Measurement and Research Department Reports 96-5

On the Sampling Interpretation of Confidence Intervals and Hypothesis Tests in the Context of Conditional Maximum Likelihood Estimation

E. Maris

Cito

Measurement and Research Department Reports 96-5

On the Sampling Interpretation of Confidence Intervals and Hypothesis Tests in the Context of Conditional Maximum Likelihood Estimation

E. Maris

Cito
Arnhem, oktober 1996

Abstract

In the context of *conditional maximum likelihood* (CML) estimation, confidence intervals can be interpreted in three different ways, depending on the sampling distribution under which these confidence intervals contain the true parameter value with a certain probability. These sampling distributions are (a) the distribution of the data given the *incidental parameters*, (b) the marginal distribution of the data (i.e., with the incidental parameters integrated out), and (c) the conditional distribution of the data given the sufficient statistics for the incidental parameters. Results on the asymptotic distribution of CML estimates under sampling scheme (c) can be used to construct asymptotic confidence intervals using only the CML estimates. This is not possible for the results on the asymptotic distribution under sampling schemes (a) and (b). However, it is shown that the *conditional* asymptotic confidence intervals are also valid under the other two sampling schemes.

Key-words: CML estimation, confidence intervals, conditional inference.

There are three ways (a, b and c) to interpret inferential statements in the context of *conditional maximum likelihood* (CML) estimation. For instance, an asymptotic 95 percent confidence interval for a CML paramter estimate can be interpreted as the interval that, asymptotically, contains the true parameter value in 95 percent of the samples from (a) the distribution of the data given the *incidental parameters* (i.e., treating them as known), or (b) the marginal distribution of the data (i.e., with the incidental parameters integrated out), or (c) the conditional distribution of the data given the sufficient statistics for the incidental parameters. Obviously, the truth of these statements depends on how the confidence intervals are constructed. In section 2, it is shown that the only confidence intervals that can be constructed using only the CML estimates, are those that are based on the asymptotic distribution of CML estimates under sampling scheme (c). However, from the point of view of the applied statistician, this sampling scheme is very unrealistic, because, under repeated sampling, one cannot keep the sufficient statistics fixed at a particular value. Fortunately, as is shown in sections 3 and 4, confidence intervals that are constructed from the asymptotic distribution under sampling scheme (c) are also valid under the other two sampling schemes. We start by giving a short description of CML estimation (section 1) and the asymptotic properties of the resulting estimators (section 2).

1 CML Estimation

The interest in CML estimation in psychometrics is due to the fact that many psychometric models contain so-called *incidental* or *nuisance parameters.* The typical data that are of interest here are the responses of a set of persons to a set of items. The models for these data usually contain person-specific parameters. These person parameters are an obstacle in proving the consistency of the item parameter esimates. The reason for this is that the number of person parameters increases with the sample size. This is why they are called incidental. Parameters whose number does not increase with the sample size are called *structural.* Obviously, item parameters are structural.

CML estimation is a solution for problems caused by incidental parameters because, instead of the usual likelihood, the conditional likelihood given sufficient statistics for the incidental parameters is maximized. We use the K-dimensional vector $\boldsymbol{\eta}$ to denote the structural parameters and the R-dimensional vector $\boldsymbol{\xi}$ to denote the incidental parameters. The data will be denoted by $\boldsymbol{X}$. In the following, we assume that the probability density function (PDF) of $\boldsymbol{X}$ belongs to the regular exponential family. This is not a restrictive assumption because CML estimation is only useful if there is a nontrivial sufficient statistic upon which to condition, and under mild regularity conditions, this is only the case if the model belongs to the exponential family (see Lehmann, 1983, pp. 44-45).

Now, this exponential family PDF can be written as follows:

$$f(\boldsymbol{X};\boldsymbol{\eta},\boldsymbol{\xi}) = a(\boldsymbol{\eta},\boldsymbol{\xi})b(\boldsymbol{X})\exp\left(\sum_{k=1}^{K}\mathcal{S}_k(\boldsymbol{X})\eta_k + \sum_{r=1}^{R}\mathcal{T}_r(\boldsymbol{X})\xi_r\right) \tag{1}$$

The a and the b in (1) are functions of, respectively, the parameters only and the data only. And the quantities $\mathcal{S}_k(\boldsymbol{X})$ and $\mathcal{T}_r(\boldsymbol{X})$ are the minimal sufficient statistics for the parameters η_k and ξ_r, respectively. As is shown in the notation, these quantities are functions of $\boldsymbol{X}$. The K-dimensional vector of $\mathcal{S}_k(\boldsymbol{X})$'s and the R-dimensional vector of $\mathcal{T}_r(\boldsymbol{X})$'s will be denoted by, respectively, $\boldsymbol{S}$ and $\boldsymbol{T}$. To prevent a possible misunderstanding, it has to be noted that K is not necessarily equal to the number of items (denoted by I) or some multiple of the number of items. Often, one of the item parameters is fixed to identify the model. For this reason, in the Rasch model K is usually equal to $(I-1)$, and in the Partial Credit model for M-category items K is usually equal to $(M-1)(I-1)$.

The next step is to define the conditional PDF of $\boldsymbol{X}$ given $\boldsymbol{T}$, the sufficient statistics for the incidental parameters. This conditional PDF is the following:

$$f(\boldsymbol{X} \mid \boldsymbol{T}; \boldsymbol{\eta}, \boldsymbol{\xi}) = \frac{f(\boldsymbol{X}; \boldsymbol{\eta}, \boldsymbol{\xi})}{f(\boldsymbol{T}; \boldsymbol{\eta}, \boldsymbol{\xi})} \tag{2}$$

In the following, we need $f(\boldsymbol{T}; \boldsymbol{\eta}, \boldsymbol{\xi})$, the PDF of $\boldsymbol{T}$. For simplicity, we assume the elements of $\boldsymbol{X}$ to be discrete. (If they are continuous, in the formula's below, one only has to replace the summation signs by integrals.) Now, $f(\boldsymbol{T}; \boldsymbol{\eta}, \boldsymbol{\xi})$ is defined as follows:

$$f(\boldsymbol{T}; \boldsymbol{\eta}, \boldsymbol{\xi}) = \sum_{\boldsymbol{X}:\boldsymbol{T}(\boldsymbol{X})=\boldsymbol{T}} f(\boldsymbol{X}; \boldsymbol{\eta}, \boldsymbol{\xi}) \tag{3}$$

The important point in (3) is the summation that runs over all data sets $\boldsymbol{X}$ that result in the vector $\boldsymbol{T}$ of sufficient statistics. Inserting (1) in (3), we get the following:

$$f(\boldsymbol{T}; \boldsymbol{\eta}, \boldsymbol{\xi}) = a(\boldsymbol{\eta}, \boldsymbol{\xi}) \exp\left(\sum_{r=1}^{R} T_r \xi_r\right) \sum_{\boldsymbol{X}:\boldsymbol{T}(\boldsymbol{X})=\boldsymbol{T}} b(\boldsymbol{X}) \exp\left(\sum_{k=1}^{K} \mathcal{S}_k(\boldsymbol{X})\eta_k\right) \tag{4}$$

This formula can be simplified by replacing the summation in the right-hand side by the function $c(\boldsymbol{T}, \boldsymbol{\eta})$.

Inserting (1) and (4) in (2) and cancelling some terms, we get the following expression for the conditional PDF of $\boldsymbol{X}$ given $\boldsymbol{T}$:

$$f(\boldsymbol{X} \mid \boldsymbol{T}; \boldsymbol{\eta}, \boldsymbol{\xi}) = \frac{b(\boldsymbol{X}) \exp\left(\sum_{k=1}^{K} \mathcal{S}_k(\boldsymbol{X})\eta_k\right)}{c(\boldsymbol{T}, \boldsymbol{\eta})} \tag{5}$$

The usefulness of this conditional PDF is due to the fact that it is independent of $\boldsymbol{\xi}$, the incidental parameters. Therefore, the $\boldsymbol{\xi}$ in the left-hand side of (5) can be omitted.

CML estimation is simply the maximization of the likelihood $l(\boldsymbol{\eta}; \boldsymbol{X}|\boldsymbol{T})$ that corresponds to the conditional PDF in (5). The maximization of this conditional likelihood is a fairly easy problem because the conditional PDF in (5) also belongs to the exponential family. The only complication that may be involved is that the conditional likelihood may have its maximum on the boundary of the parameter space. The conditions under which this may occur for the special case of the Rasch model are given by Fischer (1981).

2 Asymptotic Properties of CML Estimators

Deriving asymptotic properties of CML estimators in general is a difficult problem because of the dependence between the elements of $\boldsymbol{X}$ that is induced by the conditioning on $\boldsymbol{T}$. Andersen (1970, 1973) and Pfanzagl (1993, 1994) give results on the asymptotic properties of CML estimators in a particular class of models. However, this class is still so general that it includes all psychometric models that are of interest. This class of models is for data matrices $\boldsymbol{X}$ and incidental parameters $\boldsymbol{\xi}$ that can be divided into N parts X_n ($n = 1, \ldots, N$) and ξ_n, respectively, in such a way that the PDF of X_n only depends on ξ_n (and $\boldsymbol{\eta}$, of course). In psychometric applications, X_n usually is the response vector of the n-th subject and ξ_n is that subject's vector of person parameters.

The asymptotic properties of CML estimators have been studied under the three sampling schemes discussed in the previous. Sampling scheme (a) is sampling from $f(\boldsymbol{X}; \boldsymbol{\eta}, \boldsymbol{\xi})$, in which $\boldsymbol{\xi}$ is treated as fixed. Notice that $\boldsymbol{\xi}$ is of dimension N and that this dimension increases to infinity in the asymptotic argument. Andersen (1970, 1973) considered this sampling scheme for the case in which the ξ_n's belong to a compact space (for ξ_n's in real space, this means that they belong to a bounded subspace). For this case, he proved consistency and asymptotic normality under some regularity conditions on $f(\boldsymbol{X}|\boldsymbol{T}; \boldsymbol{\eta})$ and $f(\boldsymbol{X}; \boldsymbol{\eta})$ (Theorems 3 and 4 on p. 292 in Andersen (1970)). These regularity conditions are very similar to those of the usual proofs of consistency and asymptotic normality of ML estimators (see, e.g., Cramèr, 1946). Only his Assumption 2.1 (on p. 291 in the 1970 paper) is specific for the conditional nature of the estimation. This assumption involves that there is a positive probability of observing sufficient statistics $T_n = t_n$ (the sufficient statistic for ξ_n) for which the conditional PDF of X_n given $T_n = t_n$ depends on $\boldsymbol{\eta}$. For the Rasch model, this means that, for every person, there must be a positive probability for a sum score different from zero or the perfect score. If the ξ_n's are bounded, it is easy to see that this condition holds for the Rasch model.

Andersen's (1970) Theorem 4 also gives the asymptotic covariance matrix. In our notation, this covariance matrix is the inverse of a matrix $\boldsymbol{B}(\boldsymbol{\eta}, \boldsymbol{\xi})$, with elements $b_{kl}(\boldsymbol{\eta}, \boldsymbol{\xi})$ that are defined as follows:

$$b_{kl}(\boldsymbol{\eta}, \boldsymbol{\xi}) = \mathcal{E}_{\boldsymbol{\eta}, \boldsymbol{\xi}} \left[-\partial^2 \ln f(\boldsymbol{X} | \boldsymbol{T}; \boldsymbol{\eta}) / (\partial \eta_k)(\partial \eta_l) \right] \qquad , \tag{6}$$

in which the subscrips to $\mathcal{E}$ denote that the expectation is taken at some fixed values for $\boldsymbol{\eta}$ and $\boldsymbol{\xi}$. This matrix cannot be estimated consistently because it depends on $\boldsymbol{\xi}$.

Under sampling scheme (a), Pfanzagl (1994) gives a consistency proof for the special case of the Rasch model using a condition on the ξ_n's, different from compactness, that is not only sufficient but also necessary.

We now consider sampling scheme (b). This sampling scheme involves that the ξ_n's are considered as random variables having a common distribution. This distribution will be denoted by $g(\boldsymbol{\xi}; \boldsymbol{\lambda})$, in which $\boldsymbol{\lambda}$ denotes the parameters of this PDF. This sampling scheme is layered: first, $\boldsymbol{\xi}$ is drawn from $g(\boldsymbol{\xi}; \boldsymbol{\lambda})$, and then $\boldsymbol{X}$ is drawn from $f(\boldsymbol{X} | \boldsymbol{\xi}; \boldsymbol{\eta})$ (in which the conditioning bar denotes that $\boldsymbol{\xi}$ is now a random variable). The marginal PDF of $\boldsymbol{X}$ is then defined as follows:

$$f(\boldsymbol{X}; \boldsymbol{\eta}, \boldsymbol{\lambda}) = \int f(\boldsymbol{X} \mid \boldsymbol{\xi}; \boldsymbol{\eta}) g(\boldsymbol{\xi}; \boldsymbol{\lambda}) \, d\boldsymbol{\xi} \tag{7}$$

In his 1970 paper, Andersen notes that his proof of Theorems 3 and 4 (consistency and the asymptotic normality under sampling from $f(\boldsymbol{X}; \boldsymbol{\eta}, \boldsymbol{\xi})$) can be easily adapted for sampling from $f(\boldsymbol{X}; \boldsymbol{\eta}, \boldsymbol{\lambda})$ (see p. 292). In Andersen (1973), he actually gives this proof (see Theorem 3.2 on p. 88). Both the regularity conditions and the proof itself are parallel to those of Andersen's (1973) Theorem 3.1, which corresponds to Theorems 3 and 4 in his 1970 paper. The asymptotic covariance matrix under sampling from $f(\boldsymbol{X}; \boldsymbol{\eta}, \boldsymbol{\lambda})$ is the inverse of a matrix $\boldsymbol{B}(\boldsymbol{\eta}, \boldsymbol{\lambda})$, with elements $b_{kl}(\boldsymbol{\eta}, \boldsymbol{\lambda})$ that are defined as follows:

$$b_{kl}(\boldsymbol{\eta}, \boldsymbol{\lambda}) = \mathcal{E}_{\boldsymbol{\eta}, \boldsymbol{\lambda}} \left[-\partial^2 \ln f(\boldsymbol{X} | \boldsymbol{T}; \boldsymbol{\eta}) / (\partial \eta_k)(\partial \eta_l) \right] \qquad , \tag{8}$$

in which the subscrips to $\mathcal{E}$ denote that the expectation is taken given some fixed values for $\boldsymbol{\eta}$ and $\boldsymbol{\lambda}$. This matrix can be estimated consistently if we can also consistently estimate $\boldsymbol{\lambda}$. However, for this a separate estimation procedure is required. How this can be done for the special case of the Rasch model, is described by Andersen and Madsen (1977).

A different type of consistency proof for sampling scheme (b) is given by Pfanzagl (1993).

A line of research that is closely related to Pfanzagl's (1993) work, focusses on the question under which conditions CML and so-called *semiparametric* ML estimators are identical. Semiparametric ML estimators provide an estimate of both the structural parameters and the *distribution* of the nuisance parameters. Essential for *semi*parametric estimation is that this distribution is estimated non-parametrically. The consistency of semi-parametric estimators was shown by Kiefer and Wolfowitz (1956) under very weak regularity conditions. The relevance for the present paper lies in the fact that several authors (a) have presented the conditions under which CML and semiparametric ML estimates are identical, and (b) shown that these conditions are fulfilled with probability one as the sample size increase, in this way giving an indirect consistency proof for the CML estimators (De Leeuw & Verhelst, 1986; Follman, 1988; Lindsay, Clogg, & Grego, 1991).

Finally, we consider sampling scheme (c), which is sampling from the conditional PDF $f(\boldsymbol{X}|\boldsymbol{T};\boldsymbol{\eta})$. An important point here, is that consistency and asymptotic normality of CML estimators cannot be proved for *all* sequences of sufficient statistics $t_1, t_2, \ldots$ (t_n being a realization of T_n). The reason for this is that, for some realizations of T_n, the conditional PDF of X_n given T_n does not depend on $\boldsymbol{\eta}$. For the Rasch model, these are the response patterns with a zero or a perfect sum score. These observations provide no statistical information about $\boldsymbol{\eta}$. Therefore, consistency cannot be proved for *all* sequences, namely not for those with only a finite number of informative observations.

It is of interest to know what is the probability of the sequences $t_1, t_2, \ldots$ for which consistency and asymptotic normality cannot be proved. Now, it was shown by Andersen (1973) that, both under sampling scheme (a) (with the ξ_n's belonging to a compact space) and (b), the probability of such a sequence is zero (Theorems 4.1 and 4.2, p. 116, resp., p.117). Stated positively, Andersen proved that CML estimators are consistent and asymptotically normal under sampling from $f(\boldsymbol{X}|\boldsymbol{T};\boldsymbol{\eta})$ for all realizations of $\boldsymbol{T}$, except for a subset with probability zero.

The asymptotic covariance matrix under sampling from $f(\boldsymbol{X}|\boldsymbol{T};\boldsymbol{\eta})$ is the inverse of a matrix $\boldsymbol{B}(\boldsymbol{\eta},\boldsymbol{t})$, with elements $b_{kl}(\boldsymbol{\eta},\boldsymbol{t})$ that are defined as follows:

$$b_{kl}(\boldsymbol{\eta},\boldsymbol{t}) = \mathcal{E}_{\boldsymbol{\eta},\boldsymbol{t}}\left[-\partial^2 \ln f(\boldsymbol{X}|\boldsymbol{T}=\boldsymbol{t};\boldsymbol{\eta})/(\partial\eta_k)(\partial\eta_l)\right] \quad , \tag{9}$$

in which the subscrips to $\mathcal{E}$ denote that the expectation is taken given some fixed values for $\boldsymbol{\eta}$ and $\boldsymbol{t}$. This matrix can be estimated consistently by replacing $\boldsymbol{\eta}$ in (9) by its CML estimator. This is an important difference with (6), which cannot be estimated consistently, and (8), which can only be estimated consistently if a consistent estimator for $\boldsymbol{\lambda}$ is available.

We are in the unfortunate position that the only asymptotic confidence intervals we can construct from the CML estimates have a reference distribution that is not realistic for the applied statistician, namely $f(\boldsymbol{X}|\boldsymbol{T};\boldsymbol{\eta})$. Because he cannot keep $\boldsymbol{T}$ fixed at $\boldsymbol{t}$, its realization, repeated sampling from $f(\boldsymbol{X}|\boldsymbol{T};\boldsymbol{\eta})$ is not feasible in practice. However, in section 4, it is shown that this confidence interval is also valid under the other two reference distributions. But first, in section 3, we introduce this asymptotic confidence interval and give its interpretation under sampling from $f(\boldsymbol{X}|\boldsymbol{T};\boldsymbol{\eta})$.

3 The Asymptotic Confidence Interval under Sampling from $f(X \mid T; \eta)$

The $(1-\alpha/2) \times 100$ percent asymptotic confidence interval is the following:

$$\left(\hat{\eta}_k - U_{1-\alpha/2}\hat{\sigma}_k, \hat{\eta}_k + U_{1-\alpha/2}\hat{\sigma}_k\right) \tag{10}$$

In this expression, $U_{1-\alpha/2}$ is the $(1-\alpha/2) \times 100$-th percentile of the standardnormal distribution and $\hat{\sigma}_k$ is the estimated asymptotic sampling error of $\hat{\eta}_k$, obtained by taking the square root of the k-th diagonal element of the inverse of the expected information matrix, defined in (9), evaluated at the CML estimates. The interval in (10) will be denoted by $CI_\alpha(\hat{\eta}_k)$. Now, the interpretation of $CI_\alpha(\hat{\eta}_k)$ is given by the following probability statement:

$$\lim_{N\to\infty} P\left(\eta_k \in CI_\alpha(\hat{\eta}_k) \mid \boldsymbol{\mathcal{T}}(\boldsymbol{X}) = \boldsymbol{t}; \boldsymbol{\eta}\right) = 1 - \alpha \tag{11}$$

In this equation, the limit is taken for the number of persons (N) going to infinity. This is specified by a vector of sufficient statistics $\boldsymbol{t} = (t_1, \ldots, t_N)$ of increasing dimensionality. Successive $\boldsymbol{t}$'s correspond to data sets of increasing size. The probability statement in (11) is a direct consequence of Andersen's (1973) Theorems 4.1 and 4.2.

An alternative way of expressing (11) is possible by making use of the indicator function $I_\alpha(\boldsymbol{X}, k)$ that has the value 1 if η_k belongs to $CI_\alpha(\hat{\eta}_k)$ for the data set $\boldsymbol{X}$ and 0 if it does not:

$$\lim_{N\to\infty} \sum_{\boldsymbol{X}:\boldsymbol{\mathcal{T}}(\boldsymbol{X})=\boldsymbol{t}} I_\alpha(\boldsymbol{X}, k) f(\boldsymbol{X} \mid \boldsymbol{T} = \boldsymbol{t}; \boldsymbol{\eta}) = 1 - \alpha \tag{12}$$

4 The Asymptotic Confidence Interval under Sampling from $f(X;\eta,\xi)$ and $f(X;\eta,\lambda)$

For the following, it is convenient to have expressions for $f(\boldsymbol{X};\boldsymbol{\eta},\boldsymbol{\xi})$ and $f(\boldsymbol{X};\boldsymbol{\eta},\boldsymbol{\lambda})$ that show their relation with $f(\boldsymbol{X}|\boldsymbol{T};\boldsymbol{\eta})$. By making use of the fact that $f(\boldsymbol{X}|\boldsymbol{T},\boldsymbol{\eta},\boldsymbol{\xi})$ does not depend on $\boldsymbol{\xi}$, we can write the following:

$$f(\boldsymbol{X};\boldsymbol{\eta},\boldsymbol{\xi}) = f(\boldsymbol{X} \mid \boldsymbol{T};\boldsymbol{\eta})f(\boldsymbol{T};\boldsymbol{\eta},\boldsymbol{\xi}) \qquad , \tag{13}$$

and,

$$\begin{aligned} f(\boldsymbol{X};\boldsymbol{\eta},\boldsymbol{\lambda}) &= f(\boldsymbol{X} \mid \boldsymbol{T};\boldsymbol{\eta}) \int f(\boldsymbol{T};\boldsymbol{\eta},\boldsymbol{\xi})g(\boldsymbol{\xi};\boldsymbol{\lambda})\, d\boldsymbol{\xi} \\ &= f(\boldsymbol{X} \mid \boldsymbol{T};\boldsymbol{\eta})f(\boldsymbol{T};\boldsymbol{\eta},\boldsymbol{\lambda}), \end{aligned} \tag{14}$$

in which $f(\boldsymbol{T};\boldsymbol{\eta},\boldsymbol{\lambda})$ denotes the marginal PDF of $\boldsymbol{T}$.

We now consider $CI_\alpha(\hat{\eta}_k)$, as defined in (11), under sampling from $f(\boldsymbol{X};\boldsymbol{\eta},\boldsymbol{\lambda})$. The exposition for sampling from $f(\boldsymbol{X};\boldsymbol{\eta},\boldsymbol{\xi})$ is completely analogous: in the formula's, one only has to replace $\boldsymbol{\lambda}$ by $\boldsymbol{\xi}$. Now, using the indicator function $I_\alpha(\boldsymbol{X},k)$, the asymptotic probability of η_k belonging to $CI_\alpha(\hat{\eta}_k)$ under sampling from $f(\boldsymbol{X},\boldsymbol{\eta},\boldsymbol{\lambda})$ can be written as follows:

$$\lim_{N\to\infty}\sum_{\boldsymbol{X}} I_\alpha(\mathbf{X},k)f(\boldsymbol{X};\boldsymbol{\eta},\boldsymbol{\lambda}) \tag{15}$$

Inserting (14) in (15), it follows that the asymptotic probability of η_k belonging to $CI_\alpha(\hat{\eta}_k)$ under sampling from $f(\boldsymbol{X};\boldsymbol{\eta},\boldsymbol{\lambda})$ can be written as follows:

$$\lim_{N\to\infty}\sum_{\boldsymbol{T}}\left(\sum_{\boldsymbol{X}\,:\,\boldsymbol{T}(\boldsymbol{X})=\boldsymbol{T}} I_\alpha(\boldsymbol{X},k)f(\boldsymbol{X} \mid \boldsymbol{T};\boldsymbol{\eta})\right) f(\boldsymbol{T};\boldsymbol{\eta},\boldsymbol{\lambda}) \tag{16}$$

The important point is that the expression between large parentheses in (16) converges to $(1-\alpha)$ (see (12), except for those terms that correspond to sequences $t_1, t_2, \ldots$ that do not allow for consistent asymptotically normal CML estimates. However, this set of sequences has probability zero under sampling from $f(\boldsymbol{X};\boldsymbol{\eta},\boldsymbol{\lambda})$ (Andersen's (1973) Theorem 4.2). Because the expression between large parentheses in (16) is always between zero and one, we can safely ignore this set of probability zero. It then follows that the complete expression in (16) converges to a weighted sum of terms all of which are equal to $(1-\alpha)$. It follows that the asymptotic probability of η_k belonging to $CI_\alpha(\hat{\eta}_k)$ under sampling from $f(\boldsymbol{X};\boldsymbol{\eta},\boldsymbol{\lambda})$ equals $(1-\alpha)$. Formally,

$$\lim_{N\to\infty}\sum_{\boldsymbol{X}} I_\alpha(\mathrm{X},k)f(\boldsymbol{X};\boldsymbol{\eta},\boldsymbol{\lambda}) = 1-\alpha \tag{17}$$

5 Extension to Hypothesis Tests

It is straightforward to generalize the result in (17) to the sampling interpretation of hypothesis tests in the context of CML estimation. In particular, one can show that the asymptotic type-I error α not only holds under sampling from $f(\boldsymbol{X} \mid \boldsymbol{T};\boldsymbol{\eta})$ but also under sampling from $f(\boldsymbol{X};\boldsymbol{\eta},\boldsymbol{\xi})$ and $f(\boldsymbol{X};\boldsymbol{\eta},\boldsymbol{\lambda})$. One only has to replace the indicator function $I_\alpha(\boldsymbol{X},k)$ by $I_\alpha(\boldsymbol{X})$ and define it as an indicator for the event that the null hypothesis is *not* rejected using significance level α.

6 Conclusion and Extension

The results in this paper are of interest because of their usefulness for the applied statistician. The results show that one does not have to consider an unrealistic conditional sampling scheme in the interpretation of asymptotic confidence intervals and hypothesis tests. The interpretation that holds under sampling from $f(\boldsymbol{X} \mid \boldsymbol{T};\boldsymbol{\eta})$ also holds under sampling from $f(\boldsymbol{X};\boldsymbol{\eta},\boldsymbol{\xi})$ and $f(\boldsymbol{X};\boldsymbol{\eta},\boldsymbol{\lambda})$.

The argument that leads to these results is not specific for inference conditional on sufficient statistics. For instance, the same argument can be applied to inference in the linear model, where the conditioning is on the independent variables. In observational studies, these independent variables are random, and therefore one cannot expect their values to be constant over replications. However, all inferential statements, such as those based on confidence intervals for regression parameters and F-statistics, assume repeated sampling from the conditional distribution of the dependent variable given some fixed values for the independent variables. Essentially the same argument as the one in this paper, allows us to conclude that these inferential statements are also valid under sampling from the marginal distribution of the dependent variable (i.e., with the independent variables integrated out). Moreover, in case of the linear model, we don't have to deal with tricky asymptotics, because we know the exact sampling distribution of the regression coefficients and the F-statistics for finite samples. Therefore, we only need a matrix of independent variables that is of full column rank (or something equivalent, like estimable functions).

References

[1] Andersen, E. B. (1970). Asymptotic properties of conditional maximum likelihood estimators. *Journal of the Royal Statistical Society, B, 32,* 283-301.

[2] Andersen, E. B. (1973). *Conditional inference and models for measuring.* (unpublished Ph. D. Thesis). Copenhagen: Mentalhygiejnisk Forlag.

[3] Andersen, E. B., & Madsen, M. (1977). Estimating the parameters of the latent population distribution. *Psychometrika, 42*, 357-374.

[4] Cramèr, H. (1946). *Mathematical methods of statistics.* Princeton: Princeton University Press.

[5] De Leeuw, J., & Verhelst, N. (1986). Maximum likelihood estimation in generalized Rasch models, *Journal of Educational Statistics, 11,* 183-196.

[6] Fischer, G. H. (1981). On the existence and uniqueness of maximum likelihood estimates in the Rasch model. *Psychometrika, 53*, 525-546.

[7] Follman, D. A. (1988). Consistent estimation in the Rasch model based on nonparametric margins. *Psychometrika, 53,* 553-562.

[8] Lehmann, E. L. (1983). *Theory of point estimation.* New York: Wiley.

[9] Lindsay, B., Clogg, C. C., & Grego, J. (1991). Semiparametric estimation in the Rasch model and related exponential response models, including a simple latent class model for item analysis. *Journal of the American Statistical Association, 86,* 96-107.

[10] Pfanzagl, J. (1993). A case of asymptotic equivalence between conditional and marginal maximum likelihood estimators. *Journal of Statistical Planning and Inference, 35,* 301-307.

[11] Pfanzagl, J. (1994). On item parameter estimation in certain latent trait models. In G. H. Fischer & D. Laming (Eds.). *Contributions to mathematical psychology, psychometrics and methodology* (pp. 249-263). New York: Springer.

Recent Measurement and Research Department Reports:

96-1 H.H.F.M. Verstralen. Estimating Integer Parameters in IRT Models for Polytomous Items.

96-2 H.H.F.M. Verstralen. Evaluating Ability in a Korfball Game.

96-3 T.J.H.M. Eggen & G.J.J.M. Straetmans. Computerized Adaptive Testing for Classifying Examinees Into Three Categories.

96-4 P.F. Sanders & A.J. Verschoor. Parallel Test Construction Using Classical Item Parameters.