

Байесовская деконволюция для эксперимента СФЕРА-2

Игорь Вайман

3 мая 2021 г.

Глава 1

Описание эксперимента СФЕРА-2

Эксперимент СФЕРА-2 основан на сборе и регистрации черенковского света ШАЛ, отражённого от «экрана» – ровной заснеженной поверхности льда озера Байкал. Установка представляет собой сферическое зеркало и мозаику ФЭУ, установленную вблизи фокальной поверхности. Черенковский свет ШАЛ, рассеиваясь на снегу, проходит через диафрагму установки, отражается от зеркала, и попадает на фотокатод. При попадании фотона с длиной волны в области чувствительности фотокатода с него выбиваются 1 – 10 (уточнить) вторичных электронов, которые проходят систему динодов с разностью потенциалов между каждой парой, создавая в результате избыточный на аноде. Этот заряд стекает с анода через цепь усиления, создавая в результате напряжение на входе АЦП. Это напряжение считывается электроникой, в результате и составляя (вместе с некоторой телеметрией) экспериментальные данные.

Этот набор физических процессов мы описываем в следующих положениях:

1. В системе нет электронных шумов – единственным источником шума являются фоновые фотоны (от звёздного и зодиакального света, так как эксперимент проводится в ясные безлунные ночи). Фотоны шума попадают в установку аналогично «сигнальным» фотонам – после рассеяния на поверхности. Их поток можно приблизительно оценить по каталогам светимости звёзд и из моделей зодиакального света. (Ссылка на работу Энтиной)

2. Работу ФЭУ можно эффективно описать одной «рандомизированной» импульсной характеристикой (РИХ). Понятие и применение РИХ будет раскрыто далее, в общем виде это случайная функция времени, описывающая отклик системы на один фотон, падающий на фотокатод. Главный источник случайности в установке СФЕРА-2 – неопределённость числа вторичных фотонов, вылетевших с фотокатода, приводящая к неопределённости амплитуды импульса тока. Стоит отметить, что случайный характер ИХ проявляется именно на границе между режимом счёта фотонов и режимом измерения потока, поскольку поток фотонов уже слишком велик, чтобы разрешать отдельные импульсы, но недостаточно – чтобы случайные характеристики эффективно усреднялись. Мы также предполагаем, что случайность импульсной характеристики проявлялась независимо между отдельными фотонами и между разными ФЭУ.
3. Используя экспериментальные данные (данные, записанные АЦП, постоянную компоненту анодного тока и абсолютную калибровку ФЭУ) мы можем восстановить показания анодного тока (ссылка на работу по электронике). Неопределённость, вносимая анодной цепью, усилителем и (главным образом) дискретизацией сигнала на входе АЦП, хорошо известна и учитывается во вторую очередь.

Глава 2

Байесовская деконволюция

2.1 Общие допущения

Рассмотрим N последовательных равновеликих временных бинов. Для простоты будем считать единицей времени длительность одного бина. Тогда мы можем задать временные бины интервалами $[i-1, i]$, $i = 1 \dots N$. Обратим внимание, что при такой нумерации бинов эффект от фотонов, попавших в i -тый временной бин проявляется впервые в момент времени i . Обозначим число фотонов в каждом бине как n_i , $i = 1, \dots, N$.

Время прихода отдельного фотона относительно начала бина $t_{inbin} = \{t\}$ будем описывать случайной величиной, подразумевая, что мы не интересуемся такими детальными характеристиками сигнала. Эта величина может быть, вообще говоря, распределена произвольным образом в интервале $[0, 1)$, однако мы в простейшем случае будем считать $t_{inbin} \sim U(0, 1)$. Это оправдано для независимых друг от друга фоновых фотонов, и может служить приближением для фотонов ШАЛ в случае, если дисперсия времён прихода фотонов внутри «пакета» сильно превышает длительность временного бина. По данным модельных ливней это не всегда так, поэтому влияние неравномерности распределения времён прихода фотонов будет исследовано отдельно.

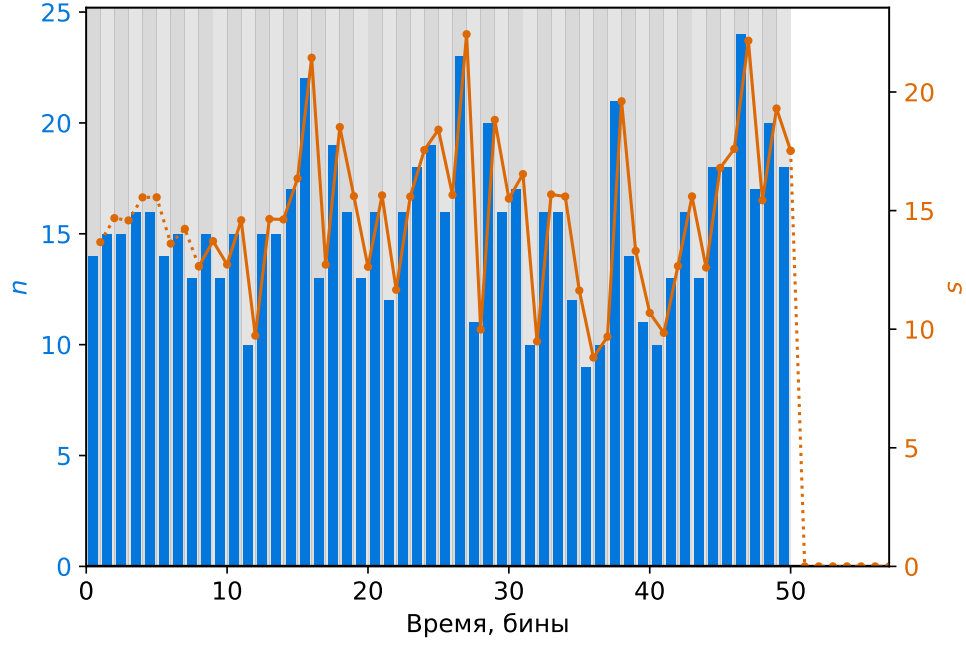


Рис. 2.1: Пример данных для задачи байесовской деконволюции. Здесь $N = 50$, количество фотонов в каждом бине выбрано из пуассоновского распределения с $\lambda = \mathbb{E}(n_i) = 15$, они показаны синими столбцами; РИХ – $A \exp(-t/2)$, обрезанная на $L = 10$, где случайный множитель $A \sim U(0.75, 1)$; оранжевые точки – выходной сигнал; пунктиром показаны участки, исключаемые из рассмотрения из-за краевых эффектов.

Предположим, что импульсная характеристика системы – случайная функция $\tilde{h}(t)$ в том смысле, что для входного сигнала, состоящего из конечного числа дельта-функций на единицу времени, каждая из этих дельта функций сворачивается с отдельной независимой реализацией $h(t) \sim \tilde{h}(t)$. Будем также считать, что любая реализация удовлетворяет условию каузальности, то есть $\forall t < 0 \quad \forall h \sim \tilde{h} \quad h(t) = 0$, и конечности во времени, то есть $\exists \tilde{L}$ такое, что $\forall t > \tilde{L} \quad \tilde{h}(t) = 0$.

Заметим, что эффект от фотонов в i -том бине проявляется в отсчётах с i по $i + \lfloor \tilde{L} \rfloor$, где $\lfloor \tilde{L} \rfloor$ – наибольшее целое число, не превышающее \tilde{L} (округление вниз). Обозначим $L \equiv \lfloor \tilde{L} \rfloor$. Тогда полный сигнал от фотонов

гарантированно содержится в отсчётах с 1 по $N + L$.

Таким образом, фотоны, приходящие на вход в течение времени $[0, N]$ будут приводить к регистрации отсчётов s_j в моменты времени $j = 1, \dots, N + L$. Иллюстрация этой постановки задачи приведена на рис. 2.1.

2.1.1 Краевые эффекты

В реальном эксперименте входные фотоны не ограничены интервалом $[0; N]$, но приходят постоянно. Модифицируем постановку задачи так, чтобы устранить краевые эффекты – плавный рост сигнала в начале и затухание в конце.

Для этого достаточно исключить из всего дальнейшего рассмотрения эти участки сигнала. Так, в реальной ситуации фотоны продолжают приходить после $t = N$, и вносят соответствующий вклад в отсчёты начиная с $N - 1$. Фотоны, пришедшие до $t = 0$, вносят вклад в отсчёты до L .

Поэтому для восстановления значений n_i , $i = 1, \dots, N$ мы будем использовать только отсчёты S_j при $j = L + 1, \dots, N$. Этот участок изображён на рис. 2.1 сплошной линией, участки по краям, исключаемые из рассмотрения – пунктиром.

2.2 Постановка задачи

Поставим задачу статистической деконволюции следующим образом, используя байесовскую терминологию (поэтому будем также называть эту процедуру байесовской деконволюцией):

Пусть дана рандомизированная импульсная характеристика системы $\tilde{h}(t)$ и значения s_j , $j = 1, \dots, N + L$. Найти апостериорные функции плотности вероятности для значений n_i , $i = 1, \dots, N$.

Заметим, что, в отличие от обычной деконволюции, мы не ставим задачу оценить исходный сигнал сам по себе, представляющий собой сумму δ -функций, но только его несколько обобщённую характеристику.

2.3 Решение

2.3.1 Выходной сигнал как реализация случайного процесса

Ясно, что в силу случайного характера отклика системы, а также агрегирования фотонов в бины, значения отсчётов $\{s_j\}$ являются реализациями некоторых случайных величин. Обозначим сами эти случайные величины как $\{S_j\}$.

Запишем S_j как сумму вкладов от фотонов разных бинов

$$S_j = \sum_{l=0}^L C(n_{j-l}, l) \quad (2.1)$$

Здесь $C(n, l)$ – случайная величина, описывающая вклад в сигнал на j -том временном отсчёте от n фотонов в бине $j - l$, иначе говоря, вклад с *задержкой* l бинов. Из выбранной схемы индексации бинов и условий каузальности и ограниченности во времени РИХ легко видеть, что $l \in [0, L]$, поскольку вклад от фотонов более ранних бинов равен нулю.

Охарактеризуем распределение $C(n, l)$. Проще всего сделать это через Монте-Карло-сэмплирование распределения этой величины. Получим сначала с произвольной точностью эмпирическую функцию плотности распределения для $C(1, l)$. Для этого сгенерируем значения $t_k \sim t_{inbin}$; и функции $h_k(t) \sim \tilde{h}(t)$ для $k = 1 \dots N_{sample}$. Выборка для $C(1, l)$ тогда будет состоять из значений $h_k(l + 1 - t_k)$. Выборка для $C(n, l)$ легко получить, проделав описанную процедуру n раз и сложив все n реализаций N_{sample} -мерных векторов выборок.

2.3.2 Грубая оценка

Перед тем, как решать задачу деконволюции в статистическом смысле, сделаем грубую оценку \vec{n} , основанную только на соотношениях между математическими ожиданиями случайных величин. Для этого применим операцию вычисления математического ожидания к обеим частям равенства (2.1).

Фотоны независимы друг от друга в пределах одного бина, из чего следует $\mathbb{E} C(n, l) = n \mathbb{E} C(1, l)$. Нетрудно заранее вычислить для данной

РИХ выборки значений $C(1, l)$ для $l = 0 \dots L$. Тогда получим, обозначая $c_l \equiv \mathbb{E} C(1, l)$,

$$\bar{S}_j = \mathbb{E} S_j = \sum_{l=0}^L \mathbb{E} C(n_{j-l}, l) = \sum_{l=0}^L n_{j-l} \mathbb{E} C(1, l) = \sum_{l=0}^L n_{j-l} c_l \quad (2.2)$$

Суммирование можно записать в матричном виде для $j = 1, \dots, N + L$:

$$\begin{pmatrix} c_0 & 0 & 0 & \dots & 0 \\ c_1 & c_0 & 0 & \dots & 0 \\ c_2 & c_1 & c_0 & & \vdots \\ c_3 & c_2 & c_1 & \ddots & 0 \\ \vdots & c_3 & c_2 & \ddots & c_0 \\ c_L & \vdots & c_3 & \ddots & c_1 \\ 0 & c_L & \vdots & \ddots & c_2 \\ \vdots & & c_L & & c_3 \\ 0 & \dots & 0 & \ddots & \vdots \\ 0 & \dots & 0 & 0 & c_L \end{pmatrix} \begin{pmatrix} n_1 \\ n_2 \\ n_3 \\ \vdots \\ n_N \end{pmatrix} = \begin{pmatrix} \bar{S}_1 \\ \bar{S}_2 \\ \vdots \\ \bar{S}_N \\ \bar{S}_{N+1} \\ \vdots \\ \bar{S}_{N+L} \end{pmatrix} \quad (2.3)$$

Для перехода к случаю непрерывного потока фотонов достаточно, как указано в разделе 2.1.1, ограничиться рассмотрением строк $j = L + 1, \dots, N$, то есть убрать первые и последние L уравнений системы.

Эта возможно несовместная система линейных уравнений, допускает решение в смысле наименьших квадратов с помощью псевдообратной матрицы Мура-Пенроуза [5]. Псевдообратная матрица C^+ для C определяется следующими условиями: (1) $CC^+C = C$, (2) $C^+CC^+ = C^+$, (3) CC^+ и C^+C – эрмитовы матрицы. Псевдообратная матрица всегда существует, и для системы $C\vec{n} = \vec{S}$ вектор $C^+\vec{S}$ даёт искомое МНК-решение системы. Приближённый численный расчёт такой матрицы можно провести, например, с помощью функции `pinv` модуля `numpy.linalg` в Python [4].

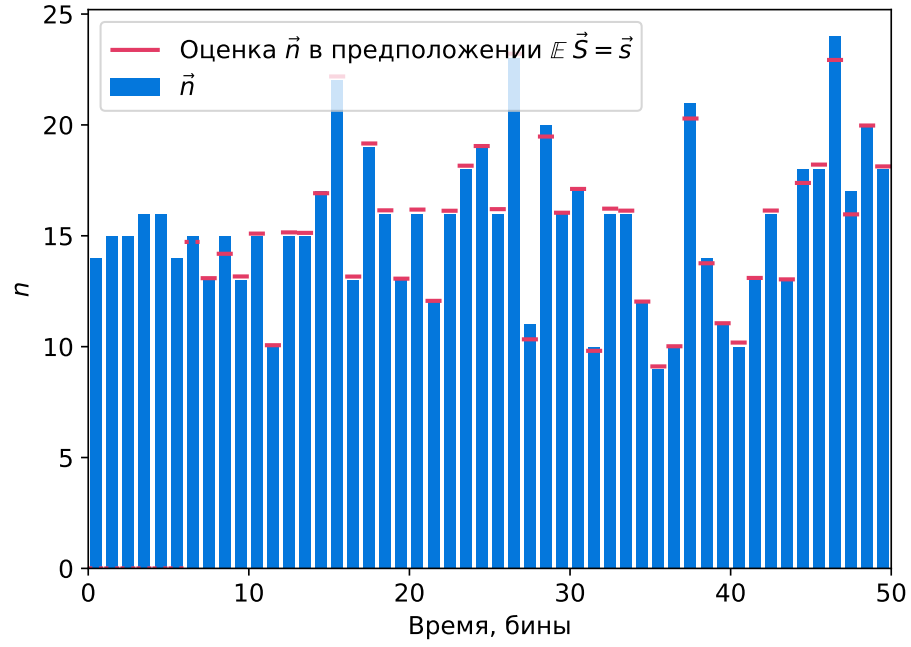


Рис. 2.2: Оценка \vec{n} в предположении, что выходной сигнал \vec{s} равен своему математическому ожиданию. Пунктиром показана область, где оценка искажена эффектом ограниченности выборки во времени.

Однако для решения системы необходимо знать $\mathbb{E} S_j$, в то время как в эксперименте мы имеем всего лишь единственную реализацию этой случайной величины s_j . Для грубой оценки остаётся положить $\mathbb{E} S_j \cong s_j$. Результат описанной процедуры приведён на рис. 2.2 фиолетовым, пунктиром отмечена область, в которой оценка искажена краевым эффектом, её необходимо отбросить.

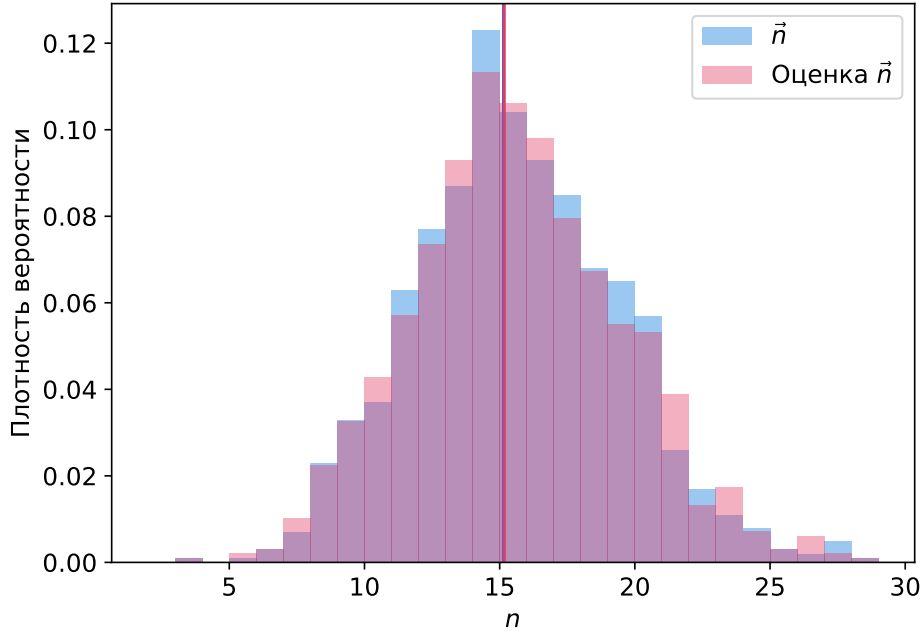


Рис. 2.3: Распределения истинных и грубо восстановленных значений \vec{n} для 1000 значений. Истинные значения выбраны из пуассоновского распределения с $\lambda = 15$, аналогично рис. 2.1 и 2.2.

Строгое исследование свойств такой «one-shot» оценки \vec{n} находится за рамками данной работы. Однако для простой численной проверки можно провести описанную процедуру для большего числа входных бинов и сравнить распределения истинных и оцененных значений \vec{n} . На рис. 2.3 приведено такое сравнение для 1000 бинов. Видно, что распределения практически совпадают, а значит отсутствует по крайней мере систематическая ошибка. В дальнейшем эта оценка будет играть роль первого приближения, или стартовой точки, на основе которой уже полными статистическим методом можно найти полное решение.

2.3.3 Решение задачи байесовской деконволюции

Теперь перейдём к решению полноценной статистической задачи, используя метод, описанный в предыдущем разделе, как первое прибли-

жение.

Запишем сначала теорему Байеса в общем виде, учитывая, что наблюдаемыми значениями является сигнал \vec{s} , а неизвестными параметрами, которые задают распределение наблюдаемых – \vec{n} :

$$P(\vec{n}|\vec{s}) = \frac{P(\vec{s}|\vec{n}) P(\vec{n})}{P(\vec{s})} \quad (2.4)$$

Поясним вероятности, входящие в выражение:

1. $P(\vec{n}|\vec{s})$ – искомое *апостериорное* распределение, описывающее (в байесовском определении вероятности) наши знания о \vec{n} после проведения измерений
2. $P(\vec{s}|\vec{n}) \equiv \mathcal{L}(\vec{s}, \vec{n})$ – функция правдоподобия, описывающая, насколько вероятна регистрация определённых значений \vec{s} при заданных параметрах \vec{n}
3. $P(\vec{n}) \equiv \pi(\vec{n})$ – *априорное* распределение \vec{n} , не требующее знаний о конкретном выходном сигнале
4. $P(\vec{s})$ – полная вероятность регистрации данного сигнала при всех возможных значениях \vec{n} . Она также называемая маргинальной вероятностью или нормировочным множителем, по определению $P(\vec{s}) = \int_{\infty} P(\vec{s}|\vec{n}) P(\vec{n}) d\vec{n}$. Эту величину можно использовать для сравнения моделей, например, если бы мы имели в распоряжении альтернативную модель генерации значений \vec{s} и хотели бы понять, какая из двух лучше описывает данные. Однако для поставленной задачи нет нужды ни вычислять, ни даже учитывать этот множитель.

Учитывая введённые обозначения, запишем

$$P(\vec{n}|\vec{s}) \propto \mathcal{L}(\vec{s}, \vec{n}) \pi(\vec{n}) \quad (2.5)$$

2.3.3.1 Выбор априорного распределения

В качестве априорного распределения будем использовать неограниченное равномерное, или *неинформативное* распределение. Методы байесовской статистики работают даже в ситуации, когда у нас нет вообще никакой информации о распределении \vec{n} до начала измерений. Формально для этого нужно положить $\pi(\vec{n}) = \text{Const} \forall \vec{n}$. Такое распределение

нельзя использовать напрямую, поскольку его невозможно отнормировать на 1. Однако тогда в выражении 2.5 можно просто исключить $\pi(\vec{n})$ из правой части.

2.3.3.2 Полное вычисление функции правдоподобия методом Монте-Карло

Функция правдоподобия $\mathcal{L}(\vec{s}, \vec{n})$ определяется как вероятность того, что данный сигнал \vec{s} получился в результате преобразования системой входного сигнала \vec{n} .

Опишем сначала Монте-Карло метод оценки правдоподобия, и затем введём упрощения, которые позволят эффективнее вычислять эту функцию, а также обоснуем корректность этих упрощений. Для этого при фиксированных \vec{n} необходимо смоделировать выборку из большого числа реализаций случайной величины \vec{S} , а затем оценить плотность вероятности в точке \vec{s} по этой выборке.

Существует несколько методов для такой оценки, самый простой из которых – многомерная гистограмма. Для этого пространство реализаций \vec{S} нужно поделить на ячейки, подсчитать число элементов выборки, попавших в каждую ячейку, и поделить на общее число элементов выборки и на объём ячейки. Нетрудно видеть, что полученное число даст оценку плотности вероятности в произвольной точке – а значит, искомую оценку функции правдоподобия.

Однако оказывается, что для поставленной задачи метод прямого вычисления функции правдоподобия плохо подходит из-за «проклятия размерности». Размерность пространства, в котором нужно оценить эмпирическую функцию плотности вероятности, равна ширине окна, в котором рассматривается входной сигнал, в бинах. Соответственно, общее число N -мерных бинов составляет g^N , где g – мощность бинирования каждого отсчёта. В этой ситуации для надёжной оценки придётся генерировать выборку сравнимого объёма.

Поэтому вместо прямого использования описанного метода аппроксимируем распределение \vec{S} более вычислительно эффективной функцией.

2.3.3.3 Аппроксимация функции правдоподобия многомерным нормальным распределением

Сначала проанализируем результат полной Монте-Карло симуляции, чтобы обосновать корректность аппроксимации.

Частные распределения Из общих соображения ясно, что распределение $S_j \forall j$ будет более или менее близко к нормальному просто в силу того, что каждая из этих величин является суммой независимых вкладов от фотонов в предыдущих бинах. Также из общих соображений можно сказать, что *чем меньше среднее число фотонов в бине, тем хуже будет аппроксимация нормальным распределением*. На рис. 2.4 приведены несколько распределений «лучшего», «медианного» и «худшего» распределений S_j для нескольких значений среднего число фотонов в бине $\bar{n} \equiv \mathbb{E} n_i$. Из предварительного анализа и независимых соображений известно, что менее 5 фотонов на бин – весьма нечастая ситуация в экспериментальных данных – но даже в этом случае видно, что самое существенное среднеквадратичное отклонение плотности распределения от гауссианы не превысит 10%. Таким образом, *частное распределение* каждой компоненты вектора \vec{S} можно приблизить нормальным распределением с эффективной погрешностью в несколько процентов для практически значимых случаев.

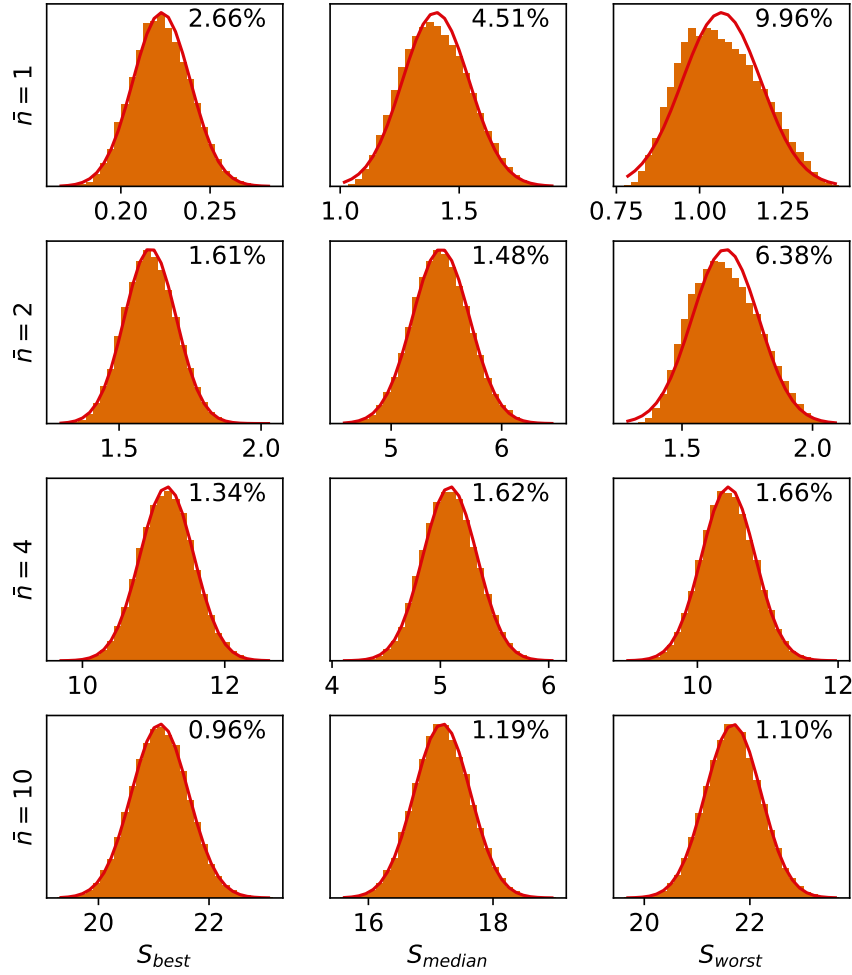


Рис. 2.4: Оценка нормальности частных распределений компонент \vec{S} для разных интенсивностей входного потока: от экстремально низких 1 и 2 до средней ожидаемой 10. «Лучший», «медианный» и «худший» бины определены эвристически по квадрату разницы выборочного среднего и медианы (чем больше эта величина, тем менее симметрично распределение). На каждом графике построена плотность нормального распределения с соответствующими μ и σ , а также приведено среднеквадратичное отклонение гистограммы от гауссианы в процентах относительно максимального значения.

Полное $(N - L)$ -мерное распределение Описанная процедура для полного многомерного распределения представляет собой непростую задачу, поэтому ограничимся здесь визуальным исследованием двумерных распределений пар значений в отсчётах, разделённых не более чем L бинами (из постановки задачи ясно, что на больших расстояниях S_j являются независимыми). На рис. 2.5 представлены такие распределения. Видно, что двумерные распределения представляют собой овалы правильной симметричной формы. Таким образом, можно положить, что *распределение вектора \vec{S} можно приблизить многомерным нормальным распределением.*

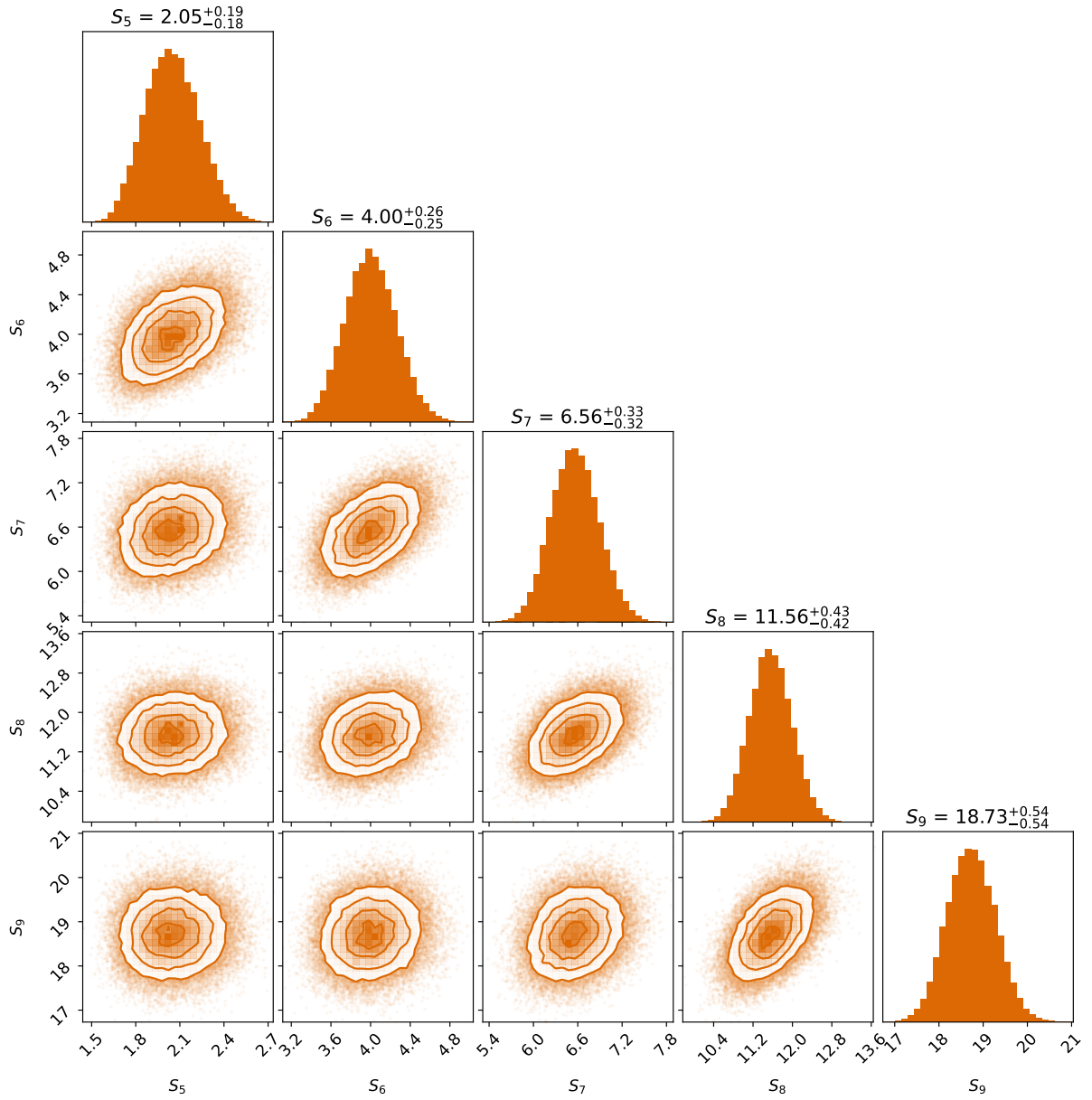


Рис. 2.5: Оценка нормальности совместных двумерных распределений для пар компонент \vec{S} . График построен с помощью библиотеки corner [2].

Таким образом, мы имеем право построить аппроксимацию функции правдоподобия $\mathcal{L}(\vec{n}) \equiv P(\vec{S}|\vec{n})$, гораздо более устойчивую и вычислительно эффективную, чем наивный метод Монте-Карло, описанный в разделе 2.3.3.2.

Многомерное нормальное распределение в общем виде задаётся вектором математических ожиданий $\vec{\mu}$ и матрицей ковариаций Σ . Функция плотности вероятности $(N + L)$ -мерного случайного вектора \vec{S} тогда вычисляется как

$$p(\vec{S}) = ((2\pi)^{N+L} \det \Sigma)^{-1/2} \exp \left(-\frac{1}{2} (\vec{S} - \vec{\mu})^T \Sigma^{-1} (\vec{S} - \vec{\mu}) \right) \quad (2.6)$$

Вектор $\vec{\mu} \equiv \mathbb{E} \vec{S}$ при данных \vec{n} вычисляется в соответствии с выражением (2.3). Вычислим произвольный элемент матрицы $\Sigma_{ij} \equiv \text{cov}(S_i, S_j)$. В силу симметричности матрицы ковариаций можно положить для определённости $i \leq j$. Запишем выражение (2.1) для двух интересующих нас элементов \vec{S} :

$$S_i = \sum_{l=0}^L C(n_{i-l}, l) \quad (2.7)$$

$$S_j = \sum_{k=0}^L C(n_{j-k}, k) \quad (2.8)$$

Источник возможной ненулевой ковариации этих величин – члены, описывающие вклады фотонов из одного бина, но на разных задержках, все прочие слагаемые независимы. Условие, описывающее такие попарно-скоррелированные члены, получается из равенства индексов n в суммах: $i - l = j - k$. Вводя $\Delta \equiv j - i$, отсюда получаем $k = l + \Delta$. Отсюда сразу следует интуитивный вывод о том, что при $\Delta > L$ ковариация будет заведомо нулевой в силу конечности РИХ во времени.

Используем следующее свойство ковариации: если x, y, ϵ, η – случайные величины, из которых только x и y являются зависимыми, то $\text{cov}(x + \epsilon, y + \eta) = \text{cov}(x, y)$. Иначе говоря, заведомо независимые члены при вычислении ковариации можно просто опустить. Прделаем это в суммах выше, получив таким образом «скоррелированные части» \hat{S}_i и \hat{S}_j такие, что $\text{cov}(S_i, S_j) = \text{cov}(\hat{S}_i, \hat{S}_j)$:

$$\hat{S}_i = \sum_{l=0}^{L-\Delta} C(n_{i-l}, l) \quad (2.9)$$

$$\hat{S}_j = \sum_{k=\Delta}^L C(n_{j-k}, k) = [k = l + \Delta] = \sum_{l=0}^{L-\Delta} C(n_{i-l}, l + \Delta) \quad (2.10)$$

Наконец, используем ещё одно свойство ковариации: если среди случайных величин x_1, x_2, y_1, y_2 независимы все пары, кроме x_1 и y_1, x_2 и y_2 , то $\text{cov}(x_1 + x_2, y_1 + y_2) = \text{cov}(x_1, y_1) + \text{cov}(x_2, y_2)$. Индуктивно обобщая на случай сумм с произвольным числом членов, и принимая во внимание, что фотоны в разных бинах независимы, получаем

$$\text{cov}(S_i, S_j) = \text{cov}(\hat{S}_i, \hat{S}_j) = \sum_{l=0}^{L-\Delta} \text{cov}(C(n_{i-l}, l), C(n_{i-l}, l + \Delta)) \quad (2.11)$$

Наконец, принимая во внимание, что сама по себе величина $C(n, l)$ есть сумма n независимых одинаково распределённых случайных величин, а между членами сумм $C(n, l)$ и $C(n, l + \Delta)$ есть только попарные корреляции в отношении 1 к 1, можно записать окончательную формулу для вычисления элемента матрицы ковариации:

$$\text{cov}(S_i, S_{i+\Delta}) = \sum_{l=0}^{L-\Delta} n_{i-l} \text{cov}(C(1, l), C(1, l + \Delta)) \quad (2.12)$$

Из формулы видно, что автокорреляция сигнала описана полностью в терминах автокорреляции РИХ и чисел фотонов в соответствующих бинах, служащих весами для вклада разных участков РИХ.

Для удобства вычислений можно записать это равенство в матричном виде для значений $\Delta \in [0; L]$. Для этого введём обозначение $\xi(l, \Delta) \equiv \text{cov}(C(1, l), C(1, l + \Delta))$ и с помощью него запишем:

$$\begin{pmatrix} \xi(L, 0) & \xi(L-1, 0) & \xi(L-2, 0) & \cdots & \xi(0, 0) \\ 0 & \xi(L-1, 1) & \xi(L-2, 1) & \cdots & \xi(0, 1) \\ 0 & 0 & \xi(L-2, 2) & \cdots & \xi(0, 2) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \xi(0, L) \end{pmatrix} \begin{pmatrix} n_{i-L} \\ n_{i-(L-1)} \\ n_{i-(L-2)} \\ \vdots \\ n_i \end{pmatrix} = \begin{pmatrix} \Sigma_{i,i} \\ \Sigma_{i,i+1} \\ \Sigma_{i,i+2} \\ \vdots \\ \Sigma_{i,i+L} \end{pmatrix} \quad (2.13)$$

Это достаточно громоздкое и неинтуитивное выражение позволяет, однако, единожды вычислить матрицу Ξ и эффективно рассчитывать матрицу ковариаций Σ при каждом заданном \vec{n} .

После исключения областей влияния краевых эффектов – первых и последних L элементов \vec{S} – нас будут интересовать только внутренняя часть матрицы Σ , то есть индексы $i, j \in [L + 1; N]$. Это удобно, так как для вычисления интересующей нас части Σ не нужно рассматривать «урезанную» матрицу Ξ , что потребовалось бы, если бы мы не исключили краевые эффекты.

2.3.3.4 Сэмплирование апостериорного распределения

Имея функцию правдоподобия \mathcal{L} и учитывая выбранные неинформативные априорные распределения, мы получаем функцию плотности апостериорной вероятности $P(\vec{n}|\vec{s}) \propto \mathcal{L}(\vec{n}, \vec{s})$, которую требуется исследовать.

Требуется получить не только наиболее оптимальное значение \vec{n} , но и оценку неопределённости этой величины. Поэтому, например, простой метод максимального правдоподобия, в котором мы нашли бы максимум \mathcal{L} в пространстве параметров, не подходит – хотя существуют методы поиска доверительных интервалов. Мы будем использовать метод Монте-Карло с марковскими цепями (Markov Chain Monte Carlo), который хорошо применим к задачам большой размерности. Суть метода сводится к запуску марковского процесса случайного блуждания в пространстве параметров \vec{n} с такой специально подобранной вероятностью перехода, зависящей от апостериорной вероятности, что в пределе это блуждание даст выборку из исследуемой функции плотности. Общей характеристикой этих методов является то, что они зависят только от отношения вероятностей в разных точках пространства параметров – именно поэтому в выражении 2.5 и во всех последующих вычислениях мы не интересовались ни полной маргинальной вероятностью наблюдаемых данных, ни нормировкой априорного распределения.

Подробный обзор и теоретическое исследование весьма широкой области МСМС-сэмплирования находятся за рамками данной работы, но могут быть найдены, например, в обзоре [1].

В данной работе будем использовать популярный алгоритм, предложенный в 2010 году в работе [3] и его имплементацию в пакете `emcee` [2] на языке `Python`.

Список литературы

- [1] Mary Kathryn Cowles и Bradley P. Carlin. “Markov Chain Monte Carlo Convergence Diagnostics: A Comparative Review”. в: *Journal of the American Statistical Association* 91.434 (июнь 1996), с. 883—904. DOI: 10.1080/01621459.1996.10476956.
- [2] Daniel Foreman-Mackey. “corner.py: Scatterplot matrices in Python”. в: *The Journal of Open Source Software* 1.2 (июнь 2016), с. 24. DOI: 10.21105/joss.00024.
- [3] Jonathan Goodman и Jonathan Weare. “Ensemble samplers with affine invariance”. в: *Communications in Applied Mathematics and Computational Science* 5.1 (янв. 2010), с. 65—80. DOI: 10.2140/camcos.2010.5.65.
- [4] Charles R. Harris и др. “Array programming with NumPy”. в: *Nature* 585.7825 (сент. 2020), с. 357—362. DOI: 10.1038/s41586-020-2649-2. URL: <https://doi.org/10.1038/s41586-020-2649-2>.
- [5] R. Penrose. “On best approximate solutions of linear matrix equations”. в: *Mathematical Proceedings of the Cambridge Philosophical Society* 52.1 (янв. 1956), с. 17—19. DOI: 10.1017/s0305004100030929.