simplified expression,

$$
\begin{aligned}
p\left(\mathbf{X}=\mathbf{x}\right) &= \frac{\exp\left(\sum_i x_i\mu_i + \sigma \sum_i \sum_j x_ix_j\right)}{\sum_{\mathbf{x}} \exp\left(\sum_i x_i\mu_i + \sigma \sum_i \sum_j x_ix_j\right)} \\
&= \frac{\left(\prod_{i=1}^k \exp\left(x_i\mu_i\right)\right)\exp\left(\sigma x_+^2\right)}{\sum_{\mathbf{x}}\left(\prod_{i=1}^k \exp\left(x_i\mu_i\right)\right)\exp\left(\sigma x_+^2\right)} \\
&= \frac{\left(\prod_{i=1}^k \exp\left(x_i\mu_i\right)\right)\exp\left(\sigma x_+^2\right)}{\sum_{s=0}^k \gamma_s\left(\boldsymbol{\mu}\right)\exp\left(\sigma s^2\right)},
\end{aligned}
\tag{5.3}
$$

where x_+ is used to denote the sum score $\sum_i x_i$, and $\gamma_s\left(\boldsymbol{\mu}\right)$ is used to denote the elementary symmetric function of order s of the vector $\boldsymbol{\mu}$. The elementary symmetric function of order s of the vector $\boldsymbol{\mu}$ is defined here as

$$
\gamma_s\left(\boldsymbol{\mu}\right) = \sum_{\mathbf{x}:\, x_+=s} \prod_{i=1}^k \exp\left(x_i\mu_i\right),
$$

where the sum ranges across all configurations of $\mathbf{x}$ for which the total score x_+ is equal to s.

Observe that the normalizing constant of this version of the Curie-Weiss model

$$\sum_{s=0}^{k} \gamma_s(\boldsymbol{\mu}) \exp\left(\sigma s^2\right),$$

is linear in the number of variables in the network. Even though this expression depends on the elementary symmetric functions $\gamma_s(\boldsymbol{\mu})$, their computation using, for example, the summation algorithm (Fischer 1974; Verhelst et al. 1984) is of a quadratic order of complexity (Baker and Harwell 1996). As a result, the computation of the normalizing constant also has a quadratic order of complexity, which is a huge improvement over the exponential order of complexity of computing the normalizing constant of the Ising model. As a result, the normalizing constant of the Curie-Weiss model can be efficiently computed.

Let $A \subset \Omega = \{1, 2, \ldots, k\}$ be a subset of the variables and $\bar{A}$ its complement, such that $\Omega = A \cup \bar{A}$. Consider the conditional distribution $p\left(\mathbf{X}^{(A)} \mid \mathbf{x}^{(\bar{A})}\right)$ of the variables in subset A given the remaining variables. We find

$$
p\left(\mathbf{X}^{(A)} = \mathbf{x}^{(A)} \mid \mathbf{x}^{(\bar{A})}\right) = \frac{p\left(\mathbf{x}^{(A)}, \mathbf{x}^{(\bar{A})}\right)}{\sum_{\mathbf{x}^{(A)}} p\left(\mathbf{x}^{(A)}, \mathbf{x}^{(\bar{A})}\right)}
$$

$$
= \frac{\left(\prod_{i \in A} \exp(x_i \mu_i)\right) \exp\left(\sigma\left[\left(x_+^{(A)}\right)^2 + 2x_+^{(A)} x_+^{(\bar{A})}\right]\right)}{\sum_{\mathbf{x}^{(A)}} \left(\prod_{i \in A} \exp(x_i \mu_i)\right) \exp\left(\sigma\left[\left(x_+^{(A)}\right)^2 + 2x_+^{(A)} x_+^{(\bar{A})}\right]\right)}
$$

$$
= \frac{\left(\prod_{i \in A} \exp\left(x_i\left[\mu_i + 2\sigma x_+^{(\bar{A})}\right]\right)\right) \exp\left(\sigma\left(x_+^{(A)}\right)^2\right)}{\sum_{r=0}^{|A|} \gamma_r\left(\boldsymbol{\mu}^{(A)} + 2\sigma x_+^{(\bar{A})}\right) \exp\left(\sigma r^2\right)},
$$

where $x_+^{(A)} = \sum_{i \in A} x_i$ denotes the sum score on the item set $A \subset \Omega$, and r the index of the rest-score that ranges from zero to the size of the set A. Note that we have used a well-known property of elementary symmetric functions (e.g., Verhelst et al. 1984). Namely, that

$$\gamma_s(\boldsymbol{\mu} + c) = \exp(sc)\, \gamma_s(\boldsymbol{\mu}),$$

such that

$$
\gamma_r\left(\boldsymbol{\mu}^{(A)} + 2\sigma x_+^{(\bar{A})}\right) = \sum_{\mathbf{x}^{(A)}: x_+^{(A)} = r} \prod_{i \in A} \exp\left(x_i\left[\mu_i + 2\sigma x_+^{(\bar{A})}\right]\right)
$$

$$
= \exp\left(r\, 2\sigma x_+^{(\bar{A})}\right) \sum_{\mathbf{x}^{(A)}: x_+^{(A)} = r} \prod_{i \in A} \exp(x_i \mu_i)
$$

is the elementary symmetric function of order r of the sub-vector $\boldsymbol{\mu}^{(A)}$ shifted by the constant $2\sigma x_+^{(\bar{A})}$. Since the conditional distribution is a Curie-Weiss model it follows that the model is *closed under conditioning*. That is, the distribution of variables in a Curie-Weiss network conditional upon a subset of the variables in the network results in a Curie-Weiss model. Furthermore, $p\left(\mathbf{X}^{(A)} = \mathbf{x}^{(A)} \mid \mathbf{x}^{(\bar{A})}\right) = p\left(\mathbf{X}^{(A)} = \mathbf{x}^{(A)} \mid x_+^{(\bar{A})}\right)$.

Of particular interest is the conditional distribution of one variable x_j conditional upon the remaining variables $\mathbf{x}^{(j)}$,

$$p\left(X_j = x_j \mid \mathbf{x}^{(j)}\right) = \frac{\exp\left(x_j \left[\mu_j + \sigma + 2\sigma x_+^{(j)}\right]\right)}{1 + \exp\left(\mu_j + \sigma + 2\sigma x_+^{(j)}\right)}, \tag{5.4}$$

which depends on the remaining variables only through the rest score $x_+^{(j)} = \sum_{i \neq j} x_i$ and does not depend on the external fields $\boldsymbol{\mu}^{(j)}$ that are associated to the $k - 1$ variables we conditioned on. In sum, Eq. (5.4) provides an analytic expression of the item-rest regressions that can be useful for assessing the fit of the Curie-Weiss model. The conditional distribution also reveals how the Curie-Weiss model operates: The field μ_j models the general tendency to respond correctly or incorrectly to item j, and the interaction parameter σ scales the influence of the remaining $k - 1$ response variables $\mathbf{x}^{(j)}$ on the response to item j. Observe that the tendency to respond correctly increases with both μ_j and σ.

While the Curie-Weiss model is closed under conditioning, it is not closed under marginalization. To see this, we derive the expression for the marginal distribution of the first $k - 1$ variables of a k variable Curie-Weiss network

$$p\left(\mathbf{X}^{(k)} = \mathbf{x}^{(k)}\right) = p\left(x_k = 1, \mathbf{X}^{(k)} = \mathbf{x}^{(k)}\right) + p\left(x_k = 0, \mathbf{X}^{(k)} = \mathbf{x}^{(k)}\right)$$

$$= \frac{\left\{\exp\left(\mu_k + \sigma \left[1 + 2x_+^{(k)}\right]\right) + 1\right\} \left(\prod_{i=1}^{k-1} \exp\left(x_i \mu_i\right)\right) \exp\left(\sigma \left(x_+^{(k)}\right)^2\right)}{\sum_{s=0}^{k} \gamma_s\left(\boldsymbol{\mu}\right) \exp\left(\sigma s^2\right)}.$$

Using the recursive property of elementary symmetric functions (Fischer 1974, p. 250):

$$\gamma_s\left(\boldsymbol{\mu}\right) = \exp\left(\mu_k\right) \gamma_{s-1}\left(\boldsymbol{\mu}^{(k)}\right) + \gamma_s\left(\boldsymbol{\mu}^{(k)}\right)$$

we can simplify this expression to

$$p\left(\mathbf{X}^{(k)} = \mathbf{x}^{(k)}\right) = \frac{\left\{\exp\left(\mu_k + \sigma \left[1 + 2x_+^{(k)}\right]\right) + 1\right\} \left(\prod_{i=1}^{k-1} \exp\left(x_i \mu_i\right)\right) \exp\left(\sigma \left(x_+^{(k)}\right)^2\right)}{\sum_{r=0}^{k-1} \left\{\exp\left(\mu_k + \sigma \left[1 + 2r\right]\right) + 1\right\} \gamma_r\left(\boldsymbol{\mu}^{(k)}\right) \exp\left(\sigma r^2\right)}, \tag{5.5}$$

Since we cannot factor out the terms that depend on variable i from the sum in the denominator due to its interaction with the rest score $x_+^{(k)}$, we are unable to simplify this expression to the form of a Curie-Weiss model. It follows that the Curie-Weiss model is *not closed under marginalization*.

5.2.2 The Curie-Weiss to Rasch Connection

Equation (5.3) can be written as:

$$p\left(\mathbf{X} = \mathbf{x}\right) = \frac{\left(\prod_{i=1}^{k} \exp\left(x_i \mu_i\right)\right)}{\gamma_{x_+}\left(\boldsymbol{\mu}\right)} \frac{\gamma_{x_+}\left(\boldsymbol{\mu}\right) \exp\left(\sigma x_+^2\right)}{\sum_{s=0}^{k} \gamma_s\left(\boldsymbol{\mu}\right) \exp\left(\sigma s^2\right)}$$
$$= p\left(\mathbf{X} = \mathbf{x} \mid X_+ = x_+\right) p\left(X_+ = x_+\right), \tag{5.6}$$

where $p\left(\mathbf{X} = \mathbf{x} \mid X_+ = x_+\right)$ can be recognized as the conditional likelihood function of the Rasch (1960) model (Andersen 1973) with item difficulties $-\mu_i$. In fact, it is readily seen that our Curie-Weiss model is a special case of the *extended Rasch model* (ERM; Tjur 1982; Cressie and Holland 1983). In general, the ERM is characterized by the following distribution

$$p\left(\mathbf{X} = \mathbf{x}\right) = \frac{\prod_{i=1}^{k} \exp(x_i \mu_i) \, \lambda_{x_+}}{\sum_{s=0}^{k} \gamma_s\left(\boldsymbol{\mu}\right) \, \lambda_s}, \tag{5.7}$$

which equals the Curie-Weiss model when $\lambda_s = \exp\left(\sigma s^2\right)$. An empirical illustration of such a quadratic relation can be found in Brinkhuis (in press, Chap. 5).

Importantly, the ERM can be expressed as a *marginal Rasch model* (MRM; Bock and Aitken 1981) iff it's score parameters λ_s form a moment sequence (Cressie and Holland 1983). That our Curie-Weiss model is an MRM can be seen from the original derivation by Kac, who used the following well-known identity (Stratonovich 1957; Hubbard 1959),

$$\exp\left(\sigma s^2\right) = \int_{\mathbb{R}} \frac{1}{\sqrt{\pi}} \exp\left(2\sqrt{\sigma} \, s \, \eta - \eta^2\right) \mathrm{d}\eta,$$

in which we replace the exponential of a square with the integral on the right hand side—the expectation $\mathbb{E}\left(\exp\left(2\sqrt{\sigma} \, s \, H\right)\right)$ of the normal random variable H. Writing the right hand side of this identity for the squared exponential in the numerator of the Curie-Weiss model gives

$$p\left(\mathbf{X}=\mathbf{x}\right)=\frac{\left(\prod_{i=1}^{k}\exp\left(x_{i}\mu_{i}\right)\right)\int_{\mathbb{R}}\frac{1}{\sqrt{\pi}}\exp\left(2\sqrt{\sigma}\,x_{+}\,\eta-\eta^{2}\right)\mathrm{d}\eta}{\sum_{s=0}^{k}\gamma_{s}\left(\boldsymbol{\mu}\right)\exp\left(\sigma s^{2}\right)}$$

$$=\int_{\mathbb{R}}\prod_{i=1}^{k}\exp\left(x_{i}\left[\mu_{i}+2\sqrt{\sigma}\eta\right]\right)\frac{\exp\left(-\eta^{2}\right)}{\sqrt{\pi}\sum_{s=0}^{k}\gamma_{s}\left(\boldsymbol{\mu}\right)\exp\left(\sigma s^{2}\right)}\mathrm{d}\eta$$

$$=\int_{\mathbb{R}}\prod_{i=1}^{k}\frac{\exp\left(x_{i}\left[\mu_{i}+2\sqrt{\sigma}\eta\right]\right)}{1+\exp\left(\mu_{i}+2\sqrt{\sigma}\eta\right)}\frac{\prod_{i=1}^{k}\left\{1+\exp\left(\mu_{i}+2\sqrt{\sigma}\eta\right)\right\}\exp\left(-\eta^{2}\right)}{\sqrt{\pi}\sum_{s=0}^{k}\gamma_{s}\left(\boldsymbol{\mu}\right)\exp\left(\sigma s^{2}\right)}\mathrm{d}\eta$$

$$=\int_{\mathbb{R}}\prod_{i=1}^{k}\frac{\exp\left(x_{i}\left[\mu_{i}+\theta\right]\right)}{1+\exp\left(\mu_{i}+\theta\right)}\frac{\prod_{i=1}^{k}\left\{1+\exp\left(\mu_{i}+\theta\right)\right\}\exp\left(-\frac{1}{4\sigma}\theta^{2}\right)}{\sqrt{4\sigma\pi}\sum_{s=0}^{k}\gamma_{s}\left(\boldsymbol{\mu}\right)\exp\left(\sigma s^{2}\right)}\mathrm{d}\theta$$

$$=\int_{\mathbb{R}}\prod_{i=1}^{k}p\left(X_{i}=x_{i}\mid\theta\right)f\left(\theta\right)\mathrm{d}\theta,$$

where we have used the change of variable $\theta=2\sqrt{\sigma}\eta$. Noting that $p\left(X_{i}=x_{i}\mid\theta\right)$ is a Rasch model with item difficulties $-\mu_{i}$, it follows that the Curie-Weiss model corresponds to a marginal Rasch model albeit with a rather peculiar[1] latent variable distribution $f\left(\theta\right)$ that depends on the item parameters. This distribution closely resembles a normal distribution or a skewed-normal distribution (Marsman et al. 2018; Haslbeck et al. 2018), depending on the value of the scaling parameter σ.

It is easy to verify that both the ERM and the MRM are closed under marginalization (e.g., Maris et al. 2015) whereas, as we showed earlier, the Curie-Weiss model is not. Specifically, the marginal distribution in Eq. (5.5) is not a Curie-Weiss model, although it is an ERM with

$$\lambda_{r}=\exp\left(\sigma r^{2}\right)\left[1+\exp\left(\mu_{k}+\sigma\left(1+2r\right)\right)\right],$$

for $r=0,\ldots,k-1$. Thus, marginalization gives us an ERM, but not a Curie-Weiss model. The reason that the ERM in Eq. (5.7) is closed under marginalization yet the Curie-Weiss model is not is because the characterization $\lambda_{s}=\exp\left(\sigma s^{2}\right)$ introduces a dependency between the score parameters that is not found in Eq. (5.7). The issue with the MRM is slightly different. Observe first that the MRM itself is closed under marginalization (Marsman et al. 2017), such that

$$p\left(\mathbf{X}^{(k)}=\mathbf{x}^{(k)}\right)=\int_{\mathbb{R}}\sum_{x_{k}}\prod_{i=1}^{k}p\left(X_{i}=x_{i}\mid\theta\right)f\left(\theta\right)\mathrm{d}\theta=\int_{\mathbb{R}}\prod_{i=1}^{k-1}p\left(X_{i}=x_{i}\mid\theta\right)f\left(\theta\right)\mathrm{d}\theta,$$

where the latent trait model is the Rasch model. In the Curie-Weiss model, the latent variable distribution $f\left(\theta\right)$ explicitly depends on the model parameters, including the

---

[1] A similar construction of the latent variable distribution can also be found in, for instance, Cressie and Holland (1983) and McCullagh (1994), and is used by Marsman et al. (in press) to generalize the Dutch Identity (Holland 1990).

field μ_k of variable k. Even though this ought not be a problem in practical applications with incomplete test designs, it does show why the marginal is not a Curie-Weiss model, since it is clear that the expectations $\mathbf{E}_f\left[p\left(\mathbf{x}\mid\Theta\right)\right]$ and $\mathbf{E}_f\left[p\left(\mathbf{x}^{(k)}\mid\Theta\right)\right]$ differ.

Compared to a regular MRM, the Curie-Weiss model has a number of convenient properties. One is that the origin of the ability distribution is identifiable so that absolute values of the item difficulties can be interpreted. Furthermore, the posterior distribution of the latent variable is available in closed form:

$$f\left(\theta\mid\mathbf{X}=\mathbf{x}\right)=f\left(\theta\mid X_+=x_+\right)=\frac{1}{\sqrt{4\sigma\pi}}\exp\left(-\frac{1}{4\sigma}\left(\theta-2\sigma x_+\right)^2\right).$$

This is a normal distribution[2] with mean $2\sigma x_+$ and variance 2σ. Sampling so-called plausible values is thus trivial in the Curie-Weiss model. If we write the association parameter in its original form, $\sigma=J/k$, the posterior is a normal distribution with mean $2J\bar{x}$ and variance $2J/k$:

$$f\left(\theta\mid\mathbf{X}=\mathbf{x}\right)=f\left(\theta\mid X_+=x_+\right)=\frac{\sqrt{k}}{\sqrt{4J\pi}}\exp\left(-\frac{k}{4J}\left(\theta-2J\bar{x}\right)^2\right),$$

Note that the posterior standard deviation now shrinks with a rate of $1/\sqrt{k}$. Note further that the posterior variance is the same for all test scores, x_+, which suggests that test information is the same for all values of the latent variable. This constant rate of information is more in line with classical test theory, for example, than with regular IRT models where information is larger for abilities close to the item difficulties.

Another convenient property of the Curie-Weiss model is that it provides analytic expressions for the distribution of observables, and in particular the item-rest regressions—e.g., Eq. (5.4)—and the distribution of test scores—e.g., the second factor in Eq. (5.6). The benefit of having analytic expressions for these distributions is that they can be used to assess the fit of the Curie-Weiss model.

5.3 Maximum Likelihood Estimation of the Curie-Weiss Model

Maximum likelihood (ML) estimation of the Curie-Weiss model has been worked out for the case of a single realization of the Curie-Weiss network with an external field μ that is the same for each variable in the network. It is easily shown that for this constrained $n=1$ case the two parameters σ and μ cannot be consistently estimated (e.g., Comets and Gidas 1991). In psychometrics and educational measurement, however, we typically have many replications of the Curie-Weiss network, i.e., $n\gg1$.

---

[2]Since the Rasch model is in the exponential family, the posterior ability distribution depends on the data only through the sufficient statistic X_+ (Dawid 1979).

We will focus here on the ML procedure for estimating μ and σ; first for the complete data case, then for the incomplete data case. Sample R-code is provided in an online repository located at https://osf.io/4m3dq/.

5.3.1 Maximum Likelihood in the Complete Data Case

The factorization in Eq. (5.6) shows that the external fields of the Curie-Weiss model can be consistently estimated with conditional maximum likelihood (CML) as used for the regular Rasch model (Andersen 1970, 1973). We may then estimate the association strength conditional upon the CML-estimates of the external fields with little loss of information (Eggen 2000). This is complicated, however, by the fact that the parameters of the Curie-Weiss model are identified, but the parameters of the conditional Rasch model are not.[3] We will therefore focus on joint estimation of the model parameters.

The complete data likelihood for the Curie-Weiss parameters μ and σ based on the responses of n pupils to k test questions is

$$L\left(\mu, \sigma \mid \mathbf{x}_1, \ldots, \mathbf{x}_n\right) = \prod_{p=1}^{n} \frac{\left(\prod_{i=1}^{k} \exp\left(x_{pi}\mu_i\right)\right) \exp\left(\sigma x_{p+}^2\right)}{\sum_{s=0}^{k} \gamma_s\left(\mu\right) \exp\left(\sigma s^2\right)}$$

$$= \frac{\exp\left(\sum_{i=1}^{k} x_{+i}\mu_i + \sigma \sum_{p=1}^{n} x_{p+}^2\right)}{\left(\sum_{s=0}^{k} \gamma_s\left(\mu\right) \exp\left(\sigma s^2\right)\right)^n},$$

and only depends on the (sufficient) statistics x_{+i}, for $i = 1, \ldots, k$, and $\sum_p x_{p+}^2$. We seek values $\hat{\mu}$ and $\hat{\sigma}$ that maximize the likelihood function, or, similarly, the roots of its gradient $\nabla \ln L\left(\mu, \sigma \mid \mathbf{x}_1, \ldots \mathbf{x}_n\right)$. The roots of the gradient can be found using iterative procedures such as the Newton-Raphson (NR) procedure. Unfortunately, NR procedures require the computation of the Hessian matrix of mixed partial (second order) derivatives in every iteration to update parameter values, which can be expensive to compute in practice. We therefore choose a *divide and conquer* strategy and maximize each of the parameters in turn, while ignoring cross-parameter dependency during optimization. Even though this might slow down convergence, it circumvents having to compute (and invert) the complete Hessian matrix in every iteration.

We first investigate the maximization of $\ln L\left(\mu, \sigma \mid \mathbf{x}_1, \ldots \mathbf{x}_n\right)$ with respect to the external field μ_i of the response variable i, fixing the parameter values of the remaining k parameters to their current estimates. The partial derivative of

---

[3]This implies that we lose a degree of freedom by factoring the joint distribution and condition on the observed test score. To consistently estimate the parameters of the Curie-Weiss model using a two-step procedure, we therefore need to estimate a shift of the external fields in the test score distributions $p\left(X_+ = x_+\right)$.

$\ln L (\boldsymbol{\mu}, \sigma \mid \mathbf{x}_1, \ldots \mathbf{x}_n)$ with respect to μ_i is

$$\frac{\partial}{\partial \mu_i} \ln L (\boldsymbol{\mu}, \sigma \mid \mathbf{x}_1, \ldots \mathbf{x}_n) = x_{+i} - n \exp (\mu_i) \frac{\sum_{s=0}^{k-1} \gamma_s \left(\boldsymbol{\mu}^{(i)}\right) \exp \left(\sigma s^2\right)}{\sum_{s=0}^{k} \gamma_s \exp \left(\sigma s^2\right)}.$$

where we have used the following well-known property of the elementary symmetric function,

$$\frac{\partial}{\partial \mu_i} \gamma_s (\boldsymbol{\mu}) = \exp (\mu_i) \gamma_{s-1} \left(\boldsymbol{\mu}^{(i)}\right).$$

Setting the derivative to zero we obtain the following closed form expression for the parameter μ_i:

$$\mu_i = \ln \left(\frac{x_{+i}}{n - x_{+i}}\right) + \ln \left(\frac{\sum_{s=0}^{k-1} \gamma_s^{(i)} \exp \left(\sigma s^2\right)}{\sum_{s=0}^{k-1} \gamma_s^{(i)} \exp \left(\sigma (s+1)^2\right)}\right),$$

where the first term is a function of the sufficient statistics and the second term is a function of the remaining k parameters of the model. In an iteration t of our numerical procedure, we fix the states of these parameters to their current estimates $\hat{\boldsymbol{\mu}}_t^{(i)}$ and $\hat{\sigma}_t$ to compute an updated value for μ_i.

What remains is the maximization of $\ln L (\boldsymbol{\mu}, \sigma \mid \mathbf{x}_1, \ldots \mathbf{x}_n)$ with respect to the association strength σ. The partial derivative of $\ln L (\boldsymbol{\mu}, \sigma \mid \mathbf{x}_1, \ldots \mathbf{x}_n)$ with respect to σ is

$$\frac{\partial}{\partial \sigma} \ln L (\boldsymbol{\mu}, \sigma \mid \mathbf{x}_1, \ldots \mathbf{x}_n) = \sum_{p=1}^{n} x_{p+}^2 - n \frac{\sum_{s=0}^{k} s^2 \gamma_s (\boldsymbol{\mu}) \exp \left(\sigma s^2\right)}{\sum_{s=0}^{k} \gamma_s (\boldsymbol{\mu}) \exp \left(\sigma s^2\right)}.$$

Setting this derivative to zero does not lead to a closed form solution for the parameter σ. We therefore propose a one-dimensional NR step:

$$\sigma = \hat{\sigma} - \frac{\frac{\partial}{\partial \sigma} \ln L (\boldsymbol{\mu}, \sigma \mid \mathbf{x}_1, \ldots \mathbf{x}_n)\big|_{\sigma = \hat{\sigma}}}{\frac{\partial^2}{\partial \sigma^2} \ln L (\boldsymbol{\mu}, \sigma \mid \mathbf{x}_1, \ldots \mathbf{x}_n)\big|_{\sigma = \hat{\sigma}}}$$

$$= \hat{\sigma} + \frac{\frac{1}{n} \sum_{p=1}^{n} x_{p+}^2 - \frac{\sum_{s=0}^{k} s^2 \gamma_s(\boldsymbol{\mu}) \exp(\hat{\sigma} s^2)}{\sum_{s=0}^{k} \gamma_s(\boldsymbol{\mu}) \exp(\hat{\sigma} s^2)}}{\frac{\sum_{s=0}^{k} s^4 \gamma_s(\boldsymbol{\mu}) \exp(\hat{\sigma} s^2)}{\sum_{s=0}^{k} \gamma_s(\boldsymbol{\mu}) \exp(\hat{\sigma} s^2)} - \left(\frac{\sum_{s=0}^{k} s^2 \gamma_s(\boldsymbol{\mu}) \exp(\hat{\sigma} s^2)}{\sum_{s=0}^{k} \gamma_s(\boldsymbol{\mu}) \exp(\hat{\sigma} s^2)}\right)^2},$$

where in an iteration t we evaluate the partial derivatives based on the current estimates $\hat{\boldsymbol{\mu}}_t$ and $\hat{\sigma}_t$ to compute our update of σ.

5.3.1.1 Asymptotic Standard Errors for the Complete Data Case

We estimate the variance-covariance matrix of our estimators for the parameters $\boldsymbol{\mu}$ and σ, by evaluating the inverse of the Fisher information matrix at the estimated values $\hat{\boldsymbol{\mu}}$ and $\hat{\sigma}$:

$$\text{Var}\left(\hat{\boldsymbol{\mu}}, \hat{\sigma}\right) \approx \left[\mathcal{I}\left(\hat{\boldsymbol{\mu}}, \hat{\sigma}\right)\right]^{-1},$$

where $\mathcal{I}\left(\hat{\boldsymbol{\mu}}, \hat{\sigma}\right)$ denotes the Fisher Information matrix. To compute the Fisher information matrix we work out (minus) the second order mixed derivatives from the Q-function. As the Curie-Weiss model is a member of the exponential family of distributions, the second derivatives will not depend on data, and we do not have to (numerically) evaluate any expectations in computing the information matrix. In Appendix 1 we present the mixed partial derivatives of the Q-function.

We compute the information matrix in four parts:

$$\mathcal{I}\left(\boldsymbol{\mu}, \sigma\right) = \begin{pmatrix} \mathcal{I}_{\boldsymbol{\mu}\cdot\boldsymbol{\mu}} & \mathcal{I}_{\boldsymbol{\mu}\cdot\sigma} \\ \mathcal{I}_{\sigma\cdot\boldsymbol{\mu}} & \mathcal{I}_{\sigma\cdot\sigma} \end{pmatrix}.$$

The main effects part $\mathcal{I}_{\boldsymbol{\mu}\cdot\boldsymbol{\mu}}$ of the information matrix may be expressed as follows:

$$\mathcal{I}_{\boldsymbol{\mu}\cdot\boldsymbol{\mu}} = \mathbf{A} - \mathbf{v}\mathbf{v}^{\mathsf{T}},$$

where the vector $\mathbf{v}$ has elements

$$v_i = \sqrt{n} \exp\left(\mu_i\right) \frac{\sum_{s=0}^{k-1} \gamma_s\left(\boldsymbol{\mu}^{(i)}\right) \exp\left(\sigma s^2\right)}{\sum_{s=0}^{k} \gamma_s\left(\boldsymbol{\mu}\right) \exp\left(\sigma\left(s+1\right)^2\right)},$$

and where $\mathbf{A}$ is a symmetric matrix with off-diagonal elements

$$A_{ij} = n \exp\left(\mu_i + \mu_j\right) \frac{\sum_{s=0}^{k-2} \gamma_s\left(\boldsymbol{\mu}^{(i,j)}\right) \exp\left(\sigma\left(s+2\right)^2\right)}{\sum_{s=0}^{k} \gamma_s\left(\boldsymbol{\mu}\right) \exp\left(\sigma s^2\right)},$$

and diagonal elements $A_{ii} = \sqrt{n}\, v_i$. The contribution of the scaling parameter $\mathcal{I}_{\sigma\cdot\sigma}$ is the scalar

$$\mathcal{I}_{\sigma\cdot\sigma} = n \frac{\sum_{s=0}^{k} s^4 \gamma_s\left(\boldsymbol{\mu}\right) \exp\left(\sigma s^2\right)}{\sum_{s=0}^{k} \gamma_s\left(\boldsymbol{\mu}\right) \exp\left(\sigma s^2\right)} - n\left(\frac{\sum_{s=0}^{k} s^2 \gamma_s\left(\boldsymbol{\mu}\right) \exp\left(\sigma s^2\right)}{\sum_{s=0}^{k} \gamma_s\left(\boldsymbol{\mu}\right) \exp\left(\sigma s^2\right)}\right)^2,$$

and we may express the vector $\mathcal{I}_{\boldsymbol{\mu}\cdot\sigma} = \mathcal{I}_{\sigma\cdot\boldsymbol{\mu}}^{\mathsf{T}}$ as

$$\mathcal{I}_{\boldsymbol{\mu}\cdot\sigma} = \mathbf{w} - \sqrt{n} \frac{\sum_{s=0}^{k} s^2 \gamma_s\left(\boldsymbol{\mu}\right) \exp\left(\sigma s^2\right)}{\sum_{s=0}^{k} \gamma_s\left(\boldsymbol{\mu}\right) \exp\left(\sigma s^2\right)} \mathbf{v},$$

where the vector $\mathbf{w}$ has elements

$$w_i = n \exp\left(\mu_i\right) \frac{\sum_{s=0}^{k-1} (s+1)^2 \, \gamma_s\left(\boldsymbol{\mu}^{(i)}\right) \, \exp\left(\sigma\,(s+1)^2\right)}{\sum_{s=0}^{k} \gamma_s\left(\boldsymbol{\mu}\right) \, \exp\left(\sigma s^2\right)}.$$

Sample R-code for computing this Fisher information matrix is provided at https://osf.io/4m3dq/.

5.3.2 Maximum Likelihood Estimation in the Incomplete Data Case

When test data are collected according to a randomized incomplete test design, the structurally missing data patterns are assumed to be missing (completely) at random (Eggen 1993; Eggen and Verhelst 2011). We may then obtain unbiased estimates of the parameters from the incomplete data likelihood (Eggen 1993; Eggen and Verhelst 2011; Rubin 1976)

$$L\left(\boldsymbol{\mu}, \sigma \mid \mathbf{x}_1^O, \ldots, \mathbf{x}_1^O\right) = \prod_{p=1}^{n} p\left(\mathbf{x}_p^O\right) = \prod_{p=1}^{n} \sum_{\mathbf{x}_p^M} p\left(\mathbf{x}_p^O, \mathbf{x}_p^M\right) = \prod_{p=1}^{n} \sum_{\mathbf{x}_p^M} p\left(\mathbf{x}_p\right),$$

where $\mathbf{x}_p^O$ denotes the observed responses for person p, $\mathbf{x}_p^M$ denotes his or her missing responses, and the complete data-likelihood $p\left(\mathbf{x}_p\right)$ is the Curie-Weiss model. When the complete data likelihood $L\left(\boldsymbol{\mu}, \sigma \mid \mathbf{x}_1, \ldots, \mathbf{x}_n\right)$ corresponds to the Curie-Weiss model, the incomplete data likelihood $L\left(\boldsymbol{\mu}, \sigma \mid \mathbf{x}_1^O, \ldots, \mathbf{x}_n^O\right)$ does not, because the Curie-Weiss model is not closed under marginalization. As a result, estimating the parameters of the model does not simplify to the procedure that we outlined for the complete data case.

To come to a tractable approach for estimating the Curie-Weiss parameters in the incomplete data case, we will use the Expectation-Maximization (EM) algorithm (Dempster et al. 1977), treating the unobserved responses $\mathbf{x}_1^M, \ldots, \mathbf{x}_n^M$ as missing (completely) at random. The EM approach alternates between two steps: an Expectation or E-step in which we compute the expected complete data log-likelihood of the parameters $\boldsymbol{\mu}$ and σ, and a maximization or M-step in which we find the parameter values $\hat{\boldsymbol{\mu}}$ and $\hat{\sigma}$ that maximize the expected log-likelihood that was found in the E-step. We will outline a general procedure for which the missing data patterns are unique to the individuals—e.g., as those obtained from computerized adaptive testing (van der Linden and Glas 2002; Eggen 2004; van der Linden and Glas 2010)—but note that the equations and their computations simplify considerably for the simple testing designs that are often encountered, where subsets of the items are allocated to "booklets" and pupils are allocated to one of these booklets (e.g., Eggen 1993).

5.3.2.1 The E-Step

Under the assumption that the responses of individual subjects are independent, we can write the complete data log-likelihood as

$$\ln L\left(\mu, \sigma \mid \mathbf{x}_1^M, \ldots, \mathbf{x}_n^M, \mathbf{x}_1^O, \ldots, \mathbf{x}_n^O\right) = \sum_{p=1}^{n} \ln L\left(\mu, \sigma \mid \mathbf{x}_p^M, \mathbf{x}_p^O\right),$$

and we may write the log-likelihood of person p as

$$\ln L\left(\mu, \sigma \mid \mathbf{x}_p^M, \mathbf{x}_p^O\right) = \sum_{i=1}^{k} \mu_i \left[x_{pi}^O + x_{pi}^M\right] + \sigma \left[x_{p+}^O + x_{p+}^M\right]^2$$

$$- \ln \left(\sum_{s=0}^{k} \gamma_s(\mu) \exp\left(\sigma s^2\right)\right),$$

where we use

$$x_{pi}^O = \begin{cases} 1 & \text{the observed response of person } p \text{ on item } i \text{ was correct,} \\ 0 & \text{the observed response of person } p \text{ on item } i \text{ was incorrect,} \\ 0 & \text{no response of person } p \text{ on item } i \text{ was observed,} \end{cases}$$

and

$$x_{pi}^M = \begin{cases} 1 & \text{the unobserved response of person } p \text{ on item } i \text{ was correct,} \\ 0 & \text{the unobserved response of person } p \text{ on item } i \text{ was incorrect,} \\ 0 & \text{a response of person } p \text{ on item } i \text{ was observed.} \end{cases}$$

The expected log-likelihood, or Q-function, can now be written as

$$Q\left(\mu, \sigma \; ; \hat{\mu}, \hat{\sigma}\right) = \sum_{p=1}^{n} \sum_{i=1}^{k} \mu_i \left[x_{pi}^O + \mathbb{E}_{\{\hat{\mu},\,\hat{\sigma}\}}\left\{X_{pi}^M \mid \mathbf{x}_p^O\right\}\right]$$

$$+ \sigma \sum_{p=1}^{n} \mathbb{E}_{\{\hat{\mu},\hat{\sigma}\}}\left\{\left[x_{p+}^O + X_{p+}^M\right]^2 \mid \mathbf{x}_p^O\right\}$$

$$- n \ln \left(\sum_{s=0}^{k} \gamma_s(\mu) \exp\left(\sigma s^2\right)\right),$$

and requires the computation of two expectations of the conditional distribution $p\left(\mathbf{x}^M \mid \mathbf{x}^O\right)$.

Since the Curie-Weiss model is closed under conditioning, we know that the conditional distribution $p_{\{\hat{\mu}, \hat{\sigma}\}} \left(\mathbf{x}^M \mid \mathbf{x}^O \right)$ is a Curie-Weiss model. Specifically,

$$
p_{\{\hat{\mu}, \hat{\sigma}\}} \left(\mathbf{x}_p^M \mid x_{p+}^O \right) = \frac{\exp \left(\sum_{i=1}^k x_{pi}^M \left[\hat{\mu}_i + 2\hat{\sigma} x_{p+}^O \right] + \hat{\sigma} \left(x_{p+}^M \right)^2 \right)}{\sum_{r=0}^{k_p^M} \gamma_r \left(\hat{\boldsymbol{\mu}}^{(o_p)} \right) \exp \left(r \, 2\hat{\sigma} x_{p+}^O \right) \exp \left(\hat{\sigma} r^2 \right)},
$$

$$
= \frac{\exp \left(\sum_{i=1}^k x_{pi}^M \left[\hat{\mu}_i + 2\hat{\sigma} x_{p+}^O \right] \right)}{\gamma_{x_{p+}^M} \left(\hat{\boldsymbol{\mu}}^{(o_p)} \right) \exp \left(x_{p+}^M \, 2\hat{\sigma} x_{p+}^O \right)} \frac{\gamma_{x_{p+}^M} \left(\hat{\boldsymbol{\mu}}^{(o_p)} \right) \exp \left(x_{p+}^M \, 2\hat{\sigma} x_{p+}^O \right) \exp \left(\hat{\sigma} \left(x_{p+}^M \right)^2 \right)}{\sum_{r=0}^{k_p^M} \gamma_r \left(\hat{\boldsymbol{\mu}}^{(o_p)} \right) \exp \left(r \, 2\hat{\sigma} x_{p+}^O \right) \exp \left(\hat{\sigma} r^2 \right)}
$$

$$
= p_{\{\hat{\mu}, \hat{\sigma}\}} \left(\mathbf{x}_p^M \mid x_{p+}^M, x_{p+}^O \right) p_{\{\hat{\mu}, \hat{\sigma}\}} \left(x_{p+}^M \mid x_{p+}^O \right)
$$

where k_p^M denotes the number of missing responses for pupil p, and $\gamma_r \left(\hat{\boldsymbol{\mu}}^{(o_p)} \right)$ denotes the elementary symmetric function of order r of the subvector $\hat{\boldsymbol{\mu}}^{(o_p)}$, i.e., the external fields corresponding to the missing observations for a person p.

The two expectations can now be found as follows. Assuming that the response X_{pi} of pupil p to question i is missing, we compute its expectation as

$$
\mathbb{E}_{\{\hat{\mu}, \hat{\sigma}\}} \left\{ X_{pi}^M \,\middle|\, \mathbf{x}_p^O \right\} = \sum_{\mathbf{x}_p^M} x_{pi}^M \, p_{\{\hat{\mu}, \hat{\sigma}\}} \left(\mathbf{x}_p^M \mid x_{p+}^O \right)
$$

$$
= \sum_{x_{pi}^M} x_{pi}^M \sum_{\mathbf{x}_p^{m(i)}} p_{\{\hat{\mu}, \hat{\sigma}\}} \left(\mathbf{x}_p^M \mid x_{p+}^O \right)
$$

$$
= \sum_{x_{pi}^M} x_{pi}^M \exp \left(\mu_i \right) \frac{\sum_{r=0}^{k_p^M - 1} \gamma_r \left(\hat{\boldsymbol{\mu}}^{(o_p, i)} \right) \exp \left((r+1) \, 2\hat{\sigma} x_{p+}^O \right) \exp \left(\hat{\sigma} (r+1)^2 \right)}{\sum_{r=0}^{k_p^M} \gamma_r \left(\hat{\boldsymbol{\mu}}^{(o_p)} \right) \exp \left(r \, 2\hat{\sigma} x_{p+}^O \right) \exp \left(\hat{\sigma} r^2 \right)}
$$

$$
= \exp \left(\hat{\mu}_i \right) \frac{\sum_{r=0}^{k_p^M - 1} \gamma_r \left(\hat{\boldsymbol{\mu}}^{(o_p, i)} \right) \exp \left((r+1) \, 2\hat{\sigma} x_{p+}^O \right) \exp \left(\hat{\sigma} (r+1)^2 \right)}{\sum_{r=0}^{k_p^M} \gamma_r \left(\hat{\boldsymbol{\mu}}^{(o_p)} \right) \exp \left(r \, 2\hat{\sigma} x_{p+}^O \right) \exp \left(\hat{\sigma} r^2 \right)},
$$

otherwise this expectation is set to zero. In the same way, we find

$$
\mathbb{E}_{\{\hat{\mu}, \hat{\sigma}\}} \left\{ \left[x_{p+}^O + X_{p+}^M \right]^2 \,\middle|\, \mathbf{x}_p^O \right\} = \frac{\sum_{r=0}^{k_p^M} \left[x_{p+}^O + r \right]^2 \gamma_r \left(\hat{\boldsymbol{\mu}}^{(o_p)} \right) \exp \left(r \, 2\hat{\sigma} x_{p+}^O \right) \exp \left(\hat{\sigma} r^2 \right)}{\sum_{r=0}^{k_p^M} \gamma_r \left(\hat{\boldsymbol{\mu}}^{(o_p)} \right) \exp \left(r \, 2\hat{\sigma} x_{p+}^O \right) \exp \left(\hat{\sigma} r^2 \right)}.
$$

Both expectations are easy to evaluate expressions of the conditional Curie-Weiss model. However, computing the basis functions for the missing item responses per pupil might be expensive. Fortunately, for incomplete test designs these functions often need only be computed per booklet, or cluster of questions, and not per individual.

5.3.3 The M-Step

The E-step effectively completes the incomplete data likelihood, and we now end up
with a maximization problem that is comparable to that for the complete data case.
We approach the maximization in a similar way as before, and use the same *divide
and conquer* strategy to maximize each of the parameters in turn, while ignoring
cross-parameter dependency during optimization.

5.3.3.1 M-Step for μ_i

The partial derivative of $Q\left(\mu, \sigma \; ; \hat{\mu}, \hat{\sigma}\right)$ with respect to μ_i is

$$
\frac{\partial}{\partial \mu_i} Q\left(\mu, \sigma \; ; \hat{\mu}, \hat{\sigma}\right) = \left[x_{+i}^O + \sum_{p=1}^{n} \mathbb{E}_{\{\hat{\mu}, \hat{\sigma}\}} \left\{ X_{pi}^M \middle| \mathbf{x}_p^O \right\} \right]
$$
$$
- n \exp\left(\mu_i\right) \frac{\sum_{s=0}^{k-1} \gamma_s \left(\mu^{(i)}\right) \exp\left(\sigma s^2\right)}{\sum_{s=0}^{k} \gamma_s \left(\mu\right) \exp\left(\sigma s^2\right)}.
$$

When we set this derivative to zero we obtain the following closed form solution for
the parameter μ_i:

$$
\mu_i = \ln \left(\frac{\left[x_{+i}^O + \sum_{p=1}^{n} \mathbb{E}_{\{\hat{\mu}, \hat{\sigma}\}} \left\{ X_{pi} \middle| \mathbf{x}_p^O \right\} \right]}{n - \left[x_{+i}^O + \sum_{p=1}^{n} \mathbb{E}_{\{\hat{\mu}, \hat{\sigma}\}} \left\{ X_{pi} \middle| \mathbf{x}_p^O \right\} \right]} \right)
$$
$$
+ \ln \left(\frac{\sum_{s=0}^{k-1} \gamma_s \left(\mu^{(i)}\right) \exp\left(\sigma s^2\right)}{\sum_{s=0}^{k-1} \gamma_s \left(\mu^{(i)}\right) \exp\left(\sigma \left(s+1\right)^2\right)} \right),
$$

again fixing the states of the remaining k parameters to their current estimates $\hat{\mu}^{(i)}$
and $\hat{\sigma}$ to compute an updated value for μ_i.

5.3.3.2 M-Step for σ

The partial derivative of $Q\left(\mu, \sigma \; ; \hat{\mu}, \hat{\sigma}\right)$ with respect to σ is

$$
\frac{\partial}{\partial \sigma} Q\left(\mu, \sigma \; ; \hat{\mu}, \hat{\sigma}\right) = \sum_{p=1}^{n} \mathbb{E}_{\{\hat{\mu}, \hat{\sigma}\}} \left\{ \left[x_{p+}^O + X_{p+}^M \right]^2 \middle| \mathbf{x}_p^O \right\}
$$
$$
- n \frac{\sum_{s=0}^{k} s^2 \gamma_s \left(\mu\right) \exp\left(\sigma s^2\right)}{\sum_{s=0}^{k} \gamma_s \left(\mu\right) \exp\left(\sigma s^2\right)}.
$$

Setting this derivative to zero does not lead to a closed form solution for the parameter σ. We therefore propose a one-dimensional NR step:

$$
\sigma = \hat{\sigma} - \frac{\frac{\partial}{\partial \sigma} Q\left(\boldsymbol{\mu}, \sigma \; ; \hat{\boldsymbol{\mu}}, \hat{\sigma}\right)\big|_{\sigma=\hat{\sigma}}}{\frac{\partial^2}{\partial \sigma^2} Q\left(\boldsymbol{\mu}, \sigma \; ; \hat{\boldsymbol{\mu}}, \hat{\sigma}\right)\big|_{\sigma=\hat{\sigma}}}
$$

$$
= \hat{\sigma} + \frac{\frac{1}{n} \sum_{p=1}^n \mathbb{E}_{\{\hat{\boldsymbol{\mu}}, \hat{\sigma}\}} \left\{ \left[x_{p+}^O + X_{p+}^M \right]^2 \middle| \mathbf{x}_p^O \right\} - \frac{\sum_{s=0}^k s^2 \gamma_s\left((\hat{\boldsymbol{\mu}})\right) \exp(\hat{\sigma} s^2)}{\sum_{s=0}^k \gamma_s\left((\hat{\boldsymbol{\mu}})\right) \exp(\hat{\sigma} s^2)}}{\frac{\sum_{s=0}^k s^4 \gamma_s\left((\hat{\boldsymbol{\mu}})\right) \exp(\hat{\sigma} s^2)}{\sum_{s=0}^k \gamma_s\left((\hat{\boldsymbol{\mu}})\right) \exp(\hat{\sigma} s^2)} - \left(\frac{\sum_{s=0}^k s^2 \gamma_s\left((\hat{\boldsymbol{\mu}})\right) \exp(\hat{\sigma} s^2)}{\sum_{s=0}^k \gamma_s\left((\hat{\boldsymbol{\mu}})\right) \exp(\hat{\sigma} s^2)} \right)^2},
$$

where we evaluate the partial derivatives on the current states of all parameters.

5.3.3.3 Asymptotic Standard Errors for the Incomplete Data Case

The *missing information principle* of Louis (1982) states that the observed information is equal to the complete information—the expression that we obtained for the complete data case—minus the missing information (see also Tanner 1996):

$$
\mathcal{I}^O\left(\boldsymbol{\mu}, \sigma\right) = \mathcal{I}\left(\boldsymbol{\mu}, \sigma\right) - \mathcal{I}^M\left(\boldsymbol{\mu}, \sigma\right).
$$

The observed Fisher information $\mathcal{I}^O$ is computed from mixed partial derivatives of the incomplete data likelihood,

$$
L\left(\boldsymbol{\mu}, \sigma \mid \mathbf{X}^O\right) = \frac{\prod_{p=1}^n \left(\prod_{i=1}^k e^{x_{pi}^O \mu_i} \right) e^{\sigma \left(x_{p+}^O\right)^2} \sum_{\mathbf{x}_p^M} \left(\prod_{i=1}^k e^{x_{pi}^M \mu_i} \right) e^{\sigma \left(\left(x_{p+}^M\right)^2 + 2 x_{p+}^M x_{p+}^O \right)}}{\left(\sum_{s=0}^k \gamma_s\left(\boldsymbol{\mu}\right) e^{\sigma s^2} \right)^n},
$$

and the missing Fisher information $\mathcal{I}^M$ is computed from the conditional distribution of the missing data given the observed data,

$$
p\left(\mathbf{X}^M \mid \mathbf{X}^O\right) = \prod_{p=1}^n \frac{\exp\left(\sum_{i=1}^k x_{pi}^M \mu_i + \sigma \left(x_{p+}^O + x_{p+}^M \right)^2 \right)}{\sum_{r=0}^{k_p^M} \gamma_r\left(\boldsymbol{\mu}^{(o_p)} \right) \exp\left(\sigma \left(x_{p+}^O + r \right)^2 \right)}.
$$

This conditional distribution is a member of the exponential family and its mixed second order partial derivatives do not depend on missing data such that we do not have to numerically evaluate any expectations in computing the missing information. We therefore compute the observed information as the difference between the complete information and the missing information. In Appendix 2 we present the mixed partial derivatives of the conditional distribution of the missing data. Sample R-code to compute the observed and missing information is made available at https://osf.io/4m3dq/.

5.4   Numerical Illustrations

In this section we provide two numerical illustrations of our procedures; one based on simulated data, and one based on real data from the 2012 Cito Eindtoets.

5.4.1   Simulated Example

We illustrate that our procedures work using a small simulated example. We have simulated the responses of $n = 10,000$ pupils to $k = 20$ test questions.[4] The external fields were sampled uniformly between -1 and $+1$, and the scale parameter σ was set to 0.05 (or $J = k\sigma = 1$). For the incomplete data case, we omitted the responses to the last five questions for 5026 randomly selected pupils and the responses to the first five questions for the 4974 remaining pupils. This mimics a two-booklet test design, where booklet one comprises of questions 1–15 and booklet two comprises of questions 6–20. The parameter estimates for this example are shown in Table 5.1.

From Table 5.1 we observe that the parameter estimates in both the complete data case and in the incomplete data case are close to their true values, and well within the range of their 95% confidence intervals. Thus, we were able to estimate the parameters of the Curie-Weiss model in both the complete and incomplete data case.

We also observe from Table 5.1 that the field estimates of questions for which no data were excluded—questions 6–15—slightly differed from their estimates obtained from the complete data case. At the same time, the (asymptotic) standard errors of all external fields increased in size, even though half of the external field parameters are estimated based on the same amount of observations. This reveals the impact of cross-parameter correlations on the obtained estimates.

5.4.2   The Cito Eindtoets 2012

As an empirical example from educational measurement, we consider an application of the Curie-Weiss model to data from the 2012 Cito Eindtoets. The data consist of the responses of $n = 133,768$ Dutch pupils at the end of primary school to $k = 200$ questions on twelve distinct topics from the Dutch primary school curriculum related to mathematics and language education.

The Cito Eindtoets data are typically not analyzed with a Rasch-type model. Since the test comprises of such distinct topics from the educational curriculum, we may *a priori* expect that the Curie-Weiss model fits poorly to the data from this test. However, in an earlier analysis of Cito Eindtoets data using a low-rank Ising model we found that the first principal component explained roughly 99% of the

---

[4]Data were generated with a Gibbs sampler (Geman and Geman 1984) utilizing the full-conditional distributions in Eq. (5.4).

Table 5.1 Parameter estimates of the Curie-Weiss model based on simulated data from $n = 10,000$ pupils responding to $k = 20$ test questions

	True value	Complete data		Incomplete data	
		MLE	(SE)	MLE	(SE)
μ_1	0.309	0.287	(0.068)	0.320	(0.089)
μ_2	−0.581	−0.590	(0.066)	−0.571	(0.085)
μ_3	−0.275	−0.310	(0.067)	−0.336	(0.086)
μ_4	0.550	0.448	(0.069)	0.426	(0.089)
μ_5	−0.726	−0.739	(0.066)	−0.736	(0.085)
μ_6	0.250	0.261	(0.068)	0.271	(0.083)
μ_7	−0.397	−0.428	(0.066)	−0.418	(0.082)
μ_8	−0.736	−0.755	(0.066)	−0.745	(0.082)
μ_9	0.372	0.355	(0.068)	0.365	(0.083)
μ_{10}	−0.551	−0.600	(0.066)	−0.590	(0.082)
μ_{11}	0.770	0.704	(0.070)	0.714	(0.085)
μ_{12}	0.578	0.481	(0.069)	0.491	(0.084)
μ_{13}	0.897	0.915	(0.072)	0.924	(0.086)
μ_{14}	0.646	0.692	(0.070)	0.702	(0.085)
μ_{15}	−0.251	−0.308	(0.067)	−0.298	(0.082)
μ_{16}	0.296	0.289	(0.068)	0.320	(0.088)
μ_{17}	0.541	0.556	(0.069)	0.606	(0.091)
μ_{18}	0.054	0.003	(0.067)	0.003	(0.087)
μ_{19}	0.181	0.163	(0.068)	0.166	(0.088)
μ_{20}	−0.766	−0.755	(0.066)	−0.742	(0.085)
σ	0.050	0.051	(0.002)	0.050	(0.002)

variation in the matrix of observed sufficient statistics (see Marsman et al. 2015). This principal component score correlated highly with the raw test score, and moreover, the estimated elements in the first eigenvector were found to be nearly constant (van den Bergh et al. 2018). Both observations suggest that a one-dimensional model such as a 2PL or Rasch-type model might show a reasonable fit to the observed data.

We assess the fit of the Curie-Weiss model to the 2012 Cito Eindtoets using the item-rest regressions in Eq. (5.4), focusing on the rest-scores with at least 25 observations to obtain stable estimates of the observed proportions. There are 200 item-rest regressions in total, one for each test question. We show the item-rest regressions of four questions (Questions 1, 7, 10, and 12 from the test) in Fig. 5.1, but provide all available item-rest regressions at https://osf.io/4m3dq/. When we investigate the item-rest regressions at https://osf.io/4m3dq/ it is clear that we did not find a good fit of the Curie-Weiss model to all of the 200 questions from the 2012 Cito Eindtoets. An example of an item-rest regression for a poor fitting question is that of Question 1, which is shown in the top-left panel of Fig. 5.1. Even though
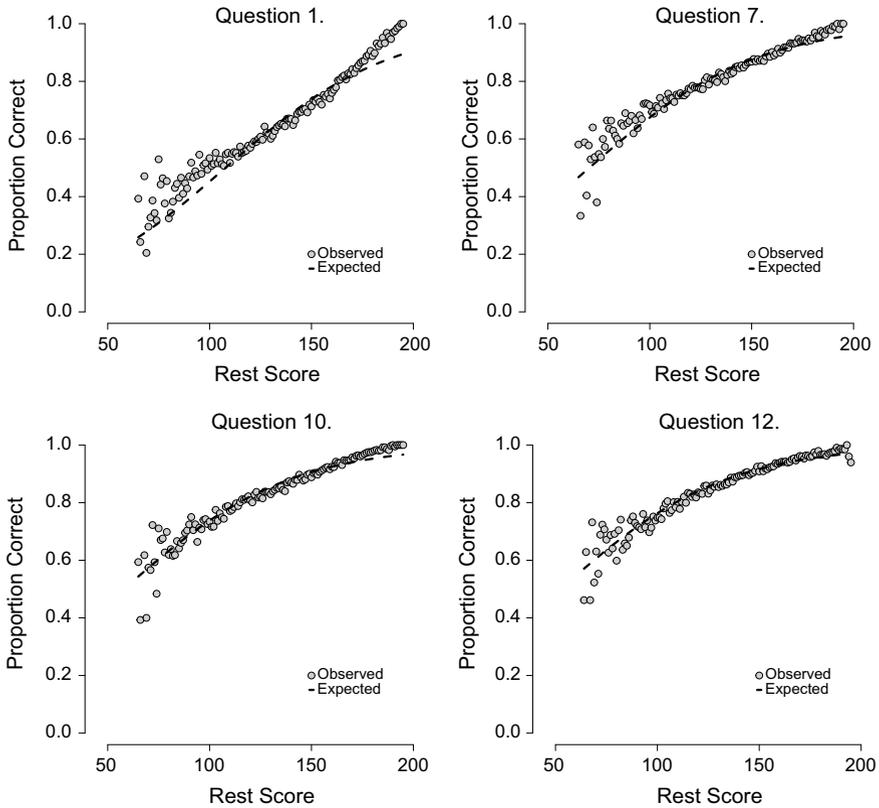
Fig. 5.1 The item-rest regressions for four questions from the Cito Eindtoets 2012

we did not find a good fit of the Curie-Weiss model to all of the questions in the Cito Eindtoets, we did observe many questions for which the estimated Curie-Weiss model did provide an at least reasonable fit. Two examples of questions for which the item-rest regression revealed a reasonable fit to the estimated Curie-Weiss model are Questions 7 and 10 in Fig. 5.1. Question 12 in Fig. 5.1 is an example of an item-rest regression that corresponds to a good fit.

Despite the fact that several of the item-rest regressions indicate a reasonable fit to the data at the item-level, it is evident that the Curie-Weiss model fits poorly to these data on a global level, as it is unable to reproduce the observed score distribution. This is illustrated in Fig. 5.2, where we compare the observed score distribution with the theoretical score distribution, e.g. Eq. (5.6). Even though it is clear that the two score distributions differ from each other, they have the same mean and variance. However, the overdispersion in the observed score distribution suggests that a more complicated model than the Curie-Weiss model, such as the low-rank Ising model that was used in Marsman et al. (2015), is likely more suited for these data.

Fig. 5.2 The observed score
distribution for the Cito
Eindtoets 2012 (gray dots)
and the theoretical score
distribution as predicted by
the estimated Curie-Weiss
model (black dots) using Eq.
(5.6)
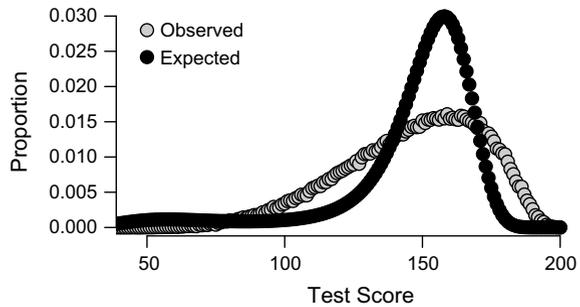
5.5 Discussion

In this chapter we have focused on the statistical analysis of the Curie-Weiss model
(Kac 1968) in the context of educational testing. In contrast to other graphical models
that are regularly used in the literature—such as the Ising model (Ising 1925; Lenz
1920)—we showed that the Curie-Weiss model is computationally tractable, which
makes it a practical tool for the large-scale applications that we often see in educa-
tional practice. One of the focal points of our statistical treatment was the analysis of
the Curie-Weiss model in the face of data that are missing at random (Rubin 1976),
such as the missing data patterns that we often observe in incomplete test designs
(Eggen 1993; Eggen and Verhelst 2011). We have developed an approach using the
EM algorithm (Dempster et al. 1977) to estimate the Curie-Weiss model on partially
observed data. (Sample R-code is made available at https://osf.io/4m3dq/.)

We have provided two illustrations of our work. Simulated data were used to
illustrate that we could recover the parameters in both the complete data case and
the incomplete data case. The example using the 2012 Cito Eindtoets data was in-
cluded for two reasons. Firstly, this example allowed us to illustrate the ease with
which the fit of the Curie-Weiss model can be investigated using the analytic expres-
sions of the item-rest regressions in Eq. (5.4), and the theoretical score distribution
in Eq. (5.6). Secondly, it allowed us to illustrate that a simple model such as the
Curie-Weiss model is able to fit complex data, at least at the item-level. The fact that
the Curie-Weiss model seems to fit reasonably well to several substantively differ-
ent questions—ranging from spelling and reading comprehension to working with
fractions—definitely warrants further investigation.

The work in this chapter is a first step in the psychometric treatment of the Curie-
Weiss model and can be generalized in several ways. For example, the factorization
in Eq. (5.6) reminds us of a two-step procedure that is used in the Cito program
SAUL (Structural Analysis of a Univariate Latent variable; Verhelst and Eggen 1989;
Verhelst and Verstralen 2002). In this program, an IRT model is analyzed and fitted
to observed data first, which in the Cito tradition comprises of either a Rasch model
or a one parameter logistic model (OPLM; Verhelst and Glas 1995): a tradition that is
pursued by the dexter R package (Maris et al. 2018). After a fitting model is obtained,

its parameters are fixed, and the influence of background variables on the latent trait is assessed. This suggests two generalizations of the Curie-Weiss model that we have analyzed here. The first is the inclusion of integer weights or discrimination's for each of the variables, as with the OPLM. The advantage of using integer weights is that this ensures that the normalizing constant of the Curie-Weiss model remains tractable. A second generalization would be the assessment of the influence of background variables on the network's structure or score distribution. One way to realize this is by explicitly modeling the scaling parameter, for instance using a regression model to analyze differences in the scaling parameter for different pupils:

$$\sigma_p = \exp\left(\boldsymbol{\beta}^\mathsf{T}\mathbf{z}_p\right),$$

where $\mathbf{z}_p$ denotes a vector of covariates that correspond to a pupil p and $\boldsymbol{\beta}$ denotes a set of regression parameters. Both generalizations would likely lead to the improved fit of the Curie-Weiss model in the Cito Eindtoets example, at both the item- and test-level. It is important to observe that the Curie-Weiss model under both generalizations remains entirely tractable.

Even though we have focused on the statistical treatment of the Curie-Weiss model from a classical perspective, it is easy to generalize our work to a Bayesian approach. One way to estimate the model in the complete data case, for example, is by means of the approximate Gibbs sampling approach of Marsman et al. (2015), that was further analyzed by Bechger et al. (2018) and Marsman et al. (2017). Another approach is based on the Gibbs sampler that Maris et al. (2015) developed for estimating the ERM, and extended by Brinkhuis (in press, Chap. 5) to the incomplete data case, where Tanis (2018) has recently shown how to adapt this Gibbs sampler for estimating the Curie-Weiss model. Data augmentation can then be used to handle incomplete data (Tanner and Wong 1987; Tanner 1996). This entails adding an additional step to the Gibbs sampler to impute missing observations based on the observed data using, for example, Eq. (5.4).

Appendix 1: Mixed Partial Derivatives Complete Data Likelihood

To compute the Fisher information matrix, we need the following mixed partial derivatives of the (complete data) likelihood:

$$\frac{\partial^2}{\partial\mu_i\partial\mu_j}L\left(\boldsymbol{\mu},\sigma\mid\mathbf{X}\right) = n\,\mathrm{e}^{\mu_i+\mu_j}\left\{\frac{\sum_{s=0}^{k-1}\gamma_s\left(\boldsymbol{\mu}^{(i)}\right)\,\mathrm{e}^{\sigma(s+1)^2}}{\sum_{s=0}^{k}\gamma_s\left(\boldsymbol{\mu}\right)\,\mathrm{e}^{\sigma s^2}}\frac{\sum_{s=0}^{k-1}\gamma_s\left(\boldsymbol{\mu}^{(j)}\right)\,\mathrm{e}^{\sigma(s+1)^2}}{\sum_{s=0}^{k}\gamma_s\left(\boldsymbol{\mu}\right)\,\mathrm{e}^{\sigma s^2}}\right.$$
$$\left.-\frac{\sum_{s=0}^{k-2}\gamma_s\left(\boldsymbol{\mu}^{(i,j)}\right)\,\mathrm{e}^{\sigma(s+2)^2}}{\sum_{s=0}^{k}\gamma_s\left(\boldsymbol{\mu}\right)\,\mathrm{e}^{\sigma s^2}}\right\}$$

$$\frac{\partial^2}{\partial \mu_i^2} L\left(\boldsymbol{\mu}, \sigma \mid \mathbf{X}\right) = n\, e^{\mu_i} \left\{ e^{\mu_i} \left(\frac{\sum_{s=0}^{k-1} \gamma_s\left(\boldsymbol{\mu}^{(i)}\right) e^{\sigma (s+1)^2}}{\sum_{s=0}^{k} \gamma_s\left(\boldsymbol{\mu}\right) e^{\sigma s^2}} \right)^2 \right.$$
$$\left. - \frac{\sum_{s=0}^{k-1} \gamma_s\left(\boldsymbol{\mu}^{(i)}\right) e^{\sigma (s+1)^2}}{\sum_{s=0}^{k} \gamma_s\left(\boldsymbol{\mu}\right) e^{\sigma s^2}} \right\}$$

$$\frac{\partial^2}{\partial \sigma^2} L\left(\boldsymbol{\mu}, \sigma \mid \mathbf{X}\right) = n \left(\frac{\sum_{s=0}^{k} s^2 \gamma_s\left(\boldsymbol{\mu}\right) e^{\sigma s^2}}{\sum_{s=0}^{k} \gamma_s\left(\boldsymbol{\mu}\right) e^{\sigma s^2}} \right)^2 - n \frac{\sum_{s=0}^{k} s^4 \gamma_s\left(\boldsymbol{\mu}\right) e^{\sigma s^2}}{\sum_{s=0}^{k} \gamma_s\left(\boldsymbol{\mu}\right) e^{\sigma s^2}}$$

$$\frac{\partial^2}{\partial \sigma \partial \mu_i} L\left(\boldsymbol{\mu}, \sigma \mid \mathbf{X}\right) = n\, e^{\mu_i} \left\{ \frac{\sum_{s=0}^{k} s^2 \gamma_s\left(\boldsymbol{\mu}\right) e^{\sigma s^2}}{\sum_{s=0}^{k} \gamma_s\left(\boldsymbol{\mu}\right) e^{\sigma s^2}} \frac{\sum_{s=0}^{k-1} \gamma_s\left(\boldsymbol{\mu}^{(i)}\right) e^{\sigma (s+1)^2}}{\sum_{s=0}^{k} \gamma_s\left(\boldsymbol{\mu}\right) e^{\sigma s^2}} \right.$$
$$\left. - \frac{\sum_{s=0}^{k-1} (s+1)^2 \gamma_s\left(\boldsymbol{\mu}^{(i)}\right) e^{\sigma (s+1)^2}}{\sum_{s=0}^{k} \gamma_s\left(\boldsymbol{\mu}\right) e^{\sigma s^2}} \right\}$$

Appendix 2: Mixed Partial Derivatives of (Conditional) Distribution of the Missing Data

To compute the missing Fisher information, we need the mixed partial derivatives of $p\left(\mathbf{X}^M \mid \mathbf{X}^O\right)$ with respect to $\boldsymbol{\mu}$ and σ. Let m_{pi} be the missing data indicator, such that

$$m_{pi} = \begin{cases} 1 & \text{if the response of pupil } p \text{ to item } i \text{ is missing,} \\ 0 & \text{if the response of pupil } p \text{ to item } i \text{ is observed.} \end{cases}$$

The mixed partial derivatives of $p\left(\mathbf{X}^M \mid \mathbf{X}^O\right)$ are then

$$\frac{\partial^2}{\partial \mu_i \partial \mu_j} p\left(\mathbf{X}^M \mid \mathbf{X}^O\right) = \sum_{p=1}^{n} m_{pi} m_{pj}\, e^{\mu_i + \mu_j}$$
$$\left\{ \frac{\sum_{r=0}^{k_p^M-1} \gamma_r\left(\boldsymbol{\mu}^{(O_p, i)}\right) e^{\sigma\left(x_p^O + r + 1\right)^2}}{\sum_{r=0}^{k_p^M} \gamma_r\left(\boldsymbol{\mu}^{(O_p)}\right) e^{\sigma\left(x_p^O + r\right)^2}} \frac{\sum_{r=0}^{k_p^M-1} \gamma_r\left(\boldsymbol{\mu}^{(O_p, j)}\right) e^{\sigma\left(x_p^O + r + 1\right)^2}}{\sum_{r=0}^{k_p^M} \gamma_r\left(\boldsymbol{\mu}^{(O_p)}\right) e^{\sigma\left(x_p^O + r\right)^2}} \right.$$
$$\left. - \frac{\sum_{r=0}^{k_p^M-2} \gamma_r\left(\boldsymbol{\mu}^{(O_p, i, j)}\right) e^{\sigma\left(x_p^O + r + 2\right)^2}}{\sum_{r=0}^{k_p^M} \gamma_r\left(\boldsymbol{\mu}^{(O_p)}\right) e^{\sigma\left(x_p^O + r\right)^2}} \right\}$$

$$\frac{\partial^2}{\partial \mu_i^2} p\left(\mathbf{X}^M \mid \mathbf{X}^O\right) = \sum_{p=1}^n m_{pi}\, \mathrm{e}^{\mu_i} \left\{ \mathrm{e}^{\mu_i} \left(\frac{\sum_{r=0}^{k_p^M - 1} \gamma_r\left(\boldsymbol{\mu}(O_p, i)\right)\, \mathrm{e}^{\sigma\left(x_p^O + r + 1\right)^2}}{\sum_{r=0}^{k_p^M} \gamma_r\left(\boldsymbol{\mu}(O_p)\right)\, \mathrm{e}^{\sigma\left(x_p^O + r\right)^2}} \right)^2 \right.$$

$$\left. - \frac{\sum_{r=0}^{k_p^M - 1} \gamma_r\left(\boldsymbol{\mu}(O_p, i)\right)\, \mathrm{e}^{\sigma\left(x_p^O + r + 1\right)^2}}{\sum_{r=0}^{k_p^M} \gamma_r\left(\boldsymbol{\mu}(O_p)\right)\, \mathrm{e}^{\sigma\left(x_p^O + r\right)^2}} \right\}$$

$$\frac{\partial^2}{\partial \sigma^2} p\left(\mathbf{X}^M \mid \mathbf{X}^O\right) = \sum_{p=1}^n m_{p+} \left\{ \left(\frac{\sum_{r=0}^{k_p^M} \left(x_p^O + r\right)^2 \gamma_r\left(\boldsymbol{\mu}(O_p)\right)\, \mathrm{e}^{\sigma\left(x_p^O + r\right)^2}}{\sum_{r=0}^{k_p^M} \gamma_r\left(\boldsymbol{\mu}(O_p)\right)\, \mathrm{e}^{\sigma\left(x_p^O + r\right)^2}} \right)^2 \right.$$

$$\left. - \frac{\sum_{r=0}^{k_p^M} \left(x_p^O + r\right)^4 \gamma_r\left(\boldsymbol{\mu}(O_p)\right)\, \mathrm{e}^{\sigma\left(x_p^O + r\right)^2}}{\sum_{r=0}^{k_p^M} \gamma_r\left(\boldsymbol{\mu}(O_p)\right)\, \mathrm{e}^{\sigma\left(x_p^O + r\right)^2}} \right\}$$

$$\frac{\partial^2}{\partial \sigma\, \partial \mu_i} p\left(\mathbf{X}^M \mid \mathbf{X}^O\right)$$

$$= \sum_{p=1}^n m_{pi}\, \mathrm{e}^{\mu_i} \left\{ \frac{\sum_{r=0}^{k_p^M} \left(x_p^O + r\right)^2 \gamma_r\left(\boldsymbol{\mu}(O_p)\right)\, \mathrm{e}^{\sigma\left(x_p^O + r\right)^2}}{\sum_{r=0}^{k_p^M} \gamma_r\left(\boldsymbol{\mu}(O_p)\right)\, \mathrm{e}^{\sigma\left(x_p^O + r\right)^2}} \frac{\sum_{r=0}^{k_p^M - 1} \gamma_r\left(\boldsymbol{\mu}(O_p, i)\right)\, \mathrm{e}^{\sigma\left(x_p^O + r + 1\right)^2}}{\sum_{r=0}^{k_p^M} \gamma_r\left(\boldsymbol{\mu}(O_p)\right)\, \mathrm{e}^{\sigma\left(x_p^O + r\right)^2}} \right.$$

$$\left. - \frac{\sum_{r=0}^{k_p^M - 1} \left(x_p^O + r + 1\right)^2 \gamma_r\left(\boldsymbol{\mu}(O_p, i)\right)\, \mathrm{e}^{\sigma\left(x_p^O + r + 1\right)^2}}{\sum_{r=0}^{k_p^M} \gamma_r\left(\boldsymbol{\mu}(O_p)\right)\, \mathrm{e}^{\sigma\left(x_p^O + r\right)^2}} \right\},$$

where terms for which $m_{pi} = 0$ are simply excluded from any computations.

References

Andersen, E. (1970). Asymptotic properties of conditional maximum-likelihood estimators. *Journal of the Royal Statistical Society. Series B (Methodological), 32*(2), 283–301. Retrieved from https://www.jstor.org/stable/2984535

Andersen, E. (1973). *Conditional inference and models for measuring* (Unpublished doctoral dissertation). Mentalhygiejnisk Forskningsinstitut.

Anderson, C., & Vermunt, J. (2000). Log-multiplicative association models as latent variable models for nominal and/or ordinal data. *Sociological Methodology*, *30*(1), 81–121. https://doi.org/10.1111/0081-1750.00076.

Anderson, C., & Yu, H.-T. (2007). Log-multiplicative association models as item response models. *Psychometrika*, *72*(1), 5–23. https://doi.org/10.1007/s11336-005-1419-2.

Baker, F., & Harwell, M. (1996). Computing elementary symmetric functions and their derivatives: A didactic. *Applied Psychological Measurement*, *20*(2), 169–192. https://doi.org/10.1177/014662169602000206.

Bechger, T. M., Maris, G. K. J., & Marsman, M. (2018). An asymptotically efficient sampler for Bayesian inference in educational measurement. arxiv:1808.03947

Bergsma, W. (1997). *Marginal models for categorical data* (Unpublished doctoral dissertation). Tilburg University.

Bergsma, W., & Rudas, T. (2002). Marginal models for categorical data. *The Annals of Statistics, 30*(1), 140–159. Retrieved from https://www.jstor.org/stable/2700006

Bock, R., & Aitken, M. (1981). Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. *Psychometrika*, *46*(4), 179–197. https://doi.org/10.1007/BF02293801.

Borsboom, D. (2008). Psychometric perspectives on diagnostic systems. *Journal of Clinical Psychology*, *64*(9), 1089–1108. https://doi.org/10.1002/jclp.

Brinkhuis, M.J.S., & Maris, G.K.J. (in press). Dynamic estimation in the extended marginal Rasch model with an application to mathematical computer-adaptive practice. *British Journal of Mathematical and Statistical Psychology*. https://doi.org/10.1111/bmsp.12157.

Chrétien, M., Gross, E., & Deser, S. (Eds.). (1968). *Statistical physics: Phase transitions and superfluidity, vol. 1, Brandeis University summer institute in theoretical physics*. New York: Gordon and Breach Science Publishers.

Comets, F., & Gidas, B. (1991). Asymptotics of maximum likelihood estimators for the Curie-Weiss model. *The Annals of Statistics, 19*(2), 557–578. Retrieved from https://www.jstor.org/stable/2242074

Cressie, N., & Holland, P. (1983). Characterizing the manifest probabilities of latent trait models. *Psychometrika*, *48*(1), 129–141. https://doi.org/10.1007/BF02314681.

Dawid, A. (1979). Conditional independence in statistical theory. *Journal of the Royal Statistical Society. Series B (Methodological)*, *41*(1), 1–31.

Dempster, A., Laird, N., & Rubin, D. (1977). Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society. Series B (Methodological), 39*(1), 1–38. Retrieved from https://www.jstor.org/stable/2984875

Eggen, T. (1993). Psychometrie in de praktijk [psychometrics in practice]. In T. Eggen & P. Sanders (Eds.), (pp. 239–284). Arnhem: Cito.

Eggen, T. (2000). On the loss of information in conditional maximum likelihood estimation of item parameters. *Psychometrika*, *65*(3), 337–362. https://doi.org/10.1007/BF02296150.

Eggen, T. (2004). *Contributions to the theory and practice of the computerized adaptive testing* (Unpublished doctoral dissertation). University of Twente.

Eggen, T., & Verhelst, N. (2011). Item calibration in incomplete testing designs. *Psicologica: International Journal of Methodology and Experimental Psychology*, *32*(1), 107–132.

Epskamp, S. (2017). *Network psychometrics* (Unpublished doctoral dissertation). University of Amsterdam.

Epskamp, S., Maris, G., Waldorp, L., & Borsboom, D. (2018). Handbook of psychometrics. In P. Irwing, D. Hughes, & T. Booth (Eds.), Network psychometrics. New York, NY: Wiley-Blackwell.

Fischer, G. H. (1974). *Einführung in die theorie psychologischer tests [introduction to the theory of psychological tests]*. Bern: Huber.

Geman, S., & Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *6*(6), 721–741. https://doi.org/10.1109/TPAMI.1984.4767596.

Gould, H., & Tobochnik, J. (2010). *Statistical and thermal physics with computer applications*. New Jersey: Princeton University.

Haslbeck, J., Epskamp, E., Marsman, M., & Waldorp, L. (2018). *Interpreting the Ising model: The input matters*. Retrieved from arxiv:1811.02916 (ArXiv preprint.)

Hessen, D. (2012). Fitting and testing conditional multinormal partial credit models. *Psychometrika*, *77*(4), 693–709. https://doi.org/10.1007/s11336-012-9277-1.

Holland, P. (1990). The Dutch Identity: A new tool for the study of item response models. *Psychometrika*, *55*(6), 5–18. https://doi.org/10.1007/BF02294739.

Hubbard, J. (1959). Calculation of partition functions. *Physical Review Letters*, *3*(2), 77–78. https://doi.org/10.1103/PhysRevLett.3.77.

Ip, E. (2002). Locally dependent latent trait model and the Dutch Identity revisited. *Psychometrika*, *67*(3), 367–386. https://doi.org/10.1007/BF02294990.

Ising, E. (1925). Beitrag zur theorie des ferromagnetismus. *Zeitschrift für Physik*, *31*(1), 253–258. https://doi.org/10.1007/BF02980577.

Kac, M. (1968). Statistical physics: Phase transitions and superfluidity, vol. 1, Brandeis University Summer Institute in Theoretical Physics. In: M. Chrétien, E. Gross, & S. Deser (Eds.), pp. 241–305). New York: Gordon and Breach Science Publishers.

Kac, M., & Thompson, C. (1966). On the mathematical mechanism of phase transition. *Proceedings of the National Academy of Sciences of the United States of America*, *55*(4), 676–683. https://doi.org/10.1073/pnas.55.4.676.

Kochmański, M., Paszkiewicz, T., & Wolski, S. (2013). Curie-Weiss magnet-a simple model of phase transition. *European Journal of Physics*, *34*(6), 1555–1573.

Lenz, W. (1920). Beiträge zum verständnis der magnetischen eigenschaften in festen körpern. *Physikalische Zeitschrift*, *21*, 613–615.

Louis, T. (1982). Finding observed information using the EM algorithm. *Journal of the Royal Statistical Society. Series B (Methodological), 44*(2), 226–233. Retrieved from https://www.jstor.org/stable/2345828

Maris, G., Bechger, T., Koops, J., & Partchev, I. (2018). dexter: Data management and analysis of tests [Computer software manual]. Retrieved from https://CRAN.R-project.org/package=dexter (R package version 0.8.1)

Maris, G., Bechger, T., & San Martin, E. (2015). A Gibbs sampler for the (extended) marginal Rasch model. *Psychometrika*, *80*(4), 859–879. https://doi.org/10.1007/s11336-015-9479-4.

Marsman, M., Borsboom, D., Kruis, J., Epskamp, S., van Bork, R., Waldorp, L., ... Maris, G. (2018). An introduction to network psychometrics: Relating Ising network models to item response theory models. *Multivariate Behavioral Research*, *53*(1), 15–35. https://doi.org/10.1080/00273171.2017.1379379.

Marsman, M., Maris, G., Bechger, T., & Glas, C. (2015). Bayesian inference for low-rank Ising networks. *Scientific Reports*, *5*(9050). https://doi.org/10.1038/srep09050

Marsman, M., Maris, G., Bechger, T., & Glas, C. (2017). Turning simulation into estimation: Generalized exchange algorithms for exponential family models. *PLoS One*, *12*(1), 1–15 (e0169787). https://doi.org/10.1371/journal.pone.0169787

Marsman, M., Sigurdardóttir, H., Bolsinova, M., & Maris, G.K.J. (in press). Characterizing the manifest probability distributions of three latent trait models for accuracy and response time. *Psychometrika*. https://doi.org/10.1007/s11336-019-09668-3.

Marsman, M., Waldorp, L., & Maris, G. (2017). A note on large-scale logistic prediction: Using an approximate graphical model to deal with collinearity and missing data. *Behaviormetrika*, *44*(2), 513–534. https://doi.org/10.1007/s41237-017-0024-x.

McCullagh, P. (1994). Exponential mixtures and quadratic exponential families. *Biometrika*, *81*(4), 721–729. https://doi.org/10.1093/biomet/81.4.721.

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

Rubin, D. (1976). Inference and missing data. *Biometrika*, *63*(3), 581–592. https://doi.org/10.1093/biomet/63.3.581.

Stratonovich, R. (1957). On a method of calculating quantum distribution functions. *Soviet Physics Doklady*, *2*, 416.

Tanis, C. (2018). *An introduction of the Curie-Weiss model in educational measurement*.

Tanner, M. (1996). *Tools for statistical inference: Methods for the exploration of posterior distributions and likelihood functions*. New York, NY: Springer. https://doi.org/10.1007/978-1-4612-4024-2

Tanner, M., & Wong, W. (1987). The calculation of posterior distributions by data augmentation. *Journal of the American Statistical Association*, *82*(398), 528–540. https://doi.org/10.2307/2289457.

Tjur, T. (1982). A connection between Rasch's item analysis model and a multiplicative poisson model. *Scandinavian Journal of Statistics*, *9*(1), 23–30.

van den Bergh, D., Bechger, T., & Marsman, M. (2018). Confirmatory item response theory using contrasts. *Manuscript in preparation*.

van der Linden, W., & Glas, C. (Eds.). (2002). *Computerized adaptive testing: Theory and practice*. New York: Kluwer Academic Publishers. https://doi.org/10.1007/0-306-47531-6.

van der Linden, W., & Glas, C. (Eds.). (2010). *Elements of adaptive testing*. New York: Springer. https://doi.org/10.1007/978-0-387-85461-8.

van der Maas, H., Dolan, C., Grasman, R., Wicherts, J., Huizenga, H., & Raijmakers, M. (2006). A dynamical model of general intelligence: The positive manifold of intelligence by mutualism. *Psychological Review*, *113*(4), 842–861. https://doi.org/10.1037/0033-295X.113.4.842.

Verhelst, N., & Eggen, T. (1989). *Psychometrische en statistische aspecten van peilingsonderzoek [psychometric and statistical aspects of survey research] (PPON-rapport)*. Arnhem: Cito.

Verhelst, N., & Glas, C. (1995). Rasch models. In G. H. Fisher & I. W. Molenaar (Eds.), (pp. 215–237). New York, NY: Springer.

Verhelst, N., Glas, C., & van der Sluis, A. (1984). Estimation problems in the Rasch model: The basic symmetric functions. *Computational Statistics Quarterly*, *1*(3), 245–262.

Verhelst, N., & Verstralen, H. (2002). Structural analysis of a univariate latent variable (SAUL): Theory and a computer program [Computer software manual]. Arnhem.