Judging a Mathematics Contest

Daniel A. Calderón Brennan
Philip J. Darcy
David T. Tascione
St. Bonaventure University
St. Bonaventure, NY 14778

Advisor: Albert G. White

Overview

Our model is based on breaking the problem down into four main areas and dealing with each: the distribution of papers among the judges, scoring methods, the number of papers to eliminate per round and the number of rounds, and the performance of the model with larger numbers of papers.

In each component, we focused on the goals of maintaining fairness and variety in all judging procedures, eliminating as many papers as possible in each round, minimizing the number of rounds, and, most important, seeing that no one goal was attained at the expense of any other.

The papers were coded and sent on from judge to judge without any prior knowledge of that particular participant or paper. There is no way to avoid having a judge read a paper twice, but a judge will not read a paper twice in a row until perhaps the final two rounds.

Assumptions

- Budget constraints affect only the number of judges.
- Time constraints affect only the number of papers that each judge can read.
- An approximate "absolute" ranking system exists among the judges; i.e., if every paper were to be scored or ranked by each judge, the results of each judge would generally agree with every other (allowing for a few places where consecutive papers may be "flip-flopped").
- All papers are eligible to win (none disqualified for cheating, missing sections, etc.).
- Judges need not be in the same location, but a copy of each paper (electronic or hard copy) is readily available to each judge.
- Judges remain ignorant of other judges' opinions on all papers.

- There is no way to avoid having a judge read a paper twice (nonconsecutively) during the reading process, but re-reading has no effect on judges' opinions of papers (i.e., a judge rates a paper as effectively on a second read as on the first).
- A minimum of 5 judges per 100 papers are needed.

Developing the Model

- Paper distribution:
  - Maintain judges' ignorance of other judges' opinions of papers.
  - Distribute papers so that the top $2W$ (6) are kept in competition throughout the contest.
  - Distribute papers such that no judge sees the same paper in consecutive rounds until the final two rounds (ensures efficiency in judging and fairness, as multiple opinions of papers are necessary).
  - Computer distribution both frees one human to judge instead of distributing papers and accomplishes the above tasks.
- Equating numerical judging systems:
  - Deals with the possibility of systematic bias in use of a numerical scoring scheme.
  - For each judge in the numerical scoring round, we adjust scores so that the highest score is equated to 100% and the others are adjusted proportionally; for example, scores of 92, 86, 89, 79 are adjusted to the same values divided by 92.
  - Given our assumption of an absolute ranking scheme and that a numerical scheme will closely follow, this method in essence allows us to put a numerical value on a judge's ranking so that it may be compared with ranks of other judges.
- Cuts at end of each round; number of rounds; total papers read per judge:
  - As few rounds as possible (time / budget constraints)
  - As few reads per judge as possible (time / budget constraints)
  - Keep the top $2W$ papers intact
- Cut approximate 44% of papers remaining in each round. When possible, leave a multiple of J intact, so that only in the first round do some judges read more than others. In the first round, do not cut to fewer than $2W + 1$ papers, in case the best $2W$ papers go to the same judge. After the first round, distribution of papers will prevent this problem from occurring again (fairness).

Methods for Distribution

The method for redistributing the papers after the first round is based upon a matrix created by the judges, each of whom enters in a column, from top to bottom, the numbers of the paper read, in decreasing order of rank. Then the first row consists of the highest-ranked papers of each judge.

Using falling diagonals of the matrix, it is possible to ensure that in the next round no judge receives any paper just read in the previous round. Further, it is possible to ensure that one judge does not receive all $2W$ top papers, which would force them to be cut the next round. The matrix below illustrates the method for redistributing the papers.

$$\begin{array}{c|cccccccc|}
 & J_8 & J_1 & J_2 & J_3 & J_4 & J_5 & J_6 & J_7 \\
 & a_{11} & a_{12} & a_{13} & a_{14} & a_{15} & a_{16} & a_{17} & a_{18} \\
J_7 & a_{21} & a_{22} & a_{23} & a_{24} & a_{25} & a_{26} & a_{27} & a_{28} \\
J_6 & a_{31} & a_{32} & a_{33} & a_{34} & a_{35} & a_{36} & a_{37} & a_{38} \\
J_5 & a_{41} & a_{42} & a_{43} & a_{44} & a_{45} & a_{46} & a_{47} & a_{48} \\
J_4 & a_{51} & a_{52} & a_{53} & a_{54} & a_{55} & a_{56} & a_{57} & a_{58} \\
J_3 & a_{61} & a_{62} & a_{63} & a_{64} & a_{65} & a_{66} & a_{67} & a_{68} \\
J_2 & a_{71} & a_{72} & a_{73} & a_{74} & a_{75} & a_{76} & a_{77} & a_{78}
\end{array}$$

From the diagonals of the matrix that run downward from left to right, we get the assignments of papers to judges for the next round:

$$\begin{array}{ll}
J_1: & a_{12},\ a_{23},\ a_{34},\ a_{45},\ a_{56},\ a_{67},\ a_{78} \\
J_2: & a_{13},\ a_{24},\ a_{35},\ a_{46},\ a_{57},\ a_{68},\ a_{71} \\
J_3: & a_{14},\ a_{25},\ a_{36},\ a_{47},\ a_{58},\ a_{61},\ a_{72} \\
J_4: & a_{15},\ a_{26},\ a_{37},\ a_{48},\ a_{52},\ a_{62},\ a_{73} \\
J_5: & a_{16},\ a_{27},\ a_{38},\ a_{41},\ a_{53},\ a_{63},\ a_{74} \\
J_6: & a_{17},\ a_{28},\ a_{31},\ a_{42},\ a_{54},\ a_{64},\ a_{75} \\
J_7: & a_{18},\ a_{21},\ a_{32},\ a_{43},\ a_{55},\ a_{65},\ a_{76} \\
J_8: & a_{11},\ a_{22},\ a_{33},\ a_{44},\ a_{56},\ a_{66},\ a_{77}
\end{array}$$

The Judges rank-order the newly distributed papers, and again a matrix is used to redistribute the papers that make the cut:

$$\begin{pmatrix}
a_{11} & a_{12} & a_{13} & a_{14} & a_{15} & a_{16} & a_{17} & a_{18} \\
a_{21} & a_{22} & a_{23} & a_{24} & a_{25} & a_{26} & a_{27} & a_{28} \\
a_{31} & a_{32} & a_{33} & a_{34} & a_{35} & a_{36} & a_{37} & a_{38} \\
a_{41} & a_{42} & a_{43} & a_{44} & a_{45} & a_{46} & a_{47} & a_{48}
\end{pmatrix}$$

Again, diagonals are used to redistribute four papers to each judge.

$$\begin{array}{lllll}
J_1: & a_{12}, & a_{23}, & a_{34}, & a_{45} \\
J_2: & a_{13}, & a_{24}, & a_{35}, & a_{46} \\
J_3: & a_{14}, & a_{25}, & a_{36}, & a_{47} \\
J_4: & a_{15}, & a_{26}, & a_{37}, & a_{48} \\
J_5: & a_{16}, & a_{27}, & a_{38}, & a_{41} \\
J_6: & a_{17}, & a_{28}, & a_{31}, & a_{42} \\
J_7: & a_{18}, & a_{21}, & a_{32}, & a_{43} \\
J_8: & a_{11}, & a_{22}, & a_{33}, & a_{44}
\end{array}$$

The next round of judging results in only two papers being passed by each judge. For this round, we convert the 2×2 matrix of the papers that pass this cut into a 4×4 matrix as follows:

$$\left(\begin{array}{cccc|cccc} a_{12} & a_{13} & a_{14} & a_{15} & a_{16} & a_{17} & a_{18} & a_{11} \\ a_{23} & a_{24} & a_{25} & a_{26} & a_{27} & a_{28} & a_{21} & a_{22} \end{array}\right) \longrightarrow \begin{pmatrix} a_{12} & a_{13} & a_{14} & a_{15} \\ a_{16} & a_{17} & a_{18} & a_{11} \\ a_{23} & a_{24} & a_{25} & a_{26} \\ a_{27} & a_{28} & a_{21} & a_{22} \end{pmatrix}$$

We split each row between columns 4 and 5 to create four rows of four elements. The papers are once again redistributed among the judges for numerical scoring. We must distribute each paper to two judges, according to our judging scheme, and we have devised a distribution method that ensures that no judge shares more than one paper to be read with any other (so that each paper may be numerically scored and therefore weighted against the widest possible range of papers allowable; in this case, each paper is scored against seven others). We do this first by adopting the diagonal distribution scheme that we have used throughout this problem and assigning the groups of four papers to four of the judges. Then, once we have distributed each paper once, we distribute the papers a second time by assigning the next four judges one column of the matrix as it stands above. In this case, we really cannot prevent a judge from reading a same paper as in the previous round, but the different scoring system calls for a more in-depth look at the papers anyway, and previous knowledge of a paper should not hinder the fairness of the distribution / scoring scheme.

In the final round, each judge reads the final eight papers and ranks them. The papers are then scored according to low-rank-sum method described in the **Judging Methods** section. The top three papers, the ones with the lowest rank sum, are the winners.

Number of Rounds, Papers Read per Judge, Paper Elimination

We must limit the number of rounds of judging and the number of papers read by each judge by eliminating the greatest possible number of papers per round, while protecting papers that have an earnest chance of winning from being eliminated in a round in which too many papers are eliminated. This amounts to protecting the "best" $2W$ papers until they may compete against each other in the final round of judging.

In the first round, we call for elimination of only so many papers as to leave $2W + 1$ papers from each judge remaining in the competition. This ensures against the unlikely occurrence that a single judge receives the top $2W$ papers in the initial round. After this round, the methods of paper distribution prevent this from happening. A maximum of W papers may be given to a single judge in the second round, which calls for elimination of all but $W + 1$ papers; and in the third round, elimination of one-half of the remaining papers is allowable, since by this time the distribution scheme has spread out the $2W$ "best" papers enough to protect them from being eliminated.

Now that the papers have been thinned out, and the "cream" has been allowed to rise to the top, we may begin more forceful elimination measures. We make fourth round one in which the judges assign numerical grades to papers, with only the top $2W + 2$ papers surviving elimination The scoring procedure is discussed in the **Judging Methods** section; it suffices here to say here that the judging saves the top papers from being eliminated while allowing a drastic reduction in the number of papers remaining.

With only $2W + 2$ papers left in the running after the fourth round, we can have each judge read each of them and rank them in order, with the W papers ranked consistently high enough emerging as victors.

The judges have been subjected to the fewest number of rounds and the fewest possible number of papers to read.

Judging Methods

Our model uses two methods of judging, rank-ordering and numerical judging. Rank-ordering offers the advantage that there should not be any bias to interfere with the ranking. We base this on the assumption of an absolute ranking system, i.e., if the judges were to read all the papers, they would agree to an absolute ranking of the papers (with some reversals of consecutive papers). The ranking of any subset of papers must also conform to the absolute ranking, that is, the process of ranking must be order-preserving. The disadvantage with rank-ordering is that if a judge were to receive all the top papers, some of the top papers would get cut.

The other method, numerical grading, has the disadvantage of allowing for systematic bias. If a judge on average gives lower scores, our adjustment procedure compensates.

Since we use both methods of judging, we developed corrective measures for both to prevent these problems. The simpler of the two methods to correct is the rank-ordering method. We base the redistribution of papers that passed the cut on the rank that they had in the previous cut, ensuring that the best

$2W+1$ papers pass the first rank cut. After that cut, the redistribution prevents any judge from receiving a majority of the top $2W$ papers until the numerical grading round.

To prevent bias from removing one of the top $2W$ papers at the numerical round was a difficult task. We decided finally to base the numerical grading on a percentage curve based on the grades that the judges assign to all the papers. The top-scoring paper in a given judge's group determines the 100% paper, with the each of the remaining papers curved assigned its percentage of the top score. The distribution of the papers in the scoring round ensures that any two judges will have only one paper in common and that every paper is graded twice. Therefore every paper will have two "curved" percentage grades. These two grades are averaged together, yielding an overall weighted percentage ranking. The top $2W+2$ papers pass this cut in a further correction to ensure that the top $2W$ papers pass into the final round.

The eight papers in the final round are scored by using low-rank scoring. One of the eight judges assigns a number respectively to a paper, 1 (outstanding) through 8 (poor). At the end of the final round, the scores are added together and the paper with the lowest score wins.

Generalizing the Model

Since we have assumed that the ratio of judges to papers is greater than .05, we may test our model's growth rate, in terms of reads per judge and percentage of total papers read per judge, for larger ratios. We examined the changes as the number of papers doubles from 100 to 200, and then to 400, for ratios of .05 and .10. We assumed that each time P doubles, so does W.

Table 1 gives the details of our analysis, for a variety of different situations. For a ratio of .10, we see a tremendous increase in efficiency over the situation for a ratio of .05! Each judge reads about one-third of the total papers instead of more than half.

Demonstration of the Model

[EDITOR'S NOTE: At this point the authors illustrate in detail the application of their model to two sets of data devised by them, each with $P=100$, $J=8$, and $W=3$. In the first example, the top three papers in fact win. The second example shows what happens when one judge receives all $2W$ top papers in the first round: The top three papers still win. For reasons of space, we omit these extended examples.]

Table 1.

Analysis of contest situations for ratios of judges to papers of .05 and .10.

Ratio	P	J	W	Round	Papers	Papers / judge	Method	Eliminated / judge
.05	100	5	3	1	100	20	rank	8
				2	60	12	rank	4
				3	40	8	rank	4
				4	20	8	score	bottom 14 averages
				5	8	8	rank	5 highest rank sums
						Total: 56 (56%)		
	200	10	6	1	200	20	rank	8
				2	120	12	rank	4
				3	80	8	rank	4
				4	40	8	score	bottom 26 averages
				5	14	14	rank	8 highest rank sums
						Total: 62 (31%)		
	400	5	12	1	400	20	rank	8
				2	240	12	rank	4
				3	160	8	rank	4
				4	80	8	score	bottom 54 averages
				5	26	26	rank	14 highest rank sums
						Total: 74 (18.5%)		
.10	100	10	3	1	100	10	rank	4
				2	60	6	rank	2
				3	40	4	rank	2
				4	20	4	score	bottom 12 averages
				5	8	8	rank	5 highest rank sums
						Total: 32 (32%)		
	200	20	6	1	200	10	rank	4
				2	120	6	rank	2
				3	80	4	rank	2
				4	40	4	score	bottom 26 averages
				5	14	14	rank	6 highest rank sums
						Total: 38 (19%)		
	400	40	12	1	400	10	rank	4
				2	240	6	rank	2
				3	160	4	rank	2
				4	80	4	score	bottom 54 averages
				5	26	14	rank	14 highest rank sums
						Total: 50 (12.5%)		

Strengths and Weaknesses

Strengths

- Paper distribution: The distribution method keeps the required $2W$ papers in circulation while eliminating the maximum number of papers at each round. It also allows for a wide variety of judges' opinions to be given on the top $2W$ papers.
- Judging methods: The numerical grades overcome systematic bias by the judges. The rank-ordering allows the maximum number of papers to be culled in early rounds while preserving the top $2W$ papers.
- Elimination methods: We minimize the number of rounds and the number of papers judges must read, while maintaining fairness.
- Growth rate: For a given ratio of papers to judges, as the number of papers increases, the percentage of papers read per judge diminishes rapidly.

Weaknesses

- Paper distribution: In the worst-case scenario, we cannot prevent a single judge from receiving all of the top P/J papers and eliminating the bottom portion of those.
- Judging methods: The numerical grading adjustments requires redundant readings, and extra time is consumed in applying the adjustments. Further, a less-qualified paper may receive a higher weighted average than a paper with a higher absolute ranking, if the lower-ranked paper is judged against papers with even lower ranks and the higher-ranked paper is judged against papers with even higher ranks.
- Elimination methods: Entering the final round, too few papers may be eliminated. This apparent inefficiency is necessary to carry the top $2W$ papers into the final round.
- Growth rate: With a low judge-to-total-papers ratio and a relatively low number of papers, the percentage of total papers read by each judge is quite high. The total number of papers read per judge increases even as the percentage drops.

References

Decker, Rich and Stuart Hirschfield. 1995. *The Object Concept: An Introduction to Computer Programming Using C++*. Boston, MA: PWS Publishing.

The selection sort that we used comes from this book.