On the Equivalence of a Likelihood Ratio of Drasgow, Levine, and Zickar (1996) and the Statistic Based on the Neyman–Pearson Lemma of Belov (2016)

Sandip Sinharay¹

Abstract

Levine and Drasgow (1988) suggested an approach based on the Neyman–Pearson lemma to detect examinees whose response patterns are “aberrant” due to cheating, language issues, and so on. Belov (2016) used the approach of Levine and Drasgow (1988) to suggest a statistic based on the Neyman–Pearson Lemma (SBNPL) to detect item preknowledge when the investigator knows which items are compromised. This brief report proves that the SBNPL of Belov (2016) is equivalent to a statistic suggested for the same purpose by Drasgow, Levine, and Zickar 20 years ago.

Keywords

test security, item preknowledge, item response theory

Item preknowledge refers to some examinees having prior access to test items and/or answers before taking the test. The test items involved in such a case are referred to as “compromised.” Levine and Drasgow (1988) suggested an approach based on the Neyman–Pearson lemma to detect examinees whose response patterns are “aberrant” due to cheating, language issues, and so on. Drasgow, Levine, and Zickar (1996) applied the approach of Levine and Drasgow (1988) to detect several types of aberrant responding including cheating on a known subset of items. Belov (2014, 2016) used the approach of Levine and Drasgow (1988) to suggest a *statistic based on the Neyman–Pearson Lemma* (SBNPL) to detect item preknowledge when the investigator knows which items are compromised. Note that a description of SBNPL is provided only in the Online Appendix of Belov (2016). The goal of this note is to theoretically prove the hitherto unknown result that the SBNPL of Belov (2014, 2016) is essentially equivalent to the likelihood ratio that Drasgow et al. (1996) suggested for detecting cheating on a known subset of items.

¹Pacific Metrics Corporation, Monterey, CA, USA

Corresponding Author:

Sandip Sinharay, Educational Testing Service, 660 Rosedale Road, Princeton, NJ 08541, USA.
Email: ssinharay@ets.org

Notations

Let x_i denote the score of an examinee (with true ability of θ) on item $i, i = 1, 2, \dots, I$, of a test, and let $P_i(\theta)$ denote the probability of a correct answer on item i for ability θ .

Let c denote the (known) set of compromised items administered to the examinee, and x_c denote the collection of scores of the examinee on the items in c ; that is, $x_c = \{x_i, i \in c\}$. Let u denote the set of uncompromised items administered to the examinee. Let x_u denote the collection of scores of the examinee on the items in u . Let x denote the collection of scores of the examinee on all the items, that is, to the items in c and u ; thus, x is the union of x_c and x_u .

Suppose that among the answers of the examinee to compromised items, n_{1c} denotes the number of correct answers and n_{1w} denotes the number of incorrect answers. That is,

$$\sum_{i \in c} x_i = n_{1c} \quad \text{and} \quad \sum_{i \in c} (1 - x_i) = n_{1w}. \quad (1)$$

Note that the case of known compromised items is probably uncommon, but a real example of such a case can be found in several chapters of Cizek and Wollack (in press) and in Sinharay (2016).

The Approach of Levine and Drasgow (1988)

In the approach of Levine and Drasgow (1988), one assumes a probability model $P_{\text{Normal}}(X=x)$ to describe the item-scores of the examinees whose item-responses are normal and another model $P_{\text{Aberrant}}(X=x)$ to describe the item-scores of the examinees whose item-responses are aberrant. Then one computes the likelihood ratio λ as

$$\lambda = \frac{P_{\text{Aberrant}}(X=x)}{P_{\text{Normal}}(X=x)}, \quad (2)$$

and flags an examinee as aberrant if $\lambda > k$ where k is an appropriate critical value that can be obtained by simulating data for several normal examinees. Typically, $P_{\text{Normal}}(X=x)$ is computed as the marginal probability of x under an item response theory (IRT) model. For example, Levine and Drasgow (1988, p. 170) stated that

$$P_{\text{Normal}}(X=x) = \int_{\theta} f(\theta) \prod_{i=1}^I P_i(\theta)^{x_i} [1 - P_i(\theta)]^{(1-x_i)} d\theta, \quad \text{for } f(\theta) = \frac{1}{\sqrt{2\pi}} e^{-\frac{\theta^2}{2}}. \quad (3)$$

The choice of the model for $P_{\text{Aberrant}}(X=x)$ is more complicated and depends on the type of aberrance that is of interest. Levine and Drasgow (1988) suggested two forms of $P_{\text{Aberrant}}(X=x)$, one each to detect spuriously high aberrance and spuriously low aberrance. Of those, the one to detect spuriously high aberrance, referred to as the “spurious high model,”¹ can be used to detect item preknowledge. However, the “spurious high model” involves the assumption that it is unknown exactly which item-scores are spurious/aberrant. Therefore, the computation of $P_{\text{Aberrant}}(X=x)$ for the spurious high model involves the summation over many terms and is cumbersome.

The Likelihood Ratio of Drasgow et al. (1996) for Cheating on a Known Subset of Items

Drasgow et al. (1996) considered under their “Cheating on a Known Subset of Items” section (p. 54) the case where a test includes some items that were previously administered and some

items that are new and the investigator knows which items are previously administered and which are new. They stated that

... exams are sometimes stolen or otherwise compromised (e.g., one person memorizes the first three items, another person memorizes the fourth through sixth items ...). Consequently, answer keys are sometimes available to individuals who wish to cheat during a test readministration. (p. 54)

Thus, their approach covers item preknowledge though they did not explicitly use the term “item preknowledge.” Let us henceforth refer to the “new” items in their approach as “uncompromised” and let us refer to the “previously administered” items in their approach as “compromised.” Also note that the investigator knows which items are compromised in their approach.

Drasgow et al. (1996) suggested computing $P_{\text{Aberrant}}(X=x)$ so that the item-scores on the uncompromised items follow a standard IRT model, but those on the compromised items are 1 with a large probability (p). They defined $P_{\text{Aberrant}}(X=x)$ in their Equations 10 and 11 as

$$P_{\text{Aberrant}}(X=x) = p^{n_{1c}} (1-p)^{n_{1w}} \int_{\theta} f(\theta) \prod_{i \in u} P_i(\theta)^{x_i} [1 - P_i(\theta)]^{(1-x_i)} d\theta. \quad (4)$$

Then, using Equations 2 to 4, Drasgow et al. (1996, p. 55) defined the likelihood ratio for detecting cheating on a known subset of items as

$$\frac{p^{n_{1c}} (1-p)^{n_{1w}} \int_{\theta} f(\theta) \prod_{i \in u} P_i(\theta)^{x_i} [1 - P_i(\theta)]^{(1-x_i)} d\theta}{\int_{\theta} f(\theta) \prod_{i=1}^I P_i(\theta)^{x_i} [1 - P_i(\theta)]^{(1-x_i)} d\theta}. \quad (5)$$

Zickar and Drasgow (1996) used a similar approach to detect faking on a personality instrument. Thus, the likelihood ratio in Equation 5 is a special case of (and involves much less computation than) λ in Equation 2.

The SBNPL of Belov (2016)

Let us consider a grid of 101 equispaced values $\theta_1 = -5, \theta_2 = -4.9, \dots, \theta_{100} = 4.9, \theta_{101} = 5$. Belov (2016) defined the SBNPL as

$$SBNPL = \frac{\sum_{j=1}^{101} f(\theta_j) \prod_{i \in u} P_i(\theta_j)^{x_i} [1 - P_i(\theta_j)]^{(1-x_i)} \prod_{i \in c} 0.95^{x_i} (1 - 0.95)^{(1-x_i)}}{\sum_{j=1}^{101} f(\theta_j) \prod_{i=1}^I P_i(\theta_j)^{x_i} [1 - P_i(\theta_j)]^{(1-x_i)}}. \quad (6)$$

Then, the SBNPL can be expressed as

$$SBNPL = \frac{0.95^{n_{1c}} (1 - 0.95)^{n_{1w}} \sum_{j=1}^{101} f(\theta_j) \prod_{i \in u} P_i(\theta_j)^{x_i} [1 - P_i(\theta_j)]^{(1-x_i)}}{\sum_{j=1}^{101} f(\theta_j) \prod_{i=1}^I P_i(\theta_j)^{x_i} [1 - P_i(\theta_j)]^{(1-x_i)}}, \quad (7)$$

Because an integral is the limit of a sum over a large number of values,

$$\text{SBNPL} \approx \frac{0.95^{n_{lc}} (1 - 0.95)^{n_{lw}} \int_{\theta} f(\theta) \prod_{i \in u} P_i(\theta)^{x_i} [1 - P_i(\theta)]^{(1-x_i)} d\theta}{\int_{\theta} f(\theta) \prod_{i=1}^I P_i(\theta)^{x_i} [1 - P_i(\theta)]^{(1-x_i)} d\theta} . \quad (8)$$

The denominator of the right-hand side (RHS) of the above equation is identical to the RHS of Equation 3, and the numerator represents a model for $P_{\text{Aberrant}}(X=x)$ for item preknowledge on a known set of compromised items.

The Equivalence of SBNPL and the Likelihood Ratio of Drasgow et al. (1996)

A comparison of Equations 8 and 5 reveals that the SBNPL is essentially equivalent to the likelihood ratio of Drasgow et al. (1996) for detecting cheating on a known subset of items for $p = .95$. Also note that Drasgow et al. (1996, p. 50) used numerical integration to approximate their integrals as did Belov (2016)—so the SBNPL of Belov (2016) and the likelihood ratio of Drasgow et al. (1996) are computed in identical manners.

Author's Note

Sandip Sinharay is now employed at Educational Testing Service, Princeton, New Jersey, USA.

Acknowledgment

The author would like to thank Hua Hua Chang, the editor, the two anonymous reviewers, and Matthew Johnson for several helpful comments that led to a significant improvement of the article.

Declaration of Conflicting Interests

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

Funding

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

Note

1. Levine and Drasgow (1988) actually considered the case when 10% item-scores are spurious, but any other percentage only requires trivial changes to the approach.

References

- Belov, D. I. (2014). *Comparison study of item preknowledge detectors* (LSAC Research Report Series No. 14-01). Newtown, PA: Law School Admission Council.
- Belov, D. I. (2016). Comparing the performance of eight item preknowledge detection statistics. *Applied Psychological Measurement*, 40, 83-97.
- Cizek, G. J., & Wollack, J. A. (in press). *Handbook of detecting cheating on tests*. Washington, DC: Routledge.

- Drasgow, F., Levine, M. V., & Zickar, M. J. (1996). Optimal identification of mismeasured individuals. *Applied Measurement in Education, 9*, 47-64.
- Levine, M. V., & Drasgow, F. (1988). Optimal appropriateness index. *Psychometrika, 53*, 161-176.
- Sinharay, S. (2016). Asymptotic corrections of standardized extended caution indices. *Applied Psychological Measurement, 40*, 418-433.
- Zickar, M. J., & Drasgow, F. (1996). Detecting faking on a personality instrument using appropriateness measurement. *Applied Psychological Measurement, 20*, 71-87.