

Кое-что про
вероятность

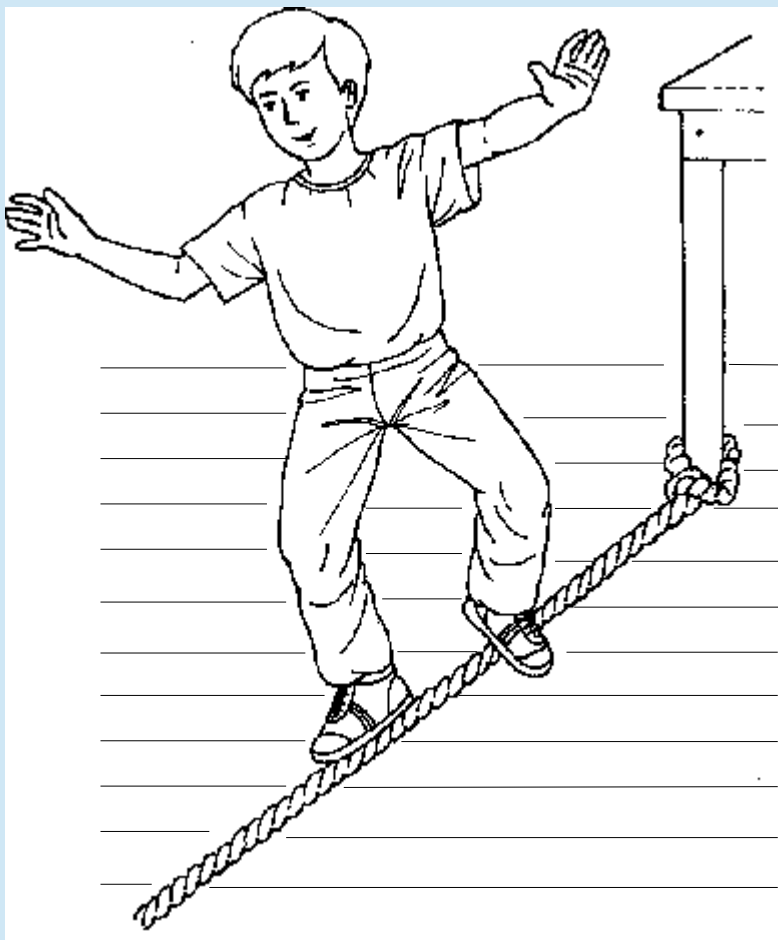
Кое-что про *вероятность*

- «Вероятность есть **здравый смысл** и немного математики» (Лаплас)
- Источники теории вероятностей ...

Кое-что про *вероятность*

- «Вероятность есть здравый смысл и немного математики» (Лаплас)
- Источники теории вероятностей:
 - Азартные игры
 - Страхование дело
 - ... ДЕНЬГИ
- Вероятность (элементарных событий) – то, насколько Вы готовы рискнуть – **СУБЪЕКТИВНА**
- Объективная вероятность – **квантовая физика.**

Априорная вероятность



Матожидание потерь ~ 0



Матожидание потерь = ОЙ-ЁЁ

Пример

- Бросили монету 3 раза и 3 раза выпал орел
- Какова вероятность выпадения орла?

$$P = \frac{N_o}{N} = \frac{3}{3} = 1$$

- Наверное — ерунда

Формула Байеса

$$Prob(\theta|D) = \frac{Prob(D|\theta) \cdot Prob(\theta)}{Prob(D)}$$

D — наблюдение (3 орла)

θ — вероятность выпадения орла (свойство монеты)

$Prob(\theta)$ — априорное распределение вероятностей орла
(свойства монеты!)

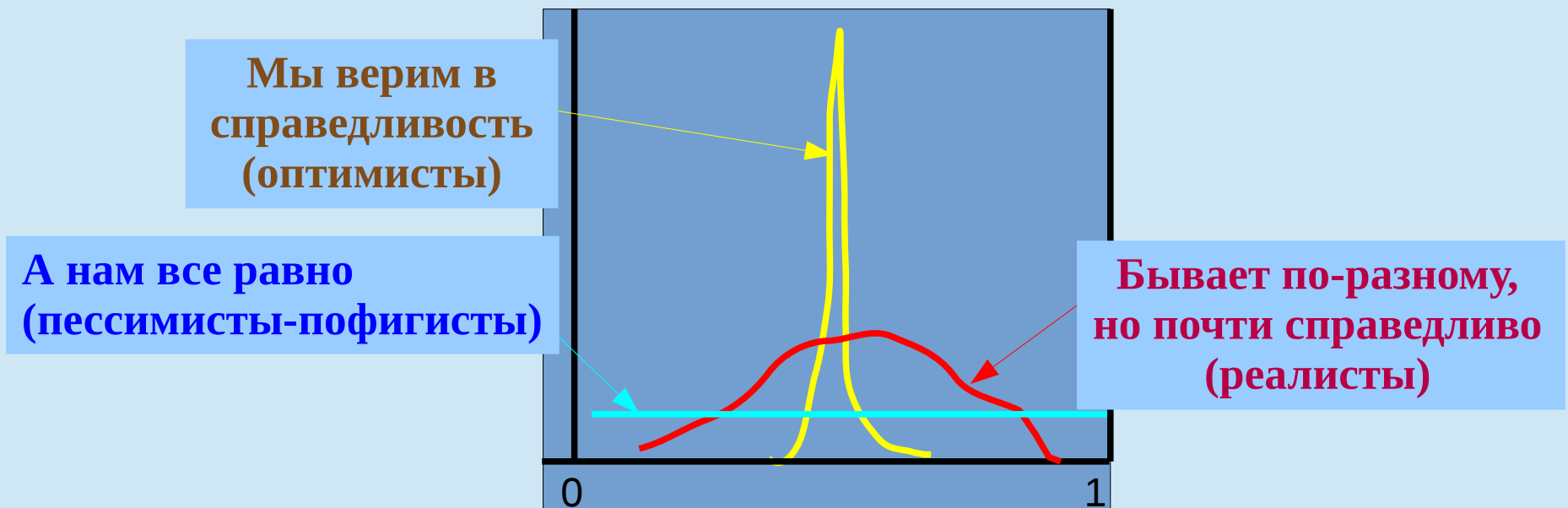
$Prob(\theta|D)$ — апостериорное распределение вероятностей
орла (свойства монеты!)

$Prob(D)$ — из условия нормировки

$$\int Prob(\theta|D) d\theta = 1 \rightarrow Prob(D) = \int Prob(D|\theta) Prob(\theta) d\theta$$

Что есть априорное распределение?

Априорное распределение – наше представление о мире



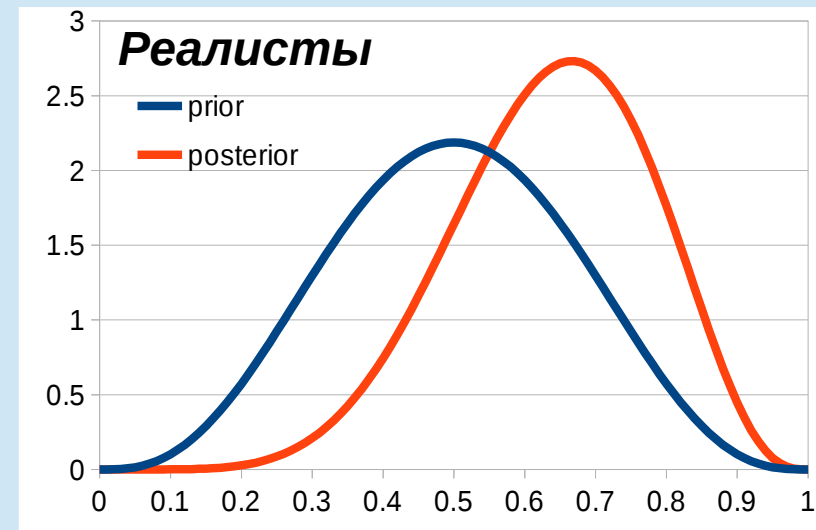
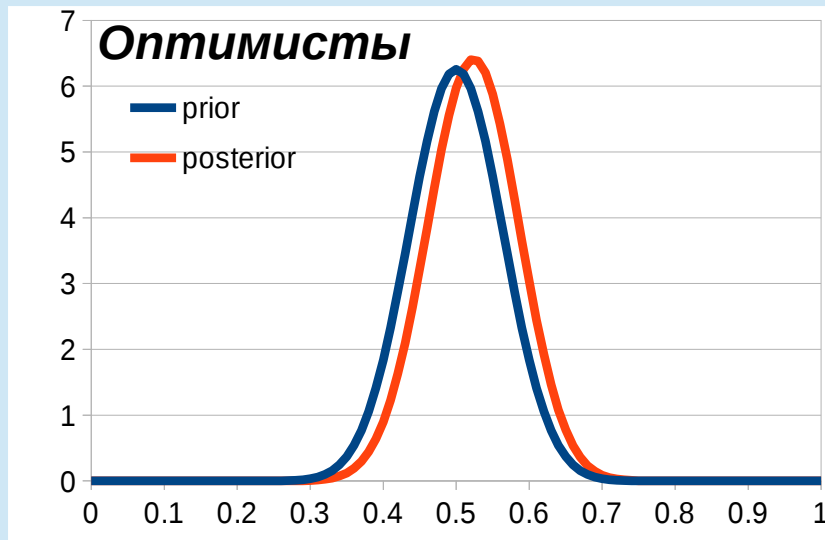
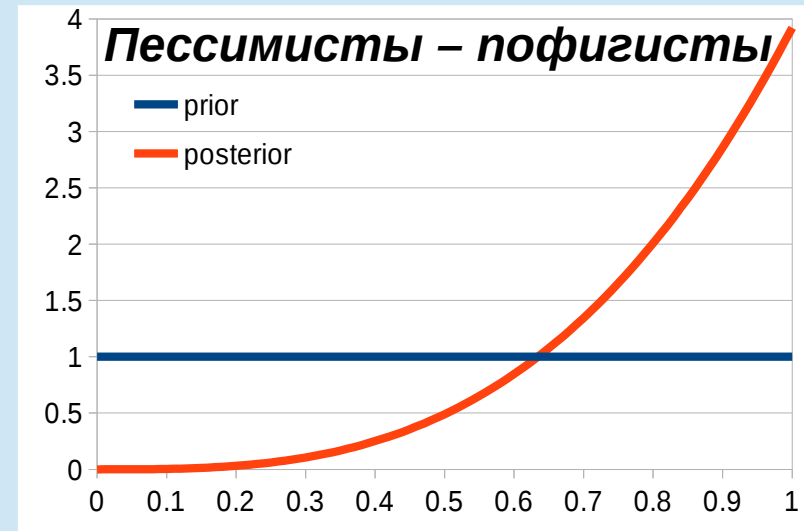
Априорное и апостериорное распределения

$$Prob(\theta|D) = \frac{Prob(D|\theta) \cdot Prob(\theta)}{Prob(D)}$$

$$Prob(D|\theta) = \theta^3 \quad \text{«А нам все равно»} \rightarrow \theta=1$$

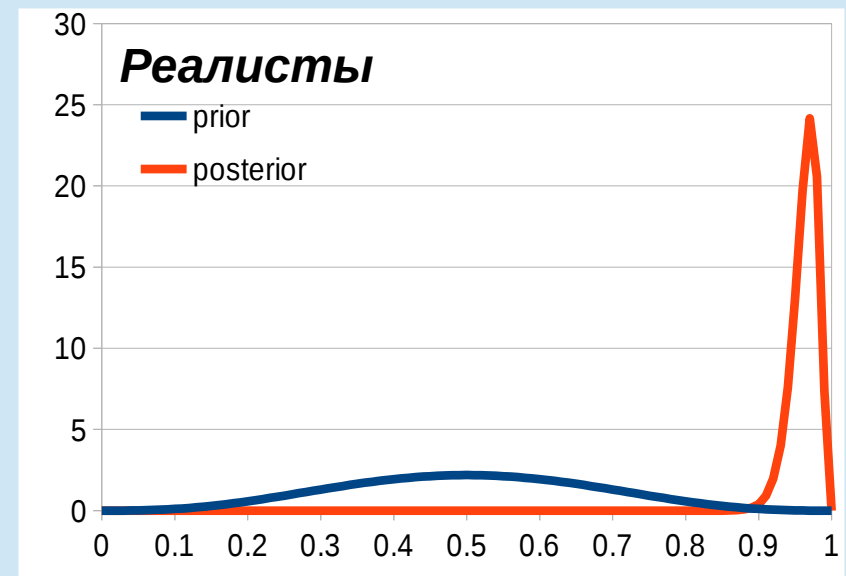
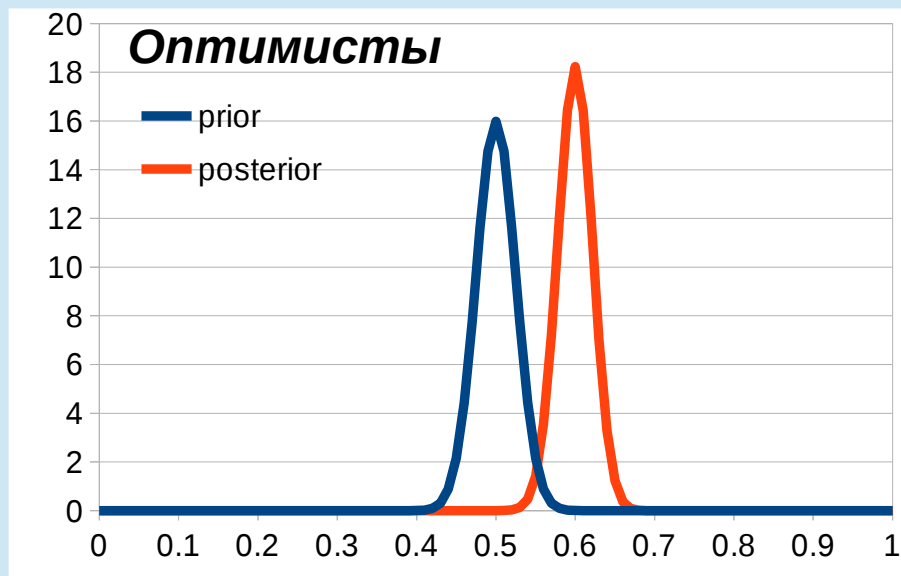
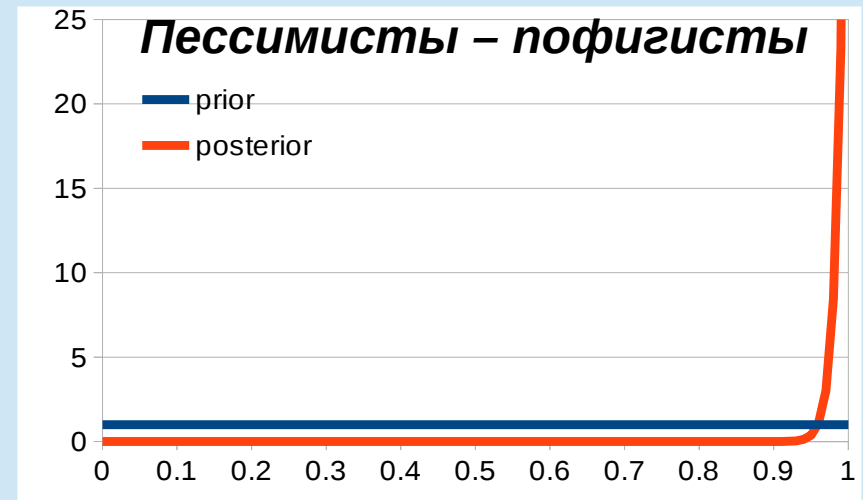
$$Prob(D) = \int Prob(D|\theta) Prob(\theta) d\theta = \int \theta^3 \cdot 1 d\theta = 1/4 \theta^4$$

$$Prob(\theta|D) = \frac{\theta^3}{1/4} = 4 \cdot \theta^3$$



Априорное и апостериорное распределения

- Если много (100) раз выпал орел



Оценки апостериорной вероятности

$$\text{Prob}(\theta|D) = \frac{\theta^3}{1/4} = 4 \cdot \theta^3$$

Это распределение, а хотелось бы число

Математическое ожидание $E(\theta)$:

$$E(\theta) = \int \text{Prob}(\theta) \cdot \theta d\theta$$

$$E(\theta) = \frac{\int \text{Prob}(D|\theta) \cdot \text{Prob}(\theta) \cdot \theta d\theta}{\int \text{Prob}(D|\theta') \cdot \text{Prob}(\theta') d\theta'}$$

В нашем случае

$$E(\theta) = \frac{\int \theta^3 \cdot 1 \cdot \theta d\theta}{\int \theta'^3 \cdot 1 d\theta'} = \frac{1/5}{1/4} = \frac{4}{5} = \frac{3+1}{3+2}$$

Априорное распределение

- Много параметров $\{\theta_1, \dots, \theta_n\}$ — например вероятности встречаемости аминокислот
- Распределение Дирихле

$$p(\theta_1, \dots, \theta_n) = Z^{-1} \prod \theta_i^{\alpha_i} \quad \sum \theta_i = 1$$

- Параметры этого распределения α_i называют ***псевдоотсчетами (pseudo counts)***.

В нашем случае:

$$E(\theta) = \frac{3 + \alpha_1}{3 + \alpha_1 + \alpha_2}$$

Если n_o орлов при N испытаниях

$$E(\theta) = \frac{n_o + \alpha_1}{N + \alpha_1 + \alpha_2}$$

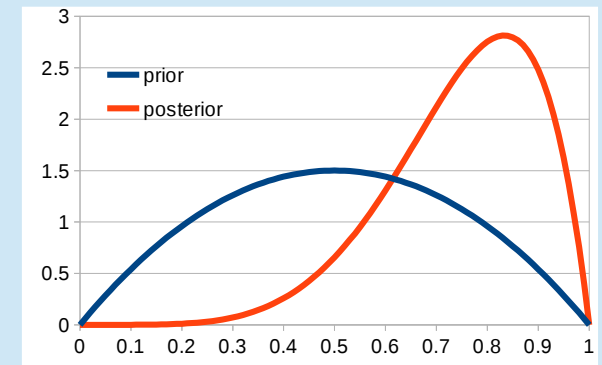
Как бы наблюдения

Априорное и апостериорное распределения

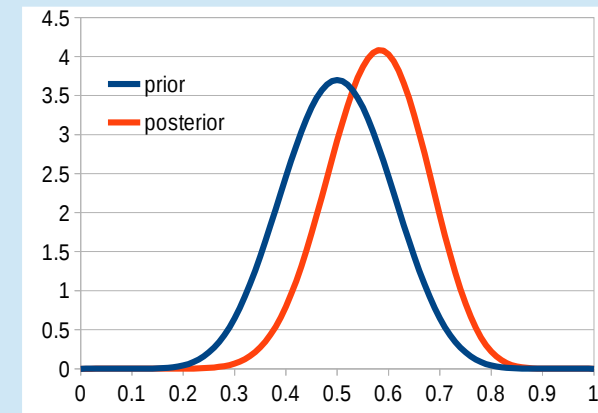
- Часто в качестве *prior* используют распределение Дирихле.
- На графиках показаны распределения для случая 4-х орлов при 4-х бросаниях монеты. θ – вероятность орла
- **Синяя линия** – распределение Дирихле $P(\theta) = \theta^{\alpha_1} (1-\theta)^{\alpha_2}$ (априорное распределение)
- **Красная линия** – апостериорная вероятность выпадения орла $P(\theta | D)$

Опыт меняет наше представление о мире

$\alpha_1=1, \alpha_2=1$



$\alpha_1=10, \alpha_2=10$



Оценка параметров по результатам

- Пусть у нас есть наблюдение D и некоторый набор параметров распределения θ , которые мы хотим оценить (см. пример про 3 орла). Кроме того, у нас есть представление о том, как эти параметры распределены (*prior*)
- Апостериорное распределение вероятностей параметров получаем из теоремы Байеса:

$$Prob(\theta|D) = \frac{Prob(D|\theta) \cdot Prob(\theta) d\theta}{\int Prob(D|\theta') Prob(\theta') d\theta'}$$