Chapter 5

Asymptotic Properties of Item Parameter Estimates Using MMLE

A large area of research in the field of Item Response Theory concerns the estimation of item parameters and examinee abilities from a data set for different unidimensional parametric IRF models. Many of the methods used to estimate these item and examinee parameters rely on the maximization of a specific likelihood function. The asymptotic theory of maximum likelihood estimates is well developed in the statistical literature when the number of parameters to be estimated is fixed as the sample size tends to infinity. However, to determine the asymptotic properties of both item parameter and examinee ability estimates, both the number of items and the number of examinees must tend to infinity. Thus, standard theorems on the consistency and asymptotic normality of maximum likelihood estimates do not apply (Neyman & Scott, 1948) to the estimation of both item parameters and examinee abilities for unidimensional parametric IRF models.

The purpose of this chapter is to prove the consistency and asymptotic normality of the item parameter estimates obtained from the Marginal Maximum Likelihood Estimation (MMLE) procedure (Bock & Lieberman, 1970) for the 1PL, 2PL and 3PL models as both the number of items and the number of examinees tends to infinity. The theorems for consistency and asymptotic normality of the item parameter estimates depend upon some general conditions on the model and on the rate of growth of the number of items relative
to the number of examinees. One should be aware the results in this chapter do not extend
to the item parameter estimates obtained from applying the EM algorithm to the MMLE
procedure as was described in Bock & Aitkin, (1981), and implemented in the program
BILOG (Mislevy & Bock, 1990).

5.1 Review of the MMLE Procedure

The Marginal Maximum Likelihood Estimation (MMLE) procedure was developed by Bock
& Lieberman (1970) to estimate item parameters and examinee abilities from an examinee
response data set. Let $P_i(\theta)$ denote the IRF for item $i$ from a particular parametric IRF
model and let $n$ denote the number of test items. Then, the probability a randomly selected
examinee with ability $\theta$ obtains a particular response pattern $u$ is defined as

$$P(u|\theta) = \prod_{i=1}^{n} (P_i(\theta))^{u_i} (1 - P_i(\theta))^{1-u_i}. \quad (5.1)$$

If one assumes the examinees are a random sample from a population with ability density
$g(\theta)$, the unconditional or marginal probability of a particular response pattern $u$ is given
by

$$P(u) = \int P(u|\theta)g(\theta)d\theta \quad (5.2)$$

Since the examinee ability $\theta$ has been integrated out in equation (5.2), $P(u)$ is a function of
only the item parameters of the test’s IRFs.

If $J$ examinees are randomly sampled from a population with ability density $g(\theta)$, the
likelihood of the marginal probabilities $P(u)$ over all $J$ observed response patterns is defined
as

$$\prod_{j=1}^{J} P(u_j). \quad (5.3)$$
Taking the log of the likelihood function gives

\[ L_U = \sum_{j=1}^{J} \log P(u_j) \]  

(5.4)

The MMLE procedure obtains item parameter estimates by maximizing the marginal log likelihood function from equation (5.4) with respect to the item parameters. Bock & Lieberman (1970) used the Newton-Raphson method to find a maximum likelihood estimate of the item parameters.

Using the item parameter estimates in place of the true parameter values, ability estimates for each examinee are then obtained using one of several different methods. The traditional method of estimating examinee abilities is to maximize the log likelihood function,

\[ L_C = \ln P(u|\theta) \]  

(5.5)

with respect to \( \theta \) for each observed response pattern \( u \). Another method finds an estimate of \( \theta \) based on the posterior distribution of \( \theta \) given the response pattern \( u \), defined as

\[ P(\theta|u) = \frac{P(u|\theta)g(\theta)}{\int P(u|\theta)g(\theta)d\theta}. \]  

(5.6)

Generally, the ability estimate for an examinee with response pattern \( u \) is determined by either finding the mean or mode of the posterior distribution of \( \theta \) given the response pattern \( u \).

### 5.2 Asymptotic Results for Item Parameter Estimates

The asymptotic properties of the item parameter estimates obtained from the MMLE procedure are presented below. The conditions (C0) through (C5) on the model and Theorems 5.1, 5.2, and 5.3 on the consistency and asymptotic normality of the item parameters esti-
mates for the 1PL, 2PL and 3PL models are directly adapted from the theorems in He & Shao (2000).

Using the notation in He & Shao (2000), let \( \tau \) denote a particular item parameter from the model and let \( \bm{\tau} \) denote the vector of item parameters from the model. Thus, \( \bm{\tau} = (b_1, b_2, \ldots, b_n) \) for the 1 PL model, \( \bm{\tau} = (a_1, b_1, a_2, b_2, \ldots, a_n, b_n) \) for the 2PL model, and \( \bm{\tau} = (a_1, b_1, c_1, a_2, b_2, c_2, \ldots, a_n, b_n, c_n) \) for the 3PL model. Let \( m \) denote the length of the vector \( \bm{\tau} \). Thus, \( m \) is the total number of item parameters in the model and equals \( n, 2n \), or \( 3n \) depending on the model used for the IRFs. Finally, denote the number of examinees as \( J \).

Let \( \rho(u_j, \bm{\tau}) = \log P(u_j) \) and denote the derivative of \( \rho \) with respect to the vector \( \bm{\tau} \) as the vector \( \psi(u_j; \bm{\tau}) \). Thus, the \( i \)th component of the vector \( \psi(u_j; \bm{\tau}) \) is

\[
\psi_i(u_j, \bm{\tau}) = \frac{\partial \log P(u_j)}{\partial \tau_i}
\]

for \( i = 1, \ldots, m \). Let \( \hat{\bm{\tau}}_J \) be a maximizer of the log-likelihood function, \( \sum_{j=1}^{J} \log P(u_j) = \sum_{j=1}^{J} \rho(u_j, \bm{\tau}) \) and let \( \bm{\tau}_0 \) be the vector of true item parameters. Define

\[
\eta_j(\bm{\tau}', \bm{\tau}) = \psi(u_j, \bm{\tau}') - \psi(u_j, \bm{\tau}) - E[\psi(u_j, \bm{\tau}') - \psi(u_j, \bm{\tau})]
\]

Finally, define the set \( S_m = \{ \bm{a} : \| \bm{a} \| = 1 \} \), where \( \| \bm{a} \| \) is the \( L_2 \) norm of the vector \( \bm{a} \). The following conditions on the model are required for Theorems 5.1, 5.2 and 5.3.

(C0) \( \| \sum_{j=1}^{J} \psi(u_j, \hat{\bm{\tau}}_J) \| = o_P(J^{1/2}) \)

(C1) There exists a constant \( C \) such that for all \( 0 < d \leq 1 \)

\[
\max_{\bm{\tau}} E_{\bm{\tau}} \sup_{\bm{\tau}' : \| \bm{\tau}' - \bm{\tau} \| \leq d} \| \eta_j(\bm{\tau}', \bm{\tau}) \|^2 \leq J^C d^2
\]
(C2) For any $\alpha \in S_m$ and any constant $B > 0$,

$$\sup_{\tau' : \|\tau' - \tau\| \leq B(m/J)^{1/2}} \sum_{j=1}^{J} E_{\tau} \left( \alpha^T \eta_j(\tau', \tau) \right)^2 = O(n^3)$$

(C3) $\sup_{\alpha \in S_m, \tau : \|\tau' - \tau\| \leq B(m/J)^{1/2}} \sum_{j=1}^{J} \left( \alpha^T \eta_j(\tau', \tau) \right)^2 = O_P(n^3)$

(C4) $\| \sum_{j=1}^{J} \psi(\mathbf{u}_j, \tau_o) \|^2 = O_P(nJ)$

(C5) Define the minimum eigenvalue of a matrix $F$ as $\lambda_{\min}(F)$. There exists a sequence of $m$ by $m$ matrices $D_j$ with $\sup_j \lambda_{\min}(D_j) < 0$ such that for any constant $B > 0$ and uniformly in $\alpha \in S_m$,

$$\sup_{\|\tau - \tau_o\| \leq B(m/J)^{1/2}} \left( \alpha^T \sum_{j=1}^{J} E_{\tau_o} \{ \psi(\mathbf{u}_j; \tau) - \psi(\mathbf{u}_j; \tau_o) \} - J \alpha^T D_j (\tau - \tau_o) \right)^2 = o(J)$$

The consistency result for the maximum likelihood estimate $\hat{\tau}_J$ of the log likelihood function in Theorem 5.1 below applies to the item parameter estimates for the 1PL model obtained from the MMLE procedure.

**Theorem 5.1 (He & Shao, 2000)** If $\hat{\tau}_J$ is a maximizer of $\sum_{j=1}^{J} \rho(\mathbf{u}_j; \tau)$ and $\rho$ is concave in $\tau$, then under Assumptions (C0) through (C5) when $n^3 = o(J/ \log J)$,

$$\| \hat{\tau}_J - \tau_o \| = O_P(n/J)$$

For the 2PL and 3PL models, since the $\rho$ function is not concave in $\tau$, the maximum likelihood estimate $\hat{\tau}_J$ of the log-likelihood function for the 2PL and 3PL models will not be unique for each $J$. As a result, the sequence of observed estimators $\hat{\tau}_J$ may not be consistent. However, there does exist some subsequence of estimators $\hat{\tau}_{J'}$ that is consistent. This result is stated in Theorem 5.2.
Theorem 5.2 If $\hat{\tau}_J$ is a maximizer of $\sum_{j=1}^{J} \rho(u_j, \tau)$, then under Assumptions (C0) through (C5) when $n^3 = o(J/\log J)$, there exists a subsequence of estimators $\hat{\tau}_{J'}$ of the log likelihood function such that

$$\|\hat{\tau}_{J'} - \tau_o\| = O_P(n/J)$$

For any consistent estimator $\hat{\tau}_J$, Theorem 5.3 below gives the asymptotic normality for the distribution of $\alpha^T(\hat{\tau}_J - \tau_o)$ for any $\alpha \in S_m$.

Theorem 5.3 (He & Shao, 2000) Denote the covariance matrix of the vector $\psi(u_j, \tau_o)$ as $\Delta$. Under Assumptions (C0) through (C5) with $n^4 = o(J/\log J)$, for any consistent estimator $\hat{\tau}_J$,

$$\frac{\sqrt{m}\alpha^T(\hat{\tau}_J - \tau_o)}{\alpha^TD_j^{-1}\Delta D_j^{-T}\alpha} \rightarrow N(0, 1)$$

5.3 Proof of Theorems 5.1 - 5.3

In order to prove the results presented in Section 5.2 hold for the item parameter estimates obtained from the MMLE procedure, the following regularity conditions are needed for the 2PL model.

**Regularity Condition 1** There exists a positive constant $M_1$ such that $a_i \leq M_1$ for all $i$.

**Regularity Condition 2** There exists a positive constant $M_2$ such that $|b_i| \leq M_2$ for all $i$.

**Regularity Condition 3** The first absolute moment of the posterior distribution of $\theta$ is bounded for all observed response patterns $u_j$.

**Regularity Condition 4** The second moment of the posterior distribution of $\theta$ is bounded for all observed response patterns $u_j$.

In addition to conditions 1 through 4, two additional regularity conditions are needed for the 3PL model.
Regularity Condition 5 There exists a positive constant $M_3$ such that $c_i \leq 1 - M_3$ for all $i$.

Regularity Condition 6 There exists a positive constant $\epsilon$ such that $P_i(\theta) > \epsilon$ for all $\theta$. This regularity condition requires that either $c_i$ is bounded away from 0 for all $i$, or that the range of $\Theta$ is finite.

Using the regularity conditions, Results 5.1 through 5.4 below hold for the 1PL, 2PL, and 3PL models.

Result 5.1 The partial derivatives of $\log P(u_j)$ with respect to each item parameter $\tau$ are bounded for all observed response patterns $u_j$.

Result 5.2 The partial derivatives of $\log P(u_j)$ with respect to each pair of item parameters $\tau$ and $\tau'$ are bounded for all observed response patterns $u_j$.

Result 5.3 The expected value of the partial derivatives of $\log P(u_r)$ with respect to each combination of item parameters $\tau$, $\tau'$ and $\tau''$ are bounded for all observed response patterns $u_j$.

Result 5.4 Let the matrix $D_J$ be defined by

$$D_J(i, l) = \frac{1}{J} \sum_{j=1}^{J} E_{\tau} \frac{\partial \log P(u_j)}{\partial \tau_i \partial \tau_l}$$  \hspace{1cm} (5.8)

Then, $\sup_J \lambda_{\min}(D_J) < 0$.

Proofs of Results 5.1 through 5.4 are given in Section 5.4. Using Results 5.1 through 5.4, the proofs of conditions (C0) through (C5) are given below. Throughout the proofs, let $A$ denote some positive constant. The actual value of the constant $A$ might change from line to line in the proofs.
**Proof of (C0)** Since $\rho$ is differentiable with respect to $\tau$ for all $\tau$, the vector of first partial derivatives of the log-likelihood function evaluated at a maximizer $\hat{\tau}_J$ is $0$. Hence, $
abla J \psi(u_j, \hat{\tau}_J) = 0$ and $\|\nabla J \psi(u_j, \hat{\tau}_J)\| = 0$.

**Proof of (C1)**

$$\|\psi(u_j, \tau') - \psi(u_j, \tau)\|^2 = \sum_{i=1}^{m} (\psi_i(u_j, \tau') - \psi_i(u_j, \tau))^2.$$  \hspace{1cm} (5.9)

Let the vector $\psi_i^{(1)}(u_j, \tau)$ be the gradient of $\psi_i(u_j, \tau)$ with respect to the vector $\tau$. Thus, the $l$th component of the vector $\psi_i^{(1)}(u_j, \tau)$ is

$$\psi_{i,l}^{(1)}(u_j, \tau) = \frac{\partial^2 \log P(u_j)}{\partial \tau_l \partial \tau_l}$$

Using the Mean Value Theorem, $\psi_i(u_j, \tau') - \psi_i(u_j, \tau) = (\psi_i^{(1)}(u_j, \tau^*)^T (\tau' - \tau)$ where $\tau^*$ is a point in the interior of the line segment between $\tau'$ and $\tau$. Substituting this into equation 5.9 gives

$$\sum_{i=1}^{m} (\psi_i(u_j, \tau') - \psi_i(u_j, \tau))^2 = \sum_{i=1}^{m} \left[ (\psi_i^{(1)}(u_j, \tau^*))^T (\tau' - \tau) \right]^2$$

$$= \sum_{i=1}^{m} \left( \sum_{l=1}^{m} \psi_{i,l}^{(1)}(u_j, \tau^*) (\tau'_l - \tau_l) \right)^2$$

$$\leq \sum_{i=1}^{m} \left( \sum_{l=1}^{m} (\psi_{i,l}^{(1)}(u_j, \tau^*))^2 \right) \left( \sum_{l=1}^{m} (\tau'_l - \tau_l)^2 \right)$$

By Result 5.2, we have

$$\sum_{i=1}^{m} \left( \sum_{l=1}^{m} (\psi_{i,l}^{(1)}(u_j, \tau^*))^2 \right) \left( \sum_{l=1}^{m} (\tau'_l - \tau_l)^2 \right) \leq Am \sum_{i=1}^{m} \left( \sum_{l=1}^{m} (\tau'_l - \tau_l)^2 \right)$$

$$= Am^2 \|\tau' - \tau\|^2$$

$$= An^2 \|\tau' - \tau\|^2$$
Thus,

$$\max_j E_\mathbf{r} \sup_{\mathbf{r}' : \|\mathbf{r}' - \mathbf{r}\| \leq d} \|\eta_j(\mathbf{r}', \mathbf{r})\|^2 \leq An^2 d^2$$

Therefore, (C1) holds if $n^2 \leq J^C$ for some constant $C$. Clearly, in practice, condition (C1) will hold for $C \geq 2$.

Proof of (C2)

$$(\mathbf{a}^T[\psi(u_j, \mathbf{r}') - \psi(u_j, \mathbf{r})])^2 = \left(\sum_{i=1}^{m} \alpha_i[\psi_i(u_j, \mathbf{r}') - \psi_i(u_j, \mathbf{r})]\right)^2$$

$$\leq \left(\sum_{i=1}^{m} \alpha_i^2\right) \left(\sum_{i=1}^{m} (\psi_i(u_j, \mathbf{r}') - \psi_i(u_j, \mathbf{r}))^2\right)$$

$$= \|\mathbf{a}\|^2 \sum_{i=1}^{m} (\psi_i(u_j, \mathbf{r}') - \psi_i(u_j, \mathbf{r}))^2$$

$$= \sum_{i=1}^{m} (\psi_i(u_j, \mathbf{r}') - \psi_i(u_j, \mathbf{r}))^2$$

By the same steps used in proving (C1),

$$\sum_{i=1}^{m} (\psi_i(u_j, \mathbf{r}') - \psi_i(u_j, \mathbf{r}))^2 \leq An^2 \|\mathbf{r}' - \mathbf{r}\|^2$$

Thus,

$$\sup_{\mathbf{r}' : \|\mathbf{r}' - \mathbf{r}\| \leq B(m/J)^{\ell^2}} \sum_{j=1}^{J} E_\mathbf{r} (\mathbf{a}^T \eta_j(\mathbf{r}', \mathbf{r}))^2 \leq An^2 J (m/J)$$

$$= An^3$$

Therefore, (C2) holds.

Proof of (C3) Follows directly from the proof of (C2).
Proof of (C4)

\[
\| \sum_{j=1}^{J} \psi(u_j, \tau_o) \|^2 = \sum_{i=1}^{m} \left( \sum_{j=1}^{J} \psi(u_j, \tau_o) \right)^2 \\
= \sum_{i=1}^{m} \left( \sum_{j=1}^{J} \psi(u_j, \tau_o) \right)^2 + 2 \sum_{j=1}^{J} \sum_{j'=j+1}^{J} \psi(u_j', \tau_o) \psi(u_j, \tau_o) 
\]

Since responses from different examinees are independent, \( \psi(u_j, \tau_o) \) and \( \psi(u_{j'}, \tau_o) \) are independent for all \( j \neq j' \). Since \( E[\psi(u_j, \tau_o)] = 0 \) for all \( j \), the term

\[
2 \sum_{j=1}^{J} \sum_{j'=j+1}^{J} \psi(u_j', \tau_o) \psi(u_j, \tau_o) = O_P(1)
\]

By Result 5.1, since \( \psi(u_j, \tau_o)^2 \) is bounded for all \( j \), the term \( \sum_{j=1}^{J} \psi(u_j, \tau_o)^2 \) is \( O_P(J) \).

Thus,

\[
\sum_{i=1}^{m} \left( \sum_{j=1}^{J} \psi(u_j, \tau_o) \right)^2 = \sum_{i=1}^{m} O_P(J) = O_P(mJ) = O_P(nJ)
\]

Thus, condition (C4) holds.

Proof of (C5)

\[
\alpha^T \left( \sum_{j=1}^{J} E_{\tau_o} \{ \psi(u_j, \tau) - \psi(u_j, \tau_o) \} \right) = \sum_{i=1}^{m} \alpha_i \left( \sum_{j=1}^{J} E_{\tau_o} \{ \psi_i(u_j, \tau) - \psi_i(u_j, \tau_o) \} \right)
\]

Let the matrix \( \psi_i^{(2)}(u_j, \tau) \) denote the second derivatives of \( \psi_i(u_j, \tau^*) \) with respect to the vector \( \tau \). Using Taylor’s Theorem, we have

\[
\psi_i(u_j, \tau) - \psi_i(u_j, \tau_o) = (\psi_i^{(1)}(u_j, \tau_o))^T (\tau - \tau_o) + \frac{1}{2} (\tau - \tau_o)^T \psi_i^{(2)}(u_j, \tau^*) (\tau - \tau_o)
\]

Substituting into equation (5.11) gives
\[
\begin{align*}
\sum_{i=1}^{m} \alpha_i \left[ \left( \sum_{j=1}^{J} E_{\tau_o} \{ \psi_i^{(1)}(u_j, \tau_o) \} \right)^T (\tau - \tau_o) \\
+ \frac{1}{2} (\tau - \tau_o)^T \left( \sum_{j=1}^{J} E_{\tau_o} \{ \psi_i^{(2)}(u_j, \tau^*) \} \right) (\tau - \tau_o) \right]
\end{align*}
\]

Define the matrix \( D_J \) by equation \((5.8)\). By Result 5.4, \( \sup_J \lambda_{m,m}(D_J) < 0 \). The quantity \( J\alpha^T D_J (\tau - \tau_o) \) is equivalent to

\[
J\alpha^T D_J (\tau - \tau_o) = \sum_{i=1}^{m} \alpha_i \left( \sum_{j=1}^{J} E_{\tau_o} \{ \psi_i^{(1)}(u_j, \tau_o) \} \right)^T (\tau - \tau_o)
\]

Taking the difference between \( \alpha^T \sum_{j=1}^{J} E_{\tau_o} \{ \psi(u_j, \tau) - \psi(u_j, \tau_o) \} \) and \( J\alpha^T D_J (\tau - \tau_o) \) leaves the term

\[
\begin{align*}
&\left( \sum_{i=1}^{m} \alpha_i (\tau - \tau_o) \right)^T \left( \sum_{j=1}^{J} E_{\tau_o} \{ \psi_i^{(2)}(u_j, \tau^*) \} \right) (\tau - \tau_o) \leq \left( \sum_{i=1}^{m} \alpha_i^2 \right) \left( \sum_{i=1}^{m} (\tau - \tau_o)^T \left( \sum_{j=1}^{J} E_{\tau_o} \{ \psi_i^{(2)}(u_j, \tau^*) \} \right) (\tau - \tau_o) \right)^2 \\
&= \| \alpha \|^2 \sum_{i=1}^{m} \left[ (\tau - \tau_o)^T \left( \sum_{j=1}^{J} E_{\tau_o} \{ \psi_i^{(2)}(u_j, \tau^*) \} \right) (\tau - \tau_o) \right]^2 \\
&= \sum_{i=1}^{m} \left[ (\tau - \tau_o)^T \left( \sum_{j=1}^{J} E_{\tau_o} \{ \psi_i^{(2)}(u_j, \tau^*) \} \right) (\tau - \tau_o) \right]^2
\end{align*}
\]

By Result 5.3, we have

\[
\begin{align*}
&\sum_{i=1}^{m} \left[ (\tau - \tau_o)^T \left( \sum_{j=1}^{J} E_{\tau_o} \{ \psi_i^{(2)}(u_j, \tau^*) \} \right) (\tau - \tau_o) \right]^2 \\
&\leq AJ^2 \sum_{i=1}^{m} \left[ (\tau - \tau_o)^T (\tau - \tau_o) \right]^2 \\
&= AJ^2 m \| \tau - \tau_o \|^4 = AJ^2 n \| \tau - \tau_o \|^4
\end{align*}
\]
Thus,

\[
\sup_{\|\boldsymbol{\tau} - \boldsymbol{\tau}_0\| \leq B(m/J)^{1/2}} \left( \alpha^T \sum_{j=1}^{J} E_{\boldsymbol{\tau}_0} \left[ \psi(\mathbf{u}_j, \boldsymbol{\tau}) - \psi(\mathbf{u}_j, \boldsymbol{\tau}_0) \right] - J \alpha^T D_J(\boldsymbol{\tau} - \boldsymbol{\tau}_0) \right)^2 \leq AJ^2 n(m/J)^2 = An^3
\]

To prove Theorems 5.1 and 5.2, \( n^3 = o(nJ) \) or \( n^2 = o(J) \), and to prove Theorem 5.3, \( n^3 = o(J) \). For Theorems 5.1 and 5.2, \( n^3 = o(J / \log J) \) implies \( n^2 = o(J) \), and for Theorem 5.3, \( n^4 = o(J / \log J) \) implies \( n^3 = o(J) \). Thus, (C5) holds.

The final assumption of Theorem 5.1 is that the function \( \rho(\mathbf{u}_j, \boldsymbol{\tau}) \) is concave in \( \boldsymbol{\tau} \). The \( \rho \) function is concave in \( \boldsymbol{\tau} \) for the 1PL model, but not for the 2PL or 3PL models. Proof of Result 5.5 below is given in Section 5.4.

**Result 5.5** For the 1PL model, the function \( \rho(\mathbf{u}_j, \boldsymbol{\tau}) \) is concave in \( \boldsymbol{\tau} \).

### 5.4 Proof of Results 5.1 - 5.5

Let \( P_i(\theta) \) denote the value of the 1PL function from equation (1.8). The derivatives of the function \( \log P(\mathbf{u}|\theta) \) with respect to the item parameters for the 1PL model are

\[
\frac{\partial \log P(\mathbf{u}|\theta)}{\partial b_i} = -(u_i - P_i(\theta))
\]

\[
\frac{\partial^2 \log P(\mathbf{u}|\theta)}{\partial b_i^2} = -P_i(\theta)(1 - P_i(\theta))
\]

All other second derivatives are 0.

Let \( P_i(\theta) \) denote the value of the 2PL function from equation (1.9). The derivatives of the function \( \log P(\mathbf{u}|\theta) \) with respect to the item parameters for the 2PL model are

\[
\frac{\partial \log P(\mathbf{u}|\theta)}{\partial a_i} = (\theta - b_i)(u_i - P_i(\theta))
\]
\[
\frac{\partial \log P(u|\theta)}{\partial b_i} = -a_i (u_i - P_i(\theta))
\]
\[
\frac{\partial^2 \log P(u|\theta)}{\partial a_i^2} = (\theta - b_i)^2 P_i(\theta)(1 - P_i(\theta))
\]
\[
\frac{\partial^2 \log P(u|\theta)}{\partial b_i^2} = -a_i^2 P_i(\theta)(1 - P_i(\theta))
\]
\[
\frac{\partial^2 \log P(u|\theta)}{\partial a_i \partial b_i} = -a_i (\theta - b_i) P_i(\theta)(1 - P_i(\theta)) - (u_i - P_i(\theta))
\]

All other second derivatives are zero.

Let \( P^*_i(\theta) \) denote the value of the 2PL function from equation (1.9) and let \( P_i(\theta) \) denote the value of the 3PL function from equation (1.10). Define the variable \( W_i(\theta) \) as

\[
W_i(\theta) = \frac{P^*_i(\theta)(1 - P^*_i(\theta))}{P_i(\theta)(1 - P_i(\theta))}
\]

and the variable \( Z_i(\theta) \) as

\[
Z_i(\theta) = (1 - 2P^*_i(\theta)) - (1 - c_i) W_i(\theta)(1 - 2P_i(\theta))
\]

The derivatives of the function \( \log P(u|\theta) \) with respect to the item parameters for the 3PL model are

\[
\frac{\partial \log P(u|\theta)}{\partial a_i} = (1 - c_i) W_i(\theta)(\theta - b_i)(u_i - P_i(\theta))
\]
\[
\frac{\partial \log P(u|\theta)}{\partial b_i} = -(1 - c_i) a_i W_i(\theta)(u_i - P_i(\theta))
\]
\[
\frac{\partial \log P(u|\theta)}{\partial c_i} = (1 - c_i)^{-1} (u_i - P_i(\theta))
\]
\[
\frac{\partial^2 \log P(u|\theta)}{\partial a_i^2} = (1 - c_i)(\theta - b_i) W_i(\theta) Z_i(\theta)(u_i - P_i(\theta))
\]
\[
\quad - (1 - c_i)^2 (\theta - b_i)^2 W_i(\theta) P^*_i(\theta)(1 - P^*_i(\theta))
\]
\[
\frac{\partial^2 \log P(u|\theta)}{\partial b_i^2} = (1 - c_i) a_i^2 (u_i - P_i(\theta)) W_i(\theta) Z_i(\theta) - (1 - c_i)^2 a_i^2 W_i(\theta) P^*_i(\theta)(1 - P^*_i(\theta))
\]
\[
\frac{\partial^2 \log P(u|\theta)}{\partial c_i^2} = (1 - c_i)^{-2} \frac{(u_i - P_i(\theta))}{P_i(\theta)} - (1 - c_i)^{-2} \frac{u_i(1 - P^*_i(\theta))}{(P_i(\theta))^2}
\]

102
\[
\frac{\partial^2 \log P(u|\theta)}{\partial a_i \partial b_i} = -(1 - c_i)W_i(\theta)(u_i - P_i(\theta)) - a_i(1 - c_i)(\theta - b_i)W_i(\theta)Z_i(\theta)(u_i - P_i(\theta)) + a_i(1 - c_i)^2(\theta - b_i)W_i(\theta)P_i^*(\theta)(1 - P_i^*(\theta))
\]
\[
\frac{\partial^2 \log P(u|\theta)}{\partial a_i \partial c_i} = -\frac{\left(\theta - b_i\right)u_iP_i^*(\theta)(1 - P_i^*(\theta))}{(P_i(\theta))^2}
\]
\[
\frac{\partial^2 \log P(u|\theta)}{\partial b_i \partial c_i} = a_i\frac{u_iP_i^*(\theta)(1 - P_i^*(\theta))}{(P_i(\theta))^2}
\]

All other second derivatives are 0.

Denote the posterior distribution of \( \theta \) given a response pattern \( u \) as \( P(\theta|u) \). The first derivative of the function \( \log P(u) \) with respect to an item parameter \( \tau \) is

\[
\frac{\partial \log P(u)}{\partial \tau} = \int \frac{\partial \log P(u|\theta)}{\partial \tau} P(\theta|u) d\theta
\]

and the second derivative of the function \( \log P(u) \) with respect to item parameters \( \tau \) and \( \tau' \) is

\[
\frac{\partial^2 \log P(u)}{\partial \tau \partial \tau'} = \int \left[ \frac{\partial^2 \log P(u|\theta)}{\partial \tau \partial \tau'} + \left(\frac{\partial \log P(u|\theta)}{\partial \tau}\right)\left(\frac{\partial \log P(u|\theta)}{\partial \tau'}\right) \right] P(\theta|u) d\theta
\]

\[
-\left(\frac{\partial \log P(u)}{\partial \tau}\right)\left(\frac{\partial \log P(u)}{\partial \tau'}\right)
\]

**Proof of Result 5.1.** For all response patterns \( u \) and all item parameters \( \tau \), we have

\[
\left| \frac{\partial \log P(u)}{\partial \tau} \right| \leq \int \left| \frac{\partial \log P(u|\theta)}{\partial \tau} \right| P(\theta|u) d\theta
\]

To prove Result 5.1 holds, we need to show the right side of equation (5.15) is bounded for all item parameters \( \tau \) and all response patterns \( u \). For the 1PL model,

\[
\int \left| \frac{\partial \log P(u|\theta)}{\partial b_i} \right| P(\theta|u) d\theta = \int |(u_i - P_i(\theta))| P(\theta|u) d\theta \leq \int P(\theta|u) d\theta = 1
\]
Thus, Result 5.1 holds for the 1PL model. For the 2PL model,

\[
\int \frac{\partial \log P(u|\theta)}{\partial b_i} P(\theta|u) d\theta = a_i \int |(u_i - P_i(\theta))| P(\theta|u) d\theta \leq a_i \\
\int \frac{\partial \log P(u|\theta)}{\partial a_i} P(\theta|u) d\theta = \int |(\theta - b_i)(u_i - P_i(\theta))| P(\theta|u) d\theta \\
\leq \int |(\theta - b_i)| P(\theta|u) d\theta \\
\leq \int |\theta| P(\theta|u) d\theta + |b_i| \int P(\theta|u) d\theta \\
= \int |\theta| P(\theta|u) d\theta + |b_i|
\]

By the regularity conditions, \(a_i\) and \(b_i\) is bounded for all \(i\) and the first moment of the posterior distribution of \(\theta\) is bounded for all response patterns \(u\). Thus, Result 5.1 holds for the 2PL model. For the 3PL model, when \(c_i = 0\), \(W_i(\theta) = 1\) for all \(\theta\). If \(c_i > 0\), \(W_i(\theta)\) is an increasing function in \(\theta\) with

\[
\lim_{\theta \to -\infty} W_i(\theta) = 0 \quad \text{and} \quad \lim_{\theta \to \infty} W_i(\theta) = \frac{1}{1 - c_i}
\]

Thus, \(W_i(\theta)\) is bounded by \(1/(1 - c_i)\) for all \(\theta\). For the 3PL model,

\[
\int \frac{\partial \log P(u|\theta)}{\partial b_i} = a_i(1 - c_i) \int W_i(\theta)|(u_i - P_i(\theta))| P(\theta|u) d\theta \leq a_i \\
\int \frac{\partial \log P(u|\theta)}{\partial a_i} = (1 - c_i) \int W_i(\theta)|(\theta - b_i)|(u_i - P_i(\theta))| P(\theta|u) d\theta \\
\leq \int |\theta| P(\theta|u) d\theta + |b_i| \\
\int \frac{\partial \log P(u|\theta)}{\partial c_i} = (1 - c_i)^{-1} \int \frac{|u_i - P_i(\theta)|}{P_i(\theta)} P(\theta|u) d\theta \\
\leq (1 - c_i)^{-1} \int (P_i(\theta))^{-1} P(\theta|u) d\theta \\
\leq (1 - c_i)^{-1} \epsilon^{-1}
\]
By the regularity conditions, $a_i$ and $b_i$ are bounded for all $i$, $c_i$ is bounded away from 1 for all $i$, the first moment of the posterior distribution of $\theta$ is bounded for all response patterns $u$, and $\epsilon^{-1}$ is bounded. Thus, Result 5.1 holds for the 3PL model.

**Proof of Result 5.2.** Since Result 5.1 holds, the third term in equation (5.14) is bounded for all item parameters $\tau$ and $\tau'$ and all response patterns $u$. Thus, to prove Result 5.2 holds, we must show the terms

$$\int \left| \frac{\partial^2 \log P(u|\theta)}{\partial \tau \partial \tau'} \right| P(\theta|u) d\theta \quad \text{and} \quad \int \left| \frac{\partial \log P(u|\theta)}{\partial \tau} \right| \left| \frac{\partial \log P(u|\theta)}{\partial \tau'} \right| P(\theta|u_r) d\theta$$

(5.16)

are bounded for all item parameters $\tau$ and $\tau'$ and all response patterns $u$. When the item parameters $\tau$ and $\tau'$ are from different items, the partial derivative of $\log P(u|\theta)$ with respect to $\tau$ and $\tau'$ is 0. Thus, to show the first term above is bounded, we need to show the partial derivative of $\log P(u|\theta)$ with respect to $\tau$ and $\tau'$ is bounded when $\tau$ and $\tau'$ are from the same item. For the 1PL model,

$$\int \left| \frac{\partial^2 \log P(u|\theta)}{\partial b_i^2} \right| = \int |P_i(\theta)(1 - P_i(\theta))|P(\theta|u)d\theta \leq 1$$

Thus, the first term is bounded for all item parameters $\tau$ and $\tau'$ and for all response patterns $u$ for the 1PL model. For the 2PL model,

$$\int \left| \frac{\partial^2 \log P(u|\theta)}{\partial b_i^2} \right| = a_i^2 \int |P_i(\theta)(1 - P_i(\theta))|P(\theta|u)d\theta \leq a_i^2$$

$$\int \left| \frac{\partial^2 \log P(u|\theta)}{\partial a_i^2} \right| = \int |(\theta - b_i)|^2 P_i(\theta)(1 - P_i(\theta))P(\theta|u)d\theta$$

$$\leq \int \theta^2 P(\theta|u)d\theta + 2|b_i| \int |\theta|P(\theta|u)d\theta + b_i^2$$

$$\int \left| \frac{\partial^2 \log P(u|\theta)}{\partial a_i \partial b_i} \right| = a_i \int |(\theta - b_i)|P_i(\theta)(1 - P_i(\theta)) - (u_i - P_i(\theta))|P(\theta|u)d\theta$$

$$\leq a_i \int |(\theta - b_i)|P(\theta|u)d\theta + \int |(u_i - P_i(\theta))|P(\theta|u)d\theta$$

$$\leq a_i \int |\theta|P(\theta|u)d\theta + a_i |b_i| + 1$$
By the regularity conditions, the first term is bounded for all item parameters $\tau$ and $\tau'$ and for all response patterns $\mathbf{u}$ for the 2PL model. For the 3PL model, if $c_i = 0$, the function $Z_i(\theta) = 0$ for all $\theta$. If $c_i > 0$, then $Z_i(\theta)$ is a decreasing function in $\theta$ with

$$
\lim_{\theta \to -\infty} Z_i(\theta) = 1 \quad \text{and} \quad \lim_{\theta \to \infty} Z_i(\theta) = 0
$$

Thus, $Z_i(\theta) \leq 1$ for all $\theta$. For the 3PL model,

$$
\int \left| \frac{\partial^2 \log P(\mathbf{u}|\theta)}{\partial b_i^2} \right| = (1 - c_i)^2 a_i^2 \int W_i(\theta) P_i^*(\theta)(1 - P_i^*(\theta)) P(\theta|\mathbf{u}) d\theta \\
\quad + (1 - c_i) a_i^2 \int W_i(\theta) Z_i(\theta) (u_i - P_i(\theta)) |P(\theta|\mathbf{u})| d\theta \\
\leq a_i^2 + a_i^2
$$

$$
\int \left| \frac{\partial^2 \log P(\mathbf{u}|\theta)}{\partial a_i^2} \right| = (1 - c_i)^2 \int |(\theta - b_i)|^2 W_i(\theta) P_i^*(\theta)(1 - P_i^*(\theta)) P(\theta|\mathbf{u}) d\theta \\
\quad + (1 - c_i) \int |(\theta - b_i)| W_i(\theta) Z_i(\theta) |u_i - P_i(\theta)| P(\theta|\mathbf{u}) d\theta \\
\leq \int \theta^2 P(\theta|\mathbf{u}_r) d\theta + 2|b_i| \int |\theta| P(\theta|\mathbf{u}) d\theta + b_i^2 + \int |\theta| P(\theta|\mathbf{u}_r) d\theta + |b_i|
$$

$$
\int \left| \frac{\partial^2 \log P(\mathbf{u}|\theta)}{\partial c_i^2} \right| = (1 - c_i)^{-2} \int \left( \frac{|u_i - P_i(\theta)|}{P_i(\theta)} \right) P(\theta|\mathbf{u}) d\theta \\
\quad + (1 - c_i)^{-2} \int \left( \frac{u_i(1 - P_i^*(\theta))}{P_i^*(\theta)^2} \right) P(\theta|\mathbf{u}) d\theta \\
\leq (1 - c_i)^{-2} e^{-1} + (1 - c_i)^{-2} e^{-2}
$$

$$
\int \left| \frac{\partial^2 \log P(\mathbf{u}|\theta)}{\partial a_i \partial b_i} \right| = (1 - c_i) \int W_i(\theta) |(u_i - P_i(\theta))| P(\theta|\mathbf{u}) d\theta \\
\quad + a_i (1 - c_i) \int |(\theta - b_i)| W_i(\theta) Z_i(\theta) |u_i - P_i(\theta)| P(\theta|\mathbf{u}) d\theta \\
\quad + a_i (1 - c_i)^2 \int |(\theta - b_i)| W_i(\theta) P_i^*(\theta)(1 - P_i^*(\theta)) P(\theta|\mathbf{u}) d\theta \\
\leq 1 + 2 \left( a_i \int |\theta| P(\theta|\mathbf{u}) d\theta + |b_i| \right)
$$

$$
\int \left| \frac{\partial^2 \log P(\mathbf{u}|\theta)}{\partial b_i \partial c_i} \right| = a_i \int \frac{u_i P_i^*(\theta)(1 - P_i^*(\theta))}{(P_i(\theta))^2} P(\theta|\mathbf{u}) d\theta \\
\leq a_i e^{-2}
$$
\[
\int \left| \frac{\partial^2 \log P(u|\theta)}{\partial a_i \partial c_i} \right| = \int \left| (\theta - b_i) u_i P_i^+(\theta) (1 - P_i^+(\theta)) \right| \frac{P(\theta|u)}{\left( P_i(\theta) \right)^2} d\theta \\
\leq \epsilon^{-1} \int |\theta| P(\theta|u) d\theta + \epsilon^{-1} |b_i|
\]

By the regularity conditions, the first term is bounded for all item parameters \( \tau \) and \( \tau' \) and for all response patterns \( u \) for the 3PL model. In a similar manner, the second term in (5.16) can be shown to be bounded for all item parameters \( \tau \) and \( \tau' \) and all response patterns \( u \) for the 1PL model, and for the 2PL and 3PL model under the regularity conditions. Thus, Result 5.2 holds.

**Proof of Result 5.3.** For the 1PL, 2PL, and 3PL models, the expected value of the partial derivative of \( \log P(u) \) with respect to the item parameters \( \tau, \tau', \) and \( \tau'' \) is equivalent to

\[
E \left( \frac{\partial^3 \log P(u)}{\partial \tau \partial \tau' \partial \tau''} \right) = -E \left[ \left( \frac{\partial^2 \log P(u)}{\partial \tau \partial \tau'} \right) \left( \frac{\partial^2 \log P(u)}{\partial \tau \partial \tau''} \right) \right] \tag{5.17}
\]

By Result 5.2, the derivatives of \( \log P(u) \) with respect to \( \tau \) and \( \tau' \) are bounded for all item parameters \( \tau \) and \( \tau' \) and response patterns \( r \). Hence, the product of the second derivatives on the right side of equation (5.17) is bounded making the expectation bounded as well. Thus, Result 5.3 holds for the 1PL, 2PL and 3PL models.

**Proof of Result 5.4.** The sum of the eigenvalues for any matrix \( B \) is the sum of the diagonal terms of the matrix, also called the trace of matrix \( B \). The trace of the matrix \( D_J \) is

\[
\frac{1}{J} \sum_{i=1}^{m} \sum_{j=1}^{J} E_{\tau_i} \frac{\partial^2 \log P(u)}{\partial \tau_i^2}
\]

Since each term in the double sum is negative for all response patterns \( u \) and all item parameters \( \tau_i \), the trace of the matrix \( D_J \) is negative for all \( J \). Then, for each \( J \), there must exist at least one negative eigenvalue of the matrix \( D_J \). Hence, the minimum eigenvalue of the matrix \( D_J \) is negative for all \( J \). Thus, Result 5.4 holds.

**Proof of Result 5.5.** To show the \( \rho \) function is concave in the item parameters \( \tau \), we
will first show the function \( \log P(u|\theta) \) is concave in the item parameters \( \tau \). For the function \( \log P(u|\theta) \), the terms in the matrix of second derivatives with respect to the item parameters \( \tau \) are

\[
B(i, l) = \begin{cases} 
-P_i(\theta)(1 - P_i(\theta)) : i = l \\
0 : i \neq l
\end{cases}
\]

For the 1PL model, the matrix of second derivatives of the function \( \log P(u|\theta) \) is negative definite. Therefore, the function \( \log P(u|\theta) \) is concave in the item parameters \( \tau \).

If the function \( \log P(u|\theta) \) is concave, the function \( P(u|\theta) \) is concave as well. Since \( g(\theta) \) is a density function, and by definition non-negative for all \( \theta \), the function \( P(u) \) is also concave in the item parameters \( \tau \). Finally, if the function \( P(u) \) is concave, so is the log of the function. Thus, for the 1PL model, the function \( \rho = \log P(u) \) is concave.
References


112


Vita

Amy Goodwin Froelich was born in Moline, Illinois on May 26, 1973. She graduated from the University of Illinois at Urbana-Champaign in May, 1994, with a B.S. in Secondary Mathematics Education. She graduated with the distinction, Highest University Honors, Bronze Tablet and was named a Bagley Scholar by the College of Education.

After teaching mathematics at United Township High School in East Moline, Illinois, Amy returned to the University of Illinois at Urbana-Champaign in August, 1995 to pursue a Ph.D. in Statistics. While at the university, she served as a Teaching Assistant for both Statistics 100 and Statistics 310. In the Fall of 1996, she was named to the Incomplete List of Instructors Rated as Excellent by their Students for teaching Statistics 100. She also worked as a consultant in the Illinois Statistics Office.

Since 1997, Amy has worked as a Research Assistant in the Statistical Laboratory for Educational and Psychological Measurement under the direction of Professor William Stout. Her main areas of research have been improvements to a statistical procedure assessing latent model unidimensionality of both dichotomous and polytomous test items, developing a statistical procedure to assess the latent model unidimensionality of Computer Adaptive Test (CAT) items, proving asymptotic properties of item parameter estimates obtained from the marginal maximum likelihood estimation method, and developing methods to determine the amount of test bias across several different test administrations using category-based test item bundles.