Байесовский подход

История

- Английский священник
- Родился то ли в 1701, то ли в 1702 году
- О жизни известно довольно мало
- Даже неясно, как он на самом деле выглядел
- Портрет Байеса, которым часто иллюстрируют книги, на самом деле может принадлежать другому человеку
- Написал две работы: одну по богословию, одну по математике

Томас Байес
(это неточно)

История

- Математическая работа Байеса была опубликована только после его смерти
- Чуть позже один из отцов-основателей теории вероятностей, Лаплас, сформулировал теорему Байеса в том виде, в каком мы её сегодня знаем

Пьер-Симон Лаплас
(это точно)

ru.wikipedia.org

Что происходит когда встаёт солнце

- Человек впервые попал в наш мир и впервые видит, как встаёт солнце
- Бедняга ничего не знает про солнце и не понимает, насколько восход солнца – обычное явление

Кадр из фильма «Звездные войны эпизод 4: новая надежда», реж. Джордж Лукас, Lucasfilm, 20th Century Fox

Что происходит когда встаёт солнце

- Поначалу он ничего не знает о солнце и думает, что вероятность его восхода p может быть какой угодно
- Каждый день новоиспечённый житель Земли видит, как восходит солнце, его вера в это укрепляется по формуле:

$$\frac{k+1}{k+2}$$

- Вероятность становится всё ближе к единице, но полной уверенности, что солнце завтра взойдёт, не будет никогда

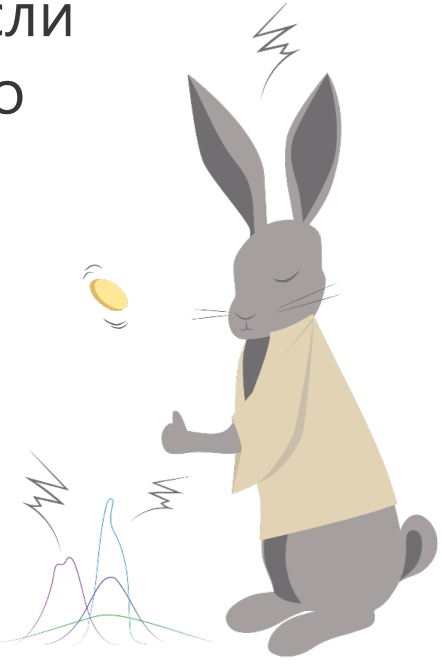
➤ *https://en.wikipedia.org/wiki/Sunrise_problem*

Снова про наше незнание

- Формулу для нашей уверенности в восходе солнца можно получить по теореме Байеса
- Теорема Байеса – простое правило обновления уровня доверия к гипотезе, при получении новых доказательств
- На самом деле мы используем этот принцип ежедневно, он очень естественен

Байесовский взгляд на вероятность

- **Лаплас:** детерминизм, мы могли бы идеально прогнозировать вселенную, если бы измерили точное положение каждого атома, но издержки этого огромны
- Между совершенством природы и несовершенством человеческого познания огромный разрыв
- **Неопределённость –** результат этого разрыва
- Случайность возникает из-за нашего незнания, а **вероятность – способ его измерить**
- **Вероятность – субъективна**



Объективная случайность

Кадр из фильма «Прибытие», реж. Дени Вильнев, FilmNation Entertainment, Lava Bear Films, 21 Laps Entertainment, Paramount Pictures

История

- В науке нет места субъективности
- Можно оценивать вероятность только тех событий, которые происходят более одного раза
- Фишер впервые употребил термин “байесовская статистика” в своих работах
- Разработка метода максимального правдоподобия, процедур проверки гипотез и всего того статистического арсенала, который мы изучали

Рональд Фишер
(это почти наверное)

Почему байесовская статистика умерла

- Вычислительные сложности
- Даже для простых моделей возникают сложные интегралы и распределения
- Не для каждого распределения можно вывести формулы
- Критика “субъективного” подхода: можно взять разные априорные распределения и получить противоположные результаты

Почему байесовская статистика воскресла

- Развитие компьютеров
- MCMC и другие алгоритмы позволили работать с любыми априорными распределениями
- Оказалось, что многие частотные алгоритмы – частный случай байесовских
- Байесовские идеи оказались полезны в разных областях современного машинного обучения

План

- Вспомним формулу Байеса для дискретного и непрерывного случаев
- Осознаем, что Байесовские методы – лишь новый способ взглянуть на старые модели
- Обсудим сопряжённые распределения и алгоритм Монте-Карло по схеме Марковской цепи (MCMC)
- Посмотрим как байесовский вывод связан с методом максимального правдоподобия
- Оценим несколько простых моделей

Две статистики

Как мы строили модели раньше

1. Мы предполагали, что у нас есть какой-то параметр θ, который надо оценить (константа)
2. У нас есть данные $y_1, \dots, y_n$, связанные с этим параметром
3. У нас есть модель, описывающая связь $y_1, \dots, y_n$ с θ
4. У нас есть методы для оценки $\hat{\theta}(y_1, \dots, y_n)$
5. У нас есть набор теорем, который говорит, как распределена величина $\hat{\theta}$
6. На основе этих теорем мы можем проверять гипотезы и строить доверительные интервалы

Схема математической статистики

Выборка: $x_1, \ldots, x_n$ Параметр: θ

$\hat{\theta} \longrightarrow f_{\hat{\theta}}(t)$

Точность оценки, прогнозов

доверительные интервалы

Ответы на вопросы

проверка гипотез

Как оценить

- Метод моментов
- Метод максимального правдоподобия

Хорошие свойства

- Несмещенная
- Состоятельная
- Эффективная

Союзники

Асимптотические (при большом n)

- ЦПТ
- Дельта-метод

Точные

- Теорема Фишера
- $\chi_n^2, t_n, F_{n,k}$
- Ещё союзники!

Подмена понятий

- Нам хотелось получить ответ на вопрос:

$$\mathbb{P}(\theta > 3 \mid y) - ?$$

- Мы не могли ответить на этот вопрос, так как θ – это неизвестная константа

- Для того, чтобы выкрутиться из такой ситуации, мы формулировали гипотезу $\mathrm{H}_0\colon \theta = 3$ и альтернативу $\mathrm{H}_0\colon \theta > 3$

- Статистический тест помогал нам понять, значимо ли отличие θ от 3, но при этом он ничего не говорил о том, насколько это различие вероятно

Подмена понятий

- Мы могли найти $\widehat{\mathbb{P}}(\hat{\theta} > 3 \mid y)$, но это оценка для вероятности оценки, а нам хочется работать с истинным значением θ

- Кроме того, мы проверяли конкретную гипотезу, а все остальные существующие гипотезы мы отбрасывали

- Байесовский подход позволяет получить именно $\mathbb{P}(\theta > 3 \mid y)$, а также вычисляет апостериорную вероятность каждой из множества гипотез

Как мы будем строить модели теперь

1. Не будем отделять друг от друга θ и $\hat{\theta}$, будем смотреть на θ как на случайную величину, а не константу

2. Своё знание о параметре θ будем формализовывать в априорном распределении $f(\theta)$

3. После будем собирать выборку и по формуле Байеса получать апостериорное распределение

$$f(\theta \mid y) = \frac{f(y \mid \theta) \cdot f(\theta)}{f(y)}$$

4. Используя апостериорное распределение, мы сможем искать ответы на наши вопросы

Байесовский подход

- Байесовский подход – это другой подход к оцениванию параметров, а не другая модель
- В его рамках можно оценивать любые классические модели

Частотный подход и Байесовский подход

- θ – неизвестен, константа
- Данные: $y_1, \dots, y_n$
- Модель: описывает связь $y_1, \dots, y_n$ и θ
- Оценивание: ММП, ММ
- Теоремы: ЦПТ, дельта-метод…
- Прогнозы и гипотезы

- θ – неизвестен, случаен, наше незнание описывает $f(\theta)$
- Данные: $y_1, \dots, y_n$
- Модель: описывает связь $y_1, \dots, y_n$ и θ
- Байесовский вывод: $f(\theta \mid y)$
- Выводы делаем по распределению
- Можно получить распределение для $\hat{y}_{n+1}$

Петя – повелитель карасей

Формула Байеса

$$f(\theta \mid y) = \frac{f(y \mid \theta) \cdot f(\theta)}{f(y)}$$

- $f(\theta \mid y)$ – апостериорное распределение
- $f(\theta)$ – априорное распределение
- $f(y \mid \theta)$ – правдоподобие
- $f(y)$ – ”полная вероятность” выборки

Формула Байеса

$$f(\theta \mid y) = \frac{f(y \mid \theta) \cdot f(\theta)}{\int f(y \mid \theta) \cdot f(\theta) d\theta}$$

Формула полной вероятности

- $f(\theta \mid y)$ – апостериорное распределение
- $f(\theta)$ – априорное распределение
- $f(y \mid \theta)$ – правдоподобие
- $f(y)$ – "полная вероятность" выборки

Формула Байеса

$$f(\theta \mid y) = \frac{f(y \mid \theta) \cdot f(\theta)}{\int f(y \mid \theta) \cdot f(\theta) d\theta}$$

Константа
относительно θ

Формула Байеса

$$f(\theta \mid y) = \frac{f(y \mid \theta) \cdot f(\theta)}{\int f(y \mid \theta) \cdot f(\theta) d\theta}$$

Константа относительно θ

$$f(\theta \mid y) \propto f(y \mid \theta) \cdot f(\theta)$$

Формула Байеса

$$f(\theta \mid y) = \frac{f(y \mid \theta) \cdot f(\theta)}{\int f(y \mid \theta) \cdot f(\theta) d\theta}$$

Константа относительно θ

$$f(\theta \mid y) \propto f(y \mid \theta) \cdot f(\theta)$$

На практике ей часто пренебрегают, делают вычисления с точностью “до константы”, а после восстанавливают

Формула Байеса

$$f(\theta \mid y) = \frac{f(y \mid \theta) \cdot f(\theta)}{\int f(y \mid \theta) \cdot f(\theta) d\theta}$$

Константа относительно θ

$$f(\theta \mid y) \propto f(y \mid \theta) \cdot f(\theta)$$

На практике ей часто пренебрегают, делают вычисления с точностью “до константы”, а после восстанавливают

- Иногда константу восстановить сложно из-за сложного интеграла
- **Байес для бедных:** не восстанавливать константу, найти моду апостериорного распределения и использовать её в качестве точечной байесовской оценки

Формула Байеса

$$f(\theta \mid y) = \frac{f(y \mid \theta) \cdot f(\theta)}{\int f(y \mid \theta) \cdot f(\theta) d\theta}$$

$$f(\theta \mid y) \propto f(y \mid \theta) \cdot f(\theta)$$

Формула Байеса задаёт правило, по которому мы меняем наши представления о неизвестном параметре θ при поступлении новой информации в виде выборки

$$Posterior = \frac{Likelihood \cdot Prior}{Evidence}$$

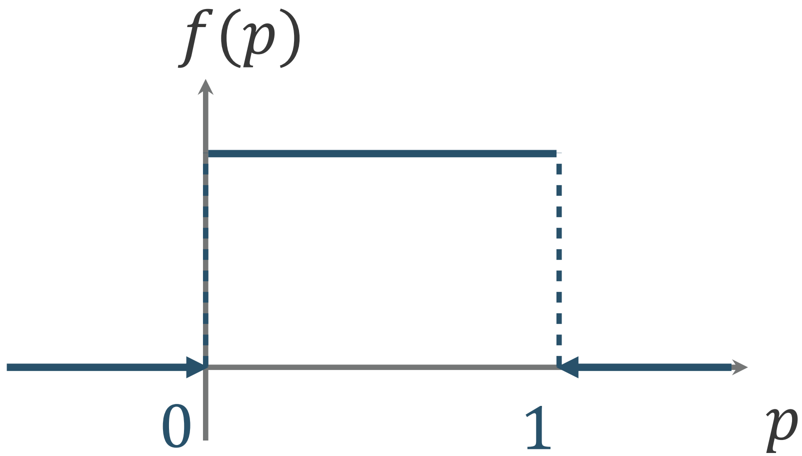
Ловля карасей

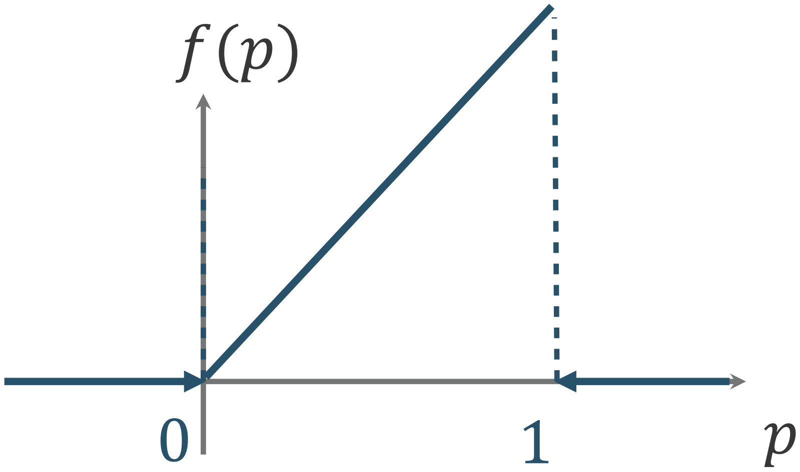
Пример: в озере около деревни Гипотезово живут караси и щуки, Петя поймал карася, щуку и ещё одного карася, а затем задумался о том, какая доля карасей водится в озере

$$y_1 = 1, y_2 = 0, y_3 = 1 \qquad\qquad p - ?$$

- Знаем о параметре p только то, что он изменяется от 0 до 1
- Возьмём в качестве априорного – равномерное распределение
- Если у Пети есть бабушка-рыбак и она говорит, что караси встречаются чаще, чем щуки, мы можем выразить это знание другим априорным распределением

Априорные распределения





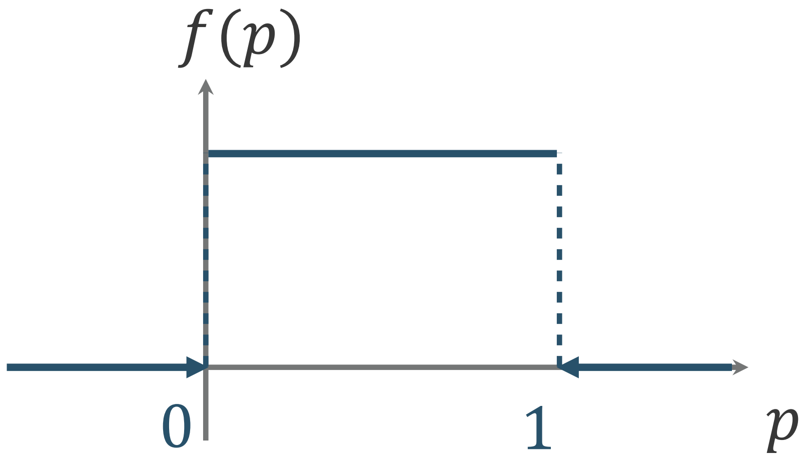
$$f(p) = \begin{cases} 1, p \in [0;1] \\ 0, p \notin [0;1] \end{cases}$$

$$f(p) = \begin{cases} 2p, p \in [0;1] \\ 0, p \notin [0;1] \end{cases}$$

$$p \sim U[0;1]$$

Байесовский вывод: равномерное

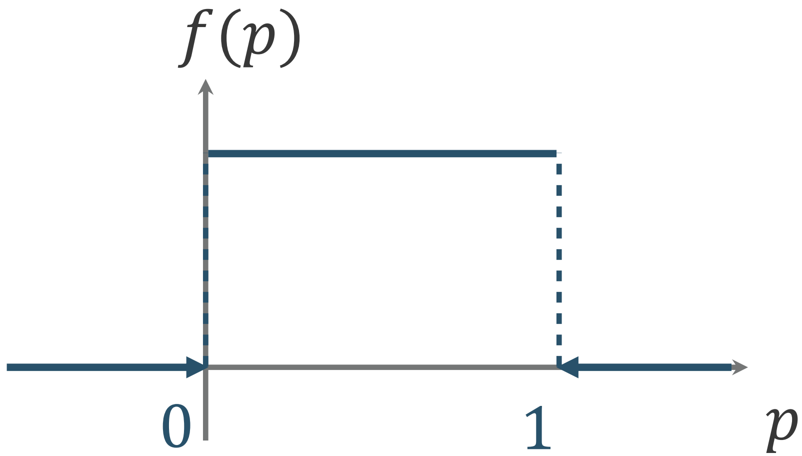
$f(p \mid y_1, y_2, y_3) =$



$$f(p) = \begin{cases} 1, p \in [0; 1] \\ 0, p \notin [0; 1] \end{cases}$$

$$p \sim U[0; 1]$$

Байесовский вывод: равномерное



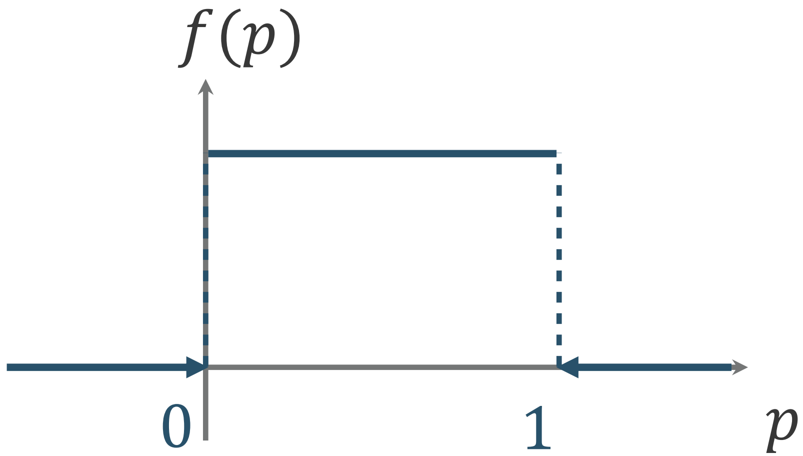
$$f(p \mid y_1, y_2, y_3) =$$

$$= \frac{f(y_1, y_2, y_3 \mid p) \cdot f(p)}{f(y_1, y_2, y_3)} \propto$$

$$f(p) = \begin{cases} 1, p \in [0; 1] \\ 0, p \notin [0; 1] \end{cases}$$

$$p \sim U[0; 1]$$

Байесовский вывод: равномерное



$$f(p \mid y_1, y_2, y_3) =$$

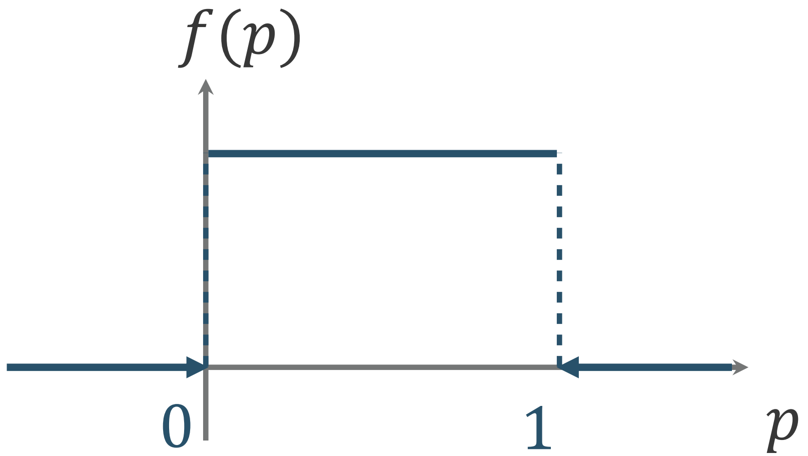
$$= \frac{f(y_1, y_2, y_3 \mid p) \cdot f(p)}{f(y_1, y_2, y_3)} \propto$$

$$\propto f(y_1, y_2, y_3 \mid p) \cdot f(p) =$$

$$f(p) = \begin{cases} 1, p \in [0; 1] \\ 0, p \notin [0; 1] \end{cases}$$

$$p \sim U[0; 1]$$

Байесовский вывод: равномерное



$$f(p) = \begin{cases} 1, p \in [0;1] \\ 0, p \notin [0;1] \end{cases}$$

$$p \sim U[0;1]$$

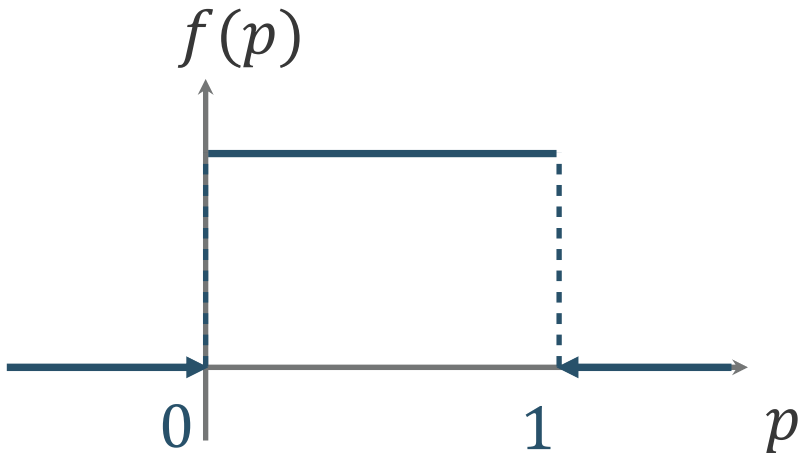
$$f(p \mid y_1, y_2, y_3) =$$

$$= \frac{f(y_1, y_2, y_3 \mid p) \cdot f(p)}{f(y_1, y_2, y_3)} \propto$$

$$\propto f(y_1, y_2, y_3 \mid p) \cdot f(p) =$$

$$= p^2 \cdot (1 - p) \cdot 1$$

Байесовский вывод: равномерное



$$f(p \mid y_1, y_2, y_3) =$$

$$= \frac{f(y_1, y_2, y_3 \mid p) \cdot f(p)}{f(y_1, y_2, y_3)} \propto$$

$$\propto f(y_1, y_2, y_3 \mid p) \cdot f(p) =$$

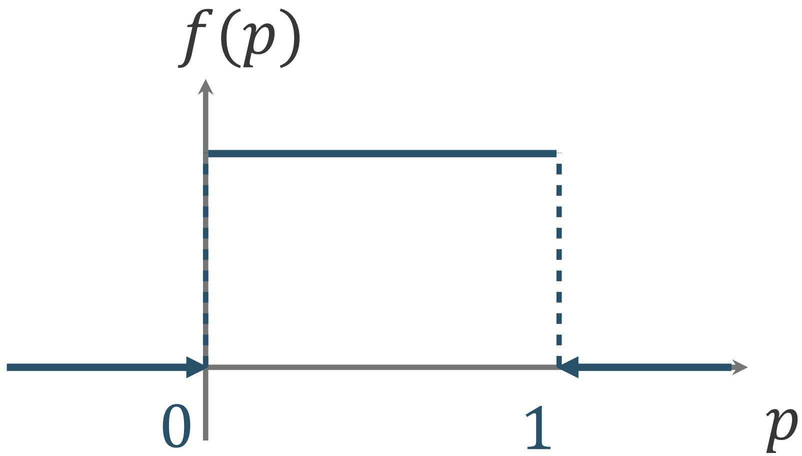
$$= p^2 \cdot (1 - p) \cdot 1$$

$$f(p) = \begin{cases} 1, p \in [0;1] \\ 0, p \notin [0;1] \end{cases}$$

$$f(p \mid y) = \begin{cases} C \cdot p^2(1-p), p \in [0;1] \\ 0, p \notin [0;1] \end{cases}$$

$$p \sim U[0;1]$$

Байесовский вывод: равномерное



$$f(p \mid y_1, y_2, y_3) =$$

$$= \frac{f(y_1, y_2, y_3 \mid p) \cdot f(p)}{f(y_1, y_2, y_3)} \propto$$

$$\propto f(y_1, y_2, y_3 \mid p) \cdot f(p) =$$

$$= p^2 \cdot (1 - p) \cdot 1$$

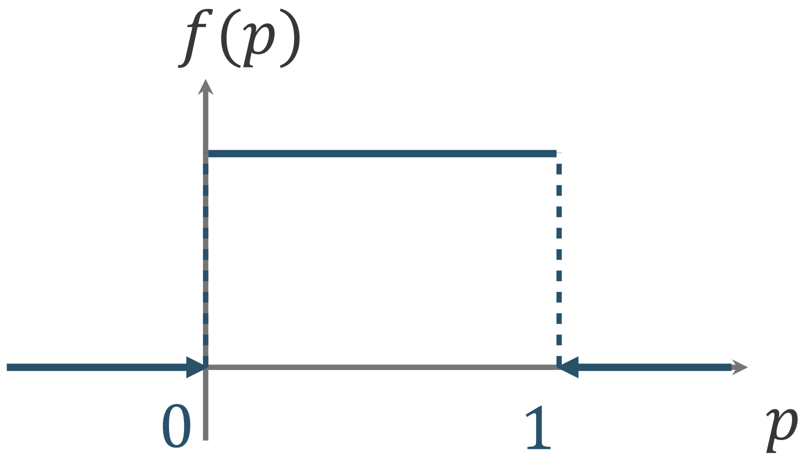
$$f(p) = \begin{cases} 1, p \in [0; 1] \\ 0, p \notin [0; 1] \end{cases}$$

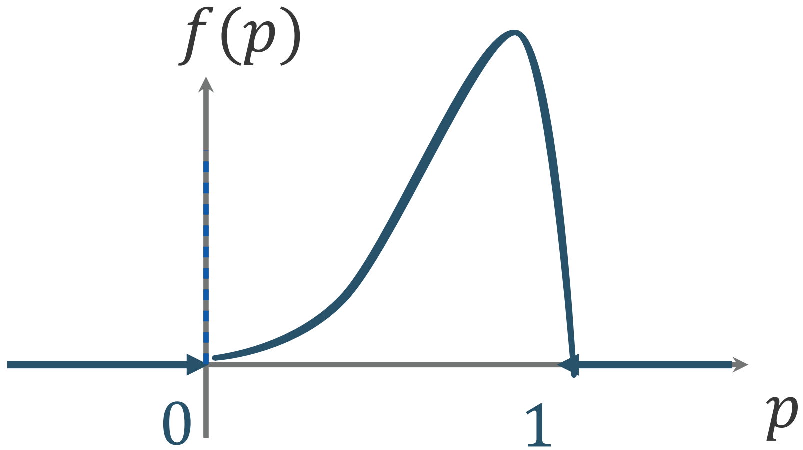
$$f(p \mid y) = \begin{cases} C \cdot p^2(1 - p), p \in [0; 1] \\ 0, p \notin [0; 1] \end{cases}$$

$$p \sim U[0; 1]$$

$$\int_0^1 C \cdot p^2(1 - p) = 1 \Rightarrow C = 12$$

Байесовский вывод: равномерное



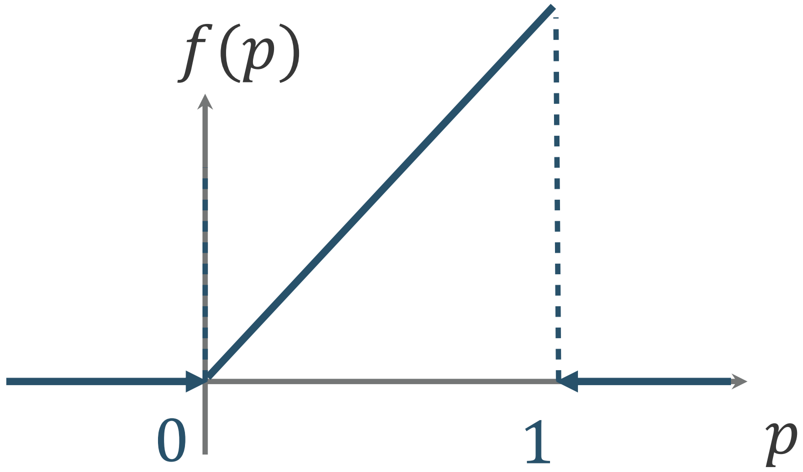


$$f(p) = \begin{cases} 1, p \in [0;1] \\ 0, p \notin [0;1] \end{cases}$$

$$f(p \mid y) = \begin{cases} 12 \cdot p^2(1-p), p \in [0;1] \\ 0, p \notin [0;1] \end{cases}$$

$$p \sim U[0;1]$$

Байесовский вывод: распределение Бабушки



$$f(p \mid y_1, y_2, y_3) =$$

$$= \frac{f(y_1, y_2, y_3 \mid p) \cdot f(p)}{f(y_1, y_2, y_3)} \propto$$

$$\propto f(y_1, y_2, y_3 \mid p) \cdot f(p) =$$

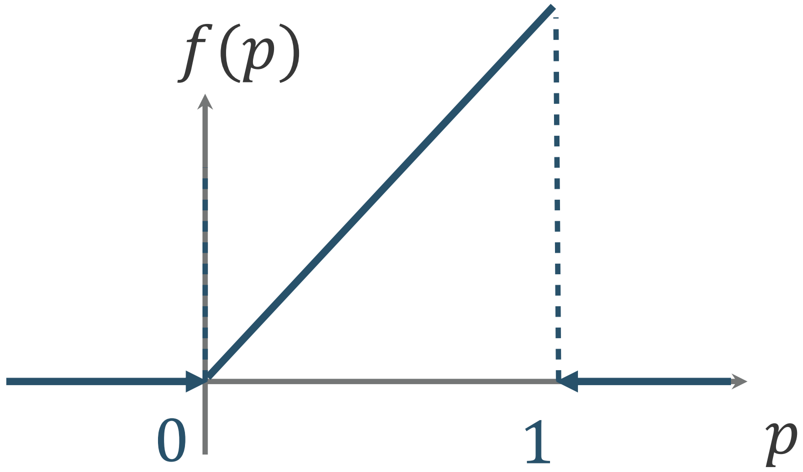
$$= p^2 \cdot (1 - p) \cdot 2p$$

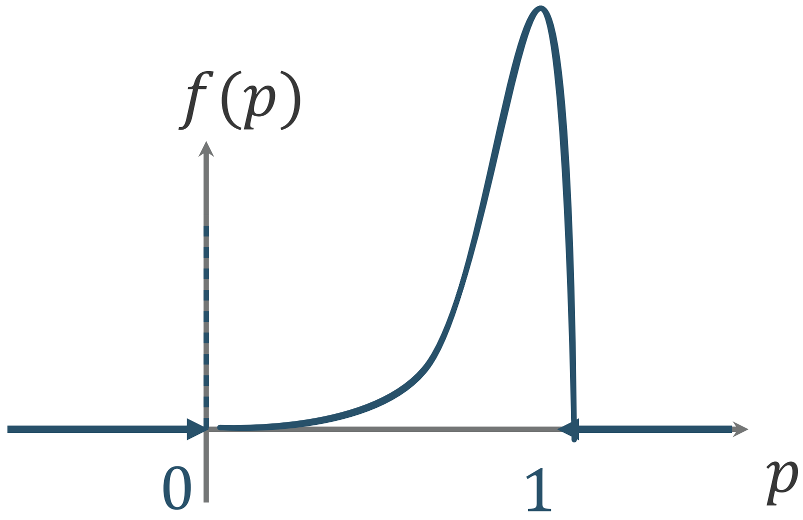
$$f(p) = \begin{cases} 2p, p \in [0; 1] \\ 0, p \notin [0; 1] \end{cases}$$

$$f(p \mid y) = \begin{cases} C \cdot p^3(1 - p), p \in [0; 1] \\ 0, p \notin [0; 1] \end{cases}$$

$$\int_0^1 C \cdot p^3(1 - p) = 1 \Rightarrow C = 20$$

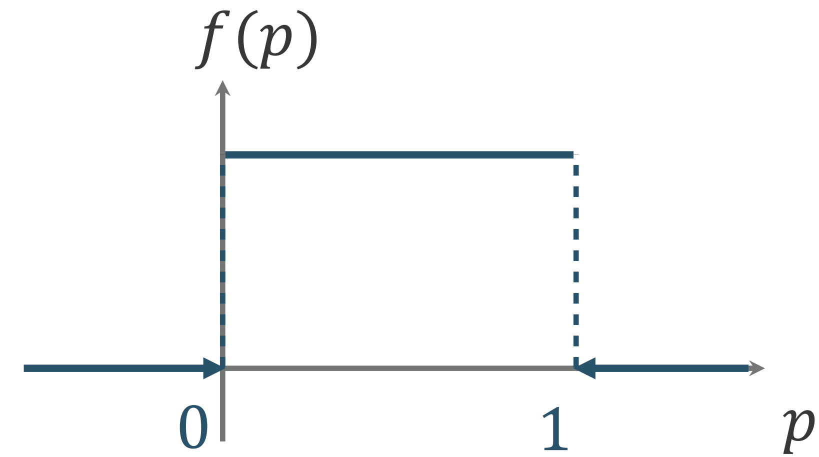
Байесовский вывод: распределение Бабушки

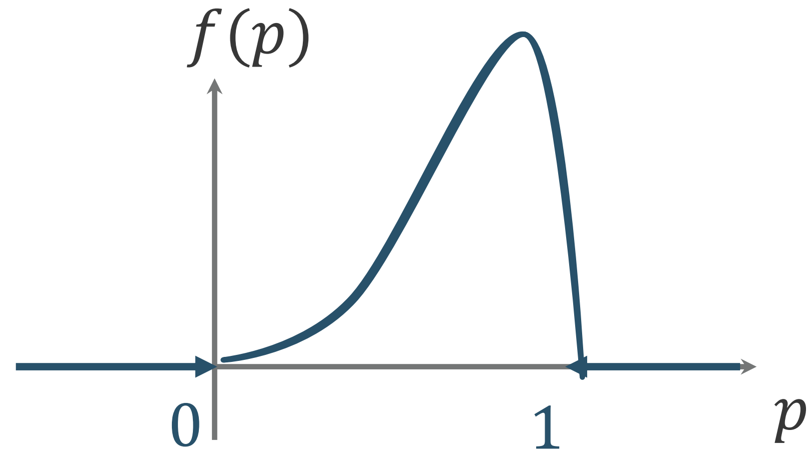


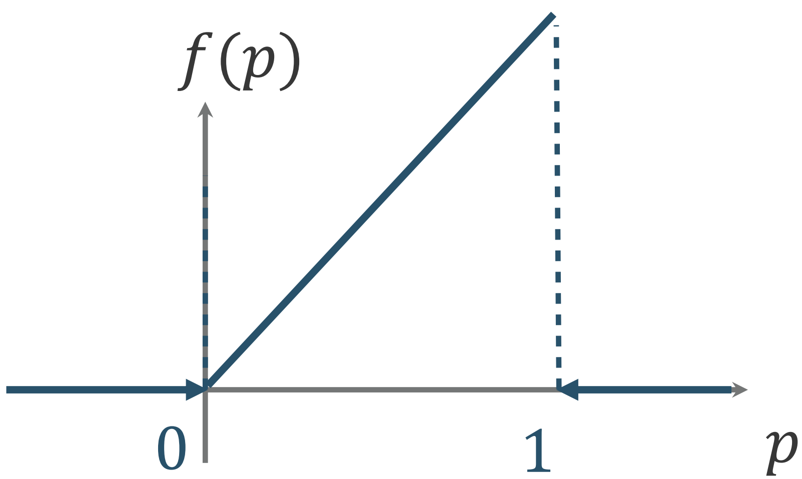


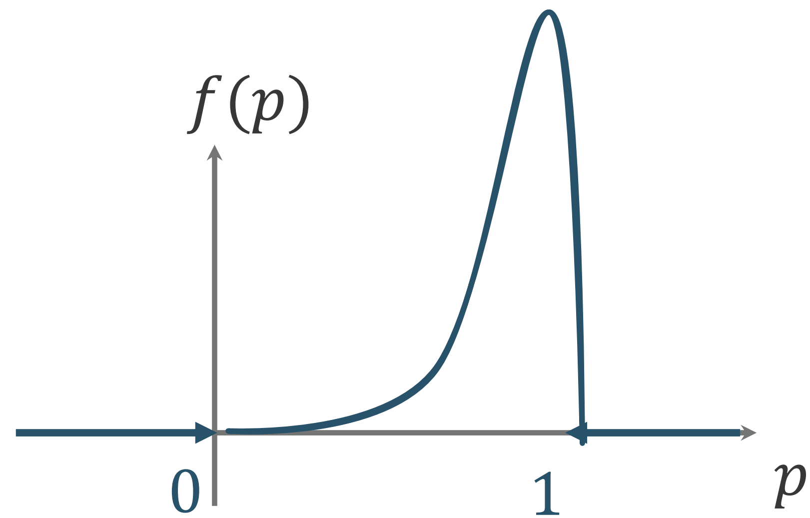
$$f(p) = \begin{cases} 2p, p \in [0; 1] \\ 0, p \notin [0; 1] \end{cases} \qquad f(p \mid y) = \begin{cases} 20\, p^3 (1-p), p \in [0; 1] \\ 0, p \notin [0; 1] \end{cases}$$

Апостериорные распределения





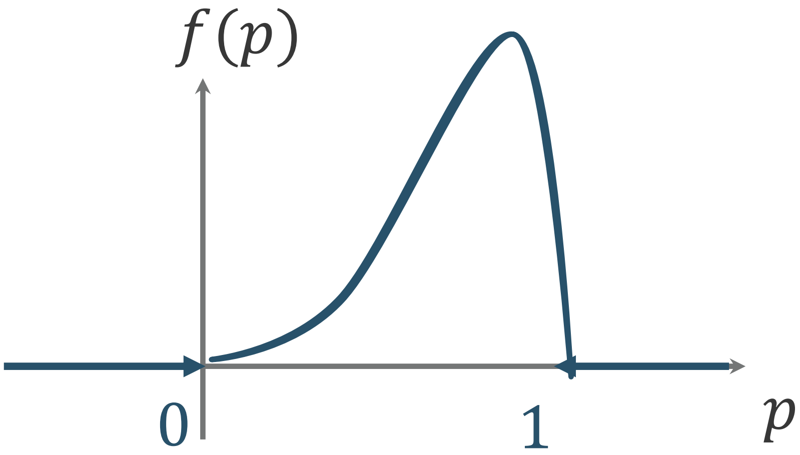




Апостериорные распределения

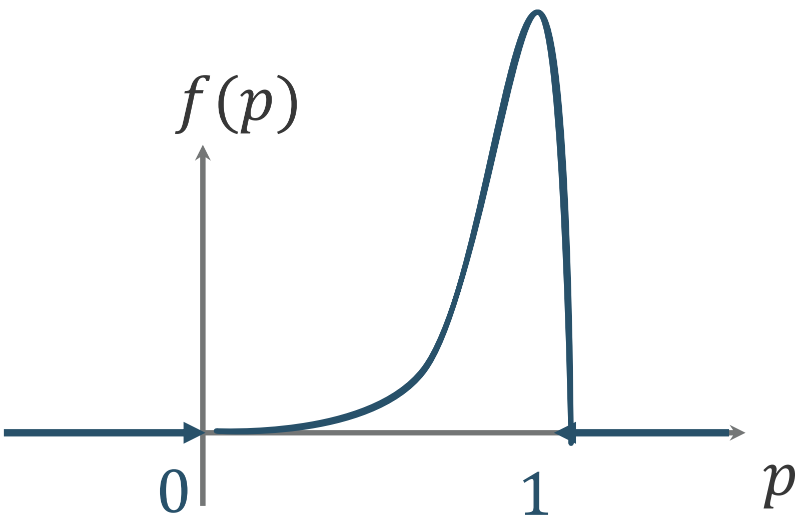
$$\mathbb{E}(p \mid y) = 0.6$$

$$\mathbb{P}(p > 0.5 \mid y) \approx 0.68$$



$$\mathbb{E}(p \mid y) = 0.66$$

$$\mathbb{P}(p > 0.5 \mid y) \approx 0.81$$



❗ Информация от бабушки сдвинула апостериорное распределение сильнее

Как выбрать априорное распределение

- Априорное распределение выбирается до сбора данных
- С помощью априорного распределения мы пытаемся описать своё незнание
- Оно отбрасывает заведомо неверные значения параметра
- Вы должны быть готовы сделать денежную ставку на выбранное вами априорное распределение
- Нельзя подстраивать априорное распределение под данные
- На выходе мы получаем целое апостериорное распределение, с помощью которого можем отвечать на разные вопросы

Байесовский интервал

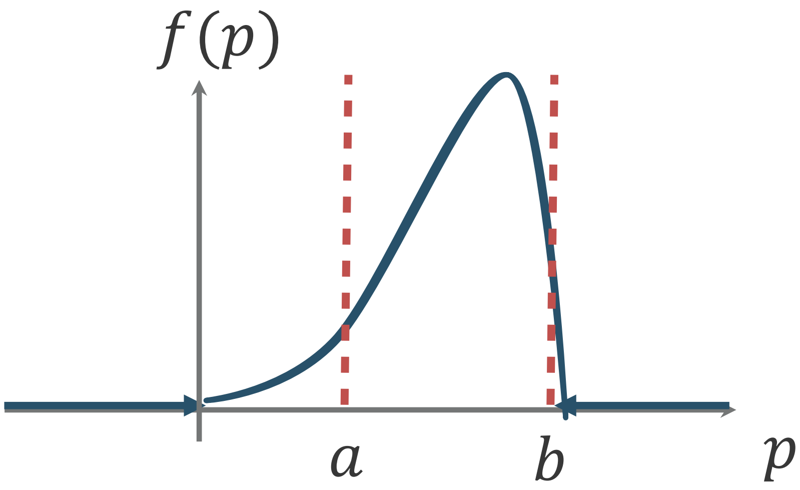
- В частотном подходе мы строили доверительные (confidence) интервалы
- В байесовском подходе интервальные оценки делаются с помощью байесовских (bayesian или credible) интервалов

Доверительный интервал: истинный параметр – константа, границы интервала – случайные величины; с вероятностью 0.95 границы накрывают истинное значение

Байесовский интервал: истинный параметр – случайная величина, он лежит с вероятностью 0.95 в этом интервале

Байесовский интервал

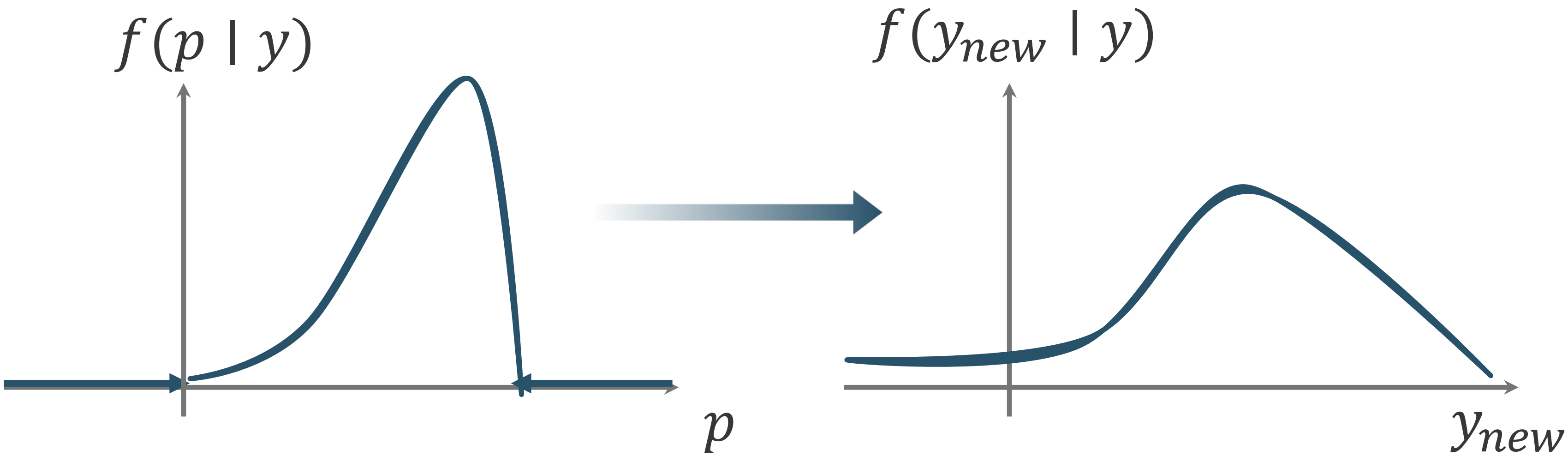
- Самый короткий байесовский интервал называется HPD (highest probability density interval)



$$\mathbb{P}(\, a \le p \le b \mid y\,) = 1 - \alpha$$

Прогнозирование

- Когда у нас есть апостериорное распределение для параметра, мы можем построить прогноз для того, какая рыба будет поймана в озере следующей:



❗ Хотим получить из апостериорного распределения для p, распределение для новых наблюдений

Прогнозирование

$$f(y_{new} \mid y)$$

Прогнозирование

$$f(y_{new} \mid y) = \int f(y_{new}, p \mid y)\, dp$$

Прогнозирование

$$f(y_{new} \mid y) = \int f(y_{new}, p \mid y)\, dp$$

$$= \int f(y_{new} \mid p) \cdot f(p \mid y)\, dp$$

Прогнозирование

$$f(y_{new} \mid y) = \int f(y_{new}, p \mid y)\, dp$$

$$= \int f(y_{new} \mid p) \cdot f(p \mid y)\, dp$$

модель

апостериорное распределение

Прогнозирование

$$f(y_{new} \mid y) = \int f(y_{new}, p \mid y)\, dp$$

$$= \int f(y_{new} \mid p) \cdot f(p \mid y)\, dp$$

модель

апостериорное
распределение

$$\mathbb{P}(y_{new} = 1 \mid y) = \int p \cdot f(p \mid y)\, dp = \mathbb{E}(p \mid y)$$

Прогнозирование

$$f(y_{new} \mid y) = \int f(y_{new}, p \mid y)\, dp$$

$$= \int f(y_{new} \mid p) \cdot f(p \mid y)\, dp$$

модель

апостериорное распределение

$$\mathbb{P}(y_{new} = 1 \mid y) = \int p \cdot f(p \mid y) dp = \mathbb{E}(p \mid y) = 0.6$$

для равномерного априорного

Прогноз:

y_{new}	0	1
$\mathbb{P}$	0.4	0.6

Точечные оценки

- Байесовский вывод даёт нам гораздо больше, чем просто точечную оценку
- В апостериорном распределении содержится вся информация о параметре p, с помощью него можно отвечать на любые вопросы
- Если нам нужна точечная оценка, то её выбор зависит
  от того, как нас накажут за ошибку

Что лучше потерять: руку или ногу?

- Если нас накажут квадратичной функцией потерь

$$L(\beta_F) = (\beta_F - \beta)^2,$$

в качестве точечной оценки лучше всего взять апостериорное математическое ожидание $\mathbb{E}(\beta \mid y)$

- Если нас накажут абсолютной функцией потерь

$$L(\beta_F) = |\beta_F - \beta|,$$

в качестве точечной оценки лучше всего взять апостериорную медиану $Med(\beta \mid y)$

Что лучше потерять: руку или ногу?

- Если нас наказывают по принципу: всё или ничего, в качестве точечной оценки уместно взять моду апостериорного распределения $Mod(\beta \mid y)$

Связь с правдоподобием

$$f(\theta \mid y) \propto f(y \mid \theta) \cdot f(\theta)$$

$$Posterior \propto Likelihood \cdot Prior$$

- Если априорное распределение равномерное, то апостериорное распределение пропорционально правдоподобию
- Если в качестве точечной оценки взять моду апостериорного распределения $Mod(\beta \mid y)$, получится метод максимального правдоподобия

Связь с регуляризацией

$$f(\theta \mid y) \propto f(y \mid \theta) \cdot f(\theta)$$

Прологарифмируем:

$$\ln f(\theta \mid y) \propto \ln f(y \mid \theta) + \ln f(\theta)$$

Связь с регуляризацией

$$f(\theta \mid y) \propto f(y \mid \theta) \cdot f(\theta)$$

Прологарифмируем:

$$\ln f(\theta \mid y) \propto \ln f(y \mid \theta) + \ln f(\theta)$$

Из логарифма правдоподобия можно получить функцию потерь

Задаёт наше мнение о параметре и добавляется к функции потерь (регуляризация)

Резюме

- Байесовский подход позволяет задавать естественные вопросы: какова вероятность того, что параметр больше заданного порога?
- На выходе получаем в качестве оценки целое распределение
- Гибкость априорных распределений: можно взять любое
- Сложность заключается в том, что даже для простых моделей приходится брать много интегралов

Сопряжённые распределения

Формула Байеса

$$f(\theta \mid y) = \frac{f(y \mid \theta) \cdot f(\theta)}{f(y)}$$

$$f(\theta \mid y) \propto f(y \mid \theta) \cdot f(\theta)$$

- $f(y \mid \theta)$ – апостериорное распределение
- $f(\theta)$ – априорное распределение
- $f(y \mid \theta)$ – правдоподобие
- $f(y)$ – "полная вероятность" выборки

Маша и сопряжённые распределения

- Когда мы работали с Машей и Медведями, априорное и апостериорное распределения оказались нормальными

Априорно: $m \sim N(1, 4^2)$ **Модель:** $y \mid m \sim N(m, 2^2)$

Апостериорно: $m \mid y \sim N\left(-{}^{1}/_{9}, \left({}^{4}/_{3}\right)^2\right)$

Маша и сопряжённые распределения

- Когда мы работали с Машей и Медведями, априорное и апостериорное распределения оказались нормальными

Априорно: $m \sim N(1, 4^2)$ **Модель:** $y \mid m \sim N(m, 2^2)$

Апостериорно: $m \mid y \sim N\left(-{}^{1}/_{9}, \left({}^{4}/_{3}\right)^2\right)$

Все распределения – нормальные! Ну не чудо ли?

Маша и сопряжённые распределения

- Когда мы работали с Машей и Медведями, априорное и апостериорное распределения оказались нормальными

Априорно: $m \sim N(\mu, \tau^2)$ **Модель:** $y \mid m \sim N(m, \sigma^2)$

Апостериорно: $m \mid y \sim N(\tilde{\mu}, \tilde{\tau}^2)$

- Для общего случая можно даже вывести формулы пересчёта параметров априорного нормального в апостериорное нормальное:

Маша и сопряжённые распределения

- Когда мы работали с Машей и Медведями, априорное и апостериорное распределения оказались нормальными

Априорно: $m \sim N(\mu, \tau^2)$ **Модель:** $y \mid m \sim N(m, \sigma^2)$

Апостериорно: $m \mid y \sim N(\tilde{\mu}, \tilde{\tau}^2)$

- Для общего случая можно даже вывести формулы пересчёта параметров априорного нормального в апостериорное нормальное:

$$\tilde{\mu} = \frac{\mu \cdot \sigma^2/n}{\sigma^2/n + \tau^2} + \frac{\bar{y}}{\sigma^2/n + \tau^2} \qquad \tilde{\tau} = \frac{\tau^2 \cdot \sigma^2/n}{\sigma^2/n + \tau^2}$$

Петя и сопряжённые распределения

Петя, рыбалка, караси

$p - ?$

$y_1, y_2, \dots y_n$

Петя и сопряжённые распределения

Петя, рыбалка, караси

$p - ?$

$y_1, y_2, \ldots y_n$

Априорно: $p \sim U[0; 1]$

$f(p \mid y)$

Петя и сопряжённые распределения

Петя, рыбалка, караси

$p - ?$

$y_1, y_2, \dots y_n$

Априорно: $p \sim U[0;1]$

$$f(p \mid y) \propto f(y \mid p) \cdot f(p) =$$

Петя и сопряжённые распределения

Петя, рыбалка, караси

$p - ?$

$y_1, y_2, \dots y_n$

Априорно: $p \sim U[0;1]$

$$f(p \mid y) \propto f(y \mid p) \cdot f(p) =$$

$$= p^{\sum y_i} \cdot (1-p)^{n - \sum y_i} \cdot 1$$

Петя и сопряжённые распределения

Петя, рыбалка, караси $p - ?$ $y_1, y_2, \dots y_n$

Априорно: $p \sim U[0;1]$

$$f(p \mid y) \propto f(y \mid p) \cdot f(p) =$$

$$= p^{\sum y_i} \cdot (1-p)^{n-\sum y_i} \cdot 1 = p^{n\bar{y}} \cdot (1-p)^{n-n\bar{y}}$$

Петя и сопряжённые распределения

Петя, рыбалка, караси $\quad p - ? \quad y_1, y_2, \dots y_n$

Априорно: $p \sim U[0;1]$

$$f(p \mid y) \propto f(y \mid p) \cdot f(p) =$$

$$= p^{\sum y_i} \cdot (1-p)^{n-\sum y_i} \cdot 1 = p^{n\bar{y}} \cdot (1-p)^{n-n\bar{y}}$$

Если попробовать восстановить константу:

$$f(p \mid y) = \frac{1}{\mathrm{B}(\bar{y}+1, n-n\bar{y}+1)} \cdot p^{n\bar{y}} \cdot (1-p)^{n-n\bar{y}}$$

Петя и сопряжённые распределения

Петя, рыбалка, караси $\qquad p - ? \qquad y_1, y_2, \dots y_n$

Априорно: $p \sim U[0;1]$

$$f(p \mid y) \propto f(y \mid p) \cdot f(p) =$$

$$= p^{\sum y_i} \cdot (1-p)^{n-\sum y_i} \cdot 1 = p^{n\bar{y}} \cdot (1-p)^{n-n\bar{y}}$$

Если попробовать восстановить константу:

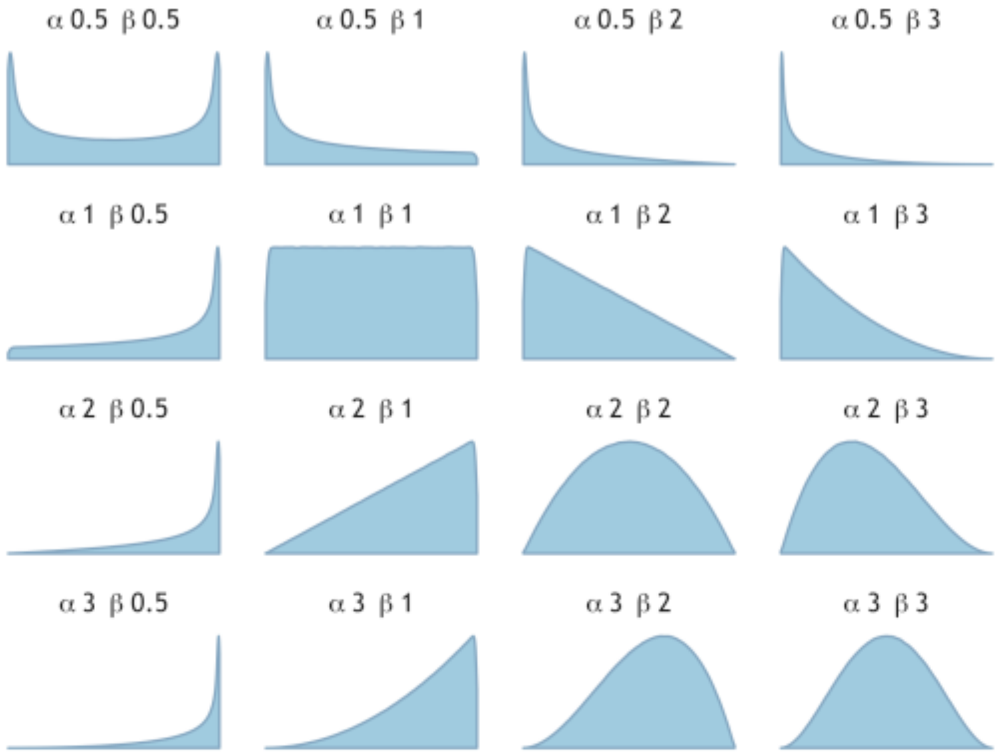
$$f(p \mid y) = \frac{1}{\mathrm{B}(\bar{y}+1, n-n\bar{y}+1)} \cdot p^{n\bar{y}} \cdot (1-p)^{n-n\bar{y}}$$

Просто какая-то константа (бета-функция)

Бета-распределение

- Бета-распределение очень эластичное и хорошо подходит для моделирования долей

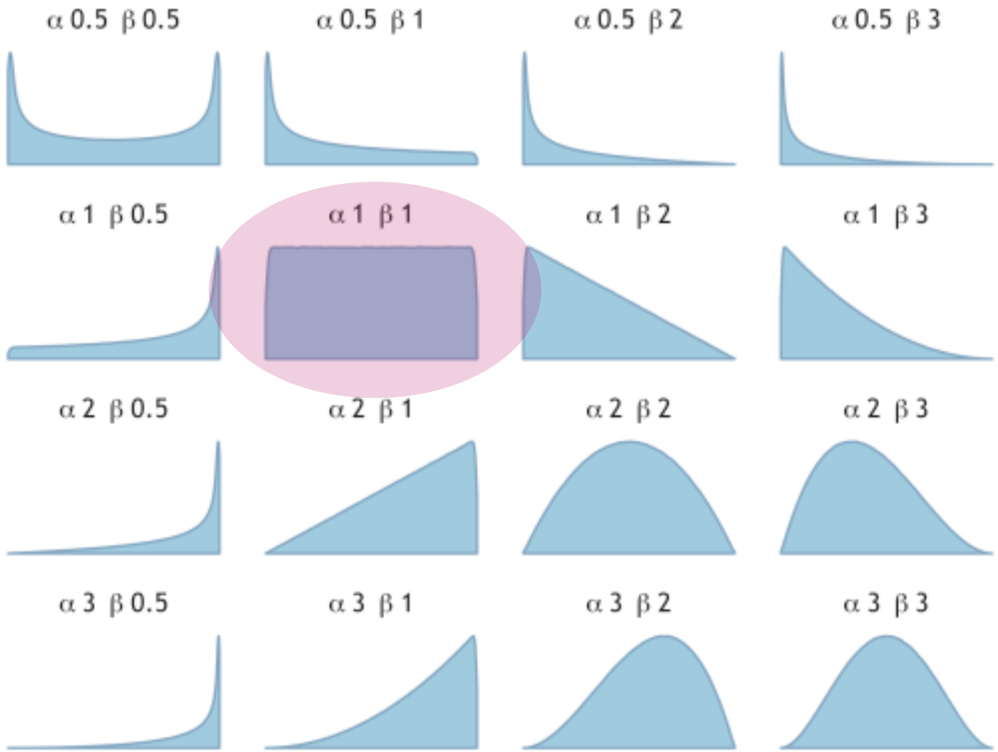
$$f(p) = \frac{1}{\mathrm{B}(\alpha, \beta)} \cdot p^{\alpha-1} \cdot (1-p)^{\beta-1}$$



Бета-распределение

- Бета-распределение очень эластичное и хорошо подходит для моделирования долей

$$f(p) = \frac{1}{\mathrm{B}(\alpha, \beta)} \cdot p^{\alpha-1} \cdot (1-p)^{\beta-1}$$



Равномерное распределение входит в это семейство распределений

Петя и сопряжённые распределения

Петя, рыбалка, караси $p - ?$ $y_1, y_2, \dots y_n$

Априорно: $p \sim \mathrm{B}(\alpha, \beta)$

$$f(p) = \frac{1}{\mathrm{B}(\alpha, \beta)} \cdot p^{\alpha-1} \cdot (1-p)^{\beta-1}$$

Петя и сопряжённые распределения

Петя, рыбалка, караси

$p - ?$

$y_1, y_2, \dots y_n$

Априорно: $p \sim \mathrm{B}(\alpha, \beta)$

$$f(p) = \frac{1}{\mathrm{B}(\alpha, \beta)} \cdot p^{\alpha-1} \cdot (1-p)^{\beta-1}$$

$$f(p) \propto \cdot\, p^{\alpha-1} \cdot (1-p)^{\beta-1}$$

Константой всегда можно пренебречь

Петя и сопряжённые распределения

Петя, рыбалка, караси $\quad p - ? \quad y_1, y_2, \dots y_n$

Априорно: $p \sim \mathrm{B}(\alpha, \beta)$

$$f(p) = \frac{1}{\mathrm{B}(\alpha, \beta)} \cdot p^{\alpha - 1} \cdot (1 - p)^{\beta - 1}$$

$$f(p) \propto \cdot\, p^{\alpha - 1} \cdot (1 - p)^{\beta - 1}$$

$$f(p \mid y) \propto f(y \mid p) \cdot f(p) =$$

Петя и сопряжённые распределения

Петя, рыбалка, караси $\quad p - ? \quad y_1, y_2, \dots y_n$

Априорно: $p \sim \mathrm{B}(\alpha, \beta)$

$$f(p) = \frac{1}{\mathrm{B}(\alpha, \beta)} \cdot p^{\alpha-1} \cdot (1-p)^{\beta-1}$$

$$f(p) \propto \cdot\, p^{\alpha-1} \cdot (1-p)^{\beta-1}$$

$$f(p \mid y) \propto f(y \mid p) \cdot f(p) =$$

$$= p^{\sum y_i} \cdot (1-p)^{n-\sum y_i} \cdot p^{\alpha-1} \cdot (1-p)^{\beta-1}$$

Петя и сопряжённые распределения

Петя, рыбалка, караси $\quad p - ? \quad y_1, y_2, \dots y_n$

Априорно: $p \sim \mathrm{B}(\alpha, \beta)$

$$f(p) = \frac{1}{\mathrm{B}(\alpha, \beta)} \cdot p^{\alpha-1} \cdot (1-p)^{\beta-1}$$

$$f(p) \propto \cdot\, p^{\alpha-1} \cdot (1-p)^{\beta-1}$$

$$f(p \mid y) \propto f(y \mid p) \cdot f(p) =$$

$$= p^{\sum y_i} \cdot (1-p)^{n-\sum y_i} \cdot p^{\alpha-1} \cdot (1-p)^{\beta-1}$$

$$= p^{n\bar{y}+\alpha-1} \cdot (1-p)^{n-n\bar{y}+\beta-1}$$

Петя и сопряжённые распределения

Петя, рыбалка, караси $\quad p - ? \quad y_1, y_2, \ldots y_n$

Априорно: $p \sim \mathrm{B}(\alpha, \beta)$

$$f(p) = \frac{1}{\mathrm{B}(\alpha, \beta)} \cdot p^{\alpha-1} \cdot (1-p)^{\beta-1}$$

$$f(p) \propto \cdot\, p^{\alpha-1} \cdot (1-p)^{\beta-1}$$

Если восстановим константу, получим бета-распределение

$$f(p \mid y) \propto f(y \mid p) \cdot f(p) =$$

$$= p^{\sum y_i} \cdot (1-p)^{n-\sum y_i} \cdot p^{\alpha-1} \cdot (1-p)^{\beta-1}$$

$$= p^{n\bar{y}+\alpha-1} \cdot (1-p)^{n-n\bar{y}+\beta-1}$$

Петя и сопряжённые распределения

Петя, рыбалка, караси $\qquad p - ? \qquad y_1, y_2, \dots y_n$

Априорно: $p \sim \mathrm{B}(\alpha, \beta)$

$$f(p) \propto \cdot\, p^{\alpha-1} \cdot (1-p)^{\beta-1}$$

Апостериорно: $p \sim \mathrm{B}(\alpha + n\bar{y}, \beta + n - n\bar{y})$

$$f(p \mid y) \propto p^{n\bar{y}+\alpha-1} \cdot (1-p)^{n-n\bar{y}+\beta-1}$$

Петя и сопряжённые распределения

Петя, рыбалка, караси $\quad p - ? \quad y_1, y_2, \dots y_n$

Априорно: $p \sim \mathrm{B}(\alpha, \beta)$

Апостериорно: $p \sim \mathrm{B}\left(\tilde{\alpha}, \tilde{\beta}\right)$

Формулы пересчёта:

$$\tilde{\alpha} = \alpha + n\bar{y}$$

$$\tilde{\beta} = \beta + n - n\bar{y}$$

Пример: сортировка комментариев на Reddit

- Нужно отсортировать комментарии
- На комментарии ставят голоса “за (upvotes)” и “против (downvotes)”
- Как определить какие комментарии лучшие?
- Нам нужно оценить “фактическое” число голосов “за”
- Посчитать разницу или долю недостаточно
- Если комментарий оценило мало людей, мы не можем быть уверены в нашей оценке
- Хочется учесть эту неопределённость

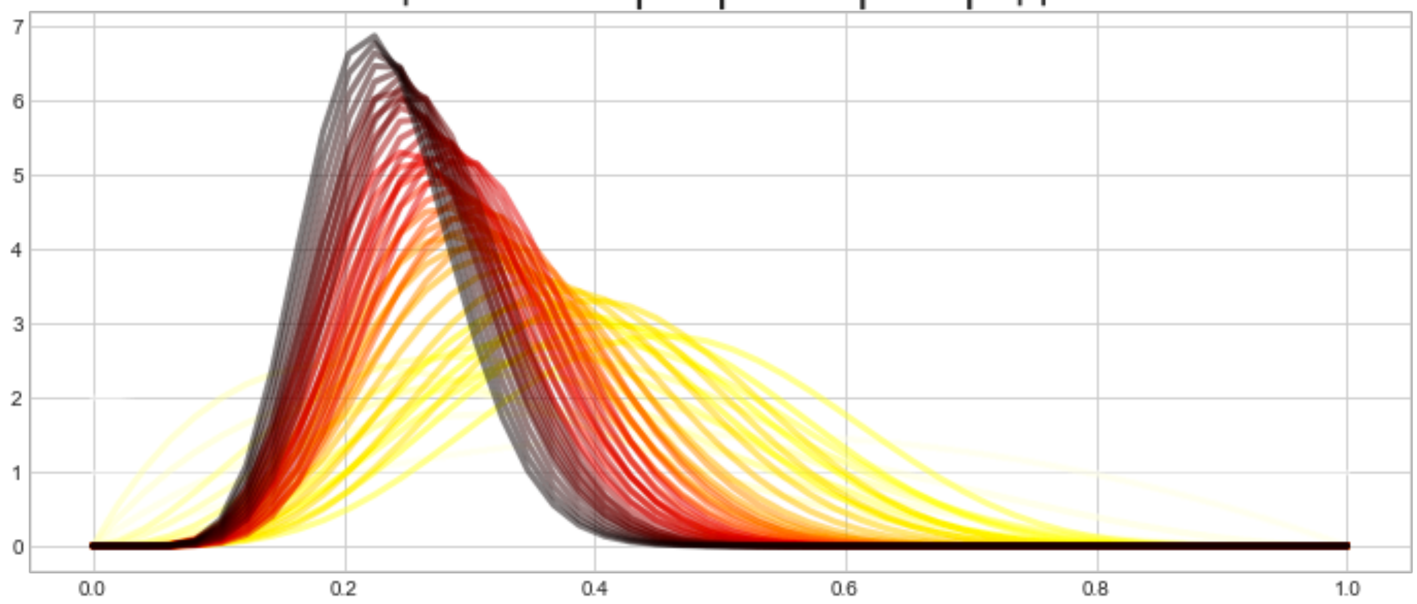
Пример: сортировка комментариев на Reddit

- Пусть голоса за это единицы, а против – нули
- Нам нужно оценить долю голосов за p

Априорно: $p \sim U[0;1] = \mathrm{B}(1,1)$

Апостериорно: $p \sim \mathrm{B}(\tilde{\alpha}, \tilde{\beta})$

Эволюция апостериорного распределения

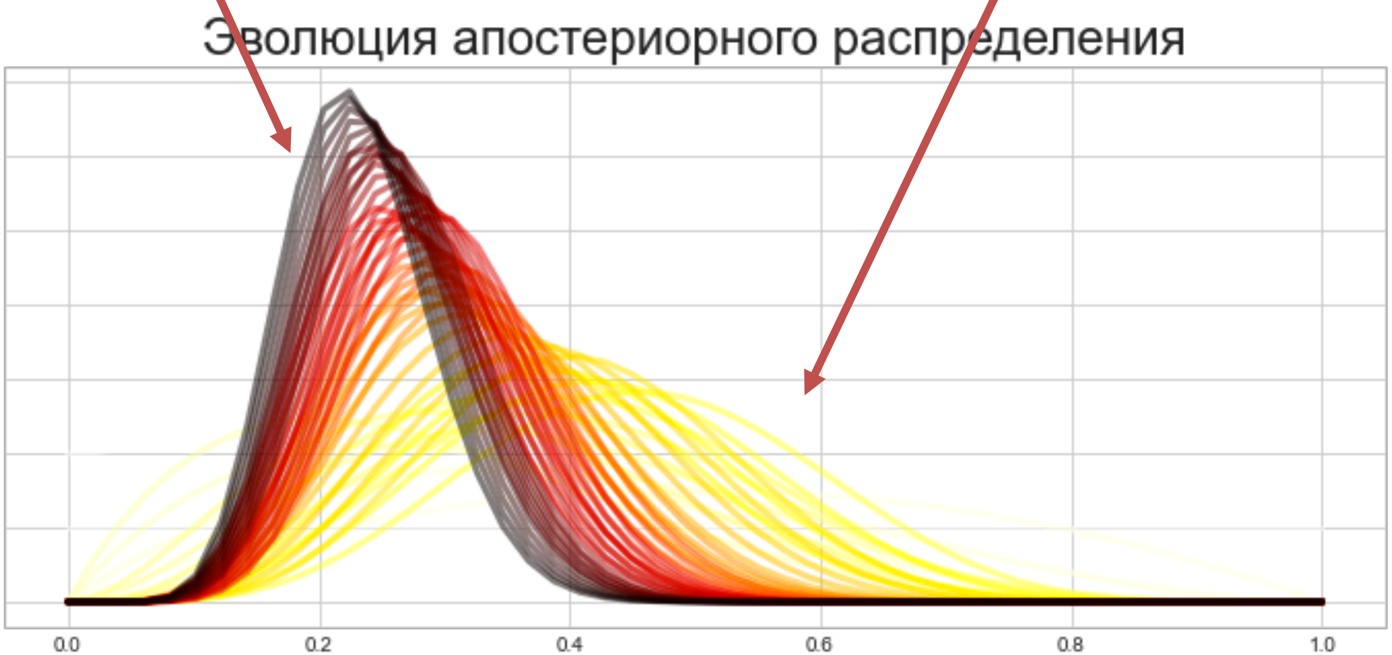


Пример: сортировка комментариев на Reddit

- При каждом новом “за” или “против” обновляем параметры распределения

Постепенно наша уверенность в оценке растёт

Поначалу дисперсия p большая

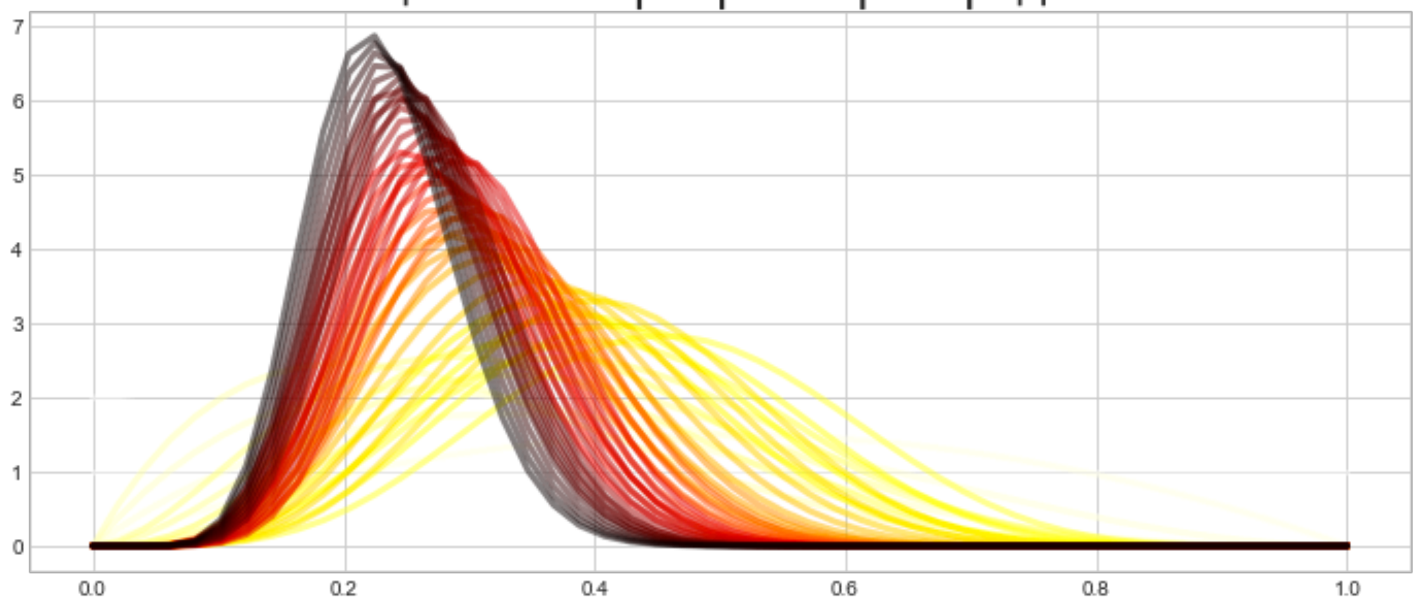


Пример: сортировка комментариев на Reddit

- При каждом новом “за” или “против” обновляем параметры распределения

Мы можем обновлять нашу модель по мере поступления данных, не делая расчётов с нуля

Эволюция апостериорного распределения



Пример: сортировка комментариев на Reddit

- Для сортировки будем использовать 5% квантиль апостериорного распределения
- Для него можно вывести формулу:

$$\frac{\tilde{\alpha}}{\tilde{\alpha}+\tilde{\beta}} - 1.65 \cdot \sqrt{\frac{\tilde{\alpha}\tilde{\beta}}{\left(\tilde{\alpha}+\tilde{\beta}\right)^2(\tilde{\alpha}+\tilde{\beta}+1)}}$$

- Такая сортировка означает максимальную осторожность в объявлении комментария лучшим
- Такую систему можно расширить на оценки с присваиванием звёзд и другие рейтинги

Резюме

- Полный байесовский вывод можно сделать не для всех распределений , такие распределения встречаются довольно редко ☹
- Основная проблема возникает с вычислением константы из знаменателя (сложные интегралы)
- **Сопряжённые распределения –** одна из ситуаций, когда его можно сделать в аналитическом виде
- Даже для простого случая с долями формулы оказываются сложноватыми
- Чтобы оценить модель с произвольным априорным распределением, нужны приближённые алгоритмы

➤ *https://en.wikipedia.org/wiki/Conjugate_prior*

MCMC-алгоритмы

Вычисление апостериорного распределения

- Аналитический подход: возникают вычислительные сложности, часто аналитически вычислить распределение невозможно

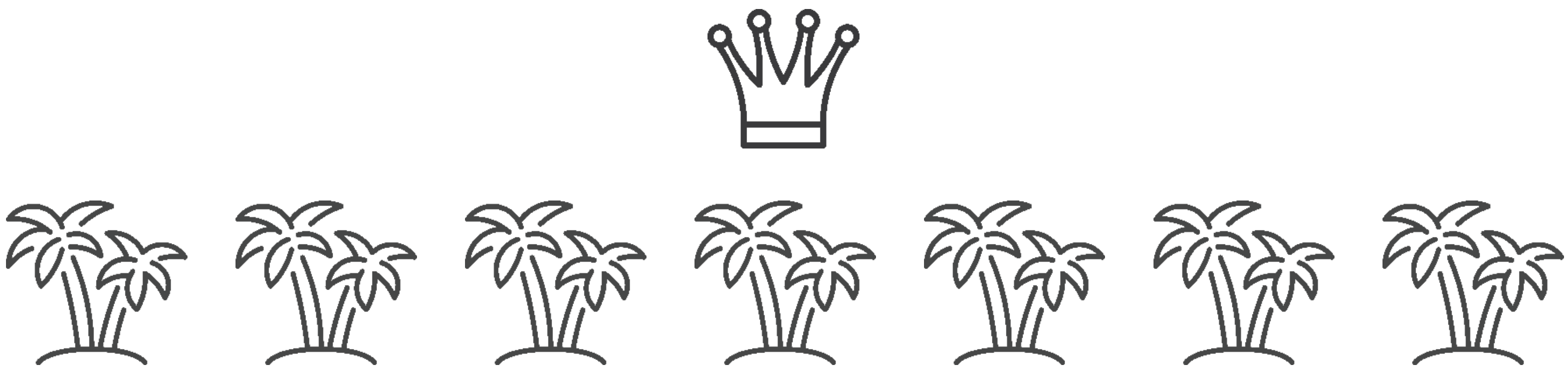
- Можно попробовать аппроксимировать неизвестное распределение каким-нибудь похожим на него

- Можно попробовать алгоритма Монте-Карло по схеме Марковской цепи (Markov Chain Monte Carlo, MCMC)

❗ MCMC сгенерирует нам выборку из апостериорного распределения, в аналитическом виде формулу мы знать не будем

MCMC

Король Марков



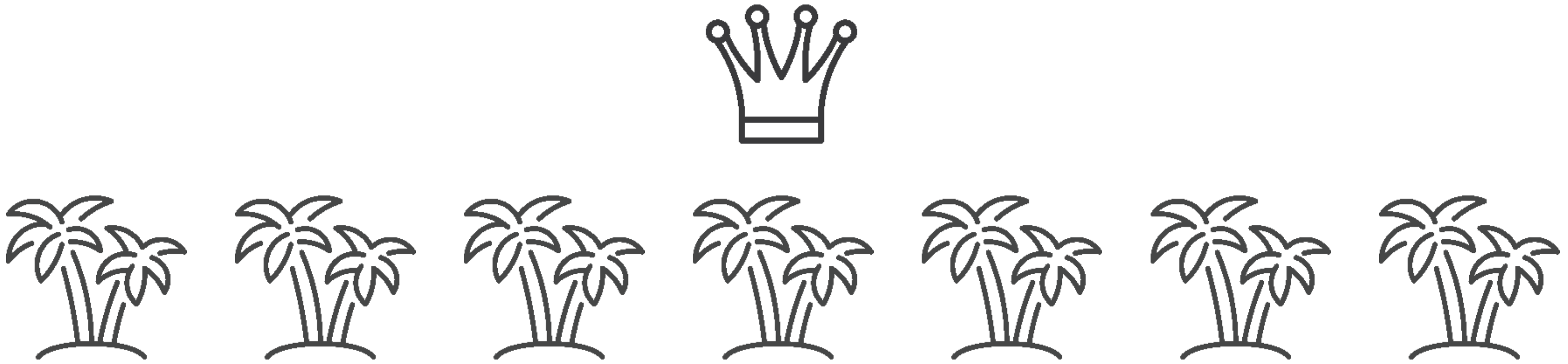
Архипелаг Метрополиса

MCMC

Обязательство: король обещает посетить каждый остров пропорционально численности его населения

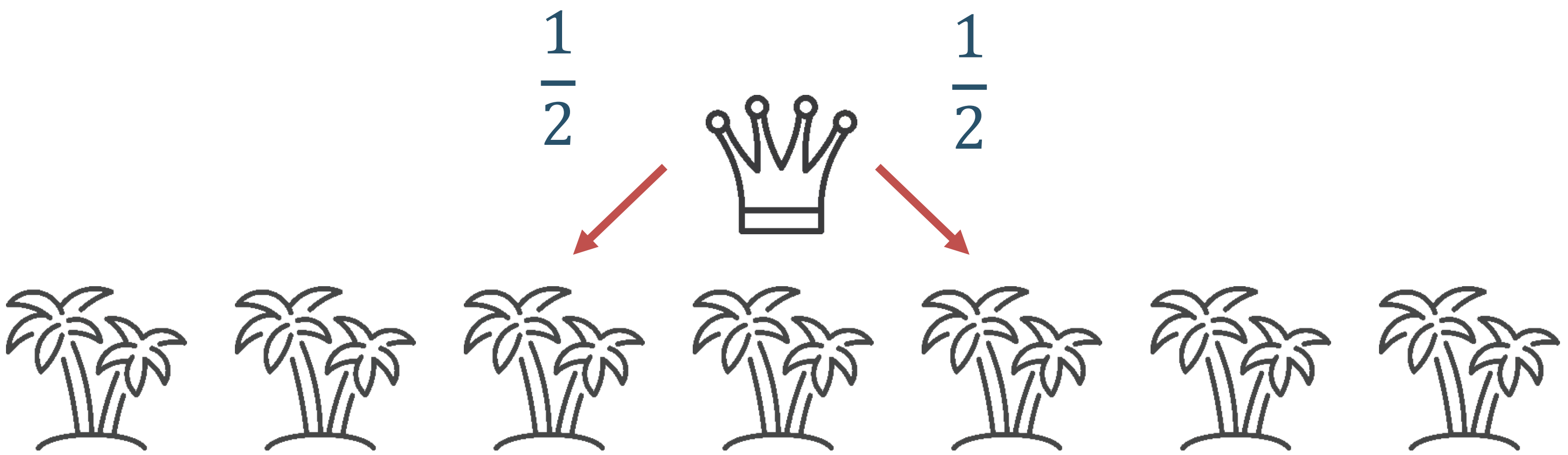
MCMC

0. Король приезжает на случайный первый остров



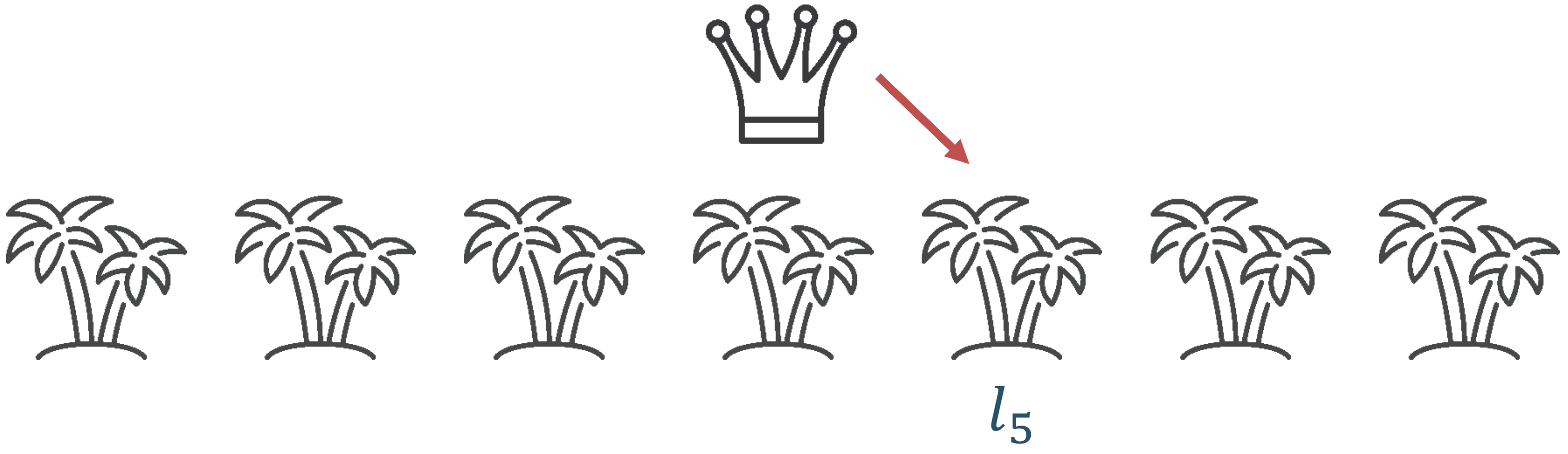
MCMC

0. Король приезжает на случайный первый остров
1. Король с вероятностью 0.5 выбирает один из соседних к нему островов



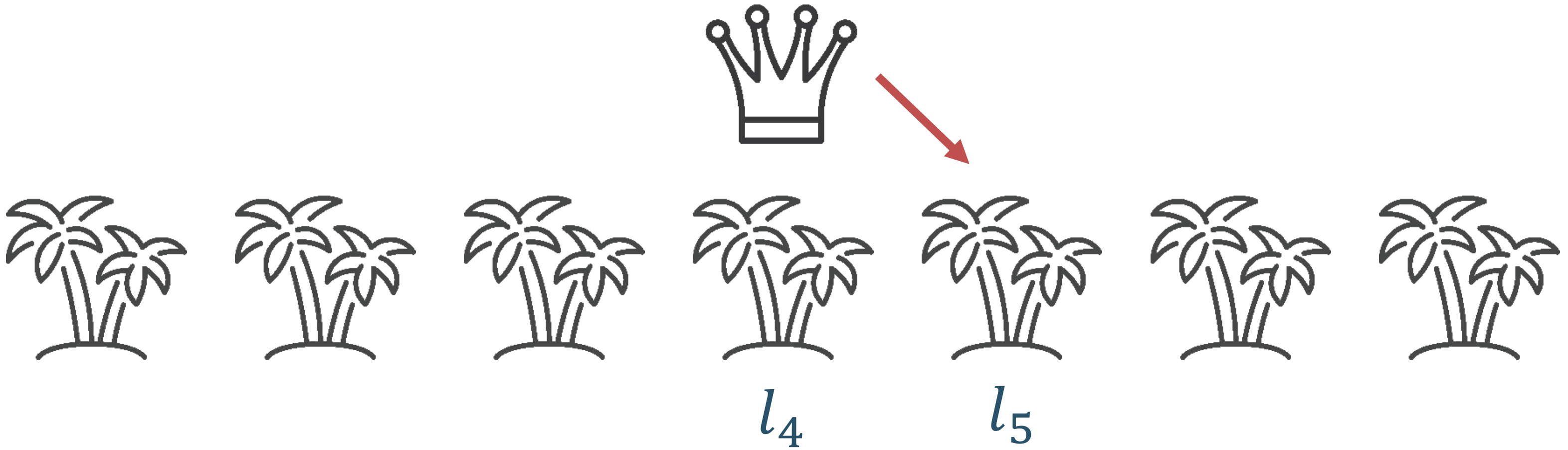
MCMC

0. Король приезжает на случайный первый остров
1. Король с вероятностью 0.5 выбирает один из соседних к нему островов
2. Находит популяцию выбранного им острова



MCMC

0. Король приезжает на случайный первый остров
1. Король с вероятностью 0.5 выбирает один из соседних к нему островов
2. Находит популяцию выбранного им острова
3. Находит популяцию текущего острова



MCMC

0. Король приезжает на случайный первый остров
1. Король с вероятностью 0.5 выбирает один из соседних к нему островов
2. Находит популяцию выбранного им острова
3. Находит популяцию текущего острова
4 Переезжает на новый остров с вероятностью ${}^{l_5}/_{l_4}$

MCMC

0. Король приезжает на случайный первый остров
1. Король с вероятностью 0.5 выбирает один из соседних к нему островов
2. Находит популяцию выбранного им острова
3. Находит популяцию текущего острова

4 Переезжает на новый остров с вероятностью ${}^{l_5}/_{l_4}$

5. Повторяет пункты 1-4 заново

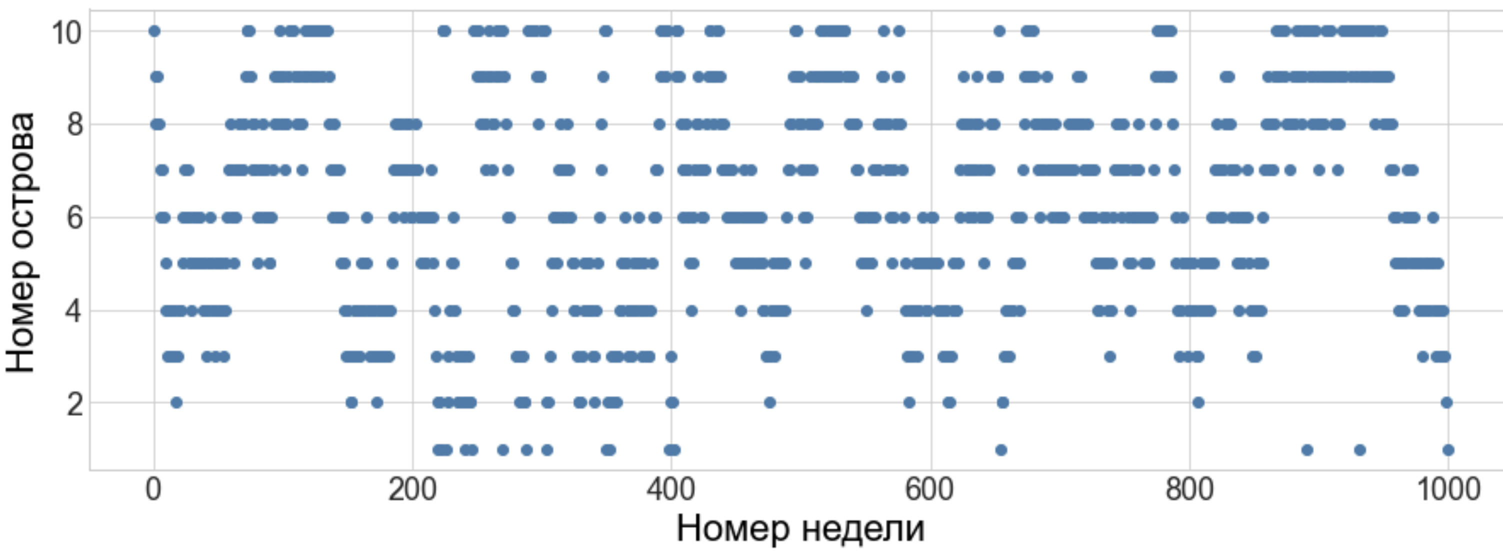
MCMC

0. Король приезжает на случайный первый остров
1. Король с вероятностью 0.5 выбирает один из соседних к нему островов
2. Находит популяцию выбранного им острова
3. Находит популяцию текущего острова
4 Переезжает на новый остров с вероятностью ${}^{l_5}/_{l_4}$
5. Повторяет пункты 1-4 заново

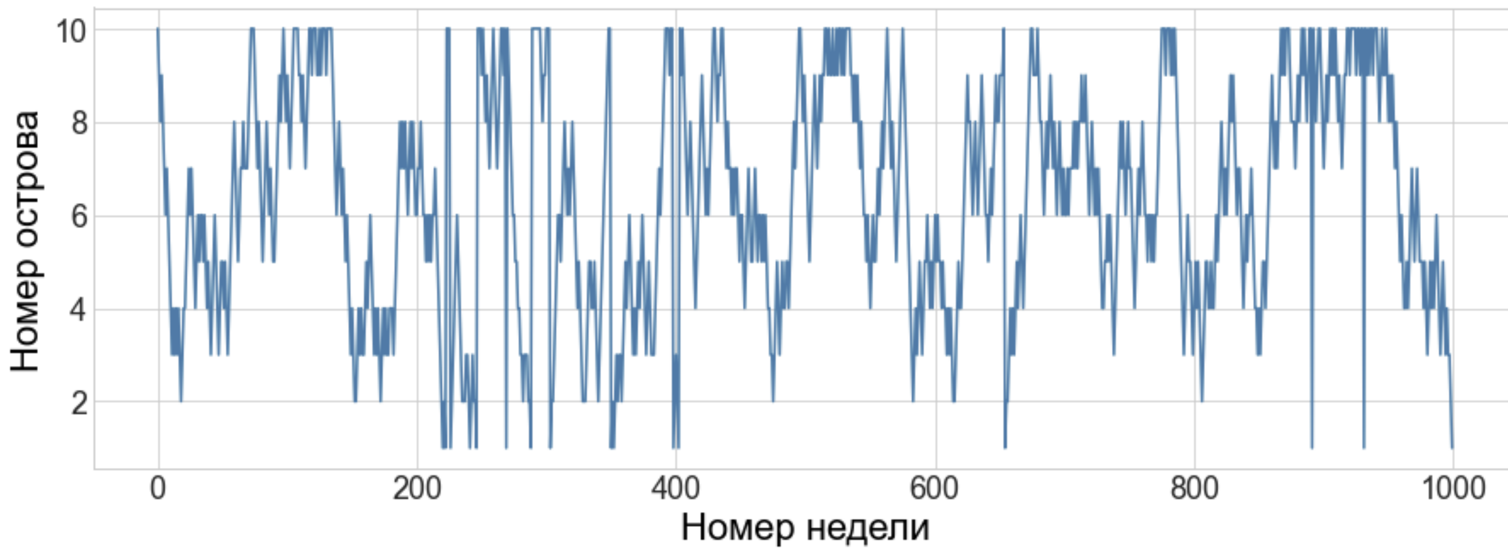
❗ Эта процедура гарантирует посещение каждого острова пропорционально его населению в долгосрочной перспективе

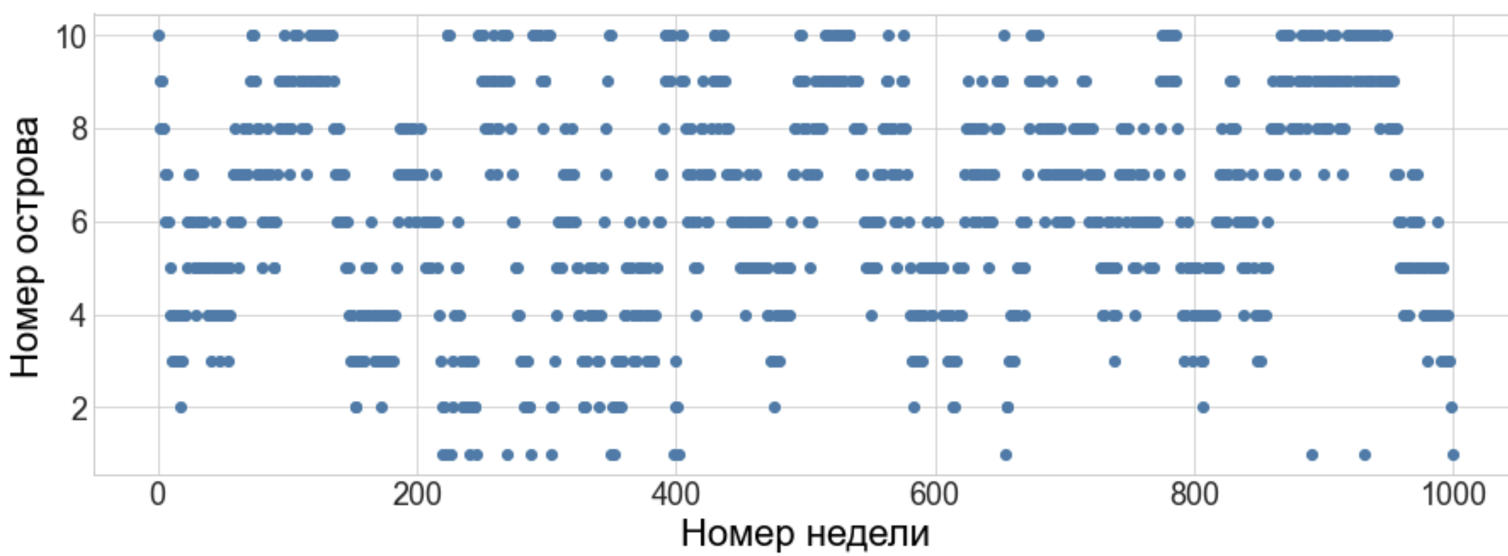
Марковская цепь из визитов короля

- Пусть у нас есть 10 островов с населением 1,2,3,...,10 человек
- Запустим алгоритм и нарисуем на картинке в какой момент на каком острове был король



Марковская цепь из визитов короля





Марковская цепь из визитов короля

спустя 200 недель

спустя 500 недель

спустя 10^5 недель

Пропорция посещений сходится к населению островов

Алгоритм Метрополиса

- Острова – возможные значения неизвестного параметра
- Число жителей – значения плотности апостериорного распределения
- Правила переезда на новый остров можно задавать по-разному ⇒ различные алгоритмы
- Главное – подобрать их так, чтобы итоговое распределение сошлось к апостериорному

Алгоритм Метрополиса

- Алгоритм Метрополиса – простейшая версия алгоритма Монте-Карло по схеме Марковской цепи (Markov chain Monte Carlo, MCMC)
- **Цепь –** последовательное извлечение значений
- **Цепь Маркова –** для текущего значения важно только предыдущее, остальная история неважна
- **Монте-Карло –** симуляция идёт случайно

Андрей Андреевич Марков

MCMC-стратегии

- Алгоритм Метрополиса – дедушка всех современных алгоритмов
- Алгоритм Метрополиса-Гастингса – более универсальный алгоритм
- Алгоритм Гиббса – более эффективная версия для простых случаев

❗ Разрабатываются новые алгоритмы

➤ *https://chi-feng.github.io/mcmc-demo/*

Алгоритмы для MCMC

- У нас есть априорные распределения и данные
- Мы хотим корректировать априорные распределения, используя данные

Алгоритмы для MCMC

Если не вдаваться в подробности, большинство алгоритмов можно описать следующим образом:

1. Начать с текущего места
2. Предложить перемещение на новое место
3. Принять/отвергнуть это новое место в зависимости от его соответствия данным и априорным распределениям
4. Если предложение принимается, перейти на новое место, перейти к шагу 1, если не принимается, сразу к шагу 1
5. После большого числа итераций вернуть все одобренные места

Сходимость

- Чтобы нащупать корректное апостериорное распределение, цепь должна прогреться: на это уходят первые несколько тысяч шагов
- Когда цепь прогрелась, апостериорное распределение сильно не изменяется
- Иногда алгоритм не сходится
- Хорошие цепи должны сходиться к одному и тому же апостериорному распределению
- Обычно алгоритм запускают несколько раз и смотрят, куда он сошёлся

mystickermania.com

Сходимость

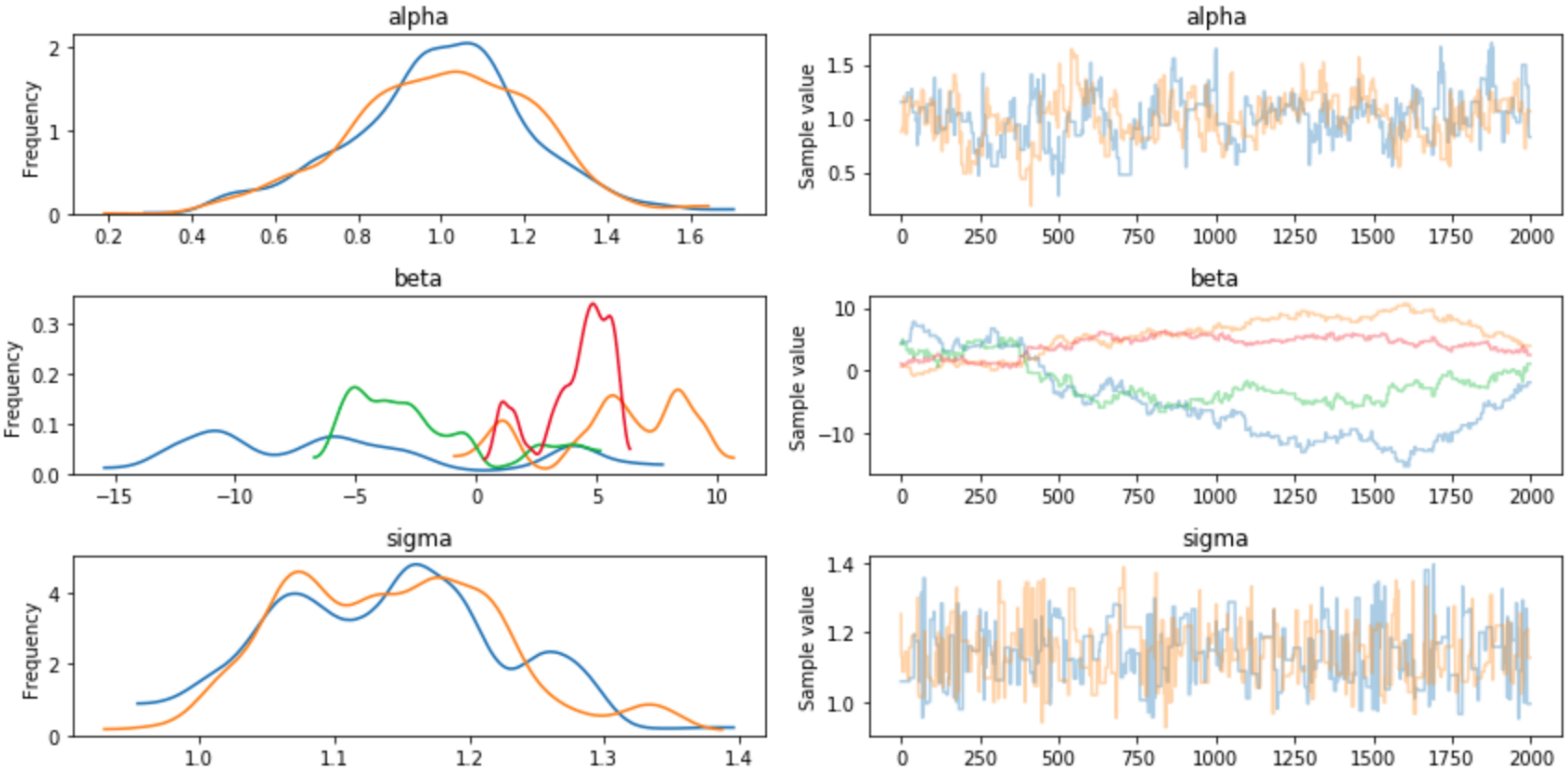
- Для анализа надо брать не всю цепь, а только ту её часть, которая получена после прогревания

- Более того, нужно чтобы выборка из апостериорного распределения была независима, поэтому цепь обычно дополнительно прореживают, чтобы не было автокорреляции

mystickermania.com

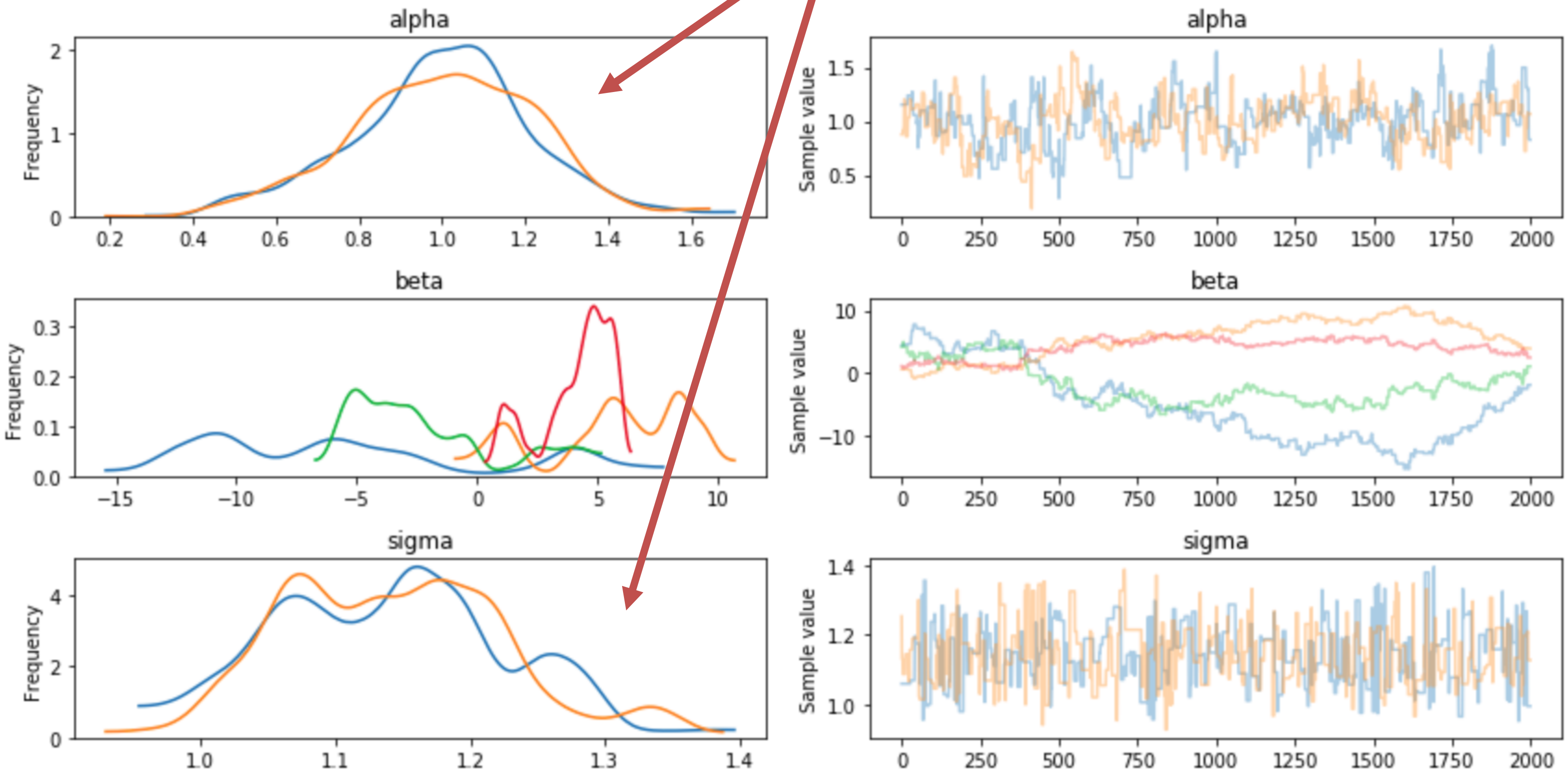
Графики трассировки

- Графики трассировки показывают часть проблем



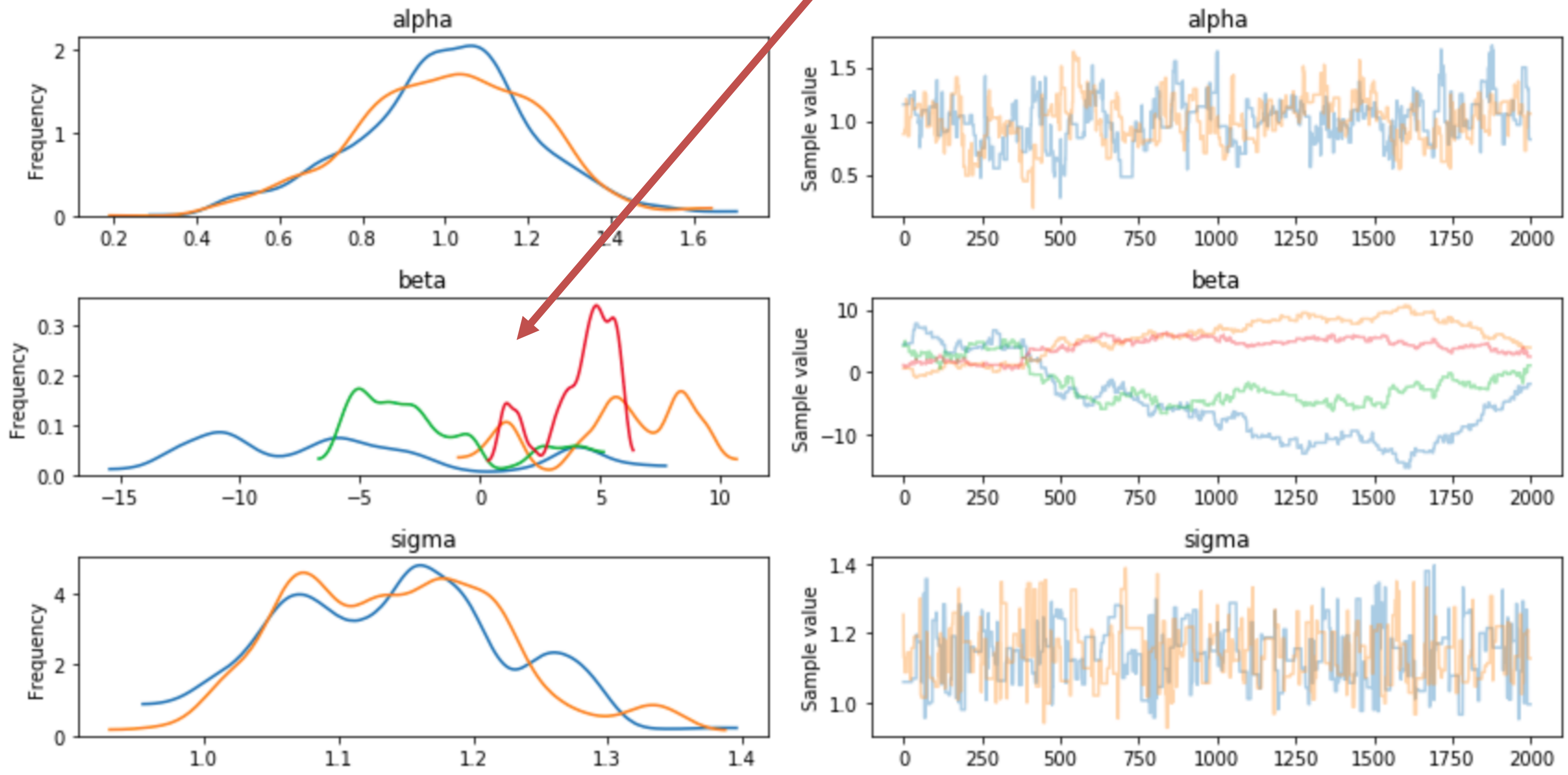
Графики трассировки

Апостериорные распределения для разных прогонов алгоритма похожие



Графики трассировки

С параметром beta есть проблемы



Графики трассировки

- Возможно, что 2000 итераций не хватает для прогревания марковской цепи
- Возможно, что есть какие-то проблемы с выборкой
- Возможно, что мы некорректно задали модель

Народная теорема статистических рассчётов:

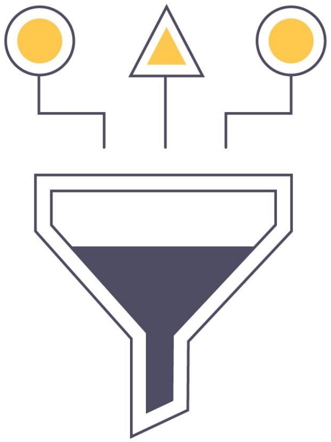
> Если вы столкнулись с вычислительными сложностями, вероятно, с вашей моделью что-то не в порядке.
>
> *(Andrew Gelman)*

Резюме

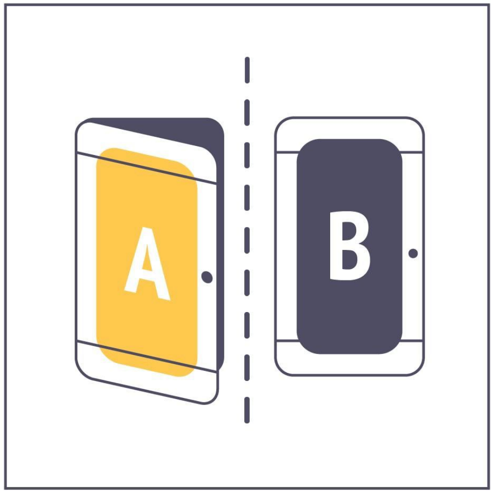
- Для произвольных априорных распределений часто сделать аналитические выкладки невозможно
- Из-за этого апостериорные распределения часто ищут с помощью численных методов, основанных на марковских цепях
- Итоговое апостериорное распределение – выборка из него

Последнее видео

Почти последний слайд



- Сбор данных
- Предобработка
- Визуализация
- ЗБЧ и ЦПТ
- Описательная статистика



- Среднее
- Схема матстата
- Доверительные интервалы
- Гипотезы
- АБ-тесты

- Максимальное правдоподобие
- Как устроен мир
- Что будет завтра
- Временные ряды
- Байесовские методы

Спасибо!

Вычитка материалов и мудрые советы:

- Лобачёва Екатерина

Помощь с материалами:

- Сергеев Дмитрий
- Демешев Борис
- Лобачёва Екатерина

Ассистенство:

- Романенко Александра
- Айрапетян Жирайр

Съёмки, их организация, монтаж и отсмотр материала:

- Капчинский Анатолий
- Савилова Ксения

Организация всех процессов:

- Застрогина Екатерина
- Сапунова Алёна

А также армии бета-тестеров ☺