ВВЕДЕНИЕ

Статистические методы используются во многих областях науки. И многие дисциплины, имеющие дело с реальной жизнью (вроде экономики, медицины и т.п.), требуют обработки различных числовых данных. Числовые данные в свою очередь являются объектом изучения математической статистики. Так вот одним из методов работы с такими данными является проверка гипотез.

В данной работе мною был рассмотрен последовательный тест для проверки двух простых гипотез. В последовательной проверке, в отличие от других способов, количество наблюдений, необходимых для принятия решения, зависит от результатов самих наблюдений, и, по факту, является случайной величиной. Работает данный метод следующим образом: после каждого наблюдения требуется принять решение – принять одну из гипотез или провести дополнительное наблюдение. Если одна из двух гипотез принята – проверка заканчивается, если нет – производится следующее наблюдение. На основе уже произведённых наблюдений принимается решение в каждом следующем случае повторного наблюдения до принятия одной из гипотез.

Преимущество последовательного анализа заключается в отсутствии необходимости работать со всей выборкой целиком, поскольку она формируется в ходе самого эксперимента. Таким образом, для работы со случайными величинами достаточно лишь знать, в соответствии с каким распределением приходят случайные величины, что может быть полезно при синхронной обработке данных (т.е. сразу при их поступлении), чтобы можно было уточнять параметры распределения в процессе работы чего-либо, не дожидаясь её завершения. В частности, такая ситуация может сложиться в теории массового обслуживания. Можно также отметить, что простейший поток в теории называется потоком Пуассона (число событий, выпадающих на интервал некоторой длины распределено по Закону Пуассона), а в данной работе были рассмотрены как раз простые гипотезы о параметре распределения Пуассона.

Кроме того, в последствии было исследовано поведение теста для полиномиально распределённой случайной величины, и, на её примере, подробнее рассмотрено, каким образом исходные параметры влияют на работу теста, а также какие существуют способы повышения эффективности работы последовательного теста.

ГЛАВА 1 ПОСТАНОВКА ЗАДАЧИ. ВВОДНЫЕ ОПРЕДЕЛЕНИЯ

1.1 Постановка задачи

1. Реализовать на компьютере последовательный тест для проверки двух простых гипотез о параметре распределения Пуассона.

2. Оценить методом Монте-Карло вероятности ошибочных решений реализованного теста по схеме:

* задали ;
* вычислили значения порогов;
* запустили 1000 раз тест при справедливой гипотезе ();
* подсчитали число неверных решений, разделили на 1000.

3. Оценить методом Монте-Карло математические ожидания числа случайного наблюдений до остановки теста при справедливых нулевой и первой гипотезах по схеме:

* задали;
* вычислили значения порогов;
* запустили 1000 раз тест при справедливой гипотезе ();
* подсчитали суммарное число наблюдений, разделили на 1000.

4. Построить таблицу полученных результатов при разных α, β из множества {0.1, 0.05, 0.01}. Сравнить полученные результаты.

5. Реализовать модель тестов с искажениями во входных данных, построить графики, отражающие поведение тестов в условиях данных искажений.

1) Искажения вида

а)

б)

где ;

2) Искажения вида

а)

б)

таким образом математическое ожидание распределений, генерирующих помехи, совпадает с таковым у тестируемого распределения.

6. По полученным графикам проанализировать поведение тестов в зависимости от параметров, влияющих на степень искажения входных данных, сравнить две модели помех.

7. Применить данные методы исследования к другим, более сложным распределениям, например, полиномиальному.

## 1.2 Актуальность и практическая значимость

Статистические методы применяются в огромном числе различных видах деятельности, подразумевающих обработку числовых данных. И данная работа позволяет получить представление о том, каким образом повысить вероятность принять верную из гипотез.

В данной работе исследовались гипотезы о параметре распределения Пуассона, которое находит применение в т.ч. и в изучаемой нами теории массового обслуживания, а также во множестве других сфер научной и практической деятельности.

## 1.3 Общие теоретические положения

Поскольку мы исследуем последовательный тест двух простых гипотез, начнём с основных определений:

Простая гипотеза – статистическая гипотеза, однозначно определяющая распределение (предположение о значении некоторого параметра; в нашем случае – параметр распределения Пуассона).

Уровень значимости теста — вероятность отклонить гипотезу , если на самом деле она верна.

Последовательная проверка (тест) двух простых гипотез – метод, при котором после каждого испытания принимается, по ранее сформулированному правилу, одно из трёх решений: 1) принять гипотезу ; 2) отклонить гипотезу ; 3) продолжить наблюдение. Причём каждое следующее наблюдение основывается на всех предыдущих. В дальнейшем при подсчёте наблюдений учитывались количества именно этих дополнительных наблюдений.

нулевая гипотеза, конкурирующая. Тогда:

Ошибка первого рода – отвержение верной гипотезы .

Ошибка второго рода – принятие ложной гипотезы (отвержение верной ).

Искажения – часть входных данных, распределённых не по предполагаемому закону, а по некоторой другой модели, будь то другой параметр распределения или всё распределение вовсе.

## 1.4 Общая модель задачи

В данной работе сначала рассматривается Пуассоновская модель распределения случайных величин, поэтому в первых трёх главах работа проводится с дискретными случайными величинами. Далее было рассмотрено поведение последовательных тестов на полиномиальном распределении случайных величин, применены схожие методы исследования с поправкой на отличия вида отдельно взятых случайных величин.

На вход подаются гипотезы (некоторые их значения) и их уровни значимости. Затем эти гипотезы проверяются (по правилу\*) на некоторой сгенерированной последовательности случайных величин, распределённых по закону Пуассона с параметром, равным одной из проверяемых гипотез.

Далее с помощью метода Монте-Карло (описан в условии) получаем оценки числа неверно принятых решений и математического ожидания числа наблюдений (всего проводится 1000 испытаний).

И наконец, строим таблицу полученных значений вероятностей ошибок первого и второго рода для (всего 9 значений).

\*Правило(алгоритм) проверки гипотез:

Пусть есть гипотеза, что , – гипотеза, что . Тогда:

1. По заданным вычисляем значения порогов (:

2. Вычисляем значения функции плотности вероятности от проверяемых гипотез и очередной случайной величины:

3. Считаем

где – вероятность получения выборки, когда справедлива гипотеза соответственно (.

4. Принимаем решение:

Если – отклоняем гипотезу ;

Если – принимаем гипотезу ;

Иначе (если ) – проводим дополнительное наблюдение. Далее в работе под количеством наблюдений подразумевается именно количество этих дополнительных наблюдений. Чтобы перейти от них к общему числу – достаточно лишь добавить число запусков теста.

Данный процесс и называется последовательным критерием отношений вероятности для проверки гипотезы относительно . Он был рассмотрен во второй части работы.

Во третьей части было рассмотрено поведение тестов Вальда в случае появления искажений двух видов. В первом случае искажения были представлены двумя моделями (либо распределены по тому же закону с другим параметром, либо сразу шли в тест):

а)

б)

где ;

Чтобы понять, есть в наблюдении искажение или нет, была введена вспомогательная случайная величина . Если , то искажений нет. Если – есть искажение. Очевидно, что чем больше , тем больше искажений.

В четвёртой части был рассмотрен другой тип искажений, для их генерации использовались другие целочисленные распределения – Биномиальное и Геометрическое:

а)

б)

В последней части были рассмотрена проблематика прошлых частей, где в качестве основного распределения было взято полиномиальное распределение и проведено через аналогичные этапы исследования.

ГЛАВА 2 ПОСЛЕДОВАТЕЛЬНЫЕ СТАТИСТИЧЕСКИЕ ТЕСТЫ В АНАЛИЗЕ СЛУЧАЙНЫХ ВЕЛИЧИН, РАСПРЕДЕЛЕННЫХ ПО ЗАКОНУ ПУАССОНА

Для каждого набора параметров было проведено по 5 тестов, – полученные вероятности ошибки первого (второго) рода соответственно; – матожидания числа дополнительных наблюдений до принятия одной из гипотез.

Итак, перейдём к полученным результатам:

1) , результаты в таблице 2.1:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  |  |  |  |
| Тест 1 | 0.026 | 0.016 | 0.436 | 0.594 |
| Тест 2 | 0.022 | 0.019 | 0.471 | 0.558 |
| Тест 3 | 0.022 | 0.023 | 0.453 | 0.568 |
| Тест 4 | 0.019 | 0.019 | 0.446 | 0.587 |
| Тест 5 | 0.019 | 0.016 | 0.463 | 0.552 |
| **средние** | **0.0216** | **0.0186** | **0.4538** | **0.5718** |

Таблица 2.1

Отклонения уровней значимости от реальных вероятностей ошибок:

Стоит отметить, что вероятность ошибки понижается с уменьшением причём зачастую независимо (таблица 2.2):

|  |  |  |  |
| --- | --- | --- | --- |
| α β | 0.1 | 0.05 | 0.01 |
| 0.1 |  |  |  |
| 0.05 |  |  |  |
| 0.01 |  |  |  |

Таблица 2.2

2) – увеличили разницу м-у :

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  |  |  |  |
| Тест 1 | 0.019 | 0.035 | 0.277 | 0.582 |
| Тест 2 | 0.015 | 0.045 | 0.288 | 0.587 |
| Тест 3 | 0.016 | 0.035 | 0.292 | 0.547 |
| Тест 4 | 0.015 | 0.045 | 0.312 | 0.586 |
| Тест 5 | 0.011 | 0.042 | 0.313 | 0.62 |
| **средние** | **0.0152** | **0.0404** | **0.2964** | **0.5844** |

Таблица 2.3

Отклонения уровней значимости от реальных вероятностей ошибок:

Интересно, что в первом случае как и здесь, но фактическая вероятность ошибки первого рода в первом пункте несколько выше, чем у ошибок второго рода; тогда как здесь разница уже почти 3 раза и имеет тот же знак, что и между оценками. Также стоит отметить, что с увеличением выросло и его отклонение от реального значения.

3) – здесь мы уменьшили разницу между гипотезами (оценки оставили как в пункте 1):

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  |  |  |  |
| Тест 1 | 0.059 | 0.095 | 68.395 | 77.506 |
| Тест 2 | 0.064 | 0.083 | 71.242 | 78.627 |
| Тест 3 | 0.064 | 0.073 | 73.081 | 78.28 |
| Тест 4 | 0.057 | 0.086 | 71.042 | 76.33 |
| Тест 5 | 0.061 | 0.094 | 69.198 | 78.993 |
| **средние** | **0.061** | **0.0862** | **70.5916** | **77.9472** |

Таблица 2.4

Отклонения уровней значимости от реальных вероятностей ошибок:

Что интересно, здесь фактические значения стали ближе к теоретическим. Видим также существенный скачок среднего числа наблюдений до принятия одной из гипотез – поскольку при меньшей разнице между параметрами, сумма делает меньшие шаги из области, при получении значения в которой требуется произвести очередное наблюдение.

Затем мы снова рассмотрели влияние уровней значимости – здесь наибольшая точность оценок получилась при , а не 0.01 как в первом тесте. Но разница с фактическим значением уменьшается по мере уменьшения , это легко видеть хотя бы потому, что наибольшая разница всегда для , а наименьшая получилась при меньших их значениях в обоих тестах.

|  |  |  |  |
| --- | --- | --- | --- |
| α β | 0.1 | 0.05 | 0.01 |
| 0.1 |  |  |  |
| 0.05 |  |  |  |
| 0.01 |  |  |  |

Таблица 2.5

Сравним также отклонения диагональных элементов этой таблицы

с соответствующими из первого теста:

Легко видеть, что во всех трёх случаях отклонения оценок от реальных результатов в первом тесте больше, чем во втором. Значит уменьшение разницы между гипотезами существенно повышает точность оценок (уровней значимости).

ГЛАВА 3 ПОСЛЕДОВАТЕЛЬНЫЕ ТЕСТЫ ДЛЯ НАБЛЮДЕНИЙ С ИСКАЖЕНИЯМИ ИЗ РАСПРЕДЕЛЕНИЙ ТОГО ЖЕ КЛАССА И С ФИКСИРОВАННЫМ ВЫБРОСОМ

Рассмотрим теперь поведение тестов, в случае появления искажений.

Общие параметры, взятые для проведения всех тестов этого раздела:

Поскольку шаг , то для каждого тест запускался 50000 раз, поэтому накопленные ошибки не могут превысить этот порог.

**3.1. Искажения первого типа**

Здесь распределение имеет следующий вид:

Как и ранее, рассмотрим два случая, когда верна одна из гипотез:

* Верна гипотеза

Количество ошибок первого рода растёт с увеличением тем быстрее, чем больше разница (рисунок 3.1, справа).

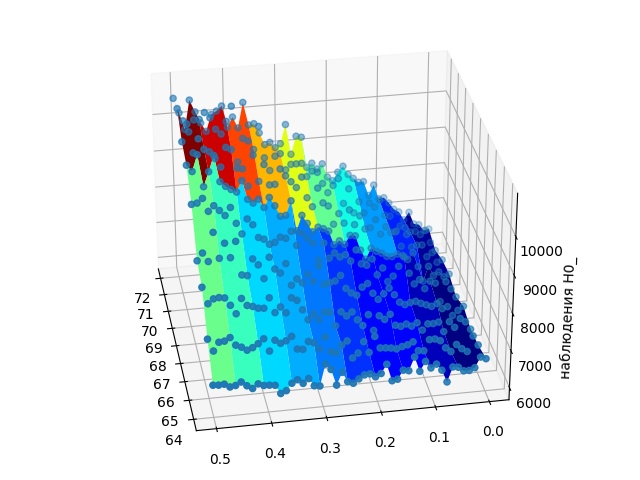
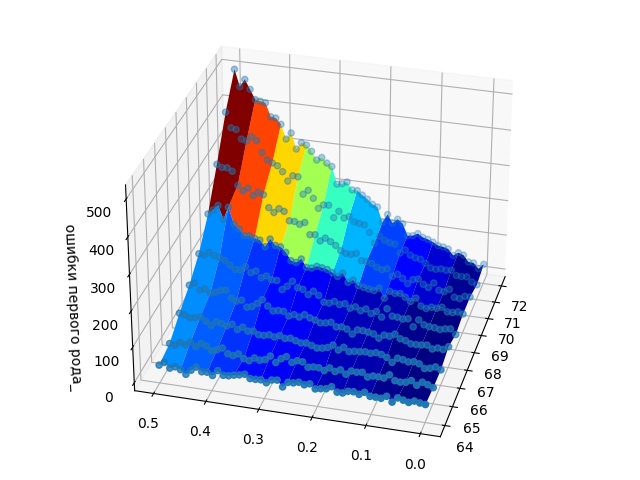
**

Рисунок 3.1 – результаты работы тестов при верной первой гипотезе

Причём, если , то, по сути, у нас в модели нет искажений, поэтому вне зависимости от , количество ошибок не меняется. Также можно обратить внимание на накопленные наблюдения и ошибки – с увеличением зависимость количества ошибок от постепенно переходит от линейной к экспоненциальной (рисунок 3.2, справа):

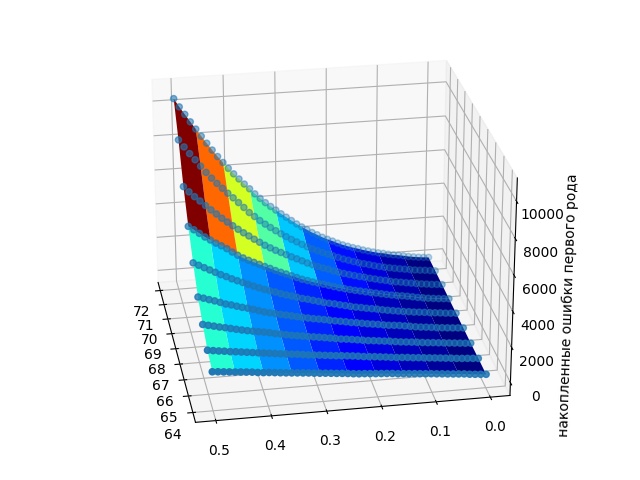
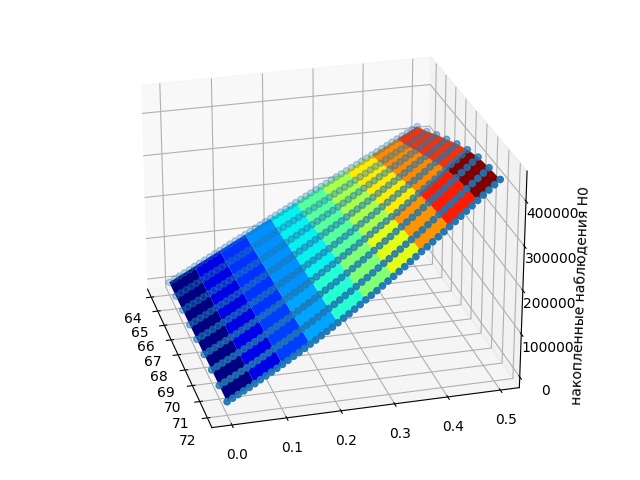


Рисунок 3.2 – накопленные результаты работы тестов при верной первой гипотезе

Количество же наблюдений растёт по тем быстрее, чем больше . Зависимость количества наблюдений от близка к линейной, с увеличением меняется, по сути, угол наклона прямой, что также можно видеть и по накопленным наблюдениям.

* Верна конкурирующая гипотеза

Имеем в целом симметричную ситуацию относительно : можно наблюдать те же тренды, количества наблюдений и ошибок практически совпадают:

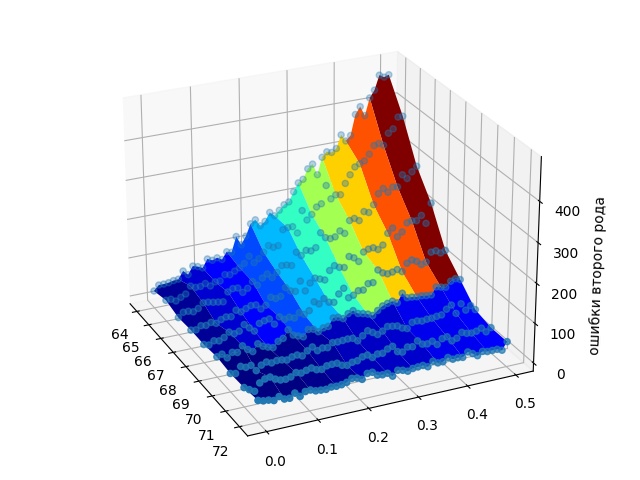
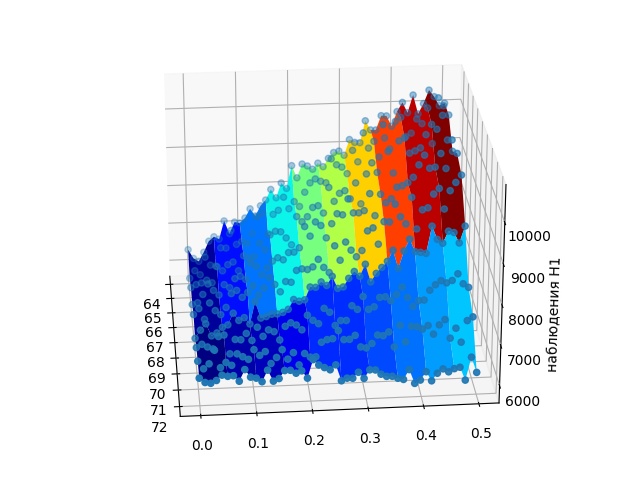


Рисунок 3.3 – результаты работы тестов при верной второй гипотезе

**3.2. Искажения второго типа**

Здесь распределение имеет следующий вид:

т.е. с вероятностью в качестве плотности засоряющего распределения добавляется дельта-функция Дирака в точке .

* Верна гипотеза , имеем:

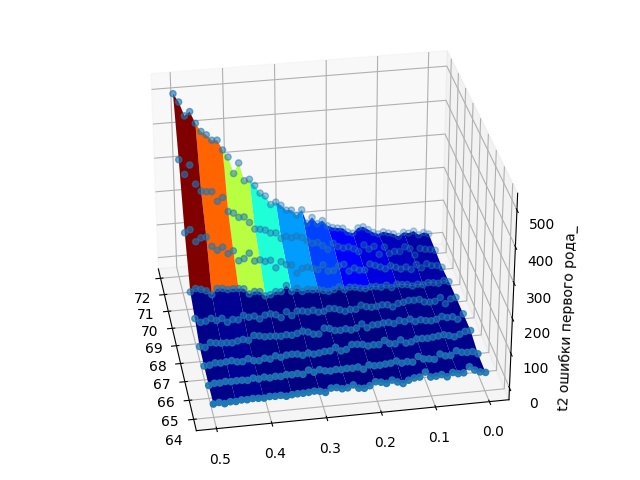
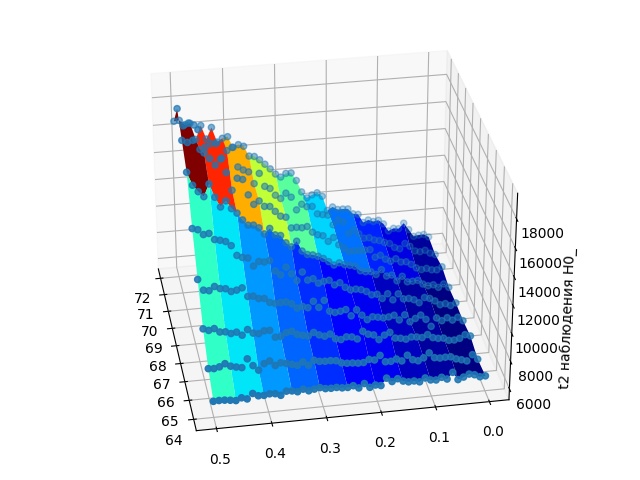
****

Рисунок 3.4 – результаты тестов при верной первой гипотезе

Ситуация схожая, однако здесь для принятия решения потребовалось гораздо больше наблюдений (порой почти вдвое больше), а также количество ошибок возрастает заметно резче, чем в предыдущей модели искажений.

* Верна конкурирующая гипотеза , имеем:

Ситуация снова симметрична относительно в обоих случаях (рисунок 3.5).

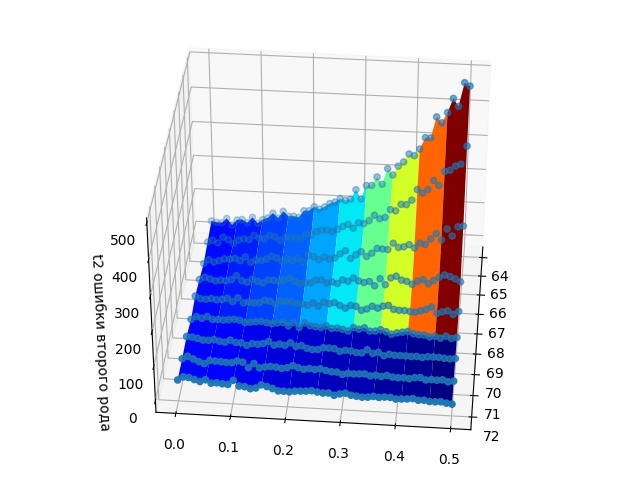
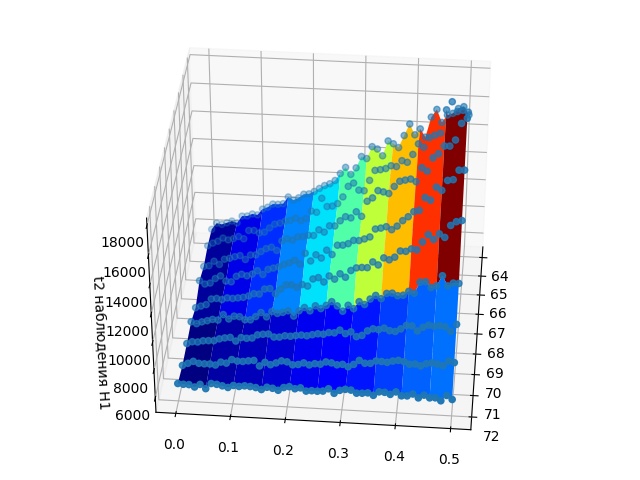
****

Рисунок 3.5 – результаты тестов при верной второй гипотезе

Однако здесь уже просматривается некоторый «изгиб» графика количеств наблюдений от линейного к экспоненциальному, а также ярче выражено аналогичное явление для ошибок (рисунок 3.6):

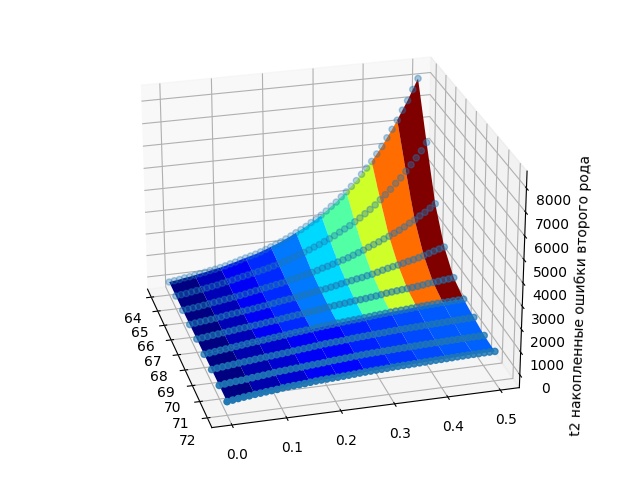
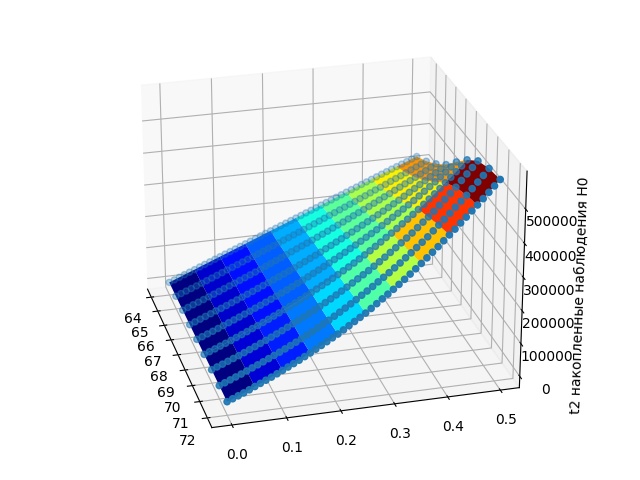
****

Рисунок 3.6 – накопленные результаты тестов при верной второй гипотезе

В накопленных значениях также можно увидеть подтверждение того, что потребовалось большее число наблюдений и что количество ошибок растёт более резко. Однако, стоит также отметить, что, во сколько раз накопленное число наблюдений больше, чем для первой модели искажений, во столько раз накопленное число ошибок меньше во второй (~1/4), таким образом в данной модели тест реже ошибался за счёт увеличения числа наблюдений.

Можно также заметить, что здесь тест Вальда достаточно устойчив к искажениям во входных данных — количество ошибок серьёзно увеличивается только по приближении количества искажений к половине от всех данных, что также можно наблюдать и по средним значениям параметров (таблица 3.1):

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Модель искажений 1 | | | | Модель искажений 2 | | | |
| θ\* | верна H0: θ=64 | | верна H1: θ=72 | | верна H0: θ=64 | | верна H1: θ=72 | |
| наблю  дения | ошибки 1  рода | наблю  дения | ошибки 2 рода | наблю  дения | ошибки 1 рода | наблю  дения | ошибки 2 рода |
| 64 | 6412.08 | 29.88 | 8582.26 | 197.62 | 6454.02 | 15.1 | 10778.7 | 162.2 |
| 65 | 6761.74 | 34.94 | 8407.42 | 166.7 | 6937.14 | 15.22 | 10692.46 | 115.34 |
| 66 | 7110.68 | 46.04 | 8180.1 | 132.06 | 7558.18 | 18.24 | 10318.48 | 71.18 |
| 67 | 7513.72 | 61.84 | 8006.14 | 100.04 | 8297.98 | 22.2 | 9602.16 | 43.74 |
| 68 | 7927.88 | 80.78 | 7675.8 | 76.58 | 9246.76 | 29.48 | 8842.74 | 28.34 |
| 69 | 8138.2 | 107.98 | 7254.64 | 58.42 | 10156.26 | 44.44 | 7960.12 | 21.72 |
| 70 | 8435.64 | 137.74 | 6898.3 | 46.36 | 10874.98 | 75.88 | 7340.92 | 18.3 |
| 71 | 8641.14 | 175.5 | 6548.3 | 36.1 | 11220.42 | 119.08 | 6711.48 | 15.4 |
| 72 | 8676.28 | 212.3 | 6242.58 | 29.56 | 11372.58 | 170.92 | 6255.74 | 14.8 |

Таблица 3.1 - средние количества наблюдений и ошибок 1/2 рода

ГЛАВА 4 ПОСЛЕДОВАТЕЛЬНЫЕ ТЕСТЫ С ИСКЖЕНИЯМИ, РАСПРЕДЕЛЕННЫМИ ПО ДРУГИМ ЗАКОНАМ

Здесь мы проверим 2 гипотезы в условиях наличия искажений, распределённых по некоторому другому закону. Для начала вспомним, как ведёт себя модель без искажений во входных данных. Как и ранее, тест был запущен 1000 раз, для каждого запуска считались ошибки и дополнительные наблюдения, рассмотренные гипотезы и заданные уровни значимости:

Результаты для каждой из гипотез приведены на рисунках 4.1 и 4.2:

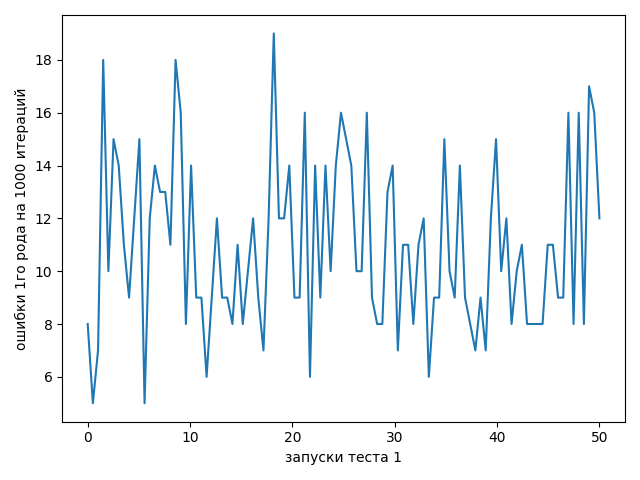
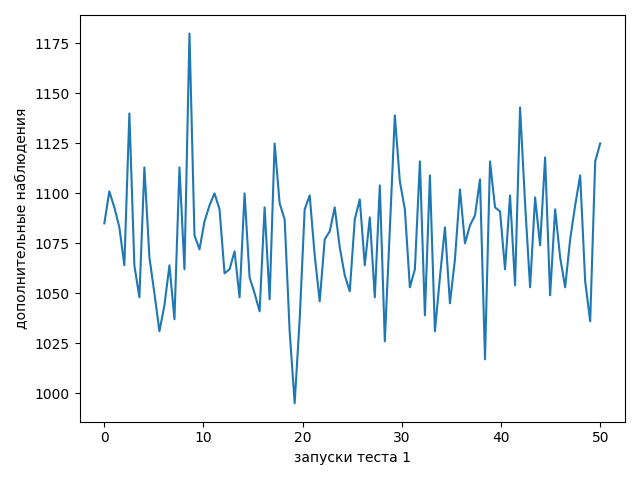


Рисунок 4.1 – результаты при первой верной гипотезе

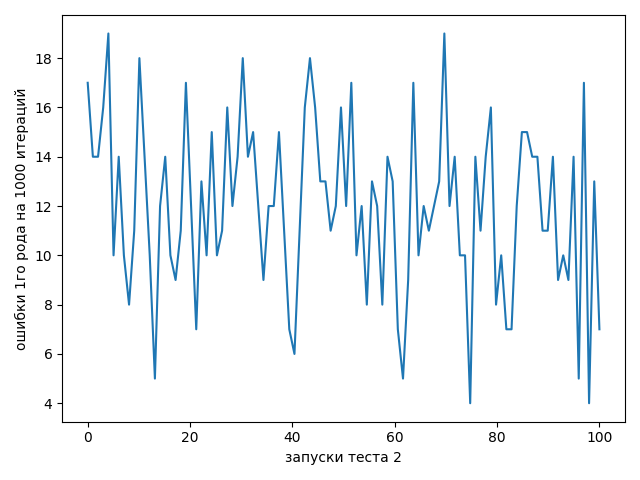
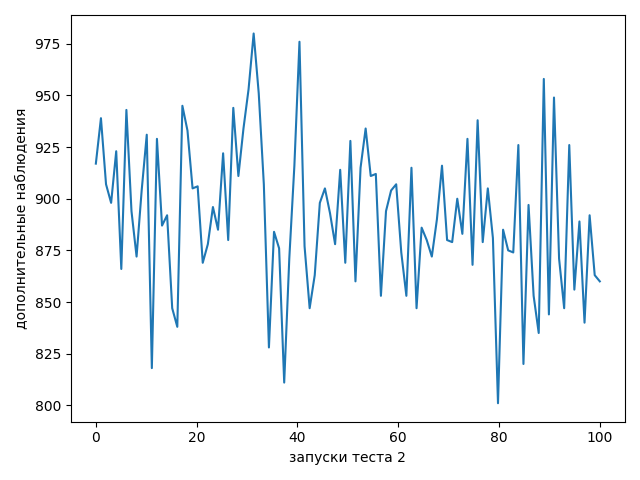


Рисунок 4.2 – результаты при верной второй гипотезе

Видим, что процент ошибочных решений от общего числа проведённых тестов примерно соответствует уровню значимости, как и было установлено во второй главе.

## 4.1. Искажения первого типа

В этом случае распределение моделируется следующим образом:

В первой рассмотренной модели с искажениями засоряющие данные были распределены по биномиальному закону, результаты на рисунке 4.3:

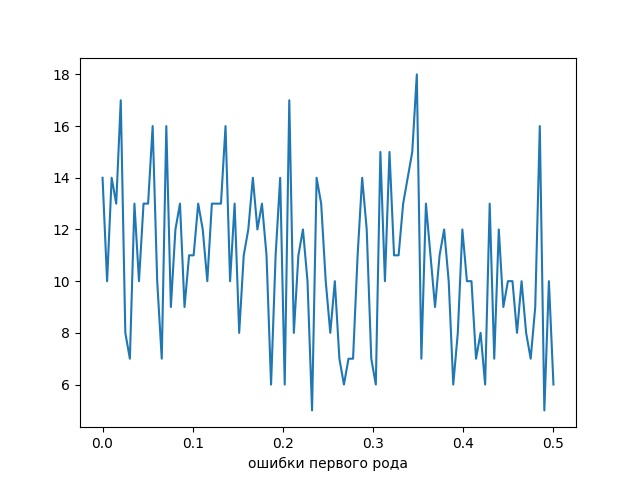
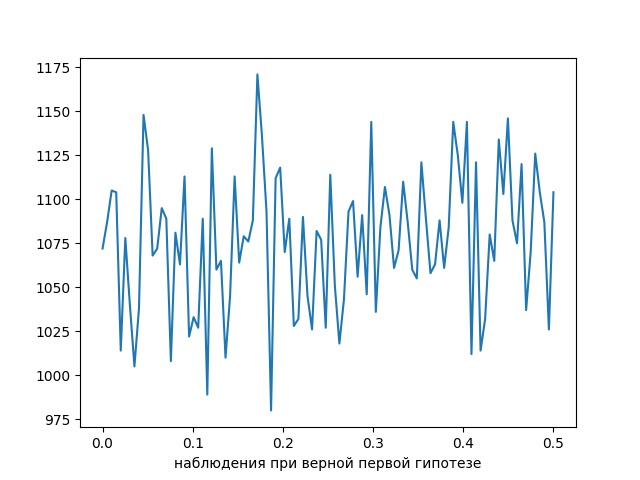


Рисунок 4.3 – зависимость результатов от числа искажений, распределенных биномиально

В целом видим аналогичную картину (если верна другая гипотеза, результаты такие же), это можно объяснить схожестью оригинального распределения (Пуассона) и генерирующего помехи (Биномиального):

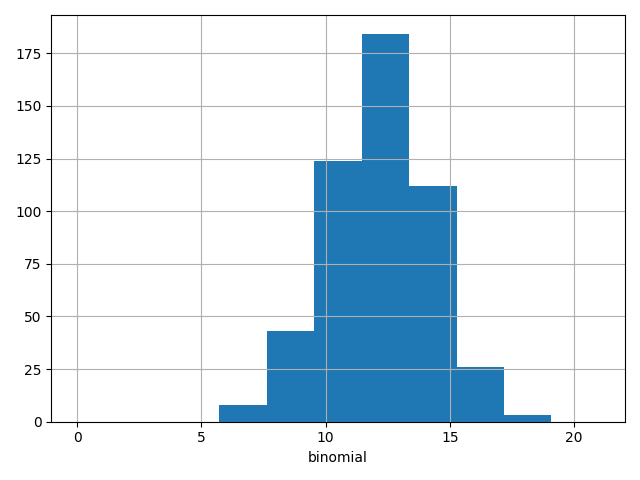
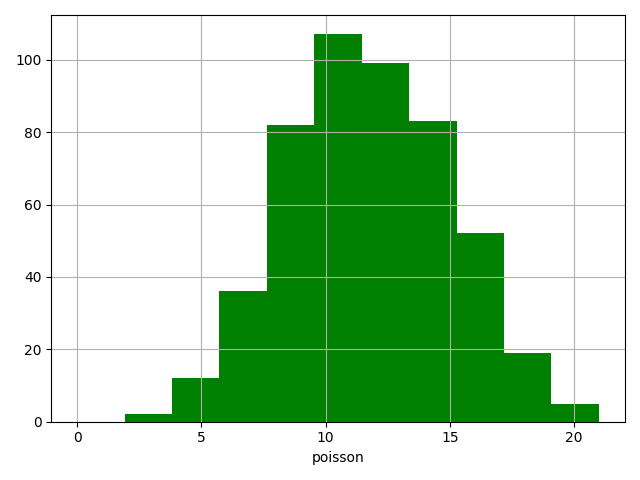


Рисунок 4.4 – сравнение выборок распределений Пуассона (слева) и биномиального (справа)

В данном случае для простоты и наглядности сравнения, у рассматриваемых распределений (и геометрического), определено матожидание, равное 12, выборка – 1000 случайных значений.

А вот с геометрическим распределением помех дела обстоят несколько интереснее (рисунок 4.5):

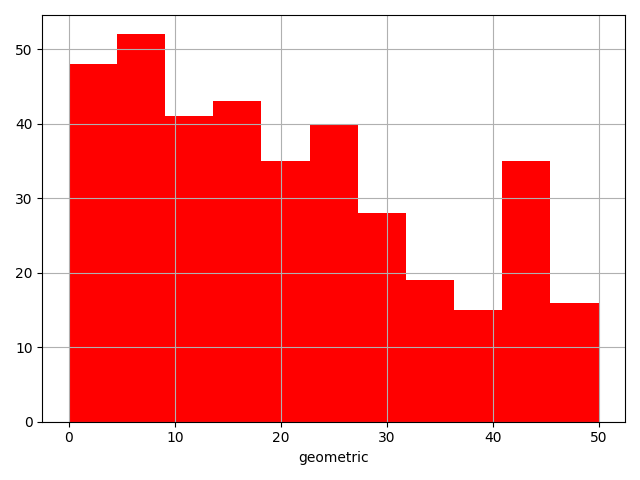


Рисунок 4.5 – графическое представление выборки, распределённой по геометрическому закону

## 4.2. Искажения второго типа

Распределение имеет вид

Рассмотрим поведение тестов при значениях параметров

.

Если верная первая гипотеза:

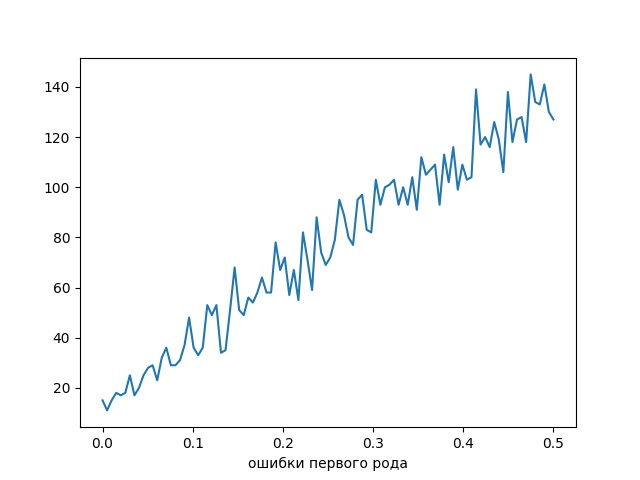
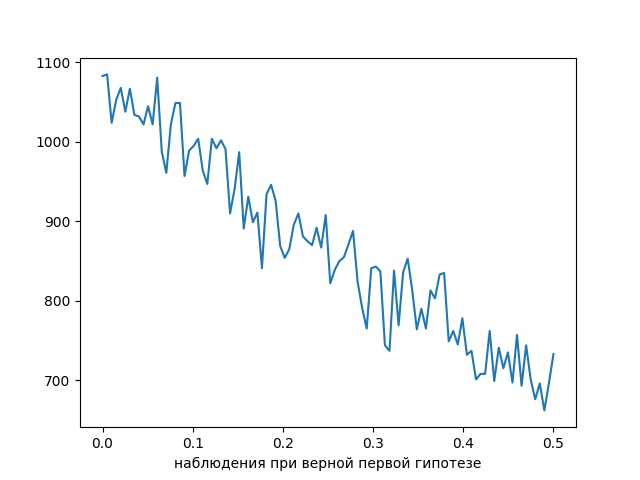


Рисунок 4.6 – зависимость значений параметров от количества искажений, верна

Здесь видим тренд снижения количества наблюдений и роста количества ошибок с ростом количества искажённых случайных величин, таким образом тест принимал решение раньше и заметно чаще ошибался.

И при верной второй гипотезе:

Имеем схожий тренд при меньшем числе наблюдений и большем числе ошибок (рисунок 3.7).

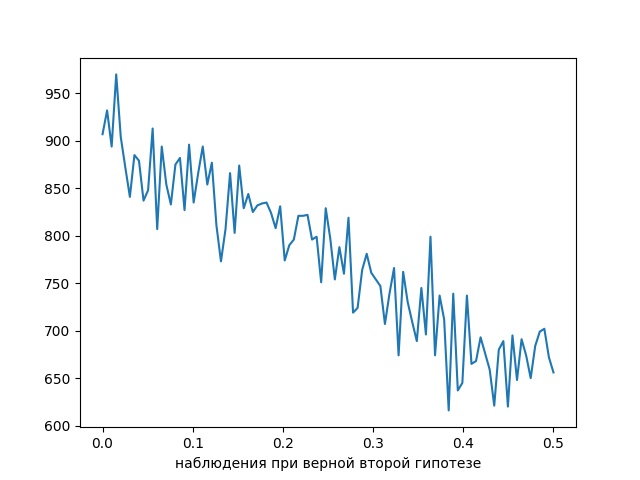
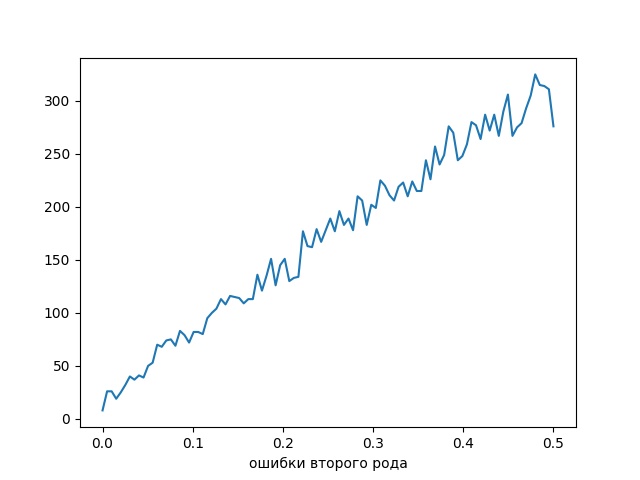


Рисунок 4.7 – зависимость значений параметров от количества искажений, верна

Как видим, кроме тренда по количеству искажённых данных, есть также тренд по значению второй гипотезы – чем оно больше, тем больше ошибок второго рода:

При значениях параметров и верной первой гипотезе:

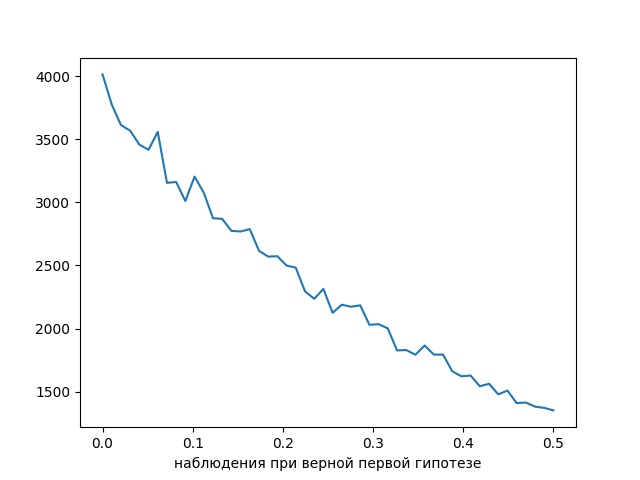
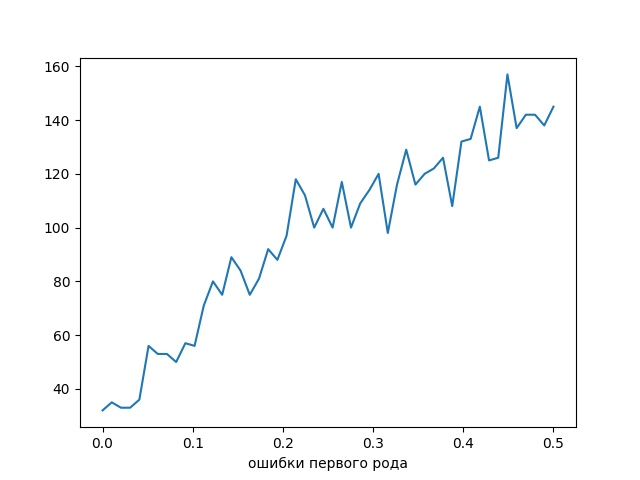
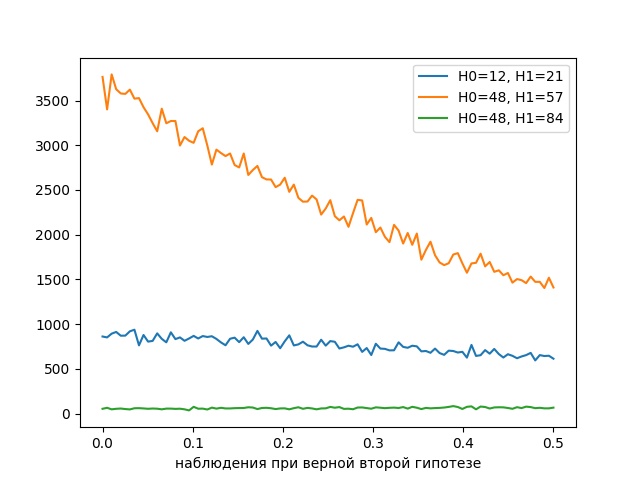


Рисунок 4.8 – зависимость значений параметров от количества искажений, верна

На рисунке 4.9 видим, что для первой гипотезы изменилось лишь число наблюдений, в то время как для второй, при относительно таком же числе наблюедний, резко выросло число ошибочных решений (жёлтая линия). Если же увеличить их пропорционально, то резко упадёт количество наблюдений при относительном сохранении количества ошибочных решений (зелёная линия), а как ранее было установлено – это и ожидается при увеличении разницы между гипотезами:



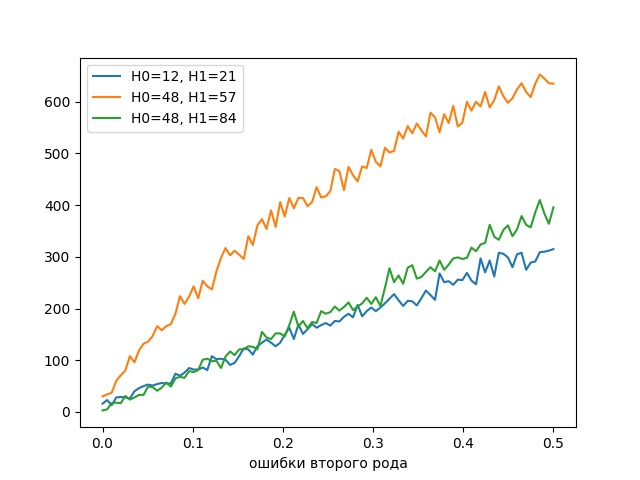


Рисунок 4.9 – сравнительные результаты для различных наборов параметров

Такое поведение можно объяснить относительной «скученностью» около нуля случайных значений, распределённых по геометрическому закону.

ГЛАВА 5 ПОСЛЕДОВАТЕЛЬНЫЕ СТАТИСТИЧЕСКИЕ ТЕСТЫ В АНАЛИЗЕ ДАННЫХ, РАСПРЕДЕЛЁННЫХ ПО ПОЛИНОМИАЛЬНОМУ ЗАКОНУ. ЭФФЕКТИВНОСТЬ ПОСЛЕДОВАТЕЛЬНЫХ ТЕСТОВ

В данном разделе для проведения дальнейший исследований было выбрано полиномиальное распределение и на его примере подробно изучено, какие параметры распределения сказываются на поведении последовательных тестов и каким образом это влияние выражено. За основу для проведения исследования была взята модель распределения “игральная кость” – вектор из 6 значений, каждое задаёт вероятность выпадения каждой из 6 граней кубика. В качестве первой гипотезы был выбран “идеальный” кубик, конкурирующая гипотеза варьировалась.

## 5.1 Исследование полиномиального распределения

Для проведения исследований в этом классе случайных величин, нам нужно несколько модифицировать полученный ранее алгоритм, ведь сейчас случайная величина представлена вектором, а не числом. Теперь, плотность вероятности считается следующим образом:

где – очередная случайная величина и гипотеза соответственно, представленные векторами длины 6, по количеству граней кости; – число бросков кости, задаётся перед проведением теста. Остальные формулы те же. Аналогично для конкурирующей гипотезы.

Итак, на первом шаге были рассмотрены различные варианты конкурирующей гипотезы, как и ранее, согласно методу Монте-Карло, тест запускался 1000 раз, в результате каждого теста записывалось число дополнительных наблюдений до принятия одной из гипотез, а также число ошибочных решений. Во всех случаях нулевая гипотеза имела вид

Полученные в первом случае результаты отображены в таблице 4.1:

| Средние значения характеристик, набор A | | | |
| --- | --- | --- | --- |
| **Верна гипотеза** | **Наблюдения** | **Ошибки** | **Вероятность ошибки** |
| 1 | 3693.746154 | 22.307692 | 0.022308 |
| 2 | 3410.823077 | 25.707692 | 0.025708 |

Таблица 5.1 – Характеристики для 1-го набора параметров

Здесь, в связи с увеличением разницы между гипотезами, закономерно падает число ошибочных решений и дополнительных наблюдений. Полученные в этом случае результаты отображены далее в таблице 4.2:

| Средние значения характеристик, набор Б | | | |
| --- | --- | --- | --- |
| **Верна гипотеза** | **Наблюдения** | **Ошибки** | **Вероятность ошибки** |
| 1 | 1128.676923 | 16.538462 | 0.016538 |
| 2 | 931.246154 | 16.092308 | 0.016092 |

Таблица 5.2 – Характеристики для 2-го набора параметров

В данном случае были получены результаты близкие набору параметров а), однако с меньшим числом дополнительных наблюдений, приведены в таблице 4.3:

| Средние значения характеристик, набор В | | | |
| --- | --- | --- | --- |
| **Верна гипотеза** | **Наблюдения** | **Ошибки** | **Вероятность ошибки** |
| 1 | 2411.876923 | 22.061538 | 0.022062 |
| 2 | 2380.084615 | 23.100000 | 0.023100 |

Таблица 5.3 – Характеристики для 3-го набора параметров

Как и можно было предположить, от порядка вероятностей выпадения полученный результат работы теста не зависит, приведён в таблице 4.4:

| Средние значения характеристик, набор Г | | | |
| --- | --- | --- | --- |
| **Верна гипотеза** | **Наблюдения** | **Ошибки** | **Вероятность ошибки** |
| 1 | 2419.607692 | 22.146154 | 0.022146 |
| 2 | 2371.615385 | 22.746154 | 0.022746 |

Таблица 5.4 – Характеристики для 4-го набора параметров

Таким образом, в дальнейшем за основную была взята модель из случая а), т.к. характеристики устойчивости теста в ней были выражены наиболее ярко.

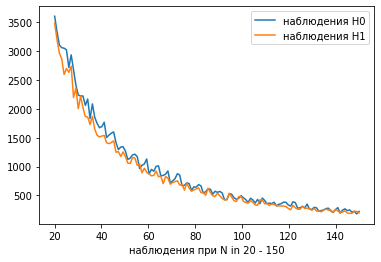
На втором шаге было рассмотрено влияние уровней значимости на поведение теста, результаты представлены в таблице 4.5:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Уровни значимо-сти | Верна первая гипотеза | | | Верна вторая гипотеза | | |
| наблюдения | ошибки | вероятность  ошибки | наблюдения | ошибки | вероятность  ошибки |
| a = 0.05,  b = 0.05 | 3691.369 | 22.054 | 2.2% | 3420.269 | 26.085 | 2.6% |
| a = 0.05,  b = 0.1 | 2747.369 | 23.4 | 2.34% | 3111.661 | 54.99 | 5.499% |
| a = 0.1,  b = 0.05 | 3421.438 | 44.285 | 4.43% | 2620.876 | 27.885 | 2.79% |
| a = 0.05,  b = 0.01 | 5667.384 | 22.485 | 2.25% | 3571.907 | 5.531 | 0.55% |
| a = 0.01,  b = 0.05 | 3849.492 | 4.862 | 0.49% | 5142.876 | 26.48 | 2.65% |
| a = 0.1,  b = 0.01 | 5437.492 | 46.37 | 4.6% | 2783.946 | 5.1 | 0.51% |

Таблица 5.5 – Средние значения характеристик при различных уровнях значимости

Как и ожидалось, количество принимаемых ошибочных решений непосредственно связано с уровнями значимости и сразу реагирует на малейшие их изменения – во сколько раз отличаются уровни значимости, примерно во столько же раз отличаются полученные вероятности принятия ошибочного решения.

На третьем шаге была рассмотрена зависимость поведения тестов от количества бросков кубика для получения отдельной случайной величины, полученные характеристики представлены на рисунке 5.1:

****

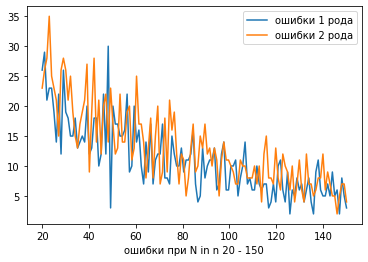
****

Рисунок 5.1 – Зависимость характеристик от числа бросков кости

Видим явную тенденцию к повышению точности и эффективности теста при увеличении количества бросков кубика. Таким образом здесь мы можем влиять на эффективность принятия решений тестом, не меняя как гипотез, так и самого теста. На практике это значит, что чем лучше подготовим данные – тем лучше будет работать последовательный тест. В данном случае можно установить примерную планку для хорошей работы теста в 80-100 бросков кубика – далее количество ошибок снижается не так значительно, зато заметно возрастает количество вычислений и риск переполнения, ведь формула плотности вероятности использует факториал.

На четвёртом шаге было рассмотрено поведение теста в условиях наличия искажений, распределеных со значением параметра, пробегающим значения из отрезка между двумя гипотезами:

Данный отрезок разбивался на 10 частей, количество искажений изменялось от 0 до половины наблюдаемых значений за 50 шагов. Количество бросков кости – . Также в силу большого количества вычислений, итераций было взято 500 вместо 1000, результаты представлены на рисунке 5.2:

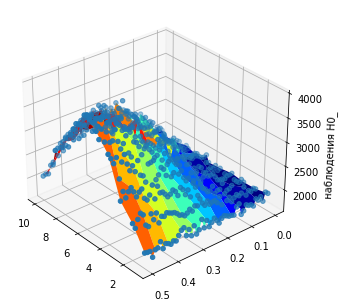
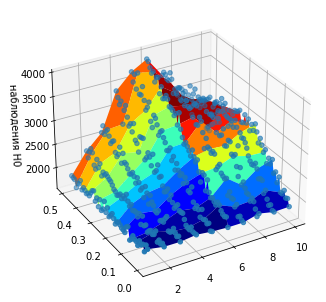
****

Рисунок 5.2 – Зависимость числа наблюдений от числа помех и значения параметра искажающего распределения

Здесь, видим отличный от распределения Пуассона тренд на увеличение количества наблюдений с увеличением числа искажений: тем ярче он выражен, чем ближе к середине отрезка между значениями проверяемых гипотез значение параметра искажающего распределения. В случае распределения Пуассона, количество наблюдений увеличивалось по мере приближения значения параметра искажающего распределения к значению параметра конкурирующей гипотезы.

Среднее число ошибок растёт достаточно равномерно по мере отклонения значения параметра распределения помех от , рисунок 5.3:

Рисунок 5.3 – Измеренная средняя вероятность ошибки 1-го рода

Общее число ошибок первого рода в целом соответствует полученным ранее результатам – имеем те же тренды как по количеству, так и по «силе» искажений, полученные характеристики представлены на рисунке 5.4:

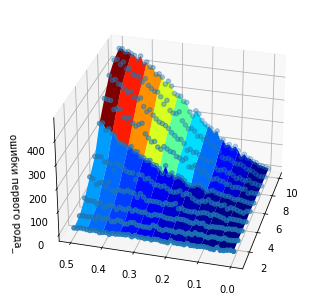
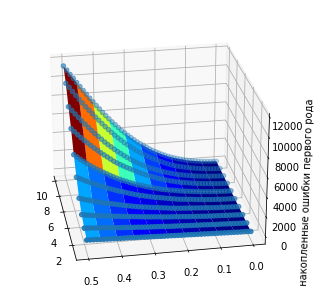
 

Рисунок 5.4 – Количественные характеристики ошибок 1-го рода на 500 запусков теста

Стоит заметить, что несмотря на аналогичные распределению Пуассона результаты в отсутствии искажений (количество ошибок ~2.5% при уровнях значимости ), здесь в худшем случае имеем до 90% ошибочных решений – из 500 итераций было получена максимально 461 ошибка, что свидетельствует о низкой устойчивости последовательного теста к искажениям в случае полиномиального распределения.

Для верной второй гипотезы ситуация в целом аналогичная, с несколько большим числом наблюдений и, что важнее и выражено ярче, большим числом ошибочных решений, рисунок 5.5:

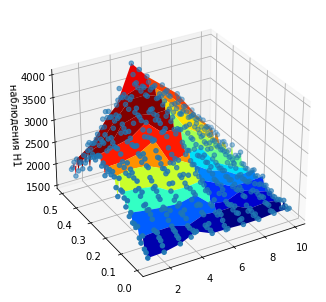
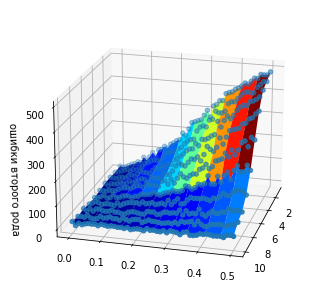
 

Рисунок 5.5 – Полученные характеристики при верной второй гипотезе, N = 20

Также стоит отметить, что увеличение количества бросков здесь уже не помогает, ведь оно влияет и на число искажений. При и 1000 запусках получили (рисунок 5.6):

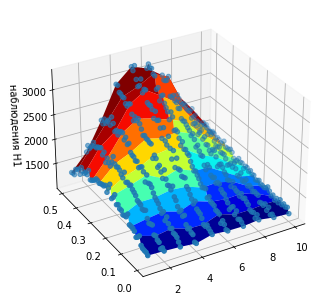
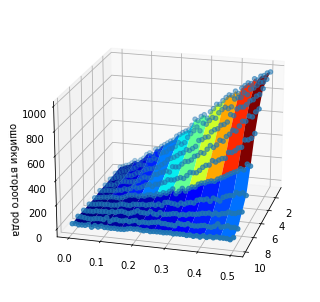
 

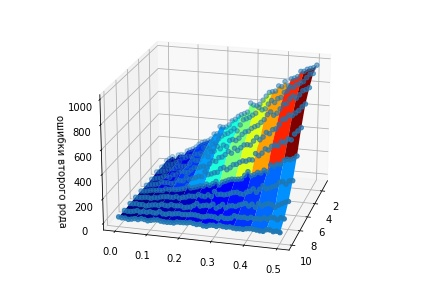
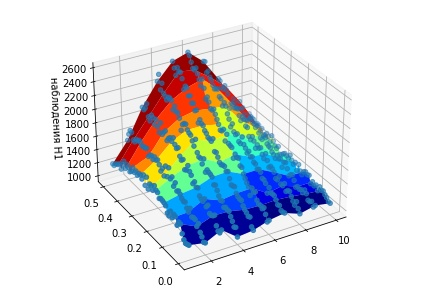
Рисунок 5.6 – Полученные характеристики при верной второй гипотезе, N = 50

Таким образом встаёт вопрос поиска других способов повышения эффективности теста.

## 5.2 Эффективность последовательных тестов

На последнем шаге предыдущей части была получена достаточно низкая эффективность теста, а значит, стоит рассмотреть способы модификации теста. Для начала, поскольку количество наблюдений растёт достаточно резко, а серьёзного эффекта от этого не наблюдается, попробуем ввести «ускоритель теста» – коэффициент, на который будем домножать каждое очередное значение в формуле из первой главы и который ускорит выход значения расчётного параметра из зоны безразличия. Причём этот коэффициент будет становиться тем больше, чем больше дополнительных наблюдений проведено. Будем проводить эксперименты при и числе итераций равном 1000, верна .

Попробуем для начала взять геометрическую прогрессию в качестве этого коэффициента с множителем 1.1, полученные результаты на рисунке 5.7:

Рисунок 5.7 – Полученные характеристики с «ускоряющим коэффициентом»

Видим уменьшение числа наблюдений (до 400 для самого большого количества наблюдений) при относительном сохранении ситуации в отношении ошибок. Значит всё-таки имеет смысл рассмотреть зависимость количества наблюдений от «ускоряющего коэффициента.

Зададим параметры модели для проведения эксперимента. Зафиксируем в качестве искажающей величины значение – середина отрезка , т.е. на предыдущем рисунке это линии из точек в районе значений 5-6 по нижней целочисленной оси. Там мы имели до 2500 наблюдения и до 400 ошибок второго рода в зависимости от числа искажений. Рассмотрим теперь вместо изменения значения параметра, искажающего распределение изменение «ускоряющего коэффициента». Для проверки был взят вектор таких коэффициентов

В полученных результатах отчётливо видно снижение дополнительных наблюдений, в нашем случае заметный эффект достигается для значения коэффициента 2.5. Значит с точки зрения наблюдений, коэффициент работает как нужно, результаты представлены на рисунке 5.8:

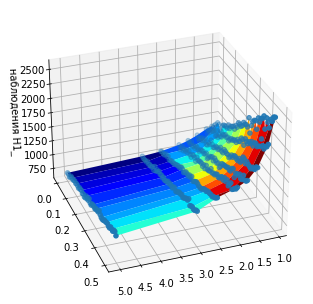
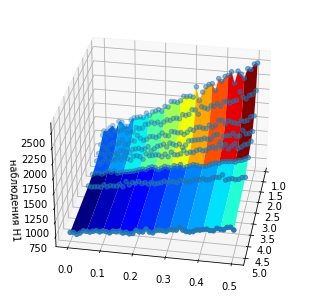
****

Рисунок 5.8 – Зависимость числа наблюдений от значения «ускоряющего коэффициента»»

А что же у нас по ошибкам? Здесь ситуация интереснее: чем больше «ускоряющий коэффициент», тем больше у нас становится ошибочных решений в отсутствии искажений. Однако при увеличении числа «искажённых» случайных величин, эта разница нивелируется и при половине искажённых данных почти не заметна, рисунок 5.9:

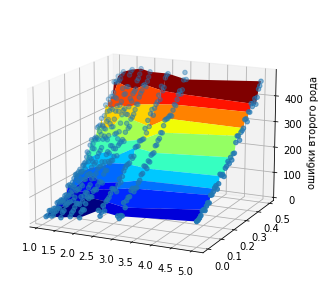
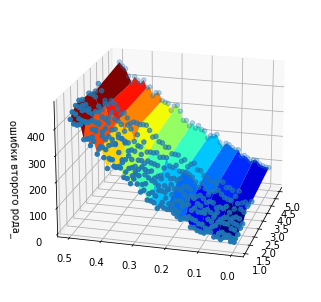


Рисунок 5.9 – Зависимость числа ошибок от значения «ускоряющего коэффициента»

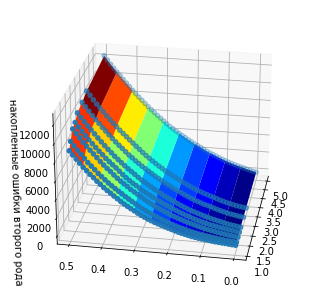
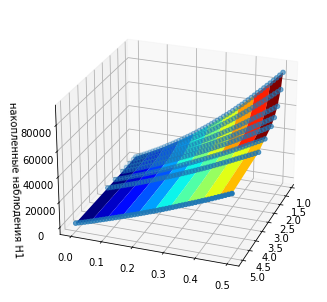


Рисунок 5.10 – Накопленные характеристики в зависимости от значения коэффициента

По представленным на рисунке 5.10 накопленным результатам видно, что при последних 3-х значениях коэффициентах достаточно резко возрастает общее число ошибок второго рода, в то время как количество дополнительных наблюдений снижается не настолько сильно. Значит в нашем случае т.н. «оптимальным» значением коэффициента будет что-то в районе 2.5, возможно даже число .

Поскольку мы убедились, что с помощью этого коэффициента можно снизить количество наблюдений, существенно не повлияв на число ошибок, то можно также поставить задачу о нахождении оптимального значения этого коэффициента, однако сейчас мы это пропустим.

С другой стороны, в отсутствии искажений мы заметили некоторый скачок ошибочных решений, если значение этого коэффициента было достаточно велико, причём это существенно не влияло на ошибки в случае половины случайных величин, распределённых по «засоряющему» закону. Таким образом, вероятно, с помощью данного коэффициента с некоторыми доработками можно узнавать о наличии искажений в обрабатываемых нами данных. Это также может стать темой для дальнейших исследований.

ЗАКЛЮЧЕНИЕ

В ходе проведённых экспериментов было установлено, что теоретические вероятности ошибок в выборе верной гипотезы (уровни значимости) практически всегда отличаются от практических, причём чаще в большую сторону. Также они влияют на число наблюдений (при большей оценке происходит больше наблюдений), требуемых для принятия одной из гипотез. Например, при изменении значений уровней значимости в разные стороны, фактические вероятности ошибки в большинстве случаев изменятся в таком же ключе.

Также было установлено, что уровни значимости дают тем более точную оценку, чем ближе значения гипотез друг к другу, но практически никогда не совпадают с фактическими вероятностями ошибок. Важно отметить, что зачастую точность повышается за счёт значительного увеличения числа наблюдений до принятия одной из гипотез, что может повлиять на время вычислений.

Таким образом, для повышения шанса принять верную гипотезу о параметре некоторого распределения (в нашем случае – Пуассона, а также Полиномиального), гораздо важнее подобрать сами гипотезы, нежели их уровни значимости. Если же высокая точность нам не особо важна, то можно ограничится лишь выбором последних не трогая гипотезы. В таком случае лучше всё-таки стараться задавать уровни значимости как можно меньше – тогда (как было установлено для обоих распределений) и фактическая вероятность ошибки будет меньше.

Стоит отметить и тот факт, что с уменьшением разницы между гипотезами повышается точность оценок возможной ошибки – т.е. уменьшается средняя разница между уровнем значимости и фактической вероятностью ошибки. Но также значительно повышается требуемое число наблюдений, что может существенно повлиять на скорость вычислений.

Что касается искажений во входных данных, то для первого типа, когда искажения распределены по некоторому закону, это вызывает больше ошибок, чем если они идут напрямую в тест. Но такое получается за счёт увеличения числа наблюдений, как видно на примере модели 1.2. Для второго типа искажений было установлено, что разница законов распределения напрямую отражается на разнице в принятии решений. Стоит также отметить и устойчивость теста к искажениям в первом случае – если их не более 20%, то тест работает с достаточно высокой точностью для обеих моделей искажений, серьёзные неточности возникают только при сильных искажениях. А поскольку повышенное число наблюдений дало более высокую точность, возможно стоит искусственно повысить «помехоустойчивость» теста, проводя дополнительные наблюдения после принятия первого решения (например, принимаем решения 3 раза и по ним делаем вывод). Для второго типа искажений ещё значительную роль играет характер помех – так, для искажений, распределённых по геометрическому закону, уже при 10% помех количество ошибочных решений было около 20%.

Также было установлено, что тест Вальда достаточно устойчив к искажениям во входных данных, кроме случаев, когда искажённые данные серьёзно отличаются от реальных (например, принадлежат другому распределению с совершенно другими свойствами) – тогда это сразу отражается на результатах теста. Серьёзные ошибки в принятии решений возникают в достаточно небольшом интервале от количества искажений, сопоставимом с количеством неискажённых данных. Если искажений ещё больше, то, возможно, стоит рассматривать модель с искажениями в качестве основной. Таким образом тест может эффективно принимать решения примерно для половины значений . А в случае серьёзного отличия искажённых данных справедливо предположить, что, проведя ряд наблюдений для разных входных параметров, можно установить характер искажений во входных данных. В таком случае мы сможем с высокой точностью «отфильтровывать» нужные данные, игнорируя искажённые.

На последнем шаге исследования был разработан механизм, позволяющий несколько повысить эффективность последовательного теста при наличии искажений путем ускорения принятия решения с помощью «ускоряющего коэффициента». Также было установлено, что этот коэффициент тем меньше влияет на получаемые результаты, чем больше искажений есть во входных данных, таким образом добавлением этого коэффициента можно примерно определить количество искаженных данных. Определение эффективности отдельных значений и попытку найти оптимальное значение такого коэффициента можно вынести в отдельный объект дальнейших исследований.