Преподавание теории вероятностей в школе*

Г.И.Фалин
д.ф.м.н., проф.
МГУ им М.В.Ломоносова
механико-математический факультет
кафедра теории вероятностей
Москва 119991
http:// math.msu.su/~falin

В статье обсуждаются основные идеи, лежащие в основе теории вероятностей, и предлагается использовать для преподавания теории вероятностей на элементарном уровне частотное определение вероятности случайного события. Показано, как с помощью этого подхода получить основные формулы элементарной теории вероятностей, мотивированно ввести классическое определение вероятности, геометрические вероятности, условные вероятности, понятие независимости. Доказываются формула полной вероятности и формула Бейеса. Приведены примеры простейших применений этих понятий и результатов, которые доступны школьникам и будут полезны в их будущей жизни.

Ключевые слова: предмет теории вероятностей, частотное определение вероятности, теорема сложения, классическое определение вероятности, геометрическая вероятность, условная вероятность, независимость

---

* Расширенный вариант лекции, прочитанной в МГУ 21 ноября 2013 г. учителям московских школ. Частично опубликованно в журнале *Математика в школе*, 2014, № 2, стр. 28-37, № 3, стр. 55-64, № 4, стр. 32-44.

Введение

Теория вероятностей в современном её понимании была создана в 1933 г. профессором МГУ Андреем Николаевичем Колмогоровым (1903-1987), который опубликовал основополагающую монографию [1] (в 1933 г. на немецком, а в 1936 г. на русском языке). А.Н.Колмогоров и другие классики современной теории вероятностей не только занимались научными исследованиями. Все они были выдающимися педагогами и опубликовали признанные во всём мире учебные пособия. Целый ряд книг и статей посвящён преподаванию теории вероятностей на элементарном уровне. Особое место среди последних занимает небольшая книга [2] коллег и соавторов А.Н.Колмогорова – Бориса Владимировича Гнеденко (1912-1995) и Александра Яковлевича Хинчина (1894-1959) (тоже профессоров МГУ, которые сами признаны во всём мире как классики теории вероятностей) под названием «Элементарное введение в теорию вероятностей». Первое издание этой книги появилось в 1946 году. С тех пор она многократно переиздавалась и переведена на многие языки (на русском языке последнее, 14-е издание, вышло в 2016 году).

Ценные идеи о преподавании теории вероятностей содержатся в статье А.Н.Колмогорова [3], статьях «Вероятность» [4] (автор – А.Н.Колмогоров) и «Вероятностей теория» [5] (авторы – профессора МГУ Юрий Васильевич Прохоров (1929-2013) и Борис Александрович Севастьянов (1923-2013)) из Математической Энциклопедии, начальных параграфах классического университетского учебника Б.В.Гнеденко [6] (эти параграфы являются незначительной переработкой неопубликованных рукописей А.Н.Колмогорова), в книгах крупнейших зарубежных учёных: Вильяма Феллера (William Feller, 1906 – 1970) [7], Гаральда Крамера (Harold Cramér, 1893 – 1985) [8] и т.д.

Тщательный анализ всех этих трудов ясно показывает, что на начальном этапе изучения излагать теорию вероятностей нужно исходя из понятия частоты. Эту физическую теорию вероятностей называют *наивной* (в самом положительном смысле этого слова) – только она может служить настоящей, а не формальной, основой для последующего изучения теории вероятностей на более высоком уровне как строгой математической науки и реальных, а не надуманных, приложений теории вероятностей к задачам реального мира.

Последующее изложение в значительной мере следует книге Б.В.Гнеденко и А.Я.Хинчина [2] и статье А.Н.Колмогорова [3].

1. О предмете математики

В соответствии с классическим определением Фридриха Энгельса,

«… математика имеет своим объектом пространственные формы и количественные отношения действительного мира …»

(Ф.Энгельс. Анти-Дюринг. Диалектика природы – отдел первый, раздел III).

Практически дословно это определение повторяется в статье «Математика» из Математической энциклопедии (Москва, Изд-во «Советская энциклопедия», 1982, том 3), которая написана на основе одноимённой статьи А.Н.Колмогорова из БСЭ:

«Математика – наука о количественных отношениях и пространственных формах действительного мира.»

К сожалению, в силу разных причин к середине 20 века из специфического инструмента изучения окружающего нас мира математика в значительной мере превратилась в «игры разума», полностью сложился и сейчас как в научных исследованиях, так и в преподавании доминирует другой, схоластический[1], взгляд на математику, который характеризуется чрезмерным преувеличением роли аксиматического метода, доходящим до абсурда стремлением к «строгости» и т.д.:

«Со времён греков говорить «математика» – значит говорить «доказательство».»

(Н.Бурбаки, Теория множеств, М.: Мир, 1965);

Детальное обсуждение предмета математики, методов её преподавания на разных уровнях, других вопросов философии математики и науки увело бы нас слишком далеко от темы статьи. Поэтому мы ограничимся несколькими цитатами из работ выдающихся математиков. В ходе дальнейшего изложения мы будем руководствоваться выраженными в них мыслями.

Рассказывая о математическом анализе в своей книге [9] (глава «Анализ», раздел III.1 «Общие замечания относительно исчисления бесконечно малых») великий немецкий математик Феликс Клейн (Felix Christian Klein, 1849-1925) пишет:

«Я хотел бы предпослать замечание общего характера относительно предмета математики. Вы можете часто услышать от нематематиков…, что математика занимается исключительно выводами логических следствий из ясно заданных посылок… Совершенно иначе смотрит на дело всякий, кто сам продуктивно занимается математикой. В действительности люди, мнение которых было приведено выше, судят исключительно по той выкристаллизованной форме, в которой принято излагать

[1] Схоластика (в переносном смысле) – оторванное от жизни бесплодное умствование (*Советский Энциклопедический Словарь. 1979).*

готовые математические теории, но исследователь работает в математике, как и во всякой другой науке, совершенно иначе: он существенно пользуется своей фантазией и продвигается вперёд индуктивно, опираясь на эвристические вспомогательные средства. Можно привести немало примеров того, как великие математики находили самые важные теоремы, не будучи в состоянии строго их доказать. Неужели допустимо не ценить такое великое творчество, неужели надо в угоду приведённому выше определению математики сказать, что всё это не математика и что только те позднейшие математики, которые нашли, наконец, вылощенные доказательства теорем, – только они одни двигали математику?...

…Убедительность такого рода наивных рассуждений представляется, конечно, различным лицам весьма различной. Многие – к ним принадлежу и я сам – чувствуют себя в высшей степени ими удовлетворёнными. Другие же, будучи односторонне расположены к чисто логической стороне, находят, что такие соображения ничего не говорят, и не могут согласиться с тем, чтобы на них можно было вообще смотреть как на основание для математических рассуждений.

С другой стороны, такие наивные приёмы мышления и в настоящее время очень часто применяются всякий раз, когда хотят – в математической физике, в механике, в дифференциальной геометрии – применить какое-нибудь математическое положение; там эти приёмы, как вы все знаете, весьма целесообразны…»

Выдающийся русский математик, профессор механико-математического факультета МГУ, Владимир Игоревич Арнольд (1937-2010) в статье «О преподавании математики» [10] отмечает:

«Математика — это та часть физики, в которой эксперименты дёшевы… В середине двадцатого века была предпринята попытка разделить математику и физику. Последствия оказались катастрофическими. Выросли целые поколения математиков, незнакомых с половиной своей науки и, естественно, не имеющих никакого представления ни о каких других науках. Они начали учить своей уродливой схоластической псевдоматематике сначала студентов, а потом и школьников…Поскольку ни для преподавания, ни для приложений в каких-либо других науках схоластическая, отрезанная от физики, математика не приспособлена, результатом оказалась всеобщая ненависть к математикам — и со стороны несчастных школьников (некоторые из которых со временем стали министрами), и со стороны пользователей. Уродливое здание, построенное замученными комплексом неполноценности математиками-недоучками, не сумевшими своевременно познакомиться с физикой, напоминает стройную аксиоматическую теорию нечётных чисел. Ясно, что такую теорию можно создать и заставить учеников восхищаться совершенством и внутренней непротиворечивостью возникающей структуры (в которой определена, например, сумма нечётного числа слагаемых и произведение любого числа сомножителей). Чётные же числа с этой сектантской точки зрения можно либо объявить ересью, либо со временем

ввести в теорию, пополнив её (уступая потребностям физики и реального мира) некоторыми «идеальными» объектами. К сожалению, именно подобное уродливое извращённое построение математики господствовало в преподавании математики в течение десятилетий. Возникнув первоначально во Франции, это извращение быстро распространилось на обучение основам математики сперва студентов, а потом и школьников всех специальностей (сперва во Франции, а потом и в других странах, включая Россию).»

Именно как часть физики, т.е. науки о природе, изучающей наиболее общие свойства окружающего нас реального мира, а не «игры разума», понимали выдающиеся учёные и теорию вероятностей.

Английский логик и философ, основоположник частотного подхода к вероятности, Джон Венн (John Venn, 1834 – 1923; школьникам его имя знакомо по диаграммам Венна) в предисловии к своей книге «Логика случая» [11] пишет:

«Мнение, что вероятность – это раздел математики, а не ветвь общей науки о наблюдаемых фактах, ошибочно…»

Рихард Мизес (Richard von Mises, 1883 – 1953), один из наиболее выдающихся специалистов в области теории вероятностей, в [12] отмечает:

«Предметом теории вероятностей, как и всякой другой ветви естествознания, служат не какие-либо спекуляции, не мнения, не «формы рассудка», но исключительно доступные наблюдению факты.»

Вспомним также, как выдающийся немецкий математик Давид Гильберт (David Hilbert, 1862-1943) в своём знаменитом докладе на Втором Международном Конгрессе математиков в Париже 8 августа 1900 года сформулировал Шестую Проблему (её решение дал в 1933 г. А.Н.Колмогоров, предложив ставшую теперь общепринятой аксиоматику теории вероятностей):

«6. Математическое изложение аксиом физики. … задача об аксиоматическом построении… тех физических дисциплин, в которых уже теперь математика играет выдающуюся роль: это в первую очередь теория вероятностей…» (цитируется по [13]).

2. Основные понятия теории вероятностей

2.1 Эксперимент и событие

Когда мы говорим о вероятности, то всегда имеем в виду вероятность наступления того или иного *события*. Понятие события является одним из основных понятий как общей аксиоматической теории вероятностей, так и наивной элементарной.

В общей аксиоматической теории вероятностей оно вводится через систему аксиом: «событие» (точнее говоря, «случайное событие») – это элемент сигма-алгебры $\mathbb{F}$ подмножеств пространства элементарных событий Ω, на которой определена вероятностная мера P.

В элементарной теории вероятностей понятие события поясняется на примерах, а события описываются обычными фразами на русском языке. После того, как событие описано, для краткости его можно обозначить большой буквой латинского алфавита. Вводить понятие пространства элементарных событий (в другой терминологии, множества исходов эксперимента), отождествлять событие с множеством благоприятствующих ему исходов эксперимента, а операции над событиями – с соответствующими операциями над множествами следует на более поздних этапах изучения предмета. В статье [3] А.Н.Колмогоров пишет: *«начинающему полезнее держаться наглядного понимания терминов «событие» и «вероятность»…»*. Более того, при решении реальных задач (я не имею в виду упражнения для школьников или студентов) практически никогда не оперируют с событиями как с множествами.

Более того, при решении реальных задач (я не имею в виду упражнения для школьников или студентов) практически никогда не оперируют с событиями как с множествами. Профессор кафедры теории вероятностей механико-математического факультета МГУ Валерий Николаевич Тутубалин, который является не только крупным учёным, но и выдающимся педагогом и методистом, в своём учебнике [7] (на стр. 14) пишет:

«Лишь в учебниках теории вероятностей действует схема, согласно которой для реального явления нужно сначала построить модель из Ω *и* $\mathsf{P}(\omega)$ … *Фактически вероятности одних событий находятся по вероятностям других событий без полного описания (иногда)* Ω *и (как правило)* $\mathsf{P}(\omega)$*, а прямо путём обращения к формулам исчисления.»*

Ясное понимание соотношений между разными событиями и основных операций над событиями безусловно необходимо для усвоения теории вероятностей (подробнее об этом мы поговорим позже, перед тем, как начнём доказывать основные теоремы теории вероятностей).

В качестве примера события можно привести следующее: «утром жидкость в бачке омывателя стёкол автомобиля находится в жидком состоянии» – наступление или ненаступление этого события чрезвычайно важно для любого автомобилиста в холодное время года. Говоря об этом событии, как и о любом другом, мы неявно предполагаем, что нам известно намного больше о ситуации, для которой оно рассматриватся. Мы знаем, например, что в бачок залита обычная вода, на улице $-2^{\circ}C$, атмосферное давление равно 760 мм ртутного столба. Из физики известно, что при этих условиях вода замёрзнет. Но если на улице будет $+2^{\circ}C$ или в бачок залита специальная смесь, то жидкость в бачке омывателя будет находится в жидком состоянии. Иначе говоря, всегда нужно ясно и полностью указывать *комплекс условий* $\mathfrak{S}$, при которых рассматривается наступление или ненаступление интересующего нас события A. В комплекс условий $\mathfrak{S}$ мы будем включать и указание на то, *что именно мы наблюдаем* (в нашем случае – агрегатное состояние вещества, а не его, например, цвет). Реализацию комплекса условий $\mathfrak{S}$ называют *экспериментом*. Ниже для упрощения речи мы будем использовать термины «комплекс условий $\mathfrak{S}$» и «эксперимент $\mathfrak{S}$» как синонимы.

2.2 Детерминированные события и эксперименты

В только что рассмотренном примере при выполнении описанного комплекса условий (характер жидкости, температура, давление) мы совершенно точно знаем, произойдёт или нет интересующее нас событие – такие события называют *детерминированными*[2] (по отношению к заданному комплексу условий $\mathfrak{S}$). Можно сказать, что эксперимент по определению агрегатного состояния жидкости в бачке омывателя при выполнении условий «в бачок залита обычная вода, на улице $-2^{\circ}C$, атмосферное давление – 760 мм ртутного столба» приводит к детерминированному исходу, т.к. результат его совершенно точно известен (вода замёрзнет).

2.3 Случайные события и эксперименты

Детерминированным событиям противостоят случайные события. Событие называется *случайным* по отношению к заданному комплексу условий $\mathfrak{S}$, если до осуществления этого комплекса условий нельзя определённо сказать, произойдёт это событие или нет. Рассмотрим, например, следующий «эксперимент» (он описан в задаче № 4 раздела 1 демонстрационного варианта ЕГЭ 2018 года; эта задача включается в демонстрационные варианты начиная с 2012 года). Школьник идёт на экзамен по биологии. Программа включает 25 вопросов, из которых 2 (в них есть вопросы о грибах) школьник не выучил. На экзамене ему достаётся один из 25 вопросов (например, на стол выкладывают 25 билетов и школьник берёт один из них). Если он отвечает на этот вопрос, экзамен сдан, если нет, то провален (в оригинальной формулировке демоварианта ЕГЭ 2018 года речь идёт о том, что школьник не получит вопрос про грибы; в другие годы – что получит). Ясно, что событие «экзамен сдан» является случайным по отношению к заданному комплексу условий, т.к. заранее нельзя знать, какой именно вопрос достанется школьнику.

---

[2] Англ. determinate (от латинского determinatus) – определённый, точно установленный.

2.4 Уникальные случайные эксперименты

Характерной особенностью последнего «эксперимента» (экзамена по биологии, который сдаёт конкретный школьник в конкретный день) является его *уникальность*. Уникальность эксперимента означает, что комплекс условий $\mathfrak{S}$ можно повторить только один раз – при повторной попытке сдать экзамен знания школьника будут другими (он либо выучит новые темы, либо забудет часть материала, который учил раньше).

Исключительно важно понимать следующее:

Если речь идёт о событии, связанном с уникальным экспериментом, то сказать можно только одно: оно либо произойдёт, либо не произойдёт. Уникальные эксперименты со случайным результатом и, соответственно, уникальные случайные события не являются предметом теории вероятностей.

Поэтому, например, в рассмотренном выше примере не имеет смысла говорить о вероятности того, что в билете, который достанется школьнику, не будет вопроса о грибах. Можно приписывать такому событию какие угодно «вероятности», но это не будет иметь никакого объективного, научного смысла в математической теории вероятностей.

Важно, однако, подчеркнуть, что *вероятность и случайность* являются предметом научного анализа не только в теории вероятностей. В более широком смысле (что включает и уникальные эксперименты) эти и другие сопутствующие понятия (*достоверность, неопределённость* и т.п.) изучают философия, логика, психология и другие науки. Математику особенно полезно ознакомиться с уже упомянутой книгой Мизеса [12], книгами [14], [15] выдающегося французского математика Эмиля Бореля (Émile Borel, 1871 – 1956), главой XI «Исчисление вероятностей» из книги [16] французского математика, физика и философа науки Анри Пуанкаре (Henri Poincaré, 1854 – 1912) – одного из величайших математиков в истории, разделом «Введение. Природа теории вероятностей» из классическое руководства Феллера [7]. В этих публикациях разбираются смысл слов «случайность», «вероятность» и т.п. в обычном языке и в математической теории, границы применимости теории вероятностей и многие другие принципиальные идеи, лежащие в основе теории вероятностей. Без хорошего понимания этих вопросов немыслимы ни преподавание, ни применение теории вероятностей.

2.5 Массовые однородные эксперименты

Итак, если про событие можно сказать только то, что оно случайно, то теория вероятностей в современном её понимании неприменима для его описания. Что же изучает теория вероятностей?

Предметом теории вероятностей являются только случайные события, связанные с массовыми однородными экспериментами, т.е. такими экспериментами, которые могут быть много раз повторены в примерно одних и тех же условиях.

Исторические примечания.

(1) Этот взляд на явления, которые являются предметом теории вероятностей, впервые был ясно выражен в середине 19 века Венном в большом философском труде «Логика случая» [11]. В §3 главы I он пишет:

«Среди вещей, имеющих отношение к вероятности, фундаментальная концепция, которую читатель должен зафиксировать в уме настолько ясно, насколько это возможно … – это понятие серии».

Под серией («series») Венн понимает *«большое количество или ряд следующих друг за другом подобных объектов»* («a large number or succession of objects»), а в §6 гл. I уточняет:

«… это серия исходов, которые имеют некоторое число общих черт или свойств, – без этого они не были бы собраны вместе. Но одновременно между ними имеются и различия; некоторые свойства могут быть обнаружены в одних исходах и отсутствуют в других.» □

(2) Через 50 лет, в 1919 году, частотный взгляд Венна на вероятность развил и дополнил Рихард Мизес в математической статье [17]. В этой статье Мизес чётко сформулировал идеи, высказанные ещё Венном, и предложил (первый) аксиоматический подход к вероятности, основанный на понятии частоты. Сейчас именно с именем Мизеса связывают частотную теорию вероятностей. Впрочем, Britannica отмечает, что *«Венн представил первое систематическое изложение частотной теории верояностей».* В опубликованной десятью годами позже книге [12], где он изложил свой подход для нематематиков, Мизес пишет (стр.14)

*«… при помощи рационального понятия вероятности … мы считаем возможным охватить только такие случаи, в которых дело идёт о явлении, могущем быть многократно повторенным; о явлении, осуществлявшемся в огромном количестве экземпляров; физически говоря – о практически **неограниченном ряде однородных** наблюдений.»*

Такую *«совокупность событий или явлений, которые отличаются друг от друга каким-нибудь доступным наблюдению признаком»* Мизес на стр. 16 называет коллективом («kollektiv» в оригинальном издании на немецком языке, «collective» в английском издании его книги) и ещё раз подчёркивает: ***«Сперва должен быть коллектив, тогда только можно говорить о вероятностях».*** □

2.6 Устойчивость частот. Определение вероятности

Наличие массового однородного случайного эксперимента – это не единственное требование для того, чтобы начинать разговор о вероятностях. Чтобы указать важнейшее дополнительное условие (которое одновременно является определением вероятности случайного события), нам потребуется понятие *частоты наступления события*.

Итак, рассмотрим некоторый эксперимент, который можно (теоретически) повторить сколько угодно раз, причём в неизменных условиях. Проведём некоторое число n экспериментов и обозначим через $n(A)$ число экспериментов, в результате которых наступило интересующее нас событие A. Число $n(A)$ назовём *абсолютной* частотой наступления события A в серии, а отношение $\nu_n(A) = n(A)/n$ – *относительной* частотой (или просто частотой) наступления события A в серии.

Теперь мы можем сформулировать то условие, которому должен удовлетворять массовый однородный эксперимент, для того, чтобы можно было говорить о вероятности связанного с ним случайного события.

Условие устойчивости частот. ***Если провести несколько длинных серий экспериментов (возможно разной длины), то частоты наступления события A в разных сериях должны быть примерно равны между собой, т.е. группироваться вокруг некоторого числа*** **P(*A*)*****, которое называют вероятностью случайного события A при осуществлении данного комплекса условий*** $\mathfrak{S}$***. По мере роста длин этих серий отличия между частотами (или, что то же самое, отклонения частот от числа*** **P(*A*)*****) должны уменьшаться.***

Данное только что определение вероятности случайного события называют *частотным определением*, но следует понимать, что это – **единственное определение вероятности случайного события, разумное с естественно-научной точки зрения**; **никакого другого определения нет и быть не может.**

«…о вероятности нельзя сказать ничего больше следующего:

- *вероятность P (A/S) есть число, вокруг которого, при условиях S и при предусмотренных этими условиями способах формирования серий, имеют тенденцию группироваться частоты, причём*
- *при возрастании численности этих серий в разумных пределах, не нарушающих однородности условий, эта тенденция проявляется со все большей отчётливостью и точностью, достигая достаточных в данной конкретной обстановке надёжности и точности при достижимых численностях серий.»*

(А.Н.Колмогоров, [3], стр.275).

Таким образом, позитивный смысл понятия вероятности заключается в следующем. Если нам известна вероятность $\mathsf{P}(A)$ некоторого случайного события A, то в большой серии из n однотипных случайных экспериментов это событие произойдёт примерно

$n\mathsf{P}(A)$ раз. Знание вероятности $\mathsf{P}(A)$ не позволяет сделать никакого вывода о наступлении или нет события A в конкретном эксперименте[3].

Важно понимать, что частота – это не приближённое значение вероятности, которая определяется каким-то другим способом. Смысл слова «вероятность» как научного термина теории вероятностей определяется через частоту. Соответственно, только через частоту *в конечном итоге* должно определяться и значение вероятности (если оно необходимо для решения конкретной реальной задачи). В школьных учебниках часто пишут, что-то вроде следующего: «При большом количестве опытов относительная частота события, как правило, мало отличается от вероятности этого события» (стр.343; С.М.Никольский, М.К.Потапов, Н.Н.Решетников, А.В.Шевкин. Алгебра и начала математического анализа. 10 кл. Учебник для общеобразовательных учреждений (МГУ – школе). 8 изд. М.:Просвещение, 2009), но совершенно неясно, на какой вопрос отвечает это утверждение – на вопрос «Что такое вероятность?» или на вопрос «Чему равна вероятность?»

Следует отметить, что после прочтения любого комплекта учебников по математике складывается весьма хаотическая картина темы «теория вероятностей и статистика», упорядочить которую практически невозможно. Мой опыт работы со школьниками показывает, что после изучения теории вероятностей (по любому комплекту учебников) школьник усваивает (если вообще что-то усваивает) примерно следующее:

- *по определению* вероятность – это отношение числа исходов, благоприятствующих наступлению события, к общему числу исходов (если эти исходы равновозможны; впрочем, школьники обычно на это условие не обращают внимания, считая, что оно выполнено автоматически);
- если эксперимент заключается в бросании точки на какую-то область, то вводится особая вероятность – геометрическая; для этой особой вероятности даётся особое определение, не связанное с «классическим определением»;
- существует ещё «высоконаучное» аксиоматическое определение Колмогорова, которое изучают в университетах;
- при многократном повторении эксперимента частота наступления события приближается к его вероятности (определение которой никак с частотой не связано);
- частотный подход – это лишь способ приближённого вычисления вероятности.

На ОГЭ и ЕГЭ по теме «Теория вероятностей» предлагаются почти исключительно чрезвычайно простые задачи на «классическое определение». Имея в виду значение, которое и школьники, и учителя придают этим экзаменам, в результате изучения основ теории вероятностей в школе ученик в лучшем случае зафиксирует именно «классическое определение» (но может быть лучше было бы вообще ничего не знать про вероятности?).

[3] Из этого утверждения есть только одно исключение. Если вероятность некоторого события очень мала, то мы считаем, что в результате единичного эксперимента оно практически не может произойти. Тем не менее, чтобы говорить о вероятности такого события, его нужно включить в гипотетический стохастический эксперимент.

Важно подчеркнуть, что для любой реальной ситуации устойчивость частот может быть подтверждена ***только*** *экспериментально и* ***только*** *с той или иной точностью*. Тот факт, что некоторый эксперимент можно повторить неограниченное число раз в неизменных условиях, ещё не означает, что любое связанное с ним событие обладает свойством устойчивости частот (т.е. имеет вероятность).

«Предположение, что при данных условиях для данного события вероятность (т.е. вполне определённая нормальная доля числа появлений данного события при большом числе повторений данных условий) существует, является гипотезой, которая в каждом отдельном вопросе требует специальной проверки или обоснования.»

(А.Н.Колмогоров, [4]).

А.Н.Колмогоров в статье [3] массовые однородные эксперименты, обладающие свойством устойчивости частот, называет *вероятностно-случайными* или *стохастическими*[4]. Термин *случайное событие* используется в теории вероятностей ***только*** применительно к стохастическим экспериментам, а термин «событие» употребляется как сокращённая форма термина «случайное событие». Мы не можем отдельно определить термины «случайное событие» (в смысле теории вероятностей) и «вероятность». Вероятностно-случайное событие – это случайное событие, которое имеет вероятность (что предполагает возможность неограниченного повторения эксперимента в неизменных условиях), а вероятность есть только у вероятностно-случайного события (у случайных событий, связанных с уникальными экспериментами, вероятности нет).

Авторы классических современных учебных пособий по теории вероятностей считают необходимым особо подчеркнуть особый смысл, который вкладывает в термины «вероятность», «случайное событие» и т.п. теория вероятностей. Так Б.В.Гнеденко в §1 главы 1 учебника [6] пишет:

«…теория вероятностей занимается изучением не любых событий, которые в житейской практике называются случайными, а только тех из них, которые обладают определёнными свойствами.

Прежде всего, она ограничивается изучением лишь тех событий, которые в принципе могут быть осуществлены неограниченное число раз, притом в неизменных условиях…

Далее, теория вероятностей занимается лишь теми событиями, которые обладают так называемой статистической устойчивостью или, иначе, устойчивостью частот…

[4] Англ. stochastic (от греческого слова stochastikós – умеющий угадывать). В теории вероятностей термин «стохастический» используют (как синоним) вместо слов «случайный», «вероятностный», если возможна их (ошибочная) трактовка в обычном житейском смысле.

…подчеркнём ту мысль, что теория вероятностей не занимается изучением уникальных событий, которые не допускают повторений. Так, не имеет смысла говорить о том, какова вероятность, что данный студент сдаст экзамен по теории вероятностей на ближайшей экзаменационной сессии, поскольку здесь речь идёт о единичном событии, повторить которое в тех же самых условиях нет возможности. Мы можем об этом событии высказывать лишь некоторые субъективные суждения, основанные на нашем знакомстве со знаниями этого студента. Теория же вероятностей ставит перед собой задачу изучения объективных закономерностей, которые не зависят от субъективных суждений того или иного лица. И как бы ни были интересны вопросы, касающиеся единичных, неповторимых событий, теория вероятностей к ним не имеет отношения, если только относительно них нет возможности провести длительные независимые испытания в одинаковых условиях…»

В.Феллер [7] выражает ту же фундаментальную идею следующим образом (§3 Введения «Природа теории вероятностей»):

«Успех современной математической теории вероятностей приобретён следующей ценой: теория ограничивается лишь одной стороной «случайности»…

В нашей системе нет места для догадок о вероятности того, что завтра взойдёт Солнце. Прежде чем говорить о такой вероятности, мы должны были бы условиться о (идеализированной) модели опыта, описание которой предположительно начиналось бы так: «случайным образом выбирается один из бесконечного множества миров…». ***Для того, чтобы построить такую модель, не требуется особого воображения, но она представляется и неинтересной, и незначительной.****»*

Известный специалист по теории вероятностей, заведующий кафедрой теории вероятностей механико-математического факультета МГУ Альберт Николаевич Ширяев (р. 1934 г.) во Введении к учебнику [17] пишет:

«Предметом теории вероятностей является математический анализ случайных явлений – эмпирических феноменов, которые (при заданном «комплексе условий») могут быть охарактеризованы тем, что

для них отсутствует детерминистическая регулярность (наблюдения над ними не всегда приводят к одним и тем же исходам)

и в то же самое время

они обладают некоторой статистической регулярностью (проявляющейся в статистической устойчивости частот).*»*

Чтобы ещё раз объяснить суть понятия «вероятность события», рассмотрим следующую ситуацию. Вы хотите приобрести автомобиль и перед вами в автосалоне стоит два автомобиля – один марки «А», второй – «Б». У них примерно одинаковые технические характеристики и цена. Вы знаете, что автомобиль марки «А» в течение трёх лет (обычный гарантийный срок) потребует существенного ремонта с вероятностью 1%, а автомобиль марки «Б» – 5%. В этой ситуации большинство людей предпочтёт приобрести автомобиль марки «А». Но какой *объективный* смысл имеют указанные вероятности отказа? Если, допустим, автозаводы, выпускающие эти автомобили, производят в год по 100 тысяч машин, то фраза «вероятность существенного ремонта в в течение трёх лет для автомобиля марки «А» равна 1%» означает *только* следующее: из 100 тысяч выпущенных машин этой марки *примерно* 1 тысяча машин потребуют ремонта в течение гарантийного срока, а 99 тысяч – нет. Узнать точные цифры (920 сломалось, а 99080 – нет, или 1072 сломалось, а 98928 – нет, и т.п.) можно только по окончании трёхлетнего периода. Аналогично, фраза «вероятность существенного ремонта в течение трёх лет для автомобиля марки «Б» равна 5%» означает *только* следующее: из 100 тысяч выпущенных машин этой марки *примерно* 5 тысяч машин потребуют ремонта в течение гарантийного срока, а 95 тысяч – нет. Важно понимать, что эта информация совершенно ничего не говорит о качестве тех двух конкретных автомобилей, из которых вы выбираете – она характеризует большие партии машин в целом. Если вы решите купить «более надёжный» автомобиль марки «А», то вам вполне может попасться один из примерно 1000 автомобилей, которые потребуют ремонта. Если вы решите купить «менее надёжный» автомобиль марки «Б», то вам вполне может попасться один из примерно 95000 автомобилей, которые не потребуют ремонта. Поскольку для вас этот эксперимент является уникальным, сказать можно только одно – ваш автомобиль либо сломается во время гарантийного срока, либо нет. В этой ситуации единственно разумным будет детальная диагностика всех систем и узлов каждого из двух предложенных автомобилей, после чего всякая случайность исчезнет – если всё нормально, вы совершенно точно будете знать, что ваш автомобиль не сломается. Но кому же тогда интересны эти вероятности? Ответ очевиден – директорам заводов. Директор первого завода, зная, что примерно 1000 автомобилей потребует гарантийного ремонта может (и должен) заранее зарезервировать соответствующую сумму (если, скажем, в среднем ремонт автомобиля обходится в 1 тысячу долларов, то 1 миллион долларов). Аналогично, директор второго завода, зная, что примерно 5000 автомобилей потребует гарантийного ремонта может (и должен) заранее зарезервировать соответствующую сумму (5 миллионов долларов, если в среднем ремонт обходится в 1 тысячу долларов). Итак, теория вероятностей интересна не потребителям, а директорам. Ранее мы это подчеркнули, сказав, что отличительной чертой случайных событий, изучаемых теорией вероятностей, является следующее: «Осуществилось или нет изучаемое случайное событие A в том или ином конкретном эксперименте из большой серии из n однотипных экспериментов, нам совершенно безразлично – важно только общее число $n(A)$ наступлений изучаемого события A в этой серии».

Этот пример, кстати, показывает, что *абсолютно точное* значение вероятности события в *реальных* ситуациях вовсе не является обязательным; разумная погрешность вполне допустима. Если вероятность отказа автомобиля марки «А» не 1%, а, скажем, 1,03% или 0,98%, то что это изменит в действиях директора завода? Ровным счётом ничего, т.к. с практической точки зрения нет никакой разницы между суммами 1 миллион, 1 миллион 30 тысяч, 980 тысяч.

Имея в виду сказанное выше, вернёмся ещё раз к рассмотренной выше задаче демоварианта ЕГЭ-2018. По мнению авторов этой задачи вероятность того, что школьник сдаст экзамен равна 0.92 (число билетов, не содержащих вопроса о грибах, равно $25-2=23$, так что 23/25=0.92 – искомая вероятность). Однако, в соответствии со сказанным выше о предмете теории вероятностей и смысле термина «вероятность», чтобы придать этой задаче смысл, нужно описать ситуацию, когда школьник много раз (например, 100 раз) сдаёт *один и тот же* экзамен, с *одним и тем же* комплектом билетов, с *неизменными* знаниями и т.д. Тогда задача утверждает, что он сдаст этот экзамен примерно 92 раза и провалит 8 раз – абсурдность такой ситуации очевидна любому здравомыслящему человеку. Ведь школьник сдаёт экзамен только один раз, а если несколько, то с новыми знаниями и до первого успеха (можно поставить вопрос, например, о среднем времени до первого успеха, но разумнее задача от этого не станет). Повторим ещё раз: теория вероятностей применима ТОЛЬКО к МАССОВЫМ экспериментам, которые повторяются много раз в НЕИЗМЕННЫХ условиях, причём результат всей массовой операции обладает свойством УСТОЙЧИВОСТИ ЧАСТОТ. Есть, пожалуй, только один способ, хотя и довольно искусственный, спасти эту задачу. Нужно сказать, что большая группа школьников (например, 100 человек) собирается сдавать экзамен по биологии, причём из 25 билетов ВСЕ школьники не выучили 2 билета про грибы и хорошо выучили 23 остальных билета. Тогда вероятность успеха на экзамене, равная 0,92, имеет смысл: она означает, что примерно 8 человек не сдадут экзамен.

Ясное понимание разницы между уникальными случайными явлениями и стохастическими явлениями безусловно необходимо при изучении теории вероятностей. К сожалению, в школьных учебниках, определяя случайные события так же, как это сделали мы, пишут, что теория вероятностей – это наука, изучающая закономерности случайных событий[5]. Эта, казалось бы незначительная, терминологическая разница играет принципиальную роль и непонимание этого приводит к тому, что в школьных учебниках и на экзаменах предлагается масса абсурдных задач, вроде рассмотренной выше. Я думаю, что это является следствием неверного понимания сути аксиоматического подхода А.Н.Колмогорова к построению теории вероятностей (немного подробнее об этом мы поговорим позже).

Исторические примечания

(1) Первым ясно сформулировал частотное определение вероятности Джон Венн в [11]. В §6 гл. I он дополнил определение серии как набора однородных объектов, некоторые свойства которых меняются от объекта к объекту без всякой системы, следующим требованием:

«… для науки о Вероятности … отличительная характеристика заключается в том, что случайные свойства, в отличие от постоянных, встречаются ***в некоторой постоянной доле от целого числа случаев****. В каждом отдельном случае мы не можем сказать, встретятся они или нет, но по мере того как мы рассматриваем больше случаев, мы обнаруживаем растущую однородность. Мы обнаруживаем, что отношение случаев, в которых они встречаются, к общему числу случаев постепенно колеблется всё*

[5] На самом деле, теория вероятностей – это «математическая наука, позволяющая по вероятностям одних случайных событий находить вероятности других случайных событий, связанных каким-либо образом с первыми» (см. статью «Вероятностей теория» [5] из Математической энциклопедии).

меньше и меньше и непрерывно приближается к некоторому явно фиксированному значению.»

Именно это предельное значение Венн рассматривает в качестве вероятности события и в начале §9 гл.I пишет: *«Я убеждён, что ряд вышеупомянутого вида является основой, на которой должны базироваться все правила Вероятности».*

В §3 гл.I Венн иллюстрирует свойство устойчивости частот следующими примерами:

«Предположим, что мы подбрасываем пенни большое число раз; результаты последовательных бросаний могут рассматриваться как серия. Отдельные бросания из этой серии кажутся происходящими в полном беспорядке; именно этот беспорядок является причиной нашей неопределённости относительно результатов бросаний. Иногда выпадает орёл, иногда – решка; иногда мы наблюдаем повторение одной и той же стороны монеты, иногда нет. До тех пока мы ограничиваем наше наблюдение несколькими бросаниями за один эксперимент, серия кажется просто хаотичной. Но когда мы рассматриваем результат длинной последовательности, мы обнаруживаем поразительное отличие; постепенно начинает проявляться некоторый вид порядка, который в конце концов обретает отчётливую и изумительную форму. Мы обнаруживаем в этом случае, что орлы и решётки появляются примерно в равном числе, что аналогичные последовательности из различных сторон монеты также демонстрируют аналогичное свойство. Короче говоря, несмотря на индивидуальный беспорядок начинает появляться суммарный порядок.

Точно так же, если мы расматриваем продолжительность человеческой жизни, различные люди … составляют серию, демонстрирующую те же свойства. Продолжительность отдельной жизни, как известно, неопределённа, но средняя продолжительность жизни для группы людей становится почти в такой же степени определённой … » □

(2) Мизес в своей основополагающей статье [17] изложил частотное определение вероятности Венна на привычном для математиков языке в следующем виде. Пусть K – некоторый коллектив, т.е. бесконечная последовательность $e_1, e_2, e_3, \ldots$ объектов, которые мы кратко будем называть «элементами». Каждый элемент имеет некоторый «признак», причём не все элементы имеют один и тот признак. Пусть N_A – количество элементов коллектива среди первых N элементов, обладающих признаком A. Коллектив является объектом теории вероятностей, если для него выполнены две аксиомы. Первая аксиома требует, чтобы для любого признака A при $N \to \infty$ существовал предел отношения $\frac{N_A}{N}$. Этот предел мы называем вероятностью появления характеристики A в коллективе K и обозначаем W_A (от немецкого «Wahrscheinlichkeit» – «вероятность»): $W_A = \lim_{N\to\infty} \frac{N_A}{N}$. Вторая аксиома – это требование иррегулярности, о котором мы будем говорить позже.

В книге [12] Мизес изложил свой подход на менее формальном языке. На стр.20 он пишет: *«коллектив есть массовое явление или повторный процесс, короче – ряд единичных наблюдений,* ***при котором обосновано предположение, что относительная частота появления каждого единичного наблюдаемого признака стремится к определённому предельному значению****»* и на стр.20-21 даёт определение вероятности: *«Это предельное значение мы называем «вероятностью появления признака в пределах коллектива».»* Таким образом, Мизес не мыслит вероятности вне коллектива: «*Сперва должен быть коллектив, тогда только можно говорить о вероятностях*».

На стр. 21 Мизес ещё раз подчёркивает: *«о вероятности исхода какой-нибудь битвы в нашем смысле не может быть и речи. Здесь нет налицо коллектива, и наш способ определения непригоден, точно так же, как если бы мы поставили перед физиком задачу измерить механической меркой «работу», которую проделывает актёр, исполняя свою роль.»*

□

Используя описанную в начале этого раздела схему с массовым однородным экспериментом, проводимым в неизменных условиях $\mathfrak{S}$, частотный подход Венна-Мизеса можно было бы описать следующим образом.

Проведём n ***раз наш эксперимент и обозначим через*** $n(A)$ ***число опытов, в которых наблюдалось событие*** A***. Если по мере роста числа n экспериментов частота*** $\nu_n(A) = \dfrac{n(A)}{n}$ ***появления события*** A ***приближается к некоторому числу*** $\mathsf{P}(A)$***, то это число мы и называем вероятностью*** $\mathsf{P}(A)$ ***случайного события A при осуществлении данного комплекса условий*** $\mathfrak{S}$***. Иначе говоря,***

$$\mathsf{P}(A) = \lim_{n\to\infty} \nu_n(A) \equiv \lim_{n\to\infty} \frac{n(A)}{n}. \tag{1}$$

Именно этого определения вероятности события мы и будем придерживаться (позже мы уточним его, включив требование отсутствия регулярности).

В средней школе при определении вероятности соотношением (1) лучше не использовать понятие предела в том смысле, в котором оно определено в курсе математического анализа, а применять равенство $a = \lim_{n\to\infty} x_n$ просто как краткую форму записи фраз «последовательность чисел x_n по мере роста n неограниченно приближается к числу a», «при большом *n* значение переменной x_n приблизительно равно a, причём при неограниченном росте *n* приближённое равенство $x_n \approx a$ превращается в точное», «при достаточно большом n практически верно точное равенство $x_n = a$» и т.п. При этом запись $n \to \infty$ следует понимать просто как краткую форму записи фраз «при большом числе *n*», «при неограниченном росте *n*» и т.п. Соответственно, определение вероятности (1) можно интерпретировать, например, так: «при большом числе n экспериментов частота $\nu_n(A)=\dfrac{n(A)}{n}$ появления события A приблизительно равна $\mathsf{P}(A)$, причём при неограниченном росте n приближённое равенство $\nu_n(A)\approx\mathsf{P}(A)$ превращается в точное» или «при достаточно большом числе экспериментов практически верно точное равенство $\nu_n(A)=\mathsf{P}(A)$» и т.п.

Рихард Мизес на стр.19-20 книги [12] предлагает объяснить смысл фразы «вероятность – это предел частот при неограниченном росте числа экспериментов» следующим образом (разбиение на абзацы приведённой ниже цитаты – наше):

«Мы будем вычислять относительную частоту признака, т.е. частное от деления числа его появлений на общее число всех наблюдений, с определённым числом десятичных знаков, не заботясь о том, каковы будут дальнейшие знаки: будут ли они нулями или другими цифрами. Если мы, таким образом, достаточно долго проследим повторное явление, например игру в орлянку, и время от времени будем вычислять частоту выпадения герба с установленной точностью, то может статься, что, начиная с какого-либо момента частота перестанет изменяться.

Если мы ограничимся только одним десятичным знаком, то обычно сравнительно скоро, может быть после 500 бросаний, мы увидим, что не наступает более никаких изменений, что относительная частота с точностью до одного знака точно составляет, скажем, 0,5.

Если вычисляешь два знака, то приходится пойти дальше… и может быть таким образом найдём, что при этом, начиная с 8000 бросаний, на втором месте десятичной дроби всегда стоит «0», и, следовательно, относительная частота составляет 0,50.

Конечно, такой опыт нельзя продолжать до бесконечности. Если два человека будут достаточно искусно работать вместе, то они могут довести число наблюдений до 1000 в час. Теперь предположим, что на это дело потрачено 10 часов и замечено, что в последние два часа относительная частота остановилась на 0,50, не принимая во внимание дальнейших знаков. Более точный исследователь вычислит при этом и третий знак частного и, может быть, найдёт, что он в самый последний час хотя и обнаруживал изменения, но небольшие.

Такой результат опыта естественно приводит научно мыслящий ум к ***предположению****, что если всё дальше и дальше продолжать игру при тех же условиях, то третий, четвёртый и наконец каждый следующий знак относительной частоты в конце концов примет определённое значение.*

Положение вещей, соответствующее этому предположению, мы коротко выражаем в словах: относительная частота признака «герб» ***стремится к предельному значению****.»*

Аппелируя к понятию движения, которое интуитивно ясно каждому человеку, можно проиллюстрировать понятие вероятности как предела частот следующим образом.

Проведём некоторый гипотетический эксперимент, скажем, $n = 20$ раз (в неизменных условиях). Результат этой серии экспериментов приведёт к такой, например, последовательности: **АА***А**АА****А*А***. Появление буквы A на втором, третьем, седьмом и т.д. местах означает, что в соответствующих экспериментах наблюдалось событие A. Звёздочка означает, что в соответствующих экспериментах обытие A не наблюдалось (т.е. наблюдалось событие $\bar{A}$, противоположное событию A). Поэтому *в этой* серии из $n = 20$ экспериментов частота наступления события A есть $\nu_{20}(A) = \dfrac{7}{20} = 0.35$. *В другой* серии из $n = 20$ экспериментов, *проведённых в тех же условиях*, частота наступления события A *может* быть иной (и скорее всего будет иной); при другом числе экспериментов частота тем более может быть иной. В таблице показано, как менялась частота $\nu_n(A)$ по мере роста числа n проведённых экспериментов; в первом столбце указан номер эксперимента, во втором – его результат, в третьем – частота появления события A в первых n экспериментах (десятичное значение округлено до сотых).

n		$\nu_n(A)$	n		$\nu_n(A)$	n		$\nu_n(A)$	n		$\nu_n(A)$
1	*	0	6	*	2/6=0.33	11	*А*	5/11=0.45	16	*А*	6/16=0.38
2	*А*	1/2=0.50	7	*А*	3/7=0.43	12	*	5/12=0.42	17	*	6/17=0.35
3	*А*	2/3=0.67	8	*	3/8=0.38	13	*	5/13=0.38	18	*А*	7/18=0.39
4	*	2/4=0.50	9	*	3/9=0.33	14	*	5/14=0.36	19	*	7/19=0.37
5	*	2/5=0.40	10	*А*	4/10=0.40	15	*	5/16=0.31	20	*	7/20=0.35

Если последовательные значения частоты отмечать на числовой оси, то мы увидим точку, которая скачками движется по отрезку $[0;1]$. Вначале эти скачки довольно значительны, а затем становятся всё меньше (см. рис.1). Равенство (1) означает, что эта движущаяся точка стремится к некоторому предельному положению, которое и объявляется вероятностью события A.

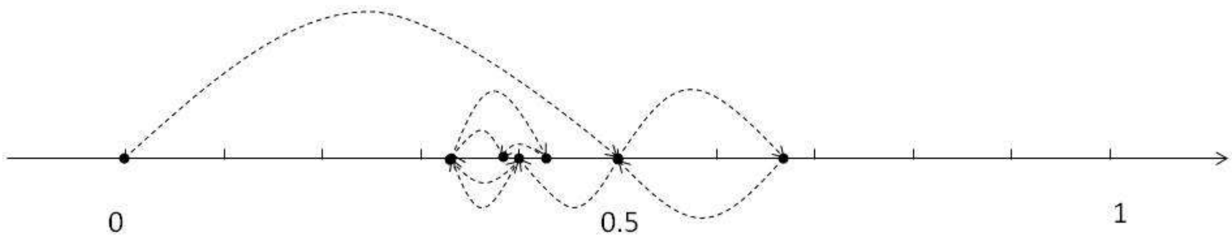

Рис. 1

Конечно, соотношение (1) не позволяет найти точное числовое значение вероятности события. Но для практического применения теории вероятностей *точные* значения вероятностей событий и не нужны; вполне достаточно их приближённого значения, тем более, что математическая теория вероятностей сама является лишь моделью, т.е. приближённым образом, реальной ситуации. Более того, для развития *теории* вероятностей, т.е. доказательства разнообразных формул для вероятностей событий, которые тем или иным способом связаны между собой, важно лишь то, что эти вероятности *существует*.

Теория Мизеса в её первоначальном виде естественно ставит ряд дополнительных вопросов, изучению которых посвящены многочисленные современные исследования по логике, теории алгоритмов, самой теории вероятностей и т.д. (см, например, [19]). Многие неспециалисты интерпретировали эти вопросы как доказательство ошибочности подхода Мизеса и стали абсолютизировать подход Колмогорова [1], основанный на теории меры.

Фактически этот же подход используется при изложении теории вероятностей в школе. Однако по поводу своей аксиоматики сам Колмогоров отмечал: *«В изложении необходимых предпосылок для приложимости теории вероятностей к миру действительных событий автор в значительной мере следует выводам Мизеса …»* ([1], стр. 12).

Вне всякого сомнения, методически частотный подход является лучшим способом начать преподавание теории вероятностей. Именно так излагают теорию вероятностей на элементарном уровне Б.В.Гнеденко и А.Я.Хинчин [2], А.Н.Колмогоров [3].

Д.Пойа в книге [21], гл. XIV, говоря о теории вероятностей, на стр. 281 отмечает:

«… её можно представлять себе и вводить различными путями. Некоторые авторы рассматривают её как чисто математическую теорию, другие – как вид или ветвь логики, а ещё одни – как часть науки о природе. …мы выбираем точку зрения, которая наиболее удобна в огромном большинстве приложений и которой начинающий может овладеть наиболее быстро. Мы рассматриваем … теорию вероятностей как часть науки о природе, как теорию некоторых доступных наблюдению явлений, ***случайных массовых явлений****. … настоящее изложение следует взглядам Р.Мизеса… »*

на стр. 290: *«Теория вероятностей, как мы её понимаем, есть часть науки о природе, теория массовых случайных явлений.*

Наиболее поразительным достижением физических наук является предсказание. Астрономы предсказывают солнечные и лунные затмения, положение планет и возращение комет, которые на несколько лет ускользают от наблюдений. Великому астроному (Леверье) удалось даже предсказать положение планеты (Нептун), самое существование которой ранее было неизвестно. Теория вероятностей с известным успехом предсказывает частоты в некоторых массовых явлениях.

Астрономы основывают свои предсказания на прежних наблюдениях, на законах механики, законе тяготения и на длинных трудных вычислениях. Любая из физических *наук основывает свои предсказания на какой-то теории или, можно сказать, на каком-то предположении, так как никакая теория не является несомненной, и, таким образом, каждая теория есть более или менее разумное, более или менее хорошо подкреплённое предположение. Пытаясь предсказать частоты в некотором случайном массовом явлении исходя из теории вероятностей, мы должны сделать какое-то теоретическое допущение относительно этого явления. Такое допущение, которое должно быть выражено на языке вероятности, называется* ***статистической гипотезой****.»*

Важное замечание об устойчивости частот. Понятие устойчивости частот, кроме равенства (1) включает ещё одно важное условие, которое мы объясним на примере.

Очень часто на кабинетах врачей можно увидеть следующее объявление: «в понедельник, среду, пятницу приём утром, а во вторник, четверг, субботу – вечером» (будем считать для простоты, что у врача один выходной в неделю – воскресенье). В этой ситуации ни один человек не скажет, что время приёма (утро или вечер) является случайным. Вот если бы врач написал что-нибудь вроде следующего: «каждое утро, после завтрака, я бросаю монету и если выпадет орёл, то приём будет утром, а если решка – то вечером». Тогда никто бы не усомнился, что время приёма (утро или вечер) является случайным.

Предположим, что начиная с сегодняшнего числа мы стали фиксировать, наступило или нет событие $A =$ «врач принимает утром» – если оно наступило, напишем 1, а если нет – 0. Рассматривая только рабочие дни и считая для определённости, что сегодня понедельник, в первом случае мы получим следующую последовательность: $101010101010\ldots$ Устроена эта последовательность очень просто – на нечётных местах стоят единицы, а на чётных нули. Если $n = 2k$ – чётное, то эта последовательность состоит из k пар вида (1,0). Поэтому число появлений события A после первых n опытов равно $k = \frac{n}{2}$, а частота $\nu_n(A)$ наступления события A равна $1/2$. Если же $n = 2k+1$ – нечётное, то эта последовательность состоит из k пар вида (1,0) и ещё одной единицы. Поэтому число появлений события A после первых n экспериментов равно $k+1 = \frac{n+1}{2}$, а частота $\nu_n(A)$ наступления события A равна $\frac{1}{2} + \frac{1}{2n}$. Ясно, что при $n \to \infty$ частота $\nu_n(A)$ имеет предел, равный $1/2$. Итак, в явно неслучайном эксперименте наблюдается устойчивость частот, но вряд ли разумно назвать предел частот, равный $1/2$, вероятностью наступления события «приём утром».

Во втором случае, когда врач устанавливает время приёма, бросая монету, мы получим «хаотичную» последовательность вида $0110101100001011\ldots$ (а может быть такую: $1100010101110011\ldots$ или ещё что-нибудь в этом духе) и никто не сомневается в том, что

- мы имеем дело со случайным экспериментом;
- имеет смысл говорить о вероятности события «приём утром», причём эта вероятность равна $1/2$ (т.к. частоты $\nu_n(A)$, очевидно, имеют предел, равный $1/2$).

Итак, и в первой, и во второй ситуации:

- мы имеем дело с (теоретически) бесконечной последовательностью опытов, проводимых в одинаковых условиях;
- частота наступления интересующего нас события стремится к пределу (равному $1/2$) при неограниченном росте числа опытов.

В чём же разница между ними? Почему только во втором случае мы считаем эксперимент случайным?

В двух словах ответ на эти вопросы заключается в следующем. Выберем из бесконечной последовательности 0 и 1, описывающей результаты отдельных опытов, бесконечную *под*последовательность. Выбрать подпоследовательность – значит указать номера тех опытов, результаты которых будут включены в подпоследовательность; результаты опытов, не включённых в подпоследовательность мы будем игнорировать. Например, мы можем рассматривать только результаты экспериментов с нечётными номерами. Возможны и более сложные правила, например, мы можем брать результаты только тех экспериментов, перед которыми наступило некоторое другое событие.

Если эксперимент является действительно случайным, то для любой такой подпоследовательности частота наступления интересующего нас события A имеет тот же предел $\mathsf{P}(A)$, что и для исходной последовательности. Если же для некоторой подпоследовательности предел частот будет другим, то это означает определённую закономерность, регулярность, в результатах бесконечной серии опытов. Соответственно, мы не называем эксперимент случайным.

Следует отметить, что даже для явно случайного эксперимента можно указать такие правила формирования подпоследовательности, что предел частот изменится. Рассмотрим, например, последовательность 0 и 1, соответствующую бесконечной серии бросаний правильной монеты (1 означает появление «орла»). Если мы будем игнорировать все 0, то получим подпоследовательность из одних единиц, для которой предел частот равен 1 (а не $1/2$ как для исходной последовательности). Поэтому включение или невключение результата отдельного опыта в подпоследовательность не должно зависеть от самого этого результата.

Вопрос о том, какие правила образования подпоследовательностей допустимы, является *чрезвычайно* сложным. Поискам ответа на него посвящены многочисленные научные исследования (см., например, [19]). На физическом уровне замечательное объяснение дал Мизес на стр. 31 книги [12] (раздел «Иррегулярность коллектива»):

«… невозможность оказать влияние на шансы выигрыша системой выбора, ***невозможность системы игры*** *есть характернейшее и решающее свойство тех рядов наблюдений или массовых явлений, которые составляют подлинный предмет исчисления вероятностей. К коллективу, к которому должно быть применимо исчисление вероятностей, мы предъявляем два требования: во-первых, относительная частота, с которой выступает в ряду какой-либо определённый признак, должна обладать предельным значением; во-вторых,* ***это предельное значение должно оставаться неизменным, если из всей последовательности произвольно выбрать любую часть и рассматривать в дальнейшем только эту часть.*** *Само собой разумеется, что выбранная частичная последовательность должна быть безграничной, как и сама основная*

последовательность, например так, что оставляется каждый второй или каждый десятый член совокупности, или каждый, нумер которого есть полный квадрат, или простое число, и т. д. В остальном при составлении выбора фантазии предоставляется полная свобода. Нужно только иметь возможность решать вопрос о принадлежности или непринадлежности каждого отдельного бросания к выбранной частичной последовательности, ***независимо от результата этого бросания****, т.е. при помощи некоторого правила, относящегося только к нумеру данного бросания, и установленного прежде, чем стал известен его исход. Поэтому мы кратко называем выбор в том смысле, какой здесь подразумевается, «выбором нумеров» и, следовательно, требуем, чтобы предельные значения относительных частот внутри коллектива были* ***нечувствительны по отношению к любому выбору нумеров.****»*

Резюмируя всё сказанное выше, можно сказать, что в теории вероятностей, рассматриваемой как наука о специфических закономерностях реального мира, фраза «событие A имеет вероятность $\mathsf{P}(A)$» означает следующее:

1. можно очень много раз, практически неограниченно, повторять комплекс условий S, относительно которого рассматривается событие A (без существенного изменения этого комплекса);

2. после каждого эксперимента можно совершенно точно установить, осуществилось или нет событие A;

3. при большом числе n реализаций комплекса условий S число $n(A)$ тех экспериментов, в которых наблюдалось событие A, примерно равно $n\mathsf{P}(A)$;

4. если из большой серии результатов эксперимента при помощи заранее установленного правила (учитывающего отмеченные выше ограничения) выбрать подсерию большой длины m, то для этой подсерии число тех экспериментов, в которых наблюдалось событие A, также примерно равно $m\mathsf{P}(A)$.

2.7 О приложениях теории вероятностей

Преподавание теории вероятностей (как, впрочем, и любого другого раздела науки) немыслимо без иллюстрации её пользы для решения реальных задач естествознания, техники, бизнеса и т.д. К сожалению, в настоящее время университетские курсы теории вероятностей являются лишь абстрактными математическими курсами. Эта деградация науки о случайном началась уже давно.

Ещё в 1923 году профессор Московского университета Владимир Александрович Костицын в Предисловии к книге Бореля [13] писал:

«… нужно всегда хорошо знать и помнить, какова ценность законов случая, какова граница их приложимости… К сожалению, существующие курсы теории вероятности

излагают лишь формально-математическую сторону дела и ничего не говорят о значении теории вероятности как метода научного познания.»

В 1977 г. один из лучших специалистов по теории вероятностей, профессор кафедры теории вероятностей механико-математического факультета МГУ Валерий Николаевич Тутубалин (р. 1936 г.) [20], отмечал: *«начинает обрисовываться традиция (полностью господствующая сейчас в преподавании математического анализа и ряда других математических дисциплин), которая резко отделяет математическое и прикладное содержание. В начале века учебники теории вероятностей были полны реальных примеров статистических данных; в новых же учебниках такие примеры исчезают.»*

Нельзя сказать, что авторы школьных учебников не понимают сказанного выше. Поэтому в школьных учебниках (формально) достаточно примеров применения теории вероятностей. Но что это за применения?

Стандартными примерами экспериментов, которые можно неограниченно повторять в неизменных условиях и которые обладают свойством устойчивости частот, связанных с ним случайных событий, являются бросание игрального кубика, монеты, выбор карты из колоды, рулетка. Число задач в школьных учебниках на расчёт вероятностей событий, связанных с этими азартными играми, настолько велико, что поневоле закрадывается мысль, не финансировала ли включение теории вероятностей в школьный курс игорная мафия.

Между тем оправдать изучение теории вероятностей в средней школе можно только серьёзными (а не надуманными) применениями её результатов в физике, химии, экономике и т.д.

Б.В.Гнеденко и А.Я.Хинчин в книге [2], А.Н.Колмогоров в статье [1] используют для иллюстрации понятий и теорем теории вероятностей простейшие задачи теории стрельбы. Это и не удивительно. Только что окончилась Великая Отечественная Война, началась война холодная и важность таких задач была очевидна всем. Наряду с этим использовались примеры из массового промышленного производства, что также вполне объяснимо (и чрезвычайно актуально сейчас).

Одной из самых важных областей применения теории вероятностей является страхование. Школьники вполне могут понять, как определяется стоимость простейших договоров страхования.

В приложениях теории вероятностей важно понимать следующее. Для реальных задач устойчивость частот наступления тех или иных событий, т.е. существование вероятностей этих событий, и значения вероятностей устанавливаются только в ходе экспериментов. Это даёт основание применять для расчёта вероятностей более сложных событий, связанных с изучаемым экспериментом, теоремы математической теории вероятностей. Однако, поскольку в реальности устойчивость частот и сами значения вероятностей исходных событий можно установить только приблизительно, нельзя гарантировать, что выводы, полученные с помощью этих теорем, применительно к изучаемому эксперименту верны хотя бы приблизительно (лучше сказать, с той степенью точности, с которой установлена устойчивость частот) – при удлиннении цепочки логических умозаключений и увеличении числа операций, производимых с исходными вероятностями (которые в реальных задачах ***всегда*** известны *только приблизительно*), точность получаемых значений и достоверность окончательных выводов уменьшается. Поэтому всегда нужно проверять, хотя бы частично, соответствие теоретических выводов реально наблюдаемым фактам. В этом отношении теория вероятностей ничем не отличается от других математических наук. Очень хорошо сказал по этому поводу знаменитый математик В.И.Арнольд [10]:

«При построении модели происходит следующая идеализация: некоторые факты, известные лишь с некоторой долей вероятия или лишь с некоторой точностью, признаются «абсолютно» верными и принимаются за «аксиомы». Смысл этой «абсолютности» состоит ровно в том, что мы позволяем себе оперировать с этими «фактами» по правилам формальной логики, объявляя «теоремами» всё то, что из них можно вывести. Понятное дело, что ни в какой реальной деятельности полностью полагаться на подобные дедукции невозможно. Причиной является хотя бы то, что параметры изучаемых явлений никогда не бывают известными нам абсолютно точно, а небольшое изменение параметров (например, начальных условий процесса) может совершенно изменить результат…Совершенно таким же образом небольшое изменение аксиом (в которых ведь мы точно уверены быть не можем) способно, вообще говоря, привести к иным выводам, чем дают выведенные из принятых аксиом теоремы. И чем длиннее и искуснее цепь выводов («доказательств»), тем менее надёжен окончательный результат…«Тонкий яд математического образования» (по выражению Ф.Клейна)… состоит именно в том, что абсолютизируемая модель отрывается от реальности и перестаёт с нею сравниваться…»

3. Операции над событиями

Основное содержание теории вероятностей заключается в разработке методов вычисления вероятностей одних случайных событий (относительно сложных) с помощью вероятностей других случайных событий (более простых), которые как-то связаны с первыми. Вероятности вторых, более простых, случайных событий в подавляющем большинстве реальных приложений теории вероятностей оценивают исходя из экспериментальных данных, проводя массовые однородные эксперименты. После этого с помощью формул теории вероятностей вычисляют вероятности более сложных событий (слово «случайных» в теории вероятностей обычно опускают), связанных с более простыми событиями, уже не проводя никаких экспериментов. Следует, однако, отметить, что общая идеология применения математики требует оценки точности полученных таким способом выводов, в частности, за счёт проведения тестовых экспериментов.

Поэтому перед изложением формул теории вероятностей нужно определить (а школьнику *хорошо усвоить*) основные типы связи между событиями. Можно даже сказать, что в изучении основ теории вероятностей школьнику удастся продвинуться ровно настолько, насколько далеко он продвинется в изучении свойств событий и операций над ними. Обратим внимание на то, что понятие события и операции над событиями имеют смысл и для уникальных экспериментов (но говорить о вероятности события можно только для стохастических экспериментов). Как мы отмечали, при начальном изучении теории вероятностей термин «событие» лучше понимать на бытовом уровне. Конспективно эта своеобразная «теория событий» может быть изложена, например, следующим образом.

Если в результате реализации комплекса условий $\mathfrak{S}$ события A и B наступают или нет одновременно, то такие события называют *равносильными* или *равными* и пишут: $A = B$. В реальных ситуациях равносильные события выражают один и тот результат эксперимента, но разными словами. Если, например, «эксперимент» заключается в сдаче экзамена, то событие «экзамен не сдан» равносильно событию «оценка за экзамен – 2»:

«экзамен не сдан»= «оценка за экзамен – 2».

Если при наступлении события A (после реализации комплекса условий $\mathfrak{S}$) обязательно наступает и событие B, то событие B называется *следствием* события A (говорят также, «A *влечёт* B», «A *– частный случай* B», «A *благоприятствует* B») и пишут: $A \subset B$. Если, например, «эксперимент» заключается в сдаче экзамена, то событие «оценка за экзамен – 4» влечёт событие «экзамен сдан»:

«оценка за экзамен – 4» $\subset$ «экзамен сдан».

Событие, которое происходит при каждой реализации комплекса условий $\mathfrak{S}$, называется *достоверным событием*. По смыслу этого определения достоверных событий может быть много, но все достоверные события равносильны во введённом выше смысле и фактически разными словами выражают один и тот результат эксперимента. Например, если эксперимент заключается в сдаче экзамена, то достоверными будут события:

«получена оценка 2 или 3 или 4 или 5», «полученная оценка или положительна, или отрицательна», «экзамен сдан или провален». Можно поэтому сказать, что достоверное событие (для данного комплекса условий) только одно; его обозначают Ω.

Событие, которое не происходит ни при одной реализации комплекса условий $\mathfrak{S}$, называется *невозможным событием*. Все невозможные события равносильны во введённом выше смысле и фактически выражают один и тот результат эксперимента, но разными словами. Можно поэтому сказать, что невозможное событие (для данного комплекса условий) только одно; его обозначают $\varnothing$, что и пустое множество. Если событие A не является невозможным, мы часто будем называть его *непустым* и писать $A \neq \varnothing$.

Проверка равенства двух событий обычно проводится с помощью следующего простого утверждения.

Утверждение 1. События A и B равны тогда и только тогда, когда при наступлении события A наступает и событие B, а при наступлении события B наступает и событие A:

$$A = B \Leftrightarrow (A \subset B \text{ и } B \subset A). \qquad (2)$$

Для каждого события A можно рассмотреть *противоположное* (или *дополнительное) событие* $\bar{A}$, которое заключается в том, что событие A не происходит. Для достоверного события противоположным будет невозможное, а для невозможного – достоверное: $\bar{\Omega} = \varnothing$, $\bar{\varnothing} = \Omega$. Дополнительное событие также называют *отрицанием* события A, *дополнением* A и т.п.

Событие, заключающееся в одновременном наступлении событий A и B, называется их *произведением* (используются и термины *пересечение*, *совмещение*) и обозначается AB или $A \cap B$. Например, если эксперимент заключается в фиксации числа дорожно-транспортных происшествий за день, то совмещение событий «число ДТП меньше 5» и «число ДТП больше 3» можно выразить словами «число ДТП равно 4»:

$$\text{«число ДТП меньше 5»} \cap \text{«число ДТП больше 3»} = \text{«число ДТП равно 4»}.$$

Аналогично можно определить произведение любого числа событий.

Если события A и B не могут наступить одновременно, то они называются *несовместными* (или *несовместимыми*). Это же определение можно выразить словами «одновременное наступление событий A и B является невозможным событием», что позволяет определить несовместные события равенством $AB = \varnothing$.

Событие, заключающееся в том, что событие A происходит, а событие B нет, называется *разностью* событий A и B и обозначается $A \setminus B$. Нетрудно сообразить, что $A \setminus B = A\bar{B}$.

Событие, заключающееся в наступлении хотя бы одного из событий A и B, называется их *объединением* (используются и другие термины, например, *соединение*) и обозначается $A \bigcup B$; при этом считается, что одновременное наступление событий A и B (если оно вообще возможно) также означает и наступление события $A \bigcup B$.

Аналогично можно определить объединение любого числа событий. Если, например, «эксперимент» заключается в сдаче экзамена, то сумму событий «оценка 3», «оценка 4», «оценка 5» можно выразить словами «экзамен сдан»:

$$\text{«оценка 3»} \bigcup \text{«оценка 4»} \bigcup \text{«оценка 5»} = \text{«экзамен сдан»}.$$

Если в объединении $A_1 \cup A_2 \cup \ldots \cup A_k$ события $A_1, A_2, \ldots, A_k$ попарно несовместны, то мы вместо термина «объединение» мы будем использовать термин «сумма», а вместо символа $\bigcup$ обычный символ сложения «+». Иначе говоря, фраза «сумма событий $A_1, A_2, \ldots, A_k$» и запись $A_1 + A_2 + \ldots + A_k$ автоматически предполагают, что накакие два события из списка $A_1, A_2, \ldots, A_k$ не могут наступить одновременно. Такое соглашение принято, например, в [1], [18]. В других пособиях, например, [6], термин «сумма событий» и символ «+» используются просто как синонимы термина «объединение событий» и символа $\bigcup$ соответственно; если же речь идёт об объединении попарно несовместных событий, то это оговаривается особо.

Если какое-то событие A можно представить как сумму попарно несовместных событий $A_1, \ldots, A_k$, то говорят, что событие A *разбивается* (или *раскладывается*) *на частные случаи* $A_1, \ldots, A_k$. Можно также сказать, что набор событий $A_1, \ldots, A_k$ – это *разбиение* (*разложение*) события A.

Набор событий $A_1, \ldots, A_N$ называют *полной системой событий*, если объединение событий $A_1, \ldots, A_N$ является достоверным событием. Иначе говоря, система событий $A_1, \ldots, A_N$ является полной, если в результате эксперимента обязательно произойдёт одно из этих событий. При этом нельзя исключить одновременного наступления нескольких событий из набора $A_1, \ldots, A_N$. На языке формул определение полной системы событий записывается равенством $A_1 \cup A_2 \cup \ldots \cup A_N = \Omega$ или, если символы $\bigcup$ и $+$ используются как равносильные, равенством $A_1 + A_2 + \ldots + A_N = \Omega$.

Полная система попарно несовместных событий $A_1, \ldots, A_N$ – это *разбиение* достоверного события; в этой ситуации сами события $A_1, \ldots, A_N$ называются *исходами эксперимента* или *элементарными событиями*. Итак, разбиение Ω – это такой набор событий, что в результате эксперимента обязательно произойдёт одно из этих событий и притом только одно (никакие два события системы не могут произойти одновременно); однако не исключено, что какое-то из событий системы никогда не произойдёт. Если символ «+» для обозначения объединения предполагает попарную независимость слагаемых, то определение разбиения достоверного события (или, другими словами, полной системы несовместных событий) записывается равенством $A_1 + A_2 + \ldots + A_N = \Omega$.

Следует иметь в виду, что в ряде учебников, например, в [2], понятие полной системы уже включает требование попарной несовместности составляющих систему событий.

Предположим, что для некоторого эксперимента мы разложили достоверное событие на исходы $A_1, \ldots, A_N$. Если событие A является суммой каких-то k исходов $A_{n_1}, \ldots, A_{n_k}$, т.е. наступление A означает наступление одного и только одного из этих исходов, то эти исходы *благоприятствуют* наступлению события A; остальные $N-k$ исходов не благоприятствуют наступлению события A.

Для развития сколь-нибудь содержательной теории безусловно необходимы следующие простейшие свойства операций над событиями.

Утверждение 2. Событие, противоложное к событию, которое само является дополнением к некоторому событию A, равносильно исходному событию A:

$$\overline{\overline{A}} = A. \tag{3}$$

Утверждение 3. Событие, противоложное к объединению событий A и B, равно произведению событий, противоположных к A и B:

$$\overline{A \cup B} = \overline{A} \cap \overline{B} \Leftrightarrow \overline{A \cup B} = \overline{A}\overline{B}. \tag{4}$$

Утверждение 4. Событие, противоложное к произведению событий A и B, равно объединению событий, противоположных к A и B:

$$\overline{A \cap B} = \overline{A} \cup \overline{B} \Leftrightarrow \overline{AB} = \overline{A} \cup \overline{B}. \tag{5}$$

Утверждение 5. Произведение некоторого события A и объединения событий B и C равно объединению двух произведений – произведения событий A и B и произведения событий A и C:

$$A(B \cup C) = (AB) \cup (AC). \tag{6}$$

4. Основные теоремы элементарной теории вероятностей

В этом разделе мы получим основные результаты элементарной теории вероятностей с помощью частотного определения вероятности.

Теорема 1. Вероятность $\mathsf{P}(A)$ любого случайного события A удовлетворяет неравенствам:

$$0 \le \mathsf{P}(A) \le 1. \tag{7}$$

Доказательство. Проведём n раз эксперимент, относительно которого рассматривается событие A, и обозначим через $n(A)$ число экспериментов, в которых наблюдалось это событие. Ясно, что число $n(A)$ неотрицательно и не превосходит общего числа проведённых экспериментов: $0 \le n(A) \le n$. Поэтому частота $\nu_n(A) = \dfrac{n(A)}{n}$ появления события A удовлетворяет двойному неравенству $0 \le \nu_n(A) \le 1$. При достаточно большом числе испытаний частота $\nu_n(A)$ практически равна вероятности $\mathsf{P}(A)$, так что и вероятность $\mathsf{P}(A)$ удовлетворяет двойному неравенству $0 \le \mathsf{P}(A) \le 1$.

Можно провести более аккуратное рассуждение: если бы точка $\mathsf{P}(A)$ лежала вне отрезка $[0;1]$, то при достаточно большом числе испытаний и частоты $\nu_A(n)$ (которые неограниченно приближаются к $\mathsf{P}(A)$) лежали бы вне этого отрезка, что невозможно.

Эти наглядные рассуждения можно заменить формальным переходом к пределу при $n \to +\infty$ в двойном неравенстве $0 \le \dfrac{n(A)}{n} \le 1$:

$$0 \le n(A) \le n \Leftrightarrow 0 \le \frac{n(A)}{n} \le 1 \Rightarrow 0 \le \lim_{n\to+\infty} \frac{n(A)}{n} \le 1 \Leftrightarrow 0 \le \mathsf{P}(A) \le 1.$$

Теорема 2. Вероятность $\mathsf{P}(\Omega)$ достоверного события равна 1:

$$\mathsf{P}(\Omega) = 1, \tag{8}$$

а вероятность $\mathsf{P}(\varnothing)$ невозможного события равна 0:

$$\mathsf{P}(\varnothing) = 0. \tag{9}$$

Доказательство. Проведём n раз эксперимент $\mathfrak{S}$. Поскольку событие Ω является достоверным, число экспериментов, в которых оно наблюдалось, равно общему числу экспериментов: $n(\Omega)=n$. Поэтому частота $\nu_n(\Omega)=\dfrac{n(\Omega)}{n}$ появления достоверного события равна 1. При росте n эта частота не будет меняться и всё время будет равна 1. Поэтому предельное значение частот $\nu_n(\Omega)$, т.е. число $\mathsf{P}(\Omega)$, также будет равно 1.

Для невозможного события рассуждения аналогичны:

$$n(\varnothing)=0 \Leftrightarrow \nu_n(\varnothing)\equiv\frac{n(\varnothing)}{n}=0 \Rightarrow \lim_{n\to+\infty}\nu_n(\varnothing)=0 \Leftrightarrow \mathsf{P}(\varnothing)=0.$$

Теорема 3. Вероятность $\mathsf{P}\left(\bar{A}\right)$ события, противоположного событию A, даётся формулой

$$\mathsf{P}\left(\bar{A}\right)=1-\mathsf{P}(A). \qquad (10)$$

Доказательство. Проведём n раз эксперимент $\mathfrak{S}$. В некоторых из этих экспериментов будет наблюдаться событие A – их общее число мы обозначаем $n(A)$. В оставшихся $n-n(A)$ экспериментах будет наблюдаться противоположное событие $\bar{A}$. Поэтому частота $\nu_n\left(\bar{A}\right)=\dfrac{n\left(\bar{A}\right)}{n}$ появления события $\bar{A}$ равна $\dfrac{n-n(A)}{n}=1-\dfrac{n(A)}{n}=1-\nu_n(A)$.

При росте n частоты $\nu_n(A)$ и $\nu_n\left(\bar{A}\right)$ будут неотличимы от соответствующих вероятностей ($\mathsf{P}(A)$ и $\mathsf{P}\left(\bar{A}\right)$ соответственно), так что равенство $\nu_n\left(\bar{A}\right)=1-\nu_n(A)$ превратится в равенство (10).

Эти наглядные рассуждения можно заменить формальным «переходом к пределу при $n\to+\infty$» в равенстве $\nu_n\left(\bar{A}\right)=1-\nu_n(A)$:

$$n\left(\bar{A}\right)=n-n(A) \Leftrightarrow \nu_n\left(\bar{A}\right)=1-\nu_n(A) \Rightarrow \lim_{n\to+\infty}\nu_n\left(\bar{A}\right)=1-\lim_{n\to+\infty}\nu_n(A) \Leftrightarrow \mathsf{P}\left(\bar{A}\right)=1-\mathsf{P}(A).$$

Теорема 4 (теорема сложения вероятностей). Если события A и B несовместны, т.е. $AB=\varnothing$, то вероятность их суммы[6] равна сумме вероятностей:

$$\mathsf{P}(A+B)=\mathsf{P}(A)+\mathsf{P}(B). \qquad (11)$$

Доказательство. Проведём n раз эксперимент $\mathfrak{S}$. В некоторых из этих экспериментов будет наблюдаться событие A (их общее число мы обозначаем $n(A)$), а в

[6] Термин «сумма» по нашему соглашению уже предполагает несовместность событий A и B.

некоторых из оставшихся – событие B (их общее число мы обозначаем $n(B)$). Поскольку события A и B не могут наступать одновременно, число $n(A+B)$ экспериментов, которые закончились одним из этих событий, равно $n(A)+n(B)$. После деления равенства $n(A+B)=n(A)+n(B)$ на n мы получим аналогичное равенство для частот: $\nu_n(A+B)=\nu_n(A)+\nu_n(B)$. При росте n частоты будут всё меньше отличаться от соответствующих вероятностей, так что равенство $\nu_n(A+B)=\nu_n(A)+\nu_n(B)$ превратится в равенство (11).

Эти наглядные рассуждения можно заменить формальным «переходом к пределу при $n\to+\infty$» в равенстве $\nu_n(A+B)=\nu_n(A)+\nu_n(B)$:

$$n(A+B)=n(A)+n(B)\Leftrightarrow \nu_n(A+B)=\nu_n(A)+\nu_n(B)\Rightarrow$$
$$\lim_{n\to+\infty}\nu_n(A+B)=\lim_{n\to+\infty}\nu_n(A)+\lim_{n\to+\infty}\nu_n(B)\Leftrightarrow \mathsf{P}(A+B)=\mathsf{P}(A)+\mathsf{P}(B).$$

Замечание. Практически дословно повторяя проведённые рассуждения нетрудно показать, что теорема сложения вероятностей верна для любого конечного числа попарно несовместных событий: если события $A_1,\ldots,A_N$ попарно несовместны, то

$$\mathsf{P}(A_1+\ldots+A_N)=\mathsf{P}(A_1)+\ldots+\mathsf{P}(A_N). \qquad (12)$$

Теорема 5. Для любых двух событий A и B верно равенство

$$\mathsf{P}(A\cup B)=\mathsf{P}(A)+\mathsf{P}(B)-\mathsf{P}(AB) \qquad (13)$$

Доказательство. Проведём n раз эксперимент $\mathfrak{S}$. Эксперименты, в которых наблюдалось событие $A\cup B$ (их общее число мы обозначаем $n(A\cup B)$), т.е. хотя бы одно из событий A или B, можно разбить на три непересекающиеся группы.

В первую включим те эксперименты, в которых наблюдается событие A, но не наблюдается событие B (иначе говоря, наблюдалась разность событий A и B, т.е. эти эксперименты приводят к событию $A\setminus B=A\bar{B}$). Общее число экспериментов из этой группы равно $n(A\setminus B)$.

Во вторую включим те эксперименты, в которых наблюдается событие B, но не наблюдается событие A (иначе говоря, эти эксперименты приводят к событию $B\setminus A$). Общее число экспериментов из этой группы равно $n(B\setminus A)$.

В третью включим те эксперименты, в которых наблюдается и событие A, и событие B (иначе говоря, эти эксперименты приводят к событию AB). Общее число экспериментов из этой группы равно $n(AB)$.

Очевидно, что верно равенство: $n(A\cup B)=n(A\setminus B)+n(B\setminus A)+n(AB)$.

Далее, нетрудно сообразить, что объединение первой и третьей из перечисленных выше групп даст группу всех экспериментов, в которых наблюдалось событие A. Поэтому верно равенство $n(A)=n(A\setminus B)+n(AB)$.

Аналогичные соображения показывают, что верно равенство $n(B)=n\left(B\setminus A\right)+n(AB)$.

Теперь мы имеем:

$$\begin{aligned} n\left(A\cup B\right) &= n\left(A\setminus B\right)+n\left(B\setminus A\right)+n(AB) \\ &= \left[n\left(A\setminus B\right)+n\left(AB\right)\right]+\left[n\left(B\setminus A\right)+n(AB)\right]-n(AB) \\ &= n(A)+n(B)-n(AB). \end{aligned}$$

После деления на n мы получим равенство для частот:

$$\nu_n(A\cup B)=\nu_n(A)+\nu_n(B)-\nu_n(AB).$$

При росте n частоты будут неотличимы от соответствующих вероятностей, так что последнее равенство превратится в (13).

Как и в предыдущей теореме, (неформальный) последний шаг рассуждений можно заменить формальным «переходом к пределу при $n\to+\infty$» в равенстве $\nu_n(A\cup B)=\nu_n(A)+\nu_n(B)-\nu_n(AB)$.

Теорема 6. Если события $A_1,\ldots,A_N$ образуют полную систему попарно несовместных событий, то сумма их вероятностей равна 1:

$$\mathsf{P}\left(A_1\right)+\ldots+\mathsf{P}\left(A_N\right)=1. \tag{14}$$

Доказательство. Проведём n раз эксперимент $\mathfrak{S}$. Из этих экспериментов $n\left(A_1\right)$ экспериментов приведут к событию A_1, $n\left(A_2\right)$ – к событию A_2, …, $n\left(A_N\right)$ – к событию A_N. В каждом эксперименте наблюдалось одно и притом *только* одно из событий нашей системы $A_1,\ldots,A_N$. Поэтому $n\left(A_1\right)+\ldots+n\left(A_N\right)=n.$ После деления на n мы получим следующее равенство для частот: $\nu_n\left(A_1\right)+\ldots+\nu_n\left(A_N\right)=1.$ При больших n частоты будут неотличимы от соответствующих вероятностей, так что последнее равенство превратится в равенство (14).

Число подобных теорем легко умножить, но мы обратим внимание на то, что некоторые из них можно получить из других, *не вдаваясь в то, какой смысл имеет символ* $\mathsf{P}(A)$. Иначе говоря, если справедливы некоторые из полученных формул, то другие также верны, *вне зависимости от того, что мы понимаем под вероятностью события*.

Например,

(1) из равенства $\mathsf{P}\left(\bar{A}\right)=1-\mathsf{P}(A)$ (теорема 3) и неотрицательности вероятности события $\bar{A}$ (первая часть теоремы 1) мы получаем неравенство $\mathsf{P}(A)\le 1$ (вторая часть теоремы 1);

(2) поскольку события A и $\bar{A}$ несовместны и в сумме дают достоверное событие (т.е. $A+\bar{A}=\Omega$), из формул (11) и (8) следует формула (10);

(3) если события $A_1,\dots,A_N$ образуют полную систему несовместных событий, т.е. $A_1+\dots+A_N=\Omega$, то формула (14) немедленно следует из формул (12) и (8).

Более того, мы можем получать и новые формулы, связывающие между собой вероятности разных событий, *не вдаваясь в смысл понятия вероятность события, а рассматривая вероятность события просто как некоторое число, приписываемое этому событию, т.е. как некоторую меру события*. В качестве примера приведём следующие утверждения.

Теорема 7. Вероятность объединения двух событий не превосходит суммы их вероятностей:

$$\mathsf{P}\left(A\cup B\right)\le\mathsf{P}(A)+\mathsf{P}(B). \tag{15}$$

Доказательство. Рассмотрим общую теорему сложения вероятностей, выражаемую формулой (13). В силу теоремы 1 вероятность любого события, в том числе и вероятность события AB, неотрицательна. Поэтому правая часть формулы (13) не превосходит $\mathsf{P}(A)+\mathsf{P}(B)$, что и доказывает неравенство (15).

Теорема 8. Если события $A_1,\dots,A_N$ образуют полную систему несовместных событий, то для вероятности любого события A верно равенство:

$$\mathsf{P}(A)=\mathsf{P}(AA_1)+\dots+\mathsf{P}(AA_N). \tag{16}$$

Доказательство. События $AA_1,\dots,AA_N$ попарно несовместны (т.к. попарно несовместны события $A_1,\dots,A_N$) и в сумме дают событие A (т.к. события $A_1,\dots,A_N$ в сумме дают достоверное событие). Применяя формулу (12), мы получим утверждение теоремы.

Теорема 9. Если события $A_1,\dots,A_N$ образуют полную систему попарно несовместных событий и их вероятности равны, то вероятность каждого из этих событий равна $\dfrac{1}{N}$.

Доказательство. Пусть p – общее значение вероятностей $\mathsf{P}\left(A_1\right),\dots,\mathsf{P}\left(A_N\right)$. Тогда равенство (14) примет вид $Np=1$, откуда и следует утверждение теоремы.

5. «Классическое определение вероятности»

Предположим, что для некоторого эксперимента мы выделили конечную полную систему событий $A_1,\ldots,A_N$ и рассматриваем эти события как набор возможных исходов (элементарных событий) эксперимента. Предположим далее, что все исходы $A_1,\ldots,A_N$ равновозможны, т.е. их вероятности равны: $P(A_1)=\ldots=P(A_N)$.

Допустим, далее, что событие A является суммой нескольких элементарных событий, которые мы обозначим $A_{n_1},\ldots,A_{n_k}$: $A=A_{n_1}+\ldots+A_{n_k}$. Число k исходов, которые в сумме дают событие A, мы обозначим через $N(A)$, чтобы подчеркнуть, что k определяется самим событием A. Элементарные события $A_{n_1},\ldots,A_{n_k}$ и только они *благоприятствуют* наступлению события A.

Теорема 10 («классическое определение вероятности»). В описанной выше ситуации вероятность события A равна отношению числа элементарных событий, благоприятствующих его наступлению, к общему числу элементарных событий:

$$P(A)=\frac{N(A)}{N}. \tag{17}$$

Доказательство. Из Теоремы 9 мы знаем, что общее значение вероятностей элеменарных событий равно $\frac{1}{N}$. В силу теоремы сложения вероятностей

$$P(A)=P\left(A_{n_1}\right)+P\left(A_{n_2}\right)+\ldots+P\left(A_{n_k}\right).$$

В правой части этого равенства стоит $k\equiv N(A)$ слагаемых, каждое из которых равно $\frac{1}{N}$.

Значит, $P(A)=N(A)\cdot\frac{1}{N}=\frac{N(A)}{N}$. □

Хотя в щкольных учебниках и упоминается частотный подход к вероятности, фактически именно соотношение (17) берётся в качестве определения вероятности случайного события. На его основе вводятся более сложные понятия (например, независимости событий) и доказываются все теоремы. В связи с этим необходимо сделать несколько важных комментариев.

(1) «Классическое определение вероятности» определяет вероятность события только в рамках классической модели, т.е. элементарной модели с конечным числом «равновозможных» исходов. Иначе говоря, *«согласно этому определению, просто не существует вероятности там, где нет равновозможных случаев»* (цитата из книги Мизеса

[12], стр.84). Например, нельзя говорить о вероятности того, что при бросании неправильной монеты выпадет орёл. □

(2) В учебниках по теории вероятностей, базирующихся на «классическом определении вероятности», именно на основе этого «определения» вводятся более сложные понятия (прежде всего независимость) и доказываются основные теоремы (вроде теоремы сложения). Но затем эти понятия и теоремы применяются и в случаях, когда эксперимент невозможно описать моделью с конечным числом «равновозможных» исходов, что совершенно незаконно (кроме того, надо иметь в виду и затронутый выше вопрос о смысле вероятности в этой ситуации).

Возьмём, например, учебник «Алгебра и начала математического анализа. 10 кл.» из серии «МГУ – школе» (авторы С.М.Никольский, М.К.Потапов, Н.Н.Решетников, А.В.Шевкин, 8-е изд. – М.: Просвещение, 2009).

В разделе 12.1, стр. 336, даётся определение вероятности события: «Вероятность события A определим как отношение числа m случаев, благоприятствующих событию A, к общему числу n рассматриваемых случаев, т.е.

$$P(A) = \frac{m}{n}.»$$

В разделе 12.2 с помощью этого определения доказываются основные свойства вероятности.

В разделе 13.2 даётся определение условной вероятности: «Условной вероятностью события B при условии, что произошло событие A, называют отношение числа случаев, благоприятствующих событию AB, к числу случаев, благоприятствующих событию A.» В наших обозначениях это определение даётся формулой $P(B \mid A) = \dfrac{N(AB)}{N(A)}$. Позже мы расскажем, как следует определять условные вероятности, и увидим, что на самом деле эта формула является теоремой (имеющей смысл и справедливой только для экспериментов с конечным числом равновозможных исходов).

Далее, в том же разделе 13.2 на стр.346 определяется понятие независимости: «События A и B (в рассматриваемом опыте) называют независимыми, если справедливо равенство $P(AB) = P(A)P(B)$». Имея в виду содержание предыдущих разделов, следует подчеркнуть, что при таком изложении теории вероятности о независимых событиях можно говорить только в рамках классической модели. Само определение независимости фактически сводится к равенству $N(AB) \cdot N = N(A) \cdot N(B)$.

Но в конце раздела 13.2, на стр.347, авторы с помощью теории, развитой **только** для экспериментов с конечным числом равновозможных исходов решают следующую задачу:

«Пример 5. Два стрелка стреляют по мишени. Вероятность попадания в мишень первым стрелком 0,6, а вторым – 0,5. Считая, что попадание в мишень каждого из стрелков является независимым событием (т.е. вероятность попадания в мишень каждым стрелком не зависит от попадания или непопадания в мишень другим стрелком), определим вероятность попадания в мишень обоими стрелками; хотя бы одним стрелком.»

Что здесь понимать под вероятностями 0,6 и 0,5, независимостью (тем более, что в учебнике определена независимость *событий*, а не *вероятностей* событий), на каком основании применяются теоремы сложения и умножения вероятностей авторы учебника не объясняют. □

(3) «Классическое определение вероятности» сводит понятие вероятности к понятию «равновероятности». По смыслу слова «равновероятность» события A и B равновероятны, если их вероятности равны – но как это установить, если само понятие вероятности ещё не определено. Понимая это обстоятельство, авторы школьных (и не только) учебников заменяют слово «равновероятные» на «равновозможные». Ясно, что это совершенно не проясняет ситуацию. Тем не менее, некоторые учебники считают эту замену слов достаточным объяснением. Другие авторы идут немного дальше и «определяют» термин «равновозможность» более пространными фразами. Например, в упомянутом выше учебнике для 10 класса «события называют равновозможными», если «нет никаких оснований предполагать, что одно из событий может произойти предпочтительнее, чем другое» – ясно, что это просто попытка отделаться от неудобного вопроса, а вовсе не объяснение смысла термина. Иногда ссылаются на авторитет великих предков: «Такое определение вероятности было впервые дано в работах французского математика Лапласа и называется классическим.»[7] Понимая смутность подобных «объяснений», которые камуфлируют отсутствие всякой мотивации, часто, в обычных традициях схоластической математики, это классическое «определение» вероятности вводится вообще без всяких объяснений: «Определение. Вероятностью события называется …». □

(4) Классическое «определение» вероятности в школьных учебниках часто мотивируется разговорами о *шансах* наступления события и о вероятности события как *мере достоверности* его наступления (см., например, §23 Алгебры для 9 кл. Ш.А.Алимова, Ю.М.Колягина и др. – 16-е изд. – М.: Просвещение, 2011). Это, как минимум, приводит к возможности говорить о вероятностях уникальных событий, т.е. совершенно ложному взгляду на теорию вероятностей. В худшем случае возникает впечатление, что вероятность – это числовая мера «степени правдоподобия», т.е. числовая мера психологического состояния человека. □

(5) На самом деле никакой проблемы с мотивацией «классического определения вероятности» нет, т.к. при правильном взгляде на теорию вероятностей «классическое определение вероятности» является *теоремой*, вытекающей из простой теоремы сложения вероятностей – именно по этой причине мы использовали кавычки, когда писали об этом «определении» (если брать в кавычки только слово «определение», наша позиция будет выражена яснее). На это обращает внимание Б.В.Гнеденко, который в своём учебнике [6] (§2 главы 1) пишет по поводу классического «определения» вероятности: «*…оно в*

[7] На самом деле классическое определение вероятности ввёл в науку Я.Бернулли за 100 лет до Лапласа. Лаплас лишь подчеркнул важность явного указания на равновозможность исходов – см. «Очерк истории вероятностей» Б.В.Гнеденко в Дополнении к [5] или «Математическая теория вероятностей. Очерк истории становления» А.Н.Ширяева в Приложении в 3-му изданию книги [1].

действительности является не определением, а скорее методом вычисления вероятностей во вполне определённых и сильно ограниченных условиях». □

(6) Конечно, есть случаи, когда классическое «определение» вероятности можно использовать как определение, но они находятся далеко вне рамок школьного курса.

Например, это определение можно использовать для доказательства непротиворечивости системы аксиом Колмогорова через указание объекта, который им удовлетворяет. Сам Колмогоров использовал одноэлементное пространство Ω: $\Omega = \{\omega\}$, алгебру событий из двух элементов: $\varnothing$ и Ω, меру $P(\varnothing) = 0$, $P(\Omega) = 1$, что, конечно, является классическим определением вероятности (с равным успехом можно было взять любое конкретное распределение на конкретном конечном множестве). □

(7) Важным и печальным результатом построения школьного курса на базе классического «определения» вероятности является отмеченное ранее большое число задач в школьных учебниках на расчёт вероятностей событий, связанных с азартными играми. Конечно, «игорная мафия» тут ни при чём. Это просто врождённый порок классического «определения» вероятности.

В.Феллер, [7], гл.I, §7, отмечает: *«Полезность пространств с равновероятными элементарными событиями проявляется в основном при изучении игр и в комбинаторном анализе».*

Г.Крамер пишет в [8], пункт 13.5: *«Слабость этого определения очевидна. Прежде всего, неизвестно, каким способом решать, можно ли считать два случая равновозможными. Более того, трудно, а по мнению некоторых даже невозможно, точно указать разбиение на равновозможные случаи результатов опытов, не принадлежащих к азартным играм».* □

(8) Единственное достоинство (если это можно считать достоинством) классического «определения» вероятностей заключается в том, что расчёт вероятности события сводится к подсчёту общего числа исходов и числа исходов, благоприятствующих данному событию – комбинаторным задачам, для решения которых в «стандартной» математике разработаны разнообразные методы. В результате теория вероятностей рассматривается как область «практического применения» комбинаторики, а не как специфический раздел науки.

Комбинаторика в школьном курсе сама по себе заслуживает отдельного разговора. Вся её мудрость содержится в сказке Виталия Губарева «Королевство кривых зеркал»: «На ста площадях по сто зеркал! … сколько будет всего зеркал?... – Десять тысяч, – раздался тонкий голосок.» Непосредственное применение этого *правила произведения* даёт все основные формулы школьной комбинаторики, а единственное реальное приложение этих формул – расчёт вероятностей событий, связанных с азартными играми (не считать же приложениями «занимательные» задачи вроде раскрашивания флага или рассаживания гостей за столом). Зачем же вводить комбинаторику и растолковывать её на примере азартных игр и «занимательных» задач? Только потому, что решение этих задач 400 лет тому назад помогло становлению теории вероятностей? Я думаю, что объяснить маниакальную любовь к классическому «определению» вероятности в школьных учебниках можно только любовью их авторов к комбинаторике, которая, в свою очередь, видимо, связана с воспоминаниями о счастливом олимпиадном детстве в физико-математических школах. □

(9) Но есть одно несомненное достоинство классического «определения» вероятностей, о котором, правда, предпочитают не говорить. Для простых школьных задач (вроде упомянутой выше задачи ЕГЭ про школьника, который не выучил все билеты по биологии) подсчёт общего числа исходов и числа исходов, благоприятствующих данному событию, производится «на пальцах», так что теория вероятностей сводится к выполнению нескольких арифметических действий. Поэтому совсем не удивительно, что подобную задачу по «теории вероятностей» на ЕГЭ верно решает около 80% школьников (в то время как меньше половины школьников смогли решить следующую лёгкую задачу на проценты: «Десять одинаковых рубашек дешевле куртки на 6%. На сколько процентов пятнадцать таких же рубашек дороже куртки?»). Понятно и зачем такие задачи предлагают на ЕГЭ – это позволяет отрапортовать об успешном усвоении школьниками «стохастической линии». Задача по «теории вероятностей» и аналогичные в идейном плане (в смысле бессодержательности) задачи на решение тривиальных показательных, логарифмических, тригонометрических уравнений, вычисление площадей по клеткам и т.п. преследуют единственную цель: создать иллюзию того, что школьник получает аттестат зрелости за знание «сложных» разделов математики (а не за зубривание нескольких формул и последующие арифметические вычисления с целыми числами). □

Очень хорошо вопрос о «классическом определении вероятности» разобран Мизесом [12], стр. 80 – 101. Мы рекомендуем читателю подробно изучить его аргументы.

6. Геометрическая вероятность

Теорема 11. Предположим, что эксперимент заключается в случайном бросании точки на некоторую ограниченную область D; её площадь мы обозначим $S(D)$. При этом под «случайностью» будем понимать то, что для любых двух квадратов однакового размера, целиком лежащих внутри области D, вероятность попадания в эти квадраты одинакова («ни одна из частей области не притягивает брошенную точку больше, чем другая»). Тогда для любой части A этой области вероятность $P(A)$ того, что брошенная точка попала именно в эту часть, равна отношению площади этой части к площади всей области D:

$$P(A) = \frac{S(A)}{S(D)}. \tag{18}$$

Доказательство. Для доказательства слегка модифицируем стандартный приём, используемый в школьном курсе геометрии при определении площади фигуры (см., например, А.П.Киселёв. Геометрия. Планиметрия. Учебник для 7-9 классов общеобразовательных школ. СПб.: Специальная Литература, 1999. Раздел пятый «Измерение площадей», пункт 239 «Понятие об измерении площади»).

Опишем вокруг области D квадрат K, разделим каждую его сторону на большое число n равных отрезков и проведём через концы этих отрезков прямые, параллельные сторонам квадрата. В результате квадрат разобьётся на n^2 маленьких квадратиков; площадь одного такого квадратика равна $\dfrac{S(K)}{n^2}$. Это построение проиллюстрировано на рис.2, где большой овал, нарисованный штрихованной линией, представляет область D, а небольшая вытянутая фигура в центре – область A.

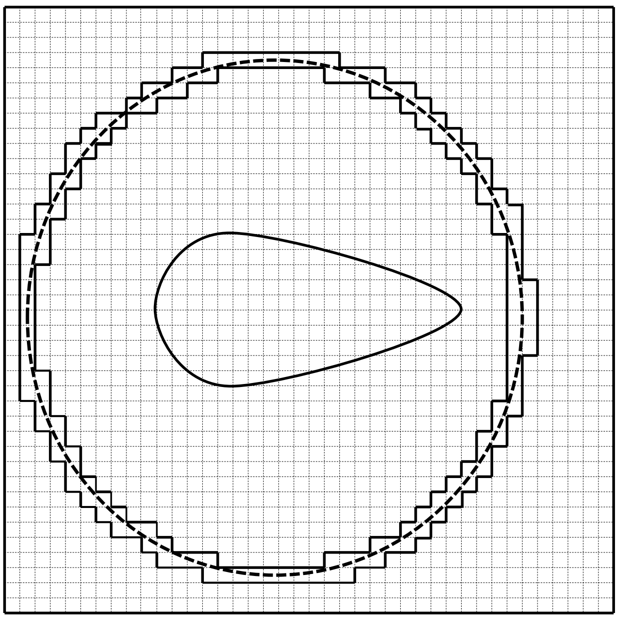

Рис. 2

Часть этих квадратиков (их число обозначим через *N*) расположена внутри области *D*, часть лежит вне области *D*, а оставшиеся квадратики частично лежат внутри области *D*, а частично выходят за пределы этой области (через эти квадратики проходит линия, ограничивающая область *D*; их число обозначим через *l*). Квадратики из первой группы образуют ступенчатую фигуру D_n, лежащую внутри области *D*. Её площадь равна $S(D_n) = \frac{N}{n^2} \cdot S(K)$. Квадратики из третьей группы образуют ступенчатую фигуру B_n, которая покрывает границу области *D*. Её площадь равна $S(B_n) = \frac{l}{n^2} \cdot S(K)$.

Ступенчатая фигура B_n при очень большом числе *n* будет практически неотличима от границы области *D*. Поскольку площадь этой границы равна 0, при очень большом числе *n* практически верно равенство: $S(B_n) = 0$. Ступенчатая фигура D_n при $n \to \infty$ будет практически неотличима от области *D*. Поэтому при $n \to \infty$ можно заменить область *D* ступенчатой фигурой D_n. Соответственно, можно считать, что практически верно равенство $S(D) = \frac{N}{n^2} \cdot S(K)$, т.е можно считать, что область *D* состоит из $N = \frac{S(D)}{S(K)} \cdot n^2$ равных маленьких квадратиков. Перенумеруем эти квадратики последовательными натуральными числами от 1 до *N* и обозначим через A_i событие, заключающееся в том, что точка, брошенная на область *D*, попадёт в *i*-й маленький квадратик, $i = 1, 2, \ldots, N$. Эти события образуют полную систему событий и по условию равновероятны. В силу Теоремы 9 вероятность каждого из них равна $\frac{1}{N} = \frac{S(K)}{S(D) \cdot n^2}$.

Теперь займёмся областью A. Рассуждения, аналогичные проведённым выше, позволяют считать, что область A состоит из $k = \frac{S(A)}{S(K)} \cdot n^2$ равных маленьких квадратиков; пусть их номера будут $i_1, \ldots, i_k$. Попадание точки, брошенной на область *D*, в область *A* означает попадание в один из этих квадратиков. Иначе говоря, $A = A_{i_1} \cup \ldots \cup A_{i_k}$. Поскольку события $A_1, \ldots, A_N$ попарно несовместны, таковы же и события $A_{i_1}, \ldots, A_{i_k}$. Применяя теорему сложения вероятностей, мы получим:

$$P(A) = P\left(A_{i_1}\right) + \ldots + P\left(A_{i_k}\right) = k \cdot \frac{1}{N} = \frac{S(A)}{S(K)} \cdot n^2 \cdot \frac{S(K)}{S(D) \cdot n^2} = \frac{S(A)}{S(D)},$$

что и требовалось доказать.

Замечание 1. Приём, применённый для доказательства теоремы 11, позволяет доказать и соответствующую «одномерную» теорему (рассуждения даже проще):

Теорема 12. Предположим, что эксперимент заключается в случайном бросании точки на некоторый отрезок *D* длины *l*(*D*). При этом под «случайностью» будем понимать то, что для любых двух отрезков одинаковой длины, целиком лежащих внутри отрезка *D*, вероятность попадания точки на эти отрезки одинакова («ни одна из частей отрезка D не притягивает брошенную точку больше, чем другая»). Тогда для любого отрезка A, лежащего внутри основного отрезка D, вероятность $P(A)$ того, что брошенная точка попала именно на эту часть основного отрезка, равна отношению длины этой части к длине основного отрезка:

$$P(A) = \frac{l(A)}{l(D)}. \tag{19}$$

7. Об аксиматическом построении теории вероятностей

В нашей статье мы начали изложение основных теорем теории вероятностей с того, что доказали небольшое число формул с помощью представления о вероятности события как о предельном значении частоты его наступления в длинной серии экспериментов, проводимых в неизменных условиях. Затем мы привели примеры теорем, доказательство которых базируется на этих ранее доказанных простых свойствах вероятностей событий. При этом нам было совершенно неважно, какой смысл имеет символ $P(A)$. Иначе говоря, мы смогли получить новые формулы, связывающие между собой вероятности разных событий, *не вдаваясь в смысл понятия «вероятность события», а рассматривая вероятность события просто как некоторое число, приписываемое этому событию, т.е. как некоторую меру события*. Позже мы увидим и другие примеры результатов такого рода.

Поэтому можно взять некоторые свойства случайных событий и их вероятностей в качестве аксиом и после этого развивать теорию вероятностей, не принимая в расчёт физический смысл понятий «случайное событие» и «вероятность события». Эта задача облегчилась после создания в конце 19 – начале 20 веков теории множеств и теории меры, когда специалисты по теории вероятностей обратили внимание на то, что операции над событиями аналогичны операциям над множествами, свойства вероятностей событий аналогичны свойствам меры Лебега, математическое ожидание случайной величины аналогично интегралу, некоторые свойства независимых случайных величин похожи на свойства ортогональных функций и т.д. Стало ясно, что[8] :

(1) эксперимент можно отождествить с множеством Ω его исходов;

(2) события, наблюдаемые в результате эксперимента, можно отождествить с подмножествами множества Ω, а операции над событиями – с обычными теоретико-множественными операциями. При этом важно, что набор $\mathbb{F}$ «событий» является σ-алгеброй, т.е. должен подчиняться следующим условиям:

1. $\Omega \in \mathbb{F}$,
2. $A \in \mathbb{F} \Rightarrow \bar{A} \in \mathbb{F}$,
3. $A_1, A_2, \ldots \in \mathbb{F} \Rightarrow \bigcup_n A_n \in \mathbb{F}$.

«Событиями» называются не все подмножества Ω, а лишь те из них, которые входят в набор $\mathbb{F}$.

[8] Ниже приводится стандартная система аксиом теории вероятностей. Следует отметить, что для исключения некоторых патологических ситуаций на *вероятностное пространство* .$(\Omega, \mathbb{F}, P)$. накладывают дополнительные требования, например, условие полноты: $A \in \mathbb{F}, P(A) = 0, B \subset A \Rightarrow B \in \mathbb{F}$.

(3) вероятность «события», т.е. множества из класса $\mathbb{F}$, можно рассматривать как меру этого множества; при этом важно, что эта мера обладает свойствами:

1. $P(A) \geq 0$,
2. $P(\Omega) = 1$;
3. если «события» из набора $A_1, A_2, \ldots$ попарно непересекаются, то $P\left(A_1 \cup A_2 \cup \ldots\right) = P\left(A_1\right) + P\left(A_2\right) + \ldots$.

Если пространство элементарных событий Ω состоит из конечного числа *N* элементов $\omega_1, \ldots, \omega_N$, то в качестве событий обычно рассматриваются все подмножества Ω (т.е. класс $\mathbb{F}$ событий – это множество всех подмножеств Ω – в этом случае аксиомы 1-3 из пункта 2 общей аксиоматики, очевидно, выполнены), а вероятностная мера *P* задаётся с помощью *N* неотрицательных чисел $p_1, \ldots, p_N$, в сумме дающих 1, соотношением:

$$P(A) = \sum_{i:\omega_i \in A} p_i. \tag{20}$$

Нетрудно видеть, что мера, определённая соотношением (20), удовлетворяет аксиомам 1-3 из пункта 3 общей аксиоматики. Если событие *A* состоит из одной точки ω_k, т.е. $A = \{\omega_k\}$, то $P(A) = p_k$. Иначе говоря, числа $p_1, \ldots, p_N$ можно *интерпретировать* как вероятности элементарных событий $\omega_1, \ldots, \omega_N$ (следует отметить, что при формальном подходе элементарное событие не является событием). Это позволяет выразить определение (20) следующей фразой: «вероятность события – это сумма вероятностей исходов, благоприятствующих этому событию».

Таким образом, элементарная теория вероятностей (когда пространство элементарных событий конечно) базируется на следующих аксиомах для «вероятностей» $p_1, \ldots, p_N$ элементарных событий $\omega_1, \ldots, \omega_N$:

$$p_1, \ldots, p_N \geq 0, \tag{21}$$

$$p_1 + \ldots + p_N = 1. \tag{22}$$

В элементарной теории свойство аддитивности вероятности, которое в общей теории принимается в качестве аксиомы, доказывается (для конечного числа событий) и называется *теоремой сложения вероятностей*: если $A_1 \cap A_2 = \varnothing$, то $P\left(A_1 \cup A_2\right) = P\left(A_1\right) + P\left(A_2\right)$.

Предположим теперь, что в рамках элементарной теории вероятностей (пространство элементарных событий конечно) все элементарные события равновероятны, т.е. $p_1 = \ldots = p_N$ (обычно это мотивируют «симметрией» элементарных событий $\omega_1, \ldots, \omega_N$). Тогда аксиома (22) влечёт, что $p_i \equiv \frac{1}{N}$. Это, в свою очередь, приводит к тому, что определение (20) превратится в «классическое определение вероятности»:

$$P(A) = \frac{N(A)}{N}, \tag{23}$$

где *N*(*A*) – число элементарных событий, из которых состоит событие *A*.

Описанная аксиоматика теории вероятностей позволяет построить всю дальнейшую теорию вероятностей как абстрактную математическую теорию. Более того, содержательную теорию можно построить и на базе упрощённой аксиоматики в случае конечного пространства элементарных событий. Поскольку физический смысл множеств Ω, $\mathbb{F}$, меры P не играет при этом никакой роли, абстрактную математическую теорию вероятностей можно применять и для решения задач, совершенно не связанных с теорией вероятностей в обычном смысле этого слова. Именно с этой особенностью абстрактной математической теории вероятностей и связано появление абсурдных задач, вроде рассмотренной выше задачи про вероятность успешной сдачи экзамена. Действительно, коль скоро вероятность – это просто определённая мера на пространстве элементарных событий, ничто не мешает рассмотреть в качестве модели экзамена классическую модель (23): пространство элементарных событий Ω – это множество $\{1,2,\ldots,25\}$ номеров билетов, алгебра событий $\mathbb{F}$ – все подмножества Ω, мера события $A \subset \Omega$ – отношение числа элементов A к числу элементов Ω. Если билеты, которые школьник не знает, имеют номера 24 и 25, то интересующее нас событие «экзамен сдан» мы отождествляем с множеством $A = \{1,2,\ldots,23\}$, мера которого, т.е. вероятность, равна $\frac{23}{25}$. Из сказанного ранее о смысле теории вероятностей как науки о специфических явлениях окружающего нас реального мира совершенно ясно, что это формальное применение абстрактной математической теории вероятностей к реальной ситуации не имеет никакого содержательного смысла. Оно совершенно незаконно и является чистейшей схоластикой (как и большинство школьных задач на применение классического «определения» вероятности).

Когда А.Н.Колмогоров создавал аксиоматическую теорию вероятностей на базе теории меры множеств, он исходил прежде всего из простоты системы аксиом, их связи с другими разделами математики, удобства логически строгого получения более сложных результатов. Но если не стоять на формальной, схоластической, позиции, то перед введением системы аксиом её нужно мотивировать. В случае теории вероятностей прежде всего нужно объяснить, *почему* мы требуем выполнения свойств 1-3 из пункта 3. Сделать это можно *только* с помощью частотного подхода – именно так мотивировал сам А.Н.Колмогоров аксиомы вероятности в своей основополагающей работе [1] (§2 главы 1). Кроме того, повторим ещё раз слова Колмогорова, основой для применения теории вероятностей к реальному миру является именно частотное определение вероятности.

8. Условная вероятность

8.1 Условная вероятность. Теорема умножения

Вероятность $P(A)$ события A характеризует частоту его наступления в большой серии однотипных экспериментов, проводимых в неизменных условиях: если число n проведённых экспериментов очень велико, то доля $\nu_n(A)=\dfrac{n(A)}{n}$ тех экспериментов, в которых наблюдалось событие A, относительно мало отличается от вероятности $P(A)$.

В многих важных ситуациях интерес представляет частота наступления интересующего нас события A не во всех экспериментах, а лишь в тех из них, в которых наблюдалось некоторое другое событие B. Например, выпуск какой-нибудь массовой однородной продукции можно рассматривать как серию однотипных «экспериментов» («экспериментом» является окончание производства очередного изделия); при этом основной интерес представляет событие A=«изделие не удовлетворяет установленным стандартам». Вероятность этого события позволяет оценить число изделий ненадлежащего качества в большой партии произведённых изделий и характеризует процесс производства. Но в торговую сеть поставляются только изделия, которые прошли определённый контроль качества. В процессе этого контроля бóльшая часть изделий ненадлежащего качества отбраковывается (при этом ошибочно и некоторые хорошие изделия могут быть признаны бракованными). Ясно, что торговлю интересует лишь доля нестандартных изделий среди изделий, поступивших в продажу (не меньший интерес эта доля представляет и для руководства предприятия), т.е. частота наступления события A=«изделие не удовлетворяет установленным стандартам» после тех «экспериментов», в которых наблюдалось событие B=«изделие прошло итоговый контроль качества».

Сказанное выше фактически означает, что в комплекс условий $\mathfrak{S}$, относительно которого рассматривается событие A, добавляется ещё одно условие: «осуществилось событие B». Обозначим этот новый комплекс условий (эксперимент) через $\mathfrak{S}+B$ и назовём его условным экспериментом при условии, что событие B произошло.

Определение. Вероятность события A при усуществлении комплекса условий $\mathfrak{S}+B$ называется *условной вероятностью события* A *при условии, что событие* B *произошло*, и обозначается $P(A\,|\,B)$.

Часто используют более короткий термин и говорят, что $P(A\,|\,B)$ – это *условная вероятность события* A *при условии* B.

Как мы отмечали ранее, говорить о вероятности любого события A можно только, если явно указан комплекс условий (эксперимент) $\mathfrak{S}$, после реализации которого устанавливается наступление или ненаступление этого события. В этом смысле все вероятности являются «условными» и, возможно, при более формальном изложении теории вероятностей следовало бы писать не $P(A)$, а $P(A \mid \mathfrak{S})$. Тогда условная вероятность $P(A \mid B)$ – это просто $P(A \mid \mathfrak{S}+B)$. Но поскольку комплекс условий $\mathfrak{S}$ обычно фиксируется в начале решения любой задачи по теории вероятностей, явное указание на комплекс условий $\mathfrak{S}$ и прилагательное «условная» опускают и говорят просто о вероятности $P(A)$. В данном выше определении термина «условная вероятность» прилагательное «условная» подчёркивает тот факт, что к фиксированному комплексу условий $\mathfrak{S}$ добавляется ещё некоторое условие (наступление события B). Если же мы хотим подчеркнуть, что никаких дополнительных условий к фиксированному комплексу условий $\mathfrak{S}$ не добавляется, то вместо термина «вероятность» используют термин «безусловная вероятность». С другой стороны, условная вероятность является вероятностью в смысле определения из первой части нашей статьи (просто усложняется комплекс условий). Поэтому для условных вероятностей (при фиксированном условии B) верны все теоремы, установленные во второй части нашей статьи.

Данное выше определение условной вероятности позволяет выразить её через безусловные вероятности. Именно, справедлива следующая

Теорема 13. Если вероятность события B не равна 0, то условная вероятность события A при условии B равна частному от деления вероятности совместного наступления событий A и B на вероятность условия:

$$P(A \mid B) = \frac{P(AB)}{P(B)}. \tag{24}$$

Доказательство. Чтобы найти вероятность $P(A \mid B)$, нужно:

1. провести k раз условный эксперимент $\mathfrak{S}+B$,
2. подсчитать число $n(A \mid B)$ экспериментов, в которых наблюдалось событие A, и найти условную частоту $\nu_k(A \mid B) = \dfrac{n(A \mid B)}{k}$ появления этого события,
3. найти предельное значение этой частоты при $k \to \infty$ – это и будет условной вероятностью $P(A|B)$:

$$P(A \mid B) = \lim_{k\to\infty} \nu_k(A \mid B). \tag{25}$$

Поскольку событие B наступает случайно, провести много раз условный эксперимент $\mathfrak{S}+B$ мы можем только проведя много раз безусловный эксперимент $\mathfrak{S}$ и отбирая те эксперименты, в которых наблюдалось событие B. Итак, допустим, что мы провели безусловный эксперимент $\mathfrak{S}$, скажем, n раз. В $k = n(B)$ из них осуществится событие B. Ясно, что появление события A в этих условных экспериментах означает одновременное наступление событий A и B в исходной серии из n безусловных экспериментов, т.е. $n(A \mid B) = n(AB)$. Поэтому

$$\nu_k(A \mid B) \equiv \frac{n(A \mid B)}{k} = \frac{n(AB)}{n(B)} = \frac{n(AB)/n}{n(B)/n} = \frac{\nu_n(AB)}{\nu_n(B)}. \tag{26}$$

Если число n будет неограниченно увеличиваться, то неограниченно будет расти и число $k = n\left(B\right) \approx nP(B)$. В силу определения (25) условной вероятности левая часть равенства (26) практически не будет отличаться от условной верятности $P(A \mid B)$. Правая же часть, в силу определения вероятности из части 1 нашей статьи, практически не будет отличаться от отношения безусловных вероятностей $P(AB)$ и $P(B)$. Этот предельный переход в равенстве (26) и доказывает справедливость соотношения (24).

Замечание 1. Если вероятность события B равна 0, то условная вероятность $P(A \mid B)$ не определяется.

Замечание 2. Соотношение (24) можно записать и в виде

$$P(AB) = P(B)P(A \mid B). \tag{27}$$

В такой форме оно называется *теоремой умножения вероятностей*.

При решении реальных задач очень часто из каких-нибудь соображений условная вероятность $P(A|B)$ считается известной, а Теорема 13 применяется в виде теоремы умножения вероятностей для вычисления $P(AB)$.

Замечание 2. В формальной математической теории вероятностей соотношение (24) принимается в качестве определения условной вероятности, а приведённое нами доказательство рассматривается как мотивация этого определения.

Рассмотрим теперь классическую вероятностную схему эксперимента с конечным числом N равновероятных элементарных исходов. В этой ситуации $P(B) = \dfrac{N(B)}{N}$, $P(AB) = \dfrac{N(AB)}{N}$, где $N\left(B\right)$ – число элементарных событий, благоприятствующих наступлению условия B, $N\left(AB\right)$ – число элементарных событий, благоприятствующих совместному наступлению интересующего нас события A и условия B. Поэтому равенство (24) примет вид:

$$P(A \mid B) = \frac{N(AB)}{N(B)}. \tag{28}$$

В учебниках, ориентирующихся на классическое определение вероятности, это соотношение принимается в качестве определения условной вероятности. Мотивируется это «определение» рассуждениями о «шансах» на наступление события, что, как мы уже отмечали, порочно в принципе.

Чтобы проиллюстрировать понятие условной вероятности, рассмотрим два простых примера. Первый взят из демографии, которая наряду со страхованием жизни является одной из самых важных областей применения теории вероятностей. Дело в том, что наблюдение за продолжительностью жизни или смертностью, скажем, в возрасте 40 лет отдельного человека является своеобразным экспериментом. Людей много, так что этот эксперимент массовый. Хотя разные люди с точки зрения эксперимента под названием «жизнь» неотличимы друг от друга, чтобы обеспечить однородность при повторении эксперимента нужно брать не бросто большую группу людей, а отдельно рассматривать мужчин и женщин, если брать мужчин, то примерно в одном возрасте, с одинаково хорошим (или одинаково плохим) здоровьем и т.д. Наблюдения за смертностью на протяжении нескольких веков показывают устойчивость относительных частот различных событий, харатеризующих продолжительность жизни человека, на достаточно длительных промежутках времени (на очень длинных промежутках характеристики продолжительности жизни меняются, например, средняя продолжительность жизни постоянно растёт). Впервые это было установлено при изучении смерности населения Лондона, Парижа, Бреславля, которое провели в 17 веке английские учёные Джон Гронт (John Graunt, 1620 – 1674), Уильям Пети (William Petty, 1623 – 1687), Эдмонд Хали (Edmond Halley, 1656 – 1742); их исследования, наряду с изучением в 16-17 вв. азартных игр Кардано, Тарталья, Паскалем, Ферма и другими учёными, привели к возникновению современной теории вероятностей и статистики.

Пример 1. В таблице 1 приведены вероятности $s(x)$ того, что мужчина, родившийся в 2006 году, проживёт по меньшей мере x лет, т.е. доживёт до возраста x лет (эти данные взяты из Демографического ежегодника России за 2012 год). Функция $s(x)$ называется *функцией выживания* (для мужчин). Если обозначить через T продолжительность жизни новорожденного младенца мужского пола, то функцию выживания можно рассматривать как вероятность события $\{T > x\}$.

Таблица 1

x	*s*(*x*)	*x*	*s*(*x*)	*x*	*s*(*x*)
1	0.98859	20	0.97357	55	0.65465
2	0.98743	25	0.95691	60	0.55972
3	0.98671	30	0.92776	65	0.45497
4	0.98607	35	0.89219	70	0.34521
5	0.98548	40	0.85232	75	0.23579
10	0.98338	45	0.80120	80	0.13769
15	0.98108	50	0.73680	85	0.06875

Из этой таблицы видно, что вероятность дожития до 75 лет, т.е. вероятность события $A = \{T > 75\}$, равна примерно 24%, а вероятность дожития до 65 лет, т.е. вероятность события $B = \{T > 65\}$, – примерно 45%. Это означает, что из большой группы новорожденных младенцев мужского пола, скажем из 100 тысяч новорожденных, свой 65-й день рождения отметит чуть больше 45 тысяч из них, а 75-летний юбилей отпразнуют лишь около 24 тысяч. Если пенсионный возраст для мужчин установлен на уровне 65 лет, то это означает, что больше половины мужчин не доживёт до пенсии. Предположим теперь, что мужчина дожил до пенсионного возраста 65 лет, т.е. наступило событие $B = \{T > 65\}$.

Найдём вероятность $P(A \mid B) \equiv P(T > 75 \mid T > 65)$ дожития до 75-летнего возраста при этом условии. В соответствии с формулой (24) мы должны найти вероятность $P(AB) \equiv P(T > 75 \text{ и } T > 65)$ совместного наступления событий A и B, а затем разделить её на вероятность $P(B) \equiv P(T > 65)$ условия. Фраза «человек дожил до 65 лет и до 75 лет» выражает то же самое, что и фраза «человек дожил до 75 лет». Иначе говоря, событие $\{T > 75 \text{ и } T > 65\}$ равносильно событию $\{T > 75\}$. Поэтому искомая вероятность равна

$$\frac{P(T > 75)}{P(T > 65)} \equiv \frac{s(75)}{s(65)} = \frac{0.23579}{0.45497} \approx 0.518254.$$

Итак, условная вероятность дожития до 75 лет (при условии, что мужчина дожил до 65 лет) в 2 с лишним раза больше безусловной (для новорожденного младенца мужского пола). Тем не менее, этот расчёт показывает, что не только лишь половина мужчин доживёт до пенсионного возраста 65 лет, но и из вышедших на пенсию половина не проживёт и 10 лет[9].

Пример 2. В школе учится 1200 школьников. Осенью 800 из них согласились сделать прививку от гриппа, а 400 отказались. Весной выяснилось, что гриппом переболели 121 человек, из которых 74 человека делали прививку, а 47 – нет.

В общей сложности из 1200 школьников гриппом переболели 121 человек. Поэтому безусловная частота наступления события $A =$ «школьник заболел гриппом» равна $\frac{121}{1200} \approx 10.08\%$. Поскольку число «экспериментов» $n = 1200$ достаточно большое, можно считать, что безусловная вероятность этого события примерно равна 10%: $P(A) \approx 10\%$.

В группе школьников, делавших прививку, доля заболевших равна $\frac{74}{800} = 9.25\%$.

На языке теории вероятностей это означает, что условная частота наступления события

$$A = \text{«школьник заболел гриппом»}$$

при условии

$$B = \text{«школьник делал прививку от гриппа»}$$

равна 9.25%. Поскольку число «экспериментов» $k = 800$ достаточно большое, можно считать, что условная вероятность события A при условии наступления события B примерно равна 9%: $P(A \mid B) \approx 9\%$.

В группе школьников, не делавших прививку, доля заболевших равна $\frac{47}{400} = 11.75\%$. На языке теории вероятностей это означает, что условная частота

[9] Анализ проблемы пенсионного возраста, конечно, намного сложнее. Прежде всего нужно учитывать не всех новорожденных, а лишь тех, кто дожил до трудоспособного возраста. Далее, нужно рассчитывать среднее время работы до выхода на пенсию, среднее остаточное время жизни вышедших на пенсию, учитывать выход на пенсию по инвалидности, сумму пенсионных накоплений с учётом инвестиционной доходности и т.д. Расчёты такого рода проводят пенсионные актуарии – специалисты по применению математических и статистических методов для анализа рисков в страховом деле.

наступления события $A =$ «школьник заболел гриппом» при условии $\bar{B} =$ «школьник прививку от гриппа не делал» равна 11.75%. Поскольку число «экспериментов» $l = 400$ достаточно большое, можно считать, что условная вероятность $P\left(A \mid \bar{B}\right) \approx 12\%$.

Ниже мы обсудим, можно ли на основании этих расчётов сделать вывод о том, что прививка снижает риск заболеть гриппом.

8.2 Формула полной вероятности

Предположим, что события $A_1, \ldots, A_N$ в результате эксперимента обязательно произойдёт одно из этих событий и притом только одно. Используя нашу терминологию, можно сказать, что события $A_1, \ldots, A_N$ являются полной системой попарно несовместных событий или что они образуют разбиение достоверного события на частные случаи $A_1, \ldots, A_N$. Тогда, как мы установили в Теореме 8, для вероятности любого события A верно равенство:

$$P(A) = P(AA_1) + \ldots + P(AA_N).$$

Допустим, что вероятности событий $A_1, \ldots, A_N$ положительны. Тогда к вероятностям $P(AA_1), \ldots, P(AA_N)$, стоящим в правой части, можно применить теорему умножения (27):

$$P(A) = P(A_1) \cdot P(A \mid A_1) + \ldots + P(A_N) \cdot P(A \mid A_N). \qquad (29)$$

Полученное соотношение называется *формулой полной вероятности*.

Польза от формулы полной вероятности (29), так же как и от теоремы умножения вероятностей (27), связана с тем, что на практике для некоторых событий легче определить условные вероятности, чем безусловные. Чтобы показать возможные применения формулы полной вероятности, решим две несложные задачи.

Задача 1. Завод-изготовитель некоторого вида электрических трансформаторов в ходе заводских испытаний установил, что вероятность отказа трансформатора в обычный день равна 0.05, а в жаркий – 0.1. Вероятность того, что в селе, где установлен трансформатор, летний день является жарким равна 0.4. Какова безусловная вероятность отказа трансформатора за один летний день?

Решение. Для произвольного летнего дня рассмотрим события: $A_1 = \{\text{день жаркий}\}$, $A_2 = \{\text{день обычный}\}$, $A = \{\text{трансформатор отказал}\}$. Мы знаем, что $P(A_1) = 0.4$. Событие A_2 противоположно событию A_1; значит, $P(A_2) = 0.6$. Кроме того, мы знаем, что $P(A \mid A_1) = 0.1$, $P(A \mid A_2) = 0.05$. Поскольку события A_1 и A_2 образуют полную систему, можно применить формулу полной вероятности:

$$P(A) = P(A_1) \cdot P(A \mid A_1) + P(A_2) \cdot P(A \mid A_2) = 0.4 \cdot 0.1 + 0.6 \cdot 0.05 = 0.07.$$

Задача 2. Администация школы хочет оценить долю p школьников, которые курят. Для того, чтобы преодолеть понятное нежелание школьников искренне отвечать на этот вопрос, используется следующая процедура. Школьнику задаётся вопрос: «Ты куришь?» Перед тем, как отвечать, школьник бросает игральную кость (на которую учитель, проводящий опрос, не смотрит). Если выпадает 1,2,3 или 4, то школьник отвечает на вопрос правдиво, а если выпадет 5 или 6, то нет. По итогам опроса выяснилось, что 37% школьников ответили «да». Какая часть школьников курит?

Решение. Введём в рассмотрение события: $A_1 = \{\text{школьник курит}\}$, $A_2 = \{\text{школьник не курит}\}$, $A = \{\text{школьник ответил «да»}\}$. Искомая доля p школьников, которые курят – это вероятность события A_1: $P(A_1) = p$. Поскольку событие A_2 является противоположным событием к событию A_1, $P(A_2) = 1 - p$.

Допустим, что школьник курит. Тогда он ответит «да», если при бросании кости выпадет 1,2,3 или 4. Вероятность этого события равна $\frac{2}{3}$. Значит, $P(A \mid A_1) = \frac{2}{3}$.

Допустим, что школьник не курит. Тогда он ответит «да», если при бросании кости выпадет 5 или 6. Вероятность этого события равна $\frac{1}{3}$. Значит, $P(A \mid A_2) = \frac{1}{3}$.

Тот факт, что 37% школьников ответили «да», означает, что $P(A) = 0.37$. Поскольку события A_1 и A_2 образуют полную систему несовместных событий, можно применить формулу полной вероятности:

$$P(A) = P(A_1) \cdot P(A \mid A_1) + P(A_2) \cdot P(A \mid A_2) \Leftrightarrow 0.37 = p \cdot \frac{2}{3} + (1-p) \cdot \frac{1}{3},$$

откуда $p = 11\%$.

8.3 Формула Бейеса

Пусть, как и при выводе формулы полной вероятности, события $A_1, \ldots, A_N$ образуют полную систему попарно несовместных событий и их вероятности положительны, A – некоторое событие, вероятность которого также отлична от 0. Допустим, что событие A осуществилось. Найдём условную вероятность события A_k при этом условии. Начнём с определения условной вероятности:

$$P(A_k \mid A) = \frac{P(A_k A)}{P(A)}.$$

Преобразуем вероятность $P(A_k A)$, стоящую в числителе дроби, по теореме умножения (27): $P(A_k \mid A) = \dfrac{P(A \mid A_k) \cdot P(A_k)}{P(A)}$, а вероятность $P(A)$ – по формуле полной вероятности (29):

$$P(A_k \mid A) = \frac{P(A_k) \cdot P(A \mid A_k)}{P(A_1) \cdot P(A \mid A_1) + \ldots + P(A_N) \cdot P(A \mid A_N)}. \quad (30)$$

Полученное равенство называется формулой Бейеса (Байеса) (Thomas Bayes, 1702-1761 – английский математик).

Чтобы показать возможные применения формулы Бейеса, решим простую задачу.

Задача 3. При производстве изделий некоторого типа из-за случайных неоднородностей используемого сырья и случайных ошибок рабочих часть изделий не удовлетворяет установленным стандартам качества. Статистический анализ выборочных партий изделий позволил установить, что вероятность события

$A =$ «изготовленное изделие не удовлетворяет установленным стандартам»

равна 0.1. Перед поставкой изделий потребителям проводится укоренная проверка их качества. Если изделие нестандартное, то оно отбраковывается с вероятностью 0.95, а с вероятностью 0.05 ошибочно признаёся стандартным. Ошибки возможны и при проверке стандартных изделий: стандартное изделие проходит эту проверку с вероятностью 0.98, а с вероятностью 0.02 ошибочно признаётся нестандартным. Оцените долю нестандартных изделий в большой партии изделий, поставленной потребителям.

Решение. Доля нестандартных изделий в большой партии изделий, поставленной потребителям, примерно равна условной вероятности события A при условии наступления события $B =$ «изделие успешно прошло проверку качества». События $A_1 = A$ и $A_2 = \overline{A}$ образуют полную систему и несовместны. Поэтому по формуле Бейеса (30) мы имеем:

$$P(A \mid B) \equiv P(A_1 \mid B) = \frac{P(A_1) \cdot P(B \mid A_1)}{P(A_1) \cdot P(B \mid A_1) + P(A_2) \cdot P(B \mid A_2)}$$
$$= \frac{0.1 \cdot 0.05}{0.1 \cdot 0.05 + 0.9 \cdot 0.98} \approx 0.56\%.$$

9. Независимые события

9.1 Определение независимости в теории вероятностей

Понятие независимости случайных событий, которое мы сейчас определим, является, видимо, самым важным понятием теории вероятностей. Именно оно выделяет теорию вероятностей из числа других математических дисциплин. Действительно интересные и важные теоремы и приложения теории вероятностей начинаются только после того, как введено понятие независимости случайных событий (и случайных величин).

Перед тем, как определить независимость событий, рассмотрим слегка изменённый вариант Примера 2.

Пример 3. В школе учится 1200 школьников. Осенью 800 из них согласились сделать прививку от гриппа. Весной выяснилось, что из них гриппом переболели 93 человека, в то время как из 400 человек, не делавших прививку, гриппом переболели 47 человек.

В группе школьников, делавших прививку, доля заболевших равна $\frac{93}{800} = 11.625\%$. На языке теории вероятностей это означает, что условная частота наступления события $A =$ «школьник заболел гриппом» при условии, что произошло событие $B =$ «школьник делал прививку от гриппа» равна 11.625%. Поскольку число «экспериментов» k=800 достаточно большое, можно считать, что условная вероятность события A при условии B примерно равна 12%: $P(A \mid B) \approx 12\%$.

В общей сложности из 1200 школьников гриппом переболели 140 человек. Поэтому безусловная частота наступления события $A =$ «школьник заболел гриппом» равна $\frac{140}{1200} \approx 11.67\%$. Поскольку число «экспериментов» n=1200 достаточно большое, можно считать, что безусловная вероятность события A примерно равна 12%: $P(A) \approx 12\%$.

Сопоставляя *безусловную вероятность* заболеть гриппом, $P(A) \approx 12\%$, и *условную вероятность*, $P(A \mid B) \approx 12\%$, при условии, что школьник делал прививку, можно сказать, что $P(A \mid B) = P(A)$, т.е. прививка *практически не меняет долю* заболевших. Именно в этом и **только** в этом смысле в теории вероятностей говорят, что событие $A =$ «школьник заболел гриппом» не зависит от события $B =$ «школьник делал прививку от гриппа».

Имея в виду рассмотренный пример, для произвольного стохастического эксперимента $\mathfrak{S}$, в результате которого могут наступить или нет два случайных события A и B, дают следующее определение (мы считаем ниже, что $P(B)>0$).

Определение. *Событие A не зависит от события B, если верно равенство*

$$P(A\,|\,B) = P(A). \tag{31}$$

С помощью (24) равенство (31) можно записать как:

$$P(AB) = P(A)\cdot P(B). \tag{32}$$

Предполагая, что $P(A) > 0$, соотношение (32) можно переписать и в виде: $\dfrac{P(AB)}{P(A)} = P(B)$ или, применяя определение условной вероятности $P(B\,|\,A)$, в виде: $P(B\,|\,A) = P(B)$. Таким образом, если событие A не зависит от события B, то и событие B не зависит от события A. Поэтому можно говорить о *взаимно независимых событиях*.

Равенство (32) имеет смысл и для событий, вероятность которых равна 0. При этом считается, что событие нулевой вероятности и любое другое событие независимы Поэтому обычно именно его принимают в качестве определения (взаимной) независимости событий.

Определение. *События A и B называются взаимно независимыми, если верно равенство* $P(AB) = P(A)\cdot P(B)$. *Если это равенство не выполнено, то события* A *и* B *называются зависимыми.*

Взаимозависимость двух случайных событий нельзя путать с хорошо известной из школьного курса взаимнооднозначной *функциональной зависимостью* между двумя переменными, когда значение одной переменной однозначно определяет значение другой. Если, например, события

A = «школьник заболел гриппом» и B = «школьник делал прививку от гриппа»

зависимы, это вовсе не означает, что информация о том, что школьник сделал прививку, позволяет совершенно точно сказать, что он не заболеет гриппом. Факт зависимости этих событий означает *только* следующее. Заболеваемость (т.е. доля заболевших) гриппом среди школьников, делавших прививку, будет отличаться (очевидно, в лучшую сторону) от средней заболеваемости по школе. Чтобы подчеркнуть это обстоятельство, в теории вероятностей иногда вместо термина *независимые события* используют термины *стохастически независимые события* или *статистически независимые события*.

Чтобы лучше понять суть понятия независимости событий, докажем простую теорему.

Теорема 14. 4 пары событий: A, B; $A, \bar{B}$; $\bar{A}, B$; $\bar{A}, \bar{B}$, независимы или нет одновременно.

Доказательство. Независимость событий A, B равносильна тому, что разность $D_{AB} = P(AB) - P(A)\cdot P(B)$ равна 0. Аналогично, независимость событий $A, \bar{B}$ ($\bar{A}, B$ или

$\bar{A},\bar{B}$) равносильна тому, что разность $D_{A\bar{B}} = P(A\bar{B}) - P(A)\cdot P(\bar{B})$ (соответственно, $D_{\bar{A}B} = P\left(\bar{A}B\right) - P\left(\bar{A}\right)\cdot P\left(B\right)$ или $D_{\bar{A}\bar{B}} = P\left(\bar{A}\bar{B}\right) - P\left(\bar{A}\right)\cdot P\left(\bar{B}\right)$) равна 0.

Поскольку события B и $\bar{B}$ образуют полную систему событий, верно равенство: $P(A) = P(AB) + P(A\bar{B})$. Кроме того, $P\left(\bar{A}\right) = 1 - P\left(A\right)$ (см. Теоремы 3 и 8 из второй части статьи). Поэтому

$$\begin{aligned} D_{AB} &= P(A) - P(A\bar{B}) - P(A)\cdot P(B) = P(A)\cdot\left(1 - P(B)\right) - P(A\bar{B}) \\ &= P(A)\cdot P\left(\bar{B}\right) - P(A\bar{B}) = -D_{A\bar{B}}. \end{aligned}$$

Если равенство $D_{AB} = -D_{A\bar{B}}$ применить не к паре событий A и B (именно в этом порядке), а к паре событий B и A, мы получим равенство $D_{AB} = -D_{\bar{A}B}$, а если к паре событий $\bar{A}$ и B, то равенство $D_{\bar{A}B} = -D_{\bar{A}\bar{B}}$. Поэтому если одна из разностей D_{AB}, $D_{A\bar{B}}$, $D_{\bar{A}B}$, $D_{\bar{A}\bar{B}}$ равна 0 (что означает независимость соответствующей пары событий), то равны 0 и остальные.

9.2 Способы проверки независимости двух событий

В простых случаях, например, в классической вероятностной модели эксперимента с конечным числом равновозможных событий, вероятности $P\left(AB\right)$, $P\left(A\right)$, $P\left(B\right)$ можно подсчитать и проверить справедливо или нет равенство (32), установив тем самым независимость или зависимость событий A и B.

Предположим, например, что эксперимент заключается в бросании игральной кости. Рассмотрим два события:

$$A = \left\{\text{выпало чётное число очков}\right\},\ B = \left\{\text{выпало не больше 3 очков}\right\}.$$

Ясно, что $P(A) = 3/6 = 1/2$, $P(B) = 3/6 = 1/2$. Одновременное наступление событий A и B означает, что выпало 2 очка, так что $P\left(AB\right) = 1/6$ и, значит, $P\left(AB\right) \neq P\left(A\right)\cdot P(B)$. По определению это означает, что события A и B зависимы.

Если бы вместо события B мы рассмотрели событие $C = \left\{\text{выпало не больше 2 очков}\right\}$, то, как и раньше, одновременное наступление событий A и C означает, что выпало 2 очка, так что $P\left(AC\right) = 1/6$. Но $P\left(C\right) = 2/6 = 1/3$, а $P\left(A\right) = 3/6 = 1/2$ и, значит, $P\left(AC\right) = P\left(A\right)\cdot P(C)$. По определению это означает, что события A и C независимы.

В **реальных** (а не теоретических) приложениях проверка независимости событий непосредственно по определению невозможна в принципе, т.к. само понятие стохастического эксперимента является лишь **моделью** определённого типа наблюдаемых явлений окружающего нас мира. Для любого **реального** эксперимента сам факт, что он является стохастическим, и утверждение о том, что вероятность анализируемого события даётся некоторым конкретным числом, означают лишь одно: эта модель может описывать изучаемое явление. Но справедливость этого утверждения и точность модели могут быть установлены **только** экспериментально.

Если мы уже установили, что анализируемый реальный эксперимент можно считать стохастическим, то экспериментальная проверка независимости интересующих нас событий A и B сводится к следующему:

- найти вероятности событий как частоты их появления в достаточно большой серии экспериментов: $P(A) \approx \nu_n(A)$, $P(B) \approx \nu_n(B)$, $P(AB) \approx \nu_n(AB)$;
- проверить справедливость равенства (32), установив верно или нет приближённое равенство $\nu_n(AB) \approx \nu_n(A) \cdot \nu_n(B)$.

Формально говоря, приближённое равенство не может быть истинным или ложным; для него эти понятия вообще не имеют смысла. Поэтому вопрос о справедливости приближённого равенства $\nu_n(AB) \approx \nu_n(A) \cdot \nu_n(B)$ сводится к следующему вопросу:

какой разностью $\Delta_{AB} = \nu_n(AB) - \nu_n(A) \cdot \nu_n(B)$ можно пренебречь (объяснив её случайными колебаниями) и считать события A и B независимыми? Иначе говоря, насколько сильно должны отличаться величины $\nu_n(AB)$ и $\nu_n(A) \cdot \nu_n(B)$, чтобы мы считали приближённое равенство $\nu_n(AB) \approx \nu_n(A) \cdot \nu_n(B)$ ложным и, следовательно, события A и B – зависимыми?

Хотя строгий математический анализ этой проблемы чрезвычайно сложен и проводится только в университетском курсе математической статистики, смысл этого анализа очень легко понять.

Прежде всего, как мы установили, 4 пары событий: A, B; $A, \bar{B}$; $\bar{A}, B$; $\bar{A}, \bar{B}$ независимы или нет одновременно. Поэтому при взаимной независимости событий A и B маленькими должны быть 4 разности:

$$\Delta_{AB} = \nu_n(AB) - \nu_n(A) \cdot \nu_n(B),\ \Delta_{A\bar{B}} = \nu_n\left(A\bar{B}\right) - \nu_n\left(A\right) \cdot \nu_n\left(\bar{B}\right),$$
$$\Delta_{\bar{A}B} = \nu_n\left(\bar{A}B\right) - \nu_n\left(\bar{A}\right) \cdot \nu_n\left(B\right),\ \Delta_{\bar{A}\bar{B}} = \nu_n\left(\bar{A}\bar{B}\right) - \nu_n\left(\bar{A}\right) \cdot \nu_n\left(\bar{B}\right),$$

а, значит, и их сумма.

Простыми алгебраическими преобразованиями легко показать, что $\Delta_{A\bar{B}} = -\Delta_{AB}$, $\Delta_{\bar{A}B} = -\Delta_{AB}$, $\Delta_{\bar{A}\bar{B}} = \Delta_{AB}$. Поэтому сумма $\Delta_{AB} + \Delta_{A\bar{B}} + \Delta_{\bar{A}B} + \Delta_{\bar{A}\bar{B}}$ обязательно равна 0, так что перед суммированием нужно «стереть знаки» у слагаемых. Самый естественный способ сделать это – взять модули слагаемых. Ещё одна возможность – возвести слагаемые в квадрат. Хотя эта операция совсем не такая естественная, как взятие модуля, именна она приводит к решению проблемы. Итак, вместо величин Δ_{AB}, $\Delta_{\bar{A}B}$, $\Delta_{A\bar{B}}$, $\Delta_{\bar{A}\bar{B}}$ будем рассматривать их квадраты.

Далее, обратим внимание на то, что, например, величина $\Delta_{AB} = \nu_n(AB) - \nu_n(A)\cdot\nu_n(B)$ может быть маленькой просто потому, что малы величины $\nu_n(A), \nu_n(B)$ (поскольку $n(AB) \le \min\big(n(A), n(B)\big)$, тогда автоматически мала и величина $\nu_n(AB)$). Поэтому, как это обычно делают при сравнении величин, нужно рассматривать не *абсолютную* разность $\Delta_{AB} = \nu_n(AB) - \nu_n(A)\cdot\nu_n(B)$, а *относительную*. Правда, не совсем ясно, что брать в качестве базы сравнения. Здравый смысл говорит, что это должна быть некоторая «средняя» величина между $\nu_n(A)$ и $\nu_n(B)$. В теории вероятностей установлено, что в качестве этой «средней» нужно взять среднее геометрическое $\sqrt{\nu_n(A)\cdot\nu_n(B)}$, т.е. вместо Δ_{AB} следует рассматривать $\dfrac{\Delta_{AB}}{\sqrt{\nu_n(A)\cdot\nu_n(B)}}$.

Кроме того, нужно учесть, что по мере роста числа *n* экспериментов колебания частот вокруг соответствующих вероятностей уменьшаются. Иначе говоря, если, например, при $n = 1000$ мы бы считали относительную погрешность, скажем, в 5% приемлемой, то при $n = 10000$ нужно требовать гораздо меньшей погрешности. В теории вероятностей установлено, что для учёта числа экспериментов нужно относительную погрешность $\dfrac{\Delta_{AB}}{\sqrt{\nu_n(A)\cdot\nu_n(B)}}$ делить ещё на $\dfrac{1}{\sqrt{n}}$, т.е. в качестве меры отличия между $\nu_n(AB)$ и $\nu_n(A)\cdot\nu_n(B)$ следует рассматривать $\dfrac{\Delta_{AB}}{\sqrt{\nu_n(A)\cdot\nu_n(B)}}\cdot\sqrt{n}$.

Вспоминая сделанное ранее замечание о необходимости «стереть знаки» у величин Δ_{AB}, $\Delta_{\bar{A}B}$, $\Delta_{A\bar{B}}$, $\Delta_{\bar{A}\bar{B}}$, мы приходим к следующему выводу:

при независимости событий A и B величина

$$X^2 = n\cdot\left(\frac{\Delta^2_{AB}}{\nu_n(A)\cdot\nu_n(B)} + \frac{\Delta^2_{A\bar{B}}}{\nu_n(A)\cdot\nu_n(\bar{B})} + \frac{\Delta^2_{\bar{A}B}}{\nu_n(\bar{A})\cdot\nu_n(B)} + \frac{\Delta^2_{\bar{A}\bar{B}}}{\nu_n(\bar{A})\cdot\nu_n(\bar{B})}\right) \qquad (33)$$

не должна быть большой.

Исключительно важная (и сложная) теорема теории вероятностей утверждает, что если события A и B независимы, то при большом числе экспериментов вероятность того, что величина X^2 превысит значение 4, примерно равна 0.05, вне зависимости от того, каковы вероятности событий A и B (именно для достижения этой независимости от неизвестных вероятностей $P(A)$, $P(B)$ и делались не совсем естественные преобразования вроде возведения в квадрат для «стирания знаков»). Поэтому, если $X^2 > 4$, то имеются веские основания считать, что данные эксперимента противоречат гипотезе о независимости событий A и B, причём чем больше X^2, тем больше достоверность этого вывода. Если же $X^2 \le 4$, то разницу между $\nu_n(AB)$ и $\nu_n(A)\cdot\nu_n(B)$ можно объяснить фактором случайности, что не позволяет отвергнуть гипотезу о независимости.

С помощью соотношений $\Delta_{A\bar{B}} = -\Delta_{AB}$, $\Delta_{\bar{A}B} = -\Delta_{AB}$, $\Delta_{\bar{A}\bar{B}} = -\Delta_{AB}$ легко получить более удобные формулы для подсчёта величины X^2:

$$X^2 = n \cdot \frac{\left[\nu_n(AB) - \nu_n(A)\cdot\nu_n(B)\right]^2}{\nu_n(A)\cdot\nu_n\left(\bar{A}\right)\cdot\nu_n(B)\cdot\nu_n\left(\bar{B}\right)} = n \cdot \frac{\left[n\left(AB\right)\cdot n\left(\bar{A}\bar{B}\right) - n\left(A\bar{B}\right)\cdot n\left(\bar{A}B\right)\right]^2}{n\left(A\right)\cdot n\left(\bar{A}\right)\cdot n\left(B\right)\cdot n\left(\bar{B}\right)}. \quad (34)$$

Рассмотрим ещё раз Пример 2, где мы имели дело с двумя событиями: $A =$ «школьник заболел гриппом», $B =$ «школьник делал прививку от гриппа». Как мы вычислили,

$$\nu_n(A) = \frac{121}{1200} \approx 10.08\%, \nu_n(A \mid B) = \frac{74}{800} = 9.25\%, \ \nu_n\left(A \mid \bar{B}\right) = \frac{47}{400} = 11.75\%.$$

На первый взгляд из этих значений вытекает, что прививка от гриппа уменьшает заболеваемость гриппом. Но не исключено, что различия между безусловной частотой $\nu_n(A)$ и условными частотами $\nu_n(A \mid B)$, $\nu_n\left(A \mid \bar{B}\right)$ можно объяснить влиянием случая. Если это действительно так, то утверждать, что события A и B зависимы, нельзя.

Чтобы ответить на этот вопрос, применим описанную теорию. Чтобы не запутаться в числах, удобно результаты наблюдений оформить в виде таблицы следующего вида (её называют *таблицей сопряжённости признаков*).

Таблица 2

	B	$\bar{B}$	Итого
A	**74**	**47**	*121*
$\bar{A}$	**726**	**353**	*1079*
Итого	*800*	*400*	1200

В клетку таблицы, стоящую на пересечении строки A и столбца B вписывается число $n\left(AB\right) = 74$ и т.д. На пересечении строки A и столбца «Итого» стоит сумма чисел 74 и 47, т.е. число $n(A) = 121$, и т.д. В правом нижнем углу стоит общее число экспериментов $n = 1200$.

Для подсчёта числителя второй дроби в правой части формулы (34) нужно взять числа, отмеченные **жирным** шрифтом, перемножить их по диагоналям, а затем вычесть эти произведения. Для подсчёта знаменателя нужно перемножить числа, отмеченные *курсивом*. Поэтому делать эти вычисления очень легко. В нашем случае величина $X^2 \approx 1.84$. Так как $X^2 < 4$, наблюдаемые различия между безусловной частотой $\nu_n(A)$ и условными частотами $\nu_n(A \mid B)$, $\nu_n\left(A \mid \bar{B}\right)$ не противоречат гипотезе о независимости событий

$A =$ «школьник заболел гриппом», $B =$ «школьник делал прививку от гриппа» и могут быть объяснены случайными колебаниями частот вокруг соответствующих вероятностей.

Следует подчеркнуть, что проведённые рассуждения вовсе не позволяют утверждать, что анализируемые события A и B независимы; мы можем сказать только одно – отвергнуть эту гипотезу нет оснований. Если провести дополнительные статистические исследования, то новые данные могут привести к тому, что мы отвергнем гипотезу о независимости заболеваемости гриппом от прививки. Если же это не произойдёт, то мы укрепимся в уверенности, что прививка влияет (очевидно в положительную сторону) на заболеваемость, и можем считать эту гипотезу «доказанной» (если только новые данные не потребуют отвергнуть гипотезу).

При изучении различных школьных предметов и в повседневной жизни школьники могут найти много ситуаций, когда большая группа объектов (например, людей) классифицируется по двум признакам (назовём их A и B) и возникает проблема, могут ли эти признаки рассматриваться как независимые друг от друга. Например, десятиклассников школы можно разделить на тех, у кого по алгебре за четверть стоит оценка 4 или 5, и тех, у кого эта оценка 3. В качестве второго признака классификации можно взять оценки за четверть тех же школьников, например, по истории. Исследуя вопрос, есть или нет зависимость между успехами по алгебре и успехами по истории, школьники будут находить частоты наступления разных событий, оценивать их безусловные и условные вероятности, сравнивать эти вероятности между собой и, что самое главное, делать ясные, практически интересные выводы. Понять суть теории вероятностей и статистики и продемонстрировать пользу от их изучения можно только при исследовании подобных маленьких проблем, а вовсе не при решении тривиальных задач на расчёт вероятностей событий, связанных с азартными играми, и, тем более, не при решении нелепых задач на расчёт «вероятностей» уникальных событий.

Для многих реальных ситуаций часто факт независимости двух случайных событий A и B устанавливают без экспериментальной проверки, а с помощью интуитивных соображений о том, что наступление одного из этих событий никак не влияет на возможность и тем самым *на частоту* наступления другого. Например, если событие A заключается в том, что рост школьника больше 165 см, а наступление события B означает, что на ЕГЭ по русскому языку он получил больше 50 баллов, то без всякой математики ясно, что эти события независимы в смысле теории вероятностей. Однако, строго говоря, подобные «очевидные» выводы нужно подтверждать экспериментально, причём чем важнее рассматриваемая ситуация, тем тщательнее должен быть анализ результатов статистических наблюдений.

Литература

1. А.Н.Колмогоров. Основные понятия теории вероятностей. 1-е изд. 1933 (нем), 1936 (рус.), 2-е изд., 1974г., 3-е изд. М.:Фазис, 1998.

2. Б.В.Гнеденко, А.Я.Хинчин. Элементарное введение в теорию вероятностей. 14-е изд., М.: УРСС, 2016. ISBN 978-5-9710-3300-4

3. А.Н.Колмогоров «Теория вероятностей». В кн.: Математика, её содержание, методы и значение. Изд-во АН СССР, Москва, 1956, Том 2. Глава XI, стр 252-284.

4. А.Н.Колмогоров. Статья «Вероятность». Математическая энциклопедия. Москва, Изд-во «Советская энциклопедия», 1982.

5. Ю.В.Прохоров, Б. А. Севастьянов. Статья «Вероятностей теория». Математическая энциклопедия. Москва, Изд-во «Советская энциклопедия», 1982.

6. Б.В.Гнеденко. Курс теории вероятностей. 8-е изд., М.: УРСС, 2005.

7. В.Феллер. Введение в теорию вероятностей и её приложения. Том 1. М.:Мир, 1984.

8. Г.Крамер. Математические методы статистики. 2-е изд. М.:Мир, 1975.

9. Ф.Клейн. «Элементарная математика с точки зрения высшей. том 1. Арифметика. Алгебра. Анализ.» М.: Наука, 1987.

10. В. И. Арнольд. О преподавании математики. Успехи математических наук, 1998, 53, вып.1, стр.229-234.

11. John Venn. The Logic of Chance: An Essay on the Foundations and Province of the Theory of Probability, with Especial Reference to Its Application to Moral and Social Science. Macmillan and Co., London – Cambridge, 1866.

12. Рихард Мизес. Вероятность и статистика: Пер. с нем./ Под ред. и с предисл. А.Я.Хинчина. Изд. 5-е. – М.: Книжный дом «ЛИБРОКОМ», 2009.

13. Б.В.Гнеденко К шестой проблеме Гильберта. В кн. Проблемы Гильберта. Под ред. П.С.Александрова. М.:Наука, 1969.

14. Э.Борель. Случай. Государственное издательство, Москва-Петроград, 1923.

15. Э.Борель. Вероятность и достоверность. Изд. 3-е, Изд-во «Наука», Москва. 1969.

16. А.Пуанкаре. О науке. 2-е изд. М.: Наука, 1990.

17. R. von Mises, Grundlagen der Wahrscheinlichkeitsrechnung. Mathematische Zeitschrift, 1919, vol. 5, pp. 52-99.

18. А.Н.Ширяев. Вероятность – 1. М.: МЦНМО, 2007.

19. В.А.Успенский, А.Л.Семёнов, А.Х.Шень. Может ли (индивидуальная) последовательность нулей и единиц быть случайной? Успехи математических наук, 1990, том 45, вып.1, стр.105-162.

20. В.Н.Тутубалин. Границы применимости (вероятностно-статистические методы и их возможности). Изд-во «Знание», Москва, 1977 (Новое в жизни, науке, технике. Серия «Математика, кибернетика», 1/1977).

21. Д.Пойа. Математика и правдоподобные рассуждения. Издание второе. Изд-во «Наука», Москва, 1975.