Documentation belonging to the ShadowCAT package

August 2016

Rivka M. de Vries
Alwin W. Stegeman
Karel A. Kroeze

The ShadowCAT R package allows computer adaptive testing of patients based on several IRT models, with or without the Shadow Testing adaptive test assembly procedure by van der Linden (2000). The goal is to obtain an accurate estimate of a subject's score on one ore more latent traits denoted theta, while minimizing the number of items that need to be administered to the subject. Item characteristics, like discrimination and difficulty parameters, are assumed to be known. The ShadowCAT procedure basically consists of two iterative components:

1. Select item that contains the maximum information for subject, possibly within constraints, and administer this item
2. Update the theta estimate based on the subject's new response

This procedure continues until the minimum number of items has been reached and (1) the maximum number of items has been reached, (2) the target variance has been reached, or (3) the theta estimate is far enough below a cutoff value at the current iteration. Below the two iterative steps, selecting a new item and updating the theta estimate, are described in more detail.

Select item with maximum information

The first "burn-in" items are selected independently of the theta estimate. They are either determined by the user or drawn randomly.

After the first burn-in items have been administered, the item containing maximum information given the current theta estimate is selected from the yet available items. It is possible to add constraints to this selection process, for instance to ensure that no more than five depression items are administered. In the latter case, item selection proceeds according to the Shadow Testing adaptive test assembly procedure by van der Linden (2000).

In order to find the item with maximum information, (a summary of the) item information for all available items needs to be computed. This computation consists of three steps:

(1) finding Fisher information for each item; (2) computing a summary of the Fisher information for each available item; and (3) selecting the item with the maximum information from the available items, possibly within the specified constraints. The computations are described in detail below.

Compute Fisher information for each item

The computation of Fisher information depends on how the probability of scoring in a certain item category given a latent trait is modeled. The four probability models covered by ShadowCAT are described in Appendix A. Below, the computation of Fisher information is described for each model, with parameters defined as in Appendix A.

Three-parameter logistic model. Let $\mathbf{q}$ be a vector of length K equal to

$$\mathbf{q} = \frac{\mathbf{p}_{*,1}}{\mathbf{p}_{*,2}} * \left(\frac{\mathbf{p}_{*,2} - \mathbf{c}}{\mathbf{1} - \mathbf{c}}\right) * \left(\frac{\mathbf{p}_{*,2} - \mathbf{c}}{\mathbf{1} - \mathbf{c}}\right),$$

where * denotes elementwise multiplication, the divisions are elementwise too, and $\mathbf{1}$ denotes a vector of ones of length K. Then the Fisher information matrix $\mathbf{S}_k$ for item k for the three-parameter logistic model is equal to

$$\mathbf{S}_k = \mathbf{a}_{k,*}\mathbf{a}_{k,*}^T q_k \, .$$

Generalized partial credit model. The Fisher information matrix $\mathbf{S}_k$ for item *k* for the generalized partial credit model is equal to

$$\mathbf{S}_k = \mathbf{a}_{k,*}\mathbf{a}_{k,*}^T(-h_k) \, .$$

Sequential model. Let $\mathbf{q}$ be a vector of length *K*, with element *k* equal to

$$q_k = \sum_{i=1}^{m_k+1}\left[p_{k,i}\sum_{j=2}^{i+1} f_{k,j}\ (1 - f_{k,j})\right].$$

Then the Fisher information matrix $\mathbf{S}_k$ for item *k* for the sequential model is equal to

$$\mathbf{S}_k = \mathbf{a}_{k,*}\mathbf{a}_{k,*}^T q_k \, .$$

Graded response model. Let $\mathbf{q}$ be a vector of length *K*, with element *k* equal to

$$q_k = \sum_{i=1}^{m_k+1}\left[p_{k,i}\left(f_{k,i}\left(1 - f_{k,i}\right) + f_{k,i+1}\left(1 - f_{k,i+1}\right)\right)\right].$$

Then the Fisher information matrix $\mathbf{S}_k$ for item *k* for the graded response model is equal to

$$\mathbf{S}_k = \mathbf{a}_{k,*}\mathbf{a}_{k,*}^T q_k \, .$$

Compute summary of Fisher information for each available item
The computation of the summarized Fisher information depends on how the Fisher information should be summarized. Summary options are determinant and posterior determinant (Segall, 1996; Segall, 2000), trace and posterior trace (Van der Linden, 1999), and posterior expected Kullback-Leibler information (Chang & Ying, 1996; Mulder & Van der Linden, 2010; Wang, Chang, & Boughton, 2010), where prior distributions may be normal or uniform. Computations are described below for each type of information summary. Here, $\mathbf{S}_k$ is the $Q \times Q$ matrix containing Fisher information for item *k*, $\mathbf{o}$ is a vector of length *J* containing the item indices of administered items, $\mathbf{z}$ a vector of length $(K - J)$ containing the item indices of available items, and $\mathbf{\Psi}$ the $Q \times Q$ covariance matrix of the normal prior on $\boldsymbol{\theta}$.

Determinant. When the information summary is the determinant, the item information u_i for available item z_i is equal to

$$u_i = \det\left(\boldsymbol{S}_{z_i} + \sum_{k=o_1}^{o_J}\boldsymbol{S}_k\right),$$

for i = 1 … (K – J).

Posterior determinant. When the information summary is the posterior determinant and the prior distribution is normal, the item information u_i for available item z_i is equal to

$$u_i = \det\left(\boldsymbol{S}_{z_i} + \boldsymbol{\Psi}^{-1} + \sum_{k=o_1}^{o_J} \boldsymbol{S}_k\right),$$

for i = 1 … (K – J). When the prior distribution is uniform, the posterior determinant is equal to the determinant.

Trace. When the information summary is the trace, the item information u_i for available item z_i is equal to

$$u_i = \text{trace}\left(\boldsymbol{S}_{z_i} + \sum_{k=o_1}^{o_J} \boldsymbol{S}_k\right),$$

for i = 1 … (K – J).

Posterior trace. When the information summary is the posterior trace and the prior distribution is normal, the item information u_i for available item z_i is equal to

$$u_i = \text{trace}\left(\boldsymbol{S}_{z_i} + \boldsymbol{\Psi}^{-1} + \sum_{k=o_1}^{o_J} \boldsymbol{S}_k\right),$$

for i = 1 … (K – J). When the prior distribution is uniform, the posterior trace is equal to the trace.

Posterior expected Kullback-Leibler information. For the information summary posterior expected Kullback-Leibler information, let $\boldsymbol{\Lambda}$ be a 21 × 1 matrix containing hypothetical $\boldsymbol{\theta}$ values equal to [-4, -3.6, …, 3.6, 4] (1 dimension), or a $9^Q \times Q$ matrix with each row containing a unique combination of the hypothetical $\boldsymbol{\theta}$ values contained by Q vectors (>1 dimensions). With normal prior, each of the Q vectors is equal to [-4, -3, …, 3, 4]. With uniform prior, each of the Q vectors is equal to an equally spaced sequence ranging from lower to upper bound. Also, let $p(\boldsymbol{\theta} = \Lambda_{j,*})$ be the posterior probability that $\boldsymbol{\theta}$ is equal to $\Lambda_{j,*}$, given the responses to the administered items, $\eta_{k,h}$ the probability of scoring in category h on item k given that $\boldsymbol{\theta}$ is equal to $\Lambda_{j,*}$, and $p_{k,h}$ the probability of scoring in category h on item k given that $\boldsymbol{\theta}$ is equal to the expected a posteriori estimate of $\boldsymbol{\theta}$. Then the item information u_i for available item z_i is equal to

$$u_i = \sum_{j=1}^{7^Q} \left[p(\boldsymbol{\theta} = \Lambda_{j,*}) \sum_{k} \sum_{h=1}^{m_k+1} p_{k,h} \left(\log[p_{k,h}] - \log[\eta_{k,h}] \right) \right],$$

where the sum over k consists of all administered items and available item z_i , for i = 1 … (K – J).

Select maximum information item

If no constraints are added, the item with maximum information is selected from the available items that load on uncompleted dimensions (i.e., dimensions for which the variance target has not been reached yet). If constraints are added, the item with maximum information within the constraints is selected from the available items. These constraints are satisfied using linear programming.

In the linear programming model, the objective function coefficients are the summarized item information values with zeros for administered items, and the vector that is to be determined – the outcome variable – is forced to consist of zeros and ones. The constraint coefficients matrix, with number of columns equal to the number of items in the item bank, always consists of two default rows and at least one additional row specified by the user. The first default row is equal to **1**, of which the inner product with the outcome variable should equal the maximum number of items allowed. The other default row contains ones for administered items and zeros for available items, of which the inner product with the outcome variable should equal the number of administered items. Additional rows contain item characteristics for each added constraint (e.g., containing ones for items that belong to a certain class, like depression, and zeros elsewhere, or containing the estimated completion time for each item), of which the inner product with the outcome variable should obey a rule defined by the user (e.g., it should be smaller than 5, or between 5 and 10). Based on the objective function and constraints, the outcome variable is determined which contains ones for items that can be selected and zeros for items that should not be selected. The item with maximum information is then selected from the available items that can be selected according to the constraints.

If there is more than one item with the maximum value, one of them is selected at random.

Update theta estimate

After the first burn-in items have been administered, the theta estimate is updated each time a new item has been administered. The new estimate of theta is computed either by maximizing the likelihood of theta, maximizing the posterior distribution, or computing the expected a posteriori estimate. The prior distribution may be normal or uniform.

Maximization of the likelihood without specified bounds, and the posterior distribution with normal prior, is performed using the nlm() function (R Core Team, 2016). Maximization of the likelihood with specified bounds and the posterior distribution with uniform prior is performed with the constrOptim() function (R Core Team, 2016). The expected a posteriori estimate $\hat{\theta}_{EAP}$ is computed via Riemannsum, with grid points adapted to the current estimate and variance. With normal prior, it can also be computed via multi dimensional Gauss-Hermite quadrature using the MultiGHQuad R package (Kroeze, 2015).

Just like the estimate of theta, the accompanying covariance matrix of theta is updated with each newly administered item. For the maximum likelihood estimate and the maximum a posteriori estimate with uniform prior, the covariance matrix is computed as

$$Cov(\boldsymbol{\theta}) = \left[\sum_{k=o_1}^{o_J} \boldsymbol{S}_k\right]^{-1},$$

with parameters defined as in the previous sections. For the maximum a posteriori estimate with normal prior, the covariance matrix is computed as

$$Cov(\boldsymbol{\theta}) = \left[\boldsymbol{\Psi}^{-1} + \sum_{k=o_1}^{o_J} \boldsymbol{S}_k\right]^{-1},$$

with parameters defined as in the previous sections. For the expected a posteriori estimate, the covariance matrix is again computed via Riemannsum or, in case of normal prior, via Gauss-Hermite quadrature using the MultiGHQuad R package (Kroeze, 2015).

References

Chang, H.-H, & Ying, Z. (1996). A Global Information Approach to Computerized Adaptive Testing. Applied Psychological Measurement, *20*(3), 213 – 229. doi: 10.1177/014662169602000303.

Kroeze, K. A. (2015). MultiGHQuad: Multidimensional Gauss-Hermite Quadrature. R package version 1.1.

Mulder, J., & van der Linden, W. J. (2010). Multidimentional Adaptive testing with Kullback-Leibler information item selection. In W. J. van der Linden & C. A. W. Glas (Eds.), Elements of adaptive testing (pp. 77–101). New York: Springer.

Muraki, E. (1992). A Generized Partial Credit Model: Application of an EM Algorithm. Applied Psychological Measurement, *16*(2), 159 – 176. Doi: 10.1177/014662169201600206.

R Core Team (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Samejima, F. (1970). Estimation of latent trait ability using a response pattern of graded scores. Psychometrika, 35(1), 139 – 139. Doi: 10.1007/BF02290599.

Segall, D. O. (1996). Multidimensional adaptive testing. *Psychometrika*, *61*(2), 331–354. doi:10.1007/BF02294343.

Segall, D. O. (2000). Principles of multidimensional adaptive testing. In W. J. van der Linden & en C. A. W. Glas (Eds.), Computerized adaptive testing: Theory and practice (pp. 53 – 74). Dordrecht: Kluwer Academic Publishers.

Tutz, G. (1986). Bradley-Terry-Luce model with an ordered response. Journal of Mathematical Psychology, *30*(1), 306 – 316. doi: 10.1016/0022-2496(86)90034-9.

Van der Linden, W. J. (1999). Multidimensional Adaptive Testing with a Minimum Error-Variance Criterion. *Journal of Educational and Behavioral Statistics*, *24*(4), 398–412. doi:10.3102/10769986024004398.

Van der Linden, W. J. (2000). Constrained adaptive testing with shadow tests. In W. J. van der Linden & C. A. W. Glas (Eds.), Computerized adaptive testing: Theory and practice (pp. 27-52). Dordrecht, the Netherlands: Kluwer Academic Publishers.

Wang, C., Chang, H.-H., & Boughton, K. A. (2010). Kullback-Leibler Information and Its Applications in Multi-Dimensional Adaptive Testing. *Psychometrika*, *76*(1), 13–39. doi:10.1007/s11336-010-9186-0.

Appendix A: Probability Models, Likelihoods, Posteriors, and Derivatives

The probability models for scoring in a certain item category given a latent trait, the likelihood function and posterior distribution of the latent trait, and their first and second derivatives are presented below for the three-parameter logistic model (Segall, 2000), generalized partial credit model (Muraki, 1992), sequential model (Tutz, 1986), and the graded response model (Samejima, 1970). In all formulas, let $\mathbf{A}$ be a K (number of items) × Q (number of dimensions) matrix containing the item discrimination parameters, $\mathbf{B}$ a $K \times M$ (maximum number of item steps) matrix containing the difficulty parameters, $\mathbf{c}$ a vector of length K containing the guessing parameter for each item k, $\mathbf{m}$ a vector of length K containing the number of item steps for each item k, $\mathbf{r}$ a vector of length K containing the item responses (ranging from 1 to the number of categories), $\boldsymbol{\theta}$ a vector of length Q containing the latent trait values, $\boldsymbol{\mu}$ and $\boldsymbol{\Psi}$ the mean and covariance matrix of the normal prior on $\boldsymbol{\theta}$, respectively, and $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ the lower and upper bounds of the uniform prior on $\boldsymbol{\theta}$, respectively.

Three-parameter logistic model

The three parameter logistic model always assumes two answer categories, and hence $\mathbf{B}$ has only a single column of difficulty parameters. For the three-parameter logistic model, the probability $p_{k,2}$ that a score on item k falls in category 2 is equal to

$$p_{k,2} = \mathrm{c}_k + \frac{(1 - \mathrm{c}_k)}{1 + \exp\left[-\sum_q a_{k,q}(\theta_q - b_k)\right]}.$$

The probability $p_{k,1}$ that a score on item k falls in category 1 is equal to 1 – $p_{k,2}$. The likelihood function of θ is equal to

$$L(\boldsymbol{\theta}) = \prod_k p_{k,\mathrm{r}_k}.$$

When the estimation is Bayesian (maximum a posteriori or expected a posteriori estimation), the likelihood function is multiplied with the prior distribution of $\boldsymbol{\theta}$ in order to obtain the posterior distribution $p(\boldsymbol{\theta})$:

$$p(\boldsymbol{\theta}) \propto L(\boldsymbol{\theta}) \exp\left[\frac{-(\boldsymbol{\theta} - \boldsymbol{\mu})^T \boldsymbol{\Psi}^{-1}(\boldsymbol{\theta} - \boldsymbol{\mu})}{2}\right]$$

when the prior is normal and

$$p(\boldsymbol{\theta}) \propto L(\boldsymbol{\theta})$$

with $\boldsymbol{\alpha} \leq \boldsymbol{\theta} \leq \boldsymbol{\beta}$, when the prior is uniform. The first derivative of the log likelihood function $\mathbf{d}(\log L(\boldsymbol{\theta}))$ is equal to

$$\mathbf{d}(\log L(\boldsymbol{\theta})) = \left[\frac{(\mathbf{p}_{*,2} - \mathbf{c}) * (\mathbf{r} - \mathbf{p}_{*,2})}{(\mathbf{1} - \mathbf{c}) * \mathbf{p}_{*,2}}\right]^T \mathbf{A},$$

where * denotes elementwise multiplication, the division is elementwise too, and $\mathbf{1}$ is a vector of ones of length K. The first derivative of the log posterior distribution $\mathbf{d}(\log p(\boldsymbol{\theta}))$ is equal to

$$\mathbf{d}(\log p(\boldsymbol{\theta})) = \mathbf{d}(\log L(\boldsymbol{\theta})) - (\boldsymbol{\Psi}^{-1}\boldsymbol{\theta})^T + (\boldsymbol{\Psi}^{-1}\boldsymbol{\mu})^T$$

when the prior is normal and

$$\mathbf{d}(\log p(\boldsymbol{\theta})) = \mathbf{d}(\log L(\boldsymbol{\theta}))$$

with $\boldsymbol{\alpha} \leq \boldsymbol{\theta} \leq \boldsymbol{\beta}$, when the prior is uniform. For the second derivative, let the vector $\mathbf{g}$ be equal to

$$\mathbf{g} = \frac{\mathbf{p}_{*,1} * (\mathbf{p}_{*,2} - \mathbf{c}) * (\mathbf{c} * \mathbf{r} - \mathbf{p}_{*,2} * \mathbf{p}_{*,2})}{\mathbf{p}_{*,2} * \mathbf{p}_{*,2} * (\mathbf{1} - \mathbf{c}) * (\mathbf{1} - \mathbf{c})},$$

where the division is elementwise. Then the second derivative of the log likelihood function $\mathbf{D}(\log L(\boldsymbol{\theta}))$ is equal to

$$\mathbf{D}(\log L(\boldsymbol{\theta})) = \sum_k \mathbf{a}_{k,*}\mathbf{a}_{k,*}^T\, g_k = \mathbf{A}^T \operatorname{diag}(\mathbf{g})\mathbf{A},$$

where $\operatorname{diag}(\mathbf{g})$ denotes the diagonal matrix with $\mathbf{g}$ as its diagonal. The second derivative of the log posterior distribution $\mathbf{D}(\log p(\boldsymbol{\theta}))$ is equal to

$$\mathbf{D}(\log p(\boldsymbol{\theta})) = \mathbf{D}(\log L(\boldsymbol{\theta})) - \boldsymbol{\Psi}^{-1}$$

when the prior is normal and

$$\mathbf{D}(\log p(\boldsymbol{\theta})) = \mathbf{D}(\log L(\boldsymbol{\theta}))$$

with $\boldsymbol{\alpha} \leq \boldsymbol{\theta} \leq \boldsymbol{\beta}$, when the prior is uniform.

Generalized partial credit model

Let $\mathbf{G}$ be a $K \times M$ matrix, with elements equal to

$$g_{k,i} = \exp[(i-1)\,(\mathbf{a}_{k,*}^T\boldsymbol{\theta}) - b_{k,i}],$$

for $i = 2 \ldots (m_k + 1)$. Then the probability $p_{k,i}$ that a score on item k falls in category i is equal to:

$$p_{k,i} = \frac{g_{k,i}}{1 + \sum_{j=2}^{m_k+1} g_{k,j}},$$

for $i = 2 \ldots (m_k + 1)$. The probability $p_{k,1}$ that a score on item k falls in category 1 is equal to:

$$p_{k,1} = 1 - \sum_{i=2}^{m_k+1} p_{k,i} \,.$$

The likelihood function of $\boldsymbol{\theta}$ is equal to

$$L(\boldsymbol{\theta}) = \prod_k p_{k,r_k} \,.$$

When the estimation is Bayesian, the likelihood function is multiplied with the prior distribution of $\boldsymbol{\theta}$ in order to obtain the posterior distribution, just as for the three parameter logistic model. For the first derivative, let $\mathbf{y}$ be a vector of length K, with element k equal to

$$y_k = r_k - \sum_{i=1}^{m_k+1} i\, p_{k,i} \,.$$

Then the first derivative of the log likelihood function $\mathbf{d}(\log L(\boldsymbol{\theta}))$ is equal to

$$\mathbf{d}(\log L(\boldsymbol{\theta})) = \mathbf{y}^T \mathbf{A} \,.$$

The first derivative of the log posterior distribution $\mathbf{d}(\log p(\boldsymbol{\theta}))$ is equal to

$$\mathbf{d}(\log p(\boldsymbol{\theta})) = \mathbf{d}(\log L(\boldsymbol{\theta})) - (\boldsymbol{\Psi}^{-1}\boldsymbol{\theta})^T + (\boldsymbol{\Psi}^{-1}\boldsymbol{\mu})^T$$

when the prior is normal and

$$\mathbf{d}(\log p(\boldsymbol{\theta})) = \mathbf{d}(\log L(\boldsymbol{\theta}))$$

with $\boldsymbol{\alpha} \leq \boldsymbol{\theta} \leq \boldsymbol{\beta}$, when the prior is uniform. For the second derivative, let $\mathbf{h}$ be a vector of length K, with element k equal to

$$h_k = -\sum_{i=1}^{m_k+1} \left[\left(i\, p_{k,i}\right) \left(i - \sum_{j=1}^{m_k+1} \left(j\;\, p_{k,j}\right) \right) \right].$$

Then the second derivative of the log likelihood function $\mathbf{D}(\log L(\boldsymbol{\theta}))$ is equal to

$$\mathbf{D}(\log L(\boldsymbol{\theta})) = \sum_k \mathbf{a}_{k,*}\mathbf{a}_{k,*}^T\, h_k = \mathbf{A}^T \mathrm{diag}(\mathbf{h})\mathbf{A}.$$

The second derivative of the log posterior distribution $\mathbf{D}(\log p(\boldsymbol{\theta}))$ is equal to

$$\mathbf{D}(\log p(\boldsymbol{\theta})) = \mathbf{D}(\log L(\boldsymbol{\theta})) - \boldsymbol{\Psi}^{-1}$$

when the prior is normal and

$$\mathbf{D}(\log p(\boldsymbol{\theta})) = \mathbf{D}(\log L(\boldsymbol{\theta}))$$

with $\boldsymbol{\alpha} \leq \boldsymbol{\theta} \leq \boldsymbol{\beta}$, when the prior is uniform.

Sequential model
Let $\mathbf{F}$ be a $K \times (M + 2)$ matrix, with elements equal to

$$f_{k,i} = \frac{\exp[\mathbf{a}_{k,*}^T \boldsymbol{\theta} - b_{k,i-1}]}{1 + \exp[\mathbf{a}_{k,*}^T \boldsymbol{\theta} - b_{k,i-1}]},$$

for $i = 2 \ldots (m_k + 1)$, the first column equal to $\mathbf{1}$, and the remaining elements equal to 0. Then the probability $p_{k,i}$ that a score on item k falls in category i is equal to:

$$p_{k,i} = \left(1 - f_{k,i+1}\right) \prod_{j=1}^{i} f_{k,j} ,$$

with $i = 1 \ldots (m_k + 1)$. The likelihood function of $\boldsymbol{\theta}$ is equal to

$$L(\boldsymbol{\theta}) = \prod_{k} p_{k,r_k} .$$

When the estimation is Bayesian, the likelihood function is multiplied with the prior distribution of $\boldsymbol{\theta}$ in order to obtain the posterior distribution, as before.

For the first derivative, let $\mathbf{s}$ be a vector of length K, with element k equal to

$$s_k = -f_{k,r_k+1} + \sum_{i=1}^{r_k} (1 - f_{k,i}) ,$$

Then the first derivative of the log likelihood function $\mathbf{d}(\log L(\boldsymbol{\theta}))$ is equal to

$$\mathbf{d}(\log L(\boldsymbol{\theta})) = \mathbf{s}^T \mathbf{A} .$$

The first derivative of the log posterior distribution $\mathbf{d}(\log p(\boldsymbol{\theta}))$ is equal to

$$\mathbf{d}(\log p(\boldsymbol{\theta})) = \mathbf{d}(\log L(\boldsymbol{\theta})) - (\boldsymbol{\Psi}^{-1}\boldsymbol{\theta})^T + (\boldsymbol{\Psi}^{-1}\boldsymbol{\mu})^T$$

when the prior is normal and

$$\mathbf{d}(\log p(\boldsymbol{\theta})) = \mathbf{d}(\log L(\boldsymbol{\theta}))$$

with $\boldsymbol{\alpha} \leq \boldsymbol{\theta} \leq \boldsymbol{\beta}$, when the prior is uniform. For the second derivative, let $\mathbf{v}$ be a vector of length K, with element k equal to

$$v_k = -f_{k,r_k+1}(1 - f_{k,r_k+1}) - \sum_{i=1}^{r_k} f_{k,i}\,(1 - f_{k,i})$$

Then the second derivative of the log likelihood function $\mathbf{D}(\log L(\boldsymbol{\theta}))$ is equal to

$$\mathbf{D}(\log L(\boldsymbol{\theta})) = \sum_k \mathbf{a}_{k,*}\mathbf{a}_{k,*}^T\, v_k = \mathbf{A}^T \mathrm{diag}(\mathbf{v})\mathbf{A}.$$

The second derivative of the log posterior distribution $\mathbf{D}(\log p(\boldsymbol{\theta}))$ is equal to

$$\mathbf{D}(\log p(\boldsymbol{\theta})) = \mathbf{D}(\log L(\boldsymbol{\theta})) - \boldsymbol{\Psi}^{-1}$$

when the prior is normal and

$$\mathbf{D}(\log p(\boldsymbol{\theta})) = \mathbf{D}(\log L(\boldsymbol{\theta}))$$

with $\boldsymbol{\alpha} \leq \boldsymbol{\theta} \leq \boldsymbol{\beta}$, when the prior is uniform.

Graded response model

Let $\mathbf{F}$ be defined as for the sequential model. Then the probability $p_{k,i}$ that a score on item k falls in category i is equal to:

$$p_{k,i} = f_{k,i} - f_{k,i+1}\,,$$

with $i = 1 \ldots (m_k + 1)$. The likelihood function of $\boldsymbol{\theta}$ is equal to

$$L(\boldsymbol{\theta}) = \prod_k p_{k,r_k}\,.$$

When the estimation is Bayesian, the likelihood function is multiplied with the prior distribution of $\boldsymbol{\theta}$ in order to obtain the posterior distribution, as before. For the first derivative, let $\mathbf{w}$ be a vector of length K, with element k equal to

$$w_k = 1 - f_{k,r_k} - f_{k,r_k+1}\,.$$

Then the first derivative of the log likelihood function $\mathbf{d}(\log L(\boldsymbol{\theta}))$ is equal to

$$\mathbf{d}(\log L(\boldsymbol{\theta})) = \mathbf{w}^T\mathbf{A}\,.$$

The first derivative of the log posterior distribution $\mathbf{d}(\log p(\boldsymbol{\theta}))$ is equal to

$$\mathbf{d}(\log p(\boldsymbol{\theta})) = \mathbf{d}(\log L(\boldsymbol{\theta})) - (\boldsymbol{\Psi}^{-1}\boldsymbol{\theta})^T + (\boldsymbol{\Psi}^{-1}\boldsymbol{\mu})^T$$

when the prior is normal and

$$\mathbf{d}(\log p(\boldsymbol{\theta})) = \mathbf{d}(\log L(\boldsymbol{\theta}))$$

with $\boldsymbol{\alpha} \leq \boldsymbol{\theta} \leq \boldsymbol{\beta}$, when the prior is uniform. For the second derivative, let $\mathbf{x}$ be a vector of length K, with element k equal to

$$x_k = -(f_{k,r_k}\left(1 - f_{k,r_k}\right) + f_{k,r_k+1}\left(1 - f_{k,r_k+1}\right)) .$$

Then the second derivative of the log likelihood function $\mathbf{D}(\log L(\boldsymbol{\theta}))$ is equal to

$$\mathbf{D}(\log L(\boldsymbol{\theta})) = \sum_k \mathbf{a}_{k,*}\mathbf{a}_{k,*}^T\, x_k = \mathbf{A}^T \operatorname{diag}(\mathbf{x})\mathbf{A} .$$

The second derivative of the log posterior distribution $\mathbf{D}(\log p(\boldsymbol{\theta}))$ is equal to

$$\mathbf{D}(\log p(\boldsymbol{\theta})) = \mathbf{D}(\log L(\boldsymbol{\theta})) - \boldsymbol{\Psi}^{-1}$$

when the prior is normal and

$$\mathbf{D}(\log p(\boldsymbol{\theta})) = \mathbf{D}(\log L(\boldsymbol{\theta}))$$

with $\boldsymbol{\alpha} \leq \boldsymbol{\theta} \leq \boldsymbol{\beta}$, when the prior is uniform.