

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ  
УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ  
«МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ  
имени М.В.ЛОМОНОСОВА»

ФИЗИЧЕСКИЙ ФАКУЛЬТЕТ

КАФЕДРА ФИЗИКИ ЧАСТИЦ И КОСМОЛОГИИ

ДИПЛОМНАЯ РАБОТА

**«СТАТИСТИЧЕСКИЕ КРИТЕРИИ ДЛЯ ОБНАРУЖЕНИЯ  
ИСТОЧНИКОВ СИГНАЛА В УСЛОВИЯХ НЕОПРЕДЕЛЕННОГО  
ФОНА»**

Выполнил студент

643 группы

Бабёнов Глеб Игоревич

\_\_\_\_\_

подпись студента

Научный руководитель:

к.ф.-м.н. Сатунин Пётр Сергеевич

\_\_\_\_\_

подпись научного руководителя

Допущена к защите

Зав.кафедрой \_\_\_\_\_

подпись зав.кафедрой

Москва

2026

# Оглавление

<b>ВВЕДЕНИЕ</b> . . . . .	<b>2</b>
0.1 Аннотация . . . . .	2
0.2 Постановка задачи . . . . .	3
0.3 Применение теоремы Уилкса . . . . .	4
<b>1. БАЙЕСОВСКОЕ РЕШЕНИЕ</b> . . . . .	<b>7</b>
1.1 Введение в методику . . . . .	7
1.2 Теоретическое описание . . . . .	7
1.3 Применение к задаче . . . . .	10
<b>2. ВЫБОР МЕТОДА</b> . . . . .	<b>13</b>
2.1 Проверка корректности работы методов . . . . .	13
2.2 Сравнение методов . . . . .	16
<b>3. ДОПОЛНИТЕЛЬНЫЕ ПРОВЕРКИ МЕТОДА</b> . . . . .	<b>18</b>
3.1 Проверка метода на робастность . . . . .	18
3.2 Peeking problem vs Байес . . . . .	21
3.3 Дизайн эксперимента . . . . .	23
<b>4. ПРАКТИЧЕСКОЕ ПРИМЕНЕНИЕ</b> . . . . .	<b>25</b>
<b>ВЫВОДЫ</b> . . . . .	<b>27</b>
<b>ЗАКЛЮЧЕНИЕ</b> . . . . .	<b>29</b>
<b>СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ</b> . . . . .	<b>30</b>

## ВВЕДЕНИЕ

### 0.1 Аннотация

В астрофизике, особенно в гамма-астрономии и физике космических лучей, существует так называемая «проблема on/off». Она возникает, когда ученые пытаются обнаружить слабый сигнал от астрофизического источника на фоне шума (фона), параметры которого точно неизвестны. Суть проблемы заключается в том, как корректно и наилучшим образом оценить статистическую значимость сигнала от источника при ограниченном количестве событий и наличии фоновой помехи. Неправильная статистическая обработка может привести к ложным открытиям или, наоборот, к пропуску реальных источников. Данная работа как раз и посвящена поиску наиболее оптимального решения данной задачи.

Для проверки статистических гипотез в физике чаще всего пользуются классическими частотными критериями с расчетом  $p$ -value и последующим его сравнением с заранее выбранным уровнем значимости  $\alpha$ . Этот подход, будучи наиболее привычным и понятным для всех, имеет ряд ограничений, которыми часто пренебрегают.

К примеру, наиболее популярным методом для on/off-задачи является метод, основанный на теореме Уилкса, предложенный в 1983 году учеными Ли и Ма [1]. В своей работе авторы предложили более мощный подход к проверке гипотезы, чем классический  $Z$ -критерий Фишера. Однако и у этого метода есть слабости: теорема Уилкса требует большого (вообще говоря, стремящегося к бесконечности) объема выборки для корректной аппроксимации статистики, что не всегда выполняется в реальном эксперименте.

В текущем исследовании мы попробуем пойти дальше и решить проблему с количеством событий подхода Ли и Ма, свойственную гамма-астрономии.

Мы погрузимся в альтернативный, байесовский подход, лишенный требования большого числа наблюдений.

Небольшая ремарка: классические исследования статистических критериев обычно проводятся путем сравнения распределений  $p$ -value. В данной работе автор позволил себе использовать сравнения зависимостей ошибок I и II рода из-за специфики байесовского подхода и для большей интерпретируемости.

## 0.2 Постановка задачи

В данной работе мы хотим сделать вывод о наличии какого-либо сигнала на неизвестном фоне путем обзора двух областей – с источником и без. Это и есть классическая постановка on/off-задачи.

Пусть время обзора области с источником пропорционально времени обзора области без источника:  $t_{\text{on}} = \epsilon t_{\text{off}}$ , то есть область без источника обзревается в  $\frac{1}{\epsilon}$  раз дольше, чем область с источником. Это также можно интерпретировать как площадь наблюдаемой области по отношению к площади другой.  $N_{\text{on}}, N_{\text{off}}$  – данные измерения – число событий в области с источником и без, соответственно. Для статистического описания наблюдений будем использовать модель с распределением Пуассона:

$$P_{\text{on, off}} = \text{Pois}(N_{\text{on, off}} | \langle N_{\text{on, off}} \rangle) = \frac{\langle N_{\text{on, off}} \rangle^{N_{\text{on, off}}} e^{-\langle N_{\text{on, off}} \rangle}}{N_{\text{on, off}}!},$$

где  $\langle N_{\text{on}} \rangle = s + \epsilon b$ ,  $\langle N_{\text{off}} \rangle = b$  – параметры модели, причем  $s$  – переменная, отвечающая за сигнал, а  $b$  – за фон. Функция правдоподобия в данном случае примет следующий вид:

$$L = \frac{(s + \epsilon b)^{N_{\text{on}}}}{N_{\text{on}}!} e^{-(s + \epsilon b)} \frac{b^{N_{\text{off}}}}{N_{\text{off}}!} e^{-b}. \quad (0.1)$$

Мы проверяем нулевую гипотезу  $H_0: s = 0$  (отсутствие сигнала) против альтернативной  $H_1: s > 0$  (сигнал есть).

Найдем оценку параметра  $\hat{b}_c$  фона, соответствующую условному максимуму функции правдоподобия  $L_c$  при условии, что нулевая гипотеза

верна:  $s = 0$ . Получим следующее выражение:

$$L_c = \frac{(\epsilon \hat{b}_c)^{N_{\text{on}}}}{N_{\text{on}}!} e^{-\epsilon \hat{b}_c} \frac{\hat{b}_c^{N_{\text{off}}}}{N_{\text{off}}!} e^{-\hat{b}_c},$$

при ОМП  $\hat{b}_c = \frac{N_{\text{on}} + N_{\text{off}}}{\epsilon + 1}$ .

Далее построим отношение функций правдоподобия вида:

$$\begin{aligned} \Lambda &= \frac{L(X|H_0)}{L(X|H_1)} \\ &= \frac{L(N_{\text{on}}, N_{\text{off}}|s = 0, b = \hat{b}_c)}{L(N_{\text{on}}, N_{\text{off}}|s = \hat{s}, b = \hat{b})} \\ &= \frac{L(N_{\text{on}}, N_{\text{off}}|s = 0, b = \hat{b}_c)}{L(N_{\text{on}}, N_{\text{off}}|s = N_{\text{on}} - \epsilon N_{\text{off}}, b = N_{\text{off}})}. \end{aligned}$$

Для нашего случая получаем:

$$\Lambda = \left( \frac{\epsilon}{\epsilon + 1} \left( \frac{N_{\text{on}} + N_{\text{off}}}{N_{\text{on}}} \right) \right)^{N_{\text{on}}} \cdot \left( \frac{1}{\epsilon + 1} \left( \frac{N_{\text{on}} + N_{\text{off}}}{N_{\text{off}}} \right) \right)^{N_{\text{off}}},$$

тогда некоторая статистика  $S$  вида  $-2 \ln \Lambda$ , о которой пойдет речь ниже, равна [1]

$$S = 2 \left[ N_{\text{on}} \ln \left( \frac{\epsilon + 1}{\epsilon} \left( \frac{N_{\text{on}}}{N_{\text{on}} + N_{\text{off}}} \right) \right) + N_{\text{off}} \ln \left( (\epsilon + 1) \left( \frac{N_{\text{off}}}{N_{\text{on}} + N_{\text{off}}} \right) \right) \right]. \quad (0.2)$$

### 0.3 Применение теоремы Уилкса

Наиболее популярным методом решения данной задачи является теорема Уилкса [2]. Вот ее формулировка:

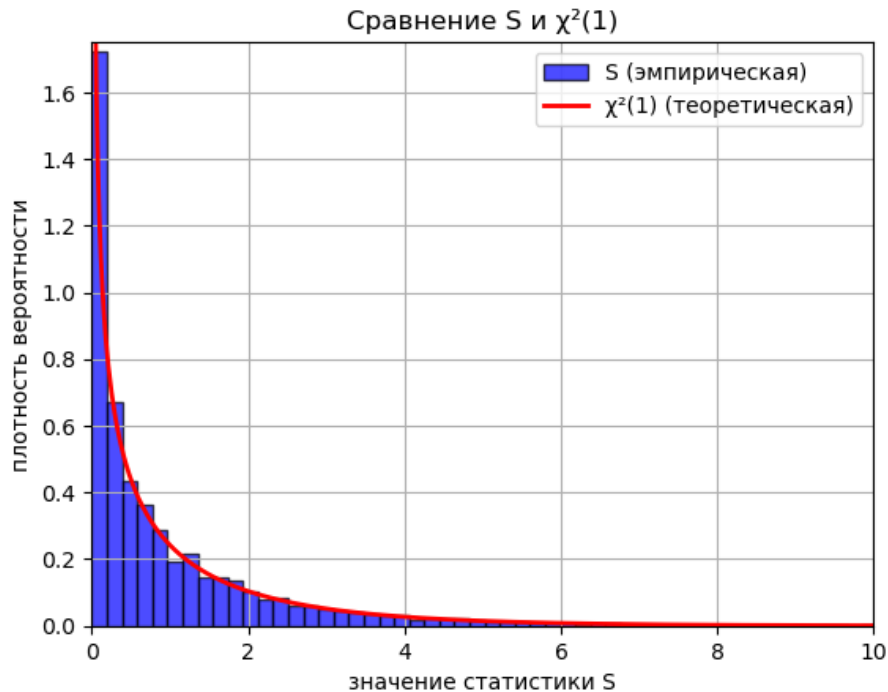
**Пусть** у нас есть набор данных  $X = (x_1, \dots, x_n)$ , параметров  $\Theta = (s, b)$  и нулевая гипотеза  $H_0: s = s_0$ .

**Тогда**

$$S = -2 \ln \Lambda \stackrel{n \rightarrow \infty}{\sim} \chi_1^2,$$

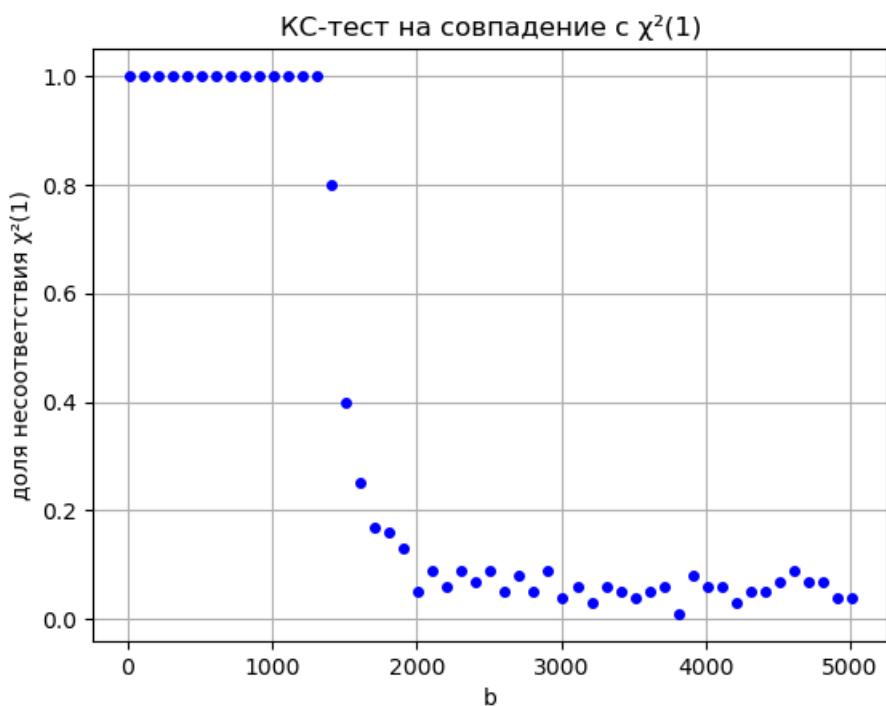
где  $\Lambda = \frac{L(X|s = s_0, b = \hat{b})}{L(X|s = \hat{s}, b = \hat{b})}$  – отношение функции правдоподобия при нулевой гипотезе к функции правдоподобия при альтернативной, а  $\chi_1^2$  – распределение хи-вадрат с одной степенью свободы.

Как мы видим, теорема требует большого количества наблюдений, чего не всегда удается достичь в реальных экспериментах, поэтому проверим данное утверждение методом Монте-Карло с  $s = 0$  (то есть при условии, что нулевая гипотеза верна):



Здесь мы брали значение параметра распределения Пуассона для фона равным 50, и по графику кажется, что теорема Уилкса выполняется. Однако давайте проведем более строгую проверку при помощи критерия Колмогорова-Смирнова, позволяющего статистически сравнивать распределения.

Построим зависимость доли несоответствия теоретическому распределению хи-квадрат с одной степенью свободы нашего эмпирического распределения статистики.



Теперь видно, что «бесконечность» из теоремы Уилкса при заданном параметре  $\epsilon = 0.2$  начинается примерно с 2000 событий, и, вообще говоря, утверждение неверно при меньшем фоне. Однако важно отметить, что распределение хи-квадрат с одной степенью свободы имеет специфическую структуру с очень высокой плотностью около нуля и большим хвостом справа. В таком случае КС-тест может быть не самым мощным критерием, поскольку он наиболее чувствителен к различиям в центральной части распределений – здесь можно было бы применить, к примеру, критерий Андерсона-Дарлинга. Однако для упрощения анализа мы пойдем другим путем: рассчитаем классическое p-value (вероятность получить статистику теста такую же или более экстремальную, чем наблюдаемое значение, при условии, что нулевая гипотеза верна), применим известный критерий проверки гипотез – сравнение p-value с заранее зафиксированным уровнем значимости  $\alpha$  – проверим, как он отрабатывает на реальных данных и сравним с байесовским подходом, который не требует большого количества наблюдений.

# 1. БАЙЕСОВСКОЕ РЕШЕНИЕ

## 1.1 Введение в методику

В байесовском подходе степень нашей разумной уверенности в справедливости некоторого утверждения можно численно выразить в виде вероятности. Помимо этого, особенность метода заключается в том, что для принятия решения используется сразу два типа информации: априорная, выражающая наши знания о модели на момент начала исследования, и сами данные, полученные в ходе эксперимента.

Априорная информация представлена в виде некоторого априорного распределения вероятностей анализируемого неизвестного параметра, которое описывает степень уверенности в том, что он примет определенное значение еще до начала сбора экспериментальных данных. По мере же поступления данных это распределение уточняется, трансформируясь из априорного распределения в апостериорное. Для этой трансформации используется формула [4]:

**Предположим**, что  $A_1, A_2, \dots, A_N$  образуют полную систему событий, а событие  $B$  имеет ненулевую вероятность:  $P(B) > 0$ .

**Тогда**

$$P(A_k|B) = \frac{P(B|A_k)P(A_k)}{\sum_i P(B|A_i)P(A_i)}$$

определяет правило вычисления условной вероятности события  $A_k$  – при условии, что событие  $B$  уже имело место – по безусловной вероятности события  $A_k$  и условным вероятностям  $P(B|A_k)$ .

## 1.2 Теоретическое описание

В нашем случае событие  $A$  заключается в том, что набор параметров распределения  $\Theta$  принимает определенные значения, а событие  $B$  отвечает за

конкретную реализацию наблюдаемых величин  $X = (x_1, \dots, x_n)$ . То есть мы при помощи формулы Байеса, по сути, переходим от распределений случайных величин к распределениям параметров, что позволяет нам исследовать их напрямую: получать точечные или интервальные оценки или делать выводы о справедливости заранее выдвинутой гипотезы о значении параметров.

Поскольку в нашем случае мы работаем с распределением Пуассона, то параметр  $\Theta = \lambda$  может изменяться непрерывно от 0 до  $\infty$ , так что перепишем формулу Байеса в следующем виде:

$$p(\lambda|X) = \frac{p(X|\lambda)\pi(\lambda)}{\int_0^{\infty} p(X|\lambda)\pi(\lambda)d\lambda}, \quad (1.1)$$

где  $\pi(\lambda)$  – это плотность априорного распределения параметра  $\lambda$  или просто приор, а  $p(\lambda|X)$  – плотность апостериорного распределения.

Приор позволяет учитывать заранее известные нам факты: к примеру, ограничения параметров, которые уже были получены ранее другими методами. А исходные статистические данные  $X$  порождаются законом распределения  $p(X|\lambda)$ .

В случае, если мы не располагаем никакой априорной информацией о параметре  $\lambda$ , следуя Джеффрису [3], можно придерживаться следующих правил:

- если оцениваемый параметр  $\theta$  теоретически может принимать значения на конечном от  $\theta_{\min}$  до  $\theta_{\max}$  или бесконечном от  $-\infty$  до  $+\infty$  интервалах, то мы полагаем, что он распределен равномерно на этих интервалах – то есть  $\pi(\theta) = \text{const}$ ;
- если же  $\theta$  принимает только положительные значения  $\theta > 0$ , то мы полагаем на всей числовой прямой  $(-\infty, +\infty)$ :  $\pi(\ln \theta) = \text{const}$  – таким образом, мы обеспечиваем инвариантность относительно масштабов.

Поскольку в нашем случае  $\lambda > 0$  и мы не обладаем никакой априорной информацией о нем, давайте получим конкретный вид для  $\pi(\lambda)$ .

$$F_\lambda(x) = P(\lambda < x) = P(\ln \lambda < \ln x) = F_{\ln \lambda}(\ln x)$$

– функция распределения параметра  $\lambda$ . Таким образом, получаем

$$p(x) = \frac{\partial F_\lambda(x)}{\partial x} = \frac{\partial F_{\ln \lambda}(\ln x)}{\partial \ln x} \cdot \frac{\partial \ln x}{\partial x} = p_{\ln \lambda}(\ln x) \cdot \frac{1}{x},$$

а по условию  $p_{\ln \lambda}(\ln x) = \pi(\ln \lambda) = \text{const}$ , то есть:

$$\pi(\lambda) \sim \frac{1}{\lambda}. \quad (1.2)$$

Далее нам необходимо ввести одно понятие:

Семейство априорных распределений  $G = \{\pi(\lambda; D)\}$  является сопряженным по отношению к наблюдаемой генеральной совокупности  $p(X|\lambda)$ , если и апостериорное распределение  $p(\lambda|X)$ , вычисленное по формуле (1.1), снова принадлежит к этому же семейству  $G$ .

Здесь в качестве функции  $p(X|\lambda)$  может выступать функция правдоподобия  $L(X|\lambda)$ .

Также запишем условие существования сопряженного семейства априорных распределений [4]:

Если функция правдоподобия  $L(X|\lambda)$  представима в виде:

$$L(X|\lambda) = \nu(T_1(X), \dots, T_m(X); \lambda) \cdot \psi(X), \quad (1.3)$$

где  $X = (x_1, \dots, x_n)$  – набор наблюдений, а  $T(X)$  и  $\psi(X)$  – некоторые их функции, не зависящие от  $\lambda$ , то существует семейство  $G = \{\pi(\lambda; D)\}$  априорных распределений, сопряженных с  $L(X|\lambda)$ . Причем  $T(X) = (T_1(X), \dots, T_m(X))$  –  $m$ -мерная достаточная статистика.

Давайте проверим это условие для распределения Пуассона, фигурирующего в нашей задаче:

$$P(x|\lambda) = \text{Pois}(x|\lambda) = \frac{\lambda^x}{x!} e^{-\lambda},$$

отсюда функция правдоподобия равна

$$L(x_1, \dots, x_n | \lambda) = \prod_{i=1}^n \frac{\lambda^{x_i}}{x_i!} e^{-\lambda} = e^{-n\lambda} \cdot \lambda^{\sum_{i=1}^n x_i} \cdot \prod_{i=1}^n \frac{1}{x_i!}. \quad (1.4)$$

Сравнивая это с (1.3), имеем:  $m = 1$ ,  $T(x_1, \dots, x_n) = \sum_{i=1}^n x_i$  – достаточная статистика, что подтверждает существование сопряженного с  $L$  распределения параметра  $\lambda$ .

Наконец, давайте получим апостериорное распределение для нашей функции правдоподобия. Как было показано выше (1.4), функция правдоподобия  $L(x_1, \dots, x_n | \lambda) \sim e^{-n\lambda} \cdot \lambda^{\sum_{i=1}^n x_i}$ , таким образом, получаем:

$$p(\lambda | X) \sim \pi(\lambda) L(X | \lambda) \sim e^{-n\lambda} \cdot \lambda^{\sum_{i=1}^n x_i - 1}, \quad (1.5)$$

поскольку  $\pi(\lambda) \sim \frac{1}{\lambda}$  (1.2). Правая часть здесь с точностью до нормирующего множителя представляет собой плотность гамма-распределения [4]

$$p(\lambda) = \frac{\beta^\alpha}{\Gamma(\alpha)} e^{-\beta\lambda} \lambda^{\alpha-1}, \quad \lambda > 0.$$

### 1.3 Применение к задаче

Теперь мы можем приступить к решению поставленной задачи описанным выше методом. Пусть у нас есть два значения  $\lambda_{\text{on}}$  и  $\lambda_{\text{off}}$ , которые мы хотим сравнить для проверки гипотезы о том, что в области  $S_{\text{on}}$  присутствует сигнал. Как мы показали в предыдущем параграфе (1.5), наши параметры генерируются из гамма-распределений:

$$\lambda_{\text{on, off}} \sim \text{Gamma}(\alpha_{\text{on, off}}, \beta_{\text{on, off}}),$$

где  $\alpha_{\text{on, off}} = N_{\text{on, off}}$ ,  $\beta_{\text{on}} = 1$ ,  $\beta_{\text{off}} = \frac{1}{\epsilon}$ , поскольку область  $S_{\text{off}}$  наблюдается в  $\frac{1}{\epsilon}$  раз чаще, чем  $S_{\text{on}}$ .

Посчитаем байесовскую вероятность  $P(H_1|X)$ , то есть вероятность того, что  $\lambda_{\text{on}}$  больше  $\lambda_{\text{off}}$ :

$$P(\lambda_{\text{on}} > \lambda_{\text{off}}) = \int_0^{\infty} d\lambda_{\text{off}} \int_{\lambda_{\text{off}}}^{\infty} d\lambda_{\text{on}} \frac{\beta_{\text{on}}^{\alpha_{\text{on}}}}{\Gamma(\alpha_{\text{on}})} e^{-\beta_{\text{on}}\lambda_{\text{on}}} \lambda_{\text{on}}^{\alpha_{\text{on}}-1} \frac{\beta_{\text{off}}^{\alpha_{\text{off}}}}{\Gamma(\alpha_{\text{off}})} e^{-\beta_{\text{off}}\lambda_{\text{off}}} \lambda_{\text{off}}^{\alpha_{\text{off}}-1},$$

используя определение верхней неполной гамма-функции, перейдем к одномерному интегралу:

$$P(\lambda_{\text{on}} > \lambda_{\text{off}}) = \int_0^{\infty} \Gamma(\alpha_{\text{on}}, \beta_{\text{on}}\lambda_{\text{on}}) \frac{\beta_{\text{off}}^{\alpha_{\text{off}}} e^{-\beta_{\text{off}}\lambda_{\text{off}}} \lambda_{\text{off}}^{\alpha_{\text{off}}-1}}{\Gamma(\alpha_{\text{off}})} d\lambda_{\text{off}}.$$

Используя тождество  $\Gamma(1, x) = e^{-x}$  и рекурсивное соотношение  $\Gamma(a + n, x) = \Gamma(a, x) + x^a e^{-x} \sum_{k=0}^{n-1} \frac{x^k}{\Gamma(a + k + 1)}$ , получим для  $\Gamma(n, x)$ :

$$\Gamma(n, x) = e^{-x} \sum_{k=0}^{n-1} \frac{x^k}{\Gamma(k + 1)}.$$

После этого можем переписать нашу вероятность следующим образом:

$$P(\lambda_{\text{on}} > \lambda_{\text{off}}) = \int_0^{\infty} d\lambda_{\text{off}} e^{-\beta_{\text{on}}\lambda_{\text{off}}} \left( \sum_{k=0}^{\alpha_{\text{on}}-1} \frac{(\beta_{\text{on}}\lambda_{\text{off}})^k}{\Gamma(k + 1)} \right) \frac{\beta_{\text{off}}^{\alpha_{\text{off}}} \lambda_{\text{off}}^{\alpha_{\text{off}}-1} e^{-\beta_{\text{off}}\lambda_{\text{off}}}}{\Gamma(\alpha_{\text{off}})},$$

откуда, после смены порядка суммирования и интегрирования и последующего интегрирования, можно получить следующее выражение:

$$P(\lambda_{\text{on}} > \lambda_{\text{off}}) = \sum_{k=0}^{\alpha_{\text{on}}-1} \beta_{\text{on}}^k \beta_{\text{off}}^{\alpha_{\text{off}}} (\beta_{\text{on}} + \beta_{\text{off}})^{-(k+\alpha_{\text{off}})} \frac{\Gamma(k + \alpha_{\text{off}})}{\Gamma(k + 1)\Gamma(\alpha_{\text{off}})} \cdot \frac{k + \alpha_{\text{off}}}{k + \alpha_{\text{off}}}.$$

Далее, используя свойство гамма-функции  $\Gamma(z + 1) = z\Gamma(z)$  и вводя бета-функцию  $B(z_1, z_2) = \frac{\Gamma(z_1)\Gamma(z_2)}{\Gamma(z_1 + z_2)}$ , получим итоговый ответ:

$$P(\lambda_{\text{on}} > \lambda_{\text{off}}) = \sum_{k=0}^{\alpha_{\text{on}}-1} \frac{(\beta_{\text{on}} + \beta_{\text{off}})^{-(k+\alpha_{\text{off}})} \beta_{\text{on}}^k \beta_{\text{off}}^{\alpha_{\text{off}}}}{(k + \alpha_{\text{off}})B(k + 1, \alpha_{\text{off}})},$$

перепишывая с учетом наших обозначений  $\alpha_{\text{on, off}} = N_{\text{on, off}}$ ,  $\beta_{\text{on}} = 1$ ,  $\beta_{\text{off}} = \frac{1}{\epsilon}$ , получим:

$$P(\lambda_{\text{on}} > \lambda_{\text{off}}) = \sum_{k=0}^{N_{\text{on}}-1} \frac{(1 + \frac{1}{\epsilon})^{-(k+N_{\text{off}})}}{\epsilon^{N_{\text{off}}}(k + N_{\text{off}})B(k + 1, N_{\text{off}})}. \quad (1.6)$$

Это вероятность того, что параметр распределения Пуассона в области  $S_{\text{on}}$  больше, чем в области  $S_{\text{off}}$ , то есть что сигнал действительно присутствует на фоне заранее неизвестного фона. Как говорилось ранее, байесовский подход позволяет получить более «осязаемые» понятия, нежели частотный. Помимо этого, важно отметить, что он аналитически точный и не требует асимптотик.

Однако стоит обратить внимание на случай  $N_{\text{on}} = 0$ , когда в области он вообще не было ни одного события и наша формула не работает. Такие случаи, очевидно, нам не интересны и мы можем сказать, что сигнал отсутствует, даже не прибегая к использованию критериев.

Введем критерий, на основании которого мы будем принимать решение. Будем говорить, что сигнал есть (то есть отвергать нулевую гипотезу  $H_0: s = 0$ ) на уровне значимости  $\alpha = 5\%$ , если:

$$P(\lambda_{\text{on}} > \lambda_{\text{off}}) > 1 - 0.05, \quad (1.7)$$

то есть когда вероятность того, что сигнал есть, больше 95%.

Полученный критерий в общем случае формулируется следующим образом:

Мы отвергаем  $H_0$  на уровне значимости  $\alpha$ , если  $P(H_1|X) > 1 - \alpha$ .

## 2. ВЫБОР МЕТОДА

### 2.1 Проверка корректности работы методов

Для сравнения качества частотного и байесовского подходов к проверке гипотез о наличии сигнала на неизвестном фоне будем генерировать  $N_{\text{on}}$  и  $N_{\text{off}}$  из следующих распределений Пуассона:

$$N_{\text{on}} \sim \text{Pois}(\epsilon b), \quad N_{\text{off}} \sim \text{Pois}(b),$$

где  $b$  – параметр фона, а  $s = 0$ , то есть никакого сигнала на самом деле нет.

Далее воспользуемся нашими критериями с использованием теоремы Уилкса и последующим расчетом p-value (0.1) и с использованием байесовской вероятности (1.7). Сравним для них ошибку I рода (частоту появления ложноположительных выводов) для различных уровней значимости  $\alpha$  и различных параметров  $b$ .

Вспомним теорему из математической статистики:

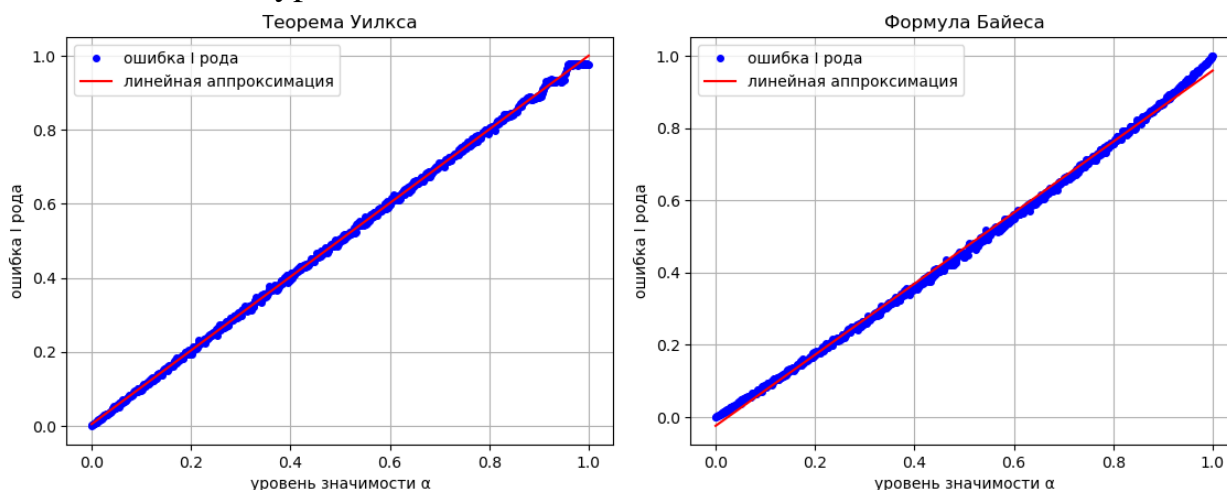
**Пусть**  $T$  – тестовая статистика и  $\text{p-value} = P(T \geq T_{\text{obs}} | H_0)$  – соответствующее p-value.

**Тогда** при условии, что  $H_0$  верна:

$$\text{p-value} \sim \text{Uniform}(0, 1).$$

Для проверки этого утверждения в нашем случае достаточно построить зависимость ошибки I рода от уровня значимости  $\alpha$ . Поскольку  $H_0$  точно верна в силу нашей генерации, эта зависимость должна быть линейная с коэффициентом 1.

Выберем для общности  $b = 50$  и сгенерируем по 10 000 вариантов  $N_{\text{он}}$  и  $N_{\text{офф}}$  для каждого уровня значимости  $\alpha$  от 0 до 1 с шагом 0.001.



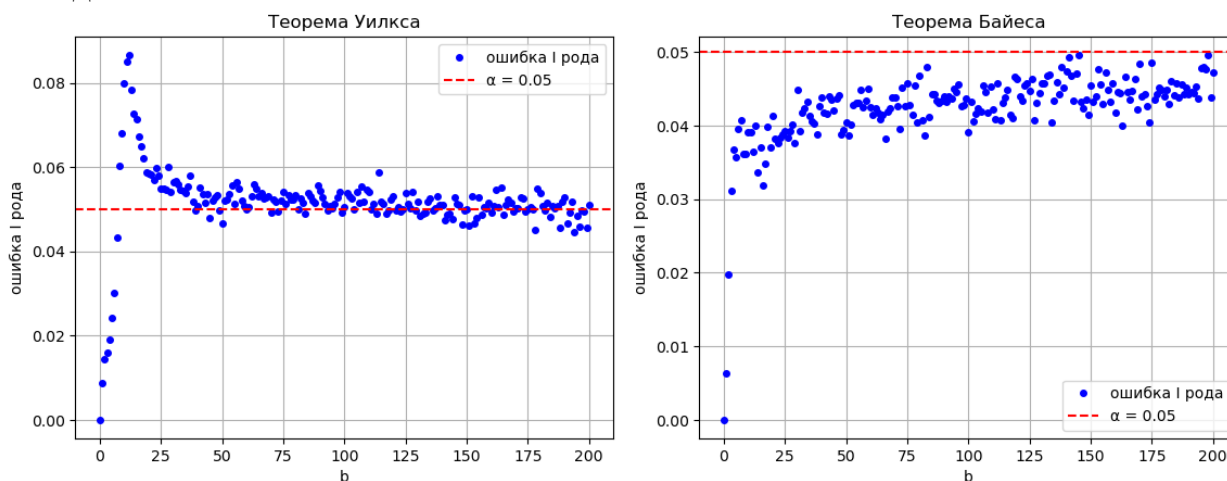
Видно, что зависимости в обоих случаях, действительно, линейные, хотя в случае теоремы Байеса она слегка загибается. Здесь также была проведена линейная аппроксимация  $y = a_1x + a_0$  при помощи МНК:

- для первого случая  $a_1 = 0.9960, a_0 = 0.0046, R^2 = 0.9996$ ;
- для второго случая  $a_1 = 0.9833, a_0 = -0.0245, R^2 = 0.9979$ .

Эти числа так же демонстрируют небольшое преимущество на стороне теоремы Уилкса.

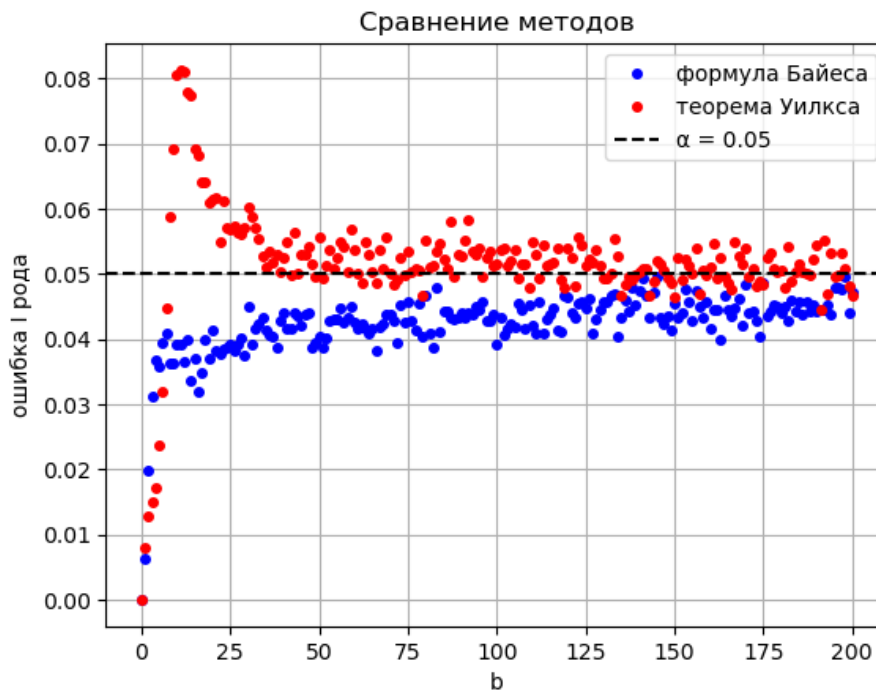
Теперь построим зависимости ошибки I рода от параметра  $b$ . Очевидно, что в идеальной картине ошибка не должна зависеть от  $b$ .

Для примера зафиксируем уровень значимости  $\alpha = 0.05$ , а  $b$  будем менять от 1 до 200 с шагом 1.



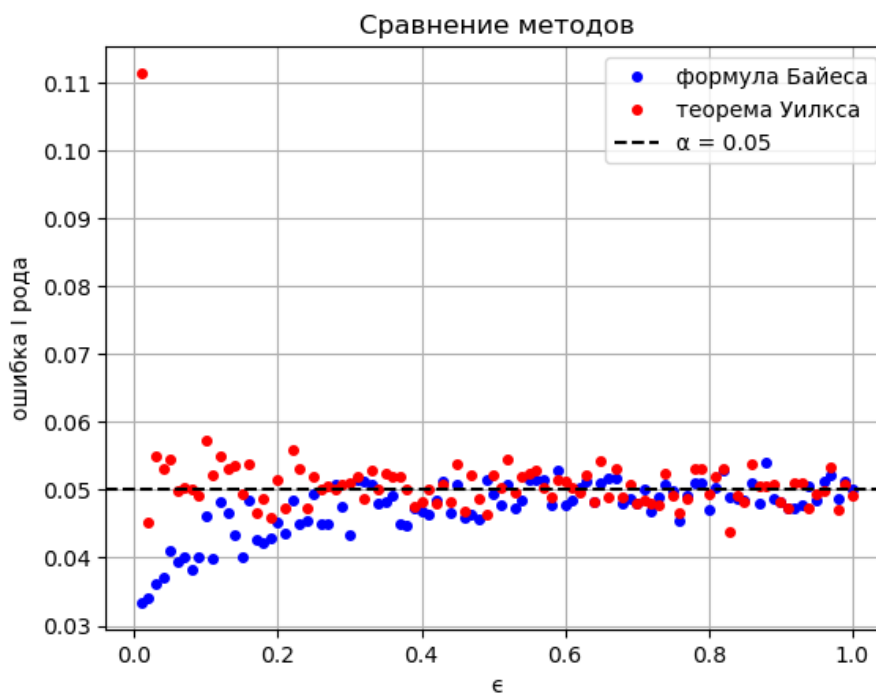
Видно, что оба распределения для малых значений параметра фона ведут себя не совсем адекватно, а затем выходят на некоторое плато около 0.05. Причем в случае использования байесовского подхода ошибка I рода лежит преимущественно ниже уровня значимости, в то время как в случае теоремы Уилкса приближается сверху, то есть в частотном подходе число ложноположительных результатов будет завышено.

Для наглядности сравнения также построим последние два распределения на одном графике:



На самом деле, заниженная вероятность ошибки I рода для байесовского подхода в области малых значений фона  $b$  связана с методом Монте-Карло. Он генерирует значения числа событий для областей on и off из распределения Пуассона с соответствующим заданным параметром. При малых параметрах довольно часто возникают случаи, когда  $N_{\text{on}} = 0$ , что, как мы писали выше, «ломает» нашу формулу для вероятности, а численная реализация возвращает 0. В случае же ненулевых значений  $N_{\text{on}}$  методу можно доверять, поскольку он использует аналитическую формулу без асимптотик.

Также для полноты картины проверим, насколько наши методы чувствительны к изменениям параметра  $\epsilon$  при фиксированном значении фона  $b = 200$ :



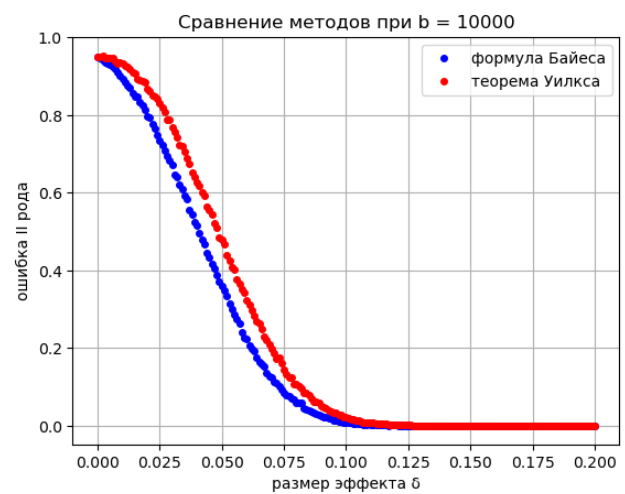
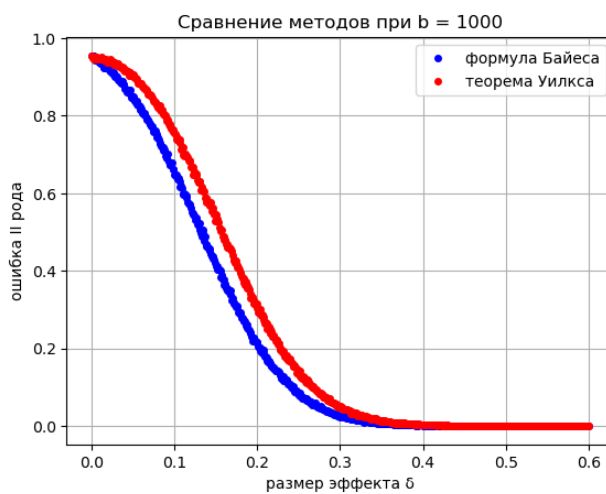
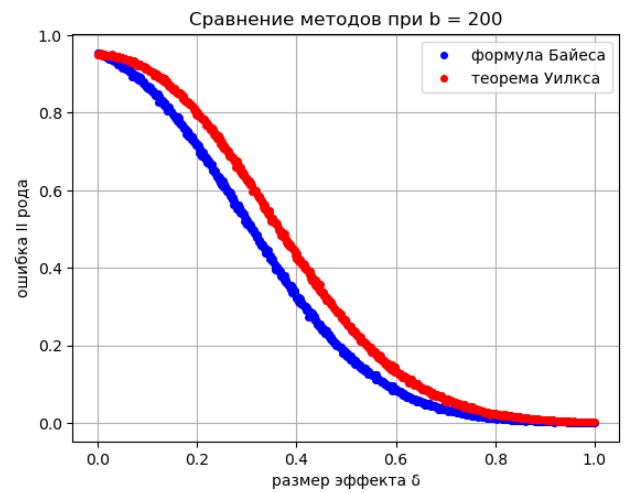
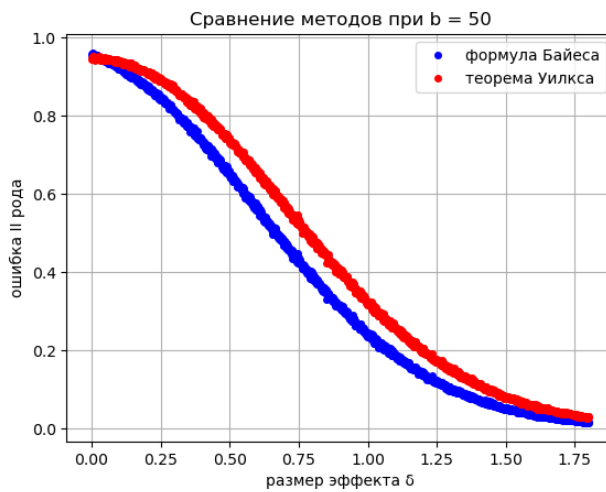
Критерии, в целом, неплохо контролируют ошибку I рода, не считая области совсем малых значений  $\epsilon$ .

Таким образом, в целом, при достаточно больших параметрах фона  $b$  оба метода работают корректно, а байес имеет преимущество еще и для редких событий. Давайте теперь попробуем сравнить эти методы для остальной области значений параметров, где оба метода ведут себя адекватно.

## 2.2 Сравнение методов

Ошибка II рода – это вероятность отклонить альтернативную гипотезу, когда она на самом деле верна. Сравнить методы мы будем при помощи сравнения их ошибок II рода: постепенно добавляя эффект в  $N_{\text{он}}$ , посмотрим, как от этого изменяется ошибка. Проверим это для нескольких вариантов параметра фона  $b$ , так как поведение метода может отличаться. Увеличивать значение параметра  $\lambda_{\text{он}}$  будем следующим образом:  $\lambda_{\text{он}} = \epsilon b \cdot (1 + \delta)$ , то есть  $s = \epsilon \delta b$ , где  $\delta$  изменяется от 0 до 1 с шагом 0.001.

Давайте промоделируем ситуации, когда параметр фона  $b$  лежит в диапазоне от 50 до 10 000, где оба наших метода отработали корректно.



Итак, мы можем наблюдать, что метод, основанный на формуле Байеса, в некоторой области величины сигнала более чувствителен – то есть он чаще «замечает» сигнал, когда он действительно присутствует. При дальнейшем увеличении сигнала оба метода обрабатывают одинаково. Таким образом, предпочтительнее использовать байесовский подход как более мощный метод поиска сигнала в условиях неопределенного фона.

### 3. ДОПОЛНИТЕЛЬНЫЕ ПРОВЕРКИ МЕТОДА

В предыдущей главе мы определили, что байесовский подход превосходит частотный. Теперь давайте проведем ряд проверок, которые позволят нам более уверенно применять нашу разработку в реальных исследованиях.

#### 3.1 Проверка метода на робастность

Робастность – свойство статистического метода, характеризующее независимость влияния на результат исследования различного рода выбросов, устойчивости к помехам.

Мы живем в реальном мире, где идеальная модель может не соответствовать действительности. В связи с этим хочется проверить, насколько наш метод устойчив к несовпадению предпосылок модели с реальностью. Самым распространенным отклонением в статистических методах является отклонение от теоретического распределения.

Как известно, для распределения Пуассона  $\mathbb{E}[X] = \text{Var}[X] = \lambda$ . Давайте добавим «овердисперсию» к нашей случайной величине:  $\text{Var}[X] = \mathbb{E}[X] + \kappa\lambda^2$ , где  $\mathbb{E}[X] = \lambda$ , а при  $\kappa = 0$  мы возвращаемся к исходному распределению Пуассона. Сделать это можно путем соответствующей репараметризации отрицательного биномиального распределения  $X \sim \text{NB}(r, p)$ :

$$\begin{cases} r = \frac{1}{\kappa} \\ p = \frac{1}{1 + \kappa\lambda} \end{cases},$$

для которого исходно  $\mathbb{E}[X] = \frac{r(1-p)}{p}$  и  $\text{Var}[X] = \frac{r(1-p)}{p^2}$ .

Этот переход и граничный случай с  $\kappa = 0$  обосновываются математически, однако «на пальцах»  $r$  – это число «успехов», которое мы ждем в испытании бернулли до остановки эксперимента, а  $p$  – вероятность

этого успеха. Случайная величина  $X$  же показывает, сколько за это время будет «неудач». Таким образом, при условиях  $r \rightarrow \infty$  и  $p \rightarrow 1$  (то есть  $q = 1 - p \rightarrow 0$ ), соответствующих переходу  $\kappa \rightarrow 0$ , мы и получаем распределение Пуассона для числа «неудач».

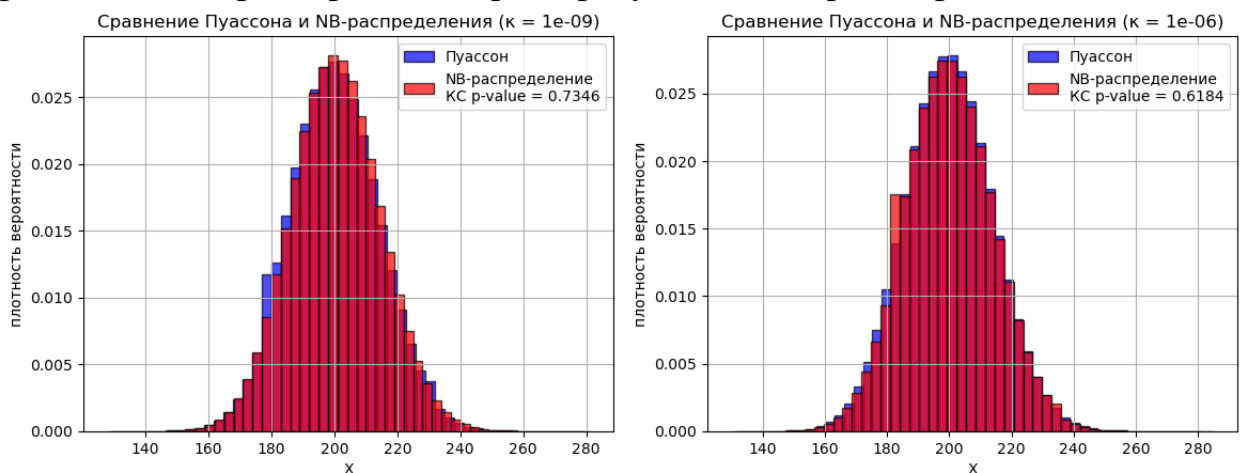
С физической точки зрения есть множество факторов, свойственных астрономическим наблюдениям, которые будут искажать наше распределение: атмосферные условия (прозрачность, давление) меняются, меняется эффективность регистрации, а интенсивность фона от космических лучей может варьироваться в зависимости от времени суток или солнечной активности. И все это можно учесть за счет перехода к тому, что сам параметр  $\lambda$  является некоторой случайной величиной.

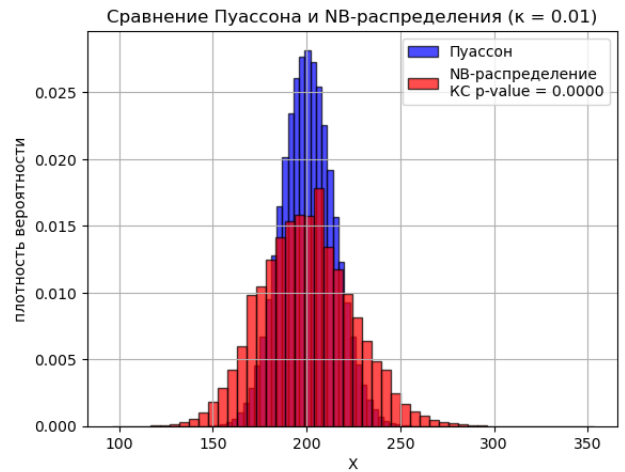
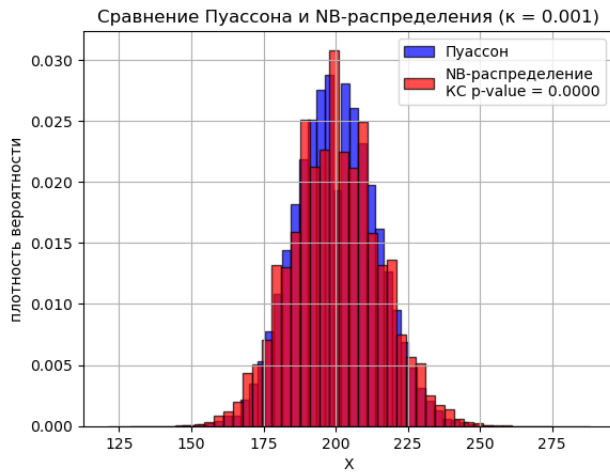
Отрицательное биномиальное распределение как раз и получается из распределения Пуассона:

$$P_{\text{NB}}(X) = \int_0^{\infty} P_{\text{Poisson}}(X|\lambda) \cdot \rho_{\text{Gamma}}(\lambda) d\lambda,$$

где параметр  $\lambda$  имеет гамма-распределение, достаточно популярное в астрономии.

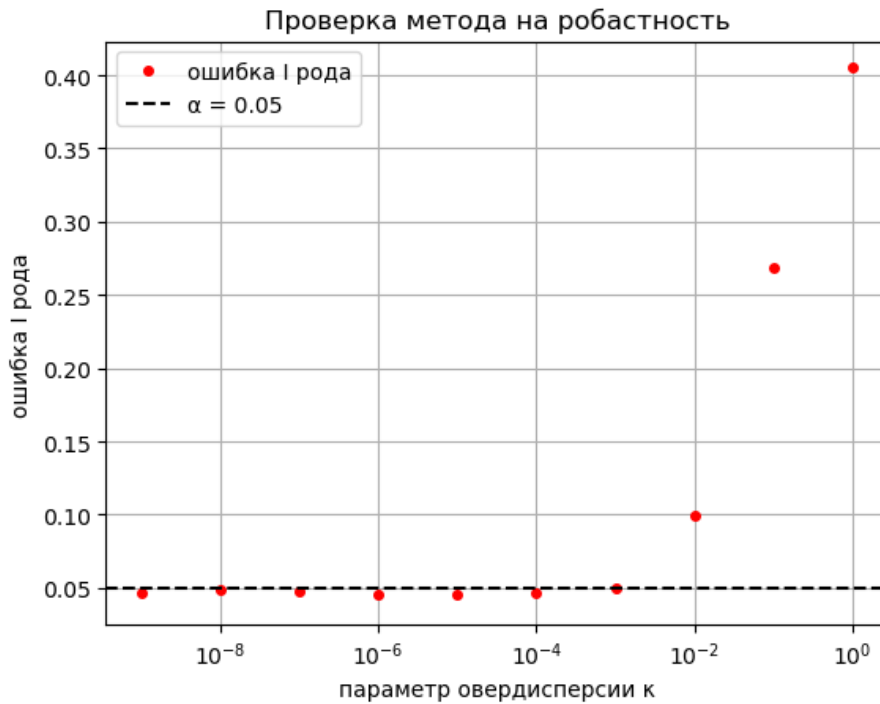
Давайте построим NB-распределения с  $\lambda = 200$  (поскольку при таком значении Байес уже достаточно уверенно контролировал ошибку I рода), но с различными параметрами  $\kappa$ , характеризующими овердисперсию.





Можем наблюдать, что, начиная с  $\kappa \sim 10^{-3}$ , двухвыборочный критерий Колмогорова-Смирнова начинает различать распределения, а с  $\kappa \sim 10^{-2}$  разница уже видна невооруженным глазом – хвосты становятся тяжелее.

Теперь построим график, демонстрирующий робастность нашего метода:



Видно, что даже при  $\kappa \sim 10^{-3}$  метод все еще остается достаточно надежным. То есть при фоне, равном примерно 200 событий, метод можно использовать даже когда мы не учли дополнительные 20% дисперсии!

### 3.2 Peeking problem vs Байес

В некоторых работах (к примеру, [5]) исследователи полагают, что байесовский подход способен без дополнительных корректировок бороться с так называемой *peeking problem* или проблемой подглядывания, которая заключается в том, что каждый раз, когда мы смотрим на данные, планируя принять решение об остановке эксперимента при достижении статистической значимости, мы таким образом увеличиваем ошибку I рода. Здесь мы при помощи моделирования экспериментов разберемся, так ли это, и поймем, какие есть ограничения, поскольку в науке часто останавливают исследования после совершения открытия.

Моделирование заключается в следующем. Положим, для примера, что мы наблюдаем области on и off в течение месяца. В области on заведомо нет источника, то есть  $s = 0$ , фон постоянен во времени и имеет фиксированное значение параметра  $b = \text{const}$  в рамках суток, а его значения за разные дни независимы друг от друга. При помощи Монте-Карло моделирования оценим вероятность ошибки I рода в двух сценариях: когда мы «подглядываем» каждый день и останавливаем эксперимент в момент достижения статистической значимости; и когда мы ждем все 30 суток и только потом применяем критерий к нашим набранным данным.

Наш эксперимент демонстрирует, что байесовский метод так же не способен контролировать ошибку I рода при подглядывании, как и частотный подход: вероятность ошибки I рода без подглядывания составила 4.8% на уровне значимости 5%, что является вполне корректным результатом. Однако, когда мы начали останавливать эксперимент в момент достижения статистической значимости, ошибка выросла до 22.32%. Для сравнения, теорема Уилкса показала похожие результаты: 4.95% и 28.33%.

Давайте взглянем на иллюстрацию данной проблемы в байесовском случае:



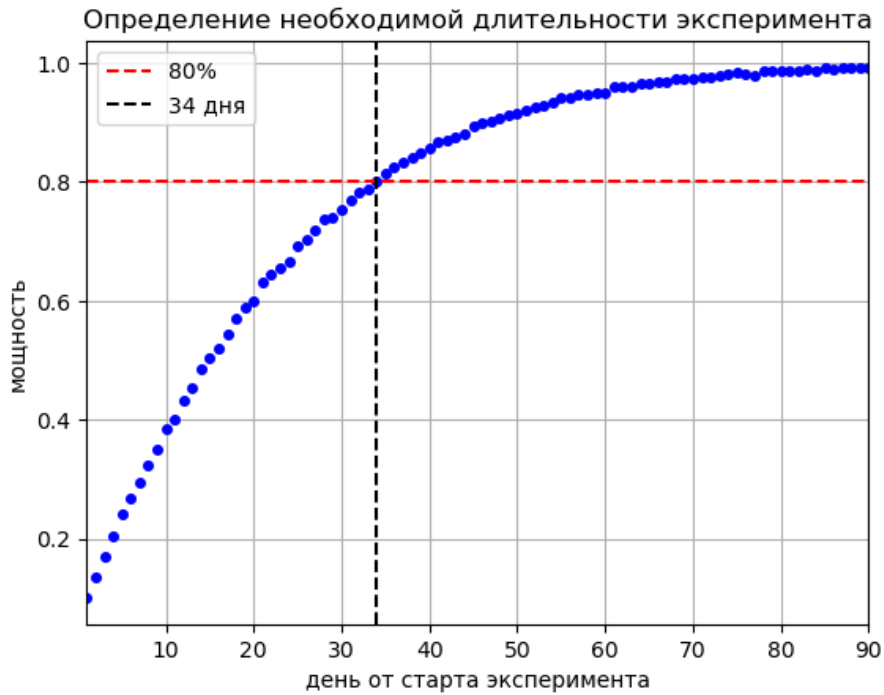
Если бы мы решили остановить сбор данных в районе двадцатого дня с начала эксперимента, то мы совершили бы ложное открытие.

### 3.3 Дизайн эксперимента

Тут может возникнуть вопрос: как нам планировать длительность эксперимента? В самых популярных частотных подходах, таких как Z-критерий Фишера, есть готовая аналитическая формула, которая при заданных мощности и уровне значимости позволяет рассчитать необходимое количество наблюдений. В байесовском подходе такой формулы нет, однако тут нам в который раз приходит на помощь Монте-Карло.

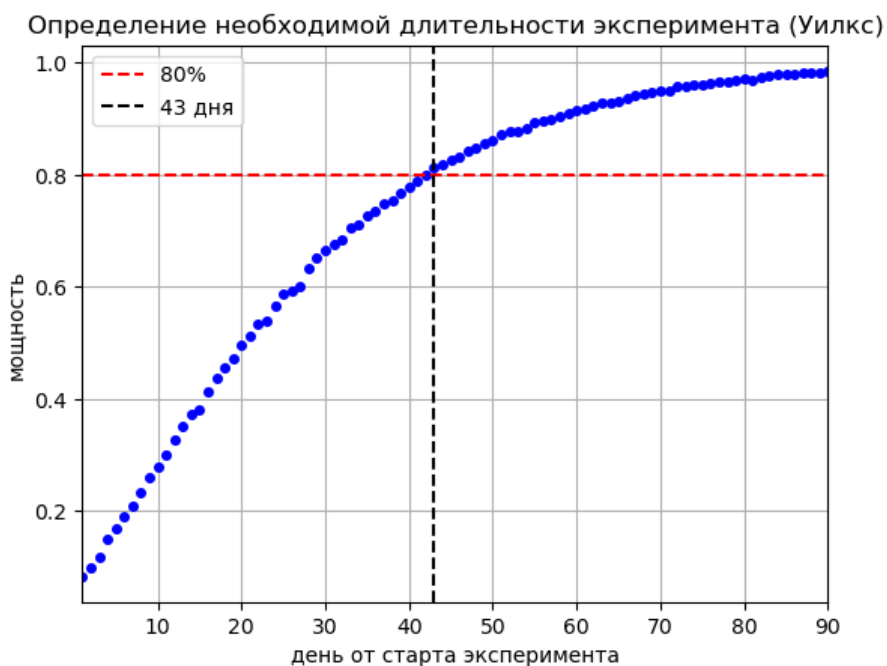
Сформулируем пошаговую инструкцию для определения длительности наблюдения:

1. выбираем уровень значимости, к примеру,  $\alpha = 5\%$ ;
2. определяемся с мощностью, которая нам необходима, то есть с вероятностью «поймать» эффект, если сигнал действительно есть: к примеру,  $\beta = 80\%$ ;
3. по данным наблюдений оцениваем величину фона на основании количества фоновых событий, приходящихся на день,  $N_{\text{off}} = 20$ ;
4. определяемся с тем, какой минимальный сигнал нам вообще имеет смысл детектировать – это может, к примеру, зависеть от чувствительности наших детекторов: положим, все, что меньше в среднем  $s = 1$  за день, нас не интересует;
5. строим график зависимости мощности (или ошибки II рода) от количества дней наблюдений;
6. по графику определяем необходимую длительность:



Таким образом, мы видим, что при озвученных вводных нам нужно набирать статистику в течение 34 дней. На самом деле, лучше заложить время с запасом на случай, если мы недостаточно хорошо оценили  $N_{\text{off}}$  или использовали слишком большую оценку  $s$ .

Для сравнения, если бы мы использовали критерий, основанный на теореме Уилкса, нам бы потребовалось на 9 дней больше для достижения такой же мощности при аналогичных вводных:



#### 4. ПРАКТИЧЕСКОЕ ПРИМЕНЕНИЕ

В данной главе мы применим разработанный байесовский критерий к реальным наблюдательным данным гамма-астрономии и посмотрим, какие выводы сделает наш новый метод относительно известных наблюдений. Наблюдения будут браться из нескольких источников: [6], [7], [8], [9], [10]. В первой представлены измерения спектра Крабовидной туманности телескопами HEGRA и H.E.S.S., в остальных – наблюдения гамма-всплесков установками VERITAS и Fermi-LAT. В нашей совокупной выборке присутствуют как предельно малые, так и умеренные количества событий.

Таблица 4.1: Результаты анализа гамма-всплесков.  $P(\lambda_{\text{on}} > \lambda_{\text{off}})$  – байесовская вероятность наличия сигнала, вычисляемая по формуле (1.6);  $S$  – статистика из теоремы Уилкса (0.2) и соответствующее ей p-value. Зеленым выделены наблюдения, демонстрирующие статистическую значимость наличия сигнала на уровне значимости  $\alpha = 1\%$  (C.L. = 99%); серым – отсутствие сигнала на уровне значимости  $\alpha = 1\%$  (C.L. = 99%).

Источник	$N_{\text{on}}$	$N_{\text{off}}$	$\epsilon$	$P(\lambda_{\text{on}} > \lambda_{\text{off}})$	$S$	p-value
HEGRA	36	104	0.2	0.996	7.32	0.007
H.E.S.S.	4	1	0.095	> 0.999	14.73	$10^{-4}$
GRB 070419A	2	14	0.057	0.808	1.18	0.278
GRB 070521	3	113	0.057	0.049	2.20	0.138
GRB 070612B	3	21	0.066	0.832	1.29	0.256
GRB 080310	3	23	0.128	0.449	$9 \cdot 10^{-4}$	0.976
GRB 080604	2	40	0.063	0.293	0.11	0.741
GRB 080607	4	16	0.112	0.883	1.74	0.187
GRB 080825C	15	19	0.063	> 0.999	40.43	$2 \cdot 10^{-10}$
GRB 081024A	1	7	0.142	0.395	$3 \cdot 10^{-5}$	0.996
GRB 090418A	3	16	0.123	0.685	0.41	0.523
GRB 090429B	2	7	0.106	0.825	1.26	0.262
GRB 090515	4	24	0.126	0.642	0.25	0.616

Несмотря на то, что в случае с теоремой Уилкса не стоит сильно доверять позициям с фоном меньше хотя бы 25 событий (2.1), оба метода показывают одинаковые результаты на этих данных. Однако это лишь несколько экспериментов: как мы показывали выше, байес демонстрирует большую чувствительность и является более приоритетным в использовании методом.

Помимо прочего, в байесовском подходе мы получаем интерпретируемую величину – вероятность того, что альтернативная гипотеза верна. Это позволяет нам трактовать наборы данных не только с точки зрения статистической значимости, но и степени уверенности в нашей гипотезе. К примеру, исходя из наших результатов (4.1): GRB 070419A, GRB 070612B, GRB 080607 и GRB 090429B имеют достаточно высокую вероятность наличия сигнала – такие значения могут рассматриваться как указание на возможный сигнал, который не фиксируется классическим бинарным критерием.

Таким образом, мы продемонстрировали согласованность принципиально разных методов для ярко выраженных источников сигнала, а также обсудили возможную интерпретацию результатов байесовского подхода.

## ВЫВОДЫ

В рамках данной дипломной работы было проведено детальное сравнение классического частотного и байесовского подходов к проблеме обнаружения источника сигнала в условиях неопределенного фона. Особое внимание уделялось корректности работы методов в режиме малой статистики, характерной для реальных астрофизических наблюдений, в частности, в гамма-астрономии.

В ходе исследования были получены следующие ключевые результаты.

Методом Монте-Карло подтверждено, что широко распространенный критерий Ли-Ма, опирающийся на теорему Уилкса об асимптотическом поведении отношения правдоподобия, демонстрирует некорректное поведение при малом количестве наблюдаемых событий. В частности, при параметре фона  $b \lesssim 25$  (при  $\epsilon = 0.2$ ) реальное распределение тестовой статистики значительно отклоняется от аппроксимации  $\chi_1^2$ , что приводит к неконтролируемому росту ошибки I рода.

Детально разработан и применен альтернативный байесовский подход, основанный на прямом вычислении вероятности гипотезы о наличии сигнала  $P(\lambda_{\text{on}} > \lambda_{\text{off}})$  по формуле (1.6). Используя сопряженное гамма-распределение для пуассоновской модели данных, было получено точное аналитическое выражение для этой вероятности, не требующее аппроксимаций больших выборок.

Проведен сравнительный анализ чувствительности методов, который показал, что при должном контроле ошибки I рода байесовский критерий обладает большей статистической мощностью (меньшей ошибкой II рода) в области слабых сигналов. Это преимущество исчезает с ростом интенсивности сигнала, где оба подхода сходятся к достоверному обнаружению.

Исследована робастность байесовского метода по отношению к нарушению предположения о пуассоновской природе фона. Введение дополнительной дисперсии, моделируемой отрицательным биномиальным распределением, показало, что байесовский критерий сохраняет устойчивость при умеренных отклонениях. Это делает его предпочтительным инструментом анализа реальных наблюдательных данных.

На примере симуляции процедуры «подглядывания» показано, что наивное применение байесовского критерия не решает проблему роста ошибки I рода при последовательном анализе данных, аналогично частотному. Эффект роста вероятности ложного открытия при многократной проверке гипотез требует использования дополнительных корректирующих процедур.

Практическое применение разработанного метода к реальным наблюдениям гамма-всплесков в целом подтвердило выводы классического анализа для статистически значимых источников. Однако это также позволило подсветить кандидатов на дополнительные исследования на основании высокой, пусть и не превосходящей порог, вероятности наличия сигнала.

## ЗАКЛЮЧЕНИЕ

Таким образом, байесовский подход к проблеме on/off является не просто математической альтернативой, а дает исследователю более интерпретируемую, чувствительную и устойчивую меру уверенности в обнаружении сигнала, что особенно важно на границе чувствительности современных астрофизических экспериментов.

Непосредственное и повсеместное внедрение разработанного аналитического байесовского критерия в стандартный набор инструментов обработки данных обсерваторий позволит вывести исследования на следующий уровень.

Перспективой для дальнейших исследований служит использование информативного приора (в данной работе использовался неинформативный приор Джеффриса (1.2)), поскольку в реальных задачах мы часто обладаем определенной информацией об источниках, а информативные приоры позволяют повысить чувствительность анализа. А также необходимо разработать решение проблемы «подглядывания» путем создания и калибровки байесовских стоп-правил, аналогичных последовательным критериям в частотных методах проверки гипотез.

## СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Ti-pei Li and Yu-qian Ma. ANALYSIS METHODS FOR RESULTS IN GAMMA-RAY ASTRONOMY. The Astrophysical Journal, 272:317-324, 1983 September 1.
2. Wilks, S. 1962, Mathematical Statistics (New York: Wiley).
3. Jeffreys H. (1957). Scientific Inference. 2nd ed. Cambridge University Press.
4. Журнал «Прикладная эконометрика» № 1(9) 2008.
5. James Berger. A Comparison of Testing Methodologies. Duke University, Durham NC, USA.
6. Grigory Rubtsov, Petr Satunin, Sergey Sibiryakov. Constraints on violation of Lorentz invariance from atmospheric showers initiated by multi-TeV photons. arXiv:1611.10125v2 [astro-ph.HE] 26 Dec 2016.
7. A.A.Abdo et al., Astrophys.J. 707 (2009) 580.
8. V.A.Acciari et al., Astrophys.J. 743 (2011) 62.
9. M.L.Knoetig, Astrophys.J. 790 (2014) 106.
10. D.Casadei, Astrophys.J. 798 (2015) 5.