Research Report

Limits on Log Cross-Product Ratios for Item Response Models

Shelby J. Haberman

Paul W. Holland

Sandip Sinharay

Research & Development

May 2006
RR-06-10

Limits on Log Cross-Product Ratios for Item Response Models

Shelby J. Haberman, Paul W. Holland, and Sandip Sinharay
ETS, Princeton, NJ

May 2006

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:

http://www.ets.org/research/contact.html

Abstract

Bounds are established for log cross-product ratios (log odds ratios) involving pairs of items for item response models. First, expressions for bounds on log cross-product ratios are provided for unidimensional item response models in general. Then, explicit bounds are obtained for the Rasch model and the two-parameter logistic (2PL) model. Results are also illustrated through an example from a study of model-checking procedures. The bounds obtained can provide a basis for assessment of goodness of fit of these models.

Key words: Rasch model, 2PL model, cumulant generating function

1 Introduction

Latent-variable models for item responses have strong implications for customary descriptive measures from contingency table analysis such as log cross-product ratios (Section 2). In the case of the Rasch model, this issue has been suggested by simulation studies intended to explore model diagnostics for the Rasch model (Sinharay & Johnson, 2003). These studies indicated that the log cross-product ratios predicted by the Rasch model showed remarkably little variability among different pairs of items. This note seeks to explain the observed results from simulation and to provide general bounds for log-cross product ratios for some familiar item response models. These bounds have importance in derivation of starting values for algorithms for item response analysis and for checking of models. Section 2 provides the required theoretical results. Section 3 illustrates their application to an example concerned with model checking. Section 4 considers application of results to point-biserial correlations and tables of log cross-product ratios.

2 Theoretical Results

The desired bounds for log cross-product ratios can be obtained without loss of generality by study of just two items because all items are conditionally independent given the ability parameters of an item response model. General expressions for log cross-product ratios will be provided for general one-dimensional item response models. Results will be illustrated by use of the Rasch and 2PL models.

Consider the relationship between two item responses X_1 and X_2. Let each X_j be a random variable with values 0 or 1, and let $\mathbf{X} = (X_1, X_2)$, and let $p(\mathbf{x}) = p(x_1, x_2)$ be the probability that $\mathbf{X} = \mathbf{x} = (x_1, x_2)$ for a two-dimensional vector $\mathbf{x}$ with coordinates 0 or 1. In the analysis of contingency tables, the *log cross-product ratio* (also known as *log odds ratio*) of X_1 and X_2 is

$$\gamma = \log \frac{p(0,0)p(1,1)}{p(1,0)p(0,1)}. \tag{1}$$

The coefficient γ is positive if, and only if, X_1 and X_2 are positively correlated, for the correlation of X_1 and X_2 is

$$\rho = \frac{p(1,1)p(0,0) - p(1,0)p(0,1)}{[p_1(0)p_1(1)p_2(0)p_2(1)]^{1/2}},$$

where $p_j(x)$ is the probability that $X_j = x$ (Bishop, Fienberg, & Holland, 1975, p. 381). In general, the formula for the correlation implies that ρ is between $1 - \exp(-\gamma)$ and $\exp(\gamma) - 1$.

In this note, the implications of item response models on the parameter γ are considered for models with a one-dimensional ability variable. One basic result, Theorem 1, is that γ must be positive for an item response model with an ability distribution not concentrated at a single point and with monotone increasing item characteristic curves for the items. Theorems 2 and 3 provide bounds on γ for the Rasch model. Similar bounds are also considered for the 2PL model. The bounds are especially easy to apply if the ability distribution is normal.

To define the implications of an item response model on the log cross-product ratio γ, let the ability variable θ be a real random variable with distribution function F, let the X_j be conditionally independent given θ, let $P_j(\theta) > 0$ be the probability that $X_j = 1$ given θ, let $Q_j(\theta) = 1 - P_j(\theta) > 0$ be the corresponding probability that $X_j = 0$ given θ, and let

$$\lambda_j(\theta) = \log \frac{P_j(\theta)}{Q_j(\theta)}$$

be the item logit function (ILF) of X_j. Then the probability $p(\mathbf{x})$ satisfies

$$p(\mathbf{x}) = \int \prod_{j=1}^{2} P_j^{x_j} Q_j^{1-x_j} dF. \tag{2}$$

The log cross-product ratio can be expressed in terms of the cumulant generating function of the item logit functions λ_j conditional on X_1 and X_2 both being 0. To verify this claim, let $\boldsymbol{\lambda}$ be the vector of λ_j, $1 \le j \le 2$, and let $\mathbf{x}'\boldsymbol{\lambda}$ be $\sum_{j=1}^{2} x_j \lambda_j$. Let V be $Q_1 Q_2$. Let t_1 and t_2 be real numbers, and let $\mathbf{t}$ be the two-dimensional vector with coordinates t_1 and t_2. Let $\mathbf{0} = (0, 0)$. Let

$$M(\mathbf{t}) = M(t_1, t_2) = [p(\mathbf{0})]^{-1} \int \exp(\mathbf{t}'\boldsymbol{\lambda}) V dF \tag{3}$$

be the moment generating function at $\mathbf{t}$ of $\boldsymbol{\lambda}(\theta)$ given $\mathbf{X} = \mathbf{0}$, and let $C(\mathbf{t}) = C(t_1, t_2) = \log M(\mathbf{t})$ be the conditional cumulant generating function of $\boldsymbol{\lambda}(\theta)$ at $\mathbf{t}$ given that $\mathbf{X} = \mathbf{0}$. The Dutch identity (Holland, 1990) states that

$$p(\mathbf{x}) = p(\mathbf{0}) M(\mathbf{x}). \tag{4}$$

By (4),

$$\gamma = C(1,1) - C(1,0) - C(0,1) + C(0,0). \tag{5}$$

Thus the log cross-product ratio can be expressed in terms of the cumulant generating function C.

The cumulant generating function C is closely related to the conditional cumulants of $\boldsymbol{\lambda}(\theta)$ given $\mathbf{X} = \mathbf{0}$. Let I be the set of integer pairs $\mathbf{i} = (i_1, i_2)$ such that i_1 and i_2 are nonnegative and at least one of i_1 and i_2 is positive, and let J be the set of $\mathbf{i}$ in I with both i_1 and i_2 positive. Provided that, for some real $r_M > 0$, $M(\mathbf{t})$ is finite whenever $|\mathbf{t}|^2 = t_1^2 + t_2^2 < r_M^2$, there exists an $r_C > 0$ such that, for $|\mathbf{t}| < r_C$,

$$C(\mathbf{t}) = \sum_{\mathbf{i} \in I} \frac{\kappa_{\mathbf{i}} t_1^{i_1} t_2^{i_2}}{i_1! i_2!}.$$

The coefficient $\kappa_{\mathbf{i}} = \kappa_{i_1 i_2}$ is the conditional product cumulant of $\boldsymbol{\lambda}(\theta)$ given $\mathbf{X} = \mathbf{0}$ corresponding to the conditional expectation of $\lambda_1^{i_1}(\theta)\lambda_2^{i_2}(\theta)$ given $\mathbf{X} = \mathbf{0}$. For example, κ_{01} is the conditional mean of $\lambda_1(\theta)$, κ_{20} is the conditional variance of $\lambda_1(\theta)$, and $\kappa_{11}(\theta)$ is the conditional covariance of $\lambda_1(\theta)$ and $\lambda_2(\theta)$. If $r_C > 2^{1/2}$, then γ has the power series expansion

$$\gamma = \sum_{\mathbf{i} \in J} \frac{\kappa_{\mathbf{i}}}{i_1! i_2!}.$$

The expansion suggests a crude approximation of γ by the conditional covariance κ_{11} of $\lambda_1(\theta)$ and $\lambda_2(\theta)$, with a more refined approximation by $\kappa_{11} + (\kappa_{21} + \kappa_{12})/2$. If the conditional distribution of $\boldsymbol{\lambda}(\theta)$ given $X_1 = X_2 = 0$ is bivariate normal, then γ is exactly equal to κ_{11}. For any distribution of θ, if $\lambda_1(\theta)$ or $\lambda_2(\theta)$ is constant, so that X_1 or X_2 is independent of θ, then $\gamma = \kappa_{11} = 0$.

A more general expression for γ can be derived by consideration of a new random vector based on $\boldsymbol{\lambda}$. This result is always available. Standard convexity properties of moment generating functions imply that $M(\mathbf{t})$ is finite for $0 \le t_j \le 1$, $1 \le j \le 2$. As a consequence, if $0 < t_j < 1$ for $1 \le j \le 2$ and $A(\mathbf{t})$ is a random variable with distribution function

$$[M(\mathbf{t})p(\mathbf{0})]^{-1} \int_{-\infty}^{x} \exp(\mathbf{t}'\boldsymbol{\lambda}) V dF$$

at x real, then the two-dimensional random vector $\boldsymbol{\Lambda}(\mathbf{t})$ with coordinates $\Lambda_j(\mathbf{t}) = \lambda_j(A(\mathbf{t}))$ for $1 \le j \le 2$ has finite moments of all orders. In particular, the expectation $\mu_j(\mathbf{t})$ of $\Lambda_j(\mathbf{t})$,

$1 \leq j \leq 2$, and the covariance $\tau(\mathbf{t})$ of $\Lambda_1(\mathbf{t})$ and $\Lambda_2(\mathbf{t})$ are defined and finite. The moment generating function of $\mathbf{\Lambda}(\mathbf{t})$ is $M(\mathbf{t}+\mathbf{u})/M(\mathbf{t})$ at $\mathbf{u}$ if $0 < u_j + t_j < 1$ for $1 \leq j \leq 2$. To explore required integrals derived from C, let $\mathbf{T}$ be uniformly distributed on the unit square S of $\mathbf{t}$ with $0 < t_j < 1$ for $1 \leq j \leq 2$.

Use of the mean value theorem of calculus shows that

$$\gamma = E(\tau(\mathbf{T}))$$

is the expected conditional covariance of $\Lambda_1(\mathbf{T})$ and $\Lambda_2(\mathbf{T})$ given $\mathbf{T}$.

The following theorem provides a simple condition that ensures that log cross-product ratio γ is positive. It is already known that $\gamma \geq 0$ (Holland & Rosenbaum, 1986).

Theorem 1 *Let λ_1 and λ_2 be monotone increasing functions. Assume that no constant c exists such that θ is c with a probability of 1. Then γ is positive.*

Proof. It suffices to show that $\tau(\mathbf{t})$ is positive. To verify that $\tau(\mathbf{t})$ is positive, let $A'(\mathbf{t})$ be a random variable independent of $A(\mathbf{t})$ with the same distribution as $A(\mathbf{t})$. Then $2\tau(\mathbf{t})$ is the expected value of the product

$$U = [\lambda_1(A(\mathbf{t})) - \lambda_1(A'(\mathbf{t}))][\lambda_2(A(\mathbf{t})) - \lambda_2(A'(\mathbf{t}))].$$

If λ_1 and λ_2 are monotone increasing, then U is nonnegative, and the probability is positive that U is positive. $\|$

In the special case of a Rasch model with item difficulties β_j for j from 1 to 2, $\lambda_j(\theta) = \theta - \beta_j$, so that $\tau(\mathbf{t})$ reduces to the variance of $A(\mathbf{t})$, and γ is the expected conditional variance of $A(\mathbf{T})$ given $\mathbf{T}$. If the conditional moment generating function of θ given $\mathbf{X} = \mathbf{0}$ is finite in an open interval that includes 0, if κ_i denotes the ith conditional cumulant of θ given $\mathbf{X} = \mathbf{0}$, and if the conditional cumulant generating function K of θ given $\mathbf{X} = \mathbf{0}$ satisfies $K(t)$ is finite and

$$K(t) = \sum_{i=1}^{\infty} \kappa_i t^i / i!$$

for $|t| < r_K$ for some $r_K > 2^{1/2}$, then

$$\gamma = \sum_{i=2}^{\infty} (2^i - 2)\kappa_i / i!.$$

If the conditional distribution of θ given $\mathbf{X} = \mathbf{0}$ is a normal distribution with variance σ_c^2, then $\gamma = \sigma_c^2$.

If θ has a continuous distribution with a positive density f that is twice differentiable, then a lower bound on γ may be obtained as in the following theorem.

Theorem 2 *Let the Rasch model hold, and let θ have continuous positive and twice differentiable density f. Assume that real $\delta > 0$ and real $c > 0$ exist such that the derivative f_1 of f satisfies the condition that*

$$|f_1(z + a)| < cf(z),\ -\infty < z < \infty, |a| < \delta.$$

Let $g = \log f$, let g_1 be the derivative of g, and let g_2 be the second derivative of g. Let $-g_2(A(\mathbf{t}))$ have a finite positive variance $\eta(\mathbf{t})$ for each $\mathbf{t}$ in S. Let $\eta_j(\mathbf{t})$ be the expectation of $P_j(A(\mathbf{t})Q_j(A(\mathbf{t}))$ for j equal 1 or 2. Then

$$\gamma \geq E([\eta(\mathbf{T}) + \eta_1(\mathbf{T}) + \eta_2(\mathbf{T})]^{-1}).$$

Proof. For $\mathbf{t}$ in S, the density of $A(\mathbf{t})$ is

$$h = H^{-1} f \prod_{j=1}^{2} P_j^{t_j} Q_j^{1-t_j},$$

where

$$H = \int f \prod_{j=1}^{2} P_j^{t_j} Q_j^{1-t_j},$$

so that $e = \log h$ has first derivative

$$e_1 = g_1 - \sum_{j=1}^{2} [(1 - t_j) - Q_j]$$

and second derivative

$$e_2 = g_2 - \sum_{j=1}^{2} P_j Q_j.$$

The first derivative of h is

$$h_1 = e_1 h.$$

Consider estimation of the expectation of a random variable Z under the model that $Z-\alpha$ has the distribution of $A(\mathbf{t})$ for some real α with $|\alpha|<\delta$. Apply the Cramér-Rao inequality (Cramér, 1946, p. 475). Then elementary calculations show that

$$\tau(\mathbf{t}) \geq [\eta(\mathbf{t})+\eta_1(\mathbf{t})+\eta_2(\mathbf{t})]^{-1}.$$

The conclusion follows. $\|$

In many cases, an upper bound on γ may be established as in the following theorem.

Theorem 3 *Let the Rasch model hold, and let θ have continuous positive and twice differentiable density f. Let $g=\log f$, let g_1 be the derivative of g, and let g_2 be the second derivative of g. For some constant $b>0$, let $g_2 \leq -b$. Then $\gamma < 1/b$.*

Proof. From the proof of Theorem 2, it follows that $e_2 < -b$. As a consequence, $\log h$ must achieve a maximum at some real z.

For any random variable Y with mean μ and variance σ^2,

$$E([Y-z]^2) = \sigma^2 + (z-\mu)^2.$$

Thus the variance $\tau(\mathbf{t})$ of $A(\mathbf{t})$ does not exceed the expected value of $[A(\mathbf{t})-z]^2$, so that

$$\tau(\mathbf{t}) \leq \int (uvw),$$

where, for a real,

$$u(a) = (a-z)^2,$$

$$v(a) = h(a)/w(a),$$

and

$$w(a) = (2\pi/b)^{-1/2}\exp[-bu(a)/2].$$

Because h and w are density functions,

$$\int w = 1$$

and

$$\int (vw) = 1.$$

Because w is the density function of the normal distribution with mean z and variance b^{-1},

$$\int (uw) = b^{-1}.$$

Use of standard formulas for changes of variables yields

$$\int (uvw) = \int_0^\infty (u^* v^* w^*),$$

$$\int_0^\infty (v^* w^*) = 1,$$

$$\int_0^\infty w^* = 1,$$

and

$$\int_0^\infty (u^* w^*) = b^{-1},$$

where, for $a > 0$,

$$u^*(a) = a,$$

$$v^*(a) = [v(z + a^{1/2}) + v(z - a^{1/2}]/2,$$

and

$$w^*(a) = (2a\pi/b)^{-1/2} \exp(-ba/2)$$

(Cramér, 1946, p. 168).

It follows that, for any real d,

$$\int_0^\infty (u^* - b^{-1})(v^* - d)w^* = \int_0^\infty (u^* v^* w^*) - b^{-1}.$$

The definition of z and the assumptions of the theorem imply that v^* is a decreasing function, so that the choice of $d = v^*(b^{-1})$ implies $y = (u^* - b^{-1})(v^* - d)$ is negative except at b^{-1}, and $y(b^{-1}) = 0$. Thus $\tau(\mathbf{t})$ is less than b^{-1}. It follows that γ is less than b^{-1}. $\|$

In the important special case of θ with a normal distribution with mean μ and variance σ^2, g_2 in Theorem 2 is the constant $-1/\sigma^2$, so that $\eta(\mathbf{t}_2) = 1/\sigma^2$ and $\tau(\mathbf{t})$ and γ are both at least $2\sigma^2/(2 + \sigma^2)$. On the other hand, it also follows from Theorem 3 that γ is less than σ^2. So, for example, for $\sigma^2 = 1$, the lower and upper bounds on γ are $2/3$ and 1, respectively. For fixed μ and σ^2, if $|\beta_j|$ approaches ∞ for j equals 1 and 2, then γ approaches σ^2. The proof of this claim is an application of Scheffé's theorem (Scheffé, 1947). For example, if β_1 and β_2 both approach $-\infty$, then multiplication of the numerator and denominator by

$$\exp[-(1-t_1)\beta_1 - (1-t_2)\beta_2]$$

shows that, in the proof of Theorem 2, h converges to the density of a normal random variable with mean $\mu - (1-t_1) - (1-t_2)$ and variance σ^2. It follows that $\eta_j(\mathbf{t}_2)$ converges to 0, so that the lower bound for γ converges to σ^2. Because σ^2 is also the upper bound for γ, γ converges to σ^2. Minor variations on the same argument apply if some β_j approaches ∞.

The arguments in Theorems 2 and 3 are readily applied to the 2PL model of item response theory. In this case, $\lambda_j(\theta) = a_j(\theta - \beta_j)$ for an item difficulty β_j and an item discrimination $a_j > 0$. The definition of $A(\mathbf{t})$ is changed due to the new definition of the λ_j; however, the remaining changes are quite limited. In Theorem 2, the lower bound becomes

$$a_1 a_2 E([\eta(\mathbf{T}) + a_1^2\eta_1(\mathbf{T}) + a_2^2\eta_2(\mathbf{T})]^{-1}),$$

and the upper bound becomes a_1a_2/b in Theorem 3. In the case of θ with a normal distribution with mean μ and variance σ^2, a lower bound is

$$\frac{4a_1a_2\sigma^2}{4 + (a_1^2 + a_2^2)\sigma^2}$$

and an upper bound is $a_1a_2\sigma^2$. The previous result for the Rasch model is obtained with $a_1 = a_2 = 1$.

3 Example

In a study of assessment of fit of common models in item response theory (Sinharay & Johnson, 2003), prediction of log cross-product ratios for item pairs was examined under the Rasch model for $\mu = 0$ and $\sigma^2 = 1$. The authors reported that the predicted log cross-product ratios (from the Rasch model) among the item pairs fall within a very narrow range, all around 0.73 in their limited simulation study. These results are consistent with the bounds of 2/3 and 1 established in Section 2. To corroborate their findings, numerical integration was employed to compute γ using (1) for this case (Rasch model with $\mu = 0$ and $\sigma^2 = 1$) with β_1 and β_2 on a grid of integer values between -4 and 4. The values of the log cross-product ratios are summarized in Table 1. As evident in Table 1, the smallest γ is observed for $\beta_1 = \beta_2 = 0$. The largest values are obtained for β_1 and β_2 large in magnitude and opposite in sign. The log cross-product ratios show little variation, all falling between 0.71 and 0.90.

Table 1.

Predicted Log Cross-Product Ratios for Item Pairs Under the Rasch Model

Difficulty	Difficulty of Item 2								
of Item 1	-4	-3	-2	-1	0	1	2	3	4
-4	0.90	0.87	0.83	0.81	0.81	0.84	0.88	0.92	0.95
-3	0.87	0.84	0.80	0.78	0.78	0.81	0.85	0.89	0.92
-2	0.83	0.80	0.77	0.75	0.75	0.77	0.81	0.85	0.88
-1	0.81	0.78	0.75	0.72	0.71	0.73	0.77	0.81	0.84
0	0.81	0.78	0.75	0.71	0.70	0.71	0.75	0.78	0.81
1	0.84	0.81	0.77	0.73	0.71	0.72	0.75	0.78	0.81
2	0.88	0.85	0.81	0.77	0.75	0.75	0.77	0.80	0.83
3	0.92	0.89	0.85	0.81	0.78	0.78	0.80	0.84	0.87
4	0.95	0.92	0.88	0.84	0.81	0.81	0.83	0.87	0.90

4 Conclusion

Although results are presented for two items, they obviously apply to more general item response models. Given any one-dimensional model for $J > 2$ binary responses in which local independence holds, the model applies to any two responses. For example, consider $J \geq 2$ items X_j with values 0 or 1. If the X_j are conditionally independent given a normally distributed random variable θ with mean 0 and variance $\sigma^2 > 0$ and if the conditional probability $P_j(\theta)$ that $X_j = 1$ given θ is

$$P_j(\theta) = [1 + \exp(-\theta + \beta_j)]^{-1}$$

for some real β_j, then X_j and X_k are positively correlated for each j and k. Thus X_j is positively correlated with the sum $S = \sum_{k=1}^{J} X_k$, so that the point-biserial correlation is positive for X_j and S. The log cross-product ratio for each pair X_j and X_k is less than σ^2, so that the correlation of X_j and X_k is less than $\exp(\sigma^2) - 1$.

The bounds obtained can provide a basis for elementary model checking. Log cross-product ratios, pairwise item correlations, and point-biserial correlations are readily estimated without use of model assumptions. If one observes negative estimates of item-pair correlations, cross-product ratios or point-biserial correlations for a data set, one can conclude even before fitting an item response model that the data are clearly incompatible with any one-dimensional item response model. Further, observed values of marginal log cross-product ratios that are clearly outside the bounds (computation of which require fitting an item response model) suggested in this paper will indicate the misfit of the item response model employed.

References

Bishop, Y. M. M., Fienberg, S. E., & Holland, P. W. (1975). *Discrete multivariate analysis: Theory and practice.* Cambridge, MA: MIT Press.

Cramér, H. (1946). *Mathematical methods of statistics.* Princeton, NJ: Princeton University Press.

Holland, P. W. (1990). The Dutch identity: A new tool for the study of item response models. *Psychometrika*, *55*, 5–18.

Holland, P. W., & Rosenbaum, P. R. (1986). Conditional association and unidimensionality in montone latent variable models. *The Annals of Statistics*, *14*, 1523–1543.

Scheffé, H. (1947). A useful convergence theorem for probability distributions. *Annals of Mathematical Statistics*, *18*, 434–438.

Sinharay, S., & Johnson, M. (2003). *Simulation studies applying posterior predictive model checking for assessing fit of the common item response theory models* (ETS RR-03-28). Princeton, NJ: ETS.