Olympiads in Informatics, 2015, Vol. 9, 139–145
DOI: http://dx.doi.org/10.15388/ioi.2015.11

The Estimation of Winners' Number of the Olympiads' Final Stage

Aleksandr MAIATIN, Pavel MAVRIN, Vladimir PARFENOV,
Oksana PAVLOVA, Dmitrii ZUBOK
ITMO University, Saint-Petersburg, Russia
e-mails: mayatin@mail.ifmo.ru; pavel.mavrin@gmail.com; parfenov@mail.ifmo.ru; pavlova.ifmo@gmail.com; zubok@mail.ifmo.ru.

Abstract. It is a complex and actual task to determine promising candidates for Olympiad's final stage from the participants' number of remote qualifying stage on the basis of their points scored during the qualifying stage. In this paper estimations of winners and awardees' number of the Olympiad final stage are made depending on the value of the passing score of the Olympiad's final stage. As part of the mathematical approach of mass Olympiad participants' results evaluation, data on indices of problems' solvability and participants' distribution in the qualifying stage according to the type of solved problems and scored points are used. The proposed approach can be applied by methodical committee and the jury of massive Olympiad in the process of the contest problems' development and determination of the passing score into the Olympiads' final stage.

Keywords: mathematical statistics, Olympiads, the criteria for results' evaluation.

1. Introduction.

The popular form of organizing Olympiads in Informatics and programming around the world consists of two stages – remote qualifying stage and the final intramural. The features of the remote stage are a great number of participants and a similar set of problems. The originality and novelty of problems are mostly inherent to the final stage. After the qualifying stage the Organizing Committee of the Olympiad decides whom to admit to the final. This decision is often based on a simple ranking of participants' scores and methodically founded principle of identifying promising or prospective participants. One can consider the time for solution of the problem or the number of attempts to solve the problem in the case of Olympiads in programming.

Methodists have a few problems. Firstly they have to decide what types of problems to devise, the number of problems and how to determine the estimation of their difficulty using points. The crucial rule here is the tradition of Olympiads. The second problem – how to determine the number of points p_{th} for qualifying participants into the final stage. And the third one – how to determine the winners of the finals. The last two issues are resolved with a help of expert estimation.

Let's proceed from the fact that the aim of the Olympiad or one of the stages of the competition is to identify as many talented teenagers as possible. But the more the finalists are, the more resources are involved into the Olympiad conducting process. It is necessary to find a compromise. So we must be able to build more or less reliable forecast of the winners' number depending on the participants' number.

One can apply optimization techniques or methods of game theory, but we use the method of extrapolation and linear regression.

One of the important steps that must be carried out is to determine the estimation of the complexity of the problem using points. One can consider a few approaches to the determination of the numerical weights of typical problems of qualifying stage: expert estimation (Option "A"), which is a priori and estimation by solvability index of respective problems of the qualifying stage (Option "B"), which is a posteriori estimation.

Nowadays it is an actual issue to determine the passing score "p" in the final stage of the Olympiad, as among those who are not admitted to the final stage could be the ones who would be able to cope with the proposed problems and could become winners and awardees. To solve this problem it is required to estimate the probability of problems solving by the participants of the final stage, and, accordingly, it is required to estimate the number of winners and awardees of the final stage depending on their results in the qualifying stage and the expected set and level of problems complexity of the final stage.

2. Method of Estimating the Passing Score

First of all each problem of the final stage should match the group of the problems of qualifying stage according to theme and difficulty level. For these groups of problems of qualifying stage one can calculate average index of solvability according to the complete data. Then let us make a table of correspondence between the indices solvability of problems of the final stage and the average solvability indices corresponding groups of qualifying stage problems.

To estimate the number of winners and awardees of the final stage of the Olympiads depending on the passing score at $p < p_{th}$, method of extrapolation is used (Krug *et al.*, 1977).

Let K be the maximum number of qualifying stage points, N is the maximum number of the final stage points. Let's consider the random variables X and $Y^{(n)}$, $0 \le n \le \mathrm{N}$. The value x_k of random variable X is the number of participants with k points after the qualifying stage and n points in the final stage. Thus, the value of $(y_k^{(n)}, x_k)$ of two-dimensional random variable $(Y^{(n)}, X)$ is considered. To estimate the number of winners and awardees of the final stage, depending on the passing score when $p < p_{th}$ let us consider a linear regression of the random variable $Y^{(n)}$ to X (Krug *et al.*, 1977; Elfving, 1952). Let us use a linear regression equation in the form (Cramer, 1975)

$$y = m_Y + \frac{\rho_{XY}\sigma_Y}{\sigma_X}\left(x - m_X\right), \tag{1}$$

where m_X, m_Y – sample average values of random variables X и Y, σ_X, σ_Y – sample standard deviations of random variables X and Y, ρ_{XY} – sample correlation coefficient of the random variables X and Y. Equation (1) gives an estimate of the values of a random variable, determined best by the theoretical regression equation in terms of the principle of the least square (Cramer, 1975; Ayvazian *et al.*, 1983).

Like the formula (1), the expressions for the partial empirical equations of linear regression for estimation of the values $Y^{(n)}$ have the form:

$$y_k^{(n)} = m_{Y^{(n)}} + \frac{\rho_{XY}^{(n)}\sigma_{Y^{(n)}}}{\sigma_X}\left(x_k - m_X\right), \tag{2}$$

where m_X, $m_{Y^{(n)}}$ – sample average values of random variables X and Y, $\sigma_X, \sigma_{Y^{(n)}}$ – sample standard deviations of random variables X and Y, where $\rho_{XY}^{(n)}$ – sample partial correlation coefficient of the random variables X and $Y^{(n)}$ with fixed n.

Further, based on the calculated value of the random variables $Y^{(n)}$, let us construct a family $\{\Psi_p\}$ of distributions of the participants' number according to scored points in the final stage, depending on the passing score p. For each of the distributions Ψ_p random variable $Y^{(n)}$ values are calculated λ_p, μ_p. Here p is the passing score into the final stage, λ_p is the number of stage winners, μ_p is the number of stage winners and awardees.

3. Applying of Passing Score Estimation for Determining the Number of the Olympiads' Final Stage Winners

Every year in the series of Olympiads in Informatics, held by ITMO University, about 4,000 students from 11th form of all federal regions of the Russian Federation participate in them. Olympiads are divided into two stages: the qualifying (remote) and final (intramural). Qualifying stage is divided into three rounds and lasts about four months. During the qualifying stage the participants are offered problems of all relevant school curriculum of computer science. To take part in the final stage those participants are admitted who received the passing score, which is established by Olympiad's jury.

Methodical Olympiad committee develop problems for qualifying and final stages, according to the criteria of determining the winners and awardees approved by the organizing committee. Thus, in accordance with the criteria, winner is the participant who decides both creative problems of programming technologies with proper solutions for at least 9 out of 10 problems on general issues of computer science & ICT at reproduction and usage levels. Awardee is the participant who solves 9 out of 10 problems on general issues of computer science & ICT, and neither managed nor solved creative problems. Thus, it can be stated that Olympiad's awardee knows perfectly well Computer Science & ICT at the levels of reproduction and use, but has not yet reached the level of creative mastery of the subject. Methodical committee and Olympiad's jury determine points for the final stage' problems, taking into account the above mentioned criteria. In determining the winners and awardees of the Olympiad's stage it is necessary to consider the following additional requirements: the winners' and awardees' number should not exceed 45% and the winners' number should not exceed 10% of the participants' number of that stage.

In a series of 2009–2010 academic year Olympiads, participants admitted to the final stage were those who scored 23 or more points out of 90 possible. Thus, the passing score is $p_{th} = 23$. The participants' number of the final stage is 931 students of 11th form. Virtually, less than a third of participants were admitted to participate in the final stage. Fig. 1 shows the distribution of participants according to scored points following the results of the qualifying stage among 11th form students in the series of 2009–2010 academic year Olympiads in Informatics.

Each problem of the final stage corresponds to a set of qualifying stage problems according to the theme and level of complexity. The average solvability index is estimated for this set of qualifying stage problems. The solvability indices of the final stage problems are compared with the average solvability indices of the corresponding sets of qualifying stage problems. The results are shown in Table 1.

The analysis shows that there is no direct correlation between the solvability index of the qualifying stage problems and solvability index of the corresponding types of the final stage problems. To estimate the number of final stage winners and prize-winners, depending on the value of the passing score when $p < 23$, it is necessary to determine the principles of assigning points for the final stage problems. Several approaches are

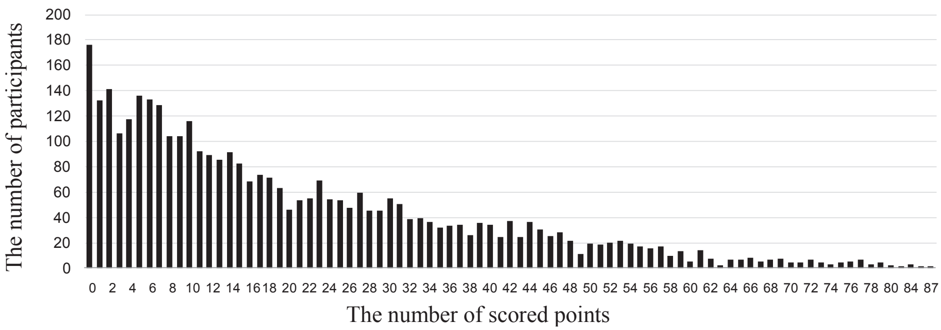

Fig. 1. Histogram of qualifying stage participants' distribution according to scored points.

Table 1

Problems' solvability indices according to Olympiad's stages

Problem №	1	2	3	4	5	6	7	8	9	10	11	12
Final stage	45,86	43,39	35,23	44,68	44,36	68,85	29,43	41,03	40,28	53,38	26,42	27,71
Qualifying stage	29,43	39,42	33,30	33,51	8,92	40,49	29,32	34,59	32,01	46,62	15,79	17,08

considered: expert evaluation (option "A"), evaluation according to solvability index of corresponding problems of the qualifying stage (Option "B").

Using the data of Table 2 and the actual results of the participants, participants' distribution of the final stage is made according to scored points for options "A" and "B" points' distribution according to problems (Fig. 2 and Fig. 3).

Besides, the error estimation of the options «A» and «B» in terms of principle ranking participants according to scored points, rather than types of solved problems. The

Table 2

Scores distribution according to problems

Problem №	1	2	3	4	5	6	7	8	9	10	11	12
	Number of points											
Option "A"	1	2	1	1	2	2	2	1	2	1	3	3
Option "B"	5	3	4	4	10	2	5	4	4	1	8	8
Option "C"	5	6	7	6	7	1	9	6	7	4	10	10

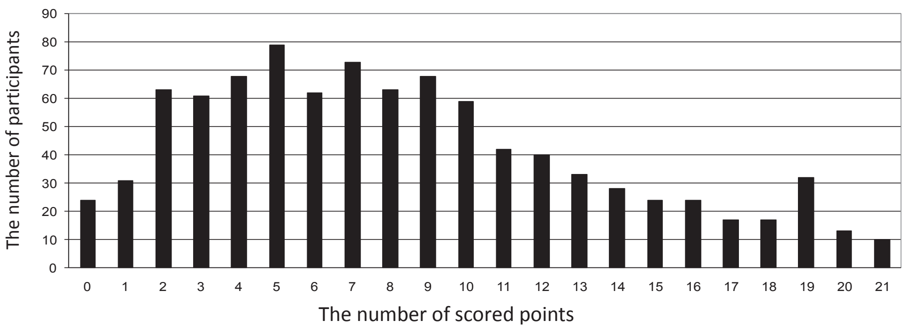

Fig. 2. Histogram of the final stage participants' distribution according to the number of scored points. The participants' number – 931. Option "A" of distribution points according to the problems (see: Table 2).

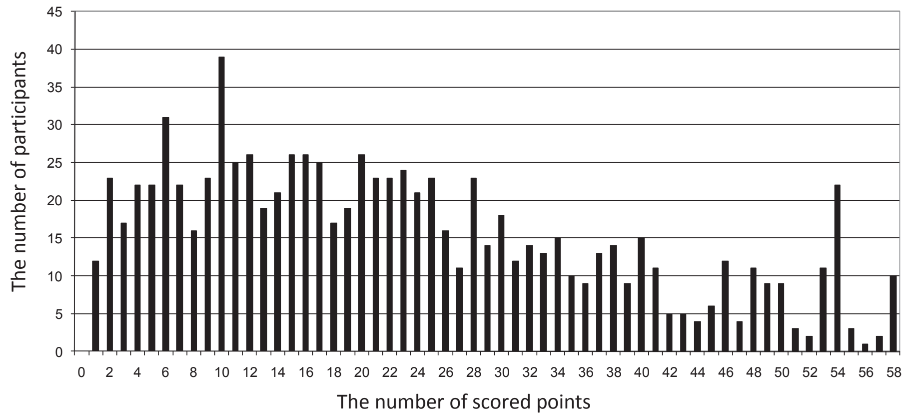

Fig. 3. Histogram of the final stage participants' distribution according to the number of scored points. The participants' number – 931. Option "B" of distribution points according to the problems (see: Table 2).

error/ inaccuracy occurs when the winner or the awardee is determined by the number of scored points, rather than composition of solved problems and means that the participant actually decided the set of problems meeting the winner criteria, but because of scored points attributed to the awardees, and vice versa.

"B" is preferred option of the two options «A» and «B», because it gives a smaller amount of error in comparison with the option «A» (see: Table 3).

The calculations of the values of the random variable allow with regard for the results of the qualifying stage – problems' solvability index and participants' distribution according to scored points – to estimate when $p < 23$ the number of the final stage winners

Data in Fig. 4 show the steady increase in the number of the final stage winners with a decrease in the values of a passing score.

The histogram in Fig. 5 shows the calculated values of the final stage winners' percentage depending on the passing score value obtained with the help of the distribution of

Table 3

The error/inaccuracy of the method of identifying winners and awardees' groups composition

	Error: winner-awardee	Overall error
Option "A"	89 participants	144 participants
Option "B"	35 participants	52 participants

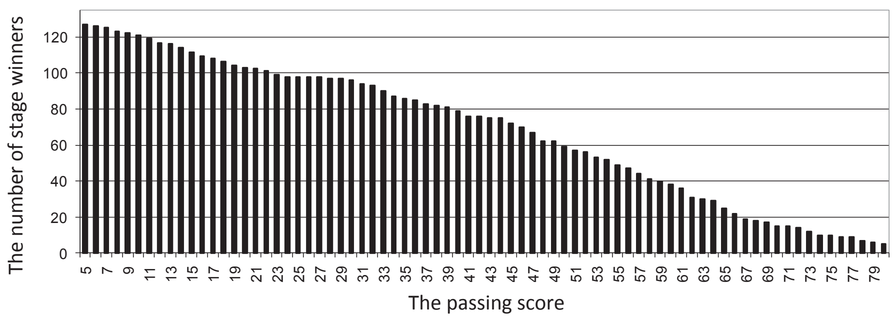

Fig. 4. The calculated number of winners depending on the value of the passing score.

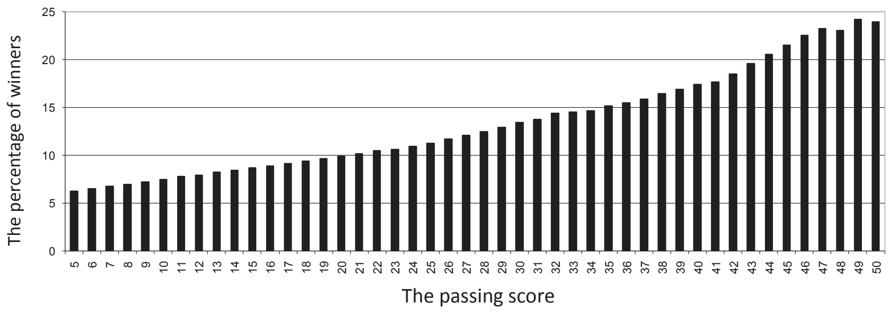

Fig. 5. Percentage of stage winners.

qualifying round participants according to points (see: Fig. 1), problems 'solvability indices (see: Table 1) and option "B" points distribution according to problems (see: Table 3).

4. Conclusion

In this paper a method of estimating the amount of the final stage winners of the Olympiad is proposed, depending on the value of passing score p in the final stage of the Olympiad.

It is obvious that there is a correlation between distribution of final stage participants according to scored points and distribution of qualifying stage participants according to scored points and an additional parameter – passing score in the final stage.

The results can be used to create methods for the development and estimation of problems' complexity of mass competition and Olympiad's finals.

References

Ayvazian S.A., Enyukov I.S., Meshalkin L.D. (1983). *Applied Statistics. Fundamentals of Modeling and Primary Data Processing*. Moscow, Finance and Statistics.

Bespalko V.P. (2002). *Education and Training with Computers* (*Pedagogy of the Third Millennium*). Moscow, Publishing House of the Moscow Psychological and Social Institute.

Cramer G. (1975). *Mathematical Methods of Statistics*. Moscow, Mir.

Elfving G. (1952). Optimum allocation in linear regression theory. *Ann. Math. Statist*, 23, 255–262. `http://projecteuclid.org/euclid.aoms/1177729442`

Krug G.K., Sosulin Y.A., Fatuev V.A. (1977). *The Planning of Experiments in Problems Identification and Extrapolation*. Moscow, Science.

A. Maiatin holds a Ph.D. in Pedagogical Sciences. He has been a deputy chairman of the methodical committee of open Olympiad in Informatics since 2008. His research interests other than education in computer science include an IT-infrastructure management based on technologies.

P. Mavrin is ACM ICPC World Champion in 2004, Silver medalist in Informatics of IOI in 2002, a member of the Technical Committee of IOI and ACM ICPC, an academic of Computer Technology chair of ITMO University where he teaches gifted school and university students.

V. Parfenov is a professor and dean of the faculty of information technologies and software engineering of ITMO University, a member of international organizing committee of the world programming championship, director of semifinal competitions in North-Eastern European region. He is one of the main organizers and creators of national and international Olympiads in computer science and programming for students and pupils in Russia. He has contributed a lot to the formation of educational system of search and training gifted in mathematics, physics, computer science and programming students since the year 2000.

O. Pavlova is an academic of Computer Technology chair of ITMO University. She has been teaching students having talent for mathematics, physics, computer science and programming since the year 2000. Nowadays her main research interest is the development of adolescents by investing in their education, in particular, by participating in Olympiads which considered one of the greatest contributions to children and adolescents' development.

D. Zubok holds a Ph.D. in physics and mathematics. He has been the deputy dean of information technology and software engineering department since 2006. Since 2008, the Executive secretary of the organizing committee of open Olympiad in Informatics and mathematics for school students. His current interests include computer science and data analysis.