
Математическое моделирование, численные методы и комплексы программ

УДК:519.711

DOI: 1030987/article_5bfd98c46c9ec9.95927169

Е.С. Гарбузьяк, О.В. Сташкова, О.В. Шестопад, Л.Я. Козак

ОЦЕНКА СТАТИСТИЧЕСКИХ ХАРАКТЕРИСТИК МАТЕМАТИЧЕСКОЙ МОДЕЛИ МЕТОДОМ ПУГАЧЕВА

В статье рассмотрено современное состояние исследуемой задачи, описан используемый алгоритм решения задачи по методу коррелированных процессов, проведен обзор других методов решения подобных задач, а также приведено описание инструментов, с помощью которых был разработан программный продукт; детально описываются дополнительные ресурсы, используемые при разработке программного продукта, приводится его конечная структура и описание пользовательских классов и методов; содержатся результаты тестирования и анализ полученных результатов.

Ключевые слова: статистические оценки; метод коррелированных процессов; метод Пугачёва.

E.S. Garbuznyak, O.V. Stashkova, O.V. Shestopal, L.Y. Kozak

ESTIMATION OF STATISTICAL CHARACTERISTICS OF THE MATHEMATICAL MODEL BY PUGACHEV'S METHOD

The purpose and objectives of the study determined the structure of the article. It consists of an introduction, two sections, a conclusion, a list of references and an appendix.

The first section is considered with the current state of the problem under study, describes the algorithm used to solve the problem using the method of correlating processes, reviews other methods for solving similar problems, and describes the tools used to develop the software product.

The second section is devoted to the development of a software product. It describes in more detail the additional resources used in the development, the final structure of the software product and a description of the user classes and methods.

Also, the second section contains the results of testing and analysis of the results. The conclusion contains the results of the work done, the features of the developed software product are described, practical value and prospects are indicated.

Keywords: statistical assessments; method of correlating processes; method of Pugachev.

Введение

В связи с усложнением современных технических систем, повышением требований к ним существует практическая потребность в повышении надёжности проектных и поверочных расчётов. Проблема учёта погрешности исходных данных задачи является одной из наиболее актуальных, что объясняется необходимостью эффективного получения статистических оценок вероятностных характеристик с заданной точностью, а также с минимальными потерями машинного времени.

Объект исследования – получение статистических оценок вероятностных характеристик с заданной точностью.

Предмет исследования – использование метода коррелированных процессов для сокращения времени на получение статистических оценок.

Целью исследования является разработка программного продукта, реализующего оценку статистической характеристики математической модели методом Пугачёва.

1. Современное состояние исследуемой задачи

Из классических методов, позволяющих при соответствующих условиях учитывать погрешность параметров модели, оптимальным является метод малого параметра. Точность этого метода существенно зависит от величины погрешности исходных данных. Другим его недостатком является отсутствие гарантий достоверности результатов вычислений.

Относительно новым подходом к решению рассматриваемой проблемы является применение двусторонних и интервальных методов [2, 6, 7]. Однако известные интервальные методы являются эффективными только в тех случаях, когда погрешность модели достаточно мала.

Описываемый в [5] подход основан на использовании вариационного принципа. Он может применяться как в случае систем с сосредоточенными, так и с распределенными параметрами, и накладывает существенно меньшие ограничения на величину неустранимой погрешности.

Перечисленные методы позволяют решить лишь одну из задач теории допусков, а именно, задачу расчёта так называемого «наихудшего случая». Кроме того, интервальные методы неэффективны при решении стохастических задач, в которых случайные параметры имеют большие дисперсии.

Более широкими возможностями обладают методы статистического моделирования (Монте-Карло), которые широко используются на практике [1, 2, 3, 8]. Однако их применение часто приводит к большим затратам машинного времени. Это связано с тем, что при использовании методов Монте-Карло приходится многократно решать одну и ту же задачу, но при разных значениях исходных данных.

Для сокращения времени на получение статистических оценок вероятностных характеристик с заданной точностью предложено много методов: метод расслоенной выборки, метод существенной выборки, метод коррелированных процессов, метод выборки с «оценкой по отношению», метод случайных квадратурных формул, метод выделения главной части [2, 3, 8]. В основном повышение эффективности в этих методах достигается за счёт понижения дисперсии статистической оценки. В методе расслоенной выборки повышение точности оценки достигается за счёт выбора наибольшего числа случайных величин из тех областей пространства значений случайных величин, где результаты моделирования имеют наибольший разброс. При реализации расслоенной выборки пространство значений случайных величин разбивается на несколько попарно непересекающихся подмножеств (слоёв) и из каждого слоя берется простая случайная выборка фиксированного объёма. Использование специального вида функции для вычисления вероятностной характеристики позволяет гарантировать несмещённость её оценки, а объёмы выборок в слоях определяются так, чтобы обеспечить дисперсию оценки, меньшую, чем при простой случайной выборке.

Метод существенной выборки близок к методу расслоенной выборки, но здесь выбор точек регулируется не заданием числа точек в областях, а специальной функцией плотности вероятностей.

Одним из универсальных подходов к уменьшению дисперсии оценок является метод случайных квадратурных формул, представляющий собой обобщение обычного метода Монте-Карло.

Выборка с «оценкой по отношению» так же, как и корреляционная выборка, требует для своей реализации функцию, хорошо аппроксимирующую оператор исходной модели с известными значениями вероятностных характеристик.

Эффективность ряда вариантов метода расслоенной выборки также существенно зависит от «близости» упрощённой и исходной модели исследуемого процесса.

Таким образом, технология наиболее употребительных методов ускорения статистического моделирования предполагает построение некоторой упрощённой модели.

Это требует учёта множества особенностей, с которыми приходится иметь дело при решении задач из конкретных предметных областей.

Рассмотрим метод коррелированных процессов применительно к задачам статистического моделирования технологических процессов с приближённо заданными или случайными параметрами.

2. Обзор методов решения подобных задач

2.1. Общее описание методов

Теоретические основы методов статистического моделирования сводятся к одному базовому положению математической статистики – теореме Хинчина. Принцип статистического моделирования заключается в том, чтобы нивелировать случайность (как категорию чисто статическую) за счёт проведения большого количества опытов и определить таким образом искомую величину с той или иной степенью уверенности.

Для определения количественных характеристик следует опираться на центральную предельную теорему, согласно которой случайная величина $\bar{x}_n = \frac{x_1+x_2+\dots+x_n}{n}$ распределена по нормальному закону с математическим ожиданием, равным математическому ожиданию величины x и дисперсией $D_{\bar{x}_n} = \frac{\sigma_{x_n}^2}{n}$.

Количество опытов, необходимое для достижения заданной погрешности с заданной доверительной вероятностью при прямом статистическом моделировании:

$$N_{\text{треб}} = \frac{\alpha_{\text{дов}}^2}{\varepsilon_{\text{доп}}^2}, \quad (1)$$

где $\alpha_{\text{дов}}$ – доверительный интервал; $\varepsilon_{\text{доп}}$ – допустимая погрешность.

Модель исследуемой системы задается в виде системы уравнений:

$$X_i = X(V), i = 1, 2, \dots, n \quad (2)$$

где $V = (V_1, V_2, \dots, V_m)$ – вектор случайных параметров; m – их количество; n – число опытов при статистическом моделировании.

Таким образом, задача статистического моделирования сводится к оценке математического ожидания $m_x = M[X(V)]$.

Для сокращения количества проводимых экспериментов применяются различные методы, некоторые из которых, описанные ниже.

2.2. Метод выделения главной части

Решение системы $X(V)$, которое, возможно, не может быть найдено аналитически, заменяют приближенным выражением $Y(V)$, удобным для аналитических преобразований.

Вводится новая переменная $Z(V) = X(V) - Y(V)$ и в системе уравнений (2) выполняется замена переменной X на Z путём следующей подстановки: $X(V) = Z(V) + Y(V)$.

Оценка искомого математического ожидания определяется в виде $m_x = M[Y(V)] + M[Z(V)]$, где первое слагаемое может быть найдено аналитически, а второе определяется по методу статистического моделирования. При удачном выборе функции $Y(V)$ дисперсия случайной величины $Z(V)$ может оказаться существенно меньше, чем дисперсия $X(V)$, что и приведёт к сокращению требуемого количества опытов.

2.3. Метод выборки по группам

В соответствии с данным методом область G возможных значений случайного вектора разбивается на K непересекающихся областей G_k : $G = G_1 \cup G_2 \cup \dots \cup G_k \cup \dots \cup G_K$.

Метод предполагает проведение статистического моделирования для каждой из областей G_k с использованием вектора случайных параметров плотностей распределения

вероятностей:

$$f_{V^{(k)}}(V) = \frac{f_V(V)}{p_k} \quad (3)$$

где $f_V(V)$ – плотность распределения вероятности на всей области определения вектора V ; $f_{V^{(k)}}(V)$ – плотности распределения отдельных слоёв; p_k – вероятность попадания случайного вектора V в область G_k .

Для каждой области определяется оценка математического ожидания и дисперсия оценки как усреднение значения математических ожиданий всех областей.

3. Описание применяемого алгоритма

Обозначим через λ искомый n -мерный вектор вероятностных характеристик исходной системы, а через μ m -мерный вектор вероятностных характеристик упрощённой системы: $\lambda = M[R]$; $\mu = M[S]$, где $M[\dots]$ – математическое ожидание величины, входящей в скобки; R и S – n - и m -мерные векторы, компоненты которых представляют собой некоторые функции от значений процессов соответственно в исходной и упрощённой системах.

Предположим, что с исходной и упрощённой системами проведено в одинаковых условиях N независимых между собой вычислительных экспериментов по методу Монте-Карло. Статистические значения λ^* и μ^* векторов λ и μ , найденные по этим N экспериментам, будут равны:

$$\lambda^* = \frac{1}{N} \sum_{j=1}^N R_j; \quad (4)$$

$$\mu^* = \frac{1}{N} \sum_{j=1}^N S_j. \quad (5)$$

Не нарушая общности, будем считать, что эксперименты с одинаковыми внешними воздействиями на исходную и упрощённую системы имеют одинаковый номер, т.е. значения R_j и S_j получены при одних и тех же воздействиях.

Предполагается, что точное значение вектора μ вероятностных характеристик упрощённой системы может быть найдено аналитически или каким-либо подходящим приближённым методом, но с высокой точностью (например, при помощи однофакторной модели исходной системы).

Метод коррелированных процессов основан на оптимальной оценке λ_0 вектора λ по значениям векторов λ^* , μ^* и μ , для чего отыскивается оценка вероятностных характеристик исходной системы по статистическим значениям вероятностных характеристик исходной и упрощённой системы. Таким образом, при статистическом моделировании используются результаты аналитического упрощённого исследования [8]. Искомая оценка имеет вид:

$$\lambda_{i0} = \lambda_i^* - K_{R_i S} K_{SS}^{-1} (\mu^* - \mu), \quad (6)$$

где $\lambda_i^* = \frac{1}{N} \sum_{j=1}^N R_{ij}$; $K_{R_i S} = M[R_i (S - \mu)^T]$; $K_{SS} = M[(S - \mu)(S - \mu)^T]$,

λ_{i0} и λ_i^* – i -е компоненты векторов λ_0 и λ^* соответственно; R_{ij} – i -я компонента вектора R_j .

Разность $\mu^* - \mu$ равна статистической ошибке в определении вектора μ , которую удаётся найти благодаря знанию точного или практически точного значения μ этого вектора.

Величина $K_{R_i S} K_{SS}^{-1} (\mu^* - \mu)$ представляет собой значение ошибки, пересчитанное для величины λ_i^* с учётом корреляционной связи λ_i^* и μ^* . Для получения оценки λ_{i0} это пересчитанное значение ошибки вычитается из статистического значения λ_i^* величины λ_i . В результате оценка λ_{i0} получается более точной, чем статистическое значение λ_i^* .

Оценка каждой из компонент вектора λ может быть произведена независимо от оценки других компонент.

Для практического использования формулы (6) необходимо знать корреляционные матрицу-строку $K_{R_i S}$ и матрицу K_{SS} .

Так как упрощённая система может исследоваться аналитическим или во многих случаях экономичным приближенным методом (по крайней мере для определения вектора μ), то возможно экономичное вычисление аналитическим или приближенным методом и с высокой точностью значения корреляционной матрицы K_{SS} . Однако в некоторых случаях определение K_{SS} может оказаться затруднительным. При этом вместо K_{SS} можно найти её статистическое значение K_{SS}^* , вычисленное по тем же N экспериментам, по которым были найдены λ^* и μ^* , т.е. по формуле $K_{SS}^* = \frac{1}{N} \sum_{j=1}^N (S_j - \mu)(S_j - \mu)^T$.

Исходная система не поддается аналитическому исследованию, поэтому вместо корреляционной матрицы-строки $K_{R_i S}$ можно найти только её статистическое значение $K_{R_i S}^*$ по формуле $K_{RS}^* = \frac{1}{N} \sum_{j=1}^N (R_j - \mu)(S_j - \mu)^T$.

Важным моментом является построение приближенной модели. Как правило, это достигается посредством линеаризации или уменьшения детальности, числа факторов и т.п. исходной модели.

4. Разработка программного продукта

Для разработки программного продукта был выбран язык программирования *Python*, который является мощным инструментом для создания программ самого разнообразного назначения. Платформой для реализации программного продукта стала *PyCharmCommunity*.

Структура программного продукта представлена несколькими Python-пакетами и несколькими скриптами:

- пакет виртуального окружения *env*;
- пакет *forms*;
- пакет *lib*;
- скрипт *main.py*;
- скрипт *main_console.py*.

Кроме этого, программный продукт содержит некоторые служебные файлы и папки, не влияющие на работу программы, но использованные для контроля версий или настроек проекта.

Исходными данными для работы программы являются две матрицы выборок: малой и большой, расположенные на листе *MicrosoftExcel* формата **.xls*. Для корректной работы программ необходимо указать положение матриц на листе, поскольку один лист может содержать не только данные, но и описательную информацию, которая не используется в расчётах. Кроме того, необходимо задать диапазон ячеек для записи результатов вычислений, представляющие собой вектор значений.

В результате работы программного продукта получается файл формата **.xls*, который представляет собой копию исходного файла с дописанным вектором результата в указанную ячейку.

5. Основные скрипты программы и виртуального окружения

Скрипты *main.py* и *main_console.py* используются для запуска программного продукта. Скрипт *main.py* инициализирует интерфейс программного продукта и передаёт ему управление работой программы, а скрипт *main_console.py* используется в частности для отладки. Виртуальное окружение позволяет хранить только те модули и библиотеки, которые используются данным проектом. В данном случае каталог *env* хранит копию

интерпретатора *Python*, менеджер пакетов *pip* 8.1.1, а также все дополнительные модули и библиотеки, от которых зависит данный проект.

Графический пользовательский интерфейс реализован в среде *QtDesigner*, а затем компилировался в файлы с кодом *Python*. Его реализация в структуре программного продукта представлена пакетом *forms*.

Файл *mainform.py* представляет собой скомпилированный класс формы *mainform.ui* и реализует интерфейс программного продукта (рис. 1).

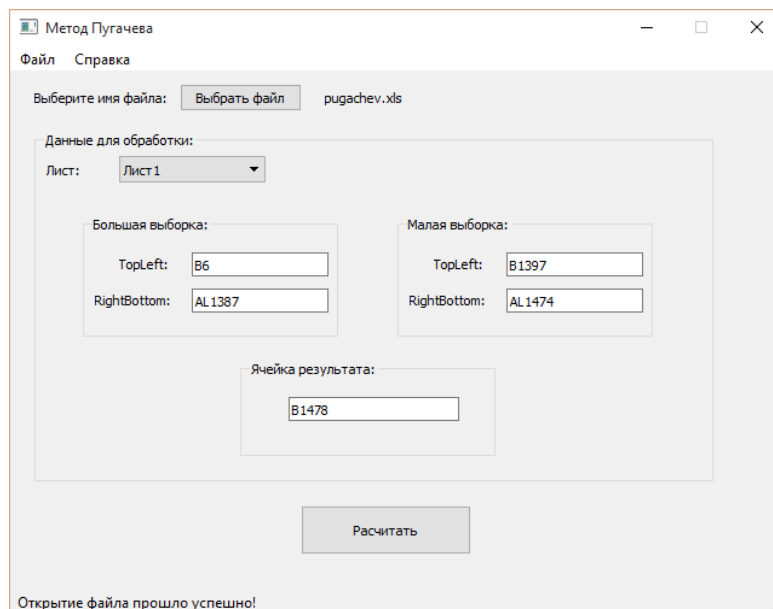


Рис. 1. - Главная форма с выбранным именем файла

Реализация логики программы и непосредственно метода Пугачёва осуществлена в пакете *lib*. Этот пакет содержит 3 файла: *__init__.py*, *functions.py*, *io.py*.

Файл *__init__.py* является инициализирующим пакет. Также в нём находится реализация главного фасадного класса *PugachevMethod*.

Остальные файлы пакета (*functions.py* и *io.py*) используются классом *PugachevMethod*. Файл *functions.py* содержит вспомогательные функции, не относящиеся к классам. Файл *io.py* содержит класс *ExcelManager*, который отвечает за считывание и запись данных в файлы **.xls*. Класс *PugachevMethod* использует функционал класса *ExcelManager* для получения и возвращения данных.

Всего реализовано три класса:

- *ExcelManager*: реализует ввод/вывод данных из файла **.xls*.
- *PugachevMethod*: реализует работу алгоритма для внешних интерфейсов.
- *UiMainWindowRealization* содержит реализацию действий, происходящих на форме графического интерфейса.

6. Тестирование программного продукта и анализ полученных результатов

В результате тестирования программного продукта были сделаны следующие выводы:

1. Наиболее точный результат достигается при обработке квадратных матриц.
2. Для получения актуального значения данного коэффициента требуется брать более семи значений для вычисления μ^* . В результате для вычисления вышеуказанного коэффициента был реализован ввод дополнительной матрицы, с использованием которой будет осуществляться расчет μ^* .

При тестировании на различных операционных системах проблем не выявлено, что связано с использованием такого инструмента реализации, как язык программирования

Python.

Тестирование проводилось на наборе статистических данных, полученных в ходе многофакторного неустойчивого технологического процесса. Таблица 1 и таблица 2 содержат матрицы R и S , компоненты которых представляют собой некоторые функции от значений процессов соответственно в исходной и упрощённой системах.

Таблица 1 - Фрагмент исходной R -матрицы

№	Par1	Par2	Par3	Par4	Par5	Par6	Par7	Par8	...	Par36	Y
1	9,80	139,19	115,1	1595	1012	54	44	698		1746,3	77,96
2	11,62	139,19	127,9	1578	891	55	45	780		1747,6	79,39
3	15,70	138,26	103,8	1641	1203	117	44	919		1760,8	74,32
4	14,58	141,00	113,7	1628	981	54	45	755		1628,0	87,56
5	14,30	145,52	130,8	1636	1154	75	49	950		1759,8	84,06
6	15,20	137,76	105,2	1652	1332	51	43	756		1759,2	81,87
		
1375	15,00	140,17	116,6	1624	1211	55	48	682		1758,5	79,64

Таблица 2 - Фрагмент исходной S -матрицы

№	Par1	Par2	Par3	Par4	Par5	Par6	Par7	Par8	...	Par36	Y
1	8,38	138,76	114,69	1465,62	1135,79	51,18	44,77	699,17		1723,74	72,06
2	8,61	140,50	114,35	1457,64	1138,06	52,28	42,57	629,97		1732,49	72,72
3	8,84	138,06	116,36	1470,50	1140,32	53,37	44,05	720,30		1777,45	73,37
4	9,07	138,23	114,79	1469,71	1142,59	54,46	45,45	798,00		1729,95	74,03
5	9,30	140,93	111,23	1497,01	1144,86	55,55	43,98	665,44		1688,25	74,69
6	9,53	138,10	115,99	1497,24	1147,13	56,64	44,49	703,94		1736,32	75,35
7	9,76	140,95	115,57	1509,93	1149,39	57,73	43,44	638,71		1757,53	76,00
		
78	10,45	142,53	115,53	1495,51	1138,06	50,09	45,07	691,79		1729,48	72,72

Средствами разработанного программного продукта рассчитаны искомый вектор вероятностных характеристик исходной системы λ и вектор вероятностных характеристик упрощённой системы μ . Статистические значения λ^* и μ^* векторов λ и μ , найденные по этим 78 экспериментам, будут равны (табл. 3).

Таблица 3 - Полученные оценки вероятностных характеристик

№	Par1	Par2	Par3	Par4	Par5	Par6	Par7	Par8	...	Par36	Y
μ^*	5,18	137,00	117,06	1645,54	1235,67	59,98	46,81	694,03	...	1752,5	78,03
λ^*	11,61	138,98	115,42	1476,48	1131,75	58,19	44,69	701,87		1740,76	77,90

Выводы

Существенным преимуществом метода является то, что даже если приближённая модель весьма неточна, точность статистического метода будет не хуже, чем при применении обычного метода Монте-Карло. Для повышения точности расчёта необходимо, чтобы существовала корреляция между оцениваемыми параметрами исходной и приближённой моделей.

В качестве основного недостатка метода в [3] отмечается существование систематической ошибки при статистической оценке параметров.

Практическая ценность разработанного программного продукта состоит в том, что он может быть использован для сокращения количества экспериментов в процессе получения адекватной модели.

Список литературы:

1. Беляков Ю.Н., Курмаев Ф.А., Баталов Б.В. Методы статистических расчетов микросхем на ЭВМ. М.: Радио и Связь, 1985. 232 с.
2. Бахвалов Н.С. Численные методы. М.: Наука, 1973. 632 с.
3. Васильев Д.В., Сабинин О.Ю. Ускоренное статистическое моделирование систем управления. Л.: Энергоатомиздат. Ленингр. отд-ние, 1987. 136 с.
4. Гехер К. Теория чувствительности и допусков электронных цепей. М.: Сов. Радио, 1973. 200 с.
5. Пугачёв В.Н. Комбинированные методы определения вероятностных характеристик. М.: Сов.радио, 1973. 256 с.
6. SCI-ARTICLE // Электрон. дан. Режим доступа URL: <http://sci-article.ru/stat.php?i=1492627101>.

References:

1. Belyakov Yu.N., Kurmaev F.A., Batalov B.V. (1985). Methods of statistical computing of microcircuits on a computer. Moscow: Radio and Communication. [in Russian language]
2. Bakhvalov N.S.(1973). Numerical methods. Moscow: Nauka. [in Russian language]
3. Vasiliev DV, Sabinin O.Yu. (1987) Accelerated statistical modeling of control systems. Leningrad: Energoatomizdat. [in Russian language]
4. Geher K. (1973). Theory of sensitivity and tolerances of electronic circuits. Moscow: Sov. Radio. [in Russian language]
5. Pugachev V.N. (1973). Combined methods for determining probabilistic characteristics. Moscow: Sov. Radio. [in Russian language]
6. SCI-ARTICLE. Available at: <http://sci-article.ru/stat.php?i=1492627101>. [in Russian language]

*Статья поступила в редколлегию 31.01.18.
Рецензент: к.т.н., доцент Брянского государственного технического университета
Подвесовский А.Г.
Статья принята к публикации 07.09.18*

Сведения об авторах:

Гарбузняк Елена Сергеевна

Ст. преподаватель
Филиал Приднестровского государственного университета им. Т.Г. Шевченко в г. Рыбнице
тел.: 00373-555-23912
E-mail: goldfenix@mail.ru

Сташкова Ольга Витальевна

Ст. преподаватель
Филиал Приднестровского государственного университета им. Т.Г. Шевченко в г. Рыбнице
тел.: 00373-555-23912
E-mail: stashkova.ola@mail.ru

Шестопад Оксана Викторовна

Ст. преподаватель
Филиал Приднестровского государственного университета им. Т.Г. Шевченко в г. Рыбнице
тел.: 00373-555-23912
E-mail: oksanashes@gmail.com

Козак Людмила Ярославовна

Кандидат технических наук, доцент
Филиал Приднестровского государственного университета им. Т.Г. Шевченко в г. Рыбнице
тел.: 00373-555-4-39-21
E-mail: Ludmilayaroslavovna@gmail.com

Information about authors:

Garbuznyak Elena

senior lecturer
Rybnitsa branch of Pridnestrovian Shevchenko State University
phone.: 00373-555-23912
E-mail: goldfenix@mail.ru

Stashkova Olga

senior lecturer
Rybnitsa branch of Pridnestrovian Shevchenko State University
phone.: 00373-555-23912
E-mail: stashkova.ola@mail.ru

Shestopal Oksana

senior lecturer
Rybnitsa branch of Pridnestrovian Shevchenko State University
phone.: 00373-555-23912
E-mail: oksanashes@gmail.com

Kozak Ludmila

Candidate of technical sciences, principal lecturer
Rybnitsa branch of Pridnestrovian Shevchenko State University
phone.: 00373-555-4-39-21
E-mail: Ludmilayaroslavovna@gmail.com