The Philippine Statistician, 2003
Vol. 52, Nos. 1-4, pp. 59-67

Cheating Detection Using the S_2 Copying Index

Leonardo S. Sotaridona[1]

received November, 2002; revised March, 2003

ABSTRACT

Cheating in examination is a serious problem. Some of the unfavorable consequences includes: invalidating the testing result, creating unfairness to the other examinees, and producing inaccurate estimate of item parameters. Several copying indices were developed that can be used to detect cheating in examinations. Recent study showed that the S_2 index (Sotaridona & Meijer, 2003) provides good promise for detecting answer copying. This paper presents an S-Plus function that implements the procedure for computing the S_2 index. An overview of the S_2 index is presented.

Key words: copying index, cheating, S_2 index, Poisson distribution.

1. INTRODUCTION

In academic examinations, multiple-choice tests are often used because they provide an efficient and reliable way of measuring an examinee's proficiency level. A serious problem that may invalidate the test scores is that examinees copy the answers from other examinees. To detect such behavior, both observational and statistical methods can be used (Cizek, 1999). Observational methods rely on a human observer to make an inference that answer copying has occurred, either through the behavior of a test taker (e.g., the glance of a test taker in the direction of another) or through physical evidence (e.g., confiscated cheat sheets). Statistical methods address cheating by evaluating whether the probability of examinees' answers being identical is sufficiently smaller than the probability of similar answers occuring by chance alone.

Several copying indices have been proposed to detect or back up allegation of answer copying. These copying indices are usually based on the similarity of the responses of an examinee suspected of copying answers (examinee c or the copier) with the responses of another examinee (examinee s or the source). Examples are the K index (Holland, 1996) and its variants—K_2 (Sotaridona & Meijer, 2002), the B_m index (Bay, 1995), the g_2 index (Frary, Tideman, & Watts, 1977), and the ω-index (Wollack, 1997). Cizek (1999) gives a comprehensive review of different copying indices.

Although the evidence of answer copying is stronger if based on matching incorrect response than if based on the matching correct response, a lost of information is incurred if matching correct responses are discarded. The S_2 index proposed by Sotaridona and Meijer (2003) incorporates the information in the matching correct answer in addition to the matching incorrect answer. This was done by introducing a suitable weight function that captures the differential information that are contained in each matching correct answer.

[1] Mathematics Department, Western Visayas College of Science and Technology, Iloilo City. Email: l.s.sotaridona@edte.utwente.nl

This paper presents an S-Plus function that implements the S_2 index of copying. The next section gives some details of the S_2 index. An elaborate treatment of S_2 is found in Sotaridona and Meijer (2003).

2. THE S_2 INDEX

Copying indices that are based solely on the matching incorrect answers, such as the K and K_2 indices, discard the additional information about copying that are available in the matching correct answers. By excluding the number of matching correct answers in the analysis of answer copying, it carries an implicit assumption that c completely knows the answer to item $i=1, 2, ..., I$ whenever c and s give a correct response to item i. However, this is not always the case. An examinee may obtain the correct answer to an item by copying or by guessing.

Note that the K and K_2 indices are not sensitive to a copier who is copying only the correct answers of the source. This may be the case when s and c are friends and s shares his or her answers to c on items where he or she is almost sure of the correct answers. Another example is a high-stakes examination where c may bribe s for sharing his/her correctly answered items to c.

The new copying index S_2 was proposed to overcome this limitation. The S_2 index incorporates information about copying that are contained in the matching correct answers in addition to the information in the matching incorrect answers. Note that as used in K and K_2, the evidence of answer copying is 1 if s and c choose the same wrong option to an item, and 0 if they are both correct or their response to an item did not match. For S_2 however, the amount of evidence is 1 if s and c choose the same wrong option to an item, it is δ (to be described below) if s and c are both correct to an item, and 0 otherwise. The variable δ quantifies the amount of correct-answer copying information to an item for a particular s-c pair.

Define the number-incorrect group $r=1, 2, ..., R$ such that examinees $j=1, 2, ..., J_r$ have the same number of wrong answers, and c' indicate the group membership of c. The number of examinees in the number-incorrect group r is denoted by J_r so that $J_{c'}$ is the number of examinees with the same number of wrong answers as examinee c. Consequently, the two-letter index rj will be used to indicate an examinee j in number incorrect group r. Let U_{irj} be the response of examinee rj to item i and let W_s be the set of items, of size w_s, that were answered incorrectly by s.

For each examinee rj, an indicator variable A_{irj} equal to 1 if $U_{irj} = U_{is}$, and 0 otherwise. The item response of s is index by is indicating that s does not belong to any number-incorrect group r. The number of matching incorrect answers of rj and s, denoted by M_{rj} is then defined as

$$M_{rj} = \sum_{i \in W_s} A_{irj}.$$

For a particular s-c pair, M_{rj} is observed for each examinee rj. For simplicity, M_{rj} will be denoted by M when it is not necessary to identify the examinee. Furthermore, let i^* denotes an item that was answered correctly by s, and U_{i^*rj} the response of examinee rj to item i^*. Then, δ_{i^*rj} gives the estimate of copying information on item i^* by examinee rj. The value of δ_{i^*rj} satisfies the inequality

$$1 \geq \delta_{i^*rj} \geq 0$$

that is, $\delta_{i^*rj}=0$ if rj knows the correct answer to item i^* and $\delta_{i^*rj}=1$ if rj is completely ignorant about the correct answer to item i^* (see conditions 1-2 below). The problem is to quantify the amount of knowledge that rj has on i^*. To do this, it is necessary to obtain the probability of rj answering item i^* correctly. Let this probability be P_{i^*rj}. The P_{i^*rj} can be estimated as the proportion of examinees in number-incorrect group r getting the correct answer to item i^*, that is:

$$\hat{P}_{i^*rj} = \frac{\sum_{j=1}^{J_r} A_{i^*rj}}{J_r} \tag{1}$$

where A_{i^*rj} denotes an indicator variable equal to 1 if $U_{i^*rj}=u_{i^*s}$, and 0 otherwise. An alternative to Eq. (1) is using an item response theory (IRT) model (van der Linden & Hambleton, 1997). A copying index based on an IRT model was proposed by Wollack (1997).

Given the estimate of P_{i^*rj}, what remains is to transform this estimate into δ_{i^*rj}. A suitable transformation function, $f(P_{i^*rj})$, satisfy the following conditions:

1. $f(P_{i^*rj})$ approaches 0 as P_{i^*rj} approaches 1; that is, the evidence of answer copying deminishes as P_{i^*rj} approaches 1.

2. $f(P_{i^*rj})$ approaches 1 as P_{i^*rj} approaches 0; that is, the evidence of answer copying approaches 1 if the suspected copier is correct to an item despite low probability of getting the correct answer to such an item.

3. Test with different number of options must have different weight function. Let f and f' be two different weight functions and i^* and $i^{*\prime}$ are items taken from two tests with number of options V and V' such that $V<V'$. Then it holds that $f(P_{i^*rj})>f'(P_{i^*rj})$ whenever $P_{i^*rj}=P_{i^{*\prime}rj}$.

The basis for conditions 1-2 should be clear from the above discussions. Condition 3 arises from the idea that multiple-choice tests with different number of options should have different transformation functions that differ by a factor that is a function of the number of options. This calls for a function that account for the probability of guessing to an item as a scaling factor.

For notational convenience, let g denotes the probability of getting the correct answer to item i by guessing. Note, an often used value of g is 0.20 for a 5-option test and 0.25 for a 4-option test. A sensible function satisfying conditions 1-3 is shown in Eq. (2).

$$\delta_{i^*rj} = f(P_{i^*rj}) = d_1 e^{d_2 P_{i^*rj}}, \tag{2}$$

where

$$d_2 = -\left(\frac{1+g}{g}\right) \text{ and } d_1 = \left(\frac{1+g}{1-g}\right)^{d_2 P_{i*rj}}.$$

Eq. (2) is a monotone decreasing function of P_{i*rj} with g a scaling constant. Let M^*_{rj} denotes the sum of the number of matching incorrect answers and weighted matching correct answers by examinee rj and examinee s. The expression for M^*_{rj} is given by

$$M^*_{rj} = M_{rj} + \sum_{i*} \delta_{i*rj}. \tag{3}$$

In Eq. 3, the contribution of each item to the value of M^*_{rj} is 0 if the response of rj did not match that of s, 1 if the wrong response of rj matches that of s, and δ_{i*rj} if the correct response of rj matches that of s. The value of M^*_{rj} would be large if most of the incorrect responses of rj matches the wrong responses of s or if P_{i*rj} is small and most of the correct responses of rj matches the correct responses of s. The larger the value of M^*_{rj} relative to the number of items, the stronger the evidence of answer copying.

Note that if there are no matching correct answers between s and rj the second term in Eq. 3 sum up to zero and $M^*_{rj}=M_{rj}$. Hence, M_{rj} becomes a special case of M^*_{rj}. On the other hand, if there are no matching incorrect items but only matching correct answers, then $M_{rj}=0$ and

$$M^*_{rj}=\sum_{i*} \delta_{i*rj}.$$

Thus, while M_{rj} is only sensitive to incorrect answer copying, M^*_{rj} is sensitive to both correct and incorrect answer copying.

In reality, the random variable M^*_{rj} is a nonnegative real-valued random variable. M^*_{rj} was treated as an integer by rounding it off to the nearest integer. Although some error is introduced by doing this, the effect is expected to be of minor influence on the effectiveness of the statistic. A Poisson distribution was assumed for M^*_{rj} and the loglinear model was used to estimate its mean μ. Results of a simulation study (Sotaridona & Meijer, 2003) showed that the Poisson distribution has a reasonable fit for M^*_{rj}. Given the estimate of μ, the S_2 index is defined as

$$S_2 = \sum_{w=m^*_{c'c}}^{I} \frac{e^{-\hat{\mu}_{c'}} \hat{\mu}_{c'}^{w}}{w!}$$

where $m^*_{c'c}$ is the sum of the number of matching incorrect and weighted matching correct answers between c and s, and $e \approx 2.7183$. The smaller the value of S_2, the more likely that answer copying occurred. In general, the S_2 can be treated as a one-sided statistical test of answer copying where the analyst has to specify the level of significance α. The null hypothesis being that S_2 is greater than α against the alternative hypothesis that S_2 is less than or equal to α. For allegation of answer copying in a high-stakes examination with the purpose of nullifying the score of c when other evidence of answer copying corroborated such allegation, very small value of α such as .0001 is necessary because the effect of Type I error is very detrimental to the refutation of the alleged examinee.

3. THE S-PLUS FUNCTION

Below is a function written in S-Plus (S-PLUS 2000, MathSoft, Inc.) that implements the procedure for computing the S_2 index.

```
S2.index <-function(k, n, X, sou, sub, v, m)
{
        # The "S2.index function" implements the S2 index proposed by Sotaridona
        # and Meijer (in press).

        #Programmed by: Leonardo S. Sotaridona
        #Date:          November 2002

        #Reference:     Sotaridona, L. S., & Meijer, R. R. (in press). Two new
        #               statistics to detect answer copying. Journal of Educational
        #               Measurement.

        # NOTATIONS
        #  k - the number of items
        #  n - the number of examinees
        #  X - a matrix of response pattern (actual response, not the 0/1). The
        #     rows of X for the items and the columns for the examinees
        #  sou - the column number identifying the source's location in X
        #  sub - the column number identifying the copier's location in X
        #  v - a vector of answer keys, the length of v is the same is the number
        #        of rows of X
        #  m - the number of options

# (1) Number or wrong (nw) and number of matching wrong responses (mw)
        nw      <- matrix(0, 1, n)
        mw      <- matrix(0, 1, n)
        for(i in 1:n) {
                mw[, i]        <-      sum(X[, sou] == X[, i] & X[, sou] != v)
                nw[, i] <-     k-sum(X[, i] ==v)
        }

        u <- matrix(sort(unique(nw)), 1)        # unique number of wrong, needed in
                                                # conditioning

        r <- ncol(u)                            # r is the number of groups
```

```
# (2) Identify examinees with the same number of wrong

        S <- matrix(NA,r,n)
                for(i in 1:r) {
                        for(j in 1:n) {
                                if(u[1,i]==nw[j]) S[i,j] <- j
                }
        }

        Usi <- X[,sou]                  # Usi is the response pattern of the source

# (3) Computes Pij for each group r

        dummy1 <-matrix(0,k,r)
        for(i in 1:r) {
                ind     <- S[i,][S[i,]!="NA"]
                respp1  <- X[,ind]

        # Within a certain group r, determine the examinees with matching
        # answer as the source in every item, note, the correct responses are included

                F1 <- respp1-Usi
                F2 <- F1*2

                F2[F2==0]       <- 1; F2[F2!=1] <- 0
                dim(F2)         <- c(k,length(ind))

                F3 <- apply(F2,1,sum)

                # Prij
                        Prij <- F3/length(ind)

                # weighted evidence is in q.i
                        q.i <- round( ((((m+1)/(m-1))*2.718)^(-(m+1)*Prij)) ,4)

                dummy1[,i]      <- q.i
        }

# (4) Reflecting the weight of evidence

        weight <- matrix(0,k,n)
        for(i in 1:r) {
                for(j in 1:n) {
                        if(u[1,i]==nw[j])
                                weight[,j] <- dummy1[,i]
                }
        }
```

(5) Computing M^*_{rj} (sum of weighted evidence for all the matching correct items)

```
D2 <- weight
XX <- X                          # X is an input matrix of response
                                 # patterns

XX[,sou][ XX[,sou]!=v ] <- 0         # source's incorrect answer is
                                     # replaced with 0
XX     <- XX*XX[,sou]                # just a trict

XX[XX==(v*v)]  <-1                   # source's correct response
                                     # is replaced with 1
XX[XX!=1]  <-0


# the sum of weighted evidence is in Mrij=M*ij

    Mrij <- matrix( apply(XX*D2,2,sum) ,1 ,n)

# combining evidence: matching wrong plus weighted
# matching correct

    mw1 <- mw + ceiling(Mrij)
```

(6) Estimation of the loglinear parameters

```
u1 <- matrix(0, 1, r)
z <- matrix(0, 1, r)
for(i in 1:r) {
    for(j in 1:n) {
        if(u[,i] == nw[j]) {

            u1[, i] <- 1 + u1[, i]          # this counts the number
                                            # of examinees with the
                                            # same number-wrong

            z[, i] <- z[, i] + mw[j]        # this computes the sum
                                            # of match wrong which
                                            # which is needed to
        }                                   # obtained the mu
    }
}
```

```
#mu is the mean number of match-wrong for each group r
        mu <- z/u1

uu1 <- matrix(0, 1, r)
zzz <- matrix(0, 1, r)
for(i in 1:r) {
        for(j in 1:n) {
                if(u[, i] == nw[j]) {
                        uu1[, i] <- 1 + uu1[, i] # this counts the number of
                                                 # examinees with the same
                                                 # number-wrong
                        zzz[, i] <- zzz[, i] + mw1[j]   # this computes the sum
                                                        # of match wrong (plus
                }                                       # weighted match-correct)
        }                                               # which is needed to
                                                        # obtained the mean mu
}

# Mean mu1
        mu1 <- zzz/uu1


# Estimation of the mean mu using loglinear model

################################################################################
ff is an array which stores the results of fitting the loglinear model
##      ff[[1]][1] is the first element of the first array in ff3
##      pred is the "predicted value" assuming the model holds; also called the fitted ##
        value
################################################################################

x <-u[1,]       # transform to become a vector, note that u and mu1 are in
                # matrix form
y <-mu[1,]

# Fitting the loglinear model using the "glim" function in S-Plus

        ff <- glim(x, y, error="poisson", link="log", resid="p")

#Estimate of the parameters of the loglinear model
        B0 <- ff3[[1]][1]
        B1 <- ff3[[1]][2]

#Predicted value
        pred <- exp(B0 + B1*u)

#reflecting the predicted value to all examinees

        pred.i <-matrix(0, 1, n)
```

```
            for(i in 1:r) {
                    for(j in 1:n) {
                            if(u[, i] == nw[j])
                                    pred.i[, j] <- pred[i]
                    }
            }

            S2 <- 1 - ppois( mw1[sub], pred.i[sub])

    Out <-list( S2 =S2)
}
# end
```

ACKNOWLEDGMENT

The author thanks the anonymous referee for the comments and suggestions which are very helpful in improving the manuscript.

References

BAY, L. G. (1995). "Detection of Cheating on Multiple-choice Examinations." Paper presented at the annual meeting of the American Educational Research Association, San Francisco.

CIZEK, G. J. (1999). *Cheating on Tests: How to Do It, Detect It, and Prevent It.* NJ: Lawrence Erlbaum.

FRARY, R. B., TIDEMAN, T. N., & WATTS, T. M. (1977). "Indices of Cheating on Multiple-Choice tests." *Journal of Educational Statistics*, 6, 152-165.

HOLLAND, P. W. (1996). "Assessing Unusual Agreement Between the Incorrect Answers of Two Examinees Using the K-index: Statistical Theory and Empirical Support." *ETS Technical Report No. 96-4*. Princeton, NJ: Educational Testing Service.

S-PLUS 2000 *Programmer's Guide and Software*, Data Analysis Products Division, MathSoft, Seatle, WA.

SOTARIDONA, L. S., & MEIJER, R. R. (2002). "Statistical Properties of the K-index for Detecting Answer Copying." *Journal of Educational Measurement*, 39: 115-132.

SOTARIDONA, L. S., & MEIJER, R. R. (2003). "Two New Statistics to Detect Answer Copying." *Journal of Educational Measurement.* March Issue

VAN DER LINDEN, W. J. & HAMBLETON, R. K. (1997). *Handbook of Modern Item Response Theory.* NY: Springer-Verlag.

WOLLACK, J. A. (1997). "A Nominal Response Model Approach to Detect Answer Copying." *Applied Psychological Measurement,* 21, 307-320.