Уилл Курт. Байесовская статистика

Байесовская статистика недостаточно представлена в русскоязычной литературе. На мой взгляд, это одна из важнейших тем, выходящая далеко за рамки математики и статистики. Как ни странно, но человек является идеальным байесовским наблюдателем. Т.е., мы имеем суждения по множеству вопросов. Постоянно наблюдаем новые события. И корректируем свои первоначальные суждения на основе вновь увиденного. Байесовских подход позволяет выразить этот цикл научения количественно.

Книга Уила Курта даст вам понимание байесовской статистики буквально «на пальцах» — с помощью простых объяснений и ярких примеров. Прикладные задачи и упражнения помогут закрепить материал и заложить фундамент для работы с широким спектром задач: от невероятных текущих событий до ежедневных сюрпризов делового мира.

Уилл Курт. Байесовская статистика: Star Wars, LEGO, резиновые уточки и многое другое. – СПб.: Питер, 2021. – 304 с.

ЧАСТЬ I. ВВЕДЕНИЕ В ТЕОРИЮ ВЕРОЯТНОСТЕЙ

Глава 1. Байесовские рассуждения

Байесовская статистика очень хорошо согласуется с тем, как люди рассуждают в обычной жизни, делая выводы из имеющихся сведений. Сложность лишь в том, чтобы разбить этот процесс на шаги и привести к математической форме. Байесовские рассуждения повторяют ход ваших мыслей, когда вы сталкиваетесь с новой ситуацией – вскрыть имеющиеся предположения и обновить представления о мире на основе события. Таким образом цикл байесовского анализа:

1) получили данные;
2) сформулировали гипотезу;
3) пересмотрели свои представления, основываясь на новых данных.

У вас есть априорные предположения о том, как устроен мир. Наши априорные предположения — набор представлений, сформированных за годы жизни. Формула условной вероятности:

P(событие|априорные предположения) =…

Читается: вероятность *события* при условии, что выполнены *априорные предположения*, равна…

Например,

P(яркий свет за окном, тарелкообразный объект в небе | дело происходит на Земле) = очень низкая

Читается так: «Вероятность наблюдать яркий свет за окном и тарелкообразный объект в небе при условии, что дело происходит на Земле, очень низкая». Чтобы объяснить, что мы увидели, следует выдвинуть некоторую гипотезу — модель мира, которая даст какое-то предсказание.

Говоря о гипотезах в байесовской статистике, мы интересуемся, насколько хорошо они предсказывают наблюдаемые нами данные. Когда после увиденного вы думаете: «НЛО!» — то выдвигаете гипотезу.

(1) H_1 = НЛО у меня во дворе!

Так как гипотеза H_1 предсказывает данные D (здесь D просто краткая запись «мы видели яркий свет и тарелкообразный объект»), то, когда мы наблюдаем эти данные при условии верности гипотезы, их вероятность повышается. Формально это записывается так:

(2) $P(D|H_1,X) >> P(D|X)$

где X – наш прошлый опыт. Читается так: «Вероятность увидеть яркий свет за окном и тарелкообразный объект в небе при условии, что это НЛО, и при моем прошлом опыте намного больше, чем просто увидеть яркий свет за окном и тарелкообразный объект в небе без объяснений. Фактически, гипотеза объясняет имеющиеся данные.

Однако с учетом вашего предыдущего опыта гипотеза все еще смотрится диковато. Чтобы прийти к более надежным выводам, нужно собрать больше данных. Вы выглядываете в окно, и видите, что «тарелка» удерживается канатами, замечаете оператора с камерой, слышите хлопок и крик: «Стоп! Снято!» Исходная гипотеза – H_1 = Приземлилось НЛО – не подтвердилась. Зато родилась новая гипотеза H_2 = За окном снимают кино.

Рассмотреть две альтернативные гипотезы — значит сравнить теории, используя имеющиеся данные. Ключевой момент — сравнить, насколько хорошо две гипотезы объясняют наблюдаемые данные. Говоря, что вероятность наших данных при условии второй гипотезы намного больше, чем при условии первой, мы сообщаем, что вторая гипотеза объясняет наблюдения лучше. Это подводит нас к сути байесовского анализа: проверкой убеждений является то, насколько хорошо они объясняют мир. Мы считаем, что некоторые представления правильнее других, поскольку они лучше объясняют наблюдаемые вокруг явления. Математически мы выражаем нашу идею как отношение двух вероятностей:

$$(3)\ \frac{P(D_{\text{обновл.}}|H_2,X)}{P(D_{\text{обновл.}}|H_1,X)}$$

Большое отношение, например 1000, означает, что H_2 объясняет данные в 1000 раз лучше, чем H_1. Так как H2 объясняет данные во много раз лучше, чем H1, мы меняем наши представления с H_1 на H_2.

Глава 2. Измеряем неопределенность

Вероятность — мера нашей убежденности в чем-либо. Если мы определи P(X) как число, то оно покажет, насколько мы убеждены в X. В каком-то смысле вероятность — расширение логики. В логике у нас есть истина и ложь — обе выражают абсолютную убежденность. Вероятности расширяют логику до промежуточных значений между истиной и ложью.

Самый простой способ вычислить вероятность — посчитать отношение положительных исходов ко всем возможным исходам. В теории вероятностей множество всех событий обозначается Ω. Например, для броска монеты Ω = {орел, решка}.

Для вычисления вероятностей в более сложных случаях можно используют ставки. Например, ставка на лошадь 12 к 1 означает, что, поставив 1 доллар, вы получите от букмекера 12, если лошадь выиграет. Легко перейти от ставок к вероятностям:

$$(4)\ P(H) = \frac{O(H)}{1 + O(H)}$$

где O — ставка.

Столкнувшись на практике с каким-то абстрактным представлением, спрашивайте себя, сколько вы бы поставили на его верность.

Глава 3. Логика неопределенности

В классической логике важнейшую роль играют три операции: И, ИЛИ, НЕ. Операция НЕ для вероятностных рассуждений:

(5) P(¬X) = 1 – P(X)

Операция И соответствует вероятности наступления обоих событий, и равна произведению вероятностей:

(6) P(A, B) = P(A) * P(B)

Правило суммы вероятностей (операция ИЛИ):

(7) P(A ИЛИ B) = P(A) + P(B) – P(A, B)

Мы складываем вероятности обоих событий по отдельности и вычитаем вероятность одновременного их наступления. Заметим, что для взаимоисключающих событий P(A, B) = 0, так что формула упрощается: P(A ИЛИ B) = P(A) + P(B).

Глава 4. Как получить биноминальное распределение

Биномиальное распределение используется для подсчета числа успешных исходов, когда мы знаем число попыток и вероятность успеха. Приставка «би» означает, что возможных исходов два: событие происходит или нет. Каждое биномиальное распределение имеет три параметра: k – число интересующих нас исходов; n — число попыток; p — вероятность интересующего нас исхода. Например, вероятность выкинуть двух орлов за три броска монеты: k = 2 – число интересующих нас исходов (здесь – орлов); n = 3 – общее число бросков; p = 1/2 – вероятность выкинуть орла при броске. Биномиальное распределение обозначается B (k; n, p). В нашем примере B(2; 3, ½). И это лишь одно из значений распределения для заданных n и p. Распределение в целом обозначается B(n, p).

$$(8)\ B(k;n,p) = \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k}$$

где $\binom{n}{k}$ – число сочетаний из n по k:

$$(9)\ \binom{n}{k} = C_n^k = \frac{n!}{k! \cdot (n-k)!}$$

Функция, описывающая распределение, называется функцией вероятности (PMF, probability mass function). Она позволяет вычислить вероятность для любого k при фиксированных n и p. Например, для 10 бросков монеты биномиальное распределение имеет вид:

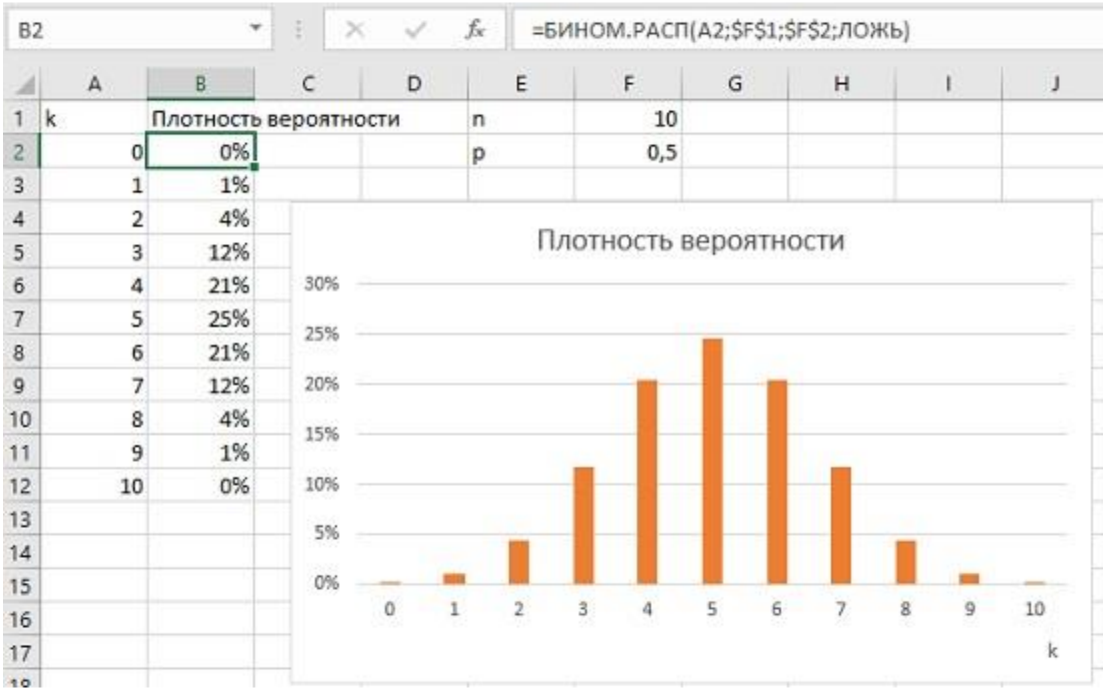

Рис. 1. Биноминальное распределение для 10 бросков монеты

Глава 5. Бета-распределение

Бета-распределение используется для оценки вероятности события, когда у вас есть опыт наблюдения некоторого числа испытаний. Например, вы сделали 100 бросков монетки, в которых орел выпал 40 раз. Для оценки вероятности выпадения орла в будущем следует использовать бета-распределение.

Изучая бета-распределение, мы также обсудим разницу между теорией вероятностей и статистикой. В литературе вероятности событий часто заданы. В реальности так бывает редко. Но можно использовать имеющиеся данные для оценки вероятностей. Задача нахождения вероятностей по данным называется *статистическим выводом*.

В отличие от биномиального распределения, распадающегося на дискретный набор значений, бета-распределение определено на сплошном интервале. Плотность вероятности бета-распределения:

$$(10)\ Beta(p;\alpha,\beta)=\frac{p^{\alpha-1}\cdot(1-p)^{\beta-1}}{beta(\alpha,\beta)}$$

где p – вероятность события или вероятность гипотезы; p не известна; мы строим кривую плотности вероятности бета-распределения, чтобы найти самую вероятную гипотезу; α показывает, сколько раз произошло интересующее нас событие; β – сколько раз оно не произошло; общее число испытаний равно α + β.

Мы вычитаем 1 из показателя степени и делим на бета-функцию для нормализации (то есть для того, чтобы распределение давало в сумме 1). Бета-функция — это интеграл по p от 0 до 1 (фактически – сумма всех возможных значений функции плотности распределения для всех гипотез p от 0 до 1):

$$(11)\ beta(\alpha,\beta)=\int_0^1 x^{\alpha-1}\cdot(1-x)^{\beta-1}\cdot dx$$

Фактически бета-распределение описывает, насколько хорошо разные биномиальные распределения, каждое со своей собственной вероятностью p, описывают наши наблюдаемые данные.

Допустим вы хотите определить, честная ли монета – то есть равны ли для нее вероятности орла и решки. Вы подбрасываете монету 20 раз и получаете 9 орлов и 11 решек. Какова вероятность того, что монетка честная? Честной будем считать монету, для которой вероятность орла находится в диапазоне 0,45–0,55. В Excel можно воспользоваться формулой =БЕТА.РАСП(0,55;9;11;ИСТИНА)-БЕТА.РАСП(0,45;9;11;ИСТИНА)=31%. Данные – лучший способ убедиться в верности своих утверждений. Вы еще 200 раз подбрасываете монетку и в итоге получаете 109 орлов и 111 решек. Какова теперь вероятность того, что монетка честная? =БЕТА.РАСП(0,55;109;111;ИСТИНА)-БЕТА.РАСП(0,45;109;111;ИСТИНА)=86%. Теперь мы на 86 % уверены, что монетка честная.

ЧАСТЬ II. БАЙЕСОВСКИЕ И АПРИОРНЫЕ ВЕРОЯТНОСТИ

Глава 6. Условная вероятность

Условные вероятности записываются так: P(A|B), вероятность A при условии B. Условные вероятности – важнейший инструмент статистики, они позволяют показать, как наши представления изменяются на основании поступившей информации.

Рассмотрим дальтонизм. В популяции 4,25% дальтоников, и в большинстве случаев причина в наследственности. Дальтонизм возникает из-за дефекта в одном из генов X-хромосомы. Так как у мужчин одна X-хромосома, а у женщин — две, мужчины в 16 раз чаще страдают дальтонизмом из-за дефектного гена:

P (дальтоник) = 0,0425;

P (дальтоник|женщина) = 0,005;

P (дальтоник|мужчина) = 0,08.

Формулы (6) и (7) для операций И/ИЛИ были выведены для независимых событий. Если события связаны, операция И подчиняется правилу умножения:

(12) P(A, B) = P(A) * P(B|A)

Правило суммы вероятностей (операция ИЛИ):

(13) P(A ИЛИ B) = P(A) + P(B) – P(A) * P(B|A)

Теорема Байеса

Один из самых замечательных трюков при работе с условными вероятностями – перемена мест условия и зависящего от него события, то есть использование вероятности P(A|B) для вычисления P(B|A).

Например, мы знаем, что среди мужчин дальтоников 8% или P(дальтоник|мужчина) = 0,08. А среди женщин – только 0,5% или P(дальтоник|женщина) = 0,005. Как найти P(мужчина|дальтоник)? Т.е., какова вероятность что, если человек дальтоник, то он мужчина?

Для начала определим, сколько дальтоников в стране? Допустим, всё население = N, тогда дальтоников = P(дальтоник)*N. А сколько среди них мужчин? P(мужчина)*P(дальтоник|мужчина)*N. Таким образом, если человек дальтоник, то вероятность, что он мужчина:

$$(14)\ P(\text{мужчина}|\text{дальтоник}) = \frac{P(\text{мужчина}) \cdot P(\text{дальтоник}|\text{мужчина}) \cdot N}{P(\text{дальтоник}) \cdot N} = \frac{0{,}5 \cdot 0{,}08}{0{,}0425} = 94\%$$

В общем виде теорема Байеса:

$$(15)\ P(\text{A}|\text{B}) = \frac{P(\text{A}) \cdot P(\text{B}|\text{A})}{P(\text{B})}$$

Теорема Байеса позволяет обратить вероятность увиденного в рамках наших представлений – P(наблюдения|представления) – в убежденность/вероятность наших представлений, если реализовалось увиденное – P(представления|наблюдения).

Теорема Байеса позволяет нам получить из данных и исходных представлений о мире оценку нашей уверенности в своих представлениях при этих данных. Часто наши представления P(A) в теореме Байеса взяты «с потолка». Мы ожесточенно спорим, уменьшит ли больший контроль продажи оружия число насильственных преступлений, помогает ли тестирование улучшению качества образования и хороша ли очередная реформа здравоохранения. Но мы редко думаем, как нас — или оппонентов — должны переубеждать данные. Теорема Байеса помогает понять, как данные меняют нашу уверенность в той или иной идее.

Глава 8. Априорная и апостериорная вероятности и правдоподобие в теореме Байеса

Теорема Байеса позволяет численно выразить, насколько наблюдаемые данные влияют на наши представления. При этом вероятность P(предположения|данные) показывает, насколько сильно мы держимся за наши предположения при условии имеющихся данных. Эта часть формулы называется апостериорной вероятностью. Именно ее мы ищем с помощью теоремы Байеса.

При этом вероятность имеющихся данных при условии наших предположений, P(данные|предположения), известна как правдоподобие, поскольку показывает, насколько правдоподобны данные (при условии имеющихся предположений). Третий компонент, вероятность имеющихся предположений – P(предположения) – называется априорной вероятностью и характеризует нашу убежденность до того, как мы увидели данные. Чтобы апостериорная вероятность лежала между 0 и 1, произведение правдоподобия и априорной вероятности следует поделить на вероятность данных – P(данные), т.е., нормализовать.

Рис. 2. Компоненты теоремы Байеса

Априорную вероятность и правдоподобие можно оценить или узнать из открытых источников. А вот вероятность данных узнать сложнее. Но часто это и не требуется, так как можно сравнить апостериорные вероятности для нескольких гипотез. Для этого пригодится следующая версия теоремы Байеса:

$(16)\ P(H|D) \propto P(H)*P(D|H)$

Ее можно прочитать так: «Апостериорная вероятность P(H|D) — вероятность гипотезы при наблюдении данных — пропорциональна априорной вероятности гипотезы P(H), умноженной на вероятность данных для гипотезы H, P(D|H)».

С такой записью теоремы Байеса отношение апостериорных вероятностей двух гипотез можно найти без информации о вероятности данных P(D):

$$(17)\ \frac{P(H_1|D)}{P(H_2|D)} = \frac{P(H_1) \cdot P(D|H_1)}{P(H_2) \cdot P(D|H_2)}$$

Формула (16) очень полезна, когда нужно сравнить вероятность двух идей, но нет возможности узнать P(D). Невозможно найти само по себе значение вероятности для гипотезы, но все еще можно сравнивать гипотезы по теореме Байеса. Сравнение гипотез означает, что мы всегда можем увидеть, во сколько раз одно объяснение лучше, чем другое.

Глава 9. Байесовские априорные вероятности и распределение вероятностей

Априорные вероятности являются наиболее спорным аспектом теоремы Байеса, потому что их часто считают субъективными. Но на практике они нередко иллюстрируют, как применять жизненно важную справочную информацию, чтобы обосновать неопределенную ситуацию.

Использование вероятностных распределений вместо отдельных (детерминированных) значений полезно по двум причинам. Во-первых, в действительности часто существует широкий спектр возможных убеждений, которые можно было бы иметь и рассматривать. Во-вторых, представление диапазонов вероятностей позволяет заявить об уверенности в ряде гипотез.

В качестве примера возьмем одну из ошибок статистического анализа из эпизода «Звездные войны: Империя наносит ответный удар». Когда Хан Соло, пытаясь уклониться от вражеских истребителей, направляет «Тысячелетний Сокол» в астероидное поле, всезнающий C-3PO сообщает Хану, что вероятность не на его стороне. C-3PO говорит: «Сэр, возможность успешного преодоления области астероидов — 1 к 3720!» Проблема анализа C-3PO состоит в том, что его данные относятся ко всем пилотам, но Хан сильно отличается от среднего пилота. Будем считать, что наша вера в то, что Хан выживет, составляет 20 000 к 1.

Воспользуемся бета-распределениями для правдободобия (данные C-3PO) и для априорной вероятности «крутости» Хана: =БЕТА.РАСП(2;7440;ЛОЖЬ) и =БЕТА.РАСП(20000;1;ЛОЖЬ)

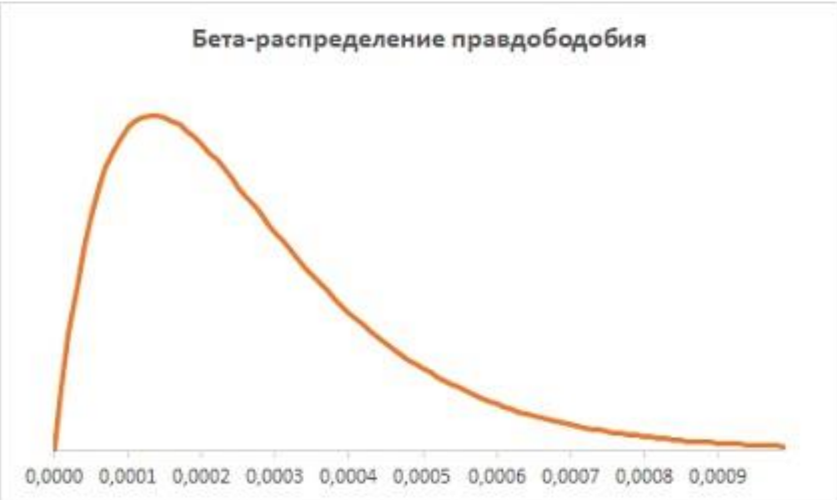

Рис. 3. Бета-распределение правдоподобия

Рис. 4. Бета-распределение априорной вероятности

Мы установили, во что верит С-3РО (правдоподобность), и смоделировали наши собственные убеждения о Хане (априорная вероятность). Объединяя убеждения с помощью формулы (16) мы создаем апостериорное распределение. В этом случае апостериорная вероятность моделирует чувство неопределенности. Есть простой способ объединить бета-распределения, которые дадут нормализованную апостериорную вероятность при наличии только правдоподобия и априорной вероятности. Альфа итогового распределения равно сумме альф, а бета – сумме бет. Beta(2 + 20 000, 7400 + 1):

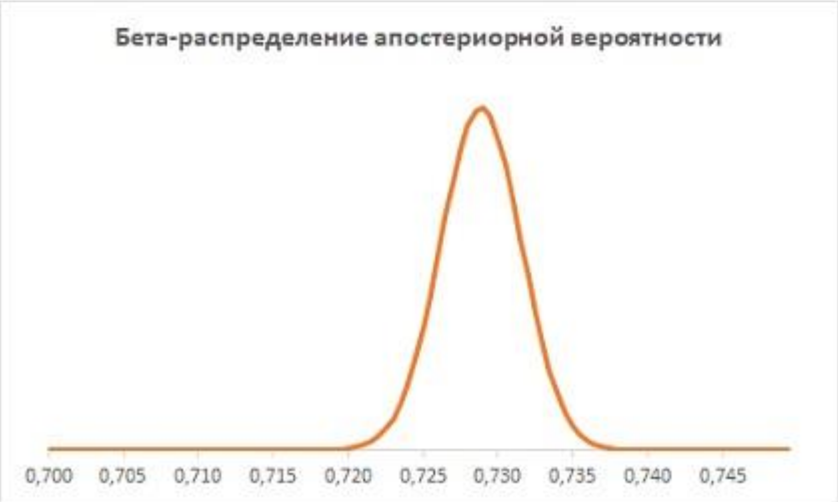

Рис. 5. Бета-распределение апостериорной вероятности

Наше апостериорное убеждение — шанс выживания примерно 73%. Весьма полезно, что у нас не просто одно значение – 73%, а полное распределение возможных убеждений. На практике использование распределения помогает проявлять гибкость относительно силы наших убеждений.

Часть III. ОЦЕНКА ПАРАМЕТРОВ

Глава 10. Введение в усреднение и оценку параметров

Допустим у нас есть неизвестное значение, которое нужно оценить, и мы можем использовать наблюдения, чтобы сделать предположение. Эти неизвестные значения называются параметрами, а процесс выбора наилучшего значения этих параметров — оценкой параметров. Одна из лучших методик работы с неопределенностью – усреднение измерений. К сожалению, в статистике часто допускают ошибку — слепое применение процедур без их понимания. Вероятность – наш инструмент для рассуждений о неопределенности, а оценка параметров, возможно, является наиболее распространенным процессом для решения проблем неопределенности.

Почему усреднение работает? Ошибки в измерении имеют тенденцию взаимно компенсировать друг друга. На практике не всегда можно отобрать все пространство возможных измерений, но чем больше выборка, тем большее количество ошибок будет устранено и тем ближе оценка будет к истинному значению.

Несмотря на то что среднее значение является очень простой и общеизвестной оценкой параметров, им можно легко злоупотребить, что приведет к весьма странным результатам. Например, мы можем измерить глубину снега в нескольких точках двора для оценки, сколько снега выпало прошлой ночью, прежде чем ветер разметал его. Однако при измерении «среднего роста» нет «нормального» человека, и различия в росте, которые мы наблюдаем, не являются ошибками — это действительно разные величины. Человек имеет рост 1,67 метра не потому, что какая-то его часть сместилась на человека ростом 1,92 метра!

Глава 11. Измерение разброса данных

Поскольку разброс наблюдений связан с неопределенностью в измерении, нужно иметь возможность количественно оценить его. Наиболее широко используются две характеристики разброса: дисперсия и стандартное отклонение.

Глава 12. Нормальное распределение

Истинная цель оценки параметров заключается не просто в оценке значения, а в том, чтобы назначить вероятность для диапазона возможных значений. Это позволяет проводить более сложные рассуждения с неопределенными значениями.

Нормальное распределение — это непрерывное распределение вероятностей, которое наилучшим образом описывает силу возможных убеждений в значении неопределенного измерения, учитывая известное среднее значение и стандартное отклонение. Оно принимает значения μ и σ (среднее значение и стандартное отклонение) в качестве двух параметров, и имеет форму колокола. Стандартное нормальное распределение с $\mu = 0$ и $\sigma = 1$:

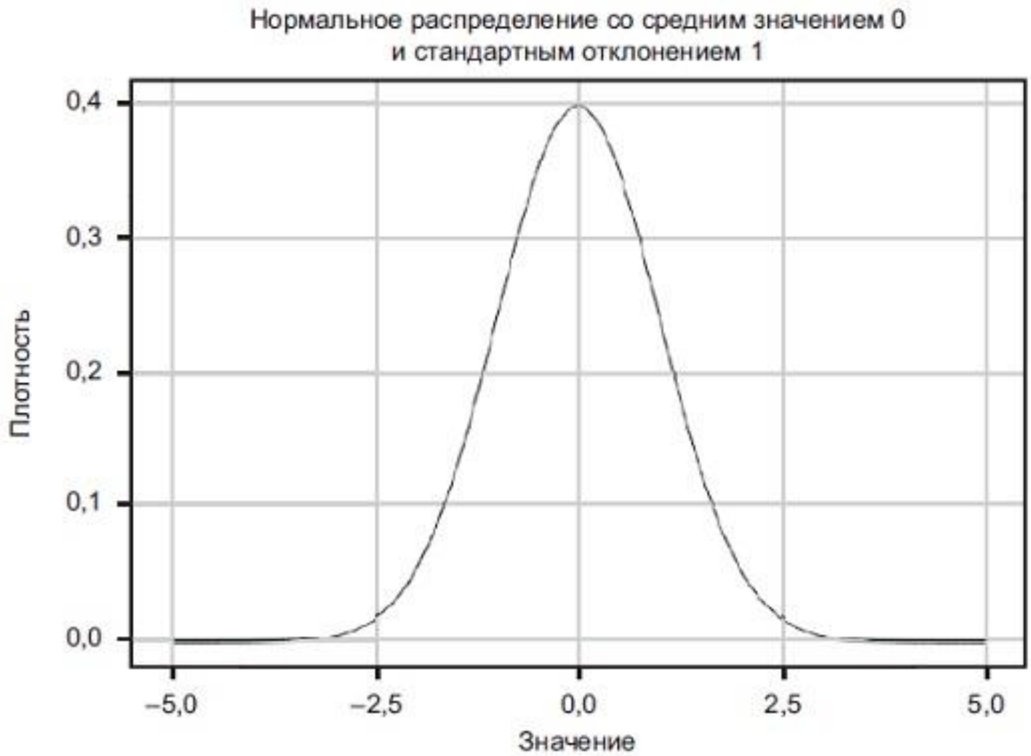

Рис. 6. Стандартное нормальное распределение

Ширина нормального распределения определяется его стандартным отклонением. По мере того как стандартное отклонение уменьшается, уменьшается и ширина нормального распределения. Нормальное распределение отражает то, насколько сильно мы верим в среднее значение. Если наши наблюдения разбросаны сильнее, мы верим в более широкий диапазон возможных значений и меньше доверяем среднему значению. И наоборот, если наши наблюдения имеют небольшое σ, мы считаем оценку довольно точной.

Глава 13. Инструменты оценки параметров: PDF, CDF и квантильная функция

Рассмотрим функцию плотности вероятности (probability density function, PDF) и интегральную функцию распределения (cumulative distribution function, CDF). Последняя хорошо справляется с оценкой вероятности диапазонов значений. Также изучим квантили, которые делят распределения вероятностей на части с равными вероятностями. Например, процентиль — это 100-я квантиль, то есть он делит распределение вероятностей на 100 равных частей.

Предположим, вы ведете блог и хотите знать вероятность того, что посетитель блога подпишется на вашу рассылку. Оказалось, что для первых 40 000 посетителей, вы получили 300 подписчиков. PDF в этом примере – это бета-распределение с α = 300 и β = 39 700. В Excel для визуализации бета-распределения можно использовать функцию:

=БЕТА.РАСП(***x***;α;β;ЛОЖЬ) для плотности вероятности PDF, и

=БЕТА.РАСП(***x***;α;β;ИСТИНА) для интегральной вероятности CDF.

	A	B	C	D
1				
2	Число удачных исходов	α	300	
3	Число неудачных исходов	β	39 700	
4				
5		p	Функция г	Интегральная вероятность, CDF
6		0,0060	1	0,0001
7		0,0060	1	0,0001
8		0,0060	1	0,0001
9		0,0060	1	0,0001
10		0,0060	2	0,0002
11		0,0061	2	0,0002
12		0,0061	2	0,0002
13		0,0061	2	0,0002
14		0,0061	2	0,0002
15		0,0061	3	0,0003
16		0,0061	3	0,0003
17		0,0061	3	0,0003
18		0,0061	3	0,0003
19		0,0061	4	0,0004
20		0,0061	4	0,0004
21		0,0061	4	0,0005
22		0,0062	5	0,0005
23		0,0062	5	0,0005
24		0,0062	6	0,0006
25		0,0062	6	0,0007
26		0,0062	7	0,0007
27		0,0062	7	0,0008
28		0,0062	8	0,0009
29		0,0062	8	0,0009
30		0,0062	9	0,0010
31		0,0062	10	0,0011
32		0,0063	11	0,0012
33		0,0063	11	0,0013
34		0,0063	12	0,0015

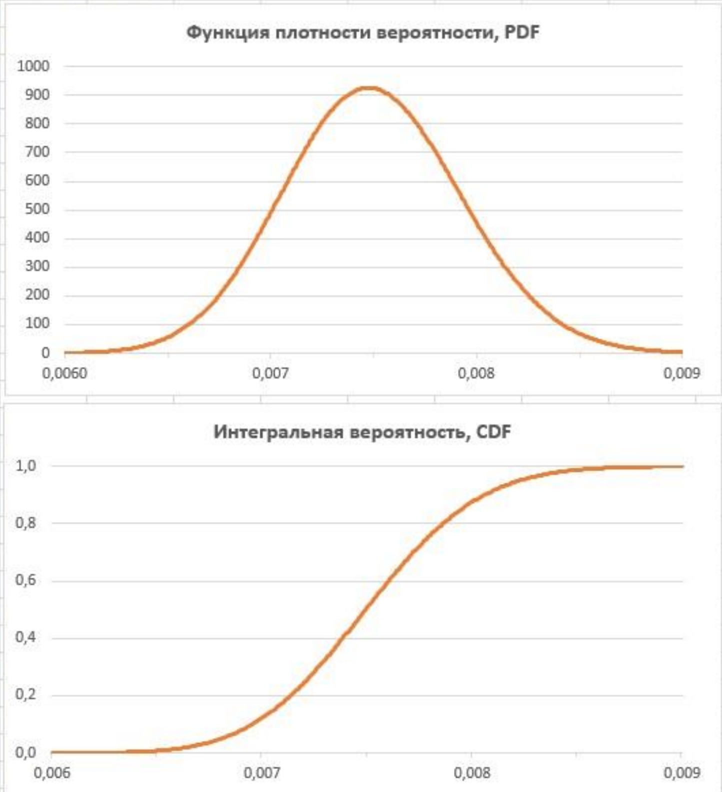

Рис. 7. Плотность вероятности PDF, и интегральная вероятность CDF бета-распределения

Глава 14. Оценка параметров с априорными вероятностями

Продолжим предыдущий пример. По данным вашего провайдера, для блогов, относящихся к той же категории, что и ваш, только 2,4% людей, открывающих письма, переходят к контенту. В байесовских терминах данные, которые мы наблюдаем, это правдоподобие, а информация от провайдера – априорная вероятность. Воспользуемся методом, описанным в главе 9, чтобы объединить априорную вероятность и правдоподобие.

Провайдер сообщил, что из первых 100 человек, открывших письмо, 25 перешли на ссылке. Мы обновили наши априорные убеждения (2,4%):

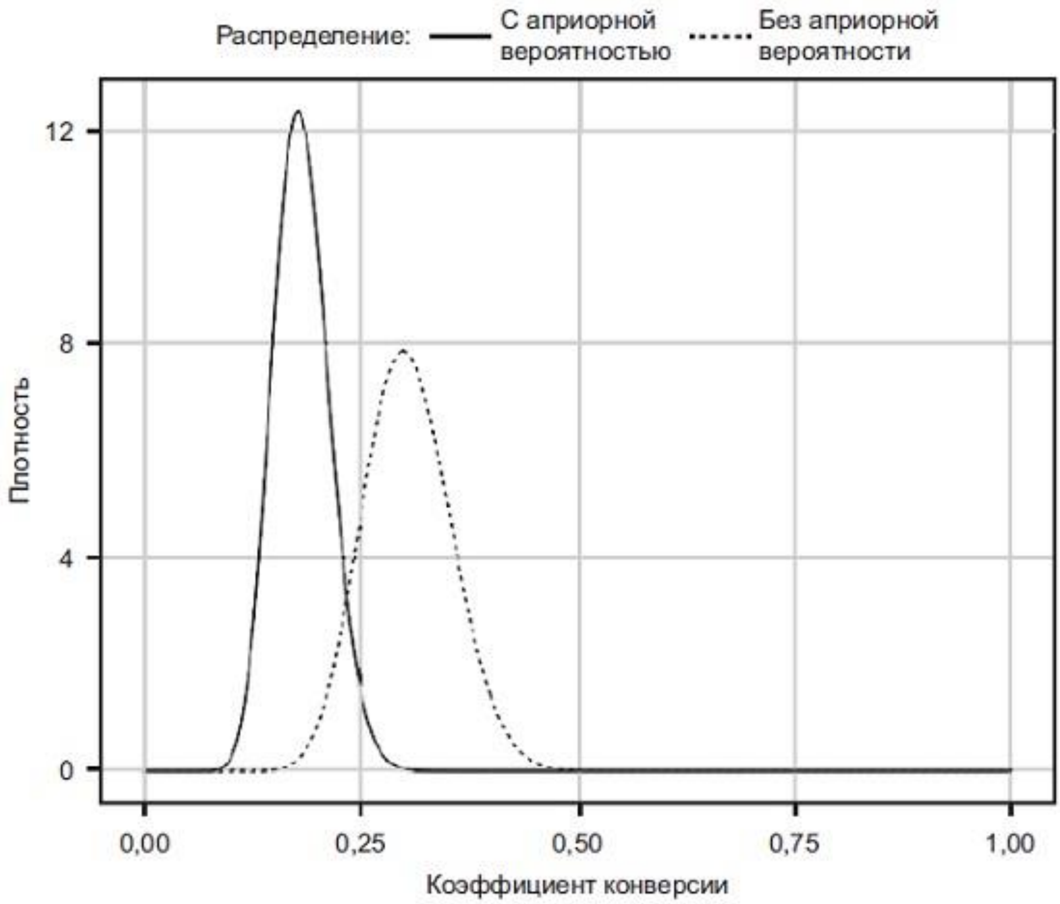

Рис. 8. Оценка коэффициента конверсии после 100 наблюдений

Мы подождали еще, и узнали, что из 300 подписчиков 86 перешли по ссылке. Мы еще раз обновили наши убеждения:

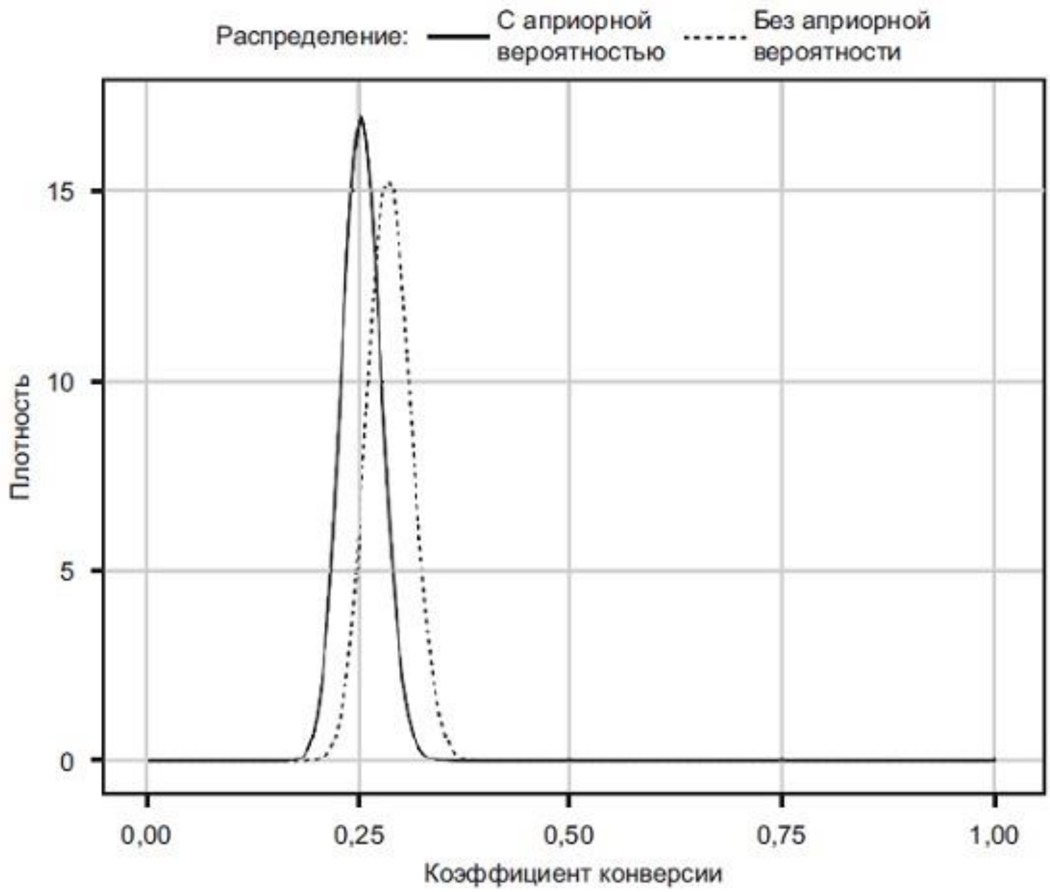

Рис. 9. Оценка коэффициента конверсии после 300 наблюдений

То, что мы наблюдаем здесь, является наиболее важным моментом в байесовской статистике: чем больше данных собирается, тем меньше влияние априорных убеждений на апостериорные. Тем ближе бета-распределения с и без априорных данных.

Существует ли справедливая априорная вероятность, если ничего не известно? В некоторых школах статистики считают, что при оценке параметров без какой-либо другой априорной вероятности к α и β нужно добавлять 1. Это соответствует использованию очень слабой априорной вероятности, которая считает, что каждый результат одинаково вероятен: Beta(1,1). Аргумент заключается в том, что это «самая справедливая» (то есть самая слабая) априорная вероятность, которую можно придумать в отсутствие информации. Но иногда такой подход дает странные результаты.

Стоит отметить, что вопрос, использовать ли Beta(1,1), имеет давнюю историю. Томас Байес с горем пополам верил в Beta(1,1), великий математик Пьер-Симон Лаплас был совершенно уверен, что Beta(1,1) имеет право на жизнь, а известный экономист Джон Мейнард Кейнс считал, что использование Beta(1,1) настолько нелепо, что дискредитирует всю байесовскую статистику!

ЧАСТЬ IV. ПРОВЕРКА ГИПОТЕЗ: СЕРДЦЕ СТАТИСТИКИ

Глава 15. От оценки параметров к проверке гипотез: создание байесовских A/B-тестов

С помощью A/B-теста проверим две гипотезы: удаление картинки из мейла увеличит коэффициент переходов; удаление картинки навредит кликабельности. Предположим, что у нас 600 подписчиков. Поскольку мы хотим использовать знания, полученные в ходе эксперимента, то проведем тест только на 300 из них. Таким образом, мы можем отправить оставшимся 300 подписчиков письмо, которое сочтем более эффективным.

Триста тестируем человек разделим на две группы: A и B. Группа A получит обычное письмо с большой картинкой вверху, а группа B получит письмо без картинки. Гипотеза такая: более простое письмо меньше будет похоже на спам и побудит пользователей переходить по ссылке.

Для простоты мы будем использовать одну и ту же априорную вероятность для обоих вариантов – 30%. Мы выберем слабую версию априорного распределения, что даст широкий диапазон коэффициентов конверсии. Остановимся на Beta(3,7):

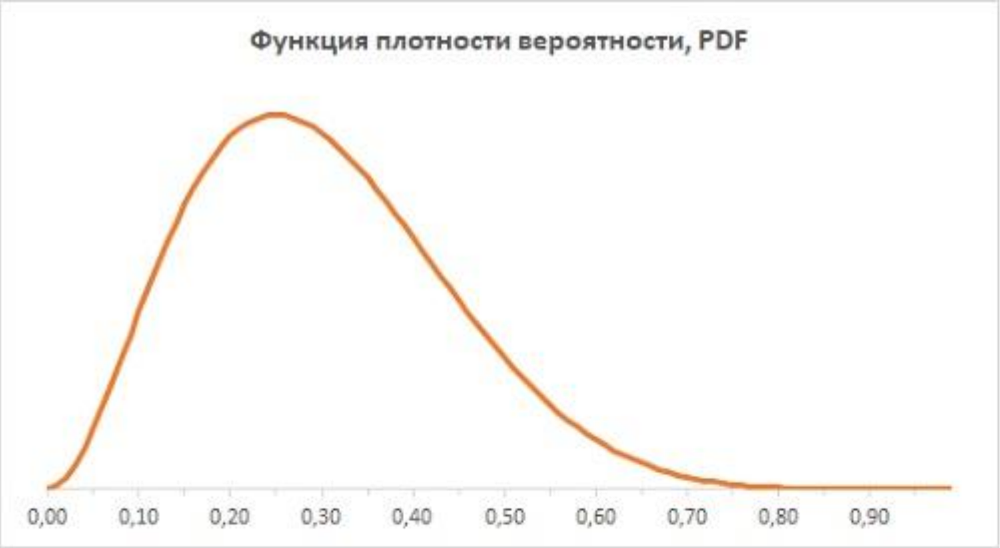

Рис. 10. Слабое априорное убеждение в коэффициенте конверсии 0,3

Мы отправили письма и получили следующие результаты: с письмом A перешли по ссылке 36; B – 50. Объединим каждый вариант по отдельности с априорным бета-распределением:

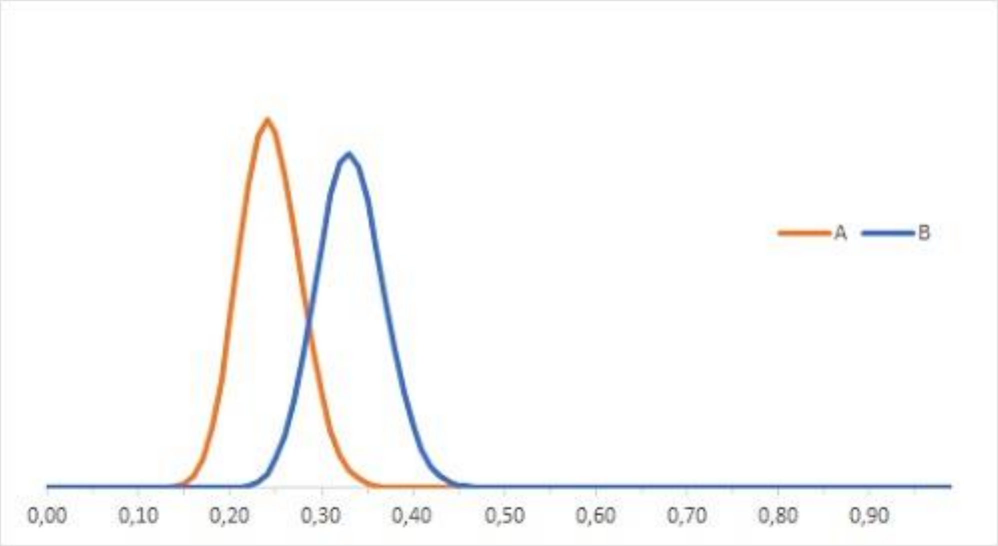

Рис. 11. Апостериорные вероятности А/В-теста

Бета-распределение для варианта В явно расположено правее, но также имеется и зона перекрытия. Насколько мы можем быть уверены, что В – лучший вариант? Можно воспользоваться моделированием по методу [Монте-Карло](#) (подробности см. в приложенном файле Excel на листе «Рис. 11»). С вероятностью 96% вариант В лучше варианта А.

Более того, моделирование методом Монте-Карло позволяет оценить насколько вариант В лучше варианта А:

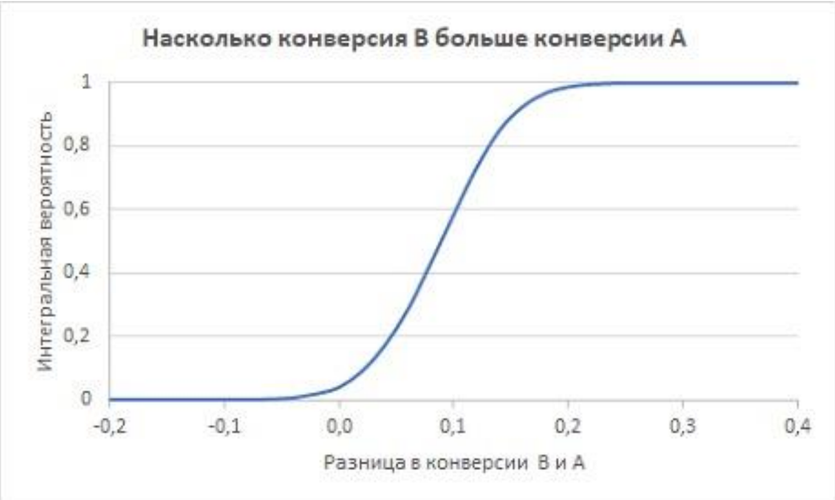

Рис. 12. Интегральная функция распределения (CDF) возможных улучшений

Глава 16. Введение в коэффициент Байеса и апостериорные шансы: конкуренция идей

Коэффициент Байеса – это формула, которая проверяет достоверность одной гипотезы, сравнивая ее с другой. В результате мы видим, во сколько раз одна гипотеза вероятнее, чем другая. Разделим соотношение (17)

$$(17)\ \frac{P(H_1) \cdot P(D|H_1)}{P(H_2) \cdot P(D|H_2)}$$

… на две части: коэффициент правдоподобия, или коэффициент Байеса:

$$(17а)\ \frac{P(D|H_1)}{P(D|H_2)}$$

… и коэффициент априорных вероятностей:

$$(17б)\ \frac{P(H_1)}{P(H_2)}$$

Коэффициент Байеса и апостериорные шансы можно использовать как форму проверки гипотезы, в которой каждый тест является соревнованием двух идеей. Предположим, у вашего друга в сумке лежат три шестигранных кубика. Один – утяжеленный (H_1). В половине случаев при подбрасывании выпадает шестерка. Два других кубика – обычные (H_2). Друг достает наугад кубик и бросает 10 раз. Выпали: 6, 1, 3, 6, 4, 5, 6, 1, 2, 6. Является ли кубик утяжеленным?

Первый шаг – вычисление P(D|H), или правдоподобия H_1 и H_2, учитывая наблюдаемые данные. Если кубик утяжеленный, вероятность выпадения шестерки равна 1/2, а вероятность выпадения любой иной цифры также равна 1/2. Это означает, что вероятность увидеть эти данные при использовании утяжеленного кубика равна:

$$(18)\ P(D|H_1) = \left(\frac{1}{2}\right)^4 \cdot \left(\frac{1}{2}\right)^6 = 0{,}00098$$

Для обычного кубика такой набор данных имеет вероятность:

$$(19)\ P(D|H_2) = \left(\frac{1}{6}\right)^4 \cdot \left(\frac{5}{6}\right)^6 = 0{,}00026$$

Коэффициент Байеса:

$$(20)\ \frac{P(D|H_1)}{P(D|H_2)} = \frac{0{,}00098}{0{,}00026} = 3{,}78$$

Это означает, что H_1 (кубик нечестный) объясняет наблюдаемые данные почти в четыре раза лучше, чем H_2. Основываясь на распределении игральных костей в сумке, мы знаем, априорные вероятности для каждой гипотезы: $P(H_1) = 1/3$, $P(H_2) = 2/3$. Таким образом, отношение апостериорных шансов:

$$(21)\ \frac{P(H_1) \cdot P(D|H_1)}{P(H_2) \cdot P(D|H_2)} = \frac{1/3}{2/3} \cdot 3{,}78 = 1{,}89$$

Глава 19. От проверки гипотез к оценке параметров

В главе 15 мы показали, как превратить задачу оценки параметров в проверку гипотез. В этой главе мы сделаем обратное: рассматривая практически непрерывный диапазон возможных гипотез, мы можем использовать коэффициент Байеса и апостериорные шансы (проверка гипотезы) в качестве формы оценки параметров! Этот подход позволяет оценивать более двух гипотез и предоставляет простую структуру для оценки любого параметра.

Предположим, прогуливаясь по ярмарке, вы замечаете, что кто-то спорит с работником ярмарки возле бассейна с маленькими резиновыми уточками. Вы подходите ближе и слышите, как человек кричит: «Вы жулики! Вы сказали, что шанс получить приз 1 к 2. Я выловил 20 уток и получил только один приз! Получается, что шанс на приз всего лишь 1 к 20!»

Вы создаете две гипотезы. H_1 — утверждение работника, что вероятность выигрыша равна 1/2, и H_2 — утверждение сердитого клиента, что вероятность выигрыша составляет всего 1/20:

H_1: P(выигрыша) = 1/2

H_2: P(выигрыша) = 1/20

Вы решаете посмотреть следующие 100 игр и использовать их в качестве данных. После того как клиент подобрал 100 уток, вы заметили, что 24 из них получили призы.

Поскольку у нас нет мнения ни о претензии клиента, ни о заявлениях работника, мы не будем беспокоиться об априорных шансах или вычислении полных апостериорных шансов. Просто вычислим коэффициент Байеса:

$$(22)\ \frac{P(D|H_2)}{P(D|H_1)} = \frac{(0{,}05)^{24} \cdot (1-0{,}05)^{76}}{(0{,}5)^{24} \cdot (1-0{,}5)^{76}} = \frac{1}{653}$$

Мы взяли соотношение с точки зрения H_2/H_1, чтобы результат сообщал нам, во сколько раз гипотеза клиента лучше объясняет данные, чем гипотеза работника. Коэффициент Байеса указывает, что H_1, гипотеза работника, объясняет данные в 653 раза лучше, чем H_2; это означает, что гипотеза работника (вероятность получить приз при вылавливании уточки составляет 0,5) является более вероятной.

Результат должен насторожить. Очевидно, что вероятность получить только 24 приза, когда было выловлено 100 уточек, кажется маловероятной, если истинная вероятность выигрыша равна 0,5. Мы можем использовать плотность биномиального распределения, которая сообщит вероятность получить 24 или меньше призов, предполагая, что вероятность получения приза действительно равна 0,5: =БИНОМ.РАСП(24;100;0,5;ИСТИНА)= 0,0000091%

Вероятность получения 24 или менее призов при истинной вероятности выигрыша 0,5 чрезвычайно мала. Хотя гипотеза работника, объясняет данные гораздо лучше, чем данные клиента, мы всё равно не верим в нее.

Проблема в том, что хотя интуитивно и понятно, что работник ошибается в своей гипотезе, альтернативная гипотеза клиента слишком экстремальна, чтобы быть верной. Что, если вероятность выигрыша равна 0,2? Мы назовем эту гипотезу H_3. Теперь отношение правдоподобия

$$(23)\ \frac{P(D|H_3)}{P(D|H_1)} = \frac{(0{,}2)^{24} \cdot (1-0{,}2)^{76}}{(0{,}5)^{24} \cdot (1-0{,}5)^{76}} = 917\ 399$$

… явно в пользу H_3. Но задачу мы не решили. Что, если есть еще лучшая гипотеза?

Воспользуемся инструментом Excel *Анализ "что-если"* –> *Таблица данных*.

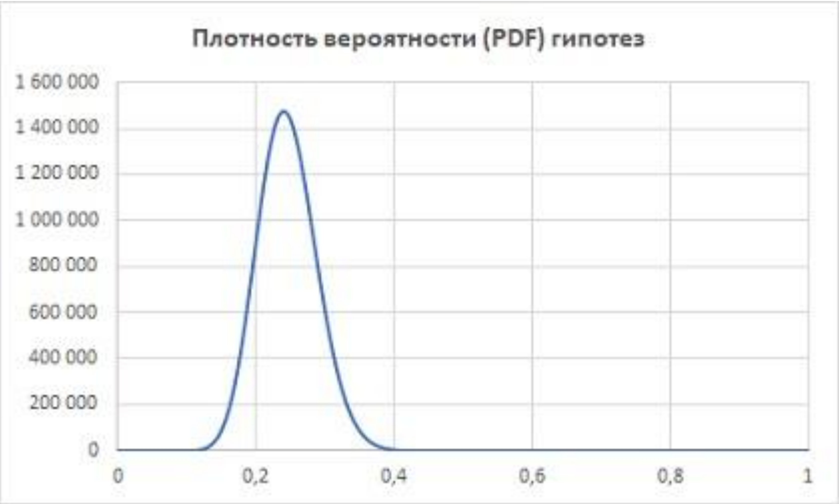

Рис. 13. Вероятность различных гипотез

Итак, мы получили распределение различных объяснений (гипотез) для наблюдаемых данных. Видно, что наибольшее отношение правдоподобия дает гипотеза H_4: P(выигрыша) = 0,24.

Из главы 10 вы узнали, что использование среднего значения наблюдаемых данных часто является хорошим способом оценки параметров. Здесь мы просто выбрали гипотезу, которая объясняет данные наилучшим образом, потому что сейчас нет способа взвесить оценки по вероятности их появления.

Любопытно, что проведенный анализ в точности соответствует бета-распределению с параметрами данных $\alpha=24$ и $\beta=76$ и априорной вероятностью $\alpha=1$ и $\beta=1$:

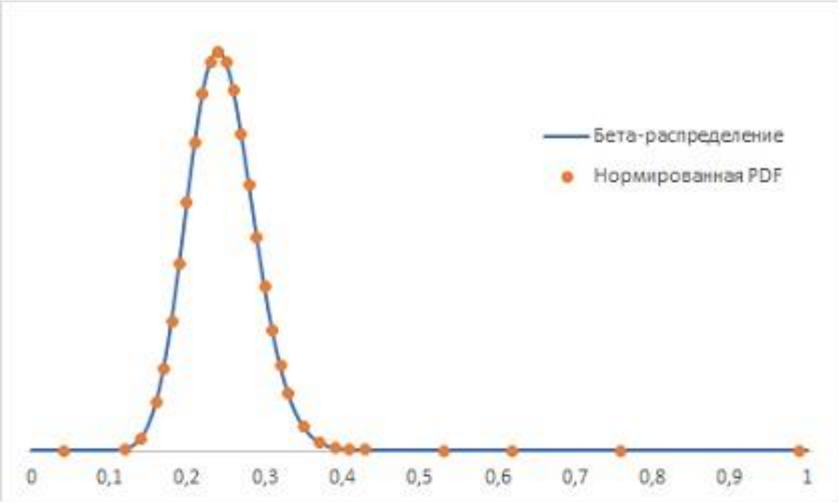

Рис. 14. Бета-распределение Beta(24+1;76+1) на фоне плотности распределения гипотез