Раскин М. А. «Условные вероятности. . »

L:\materials\raskin

Мы рассматриваем ситуацию, дальнейшее развитие которой мы не можем предсказать точно. При этом некоторые исходы (сценарии развития) для текущей задачи не отличаются. Например, при бросании монеты мы разбиваем исходы на выпадение орла и выпадение решки, не обращая внимания на то, куда в точности упала монета.

Определение 1. *Событие* — это множество возможных исходов.

Нас интересует некоторый набор свойств исходов (например, какой стороной упала монета). Можно рассмотреть *элементарные события* — множества исходов, каждое из которых получается при фиксации значений всех рассматриваемых свойств. Например, представим, что кидается два кубика, интересуют только выпавшие числа, и неправильные исходы (например, укатившиеся со стола кубики потерялись) почему-то невозможны (или просто достаточно маловероятны, чтобы их считали невозможными и считали погрешность от этого допустимой). В этой ситуации у нас будет 36 элементарных событий (например, утверждение ‘‘первый кубик выпал единицей вверх, а второй — тройкой’’ задаёт одно из этих 36 множеств исходов). Разумеется, множество элементарных исходов может быть бесконечным (например, мы бросаем монетку до первого орла, сколько бы раз её ни пришлось бросить).

Два элементарных события не пересекаются или совпадают, так как если они различны, то есть какое-то свойство, принимающее разные значения для всех исходов в одном событии и всех исходов во втором событии.

Будем рассматривать дискретный случай — то есть случай, когда элементарных событий конечное или счётное число.

В дальнейшем, мы будем считать, что все рассматриваемые события представляются в виде объединения элементарных. В принципе, для примера с бросанием кубиков можно было бы рассмотреть событие ‘‘первый кубик упал раньше второго’’, но мы в этом примере уже условились игнорировать всё, кроме выпавших чисел.

Определение 2. *Невозможным* событием называется пустое множество (подразумевается, что оно построено как множество исходов, в которое не попало ни одного исхода). *Достоверным* событием называется событие, состоящее из всех рассматриваемых исходов.

Определение 3. Мы говорим, что задано распределение вероятностей, если каждому из элементарных событий приписано действительное неотрицательное число (*вероятность*) и сумма вероятностей всех элементарных событий равна 1. Про свойство, вероятность выполнения которого равна 1, говорят, что оно выполнено почти наверное.

Определение 4. Пусть дано событие, являющееся объединением элементарных. Его вероятностью называется сумма вероятностей входящих в него элементарных событий. Будем обозначать вероятность события A как $Pr(A)$.

Замечание 1. Рассмотрим в последний раз события, не являющиеся объединением элементарных, и скажем, что их вероятность никак не определена.

Замечание 2. Можно задать вопрос, как связаны вероятности с реальностью. Если мы повторяем один и тот же эксперимент много раз, то частота события (доля экспериментов, в которых событие произошло) стремится к вероятности. К сожалению, это верно только с вероятностью 1 (а мы ещё не решили, стоит ли этому верить), а любые оценки для конечного числа повторений верны с вероятностью, меньшей 1.

У нас нет и не будет никакого внешнего выражения надёжности предсказаний теории вероятностей. Но опыт показывает, что их использование может принести пользу.

Определение 5. *Условная вероятность* события A при условии события B равна $Pr(A \mid B) := \frac{Pr(A \cap B)}{Pr(B)}$. Разумеется, событие B должно иметь ненулевую вероятность.

События A и B *независимы*, если $Pr(A \mid B) = Pr(A)$. Это равносильно $Pr(A \cap B) = Pr(A)Pr(B)$.

Теорема 1. Формула Байеса. $Pr(A \mid B) = \frac{Pr(B|A)Pr(A)}{Pr(B)}$.

Доказательство 1. $Pr(A \mid B) = \frac{Pr(A \cap B)}{Pr(B)} = \frac{Pr(B|A)Pr(A)}{Pr(B)}$

Теорема 2. Формула полной вероятности. Если $B_1 \cup \ldots \cup B_n \supset A$, то $Pr(A) = \sum\limits_{k=1}^{n} Pr(A \mid B_k)Pr(B_k)$.

Доказательство 2. $Pr(A) = \sum\limits_{k=1}^{n} Pr(A \cap B_k) = \sum\limits_{k=1}^{n} Pr(A \mid B_k)Pr(B_k)$

Определение 6. *Случайная величина* — это функция из элементарных событий в действительные числа (другими словами, это функция на исходах, постоянная на каждом элементарном событии). Две случайных величины ξ и η называются независимыми, если для любых x и y события $\xi = x$ и $\eta = y$ независимы.

Пример 1. Номер выпавшей при бросании грани кубика является случайной величиной.

Пример 2. Константой называется случайная величина, принимающая на всех элементарных событиях одно и тоже значение.

Определение 7. *Математическим ожиданием* случайной величины ξ называется $E\xi := \sum\limits_i Pr(\sigma_i)\xi(\sigma_i)$.

Теорема 3. Математическое ожидание суммы двух случайных величин равно сумме математических ожиданий. Математическое ожидание константы равно ей же.

Теорема 4. Если ξ и η независимы, то $E\xi\eta = E\xi E\eta$.

Доказательство 3. Обозначим за S множество элементарных событий.

Перегруппируем слагаемые в определении математического ожидания.

$$E\xi = \sum_i Pr(\sigma_i)\xi(\sigma_i) = \sum_{x \in \xi(S)} \sum_{\xi(\sigma)=x} xPr(\sigma) = \sum_x xPr(\xi = x)$$

.

Тогда
$$E\xi\eta = \sum_{z\in(\xi\eta)(S)} zPr(\xi\eta = z) = \sum_{z\in(\xi\eta)(S)} \sum_{x\in\xi(S),y\in\eta(S),xy=z} Pr(\xi = x и \eta = y)$$
. Заметим, что внешняя сумма происходит по всем возможным как произведение пары z, а внутренняя ограничена условием, что произведение навно текущему z. В результате получается сумма по всем парам принимаемых значений. Далее, $Pr(\xi = x и \eta = y) = Pr(\xi = x)Pr(\eta = y)$ в силу независимости. В результате получаем $\sum\limits_{x\in\xi(S),y\in\eta(S)} xPr(\xi = x)yPr(\eta = y) = E\xi E\eta$, ч. т. д.

Ясно, что математическое ожидание не описывает отклонение случайной величины от среднего значения. $E(\xi - E\xi) = E\xi - EE\xi = 0$ (первое равенство — ожидание суммы, второе — ожидание константы). Можно рассмотреть ожидание квадрата отклонения, $E(\xi - E\xi)^2$. Если ξ принимает каждое значение с вероятностью меньше 1 (то есть не является константой почти наверное), то сумма для этого ожидания содержит ненулевое слагаемое, и состоит только из неотрицательных членов, а, следовательно, не равна 0.

Определение 8. *Дисперсией* случайной величины ξ называется $D\xi = E(\xi - E\xi)^2$.

Дисперсия равна $E(\xi - E\xi)^2 = E(\xi^2 - 2\xi E\xi + E^2\xi) = E\xi^2 - E^2\xi$ (мы пользуемся тем, что $E\xi$ — константа, поэтому оно независимо с ξ и имеет ожидание $E\xi$).

Теорема 5. Дисперсия суммы независимых величин равна сумме дисперсий.

Доказательство 4.

$$E(\xi+\eta)^2 - E^2(\xi+\eta) = E(\xi^2+2\xi\eta+\eta^2) - (E\xi+E\eta)^2 = E\xi^2+2E\xi E\eta+E\eta^2 - E^2\xi - 2E\xi E\eta - E^2\eta = D\xi + D\eta$$

Для зависимых величин (например, ξ и $-\xi$) это не выполняется. Независимость использована в одном месте — при замене ожидания произведения (в ожидании квадрата суммы) на произведение ожиданий.

Теорема 6. $Da\xi = a^2 D\xi$

Пусть есть две случайные величины. Мы хотим выяснить, верно ли, что они либо обе принимают большие значения, либо обе принимают маленькие. Разумеется, ‘‘большие’’ и ‘‘маленькие’’ значения случайной величины имеют не абсолютный смысл, поэтому можно потребовать, чтобы добавление константы и умножение на положительную константу одной из величин не меняло этой характеристики. От сдвига можно избавиться, вычтя из величины её ожидание. От умножения — поделив величину на корень из её дисперсии.

Определение 9. *Корреляция* случайных величин ξ и η равна $\frac{E(\xi - E\xi)(\eta - E\eta)}{\sqrt{D\xi D\eta}}$.

Теорема 7. Корреляция величины с самой собой равна 1. Корреляция независимых величин равна 0. Корреляция ξ и $-\xi$ равна -1. Корреляция по модулю не больше 1.

Доказательство 5. Первые три утверждения проверяются лобовым счётом.

Четвёртое утверждение означает, в сущности, что $\left|\frac{\sum a_i b_i p_i}{\sqrt{\sum p_i a_i^2 \sum p_i b_i^2}}\right| \leqslant 1$, при $p_i \leqslant 0$. Это неравенство Коши-Буняковского (добавьте имена к названию по вкусу), оно, пожалуй, проще всего доказывается из того, что $a_i^2 t^2 - 2a_i b_i t + b_i^2 \geqslant 0$, это можно просуммировать с коэффициентами p_i, получится $t^2 \sum p_i a_i^2 - 2t \sum |a_i b_i| p_i + \sum p_i b_i^2 \geqslant 0$, тогда дискриминант неположителен.

В прошлый раз мы говорили, что дисперсия некоторым образом характеризует разброс значений случайной величины. Покажем, какой точный смысл можно придать этому утверждению.

Теорема 8. (Неравенство Чебышёва) Пусть ξ — случайная величина. Тогда для любого $k > 1$ выполнено $Pr(|\xi - E\xi| \geqslant k\sqrt{D\xi}) \leqslant \frac{1}{k^2}$.

Доказательство 6.

Лемма 1. Пусть для всех исходов $\xi \geqslant \eta$. Тогда $E\xi \geqslant E\eta$.

Доказательство 7. Напишем ожидания как суммы по элементарным событиям произведений вероятности и значения. Тогда в каждом слагаемом первой суммы первый сомножитель неотрицательный и такой же, как во второй, а второй не меньше.

Рассмотрим для заданных ξ и k из условия теоремы величину η, равную 0 при $|\xi - E\xi| < k\sqrt{D\xi}$ и 1 иначе. Очевидно, что $E\eta = Pr(|\xi - E\xi| \geqslant k\sqrt{D\xi})$. Рассмотрим также величину $\rho = \frac{1}{k^2 D\xi}(\xi - E\xi)^2$. Заметим, что $\rho \geqslant \eta$. Тогда $Pr(|\xi - E\xi| \geqslant k\sqrt{D\xi})$ равно $E\eta \leqslant E\rho = \frac{1}{k^2 D\xi} E(\xi - E\xi)^2 = \frac{1}{k^2}$, что и требовалось.

Теорема 9. (Закон больших чисел) Пусть мы повторяем эксперимент (отдельные повторы независимы) и считаем долю испытаний, в которой произошло данное событие A, имеющее вероятность p. Тогда вероятность того, что при n повторениях отклонение доли от вероятности превысит заданное $\varepsilon > 0$ стремится к нулю с ростом n.

Доказательство 8. Рассмотрим случайные величины ν_k, каждая из которых равна 1, если в k-м испытании произошло событие A и ν, равную сумме все ν_k с 1-го по n-е, делённой на n. Каждая из ν_k имеет математическое ожидание p и дисперсию $p(1-p) \leqslant$. Ожидание ν равно p. Дисперсия суммы ν_k в силу независимости величин равна $np(1-p)$, а дисперсия ν — по формуле выноса константного множителя из дисперсии — равна $\frac{p(1-p)}{n}$. Теперь можно применить к ν неравенство Чебышёва. Так как дисперсия стремится к нулю, то вероятность фиксированного отклонения тоже стремится к нулю.

Рассмотрим произвольное множество исходов, Ω, вероятностное пространство на нём и событие в нём B, имеющее положительную вероятность. Тогда можно считать, что B само является вероятностным пространством, так как условные вероятности при условии B обладают всеми свойствами вероятностей. Если у нас есть случайная величина ξ, то есть функция на исходах, то можно рассмотреть её ограничение на множество исходов B как случайную величину. Её математическое ожидание назовём математическим ожиданием ξ при условии B.

Определение 10. Пусть у нас есть множество событий, имеющих положительную вероятность, $\{B_1, \ldots, B_n, \ldots\}$. *условным математическим ожиданием* величины ξ при условии множества событий $\{B_1, \ldots, B_n, \ldots\}$ называется отображение, переводящее B_k в математическое ожидание ξ при условии B_k.

Если $B_1, \ldots, B_n, \ldots$ попарно не пересекаются, условное математическое ожидание является функцией не только на событиях, но и на самих исходах (никакому исходу не поставлено в соответствие два разных значения), определённой на каком-то событии, то есть случайной величиной.

Определение 11. Пусть η, ξ —случайные величины. Рассмотрим B — множество событий вида $''\eta$ принимает значение x''. Условное математическое ожидание ξ при условии η, $E(\xi|\eta) = E(\xi|B)$.

Основные свойства условного математического ожидания: $E(\xi|\xi) = \xi$, так как усреднение происходит по событиям, при условии которых ξ является константой. Кроме того, $E(\xi|\eta) = E\xi$ при независимых ξ и η, так как фиксация значения, принятого η, не меняет вероятности принятия значений ξ.

Приведём пример задачи, в которой ответ удобно определять через УМО. Пусть есть случайная величина ξ, которую мы хотим оценить по результатам эксперимента, и случайная величина η, которую нам сообщают. При этом будем считать, что мы сообщаем как оценку ξ величину $f(\eta)$ и выбираем функцию f для минимизации $E(f(\eta) - \xi)^2$.

Найдём, какое y нам выгодно сообщить при условии $\eta = x$. Вычислим

$$E((y-\xi)^2|\eta = x) = E(y^2 - 2y\xi + \xi^2|\eta = x) = y^2 - 2yE(\xi|\eta = x) + E(\xi^2|\eta = x) =$$
$$= (y - E(\xi|\eta = x))^2 + E(\xi^2|\eta = x) - E^2(\xi|\eta = x) = (y - E(\xi|\eta = x))^2 + D(\xi|\eta = x)$$

Ясно, что минимум достигается при $y = E(\xi|\eta = x)$. Выбирая таким образом значения f, получаем, что при наилучшей f будет выполнено $f(\eta) = E(\xi|\eta)$.