Министерство образования Республики Беларусь
Учреждение образования
"Гомельский государственный университет
имени Франциска Скорины"

Ю.В. Малинковский

ТЕОРИЯ ВЕРОЯТНОСТЕЙ И МАТЕМАТИЧЕСКАЯ СТАТИСТИКА (ЧАСТЬ 1. ТЕОРИЯ ВЕРОЯТНОСТЕЙ)

Учебное пособие

Допущено Министерством образования Республики Беларусь в качестве учебного пособия для студентов математических специальностей учреждений, обеспечивающих получение высшего образования

Гомель 2004

УДК 519.2(075.8)
ББК 22.17 Я73
М 191

Рецензенты:
кафедра теории функций, функционального анализа, теории вероятностей и прикладной математики Гродненского государственного университета им. Янки Купалы,
заведующий кафедрой теории вероятностей и математической статистики БГУ д-р физ.-мат. наук, профессор Н.Н.Труш

Малинковский Ю.В.

Теория вероятностей и математическая статистика (часть 1. Теория вероятностей): Учебное пособие / Ю.В.Малинковский. — Гомель: УО "ГГУ им. Ф.Скорины", 2004. — 355 с.

ISBN

В основу данного учебного пособия положены курсы лекций по теории вероятностей и математической статистике, которые читаются автором на математическом факультете Гомельского государственного университета им. Ф.Скорины начиная с 1973 года в рамках специальности "Математика". В первой части изложены основы теории вероятностей.

Предназначено для студентов математических специальностей ВУЗов. Будет полезно для студентов других факультетов, изучающих теорию вероятностей и математическую статистику.

Библиогр.: 17 назв.

ISBN

ПОСВЯЩАЕТСЯ 30–летию
Белорусской школы по
теории вероятностей и
математической статистике
и ее создателю Г.А.Медведеву

Предисловие

Начало развития исследований по теории вероятностей и математической статистике в Республике Беларусь связано с приглашением в 1974 году профессора Геннадия Алексеевича Медведева на должность заведующего открывшейся кафедры теории вероятностей и математической статистики БГУ. Им создана известная белорусская вероятностная школа, включающая такие направления исследований как теория массового обслуживания, теория случайных процессов, многомерный статистический анализ, теория стационарных процессов и временных рядов, финансовая и актуарная математика и др. Широкую известность в мире получила Белорусская школа по теории массового обслуживания, которая регулярно проводится в виде тематических международных конференций начиная с 1985 года. Последняя, 17-ая конференция, под названием "Современные математические методы анализа и оптимизации телекоммуникационных сетей" проводилась в Гомельском государственном университете им. Ф.Скорины 23 — 25 сентября 2003 года. Интересы дальнейшего развития белорусской вероятностной школы диктуют создание учебников и пособий по различным разделам фундаментального и прикладного вероятностного анализа.

Данное пособие написано на основе лекций по общему двухсеместровому курсу теории вероятностей и математической статистики, которые читаются ее автором на математическом факультете ГГУ им. Ф.Скорины начиная с 1973 года. Параллельно в функциональном анализе в течение первого семестра студенты изучают теорию меры и интеграла Лебега. Поэтому имеется возможность использования получаемых там студентами сведений. При преподавании теории вероятностей возникают трудности, связанные с тем, что, с одной стороны, необходимо научить студентов построению вероятностных моделей случайных экспериментов, возникающих на практике, и развить у них вероятностную интуицию, а, с другой стороны, дать достаточно строгое аксиоматическое изложение, использующее современные идеи функционального анализа. В настоящем пособии изложение основывается на некотором компромиссе между двумя указанными противоречивыми тенденциями.

Книга состоит из двух частей-глав: глава 1 посвящена собственно теории вероятностей, а глава 2 — математической статистике. В первых четырех параграфах главы 1 рассматривается аксиоматика теории вероятностей, включающая, в частности, определение вероятности в дискретном пространстве элементарных исходов, классическое и геометрическое определения вероятности. Для удобства изложены элементы комбинаторики с полным выводом формул для подсчета наиболее часто встречающихся комбинаторных схем. Автор не считает нужным рассмотрение очень сложных комбинаторных задач на классическое определение вероятности, поскольку это может выбить почву у обучающихся и к тому же создать неверное представление о теории вероятностей в целом. В 5-ом параграфе рассматривается условная вероятность и независимость событий, в 6-ом и 7-ом — схема Бернулли и предельные теоремы для нее. Изучению случайных величин, законов распределения и типов случайных величин посвящены параграфы 8 – 10, изучению многомерных случайных величин — 11-ый параграф. 12-ый параграф посвящен важному понятию независимости случайных величин, 13-ый параграф — законам распределения функций от случайных величин, 14-ый параграф — числовым характеристикам случайных величин. В 15-ом и 16-ом параграфах изучаются виды сходимости случайных величин и законы больших чисел, включая вероятностное доказательство теоремы Вейерштрасса о равномерном приближении непрерывной на отрезке функции многочленами. 17-ый параграф посвящен изучению таких преобразований распределений случайных величин как производящая функция, характеристическая функция, преобразование Лапласа-Стилтьеса. Наконец в последнем, 18-ом параграфе, рассмотрена центральная предельная теорема и ее применение к приближенному вычислению интегралов.

Очень надеюсь, что книга будет полезна как для профессиональных математиков, так и для всех исследователей, которые применяют вероятностно-статистические методы в физике, химии, биологии, технике, экономике, финансах и т.д.

В заключение, автор выражает глубокую благодарность членам кафедры теории функций, функционального анализа, теории вероятностей и прикладной математики Гродненского государственного университета им. Янки Купалы, взявшим на себя нелегкий труд рецензирования данного пособия, и рецензенту Н.Н.Трушу за весьма полезные замечания.

г.Гомель, июнь 2004 г. *Ю.Малинковский*

ГЛАВА 1. ТЕОРИЯ ВЕРОЯТНОСТЕЙ

Введение

Теория вероятностей — математическая наука, изучающая закономерности случайных явлений, в которых проявляется статистическая регулярность. Сказанное можно пояснить на примере подбрасывания "правильной" монеты. Ясно, что заранее невозможно с определенностью предсказать исход каждого подбрасывания. Результаты отдельных экспериментов носят крайне нерегулярный характер, и кажется, что нет возможности познать какие-нибудь закономерности, связанные с такими экспериментами. Тем не менее, если монету подбросить много раз, то проявляется вполне определенная статистическая регулярность — относительная частота выпадения "герба" (отношение $N(г)/N$, где N — число подбрасываний, $N(г)$ — число выпадений "герба") будет близка к $1/2$.

Статистическая устойчивость частот, проявляющаяся во многих других экспериментах, приводит к тому, что можно количественно оценить степень возможности появления того или иного события A, происходящего в результате экспериментов. Теория вероятностей принимает существование у события A определенной числовой характеристики $P(A)$, называемой вероятностью этого события. Назовем отношение

$$P_N(A) = N(A)/N, \qquad (B1)$$

где N — число проведенных экспериментов, $N(A)$ – число экспериментов, в которых появилось событие A, **относительной частотой события** A. Опытным путем было установлено, что для широкого класса явлений с ростом числа экспериментов N относительная частота $P_N(A)$ стабилизируется вокруг некоторого числа $P(A)$, которое естественно принять за вероятность события A. Для рассмотренного выше примера это означает, что вероятность выпадения "герба" при подбрасывании "правильной" монеты надо считать равной $1/2$.

Понятие вероятности является одной из наиболее мощных концепций современной науки. Все современное естествознание пронизано вероятностными идеями. Фундаментальное значение вероятностных идей и представлений в современной науке неоднократно подчеркивалось выдающимися учеными. Например, Н. Винер связывает радикальное становление вероятности в науке, развитие вероятностной точки зрения на устройство мира и основания многих наук с именем Гиббса.

Вероятность вошла в физику в процессе разработки молекулярно-кинетической теории. Но еще большее значение вероятностные идеи приобрели в современной физике, например, в квантовой механике, которая была бы вообще невозможна без понятия вероятности. Многие разделы биологии используют вероятностные модели, особенно генетика, так как механизм наследственности имеет вероятностную природу. Теория вероятностей является теоретической базой таких прикладных разделов современной науки как теория надежности, теория массового обслуживания, теория финансов (особенно в части, связанной с описанием стоимости ценных бумаг, в теории страхования и пенсионных фондов).

Играя фундаментальную роль в современной науке, понятие вероятности является существенным элементом современной философии.

Развитие теории вероятностей можно разбить на следующие этапы.

1. *Предыстория теории вероятностей.* В этот период ставились и решались элементарные задачи, которые позже отнесли к теории вероятностей. Никаких специальных методов в этот период не возникло.

С вероятностными представлениями были знакомы античные ученые, например, Демокрит, Лукреций Кар, которые предвидели строение материи с беспорядочным движением мелких частиц, рассуждали о равновозможных исходах. Но античная наука не дошла до выделения понятия вероятности.

В средние века в работах Л.Пачоли, Н.Тартальи и Д.Кардано делается попытка выделить новое понятие — отношение шансов — при решении ряда задач, прежде всего комбинаторных. Но такие попытки встречаются только эпизодически.

2. *Возникновение теории вероятностей как науки.* Возникновение теории вероятностей как науки относится к середине $XVIII$ века и связано с именами Паскаля (1623-1662), Ферма (1601-1665) и Гюйгенса (1629-1695). Общие методы решения задач, связанных с подсчетом шансов в азартных играх, были даны в знаменитой переписке между Паскалем и Ферма и в первой книге по теории вероятностей "О расчетах в азартной игре" Гюйгенса. Именно в этот период вырабатывается важное понятие "математическое ожидание", устанавливаются теоремы сложения и умножения вероятностей. В это время теория вероятностей находит свои первые применения в демографии, страховом деле, в оценке ошибок

наблюдений.

3. *Фактическая история теории вероятностей* начинается с работы Я.Бернулли (1654-1705) "Искусство предположений", в которой была строго доказана первая предельная теорема теории вероятностей — закон больших чисел, и работы де Муавра (1667-1754) "Аналитическая смесь", в которой впервые была сформулирована и доказана центральная предельная теорема теории вероятностей в симметричной схеме Бернулли. Я.Бернулли был, вероятно, первым, кто осознал важность рассмотрения бесконечных последовательностей повторных испытаний и кто делал четкое различие между понятием вероятности события и частоты его появления. Де Муавру принадлежит заслуга в определении таких понятий, как независимость, математическое ожидание, условная вероятность.

В 1812 г. выходит большой трактат Лапласа (1749-1827) "Аналитическая теория вероятностей", в котором он излагает свои собственные результаты в области теории вероятностей, а также результаты своих предшественников. В частности, он обобщил теорему Муавра на несимметричный случай схемы Бернулли. Лаплас, применив вероятностные методы к теории ошибок наблюдений, высказал плодотворную идею об ошибке наблюдения как суммарном эффекте большого числа независимых элементарных ошибок.

К этому же периоду в развитии теории вероятностей, когда центральное место в исследованиях занимали предельные теоремы, относятся работы Пуассона (1781-1840) и Гаусса (1777-1855). С именем Пуассона связано понятие распределения Пуассона и процесса Пуассона.

Гауссу принадлежит заслуга создания теории ошибок и, в частности, обоснование одного из ее основных принципов — метода наименьших квадратов.

4. *Петербургский период.* Следующий период развития теории вероятностей связан прежде всего с Петербургской школой, особенно такими ее яркими представителями как П.Л.Чебышев (1821-1894), А.А.Марков (1856-1922), А.М.Ляпунов (1857-1918), которые создали эффективные методы доказательства предельных теорем для сумм независимых произвольно распределенных случайных величин. Ляпунов создал мощный математический аппарат — теорию характеристических функций, Марков ввел понятие цепи; теперь без цепей Маркова и марковских процессов невозможно представить себе развитие современной науки.

5. *Современный период.* Современный период в развитии теории вероятностей начинается с установления аксиоматики. Первые работы в этом направлении принадлежат С.Н.Бернштейну, Р.Мизесу и Э.Борелю. В 1933 г. вышла книга Колмогорова "Основные понятия теории вероятностей", в которой была предложена аксиоматика, получившая всеобщее признание и которая используется в настоящее время. К данному моменту теория вероятностей складывается из большого числа самостоятельных научных дисциплин. На вероятностной базе построены математическая статистика, теория случайных процессов, теория информации, а также такие прикладные разделы как теория надежности и теория массового обслуживания, финансовая и актуарная математика.

§ 1. Пространство элементарных исходов. Событие. Относительная частота

В любой математической дисциплине имеются первичные понятия, которые невозможно определить через другие. Например, в геометрии такими понятиями являются точка, прямая и плоскость. Не является исключением и теория вероятностей. В теории вероятностей первичным понятием, не определяемым через другие, является понятие **пространства элементарных исходов** $\Omega = \{\omega\}$ — некоторого множества, состоящего из элементарных исходов ω. Элементарные исходы соответствуют, как правило, единственно возможным неразложимым результатам некоторого случайного эксперимента, хотя это и необязательно. Смысл введенного понятия можно выяснить, рассматривая конкретные примеры.

Пример 1. Пусть эксперимент заключается в подбрасывании "правильной" монеты. В этом случае $\Omega = \{г, р\}$ состоит из двух исходов $\omega_1 = г$, $\omega_2 = р$, соответствующих выпадению "герба" и "решки" соответственно.

Пример 2. При подбрасывании игральной кости возможны шесть элементарных исходов: ω_1 — выпадение 1-го очка, ω_2 — выпадение 2-х очков, и т.д., ω_6 — выпадение 6-ти очков. В этом случае пространство элементарных исходов $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4, \omega_5, \omega_6\}$. Удобно обозначать исход ω_i просто i, тогда $\Omega = \{1, 2, 3, 4, 5, 6\}$, где i соответствует выпадению i очков.

В этом эксперименте можно ввести другое пространство элементарных исходов $\widetilde{\Omega} = \{\omega_1, \omega_2\}$, где ω_1 соответствует выпадению четного числа очков, ω_2 соответствует выпадению нечетного числа очков. Здесь исходы уже не являются неразложимыми, например, исход ω_1 раскладывается на исходы 2,4,6 из пространства Ω. Однако в большинстве случаев удобно рассматривать пространство Ω, состоящее из неразложимых исходов, а не пространство $\widetilde{\Omega}$.

Пример 3. При подбрасывании двух "правильных" монет в качестве элементарного исхода удобно взять пару $\omega = (i, j)$, где $i =$*г*, если на первой монете выпал "герб", и $i =$*р*, если на первой монете выпала "решка", $j =$*г*, если на второй монете выпал "герб", и $j =$*р*, если на второй монете выпала "решка". Пространство элементарных исходов

$$\Omega = \{(г,г),\ \ (г,р),\ \ (р,г),\ \ (р,р)\}$$

является декартовым произведением пространства примера 1 на себя.

В качестве пространства элементарных исходов можно было бы также взять $\widetilde{\Omega} = \{2г,\ \ 2р,\ \ \text{смесь}\}$, где $2г$ — выпадение двух "гербов", $2р$ — выпадение двух "решек", "смесь" — выпадение на одной из монет "герба", а на другой — "решки". Фактически здесь $2г$ совпадает с $(г,г)$ из пространства Ω, $2р$ — с $(р,р)$, а "смесь" получается соединением исходов $(г,р)$ и $(р,г)$. Однако пространство $\widetilde{\Omega}$ никогда не используется, так как его исходы имеют различные вероятности.

Пример 4. При подбрасывании двух игральных костей в качестве элементарного исхода также удобно взять пару $\omega = (i, j)$, где i — число очков, выпавших на 1-й кости, j — число очков, выпавших на 2-й кости. Поэтому пространство элементарных исходов состоит из 36 элементарных исходов: $\Omega = \{(1,1),(1,2),(1,3),(1,4),(1,5),(1,6),(2,1),(2,2),(2,3),$ $(2,4),(2,5),(2,6),(3,1),(3,2),(3,3),(3,4),(3,5),(3,6),(4,1),$ $(4,2),(4,3),(4,4),(4,5),(4,6),(5,1),(5,2),(5,3),(5,4),(5,5),$ $(5,6),(6,1),(6,2),(6,3),(6,4),(6,5),(6,6)\}$ и является декартовым произведением пространства примера 2 на себя.

Пример 5. Пусть на отрезок $[a,b]$ "наудачу" бросается точка. За исход такого эксперимента естественно взять координату точки, в которую она попадет. Так как точка может попасть в любую точку отрезка $[a,b]$, то $\Omega = \{\omega : \omega \in [a,b]\} = [a,b]$.

Пример 6. Пусть на отрезок $[a,b]$ "наудачу" бросаются две точки. За исход можно взять пару $\omega = (x,y)$, где x — координата первой точки, y — координата второй точки. Поэтому

$$\Omega = \{\omega = (x,\, y) :\, a \leq x \leq b,\, a \leq y \leq b\}$$

является декартовым произведением пространства примера 5, т.е. отрезка $[a,b]$, на себя. Иными словами, Ω — квадрат в координатной плоскости $0xy$.

Пример 7. При определении времени жизни элементарной частицы пространство элементарных исходов представляет собой полупрямую $\Omega = [0,\, \infty)$.

Пример 8. Пусть по паспортным данным маневровый локомотив может без капитального ремонта использоваться в течение времени T. Эксперимент состоит в наблюдении времени его бесперебойной работы; тогда элементарный исход ω — время бесперебойной работы, пространство элементарных исходов $\Omega = [0, T]$.

Предположим временно, что пространство элементарных исходов конечно или счетно. Будем называть такое пространство **дискретным**.

Определение 1.1. (Случайным) **событием** называется любое подмножество A пространства элементарных исходов. Элементы множества A называются **элементарными исходами, благоприятствующими появлению события** A или **благоприятствующими событию** A.

Следующее определение нельзя назвать, строго говоря, математическим, поскольку оно использует не определенное ранее понятие "осуществления" элементарного исхода. Это определение служит лишь для того, чтобы связать математическую теорию с возможностью применения при решении практических задач. Оно дает возможность перевода с языка теории на язык практических приложений.

Определение 1.2. Говорят, что **событие** A **происходит** или **осуществляется**, если в результате эксперимента осуществляется элементарный исход, благоприятствующий событию A.

Пример 9. Пусть эксперимент состоит в подбрасывании игральной кости, т.е. $\Omega = \{1, 2, 3, 4, 5, 6\}$. Событию $A = \{1, 3, 5\}$ благоприятствует выпадение 1, 3 или

5 очков, и оно происходит, если на игральной кости выпадет 1, 3 или 5 очков. Другими словами, событие A заключается в выпадении нечетного числа очков. Аналогично, событие $B = \{2, 4, 6\}$ состоит в выпадении четного числа очков.

Пример 10. Пусть в примере 8 время измеряется в сутках и пусть $A = [3,\ 20]$ (предполагается, что $T \geq 20$). Тогда событие A заключается в том, что локомотив проработает бесперебойно от 3-х до 20-ти суток. Если бы оказалось, что $T < 20$, то событие A состояло бы в бесперебойной работе от 3 до T суток.

События будем обозначать прописными латинскими буквами $A, B, C, D, H_1, H_2, H_3$ и т.п.

Все пространство Ω является подмножеством самого себя и, согласно определению 1.1, является событием. Оно называется **достоверным событием**. Так как ему благоприятствует любой элементарный исход, которым может окончиться эксперимент (любое $\omega \in \Omega$), то Ω происходит всегда (конечно, в условиях, когда эксперимент описывается Ω). Таким образом, пространство элементарных исходов Ω выступает в двух качествах: в качестве собственно множества всех элементарных исходов, которые могут произойти в результате эксперимента, и в качестве достоверного события, т.е. события, которое происходит в каждом таком эксперименте.

Пустое множество $\emptyset$ также является подмножеством Ω, т.е. является событием, которое называется **невозможным**. Каким бы исходом ни закончился эксперимент, то, поскольку $\omega \notin \emptyset$, невозможное событие никогда не происходит.

События Ω и $\emptyset$ являются единственно возможными неслучайными событиями (Ω происходит всегда, $\emptyset$ никогда не происходит). Все остальные события являются случайными, т.е. могут как произойти, так и не произойти в результате эксперимента.

Введем операции над событиями. Они вводятся как соответствующие операции над множествами (ведь событие, согласно определению 1.1, является подмножеством Ω).

Определение 1.3. Суммой событий A **и** B называется объединение $A \cup B$ множеств A и B (состоящее из элементарных исходов, принадлежащих хотя бы одному из множеств A или B).

Согласно определению 1.2, $A \cup B$ происходит тогда и только тогда, когда происходит $\omega \in A \cup B$, т.е. ω, принадлежащее хотя бы одному из множеств A или B. В силу определения 1.2 это означает, что происходит хотя бы одно из событий A или B. Таким образом, *сумма событий* $A \cup B$ *– это событие, которое происходит тогда и только тогда, когда происходит хотя бы одно из событий* A *или* B.

Долгое время теория вероятностей и теория множеств развивались независимо. При этом в теории вероятностей первичным понятием являлось событие, понятие пространства элементарных исходов не вводилось вообще. А суммой событий A и B называлось событие, состоящее в наступлении хотя бы одного из событий A или B. Отсутствие понятия пространства элементарных исходов приводило к известным парадоксам. Для создания аксиоматики и, в частности, преодоления этих парадоксов А.Н. Колмогоровым была предложена теоретико-множественная база, лежащая в основе современной теории вероятностей.

Определение 1.3 может быть обощено на произвольное число слагаемых.

Если I — множество индексов произвольной мощности, то **суммой событий** A_α, $\alpha \in I$, называется объединение $\bigcup\limits_{\alpha \in I} A_\alpha$. Очевидно, оно происходит тогда и только тогда, когда происходит хотя бы одно из событий A_α, $\alpha \in I$.

Определение 1.4. Произведением событий A **и** B называется пересечение $A \cap B$ множеств A и B (состоящее из элементарных исходов, принадлежащих A и B).

В силу определения 1.2 произведение $A \cap B$ — это событие, состоящее в одновременном появлении событий A и B. Произведение событий часто обозначается как AB или $A \cdot B$. Таким образом,

$$A \cap B = A \cdot B = AB.$$

Произведением событий A_α, $\alpha \in I$, называется пересечение $\bigcap\limits_{\alpha \in I} A_\alpha$ множеств A_α. Очевидно, $\bigcap\limits_{\alpha \in I} A_\alpha$ происходит тогда и только тогда, когда одновременно происходят все события A_α.

Определение 1.5. Разностью событий A **и** B называется разность $A \setminus B$ множеств A и B (множество ω, принадлежащих A, но не принадлежащих B).

В силу определения 1.2 $A \setminus B$ происходит тогда и только тогда, когда A происходит, а B не происходит.

Определение 1.6. Если $A \subset B$ (A является подмножеством B), то говорят, что **событие** A **влечет событие** B или что B **следует из** A.

Пусть произошло событие A. По определению 1.2 это означает, что осуществился $\omega \in A$. Так как $A \subset B$, то это $\omega \in B$. Значит осуществился элементарный исход $\omega \in B$, т.е. произошло B. Итак, всякий раз, когда происходит событие A, происходит также и событие B. Обратно, пусть последнее утверждение выполняется. Каким бы элементарным исходом ω ни закончилось событие A, то в силу того, что $A \subset B$, это $\omega \in B$. Итак, осуществилось $\omega \in B$, т.е. произошло событие B. Таким образом, $A \subset B$ *тогда и только тогда, когда каждый раз при появлении* A *происходит также и событие* B.

Определение 1.7. События A и B называются **несовместимыми** или **несовместными**, если множества A и B не пересекаются, т.е. $AB = \emptyset$.

Пользуясь определением 1.2, переведем это определение на традиционный вероятностный язык. Событие AB заключается в одновременном осуществлении событий A и B и оно является невозможным. Поэтому A и B несовместимы, если и только если они *не могут произойти* *одновременно.*

Определение 1.8. Событие $\bar{A} = \Omega \setminus A$, являющееся дополнением к множеству A в пространстве Ω, называется **противоположным** к событию A.

С помощью определения 1.2 легко проверяется (обязательно проверьте), что $\bar{A}$ происходит тогда и только тогда, когда A не происходит.

Примем следующее соглашение. Если A и B несовместимы, то сумму этих событий будем обозначать $A + B$. Точно так же, если события семейства $\{A_\alpha, \alpha \in I\}$ попарно несовместимы, то сумму событий $\bigcup\limits_{\alpha\in I} A_\alpha$ будем обозначать $\sum\limits_{\alpha\in I} A_\alpha$. В теоретико-множественной интерпретации это означает, что знаком "плюс" или "$\sum$" обозначается объединение попарно непересекающихся множеств. Если $A \subset B$, то разность $B \setminus A$ будем обозначать $B - A$ (выражаясь фигурально, знак "минус" применяется, когда из события B вычитается "меньшее" событие A).

Определение 1.9. Говорят, что конечный (счетный) набор событий $H_1, H_2, \ldots, H_n$ ($H_1, H_2, \ldots$) **образует полную группу**, если

$$H_1 + H_2 + \ldots + H_n = \Omega \quad (\sum_{k=1}^{\infty} H_k = \Omega).$$

В теории множеств мы бы выразили этот факт так: события образуют полную группу, если они *образуют разбиение пространства элементарных исходов на непере-*

секающиеся части. Говоря, используя определение 1.2, вероятностным языком, устанавливаем, что события образуют полную группу, если *хотя бы одно из них обязательно происходит, но никакие два из них не могут произойти* в результате рассматриваемого эксперимента. Так как H_i — один из возможных противоречащих друг другу вариантов, которыми может закончиться случайный эксперимент, то события, образующие полную группу, часто называют **гипотезами**.

Заметим, что события A и $\bar{A}$, противоположные друг к другу, образуют полную группу из двух гипотез.

Пример 11. Рассмотрим $\Omega = \{1,2,3,4,5,6\}$, соответствующее подбрасыванию "правильной" игральной кости, и пусть $A = \{1,3,5\}$, $B = \{2,4\}$, $H_1 = \{1,2\}$, $H_2 = \{3,4\}$, $H_3 = \{5,6\}$. Противоположным к событию A, состоящему в выпадении нечетного числа очков, является выпадение четного числа очков $\bar{A} = \{2,4,6\}$. События H_1, H_2, H_3 образуют полную группу. События B и H_3 являются несовместимыми. Кроме того $BH_1 = \{2\}$ состоит в выпадении двух очков, а $\bar{A} - B = \{6\}$ состоит в выпадении шести очков.

Поскольку события — подмножества пространства элементарных исходов, а операции над событиями — соответствующие операции над множествами, то для операций над событиями сохраняются все известные тождества из теории множеств. Например, для событий справедливы **соотношения двойственности (формулы де Моргана)**:

$$\overline{AB} = \bar{A} \cup \bar{B}, \quad \overline{A \cup B} = \bar{A}\bar{B};$$

$$\overline{\bigcup_{\alpha\in I} A_\alpha} = \bigcap_{\alpha\in I} \bar{A}_\alpha, \qquad \overline{\bigcap_{\alpha\in I} A_\alpha} = \bigcup_{\alpha\in I} \bar{A}_\alpha.$$

Как нами упоминалось, вероятность — идеализированная относительная частота, т.е. то число, вокруг которого стабилизируется относительная частота при большом числе экспериментов. Поэтому в качестве аксиом вероятности А.Н. Колмогоров предложил взять фундаментальные свойства, которым удовлетворяет относительная частота. Вот они:

I. Относительная частота появления случайного события неотрицательна:

$$P_N(A) \geq 0.$$

Действительно, это следует из определения относительной частоты (В1), так как $N > 0$, а $N(A) \geq 0$ для любого события A.

II. Относительная частота появления достоверного события Ω равна единице:

$$P_N(\Omega) = 1.$$

В самом деле, так как Ω происходит в каждом эксперименте, то $N(\Omega) = N$, и по формуле (В1) $P_N(\Omega) = N(\Omega)/N = N/N = 1$.

III. Если события A и B несовместимы, то относительная частота появления их суммы равна сумме частот появления этих событий:

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

Действительно, так как A и B не могут произойти в одном эксперименте одновременно, то число экспериментов, в которых происходит хоть одно из них, складывается из числа экспериментов, в которых происходит каждое

из них:
$$N(A + B) = N(A) + N(B).$$
Следовательно, по формуле (B1),
$$P_N(A + B) = N(A + B)/N = N(A)/N + N(B)/N =$$
$$= P_N(A) + P_N(B).$$

Свойства I - III являются фундаментальными в том смысле, что любые другие мыслимые свойства относительной частоты могут быть доказаны, исходя из свойств I - III. Например, относительная частота обладает следующим свойством: если $A \subset B$, то $P_N(A) \leq P_N(B)$. Действительно, если всякий раз, когда происходит A, происходит также и B, то, очевидно, $N(A) \leq N(B)$, откуда, в силу (B1), $P_N(A) \leq P_N(B)$. Это свойство следует из свойств I, III, так как из $A \subset B$ следует, что $B = A+(B-A)$, а, по свойству III, $P_N(B) = P_N(A)+P_N(B-A)$. В силу свойства I второе слагаемое в последней сумме является неотрицательным, следовательно $P_N(B) \geq P_N(A)$.

А.Н. Колмогоров предложил аксиомы для вероятности, которые повторяют фундаментальные свойства I - III относительной частоты. При этом он потребовал вместо свойства III выполнение для вероятности более сильного свойства **счетной аддитивности**:
$$\text{IV. } P(\sum_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} P(A_n).$$
Хотя в таком виде аксиома не вытекает из свойств относительной частоты, она необходима для привлечения аппарата современной математики (в частности, теории меры). Без этой аксиомы пропало бы богатство теории

вероятностей и невозможно было бы получить нетривиальные результаты. А приемлемость этой аксиомы подтверждается тем, что результаты, полученные с ее использованием, ни разу не приводили к противоречиям с результатами экспериментов.

В следующем параграфе мы построим вероятностную модель наиболее простых случайных экспериментов, которые описываются дискретным пространством элементарных исходов. Аксиоматическое введение вероятности для случая пространства элементарных исходов произвольной мощности отложим на более позднее время.

§ 2. Вероятность в дискретном пространстве элементарных исходов. Классическое определение вероятности

Напомним, что не более чем счетное пространство элементарных исходов Ω мы назвали **дискретным** и что именно для случайного эксперимента, связанного с таким пространством, были введены события и операции над ними. В этом и следующем параграфах все рассматриваемые пространства элементарных исходов предполагаются дискретными.

Множество $\{\omega\}$, состоящее из единственного элементарного исхода ω, является событием и должно иметь некоторую вероятность $P(\{\omega\})$, которую мы обозначим $p(\omega)$ и назовем **вероятностью элементарного исхода** ω. Так как относительная частота обладает свойством I, мы должны потребовать, чтобы эта вероятность была неотрицательной. В силу свойства II относительной частоты мы должны потребовать, чтобы вероятность достоверного события равнялась единице. Так как $\Omega = \sum\limits_{\omega\in\Omega} \{\omega\}$, то по свойству III (если пространство конечно) или усиленному свойству счетной аддитивности IV (если пространство счетно)

$$1 = P(\Omega) = \sum_{\omega\in\Omega} P(\{\omega\}) = \sum_{\omega\in\Omega} p(\omega).$$

Так как любое событие может быть представлено в виде конечной или счетной суммы одноточечных множеств, из которых оно состоит, т.е. $A = \sum\limits_{\omega\in A}\{\omega\}$, то по свойству III или IV

$$P(A) = \sum_{\omega\in A} p(\omega).$$

Эти соображения приводят к следующему способу введения вероятности в дискретном пространстве элементарных исходов.

Конструктивно-аксиоматическое определение вероятности.

Каждому элементарному исходу ω ставится в соответствие число $p(\omega) \geq 0$, называемое **вероятностью элементарного исхода** ω так, что при этом $\sum\limits_{\omega\in\Omega} p(\omega) = 1$. **Вероятностью произвольного события** A называется сумма вероятностей элементарных исходов, которые благоприятствуют событию A:

$$P(A) = \sum_{\omega\in A} p(\omega). \tag{2.1}$$

Пусть $\mathcal{F}$ — совокупность всех подмножеств Ω, т.е. всех возможных событий. Тройка $(\Omega, \mathcal{F}, P)$, где P — вероятность, введенная с помощью конструктивно-аксиоматического определения, называется **вероятностным пространством** или **вероятностной моделью случайного эксперимента**.

Итак, конструктивно-аксиоматическое определение по формуле (2.1) ставит каждому событию A, т.е. каждому $A \in \mathcal{F}$, в соответствие действительное число $P(A)$ — вероятность этого события. При этом

I. $P(A) \geq 0$ для любого события A.

II. $P(\Omega) = 1$.

III. Если события $A_1, A_2, \ldots$ попарно несовместимы, то

$$P(\sum_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} P(A_n).$$

В самом деле, I следует из того, что $p(\omega) \geq 0$ и (2.1), II следует из того, что $\sum\limits_{\omega \in \Omega} p(\omega) = 1$ и (2.1). Наконец, III следует из того, что

$$P(\sum_{n=1}^{\infty} A_n) = \sum_{\omega \in \sum\limits_{n=1}^{\infty} A_n} p(\omega) = \sum_{\omega \in A_1} p(\omega) + \sum_{\omega \in A_2} p(\omega) + \ldots =$$

$$= P(A_1) + P(A_2) + \ldots .$$

Здесь мы воспользовались тем, что для сходящихся рядов с неотрицательными членами можно как угодно группировать члены ряда в блоки, состоящие из конечного или счетного числа слагаемых.

Обратно, пусть каждому событию A поставлено в соответствие число $P(A)$, т.е. задана функция $P : \mathcal{F} \longrightarrow \mathbb{R}$, которая удовлетворяет аксиомам I – III (аксиоматическое определение вероятности). Тогда, в частности, и каждому одноточечному событию $\{\omega\}$ поставлено в соответствие число $P(\{\omega\})$, которое мы переобозначаем как $p(\omega)$, причем в силу аксиомы I $p(\omega) \geq 0$. Так как $\Omega = \sum\limits_{\omega \in \Omega} \{\omega\}$, то по аксиомам II и III $1 = P(\Omega) = \sum\limits_{\omega \in \Omega} P(\{\omega\}) = \sum\limits_{\omega \in \Omega} p(\omega)$. При этом для события $A = \sum\limits_{\omega \in A} \{\omega\}$ в силу аксиомы III будет выполняться равенство (2.1).

Итак, в случае дискретного пространства элементарных исходов конструктивно-аксиоматическое и аксиоматическое определения вероятности эквивалентны.

Рассмотрим следующий частный случай. Пусть

1) пространство элементарных исходов $\Omega = \{\omega_1,\ \omega_2,\ \dots,\ \omega_n\}$ конечно;

2) элементарные исходы равновозможны (равновероятны), т.е. $p(\omega_1) = p(\omega_2) = \ldots = p(\omega_n)$.

Поскольку $\sum\limits_{\omega_i \in \Omega} p(\omega_i) = p(\omega_1) + p(\omega_2) + \ldots + p(\omega_n) = 1$, а слагаемые в сумме одинаковы, то $p(\omega_i) = 1/n$. Если событию A благоприятствует k элементарных исходов, т.е. $A = \{\omega_{i_1},\ \omega_{i_2},\ \dots,\ \omega_{i_k}\}$, то по конструктивно-аксиоматическому определению (2.1)

$$P(A) = p(\omega_{i_1}) + p(\omega_{i_2}) + \ldots + p(\omega_{i_k}) = \frac{1}{n} + \frac{1}{n} + \ldots + \frac{1}{n} = \frac{k}{n}.$$

Итак, конструктивно-аксиоматическое определение превращается для данного случая в так называемое

Классическое определение вероятности. Если пространство элементарных исходов конечно, а все исходы равновозможны, то **вероятностью события** A называется отношение числа элементарных исходов, благоприятствующих появлению события A, к числу всех возможных элементарных исходов:

$$P(A) = \frac{k}{n}.$$

Пример 1. При подбрасывании "правильной" игральной кости $\Omega = \{1, 2, 3, 4, 5, 6\}$. Пусть $A = \{1, 3, 5\}$, $\bar{A} = \{2, 4, 6\}$ — события, состоящие соответственно в выпадении нечетного и четного числа очков. Каждому из них благоприятствует 3 элементарных исхода. Согласно классическому определению вероятности $P(A) = P(\bar{A}) = 3/6 = 1/2$.

Пример 2. При подбрасывании двух "правильных" игральных костей всего имеется 36 элементарных исходов (см. пример 4 параграфа 1), из которых 6 исходов благоприятствует появлению события $A = \{(1,6), (6,1), (2,5), (5,2), (3,4), (4,3)\}$, состоящего в том, что сумма выпавших на костях очков равна 7. Поэтому $P(A) = 6/36 = 1/6$.

Пример 3. Подбрасывается "правильная" монета до тех пор, пока впервые не выпадет герб. Возможные исходы данного эксперимента таковы: $\omega_1 =$(*г*) — герб выпадает при первом подбрасывании и эксперимент прекращается, $\omega_2 =$(*р,г*) — при первом подбрасывании выпадает решка, а при втором подбрасывании выпадает герб, и эксперимент прекращается, $\omega_3 =$(*р,р,г*) — при первых двух подбрасываниях выпадает решка, а при третьем подбрасывании выпадает герб, и эксперимент прекращается и т.д. Элементарный исход, k-й по счету, имеет вид $\omega_k =(р,р,\ldots,р,г)$ — при первых $k-1$ подбрасываниях выпадает решка, а при k-м подбрасывании выпадает герб, после чего эксперимент прекращается. Теоретически можно допустить еще один исход $\omega_\infty =(р,р,\ldots,р,\ldots)$ — герб никогда не выпадет и эксперимент будет продолжаться бесконечно. На интуитивном уровне такой исход вряд ли возможен и потому должен иметь нулевую вероятность.

Каким же образом определить вероятности элементарных исходов дискретного пространства $\Omega = \{\omega_1, \omega_2, \ldots, \omega_\infty\}$? Для определения вероятности исхода ω_1 рассмотрим случайный эксперимент, состоящий в однократном подбрасывании монеты, в котором вероятность выпадения герба равна $1/2$. Положим $p(\omega_1) = 1/2$. Аналогично, для определения вероятности ω_k рассмотрим случайный эксперимент, состоящий в подбрасывании монеты k раз. Пространство элементарных исходов такого эксперимента состоит из исходов $(i_1, i_2, \ldots, i_k)$, где каждое i_l может принимать только два возможных значения $г$ или $р$, т.е. состоит из 2^k элементарных исходов. В этом пространстве исходу ω_k благоприятствует единственный исход, поэтому его вероятность равна $1/2^k = 2^{-k}$. Положим $p(\omega_k) = 2^{-k}$. Итак, для определения вероятностей исходов вероятностного пространства, соответствующего подбрасыванию монеты до первого выпадения герба, нам пришлось находить эти вероятности в совершенно других вероятностных пространствах, соответствующих k-кратному подбрасыванию монеты при различных k. Пока не будем определять вероятность исхода ω_∞. Согласно конструктивно-аксиоматическому определению вероятности сумма вероятностей всех исходов пространства должна равняться 1, т.е.

$$p(\omega_\infty)+\sum_{k=1}^{\infty} p(\omega_k) = p(\omega_\infty)+\sum_{k=1}^{\infty} 2^{-k} = p(\omega_\infty)+\frac{1/2}{1-1/2} = 1,$$

откуда $p(\omega_\infty) = 0$. Это вполне согласуется с нашими интуитивными представлениями, о которых говорилось выше.

Вероятность события $A = \{\omega_1, \omega_3, \omega_5, \dots\}$, состоящего в том, что эксперимент закончится на нечетном шаге, равна

$$P(A) = p(\omega_1)+p(\omega_3)+p(\omega_5)+\ldots = 1/2+1/2^3+1/2^5+\ldots =$$

$$= \frac{1/2}{1-1/4} = 2/3.$$

Пример 4. В партии из N деталей имеется M стандартных. Наудачу отобрано n деталей. Какова вероятность того, что среди них окажется ровно m стандартных деталей?

Для решения поставленной задачи прежде всего надо найти число всех возможных исходов эксперимента, т.е. число способов выбора n деталей из N деталей. Решить такую задачу без знания элементов комбинаторики очень трудно. Поэтому мы оставим этот пример на будущее. Для тех, кто знаком с комбинаторикой, скажем, что искомая вероятность равна

$$P(A) = \frac{C_M^m \, C_{N-M}^{n-m}}{C_N^n},$$

где C_N^n – число сочетаний из N элементов по n элементов.

§ 3. Элементы комбинаторики

Начнем изложение с двух вспомогательных утверждений — лемм 3.1 и 3.2, которые составляют так называемое **основное правило комбинаторики.**

Лемма 3.1. *Из m элементов первой группы $a_1, a_2, \ldots, a_m$ и n элементов второй группы $b_1, b_2, \ldots, b_n$ можно составить ровно mn различных упорядоченных пар (a_i, b_j), содержащих по одному элементу из каждой группы.*

Доказательство. Выпишем все возможные пары, причем в первой строке укажем все пары, содержащие a_1, во второй строке – все пары, содержащие a_2, и т.д. Получим таблицу

$$(a_1, b_1), (a_1, b_2), (a_1, b_3), \ldots, (a_1, b_n),$$

$$(a_2, b_1), (a_2, b_2), (a_2, b_3), \ldots, (a_2, b_n),$$

$$(a_3, b_1), (a_3, b_2), (a_3, b_3), \ldots, (a_3, b_n),$$

$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

$$(a_m, b_1), (a_m, b_2), (a_m, b_3), \ldots, (a_m, b_n).$$

Поскольку в ней m строк и n столбцов, то всего в ней mn пар. Значит, число всевозможных пар равно mn. $\square$

Пример 1. Рассмотрим две группы элементов: $\spadesuit$ — пики, $\clubsuit$ — трефы, $\diamondsuit$ — бубны, $\heartsuit$ — черви и 6, 7, 8, 9, 10, валет, дама, король, туз. По лемме 3.1 число пар будет равно 4·9=36. Это как раз — число карт в колоде (каждая карта определяется парой элементов (масть, значение)).

Пример 2. Подбрасываются две игральные кости. Элементарный исход определяем как пару $\omega = (i,\, j)$, где i — число очков, выпавших на первой кости, j — число очков, выпавших на второй кости. Тогда i выбирается из группы 1, 2, 3, 4, 5, 6, j выбирается из этой же группы. По лемме 3.1 число всех элементарных исходов (т.е. всевозможных пар $(i,\, j)$) будет равно $6 \cdot 6 = 6^2 = 36$.

Лемма 3.2. *Из n_1 элементов первой группы $a_1, a_2, \ldots, a_{n_1}$, n_2 элементов второй группы $b_1, b_2, \ldots, b_{n_2}$, и т.д., наконец, n_k элементов k-й группы $x_1, x_2, \ldots, x_{n_k}$ можно составить ровно $n_1 \cdot n_2 \cdot \ldots \cdot n_k$ различных упорядоченных комбинаций вида $(a_{j_1}, b_{j_2}, \ldots, x_{j_k})$, содержащих по одному элементу из каждой группы.*

Доказательство. Доказательство проведем с помощью математической индукции по числу групп k. При $k = 2$ лемма 3.2 справедлива, так как превращается в лемму 3.1. Предположим, что лемма 3.2 справедлива для k групп, фигурирующих в ее условии. Добавим к ним еще одну $(k+1)$-ю группу $y_1,\, y_2,\, \ldots,\, y_{n_{k+1}}$. Комбинацию $(a_{j_1}, b_{j_2}, \ldots, x_{j_k}, y_{j_{k+1}})$ можно рассматривать как пару, состоящую из сложного элемента $(a_{j_1}, b_{j_2}, \ldots, x_{j_k})$ и простого элемента $y_{j_{k+1}}$. По предположению индукции число сложных элементов равно $n_1 \cdot n_2 \cdot \ldots \cdot n_k$, число простых элементов равно n_{k+1}. По лемме 3.1 число таких пар, а, значит, и число комбинаций элементов из $k + 1$ групп равно $(n_1 \cdot n_2 \cdot \ldots \cdot n_k) \cdot n_{k+1} = n_1 \cdot n_2 \cdot \ldots \cdot n_k \cdot n_{k+1}$. Итак, предположив, что лемма 3.2 верна для k групп, мы доказали, что она верна и для $k+1$ групп. На основании принципа математической индукции это завершает доказательство. □

Пример 3. Подбрасываются "наудачу" три игральные кости; элементарный исход $\omega = (i, j, k)$, где i — число выпавших очков на 1-й, j — число выпавших очков на 2-й, k — число выпавших очков на 3-й кости. По лемме 3.2 число таких элементарных исходов равно $6 \cdot 6 \cdot 6 = 6^3 = 216$.

Пример 4. Из пункта A в пункт B проходит 15 дорог, из пункта B в пункт C — 4 дороги, из пункта C в пункт D проходит 5 дорог. При этом все дороги, ведущие из A в D, проходят сначала через B, а затем через C. По лемме 3.2 из пункта A в пункт D проходит $15 \cdot 4 \cdot 5 = 300$ дорог.

Отметим, что на языке теории множеств лемму 3.2 можно переформулировать следующим образом.

Лемма 3.2′**.** *Если $A_1, A_2, \ldots, A_k$ — конечные множества с мощностями $n_1, n_2, \ldots, n_k$ соответственно, то мощность декартового произведения $A_1 \times A_2 \times \ldots \times A_k$ равна $n_1 \cdot n_2 \cdot \ldots \cdot n_k$.*

Предположим, что имеется конечное множество $\{a_1, a_2, \ldots, a_n\}$, из которого выбирается k элементов, которые мы запишем в порядке их появления как $(a_{j_1}, a_{j_2}, \ldots, a_{j_k})$ и назовем **выборкой объема** k **из** n **элементов**. По способу выбора элементов выборки делятся на **выборки с возвращением** и **без возвращения**. При выборе с возвращением сначала выбирается элемент a_{j_1} из первоначального множества, затем элемент a_{j_1} возвращается в него, после чего элемент a_{j_2} также выбирается из первоначального множества и т.д. При этом объем выборки k может принимать любые значения из множества $\{1, 2, \dots\}$. При выборе без возвращения после выбора первого элемента a_{j_1} элемент a_{j_2} выбирается из оставшихся элементов в первоначальном множестве и т.д. В таком случае всегда $k \leq n$, так как после выбора n элементов в первоначальном множестве ничего не останется и дальнейший выбор становится невозможным. В зависимости от того, учитывается в выборке порядок следования элементов или не учитывается, выборки делятся на упорядоченные и неупорядоченные. Если выборки с одинаковым составом элементов, отличающиеся порядком их следования, считать различными, то такие выборки называются **упорядоченными**. Если отождествлять выборки с одинаковым составом, то такие выборки называются **неупорядоченными**. Таким образом, возможны четыре типа выборок: упорядоченные с возвращением, упорядоченные без возвращения, неупорядоченные с возвращением и неупорядоченные без возвращения. Для наглядности в таблице 1 указаны все возможные выборки объема $k = 2$ из множества $\{1, 2, 3\}$. Обозначим числа возможных выборок каждого типа таким образом, как указано в таблице 2.

По другому упорядоченные выборки без возвращения называются размещениями, а неупорядоченные выборки без возвращения — сочетаниями. Таким образом, A_n^k — число размещений из n элементов по k, C_n^k — число сочетаний из n элементов по k.

Табл. 1. Выборки объема 2 из множества $\{1, 2, 3\}$.

Выборки	упорядоченные	неупорядоченные
с возвращением	(1,1) (1,2) (1,3) (2,1) (2,2) (2,3) (3,1) (3,2) (3,3)	(1,1) (1,2) (1,3) (2,2) (2,3) (3,3)
без возвращения	(1,2) (1,3) (2,1) (2,3) (3,1) (3,2)	(1,2) (1,3) (2,3)

Табл. 2. Числа выборок объема k из множества $\{a_1, a_2, \ldots, a_n\}$.

Выборки	упорядоченные	неупорядоченные
с возвращением	$\overline{A_n^k}$	$\overline{C_n^k}$
без возвращения	A_n^k	C_n^k

Упорядоченные выборки с возвращением по другому называют размещениями с повторениями, а неупорядоченные выборки с возвращением — сочетаниями с повторениями. Поэтому $\overline{A_n^k}$ — число размещений из n элементов по k с повторениями, $\overline{C_n^k}$ — число сочетаний из n элементов по k с повторениями.

Выведем формулы для подсчета числа выборок каждого типа. Рассмотрим произвольную выборку объема k из n элементов $(a_{j_1}, a_{j_2}, \ldots, a_{j_k})$. Если выборка упорядоченная с возвращением, то каждый из ее элементов выбирается из группы $a_1, a_2, \ldots, a_n$, содержащей

n элементов. По лемме 3.2 число таких выборок равно $n \cdot n \cdot \ldots \cdot n = n^k$. Таким образом,

$$\overline{A_n^k} = n^k. \tag{3.1}$$

Если выборка упорядоченная без возвращения, то ее первый элемент a_{j_1} выбирается из группы $a_1, a_2, \ldots, a_n$, состоящей из n элементов, второй ее элемент a_{j_2} выбирается из оставшейся после этого группы из $n-1$ элементов и т.д. Последний элемент выборки a_{j_k} выбирается из группы $n-(k-1) = n-k+1$ элементов, оставшихся в первоначальной группе после удаления из нее $k-1$ элементов для образования предыдущих элементов $a_{j_1}, a_{j_2}, \ldots, a_{j_{k-1}}$ выборки. По лемме 3.2 число таких выборок равно $n \cdot (n-1) \cdot \ldots \cdot (n-k+1)$. Поэтому

$$A_n^k = n \cdot (n-1) \cdot (n-2) \cdot \ldots \cdot (n-k+1). \tag{3.2}$$

Умножив и разделив (3.2) на $(n-k)! = (n-k)(n-k-1)\ldots 2 \cdot 1$, получим другую формулу для подсчета числа размещений:

$$A_n^k = \frac{n!}{(n-k)!}. \tag{3.3}$$

Совокупность k различных элементов, записанных в произвольном порядке, называется **перестановкой** из k элементов. Для образования перестановки необходимо выбрать без возвращения все k элементов и записать их в порядке появления. Поэтому число перестановок из k элементов равно числу упорядоченных выборок объема k из k элементов:

$$P_k = A_k^k = k! \tag{3.4}$$

Рассмотрим произвольное сочетание $(a_{j_1}, a_{j_2}, \ldots, a_{j_k})$. Ему соответствует ровно P_k различных размещений с таким же составом элементов.

Поэтому число размещений из n элементов по k в P_k раз больше числа сочетаний из n элементов по k:

$$A_n^k = C_n^k \cdot P_k, \text{ откуда } C_n^k = \frac{A_n^k}{P_k}. \tag{3.5}$$

Подставляя (3.2)-(3.4) в соотношение (3.5), получим следующие две формулы для подсчета числа сочетаний из n по k:

$$C_n^k = \frac{n \cdot (n-1) \cdot (n-2) \cdot \ldots \cdot (n-k+1)}{k!}, \tag{3.6}$$

$$C_n^k = \frac{n!}{k!\,(n-k)!}. \tag{3.7}$$

Рассмотрим произвольное сочетание с повторением (неупорядоченную выборку с возвращением) $(a_{j_1}, a_{j_2}, \ldots, a_{j_k})$. Предположим, что в ней элемент a_1 встречается k_1 раз, элемент a_2 встречается k_2 раз, и т.д., наконец элемент a_n встречается k_n раз. Очевидно, что $k_1 + k_2 + \ldots + k_n = k$ и что каждое сочетание с повторением полностью определяется набором $(k_1, k_2, \ldots k_n)$, удовлетворяющим этому условию. Закодируем этот набор последовательностью единиц и нулей, ставя сначала k_1 единиц, затем один разделительный ноль, затем ставя k_2 единиц, затем один разделительный ноль и т.д., наконец ставя разделительный ноль, а после него число единиц, равное k_n. Между наборами и кодами, а следовательно, между сочетаниями с повторениями и кодами устанавливается таким образом биекция. Например, сочетанию с повторениями $(a_1, a_1, \ldots, a_1)$ соответствует код $1, 1, \ldots, 1, 0, 0, \ldots, 0$, состоящий из k единиц и $n-1$ нулей, а сочетанию с повторениями $(a_n, a_n, \ldots, a_n)$ соответствует код $0, 0, \ldots, 0, 1, 1, \ldots, 1$, состоящий из $n-1$ нулей и k единиц. Поэтому число сочетаний с повторениями совпадает с числом всех кодов. В каждом коде содержится $n+k-1$ символов: k единиц и $n-1$ нулей. Следовательно, число кодов равно числу способов, с помощью которых для $n+k-1$ вакантных мест можно выбрать k мест для единиц, а остальные места заполнить нулями. Так как эти места не упорядочиваются и выбираются без возвращения, то их число равно C_{n+k-1}^k. Таким образом, число сочетаний с повторениями из n по k равно:

$$\overline{C_n^k} = C_{n+k-1}^k. \tag{3.8}$$

Пример 5. Сколькими способами можно купить 7 пирожных, если имеются пирожные 4 сортов? Предполагая, что в магазине имеется не менее 7 пирожных каждого сорта, мы можем считать, что множество состоит из 4 пирожных (по одному каждого сорта) и мы выбираем из него с возвращением 7 пирожных. Очевидно, выборки неупорядоченные, так как для нас не имеет значения, в каком порядке выбираются пирожные. Поэтому пирожные можно купить

$$\overline{C_4^7} = C_{4+7-1}^7 = C_{10}^7 = C_{10}^3 = \frac{10 \cdot 9 \cdot 8}{2 \cdot 3} = 120$$

способами.

Свойства сочетаний.

1. $C_n^k = C_n^{n-k}$.
2. $C_n^0 = C_n^n = 1$.
3. $C_n^1 = C_n^{n-1} = n$.
4. $C_n^{k-1} + C_n^k = C_{n+1}^k$.

Свойства 1-3 следуют непосредственно из формулы (3.7) или (3.6). Свойство 4 выполняется, так как на основании (3.7)

$$C_n^{k-1} + C_n^k = \frac{n!}{(k-1)!\,(n-k+1)!} + \frac{n!}{k!\,(n-k)!} =$$

$$= \frac{n!}{k!\,(n-k+1)!}\,[k+(n-k+1)] == \frac{(n+1)!}{k!\,(n+1-k)!} = C_{n+1}^k.$$

Отметим, что свойство 1 удобно применять при вычислениях "вручную", когда используется формула (3.6) и $k > n/2$. Множители в числителе (3.6) начинаются с n и убывают на 1, а их число равно k. Например, поскольку $6 > 8/2$, то

$$C_8^6 = C_8^2 = \frac{8 \cdot 7}{2} = 28.$$

Примем по определению, что $C_0^0 = 1$. На свойстве 4 основано построение так называемого треугольника Паскаля — таблицы, позволяющей быстро находить C_n^k при не очень больших n и k. Она имеет следующий вид:

$$
\begin{array}{ccccccccccc}
 & & & & & C_0^0 & & & & & \\
 & & & & C_1^0 & & C_1^1 & & & & \\
 & & & C_2^0 & & C_2^1 & & C_2^2 & & & \\
 & & C_3^0 & & C_3^1 & & C_3^2 & & C_3^3 & & \\
 & C_4^0 & & C_4^1 & & C_4^2 & & C_4^3 & & C_4^4 & \\
C_5^0 & & C_5^1 & & C_5^2 & & C_5^3 & & C_5^4 & & C_5^5 \\
\multicolumn{11}{c}{\dotfill}
\end{array}
$$

В силу свойства 2 по боковым сторонам треугольника пишем единицы, а на основании свойства 4 каждое число внутри треугольника получаем сложением двух соседних лежащих над ним чисел:

$$
\begin{array}{ccccccccccc}
 & & & & & 1 & & & & & \\
 & & & & 1 & & 1 & & & & \\
 & & & 1 & & 2 & & 1 & & & \\
 & & 1 & & 3 & & 3 & & 1 & & \\
 & 1 & & 4 & & 6 & & 4 & & 1 & \\
1 & & 5 & & 10 & & 10 & & 5 & & 1 \\
\multicolumn{11}{c}{\dotfill}
\end{array}
$$

Теорема 3.3. *Справедлива формула бинома Ньютона*

$$(x+y)^n = \sum_{k=0}^{n} C_n^k \, x^k \, y^{n-k}. \tag{3.9}$$

Доказательство. Доказательство проведем с помощью математической индукции по n. При $n = 1$ формула (3.9), очевидно, верна. Предположим, что она верна для натурального числа n. Докажем, что она верна для $n+1$. Действительно, в силу свойств 2 и 4 сочетаний,

$$(x+y)^{n+1} = (x+y)\cdot(x+y)^n = (x+y)\sum_{k=0}^{n} C_n^k \, x^k \, y^{n-k} =$$

$$= \sum_{k=0}^{n} C_n^k \, x^{k+1} \, y^{n-k} + \sum_{k=0}^{n} C_n^k \, x^k \, y^{n-k+1} = x^{n+1} + y^{n+1} +$$

$$+ \sum_{k=0}^{n-1} C_n^k \, x^{k+1} \, y^{n-k} + \sum_{k=1}^{n} C_n^k \, x^k \, y^{n-k+1} = x^{n+1} + y^{n+1} +$$

$$+ \sum_{k=1}^{n} C_n^{k-1} \, x^k \, y^{n-k+1} + \sum_{k=1}^{n} C_n^k \, x^k \, y^{n-k+1} = x^{n+1} + y^{n+1} +$$

$$+ \sum_{k=1}^{n} (C_n^{k-1} + C_n^k) \, x^k \, y^{n-k+1} = x^{n+1} + y^{n+1} +$$

$$+ \sum_{k=1}^{n} C_{n+1}^k \, x^k \, y^{(n+1)-k} = \sum_{k=0}^{n+1} C_{n+1}^k \, x^k \, y^{(n+1)-k}.$$

На основании принципа математической индукции заключаем, что доказываемая формула верна для любого натурального числа n. □

Теорема 3.4. *Имеет место полиномиальное разложение*

$$(x_1 + x_2 + \ldots + x_k)^n =$$

$$= \sum_{r_1+r_2+\ldots+r_k=n} \frac{n!}{r_1!\,r_2!\,\ldots\,r_k!}\, x_1^{r_1}\, x_2^{r_2}\, \ldots\, x_k^{r_k}, \tag{3.10}$$

где суммирование ведется по всем целым неотрицательным числам $r_1, r_2, \ldots, r_k$, *в сумме дающим* n.

Доказательство. Доказательство проведем с помощью математической индукции по k. При $k = 2$ формула, очевидно, верна, так как превращается в формулу бинома Ньютона (3.9). Предположим, что доказываемое равенство выполнено для натуральных чисел, меньших либо равных k. Докажем, что оно верно для числа $k+1$. Действительно, в силу формулы бинома Ньютона (3.9),

$$(x_1 + x_2 + \ldots + x_k + x_{k+1})^n = \sum_{s=0}^{n} C_n^s \Bigl(\sum_{i=1}^{k} x_i\Bigr)^s x_{k+1}^{n-s} =$$

$$= \sum_{s=0}^{n} \frac{n!}{s!\,(n-s)!} \Bigl(\sum_{r_1+\ldots+r_k=s} \frac{s!}{r_1!\,\ldots\,r_k!} x_1^{r_1}\,\ldots\,x_k^{r_k} \Bigr)\, x_{k+1}^{n-s}.$$

Сделаем замену индексов суммирования, вводя $r_{k+1} = n - s$. Тогда при изменении s от 0 до n число r_{k+1} будет изменяться от n до 0. Поэтому

$$(x_1 + x_2 + \ldots + x_k + x_{k+1})^n = \sum_{r_{k+1}=0}^{n} \frac{n!}{(n-r_{k+1})!\,r_{k+1}!} \cdot$$

$$\cdot \sum_{r_1+r_2+\ldots+r_k=n-r_{k+1}} \frac{(n-r_{k+1})!}{r_1!\,r_2!\,\ldots\,r_k!}\, x_1^{r_1}\, x_2^{r_2}\, \ldots\, x_k^{r_k}\, x_{k+1}^{r_{k+1}} =$$

$$= \sum_{r_{k+1}=0}^{n} \sum_{r_1+\ldots+r_k=n-r_{k+1}} \frac{n!}{r_1! \ldots r_k! \, r_{k+1}!} x_1^{r_1} \ldots x_k^{r_k} \, x_{k+1}^{r_{k+1}} =$$

$$= \sum_{r_1+r_2+\ldots+r_{k+1}=n} \frac{n!}{r_1! \, r_2! \ldots r_{k+1}!} x_1^{r_1} x_2^{r_2} \ldots x_{k+1}^{r_{k+1}}.$$

Теорема доказана. □

Теорема 3.5. *Пусть $r_1, r_2, \ldots, r_k$ — целые неотрицательные числа, причем $r_1 + r_2 + \ldots + r_k = n$. Тогда число способов, с помощью которых n элементов можно разбить на k групп так, чтобы в первую группу вошло r_1 элементов, во вторую группу — r_2 элементов и т.д., наконец в k-ю группу вошло r_k элементов, равно*

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

Доказательство. Будем выбирать сначала r_1 элементов первой группы из n злементов, что можно сделать $C_n^{r_1}$ способами. Остается $n - r_1$ злементов, из которых r_2 элементов второй группы можно выбрать $C_{n-r_1}^{r_2}$ способами. После этого остается $n-r_1-r_2$ элементов, из которых r_3 элементов третьей группы можно выбрать $C_{n-r_1-r_2}^{r_3}$ способами. Продолжая далее такой процесс, мы после выбора $(k-2)$-й группы будем иметь $n-r_1-r_2-\ldots-r_{k-2}$ элементов, из которых r_{k-1} элементов предпоследней $(k-1)$-й группы можно выбрать $C_{n-r_1-r_2-\ldots-r_{k-2}}^{r_{k-1}}$ способами. Останутся $n - r_1 - r_2 - \ldots - r_{k-1} = r_k$ элементов, которые единственным способом образуют последнюю k-ю группу. По основному правилу комбинаторики (лемме 3.2) число разбиений n элементов на k групп указанным в формулировке теоремы способом будет равно

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

$$= \frac{n!}{r_1!\,(n-r_1)!}\,\frac{(n-r_1)!}{r_2!\,(n-r_1-r_2)!}\,\frac{(n-r_1-r_2)!}{r_3!\,(n-r_1-r_2-r_3)!}\,\ldots$$

$$\ldots\frac{(n-r_1-r_2-\ldots-r_{k-2})!}{r_{k-1}!\,(n-r_1-r_2-\ldots-r_{k-1})!} = \frac{n!}{r_1!\,r_2!\,\ldots\,r_k!}.$$

□

Теорема 3.5 устанавливает смысл полиномиальных коэффициентов в разложении (3.10).

Пример 6. 1. Сколькими способами n различимых дробинок можно разместить по k занумерованным урнам так, чтобы в 1-ю урну попало r_1 дробинок, во 2-ю урну — r_2 дробинок и т.д., наконец, в k-ю урну попало r_k дробинок $(r_1 + r_2 + \ldots + r_k = n)$?

2. Сколькими способами n различимых дробинок можно разместить по k занумерованным урнам?

3. Сколькими способами n неразличимых дробинок можно разместить по k занумерованным урнам?

По теореме 3.5 число способов пункта 1 примера равно

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

Для ответа на пункт 2 примера необходимо сложить числа способов пункта 1 по всем целым неотрицательным $r_1, r_2, \ldots, r_k$, дающим в сумме n:

$$\sum_{r_1+r_2+\ldots+r_k=n} \frac{n!}{r_1!\,r_2!\,\ldots\,r_k!}.$$

Но по формуле (3.10) эта сумма равна

$$(1+1+1+\ldots+1)^n = k^n.$$

Таким образом, число способов в пункте 2 равно k^n. Можно получить этот результат более простым способом. Пусть различимые дробинки занумерованы как

$a_1,\, a_2,\, \ldots\, ,\, a_n$. Исход рассматриваемого эксперимента однозначно определяется набором $(i_1,\, i_2,\, \ldots\, ,\, i_n)$, где i_l — номер урны, в которую попала дробинка a_l. Каждое i_l выбирается из множества k урн с возвращением и упорядоченным образом. Поэтому число исходов такого эксперимента совпадает с числом упорядоченных выборок объема n с возвращением из k элементов, т.е. равно

$$\overline{A_k^n} = k^n.$$

Для ответа на пункт 3 заметим, что исход эксперимента определяется таким же набором, как и в пункте 2, за исключением того, что выборка неупорядоченная. Поэтому число исходов такого эксперимента совпадает с числом неупорядоченных выборок объема n с возвращением из k элементов, т.е. равно

$$\overline{C_k^n} = C_{k+n-1}^n.$$

Вернемся к примерам на классическое определение вероятности. В предыдущем параграфе рассматривался следующий

Пример 7. В партии из N деталей имеется M стандартных. Наудачу отобрано n деталей. Какова вероятность того, что среди них окажется ровно m стандартных деталей?

Любую задачу на классическое определение вероятности следует начинать с построения пространства элементарных исходов Ω. Подразумевается, что детали выбираются без возвращения. Когда выбирается n деталей из деталей $a_1, a_2, \ldots\, , a_M,\, b_1, b_2, \ldots\, , b_{N-M}$, для нас порядок, в котором они появляются, не имеет значения, важно лишь какие детали появятся (здесь для удобства стандартные детали обозначены буквой a, а нестандартные — буквой b). Поэтому элементарными исходами являются неупорядоченные выборки объема n из N элементов (сочетания). Следовательно, всего в пространстве Ω имеется C_N^n исходов. Благоприятные исходы имеют вид $(a_{j_1}, a_{j_2} \dots , a_{j_m}, b_{i_1}, b_{i_2}, \dots , b_{i_{n-m}})$. Стандартные детали можно выбрать C_M^m способами, нестандартные — C_{N-M}^{n-m}. По основному правилу комбинаторики (лемме 3.2) число благоприятных исходов равно

$$C_M^m C_{N-M}^{n-m}.$$

Искомая вероятность равна

$$P(A) = \frac{C_M^m C_{N-M}^{n-m}}{C_N^n}.$$

Набор вероятностей $\{P(m),\ m = m_0, m_0+1, \dots , m_1\}$, где

$$P(m) = \frac{C_M^m C_{N-M}^{n-m}}{C_N^n},$$

$$m_0 = \max(0, M - N + n),\ m_1 = \min(M, n),$$

называется **гипергеометрическим распределением**.

Пример 8. При игре в спортлото из 49 видов спорта, занумерованных как 1, 2, ... , 49, зачеркиваются любые 6 видов. При розыгрыше 49 перенумерованных шаров тщательно перемешиваются, после чего из них "наудачу" выбирается первый шар. Оставшиеся шары вновь перемешиваются, после чего из них "наудачу" выбирается второй шар и т.д. до тех пор, пока не будет выбрано 6 шаров. Если все 6 зачеркнутых Вами номеров совпадут с разыгранными, то Вы получаете максимальный выигрыш, если из зачеркнутых номеров ровно три любые номера совпадут с какими то тремя из разыгранных, то Вам достанется минимальный выигрыш. Какова вероятность максимального и минимального выигрыша на один билет спортлото?

Очевидно, данная задача — частный случай предыдущей. Действительно, у нас имеется 49 видов спорта (деталей), из них 6 выигрышных (стандартных). Выбирается 6 номеров спорта (деталей). Если событие A состоит в получении максимального, а событие B — в получении минимального выигрыша, то A — из выбранных 6 деталей все 6 — стандартные, B — из выбранных 6 деталей ровно 3 стандартные. Следовательно,

$$P(A) = \frac{1}{C_{49}^6} = 7.151 \cdot 10^{-8}; \quad P(B) = \frac{C_6^3 \cdot C_{43}^3}{C_{49}^6} = 0.017650.$$

Пример 9. В лифт 11-этажного дома на первом этаже вошло 7 человек. Предполагается, что каждый может выйти на любом из этажей, начиная со второго. Какова вероятность того, что а) все выйдут на 7-м этаже; б) все выйдут на одном и том же этаже; в) все выйдут на разных этажах?

Исход данного вероятностного эксперимента определяется указанием номеров этажей, выбранных каждым из лиц. Поэтому в качестве элементарного исхода возьмем $\omega = (a_1, a_2, \dots, a_7)$ — выборку из множества всех допустимых этажей $\{2, 3, \dots, 11\}$. При этом выборки следует рассматривать как упорядоченные ("Коля вышел на 2-м, а Петя — на 3-м" и "Петя вышел на 2-м, а Коля — на 3-м" есть не одно и то же) и с возвращением (некоторые лица в принципе могут выйти на одном и том же этаже). Итак, $\Omega = \{(a_1, a_2, \dots, a_7) : \quad a_i =$

2, 3, ... , 11 ($i = 1, 2, \ldots, 7$)}. Поэтому число всех возможных исходов равно

$$n = \overline{A_{10}^7} = 10^7.$$

Пусть

$$A = \{\text{все выйдут на 7 этаже}\},$$

$$B = \{\text{все выйдут на одном и том же этаже}\},$$

$$C = \{\text{все выйдут на разных этажах}\}.$$

Событию A благоприятствует единственный исход $(7, 7, 7, 7, 7, 7, 7)$, событию B — 10 исходов $(2, 2, 2, 2, 2, 2, 2)$, $(3, 3, 3, 3, 3, 3, 3)$, $\ldots$, $(11, 11, 11, 11, 11, 11, 11)$, событию C — A_{10}^7 исходов. Следовательно,

$$P(A) = \frac{1}{10^7}; \quad P(B) = \frac{10}{10^7} = 10^{-6}; \quad P(C) = \frac{A_{10}^7}{10^7} = 0.06.$$

Пример 10. В чулане находится n пар ботинок. Из них наудачу выбираются $2r$ ботинок ($2r < n$). Какова вероятность того, что среди выбранных ботинок отсутствуют парные?

Разумеется выбор ботинок осуществляется без возвращения. Нас интересует состав выбранных ботинок; порядок, в котором они выбираются, не имеет значения. Поэтому надо пользоваться неупорядоченными выборками без возвращения, т.е. сочетаниями. Всего $2r$ ботинок из $2n$ ботинок можно выбрать C_{2n}^{2r} способами. Для подсчета числа элементарных исходов, благоприятствующих событию $A = \{$среди выбранных ботинок отсутствуют парные$\}$, мысленно выберем $2r$ пар из имеющихся n пар, что можно сделать C_n^{2r} способами, после чего в каждой паре можно выбрать один из ботинков 2 способами. По основному правилу комбинаторики число благоприятных исходов равно $2^{2r}C_n^{2r}$. Поэтому искомая вероятность равна

$$P(A) = \frac{2^{2r}C_n^{2r}}{C_{2n}^{2r}}.$$

При решении задач на классическое определение вероятности пространство элементарных исходов можно выбирать по разному.

Пример 11. n друзей наудачу рассаживаются за круглым столом. Найти вероятность того, что два фиксированных лица A и B сядут рядом, причем B слева от A.

1 способ. Всего n друзей можно рассадить $n!$ способами. Лицо A можно посадить на любое из n мест n способами, тогда B садится слева от A единственным способом; остаются $n-2$ лица, которые можно разместить на оставшиеся $n-2$ места $(n-2)!$ способами. Поэтому число благоприятных исходов равно $n \cdot 1 \cdot (n-2)!$. искомая вероятность есть

$$P(A) = \frac{n(n-2)!}{n!} = \frac{1}{n-1}.$$

2 способ. Интуитивно понятно, что лицо A можно посадить на фиксированное место. Тогда остальные лица можно разместить $(n-1)!$ способами. Лицо B можно посадить слева от A единственным способом; остаются $n-2$ лица, которые можно разместить на оставшиеся $n-2$ места $(n-2)!$ способами. Поэтому число благоприятных исходов равно $1 \cdot (n-2)!$, а искомая вероятность есть

$$P(A) = \frac{(n-2)!}{(n-1)!} = \frac{1}{n-1}.$$

3 способ. Интуитивно понятно, что размещение лиц, отличных от A и B не имеет значения. Всего лица A и B можно рассадить $A_n^2 = n(n-1)$ способами. Для благоприятного случая лицо A можно посадить n способами, тогда B единственным способом садится слева от A. Поэтому число благоприятных исходов равно $n \cdot 1$. Искомая вероятность есть

$$P(A) = \frac{n}{n(n-1)} = \frac{1}{n-1}.$$

4 способ. Интуитивно ясно, что лицо A можно посадить на фиксированное место и интересоваться, как относительно него сядет B, не рассматривая способы размещения остальных лиц. Лицо B можно посадить на любое из оставшихся после A мест $n-1$ способами; благоприятный способ — один, когда B сядет слева от A. Искомая вероятность есть

$$P(A) = \frac{1}{n-1}.$$

Приведем пример, показывающий предпочтительность решения задачи из примера 11 первым способом.

Пример 12. Наудачу подбрасываются две правильные монеты. Какова вероятность выпадения разных сторон на этих монетах?

1 способ. В качестве пространства элементарных исходов возьмем $\Omega = \{(г,г),(г,р),(р,г),(р,р)\}$, где первая позиция в паре — исход подбрасывания первой монеты, вторая позиция — исход подбрасывания второй монеты. Событию $A = \{$выпадение разных сторон$\}$ благоприятствует два исхода $(г,р),(р,г)$. Поэтому $P(A) = \frac{2}{4} = \frac{1}{2}$.

2 способ. В качестве пространства элементарных исходов возьмем $\tilde{\Omega} = \{2г, 2р, смесь\}$, где 2*г* — выпадение двух "гербов", $2p$ – выпадение двух "решек", *смесь* — выпадение на одной из монет "герба", а на другой — "решки". Событию A благоприятствует единственный исход "смесь", поэтому его вероятность $P(A) = \frac{1}{3}$.

Разумеется 2 способ решения неправомочен, поскольку проведенная факторизация Ω (исходы (*г,р*) и (*р,г*) объединены в исход "смесь") привела к тому, что в пространстве $\tilde{\Omega}$ исходы не равновероятны. Следовательно, классическое определение неприменимо.

Возвращаясь к примеру 11, отметим, что факторизация пространства элементарных исходов способа 1 в способах 2-4 заключалась в объединении в новые исходы одинакового числа равновероятных исходов. Поэтому новые исходы оказались также равновероятными. Чтобы не оказаться в ситуации 2-го способа примера 12, рекомендуется в качестве элементарных исходов сразу выбирать "наиболее мелкие, неделимые" исходы случайного эксперимента.

§ 4. Геометрическая вероятность. Аксиоматика теории вероятностей

Предположим, что на отрезок $\Omega = [a, b]$ "наудачу" бросается точка. Под словами "наудачу" подразумевается тот факт, что вероятности попадания точки в различные части отрезка одинаковой длины совпадают. В частности, должны совпадать вероятности попадания в одноточечные множества. Припишем каждому одноточечному множеству $\{\omega\}$ вероятность $P(\{\omega\}) = p(\omega) = p$, подобно тому, как это делалось при конструктивно-аксиоматическом определении вероятности в дискретном пространстве элементарных исходов. Если A — множество рациональных чисел отрезка $[a, b]$, то в силу того, что $A \subset [a, b]$, имеем

$$P(A) \leq P(\Omega) = 1. \tag{4.1}$$

Если $p > 0$, то $P(A) = \sum\limits_{\omega \in A} p(\omega) = p + p + p + \ldots = \infty$, что противоречит (4.1). Поэтому одноточечным множествам мы должны приписать нулевую вероятность. В силу счетной аддитивности вероятности нулевую вероятность будут иметь все не более чем счетные подмножества отрезка $[a, b]$. Кроме того, единичную вероятность будет иметь Ω и все его подмножества, отличающиеся от него на не более чем счетное множество точек. Для определения вероятностей всех иных несчетных подмножеств Ω информации о том, что $p = 0$, недостаточно, так как операция суммирования определяется в математике только для конечного (обычная сумма) и счетного (сумма ряда) числа слагаемых. В частности, мы не можем определить вероятность попадания брошенной точки на отрезок $[a, \frac{a+b}{2}]$, хотя интуитивно ясно, что эта вероятность равна $\frac{1}{2}$. При классическом определении в качестве вероятности бралось отношение числа благоприятных исходов к числу всех возможных исходов. В рассматриваемом случае оба этих числа бесконечны. Поэтому естественно заменить эти числа на соответствующие длины. Обобщением понятия длины является мера Лебега. Однако не все подмножества отрезка $[a, b]$ измеримы по Лебегу. Следовательно, мы не можем определить вероятность брошенной точки в множество, не измеримое по Лебегу. Естественно поэтому событиями называть не все подмножества $[a, b]$, а только те из них, которые являются измеримыми. Хорошо известно, что измеримые подмножества отрезка $[a, b]$ образуют σ-алгебру, которую мы обозначим $\mathcal{F}$. Приходим к так называемому **геометрическому определению вероятности**.

Определение 4.1. Событием называется элемент A σ-алгебры $\mathcal{F}$. **Вероятностью попадания** наудачу брошенной на отрезок $[a, b]$ точки **в множество** A называется число

$$P(A) = \frac{\mu(A)}{b-a},$$

где $\mu(A)$ — мера Лебега множества A.

Это определение естественным образом обобщается на случай, когда точка бросается в n-мерном пространстве. Пусть Ω — множество в $\mathbb{R}^n$, имеющее конечную n-мерную меру Лебега $\mu(\Omega)$, $\mathcal{F}$ — σ-алгебра измеримых по Лебегу подмножеств множества Ω. Элементы $\mathcal{F}$ называются **событиями**. Если n-мерная точка "наудачу" бросается в Ω, то **вероятность ее попадания** в $A \in \mathcal{F}$ определяется в соответствии с правилом **геометрической вероятности**:

$$P(A) = \frac{\mu(A)}{\mu(\Omega)},$$

где μ — n-мерная мера Лебега.

Отметим, что в приложениях как правило встречаются множества, имеющие n-мерный обем, который и выступает в качестве меры Лебега. Например, в $\mathbb{R}^1$ это есть длина, в $\mathbb{R}^2$ — площадь, в $\mathbb{R}^3$ — объем.

Пример 1. Студент и студентка договорились встретиться в определенном месте между 12 и 13 часами. Пришедший ждет другого в течение 20 минут, после чего уходит. Предполагая, что каждый из них выбирает момент своего прихода в течение указанного часа "наудачу", найдите вероятность того, что встреча состоится.

Для удобства будем называть 12 и 13 часов соответственно 0 и 1 часами дня (для того, чтобы отсчет времени начинался от 0). В качестве элементарного исхода возьмем упорядоченную пару $\omega = (x, y)$, где x —- момент прихода студента, y — момент прихода студентки. Тогда пространство элементарных исходов представляет собою квадрат $\Omega = \{\omega = (x, y) : \ x, y \in [0, 1]\}$. Событие, состоящее в том, что встреча состоится, представляет из себя те точки единичного квадрата Ω, для которых разность между моментами прихода студента и студентки по модулю не превосходит 1/3часа (=20 мин.): $A = \{\omega = (x, y) : \ |y - x| \leq \frac{1}{3}, \ x, y \in [0,\ 1]\}$. На Рис.1

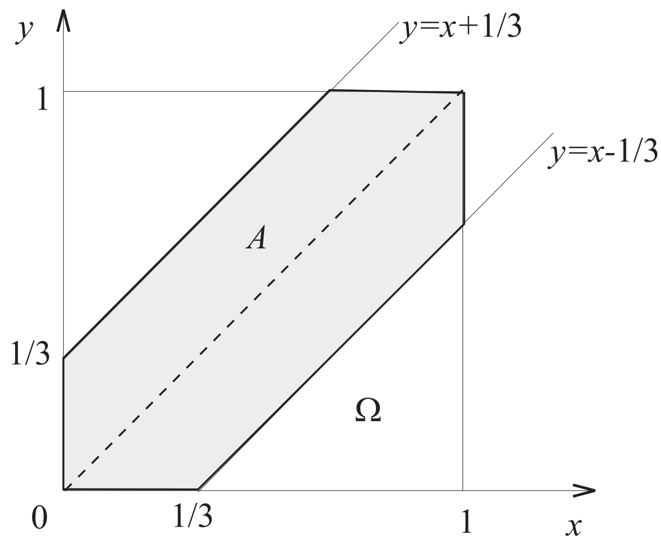


Рис.1

событие A изображено в виде заштрихованной части квадрата. По правилу геометрической вероятности искомая вероятность равна отношению площадей:

$$P(A) = \frac{S(A)}{S(\Omega)} = S(\Omega) - S(\overline{A}) = 1 - \left(\frac{2}{3}\right)^2 = \frac{5}{9} > \frac{1}{2}.$$

Заметим, что если бы студент и студентка договорились ждать друг друга в течение 15 минут, то вероятность встречи была бы равна

$$P(A) = 1 - \left(\frac{3}{4}\right)^2 = \frac{7}{16} < \frac{1}{2}.$$

Отметим, что выбор элементарных исходов при решении задач на геометрическую вероятность не является однозначным. Например, в примере 1 в качестве элементарного исхода можно было взять $\omega = (x, y)$, где x — момент прихода первого из лиц, y — момент прихода второго лица. Тогда $\Omega = \{\omega = (x, y) : \ 0 \le x \le y \le 1\}$, $A = \{\omega = (x, y) \in \Omega : \ y - x \le \frac{1}{3}\}$. Тем не менее следует отметить, что при выборе пространства элементарных исходов в задачах на геометрическую вероятность необходимо проявлять осторожность.

Мы убедились в том, что даже в простейшем случае несчетного (недискретного) пространства элементарных исходов в качестве событий надо брать не произвольные подмножества Ω, а только те, которые образуют некоторую σ-алгебру. В общем случае в теории вероятностей принято фиксировать определенную σ-алгебру подмножеств пространства элементарных исходов $\mathcal{F}$, элементы которой называют событиями. При этом для подмножеств из Ω, не принадлежащих $\mathcal{F}$, вероятность не определяется. Заметим, что случай дискретного пространства вписывается в рассматриваемую схему, если в качестве $\mathcal{F}$ взять совокупность всех подмножеств из Ω. Так как понятие σ-алгебры используется в основаниях теории вероятностей, напомним его определение и некоторые свойства.

Определение 4.2. Система $\mathcal{A}$ подмножеств множества Ω называется **алгеброй**, если

1. $\Omega \in \mathcal{A}, \ \emptyset \in \mathcal{A}$.

2. Если $A, B \in \mathcal{A}$, то $A \cup B, \ AB, \ A \setminus B \in \mathcal{A}$.

Таким образом, алгебра — система подмножеств Ω, содержащая Ω и замкнутая относительно конечного числа теоретико-множественных операций (мы не стремимся в определениях алгебры и σ-алгебры использовать минимальное число свойств, достаточных для их определения).

Определение 4.3. Система $\mathcal{F}$ подмножеств множества Ω называется **σ-алгеброй**, если

1. $\mathcal{F}$ — алгебра.

2. Если $A_n \in \mathcal{F}$ $(n = 1, 2, \dots)$, то $\bigcup\limits_{n=1}^{\infty} A_n \in \mathcal{F}$, $\bigcap\limits_{n=1}^{\infty} A_n \in \mathcal{F}$.

Поэтому σ-алгебра — система подмножеств Ω, содержащая Ω и замкнутая относительно не более чем счетного числа теоретико-множественных операций.

Пример 2. Система $\mathcal{F}_* = \{\emptyset, \Omega\}$, состоящая из двух подмножеств Ω, является σ-алгеброй (и тем более алгеброй). Эта самая "бедная" σ-алгебра называется **тривиальной**. Если $\mathcal{F}$ — любая σ-алгебра подмножеств Ω, то $\mathcal{F}_* \subset \mathcal{F}$.

Пример 3. Система $\mathcal{F}^* = \{A : A \subset \Omega\}$, состоящая из всех подмножеств Ω, является σ-алгеброй (и тем более алгеброй). Это самая "богатая" σ-алгебра. Если $\mathcal{F}$ — любая σ- алгебра подмножеств Ω, то $\mathcal{F} \subset \mathcal{F}^*$.

Пример 4. Система $\mathcal{F}_A = \{\emptyset, \Omega, A, \overline{A}\}$, состоящая из четырех подмножеств Ω, где $A \subset \Omega$, называется **σ-алгеброй, порожденной множеством** A. Если $\mathcal{F}^A$ — любая σ-алгебра подмножеств Ω, содержащая A, то $\mathcal{F}_A \subset \mathcal{F}^A$.

Пример 5. Пусть $D = \{D_1, \quad D_2, \quad D_3, \quad \dots\}$ — счетное разбиение Ω на непустые множества ($\Omega = D_1 + D_2 + D_3 + \dots$). Тогда σ-алгебра подмножеств Ω, состоящая из всех конечных или счетных объединений множеств системы D, называется **σ-алгеброй, порожденной разбиением** D и обозначается $\sigma(D)$.

Лемма 4.4. *Пусть X — некоторая система подмножеств множества Ω. Тогда существуют наименьшая алгебра $\alpha(X)$ и наименьшая σ-алгебра $\sigma(X)$, содержащие систему X.*

Доказательство. По крайней мере одна алгебра и σ-алгебра, содержащая X, существует (например,

$\mathcal{F}^*$). Пусть $\alpha(X)$ — пересечение всех алгебр, а $\sigma(X)$ — пересечение всех σ-алгебр, содержащих систему X. Очевидно, $\alpha(X)$ — алгебра (как пересечение алгебр), $\sigma(X)$ — σ-алгебра (как пересечение σ-алгебр). Они являются минимальными. Например, поскольку если $A \in \sigma(X)$, то A принадлежит любой σ-алгебре $\mathcal{F}$, содержащей X, т.е. $\sigma(X) \subset \mathcal{F}$, следовательно $\sigma(X)$ — минимальна. $\square$

Отметим, что $\sigma(D)$ в примере 5 как раз и является наименьшей σ-алгеброй, содержащей систему D.

Пусть Ω — пространство элементарных исходов, соответствующее некоторому случайному эксперименту, $\mathcal{F}$ — некоторая фиксированная σ-алгебра подмножеств множества Ω.

Определение 4.5. **Событием** называется любой элемент σ-алгебры $\mathcal{F}$.

В отличие от частного случая дискретного пространства элементарных исходов в общем случае событиями будут являться не все подмножества Ω, а только те из них, которые входят в σ-алгебру $\mathcal{F}$. И только для них будет определяться понятие вероятности. В случае дискретного пространства в качестве $\mathcal{F}$ обычно берется совокупность всех подмножеств $\mathcal{F}^*$ пространства Ω. Таким образом, случай дискретного пространства — частный случай рассматриваемой ниже общей схемы.

Определение 4.6. Пара $(\Omega,\ \mathcal{F})$, где Ω — некоторое множество, $\mathcal{F}$ — некоторая σ-алгебра подмножеств множества Ω, называется **измеримым пространством**.

Каждому событию A (т.е. каждому $A \in \mathcal{F}$) ставится в соответствие действительное число $P(A)$, называемое **вероятностью события** A, таким образом, что при этом выполнены следующие аксиомы:

I. $P(A) \geq 0$ для любого события A (аксиома **неотрицательности**);

II. $P(\Omega) = 1$ (аксиома **нормированности**);

III. $P(\sum\limits_{n=1}^{\infty} A_n) = \sum\limits_{i=1}^{\infty} P(A_n)$ (аксиома **счетной аддитивности**).

Как уже говорилось в параграфе 1, аксиомы вероятности представляют не что иное, как математическое отражение фундаментальных свойств относительной частоты за исключением того, что для возможности применения теории меры и интеграла Лебега свойство конечной аддитивности относительной частоты расширено до свойства счетной аддитивности вероятности. Поскольку к настоящему моменту не получено ни одного результата, основанного на приведенной аксиоматике и противоречащего экспериментальным данным, такое расширение вполне приемлемо.

Тройка $(\Omega, \mathcal{F}, P)$, где Ω — пространство элементарных исходов, $\mathcal{F}$ — некоторая σ-алгебра подмножеств множества Ω, P — вероятность (вероятностная мера) на $\mathcal{F}$, называется **вероятностным пространством** или **вероятностной моделью случайного эксперимента**. При этом вероятность P — это действительная функция, определенная на классе всех событий $\mathcal{F}$ и удовлетворяющая аксиомам I-III. В математике неотрицательная счетноаддитивная функция μ, определенная на некоторой σ-алгебре и такая, что $\mu(\emptyset) = 0$, называется мерой. Мера, для которой $\mu(\Omega) < \infty$, называется вполне конечной. Единственное, что выделяет вероятность или вероятностную меру из класса всех мер на $\mathcal{F}$, это то, что $P(\Omega) = 1$. В теории меры вероятностное пространство $(\Omega, \mathcal{F}, P)$ мы бы назвали **измеримым пространством с мерой** P.

Напомним один важный факт из теории меры, которым мы часто будем пользоваться. Мера μ на измеримом пространстве $(\Omega, \mathcal{F})$ называется σ-**конечной**, если существует последовательность $\{A_n\}$ измеримых (т.е. принадлежащих $\mathcal{F}$) множеств A_n таких, что $\Omega = \bigcup\limits_{n=1}^{\infty} A_n$ и $\mu(A_n) < \infty$.

Теорема о продолжении меры. *Любая σ-конечная мера μ может быть единственным способом продолжена с алгебры $\mathcal{A}$ на наименьшую σ-алгебру $\mathcal{F} = \sigma(\mathcal{A})$, содержащую алгебру $\mathcal{A}$.*

Напомним, что это означает. Первоначально мера μ определена на алгебре $\mathcal{A}$. Тогда существует мера μ_1, определенная на $\mathcal{F}$ и такая, что $\mu_1(A) = \mu(A)$ для всех $A \in \mathcal{A}$, которая и называется продолжением меры μ с алгебры $\mathcal{A}$ на σ-алгебру $\mathcal{F}$. Единственность продолжения означает, что если μ_1 и μ_2 — две меры на $\mathcal{F}$ такие, что $\mu_1(A) = \mu_2(A) = \mu(A)$ для всех $A \in \mathcal{A}$, то $\mu_1(A) = \mu_2(A)$ для всех $A \in \mathcal{F}$.

Замечание 1. Если предположить, что в случае дискретного пространства элементарных исходов вероятность определена не для всех его подмножеств, а лишь на некоторой σ-алгебре, то нетрудно убедиться, что вероятность может быть продолжена и притом бесконечным числом способов на σ-алгебру всех подмножеств пространства элементарных исходов. Например, если бросается игральная кость, то $\Omega = \{1, 2, 3, 4, 5, 6\}$. Предположим, что вероятность определена на σ-алгебре

$$\mathcal{F} = \{\phi, \Omega, A, \overline{A}\},$$

где $A = \{1, 3, 5\}$ — выпадение нечетного числа очков, $\overline{A} = \{2, 4, 6\}$ — выпадение четного числа очков. Пусть

$$P(\phi) = 0,\ P(\Omega) = 1,\ P(A) = 1/2,\ P(\overline{A}) = 1/2.$$

Если элементарным исходам 1,2,3 приписать произвольные вероятности $p_1 \geq 0,\ p_2 \geq 0\ p_3 \geq 0$ такие, что $p_1 + p_2 + p_3 = P(A) = 1/2$, а исходам 4,5,6 — вероятности $p_4 \geq 0,\ p_5 \geq 0,\ p_6 \geq 0$ такие, что $p_4 + p_5 + p_6 = P(\overline{A}) = 1/2$, то полагая

$$P(A) = \sum_{i \in A} p_i, \quad A \subset \Omega,$$

мы продолжим первоначально заданную на $\mathcal{F}$ вероятность на σ-алгебру всех подмножеств пространства Ω. Для обобщения этой идеи на случай произвольного дискретного пространства Ω решите следующую задачу.

Задача 1. Докажите, что если Ω дискретно, то любая σ-алгебра его подмножеств порождается не более чем счетным разбиением Ω на непересекающиеся части.

Если теперь $\mathcal{F}$ порождается разбиением $D_1,\ D_2,\ \ldots \quad (D_1 + D_2 + ... = \Omega)$, то для продолжения вероятности с σ-алгебры, порожденной этим разбиением, на σ-алгебру всех подмножеств Ω достаточно на каждом атоме D_k разбиения приписать каждому входящему в него элементарному исходу произвольным образом вероятность $p(\omega)$, лишь бы выполнялось $p(\omega) \geq 0$ и $\sum\limits_{\omega \in D_k} p(\omega) = P(D_k)$. Вероятность для любого $A \subset \Omega$ определяем согласно конструктивно-аксиоматическому определению вероятности.

Итак, не ограничивая общности, в случае дискретного пространства элементарных исходов можно считать измеримыми все подмножества Ω (т.е. в качестве σ-алгебры $\mathcal{F}$ брать совокупность всех подмножеств Ω). Это оправдывает наше определение события в дискретном пространстве элементарных исходов как произвольного подмножества Ω.

Свойства вероятности.

1. *Вероятность невозможного события равна нулю*: $P(\emptyset) = 0$.

Для того, чтобы неформальное представление о вероятности как об "идеальной" частоте глубже проникло в сознание и развивало вероятностную интуицию, полезно перед формальным доказательством каждого свойства провести его "частотное" обоснование. Так как невозможному событию не благоприятствует ни один возможный исход, то относительная частота невозможного события равна 0.

Доказательство. Поскольку

$$\Omega = \Omega + \emptyset + \emptyset + \emptyset + \dots ,$$

то используя аксиомы II и III, получим:

$$1 = P(\Omega) = P(\Omega) + P(\emptyset) + P(\emptyset) + P(\emptyset) + \dots ,$$

откуда

$$P(\emptyset) + P(\emptyset) + P(\emptyset) + \ldots = 0.$$

С учетом того, что $P(\emptyset) \geq 0$ по аксиоме I, из последнего равенства следует, что $P(\emptyset) = 0$. □

2. *Вероятность обладает свойством* **конечной аддитивности**:

$$P(\sum_{k=1}^{n} A_k) = \sum_{k=1}^{n} P(A_k).$$

Задача 2 (очень простая). Докажите, что это свойство эквивалентно конечной аддитивности для двух слагаемых:

$$P(A_1 + A_2) = P(A_1) + P(A_2).$$

Это свойство для относительной частоты проверялось ранее.

Доказательство. Для $k = n+1$, $n+2$, $\ldots$ введем $A_k = \emptyset$. На основании аксиомы III и свойства 1

$$P(\sum_{k=1}^{n} A_k) = P(\sum_{k=1}^{\infty} A_k) = \sum_{k=1}^{\infty} P(A_k) = \sum_{k=1}^{n} P(A_k).$$

□

3. *Если* $A \subset B$, *то* $P(B - A) = P(B) - P(A)$.

Так как $A \subset B$, то всякий раз, когда A происходит, происходит B. Поэтому из N проводимых экспериментов число тех из них, в которых происходит $B - A$, т.е. происходит B, но не происходит A, равно разности между числом экспериментов, в которых происходит B, и числом экспериментов, в которых происходит A, т.е. $N(B) - N(A)$. По формуле (В1)

$$P_N(B-A) = \frac{N(B-A)}{N} = \frac{N(B)-N(A)}{N} =$$

$$= \frac{N(B)}{N} - \frac{N(A)}{N} = P_N(B) - P_N(A),$$

т.е. относительная частота удовлетворяет этому свойству.

Доказательство. Если $A \subset B$, то $B = A + (B - A)$. По свойству 2 конечной аддитивности $P(B) = P(A) + P(B - A)$, откуда следует доказываемое свойство. $\square$

4. *Если $A \subset B$, то $P(A) \leq P(B)$.*

Это свойство для относительной частоты проверялось ранее.

Доказательство. В силу предыдущего свойства 3 $P(B-A) = P(B) - P(A)$, но по аксиоме I неотрицательности $P(B - A) \geq 0$, т.е. $P(B) - P(A) \geq 0$, откуда $P(A) \leq P(B)$. $\square$

Напомним, что симметрической разностью множеств A_1 и A_2 называется множество

$$A_1 \Delta A_2 = (A_1 \setminus A_2) + (A_2 \setminus A_1).$$

5. *Если $P(A_1 \Delta A_2) = 0$ то $P(A_1) = P(A_2)$. Обратное, вообще говоря, неверно; но если $A_1 = A_2$, то $P(A_1 \Delta A_2) = 0$.*

Доказательство. По свойству 2 конечной аддитивности

$$P(A_1 \Delta A_2) = P(A_1 \setminus A_2) + P(A_2 \setminus A_1) = 0.$$

По аксиоме I оба слагаемых неотрицательны; поэтому каждое из них равно 0. Но $A_1 \setminus A_2 = A_1 - (A_1A_2)$ и по свойству 3 $P(A_1 \setminus A_2) = P(A_1) - P(A_1A_2) = 0$, откуда $P(A_1) = P(A_1A_2)$. Аналогично, $P(A_2) = P(A_1A_2)$. Отсюда $P(A_1) = P(A_2)$.

Задача 3. Приведите пример, показывающий, что из $P(A_1) = P(A_2)$ не следует $P(A_1 \Delta A_2) = 0$.

6. *Для любого события A выполняется неравенство*

$$0 \le P(A) \le 1.$$

Так как $0 \le N(A) \le N$, то для относительной частоты это свойство выполняется.

Доказательство. По аксиоме I неотрицательности $P(A) \ge 0$. Так как любое событие $A \subset \Omega$, то по свойству 4 имеем $P(A) \le P(\Omega)$, но в силу аксиомы II $P(\Omega) = 1$. $\square$

7. *Имеет место* **обобщенная теорема сложения**:

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

Для того, чтобы посчитать число экспериментов, в которых произошло событие $A \cup B$, т.е. хотя бы одно из событий A или B, надо сложить число экспериментов, в которых произошло A, с числом экспериментов, в которых произошло B. При этом эксперименты, в которых они происходили одновременно, учитываются в обоих слагаемых. Поэтому из полученной суммы надо вычесть число таких экспериментов. Таким образом,

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

Следовательно, для относительных частот свойство 7 выполняется.

Доказательство. Очевидно, $A \cup B = A + (B - AB)$. В силу конечной аддитивности вероятности (свойство 2)

$$P(A \cup B) = P(A + (B - AB)) = P(A) + P(B - AB). \quad (4.2)$$

Так как $AB \subset B$, то по свойству 3

$$P(B - AB) = P(B) - P(AB). \quad (4.3)$$

Доказательство завершится, если подставить (4.3) в (4.2). □

Последнее свойство легко обобщается на произвольное конечное число слагаемых.

7′. $P(A_1 \cup A_2 \cup \ldots \cup A_n) = P(A_1) + P(A_2) + \ldots + P(A_n) -$

$$-P(A_1A_2) - P(A_1A_3) - \ldots - P(A_{n-1}A_n) + \ldots +$$
$$+(-1)^{n+1} P(A_1A_2 \ldots A_n).$$

Например, для трех слагаемых свойство имеет вид

$$P(A_1 \cup A_2 \cup A_3) = P(A_1) + P(A_2) + P(A_3) -$$
$$-P(A_1A_2) - P(A_1A_3) - P(A_2A_3) + P(A_1A_2A_3).$$

Задача 4. Докажите свойство 7′.

8. *Вероятность противоположного события равна*

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

Так как во всяком эксперименте происходит одно и только одно из событий A, $\overline{A}$, то число экспериментов, в которых происходит $\overline{A}$, получается вычитанием из числа всех экспериментов числа экспериментов, в которых происходит A:

$$N(\overline{A}) = N - N(A).$$

Поэтому свойство 8 выполняется для относительной частоты.

Доказательство. По определению противоположного события $\overline{A} = \Omega - A$, причем $A \subset \Omega$; по свойству 3 $P(\overline{A}) = P(\Omega) - P(A)$. Остается только учесть, что $P(\Omega) = 1$ по аксиоме II. $\square$

9. *Если события H_1, H_2, ... , H_n образуют полную группу, то*

$$P(H_1) + P(H_2) + \ldots + P(H_n) = 1.$$

Частотная интерпретация этого свойства такая же, как для свойства 8, но только не для двух, а для n событий.

Доказательство. По определению полной группы $H_1 + H_2 + \ldots + H_n = \Omega$; по свойству 2 конечной аддитивности

$$P(H_1) + P(H_2) + \ldots + P(H_n) = P(\Omega) = 1. \qquad \square$$

Отметим, что свойство 8 является частным случаем свойства 9 при $n = 2$, поскольку A и $\overline{A}$ являются противоположными событиями тогда и только тогда, когда они образуют полную группу из двух событий. Заметим также, что свойство 9 имеет место и для случая полной группы из счетного числа событий (в доказательстве вместо конечной аддитивности надо применить счетную аддитивность).

10. *Вероятность обладает свойством* **полуаддитивности**:

$$P(\bigcup_{n=1}^{\infty} A_n) \le \sum_{n=1}^{\infty} P(A_n).$$

Замечание 2. Это свойство выполняется и для конечного числа событий.

Очевидно, число экспериментов, в которых происходит хотя бы одно из событий, не превосходит суммы числа экспериментов, в которых происходит каждое событие. Поэтому для конечного числа событий это свойство для относительных частот выполняется.

Доказательство. Так как

$$\bigcup_{n=1}^{\infty} A_n = A_1 + \overline{A_1}\, A_2 + \overline{A_1}\,\overline{A_2}\, A_3 + \dots ,$$

то в силу счетной аддитивности и свойства 4

$$P(\bigcup_{n=1}^{\infty} A_n) = P(A_1) + P(\overline{A_1}\, A_2) + P(\overline{A_1}\,\overline{A_2}\, A_3) + \dots \le$$

$$\le P(A_1) + P(A_2) + P(A_3) + \dots . \qquad \square$$

Прежде, чем формулировать следующее свойство, введем важное для теории вероятностей понятие сходимости последовательности множеств.

Определение 4.7. Говорят, что последовательность множеств $\{A_n\}$ **сходится к множеству** A **снизу** и пишут $A_n \uparrow A$, если $A_1 \subset A_2 \subset A_3 \subset \ldots$ и $A = \bigcup\limits_{n=1}^{\infty} A_n$. Говорят, что последовательность множеств $\{A_n\}$ **сходится к множеству** A **сверху** и пишут $A_n \downarrow A$, если $A_1 \supset A_2 \supset A_3 \supset \ldots$ и $A = \bigcap\limits_{n=1}^{\infty} A_n$.

11. **Свойство непрерывности вероятности**. *Вероятность непрерывна снизу, т.е. если $A_n \uparrow A$, то $P(A_n) \to P(A)$. Вероятность непрерывна сверху, т.е. если $A_n \downarrow A$, то $P(A_n) \to P(A)$.*

Это свойство не может быть интерпретировано в терминах относительной частоты, так как связано с бесконечными последовательностями множеств.

Доказательство. Поскольку $A_n \in \mathcal{F}$, то для обоих случаев $A \in \mathcal{F}$. Пусть $A_n \uparrow A$. Так как $A_n \subset A_{n+1}$, то $A = \bigcup\limits_{n=1}^{\infty} A_n = A_1 + (A_2 - A_1) + (A_3 - A_2) + \ldots$. По аксиоме III счетной аддитивности и свойству 3

$$P(A) = P(A_1) + P(A_2 - A_1) + P(A_3 - A_2) + \ldots = P(A_1) +$$

$$+ [P(A_2) - P(A_1)] + + [(P(A_3) - P(A_2)] + \ldots = \lim_{n\to\infty} P(A_n)$$

(ведь частная сумма последнего ряда равна $P(A_n)$). Очевидно, сходимость $P(A_n)$ к $P(A)$ — монотонная.

Пусть теперь $A_n \downarrow A$. Нетрудно понять из определения 4.7, что тогда $\overline{A}_n \uparrow \overline{A}$. По доказанному $P(\overline{A}_n) \to P(\overline{A})$. В силу свойства 8 это означает, что $1 - P(A_n) \to 1 - P(A)$, откуда немедленно следует, что $P(A_n) \to P(A)$. $\square$

Задача 5. Докажите, что если P — конечно-аддитивная функция множеств, определенная на σ-алгебре $\mathcal{F}$, то следующие 4 условия эквивалентны:

1) P счетно-аддитивна;
2) P непрерывна снизу, т.е.

$$\text{если}\ \ A_n \in \mathcal{F},\ \ A_n \uparrow A,\ \ \text{то}\ \ \lim_{n\to\infty} P(A_n) = P(A);$$

3) P непрерывна сверху, т.е.

$$\text{если}\ \ A_n \in \mathcal{F},\ \ A_n \downarrow A,\ \ \text{то}\ \ \lim_{n\to\infty} P(A_n) = P(A);$$

4) P непрерывна в "нуле" (пустом множестве), т.е.

$$\text{если}\ \ A_n \in \mathcal{F},\ \ A_n \downarrow \emptyset,\ \ \text{то}\ \ \lim_{n\to\infty} P(A_n) = 0.$$

Подводя итог, можно сказать, что *вероятность или вероятностная мера — мера, определенная на σ-алгебре $\mathcal{F}$ измеримого пространства* $(\Omega,\ \mathcal{F})$, *удовлетворяющая условию* $P(\Omega) = 1$ *и, следовательно, обладающая всеми свойствами вполне конечной меры.*

§ 5. Условная вероятность. Независимость событий. Формула полной вероятности и формула Байеса

Часто бывает необходимо вычислить вероятность события B в предположении, что наступило некоторое событие A, имеющее отличную от нуля вероятность. Для этого естественно рассматривать только те эксперименты, в которых происходит A. Предположим, что в N экспериментах A произошло $N(A)$ раз, AB произошло $N(AB)$ раз. Если рассматриваются только те эксперименты, в которых происходит A, то среди них событие B происходит $N(AB)$ раз. Поэтому относительная частота B при условии, что A наступило, равна

$$P_N(B/A) = \frac{N(AB)}{N(A)} = \frac{\frac{N(AB)}{N}}{\frac{N(A)}{N}}. \qquad (5.1)$$

Поскольку при больших N относительная частота $\frac{N(AB)}{N}$ стабилизируется около вероятности $P(AB)$, а относительная частота $\frac{N(A)}{N}$ — около вероятности $P(A)$, то, как видно из (5.1), условная относительная частота $P_N(B/A)$ будет стабилизироваться около числа $P(B/A) = \frac{P(AB)}{P(A)}$. Это показывает естественность следующего определения условной вероятности.

Определение 5.1. Пусть $(\Omega, \mathcal{F}, P)$ — вероятностное пространство, A и B — события (т.е. $A, B \in \mathcal{F}$), причем $P(A) \neq 0$. **Условной вероятностью события B относительно A или при условии A** называется число

$$P(B/A) = \frac{P(AB)}{P(A)}. \qquad (5.2)$$

Теорема 5.2. *При фиксированном A таком, что $P(A) \neq 0$, условная вероятность $P(B/A)$ как функция множества $B \in \mathcal{F}$ является вероятностью (вероятностной мерой) на $\mathcal{F}$.*

Доказательство. По формуле (5.2) каждому $B \in \mathcal{F}$ ставится в соответствие действительное число $P(B/A)$. Остается проверить выполнение аксиом I-III. Так как $P(AB) \geq 0$, $P(B) > 0$, то для любого $B \in \mathcal{F}$ будет

$P(B/A) \geq 0$, т.е. справедлива аксиома I. Далее, поскольку $A\Omega = A$, то

$$P(\Omega/A) = \frac{P(A\Omega)}{P(A)} = \frac{P(A)}{P(A)} = 1.$$

Наконец, если $B_1, B_2, B_3, \dots$ попарно не пересекаются, то тем более попарно не пересекаются $AB_1, AB_2, AB_3, \dots$ как их части. Поэтому $A(B_1 + B_2 + B_3 + \dots) = AB_1 + AB_2 + AB_3 + \dots$, следовательно,

$$P\Big(\sum_{n=1}^{\infty} B_n/A\Big) = \frac{P\Big(A\Big(\sum\limits_{n=1}^{\infty} B_n\Big)\Big)}{P(A)} = \frac{P\Big(\sum\limits_{n=1}^{\infty} AB_n\Big)}{P(A)} =$$

$$= \frac{\sum\limits_{n=1}^{\infty} P(AB_n)}{P(A)} = \sum_{n=1}^{\infty} \frac{P(AB_n)}{P(A)} = \sum_{n=1}^{\infty} P(B_n/A),$$

т.е. выполнена аксиома III. $\square$

Таким образом, условная вероятность при фиксированном A обладает всеми свойствами (обычной) вероятности. Например, $P(\overline{B}/A) = 1 - P(B/A)$; если $B_n \uparrow B$, то $P(B_n/A) \to P(B/A)$. Итак, если рассматривать только те эксперименты, в которых происходит событие A, то их математической моделью является вероятностное пространство $(\Omega, \mathcal{F}, P(\cdot/A))$. Иногда бывает удобно рассматривать сужение меры $P(\cdot/A)$ на множество A. В этом случае в качестве σ-алгебры берется $A \cap \mathcal{F} = \{A \cap C : C \in \mathcal{F}\}$, и тогда вероятностное пространство — $(A, A \cap \mathcal{F}, P(\cdot/A))$.

Полезно рассмотреть следующую интерпретацию условной вероятности в частном случае конечного пространства равновероятных элементарных исходов $\Omega =$

$\{\omega_1, \omega_2, \dots, \omega_n\}$. Пусть событию A благоприятствует n_A элементарных исходов, событию AB — n_{AB} элементарных исходов. По определению условной вероятности

$$P(B/A) = \frac{P(AB)}{P(A)} = \frac{n_{AB}/n}{n_A/n} = \frac{n_{AB}}{n_A}. \qquad (5.3)$$

Если рассматривать только те эксперименты, в которых A происходит, то n_A — число всех возможных исходов, n_{AB} — число исходов, благоприятствующих B. Таким образом, вычисление условной вероятности по формуле (5.3) происходит по тому же самому классическому определению, что и вычисление (безусловной) вероятности: это отношение числа благоприятствующих исходов к числу всех возможных исходов. Только при этом исключаются эксперименты, в которых A не происходит.

Из определения 5.1 следует, что

$$P(AB) = P(A) \cdot P(B/A). \qquad (5.4)$$

Формулу (5.4) называют **формулой** или даже **теоремой умножения**. Такое "громкое" название элементарного следствия определения условной вероятности исторически оправдано, так как первоначально условная вероятность вводилась и долго использовалась только в условиях классического определения вероятности с помощью равенства (5.3). При этом (5.4) требовало доказательства и потому называлось теоремой.

Теорема умножения для N событий. *Если $A_1, A_2, \dots, A_n$ — события и $P(A_1A_2\dots A_{n-1}) \neq 0$, то*

$$P(A_1A_2\dots A_n) = P(A_1)P(A_2/A_1)P(A_3/A_1A_2)\dots$$
$$\dots P(A_n/A_1A_2\dots A_{n-1}).$$

Доказательство получается с помощью последовательного применения формулы умножения (5.4):

$$P(A_1 \dots A_{n-1}A_n) = P(A_1 \dots A_{n-1})P(A_n/A_1 \dots A_{n-1}) =$$
$$= P(A_1 \dots A_{n-2})P(A_{n-1}/A_1 \dots A_{n-2})P(A_n/A_1 \dots A_{n-1}) =$$
$$= P(A_1A_2)P(A_3/A_1A_2)\dots P(A_{n-1}/A_1A_2 \dots A_{n-2})\cdot$$
$$\cdot P(A_n/A_1A_2 \dots A_{n-1}) = P(A_1)P(A_2/A_1)P(A_3/A_1A_2)\dots$$
$$\dots P(A_n/A_1A_2 \dots A_{n-1}). \qquad \square$$

Пример 1. Студент знает 20 из 25 вопросов. Какова вероятность того, что он ответит на три наудачу поставленные ему преподавателем вопроса?

Введем события: $A = \{$студент ответит на три вопроса$\}$, $A_i = \{$студент ответит на i-й по счету вопрос$\}$, $i = 1, 2, 3$. Очевидно, $A = A_1A_2A_3$. Если студент ответил на 1-й вопрос, то всего остается 24 "вакантных" вопроса, из них студент знает 19, поэтому $P(A_2/A_1) = 19/24$. Если студент ответил на первые два вопроса, то остается 23 "вакантных" вопроса, из них студент знает 18, следовательно, $P(A_3/A_1A_2) = 18/23$. По теореме умножения для 3-х событий

$$P(A) = P(A_1A_2A_3) = P(A_1)P(A_2/A_1)P(A_3/A_1A_2) =$$
$$= \frac{20}{25} \cdot \frac{19}{24} \cdot \frac{18}{23} = \frac{57}{115} < \frac{1}{2}.$$

Заметим, что решение этой задачи с помощью классического определения приводит к тому же результату (что неудивительно) в форме

$$P(A) = \frac{C_{20}^3}{C_{25}^3} = \frac{57}{115}.$$

Рассмотрим пример 10 из третьего параграфа, к которому применялось ранее классическое определение:

Пример 2. В чулане находится n пар ботинок. Из них наудачу выбираются $2r$ ботинок ($2r < n$). Какова вероятность того, что среди выбранных ботинок отсутствуют парные?

Введем события: $A = \{$среди выбранных ботинок отсутствуют парные$\}$, $A_i = \{$среди первых i выбранных ботинок отсутствуют парные$\}$, $i = 2, 3, \ldots, 2r$. Очевидно, $A = A_2 A_3 \ldots A_{2r}$. По теореме умножения

$$P(A) = P(A_2 A_3 \ldots A_{2r}) = P(A_2)P(A_3/A_2)P(A_4/A_2A_3)\ldots$$

$$\ldots P(A_{2r-1}/A_2A_3\ldots A_{2r-2})P(A_{2r}/A_2A_3\ldots A_{2r-1}) =$$

$$= \frac{2n-2}{2n-1}\cdot\frac{2n-4}{2n-2}\cdot\frac{2n-6}{2n-3}\cdot\ldots\cdot\frac{2n-4r+4}{2n-2r+2}\cdot\frac{2n-4r+2}{2n-2r+1} =$$

$$= \frac{2^{2r-1}(n-1)(n-2)\ldots(n-2r+2)}{(2n-1)(2n-2)\ldots(2n-2r+1)} = \frac{2^{2r}n!(2n-2r)!}{(2n)!(n-2r)!} =$$

$$= \frac{2^{2r}C_n^{2r}}{C_{2n}^{2r}}.$$

Определение 5.3. События A и B называются **независимыми**, если $P(AB) = P(A)P(B)$.

Пример 3. Наудачу подбрасываются две правильные монеты. Пусть $A = \{$выпадение герба на первой монете$\}$, $B = \{$выпадение герба на второй монете$\}$.

В данном случае $\Omega = \{(г,г),(г,р),(р,г), (р,р)\}$, $A = \{(г,г),(г,р)\}$, $B = \{(г,г), (р,г)\}$, $AB = \{(г,г)\}$. Поэтому

$$P(AB) = \frac{1}{4} = \frac{1}{2} \cdot \frac{1}{2} = \frac{2}{4} \cdot \frac{2}{4} = P(A)P(B),$$

т.е. A и B независимы.

Пример 4. В прямоугольник

$$\Omega = \{(x, y) \in \mathbb{R}^2 : \ \ 0 \leq x \leq 6, \ 0 \leq y \leq 3\}$$

"наудачу" бросается точка (слово "наудачу" означает, что имеет место определение геометрической вероятности). Пусть

$$A = \{\text{точка попадет в прямоугольник } A = \{(x, y) \in \mathbb{R}^2 :$$

$$0 \leq x \leq 6, \ 0 \leq y \leq 2\}\},$$

$$B = \{\text{точка попадет в прямоугольник } B = \{(x, y) \in \mathbb{R}^2 :$$

$$2 \leq x \leq 4, \ 0 \leq y \leq 3\}\}.$$

Тогда $AB = \{(x, y) : \ 2 \leq x \leq 4, \ 0 \leq y \leq 2\}$. Если $S(\Pi)$ — площадь прямоугольника Π, то

$$P(AB) = \frac{S(AB)}{S(\Omega)} = \frac{2 \cdot 2}{3 \cdot 6} = \frac{2}{9}, \quad P(A) = \frac{S(A)}{S(\Omega)} = \frac{2 \cdot 6}{3 \cdot 6} = \frac{2}{3},$$

$$P(B) = \frac{S(B)}{S(\Omega)} = \frac{2 \cdot 3}{3 \cdot 6} = \frac{1}{3},$$

т.е. $P(AB) = P(A)P(B)$. Значит A и B независимы (см. Рис.2).

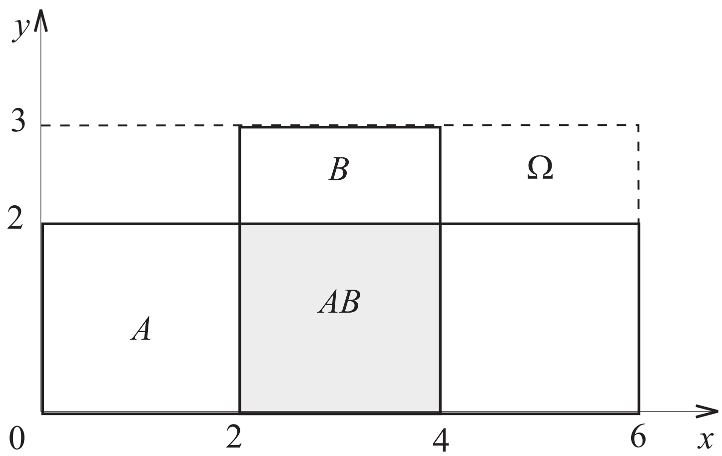


Рис.2

Свойства.

1. Любое событие и событие, вероятность которого равна нулю, независимы.

Доказательство. Пусть A — любое событие, $P(B) = 0$. Так как $AB \subset B$, то $P(AB) \leq P(B) = 0$, откуда $P(AB) = 0$. Поэтому $P(AB) = 0 = P(A) \cdot 0 = P(A)P(B)$. □

В частности, невозможное событие не зависит от любого события и даже от самого себя.

2. Любое событие и событие, вероятность которого равна единице, независимы.

Доказательство. Пусть A — любое событие, $P(B) = 1$. Так как $P(A) = P(AB) + P(A\overline{B})$, $P(A\overline{B}) \leq P(\overline{B}) = 1 - P(B) = 0$, то $P(AB) = P(A) = P(A) \cdot 1 = P(A)P(B)$, т.е. A и B независимы. □

В частности, достоверное событие не зависит от любого события и даже от самого себя.

3. Событие не зависит от самого себя тогда и только тогда, когда его вероятность равна нулю или единице (т.е. оно отличается от $\emptyset$ или от Ω на множество вероятности ноль).

Доказательство. Если событие имеет вероятность 0 или 1, то оно не зависит от самого себя по свойствам 1 и 2. Обратно, пусть событие A не зависит от самого себя, т.е. $P(AA) = P(A)P(A)$. Тогда $P(A) = [P(A)]^2$, откуда $P(A)$ равно 0 или 1. $\square$

4. Пусть $A, B \in \mathcal{F},\ P(A) \neq 0$. События A и B независимы тогда и только тогда, когда $P(B/A) = P(B)$.

Доказательство. Если A и B независимы, то, используя определения условной вероятности и независимости, получим

$$P(B/A) = \frac{P(AB)}{P(A)} = \frac{P(A)P(B)}{P(A)} = P(B).$$

Обратно, пусть $P(B/A) = P(B)$. По теореме умножения (5.4)

$$P(AB) = P(A)P(B/A) = P(A)P(B),$$

т.е. A и B независимы. $\square$

Замечание 1. Данное свойство долгое время служило определением независимости. Если $P(A) \neq 0$, то B не зависит от A, когда вероятность события B не зависит от того, произошло событие A или нет. Иными словами, на вероятность события B дополнительная информация о том, что произошло событие A, не оказывает никакого влияния. Однако при определении независимости событий с помощью свойства 4 теряется вырожденный случай $P(A) = 0$. Поэтому в настоящее время принято более симметричное по отношению к A и B определение независимости 5.3. В связи с вышеизложенным весьма правдоподобным является следующий принцип, который очень часто используется на практике: **если события A и B не связаны причинно, то они независимы**. Например, если $A = \{$ в течение ближайших суток произойдет извержение вулкана на Камчатке$\}$, $B = \{$в течение ближайших суток в Нью-Йорке осуществится террористический акт$\}$, то A и B независимы, так как не связаны причинно. Приведенный принцип является практическим правилом, а не математическим утверждением хотя бы потому, что по крайней мере на данный момент невозможно дать строгое математическое определение понятия причинной связанности событий. Поскольку применение этого принципа пока не приводило к противоречиям с опытом, его можно считать разумным. Отметим, что на самом деле определение 5.3 независимости событий даже шире: можно привести примеры причинно связанных событий, независимых в смысле определения 5.3.

5. Если события A и B независимы, то A и $\overline{B}$ независимы.

С точки зрения сформулированного принципа это утверждение весьма правдоподобно.

Доказательство. Так как $AB \subset A$, $A\overline{B} = A - AB$, то по свойству 3 из четвертого параграфа $P(A\overline{B}) = P(A) - P(AB) = P(A) - P(A)P(B) =$

$P(A)(1 - P(B)) = P(A)P(\overline{B})$, т.е. A и $\overline{B}$ независимы. $\square$

Как показал многолетний опыт преподавания, студенты часто путают понятия независимости и несовместимости событий. В связи с этим предлагается следующая задача.

Задача 1. Установите необходимые и достаточные условия для того, чтобы несовместимые события A и B были независимыми. Ответ дайте в форме: по крайней мере одно из событий A или B обладает некоторым свойством.

Определение 5.4. События $A_1, A_2, \ldots, A_n$ называются **независимыми (в совокупности)**, если

$$P(A_iA_j) = P(A_i)P(A_j) \quad \forall\, i,\, j = \overline{1,n};\ i \neq j; \qquad (5.5)$$

$$P(A_iA_jA_k) = P(A_i)P(A_j)P(A_k) \quad \forall\, i,\, j,\, k = \overline{1,n};$$

$$i \neq j;\ i \neq k;\ j \neq k;$$

$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

$$P(A_1A_2\ldots A_n) = P(A_1)P(A_2)\ldots P(A_n).$$

Естественно назвать события, удовлетворяющие (5.5), **попарно независимыми**. На первый взгляд может показаться, что для независимости в совокупности достаточно попарной независимости. На самом деле это не так, что показывает классический пример

Бернштейна.

Пример 4. Предположим, что грани тетраэдра окрашены следующим образом: первая — в красный цвет, вторая — в синий цвет , третья — в зеленый цвет, четвертая — во все три цвета. Тетраэдр подбрасывается наудачу и выпавшей считается грань, попадающая в основание. Введем события: $A = \{$ на выпавшей грани имеется красный цвет$\}$, $B = \{$на выпавшей грани имеется синий цвет$\}$, $C = \{$на выпавшей грани имеется зеленый цвет$\}$. Очевидно,

$$P(AB) = P(AC) = P(BC) = \frac{1}{4} = \frac{2}{4} \cdot \frac{2}{4} =$$

$$= P(A)P(B) = P(A)P(C) = P(B)P(C),$$

т.е. события A, B и C попарно независимы. Однако

$$P(ABC) = \frac{1}{4} \neq \frac{1}{8} = P(A)P(B)P(C),$$

т.е. A, B и C не являются независимыми в совокупности.

Теорема 5.5. *Пусть событие A может происходить только с одним из событий H_1, H_2, … , H_n. Тогда имеет место* **формула полной вероятности***:*

$$P(A) = \sum_{i=1}^{n} P(H_i)P(A/H_i).$$

Доказательство. По условию A происходит тогда и только тогда, когда происходит хотя бы одно из событий AH_1, AH_2, … , AH_n, т.е. $A = AH_1 \cup AH_2 \cup \ldots \cup AH_n$. Так как A может происходить только с одним из событий

$H_1, H_2, \ldots, H_n$, то $AH_1, AH_2, \ldots, AH_n$ попарно несовместимы, следовательно, $A = AH_1 + AH_2 + \ldots + AH_n$. В силу конечной аддитивности вероятности $P(A) = P(AH_1) + P(AH_2) + \ldots + P(AH_n)$. Используя теорему умножения, получаем: $P(A) = P(H_1)P(A/H_1) + P(H_2)P(A/H_2) + \ldots + P(H_n)P(A/H_n)$. □

Замечание 2. Формально в условии теоремы надо добавить, что $P(H_i) \neq 0, \; i = \overline{1, n}$, для того, чтобы были определены условные вероятности $P(A/H_i)$. Однако в этом случае формула полной вероятности применима, если считать, что произведение $P(H_i) = 0$ на неопределенную величину $P(A/H_i)$ равно нулю.

Замечание 3. Условия теоремы 5.5 заведомо выполняются, если A может происходить хотя бы с одним из событий $H_1, H_2, \ldots, H_n$, являющихся попарно несовместимыми. Чаще всего формула полной вероятности применяется, когда $H_1, H_2, \ldots, H_n$ образуют полную группу, т.е. $\sum\limits_{i=1}^{n} H_i = \Omega$. Тогда тем более выполнено $A = \sum\limits_{i=1}^{n} AH_i$, т.е. условия теоремы 5.5. Поскольку в этом случае в результате эксперимента должно произойти одно и только одно из событий $H_1, H_2, \ldots, H_n$, то эти события часто называют **гипотезами**. Таким образом, для определения вероятности события A надо сложить по всем возможным гипотезам произведения вероятностей гипотез на условные вероятности A при этих гипотезах.

Замечание 4. Формула полной вероятности остается справедливой и в случае счетного числа событий $H_1, H_2, \ldots$. Просто при доказательстве надо вместо конечной аддитивности использовать счетную аддитивность вероятности.

Формула полной вероятности приведена в форме теоремы умышленно с тем, чтобы подчеркнуть ее исключительную важность для всей теории вероятностей (ведь доказательство тривиально). Эта формула является основным инструментом ("рабочей лошадкой") в большинстве теоретических исследований и разнообразных приложениях теории вероятностей к практическим задачам.

Пример 5. Из первой урны, содержащей 2 белых и 3 черных шара, наудачу выбирается шар и перекладывается во вторую урну, содержащую 2 белых и 2 черных шара. Шары во второй урне тщательно перемешиваются, после чего из нее наудачу выбирается шар. Какова вероятность, что он белый?

Пусть $A = \{$из второй урны выбран белый шар$\}$. Сразу вероятность A трудно определить в силу того, что нам неизвестно, какой шар переложен в нее из первой урны. Но если цвет переложенного шара известен, то эта неопределенность снимается и вероятность A становится известной. Поэтому введем две гипотезы: $H_1 = \{$из первой урны во вторую переложен белый шар$\}$, $H_2 = \overline{H_1} = \{$из первой урны во вторую переложен черый шар$\}$. Очевидно, $P(H_1) = \frac{2}{5}$, $P(H_2) = \frac{3}{5}$, $P(A/H_1) = \frac{3}{5}$, $P(A/H_2) = \frac{2}{5}$; по формуле полной вероятности

$$P(A) = P(H_1)P(A/H_1)+P(H_2)P(A/H_2) = \frac{2}{5}\cdot\frac{3}{5}+\frac{3}{5}\cdot\frac{2}{5} = \frac{12}{25}.$$

Пример 6. Случайное блуждание и задача о разорении игрока.

Предположим, что в дискретные моменты времени $1, 2, \ldots$ частица может блуждать по целочисленным точкам прямой. Если в некоторый целочисленный момент она находилась в точке i, то в следующий целочисленный момент она совершает скачок в точку $i+1$ с вероятностью p, в точку $i-1$ — с вероятностью q или остается на месте с вероятностью r $(p+q+r=1,\ p,q,r \geq 0)$. Движение может происходить на всей прямой, на полупрямой или на отрезке. В последнем случае говорят об экранах, поставленных на пути движения частицы. Если частица, попав в такую точку, навсегда остается в ней, то такой экран называется **поглощающим**. Если движение частицы происходит справа от экрана j и в некоторый момент она попадает в него, а в следующий момент перескакивает в точку $j+1$, то экран j называется **отражающим**. При движении частицы слева от отражающего экрана j и попадании в него в некоторый дискретный момент, в следующий момент частица перескакивает в точку $j-1$. Возможен и промежуточный вариант так называемого **упругого жесткого экрана**, попав в который в некоторый момент частица с положительной вероятностью s остается в нем навсегда, а с положительной вероятностью $1-s$ в следующий момент времени перескакивает в соседнее с экраном состояние, из которого она попала в этот экран.

Для того, чтобы не рассматривать специальные тривиальные случаи, в дальнейшем будем предполагать, что $pq > 0$.

Рассмотрим блуждание частицы с двумя поглощающими экранами 0 и a $(a > 0$ — целое число). Предположим, что частица находится в состоянии i $(i = \overline{1, a-1})$.

Обозначим через A событие, состоящее в том, что частица когда-нибудь поглотится в состоянии 0 (возможно еще, что частица поглотится в состоянии a; интуитивно ясно, что третья возможность, что частица никогда не поглотится и все время будет блуждать по состояниям $1, 2, \ldots, a-1$ равна нулю). Обозначим через p_i условную вероятность A при условии, что частица находится в состоянии i. Введем следующие три гипотезы: $H_1 = \{$после первого скачка частица перейдет в состояние $i+1\}$; $H_2 = \{$после первого скачка частица перейдет в состояние $i-1\}$; $H_3 = \{$после первого скачка частица останется в состоянии $i\}$. Очевидно, что они образуют полную группу. Пусть $B_i = \{$ в начальный момент частица находится в состоянии $i\}$. Как мы знаем, при фиксированном B_i условная вероятность $P(A/B_i)$ является вероятностной мерой. По формуле полной вероятности для этой меры (а не для первоначальной вероятности $P(A)$)

$$P(A/B_i) = P(H_1/B_i)P(A/H_1B_i) + P(H_2/B_i)P(A/H_2B_i) + \\ + P(H_3/B_i)P(A/H_3B_i),$$

что в наших обозначениях запишется как

$$p_i = pp_{i+1} + qp_{i-1} + rp_i, \qquad i = 1, 2, \ldots, a-1. \tag{5.6}$$

Здесь неявно использовано марковское свойство процесса случайного блуждания, присущее так называемым **марковским процессам**. Заключается оно в данном случае в том, что $P(A/H_1B_i)$, например, заменяется на p_{i+1}. Событие H_1B_i заключается в одновременном осуществлении двух событий: начальное состояние частицы есть i и после первого шага она перескочит из точки i в точку $i+1$. Замена $P(A/H_1B_i)$ на p_{i+1} означает, что вероятность события A зависит только от текущего состояния $i+1$ и не зависит от прошлых значений блуждания на предыдущих шагах. А зто обусловлено тем, что на каждом шаге частица совершает скачки вправо, влево или остается на месте с одними и теми же вероятностями, зависящими от текущего состояния и не зависящими от положения частицы на предыдущих шагах. Марковское свойство случайного блуждания состоит, таким образом, в том, что будущее положение частицы при фиксированном настоящем и прошлом ее положениях не зависит от прошлых положений, а зависит только от текущего ее положения. Если начальное состояние частицы 0, то она уже поглотилась в нуле, т.е. $p_0 = 1$. Если же начальное состояние частицы a, то она уже поглотилась в a, т.е. не может поглотиться в 0, следовательно $p_a = 0$. Поэтому к линейным разностным уравнениям второго порядка с постоянными коэффициентами (5.6) надо добавить граничные условия

$$p_0 = 1, \;\; p_a = 0. \tag{5.7}$$

Решение (5.6) ищем в виде $p_i = \lambda^i$. Подставляя его в (5.6) и деля обе части на λ^{i-1}, получим характеристическое уравнение

$$\lambda = p\lambda^2 + q + r\lambda$$

или

$$p\lambda^2 - (p+q)\lambda + q = 0,$$

откуда

$$\lambda_1 = \mu; \quad \lambda_2 = 1,$$

где

$$\mu = \frac{q}{p}.$$

Поскольку при $p \neq q$ характеристическое уравнение не имеет кратных корней, то в этом случае общее решение (5.6) запишется в виде

$$p_i = C_1\left(\frac{q}{p}\right)^i + C_2, \qquad \mu \neq 1.$$

Учитывая начальные условия (5.7), получим:

$$C_1 + C_2 = 1, \quad C_1\left(\frac{q}{p}\right)^a + C_2 = 0,$$

откуда

$$C_1 = -\frac{1}{\mu^a - 1}; \qquad C_2 = \frac{\mu^a}{\mu^a - 1},$$

следовательно,

$$p_i = \frac{\mu^a - \mu^i}{\mu_a - 1} \text{ при } \mu \neq 1. \tag{5.8}$$

Если же $p = q$, то характеристическое уравнение имеет кратный корень 1, поэтому общее решение (5.6) имеет вид

$$p_i = C_1 i + C_2.$$

Учитывая начальные условия (5.7), находим

$$C_1 = -\frac{1}{a}; \qquad C_2 = 1.$$

Поэтому

$$p_i = 1 - \frac{i}{a} \text{ при } \mu = 1. \tag{5.9}$$

Заметим, что если обозначить через q_i, r_i условные вероятности того, что частица поглотится в a и что она вообще не поглотится соответственно при условии, что она начинает движение из состояния i, то по формуле полной вероятности q_i и r_i будут удовлетворять тому же уравнению (5.6), что и p_i. Изменятся только граничные условия:

$$q_0 = 0, \quad q_a = 1; \quad r_0 = 0; \quad r_a = 0.$$

Поэтому

$$q_i = 1 - \frac{\mu^a - \mu^i}{\mu^a - 1} = \frac{\mu^i - 1}{\mu^a - 1} \ \text{ при } \ \mu \neq 1,$$

$$q_i = \frac{i}{a} \ \text{ при } \ \mu = 1,$$

$$r_i = 0 \ \text{ при всех } \ \mu.$$

Отметим, что решение при $\mu = 1$ может быть получено из решения при $\mu \neq 1$ предельным переходом при $\mu \to 1$:

$$p_i = \lim_{\mu \to 1} \frac{\mu^a - \mu^i}{\mu^a - 1} = 1 - i/a.$$

К рассмотренному случайному блужданию может быть сведена следующая классическая задача о разорении игрока. Подбрасывается монета с вероятностями выпадения герба и решки p и q соответственно ($p + q = 1$). Предположим, что один из игроков все время загадывает герб. Если угадывает, то получает от соперника рубль, если ошибается, то отдает рубль сопернику. Пусть a — суммарный капитал игроков, i — начальный капитал первого игрока. Игра ведется до полного разорения одного из соперников. Тогда, очевидно, что вероятность разорения первого игрока совпадает с вероятностью поглощения в нуле p_i описанного выше случайного блуждания, в котором $r = 0$.

Зададимся вопросом: какими должны быть p и q для того, чтобы игра была "справедливой"? Для этого надо, чтобы вероятности разорения для соперников были одинаковыми, т.е. $p_i = 1/2$. Согласно (5.8), искомое уравнение запишется в виде

$$\frac{\mu^a - \mu^i}{\mu^a - 1} = \frac{1}{2}.$$

Пусть, например, a — четное, $i = \frac{a}{2}$. Тогда уравнение для определения "справедливой" игры принимает форму

$$\mu^a - 2\mu^{\frac{a}{2}} + 1 = 0,$$

т.е. $\mu = 1$, другими словами $p = q = \frac{1}{2}$, что интуитивно понятно (если игроки имеют одинаковый начальный капитал $a/2$, то на каждом шаге у них должны быть одинаковые шансы выигрыша и проигрыша).

Теорема 5.6. *Пусть выполнены условия теоремы 5.5 и A — событие, имеющее ненулевую вероятность. Тогда имеет место формула Байеса*

$$P(H_j/A) = \frac{P(H_j)P(A/H_j)}{P(A)} = \frac{P(H_j)P(A/H_j)}{\sum\limits_{i=1}^{n} P(H_i)P(A/H_i)}, \; j = \overline{1,n}.$$

Доказательство. По определению условной вероятности

$$P(H_j/A) = \frac{P(H_jA)}{P(A)} = \frac{P(H_j)P(A/H_j)}{P(A)}.$$

Остается в знаменателе заменить $P(A)$ выражением из формулы полной вероятности. □

Замечание 5. Вероятности гипотез $P(H_1), P(H_2), \ldots, P(H_n)$ часто называют **априорными** вероятностями гипотез (т.е. данными свыше или доопытными вероятностями гипотез). Если стало известно, что в результате рассматриваемого эксперимента произошло некоторое событие A, то с учетом этой дополнительной информации априорные вероятности переоцениваются и заменяются вероятностями $P(H_1/A), P(H_2/A), \ldots, P(H_n/A)$, которые называются **апостериорными** вероятностями гипотез (т.е. послеопытными вероятностями гипотез для тех опытов, экспериментов, которые привели к наступлению события A).

Пример 7. Пусть известно, что в условиях примера 5 из второй урны после перекладывания в нее шара из первой урны появился белый шар. Какова вероятность того, что из первой урны во вторую был переложен белый шар?

Здесь речь идет об апостериорной вероятности гипотезы H_1 при дополнительной информации о появлении белого шара из второй урны. По формуле Байеса

$$P(H_1/A) = \frac{P(H_1)P(A/H_1)}{P(A)} = \frac{\frac{2}{5} \cdot \frac{3}{5}}{\frac{12}{25}} = \frac{1}{2}.$$

Таким образом, после переоценки априорной вероятности $P(H_1) = \frac{2}{5}$ и замены ее апостериорной вероятностью, она увеличилась до $1/2$. Это — результат действия дополнительной информации о том, что из первой урны во вторую был переложен белый шар.

§ 6. Произведение вероятностных пространств. Схема независимых испытаний Бернулли. Полиномиальная схема

Под испытанием будем понимать случайный эксперимент, который описывается некоторым вероятностным пространством $(\Omega, \mathcal{F}, P)$. Поэтому можно отождествить испытание с этим вероятностным пространством. Если Ω конечно, то в качестве σ-алгебры $\mathcal{F}$ мы договорились брать совокупность всех подмножеств множества Ω. В силу такого фиксированного выбора $\mathcal{F}$ в обозначении вероятностного пространства с конечным Ω его можно опустить. Поэтому в дальнейшем будем обозначать такое вероятностное пространство символом $[\Omega,\ P]$.

Пусть имеется n испытаний, т.е. n вероятностных пространств

$$[\Omega_1,\ P_1],\ [\Omega_2,\ P_2],\ \dots,\ [\Omega_n,\ P_n].$$

Здесь будет предполагаться, что $\Omega_1,\ \Omega_2,\ \dots,\ \Omega_n$ конечны. Для регистрации результата n этих испытаний в качестве элементарного исхода естественно взять $\omega = (\omega_1, \omega_2, \dots, \omega_n)$, где $\omega_1 \in \Omega_1, \omega_2 \in \Omega_2, \dots, \omega_n \in \Omega_n$. Таким образом, пространство элементарных исходов, описывающее n различных случайных экспериментов, есть $\Omega = \Omega_1 \times \Omega_2 \times \dots \times \Omega_n$. Например, если проводятся два испытания: сначала подбрасывается монета, а затем игральная кость, то $\Omega_1 = \{\textit{г}, \textit{р}\}$, а $\Omega_2 = \{1,2,3,4,5,6\}$. При этом пара $(\textit{г},3)$ регистрирует элементарный исход из $\Omega = \Omega_1 \times \Omega_2$, означающий, что на монете выпал герб, а на игральной кости выпало 3 очка. Про σ-алгебру событий пространства Ω можно опять не беспокоиться, поскольку оно конечно как прямое произведение конечных пространств. Как мы знаем, на каждом из конечных пространств Ω_i вероятности можно ввести с помощью конструктивно-аксиоматического определения: каждому ω_i ставится в соответствие число $p_i(\omega_i) \geq 0$, так что при этом

$$\sum_{\omega_i \in \Omega_i} p_i(\omega_i) = 1 \quad , i = \overline{1, n},$$

а вероятность в i-м испытании определяется как

$$P_i(A_i) = \sum_{\omega_i \in A_i} p_i(\omega_i), \quad A_i \subset \Omega_i.$$

Как ввести вероятность в конечном пространстве элементарных исходов

$$\Omega = \Omega_1 \times \Omega_2 \times \ldots \times \Omega_n =$$

$$= \{\omega = (\omega_1, \omega_2, \ldots, \omega_n) : \ \omega_1 \in \Omega_1, \omega_2 \in \Omega_2, \ldots, \omega_n \in \Omega_n\},$$

описывающем n испытаний $[\Omega_1, P_1], [\Omega_2, P_2], \ldots, [\Omega_n, P_n]$? Это можно сделать согласно конструктивно-аксиоматическому определению вероятности произвольным образом, ставя каждому $\omega \in \Omega$ в соответствие число $p(\omega) \geq 0$ так, что при этом

$$\sum_{\omega \in \Omega} p(\omega) = 1.$$

Однако если мы хотим ввести вероятность таким образом, чтобы испытания были независимыми, мы потребуем, чтобы эта вероятность определялась так:

$$p(\omega) = p_1(\omega_1) p_2(\omega_2) \ldots p_n(\omega_n) \tag{6.1}$$

для любого $\omega = (\omega_1, \ \omega_2, \ \ldots, \ \omega_n) \in \Omega$. Вероятность любого события $A \subset \Omega$ при этом есть

$$P(A) = \sum_{\omega \in \Omega} p(\omega).$$

Определение 6.1. Вероятность P, определенная с помощью (6.1), называется **прямым произведением вероятностей** P_i и обозначается $P = P_1 \times P_2 \times \ldots \times P_n$. Испытания $[\Omega_1,\ P_1]$, $[\Omega_2,\ P_2]$, $\ldots$, $[\Omega_n,\ P_n]$ называются **независимыми**, если на вероятностном пространстве $\Omega = \Omega_1 \times \Omega_2 \times \ldots \times \Omega_n$ выбрана вероятностная мера $P = P_1 \times P_2 \times \ldots \times P_n$.

Пусть событие имеет вид $A = A_1 \times A_2 \times \ldots A_n$, где $A_i \subset \Omega_i, \quad i = \overline{1, n}$. Оно происходит тогда и только тогда, когда в первом испытании происходит A_1, во втором испытании происходит A_2 и т.д., наконец в последнем n-м испытании происходит A_n. События такого вида назовем прямоугольниками. Если испытания независимы, т.е. выполнено (6.1), то

$$P(A) = \sum_{\omega\in\Omega} p(\omega) = \sum_{\omega_1\in\Omega_1} \sum_{\omega_2\in\Omega_2} \ldots \sum_{\omega_n\in\Omega_n} p_1(\omega_1)\, p_2(\omega_2)\ \ldots$$

$$\ldots\ p_n(\omega_n) = \sum_{\omega_1\in\Omega_1} p_1(\omega_1) \sum_{\omega_2\in\Omega_2} p_2(\omega_2)\ \ldots \sum_{\omega_n\in\Omega_n} p_n(\omega_n) =$$

$$= P_1(A_1)P_2(A_2)\ldots P_n(A_n). \qquad (6.2)$$

Сопоставим каждому событию $A_i \subset \Omega_i$ событие

$$\widetilde{A}_i = \Omega_1 \times \Omega_2 \times \ldots \times \Omega_{i-1} \times A_i \times \Omega_{i+1} \times \ldots \times \Omega_n \subset \Omega =$$

$$= \Omega_1 \times \Omega_2 \times \ldots \times \Omega_n.$$

Из (6.2) следует, что $P(\widetilde{A}_i) = P_i(A_i)$. Таким образом, между событиями A_i из вероятностного пространства $[\Omega_i, P_i]$ (соответствующего i-му испытанию) и событиями

$\widetilde{A}_i$ из вероятностного пространства $[\Omega, P]$ (соответствующего n испытаниям) устанавливается естественный изоморфизм, сохраняющий вероятности изоморфных событий. Поскольку

$$A = A_1 \times A_2 \times \ldots \times A_n = \bigcap_{k=1}^{n} \widetilde{A}_k,$$

то независимость испытаний, в силу (6.2), эквивалентна тому, что

$$P(\bigcap_{k=1}^{n} \widetilde{A}_k) = \prod_{k=1}^{n} P(\widetilde{A}_k),$$

т.е. тому, что образы $\widetilde{A}_1, \quad \widetilde{A}_2, \quad \ldots \widetilde{A}_n$ событий $A_1, \quad A_2, \quad \ldots A_n$ при рассматриваемом изоморфизме являются независимыми событиями вероятностного пространства $[\Omega,\ P]$.

Замечание 1. Если испытания описываются произвольными вероятностными пространствами $(\Omega_1,\ \mathcal{F}_1,\ P_1)$, $(\Omega_2,\ \mathcal{F}_2,\ P_2)$, $\ldots$, $(\Omega_n,\ \mathcal{F}_n,\ P_n)$, то дело обстоит значительно сложнее. В качестве пространства элементарных исходов, соответствующего n испытаниям, снова берется $\Omega = \Omega_1 \times \Omega_2 \times \ldots \times \Omega_n$. Оказывается, что прямое произведение $\mathcal{F}_1 \times \mathcal{F}_2 \times \ldots \times \mathcal{F}_n$ σ-алгебр $\mathcal{F}_1, \mathcal{F}_2, \ldots, \mathcal{F}_n$ не является σ-алгеброй (здесь под прямым произведением σ-алгебр понимается класс множеств, состоящий из всех прямоугольников $A_1 \times A_2 \times \ldots \times A_n$, где $A_1 \in \mathcal{F}_1,\ A_2 \in \mathcal{F}_2,\ \ldots,\ A_n \in \mathcal{F}_n$). В качестве σ-алгебры $\mathcal{F}$, соответствующей n испытаниям, берется наименьшая σ-алгебра, содержащая прямое произведение $\mathcal{F}_1 \times \mathcal{F}_2 \times \ldots \times \mathcal{F}_n$. На прямом произведении

$\mathcal{F}_1 \times \mathcal{F}_2 \times \ldots \times \mathcal{F}_n$ определяется функция P с помощью равенства, аналогичного (6.2):

$$P(A) = P_1(A_1)P_2(A_2)\ldots P_n(A_n), \tag{6.3}$$

$$\text{если } A = A_1 \times A_2 \times \ldots \times A_n, \text{где } A_1 \in \mathcal{F}_1,$$
$$A_2 \in \mathcal{F}_2,\ \ldots,\ A_n \in \mathcal{F}_n.$$

Можно доказать, что существует и притом единственная мера P на $\mathcal{F}$, совпадающая с функцией множеств на $\mathcal{F}_1 \times \mathcal{F}_2 \times \ldots \times \mathcal{F}_n$, определенной с помощью равенства (6.3). Эта вероятностная мера называется прямым произведением вероятностных мер $P_1, P_2, \ldots, P_n$ и обозначается $P = P_1 \times P_2 \times \ldots \times P_n$. При этом вероятностное пространство $(\Omega, \mathcal{F}, P)$ называется прямым произведением вероятностных пространств $(\Omega_1, \mathcal{F}_1, P_1)$, $(\Omega_2, \mathcal{F}_2, P_2)$, $\ldots$, $(\Omega_n, \mathcal{F}_n, P_n)$. В этом случае испытания называются независимыми.

Определение 6.2. Конечное вероятностное пространство $[\Omega, P]$, где $\Omega = \Omega_1 \times \Omega_2 \times \ldots \times \Omega_n$, $P = P_1 \times P_2 \times \ldots \times P_n$, называется **произведением вероятностных пространств** $[\Omega_1, P_1]$, $[\Omega_2, P_2]$, $\ldots$, $[\Omega_n, P_n]$.

Таким образом, произведение вероятностных пространств служит математической моделью независимых испытаний. При этом, если испытания причинно не связаны, то причинно не связаны любые события $A_1 \subset \Omega_1$, $A_2 \subset \Omega_2$, $\ldots$, $A_n \subset \Omega_n$. Но тогда не связаны причинно и изоморфные им события $\widetilde{A}_1 \subset \Omega$, $\widetilde{A}_2 \subset \Omega$, $\ldots$, $\widetilde{A}_n \subset \Omega$. Согласно сформулированному при изучении независимых событий нематематическому принципу отсюда следует, что они суть независимые события, что эквивалентно (6.2), а следовательно (6.1). Мы можем теперь переформулировать упомянутый практический принцип (нематематический) следующим образом:

Если испытания причинно не связаны между собой, то они автоматически независимы в смысле определения 6.1.

Перейдем к так называемой схеме Бернулли. Сначала дадим неформальное (нематематическое) определение.

Определение 6.3. Повторные независимые испытания, в каждом из которых возможны два исхода, вероятности которых не меняются от испытания к испытанию, называются **испытаниями Бернулли**.

Обозначим эти исходы буквами У и Н и назовем **успехом** и **неудачей**. В данном случае $\Omega_1 = \Omega_2 = \ldots \Omega_n = \{\text{У,Н}\}$. Пусть $p_1(\text{У}) = p,\ p_1(\text{Н}) = q$. Очевидно, $p \geq 0,\ q \geq 0,\ p + q = 1$. Испытания идентичны в том смысле, что все n испытаний описываются одним и тем же вероятностным пространством $[\Omega_1,\ P_1]$. Дадим теперь формальное определение.

Определение 6.3$'$. Схемой из n независимых испытаний Бернулли называется вероятностное пространство $[\Omega_1^n,\ P_1^n]$, где $\Omega_1 = \{\text{У,Н}\}$.

Здесь Ω_1^n — n-кратное декартово произведение Ω_1 на себя, P_1^n — n-кратное прямое произведение P_1 на себя.

Пример 1. Пусть две грани правильной игральной кости окрашены в красный цвет, а четыре — в синий. Кость подбрасывается 10 раз. Очевидно, что это — испытания Бернулли с $n = 10$. Если за успех принять выпадение красной грани, а за неудачу — выпадение синей, то $p = 1/3,\ q = 2/3$.

Обозначим через $\omega = (\omega_1,\ \omega_2,\ \dots,\ \omega_n)$ элементарный исход пространства Ω_1^n, описывающий результат n независимых испытаний Бернулли. Здесь каждое ω_i может принимать одно из двух возможных значений У или Н. Например, элементарный исход (У,У,У,Н,Н,...,Н) означает, что в первых трех испытаниях наступил успех, а в остальных $n-3$ испытаниях произошли неудачи. Так как элементарный исход представляет собой упорядоченную выборку с возвращением объема n из множества {У,Н}, то общее число элементарных исходов в пространстве Ω_1^n равно 2^n.

Обозначим через $\mu = \mu_n$ число успехов в n испытаниях. Указать, какое именно значение примет μ до проведения испытаний, мы не можем, так как не знаем какой именно исход осуществится в результате n испытаний. Но если результат этих испытаний известен, то, очевидно, μ становится полностью известным:

$$\mu = \mu(\omega) = \text{число букв У в конечной последовательности}$$

$$\omega = (\omega_1,\ \omega_2,\ \dots,\ \omega_n).$$

Таким образом, $\mu = \mu(\omega)$ является функцией, определенной на пространстве элементарных исходов и принимающей действительные значения. Такие функции называют **случайными величинами**. Позже мы систематически изучим их. Итак, число успехов в n испытаниях — случайная величина.

Обозначим через $P_n(k) = P\{\omega:\ \mu_n(\omega) = k\} = P\{\mu = k\}$ вероятность того, что в n испытаниях Бернулли произойдет ровно k успехов (более точно вероятность множества тех элементарных исходов $\omega = (\omega_1,\ \omega_2,\ \ldots,\ \omega_n)$ из Ω_1^n, на которых $\mu(\omega)$ принимает значение k, поскольку вероятность P определена на подмножествах Ω_1^n). С практической точки зрения интерес представляет именно эта вероятность. Например, если речь идет об игре двух соперников в шахматы, то нас интересует не вероятность распределения выигрышей и проигрышей определенного игрока по n партиям, а общее число выигранных им партий. Событию $\{\mu = k\}$ благоприятствуют те элементарные исходы $\omega = (\omega_1,\ \omega_2,\ \ldots,\ \omega_n)$ из Ω, в которых среди $\omega_1,\ \omega_2,\ \ldots,\ \omega_n$ ровно k букв У, а остальные — буквы Н. По конструктивно-аксиоматическому определению вероятности

$$P_n(k) = P\{\mu = k\} =$$

$$= \sum_{\text{по всем } \omega \text{ таким, что среди } \omega_1,\, \omega_2,\, \ldots,\, \omega_n \text{ ровно } k \text{ букв У}} p(\omega). \tag{6.4}$$

Но для ω, фигурирующих в сумме (6.4), в силу определения 6.1 выполнено соотношение (6.1); следовательно для всех них вероятности одинаковы и равны

$$p(\omega) = p^k q^{n-k}.$$

Поэтому сумма слагаемых (6.4) равна произведению числа слагаемых на отдельное слагаемое:

$$P_n(k) = [\text{число слагаемых}]p^k q^{n-k}.$$

Число слагаемых равно числу способов, с помощью которых из n мест для $\omega_1,\ \omega_2,\ \ldots,\ \omega_n$ можно выбрать k мест для буквы У, а остальные места заполнить буквами Н, т.е. C_n^k. Поэтому окончательно получим:

$$P_n(k) = C_n^k p^k q^{n-k}, \quad k = \overline{0, n}. \tag{6.5}$$

Формула (6.5) называется **формулой Бернулли**. Набор вероятностей $\{C_n^k p^k q^{n-k}, \quad k = \overline{0,n}\}$ называется **биномиальным распределением вероятностей**. Так как одно из событий $\{\mu = k\}, \quad k = \overline{0,n}$, обязательно происходит в схеме Бернулли, а никакие попарно различные из них не могут произойти одновременно, то они образуют полную группу; по свойству вероятности сумма вероятностей этих событий равна единице:

$$\sum_{k=0}^{n} P_n(k) = 1.$$

Обязательное выполнение этого равенства вполне согласуется с тем, что по формуле бинома Ньютона

$$\sum_{k=0}^{n} C_n^k p^k q^{n-k} = (p+q)^n = 1^n = 1.$$

Пример 2. Два равносильных противника играют в шахматы. Если партия заканчивается ничьей, то она считается недействительной и аннулируется. Что вероятнее, выиграть а)две партии из четырех или три партии из шести; б)не менее двух партий из четырех или трех партий из шести?

Равносильность противников означает, что вероятность выигрыша и проигрыша в одной результативной партии для каждого из соперников равна $p = q = 0.5$. По формуле Бернулли

$$P_4(2) = C_4^2 (0.5)^2 (0.5)^2 = 6/16;$$

$$P_6(3) = C_6^3 (0.5)^3 (0.5)^3 = 5/16,$$

т.е. вероятнее выиграть две партии из четырех, чем три партии из шести. Далее,

$$P(\mu_4 \geq 2) = P(\{\mu = 2\} + \{\mu = 3\} + \{\mu = 4\}) = P_4(2) +$$

$$+ P_4(3) + P_4(4) = (C_4^2 + C_4^3 + C_4^4)(0.5)^4 = 11/16 = 22/32;$$

$$P(\mu_6 \geq 3) = P_6(3) + P_6(4) + P_6(5) + P_6(6) = 1 - P_6(0) -$$

$$- P_6(1) - P_6(2) = 1 - (C_6^0 + C_6^1 + C_6^2)(0.5)^6 = 21/32.$$

Таким образом, вероятнее выиграть не менее двух партий из четырех, чем не менее трех партий из шести.

Задача 1. Пусть k_0 — наивероятнейшее число успехов в n испытаниях Бернулли, т.е. то значение k, для которого $P_n(k)$ принимает наибольшее значение при $k = \overline{0, n}$. Покажите, что k_0 может быть найдено из двойного неравенства

$$np + q \leq k_0 \leq np + p.$$

Указание. Составьте отношение $\frac{P_n(k+1)}{P_n(k)}$ и исследуйте с его помощью функцию $P_n(k)$ целочисленного аргумента k на возрастание, постоянство и убывание.

Рассмотрим одно обобщение схемы Бернулли, когда в одном испытании возможны не два исхода, а несколько.

Определение 6.4. Повторные независимые испытания, в каждом из которых возможны $k \geq 2$ исходов, вероятности которых не меняются от испытания к испытанию, называются **испытаниями полиномиальной (мультиномиальной) схемы**.

В частности, испытания Бернулли являются испытаниями биномиальной схемы ($k = 2$). Обозначим исходы отдельного испытания через $u_1, u_2, \dots, u_k$. Более строгое с математической точки зрения определение полиномиальной схемы выглядит так.

Определение 6.4$'$. Полиномиальной (мультиномиальной) схемой из n независимых испытаний называется вероятностное пространство $[\Omega_1^n, \ \ P_1^n]$, где $\Omega_1 = \{u_1, u_2, \dots, u_k\}$.

Обозначим вероятности исходов в отдельном испытании через

$$p_1 = p_1(u_1), \ p_2 = p_1(u_2), \ \dots, \ p_k = p_1(u_k).$$

Очевидно, что

$$p_1, p_2, \dots, p_k \geq 0 \quad \text{и} \quad p_1 + p_2 + \ldots + p_k = 1.$$

Обозначим через $\omega = (\omega_1, \ \omega_2, \ \dots, \ \omega_n)$ элементарный исход пространства Ω_1^n, описывающий результат n независимых испытаний полиномиальной схемы. Здесь каждое ω_i может принимать одно из k возможных значений $u_1, u_2, ..., u_k$. Очевидно, в пространстве Ω_1^n имеется k^n элементарных исходов. При этом вероятности элементарных исходов в силу независимости испытаний по-прежнему определяются равенством (6.1), которое в данном случае принимает вид

$$p(\omega) = p_1(\omega_1)p_1(\omega_2)\dots p_1(\omega_n). \tag{6.6}$$

Введем следующие случайные величины: μ_1 — число исходов u_1 в n испытаниях, μ_2 — число исходов u_2 в n испытаниях и т.д., наконец, μ_k — число исходов u_k в n испытаниях. Более формально, как функции на пространстве элементарных исходов, они определяются так:

$$\mu_1 = \mu_1(\omega) = [\text{число символов } u_1 \text{ среди } \omega_1,\, \omega_2,\, \ldots,\, \omega_n],$$

$$\mu_2 = \mu_2(\omega) = [\text{число символов } u_2 \text{ среди } \omega_1,\, \omega_2,\, \ldots,\, \omega_n]$$

и т.д., наконец,

$$\mu_k = \mu_k(\omega) = [\text{число символов } u_k \text{ среди } \omega_1,\, \omega_2,\, \ldots,\, \omega_n].$$

Замечание 2. В теории вероятностей принято обозначать $\{\omega : \ R\}$ множество тех элементарных исходов $\omega \in \Omega$, для которых выполнено свойство R. В функциональном анализе это множество чаще обозначают как $\{\omega |\ R\}$. Но в теории вероятностей уже есть обозначения $P(A \setminus B)$ для вероятности разности событий A и B, $P(A/B)$ для условной вероятности события A при условии B. Если обозначить через $P(\omega |\ R)$ вероятность множества тех элементарных событий, для которых выполнено свойство R, то в наших обозначениях будут использоваться три вида черточек, что может внести путаницу.

Рассмотрим событие $\{\omega : \ \mu_1(\omega) = r_1,\ \mu_2(\omega) = r_2,\ \ldots,\ \mu_k(\omega) = r_k\}$, где $r_1, r_2, \ldots, r_k$ — целые неотрицательные числа, дающие в сумме n (в иных случаях множество таких ω будет пустым множеством (невозможным событием)). Ему благоприятствуют те и только те элементарные исходы $\omega = (\omega_1,\ \omega_2,\ \ldots,\ \omega_n)$, для которых среди $\omega_1,\ \omega_2,\ \ldots,\ \omega_n$ будет ровно r_1 символов u_1,

r_2 символов u_2, и т.д., наконец, r_k символов u_k. Согласно конструктивно-аксиоматическому определению вероятности в дискретном пространстве вероятность этого события будет равна

$$P\{\omega:\ \mu_1(\omega)=r_1,\ \mu_2(\omega)=r_2,\ \dots,\ \mu_k(\omega)=r_k\} =$$

$$= \sum_{\omega:\ \mu_1(\omega)=r_1,\ \mu_2(\omega)=r_2,\ \dots,\ \mu_k(\omega)=r_k} p(\omega).$$

Так как испытания независимы, то в силу (6.6) вероятности, фигурирующие в этой сумме, совпадают и равны

$$p(\omega) = p_1^{r_1} p_2^{r_2} \dots p_k^{r_k}.$$

Поэтому сумма будет равна произведению числа слагаемых на отдельное слагаемое:

$$P\{\mu_1 = r_1,\ \mu_2 = r_2,\ \dots,\ \mu_k = r_k\} =$$

$$= [\text{число слагаемых}]\ p_1^{r_1} p_2^{r_2} \dots p_k^{r_k}.$$

Но число слагаемых совпадает с числом способов, с помощью которых множество $\{\omega_1, \omega_2, \dots, \omega_n\}$ можно разбить на k групп так, чтобы в первую группу вошло ровно r_1 элементов для символа u_1, во вторую группу вошло ровно r_2 элементов для символа u_2 и т.д., наконец, в k-ю группу вошло ровно r_k элементов для символа u_k. По теореме 3.5 число таких способов равно

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

Таким образом, вероятность того, что в n испытаниях полиномиальной схемы произойдет ровно r_1 исходов u_1, ровно r_2 исходов u_2 и т.д., наконец, ровно r_k исходов u_k, равна

$$P\{\mu_1 = r_1,\ \mu_2 = r_2,\ \ldots,\ \mu_k = r_k\} = \\ = \frac{n!}{r_1!r_2!\ldots r_k!} p_1^{r_1} p_2^{r_2} \ldots p_k^{r_k}, \qquad (6.7)$$

где $r_1, r_2, \ldots, r_k$ — целые неотрицательные числа, дающие в сумме n. Так как события $\{\mu_1 = r_1,\ \mu_2 = r_2,\ \ldots,\ \mu_k = r_k\}$ для различных неотрицательных значений $r_1, r_2, \ldots, r_k$, дающих в сумме n, образуют полную группу, то

$$\sum_{r_1 + r_2 + \ldots + r_k = n} \frac{n!}{r_1!\, r_2!\, \ldots\, r_k!} p_1^{r_1} p_2^{r_2} \ldots p_k^{r_k} = 1.$$

Согласно формуле (3.10) эта сумма равна

$$(p_1 + p_2 + \ldots + p_k)^n = 1^n = 1.$$

Это косвенно подтверждает справедливость формулы (6.7), которая называется **полиномиальной** или **мультиномиальной формулой**.

Пример 3. На отрезок $[0, 1]$, разбитый на три части $[0, 1/4)$, $[1/4,\ 3/4)$, $[3/4, 1]$, "наудачу" брошено 5 точек. Какова вероятность того, что на первые два промежутка разбиения попадет по две точки, а на последний промежуток — одна точка?

Когда в той или иной задаче говорится о бросании нескольких точек на отрезок, то подразумевается, что для каждой отдельной точки вероятность попадания в измеримое по Лебегу подмножество отрезка $[0, 1]$ определяется по правилу геометрической вероятности, а испытания, связанные с бросанием различных точек, независимы. Пусть μ_1 — число точек, попавших в $[0, 1/4)$, μ_2 —

число точек, попавших в $[1/4,\ 3/4)$, μ_3 — число точек, попавших в $[3/4, 1]$. По формуле (6.7) искомая вероятность равна

$$P\{\mu_1 = 2,\ \mu_2 = 2,\ \mu_3 = 1\} = \frac{5!}{2!2!1!}(1/4)^2(1/2)^2 1/4 =$$

$$= 15/128.$$

§ 7. Предельные теоремы в схеме Бернулли

Вероятность того, что изделие данного завода является бракованным, равна 0.005. Какова вероятность того, что из 10 000 наудачу взятых изделий этого завода бракованными окажутся а) 40 деталей; б) не более 70 деталей?

Согласно формуле Бернулли вероятность того, что бракованными окажутся 40 деталей, равна

$$P_{10000}(40) = C_{10000}^{40}(0.005)^{40}(0.995)^{9960}.$$

Вероятность того, что бракованными окажутся не более 70 деталей, равна

$$P\{\mu \le 70\} = \sum_{k=0}^{70} P_{10000}(k) =$$

$$= \sum_{k=0}^{70} C_{10000}^{k}(0.005)^{k}(0.995)^{10000-k}.$$

Видно, что для подсчета этих вероятностей требуется провести достаточно громоздкие вычисления. Даже при использовании компьютера возникнут сложности с вычислением C_{10000}^k, поскольку придется умножать друг на друга большое количество множителей порядка 10^4. В ситуации рассмотренного примера можно применить предельную теорему, предложенную Пуассоном. Заслуга Пуассона состоит не в доказательстве приводимого ниже результата, ибо оно является простым упражнением по математическому анализу, а в том, что он предложил асимптотику вероятности $P_n(k)$ в схеме Бернулли при стремлении n к бесконечности определенным способом, в котором p зависит от n.

Теорема 7.1 (Предельная теорема Пуассона). *Если в схеме Бернулли $n \to \infty$, $p \to 0$ таким образом, что $np = \lambda$ остается постоянным, то*

$$P_n(k) \to P(k) = \frac{\lambda^k}{k!} e^{-\lambda}, \quad k = 1, 2, \dots .$$

Доказательство. По условию теоремы $p = \frac{\lambda}{n}$, поэтому $q = 1 - p = 1 - \frac{\lambda}{n}$. Следовательно,

$$\lim_{n\to\infty} C_n^k p^k q^{n-k} = \lim_{n\to\infty} \frac{n(n-1)\dots(n-k+1)}{k!} \left(\frac{\lambda}{n}\right)^k \cdot$$

$$\cdot \left(1 - \frac{\lambda}{n}\right)^{n-k} = \frac{\lambda^k}{k!} \lim_{n\to\infty} \left(1 - \frac{1}{n}\right)\left(1 - \frac{2}{n}\right)\dots\left(1 - \frac{k-1}{n}\right)\cdot$$

$$\cdot \left(1 - \frac{\lambda}{n}\right)^{-k} \left(1 - \frac{\lambda}{n}\right)^{n} = \frac{\lambda^k}{k!} \lim_{n\to\infty} \left(1 - \frac{\lambda}{n}\right)^{n} = \frac{\lambda^k}{k!} e^{-\lambda}.$$

$\square$

Замечание 1. Теоремой Пуассона можно пользоваться для приближенного вычисления вероятности $P_n(k)$ при больших n и малых p. Для удобства сформулируем не вполне четкое с математической точки зрения следствие.

Следствие 7.2. Если в схеме Бернулли n велико, а p мало так, что при этом $\lambda = np$ не мало и не велико, то

$$P_n(k) \approx \frac{\lambda^k}{k!} e^{-\lambda}.$$

Отметим также, что теоремой Пуассона можно пользоваться, когда n велико, а p близка к единице. В этом случае надо успех переименовать на неудачу, а неудачу — на успех. Тогда вероятность "нового" успеха будет близка к нулю.

Пример 1. Коммутатор обслуживает 100 абонентов. Вероятность того, что в течение минуты абонент позвонит на коммутатор, равна 0.01. Найдите вероятность того, что в течение минуты на коммутатор позвонят а)ровно 3 абонента; б)менее трех абонентов.

Считая, что абоненты звонят независимо друг от друга и принимая за успех звонок отдельного абонента, получаем схему Бернулли из $n = 100$ испытаний с $p = 0.01$, $q = 0.99$. Так как $n = 100$ велико, $p = 0.01$ мало, а $\lambda = np = 1$ не мало и не велико, то по следствию 1

$$a) P_{100}(3) \approx \frac{1^3}{3!} e^{-1} = 0.0613;$$

$$\text{б)} P\{\mu < 3\} = P_{100}(0) + P_{100}(1) + P_{100}(2) \approx$$

$$\approx \frac{1^0}{0!} e^{-1} + \frac{1^1}{1!} e^{-1} + \frac{1^2}{2!} e^{-1} = 0.9197.$$

Что же делать, когда вероятность успеха p далеко отстоит от концов отрезка $[0, 1]$? В этом случае на помощь приходит так называемая теорема Муавра-Лапласа, доказанная впервые для случая симметричной схемы Бернулли (т.е. при $p = q = 1/2$) Муавром, а затем обобщенная на случай произвольного $0 < p < 1$ Лапласом. Прежде, чем переходить к формулировкам соответствующих теорем, напомним некоторые факты из математического анализа. Последовательности $\{a_n\}$ и $\{b_n\}$ называются эквивалентными (при $n \to \infty$), если

$$\lim_{n\to\infty} \frac{b_n}{a_n} = 1. \tag{7.1}$$

Факт эквивалентности последовательностей $\{a_n\}$ и $\{b_n\}$ обозначается так: $b_n \sim a_n$. Если дробь, стоящая под знаком предела в (7.1), зависит от некоторой последовательности $\{x_n\}$, а стремление к пределу в (7.1) равномерно относительно всех $x_n \in X$, где X — некоторое множество (т.е. для любого $\varepsilon > 0$ существует натуральное число N такое, что для всех $n > N$ и для всех $x_n \in X$ выполняется неравенство $|b_n/a_n - 1| < \varepsilon$), то пишут $b_n \sim a_n$ равномерно относительно $x_n \in X$ (или равномерно по $x_n \in X$). Известно, что если $a_n \sim b_n$ равномерно относительно $x_n \in X$, а $c_n \sim d_n$ равномерно относительно $x_n \in X$, то и равномерно относительно $x_n \in X$ выполняется:

$$a_n c_n \sim b_n d_n; \qquad \frac{a_n}{c_n} \sim \frac{b_n}{d_n}.$$

Для неформального понимания доказательства приводимой далее теоремы полезно вспомнить совсем простое доказательство этих свойств. Напомним также, что согласно общему определению $o(a_n)$ при $n \to \infty$ для бесконечно малой величины при $n \to \infty$ можно использовать обозначение $o(1)$.

Локальная предельная теорема Муавра-Лапласа.

Пусть в схеме Бернулли $0 < p < 1$, *тогда при* $n \to \infty$ *равномерно по всем* x_k *вида* $x_k = \frac{k-np}{\sqrt{npq}}$, *принадлежащим конечному отрезку* [a,b],

$$P_n(k) \sim \frac{1}{\sqrt{npq}}\varphi(x_k),$$

где $\varphi(x) = \frac{1}{\sqrt{2\pi}} \exp\{-\frac{x^2}{2}\}$ — *так называемая плотность стандартного нормального распределения.*

Д о к а з а т е л ь с т в о. Воспользуемся известной из математического анализа асимптотической формулой Стирлинга:

$$n! \sim \sqrt{2\pi n}\ \ n^n e^{-n} \qquad \text{при } n \to \infty. \tag{7.2}$$

По формуле Бернулли

$$P_n(k) = C_n^k p^k q^{n-k} = \frac{n!}{k!(n-k)!} p^k q^{n-k}. \tag{7.3}$$

По условию

$$x_k = \frac{k-np}{\sqrt{npq}} \in [a,b],$$

поэтому при $n \to \infty$ равномерно относительно $x_k \in [a,b]$ будет

$$k = np + x_k\sqrt{npq} \to \infty, \ \ n-k = nq - x_k\sqrt{npq} \to \infty.$$

Поэтому из (7.3) с учетом (7.2) получим, что равномерно относительно $x_k \in [a,b]$ будет выполняться

$$P_n(k) \sim \frac{\sqrt{2\pi n}\ n^n e^{-n}}{\sqrt{2\pi k}\ k^k e^{-k} \sqrt{2\pi(n-k)}\ (n-k)^{n-k} e^{-(n-k)}} =$$

$$= \frac{1}{\sqrt{2\pi}} \sqrt{\frac{n}{k(n-k)}} \Big(\frac{np}{k}\Big)^k \Big(\frac{nq}{n-k}\Big)^{n-k}.$$

Пользуясь формулой Тейлора

$$\ln(1+x) = x - \frac{x^2}{2} + o(x^2), \qquad x \to 0,$$

и тем, что

$$\frac{k}{np} = 1 + x_k \sqrt{\frac{q}{np}}; \quad \frac{n-k}{nq} = 1 - x_k \sqrt{\frac{p}{nq}},$$

получим:

$$\ln\Big(\frac{np}{k}\Big)^k = -\ln\Big(\frac{k}{np}\Big)^k = -k \ln\frac{k}{np} = -(np + x_k\sqrt{npq})\cdot$$

$$\ln\Big(1+x_k\sqrt{\frac{q}{np}}\Big) = -(np+x_k\sqrt{npq})\Big(x_k\sqrt{\frac{q}{np}} - \frac{qx_k^2}{2np} + o\Big(\frac{1}{n}\Big)\Big) =$$

$$= -x_k\sqrt{npq} - qx_k^2 + \frac{qx_k^2}{2} + o(1) = -x_k\sqrt{npq} - \frac{qx_k^2}{2} + o(1).$$

Аналогично, равномерно по $x_k \in [a,b]$

$$\ln\Big(\frac{nq}{n-k}\Big)^{n-k} = -\ln\Big(\frac{n-k}{nq}\Big)^{n-k} = -(n-k)\ln\frac{n-k}{nq} =$$

$$= -(nq - x_k\sqrt{npq}) \ln\Big(1 - x_k\sqrt{\frac{p}{nq}}\Big) =$$

$$= -(nq - x_k\sqrt{npq})\Big(-x_k\sqrt{\frac{p}{nq}} - \frac{px_k^2}{2nq} + o\Big(\frac{1}{n}\Big)\Big) =$$

$$= x_k\sqrt{npq} - px_k^2 + \frac{px_k^2}{2} + o(1) = x_k\sqrt{npq} - \frac{px_k^2}{2} + o(1).$$

Поэтому равномерно по таким x_k

$$\ln\left(\frac{np}{k}\right)^k\left(\frac{nq}{n-k}\right)^{n-k} = -\frac{x_k^2}{2} + o(1).$$

Следовательно,

$$\left(\frac{np}{k}\right)^k\left(\frac{nq}{n-k}\right)^{n-k} = \exp\Big\{-\frac{x_k^2}{2}\Big\}\exp\{o(1)\} \sim$$

$$\sim \exp\Big\{-\frac{x_k^2}{2}\Big\}$$

равномерно по $x_k \in [a,b]$. Поскольку равномерно относительно таких x_k

$$\sqrt{\frac{n}{k(n-k)}} \sim \sqrt{\frac{n}{npnq}} = \frac{1}{\sqrt{npq}},$$

то равномерно по $x_k \in [a,b]$

$$P_n(k) \sim \frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{npq}}e^{-\frac{x_k^2}{2}} = \frac{1}{\sqrt{npq}}\varphi(x_k).$$

□

Следствие 7.3. Если в схеме Бернулли n велико, а вероятность успеха p не очень близка к концам интервала (0,1) ($\sqrt{npq}$ велико), то

$$P_n(k) \approx \frac{1}{\sqrt{npq}}\varphi(x_k), \text{ где } x_k = \frac{k-np}{\sqrt{npq}}.$$

Поскольку

$$P\Big\{a \le \frac{\mu - np}{\sqrt{npq}} \le b\Big\} = \sum_{k:\ a \le x_k \le b} P_n(k) \approx$$

$$\approx \sum_{k:\, a \le x_k \le b} \varphi(x_k)\Delta x_k \approx \int\limits_a^b \varphi(x)dx,$$

где $\Delta x_k = \frac{1}{\sqrt{npq}}$, то весьма правдоподобен следующий результат:

Интегральная предельная теорема Муавра-Лапласа. *Пусть в схеме Бернулли* $0 < p < 1$. *Тогда равномерно по всем* a *и* b *таким, что* $-\infty \le a < b \le +\infty$,

$$\lim_{n\to\infty} P\Big\{a \le \frac{\mu - np}{\sqrt{npq}} \le b\Big\} = \int\limits_a^b \varphi(x)dx = \varPhi(b) - \varPhi(a) =$$

$$= N(b) - N(a).$$

Здесь

$$\varPhi(x) = \int\limits_0^x \varphi(t)dt \quad \text{— так называемая функция Лапласа,}$$

$N(x) = \int\limits_{-\infty}^{x} \varphi(t)dt$ — так называемая функция стандартного нормального распределения.

Доказательство этой теоремы будет являться следствием гораздо более общей так называемой центральной предельной теоремы и будет дано позже.

В силу интегральной теоремы вероятность того, что число успехов может принимать целые значения от k_1 до k_2 включительно, равна

$$P_n(k_1, k_2) = P\{k_1 \le \mu \le k_2\} =$$

$$= P\Big\{\frac{k_1 - np}{\sqrt{npq}} \le \frac{\mu - np}{\sqrt{npq}} \le \frac{k_2 - np}{\sqrt{npq}}\Big\} \approx$$

$$\approx \Phi\Big(\frac{k_2 - np}{\sqrt{npq}}\Big) - \Phi\Big(\frac{k_1 - np}{\sqrt{npq}}\Big).$$

Получается следующее

Следствие 7.4. *Если в схеме Бернулли вероятность успеха p не очень близка к 0 или 1, то при больших n ($\sqrt{npq}$ велико)*

$$P_n(k_1, k_2) \approx \Phi(x_{k_2}) - \Phi(x_{k_1}),$$

где $x_k = \frac{k-np}{\sqrt{npq}}$.

Пусть в результате некоторого случайного эксперимента может появиться или не появиться событие A. Всегда можно факторизовать исходное пространство элементарных исходов, объединяя исходы, благоприятствующие A, в исход A, а исходы, не благоприятствующие A, в исход $\overline{A}$. Таким образом, в качестве σ-алгебры берем $\mathcal{F} = \{\Omega, \phi, A, \overline{A}\}$. Считая, что A — успех, $\overline{A}$ — неудача, получаем, что вероятность произвольного события A можно отождествить с вероятностью успеха: $p = P(A)$. Тогда μ — число успехов в n испытаниях — это число появлений события A в n испытаниях, $\frac{\mu}{n}$ — относительная частота появления события A в n испытаниях. При аксиоматизации теории вероятностей мы стремились к тому, чтобы при больших n относительная частота $\frac{\mu}{n}$ стабилизировалась около вероятности p. А как оценить, насколько относительная частота может отклоняться от вероятности. В силу интегральной теоремы Муавра-Лапласа

$$P\Big\{\Big|\frac{\mu}{n} - p\Big| \le \varepsilon\Big\} = P\{-n\varepsilon \le \mu - np \le n\varepsilon\} =$$

$$= P\Big\{ - \varepsilon\sqrt{\frac{n}{pq}} \le \frac{\mu - np}{\sqrt{npq}} \le \varepsilon\sqrt{\frac{n}{pq}}\Big\} \approx$$

$$\approx \Phi\Big(\varepsilon\sqrt{\frac{n}{pq}}\Big) - \Phi\Big(-\varepsilon\sqrt{\frac{n}{pq}}\Big) = 2\Phi\Big(\varepsilon\sqrt{\frac{n}{pq}}\Big).$$

Итак, получили удобную в приложениях формулу для оценки степени отклонения относительной частоты события от вероятности этого события:

$$P\Big\{\Big|\frac{\mu}{n} - p\Big| \le \varepsilon\Big\} \approx 2\Phi\Big(\varepsilon\sqrt{\frac{n}{pq}}\Big).$$

Так как правая часть этого приближенного равенства при любом $\varepsilon > 0$ стремится к 1, то весьма правдоподобно, что для всякого $\varepsilon > 0$

$$\lim_{n\to\infty} P\Big\{\Big|\frac{\mu}{n} - p\Big| \le \varepsilon\Big\} = 1.$$

Это так называемый закон больших чисел в форме Бернулли, который будет строго доказан позже. Выполнение этого закона показывает разумность введенной нами аксиоматики теории вероятностей, ибо в ее рамках показано, что в определенном смысле относительная частота события стабилизируется вокруг его вероятности, а именно с вероятностью, сколь угодно близкой к единице, она сколь угодно мало отклоняется от этой вероятности.

§ 8. Борелевские множества и борелевские функции. Пополнение вероятностного пространства

Этот параграф является вспомогательным и необходим для понимания вводимого далее понятия случайной величины. Часть материала известна из курса функционального анализа, другая часть является дальнейшим ее развитием.

Пусть X — метрическое пространство. В силу леммы 4.4 для любого семейства подмножеств пространства X, в частности, для семейства всех открытых множеств из X, существует наименьшая σ-алгебра, содержащая это семейство.

Определение 8.1. Наименьшая σ-алгебра, содержащая все открытые множества метрического пространства X, называется **σ-алгеброй борелевских множеств** пространства X, а множества из нее — **борелевскими**.

Будем через $\mathcal{B}(X)$ обозначать эту σ-алгебру. Как всякая σ-алгебра, $\mathcal{B}(X)$ содержит $\emptyset$ и все пространство X. По определению 8.1 она содержит все открытые множества пространства X. Так как любое замкнутое множество F представляется в виде $F = X \setminus G$, где G — открытое множество, то, по определению σ-алгебры, $F \in \mathcal{B}(X)$. Итак, все замкнутые множества являются борелевскими. Нетрудно понять, что мы могли бы дать эквивалентное определение $\mathcal{B}(X)$ как наименьшей σ-алгебры, содержащей все замкнутые множества пространства X.

Поскольку числовая прямая $\mathbb{R}$ является метрическим пространством, можно рассматривать σ-алгебру борелевских множеств числовой прямой, которую для краткости обозначим через $\mathcal{B}$, т.е. $\mathcal{B} = \mathcal{B}(\mathbb{R})$. Зададимся вопросом, насколько широк класс борелевских множеств.

1. *Одноточечные множества являются борелевскими.*

Действительно, одноточечные множества являются замкнутыми. Возможны и другие мотивировки, например,

$$\{a\} = \bigcap_{n=1}^{\infty} \Big(a - \frac{1}{n},\ a + \frac{1}{n}\Big),$$

а так как интервалы, стоящие под знаком пересечения — открытые множества, т.е. борелевские, то по определению σ-алгебры $\{a\} \in \mathcal{B}$.

2. *Множество рациональных точек — борелевское.*

Действительно, оно является счетным объединением одноточечных множеств.

3. *Множество иррациональных точек — борелевское.*

Действительно, оно является дополнением к множеству рациональных точек.

4. *Любые промежутки* $(-\infty, +\infty)$, $(-\infty, a)$, $(-\infty, a]$, $(a, +\infty)$, $[a, +\infty)$, (a, b), $[a, b]$, $[a, b)$, $(a, b]$ — *борелевские множества.*

Это очевидно, и Вы должны уметь доказывать. Например, $[a, b) \in$ ß, так как $[a, b) = (a, b) \cup \{b\}$.

5. *Если* $f(x)$ — *непрерывная функция, то множества*

$$\{x :\ f(x) < c\},\ \{x :\ f(x) \le c\},\ \{x :\ a \le f(x) < b\}$$

и т.п., — борелевские.

Действительно, например, $\{x : f(x) < c\} = f^{-1}((-\infty,\ c))$ является открытым, т.е. борелевским, поскольку при непрерывном отображении прообраз открытого множества $(-\infty,\ c)$ открыт. Далее,

$$\{x :\ a \le f(x) < b\} = \{x :\ a < f(x) < b\} \cup \{x :\ f(x) = a\}.$$

Первое слагаемое в объединении — открытое множество, т.е. борелевское, как прообраз открытого множества (a, b) при непрерывном отображении f, второе слагаемое — замкнутое множество, т.е. борелевское, как прообраз замкнутого множества $\{a\}$ при непрерывном отображении f. Так как $\mathcal{B}$ — алгебра, то и объединение будет принадлежать $\mathcal{B}$.

Довольно трудно привести пример множества, не являющегося борелевским. Все множества, которые могут быть сконструированы из множеств, фигурирующих в свойствах 1-5, с помощью конечного или счетного числа теоретико-множественных операций, являются борелевскими. Для построения примера не борелевского множества нельзя обойтись без аксиомы выбора, подобно тому, как строится пример множества, не измеримого по Лебегу. Понятно, что все множества, имеющие какой-то практический интерес, являются борелевскими. Поэтому в теории вероятностей на числовой прямой в качестве основной берется σ-алгебра борелевских множеств, что с запасом обеспечивает все возможные прикладные потребности.

В дальнейшем нам понадобится другая конструкция σ-алгебры $\mathcal{B}$, основанная не на открытых множествах, а на полуинтервалах. Пусть $[a, b) = \{x : \ a \leq x < b\}$ для всех $-\infty \leq a \leq b \leq +\infty$. Здесь введено обозначение $[-\infty, a)$ для интервала $(-\infty, a)$ и обозначение $[-\infty, +\infty)$ для $(-\infty, +\infty)$, чтобы не делать лишних оговорок в приводимой ниже конструкции. Таким образом, вся числовая прямая $\mathbb{R}$ и пустое множество $\phi = [a, a)$ также являются полуинтервалами такого вида. Пусть $\mathcal{A}$ — совокупность всех подмножеств $\mathbb{R}$, которые могут быть представлены в виде конечных объединений попарно не пересекающихся полуинтервалов такого вида, т.е.

$$\mathcal{A} = \{A:\ \forall n \in \mathbb{N},\ a_i \le b_i,\ i = \overline{1,n};\ A = \sum_{i-1}^{n} [a_i, b_i)\}.$$

Задача 1. Покажите, что $\mathcal{A}$ — алгебра.

В дальнейшем будем называть эту алгебру **алгеброй объединений полуинтервалов**.

Однако $\mathcal{A}$ не является σ-алгеброй, так как $\forall n \in \mathbb{N}$ $\left[\frac{1}{n}, 1\right) \in \mathcal{A}$, однако

$$\bigcup_{n=1}^{\infty} \left[\frac{1}{n}, 1\right) = (0, 1) \notin \mathcal{A}.$$

Лемма о строении σ-алгебры борелевских множеств на $\mathbb{R}$. *σ-алгебра борелевских множеств числовой прямой совпадает с наименьшей σ-алгеброй, содержащей алгебру $\mathcal{A}$, т.е.*

$$\mathcal{B} = \sigma(\mathcal{A}).$$

Д о к а з а т е л ь с т в о. 1. Докажем сначала, что $\sigma(\mathcal{A}) \subset \mathcal{B}$. Множества из алгебры $\mathcal{A}$ — борелевские по свойству 4 и определению алгебры, т.е. $\mathcal{A} \subset \mathcal{B}$. Итак, $\mathcal{B}$ — σ-алгебра, содержащая $\mathcal{A}$, но $\sigma(\mathcal{A})$ — наименьшая σ-алгебра, содержащая $\mathcal{A}$. Поэтому $\sigma(\mathcal{A}) \subset \mathcal{B}$.

2. Теперь докажем, что $\mathcal{B} \subset \sigma(\mathcal{A})$. Заметим, что любой открытый интервал (a, b) содержится в $\sigma(\mathcal{A})$. Действительно, $\forall n \in \mathbb{N}$ будет $\left[a + \frac{1}{n}, b\right) \in \mathcal{A} \subset \sigma(\mathcal{A})$, а так как $\sigma(\mathcal{A})$ — σ-алгебра, то

$$(a, b) = \bigcup_{n=1}^{\infty} \left[a + \frac{1}{n}, b\right) \in \sigma(\mathcal{A}).$$

Из теории функций известно, что любое открытое множество на действительной прямой представляется в виде не более чем счетного объединения попарно непересекающихся интервалов. Пусть W — совокупность всех открытых множеств $\mathbb{R}$, $G \in W$. Согласно вышесказанному,

$$G = \sum_{i \in I} (a_i, b_i),$$

где $I \subset \mathbb{N}$. Поскольку, как выше доказано, $(a_i, b_i) \in \sigma(\mathcal{A})$ и $\sigma(\mathcal{A})$ — σ-алгебра, то $G \in \sigma(\mathcal{A})$. Поэтому $W \subset \sigma(\mathcal{A})$. По определению, $\mathcal{B}$ — наименьшая σ-алгебра, содержащая W, а $\sigma(\alpha)$ — некоторая σ-алгебра, содержащая W. Следовательно, $\mathcal{B} \subset \sigma(\mathcal{A})$. □

Так как отрезок $[0, 1]$ — метрическое пространство, то можно рассматривать σ-алгебру его борелевских множеств $\mathcal{B}([0, 1])$. Для нее верны те же свойства, что и для $\mathcal{B}(\mathbb{R})$, с той лишь оговоркой, что все множества берутся из $[0, 1]$, а роль единицы в этой σ-алгебре играет не $\mathbb{R}$, а $[0, 1]$. Любое борелевское множество из $\mathcal{B}([0, 1])$ можно представить как $B \cap [0, 1]$, где $B \in \mathcal{B}$. Таким образом, $\mathcal{B}([0, 1]) = \mathcal{B} \cap [0, 1]$. Если рассмотреть как и в случае борелевских множеств на прямой алгебру объединений полуинтервалов $\mathcal{A}([0,1])$, составленную точно таким же образом, что и $\mathcal{A}$, только интервалы $[a_i, b_i)$ должны браться из $[0,1]$, то легко доказывается, что $\mathcal{B}([0,1]) = \sigma(\mathcal{A}([0,1]))$.

Определение 8.2. Пусть D — борелевское множество. Функция $f : D \to \mathbb{R}$ называется **измеримой по Борелю** или **борелевской**, если прообраз борелевского множества является борелевским множеством, т.е. $\forall B \in \mathcal{B}$ выполняется $f^{-1}(B) \in \mathcal{B}$.

Теорема 8.3. *Если $f : D \to \mathbb{R}$ — непрерывная функция, то f — борелевская функция.*

Д о к а з а т е л ь с т в о. Рассмотрим класс подмножеств числовой прямой

$$\mathfrak{M} = \{B : \ B \subset R, \ f^{-1}(B) \in \mathcal{B}\}.$$

Сначала докажем, что $\mathfrak{M}$ — σ-алгебра подмножеств $\mathbb{R}$. Действительно, $f^{-1}(\mathbb{R}) = D \in \mathcal{B}$, следовательно, $\mathbb{R} \in \mathfrak{M}$. Далее, $f^{-1}(\emptyset) = \emptyset \in \mathcal{B}$, поэтому $\phi \in \mathfrak{M}$. Если $A, B \in \mathfrak{M}$, то $f^{-1}(A)$, $f^{-1}(B) \in \mathcal{B}$, следовательно,

$$f^{-1}(A \cup B) = f^{-1}(A) \cup f^{-1}(B) \in \mathcal{B},$$

$$f^{-1}(A \cap B) = f^{-1}(A) \cap f^{-1}(B) \in \mathcal{B},$$

$$f^{-1}(A \setminus B) = f^{-1}(A) \setminus f^{-1}(B) \in \mathcal{B}.$$

Поэтому $A \cup B$, $A \cap B$, $A \setminus B \in \mathfrak{M}$. Значит, $\mathfrak{M}$ — алгебра. Если $B_1, B_2, \ldots \in \mathfrak{M}$, то $f^{-1}(B_1), f^{-1}(B_2), \ldots \in \mathcal{B}$, следовательно,

$$f^{-1}(B_1 \cup B_2 \cup \ldots) = f^{-1}(B_1) \cup f^{-1}(B_2) \cup \ldots \in \mathcal{B},$$

$$f^{-1}(B_1 \cap B_2 \cap \dots) = f^{-1}(B_1) \cap f^{-1}(B_2) \cap \dots \in \mathcal{B}.$$

По определению $\mathfrak{M}$ это означает, что $\bigcup\limits_{n=1}^{\infty} B_n \in \mathfrak{M}$, $\bigcap\limits_{n=1}^{\infty} B_n \in \mathfrak{M}$. Поэтому алгебра $\mathfrak{M}$ является σ-алгеброй.

Пусть, как и раньше, W — совокупность всех открытых множеств $\mathbb{R}$. Так как f — непрерывная функция, то прообраз открытого множества — открытое множество, т.е. борелевское. Таким образом, $\forall G \in W$ выполняется $f^{-1}(G) \in \mathcal{B}$, т.е. $G \in \mathfrak{M}$. Итак, $W \subset \mathfrak{M}$, а так как $\mathfrak{M}$ — σ-алгебра, содержащая W, то она содержит наименьшую σ-алгебру, содержащую W, т.е. $\mathcal{B}$. Поскольку $\mathcal{B} \subset \mathfrak{M}$, то $\forall B \in \mathcal{B}$ выполняется $B \in \mathfrak{M}$, т.е. $f^{-1}(B) \in \mathcal{B}$. Последнее, согласно определению 8.2, означает, что f — борелевская функция. $\square$

Так как n-мерное евклидово пространство $\mathbb{R}^n$ — метрическое пространство, то можно говорить о σ-алгебре его борелевских множеств $\mathcal{B}(\mathbb{R}^n) = \mathcal{B}^n$. Пусть $D \subset \mathbb{R}^n$ — n-мерное борелевское множество (на самом деле D — единица σ-алгебры $D \cap \mathcal{B}^n$).

Определение 8.4. Функция $f: \ D \to \mathbb{R}$ называется **измеримой по Борелю** или **борелевской**, если $\forall B \in \mathcal{B}$ выполняется $f^{-1}(B) \in \mathcal{B}^n$ (на самом деле $f^{-1}(B) \in D \cap \mathcal{B}^n$).

Теорема 8.5. *Если $f: D \to \mathbb{R}$ — непрерывная функция, то f — борелевская функция.*

Доказательство почти дословно повторяет доказательство теоремы 8.3.

Задача 2. Докажите теорему 8.5.

Пусть $(\Omega, \mathcal{F}, P)$ — вероятностное пространство. Может оказаться, что у множества $A \in \mathcal{F}$, имеющего нулевую вероятность ($P(A) = 0$), существуют подмножества, не принадлежащие $\mathcal{F}$, и, следовательно, не имеющие вероятности. Для того, чтобы избавиться от указанного недостатка, введем следующее определение. Множество $N \subset \Omega$ называется P-**нулевым**, или **просто нулевым**, если существует множество $A \in \mathcal{F}$ такое, что $N \subset A$ и $P(A) = 0$.

Определение 8.6. Вероятностное пространство $(\Omega, \mathcal{F}, P)$ называется **полным**, если $\mathcal{F}$ содержит все нулевые множества.

Полное вероятностное пространство обладает тем свойством, что любое подмножество N множества A нулевой вероятности измеримо (т.е. принадлежит $\mathcal{F}$), и, следовательно, $P(N) = 0$.

Рассмотрим следующую процедуру пополнения вероятностного пространства. Пусть $\mathfrak{N}$ — класс нулевых множеств, $\overline{\mathcal{F}} = \{A \cup N : \ A \in \mathcal{F}, \ \ N \in \mathfrak{N}\}$. Определим на $\overline{\mathcal{F}}$ числовую функцию $\overline{P}$, полагая

$$\overline{P}(A \cup N) = P(A).$$

Докажем, что такое определение функции $\overline{P}$ корректно. Для этого необходимо показать, что если множество из $\overline{\mathcal{F}}$ имеет два представления $A_1 \cup N_1$ и $A_2 \cup N_2$, где $A_1, \ A_2 \in \mathcal{F}, \quad N_1, \ N_2 \in \mathfrak{N}$, то

$P(A_1) = P(A_2)$. Из $A_1 \cup N_1 = A_2 \cup N_2$ следует, что $(A_1 \setminus A_2) + (A_2 \setminus A_1) \subset [(A_2 \cup N_2) \setminus A_2] \cup [(A_1 \cup N_1) \setminus A_1] \subset N_1 \cup N_2$. Так как $P(N_1 \cup N_2) = 0$, то отсюда следует, что $P(A_1 \Delta A_2) = P((A_1 \setminus A_2) + (A_2 \setminus A_1)) = 0$. Следовательно, по свойству вероятности 4 пятого параграфа $P(A_1) = P(A_2)$.

Замечание 1. Когда давалось определение алгебры и σ-алгебры, мы не стремились минимизировать число требований, накладываемых на множества этих классов. Однако при доказательствах такая "расточительность" выливается в их усложнение. Поэтому приведем удобные необходимые и достаточные условия для того, чтобы множества некоторого семейства $\mathcal{F}$ подмножеств множества Ω образовывали σ-алгебру. Вот они:

1. $\emptyset \in \mathcal{F}$.
2. Если $A, B \in \mathcal{F}$, то $\overline{A} \in \mathcal{F}$, $A \cup B \in \mathcal{F}$.
3. Если $A_1, A_2, \ldots, \in \mathcal{F}$, то $\bigcup\limits_{n=1}^{\infty} A_n \in \mathcal{F}$.

Необходимость этих условий очевидна. Для доказательства достаточности необходимо аккуратно проверить все пункты определения σ-алгебры, используя соотношения двойственности.

Задача 3. Докажите достаточность этих условий.

Задача 4. Решите задачу 1 с помощью проверки условий 1.-3.

Задача 5. Модифицируйте часть доказательства теоремы 8.3, связанную с тем, что $\mathfrak{M}$ — σ-алгебра, используя условия 1. -3.

Решение задач 3-5 рекомендуется провести обязательно в связи с тем, что их решение закрепляет стандартные навыки, которые будут использованы в дальнейшем и которые играют важную роль в теории вероятностей.

Теорема 8.7.

1. $\overline{\mathcal{F}}$ *— σ-алгебра, которая является наименьшей σ-алгеброй, содержащей $\mathcal{F}$ и $\mathfrak{N}$.*

2. Функция $\overline{P}$ *является вероятностной мерой на* $\overline{\mathcal{F}}$.

3. Вероятностное пространство $(\Omega,\ \overline{\mathcal{F}},\ \overline{P})$ *является полным.*

Доказательство. 1. По определению нулевого множества $\emptyset \in \mathfrak{N}$. Поэтому $\emptyset \in \overline{\mathcal{F}}$, поскольку $\emptyset = \emptyset + \emptyset$, причем первое слагаемое принадлежит $\mathcal{F}$, а второе — $\mathfrak{N}$.

Пусть $A, B \in \overline{\mathcal{F}}$, т.е. $A = A_1 \cup N_1, \quad B = B_1 \cup N_2$, где $A_1,\ B_1 \in \mathcal{F}, \quad N_1,\ N_2 \in \mathfrak{N}$. Очевидно, объединение нулевых множеств — нулевое множество (если $N_1,\ N_2$ — нулевые множества, т.е. существуют $C_1,\ C_2 \in \mathcal{F}$ такие, что $N_1 \subset C_1$, $N_2 \subset C_2$ и $P(C_1) = P(C_2) = 0$, то $N_1 \cup N_2 \subset C_1 \cup C_2$, причем $P(C_1 \cup C_2) = 0$, поскольку $P(C_1 \cup C_2) \leq P(C_1) + P(C_2)$). Следовательно, $A \cup B = (A_1 \cup B_1) \cup (N_1 \cup N_2)$, причем $A_1 \cup B_1 \in \mathcal{F},\ N_1 \cup N_2 \in \mathfrak{N}$. Таким образом, $A \cup B \in \mathfrak{N}$, т.е. класс множеств $\overline{\mathcal{F}}$ замкнут относительно конечных объединений.

Похожим образом доказывается замкнутость этого класса относительно счетных объединений. Пусть $A_1,\ A_2,\ \ldots \in \overline{\mathcal{F}}$, т.е. $A_1 = D_1 \cup N_1,\ A_2 = D_2 \cup N_2,\ \ldots$, где $D_1,\ D_2,\ \ldots \in \mathcal{F}, \quad N_1,\ N_2,\ \ldots \in \mathfrak{N}$. Счетное объединение нулевых множеств — нулевое множество (доказательство аналогично проведенному выше для двух множеств с заменой свойства конечной полуаддитивности вероятности свойством счетной полуаддитивности 10 четвертого параграфа). Далее,

$$\bigcup_{n=1}^{\infty} A_n = \big(\bigcup_{n=1}^{\infty} D_n\big) \cup \big(\bigcup_{n=1}^{\infty} N_n\big),$$

причем $\bigcup\limits_{n=1}^{\infty} D_n \in \mathcal{F}$, $\bigcup\limits_{n=1}^{\infty} N_n \in \mathfrak{N}$, т.е. $\bigcup\limits_{n=1}^{\infty} A_n \in \overline{\mathcal{F}}$.

Пусть $A \in \overline{\mathcal{F}}$, т.е. $A = A_1 \cup N$, где $A_1 \in F$, $N \in \mathfrak{N}$. Так как N — нулевое множество, то найдется $A_2 \in \mathcal{F}$ такое, что $N \subset A_2$ и $P(A_2) = 0$. Очевидно,

$$\overline{A} = \overline{A_1 \cup A_2} + A_2(\overline{A_1 \cup N}),$$

причем $\overline{A_1 \cup A_2} \in \mathcal{F}$, $A_2(\overline{A_1 \cup N}) \in \mathfrak{N}$, поскольку $A_2(\overline{A_1 \cup N}) \subset A_2$, а $P(A_2) = 0$. Следовательно, $\overline{A} \in \overline{\mathcal{F}}$, т.е. $\overline{\mathcal{F}}$ замкнут относительно перехода к дополнительному множеству.

Таким образом, $\overline{\mathcal{F}}$ — σ-алгебра. Из ее построения видно, что $\overline{\mathcal{F}}$ — наименьшая σ-алгебра, содержащая $\mathcal{F}$ и $\mathfrak{N}$.

2. Покажем, что $\overline{P}$ — вероятностная мера на $\overline{\mathcal{F}}$. Аксиома неотрицательности очевидна. Поскольку $\Omega = \Omega + \emptyset$, то $\overline{P}(\Omega) = P(\Omega) = 1$. Остается доказать счетную аддитивность P. Пусть $A_n \in \overline{\mathcal{F}}$, $n = 1, 2, \ldots$, т.е. $A_n = D_n \cup N_n$, где $D_n \in \mathcal{F}$, $N_n \in \mathfrak{N}$. Пусть множества A_n, $n = 1, 2, \ldots$ попарно не пересекаются, тогда множества D_n, $n = 1, 2, \ldots$ попарно не пересекаются и множества N_n, $n = 1, 2, \ldots$ попарно не пересекаются. При этом

$$\sum_{n=1}^{\infty} A_n = \sum_{n=1}^{\infty} (D_n \cup N_n) = \big(\sum_{n=1}^{\infty} D_n\big) \cup \big(\sum_{n=1}^{\infty} N_n\big),$$

где $\sum\limits_{n=1}^{\infty} D_n \in \mathcal{F}$, $\sum\limits_{n=1}^{\infty} N_n \in \mathfrak{N}$. Следовательно,

$$\overline{P}\Big(\sum_{n=1}^{\infty} A_n\Big) = \overline{P}\Big(\Big(\sum_{n=1}^{\infty} D_n\Big) \cup \Big(\sum_{n=1}^{\infty} N_n\Big)\Big) = P\Big(\sum_{n=1}^{\infty} D_n\Big) =$$

$$= \sum_{n=1}^{\infty} P(D_n) = \sum_{n=1}^{\infty} \overline{P}(A_n).$$

3. Покажем, что вероятностное пространство $(\Omega, \overline{\mathcal{F}}, \overline{P})$ — полное. Пусть $C \in \mathcal{F}$ — нулевое множество, т.е. существует $A \cup N$ такое, что $A \in \mathcal{F}$, $N \in \mathfrak{N}$, $C \subset A \cup N$ и $\overline{P}(A \cup N) = 0$. Тогда $P(A) = 0$. Так как N — нулевое, то существует $B \in \mathcal{F}$, такое, что $N \subset B$ и $P(B) = 0$. Очевидно, $C \subset A \cup B$, $P(A \cup B) \leq P(A) + P(B) = 0$. Значит, $P(C) = 0$, т.е. $C \in \mathfrak{N}$, следовательно, $C = \emptyset + C$, причем $\emptyset \in \mathcal{F}$, $C \in \mathfrak{N}$. Но это означает, что $C \in \overline{\mathcal{F}}$, т.е. любое нулевое множество содержится в $\overline{\mathcal{F}}$. $\square$

Мера $\overline{P}$ на $\overline{\mathcal{F}}$, фигурирующая в теореме, называется **пополнением меры** P на $\mathcal{F}$. Обычно ее обозначают той же буквой P, что и первоначальную меру. Эта процедура использовала аксиому нормированности вероятности P только в том месте, где доказывалось, что $\overline{P}(\Omega) = 1$ и потому может быть применена к любой мере (не обязательно вероятностной).

Оказывается, что мера Лебега на $\mathbb{R}$ может быть получена следующим образом. Сначала определяется мера Лебега на полуалгебре всех полуинтервалов вида $[a, b)$, где $-\infty \le a \le b \le +\infty$, с помощью равенства

$$\mu([a,b)) = b - a.$$

затем эта мера единственным образом продолжается на алгебру объединений полуинтервалов $\mathcal{A}$ с помощью соотношения

$$\mu(A) = \sum_{i=1}^{n} \mu([a_i, b_i)),$$

если $A = \sum\limits_{i=1}^{n} [a_i, b_i)$. Так как эта мера на $\mathcal{A}$ σ-конечна, то по теореме о продолжении меры, она единственным образом продолжается на наименьшую σ-алгебру, содержащую $\mathcal{A}$, т.е. на σ-алгебру борелевских множеств $\mathcal{B}$ (см. лемму о строении σ-алгебры борелевских множеств на $\mathbb{R}$). В продвинутых курсах по теории меры доказывается, что если пополнить меру μ, определенную на $\mathcal{B}$, с помощью описанной выше процедуры, то наименьшая σ-алгебра, содержащая $\mathcal{B}$ и все нулевые множества из $\mathcal{B}$, совпадет с σ-алгеброй измеримых по Лебегу множеств, а мера Лебега на этой σ-алгебре является пополнением меры μ, определенной на $\mathcal{B}$.

§ 9. Случайная величина. Распределение и функция распределения. Закон распределения

Одним из основных понятий теории вероятностей является понятие случайной величины. Прежде, чем переходить к формальному определению, рассмотрим ряд примеров таких величин.

1. Число вызовов, поступающих от абонентов на телефонную станцию в течение некоторого времени.

2. Число появлений успеха в схеме из n независимых испытаний Бернулли.

3. Скорость молекулы газа в некоторый момент времени.

4. Координаты точки попадания при выстреле по круговой мишени.

Со случайными величинами приходится иметь дело в самых различных областях науки и техники. Возникает проблема создания методов их изучения. Приведенные примеры показывают, что мы имеем дело с величинами, принимающими действительные числовые значения, причем каждая из величин под влиянием случайных обстоятельств в состоянии принимать различные значения. Значение, которое принимает случайная величина, меняется от опыта к опыту. Оставаясь на позициях качественного описания, можем сказать, что случайная величина — это переменная величина, значения которой зависят от случая. Попробуем формализовать это определение. Пусть $(\Omega, \ \mathcal{F}, \ P)$ — вероятностное пространство, являющееся математической моделью рассматриваемого случайного эксперимента. При одном исходе эксперимента случайная величина принимает одно значение, при другом, вообще говоря, — иное. Таким образом, каждому элементарному исходу $\omega \in \Omega$ ставится в соответствие некоторое действительное число. Например, если μ — число успехов в n независимых испытаниях Бернулли, то μ ставит элементарному исходу $\omega = (\omega_1, \omega_2, \dots, \omega_n)$ число букв "У" в конечной последовательности $\omega_1, \omega_2, \dots, \ \omega_n$. Таким образом, $\mu = \mu(\omega)$ есть функция, определенная на пространстве элементарных исходов. Естественно поэтому считать случайной величиной действительную функцию $\xi = \xi(\omega)$, определенную на пространстве элементарных исходов Ω. Другими словами, $\xi : \Omega \to \mathbb{R}$ — отображение пространства элементарных исходов в числовую прямую. Но этого мало. На практике в качестве наблюдаемых событий являются попадания значений случайной величины в те или иные простые множества, например, можно наблюдать событие $\{a < \xi < b\}$, состоящее в том, что значения случайной величины попадают в интервал (a, b). Относительная частота такого события может быть измерена (хотя бы принципиально). Следовательно должна существовать вероятность такого события. Но вероятность определена на σ-алгебре $\mathcal{F}$. Поэтому, чтобы оно действительно являлось событием, необходимо и достаточно, чтобы множество тех элементарных исходов, для которых $\xi(\omega)$ принадлежит (a, b), принадлежало σ-алгебре $\mathcal{F}$. Поэтому естественно ввести следующее

Определение 9.1. **(Действительной) случайной величиной** называется действительная функция $\xi : \Omega \to \mathbb{R}$, измеримая относительно σ-алгебры $\mathcal{F}$. При этом под **измеримостью относительно** $\mathcal{F}$ понимается выполнение одного из следующих двух эквивалентных условий:

1. $\forall x \in \mathbb{R}$ выполняется $\{\omega : \xi(\omega) < x\} \in \mathcal{F}$;
2. $\forall B \in \mathcal{B}$ выполняется
$\xi^{-1}(B) = \{\omega : \xi(\omega) \in B\} \in \mathcal{F}$.

На первый взгляд может показаться, что второе условие требует больше, чем первое. Действительно, из 2. следует 1. как частный случай, если в 2. положить $B = (-\infty, x)$. На самом деле из 1. также следует 2. за счет того, что семейство множеств B, для которых $\{\omega : \xi(\omega) \in$

$B\} \in \mathcal{F}$ образует σ-алгебру, которая содержит множества вида $(-\infty, x)$, а, значит, содержит $\mathcal{B}$. Проведем более подробное доказательство.

Д о к а з а т е л ь с т в о того, что из 1. следует 2. Рассмотрим семейство множеств

$$\mathfrak{M} = \{B : \ B \subset \mathbb{R}, \ \xi^{-1}(B) \in \mathcal{F}\}.$$

Докажем, что $\mathfrak{M}$ — σ-алгебра. Из того, что $\xi^{-1}(\emptyset) = \emptyset \in \mathcal{B}$ следует, что $\emptyset \in \mathfrak{M}$. Далее, если $A, \ B \in \mathfrak{M}$, т.е. $\xi^{-1}(A), \ \xi^{-1}(B) \in \mathcal{F}$, то $\xi^{-1}(\overline{A}) = \xi^{-1}(\mathbb{R} - A) = \xi^{-1}(\mathbb{R}) \setminus \xi^{-1}(A) = \Omega - \xi^{-1}(A) \in \mathcal{F}$, т.е. $\overline{A} \in \mathfrak{M}$. Кроме того, из $\xi^{-1}(A), \ \xi^{-1}(B) \in \mathcal{F}$ следует, что $\xi^{-1}(A \cup B) = \xi^{-1}(A) \cup \xi^{-1}(B) \in \mathcal{F}$, т.е. $A \cup B \in \mathfrak{M}$. Итак, $\mathfrak{M}$ — алгебра. Если к тому же для $n = 1, 2, \ldots \quad A_n \in \mathfrak{M}$, т.е. $\xi^{-1}(A_n) \in \mathcal{F}$, то $\xi^{-1}(\bigcup\limits_{n=1}^{\infty} A_n) = \bigcup\limits_{n=1}^{\infty} \xi^{-1}(A_n) \in \mathcal{F}$, а, значит, $\bigcup\limits_{n=1}^{\infty} A_n \in \mathfrak{M}$. Поэтому $\mathfrak{M}$ — σ-алгебра.

В силу 1. $\quad \xi^{-1}((-\infty, \ \ x)) \in \mathcal{F}$, следовательно, $(-\infty, \ x) \in \mathfrak{M} \quad \forall x \in \mathbb{R}$ и так как $\mathfrak{M}$ — σ-алгебра, то $(a, b) = (-\infty, b) - (\infty, a) \in \mathfrak{M}$, следовательно, множества вида $A = \sum\limits_{i=1}^{n} [a_i, \ b_i)$ принадлежат $\mathfrak{M}$. Значит, алгебра полуинтервалов $\mathcal{A} \subset \mathfrak{M}$. Так как $\mathfrak{M}$ — σ-алгебра, то $\sigma(\mathcal{A}) \subset \mathfrak{M}$, т.е. $\mathcal{B} \subset \mathfrak{M}$. Последнее означает, что $\forall B \in \mathcal{B}$ выполняется $B \in \mathfrak{M}$, что, по определению $\mathfrak{M}$, означает, что $\xi^{-1}(B) \in \mathcal{F}$, т.е. выполняется 2. $\quad \square$

Заметим, что с точки зрения отображений измеримых пространств случайная величина ξ есть $\mathcal{F}/\mathcal{B}$-измеримое отображение измеримого пространства $(\Omega, \ \mathcal{F})$ в измеримое пространство $(\mathbb{R}, \ \mathcal{B})$. Действительно, условие измеримости 2. эквивалентно тому, что $\xi^{-1}(\mathcal{B}) \subset \mathcal{F}$.

Отметим, что если пространство элементарных исходов Ω дискретно, то случайная величина — любая действительная функция $\xi = \xi(\omega)$, определенная на пространстве элементарных исходов Ω. Действительно, в этом случае мы договорились в качестве σ-алгебры $\mathcal{F}$ брать класс всех подмножеств множества Ω. Поэтому любое подмножество, в частности, подмножество $\{\omega : \ \xi(\omega) \in B\}$ принадлежит $\mathcal{F}$, и требование измеримости автоматически выполняется для любой функции $\xi = \xi(\omega)$.

Определение 9.2. **Функцией распределения** случайной величины ξ называется действительная функция действительной переменной $F = F_\xi : \ \mathbb{R} \to \mathbb{R}$, определенная равенством

$$F(x) = F_\xi(x) = P\{\omega : \ \xi(\omega) < x\} = P\{\xi < x\}, \quad x \in \mathbb{R}.$$

Определение 9.3. **Распределением** случайной величины ξ называется функция множеств $P_\xi : \ \mathcal{B} \to \mathbb{R}$, заданная равенством

$$P_\xi(B) = P(\xi^{-1}(B)) = P\{\omega : \ \xi(\omega) \in B\} = P\{\xi \in B\},$$
$$B \in \mathcal{B}. \tag{9.1}$$

Подчеркнем, что областью определения $F(x)$ является вся числовая прямая, областью определения $P_\xi(B)$ — σ-алгебра борелевских множеств числовой прямой. Индекс ξ у функции распределения ставится в том случае, когда рассматривается несколько случайных величин и надо подчеркнуть, к какой именно из них относится рассматриваемая функция распределения. События $\{\xi < x\}$ и $\{\xi \in B\}$ служат краткими обозначениями соответственно событий $\{\omega : \xi(\omega) < x\}$ и $\{\omega : \xi(\omega) \in B\}$ и для краткости речи соответственно называются "вероятностью того, что ξ принимает значение, меньшее x" и "вероятностью того, что ξ попадает в множество B". Еще раз подчеркнем, что измеримость случайной величины относительно σ-алгебры $\mathcal{F}$ необходима, в частности, для того, чтобы были корректными определения 9.1 и 9.2, т.е. чтобы множества, стоящие под знаком вероятности в этих определениях, являлись событиями (принадлежали $\mathcal{F}$), и их вероятности были определены.

Пример 1. Наудачу бросается "правильная" монета. Перед ее подбрасыванием игрок загадывает выпадение герба. Если выпадет герб, то игрок получит выигрыш одной денежной единицы (д.е.), если выпадет решка, то игрок отдаст сопернику одну д.е. Пусть ξ — выигрыш игрока в одной такой партии. В данном случае $\Omega = \{г, р\}$, причем $\xi(г) = 1$ д.е. и $\xi(р) = -1$д.е. Так как Ω дискретно, то измеримость выполнена автоматически. Итак, $\xi = \xi(\omega)$ — случайная величина.

Если $x \le -1$, то, поскольку $\{\omega : \xi(\omega) < x\} \subset \{\omega : \xi(\omega) < -1\}$, имеем $F(x) = P\{\omega : \xi(\omega) < x\} \le P\{\omega : \xi(\omega) < -1\} = P(\emptyset) = 0$. Итак, $F(x) = 0$ при $x \le -1$.

Если $-1 < x \le 1$, то $\{\omega : \xi(\omega) < x\} = \{\omega : \xi(\omega) = -1\} = \{р\}$, следовательно, $F(x) = P\{\omega : \xi(\omega) < x\} = P(\{р\}) = 1/2$.

Наконец, если $x > 1$, то $\{\omega : \xi(\omega) < x\} = \Omega$, поэтому $F(x) = P\{\omega : \xi(\omega) < x\} = P(\Omega) = 1$.

Таким образом, функция распределения случайной величины ξ равна

$$F(x) = \begin{cases} 0, & \text{если } x \le -1, \\ \frac{1}{2}, & \text{если } -1 < x \le 1, \\ 1, & \text{если } x > 1. \end{cases}$$

Отметим, что эта функция — неубывающая, непрерывная слева (в каждой точке), имеет предел, равный нулю, при $x \to -\infty$ и предел, равный единице, при $x \to +\infty$.

Распределение случайной величины ξ равно

$$P_\xi(B) = \begin{cases} 0, & \text{если } -1 \notin B,\ 1 \notin B, \\ \frac{1}{2}, & \text{если } -1 \in B,\ 1 \notin B \text{ или } -1 \notin B,\ 1 \in B, \\ 1, & \text{если } -1 \in B,\ 1 \in B. \end{cases}$$

Пример 2. На отрезок $[a, b]$ наудачу бросается точка, $\xi = \xi(\omega) \equiv \omega$ — ее координата ω. Эта случайная величина ставит в соответствие каждому элементарному исходу $\omega \in \Omega = [a, b]$ сам этот элементарный исход. Так как $\mathcal{F} = Leb([a, b])$ — σ-алгебра измеримых по Лебегу подмножеств отрезка $[a, b]$, и она содержит σ-алгебру $\mathcal{B}([a, b])$ борелевских множеств отрезка $[a, b]$, то

$$\forall B \in \mathcal{B} \text{ прообраз } \xi^{-1}(B) = \{\omega \in \Omega : \ \xi(\omega) \in B\} =$$

$$= \{\omega \in \Omega : \ \omega \in B\} = B \cap [a, b] \in \mathcal{B}([a, b]) \subset Leb([a, b]),$$

и $\xi(\omega)$ измерима относительно $\mathcal{F} = Leb([a, b])$, т.е. является случайной величиной.

Если $x \le a$, то событие $\{\omega : \ \xi(\omega) < x\}$ является невозможным, поэтому $F(x) = P\{\omega : \ \xi(\omega) < x\} = 0$.

Если $a < x \leq b$, то, в соответствие с геометрическим определением вероятности, $F(x) = P\{\omega : \quad \omega < x\} = \frac{\mu([a,x))}{b-a} = \frac{x-a}{b-a}$.

Наконец, если $x > b$, то событие $\{\omega : \quad \xi(\omega) < x\}$ является достоверным, следовательно, $F(x) = P(\Omega) = 1$.

Таким образом, функция распределения случайной величины ξ равна

$$F(x) = \begin{cases} 0, & \text{если } x \leq a, \\ \frac{x-a}{b-a}, & \text{если } a < x \leq b, \\ 1, & \text{если } x > b. \end{cases}$$

Про случайную величину с такой функцией распределения говорят, что она **равномерно распределена на отрезке** $[a, b]$ или что она **имеет равномерное распределение на отрезке** $[a, b]$. Итак, координата брошенной наудачу на отрезок $[a, b]$ точки имеет равномерное распределение на этом отрезке. Как и в предыдущем примере отметим, что функция равномерного распределения — неубывающая, непрерывная слева (даже непрерывная), имеет предел, равный нулю, при $x \to -\infty$ и предел, равный единице, при $x \to +\infty$. Как мы убедимся далее, эти свойства присущи любой функции распределения и в определенном смысле являются характеристическими.

Распределение ξ равно

$$P_\xi(B) = \frac{\mu(B \cap [a, b])}{b-a}, \qquad B \in \mathcal{B},$$

где $\mu(B)$ — мера Лебега множества B.

Замечание 1. В классической схеме вероятность события равна нулю тогда и только тогда, когда это событие является невозможным. Действительно, в силу классического определения вероятности, $P(A) = \frac{k}{n} = 0$ в том и только том случае, если $k = 0$, т.е. событию не благоприятствует ни один исход, следовательно, $A = \emptyset$. В случае, когда Ω бесконечно, нулевую вероятность могут иметь события, отличные от невозможного. Например, если ξ равномерно распределена на отрезке $[a, b]$, то $\forall x \in \mathbb{R}$ выполняется $P\{\xi = x\} = 0$. На первый взгляд имеется парадоксальная ситуация: вся числовая прямая состоит из точек, вероятность попасть в каждую из которых равна нулю, причем $1 = P\{\xi \in \mathbb{R}\} = \sum\limits_{x \in \mathbb{R}} P\{\xi = x\} = 0$. Так рассуждать нельзя. Во-первых, нельзя применять свойство счетной аддитивности к континуальному числу слагаемых, во-вторых, понятие суммы определяется либо для конечного числа слагаемых (обычная сумма), либо для счетного (сумма ряда). В континуальном случае роль суммы выполняет интеграл. И, действительно, в дальнейшем, при изучении случайных величин такого же типа как равномерно распределенная, мы используем интегралы. Здесь уместно провести аналогию с механикой. При непрерывном распределении массы по объему каждая отдельная материальная точка тела имеет нулевую массу, и тело складывается из континуального числа таких точек, что не мешает этому телу иметь конечную ненулевую массу. Для этого случая в механике вводится понятие плотности массы, с помощью которой масса всего тела выражается интегралом от плотности по объему этого тела. В дальнейшем будет применен аналогичный подход и в теории вероятностей.

Теорема 9.4. *Распределение вероятностей P_ξ является вероятностью (вероятностной мерой) на измери-*

мом пространстве $(\mathbb{R}, \mathcal{B})$.

Д о к а з а т е л ь с т в о. Функция P_ξ по формуле (9.1) ставит каждому $B \in \mathcal{B}$ в соответствие действительное число $P_\xi(B)$. При этом

1) $\forall B \in \mathcal{B} \quad P_\xi(B) = P(\xi^{-1}(B)) \geq 0$;

2) $P_\xi(\mathbb{R}) = P(\xi^{-1}(\mathbb{R})) = P(\Omega) = 1$;

3) если события $B_1, B_2, \ldots \in \mathcal{B}$ и попарно несовместимы, то, поскольку прообразы $\xi^{-1}(B_1), \xi^{-1}(B_2), \ldots$ непересекающихся множеств $B_1, B_2, \ldots$ сами не пересекаются, $P_\xi(\sum\limits_{n=1}^{\infty} B_n) = P(\xi^{-1}(\sum\limits_{n=1}^{\infty} B_n)) = P(\sum\limits_{n=1}^{\infty} \xi^{-1}(B_n)) = \sum\limits_{n=1}^{\infty} P(\xi^{-1}(B_n)) = \sum\limits_{n=1}^{\infty} P_\xi(B_n)$, $\square$

Так как $\{\omega : \ a \leq \xi(\omega) < b\} = \{\omega : \ \xi(\omega) < b\} - \{\omega : \xi(\omega) < a\}$, причем $\{\omega : \ \xi(\omega) < a\} \subset \{\omega : \ \xi(\omega) < b\}$, то

$$P\{a \leq \xi < b\} = P\{\omega : \ a \leq \xi(\omega) < b\} = F(b) - F(a). \tag{9.2}$$

Свойства функции распределения.

1. $0 \leq F(x) \leq 1$.

Это свойство очевидно, поскольку $F(x)$ является вероятностью при фиксированном x.

2. Функция распределения является неубывающей.

Действительно, если $x_1, x_2 \in \mathbb{R}, \quad x_1 < x_2$, то $\{\omega : \xi(\omega) < x_1\} \subset \{\omega : \ \xi(\omega) < x_2\}$ (если первое событие происходит, то происходит и второе). По свойству вероятности 4 из четвертого параграфа

$$F(x_1) = P\{\omega : \ \xi(\omega) < x_1\} \leq P\{\omega : \ \xi(\omega) < x_2\} = F(x_2).$$

В силу свойств 1 и 2 функция F является монотонной и ограниченной. Поэтому существуют ее пределы на $-\infty$ и $+\infty$, которые мы обозначим как

$$F(-\infty) = \lim_{x \to -\infty} F(x); \quad F(+\infty) = \lim_{x \to +\infty} F(x).$$

3. $F(-\infty) = 0, \quad F(+\infty) = 1.$

На интуитивном уровне эти свойства понятны: если $x \to -\infty$, то событие $\{\xi < x\}$ стремится к событию $\{\xi < -\infty\}$, которое является невозможным, поэтому

$$F(-\infty) = P\{\xi < -\infty\} = P(\emptyset) = 0.$$

аналогично, если $x \to +\infty$, то событие $\{\xi < x\}$ стремится к событию $\{\xi < +\infty\}$, которое является достоверным, поэтому

$$F(+\infty) = P\{\xi < +\infty\} = P(\Omega) = 1.$$

Это, конечно, не может служить доказательством, так как предельный переход, совершенный под знаком вероятности, требует дополнительного обоснования.

Д о к а з а т е л ь с т в о. В силу существования пределов достаточно доказать, что

$$\lim_{n \to +\infty} F(-n) = 0, \quad \lim_{n \to +\infty} F(+n) = 1.$$

Для этого мы рассмотрим последовательность множеств $\{A_n\}$, где $A_n = \{\omega : \quad -n \le \xi(\omega) < n\}$. Так как $A_n \subset A_{n+1}$ и $\bigcup\limits_{n=1}^{\infty} A_n = \Omega$, то, по определению сходимости последовательностей множеств, $A_n \uparrow \Omega$. По свойству непрерывности вероятности, $P(A_n) \to 1$, что в силу (9.2) означает: $F(+\infty) - F(-\infty) = 1$. В силу свойства 1 оба числа $F(+\infty)$ и $F(-\infty)$ неотрицательны, а их разность равна 1. Это возможно в том и только том случае, если первое из них равно 1, а второе — 0.

4. Функция распределения непрерывна слева (в каждой точке $\mathbb{R}$).

Функция F монотонна и ограничена, поэтому она может иметь разрывы только первого рода. Следовательно в каждой точке она имеет односторонние пределы. Итак, для любого $a \in \mathbb{R}$ существует предел $F(a+0) = \lim\limits_{x\to a+0} F(x)$. Требуется доказать, что $F(a+0) = F(a)$. Для этого достаточно доказать, что для какой-то последовательности $x_n \to a+0$ выполняется $F(x_n) \to F(a)$. Таким образом, достаточно доказать, что $F\left(a - \frac{1}{n}\right) \to F(a)$ при $n \to \infty$. Рассмотрим последовательность множеств B_n, где $B_n = \left\{\omega : \quad \xi(\omega) < a - \frac{1}{n}\right\}$. Если $B = \{\omega : \quad \xi(\omega) < a\}$, то, очевидно, $B_n \subset B_{n+1}$ и $\bigcup\limits_{n=1}^{\infty} B_n = B$, т.е. $B_n \uparrow B$. По свойству непрерывности вероятности это значит, что $P(B_n) \to P(B)$, т.е. $F\left(a - \frac{1}{n}\right) \to F(a)$.

5. $P\{\xi = a\} = P\{\omega : \ \xi(\omega) = a\} = F(a+0) - F(a)$, т.е. вероятность значения случайной величины равна скачку функции распределения при этом значении.

Рассмотрим последовательность множеств C_n, где $C_n = \left\{\omega : \quad a \le \xi < a + \frac{1}{n}\right\}$. Очевидно, $C_n \supset C_{n+1}$ и $\bigcap\limits_{n=1}^{\infty} C_n = C$, где $C = \{\omega : \quad \xi(\omega) = a\}$. Итак, $C_n \downarrow C$; по свойству непрерывности вероятности, $P(C_n) \to P(C)$, т.е.

$P\Big\{\omega: \ \ a \le \xi < a + \frac{1}{n}\Big\} \to P\{\omega: \ \ \xi(\omega) = a\}$. В силу (9.2) это означает, что $F\Big(a + \frac{1}{n}\Big) - F(a) \to P\{\omega: \ \ \xi(\omega) = a\}$, т.е. $P\{\omega: \ \ \xi(\omega) = a\} = F(a+0) - F(a)$.

Следствие 9.5. *Наряду с (9.2) выполняются равенства*

$$P\{a < \xi < b\} = F(b) - F(a+0),$$

$$P\{a \le \xi \le b\} = F(b+0) - F(a),$$

$$P\{a < \xi \le b\} = F(b+0) - F(a+0).$$

Действительно, $P\{a < \xi < b\} = P\{a \le \xi < b\} - P\{\xi = a\} = F(b) - F(a) - [F(a+0) - F(a)] = F(b) - F(a+0)$.

Задача 1. Докажите две другие формулы.

Следствие 9.6. *Если F непрерывна в точке a, то* $P\{\xi = a\} = 0$.

В частности,

Следствие 9.7. *Если F непрерывна, то для любого $x \in \mathbb{R}$ выполняется* $P\{\xi = x\} = 0$.

Замечание 2. Мы определили функцию распределения равенством $F(x) = P\{\xi < x\}$. Не менее часто функцию распределения определяют как $F(x) = P\{\xi \le x\}$. При таком ее определении все свойства, кроме 4,5, остаются в силе. Свойство 4 превращается в непрерывность F справа, а свойство 5 в $P\{\xi = a\} = F(a) - F(a-0)$. Формула (9.2) превращается в $P\{a < \xi \le b\} = F(b) - F(a)$, формулы в следствии 9.5 соответствующим образом изменяются. Следствия 9.6, 9.7 сохраняются.

Теорема о существовании случайной величины с заданной функцией распределения. *Пусть функция* $F: \mathbb{R} \to \mathbb{R}$

1) *не убывает,*

2) *непрерывна слева,*

3) $F(-\infty) = 0, \quad F(+\infty) = 1.$

Тогда существует вероятностное пространство $(\Omega,\ \mathcal{F},\ P)$ *и случайная величина* $\xi = \xi(\omega)$, *определенная на нем такая, что*

$$F_\xi(x) = F(x).$$

Д о к а з а т е л ь с т в о этой теоремы выносится в приложение. Отметим, что она дает возможность построения неограниченного числа примеров случайных величин. Для этого достаточно придумать функцию $F: \mathbb{R} \to \mathbb{R}$, удовлетворяющую условиям 1)-3) теоремы.

Определение 9.8. **Законом распределения** случайной величины ξ называется любая функция, однозначно определяющая ее функцию распределения.

В частности, сама функция распределения является законом распределения. Если известно распределение P_ξ, то функция распределения может быть найдена как

$$F(x) = P\{\xi < x\} = P\{\xi \in (-\infty,\ x)\} = P_\xi((-\infty,\ x)).$$

Оказывается, имеет место обратное утверждение.

Теорема 9.9. *Распределение вероятностей* P_ξ *однозначно определяется функцией распределения* $F_\xi(x)$.

Доказательство. Использование (9.2) показывает, что распределение однозначно определяется функцией распределения на полуалгебре интервалов вида $[a, b)$ как

$$P_\xi([a, b)) = F(b) - F(a).$$

В свою очередь, распределение на алгебре объединений полуинтервалов $\mathcal{A}$ однозначно определяется распределением на полуалгебре интервалов вида $[a, b)$ как

$$P_\xi(A) = \sum_{i=1}^{n} P_\xi([a_i, b_i)), \quad \text{где} \;\; A = \sum_{i=1}^{n} [a_i, b_i).$$

Так как распределение — вероятностная мера, то, по теореме о продолжении меры, на наименьшей σ-алгебре, содержащей алгебру $\mathcal{A}$, т.е. на σ-алгебре борелевских множеств, распределение однозначно определяется своими значениями на алгебре $\mathcal{A}$. Поэтому $P_\xi(B)$ определяется для всех $B \in \mathcal{B}$ значениями функции распределения. □

Таким образом, распределение и функция распределения однозначно определяют друг друга. Чаще пользуются функцией распределения, поскольку привычнее пользоваться действительными функциями действительного переменного, а не мерами. Однако иногда удобнее использовать распределение.

В силу вышесказанного, закон распределения можно определить как любую функцию, однозначно определяющую распределение. Итак, функция распределения и распределение являются универсальными законами распределения (т.е. годящимися для любой случайной величины). Чуть позже мы познакомимся со специфическими законами распределения, которые характеризуют определенные типы случайных величин.

Случайная величина $\xi = \xi(\omega)$, определенная на вероятностном пространстве $(\Omega, \mathcal{F}, P)$, индуцирует (порождает) новое вероятностное пространство $(\mathbb{R}, \mathcal{B}, P_\xi)$, которое называется **координатным пространством** случайной величины ξ. При этом $\mathbb{R}$ также называется **координатным пространством элементарных исходов**. На ξ можно смотреть как на элементарный исход в $\mathbb{R}$, а множествам элементарных исходов $B \in \mathcal{B}$ приписывается вероятность $P_\xi(B)$. Во многих задачах теории вероятностей важна не сама случайная величина, а ее закон распределения. В таком случае можно забыть о "функциональном" происхождении случайной величины, отождествляя ее с элементарным исходом координатного пространства $(\mathbb{R}, \mathcal{B}, P_\xi)$ и называя ее **координатной случайной величиной**. Такой огрубленный взгляд на случайную величину особенно уместен в математической статистике, когда имеются результаты ее измерений, а вероятностное пространство, на котором она определена как функция от ω, как правило, неизвестно.

Определение 9.10. Случайные величины ξ и η называются **эквивалентными**, если они совпадают почти всюду относительно вероятностной меры P, т.е.

$$P\{\omega : \ \xi(\omega) \neq \eta(\omega)\} = 0.$$

Другими словами, они совпадают на множестве полной вероятности:

$$P\{\omega : \ \xi(\omega) = \eta(\omega)\} = 1.$$

Задача 2. Покажите, что отношение эквивалентности между случайными величинами действительно является отношением эквивалентности, т.е., что оно рефлексивно, симметрично и транзитивно.

Поэтому множество всех случайных величин разбивается на непересекающиеся классы эквивалентных между собой случайных величин. Поскольку

$$P\{\xi \in B\} = P\{\xi \in B,\ \xi = \eta\} + P\{\xi \in B,\ \xi \neq \eta\} =$$
$$= P\{\xi \in B,\ \xi = \eta\} = P\{\eta \in B,\ \xi = \eta\} =$$
$$= P\{\eta \in B,\ \xi = \eta\} + P\{\eta \in B,\ \xi \neq \eta\} = P\{\eta \in B\},$$

то распределения случайных величин внутри одного класса совпадают. Итак, если случайные величины эквивалентны, то их законы распределения совпадают. Обратное, вообще говоря неверно, что показывает приводимый ниже пример. Разумеется, всюду нами предполагалось, что случайные величины определены на одном и том же вероятностном пространстве.

Пример 3. Рассмотрим вероятностное пространство $(\Omega,\ \mathcal{F},\ P)$, где $\Omega = [0,1]$, $F = \mathcal{B}([0,1])$ — σ-алгебра борелевских множеств отрезка $[0,1]$, P — мера Лебега. Оно соответствует бросанию наудачу точки на отрезок $[0,1]$. Пусть $\xi_1(\omega) \equiv \omega$, $\xi_2(\omega) \equiv 1 - \omega$,

$$\xi_3(\omega) \equiv \begin{cases} 2\omega, & \text{если } 0 \leq \omega \leq \frac{1}{2}, \\ 2(1-\omega), & \text{если } \frac{1}{2} < \omega \leq 1. \end{cases}$$

Очевидно, что для $0 \leq x \leq 1$

$$F_{\xi_1}(x) = P\{\omega < x\} = x,$$

$$F_{\xi_2}(x) = P\{1-\omega < x\} = P\{\omega > 1-x\} = x,$$

$$F_{\xi_3}(x) = P\{\omega:\ \xi_3(\omega) < x\} =$$

$$= P\{\omega:\ \xi_3(\omega) < x,\ 0 \le \omega \le \frac{1}{2}\} +$$

$$+ P\{\omega:\ \xi_3(\omega) < x,\ \frac{1}{2} < \omega \le 1\} = P\{2\omega < x,\ 0 \le \omega \le \frac{1}{2}\} +$$

$$+ P\{2(1-\omega) < x,\ \frac{1}{2} < \omega \le 1\} = P\{\omega < \frac{x}{2}\} + P\{\omega > 1 - \frac{x}{2}\} =$$

$$= \frac{x}{2} + \frac{x}{2} = x.$$

Очевидно, все три функции обращаются в ноль, если $x < 0$, и равны единице, если $x > 1$. Таким образом,

$$F_{\xi_1}(x) = F_{\xi_2}(x) = F_{\xi_3}(x) = \begin{cases} 0, & \text{если } x < 0, \\ x, & \text{если } 0 \le x \le 1, \\ 1, & \text{если } x > 1. \end{cases}$$

Итак все три величины имеют один и тот же равномерный закон распределения на отрезке $[0,1]$. Сами эти величины существенно отличаются. Например, при отображении ξ_1 все точки неподвижны, при отображении ξ_2 неподвижна единственная точка $\omega = 1/2$, при отображении ξ_3 неподвижны две точки $\omega = 0$ и $\omega = 2/3$. Таким образом, один и тот же закон распределения могут иметь существенно отличающиеся случайные величины. Часто, проявляя небрежность речи, говорят, что случайная величина определяется своим законом распределения. Как показывает приведенный пример, по закону распределения однозначно восстановить случайную величину невозможно. Поэтому правильнее говорить, что вероятностное поведение случайной величины определяется ее распределением, а не сама эта величина.

Очень часто приходится рассматривать различные функции от случайной величины. Когда эти функции сами являются случайными величинами?

Лемма о борелевской функции от случайной величины. *Пусть ξ — случайная величина, $f(x)$ — борелевская функция. Тогда $\eta = f(\xi)$ — случайная величина.*

Доказательство. В несколько иной форме эта лемма доказывается в курсах функционального анализа. Очевидно, $\eta(\omega) = f(\xi(\omega)) : \quad \Omega \to \mathbb{R}$ — сложная функция. Остается доказать ее измеримость относительно σ-алгебры $\mathcal{F}$. Имеем:

$$\eta^{-1}(B) = \{\omega : \ \eta(\omega) \in B\} = \{\omega : \ f(\xi(\omega)) \in B\} = \\ = \{\omega : \ \xi(\omega) \in f^{-1}(B)\} \in \mathcal{F}$$

для любого $B \in \mathcal{B}$. Здесь использовано то, что ξ измерима относительно $\mathcal{F}$ и что $f^{-1}(B)$ — борелевское множество, если B — борелевское множество. $\square$

Задача 3. Докажите, что если $\xi_1, \ \xi_2, \ \ldots, \ \xi_n$ — случайные величины, определенные на одном и том же вероятностном пространстве, $f(x_1, \ x_2, \ \ldots, \ x_n)$ — борелевская функция n переменных, то $\eta = f(\xi_1, \ \xi_2, \ \ldots, \ \xi_n)$ — случайная величина.

В заключение, введем еще одно важное понятие. Наименьшая σ-алгебра, относительно которой измерима случайная величина:

$$\mathcal{F}_\xi = \{\xi^{-1}(B), \ \ B \in \mathcal{B}\} = \xi^{-1}(\mathcal{B})$$

называется σ-алгеброй, **порожденной случайной величиной** ξ. Ее можно трактовать как совокупность всех наблюдаемых событий в эксперименте, связанном с измерением ξ (т.е. всех подмножеств, относительные частоты которых могут быть подсчитаны по результатам измерений ξ).

§ 10. Типы случайных величин

По типам законов распределения все случайные величины делятся на "чистые" (дискретные, абсолютно непрерывные и сингулярные) и "смешанные", законы распределения которых являются их линейными комбинациями ("смесью" чистых законов распределения). Поэтому изучение начнем с основных, чистых распределений.

I. Дискретные распределения.

Определение 10.1. Случайная величина $\xi = \xi(\omega)$, ее распределение P_ξ и функция распределения F_ξ называются **дискретными**, если мера P_ξ сосредоточена в конечном или счетном множестве точек X.

Последнее означает, что $P_\xi(X) = 1$ или, что эквивалентно, $P_\xi(\overline{X}) = 0$. Если в X имеются одноточечные подмножества $\{x\}$, для которых $P_\xi(\{x\}) = 0$, то, перебрасывая их из X в $\overline{X}$, добьемся, чтобы $P_\xi(X) = 1$ и $\forall x \in X$ выполнялось $P_\xi(\{x\}) > 0$. Называя для удобства значения из X — возможными значениями, а значения из $\overline{X}$ — невозможными значениями случайной величины ξ, мы можем сформулировать эквивалентное определение.

Определение 10.2. Случайная величина $\xi = \xi(\omega)$, ее распределение P_ξ и функция распределения F_ξ называются **дискретными**, если множество возможных значений случайной величины ξ конечно или счетно.

Обозначим через $x_1, x_2, \ldots, x_k$, если X — конечно, и через $x_1, x_2, \ldots$, если X — счетно, — множество возможных значений случайной величины ξ. Пусть $p_i = P\{\omega : \ \xi(\omega) = x_i\} = P\{\xi = x_i\}$ — вероятность принятия случайной величиной значения $x_i \quad (i = 1, 2, \ldots, k$ или $i = 1, 2, \ldots)$.

Определение 10.3. Соответствие между значениями случайной величины и вероятностями этих значений $x_i \leftrightarrow p_i$ называется **рядом распределения** дискретной случайной величины ξ.

Обычно ряд распределения задают в виде таблицы

ξ	x_1	x_2	$\ldots\ldots$	x_k
P	p_1	p_2	$\ldots\ldots$	p_k

,

если X конечно, или таблицы

ξ	x_1	x_2	x_3	$\ldots\ldots$
P	p_1	p_2	p_3	$\ldots\ldots$

,

если X — счетно.

Так как $\sum\limits_i p_i = \sum\limits_i P\{\omega : \ \xi(\omega) = x_i\} = P\{\sum\limits_i \{\omega : \xi(\omega) = x_i\}\} = P\{\omega : \ \xi(\omega) \in X\} = P_\xi(X) = 1$, то сумма вероятностей в нижней строке ряда распределения равна 1.

Утверждение 10.4. *Ряд распределения дискретной случайной величины является ее законом распределения.*

Д о к а з а т е л ь с т в о. Так как событие $\{\xi < x\}$ происходит тогда и только тогда, когда происходит хотя бы одно из событий $\{\xi = x_i\}$ для $x_i \in X$ с $x_i < x$, то $\{\xi < x\} = \bigcup\limits_{i:\, x_i < x} \{\xi = x_i\}$, а поскольку для различных индексов i события, стоящие под знаком объединения несовместимы, то $\{\xi < x\} = \sum\limits_{i:\, x_i < x} \{\xi = x_i\}$. В силу аддитивности вероятности,

$$F(x) = P\{\xi < x\} = \sum_{i:\, x_i < x} P\{\xi = x_i\} = \sum_{i:\, x_i < x} p_i. \qquad (10.1)$$

С помощью (10.1) $F(x)$ однозначно выражается через ряд распределения, т.е. ряд распределения является законом распределения. □

Заметим, что столь же легко через ряд распределения выражается распределение дискретной случайной величины:

$$P_\xi(B) = P\{\omega : \xi(\omega) \in B\} = P\Big(\sum_{i:\, x_i \in B} \{\omega : \xi(\omega) = x_i\}\Big) =$$

$$= \sum_{i:\, x_i \in B} p_i. \qquad (10.2)$$

Подчеркнем, что ряд распределения является не универсальным, а специальным законом распределения, который годится только для дискретных случайных величин. В то же время для таких величин он является наиболее естественным и удобным. Действительно, например, функция распределения дискретной случайной величины имеет по сравнению с рядом распределения более сложную структуру: из (10.1) видно, что она является ступенчатой функцией, постоянной между соседними значениями случайной величины, со скачками в точках x_i, равными p_i, и непрерывной слева.

Теорема о существовании дискретной случайной величины с заданным рядом распределения. *Пусть* $x_1, x_2, \dots$ *— произвольные попарно различные действительные числа,* $p_1, p_2, \dots$ *— произвольные положительные числа, дающие в сумме единицу. Тогда существует вероятностное пространство* $(\Omega,\ \mathcal{F},\ P)$ *и дискретная случайная величина* $\xi = \xi(\omega)$*, определенная на нем такая, что ее ряд распределения имеет вид*

ξ	x_1	x_2	x_3	......
P	p_1	p_2	p_3	......

.

Понятно, каким образом видоизменить формулировку теоремы для случая, когда множество возможных значений случайной величины конечно.

Доказательство. Можно свести доказательство к теореме о существовании случайной величины с заданной функцией распределения из девятого параграфа. По заданным в условии теоремы системам чисел построим функцию $F(x)$ с помощью формулы (10.1). Очевидно, что эта функция неубывающая, непрерывна слева. Независимо от того, существуют ли среди чисел $p_1, p_2, \dots$ минимальное и максимальное, $F(-\infty) = 0$ как остаток сходящегося ряда с положительными членами и $F(+\infty) = 1$ как предел частной суммы ряда, который по условию теоремы сходится к 1. Итак, для $F(x)$ выполнены все условия теоремы о существовании случайной величины с заданной функцией распределения. Поэтому существует вероятностное пространство $(\Omega,\ \mathcal{F},\ P)$ и случайная величина $\xi = \xi(\omega)$, определенная на нем такая, что ее функция распределения $F_\xi(x) = F(x)$. Нетрудно понять, что случайная величина с функцией распределения $F(x)$ дискретна и имеет искомый ряд распределения.

Так как теорема о существовании случайной величины с заданной функцией распределения в основном тексте не доказывалась, дадим непосредственное не опирающееся на нее доказательство. К тому же это полезно для уяснения того, как строится соответствующее вероятностное пространство. При доказательстве любых теорем о существовании случайной величины для построения, как правило, используется координатное пространство, что позволяет наиболее простым способом сконструировать требуемую случайную величину как тождественную (элемент координатного пространства).

Итак, в качестве элементарного исхода возьмем будущее возможное значение конструируемой случайной величины, т.е. значение x_i из заданного в условии теоремы набора чисел. Таким образом, $\Omega = \{x_1,\ x_2,\ \ldots,\}$. Так как это пространство счетное, то в качестве σ-алгебры событий $\mathcal{F}$ необходимо взять совокупность всех подмножеств Ω. Поскольку числа p_1, $p_2, \ldots$ по условию теоремы обладают всеми свойствами вероятностей элементарных исходов при конструктивно-аксиоматическом определении вероятности в дискретном пространстве элементарных исходов, то мы можем приписать каждому элементарному исходу x_i вероятность p_i и определить вероятность любого события (элемента $\mathcal{F}$) в соответствии с конструктивноаксиоматическим определением как

$$P(A) = \sum_{x_i \in A} p_i, \quad A \in \mathcal{F}.$$

Итак, вероятностное пространство построено. Определим случайную величину на нем как тождественную, т.е. $\xi(x_i) = x_i \quad \forall x_i \in \Omega$. Как мы помним, измеримость ее относительно $\mathcal{F}$ выполняется автоматически. Очевидно, она является дискретной случайной величиной с рядом распределения

ξ	x_1	x_2	x_3	$\ldots\ldots$
P	p_1	p_2	p_3	$\ldots\ldots$

.

$\square$

Замечание 1. Разумеется, координатное пространство при доказательстве берется только для его простоты. Действительно, можно было определить другую случайную величину $\eta(\omega)$, заданнную на вероятностном пространстве $(\Omega, \mathcal{F}, P)$, где $\Omega = [0,1], \quad F = \mathcal{B}([0,1])$ — σ-алгебра борелевских множеств отрезка $[0,1]$ (с таким же успехом можно было бы взять ее пополнение, т.е. σ-алгебру измеримых по Лебегу подмножеств отрезка $[0,1]$), а в качестве P — меру Лебега (иными словами η имеет равномерное распределение на отрезке $[0,1]$, например, является координатой наудачу брошенной на $[0,1]$ точки). Отрезок $[0,1]$ разбит на счетное число попарно непересекающихся измеримых частей (т.е. принадлежащих $\mathcal{F}$) $\Omega_1, \Omega_2, \ldots$ с мерами Лебега $p_1, p_2, \ldots$ соответственно. Тогда случайная величина $\eta = \eta(\omega)$, ставящая в соответствие каждому $\omega \in \Omega_i$ число $p_i, \quad i = 1, 2, \ldots,$ также является дискретной с тем же рядом распределения, что и координатная случайная величина, использованная при доказательстве.

Замечание 2. Доказанная теорема позволяет конструировать неограниченное число примеров дискретных случайных величин.

Часто встречающиеся дискретные распределения

1. *Вырожденное распределение.*

Определенная на любом вероятностном пространстве случайная величина $\xi(\omega) \equiv C$, называемая **постоянной** или **неслучайной величиной**, является дискретной и принимает единственное значение C с вероятностью 1 (проверьте ее измеримость). Ее функция распределения, очевидно, равна

$$F(x) = \begin{cases} 0, & \text{если } x \le C, \\ 1, & \text{если } x > C. \end{cases}$$

Распределение этой случайной величины называется **вырожденным**.

2. *Распределение Бернулли.*

Говорят, что случайная величина ξ имеет **распределение Бернулли**, если она дискретна с рядом распределения

ξ	1	0
P	p	q

,

где $p,\ q > 0, \quad p + q = 1$.

Такая случайная величина существует, ибо для нее выполняются условия последней теоремы. На самом деле, если в схеме Бернулли рассмотреть отдельное испытание, то ξ — число успехов в этом испытании.

3. *Дискретное равномерное распределение.*

Говорят, что случайная величина ξ имеет **дискретное равномерное распределение**, если она принимает конечное число возможных значений с одинаковой вероятностью, т.е. ряд ее распределения имеет вид

$$\begin{array}{c||c|c|c|c||} \xi & x_1 & x_2 & \ldots\ldots & x_k \\ \hline P & \frac{1}{k} & \frac{1}{k} & \ldots\ldots & \frac{1}{k} \end{array} \quad .$$

Она существует по последней теореме.

4. *Биномиальное распределение.*

Говорят, что случайная величина ξ имеет **биномиальное распределение** с параметрами n, p, где $n \in \mathbb{N}$, $0 < p < 1$, если она может принимать лишь значения $0, 1, \ldots, n$ с вероятностями

$$p_k = P\{\xi = k\} = C_n^k p^k (1-p)^{n-k}, \qquad k = 0, 1, \ldots, n.$$

Очевидно, все $p_k > 0$. По формуле бинома Ньютона

$$\sum_{k=0}^{n} p_k = \sum_{k=0}^{n} C_n^k p^k (1-p)^{n-k} = [p + (1-p)]^n = 1.$$

Значит, такая случайная величина существует. В этом можно убедиться и непосредственно, вспоминая, что, в силу формулы Бернулли (6.5), биномиальное распределение будет иметь число успехов μ в схеме n независимых испытаний Бернулли с вероятностью успеха p в отдельном испытании.

5. *Геометрическое распределение.*

Говорят, что случайная величина ξ имеет **геометрическое распределение**, если ее возможными значениями являются 0, 1, ..., причем

$$p_k = P\{\xi = k\} = q^k p, \qquad \text{где } q = 1 - p, \qquad 0 < p < 1,$$

$$k = 0,\, 1,\, \dots \, .$$

Так как

$$\sum_{k=0}^{\infty} p_k = p \sum_{k=0}^{\infty} q^k = \frac{p}{1-q} = 1,$$

то такая случайная величина существует.

Нетрудно понять, что геометрическое распределение будет иметь случайная величина ν, равная числу испытаний, проведенных до первого появления успеха, в бесконечной последовательности независимых испытаний Бернулли.

6. *Распределение Пуассона.*

Говорят, что случайная величина ξ имеет **распределение Пуассона с параметром** λ, где $\lambda > 0$, если ее возможными значениями являются 0, 1, ..., причем

$$p_k = P\{\xi = k\} = \frac{\lambda^k}{k!} e^{-\lambda}, \qquad k = 0,\, 1,\, \dots \, .$$

Поскольку

$$\sum_{k=0}^{\infty} p_k = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^{-\lambda} e^{\lambda} = 1,$$

то случайная величина с распределением Пуассона существует. В силу предельной теоремы Пуассона (теоремы

7.1) распределение Пуассона является предельным для биномиального распределения, когда $n \to \infty$, $p \to 0$ таким образом, что $\lambda = np$ остается постоянным.

7. *Гипергеометрическое распределение.*

Пусть N, M, n — натуральные числа, m — целое неотрицательное число, причем $n \leq N$, $m \leq M$, $M \leq N$, $m_0 = \max(0, M - N + n)$, $m_1 = \min(M, n)$.

Говорят, что случайная величина ξ имеет **гипергеометрическое распределение**, если ее возможными значениями являются $m_0, m_0 + 1, \ldots, m_1$, причем

$$p_m = P\{\xi = m\} = \frac{C_M^m C_{N-M}^{n-m}}{C_N^n}, \quad m = m_0, m_0 + 1, \ldots, m_1.$$

Как мы помним (см. пример 7 из третьего параграфа), случайной величине ξ можно придать следующий смысл. Пусть в партии из N деталей имеется M стандартных. Тогда p_m есть вероятность того, что среди n наудачу извлеченных из этой партии деталей имеется ровно m стандартных. Поэтому ξ — число стандартных деталей среди n извлеченных. Существование случайной величины ξ доказано. Поэтому должно выполняться

$$\sum_{m=m_0}^{m_1} p_m = \sum_{m=\max(0, M-N+n)}^{\min(M,n)} \frac{C_M^m C_{N-M}^{n-m}}{C_N^n} = 1. \tag{10.3}$$

Конечно, это комбинаторное тождество можно доказать аналитически, не пользуясь теорией вероятностей (например, можно воспользоваться полиномиальной формулой (3.10)). Но это увело бы нас в сторону от основного направления наших исследований.

Задача 1. Доказажите (10.3) аналитически.

Мы фактически доказали (10.3) с помощью вероятностного метода. Этот метод можно эффективно применять при получении и доказательстве самых разнообразных комбинаторных тождеств.

II. Абсолютно непрерывные распределения.

1. Элементарный подход.

Вспомним, как в механике определяется линейная плотность массы (или плотность распределения массы) конечного стержня. Предположим, что имеется стержень длины l и массы m, который мы будем представлять отрезком $[0, l]$ оси Ox. Обозначим через $m(x)$ часть массы стержня на участке $[0, x)$ (точку x можно включать или не включать, поскольку ее масса равна нулю). Масса участка стержня $[x,\, x + \Delta x)$ равна $M([x,\, x + \Delta x)) = m(x + \Delta x) - m(x)$. Средняя масса, приходящаяся на единицу длины на этом участке равна

$$\frac{M([x,\, x + \Delta x))}{\Delta x} = \frac{m(x + \Delta x) - m(x)}{\Delta x}.$$

Предел средней массы, приходящейся на единицу длины этого участка, когда участок стягивается в точку x,

$$\rho(x) = \lim_{\Delta x \to 0} \frac{M([x,\, x + \Delta x))}{\Delta x} =$$

$$= \lim_{\Delta x \to 0} \frac{m(x + \Delta x) - m(x)}{\Delta x} = m'(x),$$

в механике называется линейной плотностью массы стержня в точке x. Из этого определения немедленно следует известная формула, выражающая массу стержня через линеную плотность:

$$m = \int\limits_0^l \rho(x)dx.$$

Поступим аналогичным образом с вероятностью (с вероятностной массой). Единственное отличие состоит в том, что вероятностная масса распределена по "стержню бесконечной длины" — числовой прямой $\mathbb{R}$, и вся она равна 1, поскольку $P\{\omega : \xi(\omega) \in \mathbb{R}\} = 1$.

Предположим, что функция распределения случайной величины ξ имеет на $\mathbb{R}$ непрерывную производную $F'(x)$. Вероятность попадания случайной величины ξ в промежуток $[x, x + \Delta x)$ равна $P\{\omega : \ x \le \xi(\omega) < x + \Delta x\} = F(x + \Delta x) - F(x)$. Средняя вероятность попадания на единицу длины на участке $[x, x + \Delta x)$ есть

$$\frac{P\{x \le \xi < x + \Delta x\}}{\Delta x} = \frac{F(x + \Delta x) - F(x)}{\Delta x}.$$

В силу нашего предположения при каждом x существует конечный предел,

$$p(x) = \lim_{\Delta x \to 0} \frac{P\{x \le \xi < x + \Delta x\}}{\Delta x} =$$

$$= \lim_{\Delta x \to 0} \frac{F(x + \Delta x) - F(x)}{\Delta x} = F'(x),$$

который является непрерывной функцией. Предел этой средней вероятности $p(x)$ естественно назвать **плотностью вероятности**. Отметим, что при элементарном подходе из неубывания функции распределения F следует неотрицательность плотности вероятности.

Из определения плотности вероятности следует, что

$$P\{x \le \xi < x + \Delta x\} = p(x)\Delta x + o(\Delta x) \quad \text{при} \quad \Delta x \to 0.$$

Поэтому с точностью до бесконечно малой более высокого порядка, чем Δx, имеем

$$P\{x \le \xi < x + \Delta x\} \approx p(x)\Delta x.$$

Вместо приближенного равенства пишут точное, заменяя приращение Δx на дифференциал dx и понимая под $P\{x \le \xi < x + dx\}$ дифференциал функции $F(x)$:

$$P\{x \le \xi < x + dx\} = p(x)dx.$$

Произведение $p(x)dx$ называют **элементом вероятности**.

Из равенства $p(x) = F'(x)$ следует, что

$$F(x) = \int\limits_{-\infty}^{x} p(t)dt + C.$$

По свойствам функции распределения $F(-\infty) = 0$, откуда $C = 0$. Таким образом,

$$F(x) = \int\limits_{-\infty}^{x} p(t)dt. \tag{10.4}$$

Из (10.4) следует, что

$$P\{a \le \xi < b\} = F(b) - F(a) = \int\limits_{a}^{b} p(x)dx. \tag{10.5}$$

Для целей теории вероятностей достаточно такого определения плотности вероятности, чтобы с ее помощью для достаточно "хороших" множеств (а именно, борелевских) уметь находить вероятность попадания в них, подобно тому, как это может быть сделано для полуинтервалов с помощью соотношения (10.5). Естественно, нам надо постараться избавится от довольно жесткого ограничения, состоящего в непрерывной дифференцируемости F. Зададимся целью определить $p(x)$ как такую функцию, для которой выполняется (10.4). В силу критерия Лебега интегрируемости по Риману, функция $p(x)$ определяется равенством (10.4) неоднозначно. Если изменить значения $p(x)$ в точках множества лебеговой меры нуль, то значение интеграла (10.4) не изменится. Поэтому в качестве плотности вероятности может быть выбрана любая такая функция. Для удобства ее ограничивают единственным требованием — она должна быть неотрицательна. Если идти в этом направлении дальше, то можно существенно ослабить требование непрерывной дифференцируемости $F(x)$. Однако этого делать не нужно. Так как у нас уже появились в рассуждениях множества меры нуль, то естественно вместо интеграла Римана в определении плотности вероятности использовать интеграл Лебега. К тому же существенно расширяется класс интегрируемых функций.

2. Подход, основанный на неопределенном интеграле Лебега.

Напомним, что функция f, определенная на отрезке $[a, b]$, называется абсолютно непрерывной на этом отрезке, если для любого $\varepsilon > 0$ найдется такое $\delta > 0$, что для всякой конечной системы попарно непересекающихся интервалов $(a_k, b_k) \subset [a, b]$, $k = 1, 2, \ldots, n$, с суммой длин, меньшей $\delta > 0$,

$$\sum_{k=1}^{n}(b_k - a_k) < \delta,$$

выполняется неравенство

$$\sum_{k=1}^{n}|f(b_k) - f(a_k)| < \varepsilon.$$

Известно, что всякая абсолютно непрерывная на отрезке функция равномерно непрерывна и имеет ограниченное изменение. Если функция интегрируема на отрезке $[a, b]$ по Лебегу, то мы будем обозначать интеграл Лебега от нее по отрезку $[a, x]$ через

$$F(x) = \int_{a}^{x} f(t)dt, \qquad x \in [a, b]. \tag{10.6}$$

Известно, что функция F абсолютно непрерывна на отрезке $[a, b]$ тогда и только тогда, когда для любого $x \in [a, b]$ она представляется в виде интеграла с переменным верхним пределом (10.6) от некоторой функции f. При этом почти всюду (относительно меры Лебега) производная от интеграла с переменным верхним пределом (10.6) совпадает с $f(x)$. Отметим, что если задана функция F, то из уравнения (10.6) функция f находится неоднозначно, с точностью до значений на множестве лебеговой меры нуль. Пользуясь этими наводящими соображениями, введем следующее

Определение 10.5. Случайная величина $\xi = \xi(\omega)$, ее распределение P_ξ и функция распределения F_ξ называются **абсолютно непрерывными**, если существует функция $p_\xi : \mathbb{R} \to [0, \infty)$ такая, что функция распределения случайной величины ξ представляется для любого $x \in \mathbb{R}$ в виде

$$F_\xi(x) = \int\limits_{-\infty}^{x} p_\xi(t) dt. \tag{10.7}$$

Любая неотрицательная функция p_ξ, удовлетворяющая (10.7), называется **плотностью вероятности** или **плотностью распределения вероятностей** абсолютно непрерывной случайной величины ξ.

Как и в случае функции распределения мы иногда будем писать у плотности вероятности индекс ξ, а иногда не будем, так, что $p_\xi(x) = p(x)$.

Покажем, что из (10.7) следует, что для любого борелевского множества B

$$P_\xi(B) = P\{\omega : \xi(\omega) \in B\} = P\{\xi \in B\} = \int\limits_{B} p_\xi(x) dx. \tag{10.8}$$

Левая часть (10.8), как ранее мы видели, является мерой на σ-алгебре борелевских множеств $\mathcal{B}$. Из теории меры Вы знаете, что интеграл Лебега от неотрицательной функции $p_\xi(x)$ является мерой множества интегрирования на σ-алгебре измеримых по Лебегу множеств, тем более, этот интеграл будет мерой на входящей в нее σ-алгебре борелевских множеств $\mathcal{B}$. Итак, для доказательства (10.8), надо доказать совпадение этих двух мер на $\mathcal{B}$. В силу (10.7) эти меры совпадают на борелевских множествах вида $B = (-\infty,\, x)$. Так как для любой меры μ выполняется $\mu([a,\, b)) = \mu((-\infty,\, b)) - \mu((-\infty,\, a))$, то эти меры совпадают на полуинтервалах $[a,\, b)$. В силу конечной аддитивности эти меры совпадают на алгебре объединений полуинтервалов $\mathcal{A}$. По теореме о продолжении меры эти меры совпадут и на наименьшей σ-алгебре, содержащей алгебру $\mathcal{A}$, т.е. на σ-алгебре борелевских множеств $\mathcal{B}$.

Обратно, из (10.8) следует (10.7), если в качестве B взять $(-\infty,\, x)$. Поэтому определение 10.5 эквивалентно следующему

Определение 10.6. Случайная величина $\xi = \xi(\omega)$, ее распределение P_ξ и функция распределения F_ξ называются **абсолютно непрерывными**, если существует функция $p_\xi :\ \mathbb{R} \to [0, \infty)$ такая, что распределение случайной величины ξ представляется для любого $B \in \mathcal{B}$ в виде

$$P_\xi(B) = \int\limits_B p_\xi(x) dx. \tag{10.9}$$

Любая неотрицательная функция p_ξ, удовлетворяющая (10.9), называется **плотностью вероятности** или **плотностью распределения вероятностей** абсолютно непрерывной случайной величины ξ.

3. Подход, основанный на теореме Радона-Никодима.

Пусть μ и ν — две меры, определенные на одной и той же σ- алгебре $\mathcal{F}$. Напомним, что мера μ называется абсолютно непрерывной относительно меры ν (обозначение $\mu << \nu$), если из $\nu(A) = 0$ следует $\mu(A) = 0$. Теорема Радона-Никодима утверждает, как мы помним, что

$\mu << \nu$ тогда и только тогда, когда существует неотрицательная суммируемая по мере ν функция p такая, что для любого $A \in \mathcal{F}$ выполняется

$$\mu(A) = \int\limits_A p\, d\nu.$$

Функция p определяется этим равенством неоднозначно с точностью до значений на множестве нулевой ν-меры, обозначается $p = \frac{d\mu}{d\nu}$ и называется производной или плотностью Радона-Никодима меры μ по мере ν.

Поэтому определение 10.6 можно перефразировать следующим образом. Случайная величина $\xi = \xi(\omega)$, ее распределение P_ξ и функция распределения F_ξ называются **абсолютно непрерывными**, если распределение P_ξ абсолютно непрерывно относительно меры Лебега на измеримом пространстве $(\mathbb{R}, \mathcal{B})$. **Плотностью вероятности** или **плотностью распределения вероятностей** называется производная Радона-Никодима $p_\xi(x) = \frac{dP_\xi}{dx}$ меры P_ξ по мере Лебега x.

Свойства плотности вероятности.

1. Плотность вероятности определена неоднозначно, с точностью до значений на множестве лебеговой меры ноль.

Это следует из теоремы Радона-Никодима.

2. Так как плотность можно выбрать в классе эквивалентных между собой функций, договоримся выбирать ее всегда таким образом, что если в данной точке существует ее непрерывная эквивалентная модификация, то

мы будем брать именно ее. При этом условии, если x — точка непрерывности $p(x)$, то $p(x) = F'(x)$.

В частности, если существует непрерывная на всей числовой прямой модификация плотности вероятности, то именно ее выбираем в качестве этой плотности. При этом для всех $x \in \mathbb{R}$ выполняется $F'(x) = p(x)$ (сравните с элементарным подходом к введению плотности вероятности).

3. Почти всюду существует производная функции распределения, причем почти всюду

$$p(x) = F'(x).$$

Это следует из теории интеграла Лебега с переменным верхним пределом.

4. $p(x) \geq 0 \quad \forall x \in \mathbb{R}$.

Так как $F(x)$ не убывает, то из предыдущего свойства неотрицательность $p(x)$ выполняется почти всюду. По нашей договоренности в точках множества меры ноль, в которых $p(x)$ можно задать произвольно, мы доопределяем ее неотрицательными значениями.

5.

$$\int\limits_{-\infty}^{\infty} p(x)dx = 1.$$

Это свойство следует из (10.7), если x устремить к бесконечности и использовать свойство $F(+\infty) = 1$, или из (10.8), если взять $B = \mathbb{R}$.

Свойства 4 и 5 являются аналогом для абсолютно непрерывного случая свойств $p_i \geq 0$ и $\sum\limits_i p_i = 1$ ряда распределения дискретной случайной величины. Фигурально выражаясь, можно сказать, что вся единичная вероятностная масса, сосредоточенная для дискретного случая в конечном или счетном множестве точек "размазывается" с помощью плотности распределения по всей числовой прямой (или ее части).

В силу (10.7) и (10.8) плотность вероятности однозначно определяет функцию распределения и распределение, т.е. является законом распределения. Подчеркнем, что этот закон не является универсальным, а годится только для абсолютно непрерывных величин.

Теорема о существовании абсолютно непрерывной случайной величины с заданной плотностью распределения вероятностей.

Пусть задана функция $p: \mathbb{R} \to \mathbb{R}$, *удовлетворяющая свойствам*

1. $p(x) \geq 0 \ \ \forall x \in \mathbb{R}$;
2. $\int\limits_{-\infty}^{+\infty} p(x)dx = 1$.

Тогда существует вероятностное пространство $(\Omega, \mathcal{F}, P)$ *и абсолютно непрерывная случайная величина* $\xi = \xi(\omega)$, *определенная на нем, такая, что ее плотность распределения есть заданная функция*: $p_\xi(x) = p(x)$.

Д о к а з а т е л ь с т в о. Как и в дискретном случае, можно свести доказательство к теореме о существовании случайной величины с заданной функцией распределения из девятого параграфа. По заданной в условии теоремы функции $p(x)$ построим функцию $F(x)$ с помощью фор-

мулы
$$F(x) = \int\limits_{-\infty}^{x} p(x)dx. \tag{10.10}$$

Очевидно, что эта функция неубывающая, непрерывная. Так как остаток сходящегося интеграла (10.10) стремится на $-\infty$ к нулю, то $F(-\infty) = 0$. Устремляя x к $+\infty$ и используя условие 2 теоремы, получаем, что $F(+\infty) = 1$. Поскольку для F выполнены все условия теоремы о существовании случайной величины с заданной функцией распределения, то существует вероятностное пространство $(\Omega, \mathcal{F}, P)$ и случайная величина $\xi = \xi(\omega)$, определенная на нем, с функцией распределения (10.10). По определению 10.5 эта случайная величина абсолютно непрерывна с плотностью вероятности $p_\xi(x) = p(x)$.

Так как теорема о существовании случайной величины с заданной функцией распределения в основном тексте не доказывалась, дадим независимое от нее доказательство. Построим вероятностное пространство $(\Omega, \mathcal{F}, P)$, где $\Omega = \mathbb{R},\ \mathcal{F} = \mathcal{B}$, а
$$P(B) = \int\limits_{B} p(x)dx. \tag{10.11}$$

Так как $p(x) \geq 0$, то $\forall B \in \mathcal{B}$ выполняется $P(B) \geq 0$. В силу условия 2 теоремы $P(\mathbb{R}) = 1$. Счетная аддитивность меры P следует из того, что интеграл Лебега (10.11) является счетно аддитивной функцией от множества интегрирования из σ-алгебры измеримых по Лебегу множеств, а, следовательно, от множества интегрирования из ее под-σ-алгебры $\mathcal{B}$. Итак P — вероятность. Остается ввести случайную величину $\xi(\omega) = \omega$ (являющуюся координатной).

Тогда, в силу (10.11),

$$P_\xi(B) = P\{\omega: \xi(\omega) \in B\} = P\{\omega: \omega \in B\} = P(B) =$$

$$= \int_B p(x)dx.$$

По определению 10.6 случайная величина ξ абсолютно непрерывна и имеет плотность распределения $p_\xi(x) = p(x)$. □

Замечание 3. Интеграл в равенстве (10.7) понимается как интеграл Лебега. Когда этот интеграл существует как несобственный интеграл Римана, то, в силу неотрицательности подынтегральной функции, он совпадет с рассматриваемым интегралом Лебега. Если подынтегральная функция не являлась бы неотрицательной, то из сходимости несобственного интеграла Римана не следовало бы существование этого интеграла как интеграла Лебега. Сама конструкция интеграла Лебега такова, что он существует тогда и только тогда, когда существует интеграл от модуля рассматриваемой функции. В общем случае, если в (10.7) подынтегральная функция принимает как положительные, так и отрицательные значения, то *из абсолютной сходимости интеграла (10.7) в смысле Римана следует его существование в смысле Лебега. При этом интегралы Римана и Лебега совпадают.* В дальнейшем, как правило, будет встречаться именно такой случай. Поэтому в практических ситуациях возможность замены интеграла Лебега соответствующим несобственным интегралом Римана законна, если этот интеграл абсолютно сходится.

Замечание 4. Иногда, особенно в технической литературе, дискретную случайную величину также задают с помощью "плотности вероятности". Если величина задана рядом распределения

ξ	x_1	x_2	x_3	......
P	p_1	p_2	p_3	......

,

то ее плотность записывают в виде

$$p(x) = \sum_{i=1}^{\infty} p_i \delta(x - x_i),$$

где $\delta(x)$ — так называемая дельта-функция Дирака, т.е. такая обобщенная функция, что $\delta(x) \geq 0$ и $\int\limits_{-\infty}^{+\infty} \delta(x)dx = 1$. Из теории обобщенных функций (распределений Шварца) известно, что

$$\int\limits_B f(x)\delta(x - x_0)dx = f(x_0)I_B(x_0),$$

где $I_B(x) = 1$, если $x \in B$, и $I_B(x) = 0$ в противоположном случае. Применение такого подхода является определенным формализмом, который удобен части инженерно-технических работников для того, чтобы не запоминать формулы типа (10.1),(10.2), а всегда использовать вместо них формулы (10.7), (10.8). При этом указанные работники, как правило, пользуются дельта-функцией формально, представляя ее как некий импульс, не понимая при этом, что это не есть функция в обычном смысле, а интеграл не является интегралом в общепринятом понимании. Зато они хорошо владеют формализмом преобразований интегралов, которые содержат дельта-функции, умеют производить с ними простейшие операции типа дифференцирования. Так, например, при подсчете вероятности попадания значений дискретной случайной величины в некоторое борелевское множество они поступают приблизительно таким образом:

$$P\{\xi \in B\} = \int\limits_B p(x)dx = \int\limits_B \sum_{i=1}^{\infty} p_i \delta(x - x_i)dx =$$

$$= \sum_{i=1}^{\infty} p_i \int\limits_B \delta(x - x_i)dx = \sum_{i=1}^{\infty} p_i I_B(x_i) = \sum_{i:\, x_i \in B} p_i.$$

Профессиональные математики, как правило, такой подход не используют.

Часто встречающиеся абсолютно непрерывные распределения

1. *Равномерное распределение на отрезке.*

Это распределение уже встречалось, когда рассматривалась координата ξ точки, наудачу брошенной на отрезок $[a, b]$. Функция ее распределения равнялась

$$F(x) = \begin{cases} 0, & \text{если } x \le a, \\ \frac{x-a}{b-a}, & \text{если } a < x \le b, \\ 1, & \text{если } x > b. \end{cases}$$

Предположим, что рассматриваемая функция абсолютно непрерывна. По свойству плотности вероятности почти всюду

$$p(x) = F'(x) = \begin{cases} \frac{1}{b-a}, & \text{если } a < x < b, \\ 0, & \text{если } x \notin [a, b]. \end{cases}$$

Так как множество, состоящее из двух точек a и b, в которых производная от F не определена, имеет нулевую меру Лебега, то мы можем определить плотность распределения в этих точках произвольно, лишь бы она оставалась неотрицательной. Для определенности будем считать, что

$$p(x) = \begin{cases} \frac{1}{b-a}, & \text{если } a \leq x \leq b, \\ 0, & \text{если } x \notin [a,b]. \end{cases} \tag{10.12}$$

Для доказательства абсолютной непрерывности ξ и $F(x)$ остается, например, проверить выполнение определения 10.5, т.е. установить, что рассмотренные $F(x)$ и $p(x)$ удовлетворяют при каждом $x \in R$ соотношению

$$F(x) = \int\limits_{-\infty}^{x} p_(t)dt.$$

Задача 2. Проверьте, что это так.

Поэтому эквивалентное определение выглядит так: случайная величина **равномерно распределена на отрезке** $[a,b]$, если она абсолютно непрерывна и имеет плотность распределения вида (10.12).

Проверять существование такой случайной величины с помощью теоремы о существовании абсолютно непрерывной случайной величины с заданной плотностью распределения нет необходимости, так как такой величиной является координата наудачу брошенной на отрезок $[a,b]$ точки. Однако, если бы мы исходили только из последнего определения равномерного распределения, то должны были бы проверить, что выполняются условия 1 и 2 этой теоремы, что, впрочем, очевидно.

2. *Показательное распределение.*

Говорят, что случайная величина ξ имеет **показательное** или **экспоненциальное распределение** с параметром λ, где $\lambda > 0$, если она абсолютно непрерывна и имеет плотность распределения вероятностей вида

$$p(x) = \begin{cases} \lambda e^{-\lambda x}, & \text{если } x \geq 0, \\ 0, & \text{если } x < 0. \end{cases}$$

Такая случайная величина действительно существует, так как

$$p(x) \geq 0, \qquad \int\limits_{-\infty}^{+\infty} p(x)dx = \int\limits_{0}^{+\infty} \lambda e^{-\lambda x} dx = 1.$$

Используя (10.7), находим соответствующую функцию распределения:

$$F(x) = \begin{cases} 1 - e^{-\lambda x}, & \text{если } x > 0, \\ 0, & \text{если } x \leq 0. \end{cases}$$

Отметим, что так как $F(0) = P\{\xi < 0\} = 0$, то показательное распределение испльзуется для описания вероятностного поведения неотрицательных (по смыслу) случайных величин. Например, часто таким распределением характеризуют длительность некоторого случайного процесса.

Лемма об отсутствии "памяти" (последействия) у показательного распределения.

Если случайная величина ξ имеет показательное распределение, то для любого $\tau \geq 0$

$$P\{\xi < t + \tau / \xi \geq \tau\} = P\{\xi < t\}.$$

Доказательство. По определению условной вероятности для $t, \tau > 0$

$$P\{\xi < t+\tau/\xi \geq \tau\} = \frac{P\{\tau \leq \xi < t+\tau\}}{P\{\xi \geq \tau\}} = \frac{F(t+\tau) - F(\tau)}{1 - F(\tau)} =$$

$$= \frac{e^{-\lambda\tau} - e^{-\lambda(t+\tau)}}{e^{-\lambda\tau}} = 1 - e^{-\lambda t} = F(t) = P\{\xi < t\}.$$

Если же $t \leq 0$, то обе части доказываемого равенства также равны, ибо обращаются в ноль. □

Рассмотрим работу некоторого элемента, например, электролампочки. Предположим, что время "жизни" ξ этого элемента имеет показательное распределение. Пусть в момент τ (время отсчитывается с момента начала работы элемента) элемент еще не вышел из строя, т.е. произошло событие $\{\xi \geq \tau\}$. Тогда его остаточное время "жизни" $\xi_\tau = \xi - \tau$ имеет функцию распределения

$$F_{\xi_\tau}(t) = P\{\xi_\tau < t\} = P\{\xi - \tau < t/\xi \geq \tau\} =$$

$$= P\{\xi < t\} = F_\xi(t).$$

Таким образом, остаточное время имеет тот же закон распределения, что и полное время "жизни" элемента. Другими словами, элемент как бы "забывает", сколько времени он уже существовал, и его остаточное время управляется точно таким же вероятностным механизмом, которым управляется полное время его "жизни". Можно доказать, что свойство отсутствия памяти является характеристическим для показательного распределения: других случайных величин, кроме показательно распределенных, которые удовлетворяли бы свойству отсутствия памяти, не существует.

Благодаря этому свойству показательное распределение описывает времена пребывания в состояниях так называемых марковских процессов с не более чем счетным множеством состояний. Поэтому оно находит широкое применение в физике (например, при описании радиактивных явлений), биологии и химии (например, при изучении биологических популяций), теории надежности, теории массового обслуживания, актуарной науке, технике.

3. *Распределение Коши.*

Предположим, что артиллерийское орудие расположено на единичном расстоянии от бесконечно длинной стены (см. Рис.3).

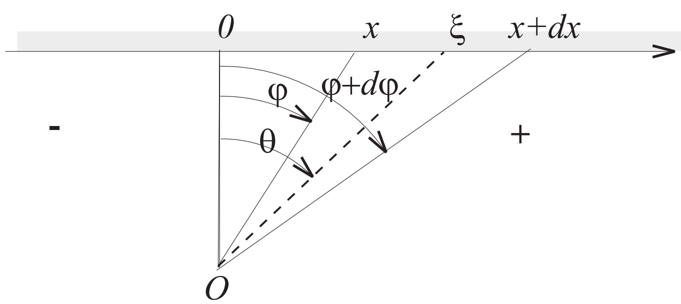


Рис.3

Лафет орудия вращается с постоянной угловой скоростью от положния, когда снаряд идет параллельно стене влево, до положения, когда он идет параллельно стене вправо, и между ними все время попадает в стену. Затем лафет вращается в противоположном направлении, затем все повторяется и т.д. На каждом полуобороте в наудачу выбранный момент из орудия производится выстрел. Пусть ξ — расстояние от основания перпендикуляра, опущенного из орудия на стену и взятого со знаком "+", если снаряд отклонился вправо, и со знаком "−", если снаряд отклонился влево. Пусть случайная величина θ — угол, на который повернулся ствол орудия относительно перпендикуляра, отсчитываемый по часовой стрелке. Очевидно, $-\pi/2 \leq \theta \leq +\pi/2$ и $\xi = tg\,\theta$. Так как скорость вращения лафета постоянна, а момент выстрела выбирается наудачу, то случайная величина θ равномерно распределена на отрезке $[-\pi/2,\, +\pi/2]$, следовательно ее плотность вероятности равна

$$p_\theta(\varphi) = \begin{cases} \frac{1}{\pi}, & \text{если } -\pi/2 < \varphi < +\pi/2, \\ 0, & \text{если } x \notin (-\pi/2,\, +\pi/2). \end{cases}$$

Пусть попаданию угла θ в диапазон $(\varphi,\, \varphi + d\varphi)$ соответствует попадание ξ в диапазон $(x,\, x + dx)$. Очевидно, $x = tg\,\varphi$, $\varphi = arctg\,x$, $d\varphi = \frac{dx}{1+x^2}$. Так как $P\{\varphi < \theta < \varphi + d\varphi\} = P\{x < \xi < x + dx\}$, то

$$p_\theta(\varphi)d\varphi = p_\xi(x)dx,$$

откуда

$$p_\xi(x) = \frac{1}{\pi(1+x^2)}.$$

Это распределение называется **распределением Коши**, а случайная величина с таким распределением — **распределенной по закону Коши**.

Рассмотренные рассуждения, основанные на понятии элемента вероятности, можно сделать строгими, если использовать следующее утверждение.

Теорема о плотности вероятности функции от случайной величины.

Пусть (a,b) *— конечный или бесконечный интервал,* $f:(a,b)\to\mathbb{R}$ *— строго монотонная дифференцируемая функция,* ξ *— абсолютно непрерывная случайная величина, распределение которой сосредоточено на* (a,b) *(т.е.* $P\{\xi\in(a,b)\}=1$*). Тогда случайная величина* $\eta=f(\xi)$ *абсолютно непрерывна и имеет плотность распределения вероятности*

$$p_\eta(x)=p_\xi(f^{-1}(x))|[f^{-1}(x)]'|. \tag{10.12}$$

Прежде, чем доказывать теорему, приведем пример, иллюстрирующий стандартный способ рассуждений. Пусть случайная величина ξ абсолютно непрерывна, $\eta=\xi^2$. Так как $f(x)=x^2$ — непрерывная функция, т.е. борелевская, то η — случайная величина. Выразим функцию распределения η через функцию распределения ξ. Если $x>0$, то

$$F_\eta(x)=P\{\eta<x\}=P\{\xi^2<x\}=P\{-\sqrt{x}<\xi<\sqrt{x}\}=$$

$$=F_\xi(\sqrt{x})-F_\xi(-\sqrt{x}).$$

Если $x\le 0$, то, очевидно, $F_\eta(x)=0$. Предположим, что η также абсолютно непрерывна. Дифференцируя полученные соотношения, получим:

$$p_\eta(x)=\begin{cases}\frac{1}{2\sqrt{x}}[p_\xi(\sqrt{x})+p_\xi(-\sqrt{x})], & \text{если } x>0,\\ 0, & \text{если } x\le 0.\end{cases}$$

Разумеется, это рассуждение не является исчерпывающим, поскольку абсолютная непрерывность η не доказана.

Задача 3. Докажите, что случайная величина $\eta = \xi^2$ абсолютно непрерывна, если ξ абсолютно непрерывна.

Доказательство. Так как $f(x)$ дифференцируема, то она непрерывна, т.е. является борелевской функцией. По лемме о борелевской функции от случайной величины из девятого параграфа $\eta = f(\xi)$ — случайная величина.

1-й способ доказательства.

1. Пусть функция f возрастает. Поскольку в этом случае событие $\{f(\xi) < x\}$ происходит тогда и только тогда, когда происходит событие $\{\xi < f^{-1}(x)\}$, то

$$F_\eta(x) = P\{\eta < x\} = P\{f(\xi) < x\} =$$

$$= P\{\xi < f^{-1}(x)\} = F_\xi(f^{-1}(x)),$$

и дифференцируя полученное равенство, находим:

$$p_\eta(x) = p_\xi(f^{-1}(x))[f^{-1}(x)]'.$$

Так как f возрастает, то f^{-1} также возрастает, следовательно, $[f^{-1}(x)]' > 0$, т.е. $|[f^{-1}(x)]'| = [f^{-1}(x)]'$, и (10.12) доказано.

2. Пусть функция f убывает. Поскольку в этом случае событие $\{f(\xi) < x\}$ происходит тогда и только тогда, когда происходит событие $\{\xi > f^{-1}(x)\}$, то

$$F_\eta(x) = P\{\eta < x\} = P\{f(\xi) < x\} = P\{\xi > f^{-1}(x)\} =$$

$$= P\{\xi > f^{-1}(x)\} + P\{\xi = f^{-1}(x)\} = P\{\xi \ge f^{-1}(x)\} =$$
$$= 1 - P\{\xi < f^{-1}(x)\} = 1 - F_\xi(f^{-1}(x)),$$

откуда

$$p_\eta(x) = -p_\xi(f^{-1}(x))[f^{-1}(x)]'.$$

Так как f убывает, то $[f^{-1}(x)]' < 0$, т.е. $|[f^{-1}(x)]'| = -[f^{-1}(x)]'$, и (10.12) доказано.

На самом деле (10.12) выполнено почти всюду. Но можно записать точное равенство, выбрав именно такой вариант плотности вероятности из класса эквивалентных между собой плотностей. Остается доказать абсолютную непрерывность η. Так как $p_\eta(x) \ge 0$ и

$$\int\limits_{-\infty}^{+\infty} p_\eta(x)dx = \int\limits_{-\infty}^{+\infty} p_\xi(f^{-1}(x))|[f^{-1}(x)]'|dx =$$

$$= \int\limits_a^b p_\xi(y)dy = \int\limits_{-\infty}^{+\infty} p_\xi(y)dy = 1,$$

то существует абсолютно непрерывная случайная величина с плотностью $p_\eta(x)$, которой соответствует функция распределения $F_\eta(x)$. Это и означает абсолютную непрерывность η.

2-й способ доказательства.

Здесь удобно рассмотреть **неориентированный** интеграл (интеграл Лебега), в котором производится замена $y = f(x)$, $x = f^{-1}(y)$ (в силу монотонности f с помощью нее устанавливается взаимно-однозначное соответствие между (a, b) и $(\min\{f(a), f(b)\}, \max\{f(a), f(b)\})$):

$$P\{\eta \in B\} = P\{f(\xi) \in B\} = P\{\xi \in f^{-1}(B)\} =$$

$$= \int\limits_{f^{-1}(B)} p_\xi(x)dx.$$

Производя в интеграле Лебега замену переменной $y = f(x)$, $x = f^{-1}(y)$, получим:

$$P\{\eta \in B\} = \int\limits_B p_\xi(y)|[f^{-1}(y)]'|dy.$$

В силу определения 10.6, это означает, что случайная величина η абсолютно непрерывна и имеет искомую плотность. □

Замечание 5. Если интеграл существует в смысле Римана, то во 2-м способе доказательства используется замена переменной в неориентированном интеграле Римана. Обычно, в силу сложившейся традиции, одномерные интегралы Римана рассматриваются как ориентированные, в то время как кратные интегралы Римана являются неориентированными. Поэтому теорема о замене переменной выглядит, в некотором смысле, проще для кратных интегралов, чем для обычных. Здесь как раз применяется ее модификация для одномерного случая.

Теперь понятно, как решить задачу 3.

Возвращаясь к распределению Коши, вспомним, что $\xi = tg\,\theta$. Так как функция $y = tg\,x$ возрастает и дифференцируема на $(-\pi/2,\ \pi/2)$, обратная функция $x = f^{-1}(y) = arctg\,y$, то по доказанной теореме

$$p_\xi(x) = p_\theta(arctg\,x)|[arctg\,x]'| = \frac{1}{\pi(1+x^2)}.$$

4. *Нормальное распределение.*

Говорят, что случайная величина ξ имеет **нормальное распределение с параметрами** a, σ^2 ($a \in R$, $\sigma > 0$), если она абсолютно непрерывна и имеет плотность вероятности вида

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-a)^2}{2\sigma^2}}.$$

Такая случайная величина существует, так как $p(x) > 0 \ \ \forall x \in \mathbb{R}$ и

$$\int\limits_{-\infty}^{+\infty} p(x)dx = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{+\infty} e^{-\frac{(x-a)^2}{2\sigma^2}} dx = \frac{1}{\sqrt{\pi}} \int\limits_{-\infty}^{+\infty} e^{-t^2} dt = 1.$$

Здесь мы сделали замену переменной $\frac{x-a}{\sigma\sqrt{2}} = t$ и воспользовались известным из математического анализа интегралом Пуассона

$$\int\limits_{-\infty}^{+\infty} e^{-t^2} dt = \sqrt{\pi}.$$

Для краткости записи фразу "ξ имеет нормальное распределение с параметрами a, σ^2" будем писать как $\xi \sim N(a, \sigma^2)$.

С помощью стандартных методов математического анализа легко проверяется (проверьте), что

1) $p(x)$ симметрична относительно прямой $x = a$;

2) в точке $x = a$ она достигает максимума, который равен $\frac{1}{\sqrt{2\pi}\sigma}$;

3) в промежутке $(-\infty, a]$ она возрастает, а в промежутке $[a, +\infty)$ убывает;

4) точки $a-\sigma$, $a+\sigma$ являются точками перегиба; при этом в прмежутках $(-\infty,\, a-\sigma]$, $[a+\sigma,\, +\infty)$ она выпукла вниз, а в промежутке $[a-\sigma,\, a+\sigma]$ выпукла вверх;

5) ось абсцисс является горизонтальной асимптотой при $x \to \infty$ ($\lim\limits_{x\to+\infty} p(x) = \lim\limits_{x\to-\infty} p(x) = 0$);

6) площадь, заключенная между графиком и осью абсцисс, равна 1.

Определение 10.7. Говорят, что случайная величина ξ имеет **стандартное нормальное распределение**, если $\xi \sim N(0,1)$.

Плотность стандартного нормального распределения (плотность вероятности случайной величины ξ_0, имеющей стандартное нормальное распределение) обозначают через

$$\varphi(x) = p_{\xi_0}(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}.$$

Эта функция уже встречалась нам в формулировке локальной предельной теоремы Муавра-Лапласа. Функцию стандартного нормального распределения (функцию распределения случайной величины ξ_0, имеющей стандартное нормальное распределение) будем обозначать через

$$N(x) = P\{\xi_0 < x\} = \int\limits_{-\infty}^{x} \varphi(t)dt.$$

Эта функция и связанная с ней функция Лапласа

$$\varPhi(x) = P\{0 < \xi_0 < x\} = \int\limits_{0}^{x} \varphi(t)dt$$

встречались нам ранее в формулировке интегральной предельной теоремы Муавра-Лапласа. Теоремы Муавра-Лапласа фактически утверждали, что нормальное распределение является предельным для биномиального распределения. Таким образом, в зависимости от способа стремления n к бесконечности предельным для биномиального распределения может служить как нормальное распределение, так и распределение Пуассона.

Так как $\varphi(-x) = \varphi(x)$, то

$$\Phi(-x) = \int_0^{-x} \varphi(t)dt = \int_0^{-x} \varphi(-t)dt = -\int_0^{x} \varphi(t)dt = -\Phi(x).$$

Таким образом, функция $\varphi(x)$ — четная, а функция $\Phi(x)$ — нечетная. Далее, для $x \geq 0$

$$N(x) = \int_{-\infty}^{x} \varphi(t)dt = \int_{-\infty}^{0} \varphi(t)dt + \int_{0}^{x} \varphi(t)dt =$$

$$= 1/2 \int_{-\infty}^{+\infty} \varphi(t)dt + \Phi(x) = 1/2 + \Phi(x),$$

а для $x < 0$

$$N(x) = \int_{-\infty}^{x} \varphi(t)dt = \int_{-\infty}^{0} \varphi(t)dt - \int_{x}^{0} \varphi(t)dt =$$

$$= 1/2 + \int_{0}^{x} \varphi(t)dt = 1/2 + \Phi(x).$$

Итак, для любого x

$$N(x) = 1/2 + \Phi(x).$$

Имеем

$$N(x) + N(-x) = 1/2 + \Phi(x) + 1/2 + \Phi(-x) = 1.$$

Имеются таблицы функций $\varphi(x)$, $\Phi(t)$, $N(x)$. Часть авторов пользуется таблицами функции распределения $N(x)$, часть авторов — таблицами функции Лапласа $\Phi(x)$. При этом, как правило, и ту и другую функцию обозначают одинаково как $\Phi(x)$. Поэтому при использовании тех или иных источников таблиц надо быть внимательным и смотреть не на обозначение, а на пределы интегрирования.

Лемма о нормальном распределении.

Если $\xi \sim N(a, \sigma^2)$, $\eta = A\xi+B$, *где* $A \neq 0$, *— линейная функция от* ξ, *то* $\eta \sim N(Aa + B, A^2\sigma^2)$.

В частности, если $\xi \sim N(a, \sigma^2)$, *то* $\xi_0 = \frac{\xi-a}{\sigma} \sim N(0,1)$.

Доказательство. При $A \neq 0$ функция $y = f(x) = Ax + B$ является строго монотонной дифференцируемой функцией. Обратная функция имеет вид $x = f^{-1}(y) = \frac{y-B}{A}$, причем $[f^{-1}(y)]' = \frac{1}{A}$. По теореме о плотности вероятности функции от случайной величины случайная величина η абсолютно непрерывна и имеет плотность распределения вероятностей

$$p_\eta(y) = p_\xi(f^{-1}(y))|[f^{-1}(y)]'| = p_\xi\Big(\frac{y-B}{A}\Big)\frac{1}{|A|} =$$

$$= \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(\frac{y-B}{A}-a)^2}{2\sigma^2}} \frac{1}{|A|} = \frac{1}{\sqrt{2\pi}\sigma|A|} e^{-\frac{[y-(Aa+B)]^2}{2\sigma^2 A^2}},$$

т.е. $\eta \sim N(Aa+B,\, A^2\sigma^2)$.

Если, в частности, $\eta = \xi_0 = \frac{\xi - a}{\sigma}$, то $A = 1/\sigma$, $B = -a/\sigma$, т.е. $Aa + B = 0$, $A^2\sigma^2 = 1$, поэтому $\xi_0 \sim N(0,1)$. □

Следствие 10.8. *Если $\xi \sim N(a,\, \sigma^2)$, то эта случайная величина может быть представлена в виде*

$$\xi = \sigma\xi_0 + a,$$

где $\xi_0 \sim N(0,1)$.

Пусть $\xi \sim N(a,\, \sigma^2)$. Как найти вероятность

$$P\{\alpha < \xi < \beta\}?.$$

Так как ξ абсолютно непрерывна, то $P\{\xi = \alpha\} = 0$. Поэтому

$$P\{\alpha < \xi < \beta\} = P\{\alpha \le \xi < \beta\} = F_\xi(\beta) - F_\xi(\alpha).$$

Чтобы воспользоваться этой формулой, нам необходимо бы было иметь таблицы функций нормального распределения при любых значениях параметров a, σ^2. К счастью, лемма о нормальном распределении позволяет обойтись без этого неудобного шага и пользоваться таблицами стандартного нормального распределения. В самом деле,

$$P\{\alpha < \xi < \beta\} = P\{\alpha - a < \xi - a < \beta - a\} =$$

$$= P\left\{\frac{\alpha - a}{\sigma} < \frac{\xi - a}{\sigma} < \frac{\beta - a}{\sigma}\right\} = P\left\{\frac{\alpha - a}{\sigma} < \xi_0 < \frac{\beta - a}{\sigma}\right\} =$$

$$= N\Big(\frac{\beta - a}{\sigma}\Big) - N\Big(\frac{\alpha - a}{\sigma}\Big) = \Phi\Big(\frac{\beta - a}{\sigma}\Big) - \Phi\Big(\frac{\alpha - a}{\sigma}\Big).$$

В частности, если $\xi \sim N(a,\, \sigma^2)$, то

$$P\Big\{\Big|\frac{\xi - a}{\sigma}\Big| < \varepsilon\Big\} = 2\Phi(\varepsilon).$$

Мы видели, что нормальное распределение является предельным для биномиального распределения. Позже мы обобщим этот результат и докажем, что нормальное распределение выступает при достаточно общих предположениях в качестве предельного распределения для сумм так называемых независимых случайных величин. Этот результат — центральная предельная теорема теории вероятностей. Он теоретически обосновывает частое появление нормального закона распределения в практических ситуациях.

III. Сингулярные распределения.

Определение 10.9. Точка x называется **точкой роста** функции распределения $F(x)$, если

$$\forall \varepsilon > 0 \ \text{ выполняется } \ F(x + \varepsilon) - F(x - \varepsilon) > 0.$$

Определение 10.10. Случайная величина ξ, ее распределение P_ξ и функция распределения F_ξ называются **сингулярными**, если

1) F непрерывна;

2) множество точек роста F имеет лебегову меру ноль.

Если x не является точкой роста, то найдется $\varepsilon > 0$ такое, что во всем промежутке $[x - \varepsilon, x + \varepsilon]$ функция распределения будет постоянна и, следовательно, ее производная в точке x будет равна нулю. Итак, если ξ сингулярна, то $F(x)$ непрерывна, а $F'(x) = 0$ почти всюду. Такая случайная величина не может быть абсолютно непрерывной, так как тогда ее плотность вероятности $p(x) = 0$ почти всюду, поэтому $\int\limits_{-\infty}^{+\infty} p(x)dx = 0$, что невозможно для абсолютно непрерывного случая. Объясняется это известным фактом из теории неопределенного интеграла Лебега: функция восстанавливается по своей производной в виде неопределенного интеграла тогда и только тогда, когда она абсолютно непрерывна.

Пример 1. **Канторова лестница**. Возьмем отрезок $[0, 1]$ и построим функцию $F(x)$ следующим образом (см. Рис.4). Положим $F(x) = 0$ при $x < 0$ и $F(x) = 1$ при $x > 1$. Отрезок $[0, 1]$ разбивается на три части равной длины $[0, 1/3]$, $(1/3, 2/3)$, $[2/3]$. На срединном интервале полагаем $F(x) = 1/2$. Оставшиеся два отрезка снова разбиваются на три части равной длины каждый, и на срединных интервалах $F(x)$ полагается равной соответственно $1/4$ и $3/4$. Каждый из оставшихся промежутков снова делится на три части равной длины и на средних интервалах $F(x)$ определяется как постоянная, равная среднему арифметическому между соседними, уже определенными значениями $F(x)$, и т.д.

Таким образом, F определена во всех точках отрезка $[0, 1]$ за исключением точек канторова совершенного множества K. Это множество, как мы помним, одновременно и "большое" и "маленькое", так как имеет мощность континуума и лебегову меру ноль. Каждая точка K является

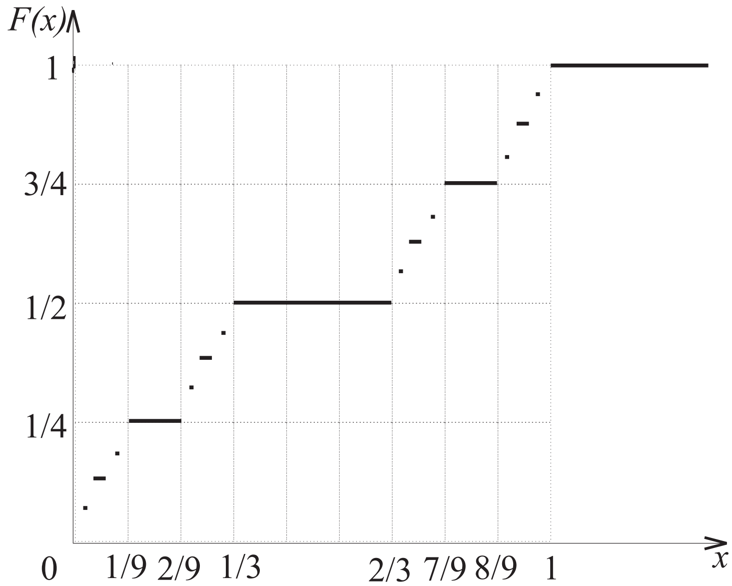


Рис.4

предельной для точек $\overline{K} = [0,1] \setminus K$. Поэтому в точках K мы можем доопределить $F(x)$ по непрерывности. Итак, $F(x)$ непрерывна. Очевидно, построенная функция F не убывает, непрерывна, $F(-\infty) = 0,\ F(+\infty) + 1$, т.е. является функцией распределения некоторой случайной величины. Пусть N — множество точек роста F, очевидно, $N \subset K$, откуда $\overline{K} \subset \overline{N} = [0,1] \setminus N$, следовательно,

$$\mu(\overline{N}) \geq \mu(\overline{K}) = \frac{1}{3} + \frac{2}{9} + \frac{4}{27} + \ldots = \frac{1/3}{1 - 2/3} = 1,$$

где μ — мера Лебега. Поэтому $\mu(\overline{N}) = 1$, откуда $\mu(N) = 0$. Итак, F — непрерывна, а множество ее точек роста имеет меру ноль. Кроме того, эта функция не убывает, $F(-\infty) = 0$, $F(+\infty) = 1$. Следовательно, F — функция распределения некоторой случайной величины, и она сингулярна.

Приведем без доказательства известную теорему Лебега.

Теорема Лебега. *Функция распределения любой случайной величины может быть представлена в виде*

$$F(x) = p_1 F_1(x) + p_2 F_2(x) + p_3 F_3(x),$$

где $p_1, p_2, p_3 \geq 0$, $p_1 + p_2 + p_3 = 1$ *, а* F_1 *— дискретная функция распределения,* F_2 *— абсолютно непрерывная функция распределения,* F_3 *— сингулярная функция распределения.*

Таким образом, распределение любой случайной величины либо принадлежит к "чистому" типу (является дискретным, абсолютно непрерывным или сингулярным), либо является взвешенной линейной комбинацией этих трех типов. Отметим, что на практике сингулярная составляющая, как правило отсутствует. Поэтому прикладникам нет даже необходимости знать что-либо о сингулярном типе. Они сталкиваются с дискретными, абсолютно непрерывными распределениями и немного реже со смесями дискретных и абсолютно-непрерывных распределений. А в учебниках для не математиков, как правило, рассматриваются только дискретные и абсолютно непрерывные случайные величины и даже не рассматриваются смеси их распределений.

§ 11. Случайный вектор. Совместное распределение случайных величин

Во многих вопросах теории вероятностей возникает необходимость исследования нескольких случайных величин, которые могут так или иначе влиять друг на друга. В таком случае знания законов их распределения, вообще говоря, недостаточно для описания взаимодействия между ними. Для этого необходимо задать так называемый закон совместного распределения, что удобнее всего сделать с помощью введения понятия случайного вектора.

Пусть $\xi_1 = \xi_1(\omega), \xi_2 = \xi_2(\omega), \dots, \xi_n = \xi_n(\omega)$ — случайные величины, определенные на одном и том же вероятностном пространстве $(\Omega, \mathcal{F}, P)$.

Определение 11.1. Отображение $\xi : \Omega \to \mathbb{R}^n$, ставящее в соответствие каждому $\omega \in \Omega$ n-мерный вектор $\xi(\omega) = (\xi_1(\omega), \xi_2(\omega), \dots, \xi_n(\omega))$ из $\mathbb{R}^n$ называется (n-мерным) **случайным вектором** или (n-мерной) **случайной точкой**.

Для краткости записи мы часто будем обозначать случайный вектор как $\xi = (\xi_1, \xi_2, \dots, \xi_n)$. Рассмотрим n-мерные параллелепипеды вида $[a_1, b_1) \times [a_2, b_2) \times \dots \times [a_n, b_n)$, где принята та же оговорка, которая применялась нами в одномерном случае, что $[-\infty, b)$ и $[-\infty, +\infty)$ обозначают соответственно интервалы $(-\infty, b)$ и $(-\infty, +\infty)$. Пусть алгебра $\mathcal{A}$ состоит из всех конечных объединений попарно непересекающихся параллелепипедов указанного вида. Как и в одномерном случае, можно доказать, что $\mathcal{B}^n = \mathcal{B}(\mathbb{R}^n) = \sigma(\mathcal{A})$, т.е., что σ-алгебра n-мерных борелевских множеств совпадает с наименьшей σ-алгеброй, содержащей алгебру $\mathcal{A}$. С помощью этого факта легко доказать, что определение 11.1 эквивалентно следующему.

Определение 11.2. Отображение $\xi : \Omega \to \mathbb{R}^n$ называется **случайным вектором**, если оно измеримо относительно σ-алгебры $\mathcal{F}$, т.е.

$$\forall B \in \mathcal{B}^n \text{ выполняется } \xi^{-1}(B) \in \mathcal{F}.$$

Действительно, из определения 11.2 следует определение 11.1. Ведь множество вида $B = \mathbb{R} \times \mathbb{R} \times \ldots \times \mathbb{R} \times (-\infty, x_i) \times \mathbb{R} \times \ldots \times \mathbb{R} \in \mathcal{B}^n$, следовательно,

$$\xi^{-1}(B) = \{\omega : \xi_1(\omega) \in \mathbb{R}, \ldots, \xi_{i-1}(\omega) \in \mathbb{R}, \xi_i(\omega) < x_i,$$
$$\xi_{i+1}(\omega) \in \mathbb{R}, \ldots, \xi_n(\omega) \in \mathbb{R}\} =$$
$$= \{\omega : \xi_i(\omega) < x_i\} \in \mathcal{F},$$

Поэтому каждая ξ_i $(i = 1, 2, \ldots, n)$ — случайная величина.

Докажем теперь, что из определения 11.1 следует определение 11.2. Рассмотрим семейство множеств

$$\mathfrak{M} = \{B \in \mathbb{R}^n : \xi^{-1}(B) \in \mathcal{F}\},$$

которое, как нетрудно видеть, образует σ-алгебру.

Задача 1. Докажите, что $\mathfrak{M}$ — σ-алгебра.

Поскольку

$$\xi^{-1}((-\infty, x_1) \times \ldots \times (-\infty, x_n)) =$$

$$= \{\omega:\ \xi_1(\omega) < x_1,\ \ldots,\ \xi_n(\omega) < x_n\} =$$

$$= \bigcap_{i=1}^{n} \{\omega:\ \xi_i(\omega) < x_i\} \in \mathcal{F},$$

все параллелепипеды вида $[a_1,\, b_1) \times [a_2,\, b_2) \times \ldots \times [a_n,\, b_n)$ выражаются с помощью конечного числа теоретико-множественных операций через множества вида $(-\infty, x_1) \times \ldots \times (-\infty, x_n)$, а операция взятия прообраза сохраняет все теоретико-множественные операции, то все вышеуказанные параллелепипеды принадлежат σ-алгебре $\mathfrak{M}$. Отсюда следует, что $\mathcal{A} \subset \mathfrak{M}$. Так как $\mathfrak{M}$ — σ-алгебра, содержащая $\mathcal{A}$, то она содержит наименьшую σ-алгебру, содержащую $\mathcal{A}$, т.е. $\mathcal{B}^n$. Итак, $\mathcal{B}^n \subset \mathfrak{M}$. По определению $\mathfrak{M}$ это означает, что $\forall B \in \mathcal{B}^n$ выполняется $\xi^{-1}(B) \in \mathcal{F}$, т.е. имеет место определение 11.1. $\square$

Определение 11.3. Вероятностная мера, определенная на σ-алгебре n-мерных борелевских множеств $\mathcal{B}^n$ с помощью равенства

$$P_\xi(B) = P_{\xi_1, \xi_2, \ldots, \xi_n}(B) =$$

$$= P\{\omega:\ \xi(\omega) \in B\} = P\{\xi \in B\} = P(\xi^{-1}(B)),$$

называется **распределением случайного вектора** ξ или **совместным распределением случайных величин** $\xi_1, \xi_2, \ldots, \xi_n$. Функция $F = F_\xi = F_{\xi_1, \xi_2, \ldots, \xi_n} :\ \mathbb{R}^n \to \mathbb{R}$ вида

$$F(x) = F_\xi(x) = F_{\xi_1, \xi_2, \ldots, \xi_n}(x_1, x_2, \ldots, x_n) =$$

$$= P\{\omega:\ \xi(\omega) < x\} =$$

$$= P\{\omega : \ \xi_1(\omega) < x_1, \ \xi_2(\omega) < x_2, \ \dots, \ \xi_n(\omega) < x_n\} =$$
$$= P\{\xi_1 < x_1, \ \xi_2 < x_2, \ \dots, \ \xi_n < x_n\}$$

называется **функцией распределения случайного вектора** ξ или **функцией совместного распределения (совместной функцией распределения) случайных величин** $\xi_1, \xi_2, \dots, \xi_n$.

Это определение корректно, так как множества, стоящие под знаком вероятности, являются событиями (т.е. принадлежат $\mathcal{F}$). То, что распределение является вероятностной мерой, доказывается точно так же, как и для случайной величины.

Задача 2. Докажите это.

Свойства функции распределения.

1. $0 \le F(x) \le 1$.

2. Функция распределения непрерывна слева по каждой переменной.

Доказательство аналогично доказательству сходного свойства функции распределения случайной величины.

Задача 3. Докажите свойство 2.

3. Функция распределения удовлетворяет условиям согласованности:

а) если $i_1, i_2, \dots, i_n$ — произвольная перестановка элементов множества $\{1, 2, \dots, n\}$, то

$$F_{\xi_{i_1}, \xi_{i_2}, \dots, \xi_{i_n}}(x_{i_1}, x_{i_2}, \dots, x_{i_n}) = F_{\xi_1, \xi_2, \dots, \xi_n}(x_1, x_2, \dots, x_n);$$

б) для любого $i = 1, 2, \ldots, n$ выполняется следующее:

$$\lim_{x_i \to -\infty} F_{\xi_1, \ldots, \xi_n}(x_1, \ldots, x_n) = 0;$$

$$\lim_{x_i \to +\infty} F_{\xi_1, \ldots, \xi_n}(x_1, \ldots, x_n) =$$

$$= F_{\xi_1, \ldots, \xi_{i-1}, \xi_{i+1}, \xi_{i+2}, \ldots, \xi_n}(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n).$$

Свойство а) очевидно и выражает коммутативность операции пересечения множеств. Свойство б) доказывается с помощью свойства непрерывности вероятности точно так же, как свойство функции распределения случайной величины $F(-\infty) = 0,\ F(+\infty) = 1$.

Задача 4. Докажите свойство 3.

Впрочем, если первое свойство б) доказано, то второе свойство б) можно доказать с помощью аксиомы счетной аддитивности:

$$F_{\xi_1, \ldots, \xi_{i-1}, \xi_{i+1}, \xi_{i+2}, \ldots, \xi_n}(x_1, \ldots, x_{i-1}, x_{i+1}, x_{i+2}, \ldots, x_n) =$$

$$= P\{\xi_1 < x_1, \ldots, \xi_{i-1} < x_{i-1}, \xi_{i+1} < x_{i+1}, \xi_{i+2} < x_{i+2}, \ldots,$$

$$\xi_n < x_n\} = P\{\xi_1 < x_1, \ldots, \xi_{i-1} < x_{i-1}, -\infty < \xi_i < +\infty,$$

$$\xi_{i+1} < x_{i+1}, \ldots, \xi_n < x_n\} = P\{\xi_1 < x_1, \ldots, \xi_{i-1} < x_{i-1},$$

$$\sum_{k=-\infty}^{+\infty} \{k \le \xi_i < k+1\}, \xi_{i+1} < x_{i+1}, \ldots, \xi_n < x_n\} =$$

$$= F_{\xi_1, \ldots, \xi_n}(x_1, \ldots, x_{i-1}, +\infty, x_{i+1}, \ldots, x_n).$$

Последнее равенство получается по определению суммы ряда.

Следствие 11.4.

$$F(+\infty, \dots, +\infty) =$$
$$= \lim_{x_1 \to +\infty} \lim_{x_2 \to +\infty} \dots \lim_{x_n \to +\infty} F(x_1, \dots, x_n) = 1.$$

4. Функция распределения случайного вектора не убывает по каждой переменной.

Действительно, так как если $x_i < y_i$, то

$$\{\omega:\ \xi_1(\omega) < x_1, \dots, \xi_{i-1}(\omega) < x_{i-1}, \xi_i(\omega) < x_i,$$
$$\xi_{i+1}(\omega) < x_{i+1}, \dots, \xi_n(\omega) < x_n\} \subset$$
$$\subset \{\omega:\ \xi_1(\omega) < x_1, \dots, \xi_{i-1}(\omega) < x_{i-1}, \xi_i(\omega) < y_i,$$
$$\xi_{i+1}(\omega) < x_{i+1}, \dots, \xi_n(\omega) < x_n\},$$

и вероятность первого события не превосходит вероятности второго, т.е.

$$F(x_1, \dots, x_{i-1}, x_i, x_{i+1}, \dots, x_n) \leq$$
$$\leq F(x_1, \dots, x_{i-1}, y_i, x_{i+1}, \dots, x_n).$$

Оказывается, это свойство может быть усилено следующим образом. Сначала рассмотрим двумерный случайный вектор $\xi = (\xi_1, \xi_2)$. Пусть $\Pi = [a_1, b_1) \times [a_2, b_2)$ — прямоугольник на плоскости. Вероятность попадания в него двумерной случайной точки ξ равна

$$P\{\xi \in \Pi\} = P\{a_1 \leq \xi_1 < b_1,\ a_2 \leq \xi_2 < b_2\} =$$
$$= P\{a_1 \leq \xi_1 < b_1,\ \xi_2 < b_2\} - P\{a_1 \leq \xi_1 < b_1,\ \xi_2 < a_2\} =$$
$$= F(b_1, b_2) - F(a_1, b_2) - [F(b_1, a_2) - F(a_1, a_2)] =$$

$$= F(b_1, b_2) - F(a_1, b_2) - F(b_1, a_2) + F(a_1, a_2).$$

Для функции двух действительных переменных f введем разностные операторы Δ_{a_1,b_1} и Δ_{a_2,b_2} следующим образом:

$$\Delta_{a_1,b_1} f(x_1, x_2) = f(b_1, x_2) - f(a_1, x_2),$$

$$\Delta_{a_2,b_2} f(x_1, x_2) = f(x_1, b_2) - f(x_1, a_2).$$

С их помощью вероятность попадания в прямоугольник двумерной случайной точки запишется в виде

$$P\{\xi \in \Pi\} = \Delta_{a_1,b_1} \Delta_{a_2,b_2} F(x_1, x_2).$$

Так как вероятность попадания в прямоугольник неотрицательна, то для двумерного случая желаемое свойство принимает форму

$$\Delta_{a_1,b_1} \Delta_{a_2,b_2} F(x_1, x_2) \geq 0$$

для любых действительных чисел a_1, a_2, b_1, b_2 таких, что $a_1 \leq b_1, \;\; a_2 \leq b_2$.

Возвращаясь к n-мерному вектору, для функции n переменных f при $i = 1, 2, \ldots, n$ введем разностные операторы Δ_{a_i,b_i} для любых действительных чисел $a_i \leq b_i$ как

$$\Delta_{a_i,b_i} f(x_1, x_2, \ldots, x_n) = f(x_1, \ldots, x_{i-1}, b_i, x_{i+1}, \ldots, x_n) -$$

$$- f(x_1, \ldots, x_{i-1}, a_i, x_{i+1}, \ldots, x_n).$$

Простой подсчет показывает, что вероятность попадания случайного вектора в параллелепипед $\Pi = [a_1, b_1) \times [a_2, b_2) \times \ldots [a_n, b_n)$ равна

$$P\{\xi \in \Pi\} = \Delta_{a_1,b_1} \Delta_{a_2,b_2} \ldots \Delta_{a_n,b_n} F(x_1, x_2, \ldots, x_n). \tag{11.1}$$

Поскольку эта вероятность неотрицательна, то получается свойство

5. Для любых действительных чисел $a_1, a_2, \dots, a_n, b_1, b_2, \dots, b_n$ таких, что $a_1 \le b_1$, $a_2 \le b_2$, ..., $a_n \le b_n$

$$\Delta_{a_1,b_1}\Delta_{a_2,b_2}\dots\Delta_{a_n,b_n}F(x_1, x_2, \dots, x_n) \ge 0. \qquad (11.2)$$

Задача 5. Докажите формулу (11.1).

Теорема о существовании случайного вектора с заданным распределением.

Пусть $F : \mathbb{R}^n \to \mathbb{R}$ — функция n переменных, удовлетворяющая свойствам:

1) F — тотально возрастающая в широком смысле функция, т.е. для любых действительных чисел $a_1, a_2, \dots, a_n, b_1, b_2, \dots, b_n$ таких, что $a_1 \le b_1$, $a_2 \le b_2$, ..., $a_n \le b_n$

$$\Delta_{a_1,b_1}\Delta_{a_2,b_2}\dots\Delta_{a_n,b_n}F(x_1, x_2, \dots, x_n) \ge 0;$$

2) для любого $i = 1, 2, \dots, n$

$$\lim_{x_i \to -\infty} F_{\xi_1,\dots,\xi_n}(x_1, \dots, x_n) = 0;$$

$$F(+\infty, \dots, +\infty) =$$

$$= \lim_{x_1 \to +\infty} \lim_{x_2 \to +\infty} \dots \lim_{x_n \to +\infty} F(x_1, \dots, x_n) = 1;$$

3) F непрерывна слева по каждой переменной.

Тогда существует вероятностное пространство $(\Omega, \mathcal{F}, P)$ *и* n*-мерный случайный вектор* ξ*, определенный на нем, такой, что его функция распределения совпадает с заданной функцией* F:

$$F_\xi(x) = F(x), \quad x \in \mathbb{R}^n.$$

Примем эту теорему без доказательства.

Распределение случайного вектора однозначно определяет его функцию распределения по формуле

$$F_{\xi_1,\xi_2,\dots\xi_n}(x_1, x_2, \dots, x_n) = P(\xi \in I),$$

где $I = (-\infty, x_1) \times (-\infty, x_2) \times \dots (-\infty, x_n)$. Как и для случайных величин, для случайных векторов может быть доказано обратное утверждение: *функция распределения случайного вектора однозначно определяет его распределение.* Итак, распределение и функция распределения случайного вектора однозначно определяют друг друга.

Определение 11.5. Любая функция, однозначно определяющая распределение (функцию распределения), называется **законом распределения случайного вектора**.

Закон распределения позволяет находить вероятность попадания случайного вектора в любое n-мерное борелевское множество.

Задача 6. Докажите, что функция распределения случайного вектора однозначно определяет его распределение.

Определение 11.6. Случайный вектор ξ, его распределение P_ξ и функция распределения F_ξ называются **дискретными**, если мера P_ξ сосредоточена в конечном или счетном множестве точек.

Как и для случайных величин, для случайных векторов с помощью аналогичных рассуждений можно прийти к эквивалентному определению.

Определение 11.7. Случайный вектор ξ, его распределение P_ξ и функция распределения F_ξ называются **дискретными**, если множество возможных значений случайного вектора ξ конечно или счетно.

Очевидно, координаты случайного дискретного вектора являются дискретными случайными величинами.

Закон распределения дискретного случайного вектора удобно задавать в виде соответствия между возможными значениями его координат $x_1, x_2, \ldots, x_n$ и совместными вероятностями этих значений $P\{\xi_1 = x_1, \xi_2 = x_2, \ldots \xi_n = x_n\}$. Это — закон распределения, так как для любого n-мерного борелевского множества B

$$P\{\xi \in B\} = \sum_{(x_1, x_2, \ldots, x_n) \in B} P\{\xi_1 = x_1, \xi_2 = x_2, \ldots \xi_n = x_n\}.$$

Для иллюстрации рассмотрим двумерный дискретный случайный вектор $\xi = (\xi_1, \xi_2)$. Пусть $x_1, x_2, \ldots$ — возможные значения первой его координаты ξ_1, $y_1, y_2, \ldots$ — возможные значения второй координаты ξ_2. Обозначим через $p_{ij} = P\{\xi_1 = x_i, \xi_2 = y_j\}$ — вероятности возможных значений случайного вектора.

Определение 11.8. Соответствие между значениями случайного вектора и их вероятностями, заданное с помощью таблицы

$\xi_1 \backslash \xi_2$	y_1	y_2	y_3	....
x_1	p_{11}	p_{12}	p_{13}	....
x_2	p_{21}	p_{22}	p_{23}	....
x_3	p_{31}	p_{32}	p_{33}	....
....	....	....	....	....

,

называется **рядом** или **матрицей распределения** двумерного дискретного вектора.

Поскольку

$$\{\xi_1 = x_i\} = \sum_j \{\xi_1 = x_i \ \xi_2 = y_j\},$$

то

$$p_{i.} = P\{\xi_1 = x_i\} = \sum_j p_{ij}. \qquad (11.3)$$

Аналогично,

$$p_{.j} = P\{\xi_2 = y_j\} = \sum_i p_{ij}. \qquad (11.4)$$

Очевидно, что

$$\sum_{i,\,j} p_{ij} = \sum_i p_{i.} = 1.$$

С помощью (11.3) и (11.4) по ряду распределения двумерного вектора определяются ряды распределений его координат. Это обстоятельство удобно учесть в матрице распределения вектора, суммируя вероятности в ее строках, а затем в столбцах, и добавляя полученный столбец и строку к этой матрице. Расширенная матрица содержит как информацию о распределении вектора, так и информацию о распределении его координат:

$\xi_1 \backslash \xi_2$	y_1	y_2	y_3	....	
x_1	p_{11}	p_{12}	p_{13}	....	$p_{1.}$
x_2	p_{21}	p_{22}	p_{23}	....	$p_{2.}$
x_3	p_{31}	p_{32}	p_{33}	....	$p_{3.}$
....	....	....	....	....	....
	$p_{.1}$	$p_{.2}$	$p_{.3}$	....	

Определение 11.9. Случайный вектор ξ, его распределение P_ξ и функция распределения F_ξ называются

абсолютно непрерывными, если существует неотрицательная борелевская функция

$$p(x) = p_\xi(x) = p(x_1, x_2, \dots, x_n) = p_{\xi_1, \xi_2, \dots, \xi_n}(x_1, x_2, \dots, x_n)$$

такая, что для любого n-мерного борелевского множества B

$$P_\xi(B) = P\{\xi \in B\} =$$

$$= \int_B \cdots \int p(x_1, x_2, \dots, x_n) dx_1 dx_2 \dots dx_n. \qquad (11.5)$$

При этом функция $p(x)$, определенная с помощью (11.5) с точностью до значений на множестве n-мерной лебеговой меры ноль, называется **плотностью распределения вероятностей**, либо **плотностью вероятности случайного вектора** ξ, либо **плотностью совместного распределения случайных величин** $\xi_1, \xi_2, \dots, \xi_n$.

Плотность распределения обладает следующими свойствами:

1) $p_\xi(x) \geq 0$ для любого $x \in \mathbb{R}^n$;

2) $\int\limits_{\mathbb{R}^n} \dots \int p_\xi(x) dx_1 dx_2 \dots dx_n = 1$.

Свойство 2) следует из (11.5), если положить $B = \mathbb{R}^n$. Беря в (11.2), в частности, $B = (-\infty, x_1) \times (-\infty, x_2) \times \dots (-\infty, x_n)$, получим:

$$F_{\xi_1, \xi_2, \dots, \xi_n}(x_1, x_2, \dots, x_n) =$$

$$= \int\limits_{-\infty}^{x_1} \int\limits_{-\infty}^{x_2} \dots \int\limits_{-\infty}^{x_n} p_{\xi_1, \xi_2, \dots, \xi_n}(t_1, t_2, \dots, t_n) dt_n \dots dt_2 dt_1. \qquad (11.6)$$

Из (11.5) или из (11.6) следует, что плотность распределения является законом распределения абсолютно непрерывного случайного вектора. Кроме того, из (11.6) с помощью последовательного дифференцирования по $x_1, x_2, \ldots, x_n$ вытекает, что почти всюду (относительно n-мерной меры Лебега)

$$p(x_1, x_2, \ldots, x_n) = \frac{\partial^n F(x_1, x_2, \ldots, x_n)}{\partial x_1 \partial x_2 \ldots \partial x_n}.$$

Может быть доказана следующая

Теорема о существовании абсолютно непрерывного случайного вектора с заданной плотностью распределения вероятностей.

Пусть функция $p : \mathbb{R}^n \to \mathbb{R}$ обладает свойствами 1), 2). Тогда существует вероятностное пространство $(\Omega, \mathcal{F}, P)$ и абсолютно непрерывный случайный вектор $\xi = (\xi_1, \xi_2, \ldots, \xi_n)$, определенный на нем такой, что его плотность распределения вероятностей совпадает с заданной функцией:

$$p_{\xi_1, \xi_2, \ldots, \xi_n}(x_1, x_2, \ldots, x_n) = p(x_1, x_2, \ldots, x_n).$$

Если плотность распределения непрерывна, то, применяя теорему о среднем для интеграла (11.5) с $B = [x_1, x_1 + \Delta x_1) \times [x_2, x_2 + \Delta x_2) \times \ldots \times [x_n, x_n + \Delta x_n)$, получим с точностью до бесконечно малых более высокого порядка малости, чем $\Delta x_1 \Delta x_2 \ldots \Delta x_n$, что

$$P\{x_1 \le \xi_1 < x_1 + \Delta x_1, x_2 \le \xi_2 < x_2 + \Delta x_2, \ldots x_n \le \xi_n <$$

$$< x_n + \Delta x_n\} = p(x_1, x_2, \ldots, x_n) \Delta x_1 \Delta x_2 \ldots \Delta x_n.$$

Пусть $\xi = (\xi_1, \xi_2)$ — двумерный абсолютно непрерывный случайный вектор. Поскольку

$$P\{\xi_1 \in B\} = P\{\xi_1 \in B,\ \xi_2 \in \mathbb{R}\} = P\{\xi \in B \times \mathbb{R}\} =$$
$$= \iint\limits_{B\times\mathbb{R}} p_{\xi_1,\xi_2}(x_1, x_2) dx_1 dx_2 = \int\limits_B \Big[\int\limits_{\mathbb{R}} p_{\xi_1,\xi_2}(x_1, x_2) dx_2\Big] dx_1,$$

то по определению абсолютно непрерывной случайной величины это означает, что ξ_1 абсолютно непрерывна и имеет плотность распределения вероятностей

$$p_{\xi_1}(x_1) = \int\limits_{-\infty}^{+\infty} p_{\xi_1,\xi_2}(x_1, x_2) dx_2.$$

Аналогично, вторая координата абсолютно непрерывного случайного вектора является абсолютно непрерывной случайной величиной с плотностью распределения вероятностей

$$p_{\xi_2}(x_2) = \int\limits_{-\infty}^{+\infty} p_{\xi_1,\xi_2}(x_1, x_2) dx_1.$$

Последние две формулы являются непрерывным аналогом формул (11.3) и (11.4).

Отметим, что по индукции отсюда следует, что любая подгруппа координат абсолютно непрерывного вектора образует абсолютно непрерывный вектор, плотность распределения которого получается с помощью интегрирования плотности распределения вектора ξ по подпространству переменных, соответствующих выброшенным координатам случайного вектора. Например, для $n < m$

$$p_{\xi_1,\dots,\xi_m}(x_1, \dots, x_m) = \int\limits_{\mathbb{R}^{n-m}} \cdots \int p_\xi(x) dx_{m+1} dx_{m+2} \dots dx_n.$$

§ 12. Независимость случайных величин

Очень важное значение в теории вероятностей придается понятию независимости случайных величин.

Определение 12.1. Случайные величины $\xi_1, \xi_2, \dots, \xi_n$ называются **независимыми**, если для любых действительных $x_1, x_2, \dots, x_n$

$$F_{\xi_1,\xi_2,\dots,\xi_n}(x_1,x_2,\dots,x_n) = \\ = F_{\xi_1}(x_1)F_{\xi_2}(x_2)\dots F_{\xi_n}(x_n). \tag{12.1}$$

Положим в (12.1) $x_n = k$, где k — натуральное число. Тогда

$$F_{\xi_1,\xi_2,\dots,\xi_n}(x_1,x_2,\dots,x_{n-1},k) = F_{\xi_1}(x_1)F_{\xi_2}(x_2)\dots \\ \dots F_{\xi_{n-1}}(x_{n-1})F_{\xi_n}(k). \tag{12.2}$$

Рассмотрим последовательность событий вида $A_k = \{\omega : \xi_1(\omega) < x_1,\ \xi_2(\omega) < x_2, \dots, \xi_{n-1}(\omega) < x_{n-1},\ \xi_n(\omega) < k\}$. Очевидно, $A_k \subset A_{k+1}$ и $\bigcap\limits_k A_k = A = \{\omega : \quad \xi_1(\omega) < x_1,\ \xi_2(\omega) < x_2,\ \dots,\ \xi_{n-1}(\omega) < x_{n-1}\}$, т.е. $A_k \uparrow A$. По свойству непрерывности вероятности $P(A_k) \to P(A)$. Поэтому левая часть (12.2) сходится к $P(A) = F_{\xi_1,\xi_2,\dots,\xi_{n-1}}(x_1,x_2,\dots,x_{n-1})$. По свойству функции распределения $F_{\xi_n}(k) \to 1$ при $k \to \infty$. Переходя в (12.2) к пределу, получим:

$$F_{\xi_1,\xi_2,\dots,\xi_{n-1}}(x_1,x_2,\dots,x_{n-1}) = \\ = F_{\xi_1}(x_1)F_{\xi_2}(x_2)\dots F_{\xi_{n-1}}(x_{n-1}).$$

Это означает, что случайные величины $\xi_1, \xi_2, \ldots, \xi_{n-1}$ независимы. Легко понять, что отсюда следует

Свойство 1. Если случайные величины $\xi_1, \xi_2, \ldots, \xi_n$ *независимы, то для любого множества не менее двух индексов* $1 \le i_1 < i_2 < \ldots < i_k \le n$ *случайные величины* $\xi_{i_1}, \xi_{i_2}, \ldots, \xi_{i_k}$ *независимы.*

В этом существенное различие от независимости n событий, при определении которой кроме равенства $P(A_1A_2\ldots A_n) = P(A_1)P(A_2)\ldots P(A_n)$ требовались аналогичные равенства для любых подгрупп событий. Здесь же равенства для подгрупп автоматически вытекают из одного этого равенства.

Наше определение независимости случайных величин означает, что для любых действительных $x_1, x_2, \ldots, x_n$ события $\{\xi_1 < x_1\}, \{\xi_2 < x_2\}, \{\xi_n < x_n\}$ независимы.

Если случайная величина как функция от $\omega \in \Omega$ постоянна, т.е. $\xi(\omega) \equiv C$, то мы будем называть ее **неслучайной** или **постоянной** величиной.

Свойство 2. Любая случайная и неслучайная величины независимы между собой.

Д о к а з а т е л ь с т в о. Пусть $\xi = \xi(\omega)$ — случайная величина, $\eta(\omega) \equiv C$ — неслучайная величина. Очевидно,

$$F_C(y) = P\{\omega :\ \ C < y\} = \begin{cases} P(\emptyset) = 0, & \text{если } y \le C, \\ P(\Omega) = 1, & \text{если } y > C. \end{cases}$$

Поэтому совместная функция распределения ξ и C равна

$$F_{\xi,\,C}(x,\ y) = P\{\xi < x,\ C < y\} =$$

$$= \begin{cases} P\{\{\xi < x\} \cap \emptyset\} = 0, & \text{если } y \le C, \\ P\{\{\xi < x\} \cap \Omega\} = P\{\xi < x\}, & \text{если } y > C \end{cases} =$$

$$= P\{\xi < x\} \begin{cases} 0, & \text{если } y \le C, \\ 1, & \text{если } y > C \end{cases} = F_\xi(x) F_C(y).$$

Согласно определению 12.1 это и означает, что ξ и C независимы.

□

Наше определение независимости означает, что

$$P\{\omega : \ \xi_1(\omega) \in B_1, \ \xi_2(\omega) \in B_2, \ \dots , \ \xi_n(\omega) \in B_n\} = \\ = P\{\omega : \ \xi_1(\omega) \in B_1\} P\{\omega : \ \xi_2(\omega) \in B_2\} \dots \\ \dots P\{\omega : \ \xi_n(\omega) \in B_n\} \qquad (12.3)$$

для любых борелевских множеств вида $B_1 = (-\infty, x_1)$, $B_2 = (-\infty, x_2)$, $B_n = (-\infty, x_n)$. Оказывается, что конкретный вид борелевских множеств в определении 12.1 можно не учитывать, а именно, имеет место следующий

Критерий независимости случайных величин. *Для того, чтобы случайные величины* $\xi_1, \ \xi_2, \ \dots , \ \xi_n$ *были независимыми, необходимо и достаточно, чтобы для любых борелевских множеств* $B_1, \ B_2, \ \dots , \ B_n$ *выполнялось* (12.3).

Д о к а з а т е л ь с т в о. *Достаточность.* Из (12.3) следует (12.1), если в (12.3) взять $B_1 = (-\infty, x_1)$, $B_2 = (-\infty, x_2)$, $\dots$ $B_n = (-\infty, x_n)$.

Необходимость. Для простоты будем считать, что $n = 2$. Общий случай получается с помощью математической индукции по n. Пусть выполнено (12.1):

$$F_{\xi_1, \xi_2}(x_1, x_2) = F_{\xi_1}(x_1) F_{\xi_2}(x_2). \qquad (12.4)$$

Требуется доказать, что для любых борелевских множеств $B_1,\ B_2$

$$P\{\xi_1 \in B_1,\ \xi_2 \in B_2\} = P\{\xi_1 \in B_1\}P\{\xi_2 \in B_2\} \qquad (12.5)$$

Очевидно,

$$P\{a_1 \leq \xi_1 < b_1,\ a_2 \leq \xi_2 < b_2\} = F_{\xi_1,\,\xi_2}(b_1,\ b_2) -$$

$$-F_{\xi_1,\,\xi_2}(a_1,\ b_2) - F_{\xi_1,\,\xi_2}(b_1,\ a_2) + F_{\xi_1,\,\xi_2}(a_1,\ a_2) =$$

$$= F_{\xi_1}(b_1)F_{\xi_2}(b_2) - F_{\xi_1}(a_1)F_{\xi_2}(b_2) - F_{\xi_1}(b_1)F_{\xi_2}(a_2) +$$

$$+F_{\xi_1}(a_1)F_{\xi_2}(a_2) = [F_{\xi_1}(b_1) - F_{\xi_1}(a_1)]F_{\xi_2}(b_2) -$$

$$-[F_{\xi_1}(b_1) - F_{\xi_1}(a_1)]F_{\xi_2}(a_2) = [F_{\xi_1}(b_1) - F_{\xi_1}(a_1)]\cdot$$

$$\cdot[F_{\xi_2}(b_2) - F_{\xi_2}(a_2)] = P\{a_1 \leq \xi_1 < b_1\}P\{a_2 \leq \xi_2 < b_2\}.$$

Итак, (12.5) доказано для любых борелевских множеств вида $B_1 = [a_1,\ b_1),\ \ B_2 = [a_2,\ b_2)$. Докажем теперь, что (12.5) выполняется для любых борелевских множеств B_1 вида $B_1 = [a_1,\ b_1)$ и любых борелевских множеств B_2. При фиксированном $B_1 = [a_1,\ b_1)$ как функции множества B_2 обе части являются мерами на σ-алгебре борелевских множеств $\mathcal{B}$. По доказанному эти меры совпадают на полуинтервалах $[a_2,\ b_2)$. Для любого $A = \sum\limits_{i=1}^{n}[a_i,\ b_i)$ из алгебры $\mathcal{A}$ произвольная мера μ определяется своими значениями на полуинтервалах $[a_i,\ b_i)$ с помощью формулы

$$\mu(A) = \sum_{i=1}^{n} \mu([a_i,\ b_i)).$$

Поэтому меры, стоящие в левой и правой частях (12.5), совпадают на алгебре $\mathcal{A}$. По теореме о продолжении меры они совпадут на $\sigma(\mathcal{A}) = \mathcal{B}$, т.е. на σ-алгебре борелевских множеств. Тем самым (12.5) доказано для всех B_1 вида $B_1 = [a_1,\ b_1)$ и любых борелевских B_2.

Теперь фиксируем $B_2 \in \mathcal{B}$. Как функции множества B_1 обе части (12.5) являются мерами на $\mathcal{B}$, причем уже доказано, что они совпадают на множествах вида $B_1 = [a_1,\ b_1)$. Поэтому они совпадают на алгебре $\mathcal{A}$. По теореме о продолжении меры они совпадут и на $\mathcal{B} = \sigma(\mathcal{A})$. Значит, (12.5) будет выполняться для любых борелевских множеств B_1 и B_2.

$\square$

Для установления независимости случайных величин на практике пользуются следующим принципом: если случайные величины не связаны причинно, то они независимы. Этот принцип подтверждается многолетним успешным применением и ни разу не приводил к противоречиям. Но он не является математическим, так как мы не можем дать формального математического определения причинной несвязанности случайных величин. На самом деле определение независимости случайных величин значительно шире их причинной несвязанности, что показывает следующая

Задача 1. Докажите, что если ξ равномерно распределена на отрезке $[0, 1]$, то в ее разложении в двоичную дробь

$$\xi = \frac{\xi_1}{2} + \frac{\xi_2}{2^2} + \frac{\xi_3}{2^3} + \dots$$

случайные величины $\xi_1,\ \xi_2,\ \dots,\ \xi_n$ независимы при любом натуральном n. Но эти случайные величины причинно связаны по своему происхождению.

Если $\xi_1, \ \xi_2, \ \dots, \ \xi_n$ причинно не связаны, то применение указанного принципа ведет к тому, что случайные величины $f_1(\xi_1), \ f_2(\xi_2), \ \dots, \ f_n(\xi_n)$, где $f_1, \ f_2, \ \dots, \ f_n : \mathbb{R} \to \mathbb{R}$, независимы. Действительно, имеет место

Теорема 12.2. *Если случайные величины $\xi_1, \ \xi_2, \ \dots, \ \xi_n$ независимы, $f_1, \ f_2, \ \dots, \ f_n : \mathbb{R} \to \mathbb{R}$ — борелевские функции, то случайные величины $\eta_1 = f_1(\xi_1), \ \eta_2 = f_2(\xi_2), \ \dots, \ \eta_n = f_n(\xi_n)$ независимы.*

Д о к а з а т е л ь с т в о. Борелевость функций — чисто техническое требование, которое обеспечивает то, что функции $\eta_1(\omega), \ \eta_2(\omega), \ \dots, \ \eta_n(\omega)$ являются случайными величинами, т.е. измеримы относительно σ-алгебры $\mathcal{F}$. Используя критерий независимости (необходимость) и определение борелевской функции, получим, что для любых борелевских множеств $B_1, \ B_2, \ \dots, \ B_n$

$$P\{\omega : \ \eta_1(\omega) \in B_1, \ \eta_2(\omega) \in B_2, \ \dots, \ \eta_n(\omega) \in B_n\} =$$

$$= P\{\omega : \ f_1(\xi_1(\omega)) \in B_1, \dots, f_n(\xi_n(\omega)) \in B_n\} =$$

$$= P\{\omega : \ \xi_1(\omega) \in f_1^{-1}(B_1), \dots, \xi_n(\omega) \in f_n^{-1}(B_n)\} =$$

$$= P\{\omega : \ \xi_1(\omega) \in f_1^{-1}(B_1)\} \dots P\{\omega : \ \xi_n(\omega) \in f_n^{-1}(B_n)\} =$$

$$= P\{\omega : \ f_1(\xi_1(\omega)) \in B_1\} \dots P\{\omega : \ f_n(\xi_n(\omega)) \in B_n\} =$$

$$= P\{\omega : \ \eta_1(\omega) \in B_1\} P\{\omega : \ \eta_2(\omega) \in B_2\} \dots$$

$$\dots P\{\omega : \ \eta_n(\omega) \in B_n\}.$$

По критерию независимости (достаточность) это означает, что случайные величины $\eta_1, \ \eta_2, \ \dots, \ \eta_n$ независимы. $\square$

Для абсолютно непрерывного случая самым удобным законом распределения является плотность распределения вероятностей. Поэтому желательно иметь определение независимости, выраженное с помощью плотностей распределения. Если распределение случайного вектора $\xi = (\xi_1,\ \xi_2,\ \dots,\ \xi_n)$ абсолютно непрерывно, то **критерий независимости** случайных величин $\xi_1,\ \xi_2,\ \dots,\ \xi_n$ выглядит следующим образом.

Теорема 12.3. *Если случайные величины $\xi_1,\ \xi_2,\ \dots,\ \xi_n$ абсолютно непрерывны и независимы, то случайный вектор $\xi = (\xi_1,\ \xi_2,\ \dots,\ \xi_n)$ абсолютно непрерывен, причем почти всюду*

$$p_{\xi_1,\xi_2,\dots,\xi_n}(x_1, x_2, \dots, x_n) = p_{\xi_1}(x_1)p_{\xi_2}(x_2)\dots p_{\xi_n}(x_n). \tag{12.6}$$

Обратно, если случайный вектор $\xi = (\xi_1,\ \xi_2,\ \dots,\ \xi_n)$ абсолютно непрерывен, то случайные величины $\xi_1,\ \xi_2,\ \dots,\ \xi_n$ абсолютно непрерывны, а если к тому же почти всюду выполнено (12.6), то они независимы.

Таким образом равенство (12.6) может служить определением независимости абсолютно непрерывных случайных величин.

Доказательство. Пусть случайные величины $\xi_1,\ \xi_2,\ \dots,\ \xi_n$ абсолютно непрерывны и независимы. В силу их абсолютной непрерывности

$$F_{\xi_i}(x_i) = \int\limits_{-\infty}^{x_i} p_{\xi_i}(t_i)dt_i, \quad i = \overline{1, n},$$

а в силу независимости функция их совместного распределения имеет вид

$$F_{\xi_1,\,\xi_2,\,\ldots,\,\xi_n}(x_1,\ x_2,\ \ldots\ ,\ x_n) = F_{\xi_1}(x_1)F_{\xi_2}(x_2)\ldots F_{\xi_n}(x_n) =$$

$$= \int\limits_{-\infty}^{x_1} p_{\xi_1}(t_1)dt_1 \ \int\limits_{-\infty}^{x_2} p_{\xi_2}(t_2)dt_2\ \ldots\ \int\limits_{-\infty}^{x_n} p_{\xi_n}(t_n)dt_n =$$

$$= \int\limits_{-\infty}^{x_1}\int\limits_{-\infty}^{x_2}\ldots\int\limits_{-\infty}^{x_n} p_{\xi_1}(t_1)p_{\xi_2}(t_2)\ldots p_{\xi_n}(t_n)dt_n\ \ldots dt_2\, dt_1.$$

Это означает, что случайный вектор ξ абсолютно непрерывен, а функция $p_{\xi_1}(x_1)p_{\xi_2}(x_2)\ldots p_{\xi_n}(x_n)$ является плотностью его распределения вероятностей, т.е. выполнено (12.6).

Обратно, пусть случайный вектор ξ абсолютно непрерывен. Как показано в конце предыдущей темы, отсюда следует, что его координаты абсолютно непрерывны. Если выполнено (12.6), то

$$F_{\xi_1,\,\xi_2,\,\ldots,\,\xi_n}(x_1,\ x_2,\ \ldots\ ,\ x_n) =$$

$$= \int\limits_{-\infty}^{x_1}\int\limits_{-\infty}^{x_2}\ldots\int\limits_{-\infty}^{x_n} p_{\xi_1,\xi_2,\ldots,\xi_n}(t_1,t_2,\ldots,t_n)dt_n\ \ldots dt_2\, dt_1 =$$

$$= \int\limits_{-\infty}^{x_1}\int\limits_{-\infty}^{x_2}\ldots\int\limits_{-\infty}^{x_n} p_{\xi_1}(t_1)p_{\xi_2}(t_2)\ldots p_{\xi_n}(t_n)dt_n\ \ldots dt_2\, dt_1 =$$

$$= \int\limits_{-\infty}^{x_1} p_{\xi_1}(t_1)dt_1 \ \int\limits_{-\infty}^{x_2} p_{\xi_2}(t_2)dt_2\ \ldots\ \int\limits_{-\infty}^{x_n} p_{\xi_n}(t_n)dt_n =$$

$$= F_{\xi_1}(x_1)F_{\xi_2}(x_2)\ldots F_{\xi_n}(x_n),$$

т.е. $\xi_1,\ \xi_2,\ \dots,\ \xi_n$ независимы. □

В приложениях часто бывает удобным следующее

Следствие 12.4. *Если случайные величины $\xi_1,\ \xi_2,\ \dots,\ \xi_n$ абсолютно непрерывны, не связаны между собой функциональной зависимостью и почти всюду*

$$p_{\xi_1,\xi_2,\dots,\xi_n}(x_1,x_2,\dots,x_n) = f_1(x_1)f_2(x_2)\dots f_n(x_n),$$

где каждая функция $f_i(x_i)$ зависит только от x_i, $i=\overline{1,n}$, то случайные величины $\xi_1,\ \xi_2,\ \dots,\ \xi_n$ независимы, а $f_1(x_1),\ f_2(x_2),\ \dots,\ f_n(x_n)$ совпадают с точностью до постоянных множителей с плотностями распределений $p_{\xi_1}(x_1),\ p_{\xi_2}(x_2),\ \dots,\ p_{\xi_n}(x_n)$ соответственно.

Д о к а з а т е л ь с т в о. Действительно, например, в условиях следствия

$$p_{\xi_1}(x_1) = \int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty}\dots\int\limits_{-\infty}^{\infty} p_{\xi_1,\xi_2,\dots,\xi_n}(x_1,x_2,\dots,x_n)dx_2\,dx_3\,\dots$$

$$\dots\,dx_n = \int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty}\dots\int\limits_{-\infty}^{\infty} f_1(x_1)f_2(x_2)\dots f_n(x_n)dx_2\,dx_3\,\dots$$

$$\dots\,dx_n = C_1f_1(x_1),$$

где

$$C_1 = \int\limits_{-\infty}^{\infty} f_2(x_2)dx_2\dots\int\limits_{-\infty}^{\infty} f_n(x_n)dx_n.$$

Аналогично доказывается, что $p_{\xi_i}(x_i) = C_if_i(x_i)$, $i=\overline{2,n}$. Поэтому

$$p_{\xi_1,\xi_2,\dots,\xi_n}(x_1,x_2,\dots,x_n) = Cp_{\xi_1}(x_1)p_{\xi_2}(x_2)\dots p_{\xi_n}(x_n),$$

где $C^{-1} = C_1 C_2 \dots C_n$. Интегрируя обе части по $\mathbb{R}^n$, находим, что $C = 1$. По теореме 12.3 случайные величины являются независимыми. □

Для дискретных случайных величин критерий независимости выглядит следующим образом.

Теорема 12.5. *Если случайные величины $\xi_1,\ \xi_2,\ \dots,\ \xi_n$ дискретны и независимы, то случайный вектор $\xi = (\xi_1,\ \xi_2,\ \dots,\ \xi_n)$ также является дискретным, причем для любых действительных $x_1,\ x_2,\ \dots,\ x_n$*

$$P\{\xi_1 = x_1,\ \xi_2 = x_2,\ \dots,\ \xi_n = x_n\} =$$

$$= P\{\xi_1 = x_1\} P\{\xi_2 = x_2\} \dots P\{\xi_n = x_n\}. \qquad (12.7)$$

Обратно, если случайный вектор $\xi = (\xi_1,\ \xi_2,\ \dots,\ \xi_n)$ является дискретным, то случайные величины $\xi_1,\ \xi_2,\ \dots,\ \xi_n$ также дискретны, а если к тому же выполнено (12.7), то они независимы.

Доказательство аналогично доказательству теоремы 12.3 с заменой интегралов на соответствующие суммы.

Задача 2. Докажите теорему 12.5.

Замечание 1. Если какое-нибудь x_i не является возможным значением случайной величины ξ_i, то равенство (12.7) выполняется автоматически, поскольку в левой части будет стоять вероятность невозможного события, а в правой части i-й множитель будет выражать вероятность невозможного события. Поэтому в теореме 12.5 для независимости достаточно требовать выполнения (12.7) только для возможных значений случайных величин, а не для всех действительных значений.

Следствие 12.6. *Если случайные величины $\xi_1,\ \xi_2,\ \dots,\ \xi_n$ дискретны, не связаны между собой функциональной зависимостью и для любых значений $x_1, x_2, \dots, x_n$ случайных величин $\xi_1,\ \xi_2,\ \dots,\ \xi_n$ соответственно*

$$P\{\xi_1 = x_1,\ \xi_2 = x_2,\ \dots,\ \xi_n = x_n\} = f_1(x_1) f_2(x_2) \dots f_n(x_n),$$

где каждая функция $f_i(x_i)$ зависит только от x_i, $\quad i = \overline{1,n}$, то случайные величины $\xi_1,\ \xi_2,\ \dots,\ \xi_n$ независимы, а $f_1(x_1),\ f_2(x_2),\ \dots,\ f_n(x_n)$ совпадают с точностью до постоянных множителей с вероятностями $P\{\xi_1 = x_1\},\ P\{\xi_2 = x_2\},\ \dots,\ P\{\xi_n = x_n\}$ соответственно.

Доказательство. Действительно, например,

$$P\{\xi_n = x_n\} = \sum_{x_1} \sum_{x_2} \dots \sum_{x_{n-1}} P\{\xi_1 = x_1,\ \xi_2 = x_2,\ \dots$$

$$\dots,\ \xi_n = x_n\} = C_n f_n(x_n),$$

где $C_n = \sum\limits_{x_1} f_1(x_1) \sum\limits_{x_2} f_2(x_2) \dots \sum\limits_{x_{n-1}} f_{n-1}(x_{n-1})$. Аналогично, $P\{\xi_i = x_i\} = C_i f_i(x_i), \quad i = \overline{1, n-1}$. Поэтому

$$P\{\xi_1 = x_1,\ \xi_2 = x_2,\ \dots,\ \xi_n = x_n\} =$$

$$= C P\{\xi_1 = x_1\} P\{\xi_2 = x_2\} \dots P\{\xi_n = x_n\},$$

где $C^{-1} = C_1 C_2 \dots C_n$. Суммируя обе части по всем возможным значениям $x_1, x_2, \dots, x_n$ случайных величин $\xi_1,\ \xi_2,\ \dots,\ \xi_n$ соответственно, убеждаемся, что $C = 1$. □

Заметим, что в следствиях 12.4 и 12.6 условие, чтобы случайные величины не были связаны функциональной зависимостью, является весьма существенным. Действительно, пусть $X = \{0, 1, \ldots, K\}$, дискретные случайные величины $\xi_1,\ \xi_2,\ \ldots,\ \xi_N$ принимают возможные значения в множестве X, причем для любого $k \in X$ и любого $i = \overline{1, N}$ вероятность $P\{\xi_i = k\} > 0$. Пусть также эти случайные величины связаны линейной функциональной зависимостью $\xi_1 + \xi_2 + \ldots + \xi_N = K$ и известно, что для всех $i_1, i_2, \ldots, i_N \in X$ таких, что $i_1 + i_2 = \ldots + i_N = K$,

$$P\{\xi_1 = i_1,\ \xi_2 = i_2,\ \ldots,\ \xi_N = i_N\} =$$
$$= f_1(i_1)\ f_2(i_2)\ \ldots\ f_N(i_N), \tag{12.8}$$

где функция f_k зависит только от i_k. Так как $\xi_1 + \xi_2 + \ldots + \xi_N = K$, то случайные величины $\xi_1,\ \xi_2,\ \ldots,\ \xi_N$ не являются независимыми. Действительно,

$$P\{\xi_1 = K\} = P\{\xi_1 = K,\ \xi_1 + \xi_2 + \ldots + \xi_N = K\} =$$
$$= P\{\xi_1 = K,\ \xi_2 + \ldots + \xi_N = 0\} =$$
$$= P\{\xi_1 = K,\ \xi_2 = 0,\ \ldots,\ \xi_N = 0\}.$$

Если бы случайные величины были бы независимыми, то из последнего равенства следовало бы, что

$$P\{\xi_1 = K\} = P\{\xi_1 = K\} P\{\xi_2 = 0\} \ldots P\{\xi_N = 0\}.$$

По условию $P\{\xi_1 = K\} > 0$, поэтому

$$P\{\xi_2 = 0\} \ldots P\{\xi_N = 0\} = 0,$$

откуда при некотором i выполнится $P\{\xi_i = 0\} = 0$, что противоречит нашим предположениям.

В чем же причина того, что в случае дополнительной функциональной зависимости между случайными величинами, равенство (12.8) не приводит к независимости? Если бы функциональной зависимости $\xi_1+\xi_2+\ldots+\xi_N = K$ не существовало, то как в доказательстве следствия 2 суммированием (2.7) по всем $i_1,\ldots,i_{k-1},i_{k+1},\ldots,i_N \in X$ мы бы получили, что $P\{\xi_k = i_k\} = C_k f_k(i_k)$. При наличии указанной функциональной зависимости, например,

$$P\{\xi_1 = i_1\} = \sum_{i_2,\ldots,i_N \in X:\; i_1+i_2+\ldots+i_N=K} f_1(i_1)\; f_2(i_2)\; \ldots$$

$$\ldots\; f_N(i_N) = C_1(i_1) f_1(i_1),$$

где

$$C_1(i_1) = \sum_{i_2,\ldots,i_N \in X:\; i_2+\ldots+i_N=K-i_1} f_2(i_2)\; \ldots\; f_N(i_N)$$

уже зависит от i_1 (например, $C_1(i_1) = 0$ при $i_1 > K$, $\quad C_1(K) = f_2(0)\ldots f_N(0)$ и т.д.).

Итак, при наличии указанной зависимости множители в (12.8) существенно отличаются от $P\{\xi_k = i_k\}$, а не просто постоянным множителем, как это было при отсутствии функциональной зависимости.

В заключение дадим определение независимости случайных величин через понятие независимости σ-алгебр.

Определение 12.7. σ-алгебры $\mathcal{F}_1,\mathcal{F}_2,\ldots,\mathcal{F}_n$ называются **независимыми**, если для любых $A_1 \in \mathcal{F}_1,\ A_2 \in \mathcal{F}_2,\ \ldots,\ A_n \in \mathcal{F}_n$ события $A_1,A_2,\ldots,A_n$ независимы.

Приведем без доказательства еще один критерий независимости случайных величин: *для того, чтобы случайные величины* $\xi_1, \xi_2, \dots, \xi_n$ *были независимыми, необходимо и достаточно, чтобы были независимы порожденные ими* σ-алгебры $\mathcal{F}_{\xi_1} = \{\xi_1^{-1}(B), \quad B \in \mathcal{B}\} = \xi_1^{-1}(\mathcal{B})$, $\mathcal{F}_{\xi_2} = \{\xi_2^{-1}(B), \quad B \in \mathcal{B}\} = \xi_2^{-1}(\mathcal{B})$, $\dots$, $\mathcal{F}_{\xi_n} = \{\xi_n^{-1}(B), \ B \in \mathcal{B}\} = \xi_n^{-1}(\mathcal{B})$.

§ 13. Преобразование случайных величин. Свертка распределений

Во многих вопросах возникает необходимость по закону совместного распределения каких-то случайных величин находить закон совместного распределения некоторых функций от них.

Пусть n гладких функций (т.е. имеющих непрерывные частные производные первого порядка по всем переменным)

$$y_1 = f_1(x_1, x_2, \dots, x_n),$$
$$y_2 = f_2(x_1, x_2, \dots, x_n),$$
$$\dots\dots\dots\dots\dots\dots\dots$$
$$y_n = f_n(x_1, x_2, \dots, x_n)$$

задают взаимно-однозначное отображение $f: \mathbb{R}^n \to \mathbb{R}^n$ с якобианом

$$I = Df(x) = det\Big(\frac{\partial f_i}{\partial x_j}\Big)_{i,j=1}^n = \frac{D(y_1, y_2, \dots, y_n)}{D(x_1, x_2, \dots, x_n)} \neq 0.$$

Как известно из математического анализа, в этом случае существует обратное отображение f^{-1} с якобианом $Df^{-1}(y)$, удовлетворяющим равенству

$$Df^{-1}(f(x))Df(x) = 1.$$

Пусть $\xi = (\xi_1, \xi_2, \dots, \xi_n)$ — абсолютно непрерывный случайный вектор и пусть $\eta = f(\xi)$, т.е.

$$\eta_1 = f_1(\xi_1, \xi_2, \dots, \xi_n),$$

$$\eta_2 = f_2(\xi_1, \xi_2, \dots, \xi_n),$$

$$\dots\dots\dots\dots\dots\dots\dots$$

$$\eta_n = f_n(\xi_1, \xi_2, \dots, \xi_n).$$

Пусть B — n-мерное борелевское множество, тогда

$$P\{\eta \in B\} = P\{f(\xi) \in B\} = P\{\xi \in f^{-1}(B)\} =$$

$$= \int\limits_{f^{-1}(B)} \cdots \int p_\xi(x) dx_1 \; dx_2 \; \dots, \; dx_n.$$

Производя в этом кратном интеграле замену переменных $y = f(x)$ или, что то же самое, $x = f^{-1}(y)$, получим:

$$P\{\eta \in B\} = \int\limits_{B} \cdots \int p_\xi(f^{-1}(y)) |Df^{-1}(y)| dy_1 \; dy_2 \; \dots \; dy_n.$$

По определению абсолютно непрерывного вектора это означает, что случайный вектор η абсолютно непрерывен, причем $p_\xi(f^{-1}(y))|Df^{-1}(y)|$ является его плотностью распределения вероятностей. Таким образом, нами доказана следующая

Теорема 13.1. *Если ξ — абсолютно непрерывный вектор с плотностью распределения вероятностей $p_\xi(x)$, f — преобразование $\mathbb{R}^n$ с помощью гладких функций с якобианом $Df(x) \neq 0$, то случайный вектор $\eta =$*

$f(\xi)$ также абсолютно непрерывен, а плотность его распределения вероятностей равна

$$p_\eta(x) = p_\xi(f^{-1}(x))|Df^{-1}(x)|. \qquad (13.1)$$

Отметим, что теорема о плотности вероятности функции от случайной величины (см. (10.12)) является частным случаем данной при $n = 1$.

Пусть известна плотность $p_{\xi,\eta}(x, y)$ совместного распределения вероятностей случайных величин ξ и η. Найдем плотность распределения вероятностей случайной величины $\zeta = \xi + \eta$. Функция ее распределения

$$F_\zeta(x) = P\{\zeta < x\} = P\{\xi + \eta < x\} = \iint\limits_{y+z<x} p_{\xi,\eta}(y, z)dydz.$$

Производя в этом интеграле замену переменных $u = y$, $v = z + y$, получим:

$$F_\zeta(x) = \iint\limits_{v<x} p_{\xi,\eta}(u, v-u)dudv = \int\limits_{-\infty}^{x} \Big[\int\limits_{-\infty}^{+\infty} p_{\xi,\eta}(u, v-u)du\Big]dv.$$

По определению абсолютно непрерывной величины это значит, что ζ абсолютно непрерывна с плотностью распределения вероятностей

$$p_\zeta(x) = \int\limits_{-\infty}^{+\infty} p_{\xi,\eta}(u, x-u)du. \qquad (13.2)$$

Меняя роли ξ и η, аналогично получим:

$$p_\zeta(x) = \int\limits_{-\infty}^{+\infty} p_{\xi,\eta}(x-v, v)dv. \qquad (13.3)$$

Любая из формул (13.2), (13.3) называется **формулой свертки**. Отметим, что ее можно получить с помощью (13.1), примененной к преобразованию $\zeta = \xi + \eta$, $\gamma = \xi$, и формулы, выражающей плотность распределения координат случайного вектора через плотность распределения этого вектора.

Задача 1. Получите формулу свертки этим способом.

Отметим, что для нахождения закона распределения суммы двух случайных величин по формуле свертки в общем случае знания законов распределения этих величин недостаточно. Требуется закон совместного распределения этих величин. Но в важном частном случае, когда случайные величины независимы, закон их совместного распределения полностью определяется законами распределения этих величин. Так, например, для рассматриваемого абсолютно непрерывного случая при условии независимости выполняется

$$p_{\xi,\eta}(x,y) = p_\xi(x)p_\eta(y),$$

так что для независимых случайных величин формула свертки принимает следующий вид:

$$p_\zeta(x) = \int\limits_{-\infty}^{+\infty} p_\xi(u)p_\eta(x-u)du = \int\limits_{-\infty}^{+\infty} p_\xi(x-v)p_\eta(v)dv. \quad (13.4)$$

При этом распределение случайной величины ζ называют **сверткой распределений случайных величин** ξ и η.

Пример 1. Пусть $\xi_1,\ \xi_2,\ \ldots,\ \xi_n$ независимы и имеют показательное распределение с параметром λ, т.е.

$$p_{\xi_i}(x) = \begin{cases} \lambda e^{-\lambda x}, & \text{если } x > 0, \\ 0, & \text{если } x \le 0. \end{cases}$$

По формуле свертки (13.4) для $x > 0$

$$p_{\xi_1+\xi_2}(x) = \int\limits_0^x p_{\xi_1}(u) p_{\xi_2}(x-u)du = \int\limits_0^x \lambda e^{-\lambda u} \lambda e^{-\lambda(x-u)} du =$$

$$= \lambda^2 e^{-\lambda x} \int\limits_0^x du = \lambda^2 x e^{-\lambda x},$$

а для $x \le 0$, очевидно, $p_{\xi_1+\xi_2}(x) = 0$. В силу независимости ξ_1, ξ_2, ξ_3 случайные величины $\xi_1 + \xi_2$ и ξ_3 также будут независимыми (почему?). Поэтому для $x > 0$ по формуле свертки (13.4)

$$p_{\xi_1+\xi_2+\xi_3}(x) = p_{(\xi_1+\xi_2)+\xi_3}(x) =$$

$$= \int\limits_0^x p_{\xi_1+\xi_2}(u) p_{\xi_3}(x-u)du = \int\limits_0^x \lambda^2 u e^{-\lambda u} \lambda e^{-\lambda(x-u)} du =$$

$$= \lambda^3 e^{-\lambda x} \int\limits_0^x u du = \frac{\lambda^3 x^2}{2} e^{-\lambda x},$$

для $x \le 0$, очевидно, $p_{\xi_1+\xi_2+\xi_3}(x) = 0$. По индукции

$$p_{\xi_1+\xi_2+\ldots+\xi_n}(x) = \begin{cases} \frac{\lambda^n x^{n-1}}{(n-1)!} e^{-\lambda x}, & \text{если } x > 0, \\ 0, & \text{если } x \le 0. \end{cases}$$

Определение 13.2. Распределение с плотностью такого вида, т.е. свертка n показательных распределений с параметром λ, называется **распределением Эрланга n-го порядка с параметром λ**.

При рассмотрении данного примера использовано следующее

Утверждение 13.3. *Пусть $\xi_1, \quad \xi_2, \quad \dots, \quad \xi_n$ — независимые случайные величины, $1 \leq k < n$, f, g — борелевские функции соответствующего числа переменных. Тогда случайные величины $f(\xi_1, \xi_2, \dots, \xi_k)$ и $g(\xi_{k+1}, \xi_{k+2}, \dots, \xi_n)$ независимы.*

Задача 2. Докажите это утверждение для абсолютно непрерывных случайных величин.

Задача 3. Докажите, что свертка нормальных распределений с параметрами (a, σ^2) и (b, τ^2) соответственно является нормальным распределением с параметрами $(a + b, \sigma^2 + \tau^2)$.

Как мы убедимся в дальнейшем, утверждение задачи 3 проще получить с помощью характеристических функций, теория которых будет изложена чуть позже и которая более приспособлена для решения подобных задач.

Определение 13.4. Говорят, что случайная величина **имеет гамма-распределение с параметрами** (α, p), если она абсолютно непрерывна и имеет плотность распределения вероятностей вида

$$p(x) = \begin{cases} \frac{\alpha^p}{\Gamma(p)} e^{-\alpha x} x^{p-1}, & \text{если } x > 0, \\ 0, & \text{если } x \leq 0., \end{cases}$$

где $\Gamma(p) = \int\limits_0^{+\infty} t^{p-1} e^{-t} dt$ — гамма-функция Эйлера.

Задача 4. Докажите, что свертка гамма-распределений с параметрами (α, p) и (α, q) соответственно есть гамма-распределение с параметрами $(\alpha, p+q)$.

Пусть $\xi \sim N(0,1)$, тогда для $x > 0$

$$F_{\frac{\xi^2}{2}}(x) = P\left\{\frac{\xi^2}{2} < x\right\} = P\{-\sqrt{2x} < \xi < \sqrt{2x}\} =$$

$$= N(\sqrt{2x}) - N(-\sqrt{2x}) = \Phi(\sqrt{2x}) - \Phi(-\sqrt{2x}) =$$

$$= 2\Phi(\sqrt{2x}) = 2 \int\limits_0^{\sqrt{2x}} \varphi(t) dt.$$

Поэтому плотность распределения вероятностей случайной величины $\frac{\xi^2}{2}$, получающаяся из функции распределения с помощью дифференцирования, для положительных значений x равна

$$p_{\frac{\xi^2}{2}}(x) = 2\varphi(\sqrt{2x})\sqrt{2}\frac{1}{2\sqrt{x}} = \frac{1}{\sqrt{\pi x}} e^{-x} = \frac{1}{\Gamma(1)} x^{\frac{1}{2}-1} e^{-x}.$$

Для отрицательных x она равна нулю. Таким образом, случайная величина $\frac{\xi^2}{2}$ имеет гамма-распределение с параметрами $(1, 1/2)$.

Определение 13.5. Говорят, что случайная величина

$$\chi^2 = \xi_1^2 + \xi_2^2 + \ldots + \xi_n^2,$$

где $\xi_1, \xi_2, \ldots, \xi_n$ — независимые случайные величины, каждая из которых имеет стандартное нормальное распределение, имеет **распределение хи-квадрат с** n **степенями свободы**.

Поэтому распределение $\frac{\chi^2}{2}$ представляет n-кратную свертку с собой гамма-распределения с параметрами $(1, 1/2)$. В силу задачи 4 $\frac{\chi^2}{2}$ имеет гамма-распределение с параметрами $(1, n/2)$, т.е. для $x > 0$

$$p_{\frac{\chi^2}{2}}(x) = \frac{1}{\Gamma(\frac{n}{2})} x^{\frac{n}{2}-1} e^{-x}.$$

Следовательно, плотность распределения хи-квадрат с n степенями свободы имеет вид

$$k_n(x) = p_{\chi^2}(x) = \frac{1}{2} p_{\frac{\chi^2}{2}}\left(\frac{x}{2}\right) =$$

$$= \begin{cases} \frac{1}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})} x^{\frac{n}{2}-1} e^{-\frac{x}{2}}, & \text{если } x > 0, \\ 0, & \text{если } x \leq 0. \end{cases}$$

Формулу свертки можно вывести для распределения суммы независимых дискретных случайных величин. Для простоты выведем формулу свертки для случая целочисленных неотрицательных случайных величин. Пусть ξ и η — целочисленные неотрицательные независимые случайные величины, т.е. их возможные значения лежат в множестве $\{0, 1, \ldots\}$. Очевидно, возможные значения $\zeta = \xi + \eta$ также принадлежат этому множеству, причем

$$P\{\zeta = n\} = P\{\xi + \eta = n\} = \sum_{k=0}^{n} P\{\xi + \eta = n,\ \xi = k\} =$$

$$= \sum_{k=0}^{n} P\{\xi = k, \eta = n - k\} = \sum_{k=0}^{n} P\{\xi = k\} P\{\eta = n - k\}. \tag{13.5}$$

Это и есть формула свертки для целочисленных величин, которая может быть записана также в виде

$$P\{\zeta = n\} = \sum_{l=0}^{n} P\{\xi = n - l\} P\{\eta = l\}. \tag{13.6}$$

Пример 2. Формулы свертки широко применяются в финансовой математике, например, в теории страхования. Простейшая модель функционирования страховой компании — модель индивидуального риска — базируется на следующих предположениях:

1) анализируется короткий промежуток времени, например, один год (т.е. можно не учитывать инфляцию и доход от инвестирования);

2) число договоров страхования N фиксировано и неслучайно;

3) плата за страховку вносится в начале анализируемого периода; никаких поступлений в течение этого периода нет;

4) наблюдаются отдельные договора страхования с известным законом распределения связанных с ними индивидуальных исков $\xi_i, \quad i = \overline{1, N}$.

Поскольку иск измеряется в денежных, т.е. целых единицах, то ξ_i — неотрицательная целочисленная случайная величина, например, ее закон распределения может быть таким:

$$\begin{array}{c||c|c|c||} \xi_i & 0 & b_1 & b_2 \\ \hline P & p & q_1 & q_2 \end{array} \quad . \tag{13.7}$$

$$p + q_1 + q_2 = 1, \; b_1 > b_2,$$

где $\xi_i = 0$, если i-е лицо не умрет в течение года, $\xi_i = b_1$ руб., если лицо умрет в результате несчастного случая, и $\xi_i = b_2$ руб., если лицо умрет от "естественных" причин.

Предположим, что иски от N договоров $\xi_1, \xi_2, \ldots, \xi_N$ — независимые случайные величины (исключаются катастрофические несчастные случаи, влекущие иск сразу по нескольким договорам; в противном случае некоторые из ξ_i будут связаны причинно, поэтому вряд ли будут независимыми).

Фундаментальный интерес для компании представляет вероятность ее разорения. Разорение страховой компании определяется суммарным иском $S = \xi_1 + \xi_2 + \ldots + \xi_N$ к ней. Пусть u — капитал компании. Компания разорится, если суммарный иск S к ней превзойдет ее резервы u. Поэтому вероятность разорения компании равна

$$R = P\{\xi_1 + \xi_2 + \ldots + \xi_N > u\} = \sum_{n=u+1}^{\infty} P\{\xi_1 + \xi_2 + \ldots +$$

$$+\xi_N = n\} = 1 - \sum_{n=0}^{u} P\{\xi_1 + \xi_2 + \ldots + \xi_N = n\},$$

где $P\{\xi_1 + \xi_2 + \ldots + \xi_N = n\}$ можно найти с помощью N-кратной свертки распределения случайной величины

ξ_i (например, распределения (13.7)) с собой с помощью (13.5) или (13.6).

Пример 3. Если (ξ, η) — абсолютно непрерывный случайный вектор, то ξ и η — абсолютно непрерывные случайные величины, следовательно, $P\{\eta = 0\} = 0$. Так как ξ и η определены на одном и том же вероятностном пространстве, то мы можем определить на нем случайную величину $\zeta = \frac{\xi}{\eta}$, задавая ее значения на множестве $\{\omega : \ \eta(\omega) = 0\}$ вероятности нуль произвольным образом, лишь бы сохранялась измеримость относительно σ-алгеюры $\mathcal{F}$. Например, можно считать, что на этом множестве она принимает несобственное значение ∞.

Найдем распределение частного ζ:

$$F_\zeta(x) = P\Big\{\frac{\xi}{\eta} < x\Big\} = \iint\limits_{\frac{u}{v} < x} p_{\xi,\eta}(u, v)dudv.$$

Производя замену переменных $\frac{u}{v} = t, v = s$ в этом интеграле, получим:

$$F_\zeta(x) = \int\limits_{-\infty}^{x} \Big[\int\limits_{-\infty}^{+\infty} p_{\xi,\eta}(st, s)|s|ds\Big]dt.$$

По определению абсолютно непрерывной случайной величины это означает, что ζ абсолютно непрерывна, а плотность ее распределения выражается через плотность совместного распределения ξ и η следующим образом:

$$p_\zeta(x) = \int\limits_{-\infty}^{+\infty} p_{\xi,\eta}(st, s)|s|ds.$$

В частности, если ξ и η независимы, то

$$p_\zeta(x) = \int\limits_{-\infty}^{+\infty} p_\xi(st)p_\eta(s)|s|ds. \tag{13.8}$$

Определение 13.6. Отношение

$$t_n = \frac{\zeta}{\sqrt{\frac{\chi^2}{n}}},$$

в котором числитель и знаменатель — независимые случайные величины, $\zeta \sim N(0,1)$, а χ^2 — случайная величина хи-квадрат с n степенями свободы, называется **отношением Стьюдента с n степенями свободы**, а распределение этого отношения — **распределением Стьюдента с n степенями свободы**.

Пользуясь критерием независимости абсолютно непрерывных величин и формулой (13.8), нетрудно вывести, что плотность распределения Стьюдента с n степенями свободы имеет вид

$$s_n(x) = p_{t_n}(x) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{\pi}\Gamma\left(\frac{n}{2}\right)}\left(1+x^2\right)^{-\frac{n+1}{2}}. \tag{13.9}$$

Задача 5. Выведите формулу (13.9).

Так как $s_n(x)$ — четная функция, то соответствующая функция распределения

$$S_n(x) = \int\limits_{-\infty}^{x} s_n(t)dt$$

обладает следующим свойством:

$$S_n(x) + S_n(-x) = 1.$$

§ 14. Числовые характеристики случайных величин

Полной вероятностной характеристикой случайной величины, позволяющей находить вероятности ее попадания в произвольные борелевские множества, является закон распределения. Установить такой закон на практике зачастую сложно, да и не всегда в приложениях он необходим. Во многих случаях достаточно иметь определенные числовые характеристики, которые характеризуют случайную величину и ее закон распределения лишь частично. Введем наиболее важные из них.

Прежде всего попытаемся определить понятие среднего значения случайной величины. Пусть ξ — дискретная случайная величина с рядом распределения

$$\begin{array}{c||c|c|c|c||} \xi & x_1 & x_2 & \ldots\ldots & x_k \\ \hline P & p_1 & p_2 & \ldots\ldots & p_k \end{array} \quad . \tag{14.1}$$

Предположим, что проведено N независимых измерений величины ξ, в которых значение x_1 появилось N_1 раз, значение x_2 появилось N_2 раз и т.д., наконец, значение x_k появилось N_k раз. Очевидно, $N_1 + N_2 + \ldots + N_k = N$. Среднее арифметическое результатов измерений равно

$$\overline{\xi} = \frac{N_1 x_1 + N_2 x_2 + \ldots + N_k x_k}{N}.$$

Заманчиво принять его за среднее значение случайной величины, но этого сделать нельзя. Действительно, это значение зависит как от числа проведенных измерений N, так и от числа появлений N_i каждого из значений x_i. Мы же хотим, чтобы среднее значение случайной величины определялось только законом распределения и не менялось при разных сериях экспериментов. Если N велико, то, в силу устойчивости относительной частоты, относительная частота события $\{\xi = x_i\}$ стабилизируется около его вероятности:

$$\frac{N_i}{N} \sim p_i.$$

Следовательно,

$$\overline{\xi} = \frac{N_1}{N}x_1 + \frac{N_2}{N}x_2 + \ldots + \frac{N_k}{N}x_k \sim p_1x_1 + p_2x_2 + \ldots + p_kx_k.$$

Итак, среднее арифметическое стабилизируется около числа $\sum\limits_{i=1}^{k} p_ix_i$, которое не зависит от проводимых измерений, а зависит только от закона распределения ξ. Сказанное обосновывает естественность следующего определения.

Определение 14.1. **Математическим ожиданием** или **средним значением** дискретной случайной величины с рядом распределения (14.1) называется число

$$M\xi = \sum_{i=1}^{k} p_ix_i.$$

Из механики известно, что если в точках с координатами $x_1, x_2, \ldots, x_k$ на числовой оси Ox помещены соответственно массы $m_1, m_2, \ldots, m_k$, то координата центра тяжести этой системы материальных точек есть

$$x_c = \frac{\sum\limits_{i=1}^{k} m_i x_i}{\sum\limits_{i=1}^{k} m_i}.$$

Закон распределения (14.1) задает распределение "вероятностной массы", равной 1 (ведь $\sum\limits_{i=1}^{k} p_i = 1$), по точкам с координатами $x_1, x_2, \ldots, x_k$. При этом математическое ожидание совпадает с координатой центра тяжести этой системы точек.

В случае, когда величина принимает счетное число значений, определение 14.1 модифицируем следующим образом.

Определение 14.2. **Математическим ожиданием** или **средним значением** дискретной случайной величины с рядом распределения

ξ	x_1	x_2	x_3	$\ldots\ldots$
P	p_1	p_2	p_3	$\ldots\ldots$

называется число

$$M\xi = \sum_{i=1}^{\infty} p_i x_i,$$

при условии, что последний ряд сходится абсолютно. Если же ряд расходится или сходится условно, то говорят, что **случайная величина ξ не имеет конечного математического ожидания**.

Почему в определении 14.2 требуется абсолютная сходимость? Дело в том, что если занумеровать значения случайной величины по другому, то это будет соответствовать перестановке членов ряда, фигурирующего в этом определении. По теореме Римана члены условно сходящегося ряда можно так поменять местами, что сумма полученного ряда будет равна любому заданному числу, в том числе несобственному (т.е. ∞). Если в определении 14.2 требовать только сходимость, то в зависимости от способа нумерации значений случайной величины мы бы получали разные значения математического ожидания (в том числе и бесконечные) одной и той же случайной величины. Если же ряд сходится абсолютно, то по теореме Коши при перестановке членов ряда его сумма сохраняется. Таким образом, определение 14.2 корректно, поскольку не зависит от способа нумерации значений случайной величины.

Предположим, что $\nu = \nu(\omega)$ — неотрицательная целочисленная случайная величина (т.е. возможные ее значения принадлежат множеству $\{0, 1, \dots\}$). Тогда

$$M\nu = \sum_{n=1}^{\infty} P\{\mu \geq n\}.$$

Действительно,

$$M\nu = \sum_{k=0}^{\infty} kP\{\mu = k\} = \sum_{k=1}^{\infty} kP\{\mu = k\} =$$

$$= \sum_{k=1}^{\infty} \sum_{n=1}^{k} P\{\mu = k\} = \sum_{n=1}^{\infty} \sum_{k=n}^{\infty} P\{\mu = k\} = \sum_{n=1}^{\infty} P\{\mu \geq n\}. \tag{14.2}$$

Изменение порядка суммирования законно в силу неотрицательности слагаемых, входящих в суммы рассмотренных выше рядов.

Задача 1. Покажите, что если μ — целочисленная случайная величина, принимающая значения в множестве $\{\dots, -2, -1, 0, 1, 2, \dots\}$, то

$$M\nu = \sum_{n=1}^{\infty} P\{|\mu| \geq n\}. \qquad (14.3)$$

Пусть теперь ξ — абсолютно непрерывная случайная величина с плотностью распределения $p(x)$. Теперь "вероятностная масса" размазана по числовой прямой (вероятность попадания в каждую точку равна нулю, а вероятность попадания в промежуток $[x, x + dx)$ приближенно равна $p(x)dx$). Поэтому в определении 14.2 значение x_i надо заменить на произвольное значение $x \in \mathbb{R}$, вероятность $p_i = P\{\xi = x_i\}$ — на элемент вероятности $p(x)dx$, а операцию суммирования — на операцию интегрирования. Это обосновывает разумность следующего определения.

Определение 14.3. **Математическим ожиданием** или **средним значением** абсолютно непрерывной случайной величины с плотностью распределения $p(x)$ называется число

$$M\xi = \int_{-\infty}^{+\infty} xp(x)dx,$$

при условии, что последний интеграл абсолютно сходится. Если интеграл расходится или условно сходится, то говорят, что **случайная величина ξ не имеет конечного математического ожидания**.

Задача 2. Покажите, что для неотрицательной абсолютно непрерывной величины

$$M\xi = \int_0^\infty [1 - F(x)]dx, \tag{14.4}$$

а для абсолютно непрерывной величины

$$M\xi = -\int_{-\infty}^0 F(x)dx + \int_0^\infty [1 - F(x)]dx. \tag{14.5}$$

Сравните (14.2), (14.3) с (14.4), (14.5).

Позже мы выведем (14.4) для произвольной случайной величины.

В теории надежности рассматривают время жизни некоторого элемента (например, электролампочки), в теории страхования — время жизни человека. Это — неотрицательная случайная величина ξ с функцией распределения $F(x)$. В качестве основной характеристики в этих теориях берется не функция распределения, а дополнительная функция $Q(x) = 1 - F(x) = P\{\xi \geq x\}$, называемая в теории надежности функцией надежности, а в теории страхования — функцией выживания. В силу (14.4)

$$M\xi = \int_0^\infty Q(x)dx.$$

Примеры.

1. Неслучайная величина C может рассматриваться как случайная с рядом распределения

ξ	C
P	1

.

Поэтому ее математическое ожидание $MC = C$.

2. Для случайной величины с распределением Бернулли

ξ	1	0
P	p	q

математическое ожидание равно $M\xi = p \cdot 1 + q \cdot 0 = p$.

3. Пусть μ имеет биномиальное распределение с параметрами n, p. Согласно определению 14.1

$$M\mu = \sum_{k=1}^{n} kC_n^k p^k q^{n-k}.$$

Используя следующее свойство сочетаний:

$$kC_n^k = k\frac{n!}{k!(n-k)!} = n\frac{(n-1)!}{(k-1)!(n-k)!} = nC_{n-1}^{k-1}$$

и формулу бинома Ньютона, получим

$$M\mu = n\sum_{k=1}^{n} C_{n-1}^{k-1} p^k q^{n-k} = np\sum_{l=0}^{n-1} C_{n-1}^{l} p^l q^{n-1-l} =$$

$$= np(p+q)^{n-1} = np.$$

4. Пусть ξ имеет распределение Пуассона с параметром λ. Тогда

$$M\xi = \sum_{k=0}^{\infty} k\frac{\lambda^k}{k!}e^{-\lambda} = \sum_{k=1}^{\infty} k\frac{\lambda^k}{k!}e^{-\lambda} = \lambda e^{-\lambda}\sum_{k=1}^{\infty}\frac{\lambda^{k-1}}{(k-1)!} =$$

$$= \lambda e^{-\lambda}e^{\lambda} = \lambda.$$

Таким образом, параметр λ в распределении Пуассона играет роль математического ожидания.

5. Для равномерно распределенной на отрезке $[a,b]$ случайной величины, которая равновероятно принимает значения на промежутках одинаковой длины, где бы они ни находились внутри $[a,b]$, интуиция подсказывает, что среднее значение должно находиться в середине отрезка $[a,b]$. Действительно,

$$M\xi = \int_{-\infty}^{+\infty} xp(x)dx = \int_{a}^{b} x\,\frac{1}{b-a}\,dx = \frac{a+b}{2}.$$

6. Если ξ имеет показательное распределение с параметром λ, то

$$M\xi = \int_{-\infty}^{+\infty} xp(x)dx = \lambda\int_{0}^{+\infty} xe^{-\lambda x}dx = \frac{1}{\lambda}.$$

7. Если $\xi \sim N(0,1)$, то в силу четности плотности стандартного нормального распределения $\varphi(x) = \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$ интеграл от нечетной функции в симметричных пределах

$$M\xi = \int_{-\infty}^{+\infty} x\varphi(x)dx = 0.$$

8. Покажем, что прием, примененный в предыдущем примере, требует осторожности. Пусть ξ распределена по закону Коши, т.е. $p(x) = \frac{1}{\pi(1+x^2)}$. Если рассуждать так же, то, в силу четности $p(x)$,

$$M\xi = \int\limits_{-\infty}^{+\infty} xp(x)dx = 0.$$

Это рассуждение неверно, поскольку рассматриваемый интеграл расходится (проверьте это). Оно верно только по отношению к главному значению интеграла, т.е.

$$V.p. \int\limits_{-\infty}^{+\infty} xp(x)dx = \lim_{A\to+\infty} \int\limits_{-A}^{A} xp(x)dx = 0.$$

Согласно определению 14.2 случайная величина ξ не имеет математического ожидания. Рассуждения предыдущего примера правомочны, поскольку рассматриваемый в нем интеграл сходится абсолютно.

Задача 3. Найдите математическое ожидание случайной величины, имеющей

а) геометрическое распределение;

б) гипергеометрическое распределение;

в) нормальное распределение с параметрами (a, σ^2);

г) гамма-распределение.

Перейдем к общему определению математического ожидания. Пусть $\xi = \xi(\omega)$ — случайная величина, определенная на вероятностном пространстве $(\Omega, \mathcal{F}, P)$.

Определение 14.4. **Математическим ожиданием** или **средним значением** случайной величины ξ называется число

$$M\xi = \int\limits_{\Omega} \xi(\omega) dP = \int\limits_{\Omega} \xi(\omega) dP(\omega) = \int\limits_{\Omega} \xi(\omega) P(d\omega).$$

Здесь использованы три различных обозначения для интеграла Лебега от измеримой относительно σ-алгебры $\mathcal{F}$ функции $\xi(\omega)$ по мере P.

Замечание 1. Напомним определение интеграла Лебега. Пусть ξ — дискретная случайная величина с конечным множеством возможных значений $\{x_1, x_2, \dots, x_k\}$, $A_l = \{\omega : \xi(\omega) = x_l\}$. Очевидно, $A_1 + A_2 + \ldots + A_k = \Omega$. Если

$$I_A = I_A(\omega) = \begin{cases} 1, & \text{если } \omega \in A, \\ 0, & \text{если } \omega \notin A \end{cases}$$

— индикатор множества A (события A), то случайная величина ξ может быть записана в виде

$$\xi(\omega) = \sum_{l=1}^{k} x_l I_{A_l}(\omega)$$

и по терминологии теории меры называется простой функцией. Интеграл Лебега от простой функции определяется как

$$\int\limits_{\Omega} \xi(\omega) dP = \sum_{l=1}^{k} x_l P(A_l).$$

При этом доказывается, что это определение корректно, т.е. не зависит от способа представления простой функции в виде линейной комбинации индикаторных функций. Таким образом, определение интеграла от простой функции совпадает с определением 14.1 математического ожидания от дискретной случайной величины с конечным множеством возможных значений.

Далее интеграл Лебега определяется для неотрицательных функций $\xi = \xi(\omega)$, измеримых относительно $\mathcal{F}$ (неотрицательных случайных величин в нашей терминологии). Для любой такой функции существует неубывающая последовательность неотрицательных простых функций (ξ_n) такая, что $\xi_n \to \xi$ при $n \to \infty$, например, можно взять

$$\xi_n(\omega) = f_n(\xi(\omega)) = \sum_{l=1}^{n2^n} \frac{l-1}{2^n} I_{\{\omega\colon \frac{l-1}{2^n} \le \xi(\omega) < \frac{l}{2^n}\}} + nI_{\{\omega\colon \xi(\omega) \ge n\}}. \tag{14.6}$$

Так как

$$\int\limits_{\Omega} \xi_n(\omega) dP \le \int\limits_{\Omega} \xi_{n+1}(\omega) dP,$$

то существует конечный или бесконечный предел такой последовательности интегралов. По определению интегралом Лебега от неотрицательной измеримой относительно $\mathcal{F}$ функции называется

$$\int\limits_{\Omega} \xi(\omega) dP = \lim_{n\to\infty} \int\limits_{\Omega} \xi_n(\omega) dP,$$

причем доказывается, что этот предел не зависит от способа выбора неубывающей последовательности простых функций, сходящихся к ξ. В случае, когда этот предел бесконечен, в зависимости от удобства считают, что интеграл Лебега не существует или что он равен $+\infty$.

Отметим, что данному определению интеграла Лебега от неотрицательной измеримой функции эквивалентно следующее определение:

$$\int\limits_{\Omega} \xi(\omega) dP = \sup_{\{\eta \le \xi\}} \int\limits_{\Omega} \eta(\omega) dP,$$

где точная верхняя грань берется по множеству простых неотрицательных измеримых функций, не превосходящих ξ.

Наконец, если $\xi(\omega)$ — произвольная функция, измеримая относительно F (произвольная случайная величина в нашей терминологии), то по определению

$$\int\limits_{\Omega} \xi(\omega) dP = \int\limits_{\Omega} \xi^+(\omega) dP - \int\limits_{\Omega} \xi^-(\omega) dP,$$

где $\xi^+ = \max(\xi, 0)$, $\xi^- = -\min(\xi, 0)$. В зависимости от удобства считают, что либо а) если хотя бы один из интегралов от ξ^+ или от ξ^- бесконечен, то интеграл Лебега от ξ не существует, либо б) если оба интеграла бесконечны, то интеграл Лебега от ξ не существует; если первый интеграл бесконечен, а второй конечен, то интеграл Лебега от ξ равен $+\infty$; если первый интеграл конечен, а второй бесконечен, то интеграл Лебега от ξ равен $-\infty$.

Отметим, что конечный интеграл Лебега от измеримой функции $\xi(\omega)$ существует тогда и только тогда, когда существует конечный интеграл Лебега от $|\xi(\omega)|$. Применительно к нашему случаю конечное математическое ожидание $M\xi$ существует тогда и только тогда, когда существует конечное математическое ожидание

$M|\xi|$. А то, что случайная величина ξ имеет конечное математическое ожидание, означает, что $\xi(\omega) \in L_1(\Omega, \mathcal{F}, P)$.

Напомним те свойства интеграла, которые понадобятся в дальнейшем. Для сокращения записи используем обозначение $M\xi$. В силу общего определения математического ожидания свойства интеграла Лебега от случайной величины по вероятностной мере P — свойства математического ожидания.

Свойства математического ожидания.

1. *Математическое ожидание неслучайной величины равно этой неслучайной величине*: $MC = C$.

2. *Линейность:* $M(a\xi + b\eta) = aM\xi + bM\eta$, *где* a, b — *постоянные. В частности, постоянную (неслучайную) величину можно выносить за знак математического ожидания.*

3. *Монотонность: если* $\xi \geq 0$, *то* $M\xi \geq 0$.

Очевидными следствиями свойств 1,2 являются: а) если $\xi \leq \eta$, то $M\xi \leq M\eta$; б) $|M\xi| \leq M|\xi|$.

Остальные свойства не вытекают непосредственно из элементарных свойств интеграла Лебега и требуют дополнительных рассуждений.

4. **Неравенство Чебышева (для математического ожидания)**. *Если случайная величина* ξ *имеет математическое ожидание, то для любого* $\varepsilon > 0$

$$P\{|\xi| \geq \varepsilon\} \leq \frac{M|\xi|}{\varepsilon}.$$

Д о к а з а т е л ь с т в о. Разобьем пространство Ω на две непересекающиеся части:

$$\Omega = \{\omega : \ |\xi(\omega)| \geq \varepsilon\} + \{\omega : \ |\xi(\omega)| < \varepsilon\}.$$

В силу свойств конечной аддитивности как функции множества интегрирования, монотонности и свойства

$$\int\limits_A dP = P(A)$$

интеграла Лебега получим:

$$M|\xi| = \int\limits_\Omega |\xi(\omega)| dP = \int\limits_{\{\omega:\ |\xi(\omega)| \geq \varepsilon\}} |\xi(\omega)| dP +$$

$$+ \int\limits_{\{\omega:\ |\xi(\omega)| < \varepsilon\}} |\xi(\omega)| dP \geq \int\limits_{\{\omega:\ |\xi(\omega)| \geq \varepsilon\}} |\xi(\omega)| dP \geq$$

$$\geq \varepsilon \int\limits_{\{\omega:\ |\xi(\omega)| \geq \varepsilon\}} dP = \varepsilon P\{\omega : \ |\xi(\omega)| \geq \varepsilon\}.$$

Так как $\varepsilon > 0$, то неравенство Чебышева следует из последнего неравенства. □

5. *Если* $M|\xi| = 0$*, то* $\xi = 0$ *с вероятностью* 1.

Д о к а з а т е л ь с т в о. Из неравенства Чебышева следует, что для любого $\varepsilon > 0$ выполняется $P\{|\xi| \geq \varepsilon\} = 0$. Поэтому $P\{|\xi| < \varepsilon\} = 1$. Следовательно, $P\{|\xi| \leq 0\} = 1$. Значит, $|\xi| = 0$ с вероятностью 1. □

Замечание 2. Внимательный читатель обратил внимание на то, что переход "для любого $\varepsilon > 0$ выполняется

$P\{|\xi| < \varepsilon\} = 1$" влечет "$P\{|\xi| \leq 0\} = 1$" сделан на интуитивном уровне и требует доказательства. Так как $P\{|\xi| < \varepsilon\} = 1$ для любого $\varepsilon > 0$, то, в частности, для любого натурального n будет $P\{|\xi| < \frac{1}{n}\} = 1$. Но $\{\omega : \ |\xi(\omega)| < \frac{1}{n}\} \downarrow \{\omega : \ |\xi(\omega)| = 0\}$ (проверьте!). По свойству непрерывности вероятности отсюда следует, что $P\{\omega : \ |\xi(\omega)| < \frac{1}{n}\} \to P\{\omega : \ |\xi(\omega)| = 0\}$. С другой стороны, предел последовательности единиц равен единице, поэтому $P\{\omega : \ |\xi(\omega)| = 0\} = 1$. $\square$

Следствие 14.5. *Если $\xi \geq 0$ и $M\xi = 0$, то $\xi = 0$ с вероятностью 1.*

6. **Неравенство Иенсена**. *Пусть $g(x)$ — выпуклая вниз борелевская функция и $M|\xi| < \infty$. Тогда*

$$g(M\xi) \leq Mg(\xi). \tag{14.7}$$

Д о к а з а т е л ь с т в о. Известно, что если $g(x)$ — выпуклая вниз функция, то для любого $x_0 \in \mathbb{R}$ найдется $c = c(x_0)$ такое, что для любого $x \in \mathbb{R}$

$$g(x) \geq g(x_0) + c(x - x_0),$$

т.е. через любую точку $(x_0, g(x_0))$ на графике $g(x)$ можно провести прямую так, чтобы график $g(x)$ лежал выше этой прямой. При этом никаких ограничений на гладкость $g(x)$ не надо накладывать. Но если функция дифференцируема, то в качестве указанной прямой можно взять касательную. Полагая в последнем неравенстве $x = \xi$ и $x_0 = M\xi$, получим

$$g(\xi) \geq g(M\xi) + c(\xi - M\xi).$$

По следствию а) свойства 3 $Mg(\xi) \geq g(M\xi)$.

□

7. **Нераверство Ляпунова**. *Если* $0 < s < t$, *то*

$$(M|\xi|^s)^{\frac{1}{s}} \leq (M|\xi|^t)^{\frac{1}{t}}.$$

Доказательство. Полагая $r = t/s, \quad \eta = |\xi|^s$ и применяя неравенство Иенсена к функции $g(x) = |x|^r$, получим $|M\eta|^r \leq M|\eta|^r$, т.е.

$$(M|\xi|^s)^{r/s} \leq M|\xi|^t,$$

что доказывает неравенство Ляпунова. □

Замечание 3. В функциональном анализе через $L_t(\Omega, \mathcal{F}, P)$ обозначается линейное нормированное пространство функций $f(\omega)$, таких, что $f^t(\omega)$ интегрируема по Лебегу по мере P. Норма функции в нем определяется как

$$||f|| = \Big[\int\limits_\Omega |f(\omega)|^t dP\Big]^{\frac{1}{t}}.$$

В этих обозначениях неравенство Ляпунова принимает вид

$$||\xi||_{L_s} \leq ||\xi||_{L_t} \qquad \text{при } \; s \leq t.$$

Из неравенства Ляпунова следует, что если $s < t$, то $L_s \subset L_t$.

8. *Если* ξ *и* η *— независимые случайные величины с конечными математическими ожиданиями, то случайная величина* $\xi\eta$ *имеет конечное математическое ожидание, причем*

$$M(\xi\eta) = M\xi \cdot M\eta.$$

Д о к а з а т е л ь с т в о. Докажем это свойство для простых независимых случайных величин

$$\xi(\omega) = \sum_{i=1}^{k} x_i I_{\{\omega:\,\xi(\omega)=x_i\}}(\omega), \quad \eta(\omega) = \sum_{j=1}^{l} y_j I_{\{\omega:\,\eta(\omega)=y_j\}}(\omega).$$

Тогда $\xi\eta$ — также простая случайная величина:

$$\xi\eta = \sum_{i=1}^{k}\sum_{j=1}^{l} x_i y_j I_{\{\omega:\,\xi(\omega)=x_i,\,\eta(\omega)=y_j\}}.$$

По определению интеграла Лебега от простой функции

$$M(\xi\eta) = \sum_{i=1}^{k}\sum_{j=1}^{l} x_i y_j P\{\omega:\ \xi(\omega)=x_i,\ \eta(\omega)=y_j\}.$$

По критерию независимости дискретных случайных величин при любых $x_i,\ y_j$

$$P\{\omega:\ \xi(\omega)=x_i,\ \eta(\omega)=y_j\} =$$

$$= P\{\omega:\ \xi(\omega)=x_i\}P\{\omega:\ \eta(\omega)=y_j\}.$$

Поэтому

$$M(\xi\eta) = \sum_{i=1}^{k}\sum_{j=1}^{l} x_i y_j P\{\omega:\ \xi(\omega)=x_i\}P\{\omega:\ \eta(\omega)=y_j\} =$$

$$= \sum_{i=1}^{k} x_i P\{\omega:\ \xi(\omega)=x_i\} \sum_{j=1}^{l} y_j P\{\omega:\ \eta(\omega)=y_j\} = M\xi{\cdot}M\eta.$$

Пусть теперь ξ и η — неотрицательные независимые случайные величины. Для них существуют неубывающие последовательности простых случайных величин $\xi_n = f_n(\xi)$ и $\eta_n = f_n(\eta)$ такие, что $\xi_n \to \xi$ и $\eta_n \to \eta$ при $n \to \infty$, где борелевские функции $f_n(x)$ определены с помощью (14.6). Так как ξ и η независимы, то при любых n и m случайные величины $\xi_n = f_n(\xi)$ и $\eta_m = f_m(\eta)$ независимы как борелевские функции от независимых случайных величин. Поскольку $(\xi_n\eta_n)$ — неубывающая последовательность простых случайных величин, причем $\xi_n\eta_n \to \xi\eta$, то по определению интеграла Лебега от неотрицательных измеримых функций

$$M(\xi\eta) = \lim_{n\to\infty} M(\xi_n\eta_n) = \lim_{n\to\infty} M\xi_n \cdot M\eta_n = M\xi \cdot M\eta.$$

Наконец, если ξ и η — любые независимые случайные величины, то $\xi^+ = f(\xi)$, $\xi^- = g(\xi)$, $\eta^+ = f(\eta)$, $\eta^- = g(\eta)$, где борелевские функции f и g определены равенствами $f(x) = \max(x, 0)$, $g(x) = -\min(x, 0)$. Поэтому будут независимыми между собой неотрицательные случайные величины в каждой из пар (ξ^+, η^+), (ξ^+, η^-), (ξ^-, η^+), (ξ^-, η^-). Следовательно,

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

$$-M(\xi^-\eta^+) + M(\xi^-\eta^-) = M\xi^+ \cdot M\eta^+ - M\xi^+ \cdot M\eta^- -$$

$$-M\xi^- \cdot M\eta^+ + M\xi^- \cdot M\eta^- = M(\xi^+ - \xi^-) \cdot M(\eta^+ - \eta^-) =$$

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

□

В курсе функционального анализа доказывается следующая

Теорема о замене переменной в интеграле Лебега.

Пусть $\xi = \xi(\omega)$ — $\mathcal{F}/\mathcal{B}$-измеримое отображение вероятностного пространства $(\Omega, \mathcal{F}, P)$ в измеримое пространство $(X, \mathcal{B})$, P_ξ — вероятностная мера на $(X, \mathcal{B})$, индуцируемая $\xi = \xi(\omega)$, т.е.

$$P_\xi(B) = P\{\xi^{-1}(B)\} = P\{\omega : \ \xi(\omega) \in B\}, \qquad B \in \mathcal{B}.$$

Тогда для любой $\mathcal{B}$-измеримой функции $f = f(x)$, $x \in X$

$$\int\limits_B f(x) dP_\xi = \int\limits_{\xi^{-1}(B)} f(\xi(\omega)) dP, \qquad B \in \mathcal{B}, \tag{14.8}$$

в предположении, что какой-то один из этих двух интегралов существует.

1. В частности, если ξ — случайная величина, определенная на вероятностном пространстве $(\Omega, \mathcal{F}, P)$, то она задает указанное в теореме отображение в измеримое пространство $(R, \mathcal{B})$, где $\mathcal{B}$ — σ-алгебра борелевских множеств $\mathbb{R}$, а индуцированная на нем мера совпадает с распределением P_ξ этой случайной величины. Поэтому для любой борелевской функции $f(x)$ и любого борелевского множества B выполняется (14.8). Беря $B = \mathbb{R}$ и учитывая, что $\xi^{-1}(\mathbb{R}) = \Omega$, получим

$$\int\limits_\Omega f(\xi(\omega)) dP = \int\limits_{\mathbb{R}} f(x) dP_\xi.$$

Поскольку распределение P_ξ — это мера, которая порождается функцией распределения $F(x)$ случайной величины ξ, то в обозначении этого интеграла принято вместо меры писать эту функцию и называть его интегралом

Лебега-Стилтьеса. Таким образом,

$$\int\limits_{\Omega} f(\xi(\omega))dP = \int\limits_{-\infty}^{+\infty} f(x)dF(x). \tag{14.9}$$

Мы будем говорить, что (14.9) задает замену переменной $\xi(\omega) = x$.

Если интеграл в правой части (14.9) абсолютно сходится как несобственный интеграл Римана-Стилтьеса, то, как известно из теории меры, он существует как интеграл Лебега-Стилтьеса, причем значения этих интегралов совпадают.

2. В другом частном случае, если ξ — случайный n-мерный вектор, определенный на вероятностном пространстве $(\Omega, \mathcal{F}, P)$, то он задает указанное в теореме отображение в измеримое пространство $(\mathbb{R}^n, \mathcal{B}^n)$, где $\mathcal{B}^n$ — σ-алгебра борелевских множеств $\mathbb{R}^n$, а индуцированная на нем мера совпадает с распределением $P_\xi = P_{\xi_1,\xi_2,\dots,\xi_n}$ этого случайного вектора. Поэтому для любой борелевской функции $f(x_1, x_2, \dots, x_n)$ и любого n-мерного борелевского множества B выполняется (14.8). Беря $B = \mathbb{R}^n$ и учитывая, что $\xi^{-1}(\mathbb{R}^n) = \Omega$, получим

$$\int\limits_{\Omega} f(\xi_1(\omega), \xi_2(\omega), \dots, \xi_n(\omega))dP =$$

$$= \int\limits_{\mathbb{R}^n} f(x_1, x_2, \dots, x_n)dP_{\xi_1,\xi_2,\dots,\xi_n}. \tag{14.10}$$

Мы будем говорить, что (14.10) задает замену переменных $\xi_1(\omega) = x_1, \xi_2(\omega) = x_2, \dots, \xi_n(\omega) = x_n$.

Как соотносится общее определение математического ожидания как интеграла Лебега с теми частными определениями, с которых начиналось наше изложение? Выше мы уже видели, что для дискретной случайной величины с конечным множеством значений определение математического ожидания как интеграла Лебега совпало с определением 14.1. Если ξ — дискретная случайная величина со счетным множеством значений, то ее можно записать в виде

$$\xi = \sum_{i=1}^{\infty} x_i I_{\{\omega:\, \xi(\omega)=x_i\}}(\omega).$$

При этом

$$|\xi| = \sum_{i=1}^{\infty} |x_i| I_{\{\omega:\, \xi(\omega)=x_i\}}(\omega).$$

Математическое ожидание как интеграл Лебега существует тогда и только тогда, когда существует интеграл от модуля ξ. Но последний интеграл существует тогда и только тогда, когда сходится ряд

$$\sum_{i=1}^{\infty} |x_i| P\{\omega :\ \xi(\omega) = x_i\}.$$

По свойствам интеграла Лебега

$$M\xi = \int\limits_{\Omega} \sum_{i=1}^{\infty} x_i I_{\{\omega:\, \xi(\omega)=x_i\}}(\omega) dP =$$

$$= \sum_{i=1}^{\infty} x_i P\{\omega :\ \xi(\omega) = x_i\} = \sum_{i=1}^{\infty} x_i p_i,$$

если последний ряд абсолютно сходится. Это совпадает с определением 14.2.

Если же ξ абсолютно непрерывна с плотностью распределения $p(x)$, то, производя в интеграле Лебега $M\xi$ замену $\xi(\omega) = x$, получим (см. (14.9))

$$\int_{\Omega} \xi(\omega) dP = \int_{-\infty}^{+\infty} x dF(x). \tag{14.11}$$

Так как почти всюду $dF(x) = p(x)dx$, то

$$M\xi = \int_{-\infty}^{+\infty} x dF(x) = \int_{-\infty}^{+\infty} x p(x) dx.$$

Предположим, что выполнены условия существования математического ожидания в определении 14.3, т.е. последний интеграл предполагался существующим как абсолютно сходящийся несобственный интеграл Римана. Тогда (14.11) существует как абсолютно сходящийся интеграл Римана-Стилтьеса, значение которого совпадает со значением интеграла Лебега-Стилтьеса. Итак, в предположении абсолютной сходимости получилось определение 14.3. Заметим, что общее определение даже в абсолютно непрерывном случае позволяет несколько расширить определение 14.3, если интеграл в этом определении рассматривать как интеграл Лебега по мере Лебега на прямой. Действительно, рассмотрим плотность вероятности следующего вида:

$$p(x) = \begin{cases} 1, & \text{если } x \in [0,1] \text{ и } x \text{ — иррационально}, \\ 0, & \text{в противном случае.} \end{cases}$$

Очевидно, $p(x) \geq 0$, и поскольку $p(x)$ почти всюду совпадает с плотностью равномерного распределения на от-

резке $[0, 1]$, то
$$\int\limits_{-\infty}^{+\infty} p(x)dx = 1,$$
т.е. $p(x)$ действительно является плотностью распределения некоторой абсолютно непрерывной случайной величины ξ. При этом
$$M\xi = \int\limits_{-\infty}^{+\infty} xp(x)dx = \int\limits_{0}^{1} xp(x)dx = \int\limits_{0}^{1} xdx = 1/2,$$
если рассматривать этот интеграл как интеграл Лебега, поскольку почти всюду $xp(x) = x$ на $[0, 1]$. В то же время последний интеграл является расходящимся как несобственный интеграл Римана (например, по критерию Лебега интегрируемости по Риману). Таким образом, по определению 14.3 математическое ожидание не существует, а по общему определению математического ожидания оно равно 1/2, что больше согласуется с интуицией (поскольку плотность вероятности почти всюду совпадает с плотностью равномерного распределения, то математическое ожидание рассматриваемой величины должно совпадать с математическим ожиданием равномерно распределенной величины).

Итак, частные определения математического ожидания являются частными случаямми общего определения математического ожидания.

Из (14.9) следует, что математическое ожидание может быть определено как интеграл Стилтьеса
$$M\xi = \int\limits_{-\infty}^{+\infty} xdF_{\xi}(x).$$

Пусть ξ — неотрицательная случайная величина. Так как $F_\xi(x) = 0$ для $x \leq 0$, то

$$M\xi = \int\limits_0^{+\infty} x dF_\xi(x). \tag{14.12}$$

Поскольку этот интеграл сходится, то остаток $\int\limits_A^{\infty} x dF_\xi(x) \to 0$ при $A \to +\infty$. Так как

$$\int\limits_A^{\infty} x dF_\xi(x) \geq A \int\limits_A^{\infty} dF_\xi(x) = A(1 - F_\xi(A)) \geq 0,$$

то $A(1 - F_\xi(A)) \to 0$ при $A \to +\infty$. Интегрируя (14.12) по частям, получим

$$M\xi = -\int\limits_0^{+\infty} x d[1 - F_\xi(x)] = -x[1 - F_\xi(x)]\Big|_0^{\infty} + \int\limits_0^{+\infty} [1 - F_\xi(x)] dx.$$

Следовательно,

$$M\xi = \int\limits_0^{+\infty} [1 - F_\xi(x)] dx.$$

Задача 4. Покажите, что для любой случайной величины, имеющей математическое ожидание,

$$M\xi = -\int\limits_{-\infty}^{0} F_\xi(x) dx + \int\limits_0^{\infty} [1 - F_\xi(x)] dx.$$

Сформулируем результаты (14.9) и (14.10) в виде теорем.

Теорема о математическом ожидании функции от случайной величины.

Если $f(x)$ — борелевская функция, то для случайной величины $\eta = f(\xi)$

$$M\eta = Mf(\xi) = \int\limits_{-\infty}^{+\infty} f(x)dF(x),$$

где $F(x)$ — функция распределения ξ. В частности, если ξ — дискретная случайная величина, принимающая значения x_i с вероятностями p_i, то

$$Mf(\xi) = \sum_{x_i} f(x_i)p_i,$$

а если ξ — абсолютно непрерывнвя случайная величина с плотностью вероятности $p(x)$, то

$$Mf(\xi) = \int\limits_{-\infty}^{\infty} f(x)p(x)dx.$$

Отметим, что формула замены переменной в интеграле Лебега позволила нам вывести формулы, поэволяющие подсчитывать математическое ожидание $f(\xi)$ по закону распределения ξ без предварительного нахождения закона распределения $f(\xi)$.

Теорема о математическом ожидании функции от нескольких случайных величин.

Если $f(x_1, x_2, \dots, x_n)$ — борелевская функция, то для случайной величины $\eta = f(\xi_1, \xi_2, \dots, \xi_n)$

$$M\eta = Mf(\xi_1, \xi_2, \dots, \xi_n) = \int\limits_{R^n} f(x_1, x_2, \dots, x_n)dP_{\xi_1,\xi_2,\dots,\xi_n}$$

где $P_{\xi_1,\xi_2,\dots,\xi_n}$ *— распределение случайного вектора* ξ. *В частности, если* ξ *— дискретный случайный вектор, то*

$$Mf(\xi_1,\xi_2,\dots,\xi_n) =$$

$$= \sum_{x_1,x_2,\dots,x_n} f(x_1,x_2,\dots,x_n)P\{\xi_1 = x_1, \xi_2 = x_2 \dots, \xi_n = x_n\},$$

где сумма берется по всем возможным значениям $x_1,x_2,\dots,x_n$ *случайных величин* $\xi_1,\xi_2,\dots,\xi_n$ *соответственно, а если* ξ *— абсолютно непрерывный случайный вектор с плотностью вероятности* $p_{\xi_1,\xi_2,\dots,\xi_n}(x_1,x_2,\dots,x_n)$, *то*

$$Mf(\xi_1,\xi_2,\dots,\xi_n) =$$

$$= \int\limits_{\mathbb{R}^n}\!\!\cdots\!\!\int f(x_1,x_2,\dots,x_n)p_{\xi_1,\xi_2,\dots,\xi_n}(x_1,x_2,\dots,x_n)dx,$$

где $dx = dx_1dx_2\dots dx_n$.

Отметим, что формула замены переменной в интеграле Лебега позволила нам вывести формулы, поэволяющие подсчитывать математическое ожидание $f(\xi_1,\xi_2,\dots,\xi_n)$ по закону совместного распределения случайных величин $\xi_1,\xi_2,\dots,\xi_n$ без предварительного нахождения закона распределения случайной величины $f(\xi_1,\xi_2,\dots,\xi_n)$.

Определение 14.6. **Дисперсией** случайной величины $\xi = \xi(\omega)$ называется число

$$D\xi = M(\xi - M\xi)^2. \tag{14.13}$$

Число $\sigma_\xi = \sigma = \sqrt{D\xi}$ называется **средним квадратическим отклонением** случайной величины ξ.

Дисперсия и среднее квдратическое отклонение являются мерами разброса значений случайной величины вокруг математического ожидания. Чем они больше, тем больше разброс значений случайной величины вокруг среднего значения. И, как будет видно из дальнейшего, они равны нулю, когда разброса значений нет, т.е. величина постоянна с вероятностью 1. В практических вопросах удобнее использовать среднее квадратическое отклонение, поскольку его размерность совпадает с размерностью случайной величины, в то время как дисперсия имеет размерность квадрата случайной величины.

Из определения видно, что для существования дисперсии необходимо, как минимум, существование математического ожидания.

В силу свойств математического ожидания

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

Таким образом,

$$D\xi = M\xi^2 - (M\xi)^2. \tag{14.14}$$

Свойства дисперсии.

1. *Для любой случайной величины* $D\xi \geq 0$.

2. $D\xi = 0$ *тогда и только тогда, когда* ξ *постоянна с вероятностью* 1.

Доказательство. Если $\xi = C$ с вероятностью 1, то $\xi^2 = C^2$ с вероятностью 1. Следовательно, $M\xi = C$, $M\xi^2 = C^2$. В силу (14.14) $D\xi = 0$. Обратно, если $D\xi = M(\xi - M\xi)^2 = 0$, то по следствию 1 свойства 5 математического ожидания $(\xi - M\xi)^2 = 0$ с вероятностью 1, т.е. $\xi = M\xi = C$ с вероятностью 1. $\square$

Впрочем это свойство можно доказать, используя следующее свойство.

3. **Неравенство Чебышева (для дисперсии)**. *Для любой случайной величины, имеющей конечную дисперсию при любом* $\varepsilon > 0$

$$P\{\omega : \, |\xi(\omega) - M\xi| \geq \varepsilon\} \leq \frac{D\xi}{\varepsilon^2}.$$

Из неравенства Чебышева следует, что чем меньше дисперсия, тем меньше вероятность больших отклонений ξ от $M\xi$.

Д о к а з а т е л ь с т в о. Применяя неравенство Чебышева для математического ожидания к случайной величине $\eta = (\xi - M\xi)^2$, получим:

$$P\{|\xi - M\xi| \geq \varepsilon\} = P\{\eta \geq \varepsilon^2\} \leq \frac{M\eta}{\varepsilon^2} = \frac{D\xi}{\varepsilon^2}.$$

□

4. *Если* C *— неслучайная величина, то* $D(C\xi) = C^2 D\xi$.

Д о к а з а т е л ь с т в о. В силу свойств математического ожидания

$$D(C\xi) = M(C\xi - M(C\xi))^2 = M[C(\xi - M\xi)]^2 =$$

$$= M[C^2(\xi - M\xi)^2] = C^2 M(\xi - M\xi)^2 = C^2 D\xi.$$

□

Прежде, чем формулировать следующее свойство, дадим необходимые для этого определения.

Определение 14.7. **Ковариацией** или **коэффициентом ковариации** между случайными величинами ξ и η называется число

$$cov(\xi,\ \eta) = M[(\xi - M\xi)(\eta - M\eta)] = M(\xi\eta) - M\xi \cdot M\eta. \tag{14.15}$$

Здесь второе равенство получается из первого, если перемножить выражения в круглых скобках и применить свойство линейности математического ожидания. Из определений дисперсии и ковариации видно, что $D\xi = cov(\xi, \xi)$.

Определение 14.8. Случайные величины ξ и η называются **некоррелированными**, если $cov(\xi,\ \eta) = 0$.

Если случайные величины независимы и имеют конечные математические ожидания, то они некоррелированны.

Это следует из (14.15) и свойства математического ожидания 8. Обратное, вообще говоря, неверно, что показывает решение следующей задачи.

Задача 5. Приведите пример некоррелированных случайных величин, которые не являются независимыми.

5. *Если случайные величины* $\xi_1, \xi_2, \ldots, \xi_n$ *имеют конечные дисперсии,* $c_1, c_2, \ldots, c_n$ *— неслучайные величины, то*

$$D\sum_{i=1}^{n} c_i\xi_i = \sum_{i=1}^{n}\sum_{j=1}^{n} c_i c_j cov(\xi_i,\ \xi_j) =$$

$$= \sum_{i=1}^{n} c_i^2 D\xi_i + \sum_{i=1}^{n} \sum_{j \neq i} c_i c_j cov(\xi_i,\ \xi_j) =$$

$$= \sum_{i=1}^{n} c_i^2 D\xi_i + 2 \sum_{i=1}^{n} \sum_{j<i} c_i c_j cov(\xi_i,\ \xi_j).$$

Д о к а з а т е л ь с т в о. В силу неравенства Коши-Буняковского-Шварца

$$|cov(\xi_i,\ \xi_j)| = |M[(\xi_i - M\xi_i)(\xi_j - M\xi_j)]| \leq$$

$$\leq M(\xi_i - M\xi_i)^2 M(\xi_j - M\xi_j)^2 = D\xi_i D\xi_j$$

из существования конечных дисперсий следует существование попарных ковариаций. При этом

$$D\sum_{i=1}^{n} c_i\xi_i = M\Big[\sum_{i=1}^{n} c_i\xi_i - M\sum_{i=1}^{n} c_i\xi_i\Big]^2 = M\Big[\sum_{i=1}^{n} c_i(\xi_i - M\xi_i)\Big]^2 =$$

$$= M\Big[\sum_{i=1}^{n} c_i(\xi_i - M\xi_i)\sum_{j=1}^{n} c_j(\xi_j - M\xi_j)\Big] =$$

$$= M\sum_{i=1}^{n}\sum_{j=1}^{n} c_i c_j(\xi_i - M\xi_i)(\xi_j - M\xi_j) = \sum_{i=1}^{n}\sum_{j=1}^{n} c_i c_j cov(\xi_i,\ \xi_j),$$

что и требуется.

□

Если положить $c_1 = c_2 = \ldots = c_n = 1$ и считать, что $cov(\xi_i,\ \xi_j) = 0$ при $i \neq j$, то получим

Следствие 14.9. *Если случайные величины $\xi_1, \xi_2, \dots, \xi_n$ попарно некоррелированы, то*

$$D\sum_{i=1}^{n}\xi_i = \sum_{i=1}^{n} D\xi_i.$$

В частности, справедливо

Следствие 14.10. *Если случайные величины $\xi_1, \xi_2, \dots, \xi_n$ попарно независимы, то*

$$D\sum_{i=1}^{n}\xi_i = \sum_{i=1}^{n} D\xi_i.$$

Применяя теорему о математическом ожидании функции от случайной величины к функции $f(x) = (x - M\xi)^2$, получаем следующие формулы для подсчета дисперсии:

$$D\xi = \sum_i (x_i - M\xi)^2 p_i,$$

если ξ — дискретная случайная величина, и

$$D\xi = \int\limits_{-\infty}^{+\infty} (x - M\xi)^2 p_\xi(x)dx,$$

если ξ — абсолютно непрерывная случайная величина.

Можно также вычислять дисперсию по формуле

$$D\xi = M\xi^2 - (M\xi)^2,$$

где для дискретной случайной величины

$$M\xi^2 = \sum_i x_i^2 p_i,$$

а для абсолютно непрерывной случайной величины

$$M\xi^2 = \int_{-\infty}^{+\infty} x^2 p_\xi(x) dx.$$

Примеры.

1. Для случайной величины с распределением Бернулли $M\xi = p$. Так как $\xi^2 = \xi$, то $M\xi^2 = M\xi = p$. Поэтому $D\xi = M\xi^2 - (M\xi)^2 = p - p^2 = pq$.

2. Пусть μ имеет биномиальное распределение с параметрами (n, p). Так как распределение μ совпадает с распределением числа успехов в схеме Бернулли, а числовые характеристики случайной величины однозначно определяются ее законом распределения, то будем считать, что μ — число успехов в схеме n независимых испытаний с вероятностью успеха в одном испытании p. Ранее, пользуясь определением, мы подсчитали $M\mu$. Таким же способом может быть подсчитана и дисперсия. Однако эти числовые характеристики удобнее считать не по определению, а с помощью свойств математического ожидания и дисперсии. Применим прием, который часто используется на практике.

Пусть μ_k — число успехов в k-м испытании, очевидно

μ_k	1	0
P	p	q

,

т.е. μ имеет распределение Бернулли. Поэтому $M\mu = p$, $D\mu = pq$. Поскольку число успехов в n испытаниях складывается из чисел успехов в каждом из этих испытаний, то $\mu = \mu_1 + \mu_2 + \ldots + \mu_n$. В силу свойства линейности математического ожидания $M\mu = M\mu_1 + M\mu_2 +$

$\ldots + M\mu_n = p + p + \ldots + p = np$. Так как каждая μ_k связана только с k-м испытанием, а испытания независимы, то случайные величины $\mu_1, \mu_2, \ldots, \mu_n$ независимы. По следствию 14.10

$$D\mu = D\mu_1 + D\mu_2 + \ldots + D\mu_n = pq + pq + \ldots + pq = npq.$$

Итак,

$$M\mu = np, \quad D\mu = npq.$$

3. Пусть ξ имеет распределение Пуассона с параметром λ. Ранее мы нашли, что $M\xi = \lambda$. Далее,

$$M\xi^2 = \sum_{k=0}^{\infty} k^2 \frac{\lambda^k}{k!} e^{-\lambda} = \sum_{k=1}^{\infty} k^2 \frac{\lambda^k}{k!} e^{-\lambda} = \lambda e^{-\lambda} \sum_{k=1}^{\infty} k \frac{\lambda^{k-1}}{(k-1)!} =$$

$$= \lambda e^{-\lambda} \sum_{k=1}^{\infty} [(k-1)+1] \frac{\lambda^{k-1}}{(k-1)!} =$$

$$= \lambda e^{-\lambda} \Big[\lambda \sum_{k=2}^{\infty} \frac{\lambda^{k-2}}{(k-2)!} + \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} \Big] =$$

$$= \lambda e^{-\lambda} (\lambda e^{\lambda} + e^{\lambda}) = \lambda(\lambda+1) = \lambda^2 + \lambda.$$

Поэтому $D\xi = M\xi^2 - (M\xi)^2 = \lambda^2 + \lambda - \lambda^2 = \lambda$. Итак, для случайной величины с распределением Пуассона

$$M\xi = D\xi = \lambda,$$

т.е. параметр в распределении Пуассона играет одновременно как роль математического ожидания, так и роль дисперсии.

4. Если ξ равномерно распределена на отрезке $[a, b]$, то

$$M\xi^2 = \int_a^b x^2 \cdot \frac{1}{b-a} dx = \frac{a^2 + ab + b^2}{3},$$

следовательно,

$$D\xi = M\xi^2 - (M\xi)^2 = \frac{a^2+ab+b^2}{3} - \left[\frac{a+b}{2}\right]^2 = \frac{(b-a)^2}{12}.$$

Из этой формулы следует, что чем больше разбросаны значения равномерной величины вокруг среднего значения, т.е. чем больше длина отрезка $[a,b]$, тем больше дисперсия.

5. Если ξ имеет показательное распределение с параметром λ, то

$$M\xi^2 = \int\limits_0^\infty x^2\lambda e^{-\lambda x}dx = -\int\limits_0^\infty x^2 d(e^{-\lambda x}) = -x^2e^{-\lambda x}\Big|_0^\infty +$$

$$+2\int\limits_0^\infty xe^{-\lambda x}dx = \frac{2}{\lambda}\int\limits_0^\infty x\lambda e^{-\lambda x}dx.$$

Но $\int\limits_0^\infty \lambda xe^{-\lambda x}dx = \int\limits_{-\infty}^{+\infty} xp_\xi(x)dx = M\xi = \frac{1}{\lambda}$. Значит,

$$D\xi = M\xi^2 - (M\xi)^2 = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}.$$

6. Пусть $\xi \sim N(a,\ \sigma^2)$. По лемме о нормальном распределении она может быть представлена в виде

$$\xi = \sigma\xi_0 + a, \tag{14.16}$$

где $\xi_0 \sim N(0,\ 1)$. Ранее мы убедились в том, что $M\xi_0 = 0$. Поэтому

$$D\xi_0 = M\xi_0^2 = \int\limits_{-\infty}^{+\infty} x^2\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}dx = -\int\limits_{-\infty}^{+\infty} x\frac{1}{\sqrt{2\pi}}d(e^{-\frac{x^2}{2}}) =$$

$$= -x\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}\Big|_{-\infty}^{+\infty} + \int\limits_{-\infty}^{+\infty}\varphi(x)dx = 1.$$

В силу линейности математического ожидания из (14.16) следует, что $M\xi = \sigma M\xi_0 + a = a$. Так как неслучайная величина a не зависит от любой случайной величины, в частности, от $\sigma\xi_0$, то из (14.16), следствия 3 свойства 5 и свойства 4 вытекает, что $D\xi = \sigma^2 D\xi_0 + Da = \sigma^2 \cdot 1 + 0 = \sigma^2$. Таким образом, параметры нормального распределения $N(a,\ \sigma^2)$ имеют следующий смысл: a — это математическое ожидание, а σ^2 — это дисперсия. Поэтому σ имеет смысл среднего квадратического отклонения.

7. Так как случайная величина, распределенная по закону Коши, не имеет математического ожидания, то она не имеет и дисперсии.

Задача 6. Найдите дисперсию случайной величины, имеющей

а) геометрическое распределение;

б) гипергеометрическое распределение;

в) гамма-распределение.

Определение 14.11. **Начальным и абсолютным начальным моментами k-го порядка** называются числа

$$b_k = M\xi^k, \qquad \mu_k = M|\xi|^k.$$

Центральным и абсолютным центральным моментами k-го порядка называются числа

$$c_k = M(\xi - M\xi)^k, \qquad \nu_k = M|\xi - M\xi|^k.$$

В частности, математическое ожидание — это начальный момент первого порядка, а дисперсия — это центральный момент второго порядка. По мере увеличения порядка момента этот момент начинает играть все менее важную роль для характеристики закона распределения. Центральные моменты начинаются с момента второго порядка (центральный момент первого порядка всегда равен 0 и не является числовой характеристикой распределения). Легко выписать формулы, выражающие центральные моменты через начальные:

$$c_2 = D\xi = M\xi^2 - (M\xi)^2 = b_2 - b_1^2,$$

$$c_3 = M(\xi - b_1)^3 = M\xi^3 - 3b_1 M\xi^2 + 3b_1^2 M\xi - b_1^3 =$$
$$= b_3 - 3b_1 b_2 + 3b_1^2 b_1 - b_1^3 = b_3 - 3b_1 b_2 + 2b_1^3$$

и т.д.

Определение 14.12. **Математическим ожиданием** случайного вектора $\xi = (\xi_1, \xi_2, \dots, \xi_n)$ называется неслучайный вектор $M\xi = (M\xi_1, M\xi_2, \dots, M\xi_n)$. Матрица $||cov(\xi_i, \xi_j)||_{i,j=\overline{1,n}}$ называется **матрицей ковариаций** или **ковариационной матрицей** случайного вектора ξ.

Отметим, что на диагонали ковариационной матрицы стоят $D\xi_1, D\xi_2, \dots, D\xi_n$. Как мы убедимся позже, эта матрица симметрическая и неотрицательно определенная.

Задача 7. Покажите, что ковариационная матрица является симметрической неотрицательно определенной матрицей.

Определение 14.13. Коэффициентом корреляции между случайными величинами ξ и η называется число

$$r = r_{\xi,\eta} = \frac{cov(\xi,\eta)}{\sigma_\xi \cdot \sigma_\eta} = \frac{M[(\xi - M\xi)(\eta - M\eta)]}{\sqrt{D\xi \cdot D\eta}}.$$

Отметим, что коэффициент корреляции определен, когда случайные величины ξ и η имеют отличные от нуля дисперсии (в частности, с вероятностью 1 они не должны быть постоянными).

Выведем формулу, полезную для установления свойств коэффициента корреляции. В силу свойства 5 дисперсии,

$$D\Big(\frac{\xi}{\sigma_\xi} \pm \frac{\eta}{\sigma_\eta}\Big) = \frac{1}{\sigma_\xi^2} D\xi + \frac{1}{\sigma_\eta^2} D\eta \pm \frac{2}{\sigma_\xi \sigma_\eta} \cdot cov(\xi,\eta) = 2(1 \pm r). \tag{14.17}$$

Свойства коэффициента корреляции.

1. $|r| \le 1$.

Действительно, так как дисперсия любой случайной величины неотрицательна, то из (14.17) следует, что $1 \pm r \ge 0$, т.е. $-1 \le r \le 1$.

2. $|r| = 1$ *тогда и только тогда, когда* η *и* ξ *с вероятностью 1 связаны линейной функциональной зависимостью:* $\eta = a\xi + b, \quad a \ne 0$. *При этом* $r = 1$ *тогда и только тогда, когда* $a > 0$, *и* $r = -1$ *тогда и только тогда, когда* $a < 0$.

Пусть с вероятностью 1 $\eta = a\xi + b, \quad a \ne 0$. Тогда $D\eta = a^2 D\xi$, следовательно, $\sigma_\eta = |a|\sigma_\xi, \quad \eta - M\eta = a(\xi - M\xi)$, откуда $cov(\xi,\eta) = M[(\xi - M\xi)(\eta - M\eta)] = aM(\xi - M\xi)^2 =$

$a\sigma_\xi^2$. Следовательно,

$$r = \frac{a\sigma_\xi^2}{\sigma_\xi \cdot |a|\sigma_\xi} = \frac{a}{|a|} = \begin{cases} 1, & \text{если } a > 0, \\ -1, & \text{если } a < 0. \end{cases}$$

Обратно, пусть $|r| = 1$. Если $r = 1$, то выбирая в (14.17) знак "минус", получим:

$$D\Big(\frac{\xi}{\sigma_\xi} - \frac{\eta}{\sigma_\eta}\Big) = 0.$$

По свойству 2 дисперсии это означает, что с вероятностью 1 при некотором C

$$\frac{\xi}{\sigma_\xi} - \frac{\eta}{\sigma_\eta} = C,$$

откуда

$$\eta = \frac{\sigma_\eta}{\sigma_\xi} \cdot \xi - C \cdot \sigma_\eta,$$

причем $a = \sigma_\eta/\sigma_\xi > 0$.

Если же $r = -1$, то выбирая в (14.17) знак "плюс", получим:

$$D\Big(\frac{\xi}{\sigma_\xi} + \frac{\eta}{\sigma_\eta}\Big) = 0.$$

откуда с вероятностью 1 при некотором C

$$\eta = -\frac{\sigma_\eta}{\sigma_\xi} \cdot \xi + C \cdot \sigma_\eta,$$

причем $a = -\sigma_\eta/\sigma_\xi < 0$.

3. *Если коэффициент корреляции между независимыми случайными величинами определен, то он равен нулю.*

Это следует из доказанного нами факта, что независимые случайные величины с конечными математическими ожиданиями являются некоррелированными.

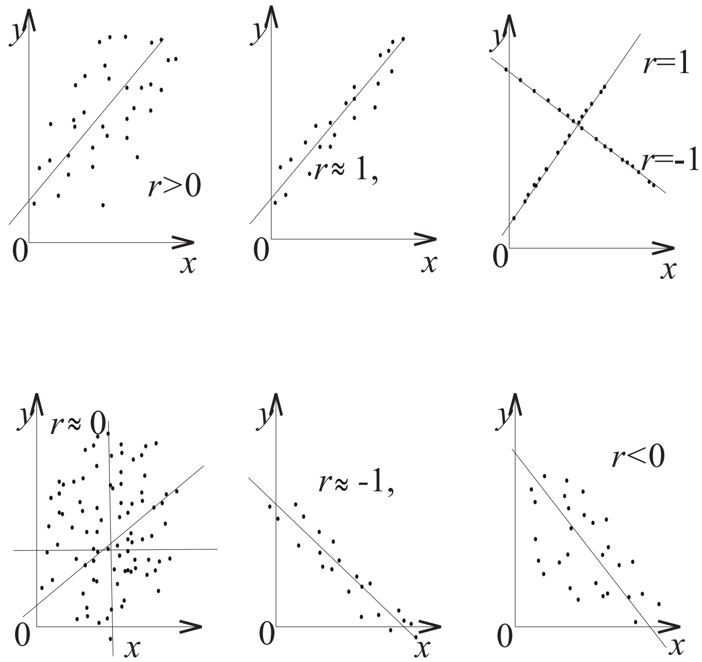


Рис.5

Коэффициент корреляции характеризует тенденцию к линейной функциональной зависимости между случайными величинами. Чем ближе $|r|$ к 1, тем больше эта тенденция; при значениях $|r|$, близких к 0, такая тенденция пропадает. В силу свойства 2 при $|r| = 1$ эта тенденция переходит в строгую линейную функциональную зависимость между случайными величинами. Качественное представление о величине коэффициента корреляции можно получить на основании статистических данных — результатах измерений (x_i, y_i) значений случайного вектора (ξ_1, ξ_2), изображенных на Рис.5.

§ 15. Виды сходимости случайных величин

Пусть $\xi, \xi_1, \xi_2, \ldots$ — случайные величины, определенные на вероятностном пространстве (Ω, F, P).

Определение 15.1. Говорят, что последовательность случайных величин (ξ_n) **сходится по вероятности** к случайной величине ξ (обозначение $\xi_n \xrightarrow{P} \xi$), если для любого $\varepsilon > 0 \quad P\{\omega : \ |\xi_n(\omega) - \xi(\omega)| \geq \varepsilon\} \to 0$ при $n \to \infty$.

Это определение совпадает с определением сходимости по вероятностной мере последовательности измеримых относительно σ-алгебры F функций $(\xi_n(\omega))$ к измеримой функции $\xi(\omega)$.

Заметим, что фигурирующее в определении 15.1 требование эквивалентно тому, что $P\{\omega : \ |\xi_n(\omega) - \xi(\omega)| < \varepsilon\} \to 1$ при $n \to \infty$. Таким образом, сходимость по вероятности означает, что при больших n с вероятностью, сколь угодно близкой к 1, ξ_n и ξ будут сколь угодно мало отличаться друг от друга.

Теорема 15.2. *Пусть* $\xi_n \xrightarrow{P} a, \quad \eta_n \xrightarrow{P} b$, *а борелевская функция двух действительных переменных* $f(x, y)$ *непрерывна в точке* (a, b). *Тогда* $f(\xi_n, \eta_n) \xrightarrow{P} f(a, b)$.

Доказательство. Так как f непрерывна в точке (a, b), то для любого $\varepsilon > 0$ найдется $\delta > 0$ такое, что если $|x - a| < \delta, \ \ |y - b| < \delta$, то $|f(x, y) - f(a, b)| < \varepsilon$. Следовательно, если $|f(\xi_n, \eta_n) - f(a, b)| \geq \varepsilon$, то нарушится хотя бы одно из неравенств $|\xi_n - a| < \delta, \ \ |\eta_n - b| < \delta$, т.е. выполнится хотя бы одно из неравенств $|\xi_n - a| \geq \delta, \ |\eta_n - b| \geq \delta$. Поэтому

$$\{\omega : \ |f(\xi_n(\omega), \eta_n(\omega)) - f(a, b)| \geq \varepsilon\} \subset$$

$$\subset \{\omega :\ |\xi_n(\omega) - a| \geq \delta\} \cup \{\omega :\ |\eta_n(\omega) - b| \geq \delta\}.$$

По свойствам вероятности

$$P\{\omega :\ |f(\xi_n(\omega), \eta_n(\omega)) - f(a,b)| \geq \varepsilon\} \leq$$

$$\leq P\{\{\omega :\ |\xi_n(\omega) - a| \geq \delta\} \cup \{\omega :\ |\eta_n(\omega) - b| \geq \delta\}\} \leq$$

$$\leq P\{\omega :\ |\xi_n(\omega) - a| \geq \delta\} + P\{\omega :\ |\eta_n(\omega) - b| \geq \delta\}.$$

Так как $\xi_n \xrightarrow{P} a$, $\eta_n \xrightarrow{P} b$, то по определению сходимости по вероятности каждое слагаемое в правой части последнего неравенства стремится к нулю. Следовательно, при каждом $\varepsilon > 0$

$$P\{\omega :\ |f(\xi_n(\omega), \eta_n(\omega)) - f(a,b)| \geq \varepsilon\} \to 0.$$

По определению сходимости по вероятности это означает, что $f(\xi_n, \eta_n) \xrightarrow{P} f(a,b)$. $\square$

Полагая $f(x,y) = x+y,\ x-y,\ xy,\ x/y$, получаем свойства предела по вероятности, аналогичные свойствам обычных пределов.

Следствие 15.3. *Если* $\xi_n \xrightarrow{P} a$, $\eta_n \xrightarrow{P} b$, *то*

$$1)\ \xi_n + \eta_n \xrightarrow{P} a+b;\ \ 2)\ \xi_n - \eta_n \xrightarrow{P} a-b;\ \ 3)\ \xi_n \cdot \eta_n \xrightarrow{P} a\cdot b;$$

$$4)\ \frac{\xi_n}{\eta_n} \xrightarrow{P} \frac{a}{b}, \quad \text{если}\ \ b \neq 0.$$

Задача 1. Покажите, что если $\xi_n \xrightarrow{P} \xi$, $\eta_n \xrightarrow{P} \eta$, то $\xi_n + \eta_n \xrightarrow{P} \xi + \eta$, $\xi_n - \eta_n \xrightarrow{P} \xi - \eta$, $\xi_n \cdot \eta_n \xrightarrow{P} \xi \cdot \eta$, а если дополнительно $P\{\omega :\ \eta(\omega) \neq 0\} = 1$, то $\xi_n/\eta_n \xrightarrow{P} \xi/\eta$.

Определение 15.4. Говорят, что последовательность случайных величин (ξ_n) **сходится почти наверное** к случайной величине ξ (обозначение $\xi_n \overset{\text{п.н.}}{\to} \xi$), если $P\{\omega : \lim\limits_{n\to\infty} \xi_n(\omega) \neq \xi(\omega)\} = 0$.

Это определение совпадает с определением сходимости почти всюду относительно вероятностной меры P последовательности измеримых относительно σ-алгебры F функций $(\xi_n(\omega))$ к измеримой функции $\xi(\omega)$.

Из теории меры известно, что из сходимости почти всюду следует сходимость по мере. Поэтому из сходимости почти наверное вытекает сходимость по вероятности.

Заметим, что фигурирующее в определении 15.4 требование эквивалентно тому, что $P\{\omega : \lim\limits_{n\to\infty} \xi_n(\omega) = \xi(\omega)\} = 1$. Таким образом, сходимость почти наверное означает, что с вероятностью 1 последовательность ξ_n сходится к ξ.

Задача 2. Покажите, что если $\xi_n \overset{\text{п.н.}}{\to} \xi$, $\eta_n \overset{\text{п.н.}}{\to} \eta$, то $\xi_n + \eta_n \overset{\text{п.н.}}{\to} \xi + \eta$, $\xi_n - \eta_n \overset{\text{п.н.}}{\to} \xi - \eta$, $\xi_n \cdot \eta_n \overset{\text{п.н.}}{\to} \xi \cdot \eta$, а если дополнительно $P\{\omega : \eta(\omega) \neq 0\} = 1$, то $\xi_n/\eta_n \overset{\text{п.н.}}{\to} \xi/\eta$.

Определение 15.5. Говорят, что последовательность случайных величин (ξ_n) **сходится в среднем порядка** p $(p \geq 1)$ к случайной величине ξ (обозначение $\xi_n \overset{L_p}{\to} \xi$), если $M|\xi_n - \xi|^p \to 0$ при $n \to \infty$.

Если $L_p = L_p(\Omega, F, P)$ — линейное нормированное пространство функций $\xi(\omega)$, для которых ξ^p интегрируема по Лебегу по мере P, с нормой $||\xi|| = (M|\xi|^p)^{\frac{1}{p}}$, то сходимость в среднем порядка p является сходимостью по этой норме. Известно также, что из сходимости почти всюду не следует сходимость в среднем порядка p, а из сходимости в среднем порядка p не следует сходимость почти всюду. Следовательно, ни одна из сходимостей (почти наверное и в среднем порядка p) не является более сильной, чем другая.

При $p = 2$ сходимость в среднем порядка p называют сходимостью в среднем квадратическом, а тот факт, что $\xi_n \xrightarrow{L_2} \xi$, записывают также в виде $\xi = \underset{n\to\infty}{\mathrm{l.\,i.\,m.}}\ \xi_n$ (limit in mean).

Задача 3. Покажите, что если $\xi_n \xrightarrow{L_p} \xi$, $\eta_n \xrightarrow{L_p} \eta$, то $\xi_n + \eta_n \xrightarrow{L_p} \xi + \eta$, $\xi_n - \eta_n \xrightarrow{L_p} \xi - \eta$, $\xi_n \cdot \eta_n \xrightarrow{L_p} \xi \cdot \eta$, а если дополнительно $P\{\omega:\ \eta(\omega) \neq 0\} = 1$, то $\xi_n/\eta_n \xrightarrow{L_p} \xi/\eta$.

Указание. Использовать тот факт, что сходимость в среднем порядка p, $p \geq 1$, является сходимостью по норме L_p.

Утверждение 15.6. *Если $\xi_n \xrightarrow{L_p} \xi$, то $\xi_n \xrightarrow{P} \xi$.*

Доказательство. В силу неравенства Чебышева для математических ожиданий

$$0 \leq P\{|\xi_n - \xi| \geq \varepsilon\} = P\{|\xi_n - \xi|^p \geq \varepsilon^p\} \leq \frac{M|\xi_n - \xi|^p}{\varepsilon^p} \to 0$$

при $n \to \infty$, что и требуется. $\square$

Определение 15.7. Говорят, что последовательность функций распределения $(F_n(x))$ **слабо сходится** или **сходится в основном** к неубывающей функции $F(x)$ (обозначение $F_n(x) \Rightarrow F(x)$), если $F_n(x) \to F(x)$

в каждой точке $x \in \mathbb{R}$, в которой предельная функция $F(x)$ непрерывна.

Заметим, что в условиях определения 15.7 предельная функция $F(x)$ не обязательно является функцией распределения. Например, если $\xi_n \equiv n$, $F_n(x) = P\{\xi_n < x\}$, то $F_n(x) \Rightarrow 0$, а функция $F(x) \equiv 0$ не является функцией распределения (так как $F(+\infty) = 0 \neq 1$). В некоторых учебниках в определении 15.7 выбрасывается требование монотонности предельной функции F, что может привести к недоразумениям. В частности, без этого требования любая последовательность функций распределения слабо сходится к функции Дирихле

$$F(x) = \begin{cases} 1, & \text{если } x \text{ рационально}, \\ 0, & \text{если } x \text{ иррационально}, \end{cases}$$

поскольку все ее точки являются точками разрыва, а множество точек непрерывности пусто. В принятом нами определении множество точек разрыва предельной функции не более чем счетно; следовательно, множество точек непрерывности всюду плотно на числовой прямой. Пусть

$$F^*(x) = \sup_{y<x,\, y\in C(F)} F(y),$$

где $C(F)$ — множество точек непрерывности $F(x)$. Функция F^* — неубываюшая и непрерывна слева, причем ее множество точек разрыва совпадает с множеством точек разрыва F. Значит, $F_n(x) \Rightarrow F(x)$ в том и только том случае, когда $F_n(x) \Rightarrow F^*(x)$ (более строгое доказательство вышесказанного приводится в приложении). Поэтому, не ограничивая общности, можно считать, что в определении слабой сходимости предельная функция

— неубывающая и непрерывна слева. Она может, следовательно, оказаться не функцией распределения только за счет нарушения свойства $F(-\infty)=0,\ \ F(+\infty)=1$.

Определение 15.8. Говорят, что последовательность случайных величин (ξ_n) **слабо сходится** или **сходится по распределению** к случайной величине ξ (обозначение $\xi_n \Rightarrow \xi$), если $F_{\xi_n}(x) \Rightarrow F_\xi(x)$.

Теорема 15.9. *Если* $\xi_n \xrightarrow{P} \xi$, *то* $\xi_n \Rightarrow \xi$.

Доказательство. Введем обозначения $F_n(x) = F_{\xi_n}(x),\ \ F(x)=F_\xi(x)$. Пусть $x'<x<x''$. Покажем, что

$$F(x') \le \liminf_{n\to\infty} F_n(x) \le \limsup_{n\to\infty} F_n(x) \le F(x''). \qquad (15.1)$$

Сначала покажем, что $F(x') \le \liminf\limits_{n\to\infty} F_n(x)$. Имеем:

$$F(x') = P\{\xi<x'\} = P\{\xi<x',\ \xi_n<x\}+$$

$$+P\{\xi<x',\ \xi_n \ge x\} \le P\{\xi_n<x\} + P\{|\xi_n-\xi| \ge x-x'\} =$$

$$= F_n(x) + P\{|\xi_n-\xi| \ge x-x'\},$$

т.е. $F(x') - P\{|\xi_n-\xi| \ge x-x'\} \le F_n(x)$. Так как $\xi_n \xrightarrow{P} \xi$, то $P\{|\xi_n-\xi| \ge x-x'\} \to 0$ при $n\to\infty$. Следовательно,

$$F(x') = \lim_{n\to\infty}[F(x') - P\{|\xi_n-\xi| \ge x-x'\}] =$$

$$= \liminf_{n\to\infty}[F(x') - P\{|\xi_n-\xi| \ge x-x'\}] \le$$

$$\le \liminf_{n\to\infty} F_n(x).$$

Теперь покажем, что $F(x'') \ge \limsup\limits_{n\to\infty} F_n(x)$. Имеем:

$$1-F(x'') = P\{\xi \ge x''\} = P\{\xi \ge x'',\ \xi_n<x\}+$$

$$+P\{\xi \geq x'',\ \xi_n \geq x\} \leq P\{\xi_n \geq x\} + P\{|\xi_n - \xi| \geq x'' - x\} =$$
$$= 1 - F_n(x) + P\{|\xi_n - \xi| \geq x'' - x\},$$

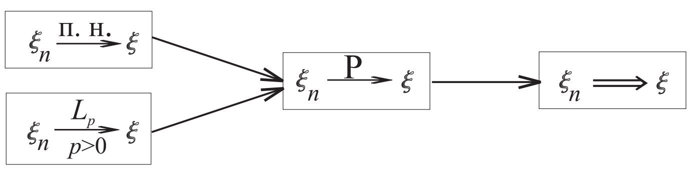
откуда $F(x'') + P\{|\xi_n - \xi| \geq x'' - x\} \geq F_n(x)$. Поэтому $F(x'') \geq \limsup\limits_{n\to\infty} F_n(x)$. Так как нижний предел всегда меньше верхнего, то (15.1) доказано.

Переходя в (15.1) к пределу при $x' \uparrow x, \quad x'' \downarrow x$, получим:

$$F(x) \leq \liminf_{n\to\infty} F_n(x) \leq \limsup_{n\to\infty} F_n(x) \leq F(x+0).$$

Если x — точка непрерывности F, то $F(x) = F(x+0)$. Следовательно, существует $\lim\limits_{n\to\infty} F_n(x) = F(x)$. Это означает, что $F_n(x) \Rightarrow F(x)$, т.е. $\xi_n \Rightarrow \xi$. □

Собирая вместе полученные выше результаты, получаем следующую диаграмму, характеризующую соотношения между введенными четырьмя видами сходимости случайных величин:



Задача 4. Приведите пример, показывающий, что из слабой сходимости не следует сходимость по вероятности.

Однако в том частном случае, когда предельная величина не случайна, эти виды сходимости эквивалентны.

Лемма о слабой сходимости к константе. *Если $\xi_n \Rightarrow C$, то $\xi_n \xrightarrow{P} C$.*

Д о к а з а т е л ь с т в о. Так как $\xi_n \Rightarrow C$, то

$$F_n(x) = F_{\xi_n}(x) \to \begin{cases} 1, & \text{если } x > C, \\ 0, & \text{если } x < C. \end{cases}$$

Следовательно, для любого $\varepsilon > 0$

$$P\{|\xi_n - C| \geq \varepsilon\} = 1 - P\{C - \varepsilon < \xi_n < C + \varepsilon\} =$$

$$= 1 - (F_n(C + \varepsilon) - F_n(C - \varepsilon + 0)) \to 1 - (1 - 0) = 0.$$

Это означает, что $\xi_n \xrightarrow{P} C$. $\square$.

Лемма о слабой сходимости функций распределения к непрерывной функции. *Если последовательность функций распределения слабо сходится к непрерывной функции, то она равномерно сходится на $\mathbb{R}$ к этой функции.*

Доказательство выносится в приложение. Заметим, что здесь уже предельная функция является функцией распределения. Действительно, из монотонности $F_n(x)$ следует монотонность $F(x)$. В силу леммы $\forall \varepsilon > 0 \quad \exists N$ такое, что $\forall x \in R$ и $\forall n > N$ выполняется $|F_n(x) - F(x)| < \varepsilon$. Переходя в последнем неравенстве к пределу при $x \to -\infty$ и $x \to +\infty$, получим, что $|-F(-\infty)| < \varepsilon$ и $|1 - F(+\infty)| < \varepsilon$. В силу произвольности $\varepsilon > 0$ отсюда следует, что $F(-\infty) = 0,\ F(+\infty) = 1$.

§ 16. Законы больших чисел

Предположим, что в одинаковых условиях проводится n независимых испытаний, в которых измеряется некоторая физическая постоянная a. Обычно измерения подвержены случайным ошибкам, так что результат k-го измерения $\xi_k = a + \delta_k$ отличается от измеряемой величины на случайную ошибку δ_k, $k = \overline{1,n}$. Пусть результаты измерений лишены систематической ошибки, т.е. случайные величины δ_k имеют нулевое математическое ожидание (если речь идет о взвешивании продавцом товара, то результаты честного продавца лишены систематической ошибки — одному недовесил, другому перевесил, но в среднем не обманывает). Очевидно, $M\xi_k = a$. Путем многовековой практики принято в качестве среднего значения измеряемой величины a брать среднее арифметическое результатов измерений $\frac{1}{n}\sum\limits_{k=1}^{n}\xi_k$. Возникает законный вопрос, насколько могут отличаться эти величины. Хотелось бы, чтобы разница между ними была мала, т.е. чтобы она при $n \to \infty$ стремилась в некотором смысле к нулю. В более общих случаях, чем в рассмотренном примере, случайные величины ξ_k могут иметь разные математические ожидания. Поэтому введем следующее определение.

Определение 16.1. Говорят, что последовательность случайных величин $\xi_1, \xi_2, \ldots$, определенных на одном и том же вероятностном пространстве $(\Omega, \mathcal{F}, P)$, **удовлетворяет закону больших чисел**, если

$$\frac{1}{n}\sum_{k=1}^{n}\xi_k - \frac{1}{n}\sum_{k=1}^{n}M\xi_k \xrightarrow{P} 0,$$

и **усиленному закону больших чисел**, если

$$\frac{1}{n}\sum_{k=1}^{n}\xi_k - \frac{1}{n}\sum_{k=1}^{n}M\xi_k \xrightarrow{\text{п.н.}} 0.$$

Так как из сходимости почти наверное следует сходимость по вероятности, то если последовательность случайных величин удовлетворяет усиленному закону больших чисел, то она удовлетворяет закону больших чисел.

Применительно к рассмотренному примеру правило брать в качестве значения измеряемой величины среднее арифметическое измерений будет обосновано, если удастся доказать, что последовательность измерений удовлетворяет закону больших чисел. Поэтому возникает вопрос о том, какие условия надо наложить на случайные величины, чтобы их последовательность удовлетворяла закону больших чисел.

Теорема 16.2 (закон больших чисел в форме Чебышева). *Если* (ξ_n) *— последовательность попарно независимых случайных величин, имеющих конечные равномерно по* n *ограниченные дисперсии (т.е. существует постоянная* C*, такая что* $D\xi_n \leq C$ *для всех* n*), то она удовлетворяет закону больших чисел.*

Д о к а з а т е л ь с т в о. В силу неравенства Чебышева и свойств дисперсиии для каждого $\varepsilon > 0$

$$0 \leq P\Big\{\Big|\frac{1}{n}\sum_{k=1}^{n}\xi_k - \frac{1}{n}\sum_{k=1}^{n}M\xi_k\Big| \geq \varepsilon\Big\} \leq$$

$$\leq \frac{D\Big(\frac{1}{n}\sum\limits_{k=1}^{n}\xi_k\Big)}{\varepsilon^2} \leq \frac{C}{n\varepsilon^2} \to 0.$$

По теореме "о двух милиционерах"

$$P\Big\{\Big|\frac{1}{n}\sum_{k=1}^{n}\xi_k - \frac{1}{n}\sum_{k=1}^{n}M\xi_k\Big| \geq \varepsilon\Big\} \to 0.$$

Согласно определению сходимости по вероятности это означает, что последовательность (ξ_n) удовлетворяет закону больших чисел. $\square$

Теорема 16.3 (закон больших чисел в форме Бернулли). *Пусть μ — число успехов в схеме Бернулли, n — число испытаний, p — вероятность успеха в одном испытании. Тогда*

$$\frac{\mu}{n} \xrightarrow{P} p.$$

Д о к а з а т е л ь с т в о. Обозначим через μ_k число успехов в k-ом испытании. Эта величина имеет распределение Бернулли, следовательно, $D\mu_k = pq = C$. Случайные величины $\mu_1, \mu_2, \ldots$ независимы и имеют равномерно ограниченные дисперсии. По закону больших чисел в форме Чебышева

$$\frac{\mu}{n} = \frac{1}{n}\sum_{k=1}^{n}\mu_k \xrightarrow{P} M\mu_k = p.$$

$\square$

Учитывая, что $pq = p(1-p) = -(p-1/2)^2 + 1/4 \leq 1/4$, в силу неравенства Чебышева получаем следующую оценку:

$$P\left\{\left|\frac{\mu}{n} - p\right| \geq \delta\right\} \leq \frac{D\left(\frac{\mu}{n}\right)}{\delta^2} = \frac{pq}{n\delta^2} \leq \frac{1}{4n\delta^2}, \qquad (16.1)$$

которая нам понадобится чуть позже.

Из этой оценки и теоремы "о двух милиционерах" сразу вытекает закон больших чисел в форме Бернулли.

Именно статистическая устойчивость относительной частоты события служила нам отправной точкой при аксиоматизации вероятности. Выполнение закона больших чисел в форме Бернулли показывает разумность введенной аксиоматики, ибо в ее рамках показано, что в определенном смысле относительная частота события стабилизируется вокруг его вероятности, а именно, с вероятностью, сколь угодно близкой к единице, она сколь угодно мало отклоняется от этой вероятности.

Определение 16.4. **Верхним пределом последовательности событий** A_n называется событие

$$\limsup A_n = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k.$$

Если это событие происходит, то происходит бесконечное число событий A_k (в противном случае для какого-то n не произойдет $\bigcup\limits_{k=n}^{\infty} A_k$). Обратно, если происходит бесконечное число A_k, то для каждого n происходит хотя бы одно из событий $A_n, A_{n+1}, \dots$. Таким образом, $\limsup A_n$ — событие, состояшее в том, что произойдет бесконечно много событий из $A_1, A_2, \dots$.

Лемма Бореля-Кантелли.

1. Если $\sum\limits_{n=1}^{\infty} P(A_n) < \infty$*, то* $P(\limsup A_n) = 0$.

2. Если $\sum\limits_{n=1}^{\infty} P(A_n) = \infty$ *и события* $A_1, A_2, \dots$ *независимы, то* $P(\limsup A_n) = 1$.

События $A_1, A_2, \dots$ называются независимыми, если для любого n события $A_1, A_2, \dots, A_n$ независимы.

Д о к а з а т е л ь с т в о. 1. Так как $\bigcup\limits_{k=n}^{\infty} A_k \downarrow \bigcap\limits_{n=1}^{\infty} \bigcup\limits_{k=n}^{\infty} A_k$, то

$$P(\limsup A_n) = P\Big(\bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k\Big) = \lim_{n\to\infty} P\Big(\bigcup_{k=n}^{\infty} A_k\Big) \leq$$

$$\leq \lim_{n\to\infty} \sum_{k=n}^{\infty} P(A_k) = 0.$$

2. Рассмотрим функцию $\varphi(x) = \ln(1-x) + x, \quad 0 \leq x < 1$. Так как $\varphi'(x) = \frac{x}{x-1} \leq 0$, то $\varphi(x)$ не возрастает, $\varphi(0) = 0$, следовательно, $\varphi(x) \leq 0$. Итак,

$$\ln(1-x) \leq -x, \quad \text{если} \ \ 0 \leq x < 1. \tag{16.2}$$

Так как события $A_1, A_2, \ldots$ независимы, то для любого $N \geq n$

$$P\Big(\bigcap_{k=n}^{N} \overline{A}_k\Big) = \prod_{k=n}^{N} P(\overline{A}_k).$$

Но $\bigcap\limits_{k=n}^{N} \overline{A}_k \downarrow \bigcap\limits_{k=n}^{\infty} \overline{A}_k$ при $N \to \infty$. Следовательно,

$$P\Big(\bigcap_{k=n}^{\infty} \overline{A}_k\Big) = \lim_{N\to\infty} P\Big(\bigcap_{k=n}^{N} \overline{A}_k\Big) = \prod_{k=n}^{\infty} P(\overline{A}_k).$$

Так как $A_k, \ \ k \geq 1$, независимы, то $P(A_k) \neq 1$. Используя (16.2), получим

$$\ln P\Big(\bigcap_{k=n}^{\infty} \overline{A}_k\Big) = \sum_{k=n}^{\infty} \ln\left[1 - P(A_k)\right] \leq -\sum_{k=n}^{\infty} P(A_k) = -\infty.$$

Поэтому $P\Big(\bigcap\limits_{k=n}^{\infty} \overline{A}_k\Big) = 0$. Переходя к противоположному событию и используя соотношение двойственности, получим $P\Big(\bigcup\limits_{k=n}^{\infty} A_k\Big) = 1$. Значит,

$$P(\limsup A_n) = \lim_{n\to\infty} P\Big(\bigcup_{k=n}^{\infty} A_k\Big) = 1.$$

$\square$

Выведем одно полезное достаточное условие сходимости почти наверное. Имеем,

$$\xi_n(\omega) \to \xi(\omega) \ \text{ тогда и только тогда, когда}$$

$$\forall\ \varepsilon > 0\ \ \exists\ \ n \in \mathbb{N}\ \text{ такое, что }\ \forall\ \ k \geq n\ \text{ выполняется}$$

$$|\xi_k(\omega) - \xi(\omega)| < \varepsilon.$$

Поэтому

$$\xi_n(\omega) \to \xi(\omega) \ \text{ тогда и только тогда, когда}$$

$$\forall\ \ m \in \mathbb{N}\ \ \exists\ \ n \in \mathbb{N}\ \text{ такое, что }\ \forall\ \ k \geq n\ \text{ выполняется}$$

$$|\xi_k(\omega) - \xi(\omega)| < \frac{1}{m}.$$

Действительно, если утверждение выполнено $\forall\varepsilon > 0$, то оно выполнено для $\varepsilon = \frac{1}{m}$, а если оно выполнено $\forall m$, то поскольку $\forall\varepsilon > 0\ \ \exists m$ такое, что $\frac{1}{m} < \varepsilon$, то $|\xi_k - \xi| < \frac{1}{m} < \varepsilon$, т.е. оно выполнено $\forall\varepsilon > 0$. Значит,

$$\xi_n(\omega) \not\to \xi(\omega)\ \text{ тогда и только тогда, когда }\ \exists\ \ m \in \mathbb{N}$$

такое, что $\forall\ \ n \in \mathbb{N}\ \ \exists\ \ k \geq n$, для которого выполняется

$$|\xi_k(\omega) - \xi(\omega)| \geq \frac{1}{n}.$$

Следовательно,

$$\{\omega:\ \xi_n(\omega) \nrightarrow \xi(\omega)\} = \bigcup_{m=1}^{\infty}\bigcap_{n=1}^{\infty}\bigcup_{k=n}^{\infty}\Big\{\omega:\ |\xi_k(\omega)-\xi(\omega)| \geq \frac{1}{m}\Big\}.$$

Достаточное условие сходимости почти наверное. *Если для любого* $\varepsilon > 0$ *сходится ряд* $\sum\limits_{n=1}^{\infty} P\{\omega : |\xi_n(\omega) - \xi(\omega)| \geq \varepsilon\}$, *то* $\xi_n \overset{\text{п.н.}}{\to} \xi$.

Д о к а з а т е л ь с т в о. Пусть $A_n^\varepsilon = \{\omega:\ |\xi_n(\omega) - \xi(\omega)| \geq \varepsilon\}$. По условию $\sum\limits_{n=1}^{\infty} P(A_n^\varepsilon) < \infty$. По лемме Бореля-Кантелли $P(\limsup A_n^\varepsilon) = 0$, т.е. $P\Big(\bigcap\limits_{n=1}^{\infty}\bigcup\limits_{k=n}^{\infty} A_k^\varepsilon\Big) = 0$. Следовательно, $\forall m \in \mathbb{N}$ выполняется $P\Big(\bigcap\limits_{n=1}^{\infty}\bigcup\limits_{k=n}^{\infty} A_k^{\frac{1}{m}}\Big) = 0$, т. е. $P\Big\{\omega:\ \bigcap\limits_{n=1}^{\infty}\bigcup\limits_{k=n}^{\infty}\Big\{\omega:\ |\xi_k(\omega) - \xi(\omega)| \geq \frac{1}{m}\Big\}\Big\} = 0 \quad \forall m \in \mathbb{N}$. Отсюда следует, что

$$P\{\omega:\ \xi_n(\omega) \nrightarrow \xi(\omega)\} =$$

$$= P\Big(\bigcup_{m=1}^{\infty}\bigcap_{n=1}^{\infty}\bigcup_{k=n}^{\infty}\Big\{\omega:\ |\xi_k(\omega) - \xi(\omega)| \geq \frac{1}{m}\Big\}\Big) \leq$$

$$\leq \sum_{m=1}^{\infty} P\Big(\bigcap_{n=1}^{\infty}\bigcup_{k=n}^{\infty}\Big\{\omega:\ |\xi_k(\omega) - \xi(\omega)| \geq \frac{1}{m}\Big\}\Big) = 0.$$

$\square$

Усиленный закон больших чисел в форме Бореля. *Если μ — число успехов в схеме n независимых испытаний Бернулли с вероятностью успеха p, то*

$$\frac{\mu}{n} \xrightarrow{\text{п.н.}} p.$$

Д о к а з а т е л ь с т в о. Пусть

$$\mu_n^* = \frac{\mu}{n} = \frac{1}{n}\sum_{k=1}^{n} \mu_k,$$

где μ_k — число успехов в k-м испытании. Тогда $M\mu_n^* = \frac{1}{n}M\mu = p$, $D\mu_n^* = \frac{1}{n^2}D\mu = \frac{pq}{n}$. В силу неравенства Чебышева

$$\sum_{k=1}^{\infty} P\{|\mu_{k^2}^* - p| \geq \varepsilon\} \leq \sum_{k=1}^{\infty} \frac{D\mu_{k^2}^*}{\varepsilon^2} = \frac{pq}{\varepsilon^2}\sum_{k=1}^{\infty}\frac{1}{k^2} < \infty;$$

по достаточному условию сходимости почти наверное $\mu_{k^2}^* \xrightarrow{\text{п.н.}} p$.

Далее, $\forall n\ \ \exists k = k(n)$ такое, что $k^2 \leq n < (k+1)^2$. При этом $0 \leq n - k^2 \leq 2k$ и $k \to \infty$ при $n \to \infty$. Следовательно,

$$|\mu_n^* - \mu_{k^2}^*| = \Big|\frac{1}{n}\sum_{j=1}^{n}\mu_j - \frac{1}{k^2}\sum_{j=1}^{k^2}\mu_j\Big| = \Big|\frac{1}{n}\sum_{j=1}^{k^2}\mu_j + \frac{1}{n}\sum_{j=k^2+1}^{n}\mu_j -$$

$$-\frac{1}{k^2}\sum_{j=1}^{k^2}\mu_j\Big| = \frac{n-k^2}{nk^2}\sum_{j=1}^{k^2}\mu_j + \frac{1}{n}\sum_{j=k^2+1}^{n}\mu_j \leq \frac{2}{nk}\sum_{j=1}^{k^2}\mu_j +$$

$$+\frac{1}{n}\sum_{j=k^2+1}^{n}\mu_j \leq \frac{2k}{n} + \frac{n-k^2}{n} \leq \frac{4k}{n} \leq \frac{4}{k}.$$

Пусть $N = \{\omega : \ \mu_{k^2}(\omega) \to p\}$. Так как $\mu_{k^2} \overset{\text{п.н.}}{\to} p$, то $P(N) = 1$. Поскольку

$$0 \le |\mu_n^* - p| \le |\mu_n^* - \mu_{k^2}^*| + |\mu_{k^2}^* - p| \le \frac{4}{k} + |\mu_{k^2}^* - p|$$

и на множестве N выполняется $\mu_{k^2} \to p$, то по теореме о двух милиционерах на множестве N будет $\mu_n^* \to p$. Это и означает, что $\mu_n^* \overset{\text{п.н.}}{\to} p$.

$\square$

Приведем без доказательства более общий усиленный закон больших чисел, частным случаем которого является закон в форме Бореля, а также неравенство, которое используется при его доказательстве.

Неравенство Колмогорова. *Пусть* $\xi_1, \xi_2, \dots, \xi_n$ *— независимые случайные величины,* $M\xi_k = 0, \quad M\xi_k^2 < \infty, \ S_k = \sum\limits_{l=1}^{k} \xi_l, \ k = \overline{1,n}$. *Тогда для любого* $\varepsilon > 0$

$$P\Big\{ \max_{1 \le k \le n} |S_k| \ge \varepsilon \Big\} \le \frac{MS_n^2}{\varepsilon^2}.$$

Усиленный закон больших чисел в форме Колмогорова. *Пусть* ξ_n *— последовательность независимых одинаково распределенных (т.е. имеющих один и тот же закон распределения) случайных величин. Для того, чтобы*

$$\frac{1}{n} \sum_{k=1}^{n} \xi_k \overset{\text{п.н.}}{\to} a,$$

где a — некоторая постоянная, необходимо и достаточно, чтобы случайные величины $\xi_1, \xi_2, \ldots$ имели конечное математическое ожидание $M\xi_k = a$.

Впрочем достаточность этого утверждения впервые была доказана Хинчиным и носит название усиленного закона больших чисел в форме Хинчина. Позже мы докажем его с помощью другого метода — метода характеристических функций.

В заключение данной темы покажем, как теория вероятностей была применена к классической проблеме математического анализа. Известная теорема Вейерштрасса гласит, что любая непрерывная на отрезке функция действительной переменной может быть равномерно приближена на этом отрезке многочленами. С помощью очевидной замены переменной любой отрезок переводится в отрезок $[0,1]$. Поэтому, не ограничивая общности, достаточно доказать эту теорему для функций, определенных на отрезке $[0,1]$.

Пусть функция $f(p)$ непрерывна на отрезке $[0,1]$. Рассмотрим схему независимых испытаний Бернулли с вероятностью успеха p. Так как по усиленному закону больших чисел $\frac{\mu}{n} \overset{\text{п.н.}}{\to} p$, а функция $f(p)$ непрерывна на $[0,1]$, то $f\left(\frac{\mu}{n}\right) \overset{\text{п.н.}}{\to} f(p)$. Кроме того, раз f непрерывна на $[0,1]$, то она ограничена: $|f(p)| \le C$. По теореме Лебега об ограниченной сходимости (предельный переход под знаком интеграла Лебега)

$$Mf\left(\frac{\mu}{n}\right) \to f(p),$$

т.е.

$$f(p) = \lim_{n\to\infty} Mf\left(\frac{\mu}{n}\right) =$$

$$= \lim_{n\to\infty} \sum_{k=0}^{n} f\left(\frac{k}{n}\right) C_n^k p^k (1-p)^k = \lim_{n\to\infty} B_n(p),$$

где

$$B_n(p) = \sum_{k=0}^{n} f\left(\frac{k}{n}\right) C_n^k p^k (1-p)^k \qquad -$$

так называемые многочлены Бернштейна. Итак, последовательность многочленов Бернштейна сходится к непрерывной функции на $[0,1]$. Для доказательства того, что эта сходимость равномерная, необходимо более тонкое исследование. Так как $f(p)$ равномерно непрерывна на $[0,1]$, то $\forall \varepsilon > 0 \ \exists \ \delta > 0$ такое, что если $x, y \in [0,1]$, $|x-y| < \delta$, то $|f(x) - f(y)| < \varepsilon$.

Используя бином Ньютона и оценку (16.1), получим:

$$|f(p) - B_n(p)| = \Big| \sum_{k=0}^{n} \Big[f(p) - f\left(\frac{k}{n}\right) \Big] C_n^k p^k q^{n-k} \Big| \leq$$

$$\leq \sum_{k:\ |k/n-p|<\delta} \Big| f(p) - f\left(\frac{k}{n}\right) \Big| C_n^k p^k q^{n-k} +$$

$$+ \sum_{k:\ |k/n-p|\geq\delta} \Big| f(p) - f\left(\frac{k}{n}\right) \Big| C_n^k p^k q^{n-k} \leq \varepsilon \sum_{k=0}^{n} C_n^k p^k q^{n-k} +$$

$$+2C \sum_{k:\ |k/n-p|\geq\delta} C_n^k p^k q^{n-k} =$$

$$= \varepsilon + 2CP\Big\{ \Big| \frac{\mu}{n} - p \Big| \geq \delta \Big\} \leq \varepsilon + \frac{C}{2n\delta^2}. \tag{16.3}$$

Сначала по $\varepsilon > 0$ находим $\delta > 0$, а затем по ε и δ найдем такое N, чтобы $\forall n > N$ выполнялось $\frac{C}{2n\delta^2} < \varepsilon$. Тогда из (16.3) $\forall \varepsilon > 0 \ \exists N$ такое, что $\forall n > N$ выполняется

$$\sup_{p\in[0,1]} |f(p) - B_n(p)| < 2\varepsilon.$$

Это и означает равномерную сходимость $B_n(p)$ к $f(p)$.

Таким чисто вероятностным методом Бернштейном была не только передоказана теорема Вейерштрасса, но и предложены конструктивно сами многочлены, аппроксимирующие непрерывную функцию.

§ 17. Характеристические преобразования законов распределения

К характеристическим преобразованиям законов распределения (преобразованиям случайных величин) относятся такие преобразования распределений случайной величины, которые осуществляют биекцию множества распределений и множества объектов, в которые переходят распределения при данном преобразовании. В частности, преобразование от данного распределения является законом распределения. Преобразования вводятся для эффективного исследования сумм независимых случайных величин. Обычно их вводят таким образом, чтобы преобразование суммы независимых случайных величин равнялось произведению преобразований слагаемых, что дает громадное преимущество этих методов по сравнению с прямыми методами, требующими применения формулы свертки. Большинство преобразований осуществляют взаимно-непрерывную биекцию в том смысле, что если $L(\xi)$ — преобразование ξ и выбраны определенные виды сходимости в множестве случайных величин (или распределений) и в множестве преобразований, то 1) если $\xi_n \to \xi$, то $L(\xi_n) \to L(\xi)$; 2) если $L_n \to L$, то $\xi(L_n) \to \xi(L)$.

17.1. Комплекснозначные случайные величины.

В этом разделе случайные величины, которые мы до сих пор рассматривали, будем называть действительными случайными величинами.

Определение 17.1. **Комплекснозначной случайной величиной** называется отображение $\xi : \Omega \to \mathbb{C}$, ставящее в соответствие каждому $\omega \in \Omega$ комплексное число $\xi(\omega) = \xi_1(\omega) + i\xi_2(\omega)$, где $i = \sqrt{-1}$; $\xi_1 = \xi_1(\omega)$, $\xi_2 = \xi_2(\omega)$ — действительные случайные величины. При этом ξ_1 и ξ_2 называют **частями** ξ (**действительной** и **мнимой** соответственно).

Определение 17.2. **Математическим ожиданием** комплекснозначной случайной величины $\xi = \xi_1 + i\xi_2$ называется число $M\xi = M\xi_1 + iM\xi_2$.

Нетрудно понять, что все свойства математического ожидания, кроме свойств, связанных с отношением порядка в множестве действительных чисел, сохраняются и для рассматриваемого случая. Более того, сохраняются те свойства, связанные с отношением порядка, в которые входят не сами случайные величины, а их модули. Например,

$$|M\xi| \le M|\xi|.$$

Задача 1. Докажите это неравенство.

Определение 17.3. Комплекснозначные случайные величины $\xi_1, \xi_2, \dots, \xi_n$ называются **независимыми**, если любые части их являются независимыми между собой случайными величинами.

Доказанное ранее **мультипликативное свойство** математического ожидания сохраняется и в комплексном случае:

Если комплекснозначные случайные величины $\xi_1, \xi_2, \ldots, \xi_n$ *независимы, то*

$$M(\xi_1 \xi_2 \ldots \xi_n) = M\xi_1 M\xi_2 \ldots M\xi_n.$$

Д о к а з а т е л ь с т в о. Достаточно рассмотреть случай $n = 2$, а затем воспользоваться индукцией. Пусть $\xi = \xi_1 + i\xi_2$ и $\eta = \eta_1 + i\eta_2$ независимы. По определению 17.3 это означает, что независимы между собой ξ_1 и η_1, ξ_1 и η_2, ξ_2 и η_1, ξ_2 и η_2. Применяя мультипликативное свойство к действительным случайным величинам, получим:

$$M(\xi\eta) = M[(\xi_1 + i\xi_2)(\eta_1 + i\eta_2)] = M[(\xi_1\eta_1 - \xi_2\eta_2) + i(\xi_1\eta_2 +$$

$$+\xi_2\eta_1)] = M(\xi_1\eta_1 - \xi_2\eta_2) + iM(\xi_1\eta_2 + \xi_2\eta_1) =$$

$$= M\xi_1 M\eta_1 - M\xi_2 M\eta_2 + iM\xi_1 M\eta_2 + iM\xi_2 M\eta_1 =$$

$$= (M\xi_1 + iM\xi_2)(M\eta_1 + iM\eta_2) = M\xi \cdot M\eta.$$

$\square$

Определение 17.4. Комплекснозначная функция $f : \mathbb{R} \to \mathbb{C}$, ставящая в соответствие действительному числу x комплексное число $f(x) = f_1(x) + if_2(x)$, где f_1, f_2 — действительные борелевские функции, называется **комплекснозначной борелевской функцией**.

Например, функции $f_1(x) = z^x = |z|^x[\cos(x \cdot \arg z) + i \sin(x \cdot \arg z)]$ и $f_2(x) = e^{itx} = \cos tx + i \sin tx$ — комплекснозначные борелевские функции, так как обе части этих функций являются непрерывными действительными, а, значит, борелевскими функциями. Если f — комплекснозначная борелевская функция, ξ — действительная случайная величина, то, очевидно, $f(\xi)$ — комплекснозначная случайная величина.

Как и для действительного случая имеет место

Теорема о борелевских функциях от независимых случайных величин. *Если действительные случайные величины $\xi_1, \xi_2, \dots, \xi_n$ независимы, $f_1, f_2, \dots, f_n$ — комплекснозначные борелевские функции, то комплекснозначные случайные величины $f_1(\xi_1), f_2(\xi_2), \dots, f_n(\xi_n)$ независимы.*

Доказательство. Любые части комплекснозначных случайных величин $f_1(\xi_1), f_2(\xi_2), \dots, f_n(\xi_n)$ независимы как действительные борелевские функции от независимых действительных случайных величин. По определению 17.3 $f_1(\xi_1), f_2(\xi_2), \dots, f_n(\xi_n)$ независимы.

□

17.2. Производящие функции.

Пусть ξ — неотрицательная целочисленная случайная величина, т.е. величина, принимающая значения n с вероятностями p_n $(n = 0, 1, 2, \dots)$.

Определение 17.5. **Производящей функцией целочисленной случайной величины** ξ или **вероят-**

ностей p_n, $n = 0, 1, \dots$ называется функция

$$P_\xi(z) = P(z) = Mz^\xi = \sum_{n=0}^{\infty} p_n z^n.$$

Можно считать, в зависимости от удобства, производящую функцию функцией комплексной или действительной переменной. Для определенности в дальнейшем считаем $P(z)$ функцией комплексной переменной. Из определения следует, что в круге сходимости она является аналитической функцией. Поскольку $|p_n z^n| \le p_n$ при $|z| \le 1$, а $\sum\limits_{n=0}^{\infty} p_n = 1 < \infty$, то ряд, определяющий производящую функцию, сходится абсолютно и равномерно в единичном круге. Значит, производящая функция определена по крайней мере для $|z| \le 1$ и заведомо аналитична в открытом круге $|z| < 1$.

Свойства.

1. *При* $|z| \le 1$ *выполняется* $|P(z)| \le 1$, *причем* $P(1) = 1$.

Действительно, при $|z| \le 1$ имеем $|z^\xi| \le 1$, следовательно, $|P(z)| = |Mz^\xi| \le M|z^\xi| \le M1 = 1$. Кроме того, $P(1) = M1 = 1$.

2. *Производящая функция является законом распределения.*

Действительно, коэффициенты p_n степенного ряда, определяющего $P(z)$, однозначно находятся как

$$p_n = \frac{1}{n!} P^{(n)}(0).$$

Следовательно, по $P(z)$ однозначно восстанавливается ряд распределения

ξ	0	1	2	......
P	p_0	p_1	p_2	......

.

3. *Если $\xi_1, \xi_2, \ldots, \xi_n$ — независимые целочисленные случайные величины, то*

$$P_{\xi_1+\xi_2+\ldots+\xi_n}(z) = P_{\xi_1}(z)P_{\xi_2}(z)\ldots P_{\xi_n}(z).$$

Очевидно, $\xi_1+\xi_2+\ldots+\xi_n$ — целочисленная случайная величина. Функция $f(x) = z^x$, как мы видели в 17.1, является комплекснозначной борелевской функцией от x. Так как $\xi_1, \xi_2, \ldots, \xi_n$ независимы, то и $z^{\xi_1}, z^{\xi_2}, \ldots, z^{\xi_n}$ независимы. По свойствам математического ожидания

$$P_{\xi_1+\xi_2+\ldots+\xi_n}(z) = Mz^{\xi_1+\xi_2+\ldots+\xi_n} = M[z^{\xi_1}z^{\xi_2}\ldots z^{\xi_n}] =$$

$$= Mz^{\xi_1}Mz^{\xi_2}\ldots Mz^{\xi_n} = P_{\xi_1}(z)P_{\xi_2}(z)\ldots P_{\xi_n}(z).$$

4. $M\xi = P'(1), \quad D\xi = P''(1) + P'(1) - [P'(1)]^2,$ *если они существуют.*

Известно, что степенной ряд можно почленно дифференцировать в круге сходимости; поэтому

$$P'(z) = \sum_{n=0}^{\infty} np_n z^{n-1}, \tag{17.1}$$

$$P''(z) = \sum_{n=0}^{\infty} n(n-1)p_n z^{n-2}. \tag{17.2}$$

Если окажется, что радиус сходимости строго больше 1, то в этих равенствах можно сразу взять $z = 1$. Если радиус сходимости равен 1, то при условии существования конечного $M\xi = \sum\limits_{n=0}^{\infty} np_n$ ряд (17.1) сходится в точке $z = 1$ и по теореме Абеля $P'(1)$ равно пределу $P'(z)$, когда z стремится к 1, оставаясь внутри единичного круга. Значит и в этом случае в (17.1) можно положить $z = 1$. Точно так же обосновывается возможность положить $z = 1$ в (17.2), если существует конечная $D\xi$. Полагая в (17.1) и (17.2) $z = 1$, получим

$$P'(1) = \sum_{n=0}^{\infty} np_n = M\xi,$$

$$P''(1) = \sum_{n=0}^{\infty} n(n-1)p_n = M[\xi(\xi-1)] = M\xi^2 - M\xi.$$

Отсюда и следуют доказываемые равенства.

5. Мы ввели понятие производящей функции целочисленной случайной величины или вероятностей p_k, $k = 0, 1, 2, \dots$; $\sum\limits_{k-0}^{\infty} p_k = 1$. Точно также можно ввести производящую функцию $\sum\limits_{k=0}^{\infty} p_k z^k$ числовой последовательности (p_k).

Теорема непрерывности для производящих функций. *Пусть* $(P_n(z) = \sum\limits_{k=0}^{\infty} p_k^{(n)} z^k,\ n = 1, 2, \dots)$ — *последовательность производящих функций вероятностей,* $P(z) = \sum\limits_{k=0}^{\infty} p_k z^k$ — *производящая функция последовательности. Для того, чтобы все* $\lim\limits_{n\to\infty} p_k^{(n)} = p_k$, $k = 0, 1, 2, \dots$, *необходимо и достаточно, чтобы при всех*

$z, \quad |z| < 1$

$$\lim_{n\to\infty} P_n(z) = P(z).$$

Доказательство теоремы выносится в приложение.

Замечание 1. Пусть $\xi_n(\omega) \equiv n, \quad n = 1, 2, \ldots$, тогда $p_k^{(n)} = P\{\xi_n = k\} = \delta_{nk}, \quad P_n(z) = P_{\xi_n}(z) = z^n \to 0 = P(z)$, когда $|z| < 1$. Предельные вероятности $p_k = \lim\limits_{n\to\infty} \delta_{nk} = 0$ для всех $k = 0, 1, \ldots$, т.е. не образуют распределение вероятностей ($\sum\limits_{k=0}^{\infty} p_k = 0 \neq 1$). В общем случае, очевидно, $\sum\limits_{k=0}^{\infty} p_k \leq 1$. Если потребовать, чтобы $\lim\limits_{z\to 1-0} P(z) = 1$, то $\sum\limits_{k=0}^{\infty} p_k = 1$ и в пределе получается распределение вероятностей.

Случай, когда $\sum\limits_{k=0}^{\infty} p_k = p < 1$, можно интерпретировать следующим образом: предельная случайная величина является несобственной, т.е. с вероятностью $1 - p$ она принимает несобственное значение $+\infty$. В рассмотренном выше примере предельная величина принимает значение $+\infty$ с вероятностью 1. Несобственная случайная величина в отличие от случайной величины может принимать значения в расширенной числовой прямой $R \cup \{+\infty\}$ и для нее, вообще говоря, $F(+\infty) = p < 1$. Иногда рассматриваются несобственные случайные величины, принимаюшие значения в $R \cup \{-\infty\} \cup \{+\infty\}$.

Итак, между законами распределения целочисленных случайных величин и производящими функциями установлено взаимно-однозначное и взаимно-непрерывное соответствие.

Примеры.

1. Если μ — число успехов в n испытаниях Бернулли с вероятностью успеха p, то $\mu = \sum\limits_{k=0}^{n} \mu_k$, где μ_k — число успехов в k-м испытании. Очевидно, $p_0 = P\{\mu_k = 0\} = 1 - p = q, \quad p_1 = P\{\mu_k = 1\} = p, \quad p_l = 0,\ l \geq 2$, поэтому

$$P_{\mu_k}(z) = q + pz, \quad k = \overline{1, n}.$$

Так как $\mu_1, \mu_2, \dots, \mu_n$ независимы, то по свойству 3 производящих функций

$$P_\mu(z) = P_{\mu_1}(z) P_{\mu_2}(z) \dots P_{\mu_n}(z) = (pz + q)^n.$$

Отметим, что в данном случае область определения производящей функции — вся комплексная плоскость $\mathbb{C}$. Так как

$$P'_\mu(z) = np(pz+q)^{n-1}, \quad P'_\mu(1) = np,$$

$$P''_\mu(z) = n(n-1)p^2(pz+q)^{n-2}, \quad P''_\mu(1) = n(n-1)p^2,$$

то, в силу свойства 4 производящих функций,

$$M\mu = np, \quad D\mu = npq.$$

2. Пусть ξ имеет распределение Пуассона с параметром λ:

$$p_n = P\{\xi = n\} = \frac{\lambda^n}{n!} e^{-\lambda}, \quad n = 0, 1, \dots .$$

Тогда

$$P(z) = P_\xi(z) = \sum_{n=0}^{\infty} \frac{\lambda^n}{n!} e^{-\lambda} z^n = e^{-\lambda} \sum_{n=0}^{\infty} \frac{(\lambda z)^n}{n!} =$$

$$= e^{-\lambda} e^{\lambda z} = e^{\lambda(z-1)}.$$

Поскольку ряд Тейлора для экспоненты сходится к ней во всей комплексной плоскости, то область определения производящей функции — вся комплексная плоскость $\mathbb{C}$. Так как

$$P'(z) = \lambda e^{\lambda(z-1)}, \quad P'(1) = \lambda,$$
$$P''(z) = \lambda^2 e^{\lambda(z-1)}, \quad P''(1) = \lambda^2,$$

то, в силу свойства 4 производящих функций,

$$M\xi = D\xi = \lambda.$$

Производящая функция удобнее для нахождения распределения сумм независимых неотрицательных целочисленных случайных величин, чем формула свертки. Пусть, например, ξ и η — независимые случайные величины, имеющие распределения Пуассона с параметрами λ и μ соответственно. Тогда $P_\xi(z) = e^{\lambda(z-1)}$, $P_\eta(z) = e^{\mu(z-1)}$. По свойству 3 $P_{\xi+\eta}(z) = P_\xi(z) P_\eta(z) = e^{(\lambda+\mu)(z-1)}$. Но это — производящая функция распределения Пуассона с параметром $\lambda + \mu$. На основании свойства 2 $\xi + \eta$ имеет распределение Пуассона с параметром $\lambda + \mu$.

17.3. Характеристические функции.

Определение 17.6. Характеристической функцией случайной величины ξ или распределения P_ξ называется

$$f_\xi(t) = f(t) = Me^{it\xi} = \int\limits_\Omega e^{it\xi(\omega)} dP,$$

где $i = \sqrt{-1}$.

Здесь интеграл от комплекснозначной функции определяется обычным образом, таким же, как математическое ожидание от комплекснозначной случайной величины:

$$\int\limits_{\Omega} e^{it\xi(\omega)}dP = \int\limits_{\Omega} \cos t\xi(\omega)dP + i\int\limits_{\Omega} \sin t\xi(\omega)dP.$$

Производя замену переменной $\xi(\omega) = x$, соответствующую измеримому отображению $\xi: \ (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B})$, получим:

$$f(t) = \int\limits_{\mathbb{R}} e^{itx}dP_{\xi} = \int\limits_{\mathbb{R}} e^{itx}dF_{\xi}(x) = \int\limits_{-\infty}^{+\infty} e^{itx}dF_{\xi}(x). \quad (17.3)$$

Последний интеграл существует как абсолютно сходящийся интеграл Римана-Стилтьеса для любой случайной величины, ибо $|e^{itx}| = 1$, а

$$\int\limits_{-\infty}^{+\infty} 1dF_{\xi}(x) = F_{\xi}(+\infty) - F_{\xi}(-\infty) = 1.$$

Таким образом, в отличие от производящей функции, характеризующей только неотрицательные целочисленные величины, характеристическая функция является преобразованием универсальным, пригодным для любой случайной величины.

Свойства характеристических функций.

1. $f(0) = 1$.

В самом деле, $f(0) = Me^{i\cdot 0\cdot \xi} = Me^0 = M1 = 1$.

2. $|f(t)| \leq 1$.

Действительно, $|f(t)| = |Me^{it\xi}| \leq M|e^{it\xi}| = M1 = 1$.

3. $f_{a\xi+b}(t) = e^{itb} f_\xi(at)$.

Имеем, $f_{a\xi+b}(t) = Me^{it(a\xi+b)} = M[e^{itb}e^{i(at)\xi}] = e^{itb}Me^{i(at)\xi} = e^{itb}f_\xi(at)$.

4. *Характеристическая функция равномерно непрерывна на всей числовой прямой.*

Это свойство в дальнейшем не используется, поэтому рекомендуется решить следующую задачу.

Задача 2. Докажите свойство 4.

5. *Если случайные величины $\xi_1, \xi_2, \ldots, \xi_n$ независимы, то*

$$f_{\sum\limits_{k=1}^{n}\xi_k}(t) = \prod_{k=1}^{n} f_{\xi_k}(t).$$

В самом деле, так как действительные случайные величины $\xi_1, \xi_2, \ldots, \xi_n$ независимы, а комплекснозначная функция $g(x) = e^{itx}$ борелевская, то комплекснозначные случайные величины $e^{it\xi_1}$, $e^{it\xi_2}$, $\ldots$, $e^{it\xi_n}$ независимы. Используя мультипликативное свойство математических ожиданий комплекснозначных случайных величин, получим:

$$f_{\sum\limits_{k=1}^{n}\xi_k}(t) = Me^{it\sum\limits_{k=1}^{n}\xi_k} = M\prod_{k=1}^{n} e^{it\xi_k} =$$

$$= \prod_{k=1}^{n} Me^{it\xi_k} = \prod_{k=1}^{n} f_{\xi_k}(t).$$

6. *Если для $k \geq 1$ существует конечный момент k-го порядка (т.е. $M|\xi|^k < \infty$), то существует непрерывная*

производная k-го порядка от характеристической функции, причем

$$M\xi^k = (-i)^k f^{(k)}(0).$$

На основании неравенства Ляпунова (свойство 7 математического ожидания) из существования $M|\xi|^k < \infty$ вытекает существование $M|\xi|^l < \infty$ для всех $l < k$. Дифференцируя l раз по параметру t равенство

$$f(t) = \int\limits_{-\infty}^{+\infty} e^{itx} dF(x),$$

получим:

$$f^{(l)}(t) = i^l \int\limits_{-\infty}^{+\infty} x^l e^{itx} dF(x), \qquad l = \overline{1,k}.$$

Каждый из этих интегралов существует как абсолютно сходящийся интеграл Римана-Стилтьеса. Известно, что дифференцирование по параметру под знаком несобственного интеграла законно, если сам интеграл сходится, а интеграл от производной по параметру сходится равномерно. В данном случае при любом $l = \overline{1,k}$ все интегралы сходятся равномерно по признаку Вейерштрасса, поскольку

$$|x^l e^{itx}| \le |x|^l, \quad \text{а} \quad \int\limits_{-\infty}^{+\infty} |x|^l dF(x) = M|\xi|^l < \infty.$$

В частности, при $l = k$

$$f^{(k)}(t) = i^k \int\limits_{-\infty}^{+\infty} x^k e^{itx} dF(x).$$

Так как функция $x^k e^{itx}$ непрерывна по паре переменных (t, x), то последний интеграл (т.е. $f^{(k)}(t)$) непрерывен по параметру t. Наконец, полагая в последнем равенстве $t = 0$, получим

$$f^{(k)}(0) = i^k \int\limits_{-\infty}^{+\infty} x^k dF(x) = i^k M\xi^k,$$

откуда $M\xi^k = (-i)^k f^{(k)}(0)$.

7. *Если $M|\xi|^k < \infty$, то в окрестности точки $t = 0$*

$$f(t) = \sum_{l=0}^{k} \frac{(it)^l}{l!} M\xi^l + o(|t|^k).$$

(здесь нулевой член суммы равен 1).

По формуле Тейлора с остаточным членом в форме Пеано

$$f(t) = \sum_{l=0}^{k} \frac{f^{(l)}(0)}{l!} t^l + o(|t|^k).$$

Остается подставить сюда $f^{(k)}(0) = i^k M\xi^k$.

8. $\overline{f(t)} = f(-t)$, *в частности, если характеристическая функция действительна, то она четная (здесь черта обозначает операцию комплексного сопряжения).*

Действительно, $\overline{f(t)} = \overline{Me^{it\xi}} = M\overline{e^{it\xi}} = Me^{-it\xi} = Me^{i(-t)\xi} = f(-t)$.

Заметим, что если ξ — неотрицательная целочисленная случайная величина с производящей функцией $P(z)$, то ее характеристическая функция равна

$$f(t) = Me^{it\xi} = M\left(e^{it}\right)^{\xi} = P\left(e^{it}\right). \qquad (17.4)$$

Если ξ — дискретная случайная величина, принимающая значения x_k с вероятностями p_k, то для нее формула (17.3) превращается в

$$f(t) = \sum_k e^{itx_k} p_k, \tag{17.5}$$

а если ξ — абсолютно непрерывная случайная величина с плотностью распределения вероятностей $p(x)$, то — в

$$f(t) = \int_{-\infty}^{+\infty} e^{itx} p(x) dx. \tag{17.6}$$

Примеры.

3. Для случайной величины μ, имеющей биномиальное распределение с параметрами (n, p), ранее была найдена $P(z) = (pz + q)^n$. По формуле (17.4) ее характеристическая функция равна

$$f(t) = \left(pe^{it} + q\right)^n.$$

4. Для случайной величины ξ, имеющей распределение Пуассона с параметром λ, производящая функция равна $P(z) = e^{\lambda(z-1)}$, поэтому

$$f(t) = e^{\lambda\left(e^{it}-1\right)}.$$

5. Пусть ξ_0 имеет стандартное нормальное распределение. По формуле (17.6) ее характеристическая функция равна

$$f(t) = \int_{-\infty}^{+\infty} e^{itx} \varphi(x) dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{itx} e^{-\frac{x^2}{2}} dx =$$

$$= \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{+\infty} \cos(tx) e^{-\frac{x^2}{2}} dx. \tag{17.7}$$

Здесь мы воспользовались формулой Эйлера и тем, что $\frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{+\infty} \sin(tx) e^{-\frac{x^2}{2}} dx = 0$ как сходящийся интеграл от нечетной функции в симметричных пределах. Дифференцируя (17.7) и интегрируя по частям, получим:

$$f'(t) = -\frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{+\infty} x \sin(tx) e^{-\frac{x^2}{2}} dx =$$

$$= \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{+\infty} \sin(tx) d\left(e^{-\frac{x^2}{2}}\right) =$$

$$= \frac{1}{\sqrt{2\pi}} \sin(tx) e^{-\frac{x^2}{2}} \Big|_{-\infty}^{+\infty} - \frac{t}{\sqrt{2\pi}} \int\limits_{-\infty}^{+\infty} \cos(tx) e^{-\frac{x^2}{2}} dx = -tf(t).$$

Законность дифференцирования по параметру t под знаком сходящегося интеграла следует из равномерной сходимости интеграла от $x \sin(tx) e^{-\frac{x^2}{2}}$ (ведь $|x \sin(tx) e^{-\frac{x^2}{2}}| \leq |x| e^{-\frac{x^2}{2}}$, а $M|\xi_0| < \infty$). Общее решение полученного дифференциального уравнения имеет вид $f(t) = Ce^{-\frac{t^2}{2}}$, но из начального условия $f(0) = 1$ (свойство 1) получается $C = 1$. Таким образом,

$$f_{\xi_0}(t) = e^{-\frac{t^2}{2}}.$$

6. Пусть $\xi \sim N(a, \sigma^2)$. По лемме о нормальном распределении она представляется в виде $\xi = \sigma\xi_0 + a$, где

$\xi_0 \sim N(0,1)$. По свойству 3

$$f_\xi(t) = e^{iat} f_{\xi_0}(\sigma t) = e^{iat - \frac{\sigma^2 t^2}{2}}.$$

Формула (17.6) показывает, что для абсолютно непрерывной случайной величины характеристическая функция является преобразованием Фурье от плотности вероятности. Из функционального анализа известно, что если $f \in L_1$, то плотность вероятности однозначно восстанавливаетя по характеристической функции как обратное преобразование Фурье:

$$p(x) = \frac{1}{2\pi} \int\limits_{-\infty}^{+\infty} e^{-itx} f(t) dt.$$

Так как функция распределения $F(x)$ в формуле (17.3) стоит под знаком дифференциала, а форма интеграла такая же, как в (17.6), то естественно назвать характеристическую функцию $f(t)$ преобразованием Фурье-Стилтьеса от функции распределения $F(x)$. Ожидается, что и в общем случае функция распределения однозначно восстанавливается по характеристической функции.

Лемма 17.7. $\int\limits_A^B \frac{\sin u}{u} du \to \pi$ *и равномерно по* a *и* b $(a < b)$ *ограничен.*

Д о к а з а т е л ь с т в о. Первая часть следует из известного из математического анализа факта, что интеграл Дирихле

$$\int\limits_{-\infty}^{+\infty} \frac{\sin u}{u} du = \pi.$$

Далее,

$$\int\limits_0^B \frac{\sin u}{u} du = \int\limits_0^{\pi} \frac{\sin u}{u} du + \int\limits_{\pi}^{2\pi} \frac{\sin u}{u} du + \dots$$

$$\dots + \int\limits_{\pi\left\{[B/\pi]-1\right\}}^{\pi[B/\pi]} \frac{\sin u}{u} du + \int\limits_{\pi[B/\pi]}^{B} \frac{\sin u}{u} du.$$

Эти интегралы убывают по абсолютной величине, а знаки их чередуются, начиная со знака "плюс" . Поэтому

$$\int\limits_0^B \frac{\sin u}{u} du \leq \int\limits_0^{\pi} \frac{\sin u}{u} du = C_1.$$

Аналогично,

$$\int\limits_A^0 \frac{\sin u}{u} du \leq \int\limits_{-\pi}^{0} \frac{\sin u}{u} du = C_1.$$

Поэтому

$$0 < \int\limits_A^B \frac{\sin u}{u} du \leq C = 2C_1.$$

$\square$

Теорема обращения для характеристических функций. *Пусть $F(x)$ — функция распределения, $f(t)$ — соответствующая характеристическая функция.*

1. Для любых действительных a и b $(a < b)$, в которых F непрерывна,

$$F(b) - F(a) = \frac{1}{2\pi} \lim_{N\to\infty} \int\limits_{-N}^{N} \frac{e^{-ita} - e^{-itb}}{it} f(t)dt.$$

2. В частности, если $f(t) \in L_1$, то $F(x)$ абсолютно непрерывна, причем соответствующая плотность вероятности равна

$$p(x) = \frac{1}{2\pi} \int\limits_{-\infty}^{+\infty} e^{-itx} f(t)dt.$$

Д о к а з а т е л ь с т в о. 1. Обозначим

$$I_N = \frac{1}{2\pi} \int\limits_{-N}^{N} \frac{e^{-ita} - e^{-itb}}{it} f(t)dt.$$

Подставляя сюда (17.3) и меняя порядок интегрирования, получим:

$$I_N = \frac{1}{2\pi} \int\limits_{-N}^{N} \int\limits_{-\infty}^{+\infty} \frac{e^{it(x-a)} - e^{it(x-b)}}{it} dF(x)dt =$$

$$= \frac{1}{2\pi} \int\limits_{-\infty}^{+\infty} \Big[\int\limits_{-N}^{N} \frac{e^{it(x-a)} - e^{it(x-b)}}{it} dt \Big] dF(x) =$$

$$= \frac{1}{2\pi} \int\limits_{-\infty}^{+\infty} \Big[\int\limits_{-N}^{N} \frac{sint(x-a) - sint(x-b)}{t} dt \Big] dF(x) =$$

$$= \frac{1}{2\pi} \int\limits_{-\infty}^{+\infty} \Big[\int\limits_{-N(x-a)}^{N(x-a)} \frac{sinu}{u} du - \int\limits_{-N(x-b)}^{N(x-b)} \frac{sinu}{u} du \Big] dF(x).$$

Законность изменения порядка интегрирования следует из теоремы Фубини, ибо

$$\Big| \frac{e^{-ita} - e^{-itb}}{it} e^{itx} \Big| = \Big| \int\limits_{a}^{b} e^{-itx} dx \Big| \leq b - a,$$

а

$$\int\limits_{-N}^{N} \int\limits_{-\infty}^{\infty} (b-a) dF(x) dt \leq 2N(b-a) < \infty.$$

Преобразуя I_N, получим:

$$I_N = \frac{1}{2\pi} \int\limits_{-\infty}^{+\infty} \Big[\int\limits_{-N(x-a)}^{-N(x-b)} \frac{sinu}{u} du + \int\limits_{N(x-b)}^{N(x-a)} \frac{sinu}{u} du \Big] dF(x) =$$

$$= \frac{1}{\pi} \int\limits_{-\infty}^{+\infty} \Big[\int\limits_{N(x-b)}^{N(x-a)} \frac{sinu}{u} du \Big] dF(x).$$

По лемме 17.7 интеграл

$$\int\limits_{N(x-b)}^{N(x-a)} \frac{sinu}{u} du \to g(x) = \begin{cases} \pi, & \text{если } a < x < b, \\ \frac{\pi}{2}, & \text{если } x = a \text{ или } x = b, \\ 0, & \text{если } x < a \text{ или } x > b. \end{cases}$$

и ограничен равномерно. По теореме Лебега о предельном переходе в интеграле Лебега

$$\lim_{N\to\infty} I_N = \frac{1}{\pi}\int\limits_{-\infty}^{+\infty} g(x)dF(x) = \frac{1}{\pi}\int\limits_{-\infty}^{+\infty} g(x)dP_\xi =$$

$$= \frac{1}{\pi}[\pi P_\xi((a,b)) + \frac{\pi}{2}P_\xi(\{a\}) + \frac{\pi}{2}P_\xi(\{b\})] =$$

$$= P_\xi((a,b)) + \frac{1}{2}(P_\xi(\{a\}) + P_\xi(\{b\})) = F(b) - F(a),$$

поскольку a, b — точки непрерывности F, для которых $P_\xi((a,b)) = F(b) - F(a)$, $P_\xi(\{a\}) = P_\xi(\{b\}) = 0$.

2. Пусть $\int\limits_{-\infty}^{\infty} |f(t)|dt < \infty$. Обозначим

$$p(x) = \frac{1}{2\pi}\int\limits_{-\infty}^{\infty} e^{-itx} f(t)dt.$$

По теореме Лебега о предельном переходе под знаком интеграла Лебега при каждом фиксированном $x_0 \in \mathbb{R}$ выполняется $p(x) \to p(x_0)$ при $x \to x_0$, т.е. $p(x)$ непрерывная на $\mathbb{R}$ функция. Значит она интегрируема на $[a, b]$. Если a, b — точки непрерывности F, то, применяя снова теорему Фубини, получим:

$$\int\limits_a^b p(x)dx = \frac{1}{2\pi}\int\limits_{-\infty}^{\infty} f(t)\Big[\int\limits_a^b e^{-itx}dx\Big]dt =$$

$$= \frac{1}{2\pi}\lim_{N\to\infty}\int\limits_{-N}^{N} \frac{e^{-ita} - e^{-itb}}{it} f(t)dt = F(b) - F(a).$$

Так как F монотонна, то она имеет не более чем счетное множество точек разрыва. Поэтому множество ее точек непрерывности всюду плотно на расширенной числовой прямой $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty\} \cup \{+\infty\}$. Если x — произвольная точка $\mathbb{R}$, то переходя в последнем равенстве к пределу, когда $a \to -\infty$, $b \to x - 0$, причем a, b остаются точками непрерывности F, получим, что

$$\int\limits_{-\infty}^{x} p(x)dx = F(x),$$

т.е. $F(x)$ абсолютно непрерывна и $p(x)$ — соответствующая плотность распределения вероятностей.

$\square$

Следствие 17.8. *Характеристическая функция является законом распределения случайной величины. Другими словами, она однозначно определяет функцию распределения.*

Д о к а з а т е л ь с т в о. Пусть $x \in \mathbb{R}$. Если (a_n), (b_n) — произвольные последовательности точек непрерывности F, такие что $a_n \to -\infty$, $b_n \to x - 0$, то значение функции распределения в точке x однозначно определяется разностями значений функции распределения в точках ее непрерывности как

$$F(x) = \lim_{n\to\infty} [F(b_n) - F(a_n)],$$

а в свою очередь эти разности по формуле обращения однозначно определяются характеристической функцией.

$\square$

Характеристическая функция удобна для нахождения распределений сумм независимых случайных величин.

Примеры.

7. Пусть ξ и η независимы, $\xi \sim N(a, \sigma^2)$, $\eta \sim N(b, \tau^2)$. Тогда

$$f_\xi(t) = e^{ita - \frac{\sigma^2 t^2}{2}}, \qquad f_\eta(t) = e^{itb - \frac{\tau^2 t^2}{2}}.$$

На основании свойства 5 характеристических функций

$$f_{\xi+\eta}(t) = f_\xi(t) f_\eta(t) = e^{it(a+b) - \frac{(\sigma^2+\tau^2)t^2}{2}}.$$

Но это — характеристическая функция нормального распределения с параметрами $(a + b, \sigma^2 + \tau^2)$. В силу следствия 17.8 это означает, что $\xi + \eta \sim N(a + b, \sigma^2 + \tau^2)$.

Крамеру удалось доказать более сложное обратное утверждение.

Теорема Крамера. *Если сумма двух независимых случайных величин имеет нормальное распределение, то каждая из этих случайных величин имеет нормальное распределение.*

8. Пусть ξ и η независимы и имеют распределение Пуассона с параметрами λ и μ соответственно. Тогда $f_\xi(t) = e^{\lambda\left(e^{it}-1\right)}$, $f_\eta(t) = e^{\mu\left(e^{it}-1\right)}$. По свойству 5

$$f_{\xi+\eta}(t) = f_\xi(t) f_\eta(t) = e^{(\lambda+\mu)\left(e^{it}-1\right)}.$$

Но это — характеристическая функция распределения Пуассона с параметром $\lambda + \mu$. В силу следствия 17.8

$\xi + \eta$ имеет распределение Пуассона с параметром $\lambda + \mu$. Впрочем, это повторение рассуждений, которые использовались для установления этого факта с помощью производящих функций. Имеет место

Теорема Райкова. *Если сумма двух независимых случайных величин имеет распределение Пуассона, то каждая из этих случайных величин имеет распределение Пуассона.*

В заключение приведем важные для доказательства предельных теорем утверждения, устанавливающие взаимную непрерывность соответствия между функциями распределения и характеристическими функциями. При этом для функций распределения выбирается слабая сходимость, а для характеристических функций — поточечная сходимость.

Прямая теорема непрерывности для характеристических функций. *Пусть $(F_n(x))$ — последовательность функций распределения, $(f_n(t))$ — последовательность соответствующих характеристических функций, $F(x)$ — функция распределения, $f(t)$ — соответствующая характеристическая функция. Если $F_n(x) \Rightarrow F(x)$, то $f_n(t) \to f(t)$ для каждого $t \in \mathbb{R}$.*

Обратная теорема непрерывности для характеристических функций. *Пусть $(f_n(t))$ — последовательность характеристических функций, $(F_n(x))$ — соответствующая последовательность функций распределения и для каждого t $\quad f_n(t) \to f(t)$, причем $f(t)$ непрерывна в 0. Тогда $f(t)$ — характеристическая* *функция, и, если $F(x)$ — соответствующая функция распределения, то $F_n(x) \Rightarrow F(x)$.*

Доказательство теорем непрерывности выносится в приложение.

17.4. Преобразование Лапласа-Стилтьеса.

Вместо преобразования Фурье для функций, обращающихся в 0 на отрицательной полуоси, удобнее применять преобразование Лапласа. Точно так же для неотрицательных случайных величин удобнее использовать не характеристические функции, а преобразование Лапласа-Стилтьеса. Это преобразование находит разнообразные применения в актуарной математике, теории надежности, теории массового обслуживания и других прикладных дисциплинах.

Пусть ξ — неотрицательная случайная величина. Очевидно, $F_\xi(x) = 0$ для $x \leq 0$.

Определение 17.9. Преобразованием Лапласа-Стилтьеса от функции распределения $F_\xi(x)$ или **от случайной величины** ξ называется

$$\phi(s) = \phi_\xi(s) = Me^{-s\xi} = \int\limits_0^\infty e^{-sx} dF_\xi(x).$$

Аргумент s можно считать действительным или комплексным. Для определенности считаем s комплексным. Если $Re\ s \geq 0$, то $|e^{-sx}| = e^{(-Re\ s)x} \leq 1$, а $\int\limits_0^\infty dF_\xi(x) = 1$.

Таким образом, для любой неотрицательной случайной величины преобразование Лапласа-Стилтьеса определено по крайней мере в полуплоскости $Re\ s \geq 0$. Можно показать, что $\phi(s)$ является аналитической функцией по крайней мере в полуплоскости $Re\ s > 0$.

Свойства преобразования Лапласа-Стилтьеса.

1. $|\phi(s)| \leq 1$ *при* $Re\ s \geq 0$.
2. $\phi(0) = 1$.
3. *Если неотрицательные случайные величины* $\xi_1, \xi_2, \ldots, \xi_n$ *независимы, то*

$$\phi_{\xi_1+\xi_2+\ldots+\xi_n}(s) = \phi_{\xi_1}(s)\phi_{\xi_2}(s)\ldots\phi_{\xi_n}(s).$$

4. *Если существует* $M\xi^2 < \infty$*, то*

$$M\xi = -\phi'(0), \quad D\xi = \phi''(0) - [\phi'(0)]^2.$$

Доказательство — почти дословное повторение аналогичных свойств характеристических функций.

Задача 3. Докажите эти свойства.

Задача 4. Докажите, что в полуплоскости $Re\ s > 0$

$$\phi(s) = s \int\limits_0^\infty F(x)e^{-sx}dx.$$

Для преобразования Лапласа-Стилтьеса можно вывести формулу, подобную формуле обращения для характеристических функций, и установить взаимно однозначное и взаимно непрерывное соответствие между функциями распределения и преобразованиями Лапласа-Стилтьеса неотрицательных случайных величин. При этом, как и в случае характеристических функций, в множестве функций распределения выбирается слабая сходимость, а в множестве преобразований Лапласа-Стильтьеса — поточечная сходимость.

В качестве примера найдем преобразование Лапласа-Стилтьеса случайной величины ξ, имеющей распределение Эрланга k-го порядка с параметром λ. Если ξ_l, $l = \overline{1,k}$, имеют показательное распределение с параметром λ, то их преобразования Лапласа-Стилтьеса равны

$$\phi_{\xi_l}(s) = \int\limits_0^\infty e^{-sx} d(1 - e^{-\lambda x}) = \int\limits_0^\infty e^{-sx} \lambda e^{-\lambda x} dx =$$

$$= \frac{\lambda}{\lambda + s}, \qquad l = \overline{1,k}.$$

Так как $\xi = \xi_1 + \xi_2 + \ldots + \xi_k$ и слагаемые независимы, то по свойству 3

$$\phi_\xi(s) = \phi_{\xi_1}(s)\phi_{\xi_2}(s)\ldots\phi_{\xi_k}(s) = \left[\frac{\lambda}{\lambda + s}\right]^k.$$

§ 18. Центральная предельная теорема

Центральная предельная теорема является одним из высших достижений человеческого гения в теории вероятностей. Она позволяет считать во многих случаях, что при большом числе слагаемых суммы независимых случайных величин будут иметь нормальное распределение.

Попытка доказательства этой теоремы при наиболее широких предположениях оказала огромное влияние на развитие аналитических методов теории вероятностей.

Для демонстрации основных идей, применяемых в доказательствах, начнем не с этой теоремы, а установим закон больших чисел в форме Хинчина.

18.1. Закон больших чисел в форме Хинчина.

Теорема Хинчина. *Пусть* (ξ_n) *— последовательность независимых одинаково распределенных случайных величин, имеющих конечное математическое ожидание* $M\xi_n = a$. *Тогда*

$$\frac{1}{n}\sum_{k=1}^{n}\xi_k \xrightarrow{P} a.$$

Д о к а з а т е л ь с т в о. Обозначим через $f(t)$ характеристическую функцию случайных величин $\xi_1, \xi_2, \ldots$ (так как эти величины имеют одну и ту же функцию распределения, то они имеют одну и ту же характеристическую функцию). Учитывая независимость этих случайных величин и применяя свойства характеристических функций, получим:

$$f_{\frac{1}{n}\sum\limits_{k=1}^{n}\xi_k}(t) = f_{\sum\limits_{k=1}^{n}\xi_k}\left(\frac{t}{n}\right) = \left[f\left(\frac{t}{n}\right)\right]^n. \qquad (18.1)$$

Так как $M|\xi_k| < \infty$, то, применяя свойство 7 характеристических функций, при $t \to 0$ имеем

$$f(t) = 1 + ita + o(|t|).$$

Следовательно, при фиксированном $t \in \mathbb{R}$ и $n \to \infty$

$$f\left(\frac{t}{n}\right) = 1 + \frac{ita}{n} + o\left(\frac{1}{n}\right). \tag{18.2}$$

Подставляя (18.2) в (18.1) и применяя второй замечательный предел, получим, что

$$f_{\frac{1}{n}\sum\limits_{k=1}^{n}\xi_k}(t) = \left[1 + \frac{ita}{n} + o\left(\frac{1}{n}\right)\right]^n \to e^{ita} = f_a(t),$$

где $f_a(t)$ — характеристическая функция неслучайной величины a. По обратной теореме непрерывности для характеристических функций

$$F_{\frac{1}{n}\sum\limits_{k=1}^{n}\xi_k}(x) \Rightarrow F_a(x), \quad \text{т.е. } \frac{1}{n}\sum_{k=1}^{n}\xi_k \Rightarrow a.$$

По лемме о слабой сходимости к константе параграфа 15

$$\frac{1}{n}\sum_{k=1}^{n}\xi_k \xrightarrow{P} a.$$

$\square$

Сравнивая законы больших чисел в форме Чебышева и в форме Хинчина, видим, что ни один их них не вкладывается в другой. Чебышевский закон не требует одинаковой распределенности случайных величин (сильнее хинчиновского закона), но хинчиновский закон не требует существования и равномерной ограниченности дисперсий (сильнее чебышевского закона).

18.2. Центральная предельная теорема для одинаково распределенных случайных величин.

Центральная предельная теорема. *Пусть* (ξ_n) *— последовательность независимых одинаково распределенных случайных величин, имеющих конечные* $M\xi_n = a$ *и* $D\xi_n = \sigma^2$*, и пусть* $S_n = \sum\limits_{k=1}^{n} \xi_k, \quad n = 1, 2, \ldots$. *Тогда равномерно по* $x \in \mathbb{R}$

$$F_{\frac{S_n - na}{\sigma\sqrt{n}}}(x) = P\Big\{\frac{S_n - na}{\sigma\sqrt{n}} < x\Big\} \to N(x),$$

где $N(x)$ *— функция стандартного нормального распределения.*

Д о к а з а т е л ь с т в о. Не ограничивая общности будем считать, что $a = 0$ (иначе надо перейти к случайным величинам $\xi'_n = \xi_n - a$). Обозначим через $f(t)$ характеристическую функцию случайных величин $\xi_1, \xi_2, \ldots$, и, кроме того, пусть $\zeta = \frac{S_n}{\sigma\sqrt{n}}$. Учитывая независимость этих случайных величин и применяя свойства характеристических функций, получим:

$$f_{\zeta_n}(t) = f_{\S_n}\Big(\frac{t}{\sigma\sqrt{n}}\Big) = \Big[f\Big(\frac{t}{\sigma\sqrt{n}}\Big)\Big]^n. \tag{18.3}$$

Так как $M|\xi_k|^2 < \infty$, то, применяя свойство 7 характеристических функций, при $t \to 0$ имеем

$$f(t) = 1 - \frac{\sigma^2 t^2}{2} + o(|t|^2).$$

Следовательно, при фиксированном $t \in \mathbb{R}$ и $n \to \infty$

$$f\Big(\frac{t}{\sigma\sqrt{n}}\Big) = 1 - \frac{t^2}{2n} + o\Big(\frac{1}{n}\Big). \tag{18.4}$$

Подставляя (18.4) в (18.3) и применяя второй замечательный предел, получим, что

$$f_{\zeta_n}(t) = \Big[1 - \frac{t^2}{2n} + o\Big(\frac{1}{n}\Big)\Big]^n \to e^{-\frac{t^2}{2}} = f_{\xi_0}(t),$$

где $f_{\xi_0}(t)$ — характеристическая функция случайной величины $\xi_0 \sim N(0,1)$. По обратной теореме непрерывности для характеристических функций

$$F_{\zeta_n}(x) \Rightarrow N(x),$$

где $N(x)$ — функция стандартного нормального распределения. По лемме о слабой сходимости функций распределения к непрерывной функции из пятнадцатого параграфа эта сходимость равномерная. $\square$

Замечание 1. Утверждение центральной предельной теоремы можно записать в эквивалентной форме: *при $n \to \infty$ равномерно по всем α, β таким, что $-\infty \le \alpha < \beta \le +\infty$*

$$P\{\alpha \le \frac{S_n - na}{\sigma\sqrt{n}} \le \beta\} \to \int\limits_{\alpha}^{\beta} \varphi(x)dx =$$

$$= \Phi(\beta) - \Phi(\alpha) = N(\beta) - N(\alpha). \qquad (18.5)$$

Действительно, если выполнено утверждение центральной предельной теоремы, то равномерно по β при $n \to \infty$

$$P\Big\{\frac{S_n - na}{\sigma\sqrt{n}} < \beta\Big\} \to N(\beta).$$

Так как

$$\lim_{n\to\infty} P\Big\{\frac{S_n - na}{\sigma\sqrt{n}} = \beta\Big\} = \lim_{n\to\infty}\lim_{m\to\infty} P\Big\{\beta \le \frac{S_n - na}{\sigma\sqrt{n}} < \beta +$$

$$+\frac{1}{m}\Big\} = \lim_{m\to\infty}\lim_{n\to\infty} P\Big\{\beta \le \frac{S_n - na}{\sigma\sqrt{n}} < \beta + \frac{1}{m}\Big\} =$$

$$= \lim_{m\to\infty}\Big[N\Big(\beta + \frac{1}{m}\Big) - N(\beta)\Big] = 0,$$

то равномерно по β при $n \to \infty$

$$P\Big\{\frac{S_n - na}{\sigma\sqrt{n}} \le \beta\Big\} \to N(\beta).$$

Поскольку выражение в левой части не зависит от α, это стремление будет равномерным по α, β. Аналогично при $n \to \infty$ равномерно по α, β

$$P\Big\{\frac{S_n - na}{\sigma\sqrt{n}} < \alpha\Big\} \to N(\alpha),$$

Поэтому для

$$P\{\alpha \le \frac{S_n - na}{\sigma\sqrt{n}} \le \beta\} =$$

$$= P\Big\{\frac{S_n - na}{\sigma\sqrt{n}} \le \beta\Big\} - P\Big\{\frac{S_n - na}{\sigma\sqrt{n}} < \alpha\Big\}$$

выполняется (18.5).

Обратно, если выполнено (18.5), то, беря в нем $\alpha = -\infty$, $\beta = x$, получим утверждение центральной предельной теоремы.

Теперь интегральная предельная теорема Муавра-Лапласа является следствием центральной предельной теоремы. Напомним формулировку.

Интегральная теорема Муавра-Лапласа. *Пусть в схеме Бернулли* $0 < p < 1$*. Тогда равномерно по всем* a

и b таким, что $-\infty \le a < b \le +\infty$

$$\lim_{n\to\infty} P\Big\{a \le \frac{\mu - np}{\sqrt{npq}} \le b\Big\} = \int\limits_a^b \varphi(x)dx = \varPhi(b) - \varPhi(a) =$$

$$= N(b) - N(a).$$

Д о к а з а т е л ь с т в о. Если μ_k — число успехов в k-м испытании, то (μ_k) — последовательность независимых одинаково распределенных случайных величин с конечными $M\mu_k = p, \quad D\mu_k = pq$, т.е. удовлетворяет условиям центральной предельной теоремы. При этом $S_n = \mu, \ a = p, \ \sigma^2 = pq$. Остается применить центральную предельную теорему с формулировкой ее заключения в форме (18.5). $\square$

Замечание 2. В силу центральной предельной теоремы при больших n случайная величина $\frac{S_n - na}{\sigma\sqrt{n}}$ приближенно имеет стандартное нормальное распределение. По лемме о нормальном распределении случайная величина S_n имеет приближенно нормальное распределение с параметрами $(na, \ n\sigma^2)$. В этом случае говорят, что она **асимптотически нормальна** с этими параметрами.

Замечание 3. Полагая в приближенной формуле

$$P\Big\{\alpha \le \frac{S_n - na}{\sigma\sqrt{n}} \le \beta\Big\} \approx \varPhi(\beta) - \varPhi(\alpha)$$

$\alpha = -t, \ \beta = t$, получим, в частности, что

$$P\Big\{\Big|\frac{S_n - na}{\sigma\sqrt{n}}\Big| \le t\Big\} \approx 2\varPhi(t). \tag{18.6}$$

18.3. Применение центральной предельной теоремы к приближенному вычислению интегралов.

Предположим, что необходимо приближенно вычислить интеграл

$$I = \int_a^b f(x)dx,$$

где $-\infty \le a < b \le +\infty$. Если хоть один из пределов интегрирования бесконечен, то будет предполагаться, что интеграл сходится абсолютно. Пусть $p(x)$ — некоторая плотность вероятности, т.е. $p(x) \ge 0$ и $\int\limits_{-\infty}^{+\infty} p(x)dx = 1$, причем $p(x) > 0$ для $x \in [a, b]$ и $p(x) = 0$ в противном случае. Тогда

$$I = \int_a^b \varphi(x)p(x)dx, \qquad \text{где} \quad \varphi(x) = \frac{f(x)}{p(x)}.$$

Доопределяя функцию $\varphi(x)$ вне отрезка $[a, b]$ произвольным образом, получим:

$$I = \int_{-\infty}^{+\infty} \varphi(x)p(x)dx. \tag{18.7}$$

Итак, задача свелась к вычислению интеграла (18.7). Пусть $\xi_1, \xi_2, \dots, \xi_n, \dots$ — независимые случайные величины с одной и той же плотностью вероятности $p(x)$. Тогда случайные величины $\eta_1 = \varphi(\xi_1), \eta_2 = \varphi(\xi_2), \dots, \eta_n = \varphi(\xi_n), \dots$ независимы (как борелевские функции от независимых случайных величин) и одинаково распределены.

По теореме о математическом ожидании функции от случайной величины

$$M\eta_k = \int\limits_{-\infty}^{+\infty} \varphi(x)p(x)dx = I.$$

По закону больших чисел в форме Хинчина $\frac{1}{n}\sum\limits_{k=1}^{n}\eta_k \xrightarrow{P} I$. Поэтому при больших n

$$I \approx \frac{1}{n}\sum_{k=1}^{n}\eta_k = \frac{1}{n}\sum_{k=1}^{n}\varphi(\xi_k). \qquad (18.8)$$

Получили приближенную формулу для вычисления интеграла. Для компьютеров разработаны стандартные программы, позволяющие моделировать независимые случайные величины с заданной плотностью вероятности. Как мы видим, данный алгоритм вычисления очень прост. Однако желательно, чтобы контролировалась заданная точность вычислений. Для того, чтобы это стало возможным, дополнительно предположим, что существует

$$M\eta_1^2 = \int\limits_a^b [\varphi(x)]^2 p(x)dx < \infty.$$

Тогда $\sigma^2 = D\eta = M\eta^2 - (M\eta)^2 < \infty$, т.е. последовательность случайных величин (η_n) удовлетворяет условиям центральной предельной теоремы. Используя (18.6) при $t = 3$ и таблицы функции Лапласа, получим:

$$P\left\{\left|\frac{\sum\limits_{k=1}^{n}\eta_k - nI}{\sigma\sqrt{n}}\right| \leq 3\right\} \approx 2\varPhi(3) \approx 0.997.$$

Значит, практически достоверно, что

$$\Big|\frac{1}{n}\sum_{k=1}^{n}\eta_k - I\Big| \le \frac{3\sigma}{\sqrt{n}}.$$

Таким образом, при применении приближенной формулы (18.7) ошибка вычислений имеет порядок $\frac{1}{\sqrt{n}}$.

Достоинством данного метода является очень простой алгоритм вычисления, недостаток — медленная скорость сходимости (порядка $O\Big(\frac{1}{\sqrt{n}}\Big)$), в связи с чем приходится моделировать очень большое число независимых случайных величин с данной плотностью вероятности. Однако при использовании компьютера случайные величины заменяются на так называемые псевдослучайные величины, для которых, начиная с некоторого номера начинает повторяться некоторый отрезок $\eta_N, \eta_{N+1}, \dots, \eta_{N+K-1}$ последовательности (η_n). Поэтому надо заботиться о том, чтобы длина периода K и длина отрезка апериодичности $N+K-1$ были достаточно большими (чтобы при вычислении не применялись повторяющиеся значения случайных величин η_k).

Следует отметить, что построенный нами алгоритм для вычисления однократного интеграла практически не используется, а используются классические способы вычисления. Но наши рассуждения допускают простую модификацию, если надо вычислить кратный интеграл

$$I = \int\limits_{D}\cdots\int f(x_1, x_2, \dots, x_n)dx_1dx_2\dots dx_n.$$

В этом случае необходимо моделировать независимые случайные вектора $\xi_1, \xi_2, \dots$ с плотностью распределения, сосредоточенной в области D и строго положительной в этой области. Если необходимо считать интегралы большой кратности, то, как показал опыт физиков-ядерщиков, классические подходы применять невозможно, а единственный способ вычисления — применить описанную выше процедуру. Методы такого рода носят название **методов статистических испытаний** или **методов Монте-Карло**.

18.4. Центральная предельная теорема в формах Линдеберга-Феллера и Ляпунова.

Оказывается, требование одинаковой распределенности случайных величин в центральной предельной теореме можно существенно ослабить.

Пусть $\xi_1, \xi_2, \dots$ — последовательность независимых случайных величин с конечными моментами второго порядка, $a_k = M\xi_k$, $\sigma_k^2 = D\xi_k$, $S_n = \sum\limits_{k=1}^{n} \xi_k$, $D_n^2 = \sum\limits_{k=1}^{n} \sigma_k^2$, $F_k(x) = F_{\xi_k}(x)$ — функция распределения случайной величины ξ_k $(k = 1, 2, \dots)$.

Определение 18.1. Говорят, что последовательность случайных величин (ξ_n) удовлетворяет **условию Линдеберга**, если для любого $\varepsilon > 0$

$$\frac{1}{D_n^2} \sum_{k=1}^{n} \int\limits_{\{x:\, |x-a_k| \geq \varepsilon D_n\}} (x - a_k)^2 dF_k(x) \to 0 \qquad \text{при } \; n \to \infty.$$

Определение 18.2. Говорят, что случайные величины $\frac{\xi_k - a_k}{D_n}$, $k = \overline{1, n}$, $n = 1, 2, \dots$, **асимптотически ма-**

лы, если для любого $\varepsilon > 0$

$$\max_{1\le k\le n} P\Big\{\Big|\frac{\xi_k - a_k}{D_n}\Big| \ge \varepsilon\Big\} \to 0 \qquad \text{при} \quad n \to \infty.$$

Лемма 18.3. *Если последовательность случайных величин* (ξ_n) *удовлетворяет условию Линдеберга, то случайные величины* $\frac{\xi_k - a_k}{D_n}$, $k = \overline{1,n}$, $n = 1, 2, \ldots$, *асимптотически малы.*

Доказательство. Утверждение следует из следующей цепочки неравенств:

$$\frac{1}{D_n^2}\sum_{k=1}^{n}\int\limits_{\{x:\,|x-a_k|\ge\varepsilon D_n\}}(x-a_k)^2 dF_k(x) \ge$$

$$\ge \varepsilon^2 \sum_{k=1}^{n} P\{|\xi_k - a_k| \ge \varepsilon D_n\} \ge$$

$$\ge \varepsilon^2 \max_{1\le k\le n} P\Big\{\Big|\frac{\xi_k - a_k}{D_n}\Big| \ge \varepsilon\Big\}.$$

□

Определение 18.4. Говорят, что последовательность случайных величин (ξ_n) удовлетворяет **условию Ляпунова**, если для некоторого $\delta > 0$

$$\frac{1}{D_n^{2+\delta}}\sum_{k=1}^{n} M|\xi_k - a_k|^{2+\delta} \to 0 \qquad \text{при} \quad n \to \infty.$$

Лемма 18.5. *Если последовательность случайных величин (ξ_n) удовлетворяет условию Ляпунова, то она удовлетворяет условию Линдеберга.*

Доказательство. Пусть выполнено условие Ляпунова. Тогда

$$M|\xi_k - a_k|^{2+\delta} = \int\limits_{-\infty}^{+\infty} |x - a_k|^{2+\delta} dF_k(x) \geq$$

$$\geq \int\limits_{\{x\colon |x-a_k|\geq \varepsilon D_n\}} |x - a_k|^{2+\delta} dF_k(x) \geq$$

$$\geq \varepsilon^\delta D_n^\delta \int\limits_{\{x\colon |x-a_k|\geq \varepsilon D_n\}} (x - a_k)^2 dF_k(x).$$

Отсюда

$$\frac{1}{D_n^2} \sum_{k=1}^{n} \int\limits_{\{x\colon |x-a_k|\geq \varepsilon D_n\}} (x - a_k)^2 dF_k(x) \leq$$

$$\leq \frac{1}{\varepsilon^\delta} \frac{1}{D_n^{2+\delta}} \sum_{k=1}^{n} M|\xi_k - a_k|^{2+\delta}.$$

Так как выполняется условие Ляпунова, то при любом $\varepsilon > 0$ правая часть последнего неравенства стремится к нулю. Значит и его левая часть стремится к нулю, т.е. выполняется условие Линдеберга.

□

Центральная предельная теорема Линдеберга. *Если последовательность независимых случайных вели-*

чин с конечным вторым моментом удовлетворяет условию Линдеберга, то при $n \to \infty$ равномерно по $x \in \mathbb{R}$

$$P\left\{\frac{1}{D_n}\sum_{k=1}^{n}(\xi_k - a_k) < x\right\} \to N(x).$$

Доказательство этой теоремы выносится в приложение.

В силу леммы 18.5 отсюда следует

Центральная предельная теорема Ляпунова. *Если последовательность независимых случайных величин удовлетворяет условию Ляпунова, то при $n \to \infty$ равномерно по $x \in \mathbb{R}$*

$$P\left\{\frac{1}{D_n}\sum_{k=1}^{n}(\xi_k - a_k) < x\right\} \to N(x).$$

Итак, если выполнено условие Линдеберга, то $\frac{\xi_k - a_k}{D_n}$ асимптотически малы и выполняется центральная предельная теорема. Феллеру удалось обратить это утверждение. Точнее, он доказал, что если для независимых случайных величин с асимптотически малыми $\frac{\xi_k - a_k}{D_n}$ выполняется центральная предельная теорема, то эта последовательность случайных величин удовлетворяет условию Линдеберга. Объединяя результаты Линдеберга и Феллера получаем следующий результат.

Центральная предельная теорема Линдеберга-Феллера. *Пусть последовательность независимых случайных величин удовлетворяет условию асимптотической малости $\frac{\xi_k - a_k}{D_n}$. Для того, чтобы для нее* *выполнялась центральная предельная теорема, необходимо и достаточно, чтобы эта последовательность удовлетворяла условию Линдеберга.*

Замечание 4. Резюмируя полученные результаты, можно сказать, что для независимых случайных величин центральная предельная теорема выполняется, если каждое слагаемое $\frac{\xi_k - a_k}{D_n}$ вносит равномерно малый вклад в нормированную сумму $\sum\limits_{k=1}^{n} \frac{\xi_k - a_k}{D_n}$. Очень часто случайная величина, описывающая некоторое явление, является суммой большого числа независимых случайных величин. Например, ошибка взвешивания состоит из большого числа независимых вкладов: 1) за счет изменения атмосферного давления; 2) за счет положения тела на чашке весов; 3) за счет особенностей в конструкции весов; 4) за счет глазомера взвешивающего лица и т.д. По центральной предельной теореме ошибка взвешивания приближенно имеет нормальное распределение. Подобная же причина объясняет, почему нормальное распределение чаще других распределений встречается в природе и технике. Однако этот факт нельзя абсолютизировать. Известно шутливое высказывание: математик верит в нормальный закон, так как думает, что это проверено экспериментатором, а экспериментатор верит в нормальный закон, так как думает, что этот факт доказан математиком.

Приложение 1. Доказательство теоремы о существовании случайной величины с заданной функцией распределения

Теорема о существовании случайной величины с заданной функцией распределения. *Пусть функция* $F: \mathbb{R} \to \mathbb{R}$

1) *не убывает*,

2) *непрерывна слева*,

3) $F(-\infty) = 0, \quad F(+\infty) = 1.$

Тогда существует вероятностное пространство $(\Omega,\ \mathcal{F},\ P)$ *и случайная величина* $\xi = \xi(\omega)$, *определенная на нем такая, что*

$$F_\xi(x) = F(x).$$

Д о к а з а т е л ь с т в о. Построим вероятностное пространство $(\Omega,\ \mathcal{F},\ P)$. Возьмем $\Omega = \mathbb{R}$, $F = \mathcal{B}$ — σ-алгебра борелевских множеств. Определим вероятностную меру P сначала на алгебре $\mathcal{A}$ конечных объединений непересекающихся полуинтервалов, а именно, для любого $A = \sum\limits_{i=1}^{n} [a_i, b_i) \in \mathcal{A}$ положим

$$P(A) = \sum_{i=1}^{n} [F(b_i) - F(a_i)].$$

Так как F не убывает, то $P(A) \geq 0 \quad \forall A \in \mathcal{A}$, т.е. выполнена аксиома I вероятности. Далее, в силу условия 3) теоремы $P(\Omega) = P(\mathbb{R}) = F(+\infty) - F(-\infty) = 1$, т.е. выполнена аксиома II вероятности. Если $A = \sum\limits_{i=1}^{n} [a_i, b_i)$ и

$B = \sum\limits_{j=1}^{n} [c_j, d_j)$ не пересекаются, то при любых i, j полуинтервалы $[a_i, b_i)$ и $[c_j, d_j)$ не пересекаются, т.е. $A + B = \sum\limits_{i=1}^{n} [a_i, b_i) + \sum\limits_{j=1}^{n} [c_j, d_j)$. Следовательно,

$$P(A+B) = \sum_{i=1}^{n} [F(b_i) - F(a_i)] + \sum_{j=1}^{n} [F(d_j) - F(c_j)] =$$

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

т.е. P конечно аддитивна. Для того, чтобы доказать, что она счетно аддитивна, достаточно доказать, что P непрерывна в пустом множестве (нуле), т.е. что при $A_n \downarrow \emptyset$ выполняется $P(A_n) \to 0 \quad (A_n \in \mathcal{A})$. Сначала предположим, что $A_n \downarrow \emptyset$ и все множества A_n принадлежат некоторому конечному замкнутому отрезку $[-N, N]$. Так как в силу непрерывности слева функции $F(x)$

$$P([a, b')) = F(b') - F(a) \to F(b) - F(a) = P([a, b))$$

при $\quad b' \to b - 0$, то существует $[a, b')$ такой, что $[a, b'] \subset [a, b)$, а разность $P([a, b)) - P([a, b'))$ сколь угодно мала. Учитывая, что A_n — конечные суммы попарно непересекающихся $[a_i, b_i)$, получаем, что для любого $\varepsilon > 0$ и любого A_n найдется множество $B_n \in \mathcal{A}$ такое, что его замыкание $[B_n] \subset A_n$ и

$$P(A_n) - P(B_n) \le \frac{\varepsilon}{2^n}.$$

По предположению $\bigcap\limits_n A_n = \emptyset$, следовательно, $\bigcap\limits_n [B]_n = \emptyset$. Но множества $[B_n]$ замкнуты, поэтому найдется такое конечное $n_0 = n_0(\varepsilon)$, что

$$\bigcap_{n=1}^{n_0} [B_n] = \emptyset. \tag{П1.1}$$

(В самом деле, $[-N, N]$ — компактное множество, а система множеств $\{ [-N, N] \backslash [B_n] \ n = 1, 2, \dots \}$ образует открытое покрытие этого компактного множества. По лемме Гейне-Бореля существует конечное подпокрытие:

$$\bigcup_{n=1}^{n_0} ([-N, N] \setminus [B_n]) = [-N, N],$$

а значит, $\bigcap\limits_{n=1}^{n_0} [B_n] = \emptyset$).

Учитывая (П1.1) и то, что $A_{n_0} \subset A_{n_0-1} \subset \dots \subset A_1$, получаем

$$P(A_{n_0}) = P\Big(A_{n_0} \setminus \bigcap_{k=1}^{n_0} B_k\Big) + P\Big(\bigcap_{k=1}^{n_0} B_k\Big) =$$

$$= P\Big(A_{n_0} \setminus \bigcap_{k=1}^{n_0} B_k\Big) \le P\Big(\bigcup_{k=1}^{n_0} (A_k \setminus B_k)\Big) \le$$

$$\le \sum_{k=1}^{n_0} P(A_k \setminus B_k) \le \sum_{k=1}^{n_0} \frac{\varepsilon}{2^k} < \varepsilon.$$

В силу произвольности $\varepsilon > 0$ это и означает, что $P(A_n) \to 0$ при $n \to \infty$.

Остается избавиться от предположения, что все $A_n \subset [-N, N]$ для некоторого N. Так как $P([-N, N]) \to 0$ при $N \to \infty$, то по заданному $\varepsilon > 0$ можно найти такое N, что $P([-N, N]) > 1 - \frac{\varepsilon}{2}$. Тогда, поскольку

$$A_n = A_n \cap [-N, N] + A_n \cap \overline{[-N, N]},$$

то

$$P(A_n) = P(A_n \cap [-N, N]) + P(A_n \cap \overline{[-N, N]}) \le$$

$$\le P(A_n \cap [-N, N]) + \frac{\varepsilon}{2},$$

и, применяя предшествующее рассуждение (с заменой A_n на $A_n \cap [-N, N]$), получаем, что для достаточно больших n будет выполняться $P(A_n \cap [-N, N]) \le \frac{\varepsilon}{2}$. Итак, снова доказано, что $P(A_n) \to 0$ при $n \to \infty$.

Таким образом, P — вероятностная мера на $\mathcal{A}$. По теореме о продолжении меры мы можем единственным способом продолжить ее на σ-алгебру борелевских множеств $\mathcal{B}$.

Итак, вероятностное пространство $(\mathbb{R}, \mathcal{B}, P)$ построено. Остается положить $\xi(x) = x \quad \forall\, x \in \mathbb{R}$. Тогда

$$F_\xi(x) = P\{\xi < x\} = P((-\infty, x)) = F(x).$$

$\square$

Приложение 2. Доказательство леммы о слабой сходимости к непрерывной функции

Лемма о слабой сходимости функций распределения к непрерывной функции. *Если последовательность функций распределения слабо сходится к непрерывной функции, то она равномерно сходится на $\mathbb{R}$ к этой функции.*

Д о к а з а т е л ь с т в о. Так как $F(x)$ не убывает и $0 \le F(x) \le 1$, то $\forall \varepsilon > 0$ существует конечное число точек $x_1 < x_2 < \ldots < x_K$ такое, что на каждом из промежутков

$$(-\infty \equiv x_0,\ x_1),\ \ [x_1,\ x_2),\ \ldots,\ [x_K,\ x_{K+1} \equiv +\infty)$$

приращение функции $F(x)$ не превосходит ε. Пусть $x \in [x_k,\ x_{k+1})$. Так как функции $F_n(x)$ и $F(x)$ не убывают, то

$$|F_n(x)-F_n(x_k)| \le F_n(x_{k+1})-F_n(x_k) = F_n(x_{k+1})-F(x_{k+1})+$$

$$+F(x_{k+1})-F(x_k)+F(x_k)-F_n(x_k) \le |F_n(x_{k+1})-F(x_{k+1})|+$$

$$+F(x_{k+1}) - F(x_k) + |F(x_k) - F_n(x_k)|,$$

$$|F(x_k) - F(x)| \le F(x_{k+1}) - F(x_k).$$

Поэтому

$$|F_n(x) - F(x)| \le |F_n(x) - F_n(x_k)| + |F_n(x_k) - F(x_k)|+$$

$$+|F(x_k) - F(x)| \le |F_n(x_{k+1}) - F(x_{k+1})|+$$

$$+2|F_n(x_k) - F(x_k)| + 2(F(x_{k+1} - F(x_k)). \qquad (\Pi 1.1)$$

Так как $F_n(x_k) \to F(x_k)$, то $\forall \varepsilon > 0\ \ \exists N(k)$ такое, что $\forall n > N(k)$ выполняется $|F_n(x_k)-F(x_k)| < \varepsilon,\ \ |F_n(x_{k+1})-F(x_{k+1})| < \varepsilon$. Так как $x \in [x_k,\ x_k+1)$, то $F(x_{k+1})-F(x_k) < \varepsilon$.

Пусть $N = \max\limits_{0 \le k \le K} N(k)$. Тогда $\forall n > N$ и $\forall x \in \mathbb{R}$ $|F_n(x_k) - F(x_k)| < \varepsilon$. Но тогда из (П1.1) следует, что $\forall \varepsilon > 0$ $\exists N$ такое, что $\forall n > N$ и $\forall x \in \mathbb{R}$ выполняется $|F_n(x) - F(x)| < 5\varepsilon$. Это и означает равномерную сходимость $F_n(x)$ к $F(x)$. $\square$

Приложение 3. Доказательство теоремы непрерывности для производящих функций.

Теорема непрерывности для производящих функций. *Пусть* $(P_n(z) = \sum\limits_{k=0}^{\infty} p_k^{(n)} z^k,\ n = 1, 2, \dots)$ — *последовательность производящих функций вероятностей,* $P(z) = \sum\limits_{k=0}^{\infty} p_k z^k$ — *производящая функция последовательности. Для того, чтобы все* $\lim\limits_{n\to\infty} p_k^{(n)} = p_k, \quad k = 0, 1, 2, \dots,$ *необходимо и достаточно, чтобы при всех* $z, \quad |z| < 1$

$$\lim_{n\to\infty} P_n(z) = P(z).$$

Д о к а з а т е л ь с т в о. *Необходимость*. Предположим, что $\lim\limits_{n\to\infty} p_k^{(n)} = p_k, \quad k = 0, 1, 2, \dots$. Фиксируем произвольное $\varepsilon > 0$, и пусть $|z| < 1$. Тогда

$$|P_n(z) - P(z)| = \Big| \sum_{k=0}^{\infty} (p_k^{(n)} - p_k) z^k \Big| \le \sum_{k=0}^{N-1} |p_k^{(n)} - p_k| +$$

$$+2 \sum_{k=N}^{\infty} |z|^k = \sum_{k=0}^{N-1} |p_k^{(n)} - p_k| + \frac{2|z|^N}{1 - |z|}.$$

Так как $\frac{2|z|^N}{1-|z|} \to 0$ при $|z| < 1$, то по рассматриваемому $\varepsilon > 0$ найдется N такое, что $\frac{2|z|^N}{1-|z|} < \frac{\varepsilon}{2}$. Так как при найденном фиксированном N сумма $\sum\limits_{k=0}^{N-1} |p_k^{(n)} - p_k| \to 0$ при $n \to \infty$, то по ε и N найдется n_0

такое, что $\sum\limits_{k=0}^{N-1}|p_k^{(n)}-p_k|<\frac{\varepsilon}{2}$ при $n\geq n_0$. Следовательно, $\forall\,\varepsilon>0\;\;\exists\,n_0=n_0(\varepsilon)$ такое, что $|P_n(z)-P(z)|<\frac{\varepsilon}{2}+\frac{\varepsilon}{2}=\varepsilon$ при $n>n_0$. Это и означает, что $\lim\limits_{n\to\infty}P_n(z)=P(z)$ при $|z|<1$.

Достаточность. Пусть $\lim\limits_{n\to\infty}P_n(z)=P(z)$ при $|z|<1$. Из ограниченной последовательности $0\leq p_0^n\leq 1$ выбираем сходящуюся подпоследовательность $p_0^{(n,1)}\to p_0$. Из ограниченной последовательности $0\leq p_1^{(n,1)}\leq 1$ выбираем сходящуюся подпоследовательность $p_1^{(n,2)}\to p_1$ и т.д. Из последовательностей

$$\begin{array}{cccc}
p_k^{(1,1)}, & p_k^{(2,1)}, & p_k^{(3,1)} & \dots\\
p_k^{(1,2)}, & p_k^{(2,2)}, & p_k^{(3,2)} & \dots\\
p_k^{(1,3)}, & p_k^{(2,3)}, & p_k^{(3,3)} & \dots\\
\dots & \dots & \dots & \dots
\end{array}$$

выбираем диагональную подпоследовательность $p_k^{(n,n)}$, которая сходится к p_k при любом k. Предположим, что хотя бы при одном k последовательность $p_k^{(n)}$ не сходится к p_k. Тогда можно выбрать две сходящиеся к разным пределам подпоследовательности $p_k^{(n')}\to p_k^*$, $\;p_k^{(n'')}\to p_k^{**}$. По первой части теоремы $P_{n'}(z)\;\to\;P^*(z)\;=\;\sum\limits_{k=0}^{\infty}p_k^*z^k$ и $P_{n''}(z)\;\to\;P^{**}(z)\;=\;\sum\limits_{k=0}^{\infty}p_k^{**}z^k$. Так как по условию $P_n(z)\;\to\;P(z)$, то $P^*(z)\;=\;P^{**}(z)\;=\;P(z)$ и $p_k^*\;=\;p_k^{**}=p_k$, т.е. $\lim\limits_{n\to\infty}p_k^n=p_k$. Противоречие. Следовательно, при всех k последовательность $p_k^{(n)}$ сходится к p_k. $\square$

Приложение 4. Доказательство теорем непрерывности для характеристических функций.

Лемма П4.1. *Если последовательность функций распределения $(F_n(x))$ слабо сходится к неубывающей функции $F(x)$, то существует неубывающая непрерывная слева функция $0 \leq F^*(x) \leq 1$ такая, что $F_n(x) \Rightarrow F^*(x)$. При этом множества точек непрерывности F и F^* совпадают.*

Д о к а з а т е л ь с т в о. Для любой функции $g(x)$ обозначим через $C(g)$ множество ее точек непрерывности. Положим для $x \in \mathbb{R}$

$$F^*(x) = \sup_{y<x,\ y\in C(F)} F(y), \tag{П4.1}$$

Покажем, что F^* — искомая функция.

1. F^* не убывает, так как если $x_1 < x_2$, то

$$F^*(x_1) = \sup_{y<x_1,\ y\in C(F)} F(y) \leq \sup_{y<x_2,\ y\in C(F)} F(y) = F^*(x_2).$$

2. Покажем, что для любого $x_0 \in \mathbb{R}$ $F^*(x_0) = \lim\limits_{x\to x_0-0} F(x)$. В силу (П4.1) $\forall \varepsilon > 0 \ \ \exists y \in C(F),\ y < x$ такое, что $F^*(x_0) - F(y) < \varepsilon$. Обозначим $\delta = x_0 - y > 0$. Тогда $\forall \varepsilon > 0 \ \ \exists \delta = x_0 - y > 0$ такое, что из $x - \delta = y < x < x_0$ следует $F^*(x_0) - F(x) \leq F^*(x_0) - F(y) < \varepsilon$.

3. Докажем, что F^* непрерывна слева. Действительно,

$$\lim_{x\to x_0-0} F^*(x) = \lim_{x\to x_0-0} \lim_{y\to x-0} F(y) = \lim_{y\to x_0-0} F(y) = F^*(x_0).$$

4. Если $x_0 \in C(F)$, то в силу доказанного в 2

$$F^*(x_0) = \lim_{x\to x_0-0} F(x) = F(x_0).$$

5. Установим, что в каждой точке пределы справа у функций F и F^* совпадают. Так как множество точек непрерывности F всюду плотно в $\mathbb{R}$, то, применяя 4, получим

$$\lim_{x\to x_0+0} F^*(x) = \lim_{x\to x_0+0,\, x\in C(F)} F^*(x) =$$

$$= \lim_{x\to x_0+0,\, x\in C(F)} F(x) = \lim_{x\to x_0+0} F(x).$$

6. Если $x_0 \in C(F)$, то в силу 4,5 $x_0 \in C(F^*)$.

7. Пусть $x_0 \notin C(F)$, т.е. $\lim\limits_{x\to x_0+0} F(x) - \lim\limits_{x\to x_0-0} F(x) > 0$. В силу (П4.1) и 4

$$\lim_{x\to x_0+0} F^*(x) - \lim_{x\to x_0-0} F^*(x) \geq \lim_{x\to x_0+0} F^*(x) -$$

$$- \lim_{x\to x_0-0} F(x) = \lim_{x\to x_0+0} F(x) - \lim_{x\to x_0-0} F(x) > 0,$$

т.е. $x_0 \notin C(F^*)$. Из 6 и 7 следует, что множества точек непрерывности F и F^* совпадают. Так как в этих точках $F = F^*$, то $F_n \Rightarrow F^*$. Поскольку $0 \leq F_n(x) \leq 1$, то в точках непрерывности $0 \leq F(x) \leq 1$. Тогда из (П4.1) следует, что $0 \leq F^* \leq 1$. $\square$

Замечание 1. К сожалению, F^* может не оказаться функцией распределения, ибо, как мы убедились на примере $\xi_n(\omega) \equiv n$, может нарушаться свойство $F(+\infty) - F(-\infty) = 1$. Можно быть уверенным только в том, что $0 \leq F(+\infty) - F(-\infty) \leq 1$.

Лемма П4.2. *Если $F_n(x) \to F(x)$ на всюду плотном на прямой множестве D, то $F_n(x) \Rightarrow F(x)$.*

Д о к а з а т е л ь с т в о. Пусть $x',\ x'' \in D,\ \ x' < x < x''$. Очевидно, что

$$F(x') = \lim_{n\to\infty} F_n(x') \le \liminf_{n\to\infty} F_n(x) \le \limsup_{n\to\infty} F_n(x) \le$$

$$\le \lim_{n\to\infty} F_n(x'') = F(x'').$$

Поэтому

$$F(x-0) \le \liminf_{n\to\infty} F_n(x) \le \limsup_{n\to\infty} F_n(x) \le F(x+0).$$

Если x — точка непрерывности F, то крайние члены последнего неравенства совпадают с $F(x)$. Следовательно, в точках непрерывности F существует $\lim\limits_{n\to\infty} F_n(x) = F(x)$. Это и означает слабую сходимость F_n к F. $\square$

Первая теорема Хелли. *Из всякой последовательности функций распределения $(F_n(x))$ можно выбрать слабо сходящуюся подпоследовательность. Ее слабый предел $F(x)$ может быть выбран непрерывной слева неубывающей функцией, для которой* $0 \le F(x) \le 1$.

Д о к а з а т е л ь с т в о. Пусть $D = \{x_k\}$ — счетное всюду плотное множество на прямой. Из ограниченной числовой последовательности $(F_n(x_1))$ (все ее члены лежат между 0 и 1) выбираем сходящуюся подпоследовательность $(F_{1n}(x_1))$. Пусть $c_1 = \lim\limits_{n\to\infty} F_{n_1}(x_1)$. Из ограниченной числовой последовательности $F_{1n}(x_2)$ выбираем сходящуюся подпоследовательность $F_{2n}(x_2)$, предел которой обозначим c_2 и т.д. Тогда для диагональной подпоследовательности $F_{nn}(x)$ в любой точке $x_k \in D$ получим, что $F_{nn}(x_k) \to c_k$. Так как $F_{nn}(x_k) \le F_{nn}(x_l)$ при $x_k < x_l$, то в пределе при $n \to \infty$ получим, что $c_k \le c_l$. Положим

$$F(x) = \sup_{x_k \le x,\ x_k \in D} c_k.$$

Очевидно, $F(x)$ — неубывающая функция. Поскольку $F(x_k) = c_k$ для любого $x_k \in D$, то $F_{nn}(x_k) \to F(x_k)$ в точках множества D. По лемме П4.2 отсюда следует, что $F_{nn}(x) \Rightarrow F(x)$. В силу леммы П4.1 слабый предел может быть выбран непрерывной слева неубывающей функцией, для которой $0 \le F(x) \le 1$ $\square$.

Замечание 2. На самом деле можно показать, что построенная в доказательстве функция F сама удовлетворяет нужным условиям (непрерывна слева, не убывает, $0 \le F \le 1$). Доказательство этого факта будет почти дословным повторением доказательства леммы П4.1.

Замечание 3. Мы определили слабую сходимость как сходимость к неубывающей функции. Если определять слабую сходимость альтернативно, не требуя, чтобы предельная функция была неубывающей, то формулировка первой теоремы Хелли должна быть такой: *из всякой последовательности функций распределения* $(F_n(x))$ *можно выбрать подпоследовательность, слабо сходящуюся к неубывающей функции.* Действительно, при таком определении слабой сходимости любая последовательность сама уже слабо сходится, например, к функции Дирихле. Тогда и любая ее подпоследовательность будет слабо сходится к функции Дирихле. К тому же при доказательстве обратной теоремы непрерывности для характеристических функций, которое приводится ниже, требуется, чтобы предельная функция у подпоследовательности была неубывающей и непрерывной слева.

Вторая теорема Хелли. *Если $g(x)$ — непрерывная ограниченная функция, определенная на числовой прямой, $F_n(x) \Rightarrow F(x)$ и $F(+\infty) - F(-\infty) = 1$, то*

$$\lim_{n\to\infty} \int\limits_{-\infty}^{\infty} g(x)dF_n(x) = \int\limits_{-\infty}^{\infty} g(x)dF(x). \tag{П4.2}$$

Д о к а з а т е л ь с т в о. Пусть a и b $(a < b)$ — точки непрерывности функции $F(x)$. Сначала докажем, что

$$\lim_{n\to\infty} \int\limits_{a}^{b} g(x)dF_n(x) = \int\limits_{a}^{b} g(x)dF(x). \tag{П4.3}$$

Фиксируем $\varepsilon > 0$. Так как $g(x)$ равномерно непрерывна на отрезке $[a, b]$, а множество точек непрерывности $F(x)$ всюду плотно на числовой прямой R, то отрезок $[a, b]$ можно разбить точками непрерывности $a = x_0,\ x_1,\ \dots,\ x_{n-1},\ x_n = b$ функции $F(x)$ на такие отрезки $[x_{k-1}, x_k]$, что $|g(x) - g(x_k)| < \varepsilon$ при $x \in [x_{k-1}, x_k]$. Определим ступенчатую функцию

$$g_\varepsilon(x) = \sum_{k=1}^{n} g(x_k) I_{x\in[x_{k-1}, x_k)},$$

для которой $|g_\varepsilon(x) - g(x)| \le \varepsilon$ при $x \in [a, b]$, и пусть $M = \sup\limits_{x\in R} |g(x)|$. Тогда при достаточно больших n

$$\Big|\int\limits_{a}^{b} g(x)dF_n(x) - \int\limits_{a}^{b} g(x)dF(x)\Big| \le \int\limits_{a}^{b} |g(x) - g_\varepsilon(x)|dF_n(x) +$$

$$+\Big|\int\limits_a^b g_\varepsilon(x)dF_n(x)-\int\limits_a^b g_\varepsilon(x)dF(x)\Big|+\int\limits_a^b |g(x)-g_\varepsilon(x)|dF(x)\le$$

$$\le 2\varepsilon+\Big|\sum_{k=1}^n [F_n(x_k)-F_n(x_{k-1})]g_\varepsilon(x_k)-$$

$$-\sum_{k=1}^n [F(x_k)-F(x_{k-1})]g_\varepsilon(x_k)\Big| = 2\varepsilon+\Big|\sum_{k=1}^n g_\varepsilon(x_k)[F_n(x_k)-$$

$$-F_n(x_{k-1})-F(x_k)+F(x_{k-1})]\Big|\le 2\varepsilon+M\sum_{k=1}^n [|F_n(x_k)-$$

$$-F(x_k)|+|F_n(x_{k-1})-F(x_{k-1})|]\le 2\varepsilon(1+Mn).$$

Этим доказано (П4.3). Выберем $X>0$ так, чтобы $F(-X)<\frac{\varepsilon}{4}$, $1-F(X)<\frac{\varepsilon}{4}$, причем $-X$ и X были точками непрерывности $F(x)$. Поскольку $F_n(-X)\to F(-X)$, $F_n(X)\to F(X)$, то $\forall\varepsilon>0$ $\exists N$ такое, что $\forall n>N$ выполняется $F_n(-X)<\frac{\varepsilon}{2}$, $1-F_n(X)<\frac{\varepsilon}{2}$. Имеем следующую оценку:

$$\Big|\int\limits_{-\infty}^{\infty} g(x)dF_n(x)-\int\limits_{-\infty}^{\infty} g(x)dF(x)\Big|\le\Big|\int\limits_{-X}^{X} g(x)dF_n(x)-$$

$$-\int\limits_{-X}^{X} g(x)dF(x)\Big|+\Big|\int\limits_{|x|>X} g(x)dF_n(x)\Big|+\Big|\int\limits_{|x|>X} g(x)dF(x)\Big|\le$$

$$\le\Big|\int\limits_{-X}^{X} g(x)dF_n(x)-\int\limits_{-X}^{X} g(x)dF(x)\Big|+M\Big(\int\limits_{|x|>X} dF_n(x)+$$

$$+\int\limits_{|x|>X} dF(x)\Big) \le \Big|\int\limits_{-X}^{X} g(x)dF_n(x) - \int\limits_{-X}^{X} g(x)dF(x)\Big| +$$

$$+M[1-F_n(X)+F_n(-X)+F(\infty)-F(X)+F(-X)-F(-\infty)] =$$

$$= \Big|\int\limits_{-X}^{X} g(x)dF_n(x) - \int\limits_{-X}^{X} g(x)dF(x)\Big| + M[1-F_n(X)+$$

$$+F_n(-X)+1-F(X)+F(-X)] <$$

$$< \Big|\int\limits_{-X}^{X} g(x)dF_n(x) - \int\limits_{-X}^{X} g(x)dF(x)\Big| + \frac{3M\varepsilon}{2}. \qquad (\text{П}4.4)$$

На основании (П4.3) заключаем, что правая часть (П4.4) может быть сделана сколь угодно малой. Это завершает доказательство. □

Прямая теорема непрерывности для характеристических функций. *Пусть* $F_n(x)$, $F(x)$ — *функции распределения,* $f_n(t)$, $f(t)$ — *соответствующие характеристические функции. Если* $F_n(x) \Rightarrow F(x)$, *то* $f_n(t) \to f(t)$ *для каждого* $t \in \mathbb{R}$.

Д о к а з а т е л ь с т в о. Поскольку при каждом фиксированном t функция $g(x) = e^{itx}$ непрерывна и ограничена, то по второй теореме Хелли

$$f_n(t) = \int\limits_{-\infty}^{\infty} e^{itx} dF_n(x) \to \int\limits_{-\infty}^{\infty} e^{itx} dF(x) = f(t).$$

□

Лемма П4.3. *Если* $f(t)$ *— характеристическая функция случайной величины* ξ, $X > 0$, $\tau > 0$, $X\tau > 1$, *то*

$$P\{|\xi| \le X\} \ge \frac{\left|\frac{1}{2\tau}\int\limits_{-\tau}^{\tau} f(t)dt\right| - \frac{1}{\tau X}}{1 - \frac{1}{\tau X}},$$

в частности, при $X\tau = 2$

$$P\{|\xi| \le X\} \ge 2\Big|\frac{1}{2\tau}\int\limits_{-\tau}^{\tau} f(t)dt\Big| - 1. \qquad (\text{П}4.5)$$

Д о к а з а т е л ь с т в о следует из цепочки неравенств

$$\Big|\frac{1}{2\tau}\int\limits_{-\tau}^{\tau} f(t)dt\Big| = \Big|\frac{1}{2\tau}\int\limits_{-\tau}^{\tau}\int\limits_{-\infty}^{\infty} e^{itx}dF(x)dt\Big| =$$

$$= \Big|\frac{1}{2\tau}\int\limits_{-\infty}^{\infty}\Big(\int\limits_{-\tau}^{\tau} e^{itx}dt\Big)dF(x)\Big| = \Big|\frac{1}{2\tau}\int\limits_{-\infty}^{\infty}\Big(\int\limits_{-\tau}^{\tau} \cos txdt\Big)dF(x)\Big| =$$

$$= \Big|\int\limits_{-\infty}^{\infty} \frac{\sin x\tau}{x\tau}dF(x)\Big| \le \int\limits_{-\infty}^{\infty} \Big|\frac{\sin x\tau}{x\tau}\Big|dF(x) =$$

$$= \int\limits_{|x|\le X} \Big|\frac{\sin x\tau}{x\tau}\Big|dF(x) + \int\limits_{|x|>X} \Big|\frac{\sin x\tau}{x\tau}\Big|dF(x) \le$$

$$\le \int\limits_{|x|\le X} dF(x) + \frac{1}{X\tau}\int\limits_{|x|>X} dF(x) = P\{|\xi| \le X\} +$$

$$+\frac{1}{X\tau}P\{|\xi| > X\} = P\{|\xi| \le X\} + \frac{1}{X\tau}[1 - P\{|\xi| \le X\}] =$$

$$= \Big(1 - \frac{1}{X\tau}\Big) P\{|\xi| \le X\} + \frac{1}{X\tau}. \qquad \square$$

Обратная теорема непрерывности для характеристических функций. *Пусть $(f_n(t))$ — последовательность характеристических функций, $(F_n(x))$ — последовательность соответствующих функций распределения. Если при каждом $t \in \mathbb{R}$ последовательность $(f_n(t))$ сходится к некоторой функции $f(t)$, непрерывной в нуле, то $f(t)$ — характеристическая функция. Если $F(x)$ — соответствующая ей функция распределения, то $F_n(x) \Rightarrow F(x)$.*

Доказательство. По первой теореме Хелли из последовательности (F_n) можно выделить слабо сходящуюся подпоследовательность (F_{nn}), $\quad F_{nn} \Rightarrow F$. При этом F может быть выбрана непрерывной слева и неубывающей, причем $0 \le F \le 1$.

Докажем, что $F(x)$ — функция распределения. Для этого остается доказать, что $F(-\infty) = 0, \quad F(\infty) = 1$. Заметим, что так как $f_n(0) = 1, \quad |f_n(t)| \le 1$ и $f_n \to f$, то $f(0) = 1, \quad |f(t)| \le 1$. По условию теоремы f непрерывна в нуле, т.е. $\forall \varepsilon > 0 \;\; \exists \tau_0 > 0$ такое, что $|f(t) - 1| \le \frac{\varepsilon}{4}$ при $|t| \le \tau_0$. Но тогда

$$\Big|\frac{1}{2\tau}\int\limits_{-\tau}^{\tau} f(t)dt\Big| \ge 1 - \frac{\varepsilon}{4} \quad \text{при} \quad 0 < \tau \le \tau_0.$$

Далее, $f_n(t) \to f(t) \quad \forall t \in \mathbb{R}$ и $|f_n(t)| \le 1$; по теореме Лебега о мажорируемой сходимости (предельный переход под знаком интеграла Лебега)

$$\frac{1}{2\tau}\int\limits_{-\tau}^{\tau} f_n(t)dt \to \frac{1}{2\tau}\int\limits_{-\tau}^{\tau} f(t)dt \quad \text{при} \quad n \to \infty,$$

и, следовательно,

$$\Big|\frac{1}{2\tau}\int\limits_{-\tau}^{\tau} f_n(t)dt - \frac{1}{2\tau}\int\limits_{-\tau}^{\tau} f(t)dt\Big| \le \frac{\varepsilon}{4}.$$

Значит,

$$\Big|\frac{1}{2\tau}\int\limits_{-\tau}^{\tau} f_n(t)dt\Big| = \Big|\frac{1}{2\tau}\int\limits_{-\tau}^{\tau} f(t)dt - \Big[\frac{1}{2\tau}\int\limits_{-\tau}^{\tau} f(t)dt-$$

$$-\frac{1}{2\tau}\int\limits_{-\tau}^{\tau} f_n(t)dt\Big]\Big| \ge \Big|\frac{1}{2\tau}\int\limits_{-\tau}^{\tau} f(t)dt\Big|-$$

$$-\Big|\frac{1}{2\tau}\int\limits_{-\tau}^{\tau} f(t)dt - \frac{1}{2\tau}\int\limits_{-\tau}^{\tau} f_n(t)dt\Big| \ge 1 - \frac{\varepsilon}{4} - \frac{\varepsilon}{4} = 1 - \frac{\varepsilon}{2}.$$

Если $\frac{2}{\tau}$ — точка непрерывности $F_n(x)$, то в силу (П4.5) из леммы П4.3

$$P\Big\{|\xi_n| \le \frac{2}{\tau}\Big\} = F_n\Big(\frac{2}{\tau}\Big) - F_n\Big(-\frac{2}{\tau}\Big) \ge 2\Big|\frac{1}{2\tau}\int\limits_{-\tau}^{\tau} f_n(t)dt\Big| - 1 \ge$$

$$\ge 2\Big(1 - \frac{\varepsilon}{2}\Big) - 1 = 1 - \varepsilon.$$

Последнее неравенство будет выполнятся и для подпоследовательности (F_{nn}), переходя в котором к пределу при $n \to \infty$, получим

$$F\Big(\frac{2}{\tau}\Big) - F\Big(-\frac{2}{\tau}\Big) \geq 1 - \varepsilon,$$

откуда $F(+\infty) - F(-\infty) \geq 1 - \varepsilon \quad \forall \varepsilon > 0$. Таким образом, $F(+\infty) - F(-\infty) \geq 1$. Так как $0 \leq F_{nn}(x) \leq 1$, то и $0 \leq F(x) \leq 1$, что вместе дает $F(-\infty) = 0, \;\; F(+\infty) = 1$. Итак, $F(x)$ — функция распределения.

Докажем, что $F_n(x) \Rightarrow F(x)$. Пусть $F_n \not\Rightarrow F$, тогда существуют подпоследовательности $F_{n'} \Rightarrow F^*, \quad F_{n''} \Rightarrow F^{**}, \quad F^* \neq F^{**}$. По прямой теореме непрерывности для характеристических функций $f_{n'} \to f^*, \quad f_{n''} \to f^{**}$ и $f^* \neq f^{**}$, а это противоречит тому, что $f_n \to f$. $\square$

Приложение 5. Доказательство центральной предельной теоремы Линдеберга.

Центральная предельная теорема Линдеберга. *Если последовательность независимых случайных величин с конечным вторым моментом удовлетворяет условию Линдеберга, то при $n \to \infty$ равномерно по $x \in \mathbb{R}$*

$$P\Big\{\frac{1}{D_n}\sum_{k=1}^{n}(\xi_k - a_k) < x\Big\} \to N(x).$$

Д о к а з а т е л ь с т в о. Без ограничения общности можно считать $a_k = 0, \quad k = 1, 2, \ldots$ (иначе надо перейти к случайным величинам $\xi_k' = \xi_k - a_k$). Введем обозначения:

$$f_k(t) = Me^{it\xi_k}, \quad T_n = \frac{S_n}{\sqrt{DS_n}} = \frac{S_n}{D_n},$$

$$f_{S_n}(t) = Me^{itS_n}, \quad f_{T_n}(t) = Me^{itT_n}.$$

Тогда в силу свойств характеристической функции

$$f_{T_n}(t) = Me^{itT_n} = Me^{i\frac{t}{D_n}S_n} = f_{S_n}\Big(\frac{t}{D_n}\Big) = \prod_{k=1}^{n} f_k\Big(\frac{t}{D_n}\Big), \tag{П5.1}$$

и для доказательства теоремы достаточно показать, что при $n \to \infty$

$$f_{T_n}(t) \to e^{-\frac{t^2}{2}}.$$

Действительно, тогда по обратной теореме непрерывности для характеристических функций $F_{T_n}(x) \Rightarrow N(x)$, а так как функция стандартного нормального распределения $N(x)$ непрерывна, то по лемме о слабой сходимости к непрерывной функции $F_{T_n}(x)$ будет сходится равномерно к $N(x)$.

Произвольным образом зафиксируем $t \in R$ до конца доказательства. В силу разложений

$$e^{iy} = 1 + iy + \frac{\theta_1 y^2}{2},$$

$$e^{iy} = 1 + iy - \frac{y^2}{2} + \frac{\theta_2 |y|^3}{3!},$$

справедливых для каждого действительного y с $\theta_1 = \theta_1(y)$, $\theta_2 = \theta_2(y)$ такими, что $|\theta_1| \leq 1$, $|\theta_2| \leq 1$, находим, что

$$f_k(t) = Me^{it\xi_k} = \int\limits_{-\infty}^{+\infty} e^{itx} dF_k(x) = \int\limits_{|x| \geq \varepsilon D_n} \Big(1 + itx +$$

$$+ \frac{\theta_1 (tx)^2}{2}\Big) dF_k(x) + \int\limits_{|x| < \varepsilon D_n} \Big(1 + itx - \frac{t^2 x^2}{2} + \frac{\theta_2 |tx|^3}{6}\Big) dF_k(x) =$$

$$= 1 + \frac{t^2}{2} \int\limits_{|x| \geq \varepsilon D_n} \theta_1 x^2 dF_k(x) - \frac{t^2}{2} \int\limits_{|x| < \varepsilon D_n} x^2 dF_k(x) +$$

$$+ \frac{|t|^3}{6} \int\limits_{|x| < \varepsilon D_n} \theta_2 |x|^3 dF_k(x)$$

(здесь учитывалось, что, согласно предположению, $a_k = \int\limits_{-\infty}^{+\infty} x dF_k(x) = 0$).

Следовательно,

$$f_k\Big(\frac{t}{D_n}\Big) = 1 - \frac{t^2}{2D_n^2} \int\limits_{|x| < \varepsilon D_n} x^2 dF_k(x) +$$

$$+\frac{t^2}{2D_n^2}\int\limits_{|x|\geq\varepsilon D_n}\theta_1 x^2 dF_k(x)+\frac{|t|^3}{6D_n^3}\int\limits_{|x|<\varepsilon D_n}\theta_2|x|^3 dF_k(x). \tag{П5.2}$$

Так как

$$\Big|\frac{1}{2}\int\limits_{|x|\geq\varepsilon D_n}\theta_1 x^2 dF_k(x)\Big|\leq\frac{1}{2}\int\limits_{|x|\geq\varepsilon D_n}x^2 dF_k(x),$$

то

$$\frac{1}{2}\int\limits_{|x|\geq\varepsilon D_n}\theta_1 x^2 dF_k(x)=\tilde{\theta}_1\int\limits_{|x|\geq\varepsilon D_n}x^2 dF_k(x), \tag{П5.3}$$

где $\tilde{\theta}_1=\tilde{\theta}_1(t,k,n)$ и $|\tilde{\theta}_1|\leq\frac{1}{2}$.

Точно так же устанавливаем, что

$$\Big|\frac{1}{6}\int\limits_{|x|<\varepsilon D_n}\theta_2|x|^3 dF_k(x)\Big|\leq\frac{1}{6}\int\limits_{|x|<\varepsilon D_n}\frac{\varepsilon D_n}{|x|}\cdot|x|^3 dF_k(x)=$$

$$=\frac{1}{6}\int\limits_{|x|<\varepsilon D_n}\varepsilon D_n x^2 dF_k(x),$$

и, значит,

$$\frac{1}{6}\int\limits_{|x|<\varepsilon D_n}\theta_2|x|^3 dF_k(x)=\tilde{\theta}_2\int\limits_{|x|<\varepsilon D_n}\varepsilon D_n x^2 dF_k(x), \tag{П5.4}$$

где $\tilde{\theta}_2=\tilde{\theta}_2(t,k,n)$ и $|\tilde{\theta}_2|\leq\frac{1}{6}$.

Положим теперь

$$A_{kn}=\frac{1}{D_n^2}\int\limits_{|x|<\varepsilon D_n}x^2 dF_k(x),$$

$$B_{kn} = \frac{1}{D_n^2} \int\limits_{|x| \geq \varepsilon D_n} x^2 dF_k(x).$$

Тогда в силу (П5.2) – (П5.4)

$$f_k\left(\frac{t}{D_n}\right) = 1 - \frac{t^2 A_{kn}}{2} + t^2 \tilde{\theta}_1 B_{kn} + |t|^3 \varepsilon \tilde{\theta}_2 A_{kn} = 1 + C_{kn}. \tag{П5.5}$$

Заметим, что

$$\sum_{k=1}^{n} (A_{kn} + B_{kn}) = 1 \tag{П5.6}$$

и, согласно условию Линдеберга,

$$\sum_{k=1}^{n} B_{kn} \to 0 \quad \text{при} \quad n \to \infty. \tag{П5.7}$$

Поэтому для достаточно больших n

$$\max_{1 \leq k \leq n} |C_{kn}| \leq t^2 \varepsilon^2 + \varepsilon |t|^3 \tag{П5.8}$$

и

$$\sum_{k=1}^{n} |C_{kn}| \leq t^2 + \varepsilon |t|^3. \tag{П5.9}$$

Воспользуемся тем, что для любых комплексных чисел z при $|z| \geq \frac{1}{2}$

$$\ln(1 + z) = z + \theta |z|^2,$$

где $\theta = \theta(z)$ с $|\theta| \leq 1$ и $\ln$ обозначает главное значение логарифма. Тогда для достаточно больших n из (П5.5) и (П5.8) следует, что для достаточно малых $\varepsilon > 0$

$$\ln f_k\left(\frac{t}{D_n}\right) = \ln(1 + C_{kn}) = C_{kn} + \theta_{kn} |C_{kn}|^2,$$

где $|\theta_{kn}| \le 1$. Следовательно, из (П5.1)

$$\frac{t^2}{2} + \ln f_{T_n}(t) = \frac{t^2}{2} + \sum_{k=1}^{n} \ln f_k\Big(\frac{t}{D_n}\Big) =$$

$$= \frac{t^2}{2} + \sum_{k=1}^{n} C_{kn} + \sum_{k=1}^{n} \theta_{kn}|C_{kn}|^2.$$

Но

$$\frac{t^2}{2} + \sum_{k=1}^{n} C_{kn} = \frac{t^2}{2}\Big(1 - \sum_{k=1}^{n} A_{kn}\Big) + t^2 \sum_{k=1}^{n} \tilde{\theta}_1(t,k,n) B_{kn} +$$

$$+ \varepsilon |t|^3 \sum_{k=1}^{n} \tilde{\theta}_2(t,k,n) A_{kn},$$

и, в силу (П5.6), (П5.7), для любого $\delta > 0$ можно найти столь большое n_0 и $\varepsilon > 0$, что для всех $n \ge n_0$

$$\Big|\frac{t^2}{2} + \sum_{k=1}^{n} C_{kn}\Big| \le \frac{\delta}{2}.$$

Далее, в силу (П5.8) и (П5.9),

$$\Big|\sum_{k=1}^{n} \theta_{kn}|C_{kn}|^2\Big| \le \max_{1\le k\le n} |C_{kn}| \sum_{k=1}^{n} |C_{kn}| \le$$

$$\le (t^2\varepsilon^2 + \varepsilon|t|^3)(t^2 + \varepsilon|t|^3).$$

Поэтому для достаточно больших n за счет выбора $\varepsilon > 0$ можно добиться того, что

$$\Big|\sum_{k=1}^{n} \theta_{kn}|C_{kn}|^2\Big| \le \frac{\delta}{2}$$

и, следовательно,

$$\left|\frac{t^2}{2} + \ln f_{T_n}(t)\right| \le \delta.$$

Таким образом, для любого действительного t

$$f_{T_n}(t)e^{\frac{t^2}{2}} \to 1 \quad \text{при} \quad n \to \infty$$

и, значит,

$$f_{T_n}(t) \to e^{-\frac{t^2}{2}} \quad \text{при} \quad n \to \infty.$$

Доказательство завершено. $\square$

Литература

1. Боровков А.А. Теория вероятностей. — М.: Наука, 1986. 432 с.

2. Гихман И.И., Скороход А.В. Введение в теорию случайных процессов. — М.: Наука, 1977. 568 с.

3. Гнеденко Б.В. Курс теории вероятностей. — М.: Наука, 1969. 400с.

4. Зубков А.М., Севастьянов Б.А., Чистяков В.П. Сборник задач по теории вероятностей. — М.: Наука, 1989. 320с.

5. Климов Г.П. Теория вероятностей и математическая статистика. — М.: МГУ, 1983. 328с.

6. Колмогоров А.Н. Основные понятия теории вероятностей. — М.: Наука, 1974. 120с.

7. Лоэв М. Теория вероятностей. — М.: ИЛ, 1962. 720с.

8. Мешалкин Л.Д. Сборник задач по теории вероятностей. — М.: МГУ, 1963. 157с.

9. Неве Ж. Математические основы теории вероятностей. — М.: Мир, 1969. 310с.

10. Партасарати К. Введение в теорию вероятностей и теорию меры. — М.: Мир, 1983. 336с.

11. Прохоров А.В., Ушаков В.Г., Ушаков Н.Г. Задачи по теории вероятностей. — М.: Наука, 1986. 328с.

12. Сборник задач по теории вероятностей, математической статистике и теории случайных функций. // под редакцией Свешникова А.А. — М.: Наука, 1970. 656с.

13. Севастьянов Б.А. Курс теории вероятностей и математической статистики. — М.: Наука, 1982. 256с.

14. Тутубалин В.Н. Теория вероятностей. — М.: МГУ, 1972. 231с.

15. Феллер В. Введение в теорию вероятностей и ее приложения. — М.: Мир, 1967. т.1, 499с., т.2, 752с.

16. Хеннекен П.Л., Тортра А. Теория вероятностей и некоторые ее приложения. — М.: Наука, 1974. 472с.

17. Ширяев А.Н. Вероятность. — М.: Наука, 1980. 576с.

Содержание

1 Пространство элементарных исходов. Событие. Относительная частота 11

2 Вероятность в дискретном пространстве элементарных исходов. Классическое определение вероятности 23

3 Элементы комбинаторики 30

4 Геометрическая вероятность. Аксиоматика теории вероятностей 50

5 Условная вероятность. Независимость событий. Формула полной вероятности и формула Байеса 69

6 Произведение вероятностных пространств. Схема независимых испытаний Бернулли. Полиномиальная схема 90

7 Предельные теоремы в схеме Бернулли 104

8 Борелевские множества и борелевские функции. Пополнение вероятностного пространства 113

9 Случайная величина. Распределение и функция распределения. Закон распределения 126

10 Типы случайных величин 145

11 Случайный вектор. Совместное распределение случайных величин **187**

12 Независимость случайных величин **202**

13 Преобразование случайных величин. Свертка распределений **215**

14 Числовые характеристики случайных величин **227**

15 Виды сходимости случайных величин **266**

16 Законы больших чисел **273**

17 Характеристические преобразования законов распределения **285**

18 Центральная предельная теорема **311**

Приложение 1. Доказательство теоремы о существовании случайной величины с заданной функцией распределения **326**

Приложение 2. Доказательство леммы о слабой сходимости к непрерывной функции **330**

Приложение 3. Доказательство теоремы непрерывности для производящих функций **332**

Приложение 4. Доказательство теоремы непрерывности для характеристических функций **334**

Приложение 5. Доказательство центральной предельной теоремы Линдеберга **345**

Литература **351**

Учебное издание

МАЛИНКОВСКИЙ Юрий Владимирович

Теория вероятностей и математическая статистика (Часть 1. Теория вероятностей). Учебное пособие

Подписано в печать 5.07.2004. ЛВ № 02330/0133208 от 30 апреля 2004г. Формат 60×84 1/16. Бумага офсетная. Печать офсетная. Усл. п. л. 14,0 Уч.-изд. л. 20. Тираж 250 экз. Заказ № 58.

Отпечатано на полиграфической технике УО “ГГУ им. Ф.Скорины”

Лицензия ЛВ № 02330/0056611.

246019 г.Гомель, ул.Советская 104