

Задание 3

Тема: Статистические оценки

Для сдачи: 5 вопросов + 3 учебные задачи

Ограничения: 6 человек на 1 вопрос/задачу

Вопросы:

1. Какая оценка называется состоятельной, несмещенной и эффективной?
2. Могут ли быть случаи, когда можно сделать интервальную оценку, но нельзя точечную?
3. При каком распределении логарифм функции правдоподобия имеет вид параболы?
4. Как связаны метод наименьших квадратов, метод хи-квадрата и метод максимума правдоподобия?
5. В каких случаях эффективность оценки хи-квадрат и оценки методом максимума правдоподобия будут отличаться? В какую сторону?
6. Согласно телефонным опросам в [президентских выборах в США \(1936\)](#) должен был победить республиканец Альфред Лэндон, но с огромным отрывом победил демократ Франклин Рузвельт, хотя этот же опрос в выборах ранее правильно предсказал победу республиканцев. Объясните, что не так с этим опросом, каким свойством обладает его оценка.

Задачи:

Учебные:

1. При обработке данных в питоне основным пунктом является их [фитирование](#). Наиболее часто используется [curve_fit](#). Чтобы он выдавал правильные результаты, надо понимать специфику документации, для этого изучите ее. Затем сгенерируйте данные, используя нормальное распределение для генерации ошибки и полином 4 степени. Проведите своеобразный эксперимент с фитом: используйте полиномы большей, например 10-й, и меньшей, например 2-й, степени, подавайте разные параметры в [curve_fit](#): границы, ошибки, начальное положение.
2. Пусть с помощью сцинтиллятора вы измерили некий спектр, у которого есть два пика, данные можно смоделировать с помощью двух Гауссов (обоснуйте, почему можно использовать нормальное распределение). Сравните с заданными сигмами ширины на полувысотах пиков и значения, полученные с помощью фитирования. Получите точечные и интервальные оценки положений пиков.
3. Рассмотрим процесс $e^+e^- \rightarrow \mu^+\mu^-$, для которого угловое распределение мюонов имеет вид $f(x; \alpha, \beta) = \frac{1+\alpha x+\beta x^2}{2+2\beta/3}$, $x = \cos(\theta)$, θ — угол рассеяния. Допустим, что в данных заложена гауссова ошибка. Сгенерируйте данные, самостоятельно реализуйте Метод Максимального Правдоподобия и получите оценки параметров (истинными можно взять 0.5 и 0.5)
4. Постройте функцию правдоподобия для распределения Бернулли, данные сгенерируйте с помощью Бернулли, 1 будет соответствовать удачному испытанию, а 0 - неудачному. Постройте графики функции правдоподобия, логарифмической функции правдоподобия, ее производной и ее вторую, на один график следует нанести функции для нескольких сгенерированных выборок, например, 100, чтобы увидеть, какой разброс. Графики также лучше строить около "истинного" параметра θ , проделать процедуру можно для нескольких параметров. Подумайте, как полученные картинки можно сопоставить с информацией о параметре θ .
5. [Данные](#) это набор точек (x,y, yErr) вида подложка + пик. Определите положение и амплитуду пика и постройте доверительные интервалы (как минимум для положения).

Научные:

1. Скачайте данные из [гугл диска](#). Они представляют собой сигналы [SIPM](#), который был соединен со сцинтиллятором, регистрирующим протоны в ускорителе. Для загрузки используйте `data = np.load('raw_data_0.npy')`, каждая строка соответствует сигналу, то есть `data[0]` будет первым сигналом. Вам надо профитировать каждый сигнал, использовать можно аппроксимацию распределения Ландау. Так как сигналов много, то визуально все не проверить, а еще есть явно кривые сигналы, которые стоит отбросить. Для этого возьмите некую характеристику качества фита и постройте ее распределение по построенным фитам.