Вероятность в дискретном случае.

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 $P(A)$.

Замечание 1. Если сумма бесконечна, то ряд абсолютно сходится, так как все члены положительны и частичные суммы не больше 1.

Замечание 2. Рассмотрим в последний раз события, не являющиеся объединением элементарных, и скажем, что их вероятность никак не определена.

Замечание 3. 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.

Замечание 4. Есть упрощённое определение вероятности в дискретном случае как отношения числа благоприятных для события исходов к общему числу исходов. В наших терминах исходы заменяются на элементарные события, и если все они достаточно симметричны, чтобы быть равновероятными, то получается то же самое.

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

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

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

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

Замечание 5. Таким образом, если мы знаем, что событие B наступило, то наша оценка вероятности события A умножается на $\frac{P(B|A)}{P(B)}$. Например, если мы уже считаем высокой вероятность, что обвиняемый действительно виновен, то положительный результат ДНК-анализа, ошибающегося с вероятностью 10^{-6} укрепит нашу уверенность; но если это единственное доказательство, то следует понимать, что тестирование всего населения Москвы уже даст несколько подозреваемых.

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

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

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.

Польза от того, что значения — это числа, заключается в возможности выполнять арифметические операци с ними.

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

Замечание 6. Если этот ряд сходится условно, мы считаем, что математического ожидания нет, так как выделенного порядка на элементарных событиях нет. Мы не будем в дальнейшем проверять сходимость, все утверждения будут предполагать, что все участвующие ожидания сходятся.

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 $M\xi\eta = M\xi M\eta$.

Доказательство 3. Обозначим за S множество элементарных событий.

Перегруппируем слагаемые в определении математического ожидания.

$$M\xi = \sum_i P(\sigma_i)\xi(\sigma_i) = \sum_{x\in\xi(S)}\sum_{\xi(\sigma)=x} xP(\sigma) = \sum_x xP(\xi = x)$$

Тогда

$$M\xi\eta = \sum_{z\in(\xi\eta)(S)} zP(\xi\eta = z) = \sum_{z\in(\xi\eta)(S)}\sum_{x\in\xi(S),y\in\eta(S),xy=z} P(\xi = x \text{и} \eta = y)$$

Заметим, что внешняя сумма происходит по всем возможным как произведение пары z, а внутренняя ограничена условием, что произведение равно текущему z. В результате получается сумма по всем парам принимаемых значений. Далее,

$$P(\xi = x \text{и} \eta = y) = P(\xi = x)P(\eta = y)$$

в силу независимости. В результате получаем $\sum\limits_{x\in\xi(S),y\in\eta(S)} xP(\xi = x)yP(\eta = y) = M\xi M\eta$, ч. т. д.

Ясно, что математическое ожидание не описывает отклонение случайной величины от среднего значения. $M(\xi - M\xi) = M\xi - MM\xi = 0$ (первое равенство — ожидание суммы, второе — ожидание константы). Можно рассмотреть ожидание квадрата отклонения, $M(\xi - M\xi)^2$. 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 $D\xi = M(\xi - M\xi)^2$.

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

Теорема 5. Дисперсия суммы независимых величин равна сумме дисперсий.

Доказательство 4.

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

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. $Da\xi=a^2D\xi$

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 $\frac{M(\xi-M\xi)(\eta-M\eta)}{\sqrt{D\xi D\eta}}$.

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

Замечание 7. Зависимые случайные величины могут иметь нулевую корреляцию. Например, если мы бросаем две симметричные монетки и величина ξ равна числу -1 или числу 1 в зависимости от результата бросания первой монетки, а величина η принимает значение 0 при $\xi=-1$ и значения -1 и 1 в зависимости от результата бросания второй монетки при $\xi=1$.

Свойства случайной величины, такие как математическое ожидание и дисперсия, зависят только от вероятностей принятия случайной величиной каких-то значений. Стандартным способом записи этих вероятностей является функция распределения.

Определение 10. Пусть задана случайная величина ξ. *Функцией распределения* случайной величины ξ называется функция $F_\xi(x)=P(\xi\leqslant x)$.

Теорема 8. Функция распределения случайной величины стремится к значению 0 при $x\to-\infty$ и к значению 1 при $x\to\infty$. По функции распределения можно вычислить математическое ожидание и дисперсию случайной величины. При этом по функциям распределения двух случайных величин нельзя проверить их независимость.

Доказательство 6. Доказательство приведём для счётного случая.

Сумма вероятностей всех элементарных событий равна 1. При стремлении x к -1, сумма для вычисления величины $F_\xi(x)$ будет включать всё меньше слагаемых, причём каждое конкретное слагаемое будет вычеркнуто при некотором конкретном значении x. Таким образом мы получаем сумму хвоста сходящегося ряда, которая стремится к нулю. Аналогично при $x \to \infty$ мы получаем частичные суммы ряда, стремящегося к сумме 1, а они стремятся к числу 1.

Для дискретной величины принимаемые значения — это точки разрыва функции распределения, а вероятность каждого значения равна величине соответствующего разрыва. Этого достаточно для вычисления значений $M\xi$ и $D\xi$.

Заметим, что при бросании двух кубиков числа, выпавшие на кубиках, распределены одинаково и независимы. Но с теми же функциями распределения мы могли бы взять две совпадающие случайные величины, равные числу на первом кубике.

Замечание 8. Для нескольких случайных величин можно рассмотреть функцию *совместного распределения*. Это функция от нескольких аргументов, значение которой равно вероятности, что для всех k значение k-й величины оказалось менее k-го аргумента.