Южный федеральный университет

Пример байесовского вывода

Сергей Николенко
6 ноября 2025 г.

Технологии и фронтиры науки о данных

@SINECOR

Теорема Байеса в машинном обучении

Теорема Байеса в ML

- Давайте обсудим компоненты формулы Байеса в машинном обучении:
- Параметры θ, данные D:

$$p(\theta|D) = \frac{p(\theta)p(D|\theta)}{p(D)}$$

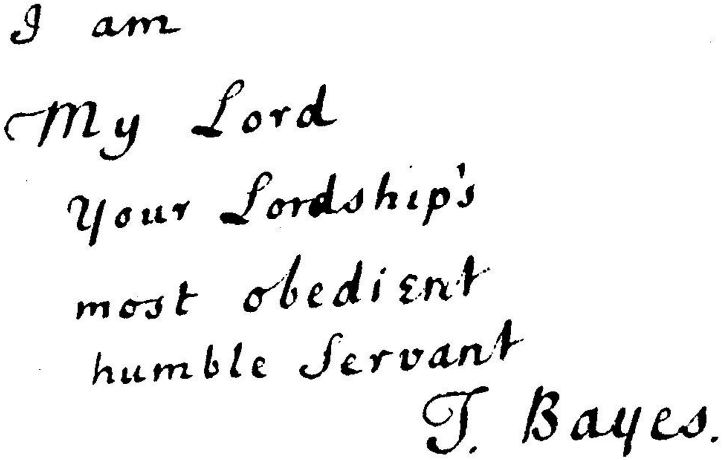
I am
My Lord
Your Lordship's
most obedient
humble Servant
T. Bayes.

Теорема Байеса в ML

- Давайте обсудим компоненты формулы Байеса в машинном обучении:
- Параметры θ, данные D:

$$p(\theta|D) = \frac{p(\theta)p(D|\theta)}{p(D)}$$

- **Правдоподобие** $p(D|\theta)$

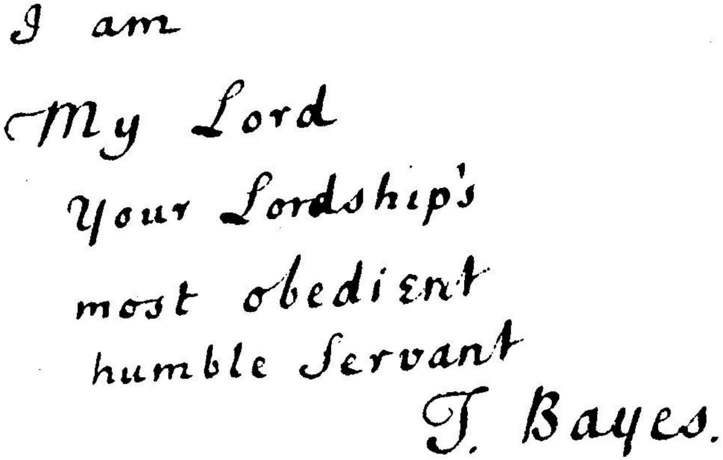


Теорема Байеса в ML

- Давайте обсудим компоненты формулы Байеса в машинном обучении:
- Параметры θ, данные D:

$$p(\theta|D) = \frac{p(\theta)p(D|\theta)}{p(D)}$$

- **Правдоподобие** p(D|θ)
- **Априорное распределение** p(θ)



Теорема Байеса в ML

- Давайте обсудим компоненты формулы Байеса в машинном обучении:
- Параметры θ, данные D:

$$p(\theta|D) = \frac{p(\theta)p(D|\theta)}{p(D)}$$

- **Правдоподобие** $p(D|\theta)$
- **Априорное распределение** $p(\theta)$
- **Апостериорное распределение** $p(\theta|D)$

Задачи байесовского вывода

$$p(\theta|D) = \frac{p(\theta)p(D|\theta)}{p(D)}$$

- Отсюда следуют основные задачи байесовского вывода
- Первая задача – найти *гипотезу максимального правдоподобия* (maximum likelihood, ML):

$$\theta_{ML} = \underset{\theta}{\operatorname{argmax}}\, p(D|\theta)$$

Задачи байесовского вывода

$$p(\theta|D) = \frac{p(\theta)p(D|\theta)}{p(D)}$$

- Отсюда следуют основные задачи байесовского вывода
- Первая задача – найти *гипотезу максимального правдоподобия* (maximum likelihood, ML):

$$\theta_{ML} = \operatorname*{argmax}_{\theta} p(D|\theta)$$

- Вторая – найти апостериорное распределение (posterior)

$$p(\theta|D) \propto p(\theta)p(D|\theta)$$

и, если будет нужно, максимальную апостериорную гипотезу (maximum a posteriori, MAP)

$$\theta_{MAP} = \arg\max_{\theta} p(\theta|D) = \arg\max_{\theta} p(\theta)p(D|\theta)$$

Задачи байесовского вывода

$$p(\theta|D) = \frac{p(\theta)p(D|\theta)}{p(D)}$$

- Отсюда следуют основные задачи байесовского вывода
- Первая задача – найти *гипотезу максимального правдоподобия* (maximum likelihood, ML):

$$\theta_{ML} = \operatorname{argmax}_{\theta} p(D|\theta)$$

- Вторая – найти апостериорное распределение (posterior)

$$p(\theta|D) \propto p(\theta)p(D|\theta)$$

и, если будет нужно, максимальную апостериорную гипотезу (maximum a posteriori, MAP)

$$\theta_{MAP} = \arg\max_{\theta} p(\theta|D) = \arg\max_{\theta} p(\theta)p(D|\theta)$$

- Самая сложная (и интересная) третья задача – найти *предсказательное распределение* (predictive distribution)

$$p(x|D) \propto \int_{\theta\in\Theta} p(x|\theta)p(\theta|D)\mathrm{d}\theta$$

Задачи байесовского вывода

$$p(\theta|D) = \frac{p(\theta)p(D|\theta)}{p(D)}$$

- Отсюда следуют основные задачи байесовского вы
- Первая задача – найти *гипотезу максимальн*
(maximum likelihood, ML):

$$\theta_{ML} = \arg$$

- Вторая – найти апост rior)

)

и, есл апостериорную гипотезу
(m

$$p(\theta|D) = \arg\max_{\theta} p(\theta)p(D|\theta)$$

- Сама (и интересная) третья задача – найти *предска тельное распределение* (predictive distribution)

$$p(x|D) \propto \int_{\theta\in\Theta} p(x|\theta)p(\theta|D)\mathrm{d}\theta$$

Давайте обсудим их на конкретном примере.

Байесовский вывод для монетки

Задачи байесовского вывода

- Рассмотрим монетку, которая подброшена N раз, и результаты этих бросков известны, т.е. данные D – это последовательность исходов: орёл, решка, решка, орёл, орёл…
- Что такое здесь параметры θ? Сколько их?

$$p(\theta|D) = \frac{p(\theta)p(D|\theta)}{p(D)}$$

Задачи байесовского вывода

- Задача машинного обучения (байесовского вывода) здесь в том, чтобы определить «нечестность» монетки, то есть её единственный параметр θ (вероятность выпадения орла), и предсказать, с какой вероятностью она выпадет орлом в следующий раз.
- Начнём с *правдоподобия*: какое будет правдоподобие $p(D|\theta)$ для датасета D из n орлов и m решек?

$$p(\theta|D) = \frac{p(\theta)p(D|\theta)}{p(D)}$$

Задачи байесовского вывода

- Задача машинного обучения (байесовского вывода) здесь в том, чтобы определить «нечестность» монетки, то есть её единственный параметр θ (вероятность выпадения орла), и предсказать, с какой вероятностью она выпадет орлом в следующий раз.
- Начнём с *правдоподобия*: какое будет правдоподобие p(D|θ) для датасета D из n орлов и m решек?

$$p(D|\theta) = \theta^n (1-\theta)^m$$

- Где оно максимизируется?

$$p(\theta|D) = \frac{p(\theta)p(D|\theta)}{p(D)}$$

Задачи байесовского вывода

- Мы вроде бы получили решение: $\theta_{ML} = \frac{n}{n+m}$
- Но, во-первых, у правдоподобия для m = 0 нет максимума, оно идёт к бесконечности.
- Во-вторых, даже если предположить, что мы как-то ввели ограничение, утверждающее, что θ должно быть вероятностью, получается, что после первого же броска мы предсказываем, что теперь монетка всегда будет выпадать той же стороной…
- Звучит странно… а что не так? Чего не хватает?

Задачи байесовского вывода

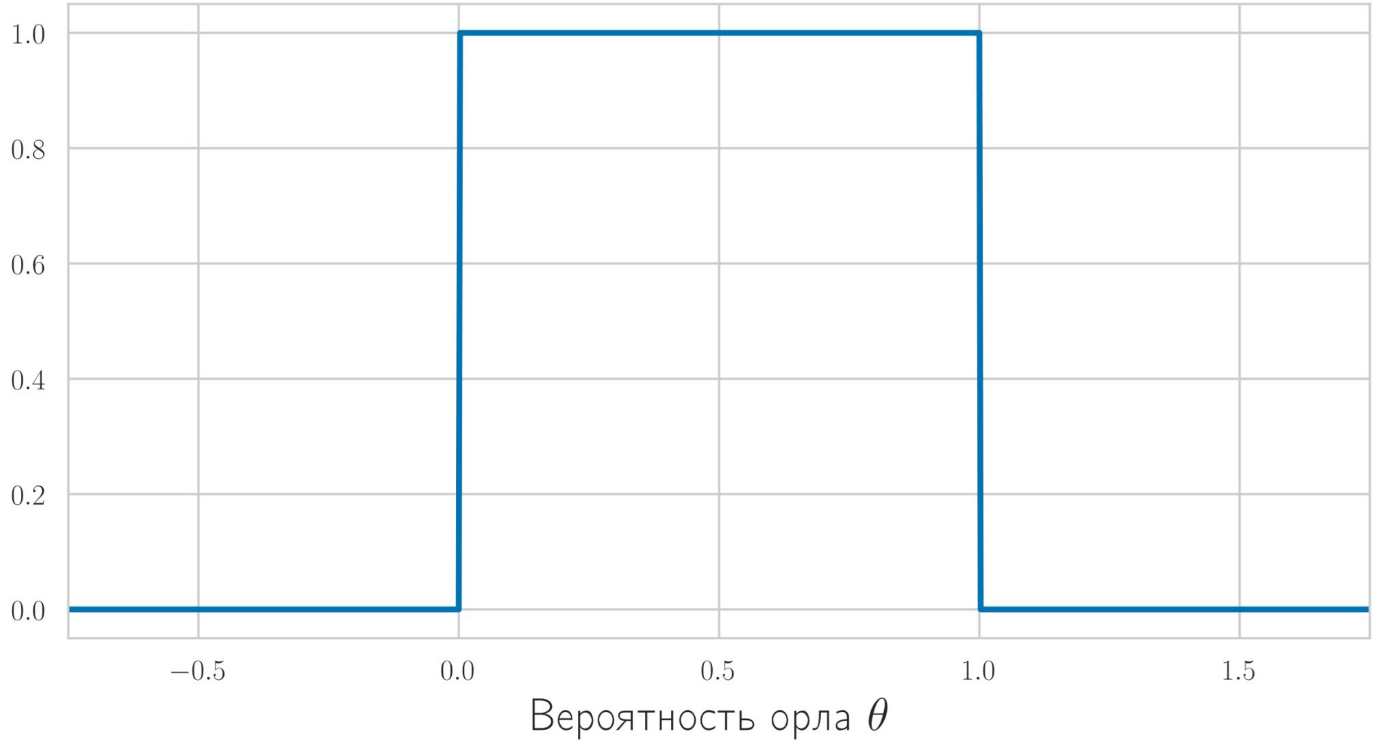
- Мы вроде бы получили решение: $\theta_{ML} = \frac{n}{n+m}$
- Но, во-первых, у правдоподобия для m = 0 нет максимума, оно идёт к бесконечности.
- Во-вторых, даже если предположить, что мы как-то ввели ограничение, утверждающее, что θ должно быть вероятностью, получается, что после первого же броска мы предсказываем, что теперь монетка всегда будет выпадать той же стороной...
- Звучит странно... а что не так? Чего не хватает?
- Не хватает нашей интуиции о том, какие на свете бывают монетки. Что выражает эту интуицию в байесовском выводе?

Задачи байесовского вывода

- Мы вроде бы получили решение: $\theta_{ML} = \frac{n}{n+m}$
- Но, во-первых, у правдоподобия для m = 0 нет максимума, оно идёт к бесконечности.
- Во-вторых, даже если предположить, что мы как-то ввели ограничение, утверждающее, что θ должно быть вероятностью, получается, что после первого же броска мы предсказываем, что теперь монетка всегда будет выпадать той же стороной...
- Звучит странно... а что не так? Чего не хватает?
- Не хватает нашей интуиции о том, какие на свете бывают монетки. Что выражает эту интуицию в байесовском выводе?
- Априорное распределение p(θ)! Давайте его зададим.

Задачи байесовского вывода

- Начнём с равномерного априорного распределения:



- Это не совсем моя интуиция о монетках, но начнём отсюда...

Задачи байесовского вывода

- Теперь мы хотим найти

$$\theta_{MAP} = \arg\max_{\theta} p(\theta|D) = \arg\max_{\theta} p(\theta)p(D|\theta)$$

- Но с равномерным априорным распределением это та же самая функция, оно добавляет только то самое ограничение θ от 0 до 1, а внутри этого правдоподобие не меняется
- То есть мы ничего не достигли? Нет, это только начало...

Задачи байесовского вывода

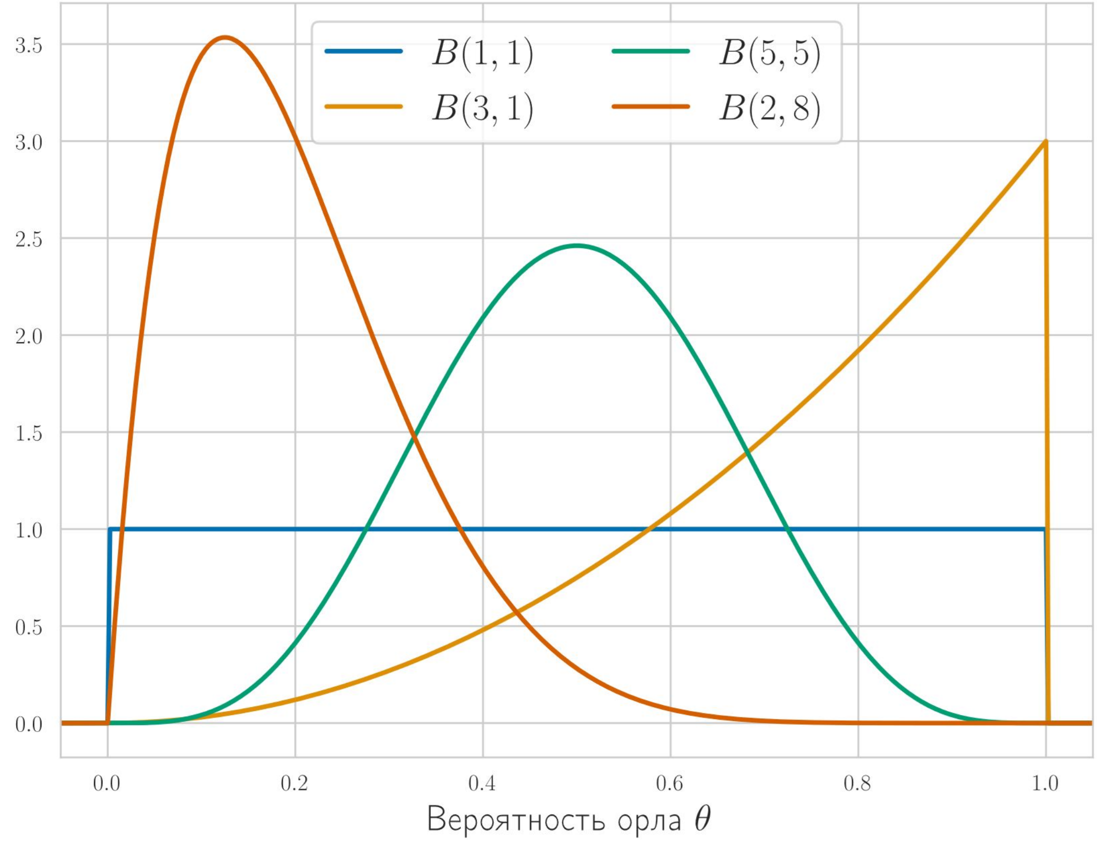
- Нормировочная константа для апостериорного распределения:

$$B(\alpha,\beta) = \int_0^1 x^{\alpha-1}(1-x)^{\beta-1}\mathrm{d}x \qquad B(\alpha,\beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$$

- *Бета-распределение* на [0, 1]:

$$\mathrm{Beta}(\theta|\alpha,\beta) = \frac{1}{B(\alpha,\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1}$$

- Давайте подсчитаем теперь с константой; от этого, правда, максимум всё равно не изменится, так зачем же всё это было нужно?



Задачи байесовского вывода

- Последняя, самая сложная задача байесовского вывода — найти *предсказательное распределение*:

$$p(\text{орёл}|D) = \int p(\text{орёл}|\theta)p(\theta|D)\mathrm{d}\theta$$

- Подсчитав его в данном случае, получим *правило Лапласа*:

$$p(\text{орёл}|D) = \frac{n+1}{n+m+2}$$

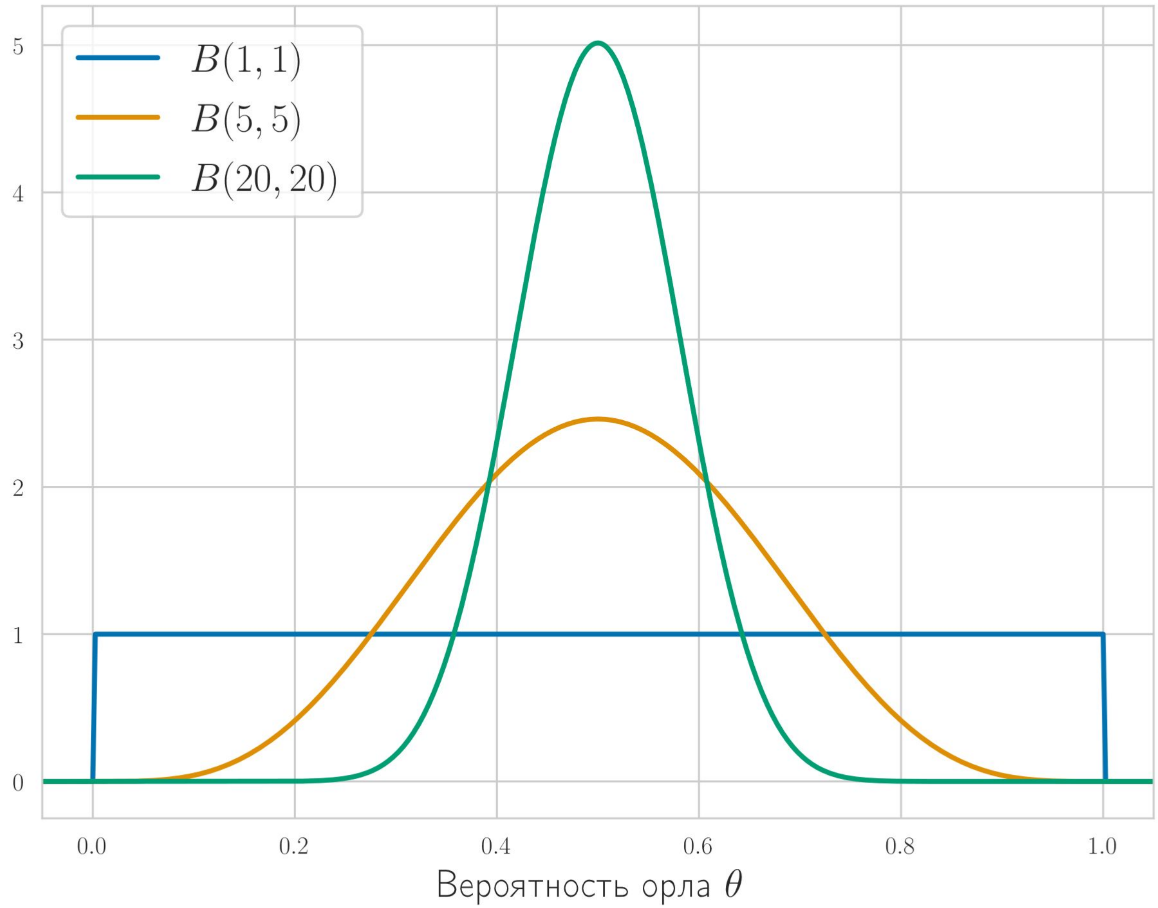
- Давайте это обсудим...

Сопряжённые априорные распределения

Бета-распределение как сопряжённое

- Мы уже выяснили, что *бета-распределение* похоже на правдоподобие монетки
- Более того, кажется, что для больших одинаковых параметров бета-распределение выражает нужную интуицию про настоящие монетки…

$$\mathrm{Beta}(\theta|\alpha,\beta) = \frac{1}{B(\alpha,\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1}$$



Бета-распределение как сопряжённое

- Давайте использовать бета-распределение как априорное:

  $p(\theta) = \mathrm{Beta}(\theta|\alpha, \beta)$

- Тогда

$p(\theta|D) \propto p(\theta)p(D|\theta) =$

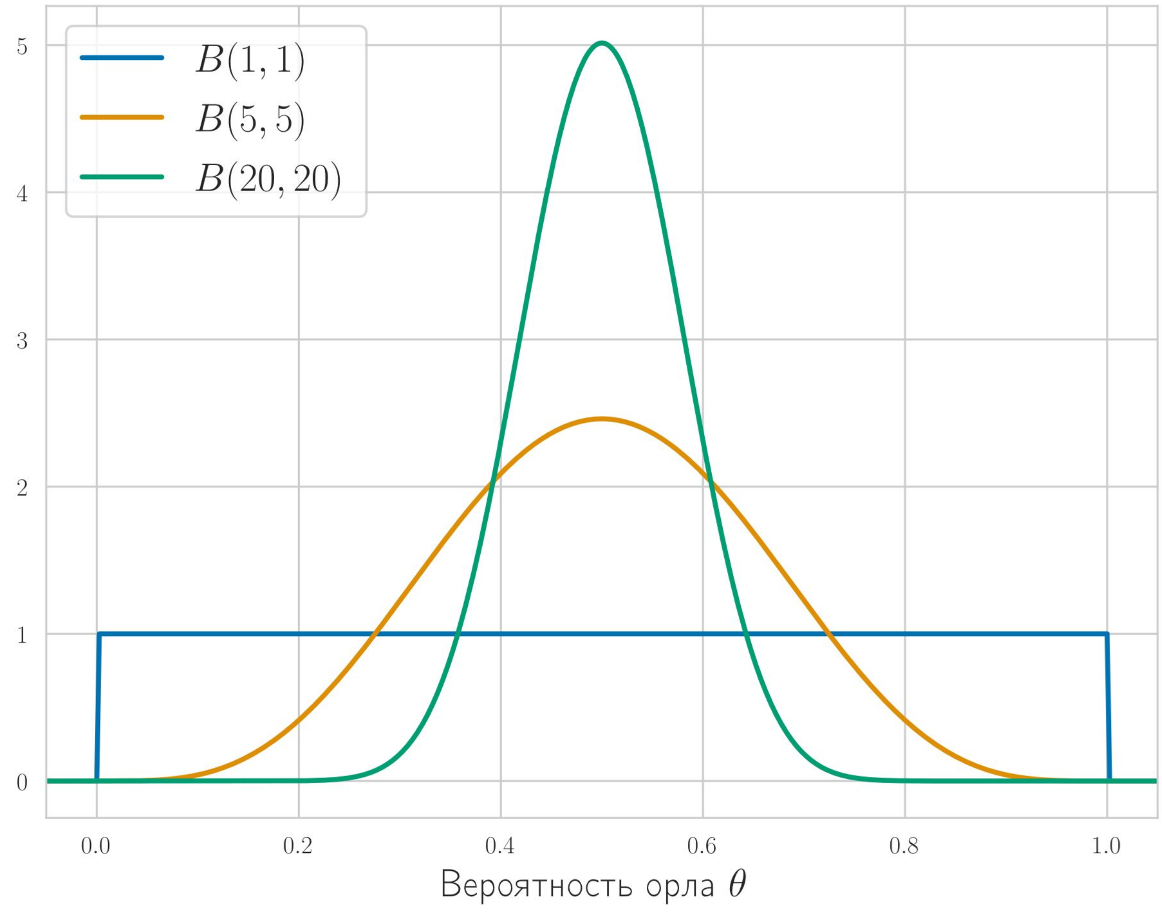
$= \mathrm{Beta}(\theta|\alpha, \beta)\theta^n(1-\theta)^m \propto$

$\propto \theta^{n+\alpha-1}(1-\theta)^{m+\beta-1},$

  и снова получается бета-распределение, теперь как апостериорное!

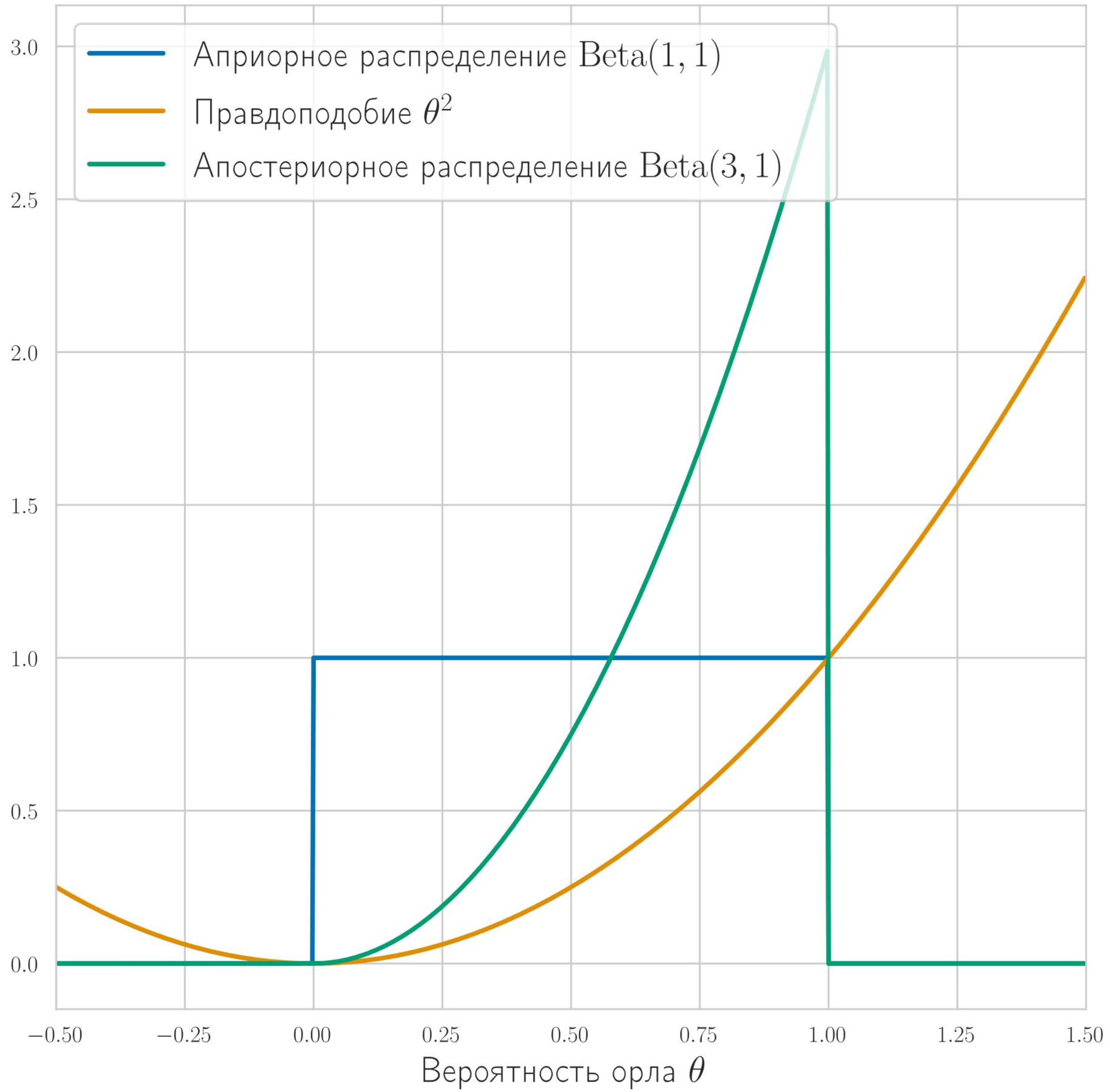
- Это пример *сопряжённых* априорных распределений (conjugate priors)

$$\mathrm{Beta}(\theta|\alpha, \beta) = \frac{1}{B(\alpha, \beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1}$$



Пример байесовского вывода

- Вот пример байесовского вывода как умножения одного распределения на другое
- Давайте посмотрим на примеры в коде и заодно обсудим один любопытный частный случай...



Южный федеральный университет

@SINECOR

Спасибо за внимание!