



*Research
Report*

**User's Guide for
SCORIGHT (Version 3.0): A
Computer Program for
Scoring Tests Built of
Testlets Including a Module
for Covariate Analysis**

**Xiaohui Wang
Eric T. Bradlow
Howard Wainer**

**User's Guide for SCORIGHT (Version 3.0):
Computer Program for Scoring Tests Built of Testlets
Including a Module for Covariate Analysis**

Xiaohui Wang

University of Virginia, Charlottesville

Eric T. Bradlow

The Wharton School of the University of Pennsylvania, Philadelphia

Howard Wainer

The National Board of Medical Examiners, Philadelphia, PA

January 2005

As part of its educational and social mission and in fulfilling the organization's nonprofit charter and bylaws, ETS has and continues to learn from and also to lead research that furthers educational and measurement research to advance quality and equity in education and assessment for all users of the organization's products and services.

ETS Research Reports provide preliminary and limited dissemination of ETS research prior to publication. To obtain a PDF or a print copy of a report, please visit:

www.ets.org/research/contact.html



Abstract

SCORIGHT is a very general computer program for scoring tests. It models tests that are made up of dichotomously or polytomously rated items or any kind of combination of the two through the use of a generalized item response theory (IRT) formulation. The items can be presented independently or grouped into clumps of allied items (testlets) or in any combination of the two. When there are testlets, the program assesses the degree of local dependence and adjusts the estimates accordingly. The estimation is accomplished within a fully Bayesian framework using Markov chain Monte Carlo procedures, which allows the easy calculation of many important characteristics of the scores that are not available with other methods. The current version of SCORIGHT, version 3.0, includes a new module that allows the user to include covariates in the analysis.

Key words: Markov chain Monte Carlo (MCMC), Bayesian, testlets, dichotomous, polytomous, item response theory (IRT)

Acknowledgments

Xiaohui Wang is an assistant professor in the Department of Statistics, University of Virginia; Eric T. Bradlow is an associate professor of marketing and statistics and the academic director of Wharton Small Business Development Center, The Wharton School of the University of Pennsylvania; and Howard Wainer is the distinguished research scientist at the National Board of Medical Examiners. SCORIGHT was prepared with support from the research allocation of Educational Testing Service and is part of a long-term project on testlet response theory that has been supported at various times by the research budgets of the Graduate Record Examinations[®], the Test of English as a Foreign Language[™], and the Joint Staff Research & Development Committee. We are grateful for the opportunity to acknowledge their support. We would also like to express our gratitude to David Thissen, who provided us with advice, and the North Carolina Department of Public Instruction, which provided us with the data from the North Carolina Test of Computer Skills that appeared in earlier versions of the SCORIGHT manual. This version of SCORIGHT includes a new module developed by and owned by the National Board of Medical Examiners (NBME). The NBME module permits the inclusion of covariates in the test scoring model. Last, we would like to thank Melissa Margolis, Catherine McClellan, and Kim Fryer for their careful and thoughtful reading of an earlier version of this manual.

1 Introduction

Since the introduction of item response theory (IRT) as a primary scoring method for standardized tests more than 30 years ago, people have been questioning the fundamental assumptions that underlie its foundation. One of the most basic tenets in IRT is that, given an individual's proficiency (θ_i), his or her item responses are *conditionally independent*. While this assumption leads to more tractable answers and computation and may be approximately true when the items are carefully written (although sequence effects put that in question), current trends in educational assessment tools make the conditional independence assumption increasingly impractical.

Specifically, as the need for richer and more highly diagnostic forms of assessment arise, test writers have moved towards tests composed either wholly or in part of testlets (Wainer & Kiely, 1987). In short, a testlet is a subset of items (one or more) that, considered as a whole, work in a unified way to measure the construct of interest. A common form of testlet is a set of items generated from a *single* stimulus (e.g., a reading comprehension passage). In a testlet, one could easily imagine that items' behavior is more highly correlated than pure unidimensional proficiency would predict as in, for example, the misreading of the passage (yielding the effect of all items within the testlet being answered apropos to a lower proficiency than presumed under a unidimensional model), a common subarea expertise, and so on. Much work has been done in this area under the name of appropriateness measurement (Levine & Drasgow, 1988), and nonparametric approaches to detecting violations of conditional independence have been proposed (Stout, 1987; Zhang & Stout, 1999).

Our research has been to step beyond detection and instead to use a parametric approach to actually model the violations of conditional independence due to testlets. We modify standard IRT models to include a random effect that is common to all item responses within a testlet but that differs across testlets. In this manner, the generalized IRT model allows fitted item responses given by an individual in a testlet to be more highly correlated than his or her corresponding responses across testlets. In addition, our parametric approach is Bayesian in that we specify prior distributions that will allow

for sharing of information across persons, testlets, and items. This parametric approach (briefly reviewed in Section 2) was first described in Bradlow, Wainer, and Wang (1999) and was subsequently extended in Wainer, Bradlow, and Du (2000) and Wang, Bradlow, and Wainer (2002). The program SCORIGHT version 3, described here, is based on Wang et al. (2002) but is extended in a number of important ways.

Specifically, SCORIGHT is a computer program designed to facilitate analysis of item response data that may contain testlets. This program is completely general in that it can handle data composed of binary or polytomously scored items that are independent or nested within testlets. More specifically, the model used for the binary data is the three-parameter logistic (3PL) model (Birnbaum, 1968), and that used for the ordinal data is Samejima's (1969) ordinal response model. In this manner, our program can be adjusted to use the standard two-parameter logistic (2PL) and ordinal models that are often fit by commercial software (e.g., BILOG, MULTILOG), differing only by a Bayesian structure, outlined here.¹

The remainder of this manual is divided into four main sections. Section 2 (below) presents an explicit description of the model that is fit to the data. Section 3 (page 7) contains specific instructions on how to use the software. Section 4 (page 25) presents examples and a description of the model output files, Section 5 (page 27) details the output files, and the final section (page 38) provides a worked-out example.

2 Models

This section describes the base probability models that are used. As the model is Bayesian in nature and can be used for both binary and polytomous items, this requires one to specify the following probability models:

1. the model for binary data,
2. the polytomous data model, and
3. the prior distributions for all parameters governing (1) and (2).

2.1 Model Specification

The models that are used in this program have two basic probability kernels that allow one to encompass both dichotomous and polytomous items. For dichotomous items, we utilized the 3PL model:

$$P(Y_{ij} = 1) = c_j + (1 - c_j)\text{logit}^{-1}(t_{ij}),$$

and for polytomous items, we utilized the ordinal response model introduced by Samejima (1969):

$$P(Y_{ij} = r) = \Phi(d_r - t_{ij}) - \Phi(d_{r-1} - t_{ij}),$$

where Y_{ij} is the response of examinee i on item j , c_j is the lower asymptote (guessing parameter) for dichotomous item j , d_r are the latent cutoffs (score thresholds) for the polytomous items $\text{logit} = \log(x/(1 - x))$, Φ is the normal cumulative density function, and t_{ij} is the latent linear predictor of score. The two-parameter dichotomous items are a special case of the 3PL model with $c_j = 0$. In this special case,

$$P(Y_{ij} = 1) = \text{logit}^{-1}(t_{ij}).$$

SCORIGHT models the extra dependence due to testlets by extending the linear score predictor t_{ij} from its standard form:

$$t_{ij} = a_j(\theta_i - b_j),$$

where a_j , b_j , and θ_i have their standard interpretations as item slope, item difficulty, and examinee proficiency, to:

$$t_{ij} = a_j(\theta_i - b_j - \gamma_{id(j)})$$

with $\gamma_{id(j)}$ denoting the testlet effect (interaction) of item j with person i that is nested within testlet $d(j)$. The extra dependence of items within the same testlet (for a given examinee) is modeled in this manner as both would share the effect $\gamma_{id(j)}$ in their score predictor. By definition, $\gamma_{id(j)} = 0$ for all independent items. Thus, to sum up the model extension here, it is the set of parameters $\gamma_{id(j)}$ that represents the difference between this model and standard approaches.

In order to combine all information across examinees, items, and testlets, a hierarchical Bayesian framework is introduced into the model. The following prior distributions for parameters $\Lambda_1 = \{h_j, b_j, q_j, \theta_i, \gamma_{id(j)}\}$ are asserted:

$$\begin{pmatrix} h_j \\ b_j \\ q_j \end{pmatrix} \sim N_3 \left(\begin{pmatrix} X_j^h \beta_h^{(3)} \\ X_j^b \beta_b^{(3)} \\ X_j^q \beta_q^{(3)} \end{pmatrix}, \begin{pmatrix} (\sigma_h^{(3)})^2 & \rho_{hb}^{(3)} \sigma_h^{(3)} \sigma_b^{(3)} & \rho_{hq}^{(3)} \sigma_h^{(3)} \sigma_q^{(3)} \\ \rho_{hb}^{(3)} \sigma_h^{(3)} \sigma_b^{(3)} & (\sigma_b^{(3)})^2 & \rho_{bq}^{(3)} \sigma_b^{(3)} \sigma_q^{(3)} \\ \rho_{hq}^{(3)} \sigma_h^{(3)} \sigma_q^{(3)} & \rho_{bq}^{(3)} \sigma_b^{(3)} \sigma_q^{(3)} & (\sigma_q^{(3)})^2 \end{pmatrix} \right) = N(\mu_{3PL}, \Sigma_{3PL})$$

for the 3PL binary items,

$$\begin{pmatrix} h_j \\ b_j \end{pmatrix} \sim N_2 \left(\begin{pmatrix} X_j^h \beta_h^{(2)} \\ X_j^b \beta_b^{(2)} \end{pmatrix}, \begin{pmatrix} (\sigma_h^{(2)})^2 & \rho_{hb}^{(2)} \sigma_h^{(2)} \sigma_b^{(2)} \\ \rho_{hb}^{(2)} \sigma_h^{(2)} \sigma_b^{(2)} & (\sigma_b^{(2)})^2 \end{pmatrix} \right) = N(\mu_{2PL}, \Sigma_{2PL})$$

for the 2PL binary items,

$$\begin{pmatrix} h_j \\ b_j \end{pmatrix} \sim N_2 \left(\begin{pmatrix} X_j^h \beta_h^{(p)} \\ X_j^b \beta_b^{(p)} \end{pmatrix}, \begin{pmatrix} (\sigma_h^{(p)})^2 & \rho_{hb}^{(p)} \sigma_h^{(p)} \sigma_b^{(p)} \\ \rho_{hb}^{(p)} \sigma_h^{(p)} \sigma_b^{(p)} & (\sigma_b^{(p)})^2 \end{pmatrix} \right) = N(\mu_{poly}, \Sigma_{poly})$$

for the polytomous items, and

$$\begin{aligned} \theta_i &\sim N(W_i \lambda, 1) \\ \gamma_{id(j)} &\sim N(0, \sigma_{d(j)}^2), \end{aligned}$$

where $h_j = \log(a_j)$, $q_j = \log(c_j/(1 - c_j))$, and X_j^h, X_j^b , and X_j^q are covariates associated with the item parameters. That is, the parameters are assumed to be distributed normally across the three different populations of items and drawn from three different distributions; one each for the 3PL binary items, the 2PL binary items, and the polytomous items. Furthermore, covariates are brought into the model in a natural way via the mean of the prior distribution of the item parameters (h_j, b_j, q_j) and the ability parameters θ_i , where the β s and λ (as in standard regressions) denote the covariate slopes. Note that the variance of the distribution for θ and the mean of the distribution for testlet effects $\gamma_{id(j)}$ are fixed to identify the model. Furthermore, if there are covariates W_i for θ , the W_i will have no intercept and be centered at 0 in order to identify the model.

To complete the model specification, a set of hyperpriors for parameters

$$\Lambda_2 = \{\lambda, \beta_h^{(3)}, \beta_b^{(3)}, \beta_q^{(3)}, \Sigma_{3PL}, \beta_h^{(2)}, \beta_b^{(2)}, \Sigma_{2PL}, \beta_h^{(p)}, \beta_b^{(p)}, \Sigma_{Poly}, \sigma_{d(j)}^2\}$$

is added to reflect the uncertainty in their values. The distributions for these parameters were chosen out of convenience as conjugate priors to Λ_1 . For the distribution of λ ,

$$\lambda \sim N(0, \sigma_\lambda^2 \mathbf{I}_m),$$

where $\sigma_\lambda = 5$ and \mathbf{I}_m is the identity matrix with dimension equal to m . For the distribution of coefficients,

$$\beta_h^{(3)} \sim \text{MVN}(0, V_a),$$

$$\beta_b^{(3)} \sim \text{MVN}(0, V_b), \text{ and}$$

$$\beta_q^{(3)} \sim \text{MVN}(0, V_q),$$

where $|V_a|^{-1} = |V_b|^{-1} = |V_q|^{-1}$ are set to 0 to give a noninformative prior. Similarly,

$$\beta_h^{(2)} \sim \text{MVN}(0, V_a) \text{ and}$$

$$\beta_b^{(2)} \sim \text{MVN}(0, V_b),$$

where $|V_a|^{-1} = |V_b|^{-1} = 0$ for the 2PL binary cases, and

$\beta_h^{(p)} \sim \text{MVN}(0, V_a)$ and

$\beta_b^{(p)} \sim \text{MVN}(0, V_b)$,

where $|V_a|^{-1} = |V_b|^{-1} = 0$ for the polytomous cases. Slightly informative hyperpriors for Σ_{3PL} , Σ_{2PL} , and Σ_{Poly} are used:

$$\Sigma_{3PL} \sim \text{Inv-Wishart}(3, M_3^{-1}),$$

$$\Sigma_{2PL} \sim \text{Inv-Wishart}(2, M_2^{-1}), \text{ and}$$

$$\Sigma_{Poly} \sim \text{Inv-Wishart}(2, M_2^{-1}),$$

where

$$M_3 = \begin{pmatrix} \frac{1}{100} & 0 & 0 \\ 0 & \frac{1}{100} & 0 \\ 0 & 0 & \frac{1}{100} \end{pmatrix}$$

and

$$M_2 = \begin{pmatrix} \frac{1}{100} & 0 \\ 0 & \frac{1}{100} \end{pmatrix}.$$

If there are no covariates for the testlet effects, the distribution for $\sigma_{d(j)}^2$ is

$$\sigma_{d(j)}^2 \sim \text{Inv-}\chi_{g_\tau}^2$$

for every testlet and $g_\tau = \frac{1}{2}$. If covariates exist for the testlet variances, the common testlet effect variance distribution is modeled as a function of testlet covariates $Z_{d(j)}$. For example, testlet covariates may include the number of words in the testlet stimuli, the type of stimuli, and so on. In this manner, one will be able to explore the factors that lead to larger interdependence (correlation) among testlet item parameters. This can have great practical importance for test design. Note that since the scale of the IRT model is fixed by the unit variance of the ability distribution, values for testlet variances that are comparable in magnitude (0.5 and above) have been shown to impact the resulting inferences (Bradlow et al., 1999).

In the case with test covariates, we therefore utilize:

$$\log(\sigma_{d(j)}^2) \sim N(Z_{d(j)}\delta, \tau^2)$$

with associated hyperpriors for

$$\Lambda_3 = \{ \delta, \tau^2 \}$$

chosen as mostly noninformative: a prior distribution for δ as $\delta \sim \text{MVN}(0, V_\delta)$ where $|V_\delta|^{-1} \equiv 0$, and a slightly informative prior for τ , $\tau^2 \sim \text{Inv-}\chi_{g_\tau}^2$ where $g_\tau = \frac{1}{2}$.

2.2 Computation

To draw inferences under this Bayesian testlet model, samples from the posterior distribution of the parameters are obtained using Markov chain Monte Carlo (MCMC) techniques. Details of the techniques are presented in Wang et al. (2002). The relevant aspects of computation (from the user's perspective) for implementing the model are described in Section 3.

The model to be utilized in fitting the data is specified by the user. The choices for the dichotomous data are to fit a 2PL or 3PL model. For polytomous items, Samejima's model for graded responses is used.

3 How To Use It

The SCORIGHT program is run in a DOS environment. The user at the keyboard starts the program and proceeds to answer a series of questions by typing an appropriate response. The responses are used to determine the location of the input data files, the details of how SCORIGHT will be run, and the location of the output files.

On the following pages, the output from SCORIGHT is printed in a **monospaced font** while everything typed by the user from the keyboard is printed in **boldface**. We have written the manual in terms of a specific set of step-by-step instructions.

3.1 Step-by-Step User Instructions for SCORIGHT

Step 1: Start SCORIGHT

In the DOS window in the subdirectory where the SCORIGHT program is installed, type:

scoright.exe

and hit Enter to start SCORIGHT. The following will then appear on the screen:

```
This program estimates the proficiency, item parameters, and testlet effects
for both dichotomous and polytomous items that could be independent or
nested within testlets using the Gibbs sampler. To run this program, you
need to provide the following information.
```

```
Please enter the number of examinees and the number of items in your dataset
separated by at least one space:
```

Step 2: Enter the Number of Examinees and Items

Now the user must respond to the request by typing two numbers separated by spaces. Spaces could be a single space, multiple spaces, a single tab, multiple tabs, or even a return key. Regarding SCORIGHT, any number or type of spaces has the same meaning. The first number to be input is the total number of examinees and the second is the total number of items. Each examinee must have item responses for all of the items. If one examinee is given an item, his or her response should be how he or she answered this item. If some items are not assigned to an examinee, the response of this examinee under these items should be coded as N, which stands for not assigned. If there are nonignorable missing data, you should preprocess the data to accommodate a model for the missing data (e.g., impute values from an appropriate kind of missing data model); otherwise, SCORIGHT will treat the Ns as ignorably missing responses and also missing completely at random.

For example, if 1,000 examinees took an 80-item test, your response to the prompt

would be:

```
Please enter the number of examinees and the number of items in your dataset
separated by at least one space: 1000 80
```

SCORIGHT would interpret this to mean that there are a total of 1,000 examinees, each of whom has responded to as many as 80 test items. If the user enters anything other than two numbers separated by spaces, SCORIGHT will display an error message and ask the user to reenter the information.

Step 3: Enter the Number of Dichotomous Items Among All the Items

The next SCORIGHT prompt is:

```
Please enter the number of dichotomous items within the total 80 items:
```

Based on the total number of items entered in Step 2, SCORIGHT asks the user how many dichotomous items (including both 3PL and 2PL binary items) there are among all the items. If the user enters **0**, that means that all items are polytomous items. If the number the user enters is less than the total number of items in the analysis, SCORIGHT will then request (in Step 4) information about which items are to be treated as 2PL dichotomous, 3PL dichotomous, or polytomous, and the total number of categories for each polytomous item. If either the number of dichotomous items entered is greater than the total number of items in the test or there is any other wrong input (such as alphabetic characters), SCORIGHT will print an error message and ask the user to reenter the response until the input is consistent. For example, assume that of the total 80 items there are 60 dichotomous items: 40 3PL binary items and 20 2PL binary items. Therefore the user types **60** after the prompt:

```
Please enter the number of dichotomous items within the total 80 items: 60
```

Step 4 (Optional): Enter the Number of 2PL Binary Items

The next prompt will not appear if the response is 0 to the above prompt (Step 3). Otherwise, SCORIGHT will prompt you for the number of 2PL dichotomous items among the total number of dichotomous items:

Please enter the number of 2PL binary response items:

The user must prepare a file (described in Step 10) to indicate the type of each item: a 2PL dichotomous item, a 3PL dichotomous item, or a polytomous item. For example, if there are 20 2PL dichotomous items among the 60 dichotomous items, the user will type **20** after the prompt:

Please enter the number of 2PL binary response items: **20**

For the input corresponding to the current example, there are three different groups of items: $40=(60-20)$ 3PL dichotomous items, 20 2PL dichotomous items, and $20=(80-60)$ polytomous items.

Step 5: Enter the Number of Testlets

The next information SCORIGHT requests is:

Enter the total number of testlets in the test:

If there are no testlets (i.e., all the items in the test are independent), enter **0**. Otherwise, enter the number of testlets. For example, if there are 20 testlets within the dataset, type **20** following the prompt:

Enter the total number of testlets in the test: **20**

Step 6: Enter the Name/Path of the Data File

The next prompt requests that the user input the name of the file that contains the response data (outcome) matrix. Enter the whole name of the file including the drive name and all subdirectory names (i.e., the entire path). For example:

Enter the name of the file that contains the test data:

c:\subdirectory\test.dat

The name of the file is case sensitive, since the program is designed for the PC or Unix platforms. The input dataset is also required to have a specific format:

1. Each examinee's data record must be contained in one row with item responses recorded sequentially.
2. Each item response must occupy only one column in the file.
3. There should be no spaces between the item responses.
4. It is not necessary that the response to the first item starts in the first column of the record, only that all persons' responses begin in the same column.
5. If there are testlets, the item responses nested within each testlet must be ordered sequentially (clustered) in the dataset.
6. The responses for dichotomous items must be coded 1 for correct answers, 0 for incorrect.
7. For polytomous items, responses start from 1 and range to the highest category. For example, if a polytomous item has a total of four different responses, the responses on the data file should be 1, 2, 3, or 4. The model does not have any restriction about the total number of categories for polytomous items; however, the current version of SCORIGHT can only handle items with total categories equal to or less than 9. This was done to keep the format of input files consistent. We suggest recoding (collapsing) the data if there are more than nine categories for any polytomous item.

Ramsay (1973) has shown that under broad conditions very little information is lost recoding a continuous variable into seven categories.

For example:

```
ID0001 41011211110131001141430100011141114120411411111010411321131113111010413114011311
ID0002 11101101000010010111431100111010001131401100011000410441110001000000213001000100
ID0003 2111021111011110114143011101113001311041110011111411441031011110010413101001411
ID0004 20111211110110001121430110111130011130411101111100411431111011111010103001000311
```

8. For items that are not assigned to the examinee or those that you want to treat as ignorably missing, the responses should be coded as N.

Step 7: Enter the Beginning and Ending Column of the Test Data

The next prompt to appear after entering the information in Step 6 is:

```
Enter the starting and ending columns of the test scores
for the data file:  8 87
```

For example, if the data are as follows:

```
ID0001 41011211110131001141430100011141114120411411111010411321131113111010413114011311
ID0002 11101101000010010111431100111010001131401100011000410441110001000000213001000100
ID0003 2111021111011110114143011101113001311041110011111411441031011110010413101001411
ID0004 20111211110110001121430110111130011130411101111100411431111011111010103001000311
```

the beginning column would be 8 and the ending column would be 87 (indicating an 80-item test). The two numbers entered should be separated by spaces. Here SCORIGHT will check the user's input. If the number of the ending column minus the beginning column plus one is not equal to the total number of items input or if some other input is incorrect, SCORIGHT will print an error message and ask the user to reenter the information until the input is consistent.

Step 8: Enter the Beginning and Ending Columns for the Testlet Items

If the user has entered **0** following the prompt:

```
Enter the total number of testlets in the test:
```

the following prompts will not appear. Otherwise, the user has to provide information about the testlets' starting and ending columns.

For example, if the first two testlets consisted of three items each starting at the 28th and 31st columns of the dataset (Step 6), the user would type:

```
Enter the starting and ending columns of Testlet #1: 28 30
```

```
Enter the starting and ending columns of Testlet #2: 31 33
```

```
...
```

The user has to complete all information about each testlet until all testlet information has been entered. The number of prompts about testlets will correspond to the number input in Step 5 (number of testlets).

Step 9: Enter the Beginning and Ending Rows of the Dataset

The next SCORIGHT prompt is:

```
Enter the starting and ending rows of the test scores: 1 1000
```

This would indicate that the data begin at the top of the file (as you can see, this is not required) and continue to Row 1000 (indicating 1,000 examinees). If the number entered for the ending row minus the starting row plus one is not equal to the total number of examinees that the user input or is otherwise input incorrectly, SCORIGHT will print out an error message and prompt the user to reenter the information until the input is consistent.

Step 10 (Optional): Create an Information File About the Items

Except for the case in which all items are 3PL dichotomous items, the user has to provide information about each item's type through another file.

This file indicates which items are 3PL dichotomous, 2PL dichotomous, or polytomous by using one character: D for 3PL dichotomous items, 2 for 2PL dichotomous items, and P for polytomous items. The D, 2, or P must be located in the first column of each record of the file followed by at least one space and then the total number of categories for this item. If the item is dichotomous, the number of categories is 2. Each item occupies one row of the file, with the first item in the starting row, until all items are described. The following is an example of part of an item input file, *index*, in c:\subdirectory\

```
D 2
D 2
D 2
P 5
2 2
P 4
...
```

This indicates that the first three items on the test are 3PL dichotomous items, the fourth is a polytomous item with five categories, the fifth is a 2PL dichotomous item, and the sixth is a polytomous item with four categories.

Step 11 (Optional): Enter the Name/Path of the Item Information File

The user is prompted for the location of the item information file (described in Step 10):

```
Enter the name of the item information file: c:\subdirectory\index
```

It is the same requirement as in Step 6 for the item response data file.

Step 12: Enter the Name/Path Where the Output Files Should Be Stored

Because SCORIGHT generates many output files, the user may put all output files within any user-specified subdirectory:

```
Please enter the name of the subdirectory (include the last backslash)
where you want to put the analysis results, and make sure that there
is no subdirectory called ‘‘ch1,’’ ‘‘ch2,’’ ... under it:  c:\result\
```

Step 13: Enter the Number of Iterations for the Gibbs Sampler

SCORIGHT uses Gibbs sampling methods for inference. For the inferences to be valid, the Gibbs sampler must have converged. The convergence rate depends on the data and the initial values of the model parameters utilized. In this step, the user must specify the number of iterations to run. Typically, this would be at least 4,000 iterations, with the potential of a much larger number (Sinharay, 2004). One way to diagnose convergence of the sampler and hence the minimum number of iterations, which we strongly recommend, is to take the dataset and run multiple Markov chains with different starting values for the parameters. Convergence would be indicated by similar output across the chains. Section 5 provides a convergence diagnostic and describes the ability to run multiple chains.

```
Enter the number of needed iterations of sampling:  4000
```

Step 14: Enter the Number of Initial Draws To Be Discarded

As mentioned in Step 13, the sampler must converge before valid inferences under the model can be obtained. Therefore, iterations (and their draws) obtained prior to convergence should be discarded for estimating quantities of interest. In this step, the user specifies the number of initial iterations of draws to be discarded for inference purposes. For example:

Enter the number of draws to be discarded: **3000**

The draws after the initial 3,000 will be recorded as output and further estimation or computation will be calculated based on these. As mentioned, convergence should be assessed to decide when the number discarded is adequate.

Step 15: Enter the Number of Times the Posterior Draws Will Be Recorded

Since the posterior draws are highly correlated (autocorrelated through time via the construction of the Markov chain) it is often sensible to record only every k-th draw (i.e., to include some gaps). The virtue of this is that if the draws kept are essentially uncorrelated, the variance of estimators can be computed using the standard formula and does not require time series modeling. Thus, when recording the posterior draws, the user can specify how often the posterior draws are recorded for their output. For example, to keep every 11th draw:

Enter the size of the gap between posterior draws: **10**

Step 16: Enter the Number of Markov Chains You Want To Run

The current version of SCORIGHT allows the user to run multiple Markov chains. This facilitates the user's proficiency to detect whether the chains have converged or not. SCORIGHT utilizes the F-test convergence criterion of Gelman and Rubin (1993) in SCORIGHT.

How many chains do you want to run?

The user can answer any desired number. Of course, the more chains run the more running time SCORIGHT will use. For example, if the user types **3**, that means the user wants to run three chains, and one estimated set of results will be output that is based on the three runs. Commonly, people run from three to five chains in order to assess

convergence. Details regarding the convergence output are given in Section 5.

Step 17: Enter Initial Values for the Parameters

The convergence of SCORIGHT may depend in part on initial starting values for the parameter values. SCORIGHT will automatically select starting values for the user unless they are input. However, the user sometimes may have some information (perhaps from the output of a different program or a previous run of SCORIGHT) that suggests a reasonable set of starting values for either abilities θ_i or item parameters a_j, b_j, c_j . This part of SCORIGHT allows the user to utilize those values. In addition, SCORIGHT allows the user (if desired) to fix the values of part or all of the item parameters if those parameters are to be treated as fixed and known (although this is counter to the Bayesian nature of SCORIGHT); yet it aligns with some maximum marginal likelihood procedures. For example, a user may wish to fix the item guessing parameters (the c s) while allowing for estimation of the remaining item parameters. Similarly, SCORIGHT allows the user to fix the values of part or all of the examinees' proficiency, θ . This capability has obvious application in the equating of multiple test forms.

For CHAIN 1:

Do you want to input the initial values for item parameters a, b, and c?

If yes, enter 1, otherwise, enter 0:

If the user answers **1** to answer the above question, the user has to: (a) prepare the initial values for all item parameters, and (b) respond to the following prompt:

Please enter the name of the file that contains the initial values of the item parameters:

The format of the file that contains the initial values of all three item parameters $a, b,$ and c must take a specific form:

1. The file must contain all the initial values for item parameters $a, b,$ and $c;$ and should have as many rows as items.

2. For each row, there should be either 0 or 1 in the first five columns. Following the first five columns in each row, there should be the initial value for item parameter a , the initial value for item parameter b , and the initial value for item parameter c .
3. Each initial value has a column width of 12. If the item is either a 2PL dichotomous item or a polytomous item, the last 12 columns for the initial value of item parameter c should be empty.
4. In the first five columns, if the user inputs **1**, it indicates that the user wants to fix the value of this item parameter throughout the analysis. If the user inputs **0**, it indicates that the values for the item parameter are just the initial values. The file requires that each initial value for each item occupies 12 columns. It is not necessary to begin at the first column or that each one starts at the same column.

For example, the file that contains:

```

1           0.65           0.01           0.12
0           0.74          -1.43
...

```

would indicate that: (a) the first item is a 3PL dichotomous item; (b) the user wants to fix the value for this item as $a = 0.65$, $b = 0.01$, and $c = 0.12$; (c) the second item is either a 2PL dichotomous item or a polytomous item; (d) the user wants to give the initial values for item parameter a and b as $a = 0.74$ and $b = -1.43$; and (e) the user wants SCORIGHT to estimate the parameters of the second item.

If the user entered more than one chain in Step 16, the above prompts will repeat for every chain. If the user does not provide the initial values, SCORIGHT will randomly generate values for them.

Step 18: Enter Initial Values for the θ s

SCORIGHT will ask whether the user wants to input the initial values for the θ s:

For CHAIN 1:

Do you want to input the initial values for proficiency parameters theta?

If yes, enter 1, otherwise, enter 0:

If the user enters **1** for the above question, the user must also respond to the following prompt (if the user enters **0** to the above, the following prompt will not be shown):

Please enter the name of the file that contains the initial values of the theta parameter:

The file containing the initial values of θ should contain as many rows as there are examinees. For each row, either **0** or **1** should be entered in the first five columns, and following the first five columns should be the initial value for examinees' proficiency θ . Each initial value has a column width of 12. In the first five columns, if the user inputs **1**, that means the user wants to fix the value of this examinee's proficiency throughout the analysis. If the user inputs **0**, that means that the value for this examinee's proficiency starts out at the initial value, but is then estimated. The requirement is that each initial value for each examinee occupies 12 columns. If the user enters covariates for the θ s (as given in Step 20), the program will estimate the coefficients for the covariates based on the θ values estimated by SCORIGHT. However, if the user fixes all the θ values for the analysis, SCORIGHT will not estimate the coefficients for the covariates even if the user inputs the covariate information for the θ s.

Step 19: Enter Covariate Information for the Item a a Parameters as

SCORIGHT asks the user about the covariate information:

Do you have covariates for item parameter a (not including intercept)?

If Yes, enter 1, otherwise enter 0:

Since there can be at most three different groups of item parameters—3PL dichotomous, 2PL dichotomous, and polytomous items—among the total items, the user should type

1 after the above question if there are covariates for at least one group of the item a parameters. That is, according to Section 2.1, there are three regressions governing the a 's for 3PL, 2PL, and polytomous items. If any of them have covariates, the user should enter **1**.

If the user enters **0** following the first question, the next prompts (given below) will not be shown; subsequent prompts therefore depend on the information the user has input. For example, if all items are 3PL dichotomous items, SCORIGHT will request information about the covariates for the 3PL dichotomous item as only. The current example has all three cases, and hence the program will present all of the following prompts:

```
Please enter the total number of covariates for parameter a (without
intercept) of the 3PL binary response items:
```

```
Please enter the total number of covariates for parameter a (without
intercept) of the 2PL binary response items:
```

```
Please enter the total number of covariates for parameter a (without
intercept) of the polytomous items:
```

The number of covariates is the number of independent variables (which does not include an intercept). If the user enters **0** at the first prompt of this step (i.e., there are no covariates at all), SCORIGHT will give the estimated intercept only at the end (i.e., the estimated mean of item parameter as for each of the three types of items). If the number of covariates is 1 or more for any specific group, SCORIGHT will give the estimated coefficients (including the intercept) for this group at the end. For other groups, if the entered number of covariates is **0**, SCORIGHT will give the estimated mean of item parameter a for the corresponding group. Thus, in summary, SCORIGHT treats each of the 3PL, 2PL, and polytomous items as separate entities that might have covariates.

The format of the file that contains the information about the covariates for the a item parameters must take a specific form. This file, containing the covariates of item

parameter a , (a) should have as many rows as items and (b) should contain the covariate values for that item in each row of the file. For example, if only 3PL dichotomous items have covariates for parameter a and no covariates exist for 2PL dichotomous items, the user could enter **0.0** or nothing (empty row) for the corresponding 2PL dichotomous items in the file. In fact, it does not matter what the user enters for this 2PL dichotomous item; the information the user entered earlier will inform SCORIGHT that this item does not have any covariates. But it is important that a line exists (either empty or containing data) in order for SCORIGHT to read the information item by item. If an item does have covariates, each covariate's value has a fixed width of 12 characters. For example, a file that contains:

```
-0.4343    -1.2203
0
-0.2167
0
...
```

would indicate that the first covariate value for parameter a of Item 1 is -0.4343, and the second covariate value is -1.2203. For the second item, one can not tell whether there is only one covariate for this item or there are no covariates from this file. However, the earlier input will inform SCORIGHT how to interpret this information. For the third item, the file indicates that there is only one covariate. One can tell that Item 3 and Item 1 are not from the same group of items (either 2PL binary items, 3PL binary items, or polytomous items), since the items from the same group should have the same number of covariates. But one can not tell whether Item 3 is from the same or a different group as Item 2. As before, the earlier input will provide SCORIGHT with this information. The covariate values do not necessarily need to begin in Column 1, but there should be at least one space between two values. Assuming that the first item is a 3PL dichotomous item with two covariates, Items 2 and 4 are 2PL dichotomous items with no covariates, and Item 3 is a polytomous item with one covariate, the user must respond to the above prompts as follows:

Please enter the total number of covariates for parameter a (without intercept) of the 3PL binary response items: **2**

Please enter the total number of covariates for parameter a (without intercept) of the 2PL binary response items: **0**

Please enter the total number of covariates for parameter a (without intercept) of the polytomous items: **1**

After responding to the above prompts, SCORIGHT will request the file name that contains the covariates of item parameter a :

Please enter the name of the file that contains the covariate information of the parameter a : **c:\subdirect\acovariates**

SCORIGHT will then request similar information about the b parameters. If there are any 3PL dichotomous items in the test, SCORIGHT will request similar information about the c parameters. If there are polytomous items and/or 2PL dichotomous items in the test, the number of rows of the covariates for item parameter c_s should be same as the number of total items (including all items). Just enter **0** or leave the row blank for the corresponding 2PL dichotomous and polytomous items. That is, each item file should contain the same number of rows as items, yet some may be blank (or have 0s) if they are not of that particular type.

Step 20: Enter Covariate Information for θ

After requesting information about the covariates of item parameters a , b , and c , SCORIGHT will request information about covariates for θ .

Do you have covariates for parameter theta? If Yes, enter 1, otherwise enter 0:

If the user enters 1, SCORIGHT will present the following two prompts (otherwise, these two prompts will not be shown):

Please enter the total number of the covariates for parameter theta (without intercept):

Please enter the name of the file that contains the covariate information for parameter theta:

Because both the θ s and their covariates W are centered at 0 (to identify the model), there is no estimated intercept for the coefficients of covariates W . The format of the file that contains the covariate information of θ is same as the file containing the covariate information of item parameters. The file has as many rows as there are examinees. Even if some examinees' θ values are fixed, the user still needs to keep the place of the corresponding row (by entering any values for that row or leaving it blank). Each row contains the total number of covariates for that examinee, and each covariate occupies 12 bytes of space (12 columns). For example, one could respond to the above two prompts as follows:

Please enter the total number of the covariates for parameter theta (without intercept): 2

Please enter the name of the file that contains the covariate information for parameter theta: c:\subdirect\thetacovariates

based on the following file:

```
-0.4343    -1.2203
-0.5465    -0.7103
-0.2167     1.0209
-0.4237     0.3562
...
```

Step 21 (Optional): Entering Covariate Information for the Testlets

If the response in Step 5 is not **0**, there must be at least one testlet in the test. The user then has to respond to the following prompts. If the response in Step 5 is **0**, these questions and prompts will not be presented.

Do you have any covariates for the testlet effects (not including intercept)?

If YES, enter 1, otherwise enter 0:

Please enter the total number of covariates for the testlet effects
(without intercept):

Please enter the name of the file that contains the covariate information
for the testlet effect variances:

The above should be answered in exactly the same way as in Steps 19 and 20 (covariates for the item parameters and θ). The format of the file containing the covariate information for the testlet is also similar. The file has as many rows as the number of testlets. Each row contains the total number of covariates for that testlet (not including the intercept). Each covariate occupies 12 bytes (columns) of space.

This completes the user input for SCORIGHT.


```
CHAIN 1 Time after 31 cycles: Tue Oct 30 09:54:51 2003
```

```
CHAIN 1 Time after 41 cycles: Tue Oct 30 09:55:14 2003
```

```
...
```

For example, this output shows that the sampler is taking roughly 23 seconds per 10 iterations. This indicates that running 4,000 iterations would take 9,200 seconds or about 2 hours and 30 minutes. Faster processors will yield shorter run lengths. Of course, the simulated dataset above is very complicated; it has polytomous, 3PL, and 2PL dichotomous items and 20 testlets, and both item parameters and testlet effects have covariates. More experience with SCORIGHT is necessary to provide efficient rules of thumb for the number of iterations required, and until such experience is amassed, caution suggests going in the direction of too many iterations rather than too few. We commonly run 5,000 iterations, and this number has usually proven to be adequate.

After completing all iterations for Chain 1, SCORIGHT prints a message indicating the completion of the analysis. Note that the values indicated here for the number of iterations, number of initial iterations to be discarded, and so on, correspond to those values input in Steps 1 - 18 above. For example:

```
For Chain 1:
```

```
The Gibbs sampling of 4000 iterations is completed.
```

```
... End of running of CHAIN 1
```

The first line indicates which chain is running. The second line and the third lines indicate the completion of the iterations for the first chain. If the user requests more than one chain in Step 16, SCORIGHT will start to print the information again for Chain 2 until the last chain is complete.

After the running of all chains is complete, SCORIGHT will print out the following message to indicate where the analysis results are stored:

The point estimates are computed from the last 1000 iterations for all 3 chains with every 10 iterations.

The theta estimates and their standard errors are in the file -- theta.est.

The item parameter estimates and their standard errors are in the file -- itemP.est.

The estimates related to testlets and their standard errors are in the file -- testlet.est.

The estimates gamma of each examinee for each testlet are in the file -- gamma.est.

The diagnosis of convergence are in the file -- Convergence.

... End of SCORIGHT analysis!

The program provides the names of the output files of the item parameter estimates and the estimates of the examinee proficiency θ values. If there are testlets, it also provides the name of the output files of the values related to the testlet.

5 Output Files and Format

In the subdirectory the user specifies (Step 12), SCORIGHT will generate several additional subdirectories. The number of subdirectories is the same as the number of chains specified in Step 16. The subdirectory names will be ch1, ch2, and so on, referring to the different chains.

In the subdirectory that the user specifies, SCORIGHT will generate several output files (these are described in more detail below):

itemP.est
theta.est
testlet.est (optional)
gamma.est (optional)
Convergence (optional)

If the user runs more than one chain, there will be a file called Convergence in the subdirectory the user specified in Step 12. The file contains the information about the diagnosis of the convergence for the whole analysis. These are described later in Section 5.1 and are based upon the Gelman-Rubin (1993) F-test diagnostic. The files testlet.est and gamma.est will be generated if there are any testlet items within the test.

5.1 *Item Parameter Output File: itemP.est*

The file itemP.est contains two parts. The first part contains the information for the estimated parameters of each item and is formatted in typical FORTRAN fashion:

I4, 1X, 1C, 6F11.4, 2X, I1, mF11.5

This information indicates the following:

- *I4*: The first four columns of each record (row) is an integer that indicates the item number.
- *1X*: An empty space.
- *1C*: One character for an item type (D for 3PL dichotomous item, 2 for 2PL dichotomous, and P for polytomous).
- *6F11.4*: Six floating point values, each occupying 11 columns, which are the estimated values of parameters a , b , and c and their estimated standard errors. If the item is either a polytomous item or a 2PL dichotomous item, the corresponding spaces for the estimated values of parameter c and its standard error include NAs. For a polytomous

item, the output following the estimated values includes an integer that indicates the total number of categories, m .

- $2(m - 1)F11.5$: A polytomous item with $2(m - 1)$ floating point values, which are the $m - 1$ estimated values of the cutoffs and the corresponding standard errors for each category. If any parameters are fixed at the initial values, the corresponding estimated standard errors will be printed as NAs.

The second part of the file itemP.est gives the estimated values for the covariate coefficients and the estimated variances and covariances of the item parameters and their corresponding standard errors. The number of covariate coefficients that each item parameter has will correspond to the number of covariates that the user has input in Step 19 plus one (an intercept). The outputs are given as follows:

Estimated regression coefficients of 3PL binary item parameters:

For item parameter h ($h = \log(a)$):

| | beta_0 | beta_1 | beta_2 |
|-------------------|--------|--------|--------|
| Estimated values: | 0.8768 | 1.7262 | 2.1374 |
| s.e.: | 0.0890 | 0.1435 | 0.1251 |

For item parameter b :

| | beta_0 | beta_1 |
|-------------------|--------|--------|
| Estimated values: | 1.9850 | 1.5249 |
| s.e.: | 0.1849 | 0.0821 |

For item parameter q ($q = \log(c/(1 - c))$):

| | beta_0 | beta_1 |
|-------------------|---------|---------|
| Estimated values: | -1.5387 | -1.0632 |
| s.e.: | 0.1446 | 0.0994 |

Estimated covariance matrix of item parameters $h(= \log(a))$, b , $q(= \log(c/(1 - c)))$:

| | SIGMA_h | RHO_hb | RHO_hq | SIGMA_b | RHO_bq | SIGMA_q |
|-------------------|---------|--------|--------|---------|--------|---------|
| Estimated values: | 0.0797 | 0.2237 | 0.1687 | 0.9400 | 0.6434 | 0.5418 |
| s.e.: | 0.0231 | 0.0549 | 0.0665 | 0.1955 | 0.1719 | 0.2086 |

Estimated coefficients of polytomous item parameters:

For item parameter h ($h = \log(a)$):

| | beta_0 | beta_1 |
|-------------------|--------|--------|
| Estimated values: | 2.8768 | 0.3242 |
| s.e.: | 0.0890 | 0.1435 |

For item parameter b :

| | beta_0 | beta_1 |
|-------------------|--------|--------|
| Estimated values: | 1.9850 | 1.5249 |
| s.e.: | 0.1849 | 0.0821 |

Estimated covariance matrix of item parameters $h(= \log(a))$ and b :

| | SIGMA_h | RHO_hb | SIGMA_b |
|-------------------|---------|--------|---------|
| Estimated values: | 0.0797 | 0.2237 | 0.1687 |
| s.e.: | 0.0231 | 0.0549 | 0.0665 |

Estimated coefficients of 2PL binary item parameter regression:

For item parameter h ($h = \log(a)$):

| | beta_0 |
|-------------------|--------|
| Estimated values: | 0.5768 |
| s.e.: | 0.0890 |

For item parameter b :

| | beta_0 | beta_1 |
|-------------------|--------|--------|
| Estimated values: | 1.9850 | 1.5249 |
| s.e.: | 0.1849 | 0.0821 |

Estimated covariance matrix of item parameters $h(= \log(a))$ and b :

| | SIGMA_h | RHO_hb | SIGMA_b |
|-------------------|---------|--------|---------|
| Estimated values: | 0.0797 | 0.2237 | 0.1687 |
| s.e.: | 0.0231 | 0.0549 | 0.0665 |

Estimated coefficients of theta regression:

For theta covariates:

| | beta_0 | beta_1 |
|-------------------|--------|--------|
| Estimated values: | 1.8872 | 1.2405 |
| s.e.: | 0.0011 | 0.0023 |

For the second part of the file itemP.est, SCORIGHT will print out the estimated values only when they are available. For example, if there are no covariates for θ , the last part in the above output, estimated coefficients of the theta regression, will not appear.

5.2 Proficiency Parameter Output File: theta.est

The file named theta.est contains the estimated value of each examinee's proficiency θ value and its corresponding standard error. This file has the same number of records (rows) as examinees. Its format is:

I6, 2F11.4

This information means the following:

- *I6*: The first six columns contain an integer indicating the examinee number.
- *2F11.4*: Two floating values, each occupying 11 columns, which are the estimated proficiency (θ) value and its estimated standard error. If the initial values of θ are supplied and fixed, there will be no estimated standard error and the corresponding spaces of the output will be printed as NAs.

5.3 Testlet Output File: *testlet.est*

If there are any testlets in the test, the file *testlet.est* will be generated. It contains the estimated variance of γ for each testlet, the estimated regression coefficients for $\log(\sigma_{d(j)}^2)$, and the estimated variance τ^2 for $\log(\sigma_{d(j)}^2)$ if there are covariates for the testlet effects. For the previous artificial (simulated) dataset:

Estimated variance of the variance of gamma for each testlet:

| | Estimated | S.E. |
|-------------|-----------|--------|
| Testlet 1: | 10.9020 | 2.3477 |
| Testlet 2: | 9.4903 | 2.2128 |
| ... | | |
| Testlet 11: | 3.5296 | 0.5179 |
| Testlet 12: | 0.5436 | 0.1103 |
| ... | | |

Estimated coefficients of log of variance of variance of gamma:

| | delta_0 | delta_1 |
|-------------------|---------|---------|
| Estimated values: | 0.5247 | -0.9121 |
| s.e.: | 1.0657 | 10.0793 |

Estimated variance of the log testlet variances -- gamma:

| | Tau |
|-------------------|--------|
| Estimated values: | 2.8750 |
| s.e.: | 1.7030 |

5.4 Testlet Output File: *gamma.est*

The file *gamma.est* contains the estimated value of γ for each examinee and each testlet. It has as many records (rows) as the number of examinees. For each record, it has the following format:

I6, DF9.4

This information means the following:

- *I6*: The first six columns contain an integer representing the examinee number.
- *D*: The total number of testlets. So there are *D* estimated γ values for each examinee.

There are several output files in each chain subdirectory. They are:

| | |
|------------------------|-------------------------|
| a_DrawsC (optional) | gamV_DrawsC (optional) |
| b_DrawsC (optional) | SIGMA_DrawsC (optional) |
| c_DrawsC (optional) | delta_DrawsC (optional) |
| t_DrawsC (optional) | tau_DrawsC (optional) |
| dr_DrawsC (optional) | lambda_DrawC (optional) |
| beta_DrawsC (optional) | |

The files *a_DrawsC*, *b_DrawsC*, and *c_DrawsC* will not be generated if the user fixes all item parameters. If the user only fixes some of the item parameters, these three files are still generated. The file *t_DrawsC* will not be generated if the user fixes the θ values throughout

the analysis. The file `dr_DrawsS` will be generated if the test has polytomous items. Files `delta_DrawsC` and `tau_DrawsC` will only be generated if there are covariates for the testlet effects. The file `lambda_DrawsC` will be generated only when there are covariates for the θ s.

5.5 Output Files Containing Parameter Draws: *.DrawsC

All file names ending with `DrawsC` contain the random draws from the posterior distribution of the corresponding parameters. The user can use this information to calculate any interesting statistic or to make further statistical inferences from them. File `a_DrawsC` contains the random draws for item parameters a_1, \dots, a_J with format:

JF11.6

where J is the total number of items that are estimated by SCORIGHT (i.e., not including those whose values are fixed throughout the analysis). There are J floating point values that are the draws. The file contains the same number of rows as the number of iterations specified (Step 12) minus the number of initial iterations that were discarded (Step 13) and divided by the gap. For the example in this manual, the total records (rows) is equal to $g = (4000 - 3000)/10 = 100$. So the files `a_DrawsC` and `b_DrawsC` contain g rows and J columns corresponding to the number of item parameters that are not fixed in the analysis. And the file `c_DrawsC` contains g rows and the number of 3PL dichotomous items whose item parameters are not fixed in the analysis. For the example inputs entered in earlier sections, these files would have $g = 100$ rows and 80 columns (80-item test) in the file `a_DrawsC` if all the item parameters are not fixed in the analysis.

The files `b_DrawsC` and `c_DrawsC` are similar to `a_DrawsC`. It should be noted that in the output for the draws of item parameter c the format should be $J_1F11.6$, where J_1 is the total number of estimated 3PL dichotomous items whose item parameters are not fixed in the analysis.

The format of file t_DrawsC is:

nF11.6

where n is the total number of examinees (if they are not fixed). There are n floating point values that are the draws of the examinees' proficiency θ values from the sampler. The total number of records (rows) of this file is the same as a_DrawsC.

The format of the file dr_DrawsC is:

KF10.6

where K is the total number of polytomous cutoffs that need to be estimated for the model. Suppose there are L polytomous items in the test, and for each polytomous item, $d_i, i = 1, \dots, L$, is the number of categories. Since the first cutoff of each polytomous item is set to 0, there are $d_i - 2$ estimated cutoffs needed for each polytomous item. Therefore,

$$K = \sum_i^L (d_i - 2)$$

cutoffs are needed in total. Each record of the file therefore contains the random draws of cutoffs for each polytomous item in sequence (i.e., the first $d_1 - 2$ are for the first polytomous item and so on). As before, the number of records (rows) is the same as a_DrawsC.

File beta_DrawsC has the format:

nF10.3

It also has g rows, which is similar to a_DrawsC and others. For each row, SCORIGHT will print the coefficients of the covariates (if there are any and just the intercept if there are not) in the following order: $\beta_h^{(3)}, \beta_b^{(3)}, \beta_q^{(3)}, \beta_h^{(p)}, \beta_b^{(p)}, \beta_h^{(2)}$, and $\beta_b^{(2)}$, according to the inclusion of items of the three different groups. If all the item parameters are fixed during

the analysis, this file will not appear.

File SIGMA_DrawsC has the format:

lF10.3

It has g rows as before. For each row, SCORIGHT will print the components of the upper triangular of each covariance matrix in the following order: Σ_{3PL} , Σ_{Poly} , and Σ_{2PL} , according to the inclusion of test items from the three different groups. If all item parameters are fixed during the analysis, this file will not appear.

File lambda_DrawsC has the format:

lF11.6

It has g rows as before. For each row, SCORIGHT will print one draw for each of the L coefficients for the covariates of the θ values. If there are no covariates for θ , this file will not be generated.

File gamV_DrawsC contains all the draws of the variance of γ for each testlet for all the iterations. For example, as input before, this file has 4,000 records. And each record has the following format:

I7, DF12.6

This means the following:

- *I7*: The first seven columns together contain an integer, indicating the iteration number.
- *D*: The total number of testlets.

Since this file contains all the draws values starting from the first iteration, it is possible to analyze how fast SCORIGHT has converged for this parameter.

File `delta_DrawsC` contains all the iterations of `gamV.est` and has the format:

I7, pF14.6

This means the following:

- *I7*: The first seven columns together contain an integer, indicating the iteration number.
- *p*: The intercept and the regression coefficients of the $p - 1$ covariates for the testlet.

File `tau_DrawsC` contains all draws for the variance of $\log(\sigma_{d(j)}^2)$. It has the format:

I7, F20.4

which is similar to `delta_DrawsC`.

5.6 Convergence Assessment

The file `Convergence` contains information that allows for a diagnosis of the convergence of the Markov chains. It only appears when the user runs more than one chain, and the diagnosis is printed out only for the higher-level parameters (i.e., means and covariances that drive the individual and item-level parameters). If there are any testlets, it will also print out the diagnosis information for all the variances of gammas. For each estimated value, SCORIGHT will print out two statistics: `post` and `confshrink`. `post` gives three quantiles, 2.5%, 50%, and 97.5%, for the target distribution based on the Student-t distribution; `confshrink` gives 50% and 97.5% quantiles of a rough upper bound on how much the confidence interval of `post` would shrink if the iterative simulation is continued forever. If both components of `confshrink` are not near 1, the user should probably run the iterative simulation further. Gelman and Rubin (1993) suggest that values of `confshrink` less than 1.2 indicate reasonable convergence.

6 Example: Equating Using SCORIGHT

This section will use a simulated (artificial) example to demonstrate some of SCORIGHT's capabilities. The simulated (artificial) data have the following structure: There are a total of 2,000 examinees and 80 items, 60 of which comprise 20 testlets and 20 that are independent. See Table 1.

Table 1.

The Structure of the Simulated Test

| Items | Structure | Item types |
|-------|-------------------|---------------------------|
| 1-3 | testlet # 1 | 2PL binary response items |
| 4-6 | testlet # 2 | 2PL binary response items |
| 7-10 | independent items | 2PL binary response items |
| 11-13 | testlet # 3 | 3PL binary response items |
| 14-16 | testlet # 4 | 3PL binary response items |
| 17-20 | independent items | 3PL binary response items |
| 21-23 | testlet # 5 | polytomous items |
| 24-26 | testlet # 6 | polytomous items |
| 27-29 | testlet # 7 | polytomous items |
| 30-32 | testlet # 8 | polytomous items |
| 33-35 | testlet # 9 | polytomous items |
| 36-38 | testlet # 10 | polytomous items |
| 39-40 | independent items | polytomous items |

Note. Items 41-80 are a repeat of the structure of Items 1-40.

There are covariates for examinees' proficiency, item parameters, and testlet effects. See Table 2.

This simulated dataset illustrates some of SCORIGHT's capabilities within a practical context. One common situation requires one to fit data that arise from multiple test forms or multiple examinee groups or both. To fit such data requires that the different datasets

Table 2.
Covariates for Examinee's Data

| | | Covariate | Number of covariates |
|--|---------------------|-----------|----------------------|
| Examinees' proficiency θ | | Yes | 1 |
| Testlet effect $\log(\sigma_{d(j)}^2)$ | | Yes | 1 |
| 2PL items | Item parameter a | Yes | 1 |
| | Items parameter b | Yes | 1 |
| 3PL items | Item parameter a | Yes | 1 |
| | Item parameter b | Yes | 1 |
| | Item parameter c | Yes | 1 |
| Polytomous items | Items parameter a | Yes | 1 |
| | Items parameter b | Yes | 1 |

be equated. Consider specific cases of four broad categories of such situations.

- *Case I – Equating randomly equivalent groups.* Two different test forms are administered to two groups of examinees that can be assumed to have been randomly assigned to the form they received. This approach is used, for example, by the Canadian military to equate the French and English versions of their placement exam. They assume that the ability distributions of Anglophone and Francophone enlistees are the same and estimate the difficulties of each test form accordingly (Wainer, 1999). This sort of design can be easily dealt with in SCORIGHT by fitting each form separately and insisting that the ability distributions for the two groups be identical, say $N(0,1)$. The items from each form will automatically be equated. An example of such a dataset would be one in which the first 1,000 examinees are administered the first 40 items, and the second 1,000 examinees are given the second 40 items. See Table 3.

- *Case II – Equating two forms with some overlapping items.* Two different test forms are administered to two groups of examinees that were not randomly assigned to the

Table 3.
Equating Randomly Equivalent Groups

| Examinees | Items | | | | | |
|-----------|-------|-------|-------|-------|-------|-------|
| | 1-10 | 11-20 | 21-40 | 41-50 | 51-60 | 61-80 |
| 1-1000 | X | X | X | | | |
| 1001-2000 | | | | X | X | X |

form they received. Because one cannot make the assumption that the two groups share the same ability distribution (as in Case I), one must use the overlapping items as an anchor link to equate the two forms. This is the situation that is commonly encountered with most professionally prepared tests that administer several different forms. Once again SCORIGHT can do this equating easily. One way is to fix the ability distribution for the two groups together as, say, $N(0,1)$, and estimate the parameters for all of the items. SCORIGHT treats unobserved responses as ignorably missing. A second approach is to fix the ability distribution of one group as, say $N(0,1)$, and estimate the parameters of the other group's ability distribution along with the item parameters. An schematic representation of a dataset with overlapping items is shown in Table 4, in which 2,000 examinees take one of two 50-item test forms, with each form having 24 testlet items and 6 independent items in common.

Table 4.
Equating Two Forms

| Examinees | Items | | | | | |
|-----------|-------|-------|-------|-------|-------|-------|
| | 1-10 | 11-20 | 21-40 | 41-50 | 51-60 | 61-80 |
| 1-1000 | X | X | X | | | X |
| 1001-2000 | | X | | X | X | X |

- *Case III – Equating two forms with no items in common but with some individuals who have taken both forms: common person equating.* This approach is often used when equating two test forms given in different languages, and there is a sample of bilingual examinees who have taken both forms (Hambleton, 1993; Sireci, 1997). If one can assume that the individuals who have taken both forms are equally able in either context, they can be used as the link to equate the forms. This situation is shown in Table 5. SCORIGHT can fit tests with this structure by using the common examinees as the equating link. This is done operationally by treating the entire administration as a single test in which Examinees 1-800 are ignorably missing responses to Items 41-80 and Examinees 1201-2000 are missing responses to Items 1-40. Linking these two disparate groups are Examinees 801 through 1200, who took all 80 items.

Table 5.
Common Person Equating

| Examinees | Items | | | | | |
|-----------|-------|-------|-------|-------|-------|-------|
| | 1-10 | 11-20 | 21-40 | 41-50 | 51-60 | 61-80 |
| 1-800 | X | X | X | | | |
| 801-1200 | X | X | X | X | X | X |
| 1201-2000 | | | | X | X | X |

- *Case IV – Equating two forms that have some items in common as well as some examinees who took all items: a hybrid situation.* In some sense, this is the most complex of the equating situations. One version, combining both Cases II and III, is depicted in Table 6. SCORIGHT can deal with this as easily as it can with any of the others. This manual will now demonstrate, in detail, exactly how to run SCORIGHT for this circumstance and include both input and output. Users who have a situation like those depicted in Cases I, II, or III can use this same setup with appropriate deletions.

Table 6.
Hybrid Situation

| Examinees | Items | | | | | |
|-----------|-------|-------|-------|-------|-------|-------|
| | 1-10 | 11-20 | 21-40 | 41-50 | 51-60 | 61-80 |
| 1-800 | X | X | X | | | X |
| 801-1200 | X | X | X | X | X | X |
| 1201-2000 | | X | | X | X | X |

What follows are the details of running SCORIGHT for Case IV. For Case IV, there are 2,000 examinees and 80 items (Step 2), which comprise 20 testlets (Step 3) with 60 testlet items and 20 independent items. This analysis also includes one covariate for each of the parameters, a , b , c , γ , and θ . The part of the data containing just the item responses is shown below with the file name case4.data (Step 4):

```
000010100011000100001115311111112211111NNNNNNNNNNNNNNNNNNNN11111145211243111111
...
10111101100000111111111522151111332313530011101101011110011125252111121142121323
...
NNNNNNNNNN00011111111NNNNNNNNNNNNNNNNNNNN111010111111111111155553545525453422553
```

Here the responses of three examinees are shown: Examinees 1, 801, and 1201. For Examinee 1, there are 60 responses for the first 40 items, and the last 20 items, Item 41 to Item 60, are not assigned. Examinees 1 to 800 have the same item response structure. For Examinee 801, all 80 items are assigned. The item response structure is the same for Examinees 801 to 1200. For Examinee 1201, Items 1-10 and Items 21-40 are not assigned. The item response structure is the same for Examinees 1201 to 2000. Since there are polytomous items, 2PL binary response items, and 3PL binary response items, a file indicating the different types of items is needed. In the example in Step 10, this file was named index, the content of which follows:

2 2
2 2
...
D 2
D 2
...
P 5
P 5
...
D 2
D 2
...
P 5
P 5
...

In this file, Lines 1-10 are 2 2 (meaning that the first 10 items are 2PL binary response items), Lines 11-20 are D 2, (meaning Items 11 to 20 are 3PL binary response items), Lines 21-40 are P 5 (meaning items 21 to 40 are polytomous response items with five categories each), Lines 41-50 are 2 2 (the same meaning as Items 1 to 10), Lines 51-60 are D 2 (the same meaning as items 11 to 20), Lines 61-80 are P 5 (the same meaning as Items 21 to 40).

Since there are covariates for examinees' proficiency θ , item parameters, and testlet effects, prepare the corresponding files as specified in Steps 19, 20, and 21. The following are the first five lines of the file that contains the covariates of examinees' proficiency θ :

-0.79447
0.42339
1.27102
0.69914
-0.34366
...

The file containing the covariates of θ , theta_covariate, has 2,000 lines, and the first 12 columns of each line contain the value of the covariate of each examinee's proficiency (Step 20). In all cases, if you have more than one covariate, say p of them, they are entered here in $pF12.5$ format. The file containing the covariates of item parameter a , a_covariate, has 80 lines in total, and the first 12 columns of each line contain the value of the covariate of each item parameter a . The files b_covariate and c_covariate are similar. Their formats are described in Step 19. The file containing the covariates of $\log(\sigma_{d(j)}^2)$, testlet_covariate, has 20 rows (the number of testlets), and the first 12 columns of each row contain the value of the single covariate.

Next are the details on how to enter the information into SCORIGHT and the results of the analysis.

This program estimates the proficiency and item parameters for both dichotomous and polytomous items that could be independent or nested within testlets using the Gibbs sampler. To run this program, you need to provide the following information.

Please enter the number of examinees and the number of items

in your dataset separated by at least one space: **2000 80**

Please enter the number of dichotomous items within the total 80 items: **40**

Please enter the number of 2PL binary response items: **20**

Enter the total number of testlets in the test: **20**

Enter the name of the file that contains the test data: **c:\equating\case4.data**

Enter the starting and ending columns of the test scores
for the data file: **1 80**

Enter the starting and ending columns of Testlet #1: **1 3**

Enter the starting and ending columns of Testlet #2: **4 6**

Enter the starting and ending columns of Testlet #3: **11 13**

Enter the starting and ending columns of Testlet #4: 14 16

Enter the starting and ending columns of Testlet #5: 21 23

Enter the starting and ending columns of Testlet #6: 24 26

Enter the starting and ending columns of Testlet #7: 27 29

...

Enter the starting and ending rows of the test scores: 1 2000

Enter the name of the item information file: c:\equating\index

Please enter the name of the subdirectory (include the last backslash) where you want to put the analysis results, and make sure that there is no subdirectory called ch1, ch2, ... in it: c:\equating\result\

Enter the number of needed iterations of sampling: 25000

Enter the number of draws to be discarded: 15000

Enter the size of the gap between posterior draws: 50

How many chains do you want to run? 2

For CHAIN 1: Do you want to input the initial values for item parameters a, b, and c? If yes, enter 1, otherwise, enter 0: 0

For CHAIN 1: Do you want to input the initial values for the proficiency parameters' theta? If yes, enter 1, otherwise, enter 0: 0

For CHAIN 2: Do you want to input the initial values for item parameters a, b, and c? If yes, enter 1, otherwise, enter 0: 0

For CHAIN 2: Do you want to input the initial values for the proficiency parameters' theta? If yes, enter 1, otherwise, enter 0: 0

Do you have covariates for item parameter a (not including intercept)?

If Yes, enter 1, otherwise enter 0: 1

Please enter the total number of covariates for Item

Parameter a (without intercept) for the 3PL binary

response items: 1

Please enter the total number of covariates for Item

Parameter a (without intercept) for the 2PL binary

response items: 1

Please enter the total number of covariates for Item

Parameter a (without intercept) for the polytomous items: 1

Please enter the name of the file

that contains the covariate information of the Item

Parameter a: c:\equating\a_covariate

Do you have covariates for item parameter b (not including intercept)?

If Yes, enter 1, otherwise enter 0: 1

Please enter the total number of covariates for Item

Parameter b (without intercept) for the 3PL binary

response items: 1

Please enter the total number of covariates for Item

Parameter b (without intercept) for the 2PL binary

response items: 1

Please enter the total number of covariates for Item

Parameter b (without intercept) of the polytomous items: 1

Please enter the name of the file

that contains the covariate information for Item

Parameter b: c:\equating\b_covariate

Do you have covariates for item parameter c (not including intercept)?

If Yes, enter 1, otherwise enter 0: 1

Please enter the total number of covariates for Item

Parameter c (without intercept) for the 3PL binary
response items: 1

Please enter the name of the file
that contains the covariate information for Item

Parameter c: c:\equating\c_covariate

Do you have covariates for item parameter theta? If Yes, enter 1
otherwise enter 0: 1

Please enter the total number of the covariates for Item

Parameter theta (without intercept): 1

Please enter the name of the file
that contains the covariate information for Item

Parameter theta: c:\equating\theta_covariate

Do you have any covariates for the testlet effects (not including intercept)?
If YES, enter 1, otherwise enter 0: 1

Please enter the total number of covariates
for the testlet effects (without intercept): 1

Please enter the name of the file
that contains the covariate information for the
testlet effect variances: c:\equating\testlet_covariate

Please check the input:

-2 means independent items,
1 means the first testlet item,
2 means the second testlet items,
... and so on

111222-2-2-2-2333444-2-2-2-2555666777888999101010-2- 2111111121212-2-2-2-2

131313141414-2-2-2151515161616171717181818191919202020-2-2

If the input is correct, enter 1, otherwise, enter 0: 1

CHAIN 1 Starting time: Sat Mar 13 17:08:02 2004

CHAIN 1 Time after 1 cycle: Sat Mar 13 17:08:05 2004
CHAIN 1 Time after 11 cycles: Sat Mar 13 17:08:30 2004
...

For Chain 1:

The Gibbs sampling of 25000 iterations is completed.
... End of running of CHAIN 1

CHAIN 2 Starting time: Sat Mar 13 18:54:50 2004
CHAIN 2 Time after 1 cycle: Sat Mar 13 18:54:53 2004
CHAIN 2 Time after 11 cycles: Sat Mar 13 18:55:19 2004
...

For Chain 2:

The Gibbs sampling of 25000 iterations is completed.
... End of running of CHAIN 2

The point estimates are computed from the last 10000 iterations
for all 2 chains with every 50 iterations

The theta estimates and their standard errors are
in file -- theta.est.

The item parameter estimates and their standard errors are
in file -- itemP.est.

The estimates related to testlets and their standard errors are
in file -- testlet.est.

The gamma estimates of each examinee for each testlet are
in file -- gamma.est.

The diagnosis of convergence are
in file -- Convergence.

... End of analysis of SCORIGHT!

All the input files are in c:\equating\. The output files are in c:\equating\result\,
which has two subdirectories, ch1 and ch2, and the following files:

itemP.est
theta.est
gamma.est
testlet.est
Convergence

Below is part of the file itemP.est:

```
#####          EST 'a'  SE('a')  EST 'b'  SE('b')  EST 'c'  SE('c')  
      1  2    2.0919   0.1701   0.7346   0.0540         NA         NA  
      ...  
     11  D    0.6017   0.1341   2.6087   0.3086   0.0559   0.0296  
      ...
```

The first line of the file are labels that describe what is printed beneath it. The first column contains the item number. The second line (row) of this file tells us that the first item is a 2PL binary response item, since in the second column there is 2, which represents a 2PL binary response item. The estimated value of item parameter a is 2.0919, and the corresponding estimated standard error is 0.1701. The estimated value of item parameter b for Item 1 is 0.7346, and its corresponding estimated standard error is 0.0540. Since it is a 2PL binary response item, there is no estimated value of item parameter c , and therefore, the next two columns are coded NA, meaning that they are not in the model. For Item 11, the D in the second column means that it is a 3PL binary response item. The estimated value of item parameter a is 0.6017, and the corresponding estimated standard error is 0.1341. The estimated value of item parameter b for Item 11 is 2.6087, and its corresponding estimated standard error is 0.3086. The estimated value of item parameter c for Item 11 is 0.0559, and its corresponding estimated standard error is 0.0296. If an item is a polytomous item, the information not only includes the estimated item parameters but also includes the estimated cutoffs. To illustrate this, let us consider Item 21:

```

#####      EST 'a'  SE('a')  EST 'b'  SE('b')  EST 'c'  SE('c')
...
      21 P   3.2151   0.3763   1.6161   0.0923        NA        NA
...

```

The P in the second column indicates that this item is polytomous. Therefore, there is no estimated value for c nor any corresponding standard error. Following the last NA on the same line, there is information about the cutoffs:

```

...
5  0.00000  NA  2.26385  0.25928  4.34023  0.39543  5.33997  0.46110
...

```

In the ninth data field, the 5 means that Item 21 has five categories. Therefore, it has four cutoffs, in which the first one is set to 0.00, and the corresponding estimated standard error is not available, NA. The estimated value of the second one is 2.26385; the estimated standard error is 0.25928. The estimated value of the third one is 4.34023; the estimated standard error is 0.39543. The estimated value of the fourth one is 5.33997; the estimated standard error is 0.46110.

Since there are covariates for item parameters and for examinees' proficiency, this file contains more analysis results for the estimated item parameters. The following describes the covariate effects for the 2PL binary response items:

Estimated coefficients of 2PL binary item parameters:

```

For item parameter  $h$  ( $h = \log(a)$ ):
                beta_0  beta_1
Estimated values:  2.2001  1.0750
s.e.:             0.1968  0.1073

```

For item parameter b :

| | beta_0 | beta_1 |
|-------------------|---------|--------|
| Estimated values: | -1.1059 | 1.3801 |
| s.e.: | 0.4091 | 0.3776 |

Estimated covariance matrix of item parameters $h(= \log(a))$, b :

| | SIGMA_h | RHO_hb | SIGMA_b |
|-------------------|---------|--------|---------|
| Estimated values: | 0.0998 | 0.0717 | 0.6978 |
| s.e.: | 0.0438 | 0.0767 | 0.2612 |

For the 2PL binary response items, there is only one covariate each for item parameters a and b . Therefore, under h ($h = \log(a)$), beta_0 is the estimated intercept, and beta_1 is the estimated coefficient, with the line underneath it the corresponding estimated standard error. The covariate estimates for the item parameter b follow similarly. Note that the covariates on $\log(a)$ and b were regressed as specified in the model. The covariate estimates for the polytomous items parallel the 2PL binary case.

The covariate analysis of the 3PL binary response items yields:

Estimated coefficients of 3PL binary item parameters:

For item parameter h ($h = \log(a)$):

| | beta_0 | beta_1 |
|-------------------|--------|--------|
| Estimated values: | 1.2721 | 0.7055 |
| s.e.: | 0.5622 | 0.3525 |

For item parameter b :

| | beta_0 | beta_1 |
|-------------------|---------|--------|
| Estimated values: | -0.5813 | 1.2895 |
| s.e.: | 0.2597 | 0.2132 |

For item parameter q ($q = \log(c/(1 - c))$):

| | | |
|-------------------|---------|--------|
| | beta_0 | beta_1 |
| Estimated values: | -2.5000 | 2.6451 |
| s.e.: | 0.3957 | 1.0977 |

Estimated covariance matrix of item parameters $h(= \log(a))$, b , $q(= \log(c/(1 - c)))$:

| | | | | | | |
|-------------------|---------|---------|--------|---------|---------|---------|
| | SIGMA_h | RHO_hb | RHO_hq | SIGMA_b | RHO_bq | SIGMA_q |
| Estimated values: | 0.3890 | -0.3614 | 0.2111 | 1.0158 | -0.0971 | 0.6375 |
| s.e.: | 0.2435 | 0.2179 | 0.1724 | 0.4006 | 0.2108 | 0.2735 |

Since there is one covariate for the examinees' proficiency θ , the corresponding covariate analysis results appear at the end of the file itemP.est. This is the estimated value for the coefficient for the covariate of θ . Note that there is no intercept for the regression of θ .

Estimated coefficients of theta:

For theta covariates:

| | |
|-------------------|--------|
| | beta_1 |
| Estimated values: | 1.8620 |
| s.e.: | 0.0289 |

The following are the first few lines from the file theta.est, which contains all the estimated values of the 2,000 examinees' proficiency, θ (which is calculated as the mean of each examinee's posterior density).

| #### | EST Theta | SE(Theta) |
|------|-----------|-----------|
| 1 | -1.4330 | 0.2590 |
| 2 | 0.4614 | 0.1691 |
| 3 | 2.2661 | 0.2497 |
| 4 | 1.3236 | 0.2084 |

```

...
1999 -1.3830 0.3130
2000 -0.6210 0.2715

```

The first column contains the examinee's ID, the second column is the estimated proficiency, and the third column is the corresponding estimated standard error.

The following lines are from the file `gamma.est`, which contains the estimated γ values for each examinee across all testlets. There are 20 testlets in this analysis, and the first line of this file reports that Examinee 1 has estimated values of γ for Testlets 1, 2, 3, and 4 as 0.4406, 0.2521, 0.2254, and -0.3168, respectively. The values of γ s for the remaining 16 testlets are not shown because of space limitations. Each subsequent line represents one examinee; thus for this analysis, this file contains 2,000 lines:

```

1 0.4406 0.2521 0.2254 -0.3168 ...
2 0.2855 -1.1420 0.2276 -2.0955 ...
...

```

The file `Convergence` contains the information about the diagnosis of the convergence of the Markov chain. It only exists when the user runs multiple chains. This example ran two chains. The first part of `Convergence` looks like the following:

```

DIAGNOSIS FOR CONVERGENCE;

post: (2.5, 50, 97.5) quantiles for the target distribution
      based on the Student-t distribution

confshrink: the 50th and 97.5th quantiles of a rough upper bound on
            how much the confidence interval of post will shrink
            if the iterative simulation is continued forever.

```

If both components of `confshrink` are not near 1, the user should probably run the iterative simulation further.

The rest of Convergence gives the posterior quantiles and statistics for

$$\Lambda_2 = \{\beta_h^{(3)}, \beta_b^{(3)}, \beta_q^{(3)}, \Sigma_{3PL}, \beta_h^{(2)}, \beta_b^{(2)}, \Sigma_{2PL}, \beta_h^{(p)}, \beta_b^{(p)}, \Sigma_{Poly}, \sigma_{d(j)}^2\}$$

For example, the following is for 2PL binary response items:

2PL binary items:

Coefficients for item parameter a :

Beta_0

Post:

| | | |
|------|------|------|
| 1.81 | 2.20 | 2.59 |
|------|------|------|

Confshrink:

| | | |
|------|------|--|
| 1.01 | 1.05 | |
|------|------|--|

Beta_1

Post:

| | | |
|------|------|------|
| 0.86 | 1.07 | 1.29 |
|------|------|------|

Confshrink:

| | | |
|------|------|--|
| 1.02 | 1.09 | |
|------|------|--|

Coefficients for item parameter b :

Beta_0

Post:

| | | |
|-------|-------|-------|
| -1.91 | -1.11 | -0.30 |
|-------|-------|-------|

Confshrink:

| | | |
|------|------|--|
| 1.00 | 1.02 | |
|------|------|--|

Beta_1

Post:

| | | |
|------|------|------|
| 0.64 | 1.38 | 2.12 |
|------|------|------|

Confshrink:

| | | |
|------|------|--|
| 1.01 | 1.05 | |
|------|------|--|

Variance Matrix of item parameters a and b :

Variance of a :

Post:

| | | |
|-------|------|------|
| 0.014 | 0.10 | 0.19 |
|-------|------|------|

Confshrink:

| | | |
|------|------|--|
| 1.00 | 1.00 | |
|------|------|--|

```

Covariance of item parameters a and b:
Post:
-0.08  0.07                                0.22
Confshrink:
  1.01  1.06

```

```

Variance of item parameter b:
Post:
  0.16  0.70                                1.24
Confshrink:
  1.04  1.06

```

If there are testlets in the analysis, the last part of Convergence shows the posterior quantiles and the corresponding statistics for each $\sigma_{d(j)}^2$. Here only the first two testlets are shown:

```

Variance of gamma for testlet:

Testlet   1:
Posterior Range:
          0.13          0.22          0.32
Confidence Range:
          1.03          1.14

          Testlet   2:
          Posterior Range:
          0.68  1.06  1.45
          Confidence Range:
          1.07  1.30

          ...

```

The other files contain the random draws from the posterior distributions for each chain and are in subdirectories ch1 and ch2. For example, the file a_DrawC in ch1 has the last 200 random draws; the number is specified by the user as the difference between the number of iterations (Step 13), the number of initial draws to discard as burn-in (Step 14), and the gap between each record (Step 15) for the 80 *a* parameters. It contains a 200×80

matrix. The i th row contains the draws of the 80 a's from the i th iteration of the sampler, and the j th column contains 200 draws from the posterior distribution of the j th test item.

This example treated the nonresponse to unadministered items as ignorable, in the sense of Little and Rubin (1987). If this assumption is correct, the three test forms are now equated, and the scores that were estimated are comparable regardless of the test form that generated them.

References

- Birnbaum, A. (1968). Some latent trait models. In F. Lord & M. Novick (Eds.), *Statistical theories of mental test scores* (pp. 397-479). Reading, MA: Addison Wesley.
- Bradlow, E. T., Wainer, H., & Wang, X. (1999). A Bayesian random effects model for testlets. *Psychometrika*, *64*, 153-168.
- Gelman, A., & Rubin, D. B. (1993). Inference from iterative simulation using multiple sequences. *Statistical Science*, *7*, 457-472.
- Hambleton, R. K. (1993). Translating achievement tests for use in cross-national studies. *European Journal of Psychological Assessment*, *9*, 57- 68.
- Levine, M. V., & Drasgow, F. (1988). Optimal appropriateness measurement. *Psychometrika*, *53*, 161-176.
- Little, R. J. A., & Rubin, D. B. (1987). *Statistical analysis with missing data*. New York: Wiley.
- Ramsay, J. O. (1973). The effect of number of categories in rating scales on precision of estimation of scale values. *Psychometrika*, *28*, 513-532.
- Rasch, G. (1980). *Probabilistic models for some intelligence and attainment tests*. University of Chicago. (Original work published 1960).
- Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika Monographs*, No. 17.
- Sinharay, S. (in press). Experiences with MCMC convergence assessment in two psychometric examples. *Journal of Educational and Behavioral Statistics*, *29*.
- Sireci, S. G. (1997). Problems and issues in linking assessments across languages. *Educational Measurement: Issues and Practice*, *16*(1), 12-19, 29.
- Stout, W. F. (1987). A nonparametric approach for assessing latent trait dimensionality. *Psychometrika*, *52*, 589-617.
- Wainer, H. (1999). Comparing the incomparable: An essay on the importance of big assumptions and scant evidence. *Educational Measurement: Issues and Practice*, *18*, 10-16.
- Wainer, H., Bradlow, E. T., & Du, Z. (2000). Testlet response theory: An analog for the

- 3-PL useful in adaptive testing. In W. J. van der Linden & C. A. W. Glas (Eds.), *Computerized adaptive testing: Theory and practice* (pp. 245-270). Boston, MA: Kluwer-Nijhoff.
- Wainer, H., & Kiely, G. (1987). Item clusters and computerized adaptive testing: A case for testlets. *Journal of Educational Measurement, 24*, 185-202.
- Wang, X., Bradlow, E. T., & Wainer, H. (2002). A general Bayesian model for testlets: Theory and applications. *Applied Psychological Measurement, 26*(1), 109-128.
- Zhang, J., & Stout, W. F. (1999). The theoretical DETECT index of dimensionality and its application to approximate simple structure. *Psychometrika, 64*, 213-249.

Notes

¹ A version of SCORIGHT that fits the one-parameter logistical (1PL) and Rasch (1980/1960) models as a special case is currently under development.