Statistical Models for the Precision of Categorical Measurement
van Wieringen, W.N.

Citation for published version (APA):

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: http://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
3 On the latent class model

The previous chapter proposes the latent class model for the outcome of a measurement system analysis experiment for measurement systems with a binary response. In this chapter, we prove first the identifiability of this model. Next, estimators for the parameters of the model are constructed, and their variance is determined. Finally, we show how confidence intervals of the estimates are constructed and how the goodness-of-fit of the latent class model is assessed.

3.1 The latent class model

We adopt the notation introduced in chapter 2. Let $Y_i$ be the reference values of objects $i = 1, \ldots, n$, which one cannot observe directly. Take $Y_i$ to be Bernoulli distributed with unknown parameter $\theta = P(Y_i = 1)$ for each $i$, the probability of the object being good. The objects are measured $\ell_j \geq 1$ times by raters $j = 1, \ldots, m$. Note that compared with chapter 2 this is a generalization of the latent class model. The random variable $X_{ij} \in \{0, 1, \ldots, \ell_j\}$ represents the number of times rater $j$ measures object $i$ as good. The distribution of $X_{ij}$ depends on $Y_i$. We let $\pi_j(1)$ be the probability that rater $j$ rates a good object as good, and $\pi_j(0)$ that he rates a bad object as good. Finally, let $X$ be the matrix containing the data from the experiment, defined as:

$$X = \begin{pmatrix} X_{11} & \cdots & X_{1m} \\ \vdots & \ddots & \vdots \\ X_{n1} & \cdots & X_{nm} \end{pmatrix}.$$  

(3.1)

The rows of $X$ are denoted by $X_1, \ldots, X_n$.

We write the likelihood function as (compare equation (2.3)):

$$L(X; \Psi) = \prod_{i=1}^{n} P(X_i = x_i)$$

$$= \prod_{i=1}^{n} \left( (1 - \theta) \prod_{j=1}^{m} \left( \frac{\ell_j}{X_{ij}} \right) (1 - \pi_j(0))^{\ell_j - X_{ij}} (\pi_j(0))^{X_{ij}} \right. $$

$$\left. + \theta \prod_{j=1}^{m} \left( \frac{\ell_j}{X_{ij}} \right) (1 - \pi_j(1))^{\ell_j - X_{ij}} (\pi_j(1))^{X_{ij}} \right)$$

$$= \prod_{i=1}^{n} \left( (1 - \theta) \ G(\pi_1(0), \ldots, \pi_m(0); X_i) \right.$$  

$$\left. + \theta G(\pi_1(1), \ldots, \pi_m(1); X_i) \right)$$

(3.2)

where $\Psi = (\theta, \pi_1(1), \ldots, \pi_m(1), \pi_1(0), \ldots, \pi_m(0))^T$, and

$$G(\pi_1(y_i), \ldots, \pi_m(y_i); X_i) = P(X_{i1} = x_{i1}, \ldots, X_{im} = x_{im}|Y_i = y_i)$$

(3.3)
is the likelihood of $X_i$, given the reference value $Y_i$.

This model treats the differences among raters as fixed effects, because the parameters $\pi_j(1)$ and $\pi_j(0)$ reflect sensitivity and specificity of the raters individually, not of a population of raters. Consequently, inferences based on the model apply solely to the particular raters who take part in the experiment. Would one want to regard the raters in the experiment as a sample from the population of raters, one should consider a random effects model.

### 3.2 Identifiability

Model (3.2) is parametric. This means that the model specifies the essential form of the probability distribution, but leaves some degrees of freedom in the form of parameters to be estimated. A parameterization is a map from the Euclidian space, domain of the parameters, to the corresponding space of distributions. A restriction on a parameterization is that it must be identifiable. We quote the definition of identifiability from Bickel and Doksum (2001):

**Definition 1** A parameterization is called *identifiable* if it is one-to-one. That is, let $\xi_1$ and $\xi_2$ be two parameter values with their corresponding distributions $P_{\xi_1}$ and $P_{\xi_2}$, then $\xi_1 \neq \xi_2$ implies $P_{\xi_1} \neq P_{\xi_2}$.

One distinguishes between two kinds of identifiability, *local* and *global*. Local identifiability means that within a small enough neighbourhood no two values for the parameters result in the same distribution. Global identifiability holds within the whole parameter domain, and guarantees uniqueness.

The latent class model (3.2) is in its general form not identifiable. To demonstrate this, we quote a theorem of Yakowitz and Spragins (see Titterington, Smith, and Makov, 1985) that specifies when a class of finite mixture distributions is identifiable. A parameterization plus the domain of the parameters define a class of distributions. The parameterization of model (3.2) induces a class of finite mixture distributions, for the latent class model is a mixture of the distributions given in equation (3.3). Let $\mathcal{F}$ be the class of distribution functions from which mixtures are formed, and define $\mathcal{G}$ as the class of finite mixtures of $\mathcal{F}$. Then:

**Theorem 1** A necessary and sufficient condition for $\mathcal{G}$ to be identifiable is that $\mathcal{F}$ is a linearly independent set over the field of real numbers, $\mathbb{R}$.

Let $\mathcal{B}$ be the class of products of binomial distributions:

$$\mathcal{B} = \left\{ \prod_{j=1}^{m} B(\pi_j(\cdot); \ell_j) \mid \pi_j(\cdot) \in [0,1]; \ m \geq 1; \ \ell_j \in \mathbb{N} \right\},$$

and define the class of mixtures of two products of binomial distributions:

$$\mathcal{M} = \{ M = \theta B_1 + (1-\theta)B_2 \mid \theta \in [0,1]; B_1, B_2 \in \mathcal{B} \}.$$ 

The latent class model (3.2) is an element of the class $\mathcal{M}$. For this model it can be shown that the condition of theorem 1 is violated. To this end choose any vector of parameters $\Psi' = (\theta', \pi_1'(1), \ldots, \pi_m'(1), \pi_1'(0), \ldots, \pi_m'(0))^T$, and define

$$\Psi^* = (1-\theta', \pi_1'(0), \ldots, \pi_m'(0), \pi_1'(1), \ldots, \pi_m'(1))^T.$$ 

(3.4)
Then, for any realization of $x$:

$$P(X = x; \Psi') = P(X = x; \Psi^*)$$

This violates the linear independency condition in theorem 1 as well as definition 1 and shows that the latent class model (3.2) is not identifiable.

We now study whether $\mathcal{B}$ contains a subclass that is identifiable, and hence under which conditions the latent class model is identifiable.

### 3.2.1 Main result

The main result is:

**Theorem 2** *For global identifiability of model (3.2) it is sufficient to require the following:

$$\theta \neq 0, \quad \theta \neq 1 \quad \text{and} \quad \pi_j(1) > \pi_j(0) \quad \text{for} \quad j = 1, \ldots, m,$$

and

$$\prod_{j=1}^{m} (\ell_j + 1) - 1 \geq 2m + 1. \quad (3.6)$$

**Remark 1:** The restricted parameter space is a connected subspace of the original $2m + 1$ dimensional unit cube, a desired property for many maximum likelihood procedures.

**Remark 2:** One can require either $\pi_j(1) > \pi_j(0)$ or $\pi_j(1) < \pi_j(0)$ for each rater $j \in A$. Choosing the former and using

$$\max\{c_1, c_2\} \geq \theta c_1 + (1 - \theta) c_2 \quad \text{for all} \quad \theta \in [0, 1],$$

it implies, when $\ell_j = 1$,

$$\pi_j(1) \geq \theta \pi_j(1) + (1 - \theta) \pi_j(0) = P(X_{ij} = 1). \quad (3.7)$$

This states that the probability of a rater measuring an object as good is less than the probability that he rates it as good given that it is good. When we still assume $\ell_j = 1$, Bayes’ theorem gives:

$$P(X_{ij} = 1|Y_i = 1)P(Y_i = 1) = P(Y_i = 1|X_{ij} = 1)P(X_{ij} = 1).$$

Combining this and equation (3.7) implies $P(Y_i = 1) \leq P(Y_i = 1|X_{ij} = 1)$. Thus, the measurement of any rater is useful in assessing the reference value of the object. Therefore we have chosen to formulate the restriction in theorem 2 such that it is in line with the philosophy of the problem of measurement system analysis.

**Remark 3:** The strict inequality $\pi_j(0) \neq \pi_j(1)$ for all $j$ arises naturally from the model, since $\pi_j(0) = \pi_j(1)$ implies:

$$P(X_{ij} = x_i|Y_i = y) = P(X_{ij} = x_i) \quad \text{for all} \quad x_i \text{ and } y.$$ 

This means that rater $j$ measures independently of the object. This violates the fundamental conditional independence assumption of the model.

**Remark 4:** It turns out that $\pi_j(1) > \pi_j(0)$ is not required for all operators. However, it is
On the latent class model

convenient (from an interpretational point of view) to state the condition in this way.

**Remark 5:** To explain intuitively condition (3.6), consider a balanced design: \( \ell_j = \ell \) for all \( j \). As the number of raters equals \( m \), the model involves \( 2m + 1 \) parameters. For this number of raters there are \((\ell + 1)^m\) different potential responses, which are subject to the restriction that the sum of the probabilities of these outcomes should equal one. Therefore the model can only be identifiable if \((\ell + 1)^m - 1 \geq 2m + 1 \). If this is not satisfied, the map from the parameter space to the outcome space is one from a \( 2m + 1 \) dimensional space to a lower dimensional one. The implicit function theorem (Stromberg, 1981) implies that local identifiability is not possible. This has also been pointed out by McHugh (1954) and Goodman (1974).

### 3.2.2 Mixed factorial moments

For the proof of theorem 2 we need the concept of mixed factorial moments. This concept is introduced here. If \( X_{i1}, \ldots, X_{im} \) are random variables, their mixed factorial moments are, for \( a_1, \ldots, a_m \in \mathbb{N} \) (we take zero to be included in \( \mathbb{N} \)):

\[
E \left( \prod_{j=1}^{m} X_{ij}(X_{ij} - 1) \cdots (X_{ij} - a_j + 1) \right).
\]

(3.8)

We define the mixed factorial moments for binomial random variables slightly differently from (3.8), namely as:

\[
\mu(a_1, \ldots, a_m) = E \left( X_{i1}^{(a_1)} \cdots X_{im}^{(a_m)} \right)
= E \left( \prod_{j=1}^{m} \frac{X_{ij}(X_{ij} - 1) \cdots (X_{ij} - a_j + 1)}{\ell_j(\ell_j - 1) \cdots (\ell_j - a_j + 1)} \right).
\]

(3.9)

This can be rewritten to:

\[
\mu(a_1, \ldots, a_m) = E \left( \prod_{j=1}^{m} \frac{X_{ij}(X_{ij} - 1) \cdots (X_{ij} - a_j + 1)}{\ell_j(\ell_j - 1) \cdots (\ell_j - a_j + 1)} \bigg| Y_i = 0 \right) P(Y_i = 0)
+ E \left( \prod_{j=1}^{m} \frac{X_{ij}(X_{ij} - 1) \cdots (X_{ij} - a_j + 1)}{\ell_j(\ell_j - 1) \cdots (\ell_j - a_j + 1)} \bigg| Y_i = 1 \right) P(Y_i = 1),
\]

and due to the conditional independence this becomes:

\[
\mu(a_1, \ldots, a_m) = P(Y_i = 0) \prod_{j=1}^{m} E \left( \frac{X_{ij}(X_{ij} - 1) \cdots (X_{ij} - a_j + 1)}{\ell_j(\ell_j - 1) \cdots (\ell_j - a_j + 1)} \bigg| Y_i = 0 \right)
+ P(Y_i = 1) \prod_{j=1}^{m} E \left( \frac{X_{ij}(X_{ij} - 1) \cdots (X_{ij} - a_j + 1)}{\ell_j(\ell_j - 1) \cdots (\ell_j - a_j + 1)} \bigg| Y_i = 1 \right).
\]

(3.10)

It is possible to give an explicit expression for (3.10). Cramér (1974) showed that if \( X \) is binomially distributed with parameters \( p \) and \( n \), the factorial moments are given by

\[
E \left( X(X - 1) \cdots (X - a + 1) \right) = n(n - 1) \cdots (n - a + 1) p^a.
\]

(3.11)
3.2 Identifiability

This enables us to rewrite (3.10). Because \( X_{ij} \) is, conditionally on the event \( Y_i = y_i \), distributed as \( B(\pi_j(y_i); \ell_j) \), we apply (3.11), which shows that (3.10) under the latent class model is equal to:

\[
\mu(a_1, \ldots, a_m) = \begin{cases} 
(1 - \theta) \prod_{j=1}^m \pi_{aj}(0) + \theta \prod_{j=1}^m \pi_{aj}(1) & \text{if } 0 \leq a_j \leq \ell_j \text{ for all } j, \\
0 & \text{if } \exists j \text{ such that } \ell_j < a_j.
\end{cases}
\]

(3.12)

Factorial moments will be denoted in the next section with the use of unit vectors \( e_j = (0, \ldots, 0, 1, 0, \ldots, 0) \) (all entries zero, except for the \( j \)-th, which is 1). For example:

\[
\mu_{2e_1 + e_2} = \mu(2, 1, 0, \ldots, 0) = E \left( X_{i1}^{(2)} X_{i2}^{(1)} X_{i3}^{(0)} \cdots X_{im}^{(0)} \right) = E \left( \frac{X_{i1}(X_{i1} - 1)X_{i2}}{\ell_1(\ell_1 - 1)\ell_2} \right)
\]

3.2.3 Proof of theorem 2

First we study some particular cases (lemma 1 through lemma 3), prerequisites to prove the general case (lemma 4). In the proof of lemma 1 through lemma 3 we start with local identifiability. Successively, we show global identifiability. We realize that the latter implies the former, however results from local identifiability will be used later in section 3.3.1.

**Lemma 1** For \( m = 1, \ell = 3 \) (one rater, three replications) model (3.2) is identifiable (under the restrictions mentioned in theorem 2).

**Proof:** Given a distribution we can construct its moments. This is done for the latent class model in section 3.2.2 (the mixed factorial moments are a linear combination of the moments). The distribution under study is identifiable if the parameters can be expressed uniquely in terms of the moments. Therefore, to show identifiability we must establish a link between the mixed factorial moments and the model parameters. For local identifiability this link must be one-to-one in a small enough neighbourhood of any point of the parameter space. This is guaranteed by the inverse function theorem (Stromberg, 1981) if the matrix of all partial derivatives (the Jacobi-matrix) of this link has a nonzero determinant (the Jacobian).

Restriction (3.6) is satisfied. Define the map \( W \) from the parameter space to the 3-dimensional unit cube:

\[
W(\Psi) = W \left( \begin{array}{c} \theta \\ \pi_1(1) \\ \pi_1(0) \end{array} \right) \rightarrow \left( \begin{array}{c} \mu_{e_1} \\ \mu_{2e_1} \\ \mu_{3e_1} \end{array} \right).
\]

(3.13)

Calculation of the Jacobian of this map yields:

\[
\begin{vmatrix} 
\pi_1(1) - \pi_1(0) & \theta & 1 - \theta \\
\pi_1(1)^2 - \pi_1(0)^2 & 2 \theta \pi_1(1) & 2(1 - \theta) \pi_1(0) \\
\pi_1(1)^3 - \pi_1(0)^3 & 3 \theta \pi_1(1)^2 & 3(1 - \theta) \pi_1(0)^2 
\end{vmatrix} = \theta (1 - \theta) (\pi_1(1) - \pi_1(0))^3.
\]

The zeros of the right hand part are \( \theta = 0, \theta = 1 \) and \( \pi_1(1) = \pi_1(0) \). These are excluded from the parameter space by the conditions (3.5). We have thus shown that the determinant of \( W \) is nonzero everywhere in the parameter domain, excluding these zeros. The inverse function theorem now gives local identifiability.
For global identifiability it must be shown that the map $W$ from parameters to the mixed factorial moments is one-to-one in the whole domain. This is done by expressing the parameters in terms of the mixed factorial moments. The solution for this particular problem is given in Blischke (1962) and is briefly outlined here. After manipulation of the factorial moments, we write (see also relations 1 and 2 in appendix A of this chapter):

$$\begin{align*}
\pi_1(1) + \pi_1(0) &= \frac{\mu_{3e_1} - \mu_{2e_1}\mu_{e_1}}{\mu_{2e_1} - \mu_{e_1}\mu_{e_1}} = b_1, \\
\pi_1(1) \pi_1(0) &= \frac{\mu_{3e_1}\mu_{e_1} - \mu_{2e_1}\mu_{2e_1}}{\mu_{2e_1} - \mu_{e_1}\mu_{e_1}} = b_2, \\
\theta &= \frac{\mu_{e_1} - \pi_1(0)}{\pi_1(1) - \pi_1(0)},
\end{align*}$$

From this we solve $\theta, \pi_1(1)$ and $\pi_1(0)$:

$$\begin{align*}
\pi_1(1) &= \frac{1}{2} \left( b_1 \pm \sqrt{b_1^2 - 4b_2} \right), \\
\pi_1(0) &= \frac{1}{2} \left( b_1 \mp \sqrt{b_1^2 - 4b_2} \right), \\
\theta &= \frac{2\mu_{e_1} - b_1 \pm \sqrt{b_1^2 - 4b_2}}{\pm 2\sqrt{b_1^2 - 4b_2}}.
\end{align*}$$

Taking $\pi_1(1) > \pi_1(0)$ from (3.5) only one solution for each parameter remains, consequently this fixes the solution for $\theta$. Thus, there is a 1-1 relationship between model parameters and mixed factorial moments, and therefore with the distribution. We have global identifiability under restriction (3.5).

**Lemma 2** For $m = 2, \ell_1 = 2, \ell_2 = 1$ (two raters, one with two replicates, the other with one) model (3.2) is identifiable (under the restrictions mentioned in theorem 2).

**Proof:** Again restriction (3.6) is satisfied. Redefine $W$ as:

$$W (\Psi) = W \begin{pmatrix} \theta \\ \pi_1(1) \\ \pi_2(1) \\ \pi_1(0) \\ \pi_2(0) \end{pmatrix} \rightarrow \begin{pmatrix} \mu_{e_1} \\ \mu_{e_2} \\ \mu_{e_1+e_2} \\ \mu_{2e_1+e_2} \end{pmatrix}.$$ 

The Jacobian of $W$ is ($DW$ is the Jacobi-matrix):

$$\det(DW) = \theta^2(1 - \theta)^2(\pi_1(1) - \pi_1(0))^4(\pi_2(1) - \pi_2(0)).$$

Restriction (3.5) excludes the zeros of the righthand side from the parameter space and, thus, ensures local identifiability.

To prove global identifiability we give, as in lemma 1, explicit expressions for the parameters in terms of the mixed factorial moments. To this end we have (see also relations 1 and 2 in appendix A of this chapter):

$$\begin{align*}
\pi_1(1) + \pi_1(0) &= \frac{\mu_{2e_1+e_2} - \mu_{2e_1}\mu_{e_2}}{\mu_{e_1+e_2} - \mu_{e_1}\mu_{e_2}} = c_1, \\
\pi_2(1) + \pi_2(0) &= \frac{\mu_{2e_1+e_2} + \mu_{2e_1}\mu_{e_2} - 2\mu_{e_1+e_2}\mu_{e_1}}{\mu_{2e_1} - \mu_{e_1}\mu_{e_1}} = c_2.
\end{align*}$$
3.2 Identifiability

\[
\frac{1}{4} (\pi_1(1) - \pi_1(0)) (\pi_2(1) - \pi_2(0)) = \mu_{e_1+e_2} - \frac{c_2}{2} \mu_{e_1} - \frac{c_1}{2} \mu_{e_2} + \frac{c_1 c_2}{4} = c_3, \quad (3.16.c)
\]

\[
4 \theta (1 - \theta) = \frac{1}{c_3} (\mu_{e_1+e_2} - \mu_{e_1} \mu_{e_2}) = c_4, \quad (3.16.d)
\]

\[
\frac{1}{4} (\pi_1(1) - \pi_1(0))^2 = \frac{1}{c_4} (\mu_{e_2} - \mu_{e_1} \mu_{e_1}) = c_5, \quad (3.16.e)
\]

We solve these equations for \( \theta, \pi_1(1), \pi_2(1), \pi_1(0) \) and \( \pi_2(0) \). Combining (3.16.a) and (3.16.e) produces solutions for \( \pi_1(1) \) and \( \pi_1(0) \). These can be substituted in \( \mu_{e_1} \) (see relation 3 in appendix A of this chapter), which results in a solution for \( \theta \). Next, taking the square root of (3.16.e) and substituting the result into (3.16.c) gives, when using (3.16.b), the solutions for \( \pi_2(1) \) and \( \pi_2(0) \). The following solutions are arrived at:

\[
\pi_1(1) = \frac{c_1}{2} \pm \sqrt{c_5}, \quad \text{(3.17.a)}
\]

\[
\pi_1(0) = \frac{c_1}{2} \mp \sqrt{c_5}, \quad \text{(3.17.b)}
\]

\[
\pi_2(1) = \frac{c_2 \sqrt{c_5} \pm 2 c_3}{2 \sqrt{c_5}}, \quad \text{(3.17.c)}
\]

\[
\pi_2(0) = \frac{c_2 \sqrt{c_5} \mp 2 c_3}{2 \sqrt{c_5}}, \quad \text{(3.17.d)}
\]

\[
\theta = \frac{2 \mu_{e_1} - c_1 \pm 2 \sqrt{c_5}}{\pm 4 \sqrt{c_5}}. \quad \text{(3.17.e)}
\]

If we observe restriction (3.5) in the assignment of the solutions to the parameters, we are left with one solution for each parameter. This proves global identifiability.

**Lemma 3** For \( m = 3, \ell_1 = 1, \ell_2 = 1 \) and \( \ell_3 = 1 \) model (3.2) is identifiable (under the restrictions mentioned in theorem 2).

**Proof:** Restriction (3.6) is satisfied. Modifying the map \( W \) – see equation (3.13) – in a natural manner to the present situation, and calculating the determinant of its Jacobian yield:

\[
\det(DW) = \theta^3(1 - \theta)^3 (\pi_1(1) - \pi_1(0))^2 (\pi_2(1) - \pi_2(0))^2 (\pi_3(1) - \pi_3(0))^2.
\]

The zeros of the righthand side are excluded from the parameter space by restriction (3.5), thus guaranteeing local identifiability.

We show the global identifiability of model (3.2) by relating its parameters one–to–one to its mixed factorial moments. To this end, we manipulate the mixed factorial moments in the following way (see also relation 4 in appendix A of this chapter):

\[
\pi_1(1) + \pi_1(0) = \gamma_1 = \frac{\mu_{e_2+e_3} \mu_{e_1} - \mu_{e_1+e_3} \mu_{e_2} + \mu_{e_1+e_2+e_3} - \mu_{e_1+e_2} \mu_{e_3}}{\mu_{e_2+e_3} - \mu_{e_2} \mu_{e_3}}. \quad (3.18)
\]

Similar relations for the other \( \pi_j(1) \) and \( \pi_j(0) \) are obtained by permutation of the indices. Then, define:

\[
A_{12} = \mu_{e_1+e_2} - \frac{\gamma_1}{2} \mu_{e_2} - \frac{\gamma_2}{2} \mu_{e_1} + \frac{\gamma_1 \gamma_2}{4} = \left( \frac{\gamma_1}{2} - \pi_1(0) \right) \left( \frac{\gamma_2}{2} - \pi_2(0) \right). \quad (3.19)
\]
We define $A_{13}$ and $A_{23}$ analogously by permuting the indices. This gives the solutions of $\pi_j(1)$ and $\pi_j(0)$ for all $j$, e.g.,

$$\pi_1(1) = \frac{\gamma_1}{2} \pm \sqrt{\frac{A_{12} A_{13}}{A_{23}}},$$

$$\pi_1(0) = \frac{\gamma_1}{2} \mp \sqrt{\frac{A_{12} A_{13}}{A_{23}}}.$$  \hspace{1cm} (3.20.a)

Permutation of the indices yields the solutions for the other $\pi_j(1)$ and $\pi_j(0)$. To find the solution for $\theta$ one selects a mixed factorial moment and substitutes the solutions for the necessary $\pi_j(y)$ (see relation 3 in appendix A of this chapter). This yields, for instance when selecting $\mu_{e_1}$:

$$\theta = \frac{\mu_{e_1} - \pi_1(0)}{\pi_1(1) - \pi_1(0)}.$$  \hspace{1cm} (3.21)

Restriction (3.5) assures that each parameter can be expressed in a unique way in terms of the mixed factorial moments. Thus, global identifiability is proved.

**Lemma 4** For $m > 3$, model (3.2) is identifiable (under the restrictions mentioned in theorem 2).

**Proof:** Adding more repetitions has no consequences for the issue of identifiability. Adding more raters has, because that introduces two additional parameters. Suppose that we have global identifiability for $m = m_0$. Now consider $m = m_0 + 1$ and adjust the map $W(\Psi)$ for the case of $m_0 + 1$ raters. Suppose that model (3.2) is not globally identifiable for $m = m_0 + 1$, then there would be $\Psi'$ and $\Psi^*$ such that $W(\Psi') = W(\Psi^*)$ with $\Psi' \neq \Psi^*$. Consider only the entries of $W(\Psi)$ that are associated with the first $m_0$ raters:

$$W : (\theta, \pi_1(1), \ldots, \pi_{m_0}(1), \pi_1(0), \ldots, \pi_{m_0}(0))$$

$$\rightarrow \left\{ \mu_{\sum_{j=1}^{m_0} a_j}, a_j \leq \ell_j \right\}$$  \hspace{1cm} (3.22)

For this restricted map (3.22) – by assumption – global identifiability holds. Thus, $\theta' = \theta^*$, $\pi_j'(1) = \pi_j^*(1)$ and $\pi_j'(0) = \pi_j^*(0)$ for $j = 1, \ldots, m_0$. If we substitute this in the equation $W(\Psi') = W(\Psi^*)$ for $m_0 + 1$ raters, this yields:

$$(\pi_{m_0+1}^*(0) - \pi_{m_0+1}^*(0)) \cdot (1 - \theta) \prod_{j=1}^{m} \pi_j^a(0)$$

$$= (\pi_{m_0+1}'(1) - \pi_{m_0+1}'(1)) \cdot \theta \prod_{j=1}^{m} \pi_j^a(1).$$  \hspace{1cm} (3.23)

This (over-determined) system of linear equations in two unknowns can have zero, one or an infinite number of solutions. Since $\theta = 0$ and $\theta = 1$ are excluded and for at least one rater $\pi_j(0) \neq \pi_j(1)$, there is only one solution, namely:

$$\pi_{m_0+1}^*(0) - \pi_{m_0+1}^*(0) = 0 = \pi_{m_0+1}'(1) - \pi_{m_0+1}'(1).$$

Therefore, $\Psi' = \Psi^*$. Thus, if for $m = m_0$ there is a one-to-one relation between the model parameters and the distribution, there is as well for $m = m_0 + 1$. Since global identifiability is
proved for $m = 3$ (lemma 3), the proof of lemma 4 follows.

**Remark 1:** Nowhere in the proof of lemma 4 we have used that $\pi_{m_0+1}(1) > \pi_{m_0+1}(0)$. Therefore, condition (3.5) in theorem 2 is sufficient for identifiability but not necessary.

**Remark 2:** For global identifiability of model (3.2) criterion (3.5) needs not be imposed on the parameters relating to rater $m_0 + 1$. To obtain solutions for parameters $\pi_{m_0+1}(1)$ and $\pi_{m_0+1}(0)$ in terms of the mixed factorial moments manipulate $\mu_{e_{m_0+1}}$ and $\mu_{e_{m_0}+e_{m_0+1}}$ (the choice of $m_0$ is arbitrary):

$$
\pi_{m_0+1}(1) = \frac{\mu_{e_{m_0}+e_{m_0+1}} - \pi_{m_0}(0) \mu_{e_{m_0+1}}}{\theta (\pi_{m_0}(1) - \pi_{m_0}(0))},
$$

(3.24.a)

$$
\pi_{m_0+1}(0) = \frac{\pi_{m_0}(1) \mu_{e_{m_0+1}} - \mu_{e_{m_0}+e_{m_0+1}}}{(1 - \theta) (\pi_{m_0}(1) - \pi_{m_0}(0))}.
$$

(3.24.b)

The righthand side of the equations above are linear in the mixed factorial moments. Thus, (3.24.a) and (3.24.b) give a one-on-one relation between $\pi_{m_0+1}(1)$ and $\pi_{m_0+1}(0)$ and the mixed factorial moments. Therefore, global identifiability still holds without (3.5) applying to the rater number $m_0 + 1$.

Combining lemmas 1 through 4 and remark 5 in section 3.2.1 gives the proof of theorem 2.

3.3 Estimation of the model parameters

In this section we develop two procedures for the estimation of the parameters of the latent class model: the method of moments and the maximum likelihood method. We also study the variance of the estimators of each procedure.

3.3.1 Method of moments

Here we apply the method of moments to obtain estimators for the parameters of the latent class model. This consists of finding expressions for all parameters in terms of the moments of the distribution. Then, estimators for the moments are substituted in these expressions, resulting in estimators for the parameters. Moreover, we give the asymptotic distribution of the parameter estimators.

Equation (3.12) gives the relations between the parameters and the mixed factorial moments of the latent class model. While proving the identifiability of the latent class model, we showed in the proof of lemma 3 that the following expressions from the relations (for $m \geq 3$) can be derived:

$$
\theta = \frac{\mu_{e_1} - \pi_1(0)}{\pi_1(1) - \pi_1(0)},
$$

(3.25.a)

$$
\pi_1(1) = \frac{\gamma_1}{2} + \sqrt{\frac{A_{12} A_{13}}{A_{23}}},
$$

(3.25.b)

$$
\pi_1(0) = \frac{\gamma_1}{2} - \sqrt{\frac{A_{12} A_{13}}{A_{23}}},
$$

(3.25.c)
\[ \pi_2(1) = \frac{\gamma_2}{2} + \frac{\sqrt{A_{12} A_{23}}}{A_{13}}, \]  
\[ \pi_2(0) = \frac{\gamma_2}{2} - \frac{\sqrt{A_{12} A_{23}}}{A_{13}}, \]  
\[ \pi_3(1) = \frac{\gamma_3}{2} + \frac{\sqrt{A_{13} A_{23}}}{A_{12}}, \]  
\[ \pi_3(0) = \frac{\gamma_3}{2} - \frac{\sqrt{A_{13} A_{23}}}{A_{12}}, \]  
\[ \pi_j(1) = \frac{\mu_{e_{j-1}+e_i} - \pi_j(0)}{\theta(\pi_j(1) - \pi_j(0))} \]  
\[ \pi_j(0) = \frac{\pi_j(1)-\nu_j}{(1-\theta)(\pi_j(1)-\pi_j(0))} \]  
for all \( j > 3 \).

see equations (3.21), (3.20.a), (3.20.b), (3.19), (3.18), (3.24.a) and (3.24.b). For the solutions of the parameters when \( m = 2 \) we refer to equations (3.16.a), (3.16.b), (3.16.c), (3.16.d), (3.16.e), (3.17.a), (3.17.b), (3.17.c), (3.17.d) and (3.17.e), and when \( m = 1 \) to equations (3.14.a), (3.14.b), (3.14.c), (3.15.a), (3.15.b) and (3.15.c). Alternative relations may be constructed by involving different moments. The relations in appendix A may be of assistance when trying to achieve this.

To arrive at estimators for the parameters we replace the mixed factorial moments by their estimators in the expressions (3.25.a) through (3.25.i). Hereto, let \( X \), as defined in (3.1), be the outcome of the measurement system analysis experiment. We estimate \( \mu(a_1,...,a_m) \) (see equation 3.9) by its mixed factorial sample moment, defined by:

\[ \hat{\mu}(a_1,...,a_m) = \frac{1}{n} \sum_{i=1}^{n} \prod_{j=1}^{m} \frac{X_{ij}(X_{ij} - 1) \cdots (X_{ij} - a_j + 1)}{\ell_j(\ell_j - 1) \cdots (\ell_j - a_j + 1)}. \]  

Substituting these in the equations above yields estimates for \( \Psi = (\theta, \pi_1(1), \ldots, \pi_m(1), \pi_1(0), \ldots, \pi_m(0))^T \), that we denote \( \hat{\Psi}(\hat{\mu}) \).

We now derive the asymptotic distribution of the moment estimators.

A sequence \( \{X_n\} \) converges in distribution to \( X \) if \( F_{X_n}(x) \rightarrow F_X(x) \) for every point \( x \) where \( F_X \) is continuous. A sequence \( \{X_n\} \) is asymptotically normal with mean \( \mu_n \) and variance \( \sigma_n^2 \) (denoted as \( X_n \) is \( AN(\mu_n, \sigma_n^2) \)) if for every \( x \) we have

\[ P \left( \frac{X_n - \mu_n}{\sigma_n} \leq x \right) \rightarrow \Phi(x) \quad \text{if} \quad n \rightarrow \infty. \]

These notions are naturally extended to the situation where \( \{X_n\} \) is a sequence of vectors.

Define

\[ R_k = \frac{1}{n} \sum_{i=1}^{n} X_i^k, \]

having mean \( \rho_k \). We quote two results from Serfling (1980).

**Theorem 3** If \( \rho_{2k} < \infty \) then the random vector \( n^{1/2}(R_1 - \rho_1, \ldots, R_k - \rho_k) \) converges - as \( n \rightarrow \infty \) - in distribution to a \( k \)-variate normal with mean vector \( (0, \ldots, 0) \) and covariance matrix \([\sigma_{ij}]_{k \times k}\), where \( \sigma_{ij} = \rho_{i+j} - \rho_i \rho_j \).
Theorem 4 Suppose that $X_n = (X_{n1}, \ldots, X_{nk})$ is $AN(\mu, n^{-1}\Sigma)$, with $\Sigma$ a covariance matrix. Let $g(X)$ be a real-valued function having a nonzero differential at $x = \mu$. Then

$$g(X_n) \sim AN\left(g(\mu), \frac{1}{n} \sum_{i=1}^k \sum_{j=1}^k \sigma_{ij} \frac{\partial g}{\partial x_i}\bigg|_{x=\mu} \frac{\partial g}{\partial x_j}\bigg|_{x=\mu}\right).$$ (3.27)

Theorems 3 and 4 specify, under the assumption that regularity conditions hold, to which distribution a function of random variables converges when the sample size increases. We point out the relevance of these results for finding out the distribution of the moment estimators constructed above. Finally, we show that the moment estimators satisfy the regularity conditions and arrive at the limiting distribution of these estimators.

We apply theorem 3 to the mixed factorial sample moments and their means, the mixed factorial moments. Under the assumption that the relevant mixed factorial moments are finite, theorem 3 states that the vector of mixed factorial sample moments is asymptotically normal. This is a prerequisite for theorem 4. The moment estimators $\hat{\Psi}(\hat{\mu})$ constructed above are functions of the mixed factorial sample moments. We apply theorem 4 to them. If we assume that $\hat{\Psi}(\mu)$ has a nonzero Jacobian (the equivalent of the nonzero differential in the multi parameter case) at $\mu = \hat{\mu}$, it follows that the moment estimators given by the substitution of (3.26) in equations (3.25.a) through (3.25.i) are asymptotically normal.

We show that the conditions of theorem 3 and theorem 4 are satisfied by the mixed factorial moments and the estimators of the parameters. For the condition of theorem 3 observe that the mixed factorial sample moments are defined such that:

$$0 \leq \hat{\mu}_{\sum_{j=1}^m a_j e_j} \leq 1 \quad \text{with } a_j \in \mathbb{N} \text{ for } j = 1, \ldots, m.$$

This implies that

$$\mu_{\sum_{j=1}^m a_j e_j} = E\left(\hat{\mu}_{\sum_{j=1}^m a_j e_j}\right) \leq E(1) = 1 < \infty.$$

Thus, the mixed factorial moments are all finite and satisfy the condition of theorem 3.

Remains to show that the conditions of theorem 4 are met. The asymptotic normality of the mixed factorial moments is given by theorem 3. The estimator function $\hat{\Psi}(\hat{\mu})$ is more-dimensional, therefore the second condition of theorem 4 changes to a nonzero determinant of the Jacobi–matrix of $\hat{\Psi}(\hat{\mu})$. To this end we have given, in appendix B, the partial derivatives of the estimators (only for the situation where $m = 1, 2, 3$ and 4). All these partial derivatives are well defined because the identifiability condition (3.5) prevents the denominators from being zero. Moreover, identifiability ensures that the map $\hat{W}$, from the model parameters to the mixed factorial moments (as defined in the proof of theorem 2), is invertible. In fact $\hat{\Psi}(\mu)$ is the inverse of $\hat{W}$. Furthermore, invertible functions have nonzero Jacobians, and the Jacobian of their inverse is the reciprocal of the Jacobian (Stromberg, 1981). Thus, as restriction (3.5) guarantees identifiability it ensures that the Jacobian of $\hat{\Psi}(\mu)$ is nonzero (confer the Jacobians calculated in lemma's 1, 2 and 3), the second condition of theorem 4.

We have shown that theorem 3 and theorem 4 apply, and thus know the distribution of the moment estimators. To obtain an expression for the variance of this distribution we specify the terms in equation (3.27). The partial derivatives of the estimators with respect to the parameters are given in appendix B. This leaves us to derive the covariances of the mixed factorial sample
moments. Let $a_j, b_j \in \mathbb{N}$ for all $j$, then

\[
\text{Cov} \left( \hat{\mu}^{n}_{\sum_{j=1}^{m} a_j, e_j}, \hat{\mu}^{n}_{\sum_{j=1}^{m} b_j, e_j} \right) =
\]

\[
= E \left( \hat{\mu}^{n}_{\sum_{j=1}^{m} a_j, e_j} \hat{\mu}^{n}_{\sum_{j=1}^{m} b_j, e_j} \right)
- E \left( \hat{\mu}^{n}_{\sum_{j=1}^{m} a_j, e_j} \right) E \left( \hat{\mu}^{n}_{\sum_{j=1}^{m} b_j, e_j} \right)
\]

\[
= P(Y = 1) E \left( \hat{\mu}^{n}_{\sum_{j=1}^{m} a_j, e_j} \hat{\mu}^{n}_{\sum_{j=1}^{m} b_j, e_j} \bigg| Y = 1 \right)
+ P(Y = 0) E \left( \hat{\mu}^{n}_{\sum_{j=1}^{m} a_j, e_j} \hat{\mu}^{n}_{\sum_{j=1}^{m} b_j, e_j} \bigg| Y = 0 \right)
- \left( P(Y = 1) E \left( \hat{\mu}^{n}_{\sum_{j=1}^{m} a_j, e_j} \bigg| Y = 1 \right) \right)
+ P(Y = 0) E \left( \hat{\mu}^{n}_{\sum_{j=1}^{m} a_j, e_j} \bigg| Y = 0 \right)
\]

\[
\times \left( P(Y = 1) E \left( \hat{\mu}^{n}_{\sum_{j=1}^{m} b_j, e_j} \bigg| Y = 1 \right) \right)
+ P(Y = 0) E \left( \hat{\mu}^{n}_{\sum_{j=1}^{m} b_j, e_j} \bigg| Y = 0 \right)
\]

\[
= \theta \text{Cov} \left( \hat{\mu}^{n}_{\sum_{j=1}^{m} a_j, e_j}, \hat{\mu}^{n}_{\sum_{j=1}^{m} b_j, e_j} \bigg| Y = 1 \right)
+ (1 - \theta) \text{Cov} \left( \hat{\mu}^{n}_{\sum_{j=1}^{m} a_j, e_j}, \hat{\mu}^{n}_{\sum_{j=1}^{m} b_j, e_j} \bigg| Y = 0 \right)
+ \theta (1 - \theta)
\]

\[
\times \left( E \left( \hat{\mu}^{n}_{\sum_{j=1}^{m} a_j, e_j} \bigg| Y = 1 \right) - E \left( \hat{\mu}^{n}_{\sum_{j=1}^{m} a_j, e_j} \bigg| Y = 0 \right) \right)
\]

\[
\times \left( E \left( \hat{\mu}^{n}_{\sum_{j=1}^{m} b_j, e_j} \bigg| Y = 1 \right) - E \left( \hat{\mu}^{n}_{\sum_{j=1}^{m} b_j, e_j} \bigg| Y = 0 \right) \right) \quad (3.28)
\]

In this we have

\[
E \left( \hat{\mu}^{n}_{\sum_{j=1}^{m} a_j, e_j} \bigg| Y = y \right) = \prod_{j=1}^{m} E \left( \hat{\mu}_{a_j, e_j} \bigg| Y = y \right) = \prod_{j=1}^{m} \pi_{a_j}^{(y)} \quad (3.29)
\]

and

\[
\text{Cov} \left( \hat{\mu}^{n}_{\sum_{j=1}^{m} a_j, e_j}, \hat{\mu}^{n}_{\sum_{j=1}^{m} b_j, e_j} \bigg| Y = y \right) =
\]

\[
= \prod_{j=1}^{m} E \left( \hat{\mu}_{a_j, e_j} \bigg| Y = y \right) \times \prod_{j=1}^{m} E \left( \hat{\mu}_{b_j, e_j} \bigg| Y = y \right)
\]

\[
\times \prod_{j=1}^{m} \text{Cov} \left( \hat{\mu}_{a_j, e_j}, \hat{\mu}_{b_j, e_j} | Y = y \right)
\]

\[
= \prod_{j=1}^{m} \pi_{a_j}^{(y)} \times \prod_{j=1}^{m} \pi_{b_j}^{(y)} \times \prod_{j=1}^{m} \text{Cov} \left( \hat{\mu}_{a_j, e_j}, \hat{\mu}_{b_j, e_j} | Y = y \right).
\quad (3.30)
\]
3.3 Estimation of the model parameters

To obtain an approximation of (3.30) the following result (Blischke, 1964) can be used:

\[
\text{Cov}(\hat{\mu}_{a_j,e_j}, \hat{\mu}_{b_j,e_j} \mid Y = y) = \frac{(\ell_j - a_j)(\ell_j - a_j - 1) \cdots (\ell_j - a_j - b_j + 2)}{n \ell_j(\ell_j - 1) \cdots (\ell_j - b_j + 1)} 
\times \left( (\ell_j - a_j - b_j + 1) (\pi_j(y))^{a_j+b_j} + \frac{a_j b_j (\pi_j(y))^{a_j+b_j-1}}{n^{a_j}(\pi_j(y))^{b_j}} \right) + O\left( \frac{1}{\ell_j^2} \right).
\]

Thus, (3.28), (3.29) and (3.30) and appendix B specify all the terms used in (3.27). This gives the distribution of the moment estimators.

We have – by means of the method of moments – obtained estimators for the parameters of the latent class model, see equations (3.25.a) through (3.25.i). Moreover, their asymptotic distribution is given in (3.27).

3.3.2 Method of maximum likelihood

In this section we line out how to find estimators by means of the maximum likelihood method. Here we briefly introduce the idea of maximum likelihood. The maximum likelihood method employ the E-M algorithm to arrive at estimators for the model parameters. Therefore, the E-M algorithm and its properties are described. Finally, the distribution of the estimators produced by the E-M algorithm is given.

We describe the method of maximum likelihood. Suppose we have a sequence of independent random variables \(X_1, X_2, \ldots, X_n\). The density of the distribution of each \(X_i\) is given by \(f(X; \xi)\), with \(\xi\) an unknown parameter. Given realizations \(x_1, \ldots, x_n\) of \(X_1, \ldots, X_n\), the likelihood is defined by:

\[
L(x_1, \ldots, x_n; \xi) = \prod_{i=1}^{n} f(x_i; \xi).
\]

(3.31)

This is a function from \(\mathbb{R}\) to \(\mathbb{R}\).

In the situation where \(X_1, \ldots, X_n\) are discrete random variables, the likelihood \(L(x_1, \ldots, x_n; \xi)\) is:

\[
L(x_1, \ldots, x_n; \xi) = \prod_{i=1}^{n} P_\xi(X_i = x_i).
\]

\(L(x_1, \ldots, x_n; \xi)\) can be regarded as a measure of how ‘likely’ \(\xi\) has produced the realizations \(x_1, \ldots, x_n\). The method of maximum likelihood consists of finding that value of \(\xi\) that produces the highest likelihood for a given sample \(X_1, \ldots, X_n\). This is denoted by:

\[
\hat{\xi} = \arg \max_{\xi} L(x_1, \ldots, x_n; \xi).
\]

One finds \(\hat{\xi}\) by solving

\[
\frac{\partial L}{\partial \xi} = 0.
\]
with the restriction that
\[
\frac{\partial^2 L}{\partial \xi^2} < 0
\]
to exclude minima and other stationary points except maxima. The thus found \( \xi \) maximizes the likelihood function and is called the maximum likelihood estimate.

The theory above can be extended to the multiple parameter situation. In this case to find the maxima one equates the gradient to zero, and imposes a similar requirement as above on the second order partial derivatives. For the latent class model the maximum likelihood estimate is defined by:
\[
\hat{\Psi} = \arg\max_{\Psi} L(\mathbf{X}; \Psi),
\]
with \( L(\mathbf{X}; \Psi) \) defined by:
\[
L(\mathbf{X}; \Psi) = \prod_{i=1}^{n} \left[ (1 - \theta) \prod_{j=1}^{m} \left( \ell_{ij} \right)^{X_{ij}} \left( 1 - \pi_j(0) \right)^{I_{ij} - X_{ij}} \left( \pi_j(0) \right)^{X_{ij}} + \theta \prod_{j=1}^{m} \left( \ell_{ij} \right)^{X_{ij}} \left( 1 - \pi_j(1) \right)^{I_{ij} - X_{ij}} \left( \pi_j(1) \right)^{X_{ij}} \right]
\]
\[
= \prod_{i=1}^{n} \left( (1 - \theta) G(\pi_1(0), \ldots, \pi_m(0); \mathbf{X}_i) \right.
\]
\[
+ \theta G(\pi_1(1), \ldots, \pi_m(1); \mathbf{X}_i)) \right),
\]
as in equation (3.2).

To find the maximum likelihood estimates of \( \Psi \), instead of applying the traditional Newton-Raphson algorithm, one uses the so-called E-M algorithm (McLachlan and Krishnan, 1997). The E-M algorithm approaches the problem of maximizing the likelihood function (3.2) indirectly by exploiting the more convenient form of a related likelihood function. This reduces the complexity of the maximum likelihood estimation. Moreover, the E-M algorithm has the appealing property that the likelihood function of interest is not decreased with each iteration. The E-M algorithm is frequently used in the context of censored data, truncated distributions and mixture distributions, among others.

The E-M algorithm makes use of the concept of ‘incomplete information’. We introduce this and show how it applies to the latent class model.

Model (3.2) involves a latent variable, which is unobserved. This lack of information can be viewed as a case of ‘incomplete information’, as only the observations of the raters are at one’s disposal, while the reference values of the objects are not given.

To deal with incomplete information, we introduce new variables:
\[
Z_{i,0} = \begin{cases} 
1 & \text{if the reference value of object } i \text{ is } 0 \\
0 & \text{if the reference value of object } i \text{ is } 1 
\end{cases}
\]
and
\[
Z_{i,1} = \begin{cases} 
1 & \text{if the reference value of object } i \text{ is } 1 \\
0 & \text{if the reference value of object } i \text{ is } 0 
\end{cases}
\]
The \( n \times 2 \)-matrix \( \mathbf{Z} \) indicates the reference values for the objects. Thus, where \( \mathbf{X} \) does not give the complete information, \( (\mathbf{X}, \mathbf{Z}) \) contains the ‘complete’ information.
For the latent class model the incomplete likelihood function is given by (3.2). The likelihood function corresponding to complete information situation is:

\[
L_c((X, Z); \Psi) = \prod_{i=1}^{n} \{ (1 - \theta) G(\pi_1(0), \ldots, \pi_m(0); X_i) \}^{Z_{i,0}} \times \{ \theta G(\pi_1(1), \ldots, \pi_m(1); X_i) \}^{Z_{i,1}} . \tag{3.32}
\]

Note that only one of \(Z_{i,0}\) and \(Z_{i,1}\) can be equal to one. Taking the logarithm, one gets:

\[
\ln(L_c((X, Z); \Psi)) \propto \sum_{i=1}^{n} \left( Z_{i,0} \ln(1 - \theta) + Z_{i,1} \ln(\theta) \right)
+ Z_{i,0} \sum_{j=1}^{m} (X_{ij} \ln(\pi_j(0)) + (\ell_j - X_{ij}) \ln(1 - \pi_j(0)))
+ Z_{i,1} \sum_{j=1}^{m} (X_{ij} \ln(\pi_j(1)) + (\ell_j - X_{ij}) \ln(1 - \pi_j(1))) .
\tag{3.33}
\]

This complete information log-likelihood function has a rather convenient form: to maximize \(L_c\) one needs to find the zeros of the partial derivatives (with respect to the individual parameters), which comes down to solving equations involving only one parameter.

The E-M algorithm – applied to the estimation of \(\Psi\) in model (3.2) – can be described as follows (McLachlan and Krishnan, 1997):

**Step 1**

Choose initial values for the estimate \(\hat{\Psi}^{(0)}\), and specify a stopping criterion. Stopping criteria usually specify a maximum number of iterations to be performed by the algorithm or a minimum distance, say \(\varepsilon\), between two successive iterations that is to be achieved:

\[
\left\| \hat{\Psi}^{(t+1)} - \hat{\Psi}^{(t)} \right\| < \varepsilon.
\]

**Step 2** (Referred to as the “E-step”)

We have no knowledge of \(Z\). However, using the current estimate of \(\hat{\Psi}^{(t)}\) we replace \(Z\) by its conditional expectation given \(X\):

\[
\hat{Z}_{i,0} = E_{\hat{\Psi}^{(t)}}(Z_{i,0} | X), \quad \hat{Z}_{i,1} = E_{\hat{\Psi}^{(t)}}(Z_{i,1} | X).
\]

In the present situation these are probabilities that are complementary, i.e., \(\hat{Z}_{i,1} = 1 - \hat{Z}_{i,0}\). In fact real values from the interval \([0, 1]\) are substituted for \(Z\), while their proper value is either 0 or 1. Furthermore, using Bayes’ theorem:

\[
E_{\hat{\Psi}^{(t)}}(Z_{i,1} | X) = P_{\hat{\Psi}^{(t)}}(Z_{i,1} | X) = P_{\hat{\Psi}^{(t)}}(Y_i = 1 | X) = \frac{P_{\hat{\Psi}^{(t)}}(X | Y_i = 1) P_{\hat{\Psi}^{(t)}}(Y_i = 1)}{P_{\hat{\Psi}^{(t)}}(X)} .
\]
We calculate this expectation and find:

\[
\hat{Z}_{i,t} = \frac{\hat{\theta}(t)G(\hat{x}_{i1}(t), \ldots, \hat{x}_{im}(t); X_i)}{\hat{\theta}(t)G(\hat{x}_{i1}(t), \ldots, \hat{x}_{im}(t); X_i) + (1 - \hat{\theta}(t))G(\hat{x}_{i1}(0), \ldots, \hat{x}_{im}(0); X_i)}. \number{3.35}
\]

**Step 3** (Referred to as the M-step)

The M-step consists of maximizing the log-likelihood function, which is a linear combination of functions of single parameters. Taking the first order partial derivatives and equating them to zero, one arrives at the following set of equations to solve:

\[
\sum_{i=1}^{n} \left( \frac{\hat{Z}_{i,t}}{\hat{\theta}(t+1)} - \frac{\hat{Z}_{i,0}}{1 - \hat{\theta}(t+1)} \right) = 0, \number{3.36.a}
\]

\[
\sum_{i=1}^{n} \left( X_{ij} \frac{\hat{Z}_{i,t}}{\hat{\pi}_{j,t+1}(1)} - (\ell_j - X_{ij}) \frac{\hat{Z}_{i,t}}{1 - \hat{\pi}_{j,t+1}(1)} \right) = 0, \number{3.36.b}
\]

\[
\sum_{i=1}^{n} \left( X_{ij} \frac{\hat{Z}_{i,t}}{\hat{\pi}_{j,t}(0)} - (\ell_j - X_{ij}) \frac{\hat{Z}_{i,t}}{1 - \hat{\pi}_{j,t}(0)} \right) = 0. \number{3.36.c}
\]

Let \( X_j \) be the \( j \)-th column of the matrix \( X \) and \( 1 = (1, \ldots, 1)^T \) of length \( n \). This yields the estimates:

\[
\hat{\theta}(t+1) = \frac{1}{n} \sum_{i=1}^{n} \hat{Z}_{i,1} = \frac{1}{n} \left(1, \hat{Z}_{t}^{(1)}\right),
\]

\[
\hat{\pi}_{j,t+1}(1) = \frac{\sum_{i=1}^{n} X_{ij} \hat{Z}_{i,t}}{\ell_j \sum_{i=1}^{n} \hat{Z}_{i,t}} = \frac{1}{\ell_j} \left(1, \hat{Z}_{t}^{(1)}\right),
\]

\[
\hat{\pi}_{j,t+1}(0) = \frac{\sum_{i=1}^{n} X_{ij} \hat{Z}_{i,t}}{\ell_j \sum_{i=1}^{n} \hat{Z}_{i,t}} = \frac{1}{\ell_j} \left(1, \hat{Z}_{t}^{(1)}\right).
\]

Thus, the next estimate \( \hat{\Psi}_{t+1} \) of the parameters \( \Psi \) is obtained.

**Step 4**

Go back to step 2 until the stopping criterion of the algorithm has been satisfied.

Applying the E-M algorithm yields the maximum likelihood estimator of \( \Psi \) for the complete information situation.

We now show that the maximum likelihood estimator produced by the E-M algorithm not only maximizes \( L_c \) but also the original incomplete information likelihood function \( L \). This is done by showing that \( L \) does not decrease after each iteration of the E-M algorithm, and that the sequence of iterated estimates converges to a stationary point. Thus, the estimate from the E-M algorithm maximizes \( L \) and converges to a maximum likelihood estimate.

By definition after each iteration of the E-M algorithm the value of \( L_c \) is increased. We show that the incomplete information likelihood function is also increased after each iteration, in formula:

\[
L_c \left(X; \hat{\Psi}_{t+1} \right) \geq L_c \left(X; \hat{\Psi}_t \right) \implies L \left(X; \hat{\Psi}_{t+1} \right) \geq L \left(X; \hat{\Psi}_t \right).
\]
As \( L_c((x,z);\Psi) = P_{\Psi}((X,Z) = (x,z)) \) and \( L(x;\Psi) = P_{\Psi}(X = x) \), one can view \( L_c((x,Z);\Psi)/L(x;\Psi) \) as the conditional density of \((X,Z)\) given \( X = x \). Then rewrite:

\[
\ln(L(X;\Psi)) = \ln(L_c((X,Z);\Psi)) - \ln(L_c((X,Z);\Psi)/L(X;\Psi)).
\]

Taking expectations on both sides with respect to the conditional distribution of \((X,Z)\) given \( X = x \), and substituting the latest iteration of \( \hat{\Psi}^{(t)} \) for \( \Psi \), one arrives at:

\[
\ln(L(x;\Psi)) = E_{\hat{\Psi}^{(t)}}(\ln(L_c((X,Z);\Psi)) | X = x) - E_{\hat{\Psi}^{(t)}}(\ln(L_c((X,Z);\Psi)/L(X;\Psi)) | X = x),
\]

because

\[
E_{\hat{\Psi}^{(t)}}(\ln(L(X;\Psi)) | X = x) = \sum_z \ln(L(x;\Psi))P_{\Psi}^{(t)}((X,Z) = (x,z)|X = x) = \ln(L(x;\Psi)) \sum_z P_{\Psi}^{(t)}((X,Z) = (x,z)|X = x) = \ln(L(x;\Psi)).
\]

Using this in the difference of successive iterations substituted in the incomplete information loglikelihood function we get:

\[
\begin{align*}
\ln & \left(L \left(x; \hat{\Psi}^{(t+1)} \right) \right) - \ln \left(L \left(x; \hat{\Psi}^{(t)} \right) \right) \\
& = E_{\hat{\Psi}^{(t)}} \left( \ln \left(L_c \left((X,Z); \hat{\Psi}^{(t+1)} \right) \right) \bigg| X = x \right) \\
& \quad - E_{\hat{\Psi}^{(t)}} \left( \ln \left(L_c \left((X,Z); \hat{\Psi}^{(t)} \right) \right) \bigg| X = x \right) \\
& \quad - E_{\hat{\Psi}^{(t)}} \left( \ln \left( \frac{L_c \left((X,Z); \hat{\Psi}^{(t+1)} \right)}{L \left(X; \hat{\Psi}^{(t+1)} \right)} \right) \bigg| X = x \right) \\
& \quad + E_{\hat{\Psi}^{(t)}} \left( \ln \left( \frac{L_c \left((X,Z); \hat{\Psi}^{(t)} \right)}{L \left(X; \hat{\Psi}^{(t)} \right)} \right) \bigg| X = x \right).
\end{align*}
\]

The difference between the first two terms on the right hand side of the equation above is (due to the M step) always non-negative. So we are only concerned about the remaining terms on the right hand side. These can be rewritten to:

\[
\begin{align*}
& E_{\hat{\Psi}^{(t)}} \left( \ln \left( \frac{L_c \left((X,Z); \hat{\Psi}^{(t+1)} \right)}{L \left(X; \hat{\Psi}^{(t+1)} \right)} \right) \bigg| X = x \right) \\
& \quad - E_{\hat{\Psi}^{(t)}} \left( \ln \left( \frac{L_c \left((X,Z); \hat{\Psi}^{(t)} \right)}{L \left(X; \hat{\Psi}^{(t)} \right)} \right) \bigg| X = x \right) \\
& \quad = E_{\hat{\Psi}^{(t)}} \left( \ln \left( \frac{L_c \left((X,Z); \hat{\Psi}^{(t+1)} \right) L \left(X; \hat{\Psi}^{(t)} \right) }{L \left(X; \hat{\Psi}^{(t+1)} \right) L_c \left((X,Z); \hat{\Psi}^{(t)} \right)} \right) \bigg| X = x \right).
\]


Because of the concavity of the logarithm function, we can apply Jensen's inequality:

\[
E_{\Psi^{(t)}} \left( \ln \left( \frac{L_c((X, Z); \hat{\Psi}^{(t+1)})}{L(X; \hat{\Psi}^{(t+1)})} \frac{L(X; \hat{\Psi}^{(t)})}{L_c((X, Z); \hat{\Psi}^{(t)})} \right) \bigg| X = x \right) 
\leq \ln \left( \frac{L_c((X, Z); \hat{\Psi}^{(t+1)})}{L(X; \hat{\Psi}^{(t+1)})} \frac{L(X; \hat{\Psi}^{(t)})}{L_c((X, Z); \hat{\Psi}^{(t)})} \right) 
= \ln \left( \sum_{z} \frac{L_c((x, z); \hat{\Psi}^{(t+1)})}{L(x; \hat{\Psi}^{(t+1)})} \frac{P_{\Psi^{(t)}}(X = x)}{P_{\hat{\Psi}^{(t+1)}}((X, Z) = (x, z))} \right) 
= \ln \left( \frac{\sum_{z} P_{\Psi^{(t+1)}}((X, Z) = (x, z))}{P_{\hat{\Psi}^{(t+1)}}(X = x)} \right) 
= \ln(1) 
= 0.
\]

Thus, it has been shown that \( L(X; \Psi) \) does not decrease after an E-M iteration.

To see that the sequence \( \{ \hat{\Psi}^{(t)} \} \) converges to a stationary value of the likelihood function, we quote a theorem by Wu (McLachlan and Krishnan, 1997).

**Theorem 5** Suppose that \( E_{\Psi_1} \left( \ln (L_c((X, Z); \Psi_2)) \right| X = x \) is continuous in both \( \Psi_1 \) and \( \Psi_2 \). Then all the limit points of any sequence \( \{ \hat{\Psi}^{(t)} \} \) generated by the E-M algorithm are stationary points of \( L(\Psi) \), and \( \{ L((X, Z); \hat{\Psi}^{(t)}) \} \) converges monotonically to some value \( L^* = L(\Psi^*) \) for some stationary point \( \Psi^* \).

The regularity condition of theorem 5 is satisfied because \( L_c(\Psi) \) is a polynomial. This allows us to apply theorem 5, and we conclude that employing the E-M algorithm yields a stationary point for the parameters of the latent class model. Moreover, due to the identifiability of the model, this is the maximum likelihood estimator.

Next we obtain the asymptotic distribution of the maximum likelihood estimator. To this end we invoke a theorem from Lehmann (1983). This theorem states that the maximum likelihood estimator is – if the likelihood function satisfies regularity conditions – asymptotically normal.

We first give the assumptions, A1 through A4, under which the theorem on the asymptotic distribution of the maximum likelihood estimators holds. Here, define the parameter space of the latent class model as

\[
\Omega = \{ \Psi | \theta \in (0, 1) \text{ and } 1 > \pi_j(1) > \pi_j(0) > 0 \text{ for all } j \}.
\]

and denote

\[
\Psi = (\theta, \pi_1(1), \ldots, \pi_m(1), \pi_1(0), \ldots, \pi_m(0)) = (\Psi_1, \ldots, \Psi_{2m+1}).
\]
3.4 Confidence intervals

Then, the assumptions are:

A1: There exists an open subset \( \omega \) of \( \Omega \) containing the true parameter point \( \Psi_0 \) such that for almost all \( X \) the density function \( L(X; \Psi) \) admits all third order partial derivatives

\[
\frac{\partial^3 L(X; \Psi)}{\partial \Psi_p \partial \Psi_q \partial \Psi_r},
\]

for all \( \Psi \in \omega \).

A2: The first order logarithmic derivatives of \( L(X; \Psi) \) satisfy the equations:

\[
E_\Psi \left( \frac{\partial}{\partial \Psi_p} \ln (L(X; \Psi)) \right) = 0 \quad \text{for all } p, \tag{3.38}
\]

and

\[
[I(X; \Psi)]_{pq} = E_\Psi \left( -\frac{\partial^2}{\partial \Psi_p \partial \Psi_q} \ln (L(X; \Psi)) \right)
\]

exists and is finite for \( p, q = 1, \ldots, 2m + 1 \) and for all \( \Psi \in \omega \).

A3: The matrix \( I(X; \Psi) \) is positive definite for all \( \Psi \in \omega \).

A4: There exists functions \( M_{pqr}(X) \) independent of \( \Psi \) such that for all \( p, q, r = 1, \ldots, 2m + 1 \) we have

\[
\left| \frac{\partial^3 \ln (L(X; \Psi))}{\partial \Psi_p \partial \Psi_q \partial \Psi_r} \right| \leq M_{pqr}(X) \quad \text{for all } \Psi \in \omega,
\]

where

\[
E_\Psi_0 (M_{pqr}(X)) < \infty.
\]

The theorem then becomes:

**Theorem 6** Let the likelihood function satisfy assumptions A1, A2, A3 and A4, then there exists a sequence of solutions \( \Psi_n \) to the likelihood equations such that \( \hat{\Psi}_n \rightarrow \Psi \) and

\[
\hat{\Psi}_n \sim AN(\Psi, n^{-1}I(\Psi)^{-1}), \tag{3.39}
\]

where \( I(\Psi)^{-1} \) is the (Fisher) information matrix.

Our situation is analogue to Boyles (2001). He simply uses the result of theorem 6 without verifying the conditions, whose verification is indeed a highly technical matter.

In the appendix C the elements of the information matrix are specified.

We have shown how to arrive at maximum likelihood estimators for the latent class model parameters by means of the E-M algorithm. Moreover, the asymptotical distribution of these estimators is given.

**3.4 Confidence intervals**

In previous sections we have constructed estimators for the parameters of the latent class model. Furthermore, we have given the variance of these estimators. These results enable us to analyze a measurement system analysis experiment. The analysis of such an experiment yields
estimates of all parameters of the latent class model. However, it is unlikely that estimates will be exactly equal to the true value of the population parameters. Moreover, different samples (experiments) will produce different estimates. To cope with the degree of uncertainty associated with point estimates they are expressed in combination with confidence intervals. These confidence intervals are also used to compare estimation procedures.

We give a definition of confidence intervals, and show how they can be constructed using the estimators and their variance obtained in the previous section 3.3:

**Definition 2** The random interval $[L(X), U(X)]$, where $L(X)$ and $U(X)$ are statistics obtained from the sampling distribution, is a $100(1 - \alpha)\%$ confidence interval for the parameter $\nu$ if

$$P(L(X) \leq \nu \leq U(X)) \geq 1 - \alpha.$$ 

Dealing with a multi-dimensional parameter vector, one can obtain confidence intervals for each parameter individually. However, this will only specify the ranges for the individual parameters irrespective of the value of the other parameters. A method that does take into account the correlation between the different estimates seems more appropriate here. We illustrate such a method.

Let the $2m + 1$ parameter vector $\Psi$ be estimated by the vector $\hat{\Psi}_n = (\hat{\Psi}_n, \ldots, \hat{\Psi}_{n,2m+1})$.

It is assumed that $\hat{\Psi}_n$ is $\mathcal{N}(\Psi, n^{-1} \Sigma_{\Psi})$, with $\Sigma_{\Psi}$ its covariance matrix that is assumed to be non-singular. Then, (Serfling, 1980, p. 140) an ellipsoidal $100(1 - \alpha)\%$ confidence region for $\Psi$ is given by:

$$P_{\Psi}(\Psi \in CI_n) \geq 1 - \alpha,$$

where

$$CI_n = \left\{ \Psi : n(\hat{\Psi}_n - \Psi)\Sigma_{\Psi}^{-1}(\hat{\Psi}_n - \Psi)^T \leq \chi_{2m+1,1-\alpha}^2 \right\}.$$ 

This confidence ellipsoid characterizes the region in the parameter space that contains the true value of the parameter with probability $1 - \alpha$.

An alternative method to obtain confidence intervals is based on the profile likelihood (Boyles, 2001). However, this only applies to the maximum likelihood estimator and cannot be used for method-of-moment estimators. Therefore, we disregard this here. It should be noted that the normal approximation can be rather crude as small sample behaviour can be rather different from the asymptotic behaviour.

In the present situation we have two estimation procedures at hand, the method of moments and the maximum likelihood procedure. Both yield estimates and for both we can construct confidence ellipsoids around these estimates. One may wish to compare these two estimation methods. Two estimation procedures are compared on the bases of the volumes of their confidence ellipsoids (it can be shown that such a comparison is independent of the choice of $\alpha$). A sequence $\left\{ \hat{\Psi}_n^{(1)} \right\}$ corresponds to asymptotically smaller confidence ellipsoids than a sequence $\left\{ \hat{\Psi}_n^{(2)} \right\}$ if and only if

$$|\Sigma_{\Psi}^{(1)}| \leq |\Sigma_{\Psi}^{(2)}|.$$ 

Thus, we have that $\left\{ \hat{\Psi}_n^{(1)} \right\}$ is better – in the sense of having smaller confidence ellipsoids – than $\left\{ \hat{\Psi}_n^{(2)} \right\}$. 


We have introduced confidence intervals as a way to deal with the uncertainty of the estimators of the latent class method. Moreover, they enable us to compare the estimators from the method of moments with the estimators from the maximum likelihood procedure. Further research should reveal which of the two estimation methods is the better, i.e., has the smaller confidence ellipsoids, and thus produces the more precise estimates.

3.5 Goodness-of-fit

In chapter 2 we have proposed the latent class model to describe the outcome of measurement system experiments with a binary measurement. We can actually test whether the model is a good description of the data from the experiment (prerequisites are a conducted experiment and from its data estimated parameters of the latent class model). We show how to assess the appropriateness of the latent class model by means of a goodness-of-fit test. To this end we introduce the goodness-of-fit statistic that is most appropriate for experiments with a binary response and give its reference distribution needed to test the hypothesis of a good fit.

The goodness-of-fit of a latent class model is evaluated by comparing the response frequencies predicted by the model to the observed response pattern frequencies. In fact the following null hypothesis is tested:

$$H_0 : \Psi = \hat{\Psi}$$

(3.40)

How likely this hypothesis is, is evaluated by a goodness-of-fit statistic.

The goodness-of-fit statistic has been generalized by Read and Cressie (1988) to the so-called power-divergence statistic, which is defined as:

$$\frac{2}{\lambda(\lambda + 1)} \sum_{x} \text{observed}_x \left[ \left( \frac{\text{observed}_x}{\text{expected}_x} \right)^\lambda - 1 \right].$$

This measures how far the empirical probability distribution diverges from the hypothesized distribution. It can be shown (using a Taylor-expansion) that the power-divergence statistic is asymptotically approximately $\chi^2$ distributed with $(\ell + 1)^m - 1 - (2m + 1)$ degrees of freedom: the total number of different response patterns minus one (for their frequencies should sum to one) and one subtracted for each parameter estimated.

There are two well-known cases of the power-divergence statistic that are most used in evaluating the $H_0$ -- as formulated in (3.40) -- when dealing with categorical data. One is the Pearson $\chi^2$ statistic, and relates to the substitution $\lambda = 1$, resulting in:

$$\sum_{x} \frac{(\text{observed}_x - \text{expected}_x)^2}{\text{expected}_x}.$$

The other is the loglikelihood statistic $G^2$, that is arrived at by taking the limit $\lambda \to 0$. This yields:

$$G^2 = 2 \sum_{x} \text{observed}_x \ln \left( \frac{\text{observed}_x}{\text{expected}_x} \right).$$

Read and Cressie point out that the Pearson $\chi^2$ statistic and the loglikelihood statistic $G^2$ are not always appropriate to evaluate the null hypothesis in (3.40). They show that for finite sample sizes the $\chi^2$ approximation is not applicable for all $\lambda$, in particular when there are cells with low expected frequencies. Read and Cressie show that the $\chi^2$ approximation of the Pearson
\( \chi^2 \) statistic and the loglikelihood \( G^2 \) statistic is sensitive to cells with low frequencies. To deal with this Read and Cressie have shown that the choice of \( \lambda = \frac{2}{3} \) yields a power-divergence statistic that is most robust against this flaw. Thus, Read and Cressie propose

\[
\frac{9}{5} \sum_{x} \text{observed}_x \left[ \left( \frac{\text{observed}_x}{\text{expected}_x} \right)^{\frac{2}{3}} - 1 \right]
\]

(3.41)

for the evaluation of \( H_0 \) in (3.40).

We adopt the statistic (3.41) for the situation where the latent class model is used to describe the outcome of a measurement system analysis experiment. Because we expect cells with low frequencies when dealing with a precise measurement system, as precise measurement systems mainly produce response patterns \( x = (1, \ldots, 1) \) and \( x = (0, \ldots, 0) \), leaving the other response patterns – where raters disagree – with low (expected) cell frequencies.

We have shown how to test whether the latent class model gives a good description of the outcome of the measurement system analysis experiment. It turns out that the test statistic in (3.41) is best for this purpose. Note that the goodness-of-fit statistic also enables us to compare the two estimation procedures – method of moments and maximum likelihood – with respect to the best fit.
Appendix A: Relation between moments and parameters

In this appendix we give relations between the mixed factorial moments of the latent class model and its parameters. These are used in the construction of moment estimators.

Define the unit vectors $e_j = (0, \ldots, 0, 1, 0, \ldots, 0)$ with the $j$-th entry equalling one and the other entries zero. Then, rewrite

$$
\mu(a_1, \ldots, a_m) = \mu_{a_1}e_1 + \ldots + a_m e_m = \mu_{\sum_{j=1}^m a_j e_j} = E \left( X_{i1}^{(a_1)} \ldots X_{im}^{(a_m)} \right),
$$

where the $X_{ij}$ ($i = 1, \ldots, n$ and $j = 1, \ldots, m$) are distributed as defined by equation (3.2).

Relation 1:

Take $a_1, \ldots, a_m, b_1, \ldots, b_m \in \mathbb{N}$. Then, after tedious algebraic computations one arrives at the following relation:

$$
\theta (1 - \theta) \left( \prod_{j=1}^{m} \pi_j^{a_j}(1) - \prod_{j=1}^{m} \pi_j^{a_j}(0) \right) \left( \prod_{j=1}^{m} \pi_j^{b_j}(1) - \prod_{j=1}^{m} \pi_j^{b_j}(0) \right) = \frac{\mu_{(a+1)e_1} - \mu_{ae_1} \cdot \mu_{e_1}}{\mu_{2e_1} - \mu_{e_1} \cdot \mu_{e_1}} = \frac{\pi_1^q(1) - \pi_1^q(0)}{\pi_1(1) - \pi_1(0)} = \sum_{k=0}^{a-1} \pi_1^{a-k-1}(1 \pi_1^k(0)).
$$

In particular one can show that for $a \in \mathbb{N}$ with $a \geq 1$:

$$
\frac{\mu_{(a+1)e_1} - \mu_{ae_1} \cdot \mu_{e_1}}{\mu_{2e_1} - \mu_{e_1} \cdot \mu_{e_1}} = \frac{\pi_1^q(1) - \pi_1^q(0)}{\pi_1(1) - \pi_1(0)} = \sum_{k=0}^{a-1} \pi_1^{a-k-1}(1 \pi_1^k(0)).
$$

The last equation is a relation between the mixed factorial moments and the parameters that no longer involves $\theta$.

Relation 2:

Take $Q \subset \{1, \ldots, m\}$, and define $a_j, b_j, c_j, d_j \in \mathbb{N}$ such that $a_j + b_j = c_j + d_j$ for all $j \in Q$. Straightforward algebraic manipulations yield the following relationship:

$$
\mu_{\sum_{j \in Q} a_je_j} \cdot \mu_{\sum_{j \in Q} b_je_j} = \mu_{\sum_{j \in Q} c_je_j} \cdot \mu_{\sum_{j \in Q} d_je_j} = \\
\theta (1 - \theta) \left( \prod_{j=1}^{m} \pi_j^{a_j}(1) \pi_j^{b_j}(0) + \prod_{j=1}^{m} \pi_j^{b_j}(1) \pi_j^{a_j}(0) \right) \\
- \prod_{j=1}^{m} \pi_j^{c_j}(1) \pi_j^{d_j}(0) - \prod_{j=1}^{m} \pi_j^{d_j}(1) \pi_j^{c_j}(0) \\
\text{If } a_j, b_j \in \{0, 1\} \text{ for all } j \in Q \text{ and } A = \{ j \in Q \mid a_j = b_j \}, \text{ the previous equation reduces to:}
$$

$$
\mu_{\sum_{j \in Q} a_je_j} \cdot \mu_{\sum_{j \in Q} (1-a_j)e_j} = \\
\mu_{\sum_{j \in Q} b_je_j} \cdot \mu_{\sum_{j \in Q} (1-b_j)e_j} = \\
$$
On the latent class model

\[
\theta (1 - \theta) \left( \prod_{j \in A} \pi_j(a_j) \cdot a - \prod_{j \in A} \pi_j(1 - a_j) \right) \\
\times \left( \prod_{j \in Q \setminus A} \pi_j(a_j) - \prod_{j \in Q \setminus A} \pi_j(1 - a_j) \right)
\]

Also, let \( A = \{ j \in \{1, \ldots, m\} \mid a_j = b_j \} \) and take \( c_j \) such that \( c_j = (a_j + b_j)/2 \) for all \( j \in A \), then:

\[
\mu_{\sum_{j=1}^m \delta_j(e_j, \delta_j e_j)} - \mu_{\sum_{j=1}^m \delta_j(e_j, e_j)} = \\
\theta (1 - \theta) \prod_{j \in A} \pi_j^{c_j}(1) \pi_j^{c_j}(0) \times \left( \prod_{j \notin A} \pi_j^{a_j}(1) \pi_j^{b_j}(0) \right) \\
+ \prod_{j \notin A} \pi_j^{a_j}(0) \pi_j^{b_j}(1) - 2 \prod_{j \notin A} \pi_j^{a_j}(1) \pi_j^{c_j}(0)
\]

In particular, this yields:

\[
\frac{\mu_{(a+b)e_1} \cdot \mu_{(a+b)e_1} - \mu_{a\cdot e_1} \cdot \mu_{b\cdot e_1}}{\mu_{a\cdot e_1} - \mu_{b\cdot e_1}} = \pi_{1}^{a-b}(1) \pi_{1}^{a-b}(0)
\]

The last equation is a relation between the mixed factorial moments and the parameters that no longer involve \( \theta \).

Relation 3:
Using any moment one can express \( \theta \) in terms of the other parameters:

\[
\mu_{\sum_{j=1}^m \delta_j(e_j, \delta_j e_j)} = \theta \prod_{j=1}^m \pi_j^{a_j}(1) + (1 - \theta) \prod_{j=1}^m \pi_j^{a_j}(0) \quad \Rightarrow
\]

\[
\theta = \frac{\mu_{\sum_{j=1}^m \delta_j(e_j, \delta_j e_j)} - \prod_{j=1}^m \pi_j^{a_j}(0)}{\prod_{j=1}^m \pi_j^{a_j}(1) - \prod_{j=1}^m \pi_j^{c_j}(0)} \quad (3.43)
\]

Relation 4:
Let \( Q \subset \{1, \ldots, m\} \) and take \( a_j, b_j \in \{0, 1\} \) for all \( j \in Q \). Define \( A = \{ j \mid a_j = 1 = b_j \} \), and

\[
c_j = \begin{cases} 
1 & \text{if } j \in A, \\
0 & \text{if } j \in Q \setminus A.
\end{cases}
\]

Then, after cumbersome manipulations, the following relationship is arrived at:

\[
\mu_{\sum_{j \in Q} \delta_j(e_j, \delta_j e_j)} - \mu_{\sum_{j=1}^m (1-a_j) e_j} - \mu_{\sum_{j \in Q} b_j e_j} - \mu_{\sum_{j \in Q} (1-b_j) e_j} + \\
\mu_{\sum_{j \in Q} e_j} - \mu_{\sum_{j \in Q} c_j e_j} - \mu_{\sum_{j \in Q} (1-c_j) e_j} = 
\]
\[
\theta (1 - \theta) \times \left( \prod_{j \in A} \pi_j^{b_j}(1) \pi_j^{1-b_j}(0) - \prod_{j \in A} \pi_j^{1-b_j}(1) \pi_j^{b_j}(0) \right) \times \\
\left( \prod_{j \in Q \setminus A} \pi_j^{b_j}(1) \pi_j^{1-b_j}(0) - \prod_{j \in Q \setminus A} \pi_j^{1-b_j}(1) \pi_j^{b_j}(0) \right) + \\
\left( \prod_{j \in A} \pi_j^{b_j}(1) \pi_j^{1-b_j}(1) - \prod_{j \in A} \pi_j^{1-b_j}(0) \pi_j^{b_j}(0) \right) \times \\
\left( \prod_{j \in Q \setminus A} \pi_j^{b_j}(1) \pi_j^{1-b_j}(1) - \prod_{j \in Q \setminus A} \pi_j^{1-b_j}(0) \pi_j^{b_j}(0) \right) 
\]

In particular, if one chooses \( b_j = 1 \) for all \( j \in A \) and if there are \( j_1, j_2 \in Q \setminus A \) such that \( b_{j_1} \neq b_{j_2} \), this simplifies to:

\[
\theta (1 - \theta) \times \left( \prod_{j \in A} \pi_j(1) - \prod_{j \in A} \pi_j(0) \right) \times \\
\left( \prod_{j \in Q \setminus A} \pi_j^{1-b_j}(1) - \prod_{j \in Q \setminus A} \pi_j^{1-b_j}(0) \right) \times \\
\left( \prod_{j \in Q \setminus A} \pi_j^{b_j}(1) + \prod_{j \in Q \setminus A} \pi_j^{b_j}(0) \right) 
\]

(3.44)
Appendix B: Partial derivatives of the moment estimators

This appendix contains the partial derivatives of the moment estimators given in equations (3.25.a) through (3.25.i). In these equations the mixed factorial moments are substituted for their estimates, the mixed factorial sample moments. The partial derivatives are evaluated at $\hat{\mu} = \mu$, and are given for the situations with 1, 2, 3 and 4 raters involved.

Remark: All partial derivatives below are well defined. This is due to the identifiability restriction (3.5), that prevents their denominators to become zero.

One rater

Define

$$\begin{align*}
\hat{\mu} &= (\hat{\mu}_e, \hat{\mu}_{2e}, \hat{\mu}_{3e}) \quad \text{and} \quad \mu = (\mu_e, \mu_{2e}, \mu_{3e}).
\end{align*}$$

The partial derivatives for the situation involving one rater are given by:

$$
\begin{align*}
\frac{\partial \pi_1(1)}{\partial \hat{\mu}_e} &= \frac{\pi_1(0)(\pi_1(0) + 2\pi_1(1))}{\theta(\pi_1(1) - \pi_1(0))^2}, \\
\frac{\partial \pi_1(1)}{\partial \hat{\mu}_{2e}} &= -\frac{(2\pi_1(0) + \pi_1(1))}{\theta(\pi_1(1) - \pi_1(0))^2}, \\
\frac{\partial \pi_1(1)}{\partial \hat{\mu}_{3e}} &= \frac{1}{\theta(\pi_1(1) - \pi_1(0))^2}.
\end{align*}
$$

and

$$
\begin{align*}
\frac{\partial \pi_1(0)}{\partial \hat{\mu}_e} &= \frac{\pi_1(1)(2\pi_1(0) + \pi_1(1))}{(1 - \theta)(\pi_1(1) - \pi_1(0))^2}, \\
\frac{\partial \pi_1(0)}{\partial \hat{\mu}_{2e}} &= -\frac{(\pi_1(0) + 2\pi_1(1))}{(1 - \theta)(\pi_1(1) - \pi_1(0))^2}, \\
\frac{\partial \pi_1(0)}{\partial \hat{\mu}_{3e}} &= \frac{1}{(1 - \theta)(\pi_1(1) - \pi_1(0))^2},
\end{align*}
$$

and

$$
\begin{align*}
\frac{\partial \theta}{\partial \hat{\mu}_e} &= -\frac{6\pi_1(1)\pi_1(0)}{(\pi_1(1) - \pi_1(0))^3}, \\
\frac{\partial \theta}{\partial \hat{\mu}_{2e}} &= \frac{3(\pi_1(1) + \pi_1(0))}{(\pi_1(1) - \pi_1(0))^3}, \\
\frac{\partial \theta}{\partial \hat{\mu}_{3e}} &= \frac{-2}{(\pi_1(1) - \pi_1(0))^3}.
\end{align*}
$$
Two raters

Define

\[ \hat{\mu} = (\hat{\mu}_{e_1}, \hat{\mu}_{2e_1}, \hat{\mu}_{e_2}, \hat{\mu}_{e_1+e_2}, \hat{\mu}_{2e_1+e_2}), \]

and

\[ \mu = (\mu_{e_1}, \mu_{2e_1}, \mu_{e_2}, \mu_{e_1+e_2}, \mu_{2e_1+e_2}). \]

The partial derivatives for the situation involving two raters are given by:

\[
\begin{align*}
\frac{\partial \pi_1(1) }{\partial \hat{\mu}_{e_1} |_{\hat{\mu} = \mu} } &= \frac{(\pi_1(1) + \pi_1(0))\pi_2(0)}{\theta (\pi_1(1) - \pi_1(0))(\pi_2(1) - \pi_2(0))}, \\
\frac{\partial \pi_1(1) }{\partial \hat{\mu}_{2e_1} |_{\hat{\mu} = \mu} } &= \frac{-\pi_2(0)}{\theta (\pi_1(1) - \pi_1(0))(\pi_2(1) - \pi_2(0))}, \\
\frac{\partial \pi_1(1) }{\partial \hat{\mu}_{e_2} |_{\hat{\mu} = \mu} } &= \frac{\pi_1(1))\pi_1(0)}{\theta (\pi_1(1) - \pi_1(0))(\pi_2(1) - \pi_2(0))}, \\
\frac{\partial \pi_1(1) }{\partial \hat{\mu}_{e_1+e_2} |_{\hat{\mu} = \mu} } &= \frac{\pi_1(1))\pi_1(0)}{\theta (\pi_1(1) - \pi_1(0))(\pi_2(1) - \pi_2(0))}, \\
\frac{\partial \pi_1(1) }{\partial \hat{\mu}_{2e_1+e_2} |_{\hat{\mu} = \mu} } &= \frac{1}{\theta (\pi_1(1) - \pi_1(0))(\pi_2(1) - \pi_2(0))},
\end{align*}
\]

and

\[
\begin{align*}
\frac{\partial \pi_1(0) }{\partial \hat{\mu}_{e_1} |_{\hat{\mu} = \mu} } &= \frac{(\pi_1(1) + \pi_1(1))\pi_2(1)}{(1-\theta)(\pi_1(1) - \pi_1(0))(\pi_2(1) - \pi_2(0))}, \\
\frac{\partial \pi_1(0) }{\partial \hat{\mu}_{2e_1} |_{\hat{\mu} = \mu} } &= \frac{-\pi_2(0)}{(1-\theta)(\pi_1(1) - \pi_1(0))(\pi_2(1) - \pi_2(0))}, \\
\frac{\partial \pi_1(0) }{\partial \hat{\mu}_{e_2} |_{\hat{\mu} = \mu} } &= \frac{\pi_1(1))\pi_1(0)}{(1-\theta)(\pi_1(1) - \pi_1(0))(\pi_2(1) - \pi_2(0))}, \\
\frac{\partial \pi_1(0) }{\partial \hat{\mu}_{e_1+e_2} |_{\hat{\mu} = \mu} } &= \frac{-\pi_2(0)}{(1-\theta)(\pi_1(1) - \pi_1(0))(\pi_2(1) - \pi_2(0))}, \\
\frac{\partial \pi_1(0) }{\partial \hat{\mu}_{2e_1+e_2} |_{\hat{\mu} = \mu} } &= \frac{1}{(1-\theta)(\pi_1(1) - \pi_1(0))(\pi_2(1) - \pi_2(0))},
\end{align*}
\]

and

\[
\begin{align*}
\frac{\partial \pi_2(1) }{\partial \hat{\mu}_{e_1} |_{\hat{\mu} = \mu} } &= \frac{2\pi_1(0)\pi_2(1)}{\theta (\pi_1(1) - \pi_1(0))^2}, \\
\frac{\partial \pi_2(1) }{\partial \hat{\mu}_{2e_1} |_{\hat{\mu} = \mu} } &= \frac{-\pi_2(0)}{\theta (\pi_1(1) - \pi_2(0))^2},
\end{align*}
\]
On the latent class model

\[
\frac{\partial \pi_2(1)}{\partial \mu_e} \bigg|_{\mu = \mu} = \frac{\pi_1^2(1)}{\theta (\pi_1(1) - \pi_1(0))^2},
\]

\[
\frac{\partial \pi_2(1)}{\partial \mu_{e1} + e_2} \bigg|_{\mu = \mu} = \frac{-2\pi_1(0)}{\theta (\pi_1(1) - \pi_1(0))^2},
\]

\[
\frac{\partial \pi_2(1)}{\partial \mu_{e2}} \bigg|_{\mu = \mu} = \frac{1}{\theta (\pi_1(1) - \pi_1(0))^2},
\]

and

\[
\frac{\partial \pi_2(0)}{\partial \mu_{e1}} \bigg|_{\mu = \mu} = \frac{2\pi_1(1)\pi_2(0)}{(1 - \theta)(\pi_1(1) - \pi_1(0))^2},
\]

\[
\frac{\partial \pi_2(0)}{\partial \mu_{e2}} \bigg|_{\mu = \mu} = \frac{-\pi_2(0)}{(1 - \theta)(\pi_1(1) - \pi_1(0))^2},
\]

\[
\frac{\partial \pi_2(0)}{\partial \mu_{2e1} + e_2} \bigg|_{\mu = \mu} = \frac{(1 - \theta)(\pi_1(1) - \pi_1(0))^2}{\theta (\pi_1(1) - \pi_1(0))^2},
\]

\[
\frac{\partial \pi_2(0)}{\partial \mu_{e1}} \bigg|_{\mu = \mu} = \frac{-2\pi_1(0)}{(1 - \theta)(\pi_1(1) - \pi_1(0))^2},
\]

\[
\frac{\partial \pi_2(0)}{\partial \mu_{e2}} \bigg|_{\mu = \mu} = \frac{1}{(1 - \theta)(\pi_1(1) - \pi_1(0))^2},
\]

and

\[
\frac{\partial \theta}{\partial \mu_{e1}} \bigg|_{\mu = \mu} = \frac{-2(\pi_1(1)\pi_2(0) + \pi_1(0)\pi_2(1))}{(\pi_1(1) - \pi_1(0))^2(\pi_2(1) - \pi_2(0))},
\]

\[
\frac{\partial \theta}{\partial \mu_{2e1}} \bigg|_{\mu = \mu} = \frac{\pi_2(1) + \pi_2(0)}{(\pi_1(1) - \pi_1(0))^2(\pi_2(1) - \pi_2(0))},
\]

\[
\frac{\partial \theta}{\partial \mu_{e2}} \bigg|_{\mu = \mu} = \frac{-2\pi_1(1)\pi_1(0)}{(\pi_1(1) - \pi_1(0))^2(\pi_2(1) - \pi_2(0))},
\]

\[
\frac{\partial \theta}{\partial \mu_{e1} + e_2} \bigg|_{\mu = \mu} = \frac{2(\pi_1(1) + \pi_1(0))}{(\pi_1(1) - \pi_1(0))^2(\pi_2(1) - \pi_2(0))},
\]

\[
\frac{\partial \theta}{\partial \mu_{2e1} + e_2} \bigg|_{\mu = \mu} = \frac{-2}{(\pi_1(1) - \pi_1(0))^2(\pi_2(1) - \pi_2(0))},
\]

Three raters

Define

\[
\hat{\mu} = (\hat{\mu}_{e1}, \hat{\mu}_{e2}, \hat{\mu}_{e3}, \hat{\mu}_{e1+e2}, \hat{\mu}_{e1+e3}, \hat{\mu}_{e2+e3}, \hat{\mu}_{e1+e2+e3}),
\]

and

\[
\mu = (\mu_{e1}, \mu_{e2}, \mu_{e3}, \mu_{e1+e2}, \mu_{e1+e3}, \mu_{e2+e3}, \mu_{e1+e2+e3}).
\]
The partial derivatives for the situation involving three raters are given by:

\[
\frac{\partial \pi_1(1)}{\partial \mu_{e_1}} = \frac{\pi_2(0) \pi_3(0)}{\pi_1(1) \pi_3(0)}, \\
\frac{\partial \pi_1(1)}{\partial \mu_{e_2}} = \frac{\pi_2(1) - \pi_2(0)}{\pi_1(1) \pi_3(0)}, \\
\frac{\partial \pi_1(1)}{\partial \mu_{e_3}} = \frac{\pi_2(1) - \pi_2(0)}{\pi_1(1) \pi_3(0)}, \\
\frac{\partial \pi_1(1)}{\partial \mu_{e_1+e_2}} = \frac{\pi_2(1) - \pi_2(0)}{\pi_1(1) \pi_3(0)}, \\
\frac{\partial \pi_1(1)}{\partial \mu_{e_1+e_3}} = \frac{\pi_2(1) - \pi_2(0)}{\pi_1(1) \pi_3(0)}, \\
\frac{\partial \pi_1(1)}{\partial \mu_{e_2+e_3}} = \frac{\pi_2(1) - \pi_2(0)}{\pi_1(1) \pi_3(0)}, \\
\frac{\partial \pi_1(1)}{\partial \mu_{e_1+e_2+e_3}} = \frac{\pi_2(1) - \pi_2(0)}{\pi_1(1) \pi_3(0)}.
\]

Permutation of the indices in the partial derivatives above yields the partial derivatives of the other \( \pi_m(i) \).
The partial derivatives of \( \theta \) are given by:

\[
\begin{align*}
\frac{\partial \theta}{\partial \mu_1} &= -\frac{\pi_2(1) \pi_3(0) + \pi_2(0) \pi_3(1)}{(\pi_1(1) - \pi_1(0)) (\pi_2(1) - \pi_2(0)) (\pi_3(1) - \pi_3(0))}, \\
\frac{\partial \theta}{\partial \mu_2} &= -\frac{\pi_1(1) \pi_3(0) + \pi_1(0) \pi_3(1)}{(\pi_1(1) - \pi_1(0)) (\pi_2(1) - \pi_2(0)) (\pi_3(1) - \pi_3(0))}, \\
\frac{\partial \theta}{\partial \mu_3} &= -\frac{\pi_1(1) \pi_2(0) + \pi_1(0) \pi_2(1)}{(\pi_1(1) - \pi_1(0)) (\pi_2(1) - \pi_2(0)) (\pi_3(1) - \pi_3(0))}, \\
\frac{\partial \theta}{\partial \mu_{e_1+e_2}} &= \frac{-2}{(\pi_1(1) - \pi_1(0)) (\pi_2(1) - \pi_2(0)) (\pi_3(1) - \pi_3(0))}.
\end{align*}
\]

Including a fourth rater

Define

\[
\hat{\mu} = (\hat{\mu}_1, \hat{\mu}_2, \hat{\mu}_3, \hat{\mu}_{e_4}, \hat{\mu}_{e_1+e_2}, \hat{\mu}_{e_1+e_3}, \hat{\mu}_{e_2+e_3}, \hat{\mu}_{e_1+e_2+e_3}),
\]

and

\[
\mu = (\mu_1, \mu_2, \mu_3, \mu_{e_4}, \mu_{e_1+e_2}, \mu_{e_1+e_3}, \mu_{e_2+e_3}, \mu_{e_1+e_2+e_3}).
\]

The estimates for the parameters \( \theta, \pi_1(1), \pi_2(1), \pi_3(1), \pi_1(0), \pi_2(0) \) and \( \pi_3(0) \) are unchanged when including a fourth rater. As a consequence their partial derivatives remain the same as for the three rater case. Moreover, the estimates for the parameters \( \theta, \pi_1(1), \pi_2(1), \pi_3(1), \pi_1(0), \pi_2(0) \) and \( \pi_3(0) \) do not involve the moments \( \mu_{e_4} \) and \( \mu_{e_1+e_2} \). Therefore their partial derivatives with respect to these two moments are zero.

The partial derivatives of the estimates of \( \pi_4(1) \) and \( \pi_4(0) \) are given by:

\[
\begin{align*}
\frac{\partial \pi_4(1)}{\partial \mu_1} &= \frac{\pi_2(1) \pi_3(0) (\pi_4(1) - \pi_4(0))}{\theta (\pi_1(1) - \pi_1(0)) (\pi_2(1) - \pi_2(0)) (\pi_3(1) - \pi_3(0))}, \\
\frac{\partial \pi_4(1)}{\partial \mu_2} &= \frac{\pi_1(1) \pi_3(0) (\pi_4(1) - \pi_4(0))}{\theta (\pi_1(1) - \pi_1(0)) (\pi_2(1) - \pi_2(0)) (\pi_3(1) - \pi_3(0))},
\end{align*}
\]

\[
\begin{align*}
\frac{\partial \pi_4(1)}{\partial \mu_e} &= \frac{\pi_2(1) \pi_3(0) (\pi_4(1) - \pi_4(0))}{\theta (\pi_1(1) - \pi_1(0)) (\pi_2(1) - \pi_2(0)) (\pi_3(1) - \pi_3(0))}, \\
\frac{\partial \pi_4(1)}{\partial \mu_{e_1+e_2}} &= \frac{\pi_1(1) \pi_3(0) (\pi_4(1) - \pi_4(0))}{\theta (\pi_1(1) - \pi_1(0)) (\pi_2(1) - \pi_2(0)) (\pi_3(1) - \pi_3(0))}.
\end{align*}
\]
\[
\frac{\partial \pi_4(1)}{\partial \mu_{e_3}} \bigg|_{\mu = \mu} = \frac{\pi_1(0) \pi_2(1) \pi_4(1) - \pi_1(0) \pi_2(0) \pi_4(1)}{\theta (\pi_1(1) - \pi_1(0)) (\pi_2(1) - \pi_2(0)) (\pi_3(1) - \pi_3(0))} + \frac{\pi_1(1) \pi_2(0) \pi_4(1) - \pi_1(1) \pi_2(1) \pi_4(0)}{\theta (\pi_1(1) - \pi_1(0)) (\pi_2(1) - \pi_2(0)) (\pi_3(1) - \pi_3(0))},
\]
\[
\frac{\partial \pi_4(1)}{\partial \mu_{e_1}} \bigg|_{\mu = \mu} = \frac{-\pi_3(0)}{\theta (\pi_3(1) - \pi_3(0))},
\]
\[
\frac{\partial \pi_4(1)}{\partial \mu_{e_1+e_2}} \bigg|_{\mu = \mu} = \frac{\theta (\pi_1(1) - \pi_1(0)) (\pi_2(1) - \pi_2(0)) (\pi_3(1) - \pi_3(0))}{\pi_2(1) \pi_4(1) - \pi_4(0)}.
\]

and
\[
\frac{\partial \pi_4(0)}{\partial \mu_{e_1}} \bigg|_{\mu = \mu} = \frac{\pi_2(0) \pi_3(1) (\pi_4(1) - \pi_4(0))}{\theta (\pi_3(1) - \pi_3(0)) (\pi_2(1) - \pi_2(0)) (\pi_3(1) - \pi_3(0))} - \frac{\pi_1(1) \pi_2(1) \pi_4(0) - \pi_1(0) \pi_2(1) \pi_4(0)}{\theta (\pi_1(1) - \pi_1(0)) (\pi_2(1) - \pi_2(0)) (\pi_3(1) - \pi_3(0))},
\]
\[
\frac{\partial \pi_4(0)}{\partial \mu_{e_2}} \bigg|_{\mu = \mu} = \frac{\pi_1(1) \pi_2(1) \pi_4(0) - \pi_1(0) \pi_2(1) \pi_4(0)}{\theta (\pi_3(1) - \pi_3(0)) (\pi_2(1) - \pi_2(0)) (\pi_3(1) - \pi_3(0))} - \frac{\pi_1(1) \pi_2(0) \pi_4(0) - \pi_1(0) \pi_2(0) \pi_4(1)}{(1 - \theta) (\pi_1(1) - \pi_1(0)) (\pi_2(1) - \pi_2(0)) (\pi_3(1) - \pi_3(0))},
\]
\[
\frac{\partial \pi_4(0)}{\partial \mu_{e_3}} \bigg|_{\mu = \mu} = \frac{1}{\theta (\pi_3(1) - \pi_3(0))},
\]
\[
\frac{\partial \pi_4(0)}{\partial \mu_{e_4}} \bigg|_{\mu = \mu} = \frac{\pi_3(1)}{(1 - \theta) (\pi_3(1) - \pi_3(0))},
\]
\[
\frac{\partial \pi_4(0)}{\partial \mu_{e_1+e_2}} \bigg|_{\mu = \mu} = \frac{1}{\theta (\pi_3(1) - \pi_3(0)) (\pi_2(1) - \pi_2(0)) (\pi_3(1) - \pi_3(0))}.
\]
Appendix C: Elements of the information matrix

In this appendix we specify the elements of the information matrix. Here, we define the information matrix for the incomplete information log-likelihood function as:

\[ I(X; \Psi) = - \frac{\partial^2}{\partial \Psi \partial \Psi^T} \ln(L(X; \Psi)) = - \left[ \frac{\partial^2}{\partial \Psi_p \partial \Psi_q} \ln(L(X; \Psi)) \right]_{1 \leq p \leq 2m+1, 1 \leq q \leq 2m+1}. \]

Similarly, define \( I_c((X, Z); \Psi) \) the information matrix of the complete information log-likelihood function.

Differentiating (3.37) twice with respect to \( \Psi \) and multiplying by -1 yield:

\[ I(X; \Psi) = I_c((X, Z); \Psi) + \frac{\partial^2}{\partial \Psi \partial \Psi^T} \ln \left( \frac{L_c((X, Z); \Psi)}{L(X; \Psi)} \right). \]

Taking the expectation of both sides over the conditional distribution of \( Z \) given \( X = x \), we arrive at:

\[ I(x; \Psi) = E_{\Psi} \left( I_c((X, Z); \Psi) \big| X = x \right) + E_{\Psi} \left( \frac{\partial^2}{\partial \Psi \partial \Psi^T} \ln \left( \frac{L_c((X, Z); \Psi)}{L(X; \Psi)} \right) \big| X = x \right). \]

It can be shown (McLachlan and Krishnan, 1997) that:

\[ -E_{\Psi} \left( \frac{\partial^2}{\partial \Psi \partial \Psi^T} \ln \left( \frac{L_c((X, Z); \Psi)}{L(X; \Psi)} \right) \big| X = x \right) = \text{Cov}_{\Psi} \left( \nabla_{\Psi} \ln(L_c((X, Z); \Psi)) \big| X = x \right) \]
\[ = E_{\Psi} \left( \nabla_{\Psi} \ln(L_c((X, Z); \Psi)) \nabla_{\Psi}^T \ln(L_c((X, Z); \Psi)) \big| X = x \right) - \nabla_{\Psi} \ln(L(x; \Psi)) \nabla_{\Psi}^T \ln(L(x; \Psi)). \]

We now obtain the observed information matrix \( I(X; \Psi) \) by combining the last two equations and evaluating these at the maximum likelihood estimator \( \hat{\Psi} \). After observing that \( \left[ \nabla_{\Psi} \ln(L(X; \Psi)) \right]_{\Psi=\hat{\Psi}} = 0 \), this means:

\[ I(X; \hat{\Psi}) = \left[ E_{\Psi} \left( I_c((X, Z); \Psi) \big| X = x \right) \right]_{\Psi=\hat{\Psi}} \]
\[ - \left[ E_{\Psi} \left( \nabla_{\Psi} \ln(L_c((X, Z); \Psi)) \nabla_{\Psi}^T \ln(L_c((X, Z); \Psi)) \big| X = x \right) \right]_{\Psi=\hat{\Psi}} = A - B. \]  

(3.45)

(3.46)

Obviously, \( A \) and \( B \) are both \((2m+1) \times (2m+1)\) matrices.

We give the entries of the information matrix by specifying the entries of the matrices \( A \)
and $B$. Here $A$ is a diagonal matrix, whose diagonal elements are given by:

$$
[A]_{11} = E_{\Psi}(I_{c,11}((X,Z); \Psi)|X=x) = E_{\Psi}\left( \frac{\partial^2}{\partial \theta^2} \ln(L_c((X,Z); \Psi) \big| X=x) \right) = E_{\Psi}\left( \frac{\sum_{i=1}^{n} \left( \frac{Z_{i,1}}{\theta^2} + \frac{Z_{i,0}}{(1 - \theta)^2} \right) X = x} \right) = \sum_{i=1}^{n} \left( \frac{E_{\Psi}(Z_{i,1}X = x)}{\theta^2} + \frac{E_{\Psi}(Z_{i,0}X = x)}{(1 - \theta)^2} \right). \tag{3.47}
$$

In a similar way one shows that for $2 \leq j \leq m + 1$:

$$
[A]_{jj} = \sum_{i=1}^{n} \left( X_{ij} \frac{E_{\Psi}(Z_{i,1}X = x)}{\pi_j^2(1)} + (\ell_j - X_{ij}) \frac{E_{\Psi}(Z_{i,1}X = x)}{(1 - \pi_j(1))^2} \right), \tag{3.48}
$$

and for $m + 2 \leq j \leq 2m + 1$

$$
[A]_{jj} = \sum_{i=1}^{n} \left( X_{ij} \frac{E_{\Psi}(Z_{i,0}X = x)}{\pi_j^2(0)} + (\ell_j - X_{ij}) \frac{E_{\Psi}(Z_{i,0}X = x)}{(1 - \pi_j(0))^2} \right). \tag{3.49}
$$

For the calculation of the matrix $A$ we use the maximum likelihood estimates of $\Psi$ and replace $E_{\Psi}(Z_{i,1}|X = x)$ by the estimate as given in (3.35).

We now have the entries of $A$, leaves us to obtain those of $B$. To calculate the elements of $B$ we use the relationship:

$$
B = \text{Cov}_{\Psi}(\nabla_{\Psi} \ln(L_c((X,Z); \Psi))|X=x).
$$

As the gradient is given by the left hand side of the equations (3.36.a), (3.36.c) and (3.36.b), we have:

$$
[B]_{11} = \text{Var}_{\Psi}\left( \sum_{i=1}^{n} \left( \frac{Z_{i,1}}{\theta} - \frac{Z_{i,0}}{1 - \theta} \right) \big| X = x \right) = \text{Var}_{\Psi}\left( \frac{-n}{1 - \theta} + \sum_{i=1}^{n} \frac{Z_{i,1}}{\theta (1 - \theta)} \big| X = x \right) = \frac{1}{\theta^2 (1 - \theta)^2} \text{Var}_{\Psi}\left( \sum_{i=1}^{n} Z_{i,1} \big| X = x \right) = \frac{1}{\theta^2 (1 - \theta)^2} \sum_{i=1}^{n} \text{Var}_{\Psi}(Z_{i,1}|X = x) = \frac{1}{\theta^2 (1 - \theta)^2} \sum_{i=1}^{n} E_{\Psi}(Z_{i,1}|X = x) (1 - E_{\Psi}(Z_{i,1}|X = x)).
$$

Analogous reasoning leads to:

$$
[B]_{pq} = [B]_{qp} = \sum_{i=1}^{n} b((p, q); i) E_{\Psi}(Z_{i,1}|X = x) (1 - E_{\Psi}(Z_{i,1}|X = x)).
$$
where \( b((p, q); i) \) is defined as below.

For \( 2 \leq q \leq m + 1 \):

\[
b((1, q); i) = \frac{X_{i,q-1} - \pi_{q-1}(1)\ell_{q-1}}{\theta (1 - \theta) \pi_{q-1}(1) (1 - \pi_{q-1}(1))}.
\]

For \( m + 2 \leq q \leq 2m + 1 \):

\[
b((1, q); i) = \frac{- (X_{i,q-m-1} - \pi_{q-m-1}(0)\ell_{q-m-1})}{\theta (1 - \theta) \pi_{q-m-1}(0) (1 - \pi_{q-m-1}(0))}.
\]

For \( 2 \leq p \leq m + 1 \) and \( 2 \leq q \leq m + 1 \):

\[
b((p, q); i) = \frac{(X_{i,p-1} - \pi_{p-1}(1)\ell_{p-1}) (X_{i,q-1} - \pi_{q-1}(1)\ell_{q-1})}{\pi_{p-1}(1) (1 - \pi_{p-1}(1)) \pi_{q-1}(1) (1 - \pi_{q-1}(1))}.
\]

For \( 2 \leq p \leq 2m + 1 \) and \( m + 2 \leq q \leq 2m + 1 \):

\[
b((p, q); i) = \frac{- (X_{i,p-m-1} - \pi_{p-m-1}(1)\ell_{p-m-1}) (X_{i,q-m-1} - \pi_{q-m-1}(0)\ell_{q-m-1})}{\pi_{p-m-1}(0) (1 - \pi_{p-m-1}(0)) \pi_{q-m-1}(0) (1 - \pi_{q-m-1}(0))}.
\]

For the calculation of the matrix \( B \) we use the maximum likelihood estimates of \( \Psi \) and replace \( \mathbb{E}_{\xi_1 | X = x} \) by the estimate as given in (3.35).

We have thus obtained all entries of matrices \( A \) and \( B \), and thus of the information matrix. This concludes the specification of the parameters of the limiting normal distribution of the maximum likelihood.