The Modified Signed Likelihood Ratio Test and its Application to Detect Cheating

Sandip Sinharay and Jens L. Jensen

Educational Testing Service
Funded by Institute of Educational Sciences Grant No. R305D170026

11 October, 2018

Outline

1. Score Differencing
2. Existing Approaches
3. Methods
4. Applications
5. Conclusions

Conference On Test Security 2018

ETS

Cheating Detection

- *Score differencing* is one of the six categories of cheating detection methods listed in Wollack and Schoenig (2018)
  - gain score analysis
  - item preknowledge detection
  - erasure analysis
  - ....

The Problem Setup

Item Set 1	Item Set 2
θ_1	θ_2
$\hat{\theta}_1$	$\hat{\theta}_2$
$\widehat{Var}(\hat{\theta}_1)$	$\widehat{Var}(\hat{\theta}_2)$

Existing Frequentist Methods

- Z or Wald test (e.g., Guo & Drasgow, 2010)
- Likelihood ratio test (LRT; e.g., Finkelman et al., 2010)
- Signed likelihood ratio test (Sinharay, 2017)
- Score test (Klauer & Rettig, 1990; Sinharay, 2017)
- Other methods such as erasure detection index (Wollack et al., 2015)

The Z Statistic

- For testing $H_0 : \theta_1 = \theta_2$ vs $H_1 : \theta_2 > \theta_1$, the Wald/$Z$ statistic is given by

$$Z = \frac{\hat{\theta}_2 - \hat{\theta}_1}{\sqrt{\widehat{\mathrm{Var}}(\hat{\theta}_2) + \widehat{\mathrm{Var}}(\hat{\theta}_1)}}$$

- For the 2PL model,

$$\widehat{\mathrm{Var}}(\hat{\theta}_1) = \left[\sum_i a_i^2 \frac{\exp[a_i(\hat{\theta}_1 - b_i)]}{(1 + \exp[a_i(\hat{\theta}_1 - b_i)])^2} \right]^{-1}$$

Conference On Test Security 2018

ETS

Log-likelihood of θ_1 and θ_2

- Log-likelihood of θ_1 and θ_2: $\ell(\theta_1, \theta_2)$
- For dichotomous items, $\ell(\theta_1, \theta_2) = \sum_i [X_i \log p_i(\theta_1) + (1 - X_i) \log(1 - p_i(\theta_1))] + \sum_j [Y_j \log p_j(\theta_2) + (1 - Y_j) \log(1 - p_j(\theta_2))]$
- Under the 2PL model, $\ell(\theta_1, \theta_2) = \sum_i [X_i a_i(\theta_1 - b_i) - \log(1 + e^{a_i(\theta_1 - b_i)})] + \sum_j [Y_j \tilde{a}_j(\theta_2 - \tilde{b}_j) - \log(1 + e^{\tilde{a}_j(\theta_2 - \tilde{b}_j)})]$

Likelihood Ratio Test Statistic

- To test $H_0 : \theta_1 = \theta_2$ vs $H_1 : \theta_1 \neq \theta_2$, the LRT statistic is given by

$$\Gamma = 2[\ell(\hat{\theta}_1, \hat{\theta}_2) - \ell(\hat{\theta}_0, \hat{\theta}_0)]$$

- The LRT not appropriate for $H_1 : \theta_2 > \theta_1$

Signed Likelihood Ratio Statistic

- To test $H_0 : \theta_1 = \theta_2$ vs $H_1 : \theta_2 > \theta_1$, one can use the signed LRT statistic (Sinharay, 2017)

$$L_s = \begin{cases} \sqrt{\Gamma} \text{ if } \hat{\theta}_2 \geq \hat{\theta}_1, \\ -\sqrt{\Gamma} \text{ if } \hat{\theta}_2 < \hat{\theta}_1 \end{cases}$$

- The L_s statistic $\sim \mathcal{N}(0, 1)$ for long subtests under $H_0 : \theta_1 = \theta_2$

Higher-order Asymptotics

Methods based on *higher-order asymptotics*

- Modified signed likelihood ratio (MSLR) test (Barndorff-Nielson, 1986)
- Lugannani-Rice approximation (Lugannani & Rice, 1980)

have excellent properties (e.g., Pierce & Peters, 1992), especially for small samples, and can be used for score differencing

MSLR Statistic

- Derivation is simple only for the exponential family of distributions
- The MSLR statistic for 2PL+GPCM:

$$L_s + \frac{1}{L_s} \log \frac{Z'}{L_s}$$

- The MSLR statistic $\sim \mathcal{N}(0,1)$ for long subtests under $H_0 : \theta_1 = \theta_2$
- R code for computing the statistic is publicly available

Results for Simulated Data

- In simulation studies, the Type I error rate of the MSLR statistic was very close to the nominal level and the power was satisfactory in comparison to Z and L_s

Level	Z stat	MSLR stat
0.001	0.0233	0.0007
0.01	0.0340	0.0096
0.05	0.0727	0.0500

Results for a Language Test

- Scores of 629 repeaters were available on two forms of an English language test
- 34 dichotomous items in each form
- The 2PL model (operationally used) was used
- The operational item parameter estimates used
- Computed the Z and MSLR statistic

Significance (At 1% level)

Z Statistic	MSLR Statistic	
	Significant	Not Significant
Significant	35 (6%)	6 (1%)
Not Significant	0 (0%)	588 (93%)

Two Examinees

Examinee	Item Set 1		Item Set 2		Z	MSLR
	Raw-1	θ_1	Raw-2	θ_2	stat	stat
1	14	-1.1	24	.5	**2.46**	**2.38**
2	18	-.9	25	.6	**2.36**	2.29

Conclusions

- A new statistic based on higher-order asymptotics suggested for score differencing
- It has a $\mathcal{N}(0, 1)$ null distribution for long tests
- The Type I error rate and power of the statistic were satisfactory in simulations
- A real data example was discussed
- Promises to be useful in cheating detection
- Reference: Sinharay and Jensen (2018), Psychometrika Online First.