

Основные понятия теории вероятностей

Предыдущие заметки (см. [оглавление](#)) были посвящены методам сбора данных, способам построения таблиц и диаграмм, а также исследованию описательных статистик. В настоящей заметке излагаются основы теории вероятностей, позволяющей распространять результаты, полученные при изучении выборок, на всю генеральную совокупность.¹

Что означает слово *вероятность*? Вероятность — это возможность наступления некоторого события. Можно говорить о вероятности того, что из колоды карт будет вынута карта черной масти, что человек предпочтет один продукт другому или что новый продукт, появившийся на рынке, будет пользоваться спросом. В каждом из этих вариантов вероятность является числовой величиной, лежащей в интервале от 0 до 1 включительно. Вероятность события, которое никогда не может произойти (невозможное событие), равна 0, а вероятность события, которое происходит постоянно (достоверное событие), равна 1.

Существует три подхода к предмету теории вероятностей: *априорная* классическая вероятность, *эмпирическая* классическая вероятность и *субъективная* вероятность. В рамках априорного классического подхода вероятность события оценивается на основе априорной информации. В простейшем случае, когда все исходы испытаний равновероятны, их вероятность определяется в соответствии с формулой:

$$(1) \text{ вероятность события} = X / T,$$

где X — количество испытаний, в которых произошло событие, T — общее количество испытаний.

Что означает эта вероятность? Например, в колоде игральных карт есть 26 карт красной и 26 карт черной масти. Предположим, что после извлечения карта возвращается в колоду. Означает ли это, что из двух извлеченных карт одна обязательно окажется черной масти? Нет, поскольку никто не может предсказать исходы нескольких последовательных испытаний. Однако, если продолжать испытания достаточно долго, количество карт черной масти, извлеченных из колоды, будет приблизительно равно 0,5.

В предыдущем примере количество успешных испытаний и общее количество испытаний известно заранее. В рамках *эмпирического* классического подхода вероятность по-прежнему лежит в интервале от 0 до 1 и является результатом деления количества успешных исходов на общее количество испытаний, но вероятность вычисляется на основе наблюдаемой информации, а не априорной. В качестве примера можно привести количество семей, действительно купивших телевизор, количество избирателей, предпочитающих отдельного политика, или количество школьников, принятых на временную работу.

Третий подход к определению вероятностей — *субъективный*. В двух предыдущих случаях вероятность вычислялась совершенно объективно на основе априорной или апостериорной информации. Субъективная вероятность представляет собой шанс, который приписывается событию конкретным человеком. Другой человек может иначе оценивать шансы этого события. Оценки субъективной вероятности различных событий, как правило, основываются на личном опыте, общественном мнении и анализе конкретной ситуации. Субъективная вероятность особенно полезна в ситуациях, когда эмпирическую вероятность события вычислить невозможно.

Выборочное пространство и события

Основным понятием теории вероятностей является событие. Чтобы лучше понять смысл этого термина, обратимся к результатам опроса 1000 семей, в которых отражены как намерения купить телевизор, так и реальные покупки (рис. 1).

¹ Используются материалы книги Левин и др. Статистика для менеджеров. — М.: Вильямс, 2004. — с. 252–260

Планировалась ли покупка?	Совершена ли покупка?		
	Да	Нет	Всего
Да	200	50	250
Нет	100	650	750
Всего	300	700	1 000

Рис. 1. Поведение покупателей широкоэкранных телевизоров

Событие — это любой возможный результат случайного эксперимента. *Элементарное событие* — это событие, которое можно описать одной характеристикой. *Совокупность всех элементарных событий* называется выборочным пространством или пространством исходов.

В данном примере пространство элементарных событий состоит из 1000 семей. Элементарные события, принадлежащие выборочному пространству, классифицируются в зависимости от разновидности исхода. Например, если нас интересуют планы семей, события формулируются так: *покупка планируется* и *покупка не планируется*. Таким образом, способ, которым разбивается выборочное пространство, зависит от вида оцениваемой вероятности.

Дополнением события A называются все события, которые не являются частью события A . Дополнение события A обозначается символом A' . Так, на рис. 1 дополнением события *покупка планируется* является событие *покупка не планируется*.

Совместное событие — это событие, которое имеет несколько характеристик. Так, на рис. 1 событие *покупка планируется и покупка совершена* является совместным, поскольку оно состоит из двух элементарных событий — *покупка планируется* и *покупка совершена*.

Таблица сопряженности признаков

Существует несколько способов изучения выборочного пространства. Мы рассмотрим метод, основанный на таблице перекрестной классификации, частным случаем которой является таблица на рис. 1. Эту таблицу также называют таблицей сопряженности признаков или факторной (ранее мы уже приводили примеры таких таблиц; см. [Представление категориальных данных в виде таблиц и диаграмм](#)). Числа, указанные в ячейках таблицы, получены в результате разделения выборочного пространства, состоящего из 1000 семей, на семьи, планировавшие покупку и действительно купившие широкоэкранный телевизор, и семьи, не планировавшие и не сделавшие этого. Так, например, 200 семей планировали покупки и осуществили свои намерения.

Безусловная вероятность

Сформулируем несколько правил, позволяющих вычислить вероятности событий. Прежде всего, вероятность должна лежать в интервале от 0 до 1. Вероятность невозможного события равна 0, а достоверного — 1. Вероятность элементарного события A называется безусловной и обозначается как $P(A)$. Например, безусловной является вероятность события *покупка планируется*. Для того чтобы вычислить вероятность этого события, следует применить формулу (1)

$$P(\text{покупка планируется}) = \frac{\text{количество семей, планирующих покупку}}{\text{общее количество семей}} = \frac{250}{1000} = 0,25$$

Таким образом, вероятность того, что отдельная семья планирует покупку широкоэкранного телевизора, равна 0,25 (25%). Эта вероятность называется *безусловной*, поскольку общее количество успешных исходов (т.е. количество семей, планирующих покупку) можно просто извлечь из соответствующей ячейки таблицы сопряженности признаков. Рассмотрим еще один пример вычисления безусловной вероятности.

Вероятность совместных событий

В то время как безусловная вероятность относится к элементарным исходам, вероятность совместного события вычисляется в ситуациях, когда происходит несколько событий. Допустим, семья планирует и действительно приобретает широкоэкранный телевизор. Напомним, что совместное событие состоит из одновременных элементарных событий. На рис. 1 указано количество семей, которые планировали купить и действительно купили широкоэкранный телевизор. Поскольку

эта группа состоит из 200 семей, вероятность события *покупка планировалась и осуществилась* вычисляется по следующей формуле:

$P(\text{покупка планировалась и осуществилась}) = \text{количество семей, планировавших покупку и осуществивших ее} / \text{общее количество семей} = 200 / 1000 = 0,20$

С помощью понятия вероятности совместного события можно иначе определить безусловную вероятность элементарного исхода. Допустим, событие B состоит из двух взаимоисключающих событий — B_1 и B_2 . Тогда вероятность исхода A можно вычислить как сумму вероятности событий A и B_1 и A и B_2 . Безусловная вероятность элементарного события:

$$(2) P(A) = P(A \text{ и } B_1) + P(A \text{ и } B_2) + \dots + P(A \text{ и } B_k)$$

где события B_1, B_2, \dots, B_k являются взаимоисключающими и исчерпывающими.

Два события называются *взаимоисключающими*, если они не могут происходить одновременно. Множество событий называется *исчерпывающим*, если обязательно происходит хотя бы одно из них. Например, события *человек является мужчиной* и *человек является женщиной* являются взаимоисключающими и исчерпывающими. Эти события никогда не происходят одновременно, и в то же время человек всегда является либо мужчиной, либо женщиной.²

Применим формулу (2) для вычисления вероятности события *покупка планировалась*: $P(\text{покупка планировалась}) = \text{количество семей, планировавших покупку и осуществивших ее} / \text{общее количество семей} + \text{количество семей, планировавших покупку и не осуществивших ее} / \text{общее количество семей} = 200 / 1000 + 50 / 1000 = 0,25$. Этот результат можно было бы получить, просто сложив количество семей, планировавших покупку, и поделив его на общее количество семей.

Общее правило сложения вероятностей

Зная вероятности события A и события A и B , можно вычислить вероятность события A или B . В формулу для вычисления вероятности этого события входят вероятности событий A , B и A и B . Как определить вероятность того, что семья планировала покупку или совершила ее? Событие *покупка либо планировалась, либо совершена* охватывает все семьи, которые планировали покупку, и все семьи, совершившие ее. Для ответа на этот вопрос придется проверить каждую ячейку таблицы сопряженности признаков (рис. 1). Ячейка, соответствующая событию *покупка планировалась, но не совершена*, относится к событию *покупка планировалась*, поскольку учитывает семьи, которые действительно планировали приобрести широкоформатный телевизор. Ячейка, соответствующая событию *покупка не планировалась, но была совершена*, относится к событию *покупка совершена*, поскольку учитывает семьи, которые действительно приобрели широкоформатный телевизор. Ячейка, соответствующая событию *покупка планировалась и совершена*, удовлетворяет обоим условиям. Таким образом, вероятность того, что семья либо планировала, либо приобрела широкоформатный телевизор, вычисляется по следующей формуле:

$$P(\text{семья либо планировала, либо приобрела широкоформатный телевизор}) = P(\text{семья планировала, но не приобрела широкоформатный телевизор}) + P(\text{семья не планировала, но приобрела широкоформатный телевизор}) + P(\text{семья планировала и приобрела широкоформатный телевизор}) = 250 / 1000 + 300 / 1000 - 200 / 1000 = 350 / 1000 = 0,35$$

Вычисление вероятности события A или B подчиняется общей формуле сложения вероятностей: вероятность события A или B равна вероятности события A плюс вероятность события B минус вероятность события A и B .

$$(3) P(A \text{ или } B) = P(A) + P(B) - P(A \text{ и } B)$$

Правило сложения вероятностей взаимоисключающих событий. В некоторых ситуациях вероятность совместного события вычитать необязательно, поскольку она равна нулю. Например, попросим 300 семей, купивших широкоформатный телевизор, ответить, где они совершили покупку (рис. 2). Какова вероятность того, что случайно выбранная семья, купившая широкоэкранный телевизор, заказала его через Интернет или по почте?

² Я бы не был столь категоричным ☺

Место покупки	Количество респондентов
В магазине	183
По Интернет	87
По почте	30

Рис. 2. Сводная таблица покупок через Интернет и по почте

P (телевизор заказан через Интернет или по почте) = P (телевизор заказан через Интернет) + P (телевизор заказан по почте) – P (телевизор заказан через Интернет и по почте) = $87 / 300 + 30 / 300 - 0 / 300 = 117 / 300 = 0,39$. Следовательно, вероятность того, что случайно выбранная семья, купившая широкоэкранный телевизор, заказала его через Интернет или по почте, равна 0,39 (39%).

Совершенно очевидно, что клиент не может совершать заказы одновременно через Интернет и по почте. Следовательно, вероятность совместного события равна нулю. Реальная покупка может быть совершена только одним способом, иначе клиенту придется платить дважды. Как указывалось ранее, события, которые не могут происходить одновременно, называются взаимоисключающими.

Правило сложения вероятностей взаимоисключающих событий. Если события A и B являются взаимоисключающими, вероятность события A или B равна вероятности события A плюс вероятность события B :

$$(4) P(A \text{ или } B) = P(A) + P(B)$$

Правило сложения вероятностей исчерпывающих событий. Поскольку по определению множество событий называется исчерпывающим, если обязательно происходит хотя бы одно из них, то сумма исчерпывающих событий всегда равна 1.

Этические проблемы и вероятность

Когда в ходе рекламной кампании используются аргументы, апеллирующие к понятию вероятности, возникают этические проблемы. К сожалению, большинство населения слабо понимает смысл любых математических концепций и неправильно интерпретирует понятие вероятности. В некоторых ситуациях неправильная интерпретация является непреднамеренной, но в других случаях рекламное агентство может нарочно вводить потенциальных клиентов в заблуждение. Рассмотрим один из типичных примеров неэтичного применения теории вероятностей — государственную лотерею, для выигрыша в которой необходимо угадать заданное количество номеров (например, 6) из большого списка номеров (например, 54). Хотя большинство участников лотереи слабо верят в выигрыш, они даже не представляют, насколько мала вероятность угадать 6, 5 или 4 номеров из 54. Учитывая это, рекламный девиз «Мы не остановимся, пока все не станут миллионерами» следует признать совершенно ложным. Разумеется, пока лотерея приносит государству миллионы долларов в виде налогов, оно никогда не остановится, однако осчастливить всех играющих оно никогда не сможет.

Другой пример потенциально неэтичных применений теории вероятностей — обещание инвестиционных фондов принести 20% прибыли с 90%-ной вероятностью. В этом случае фонд должен: 1) разъяснить, как он вычислил эту вероятность, 2) сформулировать результат в другом виде (например, 9 шансов из 10) и 3) объяснить, что произойдет с остальными 10% инвестиций, которые не принесут 20% прибыли (будут ли они потеряны вообще?).

Правила счета

Вероятность интересующего нас исхода равна количеству вариантов, в которых возникает данный исход, деленному на общее количество исходов. Из-за большого количества возможностей во многих ситуациях трудно перечислить все варианты исходов. В таких случаях вместо списка исходов следует применять правила счета. Мы рассмотрим пять разных правил счета.

Предположим, что мы 10 раз подбрасываем монету. Каково количество разных возможных исходов (последовательностей орла и решки)? **Первое правило счета.** Если в n испытаниях могут возникнуть k взаимоисключающих и исчерпывающих событий, то количество возможных исходов равно k^n . Если монета подбрасывается 10 раз, то общее количество исходов равно $2^{10} = 1024$. Если игральный кубик, имеющий шесть сторон, выбрасывается два раза, то общее количество разных исходов равно $6^2 = 36$.

Второе правило счета обобщает первое и допускает ситуации, когда количество возможных исходов изменяется от испытания к испытанию. Например, сколько автомобильных номеров можно составить из трех букв и трех цифр. Каждая из букв порождает 12 исходов (в номерах используются только буквы, имеющие аналоги в английском алфавите: А, В, Е, К, М, Н, О, Р, С, Т, У и Х), а каждая из цифр — 10 исходов. Если первое испытание порождает k_1 событий, второе — k_2 событий, а n -е испытание — k_n событий, то общее количество возможных исходов равно $k_1 * k_2 * \dots * k_n$. Таким образом, количество автомобильных номеров, которые можно составить из трех букв и трех цифр: $12 \times 12 \times 12 \times 10 \times 10 \times 10 = 1\,728\,000$.

Третье правило счета позволяет вычислить количество способов, которыми можно упорядочить заданный набор объектов. Предположим, нам нужно расставить на книжной полке шесть книг. Сколько существует способов расстановки? Для начала следует понять, что на первом месте может стоять любая из шести книг. Как только первая позиция занята, на вторую позицию остается только пять кандидатов. Продолжим эту процедуру, пока все места не окажутся занятыми. Количество способов, которыми можно упорядочить заданный набор объектов равно $n! = n * (n-1) * (n-2) * \dots * 1$, где величина $n!$ называется факториалом, причем $0! = 1$ и $1! = 1$ по определению. Количество способов, которыми можно упорядочить шесть книг, равно $6! = 6 \times 5 \times 4 \times 3 \times 2 \times 1 = 720$.

Во многих задачах необходимо знать количество способов, которыми можно упорядочить подмножество, принадлежащее определенной группе объектов. Каждый вариант упорядочения называется *перестановкой*. Для примера изменим постановку предыдущей задачи. Предположим, что на полке помещаются только четыре книги. Сколько существует способов расставить книги на полке, если общее количество книг равно шести? **Четвертое правило счета.** Перестановки: количество способов упорядочить X объектов, извлеченных из совокупности, состоящей из n объектов, равно $n! / (n - X)!$ Исходя из этой формулы, общее количество вариантов упорядочения четырех книг, выбранных среди шести книг, равно $6! / (6-4)! = 6 \times 5 \times 4 \times 3 \times 2 \times 1 / 2 \times 1 = 360$.

В заключение рассмотрим ситуации, в которых порядок следования объектов не важен, а учитывается лишь количество вариантов извлечения X объектов из совокупности, состоящей из n объектов. Это правило называется правилом *сочетаний*. **Пятое правило счета.** Сочетания: количество способов извлечения X объектов из совокупности, состоящей из n объектов, равно

$$C_X^n = \frac{n!}{X!(n-X)!}$$

Сравнивая это правило с предыдущим, легко обнаружить, что они отличаются лишь множителем $X!$ в знаменателе. Обратите внимание на то, что в предыдущей формуле (правило 4) все *перестановки* отличаются друг от друга. В последней формуле порядок следования X объектов не важен, значит, $X!$ перестановок не отличаются друг от друга. Таким образом, общее количество комбинаций четырех книг, извлеченных из совокупности, состоящей из шести книг, равно

$$C_X^n = \frac{n!}{X!(n-X)!} = \frac{6!}{4!(6-4)!} = \frac{6!}{4!2!} = \frac{6 \times 5 \times 4 \times 3 \times 2 \times 1}{4 \times 3 \times 2 \times 1 \times 2 \times 1} = 15$$

Предыдущая заметка [Ковариация и коэффициент корреляции](#)

Следующая заметка

К оглавлению [Статистика для менеджеров с использованием Microsoft Excel](#)