УДК 519.115 DOI: 10.14529/mmp200312

ВЕРОЯТНОСТНЫЕ МОДЕЛИ КОМБИНАТОРНЫХ СХЕМ

Н.Ю. Энатская, Высшая школа экономики, Московский институт электроники и математики им. А.Н. Тихонова, г. Москва, Российская Федерация

Предлагается перечислительный метод анализа комбинаторных схем в доасимптотической области изменения их параметров на основе построения их вероятностной математической модели, представляющей для каждой схемы итерационный случайный процесс последовательного бесповторного формирования всех ее исходов с определенной дисциплиной их нумерации поединичным добавлением определенных элементов схемы до данного в ней значения. В связи с важностью для проведения ряда исследований схемы бесповторности перечисления ее исходов, если она не лежит в ее природе, может достигаться путем введения в схему некоторых ограничений, не приводящих к изменению их множества, не меняющих их вероятности, и должны быть учтены. Конструкция процесса в соответствующих условиях каждой схемы наглядно изображается графом с заданными в нем вероятностями итерационных переходов, определяющих итоговое распределение на множестве ее исходов. На этой основе решаются задачи определения числа исходов схемы, установления взаимно-однозначного соответствия между номерами и видами ее исходов, называемого задачей нумерации в прямой и обратной постановках, нахождения вероятностного распределения всех ее итоговых исходов, что дает возможность их моделирования с найденным распределением разыгрывания номера исхода и последующим определением его смоделированного вида по результату решения прямой задачи нумерации. В случае отсутствия явной формулы для числа исходов схемы при определенных условиях по результатам их моделирования может быть получена его оценка с последующим уточнением по задаче нумерации. Исследования моделей комбинаторных схем на случайных процессах с введением вероятностных параметров расширяет возможности их использования. Результаты анализа схем могут иметь характер от численных методов и алгоритмов до аналитических в виде рекуррентных соотношений и явных формул.

Ключевые слова: перечислительный метод; метод графов; вероятностные модели; задача нумерации; быстрое моделирование.

Сокращения

Введем обозначения часто используемых терминов: перечислительный метод (ПМ); метод графов (МГ); задача нумерации (ЗН); быстрое моделирование (БМ).

Введение

Число исходов комбинаторной схемы дает лишь частичную информацию о них и в ряде ситуаций не исчерпывает интереса исследователя. Примерами могут служить такие области применения комбинаторики, как криптография при переборе ключей, криминалистика при проведении следственных действий, связанных с числовыми наборами, математические вычисления операций по наборам индексов с определенными ограничениями, составления всякого рода расписаний, задач управления, распределения ресурсов, социологических моделях и т.д., где необходимо получение всех возможных исходов соответствующих комбинаторных схем в явном виде. Для получения такой качественной информации об исходах схем и перевода ее в количественную предлагается ПМ, состоящий в построении ее вероятностной модели в виде итерационного случайного процесса, формирующего ее перечисляемые исходы с введением их вероятностного распределения в доасимптотической области изменения своих параметров. Он использует и дополняет широко представленные в литературе точные и асимптотические исследования комбинаторных схем (см., например, [1–3])и алгоритмический подход, который, как сказал Д. Кнут [4], приводит к более глубокому пониманию сути, т.е. процесса формирования исходов схемы. Инструментами ПМ являются МГи ЗН.

1. Метод графов

Для анализа исходов комбинаторной схемы прежде всего нужно определить его вид, который должен отражать ее специфику.

Следующим шагом реализации ПМ является организация создания качественной информации обо всех исходах схемы, состоящая в полном их перечислении. Она представляет собой исходы итерационного случайного процесса реализации комбинаторной схемы путем последовательного поединичного добавления элементов схемы до заданного значения в базовых схемах или этапов перечисления составляющих схему более простых ранее изученных схем в составных схемах. (В качестве добавляемых выбираются такие элементы или этапы, чтобы описанная процедура перечисления исходов приводила к реализации схемы.) Для наглядности этот процесс изображается графом с дугами итерационных переходов в нем и называется МГ, дающим ответ не только на *количественный* вопрос – *сколько* исходов схемы, но и на *качественный* – *какие* они? Для анализа схемы важна пучковая структура (конструкция) графа перечисления ее исходов, где пучком в графе на i-ой итерации назовем все ее исходы из одного исхода $(i-1)$-ой итерации. Размером пучка на i-ой итерации является число его образующих дуг. Пучковой структурой итерации будем называть перечень размеров ее пучков в порядке перечисления исходов. Пучковая структура графа перечисления исходов схемы – набор пучковых структур всех его итераций – используется при определении числа исходов схемы и при решении ЗН для изучения закономерностей соответствия номеров и видов исходов.

Для реализации МГ совершаются следующие действия *алгоритма*:

1) определяется поитерационно добавляемый элемент (или этап перечисления) схемы;

2) устанавливается дисциплина нумерации исходов схемы;

3) процедура полного перечисления исходов схемы наглядно представляется *графом*, в котором итерационные переходы изображаются *дугами*, на которых можно указывать вероятности переходов из состояния в состояние.

Тогда на последней итерации мы получаем перечисление всех нумерованных исходов схемы.

МГ является инструментом создания качественной информации об исходах схемы в ПМ и дает следующие *возможности* их исследования:

1) визуальное выявление закономерностей в перечне исходов;

2) непосредственный численный подсчет числа исходов;

3) процедура перечисления исходов *без повторов* дает основание для получения аналитической или алгоритмической формулы числа исходов схемы, причем для этого часто достаточно осмысление логики перечисления исходов схемы без их явного трудоемкого перечисления;

4) разные процедуры перечисления исходов *без повторов*, приводящие к разным аналитическим формулам числа исходов схемы, являются аргументацией их эквивалентности (в обход аналитической проверки);

5) непосредственное вычисление значений любых характеристик исходов схемы;

6) возможность получения исходов с учетом любых ограничений путем удалений несоответствующих исходов и траекторий в графе их перечисления для аналогичной схемы без ограничений;

7) получение вероятностного распределения исходов схемы по ее графу их перечисления путем вычисления вероятности каждого исхода, как суммы вероятностей его траекторий, определяемых вероятностями переходов по составляющим ее дугам;

8) определение структуры графа;

9) вычисление числа исходов составных схем по перечислению исходов составляющих ее схем (пояснение ниже).

Со свойством 9 связано введение новой операции по перечислению, когда число исходов основной схемы зависит от вида исхода вспомогательной схемы перечисления, указываемой под знаком операции. Например, число исходов схемы k-цикловой подстановок степени n складывается из чисел таких подстановок при каждом наборе размеров циклов – $\{d_i\}$, получающихся в схеме D деления n различимых элементов на k непустых частей, откуда для известного числа исходов основной схемы $|S(n,k)|$ (где $S(n,k)$ – число Стирлинга 1-ого рода) получаем представление

$$|S(n,k)| = \sum_{(D)} \prod_{i=1}^{k} (d_i - 1)!. \tag{1}$$

Вероятностное распределение исходов схемы и дополнительные условия. Построение случайного процесса перечисления исходов схемы состоит из выбора структуры его графа и определения вероятностей итерационных переходов в нем и порождает их итоговое вероятностое распределение. При зависимости вероятностей итерационных переходов из каждого исхода или только от своего предшествующего исхода или еще и от номера итерации этот случайный процесс представляет собой однородную или неоднородную цепь Маркова соответственно.

Если в графе перечисления исходов схемы траектории всех итоговых исходов единственны, то граф будет соответствовать бесповторному их перечислению, которое важно для перечислительного анализа комбинаторных схем и обеспечивается, когда в процедуре перебора исходов схемы задается неизменяемое различие в итоговых исходах схемы, для чего при его отсутствии вводятся дополнительные условия. Их введение, как и любых ограничений, означает удаление из начального графа недопустимых по дополнительным условиям исходов, т. е. траекторий, не обеспечивающих их выполнение, что приводит к изменению вероятностей оставшихся итерационных переходов в графе перечисления исходов схемы с их определенными вероятностями итоговых исходов. Расчет вероятностей итерационных переходов производится из системы уравнений, составленных путем приравнивания вероятностей всех траекторий графа, выраженных через вероятности итерационных переходов, их заданным вероятностям на множестве исходов схемы. Тогда число независимых уравнений в системе уравнений будет равно числу траекторий N_r минус 1, т.к. сумма вероятностей итоговых исходов схемы равна 1.

Число неизвестных в системе уравнений равно сумме чисел неизвестных по всем итерациям от 1 до r. На i-ой итерации число неизвестных равно сумме неизвестных по всем пучкам итерации. В каждом пучке любой итерации число неизвестных равно размеру пучка минус 1, т.к. сумма вероятностей переходов в каждом пучке равна 1. Тогда число неизвестных на i-ой итерации равно числу N_i исходов на i-ой итерации минус число пучков на ней, равное числу N_{i-1} исходов предыдущей $(i-1)$-ой итерации. Отсюда получаем, что число неизвестных на i-ой итерации есть $N_i - N_{i-1}$, а т.к. $N_0 = 1$, общее число неизвестных по всем r итерациям определяется из соотношения

$$\sum_{i=1}^{r}(N_i - N_{i-1}) = N_r - 1. \tag{2}$$

Общий прием решения системы уравнений, т.е. нахождения вероятностей итерационных переходов $\{p_{ij}\}$ (i, j – соответственно номера итерации и неизвестной вероятности на ней), следует из процедуры пошагового добавления элементов схемы и пучковой структуры графа перечисления ее исходов, что приводит к следующим соображениям:

1) неизвестные системы уравнений вычисляем последовательно по итерациям, начиная с последней, и по пучкам итераций, начиная с первого пучка каждой итерации;

2) неизвестные пучка промежуточной итерации находятся в условиях полученных ранее вероятностей всех итерационных переходов после нее (по п. 1)), поэтому они определяются так же, как и в первом пучке последней итерации с заменой известных вероятностей итоговых исходов процесса в нем на вычисляемые вероятности итоговых исходов пучка исследуемой итерации.

Тогда идея решения системы уравнений сводится к нахождению неизвестных вероятностей итерационных переходов первого пучка последней итерации. Рассмотрим эту задачу отдельно, упростив обозначения: пусть имеем пучок размера L с L неизвестными вероятностями переходов $x_1, \ldots, x_L$, или с $(L-1)$ неизвестным при $x_1 + \ldots + x_L = 1$, и пусть итоговые вероятности исходов для этого пучка есть $p_1, \ldots, p_L$, а P – вероятность траектории процесса до порождающего этот пучок состояния (исхода) предпоследней итерации. Тогда система уравнений для нахождения $x_1, \ldots, x_L$ имеет вид:

$$Px_1 = p_1, \ldots, Px_L = p_L, \sum_{j=1}^{L} x_j = 1, \text{ откуда получаем}$$

$$x_2/x_1 = p_2/p_1 = a_1, \ldots, x_L/x_1 = p_L/p_1 = a_{L-1}, \sum_{j=1}^{L} x_j = 1,$$

тогда из $x_1(1 + a_1 + \ldots + a_{L-1}) = 1$ получаем решение:

$$x_1 = 1/(1 + a_1 + \ldots + a_{L-1}), x_2 = a_1 x_1, \ldots, x_L = a_L x_1.$$

Далее с учетом п. 2 определяем все остальные неизвестные СУ.

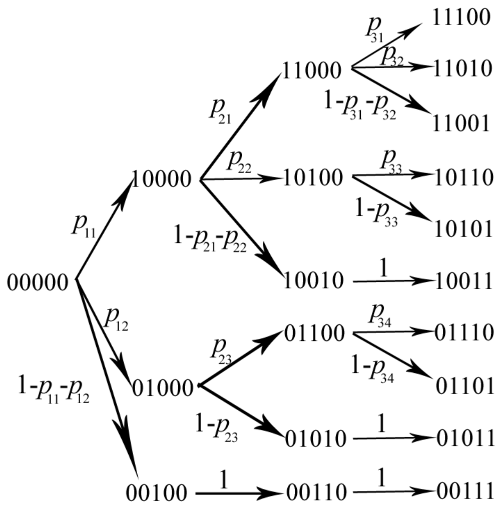
Для примера рассмотрим схему сочетаний из n элементов по r, $r < n$, как схему размещения r неразличимых частиц по n различимым ячейкам, вмещающим каждая по одной частице. Вид исхода будем задавать составами номеров непустых ячеек или n-размерной последовательностью нулей и единиц, соответствующих пустым и непустым ячейкам. Если размещать последовательно добавляемые по одной частицы по пустым ячейкам, то получим каждый исход с $r!$ повторами эа счет получения одних и тех же составов номеров непустых ячеек во всех $r!$ возможных порядках, которые несущественны при заданной неразличимости частиц. Тогда для процедуры бесповторного перечисления исходов схемы нужно исключить получение составов номеров пустых ячеек в разных порядках, т.е. ввести дополнительные условия упорядоченного, например, по возрастанию их перечисления путем размещения каждой добавленной частицы по ячейкам правее последней занятой. Это ограничение означает, что на i-ой итерации (размещения i-ой частицы) остаются только исходы с не менее, чем $(r-i)$ нулями подряд в конце исхода для того, чтобы до r-ой итерации хватило ячеек для требуемого размещения остальных $(r-i)$ частиц.

Приведем пример расчета вероятностей итерационных переходов из системы уравнений при равновероятном распределении итоговых исходов схемы.

Пример. Пусть $n = 5,\ r = 3$ (см. рисунок). Тогда

$$P(11100) = p_{11}p_{21}p_{31} = 0,1; P(11010) = p_{11}p_{21}p_{32} = 0,1;$$
$$P(11001) = p_{11}p_{21}(1 - p_{31} - p_{32}) = 0,1; P(10110) = p_{11}p_{22}p_{33} = 0,1;$$
$$P(10101) = p_{11}p_{22}(1 - p_{33}) = 0,1; P(10011) = p_{11}(1 - p_{21} - p_{22}) = 0,1;$$
$$P(01110) = p_{11}p_{21}p_{34} = 0,1; P(01101) = p_{12}p_{23}(1 - p_{34}) = 0,1;$$
$$P(01011) = p_{12}(1 - p_{23}) = 0,1; P(00111) = 1 - p_{11} - p_{21} = 0,1;$$

откуда получаем: $p_{11} = 3/5, p_{12} = 3/10, p_{21} = 1/2, p_{22} = 1/3, p_{23} = 2/3, p_{31} = 1/3, p_{32} = 1/3, p_{33} = 1/2, p_{34} = 1/2.$



Граф перечисления исходов схемы примера 1

Такой процесс соответствует неоднородной цепи Маркова, что следует из зависимости вероятности следующего исхода только от предыдущего и от номера итерации.

При отсутствии предположения о распределении вероятностей исходов схемы оно задается через постулаты о естественных вероятностях итерационных переходов с их пропорциональным (сохраняющим соотношения) пересчетом с учетом ДУ.

2. Задача нумерации для исходов схемы. Ее применение в анализе схем

Задача нумерации решается после перечисления исходов по МГ. Задача нумерации подразумевает прямую и обратную постановки: получения вида исхода по его номеру (прямая задача нумерации) и получения номера исхода по его виду (обратная задача нумерации).

Результат решения задачи нумерации может быть получено в двух видах: табличном (численном) при заданных параметрах схемы, соответствующим результату МГ, и аналитическом – в виде формул решения прямой и обратной задачи нумерации. В первом случае соответствие видов исходов схемы и их номеров требует большую память и получается только для конкретных заданных параметров схемы. Поэтому теоретическое установление этого соответствия при любых параметрах схемы будет одной из главных целей ее комбинаторного анализа.

Перечислим *возможности* использования результатов решения задачи нумерации в дальнейших комбинаторных исследованиях схемы по следующим направлениям:

1) для компактного хранения информации о видах всех исходов схемы;

2) результат решения прямой задачи нумерации дает возможность моделировать исход схемы путем разыгрывания его номера;

3) результат решения обратной задачи нумерации, т.е. формула для номера последнего, обычно очевидного по дисциплины нумерации, исхода дает формулу для числа исходов схемы;

4) результат решения прямой задачи нумерации дает возможность явного перечисления исходов без построения графа.

Пример решения задачи нумерации (другие примеры см. в [5]).

Схема перестановок [5], состоящая в установлении всех взаимных порядков n различимых элементов между собой. При перечислении исходов схемы методом графов каждый следующий добавляемый элемент ставится до, между и после всех, ранее установленных элементов.

Прямая задача нумерации решена теоремой

Теорема 1. *Дан номер $N = N_r$ перестановки R размера r. Тогда определяющие исход R значения числа M_i (порядковый номер места элемента i, $(i = \overline{1,r})$ среди элементов перестановки от 1 до i (слева направо) в данном исходе R) находятся по формуле*

$$M_i = 1 + (N_i - 1) \; mod \; i, \tag{3}$$

где N_i – номер перестановки длины i в процедуре перечисления исходов схемы, порождающей искомую перестановку длины r с данным номером $N = N_r$.

Обратная задача нумерации решена теоремой

Теорема 2. *Дана перестановка размера r или соответствующее ей исход R. Тогда его номер N определяется формулой*

$$N = \sum_{i=2}^{r-1} (M_i - 1)\frac{r!}{i!} + M_r, \tag{4}$$

где M_i – порядковый номер места элемента i, $(i = \overline{1,r})$ среди элементов перестановки от 1 до i (слева направо) в данном исходе R.

3. БМ исхода схемы

ПМ анализа исходов комбинаторных схем приводит при их известном вероятностном распределении к единому универсальному приему их моделирования без учета специфики каждой схемы, информация о которой уже содержится в нумерованном перечислении исходов. Этот прием следует из известного числа исходов схемы и из результата решения прямой задачи нумерации и состоит в разыгрывании случайного номера с известным распределением ее моделируемого исхода, по которому однозначно устанавливается его вид, являющийся смоделированным исходом.

Оценка числа исходов схемы по модели и его уточнение. В случае отсутствия точной формулы для числа исходов N_A изучаемой схемы A предлагается получение его приближенного значения $\tilde{N}$ *методом пропорций* по результатам моделирования более общей (схемы, содержащей среди своих исходов все исходы нашей схемы A) изученной схемы с известным числом исходов N с определением надежности с заданной точностью ε на примере числа исходов схемы перестановок с заданным числом инверсий – суммарным числом для всех элементов перестановки количеств элементов в ее исходе, меньших и стоящих правее каждого данного элемента.

Для проведения приближенного оценивания числа исходов схемы смоделируем N_A исходов схемы перестановок размера n с $n!$ равновероятными исходами. Для каждого исхода вычисляем число инверсий и определяем среди смоделированных количество M исходов с t инверсиями. Тогда искомое число исходов нашей схемы перестановок с t инверсиями приближенно определяется методом пропорций по формуле

$$N \approx \frac{Mn!}{N_A} = \tilde{N}. \tag{5}$$

Исследуем качество полученной оценки $\tilde{N}$ для N исходов нашей схемы, где M/N_A – наблюденная частота успеха опыта – появления ее исхода среди N_A исходов схемы перестановок. Число M можно представить в виде $M = X_1 + X_2 + \ldots + X_{N_A}$, где при $i = \overline{1, N_A}$ $\{X_i\}$ – случайные величины, имеющие распределение Бернулли с вероятностью успеха $p = N/n!$. Тогда по уточненной по неравенству Чебышева теореме Бернулли выполняются соотношения

$$\begin{aligned} \gamma = \quad & P\left(\left|\tfrac{M}{N_A} - \tfrac{N}{n!}\right| < \varepsilon^*\right) = P\left(\left|\tfrac{Mn!}{N_A} - N\right| < \varepsilon^* n! = \varepsilon\right) = \\ = \quad & P(|\tilde{N} - N| < \varepsilon) \geq 1 - \tfrac{p(1-p)}{N_A(\varepsilon^*)^2} \geq 1 - \tfrac{(n!)^2}{4N_A\varepsilon^2}, \end{aligned} \tag{6}$$

где для оценки $\tilde{N}$ числа N исходов нашей схемы ε – ее точность, а $\gamma \geq 1 - \frac{(n!)^2}{4N_A\varepsilon^2}$ – оценка ее надежности γ с этой точностью. Для получения нетривиальной оценки для γ потребуем, чтобы $N_A > (n!)^2/4\varepsilon^2$.

Более точную оценку надежности γ с заданной точностью ε оценки $\tilde{N}$ числа N исходов нашей схемы можно получить по следствию из теоремы Муавра – Лапласа из соотношений

$$\begin{aligned} \gamma \; & = P\left(\left|\tfrac{M}{N_A} - \tfrac{N}{n!}\right| \leq \varepsilon^*\right) = P\left(\left|\tfrac{Mn!}{N_A} - N\right| \leq \varepsilon^* n! = \varepsilon\right) = \\ & = P(|\tilde{N} - N| \leq \varepsilon) \approx 2\Phi\left(\varepsilon^*\sqrt{N_A/p(1-p)}\right) \geq 2\Phi(2\varepsilon\sqrt{N_A}/n!), \end{aligned} \tag{7}$$

где $\Phi(x) = (1/\sqrt{2\pi}) \int_0^x \exp^{-x^2/2} dx$ – табличная функция Лапласа.

Уточнение значения числа N производится по п. 4 *возможностей* задачи нумерации в п. 2.

Литература

1. Риордан, Д. Введение в комбинаторный анализ / Д. Риордан. – М.: Иностранная литература, 1963.

2. Сачков, В.Н. Введение в комбинаторные методы дискретной математики / В.Н. Сачков. – М.: МЦНМО, 1982.

3. Колчин, В.Ф. Случайные размещения / В.Ф. Колчин, Б.А. Севастьянов, В.П. Чистяков. – М.: Наука, 1976.

4. Кнут, Д. Искусство программирования на ЭВМ / Д. Кнут. – М.: Мир, 1976.

5. Энатская, Н.Ю. Анализ комбинаторных схем в доасимптотической области изменения параметров / Н.Ю. Энатская // Труды Карельского научного центра РАН. – 2018. – № 7. – С. 117–133.

Наталия Юрьевна Энатская, кандидат физико-математических наук, доцент, Высшая школа экономики, Московский институт электроники и математики им. А.Н. Тихонова (г. Москва, Российская Федерация), nat1943@mail.ru.

Поступила в редакцию 17 марта 2020 г.

MSC 60F15 **DOI: 10.14529/mmp200312**

PROBABILISTIC MODELS OF COMBINATORIAL SCHEMES

N.Yu. Enatskaya, Higher School of Economics, Moscow Institute of Electronics and Mathematics, Moscow, Russian Federation, nat1943@mail.ru

An enumerative method is proposed for the analysis of combinatorial schemes in the pre-asymptotic region of variation of their parameters based on the construction of their probabilistic mathematical model, which represents for each scheme an iterative random process of sequential non-repeated formation of all its outcomes with a certain discipline of their numbering by unitary addition of certain elements of the scheme to a given value in it. Due to the importance for a number of studies of the scheme of recurrence of listing its outcomes, if it does not lie in its nature, it can be achieved by introducing into the scheme some restrictions that do not lead to a change in their set, do not change their probability and should be taken into account. The design of the process under the appropriate conditions of each scheme is graphically depicted by a graph with the probabilities of iterative transitions specified in it, which determine the final distribution on the set of its outcomes. On this basis, the problems of determining the number of outcomes of a scheme, establishing a one-to-one correspondence between numbers and types of its outcomes, called the numbering problem in direct and reverse statements, and finding the probability distribution of all its final outcomes are solved, which makes it possible to model them with the found distribution of playing out the outcome number and the subsequent determination of its modeled form by the result of solving the direct numbering problem. In the absence of an explicit formula for the number of outcomes of a scheme under certain conditions, an estimate of it can be obtained from the results of their modeling, followed by refinement of the numbering problem. The study of models of combinatorial schemes on random processes with the introduction of probabilistic parameters expands the possibilities of their use. The results of the analysis of schemes can be of a nature from numerical methods and algorithms to analytical in the form of recurrence relations and explicit formulas.

Keywords: enumeration method; graph method; probabilistic models; numbering problem; rapid modelling.

References

1. Riordan D. *Introduction to Combinatorial Analysis*. N.Y., John Wiley, 1958.
2. Sachkov V.N. *Vvedenie v kombinatornye metody diskretnoy matematiki* [Introduction to Combinatorial Methods of Discrete Mathematics]. Moscow, Moscow Center for Continuous Mathematical Education, 1982. (in Russian)
3. Kolchin V.F., Sevastyanov B.A., Chistyakov V.P. *Sluchaynye razmeshcheniya* [Random Postings]. Moscow, Nauka, 1976. (in Russian)
4. Knuth D. *Iskusstvo programmirovaniya na EVM* [The Art of Computer Programming]. Moscow, Mir, 1976. (in Russian)
5. Enatskaya N.Yu. Analysis Combinatorial Schemes in the Pre-Asymptotic Region of Parameter Change. *Proceedings of the Kola Science Center of the Russian Academy of Science*, 2018, no. 7, p. 117–133. (in Russian)

Received March 17, 2020