A Multidimensional Continuous Item Response Model for Probability Testing

Yiping Zhang and Hiroshi Watanabe
Center for Research on Educational Testing

Probability Testing (PT)

• a response method for multiple-choice (MC) items (de Finetti, 1965)
• an examinee gives to each response option his/her subjective probability of its being correct as an expression of partial knowledge (examinees must understand the scoring rule)

• **Example:**
  “She certainly looks beautiful ___ a Japanese kimono.”
  
<table>
<thead>
<tr>
<th>Option</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>A: in</td>
<td>(50%)</td>
</tr>
<tr>
<td>B: for</td>
<td>(0%)</td>
</tr>
<tr>
<td>C: on</td>
<td>(20%)</td>
</tr>
<tr>
<td>D: with</td>
<td>(30%)</td>
</tr>
</tbody>
</table>
Scoring PT Responses

• The original scoring rule proposed by de Finetti (1965) was based on a decision theory, it involves the subjective probabilities for all alternative.
  \[ S = 2p_h - \sum_{r=1}^{q} p_r^2 \]

• Zhang (2007) proposed a simpler scoring rule, taking the subjective probability which was attached to the correct response option as the score for that item.
  \[ S = p_h \]

Advantages of PT

• provide a more sensitive measurement for the partial knowledge and enables more accurate measurement of the target ability than other response methods

• subject does not need a complicated judgment process to choose

• maintains all advantages of using MC items such as objectivity in scoring, flexibility in item contents, etc.

Disadvantages of PT

• examinees must understand the concept of probability

• needs training for expressing partial knowledge by subjective probability
**Notation**

Examinee: \( i = (1, 2, \cdots, N) \)

Item: \( j = (1, 2, \cdots, m) \)

Item score vector: \( s_j = p_{hj} \)

\( (P_{hj} \) is the subjective probability given to the correct response option for item \( j \)\)

The parameters of the item score distribution:

\[ S_j \sim Be(p_j, q_j) \]

Item ability score vector: \( Z_j \sim N(0,1) \)

Ability parameter: \( \Theta = (\theta_1, \theta_2, \cdots, \theta_k, \cdots, \theta_L) \)

The correlation coefficient of \( \theta_k \) and \( Z_j \): \( \rho_{kj} \)

**Transformed Item Score**

![Graph showing item score and equal percentile rank curve](image)

the item score \( 0 \leq S_j \leq 1 \) is the continuous data for \( j = (1, 2, \cdots, m) \), and the standard normal score \( Z_j \) that transformed from \( S_j \) can be thought as the ability for answering the item.
Transformed Item Score

The distribution of $S_j$ can be expressed by Beta distribution

$$Be(p_j, q_j)$$

by equation

$$K_{S_j}(p_j, q_j) = \Phi(z_j)$$

$S_j$ can be transformed into the standard normal score $Z_j$

Where,

$$\Phi(Z_j) = \int_0^{z_j} \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{w^2}{2}\right] dw$$

$$K_{S_j}(p_j, q_j) = \frac{Be(p_j, q_j)}{Be(p_j, q_j)} = \int_0^{z_j} t^{p_j-1}(1-t)^{q_j-1} dt$$

---

Model

Let $\Theta = (\theta_1, \theta_2, \ldots, \theta_g, \ldots, \theta_k)$ be the vector of the complete latent traits space, and $\theta_k \sim N(0, I), \Theta \sim N(0, C)$.

The joint probability density function of $\Theta$ and $Z_j$ can be expressed by the multivariate normal distribution as equation

$$f(Z_j, \Theta) = \frac{1}{(2\pi)^{n/2}|V_j|^{1/2}} \exp\left[-\frac{T_j^T V_j^{-1} T_j}{2}\right]$$

Where, $T_j = (Z_j, \theta_1, \theta_2, \ldots, \theta_k)^T$, $V_j = \begin{bmatrix} 1 & \Lambda_j' \\ \Lambda_j & C \end{bmatrix}$

$\Lambda_j = (\rho_{1j}, \rho_{2j}, \ldots, \rho_{sj}, \ldots, \rho_{kj})'$

$\rho_{sj}$ is the correlation coefficient vector of $\theta_s$ and $Z_j$.

And, $C = C^T$, $C^{-1} = (C^{-1}) = (C^T)^{-1}$
Model

When Θ and Λ, are given, the conditional distribution of \( Z_j \) becomes

\[
 f(Z_j | Θ) = \frac{1}{(2\pi)^{1/2}V_{j\Lambda}^{1/2}} \exp \left[ -\frac{(Z_j - u_{j\Lambda})^T V_{j\Lambda}^{-1} (Z_j - u_{j\Lambda})}{2} \right]
\]

This conditional distribution is normal distribution too.

Where, \( u_{j\Lambda} = Λ_j^T C^{-1} Θ \) and \( V_{j\Lambda} = 1 - Λ_j^T C^{-1} Λ_j \).

The marginal distribution of \( θ_g \) and \( Z_j \) can be expressed by equation

\[
 f(Z_j, θ_g) = \frac{1}{(2\pi)^{1/2}V^T\rho V^{1/2}} \exp \left[ -\frac{W_{gj}V_{gj}^{-1}W_{gj}}{2} \right]
\]

Where, \( W_{gj} = (Z_j^T θ_g)' \), \( V_g = \begin{bmatrix} 1 & ρ_{gj} \\ ρ_{gj} & 1 \end{bmatrix} \).

This is the same as the uni-dimensional model of Zhang (2007).

---

Estimation of Item Parameter

The parameters of the item score distribution

**Moment method** is thought to be a convenient and effective method when the size of the sample is large enough (Elderton & Johnson, 1969).

Because the first and the second moments equation of \( S_j \) are

\[
 E(S_j) = \frac{p_j}{p_j + q_j} \quad \text{and} \quad Var(S_j) = \frac{p_j q_j}{(p_j + q_j)^2 (p_j + q_j + 1)}
\]

If substitute \( \bar{s}_j \) and \( c_j \) as moment estimators of the population mean and the population variance, the estimators of the item parameters \( p_j \) and \( q_j \) can be derived by equation

\[
 \hat{p}_j = \frac{\bar{s}_j (\bar{s}_j - \bar{s}^2_j - c_j)}{c_j} \quad \text{and} \quad \hat{q}_j = \frac{(1 - \bar{s}_j) (\bar{s}_j - \bar{s}^2_j - c_j)}{c_j}
\]

---
Estimation of Item Parameter

The parameters of the item score distribution

The maximum likelihood method is usually more accurate than the moment method, when \( p_j \) and \( q_j \) are small or their difference is big.

The maximum likelihood equations (detailed in Chapter 25 of Johnson & Kotz & Balakrishnan, 1995) as following can be used.

\[
\psi(\tilde{p}_j) - \psi(\tilde{p}_j + \tilde{q}_j) = \frac{1}{n} \sum_{i=1}^{n} \ln(s_{ij}) \\
\psi(\tilde{q}_j) - \psi(\tilde{p}_j + \tilde{q}_j) = \frac{1}{n} \sum_{i=1}^{n} \ln(1 - s_{ij})
\]

Here, \( \psi(t) = \frac{d[\log \Gamma(t)]}{dt} = \frac{d \log \Gamma(t)}{\log \Gamma(t)} \quad t > 0 \)

Estimation of Item Parameter

\[ \Lambda_j = (\rho_{1j}, \rho_{2j}, \cdots, \rho_{gj}, \cdots, \rho_{sj})' \]

In the case of subtests unknown

Maximum likelihood method: the estimation method which is the same as factor analysis (ef. for instance, Lawley & Maxwell, 1971) can be used.

Bayesian method: the estimation method which suggested by Press and Shigemasu (1989) can be used.
Estimation of Item Parameter

\[ \Lambda_j = (\rho_{1j}, \rho_{2j}, \ldots, \rho_{sj}, \ldots, \rho_{kj})' \]

In the case of subtests known

\[ Z \] is the data matrix for subtest \( g \).

Where, \( Z \) is the data matrix for test.

\[ Q \] is the \( m \times m \) matrix for given the items to subtest \( g \).

\[ \sum_{g=1}^{k} m_g = m \]

\[ \lambda_{m,g} = (\rho_{1g}, \rho_{2g}, \ldots, \rho_{mg}) \] is the correlation coefficient vector of \( \theta_g \) and the items of subtest \( g \). It can be estimated in the same way as the method for deriving factor loading vector in one factor analysis.

Estimation of Item Parameter

\[ \Lambda_j = (\rho_{1j}, \rho_{2j}, \ldots, \rho_{sj}, \ldots, \rho_{kj})' \]

In the case of subtests known (oblique)

The ability of subtest \( g \) can be derived by equation

\[ \hat{\theta}_g = Z_g R_g^{-1} \hat{\lambda}_{m,g} \]

Where,

\[ R_g = \frac{1}{n} Z_g' Z_g \]

The correlation coefficient of \( \theta_g \) and \( Z_j \) can be derived by equation

\[ \hat{\rho}_{gj} = \frac{1}{n} Z_j' \hat{\theta}_g \]

Then,

\[ \hat{\lambda}_j = (\hat{\rho}_{1j}, \hat{\rho}_{2j}, \ldots, \hat{\rho}_{sj}, \ldots, \hat{\rho}_{kj})' \]

And, the coefficient matrix of the subtest abilities can be derived by equation

\[ \hat{C} = \frac{1}{n} \hat{\theta} \hat{\theta} \]
Estimation of Trait Value

Maximum likelihood method

The likelihood function can be written as equation

\[ l(\Theta | z_1, z_2, \ldots, z_m) \propto \exp \left[ - \sum_{j=1}^{m} \frac{(Z_j - u_{j|x^2})'V_{j|x^2}^{-1}(Z_j - u_{j|x^2})}{2} \right] \]

Because \( C = C' \), then, \( C^{-1} = (C')^{-1} \).

This likelihood function becomes maximum at

\[ \hat{\Theta} = \left[ \sum_{j=1}^{m} \left( V_{j|x^2}^{-1} A_j A_j' C^{-1} \right) \right]^{-1} \left[ \sum_{j=1}^{m} \left( V_{j|x^2}^{-1} A_j Z_j \right) \right] \]

---

Estimation of Trait Value

Bayesian estimation method

When using \( \Theta \sim N_k(u_0, D_0) \) as the prior distribution, the posterior distribution becomes

\[ p(\Theta | z_1, z_2, \ldots, z_m) \propto \exp \left[ - \frac{(\Theta - F_m)' H_m^{-1}(\Theta - F_m)}{2} \right] \]

Where, \( F_m = H_m \left( D_0^{-1} + \sum_{j=1}^{m} D_j^{-1} G_j \right) \), \( H_m^{-1} = D_0^{-1} + \sum_{j=1}^{m} D_j^{-1} \),

and \( G_j = C \Lambda_j (\Lambda_j' \Lambda_j)^{-1} Z_j \), \( D_j^{-1} = C^{-1} A_j V_{j|x^2}^{-1} A_j' C^{-1} \).

When \( Z_{m+1} \) is added.

Then \( F_{m+1} = H_{m+1} \left( D_0^{-1} u_0 + \sum_{j=1}^{m+1} D_j^{-1} G_j \right) \),

and \( H_{m+1}^{-1} = D_0^{-1} + \sum_{j=1}^{m+1} D_j^{-1} = H_m^{-1} + D_{m+1}^{-1} \).
**Item Information Function Matrix**

Fisher's information matrix is defined as a square matrix of order \( g \), the matrix is given by

\[
I_j(\Theta) = -E\left[ \frac{\partial^2 \ln f(z_j | \Theta)}{\partial \Theta^2} \right] = V_{j1}^{-1}(C^{-1} \Lambda_j \Lambda_j' C^{-1})'
\]

This is a square matrix of order \( k \).

Let \( a_j = (1 - \Lambda_j' C^{-1} \Lambda_j)^{-1/2} C^{-1} \Lambda_j \)

Then \( I_j(\Theta) = a_j a_j' \)

Here, \( a_j \) is the item discrimination parameter.

---

**Test Information Function Matrix**

The information function of \( Z_1, Z_2, \ldots, Z_m \) can be written as equation

\[
I(\Theta) = -E\left[ \frac{\partial^2 \ln l(\Theta | z_1, z_2, \ldots, z_m)}{\partial \Theta^2} \right]
\]

Because \( Z_i \) and \( Z_j \) are independent (since item information function does not depend on \( \theta \)), the amount of the test information is the sum of the amount of the items information that included in the test.

\[
I(\Theta) = \sum_{j=1}^{m} I_j(\Theta)
\]
Application of Multidimensional Continuous Model

• to raise the measurement precision for Entrance examination, Qualifying examination, etc.,

• to raise the measurement precision for Reading test, Writing test, Knowledge application test, etc.,

References


