

ПЕРМСКИЙ
ГОСУДАРСТВЕННЫЙ
НАЦИОНАЛЬНЫЙ
ИССЛЕДОВАТЕЛЬСКИЙ
УНИВЕРСИТЕТ

А. Л. Гусев, Ю. Ю. Лесникова

ТЕОРИЯ ПРИНЯТИЯ РЕШЕНИЙ

ПРОЦЕДУРЫ ГРУППОВЫХ ПРОВЕРОК



МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ
Федеральное государственное автономное
образовательное учреждение высшего образования
«ПЕРМСКИЙ ГОСУДАРСТВЕННЫЙ
НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ»

А. Л. Гусев, Ю. Ю. Лесникова

ТЕОРИЯ ПРИНЯТИЯ РЕШЕНИЙ ПРОЦЕДУРЫ ГРУППОВЫХ ПРОВЕРОК

*Допущено методическим советом
Пермского государственного национального
исследовательского университета в качестве
учебного пособия для студентов, обучающихся
по направлению подготовки бакалавров
«Прикладная математика и информатика»*



Пермь 2021

УДК 519.816(075.8)

ББК 22.18я73

Г962

Гусев А. Л.

Г962 Теория принятия решений. Процедуры групповых проверок [Электронный ресурс] : учебное пособие / А. Л. Гусев, Ю. Ю. Лесникова ; Пермский государственный национальный исследовательский университет. – Электронные данные. – Пермь, 2021. – 3,05 Мб ; 90 с. – Режим доступа: <http://www.psu.ru/files/docs/science/books/uchebnie-posobiya/gusev-lesnikova-procedury-grupovyh-proverok.pdf>. – Заглавие с экрана.

ISBN 978-5-7944-3750-8

Настоящее учебное пособие отражает современное состояние теории и практики применения процедур групповых проверок. Общая постановка задачи может быть сформулирована следующим образом. Имеется совокупность объектов, каждый из которых с некоторой вероятностью является дефектным. Требуется выявить все дефектные объекты.

Применение процедур групповых проверок позволяет снизить материальные и физические затраты при обследовании большой совокупности объектов, каждый из которых может быть признан годным или дефектным с некоторыми вероятностями. В качестве обследуемой совокупности могут выступать люди, животные, растения, приборы, химические соединения и т.д.

Авторам настоящего пособия принадлежат результаты, отраженные в гл. 4–9 пособия, которые приводятся без ссылок на литературные источники.

Издание предназначено для студентов, изучающих курс «Теория принятия решений».

УДК 519.816(075.8)

ББК 22.18я73

*Издается по решению редакционно-издательского совета
Пермского государственного национального исследовательского университета*

Рецензенты: кафедра анализа данных и искусственного интеллекта Кубанского государственного университета (зав. каф. – д-р тех. наук, доцент **А. В. Коваленко**);

доцент кафедры информационных технологий в бизнесе НИУ ВШЭ-Пермь, канд. физ.-мат. наук, доцент **Л. В. Шестакова**

ISBN 978-5-7944-3750-8

© Гусев А. Л., Лесникова Ю. Ю., 2021

© ПГНИУ, 2021

ОГЛАВЛЕНИЕ

| | |
|---|----|
| Введение | 4 |
| Глава 1. Первые процедуры групповых проверок..... | 5 |
| Глава 2. Применение групповых процедур проверок при оценивании доли дефектных объектов..... | 9 |
| Глава 3. Нестандартные постановки задач при использовании групповых процедур проверок..... | 11 |
| Глава 4. Моделирование процедуры Дорфмана с оптимальным объемом группы, минимизирующим стоимость проверок..... | 15 |
| 4.1. Построение модели | 15 |
| 4.2. Описание алгоритма и программ для моделирования | 16 |
| 4.3. Анализ результатов моделирования..... | 19 |
| 4.4. Сравнение с результатами Дорфмана | 21 |
| 4.5. Условие оправданности применения метода групповых процедур проверок..... | 27 |
| Глава 5. Моделирование процедуры Стерретта с оптимальным объемом группы, минимизирующим стоимость проверок..... | 31 |
| 5.1. Построение модели | 31 |
| 5.2. Описание алгоритма и программы..... | 33 |
| 5.3. Сравнение с результатами Стерретта | 37 |
| 5.4. Сравнение эффективности процедур Дорфмана и Стерретта с учетом стоимости объединения объектов в группы | 49 |
| Глава 6. Задача определения оптимального объема группы для выбора конечного числа годных объектов | 53 |
| Глава 7. Задача нахождения оптимального объема группы для классификации объектов на t классов ($t > 2$)..... | 55 |
| Глава 8. Две задачи выбора процедуры групповых проверок..... | 63 |
| Список литературы | 74 |
| Приложение 1. Код программы для моделирования процедуры Дорфмана . | 77 |
| Приложение 2. Результаты моделирования при использовании процедуры Дорфмана, минимизирующей стоимость проверок | 82 |
| Приложение 3. Таблица для сравнения классической процедуры Дорфмана с процедурой, в которой оптимальный объем минимизирует стоимость проверок..... | 85 |
| Приложение 4. Код программы для моделирования процедуры Стерретта . | 86 |

ВВЕДЕНИЕ

Метод групповых процедур проверок появился сравнительно недавно как решение проблемы обследования больших групп людей [4]. Благодаря своей эффективности он быстро был распространен на другие, помимо медицины, области, – биологию, технику и т. д.

Постановка задачи в общем случае выглядит следующим образом. Имеется совокупность объектов, каждый из которых с некоторой вероятностью является дефектным. Требуется выявить все дефектные объекты.

Применение процедур групповых проверок позволяет снизить материальные и физические затраты при обследовании большой совокупности. В качестве совокупности могут выступать люди [4, 14, 19, 21, 30], животные [9], растения [3, 22], приборы [1, 10], химические соединения [12, 17, 26, 29, 31]. В случае живых организмов применение групповых процедур может быть полезно во время эпидемии, чтобы как можно быстрее изолировать инфицированных. Групповые процедуры применяются также для отбора здоровых животных при проведении опытов; для проверки наличия у оборудования скрытых дефектов, обусловленных воздействиями окружающей среды, и т.д.

Однако применение групповых процедур не всегда оправдано. Связано это с тем, что для проведения процедуры часто берется объем группы, сводящий к минимуму количество проверок. Подобный подход означает, что стоимость проверки группы объектов приравнивается к стоимости проверки одного объекта. Затратами на объединение объектов в группы пренебрегают сопоставительно со стоимостью одной проверки.

Учет затрат на объединение объектов позволяет иметь дело с реальной стоимостью проведения процедуры групповых проверок и, следовательно, проводить данную процедуру тогда, когда это оказывается выгоднее индивидуальной проверки. Поэтому необходимо выявлять условия, при которых использование групповых процедур окажется менее затратным, чем индивидуальная проверка. Поскольку описанный метод направлен на снижение стоимости процесса проверки совокупности, логично подбирать такой оптимальный объем, при котором именно стоимость проверок была бы минимальной.

Одна из причин применения такого рода заключается в том, что в середине XX в., когда начали появляться первые процедуры групповых проверок, стоимость реактивов для проверки живых организмов была настолько высока, что не шла ни в какое сравнение с затратами, которые требовались на объединение образцов в группы. Следовательно, подход минимизации именно количества проверок был оправдан. Сегодня многие реактивы доступны и доступны по гораздо более низкой цене, а сам метод применяется не только в медицине, поэтому данный подход уже не всегда актуален.

ГЛАВА 1. ПЕРВЫЕ ПРОЦЕДУРЫ ГРУППОВЫХ ПРОВЕРОК

Первая статья на тему групповых процедур проверок была опубликована Робертом Дорфманом в декабре 1943 г. [4]. Предложенный метод групповых процедур проверок был описан на примере его применения в широкомасштабном проекте, которым тогда занимались система общественного здравоохранения США и служба призыва в армию. Суть данного проекта заключалась в выявлении всех людей, зараженных сифилисом, с помощью реакции Вассермана.

Основная идея метода состоит в следующем. После получения образцов крови от каждого отдельного человека эти образцы объединяются в группы и каждая группа исследуется так, будто это кровь одного человека. Если результат проверки группы отрицательный, значит, все люди, чья кровь вошла в данную группу, не инфицированы. Если положительный, то каждый человек, входящий в группу, должен быть проверен уже в индивидуальном порядке.

Р. Дорфман предложил математическую модель ожидаемого количества лабораторных проверок, в которой были приняты следующие обозначения:

p – вероятность того, что произвольно выбранный человек инфицирован (коэффициент распространенности болезни);

$1 - p$ – вероятность того, что случайно выбранный человек не инфицирован;

$(1 - p)^n$ – вероятность того, что случайно взятая группа из n человек не содержит ни одного инфицированного;

$p' = 1 - (1 - p)^n$ – вероятность того, что случайно взятая группа из n человек содержит по крайней мере одного инфицированного человека;

N/n – количество групп по n человек в популяции численности N ;

$p' N/n$ – ожидаемое количество инфицированных групп размера n в популяции численности N с коэффициентом распространенности болезни p .

Тогда предполагаемое количество проверок, которое потребуется при использовании групповых процедур, определяется как сумма количества групп N/n , а также количества человек, которых необходимо обследовать повторно в индивидуальном порядке $n(N/n)p'$:

$$E(T) = N/n + n(N/n)p' \quad (1.1)$$

Р. Дорфман приводит в своей статье такую характеристику, как относительная стоимость проверок, которая равна отношению количества химических

анализов при групповом тестировании к количеству химических анализов при индивидуальном тестировании:

$$C = T / N = \frac{n+1}{n} - (1-p)^n \quad (1.2)$$

Оптимальным объемом группы для своей процедуры Р. Дорфман считал объем, при котором достигается минимальное количество лабораторных проверок (1.1). Иные варианты расчета оптимального объема группы автор не рассматривал, хотя значение относительной стоимости проверок (1.2) также было получено.

Впоследствии другие ученые заинтересовались этой проблематикой и писали статьи со ссылкой на труд Р. Дорфмана, внося изменения в существующую модель или расширяя ее.

Низкий уровень распространенности болезни предполагает, что очередная группа объема n , взятая на проверку, с большой вероятностью окажется годной. Автор работы [21], используя данное условие, предлагает модифицированную модель для проведения процедуры групповой проверки Дорфмана.

Вместо того чтобы проверять индивидуально каждый элемент дефектной группы, в модели индивидуальные проверки проводятся только до обнаружения первого дефектного элемента [21]. Низкий уровень распространенности болезни предполагает, что подгруппа, сформированная из оставшихся непроверенных элементов, с большой вероятностью окажется годной. Если это так, то работа с выборкой завершена, в противном случае снова следует проверять элементы по отдельности до обнаружения первого дефектного элемента. Такая процедура повторяется до тех пор, пока очередная подвыборка не окажется годной.

Пусть вероятность того, что выборка, содержащая k элементов, имеет ровно i инфицированных элементов, равна $\text{Pr}_k(i)$. Тогда ожидаемое количество анализов, которых потребуется провести, чтобы выделить i дефектных элементов с помощью описанной процедуры, равно $E_k(i)$.

Дана генеральная совокупность из N элементов, процент дефектных равен p . $E(N, k, p)$ – это ожидаемое суммарное количество проверок, необходимых для обследования совокупности посредством проверки выборок из k элементов

$$E(N, k, p) = \frac{N}{k} \sum_{i=0}^k \{\text{Pr}_k(i) E_k(i)\} \quad (1.3)$$

Прежде чем $E(N, k, p)$ будет вычислено, необходимо показать, что

$$E_k(i) = \frac{i}{i+1} k + i + 1 + \frac{i}{i+1} - 2i \frac{1}{k} \quad (1.4)$$

Если в выборке не имеется дефектных элементов, одной лабораторной проверки будет достаточно, чтобы установить этот факт. Тогда согласно (1.4) $E_k(0) = 1$.

Далее в статье [21] приведена формула

$$E_k(n) = 1 + \frac{n}{k} \{1 + E_{k-1}(n-1)\} + \sum_{j=1}^{k-n} \left[\left(\prod_{i=1}^j \frac{k-(i+n-1)}{k-(i-1)} \right) \frac{n}{k-j} \{(j+1) + E_{k-[j+1]}(n-1)\} \right]. \quad (1.5)$$

Первое слагаемое в правой части формулы (1.5) представляет собой начальный групповой тест. Множитель n/k в следующем слагаемом – это вероятность того, что результат проверки первого элемента выборки положительный. Множитель $\{1 + E_{k-1}(n-1)\}$ – это общее количество тестов, необходимых для определения дефектного элемента при первом испытании и среднее количество тестов, необходимых для нахождения $(n-1)$ дефектных элементов среди оставшихся $k-1$ элементов выборки.

Вероятность того, что первые j элементов годные, равна

$$\prod_{i=1}^j [k-(i+n-1)]/[k-(i-1)],$$

в то время как вероятность того, что $(j+1)$ -й проверяемый элемент дефектный, равна $n/(k-j)$. Количество тестов, необходимое для определения первого дефектного, равно $(j+1)$, тогда $E_{k-[j+1]}(n-1)$ – предполагаемое количество тестов, которое потребуется, чтобы найти оставшихся $(n-1)$ дефектных среди $k-[j+1]$ элементов.

Уравнение (1.5) сводится к виду (1.4) при соответствующей замене значений $E_{k-[j+1]}(n-1)$, полученных из уравнения (1.5). Доказательство для формулы $E_k(i)$ следует по индукции.

Процедура групповых проверок, рассмотренная в работе [19], применяется для так называемой модифицированной биномиальной модели, или М Модели. Особенностью М Модели, отличающей ее от биномиальной, которая традиционно применяется в групповом тестировании, является измеримость результата проверки – отклика. Результат проверки одного годного объекта принимается равным нулю, а результат проверки одного дефектного объекта – равным любому положительному значению. Отклик годной группы также равен нулю, а отклик дефектной равен сумме индивидуальных откликов объектов в группе. Описанная особенность дает М Модели значительное преимущество: экспериментатор всегда сможет определить статус двух подгрупп одной дефектной группы,

сделав групповой тест только для одной из подгрупп. Рассмотрены случаи конечной и бесконечной популяций.

Как можно убедиться, методы, основанные на применении процедуры Дорфмана, достаточно просты в реализации: оптимальный объем группы легко может быть получен, причем независимо от результатов предыдущих проверок. Также проверка независимых групп может выполняться одновременно. Однако такие процедуры имеют особенность: затраты на объединение объектов в группы должны быть малы по сравнению со стоимостью одной проверки.

С. Кумар исследовал проблему разделения имеющихся N элементов на k не пересекающихся между собой категорий с помощью группового тестирования [15, 16]. k категорий были условно названы как «лучшие элементы», «вторые лучшие», ..., « k -е лучшие».

Модель, которую разработали Самуэль Коц и Мао Шисонг, учитывает вероятность возникновения ошибок при проведении лабораторных проверок [13]. В данной модели были приняты следующие допущения:

- 1) результат группового теста является измеримой величиной и определяется как сумма реакций каждого отдельного объекта, входящего в группу;
- 2) положительная реакция дефектного объекта – величина, одинаковая для всех объектов из совокупности;
- 3) положительная реакция равна 1.

Несмотря на то что в [13, 15, 16] модель групповых процедур проверок была модифицирована, здесь метод применялся также без учета стоимости объединения объектов в группы.

ГЛАВА 2. ПРИМЕНЕНИЕ ГРУППОВЫХ ПРОЦЕДУР ПРОВЕРОК ПРИ ОЦЕНИВАНИИ ДОЛИ ДЕФЕКТНЫХ ОБЪЕКТОВ

Процедуры групповых проверок используются не только для классификации объектов на годные и дефектные при сравнительно низкой вероятности дефектности, но и для оценки самой вероятности дефектности объекта. Как правило, помимо определения оценки вероятности дефектности объекта, в литературе приводятся и соответствующие оптимальные объемы групп для проведения процедуры.

Весьма распространено применение групповых процедур проверок для получения оценки коэффициента распространенности болезни [2, 5, 23, 25]. Использование методологии избыточной дисперсии для оценки уровня заражения яиц сальмонеллой и оценки распространенности вирусной инфекции желтой лихорадки у популяции комаров продемонстрировано в [5]. В работе [2] показано, как получить оценку коэффициента распространенности заболевания при условии того, что статус болезни может определяться с ошибкой.

Вычисление оценки уровня зараженности и вероятности передачи заболевания на примере различных видов растений и самих заболеваний приведены в источниках [3, 22, 24]. Здесь в качестве оптимального объема группы для проведения групповых процедур берется объем, при котором значение средней квадратичной ошибки оценки вероятности передачи заболевания оказывается наименьшим.

Другой пример практического применения групповых процедур представлен в [10]. Статья посвящена получению начальных и уточненных оценок количества скрытых дефектов, присущих новому или недавно отремонтированному оборудованию, подверженному воздействию окружающей среды (вибрации, периодическое воздействие тепловых нагрузок и т.д.).

Имеется выборка из m единиц оборудования. Каждая из единиц оборудования состоит из k независимых одинаковых частей, качество части характеризуется вероятностью π быть дефектной. Параметр π неизвестен и зависит от характеристик окружающей среды, в условиях которой планируется использовать оборудование. Особенность работы заключается в том, что объем группы для групповых процедур проверок фиксирован и известен заранее. Автор рассматривает случай, когда все k частей оборудования идентичны, то есть объем группы для проведения групповых процедур равен k . Также в статье замечено, что в случае оборудования, состоящего из частей нескольких типов, описанный метод можно использовать для каждого типа частей отдельно.

Метод применения групповых процедур проверок в работе [6] направлен на то, чтобы сохранить индивидуальную анонимность при получении оценки коэффициента распространенности болезни. Авторы подчеркивают, что, несмотря на связь используемого метода с классической процедурой, предложенной Дорфманом, цель применения метода другая: специально не идентифицировать зараженных людей, чтобы защитить их гражданские свободы. Метод получения оценки максимального правдоподобия применен на данных о заболеваемости СПИДом ряда канадских испытуемых.

ГЛАВА 3. НЕСТАНДАРТНЫЕ ПОСТАНОВКИ ЗАДАЧ ПРИ ИСПОЛЬЗОВАНИИ ГРУППОВЫХ ПРОЦЕДУР ПРОВЕРОК

Групповые процедуры проверок, применяемые для поиска неисправностей в радиоэлектронных системах (РЭС), используются как самостоятельный метод, так и в составе комплекса диагностических методов [1]. Применительно к решаемой задаче наряду с групповыми процедурами в статье описаны следующие методы: последовательных поэлементных проверок и комбинационный.

Метод последовательных поэлементных проверок предполагает, что элементы системы проверяются по одному в заранее установленном порядке. Если проверка очередного элемента показала, что он неисправен, проводится восстановление этого элемента. Затем выполняется комплексная проверка работоспособности аппаратуры. При возникновении отказа проверка продолжается с последнего неисправного элемента. Описанные действия повторяются до тех пор, пока работоспособность аппаратуры не будет восстановлена. Метод применим для систем любой конструкции, однако, несмотря на универсальность, отмечается низкая эффективность метода, обусловленная значительными временными затратами на поиск неисправностей.

Метод групповых процедур оказывается менее трудоемким. Посредством измерения одного или нескольких параметров выявляется группа элементов, содержащая неисправный элемент. Аналогичным образом неисправность локализуется до тех пор, пока не будет найден конкретный элемент, требующий восстановления. Несмотря на очевидную эффективность групповых проверок по сравнению с предыдущим методом, данный метод применим для систем, которые можно разделить на последовательно сужающиеся участки.

Аппаратура с разветвленной структурой требует более гибкого подхода к поиску неисправностей, поэтому в данной ситуации востребован комбинационный метод, сочетающий в себе поэлементные и групповые проверки. Сначала производится замер определенного набора параметров, затем в зависимости от того, значения каких параметров оказались в норме или не в норме, выявляется поврежденный элемент.

Как правило, на практике при диагностике аппаратуры применяются все три метода одновременно: «Определение неисправного устройства, тракта или сложного блока целесообразно осуществлять, используя комбинационный метод; отыскание неисправного узла (каскада) в последовательных трактах прохождения сигналов – с помощью группового метода, а неисправных элементов в узлах – поэлементных проверок» [1].

Упомянутая статья не содержит расчетов и теоретических обоснований эффективности представленных методов для решения задачи обнаружения неисправностей в радиоэлектронных системах, поэтому вряд ли можно утверждать, что метод групповых процедур менее затратный с точки зрения стоимости производимых проверок по сравнению с последовательными проверками. Следует произвести расчет стоимости проверок при использовании каждого из методов с учетом стоимости проверки одного элемента системы и стоимости проверки группы элементов, включающей накладные расходы на сборку и разборку группы элементов.

Применение групповых процедур проверок на примере аккумулятора, состоящего из 24 отдельно обсаженных ячеек, рассмотрено в работе [18]. Особенность представленной в статье процедуры заключается в необходимости так называемого повторного тестирования. В соответствии с нормами проверяемый аккумулятор должен доставлять ток силой I ампер за T минут, поддерживая напряжение 24 вольта. Первоначально каждый аккумулятор проверяется как единица, таким образом отбираются соответствующие норме или годные аккумуляторы. Затем ячейки, составляющие аккумуляторы, не соответствующие норме, проверяются отдельно, чтобы выявить, какие из них доставляют ток силой I ампер за T минут, поддерживая напряжение 1 вольт. Далее как раз необходимо повторное тестирование, чтобы собрать аккумуляторы из годных ячеек и убедиться, что полученные аккумуляторы соответствуют норме.

Кроме того, в статье [18] рассмотрена точка равновесия – вероятность дефектности объекта, при которой достигается равенство между групповой и индивидуальной проверками в рамках рассматриваемой процедуры. Точка равновесия получена для нескольких вариантов групповых процедур как с повторным тестированием, так и без него – т. е. для классической процедуры Дорфмана.

Использование процедур групповых проверок в генетике для обнаружения мутаций имеет одну принципиальную особенность: чем больше группа для проведения процедуры, тем сложнее обнаружить мутации, так как чувствительность метода анализа зависит от концентрации мутантных генов в выборке [7, 14, 30]. Основная цель исследований в данном направлении заключается в определении максимального размера группы для проведения групповых процедур, при котором чувствительность метода останется равной 100%.

Описанная особенность также была учтена при применении метода для обнаружения генетически модифицированных организмов в кукурузе [28]. Если группа достаточно большая, то генетически модифицированные организмы в ней не будут обнаружены из-за того, что их концентрация будет слишком мала. С другой стороны, использование групп меньшего размера увеличит стоимость

проведения процедуры и индивидуальная проверка может оказаться менее затратной. Оптимальный объем группы в данном случае выбирается с учетом того, что максимальная доля ГМО, которая может попасть в группу, не должна превышать пороговую долю ГМО, к которой потребитель сможет быть толерантным.

Процедуры групповых проверок регулярно применяются в фармацевтической промышленности для обнаружения среди сотен тысяч соединений таких, которые потенциально могут оказаться действующими веществами для нового лекарственного препарата. Применение групповых процедур проверок для решения данной задачи весьма осложнено по двум причинам.

Во-первых, группа, признанная негативной, может содержать активные соединения, реакция которых подавляется веществами-блокаторами. Присутствие блокаторов при проведении групповых процедур проверок было исследовано в ряде работ в области не только фармакологии [17, 31], но и медицины [8, 20].

Например, в [20] указано, что образец крови человека с высоким титром антител к антигену гепатита В может скрыть образец крови человека с низким уровнем антигена гепатита В при объединении данных образцов крови в одну группу. В статье [8] объясняется, что метод групповых процедур проверок не был принят в Соединенных Штатах для проверки образцов донорской крови по причине «небольшой потери чувствительности» метода. Авторы предполагают, что потеря чувствительности может быть вызвана присутствием блокаторов.

Вторая сложность в применении метода групповых процедур проверок заключается в том, что группа из n неактивных соединений может давать положительный отклик (что свидетельствует о наличии в группе активного вещества) вследствие синергетического эффекта. Химически несходные соединения синергичны друг с другом, если в сочетании они обнаруживают активность, а по отдельности обнаруживают слабую активность или не обнаруживают активности вовсе. Групповые проверки в условиях присутствия блокаторов и синергизма рассмотрены в [12, 26, 29].

Два критерия для создания выборок из химических соединений для обнаружения потенциальных действующих веществ в условиях присутствия блокаторов и синергизма представлены в [12]. Первый критерий пытается охватить как можно более равномерно все множество химических соединений, подлежащих проверке. Вторым критерием призван сократить количество совпадений образцов соединений, попадающих в одну выборку. В статье приводится аргументация, почему именно данные критерии важны при формировании выборки.

В работе [29] в качестве главного критерия эффективности метода рассматривается доля ложных положительных и ложных отрицательных результатов

проверок. Различные комбинации методов объединения химических соединений и схем повторного тестирования рассматриваются как с точки зрения стоимости проверок, которая выражается количеством самих проверок, так и с точки зрения минимизации ошибок, вызванных ложными положительными и ложными отрицательными результатами проверки.

Трудоёмкость и высокая стоимость процесса обнаружения новых активных соединений в фармакологической промышленности подтверждается работой [11], где эффективность применения групповых процедур проверок оценивается исходя из стоимости процесса проверки. Рассмотрены две различные функции стоимости и ряд разновидностей самих процедур: двухэтапная, трехэтапная, процедура строк и колонок. Проведено сравнение стоимости одноэтапной и многоэтапных процедур для обеих функций стоимости.

При классическом использовании групповых процедур проверок решалась задача выявления всех дефектных объектов из некоторой совокупности. А. Л. Гусев предложил другую постановку задачи: отбор из совокупности объектов определенного количества годных объектов [9].

Потребность в процедуре такого рода возникла из необходимости решить следующую задачу. В рамках проверки действия различных веществ на живой организм требовалось провести ряд опытов на мышах. Для проведения опытов из тысяч мышей нужно было отобрать несколько десятков здоровых, чтобы опыт был достоверным. Для проверки здоровья брали пробу крови, после чего можно было проверить каждую мышшь индивидуально либо применить метод групповых проверок Р. Дорфмана. В данном случае оптимальный объем группы подбирается таким образом, чтобы за одну групповую проверку можно было отобрать как можно больше здоровых мышей. Подробно процедура описана в гл. 6 этого пособия.

Применение процедур групповых проверок для отрицательного биномиального распределения продемонстрировано в [27] на примере данных по заболеваемости хламидиозом. Использование отрицательного биномиального распределения означает, что наблюдение продолжается до тех пор, пока не будет выявлена первая положительная группа. Это полезно, когда необходимо как можно скорее выявить случай проявления редкого признака. Оптимальный объем группы подбирается сведением к минимуму асимптотической дисперсии (V-правило).

ГЛАВА 4. МОДЕЛИРОВАНИЕ ПРОЦЕДУРЫ ДОРФМАНА С ОПТИМАЛЬНЫМ ОБЪЕМОМ ГРУППЫ, МИНИМИЗИРУЮЩИМ СТОИМОСТЬ ПРОВЕРОК

4.1. Построение модели

Данный параграф посвящен построению математической модели процесса проверки совокупности с помощью групповых процедур проверок, оптимальный объем группы данной модели определяется из соображений минимизации стоимости проверок.

Далее будут использованы обозначения, введенные Дорфманом.

- N – объем совокупности, в которой каждый объект является дефектным с вероятностью p ;
- $p' = (1 - (1 - p)^n)$ – вероятность получить дефектную группу, т. е. такую, в которой есть по крайней мере один дефектный объект;
- n – объем группы для проведения групповой процедуры.

В классической процедуре Дорфмана, как упоминалось ранее, оптимальный объем группы выбирается таким образом, чтобы количество проверок при заданной вероятности дефектности объекта было наименьшим. Другими словами, решалась следующая задача:

$$\begin{cases} N/n + n(N/n)p' \rightarrow \min \\ n \in \mathbb{N} \end{cases}.$$

Здесь N/n – количество групповых тестов, а $n(N/n)p'$ – количество проверок, которое нужно произвести, чтобы индивидуально проверить все объекты из дефектных групп.

Далее сформулирована постановка задачи нахождения такого объема группы для проведения групповых процедур проверок, при котором стоимость проверок будет минимальной с учетом стоимости объединения объектов в группы.

Пусть C_1 – стоимость проверки одного объекта, C_n – стоимость проверки группы из n объектов, причем $0 < C_1 < C_n$, следовательно, общая стоимость процесса обследования равна

$$C_{\text{обсл.}} = \frac{N}{n} C_n + \frac{N}{n} p' n C_1. \quad (4.1)$$

По аналогии с Дорфманом можно ввести понятие относительной стоимости проверок или стоимость проверки одного объекта в условных единицах:

$$C_{\text{отн.}} = \frac{C_{\text{обсл.}}}{N} = \frac{C_n}{n} + p' C_1. \quad (4.2)$$

Тогда оптимальный объем группы находится из решения следующей задачи:

$$\begin{cases} \frac{C_n}{nC_1} + p' \rightarrow \min \\ n \in \mathbb{N} \end{cases} \quad (4.3)$$

Предполагаемое количество проверок при этом может быть вычислено так же, как в классической процедуре Дорфмана (1.1).

Необходимо рассмотреть различные значения N , p и C_n/C_1 и сравнить стоимость индивидуального обследования со стоимостью обследования с применением процедуры групповых проверок Дорфмана.

4.2. Описание алгоритма и программ для моделирования

Моделирование было осуществлено с помощью программы, написанной на языке Python. Код программы представлен в прил.1.

Моделирование процедуры Дорфмана, рассматриваемой в данном разделе, подразумевает определение оптимального объема группы, минимизирующего стоимость проверок.

Формулы подсчета стоимости проверок (4.1) и количества проверок (1.1) используют количество групп оптимального объема N/n . На практике чаще всего оказывается, что N не кратно n . Поэтому при написании программы было использовано следующее предположение.

Пусть $k = N - [N/n^*]$, где n^* – оптимальный объем группы, $[N/n^*]$ – целая часть от деления N на n^* . Стоимость проверки группы из k элементов при этом принимается равной стоимости проверки группы из n элементов, т. е. равной C_n , причем с вероятностью $(1 - (1 - p)^k)$ данная группа окажется дефектной, тогда для выявления всех дефектных объектов потребуется провести k дополнительных индивидуальных проверок стоимостью $k \cdot C_1$.

Решение задачи поиска оптимального с точки зрения стоимости объема группы для проведения групповой процедуры фактически заключается в поиске наименьшего значения функции стоимости (4.1) на отрезке $[1, N]$, поэтому перед написанием программы следует изучить поведение функции стоимости проверок на указанном отрезке.

Функция (4.1) ведет себя по-разному на рассматриваемом отрезке в зависимости от величины соотношения C_n/C_1 и вероятности дефектности объекта p . Возможны три варианта, они продемонстрированы с помощью графиков на конкретных примерах.

На рис. 4.1 представлены графики функции стоимости (4.1) для фиксированных значений вероятности дефектности объекта и объема совокупности: $p = 0.25$ и $N = 100$, при этом рассмотрены три различных значения отношения C_n / C_1 : 1, 1.5 и 3. При $C_n / C_1 = 1$, которое соответствует классической процедуре Дорфмана, локальный минимум функции (4.1) как раз является наименьшим значением на отрезке $[1, N]$. При $C_n / C_1 = 1.5$ функция стоимости имеет локальный минимум, но очевидно, что он не является наименьшим значением на отрезке $[1, N]$, поэтому нельзя взять соответствующий этому минимуму объем группы в качестве оптимального с точки зрения стоимости проверок. Кроме того, функция стоимости (4.1) может не иметь локального минимума на отрезке $[1, N]$ вообще и монотонно убывать с ростом n , такая ситуация наблюдается при $C_n / C_1 = 3$.

Зависимость стоимости проверок от объема группы

Стоимость проверок

на 100 объектов

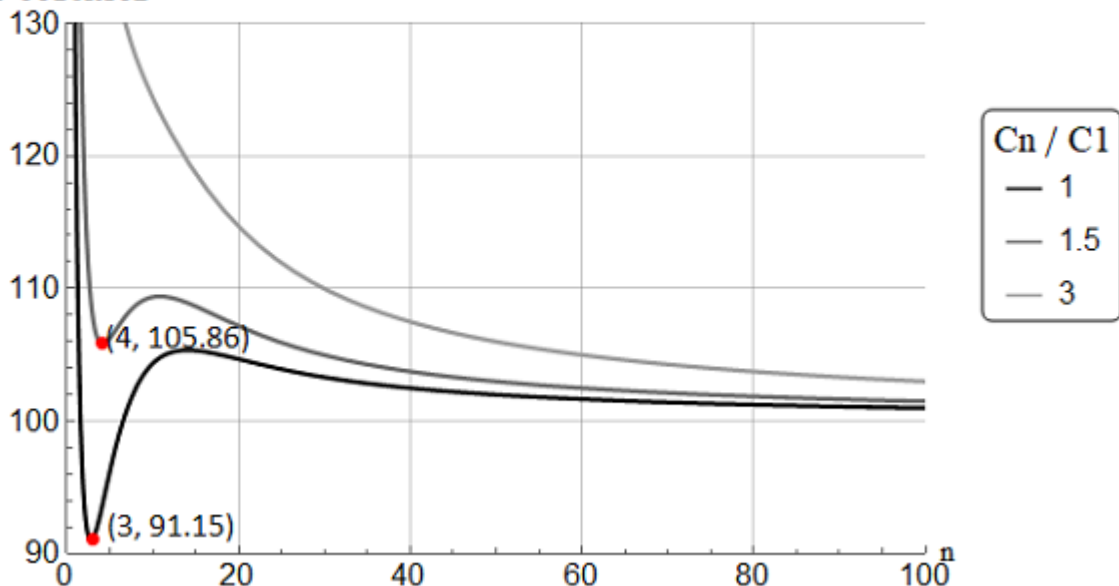


Рис. 4.1. Зависимость стоимости проверок от объема группы при $N = 100$, $p = 0,25$

Если $C_n / C_1 = 1.5$ или $C_n / C_1 = 3$ (верхние графики), применение групповых процедур невыгодно, так как наименьшее значение функции стоимости достигается на правой границе отрезка, т. е. для n , равного размеру популяции N . Такой результат предполагает проведение одного группового теста на всей совокупности и последующее индивидуальное тестирование. Так как в условиях решаемой задачи вероятность того, что все объекты совокупности годные, стремится к нулю, начальный групповой тест на совокупности не имеет смысла и в данных ситуациях следует сразу же выполнять индивидуальные проверки.

Дорфман проводил вычисления для достаточно больших значений вероятности дефектности объекта, поэтому в программе предусмотрено возникновение такой ситуации, когда применение групповых процедур проверок невыгодно в рассматриваемых условиях и дешевле будет выполнить проверку объектов индивидуально. Сначала в программе выполняется поиск локального минимума функции стоимости (4.1) на отрезке $[1, N]$, и если минимум найден, он сравнивается со стоимостью индивидуального обследования совокупности.

Программа принимает на вход следующие данные:

- вероятность дефектности объекта p ,
- объем популяции N ,
- стоимость проверки одного объекта C_1 ,
- стоимость проверки n объектов C_n .

Результатом работы программы являются следующие показатели:

- Оптимальный объем группы, минимизирующий стоимость проверки.
- Оптимальный объем группы, минимизирующий количество проверок.
- Стоимость процесса проверки при использовании групповых проверок с оптимальным объемом группы, минимизирующим стоимость процесса проверки.
- Стоимость процесса проверки при использовании классической процедуры Дорфмана.
- Стоимость процесса индивидуальной проверки.

Программа была запущена для следующих значений указанных выше параметров:

- $p = 0.01, 0.005, 0.001$
- $N = 100\ 000, 1\ 000\ 000, 5\ 000\ 000$
- $C_1 = 1$
- $C_n = C_1, 1.05C_1, 1.1C_1, 1.2C_1, 1.5C_1, 2C_1, 3C_1, 5C_1, 10C_1$.

Здесь и далее в качестве стоимости проверки одного объекта C_1 берется единица, так как в рамках решаемой задачи конкретные цены за индивидуальные и групповые тесты сами по себе не имеют значения, важна лишь величина отношения стоимости проверки группы к стоимости проверки одного объекта. Таким образом, любая цена C_1 может быть принята за единицу, тогда цены за групповые тесты можно рассматривать как доли C_1 .

Следует отметить, что ситуация, когда $C_n = C_1$, соответствует классической процедуре Дорфмана, в которой не учитываются затраты на объединение объектов в группы. Таким образом, классическая процедура Дорфмана, направленная на уменьшение количества проверок, оказывается частным случаем процедуры, минимизирующей стоимость процесса проверки.

4.3. Анализ результатов моделирования

Результаты моделирования, полученные с помощью программы, представлены в виде таблиц. Далее будут рассмотрены результаты моделирования для объема совокупности, равного 100 000 (табл. 4.1 и 4.3). С остальными результатами можно ознакомиться в прил. 2.

Колонка C_n / C_1 содержит соотношение стоимости проверки группы объектов к стоимости проверки одного объекта. Соотношение, равное 1, соответствует случаю, когда стоимость проверки группы объектов равна стоимости проверки одного объекта, т. е. для проверки объектов применяется классическая процедура Дорфмана. Оптимальный объем группы получен из решения задачи (4.3). Для каждого варианта групповой процедуры, определяемого соотношением C_n / C_1 , были вычислены оценки количества проверок, стоимости проверок при использовании групповых процедур и стоимости индивидуальных проверок. Значения в столбце «Экономия» показывают, во сколько раз индивидуальная проверка оказывается дороже групповой.

Таблица 4.1. Сравнение результатов проверки при проведении групповой и индивидуальной процедур проверок при $p = 0.001$, $N = 100000$, $C_1 = 1$

| C_n / C_1 | Опт. объем группы | Количество проверок | Стоимость групп. проверок | Стоимость индив. проверок | Экономия |
|-------------|-------------------|---------------------|---------------------------|---------------------------|----------|
| 1,00 | 32 | 6276 | 6276,00 | 100000,00 | 15,9 |
| 1,05 | 33 | 6280 | 6430,07 | 100000,00 | 15,6 |
| 1,10 | 34 | 6288 | 6580,53 | 100000,00 | 15,2 |
| 1,20 | 35 | 6300 | 6870,60 | 100000,00 | 14,6 |
| 1,50 | 39 | 6393 | 7674,17 | 100000,00 | 13,0 |
| 2,00 | 46 | 6673 | 8845,84 | 100000,00 | 11,3 |
| 2,50 | 51 | 6936 | 9876,64 | 100000,00 | 10,1 |
| 3,00 | 56 | 7235 | 10806,13 | 100000,00 | 9,3 |
| 5,00 | 73 | 8413 | 13892,72 | 100000,00 | 7,2 |
| 10,00 | 103 | 10763 | 19500,89 | 100000,00 | 5,1 |

Таблица 4.2. Сравнение результатов проверки при проведении групповой и индивидуальной процедур проверок при $p = 0.005$, $N = 100000$, $C_1 = 1$

| C_n / C_1 | Опт. объем группы | Количество проверок | Стоимость групп. проверок | Стоимость индив. проверок | Экономия |
|-------------|-------------------|---------------------|---------------------------|---------------------------|----------|
| 1,00 | 15 | 13911 | 13911,00 | 100000,00 | 7,2 |
| 1,05 | 15 | 13911 | 14243,22 | 100000,00 | 7,0 |
| 1,10 | 15 | 13911 | 14576,57 | 100000,00 | 6,9 |
| 1,20 | 16 | 13957 | 15206,89 | 100000,00 | 6,6 |
| 1,50 | 18 | 14184 | 16961,14 | 100000,00 | 5,9 |
| 2,00 | 21 | 14754 | 19515,08 | 100000,00 | 5,1 |
| 2,50 | 24 | 15502 | 21751,57 | 100000,00 | 4,6 |
| 3,00 | 26 | 16067 | 23759,68 | 100000,00 | 4,2 |
| 5,00 | 34 | 18612 | 30378,72 | 100000,00 | 3,3 |
| 10,00 | 51 | 24518 | 42166,14 | 100000,00 | 2,4 |

Таблица 4.3. Сравнение результатов проверки при проведении групповой и индивидуальной процедур проверок при $p = 0.01$, $N = 100000$, $C_1 = 1$

| C_n / C_1 | Опт. объем группы | Количество проверок | Стоимость групп. проверок | Стоимость индив. проверок | Экономия |
|-------------|-------------------|---------------------|---------------------------|---------------------------|----------|
| 1,00 | 11 | 19558 | 19558,00 | 100000,00 | 5,1 |
| 1,05 | 11 | 19558 | 20011,63 | 100000,00 | 5,0 |
| 1,10 | 11 | 19558 | 20466,18 | 100000,00 | 4,9 |
| 1,20 | 12 | 19697 | 21362,02 | 100000,00 | 4,7 |
| 1,50 | 13 | 19942 | 23787,07 | 100000,00 | 4,2 |
| 2,00 | 15 | 20661 | 27327,72 | 100000,00 | 3,7 |
| 2,50 | 17 | 21589 | 30412,59 | 100000,00 | 3,3 |
| 3,00 | 19 | 22648 | 33174,71 | 100000,00 | 3,0 |
| 5,00 | 25 | 26218 | 42217,86 | 100000,00 | 2,4 |
| 10,00 | 38 | 34375 | 58061,92 | 100000,00 | 1,7 |

Сравнение стоимостей индивидуальной и групповой проверок показало, что при рассмотренных значениях входных параметров использование групповых процедур оказывается более эффективным, чем индивидуальная проверка, даже при условии, что стоимость проверки группы выше стоимости проверки одного объекта. Однако уменьшение экономии с увеличением величины отношения C_n / C_1 позволяет предположить, существует такое значение отношения C_n / C_1 , начиная с которого применение групповых процедур теряет смысл: стоимость процесса обследования совокупности при использовании групповых процедур становится дороже стоимости процесса индивидуального обследования.

Рассмотрены сравнительно небольшие вероятности дефектности объекта, чтобы показать, как взаимные изменения параметров влияют на результаты мо-

делирования, а именно: на количество проверок и на их стоимость. Далее, в параграфе, посвященном сравнению построенной процедуры с процедурой Дорфмана, будут получены и проанализированы результаты для больших вероятностей дефектности объекта.

Увеличение оптимального объема группы с увеличением стоимости проверки группы свидетельствует о том, что объединять объекты в группы становится настолько дорого, что для достижения экономии следует объединять объекты в большие группы и, как следствие, проводить меньшее количество групповых и большее количество индивидуальных проверок по сравнению с классической процедурой Дорфмана при прочих равных условиях.

Данные в табл. 4.1 и 4.3 также позволяют сделать достаточно очевидный вывод: чем дешевле стоимость проверки группы объектов по сравнению со стоимостью проверки одного объекта, тем дешевле обходится применение групповых процедур. Однако на практике едва ли возможны ситуации, когда объединение объектов в группы и разделение группы обратно на объекты не требуют никаких затрат.

4.4. Сравнение с результатами Дорфмана

Дорфман опубликовал в своей статье [4] оптимальные объемы групп и относительные стоимости проверок (1.2) для ряда значений коэффициента распространенности болезни p . Программа, описанная в п. 4.2, была использована для сравнения результатов, полученных Дорфманом, с новой процедурой, в которой применяется оптимальный с точки зрения стоимости проверок объем группы.

Относительная стоимость проверок у Дорфмана – это среднее количество проверок, приходящееся на один объект, при конкретном значении p . Считая, что в процедуре Дорфмана стоимость проверки одного объекта равна стоимости проверки группы и равна $C_1 = C_n = 1$, можно умножить относительную стоимость проверок на стоимость одной проверки, это даст средние затраты в условных единицах, приходящиеся на один объект:

$$C_{\text{отн.}} = C_1 \cdot \frac{T}{N} = C_1 \cdot \left[\frac{n+1}{n} - (1-p)^n \right]. \quad (4.4)$$

При этом численные значения C и $C_{\text{отн.}}$ совпадают.

Те же самые показатели: оптимальный объем группы и относительная стоимость проверок в условных единицах — были получены с помощью программы и для новой модели при значениях p , взятых из статьи [1]. Оптимальный объем группы для проведения процедуры определяется в программе исходя из предпо-

ложений, описанных в п. 4.2. Средние затраты, приходящиеся на проверку одного объекта, для новой модели вычисляются в программе по формуле (4.2). Стоимость проверки одного объекта C_1 также принимается равной единице.

Сравнить оптимальный объем групп и соответствующую относительную стоимость проверок построенной модели с результатами, полученными Дорфманом, можно с помощью таблицы в прил. 3. Фрагмент данной таблицы приведен ниже (табл. 4.4). Первый столбец содержит вероятности дефектности объекта, которые Дорфман рассматривал в своей статье [4]. Затем для ряда значений C_n / C_1 получены оптимальные размеры групп для проведения групповой процедуры n и соответствующие значения относительной стоимости проверок C .

Таблица 4.4. Сравнение результатов, полученных с помощью классической процедуры Дорфмана, с результатами новой процедуры

| p | $C_n / C_1 = 1$ | | $C_n / C_1 = 1,5$ | | $C_n / C_1 = 2$ | | $C_n / C_1 = 3$ | | $C_n / C_1 = 5$ | | $C_n / C_1 = 10$ | |
|------|-----------------|------|-------------------|------|-----------------|------|-----------------|------|-----------------|------|------------------|------|
| | n | C | n | C | n | C | n | C | n | C | n | C |
| 0,01 | 11 | 0,20 | 13 | 0,24 | 15 | 0,27 | 19 | 0,33 | 25 | 0,42 | 38 | 0,58 |
| 0,02 | 8 | 0,27 | 9 | 0,33 | 11 | 0,38 | 14 | 0,46 | 19 | 0,58 | 30 | 0,79 |
| 0,03 | 6 | 0,33 | 8 | 0,40 | 9 | 0,46 | 12 | 0,56 | 16 | 0,70 | 28 | 0,93 |
| 0,04 | 6 | 0,38 | 7 | 0,46 | 8 | 0,53 | 11 | 0,63 | 15 | 0,79 | 1 | 1,00 |
| 0,05 | 5 | 0,43 | 6 | 0,51 | 8 | 0,59 | 10 | 0,70 | 14 | 0,87 | 1 | 1,00 |
| 0,06 | 5 | 0,47 | 6 | 0,56 | 7 | 0,64 | 9 | 0,76 | 14 | 0,94 | 1 | 1,00 |
| 0,07 | 4 | 0,50 | 6 | 0,60 | 7 | 0,68 | 9 | 0,81 | 14 | 1,00 | 1 | 1,00 |
| 0,08 | 4 | 0,53 | 5 | 0,64 | 6 | 0,73 | 9 | 0,86 | 1 | 1,00 | 1 | 1,00 |
| 0,09 | 4 | 0,56 | 5 | 0,68 | 6 | 0,77 | 8 | 0,90 | 1 | 1,00 | 1 | 1,00 |
| 0,10 | 4 | 0,59 | 5 | 0,71 | 6 | 0,80 | 8 | 0,94 | 1 | 1,00 | 1 | 1,00 |
| 0,12 | 4 | 0,65 | 5 | 0,77 | 6 | 0,87 | 1 | 1,00 | 1 | 1,00 | 1 | 1,00 |
| 0,13 | 3 | 0,67 | 5 | 0,80 | 6 | 0,90 | 1 | 1,00 | 1 | 1,00 | 1 | 1,00 |
| 0,15 | 3 | 0,72 | 4 | 0,85 | 6 | 0,96 | 1 | 1,00 | 1 | 1,00 | 1 | 1,00 |
| 0,20 | 3 | 0,82 | 4 | 0,97 | 1 | 1,00 | 1 | 1,00 | 1 | 1,00 | 1 | 1,00 |
| 0,25 | 3 | 0,91 | 1 | 1,00 | 1 | 1,00 | 1 | 1,00 | 1 | 1,00 | 1 | 1,00 |
| 0,30 | 3 | 0,99 | 1 | 1,00 | 1 | 1,00 | 1 | 1,00 | 1 | 1,00 | 1 | 1,00 |

Данные в табл. 4.4 свидетельствуют о том, что при высокой вероятности дефектности объекта $p \geq 0.25$ применять процедуру Дорфмана с учетом стоимости объединения объектов в группы невыгодно даже при небольшой разнице между стоимостью проверки одного объекта и стоимостью проверки группы. Невыгодно применять данную процедуру и при больших значениях C_n / C_1 , т. е. в ситуациях, когда стоимость проверки группы значительно превышает стоимость проверки одного объекта. О неэффективности групповой процедуры говорят оптимальный объем группы n и относительная стоимость проверок C , равные 1. В таких ситуациях следует проверять объекты группами по одному объекту, т. е.

индивидуально, и тогда среднее количество проверок, которое будет приходиться на один объект, составит 1.

Оценить поведение функции стоимости проверок в расчете на 100 объектов при различных значениях параметров можно по графикам на рис. 4.2 и 4.8. Графики построены для представленных в табл. 4.4 значений C_n / C_1 и некоторых значений p : 0.3, 0.25, 0.2, 0.15, 0.1, 0.05 и 0.01. При построении использованы один масштаб и одинаковый диапазон осей для n и для функции стоимости на 100 объектов, что позволяет увидеть, не только как меняется форма графиков при одной и той же вероятности дефектности объекта p , но и как меняется форма графиков при различных вероятностях дефектности объекта.

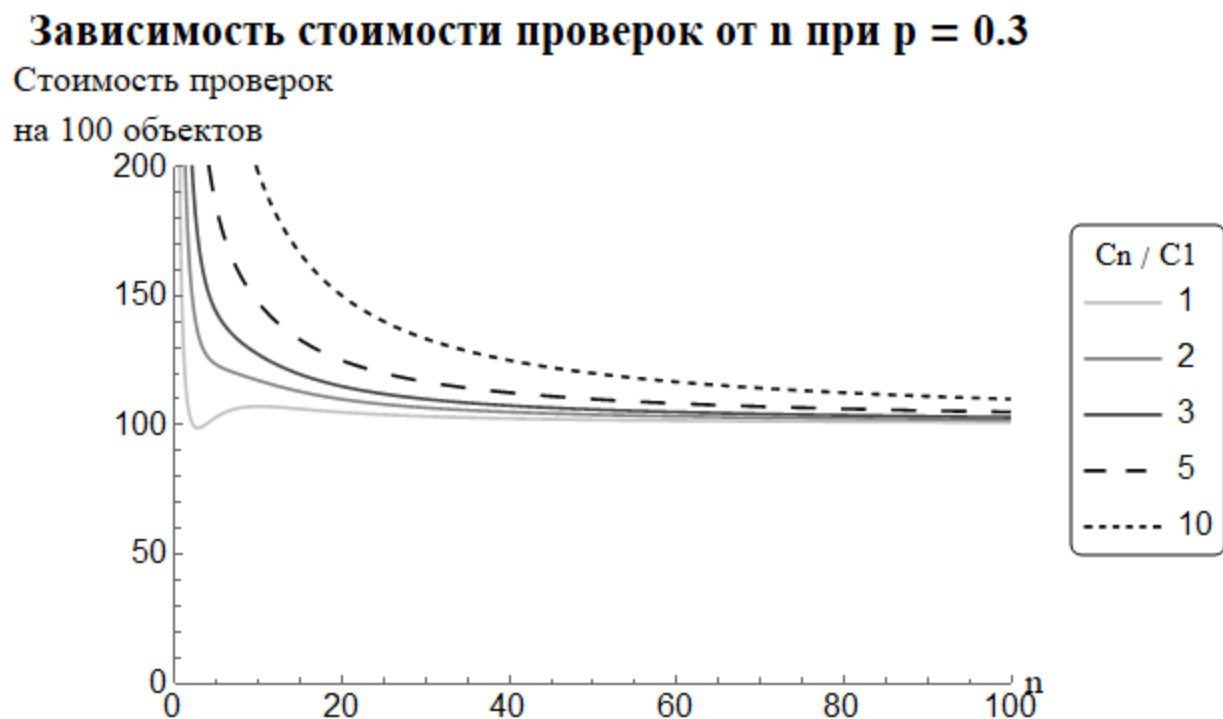


Рисунок 4.2. Зависимость стоимости проверок от объема группы для различных значений соотношения C_n / C_1 и при $p = 0.3$

Зависимость стоимости проверок от n при $p = 0.25$

Стоимость проверок

на 100 объектов

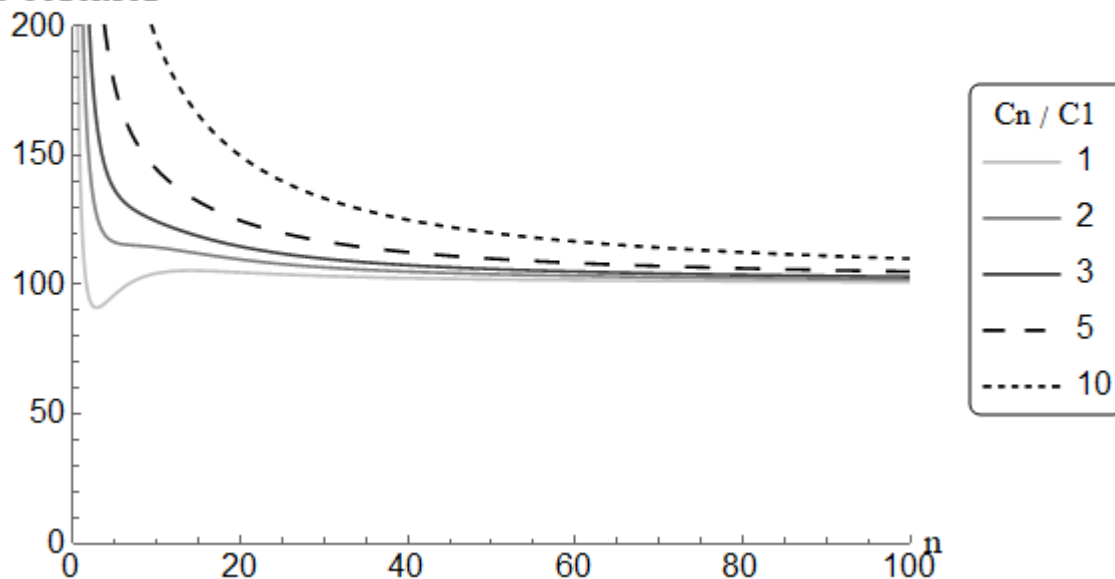


Рис. 4.3. Зависимость стоимости проверок от объема группы для различных значений соотношения C_n / C_1 и при $p = 0.25$

Зависимость стоимости проверок от n при $p = 0.2$

Стоимость проверок

на 100 объектов

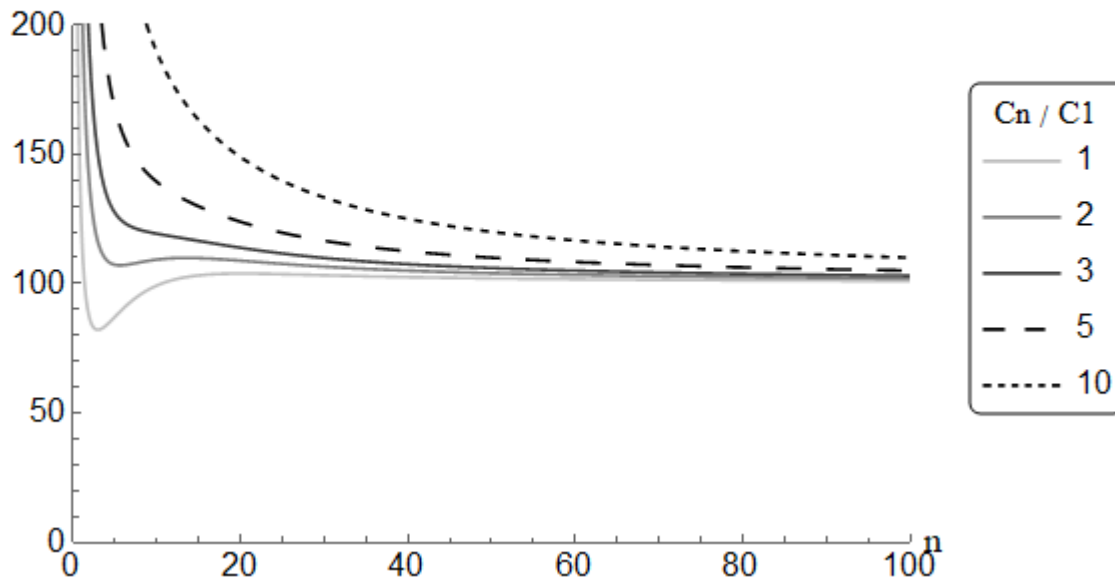


Рис. 4.4. Зависимость стоимости проверок от объема группы для различных значений соотношения C_n / C_1 и при $p = 0.2$

Зависимость стоимости проверок от n при $p = 0.15$

Стоимость проверок

на 100 объектов

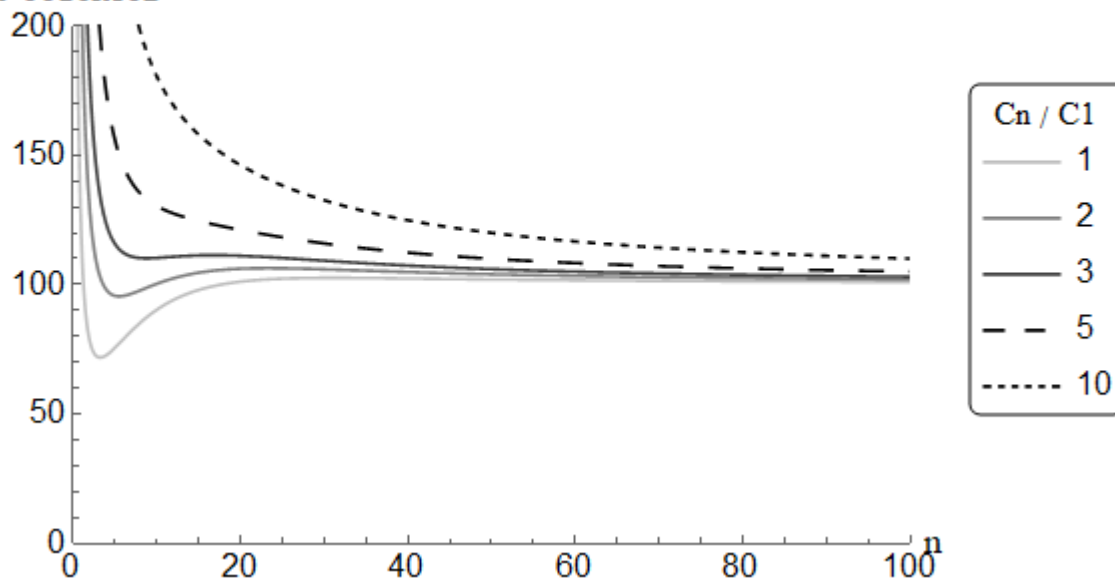


Рис. 4.5. Зависимость стоимости проверок от объема группы для различных значений соотношения C_n / C_1 и при $p = 0.15$

Зависимость стоимости проверок от n при $p = 0.1$

Стоимость проверок

на 100 объектов

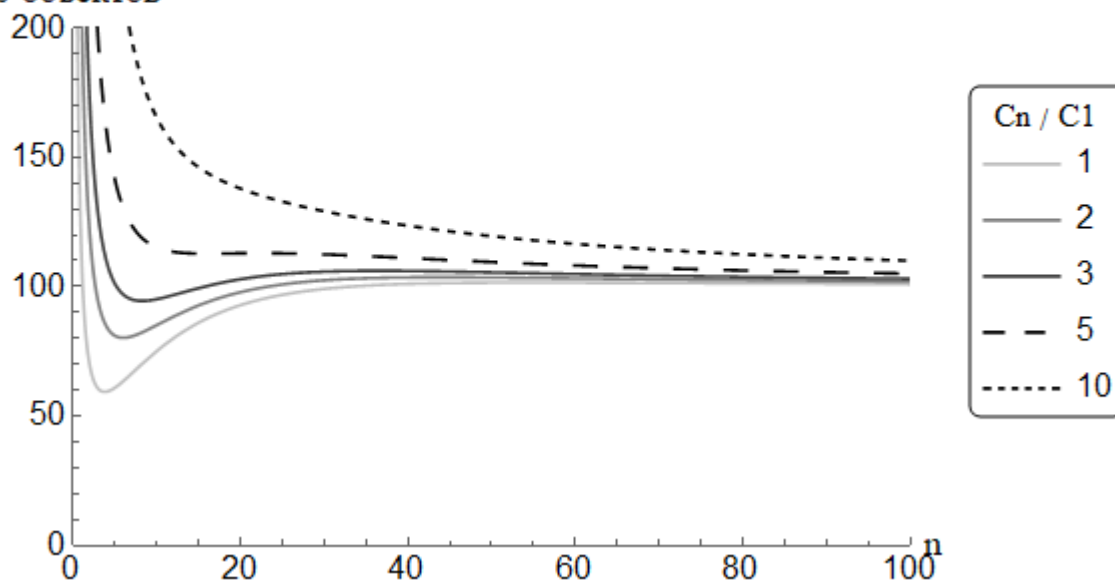


Рис. 4.6. Зависимость стоимости проверок от объема группы для различных значений соотношения C_n / C_1 и при $p = 0.1$

Зависимость стоимости проверок от n при $p = 0.05$

Стоимость проверок

на 100 объектов

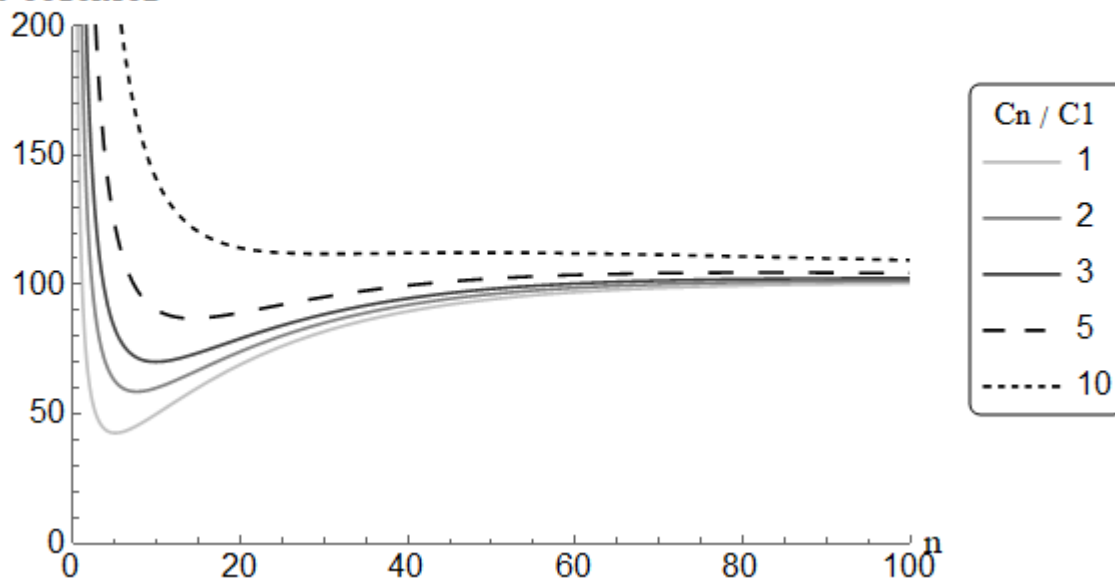


Рис. 4.7. Зависимость стоимости проверок от объема группы для различных значений соотношения C_n / C_1 и при $p = 0.05$

Зависимость стоимости проверок от n при $p = 0.01$

Стоимость проверок

на 100 объектов

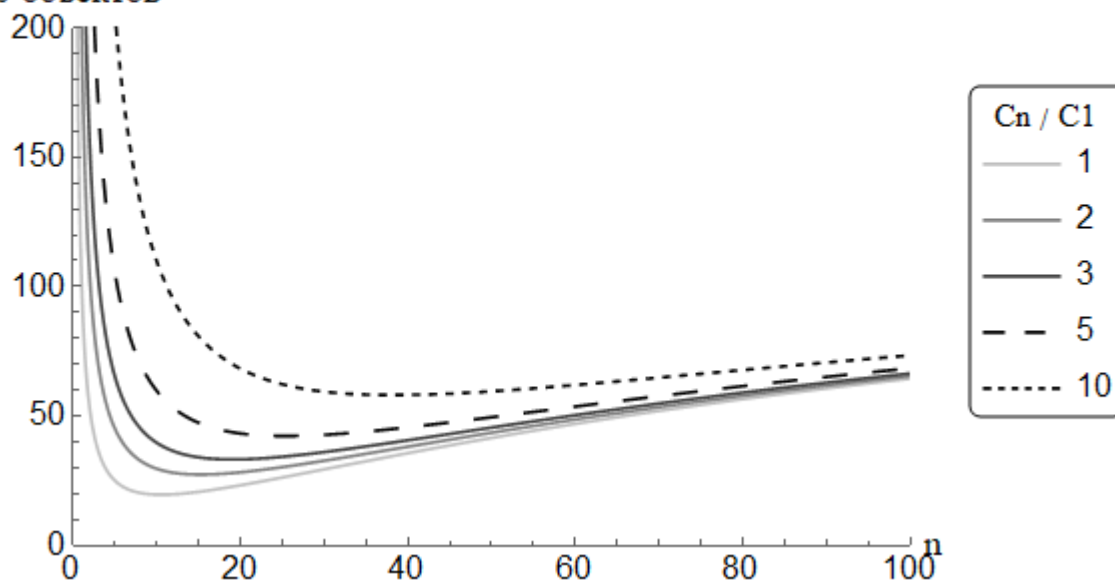


Рис. 4.8. Зависимость стоимости проверок от объема группы для различных значений соотношения C_n / C_1 и при $p = 0.01$

Посмотрев на графики рис. 4.2. и 4.8, в первую очередь можно заметить, что функция стоимости проверок (4.1) для классической процедуры Дорфмана, не учитывающей затраты на объединение объектов в группы, всегда имеет точку наименьшего значения внутри отрезка $[1, 100]$, которая является оптимальным объемом группы, позволяющим сэкономить на обследовании совокупности.

Форма функции стоимости при $C_n / C_1 > 1$ для высоких вероятностей дефектности объекта такой точки не имеет, но с уменьшением вероятности p на каждом графике начинает проявляться локальный экстремум, который впоследствии становится точкой наименьшего значения функции стоимости на отрезке $[1, 100]$. Локальный экстремум не сразу становится точкой наименьшего значения, промежуточная ситуация, когда локальный экстремум есть, а наименьшее значение достигается на правом конце отрезка, была подробно рассмотрена в параграфе 4.2. Чем выше значение C_n / C_1 , тем дольше проявляется точка наименьшего значения внутри отрезка $[1, 100]$ с уменьшением p .

4.5. Условие оправданности применения метода групповых процедур проверок

Условие, выполнение которого свидетельствует о том, что применение групповых процедур проверок для обследования совокупности объектов окажется более выгодным, чем индивидуальная проверка этих объектов, определяется величиной соотношения стоимости проверки группы объектов к стоимости проверки одного объекта C_n / C_1 . Найти граничное значение данного соотношения, при котором стоимости групповых и индивидуальных проверок будут одинаковы, можно из следующего равенства:

$$\frac{N}{n} C_n + \frac{N}{n} p^n C_1 = N \cdot C_1, \quad (4.5)$$

где левая часть – стоимость групповых проверок, а правая – стоимость индивидуальных проверок.

Графически равенство (4.5) выглядит как точка пересечения функции стоимости (4.1) с функцией стоимости индивидуальных проверок, равной величине $C_1 N$, не зависящей от объема группы n . График на рис. 4.9 построен для объема совокупности $N = 10\,000$, вероятности дефектности объекта $p = 0.001$, стоимости проверки одного объекта $C_1 = 1$ и стоимости проверки группы объектов $C_n = 367.7$.

Стоимости групповой и индивидуальной проверок

Стоимость проверок

на 10000 объектов

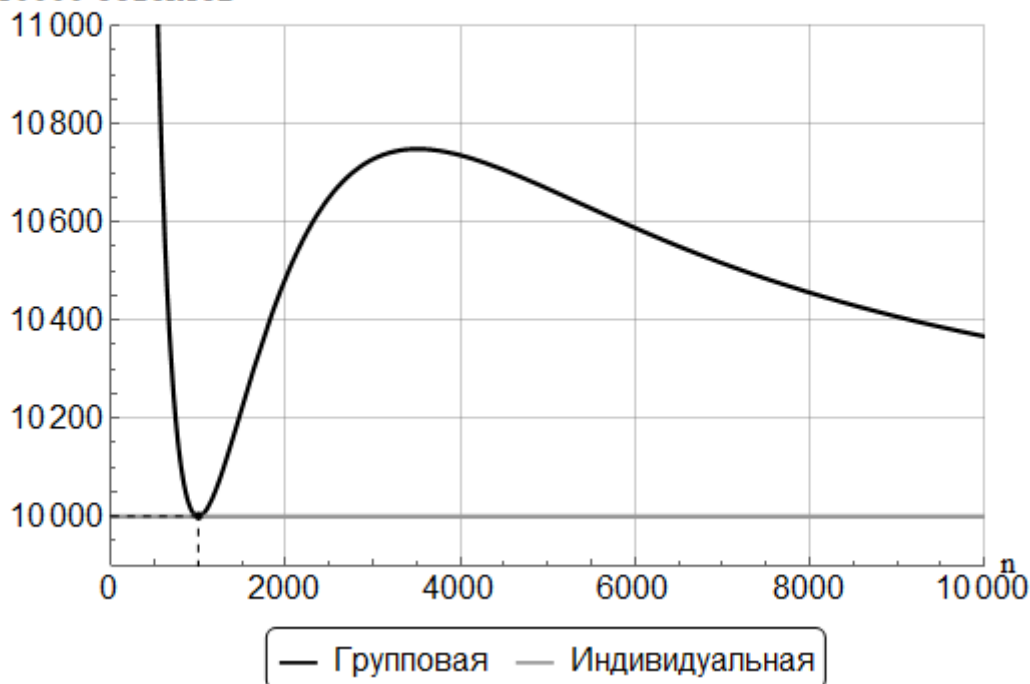


Рис. 4.9. Стоимости групповой и индивидуальной проверки (равенство (4.5) при $N = 10\,000$, $p = 0.001$, $C_1 = 1$ и $C_n = 367.7$)

Наличие у функции стоимости (4.1) локального минимума на отрезке $[1, N]$ является желаемым вариантом, если локальный минимум одновременно является наименьшим значением функции (4.1) на отрезке $[1, N]$. При таком варианте применение групповых процедур до сих пор оказывалось дешевле индивидуальных проверок. В случае, представленном на рис. 4.9, локальный минимум удовлетворяет описанным условиям, однако экстремальное значение функции стоимости (4.1) хотя и является наименьшим значением на отрезке $[1, N]$, но оно равно стоимости процесса индивидуальной проверки. Получается, что в рассмотренной ситуации выявление всех дефектных объектов в совокупности N не может обойтись дешевле, чем $C_1 \cdot N$ условных единиц.

Несмотря на то что выразить из равенства (4.5) соотношение C_n / C_1 представляется легко выполнимой задачей, полученное выражение

$$\frac{C_n}{C_1} = n \cdot (1 - p)^n$$

вряд ли можно считать ее решением, так как n – оптимальный размер группы, получаемый в результате решения задачи (4.3), который зависит в том числе и от значения соотношения C_n / C_1 .

Для рассмотренных в пп. 4.3 и 4.4 вероятностей дефектности объекта p при $N = 100\,000$ и $C_1 = 1$ были найдены приближенные решения уравнения (4.5). Полученные граничные значения соотношения C_n / C_1 определяют, во сколько раз стоимость проверки группы объектов C_n может превосходить стоимость проверки одного объекта C_1 , чтобы применение групповых процедур проверок было дешевле, чем индивидуальная проверка объектов. Если при конкретной вероятности дефектности объекта величина соотношения C_n / C_1 меньше значения, указанного в таблице, выгоднее проверять совокупность с помощью групповых процедур. Если же величина C_n / C_1 близка к значению в таблице или равна ему, стоит всерьез задуматься о предпочтении индивидуальной проверки объектов, так как полученный выигрыш в стоимости не оправдает усилий по организации объектов в группы и последующего разделения группы обратно на отдельные объекты.

Результаты расчетов приведены в табл. 4.5. Для каждой вероятности дефектности объекта p представлен также оптимальный объем группы, соответствующий граничному значению соотношения C_n / C_1 . Следует отметить, что размер совокупности N в конечном итоге влияет только на количество проверок и стоимость процесса проверки всей совокупности, поэтому результаты в табл. 4.5 справедливы для совокупности любого размера.

Табл. 4.5. Граничные значения соотношения C_n / C_1 при $N = 100000$, $C_1 = 1$

| p | Опт. объем группы | C_n / C_1 | p | Опт. объем группы | C_n / C_1 |
|-------|-------------------|-------------|------|-------------------|-------------|
| 0.001 | 1000 | 367.7000 | 0.08 | 12 | 4.4120 |
| 0.005 | 200 | 73.3920 | 0.09 | 11 | 3.8981 |
| 0.01 | 100 | 36.6035 | 0.10 | 10 | 3.4868 |
| 0.02 | 50 | 18.2085 | 0.12 | 8 | 2.8771 |
| 0.03 | 33 | 12.0777 | 0.13 | 7 | 2.6408 |
| 0.04 | 25 | 9.0100 | 0.15 | 6 | 2.2629 |
| 0.05 | 20 | 7.1698 | 0.20 | 5 | 1.6385 |
| 0.06 | 16 | 5.9452 | 0.25 | 4 | 1.2657 |
| 0.07 | 14 | 5.0687 | 0.30 | 3 | 1.0291 |

Данные, представленные в табл. 4.5, являются очередным подтверждением того, что эффективность применения метода групповых процедур проверок выше при низких вероятностях дефектности объекта. Иными словами, чем меньше вероятность дефектности объекта, тем дороже может быть стоимость

проверки группы объектов по сравнению со стоимостью проверки одного объекта, чтобы групповые процедуры проверок оставались более эффективными, чем индивидуальные проверки.

Легко заметить, что большему значению соотношения C_n / C_1 соответствует больший объем группы. Данную закономерность можно объяснить тем, что проверять группу объектов становится настолько дорого, что количество групповых тестов уменьшается с ростом C_n / C_1 и в конечном итоге приводится к минимально возможному значению: единице.

Использование достаточно большого объема группы для групповых процедур часто приводит к необходимости проверять индивидуально все объекты дефектной группы, но в рассматриваемой модели такой вариант все равно оказывается дешевле, чем вариант с меньшим объемом группы и, следовательно, большим количеством групповых тестов. Цена ошибки – получение дефектной группы, для которой необходимо индивидуальное тестирование после понесенных значительных затрат на объединение объектов в группы, – слишком высока, поэтому предпочтительнее ошибаться как можно реже.

ГЛАВА 5. МОДЕЛИРОВАНИЕ ПРОЦЕДУРЫ СТЕРРЕТТА С ОПТИМАЛЬНЫМ ОБЪЕМОМ ГРУППЫ, МИНИМИЗИРУЮЩИМ СТОИМОСТЬ ПРОВЕРОК

5.1. Построение модели

Данный параграф пособия посвящен построению математической модели процедуры Стерретта [21], в которой оптимальный объем группы определяется из соображений минимизации стоимости проверок.

Далее будут использованы следующие обозначения из статьи [21]:

- p – вероятность дефектности объекта, соответственно $q = 1 - p$ – вероятность годности объекта;
- N – размер совокупности;
- k – размер выборки для проверки совокупности N ;
- $\text{Pr}_k(i)$ – вероятность того, что в выборке k содержится i дефектных объектов;
- $E_k(i)$ – среднее количество проверок, необходимое для обнаружения в выборке k всех i дефектных объектов.

Аналогично классической процедуре Дорфмана при использовании процедуры Стерретта оптимальный объем группы определяется из решения следующей задачи:

$$\begin{cases} E(N, k, p) \rightarrow \min \\ k \in \mathbb{N} \end{cases}, \quad (5.1)$$

где $E(N, k, p) = \frac{N}{k} \sum_{i=0}^k \{\text{Pr}_k(i) E_k(i)\}$ – среднее количество проверок для обнаружения всех дефектных объектов в совокупности N .

Новая модель использует размер группы, оптимальный с точки зрения стоимости процесса проверки:

$$\begin{cases} \text{Cost}(N, k, p) \rightarrow \min \\ k \in \mathbb{N} \end{cases}. \quad (5.2)$$

Поэтому необходимо найти стоимость процесса проверки в процедуре Стерретта с учетом затрат на объединение объектов в группы $\text{Cost}(N, k, p)$.

При нахождении стоимости процесса проверки может быть использована такая же логика, какая была использована Стерреттом при получении формулы для среднего количества проверок:

$$Cost(N, k, p) = \frac{N}{k} \sum_{i=0}^k \{Pr_k(i) Cost_k(i)\}. \quad (5.3)$$

Здесь $Pr_k(i) = C_k^i p^i q^{(k-i)}$, как и в процедуре Стерретта, а выражение для $Cost_k(i)$ – средней стоимости проверок для выявления всех i дефектных объектов из выборки объема k – требуется получить.

Прежде всего следует учесть, что в процедуре Стерретта групповые тесты проводятся не только для выборок объема k , но также и для их подвыборок различного размера от 2 до $k-1$, поэтому необходимо иметь набор цен за групповые тесты. Пусть C_l – стоимость проверки одного объекта, C_n – стоимость проверки группы из n объектов для $n = \overline{2, k}$, причем $0 < C_1 < C_2 < \dots < C_{k-1} < C_k$.

Если в выборке объема k не имеется дефектных объектов, потребуется одна проверка, чтобы подтвердить годность всех k объектов. Стоимость этой проверки составляет:

$$Cost_k(0) = C_k. \quad (5.4)$$

Далее для получения выражения для $Cost_k(i)$ достаточно внести небольшие изменения в формулу (1.5): добавить цены за индивидуальные и групповые тесты. Формула (1.5) имеет вид

$$E_k(n) = 1 + \frac{n}{k} \{1 + E_{k-1}(n-1)\} + \sum_{j=1}^{k-n} \left[\left(\prod_{i=1}^j \frac{k-(i+n-1)}{k-(i-1)} \right) \frac{n}{k-j} \{(j+1) + E_{k-[j+1]}(n-1)\} \right]. \quad (1.5)$$

Первое слагаемое в правой части (1.5) представляет собой начальный групповой тест, поэтому его следует умножить на стоимость C_k проверки группы объема k . Множитель n/k в следующем слагаемом – это вероятность того, что первый проверенный элемент выборки дефектный. Первое слагаемое множителя $\{1 + E_{k-1}(n-1)\}$ – это тест, необходимый для определения дефектного элемента при первом испытании, значит, он должен быть умножен на стоимость проверки одного элемента C_1 . Второе слагаемое множителя $\{1 + E_{k-1}(n-1)\}$ – среднее количество проверок, необходимое для нахождения $(n-1)$ дефектных объектов среди оставшихся $k-1$ объектов выборки, – следует заменить на среднюю стоимость этих проверок $Cost_{k-1}(n-1)$.

Последнее слагаемое формулы (1.5) отражает все остальные варианты обнаружить первый дефектный объект, когда результат одной или нескольких первых индивидуальных проверок положительной группы отрицательный. Вероят-

ность того, что первые j объектов годные, а $(j+1)$ -й проверяемый объект дефектный, равна $\prod_{i=1}^j \left[\frac{k-(i+n-1)}{k-(i-1)} \right] \cdot \frac{n}{k-j}$. Для определения средней стоимости проверок данную вероятность необходимо умножить на стоимость $(j+1)$ индивидуальных проверок и на среднюю стоимость проверок, предназначенных для нахождения оставшихся $(n-1)$ дефектных среди $k-(j+1)$ объектов.

Таким образом, средняя стоимость процедуры Стерретта с объемом группы k для обнаружения всех дефектных объектов в совокупности N с процентом дефектных p составляет

$$\begin{aligned} Cost_k(n) = & C_k + \frac{n}{k} \{ C_1 + Cost_{k-1}(n-1) \} + \\ & + \sum_{j=1}^{k-n} \left[\left(\prod_{i=1}^j \frac{k-(i+n-1)}{k-(i-1)} \right) \frac{n}{k-j} \{ (j+1) \cdot C_1 + Cost_{k-[j+1]}(n-1) \} \right]. \end{aligned} \quad (5.5)$$

Оптимальный объем группы для процедуры Стерретта с учетом цен за индивидуальный и групповые тесты определяется в результате решения системы (5.2) при подстановке в нее полученных выражений (5.3) и (5.5).

Процедура Стерретта представляет собой оптимизированную процедуру Дорфмана для малых вероятностей дефектности объекта, поэтому учет стоимости объединения объектов в группы в данной процедуре имеет большую значимость. Выигрыш в стоимости по сравнению с процедурой Дорфмана может быть не достигнут, так как увеличение объема группы подразумевает увеличение затрат на создание этой группы.

5.2. Описание алгоритма и программы

Моделирование процедуры Стерретта с учетом цен на индивидуальные и групповые проверки было осуществлено с помощью программы, написанной на языке Python. Код программы представлен в прил. 4.

Можно заметить, что формула (5.5) для вычисления средней стоимости проверок для выборки k при условии, что в выборке есть n дефектных объектов, рекуррентная, поэтому ее удобно представить в виде рекурсивной функции, в которой рекурсия ведется по двум переменным: k и n . Обе переменные конечны, следовательно, функцию можно построить таким образом, чтобы за конечное число рекурсивных вызовов прийти к одному из частных случаев, которые вычисляются без рекурсии. Выражение (5.4) является одним из таких частных случаев. Формула (5.5), дополненная условиями остановки рекурсивного спуска, имеет следующий вид:

$$\left\{ \begin{array}{l} Cost_k(n) = C_k + \frac{n}{k} \{C_1 + Cost_{k-1}(n-1)\} + \\ + \sum_{j=1}^{k-n} \left[\left(\prod_{i=1}^j \frac{k-(i+n-1)}{k-(i-1)} \right) \frac{n}{k-j} \{ (j+1) \cdot C_1 + Cost_{k-[j+1]}(n-1) \} \right] \\ Cost_k(0) = C_k \\ Cost_0(0) = 0 \end{array} \right. \quad (5.6)$$

Практически любой рекурсивный алгоритм можно переписать с использованием цикла. Рекурсивный алгоритм на языке Python выполняется значительно дольше, чем аналогичный алгоритм с циклом, но без рекурсии, поэтому вычисления в программе реализованы с помощью цикла. Значения функции стоимости (5.6), получаемые на каждом шаге (для каждого набора k и n), сохраняются для вычисления значений данной функции на последующих шагах.

Важным моментом в моделировании процесса обследования совокупности с помощью модифицированной процедуры Стерретта является подбор необходимых цен для групповых тестов. Так как заранее сложно сказать, каким будет оптимальный с точки зрения стоимости объем группы, цены за групповые тесты удобно задать в виде функции, зависящей от объема проверяемой группы n , $n \in [1, k]$. В данной работе рассмотрены четыре различные функции:

$$C_1(n) = C_1 \cdot \left(1 + \frac{1}{2}(n-1) \right), \quad n = 1, 2, \dots, k; \quad (5.7)$$

$$C_2(n) = C_1 \cdot \left(1 + \frac{1}{4}(n-1) \right), \quad n = 1, 2, \dots, k; \quad (5.8)$$

$$C_3(n) = \begin{cases} C_3(n-1) + \frac{1}{n}, & n = 2, \dots, k; \\ C_1, & n = 1 \end{cases} \quad (5.9)$$

$$C_4(n) = \begin{cases} C_4(n-1) + \frac{1}{F_{n+1}}, & F_n = \frac{1}{\sqrt{5}} \cdot \frac{(1+\sqrt{5})^n - (1-\sqrt{5})^n}{2^n}, \quad n = 2, \dots, k \\ C_1, & n = 1 \end{cases} \quad (5.10)$$

Функции (5.7) и (5.8) являются линейными относительно n , а функции (5.9) и (5.10) имеют более сложную зависимость от n . F_n в формуле (5.10) – n -й член последовательности Фибоначчи; C_1 – стоимость проверки одного объекта.

Изучить поведение представленных функций (5.7) – (5.10) и сравнить их между собой можно с помощью графиков на рис. 5.1.

Функции стоимости проверки группы

Стоимость проверки
группы

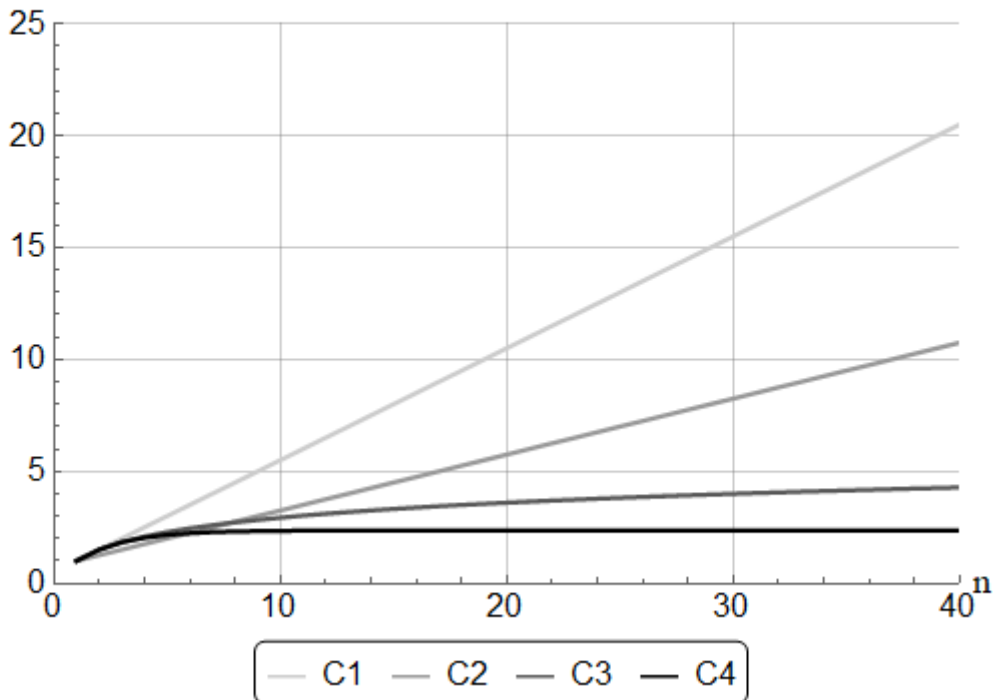


Рис. 5.1. Функции, задающие цены за групповые тесты

Рассмотрены небольшие значения n из промежутка $[1, 40]$, так как тенденция, которая просматривается на графике, сохраняется и при увеличении n .

Таким образом, цены за индивидуальные и групповые проверки задаются в программе посредством указания стоимости проверки одного объекта C_1 и функции, в соответствии с которой формируется набор стоимостей проверки групп любых возможных размеров.

Аналогично тому, как это было сделано для функции стоимости проверок (4.1) в случае использования процедуры Дорфмана, необходимо исследовать поведение функции стоимости (5.3), полученной для процедуры Стерретта. Графики на рис. 5.2 и 5.3 построены для функции стоимости (5.3) при $N = 100$ и нескольких различных значениях вероятности дефектности объекта p ; также использовались две функции, задающие цены за групповые тесты, — C_2 (5.8) и C_4 (5.10) соответственно.

Зависимость стоимости проверок от n

Стоимость проверок
на 100 объектов

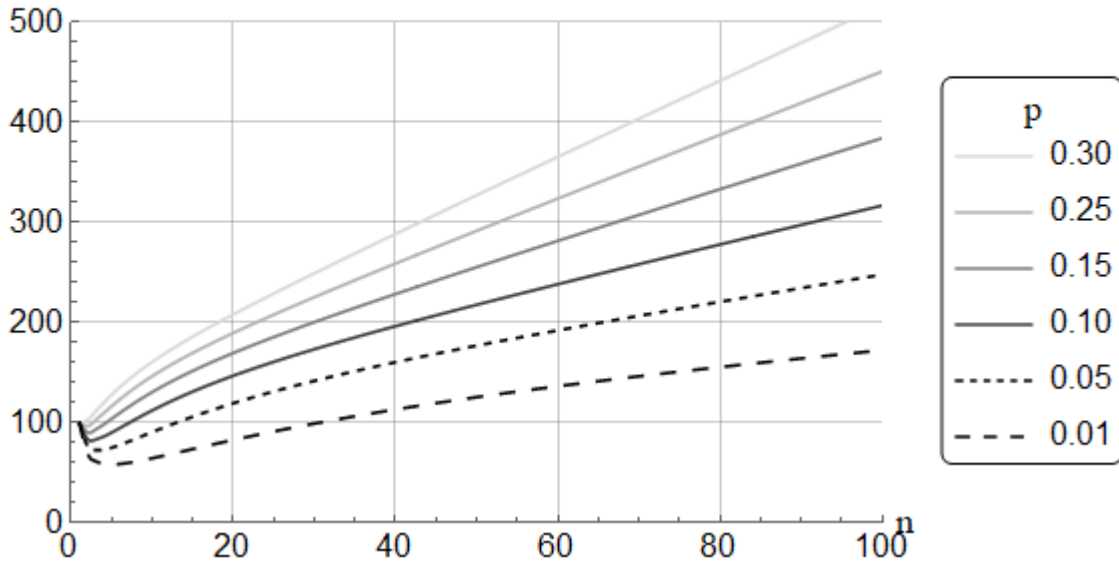


Рис. 5.2. Зависимость стоимости проверок от объема группы при $N = 100$, $C_n(n) = C_2(n)$

Зависимость стоимости проверок от n

Стоимость проверок
на 100 объектов

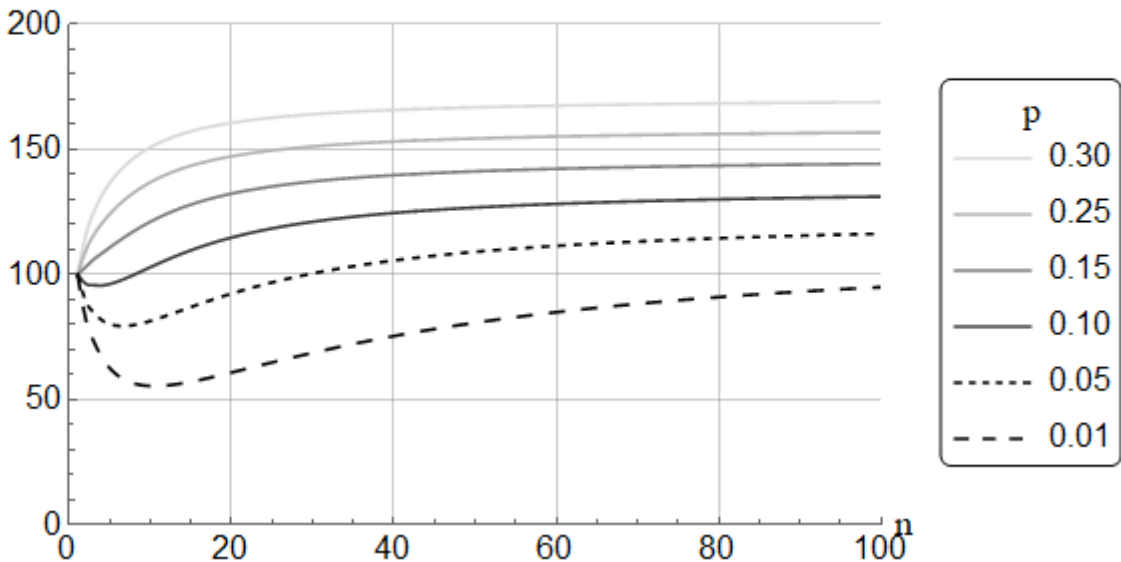


Рис. 5.3. Зависимость стоимости проверок от объема группы при $N = 100$, $C_n(n) = C_4(n)$

Поиск наименьшего значения в данном случае оказывается более легкой задачей, так как для рассматриваемых значений вероятности дефектности объекта возможны два варианта. Первый – функция (5.3) монотонно возрастает с увеличением объема группы n , и наименьшее значение находится на левой гра-

нице отрезка: при $n = 1$. Это наблюдается для больших вероятностей дефектности объекта, и в таких ситуациях следует предпочесть индивидуальную проверку. Второй – наименьшее значение функции (5.3) соответствует ее локальному минимуму на отрезке $[1, N]$, и тогда применение групповой проверки окажется более выгодным. Переход от первого варианта ко второму происходит плавно с уменьшением p . Промежуточное состояние, когда стоимость индивидуальных проверок с некоторой точностью совпадает со стоимостью групповых, возможна только для $n = 2$, что не представляет особого интереса.

Программа принимает на вход следующие данные:

- вероятность дефектности объекта p ,
- объем популяции N ,
- стоимость проверки одного объекта C_1 ,
- стоимости проверки групп объектов $C_2, C_3, \dots, C_{k-1}, C_k$.

Результатом работы программы являются следующие показатели:

- Оптимальный объем группы, минимизирующий стоимость проверки для процедуры Стерретта.
- Оптимальный объем группы, минимизирующий количество проверок для процедуры Стерретта.
- Стоимость процесса проверки 100 объектов при использовании процедуры Стерретта с оптимальным объемом группы, минимизирующим стоимость процесса проверки.
- Стоимость процесса проверки 100 объектов при использовании процедуры Стерретта с оптимальным объемом группы, минимизирующим количество проверок.

5.3. Сравнение с результатами Стерретта

Стерретт опубликовал в своей статье оптимальные объемы групп для своей процедуры и соответствующие значения количества проверок, приходящихся на 100 объектов, для некоторых вероятностей дефектности объекта. Те же самые характеристики были получены с помощью программы для новой процедуры с оптимальным объемом, минимизирующим стоимость процесса проверки. Кроме того, с помощью программы по аналогии с тем, как это было сделано для процедуры Дорфмана в параграфе 4.4, были вычислены стоимости проверок для обеих процедур. Стоимость проверки одного объекта C_1 также считалась равной единице.

Рассмотрены обычная процедура Стерратта без учета стоимости объединения объектов в группы и четыре варианта новой процедуры, которые различаются функциями зависимости цены за групповой тест от размера группы: (5.7)

— (5.10). Данные для сравнения процедур представлены в табл. 5.1. Столбец p содержит вероятности дефектности объекта, k – оптимальный объем группы, E – среднее количество проверок на 100 объектов и $Cost$ – средняя стоимость проверок 100 объектов.

Табл. 5.1а. Сравнение результатов, полученных с помощью процедуры Стерретта, с результатами новой процедуры

| p | Процедура Стерретта | | | Новая процедура с C_1 | | | Новая процедура с C_2 | | |
|-------|---------------------|-----|--------|-------------------------|-----|--------|-------------------------|-----|--------|
| | k | E | $Cost$ | K | E | $Cost$ | k | E | $Cost$ |
| 0.38 | 2 | 100 | 100.00 | 1 | 100 | 100.00 | 1 | 100 | 100.00 |
| 0.35 | 2 | 97 | 96.38 | 1 | 100 | 100.00 | 1 | 100 | 100.00 |
| 0.32 | 2 | 93 | 92.88 | 1 | 100 | 100.00 | 1 | 100 | 100.00 |
| 0.3 | 2 | 91 | 90.50 | 1 | 100 | 100.00 | 1 | 100 | 100.00 |
| 0.27 | 2 | 87 | 86.86 | 1 | 100 | 100.00 | 2 | 87 | 99.36 |
| 0.26 | 3 | 86 | 85.57 | 1 | 100 | 100.00 | 2 | 86 | 98.12 |
| 0.25 | 3 | 84 | 83.85 | 1 | 100 | 100.00 | 2 | 85 | 96.88 |
| 0.23 | 3 | 81 | 80.35 | 1 | 100 | 100.00 | 2 | 82 | 94.36 |
| 0.2 | 3 | 75 | 74.93 | 1 | 100 | 100.00 | 2 | 78 | 90.50 |
| 0.15 | 4 | 65 | 64.78 | 2 | 72 | 96.38 | 3 | 66 | 83.36 |
| 0.14 | 4 | 63 | 62.43 | 2 | 71 | 95.02 | 3 | 64 | 81.31 |
| 0.13 | 4 | 61 | 60.04 | 2 | 69 | 93.66 | 3 | 62 | 79.24 |
| 0.12 | 4 | 58 | 57.61 | 2 | 68 | 92.28 | 3 | 60 | 77.14 |
| 0.11 | 4 | 56 | 55.14 | 2 | 66 | 90.90 | 3 | 58 | 75.01 |
| 0.1 | 5 | 53 | 52.29 | 2 | 65 | 89.50 | 3 | 56 | 72.87 |
| 0.09 | 5 | 50 | 49.37 | 2 | 64 | 88.10 | 4 | 51 | 70.50 |
| 0.08 | 5 | 47 | 46.39 | 3 | 52 | 85.83 | 4 | 48 | 67.71 |
| 0.07 | 6 | 44 | 43.17 | 3 | 50 | 83.52 | 4 | 45 | 64.88 |
| 0.06 | 6 | 40 | 39.69 | 3 | 47 | 81.19 | 4 | 43 | 62.00 |
| 0.05 | 7 | 36 | 35.98 | 4 | 40 | 78.77 | 5 | 38 | 58.54 |
| 0.04 | 8 | 32 | 31.92 | 4 | 37 | 75.61 | 6 | 33 | 54.94 |
| 0.03 | 9 | 28 | 27.31 | 5 | 31 | 72.25 | 6 | 29 | 50.74 |
| 0.02 | 11 | 22 | 21.96 | 6 | 25 | 68.10 | 8 | 23 | 45.78 |
| 0.01 | 15 | 16 | 15.17 | 8 | 18 | 62.71 | 11 | 16 | 39.46 |
| 0.007 | 17 | 13 | 12.57 | 10 | 15 | 60.59 | 13 | 14 | 37.02 |
| 0.005 | 21 | 11 | 10.54 | 12 | 12 | 58.92 | 16 | 11 | 35.09 |
| 0.003 | 26 | 9 | 8.07 | 15 | 10 | 56.87 | 20 | 9 | 32.75 |
| 0.001 | 45 | 5 | 4.58 | 26 | 6 | 53.93 | 35 | 5 | 29.42 |

Табл. 5.1b. Сравнение результатов, полученных с помощью процедуры Стерретта, с результатами новой процедуры

| p | Процедура Стерретта | | | Новая процедура с C_3 | | | Новая процедура с C_4 | | |
|-------|---------------------|-----|--------|-------------------------|-----|--------|-------------------------|-----|--------|
| | k | E | $Cost$ | K | E | $Cost$ | k | E | $Cost$ |
| 0.38 | 2 | 100 | 100.00 | 1 | 100 | 100.00 | 1 | 100 | 100.00 |
| 0.35 | 2 | 97 | 96.38 | 1 | 100 | 100.00 | 1 | 100 | 100.00 |
| 0.32 | 2 | 93 | 92.88 | 1 | 100 | 100.00 | 1 | 100 | 100.00 |
| 0.3 | 2 | 91 | 90.50 | 1 | 100 | 100.00 | 1 | 100 | 100.00 |
| 0.27 | 2 | 87 | 86.86 | 1 | 100 | 100.00 | 1 | 100 | 100.00 |
| 0.26 | 3 | 86 | 85.57 | 1 | 100 | 100.00 | 1 | 100 | 100.00 |
| 0.25 | 3 | 84 | 83.85 | 1 | 100 | 100.00 | 1 | 100 | 100.00 |
| 0.23 | 3 | 81 | 80.35 | 1 | 100 | 100.00 | 1 | 100 | 100.00 |
| 0.2 | 3 | 75 | 74.93 | 1 | 100 | 100.00 | 1 | 100 | 100.00 |
| 0.15 | 4 | 65 | 64.78 | 3 | 66 | 95.72 | 4 | 65 | 95.62 |
| 0.14 | 4 | 63 | 62.43 | 3 | 64 | 93.59 | 4 | 63 | 92.93 |
| 0.13 | 4 | 61 | 60.04 | 3 | 62 | 91.43 | 5 | 61 | 89.99 |
| 0.12 | 4 | 58 | 57.61 | 4 | 58 | 88.69 | 5 | 58 | 86.78 |
| 0.11 | 4 | 56 | 55.14 | 4 | 56 | 85.89 | 6 | 57 | 83.22 |
| 0.1 | 5 | 53 | 52.29 | 5 | 53 | 82.79 | 7 | 55 | 79.48 |
| 0.09 | 5 | 50 | 49.37 | 5 | 50 | 79.39 | 7 | 51 | 75.36 |
| 0.08 | 5 | 47 | 46.39 | 6 | 47 | 75.65 | 8 | 49 | 70.97 |
| 0.07 | 6 | 44 | 43.17 | 7 | 44 | 71.60 | 9 | 46 | 66.26 |
| 0.06 | 6 | 40 | 39.69 | 7 | 40 | 67.07 | 9 | 42 | 61.13 |
| 0.05 | 7 | 36 | 35.98 | 9 | 37 | 61.96 | 10 | 38 | 55.52 |
| 0.04 | 8 | 32 | 31.92 | 10 | 33 | 56.17 | 12 | 35 | 49.28 |
| 0.03 | 9 | 28 | 27.31 | 12 | 29 | 49.43 | 14 | 30 | 42.25 |
| 0.02 | 11 | 22 | 21.96 | 16 | 24 | 41.18 | 17 | 24 | 34.00 |
| 0.01 | 15 | 16 | 15.17 | 25 | 17 | 30.06 | 23 | 17 | 23.48 |
| 0.007 | 17 | 13 | 12.57 | 30 | 15 | 25.54 | 27 | 14 | 19.44 |
| 0.005 | 21 | 11 | 10.54 | 37 | 13 | 21.90 | 32 | 12 | 16.29 |
| 0.003 | 26 | 9 | 8.07 | 50 | 10 | 17.33 | 41 | 9 | 12.47 |
| 0.001 | 45 | 5 | 4.58 | 92 | 6 | 10.47 | 70 | 5 | 7.07 |

Можно заметить, что учет стоимости объединения объектов в группы существенно увеличивает стоимость групповых проверок. Также в новой процедуре нередко требуется проводить большее количество проверок для обнаружения всех дефектных объектов, чем в процедуре Стерретта. Данный факт можно объяснить тем, что в условиях учета стоимости объединения объектов в группы при попытке минимизировать затраты на процедуру для процедуры выбирается группа большего объема и, как следствие, чаще приходится проводить индивидуальные проверки.

Функции (5.7) — (5.10), определяющие стоимость проверки группы в зависимости от ее размера, как и ожидалось, оказывают значительное влияние на результаты моделирования. Если обычная процедура Стерретта позволяет сэкономить на обследовании совокупности при всех вероятностях дефектности объекта, кроме $p = 0.38$, когда индивидуальные и групповые проверки в равной степени предпочтительны, то при учете стоимости объединения объектов в группы в нескольких ситуациях дешевле будет проверить все объекты по одному.

Так для функций (5.7), (5.9) и (5.10) при вероятностях $p \geq 0.2$ лучше не утруждаться применением групповых процедур, так как выигрыша в стоимости это не даст. Применение групповых процедур с функцией (5.8) уже при $p = 0.27$, дает, однако, определенное преимущество по сравнению с индивидуальной проверкой. Данный результат становится понятным, если обратиться к рис. 5.1, на котором изображены графики всех четырех функций. На начальном этапе при небольших значениях n (от 2 до 5) эта функция растет медленнее остальных, т. е., она дает самую маленькую цену за групповой тест. Этот факт объясняет и то, что вплоть до $p = 0.07$ новая процедура с функцией (5.8) дает наименьшую стоимость обследования совокупности по сравнению с другими вариантами новой процедуры.

Оптимальные размеры групп для новой процедуры с функциями (5.7) и (5.8) всегда оказываются меньше соответствующих размеров в процедуре Стерретта и в новой процедуре с функциями (5.9) и (5.10). Стоимость обследования при этом всегда оказывается выше.

Объясняется это тем, что функции (5.7) и (5.8) не ограничены и всегда линейно возрастают с увеличением n , а так как группа большего объема, особенно для высоких вероятностей p , предполагает необходимость много раз проверять содержащиеся в ней объекты группами меньшего объема, использовать объем группы поменьше оказывается более выгодным. И, как следствие, оптимальный размер группы для этих функций увеличивается медленнее с уменьшением вероятности дефектности объекта, чем это происходит для обычной процедуры Стерретта и новой процедуры с функциями (5.9) и (5.10).

Функция (5.9) также не ограничена, но растет она значительно медленнее, чем (5.7) и (5.8), одному и тому же приращению n соответствует все меньшее приращение функции (5.9). Функция (5.10) вообще ограничена, и цена за групповой тест при ее использовании никогда не будет больше, чем некоторая константа. Поэтому новая процедура в сочетании с этими функциями дает лучшие результаты, чем в сочетании с функциями (5.7) и (5.8).

Можно заметить, что для малых вероятностей дефектности объекта $p < 0.01$ использование функции (5.9) подразумевает значительно большие оптимальные размеры группы, чем использование функции (5.10), при этом выигрыша в стоимости это не дает: для низких вероятностей дефектности объекта наиболее выгодна новая процедура с функцией (5.10). Меньший объем группы в случае функции (5.10) по сравнению с функцией (5.9) оправдан, потому что стоимость проверки подгруппы с высокой вероятностью окажется почти такой же, как стоимость проверки группы, так как значения функции (5.10) с увеличением n достаточно быстро возрастают и потом практически перестают меняться. Следовательно, в процедуре с функцией (5.10) часто используется одна и та же близкая к предельной стоимости проверки группы и лучше взять объем группы поменьше, чтобы не делать «лишних» групповых проверок. В функции (5.9) стоимости подгрупп разного размера больше различаются между собой, поэтому при необходимости проверять подгруппы стоимость проверки подгруппы будет дешевле.

Оценить поведение функции стоимости проверок (5.3) в расчете на 100 объектов при различных значениях параметров можно по графикам на рис. 5.4 — 5.10. Графики построены для процедуры Стерретта и для новой процедуры при использовании функций (5.7) — (5.10), задающих цены за групповые тесты. Рассмотрены следующие значения вероятности дефектности объекта p : 0.3, 0.2, 0.1, 0.05, 0.01, 0.005 и 0.001. Рисунки позволяют увидеть, не только как меняется форма графиков при одной и той же вероятности дефектности объекта p , но и как меняется форма графиков при различных вероятностях дефектности объекта.

Зависимость стоимости проверок от n при $p = 0.3$

Стоимость проверок

на 100 объектов

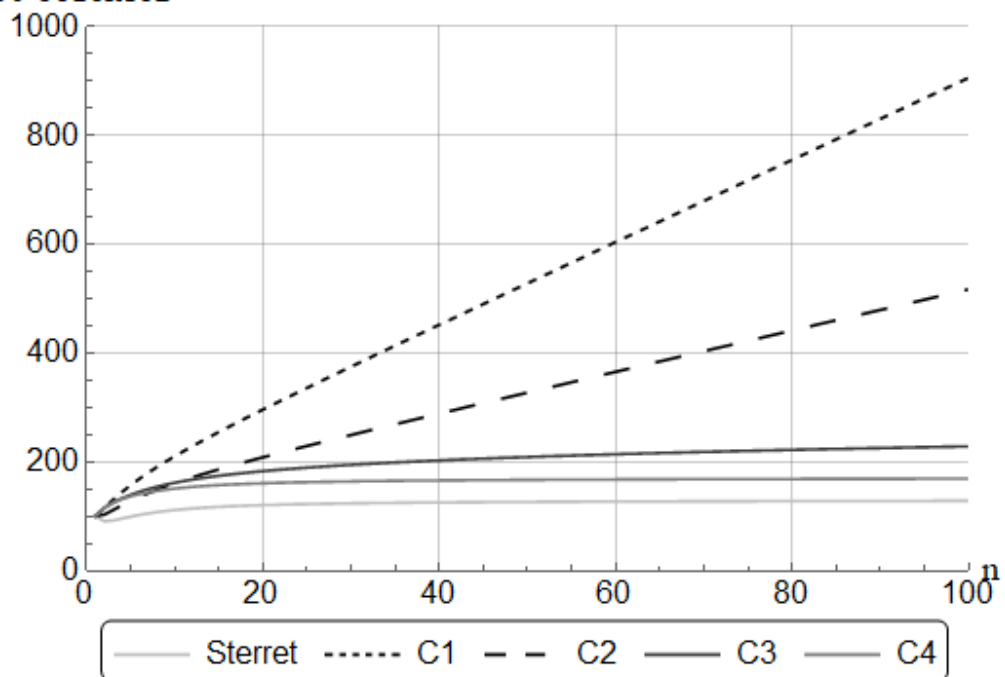


Рис. 5.4. Зависимость стоимости проверок от объема группы для различных функций $C_n(n)$ при вероятности дефектности $p = 0.3$

Зависимость стоимости проверок от n при $p = 0.2$

Стоимость проверок

на 100 объектов

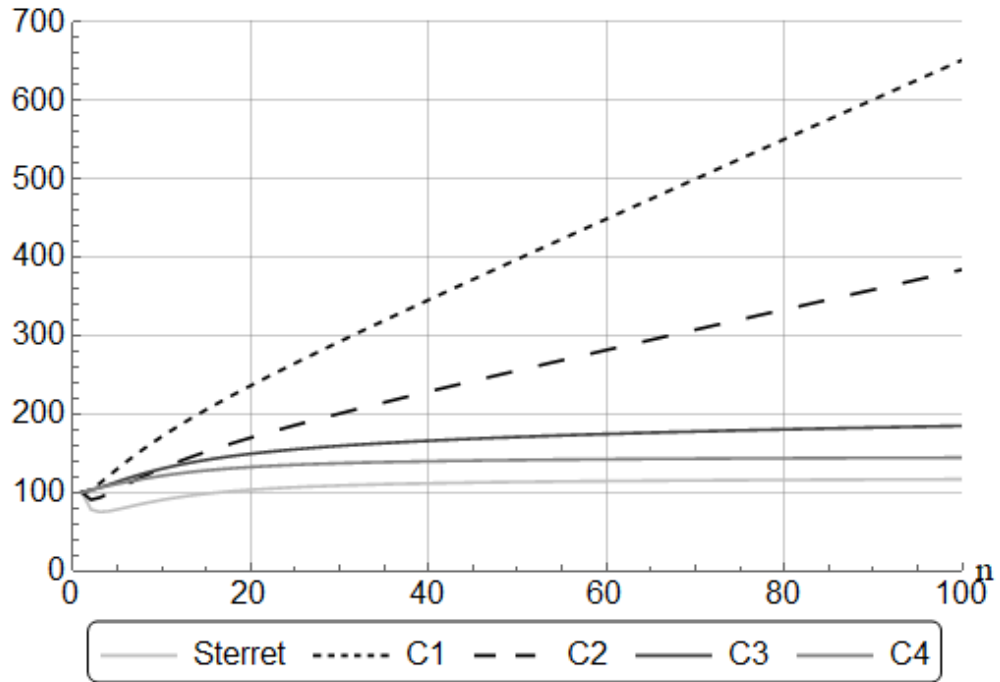


Рис. 5.5. Зависимость стоимости проверок от объема группы для различных функций $C_n(n)$ при вероятности дефектности $p = 0.2$

Зависимость стоимости проверок от n при $p = 0.1$

Стоимость проверок

на 100 объектов

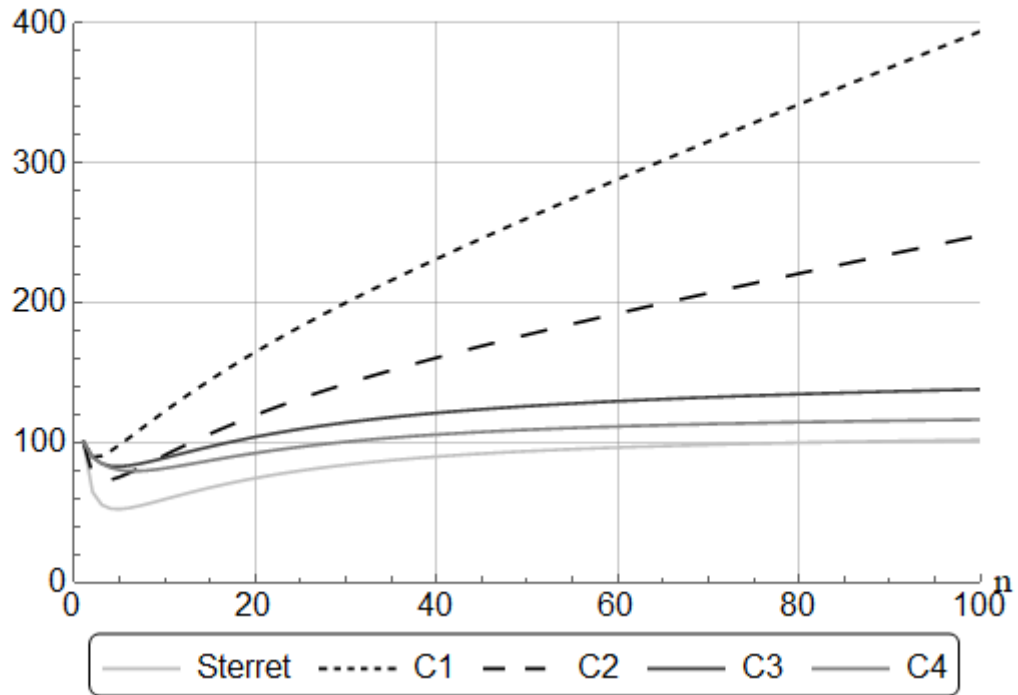


Рис. 5.6. Зависимость стоимости проверок от объема группы для различных функций $C_n(n)$ при вероятности дефектности $p = 0.1$

Зависимость стоимости проверок от n при $p = 0.05$

Стоимость проверок

на 100 объектов

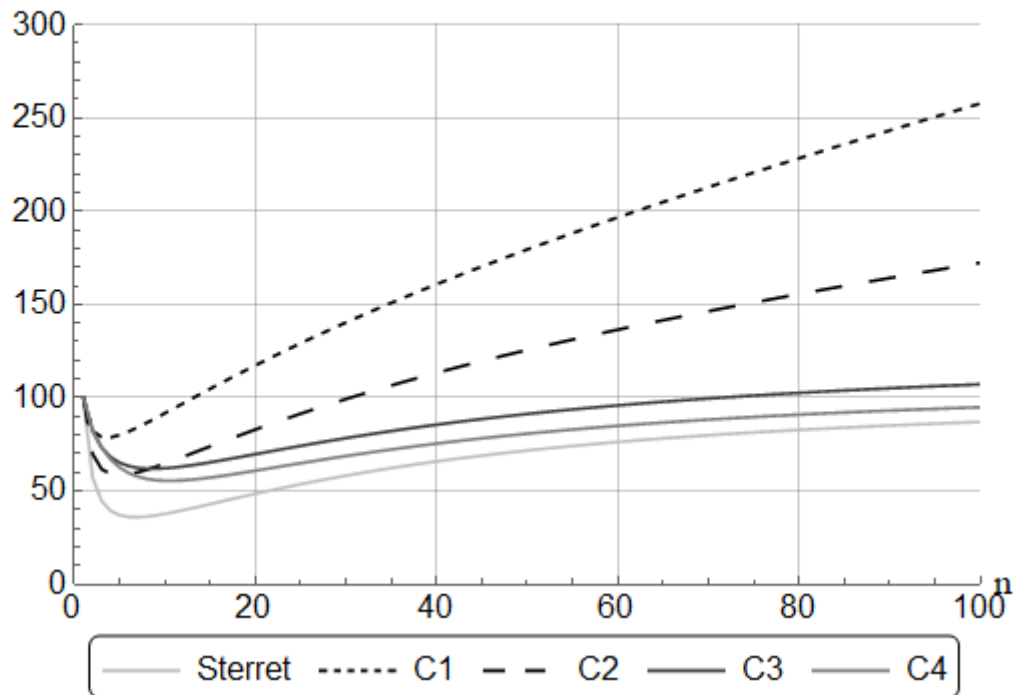


Рис. 5.7. Зависимость стоимости проверок от объема группы для различных функций $C_n(n)$ при вероятности дефектности $p = 0.05$

Зависимость стоимости проверок от n при $p = 0.01$

Стоимость проверок

на 100 объектов

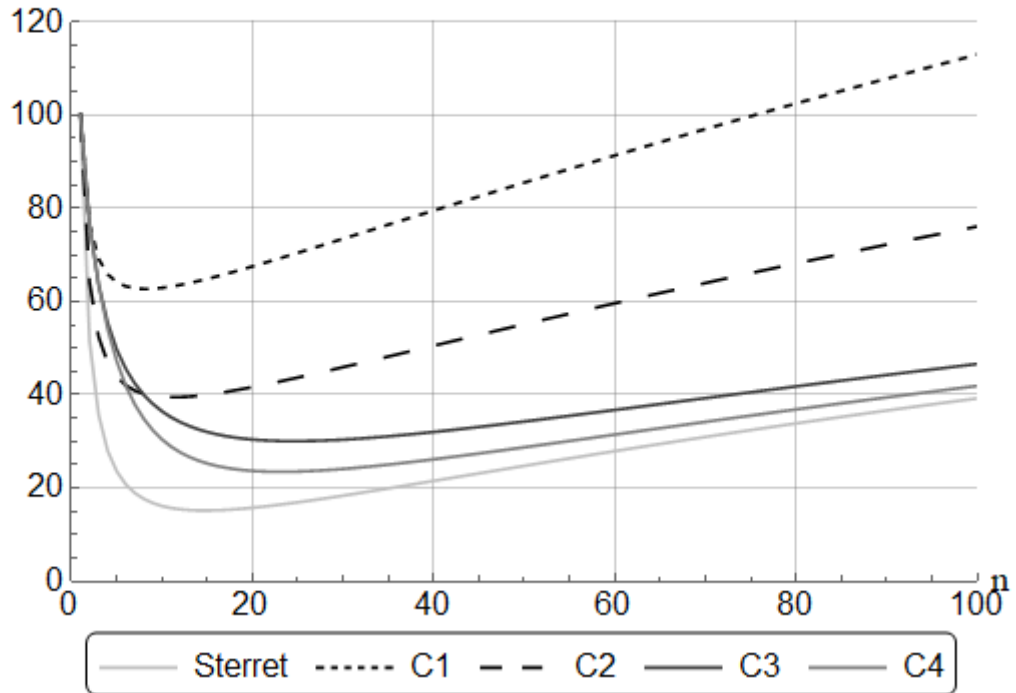


Рис. 5.8. Зависимость стоимости проверок от объема группы для различных функций $C_n(n)$ при вероятности дефектности $p = 0.01$

Зависимость стоимости проверок от n при $p = 0.005$

Стоимость проверок
на 100 объектов

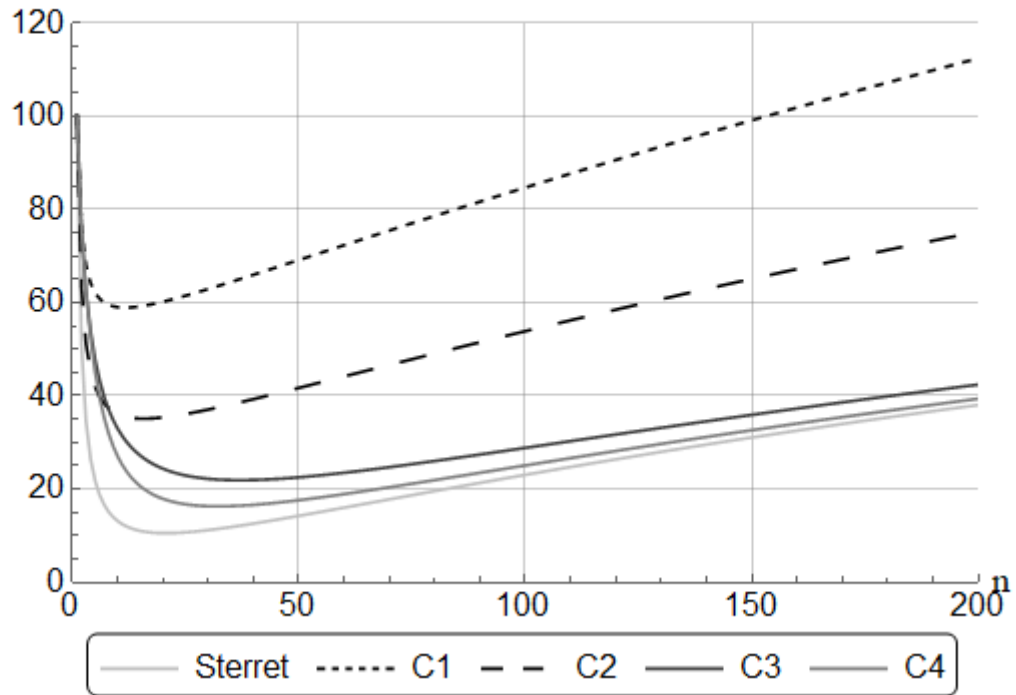


Рис. 5.9. Зависимость стоимости проверок от объема группы для различных функций $C_n(n)$ при вероятности дефектности $p = 0.005$

Зависимость стоимости проверок от n при $p = 0.001$

Стоимость проверок

на 100 объектов

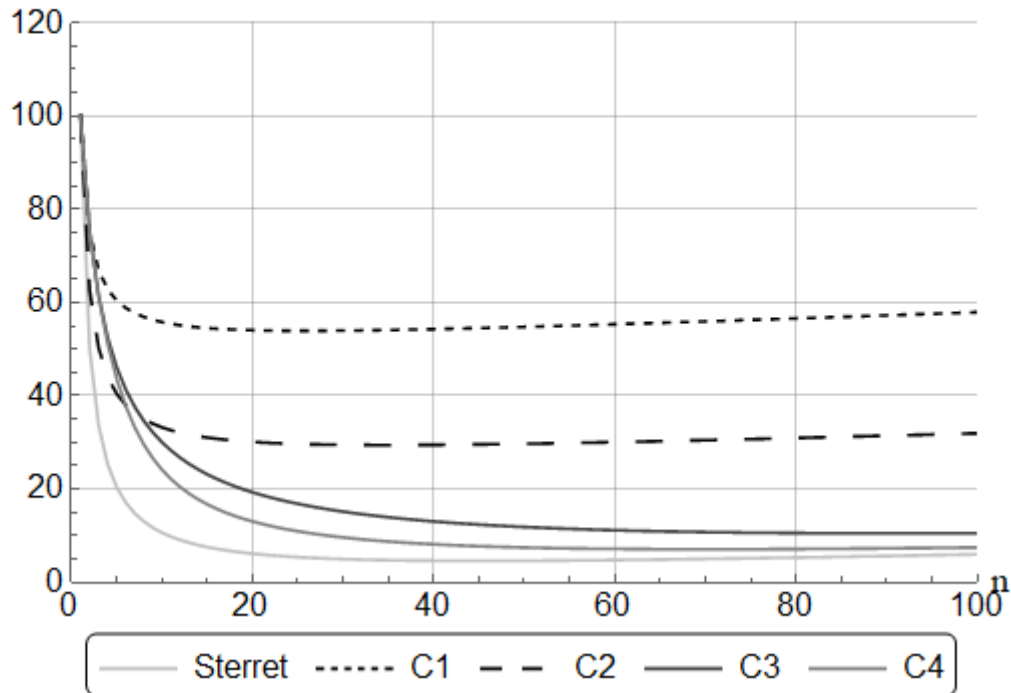


Рис. 5.10. Зависимость стоимости проверок от объема группы для различных функций $C_n(n)$ при вероятности дефектности $p = 0.001$

Прежде всего следует обратить внимание на изменение масштаба вертикальной оси. Стоимость проверок на 100 объектов значительно уменьшается при снижении вероятности дефектности объекта P . На последнем рисунке локальные экстремумы функции стоимости едва различимы, значит, при еще более низкой вероятности P экстремум окажется за пределами рассматриваемого отрезка, но это логично, поскольку в группе из 100 объектов, каждый из которых дефектный с вероятностью менее 0.001, очень трудно обнаружить дефектный объект.

Имеется следующая закономерность: с уменьшением вероятности P на каждом графике около левой границы отрезка $[1, 100]$ появляется локальный экстремум, который затем смещается к правой границе отрезка. Точка экстремума является оптимальным размером группы, так как в ней достигается наименьшее значение функции стоимости на отрезке $[1, 100]$. Смещение точки экстремума к правой границе отрезка означает, что оптимальный размер группы увеличивается и групповые процедуры становятся все дешевле по сравнению с индивидуальной проверкой.

5.4. Сравнение эффективности процедур Дорфмана и Стерретта с учетом стоимости объединения объектов в группы

Данный параграф посвящен сравнению новых моделей, построенных в параграфах 4.1 и 5.1, с целью выявить условия, в которых каждая из них дает больший выигрыш в стоимости. Таблица 5.2 построена на основе данных из табл. 4.4 и 5.1. Так как в одной из таблиц представлена средняя стоимость проверок, приходящаяся на один объект, а в другой – средняя стоимость проверок 100 объектов, стоимости в последней были пересчитаны. Столбцы с оптимальными размерами групп и стоимостью проверок на один объект для модифицированной процедуры Дорфмана названы в соответствии с использованной для моделирования величиной соотношения C_n / C_1 , а для модифицированной процедуры Стерретта – по использованной функции, задающей цены за групповые проверки $C_n(n)$.

Табл. 5.2. Сравнение результатов моделирования процедур, учитывающих стоимости объединения объектов в группы

| P | $C_n / C_1 = 1,5$ | | $C_n / C_1 = 2$ | | $C_n / C_1 = 3$ | | $C_n / C_1 = 5$ | | $C_n / C_1 = 10$ | | $C_1(n)$ | | $C_2(n)$ | | $C_3(n)$ | | $C_4(n)$ | | P |
|-------|-------------------|------|-----------------|------|-----------------|------|-----------------|------|------------------|------|----------|------|----------|------|----------|------|----------|------|-------|
| | k | Cost | k | Cost | k | Cost | k | Cost | k | Cost | k | Cost | k | Cost | k | Cost | k | Cost | |
| 0.3 | 1 | 1.00 | 1 | 1.00 | 1 | 1.00 | 1 | 1.00 | 1 | 1.00 | 1 | 1.00 | 1 | 1.00 | 1 | 1.00 | 1 | 1.00 | 0.3 |
| 0.25 | 1 | 1.00 | 1 | 1.00 | 1 | 1.00 | 1 | 1.00 | 1 | 1.00 | 1 | 1.00 | 2 | 0.97 | 1 | 1.00 | 1 | 1.00 | 0.25 |
| 0.2 | 4 | 0.97 | 1 | 1.00 | 1 | 1.00 | 1 | 1.00 | 1 | 1.00 | 1 | 1.00 | 2 | 0.91 | 1 | 1.00 | 1 | 1.00 | 0.2 |
| 0.15 | 4 | 0.85 | 6 | 0.96 | 1 | 1.00 | 1 | 1.00 | 1 | 1.00 | 2 | 0.96 | 3 | 0.83 | 3 | 0.96 | 4 | 0.96 | 0.15 |
| 0.13 | 5 | 0.80 | 6 | 0.90 | 1 | 1.00 | 1 | 1.00 | 1 | 1.00 | 2 | 0.94 | 3 | 0.79 | 3 | 0.91 | 5 | 0.90 | 0.13 |
| 0.12 | 5 | 0.77 | 6 | 0.87 | 1 | 1.00 | 1 | 1.00 | 1 | 1.00 | 2 | 0.92 | 3 | 0.77 | 4 | 0.89 | 5 | 0.87 | 0.12 |
| 0.1 | 5 | 0.71 | 6 | 0.80 | 8 | 0.94 | 1 | 1.00 | 1 | 1.00 | 2 | 0.90 | 3 | 0.73 | 5 | 0.83 | 7 | 0.79 | 0.1 |
| 0.09 | 5 | 0.68 | 6 | 0.77 | 8 | 0.90 | 1 | 1.00 | 1 | 1.00 | 2 | 0.88 | 4 | 0.71 | 5 | 0.79 | 7 | 0.75 | 0.09 |
| 0.08 | 5 | 0.64 | 6 | 0.73 | 9 | 0.86 | 1 | 1.00 | 1 | 1.00 | 3 | 0.86 | 4 | 0.68 | 6 | 0.76 | 8 | 0.71 | 0.08 |
| 0.07 | 6 | 0.60 | 7 | 0.68 | 9 | 0.81 | 14 | 1.00 | 1 | 1.00 | 3 | 0.84 | 4 | 0.65 | 7 | 0.72 | 9 | 0.66 | 0.07 |
| 0.06 | 6 | 0.56 | 7 | 0.64 | 9 | 0.76 | 14 | 0.94 | 1 | 1.00 | 3 | 0.81 | 4 | 0.62 | 7 | 0.67 | 9 | 0.61 | 0.06 |
| 0.05 | 6 | 0.51 | 8 | 0.59 | 10 | 0.70 | 14 | 0.87 | 1 | 1.00 | 4 | 0.79 | 5 | 0.59 | 9 | 0.62 | 10 | 0.56 | 0.05 |
| 0.04 | 7 | 0.46 | 8 | 0.53 | 11 | 0.63 | 15 | 0.79 | 1 | 1.00 | 4 | 0.76 | 6 | 0.55 | 10 | 0.56 | 12 | 0.49 | 0.04 |
| 0.03 | 8 | 0.40 | 9 | 0.46 | 12 | 0.56 | 16 | 0.70 | 28 | 0.93 | 5 | 0.72 | 6 | 0.51 | 12 | 0.49 | 14 | 0.42 | 0.03 |
| 0.02 | 9 | 0.33 | 11 | 0.38 | 14 | 0.46 | 19 | 0.58 | 30 | 0.79 | 6 | 0.68 | 8 | 0.46 | 16 | 0.41 | 17 | 0.34 | 0.02 |
| 0.01 | 13 | 0.24 | 15 | 0.27 | 19 | 0.33 | 25 | 0.42 | 38 | 0.58 | 8 | 0.63 | 11 | 0.39 | 25 | 0.30 | 23 | 0.23 | 0.01 |
| 0.005 | 18 | 0.17 | 21 | 0.20 | 26 | 0.24 | 34 | 0.30 | 51 | 0.42 | 12 | 0.59 | 16 | 0.35 | 37 | 0.22 | 32 | 0.16 | 0.005 |
| 0.001 | 39 | 0.08 | 46 | 0.09 | 56 | 0.11 | 73 | 0.14 | 105 | 0.19 | 26 | 0.54 | 35 | 0.29 | 92 | 0.10 | 70 | 0.07 | 0.001 |

Оказывается, что при высокой вероятности дефектности объекта $p = 0.3$ или $p = 0.25$ любая из рассмотренных процедур, учитывающих стоимость объединения объектов в группы, бесполезна, так как индивидуальная проверка оказывается дешевле. Вывод справедлив даже при условии, что стоимость проверки группы незначительно выше стоимости проверки объекта. При этом и процедура Дорфмана, и процедура Стерретта вследствие того, что игнорируют затраты на объединение объектов в группы, при обеих вероятностях оказываются эффективнее индивидуального обследования, следовательно, их использование повлечет за собой непредвиденные расходы. Данный вывод является ключевым, поскольку отвечает цели настоящей работы: дать рекомендации по предпочтению индивидуальных или групповых проверок для выявления всех дефектных объектов совокупности при учете стоимости объединения объектов в группы.

Следующий вывод, который можно сделать, совпадает с выводом Стерретта относительно большей эффективности его процедуры по сравнению с процедурой Дорфмана. А именно: если вероятность объекта мала (например, менее 0.01), тогда новая процедура на основе процедуры Стерретта эффективнее новой процедуры на основе процедуры Дорфмана. Однако данный вывод во многом зависит от функции, задающей цены за групповые тесты в модифицированной процедуре Стерретта. Если стоимость проверки группы равномерно растет с увеличением количества объектов в ней, то от модифицированной процедуры Стерретта лучше отказаться в пользу модифицированной процедуры Дорфмана. Если же при больших размерах группы затраты на их получение перестают значительно меняться с ростом n , то модифицированная процедура Стерретта будет стоить дешевле модифицированной процедуры Дорфмана.

Данный вывод тоже очень важен, так как свидетельствует о том, что в разных областях человеческой деятельности могут быть выгодны разные процедуры, так как стоимость объединения объектов в группы определяется областью решаемой задачи. Например, объединение компонентов аппаратуры или приборов имеет, скорее, линейную зависимость от количества объектов, которые необходимо объединить, к тому же является сравнительно трудоемким процессом и, как правило, требует разбора группы обратно на отдельные объекты, поэтому в данной ситуации модифицированная процедура Дорфмана предпочтительнее. Если рассматривать объединение сывороток крови, то зависимость будет также линейной, но материальных и физических затрат на создание группы требуется меньше, чем в случае с оборудованием, кроме того, отсутствует необходимость разделять группу на объекты. Поэтому в данной

ситуации модифицированная процедура Стерретта может дать лучшие результаты, чем модифицированная процедура Дорфмана.

Групповые проверки химических соединений на наличие среди них потенциальных активных веществ для новых лекарств могут быть реализованы с помощью автомата, который выполняет объединение веществ из нескольких ячеек и помещает получившийся образец в новую ячейку. Здесь затраты на создание группы любого размера практически равны. При этом проверку группы из небольшого количества соединений легче сделать вручную, а не с помощью автомата, поэтому для небольших размеров группы цены на групповые тесты будут сильно различаться. Кроме того, вероятность обнаружить активное вещество мала, поэтому процедура Стерретта хорошо подходит к данной ситуации.

Решение конкретной реальной задачи подразумевает определение более точной зависимости стоимости объединения объектов в группы от количества объектов, следовательно, применение той или иной процедуры будет более обоснованным и выгодным с точки зрения стоимости.

Разумеется, все сделанные выводы, кроме первого, не касаются случаев, когда вероятность дефектности объекта высока.

ГЛАВА 6. ЗАДАЧА ОПРЕДЕЛЕНИЯ ОПТИМАЛЬНОГО ОБЪЕМА ГРУППЫ ДЛЯ ВЫБОРА КОНЕЧНОГО ЧИСЛА ГОДНЫХ ОБЪЕКТОВ

Пусть имеется бесконечная совокупность объектов O_1, O_2, \dots , каждый из которых независимо от других с вероятностью p ($0 < p < 1$) является годным и с вероятностью $q = 1 - p$ – дефектным. Необходимо отобрать из бесконечной совокупности объектов N годных объектов ($N > 1$). Полагаем, что возможна групповая процедура проверок. То есть n объектов ($n \geq 1$) объединяются в группу и результат проверки этой группы показывает, есть или нет среди этих n объектов хотя бы один дефектный объект.

Процедуру групповых проверок построим следующим образом. Из бесконечной совокупности объектов O_1, O_2, \dots , выберем n объектов и над ними произведем групповой эксперимент. Если все объекты в группе оказались годными, то их принимаем. То есть перед следующим экспериментом необходимо отобрать $(N - n)$ годных объектов из бесконечной совокупности объектов. Если же в группе оказался хотя бы один дефектный объект, то группу из n объектов отвергаем и перед следующим экспериментом вновь необходимо выбрать N годных объектов. Таким образом, если перед i -тым экспериментом необходимо было отобрать N_i годных объектов из бесконечной совокупности и для i -того эксперимента были отобраны n_i объектов, то с вероятностью p^{n_i} перед $(i+1)$ экспериментом нужно отобрать $N_{i+1} = N - n_i$ годных объектов и с вероятностью $(1 - p^{n_i})$ нужно отобрать $N_{i+1} = N_i$ годных объектов.

Стратегия такой процедуры групповых проверок заключается в том, чтобы за один эксперимент отобрать максимальное количество годных объектов. Математическое ожидание τ – числа отобранных объектов за один эксперимент, если для эксперимента были отобраны n объектов, равняется

$$\mu(\tau) = np^n + 0 \cdot (1 - p^n) = np^n.$$

При сделанных предположениях справедлива следующая теорема.

Теорема. Оптимальным объемом группы для одной проверки вышеописанной процедуры групповых проверок является

$$n^* = \begin{cases} \lfloor -1/\ln p \rfloor, & \text{если } (-1/\ln p) \text{ — не целое число и } (p - q \lfloor -1/\ln p \rfloor) \leq 0, \\ \lfloor -1/\ln p \rfloor + 1, & \text{если } (-1/\ln p) \text{ — не целое число и } (p - q \lfloor -1/\ln p \rfloor) > 0, \\ -1/\ln p, & \text{если } (-1/\ln p) \text{ — целое число,} \end{cases}$$

где $\lfloor \cdot \rfloor$ — целая часть числа.

Доказательство. Математическое ожидание числа принятых объектов за один эксперимент, если в группу были объединены n объектов, равняется

$$\mu(\tau) = np^n,$$

где p — вероятность годности одного объекта. Найдем максимум этой функции. Для этого производную этой функции по n найдем $\mu'(\tau) = np^n \ln p$ и приравняем ее к нулю: $p^n = (1 + n \ln p) = 0$. Из этого следует, что при $n^* = -1/\ln p$ $\mu(\tau)$ принимает максимальное значение. Но n — объем группы может быть только целым числом. Тогда заметим, что функция $\mu(\tau)$ является выпуклой функцией. Следовательно, при $n = (-1/\ln p)$ — не целом числе $\mu(\tau)$ принимает максимальное значение либо при $n = \lfloor -1/\ln p \rfloor$, либо при $n = \lfloor -1/\ln p \rfloor + 1$.

Рассмотрим разность $\mu_2(\tau) - \mu_1(\tau)$, где $\mu_2(\tau)$ — математическое ожидание числа принятых объектов за один эксперимент при $n = \lfloor -1/\ln p \rfloor + 1$ и $\mu_1(\tau)$ — математическое ожидание числа принятых объектов за один эксперимент при $n = \lfloor -1/\ln p \rfloor$.

$$\mu_2(\tau) - \mu_1(\tau) = p^{\lfloor -1/\ln p \rfloor} (p - q^{\lfloor -1/\ln p \rfloor}).$$

Эта разность больше нуля при $(p - q^{\lfloor -1/\ln p \rfloor}) > 0$ и меньше или равна нулю при $(p - q^{\lfloor -1/\ln p \rfloor}) \leq 0$. То есть если $(-1/\ln p)$ является не целым числом, то оптимальным объемом группы n^* будет $\lfloor -1/\ln p \rfloor$ при $(p - q^{\lfloor -1/\ln p \rfloor}) \leq 0$ $\lfloor -1/\ln p \rfloor + 1$ при $(p - q^{\lfloor -1/\ln p \rfloor}) > 0$.

Таким образом, получаем, что оптимальный объем группы для приведенной процедуры групповых проверок равен:

$$n^* = \begin{cases} \lfloor -1/\ln p \rfloor, & \text{если } (-1/\ln p) \text{ — не целое число и } (p - q^{\lfloor -1/\ln p \rfloor}) \leq 0, \\ \lfloor -1/\ln p \rfloor + 1, & \text{если } (-1/\ln p) \text{ — не целое число и } (p - q^{\lfloor -1/\ln p \rfloor}) > 0, \\ -1/\ln p, & \text{если } \lfloor -1/\ln p \rfloor \text{ — целое число,} \end{cases} \quad (6.1)$$

что и требовалось доказать.

Следствие. При $p > 0,5$ оптимальный объем группы для групповой процедуры проверок $n^* \geq 2$.

Из (6.1) легко видеть, что n^* возрастает с ростом p . Если положим, что $\lfloor -1/\ln p \rfloor = 1$, то увидим, что разность положительна при $p > q$. Тогда, учитывая, что $p + q = 1$, получим, что при $p > 0,5$ оптимальный объем группы $n^* \geq 2$, что и требовалось доказать.

ГЛАВА 7. ЗАДАЧА НАХОЖДЕНИЯ ОПТИМАЛЬНОГО ОБЪЕМА ГРУППЫ ДЛЯ КЛАССИФИКАЦИИ ОБЪЕКТОВ НА T КЛАССОВ ($T > 2$)

Пусть на классификацию поступают объекты O_1, O_2, \dots , каждый из которых независимо от других с вероятностью $p_j (j = \overline{1, t})$ принадлежит к j -тому классу, $\sum_{j=1}^t p_j = 1$. Каждый объект $O_i (i=1, 2, \dots)$ имеет вес y_i , причем если $y_i = a_j (j = \overline{1, t})$, то объект O_i принадлежит j -тому классу. Необходимо классифицировать каждый объект.

Крайними классами будем называть классы, которые имеют наименьший и наибольший вес. Не теряя общности, будем считать, что это первый и t -й классы.

Опишем модифицированную процедуру Дорфмана. Для групповой проверки в группу объединяется n объектов. Результатом групповой проверки является установление общего веса Y . Если $Y = a_1 n$ или $Y = a_t n$, то все объекты относятся к одному из крайних классов, в противном случае объекты классифицируются индивидуально до тех пор, пока остаток объектов в группе не будет принадлежать одному из крайних классов. Может случиться так, по понятным причинам, что придется все объекты классифицировать индивидуально кроме последнего. Это вытекает из того, что зная суммарный вес n объектов и индивидуальные веса $(n-1)$ объекта, автоматически получим вес n -го объекта.

Заметим, что вообще говоря, при сильно различающихся весах $a_j (j = \overline{1, t})$ возможны случаи групповой классификации не только к крайним классам одновременно n объектов. Этот случай здесь рассматриваться не будет.

Предположим, что для классификации объектов на t классов ($t \geq 3$) применима вышеописанная модифицированная процедура групповых проверок Дорфмана. Необходимо найти оптимальный объем группы n^* для модифицированной процедуры Дорфмана, который бы минимизировал математическое ожидание общего количества проверок, приходящихся на каждый объект.

Введем следующие обозначения. Через $H(n)$ обозначим математическое ожидание общего числа проверок, необходимых для классификации на t классов n объектов, объединенных в одну группу, с помощью модифицированной процедуры Дорфмана. Через $G(n)$ обозначим математическое ожидание общего числа проверок, приходящихся на один объект, необходимых для

классификации на t классов n объектов, объединенных в одну группу, с помощью модифицированной процедуры Дорфмана. Очевидно, что $G(n) = H(n)n^{-1}$.

Положим, что $p_j = 1/t (j = \overline{1, t})$, т. е. каждый объект $O_i (i=1, 2, \dots)$ с равными вероятностями принадлежит к каждому из t классов. Тогда справедлива следующая теорема.

Теорема 7.1. Математическое ожидание общего числа проверок, необходимых для классификации на t классов n объектов, объединенных в одну группу, с помощью модифицированной процедуры Дорфмана

$$H(n) = \begin{cases} 1 & \text{при } n=1, \\ \frac{2+2(t-1)\left[\sum_{i=1}^{n-2} t^{i-1}(i+1)\right] + n(t^2-2)t^{n-2}}{t^n} & \text{при } n \geq 2 \end{cases}. \quad (7.1)$$

Доказательство. Для $n=1$ эта формула очевидна. Покажем справедливость формулы для $n=2$. Так как в группе n объектов, каждый из которых с равными вероятностями независимо от других объектов принадлежит к одному из t классов, то вероятность любого набора из n объектов равна $1/t^n$, откуда и получаем знаменатель $H(n)$ при $n \geq 2$. Числитель в этом случае должен представлять собой сумму количества проверок при каждом из наборов. Если все объекты принадлежат одному из крайних классов, то понадобится всего одна проверка. То есть одну проверку умножаем на два исхода и получаем первое слагаемое числителя. Пусть теперь не все объекты принадлежат к одному из крайних классов. Тогда если i -й объект $i = \overline{1, n-2}$ не принадлежит крайним классам, а остальные $(n-i)$ объектов принадлежат к одному из крайних классов, то потребуется $(i+1)$ проверка и таких наборов будет $2(t-1)t^{i-1}$, что отражает второе слагаемое числителя. Наконец, n проверок понадобится, если на $(n-1)$ или на n -м месте будет объект, не принадлежащий одному из крайних классов. Таких наборов будет $(t^2-2)t^{n-2}$, что отражает третье слагаемое числителя.

Просуммировав числитель в (7.1), получим более удобное выражение для $H(n)$:

$$H(n) = \begin{cases} 1 & \text{при } n=1 \\ \frac{nt^n(1-t) + 2t^{n-1} - 2}{t^n(1-t)} & \text{при } n \geq 2 \end{cases}. \quad (7.2)$$

Теорема доказана.

Теорема 7.2. Пусть вероятность принадлежности каждого объекта O_i из совокупности O_1, O_2, \dots , к одному из t классов $p_j=1/t$ ($j=1, t$), тогда оптимальный объем группы при классификации объектов для модифицированной процедуры Дорфмана $n^*=2$.

Доказательство. Выпишем функцию $G(n)$:

$$G(n) = \begin{cases} 1 & \text{при } n=1, \\ 1 + \frac{2(t^{n-1}-1)}{nt^n(1-t)} & \text{при } n \geq 2. \end{cases}$$

Рассмотрим разность $G(2) - G(1) = -t^{-2}$. При любом $t \geq 3$ эта разность отрицательна, следовательно, групповая процедура с объемом группы $n=2$ обеспечивает меньшее математическое ожидание общего числа проверок (чем индивидуальные проверки с объемом группы $n=1$), приходящихся на каждый объект при любом числе классов $t \geq 3$.

Теперь рассмотрим для $n \geq 3$ разность

$$G(n) - G(n-1) = \frac{2(1-n-t^{n-1}+nt)}{t^n n(n-1)(1-t)}.$$

Легко убедиться, что эта разность всегда положительна. Следовательно, групповая процедура с объемом $n=2$ обеспечивает меньшее математическое ожидание общего числа проверок, приходящихся на каждый объект, чем групповые процедуры с объемами групп более 2. Таким образом, оптимальный объем группы для групповой процедуры проверок $n^*=2$, что и требовалось доказать.

Приведем табл. 7.1, отражающую зависимость математического ожидания общего числа проверок, приходящихся на один объект при использовании модифицированной процедуры Дорфмана, от числа классов t , если каждый объект O_i ($i=1, 2, \dots$) равновероятно принадлежит к одному из этих классов.

Таблица 7.1. Математическое ожидание общего числа проверок

| t – количество классов | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|--------------------------|-------|-------|-------|-------|-------|-------|-------|-------|
| $G(n^*=2)$ | 0,889 | 0,938 | 0,960 | 0,972 | 0,980 | 0,984 | 0,988 | 0,990 |

Теперь рассмотрим случай, когда объект O_i ($i=1, 2, \dots$) принадлежит с вероятностью p_j ($j=1, \dots, t$) к j -тому классу, $\sum_{j=1}^t p_j = 1$.

Теорема 7.3. Математическое ожидание общего числа проверок, необходимых для классификации на t классов n объектов, объединенных в одну группу, с помощью модифицированной процедуры Дорфмана:

$$H(n) = \begin{cases} 1 & \text{при } n=1 \\ n - p_1^n - p_i^n - p_1^{n-1} - p_i^{n-1} - \dots - p_1^2 - p_i^2 & \text{при } n \geq 2 \end{cases}, \quad (7.3)$$

если p_1 и p_i – вероятности принадлежности объекта O_i ($i = 1, 2, \dots$) к крайним классам.

Доказательство. Для $n=1$ формула (7.3) очевидна. Покажем ее справедливость для $n \geq 2$.

Если все n объектов относятся к одному из крайних классов, то понадобится одна проверка. Вероятность того, что n объектов принадлежат к одному из крайних классов, равна $(p_1^n + p_i^n)$. Если на i -м ($i = \overline{1, n-2}$) месте группы из n объектов окажется объект, не принадлежащий к одному из крайних классов, а остальные $(n-i)$ объекты принадлежат к одному из крайних классов, то потребуется $(i+1)$ проверка. Вероятность такого события равна:

$$[p_1^{n-i}(1-p_1) + p_i^{n-i}(1-p_i)]$$

И наконец, в остальных случаях потребуется n проверок. Умножая вероятности описанных событий на количество необходимых проверок и суммируя эти произведения, получаем для $n \geq 2$:

$$H(n) = p_1^n + p_i^n + \sum_{i=1}^{n-2} (i+1) [(1-p_1)p_1^{n-i} + (1-p_i)p_i^{n-i}] + n \{ 1 - p_1^n - p_i^n - \sum_{i=1}^{n-2} [(1-p_1)p_1^{n-i} + (1-p_i)p_i^{n-i}] \} = n - p_1^n - p_i^n - p_1^{n-1} - p_i^{n-1} - \dots - p_1^2 - p_i^2,$$

что и требовалось доказать.

Теперь получим выражение для функции $G(n)$:

$$G(n) = \begin{cases} 1 & \text{при } n=1, \\ \frac{n - p_1^n - p_i^n - \dots - p_1^2 - p_i^2}{n} & \text{при } n \geq 2 \end{cases}.$$

Справедлива следующая теорема.

Теорема 7.4. Пусть P_j ($j = \overline{1, t}$) – вероятности принадлежности любого объекта из совокупности O_1, O_2, \dots к j -тому классу, тогда оптимальный объем группы n^* при классификации объектов для модифицированной процедуры Дорфмана будет больше $n=2$, если одна из вероятностей попасть в крайние классы p_1 или p_t больше 0,5.

Доказательство. Заметим, что для любого t и любых вероятностях p_1 и p_t справедливо

$$G(2) - G(1) < 0,$$

что означает: модифицированная процедура Дорфмана с объемом $n=2$ предпочтительнее индивидуальных проверок (объем группы $n=1$).

Нетрудно видеть, что функция $G(n)$ — выпуклая и имеет один минимум при некотором n . Рассмотрим разность

$$G(2) - G(3) = \frac{p_1^2(2p_1 - 1) + p_t^2(2p_t - 1)}{6}.$$

Эта разность положительна при $p_1 > 0,5$ или $p_t > 0,5$, стало быть, оптимальный объем группы $n^* > 2$, когда $p_1 > 0,5$ или $p_t > 0,5$.

Теорема доказана.

Теорема 7.5. Пусть выполнены условия предыдущей теоремы и $p_1 > 0,5$ или $p_t > 0,5$. Тогда оптимальным объемом группы n^* для модифицированной процедуры Дорфмана будет такое максимальное n , которое удовлетворяет неравенству

$$(n-1)p_1^n + (n-1)p_t^n - p_1^{n-1} - p_t^{n-1} - \dots - p_1^2 - p_t^2 > 0.$$

Доказательство. Как отмечалось в предыдущей теореме, $G(2) < G(1)$ и для $n > 2$ функция выпуклая, имеющая один минимум. Следовательно, оптимальным объемом для модифицированной процедуры Дорфмана будет такое n^* , которое удовлетворяет неравенству

$$G(n-1) - G(n) > 0 \text{ при } n > 3.$$

Но так как

$$G(n-1) - G(n) = (n-1)p_1^n + (n-1)p_t^n - p_1^{n-1} - p_t^{n-1} - \dots - p_1^2 - p_t^2,$$

теорема доказана.

Теперь пусть имеется m групп объектов, в каждой из которых n объектов ($m \leq n$). Причем известно, что в каждой из групп ровно один дефектный объект. Предполагаем, что любой объект может быть классифицирован либо как годный, либо как дефектный. Необходимо определить все дефектные объекты.

Самым естественным способом обнаружения дефектного объекта в каждой группе является проведение индивидуальных проверок объектов до тех пор, пока не будет обнаружен дефектный объект. Математическое ожидание количества индивидуальных проверок для нахождения единственного дефектного объекта в группе объектов объемом n равно

$$n^{-1}[1 + 2 + \dots + n] - 1 = (n^2 + n - 2) / 2n,$$

так как вероятность нахождения дефектного объекта на i -том месте ($i = \overline{1, n}$) равна i/n , а в случае нахождения дефектного объекта на n -м месте понадобится $(n - 1)$ индивидуальная проверка, так как он будет установлен из того факта, что предыдущие $(n - 1)$ объекты были годные. Следовательно, $H(nm)$ — математическое ожидание индивидуальных проверок, необходимых для выявления в m группах (каждая из которых содержит n объектов, и известно, что в каждой из групп ровно один дефектный объект) всех дефектных объектов, выражается следующей формулой:

$$H(nm) = m(n^2 + n - 2) / 2n.$$

Тогда $G(nm)$ — математическое ожидание количества индивидуальных проверок, приходящихся на один объект совокупности m групп по n объектов в каждой, равно

$$G(nm) = (n^2 + n - 2) / 2n^2,$$

так как

$$G(nm) = H(nm) / nm.$$

Ранее была описана модифицированная процедура Дорфмана, которая устанавливает точное количество дефектных объектов в группе одной групповой проверкой. Предположим, что эта процедура применима в данном случае.

Рассмотрим другой способ нахождения дефектных объектов в m группах, каждая из которых содержит n объектов ($1 < m < n$), если известно, что в каждой группе ровно по одному дефектному объекту. Для этого объекты расположим в виде матрицы $[O_{ij}]$, $i = \overline{1, m}$, $j = \overline{1, n}$. Совокупность объектов $(O_{i1}, O_{i2}, \dots, O_{in})$ назовем i -той группой-строкой, а совокупность объектов $(O_{1j}, O_{2j}, \dots, O_{mj})$ j -той группой-столбцом. Учитывая, что каждая группа-строка по условию содержит ровно один дефектный объект, перед поиском дефектных объектов индивидуальными проверками проведем групповые проверки по группам-столбцам. Для получения полной информации о содержании дефектных объектов в группах-столбцах потребуется не более $(n-1)$ проверка, так как известно общее количество дефектных объектов в матрице объектов $[O_{ij}]$, $i = \overline{1, m}$, $j = \overline{1, n}$, равное m .

Пусть случайная величина χ (число групп-столбцов, содержащих хотя бы один дефектный объект) равна l . Тогда в результате проведения групповых проверок по группам-столбцам получим новую матрицу объектов исключением из исходной матрицы групп-столбцов, не содержащих дефектных объектов и групп-столбцов, содержащих только дефектные объекты (возмо-

жен тот случай, когда все столбцы будут исключены, тогда дальнейшее проведение проверок прекращается). Пусть это матрица $[O'_{ij}]$, $i = \overline{1, m}, j = \overline{1, l}$, такая, что каждая группа-строка содержит, как и прежде, ровно один дефектный объект, а каждая группа-столбец содержит, по крайней мере, хотя бы один дефектный объект, причем $l \leq m$.

Далее индивидуальные проверки будем проводить по группам-столбцам с учетом того факта, что в каждой группе-столбце уже известно точное количество дефектных объектов. Простой перенумерацией столбцов матрицы $[O'_{ij}]$ получим такое расположение объектов матрицы, что в j_1 -том столбце последний дефектный объект (если объекты рассматриваются в следующем порядке: $O'_{1j_1}, O'_{2j_1}, \dots, O'_{mj_1}$) будет находиться раньше, чем в j_2 -м столбце при $j_1 < j_2$ ($1 \leq j_1 < j_2 \leq l$). Очевидно, что это возможно вследствие того, что в каждой строке содержится ровно один дефектный объект.

Теперь оценим сверху математическое ожидание количества индивидуальных проверок. Проверок, необходимых для обнаружения всех дефектных объектов, в матрице $[O'_{ij}]$, ($i = \overline{1, m}, j = \overline{1, l}$). Для этого рассмотрим произвольный j -тый столбец ($j = \overline{1, l-1}$). Ясно, что последний дефектный объект в этом столбце может находиться на i -том ($i = \overline{j, j+m-l}$) месте и при этом для нахождения всех дефектных объектов в этом столбце потребуется i индивидуальных проверок.

Возьмем максимальное количество проверок в каждом столбце

$$\mu_j = j + m - l. \quad (7.4)$$

В l -ом столбце проверок потребуется столько же, сколько и в $(l-1)$ -м столбце в силу способа автоматического распознавания. Действительно, если последний дефектный объект в $(l-1)$ -м столбце находился на i -том месте и для его обнаружения необходимо было m проверок, то в i -том столбце после i проверок будет установлено (так как известно общее число дефектных объектов в столбце), что объекты $O'_{li+1}, O'_{li+2}, \dots, O'_{lm}$ являются дефектными и более проверок не понадобится. Следовательно, математическое ожидание числа проверок в l -м столбце, исходя из (7.4) при $j = l-1$, равно $\mu_l = m - 1$.

Теперь ясно, что оценка сверху числа проверок для обнаружения всех дефектных объектов в первоначальной матрице $[O_{ij}]$, $i = \overline{1, m}, j = \overline{1, n}$, равна

$$H_0(nm) = \frac{1}{2}(2ml - l^2 + l + 2n - 4)$$

и достигает своего максимума при $l = \frac{1}{2}(2m+1)$:

$$H_0^*(nm) = \frac{1}{8}(4m^2 + 4m + 8n - 15).$$

Теперь выпишем формулы для $G_0(nm)$ и количества проверок, приходящихся на один объект совокупности, если поиск дефектных объектов производится групповыми и индивидуальными проверками по группам-столбцам:

$$G_0(nm) = \frac{2ml - l^2 + 2n - 4}{2nm},$$

$$G_0^*(nm) = \frac{4m^2 + 4m + 8n - 15}{8nm}.$$

Теперь возьмем производную от $G_0^*(nm)$ по m и, приравняв ее к нулю, получим, что $G_0^*(nm)$ достигает своего минимума при $m = m_0 = \sqrt{2n - 3.75}$. Так как m может принимать только целые значения, то минимум функции $G_0^*(nm)$ достигается при

$$m_* = \begin{cases} \lfloor m_0 \rfloor, & \text{если } G_0^*(n \lfloor m_0 \rfloor) \leq G_0^*(n(\lfloor m_0 \rfloor + 1)), \\ 1 + \lfloor m_0 \rfloor, & \text{если } G_0^*(n \lfloor m_0 \rfloor) > G_0^*(n(\lfloor m_0 \rfloor + 1)) \end{cases},$$

где $\lfloor \cdot \rfloor$ — целая часть числа.

Следовательно, если разность $G(nm_*)$ и $G_0^*(nm_*)$ положительна, то проведение дополнительных проверок по группам-столбцам при поиске дефектных объектов сокращает общее количество проверок, приходящихся на один объект совокупности, состоящей из n столбцов и m_* строк, в случае, когда известно, что в каждой строке ровно по одному дефектному объекту.

Глава 8. Две задачи выбора процедуры групповых проверок

Здесь будут рассмотрены две задачи выбора процедуры групповых проверок. Первая из них иллюстрирует возможности применения процедур групповых проверок при испытании блока однородных объектов с целью минимизации стоимостных затрат на испытания. Вторая задача является задачей минимизации стоимостной функции затрат при отборе объектов из бесконечной совокупности объектов, если заданы цены за один эксперимент (проверку) и за использование объекта в эксперименте.

Рассмотрим первую постановку задачи. Пусть имеется группа (блок) объектов O_1, O_2, \dots, O_n , каждый из которых с вероятностью p ($0 < p < 1$) является годным и с вероятностью $q=1-p$ является дефектным. Пусть также известны ρ_i ($i = \overline{1, n}$) стоимости проверок группы из i объектов, когда i объектов объединены в одну группу (блок), и ρ_0 – стоимость первой проверки объекта, ρ_{00} – стоимость второй проверки объекта.

Заметим, что ρ_0 не всегда равно ρ_{00} .

Рассмотрим две схемы испытаний.

Схема А. Испытывается каждый объект блока X отдельно. Понятно, что C_A – стоимость испытаний по этой схеме, приходящаяся на один элемент, равна

$$C_A = \rho_1 + \rho_0.$$

Схема В. Испытание на надежность группы из m объектов. При этом группа объектов признается годной, если l ($1 \leq l \leq m$) объектов из этой группы удовлетворяют требованиям годности блока X . В противном случае испытания проводятся индивидуально для каждого объекта из этой группы с целью выявления дефектного объекта или группы объектов и замены их на годные. При проведении испытаний по схеме В C_B – стоимость испытаний, приходящихся на один объект, равна

$$C_B = \rho_m m^{-1} + \rho_1 + \rho_0 + \rho_{00} - \rho^l (\rho_1 + \rho_{00}).$$

Очевидно, что в общем случае m может принимать несколько значений.

Оптимальной схемой испытаний будем называть схему, обеспечивающую минимум стоимости испытаний, приходящихся на один объект блока.

Рассмотрим случай, когда m единственно и $l=m$. Тогда

$$C_B = \rho_m m^{-1} + \rho_{00} + (1 - \rho^m) (\rho_1 + \rho_{00}).$$

Своей задачей будем считать выбор оптимальной схемы испытаний. Очевидно, что стоимость зависит от p . При $C_A > C_B$ оптимальной является схема В, а при $C_A \leq C_B$ оптимальной является схема А. Рассмотрим разность

$$C_A - C_B = \rho_m m^{-1} + \rho_{00} - \rho^m (\rho_1 + \rho_{00}).$$

Следовательно, при

$$p \leq \left[(\rho_m m^{-1} + \rho_{00}) / (\rho_1 + \rho_{00}) \right]^{\frac{1}{m}}$$

оптимальной схемой является схема А и при

$$p > \left[(\rho_m m^{-1} + \rho_{00}) / (\rho_1 + \rho_{00}) \right]^{\frac{1}{m}}$$

оптимальной является схема В.

Теперь рассмотрим случай, когда $m = n$ и $l = n - y$. Легко видеть, что эта схема описывает случай для блока X , состоящего из n однородных объектов. И из n однородных объектов y объектов являются резервными. Стоимость испытаний, приходящаяся на один элемент, в этом случае равна

$$C_B = \rho_n n^{-1} + \rho_0 + (1 - \rho^{n-y}) (\rho_1 + \rho_{00}).$$

Тогда $C_A - C_B = \rho_n n^{-1} + \rho_{00} - \rho^{n-y} (\rho_1 + \rho_{00})$. Вывод: при

$$p \leq \left[(\rho_n n^{-1} + \rho_{00}) / (\rho_1 + \rho_{00}) \right]^{\frac{1}{n-y}}$$

оптимальной является схема А и при

$$p > \left[(\rho_n n^{-1} + \rho_{00}) / (\rho_1 + \rho_{00}) \right]^{\frac{1}{n-y}}$$

оптимальной является схема В.

Далее рассмотрим задачу сравнения проведения испытаний по схеме В блока X с резервными объектами и без резервных объектов (предполагается, что из блока X удаляются y резервных объектов и они испытываются отдельно от блока по схеме А). Тогда в первом случае (блок с резервными объектами) стоимость испытаний, приходящихся на один объект блока X , будет вычисляться, как и ранее, а во втором случае (резервные объекты испытываются отдельно от блока) будет вычисляться по формуле

$$\begin{aligned} C_B^* &= \{ p^{n-y} [\rho_{n-y} + \rho_0 (n-y)] + (1 - \rho^{n-y}) [\rho_{n-y} + \rho_0 (n-y) + (n-y)(\rho_1 + \rho_{00})] + y(\rho_1 + \rho_{00}) \} n^{-1} = \\ &= \rho_{n-y} n^{-1} + \rho_{00} (n-y) n^{-1} + \rho_1 + \rho_0 - p^{n-y} (n-y) n^{-1} (\rho_1 + \rho_{00}). \end{aligned}$$

Рассмотрим разность

$$C_B^* - C_B = n^{-1} (\rho_{n-y} - \rho_n) - y n^{-1} \rho_{00} + y n^{-1} p^{n-y} (\rho_1 + \rho_{00}).$$

Нетрудно видеть, что при

$$p \geq \left[(y^{-1} (\rho_n - \rho_{n-y}) + \rho_{00}) / (\rho_1 + \rho_{00}) \right]^{\frac{1}{n-y}}$$

$C^*_B \geq C_B$, т. е. оптимальной является схема испытания блока с резервными объектами, и при

$$p \langle [(y^{-1}(\rho_n - \rho_{n-y}) + \rho_{00}) / (\rho_1 + \rho_{00})]^{n-y} \rangle$$

оптимальной является схема испытания без резервных объектов (резервные объекты испытываются отдельно от блока).

Таким образом, известная априорная вероятность годности произвольного однородного объекта блока X позволяет планировать испытания блока X и его объектов, минимизируя среднюю стоимость испытаний, приходящихся на один объект.

Рассмотрим вторую задачу. Пусть из бесконечной совокупности объектов O_1, O_2, \dots , каждый из которых с вероятностью p является годным ($0 < p < 1$) и с вероятностью $q = 1 - p$ является дефектным, необходимо отобрать N годных объектов. Считаем, что V_1 – стоимость использования одного объекта в проверке и V_2 – стоимость проверки известны.

Процедуру групповых проверок объектов для отбора годных объектов из бесконечной совокупности построим следующим образом: для групповой проверки объединим в группу n объектов и проверим их совместно (считаем, что проверка определяет, есть или нет в этой группе хотя бы один дефектный объект, но не устанавливает точного количества дефектных объектов). Если в группе из n объектов оказался хотя бы один дефектный объект, то группу всю отвергаем, в противном случае всю группу принимаем. Ясно, что математическое ожидание числа отобранных объектов за одну проверку, как это было показано ранее, равно

$$\mu(\tau) = np^n.$$

Тогда стоимость всех испытаний для отбора N годных объектов из бесконечной совокупности

$$C(n, p, V_1, V_2) = Np^{-n}(V_2 n^{-1} + V_1).$$

Теперь необходимо определить оптимальный объем группы для одной проверки, чтобы он минимизировал общую стоимость испытаний. Выражение для $C(n, p, V_1, V_2)$ продифференцируем по n и приравняем нулю. Обозначим $V_2 V_1^{-1} = V$. Окончательно получим, что $C(n, p, V_1, V_2)$ достигает своего минимума при

$$n_0 = \left[\left(\frac{V}{2} \right)^2 - \frac{V}{\ln p} \right]^{\frac{1}{2}} - \frac{V}{2}.$$

Так как n может принимать только целые значения, то оптимальный объем группы n_* для вышеописанной процедуры групповых проверок будет находиться по следующей формуле:

$$n_* = \begin{cases} \lfloor n_0 \rfloor, & \text{если } C(\lfloor n_0 \rfloor, p, V_1, V_2) \leq C(\lfloor n_0 \rfloor + 1, p, V_1, V_2) \\ \lfloor n_0 \rfloor + 1, & \text{в противном случае} \end{cases},$$

где $\lfloor \cdot \rfloor$ — целая часть числа.

ГЛАВА 9. ОЦЕНИВАНИЕ ПАРАМЕТРОВ

Пусть имеется m групп в каждой по n объектов. Каждый объект независимо от других с вероятностью q ($0 < q < 1$) является дефектным и с вероятностью $p = 1 - q$ – годным. Опишем процедуру групповых проверок Дорфмана.

Каждая из m групп подвергается следующей процедуре. Все n объектов проверяются вместе одновременно. Если в группе не оказалось ни одного дефектного объекта, то вся группа объектов принимается. В противном случае (есть хотя бы один дефектный объект, но точное количество дефектных объектов неизвестно) индивидуальными проверками классифицируются поочередно объекты из группы до обнаружения дефектного объекта. После этого остаток группы проверяется совместно и так далее, до тех пор пока не будут обнаружены все дефектные объекты. Считаем, что для процедуры Дорфмана имеет место принцип автоматического распознавания: если известно, что среди двух объектов есть хотя бы один дефектный и один из этих двух объектов при индивидуальном тестировании оказался годным, то второй без проверки признается дефектным.

Вышеописанную процедуру условно можно разделить на три этапа.

Первый этап – совместная проверка n объектов в каждой из m групп. Все группы на этом этапе делятся на «хорошие» и «плохие».

Хорошей группой назовем, если она не содержит дефектных объектов. Плохой назовем группу, если она содержит, хотя бы один дефектный объект.

Второй этап – нахождение одного дефектного объекта в каждой плохой группе.

Третий этап – нахождение всех дефектных объектов в плохих группах.

Пусть на первом этапе обнаружено z_0 хороших групп и $(m - z_0)$ – плохих групп. Тогда, используя функцию правдоподобия

$$\Phi(\cdot / q) = (1 - p^n)^{m - z_0} p^{nz_0},$$

можно построить оценку максимального правдоподобия для q следующего вида:

$$\psi_1(z_0) = \hat{q} = 1 - \left(\frac{z_0}{m}\right)^{\frac{1}{n}}.$$

Очевидно, что несмещенную оценку на этом этапе можно построить для функций от параметра p^n , например,

$$I_1(z_0) = \hat{p}^n = \frac{z_0}{m}.$$

Пусть по результатам индивидуальных проверок на втором этапе оказалось:

z_1 – число групп, где первой индивидуальной проверкой обнаружен дефектный объект;

z_2 – число групп, где второй индивидуальной проверкой обнаружен дефектный объект,

и так далее;

z_{n-1} – число групп, где $(n-1)$ индивидуальной проверкой обнаружен дефектный объект;

z_n – число групп, где $(n-1)$ индивидуальная проверка обнаружила годные объекты и n -й объект в силу принципа автоматического распознавания признан дефектным.

Очевидно, что $\sum_{i=1}^n z_i = m - z_0$.

Воспользуемся функцией правдоподобия

$$\Phi(\cdot / q) = p^{nz_0} q^{z_1} (pq)^{z_2} \dots (p^{n-1}q)^{z_n}$$

и получим оценку максимального правдоподобия для q как функцию от z_0, z_1, \dots, z_n :

$$\psi_2(z_0, \dots, z_n) = \hat{q} = \frac{m - z_0}{n(z_0 + z_n) + z_1 + 2z_2 + \dots + (n-1)z_{n-1}}.$$

После проведения второго этапа можно построить несмещенную оценку для q и несмещенную оценку дисперсии несмещенной оценки. Верна следующая теорема.

Теорема 9.1. После проведения второго этапа процедуры групповых проверок Дорфмана несмещенная оценка для q и несмещенная оценка дисперсии несмещенной оценки находятся соответственно по формулам

$$I_2(z_1) = \hat{q} = \frac{z_1}{m},$$

$$\hat{D}(I_2(z_1)) = \hat{D}\hat{q} = \frac{z_1(m - z_1)}{m^2(m - 1)}.$$

Доказательство. Опишем процедуру групповых проверок Дорфмана над m группами, в каждой из которых по n объектов, в виде марковского случайного блуждания со следующей матрицей вероятностей перехода:

$$\theta = \begin{matrix} & \begin{matrix} B_0 & B_1 & B_2 & \dots & B_{n-2} & B_{n-1} \end{matrix} \\ \begin{matrix} B_0 \\ B_1 \\ B_2 \\ \dots \\ B_{n-2} \\ B_{n-1} \end{matrix} & \left[\begin{array}{cccccc} B_0 & B_1 & B_2 & \dots & B_{n-2} & B_{n-1} \\ p^n & (1-p^n) & 0 & \dots & 0 & 0 \\ q(1-p^n)^{-1} & 0 & (p-p^n)(1-p^n)^{-1} & \dots & 0 & 0 \\ q(1-p^{n-1})^{-1} & 0 & 0 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots \\ q(1-p^3)^{-1} & 0 & 0 & \dots & 0 & (p-p^3)(1-p^3) \\ 1 & 0 & 0 & \dots & 0 & 0 \end{array} \right], \end{matrix}$$

где состояние B_0 – проверка n объектов совместно;
состояние B_1 – индивидуальная проверка первого объекта;
состояние B_2 – индивидуальная проверка второго объекта,
и так далее;
состояние B_{n-1} – индивидуальная проверка $(n-1)$ -го объекта.

Отметим, что процедура Дорфмана использует принцип автоматического распознавания объекта, поэтому n -й объект в любой из m групп на втором этапе процедуры индивидуально не проверяется.

После проведения второго этапа процедуры Дорфмана можно выписать матрицу наблюдений $T = [t_{ij}]$ ($i, j = \overline{0, n-1}$), где t_{ij} – число переходов марковского случайного блуждания из состояния B_i в состояние B_j . Далее можно выписать матрицу L , отвечающую цепочке событий $B_0 \rightarrow B_1 \rightarrow \dots \rightarrow B_0$, вероятность появления которой равна q . Очевидно, что $L = [l_{ij}]$ ($i, j = \overline{0, n-1}$), где все элементы матрицы равны нулю, кроме $l_{01} = l_{10} = 1$. Если начальным и конечным состоянием блужданий считать состояние B_0 , что соответствует проведению первого и второго этапов процедуры групповых проверок Дорфмана, тогда известно, что несмещенную оценку для q нужно искать по формуле

$$\hat{q} = \frac{K_{00}(T-L)}{K_{00}(T)}, \quad (9.1)$$

где $K_{i_0 j_0}(N) = A_{j_0 i_0}(M) \frac{\prod_{i=0}^{n-1} \omega_i!}{\prod_{i,j=0}^{n-1} n_{ij}!}$; $\omega_i = \sum_{j=0}^{n-1} n_{ij}$, $A_{i_0 j_0}(M)$ – алгебраическое до-

полнение элемента с индексами $(j_0 i_0)$ в матрице M , $M = [m_{ij}]$ ($i, j = \overline{0, n-1}$),

$$m_{ij} = \begin{cases} \delta_{ij}, & \text{если } \omega_i = 0 \\ \delta_{ij} - \frac{n_{ij}}{\omega_i}, & \text{если } \omega_i > 0 \end{cases}, \quad \delta_{ij} - \text{символ Кронекера.}$$

В соответствии с введенными обозначениями матрицу T запишем в следующем виде:

$$T = \begin{bmatrix} z_0 & m - z_0 & 0 & \dots & 0 & 0 \\ z_1 & 0 & m - z_0 - z_1 & \dots & 0 & 0 \\ z_2 & 0 & 0 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots \\ z_{n-2} & 0 & 0 & \dots & 0 & z_{n-1} + z_n \\ z_{n-1} + z_n & 0 & 0 & \dots & 0 & 0 \end{bmatrix}.$$

Далее найдем матрицу $(T - L)$, равную

$$T - L = \begin{bmatrix} z_0 & m - z_0 - 1 & 0 & \dots & 0 & 0 \\ z_1 - 1 & 0 & m - z_0 - z_1 & \dots & 0 & 0 \\ z_2 & 0 & 0 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots \\ z_{n-2} & 0 & 0 & \dots & 0 & z_{n-1} + z_n \\ z_{n-1} + z_n & 0 & 0 & \dots & 0 & 0 \end{bmatrix}.$$

Нетрудно видеть, что составленные матрицы M_T и M_{T-L} по матрицам T и $(T-L)$ имеют алгебраические дополнения к элементам с индексами $(0,0)$, равные единице, т. е.

$$A_{00}(M_T) = A_{00}(M_{T-L}) = 1.$$

К тому же матрицы T и $(T-L)$ отличаются только элементами с индексами $(0,1)$ и $(1,0)$. Таким образом, находим

$$\hat{q} = \frac{(m-1)!(m-z_0-1)!z_0!(m-z_0)!z_1!(m-z_0-z_1)!}{z_0!(m-z_0-1)!(z_1-1)!(m-z_0-z_1)!m!(m-z_0)!} = \frac{z_1}{m}.$$

Первая часть теоремы доказана. Для того чтобы найти $\hat{D}\hat{q}$, построим несмещенную оценку для q^2 . Тогда

$$\hat{D}\hat{q} = [\hat{q}]^2 - (q^2) \quad (9.2)$$

может быть найдена несмещенная оценка дисперсии несмещенной оценки параметра q .

Рассмотрим цепочку событий $B_0 \rightarrow B_1 \rightarrow B_0 \rightarrow B_1 \rightarrow B_0$. Повторим аналогичные рассуждения. Заметим, что в данном случае матрица $(T-L)$ примет следующий вид:

$$T - L = \begin{bmatrix} z_0 & m - z_0 - 2 & 0 & \dots & 0 & 0 \\ z_1 - 2 & 0 & m - z_0 - z_1 & \dots & 0 & 0 \\ z_2 & 0 & 0 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots \\ z_{n-2} & 0 & 0 & \dots & 0 & z_{n-1} + z_n \\ z_{n-1} + z_n & 0 & 0 & \dots & 0 & 0 \end{bmatrix}.$$

Нетрудно убедиться, что

$$(\hat{q}^2) = \frac{z_1(z_1 - 1)}{m(m - 1)}$$

и тогда из (9.1) следует

$$\hat{D}(I_2(z_1)) = \hat{D}\hat{q} = \frac{z_1(z_1 - 1)}{m(m - 1)}.$$

Теорема доказана.

Очевидно, что после проведения третьего этапа процедуры оценка максимального правдоподобия для q может быть найдена по формуле

$$\psi_3(z^*) \equiv \tilde{q} = \frac{z^*}{mn},$$

где z^* — общее число дефектных объектов, обнаруженное после проведения трех этапов процедуры.

После проведения третьего этапа процедуры Дорфмана, как и после проведения второго этапа, можно построить несмещенную оценку для q и несмещенную оценку дисперсии несмещенной оценки q . Справедлива следующая теорема.

Теорема 9.2. После проведения третьего этапа процедуры групповых проверок Дорфмана над m группами, каждая из которых содержит по n объектов, несмещенная оценка q и несмещенная оценка дисперсии несмещенной оценки соответственно равны

$$I_3(z_3) \equiv \hat{q} = \frac{z_1}{m},$$

$$\hat{D}(I_3) = \hat{D}\hat{q} = \left(\frac{z_1}{m}\right)^2 - \frac{z_1^{**} \cdot z_2^*}{m(m - z_0 - z_1)}.$$

Доказательство. Опишем процедуру групповых проверок как марковское случайное блуждание со следующей матрицей вероятностей перехода:

$$\begin{array}{c}
 B_0 \\
 B_1 \\
 B_2 \\
 B_3 \\
 \dots \\
 \dots \\
 \dots \\
 B_{2n-3} \\
 B_{2n-2}
 \end{array}
 \left[\begin{array}{cccccccc}
 B_0 & B_1 & B_2 & B_3 & \dots & B_{2n-3} & B_{2n-2} \\
 p^n & (1-p^n) & 0 & 0 & \dots & 0 & 0 \\
 0 & 0 & \frac{q}{1-p^n} & \frac{p-p^n}{1-p^n} & \dots & 0 & 0 \\
 p^{n-1} & 0 & 0 & 0 & \dots & 0 & 0 \\
 0 & 0 & 0 & 0 & \dots & 0 & 0 \\
 \dots & \dots & \dots & \dots & \dots & \dots & \dots \\
 \dots & \dots & \dots & \dots & \dots & \dots & \dots \\
 \dots & \dots & \dots & \dots & \dots & \dots & \dots \\
 \dots & \dots & \dots & \dots & \dots & \dots & \dots \\
 \frac{(p-p^2)}{1-p^2} & 0 & 0 & 0 & \dots & 0 & \frac{q}{1-p^2} \\
 1 & 0 & 0 & 0 & \dots & 0 & 0
 \end{array} \right],$$

где
 состояние B_0 – проверка n объектов совместно;
 состояние B_1 – индивидуальная проверка первого объекта в группе из n объектов;
 состояние B_2 – проверка $(n-1)$ -го объекта совместно, после того как первый объект в группе оказался дефектным;
 состояние B_3 – индивидуальная проверка второго объекта в группе из n объектов;
 и так далее;
 состояние B_{2n-3} – индивидуальная проверка $(n-1)$ -го объекта в группе;
 состояние B_{2n-2} – индивидуальная проверка n -го объекта в группе.

Введем следующие обозначения:

Z_1^* – число групп, где первый объект оказался дефектным, а остальные $(n-1)$ объект были годными;

Z_1^{**} – число групп, где первый объект оказался дефектным, а среди остальных $(n-1)$ объектов есть хотя бы один дефектный,

и так далее;

Z_{n-2}^* – число групп, где $(n-1)$ объектов оказался дефектным, а последние два объекта годные;

z_{n-2}^{**} — число групп, где $(n-1)$ объект дефектный, а среди двух последних есть хотя бы один дефектный объект.

Тогда после проведения третьего этапа процедуры легко выписать матрицу наблюдений $T=[t_{ij}]$ ($i, j = \overline{0, 2n-2}$), где t_{ij} — число переходов марковского случайного блуждания из состояния B_i в состояние B_j . Далее можно выписать матрицу L , отвечающую цепочке событий $B_0 \rightarrow B_1 \rightarrow B_2$, вероятность наступления которой равна q . Матрица $L = [l_{ij}]$ ($i, j = \overline{0, 2n-2}$) имеет все нулевые элементы, кроме $l_{01}=l_{12}=1$. Если начальным и конечным состоянием марковских случайных блужданий считать состояние B_0 , то несмещенную оценку можно искать по формуле

$$\hat{q} = \frac{K_{20}(T-L)}{K_{00}(T)},$$

где $K_{i_0 j_0}(N)$ определяется, как и ранее. Прделав аналогичные вычисления, что и в теореме 9.1, получим

$$\hat{q} = \frac{z_1^* + z_1^{**}}{m},$$

но так как $z_1 = z_1^* + z_1^{**}$ по построению, то

$$I_3(z_3) \equiv \hat{q} = \frac{z_1}{m}.$$

Для того чтобы получить несмещенную оценку дисперсии несмещенной оценки, нужно найти несмещенную оценку параметра q^2 , с этой целью можно рассмотреть цепочку событий: $B_0 \rightarrow B_1 \rightarrow B_2 \rightarrow B_3 \rightarrow B_4$. Вероятность этой цепочки равна q^2 . Матрица L для этой цепочки будет иметь четыре не нулевых элемента: $l_{01}=l_{12}=l_{23}=l_{34}=1$. Тогда

$$(\hat{q}^2) = \frac{K_{40}(T-L)}{K_{00}(T)},$$

где матрица T определена, как и прежде. Несложные вычисления позволяют

получить $(\hat{q}^2) = \frac{z_1^{**} \cdot z_2^*}{m(m - z_0 - z_1^*)}$. Тогда имеем

$$\hat{D}\hat{q} = \left(\frac{z_1}{m}\right)^2 - \frac{z_1^{**} \cdot z_2^*}{m(m - z_0 - z_1^*)},$$

что и требовалось доказать.

В заключение отметим: несмещенные оценки для q после проведения второго и третьего этапов процедуры групповых проверок Дорфмана равны, несмещенные оценки дисперсий этих несмещенных оценок различны.

СПИСОК ЛИТЕРАТУРЫ

1. *Мирунычев В. П.* Методы и способы поиска неисправностей в радиоэлектронных системах // Вологодские чтения. – 2009. – №73. – С.74–77.
2. *Aiyi L., Chunling L., Zhiwei Zh., Paul S. A.* Optimality of group testing in the presence of misclassification // *Biometrika*. – 2012. – Vol. 99, №1. – P. 245–251.
3. *Burrows P. M.* Improved estimation of pathogen transmission rates by group testing // *Phytopathology*. – 1987. – Vol. 77. – P. 36–365.
4. *Dorfman R.* The detection of defective members of large population // *Annals of Mathematical Statistics*. – 1943. – Vol. 14, № 4. – P. 436–440.
5. *Farrington C. P.* Estimating prevalence by group testing using generalized linear models // *Statistics in Medicine*. – 1992. – Vol. 11, №12. – P. 1591–1597.
6. *Gastwirth J, Hammick P.* Estimation of the prevalence of a rare disease, preserving the anonymity of the subjects by group testing: Application to estimating the prevalence of AIDS antibodies in blood donors // *Journal of Statistical Planning and Inference*. – 1989. – Vol. 22, №1. – P. 15–27.
7. *Gastwirth J.L.* The Efficiency of Pooling in the Detection of Rare Mutations // *American Journal of Human Genetics*. – 2000. – Vol. 67(4). – P. 1036–1039.
8. *Gastwirth J. L., and Johnson W. O.* Screening With Cost-Effective Quality Control: Potential Applications to HIV and Drug Testing // *Journal of the American Statistical Association*. – 1994. – Vol. 89. – P. 972–981.
9. *Gusev A.L.* The optimal number of items in a group for group testing // *Statistics & Probability Letters*. – 2012. – Vol. 82. – P. 2083–2085.
10. *Hsu L.* Group Testing with a Goal in Estimating the Number of Defects Under Imperfect Environmental Stress Screen Levels // *Communications in Statistics - Theory and Methods*. – 2005. – Vol. 34, №6. – P. 1363–1377.
11. *Jones C., Zhigljavsky A.A.* Comparison of costs for multistage group testing methods in the pharmaceutical industry // *Communications in Statistics - Theory and Methods*. – 2001. – Vol. 30:10. – P. 2189–2209.
12. *Katja S. Remlinger, Jacqueline M. Hughes-Oliver, S. Stanley Young, Raymond L. H. Lam.* Statistical Design of Pools Using Optimal Coverage and Minimal Collision // *Institutue of Statistics Mimeo Series 2549*. – 2003.
13. *Kotz S., Shisong M. L.* Effects of inspection errors on curtailed Dorfman-type procedure // *Commun. Statist.: Theory and Method*. – 1986. – Vol. 15, №3. – P. 120–123.

14. *Krook A., Stratton I. M., O'Rahilly S.* Rapid and simultaneous detection of multiple mutations by pooled and multiplex single nucleotide primer extension: application to the study of insulin-responsive glucose transporter and insulin receptor mutations in non-insulindependent diabetes // *Human Molecular Genetics*. – 1992. – Vol. 1, №6. – P. 391–395.
15. *Kumar S.* A group testing problem // *Annals of Mathematical Statistics*. – 1965. – Vol. 36, № 2. – P. 727–728.
16. *Kumar S.* Group-testing to classify all units in a multinomial sample // *Annals of Mathematical Statistics*. – 1968. – Vol. 39, №2. – P. 704.
17. *Langfeldt S. A., Hughes-Oliver J. M., Ghosh, S., and Young, S. S.* Optimal Group Testing in the Presence of Blockers // *Institute of Statistics Mimeograph Series 2297*, North Carolina State University, Dept. of Statistics. – 1997.
18. *Mundel A. B.* Group Testing // *Journal of Quality Technology*. – 1984. – Vol. 16, №4. – P. 181–188.
19. *Pfeifer C. Enis P.* Dorfman-Type Group Testing for a Modified Binomial Model // *Journal of the American Statistical Association*. – 1978. Vol. 73, No. 363. – P. 588–592.
20. *Phatarfod R. M., Sudbury A.* The Use of a Square Array Scheme in Blood Testing // *Statistics in Medicine*. – 1994. – Vol. 13. – P. 2337–2343.
21. *Sterrett A.* On the detection on defective members of large population // *Annals of Mathematical Statistics*. – 1957. – Vol. 28, № 4. – P. 1033–1036.
22. *Swallow W. H.* Group testing for estimating infection rates and On the basis of its superior bias and mean square error probabilities of disease transmission // *Phytopathology*. – 1985. – Vol. 75. – P. 882–889.
23. *Tebbs J. M., McMahan C. S., Bilder C. R.* Two-stage hierarchical group testing for multiple infections with application to the Infertility Prevention Project // *Biometrics*. – 2013. Vol. 69(4). – P. 1064–1073.
24. *Tebbs J.M., Swallow W.H.* Estimating ordered binomial proportions with the use of group testing // *Biometrika*. – 2003. – Vol. 90. – P. 471–477.
25. *Warasi M. S., Tebbs J. M., McMahan C. S., Bilder C. R.* Estimating the prevalence of multiple diseases from two-stage hierarchical pooling // *Statistics in medicine*. – 2016. – Vol. №35(21). – P. 3851–3864.
26. *Xie M., Tatsuoka K., Sacks J., Young S. S.* Group Testing With Blockers and Synergism // *Journal of the American Statistical Association*. – 2001. – Vol. 96. – P. 92–102.

27. *Xiong W.* The optimal group size using inverse binomial group testing considering misclassification // *Communications in Statistics - Simulation and Computation.* – 2016. - Vol. 45, №15. – P. 4600–4610.

28. *Yamamura K.* Estimation of the Proportion of Defective Units by Using Group Testing Under the Existence of a Threshold of Detection // *Communications in Statistics - Simulation and Computation.* – 2007. – Vol. 36, №5. – P. 949–957.

29. *Yi B.* Nonparametric, Parametric and Semiparametric Models for Screening and Decoding Pools of Chemical Compounds, unpublished Ph.D. dissertation, North Carolina State University, Dept. of Statistics. – 2002.

30. *Zarbl H., Aragaki C., Zhao L. P.* An Efficient Protocol for Rare Mutation Genotyping in a Large Population // *Genetic Testing.* – 2009. – Vol. 2, №4.

31. *Zhu L.* Statistical Decoding and Designing of Pooling Experiments Based on Chemical Structure, unpublished Ph.D. dissertation, North Carolina State University, Dept. of Statistics. – 2000.

**КОД ПРОГРАММЫ ДЛЯ МОДЕЛИРОВАНИЯ
ПРОЦЕДУРЫ ДОРФМАНА**

```
import math;
import pandas as pd;
N = 10000; #объем популяции

N_size = 3;
N_lst = [100000, 1000000, 5000000];

#массив вер-й быть годным
p_size = 3;
p = [0.99, 0.995, 0.999];

#массив вероятностей быть деф
q_size = 16;
q = [0.01, 0.02, 0.03, 0.04, 0.05, 0.06, 0.07, 0.08,
0.09, 0.1, 0.12, 0.13, 0.15, 0.2, 0.25, 0.3];

C1 = 1.0; #стоимость проверки одного объекта

#массив стоимостей проверки n объектов
C_size = 10;
Cn = [C1, 1.05 * C1, 1.1 * C1, 1.2 * C1, 1.5 * C1, 2
* C1, 2.5 * C1, 3 * C1, 5 * C1, 10 * C1];
def f(p, n, Cn, C1):
    return float(N) / float(n) * Cn / C1 + (1 - p**n)
* N;

def get_group_size(p, Cn, C1):
    opt_size = 1;
```

```

prev = f(p, opt_size, Cn, C1);
cur = f(p, opt_size + 1, Cn, C1);
while (prev > cur):
    prev = cur;
    opt_size += 1;
    if opt_size == N:
        break;
    cur = f(p, opt_size + 1, Cn, C1);
check = f(p, N, Cn, C1);
if check < prev:
    return 1, check;
else:
    return opt_size, prev;

```

```

def f_cl(p, n):
    return float(N) / float(n) + (1 - p**n) * N;

```

```

def get_group_size_cl(p):
    opt_size = 1;
    prev = f_cl(p, opt_size);
    cur = f_cl(p, opt_size + 1);
    while (prev > cur):
        prev = cur;
        opt_size += 1;
        cur = f_cl(p, opt_size + 1);
    return opt_size, prev;

```

```

def inspections_count(N_, opt_n, p):
    if (opt_n > 1):
        N_new = N_ / opt_n * opt_n;
        part1 = N_ / opt_n + math.ceil(N_new * (1.0
- p**opt_n));
        k = N_ - N_new;

```

```

        if (k > 0):
            return part1 + 1 + math.ceil((1.0 - p**k)
* k);
        else:
            return part1;
    else:
        return N_;

def inspections_cost(N_, opt_n, p, C1, Cn):
    if (opt_n > 1):
        N_new = N_ / opt_n * opt_n;
        part1 = N_ / opt_n * Cn + N_new * (1.0 -
p**opt_n) * C1;
        k = N_ - N_new;
        if (k > 0):
            return part1 + Cn + (1.0 - p**k) * C1 *
k;
        else:
            return part1;
    else:
        return N_ * C1;

def inspections_cost_cl(N_, opt_n, p):
    if (opt_n > 1):
        N_new = N_ / opt_n * opt_n;
        part1 = N_ / opt_n + math.ceil(N_new * (1 -
p**opt_n));
        k = N_ - N_new;
        if (k):
            return part1 + 1 + math.ceil((1.0 - p**k)
* k);
        else:
            return part1;

```



```

else:
    return N_;

def compare_with_Dorfman() :
    res = [];
    for k in range(C_size):
        #print k;
        for i in range(q_size):
            n, Cst = get_group_size(1-q[i], Cn[k],
C1);
            n_cl, EE = get_group_size_cl(1-q[i]);
            res.append([Cn[k] / C1, q[i], n, Cst / N,
n_cl, EE / N]);
        return res;

def modeling():
    res = [];
    for j in range(N_size):
        for i in range(p_size):
            for k in range(C_size):
                n, extra = get_group_size(p[i],
Cn[k], C1);
                insp_cost = inspec-
tions_cost(N_lst[j], n, p[i], C1, Cn[k]);
                insp_count = inspec-
tions_count(N_lst[j], n, p[i]);

                n_cl, extra =
get_group_size_cl(p[i]);
                insp_cost_cl = inspec-
tions_cost_cl(N_lst[j], n, p[i]);

                indiv_insp_cost = C1 * N_lst[j];

```

```
        res.append([N_lst[j], 1-p[i], Cn[k] /
C1, n, insp_count, insp_cost, n_cl,
                    insp_cost_cl, in-
div_insp_cost, indiv_insp_cost / insp_cost]);
    return res;
```

```
res = compare_with_Dorfman();
df = pd.DataFrame(res);
df.columns = ["Cn / C1", "p", "n*", "Cost", "n*_cl",
"C_cl"];
df.to_csv('dorfman.csv', sep=';');
```

```
res2 = modeling();
df = pd.DataFrame(res2);
df.columns = ["N", "p", "Cn / C1", "n*", "Count",
"Cost", "n*_cl", "C_cl", "C_indiv", "Saving"];
df.to_csv('mine.csv', sep=';');
```

ПРИЛОЖЕНИЕ 2

РЕЗУЛЬТАТЫ МОДЕЛИРОВАНИЯ ПРИ ИСПОЛЬЗОВАНИИ ПРОЦЕДУРЫ ДОРФМАНА, МИНИМИЗИРУЮЩЕЙ СТОИМОСТЬ ПРОВЕРОК

Таблица 2.1. Сравнение результатов проверки при проведении групповой и индивидуальной процедур проверок при $q = 0.01$, $N = 1\ 000\ 000$, $C_1 = 1$.

| C_n / C_l | Опт. объем группы | Количество проверок | Стоимость групп. проверок | Стоимость индив. Проверок | Экономия |
|-------------|-------------------|---------------------|---------------------------|---------------------------|----------|
| 1,00 | 11 | 195573 | 195571,65 | 1000000,00 | 5,1 |
| 1,05 | 11 | 195573 | 200117,15 | 1000000,00 | 5,0 |
| 1,10 | 11 | 195573 | 204662,65 | 1000000,00 | 4,9 |
| 1,20 | 12 | 196950 | 213615,6 | 1000000,00 | 4,7 |
| 1,50 | 13 | 199404 | 237864,86 | 1000000,00 | 4,2 |
| 2,00 | 15 | 206609 | 273275,20 | 1000000,00 | 3,7 |
| 2,50 | 17 | 215881 | 304116,17 | 1000000,00 | 3,3 |
| 3,00 | 19 | 226464 | 331726,62 | 1000000,00 | 3,0 |
| 5,00 | 25 | 262179 | 422178,64 | 1000000,00 | 2,4 |
| 10,00 | 38 | 343760 | 580603,69 | 1000000,00 | 1,7 |

Таблица 2.2. Сравнение результатов проверки при проведении групповой и индивидуальной процедур проверок при $q = 0.005$, $N = 1\ 000\ 000$, $C_1 = 1$.

| C_n / C_l | Опт. объем группы | Количество проверок | Стоимость групп. проверок | Стоимость индив. Проверок | Экономия |
|-------------|-------------------|---------------------|---------------------------|---------------------------|----------|
| 1,00 | 15 | 139099 | 139097,80 | 1000000,00 | 7,2 |
| 1,05 | 15 | 139099 | 142431,15 | 1000000,00 | 7,0 |
| 1,10 | 15 | 139099 | 145764,50 | 1000000,00 | 6,9 |
| 1,20 | 16 | 139569 | 152068,9 | 1000000,00 | 6,6 |
| 1,50 | 18 | 141832 | 169608,74 | 1000000,00 | 5,9 |
| 2,00 | 21 | 147534 | 195152,48 | 1000000,00 | 5,1 |
| 2,50 | 24 | 155014 | 217513,41 | 1000000,00 | 4,6 |
| 3,00 | 26 | 160653 | 237576,10 | 1000000,00 | 4,2 |
| 5,00 | 34 | 186107 | 303753,97 | 1000000,00 | 3,3 |
| 10,00 | 51 | 245187 | 421657,64 | 1000000,00 | 2,4 |

Таблица 2.3. Сравнение результатов проверки при проведении групповой и индивидуальной процедур проверок при $q = 0.001$, $N = 1\,000\,000$, $C_1 = 1$.

| C_n / C_1 | Опт. объем группы | Количество проверок | Стоимость групп. проверок | Стоимость индив. Проверок | Экономия |
|-------------|-------------------|---------------------|---------------------------|---------------------------|----------|
| 1,00 | 32 | 62759 | 62758,92 | 1000000,00 | 15,9 |
| 1,05 | 33 | 62783 | 64296,58 | 1000000,00 | 15,6 |
| 1,10 | 34 | 62858 | 65797,94 | 1000000,00 | 15,2 |
| 1,20 | 35 | 62984 | 68697,6 | 1000000,00 | 14,6 |
| 1,50 | 39 | 63912 | 76731,02 | 1000000,00 | 13,0 |
| 2,00 | 46 | 66721 | 88459,78 | 1000000,00 | 11,3 |
| 2,50 | 51 | 69354 | 98765,25 | 1000000,00 | 10,1 |
| 3,00 | 56 | 72346 | 108060,98 | 1000000,00 | 9,3 |
| 5,00 | 73 | 84132 | 138926,95 | 1000000,00 | 7,2 |
| 10,00 | 105 | 109246 | 194961,23 | 1000000,00 | 5,1 |

Таблица 2.4. Сравнение результатов проверки при проведении групповой и индивидуальной процедур проверок при $q = 0.01$, $N = 5\,000\,000$, $C_1 = 1$.

| C_n / C_1 | Опт. объем группы | Количество проверок | Стоимость групп. проверок | Стоимость индив. Проверок | Экономия |
|-------------|-------------------|---------------------|---------------------------|---------------------------|----------|
| 1,00 | 11 | 977856 | 977854,45 | 5000000,00 | 5,1 |
| 1,05 | 11 | 977856 | 1000581,75 | 5000000,00 | 5,0 |
| 1,10 | 11 | 977856 | 1023309,05 | 5000000,00 | 4,9 |
| 1,20 | 12 | 984743 | 1068075,8 | 5000000,00 | 4,7 |
| 1,50 | 13 | 997012 | 1189318,52 | 5000000,00 | 4,2 |
| 2,00 | 15 | 1033043 | 1366375,77 | 5000000,00 | 3,7 |
| 2,50 | 17 | 1079403 | 1520578,46 | 5000000,00 | 3,3 |
| 3,00 | 19 | 1132315 | 1658630,60 | 5000000,00 | 3,0 |
| 5,00 | 25 | 1310894 | 2110893,20 | 5000000,00 | 2,4 |
| 10,00 | 38 | 1718806 | 2903016,53 | 5000000,00 | 1,7 |

Таблица 2.5. Сравнение результатов проверки при проведении групповой и индивидуальной процедур проверок при $q = 0.005$, $N = 5\,000\,000$, $C_1 = 1$.

| C_n / C_1 | Опт. объем группы | Количество проверок | Стоимость групп. проверок | Стоимость индив. Проверок | Экономия |
|-------------|-------------------|---------------------|---------------------------|---------------------------|----------|
| 1,00 | 15 | 695490 | 695488,92 | 5000000,00 | 7,2 |
| 1,05 | 15 | 695490 | 712155,62 | 5000000,00 | 7,0 |
| 1,10 | 15 | 695490 | 728822,32 | 5000000,00 | 6,9 |
| 1,20 | 16 | 697845 | 760344,38 | 5000000,00 | 6,6 |
| 1,50 | 18 | 709154 | 848042,31 | 5000000,00 | 5,9 |
| 2,00 | 21 | 737660 | 975754,48 | 5000000,00 | 5,1 |
| 2,50 | 24 | 775067 | 1087566,86 | 5000000,00 | 4,6 |
| 3,00 | 26 | 803263 | 1187877,64 | 5000000,00 | 4,2 |
| 5,00 | 34 | 930533 | 1518768,60 | 5000000,00 | 3,3 |
| 10,00 | 51 | 1225934 | 2108293,14 | 5000000,00 | 2,4 |

Таблица 2.6. Сравнение результатов проверки при проведении групповой и индивидуальной процедур проверок при $q = 0.001$, $N = 5\,000\,000$, $C_1 = 1$.

| C_n / C_l | Опт. объем группы | Количество проверок | Стоимость групп. проверок | Стоимость индив. Проверок | Экономия |
|-------------|-------------------|---------------------|---------------------------|---------------------------|----------|
| 1,00 | 32 | 313795 | 313794,62 | 5000000,00 | 15,9 |
| 1,05 | 33 | 313904 | 321478,74 | 5000000,00 | 15,6 |
| 1,10 | 34 | 314284 | 328989,43 | 5000000,00 | 15,2 |
| 1,20 | 35 | 314917 | 343486,92 | 5000000,00 | 14,6 |
| 1,50 | 39 | 319548 | 383649,12 | 5000000,00 | 13,0 |
| 2,00 | 46 | 333596 | 442291,63 | 5000000,00 | 11,3 |
| 2,50 | 51 | 346769 | 493827,46 | 5000000,00 | 10,1 |
| 3,00 | 56 | 361723 | 540294,17 | 5000000,00 | 9,3 |
| 5,00 | 73 | 420660 | 694634,96 | 5000000,00 | 7,2 |
| 10,00 | 105 | 546235 | 974813,40 | 5000000,00 | 5,1 |

ТАБЛИЦА ДЛЯ СРАВНЕНИЯ КЛАССИЧЕСКОЙ ПРОЦЕДУРЫ ДОРФМАНА С ПРОЦЕДУРОЙ, В КОТОРОЙ ОПТИМАЛЬНЫЙ ОБЪЕМ МИНИМИЗИРУЕТ СТОИМОСТЬ ПРОВЕРОК

| q | $C_n / C_1 = 1$ | | $C_n / C_1 = 1,05$ | | $C_n / C_1 = 1,1$ | | $C_n / C_1 = 1,2$ | | $C_n / C_1 = 1,5$ | | $C_n / C_1 = 2$ | | $C_n / C_1 = 2,5$ | | $C_n / C_1 = 3$ | | $C_n / C_1 = 5$ | | $C_n / C_1 = 10$ | |
|------|-----------------|------|--------------------|------|-------------------|------|-------------------|------|-------------------|------|-----------------|------|-------------------|------|-----------------|------|-----------------|------|------------------|------|
| | n | C | n | C | n | C | n | C | n | C | n | C | n | C | n | C | n | C | n | C |
| 0,01 | 11 | 0,20 | 11 | 0,20 | 11 | 0,20 | 12 | 0,21 | 13 | 0,24 | 15 | 0,27 | 17 | 0,30 | 19 | 0,33 | 25 | 0,42 | 38 | 0,58 |
| 0,02 | 8 | 0,27 | 8 | 0,28 | 8 | 0,29 | 8 | 0,30 | 9 | 0,33 | 11 | 0,38 | 13 | 0,42 | 14 | 0,46 | 19 | 0,58 | 30 | 0,79 |
| 0,03 | 6 | 0,33 | 7 | 0,34 | 7 | 0,35 | 7 | 0,36 | 8 | 0,40 | 9 | 0,46 | 11 | 0,51 | 12 | 0,56 | 16 | 0,70 | 28 | 0,93 |
| 0,04 | 6 | 0,38 | 6 | 0,39 | 6 | 0,40 | 6 | 0,42 | 7 | 0,46 | 8 | 0,53 | 10 | 0,59 | 11 | 0,63 | 15 | 0,79 | 1 | 1,00 |
| 0,05 | 5 | 0,43 | 5 | 0,44 | 5 | 0,45 | 6 | 0,46 | 6 | 0,51 | 8 | 0,59 | 9 | 0,65 | 10 | 0,70 | 14 | 0,87 | 1 | 1,00 |
| 0,06 | 5 | 0,47 | 5 | 0,48 | 5 | 0,49 | 5 | 0,51 | 6 | 0,56 | 7 | 0,64 | 8 | 0,70 | 9 | 0,76 | 14 | 0,94 | 1 | 1,00 |
| 0,07 | 4 | 0,50 | 5 | 0,51 | 5 | 0,52 | 5 | 0,54 | 6 | 0,60 | 7 | 0,68 | 8 | 0,75 | 9 | 0,81 | 14 | 1,00 | 1 | 1,00 |
| 0,08 | 4 | 0,53 | 4 | 0,55 | 4 | 0,56 | 5 | 0,58 | 5 | 0,64 | 6 | 0,73 | 8 | 0,80 | 9 | 0,86 | 1 | 1,00 | 1 | 1,00 |
| 0,09 | 4 | 0,56 | 4 | 0,58 | 4 | 0,59 | 4 | 0,61 | 5 | 0,68 | 6 | 0,77 | 7 | 0,84 | 8 | 0,90 | 1 | 1,00 | 1 | 1,00 |
| 0,10 | 4 | 0,59 | 4 | 0,61 | 4 | 0,62 | 4 | 0,64 | 5 | 0,71 | 6 | 0,80 | 7 | 0,88 | 8 | 0,94 | 1 | 1,00 | 1 | 1,00 |
| 0,12 | 4 | 0,65 | 4 | 0,66 | 4 | 0,68 | 4 | 0,70 | 5 | 0,77 | 6 | 0,87 | 7 | 0,95 | 1 | 1,00 | 1 | 1,00 | 1 | 1,00 |
| 0,13 | 3 | 0,67 | 4 | 0,69 | 4 | 0,70 | 4 | 0,73 | 5 | 0,80 | 6 | 0,90 | 7 | 0,98 | 1 | 1,00 | 1 | 1,00 | 1 | 1,00 |
| 0,15 | 3 | 0,72 | 3 | 0,74 | 3 | 0,75 | 4 | 0,78 | 4 | 0,85 | 6 | 0,96 | 1 | 1,00 | 1 | 1,00 | 1 | 1,00 | 1 | 1,00 |
| 0,20 | 3 | 0,82 | 3 | 0,84 | 3 | 0,85 | 3 | 0,89 | 4 | 0,97 | 1 | 1,00 | 1 | 1,00 | 1 | 1,00 | 1 | 1,00 | 1 | 1,00 |
| 0,25 | 3 | 0,91 | 3 | 0,93 | 3 | 0,94 | 3 | 0,98 | 1 | 1,00 | 1 | 1,00 | 1 | 1,00 | 1 | 1,00 | 1 | 1,00 | 1 | 1,00 |
| 0,30 | 3 | 0,99 | 1 | 1,00 | 1 | 1,00 | 1 | 1,00 | 1 | 1,00 | 1 | 1,00 | 1 | 1,00 | 1 | 1,00 | 1 | 1,00 | 1 | 1,00 |

**КОД ПРОГРАММЫ ДЛЯ МОДЕЛИРОВАНИЯ
ПРОЦЕДУРЫ СТЕРРЕТТА**

```

import math;

p_size = 32;
p = [0.38, 0.35, 0.32, 0.3, 0.27, 0.26, 0.25, 0.23, 0.2,
0.15, 0.14, 0.13, 0.12, 0.11, 0.1, 0.09, 0.08, 0.07, 0.06,
0.05, 0.04, 0.03, 0.02, 0.01, 0.007, 0.005, 0.003, 0.001];

N = 100; #объем популяции
C1 = 1.0; #стоимость проверки одного объекта

def Cost_(k) :
    for i in range (2, k+1):
        for n in range (0, i+1):
            if (n == 0):
                tbl[n][i] = Cn(i);
            else:
                summ = 0;
                for j in range (1, i-n+1):
                    prod = 1;
                    for l in range (1, j+1):
                        prod = prod * float(i - (l + n - 1))
/ float(i - l + 1);
                    summ = summ + prod * float(n) / float(i
- j) * ((j + 1) * Cn(1) + tbl[n-1][i-j-1]);
                    tbl[n][i] = Cn(i) + float(n) / float(i) *
(Cn(1) + tbl[n-1][i-1]) + summ;

def Pr(k, n, p):
    return math.factorial(k) / math.factorial(n) / math.fac-
torial(k-n) * p**n * (1-p)**(k-n)

def get_group_size(init_size, p) :
    opt_size = init_size;

```

```

prev = S(opt_size - 1, p);
cur = S(opt_size, p);
while (prev > cur):
    print opt_size;
    prev = cur;
    opt_size += 1;
    cur = S(opt_size, p);
return opt_size - 1, prev;

def Cost2(n, k) :
    return float(n) / float(n + 1) * k + n + 1 + float(n) /
float(n + 1) - 2 * float(n) / float(k);

def E(k, p) :
    sm = 0;
    fact_k = math.factorial(k);
    for i in range(0, k+1):
        pr = fact_k / math.factorial(i) / math.factorial(k-
i) * p**i * (1-p)**(k-i);
        cst = Cost2(i, k);
        sm += pr * cst;
    return float(N) / float(k) * sm;

def get_group_size_cl(init_size, p) :
    opt_size = init_size;
    prev = E(opt_size - 1, p);
    cur = E(opt_size, p);
    while (prev > cur):
        print opt_size;
        prev = cur;
        opt_size += 1;
        cur = E(opt_size, p);
    return opt_size - 1, prev;

def get_result_set (costs, p) :
    n_init = 3;
    res = [];
    i = 1;

```



```

while i < 99 and costs[i][1] > costs[i+1][1]:
    i = i + 1;

if i == 99:
    return p, 1, N, C1 * N;
else:
    return p, costs[i][0], E(costs[i][0], p),
costs[i][1];

#fibonacci
fib = [1, 2];
for i in range(2, 10000):
    fib.append(fib[i-1] + fib[i-2]);

fib_rev = [1];
for i in range(1, 1000):
    fib_rev.append(fib_rev[i-1] + 1.0/fib[i]);

def C4(n):
    return fib_rev[n-1];

#lin 0.5
def C1(k) :
    return C1 * (1 + 0.5 * (k - 1));

#lin 0.25
def C2(k) :
    return C1 * (1 + 0.25 * (k - 1));

#seq 1/n
seq = [1];
for i in range(1, 1000):
    seq.append(seq[i-1] + 1.0 / (i+1));

def C3(n):
    return seq[n-1];

import pandas as pd;

```

```

res = [];
tmp = [];
res2 = [];
kk = 200;

n = kk + 1;
m = kk + 1;
tbl = [];
for i in range(n):
    tbl.append([0] * m);

tbl[1][1] = 1;
Cost_(kk);

for k in range (p_size):
    csts = [];
    tmp = [p[k]];
    for i in range (1, kk+1):
        sm = 0.0;
        for j in range (0, kk+1):
            if j <= i:
                sm = sm + tbl[j][i] * Pr(i, j, p[k]);
        csts.append([i, float(sm * N) / float(i)]);
        tmp.append(float(sm * N) / float(i));
    res2.append(tmp);
    p_, k_, e_, c_ = get_result_set(csts, p[k]);
    res.append([p_, k_, e_, c_]);

df = pd.DataFrame(res);
df.columns = ["p", "k", "E", "Cost"];
df.to_csv('gr_05.csv', sep=';');

df2 = pd.DataFrame(res2);
df2.to_csv('gr_05_all.csv', sep=';');

```

Учебное издание

**Гусев Андрей Леонидович
Лесникова Юлия Юрьевна**

**Теория принятия решений.
Процедуры групповых проверок**

Учебное пособие

Редактор *Л. А. Богданова*
Корректор *Л. И. Семицветова*
Компьютерная верстка: *А. Л. Гусев*

Объем данных 3,05 Мб
Подписано к использованию 27.12.2021

Размещено в открытом доступе
на сайте www.psu.ru
в разделе НАУКА / Электронные публикации
и в электронной мультимедийной библиотеке ELiS

Издательский центр
Пермского государственного
национального исследовательского университета
614990, г. Пермь, ул. Букирева, 15