

Санкт-Петербургский государственный университет
Прикладная математика и информатика
Вычислительная стохастика и статистические модели

Мейер Никита Владимирович

МОДЕЛИРОВАНИЕ ПРОЦЕССА ПЕРКОЛЯЦИИ

Выпускная квалификационная работа

Научный руководитель:
д. ф.-м. н., профессор С. М. Ермаков

Рецензент:
к. ф.-м. н. А. В. Дмитриев

Санкт-Петербург

2018

Saint Petersburg State University
Applied Mathematics and Computer Science
Computational Stochastics and Statistical Models

Meyer Nikita Vladimirovich

MODELLING OF THE PERCOLATION PROCESS

Graduation Project

Scientific Supervisor:
Professor S. M. Ermakov

Reviewer:
PhD A. V. Dmitriev

Saint Petersburg
2018

Оглавление

Введение	4
Глава 1. Постановка задачи	6
Глава 2. Моделирование методом Монте-Карло	7
2.1. Алгоритм Хошена-Копельмана	7
2.2. Несмещенная оценка для $E(K)$	8
2.3. Метод существенной выборки	8
Глава 3. Усовершенствованный алгоритм Хошена-Копельмана	11
3.1. Применение метода существенной выборки к алгоритму Хошена-Копельмана	11
3.2. Метод поиска оптимальной замещающей вероятности	12
Глава 4. Моделирование с помощью алгоритма, расширяющего класс мер	19
4.1. Алгоритм, расширяющий класс мер	19
4.2. Распределения, на которых достигнуто уменьшение дисперсии	20
4.2.1. Дискретный аналог нормального распределения	21
4.2.2. Усеченное распределение Пуассона	22
Заключение	25
Список литературы	26
Приложение А. Приложение для моделирования	27

можно смоделировать на перколяционной решетке. В начальный момент времени все узлы не заняты (непроводящие). Затем некоторые из узлов становятся проводящими, соответственно образуются проводящие кластеры. Образование перколяционного кластера соответствует электрическому «пробою», то есть фазовому переходу вещества.

В работах по теории перколяции в основном поднимаются вопросы, связанные с исследованием перколяционного кластера (кластера, соединяющего противоположные концы решетки). В этой работе исследуется только среднее число кластеров на решетке. Эта задача возникает в таких приложениях теории перколяции, как, к примеру, выведение на орбиту большого количества наноспутников [2] или обеспечение безопасности крупных компьютерных сетей [3].

Среднее число кластеров может быть получено многократным моделированием перколяционной решетки с помощью тривиального алгоритма, но из-за большого вычислительного объема встает вопрос о создании способа моделирования несмещенных оценок с меньшей дисперсией. Этому вопросу и посвящена данная работа.

Работа организована следующим образом: в Главе 1 определена формальная постановка задачи; в Главе 2 приведен алгоритм для моделирования среднего числа кластеров; в Главе 3 формулируется усовершенствованный алгоритм для моделирования несмещенных оценок для среднего числа кластеров с меньшей дисперсией; в Главе 4 алгоритм для моделирования несмещенных оценок для среднего числа кластеров с меньшей дисперсией ищется на более широком классе мер.

Глава 1

Постановка задачи

Формализуем описанную во введении задачу. Для простоты рассматривается квадратная решетка, имеющая размерность n (т.е. состоящая из n^2 узлов). На этой решетке проводится случайный эксперимент. Каждый узел заполняется независимо с вероятностью p . Результат эксперимента можно описать в виде множества $\{c_1, \dots, c_{n^2}\}$, где каждое c_j принимает одно из двух значений: либо 0, либо 1 (незаполненный и заполненный узел соответственно). Определим вероятностное пространство данного эксперимента:

- $\Omega = \{\omega : \omega = (c_1, \dots, c_{n^2}), c_j \in \{0, 1\}\}$;
- $|\Omega| = 2^{n^2}$ — количество различных элементарных исходов;
- $\mathfrak{A} = \{A | A \subseteq \Omega\}$;
- $\mathbb{P}(\{\omega_i\}) = p^m(1-p)^{n^2-m}$, где m — количество c_i , равных 1.

В условиях данной модели нас интересует количество кластеров на решетке. Для начала необходимо формально определить понятие кластера.

Определение 1. *Два заполненных узла принадлежат одному кластеру, если их можно соединить непрерывной цепочкой соседних заполненных узлов.*

Введем случайную величину $K(\omega) : \Omega \rightarrow \mathbb{N}$ как число кластеров на решетке. Запишем ее математическое ожидание в виде:

$$\mathbb{E}(K) = \sum_{i=1}^{2^{n^2}} K(\omega_i) \mathbb{P}(\omega_i),$$

где ω_i — элементарный исход; $K(\omega_i)$ — количество кластеров, соответствующее элементарному исходу; $\mathbb{P}(\omega_i)$ — вероятность элементарного исхода. K — дискретная случайная величина, распределение которой можно написать в виде $\mathbb{P}(K = X_j) = \sum_{i:K(\omega_i)=X_j} \mathbb{P}(\omega_i)$.

Получить точное значение $\mathbb{E}(K)$ затруднительно, поэтому для решения поставленной задачи используется моделирование методом Монте-Карло.

Глава 2

Моделирование методом Монте-Карло

Для моделирования методом Монте-Карло необходимо иметь алгоритм для моделирования случайной величины, распределенной как K .

2.1. Алгоритм Хошена-Копельмана

Этот алгоритм предложен в 1976 году Дж. Хошеном (J. Hoshen) и Р. Копельманом (R. Korelman) [4] и является классическим алгоритмом теории перколяции. Ниже он приведен с небольшими изменениями, чтобы соответствовать задаче вычисления среднего числа кластеров на решетке:

- **Входные данные:** размерность решетки n ; вероятность заполненного узла p .
- **Тело алгоритма:**
 1. Создаются два массива размера n^2 : пустой массив кластерных меток и массив реализаций равномерно распределенных на отрезке $(0, 1]$ псевдослучайных величин α_i .
 2. Устанавливается соответствие между узлами решетки и элементами массива кластерных меток.
 3. Поочередно сравнивая числа α_i с вероятностью заполненного узла p , будем присваивать i -му элементу массива кластерных меток значение 0 (не заполнен) или a (заполнен) соответственно до тех пор, пока не просмотрим все элементы массива. Число a определяется через значения кластерных меток в соседних узлах (одинаковые кластерные метки получают узлы, принадлежащие одному кластеру).
- **Результат работы алгоритма:** заполненный массив кластерных меток; количество различных кластерных меток в нем.

Поскольку массив кластерных меток находится в однозначном соответствии с узлами решетки, очевидно, что алгоритм позволяет получать реализацию случайной величины, распределенной как K .

2.2. Несмещенная оценка для $\mathbb{E}(K)$

Пусть K_1, \dots, K_N — независимые случайные величины, распределенные как K . Введем статистическую оценку

$$\hat{K}(K_1, \dots, K_N) = \frac{K_1 + \dots + K_N}{N}.$$

Посчитав ее математическое ожидание, получим:

$$\mathbb{E}(\hat{K}) = \frac{\mathbb{E}(K_1) + \dots + \mathbb{E}(K_N)}{N} = \mathbb{E}(K).$$

Если обозначить за σ^2 дисперсию случайной величины K , то по центральной предельной теореме

$$\frac{\sqrt{N}}{\sigma} \left(\hat{K} - \mathbb{E}(K) \right) \rightarrow N(0, 1),$$

откуда следует, что при $N \rightarrow \infty$ и $\forall x > 0$

$$P \left(|\hat{K} - \mathbb{E}(K)| < x \frac{\sigma}{\sqrt{N}} \right) \rightarrow 2\Phi(x) - 1,$$

где $\Phi(x)$ — функция распределения стандартного нормального распределения.

Данное соотношение позволяет построить асимптотический доверительный интервал уровня $1 - \gamma$ для $\mathbb{E}(K)$ при больших N :

$$P \left(\hat{K} - x_\gamma \frac{\sigma}{\sqrt{N}} < \mathbb{E}(K) < \hat{K} + x_\gamma \frac{\sigma}{\sqrt{N}} \right) \approx 1 - \gamma,$$

где x_γ — решение уравнения $2\Phi(x) - 1 = 1 - \gamma$. Отсюда видно, что построение случайной величины K' с $\mathbb{E}(K') = \mathbb{E}(K)$ и меньшей дисперсией позволяет уменьшить объем выборки N пропорционально уменьшению дисперсии при сохранении заданного уровня точности.

2.3. Метод существенной выборки

Можно рассмотреть альтернативное вероятностное пространство, в котором множество элементарных исходов и сигма-алгебра совпадают с определенным изначальной задачей:

$$\Omega = \{\omega : \omega = (c_1, \dots, c_{n^2}), c_j \in \{0, 1\}\},$$

$$\mathfrak{A} = \{A | A \subseteq \Omega\},$$

но вероятностная мера заменяется с $P(\{\omega_i\}) = p^m(1-p)^{n^2-m}$ на некую другую меру $Q : Q(\{\omega_i\}) > 0$.

Определение 2. Весом элементарного исхода ω_i при замене меры с \mathbb{P} на \mathbb{Q} называется W_i :

$$W_i = \frac{\mathbb{P}(\omega_i)}{\mathbb{Q}(\omega_i)}.$$

Введем на альтернативном вероятностном пространстве дискретную случайную величину $K'(\omega_i) : \Omega \rightarrow \mathbb{N}$, определяемую как

$$K'(\omega_i) = K(\omega_i)W_i.$$

Покажем, что $\mathbb{E}(K') = \mathbb{E}(K)$:

$$\mathbb{E}(K') = \sum_{i=1}^{2^{n^2}} \mathbb{Q}(\omega_i) K(\omega_i)W_i = \sum_{i=1}^{2^{n^2}} \mathbb{Q}(\omega_i) K(\omega_i) \frac{\mathbb{P}(\omega_i)}{\mathbb{Q}(\omega_i)} = \sum_{i=1}^{2^{n^2}} \mathbb{P}(\omega_i) K(\omega_i) = \mathbb{E}(K).$$

Введем статистическую оценку

$$\hat{K}'(K'_1, \dots, K'_N) = \frac{1}{N} \sum_{i=1}^N K'_i,$$

где K'_1, \dots, K'_N — независимые случайные величины, распределенные как K' .

Легко видеть, что

$$\mathbb{E}(\hat{K}') = \frac{\mathbb{E}(K'_1) + \dots + \mathbb{E}(K'_N)}{N} = \mathbb{E}(K') = \mathbb{E}(K).$$

В общем случае, дисперсия K' может быть как меньше, так и больше, чем дисперсия K .

Вообще говоря, возможно построить такую меру \mathbb{Q} , что дисперсия K' будет равна 0. Для примера рассмотрим перколяционную решетку размерности 2 с вероятностью заполненного узла p . Построим меру $\mathbb{Q} : \text{Var}(K') = 0$. Для этого необходимо, чтобы $K'(\omega_i) = K(\omega_i) \frac{\mathbb{P}(\omega_i)}{\mathbb{Q}(\omega_i)} = \mathbb{E}(K)$ для всех ω_i . Найдем $\mathbb{E}(K)$:

- $\mathbb{P}(K = 0) = (1 - p)^4$;
- $\mathbb{P}(K = 1) = p^4 + 4p^3(1 - p) + 4p^2(1 - p)^2 + 4p(1 - p)^3$;
- $\mathbb{P}(K = 2) = 2p^2(1 - p)^2$;
- $\mathbb{E}(K) = \sum_{i=1}^{16} K(\omega_i) \mathbb{P}(\omega_i) = p^4 + 4p^3(1 - p) + 8p^2(1 - p)^2 + 4p(1 - p)^3$.

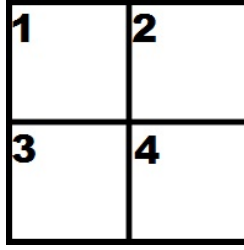


Рис. 2.1. Порядок обхода узлов.

Элементарные исходы будем записывать в виде $\omega_{i_1 i_2 i_3 i_4}$, где $i_j \in \{0, 1\}$. Порядок обхода узлов изображен на рисунке 2.1.

Сконструируем меру \mathbb{Q} следующим образом:

$$\mathbb{Q}(\omega_i) = \frac{K(\omega_i)\mathbb{P}(\omega_i)}{\mathbb{E}(K)}.$$

- $\mathbb{Q}(\omega_{1111}) = \frac{p^4}{p^4 + 4p^3(1-p) + 8p^2(1-p)^2 + 4p(1-p)^3};$
- $\mathbb{Q}(\omega_{1110}) = \mathbb{Q}(\omega_{1101}) = \mathbb{Q}(\omega_{1011}) = \mathbb{Q}(\omega_{0111}) = \frac{p^3(1-p)}{p^4 + 4p^3(1-p) + 8p^2(1-p)^2 + 4p(1-p)^3};$
- $\mathbb{Q}(\omega_{1100}) = \mathbb{Q}(\omega_{1010}) = \mathbb{Q}(\omega_{0101}) = \mathbb{Q}(\omega_{0011}) = \frac{p^2(1-p)^2}{p^4 + 4p^3(1-p) + 8p^2(1-p)^2 + 4p(1-p)^3};$
- $\mathbb{Q}(\omega_{1001}) = \mathbb{Q}(\omega_{0110}) = \frac{2p^2(1-p)^2}{p^4 + 4p^3(1-p) + 8p^2(1-p)^2 + 4p(1-p)^3};$
- $\mathbb{Q}(\omega_{1000}) = \mathbb{Q}(\omega_{0100}) = \mathbb{Q}(\omega_{0010}) = \mathbb{Q}(\omega_{0001}) = \frac{p(1-p)^3}{p^4 + 4p^3(1-p) + 8p^2(1-p)^2 + 4p(1-p)^3};$
- $\mathbb{Q}(\omega_{0000}) = 0.$

Полученная мера определяет случайную величину K' с нулевой дисперсией, но для ее построения необходимо знать значение $\mathbb{E}(K)$, то есть решение исходной задачи.

Глава 3

Усовершенствованный алгоритм Хошена-Копельмана

3.1. Применение метода существенной выборки к алгоритму Хошена-Копельмана

Приведенный в разделе 2.1 алгоритм Хошена-Копельмана позволяет задать вероятностную меру через вероятность заполненного узла q , то есть в виде $\mathbb{Q}(\{\omega_i\}) = q^m(1 - q)^{n^2 - m}$, где m — число заполненных узлов в элементарном исходе. В этой главе найдем вероятностную меру такого вида, доставляющую минимум дисперсии K' .

Замена вероятностной меры на меру вида $\mathbb{Q}(\{\omega_i\}) = q^m(1 - q)^{n^2 - m}$ эквивалентна замене вероятности заполненного узла с p на q . Вероятность q назовем замещающей вероятностью. Вероятности элементарных исходов в альтернативном вероятностном пространстве равны $q^m(1 - q)^{n^2 - m}$. Весом элементарного исхода будет

$$W_i = \frac{\mathbb{P}(\omega_i)}{\mathbb{Q}(\omega_i)} = \frac{p^m(1 - p)^{n^2 - m}}{q^m(1 - q)^{n^2 - m}}.$$

Как было показано в разделе 2.3, дискретная случайная величина на новом вероятностном пространстве $K'(\omega_i) = K(\omega_i)W_i$ будет обладать свойством $\mathbb{E}(K') = \mathbb{E}(K)$.

На практике это значит, что для моделирования решения исходной задачи с вероятностью заполненного узла p можно использовать алгоритм Хошена-Копельмана, моделирующий решетку с вероятностью заполненного узла q . Приведем схему усовершенствованного алгоритма Хошена-Копельмана:

- **Входные данные:** вероятность заполненного узла в исходной задаче p ; замещающая вероятность q ; размерность решетки n .
- **Тело алгоритма:**
 1. Применяется алгоритм Хошена-Копельмана, входными данными для которого будет q и n . Результат его работы — число кластерных меток d .
 2. Рассчитывается вес элементарного исхода $W = \frac{p^m(1-p)^{n^2-m}}{q^m(1-q)^{n^2-m}}$. Количество заполненных узлов m определяется как количество ненулевых элементов массива кластерных меток.

3. Количество кластерных меток d умножается на вес исхода:

$$b = dW.$$

- **Результат работы алгоритма:** величина b ; массив кластерных меток.

Алгоритм позволяет получать реализацию случайной величины, распределенной как K' .

3.2. Метод поиска оптимальной замещающей вероятности

Алгоритм, сформулированный в предыдущем разделе, был реализован в виде компьютерной программы на языке C#. Моделируя реализации K'_1, \dots, K'_N (независимые случайные величины, распределенные как K'), программа получает реализацию несмещенной оценки \hat{K}' . Для оценки теоретической дисперсии $\text{Var}(K')$ используется несмещенная выборочная дисперсия $S^2 = \frac{1}{N-1} \sum_{i=1}^N (K'_i - \hat{K}')^2$.

Теперь перед нами стоит задача поиска замещающей вероятности q_0 , использование которой для задания вероятностной меры в методе существенной выборки позволяет получить случайную величину K' , имеющую минимальную дисперсию среди случайных величин, полученных с помощью задания вероятностной меры в виде $\mathbb{Q}(\{\omega_i\}) = q^m(1-q)^{n^2-m}$, где m — число заполненных узлов в элементарном исходе. Такую замещающую вероятность q_0 назовем оптимальной замещающей вероятностью. Поиск оптимальной замещающей вероятности будет производиться приближенно, с помощью моделирования.

Для поиска оптимальной замещающей вероятности q_0 был реализован метод стохастической аппроксимации, который гарантирует сходимость к истинному значению, но на практике для получения значения q_0 с нужной точностью требовалось большое число итераций, поэтому был сформулирован другой метод поиска. Его будет удобно продемонстрировать на конкретном примере.

Рассмотрим моделирование решетки размерности $n = 20$ с вероятностью заполнения узла $p = 0.1$ при объеме выборки $N = 100000$. Перебирая различные замещающие вероятности q на интервале $[0.0985, 0.1051]$ с шагом 0.0003 , найдем значение, для которого достигается минимум несмещенной выборочной дисперсии S^2 . Этим значением будет точка $q = 0.1018$.

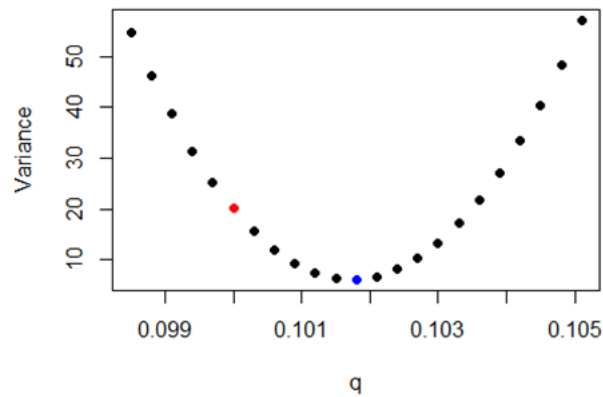


Рис. 3.1. Зависимость несмещенной выборочной дисперсии S^2 от замещающей вероятности q . Красным выделена точка, соответствующая $q = p = 0.1$ (т.е. моделированию без использования метода существенной выборки). Синим выделена точка, в которой достигается минимальная дисперсия, соответствующая $q = 0.1018$.

Если по некоторым трем точкам графика на рисунке 3.1 построить параболу, то можно видеть, что полученные результаты приблизительно описываются квадратичной функцией.

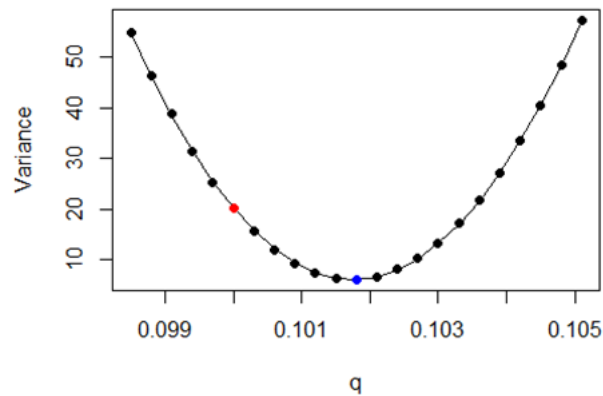


Рис. 3.2. На интервале $[0.0985, 0.1051]$ зависимость можно приближенно описать параболой $y = 4570555.04x^2 - 930172.80x + 47331.95$

Из графика на рисунке 3.3 видно, что для более широкого интервала $[0.0895, 0.1141]$ такого предположения сделать уже нельзя. Более того, наблюдается явная асимметрия точек относительно вертикальной оси, проходящей через вершину параболы.

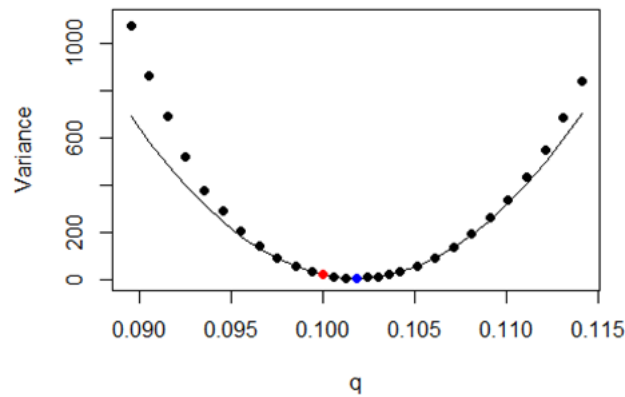


Рис. 3.3. На интервале $[0.0895, 0.1141]$ зависимость уже не описывается параболой.

В ходе исследования на решетках различной размерности n с различной вероятностью заполненного узла p были получены аналогичные результаты о характере зависимости.

Исходя из вышеизложенного, выскажем следующее предположение: на некотором интервале, содержащем вероятность заполненного узла p и оптимальную замещающую вероятность q_0 , зависимость несмещенной выборочной дисперсии S^2 от замещающей вероятности q приближенно описывается квадратичной функцией.

На основе этого предположения строится алгоритм поиска оптимальной заменяющей вероятности q_0 . Приведем схему этого алгоритма.

- **Входные данные:** размерность решетки n ; вероятность заполненного узла в исходной задаче p ; замещающая вероятность q_1 ; замещающая вероятность q_2 ; замещающая вероятность q_3 ; объем выборки N .

- **Тело алгоритма:**

1. По каждой из замещающих вероятностей q_1, q_2, q_3 с помощью усовершенствованного алгоритма Хошена-Копельмана для вероятности заполненного узла p и размерности решетки n реализуется выборка объема N . Для каждой выборки вычисляется несмещенная выборочная дисперсия S_1^2, S_2^2, S_3^2 .
2. Строится квадратичная функция $f(x) = ax^2 + bx + c$, решениями которой являются $\{(q_1, S_1^2); (q_2, S_2^2); (q_3, S_3^2)\}$. Вычисляются коэффициенты этой функции.

$$\begin{aligned}
& - \text{Коэффициент } a = \frac{q_1(S_2^2 - S_3^2) + q_2(S_3^2 - S_1^2) + q_3(S_1^2 - S_2^2)}{(q_2 - q_3)(q_3 - q_1)(q_1 - q_2)}. \\
& - \text{Коэффициент } b = \frac{q_1^2(S_3^2 - S_2^2) + q_2^2(S_1^2 - S_3^2) + q_3^2(S_2^2 - S_1^2)}{(q_2 - q_3)(q_3 - q_1)(q_1 - q_2)}. \\
& - \text{Коэффициент } c = \frac{S_1^2 q_2 q_3 (q_3 - q_2) + S_2^2 q_1 q_3 (q_1 - q_3) + S_3^2 q_1 q_2 (q_2 - q_1)}{(q_2 - q_3)(q_3 - q_1)(q_1 - q_2)}.
\end{aligned}$$

3. Координата вершины построенной параболы по оси абсцисс выражается как $q_{opt} = -\frac{b}{2a}$.
4. Координата вершины построенной параболы по оси ординат выражается как $f(q_{opt}) = c + -\frac{b^2}{4a}$.
5. По замещающей вероятности q_{opt} с помощью усовершенствованного алгоритма Хошена-Копельмана для вероятности заполненного узла p и размерности решетки n реализуется выборка объема N . Для этой выборки вычисляется несмещенная выборочная дисперсия S_0^2 .

- **Результат работы алгоритма:** координаты $(q_{opt}, f(q_{opt}))$, значение несмещенной выборочной дисперсии S_0^2 .

Полученное значение q_{opt} принимается за приближенное значение оптимальной замещающей вероятности q_0 . Точность этого метода можно оценить, сравнив $f(q_{opt})$ со значением несмещенной выборочной дисперсии S_0^2 . В рамках исследования значительным считалось отклонение $f(q_{opt})$ от S_0^2 более чем на 10%. Значительное отклонение может происходить по причине того, что по крайней мере одна из замещающих вероятностей, подающихся на вход алгоритму, не принадлежит интервалу, на котором зависимость описывается параболой. Поскольку по сделанному выше предположению вероятность заполненного узла p принадлежит такому интервалу, в качестве одной из замещающих вероятностей выгодно брать значение p . Подача на вход алгоритму двух замещающих вероятностей, лежащих слишком близко друг от друга, также может привести к значительному отклонению по причине того, что несмещенная выборочная дисперсия является оценкой, имеющей некоторую погрешность. В случае получения значительного отклонения полученное значение q_{opt} считалось неточным, и алгоритм применялся заново, а в качестве значений замещающей вероятности брались полученное неточное значение q_{opt} , вероятность заполненного узла p и значение, соответствующее $2q_{opt} - p$.

Вышеизложенный метод назовем методом приближения параболой. Теоретическое обоснование этого метода возможно после функционального описания зависимости дисперсии случайной величины K' от значения замещающей вероятности q . Из получения

такой функциональной зависимости будет следовать аналитическое выражение для $\mathbb{E}(K')$, то есть решение исходной задачи, которое мы пытаемся получить с помощью моделирования.

Несмотря на отсутствие формального обоснования, метод позволяет с большой точностью высчитывать приближенное значение q_0 . Для этого чаще всего достаточно получить значение выборочной дисперсии в 4-8 точках. Для получения сравнимой точности с помощью метода стохастической аппроксимации, как правило, требовалось более 100.

Продemonстрируем результаты, полученные с помощью изложенного метода, на решетке размерности $n = 20$ при объеме выборки $N = 50000$ для различных вероятностей заполненного узла p . На графике, приведенном на рисунке 3.4 произведено сравнение значений несмещенной выборочной дисперсии для двух методов моделирования реализации случайной величины, имеющей математическое ожидание $\mathbb{E}(K)$: обычного алгоритма Хошена-Копельмана и усовершенствованного алгоритма Хошена-Копельмана со значением замещающей вероятности, полученным с помощью метода приближения параболой.

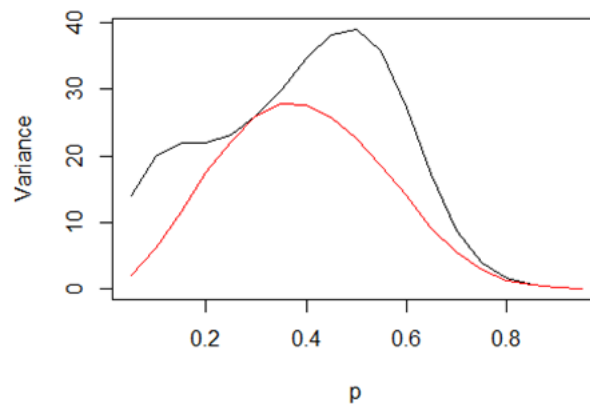


Рис. 3.4. Черным показаны значения несмещенной выборочной дисперсии S^2 , полученные с помощью обычного метода Хошена-Копельмана. Красным показаны значения S_0^2 , полученные с помощью усовершенствованного метода.

Из графика видно, что существенное снижение дисперсии может быть достигнуто не для всех значений p . Относительное значение несмещенной выборочной дисперсии S_0^2/S^2 для усовершенствованного алгоритма Хошена-Копельмана по отношению к

дисперсии, полученной из обычного алгоритма показано с помощью графика на рисунке 3.5.

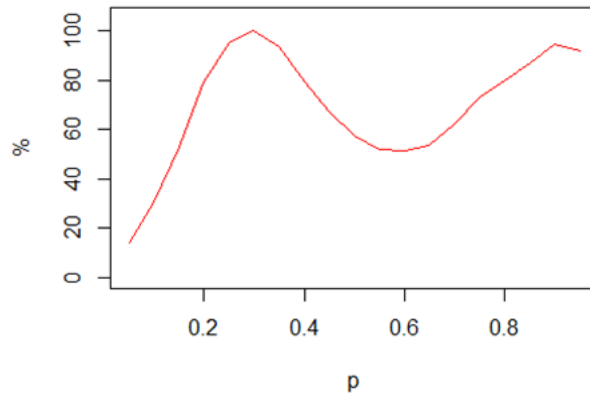


Рис. 3.5. Относительное значение S_0^2/S^2 .

График на рисунке 3.6 демонстрирует зависимость разности оптимального значения замещающей вероятности и вероятности заполненного узла $h = q_0 - p$ от вероятности заполненного узла p .

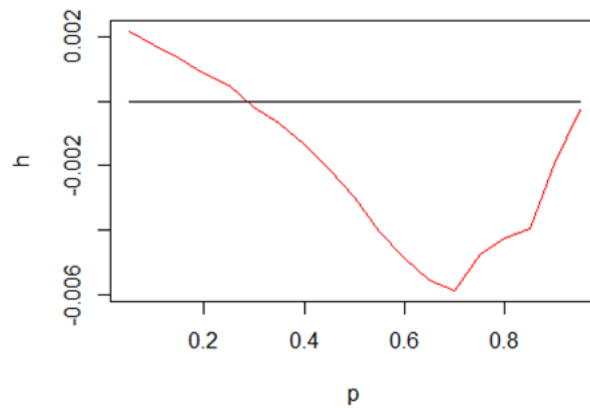


Рис. 3.6. Красным показана разность оптимального значения замещающей вероятности и вероятности заполненного узла. Черным выделена прямая, соответствующая $h = 0$.

Из этих графиков можно заметить, что для вероятности заполненного узла $p \approx 0.29$ значение оптимальной вероятности q_0 совпадает с вероятностью заполненного узла p , поэтому использование усовершенствованного алгоритма Хошена-Копельмана не дает существенного снижения дисперсии.

Также в ходе исследования было установлено, что при больших размерностях решетки n значение разности оптимального значения замещающей вероятности и вероятности заполненного узла h при фиксированном p обратно пропорциональны n^2 .

$$h \sim \frac{1}{n^2}$$

Этим можно пользоваться при практическом использовании сформулированного метода для получения приближенного оптимального значения замещающей вероятности.

В данной главе был сформулирован и реализован метод для моделирования реализации K' с помощью поиска оптимальной замещающей вероятности q_0 , задающей вероятностную меру $\mathbb{Q}_0(\{\omega_i\}) = q_0^m(1 - q_0)^{n^2 - m}$, где m — число заполненных узлов в элементарном исходе. В следующей главе будет получен алгоритм, позволяющий задавать более широкий класс вероятностных мер на перколяционной решетке.

Глава 4

Моделирование с помощью алгоритма, расширяющего класс мер

4.1. Алгоритм, расширяющий класс мер

Опишем алгоритм, позволяющий получать реализацию случайной величины, заданной на $\{\Omega, \mathfrak{A}\}$ с помощью любой вероятностной меры \mathbb{Q} следующего вида: из равенства числа заполненных узлов в двух элементарных исходах ω_i и ω_j следует $\mathbb{Q}(\omega_i) = \mathbb{Q}(\omega_j)$, а также $\forall \omega_k \mathbb{Q}(\omega_k) > 0$. Все вероятностные меры такого вида на $\{\Omega, \mathfrak{A}\}$ назовем расширенным классом мер. Алгоритм назовем алгоритмом, расширяющим класс мер.

Приведем схему этого алгоритма:

- **Входные данные:** размерность решетки n ; распределение дискретной случайной величины R , принимающей целые значения от 0 до n^2 .
- **Тело алгоритма:**
 1. По заданному распределению получается реализация m случайной величины R .
 2. Моделирование решетки, на которой случайным образом заполнено ровно m узлов. Для моделирования используем, например, «Union/find algorithm» [5], позволяющий реализовать решетку в виде массива кластерных меток.
- **Результат работы алгоритма:** массив кластерных меток; количество различных кластерных меток K_R .

Поскольку число различных решеток размерности n с числом заполненных узлов m равно $C_{n^2}^m$, алгоритм позволяет получить реализацию случайной величины, заданной как число кластеров на решетке размерности n с вероятностной мерой

$$\mathbb{Q}(\omega_i) = \frac{r_m}{C_{n^2}^m},$$

где m — число заполненных узлов элементарного исхода ω_i , r_m — вероятность того, что случайная величина R примет значение m . Вероятности $r_m > 0$ подаются на вход

алгоритму. Очевидно, для любых двух элементарных исходов ω_i и ω_j с одинаковым числом заполненных узлов m , $\mathbb{Q}(\omega_i) = \mathbb{Q}(\omega_j)$.

С помощью вероятностей $r_m > 0$ можно задать любую вероятностную меру \mathbb{Q} , удовлетворяющую свойствам

$$0 < \mathbb{Q}(\omega_i) < \frac{1}{C_{n^2}^m} \forall \omega_i;$$

$$\sum_{i=1}^{2^{n^2}} \mathbb{Q}(\omega_i) = 1.$$

Из этих свойств видно, что получен алгоритм, позволяющий задавать на решетке любую вероятностную меру из расширенного класса мер. Например, по своим вероятностным свойствам, алгоритм Хошена-Копельмана с вероятностью заполненного узла p аналогичен алгоритму расширяющему класс мер с распределением случайной величины R , заданным как $\text{Bin}(n^2, p)$.

Для моделирования среднего числа кластеров в исходной задаче на решетке размерности n с вероятностью заполненного узла p , не составляет труда реализовать метод моделирования случайной величины $K' = K_R \frac{p^m(1-p)^{n^2-m} C_{n^2}^m}{r_m}$. Из теоретических выкладок раздела 2.3 видно, что $\mathbb{E}(K) = \mathbb{E}(K')$. Компьютерная программа, осуществляющая моделирование такой случайной величины, также была реализована на языке C#. Моделируя реализации K'_1, \dots, K'_N (независимые случайные величины, распределенные как K'), программа получает реализацию несмещенной оценки \hat{K}' . Для оценки теоретической дисперсии $\text{Var}(K')$ используется несмещенная выборочная дисперсия $S^2 = \frac{1}{N-1} \sum_{i=1}^N (K'_i - \hat{K}')^2$.

4.2. Распределения, на которых достигнуто уменьшение дисперсии

С помощью программы через вероятности r_m было задано несколько распределений R на решетках различной размерности n с различными вероятностями заполненного узла p . Существенное уменьшение несмещенной выборочной дисперсии по сравнению с обычным алгоритмом Хошена-Копельмана было достигнуто на двух распределениях следующего вида.

4.2.1. Дискретный аналог нормального распределения

В качестве распределения случайной величины R рассмотрим некий дискретный аналог нормального распределения со средним $np + a$ и дисперсией σ (a и σ являются изменяемыми параметрами).

- Распределение R задано с помощью вероятностей

$$r_m = P(R = m) = \frac{e^{-\frac{(m-(np+a))^2}{2\sigma^2}}}{\sum_{i=0}^{n^2} e^{-\frac{(i-(np+a))^2}{2\sigma^2}}};$$

- Вероятность элементарного исхода

$$Q(\omega_i) = \frac{r_m}{C_{n^2}^m} = \frac{e^{-\frac{(m-(np+a))^2}{2\sigma^2}}}{\sum_{i=0}^{n^2} e^{-\frac{(i-(np+a))^2}{2\sigma^2}}} \frac{1}{C_{n^2}^m},$$

где m — количество узлов, заполненное в ω_i ;

- Весом элементарного исхода, в котором заполнено ровно m узлов будет

$$W = \frac{p^m(1-p)^{n^2-m} C_{n^2}^m}{r_m}.$$

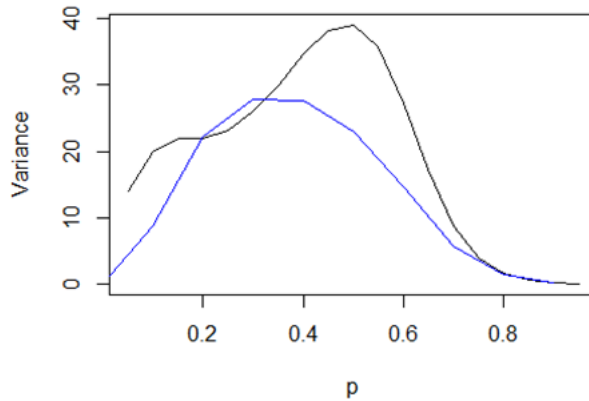


Рис. 4.1. Черным показаны значения несмещенной выборочной дисперсии, полученные с помощью обычного алгоритма Хошена-Копельмана. Синим показаны значения, полученные алгоритмом моделирования с помощью аналога дискретного нормального распределения.

Подбирая оптимальные значения параметров моделирования для перколяционной решетки размерности $n = 20$ для различных вероятностей заполненного узла p при объеме выборки $N = 50000$, были получены значения несмещенной выборочной дисперсии, которые приведены с помощью графика на рисунке 4.1.

Несмотря на то, что метод моделирования с помощью аналога дискретного нормального распределения позволяет достигнуть существенного уменьшения дисперсии, сравнение результатов его работы с результатами, полученными с помощью усовершенствованного алгоритма Хошена-Копельмана, представленное с помощью графика на рисунке 4.2, указывает на проигрышность метода моделирования с помощью аналога дискретного нормального распределения, поэтому алгоритм для поиска оптимальных параметров a и σ разрабатываться не будет.

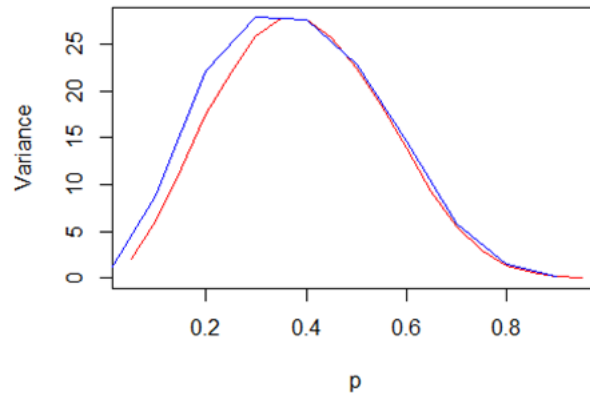


Рис. 4.2. Красным показаны значения несмещенной выборочной дисперсии, полученные с помощью усовершенствованного алгоритма Хошена-Копельмана. Синим показаны значения, полученные алгоритмом моделирования с помощью аналога дискретного нормального распределения.

4.2.2. Усеченное распределение Пуассона

В качестве распределения случайной величины R рассмотрим усеченное распределение Пуассона с параметром λ .

- Распределение R задано с помощью вероятностей

$$r_m = P(R = m) = \frac{\frac{\lambda^m}{m!}}{\sum_{i=0}^{n^2} \frac{\lambda^i}{i!}};$$

- Вероятность элементарного исхода

$$\mathbb{Q}(\omega_i) = \frac{r_m}{C_{n^2}^m} = \frac{\frac{\lambda^m}{m!}}{\sum_{i=0}^{n^2} \frac{\lambda^i}{i!}} \frac{1}{C_{n^2}^m},$$

где m — количество узлов, заполненное в ω_i ;

- Весом элементарного исхода, в котором заполнено ровно m узлов будет

$$W = \frac{p^m (1-p)^{n^2-m} C_{n^2}^m}{r_m}.$$

Подбирая оптимальное значение параметра λ для перколяционной решетки размерности $n = 20$ для малых вероятностей заполненного узла $p \leq 0.2$ при объеме выборки $N = 50000$, были получены значения несмещенной выборочной дисперсии, которые приведены с помощью графика на рисунке 4.3. Для $p > 0.2$ значения несмещенной выборочной дисперсии существенно больше, чем значения, полученные с помощью обычного алгоритма Хошена-Копельмана, поэтому рассматриваться на графике не будут.

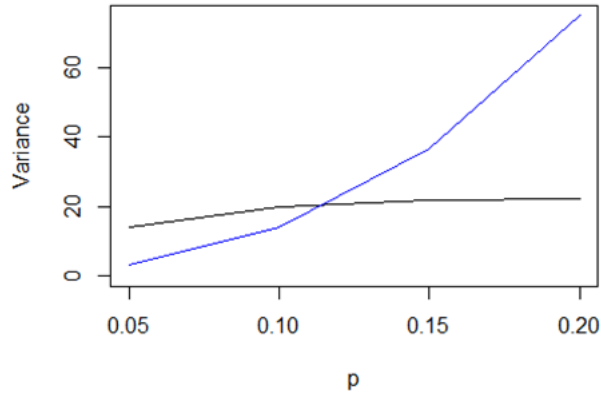


Рис. 4.3. Черным показаны значения несмещенной выборочной дисперсии, полученные с помощью обычного алгоритма Хошена-Копельмана. Синим показаны значения, полученные алгоритмом моделирования с помощью усеченного распределения Пуассона.

Алгоритм моделирования с помощью усеченного распределения Пуассона дает большие значения несмещенной выборочной дисперсии, чем усовершенствованный алгоритм Хошена-Копельмана с оптимальной замещающей вероятностью, что видно из графика на рисунке 4.4, поэтому алгоритм для поиска оптимального параметра λ разрабатываться не будет.

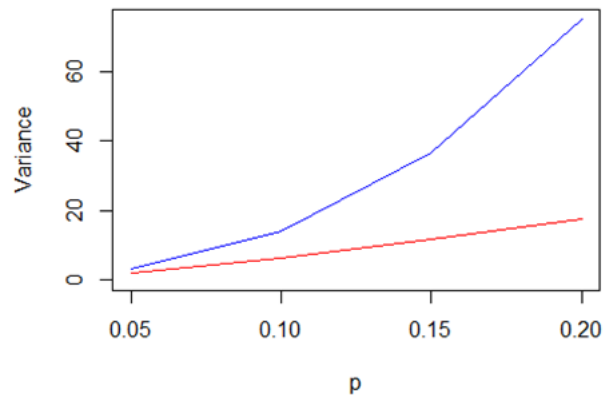


Рис. 4.4. Красным показаны значения несмещенной выборочной дисперсии, полученные с помощью усовершенствованного алгоритма Хошена-Копельмана. Синим показаны значения, полученные алгоритмом моделирования с помощью усеченного распределения Пуассона.

Несмотря на то, что расширенный класс мер гораздо шире класса мер вида $\mathbb{Q}(\{\omega_i\}) = q^m(1-q)^{n^2-m}$, никаких вероятностных мер, обеспечивающих лучшую работу метода существенной выборки, чем мера, разобранный в главе 3, в этом классе не найдено. Поиск таких мер (и, как следствие, задающих меру вероятностей $r_m = P(R = m)$) составляет отдельную обширную задачу, которая данной работой не охватывается.

Заключение

В данной работе:

- усовершенствован алгоритм Хошена-Копельмана для получения несмещенной оценки для среднего числа кластеров на решетке с меньшей дисперсией;
- разработан и реализован алгоритм поиска оптимальной замещающей вероятности для усовершенствованного алгоритма Хошена-Копельмана;
- разработан и реализован алгоритм для задания на перколяционной решетке любой вероятностной меры из расширенного класса мер;
- подобраны дискретные распределения, позволяющие получать несмещенную оценку для среднего числа кластеров с меньшей дисперсией, чем с помощью стандартного алгоритма Хошена-Копельмана.

Список литературы

1. Тарасевич Ю. Ю. Перколяция: теория, приложения, алгоритмы. — М. : Едиториал УРСС, 2002. — 112 с.
2. Мостовой Я. А. Двухфазные операции в больших сетях наноспутников // Компьютерная оптика. 2013. Т. 37. № 1. С. 120-130.
3. Mostovoy Y. A., Berdnikov V. Large scale networks security strategy // CEUR Workshop Proceedings, 2017. — Vol. 1901. — P. 187-193.
4. Hoshen J., Kopelman R. Percolation and cluster distribution. I. Cluster multiple labeling technique and critical concentration algorithm. // Phys. Rev. B 14(8), 3438-3445 (1976).
5. Cormen T. H., Leiserson C. E., Rivest R. R. Introduction to algorithms. — 3rd edition. — MIT Press, 2009.
6. Nikita Meyer. Modelling of the percolation process. — 2018. — May. — Access mode: <https://doi.org/10.5281/zenodo.1254495>.

Приложение А

Приложение для моделирования

Далее приведены скриншоты приложения на языке C# с интерфейсом Windows Forms для реализации указанных в работе алгоритмов. Приложение доступно по ссылке на репозиторий Zenodo [6].

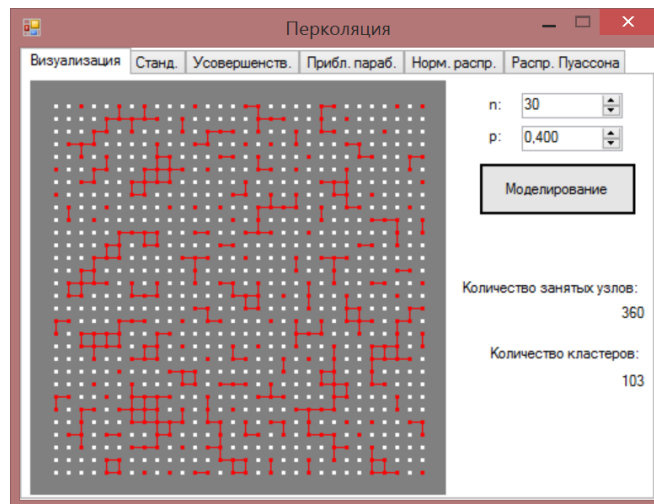


Рис. А.1. На скриншоте представлена вкладка приложения, предназначенная для визуализации случайного эксперимента на перколяционной решетке по заданным значениям размерности решетки n и вероятности заполненного узла p .

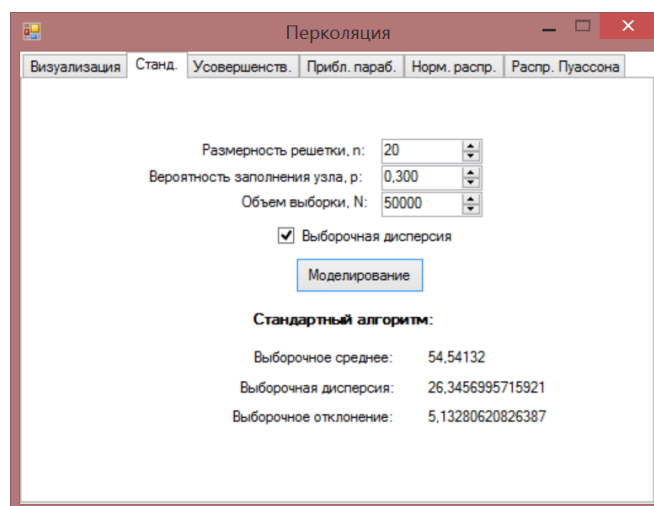


Рис. А.2. На скриншоте представлена вкладка приложения, предназначенная для реализации оценки среднего числа кластеров и вычисления несмещенной выборочной дисперсии с помощью алгоритма Хошена-Копельмана.

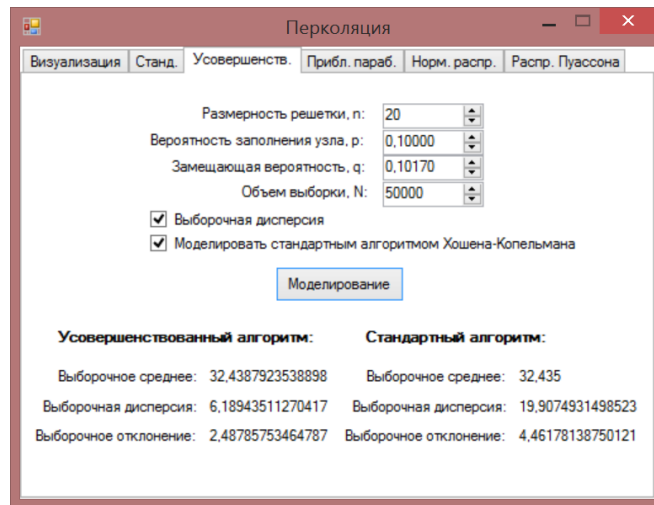


Рис. А.3. На скриншоте представлена вкладка приложения, предназначенная для реализации оценки среднего числа кластеров и вычисления несмещенной выборочной дисперсии с помощью усовершенствованного алгоритма Хошена-Копельмана, а также для сравнения этих значений со значениями, полученными с помощью стандартного алгоритма Хошена-Копельмана.

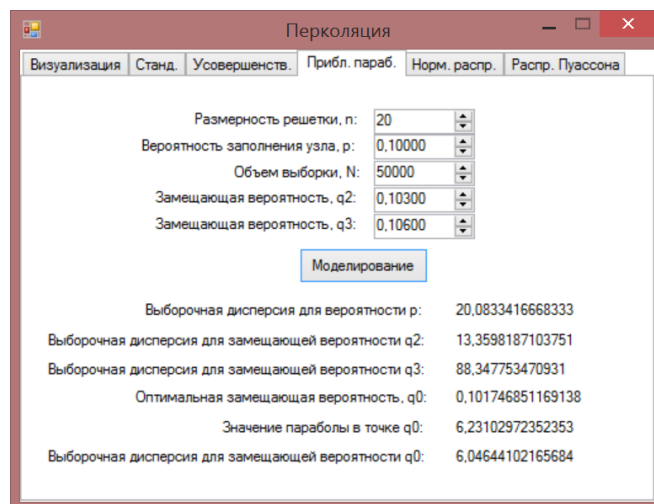


Рис. А.4. На скриншоте представлена вкладка приложения, позволяющая реализовать поиск оптимальной замещающей вероятности с помощью метода приближения параболой.

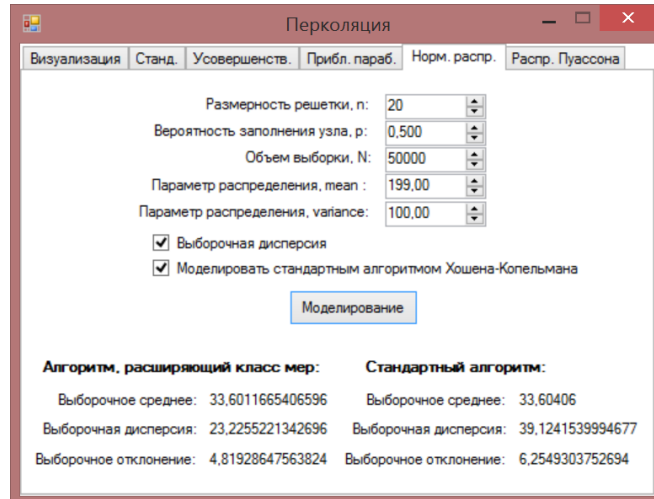


Рис. А.5. На скриншоте представлена вкладка приложения, предназначенная для реализации оценки среднего числа кластеров и вычисления несмещенной выборочной дисперсии с помощью алгоритма, расширяющего класс мер, вероятности в котором заданы с помощью дискретного аналога нормального распределения, а также для сравнения этих значений со значениями, полученными с помощью стандартного алгоритма Хошена-Копельмана.

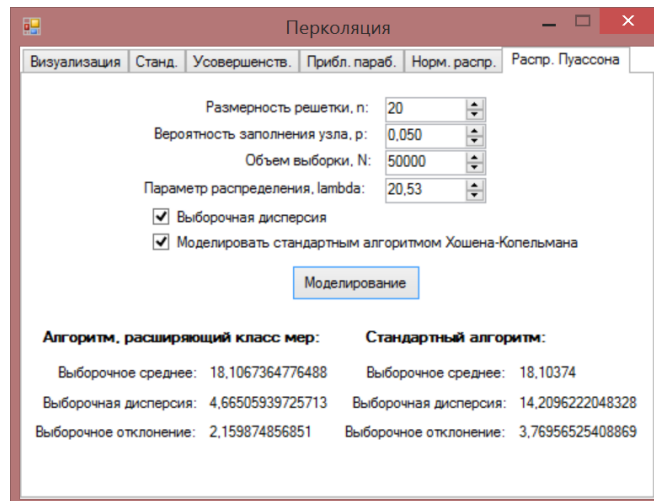


Рис. А.6. На скриншоте представлена вкладка приложения, предназначенная для реализации оценки среднего числа кластеров и вычисления несмещенной выборочной дисперсии с помощью алгоритма, расширяющего класс мер, вероятности в котором заданы с помощью усеченного распределения Пуассона, а также для сравнения этих значений со значениями, полученными с помощью стандартного алгоритма Хошена-Копельмана.