A derivation of the optimal answer-copying index and some applications

Mauricio Romero\textsuperscript{1,2} \ Álvaro Riascos\textsuperscript{2,3} \ Diego Jara\textsuperscript{2}

\textsuperscript{1}UC San Diego
\textsuperscript{2}Quantil
\textsuperscript{3}Universidad de los Andes

June 17 2014
Introduction

- In Colombia every student has to take a multiple-choice exam in the 5th, 9th and 11th grades.
- The Instituto Colombiano para la Evaluación de la Educación (ICFES) is in charge of making, distributing and applying these exams.
- Between 2010 and 2011, the ICFES used the $\kappa$ index for answer-copy detection, based on the work of Sotaridona, van der Linden, and Meijer (2006)
- Jara, Riascos, and Romero (2010) showed that in practice the $\kappa$ index has a larger type-I error rate than predicted by theory
How to choose an answer-copy index?

- All the indices in the literature are ad-hoc and there are no theoretical results justifying the use of one index over the other.
- We could use empirical or simulated type-I and type-II calculations... such as Wollack (2003) but the set of indices compared is not comprehensive (e.g. Wesolowsky (2000) index)
What we do...

- Provide theoretical foundations that justify the use of indices that reject the null hypothesis of no cheating for large values of the number of identical answers.


- Compare results with two strategies to control cheating: stricter proctoring and diversification of questions.

- Outline a procedure to detect massive cheating.
Table of Contents

1 Introduction
2 Applying Neyman-Pearson to answer copying
3 Copy Indices
4 Data
5 Monte Carlo Simulations
6 Results for different cheating strategies
   • Proctoring
   • Diversification of questions
7 Massive cheating
8 Conclusions
Multiple choice exams are frequently used as an efficient and objective way of evaluating knowledge.

But... they are more vulnerable to answer-copying than tests based on open questions.

It is normal for two answer patterns to have similarities.

Answer-copy indices try to detect similarities so unlikely that answer-copying becomes a more natural explanation than chance.
Setting up of the problem

- Test if individual suspected of cheating (denoted by \( c \)) copied from the individual who is suspected of being the source of answers (denoted by \( s \))
- Assume that there are \( N \) questions and \( n \) alternatives for each question
- \( \gamma_{cs} \rightarrow \) number of questions that \( s \) copied from \( c \).

\[
H_0 : \gamma_{cs} = 0 \\
H_1 : \gamma_{cs} > 0
\]
Setting up of the problem

- Let \( I_{csi} \) be equal to one when individuals \( c \) and \( s \) have the same answer to question \( i \) and zero otherwise

\[
M_{cs} = \sum_{i=1}^{N} I_{csi}. \tag{1}
\]

- Under \( H_0 \) \( M_{cs} \sim B(\pi_1, \ldots, \pi_N) \) (i.e. is distributed Poisson binomial)

- Let \( f_N(x; \pi_1, \ldots, \pi_N) \) be the probability mass function (pmf) at \( x \)

- If \( \pi_1 = \pi_2 = \ldots = \pi_N \) then we have a binomial distribution
Setting up of the problem

- Let $A$ denote the set of questions that student $c$ copied from $s$.
- If $|A| = k$, it means that $\gamma_{cs} = k$
- $M_{cs}$ has the following probability mass function (pmf)

$$\hat{f}_N(x; \pi_1, ..., \pi_N, A),$$

\[
\hat{f}_N(x; \pi_1, ..., \pi_N, A) = f_N(x, \pi', ..., \pi_N')
\]

s.t.

$$\pi'_i = \begin{cases} 
1 & \text{if } i \in A \\
\pi_i & \text{if } i \notin A
\end{cases}$$
Neyman-Pearson Lemma

Consider testing $H_0 : \theta = \theta_0$ against $H_1 : \theta = \theta_1$ where the pmf is $f(x|\theta_i)$, $i = 0, 1$, using a test with rejection region $R$ that satisfies

$x \in R$ if $f(x|\theta_1) > f(x|\theta_0)k$ \hspace{1cm} (2)

$x \in R^c$ if $f(x|\theta_1) < f(x|\theta_0)k$

for some $k \geq 0$, and

$\alpha = P_{H_0}(X \in R)$ \hspace{1cm} (3)

Then

1. (Sufficiency) Any test that satisfies 2 and 3 is a UMP level $\alpha$ test.
2. (Necessity) If there exists a test satisfying 2 and 3 with $k > 0$, then every UMP level $\alpha$ test is a size $\alpha$ test (satisfies 3) and every UMP level $\alpha$ test satisfies 2 except perhaps on a set $A$ satisfying

$P_{H_0}(X \in A) = P_{H_1}(X \in A) = 0$.

the test is the uniformly most powerful (UMP) level $\alpha$ test.
Simple hypothesis

- Simple hypothesis test $H_0 : A = \emptyset$ and $H_1 : A = A_1$

$$\lambda^A(x) = \frac{\hat{f}_N(x; \pi_1, \ldots, \pi_N, A)}{f_N(x; \pi_1, \ldots, \pi_N)}$$

- To find the critical value of the test we need the greatest value $c$ such that:

$$1 - P_{H_0} \left( \frac{\hat{f}_N(x; \pi_1, \ldots, \pi_N, A)}{f(x; \pi_1, \ldots, \pi_N)} < c \right) = P_{H_0} \left( \frac{\hat{f}_N(x; \pi_1, \ldots, \pi_N, A)}{f_N(x; \pi_1, \ldots, \pi_N)} > c \right) \leq \alpha$$

- How to find then UMP for more complex alternative hypothesis (in particular $H_1 : \{ A : |A| \geq 1 \}$)?

Romero, Riascos & Jara
The pmf of a poisson binomial satisfies the following inequality:

\[ f_N(x; \pi_1, \pi_2, \ldots, \pi_N)^2 > C(x)f_N(x + 1; \pi_1, \pi_2, \ldots, \pi_N)f_N(x - 1; \pi_1, \pi_2, \ldots, \pi_N) \]

where \( C(x) = \max \left( \frac{x+1}{x}, \frac{N-x+1}{N-x} \right) \)

Lemma

\( \frac{\hat{f}_N(x; \pi_1, \ldots, \pi_N, A)}{f_N(x, \pi_1, \ldots, \pi_N)} \) is increasing in \( x \in \{0, \ldots, N\} \) for all \( A \).
Complex Hypothesis

Given that \( \frac{\hat{f}(x;\pi_1,\ldots,\pi_N;A)}{f(x;\pi_1,\ldots,\pi_N)} \) is increasing in \( x \) for all \( A \) \( \Rightarrow \) for every \( c \) there exists a \( k^* \) such that

\[
P_{H_0} \left( \frac{\hat{f}_N(x; \pi_1, \ldots, \pi_N, A)}{f(x; \pi_1, \ldots, \pi_N)} < c \right) = \sum_{w=0}^{k^*} f(w, \pi_1, \ldots, \pi_N)
\]

Remark

Notice that the rejection region is the same for all \( A \), thus if we reject the null hypothesis when \( M_{cs} > k^* \), we get the UMP for all \( A \) such that \( |A| \geq 1 \).
But...

However, $\pi_i$ must be estimated somehow (it was taken as known in this section and thus in empirical applications we don’t really have the UMP) and different ways to go about this yield different results.
Table of Contents

1 Introduction
2 Applying Neyman-Pearson to answer copying
3 Copy Indices
4 Data
5 Monte Carlo Simulations
6 Results for different cheating strategies
   - Proctoring
   - Diversification of questions
7 Massive cheating
8 Conclusions
Definitions

- $\pi^j_{iv} \rightarrow$ probability student $j$ answers option $v$ on question $i$.
- $\pi_i \rightarrow$ probability two students have the same answer on question $i$.
- $\pi^c_{ivs} \rightarrow$ probability that individual $c$ answered the option $v_s$ which was chosen by $s$ in question $i$. 
Assuming independent answers: $\pi_i = \sum_{v=1}^{n} \pi_{iv}^c \pi_{iv}^s$.

Conditional on the answers of $s$, the probability that individual $c$ has the same answer as individual $s$: $\pi_i = \pi_{ivs}^c$
Critical Values: Poisson Binomial vs Normal

- $M_{cs}$ is the sum of $n$ Bernoulli variables and has mean $\sum_{i=1}^{N} \pi_i$ and variance $\sum_{i=1}^{N} \pi_i(1 - \pi_i)$.

- $\frac{M_{cs} - \sum_{i=1}^{N} \pi_i}{\sqrt{\sum_{i=1}^{N} \pi_i(1-\pi_i)}} \to d N(0, 1)$
How to calculate $\pi_{iv}^j$?

The main difference between indices is how they calculate $\pi_{iv}^j$. 
The $\omega$ index is based on the work of Wollack (1997)

$$
\pi_{iv}(\theta_j) = \frac{e^{\xi_{iv}+\lambda_{iv}\theta_j}}{\sum_{h=1}^{m} e^{\xi_{ih}+\lambda_{ih}\theta_j}},
$$

1. $\omega_1$: unconditional
2. $\omega_2$: conditional
3. $\omega_1^s$: unconditional standardized
4. $\omega_2^s$: conditional standardized
The $\gamma$ index is based on the work of Wesolowsky (2000)

- $r_i$ is the proportion of students that got the right answer in question $i$
- $c_j$ is the proportion of questions answered correctly by individual $j$
- $p_i = (1 - (1 - r_i)^{a_j})^{1/a_j} \rightarrow$ Probability that student $j$ has the correct answer in question $i$
Applying Neyman-Pearson to answer copying

Copy Indices

Monte Carlo Simulations

Results for different cheating strategies

Massive cheating

Conclusions

\( \gamma \) index

Detecting excessive similarity in answers

\( p_{ji} \)

\( r_i \)

\( a_j > 1 \)

\( a_j = 1 \)

\( a_j < 1 \)

Fig. 3. Expression (2).
\[ a_j \text{ is estimated by solving the equations } \frac{\sum_{i=1}^{N} p_i}{n} = c_j \]

To estimate the probability of incorrect options we find the proportion of students that answered each incorrect option.

1. \( \gamma_1 \): unconditional
2. \( \gamma_2 \): conditional
3. \( \gamma_1^s \): unconditional standardized
4. \( \gamma_2^s \): conditional standardized
Data

- 5th and 9th grade tests for 2009
- 12 different exams that can be classified in three broad fields, Science, Mathematics and Language for each grade (5th and 9th) for two different dates (May and September, 2009)
- Answer patterns for each student and answer keys
## Summary Statistics

<table>
<thead>
<tr>
<th>Test</th>
<th>Subject</th>
<th>Grade</th>
<th>Month</th>
<th>Questions</th>
<th>Students</th>
<th>Schools</th>
</tr>
</thead>
<tbody>
<tr>
<td>PBA5041F1</td>
<td>Math</td>
<td>5th</td>
<td>May</td>
<td>48</td>
<td>60,099</td>
<td>3,421</td>
</tr>
<tr>
<td>PBA5041F2</td>
<td>Math</td>
<td>5th</td>
<td>Oct</td>
<td>48</td>
<td>403,624</td>
<td>31,827</td>
</tr>
<tr>
<td>PBA5042F1</td>
<td>Language</td>
<td>5th</td>
<td>May</td>
<td>36</td>
<td>60,455</td>
<td>3,441</td>
</tr>
<tr>
<td>PBA5042F2</td>
<td>Language</td>
<td>5th</td>
<td>Oct</td>
<td>36</td>
<td>402,508</td>
<td>31,642</td>
</tr>
<tr>
<td>PBA5043F1</td>
<td>Science</td>
<td>5th</td>
<td>May</td>
<td>48</td>
<td>60,404</td>
<td>3,432</td>
</tr>
<tr>
<td>PBA5043F2</td>
<td>Science</td>
<td>5th</td>
<td>Oct</td>
<td>48</td>
<td>405,537</td>
<td>31,833</td>
</tr>
<tr>
<td>PBA9041F1</td>
<td>Math</td>
<td>9th</td>
<td>May</td>
<td>54</td>
<td>44,577</td>
<td>1,110</td>
</tr>
<tr>
<td>PBA9041F2</td>
<td>Math</td>
<td>9th</td>
<td>Oct</td>
<td>54</td>
<td>303,233</td>
<td>9,059</td>
</tr>
<tr>
<td>PBA9042F1</td>
<td>Language</td>
<td>9th</td>
<td>May</td>
<td>54</td>
<td>44,876</td>
<td>1,110</td>
</tr>
<tr>
<td>PBA9042F2</td>
<td>Language</td>
<td>9th</td>
<td>Oct</td>
<td>54</td>
<td>302,781</td>
<td>9,044</td>
</tr>
<tr>
<td>PBA9043F1</td>
<td>Science</td>
<td>9th</td>
<td>May</td>
<td>54</td>
<td>44,820</td>
<td>1,107</td>
</tr>
<tr>
<td>PBA9043F2</td>
<td>Science</td>
<td>9th</td>
<td>Oct</td>
<td>54</td>
<td>30,3723</td>
<td>9,053</td>
</tr>
</tbody>
</table>

Source: ICFES. Calculations: Authors.

Note: The number of schools corresponds to the number of examination rooms.
100,000 pairs are picked in such a way that for each couple, the individuals performed the exam in different examination rooms.

The answer-copy methodology is applied to these pairs, and the proportion of pairs accused of cheating is the empirical type-I error rate estimator.

To calculate the power of the index, the answer-pattern for individual \( c \) is changed by replacing \( k \) of his answers to correspond to those of the individual \( s \).

1. The level of copy, \( k \), is set, and is defined as the number of answers transferred from \( s \) to \( c \).
2. \( k \) questions are selected randomly.
3. Individual \( c \)'s answers for the \( k \) questions are changed to replicate exactly those of individual \( s \). Answers that were originally identical count as part of the \( k \) questions being changed.

We apply the answer-copy methodology to the tampered couples. The proportion of pairs accused of cheating is the power of the index for a copying level of \( k \).
## Type-I error: $\gamma$ index

<table>
<thead>
<tr>
<th>Exam</th>
<th>Subject</th>
<th>Grade</th>
<th>Month</th>
<th>$\gamma_1$</th>
<th>$\gamma_2$</th>
<th>$\gamma_1^s$</th>
<th>$\gamma_2^s$</th>
</tr>
</thead>
<tbody>
<tr>
<td>PBA5041F1</td>
<td>Mathematics</td>
<td>5th</td>
<td>May</td>
<td>0.66</td>
<td>2.20</td>
<td>0.41</td>
<td>0.76</td>
</tr>
<tr>
<td>PBA5041F2</td>
<td>Mathematics</td>
<td>5th</td>
<td>October</td>
<td>0.87</td>
<td>2.44</td>
<td>0.59</td>
<td>1.11</td>
</tr>
<tr>
<td>PBA5042F1</td>
<td>Language</td>
<td>5th</td>
<td>May</td>
<td>1.20</td>
<td>2.18</td>
<td>0.77</td>
<td>1.16</td>
</tr>
<tr>
<td>PBA5042F2</td>
<td>Language</td>
<td>5th</td>
<td>October</td>
<td>1.21</td>
<td>2.36</td>
<td>0.92</td>
<td>1.49</td>
</tr>
<tr>
<td>PBA5043F1</td>
<td>Science</td>
<td>5th</td>
<td>May</td>
<td>1.05</td>
<td>2.59</td>
<td>0.73</td>
<td>1.38</td>
</tr>
<tr>
<td>PBA5043F2</td>
<td>Science</td>
<td>5th</td>
<td>October</td>
<td>0.74</td>
<td>1.81</td>
<td>0.61</td>
<td>1.24</td>
</tr>
<tr>
<td>PBA9041F1</td>
<td>Mathematics</td>
<td>9th</td>
<td>May</td>
<td>1.38</td>
<td>1.97</td>
<td>0.96</td>
<td>1.26</td>
</tr>
<tr>
<td>PBA9041F2</td>
<td>Mathematics</td>
<td>9th</td>
<td>October</td>
<td>2.15</td>
<td>2.14</td>
<td>1.69</td>
<td>1.53</td>
</tr>
<tr>
<td>PBA9042F1</td>
<td>Language</td>
<td>9th</td>
<td>May</td>
<td>0.85</td>
<td>2.24</td>
<td>0.56</td>
<td>1.04</td>
</tr>
<tr>
<td>PBA9042F2</td>
<td>Language</td>
<td>9th</td>
<td>October</td>
<td>0.84</td>
<td>1.92</td>
<td>0.59</td>
<td>1.34</td>
</tr>
<tr>
<td>PBA9043F1</td>
<td>Science</td>
<td>9th</td>
<td>May</td>
<td>1.32</td>
<td>2.06</td>
<td>0.93</td>
<td>1.42</td>
</tr>
<tr>
<td>PBA9043F2</td>
<td>Science</td>
<td>9th</td>
<td>October</td>
<td>1.02</td>
<td>1.70</td>
<td>0.74</td>
<td>1.37</td>
</tr>
</tbody>
</table>

Source: ICFES. Calculations: Authors.

Number of copy-free couples accused of copying (for every 1000 pairs) at a 99.9% confidence level.
### Type-I error: $\omega$ index

<table>
<thead>
<tr>
<th>Exam</th>
<th>Subject</th>
<th>Grade</th>
<th>Month</th>
<th>$\omega_1$</th>
<th>$\omega_2$</th>
<th>$\omega_1^s$</th>
<th>$\omega_2^s$</th>
</tr>
</thead>
<tbody>
<tr>
<td>PBA5041F1</td>
<td>Mathematics</td>
<td>5th</td>
<td>May</td>
<td>0.42</td>
<td>1.28</td>
<td>0.23</td>
<td>0.52</td>
</tr>
<tr>
<td>PBA5041F2</td>
<td>Mathematics</td>
<td>5th</td>
<td>October</td>
<td>0.61</td>
<td>1.38</td>
<td>0.31</td>
<td>0.78</td>
</tr>
<tr>
<td>PBA5042F1</td>
<td>Language</td>
<td>5th</td>
<td>May</td>
<td>0.80</td>
<td>1.61</td>
<td>0.46</td>
<td>0.73</td>
</tr>
<tr>
<td>PBA5042F2</td>
<td>Language</td>
<td>5th</td>
<td>October</td>
<td>0.86</td>
<td>1.51</td>
<td>0.55</td>
<td>0.95</td>
</tr>
<tr>
<td>PBA5043F1</td>
<td>Science</td>
<td>5th</td>
<td>May</td>
<td>0.79</td>
<td>1.37</td>
<td>0.47</td>
<td>0.87</td>
</tr>
<tr>
<td>PBA5043F2</td>
<td>Science</td>
<td>5th</td>
<td>October</td>
<td>0.82</td>
<td>1.47</td>
<td>0.57</td>
<td>0.88</td>
</tr>
<tr>
<td>PBA9041F1</td>
<td>Mathematics</td>
<td>9th</td>
<td>May</td>
<td>0.89</td>
<td>1.53</td>
<td>0.58</td>
<td>0.89</td>
</tr>
<tr>
<td>PBA9041F2</td>
<td>Mathematics</td>
<td>9th</td>
<td>October</td>
<td>1.22</td>
<td>1.53</td>
<td>0.99</td>
<td>1.07</td>
</tr>
<tr>
<td>PBA9042F1</td>
<td>Language</td>
<td>9th</td>
<td>May</td>
<td>0.55</td>
<td>1.44</td>
<td>0.31</td>
<td>0.65</td>
</tr>
<tr>
<td>PBA9042F2</td>
<td>Language</td>
<td>9th</td>
<td>October</td>
<td>0.86</td>
<td>1.47</td>
<td>0.63</td>
<td>0.97</td>
</tr>
<tr>
<td>PBA9043F1</td>
<td>Science</td>
<td>9th</td>
<td>May</td>
<td>0.78</td>
<td>1.46</td>
<td>0.59</td>
<td>0.98</td>
</tr>
<tr>
<td>PBA9043F2</td>
<td>Science</td>
<td>9th</td>
<td>October</td>
<td>0.78</td>
<td>1.36</td>
<td>0.63</td>
<td>1.03</td>
</tr>
</tbody>
</table>

Source: ICFES. Calculations: Authors.

Number of copy-free couples accused of copying (for every 1000 pairs)
at a 99.9% confidence level
Exam PBA5041F1

Proportion of answers copied

Power

- $\gamma_1$
- $\gamma_1^s$
- $\omega_1$
- $\omega_1^s$
- $\omega_2$

Proportion of answers copied
The ICFES *randomly* assigns schools to three different samples that have different levels of proctoring.

Most of the schools are assigned to the *censal* sample where the ICFES distributes the exams to the schools and the schools perform the proctoring.

In the *controlada* the proctoring is done by the central government (i.e. the ICFES).

In the *estadistica* the proctoring is done by the regional government (Secretarias de Educacion).
<table>
<thead>
<tr>
<th></th>
<th>Controlada</th>
<th>Estadistica</th>
<th>Censal</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>5th Grade May</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No. Students</td>
<td>1,413</td>
<td>7,648</td>
<td>60,099</td>
</tr>
<tr>
<td>No. of Schools</td>
<td>141</td>
<td>680</td>
<td>3,421</td>
</tr>
<tr>
<td>Students/School</td>
<td>10.02</td>
<td>11.25</td>
<td>17.57</td>
</tr>
<tr>
<td>(0.88)</td>
<td>(0.47)</td>
<td>(0.46)</td>
<td></td>
</tr>
<tr>
<td><strong>5th Grade October</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No. Students</td>
<td>3,830</td>
<td>26,393</td>
<td>403,624</td>
</tr>
<tr>
<td>No. of Schools</td>
<td>958</td>
<td>654</td>
<td>31,827</td>
</tr>
<tr>
<td>Students/School</td>
<td>4.00</td>
<td>40.36</td>
<td>12.68</td>
</tr>
<tr>
<td>(0.13)</td>
<td>(1.36)</td>
<td>(0.11)</td>
<td></td>
</tr>
<tr>
<td><strong>9th Grade May</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No. Students</td>
<td>1,150</td>
<td>6,690</td>
<td>44,577</td>
</tr>
<tr>
<td>No. of Schools</td>
<td>75</td>
<td>351</td>
<td>1110</td>
</tr>
<tr>
<td>Students/School</td>
<td>15.33</td>
<td>19.06</td>
<td>40.16</td>
</tr>
<tr>
<td>(1.62)</td>
<td>(1.08)</td>
<td>(1.44)</td>
<td></td>
</tr>
<tr>
<td><strong>9th Grade October</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No. Students</td>
<td>3,106</td>
<td>24,387</td>
<td>303,233</td>
</tr>
<tr>
<td>No. of Schools</td>
<td>495</td>
<td>487</td>
<td>9,059</td>
</tr>
<tr>
<td>Students/School</td>
<td>6.27</td>
<td>50.08</td>
<td>33.47</td>
</tr>
<tr>
<td>(0.25)</td>
<td>(1.74)</td>
<td>(0.35)</td>
<td></td>
</tr>
</tbody>
</table>
Proportion of couples accused of copying
The SABER tests are administrated over three sessions, wherein students answer a different subject in each session.

In May, every student took the same subject at the same time.

In October, only one third of the students took the same subject in each session.
# Table of Contents

1. Introduction
2. Applying Neyman-Pearson to answer copying
3. Copy Indices
4. Data
5. Monte Carlo Simulations
6. Results for different cheating strategies
   - Proctoring
   - Diversification of questions
7. Massive cheating
8. Conclusions
To determine whether massive cheating has occurred in an examination room, multiple hypotheses must be tested.

The significance level for a multiple test

$$\alpha_{MT} = 1 - (1 - \alpha_I)^n \leq \alpha_I \cdot n$$ (assuming independence)

If this “correction” is made, in most cases the power of the test is severely diminished

Solution: Bonferroni-Type procedures that control the false positive rate
Suppose there are $H_1, ..., H_m$ hypotheses to be tested, ordered such that their $p$-values follow $P_1 \leq P_2 \leq ... \leq P_m$, where $P_i$ is the $p$-value of hypothesis $H_i$. Let $k$ be the greatest integer $i$, such that:

$$P_i \leq \frac{i}{m} p^*.$$  \hspace{1cm} (5)

$H_i$ is then rejected for every $i \in \{1, ..., k\}$. This controls for the false positive rate to a maximum of $p^*$ (Benjamini & Hochberg, 1995)
Examination rooms suspect of massive cheating
(>60% of students suspected of copying)
Table of Contents

1. Introduction
2. Applying Neyman-Pearson to answer copying
3. Copy Indices
4. Data
5. Monte Carlo Simulations
6. Results for different cheating strategies
   - Proctoring
   - Diversification of questions
7. Massive cheating
8. Conclusions
Conclusions

- Theoretical justification of the use of a variety of statistical test found in the literature to detect answer-copying.
- The most powerful test is a conditional test that models student behavior using a nominal response model and relies on the central limit theorem to find critical values.
Conclusions

- Increasing the level of proctoring in this setting can halve the prevalence of cheating.
- Randomizing the time at which each student takes each portion of the test can reduce the level of cheating by 50%.
- Methodology for detecting massive cheating while controlling for the false positive rate using a Bonferroni type procedure.
Thank you

- Gracias
- Asante Sana
- Merci
- Obrigado
- Grazie


Bibliography II


