A Note on N in Bayesian Information Criterion for Item Response Models

Applied Psychological Measurement

2018, Vol. 42(2) 169–172

© The Author(s) 2017

Reprints and permissions:

sagepub.com/journalsPermissions.nav

DOI: 10.1177/0146621617726791

journals.sagepub.com/home/apm

Sun-Joo Cho¹ and Paul De Boeck^{2,3}

Abstract

This brief report derives the N in the penalty term of the Schwarz's (1978) Bayesian information criterion (BIC) for two-parameter logistic item response models. The results in this study show that the N is the number of persons for fixed item models, whereas it is the number of observations (the Number of Persons times the Number of Items) for random item models. Given these results, the authors recommend researchers to calculate the BIC or to validate the BIC value that shows in the output of software instead of accepting the output value without a further check of implicit assumptions made for the software.

Keywords

item response theory, Bayesian information criterion, information function

It is common to use information criteria such as Schwarz's (1978) Bayesian information criterion (BIC) in model selection for item response models, as exhibited by their automatic computation in software packages. It has been observed that there are discrepancies in the formula of the BIC for item response models in software. The discrepancies concern the sample size term N in the penalty term of the BIC. As an example, for the one-parameter item response model with a random person effect and a fixed item effect using (an approximate) marginal maximum likelihood estimation, N is the number of persons in flexMIRT (Cai, 2015), Mplus (Muthén & Muthén, 1998-2015), and SAS NL MIXED (e.g., Nandakumar & Hotchkiss, 2012, for item response models), whereas it is the total number of observations (i.e., the Number of Persons $\times$ the Number of Items) in the glmer function of lme4 R package (Bates, Mächler, Bolker, & Walker, 2015). In the glmer function, the N is the total number of observations for the one-parameter item response model with a random person effect and a random item effect. However, the model cannot be fit using flexMIRT, SAS NL MIXED, and Mplus.¹

In spite of the popularity of the BIC, to the authors' knowledge, there is no derivation available for the N to be used for the BIC of item response models. Different calculation for BIC can result in different model selections and the same calculation for BIC can be misleading if that same calculation is correct for one model but not for another. The aim of this brief report

¹Vanderbilt University, Nashville, TN, USA

²Ohio State University, Columbus, OH, USA

³KU Leuven, Leuven, Belgium

Corresponding Author:

Sun-Joo Cho, Vanderbilt University, 230 Appleton Place, Nashville, TN 37203, USA.

Email: sj.cho@vanderbilt.edu

is to provide a derivation, based on the previous derivations of BIC in general (e.g., Kass & Raftery, 1995, p. 779). It focuses on two-parameter logistic (2PL) item response models for binary responses. The derivation below is applicable to the one-parameter, two-parameter, and three-parameter item response models for the binary responses.

Schwarz (1978) derived the BIC as an asymptotic approximation of the Bayesian posterior probability of a candidate model M :

$$\log P(M | \mathbf{y}) = \log P(\mathbf{y} | \hat{\boldsymbol{\theta}}, M) - \frac{K}{2} \log N + O(N^{-1}), \quad (1)$$

where $\mathbf{y}$ is data, $\boldsymbol{\theta}$ is a set of parameters, $P(\mathbf{y} | \hat{\boldsymbol{\theta}}, M)$ is the likelihood of data $\mathbf{y}$ evaluated at parameter estimates $\hat{\boldsymbol{\theta}}$. K is the number of parameters, and N is sample size. For a large sample size, the order $O(N^{-1})$ term can be dropped in Equation 1. Below, the derivation of N is presented for (a) the 2PL item response model having a random person effect and a fixed item effect and (b) the 2PL item response model having a random person effect and a random item effect. In the derivations, the notation for the candidate model M as it appears in Equation 1 is omitted for reasons of simplicity.

Random Person Effect and Fixed Item Effect

The 2PL item response model can be specified as

$$\text{logit}[P(y_{ji} = 1 | \theta_j)] = \alpha_i(\theta_j - \beta_i), \quad (2)$$

where j is an index of person ($j = 1, \dots, J$), i is an index of item ($i = 1, \dots, I$), y_{ji} is an item response, α_i is an item discrimination parameter, β_i is an item location parameter, and θ_j is a latent variable. To identify the model, a standard normal distribution can be set on θ_j . For a sample of J independent persons, the marginal probability of the data, $\mathbf{y} = [y_1, \dots, y_j, \dots, y_J]'$, is

$$P(\mathbf{y}) = \prod_{j=1}^J \int_{\boldsymbol{\theta}} \left[\prod_{i=1}^I P(y_{ji} = 1 | \theta_j; \hat{\boldsymbol{\theta}}_i)^{y_{ji}} P(y_{ji} = 0 | \theta_j; \hat{\boldsymbol{\theta}}_i)^{1-y_{ji}} \right] P(\theta_j) d\theta_j, \quad (3)$$

where $\hat{\boldsymbol{\theta}}_i = [\alpha_i, \beta_i]'$ and $P(\theta_j)$ is a prior distribution for θ_j .

The marginal probability of the data $\mathbf{y}$ can be approximated using Laplace approximation (Tierney & Kadane, 1986):

$$\log P(\mathbf{y}) = \log P(\mathbf{y} | \hat{\boldsymbol{\theta}}; \hat{\boldsymbol{\theta}}) + \log P(\hat{\boldsymbol{\theta}}) + \frac{K}{2} \log(2\pi) - \frac{1}{2} \log ||\mathbf{A}|| + O(I^{-1}), \quad (4)$$

where $\boldsymbol{\theta} = [\theta_1, \dots, \theta_J]'$ and $\mathbf{A}$ is the expected information matrix of item parameter estimates, $\hat{\boldsymbol{\theta}} = [\hat{\alpha}, \hat{\beta}]'$.

For a sample of J independent persons, the $\mathbf{A}$ in Equation 4 approaches $J \cdot \mathbf{I}_j(\boldsymbol{\theta})$, where J is the number of persons and $\mathbf{I}_j(\boldsymbol{\theta})$ is the Fisher information (or expected information) of item parameter estimates for a person j (Thissen & Wainer, 1982, p. 411):

$$\mathbf{A} = J \cdot \mathbf{I}_j(\boldsymbol{\theta}) \cong J \cdot E \left[-\ddot{P}_j(\hat{\boldsymbol{\theta}}) \right], \quad (5)$$

where $-\ddot{P}_j(\boldsymbol{\theta})$ is minus the second derivative of the loglikelihood function (called the Hessian matrix or the observed information). The matrix of $\mathbf{I}_j(\boldsymbol{\theta})$ is a $(K \times K)$ matrix (e.g., $2I \times 2I$ matrix for the 2PL item response models). Thus, $||\mathbf{A}|| \equiv ||\mathbf{I}_j(\boldsymbol{\theta})|| \equiv J^K ||\mathbf{I}_j(\boldsymbol{\theta})||$.

Inserting $\|A\| \equiv \|JI_j(\boldsymbol{\theta})\| \equiv J^K \|I_j(\boldsymbol{\theta})\|$ into Equation 4, the approximation is written as

$$\log P(\mathbf{y}) = \log P(\mathbf{y}|\boldsymbol{\theta}; \hat{\boldsymbol{\theta}}) + \log P(\boldsymbol{\theta}) + \frac{K}{2} \log(2\pi) - \frac{K}{2} \log(J) - \frac{1}{2} \log \|I_j\| + O(I^{-1}). \quad (6)$$

In Equation 6, $\log P(\boldsymbol{\theta})$, $(K/2)\log(2\pi)$, and $(1/2)\log \|I_j\|$ have an error of order $O(1)$. When ignoring the terms having an error of order $O(1)$, Equation 6 is

$$\log P(\mathbf{y}) = \log P(\mathbf{y}|\boldsymbol{\theta}; \hat{\boldsymbol{\theta}}) - \frac{K}{2} \log(J) + O(I^{-1}). \quad (7)$$

As presented in Equation 7, N in BIC is J .

Random Person Effect and Random Item Effect

Item parameters in Equation 2 can be random effects in specialized measurement problems (see De Boeck, 2008, for a discussion). For a sample of J independent persons and a sample of I independent items, the marginal probability of the data, $\mathbf{y} = [y_{11}, \dots, y_{ji}, \dots, y_{JI}]'$, is

$$P(\mathbf{y}) = \prod_{j=1}^J \prod_{i=1}^I \int \int_{\boldsymbol{\theta} \quad \boldsymbol{\xi}} \left[P(y_{ji} = 1 | \boldsymbol{\theta}_j, \boldsymbol{\xi}_i; \hat{\boldsymbol{\theta}})^{y_{ji}} \left[P(y_{ji} = 0 | \boldsymbol{\theta}_j, \boldsymbol{\xi}_i; \hat{\boldsymbol{\theta}}) \right]^{1-y_{ji}} \right] P(\boldsymbol{\theta}_j) P(\boldsymbol{\xi}_i) d\boldsymbol{\theta}_j d\boldsymbol{\xi}_i, \quad (8)$$

where $\boldsymbol{\xi}_i = [\alpha_i, \beta_i]'$ and $P(\boldsymbol{\xi}_i)$ is a prior distribution of item parameters. When item parameters are random, the log-normal distribution can be imposed on item discrimination parameters ($\alpha_i \sim \log N(0, \sigma_\alpha)$) and the normal distribution can be imposed on item location parameters ($\beta_i \sim N(0, \sigma_\beta)$). In such a case, there are three parameters ($\boldsymbol{\theta} = [\mu, \sigma_\alpha, \sigma_\beta]'$), where μ is a grand mean across persons and items, σ_α is a variance of a random item discrimination parameter, and σ_β is a variance of a random location parameter.

Using the Laplace approximation, the marginal probability of the data $\mathbf{y}$ can be approximated as follows:

$$\log P(\mathbf{y}) = \log P(\mathbf{y}|\boldsymbol{\theta}, \boldsymbol{\xi}; \hat{\boldsymbol{\theta}}) + \log P(\boldsymbol{\theta}) + \log P(\boldsymbol{\xi}) + \frac{K}{2} \log(2\pi) - \frac{K}{2} \log \|A\| + O((JI)^{-1}), \quad (9)$$

where $\boldsymbol{\xi} = [\xi_1, \dots, \xi_i, \dots, \xi_I]'$ and A is the expected information matrix of the parameter estimates, $\hat{\boldsymbol{\theta}} = [\hat{\mu}, \hat{\sigma}_\alpha, \hat{\sigma}_\beta]'$.

For a sample of independent pairs of person j and item i , A in Equation 9 approaches $(JI) \cdot I_{ji}(\boldsymbol{\theta})$, where J is the number of persons, I is the number of items, and $I_{ji}(\boldsymbol{\theta})$ is the Fisher information of parameter estimates for a person j and an item i :

$$A = (JI) \cdot I_{ji}(\boldsymbol{\theta}) \cong (JI) \cdot E \left[-\ddot{P}_{ji}(\hat{\boldsymbol{\theta}}) \right], \quad (10)$$

where $-\ddot{P}_{ji}(\hat{\boldsymbol{\theta}})$ is minus the second derivative of the loglikelihood function. When IH_{ji} is a $(K \times K)$ matrix (i.e., $K = 3$ for $\boldsymbol{\theta} = [\mu, \sigma_\alpha, \sigma_\beta]'$), $\|A\| \equiv \|(JI) \times I_{ji}\| \equiv (JI)^K \|I_{ji}\|$.

With $\|A\| \equiv \|(JI) \times I_{ji}\| \equiv (JI)^K \|I_{ji}\|$, the approximation in Equation 9 can be rewritten as

$$\log P(\mathbf{y}) = \log P(\mathbf{y}|\boldsymbol{\theta}, \boldsymbol{\xi}; \hat{\boldsymbol{\theta}}) + \log P(\boldsymbol{\theta}) + \log P(\boldsymbol{\xi}) + \frac{K}{2} \log(2\pi) - \frac{K}{2} \log(JI) - \frac{1}{2} \log \|I_{ji}\| + O((JI)^{-1}), \quad (11)$$

where $\log P(\theta)$, $\log P(\xi)$, $(K/2)\log(2\pi)$, and $(1/2)\log||I_{ji}||$ are all approximations with an error of order $O(1)$. Ignoring the $O(1)$ terms, Equation 11 results in the following equation:

$$\log P(\mathbf{y}) = \log P(\mathbf{y}|\theta, \xi; \hat{\theta}) - \frac{K}{2} \log(JI) + O((JI)^{-1}). \quad (12)$$

It follows from Equation 12 that N is JI for item response models with random item parameters.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

Note

1. In Mplus, the one-parameter item response model with a random person effect and a random item effect can be fit using Bayesian analysis so that a likelihood-based Bayesian information criterion (BIC) is not available.

References

- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67, 1-48.
- Cai, L. (2015). flexMIRT version 3.0: A numerical engine for flexible multilevel and multidimensional item analysis and test scoring [Computer Software]. Raleigh-Durham, NC: Vector Psychometric Group.
- De Boeck, P. (2008). Random item IRT models. *Psychometrika*, 73, 533-559.
- Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, 90, 773-795.
- Muthén, L. K., & Muthén, B. O. (1998-2015). *Mplus users guide* (7th ed.). Los Angeles, CA: Author.
- Nandakumar, R., & Hotchkiss, L. (2012). PROC NLMIXED: For estimating parameters of IRT models. *Applied Psychological Measurement*, 38, 404-405.
- Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, 6, 461-464.
- Thissen, D., & Wainer, H. (1982). Some standard errors in item response theory. *Psychometrika*, 47, 397-412.
- Tierney, L., & Kadane, J. B. (1986). Accurate approximations for posterior moments and marginal densities. *Journal of the American Statistical Association*, 81, 82-86.