Correlation Tests performed for the Caribou Contest

Thomas Wolf, Caribou Contests,
Parmdeep Bansal, Senbo Fan, Jan Vrbik, Mark Willoughby
Department of Mathematics, Brock University
email: ceo@cariboutests.com

December 13, 2018

1 Motivation

To make the Caribou Contest a serious and fair competition, it has to be ensured that students find the same conditions in all schools and libraries which participate, especially that collaborations during contest time are prevented. The contest already uses several measures to enhance safety:

- The order of questions is randomized for each student.
- The order of options for is randomized for each question and student.
- Students are required to input their computer screen number to identify who is sitting next to each other.
- previously selected options to questions are not visible on the screen.

As an additional measure, the procedures described below are designed to detect any collaborations between students during a contest. Based on results of these statistical tests we contact schools if necessary to ensure equal conditions at all schools and to talk to any students that appear to have collaborated massively.

2 Overview

The performed statistical analysis pairs all contestants within each school and establishes for each pair a correlation number which detects how often both contestants did choose the same wrong and same correct options. These two numbers are weighted on the basis of how many questions they got right. The average of these correlations are compared between schools and are compared with correlation values between students from different schools who could not have collaborated. Finally, all participating schools are ranked about these average correlations giving an impression how strict invigilation is handled at different schools. In addition, all pairings of any two students from any school are ranked so that one would detect the cheating of only two students in an otherwise fair environment.

Based on these two rankings (the ranking of all schools according to their average correlation and the ranking of all pairs of students from all schools with their correlation) emails are issued to school contact persons reporting the most striking incidents.

These statistics have helped to identify cases where single students helped each other but also when a supply teacher did not supervice properly and did not prevent students from looking at each others screens or even allowed all students talking to each other.

Although the statistics is based on the assumptions that all questions are similarly difficult and that wrong options are equally likely chosen which is not really the case, the findings are still valid because

- these effects are filtered out with control data: correlations between students from different schools who could not have collaborated,
- from all the correlations of pairs of students that are clearly higher than the control data, only those are reported that are absolutely inexplainable by coincidence.

3 Notation

We start with the notation. The real computation is much shorter than the notation seems to indicate. In the last section we comment on some findings and limitations of this approach.

n	number of questions in contest
u_j	number of options of question j
c_a	number of questions solved correctly by student a
w_a	number of questions solved wrongly by student a $(= n - c_a)$, not picking an option is counted as wrong
p^c_a	probability that student a answers a question correctly, $(= c_a/n)$
p^w_a	probability that student a answers a question wrongly, $(= w_a/n)$
p^c_{aj}	probability that student a picks for question j the correct option
p^w_{ajk}	probability that student a picks for question j the wrong option k, i.e. $p^w_{ajk} = 0$ if k is the correct option.
e_{ab}	number of times students a and b picked the same option no matter whether the option is right or wrong, (If both student do not pick any option to the same question then this does not count towards e_{ab}.)
r_{ab}	the number how often on average students a and b would pick the same option (if they do not communicate)
r_{abj}	the probability that students a and b pick the same option for question j (if they do not communicate)
C_{ab}	measure of correlation (i.e. of cheating) between students a and b

4 Correlations between two students

We define the measure of correlation of two student participations as

$$
\begin{aligned}
C_{ab} &= e_{ab} - r_{ab} && (1) \\
&= e_{ab} - \sum_{j=1}^{n} r_{abj} && (2) \\
&= e_{ab} - \sum_{j=1}^{n} \sum_{k=1}^{u_j} \begin{cases} p^c_{aj} p^c_{bj} & \text{if option k is correct} \\ p^w_{ajk} p^w_{bjk} & \text{if option k is wrong} \end{cases} && (3)
\end{aligned}
$$

If all questions are equally difficult even if they have different numbers of options then

$$p^c_{aj} = p^c_a = \frac{c_a}{n} \qquad (4)$$

and *if all wrong options are equally likely to be selected* then

$$p^w_{ajk} = p^w_a/(u_j - 1) = (1 - \frac{c_a}{n})/(u_j - 1) \qquad (5)$$

Formula (3) then becomes

$$
\begin{aligned}
C_{ab} &= e_{ab} - \sum_{j=1}^{n} \left(p^c_a p^c_b + \sum_{k=1}^{u_j - 1} (1 - p^c_a)(1 - p^c_b)/(u_j - 1)^2 \right) \\
&= e_{ab} - n p^c_a p^c_b - (1 - p^c_a)(1 - p^c_b) \sum_{j=1}^{n} \sum_{k=1}^{u_j - 1} 1/(u_j - 1)^2 \\
&= e_{ab} - n p^c_a p^c_b - (1 - p^c_a)(1 - p^c_b) \sum_{j=1}^{n} 1/(u_j - 1) && (6)
\end{aligned}
$$

where $p^c_a = \frac{c_a}{n}$. So all that is needed is

- to compute once $\sum_{j=1}^{n} 1/(u_j - 1)$ where u_j is the number of options of question j (currently $n = 12/15/18/21$ and $u_j = 7$ but that may change for future contests)
- for each student a
  - to count the number of correctly solved questions c_a,
  - to compute $p^c_a = c_a/n$,
- for each pair of two students a and b
  - to count the number e_{ab} of equal options picked by students a and b (if both students do not pick an option then no count) and
  - to compute C_{ab} from formula (6).

The larger the measure C_{ab} is, the more likely it is that students a, b collaborated or that at least one copied from the other in this contest.

The assumptions that all questions are equally difficult and all wrong options are equally likely to be selected are simple but not completely true. If the assumptions are wrong then C_{ab} values will be increased. For example, if one option looks totally right and all other wrong options are obviously wrong then C_{ab} will be increased by about 1 even for non-cheaters. As a consequence individual values C_{ab} have little importance, instead their relative size matters. To take care of such effects one could compute C_{ab} for strong and weak students but from different schools which will serve as a control value.

5 Correction based on varying Difficulty

5.1 Notation

One is able to get a more accurate result if one has a measure for the difficulty of each question and a measure how obvious an option is wrong.

Under the assumption that all options are equally likely the case that no option is selected is neglected but in the following this shall not be neglected, here u_j is increased by one which corresponds to the case that no option is selected.

We therefore introduce the following extra variables:

s number of students
s_{jk} number of times that for question j the option k was selected
δ_j^k =1 if option k of question j is correct else =0
s_j^c number of times that for question j the correct option was selected $(= \sum_k s_{jk}\delta_j^k)$
s_{jk}^w number of times that for question j the wrong option k was selected $(s_{jk}(1-\delta_j^k))$
s_j^w number of times that for question j a wrong option was selected $(= \sum_k s_{jk}(1-\delta_j^k))$

Not answering a question is one of the wrong options.

5.2 Constraints

For each student a the probabilities p_{aj}^c to get question j right summed over all n questions should be c_a which is the number of questions correctly answered by student a:

$$\sum_{j=1}^{n} p_{aj}^c = c_a \quad \forall a. \tag{7}$$

Similarly, for each question j

$$\sum_a p_{aj}^c = s_j^c \quad \forall j. \tag{8}$$

The probabilities p_{aj}^c should, of course, satisfy $0 \le p_{aj}^c \le 1$ and they should go monotonically with s_j^c , i.e. if $s_j^c > s_k^c$ then $p_{aj}^c > p_{ak}^c \ \ \forall a$. Similarly if $c_a > c_b$ then $p_{aj}^c > p_{bj}^c \ \ \forall j$.

As there are $n \times s$ many unknowns p_{aj}^c but only $n+s$ many equalities to be satisfied, there is no unique solution. One would want to find a function f for the probability $p_{aj}^c = f(c_a, s_j^c/s)$

that student a with strength c_a solves question j of difficulty s^w_j/s. To get a feeling for the function f one could make a 3-dimensional plot of s dots on a domain 0..24 (problems solved), 0..1 (easy..hard) (or 0..96 (points), 0..1) .[1]

One way to compute such a function $f(x,y)$ is to make a linear ansatz $f(x,y) = f_0 + f_1x + f_2y$ or a quadratic ansatz $f(x,y) = f_0 + f_1x + f_2y + f_3x^2 + f_4xy + f_5y^2$ or a higher degree ansatz and to compute a best fit to the s values of 0,1 for the s participating students which satisfies conditions 7 and 8.

The second problem is to compute the p^w_{ajk}. When p^c_{aj} is known then $p^w_{aj} = \sum_k p^w_{ajk} = 1 - p^c_{aj}$ is known. The question is how to calculate p^w_{ajk} from this. A natural approach is to compute the p^w_{ajk} through

$$p^w_{ajk} = p^w_{aj} s^w_{jk} / \sum_m s^w_{jm} \tag{9}$$

which has the benefit of satisfying the natural relations

$$p^w_{ajk} \propto s^w_{jk} \tag{10}$$

and also the necessary conditions

$$\sum_k p^w_{ajk} = p^w_{aj}. \tag{11}$$

Summations in (9) and (11) are done over all wrong options.

6 Alternative Approaches

In the above improvement we take into consideration that s^w_{jk} are not constant. But we only consider non-constant s^w_{jk} when computing the probability for the number of same options pickings. What we do not consider above is whether the option which was picked had a large s_{jk} or not, i.e. whether this option was selected often by the s students or not. Of course, if both students picked the same rarely selected option then the likelyhood of cheating is higher than if both picked a frequently selected option. What needs to be generalized is already the ansatz (1). We need to compute a probability measure for each question that cheating was done for that question and these cheating values need to be combined from all questions. When both students picked the same option then the overall cheating probability needs to increase and if they picked different options then the cheating probability needs to decrease.

Here the same explanation again with different words: Let us look at an extreme situation. Let us assume that for one question nearly all students picked wrong option k. It surely matters whether students a and b both picked option k or whether they both picked one and the same other wrong option which basically no-one else picked. In the first approach 1 we do not consider which option both students a, b picked. But that is very important. If both picked option k then this is no evidence that they cheated, if both picked another option then this is a strong indicator that they cheated.

Suggestions:

[1] Is the situation similar to the following? Given that particles in a physical system at equilibrium have a given temperature (average kinetic energy ($\equiv c_a$)) what is the probability to find them at the different energy levels? To use the Bolzmann distribution as solution one needs a relation between the energy levels in the physical model and the numbers s and s^c_j.

- The question of cheating in multiple option tests detected through correlation sounds like an every day problem which surely has been investigated before. We should do an online search about such researches.

- Can one simplify the problem and study that simpler problem first? For example, let us think of two students a, b both roling dice. What is the probability they both get an even number and the probability of both getting a 1 or 3 and the probability they both get a 5? This is a similar situation to some wrong options being more likely and some wrong options being less likely.

  In this example the probability of both picking an even number is $(3/6)^2$ and both picking 1 or 3 is $(2/6)^2$ and of both picking 5 is $(1/6)^2$. The probability of picking the same of these 3 categories is $1/4 + 1/9 + 1/36 = (9 + 4 + 1)/36 = 7/18$.

  If the categories were C1: 1 or 2, C2: 3 or 4, C3: 5 or 6 then the probabilities would be 1/9, 1/9, 1/9 so the probability of picking the same category is $1/9 + 1/9 + 1/9 = 3/9 = 6/18$. (It is natural that the probability of picking equal options is lowest when probabilities are equally spread between options.)

  So, what is the generalization to wrong options? If $s_j^w = 1 - s_j^c$ is the number of times a wrong option is taken in question j and s_{jk}^w is the number of times the wrong option k in question j is taken then the probability for 2 students taking the same wrong option k is $(s_{jk}^w / s_j^w)^2$?

  So how are these probabilities combined between different questions? Should we compute the probability that what happened could happen?

- How can one judge whether one method of computing C_{ab} is better than another method? On average the value C_{ab} should be close to zero for pairs of students a, b that can not have collaborated. Using our data this is easy to check by computing C_{ab} for 2 students from different schools.

7 A more radical Generalization to Questions of varying Difficulty and Options of varying Plausibility

7.1 Motivation and Idea

In section 4 the probability of a student to pick the correct option of a question was made dependent on the students strength which was necessary in order not to suspect cheating if each one of two students solves all questions correctly. What was ignored was that

1. questions have different difficulty and thus, for example, two students not solving a hard question or both solving a simple question is not a rare coincidence,

2. options have different appeal to students, for example, two students picking the same tempting wrong option is not a coincidence but both picking the same obviously wrong option is a rare coincidence,

3. the frequency by which a correct or wrong option is picked depends on the strength of students. For example, two strong students solving the same easy question correctly is no coincidence but two weak students solving the same easy question is a bigger coincidence. Also, an option may be obviously wrong for a strong student but not for a weak student.

In section 5 points 1 and 2 were addressed. What was not answered in section 5 was how to get a probability that a given question is correctly solved by a student with a given strength. Also the important point 3 was not addressed.

As mentioned in section 5 there are many more unknowns (all probabilities for each student to pick each option of each question) than there are constraints (7) and (8). *The key to solve all the problems is to use the available statistics of the contest. This becomes possible because we have several thousand participants in each contest. We have enough participants to determine a probability for each option of each question to be selected by students of each strength.* A question is how to characterize the strength, for example, by

- the number of points achieved, or
- the number of correctly solved questions, or by
- a range of the number of correctly solved questions?

We will use ranges for the number of correctly solved questions because

- this reduces the size of the resulting tables,
- this number will also be available in future when the ranking system will change,
- for grades 9-12 we would not have enough students to have sufficiently high numbers of students for each possible number of points achieved or each number of questions solved correctly, and because
- some numbers of points and even of numbers of solved questions have not been achieved.

Although we complicate matters by making all dependent on student strength, this also has a benefit of simplifying matters because now we do not have to distinguish between correct options and incorrect options because all option probabililties depend now on student strength.

7.2 Notation

We introduce the following variables

u_j number of options for question j
m_a index of the strength group that student a belongs to
c_a number of questions solved correctly by student a
s number of students
s^m number of students belonging to strength group m
s^m_{jk} number of students that belong to strength group m and that picked option k in question j, $1 \leq j \leq n$

7.3 Correlation Computation

The probability p_{jk}^a for student a to pick option k in question j is then

$$p_{jk}^a = s_{jk}^{m_a}/s^{m_a}. \tag{12}$$

The probability for two students a, b to pick the same option k in question j is therefore $p_{jk}^a p_{jk}^b$.

If students a and b pick the same option k in question j then a measure for the probability e_j^{ab} of cheating in question j shall be $1 - p_{jk}^a p_{jk}^b$. If both students pick different options in question j then the measure of cheating in question j shall be 0, so we get

$$e_j^{ab} = \begin{cases} 1 - p_{jk}^a p_{jk}^b & \text{if both pick the same option k} \\ \\ 0 & \text{if both pick different options} \end{cases} \tag{13}$$

and thus e_j^{ab} is dependent on the option k and is increasing as p_{jk}^a and p_{jk}^b decrease as we want it to be.

What is a measure for the total amount of correlation e^{ab} between students a and b in the whole contest?

To get a number independent of the number of questionss we define

$$e^{ab} = (\sum_{j=1}^{n} e_j^{ab})/n. \tag{14}$$

What is the expectation value r^{ab} for both students a, b if there is no cheating? The expected outcome would be the sum over all questions $\sum_{j=1}^{n}$ of the sum over all options $\sum_{k=0}^{u_j}$ in one question j times the value for that outcome. This would be

$$r^{ab} = \sum_{j=1}^{n} \sum_{k=0}^{u_j} p_{jk}^a p_{jk}^b (1 - p_{jk}^a p_{jk}^b). \tag{15}$$

The correlation value C^{ab} is therefore

$$C^{ab} = e^{ab} - r^{ab}. \tag{16}$$

Each individual e_j^{ab} can be pretty high and thus serve as a strong indication for cheating in question j. Therefore $\max_j e_j^{ab}$ could be used as a measure whether cheating occured and C^{ab} a measure for the amount of cheating that occured.

7.4 Adjustments for low Participation Numbers

For $s_{jk}^m = 1$ being a meaningful statistical measure the number of participants belonging to strength group m must be high enough. This number depends on the number of options of the question. It would be good if each strength group would have at least about 3 times as many members as there are options per question so that with 7 options there should be at least 20 members in a strength group.

For example, for grade 11/12 with the smallest participation number, for example, with 38 participants in grade 11/12 in May 2014, one could experiment with 9, 14 and 19 members per strength group. A possible procedure to establish groups of at least 10 students would be (in Pascal syntax)

```
\\ This procedure determines
\\ no_of_groups: an integer which is the number of groups
\\               of different strengths of students
\\ groups[1..no_of_groups] with components
\\ group[i,1] is the minimal # of correctly solved questions
\\            in group i
\\ group[i,2] is the maximal # of correctly solved questions
\\            in group i
\\ group[i,3] is the # of students in group i
\\
\\ Parameters are min_group_size: minimum number of students per group
\\                max_no_of_correct_solutions: depends on contest

const min_group_size=10;   \\
      max_no_of_correct_solutions=24;  \\
      max_no_of_groups=max_no_of_correct_solutions+1; \\ +1
      \\ because of the possible group of 0 correct solutions

var group: array[1..max_no_of_groups,1..3] of integer;
    no_of_groups,i:integer;

begin
 i:=0; \\ i will be the number of groups
 repeat

  i:=i+1;
  \\ initialization of group[i,...]
  if i=1 then group[i,1]:=0
         else group[i,1]:=group[i-1,2]+1;
  group[i,2]:=group[i,1];
  group[i,3]:=#_of_students_with_group[i,2]_many_correct_solutions;

  while (group[i,3]<min_group_size) and
        (group[i,2]<max_no_of_correct_solutions) do begin
   \\ increase the number of correct solutions by 1
   \\ to get more students into this group i
   inc(group[i,2]);
   group[i,3]:=group[i,3]+
               #_of_students_with_group[i,2]_many_correct_solutions;
  end;
```

```
  if group[i,3]<min_group_size then begin
   \\ The last group does not have enough students so these students
   \\ are added to the previous group which will become the last group
   group[i-1,2]:=group[i,2];
   group[i-1,3]:=group[i-1,3]+group[i,3];
   i:=i-1
  end

 until group[i,2]=max_no_of_correct_solutions; \\ i.e. all students are in
                                                \\ a group
 no_of_groups:=i

end;
```

7.5 Overview

The complete correlation statistics then consists of the following steps.

- Use the above procedure to determine the number of groups of different strengths (=`no_of_groups`) and for each group i the minimal and maximal number of correct solutions (=`group[m,1]`, `group[m,2]`) and number s^m of students in this group (=`group[m,3]`).
- Determine for each question j and option k and each group strength m the number of students S^m_{jk} from the strength group m that picked option k in question j and assign $p^m_{j,k} = S^m_{jk}/S^m$ as the probability for a student from strength group m to pick option k in question j.
- For each school and each pair of students a, b from that school compute
  - for each question j the value e^{ab} from (13) and (14)
  - r^{ab} from (15) and finally(16)
  - C^{ab} from (16)
- Put, say 40, students, each form a different school into a 'virtual control school' and compute C_{ab} for all pairs of students of this virtual school. One could create virtual schools of different strength levels in order to check whether the quality (predictive power) of C^{ab} depends on the strength level which should not be the case.
- For each school, including the virtual school determine the school maximum C^{ab} value and compute the school average of the C^{ab} values.
- Print a table with 3 columns, and as many rows as there are schools including the virtual school in the first row. The first column contains the name of the school, the second column the school average of the C^{ab}, in the 3rd column the school maximum value of C^{ab}. Each school name is a link to the scool table below, see below. The table

has for the 3 columns the 3 headers 'School Name', 'Average', 'Maximum'. Each of the 3 headers can be clicked which causes the rows to be re-sorted according to that column.

- When a school name is clicked then a window comes up which shows information of that school, i.e. a table with the highest ten C^{ab} values, one in each row, sorted by size, largest on top, each with the C^{ab} value, the two student names and their overlap time (= the difference between the earlier of the two finish times and the later of the two starting times, but not less than zero). The first column is the students name of the pair with the better rank and the 2nd column is the name of the other student with his/her rank, a column with their C^{ab} value, their overlap time (= the difference between the earlier of the two finish times and the later of the two starting times, but not less than zero), the difference between their ending times, the number of questions answered correctly through both of them, the number of questions they both answered with the same wrong option, whether at least one of both declared the other one as neighbour. To speed up the generation of this table one might show less information in this table but might make each row clickable and show that extra information for this pair in this 3rd table.

- It would be interesting to find questions j and options k from these questions where stronger student groups found fewer correct solutions than the students of a weaker group. Similarly it would be interesting to questions j and options k where students from a stronger group picked the wrong option k more often than students from a weaker group.

- Ideas to speed up

7.6 Constraints

$$\sum_m s^m = s$$

$$\sum_k s^m_{jk} = s^m, \qquad \forall m, j$$

8 Measuring the Quality of Correlation Detection

How can one determine the quality of cheating detection? The general idea is to compare the distribution of correlations of classes with potential cheating with that of known non-cheating pairs of students, for example, one from one school and one from another school. How do both distributions differ?

If the class (of possible cheaters) has n students then there are $n(n-1)/2$ many pairs. But every student can cheat only with both neighbours (unless the whole class works together). So if in the extreme case pairs of students collaborate totally then in the limit of large n the average $\frac{2}{n(n-1)} \sum_{a,b} C_{ab} \to 0$ because not more than $n/2$ pairs will show correlations.

If one alternatively takes the average of the $n/2$ largest pair correlations (which is the largest number of collaborations if everyone collaborates with one other student) or the n largest pair correlations (if everyone collaborates with 2 neighbours) then this is probably also not correct because if one has a quadratically growing number of population of pairs then the chance to have a few pairs with accidentially large correlations becomes more likely and looking a a small fraction of $n/2$ pairs from $n(n-1)/2$ pairs should produce higher correlations as $n \to \infty$.

From these two considerations we see that it is necessary to work with distributions of correlation values and not with single numbers. What should be done is to simulate the participation of a whole class and to generate a distribution of correlation values for this class under the assumption that no cheating has occured. Then one can compare the simulated C-distribution with the real C-distribution and can make a statement about the invigilation level in the class, i.e. how much cheating occured in the whole class and about individual correlation values, i.e. how likely they show cheating.

How can a non-cheating C-distribution be generated? Input: all data from section 7.2 which characterize the contest and the whole group of participants. In addition data characterizing the size and strength of the class in question:

- the number of students in the class,
- the strength distribution of the students in the class, i.e. how many students solved how many questions.

A program then simulates the contest participation for that class. In the simulation each student a picks for each question j an option k with the probability p^a_{jk} from (12). Then all correlation values of all pairs of that class are computed. This process is repeated a number of times and the correlation values are averaged over these runs. These averaged correlation values are plotted as a distribution (frequency of each C value over that C value) together with the correlation distribution of the real class and both are compared. The number of contests to be simulated and averaged has to be high enough so that the averaged curve looks smooth and Gaussian like.

9 A characterization of invigilation at schools

The statistics to be described in this section have two aims:

- to provide control values to estimate systematic errors in the C_{ab} measures,
- to compare schools in their ways how they hold the Caribou Contest.

What is a good measure for how strictly schools invigilate contests, for example, whether students were allowed to talk to each other or not?

The simplest method would be to average C_{ab} over all pairings of a school. This is not a very good measure because a student can not talk to *all* other 50 children in the class, only to his/her neighbours. So if a school had s students then one might calculate the C_{ab} values from all pairs of the school, but for computing the average one only uses the $2 \times s$ pairs with the highest C_{ab} value. This is done and the result is shown in the 'Corr. Avg.' column of the top table. The 'Max. Corr.' column shows the maximum C_{ab} value at the school. The 'Corr. Avg.' is used to remind schools with a high value to invigilate more strictly or even to disregard the results of a whole class in one contest as it has happened so for once.

As a control value we create a virtual 'control school' which contains all students in the contest. If 10,000 or 100,000 students participate in the contest then the above procedure of checking all pairings could and should not be applied. Instead, all students are sorted according to the number of points they reached in the contest and all pairs of students which are neighbour in this ranking and which are not from the same school are checked and based on these checks the average correlation value is computed.

That means that the control value is not computed in exactly the same way as the other school averages but the approximation should be good enough.

10 Discussion

- If two students would answer the same question with the same wrong option and answer no other question then cheating seems more likely than if they had also answered all other questions and picked different options for the other questions. Nevertheless, we still ignore the number of questions not answered by both students because of the following reasons.
  - Only very few students do not answer questions.
  - If students do not answer questions then they usually do not answer most of the questions and counting those questions not answered by both students a, b as correlations would give very high e_{ab} and thus high C_{ab} values which would disturb school averages.
  - Students not answering many questions perform very poorly and are not the focus of this correlation test.
- The Caribou contest on May 19, 2010 was the first with a randomized order of questions. The correlation measures dropped clearly compared to contests held before.
- With the start of school year 2010/11 contestants will be asked before starting the contest to enter the label of their own computer and the labels of the neighbouring computers. In the listing of largest correlations between 2 students one will have to indicate whether two students are neighbours or not.
- Is the school average of C_{ab} dependent on the number of students of the school? If a school has n participating students then the number of pairings goes like n^2 but we average only the top $2n$ pairs, so larger schools seem to be disadvantaged in this statistics somewhat. How big is hat dependence on n and how could one compensate this?
- If two students from the same school are very good then they will have picked the same (right) option for many questions. The above calculation is aiming to downgrade this unavoidable correlation. The ranking in the 'Control School' shows that the highest

C_{ab} are attached to pairs where both students do comparatively well, so there is still such an effect but it is not big. Only few C_{ab} values are slightly higher than one would expect which seems to be a statistical effect due to the large number of students in the 'Control School' (see the previous comment about the influence of the size of the school).

10.1 Interface

- There should be two links, one to a description of the fast first algorithm and one to the second more time consuming algorithm.

10.2 Speed Up

- Each school needs to compute only their own table. But a contact needs to see a distribution of C^{ab} averaged over each school to be able to judge how bad their own average value is. To judge how bad their max C^{ab} value is, maybe it is enough to list with the max value the list of questions where the question number background colour is normal white if both picked different options, it is red if they picked the same wrong option and the question number background is orange if the option they both picked is correct.

- To speed up computation maybe one does not compute all pairs of students from a school. It should be good enough to pair students with similar rank or similar number of correct questions or similar number of points because if 2 students with very different rank would have collaborated then the much better student did surely not benefit and the much weaker student did obviously not benefit either. Also, by checking only similar students

- One could do a fast simple computation using the first algorithm and only if that gives a value higher than some threshold then the advanced time consuming computation should be done. One then should display both values which will give us good feedback about how both algorithms compare.

10.3 Optimization, Questions

- Can one print the characters in names of French Schools correctly? Example: cole lmentaire catholique Notre-Dame

- When I click a school name in the first table of schools where the school has only 4 students, i.e. only 6 pairings it still takes a long time to come up with this small table, why?

- If I want to sort this very small table by C^{ab} value it take very long even though I can do that in my head in 2 sec, why?

- Can one sort this single school table by default according to C^{ab} and not by student A user ID? Basically always one wants this table sorted according to C^{ab} and nothing else.
- When clicking the Student A user ID a nice illustrative table comes up. Would it be possible to print the correct option values or the background of correct option values in a different colour? The reason is that picking the same correct option value is not a big coincidence but picking the same wrong option value is a bigger coincidence.
- Parm, could you run the computation with different group sizes and compare the results and thus find rules for optimal group sizes dependent on the number of participants?

11 A Very Different Fraud Check

The idea is that for solving many questions correctly it takes some time. It is allowed to guess and a guessed correct solution is as much worth as a computed correct solution. But the point is that it is very unlikely to guess the solution to many questions correctly. What would be useful is to divide the success rate by the time and compare this quotient between students. For success rate one could use the number of points but they do not properly represent the time needed to solve a question. A question worth 5 points does not just need 5/3 times more time than solving a 5 point question. If P is the number of points and T the time then the ratio is more like $T \approx P^2$ or even $T \approx C_1 + (P - C_2)^2$ where C_2 is the time needed to read and understand a question (at least 10 sec).

Acknowledgement

TW is grateful to to Brad Saxton (Brock) for programming the web interface of the first version of the correlation testing script.