

НЕМЕЦКИЕ ТАНКИ И ПРИМЕР ДЖЕЙНСА

Сергей Николенко

СПбГУ — Санкт-Петербург

14 ноября 2024 г.

Random facts:

- 14 ноября в ООН — Всемирный день борьбы с диабетом; дата выбрана потому, что 14 ноября 1891 г. родился Фредерик Бантинг, один из открывателей инсулина
- 14 ноября 332 г. до н. э. Александр Великий был провозглашён фараоном Египта, сыном Амона
- 14 ноября 1901 г. венский врач Карл Ландштейнер разделил все образцы крови на три группы: А, В и 0
- 14 ноября 1925 г. в Париже открылась выставка искусства сюрреалистов, включающая работы Макса Эрнста, Мана Рэя, Хуана Миро и Пабло Пикассо
- 14 ноября 1916 г. Павел Милюков на открытии осенней сессии Думы произнёс знаменитое: «Что это? Глупость или измена?»
- 14 ноября 1952 г. английская газета «New Musical Express» впервые опубликовала хит-парад наиболее продаваемых синглов
- 14 ноября 1981 г. Гамбия и Сенегал создали федеративное государство Сенегамбия; оно просуществовало до 1989 года



ОБУЧЕНИЕ РАВНОМЕРНОГО РАСПРЕДЕЛЕНИЯ

КАК ПОДСЧИТАТЬ НЕМЕЦКИЕ ТАНКИ

- Как во время второй мировой понять, сколько танков произвела Германия в этом году?
- Есть традиционные средства разведки: оценить число немецких танков там, где мы смогли их увидеть, и как-то экстраполировать на всё остальное
- Но нашёлся и другой неожиданный путь: предположим, что вы захватили немецкий танк (например, вывезли подбитый с поля боя); оказалось, что на разных деталях танка дотошные немецкие производители... наносят серийные номера!
- Мы можем посмотреть на, условно говоря, коленвал очередной «Пантеры» и узнать, что это коленвал «Пантеры» номер 273
- Что это нам скажет?

КАК ПОДСЧИТАТЬ НЕМЕЦКИЕ ТАНКИ

- Задача кажется несложной: даны несколько чисел, взятых из равномерного распределения на интервале $1, \dots, N$, и требуется оценить N
- Начнём с самого простого случая, когда число только одно; мы захватили танк с серийным номером m и хотим оценить, сколько всего их могло бы быть
- Функция правдоподобия:

$$p(m|N) = \begin{cases} \frac{1}{N}, & \text{если } N \geq m, \\ 0, & \text{если } N < m \end{cases}$$

- Гипотеза максимального правдоподобия — это $N = m$, тогда правдоподобие исхода m (и всех остальных) равно $\frac{1}{m}$
- Странный вывод, да? Это смещённая оценка, а нам бы несмещённую...

КАК ПОДСЧИТАТЬ НЕМЕЦКИЕ ТАНКИ

- Ответ из статистики: найти распределение максимума M из выборки, то есть оценить $\mathbb{E}[M \mid N]$, а затем подставить в качестве этого ожидания эмпирический максимум $m = \max_{i=1}^k x_i$ и решить уравнение относительно N
- Для $k = 1$ (одно число в выборке) это совсем просто: ожидание максимума совпадает с ожиданием одного x_i и равно $\frac{N+1}{2}$, а значит,

$$\hat{N} = 2m - 1$$

- Это звучит вполне интуитивно: мы ожидаем, что равномерно взятое случайное число в среднем попадёт в середину интервала, значит, мы ожидаем, что сам интервал будет примерно вдвое шире этого числа

КАК ПОДСЧИТАТЬ НЕМЕЦКИЕ ТАНКИ

- Для $k > 1$ придётся заняться комбинаторикой; распределение $p(M|N)$ равно

$$p(M = m|N) = \frac{\text{число выборов из } k \text{ чисел с максимумом } m}{\text{общее число выборов из } k \text{ чисел}}$$

- В знаменателе число сочетаний из N по k , $\binom{N}{k}$, а в числителе, соответственно, число сочетаний из $m - 1$ по $k - 1$: мы уже зафиксировали одно из чисел m , и это максимум, то есть остальные $k - 1$ число должны быть взяты от 1 до $m - 1$
- Итого получается, что

$$p(M = m|N) = \frac{\binom{m-1}{k-1}}{\binom{N}{k}}.$$

КАК ПОДСЧИТАТЬ НЕМЕЦКИЕ ТАНКИ

- А ожидание максимума, значит, равно

$$\begin{aligned}\mathbb{E}[M | N] &= \sum_{m=k}^N mp(M = m|N) = \sum_{m=k}^N m \frac{\binom{m-1}{k-1}}{\binom{N}{k-1}} = \frac{1}{\binom{N}{k-1}} \sum_{m=k}^N m \binom{m-1}{k-1} = \\ &= \frac{1}{\binom{N}{k-1}} \sum_{m=k}^N m \frac{(m-1)!}{(k-1)!(m-k)!} = \frac{k}{\binom{N}{k-1}} \sum_{m=k}^N \frac{m!}{k!(m-k)!} = \frac{k}{\binom{N}{k-1}} \sum_{m=k}^N \binom{m}{k-1};\end{aligned}$$

в предпоследнем равенстве мы умножили и разделили на k

- Воспользуемся известным тождеством о биномиальных коэффициентах: $\sum_{a=c}^b \binom{a}{c} = \binom{b+1}{c+1}$; его легко доказать индукцией по b : для $b = c$ слева и справа единица, а если мы уже доказали тождество для некоторого b , то индукционный переход выглядит так:

$$\sum_{a=c}^{b+1} \binom{a}{c} = \sum_{a=c}^b \binom{a}{c} + \binom{b+1}{c} = \binom{b+1}{c+1} + \binom{b+1}{c} = \binom{b+2}{c+1};$$

последнее равенство — свойство треугольника Паскаля

КАК ПОДСЧИТАТЬ НЕМЕЦКИЕ ТАНКИ

- Применяя доказанное тождество, получим, что

$$\begin{aligned}\mathbb{E}[M \mid N] &= \frac{k}{\binom{N}{k}} \binom{N+1}{k+1} = \\ &= k \frac{k!(N-k)!}{N!} \frac{(N+1)!}{(k+1)!(N-k)!} = (N+1) \frac{k}{k+1}.\end{aligned}$$

- Получилась очень простая формула, которую легко теперь и обратить:

$$\hat{N} = m \left(1 + \frac{1}{k}\right) - 1 = m + \left(\frac{m}{k} - 1\right).$$

- Вполне интуитивный результат: к максимуму нужно добавить «бонус», который уменьшается с ростом k

КАК ПОДСЧИТАТЬ НЕМЕЦКИЕ ТАНКИ

- А как по-байесовски? Мы хотим найти

$$p(N|M, k) = \frac{p(M|N, k) p(N|k)}{p(M|k)}.$$

- Мы уже подсчитали, что $p(M = m|N) = \binom{m-1}{k-1} / \binom{N}{k}$;
знаменатель будем считать как нормировочную константу

$$p(M|k) = \sum_{N=M}^{\infty} p(M|N, k) p(N|k).$$

- Теперь нужно выбрать априорное распределение $p(N|k)$:
какое взять? Мы бы хотели отразить в нём отсутствие информации о значении N , но равномерное распределение, как для параметра монетки, выбрать не получится, ведь нам нужно распределение на всех натуральных числах

КАК ПОДСЧИТАТЬ НЕМЕЦКИЕ ТАНКИ

- Давайте возьмём равномерное априорное распределение $p(N|k)$ на большом объемлющем интервале $k, \dots, \Omega$ (меньше k номеров при k наблюдениях получиться не может):

$$p(N|k) = \frac{1}{\Omega - k + 1} \quad \text{для всех } k \leq N, N \leq \Omega.$$

- Соответственно, получается, что для $k \leq N$ и $N < \Omega$

$$\begin{aligned} p(N|M, k) &= \frac{p(M|N, k) p(N|k)}{\sum_{n=M}^{\Omega} p(M|n, k) p(n|k)} = \frac{\frac{1}{\Omega-k+1} p(M|N, k)}{\sum_{n=M}^{\Omega} \frac{1}{\Omega-k+1} p(M|n, k)} = \\ &= \frac{\binom{M-1}{k-1} / \binom{N}{k}}{\sum_{n=M}^{\Omega} \binom{M-1}{k-1} / \binom{n}{k}} = \frac{\binom{N}{k}^{-1}}{\sum_{n=M}^{\Omega} \binom{n}{k}^{-1}}, \end{aligned}$$

а для других значений N эта вероятность равна нулю.

КАК ПОДСЧИТАТЬ НЕМЕЦКИЕ ТАНКИ

- Как выбрать верхний предел Ω ? Можно просто заведомо большой, но технически препятствий к тому, чтобы положить $\Omega = \infty$, в формулах нет, ряд $\sum_{n=M}^{\infty} \binom{n}{k}^{-1}$ сходится для $k \geq 2$
- Это называется *некорректным априорным распределением* (improper prior): распределение «расходится», но после умножения на любое правдоподобие начинает сходиться, и байесовский вывод можно вести как обычно
- Например, бета-распределение для монетки без влияния на правдоподобие, то есть $B(0, 0)$:

$$p(\theta|0, 0) \propto \frac{1}{\theta} \frac{1}{1-\theta} \quad \text{для } \theta \in (0, 1).$$

КАК ПОДСЧИТАТЬ НЕМЕЦКИЕ ТАНКИ

- А для танков, применив некорректное априорное распределение с $\Omega = \infty$, приходим к апостериорному распределению

$$p(N|M, k) = \frac{\binom{N}{k}^{-1}}{\sum_{n=M}^{\infty} \binom{n}{k}^{-1}}.$$

- Ряд просуммировать трудно, давайте просто поверим, что

$$\sum_{n=M}^{\infty} \binom{n}{k}^{-1} = \binom{M}{k}^{-1} \frac{M}{k-1}, \quad \text{а значит,}$$

$$p(N|M, k) = \frac{k-1}{M} \binom{M}{k} \binom{N}{k}^{-1} \quad \text{для } N = M, M+1, \dots$$

КАК ПОДСЧИТАТЬ НЕМЕЦКИЕ ТАНКИ

- Ответ (этот результат имеет смысл только для $k > 1$):

$$p(N|M, k) = \frac{k-1}{M} \binom{M}{k} \binom{N}{k}^{-1} \quad \text{для } N = M, M+1, \dots$$

- Получается *факториальное распределение* на $N - k$, его среднее и дисперсия известны:

$$\mathbb{E}[N | M, k] = \frac{k-1}{k-2} (M-1) \quad \text{для } k > 2,$$

$$\text{Var}[N|M, k] = \frac{(k-1)(M-1)(M-k+1)}{(k-2)^2(k-3)} \quad \text{для } k > 3.$$

ПРИМЕР ДЖЕЙНСА

ПРИМЕР ДЖЕЙНСА

- Рассмотрим простую задачу байесовского вывода: предположим, что перед нами урна, в которой N шаров, R из которых красные, а $N - R$ белые
- Мы достаём из урны n шаров, r из которых оказываются красными
- Легко найти распределение пар (n, r) по заданным N и R : вероятность вынуть из урны сначала r красных шаров подряд, а затем w белых, составляет

$$\begin{aligned} & \frac{R}{N} \frac{R-1}{N-1} \cdots \frac{R-r+1}{N-r+1} \cdot \frac{N-R}{N-r} \frac{N-R-1}{N-r-1} \cdots \frac{N-R-w+1}{N-r-w+1} = \\ &= \frac{R!(N-r)!}{(R-r)!N!} \cdot \frac{(N-R)!(N-r-w)!}{(N-R-w)!(N-r)!} = \frac{R!(N-R)!(N-r)!}{(R-r)!(N-R-w)!N!} \end{aligned}$$

- Эта вероятность получена для конкретной последовательности исходов, и она по симметрии не зависит от последовательности, то есть достаточно умножить её на число возможных таких последовательностей $\binom{n}{r}$:

$$p(r|N, R, n) = \frac{R!(N-R)!(N-n)!}{(R-r)!(N-R-w)!N!} \frac{n!}{r!(n-r)!} = \frac{\binom{R}{r} \binom{N-R}{n-r}}{\binom{N}{n}},$$

то есть гипергеометрическое распределение

- Но теперь перед нами стоит обратная задача: мы получили данные $D = (n, r)$ и хотим что-то сказать о значениях N и R
- Байесовский вывод проходит точно так же, как выше:

$$p(N, R|D) = \frac{p(D|N, R) p(R|N) p(N)}{p(D)},$$

и здесь в знаменателе стоит нормировочная константа

$$p(D) = \sum_{N=0}^{\infty} \sum_{R=0}^N p(D|N, R) p(R|N) p(N),$$

а правдоподобие представляет собой гипергеометрическое распределение

ПРИМЕР ДЖЕЙНСА

- Самый интересный вопрос: как выбрать априорное распределение $p(N, R)$? Мы его уже разложили в произведение

$$p(N, R) = p(R|N) p(N);$$

в байесовском выводе мы выбираем априорное распределение как захотим, согласно своей интуиции, но чего мы можем захотеть от этого априорного распределения, как формализовать интуицию?

- Во-первых, есть интуиция о том, что апостериорное распределение на величину N должно зависеть от числа выбранных шаров n (как в случае танков), но не должно зависеть от того, сколько из них оказались красными: с какой стати пропорция красных шаров должна влиять на их общее число?

- Более того в данном случае хотелось бы, чтобы полученные данные «обрезали» невозможные значения, то есть сообщали нам, что $N \geq n$, но не меняли нашу априорную интуицию о том, сколько именно шаров может быть в урне, если это число больше n . Иными словами, хотелось бы получить

$$p(N|D) = \begin{cases} c \cdot p(N), & \text{если } N \geq n, \\ 0 & \text{в противном случае,} \end{cases}$$

где c — нормировочная константа, не зависящая от N

- Это уже нетривиальное требование! Оно накладывает серьёзные ограничения на $p(N, R)$

- Апостериорное распределение на N равно

$$p(N|D) = \frac{p(N) p(D|N)}{p(D)} = \frac{p(N) \sum_{R=0}^N p(D|N, R) p(R|N)}{p(D)},$$

и требование выше значит, что

$$p(D|N) = \sum_{R=0}^N p(D|N, R) p(R|N) = \begin{cases} f(n, r), & \text{если } N \geq n, \\ 0 & \text{в противном случае,} \end{cases}$$

где $f(n, r)$ — это некоторая функция, зависящая от данных n и r , но не от N

ПРИМЕР ДЖЕЙНСА

- Более того, можно вспомнить, что мы говорим о гипергеометрическом распределении: для $N \geq n$

$$\sum_{R=0}^N \binom{R}{r} \binom{N-R}{n-r} p(R|N) = f(n, r) \binom{N}{n}$$

- Это уже весьма нетривиальное условие на априорное распределение $p(R|N)$!
- Мы получили его из базовой интуиции о том, что данные не должны менять наше мнение об N (это в некотором смысле независимость), но мы вряд ли смогли бы записать такое условие интуитивно, без этих рассуждений
- Разумные априорные распределения на R будут удовлетворять этому условию, и байесовский вывод по N будет тривиальным: апостериорное распределение пропорционально априорному, за исключением начального отрезка $N < n$, который данные делают невозможным

ПРИМЕР ДЖЕЙНСА

- Осталось выбрать $p(R|N)$
- Первый вариант — равномерное распределение $p(R|N) = \frac{1}{N+1}$ на интервале $0, \dots, N$
- Условие легко получить из свойств суммирования биномиальных коэффициентов: мы знаем, что $\sum_{R=0}^N \binom{R}{r} \binom{N-R}{n-r} = \binom{N+1}{n+1}$ (слева выбираем один элемент из $N+1$, фиксируя R , а потом выбираем r из R слева от него и $n-r$ из $N-R$ справа, то есть по сути выбираем $n+1$ элемент из $N+1$), а значит,

$$\begin{aligned} \sum_{R=0}^N \binom{R}{r} \binom{N-R}{n-r} p(R|N) &= \binom{N+1}{n+1} \frac{1}{N+1} = \\ &= \frac{(N+1)!(N+1-n-1)!}{(n+1)!(N+1)} = \frac{1}{n+1} \binom{N}{n}, \end{aligned}$$

то есть $f(n, r) = \frac{1}{n+1}$, и эта функция действительно не зависит от N

- Апостериорное распределение тогда получается таким:

$$p(R|D, N) = \binom{N+1}{n+1}^{-1} \binom{R}{r} \binom{N-R}{n-r};$$

уже не гипергеометрическое, потому что по R , а не по N

- Полученное апостериорное распределение удовлетворяет нескольким естественным требованиям «здравого смысла»:
 - для $n = r = 0$ это распределение превращается в априорное распределение $p(R|N) = \frac{1}{N+1}$;
 - для $n = r = 1$ (вытащили один шар, и он оказался красным) получается $p(R|D = (1, 1), N) = \frac{2R}{N(N+1)}$, что пропорционально R ; это тоже естественно, потому что в правдоподобии вероятность вытащить красный шар $p(r = 1|n = 1, R, N) = \frac{R}{N}$, и мы умножаем это правдоподобие на равномерное распределение;
 - кроме того, $\frac{2R}{N(N+1)}$ равно нулю при $R = 0$ (что логически невозможно), и это свойство сохраняется для всех $n \geq r \geq 1$

- Кроме того, можно провести байесовский вывод дальше, почти в точности как для монетки
- Для нашего апостериорного распределения

$$p(R|D, N) = \binom{N+1}{n+1}^{-1} \binom{R}{r} \binom{N-R}{n-r}$$

максимальная апостериорная гипотеза будет (проверьте в качестве упражнения)

$$R_{\text{MAP}} = \left\lfloor \frac{r}{n} (N+1) \right\rfloor,$$

что тоже звучит вполне естественно

ПРИМЕР ДЖЕЙНСА

- Но этот максимум не совпадает с математическим ожиданием:

$$\begin{aligned}\mathbb{E}[R] &= \sum_{R=0}^N R \cdot p(R|D, N) = \binom{N+1}{n+1}^{-1} \sum_{R=0}^N R \binom{R}{r} \binom{N-R}{n-r} = \\ &= \binom{N+1}{n+1}^{-1} \left(\sum_{R=0}^N (R+1) \binom{R}{r} \binom{N-R}{n-r} - \sum_{R=0}^N \binom{R}{r} \binom{N-R}{n-r} \right) = \\ &= \binom{N+1}{n+1}^{-1} \left(\sum_{R=0}^N (r+1) \binom{R+1}{r+1} \binom{N-R}{n-r} - \sum_{R=0}^N \binom{R}{r} \binom{N-R}{n-r} \right) = \\ &= (r+1) \binom{N+1}{n+1}^{-1} \binom{N+2}{n+2} - 1 = \frac{(N+2)(r+1)}{n+2} - 1,\end{aligned}$$

где мы воспользовались известным тождеством о биномиальных коэффициентах: $(R+1)\binom{R}{r} = (r+1)\binom{R+1}{r+1}$

- Для больших N , r и n этот результат, конечно, похож на $R_{\text{МАР}}$, но поправка существенна: ожидаемая доля красных шаров, остающихся в урне, составляет

$$\frac{\mathbb{E}[R] - r}{N - n} = \frac{(N + 2)(r + 1) - (n + 2) - r(n + 2)}{(N - n)(n + 2)} = \frac{r + 1}{n + 2},$$

то есть мы опять получили в точности правило Лапласа

- Можно и предсказательное распределение в виде правила Лапласа получить (опять упражнение):

$$\begin{aligned} p(\text{Red}|D, N) &= \sum_{R=0}^N R \cdot p(\text{Red}, R|D, N) = \\ &= \sum_{R=0}^N R \cdot p(\text{Red}|R, D, N) p(R|D, N) = \\ &= \binom{N+1}{n+1}^{-1} \sum_{R=0}^N \frac{R-r}{N-n} \binom{R}{r} \binom{N-R}{n-r} = \frac{r+1}{n+2}. \end{aligned}$$

- Аналогично (и снова упражнение) можно найти дисперсию апостериорного распределения:

$$\hat{R} = r + (N - n)p \pm \sqrt{\frac{p(1 - p)}{n + 3}(N + 2)(N - n)},$$

где мы обозначили $p = \frac{r+1}{n+2}$, а также подсчитали дисперсию именно оценки R , а не оценки вероятности $\frac{R-r}{N-n}$

- Оценка вероятности становится точнее с ростом n , то есть с ростом полученной выборки, причём падает до нуля, если $n = N$ — это очень интуитивно!

- А если вернуться к оценке доли шаров, то нам нужно будет разделить предыдущую оценку на $N - n$, и мы получим

$$\frac{\hat{R} - r}{N - n} = p \pm \sqrt{\frac{p(1-p)}{n+3} \frac{N+2}{N-n}}$$

- Теперь ситуация обратная: чем больше шаров мы вынули из урны, тем менее точной будет оценка доли оставшихся, а в пределе $N \rightarrow \infty$ останется как раз $p \pm \sqrt{\frac{p(1-p)}{n+3}}$, что в точности соответствует дисперсии монетки
- Но давайте исследуем и другие формы априорных распределений!

ПРИМЕР ДЖЕЙНСА

- Сначала сделаем совсем небольшое изменение:
предположим, что нам заранее известно, что в урне есть хотя бы один красный и хотя бы один белый шар, а в остальном оставим распределение равномерным: $p(R|N) = \frac{1}{N-1}$ на интервале $1, \dots, N-1$
- Тогда нужно выбросить слагаемые, соответствующие $R=0$ и $R=N$; для $R=0$ $\binom{R}{r} = [r=0]$, а для $R=N$ $\binom{N-R}{n-r} = [r=n]$, и наше тождество превращается в

$$\sum_{R=1}^{N-1} \binom{R}{r} \binom{N-R}{n-r} = \binom{N+1}{n+1} - \binom{N}{n} [r=n] - \binom{N}{n} [r=0]$$

- Соответственно, апостериорное распределение теперь

$$p(R|D, N) = \frac{\binom{R}{r} \binom{N-R}{n-r}}{\binom{N+1}{n+1} - \binom{N}{n} [r=n] - \binom{N}{n} [r=0]}$$

- Первое наблюдение: если мы получили данные с $0 < r < n$ (среди выбранных были и красные, и белые шары), то вся разница пропадает, и апостериорное распределение остаётся тем же
- Это кажется контринтуитивным: мы изменили вероятностное пространство, а распределение осталось тем же самым!
Джейнс пишет так: «И всё же после некоторого размышления мы видим, что результат верен, поскольку в этом случае из данных можно дедуктивно умозаключить, что R не может быть равным нулю или N ; поэтому не важно, говорит ли нам то же самое ещё и априорное распределение: состояние наших знаний об R в этих двух случаях одно и то же, на что указывают и теория вероятностей, и логика»

ПРИМЕР ДЖЕЙНСА

- А, например, для $r = 0$, когда мы не вытащили ни одного красного шара, результат изменится:

$$\begin{aligned} p(R|D = (0, n), N) &= \binom{N-R}{n} \left(\binom{N+1}{n+1} - \binom{N}{n} \right)^{-1} = \\ &= \binom{N}{n+1} \binom{N-R}{n}^{-1} \end{aligned}$$

для $R = 1, \dots, N-1$ и 0 вне этого интервала

- Это в точности совпадает с апостериорным распределением выше, только перенормированным на константу $\frac{N+1}{N-n}$
- Хотя на этот раз изменение в априорном распределении исключило аж максимальную апостериорную гипотезу $R = 0$, общий вид распределения на оставшейся части носителя всё равно не изменился; разные априорные распределения не обязательно приводят к разным выводам

- Идём дальше: правило Лапласа говорит, что предсказание при равномерном распределении сглаживается, причём это сглаживание в точности соответствует правдоподобию данных, в которых мы достали из урны два шара, и один из них оказался красным (эквивалентный размер выборки)
- Иначе говоря, равномерное априорное распределение даёт при выводе некоторую информацию, соответствующую выборке из $n = 2$ шаров
- Можно ли придумать полностью неинформативное (uninformative) априорное распределение, такое, чтобы никакой добавки не получалось?

- Давайте искать это распределение так: с какого распределения нужно начать, чтобы после получения данных $D = (n, r) = (2, 1)$ прийти к равномерному?
- Иначе говоря, на какое априорное распределение нужно умножить правдоподобие

$$p(r = 1 | N, R, n = 2) = \frac{\binom{R}{1} \binom{N-R}{1}}{\binom{N}{2}} = \frac{R(N-R)}{N(N-1)},$$

чтобы на выходе получилась константа?

- На первый взгляд кажется, что ничего не выйдет, ведь для $R = 0$ и $R = N$ получается $p(r = 1 | N, R, n = 2) = 0$

ПРИМЕР ДЖЕЙНСА

- Но мы только что придумали, как с этим справиться:отрежем крайние случаи в априорном распределении, и апостериорное не изменится, если $1 \leq r \leq n - 1$, а у нас данные именно такие!
- Таким образом, для априорного распределения

$$p(R|N) = \frac{\text{const}}{R(N-R)} \quad \text{для} \quad R = 1, \dots, N-1,$$

апостериорное распределение будет равно

$$\begin{aligned} p(R|D = (n, r), N) &= \frac{\text{const}}{R(N-R)} \binom{R}{r} \binom{N-R}{n-r} = \\ &= \frac{\text{const}}{r(n-r)} \binom{R-1}{r-1} \binom{N-R-1}{n-r-1}, \end{aligned}$$

- А из этого по формуле суммирования можно и нормировочную константу получить: для $R = 1, \dots, N - 1$

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

- Если подставить $n = 2$ и $r = 1$, мы получим равномерное распределение $\frac{1}{N-1}$; можно доказать, что и требуемое свойство выполняется, то есть можно вести вывод отдельно на R , и получить формулы ожидания и дисперсии апостериорного распределения (упражнение)

- Последний случай — «распределение биномиальных обезьян» (binomial monkey prior): предположим, что урна наполнялась шарами, каждый из которых имел некоторую вероятность g быть красным, причём цвет шаров выбирался независимо друг от друга (обезьяны-дальтоники бросают туда случайно выбранные мячи)
- Тогда априорное распределение будет биномиальным:

$$p(R|N) = \binom{N}{R} g^R (1-g)^{N-R}$$

- Это значит, что априори мы можем оценить число красных шаров в урне как $\hat{R} = Ng \pm \sqrt{Ng(1-g)}$

ПРИМЕР ДЖЕЙНСА

- Теперь проведём байесовский вывод; проверим наше свойство:

$$\begin{aligned} p(D|N) &= \sum_{R=0}^N p(D|N, R) p(R|N) = \sum_{R=0}^N \frac{\binom{R}{r} \binom{N-R}{n-r}}{\binom{N}{n}} \binom{N}{R} g^R (1-g)^{N-R} = \\ &= \binom{n}{r} \sum_{R=0}^N \binom{N-n}{R-r} g^R (1-g)^{N-R} = \\ &= \binom{n}{r} g^r (1-g)^{n-r} \sum_{R=0}^N \binom{N-n}{R-r} g^{R-r} (1-g)^{N-n-R} = \binom{n}{r} g^r (1-g)^{n-r}, \end{aligned}$$

где мы воспользовались тем, что (проверьте!)

$$\binom{R}{r} \binom{N-R}{n-r} \binom{N}{R} = \binom{N}{n} \binom{n}{r} \binom{N-n}{R-r},$$

а затем просуммировали по биномиальному распределению, выбирающему $R - r$ красных шаров из $N - n$ попыток; результат не зависит от N , так что свойство выполняется

- Подсчитаем апостериорное распределение:

$$p(R|D, N) = \text{const} \cdot \binom{N}{R} g^R (1-g)^{N-R} \binom{R}{r} \binom{N-r}{n-r},$$

и нормировочную константу можно подсчитать как

$$\begin{aligned} 1 &= \sum_{R=0}^N p(R|D, N) = \text{const} \binom{N}{n} \binom{n}{r} \sum_{R=0}^N \binom{N-n}{R-r} g^R (1-g)^{N-R} = \\ &= \text{const} \binom{N}{n} \binom{n}{r} g^r (1-g)^{n-r} \end{aligned}$$

для $R = r, \dots, N - n + r$

- Теперь можно выразить const и найти апостериорное распределение уже полностью, вместе с константой:

$$p(R|D, N) = \binom{N-n}{R-r} g^{R-r} (1-g)^{N-R-n+r}$$

- Отсюда можно найти и среднее, и дисперсию апостериорной оценки $\hat{R}$:

$$\hat{R} = r + (N-n)g \pm \sqrt{g(1-g)(N-n)},$$

и мы получаем оценку на долю оставшихся красных шаров

$$\frac{\hat{R} - r}{N - n} = g \pm \sqrt{\frac{g(1-g)}{N-n}}.$$

ПРИМЕР ДЖЕЙНСА

- Раньше апостериорное распределение на p определялось в первую очередь данными $D = (n, r)$, а теперь данные вообще практически не участвуют в оценках!
- Значение r для оценки теперь вообще не важно, а значение n просто постепенно уменьшает дисперсию
- Биномиальное априорное распределение *отменяет влияние данных*: апостериорная оценка совпадает с априорной и никак не зависит от r и n !
- Это выглядит странно, но, опять же, вполне интуитивно: априорное распределение с «биномиальными обезьянами» предполагало, что цвет шаров выбирается независимо друг от друга — вот и получилось, что знание цвета части шаров ничего не сообщает нам о цвете остальных
- Эта независимость оказалась «пропущена через» байесовский вывод и сохранилась в апостериорном результате

ПРИМЕР ДЖЕЙНСА

- Итак, мы рассмотрели пример об урне, из которой достают красные и белые шары, и провели байесовский вывод для нескольких разных априорных распределений:
 - обычного равномерного распределения на $R = 0, \dots, N$;
 - обрезанного равномерного, в котором исключены крайние случаи только красных и только белых шаров в урне, то есть равномерного распределения на $R = 1, \dots, N - 1$;
 - неинформативного распределения $\frac{\text{const}}{R(N-R)}$, которое не даёт дополнительного сглаживания в апостериорном распределении;
 - распределения биномиальных обезьян $p(R|N) = \binom{N}{R} g^R (1-g)^{N-R}$, которое, как выяснилось, полностью отменяет влияние данных и сохраняет независимость цвета шаров в урне друг от друга.
- Тем самым мы рассмотрели несколько важных частных случаев, которые дают представление о том, насколько разных и интересных эффектов можно добиться выбором априорного распределения; пока на дискретном примере...

СПАСИБО!

Спасибо за внимание!

