МАТЕМАТИКА В ПОЛИТЕХНИЧЕСКОМ УНИВЕРСИТЕТЕ

А.Н. ФИРСОВ

МАТЕМАТИКА

ТЕОРИЯ ВЕРОЯТНОСТЕЙ

Часть 1

Учебное пособие

Санкт-Петербург
Издательство Политехнического университета
2005

Федеральное агентство по образованию

САНКТ-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ ПОЛИТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ

А.Н. ФИРСОВ

ТЕОРИЯ ВЕРОЯТНОСТЕЙ

Часть I

УЧЕБНОЕ ПОСОБИЕ

Санкт-Петербург
2005

ББК 22.17я73
Ф 627

Фирсов А.Н. **Математика. Теория вероятностей. Ч. 1**: Учеб. пособие. СПб.: Изд-во Политехн. ун-та. 2005. 112 с. (Математика в политехническом университете. Вып. 6).
ISBN 5-7422-0944-4

Пособие написано на основе курса лекций по теории вероятностей, читаемого автором студентам третьего курса Санкт-Петербургского государственного политехнического университета, обучающимся по специальностям 553000 «Системный анализ и управление» и 071900 «Информационные системы в технике и технологии».

Данное пособие охватывает первую часть курса, а именно основные классические разделы дискретной теории вероятностей. Большое внимание уделяется логическим основам теории и характерным особенностям практического применения вероятностных методов. В книге достаточно много подробно разобранных примеров, иллюстрирующих основные понятия и методы дискретной теории вероятностей. Основной материал книги не предполагает знакомство читателя с полным вузовским курсом высшей математики, однако ориентирован на читателя, обладающего определенной математической культурой.

Пособие будет также полезно студентам техникумов и вузов с сокращенной программой по высшей математике и лицам, желающим познакомиться с основными идеями и методами теории вероятностей самостоятельно.

Печатается по решению редакционно-издательского совета Санкт-Петербургского государственного политехнического университета.

ISBN 5-7422-0944-4

© Фирсов А.Н., 2005
© Санкт-Петербургский государственный политехнический университет, 2005

Пособие написано на основе курса лекций по теории вероятностей, читаемого автором студентам третьего курса С.-Петербургского государственного политехнического университета, обучающимся по направлению 220100 «Системный анализ и управление» и специальности 230201 «Информационные системы и технологии».

Данное пособие охватывает первую часть курса, а именно основные классические разделы дискретной теории вероятностей. Большое внимание уделяется логическим основам теории и характерным особенностям практического применения вероятностных методов. В книге достаточно много подробно разобранных примеров, иллюстрирующих основные понятия и методы дискретной теории вероятностей. Основной материал книги не предполагает знакомство читателя с полным вузовским курсом высшей математики, однако ориентирован на читателя, обладающего определенной математической культурой.

Пособие будет также полезно студентам техникумов и вузов с сокращенной программой по высшей математике и лицам, желающим познакомиться с основными идеями и методами теории вероятностей самостоятельно.

В электронной версии исправлены замеченные опечатки и неточности печатного издания.

СОДЕРЖАНИЕ

ПРЕДИСЛОВИЕ

Это пособие возникло в результате обработки материала лекций по теории вероятностей, которые автор читает студентам третьего курса Санкт-Петербургского государственного политехнического университета, обучающимся по специальностям 553000 «Системный анализ и управление» и 071900 «Информационные системы в технике и технологии». Пособие охватывает первую часть курса, и посвящено элементарным основам теории. Здесь мы постараемся сформулировать те задачи, которые автор ставил перед собой при подготовке данного пособия.

По нашему глубокому убеждению, для того, чтобы успешно использовать на практике идеи и методы той или иной теории, необходимо глубоко разбираться в *основаниях* предмета, в логике и «идеологии» построения его исходных, основополагающих понятий. По отношению к теории вероятностей это справедливо вдвойне. Действительно, вероятностный подход к описанию физических (и, тем более, социальных) явлений имеет свою специфическую логику, игнорирование которой может привести к неграмотному использованию вероятностных методов и, как следствие, к грубым, иногда катастрофическим, ошибкам в выводах. Тем более, теория вероятностей как раздел математики – наука очень молодая: основы *математически удовлетворительного* подхода к построению математических моделей случайных явлений были заложены лишь в 30-х годах XX века академиком А.Н. Колмогоровым.[1] А широкому (и грамотному!) применению теории вероятностей в прикладных, в первую очередь, инженерных задачах по существу не больше полувека. Более того, само признание случайности как непременной и объективной характеристики некоторых физических явлений еще в начале прошлого века было отнюдь не всеобщим, несмотря на успехи квантовой механики. Хорошо известна дискуссия по этому поводу Нильса Бора и Альберта Эйнштейна. Последний до конца жизни (1955 г.) так и не признал, что вероятностный подход к описанию квантовых явлений имеет объективный, внутренний для этих процессов характер. Здесь нелишне вспомнить основателя кинетической

[1] Андрей Николаевич Колмогоров (1903-1987) – один из наиболее видных математиков XX века, один из немногих современных «математических полиглотов». Получил фундаментальные результаты в различных областях математики, был основателем многих направлений исследований в современной математической науке. Его аксиоматика теории вероятностей общепризнанна и лежит в основе современной трактовки этой науки.

теории газов[2] Людвига Больцмана (1844 – 1906), который, по сути, первый стал широко применять в принципы вероятностного подхода к описанию физических явлений, за что и был подвергнут жесткой критике со стороны многих современных ему известных физиков.

Суть упомянутых трудностей в том, что логические алгоритмы постановки задач, связанных со случайными явлениями весьма специфичны и довольно сильно отличаются от традиционных логических алгоритмов постановки задач алгебры и анализа. Собственно, мы здесь имеем дело с ситуацией, аналогичной постановке задач математической физики: *при формулировании* адекватной явлению или процессу математической модели важнее хорошая физическая интуиция, нежели опыт формально-математического анализа; последний же является единственно приемлемым при исследовании *уже сформулированной математической* задачи. Иными словами, *при формулировании исходных математических соотношений* (например, дифференциальных уравнений, связывающих значимые для данного явления количественные характеристики) представление о физической сути процесса и физическая интуиция позволяют ограничиваться «физическим уровнем строгости», в то время как *анализ этих математических моделей* должен проводиться уже на строго формальном, математическом уровне строгости. Поскольку логика «физического уровня строгости» и логика «математического уровня строгости» по сути своей мало совместимы, то это приводит к тому, что профессиональные математики часто с трудом ориентируются в логических построениях профессиональных физиков.[3]

В теории вероятностей – та же ситуация: постановка вероятностных задач, т.е. математическое моделирование случайных явлений, требует специфической логики анализа моделируемых процессов, что и позволяет построить адекватную математическую модель, т.е. сформулировать корректную математическую задачу. Именно эта часть постановки вероятностных задач вызывает у человека, впервые сталкивающегося с проблемой случайности, наибольшие трудности: какую именно вероятностную модель использовать в данной задаче, какие вероятностные характеристики являются в этой задаче определяющими и т.п. Более того, одним из важнейших условий эффективного использования вероятностных методов является ясное понимание того, в каких условиях эти методы «работают», а в каких нет, ибо далеко не все процессы и явления, связанные с неопределенностью, допускают адекватное исследование существующими вероятностно-статистическими методами. Это же касается и правильной интерпретации результатов, полученных такими методами.

---

[2] Больцман общепризнан также как один из основателей статистической физики – одного из важнейших и основополагающих разделов современной физики.

[3] Между прочим, профессиональные физики тоже с трудом воспринимают необходимость того строгого формализма, который является основой математического исследования.

Все, сказанное выше, и было положено в основу выбора содержания и формы изложения материала данного пособия. Мы постарались акцентировать внимание читателя на идеологию подхода и логику *постановки вероятностных задач*. Мы старались также обратить внимание читателя скорее на *основания* теории вероятностей, нежели на математический аппарат, используемый при решении вероятностных задач.

Первые шесть параграфов книги не предполагают наличия у читателя знания математики, выходящего за рамки курса средней школы. Предполагается, однако, что он обладает достаточно развитой культурой «математического мышления». Исключение составляет материал двух последних параграфов, для понимания которого требуется знание стандартного курса высшей математики, излагаемого обычно на первом году обучения технического или педагогического вуза.

В пособии достаточно много примеров, иллюстрирующих основные понятия и результаты теории. Однако, его не следует рассматривать в качестве сборника задач.

Объем излагаемого материала является, на наш взгляд, минимальным в том смысле, что если вы им не овладели, нельзя считать, что вы имеете более или менее содержательное представление о теории вероятностей.

Материал данного пособия является естественным введением в более «математизированные» разделы теории вероятностей, такие как теория непрерывных распределений и теория случайных процессов. С другой стороны, эта элементарная часть курса представляет собой некоторое замкнутое целое, и может служить основой курса теории вероятностей для техникумов и вузов с сокращенной программой по высшей математике.

Любое учебное пособие пишется не на пустом месте, тем более, когда дело касается традиционных дисциплин. В конечном счете, оно во многом является авторской переработкой (приспособленной для решения педагогических задач, которые ставит перед собой автор) материала, содержащегося в различных учебниках и монографиях. Это нормально, и данное пособие – не исключение. Естественно, немалую роль в выборе материала и формы изложения сыграли вкусы и опыт преподавания автора. В той или иной степени, в данном пособии использованы материалы и методические приемы из следующих книг:

Вентцель Е.С.. Овчаров Л.А. Теория вероятностей: Задачи и упражнения. – М.: Наука, 1969.

Гнеденко Б.В. Курс теории вероятностей. – М.: Гос. изд-во физ.-мат. лит-ры, 1961.

Гончаров В.Л. Курс теории вероятностей. – М.-Л.: ГОНТИ, 1939.

Гончаров В.Л. Теория вероятностей. – М.-Л.: Оборонгиз, 1939.

Пустыльник Е.И. Статистические методы анализа и обработки наблюдений. – М.: Наука, 1968.

Реньи А. Трилогия о математике. – М.: Мир, 1980.

Розанов Ю.А. Лекции по теории вероятностей. – М.: Наука, 1986.

Тутубалин В.Н. Теория вероятностей: Краткий курс и научно-методические замечания. – М.: Изд-во МГУ, 1972.

Феллер В. Введение в теорию вероятностей и ее приложения. Т. 1. – М.: Мир, 1967.

Санкт-Петербург, июнь 2005 г.

ВВЕДЕНИЕ

Все процессы, происходящие в природе, являются результатом взаимодействия многих факторов. Для того чтобы изучить эти процессы и в дальнейшем ими управлять, необходимо выяснить, какую роль в рассматриваемом процессе играет каждый фактор в отдельности. Так, например, изучая движение тела, необходимо выяснить, какие силы приводят его в движение, какие тормозят; наконец, каким образом само движущееся тело влияет на эти силы. Все эти факторы необходимо выразить в каких-то количественных оценках — после этого на помощь исследователю приходит могущество математических методов.

Таким образом, математические методы изучения взаимодействующих факторов требуют умения выражать действие различных факторов количественно. Чтобы получить необходимые числовые данные, нужно произвести серию наблюдений. Наблюдение — это решающее звено всякого эксперимента, всякого исследования.

Однако даже самый тщательно подготовленный эксперимент не позволяет выделить интересующий нас фактор в чистом виде. Мы не в силах изолировать многие посторонние факторы: например, изучая падающие тела, мы не можем избежать действия вращения Земли; изучая химические реакции, мы никогда не имеем дела с чистыми веществами; изучая электронные процессы, не можем вести их в абсолютном вакууме и т.д. Наконец, нужно вспомнить о различных помехах, связанных с окружающей обстановкой - ведь даже шум идущего по улице автомобиля сказывается на проводимом в лаборатории эксперименте.

Следовательно, каждое наблюдение дает нам лишь результат взаимодействия основного изучаемого фактора с многочисленными посторонними. Некоторые из этих факторов (например, вращение Земли) можно учесть, так как они сами по себе достаточно хорошо изучены. Учет других факторов (например, наличие примесей в веществах) может быть очень громоздким. Он сильно затянет эксперимент, сделает его неоправданно дорогим. Наконец, многие факторы (помехи) бывают настолько неожиданными, что их вообще нельзя учесть. Сюда же нужно отнести и те факторы, о которых на данном этапе развития науки вообще ничего не известно.

Из сказанного выше можно сделать только один вывод: *полное и точное описание какого-либо процесса возможно лишь в том случае, если известны все факторы, влияющие на этот процесс. Иными словами, такое* *описание вообще невозможно.*

К счастью, оно и не нужно.

Большинство измеряемых на практике величин обладает свойством непрерывности, т. е. их значения сплошь заполняют некоторый числовой промежуток. Однако все применяемые при этом измерительные приборы обладают некоторым пределом точности (разрешающей способностью) - минимальной разницей в значениях двух величин, которую они в состоянии обнаружить. Этот предел обычно указывается на приборах, изготовленных в заводских условиях. Например, аналитические весы, взвешивающие с точностью до 0,1 мг, не смогут различить такие веса, как 12,52 и 12,54 мг, и в обоих случаях покажут 12,5 мг. В результате все дальнейшие вычисления, связанные с этими данными, также будут содержать некоторую неточность, даже если пользоваться абсолютно точными и полными формулами, описывающими исследуемый процесс.

С другой стороны, нужно учесть, что полученные данные не всегда удается полностью использовать в дальнейшем - приходится округлять их, теряя добытую с таким трудом драгоценную точность. В этом отношении можно привести интересный пример. Некоторые математики XVIII—XIX веков увлекались вычислением числа π с высокой точностью. Математик Шенкс в шестидесятых годах позапрошлого столетия вычислил π с точностью до 707-го знака после запятой, потратив на это всю свою жизнь.

Однако подобная точность еще нигде не была использована. Так, например, чтобы вычислить с точностью до микрона длину окружности с центром на Земле и радиусом до ближайшей звезды, т. е. $R = 4{,}5$ световых лет, достаточно иметь число π с 25 знаками после запятой — даже в таком, заведомо бессмысленном вычислении 682 шенксовских знака остаются лишними!

Любое увеличение точности при измерениях сильно усложняет эксперимент. Кроме того, добавление каждого лишнего знака усложняет вычисления на 10 — 15%. Поэтому нужно всегда хорошо знать ту точность, которая потребуется от результата, не стремясь к излишней точности измерений и вычислений.

Итак, в наших наблюдениях всегда допускается некоторая «законная» неточность, величину которой можно рассчитать заранее. Благодаря этому мы можем не учитывать те посторонние факторы, действие которых намного меньше этой неточности; например, изучая движение тел на Земле, можно не учитывать силы тяготения между этими телами или кривизну Земли при малых перемещениях и т. д.

Однако и здесь возникают свои трудности. Рассмотрим для примера такой вопрос: нужно ли, изучая движение автомобиля, учитывать тепловое колебание молекул, из которых он состоит. Ответ напрашивается отрицательный. Но давайте внимательней присмотримся к движению молекул внутри твердого тела. Молекулы колеблются вокруг положения равновесия с достаточно большими скоростями, однако движение это хаотическое, так что молекулы, движущиеся в одну сторону, уравновешиваются молекулами, движущимися в противоположную сторону. При этом можно представить такое случайное стечение обстоятельств, когда большинство молекул двинется в одну сторону, не уравновешиваясь с другой стороны. Разумеется, в результате в эту же сторону резко переместится и сам автомобиль. Такая возможность вполне допустима теоретически. Как же быть: учитывать ее или нет?

Но ведь если такое стечение обстоятельств и возможно, то оно необычайно редко! — ответит читатель. Во всяком случае, оно ни разу не встречалось за всю автомобильную практику человечества.

Правильно! И в результате мы приходим к еще одному важному выводу: *даже сильные отклонения можно не учитывать, если они достаточно редки*.

При этом, правда, мы рискуем однажды попасть именно на то самое несчастное стечение обстоятельств, однако риск здесь может быть не велик, а облегчение исследований будет достаточно большое.

По этой же причине удается избежать детального исследования многочисленных непредвиденных (случайных) помех. Хотя действие каждой из них может оказаться вполне заметным, в общей массе они, как правило, уравновешивают друг друга, лишь *изредка* давая заметный суммарный эффект.

Естественно, пренебрегать можно лишь теми отклонениями, которые действительно редки — в противном случае риск будет слишком велик, равносилен беспечности. Значит, эту меру риска надо оценивать, устанавливая допустимый предел для каждых конкретных обстоятельств.

Иными словами, нужно научиться *численно характеризовать, насколько редко то или иное отклонение*.

Для того чтобы выяснить, является ли какое-то событие редким и насколько, необходимо *провести очень большее число наблюдений*, ибо *нужно иметь возможность сравнивать это событие с другими*. Частота или редкость познаются только в сравнении. Это можно подтвердить следующей юмористической историей: некий путешествующий англичанин дважды проезжал Париж и оба раза попадал в дождь; после этого он записал в дневнике: «Париж — ужасный город, здесь всегда идет дождь!».

Чем больше проведено наблюдений, тем лучше можно оценить редкость интересующего нас события. Но и этот процесс не может продолжаться бесконечно. Следовательно, при определении редкости события мы опять вынуждены идти на риск. Эту долю риска опять нужно оценивать и т. д.

Подобные рассуждения быстро завели бы нас в тупик, если бы каждое случайное событие нужно было изучать заново. Оказывается, однако, что случайные, непредвиденные события в массе своей подчиняются некоторым общим неслучайным закономерностям.

Наука, изучающая *закономерности массовых случайных событий*, называется ***теорией вероятностей***.

Математической статистикой называется наука, занимающаяся *методами обработки опытных данных*, *полученных в результате наблюдений над случайными явлениями*.

Но случайность остается случайностью, и никакие теории при наличии непредвиденных и случайных факторов *не могут давать точные и однозначные ответы*. *Основная задача математической статистики при обработке наблюдений — оценить риск той или иной ошибки в полученном результате*. Принять или не принять этот риск — дело исследователя (или того, кто использует результаты статистических оценок). В том случае, если этот риск его не устраивает, он должен найти пути его уменьшения: применить более точную методику наблюдений, устранить наиболее заметные помехи и т. д. Либо, наконец, не идти на этот риск.

§ 1. СЛУЧАЙНЫЕ СОБЫТИЯ И ИХ ВЕРОЯТНОСТИ

1.1. Понятие случайного события.

Основу изучения различных процессов, происходящих в природе, составляет выяснение всевозможных причинно-следственных связей между отдельными явлениями путем эксперимента. Осуществив по своему желанию одно или несколько первоначальных явлений-причин (в дальнейшем они называются *факторами*), экспериментатор получает возможность изучать различные появляющиеся при этом явления-следствия.

При этом, как правило, экспериментатор заранее намечает совокупность тех явлений-следствий, появления которых он ожидает и изучению которых посвящает свой эксперимент.

Изучаемые явления-следствия могут интересовать экспериментатора с самых различных сторон. Так, если исследуется вещество, полученное в результате химических реакций, то могут быть отмечены такие его характеристики, как цвет, вес, объем и многое другое. Однако самое сложное явление всегда можно мысленно разбить на такие мелкие частные явления, относительно которых остается выяснить только одно: произошли они или не произошли. Например, рассмотрев в качестве отдельных явлений всевозможные цвета спектра, мы в дальнейшем, зная о каждом цвете, осуществился он или нет, немедленно можем указать совокупный цвет полученного в результате химической реакции вещества. Точно так же, измеряя вес вещества, мы в качестве отдельных частных явлений можем рассматривать всевозможные априорные значения этого веса.

Частные явления указанного типа могут иметь различную природу, но каждое интересует нас только с одной точки зрения: произошло оно или не произошло, осуществилось или не осуществилось.

Определение 1.1. *Явления, рассматриваемые только с той точки зрения, осуществились они или не осуществились, называются* ***событиями****.*

Применительно к событиям ставится следующая ***основная задача***: *предсказать, появится ли изучаемое событие при осуществлении некоторого наперед заданного комплекса факторов (явлений-причин).*

Чтобы решить эту задачу (а также при формировании исходного комплекса факторов при постановке задачи), нужно, очевидно, знать влияние каждого фактора на интересующее нас событие. В частности, вряд ли имеет смысл включать в состав этого комплекса факторы, от которых интересующее нас событие наверняка не зависит. С другой стороны, желательно включить в комплекс те факторы, влияние которых на это событие существенно. Во всяком случае, состав комплекса факторов в большой степени зависит от научного опыта, знаний и интуиции исследователя.

Определение 1.2. *Событие, которое при заданном комплексе факторов обязательно произойдет, называется* ***достоверным событием****.*

Например, в замкнутой цепи, состоящей из хорошо соединенных проводников и исправного источника тока, появление электрического тока есть событие достоверное.

Определение 1.3. *Событие, которое не может осуществиться при заданном комплексе факторов, называется* ***невозможным событием****.*

Так, невозможным событием является электрический ток в разомкнутой цепи (при отсутствии проводимости через воздух).

Суждение о достоверности или невозможности некоторого события является категорическим суждением — именно такие суждения принято считать окончательным результатом исследования. Отсюда возникает интерес к обратной задаче: указать такие комплексы факторов, при которых о заданном событии можно сделать категорические суждения (достоверность, невозможность).

Каждое событие является результатом действия большого числа факторов. Все их нужно знать для того, чтобы суждение о событии стало категорическим. Если же заданная совокупность факторов по отношению к событию неполна, то и категорическое суждение о событии становится невозможным. Получается следующая ситуация: заданные факторы благоприятствуют событию и, значит, оно может произойти; с другой стороны, этих факторов недостаточно, чтобы гарантировать событие, и, значит, оно может и не произойти.

Определение 1.4. *Событие, которое при заданном комплексе факторов может либо произойти, либо не произойти, называется* ***случайным событием****.*

С примерами случайных событий мы встречаемся на каждом шагу. Какой номер автобуса раньше подойдет к остановке, на которой мы ожидаем; какая будет завтра погода; какой стороной упадет подброшенная вверх монета — *везде, где отсутствует полная информация, появляется случайность*.

Понятия достоверного, невозможного и случайного события являются относительными, они связаны с заданным комплексом факторов. Достаточно этот комплекс изменить, как сразу может измениться характер события. Например, для замкнутой проводящей электрической цепи с исправным источником тока наличие электрического тока в цепи есть явление достоверное. Но достаточно лишить нас сведений хотя бы об одном элементе этой цепи, и наличие тока в сложившихся условиях станет событием случайным.

Таким образом, *случайность события связана с наличием факторов, влияющих на это событие, но не вошедших в заданный комплекс*. Эти факторы по отношению к заданному комплексу называются случайными. Случайным будет любой фактор, не вошедший в заданный комплекс факторов, даже если он хорошо изучен.

Такое определение, разумеется, является только формальным — оно удобно для построения теории случайных событий. На практике чаще всего нет нужды объявлять случайными хорошо изученные факторы. В качестве случайных здесь рассматриваются обычно факторы, которые по тем или иным причинам невозможно (либо очень трудно) учесть. Эти факторы чаще всего не имеет смысла отделять друг от друга, поскольку связанные с ними причинно-следственные связи все равно не учитываются. Поэтому в дальнейшем мы, как правило, будем говорить об одном объединенном факторе случайности.

Случайность события никоим образом не связана с личными качествами исследователя — его способностями учитывать или предсказывать явления. Эта случайность связана с совершенно объективным фактом сужения комплекса факторов, обеспечивающих (или делающих невозможным) рассматриваемое событие.

Факторы, входящие в заданный комплекс, мы будем называть ***неслучайными или основными***. Влияние таких факторов на исследуемое событие должно быть строго определенно и неизменно — лишь тогда их можно включать в заданный комплекс.

1.2. Вероятность случайного события.

Для того чтобы выяснить, произойдет или не произойдет некоторое событие при заданном комплексе основных факторов, нужно, прежде всего, осуществить этот комплекс. Каждое такое осуществление принято называть *испытанием*[4]. Испытанием является, в частности, любой эксперимент, в результате которого производятся наблюдения. Ожидание автобуса, подбрасывание монеты в приводившихся примерах — тоже испытания.

Предсказать результат единичного испытания можно лишь для достоверных или невозможных событий. Случайность же события вообще не видна при единичном испытании: если событие произойдет, оно может

[4] Как синонимы используются еще термины *опыт, эксперимент, наблюдение.*

показаться нам достоверным, если не произойдет — невозможным.

Практика показывает, что события, сами по себе случайные, в большой массе, *при наличии неизменности заданного комплекса основных факторов*, начинают подчиняться некоторым неслучайным закономерностям. Эти закономерности получили название *вероятностных*, а наука, изучающая вероятностные закономерности, стала называться ***теорией вероятностей***.

Таким образом, *теория случайных событий может появиться лишь при большом числе испытаний, лишь для массовых событий*.

Какой же характер имеют эти закономерности? Рассмотрим в качестве простейшего примера испытание — подбрасывание монеты. Событием пусть будет выпадение герба. Никто не возьмется предсказывать определенно, выпадет или не выпадет герб при одном подбрасывании. Но если это испытание производится большое число раз, то мы с уверенностью можем ожидать, что герб выпадет примерно в половине случаев. Более того, если число выпадений герба сильно отличается от половины числа всех подбрасываний, то мы будем утверждать, что монета не совсем симметрична, иными словами, такое совершенно случайное событие, как выпадение герба при подбрасывании монеты, становится эталоном качества монеты! Ясно, что такая закономерность не может быть случайной.

Примеры подобного рода можно легко продолжить. При большом числе бросаний игрального кубика, на гранях которого нанесены точки в количестве от 1 до 6, каждая грань должна выпадать *в среднем* один раз из шести. Вытаскивая карту из колоды, мы примерно в четверти всех возможных случаев будем вытаскивать карту нужной масти и т. д.

Как мы видим, указанные закономерности не связаны с физической природой событий, а лишь с *числом появлений этих событий при большом числе испытаний*. Эти закономерности фактически опираются на следующую **аксиому**: *если по каким-либо соображениям возможности осуществления некоторых событий одинаковы, то эти события и происходить должны одинаково часто*. Так, у монеты есть две одинаковых грани, у кубика их шесть, у колоды карт - четыре масти с одинаковыми шансами быть вытащенными. События с одинаковыми возможностями осуществления мы будем называть в дальнейшем *равновозможными*.

Вероятностные закономерности наблюдаются и для событий с разными возможностями, хотя их и труднее предугадать. Например, если при бросании не совсем симметричной монеты из 1000 случаев герб выпал 600 раз, то мы вправе ожидать, что и в следующей тысяче подбрасываний число выпадений герба опять будет близко к 600, т.е. частота[5] этого события остается почти неизменной.

Хотя указанная закономерность проявляется лишь при большом числе испытаний, она должна быть связана со свойствами самого события, его внутренними характеристиками; в противном случае каждая серия

---

[5] ***Частотой*** события называется отношение числа испытаний, в которых событие произошло, к числу всех проведенных испытаний.

испытаний имела бы свои закономерности.

Какими же свойствами должна обладать эта внутренняя характеристика события, управляющая вероятностными закономерностями? Желательно, чтобы это было число, одинаковое для равновозможных событий и в какой-то мере связанное с частотой. Рассматривать непосредственно частоту нельзя, так как частота сама есть случайная величина, зависящая от конкретной серии испытаний. *Замечено, правда, что при очень большом числе испытаний частота почти перестает изменяться, приближаясь к некоторой величине, которую и можно принять за искомую характеристику*. В таблице 1.1 приведены для примера результаты, полученные некоторыми экспериментаторами при бросании монеты. Как видно из таблицы, частота здесь приближается к числу ½.

Таблица 1.1

Экспериментатор	Число бросаний	Число выпадений герба	Частота
Бюффон	4 040	2 048	0,5080
К. Пирсон	12 000	6 019	0,5016
К. Пирсон	24 000	12 012	0,5005

В таблице 1.2 представлено число выпадений «герба» в 100 сериях по 100 испытаний. Как мы видим, частота снова колеблется около значения **½.**

Таблица 1.2

Число испытаний	Число гербов										Общее число гербов
0— 1000	54	46	53	55	46	54	41	48	51	53	501
— 2000	48	46	40	53	49	49	48	54	53	45	485
— 3000	43	52	58	51	51	50	52	50	53	49	509
— 4000	58	60	54	55	50	48	47	57	52	55	536
— 5000	48	51	51	49	44	52	50	46	53	41	485
— 6000	49	50	45	52	52	48	47	47	47	51	488
— 7000	45	47	41	51	49	59	50	55	53	50	500
— 8000	53	52	46	52	44	51	48	51	46	54	497
— 9000	45	47	46	52	47	48	59	57	45	48	494
—10000	47	41	51	48	59	51	52	55	39	41	484

Числовая характеристика случайного события, обладающая тем свойством, что для любой достаточно большой серии испытаний частота события лишь незначительно отличается от этой характеристики, называется ***вероятностью события***.

Это определение позволяет вычислять вероятности таких событий, о структуре которых ничего неизвестно и частоту которых нельзя предсказать заранее. Например, только статистические данные за многие годы позволили найти вероятности рождения мальчиков и девочек. Оказалось, что эти вероятности отличны от ½: вероятность рождения мальчиков равна примерно 0,52.

Еще один интересный пример дает таблица частот употребления букв русского алфавита (включая «пробел» между словами), составленная на основании анализа большого количества текстов достаточно большого размера (табл. 1.3).

Таблица 1.3

—	о	е, ё	а	и	т	н	с
0,175	0,090	0,072	0,062	0,062	0,053	0,053	0,045
р	в	л	к	м	д	п	у
0,040	0,038	0,035	0,028	0,026	0,025	0,023	0,021
я	ы	з	ь, ъ	б	г	ч	й
0,018	0,016	0,016	0,014	0,014	0,013	0,012	0,010
х	ж	ю	ш	ц	щ	э	ф
0,009	0,007	0,006	0,006	0,004	0,003	0,003	0,002

Замечание. Если статистическая вероятность случайного события существует, то для нее очевидны следующие **свойства**:

- *вероятность достоверного события равна единице;*
- *вероятность невозможного события равна нулю;*
- *вероятность произвольного случайного события есть положительное число, не превосходящее единицы;*
- *вероятности равновозможных событий одинаковы.*

Статистическое определение вероятности является самым широким по числу охватываемых событий. Оно ничего не требует от события, кроме *принципиальной возможности проводить над ним сколь угодно большое число испытаний*.

Таким образом, в основе теории вероятностей лежит предположение о том, что рассматриваемые события обладают следующим ***свойством статистической устойчивости, или устойчивости частот****:*

- *имеется принципиальная возможность проводить сколь угодно большое число испытаний (в условиях неизменности заданного комплекса основных факторов), в результате которых может наблюдаться данное событие;*
- *для любой достаточно большой серии испытаний частота события лишь незначительно отличается от некоторого постоянного числа. Это число называется* ***вероятностью*** *данного события.*

Приведенное описание свойства устойчивости частот и определение вероятности вряд ли кому-либо покажутся вполне удовлетворительными. Нельзя избежать вопроса о том, насколько сильно при данном количестве испытаний частота события может отличаться от его вероятности? Вообще, каким образом следует производить проверку устойчивости частот, решая вопрос о применимости методов теории вероятностей к данному конкретному явлению? Следует отметить, что в настоящее время нет общего научного ответа на эти вопросы.

Из сказанного ясно, что вопрос о применимости вероятностных методов в каждом отдельном случае решается на интуитивном уровне (интуиция, конечно, основана на личном и общенаучном опыте). Научная добросовестность требует от каждого исследователя применения доступных методов проверки статистической устойчивости, но наличие ее редко можно вполне гарантировать.

Прогресс физики привел, однако, к тому, что теперь более чем когда-либо прежде верят в важность теории вероятностей: по существующим представлениям вероятностные модели статистической физики и явлений на квантово-механическом уровне вполне адекватно описывают реальную ситуацию. По крайней мере, *все* проведенные до настоящего времени эксперименты подтверждают эту гипотезу.

В дальнейшем изложении теории вероятностей мы оставим в стороне проблему статистической устойчивости и рассмотрим математическую модель, в которой отражены все возможные исходы испытания, причем связанные с этим испытанием вероятности будем считать известными.

В связи со сказанным, любопытно отметить, что составители задач по теории вероятностей частенько не задумываются об их содержании, и приводят задачи совершенно бессмысленные. Вот пример задачи такого рода, опубликованной в одном из сборников задач по теории вероятностей. «Охотник сидит в засаде и ждет медведя. Медведь может выскочить из-за первого куста с вероятностью 0,1, из-за второго куста – с вероятностью 0,2 и из-за третьего куста – с вероятностью 0,3. В первом случае охотник убивает его с вероятностью 0,5, во втором случае – с вероятностью 0,4, в третьем случае – с вероятностью 0,3. Найти вероятность того, что охотник убьет медведя».

В формулировке этой задачи весьма сомнительно предположение статистической устойчивости частот появления медведя из-за каждого куста, равно как и частот удачного выстрела охотника (без чего данные в задаче о вероятности не имеют смысла). Но что можно гарантировать, во всяком случае – это то, что эксперимент, давший приведенные значения вероятностей, никогда на самом деле не проводился, и вряд ли автор задачи представляет себе даже теоретически, как такой эксперимент можно поставить. В целом задача наводит на мысль о всеобъемлющем характере понятия вероятности, применимость которого не требует экспериментальной проверки (ведь каждому ясно, что эксперимента не было).[6] А поскольку изучающий теорию вероятностей знакомится с практическими примерами ее применения прежде всего из задач, то использование в преподавании задач, подобных приведенной выше, неизбежно создает у учащегося неоправданные иллюзии.

1.3. Алгебра событий.

1.3.1. В основу нашего рассмотрения мы положим деление случайных событий[7] на две группы: элементарные (или неразложимые) события и составные (или разложимые) события.

Под *элементарным событием* мы будем понимать *непосредственный, конкретный* исход испытания. Например, сказать, что сумма очков при бросании двух игральных костей равна шести, все равно, что сказать, что опыт привел к одному из исходов «(1.5), или (2.4), или (3.3), или (4.2), или (5.1)», и это перечисление разлагает событие «сумма очков равна шести» на пять элементарных событий.

Другой пример. Представим себе стрельбу по мишени. Элементарным событием будет здесь, например, конкретная пара вещественных чисел (x,y), представляющая собой координаты точки попадания. Составное событие *попадание в «десятку»* происходит в том и только в том случае, когда реализуется такое элементарное событие (x,y), при котором x и y представляют собой декартовы координаты какой-либо точки центрального круга мишени. Таких элементарных событий, очевидно, *бесконечно много*. С другой стороны, мы вполне можем под элементарным событием понимать номер круга мишени (обозначая номером «0» попадание в «молоко»). Тогда любое составное событие, например, «выбито не более n очков» будет представляться *конечным набором* элементарных событий. Мы видим, таким образом, что разделение событий на элементарные и составные весьма условно, и определяется теми задачами, которые ставятся при проведении опыта или наблюдения.

Отсюда следует, что если мы хотим говорить об «испытаниях» научно и без каких бы то ни было неясностей, то мы должны сначала условиться,

[6] Заметим, что задача, рассмотренная в примере 2 §6 (п. 6.4) и, казалось бы, схожая по содержанию, указанными недостатками не обладает.
[7] В дальнейшем слово «случайное» мы будем, как правило, опускать.

каковы же элементарные (неразложимые далее) события, представляющие собой мыслимые исходы опыта. Так, чтобы, *каждый неразложимый исход (идеализированного) опыта представлялся <u>одним и только одним</u> элементарным событием.*

Вообще дальше термином *«исход испытания» мы будем обозначать конкретный результат испытания, представляющий собой то или иное элементарное событие.*

Назовем ***пространством элементарных событий*** совокупность *всех мыслимых* элементарных событий, которые могут реализоваться в данном испытании. Соответственно ***событием*** будем называть любую совокупность (т.е. множество) элементарных событий.

Таким образом, *мы говорим, что* ***данное событие произошло****, если в результате испытания реализовалось одно из элементарных событий, его составляющих*. В этом случае мы также будем говорить, что *данный* ***исход*** *испытания* ***благоприятен*** *данному* ***событию***.

Для того, чтобы можно было развить методы *количественного описания* вероятностных закономерностей, нужно, очевидно, иметь возможность привлечь в качестве основного инструмента теоретического исследования *математические методы*. С другой стороны, чтобы эти методы можно было использовать эффективно, необходимо построить адекватную *математическую модель теории вероятностей*.

Все наше последующее изложение и будет посвящено описанию соответствующих математических моделей и методов теории вероятностей.

1.3.2. М***атематическая модель случайных событий***. ***Поле событий.*** Представленные выше рассуждения и понятия дают естественное основание использовать в качестве такой математической модели понятия и аппарат элементарной теории множеств. [8]

Прежде всего, дадим основные определения.

Определение 1.5. ***Пространством элементарных событий*** *естественно назвать произвольное абстрактное множество* Ω*, состоящее из элементов* ω*, называемых* ***элементарными событиями***.

Определение 1.6. ***Событием*** *будем называть любое подмножество пространства элементарных событий.*

Определение 1.7. ***Полем событий*** *будем называть совокупность* ***всех*** *событий, т.е. всех подмножеств множества* Ω, *включая само множество* Ω *и пустое множество* $\varnothing$.

В соответствии с этим определением, все множество Ω, одноточечное множество $\{\omega\}$ и пустое множество $\varnothing$ – тоже события.

Ниже мы увидим, что если пространство элементарных событий

[8] Мы предполагаем, что читатель знаком с основными теоретико-множественными обозначениями и операциями и их свойствами.

состоит из конечного числа n элементов, то поле событий состоит из 2^n элементов («элементами» *поля событий* являются сами события, т.е. *все* подмножества множества **Ω**, включая пустое множество Ø и само множество **Ω**).

Далее события мы будем, как правило, обозначать заглавными буквами латинского алфавита: *A*, *B*, *C*,… . Для обозначения элементарного события, представляющего собой элемент множества (события) *A*, мы будем, наряду с буквой ω, использовать соответствующую строчную букву: $a \in A, b_j \in B$ и т.п.

Очевидно, все пространство элементарных событий Ω представляет собой *достоверное событие*, а пустое множество Ø – *невозможное событие*. Для любого события *A*, очевидно, имеет место включение $A \subseteq \Omega$.

1.3.3. ***Действия со случайными событиями. Алгебра событий.*** Выясним, какой вероятностный смысл имеют при этом основные операции теории множеств.

1. *Включение множеств* $A \subset B$ в теории вероятностей соответствует утверждению, что *если происходит событие A, то обязательно происходит и событие B*.

(В теории вероятностей используется термин *событие A является частным случаем события B*).

Действительно, если происходит событие *A*, то, по определению, обязательно реализуется одно (какое-то, но только одно) элементарное событие ω из множества *A*. Но так как $A \subset B$, то ω также принадлежит и множеству B, а это и означает, опять же по определению, что событие B также произошло.

2. *Объединение множеств* $A \cup B$ (в теории вероятностей часто применяется также термин *сумма событий* А и В и используется обозначение *A*+*B*) – соответствует событию, состоящему в том, что в результате испытания *произошло или событие A, или событие B, или произошли они оба*.

3. *Пересечение множеств* $A \cap B$ (в теории вероятностей применяется также термин *произведение событий A* и *B* и используется обозначение *AB*) – в теории вероятностей соответствует событию, состоящему в том, что в результате одного и того же испытания *происходит как событие A, так и событие B*.

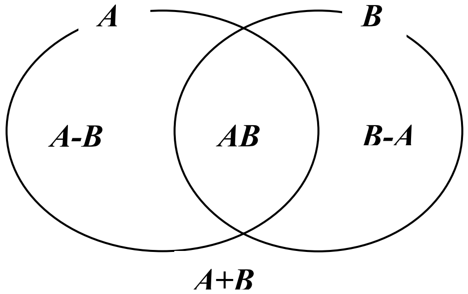
Действительно, если $\omega \in A \cap B$, то *одновременно* $\omega \in A$ и $\omega \in B$, то есть, по определению, происходит как событие *A*, так и событие *B*.

4. *Разность множеств* $A \setminus B$ (в теории вероятностей разность событий обозначается также $A - B$) – в теории вероятностей соответствует событию, состоящему в том, что в результате испытания *происходит событие A, но <u>не происходит</u> событие B*.

5. *Дополнение множества* $\overline{A} = \Omega \setminus A$, или $\overline{A} = \Omega - A$ – в теории

вероятностей соответствует событию, состоящему в том, что в результате испытания *событие* A *не происходит*. Иными словами, *событие* $\overline{A}$ *происходит тогда и только тогда, когда не происходит событие* A.

Событие $\overline{A}$ называется *событием,* ***противоположным*** *событию* A.



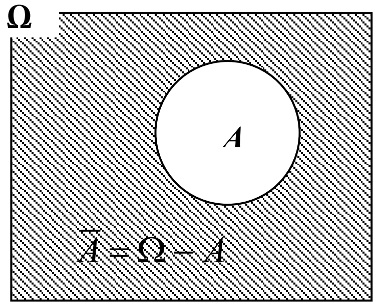


Рис. 1.1

Отметим еще следующие очевидные соотношения:

$A + \overline{A} = \Omega; \quad A\overline{A} = \varnothing; \quad \overline{\overline{A}} = A; \quad \overline{\Omega} = \varnothing; \quad \overline{\varnothing} = \Omega$.

События, для которых $AB=\varnothing$ называются ***несовместимыми*****,** ***или несовместными***. Заметим, что если $AB=\varnothing$, то $A \subset \overline{B}$ и $B \subset \overline{A}$.

В качестве **упражнения** рекомендуется доказать следующие равенства:

(1.1) $$\overline{AB} = \overline{A} + \overline{B}; \quad \overline{A + B} = \overline{A}\,\overline{B}.$$

Мы видим, что таким образом *все свойства теоретико-множественных операций элементарной теории множеств естественным образом переносятся на соответствующие операции с событиями.*[9] В частности,

(1.2) $$\begin{gathered} A + B = B + A; \quad AB = BA; \quad (A + B) + C = A + (B + C); \\ (AB)C = A(BC); \quad A(B + C) = AB + AC; \\ A + A = A; \quad AA = A; \end{gathered}$$

если $A \subset B$, то $A + B = B$ и $AB = A$, в частности, $A\Omega = A$.

Определение 1.8. *Поле событий вместе с введенными выше операциями (действиями) с событиями, называется* ***алгеброй событий***.

1.4. Основные свойства вероятности.

Попытаемся на основе представления о вероятности как пределе частоты события при неограниченном возрастании числа испытаний (п. 1.2), описать ее необходимые свойства. Это позволит нам далее при построении различных вероятностных моделей ограничивать свой выбор только теми

[9] Благодаря выбору подходящей математической модели!

моделями, в которых функция, определяющая вероятность, обладает этими свойствами.

Будем далее обозначать через A, B,… – события, обладающие свойством статистической устойчивости. Это, в частности, означает, что величина $p_n(A)=\dfrac{n_A}{n}$ (*статистическая частота* события A), где n – общее число испытаний, а n_A – число тех из них, в которых наблюдалось событие A, *стабилизируется* при неограниченном возрастании n независимо от того, о какой серии испытаний идет речь. Переводя это утверждение на математический язык, мы можем записать, что существует предел

$$(*)\qquad \lim_{n\to\infty} p_n(A)=P(A),$$

причем величина $P(A)$ не зависит от того, к какой конкретно серии испытаний относится $p_n(A)$. Эта величина $P(A)$ и называется *вероятностью события A*.[10]

Замечание. С формально-математической точки зрения, соотношение (*) означает следующее. Пусть у нас имеется последовательность серий испытаний, причем в i-й серии проводится n_i испытаний, и $n_i<n_{i+1}$, $i=1,2,\ldots$. Тогда для всякого $\varepsilon>0$ найдется такой номер i_ε, что

$$\left|p_{n_i}(A)-P(A)\right|<\varepsilon \text{ для всех } i>i_\varepsilon .$$

Выясним, на основании сформулированного определения, основные свойства функции $P(A)$. Прежде всего, очевидны следующие соотношения (ср. с замечанием после табл. 1.3):

1. $0\le P(A)\le 1$;
2. $P(\Omega)=1,\quad P(\varnothing)=0$.

Пусть теперь события A и B – несовместимы, т.е. $AB=\varnothing$. Пусть, далее, в серии n испытаний событие A наблюдалось n_A раз, а событие B – n_B раз. Рассмотрим событие $A + B$. Поскольку события A и B – несовместимы, то $n_{A+B}=n_A+n_B$. Поэтому

$$p_n(A+B)=\frac{n_{A+B}}{n}=\frac{n_A+n_B}{n}=\frac{n_A}{n}+\frac{n_B}{n}=p_n(A)+p_n(B).$$

Переходя в последнем равенстве к пределу при $n\to\infty$, получим, в силу (*), третье свойство вероятности:

3. Если $AB=\varnothing$, то $P(A+B)=P(A)+P(B)$.

[10] Величина $P(A)$ представляет собой, согласно общему определению функции, *функцию множества A*.

Следовательно, *вероятность необходимо удовлетворяет свойствам* 1 – 3.[11]

Теорема 1.1. Вероятность $P(A)$ обладает следующими свойствами:

1. $0 \le P(A) \le 1$
2. $P(\Omega) = 1, \quad P(\varnothing) = 0$
3. Если события A и B *несовместимы*, т.е. $AB = \varnothing$, то
$$P(A+B) = P(A) + P(B).$$

Замечание. Легко видеть, что если мы имеем более двух несовместимых между собой событий (*попарно несовместимые события*):

$$A_1, A_2, ..., A_n, \quad A_j A_k = \varnothing, \quad j \ne k, \quad j,k = 1,2,...,n,$$

то

$$P(A_1 + A_2 + ... + A_n) = P(A_1) + P(A_2) + ... + P(A_n). \tag{1.3}$$

Свойства вероятности 1 – 3 позволяют сформулировать следующие утверждения, *справедливые для всякой функции* $P(A)$, *удовлетворяющей этим свойствам*.

Теорема 1.2. $P(\overline{A}) = 1 - P(A)$.

Действительно, так как $\overline{A}A = \varnothing, \quad A + \overline{A} = \Omega, \quad P(\Omega) = 1$, то $P(A) + P(\overline{A}) = P(\Omega) = 1$, что и требовалось.

Теорема 1.3. Если $A \subset B$, то $P(A) \le P(B)$.

Действительно, так как $A \subset B$, то $B = A + (B - A)$ (рис. 1.2). Но, по определению разности двух событий, $A(B - A) = \varnothing$, откуда, в силу свойства 3 вероятности, $P(B) = P(A) + P(B - A)$. Но так как $P(B - A) \ge 0$ (свойство 1), получаем: $P(B) \ge P(A)$, что и требовалось.

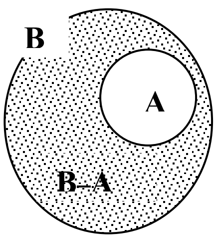


Рис. 1.2

***Лемма* 1.1.** Если A и B – два произвольных события, то
$$P(A - B) = P(A + B) - P(B);$$
$$P(B - A) = P(A + B) - P(A).$$

[11] Это, естественно не означает, что *всякая* модель вероятности, обладающая этими свойствами, годится для решения той или конкретной задачи. Эти условия являются *необходимыми*, ограничивая класс функций *P(A)*, из которых мы должны выбирать подходящую вероятностную модель для данной задачи.

Действительно, так как

$$A + B = (A - B) + B, \quad (A - B)B = \varnothing \text{, то}$$

$$P(A + B) = P(A - B) + P(B).$$

Меняя A и B местами, получим:

$$P(B + A) = P(B - A) + P(A)$$

Но так как $A+B=B+A$, из последних двух равенств и получаем утверждение леммы.

Теорема 1.4. Если A и B – два произвольных события, то

$$P(A + B) = P(A) + P(B) - P(AB).$$

Действительно, по *лемме*

(*) $$P((A + B) - AB) = P(A + B) - P(AB).$$

С другой стороны (см. рис. 1.1),

$$(A + B) - AB = (A - B) + (B - A),$$

причем

$$(A - B)(B - A) = \varnothing.$$

Поэтому,

(**) $$P((A + B) - AB) = P(A - B) + P(B - A).$$

Приравнивая правые части (*) и (**), получаем:

(+) $$P(A - B) + P(B - A) = P(A + B) - P(AB).$$

Но по *лемме*

$$P(A - B) = P(A + B) - P(B);$$

$$P(B - A) = P(A + B) - P(A)$$

Подставляя эти соотношения в левую часть (+), получим:

$$2P(A + B) - P(A) - P(B) = P(A + B) - P(AB),$$

что и доказывает теорему.

1.5. Классическая модель вероятности.

Наиболее просто изучить свойства вероятностей случайных событий, если принять предположение о *конечности пространства элементарных событий и равновозможности всех элементарных событий, составляющих это пространство.* Последнее, в частности, означает, как мы видели в п. 1.2, что *вероятности всех (элементарных) исходов опыта одинаковы.*

Ниже через $P(A)$ мы будем, как и ранее, обозначать вероятность события A.

Определение 1.9. ***Классической моделью вероятности*** *называется модель, в которой:*

1) Пространство элементарных событий Ω *состоит из конечного числа элементов:*

$\Omega = \{\omega_1, \omega_2, ..., \omega_n\}$, *$n$ – натуральное число;*

2) Вероятности всех элементарных событий одинаковы и равны

$$P(\omega_k) = \frac{1}{n}, \quad k = 1, 2, ..., n;$$

3) Вероятность события A, состоящего из m элементов, равна

$$P(A) = \frac{m}{n}, \quad A = \{\omega_{j_1}, \omega_{j_2}, ..., \omega_{j_m}\}.$$

Последний пункт формулируют еще следующим образом: *(классическая)* ***вероятность события*** *равна отношению числа исходов испытания, благоприятствующих этому событию, к общему числу возможных исходов испытания.*

Отсюда ясно, между прочим, что в классической модели *события, состоящие их одинакового числа элементарных исходов, равновозможны*.

К сказанному выше следует сделать одно важное *замечание*.

Из сказанного выше следует, что правильное описание пространства элементарных событий[12] является ключевым при решении любой вероятностной задачи. Например, при бросании двух и более игральных костей, какие элементарные события следует считать равновозможными? Ответ на этот вопрос не так очевиден, как кажется. Для одной игральной кости все просто: элементарным событием естественно считать выпадение конкретной грани, и предположение о равновозможности этих событий вполне оправданно (и подтверждается экспериментом). Иная ситуация возникает при эксперименте, состоящем в одновременном бросании нескольких (одинаковых!) игральных костей. Ключевым при попытке описать пространство элементарных событий в этом случае становится следующий вопрос: необходимо ли различать эти кости при вычислении количества элементарных исходов? Иными словами, важно ли, на какой конкретно кости выпадает то или иное число очков. Например, считаем ли мы, что при одновременном бросании трех (практически неразличимых) игральных костей, комбинации (1, 3, 4), (3, 1, 4), (1, 4, 3), (4, 1, 3), (4, 3, 1), (3, 4, 1) дают нам *шесть различных* элементарных событий, или же следует это считать *одним* элементарным исходом эксперимента. *Эксперимент показывает*, что при описании соответствующего пространства элементарных событий *кости следует различать*.

Подобный вопрос возник еще на заре зарождения теории вероятностей как науки (середина XVII века), и получил название *парадокса шевалье де Мере*. Суть его в следующем. Многократно наблюдая игру в кости, француз де Мере[13] подметил, что при одновременном бросании трех игральных костей более часто выпадает комбинация, дающая в сумме 11 очков, чем комбинация, дающая в сумме 12 очков, хотя, с его точки зрения, эти комбинации были равновозможны. Де Мере рассуждал следующим образом:

---

[12] С точки зрения адекватного соответствия этого описания реальному процессу.

[13] Вполне реальная историческая личность.

11 очков можно получить шестью различными способами (6+4+1, 6+3+2, 5+5+1, 5+4+2, 5+3+3, 4+4+3) и столькими же способами можно получить 12 очков (6+5+1, 6+4+2, 6+3+3, 5+5+2, 5+4+3, 4+4+4). Но тогда равенство числа исходов, в результате которых наступают соответствующие события A_1 и A_2, означает равенство их вероятностей $P(A_1)$ и $P(A_2)$.

Ошибка де Мере была указана знаменитым Блезом Паскалем (1623-1662). Она заключалась в том, что рассматриваемые де Мере исходы вовсе не являются равновероятными. *Нужно учитывать не только выпадающие очки, но и то, на каких именно костях они выпали.* Например, занумеровав кости и выписывая выпадающие очки в соответствующей последовательности, видно, что комбинация 6-4-1 выпадает, когда наступает один из шести исходов (6, 4, 1), (6, 1, 4), (4, 6, 1), (4, 1, 6), (1, 6, 4), (1, 4, 6), а комбинация 4-4-4 выпадает лишь при одном единственном исходе (4, 4, 4). Равновероятными в данном опыте являются исходы, описываемые тройками чисел (*a, b, c*), где *a* – число очков на первой кости, *b* – число очков на второй кости, *c* – число очков на третьей кости. Нетрудно подсчитать, что всего имеется 216 равновероятных исходов. Из них событию A_1 - «сумма выпавших очков равна 11» - благоприятствуют 27 исходов ($P(A_1) = 27/216 = 0{,}125$), а событию A_2 - «сумма выпавших очков равна 12» - благоприятствуют лишь 25 исходов ($P(A_2) = 25/216 \approx 0{,}116$). Это и объясняет подмеченную де Мере тенденцию к более частому выпадению 11 очков.

В качестве *упражнения* предлагаем читателю доказать, что классическая вероятность удовлетворяет свойствам 1 – 3 предыдущего пункта.

Замечание. Обозначим через $\boldsymbol{N}(A)$ – число элементов в произвольном конечном множестве A.

Тогда *для любых* множеств A и B

(1.4) $$N(A \cup B) = N(A) + N(B) - N(A \cap B).$$

Действительно, пусть, в рамках классической модели вероятности, A и B – произвольные события, а $\boldsymbol{\Omega}$ – соответствующее пространство элементарных событий. Тогда

$$P(A) = \frac{N(A)}{N(\Omega)},\ P(B) = \frac{N(B)}{N(\Omega)},\ P(AB) = \frac{N(AB)}{N(\Omega)},\ P(A+B) = \frac{N(A+B)}{N(\Omega)}.$$

С учетом этих соотношений, по теореме 1.4 получаем:

$$\frac{N(A+B)}{N(\Omega)} = \frac{N(A)}{N(\Omega)} + \frac{N(B)}{N(\Omega)} - \frac{N(AB)}{N(\Omega)},$$

что, очевидно, эквивалентно формуле (1.4).

1.6. Геометрическая вероятность.

Выше мы разобрали важный частный случай ***вероятностного пространства***,[14] а именно, классическую модель вероятности. Естественно, что эта простейшая модель не может «обслужить» все явления, процессы и т.п., где возможно разумное вероятностное описание, т.е. в которых предположение о статистической устойчивости событий может быть принято в качестве рабочей гипотезы. В этом пункте мы рассмотрим еще одну простую модель вероятностного пространства, в некотором смысле аналогичную классической модели, в основе которой лежит пространство элементарных событий, состоящее из бесконечного числа элементов.

Постановка задачи. Пусть, например, на плоскости имеется некоторая область G и в ней содержится некоторая другая область $g \subset G$ (см. рис. 1.3). В область G *наудачу* бросается точка и спрашивается, *чему равна вероятность того, что точка попадет в область* g. Определим точный смысл этого вопроса.

Определение 1.10. *Мы скажем, что в область* G ***наудачу*** *бросается точка, если:*

1) *брошенная точка обязательно попадает в область* G, *и может попасть в любую точку этой области*;

2) *вероятность попасть в какую-то часть* g *области* G *пропорциональна мере этой части (длине, площади и т.д.) и не зависит от ее расположения и формы.*

Иными словами, *по определению*, вероятность попадания в область

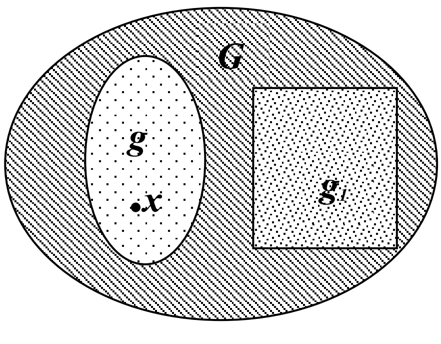


Рис. 1.3

$g \subset G$ при бросании наудачу точки x в область G равна

$$P(x \in g) = \frac{\text{мера } g}{\text{мера } G}.$$

В частности, если область $g_1 \subset G$ такова, что $\text{мера } g = \text{мера } g_1$, то

$$P(x \in g) = P(x \in g_1).$$

Описанное выше понятие вероятности носит название *геометрической вероятности*.

Заметим, что условие независимости соответствующей вероятности от формы и расположения области g, является естественным аналогом

[14] Под **вероятностным пространством** здесь и далее понимается совокупность алгебры событий и заданной на ней вероятности.

предположения о равновозможности (и равновероятности) исходов в классической модели. Кроме того, свойство *аддитивности* меры множества (длины, площади, объема)[15] позволяют легко установить основные свойства геометрической вероятности, которые, по существу, тождественны со свойствами, описанными в теореме 1.1.

Рассмотрим несколько **примеров**.

Пример 1 (задача о встрече). Два лица A и B условились встретиться в определенном месте между, например, 11 и 12 часами. Прибывший на встречу, в случае, если другой отсутствует, ждет его в течение 20 минут (но не дольше, чем до конца часа), после чего уходит. Чему равна вероятность встречи лиц A и B, если приход каждого из них в течение указанного часа может произойти наудачу, и моменты прихода независимы.[16]

Решение. Обозначим момент прихода лица A через x, а момент прихода лица B через y. Для того чтобы встреча произошла, необходимо и достаточно, чтобы

$$|x-y| \le 20.$$

Изобразим x и y как декартовы координаты на плоскости; в качестве единицы масштаба выберем минуту. Всевозможные исходы (пространство элементарных событий) изобразятся тогда точками квадрата со сторонами $0 \le x, y \le 60$; благоприятствующие встрече исходы расположатся в заштрихованной области (рис. 1.4), площадь которой равна $60^2 - 40^2 = 2000$, а искомая вероятность, следовательно, будет равна

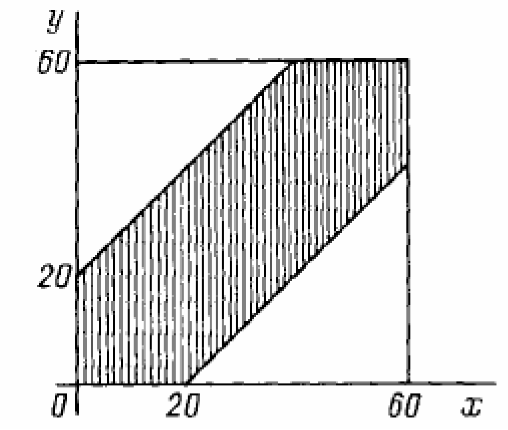


Рис. 1.4

$$p = \frac{2000}{60^2} = \frac{5}{9}.$$

Пример 2. Рассмотрим круг радиуса R, разделенный на n частей концентрическими окружностями радиусов $0 < r_1 < r_2 < ... < r_n = R$ (рис. 1.5).[17] На этот круг игроком *наудачу* бросается точка, причем, если она попадает в кольцо, образуемое окружностями радиусов

$$r_{k-1}, r_k, \quad k = 1, 2, ..., n; \quad r_0 = 0, r_n = R,$$

то игроку начисляется m_k очков.

Вопрос: каково должно быть *справедливое* распределение числа очков $m_k, k = 1, 2, ..., n$ по кольцам «мишени», если заданы радиусы r_k.

Решение. Под *справедливым* распределением очков по кольцам «мишени» естественно понимать требование выполнения соотношения:

[15] Т.е., например, площадь объединения непересекающихся плоских множеств равна сумме их площадей.
[16] Эта задача (в иной формулировке) была впервые опубликована Уайтвортом в 1886 году.
[17] Сравните с обычной «мишенью» в тире.

$$\frac{m_k}{m_{k+1}} = \frac{p_{k+1}}{p_k}$$

(*т.е., чем меньше вероятность попадания, тем большее число очков должно начисляться*).

Ключевым словом в формулировке нашей задачи является слово *наудачу*. Это означает, что *можно использовать модель геометрической вероятности*. Отсюда следует, что для оценки вероятности p_k попадания точки в k-е кольцо, достаточно знать площадь этого кольца.

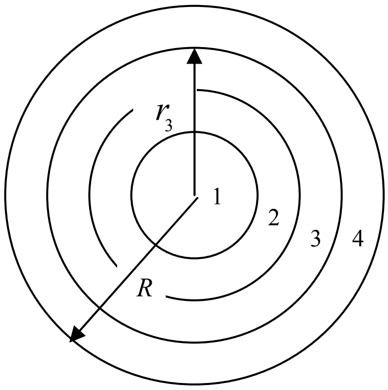


Рис. 1.5

Таким образом, должно выполняться соотношение (площадь k-го кольца равна $\pi(r_k^2 - r_{k-1}^2)$):

$$\frac{m_k}{m_{k+1}} = \frac{r_{k+1}^2 - r_k^2}{r_k^2 - r_{k-1}^2} \equiv \lambda_k, \quad k = 1, 2, ..., n-1,$$

откуда получаем: $m_{k+1} = \dfrac{1}{\lambda_k} m_k$.

В качестве *численной иллюстрации*, рассчитаем, например, вероятности попадания точки в соответствующие кольцевые области мишени[18] при следующих данных:

$$n = 4, \quad R = 4, \quad r_k - r_{k-1} = 1, \quad k = 1, 2, 3, 4; \quad r_0 = 0, \quad r_4 = 4,$$

т.е. кольца мишени имеют одинаковую ширину, равную 1. Обозначим через S_k *площади* соответствующих кольцевых областей мишени, а через A_k – событие, состоящее в попадании точки в такую область. Тогда имеем:

$$S_1 = \pi, \quad S_2 = \pi(2^2 - 1^2) = 3\pi,$$

$$S_3 = \pi(3^2 - 2^2) = 5\pi, \quad S_4 = \pi(4^2 - 3^2) = 7\pi.$$

[18] Центральный круг мы тоже считаем «кольцом» с внутренним радиусом, равным нулю.

Площадь всей мишени S равна, очевидно, 16π. Отсюда получаем (так как $P(A_k) = {S_k}/{S}$):

$$p_1 = P(A_1) = 1/16; \quad p_2 = P(A_2) = 3/16;$$
$$p_3 = P(A_3) = 5/16; \quad p_4 = P(A_4) = 7/16.$$

В частности, отсюда следует, что в нашем случае разница между очками, присуждаемыми за попадание в соответствующую зону мишени, должна составлять 2 единицы (7 очков для центральной (первой) зоны, 5 – для второй зоны и т.д.).[19]

Замечание. Последний расчет дает нам *пример модели вероятностного пространства, не являющейся классической*, ибо при конечном пространстве элементарных событий $\Omega = \{A_1, A_2, A_3, A_4\}$ (элементарный исход эксперимента – попадание в зону 1, 2, 3 или 4 соответственно), элементарные исходы *не равновероятны*!

[19] Точнее, числа очков должны относиться как 7:5:3:1 (при движении от центра к периферии мишени).

§ 2. УСЛОВНАЯ ВЕРОЯТНОСТЬ. НЕЗАВИСИМОСТЬ. ФОРМУЛА БАЙЕСА

2.1. Условная вероятность.

Оставаясь пока в рамках классической модели вероятности, рассмотрим следующий вопрос.

Пусть у нас имеется некоторая *алгебра событий* [20] $\mathfrak{M} = \{A, B, C, \ldots\}$ с заданной на ней вероятностью $P(A)$, $A \in \mathfrak{M}$.

Часто возникает вопрос, *как должны измениться вероятности P(A),* $A \in \mathfrak{M}$, *если предположить, что станет достоверным наступление одного из событий*, скажем B.

В связи с этим рассмотрим следующий **пример**.

Будем бросать две игральные кости одну за другой. Пространство элементарных событий будет здесь образовано из 36 событий, состоящих в выпадении определенного числа 1, 2, 3, 4, 5 или 6 очков на первой кости и определенного числа 1, 2, 3, 4, 5 или 6 очков на второй кости. Причем те и другие могут сочетаться друг с другом как угодно (см. табл. 2.1).[21]

Таблица 2.1

1,1	1,2	1,3	1,4	1,5	1,6
2,1	2,2	2,3	2,4	2,5	2,6
3,1	3,2	3,3	3,4	3,5	3,6
4,1	4,2	4,3	4,4	4,5	4,6
5,1	5,2	5,3	5,4	5,5	5,6
6,1	6,2	6,3	6,4	6,5	6,6

Первое число в каждой клетке этой таблицы есть число очков на первой кости, второе – на второй кости. Среди многих допустимых здесь событий отметим два. Во-первых, событие, которое мы обозначим через B, состоящее в том, что *на первой кости выпадет одно очко*. Этому благоприятствует вся первая строка таблицы, т.е. шесть клеток (см. таблицу 2.1а), так что

$$P(B) = \frac{6}{36} = \frac{1}{6} .$$

[20] См. определение 1.8 в конце пункта 1.3.
[21] Иначе, речь идет о множестве *упорядоченных пар* (m,n) натуральных чисел от 1 до 6.

Таблица 2.1а

1,1	1,2	1,3	1,4	1,5	1,6
2,1	2,2	2,3	2,4	2,5	2,6
3,1	3,2	3,3	3,4	3,5	3,6
4,1	4,2	4,3	4,4	4,5	4,6
5,1	5,2	5,3	5,4	5,5	5,6
6,1	6,2	6,3	6,4	6,5	6,6

Во-вторых, отметим событие (событие A), состоящее в том, что на обеих костях выпадет число очков, *в сумме не превосходящее* 5 (таблица 2.1б).

Таблица 2.1б

1,1	1,2	1,3	1,4	1,5	1,6
2,1	2,2	2,3	2,4	2,5	2,6
3,1	3,2	3,3	3,4	3,5	3,6
4,1	4,2	4,3	4,4	4,5	4,6
5,1	5,2	5,3	5,4	5,5	5,6
6,1	6,2	6,3	6,4	6,5	6,6

Этому благоприятствуют 10 клеток таблицы, так что

$$P(A) = \frac{10}{36} = \frac{5}{18} \ .$$

Поставим теперь вопрос следующим образом.

Представим себе, что мы бросили первую кость и получили на ней одно очко. Как изменится после этого вероятность того, что после бросания второй кости на обеих костях выпадет число очков, в сумме не превосходящее 5.

Иначе говоря, *пусть стало достоверным событие В. Как изменится после этого вероятность* $P(A)$ *события* A?

Эта, последняя, вероятность называется ***условной вероятностью события А при условии, что событие В произошло*** *и обозначается* $P(A/B)$ (читается: *вероятность А при условии В*).

Найдем эту вероятность. Предположение о том, что стало достоверным событие B, *изменяет пространство элементарных событий (т.е. достоверное событие)*, а именно *им становится само событие* B! Поэтому из всех элементарных исходов становятся возможными только те, которые составляют событие B, т.е. *первая строка* нашей таблицы (таблица 2.1а). Таких исходов, как мы знаем, шесть. Но в *первой строке* таблицы число тех клеток, которые благоприятны событию A (т.е. сумма очков на обеих костях не превосходит 5), как легко видеть, всего 4 (табл. 2.1в).

Таблица 2.1в

1,1	1,2	1,3	1,4	1,5	1,6
2,1	2,2	2,3	2,4	2,5	2,6
3,1	3,2	3,3	3,4	3,5	3,6
4,1	4,2	4,3	4,4	4,5	4,6
5,1	5,2	5,3	5,4	5,5	5,6
6,1	6,2	6,3	6,4	6,5	6,6

Таким образом,

$$P(A/B) = \frac{4}{6} = \frac{2}{3}.$$

Применим теперь аналогичное рассуждение в общем случае.

Предположение, что стало достоверным событие B, изменяет исходное пространство элементарных событий: именно, прежнее пространство элементарных событий заменяется событием B, точнее, совокупностью элементарных событий, составляющих событие B. Отсюда следует, что из общего числа первоначально благоприятных для события A исходов остаются лишь те, которые одновременно благоприятны также и событию B, то есть исходы, благоприятные произведению событий AB.

Таким образом, если исходное пространство элементарных событий Ω состоит из n элементов (элементарных событий), событие B состоит из m элементов и, наконец, событие AB состоит из k элементов, то

$$P(A/B) = \frac{k}{m} = \frac{kn}{mn} = \frac{k/n}{m/n}.$$

Но по определению

$$\frac{k}{n} = P(AB); \quad \frac{m}{n} = P(B).$$

Отсюда получаем ***основную формулу для условной вероятности***:

(2.1) $$P(A/B) = \frac{P(AB)}{P(B)}, \text{ или}$$

(2.1а) $$P(AB) = P(B)P(A/B).$$

Меняя события A и B местами и проводя аналогичные рассуждения, с учетом равенства $AB = BA$, т.е. $P(AB) = P(BA)$, получаем:

(2.2) $$P(B/A)=\frac{P(AB)}{P(A)},\ \text{или}$$

(2.2a) $$P(AB)=P(A)P(B/A).$$

Замечание. Формулы (2.1а) или (2.2а) иногда еще называют **теоремой умножения вероятностей**.

Из формул (2.1) и (2.2) следует, в частности, что

(2.3) $$P(A/B)P(B)=P(B/A)P(A).$$

В заключение данного пункта рассмотрим следующий пример.

Пример. Предположим, что в урне находится $n_ч$ – черных, $n_б$ – белых, $n_к$ – красных и $n_с$ – синих шаров. Наугад из урны извлекается один шар. Найти вероятность того, что этот шар *красный*, если стало известно, что он *не синий*.

Решение. Пусть $A_ч$, $A_б$, $A_к$ и $A_с$ – события, состоящие в том, что извлечены черный, белый, красный и синий шары соответственно. Из условия следует, что нам надо найти условную вероятность $P(A_к/\overline{A}_с)$.[22]

По формуле (2.1):

$$P(A_к/\overline{A}_с)=\frac{P(A_к\overline{A}_с)}{P(\overline{A}_с)}.$$

Но, очевидно, $A_к\overline{A}_с=A_к$ и, следовательно, $P(A_к\overline{A}_с)=P(A_к)$. Кроме того, как мы знаем, $P(\overline{A}_с)=1-P(A_с)$. Следовательно,

$$P(A_к/\overline{A}_с)=\frac{P(A_к)}{P(\overline{A}_с)}=\frac{P(A_к)}{1-P(A_с)}.$$

Подставляя в эту формулу значения вероятностей:

$$P(A_к)=\frac{n_к}{n};\quad P(A_с)=\frac{n_с}{n};\quad n=n_ч+n_б+n_к+n_с,$$

получим окончательно:

[22] Здесь событие $\overline{A}_c$ означает по определению «извлечен *не синий* шар».

$$P(A_к / \overline{A}_с) = \frac{n_к}{n - n_с} = \frac{n_к}{n_ч + n_б + n_к} \ .$$

2.2. Независимые события.

Понятие независимости событий – одно из важнейших в теории вероятностей. Большинство практически важных выводов теории вероятностей относятся именно к ситуациям, когда рассматриваемые события можно считать независимыми. Понятие независимости, как и многие понятия теории вероятностей, имеет как формально-количественную, так и качественную, интуитивную стороны.

Имея в виду изложенное в предыдущем пункте, е*стественно назвать события A и B* ***независимыми***, *если вероятность одного события не зависит от того, произошло другое событие или нет*.

Определение 2.1. Будем называть события *A и B* ***независимыми***, если

(2.4) $$P(A/B) = P(A), \quad P(B/A) = P(B).$$

Разберем в связи с этим такой **пример**.

Рассмотрим испытание, состоящее в однократном бросании одной игральной кости. Обозначим через *A* событие, состоящее в выпадении *числа очков, не превосходящего двух*, а через *B* – событие, состоящее в выпадении *четного числа очков*. Выпадению числа очков, не превосходящего двух (событие *A*) благоприятствуют *два* возможных исхода: выпадение 1 или 2 очков. Выпадению же четного числа очков, не превосходящего двух (событие *AB*) благоприятствует *один* исход – выпадение 2 очков. Поэтому, $P(B/A) = 1/2$.

С другой стороны, выпадению четного числа очков (событие *B*) благоприятствуют *три* исхода (2, 4 и 6 очков) из возможных *шести*. Поэтому, $P(B) = 3/6 = 1/2$, то есть $P(B/A) = P(B)$, и события *A и B независимы*.

Однако, если бы мы обозначили через *A* событие, состоящее в выпадении числа очков, *не превосходящего трех*, а через *B* – по-прежнему событие, состоящее в выпадении *четного числа очков*, то мы бы имели $P(B/A) = 1/3$, а $P(B) = 1/2$. В этом случае, следовательно, события *A и B* ***не*** *независимы*.

С учетом формул (2.1), (2.2) и (2.4), можно дать и такое формальное определение независимости событий.

Определение 2.1а. *События A и B называются* ***независимыми****, если выполнено равенство*

(2.5) $$P(AB) = P(A)P(B).$$

Замечание. Если события A и B независимы, то противоположные события $\overline{A}$ и $\overline{B}$ тоже независимы. Действительно, имеем (см. формулы (1.1), (2.5) и теоремы 1.2 и 1.4):

$$P(\overline{A}\overline{B}) = P(\overline{A+B}) = 1 - P(A+B) = 1 - (P(A) + P(B) - P(AB)) =$$
$$= 1 - P(A) - P(B) + P(A)P(B) = (1 - P(A))(1 - P(B)) = P(\overline{A})P(\overline{B}).$$

Определение 2.1б. *События* $A_1, A_2, ..., A_n$ *называются* ***независимыми (в совокупности)****, если для любых* k *из них* $(k \le n)$*:*

$$A_{j_1}, A_{j_2}, ..., A_{j_k}, \quad 1 \le j_1, j_2, ..., j_k \le n$$

выполнено равенство

(2.5а) $$P(A_{j_1} A_{j_2} ... A_{j_k}) = P(A_{j_1}) P(A_{j_2}) ... P(A_{j_k}).$$

Если это соотношение выполняется <u>только</u> <u>при</u> $k = 2$*, то события* $A_1, A_2, ..., A_n$ *называются* ***попарно независимыми***.

Замечания.

1) Для независимых в совокупности событий $A_1, A_2, ..., A_n$ события $\overline{A}_1, \overline{A}_2, ..., \overline{A}_n$ также независимы в совокупности. Обосновать это утверждение предлагается читателю самостоятельно.

2) Попарная независимость событий *не означает* их независимости в совокупности. Вот соответствующий **пример**. Пусть нам дана совокупность предметов, из которых одна четверть обладает тремя признаками X, Y, Z, одна четверть – только признаком X, одна четверть – только признаком Y и одна четверть – только признаком Z. Обозначим через *A, B,* и *C* соответственно события, состоящие в том, что случайно выбранный предмет окажется обладающим признаком X, Y или Z. Для вероятностей каждого из этих событий имеем: $P(A) = P(B) = P(C) = \frac{2}{4} = \frac{1}{2}$. Далее, легко видеть, что $P(AB) = P(AC) = P(BC) = \frac{1}{4}$ (поскольку двумя признаками обладают лишь предметы первой группы, а таких предметов четверть от общего количества). Отсюда видно, что события *A, B,* и *C попарно независимы*. Далее, $P(ABC) = \frac{1}{4}$,

однако $P(A)P(B)P(C)=\frac{1}{2}\cdot\frac{1}{2}\cdot\frac{1}{2}=\frac{1}{8}$. Так что события *A, B,* и *C не являются независимыми в совокупности*.

Отметим, что именно определение 2.1а, т.е. формула (2.5) (а не формулы (2.4)) *кладется в основу понятия независимости событий*. Причин такому подходу несколько.

Во-первых, бывает проще вычислить величины, входящие в формулу (2.5), нежели вычислять непосредственно условную вероятность, чтобы проверить равенства (2.4). *Во-вторых*, и это важнее, предположение о независимости событий на практике очень часто принимается как *исходное*, т.е. на основании тех или иных *априорных* предположений о характере изучаемых процессов и явлений. Типичным примером может быть априорное предположение о независимости влияния различных случайных факторов, вызывающих ошибки измерений. В ряде случаев, независимость тех или иных событий разумно принять как естественный постулат, просто по реальному смыслу поставленной задачи. Таковыми, например, являются предположения о независимости исходов двух (или более) последовательных бросаний игральной кости или монеты.

Таким образом, очень часто можно встретить утверждение такого типа: *так как события A и B независимы, то* $P(AB)=P(A)P(B)$, или: *предположим, что события A и B независимы, следовательно,* $P(AB)=P(A)P(B)$ и т.п.

Сказанное, однако, не означает, что вопрос о выяснении зависимости (независимости) событий не имеет серьезного практического значения. Наоборот, во многих случаях именно задача выяснения факта и степени зависимости тех или иных событий является основной задачей исследования. Соответствующие методы представляют собой обширный раздел *математической статистики* (так называемый *корреляционный анализ*).

Замечание. Не следует путать понятия *независимости* и *несовместимости* событий – это понятия различные по существу описываемых ими свойств. С другой стороны, между этими понятиями прослеживается определенная связь. Именно, *независимые события всегда совместимы, а несовместимые события не могут быть независимыми.*

Читателю рекомендуется, в качестве **упражнения**, строго (т.е. на основании формальных определений) обосновать эти утверждения.

2.3. Формула полной вероятности.

На практике часто встречаются задачи следующего типа.

Предположим, что на склад завезли детали с трех заводов, причем известно, что N_1 *деталей завезено с первого завода*, N_2 *деталей завезено со второго завода и* N_3 *деталей – с третьего. Известно также, что доля брака* *при изготовлении этих деталей составляет: на первом заводе –* n_1*%, на втором заводе –* n_2*% и на третьем заводе –* n_3*%. При входном контроле произвольно выбрана одна деталь. Какова вероятность того, что эта деталь будет бракованной?*

Задачи такого рода легко решаются с помощью так называемой *формулы полной вероятности*. Выведем эту формулу. Дадим сначала следующее важное для дальнейшего

Определение 2.2. *Совокупность событий* $H_1, H_2, ..., H_n$ *мы назовем* ***полной группой событий,*** *если они* ***попарно несовместимы****, т.е.*

$$H_j H_k = \varnothing, \quad j \neq k, \quad j,k = 1,2,...,n \ ,$$

и в совокупности образуют все пространство элементарных событий:

$$H_1 + H_2 + ... + H_n = \Omega .$$

Вот несколько простых **примеров.**

1. Совокупность всех возможных исходов эксперимента (элементарных событий) представляет собой полную группу событий по определению.

2. Полную группу событий образует пара событий A и $\overline{A}$ (т.е. событие A и событие, ему противоположное).

3. При бросании двух игральных костей полную группу событий образуют события, представляемые строками (соответственно столбцами) таблицы 2.1.

Теорема 2.1 (формула полной вероятности). Если A –произвольное событие, а $H_1, H_2, ..., H_n$ – полная группа событий, то справедлива следующая *формула полной вероятности*:

$$P(A) = P(A/H_1)P(H_1) + P(A/H_2)P(H_2) + ... + P(A/H_n)P(H_n), \tag{2.6}$$

или короче

$$P(A) = \sum_{k=1}^{n} P(A/H_k)P(H_k) . \tag{2.6*}$$

Действительно, так как $H_1, H_2, ..., H_n$ образуют, по условию, полную группу событий, то есть

$$H_1 + H_2 + ... + H_n = \Omega \text{ и } H_j H_k = \varnothing, \quad j \neq k, \quad j,k = 1,2,...,n ,$$

то (см. формулы (1.2)):

$$A = A\Omega = A(H_1 + H_2 + ... + H_n) = AH_1 + AH_2 + ... + AH_n ,$$

причем

$$(AH_k)(AH_j) = A(H_k H_j) = \varnothing \text{ при } k \neq j, \quad k, j = 1,2,...,n.$$

Поэтому по формуле (1.3):

(*) $\quad P(A) = P(AH_1) + P(AH_2) + ... + P(AH_n)$.

С другой стороны, по (2.1) $P(AH_k) = P(A/H_k)P(H_k)$. Подставляя это соотношение в формулу (*), мы получим (2.6), что и требовалось.

Теперь мы легко можем решить задачу, сформулированную в начале этого пункта. Заметим, прежде всего, что формулировка задачи позволяет нам принять в качестве гипотезы классическую модель вероятности.

Далее, пусть A означает событие, состоящее в том, что выбранная при входном контроле деталь оказалась *бракованной*. Полную группу событий сформируем следующим образом. Пусть будет H_k – событие, состоящее в том, что выбранная для проверки деталь произведена k-м заводом ($k = 1,2,3$). Очевидно, события H_1, H_2, H_3 образуют полную группу событий. Поэтому, по формуле полной вероятности,

(+) $\quad P(A) = P(A/H_1)P(H_1) + P(A/H_2)P(H_2) + P(A/H_3)P(H_3)$.

В силу предположения о справедливости для данной задачи классической модели вероятности, мы можем, очевидно, написать:

$$P(H_1) = \frac{N_1}{N}, \quad P(H_2) = \frac{N_2}{N}, \quad P(H_3) = \frac{N_3}{N}, \quad N = N_1 + N_2 + N_3;$$

$$P(A/H_1) = \frac{n_1}{100}, \quad P(A/H_2) = \frac{n_2}{100}, \quad P(A/H_3) = \frac{n_3}{100}.$$

Подставляя все эти выражения в формулу (+), получим окончательно:

$$P(A) = \frac{n_1 N_1 + n_2 N_2 + n_3 N_3}{100(N_1 + N_2 + N_3)}.$$

Вероятность обнаружить *полноценную* деталь будет соответственно равна

$$P(\overline{A}) = 1 - P(A) = 1 - \frac{n_1 N_1 + n_2 N_2 + n_3 N_3}{100(N_1 + N_2 + N_3)}.$$

Замечание. Из разобранной задачи легко усмотреть, что практическая ценность формулы полной вероятности возникает в том случае, когда, с одной стороны, удается выделить подходящую полную группу событий, а с другой – иметь возможность сравнительно просто найти соответствующие условные вероятности.

2.4. Формула Байеса.

Из формулы полной вероятности и формулы (2.3) легко получить простое, но важное следствие, называемое *формулой Байеса*.

Пусть, как и выше, события $H_1, H_2, ..., H_n$ образуют полную группу. Пусть далее A – некоторое событие. По формуле (2.3) мы можем написать:

$$P(H_j / A)P(A) = P(A / H_j)P(H_j), \quad j = 1, 2, ..., n.$$

Отсюда, с учетом формулы полной вероятности (2.6), получаем **формулу Байеса**:

(2.7) $$P(H_j / A) = \frac{P(A / H_j)P(H_j)}{P(A / H_1)P(H_1) + P(A / H_2)P(H_2) + ... + P(A / H_n)P(H_n)},$$

или короче:

(2.7)* $$P(H_j / A) = \frac{P(A / H_j)P(H_j)}{\sum\limits_{k=1}^{n} P(A / H_k)P(H_k)}, \quad j = 1, 2, ..., n.$$

Практическая важность формулы Байеса иллюстрируется еще такой ее интерпретацией. Предположим, что мы изучаем некоторое явление, причем у нас имеется некоторая совокупность взаимоисключающих друг друга гипотез, связанных с этим явлением. Пусть, кроме того, нам известно, что некоторое событие A, также связанное с этим явлением, может наблюдаться с той или иной априорной вероятностью, в зависимости от того, какую из гипотез принять как истинную. Предположим, наконец, что, также из некоторых априорных соображений, у нас есть представление о вероятностях реализации в действительности каждой из этих гипотез.

Задача ставится следующим образом. *В этих условиях определить вероятность реализации (т.е. справедливости) той или иной гипотезы, если в результате эксперимента наблюдалось событие A.*

Нетрудно видеть, что в такой постановке задача решается формулой Байеса *непосредственно*. Действительно, достаточно интерпретировать гипотезы как события $H_1, H_2, ..., H_n$ (что можно делать, так как относительно гипотез нас интересует только вопрос, реализуются они или нет)[23] и предположить, что они образуют в совокупности полную группу событий, т.е. что эти гипотезы взаимно исключают друг друга, но одна из них (заранее неизвестно какая) реализуется обязательно.

Опять же, как и в случае формулы полной вероятности, практическая ценность формулы Байеса усматривается лишь в тех случаях, когда вероятности, фигурирующие в правой части (2.7) могут быть сравнительно легко определены (либо заданы априори).

В качестве первого **примера** снова рассмотрим, в несколько измененной постановке, задачу, сформулированную в начале предыдущего пункта.

Именно, *предположим, что на склад завезли детали с трех заводов, причем известно, что* N_1 *деталей завезено с первого завода*, N_2 *деталей завезено со второго завода и* N_3 *деталей – с третьего. Известно также, что доля брака при изготовлении этих деталей составляет: на первом заводе –* n_1%, *на втором заводе –* n_2% *и на третьем заводе –* n_3%. *При входном контроле наудачу выбрана одна деталь. Предположим, что выбранная деталь оказалась бракованной. Какова, в этом случае, вероятность того, что деталь изготовлена, скажем, на втором заводе?*

Решение. При сохранении обозначений предыдущего пункта, из формулы Байеса непосредственно получаем:

$$P(H_2/A)=\frac{P(A/H_2)P(H_2)}{P(A/H_1)P(H_1)+P(A/H_2)P(H_2)+P(A/H_3)P(H_3)}=$$
$$=\frac{n_2N_2}{n_1N_1+n_2N_2+n_3N_3}.$$

Что и требовалось.

В заключение этого пункта рассмотрим еще один показательный **пример**. Речь идет о медицинской диагностике. Пусть в некоторую клинику обращаются больные, у которых может быть одна из болезней $H_1, H_2, ..., H_n$.

Обозначим через A комплекс симптомов для данного больного. В таком случае априорные вероятности $P(H_j)$ и условные вероятности $P(A/H_j)$ могут быть экспериментально найдены на основании собранной за прошлые годы статистики. При этом $P(H_j)$ равна примерно частоте болезни H_j среди больных данной клиники, а $P(A/H_j)$ – частота наблюдения комплекса симптомов A у больных с болезнью H_j в данной клинике. Поскольку речь идет о статистике за прошлые годы, то можно считать

[23] См. определения в пункте 1.1.

имеющиеся статистические данные почти достоверными. Применение формулы Байеса для нахождения вероятностей $P(H_j / A)$ – вероятности наличия у больного болезни H_j при условии, что у него наблюдается комплекс симптомов A, – здесь не вызывает сомнений.

§ 3. ОБОБЩЕНИЕ: ДИСКРЕТНАЯ МОДЕЛЬ ВЕРОЯТНОСТИ

3.1. Некоторые общие замечания.

В предыдущих параграфах мы видели, что для получения содержательных результатов при описании случайных явлений необходимо правильно подобрать математическую модель соответствующего ***вероятностного пространства*** (т.е. алгебры событий и определенной на ней вероятности). Выше мы рассмотрели две таких модели: классическую модель вероятности и геометрическую вероятность.

Однако, несмотря на большое количество практически интересных и важных задач, которые могут быть решены с их помощью, этими задачами далеко не исчерпываются потребности, как теории, так и практики. В частности, большинство физических задач, связанных со случайными явлениями, не подпадают под те требования, которые позволяют применить рассмотренные нами простейшие модели. Например, возникает естественный вопрос, *как* определить вероятность на алгебре событий, если исходное пространство элементарных событий бесконечно (скажем, счетное множество) и, следовательно, понятие вероятности отдельного исхода в классическом смысле явно непригодно.[24] Аналогичный вопрос возникает и тогда, когда (даже в случае конечности пространства элементарных событий) предположение о равновероятности исходов противоречит физическому смыслу задачи. Таковыми задачами являются, например, задачи *прицельной* стрельбы, задачи оценки надежности приборов, систем управления и т.п.

В этом параграфе будет описана модель вероятностного пространства, более общая, чем классическая модель. Классическая модель, как мы увидим, окажется ее частным случаем.

Но прежде сделаем важное

З*амечание*. Если проанализировать доказательства большинства теорем, приведенных в предыдущих параграфах, то мы увидим, что эти доказательства опираются не на тот факт, что в основе изложения лежит классическая модель вероятности, а на основные свойства вероятности, сформулированные в пункте 1.4. Иными словами, если этими свойствами обладает вероятность в любой другой модели вероятностного пространства, то практически все сформулированные выше теоремы остаются в силе. Более того, если формулы (2.1) или (2.2) принять как *определение условной*

[24] Действительно, в классическом понимании вероятность любого события, состоящего из *конечного* числа элементов, в этом случае будет равна нулю, что представляется бессмысленным.

вероятности, то остаются в силе и формула полной вероятности, и формула Байеса. Варианты определения независимости событий (определения 2.1, 2.1а и 2.1б) никаких изменений при этом не претерпевают, так как их формулировки инвариантны относительно выбора модели вероятностного пространства. Эта инвариантность, как легко усмотреть, имеет место и в определениях большинства остальных понятий предыдущих параграфов.

3.2. Дискретное вероятностное пространство.

Начнем с основного определения.

Пусть $\mathbf{\Omega}$ – конечное или счетное множество элементов ω_j:

$$\Omega = \{\omega_1, \omega_2, ..., \omega_n, ...\},$$

которое мы будем далее отождествлять с пространством элементарных событий, а сами ω_j – с элементарными событиями (исходами). Пусть $\mathfrak{M}$ – соответствующая алгебра событий.[25] Сопоставим далее каждому ω_j некоторое число $P(\omega_j) \equiv p_j$, $0 < p_j < 1$, $j = 1, 2, ..., n, ...$, причем потребуем, чтобы выполнялось равенство

$$p_1 + p_2 + ... + p_n + ... = 1. \tag{3.1}$$

Пусть далее $A \in \mathfrak{M}$ – некоторое событие, то есть

$$A = \left\{\omega_{j_1}, \omega_{j_2}, ..., \omega_{j_m}, ...\right\},$$

причем число элементов, составляющих *A*, может быть как конечно, так и счетно.[26]

Определение 3.1. ***Вероятностью события*** *A мы назовем величину*

$$P(A) = P(\omega_{j_1}) + P(\omega_{j_2}) + ... + P(\omega_{j_m}) + ... ,$$

или

$$P(A) = p_{j_1} + p_{j_2} + ... + p_{j_m} + ... ,$$

причем для $A = \varnothing$ *полагаем по определению* $P(\varnothing) = 0$. *Число* $P(\omega_j) \equiv p_j$ *естественно назвать* ***вероятностью элементарного события*** $\{\omega_j\}$.

Заметим, что это определение вполне согласуется с определением

---

[25] См. определение 1.8 в конце пункта 1.3.

[26] Если речь идет о счетном числе элементов, то левую часть формулы (3.1) и ей подобных, естественно, надо понимать как сумму бесконечного ряда. Читатель, незнакомый с основами математического анализа, может считать, что все рассматриваемые множества конечны, и суммы содержат конечное число слагаемых.

вероятности в классической модели. Достаточно в определении 3.1 положить $\Omega=\{\omega_1,\omega_2,...,\omega_n\}$, а $p_j \equiv 1/n$, $j=1,2,...,n$, и мы получим классическую модель вероятности (см. пункт 1.5).

Установим теперь основные свойства введенной таким образом вероятности.

Теорема 3.1. Вероятность $P(A)$ обладает следующими свойствами:

1. $0 \le P(A) \le 1$.
2. $P(\Omega)=1, \quad P(\varnothing)=0$.
3. Если события A и B *несовместимы*, т.е. $AB=\varnothing$, то

$$P(A+B)=P(A)+P(B).$$

3а. Если события

$$A_1,\ A_2,\ ...,\ A_n$$

несовместимы, т.е.

$$A_j A_k=\varnothing, \quad j\ne k, \quad j,k=1,2,...,n,$$

то

$$P(A_1+A_2+...+A_n)=P(A_1)+P(A_2)+...+P(A_n). \tag{3.2}$$

Действительно, первое свойство выполняется по определению, а второе следует из равенства (3.1).[27] Свойство 3 доказывается тем же путем, что и в теореме 1.1. Именно, пусть

$$A=\left\{\omega_{j_1},\omega_{j_2},...,\omega_{j_m},...\right\}, \quad B=\left\{\omega_{k_1},\omega_{k_2},...,\omega_{k_s},...\right\},$$

причем, в силу несовместимости событий A и B, элементы ω_r будут *все различны* для различных индексов r. Поэтому

$$A+B=\left\{\omega_{j_1},\omega_{j_2},...,\omega_{j_m},...,\omega_{k_1},\omega_{k_2},...,\omega_{k_s},...\right\},$$

откуда по определению (см. определение 3.1):

$$\begin{aligned}
&P(A+B)=P(\omega_{j_1})+P(\omega_{j_2})+...+P(\omega_{j_m})+...+P(\omega_{k_1})+P(\omega_{k_2})+...+P(\omega_{k_s})+...= \\
&=\left[P(\omega_{j_1})+P(\omega_{j_2})+...+P(\omega_{j_m})+...\right]+\left[P(\omega_{k_1})+P(\omega_{k_2})+...+P(\omega_{k_s})+...\right]= \\
&=P(A)+P(B),
\end{aligned}$$

что и требовалось.

Свойство 3а легко выводится из свойства 3 по индукции. Теорема доказана полностью.

---

[27] Ср. с замечанием после таблицы 1.3 в пункте 1.2.

Как уже было отмечено в замечании в конце предыдущего пункта, доказанная теорема позволяет перенести – по существу без изменения формулировок – практически все определения и теоремы предыдущих параграфов на случай модели дискретного вероятностного пространства.[28]

Возможность использовать модель дискретного вероятностного пространства существенно (по сравнению с классической моделью) расширяет круг практически важных задач, требующих вероятностного описания. Ниже мы в этом убедимся.

[28] Ср. с примером, разобранным в конце §1.

§ 4. НЕКОТОРЫЕ ДОПОЛНЕНИЯ И ПРИМЕРЫ

4.1. Обобщенная теорема умножения.

При решении задач часто оказываются удобными следующие обобщения формул (2.1а), (2.2а).

Терема 4.1 (*Первая обобщенная теорема умножения*). Пусть $A_1, A_2, ..., A_n$ – совокупность каких-либо событий. Тогда справедливо равенство:

(4.1)
$$P(A_1A_2...A_n) = \\ = P(A_1)\cdot P(A_2/A_1)\cdot P(A_3/A_1A_2)\cdot P(A_4/A_1A_2A_3)\cdot ...\cdot .P(A_n/A_1A_2...A_{n-1}).$$

Доказательство проще всего провести по индукции.

Для $n = 2$ теорема верна (это равенство (2.1а)). Пусть теорема верна для $n = k$. Докажем, что тогда она верна для $n = k+1$. Действительно, пусть у нас имеется $k+1$ событие $A_1, A_2, ..., A_k, A_{k+1}$; обозначим $A' = A_1A_2...A_k$. Тогда, согласно (2.1а) имеем:

$$P(A_1A_2...A_kA_{k+1}) = P(A'A_{k+1}) = P(A')P(A_{k+1}/A').$$

Но по предположению индукции

$$P(A') = P(A_1A_2...A_k) = \\ = P(A_1)\cdot P(A_2/A_1)\cdot P(A_3/A_1A_2)\cdot P(A_4/A_1A_2A_3)\cdot ...\cdot .P(A_k/A_1A_2...A_{k-1}),$$

что, после подстановки в предыдущее соотношение, и доказывает теорему.

Теорема 4.2 (*Вторая обобщенная теорема умножения*). Пусть $A_1, A_2, ..., A_n, H$ – совокупность каких-либо событий. Тогда справедливо равенство:

(4.2) $$P(A_1A_2...A_n/H) = P(A_1/H)\cdot P(A_2/A_1H)\cdot P(A_3/A_1A_2H)\cdot ...\cdot .P(A_n/A_1A_2...A_{n-1}H).$$

Действительно, по формуле (2.2) и теореме 4.2 можем написать:

$$P(A_1A_2...A_n/H)=\frac{P(A_1A_2...A_nH)}{P(H)}=\frac{P(HA_1A_2...A_n)}{P(H)}=$$
$$=\frac{P(H)\cdot P(A_1/H)\cdot P(A_2/A_1H)\cdot P(A_3/A_1A_2H)\cdot...\cdot.P(A_n/A_1A_2...A_{n-1}H)}{P(H)}=$$
$$=P(A_1/H)\cdot P(A_2/A_1H)\cdot P(A_3/A_1A_2H)\cdot...\cdot.P(A_n/A_1A_2...A_{n-1}H),$$

что и требовалось.

Замечание. Если предположить, что события $A_1, A_2, ..., A_n$ *независимы в совокупности*, то по смыслу понятия независимости естественно принять, что в этом случае должно быть выполнено равенство

(4.3) $$P(A_p/A_{j_1}A_{j_2}...A_{j_k}H)=P(A_p/H), \quad p\neq j_1, j_2, ..., j_k,$$

какое бы ни было событие H.

Из (4.2) и (4.3) получаем:

(4.4) $$P(A_1A_2...A_n/H)=P(A_1/H)\cdot P(A_2/H)\cdot P(A_3/H)\cdot...\cdot P(A_n/H).$$

В связи с равенством (4.3) *важно отметить* следующее. Выше, в пункте 2.2, мы определили независимость событий условием (2.4). Соотношение (4.3) является, по существу, естественным обобщением этого условия. Точно так же соотношение (4.4) является обобщением условия (2.5), (2.5а). Действительно, если в (4.3) положить

$$A_p=A; \quad A_{j_1}=A_{j_2}=...=A_{j_k}=B; \quad H=\Omega,$$

то получим (2.4). Если в (4.2), (4.3) положить $H=\Omega$, то из (4.4) получим (2.5а).[29]

4.2. Примеры.

Ниже мы рассмотрим несколько характерных примеров, иллюстрирующих результаты, описанные в этом и предыдущих параграфах.

Пример 1 (простейший анализ надежности систем).

1.1. Прибор состоит из трех блоков (рис. 4.1); выход из строя любого из блоков означает выход из строя прибора в целом. Блоки выходят из

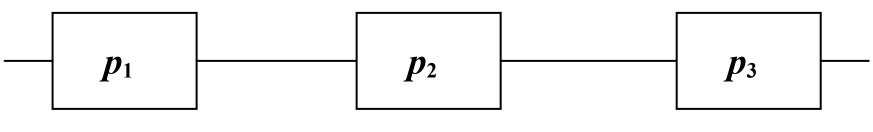


Рис. 4.1

[29] Для *независимых в совокупности* событий $A_1, A_2, ..., A_n$.

строя независимо друг от друга. Надежность (вероятность безотказной работы) каждого блока равна p_1, p_2, p_3 соответственно. Найти надежность P прибора в целом.

Замечание. Приведенную схему можно еще интерпретировать как *последовательную* цепь прохождения сигнала. Разрыв цепи (выход из строя любого из блоков) означает, что сигнал пройти не может, и прибор неработоспособен.

Решение. Обозначим через A_1, A_2, A_3 события, состоящие в том, что первый, второй и третий блоки соответственно находятся *в рабочем* состоянии. Тогда для события A, состоящего в том, что прибор оказывается работоспособным, можно написать: $A = A_1A_2A_3$. Но так как по условию события A_1, A_2, A_3 независимы, то по формуле (2.5а) получаем:

$$(4.5) \qquad P = P(A) = P(A_1A_2A_3) = p_1p_2p_3 .$$

1.2. Для повышения надежности прибора он дублируется другим точно таким же прибором (рис. 4.2). Надежность (вероятность безотказной работы) каждого прибора равна p_1, p_2 соответственно. При выходе из строя первого прибора происходит переключение на второй прибор.

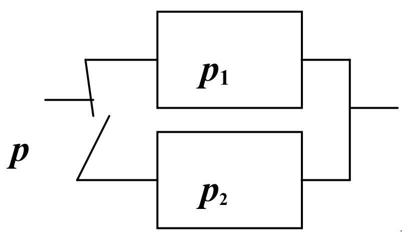


Рис. 4.2

Надежность переключающего устройства (т.е. вероятность удачного переключения) равна p. Определить надежность системы двух дублирующих друг друга приборов, если считать, что элементы системы выходят из строя независимо друг от друга.[30]

Решение. Пусть:

A – событие, состоящее в том, что система в целом работоспособна;
A_1 – событие, состоящее в том, что прибор 1 работоспособен;
A_2 – событие, состоящее в том, что прибор 2 работоспособен;
A_n – событие, состоящее в том, что переключатель работоспособен.

В данной задаче оказывается проще найти вероятность выхода системы из строя, т.е. вероятность события $\overline{A}$ – события, противоположного событию

[30] Здесь мы имеем дело, как легко видеть, с *параллельно* соединенными элементами.

A.[31] Вероятность же события A мы найдем тогда по формуле $P(A) = 1 - P(\overline{A})$.

Имеем: $\overline{A} = \overline{A}_1(\overline{A}_2 + \overline{A}_n)$ – то есть система выходит из строя тогда и только тогда, когда одновременно выходит из строя прибор 1 и либо не срабатывает переключатель, либо выходит из строя прибор 2. Но по формулам (1.1) $\overline{A}_2 + \overline{A}_n = \overline{A_2 A_n}$, а в силу независимости элементов системы

$$\begin{aligned} P(\overline{A}) &= P(\overline{A}_1)P(\overline{A}_2 + \overline{A}_n) = P(\overline{A}_1)P(\overline{A_2 A_n}) = \\ &= (1 - P(A_1))(1 - P(A_2 A_n)) = (1 - P(A_1))(1 - P(A_2)P(A_n)) = \\ &= (1 - p_1)(1 - p_2 p). \end{aligned}$$

Таким образом, для вероятности $P(A)$ получаем окончательно:

(4.6) $$P(A) = 1 - (1 - p_1)(1 - p_2 p).$$

В частности, если переключатель достоверно надежен, т. е. $p = 1$, то

(4.6а) $$P(A) = 1 - (1 - p_1)(1 - p_2).$$

Замечание. Приведенные выше примеры показывают, что если в системе элементы соединены *последовательно*, то вероятность надежной работы системы дается формулой типа (4.5). При *параллельном* соединении элементов (дублирующие устройства) следует пользоваться формулами типа (4.6), (4.6а). Проиллюстрируем это замечание следующим примером.

1.3. В технической системе дублированы не все, а только некоторые (наименее надежные) узлы (см. рис. 4.3). Надежности узлов проставлены на рисунке. Узлы выходят из строя независимо друг от друга. Определить надежность P системы.

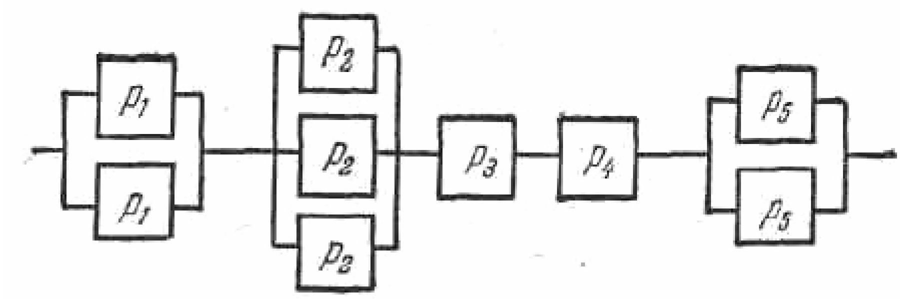


Рис. 4.3

[31] Переходить от задачи вычисления вероятности события $P(A)$ к задаче вычисления вероятности противоположного события $P(\overline{A})$ целесообразно, если событие A представляется в виде суммы двух *независимых* событий: $A = A_1 + A_2$. В этом случае $P(\overline{A}) = P(\overline{A_1 + A_2}) = P(\overline{A}_1\overline{A}_2) = P(\overline{A}_1)P(\overline{A}_2)$, так как по замечанию к определению 2.1 события $\overline{A}_1$ и $\overline{A}_2$ также независимы.

Решение. Из рисунка видно, что мы имеем дело с пятью последовательно соединенными блоками, из которых первый, второй и пятый, в свою очередь, состоят из параллельно соединенных (дублированных) узлов. Пусть A_j, $j = 1, 2, 3, 4, 5$ – события, состоящие соответственно в том, что блоки 1, 2, 3, 4 и 5 работоспособны. Пусть A – событие, состоящее в том, что система работоспособна в целом. Тогда по формулам (4.5) имеем: $P(A) = P(A_1)P(A_2)...P(A_5)$. С учетом формул (4.6а):

$$P(A_1) = 1-(1-p_1)^2; \quad P(A_2) = 1-(1-p_2)^3;$$
$$P(A_3) = p_3; \quad P(A_4) = p_4; \quad P(A_5) = 1-(1-p_5)^2.$$

Отсюда окончательно получаем:

$$P = P(A) = (1-(1-p_1)^2)(1-(1-p_2)^3)p_3p_4(1-(1-p_5)^2).$$

Дальнейшие примеры иллюстрируют применение формул полной вероятности и Байеса.

Пример 2.[32] Группа из трех самолетов совершает налет на объект. Объект защищен четырьмя батареями зенитных ракет. Каждая батарея простреливает угловой сектор размерами 60°, так что вокруг объекта оказываются защищенным только сектор в 240° из полного угла 360°. Если самолет пролетает через защищенный сектор, он попадает под обстрел; вероятность поражения самолета в этом случае равна p. Через незащищенный сектор самолет проходит беспрепятственно. Каждый самолет, прошедший к объекту, сбрасывает бомбу и поражает объект с вероятностью P. Экипажи самолетов не знаю, где расположены батареи. Найти вероятность поражения объектов для двух способов организации налета:

1) все три самолета летят по одному и тому же направлению, выбираемому случайно;

2) каждый из самолетов выбирает себе направление случайно независимо от других.

Решение. Заметим, прежде всего, что вероятность выбора самолетами незащищенного батареями направления полета равна $\frac{360^\circ - 240^\circ}{360^\circ} = \frac{1}{3}$ (геометрическая вероятность!). Поэтому для гипотез

H_1 – самолеты выбрали незащищенное направление;

H_2 – самолеты выбрали защищенное направление

[32] Этот пример достаточно сложный, но весьма показательный.

имеем соответственно: $P(H_1)=\dfrac{1}{3}$; $\quad P(H_2)=\dfrac{2}{3}$.

Далее, пусть A – событие, состоящее в том, что цель поражена. Пусть далее A_1, A_2, A_3 – события, состоящие в том, что цель поразил первый, второй и третий самолет соответственно. Пусть, наконец, B_1, B_2, B_3 – события, состоящие в том, что в результате обстрела (при пролете защищенного сектора) поражен первый, второй и третий самолет соответственно.

1) Для *первого* варианта организации налета, так как в этом случае *все* самолеты *вместе* либо проходят через незащищенный сектор, либо подвергаются обстрелу, имеем по формуле полной вероятности:

(*) $$P(A)=1-P(\overline{A})=1-(P(\overline{A}/H_1)P(H_1)+P(\overline{A}/H_2)P(H_2)).$$

Здесь $\overline{A}=\overline{A}_1\overline{A}_2\overline{A}_3$ (т.е. событие «цель *не поражена*» эквивалентно событию «*все* самолеты либо промахнулись, либо не смогли провести бомбометание, так как были подбиты»). Для *первого* варианта налета, когда все самолеты летят одним курсом (но при этом, по условию, бомбометание производится ими независимо и, следовательно, события $\overline{A}_1\overline{A}_2\overline{A}_3$ независимы в совокупности) естественно принять, что (ср. формулу (4.4)):

$$P(\overline{A}/H_k)=P(\overline{A}_1\overline{A}_2\overline{A}_3/H_k)=P(\overline{A}_1/H_k)P(\overline{A}_2/H_k)P(\overline{A}_3/H_k),\quad k=1,2.$$

При k=1, т.е. в случае беспрепятственного прохода самолетов к цели, имеем по условию задачи:

(+) $$P(\overline{A}_1/H_1)=P(\overline{A}_2/H_1)=P(\overline{A}_3/H_1)=1-P,$$

откуда

$$P(\overline{A}/H_1)=(1-P)^3.$$

Для k=2 вычисления проводим следующим образом. Имеем: $\overline{A}_j=\overline{A}_jB_j+\overline{A}_j\overline{B}_j$.[33] Но так как $B_j\subset\overline{A}_j$, то $\overline{A}_jB_j=B_j$. Поэтому

$$\overline{A}_jH_2=B_jH_2+\overline{A}_j\overline{B}_jH_2 \text{ и } P(\overline{A}_jH_2)=P(B_jH_2)+P(\overline{A}_j\overline{B}_jH_2),$$

так как события B_jH_2 и $\overline{A}_j\overline{B}_jH_2$ несовместимы (почему?). Отсюда:

[33] Напомним, что B_j – это событие, состоящее в том, что в результате обстрела подбит j-й самолет.

$$P(\overline{A}_j / H_2) = \frac{P(\overline{A}_j H_2)}{P(H_2)} = P(B_j / H_2) + P(\overline{A}_j \overline{B}_j / H_2).$$

По условию

$$P(B_j / H_2) = p; \quad P(\overline{A}_j \overline{B}_j / H_2) = \frac{P(\overline{A}_j \overline{B}_j H_2)}{P(H_2)} = \frac{P(\overline{A}_j \overline{B}_j H_2) P(\overline{B}_j / H_2)}{P(H_2) P(\overline{B}_j / H_2)} =$$

$$= \frac{P(\overline{A}_j \overline{B}_j H_2) P(\overline{B}_j / H_2)}{P(H_2 \overline{B}_j)} = P(\overline{A}_j / H_2 \overline{B}_j) P(\overline{B}_j / H_2) = (1-P)(1-p)$$

(согласно формуле (2.2): $P(H_2) P(\overline{B}_j / H_2) = P(H_2 \overline{B}_j)$). Таким образом,

(++) $$P(\overline{A}_j / H_2) = p + (1-P)(1-p) = 1 - (1-p)P.$$

Отсюда

$$P(\overline{A} / H_2) = \left[1 - (1-p)P\right]^3.$$

Окончательно в *первом* случае согласно формуле (*):

$$P_1(A) = 1 - \frac{1}{3}(1-P)^3 - \frac{2}{3}\left[1 - (1-p)P\right]^3.$$

2) Во *втором* варианте налета самолеты действуют самостоятельно, поэтому сначала надо вычислить полную вероятность поражения объекта каждым самолетом в отдельности (т.е. $P(A_j)$, $j = 1, 2, 3$), а затем использовать независимость событий $\overline{A}_1, \overline{A}_2, \overline{A}_3$.

Прежде всего, запишем (ср. формулу (*)):

(**) $$P(A_j) = 1 - P(\overline{A}_j) = 1 - (P(\overline{A}_j / H_1) P(H_1) + P(\overline{A}_j / H_2) P(H_2)).$$

Подставляя сюда формулы (+) и (++) получаем:

(***) $$p_1 = P(A_j) = \frac{1}{3} P + \frac{2}{3}(1-p)P, \quad j = 1, 2, 3.$$

Поскольку $\overline{A} = \overline{A}_1 \overline{A}_2 \overline{A}_3$ и события $\overline{A}_1, \overline{A}_2, \overline{A}_3$ независимы в совокупности, то

$$P(A)=1-P(\overline{A})=1-P(\overline{A}_1\overline{A}_2\overline{A}_3)=1-P(\overline{A}_1)P(\overline{A}_2)P(\overline{A}_3)=1-(1-p_1)^3.$$

Подставляя сюда (***), получаем окончательно для *второго* случая:

$$P_2(A)=1-\left[1-\frac{1}{3}P+\frac{2}{3}(1-p)P\right]^3.$$

Упражнение. Проверить, что вероятность поражения цели во втором случае больше, чем в первом, т.е. $P_2(A)>P_1(A)$.

Пример 3. Имеются три урны: в первой из них w_1 белых шаров и b_1 черных;[34] во второй w_2 белых шаров и b_2 черных; в третьей – w_3 белых шаров (черных нет). Некто выбирает наугад одну из урн и вынимает из нее шар. Этот шар оказался белым. Найти вероятность того, что этот шар вынут из первой урны.

Решение. Эту задачу естественно решать по формуле Байеса.

Гипотезы:

H_1 – выбрана первая урна;
H_2 – выбрана вторая урна;
H_3 – выбрана третья урна.

Поскольку выбор урн проводится *наугад*, то все гипотезы (до опыта) равновероятны:

$$P(H_1)=P(H_3)=P(H_3)=\frac{1}{3}.$$

Наблюдалось событие – обозначим его через A – появление белого шара. Условные вероятности будут равны соответственно:

$$P(A/H_1)=\frac{w_1}{w_1+b_1};\quad P(A/H_2)=\frac{w_2}{w_2+b_2};\quad P(A/H_3)=1.$$

По формуле Байеса вероятность того, что шар был вынут из первой урны:

$$P(H_1/A)=\frac{P(A/H_1)P(H_1)}{P(A/H_1)P(H_1)+P(A/H_2)P(H_2)+P(A/H_3)P(H_3)}=$$

$$=\frac{\frac{1}{3}P(A/H_1)}{\frac{1}{3}\left[P(A/H_1)+P(A/H_2)+P(A/H_3)\right]}=\frac{\frac{w_1}{w_1+b_1}}{\frac{w_1}{w_1+b_1}+\frac{w_2}{w_2+b_2}+1}.$$

[34] Обозначения выбраны по начальным буквам английских слов white – белый и black – черный. Индекс соответствует номеру урны.

Пример 4. В группе из 10 учащихся, пришедших на экзамен, 3 подготовленных отлично, 4 – хорошо, 2 – посредственно и 1 – плохо. В экзаменационных билетах имеется 20 вопросов. Отлично подготовленный учащийся может ответить на все 20 вопросов, хорошо подготовленный – на 16, посредственно – на 10, плохо подготовленный – на 5. Вызванный наугад учащийся отвечает правильно на три произвольно заданных вопроса (событие A). Найти вероятность того, что этот учащийся подготовлен: а) отлично; б) плохо.

Решение. Гипотезы:

H_1 – учащийся подготовлен отлично;
H_2 – учащийся подготовлен хорошо;
H_3 – учащийся подготовлен посредственно;
H_4 – учащийся подготовлен плохо.

Априорные (до опыта) вероятности гипотез:

$$P(H_1)=0,3;\quad P(H_2)=0,4;\quad P(H_3)=0,2;\quad P(H_4)=0,1.$$

Условные вероятности будут равны (объяснение вычислений – см. ниже *замечание*):

$$P(A/H_1)=1;\quad P(A/H_2)=\frac{16}{20}\cdot\frac{15}{19}\cdot\frac{14}{18}\approx 0,491;$$

$$P(A/H_3)=\frac{10}{20}\cdot\frac{9}{19}\cdot\frac{8}{18}\approx 0,105;\quad P(A/H_4)=\frac{5}{20}\cdot\frac{4}{19}\cdot\frac{3}{18}\approx 0,009.$$

Формула Байеса дает:

а) $P(H_1/A)=\dfrac{0,3\cdot 1}{0,3\cdot 1+0,4\cdot 0,491+0,2\cdot 0,105+0,1\cdot 0,009}=\dfrac{0,3}{0,518}\approx 0,58$;

б) $P(H_4/A)=\dfrac{0,1\cdot 0,009}{0,518}\approx 0,002$.

Замечание. Вычисление $P(A/H_2)$ обосновывается следующим образом. Обозначим через D_1, D_2, D_3 – события, состоящие соответственно в том, что учащийся правильно ответил на первый, второй и третий вопросы. Тогда: $A=D_1D_2D_3$. Для вероятности $P(A/H_2)$, следовательно, по теореме 4.2, имеем:

$$P(A/H_2)=P(D_1D_2D_3/H_2)=P(D_1/H_2)P(D_2/H_2D_1)P(D_3/H_2D_1D_2).$$

Но:

$P(D_1 / H_2) = \dfrac{16}{20}$ (так как благоприятных исходов при выборе первого вопроса для «хорошиста» – 16 из общего числа 20 возможных);

$P(D_2 / H_2 D_1) = \dfrac{15}{19}$ (так как благоприятных исходов при выборе второго вопроса *при условии удачного выбора первого* – для «хорошиста» уже 15 из общего числа оставшихся 19 возможных);

$P(D_3 / H_2 D_1 D_2) = \dfrac{14}{18}$ (так как благоприятных исходов при выборе третьего вопроса *при условии удачного выбора первого и второго* – для «хорошиста» уже 14 из общего числа оставшихся 18 возможных).

Таким образом, получаем: $P(A / H_2) = \dfrac{16}{20} \cdot \dfrac{15}{19} \cdot \dfrac{14}{18} \approx 0,491$.

Другой способ подсчета подобной вероятности, основанный на комбинаторных методах, будет указан в следующем параграфе.

Пример 5. Пусть у нас имеется две урны, причем в первой урне первоначально находится 3 белых и 3 черных шара, а во второй – 5 белых и 2 черных. Из первой урны во вторую перекладываются наудачу два шара, после чего из второй урны наудачу извлекается один шар. Требуется ответить на следующие вопросы.

1) Какова вероятность того, что извлеченный из второй урны шар будет белым?

2) Если оказалось, что извлеченный из второй урны шар – белый, то какова вероятность того, что из первой урны во вторую переложили один черный и один белый шар?

Решение. Зададим *полную группу событий* следующим образом.

H_{ww} – событие, состоящее в том, что из первой урны во вторую переложили два белых (white) шара;

H_{wb} – событие, состоящее в том, что из первой урны во вторую переложили один белый и один черный (black) шар;

H_{bb} – событие, состоящее в том, что из первой урны во вторую переложили два черных шара.

Обозначим далее:

N_{ww} – число способов, которыми можно выбрать два белых шара из первой урны;

N_{wb} – число способов, которыми можно выбрать один белый и один черный шар из первой урны;

N_{bb} – число способов, которыми можно выбрать два черных шара из первой урны;

N – число способов, которыми можно выбрать два (любых) шара из первой урны.

Элементарный подсчет дает следующие значения:

$$N_{ww} = 3; \quad N_{wb} = 9; \quad N_{bb} = 3; \quad N = 15\,.$$

Отсюда (классическая модель вероятности!) получаем:

$$P(H_{ww}) = \frac{N_{ww}}{N} = \frac{1}{5}; \quad P(H_{wb}) = \frac{N_{wb}}{N} = \frac{3}{5}; \quad P(H_{bb}) = \frac{N_{bb}}{N} = \frac{1}{5}\,.$$

Пусть далее A_w – событие, состоящее в том, что вынутый из второй урны шар будет белым. Тогда получаем:

1) по формуле полной вероятности

$$P(A_w) = P(A_w / H_{ww})P(H_{ww}) + P(A_w / H_{wb})P(H_{wb}) + P(A_w / H_{bb})P(H_{bb})\,.$$

Но

$$P(A_w / H_{ww}) = \frac{7}{9}; \quad P(A_w / H_{wb}) = \frac{6}{9} = \frac{2}{3}; \quad P(A_w / H_{bb}) = \frac{5}{9}\,,$$

так как во второй урне всего стало 9 шаров, причем в случае реализации события H_{ww} из них будет 7 белых (и 2 черных), в случае реализации события H_{wb} из них будет 6 белых (и 3 черных) и, наконец, в случае реализации события H_{bb} во второй урне будет 5 белых шаров (и 4 черных).

Таким образом, получаем:

$$P(A_w) = \frac{7}{9}\cdot\frac{1}{5} + \frac{2}{3}\cdot\frac{3}{5} + \frac{5}{9}\cdot\frac{1}{5} = \frac{2}{3}\,.$$

2) В этом случае просто используем формулу Байеса:

$$P(H_{wb} / A_w) = \frac{P(A_w / H_{wb})P(H_{wb})}{P(A_w)} = \frac{\frac{2}{3}\cdot\frac{3}{5}}{\frac{2}{3}} = \frac{3}{5}\,.$$

Ответ: $P(A_w) = \dfrac{2}{3}$, $P(H_{wb} / A_w) = \dfrac{3}{5}$.

§5. ЭЛЕМЕНТЫ КОМБИНАТОРИКИ

5.1. Введение.

В последнем примере предыдущего параграфа для вычисления вероятности интересующего нас события мы встретились с задачей следующего типа: необходимо вычислить количество способов, которыми можно выбрать m шаров из общего числа в n ($n>m$) шаров. Аналогичная задача возникает при попытке вычислить, например, вероятность того, что при раздаче из колоды в 36 карт наугад шести карт, среди этих шести окажется: ровно два туза, не менее одного туза и т.п. Для этого (если вероятность вычислять по классической модели) нам необходимо уметь вычислять сколькими *различными* способами можно из колоды 36 карт вынуть 6 карт, сколькими *различными* способами из 36 карт можно составить комбинации по 6 карт, содержащие двух тузов, не менее одного туза и т.п. Задачи такого рода и составляют, в частности, предмет раздела математики, называемого *комбинаторикой* (от латинского слова combino – соединяю).

Более точно, ***комбинаторикой*** (комбинаторным анализом) *называется раздел математики, посвященный решению задач выбора и расположения элементов некоторого, обычно конечного, множества в соответствии с заданными правилами*. Каждое такое правило определяет способы построения некоторой конструкции (набора, совокупности) элементов данного множества, называемой *комбинаторной конфигурацией*. Поэтому можно сказать, что *целью комбинаторики является изучение комбинаторных конфигураций*, в частности алгоритмы их построения, задачи на перечисление и т.п.

Основное множество (содержащее, скажем, n элементов) принято называть ***генеральной совокупностью*** (из n элементов). Комбинаторную конфигурацию, составленную из m элементов этого множества, принято называть ***выборкой из m элементов***, или ***выборкой размера m*** (из данной генеральной совокупности).

5.2. Основное правило комбинаторики.

Начнем со следующего очевидного утверждения.

Лемма 1. Пусть у нас имеется две группы элементов

$$\{a_1, a_2, ..., a_m\} \text{ и } \{b_1, b_2, ..., b_n\}$$

по m и n элементов соответственно. Тогда из этих элементов можно образовать ровно mn различных пар (a_j, b_k), содержащих по одному элементу из каждой группы.

Действительно, составим из этих пар прямоугольную таблицу, содержащую m строк и n столбцов, так, чтобы пара (a_j, b_k) стояла на пересечении j-й строки и k-го столбца. Каждая из пар (a_j, b_k) встретится в этой таблице ровно один раз, и утверждение леммы становится очевидным.

Теорема 5.1 (*основное правило комбинаторики*). Пусть даны r групп элементов:

1-я группа $\{a_1, a_2, ..., a_{n_1}\}$ - содержит n_1 элементов;

2-я группа $\{b_1, b_2, ..., b_{n_2}\}$ - содержит n_2 элементов;

………………..

r-я группа $\{x_1, x_2, ..., x_{n_r}\}$ - содержит n_r элементов.

Тогда можно образовать ровно $n_1 \cdot n_2 \cdot ... \cdot n_r$ различных комбинаций

$$\left(a_{j_1}, b_{j_2}, ..., x_{j_r}\right),$$

содержащих по одному элементу из каждой группы.

Доказательство этой теоремы легко проводится по индукции из леммы 1.

На практике бывает полезной еще следующая эквивалентная формулировка теоремы 1.

Теорема 5.1а. Пусть требуется выполнить одну за другой r операций, причем первую операцию можно выполнить n_1 способами, вторую - n_2 способами, … , r-ю - n_r способами. Тогда все r операций вместе могут быть выполнены числом способов, равным

$$n_1 \cdot n_2 \cdot ... \cdot n_r = \prod_{j=1}^{r} n_j .$$

Примеры.

1. В соревнованиях по футболу принимают участие 16 команд. Сколькими способами могут быть распределены первое и второе места?

Решение. Первое место может занять одна из 16 команд. После того, как определена команда, занявшая первое место, второе место может занять одна из оставшихся 15 команд. Следовательно, общее число способов, которыми могут быть распределены первое и второе места, равно 16×15 = 240.

2. Сколько четырехзначных чисел можно составить из цифр 0, 1, 2, 3, 4, 5, если:

а) ни одна из цифр не повторяется более одного раза;

б) цифры могут повторяться;

в) числа должны быть нечетными (цифры могут повторяться)?

Решение. а) Первой цифрой числа может быть одна из 5 цифр 1, 2, 3, 4, 5 (очевидно, 0 не может быть первой цифрой). Если первая цифра выбрана, то вторую можно выбрать 5 способами (0 в этом случае уже может быть выбран), третью – 4 способами, четвертую – 3 способами. Согласно теореме 2а, общее число способов равно 5×5×4×3 = 300.

б) Первой цифрой числа может быть одна из 5 цифр 1, 2, 3, 4, 5 (5 возможностей), для каждой из следующих трех цифр имеем 6 возможностей (0, 1, 2, 3, 4, 5). Следовательно, число искомых чисел равно 5×6×6×6 = 1080.

в) Первой цифрой может быть одна из цифр 1, 2, 3, 4, 5, второй и третьей – одна из цифр 0, 1, 2, 3, 4, 5, а последней – одна из цифр 1, 3, 5 (так как число должно быть нечетным). Следовательно, число искомых чисел равно 5×6×6×3 = 540.

3. *Сложная классификация*. Предположим, что люди классифицируются по полу, семейному положению (состоит или не состоит в браке) и профессии. Если рассматривать, например, 17 профессий, то всего мы будем иметь 2×2×17 = 68 различных классов.

4. В мастерской по изготовлению ключей есть 12 типов заготовок для ключей. Из каждой заготовки можно сделать ключ, вырезав выступы в пяти определенных местах, причем на первом месте величина выступа может принимать 2 значения, а на остальных – 3 значения. Сколько различных типов ключей может изготовить мастерская?

Решение. Для заготовки каждого типа первый выступ можно выбрать 2 способами, а остальные 4 выступа – 3 способами каждый. Таким образом, каждая заготовка допускает 2×3×3×3×3 = 162 различных комбинации выступов. Всего, следовательно, будет 162×12 = 1944 различных типа ключей.

5. *«Размещение шаров по ящикам»*. Задача состоит в размещении r различных шаров по n различным ящикам. Сколько может быть различных вариантов такого размещения (количество шаров в ящике не ограничивается, т.е. может быть от 0 до r)?

Решение. Для каждого шара ящик можно выбрать n различными способами. Следовательно, r различных шаров можно разместить по n различным ящикам n^r различными способами.

6. Сколько имеется пятизначных натуральных чисел, которые делятся на 5?

Решение. Для того чтобы целое число делилось на 5, необходимо и достаточно, чтобы оно заканчивалось цифрами 0 или 5. Таким образом, первую цифру мы можем выбрать 9 способами (1, 2, 3, 4, 5, 6, 7, 8, 9), вторую, третью и четвертую цифру – 10 способами (0, 1, 2, 3, 4, 5, 6, 7, 8, 9), а пятую цифру – 2 способами (0, 5). Поэтому искомое количество чисел будет равно 9×10×10×10×2 = 18000.

5.3. Размещения, перестановки, сочетания.

В этом пункте мы рассмотрим три основных типа выборок (комбинаторных конфигураций). Напомним, что выборка характеризуется способом построения некоторой конструкции (набора, совокупности) элементов данного множества (генеральной совокупности). В частности, выборка задается, с одной стороны, алгоритмом ее построения, а с другой – условиями, различающими одну выборку от другой. Например, выбор элементов выборки может осуществляться *последовательно*, и порядок следования элементов в выборке считается существенным для различения выборок (даже, если они состоят из одинаковых элементов: например, при определении цифр в номере выигравшего лотерейного билета). С другой стороны, в «спортлото» *порядок выпадения* номеров шаров по условию не существенен, то есть, в принципе, шары можно было бы выбирать не последовательно, как это делается при розыгрыше, а сразу «комплектом» из 5 (или 6) штук. Результат от этого не изменился бы.

Всюду ниже в этом пункте мы рассматриваем некоторую генеральную совокупность фиксированного объема n и фиксируем объем выборки ($m \leq n$ элементов). При этом элементы выборки по условию «извлекаются» из генеральной совокупности *без возвращения*, т.е. выборка не может содержать одинаковые (повторяющиеся) элементы.

1. *Размещения.*

«Размещением из n элементов по m» называется ***упорядоченная*** *выборка без возвращения объема m из генеральной совокупности, состоящей из n элементов.*

Это означает, что соответствующие выборки мы различаем как по элементному составу, так и по порядку следования элементов в выборке. Например, выборки объема 4: (1, 3, 7, 4) и (3, 7, 4, 1) из генеральной совокупности (1, 2, 3, 4, 5, 6, 7, 8, 9) объема 9 мы считаем по условию различными.

Теорема 5.2. Обозначим через P_n^m число различных размещений из n элементов по m. Тогда $P_n^m = n(n-1)(n-2)...(n-m+1)$.

Действительно, первый элемент размещения мы можем выбрать n способами, второй – (n – 1) способом, третий – (n – 2) способами, и так далее, m-й элемент мы можем выбрать (n – m + 1) способами. Теперь осталось применить основное правило комбинаторики. Теорема доказана.

Замечание. Помимо обозначения P_n^m для числа размещений часто используется обозначение A_n^m.

2. *Перестановки*.

«Перестановкой из n элементов» называется упорядоченная выборка без возвращения объема n из генеральной совокупности, состоящей из n элементов.

Из предыдущего пункта очевидно следует

Теорема 5.3. *Число различных перестановок из n элементов равно*

$$n(n-1)(n-2)\cdot\ldots\cdot 2\cdot 1 = n!.$$

Замечание. Для числа перестановок из n элементов чаще всего используется обозначение P_n.

3. *Сочетания*.

«Сочетанием из n элементов по m» называется ***неупорядоченная*** *выборка без возвращения объема m из генеральной совокупности, состоящей из n элементов.*

При *фиксированном элементном составе* размещения из n элементов по m мы имеем $m!$ различных (различаемых только по порядку следования элементов) выборок. Но поскольку в *сочетании* порядок элементов несущественен, то становится очевидной

Теорема 5.4. Обозначим через C_n^m число различных сочетаний из n элементов по m. Тогда:

$$C_n^m = \frac{P_n^m}{m!} = \frac{n(n-1)(n-2)\ldots(n-m+1)}{m!} = \frac{n!}{m!(n-m)!}. \tag{5.1}$$

Замечание. Обратим внимание читателя на то, что число различных сочетаний из n элементов по m совпадает с соответствующим *биномиальным коэффициентом* в формуле бинома Ньютона:

$$\begin{aligned}&(a+b)^n = \\ &= C_n^0 a^n + C_n^1 a^{n-1} b + C_n^2 a^{n-2} b^2 + \ldots + C_n^m a^{n-m} b^m + \\ &+ \ldots + C_n^{n-1} a b^{n-1} + C_n^n b^n.\end{aligned} \tag{5.2}$$

Отсюда, в частности, следует, что число всех подмножеств n-элементного множества равно 2^n. Действительно, поскольку в подмножества данного множества различаются только по составу элементов, то число различных k-элементных подмножеств данного множества совпадает с числом сочетаний из n элементов по k. Поэтому общее число *всех* подмножеств данного n-элементного множества (включая само это множество и пустое множество) равно

$$C_n^0 + C_n^1 + C_n^2 + \ldots + C_n^k + \ldots + C_n^{n-1} + C_n^n = (1+1)^n = 2^n.$$

5.4. Примеры.

1. У филателиста есть 8 разных марок на космическую тему и 10 разных марок на спортивную тему. Сколькими способами можно наклеить 3 марки первого вида и 3 марки второго вида в альбом на 6 пронумерованных мест?

Решение. Филателист может выбрать 3 марки из 8 числом способов, равным $C_8^3 = \dfrac{8!}{3!5!} = 56$, и 3 марки из 10 числом способов, равным $C_{10}^3 = \dfrac{10!}{3!7!} = 120$. При этом первую группу из 3 марок, он может разместить на 6 *пронумерованных* местах числом способов, равным $P_6^3 = 120$; при этом вторая группа из 3 марок размещается на оставшихся трех местах числом способов, равным 3! = 6. Таким образом, общее число различных способов, которыми можно наклеить марки в альбом, равно $56 \cdot 120 \cdot 120 \cdot 6 = 4\ 838\ 400$.

2. Сколькими способами можно переставить буквы слова «Юпитер», чтобы гласные шли в алфавитном порядке?

Решение. Число способов, которыми можно разместить в алфавитном порядке три различные гласные буквы на три места из шести, равно $C_6^3 = 20$.[35] На оставшиеся три места остальные буквы (в любом порядке) можно разместить 3! = 6 способами. Таким образом, искомое число способов равно 20×6 = 120.

3. Сколько есть пятизначных чисел, которые одинаково читаются слева направо и справа налево (например, таких, как 67876, 17071)?

Решение. Очевидно, выбор первых двух цифр такого числа однозначно определяет последние две его цифры. На первом месте может стоять любая цифра, кроме нуля (т.е. девять цифр от 1 до 9). На втором и

[35] То есть число способов, которыми можно выбрать три числа из шести без учета порядка, ибо каждой неупорядоченной тройке различных чисел соответствует один и только один упорядоченный набор, в котором эти числа идут по возрастанию.

третьем месте может стоять любая из десяти цифр (от 1 до 9 и 0). Поэтому, искомое количество чисел будет равно 9×10×10 = 900.

4. Сколькими способами можно разложить 19 различных предметов по 5 различным ящикам так, чтобы в 4 ящика легли по 4 предмета, а в оставшийся ящик – 3 предмета?

Решение. В первый ящик 4 предмета можно положить числом способов, равным $C_{19}^4 = 3876$. Во второй ящик 4 предмета можно положить, следовательно, числом способов, равным $C_{15}^4 = 1365$; в третий ящик – числом способов, равным $C_{11}^4 = 330$, в четвертый – числом способов, равным $C_7^4 = 35$. После этого, предметы для последнего ящика определяются однозначно. Наконец, мы должны учесть, что при распределении групп по 4 предмета используются 4 ящика из 5. Число различных комбинаций по 4 ящика из 5 равно $C_5^4 = 5$.

Таким образом, искомое число способов равно

$$C_{19}^4 \cdot C_{15}^4 \cdot C_{11}^4 \cdot C_7^4 \cdot C_5^4 = 305\,540\,235\,000.$$

Следующие несколько примеров более сложны, но имеют важное прикладное значение.

5. *Распределение r шаров по n ящикам*. Сколькими способами можно r неразличимых между собой шаров распределить по n различным ящикам? При этом предполагается, что число шаров в каждом ящике может принимать любое значение от 0 до r.

Решение. Будем представлять ящики как промежутки между $n+1$ черточками (черточки имитируют стенки ящиков), а шары условимся обозначать звездочками. Так, символ

|***|*||||****|

означает, что r = 8 шаров размещено по n = 6 ящикам, причем эти ящики содержат последовательно 3, 1, 0, 0, 0, 4 шаров. Такие символы всегда начинаются и кончаются черточками, но оставшиеся $n-1$ черточку и r звездочек можно разместить в произвольном порядке. Число таких различимых размещений будет равно числу способов выбора r мест среди $n+r-1$ мест (или, что то же, числу способов выбора $n-1$ мест среди $n+r-1$ мест), то есть будет равно числу сочетаний из $n+r-1$ элементов по r (или, соответственно, по $n-1$):

$$C_{n+r-1}^{r} = C_{n+r-1}^{n-1}.$$

Замечание. Если ящики считать неразличимыми, то число искомых размещений будет в n! раз меньше (почему?).

5.1. В условиях предыдущей задачи, указать число способов которыми можно r неразличимых между собой шаров распределить по n различным ящикам с условием, что ни один ящик не будет пустым. Очевидно, что при этом должно быть $r \geq n$.

Решение. Условие отсутствия пустых ящиков приводит к следующему ограничению: каждая черточка (кроме крайних) должна быть заключена между двумя звездочками. Это соответствует числу способов, которыми можно разместить $n-1$ черточку в $r-1$ промежуток между звездочками. Это число равно, очевидно,

$$C_{r-1}^{n-1}.$$

5.2. Указать число различных *целых неотрицательных* решений $r_1, r_2, ..., r_n$ уравнения

$$r_1 + r_2 + ... + r_n = r, \quad r_j \geq 0.$$

(Порядок чисел $r_1, r_2, ..., r_n$ здесь существенен: например, решения $r_1 = 2, \quad r_2 = 3$ и $r_1 = 3, \quad r_2 = 2$ считаются различными).

Решение. Очевидно, эта задача эквивалентна задаче о числе размещений r шаров по n ящикам. Таким образом, число различных целых неотрицательных решений указанного уравнения будет равно $C_{n+r-1}^{r} = C_{n+r-1}^{n-1}$.[36]

Аналогично, число различных *натуральных* решений этого уравнения будет равно C_{r-1}^{n-1} $(r \geq n)$.

6. *Разбиение на группы.*

Теорема 5.5. Пусть $r_1, r_2, ..., r_k$ - целые неотрицательные числа, причем

$$r_1 + r_2 + ... + r_k = n, \quad r_j \geq 0.$$

Тогда число способов, посредством которых *n различных* элементов могут быть разделены на *k* групп, из которых первая группа содержит r_1 элементов, вторая - r_2 элементов и т.д., равно[37]

$$\frac{n!}{r_1! r_2! ... r_k!}.$$

Обращаем внимание читателя на то, что данная задача принципиально отличается от рассмотренной в предыдущем примере задачи о распределении шаров по ящикам: в примере 5 шары считаются неразличимыми, в то время как в нашем случае элементы предполагаются различными. Не имеет значения лишь порядок элементов в каждой группе,

[36] Для читателей, знакомых с анализом, отметим, что очевидным следствием этого результата является тот факт, что этой же величине равно число различных частных производных порядка r аналитической функции n переменных: $\partial^r f(x_1, x_2, ..., x_n) \big/ \partial x_1^{r_1} \partial x_2^{r_2} ... \partial x_n^{r_n}$. Действительно, частные производные порядка r от аналитической функции n переменных не зависят от порядка дифференцирования, а зависят лишь от того, сколько раз мы дифференцируем по каждому переменному. Каждое переменное здесь играет роль «ящика», а роль «шара» играет операция дифференцирования.

[37] Случай $r_j = 0$ также включается, если учесть формулу 0! = 1.

но элементный состав групп существенен. Кроме того, порядок групп остается также существенным в том смысле, что, например, разбиения ($r_1 = 2, \quad r_2 = 3$) и ($r_1 = 3, \quad r_2 = 2$) считаются различными.

Доказательство. Выберем сначала первую группу, состоящую из r_1 элементов. Такую группу можно выбрать числом способов, равным $C_n^{r_1}$. Вторую группу объема r_2 можно, следовательно, выбрать числом способов, равным $C_{n-r_1}^{r_2}$, третью – числом способов, равным $C_{n-r_1-r_2}^{r_3}$ и т.д. После образования $(k-1)$-й группы, остается $n - r_1 - r_2 - ... - r_{k-1} = r_k$ элементов, которые и образуют последнюю группу (т.е. последняя группа выбирается единственным способом). Таким образом, согласно основному правилу комбинаторики (теорема 5.1а), интересующее нас число способов, посредством которых n *различных* элементов могут быть разделены на k групп, из которых первая содержит r_1 элементов, вторая - r_2 элементов и т.д., будет равно

$$N = C_n^{r_1} \cdot C_{n-r_1}^{r_2} \cdot C_{n-r_1-r_2}^{r_3} \cdot ... \cdot C_{n-r_1-r_2-...-r_{k-2}}^{r_{k-1}} .$$

Преобразуем последнее выражение:

$$N =$$

$$= \frac{n!}{r_1!(n-r_1)!} \cdot \frac{(n-r_1)!}{r_2!(n-r_1-r_2)!} \cdot \frac{(n-r_1-r_2)!}{r_3!(n-r_1-r_2-r_3)!} \cdot ... \cdot \frac{(n-r_1-r_2-...-r_{k-2})!}{r_{k-1}!(n-r_1-r_2-r_3-...-r_{k-1})!} =$$

$$= \frac{n!}{r_1!r_2!...r_{k-1}!r_k!},$$ так как $n - r_1 - r_2 - ... - r_{k-1} = r_k$. Теорема доказана.

В качестве *примера* применения последней теоремы рассмотрим следующую задачу. При игре в бридж между четырьмя игроками распределяют $n = 52$ карты по 13 карт каждому ($r_1 = r_2 = r_3 = r_4 = 13$). Какова вероятность того, что каждый игрок получит ровно по одному тузу?

Решение. Число вариантов распределения по 13 карт между четырьмя игроками соответствует числу способов, которыми можно 52 элемента разбить на 4 группы по 13 элементов в каждой, т.е., согласно теореме 5.5, равно $\frac{52!}{(13!)^4}$. Четыре туза могут быть распределены между четырьмя игроками $4! = 24$ способами. Оставшиеся 48 карт могут быть распределены (опять же согласно теореме 5.5) $48!/(12!)^4$ способами. Следовательно, искомая вероятность равна

$$24 \cdot \frac{48!(13!)^4}{(12!)^4 52!} = 24 \cdot \frac{48! 13^4}{52!} = 0{,}105... .$$

Любопытно, что при игре «в дурака» такая вероятность оказывается существенно меньше. Действительно, найдем вероятность того, что при раздаче четырем игрокам по 6 карт из колоды в 36 карт, каждый игрок получит ровно по одному тузу. Поскольку раздается 24 карты из 36, то нам прежде всего надо знать число способов, которыми можно выбрать 24 карты из 36. Это число равно $C_{36}^{24} = 36!/(24!12!)$.

Далее, число способов, которыми можно разбить 24 карты на 4 группы по 6 карт согласно теореме 5.5 равно $24!/(6!)^4$. Таким образом, общее число способов, которыми можно раздать четырем игрокам по 6 карт из колоды в 36 карт, равно $C_{36}^{24} \cdot \dfrac{24!}{(6!)^4}$. Четыре туза могут быть распределены между четырьмя игроками $4! = 24$ способами. Число способов, которыми можно раздать четырем игрокам по 5 карт из оставшихся 32 карт подсчитывается аналогично предыдущему, и будет равно $C_{32}^{20} \cdot \dfrac{20!}{(5!)^4}$. Таким образом, искомая вероятность равна

$$\frac{24 \cdot C_{32}^{20} \cdot \dfrac{20!}{(5!)^4}}{C_{36}^{24} \cdot \dfrac{24!}{(6!)^4}} = 24 \cdot \frac{32!12!6^4}{12!36!} = \frac{24 \cdot 1296}{33 \cdot 34 \cdot 35 \cdot 36} \approx 0,022 \ .$$

§6. ИСПЫТАНИЯ БЕРНУЛЛИ. ФОРМУЛА ПУАССОНА

6.1. Схема независимых испытаний Бернулли.

На практике часто встречается ситуация, хорошо иллюстрирующаяся следующими примерами.

Некто несколько раз подряд бросает монету. Спрашивается, можно ли заранее оценить вероятность того, что в результате n бросаний герб выпадет ровно m раз? Или: n раз бросается игральная кость; требуется оценить вероятность того, что при этом m раз выпадет 5 или 6 очков.

Очевидно, что без дополнительных предположений относительно условий проведения эксперимента однозначно ответить на эти вопросы нельзя. Так, результат, несомненно, должен зависеть от того, является ли монета (или кость) правильной, т.е. однородной и симметричной. С другой стороны, возможно ли ответить на такой вопрос: сколько раз надо бросить монету (или кость), чтобы с достаточной степенью уверенности можно было утверждать, что данная монета (или кость) *не является правильной*? Умение отвечать на такой вопрос весьма важно, например, для игорных заведений.

Естественно предположить, что если монета правильная, то вероятность появления герба при каждом бросании равна ½. Аналогично, в случае правильной кости вероятность появления 5 или 6 очков при каждом бросании равна ⅓. Иными словами, если испытаний достаточно много, то герб при бросании монеты будет появляться примерно в половине исходов, а 5 или 6 очков на кости – в одной трети случаев.

Однако все эти рассуждения основаны на интуиции. Мы же постараемся в этом параграфе описать теоретическую модель, которая позволит нам вполне обоснованно ответить на все сформулированные выше вопросы. Модель, о которой пойдет речь ниже, впервые была предложена швейцарским математиком Якобом Бернулли (1654-1705), и получила его имя.[38]

Схема независимых испытаний Бернулли. Будем производить последовательные испытания, в результате каждого из которых может наступить или не наступить некоторое событие A. Пусть при каждом отдельном испытании вероятность наступления события A одна и та же и не зависит от наступления или ненаступления этого события при других

[38] Основные результаты Я. Бернулли по теории вероятностей были опубликованы лишь после его смерти в 1713 г. Брат Я. Бернулли – Иоганн (1667-1748) и сын – Даниил (1700-1782) являлись членами Петербургской Императорской Академии Наук, и внесли большой вклад в развитие вариационного исчисления и теоретической механики.

испытаниях; обозначим эту вероятность через p. Обычно говорят, что p – это вероятность «успеха»; соответственно величина $q = 1 - p$ называется вероятностью «неудачи». Понятно, что эта терминология весьма условна. Такая модель называется *схемой (независимых) испытаний Бернулли*.

Зададимся следующим вопросом: какова вероятность того, что при проведении n испытаний «успех» (т.е. появления события A) будет наблюдаться ровно в m случаях?[39]

Эта задача решается следующим образом. Представим себе все возможные комбинации из последовательных результатов наших испытаний. Так, например, в случае 3 испытаний возможны восемь таких комбинаций, а именно:[40]

$$AAA;\quad AA\overline{A};\quad A\overline{A}A;\quad \overline{A}AA;$$
$$A\overline{A}\,\overline{A};\quad \overline{A}A\overline{A};\quad \overline{A}\,\overline{A}A;\quad \overline{A}\,\overline{A}\,\overline{A}.$$

Выделим те комбинации, в которых событие A наступает ровно m раз (и, следовательно, не наступает ровно $n - m$ раз); назовем для краткости такие комбинации *допустимыми*. Определим вероятность появления каждой отдельной допустимой комбинации. Для этого заметим, что появление допустимой комбинации представляет собой произведение n событий, а именно: m наступлений события A при одних испытаниях и $n - m$ его ненаступлений при других испытаниях. Вероятность наступления события A при каждом отдельном испытании по условию равна p; вероятность его ненаступления равна, следовательно, $q = 1 - p$. По условию наступления или ненаступления события A при различных испытаниях представляют собой независимые события; следовательно, вероятность их произведения равна произведению их вероятностей, т.е. равна величине $p^m q^{n-m} = p^m (1 - p)^{n-m}$.

Заметим теперь, что событие, состоящее в наступлении события A ровно при m испытаниях, равносильно появлению хотя бы одной из допустимых комбинаций. Так как различные допустимые комбинации представляют собой несовместимые события, искомая вероятность появления события A ровно в m испытаниях равна сумме вероятностей появления допустимых комбинаций. Поскольку вероятности появления допустимых комбинаций одинаковы, то вероятность их суммы равна величине $Kp^m q^{n-m}$, где K – число всех допустимых комбинаций. Это число равно, очевидно, числу различных способов, которыми можно выделить m мест из общего числа n мест, иными словами равно числу сочетаний из n элементов по m, т.е. равно $C_n^m = C_n^{n-m} = \dfrac{n!}{m!(n-m)!}$.

Таким образом, вероятность появления ровно m «успехов» в последовательности n независимых испытаний Бернулли равна

[39] Здесь естественно считать, что $m = 0, 1, 2, \ldots, n$.

[40] Здесь $\overline{A}$ означает событие, противоположное событию A, т.е. «неудачу».

(6.1) $$P_n(m) = b(m;\, n,\, p) = C_n^m p^m q^{n-m} = C_n^m p^m (1-p)^{n-m}\,.$$

Это и есть основная формула схемы независимых испытаний Бернулли.

Функция $b(m; n, p)$, называемая *биномиальным распределением*, или *распределением Бернулли*, определяется формулой (6.1) и дает значение вероятности m «успехов» в n испытаниях Бернулли с вероятностью «успеха» p. При фиксированных n и p она является функцией целочисленного неотрицательного аргумента m.

Испытания Бернулли – теоретическая схема, и только практика может показать, годна ли схема для описания данного физического опыта. Однако, такая ситуация, как мы видели ранее, вполне естественна при построении вероятностных моделей. При всем этом, во многих практических ситуациях использование схемы Бернулли оказывается вполне оправданным.

Приведем следующий поучительный *пример*. Американский ученый Уэлдон провел 26 306 серий испытаний по 12 бросаний одной и той же игральной кости в каждой серии, вычисляя частоту появления события («успеха»), состоящего в выпадении на кости 5 или 6 очков. Результаты его опытов приведены в таблице 6.1.

Таблица 6.1.

k	$b(k;\ 12\ \ 1\ 3)$	Наблюденная частота	$b(k;\ 12,\ 0{,}3377)$
0	0,007707	0,007033	0,007123
1	0,046244	0,043678	0,043584
2	0,127171	0,124116	0,122225
3	0,211952	0,208127	0,207736
4	0,238446	0,232418	0,238324
5	0,190757	0,197445	0,194429
6	0,111275	0,116589	0,115660
7	0,047689	0,050597	0,050549
8	0,014903	0,015320	0,016109
9	0,003312	0,003991	0,003650
10	0,000497	0,000532	0,000558
11	0,000045	0,000152	0,000052
12	0,000002	0,000000	0,000002

Если кость считать правильной, то вероятность «успеха» должна быть равна

⅓. Соответствующие теоретические значения функции $b(k;12,1/3)$ даны во второй колонке. Эксперимент показал, однако, довольно существенное отличие от теоретических значений при p = ⅓, но хорошее согласование с теоретическими значениями функции $b(k;12,0.3377)$ для $p=0.3377$. Этот результат естественно интерпретировать в том смысле, что игральная кость, использованная в эксперименте, *не является правильной*.

Это замечание имеет весьма важные практические приложения в вопросах, связанных с контролем за выполнением определенных нормативов (например, в производстве). В связи с этим, рассмотрим следующий пример.

Задача о снабжении энергией. Допустим, что n рабочих время от времени используют электрическую энергию. Чтобы получить грубое представление об ожидаемой нагрузке, представим себе, что в любой момент времени каждому рабочему с одной и той же вероятностью p может потребоваться единица энергии. Если они работают независимо, то вероятность того, что энергия потребуется одновременно k рабочим, будет равна $b(k;n,p)$ — здесь «испытанием» является проверка факта использования энергии в данный момент j-м рабочим $(j=1,2,...,n)$, а «успехом» является положительный результат проверки. Так, если один рабочий потребляет энергию в среднем 12 минут в течение часа,[41] следует положить $p=12/60=0.2$. В этом случае вероятность того, что не менее 7 из 10 рабочих будут одновременно использовать электрическую энергию, равна

$$b(7;10,0.2)+b(8;10,0.2)+b(9;10,0.2)+b(10;10,0.2)\approx 0.000864.$$

Другими словами, если снабжение рассчитано на 6 единиц энергии, то вероятность перегрузки равна 0.000864. Это означает, что одна перегрузка приходится в среднем на $1/0.000864\approx 1157$ минут, т.е. примерно на 12 часов рабочего времени. Поэтому, если перегрузки наблюдаются чаще, то это должно явиться сигналом для усиленного контроля над производственным циклом.

Следующий пример имеет несколько иной характер. При бросании двух *правильных* игральных костей, вероятность появления 12 очков равна, очевидно, $1/6^2\approx 0.0278$, т.е. в среднем одно появление за 36 бросаний. Если в казино за игорным столом в процессе игры эта пропорция существенно нарушается, то это означает либо тот факт, что кости дефектны, и их надо заменить, либо что игра идет нечестно. В любом случае, возникает основание для более тщательного наблюдения за игрой на данном игорном столе.

6.2. Обобщенная схема Бернулли.

Предположим, как и выше, что проводится серия из n независимых между собой испытаний. Однако, в отличие от предыдущего, мы предположим, что результатом каждого испытания может быть одно и

[41] Эта величина может определяться, например, производственным циклом или технологией производства.

только одно из k попарно несовместимых событий $A_1, A_2, ..., A_k$, причем вероятности появления каждого из этих событий в каждом отдельном испытании постоянны и равны соответственно

$$p_1, p_2, ..., p_k; \quad p_j > 0; \quad p_1 + p_2 + ... + p_k = 1.$$

Найдем вероятность того, что в результате n испытаний событие A_1 появится ровно m_1 раз, событие A_2 — ровно m_2 раза и т.д., событие A_k появится ровно m_k раз. Очевидно, $m_j \geq 0$, и должно выполняться равенство $m_1 + m_2 + ... + m_k = n$.

Прежде всего, отметим, что рассуждения предыдущего пункта приводят нас к выводу о том, что вероятность каждой допустимой комбинации будет равна $p_1^{m_1} p_2^{m_2} ... p_k^{m_k}$. С другой стороны, число допустимых комбинаций равно числу способов, которыми можно n элементов разбить на k групп по $m_1, m_2, ..., m_k$ элементов соответственно. Это число, согласно теореме 5.5, равно $\dfrac{n!}{m_1! m_2! ... m_k!}$.

Таким образом, искомая вероятность того, что в результате n независимых испытаний событие A_1 появится ровно m_1 раз, событие A_2 — ровно m_2 раза и т.д., событие A_k появится ровно m_k раз, будет равна

$$(6.2) \qquad \begin{cases} P_n(m_1, m_2, ..., m_k) = \dfrac{n!}{m_1! m_2! ... m_k!} p_1^{m_1} p_2^{m_2} ... p_k^{m_k}; \\ p_j > 0; \quad p_1 + p_2 + ... + p_k = 1; \\ m_j \geq 0; \quad m_1 + m_2 + ... + m_k = n. \end{cases}$$

Это и есть основная формула обобщенной схемы Бернулли.

Пример. Шахматист при игре с данным противником каждую партию выигрывает, проигрывает или сводит вничью соответственно с вероятностями $p_1 = 0.4, \quad p_2 = 0.35, \quad p_3 = 0.25$. Какова вероятность того, что в матче из 12 партий у данного шахматиста будет 5 побед, 4 поражения и 3 ничьи?

Решение. Мы в точности находимся в ситуации обобщенной схемы Бернулли с n = 12. Подставляя значения из данных задачи в формулу (6.2), получим: $P_{12}(5, 4, 3) = \dfrac{12!}{5!4!3!}(0.4)^5 (0.35)^4 (0.25)^3 \approx 0.067$.

6.3. Некоторые следствия.

Возвратимся к классической схеме Бернулли п. 6.1 и поставим следующую задачу. Пусть целые числа a, b таковы, что $0 \leq a < b \leq n$. Чему равна вероятность того, что в результате n независимых испытаний Бернулли число «успехов» будет заключено между числами a и b? Ответ на этот вопрос дается легко, поскольку допустимые комбинации для различных чисел «успехов» несовместимы. Соответствующая вероятность, очевидно, равна

$$P_n(a,b)=\sum_{k=a}^{b}C_n^k p^k q^{n-k}= \tag{6.3}$$
$$=C_n^a p^a q^{n-a}+C_n^{a+1}p^{a+1}q^{n-a-1}+C_n^{a+2}p^{a+2}q^{n-a-2}+\ldots+C_n^b p^b q^{n-b}.$$

Замечание. Для обозначения вероятности числа успехов в n испытаниях Бернулли используются различные обозначения, в зависимости от контекста рассматриваемых задач. Так, через $P_n(k<m)$ часто обозначается вероятность того, что в результате n испытаний число k успехов будет *меньше*, чем m; через $P_n(m_1\le k<m_2)$ обозначается вероятность того, в результате n испытаний число k успехов будет *больше либо равно* m_1, но *меньше* m_2; вместо обозначения $P_n(a,b)$ может использоваться обозначение $P_n(a\le k\le b)$ и т.п. Как правило, проблем с однозначным пониманием смысла подобных обозначений в контексте той или иной конкретной задачи не возникает.

Наиболее вероятное число успехов. Вычислим теперь значение числа $m=m_0$, при котором функция $b(m;n,p)$ достигает своего наибольшего значения. В этом случае число m_0 называют *наиболее вероятным числом успехов* (*в n испытаниях*).

Напомним, что функция $b(m;n,p)$ определяется как вероятность m «успехов» в n испытаниях Бернулли с вероятность «успеха» p, и вычисляется по формуле (6.1).

Рассмотрим величину

$$\frac{b(m;n,p)}{b(m-1;n,p)}=\frac{(n-m+1)p}{mq}=1+\frac{(n+1)p-m}{mq},$$

где учтено, что $q=1-p$. Отсюда видно, что функция $b(m;n,p)$ возрастает по m при $m<(n+1)p$ и убывает при $m>(n+1)p$. Имея в виду, что m_0 должно быть неотрицательным целым числом, получаем, что наиболее вероятное число «успехов» m_0 есть (единственное) неотрицательное целое число, удовлетворяющее неравенству

$$(n+1)p-1<m_0\le(n+1)p. \tag{6.4}$$

Рассмотрим теперь несколько *примеров*.

1. *Задача де Мере*. Сколько раз нужно бросить пару игральных костей, чтобы с вероятностью более ½ ожидать сумму очков, равную 12, хотя бы один раз?

Решение. Вероятность p «успеха», т.е. выпадения 12 очков, при каждом бросании одинакова и равна $p = 1/36$. Пусть n – искомое число бросаний, k – число «успехов». Тогда $P_n(k \geq 1) = 1 - P_n(k = 0)$. Но

$$P_n(k = 0) = C_n^0 p^0 (1-p)^n = \left(35/36\right)^n \approx (0.972)^n .$$

Таким образом, требуемое значение n находится из неравенства

$$(0.972)^n \leq 0.5 .$$

Решая это неравенство, получаем: $n \geq 25$.

2. Трое стрелков при стрельбе по мишени попадают в нее при одном выстреле с вероятностями 0.2, 0.3 и 0.4 соответственно. Кому из троих стрелять по мишени, определяют с помощью шести подбрасываний монеты, причем если гербов выпадет больше, чем решек, то стреляет первый стрелок, если гербов выпадет меньше, чем решек, то стреляет второй стрелок, в противном случае ─ третий стрелок. Стреляющий производит 3 выстрела. Определить вероятность того, что две пули попадут в цель.

Решение. Пусть A – событие, состоящее в том, что в мишень попадут две пули. Обозначим через B_1, B_2, B_3 ─ события, состоящие в том, что стреляет первый, второй и третий стрелок соответственно. Так как события B_1, B_2, B_3 образуют полную группу событий, то по формуле полной вероятности (2.6) имеем:

(*) $$P(A) = P(A/B_1)P(B_1) + P(A/B_2)P(B_2) + P(A/B_3)P(B_3) .$$

Вычислим по отдельности вероятности, входящие в формулу (*). Начнем с вероятностей $P(B_j)$. Поскольку предоставление права стрельбы тому или иному стрелку зависит от результатов последовательности независимых испытаний – шести подбрасываний монеты, то соответствующие вероятности должны вычисляться в соответствии со схемой Бернулли. Именно, пусть «успех» – это выпадение герба; тогда в соответствии с условиями задачи:

$$P(B_1) = P_6(k \geq 4) = C_6^4(0.5)^6 + C_6^5(0.5)^6 + C_6^6(0.5)^6 = 11/32 ;$$

$$P(B_2) = P_6(k \leq 2) = C_6^0(0.5)^6 + C_6^1(0.5)^6 + C_6^2(0.5)^6 = 11/32 ;$$

$$P(B_3) = P_6(k = 3) = C_6^3(0.5)^6 = 5/16 .$$

Далее, рассматривая стрельбу по мишени как последовательность трех независимых испытаний и считая «успехом» попадание в цель, получим:

$$P(A/B_1) = b(2; 3, 0.2) = C_3^2(0.2)^2 \cdot 0.8 = 0.096 ;$$

$$P(A/B_2) = b(2; 3, 0.3) = C_3^2(0.3)^2 \cdot 0.7 = 0.189 ;$$

$$P(A/B_3) = b(2; 3, 0.4) = C_3^2(0.4)^2 \cdot 0.6 = 0.288 .$$

По формуле полной вероятности (*) получаем окончательно

$$P(A) = 0.188 .$$

3. Каждый из $n=50$ приглашенных приходит на собрание с вероятностью $p=0.7$. Найти наиболее вероятное число m_0 людей, которые придут на собрание, и соответствующую вероятность $P_n(m_0)$.

Решение. Поскольку $P_{50}(m_0)=C_{50}^{m_0}(0.7)^{m_0}(0.3)^{50-m_0}$, то задача состоит в отыскании неотрицательного целого числа $m_0 \le 50$, доставляющего максимум функции $P_{50}(m_0)$. Мы видели выше, что такое число дается формулой (6.4). В нашем случае $n=50,\, p=0.7$. Подставляя эти значения в (6.4), получим: $51\cdot 0.7-1<m_0\le 51\cdot 0.7$, то есть $34.7<m_0<35.7$, и $m_0=35$. Вероятность $P_{50}(35)=C_{50}^{35}(0.7)^{35}(0.3)^{15}\approx 0.123$.

6.4. Формула Пуассона.

Формулы (6.1) и (6.3) дают точные значения вероятностей, связанных со схемой независимых испытаний Бернулли. Однако, вычисления по этим формулам, особенно при больших значениях *n* и *m*, весьма затруднительны. Представляет большой практический интерес получение достаточно простых приближенных формул для вычисления соответствующих вероятностей. Впервые подобную формулу вывел в 1837 году французский математик и физик Симон Пуассон (1781-1840). Ниже дается формулировка результата Пуассона.

Рассмотрим схему независимых испытаний Бернулли, в которой число испытаний *n* «относительно велико», вероятность «успеха» *p* «относительно мала», а произведение $\lambda=np$ «не мало и не велико».[42] При этих условиях справедлива формула

$$b(k;n,p)\approx\frac{\lambda^k}{k!}e^{-\lambda},\quad \lambda=np. \tag{6.5}$$

Это – *знаменитое пуассоновское приближение для биномиального распределения*. Доказательство формулы (6.5) будет дано в дополнении к настоящему параграфу.

Функция, стоящая в правой части формулы (6.5), называется *распределением Пуассона*:

$$p(k,\lambda)=\frac{\lambda^k}{k!}e^{-\lambda};\quad k=0,1,2,\ldots;\quad \lambda>0. \tag{6.6}$$

При таком обозначении $p(k,\lambda)$ будет приближенным выражением для вероятности $b(k;n,\lambda/n)$, когда n «достаточно велико».

Прежде, чем обсуждать формулу (6.5), приведем весьма показательные примеры ее использования.

В таблице 6.2 представлены значения биномиального распределения и значения распределения Пуассона при $n=100$, $p=0.01$, $\lambda=1$. Как мы

[42] Точный смысл взятых в кавычки терминов будет объяснен ниже, в частности, в § 6д.

видим, точность приближенной формулы достаточно высока.

Чем больше n, тем выше точность формулы Пуассона. Это наглядно представляет следующий пример. Вычислим вероятность p_k того, что в обществе из 500 человек ровно k человек родились в один и тот же конкретный день года. Если эти 500 человек выбраны наугад, то можно применить схему Бернулли из $n = 500$ испытаний с вероятностью «успеха» $p = 1/365$. Вычисления по точной формуле (6.1) и приближенной формуле (6.5) при $\lambda = 500/365 \approx 1{,}3699$ представлены в таблице 6.3. Как мы видим, ошибка лишь в четвертом десятичном знаке, что вполне приемлемо для практики.

Таблица 6.2.

k	$b(k;\ 100,\ 1/100)$	$p(k;\ 1)$
0	0,366032	0,367879
1	0,369730	0,367879
2	0,184865	0,183940
3	0,060999	0,061313
4	0,014942	0,015328
5	0,002898	0,003066
6	0,000463	0,000511
7	0,000063	0,000073
8	0,000007	0,000009
9	0,000001	0,000001

Таблица 6.3.

k	$b(k; 500, 1/365)$	$p(k, \lambda)$
0	0,2537	0,2541
1	0,3484	0,3481
2	0,2388	0,2385
3	0,1089	0,1089
4	0,0372	0,0373
5	0,0101	0,0102
6	0,0023	0,0023

Рассмотрим следующий типичный *пример* на применение формулы Пуассона.

Пусть известно, что вероятность «сбоя» в работе телефонной станции при каждом вызове равна 0,002. Поступило 1000 вызовов. Определить вероятность того, что при этом произойдет 7 «сбоев».

Решение. Естественно предположить, что в обычных условиях вызовы, поступающие на телефонную станцию – независимы друг от друга. Будем считать «успехом» в испытании – вызове – сбой телефонной станции. Вероятность сбоя ($p = 0,002$) можно считать «достаточно малой» величиной, а число вызовов ($n = 1000$) – «достаточно большим». Таким образом, мы находимся в условиях теоремы Пуассона. Для параметра λ получаем значение $\lambda = 0,002 \cdot 1000 = 2$. Поэтому, по формуле Пуассона, получаем для интересующей нас вероятности числа сбоев величину

$$p(7,2) = \frac{2^7}{7!} e^{-2} \approx 0,0034 .$$

Обсудим теперь *пределы применимости формулы Пуассона*. При использовании любой приближенной формулы вопрос о пределах ее применимости возникает естественным образом. При этом мы встречаемся с двумя аспектами проблемы. Во-первых, закономерен вопрос о том, в каких реальных условиях применим закон Пуассона? Опыт показывает, что простое распределение Пуассона обладает сравнительно универсальной применимостью. Вообще, с точки зрения применений, математические теоремы бывают хорошими и плохими в следующем смысле: хорошие теоремы продолжают действовать, если даже нарушить их условия, а плохие сразу перестают быть верными при нарушении условий их вывода. Теорема Пуассона (6.5) является в этом смысле хорошей и даже превосходной. Именно, закон Пуассона продолжает действовать даже тогда, когда условия схемы Бернулли нарушаются (т.е. можно допускать переменную вероятность успеха и даже не слишком сильную зависимость результатов отдельных испытаний).[43] Можно даже утверждать, что распределение Пуассона обладает сравнительно универсальной применимостью. Это надо понимать в том смысле, что если экспериментальные данные показывают, что закон Пуассона неприменим, в то время как, сообразно со здравым смыслом, он должен был бы действовать, то естественнее подвергнуть сомнению статистическую устойчивость наших данных, чем искать какой-то другой закон распределения. Иными словами, *распределение Пуассона представляет собой очень удачную математическую формулировку одного из универсальных (в рамках применимости теории вероятностей) законов природы.*

Во-вторых, возникает вопрос о порядках величин тех параметров, которые входят в формулу Пуассона, и для которых выше мы использовали расплывчатые термины «относительно велико», «относительно мало», «не мало и не велико». Опять же, разъясняющие ответы дает практика

[43] Естественно, этими особенностями распределения Пуассона не следует злоупотреблять. Например, закон Пуассона заведомо нарушается в ситуациях сильной зависимости результатов отдельных испытаний.

применения формулы (6.5). Оказывается, что формула Пуассона достаточно точна для практического применения, если число испытаний n имеет порядок нескольких десятков (лучше – сотен), а величина параметра $\lambda = np$ лежит в пределах от 0 до 10.

Для иллюстрации применения формулы Пуассона, рассмотрим еще один *пример*.

Пусть известно, что на выпечку 1000 сладких булочек с изюмом полагается 10 000 изюмин. Требуется найти распределения числа изюмин в какой-то случайным образом выбранной булочке.

Решение. Последовательность независимых испытаний мы формируем следующим образом. Всего будет $n = 10\,000$ испытаний (по числу изюмин), а именно: испытание с номером k будет состоять в том, что мы определяем, попала ли изюмина с номером k в нашу случайно выбранную булочку.[44] Тогда, поскольку всего булочек 1000, вероятность того, что k-я изюмина попала именно в нашу булочку, есть $p = 1/1000$ (при условии достаточно хорошего перемешивания теста при приготовлении булочек). Применяем теперь распределение Пуассона с параметром $\lambda = np = 10\,000 \cdot 1/1000 = 10$. Получим:

$$P_{10000}(k) \approx p(k,10) = \frac{10^k}{k!} e^{-10}.$$

В частности, вероятность того, что нам достанется булочка вовсе без изюма $(k = 0)$, равна $e^{-10} \approx 0{,}5 \cdot 10^{-4}$. Наиболее вероятное число изюмин будет, согласно формуле (6.4), равно 10. Соответствующая вероятность

$$P_{10000}(10) \approx \frac{10^{10}}{10!} e^{-10} \approx 0{,}125.$$

Пример с булочками и изюминами, несмотря на его приземленную формулировку, носит весьма общий характер. Так, вместо изюмин в булочках можно говорить, например, о числе бактерий в капле воды, взятой из хорошо перемешанного ведра. Другой пример. Предположим, что атомы радиоактивного вещества распадаются независимо друг от друга, причем в течение данного интервала времени распад данного атома происходит с вероятностью p, а всего к началу интервала времени было n атомов. Тогда для числа распадов в течение данного интервала времени получим, естественно, распределение Пуассона с параметром $\lambda = np$. Здесь k-м (мысленным) испытанием является проверка того, распался ли атом с номером k.

[44] Заметим, что на покупку булочки в магазине вполне можно смотреть как на случайный выбор.

§ 6д. ДОПОЛНЕНИЯ

В этом параграфе мы рассмотрим некоторые дальнейшие результаты, связанные с испытаниями Бернулли. Хотя эти результаты выходят за рамки «элементарных», они настолько важны, что без хотя бы краткого их рассмотрения даже элементарный курс теории вероятностей становится неполноценным.

Для чтения этого параграфа от читателя требуется владение основными понятиями высшей математики в пределах первого курса технического или педагогического ВУЗа.

6д.1. Доказательство теоремы Пуассона.

Дадим сначала точную формулировку теоремы Пуассона. Необходимость в этом возникает по следующей причине. В пункте 6.4 мы употребляли качественные термины «относительно велико», «относительно мало» и т.п. При этих весьма расплывчатых условиях была представлена *приближенная* формула (6.5). С точки зрения математики формулировки такого рода, очевидно, неприемлемы. В этом пункте мы внесем ясность в этот вопрос.

Выразить на математическом языке то утверждение, что некоторое число велико, а некоторое другое мало, можно только используя понятие предельного перехода. Следовательно, величины n и p, входящие в формулу (6.5), надо сделать переменными, стремящимися соответственно к ∞ и 0. Но в последовательности испытаний Бернулли вероятность успеха p должна быть постоянной. Поэтому приходится рассматривать не одну последовательность независимых испытаний, а *последовательность серий* независимых испытаний. Именно, в первой серии $(n=1)$ проводится всего одно испытание с вероятностью успеха p_1, число успехов в ней обозначим через m_1 (очевидно, либо $m_1=1$, либо $m_1=0$); далее, во второй серии $(n=2)$ проводится 2 испытания, каждое с вероятность успеха p_2, число успехов в ней обозначим через m_2 (очевидно, либо $m_2=1$, либо $m_2=2$, либо $m_2=0$), и т.д. В n-й серии проводится n испытаний с вероятностью успеха p_n, число успехов в ней обозначим через m_n. Теперь мы можем точно сформулировать теорему Пуассона.

Теорема 6д.1 (теорема Пуассона). Пусть $n\to\infty$, $p_n\to 0$, причем таким образом, что $np_n\to\lambda$, где λ — фиксированное положительное число. Тогда для любого фиксированного $k=0,1,2,...$ имеет место предельное соотношение

$$P_n(m_n=k)\xrightarrow[n\to\infty]{}\frac{\lambda^k}{k!}e^{-\lambda}.$$

(Здесь через $P_n(m_n=k)$ обозначена вероятность k успехов в n-й серии испытаний Бернулли: $P_n(m_n=k)=b(k;n,p_n)$ – см. п. 6.1).

Действительно, по формуле (6.1)

$$P_n(m_n=k)=b(k;n,\,p_n)=\frac{n!}{k!(n-k)!}(p_n)^k(1-p_n)^{n-k}.$$

Поскольку при $n\to\infty$, $np_n\to\lambda$, то можно написать:

$$p_n=\frac{\lambda}{n}+\frac{\alpha_n}{n},\quad \alpha_n\to 0 \text{ при } n\to\infty.$$

Далее:

(*) $$P_n(m_n=k)=\frac{n!}{k!(n-k)!}(p_n)^k(1-p_n)^{n-k}=$$

$$=\frac{n(n-1)...(n-k+1)}{k!}\left(\frac{\lambda}{n}+\frac{\alpha_n}{n}\right)^k\left(1-\frac{\lambda}{n}-\frac{\alpha_n}{n}\right)^{n-k}.$$

Вычислим отдельно:

(**) $$\left(\frac{\lambda}{n}+\frac{\alpha_n}{n}\right)^k=\frac{\lambda^k}{n^k}\left(1+\frac{\alpha_n}{\lambda}\right)^k=\frac{\lambda^k}{n^k}\beta_n,$$

где

$$\beta_n=\left(1+\frac{\alpha_n}{\lambda}\right)^k\xrightarrow[n\to\infty]{}1,\text{ так как } \alpha_n\to 0.$$

(+) $$\left(1-\frac{\lambda}{n}-\frac{\alpha_n}{n}\right)^{n-k}=\left(1-\frac{\lambda}{n}-\frac{\alpha_n}{n}\right)^{-k}\left(1-\frac{\lambda}{n}-\frac{\alpha_n}{n}\right)^{n}\xrightarrow[n\to\infty]{}e^{-\lambda}.$$

Действительно,

$$\left(1-\frac{\lambda}{n}-\frac{\alpha_n}{n}\right)^{-k}\xrightarrow[n\to\infty]{}1,$$

$$\left(1-\frac{\lambda}{n}-\frac{\alpha_n}{n}\right)^{n}=\exp\left[n\ln(1-\frac{\lambda+\alpha_n}{n})\right].$$

Далее, известно, что если γ_n – бесконечно малая величина, то $\ln(1+\gamma_n)\sim\gamma_n$. Поэтому

$$\ln\left(1-\frac{\lambda+\alpha_n}{n}\right)\sim-\frac{\lambda+\alpha_n}{n}\quad\text{при}\quad n\to\infty$$

и

$$\lim_{n\to\infty} n\ln\left(1-\frac{\lambda+\alpha_n}{n}\right)=\lim_{n\to\infty} n\left(-\frac{\lambda+\alpha_n}{n}\right)=\lim_{n\to\infty}(-\lambda-\alpha_n)=-\lambda\,.$$

Отсюда следует, что

$$\lim_{n\to\infty}\left(1-\frac{\lambda}{n}-\frac{\alpha_n}{n}\right)^n=\exp\left[\lim_{n\to\infty} n\ln(1-\frac{\lambda+\alpha_n}{n})\right]=e^{-\lambda}\,,$$

что и обосновывает равенство (+).

Наконец,

$$(++)\quad \frac{n(n-1)...(n-k+1)}{k!n^k}=\frac{1}{k!}\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)...\left(1-\frac{k-1}{n}\right)\xrightarrow[n\to\infty]{}\frac{1}{k!}\,.$$

Подставляя теперь (**), (+) и (++) в (*), получим окончательно:

$$\lim_{n\to\infty} P_n(m_n=k)=\frac{\lambda^k}{k!}e^{-\lambda}\,,$$

что и требовалось. Теорема доказана.

6д.2. Теорема Муавра-Лапласа[45] и ее приложения.

Как мы видели в §6, формула Пуассона, при всей ее важности, имеет существенное ограничение, связанное с требованием малости вероятности успеха p, т.е. она применима только для случая маловероятных событий. На практике такое ограничение часто бывает слишком жестким. Естественно поставить вопрос о выводе достаточно простого приближенного выражения для вероятности $P_n(k)=b(k;n,p)$ без ограничительных условий на величину вероятности успеха p. Этот вопрос решает так называемая теорема Муавра-Лапласа, к изложению которой мы и переходим. Доказательство этой теоремы весьма сложно и требует от читателя достаточно высокой математической культуры, поэтому мы приводить его не будем. Однако сам результат настолько важен, что его знание необходимо любому, кто так или иначе связан с использованием теории вероятностей в практической деятельности. Кроме того, сама по себе теорема Муавра-Лапласа является типичным представителем одного из важнейших для практики разделов теории вероятностей, который изучает так называемые *предельные теоремы* теории вероятностей.

Прежде, чем формулировать основную теорему этого пункта, нам необходимо ввести несколько *важных понятий и определений*.

Определение 6д.1. Функция

$$\varphi(x)=\frac{1}{\sqrt{2\pi}}e^{-x^2/2},\quad -\infty<x<\infty$$

[45] Пьер-Симон Лаплас (1749-1827) – французский астроном, математик и физик, внесший выдающийся вклад в развитие этих наук. Лапласа можно считать основоположником аналитической теории вероятностей. Им доказаны первые предельные теоремы теории вероятностей, ставшие классическими и носящие ныне его имя. Развил теорию ошибок и метод наименьших квадратов. Иностранный почетный член Петербургской Императорской Академии Наук с 1802 года.

называется *плотностью вероятности нормального распределения*, или просто *плотностью нормального распределения*.

График этой функции представлен на рисунке 6д.1.

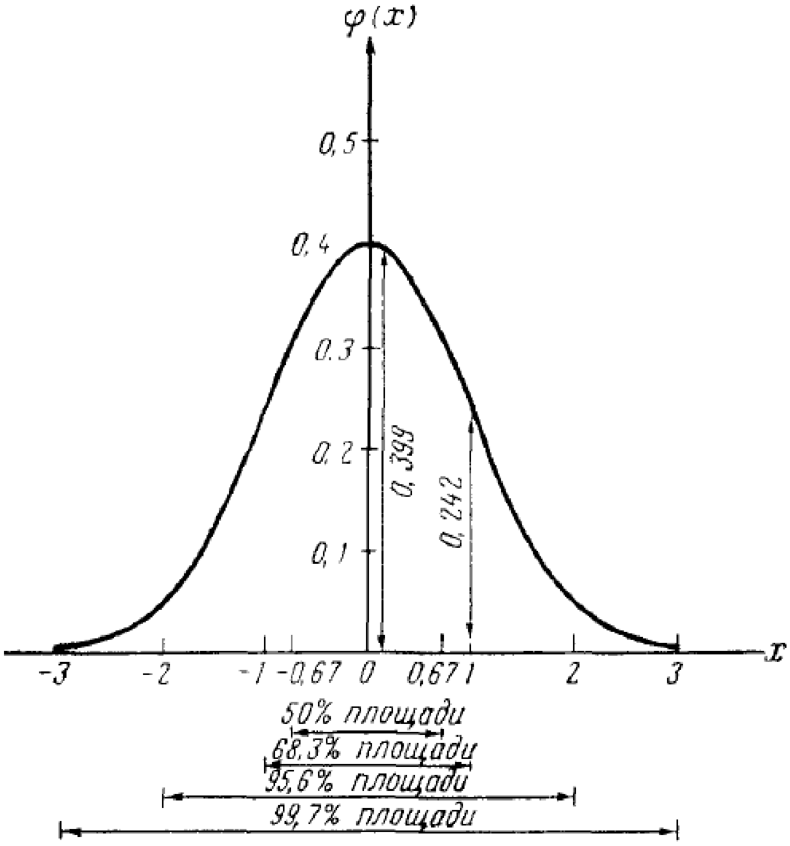


Рис. 6д.1. Плотность нормального распределения.

Определение 6д.2. Функция

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2} dt = \int_{-\infty}^{x} \varphi(t) dt, \quad -\infty \le x \le \infty$$

называется *нормальной функцией распределения,* или *функцией Лапласа*.

График этой функции представлен на рисунке 6д.2.

Легко установить следующие *свойства функции* $\Phi(x)$.

1. $\Phi(\infty) \equiv \lim_{x\to\infty} \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-t^2/2} dt = 1,$

$$\Phi(-\infty) \equiv \lim_{x\to-\infty} \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{-\infty} e^{-t^2/2} dt = 0.$$

2. Функция $\Phi(x)$ непрерывна и строго монотонно возрастает на промежутке $(-\infty, \infty)$.

3. $\frac{d\Phi(x)}{dx} = \varphi(x)$.

4. $\Phi(-x) = 1 - \Phi(x)$.

5. $\int_a^b \varphi(t)dt = \Phi(b) - \Phi(a)$ (ср. с формулой Ньютона-Лейбница).

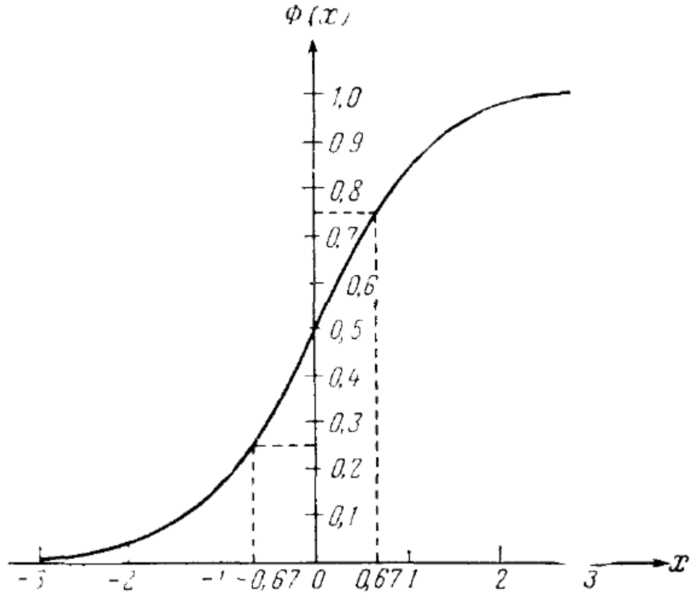


Рис. 6д2. Нормальная функция распределения.

Значения функций $\varphi(t)$ и $\Phi(t)$ представлены в таблице 6д.1.

Замечание. В литературе нередко под функцией Лапласа понимают функцию

$$F(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-t^2/2} dt, \quad -\infty \le x \le \infty.$$

Функции $\Phi(x)$ и $F(x)$ связаны очевидным равенством $\Phi(x) = \frac{1}{2} + F(x)$.

Заметим, что $F(-x) = -F(x)$. Это более простое равенство, чем аналогичное равенство для $\Phi(x)$ (свойство 4), часто и является причиной предпочтения использования $F(x)$ вместо $\Phi(x)$. При использовании таблиц функции Лапласа надо уточнять, какая именно из указанных функций представлена в таблице.

Укажем без доказательства следующее важное неравенство.

Для любого $x > 0$ справедливо неравенство:

$$\left(\frac{1}{x} - \frac{1}{x^3}\right)\varphi(x) < 1 - \Phi(x) < \frac{1}{x}\varphi(x).$$

Отсюда, в частности, следует, что при $x \to \infty$

$$1-\Phi(x) \sim \frac{1}{x}\varphi(x) = \frac{1}{\sqrt{2\pi}\cdot x}\cdot e^{-x^2/2}.$$

Таблица 6д.1.

Нормальное распределение

t	$\varphi(t)$	$\Phi(t)$	t	$\varphi(t)$	$\Phi(t)$
0,0	0,398942	0,500000	2,3	0,028327	0,989276
0,1	0,396952	0,539828	2,4	0,022395	0,991802
0,2	0,391043	0,579260	2,5	0,017528	0,993790
0,3	0,381388	0,617911	2,6	0,013583	0,995339
0,4	0,368270	0,655422	2,7	0,010421	0,996533
0,5	0,352065	0,691462	2,8	0,007915	0,997445
0,6	0,333225	0,725747	2,9	0,005953	0,998134
0,7	0,312254	0,758036	3,0	0,004432	0,998650
0,8	0,289692	0,788145	3,1	0,003267	0,999032
0,9	0,266085	0,815940	3,2	0,002384	0,999313
1,0	0,241971	0,841345	3,3	0,001723	0,999517
1,1	0,217852	0,864334	3,4	0,001232	0,999663
1,2	0,194186	0,884930	3,5	0,000873	0,999767
1,3	0,171369	0,903200	3,6	0,000612	0,999841
1,4	0,149727	0,919243	3,7	0,000425	0,999892
1,5	0,129518	0,933193	3,8	0,000292	0,999928
1,6	0,110921	0,945201	3,9	0,000199	0,999952
1,7	0,094049	0,955435	4,0	0,000134	0,999968
1,8	0,078950	0,964070	4,1	0,000089	0,999979
1,9	0,065616	0,971283	4,2	0,000059	0,999987
2,0	0,053991	0,977250	4,3	0,000039	0,999991
2,1	0,043984	0,982136	4,4	0,000025	0,999995
2,2	0,035475	0,986097	4,5	0,000016	0,999997

Сформулируем теперь основную теорему этого параграфа.

Терема 6д.1 (терема Муавра-Лапласа). Пусть m_n обозначает число успехов в n испытаниях Бернулли с вероятностью успеха в отдельном испытании p, причем p не равно 0 или 1. Пусть, как обычно, $q=1-p$. Тогда

(6д.1) $$P\left\{\frac{m_n-np}{\sqrt{npq}}<x\right\}\xrightarrow[n\to\infty]{}\Phi(x) \text{ равномерно по } x\in(-\infty,\infty).$$

Замечания.

1. В формуле (6д.1) слева стоит вероятность события, состоящего в том, что для данного числа успехов m_n выполнено неравенство, стоящее в фигурных скобках. Это неравенство можно еще переписать в виде

$$m_n < np + x\cdot\sqrt{npq}\ .$$

Кроме того, знак неравенства «<» может быть заменен знаком «≤».

2. Утверждение теоремы Муавра-Лапласа можно сформулировать еще следующим образом. Пусть

$$f_n(x) = P\left\{\frac{m_n - np}{\sqrt{npq}} < x\right\} - \Phi(x).$$

Тогда существует *не зависящая от x* бесконечно малая величина β_n (т.е. $\beta_n \to 0$ при $n \to \infty$) такая, что $\left|f_n(x)\right| \le \beta_n$ для всех $x \in (-\infty, \infty)$.

Следствием последнего замечания является приближенное равенство

(6д.2) $$P\left\{\frac{m_n - np}{\sqrt{npq}} < x\right\} \approx \Phi(x),$$

причем ошибка приближения мала равномерно по $x \in (-\infty, \infty)$ при $n \gg 1$.

Символ $n \gg 1$ (читается: «*n* много больше единицы») означает, что *n* – «достаточно большое число»; на практике это соответствует величине *n* порядка нескольких десятков и более. Подчеркнем еще раз, что, в отличие от приближенной формулы Пуассона (6.5), в формуле (6д.2) *на величину p никаких ограничений не накладывается*.

Разберем некоторые следствия соотношения (6д.2), имеющие важное прикладное значение. Для краткости записи введем обозначение

$$\gamma_n = \frac{m_n - np}{\sqrt{npq}}.$$

Тогда соотношение (6д.2) запишется в виде

(6д.3) $$P\left\{\gamma_n < x\right\} \approx \Phi(x).$$

Найдем сначала вероятность $P\left\{a < \gamma_n < b\right\}$ события $\left\{a < \gamma_n < b\right\}$. Прежде всего, отметим, что события $\left\{\gamma_n \le a\right\}$ и $\left\{a < \gamma_n < b\right\}$ несовместны, и их сумма есть событие $\left\{\gamma_n < b\right\}$. Поэтому для вероятностей имеем:

$$P\left\{\gamma_n < b\right\} = P\left\{\gamma_n \le a\right\} + P\left\{a < \gamma_n < b\right\},$$

откуда получаем:

$$P\left\{a < \gamma_n < b\right\} = P\left\{\gamma_n < b\right\} - P\left\{\gamma_n \le a\right\}.$$

Это равенство, с учетом (6д.3) и замечания 1 к теореме Муавра-Лапласа, дает:

(6д.4) $$P\left\{a < \gamma_n < b\right\} \approx \Phi(b) - \Phi(a),$$

причем в этой формуле любой (или оба) из знаков неравенства «<» может быть заменен знаком «≤».

Замечание. Возможность столь «вольного» обращения со знаками строгого и нестрогого неравенства в соотношениях типа (6д.1) – (6д.4) связана со следующим. Можно показать, что для $k = 1, 2, ...$

$$P\{\gamma_n \le x\} = \lim_{k\to+\infty} P\left\{\gamma_n < x + \frac{1}{k}\right\}.$$

Но при больших *n*:

$$P\left\{\gamma_n < x + \frac{1}{k}\right\} \approx \Phi(x + \frac{1}{k}).$$

Так как функция $\Phi(x)$ непрерывна, то $\Phi(x + \frac{1}{k}) \xrightarrow[k\to+\infty]{} \Phi(x)$. С другой стороны, как было отмечено выше, формула (6д.3) справедлива при больших *n равномерно по* $x \in (-\infty, \infty)$. Поэтому при $n \gg 1$

$$P\{\gamma_n \le x\} = \lim_{k\to+\infty} P\left\{\gamma_n < x + \frac{1}{k}\right\} \approx \lim_{k\to+\infty} \Phi(x + \frac{1}{k}) = \Phi(x).$$

Отсюда, в частности, следует, что

$$P\{\gamma_n = x\} \xrightarrow[n\to\infty]{} 0 \text{ равномерно по } x \in (-\infty, \infty).$$

Рассмотрим теперь следующую задачу. Предположим, что в результате некоторого эксперимента может появиться или не появиться событие *A*. Нас интересует вопрос о том, как экспериментально можно оценить вероятность события *A*.[46] Свяжем с этой задачей следующую схему Бернулли. Будем проводить последовательно серию из *n одинаковых* (т.е. проводимых в одинаковых условиях) экспериментов, причем так, что результаты этих экспериментов можно считать независимыми событиями. Появление в конкретном эксперименте события *A* будем считать «успехом» этого эксперимента. Таким образом, вероятность «успеха» - это то же, что и вероятность события *A*. Предположим, что в результате серии из *n* экспериментов событие *A*, т.е. «успех», наблюдалось в m_n случаях.

Обозначим через ν_n *частоту* появления события *A* в *n* экспериментах, т.е. величину

$$\nu_n = \frac{m_n}{n}.$$

Поставим следующий вопрос: сколько надо провести экспериментов, чтобы величина ν_n отличалась от вероятности *p* события *A* не более, чем на величину $\varepsilon > 0$? Теорема Муавра-Лапласа позволяет ответить на этот вопрос (естественно, в вероятностных терминах).

Преобразуем величину γ_n следующим образом:

[46] Предполагаем, конечно, что событие *A* статистически устойчиво.

$$\gamma_n = \frac{m_n - np}{\sqrt{npq}} = \frac{\sqrt{n}}{\sqrt{pq}} \cdot (\nu_n - p).$$

Пусть $\varepsilon > 0$. Чтобы использовать соотношение (6д.2), интересующее нас неравенство

$$\left|(\nu_n - p)\right| < \varepsilon \text{, или } -\varepsilon < \nu_n - p < \varepsilon$$

преобразуем следующим образом. Имеем:

$$-\varepsilon < \nu_n - p < \varepsilon \quad \Leftrightarrow \quad -\varepsilon \frac{\sqrt{n}}{\sqrt{pq}} < \gamma_n < \varepsilon \frac{\sqrt{n}}{\sqrt{pq}}.$$

Формула (6д.4) тогда дает:

$$P\left\{\left|(\nu_n - p)\right| < \varepsilon\right\} = P\left\{-\varepsilon \frac{\sqrt{n}}{\sqrt{pq}} < \gamma_n < \varepsilon \frac{\sqrt{n}}{\sqrt{pq}}\right\} \approx$$

$$\approx \Phi\left(\varepsilon \frac{\sqrt{n}}{\sqrt{pq}}\right) - \Phi\left(-\varepsilon \frac{\sqrt{n}}{\sqrt{pq}}\right) = 2\Phi\left(\varepsilon \frac{\sqrt{n}}{\sqrt{pq}}\right) - 1,$$

где использовано свойство 4 функции $\Phi(x)$: $\Phi(-x) = 1 - \Phi(x)$.

Так как для $0 < p < 1$ произведение $pq = p(1-p) \le 1/4$, то с учетом строгого возрастания функции $\Phi(x)$:

$$\Phi\left(\varepsilon \frac{\sqrt{n}}{\sqrt{pq}}\right) \ge \Phi\left(2\varepsilon\sqrt{n}\right).$$

Таким образом окончательно:

(6д.5) $$P\left\{\left|(\nu_n - p)\right| < \varepsilon\right\} \ge 2\Phi\left(2\varepsilon\sqrt{n}\right) - 1.$$

Последняя формула и решает поставленную нами выше задачу. Например, если мы хотим, чтобы с вероятностью не менее 0,95 экспериментально вычисленная частота ν_n появления события A отличалась от истинной вероятности $p = P(A)$ этого события не более, чем на 0,01, количество n требуемых экспериментов находится из уравнения

$$\Phi\left(2 \cdot 0{,}01\sqrt{n}\right) = \frac{1 + 0{,}95}{2} = 0{,}975.$$

По таблице 6д.1 находим $0{,}02\sqrt{n} \approx 2{,}0$, то есть $n \ge 10000$. Для $\varepsilon = 0{,}05$ получаем $n \ge 400$.

На практике более удобным бывает несколько иной подход к интерпретации формулы (6д.5). Перепишем ее в следующем эквивалентном виде.

$$P\{\nu_n - \varepsilon < p < \nu_n + \varepsilon\} \geq 2\Phi\left(2\varepsilon\sqrt{n}\right) - 1.$$

Назовем интервал $(\nu_n - \varepsilon, \nu_n + \varepsilon)$ *доверительным интервалом* для вероятности *p с коэффициентом доверия* α, если

$$P\{\nu_n - \varepsilon < p < \nu_n + \varepsilon\} \geq \alpha . \tag{6д.6}$$

Таким образом, коэффициент доверия α характеризует «надежность» соответствующей оценки вероятности *p* интересующего нас события.

Коэффициент доверия α связан с шириной ε доверительного интервала и числом *n* испытаний соотношением

$$2\Phi\left(2\varepsilon\sqrt{n}\right) = 1 + \alpha . \tag{6д.7}$$

Таким образом, если, например, мы ограничены числом испытаний, то формула (6д.6) показывает, что уменьшение доверительного интервала ведет к уменьшению коэффициента доверия и наоборот (т.е. требование большей близости частоты и вероятности ведет к уменьшению надежности этого результата[47]). С другой стороны, если нам необходимо выдержать соотношение между доверительным интервалом и коэффициентом доверия, мы можем оценить потребное для этого число экспериментов и т.д. Все это позволяет выбрать именно тот вариант проведения эксперимента, который наиболее подходит для того или иного конкретного случая. Так, если принять $n = 100$ и $\alpha = 0{,}95$, то из (6д.7) для ширины ε доверительного интервала получаем $\varepsilon \approx 0{,}1$.

6д.3. Последовательности зависимых испытаний. Цепи Маркова.

Выше мы подробно изучили последовательности *независимых* испытаний. В этом пункте мы рассмотрим один важный пример случая последовательности *зависимых* испытаний.

Будем производить ряд последовательных испытаний, в результате каждого из которых может наступить или не наступить некоторое событие *A*. Пусть в случае наступления события *A* при *n*-м по порядку испытании вероятность его наступления при $(n+1)$-м испытании равна *a*, а в случае его ненаступления при *n*-м испытании вероятность его наступления при $(n+1)$-м испытании равна *b*. Очевидно, $0 \leq a, b \leq 1$. Поставим следующую задачу: зная

[47] Важность представленного выше результата состоит в том, что это достаточно естественное качественное утверждение получает через формулу (6д.6) вполне определенную количественную интерпретацию.

вероятность p_1 его наступления при первом по порядку испытании, определить вероятность p_n его наступления при каждом n-м испытании.

Пусть A_n - событие, означающее наступление события A при n-м испытании. Тогда, поскольку событие A_n и противоположное ему событие $\overline{A_n}$ при каждом данном значении n образуют полную группу событий (см. §2), по формуле полной вероятности (2.6) получаем:

$$P(A_{n+1}) = P(A_n)P(A_{n+1}/A_n) + P(\overline{A_n})P(A_{n+1}/\overline{A_n}).$$

Согласно введенным выше обозначениям, $P(A_{n+1}) = p_{n+1}$, $P(A_n) = p_n$, $P(\overline{A_n}) = 1 - p_n$, $P(A_{n+1}/A_n) = a$, $P(A_{n+1}/\overline{A_n}) = b$. Поэтому

(*) $$p_{n+1} = ap_n + b(1 - p_n), \quad n = 1, 2, \ldots.$$

Найдем из уравнений (*)[48] величину вероятности p_n как функцию от величин $n\ (n \geq 2), a, b$ и p_1.

Заметим, прежде всего, что величина

$$p = \frac{b}{1 - a + b}$$

решает уравнение

(**) $$p = ap + b(1 - p),$$

в которое превращается уравнение (*) при $p_n \equiv p$ для всех n.

Для решения уравнения (*) положим $p_n = p + p_n^*$, где p – решение уравнения (**). Подставляя это в (*), получим для определения p_n^* рекуррентные соотношения

$$p + p_{n+1}^* = ap + ap_n^* + b(1 - p) - bp_n^*,$$

откуда, принимая во внимание (**):

$$p_{n+1}^* = p_n^*(a - b).$$

Последнее соотношение, очевидно, дает

$$p_{n+1}^* = p_1^*(a - b)^n.$$

Переходя от p_n^* к p_n, получим:

[48] Уравнения такого типа называются *рекуррентными*.

$$p_{n+1} - p = (p_1 - p)(a-b)^n,$$

или:

(+)
$$p_{n+1} = \frac{b}{1-a+b} + \left(p_1 - \frac{b}{1-a+b}\right)(a-b)^n,$$

что и решает нашу задачу.

Замечание. Последняя формула справедлива при $a \neq 1$, либо $b \neq 0$. В исключительном случае, когда одновременно $a = 1, b = 0$, из (*) получаем сразу $p_{n+1} = p_n = p_{n-1} = ... = p_1$. Этот тривиальный частный случай, очевидно, практически мало интересен.

Если $0 \leq a < 1$ и $0 < b \leq 1$, то общее решение (+) показывает, между прочим, при $n \to \infty$ решение имеет вполне определенный предел, не зависящий от p_1, а именно:

$$\lim_{n\to\infty} p_n = \frac{b}{1-a+b}.$$

Рассмотренный выше пример последовательности зависимых испытаний представляет собой частный случай так называемых *цепей Маркова*,[49] которые имеют большое прикладное значение.

[49] Андрей Андреевич Марков (1856-1922) – знаменитый русский математик, академик Петербургской Императорской Академии Наук с 1890 г.

§7. СЛУЧАЙНЫЕ ВЕЛИЧИНЫ

Понятие случайной величины считается, чуть ли не основным в теории вероятностей, поэтому обойти эту тему даже в элементарном курсе не представляется возможным, хотя математический аппарат, привлекаемый к изучению случайных величин, предъявляет достаточно высокие требования к уровню абстрактного мышления учащихся.

7.1. Основные понятия и определения.

Прежде, чем вводить формальные определения, попытаемся понять истоки возникновения понятия случайной величины. В качестве примера рассмотрим эксперимент по одновременному бросанию двух игральных костей. Мы уже знаем (п. 1.5), что при подсчете вероятностей выпадения того или иного суммарного числа очков, кости формально следует различать (даже, если чисто физически они неразличимы). Поэтому пространство элементарных событий данного эксперимента мы представляем таблицей 2.1, в которой каждая клетка представляет соответствующий элементарный исход. На практике, однако, нас интересует не само по себе пространство элементарных событий, а суммарные *числа очков* на костях, и соответствующие вероятности их выпадения. Составим соответствующую таблицу.

Таблица 7.1.

Число очков n_j	2	3	4	5	6	7	8	9	10	11	12
Вероятность p_j	1/36	1/18	1/12	1/9	5/36	1/6	5/36	1/9	1/12	1/18	1/36

Совершенно очевидно, что информация, содержащаяся в таблице, вполне достаточна для практических приложений.

Обратимся к примеру 2 п. 1.6 (случайный выстрел в круговую «мишень», разделенную на концентрические зоны - рис. 1.5). Элементарными событиями, т.е. возможными *исходами физического эксперимента* является попадание пули в конкретную *геометрическую точку* мишени. Нас, однако, интересует не конкретная геометрическая точка попадания, а *численное значение* r расстояния точки попадания от центра мишени, точнее, выполнение *численного* соотношения $r \in [r_{k-1}, r_k)$, где, напомним, через r_k, $k = 1, 2, 3, 4$ обозначены радиусы окружностей, делящих «мишень» на концентрические зоны. В п. 1.6 мы вычислили вероятности p_k, представляющие собой вероятности выполнения неравенств $r_{k-1} \le r < r_k$. Составим соответствующую таблицу.

Таблица 7.2.

Промежуток изменения r	$[0,1)$	$[1,2)$	$[2,3)$	$[3,4)$
Вероятности p_k	1/16	3/16	5/16	7/16

Как и выше, информация, содержащаяся в таблице, вполне достаточна для практических приложений

Таким образом, если нам задана таблица типа 7.1, 7.2, нас совершенно может не интересовать структура соответствующего пространства элементарных событий. Более того, часто мы принципиально не можем сколько-нибудь ясно описать пространство элементарных событий в конкретном эксперименте, и, следовательно, указать на уровне элементарных исходов, структуру того или иного события,[50] ибо можем не знать необходимые детали конкретного физического явления, которое мы наблюдаем с помощью эксперимента. В то же время *эффект* от реализации того или иного исхода мы наблюдать можем, например с помощью приборов, показания которых принимают то или иное значение в зависимости от того, произошло интересующее нас случайное событие или нет. На практике этого в большинстве случаев вполне достаточно. Иными словами, нас, в большинстве случаев, интересуют не случайные события сами по себе, а некоторые их *числовые характеристики*. Эти числовые характеристики часто гораздо проще измерить и описать диапазон их возможных изменений, чем пытаться разобраться (если это вообще возможно) в физической основе или сути явления, в результате которого может произойти или не произойти данное случайное событие.

Поясним сказанное следующим примером. Пусть бросается игральная кость и наблюдаемая нами числовая характеристика есть число выпавших очков. Введем пространство Ω на основании следующих соображений. Хорошо известно, что движение твердого тела вполне определяется, если в некоторый момент задать шесть параметров, определяющих его положение в пространстве (три координаты центра тяжести и три угла поворота подвижной системы координат относительно неподвижной), вместе со скоростями изменения этих параметров. Будем понимать под элементарным событием ω набор этих двенадцати чисел (измеренных в тот момент, когда мы выпускаем кость из рук), записанных с таким числом десятичных знаков, которого достаточно для определения, какой гранью кверху в конце концов остановиться кость. Тогда, зная ω, мы знаем и число выпавших очков, т.е. нужную нам числовую характеристику. Но совершенно очевидно, что

[50] Напомним, что событие есть подмножество пространства элементарных событий.

числовую характеристику – число выпавших очков – наблюдать легко, в то время как при регистрации ω возникают непреодолимые трудности. Кроме того, возможных значений ω очень много, а значений интересующей нас числовой характеристики всего шесть. Отсюда видно, насколько сильным может быть упрощение при переходе от исследования пространства элементарных событий к изучению свойств интересующих нас числовых характеристик, связанных со случайными событиями.

Таким образом, при исследовании случайных явлений, в большинстве практически важных случаев, представляют интерес величины, принимающие то или иное числовое значение в зависимости от реализации случайных событий, связанных с изучаемым случайным явлением, а не сами эти случайные события.

Величины такого рода носят название *случайных величин*. Значения случайной величины зависят от конкретной реализации случайного явления, поэтому в каждом данном эксперименте мы не можем заранее предсказать, какое конкретное значение примет случайная величина. Однако, как мы видели в разобранных выше примерах, имеет смысл говорить о *вероятностях событий, состоящих в том, что в результате эксперимента случайная величина примет то или иное конкретное значение, либо ее значение попадет в заданный интервал числовой оси*.

Поскольку, повторим, значения случайной величины зависят от конкретной реализации случайного явления, мы можем говорить о том, что ***случайная величина является функцией, заданной на пространстве элементарных событий, связанном с изучаемым случайным явлением***.

Все, сказанное выше, позволяет дать следующее формальное определение.

Определение 7.1. *Случайной величиной* ξ называется числовая функция, заданная на пространстве элементарных событий Ω:

$$\xi : \Omega \to (-\infty, \infty),$$

или

$$\xi = \xi(\omega), \quad \omega \in \Omega, \quad \xi \in (-\infty, \infty).$$

Замечания.

1. При изучении понятия случайной величины начинающий сталкивается с одной из характерных особенностей абстрактно-математического подхода. Именно, приходится отказываться от привычных представлений о том, что аргументом функции может быть только число, и приходится переключаться на методы анализа *функций нечислового (абстрактного) аргумента*.

2. Мы предположили, что случайная величина принимает вещественные значения. На самом деле, можно считать, что множество

значений случайной величины есть множество комплексных чисел, множество векторов (любой размерности), множество матриц и т.п., в зависимости от рассматриваемой задачи. Мы, однако, в нашем курсе ограничимся случаем множества вещественных чисел $\mathbf{R} = (-\infty, \infty)$.

3. Случайные величины традиционно обозначаются малыми греческими буквами $\xi, \eta, \varsigma, \ldots$, или заглавными латинскими буквами $X, Y, Z, \ldots$. Эта, традиция, однако, не является правилом.

4. Понятие случайной величины тесным образом связано с понятием случайного события, а именно: случайной величине $\xi(\omega)$ и промежутку Δ вещественной оси ставится в соответствие событие A, представляющее собой совокупность всех тех $\omega \in \Omega$, для которых величина $\xi(\omega) \in \Delta$. Записывается это следующим образом:

(7.1) $\qquad A = \{\omega : \xi(\omega) \in \Delta\}$, или просто $A = \{\xi(\omega) \in \Delta\}$.

Аналогично, событие

(7.2) $\qquad B = B(\xi, x) = \{\omega : \xi(\omega) = x\} = \{\xi(\omega) = x\}$

представляет собой совокупность всех тех $\omega \in \Omega$, для которых $\xi(\omega) = x$.

5. Для дальнейшего будет иметь большое значение множество всех значений, которые в принципе может принимать случайная величина ξ, т.е. множество всех тех вещественных чисел $x \in \mathbf{R}$, для которых существует хотя бы одно элементарное событие $\omega \in \Omega$ такое, что $\xi(\omega) = x$. Это множество всех возможных значений случайной величины ξ мы будем обозначать через $\xi(\Omega)$, или R_ξ. Формально это записывается следующим образом:

$$\xi(\Omega) = R_\xi = \{x \in \mathbf{R} : \exists \omega \in \Omega : \xi(\omega) = x\}.$$

Очевидно, $R_\xi \subseteq (-\infty, \infty)$.

Определение 7.2. Случайная величина ξ называется *дискретной*, если множество ее возможных значений конечно или счетно, т.е.

$$R_\xi = \{x_1, x_2, \ldots, x_n, \ldots\}, \quad x_j \in \mathbf{R},$$

и *непрерывной*, если множество ее значений сплошь заполняет некоторый промежуток или семейство промежутков, т.е.

$$R_\xi = \bigcup_{i=1}^{n} \Delta_i,$$

где Δ_i - непересекающиеся промежутки вещественной оси.[51]

Из рассмотренных в этом пункте примеров, первая случайная величина – дискретная, вторая – непрерывная.

Обращаем внимание читателя на то, что понятие дискретности или непрерывности случайной величины относится к свойству *области*

[51] Под *промежутком* мы понимаем здесь любое множество вида $(a,b), (a,b], [a,b), [a,b]$.

возможных значений случайной величины, а не к свойству «дискретности» или «непрерывности» пространства элементарных событий Ω, на котором определена случайная величина. Так, если во втором примере в качестве случайной величины мы будем рассматривать не расстояние точки попадания пули в мишень от центра мишени, а номер зоны попадания, то наша случайная величина станет дискретной, хотя пространство элементарных событий остается «непрерывным» - сплошной круг радиуса 4 ед.

В данном пособии мы будем рассматривать только дискретные случайные величины.

Пусть ξ – дискретная случайная величина (далее д.с.в.), возможные значения которой - конечный или счетный набор[52] *различных* вещественных чисел $x_1, x_2, ..., x_n, ...$. В случае счетного набора мы будем говорить о *последовательности* возможных значений случайной величины. Что касается пространства элементарных событий Ω, то, для упрощения изложения, мы всюду далее предполагаем его дискретным, т.е. состоящим из конечного или счетного числа элементов (см. §3).

Поскольку всякая д.с.в. связана со случайными событиями вида (7.2), то естественно говорить о вероятностях этих событий. Обозначим:

$$B_j = \left\{\omega : \xi(\omega) = x_j\right\} = \left\{\xi(\omega) = x_j\right\} = \left\{\xi = x_j\right\}, \tag{7.3}$$

$$p_j = P(B_j) = P\left\{\omega : \xi(\omega) = x_j\right\}. \tag{7.4}$$

Составим таблицу (ср. с табл.7.1):

$$\begin{pmatrix} \xi: & x_1 & x_2 & ... & x_n & ... \\ P: & p_1 & p_2 & ... & p_n & ... \end{pmatrix}. \tag{7.3}$$

Таблица (7.3) называется *распределением д.с.в. ξ.* Ясно, что эта таблица полностью характеризует поведение случайной величины ξ.

В приложениях теории вероятностей, как правило, имеют дело не с самими случайными величинами, а с их распределениями. Это связано с тем, что обычно в результате случайного эксперимента регистрируется значение случайной величины $\xi(\omega)$, но не регистрируются, как мы видели выше, каким элементарным исходом ω закончился сам опыт. Регистрируется, следовательно, значение функции, но не значение аргумента. То обстоятельство, что при разных ω случайная величина $\xi(\omega)$ может принимать одно и то же значение, неожиданно оказывается очень

[52] Читатель, незнакомый с понятиями математического анализа, может считать, что набор возможных значений случайной величины конечен. А все суммы, встречающиеся далее, состоят из конечного числа слагаемых. Для большого количества прикладных задач, связанных с дискретными случайными величинами, такое предположение вполне оправдано.

существенным: множество $\{x_1, x_2, ..., x_n, ...\}$ возможных значений случайной величины может быть гораздо проще, чем все множество Ω. Поэтому может получиться так, что мы не в состоянии узнать из опыта вероятности $P(\omega)$, но можем определить по частотам вероятности

$$p_j = P\left\{\omega : \xi(\omega) = x_j\right\}.$$

Как мы уже отмечали, значения $\{x_1, x_2, ..., x_n, ...\}$ случайной величины могут быть любыми. В то же время соответствующие вероятности p_j должны, очевидно, удовлетворять условиям: $0 \le p_j \le 1$ и

$$(*) \qquad \sum_j p_j = \sum_{x_j} P\left\{\xi = x_j\right\} = \sum_{x_j}\left[\sum_{\omega : \xi(\omega) = x_j} P(\omega)\right] = \sum_{\omega \in \Omega} P(\omega) = 1.$$

Смысл символов суммы в последней формуле вполне ясен: например, символ

$$\sum_{\omega : \xi(\omega) = x_j} P(\omega)$$

означает суммирование по всем ω таким, что $\xi(\omega) = x_j$, т.е., согласно (7.3), по всем $\omega \in B_j$. Кроме того, равенство

$$P\left\{\xi = x_j\right\} = \sum_{\omega : \xi(\omega) = x_j} P(\omega)$$

следует из того факта, что вероятность случайного события (в данном случае – B_j) есть сумма вероятностей элементарных событий, его составляющих (см. §3, п. 3.2).

Равенство (*) выражает тот естественный факт, что сумма *всех* p_j есть вероятность достоверного события.

При всех достоинствах рассмотрения распределения д.с.в. вместо ее самой, переход от случайной величины к ее распределению может быть недостаточным для удобства применения в прикладных задачах. Действительно, возможных значений $x_1, x_2, ..., x_n, ...$ может быть все же слишком много, чтобы можно было определить из опыта их вероятности p_j. Желательно, поэтому, охарактеризовать распределение несколькими параметрами, которые затем можно было бы определить экспериментально. При этом, конечно, эти параметры должны содержать достаточную для практических целей информацию о случайной величине. Эта задача чрезвычайно важна, и ей мы посвятим последующие пункты этого параграфа.

7.2. Математическое ожидание.

Определение 7.3. Пусть дана д.с.в.[53] $\xi(\omega)$. Число

$$\mathbf{M}\xi = \sum_{\omega\in\Omega} \xi(\omega)P(\omega) \tag{7.4}$$

называется *математическим ожиданием* д.с.в. ξ.

Замечание. Если пространство элементарных событий Ω *счетно*, то в формуле (7.4) справа стоит бесконечный ряд. В этом случае мы предполагаем, что он *абсолютно сходится*. Если этот ряд расходится или сходится лишь условно, мы будем говорить, что д.с.в. ξ *не имеет математического ожидания*.

Пусть $x_1, x_2, ..., x_n, ...$ - возможные значения д.с.в. ξ, а $p_1, p_2, ..., p_n, ...$- вероятности этих значений (иными словами, нам задано распределение случайной величины ξ).

Теорема 7.1. Если д.с.в. ξ имеет математическое ожидание, то

$$\mathbf{M}\xi = \sum_j x_j p_j, \tag{7.5}$$

причем в случае бесконечного числа слагаемых ряд справа абсолютно сходится.

Таким образом, математическое ожидание выражается через распределение случайной величины: нужно значения случайной величины умножить на их вероятности и сложить все полученные произведения.

Доказательство. Пусть сначала пространство элементарных событий Ω состоит из конечного числа элементов. Тогда сумма (7.4) состоит из конечного числа слагаемых и, следовательно, ее члены можно как угодно переставлять и группировать. Поэтому:

$$\mathbf{M}\xi = \sum_{\omega\in\Omega} \xi(\omega)P(\omega) = \sum_{x_j}\left[\sum_{\omega:\xi(\omega)=x_j} \xi(\omega)P(\omega)\right] = \sum_{x_j}\left[x_j \sum_{\omega:\xi(\omega)=x_j} P(\omega)\right] =$$

$$= \sum_{x_j} x_j P\{\xi = x_j\} = \sum_j x_j p_j.$$

В случае бесконечного числа слагаемых в (7.4)[54] справа стоит, по условию, абсолютно сходящийся ряд. Как известно из анализа, в абсолютно сходящемся ряде его члены можно как угодно переставлять и группировать,

[53] Напомним: сокращение д.с.в. означает «дискретная случайная величина».
[54] Т.е. когда пространство элементарных событий состоит из счетного числа слагаемых.

причем его сумма останется неизменной. Таким образом, выкладки, приведенные выше, сохраняют силу и в этом случае.

Теорема доказана.

Замечание. Повторяя выкладки, приведенные выше, получим равенство

$$\sum_{\omega \in \Omega} |\xi(\omega)| P(\omega) = \sum_j |x_j| p_j,$$

которое справедливо как для конечных, так и для бесконечных сумм. В последнем случае это позволяет утверждать, что ряды $\sum\limits_{\omega \in \Omega} \xi(\omega) P(\omega)$ и $\sum\limits_j x_j p_j$ одновременно сходятся или не сходятся абсолютно.

На равенстве (7.5) основаны различные интерпретации математического ожидания. Так, если мы в точки прямой линии с абсциссами $x_1, x_2, ..., x_n, ...$ поместим материальные точки с массами соответственно $p_1, p_2, ..., p_n, ...$, то, учитывая равенство $\sum p_j = 1$, найдем, что (7.5) дает абсциссу центра тяжести этой системы материальных точек.[55]

Выясним основные свойства математического ожидания. *Всюду ниже, говоря о нескольких случайных величинах, мы будем предполагать, что они заданны на одном и том же пространстве элементарных событий.*

Теорема 7.2. Пусть ξ и η – две случайные величины, такие, что их математические ожидания существуют. Пусть a и b – любые вещественные числа. Тогда у случайной величины $a\xi + b\eta$ также существует математическое ожидание, причем

$$\mathbf{M}(a\xi + b\eta) = a\mathbf{M}\xi + b\mathbf{M}\eta. \tag{7.6}$$

Доказательство. Имеем:

$$a\mathbf{M}\xi + b\mathbf{M}\eta = a\sum_{\omega \in \Omega} \xi(\omega) P(\omega) + b\sum_{\omega \in \Omega} \eta(\omega) P(\omega) =$$

$$= \sum_{\omega \in \Omega} \left[a\xi(\omega) + b\eta(\omega)\right] P(\omega) = \sum_{\omega \in \Omega} (a\xi + b\eta)(\omega) P(\omega) = \mathbf{M}(a\xi + b\eta).$$

Здесь через $(a\xi + b\eta)(\omega)$ обозначено значение случайной величины $a\xi + b\eta$ на элементе $\omega \in \Omega$. Для случая сумм с конечным числом слагаемых приведенные выше равенства очевидны; для случая сумм с бесконечным числом слагаемых, т.е. рядов, надо сослаться на то, что сумма двух абсолютно сходящихся рядов есть абсолютно сходящийся ряд.

[55] В соответствии с определением центра тяжести в механике.

Теорема 7.3. Пусть $f(x)$ - любая функция вещественного переменного x, ξ - случайная величина с распределением

$$(+) \qquad \begin{pmatrix} x_1 & x_2 & \dots & x_n & \dots \\ p_1 & p_2 & \dots & p_n & \dots \end{pmatrix}.$$

Тогда

$$\mathbf{M}f(\xi) = \sum_j f(x_j)p_j$$

(в предположении, что правая часть представляет сумму конечного числа слагаемых, либо является абсолютно сходящимся рядом).

Доказательство. Имеем:

$$\mathbf{M}f(\xi) = \sum_{\omega\in\Omega} f(\xi(\omega))P(\omega) = \sum_{x_j}\left[\sum_{\omega:\xi(\omega)=x_j} f(\xi(\omega))P(\omega)\right] =$$

$$= \sum_{x_j}\left[f(x_j)\sum_{\omega:\xi(\omega)=x_j} P(\omega)\right] = \sum_j f(x_j)p_j, \text{ ч.т.д.}$$

Замечание. Распространенной ошибкой является утверждение о том, что если (+) является распределением д.с.в. ξ, то

$$(++) \qquad \begin{pmatrix} f(x_1) & f(x_2) & \dots & f(x_n) & \dots \\ p_1 & p_2 & \dots & p_n & \dots \end{pmatrix}$$

будет распределением д.с.в. $f(\xi)$. Это, вообще говоря, неверно. Дело в том, что когда речь идет о возможных значениях $x_1, x_2, ..., x_n, ...$ д.с.в. ξ, то, по определению, *все эти значения предполагаются различными*, в то время, как среди величин $f(x_1), f(x_2), ..., f(x_n), ...$ могут быть и одинаковые. Поэтому, таблица (++) распределением д.с.в. $f(\xi)$ в общем случае являться не будет. Исключение составляет случай, когда *все значения* $f(x_j)$ *будут различными*, например, если функция $f(x)$ строго монотонна.

Рассмотрим несколько примеров.

Пример 1 (распределение Бернулли). Рассмотрим серию n независимых испытаний Бернулли с вероятностью успеха p. В качестве случайной величины ξ возьмем *число успехов* в этой серии. Очевидно, эта случайная величина дискретна с возможными значениями 0, 1, 2, …, n. Распределение ξ в соответствии с формулой (6.1) имеет вид:

(7.7) $$\begin{pmatrix} 0 & 1 & 2 & \dots & m & \dots & n \\ q^n & C_n^1 pq^{n-1} & C_n^2 p^2 q^{n-2} \dots & & C_n^m p^m q^{n-m} & \dots & p^n \end{pmatrix}.$$

Это и есть биномиальное распределение.

Для математического ожидания получаем по формуле (7.5):

$$\mathbf{M}\xi = \sum_{m=1}^{n} m P_n(m) = \sum_{m=1}^{n} m C_n^m p^m q^{n-m} = \sum_{m=1}^{n} m \frac{n(n-1)(n-2)\dots(n-m+1)}{m!} p^m q^{n-m} =$$

$$= \sum_{m=1}^{n} \frac{n(n-1)(n-2)\dots(n-m+1)}{(m-1)!} p^m q^{n-m} =$$

$$= np \sum_{k=0}^{n-1} \frac{(n-1)(n-2)\dots((n-1)-k+1)}{k!} p^k q^{(n-1)-k} = np(p+q)^{n-1} = np .$$

Здесь при преобразованиях принято $k = m-1$ и учтено, что $p+q=1$. Предпоследнее равенство следует из формулы бинома Ньютона (5.2).

Таким образом, для распределения Бернулли $\mathbf{M}\xi = np$.

Пример 2 (распределение Пуассона). Распределение Пуассона (6.6) доставляет нам пример дискретной случайной величины со счетным числом возможных значений:

(7.8) $$\begin{pmatrix} 0 & 1 & 2 & \dots & n & \dots \\ e^{-\lambda} & \dfrac{\lambda}{1!}e^{-\lambda} & \dfrac{\lambda^2}{2!}e^{-\lambda} & \dots & \dfrac{\lambda^n}{n!}e^{-\lambda} & \dots \end{pmatrix}.$$

Математическое ожидание

$$\mathbf{M}\xi = \sum_{k=1}^{\infty} k \frac{\lambda^k}{k!} e^{-\lambda} = e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^k}{(k-1)!} = e^{-\lambda} \sum_{m=0}^{\infty} \frac{\lambda^{m+1}}{m!} = \lambda e^{-\lambda} \sum_{m=0}^{\infty} \frac{\lambda^m}{m!} = \lambda e^{-\lambda} e^{\lambda} = \lambda .$$

Здесь мы использовали известное равенство

$$\sum_{m=0}^{\infty} \frac{\lambda^m}{m!} = e^{\lambda} .$$

Таким образом, для распределения Пуассона $\mathbf{M}\xi = \lambda$.

7.3. Дисперсия.

Математическое ожидание дает некоторую информацию о характере поведения случайной величины ξ, указывая на некоторое «среднее» ее значения, т.е. величину, возле которой «колеблются» возможные значения ξ. Это, однако, недостаточно для более или менее правильной оценки характера поведения ξ, поскольку эти «колебания» могут иметь весьма большой «размах». Например, если распределение ξ имеет вид

$$\begin{pmatrix} \xi: & -10 & 3 & -3 & 10 \\ P: & 0{,}1 & 0{,}4 & 0{,}4 & 0{,}1 \end{pmatrix},$$

то $\mathbf{M}\xi = 0$, что, конечно, не дает никакого адекватного представления о поведении величины ξ.

Поэтому оказывается необходимым ввести какую-то характеристику отклонения случайной величины от ее среднего значения.

Определение 7.4. *Дисперсией* случайной величины ξ называется число

$$\mathbf{D}\xi = \mathbf{M}\left[\left(\xi - \mathbf{M}\xi\right)^2\right], \tag{7.9}$$

то есть дисперсия – это математическое ожидание случайной величины $\left(\xi - \mathbf{M}\xi\right)^2$.

Среднеквадратическим отклонением случайной величины называется число $\sigma = \sqrt{\mathbf{D}\xi}$; таким образом, для дисперсии иногда используют обозначение σ^2 (вместо $\mathbf{D}\xi$).

Замечание. В правой части (7.9) квадратные скобки обычно опускаются; вообще $\mathbf{M}\xi^2$ означает $\mathbf{M}(\xi^2)$. Квадрат же математического ожидания $\mathbf{M}\xi$ записывается в виде $(\mathbf{M}\xi)^2$.

Почему в качестве меры отклонения случайной величины от ее математического ожидания выбрано значение $\mathbf{M}\left(\xi - \mathbf{M}\xi\right)^2$?

Формально, с тем же правом можно было бы выбрать, например, $\mathbf{M}\left|\xi - \mathbf{M}\xi\right|$ или $\mathbf{M}\left(\xi - \mathbf{M}\xi\right)^4$, или $\mathbf{M}\left|\xi - \mathbf{M}\xi\right|^3$.

Оказывается, однако, что именно дисперсия (7.9), а не какая-либо другая из возможных мер отклонения от среднего, входит в формулировку многих важных теорем теории вероятностей,[56] в частности, так называемых предельных теорем теории вероятностей, важнейшей из которых является *центральная предельная теорема*. Некоторые примеры этого рода мы

[56] Как заметил известный американский математик Майкл Спивак, многие математические теоремы становятся тривиальными, если входящие в них понятия должным образом определены.

встретим ниже.

Получим для дисперсии несколько полезных формул.

По теореме 7.3 получаем для дисперсии следующее выражение:

$$\mathbf{D}\xi = \mathbf{M}\left(\xi - \mathbf{M}\xi\right)^2 = \sum_{x_j} (x_j - \mathbf{M}\xi)^2 p_j .$$

Раскрывая скобки, будем иметь

$$\mathbf{D}\xi = \mathbf{M}\left(\xi - \mathbf{M}\xi\right)^2 = \mathbf{M}\left(\xi^2 - 2\xi\mathbf{M}\xi + (\mathbf{M}\xi)^2\right) =$$
$$= \mathbf{M}\xi^2 - \mathbf{M}\left(2\xi\mathbf{M}\xi\right) + \mathbf{M}(\mathbf{M}\xi)^2 = \mathbf{M}\xi^2 - 2(\mathbf{M}\xi)(\mathbf{M}\xi) + (\mathbf{M}\xi)^2 =$$
$$= \mathbf{M}\xi^2 - (\mathbf{M}\xi)^2 .$$

Здесь мы воспользовались теоремой 7.2 и тем фактом, что $\mathbf{M}\xi$ есть постоянное (неслучайное) число. А математическое ожидание постоянной величины равно, очевидно, этой величине, т.е. если C – константа, то $\mathbf{M}C = C$.

Таким образом, получаем полезную формулу

$$\mathbf{D}\xi = \mathbf{M}\xi^2 - (\mathbf{M}\xi)^2 = \sum_j x_j^2 p_j - \left(\sum_j x_j p_j\right)^2 . \tag{7.10}$$

Замечание. Из (7.10) очевидно, что *для постоянной величины C дисперсия равна нулю*: $\mathbf{D}C = 0$. Кроме того, с учетом свойств математического ожидания (теоремы 7.2 и 7.3), легко получаем:

$$\mathbf{D}(C\xi) = \mathbf{M}(C^2\xi^2) - (\mathbf{M}(C\xi))^2 = C^2\mathbf{M}\xi^2 - (C\mathbf{M}\xi)^2 =$$

$$= C^2\left[\mathbf{M}\xi^2 - (\mathbf{M}\xi)^2\right] = C^2\mathbf{D}\xi .$$

Обратимся к примерам.

Пример 1 (распределение Бернулли). Распределение Бернулли, или биномиальное распределение имеет вид (7.7). Вычислим дисперсию соответствующей случайной величины. Как мы видели выше, для биномиального распределения $\mathbf{M}\xi = np$. Поэтому:

$$\mathbf{D}\xi = \mathbf{M}\xi^2 - (\mathbf{M}\xi)^2 = \mathbf{M}\xi^2 - (np)^2 = \sum_m x_m^2 p_m - (np)^2 =$$
$$= \sum_{m=1}^{n} m^2 \frac{n(n-1)(n-2)...(n-m+1)}{m!} p^m q^{n-m} - (np)^2 = S - (np)^2 ,$$

где обозначено

$$S=\sum_{m=1}^{n}m^2\frac{n(n-1)(n-2)...(n-m+1)}{m!}p^m q^{n-m} .$$

Имеем:

$$S=\sum_{m=1}^{n}m\frac{n(n-1)(n-2)...(n-m+1)}{(m-1)!}p^m q^{n-m}=$$

$$\overset{k=m-1}{=} np\sum_{k=0}^{n-1}(k+1)\frac{(n-1)(n-2)...((n-1)-k+1)}{k!}p^k q^{(n-1)-k}=$$

$$=np\sum_{k=0}^{n-1}k\frac{(n-1)(n-2)...((n-1)-k+1)}{k!}p^k q^{(n-1)-k}+$$

$$+np\sum_{k=0}^{n-1}\frac{(n-1)(n-2)...((n-1)-k+1)}{k!}p^k q^{(n-1)-k}=$$

$$=np\sum_{k=1}^{n-1}\frac{(n-1)(n-2)...((n-1)-k+1)}{(k-1)!}p^k q^{(n-1)-k}+np(p+q)^{n-1}=$$

$$\overset{j=k-1}{=} \quad n(n-1)p^2\sum_{j=0}^{n-2}\frac{(n-2)...((n-2)-j+1)}{j!}p^j q^{(n-2)-j}+np=$$

$$=n(n-1)p^2(p+q)^{n-2}+np=n(n-1)p^2+np=$$

$$=(np)^2+np(1-p)=(np)^2+npq ,$$

где учтено, что $p+q=1$.

Таким образом, для *дисперсии биномиального распределения* получаем окончательно: $\mathbf{D}\xi=S-(np)^2=npq$.

Среднеквадратическое отклонение будет, следовательно, равно: $\sigma=\sqrt{npq}$.

Пример 2 (распределение Пуассона). Для распределения Пуассона (7.8), как мы знаем, $\mathbf{M}\xi=\lambda$. Поэтому:

$$\mathbf{D}\xi=\mathbf{M}\xi^2-(\mathbf{M}\xi)^2=\mathbf{M}\xi^2-\lambda^2=\sum_{k=1}^{\infty}k^2\frac{\lambda^k}{k!}e^{-\lambda}-\lambda^2=$$

$$\overset{m=k-1}{=} \lambda e^{-\lambda}\sum_{m=0}^{\infty}(m+1)\frac{\lambda^m}{m!}-\lambda^2=\lambda e^{-\lambda}\left(\sum_{m=0}^{\infty}m\frac{\lambda^m}{m!}+\sum_{m=0}^{\infty}\frac{\lambda^m}{m!}\right)-\lambda^2=$$

$$=\lambda e^{-\lambda}\sum_{m=1}^{\infty}\frac{\lambda^m}{(m-1)!}+\lambda e^{-\lambda}\sum_{m=0}^{\infty}\frac{\lambda^m}{m!}-\lambda^2 \overset{j=m-1}{=} \lambda e^{-\lambda}\sum_{j=0}^{\infty}\frac{\lambda^{j+1}}{j!}+\lambda e^{-\lambda}e^{\lambda}-\lambda^2=$$

$$=\lambda^2 e^{-\lambda}\sum_{j=0}^{\infty}\frac{\lambda^j}{j!}+\lambda-\lambda^2=\lambda^2 e^{-\lambda}e^{\lambda}+\lambda-\lambda^2=\lambda .$$

Таким образом, для распределения Пуассона $\mathbf{D}\xi=\lambda$ (т.е. дисперсия по величине совпадает с математическим ожиданием).

Среднеквадратическое отклонение будет, следовательно, равно $\sigma=\sqrt{\lambda}$.

7.4. Независимые случайные величины.

Мы уже видели, какое большое значение имеет понятие независимости для случайных событий. Оказывается, что аналогичное понятие можно естественным образом ввести и для случайных величин.

Пусть в результате опыта могут наблюдаться две случайные величины ξ и η. Слова «могут наблюдаться» мы понимаем в том смысле, что для любых двух числовых множеств A и B мы можем сказать, произошло или не произошло каждое из двух событий $\{\xi\in A\}$ и $\{\eta\in B\}$ (напомним, что так для краткости обозначаются множества $\{\omega:\xi(\omega)\in A\}$ и $\{\omega:\eta(\omega)\in B\}$).. «Независимость» случайных величин интуитивно понимается так, что, зная результат наблюдения над одной случайной величиной, мы ничего не можем сказать дополнительно о другой случайной величине. Этим мотивируется следующее

Определение 7.5. Две случайных величины ξ и η называются *независимыми*, если для любых двух числовых множеств A и B события $\{\xi\in A\}$ и $\{\eta\in B\}$ независимы, т.е. (см. §2, п. 2.2) вероятность их произведения равна произведению вероятностей этих событий:

(7.11) $$P\left(\{\xi\in A\}\cdot\{\eta\in B\}\right)=P\{\xi\in A\}\cdot P\{\eta\in B\} .$$

Замечание. Произведение событий $\{\xi\in A\}$ и $\{\eta\in B\}$, т.е. совокупность всех тех $\omega\in\Omega$, для которых *одновременно* $\xi(\omega)\in A$ и $\eta(\omega)\in B$:

$$\{\xi\in A\}\cdot\{\eta\in B\}=\{\omega:\xi(\omega)\in A,\eta(\omega)\in B\}$$

мы далее для краткости будем обозначать символом $\{\xi \in A, \eta \in B\}$. Аналогично, символом $\{\xi = a_i, \eta = b_j\}$ мы будем обозначать совокупность всех тех $\omega \in \Omega$, для которых *одновременно* $\xi(\omega) = a_i$ и $\eta(\omega) = b_j$, т.е. произведение событий $\{\xi = a_i\} = \{\omega : \xi(\omega) = a_i\}$ и $\{\eta = b_j\} = \{\omega : \eta(\omega) = b_j\}$:

$$\{\xi = a_i, \eta = b_j\} \overset{опр}{=} \{\xi = a_i\} \cdot \{\eta = b_j\} = \{\omega : \xi(\omega) = a_i, \eta(\omega) = b_j\}.$$

Пусть $x_1, x_2, ..., x_n, ...$ — возможные значения д.с.в. ξ и $P(\xi = x_i) = p_i$, $y_1, y_2, ..., y_m, ...$ — возможные значения д.с.в. η и $P(\eta = y_j) = q_j$.

Если случайные величины ξ и η – независимы, то полагая в (7.10) $A = \{x_i\}$ и $B = \{y_j\}$ (одноточечные числовые множества), получим:

$$P\left(\xi = x_i, \eta = y_j\right) = P(\xi = x_i)P(\eta = y_j) = p_i q_j. \tag{7.12}$$

Справедливо и обратное утверждение: если для всех $x_1, x_2, ..., x_n, ...$ и всех $y_1, y_2, ..., y_m, ...$ выполнено равенство (7.12), то случайные величины ξ и η – независимы. Действительно, для любых числовых множеств A и B имеем:

$$P\left(\xi \in A, \eta \in B\right) = \sum_{x_i \in A, y_j \in B} P\left(\xi = x_i, \eta = y_j\right) \overset{(7.12)}{=} \sum_{x_i \in A, y_j \in B} P(\xi = x_i)P(\eta = y_j) =$$

$$= \sum_{x_i \in A} P(\xi = x_i) \sum_{y_j \in B} P(\eta = y_j) = P(\xi \in A)P(\eta \in B),$$

что и требовалось.

Мы видим, таким образом, что справедлива следующая

Теорема 7.4. Для независимости случайных величин ξ и η необходимо и достаточно, чтобы для любых x_i и y_j было выполнено равенство (7.12).

Установим теперь некоторые свойства математического ожидания и дисперсии, связанные с понятием независимости случайных величин.

Теорема 7.5. Если случайные величины ξ и η независимы и существуют математические ожидания $\mathbf{M}\xi$ и $\mathbf{M}\eta$, то существует математическое ожидание произведения $\xi\eta$ и

$$\mathbf{M}(\xi\eta) = \mathbf{M}\xi \cdot \mathbf{M}\eta. \tag{7.13}$$

Доказательство. Имеем:

$$\mathbf{M}(\xi\eta)=\sum_{\omega\in\Omega}\xi(\omega)\eta(\omega)P(\omega)=\sum_{x_i,y_j}\left\{\sum_{\omega:\left(\substack{\xi(\omega)=x_i\\ \eta(\omega)=y_j}\right)}\xi(\omega)\eta(\omega)P(\omega)\right\}=$$

$$=\sum_{x_i,y_j}\left\{x_iy_j\sum_{\omega:\left(\substack{\xi(\omega)=x_i\\ \eta(\omega)=y_j}\right)}P(\omega)\right\}=\sum_{x_i,y_j}x_iy_jP\big(\xi=x_i,\eta=y_j\big)=$$

$$\overset{(7.12)}{=}\sum_{x_i,y_j}x_iy_jp_iq_j=(\sum_{x_i}x_ip_i)\cdot(\sum_{y_j}y_jq_j)=\mathbf{M}\xi\cdot\mathbf{M}\eta,\ \text{ч.т.д.}$$

Обратимся теперь к дисперсии и рассмотрим вопрос о дисперсии суммы $\xi+\eta$ двух случайных величин *не предполагая*, поначалу, что эти величины независимы. Вычислим:

$$\mathbf{D}(\xi+\eta)=\mathbf{M}\big[\xi+\eta-\mathbf{M}(\xi+\eta)\big]^2\overset{(7.6)}{=}\mathbf{M}\big[(\xi-\mathbf{M}\xi)+(\eta-\mathbf{M}\eta)\big]^2=$$

$$=\mathbf{M}(\xi-\mathbf{M}\xi)^2+\mathbf{M}(\eta-\mathbf{M}\eta)^2+2\mathbf{M}\big[(\xi-\mathbf{M}\xi)(\eta-\mathbf{M}\eta)\big].$$

Выражение $\operatorname{cov}(\xi,\eta)=\mathbf{M}\big[(\xi-\mathbf{M}\xi)(\eta-\mathbf{M}\eta)\big]$ называется *ковариацией* величин ξ и η. Для *независимых* случайных величин в силу (7.6) и (7.13):

$$\operatorname{cov}(\xi,\eta)=\mathbf{M}(\xi-\mathbf{M}\xi)\mathbf{M}(\eta-\mathbf{M}\eta)=0.$$

Таким образом, нами доказана

Теорема 7.6. *Для любых* случайных величин

(7.14) $$\mathbf{D}(\xi+\eta)=\mathbf{D}\xi+\mathbf{D}\eta+2\operatorname{cov}(\xi,\eta).$$

Для независимых ξ и η

(7.15) $$\mathbf{D}(\xi+\eta)=\mathbf{D}\xi+\mathbf{D}\eta.$$

Следствие. Если случайные величины $\xi_1,\xi_2,...,\xi_n$ *попарно независимы* (т.е. любые две случайные величины ξ_i и ξ_j при $i\neq j$ независимы) и дисперсии $\mathbf{D}\xi_i$ $(i=1,2,...,n)$ существуют, то

(7.15a) $$\mathbf{D}(\xi_1+\xi_2+...+\xi_n)=\mathbf{D}\xi_1+\mathbf{D}\xi_2+...+\mathbf{D}\xi_n.$$

Действительно, достаточно проверить равенство (7.15a) для трех попарно независимых случайных величин ξ,η,ς. Тогда остальное легко получится по индукции.

Итак, пусть случайные величины ξ,η,ς попарно независимы. Тогда:

(+) $$\mathbf{D}(\xi+\eta+\varsigma)=\mathbf{D}(\xi+(\eta+\varsigma))=\mathbf{D}\xi+\mathbf{D}(\eta+\varsigma)+2\mathrm{cov}(\xi,\eta+\varsigma)=$$

$$=\mathbf{D}\xi+\mathbf{D}(\eta+\varsigma)+2\mathrm{cov}(\xi,\eta+\varsigma)=\mathbf{D}\xi+\mathbf{D}\eta+\mathbf{D}\varsigma+2\mathrm{cov}(\xi,\eta+\varsigma).$$

Для $\mathrm{cov}(\xi,\eta+\varsigma)$ имеем:

$$\mathrm{cov}(\xi,\eta+\varsigma)=\mathbf{M}\left[(\xi-\mathbf{M}\xi)(\eta+\varsigma-\mathbf{M}(\eta+\varsigma))\right]=$$

$$=\mathbf{M}\left[(\xi-\mathbf{M}\xi)(\eta+\varsigma-\mathbf{M}\eta-\mathbf{M}\varsigma)\right].$$

Производя перемножение внутри квадратных скобок и используя свойства математического ожидания и попарную независимость случайных величин ξ,η,ς, получаем $\mathrm{cov}(\xi,\eta+\varsigma)=0$. Тогда формула (+) дает:

$$\mathbf{D}(\xi+\eta+\varsigma)=\mathbf{D}\xi+\mathbf{D}\eta+\mathbf{D}\varsigma,$$ ч.т.д.

7.5. Неравенство Чебышёва.[57]

В этом пункте мы познакомимся с некоторыми важными результатами, использующими понятия математического ожидания, дисперсии и независимости случайных величин.

Теорема 7.7 (неравенство Чебышёва). Пусть имеется д.с.в. ξ, у которой существуют $\mathbf{M}\xi$ и $\mathbf{D}\xi$. Тогда для любого $\varepsilon>0$ справедливо неравенство:

(7.16) $$P\left(\left|\xi-\mathbf{M}\xi\right|\ge\varepsilon\right)\le\frac{\mathbf{D}\xi}{\varepsilon^2}.$$

Замечания.

1. Обозначение $P\left(\left|\xi-\mathbf{M}\xi\right|\ge\varepsilon\right)$ есть сокращенное обозначение для вероятности

[57] Пафнутий Львович Чебышёв (1821-1894) – один из наиболее видных российских математиков второй половины XIX века, создатель петербургской научной школы, академик Петербургской Императорской Академии Наук с 1856 года.

$$P\left\{\omega : \left|\xi(\omega) - \mathbf{M}\xi\right| \geq \varepsilon\right\} = \sum_{x_i : \left|x_i - \mathbf{M}\xi\right| \geq \varepsilon} P(\xi = x_i)$$

(здесь, как и выше, $x_1, x_2, ..., x_n, ...$ — возможные значения случайной величины ξ).

2. Смысл неравенства Чебышёва состоит в том, при малых значениях дисперсии $\mathbf{D}\xi$ большие отклонения случайной величины ξ от ее математического ожидания $\mathbf{M}\xi$ маловероятны. Это подтверждает представление о дисперсии как мере отклонения случайной величины от ее математического ожидания.

Доказательство теоремы. Имеем:

$$\mathbf{D}\xi = \sum_{x_i} (x_i - \mathbf{M}\xi)^2 P(\xi = x_i).$$

Если суммирование по всем x_i заменить суммированием только по тем x_i, для которых $\left|x_i - \mathbf{M}\xi\right| \geq \varepsilon$, то сумма не увеличится. Поэтому:

$$\mathbf{D}\xi \geq \sum_{x_i : \left|x_i - \mathbf{M}\xi\right| \geq \varepsilon} (x_i - \mathbf{M}\xi)^2 P(\xi = x_i) \geq$$

$$\geq \varepsilon^2 \sum_{x_i : \left|x_i - \mathbf{M}\xi\right| \geq \varepsilon} P(\xi = x_i) = \varepsilon^2 P\left(\left|\xi - \mathbf{M}\xi\right| \geq \varepsilon\right),$$

что, очевидно, эквивалентно (7.16).

7.6. Закон больших чисел.

Случайные величины возникают в приложениях как результаты измерений, причем либо сами измерения подвержены случайным ошибкам, либо объекты измерения случайным образом выбираются из некоторой совокупности. Давно было замечено, что, в то время как результаты отдельных измерений $\xi_1, \xi_2, ..., \xi_n$ могут колебаться сильно, их *средние арифметические* $\frac{1}{n}(\xi_1 + \xi_2 + ... + \xi_n)$ обнаруживают (при больших n) гораздо бóльшую устойчивость, т.е. меняются незначительно (так называемая «устойчивость средних»). Как правило, указанная величина среднего арифметического совокупности n измерений одной и той же *физической величины* принимается за ее *среднее значение*, т.е. приближенное значение, которое можно использовать в тех вычислениях, где эта физическая величина встречается. Здесь, однако, имеется определенная тонкость, требующая разъяснения.

Дело в следующем. При измерении физической величины мы не можем заранее указать все факторы, которые в той или иной степени повлияют на показания приборов, с помощью которых мы это измерение производим. Поэтому, результат *каждого* такого *конкретного измерения* представляет собой одно из возможных значений некоторой случайной величины, которая определяется совокупностью факторов (явлений-причин), которые сопровождают *данный* эксперимент по измерению интересующей нас физической величины. При проведении следующего (независящего от предыдущего) измерения *той же самой физической величины*, даже при соблюдении одинаковости[58] условий, в которых измерения проводятся, мы уже будем иметь дело *с другой* случайной величиной, одно из возможных значений которой и будет наблюдаться при втором измерении.

Иными словами, *каждый конкретный эксперимент по измерению данной физической величины порождает свою случайную величину, одно из возможных значений которой и наблюдается экспериментатором как показание соответствующего измерительного прибора*. Таким образом, мы имеем дело с несколькими (по числу измерений) независимыми случайными величинами. Каждая такая случайная величина имеет свои собственные числовые характеристики (математическое ожидание и дисперсию), которые определяются распределениями каждой из этих случайных величин в отдельности. Интуитивно ясно, что разумнее всего в качестве значения измеряемой в эксперименте физической величины принять ее математическое ожидание, т.е. величину $\mathbf{M}\xi = \sum x_j p_j$. Однако, мы никогда непосредственно этого математического ожидания вычислить не можем, так как нам неизвестно распределение соответствующей случайной величины. На практике в качестве значения измеряемой физической величины принимают *арифметическое среднее различных измерений этой величины*, т.е. значение $(x^{(1)} + x^{(2)} + \ldots + x^{(n)})/n$, где в числителе стоят не возможные значения *одной* случайной величины, а какие-то из возможных значений *разных независимых* случайных величин (представляющих различные измерения). Естественно возникает вопрос, в какой мере такой подход является удовлетворительным, т.е. велика ли разница между математическим ожиданием $\mathbf{M}\xi$ и упомянутым арифметическим средним различных измерений?

Следующая теорема представляющая собой знаменитый *закон больших чисел* (в форме Чебышёва), дает ответ на поставленный вопрос.

Теорема 7.8. Пусть случайные величины $\xi_1, \xi_2, \ldots, \xi_n$ попарно независимы и их дисперсии ограничены в совокупности, т.е. $\mathbf{D}\xi_i \le c$ для всех *i*. Тогда для любого положительного ε при $n \to \infty$ имеем:

---

[58] В той мере, в которой такую «одинаковость» можно обеспечить.

$$P\left(\left|\frac{\xi_1+\xi_2+\ldots+\xi_n}{n}-\frac{\mathbf{M}\xi_1+\mathbf{M}\xi_2+\ldots+\mathbf{M}\xi_n}{n}\right|>\varepsilon\right)\to 0.$$

Доказательство. По формуле (7.6):

$$\frac{\mathbf{M}\xi_1+\mathbf{M}\xi_2+\ldots+\mathbf{M}\xi_n}{n}=\mathbf{M}\left(\frac{\xi_1+\xi_2+\ldots+\xi_n}{n}\right).$$

Далее, в силу попарной независимости $\xi_1,\xi_2,\ldots,\xi_n$, по формуле (7.15а), с учетом равенства $\mathbf{D}(C\xi)=C^2\mathbf{D}\xi$,[59] получаем:

$$\mathbf{D}\left(\frac{\xi_1+\xi_2+\ldots+\xi_n}{n}\right)=\frac{1}{n^2}\sum_{i=1}^{n}\mathbf{D}\xi_i\le\frac{nc}{n^2}=\frac{c}{n},$$

где последнее неравенство следует из условия теоремы. По неравенству Чебышёва получаем окончательно:

$$P\left(\left|\frac{\xi_1+\xi_2+\ldots+\xi_n}{n}-\frac{\mathbf{M}\xi_1+\mathbf{M}\xi_2+\ldots+\mathbf{M}\xi_n}{n}\right|>\varepsilon\right)=$$

$$=P\left(\left|\frac{\xi_1+\xi_2+\ldots+\xi_n}{n}-\mathbf{M}\left(\frac{\xi_1+\xi_2+\ldots+\xi_n}{n}\right)\right|>\varepsilon\right)\le$$

$$\le\frac{\mathbf{D}\left(\frac{\xi_1+\xi_2+\ldots+\xi_n}{n}\right)}{\varepsilon^2}\le\frac{c}{n\varepsilon^2}\xrightarrow[n\to\infty]{}0.$$

Теорема доказана.

Следствие. Если $\mathbf{M}\xi_1=\mathbf{M}\xi_2=\ldots=\mathbf{M}\xi_n=a$, то для любого $\varepsilon>0$ при $n\to\infty$

(7.17) $$P\left(\left|\frac{\xi_1+\xi_2+\ldots+\xi_n}{n}-a\right|>\varepsilon\right)\to 0.$$

Доказательство очевидно.

Это следствие и является выражением «устойчивости средних», о котором говорилось в начале этого пункта. Действительно, при измерении

[59] См. замечание к формуле (7.10).

одной и той же физической величины естественно предполагать, что случайные величины, характеризующие каждое конкретное измерение, имеют одинаковые распределения и, следовательно, одинаковые математические ожидания. Таким образом, при «достаточно большом количестве измерений», замена математического ожидания арифметическим средним измеренных значений физической величины вполне оправдано.

Соотношение (7.17) является крайне важным, но обладает тем же недостатком, что и теорема Пуассона: она не дает возможности достаточно ясно интерпретировать термин «достаточно большое количество измерений». Для испытаний Бернулли соответствующее уточнение дает теорема Муавра-Лапласа. В случае же закона больших чисел такую роль играет одна из важнейших теорем теории вероятностей – *центральная предельная теорема*. Доказательство этой теоремы достаточно сложно, чтобы его можно было изложить в этом пособии; мы приведем только ее «облегченную» формулировку для частного случая, относящегося к формуле (7.17).

Теорема 7.9. Пусть случайные величины $\xi_1, \xi_2, ..., \xi_n$ попарно независимы и имеют одинаковые математические ожидания и дисперсии: $\mathbf{M}\xi_i = a$ и $\mathbf{D}\xi_i = \sigma^2$ для всех *i*. Тогда для любого $\varepsilon > 0$ при $n \gg 1$

$$P\left(\left|\frac{\xi_1 + \xi_2 + ... + \xi_n}{n} - a\right| \le \varepsilon\right) \approx 2\Phi\left(\frac{\varepsilon\sqrt{n}}{\sigma}\right) - 1, \tag{7.18}$$

где $\Phi(x)$ - нормальная функция распределения (§6д, п. 6д.2).

Напомним, что эта же функция $\Phi(x)$ встречается в формулировке теоремы Муавра-Лапласа (теорема 6д.1).

Формула (7.18) позволяет связать число необходимых наблюдений (измерений) с требуемой точностью оценки математического ожидания *a* измеряемой величины. Так, при $\varepsilon\sqrt{n} = 3\sigma$ («правило 3σ») получаем

$$P\left(\left|\frac{\xi_1 + \xi_2 + ... + \xi_n}{n} - a\right| \le \varepsilon\right) \approx 0,997,$$

что вполне достаточно для практических целей. Заметим, что ε убывает, как величина $1/\sqrt{n}$. Отсюда видно, что *ошибка от замены математического ожидания измеряемой величины на* $\frac{1}{n}(\xi_1 + \xi_2 + ... + \xi_n)$ *обратно пропорциональна квадратному корню из числа наблюдений*. Это говорит о том, что иногда бывает лучше сделать более точный прибор (что позволяет уменьшить дисперсию σ^2), чем надеяться увеличить точность за счет увеличения числа наблюдений. Вообще, статистикой надо пользоваться тогда, когда исчерпаны технические возможности.

МАТЕМАТИКА
В ПОЛИТЕХНИЧЕСКОМ УНИВЕРСИТЕТЕ

Выпуск 6

Фирсов Андрей Николаевич

МАТЕМАТИКА

ТЕОРИЯ ВЕРОЯТНОСТЕЙ

Часть 1

Учебное пособие

Лицензия ЛР № 020593 от 07.08.97

Налоговая льгота — Общероссийский классификатор продукции
ОК 005-93, т. 2; 95 3005 — учебная литература

Подписано в печать 29.06.2004. Формат 60Х84/16.
Усл. печ. л. 7,0. Уч.-изд. л. 7,0. Тираж 150. Заказ 352.

Отпечатано с готового оригинал-макета, предоставленного автором,
в типографии Издательства Политехнического университета.
195251, Санкт-Петербург, Политехническая ул., 29.