Метод максимального правдоподобия

В нашем случае выберем модель, которая максимизирует функцию качества называемую правдоподобием.

Правдоподобие $\mathcal{L}(\mathcal{M}|data)$ равно вероятности (или плотности) получить наши данные $data$ в предположении, что модель реальности равна $\mathcal{M}$.

PDMI COMPUTER SCIENCE CLUB

Метод максимального правдоподобия

Посчитаем, чему равно правдоподобие для $\mathcal{M} \in \mathfrak{M}$.

$$\mathcal{L}(\mathcal{M}|data) = \mathbb{P}(data|\mathcal{M}) = \prod_{i=1}^{138} \mathbb{P}((x,y,z)_i|\mathcal{M}) = \prod_{i=1}^{138} p_{x_i}^1 p_{y_i}^2 p_{z_i}^3 = \prod_{x,w} (p_x^w)^{n_x^w},$$

где n_x^w — количество картинок x в окошке w.

Оказывается, максимум правдоподобия достигается при $p_x^w = \frac{n_x^w}{138}$, то есть

наилучшая модель из $\mathfrak{M}$ это Модель 3 с уже рассмотренными параметрами!

Метод максимального правдоподобия

Посчитаем, чему равно правдоподобие для $\mathcal{M} \in \mathfrak{M}$.

$$\mathcal{L}(\mathcal{M}|data) = \mathbb{P}(data|\mathcal{M}) = \prod_{i=1}^{138} \mathbb{P}((x,y,z)_i|\mathcal{M}) = \prod_{i=1}^{138} p_{x_i}^1 p_{y_i}^2 p_{z_i}^3 = \prod_{x,w} (p_x^w)^{n_x^w},$$

где n_x^w — количество картинок x в окошке w.

Оказывается, максимум правдоподобия достигается при $p_x^w = \frac{n_x^w}{138}$, то есть наилучшая модель из $\mathfrak{M}$ это Модель 3 с уже рассмотренными параметрами!

Но мы уже знаем, что $\mathbb{E}R(\mathcal{M}) \neq 0.92$, поэтому наилучшую модель из $\mathfrak{M}_0$ придется искать отдельно.

PDMI COMPUTER SCIENCE CLUB

Метод максимального правдоподобия

Чтобы найти лучшую модель $\mathcal{M} \in \mathfrak{M}_0$, придется максимизировать

$$\prod_{x,w} (p_x^w)^{n_x^w} \to \max_p,$$

с ограничением $\sum_{x,y,z} reward(x,y,z) p_x^1 p_y^2 p_z^3 = 0.92$.

Не совсем понятно, как это делать аналитически, но можно написать программу, которая посчитает результат численно!

PDMI COMPUTER SCIENCE CLUB

Метод максимального правдоподобия

Вот что получилось!

Вероятности лучшей модели из $\mathfrak{M} \times 138$ (то есть исходные кратности)

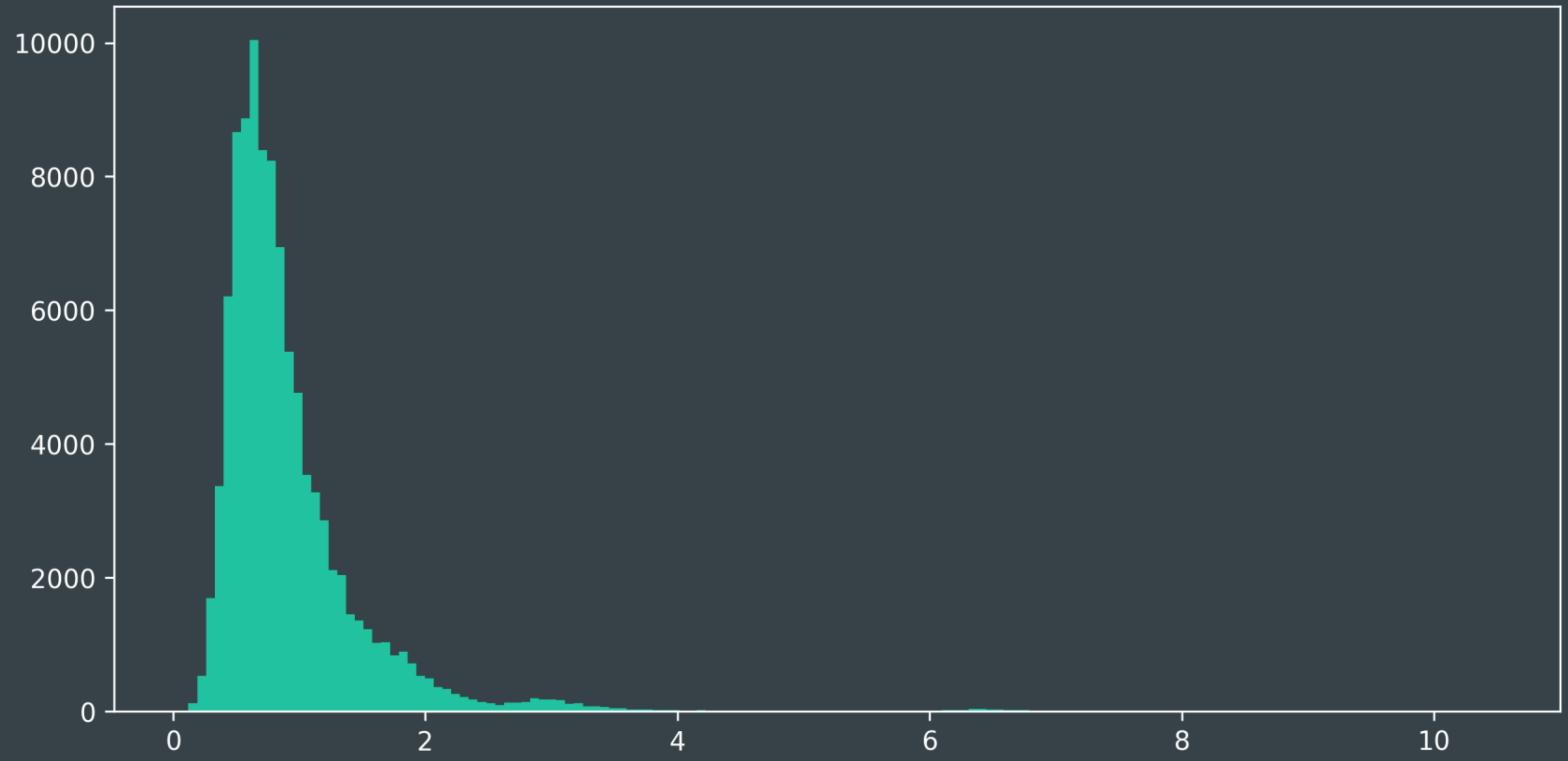
Окно	0	bar	bar x2	bar x3	7	🍒	💎 💎
1	59	49	14	6	6	1	3
2	85	8	24	16	4	0	1
3	77	39	6	1	7	3	5

Вероятности лучшей модели из $\mathfrak{M}_0 \times 138$

Окно	0	bar	bar x2	bar x3	7	🍒	💎 💎
1	56	49.7	14.6	6.2	5.9	1.1	4.4
2	80	8.7	25.5	16.9	4.1	0	1.9
3	72.9	40.1	6.4	1.1	6.8	3.3	7.4

Проверка гипотезы

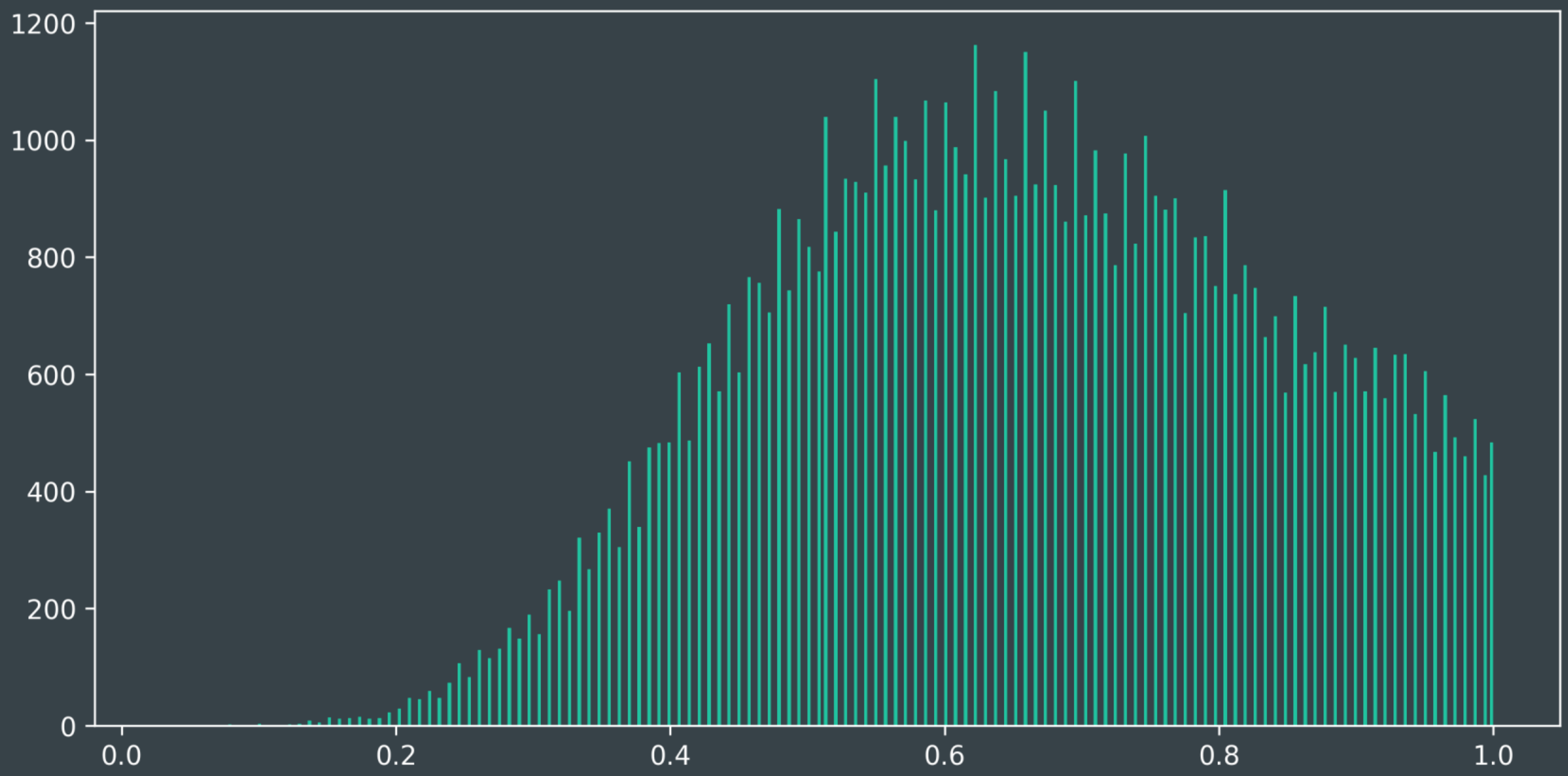
Теперь, когда мы нашли лучшую модель из $\mathfrak{M}_0$ мы можем оценить распределение среднего выборки объема 138.



PDMI COMPUTER SCIENCE CLUB

Проверка гипотезы

Рассмотрим только часть, левее 1. Видно, что средние довольно регулярно бывают меньше 0.4. Доля средних меньших 0.384 составила ≈ 0.0469.



PDMI COMPUTER SCIENCE CLUB

Проверка гипотезы

0.047 это много или мало?

Если представить толпу из 20 человек, играющих в автоматы, то примерно один из них должен получить выигрыш хуже нашего.

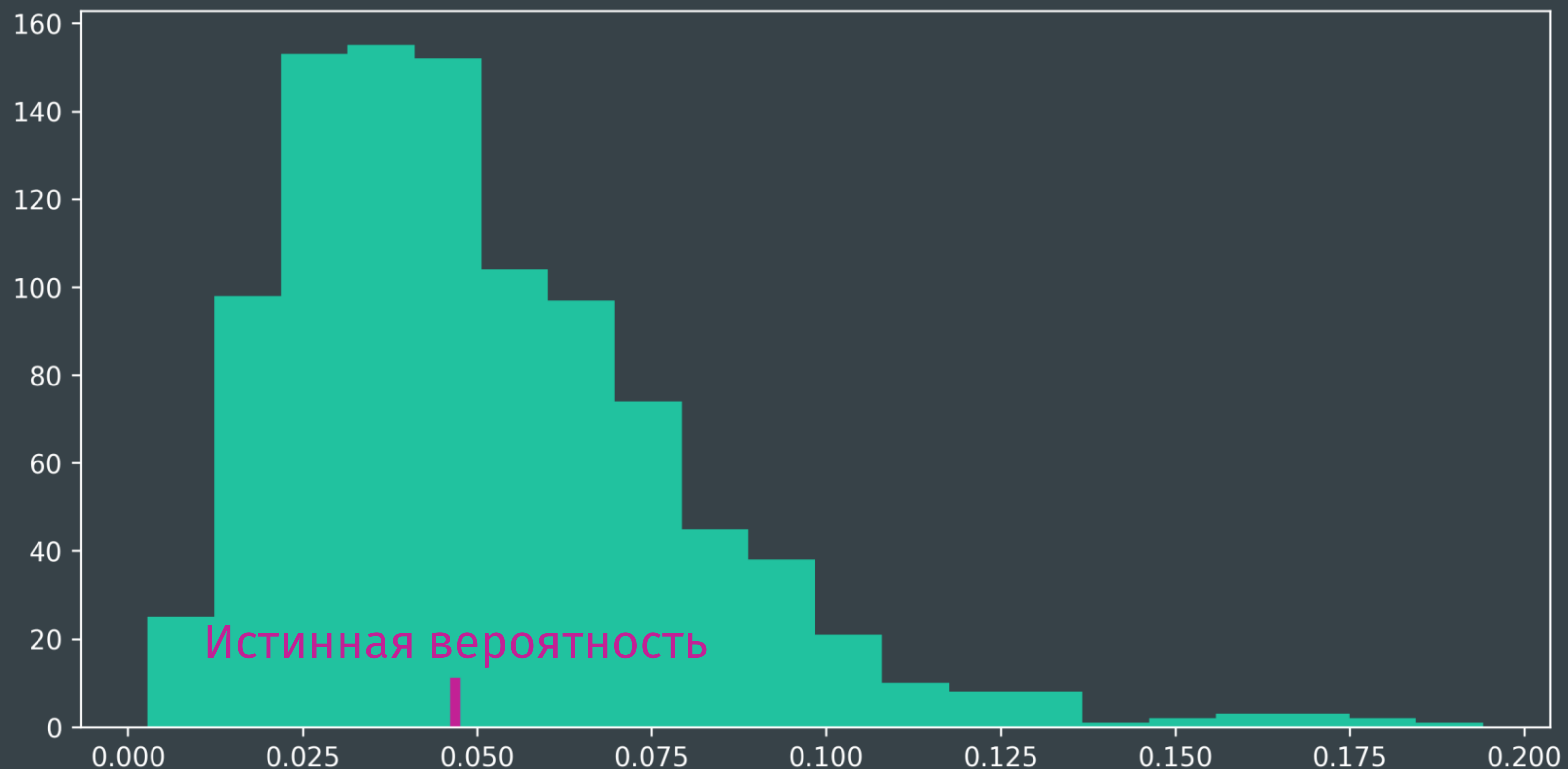
Кажется, что не очень мало. С другой стороны, мы смотрели на лучшую модель, может быть истинная модель даст сильно меньшую вероятность? Как это можно было бы проверить?

We need to go deeper...

1. Возьмем в качестве реальности найденную модель; ее истинная вероятность — 0.0469
2. Сгенерируем из модели реальности $N \cdot 138$ комбинаций
3. Для каждой выборки объема 138 найдем наилучшую модель из $\mathfrak{M}_0$ и посчитаем ее вероятность
4. Получим набор из N вероятностей

We need to go deeper...

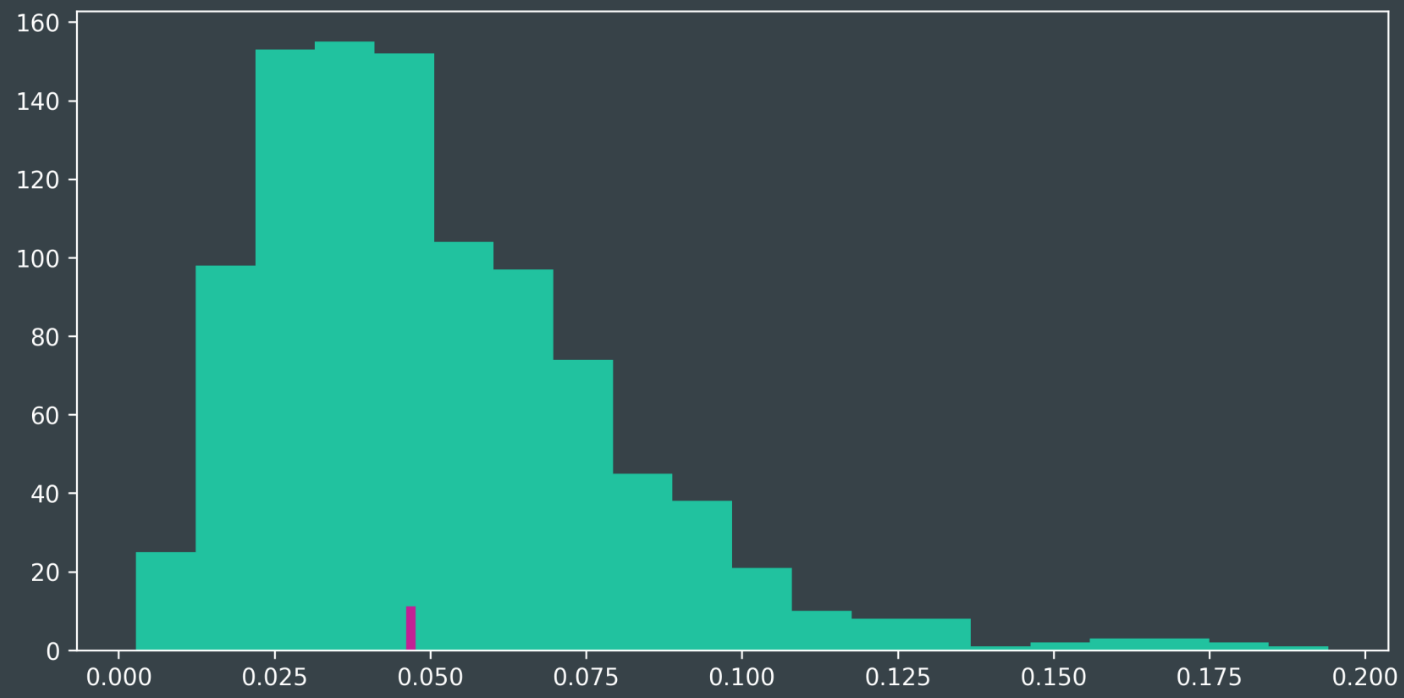
Распределение вероятностей для $N = 1000$. Каждая вероятность оценивалась с помощью 10000 выборок объема 138. Какие выводы можно сделать?



PDMI COMPUTER SCIENCE CLUB

We need to go deeper...

Отношение модельной вероятности к истинной с вероятностью 95% живет в интервале [0.26, 2.6]. Тогда обратное отношение живет в интервале [0.39, 3.78] и если наша модельная вероятность равна 0.0469, то мы можем оценить истинную вероятность с помощью интервала $[0.39, 3.78] \times 0.047 \approx [0.018, 0.18]$.



PDMI
COMPUTER
SCIENCE
CLUB

Сравнение моделей

Еще один способ проверки гипотезы связан со сравнением моделей.

В целом, сравнение моделей — это история про то, как из нескольких моделей выбрать лучшую (или по крайней мере выяснить, насколько одна модель лучше остальных).

Мы рассмотрим частный случай сравнения вложенных моделей, когда одна из моделей является подмоделью другой. В этом случае мы хотим выяснить, насколько подмодель хуже модели. Если она лишь незначительно хуже, то мы можем предпочесть ее, так как она проще, а смысл тот же.

Отношение правдоподобия

Сравним модель $\mathfrak{M}$ и подмодель $\mathfrak{M}_0 \subset \mathfrak{M}$ с помощью отношения правдоподобия (точнее, его логарифма):

$$-\log \frac{\max\limits_{\mathcal{M} \in \mathfrak{M}_0} \mathcal{L}(\mathcal{M}|data)}{\max\limits_{\mathcal{M} \in \mathfrak{M}} \mathcal{L}(\mathcal{M}|data)} = \log \max_{\mathcal{M} \in \mathfrak{M}} \mathcal{L}(\mathcal{M}|data) - \log \max_{\mathcal{M} \in \mathfrak{M}_0} \mathcal{L}(\mathcal{M}|data) \geq 0$$

Если эта величина достаточно большая, то это значит, что подмодель значительно хуже описывает данные, чем модель, и значит у нас есть основания считать, что истинное распределение плохо описывается $\mathfrak{M}_0$ в пользу альтернативы $\mathfrak{M} \setminus \mathfrak{M}_0$.

Что значит «значительно хуже»?

PDMI COMPUTER SCIENCE CLUB

Отношение правдоподобия

1. Рассмотрим лучшую модель из $\mathfrak{M}_0$ в качестве модели реальности
2. Сгенерируем $N \cdot 138$ комбинаций из модели реальности
3. Для каждой выборки объема 138 найдем максимальное правдоподобие в модели $\mathfrak{M}$ и в модели $\mathfrak{M}_0$ и вычислим минус логарифм отношения
4. Получим набор из N величин

Какова доля из этих N величин больше чем величина, построенная по исходным данным?

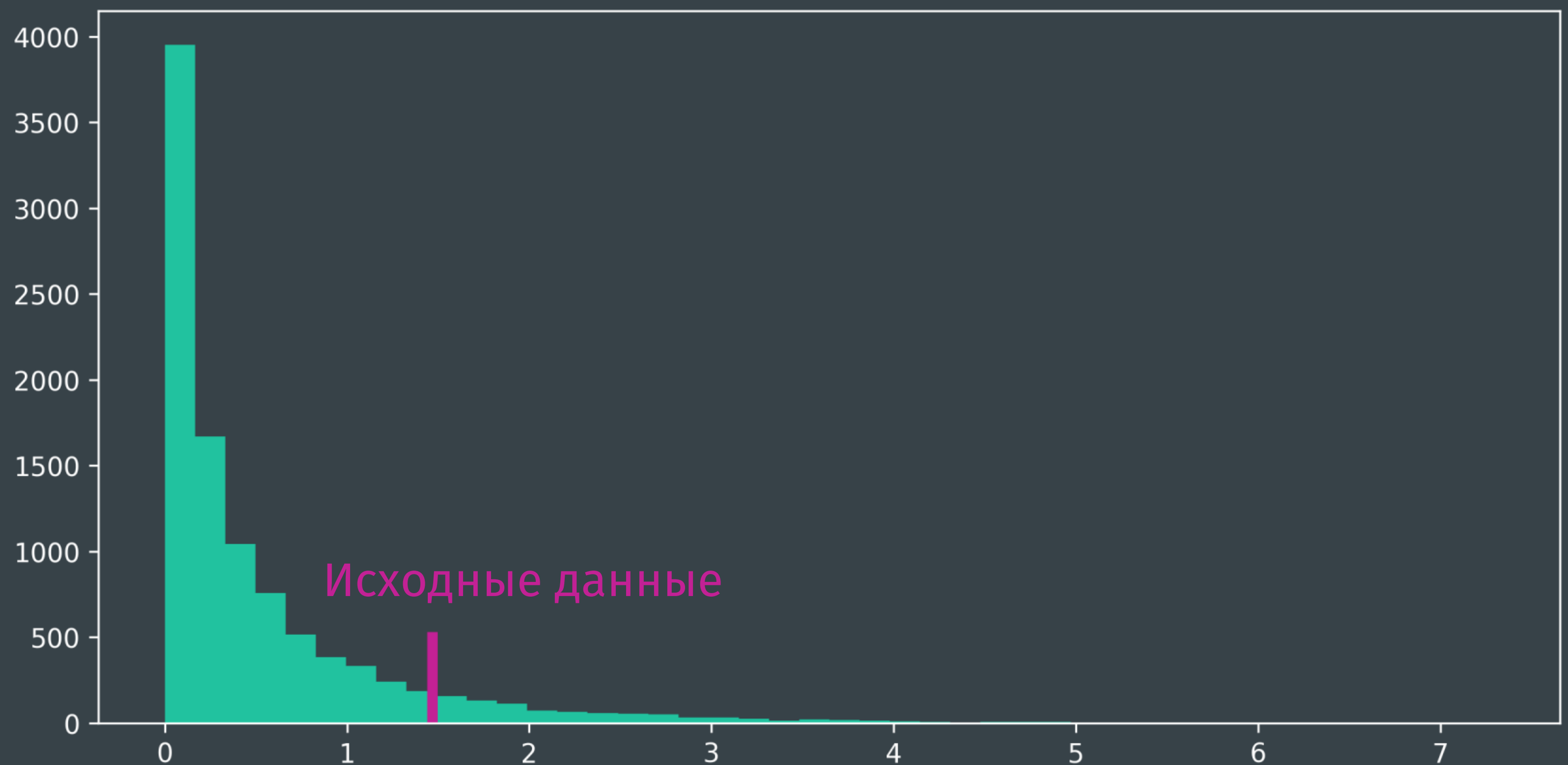
Если доля достаточно большая, то у нас нет оснований утверждать, что подмодель хуже модели, а если маленькая — есть.

PDMI COMPUTER SCIENCE CLUB

Отношение правдоподобия

Распределение минус логарифмов отношений правдоподобия для $N = 10000$.

Примерно 9.5% из них больше, чем значение на исходных данных.



PDMI COMPUTER SCIENCE CLUB

Что там с Моделью 4?

Давайте вспомним про Модель 4:

Автомат умеет возвращать любую комбинацию картинок, причем картинки в разных окошках независимы и вероятность каждой картинки одинакова для всех окошек.

Предположения:

- Окошки независимы
- Распределения окошек одинаковы

Модель описывается 7 параметрами с ограничениями: $p_x \geq 0$, $\sum_x p_x = 1$.

PDMI COMPUTER SCIENCE CLUB

Гипотеза однородности

Давайте проверим, можно ли считать, что распределения окошек одинаковы.

Гипотезы такого сорта называют гипотезами однородности — нам не важно, какое конкретно распределение у окошек, нас интересует, одинаковое оно или нет.

Как можно было решать такую задачу?

Отношение правдоподобия

Например, с помощью отношения правдоподобия.

В нашем случае большая модель $\mathfrak{M}$ это Модель 3, а подмодель $\mathfrak{M}_0$, соответствующая нулевой гипотезе, это Модель 4.

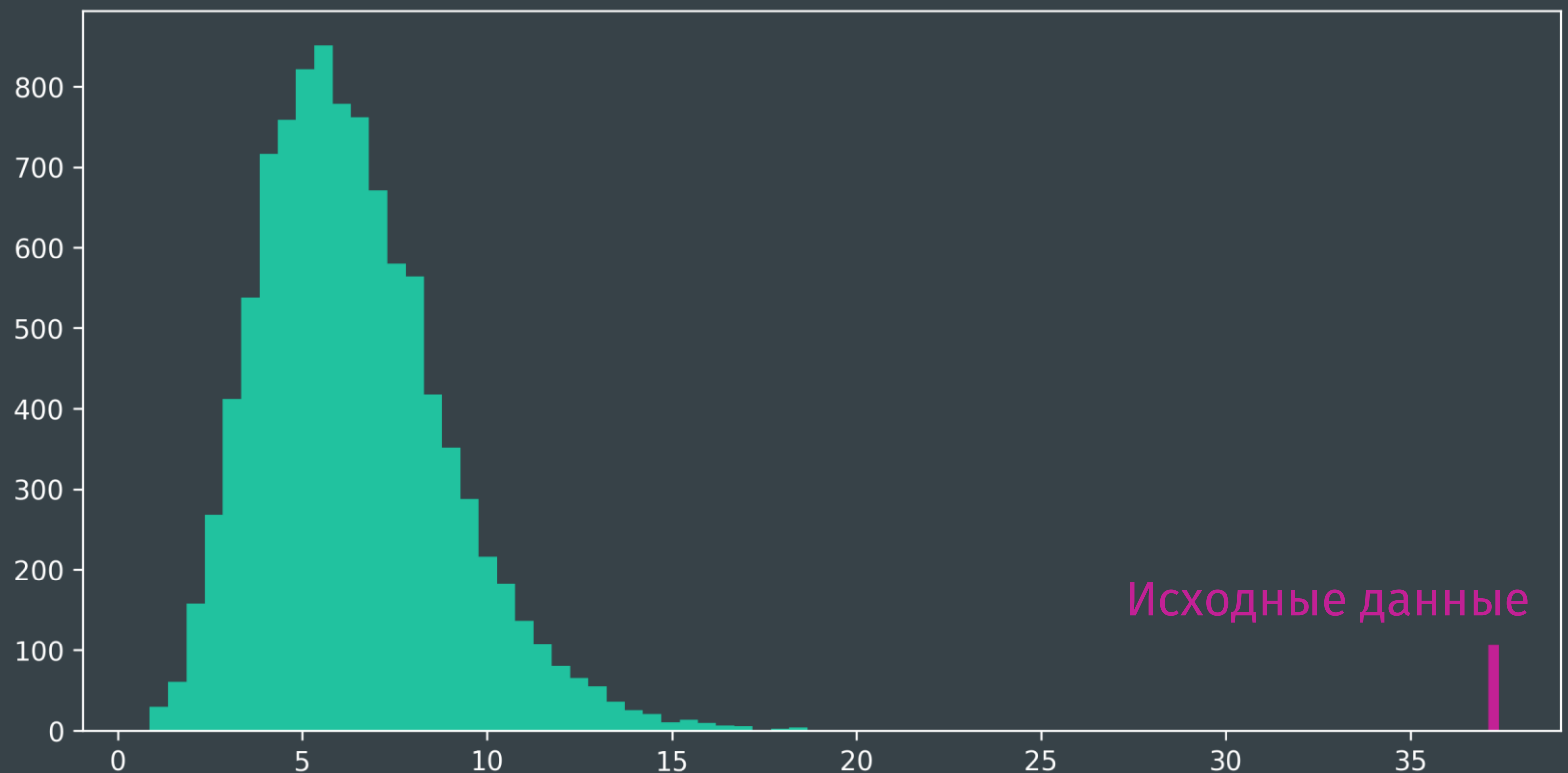
Как и раньше, рассмотрим в качестве модели реальности лучшую модель из $\mathfrak{M}_0$ и дальше посмотрим, какое распределение имеет минус логарифм.

PDMI COMPUTER SCIENCE CLUB

Отношение правдоподобия

Распределение минус логарифмов отношений правдоподобия для $N = 10000$.

Примерно так выглядит ситуация, когда вы хотите отклонить гипотезу.



PDMI COMPUTER SCIENCE CLUB

Гипотеза независимости

Раз мы так ловко отклонили гипотезу однородности, давайте проверим гипотезу независимости, на которой держится Модель 3.

Гипотеза заключается в том, что $p_{x,y,z} = p_x^1 p_y^2 p_z^3$ для всех (x, y, z).

В этот раз мы проверим гипотезу немного по-другому. Рассмотрим величину

$$\sum_{x,y,z} \frac{\left(n_{x,y,z}/138 - p_x^1 p_y^2 p_z^3\right)^2}{p_x^1 p_y^2 p_z^3},$$

где $p_x^w = n_x^w/138$. Если окошки независимы, то эта величина должна быть маленькой.

PDMI
COMPUTER
SCIENCE
CLUB

Гипотеза независимости

Распределение величин для $N = 10000$.

Примерно 14% величин больше чем у исходных данных.

PDMI COMPUTER SCIENCE CLUB

Разные вопросы

Почему мы выбирали именно такие величины? Можно ли брать другие? Как определить, какие лучше? Что значит лучше?