ВВЕДЕНИЕ В БАЙЕСОВСКИЙ ВЫВОД

Сергей Николенко

СПбГУ — Санкт-Петербург
08 сентября 2017 г.

Random facts:

- 8 сентября 1522 г. Хуан Себастьян Элькано на корабле «Виктория» вернулся на родину; из 265 человек вернулись только 18, а самого Магеллана убили люди вождя Лапу-Лапу
- 8 сентября 1566 г. закончилась героическая оборона Сигетвара; Миклош Зриньи пал, но турки фактически были остановлены: во время осады Сигетвара умер Сулейман Великолепный
- 8 сентября 1893 г. в Чикаго завершил работу первый Конгресс еврейских женщин
- 8 сентября 1966 г. вышел пилотный выпуск сериала Star Trek

Байесовский подход

Основные определения

- Нам не понадобятся математические определения сигма-алгебры, вероятностной меры, борелевских множеств и т.п.
- Достаточно понимать, что бывают дискретные случайные величины (неотрицательные вероятности исходов в сумме дают единицу) и непрерывные случайные величины (интеграл неотрицательной функции плотности равен единице).

Основные определения

- *Совместная вероятность* – вероятность одновременного наступления двух событий, $p(x, y)$; маргинализация:

$$p(x) = \sum_y p(x, y).$$

- *Условная вероятность* – вероятность наступления одного события, если известно, что произошло другое, $p(x \mid y)$:

$$p(x, y) = p(x \mid y)p(y) = p(y \mid x)p(x).$$

- *Теорема Байеса* – из предыдущей формулы:

$$p(y|x) = \frac{p(x|y)p(y)}{p(x)} = \frac{p(x|y)p(y)}{\sum_{y'} p(x|y')p(y')}.$$

- *Независимость*: x и y независимы, если

$$p(x, y) = p(x)p(y).$$

О болезнях и вероятностях

- Приведём классический пример из классической области применения статистики — медицины.
- Пусть некий тест на какую-нибудь болезнь имеет вероятность успеха 95% (т.е. 5% — вероятность как позитивной, так и негативной ошибки).
- Всего болезнь имеется у 1% респондентов (отложим на время то, что они разного возраста и профессий).
- Пусть некий человек получил позитивный результат теста (тест говорит, что он болен). С какой вероятностью он действительно болен?

О болезнях и вероятностях

- Приведём классический пример из классической области применения статистики — медицины.
- Пусть некий тест на какую-нибудь болезнь имеет вероятность успеха 95% (т.е. 5% — вероятность как позитивной, так и негативной ошибки).
- Всего болезнь имеется у 1% респондентов (отложим на время то, что они разного возраста и профессий).
- Пусть некий человек получил позитивный результат теста (тест говорит, что он болен). С какой вероятностью он действительно болен?
- Ответ: 16%.

Доказательство

- Обозначим через t результат теста, через d — наличие болезни.
- $p(t=1) = p(t=1|d=1)p(d=1) + p(t=1|d=0)p(d=0)$.
- Используем теорему Байеса:

$$p(d=1|t=1) =$$
$$= \frac{p(t=1|d=1)p(d=1)}{p(t=1|d=1)p(d=1) + p(t=1|d=0)p(d=0)} =$$
$$= \frac{0.95 \times 0.01}{0.95 \times 0.01 + 0.05 \times 0.99} = 0.16.$$

Вывод

- Вот такие задачи составляют суть вероятностного вывода (probabilistic inference).
- Поскольку они обычно основаны на теореме Байеса, вывод часто называют байесовским (Bayesian inference).
- Но не только поэтому.

Вероятность как частота

- Обычно в классической теории вероятностей, происходящей из физики, вероятность понимается как предел отношения количества определённого результата эксперимента к общему количеству экспериментов.
- Стандартный пример: бросание монетки.

Вероятность как степень доверия

- Мы можем рассуждать о том, «насколько вероятно» то, что
    - сборная России победит на чемпионате мира по футболу в 2018 году;
    - «Одиссею» написала женщина;
    - Керенский бежал за границу в женском платье;
    - ...
- Но о «стремящемся к бесконечности количестве экспериментов» говорить бессмысленно — эксперимент здесь ровно один.

Вероятность как степень доверия

- Здесь вероятности уже выступают как *степени доверия* (degrees of belief). Это байесовский подход к вероятностям (Томас Байес так понимал).
- К счастью, и те, и другие вероятности подчиняются одним и тем же законам; есть результаты о том, что вполне естественные аксиомы вероятностной логики тут же приводят к весьма узкому классу функций (Cox, 19).

Прямые и обратные задачи

- Прямая задача: в урне лежат 10 шаров, из них 3 чёрных. Какова вероятность выбрать чёрный шар?
- Или: в урне лежат 10 шаров с номерами от 1 до 10. Какова вероятность того, что номера трёх последовательно выбранных шаров дадут в сумме 12?
- Обратная задача: перед нами две урны, в каждой по 10 шаров, но в одной 3 чёрных, а в другой — 6. Кто-то взял из какой-то урны шар, и он оказался чёрным. Насколько вероятно, что он брал шар из первой урны?
- Заметьте, что в обратной задаче вероятности сразу стали байесовскими (хоть здесь и можно переформулировать через частоты).

Прямые и обратные задачи

- Иначе говоря, прямые задачи теории вероятностей описывают некий вероятностный процесс или модель и просят подсчитать ту или иную вероятность (т.е. фактически по модели предсказать поведение).
- Обратные задачи содержат *скрытые переменные* (в примере — номер урны, из которой брали шар). Они часто просят по известному поведению построить вероятностную модель.
- Задачи машинного обучения обычно являются задачами второй категории.

Определения

- Запишем теорему Байеса:

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

- Здесь $p(\theta)$ — *априорная вероятность* (prior probability), $p(D|\theta)$ — *правдоподобие* (likelihood), $p(\theta|D)$ — *апостериорная вероятность* (posterior probability), $p(D) = \int p(D \mid \theta)p(\theta)\mathrm{d}\theta$ — вероятность данных (evidence).
- Вообще, *функция правдоподобия* имеет вид

$$a \mapsto p(y|x = a)$$

  для некоторой случайной величины y.

ML vs. MAP

- В статистике обычно ищут *гипотезу максимального правдоподобия* (maximum likelihood):

$$\theta_{ML} = \arg\max{}_\theta p(D \mid \theta).$$

- В байесовском подходе ищут *апостериорное распределение* (posterior)

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

  и, возможно, *максимальную апостериорную гипотезу* (maximum a posteriori):

$$\theta_{MAP} = \arg\max{}_\theta p(\theta \mid D) = \arg\max{}_\theta p(D \mid \theta)p(\theta).$$

Постановка задачи

- Простая задача вывода: дана нечестная монетка, она подброшена N раз, имеется последовательность результатов падения монетки. Надо определить её «нечестность» и предсказать, чем она выпадет в следующий раз.
- Гипотеза максимального правдоподобия скажет, что вероятность решки равна числу выпавших решек, делённому на число экспериментов.

Постановка задачи

- Простая задача вывода: дана нечестная монетка, она подброшена N раз, имеется последовательность результатов падения монетки. Надо определить её «нечестность» и предсказать, чем она выпадет в следующий раз.
- Гипотеза максимального правдоподобия скажет, что вероятность решки равна числу выпавших решек, делённому на число экспериментов.
- То есть если вы взяли незнакомую монетку, подбросили её один раз и она выпала решкой, вы теперь ожидаете, что она всегда будет выпадать только решкой, правильно?
- Странно получается... давайте поговорим об этом поподробнее позже.

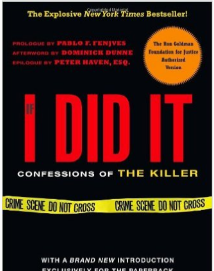
Упражнения

1. У моего знакомого два ребёнка. Будем предполагать, что пол ребёнка выбирается независимо и равновероятно, с вероятностью $\frac{1}{2}$. Две постановки вопроса:
   (1) я спросил, есть ли у него мальчики, и он ответил «да»; какова вероятность того, что один из детей – девочка?
   (2) я встретил одного из его детей, и это мальчик; какова вероятность того, что второй ребёнок – девочка?

2. Произошло убийство. На месте убийства найдена кровь, которая явно принадлежит убийце. Кровь принадлежит редкой группе, которая присутствует у 1% населения, в том числе у подсудимого.
   (1) Прокурор говорит: «Шанс, что у подсудимого была бы именно такая группа крови, если бы он был невиновен – всего 1%; значит, с вероятностью 99% он виновен». В чём не прав прокурор?
   (2) Адвокат говорит: «В городе живёт миллион человек, то есть у 10000 из них такая группа крови. Значит, всё, что говорит нам эта кровь – это что подсудимый совершил убийство с вероятностью 0.01%; никакое это не доказательство». В чём не прав адвокат?

Упражнения

3. Реальные случаи.
   (1) Прокурор указал, что O.J. Simpson уже бил жену в прошлом. Адвокат ответил: «Убивают только одну из 2500 женщин, подвергавшихся семейному насилию, так что это вообще нерелевантно». Суд согласился с адвокатом; верно ли это рассуждение?
   (2) У Sally Clark погибли два младенца; прокурор указал, что вероятность двух случаев SIDS в одной семье, которую он получил из статистики одиночных случаев, — около 1 из 73 миллионов; в чём он не прав?



Байесовский вывод для монетки

ML vs. MAP

- Мы остановились на том, что в статистике обычно ищут *гипотезу максимального правдоподобия* (maximum likelihood):
$$\theta_{ML} = \arg\max_{\theta} p(D \mid \theta).$$
- В байесовском подходе ищут *апостериорное распределение* (posterior)
$$p(\theta|D) \propto p(D|\theta)p(\theta)$$
и, возможно, *максимальную апостериорную гипотезу* (maximum a posteriori):
$$\theta_{MAP} = \arg\max_{\theta} p(\theta \mid D) = \arg\max_{\theta} p(D \mid \theta)p(\theta).$$

Постановка задачи

- Простая задача вывода: дана нечестная монетка, она подброшена N раз, имеется последовательность результатов падения монетки. Надо определить её «нечестность» и предсказать, чем она выпадет в следующий раз.

Постановка задачи

- Если у нас есть вероятность p_h того, что монетка выпадет решкой (вероятность орла $p_t = 1 - p_h$), то вероятность того, что выпадет последовательность s, которая содержит n_h решек и n_t орлов, равна

$$p(s|p_h) = p_h^{n_h}(1 - p_h)^{n_t}.$$

- Сделаем предположение: будем считать, что монетка выпадает равномерно, т.е. у нас нет априорного знания p_h.
- Теперь нужно использовать теорему Байеса и вычислить скрытые параметры.

Пример применения теоремы Байеса

- Правдоподобие: $p(p_h|s) = \frac{p(s|p_h)p(p_h)}{p(s)}$.
- Здесь $p(p_h)$ следует понимать как непрерывную случайную величину, сосредоточенную на интервале $[0, 1]$, коей она и является. Наше предположение о равномерном распределении в данном случае значит, что априорная вероятность $p(p_h) = 1$, $p_h \in [0, 1]$ (т.е. априори мы не знаем, насколько нечестна монетка, и предполагаем это равновероятным). А $p(s|p_h)$ мы уже знаем.
- Итого получается:

$$p(p_h|s) = \frac{p_h^{n_h}(1 - p_h)^{n_t}}{p(s)}.$$

Пример применения теоремы Байеса

- Итого получается:

$$p(p_h|s) = \frac{p_h^{n_h}(1-p_h)^{n_t}}{p(s)}.$$

- $p(s)$ можно подсчитать как

$$p(s) = \int_0^1 p_h^{n_h}(1-p_h)^{n_t} dp_h =$$

$$= \frac{\Gamma(n_h+1)\Gamma(n_t+1)}{\Gamma(n_h+n_t+2)} = \frac{n_h! n_t!}{(n_h+n_t+1)!},$$

но найти $\arg\max_{p_h} p(p_h \mid s) = \frac{n_h}{n_h+n_t}$ можно и без этого.

- Итого получается:

$$p(p_h|s) = \frac{p_h^{n_h}(1-p_h)^{n_t}}{p(s)}.$$

- Но это ещё не всё. Чтобы предсказать следующий исход, надо найти $p(\text{heads}|s)$:

$$p(\text{heads}|s) = \int_0^1 p(\text{heads}|p_h) p(p_h|s) dp_h =$$
$$= \int_0^1 \frac{p_h^{n_h+1}(1-p_h)^{n_t}}{p(s)} dp_h =$$
$$= \frac{(n_h+1)!n_t!}{(n_h+n_t+2)!} \cdot \frac{(n_h+n_t+1)!}{n_h!n_t!} = \frac{n_h+1}{n_h+n_t+2}.$$

- Получили правило Лапласа.

Пример применения теоремы Байеса

- Итого получается:

$$p(p_h|s) = \frac{p_h^{n_h}(1-p_h)^{n_t}}{p(s)}.$$

- Это была иллюстрация двух основных задач байесовского вывода:
    1. найти апостериорное распределение на гипотезах/параметрах:

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

    (и/или найти гипотезу максимального правдоподобия $\arg\max_\theta p(\theta \mid D)$);
    2. найти апостериорное распределение исходов дальнейших экспериментов:

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

Напоминание

- Напоминаю, что основная наша задача – как обучить параметры распределения и/или предсказать следующие его точки по имеющимся данным.
- В байесовском выводе участвуют:
    - $p(x \mid \theta)$ – правдоподобие данных;
    - $p(\theta)$ – априорное распределение;
    - $p(x) = \int_\Theta p(x \mid \theta) p(\theta) d\theta$ – маргинальное правдоподобие;
    - $p(\theta \mid x) = \frac{p(x|\theta)p(\theta)}{p(x)}$ – апостериорное распределение;
    - $p(x' \mid x) = \int_\Theta p(x' \mid \theta) p(\theta \mid x) d\theta$ – предсказание нового x'.
- Задача обычно в том, чтобы найти $p(\theta \mid x)$ и/или $p(x' \mid x)$.

- Когда мы проводим байесовский вывод, у нас, кроме правдоподобия, должно быть ещё *априорное распределение* (prior distribution) по всем возможным значениям параметров.
- Мы раньше к ним специально не присматривались, но они очень важны.
- Задача байесовского вывода – как подсчитать $p(\theta \mid x)$ и/или $p(x' \mid x)$.
- Но чтобы это сделать, сначала надо выбрать $p(\theta)$. Как выбирать априорные распределения?

Сопряжённые априорные распределения

- Разумная цель: давайте будем выбирать распределения так, чтобы они оставались такими же и *a posteriori*.
- До начала вывода есть априорное распределение $p(\theta)$.
- После него есть какое-то новое апостериорное распределение $p(\theta \mid x)$.
- Я хочу, чтобы $p(\theta \mid x)$ тоже имело тот же вид, что и $p(\theta)$, просто с другими параметрами.

Сопряжённые априорные распределения

- Не слишком формальное определение: семейство распределений $p(\theta \mid \alpha)$ называется семейством *сопряжённых априорных распределений* для семейства правдоподобий $p(x \mid \theta)$, если после умножения на правдоподобие апостериорное распределение $p(\theta \mid x, \alpha)$ остаётся в том же семействе: $p(\theta \mid x, \alpha) = p(\theta \mid \alpha')$.
- α называются *гиперпараметрами* (hyperparameters), это «параметры распределения параметров».
- Тривиальный пример: семейство всех распределений будет сопряжённым чему угодно, но это не очень интересно.

Сопряжённые априорные распределения

- Разумеется, вид хорошего априорного распределения зависит от вида распределения собственно данных, $p(x \mid \theta)$.
- Сопряжённые априорные распределения подсчитаны для многих распределений, мы приведём несколько примеров.

Испытания Бернулли

- Каким будет сопряжённое априорное распределение для бросания нечестной монетки (испытаний Бернулли)?
- Ответ: это будет бета-распределение; плотность распределения нечестности монетки θ

$$p(\theta \mid \alpha, \beta) = \frac{\theta^{\alpha-1}(1-\theta)^{\beta-1}}{B(\alpha, \beta)}.$$

Испытания Бернулли

- Плотность распределения нечестности монетки θ

$$p(\theta \mid \alpha, \beta) = \frac{\theta^{\alpha-1}(1-\theta)^{\beta-1}}{B(\alpha, \beta)}.$$

- Тогда, если мы посэмплируем монетку, получив s орлов и f решек, получится

$$p(s, f \mid \theta) = \binom{s+f}{s}\theta^s(1-\theta)^f, \text{ и}$$

$$p(\theta|s, f) = \frac{\binom{s+f}{s}\theta^{s+\alpha-1}(1-\theta)^{f+\beta-1}/B(\alpha, \beta)}{\int_0^1 \binom{s+f}{s}x^{s+\alpha-1}(1-x)^{f+\beta-1}/B(\alpha, \beta)dx} = \\ = \frac{\theta^{s+\alpha-1}(1-\theta)^{f+\beta-1}}{B(s+\alpha, f+\beta)}.$$

Испытания Бернулли

- Итого получается, что сопряжённое априорное распределение для параметра нечестной монетки θ – это

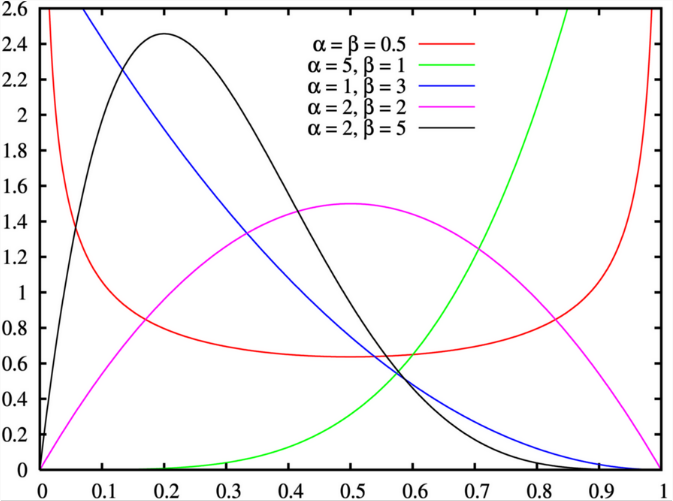
$$p(\theta \mid \alpha, \beta) \propto \theta^{\alpha-1}(1-\theta)^{\beta-1}.$$

- После получения новых данных с s орлами и f решками гиперпараметры меняются на

$$p(\theta \mid s+\alpha, f+\beta) \propto \theta^{s+\alpha-1}(1-\theta)^{f+\beta-1}.$$

- На этом этапе можно забыть про сложные формулы и выводы, получилось очень простое правило обучения (под обучением теперь понимается изменение гиперпараметров).

Бета--распределение



Мультиномиальное распределение

- Простое обобщение: рассмотрим мультиномиальное распределение с n испытаниями, k категориями и по x_i экспериментов дали категорию i.
- Параметры θ_i показывают вероятность попасть в категорию i:

$$p(x \mid \theta) = \binom{n}{x_1, \dots, x_n} \theta_1^{x_1} \theta_2^{x_2} \dots \theta_k^{x_k}.$$

- Сопряжённым априорным распределением будет распределение Дирихле:

$$p(\theta \mid \alpha) \propto \theta_1^{\alpha_1 - 1} \theta_2^{\alpha_2 - 1} \dots \theta_k^{\alpha_k - 1}.$$

Мультиномиальное распределение

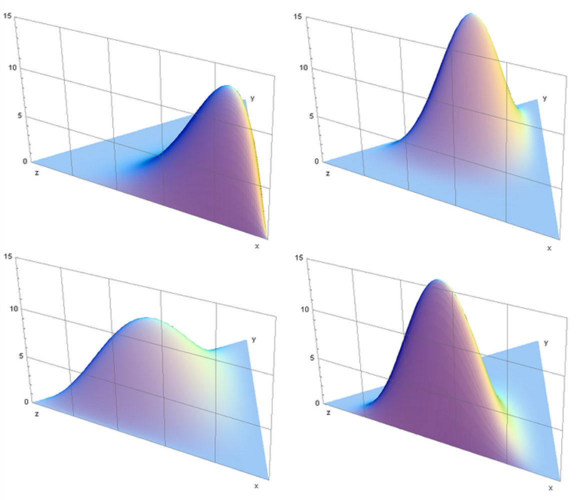
- Сопряжённым априорным распределением будет распределение Дирихле:

$$p(\theta \mid \alpha) \propto \theta_1^{\alpha_1 - 1} \theta_2^{\alpha_2 - 1} \dots \theta_k^{\alpha_k - 1}.$$

Упражнение. Докажите, что при получении данных $x_1, \dots, x_k$ гиперпараметры изменятся на

$$p(\theta \mid x, \alpha) = p(\theta \mid x + \alpha) \propto \theta_1^{x_1 + \alpha_1 - 1} \theta_2^{x_2 + \alpha_2 - 1} \dots \theta_k^{x_k + \alpha_k - 1}.$$

Распределение Дирихле



Важные распределения

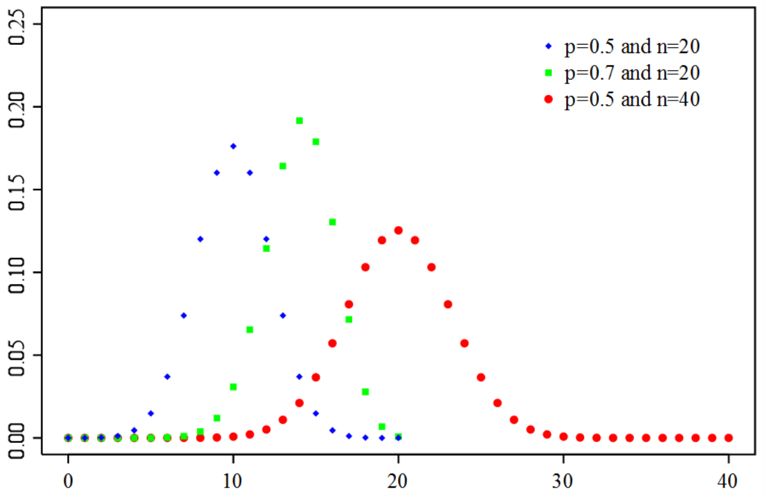
- Есть некоторое количество распределений, которые часто появляются в разных задачах и являются наиболее полезными на практике.
- Их полезно… ну, если не знать, то хотя бы быть знакомыми.
- Сейчас по ним и пробежимся.

Биномиальное распределение

- *Биномиальное распределение* возникает, когда мы подбрасываем нечестную монетку (вероятность решки p) n раз и хотим найти вероятность появления r решек.

$$p(r|p,n) = \binom{n}{r} p^r (1-p)^{n-r}.$$

Биномиальное распределение



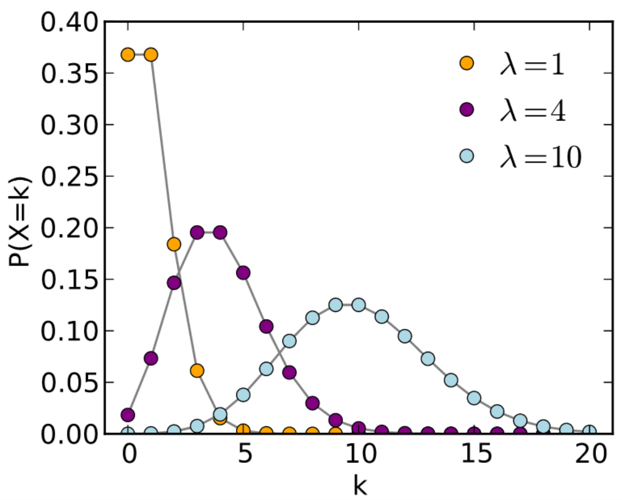
Пуассоновское распределение

- *Распределение Пуассона* возникает, когда мы хотим подсчитать количество событий за фиксированный интервал, если нам дана средняя интенсивность этих событий.
- Если ожидается в среднем λ событий за этот интервал, то вероятность того, что произойдут ровно r событий, равна

$$p(r|\lambda) = e^{-\lambda}\frac{\lambda^r}{r!}.$$

- Распределение Пуассона – это предельный случай биномиального распределения. Если n очень велико, а p очень мало, $\mathrm{Binomial}(n, p)$ будет очень похоже на $\mathrm{Poisson}(np)$.

Пуассоновское распределение



Нормальное распределение

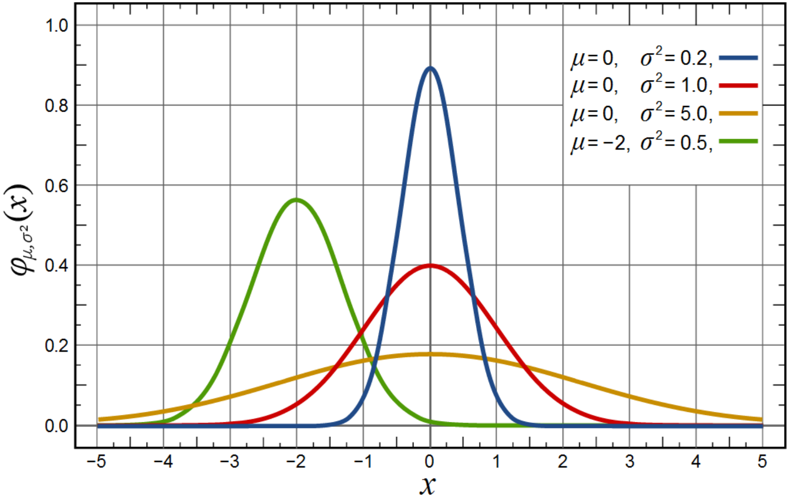
- Мы уже давно знаем нормальное распределение:

$$p(x|\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

- Очень многие процессы могут моделироваться нормальным (гауссовским) распределением; обычно возникает, когда есть некое среднее значение μ и шум вокруг него.
- Функция правдоподобия данных $x_1, \ldots, x_n$:

$$p(x_1, \ldots, x_n|\mu,\sigma) = \frac{1}{(\sigma\sqrt{2\pi})^n} e^{-\sum_{i=1}^{n} \frac{(x_i-\mu)^2}{2\sigma^2}}.$$

Нормальное распределение



Гауссиан: достаточные статистики

- Заметим, что функция эта зависит от двух параметров, а не от n:
$$p(x_1, \ldots, x_n | \mu, \sigma) = \frac{1}{(\sigma\sqrt{2\pi})^n} e^{-\frac{S + n(\bar{x}-\mu)^2}{2\sigma^2}},$$
где
$$\bar{x} = \frac{x_1 + \ldots + x_n}{n}, \quad S = \sum_{i=1}^{n} (\bar{x} - x_n)^2.$$
- Параметры $\bar{x}$ и S называются *достаточными статистиками* (sufficient statistics).

Гауссиан: ГМП

- Какие параметры лучше всего описывают данные?
- Перейдём, как водится, к логарифму:

$$\ln p(x_1, \dots, x_n | \mu, \sigma) = -n \ln(\sigma \sqrt{2\pi}) - \frac{S + n(\bar{x} - \mu)^2}{2\sigma^2}.$$

- Как выяснить, при каких параметрах функция правдоподобия максимизируется?

Гауссиан: ГМП

- Какие параметры лучше всего описывают данные?
- Перейдём, как водится, к логарифму:

$$\ln p(x_1, \ldots, x_n|\mu, \sigma) = -n \ln(\sigma\sqrt{2\pi}) - \frac{S + n(\bar{x} - \mu)^2}{2\sigma^2}.$$

- Как выяснить, при каких параметрах функция правдоподобия максимизируется?
- Взять частные производные и приравнять нулю.

Гауссиан: ГМП

- По μ:
$$\frac{\partial \ln p}{\partial \mu} = -\frac{n}{\sigma^2}(\mu - \bar{x}).$$
- То есть в гипотезе максимального правдоподобия $\mu_{ML} = \bar{x}$, независимо от S.
- Теперь нужно найти σ из гипотезы максимального правдоподобия.
- Для этого мы продифференцируем по $\ln \sigma$ — полезный приём на будущее. Кстати, $\frac{\mathrm{d}x^n}{\mathrm{d}(\ln x)} = nx^n$.

- $$\frac{\partial \ln p}{\partial \ln \sigma} = -n + \frac{\sum_{i=1}^{n}(x_i - \mu)^2}{\sigma^2}.$$
- Следовательно, в гипотезе максимального правдоподобия

$$\sigma_{ML} = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \mu)^2}{n}}.$$

Упражнение. Докажите, что это смещённая оценка, т.е. ожидание этой оценки по настоящему нормальному распределению не равно σ^2.

Распределение Стьюдента

- Распределение Стьюдента применяется, когда нужно искать доверительные интервалы на параметры нормального распределения.
- Величина $T = \frac{\bar{x}_n - \mu}{S_n/\sqrt{n}}$ распределена по закону

$$f(t) = \frac{\Gamma(n/2)}{\sqrt{\pi(n-1)}\Gamma(\frac{n-1}{2})} \left(1 + \frac{t^2}{n-1}\right)^{-n/2}.$$

- Если мы выберем число A так, чтобы $p(-A < T < A) = 1 - \alpha$, то

$$\left[\bar{x}_n - A\frac{S_n}{\sqrt{n}}, \bar{x}_n + A\frac{S_n}{\sqrt{n}}\right]$$

  будет доверительным интервалом для μ с вероятностью ошибки α.

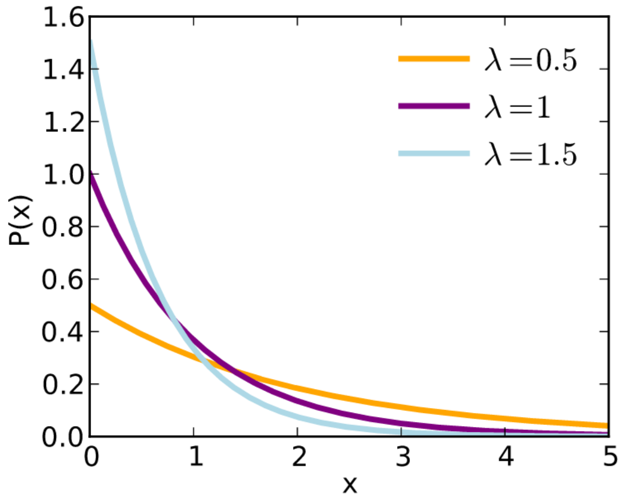
- Обычно возникает в ответах на вопрос «сколько надо ждать события».
- Дискретный вариант: вероятность того, что выпадения орла на нечестной монетке придётся ждать ровно r шагов, равна

$$p(r|p) = p^r(1-p) = (1-p)e^{-\lambda r}, \text{ где } \lambda = \ln\frac{1}{p}.$$

- Непрерывный вариант: если мы ждём события, которое происходит в среднем каждые $1/\lambda$ единиц времени (в пуассоновском процессе с интенсивностью λ), то распределение времени ожидания

$$p(x|s) = \lambda e^{-\lambda x}.$$

Экспоненциальное распределение

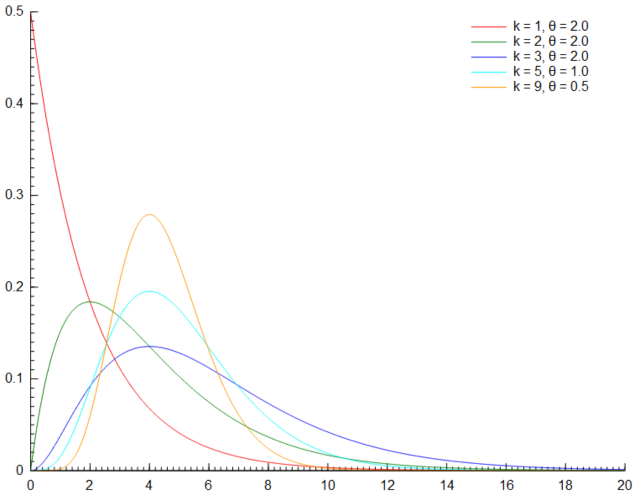


Гамма--распределение

- Может возникать как сумма нескольких экспоненциальных распределений.
- Плотность распределения

$$p(x|k, \theta) = x^{k-1} \frac{e^{-x/\theta}}{\theta^k \Gamma(k)}, \quad x > 0.$$

Гамма-распределение

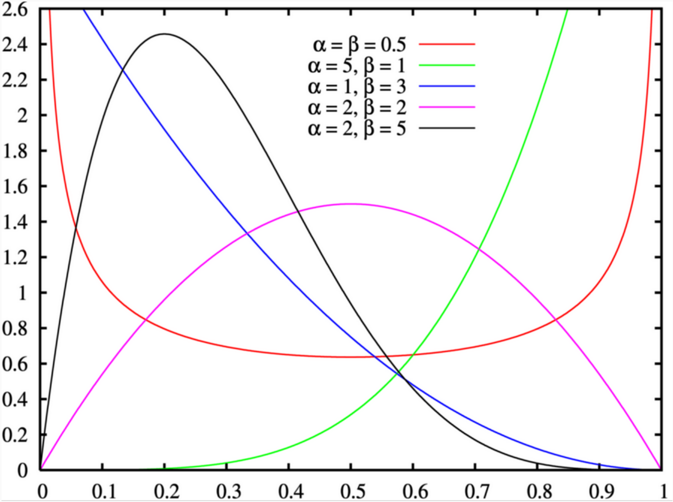


- Бета-распределение определяется над вероятностями, т.е. на интервале $[0, 1]$.
- Часто служит априорным распределением для каких-либо вероятностей; является сопряжённым априорным распределением для испытаний Бернулли:

$$\mathrm{Beta}(x|\alpha, \beta) = \frac{1}{B(\alpha, \beta)} x^{\alpha-1}(1-x)^{\beta-1}, \quad \beta(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}.$$

- $\mathrm{Beta}(i, j)$ — это распределение i-й по величине случайной величины из $i + j - 1$ случайных величин, распределённых равномерно на $[0, 1]$.

Бета--распределение



- Бета–распределение — частный случай *порядковой статистики* (order statistics).
- Если n случайных величин $X_1, \dots, X_n$ распределены одинаково с функцией распределения F, то k-я порядковая статистика $Y_k^{(n)}(y)$ — это функция распределения k-й сверху величины из этих n величин.
- Например, очевидно, что

$$Y_1^{(n)}(y) = F(y)^{n-1}.$$

Упражнение. Вывести формулу для второй порядковой статистики $Y_2^{(n)}$.

Распределение Дирихле

- Обобщение бета–распределения на многомерный случай.
- На k-мерном симплексе $\{x \mid \sum_{i=1}^{k} x_i = 1\}$ плотность распределения Дирихле с параметром $\alpha = (\alpha_1, \ldots, \alpha_k)$ равна

$$p(x|\alpha) = \frac{1}{B(\alpha)} \prod_{i=1}^{k} x_i^{\alpha_i - 1}, \quad B(\alpha) = \frac{\prod_{i=1}^{k} \Gamma(\alpha_i)}{\Gamma\left(\sum_{i=1}^{k} \alpha_i\right)}.$$

Спасибо!

Спасибо за внимание!