Estimation of Item Parameters

Ivo W. Molenaar

ABSTRACT The introduction of this chapter sketches the problem of the estimation of item parameters and the notation in the case of incomplete data. Then the joint, conditional, and marginal maximum likelihood methods are discussed. A final section briefly mentions a few other methods not based on likelihoods.

3.1 Introduction

There are several ways to obtain estimates of item parameters from a data matrix containing the dichotomously scored answers of a sample of \( n \) persons to \( k \) items. In the early publications on parameter estimation, it was assumed that this matrix is completely observed. It is a major advantage of IRT, however, that its models and methods can easily be generalized to situations where different subsets of the items are presented to certain subgroups of respondents. In many later publications, and also in this chapter, this case of data 'missing by design' is included. If data are missing for other reasons, our derivations are only valid in the case of 'ignorable non-response' (see, e.g., Rubin, 1987), which roughly means that the fact that a certain item was not responded to is not related to the parameters of person and item. Mislevy and Wu (1988) argue convincingly that this will seldom be the case when a subject decides to skip an achievement item after seeing its content. They also discuss models proposed by Lord (1974, 1983b) for achievement items not reached in a time limit test. Also in the case of attitude measurement, ignorability is often not plausible and should be investigated before the methods outlined in this chapter are applied to incomplete data where the incompleteness design was not planned in advance. In order to stay on the safe side, the present text speaks only about items not presented to a person.

If subject \( S_v \) was not asked to respond to item \( I_i \), one puts the element \( b_{vi} \) of the design matrix \( B \) equal to 0 and the corresponding element \( x_{vi} \) of the data matrix \( X \) equal to some arbitrary value \( a \) (0 < \( a < 1 \)). Thus,

\[
x_{vi} = \begin{cases} 
1 & \text{if } S_v \text{ responded positively to } I_i, \\
0 & \text{if } S_v \text{ responded negatively to } I_i, \\
 a & \text{if } I_i \text{ was not presented to } S_v,
\end{cases}
\]
and

\[ b_{vi} = \begin{cases} 
1 & \text{if } I_i \text{ was presented to } S_v, \\
0 & \text{if } I_i \text{ was not presented to } S_v.
\end{cases} \]

In the case of complete data, \( b_{vi} = 1 \) for all pairs \((v,i)\) and the formulas can be simplified. It will be assumed that the matrix \( B \) is known before the answers are obtained; see above for the more general case of ignorability, and see Glas (1988b) for a discussion of multistage testing in which a part of \( B \) depends on the answers to the routing test(s).

Let \( \beta \) denote the \( k \)-vector of item parameters, and \( \theta \) the \( n \)-vector of person parameters. For each cell of the data matrix \( X \) one has, for \( x_{vi} \in \{0, a, 1\} \) and corresponding \( b_{vi} \in \{0, 1\}, \)

\[ P(X_{vi} = x_{vi} | \theta_v, \beta_i) = \frac{\exp[b_{vi} x_{vi} (\theta_v - \beta_i)]}{[1 + \exp(\theta_v - \beta_i)]^{b_{vi}}} \]  

(3.1)

In the unobserved cells one has by definition \( x_{vi} = a \). By local independence and independence between persons, it follows that, for any choice of zero or one in the observed cells,

\[ P(X = x | \theta, \beta) = \prod_{v=1}^{n} \prod_{i=1}^{k} \frac{\exp[b_{vi} x_{vi} (\theta_v - \beta_i)]}{[1 + \exp(\theta_v - \beta_i)]^{b_{vi}}} \]  

(3.2)

If all item parameters \( \beta_i \) and all person parameters \( \theta_v \) were known, this would define a probability distribution on all \( n \times k \) matrices with entries from the set \( \{0,1\} \) for all real answers, and entries \( a \) for all missing observations.

If a test with known item parameters \( \beta_i \) is given to a fresh sample of persons, one can estimate their person parameters \( \theta_v \). This is the topic of Chapter 4. Here we deal with the situation in which both \( \theta_v \) and \( \beta_i \) are unknown and have to be estimated. The prime interest in the present chapter is on estimating the item parameters \( \beta_i \); the person parameters \( \theta_v \) act as nuisance parameters.

In this situation, one restriction must be imposed in order to obtain a unique solution. In Chapter 2 the form (2.22) was derived; above it was already assumed that the scale constant \( a \) in that formula equals 1, but one must also fix the origin of the scale, which amounts to fixing \( b \) in (2.22). It is usual to impose the restriction \( \sum_i \beta_i = 0 \); this convention is followed in the present chapter unless the contrary is stated. One could also put the population mean of the person parameters equal to 0, or select a fixed item \( I_i \) for which the item parameter equals 0. Note that the latter leads to different standard errors of the parameter estimates because they are correlated across items (Verhelst, 1993).

Two basic distinctions between estimation methods underlie the structure of this chapter. The first is whether the person parameters are

- jointly estimated with the item parameters, or
- eliminated by conditioning, or