Рогозов Юрий Иванович – Технологический институт федерального государственного автономного образовательного учреждения высшего профессионального образования «Южный федеральный университет» в г. Таганроге; e-mail: rogozov@tsure.ru; 347928, г. Таганрог, пер. Некрасовский, 44; тел.: 88634371787; кафедра системного анализа и телекоммуникаций; заведующий кафедрой; д.т.н.; профессор.

Свиридов Александр Славьевич – e-mail: sviridov@tsure.ru; кафедра системного анализа и телекоммуникаций; к.т.н.; доцент.

Rogozov Yury Ivanovich – Taganrog Institute of Technology – Federal State-Owned Autonomy Educational Establishment of Higher Vocational Education "Southern Federal University"; e-mail: rogozov@tsure.ru; 44, Nekrasovsky, Taganrog, 347928, Russia; phone: +78634371787; the department of system analysis and telecommunications; head the department; dr. of eng. sc.; professor.

Sviridov Alexander Slavevich – e-mail: sviridov@tsure.ru; the department of system analysis and telecommunications; cand. of eng. sc.; associate professor.

УДК 519.85:004.421

А.П. Попов

ИМИТАЦИЯ ПРОЦЕССА ТЕСТИРОВАНИЯ

Дано краткое описание основных принципов и предположений, на которых основана новая модель тестирования [1]. В рамках модели поиск решения тестового задания трактуется как однородный во времени (пуассоновский) стохастический процесс, а время решения тестового задания является случайной величиной, подчиняющейся гамма-распределению. Рассматриваются типичные результаты сравнения эмпирических распределений времени, необходимого для решения тестовых заданий, с зависимостью, предсказываемой теорией. Как правило, эмпирические и теоретические распределения хорошо согласуются друг с другом, но иногда это согласие нарушается. Ясно, что расхождение между теоретическими и эмпирическими распределениями вызвано малым объемом статистических данных. Это утверждение получило полное подтверждение в целой серии статистических экспериментов, имитировавших процесс тестирования. Результаты этих экспериментов позволяют дать рекомендации, которые должны учитываться в практике компьютерного тестирования.

Время поиска решения; тестовые задания; однородный во времени стохастический процесс; сравнение эмпирических и теоретических распределений времени поиска решения; статистический эксперимент; имитация процесса тестирования; практические рекомендации.

A.P. Popov

IMITATION OF TESTING PROCESS

There is given a short description of main principles and assumptions on which the new model of testing is based [1]. In the framework of model the search of test tasks solution is interpreted as time homogenous (Poisson's) stochastic process, and the time of test tasks solutions is random variable submitted to gamma distribution. There are considered the typical results of comparison of empirical distributions of times necessary for test tasks solutions with dependences predicted by theory. As a rule, empirical and theoretical distributions are in a good agreement each one to other but sometimes the agreement is broken up. Is clear the divergences between empirical and theoretical distributions are caused by small volumes of statistical data. This statement gets the full confirmation in series of statistical experiments, where the testing process is imitated. The results of these experiments allow one to give some recommendations which must be taken into account in practice of computer testing.

Times of search of solutions; test tasks; time homogenous stochastic process; comparison of empirical and theoretical distributions of time of search of solutions; statistical experiments; imitations of process of computer testing; the practical recommendations.

Введение. В работе [1] описана принадлежащая автору новая модель тестирования, в которой поиск решения тестовых заданий трактуется как однородный во времени стохастический процесс. При этом время поиска верного решения тестового задания оказывается случайной величиной, которая подчиняется хорошо известному в теории вероятностей гамма-распределению [6–7]:

$$f\left(\alpha,\lambda,t\right)=\frac{\left(\lambda t\right)^{\alpha-1}}{\Gamma\left(\alpha\right)}e^{-\lambda t}\lambda.. \tag{1}$$

В рамках модели безразмерный параметр α интерпретируется как трудность тестового задания, а имеющий размерность обратного времени параметр λ ассоциируется с уровнем подготовленности испытуемого.

Результаты эмпирической проверки адекватности модели. Новая модель прошла всестороннюю проверку [1–5], подтвердившую ее адекватность. В частности, проверялось согласие эмпирических распределений времени, затрачиваемого на поиск верного решения тестовых заданий, с предсказываемой теорией зависимостью (1).

Графики на рис. 1 получены в результате обработки данных сессии компьютерного тестирования по психологии, в котором участвовали 64 студента факультета МИиФ ПИ ЮФУ. Согласие теории с эмпирическими данными далеко не всегда проступает явно, что может вызвать сомнения в справедливости предлагаемой модели тестирования. Чтобы убедиться в безосновательности подобных сомнений, было решено провести целенаправленный статистический эксперимент, имитирующий процесс тестирования.

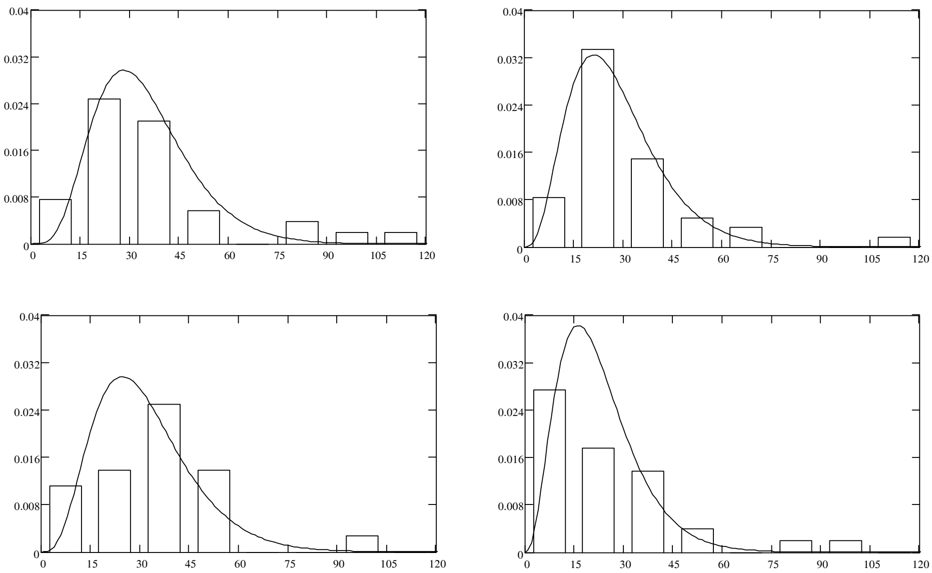

Рис. 1. Распределение времени решения ТЗ в тестах по психологии:
1) $\alpha=5{,}61$; $\lambda=0{,}163$ c^{-1}; 2) $\alpha=4{,}14$; $\lambda=0{,}148$ c^{-1}; 3) $\alpha=4{,}55$; $\lambda=0{,}143$ c^{-1};
4) $\alpha=3{,}59$; $\lambda=0{,}159$ c^{-1}

Имитация процесса тестирования. Здесь описаны результаты имитации процесса тестирования в случае, когда тест состоит из единственного задания трудности α, а уровень подготовленности всех испытуемых одинаков и равен λ. В статистических экспериментах использовался генератор последовательностей псевдослучайных величин, подчиняющихся гамма-распределению из пакета Mathcad.

1-й эксперимент. Время поиска решения тестового задания в генерируемой последовательности подчинено гамма-распределению с параметрами α=2,5; λ=0,025 $с^{-1}$. Длина последовательности совпадает с числом испытуемых, и в данном эксперименте L=25.

2-й эксперимент. Время поиска решения тестового задания в генерируемой последовательности подчинено гамма-распределению с параметрами α=2,5; λ=0,025 $с^{-1}$. Длина последовательности совпадает с числом испытуемых, и в данном эксперименте L=100.

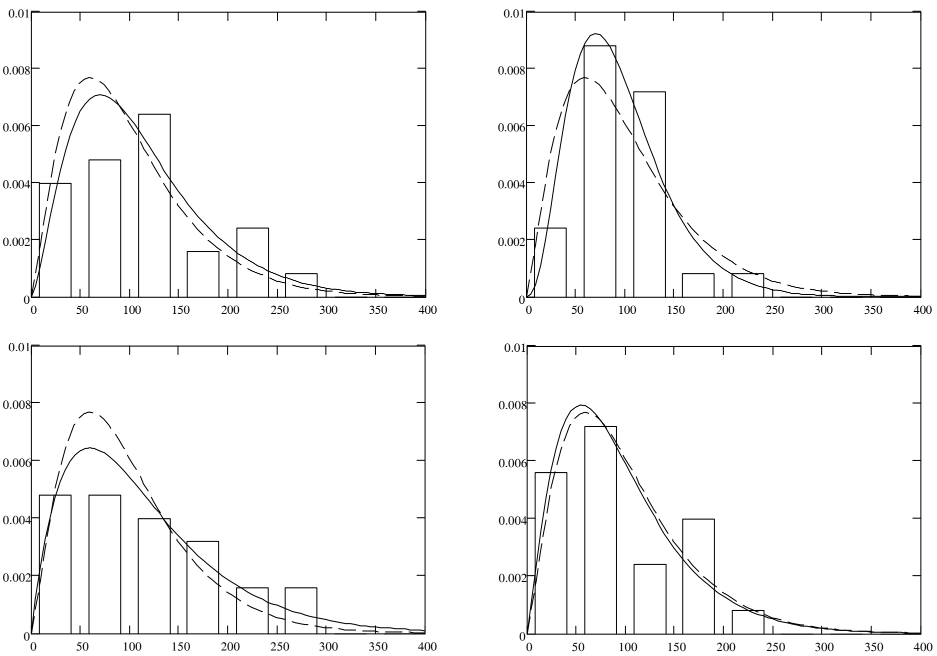

Рис. 2. Распределение времени решения ТЗ в 1-м эксперименте:
1) α=2,71; λ=0,0244 $с^{-1}$; 2) α=3,82; λ=0,0400 $с^{-1}$; 3) α=2,08; λ=0,0181 $с^{-1}$;
4) α=2,37; λ=0,0247 $с^{-1}$

Используя известные формулы [7] для математического ожидания и среднеквадратичного отклонения случайной величины, подчиняющейся гамма-распределению, можно оценить параметры эмпирического распределения, используя заранее вычисленные значения выборочного среднего и выборочного среднеквадратичного отклонения:

$$\alpha = \left(\frac{\langle t \rangle}{\sigma(t)} \right)^2 \quad \lambda = \frac{\langle t \rangle}{\sigma(t)^2}. \tag{2}$$

Именно эти оценки приведены в подписях к рис. 2 и 3.

Параметры псевдослучайных последовательностей (ПСП), генерируемых в обоих статистических экспериментах, совпадают, но число испытуемых, участвующих во 2-м эксперименте, в 4 раза больше. Это приводит к тому, что во 2-м эксперименте эмпирические и теоретические распределения времени поиска решения тестового задания отличаются заметно слабее. Во 2-м эксперименте теоретические распределения почти не отличаются и от исходного гамма-распределения с параметрами, заданными в настройках генератора ПСП (график распределения изображен на рисунках пунктиром). Наконец, найденные во 2-м эксперименте оценки параметров ПСП существенно меньше отклоняются от значений, заданных в настройках генератора. На рис. 4 и 5 показано распределение оценок параметров ПСП, полученных в результате многократного (10 000 раз) повторения статистических экспериментов. Методами математической статистики были найдены выборочные средние и выборочные среднеквадратичные отклонения этих оценок.

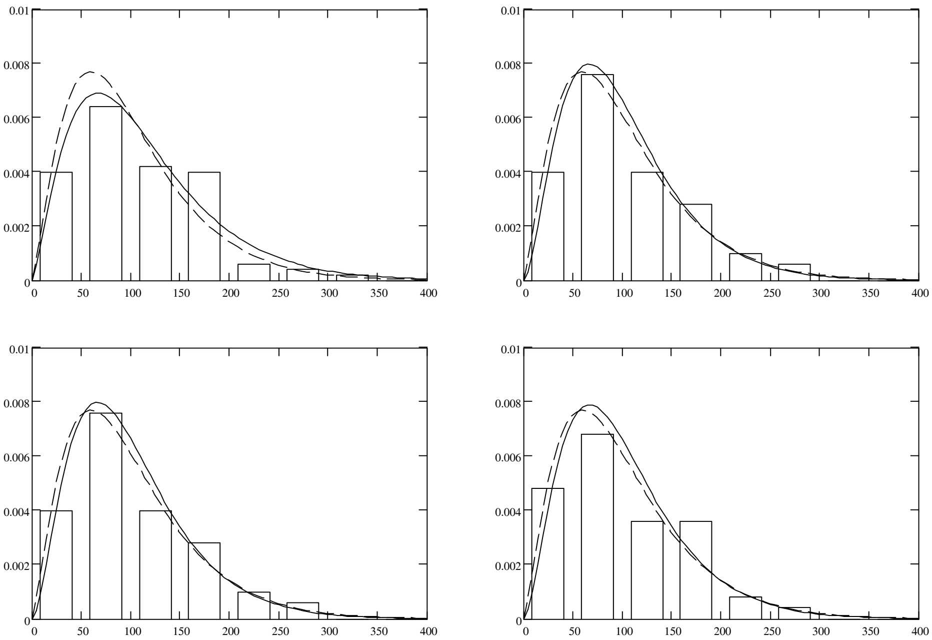

Рис. 3. Распределение времени решения ТЗ во 2-м эксперименте:
1) α=2,706; λ=0,0295 с$^{-1}$; 2) α=2,509; λ=0,0225 с$^{-1}$; 3) α=2,939; λ=0,0291 с$^{-1}$;
4) α=2,908; λ=0,0286 с$^{-1}$

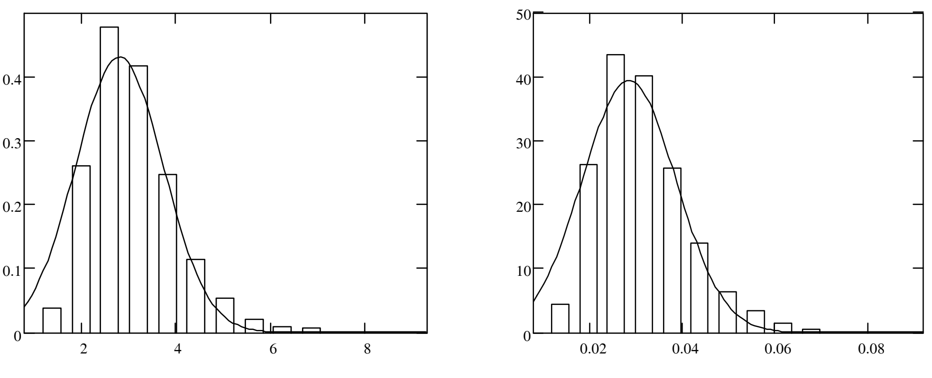

Рис. 4. Распределение оценок параметров ПСП, полученных при многократном повторении статистического 1-го эксперимента

Выборочные средние и выборочные среднеквадратичные отклонения оценок параметров ПСП, полученных при многократном повторении статистического 1-го эксперимента: $\alpha_{ср} = 2{,}81$; $\sigma(\alpha) = 0{,}93$; $\lambda_{ср} = 0,\ 0285$ с$^{-1}$; $\sigma(\lambda) = 0{,}0101$ с$^{-1}$.

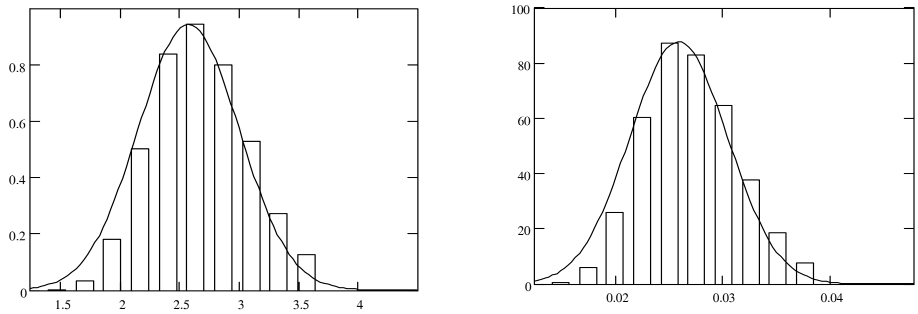

Рис. 5. Распределение оценок параметров ПСП, полученных при многократном повторении статистического 2-го эксперимента

Выборочные средние и выборочные среднеквадратичные отклонения оценок параметров ПСП, полученных при многократном повторении статистического 2-го эксперимента: $\alpha_{ср} = 2{,}58$; $\sigma(\alpha) = 0{,}42$; $\lambda_{ср} = 0{,}\ 0259$ с$^{-1}$; $\sigma(\lambda) = 0{,}0045$ с$^{-1}$.

Для объяснения результатов, полученных в статистических экспериментах, воспользуемся общим принципом математической статистики, полная и строгая формулировка которого, по-видимому, впервые была дана в 20-х годах прошлого века в работах Р. Фишера.

Принцип максимального правдоподобия. Согласно этому принципу наиболее вероятные значения параметров распределения определяются из условия максимума функции правдоподобия.

Функция правдоподобия по Фишеру в нашем случае совпадает с вероятностью появления на выходе генератора данной конкретной псевдослучайной последовательности:

$$F(\alpha,\lambda) = \prod_{j=1}^{N} f(\alpha,\lambda,t_j). \tag{3}$$

На практике, однако, удобнее иметь дело с логарифмической функцией правдоподобия:

$$L(\alpha,\lambda) = \ln F(\alpha,\lambda) = \sum_{j=1}^{N} f(\alpha,\lambda,t_j). \tag{4}$$

Подставляя в общую формулу (4) выражение (1) для плотности гамма-распределения позволяет получить явное выражение для логарифмической функции правдоподобия в рассматриваемом нами конкретном случае:

$$L(\alpha,\lambda) = N\ln\left(\frac{\lambda^{\alpha}}{\Gamma(\alpha)}\right) + \sum_{j=1}^{N}\left((\alpha-1)\ln(t_j) - \lambda t_j\right). \tag{5}$$

Необходимые условия максимума функции (5) приводят к системе нормальных уравнений правдоподобия (СНУП):

$$\begin{cases} \dfrac{\partial L}{\partial \alpha} = \sum\limits_{j=1}^{N} \ln(\lambda t_j) - N\psi(\alpha) = 0, \\ \dfrac{\partial L}{\partial \lambda} = N\dfrac{\alpha}{\lambda} - \sum\limits_{j=1}^{N} t_j = 0. \end{cases} \tag{6}$$

Очевидные преобразования позволяют привести систему (6) к виду, более пригодному для непосредственных численных расчетов:

$$\begin{cases} \psi(\alpha) - \ln(\alpha) = \dfrac{1}{N} \sum\limits_{j=1}^{N} \ln\left(\dfrac{t_j}{\langle t \rangle} \right), \\ \lambda = \dfrac{\alpha}{\langle t \rangle}. \end{cases} \tag{7}$$

Согласно известной теореме Рао–Крамера–Фреше [8–9] оценки максимального правдоподобия параметров ПСП состоятельны, их распределение асимптотически нормально, а матрица дисперсии асимптотически совпадает с обратной матрицей информации.

Матрица информации по Фишеру, с точностью до знака, равна усредненной матрице вторых производных от логарифмической функции правдоподобия по параметрам ПСП:

$$I = N \begin{pmatrix} \psi'(\alpha) & -\dfrac{1}{\lambda} \\ -\dfrac{1}{\lambda} & \dfrac{\alpha}{\lambda^2} \end{pmatrix}. \tag{8}$$

Используя (8), нетрудно получить асимптотически точные формулы для среднеквадратичных отклонений параметров:

$$\sigma(\alpha) = \sqrt{\frac{\alpha}{N(\alpha\psi'(\alpha) - 1)}} \quad \sigma(\lambda) = \lambda \sqrt{\frac{\psi'(\alpha)}{N(\alpha\psi'(\alpha) - 1)}}. \tag{9}$$

Разработанный нами эффективный численный метод решения системы (7) позволяет получить точечные оценки параметров. Вычислив затем среднеквадратичные отклонения (9) стандартными методами математической статистики, можно найти также и интервальные оценки параметров ПСП.

Следует отметить, что оценки максимального правдоподобия, получающиеся в результате решения системы (7), вообще говоря, не совпадают со статистическими оценками (2). Как показала практика, а также статистические эксперименты, нет достаточных оснований для того, чтобы отдать предпочтение какой-то одной из этих оценок.

Оценки среднеквадратичного отклонения (9) параметров ПСП оказываются заниженными в сравнении с выборочными оценками.

Например, в условиях 1-го статистического эксперимента расчет среднеквадратичных отклонений параметров в соответствии с (9) приводит к значениям $\sigma(\alpha) = 0{,}67$; $\sigma(\lambda) = 0{,}0074$ c^{-1}, что заметно меньше ранее найденных выборочных оценок $\sigma(\alpha) = 0{,}93$; $\sigma(\lambda) = 0{,}0101$ c^{-1}.

В условиях 2-го статистического эксперимента подобный расчет приводит к значениям $\sigma(\alpha) = 0{,}33$; $\sigma(\lambda) = 0{,}0037$ c^{-1}, что также несколько ниже выборочных оценок $\sigma(\alpha) = 0{,}42$; $\sigma(\lambda) = 0{,}0045$ c^{-1}.

На рис. 4–5 помимо гистограмм эмпирического распределения значений параметров ПСП, построенных по данным статистических экспериментов, показаны также графики нормальных распределений, средние значения и среднеквадратичные отклонения которых совпадают с выборочными оценками. В соответствии с теоремой Рао–Крамера–Фреше, эти распределения должны служить хорошей аппроксимацией эмпирических распределений при большой длине генерируемых ПСП, иначе говоря, при большом числе испытуемых, участвующих в тестировании. При числе испытуемых N = 25 различие между эмпирическим и нормальным распределением видно, что называется, невооруженным глазом, в частности, эмпирическое распределение отличается ярко выраженной асимметрией. Но уже при числе испытуемых N=100 эмпирическое распределение практически ничем не отличается от нормального распределения.

Заключение. Точность оценки трудности тестовых заданий существенно зависит от объема статистической выборки, т.е. от числа испытуемых, участвующих в тестировании. Чтобы добиться приемлемой точности оценки рейтинга участвующих в тестировании студентов, их общее количество должно быть не менее 100. Из этого можно сделать вывод, что для получения надежных результатов тестирование следует проводить лишь по тем дисциплинам, которые читаются в больших потоках, включающих не менее четырех полноценных академических групп.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. *Попов А.П., Богомолов А.А., Попова Л.А.* Новая математическая модель тестирования // Наука и образование. – 2005. – № 3. – С. 221.
2. *Попов А.П.* Новое направление в теории тестирования // Известия ЮФУ. Педагогические науки. – 2008. – № 1–2. – С. 24.
3. *Попов А.П., Попова Т.Ю.* Адекватность новой модели тестирования. Проверка гипотезы о распределении времени решения тестовых заданий // Материалы НМК СИТО 2009. – Ростов-на-Дону, 17–18 апреля 2009 г. – Ростов-на-Дону, 2009. – С. 234-235.
4. *Попов А.П., Акулов С.Ю., Попова Т.Ю.* Адекватность новой модели тестирования. Проверка гипотезы об аддитивности трудности тестовых заданий // Материалы НМК СИТО 2009. – Ростов-на-Дону, 17-18 апреля 2009 г. – Ростов-на-Дону, 2009. – С. 25-27.
5. *Попов А.П., Попова Т.Ю., Акулов С.Ю.* О принципиально новом направлении в теории тестирования // Грани познания: электронный журнал ВГПУ. – 2009. – № 4 (5). URL: http: // www.grani.vspu.ru.
6. *Гнеденко Б.В.* Курс теории вероятностей. – М., 1988. – 448 с.
7. *Королюк В.С., Портенко Н.И., Скороход А.В., Турбин А.Ф.* Справочник по теории вероятностей и математической статистике. – М., 1985. – 640 с.
8. *Фаддеева Л.Н., Жуков Ю.В., Лебедев А.В.* Теория вероятностей и математическая статистика. – М., 2006. – 336 с.
9. *Колемаев В.А., Староверов О.В., Турундаевский В.Б.* Теория вероятностей и математическая статистика. – М., 1991. – 400 с.

Статью рекомендовал к опубликованию д.т.н., профессор А.Н. Чернухин.

Попов Александр Петрович – Педагогический институт федерального государственного автономного образовательного учреждения высшего профессионального образования «Южный федеральный университет»; e-mail: nanosys@mail.ru; 344068, г. Ростов-на-Дону, ул. Криворожская, 57, кв. 20; тел.: +79094412895; отдел контроля качества образования; начальник.

Popov Alexander Petrovich – Federal State-Owned Autonomy Educational Establishment of Higher Vocational Education “Southern Federal University”; e-mail: nanosys@mail.ru; 57, ap. 20, Krivorozhskaya street, Rostov-on-Don, 344068, Russia; phone: +79094412895; the department of education quality control; chief.