УДК 519,21

Пытьев Ю. П., Шишмарев И. А. Курс теории вероятностей и математической статистики для физиков: Учеб. пособие. — М.: Изд-во Моск. ун-та, 1983. — 256 с.

В основу книги положен полугодовой курс лекций, читаемый авторами на физическом факультете. Кроме традиционного материала по курсу теории вероятностей большое место уделено важной для физики теории случайных процессов: марковских и стационарных. Изложение математически строгое, хотя и не основанное на использовании интеграла Лебега. Часть курса, посвященная математической статистике, наряду с традиционными вопросами содержит разделы, ориентированные на приложения к задачам автоматизации планирования, анализа и интерпретации физических экспериментов. Изложена статистическая теория измерительно-вычислительного комплекса «прибор+ЭВМ», позволяющая существенно улучшить параметры реального экспериментального оборудования путем обработки данных на ЭВМ. Включены элементы теории статистической проверки гипотез, используемые в задаче интерпретации экспериментальных данных.
Библиогр. 14 наз. Ил. 31

Рецензенты:

кафедра прикладной математики МИЭМ,
акад. АН УССР Б. В. Гнеденко

$\text{П}\frac{1702060000-161}{077(02)-83}$ 72 — 83

ОГЛАВЛЕНИЕ

ПРЕДИСЛОВИЕ

Предлагаемая книга является расширенным вариантом ротапринтного конспекта лекций, читавшихся нами в течение ряда лет на физическом факультете МГУ. При создании курса лекций и настоящего краткого учебного пособия мы ставили перед собой три основные задачи.

Во-первых, курс должен достаточно полно отражать основные понятия и факты теории вероятностей и математической статистики. Выполнение этой задачи предопределило включение в книгу традиционного материала: такого, как пространство элементарных событий, различные определения вероятности событий, последовательность независимых испытаний, предельные теоремы Пуассона и Муавра — Лапласа, случайные величины и функции распределения, математическое ожидание, дисперсия, матрица ковариаций случайных величин, законы больших чисел, характеристические функции и центральные предельные теоремы, цепи Маркова и основные типы случайных процессов, интервальные и точечные оценки параметров распределений, метод наименьших квадратов и линейный анализ регрессий, проверка статистических гипотез и т. д.

Во-вторых, курс должен быть математическим курсом, рассчитанным на математическую подготовку студентов третьего курса физических и физико-технических специальностей университетов. Решение этой задачи, по замыслу авторов, способствовало бы заполнению известной бреши между подробными и строгими курсами теории вероятностей, рассчитанными лишь на математиков и недоступными для физиков, инженеров и других прикладников, и остальными учебниками (для нематематиков), в которых, по существу, мало что доказывается и которые поэтому создают у читателя лишь «комплекс неполноценности». Хотя, конечно, последовательное изложение теории вероятностей должно опираться на теорию меры и интеграла Лебега либо ограничиваться счетными схемами, нематематику, однако, обычно бывает достаточно знать, что возникающие в ряде пунктов теории вопросы существования (вероятностных пространств, случайных величин или функций и т. д.) благополучно решаются с помощью соответствующих математических средств, и ограничиваться в подобных ситуациях проведением доказательств для дискретного случая. Мы придерживались в нашем курсе именно такой точки зрения, избегая использования теории меры и интеграла Лебега.

В-третьих, курс должен отвечать специфическим требованиям физического образования. В связи с этим в книге дано достаточно подробное изложение теории случайных процессов с разбором важных для приложений примеров. Включены вопросы статистической обработки эксперимента. Поскольку особую роль для физиков играют задачи автоматизации экспериментальных исследований, в книге приведены основы теории «прибор+ЭВМ», дающей трактовку измерительно-вычислительного комплекса как нового прибора с существенно улучшенными характеристиками. Рассматривается также ряд других вопросов, например задачи принятия статистических решений.

ЧАСТЬ 1

ТЕОРИЯ ВЕРОЯТНОСТЕЙ

ВВЕДЕНИЕ

Начнем курс теории вероятностей с описания некоторых простых экспериментов, в которых возникает интуитивное понятие вероятности. Анализ этих экспериментов позволит лучше ориентироваться в дальнейшем формальном построении теории вероятностей.

1°. Опыт с конечным числом равновероятных исходов

Рассмотрим эксперимент, который выполняется при соблюдении некоторого комплекса условий U. Будем предполагать, что при фиксированном U эксперимент может быть повторен неограниченное число раз, но при повторении результаты его могут быть различными. Иными словами, речь идет об эксперименте со случайным исходом. Теория вероятностей изучает математические модели таких экспериментов.

В качестве первого эксперимента со случайным исходом рассмотрим бросание игральной кости*. Результат эксперимента, такой, например, как выпадение одной из шести граней, можно считать непредсказуемым (случайным). Исходом эксперимента в данном случае не обязательно считать выпадение одной из граней. Можно, например, условиться, что эксперимент имеет не шесть, а лишь три исхода: A_1 — «выпадение одной из граней 1, 2 или 3», A_2 — «выпадение одной из граней 4 или 5» и, наконец, A_3 — «выпадение грани 6». Но и в этом случае удобно априори выделить в известном смысле элементарные исходы — выпадения граней, а все остальные описывать в терминах элементарных. Дело в том, что в рассматриваемом эксперименте ни один из элементарных исходов нельзя считать более предпочтительным, более вероятным, чем другой, и если n — общее число элементарных исходов (в данном случае — 6), то естественно считать элементарные исходы равновероятными и каждому приписать одинаковую вероятность, равную $1/n$. Так определенная вероятность на практике призвана оценивать частоту данного элементарного исхода в серии большого числа независимых повторений эксперимента, если под частотой понимать отношение числа появлений данного элементарного исхода к общему числу повторений эксперимента. Вслед за этим можно «вычислить» вероятность любого результата эксперимента. Именно если $n(A)$ — число элементарных исходов, приводящих к результату эксперимента A, то естественно вероятность $P(A)$ определить равенством $P(A)=n(A)/n$ — как отношение числа исходов, приводящих к A, к числу всех элементарных исходов.

* Игральная кость — куб из однородного материала, шесть граней которого перенумерованы.

Итак, вероятность выпадения каждой грани при бросании кости равна 1/6, вероятность исхода A — «выпадение либо 2, либо 3» — равна $(1+1)/6=1/6+1/6=1/3$ и т. п. При этом в понятие «вероятность» вкладывается следующий интуитивный смысл: при многократном повторении эксперимента мы ожидаем, что отношение фактического числа исходов A к общему числу повторений эксперимента будет близко к вероятности $P(A)$. В данном случае следует ожидать, что каждая грань при большом числе бросаний будет выпадать примерно в одной шестой всех исходов, а результат эксперимента A_1 будет наблюдаться втрое чаще, т. е. примерно в половине исходов.

Так это или не так, в каждом конкретном случае может свидетельствовать лишь реальный эксперимент. Накопленные на практике многочисленные наблюдения действительно подтверждают факт устойчивости частот в рассмотренном эксперименте. При большом числе бросаний частота выпадения каждой грани и на самом деле близка к 1/6. Однако не следует думать, что всякий эксперимент со случайным исходом обладает свойством устойчивости частот, или, как говорят, статистической устойчивостью. В теории вероятностей речь идет об экспериментах со свойством статистической устойчивости результатов.

2°. Геометрические вероятности

Интуитивное представление о вероятности может быть составлено также в связи со следующим мысленным экспериментом. Пусть на отрезок $[a, b]$ длины $l=b-a$ наугад бросается точка. Какова вероятность того, что она попадет на отрезок $[\alpha, \beta]$, содержащийся в $[a, b]$?

Ответ в данном случае очевиден. Поскольку вероятность попасть в $[\alpha, \beta]$ не зависит от того, где именно на $[a, b]$ расположен отрезок $[\alpha, \beta]$, то искомая вероятность $P(A)$ равна $(\beta-\alpha)/l$, т. е. отношению длин отрезков $[\alpha, \beta]$ и $[a, b]$. Здесь A обозначает факт попадания точки на $[\alpha, \beta]$. Ответ будет таким же, если вместо отрезка $[\alpha, \beta]$ выбрать любое подмножество отрезка $[a, b]$, лишь бы для него можно было определить длину и последняя равнялась $\beta-\alpha$.

Понятно, что наш вывод целиком обусловлен интерпретацией условий эксперимента, согласно которым точка бросается на $[a, b]$ наугад.

Для рассмотренного эксперимента характерно, что возможно бесконечное множество (даже континуум) элементарных исходов — попаданий точки на отрезок $[a, b]$. В таких случаях вероятность удобно задавать с помощью так называемой плотности вероятности. В примере с бросанием точки плотность вероятности $p(\cdot)$ определяется равенством $p(x)=1/l$, $x\in[a, b]$, причем $P(A)=\int_\alpha^\beta dx/l=\int_\alpha^\beta p(x)\,dx$, $\int_a^b p(x)\,dx=1$. Возвращаясь к эксперименту с игральной костью, заметим, что теперь суммированию вероятностей элементарных исходов отвечает интегрирование плотности вероятности по множеству, соответствующему исходу

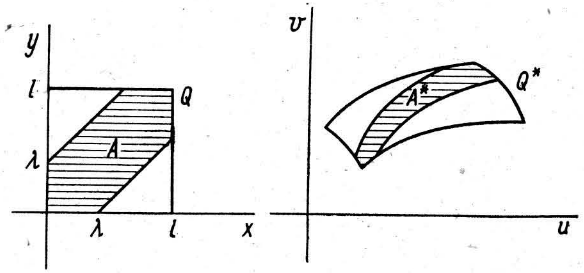


Рис. 1

A, а вероятность каждого элементарного исхода равна нулю.

Пусть на отрезок $[a, b]$ наугад бросаются две точки. Какова вероятность того, что расстояние между ними окажется не больше λ, $0\leqslant\lambda\leqslant l$? Задача, очевидно, эквивалентна следующей: в квадрат $Q=\{(x, y)\colon 0\leqslant x\leqslant l,\ 0\leqslant y\leqslant l\}$ наугад бросается точка (x, y), какова вероятность того, что $|x-y|\leqslant\lambda$? Иначе говоря, какова вероятность того, что точка попадет в заштрихованную область квадрата на рис. 1? Искомая вероятность, очевидно, равна отношению площади заштрихованной области к площади квадрата Q:

$$P(A)=\int_A \frac{dxdy}{l^2}=(l^2-(l-\lambda)^2)/l^2.$$

Заметим, что если $u=u(x, y)$, $v=v(x, y)$ — криволинейные координаты на плоскости $\{(x, y)\}$ и на плоскости $\{(u, v)\}$ квадрат Q представлен фигурой Q^*, то вероятность попасть в указанную на рис. 1 область A квадрата

Q может быть получена по формуле $P(A)=\int_A \frac{dxdy}{l^2}=$ $=\int_{A^*}\left|\frac{\partial(x, y)}{\partial(u, v)}\right|\frac{dudv}{l^2}$. В свою очередь, последнее равенство можно интепретировать следующим образом: $P(A)$ — вероятность точке (u, v) попасть в область A^*, если характер «бросаний» точки в область Q^* контролируется плотностью вероятности $p(u,v)=\frac{1}{l^2}\left|\frac{\partial(x, y)}{\partial(u, v)}\right|$, $(u, v)\in Q^*$, заданной на плоскости $\{(u, v)\}$. В область Q^* точки «бросаются» не наугад.

Рассмотрим частицу с энергией $E=mv^2/2$, движущуюся в случайном направлении. Пусть (v_1, v_2, v_3) — вектор скорости частицы в некоторой системе координат. Какова вероятность того, что $\alpha\leqslant v_1\leqslant\beta$? Искомая вероятность $P(\alpha, \beta)$ равна отношению площади заштрихованной полоски к площади сферы радиуса v, изображенных на рис. 2. Последняя равна $4\pi v^2=8\pi E/m$, а площадь полоски дается интегралом

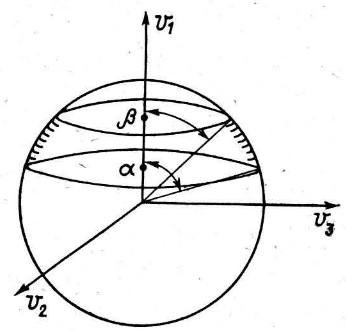


Рис. 2

$$v^2\int_0^{2\pi} d\varphi\int_{\theta_\beta}^{\theta_\alpha}\sin\theta\,d\theta=v^2\cdot 2\pi\frac{\beta-\alpha}{v}=2\pi\sqrt{\frac{2E}{m}}(\beta-\alpha),\ -v\leqslant\alpha\leqslant\beta\leqslant v.$$

Поэтому $P(\alpha, \beta)=(\beta-\alpha)\Big/\left(2\sqrt{\frac{2E}{m}}\right)$.

Если речь идет о системе n одинаковых частиц с фиксированной полной энергией $E=\sum_{i=1}^n mv_{(i)}^2/2$, то вероятность того, что $\alpha\leqslant v_{(i)1}\leqslant\beta$ может быть подсчитана вполне аналогично, если рассмотреть полоску $\alpha\leqslant v_{(i)1}\leqslant\beta$ на $3n$-мерной сфере $\sum_{i=1}^n (v_{(i)1}^2+v_{(i)2}^2+v_{(i)3}^2)=2E/m$.

Наконец, если n — число наблюдаемых частиц, движущихся в случайных направлениях, энергия каждой из которых равна $mv^2/2$ и $n_1(\alpha, \beta)/n$ — относительная доля тех из них, для которых $\alpha\leqslant v_1\leqslant\beta$, то при $n\to\infty$ вероятность любого отличия $P(\alpha, \beta)$ и $n_1(\alpha, \beta)/n$ стремится к нулю. Этот замечательный результат, известный как закон больших чисел, в дальнейшем будет доказан.

Приступим теперь к более точным построениям.

§ 1. ПРОСТРАНСТВО ЭЛЕМЕНТАРНЫХ СОБЫТИЙ. АЛГЕБРА СОБЫТИЙ

В рассмотренных во введении примерах наглядно выступают все основные моменты общей теоретико-вероятностной схемы. В этом параграфе они будут выделены и точно определены.

1°. Пространство элементарных событий

В общей теоретико-вероятностной схеме для каждого эксперимента со случайным исходом должны быть указаны все элементарные исходы, отвечающие следующему требованию: в результате эксперимента непременно происходит один и только один из этих исходов. Каждый такой исход принято называть элементарным событием; обозначать элементарные события будем буквой ω. По смыслу элементарные события неразложимы на «более элементарные».

В эксперименте с игральной костью элементарными событиями являются грани «1», «2», ..., «6». При этом считается, что не может выпасть ребро или вершина кости, хотя в принципе такое явление возможно. В эксперименте с бросанием точки на отрезок $[a, b]$ элементарным событием является точка на $[a, b]$. Соответственно в эксперименте с двумя точками элементарным событием является пара точек на $[a, b]$, или точка в квадрате $[a, b]\times[a, b]$. Наконец, в примере с частицей элементарным событием является точка на сфере радиуса v.

Множество всех элементарных событий в теории вероятностей принято называть пространством элементарных событий. Пространство элементарных событий будем обозначать буквой Ω. Элементарные события называются точками пространства элементарных событий.

Всякий результат эксперимента со случайным исходом в теории вероятностей принято называть событием. Среди всех событий элементарные события выделяются таким образом, что для каждого события A и каждого элементарного события ω известно, влечет ω наступление A или не влечет. Тем самым совокупность всех тех ω, которые влекут A, полностью характеризует A. Обратно: произвольное множество A точек $\omega\in\Omega$ можно рассматривать как событие A, которое происходит или нет в зависимости от того, принадлежит соответственно или нет множеству A элементарное событие ω, представляющее данный исход эксперимента.

Иными словами, событие A можно считать подмножеством Ω, состоящим из точек $\omega\in\Omega$, представляющих те исходы эксперимента, при которых происходит A. Далее по этой причине не делается различий между событием A и соответствующим подмножеством $A\subset\Omega$.

В приведенных во введении примерах экспериментов со случайными исходами события рассматривались как подмножества соответствующих пространств элементарных событий. В эксперименте с игральной костью были выделены три события: $A_1=\{«1», «2», «3»\}$, $A_2=\{«4», «5»\}$ и $A_3=\{«6»\}$. В этом случае, например, любое из элементарных событий

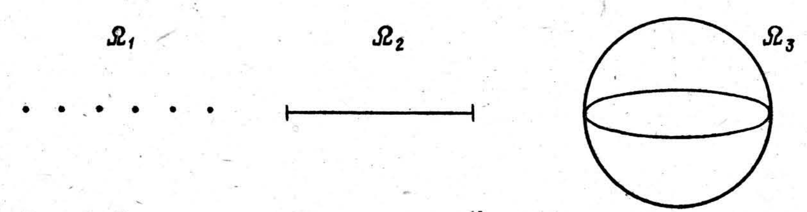


Рис. 3. Ω_1 — пространство элементарных событий в эксперименте с игральной костью, Ω_2 — в эксперименте с бросанием точки на отрезок, Ω_3 — в примере с частицей, движущейся в случайном направлении

$\omega_1=«1»$, $\omega_2=«2»$ или $\omega_3=«3»$ влечет A_1. В эксперименте с бросанием двух точек на отрезок $[a, b]$ событием является заштрихованная область A квадрата $Q=\Omega$. Наконец, в примере с частицей событием является полоска на сфере (см. рис. 1, 2, 3).

2°. Алгебра событий

Рассмотрим математические формулировки естественных операций над событиями и их теоретико-множественные аналоги. Приведенные ниже определения и свойства операций над событиями характеризуют алгебраическую структуру любой теоретико-вероятностной схемы.

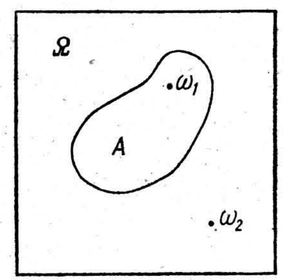


Рис. 4. Элементарное событие ω_1 влечет событие A, $\omega_1\in A$, ω_2 — не влечет, $\omega_2\bar{\in} A$

1. Если событие A происходит всякий раз, когда происходит событие B, то будем говорить, что событие A является **следствием** B, и писать $B\subset A$ или $A\supset B$. В теоретико-множественных терминах это означает, что каждая точка $\omega\in B$ содержится в A, или, иначе, B является подмножеством A. Эта связь аналогична связи между событиями и элементарными событиями: ω влечет A, если $\omega\in A$.

2. Если $A\subset B$ и $B\subset A$, то события A и B происходят или не происходят лишь одновременно. В таком случае будем писать $A=B$; множества A и B при этом совпадают.

3. Событие, состоящее в том, что не происходит событие A, называется **противоположным** событию A и обозначает-

ся $\bar{A}$. Множество $\bar{A}$ состоит из точек Ω, не принадлежащих A, и называется дополнением множества A.

4. Если событие A не содержит ни одного элементарного события, то оно называется **невозможным** и обозначается $\varnothing$. Противоположным $\varnothing$ является, очевидно, событие Ω, которое происходит всякий раз и называется **достоверным.** Наоборот, $\bar{\Omega}=\varnothing$. $\varnothing$, очевидно, пустое подмножество Ω.

5. Событие C, происходящее тогда и только тогда, когда происходят события A и B, называется **произведением**, или пересечением событий A и B, и обозначается AB или $A \cap B$. Множество C состоит из точек, принадлежащих как множеству A, так и множеству B, и называется пересечением множеств A и B. При этом $A \cap B$ обозначает пересечение A и B.

6. События A и B называются **несовместными**, если их одновременное наступление невозможно, т. е. если $A \cap B=\varnothing$. Несовместным событиям отвечают непересекающиеся множества.

7. Событие C, состоящее в наступлении хотя бы одного из событий A или B, называется объединением, или **суммой** событий A и B. Для объединения будем использовать обозначение $A \cup B$, но в том случае, когда $A \cap B=\varnothing$, условимся писать $C=A+B$. В теоретико-множественных терминах C — множество, состоящее из тех точек Ω, которые принадлежат хотя бы одному из множеств A или B. Множество $A \cup B$ также называется объединением, или суммой, множеств A и B.

Замечание. Пересечение и объединение определяется для произвольного числа событий. Например, событие $C=A \cap B \cap \ldots$ состоит в том, что происходят все события A, B, ...; событие $C=A \cup B \cup \ldots$ состоит в том, что происходит хотя бы одно из событий A, B, ... Операции объединения $\cup$ и пересечения $\cap$, очевидно, ассоциативны и коммутативны по определению: $(A \cup B) \cup C=A \cup (B \cup C)$, $A \cup B=B \cup A$, $(A \cap B) \cap C=A \cap (B \cap C)$, $A \cap B=B \cap A$ для любых событий A, B и C.

8. Событие C, состоящее в том, что событие A происходит, а событие B не происходит, называется **разностью** событий A и B и обозначается $A \setminus B$. В теоретико-множественных терминах множество $A \setminus B=A \cap \bar{B}$ состоит из точек, принадлежащих множеству A и не принадлежащих множеству B, и называется разностью множеств A и B.

Отметим простейшие свойства операций над событиями. По определению: $\bar{\bar{A}}=A$, $\bar{A}=\Omega \setminus A$ и $A \setminus B=A \cap \bar{B}=\bar{B} \setminus \bar{A}$. Кроме того, операции над событиями $\cup$ и $\cap$ взаимно дистрибутивны:

$$A \cup (B \cap C)=(A \cup B) \cap (A \cup C), \qquad (1)$$

$$A \cap (B \cup C)=(A \cap B) \cup (A \cap C).$$

Наконец, в теории вероятностей и ее приложениях важную роль играет так называемый принцип двойственности, который может быть выражен следующими соотношениями:

$$\overline{A \cup B}=\bar{A} \cap \bar{B}, \qquad (2)$$

$$\overline{A \cap B}=\bar{A} \cup \bar{B}.$$

Докажем, например, первое равенство в (2), рассматривая его как равенство событий. Событие $\overline{A \cup B}$ согласно определению 3 состоит в том, что не происходит событие $A \cup B$, состоящее в наступлении хотя бы одного из событий A или B (определение 7). Но это буквально и означает, что не происходит ни A, ни B. Другими словами, согласно определениям 3 и 5, происходит событие $\bar{A} \cap \bar{B}$. Следовательно, событие $\overline{A \cup B}$ влечет событие $\bar{A} \cap \bar{B}$: $\overline{A \cup B} \subset \bar{A} \cap \bar{B}$. Наоборот, если происходит событие $\bar{A} \cap \bar{B}$, то не происходят ни A, ни B. Следовательно, не происходит событие $A \cup B$, т. е. происходит $\overline{A \cup B}$. Поэтому $\bar{A} \cap \bar{B} \subset \overline{A \cup B}$.

Докажем теперь то же самое в терминах операций над множествами. Для этого достаточно показать, что всякий элемент множества $\overline{A \cup B}$ содержится в $\bar{A} \cap \bar{B}$ (и тем самым $\overline{A \cup B} \subset \bar{A} \cap \bar{B}$), и, наоборот, всякий элемент множества $\bar{A} \cap \bar{B}$ содержится в $\overline{A \cup B}$ (т. е. $\bar{A} \cap \bar{B} \subset \overline{A \cup B}$). Доказательство следует из системы соотношений:

$$\omega \in \overline{A \cup B} \rightleftarrows \omega \,\overline{\in}\, A \cup B \rightleftarrows \omega \,\overline{\in}\, A,\ \omega \,\overline{\in}\, B \rightleftarrows$$

$$\rightleftarrows \omega \in \bar{A},\ \omega \in \bar{B} \rightleftarrows \omega \in \bar{A} \cap \bar{B}, \qquad (3)$$

в которых символ $\rightleftarrows$ следует читать как «эквивалентно». Читая соотношения (3) слева направо, получим включение $\overline{A \cup B} \subset \bar{A} \cap \bar{B}$. Если (3) прочитать справа налево, то найдем, что $\bar{A} \cap \bar{B} \subset \overline{A \cup B}$. Следовательно, $\overline{A \cup B}=\bar{A} \cap \bar{B}$.

Разумеется, принцип двойственности справедлив для любого множества событий:

$$\overline{\bigcup_{\alpha \in S} A_\alpha}=\bigcap_{\alpha \in S} \bar{A}_\alpha; \quad \overline{\bigcap_{\alpha \in S} A_\alpha}=\bigcup_{\alpha \in S} \bar{A}_\alpha;$$

здесь символ $\bigcup\limits_{\alpha \in S}$ $\left(\bigcap\limits_{\alpha \in S}\right)$ означает объединение (пересечение) множества событий A_α, отличающихся индексом $\alpha \in S$, который может пробегать и несчетное множество значений.

К принципу двойственности следует отнести еще одно соотношение

$$A \subset B \rightleftarrows \bar{A} \supset \bar{B}, \qquad (4)$$

доказательство которого очевидно.

Роль принципа двойственности в теории вероятностей состоит в том, что для всякого утверждения, относящегося к некоторой системе событий, может быть сформулировано эквивалентное ему двойственное утверждение, в котором события должны быть заменены на соответственно противоположные, объединения — на пересечения и наоборот, и учтено соотношение (4). Например, утверждение $(A\cup B)\cap C\subset M\setminus N$ эквивалентно $\overline{(A\cup B)\cap C}\supset\overline{M\setminus N}$, а последнее может быть преобразовано к виду $(\bar{A}\cap\bar{B})\cup\bar{C}\supset N\cup\bar{M}$, поскольку $\overline{(A\cup B)\cap C}=\overline{(A\cup B)}\cup\bar{C}=(\bar{A}\cap\bar{B})\cup\bar{C}$ и $\overline{M\setminus N}=\overline{M\cap\bar{N}}=\bar{M}\cup\bar{\bar{N}}=\bar{M}\cup N$.

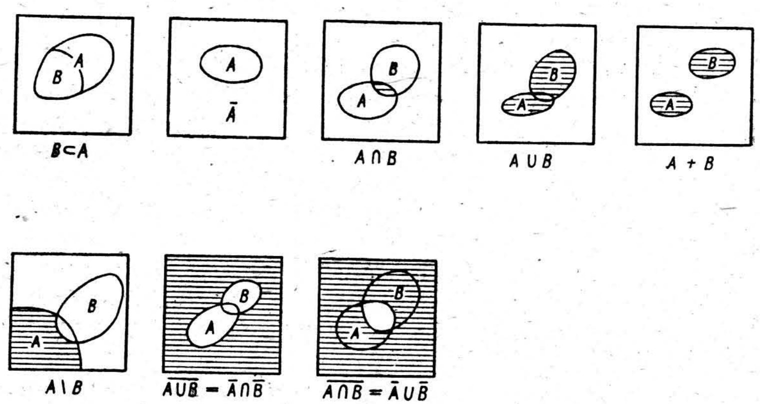


Рис. 5

Возвращаясь к соотношениям дистрибутивности (1), заметим, что и для произвольных систем событий

$$A\cup\left(\bigcap_{\alpha\in S}A_\alpha\right)=\bigcap_{\alpha\in S}(A\cup A_\alpha);\quad A\cap\left(\bigcup_{\alpha\in S}A_\alpha\right)=\bigcup_{\alpha\in S}(A\cap A_\alpha).\qquad(5)$$

В частности,

$$A\cap(B+C+D+\ldots)=A\cap B+A\cap C+A\cap D+\ldots.$$

Докажем, например, первое соотношение в (5). Событие $A\cup\left(\bigcap_{\alpha\in S}A_\alpha\right)$ происходит тогда и только тогда, когда происходит либо событие A, либо событие $\bigcap_{\alpha\in S}A_\alpha$ (т. е. все события A_α, $\alpha\in S$), либо, наконец, A и $\bigcap_{\alpha\in S}A_\alpha$. С другой стороны, событие $\bigcap_{\alpha\in S}(A\cup A_\alpha)$ происходит тогда и только тогда, когда происходят все события $A\cup A_\alpha$, $\alpha\in S$, т. е. тогда и только тогда, когда происходят либо A, либо все A_α, $\alpha\in S$, либо, наконец, когда происходят A и все A_α, $\alpha\in S$. Следовательно, события $A\cup\left(\bigcap_{\alpha\in S}A_\alpha\right)$ и $\bigcap_{\alpha\in S}(A\cup A_\alpha)$ равны.

Второе равенство в (5) читателю следует доказать самостоятельно.

Рассмотренные свойства операций над событиями носят алгебраический характер. В теории вероятностей принимается, что класс $\mathcal{F}$ всех возможных событий должен удовлетворять следующим требованиям.

1) Для каждой пары событий A и B из включений $A\in\mathcal{F}$, $B\in\mathcal{F}$ следует включение $A\cup B\in\mathcal{F}$. Иными словами, класс $\mathcal{F}$ вместе с каждой парой событий содержит их объединение.

2) Вместе с каждым событием A класс $\mathcal{F}$ содержит противоположное событие $\bar{A}$.

Если класс $\mathcal{F}$ не пуст, то отсюда следует, что $\Omega\in\mathcal{F}$, так как $\Omega=A+\bar{A}$. Следовательно, $\varnothing\in\mathcal{F}$, поскольку $\varnothing=\bar{\Omega}$. Наконец, так как согласно принципу двойственности $A\cap B=\overline{\bar{A}\cup\bar{B}}$, то класс $\mathcal{F}$ вместе с каждой парой событий A и B содержит их пересечение $A\cap B$ и разность $A\setminus B=A\cap\bar{B}$.

Класс $\mathcal{F}$ событий, удовлетворяющий условиям 1 и 2, называется алгеброй событий.

Конструкция алгебры событий позволяет охарактеризовать множество всех возможных результатов любого эксперимента со случайным исходом, если множество Ω его элементарных исходов конечно. Например, в эксперименте с игральной костью Ω состоит из шести элементарных событий, а $\mathcal{F}$ состоит из всех подмножеств Ω. Поскольку $\mathcal{F}$ содержит пустое подмножество $\varnothing$, $6=C_6^1$ одноточечных подмножеств, $15=C_6^2$ двухточечных, $20=C_6^3$ трехточечных, ..., одно (C_6^6) шеститочечное, то $\mathcal{F}$ состоит из $2^6=1+C_6^1+C_6^2+\ldots+C_6^6=64$ событий. И вообще, если Ω состоит из n элементарных событий, то $\mathcal{F}$ состоит, очевидно, из $2^n=C_n^0+C_n^1+\ldots+C_n^n$ событий.

Совсем не так просто обстоит дело в эксперименте с бросанием точки на отрезок. Здесь в качестве системы событий в дальнейшем придется выделить специальный класс подмножеств, более широкий, чем алгебра. Но, с другой стороны, при попытке использовать в качестве событий все подмножества отрезка мы столкнулись бы с серьезными трудностями.

§ 2. КЛАССИЧЕСКАЯ ТЕОРЕТИКО-ВЕРОЯТНОСТНАЯ МОДЕЛЬ

Пусть Ω — конечное или бесконечное пространство элементарных событий некоторого случайного эксперимента. Предположим, что структура эксперимента такова, что на Ω можно указать n событий A_1, A_2, ..., A_n, обладающих следующими свойствами.

1) События $A_1, \ldots, A_n$ попарно несовместны в том смысле, что никакие два из них не могут произойти одновременно. Иначе говоря, $A_i \cap A_j = \varnothing$, $i \neq j$, $i, j = 1, \ldots, n$.

2) $A_1, \ldots, A_n$ образуют полную группу событий в том смысле, что при любом исходе эксперимента хотя бы одно из них непременно происходит. Это означает, что

$$A_1 + A_2 + \ldots + A_n = \Omega.$$

3) События $A_1, \ldots, A_n$ равновероятны, или, иначе говоря, ни одно из них нельзя считать более предпочтительным, чем любое из остальных. Это, разумеется, доматематическое требование, и не следует стремиться переформулировать его в терминах каких-либо более элементарных свойств. Можно считать, что оно отражает свойство некоторой относительной симметрии событий $A_1, \ldots, A_n$, подсказанное здравым смыслом. Именно такая ситуация возникла в эксперименте с игральной костью, когда выпадения всех граней были объявлены равновероятными.

В так называемой классической схеме события $A_1, \ldots, A_n$, удовлетворяющие условиям 1—3, называются полной группой попарно несовместных, равновероятных событий.

Вероятность в классической схеме определяется лишь для тех исходов эксперимента, которые могут быть представлены в виде объединений некоторых из событий A_i, $i = 1, \ldots, n$. Именно если

$$A = A_{i_1} + \ldots + A_{i_k} \tag{1}$$

и все слагаемые в (1) различны, то вероятность события A определяется равенством

$$P(A) = k/n, \tag{2}$$

в котором k равно числу слагаемых в сумме (1). Таково классическое определение вероятности.

Для того чтобы определение (2) можно было считать корректным, достаточно доказать единственность разложения (1). Но для любого события A согласно условию 2 и свойству дистрибутивности (1.5)

$$A = A \cap \Omega = A \cap (A_1 + \ldots + A_n) = A \cap A_1 + \ldots + A \cap A_n.$$

Поэтому в рассматриваемом случае разложения (1) $A \cap A_j$ либо пусто, если j не совпадает ни с одним из i_s, $s = 1, \ldots, k$, либо $A \cap A_j = A_{i_s}$, если $j = i_s$.

Традиционно приложениям классического определения вероятности сопутствует следующая терминология. Эксперимент называют испытанием, полную группу попарно несовместных, равновероятных событий называют полной группой возможных исходов испытания, а те из возможных исходов,

из которых складывается событие A, называют исходами, благоприятствующими появлению A. В этих терминах согласно определению (2) $P(A)$ равно отношению числа исходов, благоприятствующих появлению A, к числу всех возможных исходов.

Как правило, отыскание вероятностей, основанное на классическом определении, сводится к комбинаторным вычислениям. Рассмотрим примеры, позволяющие уяснить технику вычисления классических вероятностей. Напомним вначале простейшие комбинаторные формулы.

Размещения. Перестановки. Имеется n различных объектов $x_1, \ldots, x_n$. Сколькими способами можно образовать последовательность $x_{j_1}, \ldots, x_{j_r}$, $r \leqslant n$, различных объектов?

Объект x_{j_1} можно выбрать n способами. Если x_{j_1} выбран, x_{j_2} можно выбрать $n-1$ способом из оставшихся объектов и т. д. Всего таким образом существуют $n(n-1)\ldots(n-r+1) = A_n^r$ способов образовать последовательность из r объектов, выбирая объекты из совокупности $x_1, \ldots, x_n$.

Иначе эту задачу можно сформулировать следующим образом: сколькими способами можно разместить r из n различных объектов по r местам? A_n^r называется числом размещений из n по r.

Если $r = n$, то $A_n^r = n(n-1)\ldots2\cdot1 = n!$ называется числом перестановок (n различных объектов). Поскольку для $n \geqslant 1$

$$1 < \frac{n!}{\sqrt{2\pi n}\, n^n e^{-n}} < e^{1/(12n)},$$

то $n!$ при больших n можно вычислить по формуле Стирлинга: $n! \sim \sqrt{2\pi n}\, n^n e^{-n}$. Здесь символ $\sim$ означает, что отношение левой и правой частей стремится к единице при $n \to \infty$. Для $n = 0$ полагают $0! = 1$.

Сочетания. Имеется n различных объектов $x_1, \ldots, x_n$. Выбирая объекты из $x_1, \ldots, x_n$, сколькими способами можно образовать множество X_r из r объектов, $r \leqslant n$?

Поскольку в данном случае порядок объектов, образующих множество X_r, несуществен, то искомое число способов равно $C_n^r = A_n^r/r! = \dfrac{n!}{r!(n-r)!}$. C_n^r называется числом сочетаний из n по r.

Выбор с возвращением. Имеется n различных объектов $x_1, \ldots, x_n$, из которых последовательно выбирается объект, фиксируется и возвращается обратно. Сколькими способами может быть образована выборка $x_{j_1}, \ldots, x_{j_r}$, зарегистрированная за r шагов?

Поскольку каждый раз объект может быть выбран n способами, существует всего n^r способов образовать выборку $x_{j_1}, \ldots, x_{j_r}$.

Задача о днях рождения. Предположим, что в аудитории n студентов. Какова вероятность того, что хотя бы у двоих совпадают дни рождения? Разумеется, такая постановка задачи требует уточнения. Будем считать, что в году 365 дней, у каждого студента есть день рождения, причем им может оказаться любой из 365 дней с одной и той же вероятностью.

Заметим, что для групп из $n=366$ и более человек искомая вероятность равна единице, т. е. для $n \geqslant 366$ по меньшей мере у двоих непременно совпадают дни рождения. Поэтому рассмотрим случай $n \leqslant 365$. Для группы из n человек возможно $(365)^n$ комбинаций дней рождения, поскольку таких возможностей для каждого человека 365. Все эти комбинации образуют полную группу из попарно несовместных, равновероятных событий, причем вероятность каждой комбинации равна $(365)^{-n}$. Число различных комбинаций дней рождения, в которых ни один день рождения не встречается более одного раза, равно $\bar{n}=365 \cdot 364 \cdot \ldots \cdot (365-n+1)$. Это число получено следующим образом: первый день рождения можно выбрать 365 способами, после этого второй день рождения можно выбрать $(365-1)$ способами, ..., наконец, n-й день рождения можно выбрать $(365-n+1)$ способами. Следовательно, согласно классическому определению вероятность того, что в группе из n студентов на каждый день приходится не более одного дня рождения, равна $P_n = \bar{n}(365)^{-n}$.

Подсчитав число комбинаций, в которых нет совпадающих дней рождения, нетрудно сообразить, что во всех оставшихся комбинациях имеются по меньшей мере два совпадающих дня рождения. Таких комбинаций $(365)^n - \bar{n}$. Следовательно, искомая вероятность равна $[(365)^n - \bar{n}]/(365)^n = 1-P_n$.

5	10	20	30	40	50	60	n
0,027	0,117	0,411	0,706	0,891	0,970	0,994	$1-P_n$

Как видно из приведенной таблицы, искомая вероятность становится довольно близкой к единице для групп, существенно меньших 366 человек. Такой вывод априори далеко не очевиден.

Гипергеометрическое распределение. Дана совокупность n объектов, среди которых k отмеченных (например, бракованных изделий, выигрышных билетов и т. п.). Выбирается наугад $n_1 \leqslant n$ объектов. Какова вероятность того, что среди них окажется k_1 отмеченных?

Выбрать n_1 объектов из n можно $C_n^{n_1}$ различными способами. k_1 отмеченных объектов из общего их числа k можно выбрать $C_k^{k_1}$ способами, причем каждому такому способу соответствует $C_{n-k}^{n_1-k_1}$ способов добрать еще n_1-k_1 объектов до общего числа n_1, выбирая их из $n-k$ неотмеченных. Следовательно, число способов, благоприятствующих появлению k_1 отмеченных объектов среди n_1 выбранных, равно $C_k^{k_1} C_{n-k}^{n_1-k_1}$. Поэтому искомая вероятность равна

$$P_{k,n}(k_1, n_1) = \frac{C_k^{k_1} C_{n-k}^{n_1-k_1}}{C_n^{n_1}}, \quad k_1 = 0, \ldots, \min(k, n_1). \tag{3}$$

Совокупность вероятностей (3) носит название гипергеометрического распределения.

Применим гипергеометрическое распределение к анализу вероятностей выигрышей при игре «спортлото». В данном случае $n=49$ (число наименований видов спорта), $k=6$ (число отмеченных видов спорта), $n_1=6$ (число выбранных видов спорта). Следовательно, вероятность угадывания k_1 видов спорта равна

$$P_{6,49}(k_1, 6) = C_6^{k_1} C_{49-6}^{6-k_1} / C_{49}^6.$$

В том числе вероятность угадывания всех шести видов спорта равна

$$P_{6,49}(6,6) = 1/C_{49}^6 \approx 7{,}15 \cdot 10^{-8}, \text{ а пяти} - P_{6,49}(5,6) =$$

$$= C_6^5 C_{43}^1 / C_{49}^6 \approx 1{,}84 \cdot 10^{-5}.$$

Рассмотрим несколько важных задач на размещения, возникающих при изучении некоторых систем частиц в физике и статистической механике.

Система Максвелла — Больцмана характеризуется как система r различных частиц, каждая из которых может находиться в одной из n ячеек (состояний) вне зависимости от того, где при этом находятся остальные частицы. В такой системе возможно всего n^r различных размещений r частиц по n ячейкам. Если при этом все такие размещения (состояния системы) считаются равновероятными, то говорят о статистике Максвелла — Больцмана. Вероятность каждого состояния равна n^{-r}.

Система Бозе — Эйнштейна определяется как система r неразличимых (тождественных) частиц, каждая из которых независимо от остальных может находиться в одной из n ячеек (состояний). Поскольку частицы неразличимы, каждое состояние такой системы задается «числами заполнения» $r_1, r_2, \ldots, r_n$, где r_j — число частиц в j-й ячейке. Подсчитаем число различных состояний системы (т. е. число размещений частиц, различающихся лишь числами заполнения). Состояние системы удобно условно представить, как это показано на рис. 6. Здесь изображена конфигурация из r точек (частиц) и $n+1$ черточек (границ ячеек). Понятно, что каждая такая конфигурация задает размещение неразличимых частиц по ячейкам и наоборот, если задано состояние системы в терминах чисел заполнения, то ему соответст-

| • • | • | • | | • • • |

Рис. 6

вует одна конфигурация. Каждая конфигурация, в свою очередь, полностью определяется положениями внутренних $n-1$ черточек, которые могут находиться в $n+r-1$ позициях. Следовательно, имеется всего C_{n+r-1}^{n-1} различных конфигураций и столько же различных состояний рассматриваемой системы частиц.

Если все состояния системы равновероятны, то говорят о статистике Бозе — Эйнштейна. При этом вероятность каждого состояния системы равна $1/C_{n+r-1}^{n-1}$.

Заметим, что если дополнительно потребовать, чтобы в каждом состоянии системы ни одна ячейка не оставалась пустой, то число возможных состояний системы сократится до C_{r-1}^{n-1}. При этом, разумеется, частиц должно быть не меньше, чем ячеек, $r \geqslant n$.

Чтобы получить этот результат, следует воспользоваться найденным ранее числом состояний системы без ограничений, предварительно «приклеив» к каждой из n черточек справа по одной точке, исключив последнюю $(n+1)$-ю (правую) черточку. В таком случае, переставляя черточки, мы будем получать состояния, при которых в каждой ячейке будет находиться не менее одной частицы. Теперь однако для $n-1$ черточек оказывается не $n+r-1$, а лишь $r-1$ вакантных мест (на n меньше).

Система Ферми — Дирака определяется как система Бозе — Эйнштейна, в которой дополнительно действует принцип запрета (принцип Паули), требующий, чтобы в каждой ячейке находилось не более одной частицы.

Поскольку и в этом случае частицы неразличимы, состояния системы характеризуются числами $r_j=0, 1$, $j=1, \ldots, n$. Однако в данном случае непременно $r \leqslant n$.

Задать состояние системы можно, указав заполненные ячейки. Последние можно выбрать C_n^r способами; столько же состояний системы Ферми — Дирака. Если все состояния равновероятны, то говорят о статистике Ферми — Дирака. Вероятность каждого состояния в таком случае равна $1/C_n^r$.

В классической статистической физике статистике Максвелла — Больцмана подчинены, как известно, системы молекул газа. Системы частиц с целым и полуцелым спином подчиняются соответственно статистикам Бозе — Эйнштейна и Ферми — Дирака.

Для того чтобы рассмотреть **свойства классической вероятности**, удобно несколько упростить и формализовать классическую теоретико-вероятностную модель. Будем считать события $A_1, \ldots, A_n$, образующие полную группу попарно несовместных равновероятных событий, точками $\omega_1, \ldots, \omega_n$ нового пространства элементарных событий, для которого сохраним прежнее обозначение Ω. Другими словами, положим

$$\Omega=\{\omega_1\}+\{\omega_2\}+\ldots+\{\omega_n\},$$

где $\{\omega_i\}$ обозначает множество, состоящее из одной точки ω_i (элементарного события), $i=1, \ldots, n$. Для каждого элементарного события ω_i определим вероятность

$$P(\{\omega_i\})=\frac{1}{n}, \quad i=1, \ldots, n.$$

Алгебра $\mathscr{F}$ событий в данном случае состоит из невозможного события $\varnothing$ и всевозможных объединений одноточечных множеств $\{\omega_i\}$, $i=1, \ldots, n$, всего — из $C_n^0+C_n^1+\ldots+C_n^n=2^n$ событий.

Для любого события $A \in \mathscr{F}$ вероятность $P(A)$ определим равенством $P(A)=m/n$, где m — число элементарных событий, из которых состоит A.

Формально классическая теоретико-вероятностная модель, очевидно, эквивалентна тройке $(\Omega, \mathscr{F}, P(\cdot))$, состоящей из пространства элементарных событий Ω, содержащего n точек, алгебры $\mathscr{F}$, содержащей 2^n событий, и вероятности $P(\cdot)$, определенной для всех событий из $\mathscr{F}$.

Рассмотрим свойства классической вероятности.

1) Для любого $A \in \mathscr{F}$: $0 \leqslant P(A) \leqslant 1$ (поскольку $0 \leqslant m \leqslant n$).

2) Вероятность достоверного события $A=\Omega$ равна единице (так как для $A=\Omega$ $m=n$). Вероятность невозможного события $\varnothing$ равна нулю (так как для $A=\varnothing$ $m=0$).

3) Для несовместных событий A_1 и A_2

$$P(A_1+A_2)=P(A_1)+P(A_2). \tag{4}$$

Это простейший вариант так называемой теоремы сложения вероятностей. Для доказательства равенства (4) достаточно заметить, что если m_1 и m_2 — числа элементарных событий, благоприятствующих соответственно событиям A_1 и A_2, то в силу несовместности A_1 и A_2 ($A_1 \cap A_2 = \varnothing$) сумма m_1+m_2 является числом элементарных событий, благоприятствующих A_1+A_2. Поэтому

$$P(A+B) = \frac{m_1+m_2}{n} = \frac{m_1}{n} + \frac{m_2}{n} = P(A) + P(B).$$

4) Вероятность события $\bar{A}$, противоположного A, равна

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

Доказательство следует из замечания, что $A+\bar{A}=\Omega$, и, следовательно, согласно свойствам 2, 3 $P(A+\bar{A}) = P(A) + P(\bar{A}) = 1$.

5) Если событие A влечет B, $A \subset B$, то

$$P(B \setminus A) = P(B) - P(A)$$

и $P(B) \geqslant P(A)$. Для доказательства заметим, что $B=A+B\cap\bar{A}$, причем события A и $B\cap\bar{A}$ несовместны (так как несовместные события A и $\bar{A}$ и $B\cap\bar{A}\subset\bar{A}$). Поэтому согласно свойству 3 $P(B) = P(A) + P(B\cap\bar{A})$. Отсюда следует, что $P(B) \geqslant P(A)$, так как согласно свойству 1 $P(B\cap\bar{A}) \geqslant 0$, а также $P(B\cap\bar{A}) = P(B\setminus A) = P(B) - P(A)$.

6) Для любых событий A_1 и A_2 имеет место равенство

$$P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2). \qquad (5)$$

Действительно,

$$A_1 \cup A_2 = A_1 + A_2 \setminus A_1 = A_1 + A_2 \setminus (A_1 \cap A_2).$$

А так как $A_1 \cap A_2 \subset A_2$, то равенство (5) следует из свойств 5, 3.

Равенство (5) нетрудно распространить на случай произвольного конечного числа событий. Именно вероятность того, что произойдет хотя бы одно из событий $A_1, \ldots, A_n$, равна

$$P_{n,1} = P(A_1 \cup \ldots \cup A_n) = \sum_{i=1}^{n} P(A_i) - \qquad (6)$$

$$-\sum_{i<j}^{n} P(A_i \cap A_j) + \sum_{i<j<k}^{n} P(A_i \cap A_j \cap A_k) + \ldots$$

$$\ldots + (-1)^{n-1} P(A_1 \cap \ldots \cap A_n).$$

Действительно, если $B = A_1 \cup \ldots \cup A_{n-1}$, то искомая вероятность по только что доказанному равенству (5) равна

$$P_{n,1} = P(B) + P(A_n) - P(B \cap A_n).$$

Но согласно свойству дистрибутивности (1.5)

$$B \cap A_n = (A_1 \cap A_n) \cup (A_2 \cap A_n) \cup \ldots \cup (A_{n-1} \cap A_n)$$

и если считать, что равенство (6) верно для объединения $n-1$ событий, то

$$P(B \cap A_n) = \sum_i P(A_i \cap A_n) - \sum_{i<j} P(A_i \cap A_j \cap A_n) + \ldots$$

$$\ldots + (-1)^{n-2} P(A_1 \cap \ldots \cap A_{n-1} \cap A_n),$$

где при суммировании индексы $i, j, \ldots$ пробегают значения $1, \ldots, n-1$. Теперь нетрудно увидеть, что верно и (6). Согласно (6) вероятность того, что не произойдет ни одного события из $A_1, \ldots, A_n$, равна $Q_{n,0} = 1 - P_{n,1}$.

Приведем более общие результаты: вероятность $Q_{n,m}$ того, что осуществится ровно m событий из $A_1, \ldots, A_n$, равна

$$Q_{n,m} = S_m - C^1_{m+1} S_{m+1} + C^2_{m+2} S_{m+2} + \ldots$$

$$\ldots + (-1)^{n-m} C_n^{n-m} S_n.$$

Вероятность $P_{n,m}$ того, что осуществится не менее m событий из $A_1, \ldots, A_n$, равна

$$P_{n,m} = S_m - C^1_m S_{m+1} + C^2_{m+1} S_{m+2} + \ldots + (-1)^{n-m} C^{n-m}_{n-1} S_n.$$

В этих двух формулах, которые рекомендуются читателю в качестве упражнения,

$$S_j = \sum_{i_1<i_2<\ldots<i_j}^{n} P(A_{i_1} \cap \ldots \cap A_{i_j}), \quad i_1, \ldots, i_j = 1, \ldots, n;$$

$$j = 1, \ldots, n; \quad S_0 = 1.$$

Проиллюстрируем сказанное на примере важной задачи о совпадениях. Пусть имеется n частиц и n ячеек, причем частицы и ячейки отмечены номерами $1, \ldots, n$. Частицы случайно размещаются по ячейкам, по одной в каждую ячейку, причем все такие размещения считаются равновероятными. Назовем совпадением любое событие A_i, состоящее в том, что частица с номером i попадает в ячейку с номером i. Какова вероятность $P_{n,1}$ хотя бы одного совпадения? Какова вероятность $Q_{n,m}$ ровно m совпадений?

Событию A_i (ячейка с номером i занята частицей с номером i) благоприятствуют $(n-1)!$ перестановок $n-1$ частиц по $n-1$ свободным ячейкам. Аналогично событию $A_i \cap A_j$ (ячейки с номерами i и j заняты соответственно частицами с номерами i и j) благоприятствует $(n-2)!$ перестановок и т. д. Поскольку всего возможно $n!$ размещений частиц по ячейкам, то

$$P(A_i) = (n-1)!/n!, \quad P(A_i \cap A_j) = (n-2)!/n!, \ \ldots.$$

Так как сумма S_j содержит C_n^j одинаковых слагаемых $P(A_{i_1} \cap \ldots \cap A_{i_j})$, то

$$S_j = ((n-j)!/n!)\,C_n^j = 1/j!, \quad j=1, \ldots, n.$$

Следовательно, вероятность $P_{n,1}$ хотя бы одного совпадения равна

$$P_{n,1} = 1 - 1/2! + 1/3! + \ldots + (-1)^{n-1}\, 1/n!.$$

Любопытно, что $\lim\limits_{n\to\infty} P_{n,1} = 1 - e^{-1} = 0{,}6321\ldots$ $(e=2{,}718\ldots)$, так как выражение для $P_{n,1}$ представляет сумму первых $n+1$ членов ряда для $1 - e^{-1} = 1 - \frac{1}{2!} + \frac{1}{3!} - \ldots$.

Аналогично для вероятности $Q_{n,m}$ ровно m совпадений $\lim\limits_{n\to\infty} Q_{n,m} = \frac{1}{m!}e^{-1}$, так что для больших n: $Q_{n,m} \approx \frac{1}{m!}e^{-1}$ и не зависит от n.

Заключая изучение классической теоретико-вероятностной модели, приведем поучительный пример, известный как парадокс шевалье де Мере. Пусть одновременно бросаются три игральные кости. Какая комбинация более вероятна: дающая в сумме 11 очков или 12?

По мнению де Мере, эти комбинации равновероятны, поскольку 11 очков (событие A) можно получить шестью способами: $4+4+3$, $5+3+3$, $5+4+2$, $5+5+1$, $6+3+2$, $6+4+1$, и столькими же способами можно получить 12 очков (событие B): $4+4+4$, $5+4+3$, $5+5+2$, $6+3+3$, $6+4+2$, $6+5+1$. Де Мере полагал, что поскольку число способов, позволяющих получить событие A и B, одно и то же, то равны и вероятности $P(A)$ и $P(B)$. Однако в результате многочисленных наблюдений за игрой в кости шевалье отметил, что комбинация, дающая в сумме 12 очков, выпадает реже, чем дающая 11. Он обратился за разъяснениями к знаменитому Паскалю, который указал, что рассматриваемые де Мере «способы» не равновероятны, поскольку кроме выпадающих очков следует учитывать, на каких именно костях они выпали.

Действительно, занумеруем кости и будем выписывать значения выпадающих очков в соответствии с нумерацией костей. Тогда комбинация $6+4+1$ реализуется в шести случаях (641, 614, 461, 416, 164, 146), комбинация $5+3+3$ — в трех (533, 353, 335), а комбинация $4+4+4$ — лишь в одном (444). Поскольку в данном случае, очевидно, равновероятны все $6\times6\times6=216$ исходов xyz $(x=1, \ldots, 6,\ y=1, \ldots, 6,\ z=1, \ldots, 6)$, то понятно, что «способы» де Мере не равновероятны. На самом деле 11 очкам благоприятствуют 27 исходов, а 12 очкам — 25, так что $P(A)=27/216$, $P(B)=25/216$.

Классическая теоретико-вероятностная модель была развита в течение XVII — XIX вв. на пути естественной формализации некоторых из тех интуитивных представлений о вероятности, которые обсуждались во введении. Основателями математической теории вероятностей считаются Пьер Ферма (1601—1665) и Блез Паскаль (1623—1662). Размышляя о математических проблемах, возникающих в связи с азартными играми, в 1654 г. они установили некоторые из основных положений теории вероятностей. Ознакомившись с результатами Ферма и Паскаля, в разработке проблем теории вероятностей принимает участие Христиан Гюйгенс (1629—1695) и в 1657 г. издает первый трактат по теории вероятностей «О расчетах при азартных играх». В это время Гюйгенс уже полностью отдает себе отчет в том, что на самом деле речь идет не об играх, а о глубокой математической теории. Следующий крупный шаг был сделан Якобом Бернулли (1654—1705). Его посмертный труд «Искусство предположения» содержит много новых результатов. Наконец, «Учение о случае» Авраама де Муавра (1667—1754) и фундаментальный труд «Аналитическая теория вероятностей» Пьера Симона Лапласа (1749—1827) придают этой науке в известном смысле законченный вид.

Однако после Лапласа интерес к теории вероятностей значительно упал, и в продолжение первых десятилетий XIX в. ее даже перестали относить к математическим дисциплинам. Одна из главных причин этого в том, что теория вероятностей, построенная на неудовлетворительных основаниях, изобиловала парадоксами и противоречиями. В частности, лапласовское определение вероятности события A как отношения $n(A)/n$, где n — общее число равновероятных исходов, а $n(A)$ — число исходов, влекущих A, исходило из порочного круга понятий, поскольку использовало понятие равновероятности. Кроме того, оставался широкий круг случайных явлений, которые не удавалось понять в рамках классической модели. Это относится и к задачам на геометрическую вероятность, рассмотренным во введении.

Положение самостоятельной математической дисциплины теория вероятностей достигает лишь в трудах выдающегося русского математика середины XIX в. Пафнутия Львовича Чебышева (1821—1894) и его учеников, выдающихся ученых А. А. Маркова (1856—1922) и А. М. Ляпунова (1857—1918). А в результате последующих фундаментальных исследований советских математиков А. Я. Хинчина, А. Н. Колмогорова, Е. Е. Слуцкого и С. Н. Бернштейна теория вероятностей, по существу, приобрела тот вид, какой она имеет на сегодняшний день. В частности, аксиоматика теории вероятностей, построенная академиком Колмогоровым, в настоящее время считается общепринятой.

§ 3. АКСИОМАТИЧЕСКОЕ ПОСТРОЕНИЕ ТЕОРИИ ВЕРОЯТНОСТЕЙ

Пусть Ω — пространство элементарных событий, $\mathscr{F}$ — алгебра событий (подмножеств Ω). Следующие пять условий образуют систему аксиом теории вероятностей.

1. $\mathscr{F}$ является σ-алгеброй событий.

Алгебра событий $\mathscr{F}$ называется σ-алгеброй, если для всякой последовательности событий $A_j \in \mathscr{F}$, $j=1, 2, \ldots$, их объединение $A = A_1 \cup A_2 \cup \ldots = \bigcup_1^\infty A_j$ также принадлежит $\mathscr{F}$, т. е. является событием. Согласно принципу двойственности отсюда следует, что и $B = \bigcap_1^\infty A_j \in \mathscr{F}$. Действительно, $\overline{B} = \overline{A}_1 \cup \overline{A}_2 \cup \ldots = \overline{A_1 \cap A_2 \cap \ldots}$

Подчеркнем, что речь идет лишь о счетных объединениях и пересечениях. Если A_α, $\alpha \in S$, произвольная система событий, то, например, их объединение $\bigcup_{\alpha\in S} A_\alpha$ может и не быть событием.

2. На σ-алгебре $\mathscr{F}$ определяется функция $P(\cdot)$, принимающая числовые значения $P(A) \geqslant 0$, $A \in \mathscr{F}$, называемая вероятностью и обладающая следующими свойствами.

3. Для всяких двух событий A и B, таких что $A \cap B = \varnothing$, $P(A+B) = P(A) + P(B)$ (аксиома сложения вероятностей).

Отсюда следует, что для произвольного конечного числа несовместных событий $A_1, \ldots, A_n$

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

4. Пусть события A_j, $j=1, 2, \ldots$, попарно несовместны: $A_i \cap A_j = \varnothing$, $i \neq j$, $i, j = 1, 2$ и $A = A_1 + A_2 + \ldots$. Тогда

$$P(A) = P(A_1) + P(A_2) + \ldots = \sum_{i=1}^{\infty} P(A_i) \tag{1}$$

(заметим, что согласно аксиоме 1 $A \in \mathscr{F}$).

Эта аксиома определяет счетную аддитивность вероятности. Может быть, более привычно она звучит как аксиома непрерывности вероятности. Для этого рассмотрим последовательность событий $B_1 = A_1$, $B_2 = A_1 + A_2, \ldots$. Событие A следует понимать как предел последовательности $\{B_j\}$, $A = \lim_{n\to\infty} B_n$. При этом равенство (1) можно понимать как свойство непрерывности вероятности:

$$P(A) = P(\lim_{n\to\infty} B_n) = \lim_{n\to\infty} P(B_n) =$$

$$= \lim_{n\to\infty} \sum_{j=1}^{n} P(A_j) = \sum_{j=1}^{\infty} P(A_j).$$

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

Пространство элементарных событий Ω, σ-алгебра событий $\mathscr{F}$ и вероятность $P(\cdot)$ на $\mathscr{F}$, удовлетворяющие аксиомам теории вероятностей, образуют так называемое вероятностное пространство, которое принято обозначать $(\Omega, \mathscr{F}, P)$.

Система аксиом теории вероятностей не противоречива, так как существуют Ω, $\mathscr{F}$ и $P(\cdot)$, удовлетворяющие этим аксиомам, и не полна, так как вероятность можно определить многими способами в рамках аксиом $2 \div 5$. В качестве примера укажем классическую теоретико-вероятностную модель, в которой Ω — конечное множество, $\mathscr{F}$ — алгебра (и σ-алгебра) всех подмножеств Ω и вероятность $P(\cdot)$ определена для каждого подмножества $A \in \mathscr{F}$ как отношение числа точек, образующих A, к числу всех точек Ω.

Понятно, что для произвольного пространства элементарных событий Ω система всех его подмножеств образует σ-алгебру. Но такая σ-алгебра может оказаться столь обширной, что на ней невозможно определить вероятность, удовлетворяющую свойству счетной аддитивности 4. Требование счетной аддитивности $P(\cdot)$ и стремление выбрать систему множеств $\mathscr{F}$ как можно более широкой взаимно ограничивают друг друга.

Приведем несколько примеров σ-алгебр. Если Ω — произвольное пространство элементарных событий, то $\mathscr{F} = \{\varnothing, \Omega\}$, а также множество $\mathscr{F}$ всех подмножеств Ω — тривиальные примеры σ-алгебр. Многочисленные примеры σ-алгебр могут быть получены на следующем пути. Пусть $\mathscr{A}$ — произвольная система множеств и $\Omega = \bigcup_{A\in\mathscr{A}} A$. Если $\mathscr{F}$ — σ-алгебра всех подмножеств Ω, то, очевидно, $\mathscr{A} \subset \mathscr{F}$. Следовательно, существуют σ-алгебры подмножеств Ω, содержащие $\mathscr{A}$. Если $\mathscr{F}_1$ и $\mathscr{F}_2$ — σ-алгебры подмножеств Ω, содержащие $\mathscr{A}$, то $\mathscr{F}_1 \cap \mathscr{F}_2$ также содержит $\mathscr{A}$ и, кроме того, является σ-алгеброй. Следовательно, пересечение $\bigcap_\alpha \mathscr{F}_\alpha = \mathscr{F}_{\mathscr{A}}$ всех σ-алгебр, содержащих $\mathscr{A}$, 1) является σ-алгеброй, 2) является минимальной σ-алгеброй в том смысле, что для всякой σ-алгебры $\mathscr{F}$, содержащей $\mathscr{A}$, непременно $\mathscr{F}_{\mathscr{A}} \subset \mathscr{F}$ (ибо $\mathscr{F}$ совпадает с одной из σ-алгебр $\mathscr{F}_\alpha$, как σ-алгебра, содержащая $\mathscr{A}$), 3) содержит все множества из $\mathscr{A}$, $\mathscr{A} \subset \mathscr{F}_{\mathscr{A}}$. В частности, если $\mathscr{A}$ — σ-алгебра, то $\mathscr{F}_{\mathscr{A}} = \mathscr{A}$, а если $\mathscr{A}$ состоит из двух множеств A и B, то $\mathscr{F}_{\mathscr{A}}$ состоит из восьми множеств: $\varnothing$, A, B, $A \cap B$, $A \cup B = \Omega$, $A \setminus B$, $B \setminus A$, $(A \setminus B) \cup (B \setminus A)$.

$\mathscr{F}_{\mathscr{A}}$ называется σ-алгеброй, порожденной системой множеств $\mathscr{A}$.

При определении σ-алгебры событий порождающая система событий $\mathscr{A}$, как правило, составляется из событий, наблюдаемых в эксперименте. Так, например, в эксперименте с бросанием точки на отрезок $[a, b]=\Omega$ естественно считать событием любой отрезок $[\alpha, \beta]$, содержащийся в $[a, b]$. В таком случае $\mathscr{A}$ — система всех отрезков, содержащихся в $[a, b]$. σ-алгебра $\mathscr{F}_{\mathscr{A}}$ порожденная $\mathscr{A}$, в данном случае называется σ-алгеброй борелевских подмножеств отрезка $[a, b]$, или борелевской алгеброй отрезка $[a, b]$.

Посмотрим, насколько обширна борелевская алгебра $[a, b]$. Отдельные точки $[a, b]$, как одноточечные подмножества, являются борелевскими множествами, так как все они имеют вид $[\alpha, \alpha]$. Любое счетное подмножество $[a, b]$, например множество рациональных точек, является борелевским, равно как и его дополнение. Любой интервал $(\alpha, \beta)\subset\subset[a, b]$ является борелевским множеством, поскольку $(\alpha, \beta)=\bigcup\limits_{n=1}^{\infty}\left[\alpha+\frac{1}{n}, \beta-\frac{1}{n}\right]$. Для доказательства этого равенства заметим, что если $x\in(\alpha, \beta)$, то есть $\alpha<x<\beta$, то найдется n, для которого $\alpha+1/n\leqslant x\leqslant\beta-1/n$. Следовательно, $x\in\bigcup\limits_{n=1}^{\infty}\left[\alpha+\frac{1}{n}, \beta-\frac{1}{n}\right]$. Наоборот, если выполнено последнее включение, то для некоторого $n: x\in[\alpha+1/n, \beta-1/n]$ и, следовательно, тем более $x\in(\alpha, \beta)$. Отсюда следует, что любое открытое подмножество $[a, b]$ — борелевское, поскольку всякое открытое множество, как известно, является объединением не более чем счетного числа интервалов. Все замкнутые подмножества $[a, b]$, как дополнения открытых, также борелевские. Наконец, борелевскими являются множества, полученные счетным объединением и (или) пересечением открытых и (или) замкнутых множеств, и т. д.

Класс борелевских множеств, как видно из приведенного перечисления, весьма обширен. Легко понять, что пример множества, не являющегося борелевским, не может быть простым. Однако такие множества есть, и, следовательно, не всякое подмножество $[a, b]$ является событием.

Для того чтобы завершить построение вероятностного пространства в эксперименте с бросанием точки на отрезок $[a, b]$, осталось определить вероятность на борелевской алгебре событий. Если для $[\alpha, \beta]\subset[a, b]$ положить $P([\alpha, \beta])=(\beta-\alpha)/l$, $l=b-a$, и для любых содержащихся в $[a, b]$ непересекающихся отрезков $[\alpha, \beta]$ и $[\gamma, \delta]$ определить $P([\alpha, \beta]+[\gamma, \delta])=((\beta-\alpha)+(\delta-\gamma))/l$, то $P([\alpha, \alpha])=0$, $P([\alpha, \beta))=P((\alpha, \beta])=P((\alpha, \beta))=(\beta-\alpha)/l$ и $P([a, b])=1$. Но это определение не содержит указаний на то, каким образом определить вероятность для произвольного борелевского множества. Этот вопрос полностью решается в теории меры. Можно показать, что на некоторой (лебеговской) σ-алгебре, содержащей борелевскую алгебру, можно определить

вероятность, значение которой для отрезка $[\alpha, \beta]\subset[a, b]$ равно $(\beta-\alpha)/l$.

Существует важный для приложений класс так называемых **дискретных вероятностных пространств**, теория которых достаточно проста. Пример такого вероятностного пространства естественно возникает в эксперименте с бросанием игральной кости. Рассмотрим игру в кости «до первого проигрыша». Условимся, что выпадение любого числа очков, кроме шести, означает выигрыш и продолжение игры, а выпадение шести очков означает проигрыш, и при этом игра заканчивается. Построим вероятностное пространство для такого эксперимента со случайным исходом. Начнем с того, что перечислим все исходы, приводящие к окончанию игры:

6;

16, 26, 36, 46, 56;

116, 126, 136, 146, 156, 226, ..., 256, 316, ..., 356, 416, ..., 456, 516, ..., 556;

1116, ...

Легко заметить, что имеется один исход, при котором игра заканчивается после первого бросания, 5 исходов (из 6^2), при которых игра заканчивается после второго бросания, и вообще 5^{n-1} исходов (из 6^n), при которых игра заканчивается после n-го бросания. В принципе игра может продолжаться неопределенно долго и, следовательно, не является экспериментом с конечным числом исходов, если под исходом понимать выпадение некоторого числа очков при очередном бросании.

Введем пространство элементарных событий Ω, в котором точка ω_1 обозначает исход «6», точка ω_2 обозначает совокупность исходов «16, 26, ..., 56», ω_3 соответственно «116, ..., 556» и т. д. Согласно классической модели припишем событию $\{\omega_1\}$ вероятность 1/6, событию $\{\omega_2\}$ — вероятность $5/(6^2)$, событию $\{\omega_3\}$ — вероятность $(5^2)/(6^3)$, ..., событию $\{\omega_n\}$ — вероятность $(5^{n-1})/(6^n)$. Обозначим через ω_∞ множество всех бесконечных последовательностей (очков), не содержащих шестерки, и припишем событию $\{\omega_\infty\}$ вероятность, равную нулю. Событие $\{\omega_\infty\}$ называется невероятным, но в отличие от $\varnothing$ не является невозможным.

Поскольку

$$\Omega=\{\omega_1\}+\{\omega_2\}+\ldots+\{\omega_n\}+\ldots,$$

то согласно аксиоме 4 о счетной аддитивности

$$P(\Omega)=\sum_{j=1}^{\infty}P(\{\omega_j\})=\frac{1}{6}\sum_{n=0}^{\infty}\left(\frac{5}{6}\right)^n=1,$$

причем значение $P(\Omega)$ не зависит от порядка суммирования (и порядка перечисления точек Ω), так как ряд справа сходится абсолютно.

Объявим событием любое подмножество Ω и для любого события A зададим вероятность

$$P(A) = \sum_{j:\,\omega_j \in A} P(\{\omega_j\}). \qquad (2)$$

Понятно, что это определение корректно, так как и в этом случае ряд справа сходится абсолютно, и значение $P(A)$ не зависит от порядка суммирования.

Подсчитаем, например, вероятность того, что игра закончится на четном бросании. Речь идет о вероятности события $A=\{\omega_2\}+\{\omega_4\}+\dots$, и согласно (2)

$$P(A) = \sum_{j=1}^{\infty} P(\{\omega_{2j}\}) = \frac{5}{36}\sum_{n=0}^{\infty}\left(\frac{5}{6}\right)^{2n} = \frac{5}{11}.$$

Оказывается, что вероятность (2) обладает свойством счетной аддитивности (и аддитивности) на σ-алгебре всех подмножеств Ω. Этот факт является следствием леммы о суммировании по блокам для абсолютно сходящихся рядов.

Лемма (о суммировании по блокам). Пусть $I=\{1, 2, \dots\}$ — множество всех чисел натурального ряда, I_α, $\alpha=1, 2, \dots$, — непересекающиеся подмножества I, такие, что $I=I_1+I_2+\dots$ (этих подмножеств может быть и конечное число). Пусть числовой ряд $c_1+c_2+\dots+c_n+\dots$ сходится абсолютно и его сумма равна S, $S=\sum_{k=1}^{\infty} c_k$. Если $S_\alpha = \sum_{k\in I_\alpha} c_k$, $\alpha=1, 2, \dots$, то числовой ряд $S_1+S_2+\dots+S_n+\dots$ сходится абсолютно и его сумма также равна S, $S=\sum_{\alpha=1}^{\infty} S_\alpha$.

Доказательство. Понятно, что все S_α, $\alpha=1, 2, \dots$, ограничены:

$$|S_\alpha| \leqslant \sum_{k\in I_\alpha} |c_k| \leqslant \sum_{k=1}^{\infty} |c_k| < \infty.$$

Покажем, что ряд $S_1+S_2+\dots$ сходится абсолютно и его сумма равна S. Абсолютная сходимость следует из оценки

$$\sum_{\alpha=1}^{n} |S_\alpha| \leqslant \sum_{\alpha=1}^{n}\sum_{k\in I_\alpha} |c_k| \leqslant \sum_{k=1}^{\infty} |c_k|,$$

верной для любого $n\geqslant 1$. Покажем теперь, что его сумма равна S. Зафиксируем любое $\varepsilon>0$. Следует показать, что

найдется такой номер $N_0=N_0(\varepsilon)$, что при $n\geqslant N_0$

$$\left|\sum_{\alpha=1}^{n} S_\alpha - S\right| < \varepsilon. \qquad (3)$$

Так как по условию ряд $\sum_{k=1}^{\infty} c_k$ сходится абсолютно, то для данного $\varepsilon>0$ найдется такой номер $N_1=N_1(\varepsilon)$, что при $n\geqslant N_1$

$$\left|\sum_{k=1}^{n} c_k - S\right| < \varepsilon/2, \quad \sum_{k=n+1}^{\infty} |c_k| < \varepsilon/2. \qquad (4)$$

Выберем N_2 столь большим, чтобы в сумму

$$\sum_{\alpha=1}^{N_2} S_\alpha = \sum_{\alpha=1}^{N_2}\sum_{k\in I_\alpha} c_k$$

вошли все члены c_k с номерами $k\leqslant N_1$ и положим $N_0=\max(N_1, N_2)$. Тогда при $n\geqslant N_0$, используя (4), найдем

$$\left|\sum_{\alpha=1}^{n} S_\alpha - S\right| \leqslant \left|\sum_{k=1}^{n} c_k - S\right| + \left|\sum_{\alpha=1}^{n} S_\alpha - \sum_{k=1}^{n} c_k\right| \leqslant$$

$$\leqslant \left|\sum_{k=1}^{n} c_k - S\right| + \sum_{k=N_1+1}^{\infty} |c_k| < \varepsilon/2+\varepsilon/2=\varepsilon.$$

Неравенство (3) доказано. ▲

Замечания. 1. Требование абсолютной сходимости ряда $c_1+c_2+\dots$ существенно. Например, ряд $1-1/2+1/3-1/4+\dots$ сходится (условно), но $S_1=1+1/3+1/5+\dots$ расходится.

2. Обратная теорема неверна. Если для некоторой системы подмножеств $\{I_\alpha\}$ $\sum_{\alpha=1}^{\infty} |S_\alpha| < \infty$, то отсюда не следует абсолютная сходимость ряда $c_1+c_2+\dots$. Действительно, если $S_1=1-1/2$, $S_2=1/3-1/4$, …, $S_\alpha=\frac{1}{2\alpha-1}-\frac{1}{2\alpha}=\frac{1}{2\alpha(2\alpha-1)}$, то $\sum_{\alpha=1}^{\infty} S_\alpha = \sum_{\alpha=1}^{\infty} |S_\alpha| < \infty$, но ряд $1+1/2+1/3+\dots$ расходится.

3. Если $c_k\geqslant 0$ (или $c_k\leqslant 0$), $k=1, 2, \dots$, то теорема верна в обе стороны. Действительно, пусть для некоторой системы подмножеств

$$I_\alpha,\ \alpha=1, 2, \dots, I=I_1+I_2+\dots, \quad \sum_{\alpha=1}^{\infty} S_\alpha = \sum_{\alpha=1}^{\infty} |S_\alpha| = S.$$

Тогда

$$\sum_{k=1}^{n} c_k \leqslant \sum_{\alpha=1}^{N(n)} S_\alpha \leqslant \sum_{\alpha=1}^{\infty} S_\alpha < \infty,$$

где $N(n)$ выбрано так, чтобы в сумму $S_1+...+S_{N(n)}$ вошли все c_k, $k=1$, ..., n. Отсюда следует, что последовательность частичных сумм $\sum_{k=1}^{n} c_k$, $k=1, 2, \ldots,$ монотонна и ограничена. Поэтому ряд $c_1+c_2+...$ сходится и его сумма совпадает с суммой ряда $S_1+S_2+...$ (в силу прямой теоремы).

Рассмотрим теперь общий случай дискретных вероятностных пространств.

Определение. Вероятностное пространство $(\Omega, \mathscr{F}, P)$ называется **дискретным**, если $\Omega=\{\omega_1, \omega_2, ...\}$ конечно или счетно, $\mathscr{F}$ — σ-алгебра всех подмножеств Ω (включая пустое множество $\varnothing$), вероятность $P(\cdot)$ определена для каждого одноточечного подмножества Ω:

$$P(\{\omega_j\}) = p_j \geqslant 0, \ j = 1, 2, \ldots, \sum_{j=1}^{\infty} p_j = 1. \qquad (5)$$

При этом вероятность любого события $A \in \mathscr{F}$ определяется равенством

$$P(A) = \sum_{j:\, \omega_j \in A} p_j. \qquad (6)$$

Проверим, что для дискретного вероятностного пространства выполнены аксиомы теории вероятностей. Поскольку аксиомы 1, 2 и 5, очевидно, выполнены, достаточно проверить лишь аксиомы 3 и 4. Но обе эти аксиомы следуют из леммы о суммировании по блокам.

Остановимся на свойствах вероятности, которые вытекают из аксиом. Так же как при доказательстве свойств классической вероятности, найдем, что:

1) $P(\bar{A})=1-P(A)$, так как $\bar{A}+A=\Omega$.

2) Если $A\subset B$, то $P(B\diagdown A)=P(B)-P(A)$, так как $B=A+B\diagdown A$. Следовательно, включение $A\subset B$ влечет неравенство $P(A)\leqslant P(B)$ (монотонность вероятности).

3) Для любых событий A_1, ..., A_n имеет место равенство (2.6), причем приведенное там доказательство верно и в общем случае. Сохраняют силу и выражения для вероятности $Q_{n,m}$ того, что осуществится ровно m событий из A_1, ..., A_n, а также для вероятности $P_{n,m}$ того, что осуществится не менее m событий из A_1, ..., A_n.

4) Пусть $A_1\subset A_2\subset...\subset A_n\subset...$ — последовательность событий, каждое из которых влечет все последующие. Если

$A=\bigcup_1^\infty A_j$ — событие, состоящее в том, что происходит хотя бы одно из событий A_j, $j=1, 2, ...$, то

$$P(A) = \lim_{n\to\infty} P(A_n).$$

Действительно, положим $A_0=\varnothing$. Тогда

$$A = \bigcup_1^\infty A_j = (A_1\diagdown A_0) + (A_2\diagdown A_1) + \ldots + (A_n\diagdown A_{n-1}) + \ldots$$

и

$$P(A) = \sum_{j=1}^{\infty} P(A_j\diagdown A_{j-1}) = \lim_{n\to\infty} \sum_{j=1}^{n} P(A_j\diagdown A_{j-1}) =$$

$$= \lim_{n\to\infty} \sum_{j=1}^{n} (P(A_j) - P(A_{j-1})) = \lim_{n\to\infty} P(A_n).$$

5) Если $A_1\supset A_2\supset...\supset A_n\supset...$ — последовательность событий, каждое из которых влечет все предыдущие, и $A = \bigcap_{j=1}^{\infty} A_j$ событие, состоящее в том, что происходят все события A_1, ..., A_n, ..., то

$$P(A) = \lim_{n\to\infty} P(A_n).$$

Это утверждение может быть доказано с помощью принципа двойственности. Действительно, для противоположных событий: $\bar{A}_1\subset\bar{A}_2\subset...\subset\bar{A}_n\subset...$ и

$\bar{A} = \bigcup_1^\infty \bar{A}_j$. Поэтому $P(\bar{A}) = \lim_{n\to\infty} P(\bar{A}_n)$, и, следовательно,

$$P(A) = 1 - P(\bar{A}) = \lim_{n\to\infty} (1 - P(\bar{A}_n)) = \lim_{n\to\infty} P(A_n).$$

Свойства 4 и 5 можно понимать как свойства непрерывности вероятности относительно монотонных предельных переходов. Действительно, если

$$A_1 \subset A_2 \subset \ldots \subset A_n \subset \ldots, \text{ то } \bigcup_{j=1}^{n} A_j = A_n,$$

и множество $A = \bigcup_{j=1}^{\infty} A_j$ естественно назвать пределом монотонной последовательности множеств $A_1\subset A_2\subset\ldots\subset A_n\subset\ldots$: $A = \lim_{j\to\infty} A_j$. Тогда согласно свойству 4:

$$P(A) = P(\lim_{j\to\infty} A_j) = \lim_{j\to\infty} P(A_j).$$

Точно так же, если $A_1 \supset A_2 \supset \ldots \supset A_n \supset \ldots$, то $A_n = \bigcap_{j=1}^{n} A_j$, и множество $A = \bigcap_{j=1}^{\infty} A_j$ называется пределом монотонной последовательности множеств $A_1 \supset A_2 \supset \ldots \supset A_n \supset \ldots$: $A = \lim_{j\to\infty} A_j$. В данном случае свойство 5 означает, что

$$P(A) = P(\lim_{j\to\infty} A_j) = \lim_{j\to\infty} P(A_j).$$

§ 4. УСЛОВНАЯ ВЕРОЯТНОСТЬ. НЕЗАВИСИМОСТЬ

Пусть задано вероятностное пространство $(\Omega, \mathcal{F}, P)$. Рассмотрим задачу определения вероятности события A, если известно, что произошло событие B, причем $P(B)>0$. При этих условиях пространством элементарных событий естественно считать не Ω, а B, поскольку тот факт, что B произошло, означает, что речь идет лишь о тех элементарных событиях ω, которые принадлежат множеству B. В общем случае событие $A\cap B$ влечет A, но, если известно, что событие B произошло, то при этом условии A влекут те и только те элементарные события, которые принадлежат $A\cap B$. Поскольку ранее мы условились отождествлять событие A и множество элементарных событий, влекущих A, то теперь событие A следует отождествить с множеством $A_B = A\cap B$. Можно сказать, что множество $A_B = A\cap B$ есть событие A, рассматриваемое с точки зрения, согласно которой пространством элементарных событий объявлено событие B.

На новом пространстве элементарных событий B σ-алгебра событий $\mathcal{F}_B$ определяется, или, как говорят, индуцируется, σ-алгеброй событий $\mathcal{F}$. Именно $\mathcal{F}_B$ состоит из событий вида $A_B = A\cap B$, где $A\in\mathcal{F}$. Проверим, что $\mathcal{F}_B$ действительно σ-алгебра. Пусть A_B, C_B, $C_B^j \in \mathcal{F}_B (A,\ C,\ C^j \in \mathcal{F})$, $j=1, 2, \ldots$. Тогда, используя свойства дистрибутивности операций объединения $\cup$ и пересечения $\cap$, найдем

$$A_B \cup C_B = (A\cap B)\cup(C\cap B) = (A\cup C)\cap B = (A\cup C)_B,$$

$$\bigcup_{j=1}^{\infty} C_B^j = \bigcup_{j=1}^{\infty} (C^j\cap B) = \Big(\bigcup_{j=1}^{\infty} C^j\Big)\cap B = \Big(\bigcup_{j=1}^{\infty} C^j\Big)_B,$$

$$A_B\cap C_B = (A\cap C)_B.$$

Следовательно, $A_B\cup C_B$, $\bigcup_{j=1}^{\infty} C_B^j$, $A_B\cap C_B \in \mathcal{F}_B$. Наконец, по определению $\bar{A}_B = B\setminus A_B = B\setminus A = B\cap\bar{A} = (\bar{A})_B$, поэтому $\bar{A}_B \in \mathcal{F}_B$, если $A_B\in\mathcal{F}_B$.

Введем на σ-алгебре $\mathcal{F}_B$ вероятность $P_B(\cdot)$:

$$P_B(A_B) = P(A\cap B)/P(B).$$

Без труда можно проверить, что $P_B(\cdot)$, так же как и $P(\cdot)$, удовлетворяет аксиомам 2—5 теории вероятностей.

Наглядный смысл вероятности $P_B(\cdot)$ можно пояснить с помощью классической вероятности. В этом случае $\Omega = \{\omega_1, \ldots, \omega_n\}$, причем события $\{\omega_1\}, \ldots, \{\omega_n\}$ равновероятны. Пусть $B = \{\omega_{j_1}, \ldots, \omega_{j_k}\}$, так что $P(B) = k/n$, и $A\cap B = \{\omega_{i_1}, \ldots, \omega_{i_s}\} \subset \{\omega_{j_1}, \ldots, \omega_{j_k}\}$, так что $P(A\cap B) = s/n$. Если B рассматривать как новое пространство элементарных событий, то вероятность события A_B определяется как отношение числа s исходов, благоприятствующих A_B, к общему числу исходов k. Теперь общее число исходов — число исходов, приводящих к событию B. Таким образом, в согласии с классической схемой

$$P_B(A_B) = \frac{s}{k} = \frac{s/n}{k/n} = \frac{P(A\cap B)}{P(B)}.$$

Тройка $(B, \mathcal{F}_B, P_B)$ является новым вероятностным пространством, построенным в связи с поставленной задачей. Но вероятность $P_B(\cdot)$ можно рассматривать и на σ-алгебре $\mathcal{F}$. На $\mathcal{F}$ $P_B(\cdot)$ также является вероятностью и обозначается $P(\cdot\,|\,B)$,

$$P(A\,|\,B) = \frac{P(A\cap B)}{P(B)},\quad A\in\mathcal{F}. \tag{1}$$

$P(\cdot\,|\,B)$ как функция на $\mathcal{F}$ называется условной вероятностью события A при условии, что событие B произошло.

Что касается свойств вероятности $P(\cdot\,|\,B)$, то они вполне аналогичны соответствующим свойствам вероятности $P(\cdot)$:

$P(\Omega\,|\,B) = 1$,

$P(\bar{A}\,|\,B) = 1 - P(A\,|\,B)$,

$P(A_1 + A_2\,|\,B) = P(A_1\,|\,B) + P(A_2\,|\,B)$,

так как $(A_1+A_2)\cap B = A_1\cap B + A_2\cap B$. Если воспользоваться равенством $P(\tilde{A}\cup\tilde{C}) = P(\tilde{A}) + P(\tilde{C}) - P(\tilde{A}\cap\tilde{C})$, то, полагая $\tilde{A} = A\cap B$, $\tilde{C} = C\cap B$, нетрудно проверить, что

$$P(A\cup C\,|\,B) = \frac{1}{P(B)} P((A\cup C)\cap B) =$$

$$= \frac{1}{P(B)} P((A\cap B)\cup(C\cap B)) = P(A\,|\,B) + P(C\,|\,B) - P(A\cap C\,|\,B).$$

Наконец, очевидно,

$$P(A_1 + A_2 + \ldots\,|\,B) = \sum_{j=1}^{\infty} P(A_j\,|\,B)$$

(счетная аддитивность условной вероятности).

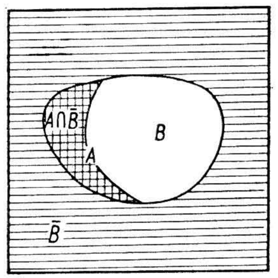
Для иллюстрации техники вычисления условных вероятностей рассмотрим пример. Известно, что вероятность аварии при запуске ракеты равна 0,1, в том числе вероятность аварии на старте — 0,09. Какова вероятность аварии в случае успешного старта? Обозначим через A событие «авария при запуске», через B — событие «авария на старте». Требуется вычислить вероятность $P(A\,|\,\bar{B})$. По смыслу ясно, что $B\subset A$. Поэтому $\bar{B}\supset\bar{A}$. Следовательно, $\bar{A}\cap\bar{B}=\bar{A}$. Это равенство позволяет подсчитать вероятность $P(\bar{A}\,|\,\bar{B})=P(\bar{A}\cap\bar{B})/P(\bar{B})=P(\bar{A})/P(\bar{B})$. Поскольку $P(A\,|\,\bar{B})+P(\bar{A}\,|\,\bar{B})=1$, то искомая условная вероятность равна $P(A\,|\,\bar{B})=1-\dfrac{1-0,1}{0,1-0,09}=\dfrac{1}{91}$. На рис. 7 приведено «графическое» решение задачи.

Из определения (1) следует так называемая **теорема умножения вероятностей**

$$P(A\cap B)=P(A\,|\,B)P(B),\quad P(B)>0. \qquad (2)$$

Вполне аналогично можно записать

$$P(A\cap B\cap C)=P(A\,|\,B\cap C)\,P(B\cap C)=$$
$$=P(A\,|\,B\cap C)\,P(B\,|\,C)\,P(C),\ \ P(B\cap C)>0.$$



$B\subset A$

Рис. 7. $P(A)=0{,}1$, $P(B)=0{,}09$, $P(A\,|\,B)=P(A\cap\bar{B})/P(\bar{B})=(0{,}1-0{,}09)/(1{,}0-0{,}09)$

С точки зрения математической этот результат тривиален. Но его роль на самом деле не математическая. Эта теорема применяется при определении вероятностей в тех случаях, когда по смыслу задачи легко определяются условные вероятности. Поясним это следующим примером. Имеются 30 экзаменационных билетов, среди которых 5 «счастливых». Кому выгоднее тянуть билет — первому или второму?

Обозначим через A событие «первый вытягивает «счастливый» билет», B — «второй вытягивает «счастливый» билет». Тогда $\bar{A}$ означает, что первый не вытянет «счастливый» билет, а $\bar{B}$ — что второй. Заметим, что $A+\bar{A}=\Omega$, поэтому $B=B\cap\Omega=B\cap A+B\cap\bar{A}$. Следовательно,

$$P(B)=P(B\cap A)+P(B\cap\bar{A})=$$
$$=P(B\,|\,A)P(A)+P(B\,|\,\bar{A})P(\bar{A}).$$

Что касается вероятностей, которые входят в эту формулу, то

$$P(A)=\frac{5}{30}=\frac{1}{6},\ \ P(\bar{A})=\frac{25}{30},\ \ P(B\,|\,A)=\frac{4}{29},$$

$$P(B\,|\,\bar{A})=\frac{5}{29}$$

и, следовательно, $P(B)=\frac{1}{6}\cdot\frac{4}{29}+\frac{5}{6}\cdot\frac{5}{29}$. Таким образом, $P(B)=1/6=P(A)$. Читателю рекомендуется выяснить, может быть третьему больше повезет?

Одним из важнейших понятий теории вероятностей является понятие независимости.

Определение. События A и B называются **независимыми**, если

$$P(A\cap B)=P(A)P(B). \qquad (3)$$

Если $P(B)>0$, то согласно (2) и (3) $P(A\,|\,B)=P(A)$, как и должно быть по смыслу условной вероятности. Аналогично, если $P(A)>0$, то $P(B\,|\,A)=P(B)$. Однако определение независимости A и B на основе равенств $P(A\,|\,B)=P(A)$, $P(B\,|\,A)=P(B)$ не эквивалентно (3), так как в (3) не предполагается существование условных вероятностей.

Из определения независимости непосредственно следует:

1) Ω и любое событие A независимы.

2) Любое событие A и событие B независимы, если $P(B)=0$. В самом деле, поскольку $A\cap B\subset B$, то $0\leqslant P(A\cap B)\leqslant P(B)=0$. Следовательно, $P(A\cap B)=0=P(A)P(B)$.

3) Если A и B независимы, то независимы также следующие пары событий: $\bar{A}$ и B, A и $\bar{B}$, $\bar{A}$ и $\bar{B}$.

Докажем, например, независимость $\bar{A}$ и B. Так как $A+\bar{A}=\Omega$, то $A\cap B+\bar{A}\cap B=B$ и, следовательно, $P(B)=P(A)P(B)+P(\bar{A}\cap B)$. Поэтому $P(\bar{A}\cap B)=P(B)\,(1-P(A))=P(B)P(\bar{A})$. Остальные утверждения доказываются аналогично.

Понятие независимости кажется несколько искусственным, так как достаточно слегка изменить одно из событий и независимость будет нарушена. Поэтому на примере дискретных вероятностных пространств покажем, какая конструкция, как правило, определяет природу независимости. Пусть

$$\Omega_1=\{\omega_1^1,\ \ldots,\ \omega_n^1,\ \ldots\},\ \ \Omega_2=\{\omega_1^2,\ \ldots,\ \omega_k^2,\ \ldots\}.$$

Как мы знаем, в вероятностном пространстве $(\Omega_1,\ \mathcal{F}_1,\ P_1)$ событиями являются все подмножества Ω, и вероятность определена для каждого элементарного события

$$P_1(\{\omega_j^1\}),\quad j=1,\ \ldots,\ n,\ \ldots\,.$$

Аналогично устроено $(\Omega_2,\ \mathcal{F}_2,\ P_2)$. Определим новое вероятностное пространство $(\Omega,\ \mathcal{F},\ P)$, в котором Ω состоит из всевозможных упорядоченных пар $\omega_{ij}=(\omega_i^1\,\omega_j^2)$:

$$\Omega=\{\omega_{ij}=(\omega_i^1\,\omega_j^2),\ \omega_i^1\in\Omega_1,\ \omega_j^2\in\Omega_2\}.$$

Так определенное Ω называется произведением Ω_1 и Ω_2 и обозначается $\Omega_1 \times \Omega_2$. Событиями назовем все подмножества Ω и соответствующую σ-алгебру обозначим $\mathcal{F} = \mathcal{F}_1 \times \mathcal{F}_2$. Наконец, вероятность на $\mathcal{F}$ определим с помощью равенства $P(\{\omega_{ij}\}) = P_1(\{\omega_i^1\}) P_2(\{\omega_j^2\})$, так что для всякого события $A \in \mathcal{F}$

$$P(A) = \sum_{\omega_{ij} \in A} P(\{\omega_{ij}\}). \tag{4}$$

Построенное вероятностное пространство называется **произведением пространств** $(\Omega_1, \mathcal{F}_1, P_1)$ и $(\Omega_2, \mathcal{F}_2, P_2)$ и обозначается $(\Omega_1, \mathcal{F}_1, P_1) \times (\Omega_2, \mathcal{F}_2, P_2)$.

Рассмотрим событие $A \in \mathcal{F}$, которое состоит из тех пар (ω_i^1, ω_j^2), в которых $\omega_i^1 \in A \in \mathcal{F}_1$, ω_j^2 произвольно, $\omega_j^2 \in \Omega_2$. Такое событие назовем **цилиндрическим** и обозначим $A = A_1 \times \Omega_2$. Согласно определению (4)

$$P(A) = \sum_{\omega_i^1 \in A_1,\ \omega_j^2 \in \Omega_2} P_1(\{\omega_i^1\}) P_2(\{\omega_j^2\}) =$$

$$= \sum_{\omega_i^1 \in A_1} P_1(\{\omega_i^1\}) \sum_{\omega_j^2 \in \Omega_2} P_2(\{\omega_j^2\}) = P_1(A_1).$$

Аналогично для $B = \Omega_1 \times A_2$ найдем $P(B) = P_2(A_2)$. Событие $A \cap B$ состоит из тех пар $(\omega_i^1 \omega_j^2)$, для которых $\omega_i^1 \in A_1$, $\omega_j^2 \in A_2$. Поэтому

$$P(A \cap B) = \sum_{\omega_i^1 \in A_1,\ \omega_j^2 \in A_2} P_1(\{\omega_i^1\}) P_2(\{\omega_j^2\}) =$$

$$= \sum_{\omega_i^1 \in A_1} P_1(\{\omega_i^1\}) \sum_{\omega_j^2 \in A_2} P_2(\{\omega_j^2\}) = P_1(A_1) P_2(A_2).$$

Следовательно, $P(A \cap B) = P(A) P(B)$, т. е. события A и B независимы.

Вероятностное пространство $(\Omega_1, \mathcal{F}_1, P_1)$ можно считать отвечающим одному вероятностному эксперименту, $(\Omega_2, \mathcal{F}_2, P_2)$ — другому, никак не связанному с первым. Тогда вероятностное пространство, отвечающее двум экспериментам, определяется как произведение $(\Omega, \mathcal{F}, P) = (\Omega_1, \mathcal{F}_1, P_1) \times (\Omega_2, \mathcal{F}_2, P_2)$. Цилиндрические подмножества Ω определяют события в одном из экспериментов при произвольном исходе другого. Пусть, например, бросаются две кости. Вероятность каждого из 36 возможных исходов определим равной 1/36. Тогда события «на первой кости выпало четное число очков, на второй — что угодно» и «на первой кости выпало что угодно, на второй — три очка» независимы. Этот результат, конечно, является следствием того, что вероятности совместных исходов определены как произведения вероятностей.

Однако в некоторых ситуациях, анализируя условия эксперимента, бывает трудно заключить, независимы те или иные события или нет. В таких случаях следует действовать согласно определению независимости. Пусть, например, из колоды, содержащей 52 карты, вытягиваются карты. Рассмотрим события: A_1 — дама, A_2 — карта пиковой масти. Тогда

$$P(A_1) = 4/52 = 1/13, \quad P(A_2) = 13/52 = 1/4,$$

$$P(A_1) P(A_2) = 1/52.$$

С другой стороны, вероятность вытянуть даму пиковой масти равна $P(A_1 \cap A_2) = 1/52$. Следовательно, события A_1 и A_2 независимы.

Определение. События $A_1, \ldots, A_n$ называются **независимыми в совокупности**, если при любом выборе различных событий $A_{i_1}, \ldots, A_{i_k}$ из данной совокупности выполняется равенство

$$P(A_{i_1} \cap A_{i_2} \cap \ldots \cap A_{i_k}) = P(A_{i_1}) P(A_{i_2}) \ldots P(A_{i_k}). \tag{5}$$

Разумеется, из независимости событий в совокупности следует их попарная независимость. Но обратное утверждение неверно. Проиллюстрируем этот факт на примере, принадлежащем Бернштейну. Пусть три грани правильного тетраэдра окрашены соответственно в красный (К), зеленый (З) и синий (С) цвета, а четвертая — в три цвета (КЗС). Ясно, что вероятность упасть тетраэдру гранью, на которой есть, скажем, красный цвет, равна $P(\text{К}) = 1/2$. Условная вероятность оказаться на этой грани красному цвету при условии, что на ней уже есть зеленый, равна $P(\text{К}|\text{З}) = 1/2$. Аналогично $P(\text{К}) = P(\text{З}) = P(\text{С}) = P(\text{К}|\text{З}) = \ldots = P(\text{С}|\text{З}) = 1/2$. Следовательно, события К, З и С попарно независимы. Однако вероятность упасть гранью, на которой есть все три цвета, равна $P(\text{К} \cap \text{З} \cap \text{С}) = 1/4 \neq 1/8$. Отсюда следует, что события К, З и С не являются независимыми в совокупности.

Пусть $(\Omega, \mathcal{F}, P)$ — вероятностное пространство, $A_1, \ldots, A_n$ — полная группа попарно несовместных событий, т. е.

$$A_i \cap A_j = \varnothing, \quad i \neq j, \quad A_1 + \ldots + A_n = \Omega.$$

Если $B \in \mathcal{F}$ — произвольное событие, то из последнего равенства следует

$$B = B \cap \Omega = B \cap A_1 + \ldots + B \cap A_n. \tag{6}$$

Поэтому

$$P(B)=\sum_{j=1}^{n} P(B\cap A_j), \tag{7}$$

или в терминах условных вероятностей (в силу равенства (2))

$$P(B)=\sum_{j=1}^{n} P(B\mid A_j)\,P(A_j), \tag{8}$$

если $P(A_j)>0$, $j=1, ..., n$. Равенство (8) называется **формулой полной вероятности.** Если полная группа попарно несовместных событий состоит из счетного множества событий $A_1, ..., A_n, ...$, то

$$B=B\cap A_1+...+B\cap A_n+...,$$

и в силу счетной аддитивности вероятности отсюда следует **формула полной вероятности для счетного множества событий**

$$P(B)=\sum_{j=1}^{\infty} P(B\mid A_j)\,P(A_j), \tag{9}$$

если $P(A_j)>0$, $j=1, 2, ...$.

Если $P(B)>0$, то согласно определению условной вероятности

$$P(A_k\mid B)=\frac{P(B\mid A_k)\,P(A_k)}{P(B)}, \tag{10}$$

или, воспользовавшись формулой (8) или (9)

$$P(A_k\mid B)=\frac{P(B\mid A_k)\,P(A_k)}{\sum\limits_{j} P(B\mid A_j)\,P(A_j)}. \tag{11}$$

Формулы (10) и (11) называются **формулами Байеса** и обычно используются следующим образом. Пусть известны вероятности $P(A_j)$ попарно несовместных событий A_j, $j=1, 2, ...$, образующих полную группу, и условные вероятности $P(B|A_j)$, $j=1, 2, ...$, события B. Вероятности $P(A_j)$, $j=1, 2, ...$, носят название априорных, или доопытных. Если в результате опыта происходит событие B, то априорные вероятности по формуле Байеса могут быть пересчитаны в апостериорные, или послеопытные, вероятности $P(A_j|B)$, $j=1, 2, ...$.

Байесовский пересчет вероятностей наглядно сводится к заменам:

$$\Omega\to B,\ A_k\to A_k\cap B,\ P(A_k)\to P(A_k\cap B)/P(B)=P(A_k\mid B).$$

В качестве иллюстрации рассмотрим задачу, известную как задача о разорении игрока. Некто при выпадении единицы получает одну копейку, при выпадении минус единицы платит одну копейку (получает минус одну копейку). По условию игра заканчивается при выполнении хотя бы одного из двух условий: либо игрок набирает капитал a копеек, либо разоряется, т. е. набирает 0 копеек. Пусть x — начальный капитал, $x\leqslant a$. Обозначим $P(x)$ вероятность разорения игрока. Будем считать, что «1» и «—1» выпадают с вероятностью 1/2. Обозначим A — разорение игрока, A_1 — выпадение «1», A_2 — выпадение «—1». Тогда $A=A\cap A_1+A\cap A_2$, и по формуле полной вероятности $P(x)=P(x|A_1)P(A_1)+P(x|A_2)P(A_2) = [P(x+1)+P(x-1)]\cdot 1/2$. Общее решение этого разностного уравнения имеет вид $P(x)=cx+d$, что совместно с граничными условиями $P(0)=1$, $P(a)=0$, дает $P(x)=1-x/a$, $0\leqslant x\leqslant a$.

§ 5. ПОСЛЕДОВАТЕЛЬНОСТЬ НЕЗАВИСИМЫХ ИСПЫТАНИЙ

Рассмотрим дискретное вероятностное пространство $(\Omega, \mathcal{F}, P)$, в котором $\Omega=\{\omega_1, \omega_2, ...\}$, $\mathcal{F}$ состоит из всех подмножеств Ω и вероятность P определена для каждого элементарного события $\omega_j\in\Omega$ равенством $P(\{\omega_j\})=p_j$, $j=1, 2, ..., \sum\limits_{j=1}^{\infty} p_j=1$. Будем считать, что вероятностное пространство $(\Omega, \mathcal{F}, P)$ сопоставлено опыту S, а $\omega_1, \omega_2, ...$ — возможные исходы S. Тогда опыту S, повторенному дважды, должно быть сопоставлено вероятностное пространство $(\Omega_2, \mathcal{F}_2, P_2)=(\Omega, \mathcal{F}, P)\times(\Omega, \mathcal{F}, P)$. Элементарными событиями теперь будут упорядоченные пары исходов $(\omega_i\omega_j)\in\Omega\times\Omega=\Omega_2$, $\mathcal{F}_2=\mathcal{F}\times\mathcal{F}$ — σ-алгеброй подмножеств Ω_2. Вероятность P_2 на $\mathcal{F}_2$ можно определить многими способами, которые, в свою очередь, определяются условиями проведения опытов. Однако если исходы первого опыта никак не влияют на исходы второго, то согласно результатам предыдущего параграфа следует положить

$$P_2(\{(\omega_i\omega_j)\})=p_ip_j,\quad i, j=1, 2, ...$$

Там было показано, что так определенная вероятность на $\mathcal{F}_2$ отвечает условию независимости опытов. События, связанные с каждым опытом, являются цилиндрическими множествами Ω_2.

Очевидно, n раз независимо повторенному опыту S отвечает вероятностное пространство $(\Omega_n, \mathcal{F}_n, P_n)=[\times(\Omega, \mathcal{F}, P)]^n$, в котором $\Omega_n=[\times\Omega]^n$, $\mathcal{F}_n=[\times\mathcal{F}]^n$, а вероятность P_n задана равенствами

$$P_n(\{(\omega_i\omega_j...\omega_k)\})=p_ip_j...p_k,\quad i, j, ..., k=1, 2, ... \tag{1}$$

Определение. Пусть $(\Omega, \mathscr{F}, P)$ — дискретное вероятностное пространство, $\Omega=\{\omega_1, \omega_2, ...\}$, $P(\{\omega_j\})=p_j$, $j=1, 2, ...$, $\sum_{j=1}^{\infty} p_j = 1$. **Последовательностью n независимых испытаний** называется вероятностное пространство $(\Omega_n, \mathscr{F}_n, P_n)$, элементарными событиями в котором являются последовательности $(\omega_{i_1}\omega_{i_2}...\omega_{i_n})$ и вероятность определена для каждого элементарного события равенством (1). Последовательность независимых испытаний называется **схемой Бернулли**, если $\Omega=\{\omega_1, \omega_2\}$, т. е. если опыт S имеет лишь два исхода.

Обычно в схеме Бернулли исход ω_1 называют успехом и обозначают соответствующую вероятность $P(\{\omega_1\})=p$, ω_2 называют неудачей, $P(\{\omega_2\})=1-p=q$. В серии из n независимых испытаний в схеме Бернулли Ω_n состоит из 2^n элементарных событий, причем

$$P_n(\{(\omega_{i_1}\omega_{i_2}\ldots\omega_{i_n})\}) = p^k q^{n-k},$$

где k — число успехов в последовательности $\omega_{i_1}, \omega_{i_2}, \ldots, \omega_{i_n}$ исходов.

В качестве примеров схемы Бернулли можно привести опыт с бросанием монеты или игральной кости. В последнем случае можно считать, что ω_1 — выпадение одного очка, ω_2 — невыпадение одного очка.

В связи со схемой Бернулли найдем вероятность того, что в серии из n испытаний успех наступит $k\leqslant n$ раз. Поскольку при этом не имеет значения, когда именно в этих испытаниях будут наблюдаться k успехов, то событие A_k, состоящее в наступлении k успехов, является объединением всех различных событий $(\omega_{i_1}\omega_{i_2}\ldots\omega_{i_n})$, в которых ω_1 встречается k раз. Всего таких элементарных событий в A_k C_n^k, а поскольку они несовместны, искомая вероятность дается равенством

$$P_n(A_k) = p_n(k) = C_n^k p^k q^{n-k}, k = 0, 1, \ldots, n. \quad (2)$$

В частности, вероятность не иметь ни одного успеха равна q^n, и, следовательно, вероятность иметь хотя бы один успех равна $1-q^n$. Совокупность вероятностей $p_n(k)$, $k=0, 1, ..., n$, называется **биномиальным распределением**. Такое название связано с тем, что вероятность (2) является общим членом разложения бинома

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

Это равенство в данном случае является отражением того факта, что в серии из n независимых испытаний Бернулли исходы, содержащие 0, 1, ..., n успехов образуют полную группу попарно несовместных событий.

Рассмотрим отношение

$$\frac{p_n(k+1)}{p_n(k)} = \frac{n!\, p^{k+1} q^{n-k-1}}{(k+1)!\,(n-k-1)!} \cdot \frac{k!\,(n-k)!}{n!\, p^k q^{n-k}} = \frac{p\,(n-k)}{q\,(k+1)}.$$

Так как неравенство $p_n(k+1)/p_n(k)>1$ эквивалентно $np-q>k$, то вероятность $p_n(k)$ возрастает при переходе от k к $k+1$, если $np-q>k$, и, наоборот, убывает при переходе от k к $k+1$, если $np-q<k$. Если же $np-q=k$, то $p_n(k)=p_n(k+1)$. Заметим, что при достаточно малых вероятностях успеха p, таких, что $np-q\leqslant 0$, вероятность $p_n(k)$ не возрастает ни при каких $k\geqslant 0$. В этом случае вероятнее всего, что при n испытаниях не будет ни одного успеха.

Возвращаясь к общей схеме последовательности независимых испытаний, рассмотрим случай конечного $\Omega=\{\omega_1, ..., \omega_r\}$ и положим $p(\{\omega_j\})=p_j$, $j=1, ..., r$. В каждом испытании теперь возможно r различных исходов, и при n испытаниях вероятность исхода $\omega_{i_1}, \ldots, \omega_{i_n}$ равна $p_1^{s_1}\ldots p_r^{s_r}$, где $s_1, s_2, ..., s_r$ — количества исходов $\omega_1, \omega_2, ..., \omega_r$ соответственно в последовательности $\omega_{i_1}, \omega_{i_2}, \ldots, \omega_{i_r}$, $s_1+s_2+...+s_r=n$. Нетрудно заметить, что при n испытаниях вероятность того, что ω_1 наблюдается s_1 раз, ω_2 — s_2 раз, ..., ω_r — s_r раз, равна

$$p_n(s_1, \ldots s_r) = \frac{n!}{s_1!\ldots s_r!} p_1^{s_1}\ldots p_r^{s_r}, s_1 + \ldots + s_r = n. \quad (3)$$

Это так называемое **полиномиальное распределение.** Название объясняется тем, что вероятность (3) является общим членом разложения полинома $(p_1+...+p_r)^n$. Множитель $n!/(s_1!...s_r!)$ равен числу различных исходов в n испытаниях, при которых ω_j наблюдается s_j раз, $j=1, ..., r$. Для того чтобы получить это число, рассмотрим последовательность $\omega_{i_1}, \omega_{i_2}, \ldots, \omega_{i_n}$, содержащую s_j исходов ω_j, $j=1, ..., r$. Всего существует $n!$ возможностей расположить исходы $\omega_{i_1}, \omega_{i_2}, \ldots, \omega_{i_n}$, однако среди них лишь $n!/(s_1!...s_r!)$ различных расположений, так как $s_1!$ перестановок s_1 исходов ω_1, ..., $s_r!$ перестановок s_r исходов ω_r приводят к эквивалентным расположениям.

С рассмотренными распределениями тесно связано так называемое отрицательно-биномиальное распределение, называемое также распределением Паскаля. Речь идет о вероятности того, что в последовательности испытаний Бернулли для достижения n успехов потребуется $n+k$ испытаний. Искомая вероятность равна

$$p(n, n+k) = C_{n-1+k}^{n-1} p^n q^k, k=0, 1, \ldots. \quad (4)$$

Действительно, выражение (4) дает вероятность того, что в серии $n+k-1$ испытаний будет ровно k неудач, а следующее, $n+k$-е, испытание приведет к успеху

$$C_{n+k-1}^{k} q^k p^{n-1} p = C_{n+k-1}^{n-1} p^n q^k.$$

Название распределения (4) связано со следующим равенством:

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

которое позволяет переписать распределение (4) в виде

$$p(n, n+k) = C_{-n}^{k} p^n (-q)^k, k = 0, 1, \dots .$$

Заметим, что

$$\sum_{k=0}^{\infty} C_{-n}^{k} (-q)^k = (1-q)^{-n} = p^{-n},$$

и, следовательно, $\sum\limits_{k=0}^{\infty} p(n, n+k) = 1$. При $n=1$ распределение (4) называется геометрическим

$$p(1, k+1) = pq^k, \quad k = 0, 1, \dots .$$

В качестве иллюстрации рассмотрим следующую задачу С. Банаха. Пусть a и b — две коробки, содержащие по n спичек. Некто пользуется ими, выбирая коробку a или b соответственно с вероятностями $p(a)=p$ или $p(b)=1-p=q$ и доставая каждый раз по одной спичке. Какова вероятность того, что в тот момент, когда выбранная коробка окажется пустой, в другой будет $r<n$ спичек?

Для решения задачи заметим, что если вынутая коробка пуста, а в другой в этот момент r спичек, то это означает, что до этого коробки вынимались $2n-r$ раз, причем одна из них — n раз. Пусть A — событие, состоящее в том, что вынутая коробка пуста. Такой коробкой может оказаться как a, так и b. Обозначим через $P(A|a)$ и $P(A|b)$ соответствующие условные вероятности. Тогда искомая вероятность равна

$$P(A) = P(A|a)p(a) + P(A|b)p(b).$$

Поскольку $P(A|a)$ — вероятность вынуть пустую коробку при условии, что вынута коробка a, то речь идет о вероятности того, что перед этим коробка a доставалась n раз из общего числа $2n-r$ выборов. Следовательно, $P(A|a) = C_{2n-r}^{n} p^n q^{n-r}$. Аналогично $P(A|b) = C_{2n-r}^{n} q^n p^{n-r}$. Поэтому $P(A) = C_{2n-r}^{n}(p^{n+1} q^{n-r} + q^{n+1} p^{n-r})$.

Следует отличать найденную вероятность от вероятности того, что в тот момент, когда одна из коробок опустела, во второй оказалось r спичек. И вообще, коробка, вынутая пустой, не обязательно опустела первой. Читателю предлагается самостоятельно подсчитать вероятность, о которой идет речь.

Рассмотрим задачу Банаха, пользуясь отрицательно-биномиальным распределением. Искомая вероятность совпадает с вероятностью того, что для $n+1$ «выбора» коробки a или b потребуется $2n-r+1$ «испытаний». Обозначим через A событие, состоящее в том, что для $n+1$ «выбора» a или b потребуется $2n-r+1$ «испытаний». Тогда $A = A\cap a + A\cap b$ и согласно формуле (4) для каждого из событий $A\cap a$ или $A\cap b$ найдем

$$P(A\cap a) = C_{2n-r}^{n} p^{n+1} q^{n-r}, \; P(A\cap b) = C_{2n-r}^{n} q^{n+1} p^{n-r},$$

что приводит к найденному ранее значению вероятности $P(A)$.

В заключение сделаем одно замечание по поводу схемы независимых испытаний, когда вероятности исходов меняются от испытания к испытанию. В таком случае серии из n независимых испытаний отвечает вероятностное пространство

$$(\Omega_n, \mathcal{F}_n, P_n) = (\Omega, \mathcal{F}, P^{(1)}) \times (\Omega, \mathcal{F}, P^{(2)}) \times \dots \times (\Omega, \mathcal{F}, P^{(n)}).$$

Всякому элементарному событию $(\omega_{i_1}\omega_{i_2}\dots\omega_{i_n})$ должна быть сопоставлена вероятность

$$P_n(\{(\omega_{i_1}\dots\omega_{i_n})\}) = p_{i_1}^{(1)} p_{i_2}^{(2)} \dots p_{i_n}^{(n)},$$

где $p_{i_k}^{(k)}$ — вероятность исхода ω_{i_k} в k-м испытании.

§ 6. РАСПРЕДЕЛЕНИЕ ПУАССОНА

Формулы биномиального распределения вероятностей приводят при больших n к очень громоздким вычислениям. Важно поэтому иметь приближенные, но зато достаточно простые формулы для вычисления соответствующих вероятностей. В частности, нередко встречаются задачи, в которых рассматривается большое число независимых испытаний, причем вероятность наступления события A при каждом отдельном испытании мала.

В этом случае вероятности $P_n(k)$ могут быть приближенно вычислены по так называемой **формуле Пуассона.** Эта формула получается как предельная для биномиального распределения, когда число испытаний стремится к бесконечности, а вероятность успеха стремится к нулю, однако в пределах одной последовательности независимых испытаний ни то, ни другое по определению невозможно. Поэтому естественно введение следующей конструкции.

Представим себе, что мы произвели некоторую серию независимых испытаний, состоящую из конечного числа испытаний, затем новую серию, затем еще новую и т. д. Мы будем иметь последовательность серий испытаний. Пусть при каждом испытании каждой серии может наступить или не наступить некоторое событие A, т. е. имеем всего два исхода, и пусть вероятность наступления A при отдельном испытании остается постоянной в пределах каждой серии (как это требуется для последовательности независимых испытаний), но может меняться от серии к серии. В этих условиях справедлива

Теорема Пуассона. Пусть дана последовательность $\{s_n\}$ серий независимых испытаний, состоящих соответственно из 1, 2, ..., n, ... испытаний, и пусть вероятность p события A при каждом испытании n-й серии равна λ/n, где λ — постоянная (не зависящая от n). Тогда вероятность $P_n(m)$ того, что число наступлений события A в n-й серии будет равно m, при $n\to\infty$ и фиксированном m стремится к $(\lambda^m/m!)e^{-\lambda}$.

Доказательство. Имеем

$$\lim_{n\to\infty} P_n(m) = \lim_{n\to\infty} \frac{n!}{m!\,(n-m)!}\, p^m (1-p)^{n-m} =$$

$$= \lim_{n\to\infty} \frac{n!}{m!\,(n-m)!} \left(\frac{\lambda}{n}\right)^m \left(1-\frac{\lambda}{n}\right)^{n-m} = \lim_{n\to\infty} \frac{\lambda^m}{m!}\left(1-\frac{\lambda}{n}\right)^n \times$$

$$\times \left\{\frac{n(n-1)\ldots(n-m+1)}{n^m}\left(1-\frac{\lambda}{n}\right)^{-m}\right\} = \frac{\lambda^m}{m!}\lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^n \times \quad (1)$$

$$\times \lim_{n\to\infty}\left\{1\left(1-\frac{1}{n}\right)\ldots\left(1-\frac{m-1}{n}\right)\left(1-\frac{\lambda}{n}\right)^{-m}\right\} = \frac{\lambda^m}{m!}\,e^{-\lambda}. \quad \blacktriangle$$

Распределение вероятностей, определяемое формулой

$$P(m) = \frac{\lambda^m}{m!}\,e^{-\lambda},\ m = 0, 1, 2, \ldots,\quad \lambda>0, \quad (2)$$

называется **распределением Пуассона** и является одним из важнейших в теории вероятностей. Эта формула получена нами как предельная для последовательности серий независимых испытаний, в которой число испытаний стремится к бесконечности, а вероятность наступления события A стремится к нулю. Поэтому ясно, что если мы хотим воспользоваться приближенной формулой

$$P_n(m) \approx \frac{\lambda^m}{m!}\,e^{-\lambda},\ \lambda = np,\ m = 0, 1, \ldots, n \quad (3)$$

для одной серии из n испытаний, то n должно быть велико, а вероятность наступления события A в этой серии мала, т. е. формула (3) применима для «редких явлений». Проведем соответствующие оценки погрешности, допускаемой при замене биномиального закона на закон Пуассона при конечных n и при $0\leqslant m\leqslant n$. Имеем согласно (1):

$$j(m, n) \equiv \ln\frac{P_n(m)}{P(m)} =$$

$$= \ln\left\{\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\ldots\left(1-\frac{m-1}{n}\right)\left(1-\frac{\lambda}{n}\right)^{n-m} e^{\lambda}\right\} =$$

$$= \sum_{k=1}^{m-1} \ln\left(1-\frac{k}{n}\right) + (n-m)\ln\left(1-\frac{\lambda}{n}\right) + \lambda. \quad (4)$$

Оценим сверху и снизу $\ln(1-x)$ при $0\leqslant x<1$:

$$\ln(1-x) = -x - \frac{x^2}{2} - \frac{x^3}{3} - \ldots < -x - \frac{x^2}{2} < -x, \quad (5)$$

$$\ln(1-x) = -x - \frac{x^2}{2} - \frac{x^3}{3} - \ldots > -x -$$

$$- \frac{x^2}{2} - \frac{x^3}{2} - \frac{x^4}{2} - \ldots = -x - \frac{x^2}{2}(1 + x + x^2 + \ldots) =$$

$$= -x - \frac{x^2}{2}\,\frac{1}{1-x} > -\frac{1}{1-x}. \quad (6)$$

Используя (5) и (6), оценим (4):

$$j(m, n) < -\frac{1}{n}\sum_{k=1}^{m-1} k + \lambda + (n-m)\left(-\frac{\lambda}{n} - \frac{\lambda^2}{2n^2}\right) = \quad (7)$$

$$= -\frac{m(m-1)}{2n} + \frac{\lambda m}{n} - \frac{\lambda^2}{2n} + \frac{m\lambda^2}{2n^2} = -\frac{(m-\lambda)^2}{2n} + \frac{m}{2n} + \frac{m\lambda^2}{2n^2}$$

и

$$j(m, n) > -\frac{1}{1-\frac{m-1}{n}}\,\frac{1}{n}\sum_{k=1}^{m-1} k + \lambda +$$

$$+ (n-m)\left(-\frac{\lambda}{n} - \frac{\lambda^2}{2n^2\left(1-\frac{\lambda}{n}\right)}\right) = -\frac{m(m-1)}{2(n-m+1)} +$$

$$+ \frac{m\lambda}{n} - \frac{\lambda^2}{2(n-\lambda)} + \frac{m\lambda^2}{2n(n-\lambda)} \geqslant -\frac{(m-\lambda)^2}{2(n-m+1)} +$$

$$+\frac{(m-\lambda-1)\,\lambda^2}{2\,(n-\lambda)\,(n-k+1)}-\frac{m\,(m-1)\,\lambda}{n(n-m+1)}. \tag{8}$$

Представляя теперь биномиальное распределение в виде

$$P_n(m)=\frac{\lambda^m}{m!}\,e^{-\lambda}\,e^{j(m,n)}, \tag{9}$$

из (7) и (8) заключаем, что множитель $e^{j(m,n)}$ отличается от единицы на величину порядка $1/n$ при $n\to\infty$ и любом фиксированном m. На практике формула (3) служит хорошим приближением, если $n\geqslant 100$, а $0\leqslant np<10$ $(m=0,\ 1,\ \ldots,\ n)$.

Обратимся к распределению Пуассона. Оно определено для всех целых неотрицательных $m=0,\ 1,\ 2,\ \ldots$

$$\sum_{m=0}^{\infty} P(m)=e^{-\lambda}\sum_{m=0}^{\infty}\frac{\lambda^m}{m!}=1. \tag{10}$$

Пусть λ фиксировано, рассмотрим поведение $P(m)$ как функции m. Поскольку $P(m)/P(m-1)=\lambda/m$, то $P(m)$ возрастает при увеличении m от нуля до $m_0=[\lambda]$ ($[\lambda]$ — целая часть числа λ), а затем убывает; $\max\limits_m P(m)=\dfrac{\lambda^{[\lambda]}}{[\lambda]!}e^{-\lambda}$. Если λ — целое, то $P(m)$ имеет максимальное значение при $m=\lambda$ и $m=\lambda-1$. Распределение Пуассона описывает вероятности во многих задачах, таких, как число отказов радиоэлектронной аппаратуры за время t, число радиоактивных атомов, распавшихся к моменту t, число вызовов на телефонной станции и т. д.

Примеры

1. В лотерее в среднем разыгрывается один выигрыш на 1000 номеров. Какова вероятность, имея 100 билетов, получить не менее 2 выигрышей?

Это схема Бернулли с $n=100$, вероятностью успеха $p=1/1000$, так что $\lambda=np=0,1$. Искомая вероятность

$$p_x=\sum_{m=2}^{100} P_{100}(m)=1-P_{100}(0)-P_{100}(1).$$

По формуле Пуассона $P_{100}(0)\approx P(0)=e^{-\lambda}\approx 0,9$, $P_{100}(1)\approx\approx P(1)=\lambda e^{-\lambda}\approx 0,09$. Таким образом, $p_x\approx 1-0,9-0,09=0,01$.

2. Вероятность зарегистрировать частицу счетчиком равна 10^{-4}. Какое наименьшее число частиц должно вылететь из источника для того, чтобы с вероятностью, не меньшей 0,99, счетчик зарегистрировал более 3 частиц?

Обозначим через n искомое число частиц, а через A — событие:

$A=$\{счетчик зарегистрировал более трех частиц\}. Имеем $p(A)=1-p(\bar A)$, и по формуле Пуассона

$$p(\bar A)=P_n(0)+P_n(1)+P_n(2)+P_n(3)\approx P(0)+P(1)+P(2)+$$
$$+P(3)=e^{-\lambda}(1+\lambda+\lambda^2/2+\lambda^3/6)\leqslant 0,01, \tag{11}$$

ибо по условию $p(A)\geqslant 0,99$. По таблицам для распределения Пуассона находим λ, удовлетворяющее неравенству (11), $\lambda=10,7$. Отсюда $n=\lambda/p=10,7\cdot 10^4$.

Для полиномиального распределения справедлив аналог теоремы Пуассона. Мы рассмотрим его на примере следующей задачи. Пусть T — область евклидова пространства, в которую наугад бросается N точек, пусть $t_0,\ t_1,\ \ldots,\ t_n$ — непересекающиеся области, такие, что $\bigcup\limits_{j=0}^{n} t_j=T$. Каждый исход опыта состоит в попадании точки в одну из областей t_j, $j=0,\ 1,\ \ldots,\ n$. Вероятности исходов равны соответственно

$$p_1=\frac{|t_1|}{|T|},\ \ldots,\ p_n=\frac{|t_n|}{|T|}, \tag{12}$$

$$p_0=1-p_1-\ldots-p_n=\frac{|T|-|t_1|-\ldots-|t_n|}{|T|},$$

где $|t_j|$ — объем области t_j, $|T|$ — объем области T. Этот пример естественно рассматривать как схему N независимых испытаний с $(n+1)$-м исходом. Обозначим через $\xi(t_j)$ — число частиц, попавших в j-ю область. Воспользовавшись полиномиальным распределением, получим

$$P\{\xi(t_1)=k_1,\ \ldots,\ \xi(t_n)=k_n,\ \xi(t_0)=k_0\}=$$
$$=\frac{N!}{k_0!\,k_1!\ldots k_n!}\,p_0^{k_0}p_1^{k_1}\ldots p_n^{k_n}, \tag{13}$$
$$k_0+k_1+\ldots+k_n=N.$$

Положим $k=k_1+\ldots+k_n$, тогда $k_0=N-k$. Считая k фиксированным, устремим $N\to\infty$. Согласно формуле Стирлинга

$$\frac{N!}{\sqrt{2\pi N}\,e^{-N}\,N^N}\to 1,\quad \frac{(N-k)!}{\sqrt{2\pi(N-k)}\,(N-k)^{N-k}\,e^{-N+k}}\to 1,$$
$$\text{при } N\to\infty. \tag{14}$$

Заметим также, что если мы обозначим через λ величину $\lambda=N/|T|$ — среднее число точек на единицу объема, то будем иметь

$$p_j^{k_j}=\left(\frac{|t_j|}{|T|}\right)^{k_j}=(|t_j|\,\lambda)^{k_j}N^{-k_j},\ j=1,\ \ldots,\ n, \tag{15}$$

$$p_0^{k_0}=\left(1-\frac{|t_1|+\ldots+|t_n|}{|T|}\right)^{N-k}=\left(1-\lambda\,\frac{|t_1|+\ldots+|t_n|}{N}\right)^{N-k}.$$

Подставляя (14) и (15) в (13), найдем

$$\lim_{N\to\infty}\frac{N!}{k_1!\ldots k_n!\,(N-k)!}\,p_1^{k_1}\ldots p_n^{k_n}p_0^{k_0}=\lim_{N\to\infty}\left[\frac{\sqrt{2\pi N}}{\sqrt{2\pi(N-k)}}\times\right.$$

$$\times\frac{N^{N-k}}{(N-k)^{N-k}}\,\frac{e^{-N}N^k}{e^{-(N-k)}}\,\frac{(|t_1|\lambda)^{k_1}}{N^{k_1}}\cdots\frac{(|t_n|\lambda)^{k_n}}{N^{k_n}}\times$$

$$\left.\times\left(1-\frac{\lambda(|t_1|+\ldots+|t_n|)}{N}\right)^{N-k}\frac{1}{k_1!\ldots k_n!}\right]=$$

$$=\frac{(|t_1|\lambda)^{k_1}\ldots(|t_n|\lambda)^{k_n}}{k_1!\ldots k_n!}\lim_{N\to\infty}\left[e^k\left(1-\frac{k}{N}\right)^{-N}\times\right.$$

$$\left.\times\left(1-\frac{\lambda(|t_1|+\ldots+|t_n|)}{N}\right)^{N-k}\right]=\tag{16}$$

$$=\frac{(|t_1|\lambda)^{k_1}\ldots(|t_n|\lambda)^{k_n}}{k_1!\ldots k_n!}\,e^{-\lambda(|t_1|+\ldots+|t_n|)}=\prod_{j=1}^{n}\frac{(|t_j|\lambda)^{k_j}}{k_j!}\,e^{-\lambda|t_j|}.$$

Таким образом, в пределе при $N\to\infty$ и фиксированном k для вероятности того, что в области t_j попадет k_j точек, мы получили выражение

$$P\{\xi(t_1)=k_1,\ldots,\xi(t_n)=k_n\}=\prod_{j=1}^{n}\frac{(|t_j|\lambda)^{k_j}}{k_j!}\,e^{-\lambda|t_j|},\tag{17}$$

представляющее собой **n-мерное распределение Пуассона.**

§ 7. ЛОКАЛЬНАЯ И ИНТЕГРАЛЬНАЯ ПРЕДЕЛЬНЫЕ ТЕОРЕМЫ МУАВРА — ЛАПЛАСА

Как мы уже отмечали в начале предыдущего параграфа, при большом числе испытаний формулы для вычисления вероятностей $P_n(m)$, отвечающих биномиальному распределению, весьма громоздки. В связи с этим надо иметь более простые приближенные формулы. Для случая, когда вероятность p успеха при каждом испытании очень мала, мы установили приближенную формулу — распределение Пуассона. Сейчас мы рассмотрим другую предельную форму биномиального распределения, считая, что вероятность p успеха отлична от нуля и единицы.

Локальная теорема Муавра—Лапласа. Если вероятность события A в n независимых испытаниях равна p, $0<p<1$, то

вероятность $P_n(m)$ того, что в этих испытаниях событие A наступит m раз, удовлетворяет при $n\to\infty$ соотношению

$$\lim_{n\to\infty}\frac{\sqrt{npq}\,P_n(m)}{\frac{1}{\sqrt{2\pi}}\,e^{-\frac{x_m^2}{2}}}=1,\tag{1}$$

$q=1-p$, $x_m=\dfrac{m-np}{\sqrt{npq}}$, $x_m\in[a,b]$, где $a<b$, a и b — любые конечные фиксированные числа.

Стремление к пределу в (1) равномерно относительно всех m, для которых $x_m\in[a,b]$.

Доказательство. Имеем

$$\sqrt{npq}\,P_n(m)=\sqrt{npq}\,\frac{n!}{m!\,(n-m)!}\,p^mq^{n-m}.$$

Воспользовавшись формулой Стирлинга,

$$k!=\sqrt{2\pi k}\,k^ke^{-k}e^{\theta_k},\ |\theta_k|\leqslant\frac{1}{12k},$$

получим

$$\sqrt{npq}\,P_n(m)=\sqrt{npq}\,\frac{1}{\sqrt{2\pi}}\,\frac{n^n\sqrt{n}\,e^{\theta_n-\theta_m-\theta_{n-m}}p^mq^{n-m}}{m^m(n-m)^{n-m}\sqrt{m}\sqrt{n-m}}=$$

$$=\frac{1}{\sqrt{2\pi}}\left[\left(\frac{np}{m}\right)^m\left(\frac{nq}{n-m}\right)^{n-m}\right]\sqrt{\frac{n^2pq}{m(n-m)}}\,e^{\theta_n-\theta_m-\theta_{n-m}}\equiv$$

$$\equiv\frac{1}{\sqrt{2\pi}}\,A_n(x_m)\,B_n(x_m)\,C_n(x_m).\tag{2}$$

Найдем пределы выражений $A_n(x_m)$, $B_n(x_m)$, $C_n(x_m)$ при $n\to\infty$.

Пусть $[a,b]$ — произвольный конечный интервал; будем рассматривать такие m, для которых $x_m\in[a,b]$. Так как $x_m=\dfrac{m-np}{\sqrt{npq}}$, то

$$m=np+x_m\sqrt{npq},\ n-m=n(1-p)-x_m\sqrt{npq}=$$

$$=nq-x_m\sqrt{npq},\ a\leqslant x_m\leqslant b.\tag{3}$$

Начнем с рассмотрения $C_n(x_m)=e^\theta$, $\theta=\theta_n-\theta_m-\theta_{n-m}$.

$$|\theta|\leqslant|\theta_n|+|\theta_m|+|\theta_{n-m}|\leqslant\frac{1}{12}\left(\frac{1}{n}+\frac{1}{np+x_m\sqrt{npq}}+\right.\tag{4}$$

$$\left.+\frac{1}{nq-x_m\sqrt{npq}}\right)=\frac{1}{12n}\left(1+\frac{1}{p+x_m\sqrt{\frac{pq}{n}}}+\frac{1}{q-x_m\sqrt{\frac{pq}{n}}}\right).$$

Из (4) в силу признака Вейерштрасса следует, что $\theta \to 0$ при $n \to \infty$ равномерно по $x_m \in [a, b]$. Таким образом,

$$C_n(x_m) \to 1 \text{ при } n \to \infty \tag{5}$$

равномерно относительно x_m, $x_m \in [a, b]$. Далее, в силу (3)

$$B_n(x_m) = \sqrt{\frac{n^2pq}{m(n-m)}} =$$

$$= \sqrt{\frac{1}{\left(1 + x_m\sqrt{\frac{q}{np}}\right)\left(1 - x_m\sqrt{\frac{p}{nq}}\right)}} \to 1, \quad n \to \infty \tag{6}$$

равномерно относительно x_m, $x_m \in [a, b]$, на основании признака Вейерштрасса.

Рассмотрим, наконец, $A_n(x_m)$. Пользуясь формулой

$$\ln(1+z) = z - \frac{z^2}{2} + O(z^3), \quad |z| < 1,$$

получим

$$\ln A_n(x_m) = -m \ln\left(\frac{m}{np}\right) - (n-m)\ln\left(\frac{n-m}{nq}\right) =$$

$$= -(np + x_m\sqrt{npq})\ln\left(1 + x_m\sqrt{\frac{q}{np}}\right) - \tag{7}$$

$$-(nq - x_m\sqrt{npq})\ln\left(1 - x_m\sqrt{\frac{p}{nq}}\right) =$$

$$= -\left[(np + x_m\sqrt{npq})\left(x_m\sqrt{\frac{q}{np}} - \frac{x_m^2 q}{2np} + O(n^{-3/2})\right) + \right.$$

$$\left. + (nq - x_m\sqrt{npq})\left(-x_m\sqrt{\frac{p}{nq}} - \frac{x_m^2 p}{2nq} + O(n^{-3/2})\right)\right] =$$

$$= -\left[x_m\sqrt{npq} + x_m^2 q - \frac{x_m^2 q}{2} + O(n^{-1/2}) - x_m\sqrt{npq} + \right.$$

$$\left. + x_m^2 p - \frac{x_m^2 p}{2} + O(n^{-1/2})\right] = -\frac{1}{2}x_m^2 + O(n^{-1/2}),$$

причем, поскольку при $n \to \infty$ $x_m\sqrt{\frac{q}{np}}$, $x_m\sqrt{\frac{p}{nq}}$ стремятся к нулю равномерно по x_m, $x_m \in [a, b]$, оценку O-членов можно взять независящей от m.

Итак, имеем

$$A_n(x_m) = e^{-\frac{1}{2}x_m^2 + O\left(\frac{1}{\sqrt{n}}\right)}, \quad n \to \infty. \tag{8}$$

Теперь утверждение (1) следует из сопоставления формул (2), (5), (6) и (8). ▲

Установленная теорема дает оценку величины $P_n(m)$ при больших n и при фиксированном m:

$$P_n(m) \approx \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{npq}} e^{-\frac{x_m^2}{2}}, \quad x_m = \frac{m - np}{\sqrt{npq}}. \tag{9}$$

Однако практически при большом числе испытаний n и не слишком малой вероятности p нас редко интересует вероятность того, что данное событие наступит точно m раз. Важно бывает уметь оценить вероятность того, что число наступлений события лежит в некоторых границах. Такую оценку можно получить с помощью интегральной предельной теоремы Муавра—Лапласа.

Интегральная предельная теорема Муавра—Лапласа. Пусть m — число наступлений события A в серии из n независимых испытаний, p — вероятность наступления A при каждом испытании, $0<p<1$, a и b — любые фиксированные числа, $a<b$. Тогда

$$\lim_{n \to \infty} P\left\{a \leqslant \frac{m - np}{\sqrt{npq}} \leqslant b\right\} = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{x^2}{2}} dx, \tag{10}$$

причем стремление к пределу равномерно относительно a и b, $-\infty < a < b < +\infty$.

Хотя доказательство этой теоремы может быть установлено на основании локальной предельной теоремы, мы предпочтем получить его как совсем простое следствие центральной предельной теоремы (см. § 11).

Практическое применение интегральной предельной теоремы основано на приближенном равенстве

$$P\left\{a \leqslant \frac{m - np}{\sqrt{npq}} \leqslant b\right\} \approx \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{x^2}{2}} dx. \tag{11}$$

Оценка соответствующей погрешности показывает, что приближенная формула (11) обеспечивает хорошую точность уже при значениях $npq \geqslant 10$ и, разумеется, тем лучшую точность, чем больше эта величина.

Рассмотрим некоторые типичные задачи, связанные с интегральной теоремой Муавра—Лапласа.

1. Заданы число испытаний n и вероятность успеха p при каждом испытании. Требуется найти вероятность того, что число успехов m будет заключено между заданными числами m_1 и m_2, $0 \leqslant m_1 < m_2 \leqslant n$.

$$P\{m_1 \leqslant m \leqslant m_2\} = P\left\{\frac{m_1 - np}{\sqrt{npq}} \leqslant \frac{m - np}{\sqrt{npq}} \leqslant \frac{m_2 - np}{\sqrt{npq}}\right\} \approx$$

$$\approx \frac{1}{\sqrt{2\pi}} \int\limits_{\frac{m_1-np}{\sqrt{npq}}}^{\frac{m_2-np}{\sqrt{npq}}} e^{-\frac{x^2}{2}}\, dx. \qquad (11^*)$$

Пример. Игральная кость бросается 12 000 раз. Какова вероятность, что число выпадений единицы будет заключено между 1900 и 2150? Здесь $n=12\,000$, $p=1/6$, $q=5/6$, $\sqrt{npq}=100\sqrt{6}$, $m_1-np=-100$, $m_2-np=150$. Искомая вероятность p_x равна

$$p_x \approx \frac{1}{\sqrt{2\pi}} \int\limits_{-\sqrt{6}}^{\frac{3\sqrt{6}}{2}} e^{-\frac{x^2}{2}}\, dx = \frac{1}{\sqrt{2\pi}} \int\limits_{0}^{\frac{3\sqrt{6}}{2}} e^{-\frac{x^2}{2}}\, dx + \frac{1}{\sqrt{2\pi}} \int\limits_{0}^{\sqrt{6}} e^{-\frac{x^2}{2}}\, dx =$$

$$= \Phi\left(\frac{3\sqrt{6}}{2}\right) + \Phi(\sqrt{6}) = 0{,}99.$$

Функция $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int\limits_0^x e^{-\frac{z^2}{2}}\, dz$ называется **интегралом ошибок**, для нее составлены подробные таблицы, поскольку $\Phi(-x) = -\Phi(x)$, значения в таблицах указаны лишь для $x \geqslant 0$. $\Phi(-x) = -\Phi(x)$
→ для такой $\boxed{\Phi}$, вообще для ошибок $\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt$,

2. Пусть заданы числа p, α и β. Требуется определить, какое наименьшее число n испытаний надо произвести для того, чтобы с вероятностью, не меньшей β, частота m/n появлений успеха отклонялась от вероятности p не больше чем на α. Таким образом, надо найти n из условия

$$P\left\{\left|\frac{m}{n} - p\right| \leqslant \alpha\right\} \geqslant \beta.$$

Поскольку

$$P\left\{\left|\frac{m}{n} - p\right| \leqslant \alpha\right\} = P\left\{-\alpha\sqrt{\frac{n}{pq}} \leqslant \frac{m-np}{\sqrt{npq}} \leqslant \alpha\sqrt{\frac{n}{pq}}\right\} \approx$$

$$\approx \frac{1}{\sqrt{2\pi}} \int\limits_{-\alpha\sqrt{\frac{n}{pq}}}^{\alpha\sqrt{\frac{n}{pq}}} e^{-\frac{x^2}{2}}\, dx = 2\Phi\left(\alpha\sqrt{\frac{n}{pq}}\right),$$

то задача состоит в определении n из условия

$$2\Phi\left(\alpha\sqrt{\frac{n}{pq}}\right) \geqslant \beta. \qquad (12)$$

Ее решение находим с помощью таблиц для интегралов ошибок.

Пример. Сколько раз надо бросить монету для того, чтобы с вероятностью, не меньшей 0,99, частота появления герба отличалась от вероятности $p=1/2$ не больше чем на 0,01.

Имеем согласно (12) $2\Phi(y) \geqslant 0{,}99$. Из таблиц находим, что $y \geqslant 2{,}58$. Таким образом, $y = \alpha\sqrt{\frac{n}{pq}} \geqslant 2{,}58$, т. е. $\sqrt{n} \geqslant$

$$\geqslant 2{,}58 \frac{\sqrt{\frac{1}{2}\cdot\frac{1}{2}}}{0{,}01} = 129 \text{ и } n \geqslant 16641.$$

3. Пусть заданы числа n, p и β. Требуется определить границы возможных отклонений частоты появления успеха от вероятности p, т. е. надо найти α, для которого

$$P\left\{\left|\frac{m}{n} - p\right| \leqslant \alpha\right\} = \beta.$$

Согласно предыдущему примеру

$$P\left\{\left|\frac{m}{n} - p\right| \leqslant \alpha\right\} \approx 2\Phi\left(\alpha\sqrt{\frac{n}{pq}}\right) = \beta,$$

отсюда по таблицам определяем α.

Пример. Вероятность попадания в цель 1/10. Сделано 100 выстрелов, в каких пределах с вероятностью 0,8 будет лежать относительная частота попаданий. Здесь $p=0{,}1$, $q=0{,}9$, $\beta=0{,}8$, $\sqrt{\frac{n}{pq}} = 33{,}33$, $2\Phi(\alpha\cdot 33{,}33) = 0{,}8$. Отсюда $\alpha=0{,}04$. Таким образом, частота m/n попаданий с вероятностью 0,8 лежит в интервале $(1/10-0{,}04,\ 1/10+0{,}04)$.

Рассмотрим в заключение еще один пример использования интегральной предельной теоремы Муавра—Лапласа.

Пример. Телефонная станция A, обслуживающая 2000 абонентов, должна соединять их с другой станцией B. Какое наименьшее число x линий должно связывать A с B, чтобы в 99% случаев вызовов нашлась свободная линия. Пусть в течение наиболее напряженного часа дня каждый абонент разговаривает с B в среднем 2 минуты.

Найдем x. Естественно рассматривать описанную ситуацию как схему Бернулли с $n=2000$, p — вероятность вызова, $p=1/30$. Число x определяется из условия: вероятность того, что число вызовов $\geqslant x$, должна быть меньше, чем 0,01, т. е.

$$P\{m \geqslant x\} < 0{,}01, \text{ или } P\{m \geqslant x\} = P\left\{\frac{m-np}{\sqrt{npq}} \geqslant \frac{x-np}{\sqrt{npq}}\right\} \approx$$

$$\approx \frac{1}{\sqrt{2\pi}} \int\limits_{\frac{x-np}{\sqrt{npq}}}^{\infty} e^{-\frac{x}{2}}\, dx = \frac{1}{2} - \Phi\left(\frac{x-np}{\sqrt{npq}}\right) < 0{,}01,$$

отсюда по таблицам

$$\frac{x-np}{\sqrt{npq}} \geqslant 2{,}327, \quad x \geqslant \frac{200}{3} + 2{,}327 \cdot 8{,}027 = 85{,}4.$$

Таким образом, $x=86$.

§ 8. СЛУЧАЙНЫЕ ВЕЛИЧИНЫ И ФУНКЦИИ РАСПРЕДЕЛЕНИЯ

1°. Случайные величины и функции распределения

Изучая схему независимых испытаний, мы, по существу, имели дело с типичным примером случайной величины, когда рассматривали число успехов в серии из n испытаний. Примерами случайных величин являются: число вызовов в единицу времени на телефонной станции, время ожидания очередного вызова, число молекул газа, продиффундировавших из одного объема газа в другой, и т. д.

Для случайной величины характерно, что мы не можем заранее указать значение, которое она примет, хотя, с другой стороны, множество ее возможных значений считается известным. Это множество может быть конечным, как в упомянутом случае числа успехов, может совпадать с положительной полупрямой $[0, \infty)$, как в случае времени ожидания, и т. д. Однако для полного задания случайной величины следует еще указать вероятности тех значений, которые она может принимать, точнее вероятность на множестве ее значений. (Хотя, конечно, вероятность задается на некоторой σ-алгебре подмножеств пространства значений, но ради краткости иногда говорят и так.)

Прежде чем приступить к изучению математического понятия случайной величины, вернемся к схеме независимых испытаний и проанализируем известную нам случайную величину — число успехов. В рассматриваемом случае пространство элементарных событий Ω состоит из 2^n элементарных событий ω-последовательностей вида $\omega = (\overset{1}{\omega_1}, \overset{2}{\omega_1}, \overset{3}{\omega_2}, \ldots, \overset{n}{\omega_1})$. В схеме испытаний Бернулли нас интересовали события A_k, $k=0, 1, \ldots, n$, где A_k составляют те последовательности, в которых успех ω_1 встречается k раз. Следовательно, A_k содержит C_n^k таких элементарных событий ω, а поскольку вероятность каждого из них есть $P_n(\{\omega\}) = p^k q^{n-k}$, то $P_n(A_k) = C_n^k p^k q^{n-k}$. Рассмотрим функцию $\xi=\xi(\omega)$, определенную на данном Ω равенствами

$$\xi(\omega) = k, \quad \omega \in A_k, \quad k=0, 1, \ldots, n. \tag{1}$$

То, что эти равенства действительно определяют $\xi(\omega)$ для каждого $\omega \in \Omega$, следует из соотношения $A_0 + A_1 + \ldots + A_n = \Omega$. Так определенная функция $\xi = \xi(\omega)$ описывает число успехов в серии из n независимых испытаний Бернулли в том смысле, что число успехов в каждой последовательности испытаний ω равно (по определению) $\xi(\omega)$. Функция $\xi=\xi(\omega)$ называется **случайной величиной.** В данном случае эта случайная величина — число успехов в серии из n испытаний Бернулли.

Обозначим через $\{\omega : \xi(\omega) = k\}$ множество тех ω, для которых $\xi(\omega) = k$. Следовательно, по определению $A_k = \{\omega : \xi(\omega) = k\}$ и $P_n(\{\omega : \xi(\omega) = k\}) = C_n^k p^k q^{n-k}$. Последние вероятности обычно записывают короче:

$$P_n(k) = C_n^k p^k q^{n-k} = P\{\xi = k\}, \quad k = 0, 1, \ldots, n \tag{2}$$

и называют **распределением вероятностей случайной величины** ξ. Таким образом, случайная величина — число успехов в серии из n испытаний Бернулли — имеет биномиальное распределение.

Формула (2) задает вероятность на алгебре всех подмножеств множества значений случайной величины ξ. Последнее в данном случае состоит из $n+1$ точек: $\overline{\Omega} = \{0, 1, \ldots, n\}$, алгебра событий состоит из всех подмножеств $\overline{\Omega}$.

Таким образом, со случайной величиной ξ оказывается связанным *новое вероятностное пространство* $(\overline{\Omega}, \overline{\mathscr{F}}, \overline{P})$, в котором пространством элементарных событий является множество значений случайной величины. $\overline{\mathscr{F}}$ — аглебра всех подмножеств $\overline{\Omega}$, а вероятность $\overline{P}$ связана с вероятностью на исходном вероятностном пространстве формулой (2): $\overline{P}(\{k\}) = P_n(k)$.

Случайная величина $\xi = \xi(\omega)$ задает отображение вероятностного пространства $(\Omega, \mathscr{F}, P)$ на вероятностное пространство $(\overline{\Omega}, \overline{\mathscr{F}}, \overline{P})$. При этом каждой точке $k \in \overline{\Omega}$ отвечает ее прообраз в Ω-множество $\{\omega : \xi(\omega) = k\} \subset \Omega$. Задание $\xi(\omega)$ эквивалентно разбиению пространства элементарных событий: $\Omega = \{\omega : \xi(\omega) = 0\} + \{\omega : \xi(\omega) = 1\} + \ldots + \{\omega : \xi(\omega) = n\}$. Утверждения «$\xi$ попадает в $\overline{A} \in \overline{\mathscr{F}}$» и «$\omega$ попадает в $A \in \mathscr{F}$» эквивалентны (рис. 8).

Характерно, что в теоретико-вероятностных задачах явная зависимость $\xi = \xi(\omega)$ от ω, как правило, не играет существенной роли.

В связи с распределением Пуассона можно рассматривать пространство элементарных событий Ω, состоящее из бесконечных последовательностей $\omega = (\omega_1, \omega_1, \omega_2, \ldots)$ испытаний. Пусть ω содержит k успехов ω_1, а событие A_k состоит из всех таких ω (которые содержат k раз ω_1). Тогда положим по определению $P(A_k)=\lambda^k e^{-\lambda}/k!$, $k=0, 1, \dots$, и определим случайную величину ξ равенствами $\xi(\omega)=k$, $\omega \in A_k$. Эта случайная величина имеет распределение Пуассона, так как $P\{\xi=k\}=(\lambda^k/k!)e^{-\lambda}$, $k=0, 1, \dots$, $\lambda>0$. (Заметим, что здесь естественно считать возможным значение $\xi(\omega)=\infty$, причем

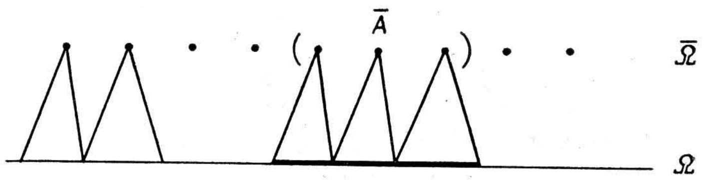


Рис. 8

$P\{\xi=\infty\}=0$.) Случайная величина $\xi=\xi(\omega)$ порождает новое вероятностное пространство $(\bar{\Omega}, \bar{\mathscr{F}}, \bar{P})$, в котором $\bar{\Omega}=\{0, 1, \dots, \infty\}$ — множество значений ξ, $\bar{\mathscr{F}}$ — σ-алгебра всех подмножеств $\bar{\Omega}$ и вероятность $\bar{P}$ определена для каждого одноточечного подмножества $\bar{\Omega}$ равенством $\bar{P}(\{k\})=(\lambda^k/k!)e^{-\lambda}$, $k=0, 1, \dots$, $\bar{P}(\{\infty\})=0$.

Это были примеры так называемых дискретных случайных величин. В общем случае случайная величина определяется следующим образом.

Определение 1. Пусть $(\Omega, \mathscr{F}, P)$ — вероятностное пространство. **Случайной величиной** ξ называется однозначная действительная функция $\xi=\xi(\omega)$, определенная на Ω, для которой множество элементарных событий вида $\{\omega: \xi(\omega)<x\}$ является событием (т. е. $\in\mathscr{F}$) для каждого действительного числа x.

В определении, таким образом, требуется, чтобы для каждого $x\in R_1$ множество $\{\omega: \xi(\omega)<x\}\in\mathscr{F}$, и это условие гарантирует, что для каждого x определена вероятность события $\{\xi<x\}: F(x)=P\{\xi<x\}$ (запись $\{\xi<x\}$ здесь и далее означает то же самое, что и $\{\omega:\xi(\omega)<x\}$).

Определение 2. Функция $F(x)=P\{\xi<x\}$, $-\infty<x<\infty$, называется **функцией распределения** случайной величины ξ.

Заметим, что, как будет видно из дальнейшего, функция $F(x)$ определяет вероятность на множестве значений ξ, являясь в то же время конструкцией, существенно более простой, чем вероятность.

Примеры

1. Пусть ξ — число успехов в серии из n испытаний Бернулли. Тогда соответствующая функция распределения определена равенством

$$F(x)=\begin{cases} 0, & x\leqslant 0, \\ \sum\limits_{k<x} C_n^k p^k q^{n-k}, & 0<x\leqslant n, \\ 1, & x>n, \text{ (рис. 9).} \end{cases} \tag{3}$$

2. Если ξ распределена по закону Пуассона, то ее функция распределения $F(x)$ имеет вид

$$F(x)=\begin{cases} 0, & x\leqslant 0, \\ \sum\limits_{k<x} \dfrac{\lambda^k e^{-\lambda}}{k!}, & x>0. \end{cases} \tag{4}$$

3. Говорят, что случайная величина ξ имеет **нормальное, или гауссово, распределение** $N(\mu, \sigma^2)$, если ее функция распределения имеет вид

$$F(x)=\frac{1}{\sigma\sqrt{2\pi}}\int\limits_{-\infty}^{x} e^{-\frac{(z-\mu)^2}{2\sigma^2}}\,dz. \tag{5}$$

При этом соответствующее ξ вероятностное пространство устроено следующим образом: $\bar{\Omega}$ — действительная прямая $-\infty<x<\infty$, σ-алгебра событий $\bar{\mathscr{F}}$—σ-алгебра борелевских множеств на прямой. Для каждого события $A\in\bar{\mathscr{F}}$

$$P(A)=\frac{1}{\sigma\sqrt{2\pi}}\int\limits_{A} e^{-\frac{(z-\mu)^2}{2\sigma^2}}\,dz. \tag{6}$$

Однако адекватное понимание этого последнего равенства возможно только в терминах интеграла Лебега. Мы ограничиваемся далее множествами A простой структуры: интервалами, конечными объединениями интервалов и т. п. В этом случае интеграл (6) можно понимать как интеграл Римана (в том числе и как несобственный).

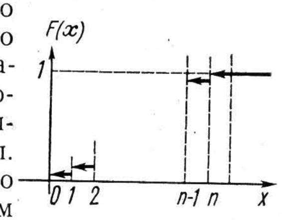


Рис. 9

Перечислим основные свойства функций распределения.

1. Для $x_2>x_1$

$$\{\xi<x_2\}=\{\xi<x_1\}+\{x_1\leqslant\xi<x_2\}.$$

Поскольку события в правой части этого равенства несовместны, то

$$P\{\xi<x_2\}=P\{\xi<x_1\}+P\{x_1\leqslant\xi<x_2\},$$

следовательно,

$$P\{x_1\leqslant\xi<x_2\}=F(x_2)-F(x_1). \tag{7}$$

Так как $P\{x_1\leqslant\xi<x_2\}\geqslant 0$, то из равенства (7) следует, что $F(x)$ — *неубывающая функция* для всех $x\in R_1$.

2. Из определения $F(x)$ следует, что

$$0\leqslant F(x)\leqslant 1,\quad x\in R_1.$$

3. $F(x)$ непрерывна слева в каждой точке $x\in R_1$, т. е.

$$F(x)=F(x-0)\equiv\lim_{x_k\uparrow x}F(x_k), \tag{8}$$

где последнее равенство — равенство по определению. Действительно, пусть $\{x_k\}$ — любая последовательность, стремящаяся к x слева, $x_k\uparrow x$, т. е. $x_1<x_2<\ldots<x_n<\ldots<x$ и $\lim\limits_{k\to\infty}x_k=x$. Событие $\xi<x$ можно представить в виде

$$\{\xi<x\}=\{\xi<x_1\}\cup\{\xi<x_2\}\cup\ldots=\{\xi<x_1\}+\{x_1\leqslant\xi<x_2\}+\ldots.$$

В силу σ-аддитивности вероятности и равенства (7) отсюда следует, что

$$F(x)=P\{\xi<x\}=P\{\xi<x_1\}+P\{x_1\leqslant\xi<x_2\}+\ldots=$$
$$=F(x_1)+[F(x_2)-F(x_1)]+\ldots+[F(x_{k+1})-$$
$$-F(x_k)]+\ldots=\lim_{k\to\infty}F(x_k),$$

и (8) доказано.

4. Правое предельное значение $F(x)$ в точке x равно $P\{\xi\leqslant x\}$, т. е.

$$F(x+0)\equiv\lim_{x_k\downarrow x}F(x_k)=P\{\xi\leqslant x\}. \tag{9}$$

Действительно, пусть $\{x_k\}$ — любая последовательность, стремящаяся к x справа, $x_k\downarrow x$, т. е. $x_1>x_2>\ldots>x_n>\ldots>x$ и $\lim\limits_{k\to\infty}x_k=x$. Тогда, снова на основании (7) и σ-аддитивности вероятности

$$\lim_{k\to\infty}F(x_k)=F(x_1)-[F(x_1)-F(x_2)]-\ldots-[F(x_k)-F(x_{k+1})]-$$
$$-\ldots=P\{\xi<x_1\}-[P\{x_2\leqslant\xi<x_1\}+\ldots+$$
$$+P\{x_{k+1}\leqslant\xi<x_k\}+\ldots]=P\{\xi<x_1\}-$$
$$-P\{x<\xi<x_1\}=P\{\xi\leqslant x\},$$

что и доказывает (9).

5. Справедливы следующие соотношения ($x_1\leqslant x_2$ — любые):

$$P\{x_1\leqslant\xi\leqslant x_2\}=F(x_2+0)-F(x_1), \tag{10}$$

так как

$$\{\xi<x_1\}+\{x_1\leqslant\xi\leqslant x_2\}=\{\xi\leqslant x_2\}.$$

В частности, если $x_2=x_1=x$,

$$P\{\xi=x\}=F(x+0)-F(x). \tag{11}$$

$$P\{x_1<\xi\leqslant x_2\}=F(x_2+0)-F(x_1+0), \tag{12}$$

так как

$$\{\xi\leqslant x_1\}+\{x_1<\xi\leqslant x_2\}=\{\xi\leqslant x_2\}.$$

И

$$P\{x_1<\xi<x_2\}=F(x_2)-F(x_1+0), \tag{13}$$

так как

$$\{\xi\leqslant x_1\}+\{x_1<\xi<x_2\}=\{\xi<x_2\}.$$

6. Ввиду неубывания $F(x)$ положим

$$F(-\infty)=\lim_{n\to\infty}F(-n),\quad F(+\infty)=\lim_{n\to\infty}F(n).$$

Тогда

$$F(-\infty)=0,\quad F(+\infty)=1. \tag{14}$$

В самом деле,

$$F(+\infty)=\lim_{n\to\infty}F(n)=F(0)+[F(1)-F(0)]+\ldots+$$
$$+[F(n+1)-F(n)]+\ldots=P\{\xi<0\}+P\{0\leqslant\xi<1\}+$$
$$+\ldots+P\{n\leqslant\xi<(n+1)\}+\ldots=P\{\xi<+\infty\}=1.$$

Аналогично $F(-\infty)=0$.

Таким образом, каждая функция распределения не убывает, непрерывна слева и удовлетворяет условиям (14). Верно и обратное утверждение: если $F(x)$, $-\infty<x<\infty$, удовлетворяет перечисленным условиям, то она может рассматриваться как функция распределения некоторой случайной величины.

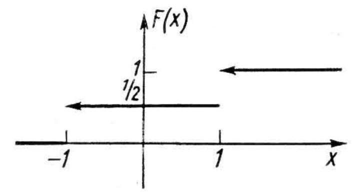


Рис. 10

Однако существует сколько угодно случайных величин с данной функцией распределения. Например, функция распределения случайной величины ξ, принимающей значения -1 и $+1$ с вероятностью 1/2, совпадает с функцией распределения $\eta=-\xi$, хотя $\xi\neq\eta$ — с вероятностью 1. Отметим также, что из неубывания произвольной функции распределения $F(x)$ и неравенства $0\leqslant F(x)\leqslant 1$ следует, что $F(x)$ имеет не более счетного числа скачков. Действительно, можно пронумеровать все скачки следующим образом: сначала все скачки, большие 1/2 (их $\leqslant 1$), затем скачки, большие 1/3 (их $\leqslant 2$), и т. д. Таким образом, $F(x)$ непрерывна всюду на R_1 за исключением не более чем счетного множества значений.

2°. Дискретные и непрерывные случайные величины

Рассмотренные выше случайные величины, распределенные нормально и по биномиальному закону, являются характерными примерами двух основных классов случайных величин: непрерывных и дискретных.

Определение 3. Случайная величина ξ называется **дискретной**, если множество ее значений конечно или счетно.

Для полной вероятностной характеристики дискретной случайной величины, принимающей значения $x_1, x_2, \ldots$, достаточно задать вероятности $p_k = P\{\xi = x_k\}$. Зная значения x_k и p_k $k=1, 2, \ldots$, можно записать функцию распределения $F(x)$ дискретной случайной величины ξ в виде

$$F(x) = \sum_{k: x_k < x} p_k. \tag{15}$$

Очевидно, $F(x)$ не зависит от способа нумерации значений случайной величины ξ. Таким образом, функция распределения любой дискретной случайной величины разрывна, возрастает скачками в точках $x = x_k$, величина скачка равна

$$F(x_k+0) - F(x_k) = p_k. \tag{16}$$

Примеры 1 и 2 в п. 1° служат примерами дискретных случайных величин.

При изучении дискретной случайной величины можно не обращаться к исходному вероятностному пространству. Достаточно рассматривать новое вероятностное пространство $(\overline{\Omega}, \overline{\mathscr{F}}, \overline{P})$, порожденное этой случайной величиной, в котором $\overline{\Omega} = \{x_1, x_2, \ldots\}$ — множество значений ξ, $\overline{\mathscr{F}}$—σ-алгебра всех подмножеств $\overline{\Omega}$ и $\overline{P}$ — вероятность на $\overline{\mathscr{F}}$, определенная для каждого одноточечного подмножества $\overline{\Omega}$ равенством

$$\overline{P}(\{x_k\}) = p_k = P\{\xi = x_k\} = P(\{\omega : \xi(\omega) = x_k\}). \tag{17}$$

С помощью (17) можно определить вероятность любого множества $A \in \overline{\mathscr{F}}$ (т. е. любого подмножества $\overline{\Omega}$):

$$\overline{P}(A) = \overline{P}\{\xi \in A\} = \sum_{k: x_k \in A} p_k = P\{\omega : \bigcup_{x_k \in A} (\xi(\omega) = x_k)\}. \tag{18}$$

Функция распределения $F(x)$ случайной величины ξ задается равенством (15), а поскольку, обратно, $p_k = \overline{P}(\{x_k\}) = F(x_k+0) - F(x_k)$, то функция распределения полностью определяет вероятность $\overline{P}$ на $\overline{\mathscr{F}}$. σ-аддитивность вероятности $\overline{P}$, как и вообще корректность определения (18), следует из леммы о суммировании по блокам (см. § 3).

Определение 4. Случайная величина ξ называется **непрерывной** (или абсолютно непрерывной), если ее функция распределения представима в виде

$$F(x) = \int_{-\infty}^{x} p(y)\, dy. \tag{19}$$

Функция $p(y)$, $-\infty < y < \infty$, называется **плотностью распределения вероятностей** (или **плотностью вероятности**) случайной величины ξ и далее предполагается неотрицательной и кусочно-непрерывной. Плотность $p(y)$ полностью определяет функцию распределения $F(x)$, а в точках непрерывности $p(y)$ определяется по функции распределения, так как в этих точках

$$p(x) = \frac{dF(x)}{dx},$$

и, таким образом, в этих точках свойство неотрицательности плотности является следствием неубывания $F(x)$.

Для любых $x_1 < x_2$

$$P\{x_1 \leqslant \xi < x_2\} = F(x_2) - F(x_1) = \int_{x_1}^{x_2} p(y)\, dy. \tag{20}$$

Если, в частности, плотность $p(y)$ непрерывна на $[x, x+\Delta x]$, $\Delta x > 0$, то согласно (20) и теореме о среднем для интеграла

$$P\{x \leqslant \xi < x + \Delta x\} = p(x)\Delta x + o(\Delta x). \tag{21}$$

Заметим, что для непрерывной случайной величины

$$P\{\xi = x\} = F(x+0) - F(x) = 0, \tag{22}$$

так как $F(x)$ непрерывна, т. е. вероятность того, что ξ примет любое фиксированное значение из R_1, равна нулю. Поэтому в формулах (20) и (21) строгие неравенства в событиях могут быть заменены на нестрогие. Очевидно, что

$$\int_{-\infty}^{\infty} p(y)\, dy = F(+\infty) = 1. \tag{23}$$

Примером непрерывной случайной величины является нормальная $N(\mu, \sigma^2)$ случайная величина с плотностью

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{24}$$

(пример 3 в п. 1°).

С непрерывной случайной величиной также связано новое вероятностное пространство $(\overline{\Omega}, \overline{\mathscr{F}}, \overline{P})$, где $\overline{\Omega}$ — действительная прямая R_1, $\overline{\mathscr{F}}$—σ-алгебра борелевских множеств на прямой. Для каждого интервала (x_1, x_2) вероятность $\bar{P}$ определяется по формуле (20). В теории меры доказывается, что тем самым вероятность $\bar{P}$ определяется и для всякого события $A \in \bar{\mathcal{F}}$, т. е. для любого борелевского множества на прямой. Отсюда следует, что и в этом случае вероятность полностью определяется функцией распределения.

В дальнейшем мы будем считать, что случайная величина ξ задана, если задано вероятностное пространство $(\bar{\Omega}, \bar{\mathcal{F}}, \bar{P})$, или, иначе, если задана функция распределения $F(x)$, когда речь идет о дискретных или непрерывных случайных величинах. Следует указать, что, конечно, существуют случайные величины, которые не являются ни дискретными, ни непрерывными, ни их комбинацией (мы имеем в виду случайные величины с функцией распределения вида:

$$F(x) = q_1 \sum_{k: x_k < x} p_k + q_2 \int_{-\infty}^{x} p(y)\, dy,$$

где

$$q_1, q_2 > 0, \; q_1 + q_2 = 1, \; \sum_k p_k = 1, \; \int_{-\infty}^{\infty} p(y)\, dy = 1).$$

Это так называемые сингулярные случайные величины. Простое приведение примера такой случайной величины требует специальной конструкции типа известной кривой Кантора (см. [1]). Сингулярные распределения представляют собой некоторую «экзотику» и в реальных задачах практически не встречаются. Мы исключаем такие случайные величины из дальнейшего изучения.

З а м е ч а н и е. Все сказанное о функциях распределения автоматически переносится на случай условных вероятностей. Если $P(B) > 0$, то $F(x|B) = P\{\xi < x|B\}$ называется **условной функцией распределения** случайной величины ξ. Она обладает всеми указанными выше свойствами функций распределения.

3°. Векторные (или многомерные) случайные величины

В задачах со случайным исходом обычно приходится учитывать взаимодействие различных случайных факторов. Это естественным образом приводит к рассмотрению многомерных случайных величин.

Определение 5. Пусть $(\Omega, \mathcal{F}, P)$ — вероятностное пространство, и $\xi_1(\omega), \xi_2(\omega), \ldots, \xi_n(\omega)$ — случайные величины, определенные на Ω. Вектор $\xi(\omega) = (\xi_1(\omega), \ldots, \xi_n(\omega))$ называется **случайным вектором**, или **n-мерной случайной величиной**, а $\xi_j(\omega)$, $j = 1, 2, \ldots, n$, называются **координатами**, или **компонентами**, случайного вектора ξ.

Поскольку все $\xi_j(\omega)$, $j = 1, 2, \ldots, n$, заданы на одном и том же вероятностном пространстве, а $\mathcal{F}$ замкнуто относительно взятия произведения конечного числа событий, то множество $\{\omega : \xi_1(\omega) < x_1, \ldots, \xi_n(\omega) < x_n\} \in \mathcal{F}$ для любого набора действительных чисел $x_1, \ldots, x_n$. Таким образом, оправдано следующее

Определение 6. Функция $F(x_1, x_2, \ldots, x_n) = P\{\xi_1 < x_1, \xi_2 < x_2, \ldots, \xi_n < x_n\}$, $x_j \in R_1$, $j = 1, 2, \ldots, n$, называется **n-мерной функцией распределения** случайной величины $\xi = (\xi_1, \xi_2, \ldots, \xi_n)$.

Ради наглядности и краткости будем рассматривать двумерные случайные величины. Геометрически двумерная функция распределения $F(x, y)$, равная

$$F(x, y) = P\{\xi < x, \eta < y\}, \tag{25}$$

задает вероятность попадания точки (ξ, η) в бесконечный прямоугольник $\xi < x$, $\eta < y$ (заштрихованная часть на рис. 11).

Коротко перечислим основные свойства двумерной функции распределения $F(x, y)$:

1) $F(x, y)$ не убывает по x и по y,
2) $F(x, y)$ непрерывна слева по каждому аргументу,
3) $F(\infty, \infty) = 1$, $F(-\infty, y) = 0$, $F(x, -\infty) = 0$, где по определению

$$F(\infty, y) = \lim_{n\to\infty} F(n, y), \; F(\infty, \infty) = \lim_{n,m\to\infty} F(n, m),$$

$$F(x, -\infty) = \lim_{n\to\infty} F(x, -n);$$

все эти пределы существуют ввиду неубывания $F(x, y)$ по x и y (доказательства свойств 1—3 аналогичны проведенным в пункте 2°),

4) $$P\{x_1 \leqslant \xi < x_2, y_1 \leqslant \eta < y_2\} = F(x_2, y_2) - F(x_1, y_2) - F(x_2, y_1) + F(x_1, y_1) \tag{26}$$

(доказательство сразу следует из определения (25) (см. рис. 12)).

5) Пользуясь двумерной функцией распределения, можно найти функции распределения координат ξ и η (так называемые маргинальные распределения):

$$F_\xi(x) = F(x, \infty), \; F_\eta(y) = F(\infty, y). \tag{27}$$

В самом деле, имеем

$$\{\xi < x\} = \sum_{k=-\infty}^{\infty} \{\xi < x, k \leqslant \eta < k+1\}.$$

Поскольку события под знаком суммы попарно несовместны, то в силу σ-аддитивности вероятности

$$F_\xi(x) = P\{\xi < x\} = \sum_{k=-\infty}^{\infty} P\{\xi < x, k \leqslant \eta < k+1\} =$$
$$= \sum_{k=-\infty}^{\infty} [F(x, k+1) - F(x, k)] =$$
$$= \lim_{N_1, N_2 \to \infty} \sum_{k=-N_1}^{N_2} [F(x, k+1) - F(x, k)] =$$
$$= \lim_{N_1, N_2 \to \infty} [F(x, N_2+1) - F(x, -N_1)] =$$
$$= F(x, \infty) - F(x, -\infty) = F(x, \infty),$$

и первая формула в (27) обоснована, другая — аналогично.

Определение 6. Случайный вектор называется **дискретным**, если каждая его координата — дискретная случайная

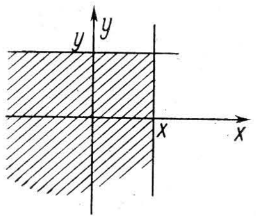


Рис. 11

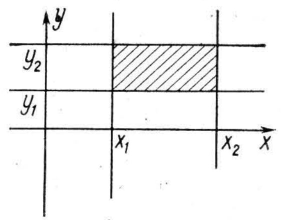


Рис. 12

величина, и **непрерывным**, если существует кусочно-непрерывная* неотрицательная функция $p(x, y)$, $x, y \in R_1$, такая, что для любых x и y

$$F(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} p(z_1, z_2)\, dz_1 dz_2. \tag{28}$$

Функция $p(x, y)$ называется **плотностью вероятности** случайного вектора (ξ, η).

Плотность вероятности обладает следующими свойствами.

1) В точках непрерывности $p(x, y)$ справедливо равенство

$$p(x, y) = \frac{\partial^2 F}{\partial x \partial y},$$

* Определение кусочно-непрерывной функции для многомерного случая см. в [5].

и, таким образом, в этих точках тот факт, что $p(x, y) \geqslant 0$, следует из неубывания $F(x, y)$ по каждой переменной x и y.

2) Для любой квадрируемой области $D \subset R_2$ имеем

$$P\{(\xi, \eta) \in D\} = \iint_D p(x, y)\, dxdy. \tag{29}$$

В самом деле, если D представляет собой произвольный прямоугольник $\Pi = \{x_1 \leqslant \xi < x_2,\ y_1 \leqslant \eta < y_2\}$, то (29) есть очевидное следствие равенств (26) и (28). В общем случае мы, как обычно, аппроксимируем область D суммой входящих и выходящих квадратов, для них справедливость (29) доказана только что, а затем устремим сторону квадратов к нулю. В силу предположенной квадрируемости области D и кусочной непрерывности $p(x, y)$ этот предельный переход приведет нас к (29).

В частности, если $p(x, y)$ непрерывна при $x_1 \leqslant x \leqslant x_1 + \Delta x$, $y_1 \leqslant y \leqslant y_1 + \Delta y$, то с помощью равенства (29) и теоремы о среднем для интеграла, получим

$$P\{x_1 \leqslant \xi \leqslant x_1 + \Delta x,\ y_1 \leqslant \eta \leqslant y_1 + \Delta y\} =$$
$$= p(x_1, y_1)\, \Delta x \Delta y + o(\Delta x \Delta y). \tag{30}$$

3) Вследствие равенства $F(\infty, \infty) = 1$ имеем

$$\iint_{-\infty}^{\infty} p(x, y)\, dxdy = 1. \tag{31}$$

4) На основании равенств (28) и (27) заключаем, что если двумерная случайная величина (ξ, η) имеет плотность, то и каждая ее компонента имеет плотность, причем

$$p_\xi(x) = \int_{-\infty}^{\infty} p(x, y)\, dy, \quad p_\eta(y) = \int_{-\infty}^{\infty} p(x, y)\, dx. \tag{32}$$

Примеры

1. Случайный вектор (ξ, η) называется **равномерно распределенным** в области D, если

$$p(x, y) = \begin{cases} 1/\mu(D), & (x, y) \in D, \\ 0, & (x, y) \overline{\in} D, \end{cases}$$

где $\mu(D)$ — площадь области D. Пусть D^* — квадрируемая область на плоскости. Тогда

$$P\{(\xi, \eta) \in D^*\} = \frac{\mu(D^* \cap D)}{\mu(D)}.$$

Это, по существу, задача на геометрические вероятности (см. § 1).

3*

2. Случайный вектор $(\xi_1, \ldots, \xi_n)$ называется **нормально распределенным**, если его плотность равна

$$p(x_1, \ldots, x_n) = \left\{\frac{\det A}{(2\pi)^n}\right\}^{1/2} \exp\left\{-\frac{1}{2}\sum_{i,j=1}^{n} a_{ij}(x_i - \mu_i)(x_j - \mu_j)\right\}, \tag{33}$$

где $A = \|a_{ij}\|$ — положительно определенная матрица (обратная к матрице ковариаций случайных величин $\xi_1, \ldots, \xi_n$, см. ниже § 9). Можно показать непосредственно интегрируя (33) и пользуясь равенствами (32) (или их аналогами, если $n > 2$), что каждая координата ξ_j, $j = 1, \ldots, n$, нормально распределенного вектора также имеет нормальное распределение.

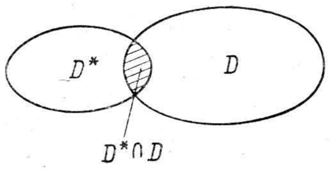


Рис. 13

В двумерном случае плотность нормального закона (33) имеет вид

$$p(x, y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-r^2}} \exp\left\{-\frac{1}{2(1-r^2)}\left[\frac{(x-a)^2}{\sigma_1^2} - \frac{2r(x-a)(y-b)}{\sigma_1\sigma_2} + \frac{(y-b)^2}{\sigma_2^2}\right]\right\}, \tag{34}$$

где

$$\sigma_1^2 = D\xi, \quad \sigma_2^2 = D\eta, \quad r = \frac{\operatorname{cov}(\xi, \eta)}{\sigma_1\sigma_2},$$

см. ниже, § 9.

Рассмотрим в заключение этого пункта понятие **плотности условного распределения.** Пусть плотность $p(x, y)$ случайного вектора (ξ, η) непрерывна. Обозначим через B событие: $B = \{y \leqslant \eta \leqslant y + \Delta y\}$; его вероятность равна $P(B) = P\{y \leqslant \eta \leqslant y + \Delta y\} = \int_y^{y+\Delta y} p_\eta(z)\,dz$, где согласно (32),

$$p_\eta(z) = \int_{-\infty}^{\infty} p(x, z)\,dx.$$

Далее, имеем

$$P\{\xi < x,\ y \leqslant \eta \leqslant y + \Delta y\} = \int_{-\infty}^{x}\int_{y}^{y+\Delta y} p(t, z)\,dt\,dz,$$

а следовательно, в силу определения условной вероятности (см. § 4), если $P(B) > 0$,

$$F_\xi(x \mid B) \equiv P\{\xi < x \mid B\} = \frac{P\{\xi < x,\ y \leqslant \eta \leqslant y + \Delta y\}}{P(B)} =$$

$$= \frac{\int_{-\infty}^{x}\int_{y}^{y+\Delta y} p(t, z)\,dt\,dz}{\int_{y}^{y+\Delta y} p_\eta(z)\,dz}. \tag{35}$$

Продифференцируем (35) по x, а затем устремим $\Delta y \to 0$ и воспользуемся теоремой о среднем для интеграла по промежутку $[y, y + \Delta y]$; в результате найдем:

$$p_\xi(x \mid y) \equiv \frac{dF_\xi(x \mid \eta = y)}{dx} = \frac{p(x, y)}{p_\eta(y)}, \quad p_\eta(y) \neq 0. \tag{36}$$

Функция $p_\xi(x \mid y)$ называется **плотностью вероятности условного распределения** ξ при условии $\eta = y$. Равенство можно переписать в виде

$$p(x, y) = p_\eta(y)\,p_\xi(x \mid y), \tag{37}$$

напоминающем по форме теорему умножения вероятностей. Если мы в (35) устремим $\Delta y \to 0$ и воспользуемся определением (36), то получим

$$F_\xi(x \mid \eta = y) = P\{\xi < x \mid \eta = y\} = \int_{-\infty}^{x} p_\xi(t \mid y)\,dt. \tag{38}$$

Интегрируя равенство (37) по переменной y, найдем в силу (32)

$$p_\xi(x) = \int_{-\infty}^{\infty} p_\eta(y)\,p_\xi(x \mid y)\,dy, \tag{39}$$

это непрерывный аналог формулы полной вероятности.

Соответствующие рассуждения справедливы и для дискректного случайного вектора (ξ, η): пусть ξ принимает значения x_i, $i = 1, 2, \ldots$, а η принимает значения y_j, $j = 1, 2, \ldots$ и $P_{ij} = P\{\xi = x_i,\ \eta = y_j\}$. Тогда аналогом формул (32) являются равенства:

$$p_i = P\{\xi = x_i\} = \sum_{j=1}^{\infty} p_{ij}, \quad q_j = P\{\eta = y_j\} = \sum_{i=1}^{\infty} p_{ij}, \quad i, j = 1, 2, \ldots. \tag{40}$$

Условные распределения вероятностей определяются следующим образом:

$$P_{i \mid j} = P\{\xi = x_i \mid \eta = y_j\} = \frac{P\{\xi = x_i, \eta = y_j\}}{P\{\eta = y_j\}} = \frac{P_{ij}}{q_j}, \tag{41}$$

$$Q_{j \mid i} = P\{\eta = y_j \mid \xi = x_i\} = \frac{P_{ij}}{p_i}, \quad i, j = 1, 2, \ldots.$$

И как следствие (40) и (41)

$$p_i = \sum_{j=1}^{\infty} q_j P_{i|j}, \quad q_j = \sum_{i=1}^{\infty} p_i Q_{j|i}. \tag{42}$$

4°. Независимость случайных величин

В § 3 мы указали конструкцию, связанную с экспериментом, повторяемым в неизменных условиях, в которой естественным образом возникают независимые события. Если в каждом таком эксперименте мы измеряем какую-либо случайную величину, например координату частицы, совершающей броуновское движение, то ее значения в различных экспериментах не зависят друг от друга. Понятие независимости случайных величин является одним из важнейших в теории вероятностей.

Определение 7. Случайные величины $\xi_1, \ldots, \xi_n$ называются **независимыми** (в совокупности), если для любых $x_1, \ldots, x_n$ события $\{\xi_1 < x_1\}, \ldots, \{\xi_n < x_n\}$ независимы в совокупности, т. е.

$$P(\{\xi_1 < x_1\} \cap \ldots \cap \{\xi_n < x_n\}) = P\{\xi_1 < x_1\} \ldots P\{\xi_n < x_n\}, \tag{43}$$

или, что то же самое,

$$F(x_1, \ldots, x_n) = F_{\xi_1}(x_1) \ldots F_{\xi_n}(x_n). \tag{44}$$

Для независимых случайных величин ξ и η имеем при любых $x_1 < x_2$ и $y_1 < y_2$

$$P\{x_1 \leqslant \xi < x_2,\ y_1 \leqslant \eta < y_2\} = P\{\xi < x_2,\ \eta < y_2\} -$$

$$- P\{\xi < x_1,\ \eta < y_2\} - P\{\xi < x_2,\ \eta < y_1\} + P\{\xi < x_1,\ \eta < y_1\} =$$

$$= P\{\xi < x_2\}[P\{\eta < y_2\} - P\{\eta < y_1\}] -$$

$$- P\{\xi < x_1\}[P\{\eta < y_2\} - P\{\eta < y_1\}] =$$

$$= P\{x_1 \leqslant \xi < x_2\} P\{y_1 \leqslant \eta < y_2\}.$$

Итак,

$$P\{x_1 \leqslant \xi < x_2,\ y_1 \leqslant \eta < y_2\} = P\{x_1 \leqslant \xi < x_2\} P\{y_1 \leqslant \eta < y_2\}, \tag{45}$$

$x_1 < x_2$, $y_1 < y_2$ — любые.

Аналогичный результат справедлив, разумеется, и для произвольного числа $\xi_1, \ldots, \xi_n$ независимых случайных величин;

$$P\left(\bigcap_{i=1}^{n} \left\{\overset{i}{x_1} \leqslant \xi_i < \overset{i}{x_2}\right\}\right) = \prod_{i=1}^{n} P\left\{\overset{i}{x_1} \leqslant \xi_i < \overset{i}{x_2}\right\}.$$

Пусть, обратно, выполнено равенство (45) при любых $x_1 < x_2$ и $y_1 < y_2$. Устремляя в (45) $x_1, y_1 \to -\infty$, получим

$$P\{\xi < x_2,\ \eta < y_2\} = P\{\xi < x_2\} P\{\eta < y_2\}.$$

Это равенство ввиду произвольности x_2 и y_2 означает независимость случайных величин ξ и η. Таким образом, условие (45) необходимо и достаточно для независимости случайных величин ξ и η.

Пусть теперь ξ и η — дискретные случайные величины, $x_1 < x_2 < \ldots < x_n < \ldots$ — значения, принимаемые ξ, и $y_1 < y_2 < \ldots < y_n < \ldots$ — значения η. Тогда

$$P\{x_k \leqslant \xi < x_{k+1}\} = P\{\xi = x_k\},\ P\{y_j \leqslant \eta < y_{j+1}\} =$$

$$= P\{\eta = y_j\},\ k, j = 1, 2, \ldots.$$

Подставляя эти выражения в (45), приходим к выводу, что условие

$$P\{\xi = x_k,\ \eta = y_j\} = P\{\xi = x_k\} P\{\eta = y_j\},\ k, j = 1, 2, \ldots, \tag{46}$$

является необходимым и достаточным условием независимости дискретных случайных величин ξ и η.

Равенство (44) служит определением независимости случайных величин $\xi_1, \ldots, \xi_n$. Однако условие независимости $\xi_1, \ldots, \xi_n$ может быть сформулировано в форме (44), даже если при этом не предполагается, что стоящие справа множители являются маргинальными функциями распределения. Пусть, например, для всех x, y функция распределения имеет вид: $F(x, y) = F_1(x) F_2(y)$, причем $\lim\limits_{x\to\infty} F_1(x) = 1$. Тогда ξ и η независимы и $F_1(x) = F_\xi(x)$, $F_2(y) = F_\eta(y)$. Действительно, $F_\eta(y) = F(\infty, y) = F_1(\infty) F_2(y) = F_2(y)$. Таким образом, $F_\eta(y) = F_2(y)$, поэтому $F_2(\infty) = 1$, а отсюда, в свою очередь, следует, что $F_1(x) = F_\xi(x)$, и ξ и η независимы в силу (44).

Если независимые случайные величины ξ и η имеют соответственно плотности вероятности $p_\xi(x)$ и $p_\eta(y)$, то вектор (ξ, η) имеет плотность $p(x, y) = p_\xi(x) p_\eta(y)$. Это равенство является следствием (44) в точках (x, y), в которых $p_\xi(x)$ и $p_\eta(y)$ непрерывны, в остальных точках оно является определением $p(x, y)$.

Важным является обратное утверждение: если плотность $p(x, y)$ случайного вектора (ξ, η) равна произведению плотностей координат, $p(x, y) = p_\xi(x) p_\eta(y)$, то ξ и η независимы, ибо в этом случае, очевидно, выполнено условие (44). Это свойство доставляет удобный критерий независимости непрерывных случайных величин (см. ниже, пример 4 в п. 5°).

Пример. Пусть $\xi \in N(\mu_1, \sigma_1^2)$, $\eta \in N(\mu_2, \sigma_2^2)$ и независимы. Запись $\xi \in N(\mu_1, \sigma_1^2)$ обозначает, что ξ — нормальная $N(\mu_1, \sigma_1^2)$ случайная величина. Плотность случайного вектора (ξ, η) равна произведению маргинальных плотностей:

$$p(x, y) = \frac{1}{2\pi\sigma_1\sigma_2} \exp\left\{-\frac{(x-\mu_1)^2}{2\sigma_1^2} - \frac{(y-\mu_2)^2}{2\sigma_2^2}\right\}.$$

Это равенство — частный случай (34), так что (ξ, η) — нормальный случайный вектор,

$$A=\begin{pmatrix}1/\sigma_1^2 & 0\\ 0 & 1/\sigma_2^2\end{pmatrix},\ \det A=\frac{1}{\sigma_1^2\sigma_2^2},\ r=0.$$

5°. Функции от случайных величин

В этом пункте мы рассмотрим функции от случайных величин. Снова ради наглядности ограничиваемся случаем двух случайных величин. Итак, пусть на вероятностном пространстве $(\Omega, \mathscr{F}, P)$ определены две случайные величины $\xi_1=\xi_1(\omega)$ и $\xi_2=\xi_2(\omega)$, $\omega\in\Omega$, и пусть $F(x_1, x_2)$ — функция распределения случайного вектора (ξ_1, ξ_2). Рассмотрим некоторые функции от случайных величин ξ_1 и ξ_2, т. е. новые случайные величины η_1 и η_2, связанные функциональными зависимостями с ξ_1 и ξ_2,

$$\begin{gathered}\eta_1=f_1(\xi_1, \xi_2)=f_1(\xi_1(\omega), \xi_2(\omega))=\eta_1(\omega),\\ \eta_2=f_2(\xi_1, \xi_2)=f_2(\xi_1(\omega), \xi_2(\omega))=\eta_2(\omega).\end{gathered}\tag{47}$$

Здесь предполагается, что f_1 и f_2 таковы, что η_1 и η_2 вновь являются случайными величинами, определенными на том же вероятностном пространстве $(\Omega, \mathscr{F}, P)$ (количество функций η_i может быть, разумеется, любым). Тем самым, η_1 и η_2 являются сложными функциями ω, заданными на Ω.

Основная задача, возникающая в этой ситуации, состоит в том, чтобы, зная функцию распределения $F(x_1, x_2)$ случайного вектора (ξ_1, ξ_2) и функции f_1 и f_2, найти функцию распределения $\Phi(y_1, y_2)$ случайного вектора (η_1, η_2). Для рассматриваемых нами дискретных и непрерывных случайных величин указанная задача решается без труда. В самом деле, имеем

$$\begin{aligned}\Phi(y_1, y_2)&=P\{\eta_1<y_1,\ \eta_2<y_2\}=\\ &=P\{f_1(\xi_1, \xi_2)<y_1,\ f_2(\xi_1, \xi_2)<y_2\},\end{aligned}\tag{48}$$

и пусть

1) (ξ_1, ξ_2) — дискретный случайный вектор, $\xi_1=x_{1i}$, $\xi_2=x_{2j}$ — значения ξ_1 и ξ_2, $P_{ij}=P\{\xi_1=x_{1i},\ \xi_2=x_{2j}\}$ — вероятности этих значений, $i, j=1, 2, \ldots$, тогда согласно (48)

$$\Phi(y_1, y_2)=\sum_{i,j\in\mathfrak{A}} p_{ij},\tag{49}$$

где суммирование распространено на множество индексов $(i, j)\in\mathfrak{A}$,

$$\mathfrak{A}=\{(i, j): f_1(x_{1i}, x_{2j})<y_1,\ f_2(x_{1i}, x_{2j})<y_2\};$$

2) (ξ_1, ξ_2) — непрерывный случайный вектор, $p(x_1, x_2)$ — его плотность вероятности, тогда согласно (48) и (29)

$$\Phi(y_1, y_2)=\iint_D p(x_1, x_2)\,dx_1\,dx_2,\tag{50}$$

где область D определяется условием: $D=\{(x_1, x_2): f_1(x_1, x_2)<y_1,\ f_2(x_1, x_2)<y_2\}$ (считаем, что f_1 и f_2 таковы, что D квадрируема).

Найдем в качестве примера функцию распределения суммы $\eta=\xi_1+\xi_2$, если задана плотность вероятности $p(x_1, x_2)$ случайного вектора (ξ_1, ξ_2). Имеем

$$\begin{aligned}\Phi(y)&=P\{\eta<y\}=P\{\xi_1+\xi_2<y\}=\iint\limits_{x_1+x_2<y} p(x_1, x_2)\,dx_1\,dx_2=\\ &=\int_{-\infty}^{\infty}dx_1\int_{-\infty}^{y-x_1}p(x_1, x_2)\,dx_2=\int_{-\infty}^{\infty}dx_2\int_{-\infty}^{y-x_2}p(x_1, x_2)\,dx_1.\end{aligned}$$

Сделав в интеграле замену переменных $x_1+x_2=z$, $x_1=x_1$, якобиан отображения $(x_1, x_2)\to(x_1, z)$ равен 1, получим

$$\Phi(y)=\int_{-\infty}^{y}dz\int_{-\infty}^{\infty}p(x_1, z-x_1)\,dx_1.\tag{51}$$

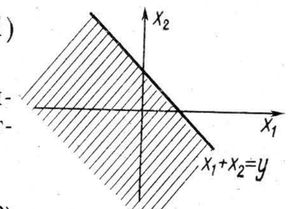


Рис. 14

Из формулы (51) следует, что случайная величина η имеет плотность вероятности

$$p_\eta(y)=\Phi'(y)=\int_{-\infty}^{\infty}p(x_1, y-x_1)\,dx_1.\tag{52}$$

Таким образом, если двумерное распределение слагаемых ξ_1 и ξ_2 имеет плотность $p(x_1, x_2)$, то и их сумма $\eta=\xi_1+\xi_2$ также имеет плотность, определяемую равенством (52).

Если случайные величины ξ_1 и ξ_2 независимы, то $p(x_1, x_2)=p_{\xi_1}(x_1)\,p_{\xi_2}(x_2)$ и (52) записывается в виде свертки функций p_1 и p_2:

$$\begin{aligned}p_\eta(y)&=\int_{-\infty}^{\infty}p_{\xi_1}(x)\,p_{\xi_2}(y-x)\,dx=\\ &=\int_{-\infty}^{\infty}p_{\xi_2}(x)\,p_{\xi_1}(y-x)\,dx=p_{\xi_1}*p_{\xi_2}.\end{aligned}\tag{53}$$

Определение закона распределения суммы по законам распределения независимых слагаемых называется **композицией** законов распределения слагаемых.

Докажем теперь теорему, которая будет часто использоваться в дальнейшем.

Теорема 1. Пусть ξ_1 и ξ_2 независимые случайные величины, а $f_1(x)$ и $f_2(x)$, $x \in R_1$, произвольные функции, такие, что $\eta_1 = f_1(\xi)$ и $\eta_2 = f_2(\xi)$ также случайные величины. Тогда η_1 и η_2 независимы, т. е. функции от независимых случайных величин являются независимыми случайными величинами.

Доказательство. Проведем доказательство для дискретных случайных величин. Пусть ξ_1 принимает значения $x_1, x_2, \ldots$ а ξ_2 — значения $y_1, y_2, \ldots$. Тогда случайные величины $\eta_1 = f_1(\xi_1)$ и $\eta_2 = f_2(\xi_2)$ также дискретны, пусть λ и μ — любые фиксированные значения η_1 и η_2. Имеем на основании (46)

$$P\{f_1(\xi_1) = \lambda,\ f_2(\xi_2) = \mu\} = \sum_{(k,l): f_1(x_k)=\lambda,\ f_2(y_l)=\mu} P\{\xi_1 = x_k,\ \xi_2 = y_l\} =$$

$$= \sum_{(k,l): f_1(x_k)=\lambda,\ f_2(y_l)=\mu} P\{\xi_1 = x_k\}\, P\{\xi_2 = y_l\} =$$

$$= \sum_{k: f_1(x_k)=\lambda} P\{\xi_1 = x_k\} \sum_{l: f_2(y_l)=\mu} P\{\xi_2 = y_l\} =$$

$$= P\{f_1(\xi_1) = \lambda\}\, P\{f_2(\xi_2) = \mu\},$$

что ввиду (46) и произвольности λ и μ означает независимость случайных величин $\eta_1 = f_1(\xi_1)$ и $\eta_2 = f_2(\xi_2)$. ▲

В заключение этого параграфа приведем несколько примеров использования установленных выше результатов.

Примеры

1. Пусть $\xi = (\xi_1, \ldots, \xi_n)$ и $\eta = (\eta_1, \ldots, \eta_n)$ — n-мерные случайные величины, причем $\eta = A\xi$ или подробно $\eta_i = \sum_{j=1}^{n} a_{ij}\xi_j$, $i = 1, \ldots, n$, где $A = \|a_{ij}\|$ — невырожденная квадратная $n \times \times n$ матрица. Пусть $p_\xi(x) = p_\xi(x_1, \ldots, x_n)$ — плотность распределения случайного вектора ξ, а D — произвольная квадрируемая область в n-мерном евклидовом пространстве R_n. Тогда по правилу замены переменных в кратном интеграле и в силу (29) получим

$$\int_D p_\eta(y)\,dy = P\{\eta \in D\} = P\{A\xi \in D\} = P\{\xi \in A^{-1}D\} =$$

$$= \int_{A^{-1}D} p_\xi(x)\,dx = \int_D p_\xi(A^{-1}y)\,|\det A^{-1}|\,dy \tag{54}$$

(здесь под $A^{-1}D$ понимается область, каждая точка которой z имеет вид $z = A^{-1}y$, где $y \in D$).

Отсюда ввиду произвольности области D следует, что

$$p_\eta(x) = p_\xi(A^{-1}x)\,|\det A^{-1}|; \tag{55}$$

мы получили формулу, связывающую плотности случайных векторов ξ и $\eta = A\xi$.

2. На отрезок $[0, a]$ случайно и независимо друг от друга падают две частицы массы m. Найти плотность распределения координаты центра тяжести системы этих двух частиц, если координаты частиц равномерно распределены (математическая формулировка того факта, что частицы падают на $[0, a]$ «случайно») на отрезке $[0, a]$.

Пусть ξ_1 и ξ_2 — координаты частиц: случайные величины ξ_1 и ξ_2 независимы; поскольку они равномерно распределены на $[0, a]$, то плотности вероятностей равны

$$p_1(x) = p_2(x) = \begin{cases} 0, & \text{если } x < 0 \text{ и } x > a, \\ \dfrac{1}{a}, & \text{если } 0 \leqslant x \leqslant a. \end{cases} \tag{56}$$

Координата центра тяжести равна $\eta = (\xi_1 + \xi_2)/2$ (массы одинаковы). Согласно (51)

$$F_\eta(y) = P\{\eta < y\} = P\{\xi_1 + \xi_2 < 2y\} =$$

$$= \int_{-\infty}^{2y} dz \int_{-\infty}^{\infty} p_1(x)\,p_2(z - x)\,dx.$$

Дифференцируя это выражение по y, находим плотность $p_\eta(y)$

$$p_\eta(y) = 2\int_{-\infty}^{\infty} p_1(x)\,p_2(2y - x)\,dx. \tag{57}$$

Так как $p_1(x)$ отлична от нуля только при $0 \leqslant x \leqslant a$, то

$$p_\eta(y) = \frac{2}{a}\int_0^a p_2(2y - x)\,dx.$$

Введем здесь новую переменную $u = 2y - x$, тогда

$$p_\eta(y) = \frac{2}{a}\int_{2y-a}^{2y} p_2(u)\,du. \tag{58}$$

Вспомним, что согласно (56) $p_2(u)$ отлична от нуля лишь при $0 \leqslant u \leqslant a$.

Ясно, что если $2y - a > a$, т. е. $y > a$, или $2y < 0$, т. е. $y < 0$, то $p_\eta(y) = 0$ (см. (58)). Пусть теперь $0 \leqslant y \leqslant a$; если $2y - a < 0$, т. е. $y < a/2$, то интегрируем в (58) от 0 до $2y$, поэтому $p_\eta(y) = \frac{2}{a}\int_0^{2y} \frac{1}{a}\,du = \frac{4}{a^2}\,y$ при $0 \leqslant y < a/2$, если же $2y -$

$-a \geqslant 0$, т. е. $y \geqslant a/2$, то интегрируем в (58) от $2y-a$ до a, поэтому

$$p_\eta(y) = \frac{2}{a}\int_{2y-a}^{a} \frac{1}{a}\,du = \frac{4}{a^2}(a-y) \text{ при } a/2 \leqslant y \leqslant a.$$

Итак, плотность распределения $p_\eta(y)$ центра тяжести равна

$$p_\eta(y) = \begin{cases} 0, & \text{если } y < 0 \text{ или } y > a, \\ \dfrac{4}{a^2}\,y, & \text{если } 0 \leqslant y < \dfrac{a}{2}, \\ \dfrac{4}{a^2}(a-y), & \text{если } \dfrac{a}{2} \leqslant y \leqslant a. \end{cases}$$

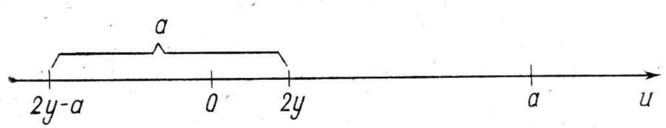


Рис. 15

3. Пусть (ξ, η) — непрерывный случайный вектор с плотностью вероятности $p(x, y)$. Найдем функцию распределения произведения $\zeta = \xi\eta$. Имеем в виду (50):

$$F_\zeta(z) = P\{\xi\eta < z\} = \int\limits_{xy<z}\int p(x, y)\,dx\,dy =$$

$$= \int_{-\infty}^{0} dx \int_{z/x}^{\infty} p(x, y)\,dy + \int_{0}^{\infty} dx \int_{-\infty}^{z/x} p(x, y)\,dy. \qquad (59)$$

Отсюда (формально) получаем выражение для плотности ζ

$$p_\zeta(z) = F'_\zeta(z) = -\int_{-\infty}^{0} \frac{1}{x}\,p\left(x, \frac{z}{x}\right)dx +$$

$$+\int_{0}^{\infty} \frac{1}{x}\,p\left(x, \frac{z}{x}\right)dx. \qquad (60)$$

4. Покажем прямым вычислением, что если случайные величины ξ и η независимы и их плотности распределения равны

$$p_\xi(x) = p_\eta(x) = \begin{cases} 0, & \text{при } x \leqslant 0, \\ e^{-x}, & \text{при } x > 0, \end{cases} \qquad (61)$$

то случайные величины $\xi+\eta$ и ξ/η также независимы.

Случайные величины ξ и η независимы, поэтому плотность вероятности $p(x_1, x_2)$ случайного вектора (ξ, η) равна $p(x_1, x_2) = p_\xi(x_1)\,p_\eta(x_2) \equiv p_1(x_1)\,p_2(x_2)$. Отсюда при любых неотрицательных x и y

$$P\left\{\xi+\eta < x,\ \frac{\xi}{\eta} < y\right\} = \iint\limits_{D} p_1(x_1)\,p_2(x_2)\,dx_1\,dx_2, \qquad (62)$$

где область D определяется условиями $D = \{(x_1, x_2) : x_1 + x_2 < x,\ x_1/x_2 < y\}$. Переходя в формуле (62) от двойного интеграла к повторному, найдем

$$P\left\{\xi+\eta < x,\ \frac{\xi}{\eta} < y\right\} = \int_{-\infty}^{0} p_2(x_2)\,dx_2 \int_{x_2 y}^{x-x_2} p_1(x_1)\,dx_1 +$$

$$+ \int_{0}^{\frac{x}{y+1}} p_2(x_2)\,dx_2 \int_{-\infty}^{x_2 y} p_1(x_1)\,dx_1 + \int_{\frac{x}{y+1}}^{\infty} p_2(x_2)\,dx_2 \int_{-\infty}^{x-x_2} p_1(x_1)\,dx_1.$$

Ввиду (61) первый интеграл в правой части равен нулю, два других преобразуются следующим образом:

$$P\left\{\xi+\eta < x, \frac{\xi}{\eta} < y\right\} =$$

$$= \int_{0}^{\frac{x}{y+1}} e^{-x_2}\,dx_2 \int_{0}^{x_2 y} e^{-x_1}\,dx_1 +$$

$$+ \int_{\frac{x}{y+1}}^{x} e^{-x_2}\,dx_2 \int_{0}^{x-x_2} e^{-x_1}\,dx_1 =$$

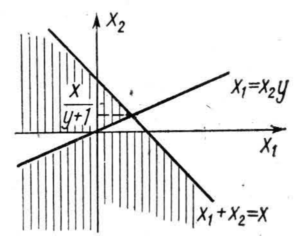


Рис. 16

$$= \int_{0}^{x} e^{-x_2}\,dx_2 - \int_{0}^{\frac{x}{y+1}} e^{-x_2(1+y)}\,dx_2 - e^{-x}x\left(1 - \frac{1}{1+y}\right) =$$

$$= 1 - e^{-x} + \frac{1}{y+1}(e^{-x} - 1) - e^{-x}\frac{xy}{y+1} =$$

$$= \frac{y}{y+1}(1 - e^{-x} - e^{-x}x). \qquad (63)$$

С другой стороны, согласно (51) и (61)

$$P\{\xi+\eta < x\} = \int_{-\infty} dz \int_{-\infty}^{\infty} p_1(x_1)\,p_2(z-x_1)\,dx_1 =$$

$$= \int\limits_{-\infty}^{x} dz \int\limits_{0}^{\infty} e^{-x_1} p_2(z-x_1)\,dx_1 =$$

$$= \int\limits_{0}^{x} dz \int\limits_{0}^{z} e^{-x_1} e^{-z+x_1}\,dx_1 = -xe^{-x} + 1 - e^{-x}. \tag{64}$$

Далее,

$$P\left\{\frac{\xi}{\eta} < y\right\} = \iint\limits_{x_1/x_2<y} p_1(x_1)\,p_2(x_2)\,dx_1\,dx_2 =$$

$$= \int\limits_{-\infty}^{0} p_2(x_2)\,dx_2 \int\limits_{x_2y}^{\infty} p_1(x_1)\,dx_1 + \int\limits_{0}^{\infty} p_2(x_2)\,dx_2 \int\limits_{-\infty}^{x_2y} p_1(x_1)\,dx_1. \tag{65}$$

Преобразуя правую часть с учетом (61), находим

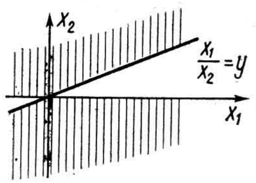


Рис. 17

$$P\left\{\frac{\xi}{\eta} < y\right\} = \int\limits_{0}^{\infty} e^{-x_2}\,dx_2 \int\limits_{0}^{x_2y} e^{-x_1}\,dx_1 =$$

$$= \int\limits_{0}^{\infty} e^{-x_2}(1-e^{-x_2y})\,dx_2 = \frac{y}{y+1}. \tag{66}$$

Сопоставляя (63), (64) и (66), получим

$$P\left\{\xi+\eta<x,\ \frac{\xi}{\eta}<y\right\} = P\{\xi+\eta<x\}\,P\left\{\frac{\xi}{\eta}<y\right\},$$

так что случайные величины $\xi+\eta$ и ξ/η независимы.

5. Докажем, что линейная функция нормальных случайных величин нормальна. Очевидно, достаточно рассмотреть линейную функцию двух случайных величин. Пусть

$$\zeta = \alpha\xi + \beta\eta + \gamma,\quad |\alpha|+|\beta|>0, \tag{67}$$

где случайный вектор (ξ, η) распределен нормально, α, β, γ — константы. Надо доказать, что ζ имеет нормальное распределение. Заметим, прежде всего, что если $\xi \in N(0, 1)$, то $(\sigma\xi+a) \in N(a, \sigma^2)$. В самом деле (пусть для определенности $\sigma>0$),

$$F_{\sigma\xi+a}(x) = P\{\sigma\xi + a < x\} = P\left\{\xi < \frac{x-a}{\sigma}\right\} =$$

$$= \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{\frac{x-a}{\sigma}} e^{-\frac{y^2}{2}}\,dy.$$

Отсюда

$$p_{\sigma\xi+a}(x) = F'_{\sigma\xi+a}(x) = \frac{1}{\sigma\sqrt{2\pi}}\,e^{-\frac{(x-a)^2}{2\sigma^2}},$$

т. е.

$$(\sigma\xi+a) \in N(a, \sigma^2).$$

Ввиду этого достаточно доказать высказанное утверждение в случае, когда $\gamma=0$, а плотность $p(x_1, x_2)$ случайного вектора (ξ, η) равна

$$p(x_1, x_2) = \frac{1}{2\pi\sqrt{1-r^2}} \exp\left\{-\frac{1}{2(1-r^2)}[x_1^2 - 2rx_1x_2 + x_2^2]\right\} \tag{68}$$

($\sigma_1^2 = D\xi = 1$, $\sigma_2^2 = D\eta = 1$, $a=b=0$, см. (34)). Используя формулу (51) для функции распределения суммы, найдем

$$F_\zeta(x) = P\{\zeta < x\} =$$

$$= P\{\alpha\xi + \beta\eta < x\} = \int\limits_{-\infty}^{x} dz \int\limits_{-\infty}^{\infty} p\left(x_1, \frac{1}{\beta}(z - \alpha x_1)\right) dx_1 \tag{69}$$

(считаем для определенности, что $\beta \neq 0$). Отсюда и из (68) для плотности $p_\zeta(x)$ случайной величины ζ получим выражение

$$p_\zeta(x) = F'_\zeta(x) = \int\limits_{-\infty}^{\infty} p\left(x_1, \frac{1}{\beta}(x - \alpha x_1)\right) dx_1 =$$

$$= \frac{1}{2\pi\sqrt{1-r^2}} \int\limits_{-\infty}^{\infty} \exp\left\{-\frac{1}{2(1-r^2)} \times\right.$$

$$\left.\times \left[x_1^2 - 2rx_1\frac{1}{\beta}(x-\alpha x_1) + \frac{1}{\beta^2}(x-\alpha x_1)^2\right]\right\} dx_1 =$$

$$= \frac{1}{2\pi\sqrt{1-r^2}} \int\limits_{-\infty}^{\infty} e^{-\frac{1}{2(1-r^2)}[Ax_1^2 - 2Bx_1 + C]}\,dx_1 =$$

$$= \frac{1}{2\pi\sqrt{1-r^2}}\,e^{-\frac{C-B^2/A}{2(1-r^2)}} \int\limits_{-\infty}^{\infty} e^{-\frac{1}{2(1-r^2)}(\sqrt{A}\,x_1 - B/\sqrt{A})^2}\,dx_1 =$$

$$= \frac{1}{\sqrt{2\pi A}}\,e^{-\frac{1}{2(1-r^2)}(C - B^2/A)}, \tag{70}$$

где

$$A = 1 + 2r\frac{\alpha}{\beta} + \frac{\alpha^2}{\beta^2}, \quad B = \frac{x}{\beta^2}(\alpha + r\beta), \quad C = \frac{x^2}{\beta^2};$$

$A>0$, так как $|r| \leqslant 1$. Но $C - B^2/A = x^2(1-r^2)/A$; на основании (70) заключаем, что

$$p_\zeta(x) = \frac{1}{\sqrt{2\pi}\sqrt{A}} e^{-\frac{x^2}{2A}}, \tag{71}$$

и, таким образом, $\zeta \in N(0, A)$.

§ 9. ЧИСЛОВЫЕ ХАРАКТЕРИСТИКИ СЛУЧАЙНЫХ ВЕЛИЧИН

1°. Основные определения. Моменты случайных величин

Случайная величина ξ считается, как известно, заданной, если задано пространство $\overline{\Omega}$ ее значений, σ-алгебра $\overline{\mathscr{F}}$ событий (подмножеств) $\overline{\Omega}$ и вероятность $\overline{P}$ на $\overline{\mathscr{F}}$. В рассмотренных ранее случаях $\overline{F}$ либо σ-алгебра борелевских подмножеств $\overline{\Omega} = R_1$, если ξ — непрерывная случайная величина, либо σ-алгебра всех подмножеств $\overline{\Omega} = \{x_1, x_2, \ldots\}$, если ξ дискретна, $x_1, x_2, \ldots$ — ее значения.

Но во многих задачах такая полная характеристика случайной величины ξ, с одной стороны, недоступна для исследователя, а с другой стороны, и необязательна, достаточно ограничиться знанием некоторых параметров распределения ξ. Такими параметрами являются различные моменты случайной величины.

Определение 1. Моментом (начальным) порядка k дискретной случайной величины ξ, принимающей значения x_i с вероятностями $P\{\xi = x_i\} = p_i$, $i = 1, 2, \ldots$, называется число

$$M\xi^k = \sum_{i=1}^{\infty} x_i^k p_i, \tag{1}$$

при условии, что ряд (1) сходится абсолютно, т. е.

$$M|\xi^k| = \sum_{i=1}^{\infty} |x_i|^k p_i < \infty. \tag{2}$$

$M|\xi^k|$ — называется **абсолютным моментом порядка k.**

Моментом (начальным) порядка k непрерывной случайной величины ξ с плотностью вероятности $p(x)$ называется число

$$M\xi^k = \int_{-\infty}^{\infty} x^k p(x)\,dx \tag{3}$$

при условии, что интеграл (3) сходится абсолютно, т. е.

$$M|\xi^k| = \int_{-\infty}^{\infty} |x|^k p(x)\,dx < \infty. \tag{4}$$

$M|\xi^k|$ — называется **абсолютным моментом порядка k.**

Привлекая интеграл Стильтьеса, можно выражения (1) и (3) объединить одной записью*

$$M\xi^k = \int_{-\infty}^{\infty} x^k dF_\xi(x). \tag{5}$$

Если $M|\xi^k|$ не существует, то говорят, что случайная величина ξ не имеет конечного момента порядка k. По определению моменты $M\xi^k$ и $M|\xi^k|$ существуют или не существуют одновременно. Требование абсолютной сходимости гарантирует возможность произвольного порядка суммирования и, следовательно, корректность определения $M\xi^k$ (в (1), например, порядок суммирования определяется порядком нумерации значений ξ).

Определение 2. Момент $M\xi$ первого порядка ($k=1$ в (1) и (3)) называется **математическим ожиданием**, или **средним значением**, случайной величины ξ.

Примеры.

1. Равномерное распределение: ξ равномерно распределена в $[a, b]$, т. е.

$$p(x) = \begin{cases} \dfrac{1}{b-a}, & x \in [a, b], \\ 0, & x \overline{\in} [a, b], \end{cases}$$

$$M\xi = \frac{1}{b-a}\int_a^b x\,dx = \frac{a+b}{2} \text{ — середина отрезка } [a, b].$$

2. Распределение Пуассона: $\xi = k$ с вероятностью

$$p_k = P\{\xi = k\} = \frac{\lambda^k}{k!}e^{-\lambda}, \ \lambda > 0, \ k = 0, 1, 2, \ldots,$$

$$M\xi = \sum_{k=0}^{\infty} k\frac{\lambda^k e^{-\lambda}}{k!} = e^{-\lambda}\sum_{k=1}^{\infty}\frac{\lambda^k}{(k-1)!} =$$

$$= \lambda e^{-\lambda}\sum_{k=0}^{\infty}\frac{\lambda^k}{k!} = \lambda.$$

* Мы далее не пользуемся свойствами интеграла Стильтьеса, и равенство (5) означает лишь краткую запись для обоих выражений (1) и (3).

3. Распределение Коши; так называется распределение случайной величины $\eta=\xi_1/\xi_2$, где ξ_1, $\xi_2 \in N(0, 1)$ и независимы. Пользуясь формулой (65) § 8 с

$$p_1(x) = p_2(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2},$$

без труда найдем

$$F_\eta(x) = \frac{1}{2} + \frac{1}{\pi} \operatorname{arctg} x, \text{ так что } p_\eta(x) = \frac{1}{\pi(1+x^2)}.$$

Поэтому $M\eta$ не существует, ибо

$$\int_{-\infty}^{\infty} \frac{|x|\,dx}{1+x^2} = \infty.$$

4. Нормальное распределение: $\xi \in N(\mu, \sigma^2)$

$$M\xi = \frac{1}{\sqrt{2\pi}\,\sigma} \int_{-\infty}^{\infty} x e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (\sigma t + \mu) e^{-\frac{t^2}{2}} dt =$$

$$= \frac{\sigma}{\sqrt{2\pi}} \int_{-\infty}^{\infty} t e^{-\frac{t^2}{2}} dt + \frac{\mu}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{t^2}{2}} dt = \mu,$$

так как первый интеграл в правой части равен нулю.

5. Некто стоит перед дверью своей квартиры и пытается открыть ее, перебирая ключи из связки; при этом испробованный ключ не исключается при дальнейших попытках. Спрашивается, сколько в среднем придется сделать попыток открывающему дверь, прежде чем он попадет домой?

Решение. Пусть ξ — число попыток, понадобившихся для открывания двери, $\xi=1, 2, \ldots$. Если в связке n ключей, из которых один — от данной двери, то событие $\xi=k$ состоит в том, что нужный ключ попадается k-м, а перед этим $k-1$ раз выбирался неподходящий ключ. Вероятность этого события

$$p_k = \overline{P}\{\xi = k\} = \left(\frac{n-1}{n}\right)^{k-1} \frac{1}{n},$$

и, следовательно,

$$M\xi = \sum_{k=1}^{\infty} k \left(\frac{n-1}{n}\right)^{k-1} \frac{1}{n} = \frac{1}{n} \sum_{k=1}^{\infty} k q^{k-1} =$$

$$= \frac{1}{n} \frac{d}{dq} \sum_{k=1}^{\infty} q^k = \frac{1}{n} \frac{d}{dq} \left(\frac{q}{1-q}\right)\Bigg|_{q=\frac{n-1}{n}} = \frac{1}{n} \cdot n^2 = n,$$

$$q = \frac{n-1}{n} < 1.$$

Предположим теперь, что испробованный ключ устраняется из дальнейших попыток. В этом случае $\xi=1, 2, \ldots, n$ и

$$\overline{P}\{\xi = k\} = \frac{n-1}{n} \cdot \frac{n-2}{n-1} \cdots$$

$$\cdots \frac{n-(k-1)}{n-(k-2)} \cdot \frac{1}{n-(k-1)} = \frac{1}{n}, \quad k = 1, \ldots, n.$$

Поэтому

$$M\xi = \sum_{k=1}^{n} \frac{1}{n} k = \frac{1}{n} \frac{n(n+1)}{2} = \frac{n+1}{2}.$$

Рассмотрим еще тот случай, когда речь идет о (безуспешных) попытках открыть дверь одним ключом, выбранным из связки наугад. Обозначим через A событие, состоящее в том, что выбранный наугад ключ подходит, $\overline{A}$ — противоположное событие. В этом случае $\xi=1, \infty$. По формуле полной вероятности:

$$p_k = \overline{P}\{\xi = x_k\} = P\{\xi = x_k | A\} P(A) + P\{\xi = x_k | \overline{A}\} P(\overline{A}),$$

и

$$M\xi = \sum_{k=1}^{2} x_k p_k = P(A) \sum_{k=1}^{2} x_k P\{\xi = x_k | A\} +$$

$$+ P(\overline{A}) \sum_{k=1}^{2} x_k P\{\xi = x_k | \overline{A}\} = M(\xi | A) P(A) + M(\xi | \overline{A}) P(\overline{A}),$$

где $M(\xi|A)$ и $M(\xi|\overline{A})$ называются условным математическим ожиданием ξ при условии A или $\overline{A}$ (см. подробнее ниже, п. 4°).

Имеем

$$P(A) = \frac{1}{n}, \quad P(\overline{A}) = \frac{(n-1)}{n}, \quad P\{\xi = 1 | A\} = 1,$$

$$P\{\xi = \infty | A\} = 0, \quad p\{\xi = 1 | \overline{A}\} = 0, \quad p\{\xi = \infty | \overline{A}\} = 1,$$

так что $M(\xi|A) = 1$, $M(\xi|\overline{A}) = \infty$ и $M\xi = \infty$, как и следовало ожидать.

6. *Петербургская игра.* Играющий бросает монету до первого выпадения герба, после чего игра прекращается и бросавший монету получает 2^k коп., где k — число бросаний монеты. Вероятность того, что герб рано или поздно появится, равна единице. Следовательно, с вероятностью единица игрок получит некоторую сумму денег. Для того чтобы ука-

зать осмысленную цену за участие в игре, следует подсчитать математическое ожидание выигрыша (средний выигрыш). Обозначим выигрыш ξ. При этом $\bar{\Omega}=\{2, 2^2, \ldots, 2^k, \ldots\}$, и, очевидно, $p_k=\bar{P}\{\xi=2^k\}=2^{-k}$, $k=1, 2, \ldots$. Следовательно,

$$M\,\xi = 2\cdot\frac{1}{2}+4\cdot\frac{1}{4}+\ldots=\infty.$$

Таким образом, играя неопределенно долго, можно рассчитывать получить сколь угодно большую сумму денег. Поэтому априори невозможно указать сумму, которая была бы адекватной платой за участие в игре, цена игры должна быть бесконечно большой.

Однако если речь идет о реальном участии в игре, то следует ограничить максимальный выигрыш. Будем считать, что максимальный выигрыш определен 2^{25} коп $\approx 335\cdot10^3$ руб. Это означает, что такая сумма будет получена, если число бросаний $n\geqslant25$. Итак, в этом случае математическое ожидание выигрыша (и цена игры)

$$M\,\xi = 2\cdot\frac{1}{2}+4\cdot\frac{1}{4}+\ldots+2^{24}\cdot\frac{1}{2^{24}}+$$

$$+\,2^{25}\left(\frac{1}{2^{25}}+\frac{1}{2^{26}}+\ldots\right)=24+\left(1+\frac{1}{2}+\ldots\right)=26 \text{ коп.}$$

Таким образом, если считать реальным, скажем, 2^{20} бросаний, что при условии затраты 3с на бросание занимает свыше месяца непрерывной игры, то средний выигрыш все-таки поразительно мал (<26 коп.). Этот факт, известный как петербургский парадокс, рассмотрен Даниилом Бернулли. Отметим, что испытания Бернулли названы в честь Якова Бернулли.

Переходим теперь к доказательству важной теоремы о математическом ожидании функций от случайных величин.

Теорема 1. Пусть ξ — дискретная случайная величина (непрерывная случайная величина), принимающая значения $x_1, x_2, \ldots$ соответственно с вероятностями $p_1, p_2, \ldots$ (имеющая плотность вероятности $p(x)$), а $\eta=f(\xi)$ — новая случайная величина, где $f(\cdot)$ — некоторая функция. Тогда математическое ожидание η равно

$$M\,\eta = Mf(\xi)=\sum_{i=1}^{\infty}f(x_i)\,p_i\quad\left(M\,\eta=Mf(\xi)=\int_{-\infty}^{\infty}f(x)\,p(x)\,dx\right),\tag{6}$$

если ряд (интеграл) (6) сходится абсолютно.

Доказательство проведем для дискретной случайной величины ξ. В этом случае случайная величина $\eta=f(\xi)$ также дискретна, ее значениями являются числа $y_1, y_2, \ldots$, где множество $y_1, y_2, \ldots$ совпадает с множеством всех различных чисел среди $f(x_1), f(x_2), \ldots$, а вероятность каждого значения y_s равна

$$q_s=\bar{P}\{\eta=y_s\}=P\{\omega: f(\xi(\omega))=y_s\}=\sum_{k:f(x_k)=y_s}p_k.\tag{7}$$

Теперь имеем

$$M\,\eta=\sum_{s=1}^{\infty}y_s q_s=\sum_{s=1}^{\infty}y_s\sum_{k:f(x_k)=y_s}p_k=$$

$$=\sum_{s=1}^{\infty}\left\{\sum_{k:f(x_k)=y_s}f(x_k)\,p_k\right\}=\sum_{i=1}^{\infty}f(x_i)\,p_i,\tag{8}$$

последнее равенство выполняется ввиду того, что каждое слагаемое $f(x_i)\,p_i$ участвует в двух последних суммах один и только один раз, поскольку все y_s различны, возможность объединения в одну сумму следует из леммы о суммировании по блокам (см. § 3), так как ряд $\sum_{i=1}^{\infty}|f(x_i)|\,p_i$ по условию сходится. ▲

Следствие. Для любой случайной величины ξ и постоянных c_k, $k=0, 1, \ldots, n$

$$M\left(\sum_{k=0}^{n}c_k\,\xi^k\right)=\sum_{k=0}^{n}c_k M\,\xi^k,\tag{9}$$

если $M|\xi^k|<\infty$, $k=1, 2, \ldots, n$.

Равенство (9) вытекает из теоремы 1, в которой $f(x)=\sum_{k=0}^{n}c_k x^k$ — полином*. Из (9), в частности, следует, что

$$Mc=c \text{ и } M(c\xi)=cM\xi,\ c \text{ — любая постоянная,}\tag{10}$$

(в последнем равенстве предполагается, что $M|\xi|<\infty$).

Замечание. В случае, когда ξ — дискретная случайная величина, принимающая значения $x_1, x_2, \ldots$ с вероятностями $p_1, p_2, \ldots$, определена на дискретном вероятностном пространстве $(\Omega, \mathscr{F}, P)$, нетрудно дать другое выражение для математического ожидания:

$$M\,\xi=\sum_{\omega_i\in\Omega}\xi(\omega_i)\,P(\omega_i),$$

* Мы не останавливаемся на доказательстве того, что непрерывная функция от случайной величины является случайной величиной, (см. [1]).

здесь $P(\omega_i)$ — вероятность элементарного исхода ω_i, и сумма распространена на все элементарные события $\omega_i \in \Omega$. Действительно,

$$\sum_{\omega_i \in \Omega} \xi(\omega_i) P(\omega_i) = \sum_{k=1}^{\infty} \Big\{ \sum_{\omega_i:\, \xi(\omega_i)=x_k} \xi(\omega_i) P(\omega_i) \Big\} =$$

$$= \sum_{k=1}^{\infty} x_k \sum_{\omega_i:\xi(\omega_i)=x_k} P(\omega_i) = \sum_{k=1}^{\infty} x_k p_k, \tag{11}$$

если последний ряд сходится абсолютно, ибо тогда применима обратная лемма о суммировании по блокам (см. § 3). Итак, в случае существования математического ожидания $M\xi$ имеем равенство

$$M\xi = \sum_{\omega_i \in \Omega} \xi(\omega_i) P(\omega_i). \tag{12}$$

Аналог формулы (12) справедлив и в самом общем случае любой случайной величины, но вместо суммы в (12) должен стоять интеграл Лебега.

Продолжим изучение моментов случайных величин.

Определение 3. Математическое ожидание

$$M(\xi - M\xi)^k \tag{13}$$

называется **k-м центральным моментом**, если существует $M|\xi - M\xi|^k$. Центральный момент второго порядка называется **дисперсией** случайной величины ξ и обозначается $D\xi$

$$D\xi = M(\xi - M\xi)^2. \tag{14}$$

Этот момент является очень удобной характеристикой разброса значений ξ около ее среднего значения — математического ожидания $M\xi$. Так как в согласии со свойством (9)

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

$$= M\xi^2 - 2M\xi \cdot M\xi + (M\xi)^2,$$

то справедлива следующая формула для дисперсии:

$$D\xi = M\xi^2 - (M\xi)^2. \tag{15}$$

Отсюда следует, в частности, что $M\xi^2 \geqslant (M\xi)^2$, поскольку $D\xi \geqslant 0$. Дисперсия имеет размерность квадрата случайной величины; для характеристики разброса, рассеивания иногда бывает удобнее пользоваться значением, размерность которого совпадает с размерностью случайной величины. Такая величина

$$\sigma\xi = \sqrt{D\xi} \tag{16}$$

называется **среднеквадратичным уклонением** ξ.

На основании определения (14) и теоремы 1 можно записать:

1) если ξ — дискретная случайная величина, $x_1, x_2, \ldots$ — ее значения, а $p_1, p_2, \ldots$ — соответствующие вероятности, то

$$D\xi = \sum_{k=1}^{\infty} (x_k - M\xi)^2 p_k, \tag{17}$$

2) если ξ — непрерывная случайная величина и $p(x)$ — ее плотность вероятности, то

$$D\xi = \int_{-\infty}^{\infty} (x - M\xi)^2 p(x)\, dx. \tag{18}$$

Примеры.

1. Нормальное распределение, $\xi \in N(\mu, \sigma^2)$. Так как $M\xi = \mu$, то в силу (18)

$$D\xi = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} (x-\mu)^2 e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} t^2 e^{-\frac{t^2}{2}} dt =$$

$$= -\frac{\sigma^2}{\sqrt{2\pi}} t e^{-\frac{t^2}{2}} \Big|_{-\infty}^{\infty} + \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{t^2}{2}} dt = \sigma^2.$$

Таким образом, параметры нормального распределения $N(\mu, \sigma^2)$: μ — математическое ожидание, σ^2 — дисперсия. Нормальное распределение полностью определяется этими двумя параметрами.

2. Распределение Пуассона: $\xi = k$, $P(\{\xi = k\}) = e^{-\lambda}\lambda^k/k!$, $\lambda > 0$, $k = 0, 1, 2, \ldots$. Было показано, что $M\xi = \lambda$. Ввиду (15) имеем

$$D\xi = \sum_{k=0}^{\infty} k^2 \frac{e^{-\lambda}\lambda^k}{k!} - \lambda^2 = e^{-\lambda} \sum_{k=0}^{\infty} k(k-1) \frac{\lambda^k}{k!} +$$

$$+ e^{-\lambda} \sum_{k=0}^{\infty} k \frac{\lambda^k}{k!} - \lambda^2 = e^{-\lambda} \lambda^2 \frac{d^2}{d\lambda^2} \Big(\sum_{k=0}^{\infty} \frac{\lambda^k}{k!} \Big) +$$

$$+ e^{-\lambda} \lambda \frac{d}{d\lambda} \Big(\sum_{k=0}^{\infty} \frac{\lambda^k}{k!} \Big) - \lambda^2 = e^{-\lambda}(\lambda^2 e^{\lambda} + \lambda e^{\lambda}) - \lambda^2 = \lambda.$$

Таким образом, единственный параметр распределения Пуассона задает как математическое ожидание, так и дисперсию.

Приведем выражения математического ожидания и дисперсии для случайной величины ξ смешанного типа (комбинации дискретного и непрерывного распределения, см. § 8):

$$M\xi = q_1 \sum_{k=1}^{\infty} x_k\, p_k + q_2 \int_{-\infty}^{\infty} x p(x)\, dx,$$

$$D\xi = q_1 \sum_{k=1}^{\infty} (x_k - M\xi)^2 p_k + q_2 \int_{-\infty}^{\infty} (x - M\xi)^2 p(x)\, dx, \qquad (19)$$

$$q_1, q_2 \geqslant 0, \quad q_1 + q_2 = 1.$$

2°. Свойства математического ожидания и дисперсии

1) Математическое ожидание суммы случайных величин равно сумме математических ожиданий

$$M(\xi+\eta) = M\xi + M\eta, \qquad (20)$$

при условии, что $M|\xi|$ и $M|\eta|$ конечны.

Доказательство.

а) (ξ, η) — дискретная случайная величина. Пусть ξ принимает значения $x_1, x_2, \ldots$, а η — значения $y_1, y_2, \ldots$. Тогда $\xi+\eta$ принимает значения $z_1, z_2, \ldots$, где все z_s различны, $z_s = x_i + y_j$, а

$$r_s = P\{\xi + \eta = z_s\} = \sum_{i,\, j: x_i + y_j = z_s} p_{ij}, \qquad (21)$$

ввиду (49), § 8; $p_{ij} = P\{\xi = x_i,\ \eta = y_j\}$. Поэтому

$$M(\xi+\eta) = \sum_{s=1}^{\infty} z_s r_s = \sum_{s=1}^{\infty} z_s \Big(\sum_{i,\, j: x_i + y_j = z_s} p_{ij} \Big) =$$

$$= \sum_{s=1}^{\infty} \Big(\sum_{i,\, j: x_i + y_j = z_s} (x_i + y_j)\, p_{ij} \Big) = \sum_{i,\, j=1}^{\infty} (x_i + y_j)\, p_{ij} = \sum_{i=1}^{\infty} x_i \sum_{j=1}^{\infty} p_{ij} +$$

$$+ \sum_{j=1}^{\infty} y_j \sum_{i=1}^{\infty} p_{ij} = \sum_{i=1}^{\infty} x_i p_i + \sum_{j=1}^{\infty} y_j q_j = M\xi + M\eta,$$

здесь $\sum_{j=1}^{\infty} p_{ij} = P\{\xi = x_i\} = p_i$ и $\sum_{i=1}^{\infty} p_{ij} = P\{\eta = y_j\} = q_j$; группировать ряды можно в силу леммы и обратной леммы о суммировании по блокам, поскольку ряды по условию сходятся абсолютно.

б) (ξ, η) — непрерывная случайная величина, $p(x, y)$ — ее плотность вероятности. Согласно (52) § 8 плотность суммы $\zeta = \xi + \eta$ имеет вид $p_\zeta(z) = \int_{-\infty}^{\infty} p(x, z-x)\, dx$. Поэтому

$$M(\xi+\eta) = \int_{-\infty}^{\infty} z p_\zeta(z)\, dz = \iint_{-\infty}^{\infty} z p(x, z-x)\, dx dz =$$

$$= \iint_{-\infty}^{\infty} (x+y)\, p(x, y)\, dx dy = \int_{-\infty}^{\infty} x dx \int_{-\infty}^{\infty} p(x, y)\, dy +$$

$$+ \int_{-\infty}^{\infty} y dy \int_{-\infty}^{\infty} p(x, y)\, dx = \int_{-\infty}^{\infty} x p_\xi(x)\, dx + \int_{-\infty}^{\infty} y p_\eta(y)\, dy = M\xi + M\eta,$$

интегралы можно переставить в силу их абсолютной сходимости. Свойство 1) доказано.

2) Из следствия (п. 1°) и свойства 1) получим свойство линейности математического ожидания:

$$M(c_1\xi + c_2\eta) = c_1 M\xi + c_2 M\eta, \qquad (22)$$

c_1, c_2 — любые постоянные, если $M|\xi|$ и $M|\eta|$ конечны.

Замечание. Если воспользоваться формулой (12), то равенства (20) и (22) можно доказать еще проще:

$$M(c_1\xi + c_2\eta) = \sum_{\omega_i \in \Omega} (c_1 \xi(\omega_i) + c_2 \eta(\omega_i))\, P\{\omega_i\} =$$

$$= c_1 \sum_{\omega_i \in \Omega} \xi(\omega_i)\, P\{\omega_i\} + c_2 \sum_{\omega_i \in \Omega} \eta(\omega_i)\, P\{\omega_i\} = c_1 M\xi + c_2 M\eta.$$

3) Если случайные величины ξ и η независимы, то

$$M\xi\eta = M\xi \cdot M\eta \qquad (23)$$

при условии, что $M|\xi|$ и $M|\eta|$ конечны.

Доказательство (в обозначениях 1)).

а) $\xi\eta$ — дискретная случайная величина со значениями $t_1, t_2, \ldots$, где все t_s различны, $t_s = x_i y_j$ и $r_s = P\{\xi\eta = t_s\} = \sum_{i,\, j: x_i y_j = t_s} p_{ij}$, причем $p_{ij} = P\{\xi = x_i,\ \eta = y_j\} = p_i q_j$, так как ξ и η независимы. Поэтому

$$M\xi\eta = \sum_{s=1}^{\infty} t_s r_s = \sum_{s=1}^{\infty} t_s \Big(\sum_{i,\, j: x_i y_j = t_s} p_i q_j \Big) =$$

$$= \sum_{s=1}^{\infty} \Big\{ \sum_{i,\, j: x_i y_j = t_s} x_i y_j p_i q_j \Big\} = \sum_{i,\, j=1}^{\infty} x_i y_j p_i q_j = \sum_{i=1}^{\infty} x_i p_i \sum_{j=1}^{\infty} y_j q_j = M\xi \cdot M\eta,$$

группировка рядов законна в силу их абсолютной сходимости и леммы о суммировании по блокам.

б) Согласно формуле (60) § 8 плотность произведения $\zeta = \xi\eta$ имеет следующий вид (с учетом того, что $p(x, y) =$

$=p_\xi(x)p_\eta(y)$ в силу независимости ξ и η):

$$p_\zeta(z)=-\int\limits_{-\infty}^{0}\frac{1}{x}p_\xi(x)p_\eta\left(\frac{z}{x}\right)dx+\int\limits_{0}^{\infty}\frac{1}{x}p_\xi(x)p_\eta\left(\frac{z}{x}\right)dx.$$

Отсюда

$$M\xi\eta=\int\limits_{-\infty}^{\infty}zp_\zeta(z)\,dz=-\int\limits_{-\infty}^{0}\frac{1}{x}p_\xi(x)\,dx\int\limits_{-\infty}^{\infty}zp_\eta\left(\frac{z}{x}\right)dz+$$

$$+\int\limits_{0}^{\infty}\frac{1}{x}p_\xi(x)\,dx\int\limits_{-\infty}^{\infty}zp_\eta\left(\frac{z}{x}\right)dz=\int\limits_{-\infty}^{0}\frac{1}{x}p_\xi(x)\,dx\int\limits_{-\infty}^{\infty}tx^2p_\eta(t)\,dt+$$

$$+\int\limits_{0}^{\infty}\frac{1}{x}p_\xi(x)\,dx\int\limits_{-\infty}^{\infty}tx^2p_\eta(t)\,dt=\int\limits_{-\infty}^{\infty}xp_\xi(x)dx\int\limits_{-\infty}^{\infty}tp_\eta(t)\,dt=M\xi M\eta,$$

перестановки интегралов законны в силу их абсолютной сходимости.

4) *Некоторые неравенства.*

а) Если $\xi\geqslant\eta$, то $M\xi\geqslant M\eta$. (24)

В самом деле, случайная величина $\zeta=\xi-\eta$ принимает неотрицательные значения, поэтому согласно определению 1 (с $k=1$) справедливо (24).

б) Неравенство Коши—Буняковского:

$$M|\xi\eta|\leqslant\sqrt{M\xi^2\cdot M\eta^2}, \tag{25}$$

если величины справа конечны.

Доказательство. Имеем $|\xi\eta|\leqslant 1/2(\xi^2+\eta^2)$. Отсюда и из (24) следует, что если $M\xi^2$ и $M\eta^2$ конечны, то конечно и $M|\xi\eta|$. Далее при любом λ

$$0\leqslant M(\lambda|\xi|+|\eta|)^2=\lambda^2M\xi^2+2\lambda M|\xi\eta|+M\eta^2.$$

Квадратный трехчлен относительно λ неотрицателен при всех λ, стало быть, его дискриминант неположителен, т. е. $(M|\xi\eta|)^2-M\xi^2\cdot M\eta^2\leqslant 0$, что и требовалось доказать. ▲

в) *Неравенство Чебышева.* Для любого $\varepsilon>0$

$$P\{|\xi|>\varepsilon\}\leqslant M|\xi|^2/\varepsilon^2, \tag{26}$$

если $M|\xi|^2$ конечно.

Доказательство. Введем случайную величину η по формуле

$$\eta=\begin{cases}0, & \text{если } |\xi|\leqslant\varepsilon,\\ \varepsilon, & \text{если } |\xi|>\varepsilon.\end{cases} \tag{27}$$

Таким образом, η — дискретная случайная величина, принимающая два значения: 0 с вероятностью $p_1=P\{|\xi|\leqslant\varepsilon\}$ и ε

с вероятностью $P\{|\xi|>\varepsilon\}$. Из определения η следует, что $\eta^2\leqslant|\xi|^2$, и в силу (24) имеем $M|\xi|^2\geqslant M\eta^2=\varepsilon^2P\{|\xi|>\varepsilon\}$, что совпадает с (26). ▲

Взяв в (26) вместо ξ случайную величину $\xi-M\xi$ и учитывая, что $M(\xi-M\xi)^2=D\xi$, запишем (26) в виде

$$P\{|\xi-M\xi|>\varepsilon\}\leqslant D\xi/\varepsilon^2, \tag{28}$$

именно это неравенство обычно называют неравенством Чебышева.

То же рассуждение с использованием случайной величины η из (27) и неравенства а) приводит к неравенству: для любого $\varepsilon>0$ $P\{|\xi|>\varepsilon\}\leqslant\frac{1}{\varepsilon}M|\xi|$, если $M|\xi|$ конечно и, в частности, если случайная величина ξ неотрицательна, то

$$P\{\xi>\varepsilon\}\leqslant\frac{1}{\varepsilon}M\xi.$$

5) Дисперсия постоянной равна нулю: $Dc=0$, так как

$$Dc=M(c-Mc)^2=M(c-c)^2=0.$$

Верно и обратное утверждение: если $D\xi=0$, то с вероятностью 1 ξ равна константе: $\xi=M\xi$. В самом деле, в силу (28) при любом $\varepsilon>0$ $P\{|\xi-M\xi|>\varepsilon\}=0$. Поэтому на основании полной аддитивности вероятности, получим

$$P\{|\xi-M\xi|>0\}=P\{|\xi-M\xi|>1\}+P\{1/2<|\xi-M\xi|\leqslant 1\}+$$
$$+\ldots+P\{1/2^k<|\xi-M\xi|\leqslant 1/2^{k-1}\}+\ldots=0,$$

и, таким образом, $P\{\xi-M\xi=0\}=1$.

6) Если $\eta=c\xi$, то $D\eta=c^2D\xi$, c — любая постоянная. Действительно,

$$D\eta=M(c\xi-Mc\xi)^2=M[c^2(\xi-M\xi)^2]=c^2D\xi.$$

7) Дисперсия суммы конечного числа попарно независимых случайных величин равна сумме дисперсий слагаемых

$$D\left(\sum_{i=1}^{n}\xi_i\right)=\sum_{i=1}^{n}D\xi_i. \tag{29}$$

В самом деле,

$$D\left(\sum_{i=1}^{n}\xi_i\right)=M\left(\sum_{i=1}^{n}\xi_i-M\sum_{i=1}^{n}\xi_i\right)^2=M\left(\sum_{i=1}^{n}(\xi_i-M\xi_i)\right)^2=$$

$$=M\sum_{i,k=1}^{n}(\xi_i-M\xi_i)(\xi_k-M\xi_k)=\sum_{i,k=1}^{n}M[(\xi_i-M\xi_i)(\xi_k-M\xi_k)]=$$

$$=\sum_{i=1}^{n}M(\xi_i-M\xi_i)^2+\sum_{\substack{i,k=1\\ i\neq k}}^{n}M(\xi_i-M\xi_i)M(\xi_k-M\xi_k)=\sum_{i=1}^{n}D\xi_i,$$

здесь мы несколько раз пользовались свойством 1), в предпоследнем равенстве учли независимость ξ_i и ξ_k при $i \neq k$ и свойство 3), а последнее равенство основано на том, что $M(\xi_i - M\xi_i) = M\xi_i - M\xi_i = 0$. Равенство (29) иногда называют *равенством Бьенеме.*

Пример. Пусть случайная величина ξ распределена по биномиальному закону. Найдем $M\xi$ и $D\xi$. Введем случайные величины ξ_k, равные числу успехов при k-м испытании в серии из n испытаний Бернулли. Если вероятность успеха при каждом испытании равна p, то ξ_k принимает два значения 0 и 1 с вероятностями: $P\{\xi_k=1\}=p$ и $P\{\xi_k=0\}=q=1-p$, $k=1, 2, \ldots, n$. Поэтому $M\xi_k = 1\cdot p + 0\cdot q = p$, $D\xi_k = M\xi_k^2 - (M\xi_k)^2 = 1\cdot p + 0\cdot q - p^2 = p - p^2 = pq$. Число успехов ξ в серии из n испытаний равно, очевидно, сумме $\xi = \xi_1 + \xi_2 + \ldots + \xi_n$. Отсюда ввиду независимости ξ_i и свойств 1) и 7) имеем

$$M\xi = \sum_{k=1}^{n} M\xi_k = pn, \quad D\xi = \sum_{k=1}^{n} D\xi_k = npq. \tag{30}$$

3°. Условное математическое ожидание

В § 8 была определена условная функция распределения $F(x|B) = P\{\xi < x|B\}$ случайной величины ξ при условии B, если $P(B) > 0$. Математическое ожидание (или среднее значение) ξ по отношению к этому условному распределению называется **условным математическим ожиданием.** Таким образом

$$M(\xi/B) = \int_{-\infty}^{\infty} x dF_\xi(x/B), \tag{31}$$

или подробно

$$M(\xi/B) = \sum_{k=1}^{\infty} x_k p_k(B), \tag{32}$$

если ξ дискретна, x_k — ее значения, а $p_k(B) = P\{\xi = x_k|B\}$ — соответствующие вероятности, $k = 1, 2, \ldots$; и

$$M(\xi/B) = \int_{-\infty}^{\infty} x p(x/B)\, dx, \tag{33}$$

если ξ непрерывна, $p(x|B)$ — условная плотность вероятности (см. § 8). В частности, если событие B состоит в том, что случайная величина η принимает некоторое значение y, $\eta = y$, то согласно формулам (36) и (41) § 8 получим

$$M(\xi|y_j) \equiv M(\xi|\eta = y_j) = \sum_{k=1}^{\infty} x_k P_{k/j} \tag{34}$$

для дискретных случайных величин, и

$$M(\xi|y) \equiv M(\xi|\eta = y) = \int_{-\infty}^{\infty} x p_\xi(x|y)\, dx \tag{35}$$

для непрерывных случайных величин.

Понятно, что так определенные условные математические ожидания обладают всеми свойствами 1)—4) обычных математических ожиданий, однако у них имеются и некоторые специфические свойства, связанные с возможностью применения к ним различных вариантов формулы полной вероятности. Так, если имеется полная группа попарно несовместных событий B_k, $k=1, 2, \ldots, n$, и $F_\xi(x|B_k)$ — соответствующие условные функции распределения, то ввиду равенства

$$F_\xi(x) = \sum_{k=1}^{n} F_\xi(x/B_k)\, P(B_k)$$

немедленно получим

$$M\xi = \sum_{k=1}^{n} P(B_k)\, M(\xi|B_k). \tag{36}$$

Поскольку правая часть равенства (36) имеет вид математического ожидания новой дискретной случайной величины, принимающей значения $M(\xi|B_k)$ с вероятностями $P(B_k)$, то естественно записать (36) в виде

$$M\xi = M\{M(\xi|B_k)\}. \tag{37}$$

Точно так же на основании формул (34) и (35) получаются равенства:

$$M\xi = \sum_{j=1}^{\infty} M(\xi/y_j)\, q_j \tag{38}$$

в дискретном случае и

$$M\xi = \int_{-\infty}^{\infty} M(\xi/y)\, p_\eta(y)\, dy \tag{39}$$

в непрерывном.

В самом деле, в первом случае умножим равенство (34) на q_j и просуммируем по всем j, получим, пользуясь формулой (41) § 8,

$$\sum_{j=1}^{\infty} M(\xi/y_j)\, q_j = \sum_{j=1}^{\infty}\sum_{k=1}^{\infty} x_k p_{kj} = \sum_{k=1}^{\infty} x_k p_k = M\xi.$$

Во втором случае умножаем (35) на $p_\eta(y)$ и, пользуясь равенством (37) § 8, интегрируем от $-\infty$ до ∞

$$\int_{-\infty}^{\infty} M(\xi/y)\,p_\eta(y)\,dy = \int_{-\infty}^{\infty} dy \int_{-\infty}^{\infty} x p_\eta(y)\,p_\xi^{\eta}(x/y)\,dx =$$

$$= \int_{-\infty}^{\infty} x\,dx \int_{-\infty}^{\infty} p(x,y)\,dy = \int_{-\infty}^{\infty} x p_\xi(x)\,dx = M\xi.$$

Проведенные выкладки оправданы при условии, что ряд (34) или интеграл (35) сходятся абсолютно.

Рассмотрим теперь условные математические ожидания $M(\xi|y_j)$ или $M(\xi|y)$, определенные формулами (34) или (35), как функции аргумента y. Этот аргумент — значения случайной величины η, и поэтому мы можем рассматривать $M(\xi|y_j)$ или $M(\xi|y)$ как новую случайную величину, зависящую определенным образом от η, и обозначать ее $M(\xi|\eta)$. При таком подходе правые части равенств (38) и (39) означают в силу теоремы 1 математическое ожидание функции $M(\xi|\eta)$ от случайной величины ξ и, следовательно, могут быть записаны в виде

$$M\xi = M\{M(\xi|\eta)\}. \tag{40}$$

Формулы (37) и (40) находят многочисленные применения в теории вероятностей и математической статистике.

Условное математическое ожидание $M(\xi|\eta)$, рассматриваемое как функция η, часто в статистике называется **функцией регрессии** величины ξ на η. Если, например, $M(\xi|\eta) = \alpha_1\eta + \alpha_2$, то говорят о линейной регрессии ξ на η, а α_1 и α_2 называют **коэффициентами регрессии.**

Рассмотрим случайную величину ξ, являющуюся **индикатором** события A:

$$\xi(\omega) = \begin{cases} 1, & \text{если } \omega \in A, \text{ т. е. если } A \text{ происходит,} \\ 0, & \text{если } \omega \overline{\in} A, \text{ т. е. если } A \text{ не происходит,} \end{cases}$$

ξ — дискретная случайная величина, $M\xi = 1\cdot P(A) + 0\cdot P(\bar{A}) = P(A)$.

Итак, $M\xi = P(A)$ и условная вероятность $M(\xi|\eta) = P(A|\eta)$, где $P(A|\eta)$ — условная вероятность события A при данном значении η. Равенство (40) в этом случае выглядит так:

$$P(A) = M\{P(A|\eta)\}. \tag{41}$$

4°. Моменты векторных случайных величин

Этот пункт посвящен изучению моментов многомерных случайных величин. Специфические свойства моментов векторных случайных величин связаны с зависимостью координат случайного вектора.

Определение 4. Математическим ожиданием вектора $\xi = (\xi_1, \ldots, \xi_n)$ называется вектор

$$M\xi = (M\xi_1, \ldots, M\xi_n), \tag{42}$$

составленный из математических ожиданий координат.

Дисперсией вектора $\xi = (\xi_1, \ldots, \xi_n)$ называется вектор

$$D\xi = (D\xi_1, \ldots, D\xi_n), \tag{43}$$

составленный из дисперсий координат.

Для многомерных случайных величин справедлив аналог теоремы 1, так что, например, если $f(x_1, \ldots, x_n)$ — произвольная функция такая, что $\eta = f(\xi)$ — новая случайная величина, то

$$M\eta = Mf(\xi) = \int_{-\infty}^{\infty}\!\!\ldots\!\int f(x_1, \ldots, x_n)\,p(x_1, \ldots, x_n)\,dx_1 \ldots dx_n \tag{44}$$

при условии, что интеграл (44) сходится абсолютно (здесь $p(x_1, \ldots, x_n)$ — плотность вероятности случайного вектора $\xi = (\xi_1, \ldots, \xi_n)$). В частности,

$$M\left(\sum_{k=1}^{n} a_k\xi_k\right) = M(a, \xi) = \sum_{k=1}^{n} a_k M\xi_k = (a, M\xi),$$

где $a = (a_1, \ldots, a_n)$ и $(\,,\,)$ — знак скалярного произведения в R_n, а если координаты ξ независимы в совокупности, то

$$M(\xi_1 \ldots \xi_n) = M\xi_1 \ldots M\xi_n.$$

Важной характеристикой n-мерной случайной величины ξ является так называемая **матрица ковариаций**, или **дисперсионная матрица**:

$$\begin{aligned} a_{ij} &= \operatorname{cov}\xi_i\xi_j \equiv M[(\xi_i - M\xi_i)(\xi_j - M\xi_j)] = \\ &= M\xi_i\xi_j - M\xi_i M\xi_j, \quad i, j = 1, \ldots, n. \end{aligned} \tag{45}$$

На главной диагонали в матрице ковариаций стоят дисперсии: $a_{ii} = D\xi_i$; $a_{ij} = a_{ji}$ называется также **корреляционным моментом** случайных величин ξ_i и ξ_j. В силу определения (45) ясно, что если ξ_i и ξ_j независимы, то $a_{ij} \equiv \operatorname{cov}\xi_i\xi_j = 0$. Таким образом, условие $a_{ij} \neq 0$ является достаточным признаком зависимости ξ_i и ξ_j. Обратное утверждение неверно: из равенства нулю $\operatorname{cov}\xi_i\xi_j$ не следует независимость ξ_i и ξ_j.

Поскольку при любом c $D(\xi_i + c\xi_j) = D\xi_i + c^2 D\xi_j + 2c\operatorname{cov}\xi_i\xi_j \geqslant 0$, то, полагая $c = -(\operatorname{cov}\xi_i\xi_j)/D\xi_j$, найдем $D\xi_i - (\operatorname{cov}\xi_i\xi_j)^2/D\xi_j \geqslant 0$, т. е. $|\operatorname{cov}\xi_i\xi_j| \leqslant \sqrt{D\xi_i D\xi_j}$. Отсюда следует существование матрицы ковариаций в случае, когда $D\xi_j < \infty$, $j = 1, 2, \ldots, n$.

Вспоминая доказательство равенства Бьенеме (29), получим для любого вектора $\xi=(\xi_1, \dots, \xi_n)$ с конечной дисперсией:

$$D\left(\sum_{i=1}^{n}\xi_i\right)=\sum_{i=1}^{n}D\xi_i+\sum_{\substack{i,\,j=1\\ i\neq j}}^{n}\operatorname{cov}\xi_i\xi_j. \qquad (46)$$

Из определения (45) видно, что корреляционный момент характеризует не только зависимость величин ξ_i и ξ_j, но и их рассеивание. Для характеристики чистой (линейной) связи между ξ_i и ξ_j вводят так называемый **коэффициент корреляции**

$$r_{ij}=\frac{\operatorname{cov}\xi_i\xi_j}{\sqrt{D\xi_i D\xi_j}}. \qquad (47)$$

Мы видели, что $|\operatorname{cov}\xi_i\xi_j|\leqslant\sqrt{D\xi_i D\xi_j}$, так что $|r_{ij}|\leqslant 1$, $r_{ii}=1$, $i, j=1, \dots, n$. На основании равенства (46) находим, что ($\sigma_i=\sqrt{D\xi_i}$, $\sigma_j=\sqrt{D\xi_j}$)

$$D\left(\frac{\xi_i}{\sigma_i}\pm\frac{\xi_j}{\sigma_j}\right)=2(1\pm r_{ij})\geqslant 0,$$

отсюда $r_{ij}=\pm 1$ тогда и только тогда, когда $D(\xi_i/\sigma_i\pm\xi_j/\sigma_j)=0$, а это, в свою очередь, в силу свойства 5) дисперсии возможно лишь тогда, когда случайная величина $\eta=\xi_i/\sigma_i\pm\xi_j/\sigma_j$ равна постоянной с вероятностью 1. Таким образом, $\xi_i=\alpha\xi_j+\beta$ (линейно связаны), если и только если $r_{ij}=\pm 1$.

Если $r_{ij}>0$, то говорят, что между ξ_i и ξ_j положительная корреляция, и это означает, что ξ_i и ξ_j имеют тенденцию возрастать и убывать одновременно. При $r_{ij}<0$ ситуация обратная. Если $r_{ij}=0$, то говорят, что случайные величины ξ_i и ξ_j **некоррелированы**, и этому свойству можно придать определенный геометрический смысл. С этой целью рассмотрим множество случайных величин с конечным моментом второго порядка $M\xi^2<\infty$. Из $M\xi^2<\infty$, $M\eta^2<\infty$ следует, что при любых постоянных a и b $M(a\xi+b\eta)^2<\infty$, поэтому множество таких случайных величин относительно операций сложения и умножения на число образуют линейное пространство L.

Определим в L квазискалярное произведение

$$(\xi, \eta)=M\xi\eta. \qquad (48)$$

$M\xi\eta$ обладает всеми свойствами скалярного произведения, за исключением одного: из $M\xi^2=0$ не следует $\xi=0$, а следует лишь, что $\xi=0$ с вероятностью 1. Это обстоятельство не мешает, однако, дать следующую геометрическую интерпретацию: матрица ковариаций (45) является матрицей квазискалярных произведений случайных величин $(\xi_i-M\xi_i)$,

$i=1, \dots, n$. При этом некоррелированность означает ортогональность в смысле квазискалярного произведения (48).

Заметим, что $(\xi_i-M\xi_i)$ является истинно «случайной» частью величины ξ_i и матрица ковариаций характеризует связь (статистическую) именно этих «случайных» частей. Равенство $|r_{ij}|=1$ означает, что $\xi_i-M\xi_i$ и $\xi_j-M\xi_j$ линейно зависимы в построенном линейном пространстве L.

Примеры. В разделе курса, посвященном математической статистике, мы будем постоянно использовать χ_n^2-распределение (распределение Пирсона), t_n-распределение (распределение Стьюдента) и $F_{k,m}$-распределение (распределение Фишера). В связи с этим вычислим здесь плотности этих распределений и их различные числовые характеристики.

1. **χ_n^2-распределением** с n степенями свободы называется распределение случайной величины $\chi_n^2=\xi_1^2+\dots+\xi_n^2$, где все $\xi_i\in N(0, 1)$ и независимы. Найдем плотность распределения χ_n^2. Случайный вектор $\xi=(\xi_1, \dots, \xi_n)$ имеет плотность $p(x_1, \dots, x_n)=1/(2\pi)^{n/2}\exp\left\{-\frac{1}{2}\sum_{i=1}^{n}x_i^2\right\}$, поэтому в силу (44)

$$F(x)=P\{\chi_n^2<x\}=\underset{\sum_{i=1}^{n}x_i^2<x}{\int\dots\int}\frac{1}{(2\pi)^{n/2}}e^{-\frac{1}{2}\sum_{i=1}^{n}x_i^2}\,dx_1\dots dx_n. \qquad (49)$$

Введем сферические координаты:

$$x_1=\rho\cos\varphi_1,$$
$$x_2=\rho\sin\varphi_1\cos\varphi_2,$$
$$\dots\dots\dots\dots,$$
$$x_n=\rho\sin\varphi_1\sin\varphi_2\dots\sin\varphi_{n-1},$$
$$-\pi/2\leqslant\varphi_i\leqslant\pi/2,\quad i=1, \dots, n-2,\quad -\pi\leqslant\varphi_{n-1}\leqslant\pi.$$

Тогда

$$F(x)=\frac{1}{(2\pi)^{n/2}}\int_0^{\sqrt{x}}\rho^{n-1}e^{-\frac{\rho^2}{2}}d\rho\, 2\underbrace{\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}}\dots\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}}}_{n-1}J(\varphi_1, \dots, \varphi_{n-1})\,d\varphi_1\dots$$

$$\ldots d\varphi_{n-1} = \frac{\omega_{n-1}}{(2\pi)^{n/2}} \int\limits_0^{\sqrt{x}} \rho^{n-1} e^{-\frac{\rho^2}{2}} d\rho, \qquad (50)$$

где ω_{n-1} — площадь поверхности единичной $(n-1)$-мерной сферы; ω_{n-1} легко вычислить исходя из формулы (50). Поскольку $F(+\infty)=1$, то

$$1 = \frac{\omega_{n-1}}{(2\pi)^{n/2}} \int\limits_0^{\infty} \rho^{n-1} e^{-\frac{\rho^2}{2}} d\rho = \frac{\omega_{n-1}}{(2\pi)^{n/2}} \Gamma\left(\frac{n}{2}\right) 2^{\frac{n}{2}-1}.$$

Таким образом,

$$\omega_{n-1} = \frac{2\pi^{N/2}}{\Gamma\left(\frac{n}{2}\right)}.$$

Подставляя ω_{n-1} в (50), находим окончательно

$$F(x) = \frac{1}{2^{\frac{n}{2}-1}\Gamma\left(\frac{n}{2}\right)} \int\limits_0^{\sqrt{x}} \rho^{n-1} e^{-\frac{\rho^2}{2}} d\rho, \quad x > 0. \qquad (51)$$

Поскольку, очевидно, $P\{\chi_n^2<x\}=0$ при $x\leqslant 0$, то $F(x)=0$, $x\leqslant 0$. Отсюда и из (51)

$$p_{\chi_n^2}(x) = F'(x) = \begin{cases} \dfrac{1}{2^{n/2}\Gamma\left(\frac{n}{2}\right)} x^{\frac{n}{2}-1} e^{-\frac{x}{2}}, & x>0, \\ 0, & x \leqslant 0. \end{cases} \qquad (52)$$

Пользуясь (52), найдем $M\chi_n^2$ и $D\chi_n^2$:

$$M\chi_n^2 = \frac{1}{2^{n/2}\Gamma\left(\frac{n}{2}\right)} \int\limits_0^{\infty} x^{\frac{n}{2}} e^{-\frac{x}{2}} dx = \frac{2^{\frac{n}{2}+1}}{2^{n/2}\Gamma\left(\frac{n}{2}\right)} \Gamma\left(\frac{n}{2}+1\right) = n, \qquad (53)$$

поскольку $\Gamma(s+1) = s\Gamma(s)$;

$$D\chi_n^2 = M(\chi_n^2)^2 - (M\chi_n^2)^2 = M(\chi_n^2)^2 - n^2,$$

и, так как

$$M(\chi_n^2)^2 = \frac{1}{2^{n/2}\Gamma\left(\frac{n}{2}\right)} \int\limits_0^{\infty} x^{\frac{n}{2}+1} e^{-\frac{x}{2}} dx =$$

$$= \frac{2^{\frac{n}{2}+2}}{2^{n/2}\Gamma\left(\frac{n}{2}\right)} \Gamma\left(\frac{n}{2}+2\right) = n^2 + 2n, \qquad (54)$$

то

$$D\chi_n^2 = 2n. \qquad (55)$$

2. **t_n-распределением** с n степенями свободы называется распределение случайной величины $t_n = \xi/\eta$, где $\xi \in N(0, 1)$, $\eta = \sqrt{\chi_n^2/n}$, χ_n^2 определено в 1) и ξ и η независимы. В силу последнего условия случайный вектор (ξ, η) имеет плотность $p(x_1, x_2) = p_\xi(x_1) p_\eta(x_2)$, а плотность частного ξ/η определяется на основании формулы (65) § 8

$$p_{t_n}(x) = F'_{t_n}(x) = \int\limits_0^{\infty} z p_\xi(zx) p_\eta(z)\, dz - \int\limits_{-\infty}^{0} z p_\xi(zx) p_\eta(z)\, dz. \qquad (56)$$

Далее,

$$F_\eta(x) = P\left\{\sqrt{\frac{\chi_n^2}{n}} < x\right\} = P\{\chi_n^2 < x^2 n\} = F_{\chi_n^2}(x^2 n).$$

Поэтому согласно (53) плотность распределения η равна

$$p_\eta(x) = F'_\eta(x) = p_{\chi_n^2}(x^2 n) \cdot 2xn =$$

$$= \begin{cases} \dfrac{n^{\frac{n}{2}}}{2^{\frac{n}{2}-1}\Gamma\left(\frac{n}{2}\right)} x^{n-1} e^{-\frac{x^2 n}{2}}, & x > 0 \\ 0, & x \leqslant 0 \end{cases} \qquad (57)$$

Ввиду (57) второй интеграл в (56) исчезает, и мы находим

$$p_{t_n}(x) = \frac{n^{\frac{n}{2}}}{2^{\frac{n}{2}-1}\Gamma\left(\frac{n}{2}\right)\sqrt{2\pi}} \int\limits_0^{\infty} z^n e^{-\frac{(zx)^2}{2} - \frac{nz^2}{2}} dz =$$

$$= \frac{n^{\frac{n}{2}} 2^{\frac{n}{2}-\frac{1}{2}}}{2^{\frac{n}{2}-1}\Gamma\left(\frac{n}{2}\right)\sqrt{2\pi}} \frac{1}{(x^2+n)^{\frac{n}{2}+\frac{1}{2}}} \int\limits_0^{\infty} t^{\frac{n-1}{2}} e^{-t} dt =$$

$$=\frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)}\frac{1}{\sqrt{\pi n}}\left(1+\frac{x^2}{n}\right)^{\frac{-(n+1)}{2}}. \tag{58}$$

С помощью (58) вычисляем Mt_n, $n\geqslant 2$,

$$Mt_n=\frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)}\frac{1}{\sqrt{\pi n}}\int_{-\infty}^{\infty}x\left(1+\frac{x^2}{n}\right)^{-\frac{n+1}{2}}dx=0 \tag{59}$$

ввиду нечетности подынтегральной функции. Теперь

$$Dt_n=Mt_n^2=M\left(\frac{\xi^2}{\chi_n^2/n}\right)=M\xi^2\cdot M\left(\frac{n}{\chi_n^2}\right), \tag{60}$$

так как ξ^2 и χ_n^2/n независимы. Обозначим $\eta=n/\chi_n^2$, имеем при $x>0$

$$F_\eta(x)=P\left\{\frac{n}{\chi_n^2}<x\right\}=P\left\{\frac{1}{x}<\frac{\chi_n^2}{n}\right\}=1-P\left\{\chi_n^2\leqslant\frac{n}{x}\right\}.$$

Отсюда, используя (52) с n/x вместо x, найдем

$$p_\eta(x)=\begin{cases}\left(\frac{n}{2}\right)^{\frac{n}{2}}\frac{1}{\Gamma\left(\frac{n}{2}\right)}e^{-\frac{n}{2x}}x^{-\left(\frac{n}{2}+1\right)}, & x>0,\\ 0, & x\leqslant 0.\end{cases} \tag{61}$$

Поэтому

$$M\left(\frac{n}{\chi_n^2}\right)=M\eta=\left(\frac{n}{2}\right)^{n/2}\frac{1}{\Gamma\left(\frac{n}{2}\right)}\int_0^\infty x^{-\frac{n}{2}}e^{-\frac{n}{2x}}dx=$$

$$=\frac{1}{\Gamma\left(\frac{n}{2}\right)}\frac{n}{2}\int_0^\infty t^{\frac{n}{2}-2}e^{-t}dt=\frac{n}{n-2}. \tag{62}$$

Сопоставляя (60) и (62) и учитывая, что $M\xi^2=D\xi=1$, получим

$$Dt_n=\frac{n}{n-2}. \tag{63}$$

3. $\mathbf{F}_{k,m}$**-распределением** с (k, m) степенями свободы называется распределение случайной величины $F_{k,m}=\frac{\chi_k^2}{k}\Big/\frac{\chi_m^2}{m}$,

где χ_k^2, χ_m^2 определены в 1) и независимы. Положим для краткости $F_{k,m}=\eta$, $\chi_k^2/k=\xi_1$, $\chi_m^2/m=\xi_2$. Ввиду независимости ξ_1 и ξ_2 плотность случайного вектора (ξ_1, ξ_2) равна $p(x_1, x_2)=p_{\xi_1}(x_1)p_{\xi_2}(x_2)$, а плотность частного ξ_1/ξ_2 равна (см. (56))

$$p_\eta(x)=\int_0^\infty zp_{\xi_1}(zx)p_{\xi_2}(z)\,dz-\int_{-\infty}^0 zp_{\xi_1}(zx)p_{\xi_2}(z)\,dz.$$

Отсюда с учетом (53) находим при $x>0$

$$p_\eta(x)=\frac{k^{\frac{k}{2}-1}m^{\frac{m}{2}-1}x^{\frac{k}{2}-1}}{2^{\frac{k+m}{2}}\Gamma\left(\frac{k}{2}\right)\Gamma\left(\frac{m}{2}\right)}\int_0^\infty z^{\frac{k+m}{2}-1}e^{-\frac{z}{2}(kx+m)}dz=$$

$$=\frac{k^{\frac{k}{2}-1}m^{\frac{m}{2}-1}x^{\frac{k}{2}-1}(kx+m)^{-\frac{k+m}{2}}}{\Gamma\left(\frac{k}{2}\right)\Gamma\left(\frac{m}{2}\right)}\Gamma\left(\frac{k+m}{2}\right), \tag{64}$$

и $p_\eta(x)=0$ при $x\leqslant 0$.

Далее, в силу (54) и (62)

$$M\eta=M\left(\frac{\chi_k^2/k}{\chi_m^2/m}\right)=M\left(\frac{\chi_k^2}{k}\right)M\left(\frac{m}{\chi_m^2}\right)=\frac{k}{k}\cdot\frac{m}{m-2}=\frac{m}{m-2}. \tag{65}$$

Найдем $D\eta$:

$$D\eta=M\eta^2-(M\eta)^2=M\eta^2-\left(\frac{m}{m-2}\right)^2. \tag{66}$$

Но

$$M\eta^2=M\left(\frac{\chi_k^2}{k}\right)^2\cdot M\left(\frac{m}{\chi_m^2}\right)^2, \tag{67}$$

где

$$M\left(\frac{\chi_k^2}{k}\right)^2=\frac{k^2+2k}{k^2} \tag{68}$$

ввиду (54). Используя (61), находим

$$M\left(\frac{m}{\chi_m^2}\right)^2=\left(\frac{m}{2}\right)^{\frac{m}{2}}\frac{1}{\Gamma\left(\frac{m}{2}\right)}\int_0^\infty x^{-\frac{m}{2}+1}e^{-\frac{m}{2x}}dx=$$

$$= \left(\frac{m}{2}\right)^{\frac{m}{2}} \frac{1}{\Gamma\left(\frac{m}{2}\right)} \left(\frac{m}{2}\right)^{-\frac{m}{2}+2} \int_0^\infty e^{-z} z^{\frac{m}{2}-3} dz =$$

$$= \left(\frac{m}{2}\right)^2 \frac{\Gamma\left(\frac{m}{2}-2\right)}{\Gamma\left(\frac{m}{2}\right)} = \frac{m^2}{(m-4)(m-2)}. \qquad (69)$$

Собирая выражения (66)—(69), получим при $m>4$

$$D\eta = \frac{2m^2}{(m-2)^2(m-4)}\left(1+\frac{m-2}{k}\right). \qquad (70)$$

§ 10. ЗАКОНЫ БОЛЬШИХ ЧИСЕЛ

Стандартное теоретико-вероятностное заключение — вероятность события A равна p — не позволяет, как правило, предсказать, произойдет событие A или нет. Исключение составляют лишь те случаи, когда p либо очень мало, либо очень близко к единице. При этом можно утверждать, что событие A практически невозможно или соответственно практически достоверно. Так, например, если подбрасывать монету 1000 раз, то событие, состоящее в выпадении герба все 1000 раз, можно считать практически невозможным, а событие, состоящее в том, что герб выпадет хотя бы один раз,— практически достоверным.

Мы рассмотрим ряд результатов теории вероятностей, известных под названием «законов больших чисел» и позволяющих делать подобные предсказания. Всякое утверждение о малости некоторой величины естественно формулировать в терминах предельного перехода. Мы введем здесь два понятия сходимости случайных величин: сходимость по вероятности и сходимость с вероятностью единица (или почти наверное). Разумеется, если речь идет о последовательности случайных величин, то предполагается, что все они определены на одном и том же вероятностном пространстве.

Определение 1. Последовательность случайных величин $\{\xi_n\}$ **сходится по вероятности** к случайной величине ξ, если для любого $\varepsilon>0$

$$\lim_{n\to\infty} P\{|\xi_n-\xi|>\varepsilon\}=0; \qquad (1)$$

обозначение $\xi_n \xrightarrow[p]{} \xi$, $n\to\infty$.

Этот тип сходимости означает, что, каково бы ни было $\varepsilon>0$, найдется число N, такое, что для всех $n\geqslant N$ вероятность неравенства $|\xi_n-\xi|>\varepsilon$ будет сколь угодно малой, или, иначе, событие $|\xi_n-\xi|>\varepsilon$ будет практически невозможным.

При этом тот или иной критерий «практической невозможности» обусловливает выбор соответствующего достаточно большого числа N.

Полезной в ряде вопросов (см. ч. 2) является следующая простая

Теорема 1. Пусть $\xi_n \xrightarrow[p]{} \xi$, а $g(x)$ — непрерывная функция, $x\in R_1$, так что $\eta=g(\xi)$ и $\eta_n=g(\xi_n)$ — случайные величины*, $n=1, 2, \ldots$. Тогда $\eta_n \xrightarrow[p]{} \eta$ при $n\to\infty$.

Доказательство. Пусть $\varkappa>0$ — любое, а $I\subset R_1$ — конечный интервал, такой, что $P\{\xi\overline{\in} I\}\leqslant\varkappa/2$. Функция $g(x)$ равномерно непрерывна на I, поэтому для любого $\varepsilon>0$ найдется такое $\delta=\delta(\varepsilon)>0$, что для всех x и y, таких, что $x\in I$ и $|x-y|<\delta$, $|g(y)-g(x)|<\varepsilon$. Таким образом,

$$|g(\xi_n)-g(\xi)|<\varepsilon, \text{ если } \xi\in I \text{ и } |\xi_n-\xi|<\delta. \qquad (2)$$

Поскольку $\xi_n \xrightarrow[p]{} \xi$ при $n\to\infty$, то для указанного δ

$$P\{|\xi_n-\xi|<\delta\}\geqslant 1-\varkappa/2, \qquad (3)$$

если $n\geqslant n_0=n_0(\delta,\varkappa)$. Из (2) и (3) следует, что

$$P\{|g(\xi_n)-g(\xi)|<\varepsilon\}\geqslant P\{|\xi_n-\xi|<\delta, \xi\in I\}\geqslant$$

$$\geqslant P\{|\xi_n-\xi|<\delta\}-P\{\xi\overline{\in} I\}\geqslant 1-\varkappa, n\geqslant n_0, \qquad (3')$$

здесь предпоследнее неравенство основано на соотношении

$$P\{|\xi_n-\xi|<\delta\}=P\{|\xi_n-\xi|<\delta, \xi\in I\}+$$

$$+P\{|\xi_n-\xi|<\delta, \xi\overline{\in} I\}\leqslant P\{|\xi_n-\xi|<\delta, \xi\in I\}+P\{\xi\overline{\in} I\}.$$

Неравенство (3′) ввиду произвольности ε и $\varkappa$ доказывает теорему. ▲

В этом параграфе мы будем постоянно использовать неравенство Чебышева (см. § 9, (28)): для любой случайной величины ξ с конечной дисперсией $D\xi$ и любого $\varepsilon>0$ справедливо неравенство

$$P\{|\xi-M\xi|>\varepsilon\}\leqslant D\xi/\varepsilon^2. \qquad (4)$$

Основываясь на неравенстве Чебышева, сформулируем удобный критерий сходимости по вероятности.

Лемма 1. Если для последовательности случайных величин $\{\xi_n\}$ $M\xi_n=0$, $D\xi_n\to 0$ при $n\to\infty$, то $\xi_n \xrightarrow[p]{} 0$.

Доказательство. В силу (4) и того, что $M\xi_n=0$, имеем для любого фиксированного $\varepsilon>0$

$$0\leqslant P\{|\xi_n|>\varepsilon\}\leqslant D\xi_n/\varepsilon^2\to 0, n\to\infty,$$

т. е. $\xi_n \xrightarrow[p]{} 0$, $n\to\infty$. ▲

* См. сноску § 9 п. 1°.

Переходим к установлению закона больших чисел в форме Чебышева.

Теорема 2 (Чебышев). Пусть $\xi_1, \xi_2, ..., \xi_n, ...$ — последовательность попарно независимых случайных величин, дисперсии которых ограничены в совокупности: $D\xi_i \leqslant c$, $i=1, 2, ...$. Тогда последовательность случайных величин $\eta_n = \frac{1}{n}\sum_{i=1}^{n}(\xi_i - M\xi_i)$ сходится по вероятности к нулю при $n \to \infty$

$$\lim_{n\to\infty} P\left\{\left|\frac{1}{n}\sum_{i=1}^{n}\xi_i - \frac{1}{n}\sum_{i=1}^{n}M\xi_i\right| > \varepsilon\right\} = 0, \varepsilon > 0 \text{ — любое.} \quad (5)$$

Доказательство. Имеем

$$M\eta_n = \frac{1}{n}\sum_{i=1}^{n}(M\xi_i - M\xi_i) = 0,$$

$$D\eta_n = \frac{1}{n^2}\sum_{i=1}^{n}D\xi_i \leqslant \frac{cn}{n^2} \to 0$$

при $n\to\infty$ (в силу попарной независимости ξ_i

$$D\sum_{i=1}^{n}(\xi_i - M\xi_i) = \sum_{i=1}^{n}D(\xi_i - M\xi_i) = \sum_{i=1}^{n}D\xi_i).$$

Отсюда на основании леммы 1 следует утверждение теоремы.

Теорему Чебышева можно записать и в виде

$$\lim_{n\to\infty} P\left\{\left|\frac{1}{n}\sum_{i=1}^{n}\xi_i - \frac{1}{n}\sum_{i=1}^{n}M\xi_i\right| \leqslant \varepsilon\right\} = 1, \varepsilon > 0 \text{ — любое.} \quad (6)$$

Следствие. Пусть в условиях теоремы Чебышева случайные величины ξ_i имеют одинаковые математические ожидания: $M\xi_1 = M\xi_2 = ... = \mu$. Тогда последовательность $\eta_n = \frac{1}{n}\sum_{i=1}^{n}\xi_i$ при $n\to\infty$ сходится по вероятности к математическому ожиданию μ:

$$\eta_n = \frac{1}{n}\sum_{i=1}^{n}\xi_i \underset{p}{\to} \mu, n \underset{p}{\to} \infty. \quad (7)$$

Это следствие теоремы Чебышева служит обоснованием правила среднего арифметического, применяемого в теории измерений, которое сводится к тому, что, повторив n раз измерение величины μ и получив в качестве результатов случайные величины $\xi_1, \xi_2, ..., \xi_n$, за приближенное значение μ принимают среднее арифметическое из наблюденных значений $\hat{\mu} = \frac{1}{n}(\xi_1 + \ldots + \xi_n)$. Если при измерениях отсутствует систематическая ошибка (т. е. все $M\xi_i = \mu$, $i=1, 2, ..., n$), то согласно закону больших чисел при достаточно больших n с вероятностью, сколь угодно близкой к 1, будет получен результат $\hat{\mu}$, произвольно мало отличающийся от истинного значения μ.

Важнейшим следствием закона больших чисел Чебышева является

Теорема 3 (Бернулли). Пусть η_n — число успехов в серии из n испытаний Бернулли и p — вероятность успеха при каждом испытании. Тогда последовательность частот $\{\eta_n/n\}$ при $n\to\infty$ сходится по вероятности к p.

Доказательство. Введем случайные величины ξ_k, равные числу успехов при k-м испытании, $k=1, 2, ...$. Тогда $\eta_n = \xi_1 + ... + \xi_n$, $M\xi_k = p$, $D\xi_k = pq$ (см. § 9). Поэтому согласно теореме Чебышева (условия которой, очевидно, выполнены для ξ_i, $i=1, 2, ...$) при любом $\varepsilon > 0$

$$\lim_{n\to\infty} P\left\{\left|\frac{\eta_n}{n} - p\right| > \varepsilon\right\} \equiv$$

$$\equiv \lim_{n\to\infty} P\left\{\left|\frac{1}{n}\sum_{k=1}^{n}\xi_k - \frac{1}{n}\sum_{k=1}^{n}M\xi_k\right| > \varepsilon\right\} = 0. \blacktriangle \quad (8)$$

В определенном смысле эта теорема может служить «аксиомой измерения», доставляя непротиворечивый способ практического определения тех вероятностей, о которых идет речь в аксиоматической теории вероятностей. Закон больших чисел Бернулли утверждает, что для фиксированного достаточно большого n очень правдоподобно, что частота η_n/n будет уклоняться от вероятности p меньше, чем на ε. Отсюда, однако, не следует, что разность $|\eta_n/n - p|$ останется малой для всех достаточно больших n. Может оказаться, что она принимает значения, близкие к единице. Теорема 3 гарантирует лишь, что эти большие отклонения могут появляться весьма редко. Для полного обоснования частотной интерпретации вероятности (см. § 1) желательно иметь теорему, обеспечивающую сходимость последовательности частот к вероятности. Мы сейчас введем некоторые новые понятия и дадим усиленный вариант теоремы Бернулли, удовлетворяющей этому требованию.

Определение 2. Последовательность случайных величин $\{\xi_n\}$ сходится к случайной величине ξ **с вероятностью 1** (или

почти наверное), если

$$P\{\omega\in\Omega:\lim_{n\to\infty}\xi_n(\omega)=\xi(\omega)\}=1, \tag{9}$$

т. е. $\xi_n(\omega)\to\xi(\omega)$ при $n\to\infty$ для всех $\omega\in\Omega$, за исключением, быть может, множества $C\subset\Omega$ нулевой вероятности, $P(C)=0$. Эта сходимость обозначается так: $\xi_n\to\xi$ п.н.

Согласно этому определению для каждого $\omega\in\Omega\setminus C$ и любого $\varepsilon>0$ $|\xi_n(\omega)-\xi(\omega)|\leqslant\varepsilon$ для всех достаточно больших n. Поэтому если обозначить через $A_{n,\varepsilon}$ событие $A_{n,\varepsilon}=\{\omega\in\Omega:$ $:|\xi_n(\omega)-\xi(\omega)|>\varepsilon\}$, $n=1, 2, \ldots$, то для любого $\varepsilon>0$ с вероятностью 1 происходит лишь конечное число событий $A_{n,\varepsilon}$. Оказывается, что это условие является и достаточным для сходимости с вероятностью 1. В самом деле, возьмем $\varepsilon=1/k$ и обозначим через B_k событие, состоящее в том, что происходит лишь конечное число событий из $A_{n,1/k}=\{\omega\in\Omega:|\xi_n(\omega)-$ $-\xi(\omega)|>1/k\}$, $n=1,2,\ldots$. По условию $P(B_k)=1$, $k=1,2,\ldots$ Очевидно, что события B_k, $k=1,2,\ldots$, образуют монотонно убывающую последовательность: $B_1\supset B_2\supset B_3\supset\ldots$. Обозначим через B событие $B=\bigcap\limits_{k=1}^{\infty}B_k$. В силу непрерывности вероятности $P(B)=\lim\limits_{k\to\infty}P(B_k)=1$, так как все $P(B_k)=1$. Из определения события B следует, что B состоит из всех таких $\omega\in\Omega$, для которых $\xi_n(\omega)\to\xi(\omega)$ при $n\to\infty$. Итак, $P(B)=1$, и высказанное выше утверждение доказано. Таким образом, $\xi_n\to\xi$ п.н. тогда и только тогда, когда для любого $\varepsilon>0$ вероятность того, что осуществляется лишь конечное число событий

$$|\xi_n-\xi|>\varepsilon,\quad n=1,2,\ldots, \tag{10}$$

равна 1.

Лемма 2 (Бореля—Кантелли). Если для последовательности $\{A_n\}$ произвольных событий A_n, $n=1,2,\ldots$, выполнено условие

$$\sum_{n=1}^{\infty}P(A_n)<\infty, \tag{11}$$

то с вероятностью 1 происходит лишь конечное число этих событий.

Доказательство. Пусть событие B_n состоит в том, что происходит хотя бы одно из событий A_k с $k\geqslant n$, т. е. $B_n=\bigcup\limits_{k=n}^{\infty}A_k$. Очевидно, что $B_1\supset B_2\supset\ldots$ Пусть, далее, событие B означает, что происходит бесконечное число событий из A_n, $n=1,2,\ldots$. Событие B наступает тогда и только тогда, когда происходят все B_n, т. е. $B=\bigcap\limits_{n=1}^{\infty}B_n$. Отсюда в силу того, что

$B_1\supset B_2\supset\ldots$ и непрерывности вероятности, получим:

$$P(B)=\lim_{n\to\infty}P(B_n). \tag{12}$$

Поскольку $B_n=\bigcup\limits_{k=n}^{\infty}A_k$, то

$$P(B_n)\leqslant\sum_{k=n}^{\infty}P(A_k). \tag{13}$$

Так как ряд (11) сходится, то его остаток $\sum\limits_{k=n}^{\infty}P(A_k)\to 0$ при $n\to\infty$ и в силу (13) $P(B_n)\to 0$ при $n\to\infty$. Отсюда и из (12) находим, что

$$P(B)=0.$$

Поэтому противоположное событие $\bar{B}$, состоящее в том, что наступает конечное число событий A_n, $n=1,2,\ldots$, имеет вероятность, равную 1, $P(\bar{B})=1$, что и требовалось доказать. ▲

Используя лемму Бореля—Кантелли, мы установим следующий усиленный вариант закона больших чисел.

Теорема 4 (усиленный закон больших чисел). Пусть $\xi_1, \xi_2, \ldots$ — последовательность попарно независимых случайных величин, для которых $M\xi_i=\mu$, $D\xi_i=\sigma^2$. Тогда при $n\to\infty$

$$\frac{1}{n}\sum_{i=1}^{n}\xi_i\to\mu \tag{14}$$

с вероятностью 1.

Доказательство. Вводя в случае необходимости новые случайные величины $\xi_i'=\xi_i-\mu$, можем считать, что $\mu=0$. Обозначим через η_k случайную величину

$$\eta_k=\sum_{i=1}^{k}\xi_i. \tag{15}$$

Нам надо доказать, что при $n\to\infty$ $(1/n)\eta_n\to 0$ п.н. Для каждого натурального n возьмем натуральное число m так, чтобы

$$m^2\leqslant n\leqslant(m+1)^2. \tag{16}$$

Так как $M\eta_k=0$, то неравенство Чебышева (4) дает

$$P\left\{\left|\frac{\eta_{m^2}}{m^2}\right|>\varepsilon\right\}\leqslant\frac{D\eta_{m^2}}{\varepsilon^2m^4}=\frac{\sigma^2}{\varepsilon^2m^2}. \tag{17}$$

Положим

$$\hat{\eta}_{m^2}=\max_{m^2+1\leqslant k\leqslant(m+1)^2}|\xi_{m^2}+\ldots+\xi_k|. \tag{18}$$

Снова применяя неравенство Чебышева (4), получим

$$P\left\{\left|\frac{\widehat{\eta}_{m^2}}{m^2}\right|>\varepsilon\right\}\leqslant\sum_{k=m^2+1}^{(m+1)^2}P\left\{\frac{|\xi_{m^2}+\ldots+\xi_k|}{m^2}>\varepsilon\right\}\leqslant$$

$$\leqslant\sum_{k=m^2+1}^{(m+1)^2}\frac{(k-m^2)\,\sigma^2}{\varepsilon^2m^4}\leqslant 2m\,\frac{(2m+1)\,\sigma^2}{\varepsilon^2m^4}\leqslant\frac{5\sigma^2}{\varepsilon^2m^2}\qquad(19)$$

(здесь в сумме $2m$ слагаемых и $(k-m^2)\leqslant 2m+1)$). В силу оценок (17) и (19) числовые ряды

$$\sum_{m=1}^{\infty}P\left\{\left|\frac{\eta_{m^2}}{m^2}\right|>\varepsilon\right\}\quad\text{и}\quad\sum_{m=1}^{\infty}P\left\{\left|\frac{\widehat{\eta}_{m^2}}{m^2}\right|>\varepsilon\right\}\qquad(20)$$

сходятся, а тогда на основании леммы Бореля—Кантелли заключаем, что с вероятностью 1 может произойти только конечное число событий $\left\{\left|\frac{\eta_{m^2}}{m^2}\right|>\varepsilon\right\}$ и $\left\{\left|\frac{\widehat{\eta}_{m^2}}{m^2}\right|>\varepsilon\right\}$, т. е. согласно критерию (10) с вероятностью 1

$$\frac{\eta_{m^2}}{m^2}\to 0\quad\text{и}\quad\frac{\widehat{\eta}_{m^2}}{m^2}\to 0\quad\text{при}\quad m\to\infty.\qquad(21)$$

Поскольку для любого n из интервала (16)

$$\left|\frac{\eta_n}{n}\right|\leqslant\left|\frac{\eta_{m^2}}{m^2}\right|+\left|\frac{\widehat{\eta}_{m^2}}{m^2}\right|,$$

то из (21) следует, что $\eta_n/n\to 0$ при $n\to\infty$ с вероятностью 1. ▲

Простым следствием доказанной теоремы является усиленный закон больших чисел Бернулли.

Теорема 5 (Борель). Пусть η_n — число успехов в серии из n независимых испытаний Бернулли, p — вероятность успеха при каждом испытании. Тогда последовательность частот $\{\eta_n/n\}$ при $n\to\infty$ сходится с вероятностью 1 к вероятности p.

Доказательство. Достаточно ввести случайные величины ξ_i, равные числу успехов в i-м испытании, $\xi_i=1, 0$, $M\xi_i=p$, $D\xi_i=pq$; $\eta_n=\sum\limits_{i=1}^{n}\xi_i$ и применить теорему 4 к η_n. ▲

Теперь мы рассмотрим один важный вариант закона больших чисел, принадлежащий Хинчину. В этом варианте не требуется существования (а тем более ограниченности в совокупности) дисперсий случайных величин.

Теорема 6 (Хинчин). Пусть одинаково распределенные случайные величины $\xi_1, \xi_2, \ldots$ попарно независимы и имеют

конечное математическое ожидание $M\xi_i=\mu$. Тогда при $n\to\infty$ $\eta_n=\frac{1}{n}\sum\limits_{k=1}^{n}\xi_k$ сходится по вероятности к μ, $\eta_n\xrightarrow[p]{}\mu$.

Доказательство. Проведем доказательство с помощью широко применяемого в теории вероятностей метода урезания: введем новые случайные величины $\bar{\xi}_k$ по формуле:

$$\bar{\xi}_k=\begin{cases}\xi_k, & \text{если } |\xi_k|\leqslant y,\\ 0, & \text{если } |\xi_k|>y,\end{cases}\qquad(22)$$

число $y>0$ будет выбрано позже. Помимо $\eta_n=\frac{1}{n}\sum\limits_{k=1}^{n}\xi_k$ рассмотрим $\bar{\eta}_n$ — среднее «урезанных» случайных величин, $\bar{\eta}_n=\frac{1}{n}\sum\limits_{k=1}^{n}\bar{\xi}_k$. Для любого $\varepsilon>0$ имеем неравенство с $\mu=M\xi_k$:

$$P\{|\eta_n-\bar{\mu}|>\varepsilon\}\leqslant P\{|\bar{\eta}_n-\bar{\mu}|>\varepsilon\}+P\{\eta_n\neq\bar{\eta}_n\},\qquad(23)$$

поскольку если наступает событие в левой части этого неравенства, то наступает хотя бы одно из событий в его правой части. Далее *,

$$\bar{\mu}=M\bar{\xi}_k=\int_{-y}^{y}x\,dF(x),\qquad(24)$$

$F(x)$ — функция распределения случайных величин ξ_k, $k=1, 2, \ldots$, поэтому

$$|\bar{\mu}-\mu|<\varepsilon,\qquad(25)$$

если $y\geqslant y_0=y_0(\varepsilon)$.

Ввиду (23) и (25) при $y\geqslant y_0$ имеем

$$P\{|\eta_n-\mu|>2\varepsilon\}\leqslant P\{|\bar{\eta}_n-\bar{\mu}|>\varepsilon\}+P\{\eta_n\neq\bar{\eta}_n\}.\qquad(26)$$

Оценим правую часть в (26). По неравенству Чебышева (4):

$$P\{|\bar{\eta}_n-\bar{\mu}|>\varepsilon\}\leqslant\frac{D\bar{\eta}_n}{\varepsilon^2}\leqslant\frac{1}{\varepsilon^2}M\bar{\eta}_n^2\leqslant\frac{1}{\varepsilon^2n^2}\sum_{k=1}^{n}M\bar{\xi}_k^2=\qquad(27)$$

$$=\frac{n}{\varepsilon^2n^2}\int_{-y}^{y}x^2dF(x)\leqslant\frac{y}{\varepsilon^2n}\int_{-y}^{y}|x|\,dF(x)\leqslant\frac{y}{\varepsilon^2n}\int_{-\infty}^{\infty}|x|\,dF(x)=\frac{Ay}{\varepsilon^2n},$$

* Здесь и ниже интеграл $\int x\,dF(x)$, как обычно, есть сокращенная запись для Σx_kp_k в дискретном случае и $\int xp(x)\,dx$ — в непрерывном.

где через A обозначена существующая по условию величина $A=\int\limits_{-\infty}^{\infty}|x|\,dF(x)$. Событие $\{\eta_n\neq\overline{\eta}_n\}$ состоит в том, что хотя бы одна случайная величина $|\xi_k|>y$, $k=1, 2, \ldots, n$. Поэтому для любого $\varepsilon_1>0$

$$P\{\eta_n\neq\overline{\eta}_n\}\leqslant\sum_{k=1}^{n}P\{|\xi_k|>y\}=n\int\limits_{|x|>y}dF(x)<$$

$$<n\frac{1}{y}\int\limits_{|x|>y}|x|\,dF(x)<\frac{n}{y}\varepsilon_1, \tag{28}$$

если $y\geqslant y_1=y_1(\varepsilon_1)$.

Подставляя (27) и (28) в (26), получим при $y\geqslant\bar{y}=\max\{y_0, y_1\}$

$$P\{|\eta_n-\mu|>2\varepsilon\}\leqslant\frac{Ay}{\varepsilon^2 n}+\frac{n}{y}\varepsilon_1. \tag{29}$$

Пусть теперь $\delta>0$ — любое, положим $y=(\varepsilon^2/A)\delta n$, a $\varepsilon_1=\delta^2/\varepsilon^2 A$. Тогда из (29) при всех достаточно больших n, $n\geqslant n_0(\delta)$, находим

$$P\{|\eta_n-\mu|>2\varepsilon\}<2\delta, \tag{30}$$

что и доказывает теорему. ▲

В заключение докажем следующую полезную теорему, в которой отсутствует предположение о попарной независимости случайных величин.

Теорема 7 (Марков). Если последовательность случайных величин $\xi_1, \xi_2, \ldots$ такова, что $M\xi_i=\mu$ и $\frac{1}{n^2}D\left(\sum_{i=1}^{n}\xi_i\right)\to 0$ при $n\to\infty$, то при $n\to\infty$ $\eta_n=\frac{1}{n}\sum_{i=1}^{n}\xi_i\to\mu$ по вероятности.

Доказательство. Положим $\zeta_n=\frac{1}{n}\sum_{i=1}^{n}\xi_i-\mu$. Тогда $M\zeta_n=0$, $D\zeta_n=\frac{1}{n^2}D\left(\sum_{i=1}^{n}\xi_i\right)\to 0$ при $n\to\infty$ и согласно лемме 1 получаем утверждение теоремы. ▲

Примеры.

1. Рассмотрим схему независимых испытаний: при каждом испытании полная группа событий состоит из $A_1, A_2, \ldots, A_r$, и вероятность наступления при каждом испытании события A_i равна p_i, $p_i\geqslant 0$, $p_1+\ldots+p_r=1$. Пусть $\eta_n^{(i)}$, $i=1, 2, \ldots, r$, случайная величина, равная числу наступлений события A_i в серии из n испытаний, тогда частота $\eta_n^{(i)}/n$ появлений события A_i при $n\to\infty$ сходится по вероятности к p_i, $i=1, 2, \ldots, r$. В самом деле, пусть ξ_k — число наступлений A_i в k-м испытании, $\xi_k=0$ или 1 с вероятностями соответственно $1-p_i$ и p_i. Поэтому

$$M\xi_k=p_i,\ D\xi_k=M\xi_k^2-(M\xi_k)^2=p_i-p_i^2=p_i(1-p_i),\ \text{а}\ \eta_n^{(i)}=\sum_{k=1}^{n}\xi_k.$$

Применяя теорему Чебышева, получим

$$\frac{\eta_n^{(i)}}{n}\xrightarrow[p]{}p_i,\quad n\to\infty,\quad i=1, 2, \ldots, r. \tag{31}$$

2. Покажем, что для пуассоновского процесса (см. ниже § 13), т. е.

$$P\{\xi(t)=k\}=\frac{(\lambda t)^k e^{-\lambda t}}{k!},\quad k=0, 1, 2, \ldots,\quad \lambda\geqslant 0,\quad t\geqslant 0,$$

имеет место сходимость по вероятности:

$$\frac{\xi(t)}{t}\xrightarrow[p]{}\lambda\quad \text{при}\ t\to\infty.$$

В самом деле, мы видели в § 9, что

$$M\xi(t)=\lambda t,\quad D\xi(t)=\lambda t.$$

Поэтому для случайной величины $\xi(t)/t$ с помощью неравенства Чебышева (4) получим при любом $\varepsilon>0$

$$P\left\{\left|\frac{\xi(t)}{t}-\lambda\right|>\varepsilon\right\}\leqslant\frac{D\left(\frac{\xi(t)}{t}\right)}{\varepsilon^2}=\frac{\lambda}{t\varepsilon^2}\to 0\quad \text{при}\ t\to\infty,$$

что и доказывает высказанное утверждение.

§ 11. ЦЕНТРАЛЬНЫЕ ПРЕДЕЛЬНЫЕ ТЕОРЕМЫ

Утверждения, полученные в форме законов больших чисел, представляют собой заключения о сходимости последовательности случайных величин $\{\xi_n\}$, $n=1, 2, \ldots$, к некоторой случайной (или неслучайной) величине ξ. Эти утверждения не дают нам никакой информации о том, как аппроксимировать распределение случайных величин ξ_n при больших n. Ответ на этот вопрос доставляют так называемые центральные предельные теоремы, в которых речь идет о новом виде сходимости последовательности случайных величин — сходимости по распределению. Основным аппаратом, используемым при изучении центральных предельных теорем, является аппарат характеристических функций, играющий важную роль и в других разделах теории вероятностей.

1°. Характеристические функции

Хорошо известна большая роль теории преобразования Фурье в анализе и в дифференциальных уравнениях. Характеристические функции — инструмент для использования преобразования Фурье в теории вероятностей.

Определение 1. Характеристической функцией (х.ф.) случайной величины ξ называется функция $f_\xi(t)$ вещественной переменной t, $t \in R_1$, определенная равенством

$$f_\xi(t) = Me^{it\xi} = \int_{-\infty}^{\infty} e^{itx}\, dF_\xi(x), \quad -\infty < t < \infty. \tag{1}$$

Вообще если ζ — комплексная случайная величина, $\zeta = \xi + i\eta$, где ξ и η — действительные случайные величины *, то по определению

$$M\zeta = M\xi + iM\eta.$$

В определении (1) интеграл понимается либо как сумма абсолютно сходящегося ряда

$$f_\xi(t) = \sum_{k=1}^{\infty} e^{itx_k} p_k, \tag{2}$$

если ξ — дискретная случайная величина, x_k — ее значения, а p_k — соответствующие вероятности, $k = 1, 2, \ldots$, либо как абсолютно сходящийся интеграл

$$f_\xi(t) = \int_{-\infty}^{\infty} e^{itx}\, p(x)\, dx, \tag{3}$$

если ξ — непрерывная случайная величина с плотностью $p(x)$. Хотя интеграл (1) представляет собой обычный интеграл Стильтьеса, мы ниже не опираемся на его специфические свойства и будем рассматривать (1) как краткую запись для выражений (2) и (3). Все дальнейшие рассуждения этого параграфа проводятся таким образом, что они одинаково применимы для случаев (2) и (3), ввиду чего мы будем использовать лишь обозначение (1).

Характеристическая функция существует для любой случайной величины, поскольку ввиду равенства $|e^{itx}| = 1$ ряд (2) и интеграл (3) сходятся абсолютно.

Очевидно, $f_\xi(0) = 1$, и $|f_\xi(t)| \leqslant 1$, $t \in R_1$.

---

* Относительно совместного распределения случайных величин ξ и η не делается никаких предположений.

Теорема 1. Пусть $\xi_1, \xi_2, \ldots, \xi_n$ — независимые в совокупности случайные величины. Тогда

$$f_{\xi_1+\ldots+\xi_n}(t) = f_{\xi_1}(t) \ldots f_{\xi_n}(t). \tag{4}$$

Доказательство.

$$f_{\xi_1+\ldots+\xi_n}(t) \equiv Me^{it(\xi_1+\ldots+\xi_n)} = M(e^{it\xi_1} \ldots e^{it\xi_n}) =$$

$$= Me^{it\xi_1} Me^{it\xi_2} \ldots Me^{it\xi_n} = f_{\xi_1}(t) f_{\xi_2}(t) \ldots f_{\xi_n}(t).$$

Здесь мы воспользовались теоремой о математическом ожидании произведения независимых случайных величин. Независимость $e^{it\xi_1}, \ldots, e^{it\xi_n}$ следует из независимости $\xi_1, \ldots, \xi_n$ (см. § 8) (тот факт, что $\eta_k = e^{it\xi_k}$ — комплексные случайные величины, очевидно, не имеет значения). ▲

В доказанной теореме сформулировано основное свойство характеристических функций, которое используется при доказательстве центральных предельных теорем.

Теорема 2.

$$f_{\sigma\xi+\mu}(t) = e^{it\mu} f_\xi(\sigma t), \quad \sigma, \mu — \text{const}.$$

Доказательство.

$$f_{\sigma\xi+\mu}(t) = Me^{it(\sigma\xi+\mu)} = e^{it\mu} Me^{it\sigma\xi} = e^{it\mu} f_\xi(\sigma t).$$

Теорема 3. Если существует момент $M\xi^k$, то k-я производная $f_\xi^{(k)}(t)$ х.ф. $f_\xi(t)$ существует, равномерно непрерывна на R_1 и

$$f_\xi^{(k)}(0) = i^k M\xi^k. \tag{5}$$

Доказательство. Имеем

$$f_\xi^{(k)}(0) = \int_{-\infty}^{\infty} e^{itx} (ix)^k\, dF_\xi(x) \big|_{t=0} = i^k M\xi^k. \tag{6}$$

Дифференцирование под знаком интеграла законно, так как из существования $M\xi^k$ следует, что $\int_{-\infty}^{\infty} |x|^k dF_\xi(x) < \infty$, и, следовательно, интеграл $\int_{-\infty}^{\infty} x^k e^{itx} dF_\xi(x)$ сходится (абсолютно) равномерно по t, $t \in R_1$. Докажем равномерную непрерывность на R_1 k-й производной $f_\xi^{(k)}(t)$

$$|f_\xi^{(k)}(t+h) - f_\xi^{(k)}(t)| \leqslant \int_{-\infty}^{\infty} |x|^k |e^{ihx} - 1|\, dF_\xi(x) =$$

$$= \int_{|x|>A} |x|^k |e^{ihx} - 1|\, dF_\xi(x) + \int_{-A}^{A} |x|^k |e^{ihx} - 1|\, dF_\xi(x) \leqslant$$

$$\leqslant 2 \int\limits_{|x|>A} |x|^k dF_\xi(x) + \int\limits_{-A}^{A} |x|^k |e^{ihx} - 1| \, dF_\xi(x). \tag{7}$$

Пусть $\varepsilon>0$ — любое. В силу существования момента k-го порядка выбором достаточно большого $A>0$ первый интеграл в правой части (7) может быть сделан меньше $\varepsilon/2$. Зафиксировав это A, возьмем $\delta>0$ столь малым, чтобы для всех h, $|h|\leqslant\delta$, второй интеграл в правой части (7) также был меньше $\varepsilon/2$ (что возможно ввиду оценки $\max\limits_{x\in[-A,A]} |e^{ihx}-1| = \max\limits_{x\in[-A,A]} \left|\int\limits_0^{hx} e^{it}dt\right| \leqslant |hA|$). Таким образом, $|f_\xi^{(k)}(t+h) - f_\xi^{(k)}(t)| < \varepsilon$, если $|h|\leqslant\delta$, сразу для всех $t\in R_1$. ▲

Из доказанной теоремы следует, в частности, что $f_\xi(t)$ равномерно непрерывна на R_1.

Теорема 4. Если характеристическая функция $f_\xi(t)$ абсолютно интегрируема на R_1, то случайная величина ξ непрерывна, а ее плотность вероятности $p(x)$ равна

$$p(x) = \frac{1}{2\pi} \int\limits_{-\infty}^{\infty} e^{-ixt} f_\xi(t)\, dt \tag{8}$$

и равномерно непрерывна на R_1.

Доказательство. Пусть x и y, $y>x$, любые два числа. Рассмотрим интеграл

$$j(x,y) = \frac{1}{2\pi} \int\limits_{-\infty}^{\infty} \frac{e^{-itx} - e^{-ity}}{it} f_\xi(t) dt = \tag{9}$$

$$= \frac{1}{2\pi} \lim_{A\to\infty} \int\limits_{-A}^{A} \frac{e^{-itx} - e^{-ity}}{it} f_\xi(t)\, dt.$$

Интеграл (9) сходится абсолютно ввиду абсолютной интегрируемости $f_\xi(t)$ и оценки

$$|e^{-itx} - e^{-ity}| = \left| \int\limits_{ty}^{tx} e^{-i\alpha} d\alpha \right| \leqslant |t(x-y)|. \tag{10}$$

Оценка (10) показывает также, что если $y=x+h$, $h>0$, то

$$j(x, x+h) \to 0 \text{ при } h\to 0. \tag{11}$$

Подставив в (9) вместо $f_\xi(t)$ выражение (1) и изменив порядок интегрирования, что возможно в силу абсолютной сходимости двойного интеграла, получим

$$j(x,y) = \frac{1}{2\pi} \lim_{A\to\infty} \int\limits_{-\infty}^{\infty} \left(\int\limits_{-A}^{A} \frac{e^{it(z-x)} - e^{it(z-y)}}{it} dt \right) dF_\xi(z) =$$

$$= \frac{1}{\pi} \lim_{A\to\infty} \int\limits_{-\infty}^{\infty} dF_\xi(z) \int\limits_0^A \left\{ \frac{e^{it(z-x)} - e^{-it(z-x)}}{2i} - \frac{e^{it(z-y)} - e^{-it(z-y)}}{2i} \right\} \frac{dt}{t} =$$

$$= \frac{1}{\pi} \int\limits_{-\infty}^{\infty} dF_\xi(z) \left[\int\limits_0^{\infty} \frac{\sin(z-x)t}{t} dt - \int\limits_0^{\infty} \frac{\sin(z-y)t}{t} dt \right]. \tag{12}$$

Последнее равенство в (12) есть следствие формулы:

$$\lim_{A\to\infty} j_0 \equiv \lim_{A\to\infty} \int\limits_{-\infty}^{\infty} F_\xi(z) \int\limits_A^{\infty} \frac{\sin\alpha t}{t} dt = 0, \ \alpha = z-x \text{ и } \alpha = z-y.$$

Для доказательства этой формулы будем считать, что x и y — точки непрерывности $F_\xi(z)$ (см. с. 62) и воспользуемся элементарной оценкой:

$$\left| \int\limits_A^{\infty} \frac{\sin\alpha t}{t} dt \right| \leqslant \begin{cases} c_0, & \text{если } A\geqslant 0, \alpha \text{ — любое;} \\ \dfrac{2}{\delta A}, & \text{если } A>0, |\alpha|\geqslant\delta>0. \end{cases}$$

Пусть $\alpha=z-x$ ($\alpha=z-y$ — аналогично). Имеем

$$|j_0| = \left| \int\limits_{|z-x|\leqslant\delta} dF_\xi(z) \int\limits_A^{\infty} \frac{\sin\alpha t}{t} dt + \int\limits_{|z-x|>\delta} dF_\xi(z) \int\limits_A^{\infty} \frac{\sin\alpha t}{t} dt \right| \leqslant$$

$$\leqslant c_0 [F_\xi(x+\delta) - F_\xi(x-\delta)] + \frac{2}{\delta A} \int\limits_{-\infty}^{\infty} dF_\xi(z) < \varepsilon,$$

ибо первое слагаемое $<\varepsilon/2$, если $\delta\leqslant\delta_0(\varepsilon)$, ввиду непрерывности F_ξ в точке x, а второе слагаемое $<\varepsilon/2$, если A достаточно велико

$$\left(A \geqslant 4\varepsilon^{-1}\delta_0^{-1}(\varepsilon), \ \int\limits_{-\infty}^{\infty} dF_\xi(z) = 1 \right).$$

Как известно,

$$\int\limits_0^{\infty} \frac{\sin\alpha t}{t} dt = \frac{\pi}{2} \operatorname{sgn}\alpha. \tag{13}$$

Поэтому квадратная скобка в правой части (12) равна нулю, если $z<x$ или $z>y$, и равна π, если $x<z<y$. Таким образом, получаем

$$j(x,y) = \int\limits_x^y dF_\xi(z). \tag{14}$$

Отсюда в силу (11) заключаем, что $F_\xi(x)$ — непрерывная функция x и (14) дает

$$j(x,y) = F_\xi(y) - F_\xi(x), \quad x, y, y>x \text{ — любые.} \tag{15}$$

Из формулы (15) вытекает равенство

$$\frac{F_\xi(x+h)-F_\xi(x-h)}{2h}=\frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{\sin th}{th}e^{-itx}f_\xi(t)\,dt. \qquad (16)$$

Покажем, что предел правой части (16) при $h\to 0$ равен

$$\frac{1}{2\pi}\int_{-\infty}^{\infty}e^{-itx}f_\xi(t)\,dt. \qquad (17)$$

В самом деле, пусть $\varepsilon>0$ — любое, имеем

$$\left|\int_{-\infty}^{\infty}\frac{\sin th}{th}e^{-itx}f_\xi(t)\,dt-\int_{-\infty}^{\infty}e^{-itx}f_\xi(t)\,dt\right|\leqslant$$

$$\leqslant 2\int_{|t|>A}|f_\xi(t)|\,dt+\int_{-A}^{A}\left|\frac{\sin th}{th}-1\right||f_\xi(t)|\,dt<\varepsilon, \qquad (18)$$

при $|h|\leqslant\delta(\varepsilon)$, поскольку первый интеграл в правой части (18) $<\varepsilon/4$, если $A>0$ достаточно велико (в силу абсолютной интегрируемости $f_\xi(t)$), а второй интеграл $<\varepsilon/2$, поскольку $|(\sin th)/th-1|\to 0$ при $h\to 0$ равномерно по $t\in[-A,A]$ при фиксированном A. Поэтому существует и предел левой части в (16), и мы получаем

$$p(x)=F'_\xi(x)=\frac{1}{2\pi}\int_{-\infty}^{\infty}e^{-itx}f_\xi(t)\,dt. \qquad (19)$$

Из представления (19) ввиду абсолютной интегрируемости $f_\xi(t)$ сразу следует равномерная непрерывность $p(x)$ при $x\in R_1$ (сравни с теоремой 2). ▲

Замечание. Из проведенного доказательства теоремы 4 легко вытекает следующее утверждение: если $F_\xi(x)$ — функция распределения, а $f_\xi(t)$ — х.ф. случайной величины ξ, то для любых точек непрерывности x и y функции $F_\xi(x)$ справедливо равенство

$$F_\xi(y)-F_\xi(x)=\frac{1}{2\pi}\lim_{A\to\infty}\int_{-A}^{A}\frac{e^{-itx}-e^{-ity}}{it}f_\xi(t)\,dt. \qquad (20)$$

Равенство (20) называется **формулой обращения**, оно позволяет находить функцию распределения по известной х.ф.

Рассмотрим несколько примеров х.ф.

1. Нормальное распределение $N(0,1)$ имеет х.ф.

$$f_\xi(t)=\int_{-\infty}^{\infty}e^{itx}\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}dx=\frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}\int_{-\infty}^{\infty}e^{-\frac{(it-x)^2}{2}}dx=e^{-\frac{t^2}{2}}. \qquad (21)$$

Если $\xi\in N(0,1)$, то $(\sigma\xi+\mu)\in N(\mu,\sigma^2)$ и согласно теореме 2

$$f_{\sigma\xi+\mu}(t)=e^{it\mu-\sigma^2t^2/2}.$$

2. Х.ф. распределения Пуассона

$$f_\xi(t)=\sum_{k=0}^{\infty}e^{itk}\frac{\lambda^k e^{-\lambda}}{k!}=e^{-\lambda}\sum_{k=0}^{\infty}(\lambda e^{it})^k\frac{1}{k!}=e^{-\lambda}e^{\lambda e^{it}}. \qquad (22)$$

3. Х.ф. биномиального распределения

$$f_\xi(t)=\sum_{k=0}^{n}e^{itk}C_n^kp^kq^{n-k}=\sum_{k=0}^{n}C_n^k(pe^{it})^kq^{n-k}=(pe^{it}+q)^n. \qquad (23)$$

В заключение этого пункта приведем для справок основные свойства многомерных характеристических функций.

Определение 2. Характеристической функцией $f_\xi(t_1,\dots,t_n)$, $(t_1,\dots,t_n)\in R_n$, **n-мерной случайной величины** $\xi=(\xi_1,\dots,\xi_n)$ называется функция

$$f_\xi(t_1,\dots,t_n)=Me^{i(t_1\xi_1+\dots+t_n\xi_n)}=$$

$$=\int_{-\infty}^{\infty}\dots\int e^{i\sum_{k=1}^{n}t_kx_k}dF_\xi(x_1,\dots,x_n). \qquad (24)$$

Из этого определения непосредственно вытекают следующие свойства х.ф. $f_\xi(t_1,\dots,t_n)$:

1) $f_\xi(0,\dots,0)=1$, $|f_\xi(t_1,\dots,t_n)|\leqslant 1$.

2) Если координаты $\xi_1,\dots,\xi_n$ случайной величины ξ независимы, то $f_\xi(t_1,\dots,t_n)=f_{\xi_1}(t_1)\dots f_{\xi_n}(t_n)$, ибо в этом случае $Me^{i(t_1\xi_1+\dots+t_n\xi_n)}=Me^{it_1\xi_1}\dots Me^{it_n\xi_n}$.

3) Х.ф. случайной величины $\eta=(\sigma_1\xi_1+\mu_1,\dots,\sigma_n\xi_n+\mu_n)$ равна

$$f_\eta(t_1,\dots,t_n)=e^{i\sum_{k=1}^{n}\mu_kt_k}f_\xi(\sigma_1t_1,\dots,\sigma_nt_n).$$

4) Х.ф. суммы координат случайной величины ξ равна

$$f_{\xi_1+\dots+\xi_n}(t)=f_\xi(t,\dots,t),\text{ ибо } f_{\xi_1+\dots+\xi_n}(t)=Me^{it(\xi_1+\dots+\xi_n)}.$$

В полной аналогии с одномерным случаем для х.ф. многомерных случайных величин могут быть доказаны формула обращения и теорема о непрерывности для х.ф. (см. ниже).

2°. Теорема о непрерывности для характеристических функций

Хорошо известно из анализа, что для финитных* функций справедливы формулы прямого и обратного преобразования Фурье: если $\varphi(x)\in C_0^\infty(R_1)$, то

$$\varphi(x)=\frac{1}{2\pi}\int_{-\infty}^{\infty}e^{itx}\tilde{\varphi}(t)\,dt, \tag{25}$$

где $\tilde{\varphi}(t)$ — преобразование Фурье функции $\varphi(x)$

$$\tilde{\varphi}(t)=\int_{-\infty}^{\infty}e^{-itx}\varphi(x)\,dx; \tag{26}$$

$\tilde{\varphi}(t)$ убывает на бесконечности быстрее любой степени и поэтому абсолютно интегрируема на R_1.

Лемма 1. Пусть $\varphi(x)$ — произвольная финитная функция, $\varphi\in C_0^\infty(R_1)$, ξ — случайная величина, а $f_\xi(t)$ — ее х.ф. Тогда

$$M\varphi(\xi)=\frac{1}{2\pi}\int_{-\infty}^{\infty}f_\xi(t)\tilde{\varphi}(t)\,dt. \tag{27}$$

Доказательство. Используя представление (25), будем иметь

$$M\varphi(\xi)=\int_{-\infty}^{\infty}\varphi(x)\,dF_\xi(x)=\frac{1}{2\pi}\int_{-\infty}^{\infty}dF_\xi(x)\int_{-\infty}^{\infty}e^{itx}\tilde{\varphi}(t)\,dt=$$

$$=\frac{1}{2\pi}\int_{-\infty}^{\infty}\tilde{\varphi}(t)\,dt\int_{-\infty}^{\infty}e^{itx}dF_\xi(x)=\frac{1}{2\pi}\int_{-\infty}^{\infty}\tilde{\varphi}(t)f_\xi(t)\,dt,$$

перестановка интегралов законна в силу абсолютной сходимости двойного интеграла. ▲

Лемма 2. Пусть последовательность $\{f_{\xi_n}(t)\}$ х.ф. случайных величин $\{\xi_n\}$ сходится при $n\to\infty$ к х.ф. случайной величины ξ равномерно по t в каждом конечном интервале $|t|\leqslant T$. Тогда для любой финитной функции $\varphi(x)$, $\varphi\in C_0^\infty(R_1)$

$$M\varphi(\xi_n)\to M\varphi(\xi) \text{ при } n\to\infty. \tag{28}$$

Доказательство. С помощью формулы (27) получим

$$|M\varphi(\xi_n)-M\varphi(\xi)|=\frac{1}{2\pi}\left|\int_{-\infty}^{\infty}[f_{\xi_n}(t)-f_\xi(t)]\tilde{\varphi}(t)\,dt\right|\leqslant$$

* Как известно, финитной называется бесконечно дифференцируемая функция с компактным носителем. Множество всех финитных на R_1 функций обозначается $C_0^\infty(R_1)$.

$$\leqslant\frac{1}{2\pi}\int_{-T}^{T}|f_{\xi_n}(t)-f_\xi(t)|\,|\tilde{\varphi}(t)|\,dt+\frac{1}{2\pi}\int_{|t|>T}|f_{\xi_n}(t)|\,|\tilde{\varphi}(t)|\,dt+$$

$$+\frac{1}{2\pi}\int_{|t|>T}|f_\xi(t)|\,|\tilde{\varphi}(t)|\,dt. \tag{29}$$

Два последних интеграла меньше $\varepsilon/3$, если $T\geqslant T_0(\varepsilon)>0$, ибо $|f_{\xi_n}(t)|\leqslant 1$, $|f_\xi(t)|\leqslant 1$, а $\int_{|t|>T}|\tilde{\varphi}(t)|\,dt<\frac{\varepsilon}{3}$, так как $\tilde{\varphi}(t)$ абсолютно интегрируема. Фиксировав такое T, видим, что первый интеграл также меньше $\varepsilon/3$, если $n\geqslant n_0(\varepsilon)$, так как $f_{\xi_n}(t)\to f_\xi(t)$ при $n\to\infty$ равномерно по t при $|t|\leqslant T$ согласно условию леммы. ▲

Теорема 5 (о непрерывности для характеристических функций). Пусть выполнены условия леммы 2, и пусть, кроме того, функция распределения $F_\xi(x)$ случайной величины ξ непрерывна. Тогда при $n\to\infty$

$$\lim_{n\to\infty}F_{\xi_n}(x)=F_\xi(x), \tag{30}$$

причем сходимость равномерна по x, $x\in R_1$.

Доказательство. Возьмем любой сегмент $[a,b]$ и любое $\varepsilon>0$. Существует финитная функция $\varphi(x)$, равная нулю вне $[a,b]$, равная 1 на $[a+\varepsilon, b-\varepsilon]$ и $0\leqslant\varphi(x)\leqslant 1$ в остальных точках.

Имеем при всех n

$$M\varphi(\xi_n)=\int_a^b\varphi(x)\,dF_{\xi_n}(x)\leqslant\int_a^b dF_{\xi_n}(x)=F_{\xi_n}(b)-F_{\xi_n}(a).$$

Отсюда, поскольку на основании леммы 2 $M\varphi(\xi_n)\to M\varphi(\xi)$ при $n\to\infty$, находим

$$\varliminf_{n\to\infty}[F_{\xi_n}(b)-F_{\xi_n}(a)]\geqslant\lim_{n\to\infty}M\varphi(\xi_n)=M\varphi(\xi)=$$

$$=\int_a^b\varphi(x)\,dF_\xi(x)\geqslant\int_{a+\varepsilon}^{b-\varepsilon}dF_\xi(x)=F_\xi(b-\varepsilon)-F_\xi(a+\varepsilon). \tag{31}$$

Аналогично, взяв финитную функцию $\bar{\varphi}(x)$, равную 1 на $[a,b]$ нулю вне $[a-\varepsilon, b+\varepsilon]$ и $0\leqslant\bar{\varphi}(x)\leqslant 1$ — в остальных точках, получим:

$$F_{\xi_n}(b)-F_{\xi_n}(a)=\int_a^b dF_{\xi_n}(x)\leqslant\int_{a-\varepsilon}^{b+\varepsilon}\bar{\varphi}(x)\,dF_{\xi_n}(x)=M\bar{\varphi}(\xi_n).$$

Отсюда, вновь используя лемму 2, находим

$$\overline{\lim_{n\to\infty}}\,[F_{\xi_n}(b)-F_{\xi_n}(a)]\leqslant\lim_{n\to\infty}M\bar\varphi(\xi_n)=M\bar\varphi(\xi)=$$

$$=\int_{a-\varepsilon}^{b+\varepsilon}\bar\varphi(x)\,dF_\xi(x)\leqslant\int_{a-\varepsilon}^{b+\varepsilon}dF_\xi(x)=F_\xi(b+\varepsilon)-F_\xi(a-\varepsilon).\quad(32)$$

Поскольку $F_\xi(x)$ непрерывна в каждой точке x, то, устремляя $\varepsilon\to0$ в (31) и (32), получим

$$\underline{\lim_{n\to\infty}}\,[F_{\xi_n}(b)-F_{\xi_n}(a)]=\overline{\lim_{n\to\infty}}\,[F_{\xi_n}(b)-F_{\xi_n}(a)]=F_\xi(b)-F_\xi(a).\quad(33)$$

Из совпадения верхнего и нижнего пределов в (33) следует существование предела и равенство

$$\lim_{n\to\infty}[F_{\xi_n}(b)-F_{\xi_n}(a)]=F_\xi(b)-F_\xi(a)\quad(34)$$

Из (34), в свою очередь, вытекает, что в каждой точке $x\in R_1$ $F_{\xi_n}(x)\to F_\xi(x)$ при $n\to\infty$. В самом деле, пусть $\delta>0$ — любое. Возьмем числа A и B так, чтобы

$$F_\xi(A)-F_\xi(B)\geqslant1-\delta\quad(35)$$

(числа A и B найдутся, ибо $F_\xi(+\infty)=1$, $F_\xi(-\infty)=0$ и $F_\xi(x)$ непрерывна). В силу (34) найдется такой номер $n_0(\delta)$, что при $n\geqslant n_0(\delta)$

$$F_{\xi_n}(A)-F_{\xi_n}(B)\geqslant1-2\delta.\quad(36)$$

Из (35) и (36) ввиду неравенства $0\leqslant F_\xi(A)$, $F_{\xi_n}(A)\leqslant1$ следует, что

$$F_\xi(B)\leqslant\delta,\ F_{\xi_n}(B)\leqslant2\delta.\quad(37)$$

Теперь

$$|F_\xi(x)-F_{\xi_n}(x)|\leqslant|[F_\xi(x)-F_\xi(B)]-$$
$$-[F_{\xi_n}(x)-F_{\xi_n}(B)]|+F_\xi(B)+F_{\xi_n}(B)\leqslant4\delta,\quad(38)$$

так как первое слагаемое справа $<\delta$ в силу (34), если $n\geqslant n_1(\delta)$, а для двух других слагаемых выполнено (37). В (38) $\delta>0$ — любое, так что мы доказали, что при каждом $x\in R_1$

$$\lim_{n\to\infty}F_{\xi_n}(x)=F_\xi(x).\quad(39)$$

Докажем, наконец, что сходимость в (39) равномерна по x на всей прямой $-\infty<x<\infty$. Пусть $\varepsilon>0$ — любое. Выберем

точки $x_1<x_2<\ldots<x_N$, $N=[1/\varepsilon]+1$, так, чтобы

$$F_\xi(x_{i+1})-F_\xi(x_i)\leqslant\varepsilon,\ F_\xi(x_1)\leqslant\varepsilon,\ 1-F_\xi(x_N)\leqslant\varepsilon,\ i=1,2,\ldots,N.\quad(40)$$

Точек x_i конечное число, поэтому в силу (39) найдется такое $n_2(\varepsilon)$, что при $n\geqslant n_2$

$$|F_\xi(x_j)-F_{\xi_n}(x_j)|\leqslant\varepsilon,\ n\geqslant n_2,\ j=1,2,\ldots,N.\quad(41)$$

Пусть x — произвольная точка R_1, тогда либо $x<x_1$, либо $x>x_N$, либо при некотором j $x_j\leqslant x\leqslant x_{j+1}$. В последнем случае

$$|F_\xi(x)-F_{\xi_n}(x)|\leqslant|F_\xi(x)-F_\xi(x_j)|+|F_\xi(x_j)-F_{\xi_n}(x_j)|+$$
$$+|F_{\xi_n}(x_j)-F_{\xi_n}(x)|\leqslant2\varepsilon+|F_{\xi_n}(x)-F_{\xi_n}(x_j)|\quad(42)$$

в силу (40), (41) и монотонности $F_\xi(x)$. Далее,

$$|F_{\xi_n}(x)-F_{\xi_n}(x_j)|\leqslant|F_{\xi_n}(x_{j+1})-F_{\xi_n}(x_j)|\leqslant|F_{\xi_n}(x_{j+1})-$$
$$-F_\xi(x_{j+1})|+|F_\xi(x_{j+1})-F_\xi(x_j)|+|F_\xi(x_j)-F_{\xi_n}(x_j)|\leqslant3\varepsilon.\quad(43)$$

Из (42) и (43)

$$|F_\xi(x)-F_{\xi_n}(x)|\leqslant5\varepsilon.$$

Аналогичная оценка верна, очевидно, и в случае $x<x_1$ или $x>x_N$ ввиду (39) и (40). Тем самым доказана равномерность сходимости в (39) на всей прямой R_1. ▲

Справедливо и обратное утверждение: если в каждой точке непрерывности $F_\xi(x)$ выполнено условие $F_\xi(x)=\lim\limits_{n\to\infty}F_{\xi_n}(x)$, то равномерно по t $f_\xi(t)=\lim\limits_{n\to\infty}f_{\xi_n}(t)$. Доказательства не приводим, так как эта теорема в дальнейшем не используется.

Из доказанной теоремы о непрерывности для х.ф. следует, конечно, что если у двух законов распределения совпадают характеристические функции, то совпадают и сами законы распределения (надо взять $f_{\xi_n}(t)\equiv f_\xi(t)$ при всех n). На этом основаны следующие

Применения.

1) Сумма $\eta=\xi_1+\xi_2$, где $\xi_1\in N(\mu_1,\sigma_1^2)$, $\xi_2\in N(\mu_2,\sigma_2^2)$ и независимы, имеет нормальное распределение $\eta\in N(\mu_1+\mu_2,\sigma_1^2+\sigma_2^2)$.

Доказательство. Имеем

$$f_{\xi_1}(t)=e^{i\mu_1t-\frac{\sigma_1^2t^2}{2}},\quad f_{\xi_2}(t)=e^{i\mu_2t-\frac{\sigma_2^2t^2}{2}},$$

и так как ξ_1 и ξ_2 независимы, то

$$f_\eta(t)=f_{\xi_1}(t)f_{\xi_2}(t)=e^{i(\mu_1+\mu_2)t-\frac{\sigma_1^2+\sigma_2^2}{2}t^2},$$

откуда и следует, что $\eta\in N(\mu_1+\mu_2,\ \sigma_1^2+\sigma_2^2)$. ▲

2) Сумма $\eta=\xi_1+\xi_2$, где ξ_1 и ξ_2 распределены по закону Пуассона с параметрами λ_1 и λ_2 и независимы, также распределена по закону Пуассона с параметром $\lambda_1+\lambda_2$.

Доказательство. Имеем

$$f_{\xi_1}(t)=e^{\lambda_1(e^{it}-1)},\quad f_{\xi_2}(t)=e^{\lambda_2(e^{it}-1)},$$

и так как ξ_1 и ξ_2 независимы, то

$$f_\eta(t)=f_{\xi_1}(t)f_{\xi_2}(t)=e^{(\lambda_1+\lambda_2)(e^{it}-1)}.$$

Отсюда и следует высказанное утверждение. ▲

3) Для распределения $\chi_n^2=\xi_1^2+\ldots+\xi_n^2$ мы ранее нашли х.ф.

$$f_{\chi_n^2}(t)=(1-2it)^{-n/2}.$$

Поскольку

$$f_{\xi_k^2}(t)=(1-2it)^{-1/2}=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}e^{itx^2}e^{-\frac{x^2}{2}}dx=$$

$$=\frac{1}{\sqrt{2\pi}}\int_0^\infty\frac{1}{y}e^{-\frac{y}{2}}e^{ity}dy,$$

то, судя по подынтегральному выражению, плотность распределения χ_n^2 имеет вид: $p_{\chi_n^2}(y)=cy^{n/2-1}e^{-y/2}$, $0\leqslant y<\infty$.

Множитель c определяется из условия нормировки $\int_0^\infty p_{\chi_n^2}(y)\,dy=1$ и равен $c=2^{-\frac{n}{2}}\frac{1}{\Gamma\left(\frac{n}{2}\right)}$ (см. также § 9).

3°. Центральные предельные теоремы

Предварительное понимание содержания центральных предельных теорем может быть получено следующим образом. Рассмотрим суммы

$$\eta_n=\xi_1+\ldots+\xi_n,\quad n=1, 2, \ldots, \tag{44}$$

независимых случайных величин, которые принимают целочисленные значения и все имеют одинаковые распределения $P\{\xi_i=m\}=p_m$, $m=0, \pm1, \ldots$ для всех i. Это распределение можно изобразить следующим образом (рис. 18): основание каждого прямоугольника равно 1, высота — p_m, так что площадь равна p_m и $\sum_{m=-\infty}^{\infty}p_m=1$.

В общем случае получим практически произвольный набор прямоугольников.

Рассмотрим теперь вместо η_n нормированные случайные величины

$$\eta_n^*=\frac{\eta_n-M\eta_n}{\sqrt{D\eta_n}}=\frac{\eta_n-\mu n}{\sigma\sqrt{n}}, \tag{45}$$

где $\mu=M\xi_i$ и $\sigma^2=D\xi_i$, $i=1, 2, \ldots$ Значениями случайной величины η_n^* являются числа $x_n(m)=(m-n\mu)/\sigma\sqrt{n}$, причем $P\{\eta_n^*=x_n(m)\}=P\{\eta_n=m\}$.

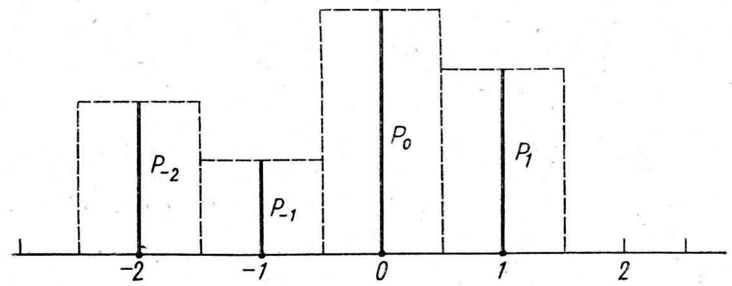


Рис. 18

Построим теперь аналогичный «график» распределения η_n^*. По оси абсцисс отложим значения $x_n(m)$, $m=0, \pm1, \ldots$, и, как и раньше, построим прямоугольники, площадь которых равна $P\{\eta_n^*=x_n(m)\}$. Поскольку длина основания теперь равна $1/\sigma\sqrt{n}$, то высоты этих прямоугольников должны быть равны $P\{\eta_n^*=x_n(m)\}\sigma\sqrt{n}=P\{\eta_n=m\}\sigma\sqrt{n}$. При достаточно большом n окажется, что верхние основания прямоугольников почти точно лягут на фиксированную кривую

$y=\frac{1}{\sqrt{2\pi}}e^{-x^2/2}$ (рис. 19), т. е.

$$\sigma\sqrt{n}\,P\{\eta_n^*=x_n(m)\}\approx\frac{1}{\sqrt{2\pi}}e^{-x_n^2(m)/2}\quad\text{при } n\to\infty.$$

При этом естественно, что

$$P\{a\leqslant\eta_n^*\leqslant b\}=\sum_{a\leqslant x_n(m)\leqslant b}P\{\eta_n^*=x_n(m)\}\approx$$

$$\approx\sum_{a\leqslant x_n(m)\leqslant b}\frac{1}{\sqrt{2\pi}}e^{-x_n^2(m)/2}\frac{1}{\sigma\sqrt{n}}\approx\frac{1}{\sqrt{2\pi}}\int_a^b e^{-x^2/2}dx, \tag{46}$$

где при замене суммы на интеграл мы считали $\Delta x=1/\sigma\sqrt{n}$. Этот факт и составляет, по существу, содержание центральных предельных теорем (которые отличаются друг от друга формулировками различных математических условий, обеспечивающих указанное выше утверждение). Мы рассмотрим вначале центральную предельную теорему в ее простейшей форме: для случая одинаково распределенных случайных величин, а затем приведем достаточно общую центральную предельную теорему Ляпунова.

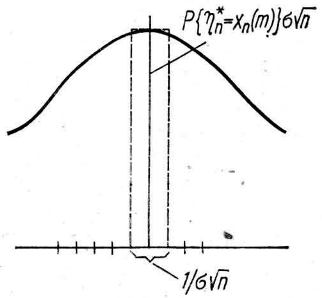


Рис. 19

Определение 3. Говорят, что последовательность случайных величин $\{\xi_k\}$, $k=1, 2, \ldots$, сходится к случайной величине ξ_0 **по распределению** или **слабо сходится**, если

$$\lim_{k\to\infty} F_k(x) = F_0(x) \tag{47}$$

в каждой точке x непрерывности $F_0(x)$, где $F_k(x)$ — функция распределения случайной величины ξ_k, $k=0, 1, 2, \ldots$.

С понятием сходимости по распределению мы уже встречались в теореме о непрерывности для характеристических функций.

Теорема 6 (центральная предельная теорема). Пусть $\{\xi_n\}$ — последовательность независимых и одинаково распределенных случайных величин, для которых $M\xi_n=\mu$, $D\xi_n=\sigma^2$, $\sigma>0$, $n=1, 2, \ldots$. Тогда последовательность

$$\eta_n = \frac{\sum\limits_{k=1}^{n}(\xi_k-\mu)}{\sigma\sqrt{n}} \tag{48}$$

сходится по распределению к $N(0, 1)$, т. е.

$$\lim_{n\to\infty} P\{\eta_n < x\} = \frac{1}{\sqrt{2\pi}}\int\limits_{-\infty}^{x} e^{-z^2/2}dz, \tag{49}$$

причем стремление к пределу в (49) равномерно по x на R_1.

Доказательство. Нам надо доказать, что равномерно по $x\in R_1$ $\lim\limits_{n\to\infty} F_n(x)=\Phi_0(x)$, где $F_n(x)$ — функция распределения случайной величины η_n, $\Phi_0(x)=\frac{1}{\sqrt{2\pi}}\int\limits_{-\infty}^{x} e^{-z^2/2}\,dz$ — функция распределения нормальной случайной величины $N(0, 1)$. В силу теоремы 5 достаточно доказать, что равномерно на каждом конечном интервале $|t|\leqslant T$, $T>0$, $f_n(t)\to e^{-t^2/2}$ при $n\to\infty$, где $f_n(t)$ — х.ф. случайной величины η_n, $e^{-t^2/2}$ — х.ф. нормального распределения. Поскольку η_n представляет собой сумму независимых случайных величин

$(\xi_k-\mu)/\sigma\sqrt{n}$, $k=1, \ldots, n$, то в силу теоремы 1 $f_n(t)=[\varphi(t/\sigma\sqrt{n})]^n$, где $\varphi(t)$ — х.ф. случайной величины $\xi_k-\mu$, $k=1, 2, \ldots$. По условию ξ_k имеет конечный момент 2-го порядка и вследствие теоремы 3 $\varphi''(t)$ непрерывна. Записывая для $\varphi(t)$ разложение по формуле Тейлора с остаточным членом в форме Пеано, будем иметь

$$\varphi\left(\frac{t}{\sigma\sqrt{n}}\right) = \varphi(0) + \varphi'(0)\frac{t}{\sigma\sqrt{n}} + \varphi''(0)\frac{t^2}{2\sigma^2 n} + o\left(\frac{t^2}{n\sigma^2}\right) =$$

$$= 1 - \frac{t^2}{2n} + o\left(\frac{t^2}{n\sigma^2}\right), \quad n\to\infty, \tag{50}$$

так как $\varphi(0)=1$, $\varphi'(0)=iM(\xi_k-\mu)=0$, $\varphi''(0)=-M(\xi_k-\mu)^2=-\sigma^2$. Отсюда

$$f_n(t) = \left[1 - \frac{t^2}{2n} + o\left(\frac{t^2}{n\sigma^2}\right)\right]^n \to e^{-t^2/2} \text{ при } n\to\infty \tag{51}$$

равномерно по t, $|t|\leqslant T$, $T>0$ — любое. ▲

Следствие (интегральная предельная теорема Муавра—Лапласа). Пусть ξ — число успехов в серии из n независимых испытаний, p — вероятность успеха при каждом испытании. Тогда при $n\to\infty$

$$\lim_{n\to\infty} P\left\{\frac{\xi-np}{\sqrt{npq}} < x\right\} = \frac{1}{\sqrt{2\pi}}\int\limits_{-\infty}^{x} e^{-z^2/2}\,dz, \tag{52}$$

причем стремление к пределу равномерно по $x\in R_1$.

Доказательство. Имеем $\xi=\xi_1+\ldots+\xi_n$, где ξ_k — число успехов при k-м испытании, так что ξ_k независимы и одинаково распределены, $P\{\xi_k=1\}=p$, $P\{\xi_k=0\}=q$, $M\xi_k=p$, $D\xi_k=pq$. Подставляя эти значения в (48), получим на основании теоремы 6 равенство (52). ▲

Естественно ожидать, что если в условиях теоремы 6 функция распределения $F(x)$ случайных величин ξ_k имеет плотность $p(x)$, то плотность $p_n(x)$ случайной величины η_n должна сходиться при $n\to\infty$ к плотности $p_0(x)$ нормального распределения. Вообще говоря, это неверно, но во всех практически интересных случаях высказанное утверждение имеет место, точнее, справедлива

Теорема 7. Пусть выполнены условия теоремы 6 и, кроме того, х.ф. $\varphi(t)$ случайных величин ξ_k абсолютно интегрируема на R_1. Тогда плотность $p_n(x)$ случайной величины $\eta_n=\sum\limits_{k=1}^{n}(\xi_k-\mu)/\sigma\sqrt{n}$ при $n\to\infty$ сходится к плотности $p_0(x)=\frac{1}{\sqrt{2\pi}}e^{-x^2/2}$ равномерно на R_1.

Доказательство. Мы видели в теореме 6, что $f_n(t)$ — х.ф. случайной величины η_n равна $f_n(t)=[\varphi(t/\sigma\sqrt{n})]^n$, где

$\varphi(t)$ — х.ф. случайной величины $\xi_k-\mu$, и, следовательно, также абсолютно интегрируема на R_1. Применяя теорему 4 к $f_n(t)$ и $e^{-t^2/2}$ — х.ф. нормального распределения, получим

$$p_n(x)-p_0(x)=\frac{1}{2\pi}\int_{-\infty}^{\infty}e^{-itx}\left[f_n(t)-e^{-t^2/2}\right]dt, \tag{53}$$

где $p_n(x)$ — плотность распределения случайной величины η_n.

Отсюда сразу для всех $x\in R_1$ получим

$$|p_n(x)-p_0(x)|\leqslant\frac{1}{2\pi}\int_{-\infty}^{\infty}|\varphi^n(t/\sigma\sqrt{n})-e^{-t^2/2}|\,dt. \tag{54}$$

Докажем, что правая часть здесь стремится к нулю при $n\to\infty$. Пусть $\varepsilon>0$ — любое. При доказательстве теоремы 6 было показано, что $|\varphi^n(t/\sigma\sqrt{n})-e^{-t^2/2}|\to 0$ при $n\to\infty$ равномерно в любом интервале $|t|\leqslant T$, $T>0$. Таким образом, для любого $T>0$ найдется такой номер $n_0=n_0(T)$, что при $n\geqslant n_0$

$$\frac{1}{2\pi}\int_{-T}^{T}|\varphi^n(t/\sigma\sqrt{n})-e^{-t^2/2}|\,dt<\frac{\varepsilon}{3}. \tag{55}$$

Поскольку $\varphi(0)=1$, $\varphi'(0)=0$ и $\varphi''(0)=-\sigma^2$ (см. теорему 6), а для функции $\psi(t)=e^{-\sigma^2t^2/4}$ имеем $\psi(0)=1$, $\psi'(0)=0$ и $\psi''(0)=-\sigma^2/2$, то существует такое $\delta>0$, что $|\varphi(t)|\leqslant e^{-\sigma^2t^2/4}$ при $|t|<\delta$. Таким образом,

$$|\varphi^n(t/\sigma\sqrt{n})|\leqslant e^{-t^2/4}\quad\text{при } |t|<\delta\sigma\sqrt{n}. \tag{56}$$

Поэтому, для части интеграла (54) по области $T\leqslant|t|\leqslant\delta\sigma\sqrt{n}$ будем иметь оценку

$$\frac{1}{2\pi}\left[\int_{T}^{\delta\sigma\sqrt{n}}\left|\varphi^n\left(\frac{t}{\sigma\sqrt{n}}\right)-e^{-\frac{t^2}{2}}\right|dt+\right.$$

$$\left.+\int_{-\delta\sigma\sqrt{n}}^{-T}\left|\varphi^n\left(\frac{t}{\sigma\sqrt{n}}\right)-e^{-\frac{t^2}{2}}\right|dt\right]\leqslant$$

$$\leqslant\frac{4}{2\pi}\int_{T}^{\delta\sigma\sqrt{n}}e^{-\frac{t^2}{4}}\,dt\leqslant\frac{2}{\pi}\int_{T}^{\infty}e^{-\frac{t^2}{4}}\,dt<\frac{\varepsilon}{3}, \tag{57}$$

если T достаточно велико.

Наконец, рассмотрим область $|t|\geqslant\delta\sigma\sqrt{n}$. В этой области аргумент функции $\varphi^n(t/\sigma\sqrt{n})$ не меньше $\delta>0$. Покажем, что существует число β, $0\leqslant\beta<1$, такое, что при $|\tau|\geqslant\delta$

$$|\varphi(\tau)|\leqslant\beta. \tag{58}$$

Действительно, в силу теоремы 4 существует непрерывная на R_1 плотность $p(x)$, такая, что

$$\varphi(t)=\int_{-\infty}^{\infty}e^{itx}p(x)\,dx. \tag{59}$$

Если в некоторой точке t $|\varphi(t)|=1$, т. е. $\varphi(t)=e^{i\alpha t}$, где α — некоторое действительное число, то

$$1=\int_{-\infty}^{\infty}e^{itx-i\alpha t}p(x)\,dx=\int_{-\infty}^{\infty}\cos(x-\alpha)t\,p(x)\,dx=$$

$$=\int_{-\infty}^{\infty}\cos yt\,p(y+\alpha)\,dy. \tag{60}$$

Поскольку $\int_{-\infty}^{\infty}p(y+\alpha)\,dy=1$, то последнее равенство можно переписать в виде

$$\int_{-\infty}^{\infty}(1-\cos yt)\,p(y+\alpha)\,dy=0. \tag{61}$$

Поскольку подынтегральная функция в (61) неотрицательна, то это равенство возможно лишь при $t=0$ (если $t\neq0$, то $p(y+\alpha)=0$ во всех точках $y\in R_1$, кроме, быть может, точек вида $y=2\pi k/t$, а тогда в силу непрерывности $p(y+\alpha)\equiv0$, что невозможно).

Итак, $|\varphi(t)|<1$ для всех $t\in R_1$, кроме $t=0$. Поскольку в силу известной леммы Римана—Лебега [5]

$$\lim_{t\to\infty}\varphi(t)=\lim_{t\to\infty}\int_{-\infty}^{\infty}e^{itx}p(x)\,dx=0,$$

то отсюда следует утверждение (58). Оценим теперь интеграл (54) по оставшейся части $|t|\geqslant\delta\sigma\sqrt{n}$. Имеем в силу (58)

$$\frac{1}{2\pi}\int_{|t|\geqslant\delta\sigma\sqrt{n}}\left|\varphi^n\left(\frac{t}{\sigma\sqrt{n}}\right)-e^{-\frac{t^2}{2}}\right|dt\leqslant$$

$$\leqslant\frac{\sigma\sqrt{n}}{2\pi}\int_{|\tau|\geqslant\delta}|\varphi^n(\tau)|\,d\tau+\frac{1}{2\pi}\int_{|t|\geqslant\delta\sigma\sqrt{n}}e^{-\frac{t^2}{2}}\,dt\leqslant$$

$$\leqslant \frac{\sigma\sqrt{n}}{2\pi}\beta^{n-1}\int\limits_{|\tau|\geqslant\delta}|\varphi(\tau)|\,d\tau+\frac{1}{2\pi}\int\limits_{|t|\geqslant\delta\sigma\sqrt{n}}e^{-\frac{t^2}{2}}dt<\frac{\varepsilon}{3}, \tag{62}$$

если n достаточно велико. Сопоставляя оценки (55), (57), (62), завершим доказательство теоремы. ▲

Теперь мы избавимся от нежелательного в ряде вопросов предположения об одинаковом распределении случайных величин ξ_k и докажем центральную предельную теорему в форме Ляпунова.

Теорема 8 (центральная предельная теорема Ляпунова). Пусть $\{\xi_n\}$ — последовательность независимых случайных величин с $M\xi_n=\mu_n$, $D\xi_n=\sigma_n^2$ и $M|\xi_n-\mu_n|^3<\infty$, $n=1, 2, \ldots$. Тогда, если выполнено условие Ляпунова

$$\frac{1}{B_n^3}\sum_{k=1}^{n}M|\xi_k-\mu_k|^3\to 0 \quad \text{при } n\to\infty, \tag{63}$$

где $B_n=\sqrt{\sum\limits_{k=1}^{n}D\xi_k}\equiv\sqrt{\sum\limits_{k=1}^{n}\sigma_k^2}$, то последовательность

$$\eta_n=\frac{\sum\limits_{k=1}^{n}(\xi_k-\mu_k)}{B_n} \tag{64}$$

сходится по распределению к $N(0, 1)$ равномерно на R_1.

Доказательство. Положим

$$\zeta_{nk}=\frac{1}{B_n}(\xi_k-\mu_k),\ k=1, 2, \ldots, n,\ f_{nk}(t) — \text{х. ф. } \zeta_{nk}. \tag{65}$$

Тогда

$$\eta_n=\sum_{k=1}^{n}\zeta_{nk},\quad M\zeta_{nk}=\frac{1}{B_n}M(\xi_k-\mu_k)=0,\ D\zeta_{nk}=\frac{1}{B_n^2}D\xi_k,$$

$$M|\zeta_{nk}|^3=\frac{1}{B_n^3}M|\xi_k-\mu_k|^3. \tag{66}$$

Если $f_n(t)$ — х.ф. η_n, то, поскольку

$$f_n(t)=\prod_{k=1}^{n}f_{nk}(t), \tag{67}$$

достаточно в силу теоремы 5 доказать, что при $n\to\infty$

$$\prod_{k=1}^{n}f_{nk}(t)\to e^{-\frac{t^2}{2}} \tag{68}$$

равномерно по t, $|t|\leqslant T$, $T>0$ — любое фиксированное. Запишем на $|t|\leqslant T$ формулу Тейлора для f_{nk} с остаточным членом Лагранжа:

$$f_{nk}(t)=1-\frac{D\xi_k}{B_n^2}\frac{t^2}{2}+R_{nk}(t), \tag{69}$$

ибо

$$f_{nk}(0)=1,\ f'_{nk}(0)=iM\zeta_{nk}=0,\ f''_{nk}(0)=-M\zeta_{nk}^2=\frac{D\xi_k}{B_n^2},$$

а для остатка $R_{nk}(t)$ имеем оценку при $k=1, \ldots, n$

$$|R_{nk}(t)|\leqslant\frac{T^3}{6}\max_{|t|\leqslant T}|f'''_{nk}(t)|\leqslant\frac{T^3}{6}M|\zeta_{nk}|^3=$$

$$=\frac{T^3}{6}\frac{1}{B_n^3}M|\xi_k-\mu_k|^3\leqslant$$

$$\leqslant\frac{T^3}{6}\frac{1}{B_n^3}\sum_{k=1}^{n}M|\xi_k-\mu_k|^3\to 0,\ n\to\infty \tag{70}$$

в силу условия (63).

Далее, имеем оценку при любом $\varepsilon>0$ и $k=1, \ldots, n$

$$\frac{D\xi_k}{B_n^2}=D\zeta_{nk}=\int_{-\infty}^{\infty}x^2\,dF_{nk}(x)\leqslant$$

$$\leqslant\varepsilon^2\int\limits_{|x|\leqslant\varepsilon}dF_{nk}(x)+\frac{1}{\varepsilon}\int\limits_{|x|\geqslant\varepsilon}|x|^3\,dF_{nk}\leqslant$$

$$\leqslant\varepsilon^2+\frac{1}{\varepsilon}M|\zeta_{nk}|^3\leqslant 2\varepsilon^2, \tag{71}$$

если n — достаточно велико, так как

$$M|\zeta_{nk}|^3=\frac{1}{B_n^3}M|\xi_k-\mu_k|^3\to 0 \text{ при } n\to\infty.$$

Положим

$$z_{nk}=-\frac{D\xi_k}{B_n^2}\frac{t^2}{2}+R_{nk}(t). \tag{72}$$

Из оценок (70) и (71) следует, что

$$|z_{nk}|\to 0 \text{ при } n\to\infty \text{ и } k=1, \ldots, n,\ |t|\leqslant T. \tag{73}$$

Равенство (69) перепишем в виде

$$f_{nk}(t)=1+z_{nk}$$

и

$$\ln f_{nk}(t)=z_{nk}+(\ln f_{nk}(t)-z_{nk}), \tag{74}$$

где взято главное значение логарифма. Из (72) и (74) находим

$$\sum_{k=1}^{n}\ln f_{nk}(t)=\sum_{k=1}^{n}z_{nk}+\sum_{k=1}^{n}(\ln f_{nk}(t)-z_{nk})=$$

$$=-\frac{t^2}{2}\frac{1}{B_n^2}\sum_{k=1}^{n}D\xi_k+\sum_{k=1}^{n}R_{nk}(t)+\sum_{k=1}^{n}(\ln f_{nk}(t)-z_{nk})=$$

$$=-\frac{t^2}{2}+\sum_{k=1}^{n}R_{nk}(t)+\sum_{k=1}^{n}(\ln f_{nk}(t)-z_{nk}). \tag{75}$$

В силу предпоследнего неравенства (70) и условия (63) имеем при $|t|\leqslant T$

$$\left|\sum_{k=1}^{n}R_{nk}(t)\right|\leqslant\frac{T^3}{6}\frac{1}{B_n^3}\sum_{k=1}^{n}M|\xi_k-\mu_k|^3\to 0,\ n\to\infty. \tag{76}$$

Для любого комплексного β с $|\beta|\leqslant 1/2$ имеем неравенство

$$|\ln(1+\beta)-\beta|=\left|\int_0^{\beta}\frac{z\,dz}{1+z}\right|\leqslant 2\left|\int_0^{\beta}|z|\,d|z|\right|\leqslant|\beta|^2, \tag{77}$$

поэтому

$$\left|\sum_{k=1}^{n}(\ln f_{nk}(t)-z_{nk})\right|\leqslant\sum_{k=1}^{n}|z_{nk}|^2\leqslant\max_{k=1,\ldots,n}|z_{nk}|\times$$

$$\times\sum_{k=1}^{n}|z_{nk}|\leqslant\max_{k}|z_{nk}|\sum_{k=1}^{n}\left(\frac{t^2}{2}\frac{D\xi_k}{B_n^2}+|R_{nk}(t)|\right)\leqslant$$

$$\leqslant\max_{k}|z_{nk}|\left(\frac{T^2}{2}+\sum_{k=1}^{n}|R_{nk}(t)|\right)\to 0,\quad n\to\infty \tag{78}$$

в силу (73) и (76).

Сопоставляя (75), (76) и (78), заключаем, что $\sum_{k=1}^{n}\ln f_{nk}(t)\to$

$\to-\dfrac{t^2}{2}$ при $n\to\infty$ равномерно по t при $|t|\leqslant T$, что равносильно (68). ▲

4°. Применения центральных предельных теорем и примеры

Отметим прежде всего случай, когда нельзя применять центральные предельные теоремы. Пусть мы хотим оценить вероятность $P\{\eta_n<x\}$, причем речь идет о тех x, при которых эта вероятность близка к нулю или единице. Если мы заменим $P\{\eta_n<x\}$ на $\Phi_0(x)$, то ошибка может быть очень большой (порядка сотен процентов) ввиду того, что хотя разность $P\{\eta_n<x\}-\Phi_0(x)$ и будет равномерно малой при всех x, но неверно, что отношение $P\{\eta_n<x\}/\Phi_0(x)\to 1$ равномерно по x, т. е. «хвосты» распределения требуют очень осторожной оценки.

Вероятность и частота. Пользуясь теоремой 6, оценим, насколько сильно может отличаться частота от вероятности в серии из n испытаний Бернулли. Оценка основана на соотношении

$$P\left\{\left|\frac{\eta_n}{n}-p\right|>\varepsilon\right\}=$$

$$=P\left\{\left|\frac{\eta_n-np}{\sqrt{npq}}\right|>\varepsilon\sqrt{\frac{n}{pq}}\right\}\approx\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{-\varepsilon\sqrt{\frac{n}{pq}}}e^{-\frac{x^2}{2}}dx+$$

$$+\frac{1}{\sqrt{2\pi}}\int_{\varepsilon\sqrt{\frac{n}{pq}}}^{\infty}e^{-\frac{x^2}{2}}dx=2\Phi_0\left(-\varepsilon\sqrt{\frac{n}{pq}}\right),$$

η_n — число успехов при n испытаниях.

Поскольку при $p+q=1$, очевидно, $pq\leqslant 1/4$, эта вероятность не превосходит $2\Phi_0(-2\varepsilon\sqrt{n})$. Поэтому

$$P\left\{\frac{\eta_n}{n}-\varepsilon<p<\frac{\eta_n}{n}+\varepsilon\right\}\approx$$

$$\approx 1-2\Phi_0\left(-\varepsilon\sqrt{\frac{n}{pq}}\right)\geqslant 1-2\Phi_0(-2\varepsilon\sqrt{n}).$$

Таким образом, зная число успехов η_n в n испытаниях Бернулли, мы можем построить интервал $\left(\frac{\eta_n}{n}-\varepsilon,\ \frac{\eta_n}{n}+\varepsilon\right)$, который будет накрывать истинное (неизвестное) значение вероятности p с любой заданной вероятностью $1-\alpha$. Для этого следует лишь выбрать $\varepsilon=\varepsilon(\alpha)$ из соотношения $2\Phi_0(-2\varepsilon\sqrt{n})=\alpha$, пользуясь таблицами нормального распределения. Тогда

$$P\left\{\frac{\eta_n}{n}-\varepsilon(\alpha)<p<\frac{\eta_n}{n}+\varepsilon(\alpha)\right\}=1-\alpha.$$

Интервал $\left(\frac{\eta_n}{n}-\varepsilon(\alpha),\ \frac{\eta_n}{n}+\varepsilon(\alpha)\right)$ называется **доверительным интервалом** для p **с уровнем доверия** $1-\alpha$ (см. подробнее об этом § 15).

Среднее арифметическое. Пусть $\xi_1, \xi_2, \dots$ — независимые случайные величины, $M\xi_k=\mu$, $D\xi_k=\sigma^2$ для всех k. Закон больших чисел (см. § 10) утверждает, что при $n\to\infty$

$$P\left\{\left|\frac{\xi_1+\dots+\xi_n}{n}-\mu\right|>\varepsilon\right\}\to 0.$$

Если $\xi_1, \xi_2, \dots$ не только попарно независимы (что достаточно для применимости закона больших чисел), но и независимы в совокупности, то можно применить теорему 6. Это дает

$$P\left\{\left|\frac{\xi_1+\dots+\xi_n}{n}-\mu\right|>\varepsilon\right\}=$$

$$=P\left\{\left|\frac{\xi_1+\dots+\xi_n-n\mu}{\sigma\sqrt{n}}\right|>\frac{\varepsilon\sqrt{n}}{\sigma}\right\}=$$

$$=P\left\{\eta_n<-\frac{\varepsilon\sqrt{n}}{\sigma}\right\}+P\left\{\eta_n>\frac{\varepsilon\sqrt{n}}{\sigma}\right\}=2\Phi_0\left(-\frac{\varepsilon\sqrt{n}}{\sigma}\right).$$

Из таблиц нормального распределения следует, что при $\frac{\varepsilon\sqrt{n}}{\sigma}=$ $=3$, т. е. при $\varepsilon=\frac{3\sigma}{\sqrt{n}}$ вероятность

$$P\left\{|\eta_n|<\frac{\varepsilon\sqrt{n}}{\sigma}\right\}=1-2\Phi_0\left(-\frac{\varepsilon\sqrt{n}}{\sigma}\right)$$

равна 0,997. Это так называемое **правило 3σ.**

Примеры.

1. Пусть ξ_k — независимые дискретные случайные величины, равные $\pm k$ с вероятностью $1/2$, $k=1, 2, \dots$. Имеем

$$M\xi_k=k\cdot\frac{1}{2}-k\cdot\frac{1}{2}=0,$$

$$D\xi_k=M\xi_k^2=k^2\cdot\frac{1}{2}+k^2\cdot\frac{1}{2}=k^2,$$

$$M|\xi_k-M\xi_k|^3=M|\xi_k|^3=k^3.$$

Легко по индукции доказываются формулы:

$$\sum_{k=1}^{n}k^2=\frac{n(n+1)(2n+1)}{6},\quad \sum_{k=1}^{n}k^3=\left[\frac{n(n+1)}{2}\right]^2.$$

Отсюда $B_n=\sqrt{\sum_{k=1}^{n}D\xi_k}=\sqrt{\frac{n(n+1)(2n+1)}{6}}$, и условие (63) Ляпунова выполнено

$$\frac{1}{B_n^3}\sum_{k=1}^{n}M|\xi_k-M\xi_k|^3=$$

$$=\left(\frac{n(n+1)(2n+1)}{6}\right)^{-3/2}\left(\frac{n(n+1)}{2}\right)^2\to 0,\quad n\to\infty.$$

Таким образом согласно теореме 8 случайные величины $\eta_n=(\xi_1+\dots+\xi_n)/B_n$ асимптотически нормальны с параметрами $(0, 1)$ при $n\to\infty$.

2. Доказать, что

$$\lim_{n\to\infty}e^{-n}\sum_{k=0}^{n}\frac{n^k}{k!}=\frac{1}{2}. \tag{79}$$

Доказательство. Рассмотрим последовательность независимых случайных величин ξ_k, $k=1, 2, \dots$, распределенных по закону Пуассона с параметром $\lambda=1$. Имеем $M\xi_k=1$, $D\xi_k=1$ и согласно теореме 6 при $n\to\infty$

$$P\{\eta_k<x\}=P\left\{\frac{\xi_1+\dots+\xi_n-n}{\sqrt{n}}<x\right\}\to\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}e^{-\frac{z^2}{2}}\,dz$$

равномерно по $x\in R_1$. Полагая здесь $x=0$, получим

$$P\{\xi_1+\dots+\xi_n<n\}=P\left\{\frac{\xi_1+\dots+\xi_n-n}{\sqrt{n}}<0\right\}\to\frac{1}{2}. \tag{80}$$

В силу применения 2) п. 2° сумма $\xi_1+\dots+\xi_n$ распределена по закону Пуассона с параметром $\lambda=n$, поэтому

$$P\{\xi_1+\dots+\xi_n<n\}=\sum_{k=0}^{n}\frac{n^k}{k!}e^{-n}-\frac{n^ne^{-n}}{n}. \tag{81}$$

Сопоставляя (80) и (81), видим, что справедливо (79).

3. Театр, вмещающий 1000 человек, имеет два разных входа. Около каждого входа имеется гардероб. Сколько мест должно быть в каждом гардеробе для того, чтобы в 99% случаев все зрители могли раздеться в гардеробе того входа,

через который они вошли? Предполагается, что зрители приходят парами и каждая пара независимо от других выбирает с вероятностью 1/2 любой из входов.

Очевидно, это схема Бернулли с числом испытаний $n=500$ — число пар, p — вероятность выбора определенного входа, $p=1/2$, $q=1-p=1/2$.

Обозначим через x число мест в каждом гардеробе, а через m — число успехов, т. е. случаев, когда пришедшей паре есть место в гардеробе соответствующего входа. Тогда x определяется из условия

$$P\left\{m \leqslant \frac{x}{2},\ n-m \leqslant \frac{x}{2}\right\} = 0{,}99.$$

Для нахождения x воспользуемся интегральной теоремой Муавра—Лапласа:

$$0{,}99 = P\left\{n-\frac{x}{2} \leqslant m \leqslant \frac{x}{2}\right\} \approx -2\Phi_0\left(\frac{\frac{x}{2}-\frac{500}{2}}{5\sqrt{5}}\right)+1.$$

Отсюда по таблицам определяем: $x=537$.

§ 12. КОНЕЧНЫЕ ОДНОРОДНЫЕ ЦЕПИ МАРКОВА

1°. Определение цепи Маркова

Понятие конечной цепи Маркова является простейшим и вместе с тем важным обобщением схемы независимых испытаний на случай, когда испытания зависимы. Напомним, что последовательностью n независимых испытаний мы называли вероятностное пространство $(\Omega_n, \mathscr{F}_n, P_n)$; элементами Ω_n являются всевозможные последовательности $(\omega_{i_1}, \ldots, \omega_{i_n})$, где $\omega_k \in \Omega$, $\Omega = (\omega_1, \omega_2, \ldots)$ — множество исходов в каждом испытании, $\mathscr{F}_n$ — σ-алгебра всех подмножеств Ω_n, при этом независимость испытаний фиксировалась определением вероятности P_n, $P_n(\omega_{i_1}, \ldots, \omega_{i_n}) = p_{i_1} \ldots p_{i_n}$, где $p_k = P\{\omega_k\}$ — вероятность исхода ω_k в каждом испытании.

Предположим теперь, что в последовательности n испытаний вероятность исхода в любом s-м испытании **зависит** от исхода $(s-1)$-го испытания и не зависит от исходов испытаний с номерами $s-2, s-3, \ldots, 1$. Обозначим эту условную вероятность

$$p_{ik} = P(\{\omega_k\} \mid \{\omega_i\}),\ 0 \leqslant p_{ik} \leqslant 1. \qquad (1)$$

В этом обозначении отражен тот факт, что вероятность события ω_k в s-м испытании при условии, что в $(s-1)$-м испытании произошло событие ω_i, не зависит от номера s. Это свойство называется **однородностью последовательности испытаний.** Пусть задано распределение вероятностей для первого испытания:

$$P(\{\omega_j\}) = a_j,\ j=1, 2, \ldots, a_j \geqslant 0,\ \sum_{j=1}^{\infty} a_j = 1. \qquad (2)$$

Тогда вероятность исхода $\{\omega_i\omega_j\}$ первого и второго испытания равна $P_2(\{\omega_i\omega_j\}) = a_i p_{ij}$ и аналогично для k испытаний, $k=2, 3, \ldots$,

$$P_k(\{\omega_{i_1} \ldots \omega_{i_k}\}) = a_{i_1} p_{i_1 i_2} \ldots p_{i_{k-1} i_k},\ i_1, i_2, \ldots, i_k = 1, 2, \ldots. \qquad (3)$$

Убедимся, что равенства (3) при $k=n$ задают вероятность на $\mathscr{F}_n$, т. е. на всех подмножествах Ω_n. Для этого достаточно проверить, что

$$\sum_{(\omega_{i_1} \ldots \omega_{i_n}) \in \Omega_n} P_n(\{\omega_{i_1} \ldots \omega_{i_n}\}) = \sum_{i_1=1, \ldots, i_n=1}^{\infty} a_{i_1} p_{i_1 i_2} \ldots p_{i_{n-1} i_n} = 1, \qquad (4)$$

так как в остальном проверка дословно повторяет случай независимых испытаний (см. § 5). Поскольку для всякого исхода ω_i в $(s-1)$-м испытании в следующем, s-м испытании непременно наблюдается один из исходов $\omega_1, \omega_2, \ldots$, то для каждого i $\sum_{j=1}^{\infty} p_{ij} = 1$. Это влечет следующую цепочку равенств:

$$\sum_{i_1=1, \ldots, i_n=1}^{\infty} a_{i_1} p_{i_1 i_2} \ldots p_{i_{n-1} i_n} =$$

$$= \sum_{i_1=1, \ldots, i_{n-1}=1}^{\infty} a_{i_1} p_{i_1 i_2} \ldots p_{i_{n-2} i_{n-1}} \sum_{i_n=1}^{\infty} p_{i_{n-1} i_n} =$$

$$= \sum_{i_1=1, \ldots, i_{n-1}=1}^{\infty} a_{i_1} p_{i_1 i_2} \ldots p_{i_{n-2} i_{n-1}} = \ldots = 1,$$

тем самым равенство (4) доказано.

Далее для простоты предполагается, что пространство Ω конечно, $\Omega = (\omega_1, \ldots, \omega_N)$.

Определение. Конечной однородной цепью Маркова, состоящей из n испытаний, называется вероятностное пространство $(\Omega_n, \mathscr{F}_n, P_n)$, в котором Ω_n — множество всех последовательностей $(\omega_{i_1}, \ldots, \omega_{i_n})$, $1 \leqslant i_1, i_2, \ldots, i_n \leqslant N$, $\mathscr{F}_n$ — алгебра всех подмножеств Ω_n и вероятность P_n определена для каждого элементарного события Ω_n равенством (3) с $k=n$, где $a_i \geqslant 0$, $\sum_{i=1}^{N} a_i = 1$ — начальное распределение, $p_{ij} \geqslant 0$, $\sum_{j=1}^{N} p_{ij} = 1$ — переходные вероятности, $i, j = 1, \ldots, N$.

Название «*переходные вероятности*» для условных вероятностей (1) связано со следующей интерпретацией цепей Маркова. Будем считать, что события ω_j, $j=1, 2, \ldots, N$, отмечают состояния некоторой системы, которая, таким образом, может находиться в одном из N состояний. Пусть в дискретные моменты времени $1, 2, \ldots$ система может менять свое состояние, переходя из состояния $\omega_{i_{s-1}}$, в котором она находилась в $(s-1)$-й момент времени, в состояние ω_{i_s} в s-й момент времени с вероятностью перехода $p_{i_{s-1} i_s}$. Если распределение вероятностей состояний в начальный момент времени задается равенствами $P(\{\omega_i\}) = a_i$, $i=1, 2, \ldots, N$, то мы получаем интерпретацию цепи Маркова в терминах состояний некоторой системы. Далее мы пользуемся как той, так и другой точкой зрения на цепи Маркова.

Совокупность вероятностей перехода p_{ij}, $i, j = 1, \ldots, N$ образует матрицу

$$\pi_1 = \begin{pmatrix} p_{11} & p_{12} & \ldots & p_{1N} \\ p_{21} & p_{22} & \ldots & p_{2N} \\ \ldots & \ldots & \ldots & \ldots \\ p_{N1} & p_{N2} & \ldots & p_{NN} \end{pmatrix}, \tag{5}$$

она называется **матрицей перехода.** По определению все элементы матрицы π_1 неотрицательны и сумма элементов каждой строки равна единице: $\sum_{j=1}^{N} p_{ij} = 1$, $i = 1, \ldots, N$. Матрицы, обладающие этими свойствами, называются **стохастическими.**

Примеры.

1. *Случайное блуждание с поглощением.* Пусть частица может передвигаться по прямой под действием случайных толчков. В точках $x=1$ и $x=N$ стоят поглощающие экраны: и пусть при каждом случайном толчке частица передвигается на единицу длины вправо с вероятностью p, влево с вероятностью q, $p+q=1$, однако, попав в точки $x=1$ и $x=N$, частица остается в этих точках. Таким образом, имеем состояния частицы: $\omega_i = \{$частица находится в точке $x=i\}$, $i=1, \ldots, N$. Переходные вероятности в этом случае равны: $p_{00}=1$, $p_{NN}=1$, $p_{ii+1}=p$, $p_{ii-1}=q$, $p_{ij}=0$ для остальных i и j. Матрица перехода имеет вид

$$\pi_1 = \begin{pmatrix} 1 & 0 & 0 & \ldots & & 0 \\ q & 0 & p & \ldots & & 0 \\ 0 & q & 0 & p & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & 0 & \ldots & q\ 0 & p \\ 0 & 0 & 0 & \ldots & 0\ 0 & 1 \end{pmatrix}.$$

Полезным представлением цепи Маркова является ее граф (рис. 20).

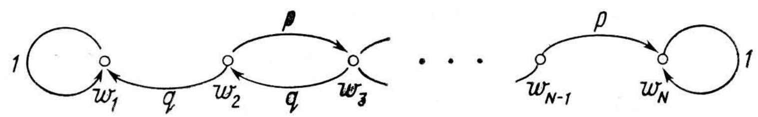


Рис. 20

2. *Случайное блуждание с отражением.* Рассмотрим ту же схему, что и в предыдущем примере с той лишь разницей, что теперь в точках $x=1$ и $x=N$ стоят отражающие экраны. В этом случае матрица переходов имеет вид

$$\pi_1 = \begin{pmatrix} 0 & 1 & 0 & \ldots & & 0 \\ q & 0 & p & \ldots & & 0 \\ 0 & q & 0 & p & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & 0 & \ldots & q\ 0 & p \\ 0 & 0 & 0 & \ldots & 0\ 1 & 0 \end{pmatrix},$$

а цепь Маркова представляется графом, изображенным на рис. 21.

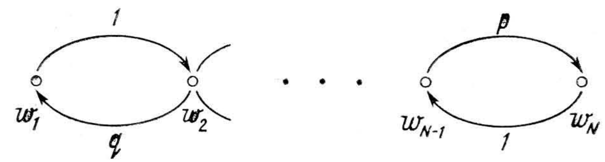


Рис. 21

3. *Схема независимых испытаний.* В этом случае $p_{ij} = P(\{\omega_j\} | \{\omega_i\}) = a_j$, т. е. вероятность состояния ω_j не зависит от исхода предыдущего испытания и совпадает с вероятностью события ω_j в начальном испытании. Матрица π_1

имеет вид

$$\pi_1 = \begin{pmatrix} a_1 & a_2 & \dots & a_N \\ a_1 & a_2 & \dots & a_N \\ \cdot & \cdot & \cdot & \cdot \\ a_1 & a_2 & \dots & a_N \end{pmatrix}.$$

Соответствующий граф представлен на рис. 22.

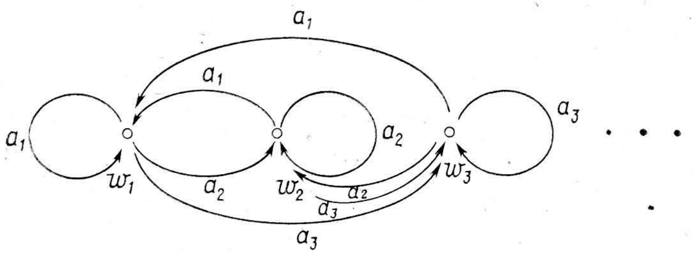


Рис. 22

2°. Вероятность перехода за n шагов

Рассмотрим вероятность перехода из состояния ω_i, которое реализовано, например, в s-м испытании, в состояние ω_j за n шагов, т. е. в состояние ω_j в $(s+n)$-м испытании. Ясно, что в силу однородности цепи Маркова эта вероятность зависит только от n (и не зависит от s). Обозначим ее $p^n{}_{ij}$. Тогда p^m_{ik}—вероятность перехода за m шагов из состояния ω_i в состояние ω_k и p^{n-m}_{kj}— вероятность перехода за $n-m$ шагов из состояния ω_k в ω_j.

Пользуясь формулой полной вероятности и учитывая, что промежуточные состояния $\omega_1, \omega_2, \dots, \omega_N$ в $(s+m)$-м испытании образуют полную группу попарно несовместных событий, найдем

$$p^n_{ij} = \sum_{k=1}^{N} p^m_{ik}\, p^{n-m}_{kj}, \tag{6}$$

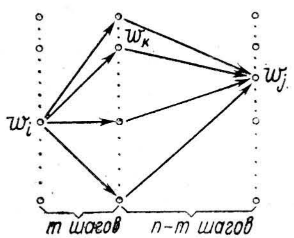


Рис. 23

здесь m — любое целое значение от 1 до $n-1$.

Обозначим через π_n матрицу, составленную из вероятностей $p^n{}_{ij}$, $i, j = 1, \dots, N$; таким образом, π_n — матрица перехода через n испытаний. Пользуясь формулой для перемножения квадратных матриц, равенство (6) можно записать в матричном виде

$$\pi_n = \pi_m \pi_{n-m}, \quad m = 1, \dots, n-1. \tag{7}$$

Применяя последовательно формулу (7), получим

$$\pi_n = \pi_1 \pi_{n-1} = \pi_1 \pi_1 \pi_{n-2} = \dots = \pi_1{}^n. \tag{8}$$

3°. Эргодичность

Матрица π_1, соответствующая схеме независимых испытаний (см. пример 3 п. 1°), обладает свойством $\pi_1 = \pi_n = \pi_1{}^n$ для любого натурального n. Интуитивно можно ожидать, что и в случае любой цепи Маркова при переходах в n шагов влияние начального распределения с ростом n должно ослабевать в том смысле, что $p^n_{ij} \to p_j$ при $n \to \infty$ независимо от i. Такая предельная матрица переходов имеет вид матрицы π_1 в схеме независимых испытаний и, следовательно, удовлетворяет условию $\pi_1{}^k = \pi_1$, $k = 1, 2, \dots$. Это свойство цепей Маркова, называемое **эргодичностью**, действительно имеет место, хотя и не во всех случаях. Далее мы укажем достаточное условие эргодичности (теорема Маркова), а сейчас рассмотрим некоторые свойства эргодических цепей Маркова. Обозначим через p^k_j — вероятность того, что в k-м испытании осуществится событие ω_j, p^k_j называется **абсолютной вероятностью.**

Пусть числа $a_1, a_2, \dots, a_N$ задают начальное распределение вероятностей состояний $\omega_1, \omega_2, \dots, \omega_N : a_i \geqslant 0$, $\sum\limits_{i=1}^{N} a_i = 1$. Согласно формуле полной вероятности распределение вероятностей состояний $\omega_1, \dots, \omega_N$ во втором испытании определяется равенством $p^2_j = \sum\limits_{k=1}^{N} a_k p_{kj}$, в третьем испытании $p^3_j = \sum\limits_{l=1}^{N} p^2_l\, p_{lj} = \sum\limits_{k=1}^{N} a_k p^2_{kj}$ и т. д., в n-м испытании задается равенством

$$p^n_j = \sum_{l=1}^{N} p^{n-1}_l p_{lj} = \sum_{k=1}^{N} a_k p^{n-1}_{kj}, \quad n = 2, 3, \dots, \quad j = 1, \dots, N. \tag{9}$$

Если при $n \to \infty$ существуют пределы

$$\lim_{n\to\infty} p^n_{kj} = p_j, \quad k, j = 1, \dots, N, \tag{10}$$

то существуют в силу (9) и пределы для абсолютных вероятностей:

$$\lim_{n\to\infty} p^n_j = \sum_{k=1}^{N} a_k \lim_{n\to\infty} p^{n-1}_{kj} = \sum_{k=1}^{N} a_k p_j = p_j, \quad j = 1, \dots, N. \tag{11}$$

Таким образом, согласно (11), если имеет место эргодичность, то существует предельное, называемое **финальным**, распределение вероятностей $p_1, \dots, p_j$ состояний $\omega_1, \dots, \omega_N$, не зависящее от начального распределения $a_1, \dots, a_N$. Переходя к пределу при $n \to \infty$ в первом равенстве формулы (9), найдем

$$p_j = \lim_{n\to\infty} p_j^n = \sum_{l=1}^{N} \lim_{n\to\infty} p_l^{n-1} p_{lj} = \sum_{l=1}^{N} p_l p_{lj}. \tag{12}$$

Таким образом, финальное распределение вероятностей $p_1, \dots, p_N$ удовлетворяет системе уравнений

$$p_j = \sum_{l=1}^{N} p_l p_{lj}, \quad j = 1, \dots, N, \tag{13}$$

и очевидным дополнительным условиям

$$p_j \geqslant 0, \quad \sum_{j=1}^{N} p_j = 1 \tag{14}$$

(которые также следуют из (9)).

Система (13) — однородная система уравнений, которая в силу стохастичности матрицы перехода π_1 имеет определитель, равный нулю. Сопоставляя (13) и первое равенство в формуле (9), приходим к выводу, что если в качестве начального распределения $a_1, \dots, a_N$ выбрать финальное распределение $p_1, \dots, p_N$, то распределение состояний $\omega_1, \dots, \omega_N$ не будет меняться от испытания к испытанию. Тем самым показано, что *финальное распределение стационарно.*

Отметим также следующий глубокий результат, характеризующий свойство эргодичности. Пусть в каждом состоянии до момента перехода в следующее состояние система находится τ секунд. Пусть, далее, A — некоторое множество состояний, T_A — время, которое система находится в состояниях из множества A, а $T = n\tau$ — общее время функционирования системы. Тогда в случае эргодичности

$$\lim_{n\tau = T\to\infty} \left(\frac{T_A}{T}\right) = \sum_{\omega_i \in A} p_i.$$

Однако доказательство этого результата потребовало бы более развитого аппарата исследования цепей Маркова.

Докажем теперь указанную выше теорему об эргодичности.

Теорема Маркова. Пусть существуют целое число $\nu \geqslant 1$ и множество J из $N_1 \geqslant 1$ значений j, такие, что

$$\min_{\substack{1\leqslant i\leqslant N \\ j\in J}} p_{ij}^{\nu} = \delta > 0. \tag{15}$$

Тогда существуют числа $p_1, \dots, p_N$, такие, что независимо от i

$$\lim_{n\to\infty} p_{ij}^n = p_j, \quad i, j = 1, \dots, N. \tag{16}$$

Доказательство. Заметим прежде всего, что матрица $\pi_n = \{p_{ij}^n\}$ — стохастическая при любом $n = 1, 2, \dots$. Действительно, в силу (6):

$$\sum_{j=1}^{N} p_{ij}^n = \sum_{j=1}^{N} \left(\sum_{k=1}^{N} p_{ik}^{n-1} p_{kj}\right) = \sum_{k=1}^{N} p_{ik}^{n-1} \sum_{j=1}^{N} p_{kj} = $$
$$= \sum_{k=1}^{N} p_{ik}^{n-1} = \dots = \sum_{k=1}^{N} p_{ik} = 1. \tag{17}$$

Обозначим

$$m_j^r = \min_{1\leqslant i\leqslant N} p_{ij}^r, \quad M_j^r = \max_{1\leqslant i\leqslant N} p_{ij}^r, \quad m_j^r \leqslant M_j^r, \quad j = 1, \dots, N.$$

Имеем

$$m_j^{r+1} = \min_i \sum_{k=1}^{N} p_{ik} p_{kj}^r \geqslant \min_i \sum_{k=1}^{N} p_{ik} m_j^r = $$
$$= m_j^r \min_i \sum_{k=1}^{N} p_{ik} = m_j^r,$$

и

$$M_j^{r+1} = \max_i \sum_{k=1}^{N} p_{ik} p_{kj}^r \leqslant \max_i \sum_{k=1}^{N} p_{ik} M_j^r = $$
$$= M_j^r \max_i \sum_{k=1}^{N} p_{ik} = M_j^r,$$

т. е.

$$m_j^1 \leqslant m_j^2 \leqslant \dots \leqslant m_j^r \leqslant M_j^r \leqslant M_j^{r-1} \leqslant \dots \leqslant M_j^2 \leqslant M_j^1.$$

Таким образом, последовательности $\{m_j^n\}$ и $\{M_j^n\}$ монотонны и ограничены, а значит, сходятся при $n \to \infty$, $\lim_{n\to\infty} m_j^n = m_j$, $\lim_{n\to\infty} M_j^n = M_j$. Мы докажем, что эти пределы совпадают: $m_j = M_j$. Если этот общий предел обозначить p_j, то, поскольку $m_j^n \leqslant p_{ij}^n \leqslant M_j^n$, $n = 1, 2, \dots$, отсюда будет следовать, что $\lim_{n\to\infty} p_{ij}^n = p_j$, что и утверждается в теореме.

Фиксируем любые две строки α и β матрицы π_ν, т. е. рассмотрим строки $p_{\alpha j}^\nu$ и $p_{\beta j}^\nu$, $j = 1, \dots, N$. Обозначим через Σ^+ суммирование по тем k, для которых $p_{\alpha k}^\nu \geqslant p_{\beta k}^\nu$, а через Σ^- суммирование по тем k, для которых $p_{\alpha k}^\nu < p_{\beta k}^\nu$; имеем

$$\sum_k{}^{+} \left(p_{\alpha k}^\nu - p_{\beta k}^\nu\right) + \sum_k{}^{-} \left(p_{\alpha k}^\nu - p_{\beta k}^\nu\right) = \sum_{k=1}^{N} p_{\alpha k}^\nu - \sum_{k=1}^{N} p_{\beta k}^\nu = 1 - 1 = 0. \tag{18}$$

Покажем, что

$$\sum_k^+ \left(p_{\alpha k}^{\nu} - p_{\beta k}^{\nu}\right) \leqslant 1 - N_1\delta. \qquad (19)$$

Действительно, пусть в $\sum\limits_k^+$ входит s, $0 \leqslant s \leqslant N_1$, значений $k \in J$, тогда в $\sum\limits_k^-$ входят остальные $N_1 - s$ значений $k \in J$. Тогда

$$\sum_k^+ \left(p_{\alpha k}^{\nu} - p_{\beta k}^{\nu}\right) = \sum_{k=1}^N p_{\alpha k}^{\nu} - \sum_k^- p_{\alpha k}^{\nu} - \sum_k^+ p_{\beta k}^{\nu} =$$

$$= 1 - \sum_k^- p_{\alpha k}^{\nu} - \sum_k^+ p_{\beta k}^{\nu} \quad \text{и} \qquad (20)$$

$$\sum_k^- p_{\alpha k}^{\nu} \geqslant \sum_{k \in J}^- p_{\alpha k}^{\nu} \geqslant \delta (N_1 - s), \quad \sum_k^+ p_{\beta k}^{\nu} \geqslant \sum_{k \in J}^+ p_{\beta k}^{\nu} \geqslant \delta s. \quad (21)$$

Подставляя (21) в (20), найдем

$$\sum_k^+ \left(p_{\alpha k}^{\nu} - p_{\beta k}^{\nu}\right) \leqslant 1 - \delta (N_1 - s) - \delta s = 1 - \delta N_1,$$

и (19) доказано. Неравенство (19) верно для любых строк α и β, поэтому и

$$\max_{1 \leqslant \alpha, \beta \leqslant N} \sum_k^+ \left(p_{\alpha k}^{\nu} - p_{\beta k}^{\nu}\right) \leqslant 1 - N_1\delta. \qquad (22)$$

Далее, при любом $n > 0$ имеем с учетом (18) и (22):

$$M_j^{\nu+n} - m_j^{\nu+n} = \max_i \sum_{k=1}^N p_{ik}^{\nu} p_{kj}^n - \min_i \sum_{k=1}^N p_{ik}^{\nu} p_{kj}^n =$$

$$= \max_{\alpha, \beta} \sum_{k=1}^N \left(p_{\alpha k}^{\nu} - p_{\beta k}^{\nu}\right) p_{kj}^n \leqslant \max_{\alpha, \beta} \left\{ \sum_k^+ \left(p_{\alpha k}^{\nu} - p_{\beta k}^{\nu}\right) M_j^n + \right.$$

$$\left. + \sum_k^- \left(p_{\alpha k}^{\nu} - p_{\beta k}^{\nu}\right) m_j^n \right\} = \max_{\alpha, \beta} \sum_k^+ \left(p_{\alpha k}^{\nu} - p_{\beta k}^{\nu}\right) \left(M_j^n - m_j^n\right) \leqslant$$

$$\leqslant (1 - N_1\delta)(M_j^n - m_j^n). \qquad (23)$$

Итак,

$$M_j^{\nu+n} - m_j^{\nu+n} \leqslant (1 - N_1\delta)(M_j^n - m_j^n), \ n \text{ — любое.}$$

Переходя к пределу в этом неравенстве при $n \to \infty$, найдем

$$0 \leqslant M_j - m_j \leqslant (1 - N_1\delta)(M_j - m_j). \qquad (24)$$

Так как $\delta > 0$, то $1 - N_1\delta < 1$ и из (24) следует равенство $M_j - m_j = 0$. ▲

П р и м е р ы.

1. Пусть ξ_k, $k = 1, 2, \ldots$, независимые целочисленные одинаково распределенные случайные величины, $P\{\xi_k = i\} = p_i$, $i = 0, 1, 2, \ldots, p_i \geqslant 0$, $\sum\limits_{i=0}^{\infty} p_i = 1$. Пусть $\eta_n = \xi_1 + \ldots + \xi_n$. Докажем, что случайные величины η_n, $n = 1, 2, \ldots$, удовлетворяют условиям (1) и (3) однородной цепи Маркова. В самом деле, $\eta_{n+1} = \eta_n + \xi_{n+1}$, так что при заданном η_n распределение η_{n+1} известно. Подсчитаем вероятности перехода:

$$p_{ij} = P\{\eta_{n+1} = j \mid \eta_n = i\} = P \mid \{\eta_n + \xi_{n+1} = j \mid \eta_n = i\} =$$

$$= P\{\xi_{n+1} = j - i\} = \begin{cases} p_{j-i}, & j \geqslant i, \\ 0, & j < i; \end{cases}$$

таким образом цепь Маркова однородна.

Матрица перехода имеет вид

$$\pi_1 = \begin{pmatrix} p_0 \, p_1 \, p_2 \ \ldots \\ 0 \ \ p_0 \, p_1 \ \ldots \\ \ldots \ldots \end{pmatrix}.$$

2. Рассмотрим цепь Маркова с двумя состояниями ω_1 и ω_2, вероятностями перехода $p_{11} = p_{22} = p$, $p_{12} = p_{21} = q$ $(0 \leqslant p \leqslant 1$, $p + q = 1)$ и начальными вероятностями: $p(\omega_1) = a_1$, $p(\omega_2) = a_2$, $a_1 + a_2 = 1$. Найдем матрицу π_n перехода через n испытаний, абсолютные вероятности p_i^n и предельные вероятности p_i, $i = 1, 2$.

Поскольку в силу (8) $\pi_n = \pi_1^n$, то нам надо возвести матрицу перехода π_1 в n-ю степень. Для этого рассмотрим $\pi_1 = \begin{pmatrix} p q \\ q p \end{pmatrix}$ как матрицу линейного оператора A в эвклидовом пространстве R_2 с базисом $f_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ и $f_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$. Найдем собственные векторы оператора A: если $e = \begin{pmatrix} \xi \\ \eta \end{pmatrix}$, то $\pi_1 e = \lambda e$, и мы имеем систему

$$\begin{cases} p\xi + q\eta = \lambda\xi, \\ q\xi + p\eta = \lambda\eta. \end{cases}$$

Характеристический определитель

$$\begin{vmatrix} p - \lambda & q \\ q & p - \lambda \end{vmatrix} = 0,$$

отсюда $\lambda_1 = 1$, $\lambda_2 = p - q$ (обозначим для удобства $p - q = c$) и собственные векторы имеют вид: $e_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$, $e_2 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$.

В базисе $\{e_1, e_2\}$ матрица $\hat{\pi}_1$ линейного оператора A диагональна $\hat{\pi}_1 = \begin{pmatrix} 1 & 0 \\ 0 & c \end{pmatrix}$, и поэтому $\hat{\pi}_n = \hat{\pi}_1^n = \begin{pmatrix} 1 & 0 \\ 0 & c^n \end{pmatrix}$. Остается вернуться к старому базису. Если B — матрица перехода от $\{e_1, e_2\}$ к $\{f_1, f_2\}$, то, как известно, $\pi_n = B^{-1}\hat{\pi}_n B$. Найдем B; если $B = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}$, то $Be = f$, т. е. $\begin{cases} \alpha + \beta = 1 \\ \gamma + \delta = 0 \end{cases}$ и $\begin{cases} \alpha - \beta = 0 \\ \gamma - \delta = 1 \end{cases}$, отсюда $B = \frac{1}{2}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$ и

$$\pi_n = B^{-1}\hat{\pi}_n B = \frac{1}{2}\begin{pmatrix} 1 + c^n & 1 - c^n \\ 1 - c^n & 1 + c^n \end{pmatrix}.$$

Таким образом,

$$p_{11}^n = p_{22}^n = \frac{1}{2}(1 + (p-q)^n), \quad p_{12}^n = p_{21}^n = \frac{1}{2}(1 - (p-q)^n),$$

$$n = 1, 2, \ldots . \tag{25}$$

Далее,

$$\begin{aligned} p_1^{n+1} &= a_1 p_{11}^n + a_2 p_{21}^n = \frac{1}{2} + \frac{1}{2}(a_1 - a_2)(p-q)^n, \\ p_2^{n+1} &= a_1 p_{12}^n + a_2 p_{22}^n = \frac{1}{2} - \frac{1}{2}(a_1 - a_2)(p-q)^n. \end{aligned} \tag{26}$$

Наконец,

$$\lim_{n\to\infty} p_{i1}^n = \lim_{n\to\infty} p_{i2}^n = \frac{1}{2}, \quad \lim_{n\to\infty} p_i^n = \frac{1}{2}, \quad i = 1, 2. \tag{27}$$

§ 13. СЛУЧАЙНЫЕ ПРОЦЕССЫ

Случайные процессы (случайные функции) — это случайные величины, зависящие от параметров, принимающих дискретное или непрерывное множество значений. В любой автоматической системе, в любой системе, связанной с управлением, приходится учитывать воздействие случайных помех, являющихся функциями времени, температуры, давления и т. д. Теория случайных процессов находит широкое применение во многих разделах физики и техники.

Определение. Пусть $(\Omega, \mathcal{F}, P)$ — вероятностное пространство, T — некоторое числовое множество. Действительная функция $\xi(t) = f(t, \omega)$, определенная при $t \in T$, $\omega \in \Omega$, называется **случайным процессом (случайной функцией)**, если при каждом $t \in T$ $f(t, \omega)$ как функция $\omega \in \Omega$ является случайной величиной.

Если $\omega = \omega_0$ — фиксировано, то $f(t, \omega_0)$, рассматриваемая как функция t, $t \in T$, называется **реализацией** случайного процесса $\xi(t)$, или **выборочной функцией.** Если же фиксировано $t = t_0$, то случайная величина $\xi(t_0)$ называется **сечением** случайного процесса в точке $t = t_0$.

В каждом сечении распределение вероятностей случайного процесса задается одномерной функцией распределения $F(t, x) = P\{\xi(t) < x\}$. Однако $F(t, x)$ не дает исчерпывающую вероятностную характеристику случайного процесса $\xi(t)$, поскольку она не учитывает зависимости случайных величин $\xi(t)$ при разных t (т. е. зависимости различных сечений случайного процесса). Более полно вероятностные свойства $\xi(t)$ описывает k-мерная функция распределения — функция распределения случайного вектора $(\xi(t_1), \ldots, \xi(t_k))$ для k сечений случайного процесса:

$$F(t_1, x_1; \ldots, t_k, x_k) = P\{\xi(t_1) < x_1, \ldots, \xi(t_k) < x_k\}.$$

Однако практическое применение находят лишь функции распределения первого и второго порядков.

Рассмотренные в предыдущих параграфах последовательности независимых случайных величин $\xi_1, \xi_2, \ldots, \xi_n$, или случайных величин, связанных в марковскую цепь, представляют собой примеры случайных процессов. Вообще любая последовательность случайных величин может быть интерпретирована как случайный процесс, при этом $T = \{1, 2, \ldots\}$. Процессы, в которых T является подмножеством множества целых чисел, называются **случайными процессами с дискретным временем**, или **случайными последовательностями.**

Переходим к изучению основных типов случайных процессов.

1.° Процесс Пуассона и процесс Винера

Эти два процесса являются наиболее важными представителями так называемых *процессов с независимыми приращениями.*

Начнем с определения процесса Пуассона, или пуассоновского потока событий. Рассмотрим некоторое событие A, которое может происходить в случайные моменты времени, и пусть $\xi(t)$ — число наступлений события A в промежутке времени длины t. Пусть выполнены условия:

1) случайные величины $\xi(t_j)$, $j = 1, 2, \ldots$, для непересекающихся промежутков времени независимы в совокупности;

2) для любого промежутка времени вероятность наступления события A в этом промежутке зависит лишь от длины этого промежутка (и не зависит от того, где на оси времени он расположен) — свойство однородности процесса по времени;

3) при $t \to 0$

$$P\{\xi(t) = 1\} = \lambda t + o(t), \ \lambda > 0; \ P\{\xi(t) > 1\} = o(t).$$

Сформулированным условиям подчиняются многие реальные явления: распад радиоактивного вещества, отказы радиоэлектронной аппаратуры, вызовы на телефонной станции, запросы на обслуживание станков и т. д.

Теорема 1. Пусть выполнены условия 1)—3). Распределение случайного вектора $(\xi(t_1), \ldots, \xi(t_j))$ для непересекающихся промежутков времени является пуассоновским:

$$P\{\xi(t_1) = k_1, \ldots, \xi(t_j) = k_j\} = \frac{(\lambda t_1)^{k_1}}{k_1!} \cdots \frac{(\lambda t_j)^{k_j}}{k_j!} e^{-\lambda(t_1 + \ldots + t_j)}. \quad (1)$$

Доказательство. Ввиду 1) достаточно доказать, что

$$P\{\xi(t) = k\} = \frac{(\lambda t)^k}{k!} e^{-\lambda t}. \quad (2)$$

Чтобы использовать условие 3), разобьем промежуток времени t на n непересекающихся интервалов Δ_i длины $\Delta = t/n$, а затем устремим $n \to \infty$. Событие $\{\xi(t) = k\}$ представимо в виде

$$\{\xi(t) = k\} = A_1 + A_2,$$

где, в свою очередь,

$$A_1 = \bigcup_{\{i_1, \ldots, i_k\}} \{\xi(\Delta_{i_1}) = 1,\ \xi(\Delta_{i_2}) = 1, \ldots, \xi(\Delta_{i_k}) = 1,$$
$$\xi(\Delta_{i_{k+1}}) = 0, \ldots, \xi(\Delta_{i_n}) = 0\}, \quad (3)$$

т. е. A_1 представляет собой сумму всех тех событий, входящих в событие $\{\xi(t) = k\}$, для которых в каждом частичном интервале Δ_i, $i = 1, \ldots, n$, событие A наступает не более одного раза, тогда как A_2 представляет собой сумму всех остальных событий, входящих в $\{\xi(t) = k\}$, т. е. таких, что по крайней мере в одном частичном интервале Δ_i событие A наступает не менее двух раз. Поскольку $A_2 \subset \bigcup\limits_{i=1}^{n} \{\xi(\Delta_i) > 1\}$, то в силу свойства 3)

$$P(A_2) \leqslant P\left(\bigcup_{i=1}^{n} \{\xi(\Delta_i) > 1\}\right) \leqslant \sum_{i=1}^{n} P\{\xi(\Delta_i) > 1\} =$$
$$= no\left(\frac{t}{n}\right) = o(1), \quad n \to \infty. \quad (4)$$

Обозначим

$$p_0 = P\{\xi(\Delta) = 0\}, \quad p_1 = P\{\xi(\Delta) = 1\}. \quad (5)$$

Согласно (3) события, входящие в A_1, попарно несовместны. Отсюда и из свойства 1) следует, что

$$P(A_1) = C_n^k\, p_1^k\, p_0^{n-k}. \quad (6)$$

Из (4) и (6) имеем

$$P\{\xi(t) = k\} = C_n^k\, p_1^k\, p_0^{n-k} + o(1), \quad n \to \infty. \quad (7)$$

Далее, ввиду 3)

$$p_0 = 1 - P\{\xi(\Delta) \geqslant 1\} = 1 - P\{\xi(\Delta) = 1\} - P\{\xi(\Delta) > 1\} =$$
$$= 1 - \lambda\Delta + o(\Delta),$$
$$p_1 = P\{\xi(\Delta) = 1\} = \lambda\Delta + o(\Delta), \quad \Delta = t/n. \quad (8)$$

Подставляя (8) в (7) и переходя к пределу при $n \to \infty$, как при выводе теоремы Пуассона (см. § 6), получим (2)

$$P\{\xi(t) = k\} = \frac{(\lambda t)^k}{k!} e^{-\lambda t}. \quad ▲$$

Аргументом t случайного процесса $\xi(t)$ является длина промежутка времени. Ввиду однородности процесса в качестве промежутка времени можно взять интервал $(0, t)$. Обозначим $\eta(t) = \xi(t)$, где аргументом процесса $\eta(t)$ является текущий момент времени t, так что, например, $\eta(t)$ выражает число α-частиц, испущенных источником к моменту t, или число вызовов на телефонной станции, поступивших к моменту времени t.

Случайный процесс $\eta(t)$ называется процессом Пуассона. Точное определение таково.

Определение. Случайный процесс $\eta(t)$, $0 \leqslant t < \infty$, называется **процессом Пуассона**, или **пуассоновским потоком событий**, если:

1) для любых $0 \leqslant t_1 < t_2 < \ldots < t_n$ случайные величины $\eta(t_2) - \eta(t_1), \ldots, \eta(t_n) - \eta(t_{n-1})$ независимы в совокупности (такие процессы называются **процессами с независимыми приращениями**);

2) случайная величина $\eta(t) - \eta(s)$ имеет распределение Пуассона с параметром $\lambda(t - s)$:

$$P\{\eta(t) - \eta(s) = k\} = \frac{(\lambda(t - s))^k}{k!} e^{-\lambda(t-s)}.$$

Если дополнительно потребовать, чтобы $\eta(0) = 0$, то говорят, что процесс начинается в нуле. Если условиться считать каждую выборочную функцию процесса непрерывной справа, то последние являются целочисленными функциями, возрастающими только целочисленными скачками.

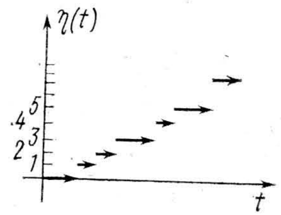


Рис. 24

Обозначим через τ время ожидания первого события в пуассоновском потоке событий. Найдем функцию распределения случайной величины τ. Ясно, что $P\{\tau \geqslant t\} = P\{\eta(t) - \eta(0) = 0\} = e^{-\lambda t}$, и, таким образом, функция распределения времени ожидания равна:

$$F_\tau(t) = P\{\tau < t\} = 1 - e_{-\lambda t} \qquad (9)$$

Следовательно, время ожидания τ распределено с плотностью вероятности

$$p_\tau(t) = \lambda e^{-\lambda t}, \quad 0 \leqslant t < \infty. \qquad (10)$$

Точно так же, если время ожидания отсчитывается от произвольного момента t_1, то

$$P\{\tau \geqslant t\} = P\{\eta(t+t_1) - \eta(t_1) = 0\} = e^{-\lambda t}. \qquad (11)$$

Поэтому время ожидания очередного события не уменьшается от того, что это событие уже ожидалось, скажем, и до момента t_1: в самом деле, ввиду равенства событий $\{\tau - t_1 \geqslant t\} \cap \{\tau \geqslant t_1\} = \{\tau - t_1 \geqslant t\}$ имеем

$$P\{\tau - t_1 \geqslant t \mid \tau \geqslant t_1\} = \frac{P\{\tau - t_1 \geqslant t\}}{P\{\tau \geqslant t_1\}} = \frac{e^{-\lambda(t+t_1)}}{e^{-\lambda t_1}} =$$

$$= e^{-\lambda t} = P\{\tau \geqslant t\}. \qquad (12)$$

Вероятностный смысл параметра λ определяется равенством

$$M\tau = \int_0^\infty t p_\tau(t)\, dt = \int_0^\infty t\lambda e^{-\lambda t}\, dt = \frac{1}{\lambda}, \qquad (13)$$

т. е. λ — среднее число событий, происходящих в единицу времени.

Обратимся к примеру с радиоактивным веществом. Вероятность распада для каждого атома к моменту t равна $P\{\tau < t\} = 1 - e^{-\lambda t}$. Следовательно, к моменту t в среднем распадается $1 - e^{-\lambda t}$ часть всего вещества, и так называемая постоянная t_0 полураспада определяется из равенства $1 - e^{-\lambda t_0} = 1/2$, т. е. $t_0 = (\ln 2)/\lambda$; для радия экспериментально найдено, что $t_0 = 1590$ лет.

Определение. Винеровским процессом $\xi(t)$, $0 \leqslant t < \infty$, начинающимся в нуле, называется случайный процесс, обладающий свойствами:

1) для любых $0 \leqslant t_0 < t_1 < \ldots < t_n$ случайные величины $\eta_1 = \xi(t_1) - \xi(t_0)$, ..., $\eta_n = \xi(t_n) - \xi(t_{n-1})$ независимы;

2) случайная величина $\xi(t) - \xi(s)$ ($0 \leqslant s < t$ — любые) имеет нормальное распределение $N(0, t-s)$;

3) $\xi(0) = 0$.

Согласно определению случайный вектор $\eta = (\eta_1, \ldots, \eta_n)$ распределен нормально с плотностью

$$p_\eta(x_1, \ldots, x_n) = \prod_{j=1}^n \{2\pi(t_j - t_{j-1})\}^{-1/2} \exp\left\{-\frac{x_j^2}{2(t_j - t_{j-1})}\right\}. \qquad (14)$$

Полагая $t_0 = 0$, в силу условия $\xi(0) = 0$ найдем $\xi(t_1) = \eta_1$, $\xi(t_2) = \eta_1 + \eta_2$, ..., $\xi(t_n) = \eta_1 + \ldots + \eta_n$, т. е. случайные векторы $\eta = (\eta_1, \ldots, \eta_n)$ и $\xi = (\xi(t_1), \ldots, \xi(t_n))$ связаны невырожденным линейным преобразованием с определителем, равным единице. Поэтому случайный вектор ξ также нормален и имеет плотность вероятности $p_\xi(x_1, \ldots, x_n)$, задаваемую формулой

$$p_\xi(x_1, \ldots, x_n) =$$

$$= \prod_{j=1}^n \{2\pi(t_j - t_{j-1})\}^{-1/2} \exp\left\{-\frac{(x_j - x_{j-1})^2}{2(t_j - t_{j-1})}\right\}, \quad x_0 = 0.$$

Представим себе частицу, взвешенную в однородной жидкости. Она испытывает хаотические столкновения с молекулами жидкости и в результате беспорядочно движется. Это так называемое *броуновское движение* частицы. Рассмотрим дискретную модель броуновского движения, в которой частица меняет свое положение лишь скачками в фиксированные моменты времени, кратные Δt. Обозначим величину скачка Δx. Тогда, находясь в момент t в точке x, частица в следующий момент $t + \Delta t$ с равными вероятностями окажется в одной из соседних точек $x - \Delta x$ или $x + \Delta x$. Модель броуновского движения можно получить в результате предельного перехода при $\Delta t \to 0$, $\Delta x \to 0$. Смещение частицы к моменту $t + s$ равно

$$\xi(t+s) = [\xi(t+s) - \xi(s)] + [\xi(s) - \xi(0)], \quad \xi(0) = 0, \qquad (15)$$

причем в рассматриваемой модели случайные величины в скобках, очевидно, независимы для любых s, $t \geqslant 0$ и, более того, распределение вероятностей $\xi(t+s) - \xi(s)$ точно такое же, как и $\xi(t) - \xi(0) = \xi(t)$. Поэтому

$$D\xi(t+s) = D[\xi(t+s) - \xi(s)] + D[\xi(s) - (0)] =$$

$$= D\xi(t) + D\xi(s).$$

Отсюда следует, что дисперсия является линейной функцией t

$$D\xi(t) = \sigma^2 t, \quad 0 \leqslant t < \infty, \qquad (16)$$

где постоянная σ^2 называется *коэффициентом диффузии*.

Будем считать, что в начальный момент броуновская частица находилась в нуле: $\xi(0) = 0$. Тогда к моменту t она совершит $n = t/\Delta t$ скачков, и если обозначить через ξ_k случайную величину, принимающую значения $\pm\Delta x$ с вероятностью $1/2$, то $\xi(t) = \sum_{k=1}^{t/\Delta t} \xi_k$. Поэтому $D\xi(t) = \sum_{k=1}^{t/\Delta t} D\xi_k$, а поскольку $D\xi_k = \frac{1}{2}(\Delta x)^2 + \frac{1}{2}(\Delta x)^2 = (\Delta x)^2$, то

$$D\xi(t) = (\Delta x)^2 \cdot \frac{t}{\Delta t}. \qquad (17)$$

Сравнивая (16) и (17), найдем, что $\sigma^2=(\Delta x)^2/\Delta t$. Рассмотрим случайную величину

$$\xi_n(t)=\frac{\xi(t)}{\sqrt{D\xi(t)}}=\frac{\xi(t)}{\sigma\sqrt{t}}=\frac{1}{\sigma\sqrt{t}}\sum_{k=1}^{\frac{t}{\Delta t}}\xi_k,\quad n=\frac{t}{\Delta t}. \tag{18}$$

Поскольку $M\xi_n(t)=0$ и $D\xi_n(t)=1$, то в силу центральной предельной теоремы при $\Delta t\to 0$ и $\Delta x\to 0$, так что $(\Delta x)^2/\Delta t=\sigma^2$, $\xi(t)$ сходится по распределению к нормальному закону. Поэтому для предельной случайной величины $\tilde{\xi}(t)$, которую естественно рассматривать как «непрерывную» модель броуновского движения, получим

$$P\left\{x_1\leqslant\frac{\tilde{\xi}(t)}{\sigma\sqrt{t}}\leqslant x_2\right\}=\frac{1}{\sqrt{2\pi}}\int_{x_1}^{x_2}e^{-x^2/2}\,dx. \tag{19}$$

Мы нашли распределение вероятностей $\tilde{\xi}(t)$ — положения броуновской частицы в момент t. То же распределение справедливо и для любого смещения частицы за время t, ибо в силу однородности процесса броуновского движения приращение $\tilde{\xi}(t+s)-\tilde{\xi}(s)$ имеет такое же распределение вероятностей, что и приращение $\tilde{\xi}(t)-\tilde{\xi}(0)=\tilde{\xi}(t)$, и, стало быть,

$$P\left\{x_1\leqslant\frac{\tilde{\xi}(t+s)-\tilde{\xi}(s)}{\sigma\sqrt{t}}\leqslant x_2\right\}=\frac{1}{\sqrt{2\pi}}\int_{x_1}^{x_2}e^{-x^2/2}dx. \tag{20}$$

Таким образом, броуновское движение частицы — винеровский процесс.

В общем случае винеровский процесс, в отличие от рассмотренного нами, определяется так, что случайная величина $\xi(t)-\xi(s)$ распределена $N(\theta(t-s),\ \sigma^2(t-s))$, где θ — *коэффициент сноса*, σ^2 — *коэффициент диффузии*. Ранее мы ввели так называемый **стандартный винеровский процесс**: $\xi(t)-\xi(s)\in N(0,\ t-s)$. Общий случай получается в результате преобразования $\xi(t)\to\sigma\xi(t)+\theta(t)$.

Оба рассмотренных процесса, пуассоновский и винеровский, являются примерами однородных процессов с независимыми приращениями $\xi(t_1)-\xi(t_0)$, $\xi(t_2)-\xi(t_1),\ldots$. **Однородным**, как уже указывалось выше, называется случайный процесс, у которого распределение $\xi(t)-\xi(s)$ зависит только от разности $t-s$. Отличную от условий 1)—3) характеризацию винеровского процесса (так сказать, «внутреннюю») мы дадим при рассмотрении непрерывных марковских процессов. Для винеровского процесса 1)—3) можно показать [1], что его траектории (т. е. реализации) непрерывны с вероятностью 1. Это позволяет ввести для винеровского процесса (в

частности, непрерывного броуновского движения) важную случайную величину τ_x-момент времени первого достижения траекторией точки x, т. е. τ_x — время ожидания события: первый раз $\xi(t)=x$, $\xi(0)=0$. Пусть $x>0$. Частица может оказаться правее точки x, $\xi(t)\geqslant x$, лишь при условии, что до этого она побывала в точке x (так как траектория непрерывна), т. е. $\tau_x\leqslant t$. Таким образом, событие $\{\xi(t)\geqslant x\}$ влечет событие $\{\tau_x\leqslant t\}$. Поэтому

$$P\{\xi(t)\geqslant x\mid\tau_x\leqslant t\}=\frac{P\{\xi(t)\geqslant x\}}{P\{\tau_x\leqslant t\}}. \tag{21}$$

Очевидно, если в момент τ_x, $\tau_x\leqslant t$, частица находится в точке x, то после выхода из x к моменту t частица с вероятностью $^1/_2$ окажется правее x, с той же вероятностью она окажется левее x. Поэтому $P\{\xi(t)\geqslant x\mid\tau_x\leqslant t\}=1/2$, и из (21) находим

$$P\{\tau_x\leqslant t\}=2P\{\xi(t)\geqslant x\}=2P\left\{\frac{\xi(t)}{\sigma\sqrt{t}}\geqslant\frac{x}{\sigma\sqrt{t}}\right\}=$$

$$=\sqrt{\frac{2}{\pi}}\int_{\frac{x}{\sigma\sqrt{t}}}^{\infty}e^{-\frac{z^2}{2}}dz,\quad t>0. \tag{22}$$

Отсюда плотность вероятности τ_x равна

$$p_{\tau_x}(t)=\frac{x}{\sqrt{2\pi}\,\sigma t^{3/2}}e^{-\frac{x^2}{2\sigma^2 t}},\quad 0\leqslant t<\infty. \tag{23}$$

Рассмотрим теперь другую важную случайную величину: $\hat{\xi}(t)$ — *максимальное смещение* броуновской частицы за время t, $\hat{\xi}(t)=\max\limits_{0\leqslant s\leqslant t}\xi(s)$, $\xi(0)=0$. Имеем согласно (22)

$$P\{\hat{\xi}(t)\geqslant x\}\equiv P\{\max_{0\leqslant s\leqslant t}\xi(s)\geqslant x\}=P\{\tau_x\leqslant t\}=$$

$$=\sqrt{\frac{2}{\pi}}\int_{\frac{x}{\sigma\sqrt{t}}}^{\infty}e^{-\frac{z^2}{2}}dz. \tag{24}$$

Отсюда, поскольку $\sqrt{\frac{2}{\pi}}\int_0^\infty e^{-\frac{z^2}{2}}dz=1$,

$$F_{\hat{\xi}}(x)=1-P\{\hat{\xi}(t)\geqslant x\}=\sqrt{\frac{2}{\pi}}\int_0^{\frac{x}{\sigma\sqrt{t}}}e^{-\frac{z^2}{2}}dz, \tag{25}$$

и, значит, $\hat{\xi}(t)$ имеет плотность вероятности

$$p_{\hat{\xi}(t)}(x) = \frac{1}{\sigma}\sqrt{\frac{2}{\pi t}}\,e^{-\frac{x^2}{2\sigma^2 t}},\quad 0 \leqslant x < \infty \tag{26}$$

$(p_{\hat{\xi}(t)}(x) = 0$ при $x < 0$, так как $\hat{\xi}(t) = \max\limits_{0 \leqslant s \leqslant t} \xi(s) \geqslant \xi(0) = 0)$.

Распределение (26) называется удвоенным нормальным законом, поскольку ввиду (22) и (24) $P\{\hat{\xi}(t) \geqslant x\} = 2P\{\xi(t) \geqslant x\}$.

2.° Марковские процессы

Мы уже рассматривали частный случай марковских процессов — конечные однородные цепи Маркова. В этом пункте мы изучим основные свойства марковских процессов с непрерывным временем.

Определение. Случайный процесс $\xi(t)$ называется **марковским**, если для любого момента времени t_1 при известном значении $\xi(t_1)$ случайные величины $\xi(t)$ с $t > t_1$ не зависят от случайных величин $\xi(s)$ с $s < t_1$. Таким образом, марковские процессы (*процессы без последействия*) характеризуются тем, что вероятностные свойства процесса в момент $t \geqslant t_1$ определяются состоянием в момент t_1 и не зависят от состояний процесса до момента t_1.

Рассмотрим вначале марковские процессы с непрерывным временем и с конечным или счетным множеством состояний $\Omega = (\omega_1, \omega_2, \ldots)$. Остановимся подробно на случае конечного числа состояний, он отличается от цепей Маркова тем, что здесь время изменяется непрерывно и переход системы из одного состояния ω_i в другое ω_j происходит в любой момент времени. Введем *вероятности перехода* $p_{ij}(s, t)$:

$$p_{ij}(s, t) = P\{\xi(t) = \omega_j \mid \xi(s) = \omega_i\},\quad p_{ij}(s, t) \geqslant 0. \tag{27}$$

Очевидно,

$$\sum_{j=1}^{N} p_{ij}(s, t) = 1,\quad s \leqslant t,\quad p_{ii}(t, t) = 1,\quad p_{ij}(t, t) = 0,\quad i \neq j. \tag{28}$$

Обозначим через a_i начальное распределение вероятностей, $t = t_0$,

$$a_i \geqslant 0,\quad \sum_{i=1}^{N} a_i = 1, \tag{29}$$

а через $p_j(t)$ — *абсолютную вероятность*, т. е. вероятность того, что система будет в состоянии ω_j в момент t, $t \geqslant t_0$. В силу формулы полной вероятности очевидны следующие равенства:

$$p_j(t) = \sum_{k=1}^{N} a_k p_{kj}(t_0, t),\quad t \geqslant t_0 \tag{30}$$

$$p_{ij}(s, t) = \sum_{k=1}^{N} p_{ik}(s, t_1)\,p_{kj}(t_1, t),\quad s < t_1 < t. \tag{31}$$

Теорема 2. Пусть переходные вероятности $p_{ij}(s, t)$ имеют частные производные по t и s. Тогда при $s \leqslant t$

$$\frac{\partial p_{ij}(s, t)}{\partial t} = \sum_{k=1}^{N} p_{ik}(s, t)\,A_{kj}(t), \tag{32}$$

$$\frac{\partial p_{ij}(s, t)}{\partial s} = -\sum_{k=1}^{N} A_{ik}(s)\,p_{kj}(s, t), \tag{33}$$

где

$$A_{ik}(t) = \left[\frac{\partial p_{ik}(t, u)}{\partial u}\right]\Bigg|_{u=t},\quad A_{ik}(t) \geqslant 0,\quad i \neq k,\quad A_{kk}(t) \leqslant 0,$$
$$\sum_{k=1}^{N} A_{ik}(t) = 0,\quad i = 1, \ldots, N. \tag{34}$$

Доказательство. Для $t > s$ имеем в силу (31)

$$\frac{\partial p_{ij}(s, t)}{\partial t} = \lim_{\Delta \to 0} \frac{p_{ij}(s, t+\Delta) - p_{ij}(s, t)}{\Delta} =$$
$$= \lim_{\Delta \to 0} \frac{1}{\Delta}\left[\sum_{k=1}^{N} p_{ik}(s, t)\,p_{kj}(t, t+\Delta) - p_{ij}(s, t)\right] =$$
$$= \lim_{\Delta \to 0}\left[\sum_{\substack{k=1\\ k \neq j}}^{N} p_{ik}(s, t)\,\frac{p_{kj}(t, t+\Delta)}{\Delta} + p_{ij}(s, t)\,\frac{p_{jj}(t, t+\Delta) - 1}{\Delta}\right]. \tag{35}$$

Рассмотрим выражение в последней квадратной скобке. В силу (28) и дифференцируемости $p_{ij}(s, t)$ по t имеем

$$\lim_{\Delta \to 0} \frac{p_{jj}(t, t+\Delta) - 1}{\Delta} = \frac{\partial p_{jj}(t, u)}{\partial u}\bigg|_{u=t} \equiv A_{jj}(t) \tag{36}$$

$$\lim_{\Delta \to 0} \frac{p_{kj}(t, t+\Delta)}{\Delta} = \frac{\partial p_{kj}(t, u)}{\partial u}\bigg|_{u=t} \equiv A_{kj}(t),\quad k \neq j.$$

Подставляя эти выражения в (35), получим первую систему уравнений (32). Сопоставляя (36) и (28), заключаем ввиду неравенств $0 \leqslant p_{ij}(t, u) \leqslant 1$, что $A_{ik} \geqslant 0$, $i \neq k$, $A_{kk} \leqslant 0$, $\sum\limits_{k=1}^{N} A_{ik}(t) = 0$, и (34) также доказано. Далее, при $t > s$ имеем в силу (31)

$$\frac{\partial p_{ij}(s, t)}{\partial s} = \lim_{\Delta \to 0} \frac{p_{ij}(s+\Delta, t) - p_{ij}(s, t)}{\Delta} =$$

$$= \lim_{\Delta\to 0} \frac{1}{\Delta}\left[p_{ij}(s+\Delta, t) - \sum_{k=1}^{N} p_{ik}(s, s+\Delta)\, p_{kj}(s+\Delta, t)\right] =$$

$$= -\lim_{\Delta\to 0}\left[\frac{p_{ii}(s, s+\Delta)-1}{\Delta} p_{ij}(s+\Delta, t) + \sum_{\substack{k=1\\ k\neq i}}^{N} \frac{p_{ik}(s, s+\Delta)}{\Delta} p_{kj}(s+\Delta, t)\right]. \qquad (37)$$

Отсюда, рассуждая как и выше, получим вторую систему дифференциальных уравнений (33). ▲

Уравнения (32) называются **прямой**, а уравнения (33) **обратной системой уравнений Колмогорова.**

Физический смысл функций $A_{ij}(t)$ состоит в том, что $A_{ij}(t)dt$ есть вероятность перехода из состояния ω_i в состояние ω_j за время от t до $t+dt$. Если функции $A_{ij}(t)$ непрерывны, то функции $p_{ij}(s, t)$ представляют единственное решение системы уравнений (32), удовлетворяющее начальным условиям $p_{ii}(s, s)=1$, $p_{ij}(s, s)=0$, $i\neq j$. Таким образом, рассматриваемый марковский процесс полностью определяется заданием функций $A_{ij}(t)$. Нетрудно показать также, что если заданы любые непрерывные функции $A_{ij}(t)$, удовлетворяющие условиям $A_{ij}\geqslant 0$, $i\neq j$, $A_{ij}\leqslant 0$, $\sum_{j=1}^{N} A_{ij}=0$, то решение $p_{ij}(s, t)$ системы (32) при начальных условиях $p_{ii}(s, s)=1$, $p_{ij}(s, s)=0$, $i\neq j$, будет неотрицательным ($p_{ij}(s, t)\geqslant 0$), справедливо равенство (31), так что $p_{ij}(s, t)$ будут определять некоторый марковский процесс.

Отправляясь от прямой системы уравнений Колмогорова, установим дифференциальные уравнения для абсолютных вероятностей $p_i(t)$. Если a_i — начальное распределение вероятностей (29), то

$$p_i(t) = \sum_{j=1}^{N} a_j p_{ji}(t_0, t). \qquad (38)$$

Дифференцируя (38) по t и используя сначала (32), а затем (38), найдем

$$\frac{dp_i(t)}{dt} = \sum_{j=1}^{N} a_j \sum_{k=1}^{N} p_{jk}(t_0, t) A_{ki}(t) = \sum_{k=1}^{N} A_{ki}(t) p_k(t). \qquad (39)$$

Таким образом, получаем систему уравнений

$$\frac{dp_i(t)}{dt} = \sum_{k=1}^{N} A_{ki}(t) p_k(t), \quad i=1, \ldots, N. \qquad (40)$$

Начальные условия для этой системы таковы: $p_i(t_0)=a_i$, $i=1, \ldots, N$.

Для однородного марковского процесса переходные вероятности $p_{ij}(s, t)$ зависят лишь от разности $t-s$. В этом случае согласно определению (34) A_{ij} — константы, и система уравнений (32) принимает вид

$$\frac{dp_{ij}(t)}{dt} = \sum_{k=1}^{N} p_{ik}(t) A_{kj}, \quad i, j=1, \ldots, N. \qquad (41)$$

Замечание. В случае счетного числа состояний $\Omega=(\omega_1, \omega_2, \ldots)$ прямая и обратная системы уравнений Колмогорова (32) и (34) остаются справедливыми, но для их обоснования надо дополнительно требовать равномерную сходимость соответствующих рядов [6].

Примеры.

1. *Двусторонняя реакция.* Система может находиться в двух состояниях ω_1 и ω_2 (ω_1 — нераспавшаяся частица, ω_2 — распавшаяся). Пусть возможен как процесс распада — переход из ω_1 в ω_2 с вероятностью $\alpha\Delta t+o(\Delta t)$ за время Δt, так и процесс восстановления — переход из ω_2 в ω_1 с вероятностью $\beta\Delta t+o(\Delta t)$ за время Δt. Итак, в этом случае $A_{12}=\alpha$, $A_{21}=\beta$, а, стало быть, $A_{11}=-\alpha$, $A_{22}=-\beta$. Уравнения (40) дают

$$\frac{dp_1(t)}{dt} = -\alpha p_1(t) + \beta p_2(t), \quad \frac{dp_2(t)}{dt} = \alpha p_1(t) - \beta p_2(t). \qquad (42)$$

Пусть задано начальное распределение вероятностей

$$a_1=1, \; a_2=0. \qquad (43)$$

Подставляя в первое уравнение (42) $p_2(t)=1-p_1(t)$, найдем $\frac{dp_1(t)}{dt} = -(\alpha+\beta)p_1(t)+\beta$, отсюда получим решение

$$p_1(t) = e^{-(\alpha+\beta)t} + \frac{\beta}{\alpha+\beta}\left(1-e^{-(\alpha+\beta)t}\right),$$

$$p_2(t) = \frac{\alpha}{\alpha+\beta}\left(1-e^{-(\alpha+\beta)t}\right). \qquad (44)$$

При $t\to\infty$ $p_1(t)\to\beta/(\alpha+\beta)$, $p_2(t)\to\alpha/(\alpha+\beta)$, т. е. процесс эргодичен. Если восстановление невозможно (например, радиоактивный распад), то $\beta=0$ и

$$p_1(t)=e^{-\alpha t}, \; p_2(t) = 1-e^{-\alpha t}.$$

2. *Двухпозиционное реле.* Пусть двухпозиционное реле находится под воздействием случайной последовательности управляющих импульсов, имеющих с одинаковой вероятностью знаки плюс или минус, причем положительный импульс создает или сохраняет состояние ω_1, а отрицательный — со-

стояние ω_2. Тогда $A_{12}=A_{21}=\alpha$, и, значит, согласно (34) $A_{11}=A_{22}=-\alpha$. Это частный случай предыдущего примера: $\beta=\alpha$. Найдем переходные вероятности. Решая систему уравнений (41) с начальными условиями $p_{ik}(t_0)=\delta_{ik}$, i, $k=1,2$, получим

$$p_{12}(t)=p_{21}(t)=\frac{1}{2}(1-e^{-2\alpha(t-t_0)}),$$

$$p_{11}(t)=p_{22}(t)=\frac{1}{2}(1+e^{-2\alpha(t-t_0)}). \qquad (45)$$

При $t\to\infty$ все вероятности перехода стремятся к значению 1/2.

3. *Пуассоновский поток требований.* Пусть на некоторую систему обслуживания поступают требования так, что $\xi(t)$ — число требований за время t — образует однородный марковский процесс со счетным числом состояний $\omega_i=0, 1, 2, \ldots$. Из состояния i система непосредственно может перейти только в состояние $i+1$, $i=0, 1, 2, \ldots$. Таким образом, $A_{i,i+1}=\alpha$, остальные $A_{ij}=0$ при $i\neq j$ и согласно (34) $A_{ii}=-\alpha$. Для абсолютных вероятностей $p_j(t)$, т. е. вероятностей того, что за время t поступит j требований, имеем бесконечную систему уравнений (40)

$$\frac{dp_i(t)}{dt}=\sum_{k=0}^{\infty}A_{ki}p_k(t) \qquad (46)$$

с начальными условиями

$$p_0(0)=1,\ p_i(0)=0,\ i=1, 2, \ldots. \qquad (47)$$

Перепишем систему (46) подробней

$$\frac{dp_0(t)}{dt}=-\alpha p_0(t), \qquad (48)$$

$$\frac{dp_i(t)}{dt}=\alpha p_{i-1}(t)-\alpha p_i(t),\ i=1, 2, \ldots.$$

Система (48) легко решается с помощью введения новых функций $q_k(t)$ по формуле $p_k(t)=e^{-\alpha t}q_k(t)$, $k=0, 1, 2, \ldots$. Тогда

$$\frac{dq_0(t)}{dt}=0, \qquad (49)$$

$$\frac{dq_i(t)}{dt}=\alpha q_{i-1}(t),\ i=1, 2, \ldots,$$

с начальными условиями

$$q_0(0)=1,\ q_i(0)=0,\ i=1, 2, \ldots.$$

Интегрируя последовательно уравнения (49) при $i=0, 1, 2, \ldots$, получим

$$q_0(t)=1,\ q_1(t)=\alpha t, \ldots, q_k(t)=\frac{(\alpha t)^k}{k!}, \ldots$$

и

$$p_k(t)=\frac{(\alpha t)^k}{k!}e^{-\alpha t},\ k=0, 1, 2, \ldots, \qquad (50)$$

т. е. рассмотренный поток требований является пуассоновским.

4. Пусть однородный марковский процесс обладает свойством эргодичности: $p_{ij}(t)\to p_j$ при $t\to\infty$. Тогда, как видно из системы (41), стационарные вероятности p_j удовлетворяют системе уравнений

$$\sum_{i=1}^{\infty}p_iA_{ik}=0,\ k=1, 2, \ldots,$$

кроме того, выполнено условие нормировки $\sum\limits_{i=1}^{\infty}p_i=1$.

Пусть, например, $A_{i,i+1}=A$, $A_{i+1,i}=B$, $B>A$, $A_{11}=-A$, $A_{ii}=-(A+B)$, $i>1$, $A_{ij}=0$ для других i и j. Тогда указанная выше система уравнений имеет вид

$$Bp_2-Ap_1=1,\ k=1,$$

$$Bp_{k+1}-Ap_k=Bp_k-Ap_{k-1},\ k=2, 3, \ldots.$$

Отсюда $Bp_k=Ap_{k-1}$, $k=2, 3, \ldots$, т. е. $p_k=(A/B)^{k-1}p_1$ и условие нормировки $\sum\limits_{i=1}^{\infty}p_i=1$ дает $p_1=1-A/B$. Таким образом, стационарные вероятности в этом случае равны

$$p_k=\left(1-\frac{A}{B}\right)\left(\frac{A}{B}\right)^{k-1},\ k=1, 2, \ldots.$$

Теперь мы рассмотрим марковские процессы $\xi(t)$ с непрерывным множеством состояний. Здесь наиболее интересным и важным для физики является случай процессов, у которых n-мерная функция распределения с любым n имеет плотность распределения вероятностей. Точнее: пусть $\xi(t)$ — случайный процесс, $t\in T$, и пусть при каждом наборе моментов времени $t_1, t_2, \ldots, t_n\in T$ n-мерная случайная величина $(\xi(t_1), \ldots, \xi(t_n))$ имеет n-мерную плотность вероятности $p_n(t_1, x_1; \ldots; t_n, x_n)$. Эта плотность обладает двумя очевидными свойствами:

1) $p_n(t_1, x_1; \ldots; t_n, x_n)$ симметрична относительно любых перестановок пар аргументов (t_i, x_i), ибо $p_n(t_1, x_1; \ldots; t_n, x_n)dx_1\ldots dx_n$ выражает вероятность совместного осуществления событий $x_i\leqslant\xi(t_i)\leqslant x_i+dx_i$, $i=1, \ldots, n$, и, стало быть не зависит от порядка их перечисления;

2) все конечномерные плотности p_n для различных n должны быть согласованы в том смысле, что плотность любого k-мерного распределения при $k<n$ определяется с помощью n-мерного распределения:

$$p_k(t_1, x_1; \dots ; t_k, x_k) = \underbrace{\int \dots \int_{-\infty}^{\infty}}_{n-k} p_n(t_1, x_1; \dots ; t_k, x_k;$$

$$t_{k+1}, x_{k+1}; \dots ; t_n, x_n)\, dx_{k+1} \dots dx_n. \tag{51}$$

Согласно определению плотности условной вероятности (см. § 8) имеем равенство

$$p_n(t_1, x_1; \dots; t_n, x_n) = $$
$$= p_{n-1}(t_1, x_1; \dots; t_{n-1}, x_{n-1}) q_n(t_n, x_n | t_1, x_1; \dots; t_{n-1}, x_{n-1}). \tag{52}$$

Свойство марковости процесса означает, как уже известно, что вероятностные свойства процесса в момент t_n определяются состоянием в момент t_{n-1} и не зависят от протекания процесса в предшествующие моменты времени, в силу чего

$$q_n(t_n, x_n | t_1, x_1; \dots; t_{n-1}, x_{n-1}) = q(t_n, x_n | t_{n-1}, x_{n-1}). \tag{53}$$

Условную плотность вероятности $q(t, x|\tau, y)$ называют **переходной плотностью вероятности.**

Сопоставляя (52) и (53), находим

$$p_n(t_1, x_1; \dots; t_n, x_n) =$$
$$= p_{n-1}(t_1, x_1; \dots; t_{n-1}, x_{n-1}) q(t_n, x_n | t_{n-1}, x_{n-1}), \tag{54}$$

а применяя эту формулу последовательно для n, $n-1, \dots, 2$, получим

$$p_n(t_1, x_1; \dots; t_n, x_n) =$$
$$= p_1(t_1, x_1) q(t_2, x_2 | t_1, x_1) \dots q(t_n, x_n | t_{n-1}, x_{n-1}). \tag{55}$$

Равенство (55) означает, что для задания n-мерной плотности вероятности марковского процесса достаточно знать лишь две функции: одномерную плотность $p_1(t, x)$ и переходную плотность вероятности $q(t, x|\tau, y)$.

Основным в теории непрерывных марковских процессов является уравнение Смолуховского (оно также называется уравнением Колмогорова—Чепмена):

$$q(t, x | t_0, x_0) = \int_{-\infty}^{\infty} q(t, x | \tau, y) q(\tau, y | t_0, x_0)\, dy \tag{56}$$

для любых трех моментов времени $t_0 < \tau < t$, $t_0, \tau, t \in T$. В самом деле, согласно (55) имеем

$$p_3(t_0, x_0; \tau, y; t, x) = p_1(t_0, x_0) q(\tau, y | t_0, x_0) q(t, x | \tau, y) \tag{57}$$

и

$$p_2(t_0, x_0; t, x) = p_1(t_0, x_0) q(t, x | t_0, x_0). \tag{58}$$

В силу условия согласования (51)

$$p_2(t_0, x_0; t, x) = \int_{-\infty}^{\infty} p_3(t_0, x_0; \tau, y; t, x)\, dy. \tag{59}$$

Подставляя (57) и (58) в (59) и сокращая обе части равенства на $p_1(t_0, x_0)$, получим (56).

Для *однородного* марковского процесса плотность вероятности перехода зависит только от разности моментов времени $q(t, x|\tau, y) = q(x | t-\tau, y)$. В этом случае уравнение Смолуховского (56) имеет вид

$$q(x | t - t_0, x_0) = \int_{-\infty}^{\infty} q(x | t - \tau, y) q(y | \tau - t_0, x_0)\, dy. \tag{60}$$

Переходим к выводу уравнения Эйнштейна—Фоккера—Планка. В дальнейшем при формулировке условий и интерпретации результатов нам будет удобно использовать терминологию, связанную с блужданием частицы под действием случайных толчков. Эта интерпретация тем более уместна, что мы ограничиваемся рассмотрением так называемых *диффузионных процессов,* которые характеризуются выполнением следующих трех условий.

$$1)\ \lim_{\Delta \to 0} \frac{1}{\Delta} \int_{-\infty}^{\infty} (x - y) q(t + \Delta, x | t, y)\, dx = A(y, t); \tag{61}$$

интеграл в этом выражении означает условное среднее значение перемещения частицы за время Δ из фиксированной точки y, так что $A(y, t)$ — средняя скорость изменения состояния в точке y в момент времени t, $A(y, t)$ называется *коэффициентом сноса.*

$$2)\ \lim_{\Delta \to 0} \frac{1}{\Delta} \int_{-\infty}^{\infty} (x - y)^2 q(t + \Delta, x | t, y)\, dx = B(y, t) > 0; \tag{62}$$

здесь интеграл дает меру разброса конечных точек x относительно исходной точки y. Условие 2) означает, что этот разброс растет при малых Δ пропорционально Δ, т. е. по диффузионному закону. Коэффициент $B(y, t)/2$ называется *коэффициентом диффузии.*

$$3)\ \lim \frac{1}{\Delta} \int_{-\infty}^{\infty} |x - y|^3 q(t + \Delta, x | t, y)\, dx = 0. \tag{63}$$

Это условие означает, что в малых промежутках времени вероятность больших значений $|x-y|$ мала (в самом деле, при малых Δ интеграл в (62) больше интеграла в (63), т. е. основную роль в интегралах играют малые значения $|x-y|$). С физической точки зрения именно это условие позволяет рассматривать $\xi(t)$ как непрерывно (во времени) меняющуюся величину, т. е. как среднее за время, гораздо большее промежутка между последовательными случайными толчками.

Теорема 3. Пусть выполнены условия (61)—(63) и $q(t, x|t_0, x_0) \in C^{(2,1)}(R_1 \times T)$ как функция x и t, $A(x, t) \in C^{(1,0)}(R_1 \times T)$, $B(x, t) \in C^{(2,0)}(R_1 \times T)$. Тогда переходная плотность вероятности $q(t, x|t_0, x_0)$ как функция x и t удовлетворяет уравнению

$$\frac{\partial q(t, x|t_0, x_0)}{\partial t} = -\frac{\partial}{\partial x}(A(x, t)\, q(t, x|t_0, x_0)) + \\ + \frac{1}{2}\frac{\partial^2}{\partial x^2}(B(x, t)\, q(t, x|t_0, x_0)). \qquad (64)$$

Доказательство. Умножив уравнение Смолуховского (56) на произвольную финитную функцию $\varphi(x) \in C_0^\infty(R_1)$, проинтегрируем его по x от $-\infty$ до ∞ и, предварительно изменив порядок интегрирования (что возможно в силу финитности $\varphi(x)$), представим $\varphi(x)$ в виде разложения по формуле Тейлора:

$$\int_{-\infty}^{\infty} \varphi(x)\, q(t+\Delta, x|t_0, x_0)\, dx =$$

$$= \int_{-\infty}^{\infty} \varphi(x)\, dx \int_{-\infty}^{\infty} q(t+\Delta, x|t, y)\, q(t, y|t_0, x_0)\, dy =$$

$$= \int_{-\infty}^{\infty} q(t, y|t_0, x_0)\, dy \int_{-\infty}^{\infty} q(t+\Delta, x|t, y) \times$$

$$\times \left\{\varphi(y) + \varphi'(y)(x-y) + \frac{\varphi''(y)}{2}(x-y)^2 + \varphi'''(z)\frac{(x-y)^3}{6}\right\} dx,$$

$$z \in (y, x), \quad \Delta > 0. \qquad (65)$$

Поскольку

$$\int_{-\infty}^{\infty} q(t+\Delta, x|t, y)\, dx = 1,$$

то после деления на Δ преобразуем (65) к виду

$$\int_{-\infty}^{\infty} \varphi(x)\frac{q(t+\Delta, x|t_0, x_0) - q(t, x|t_0, x_0)}{\Delta}\, dx =$$

$$= \int_{-\infty}^{\infty} q(t, y|t_0, x_0)\, dy \left\{\varphi'(y)\frac{1}{\Delta}\int_{-\infty}^{\infty}(x-y)\, q(t+\Delta, x|t, y)\, dx + \right.$$

$$+ \frac{1}{2}\varphi''(y)\frac{1}{\Delta}\int_{-\infty}^{\infty}(x-y)^2 q(t+\Delta, x|t, y)\, dx +$$

$$\left. + \frac{1}{6}\frac{1}{\Delta}\int_{-\infty}^{\infty}(x-y)^3\varphi'''(z)\, q(t+\Delta, x|t, y)\, dx\right\}. \qquad (66)$$

В силу финитности $\varphi(x)$ и непрерывности подинтегральных функций можно в равенстве (66) перейти к пределу при $\Delta \to 0$ под знаком интеграла. Получим с учетом условий (61)—(63)

$$\int_{-\infty}^{\infty} \varphi(x)\frac{\partial q(t, x|t_0, x_0)}{\partial t}\, dx =$$

$$= \int_{-\infty}^{\infty} q(t, y|t_0, x_0)\left\{\varphi'(y)\, A(y, t) + \frac{1}{2}\varphi''(y)\, B(y, t)\right\} dy. \qquad (67)$$

Последний член в (66) пропадает ввиду (63):

$$\lim_{\Delta\to 0}\frac{1}{\Delta}\left|\int_{-\infty}^{\infty}(x-y)^3 q(t+\Delta, x|t, y)\,\varphi'''(z)\, dx\right| \leqslant$$

$$\leqslant c\lim_{\Delta\to 0}\frac{1}{\Delta}\int_{-\infty}^{\infty}|x-y|^3 q(t+\Delta, x|t, y)\, dx = 0,$$

$$c = \sup_{z\in R_1}|\varphi'''(z)| < \infty.$$

Правую часть в (67) проинтегрируем по частям, получим с учетом финитности $\varphi(x)$:

$$\int_{-\infty}^{\infty} q(t, y|t_0, x_0)\left\{\varphi'(y)\, A(y, t) + \frac{1}{2}\varphi''(y)\, B(y, t)\right\} dy =$$

$$= -\int_{-\infty}^{\infty}\varphi(y)\left[\frac{\partial}{\partial y}(qA) - \frac{1}{2}\frac{\partial^2}{\partial y^2}(qB)\right] dy. \qquad (68)$$

Подставляя (68) в (67), найдем

$$\int_{-\infty}^{\infty}\varphi(x)\left\{\frac{\partial q(t, x|t_0, x_0)}{\partial t} + \frac{\partial}{\partial x}(q(t, x|t_0, x_0)\, A(x, t)) - \right.$$

$$-\frac{1}{2}\frac{\partial^2}{\partial x^2}(q(t, x\,|\,t_0, x_0)B(x, t))\Big\}\,dx=0. \qquad (69)$$

Поскольку здесь $\varphi(x)$ — любая финитная функция, а выражение в фигурных скобках в (69) является непрерывной функцией x, то из (69) следует справедливость равенства (64) ▲

Уравнение (64) называется уравнением Эйнштейна—Фоккера—Планка или прямым (первым) уравнением Колмогорова. Это — уравнение параболического типа. По самому смыслу переходной плотности вероятностей следует искать решения уравнения (64), удовлетворяющие условиям

$$q(t, x\,|\,t_0, x_0)\geqslant 0,\quad \int_{-\infty}^{\infty} q(t, x\,|\,t_0, x_0)\,dx=1,$$

$$\int_{-\infty}^{\infty} q(t, x\,|\,t_0, x_0)\,dx_0=1 \qquad (70)$$

и начальному условию (см. ниже формулу (77))

$$q(t, x\,|\,t_0, x_0)\big|_{t=t_0}=\delta(x-x_0). \qquad (71)$$

Переходим теперь к выводу обратного (второго) уравнения Колмогорова

Теорема 4. Пусть выполнены условия (61)—(63) и $q(t, x|t_0, x_0)\in C^{(3,1)}(R_1\times T)$ как функция x_0 и t_0, причем $\sup\limits_{x_0\in R_1}\left|\frac{\partial^3 q}{\partial x_0^3}\right|<\infty$, $t\neq t_0$. Тогда переходная плотность вероятности как функция x_0 и t_0 удовлетворяет уравнению

$$\frac{\partial q(t, x\,|\,t_0, x_0)}{\partial t_0}=-A(x_0, t_0)\frac{\partial q(t, x\,|\,t_0, x_0)}{\partial x_0}-$$

$$-\frac{B(x_0, t_0)}{2}\frac{\partial^2 q(t, x\,|\,t_0, x_0)}{\partial x_0^2}. \qquad (72)$$

Доказательство. Взяв $\Delta>0$ так, что $t_0+\Delta<t$, разложим по формуле Тейлора переходную плотность вероятности под знаком интеграла в уравнении Смолуховского (56):

$$q(t, x\,|\,t_0, x_0)=\int_{-\infty}^{\infty} q(t, x\,|\,t_0+\Delta, y)\,q(t_0+\Delta, y\,|\,t_0, x_0)\,dy=$$

$$=\int_{-\infty}^{\infty} q(t_0+\Delta, y\,|\,t_0, x_0)\Big\{q(t, x\,|\,t_0+\Delta, x_0)+$$

$$+\frac{\partial q(t, x\,|\,t_0+\Delta, x_0)}{\partial x_0}(y-x_0)+\frac{1}{2}\frac{\partial^2 q(t, x\,|\,t_0+\Delta, x_0)}{\partial x_0^2}(y-x_0)^2+$$

$$+\frac{1}{6}\frac{\partial^3 q(t, x\,|\,t_0+\Delta, z)}{\partial x_0^3}(y-x_0)^3\Big\}\,dy,\quad z\in(y, x_0). \qquad (73)$$

Поскольку

$$\int_{-\infty}^{\infty} q(t_0+\Delta, y\,|\,t_0, x_0)\,dy=1,$$

то на основании соотношения (73) получим

$$\frac{q(t, x\,|\,t_0, x_0)-q(t, x\,|\,t_0+\Delta, x_0)}{\Delta}=$$

$$=\frac{\partial q(t, x\,|\,t_0+\Delta, x_0)}{\partial x_0}\frac{1}{\Delta}\int_{-\infty}^{\infty} q(t_0+\Delta, y\,|\,t_0, x_0)(y-x_0)\,dy+$$

$$+\frac{1}{2}\frac{\partial^2 q(t, x\,|\,t_0+\Delta, x_0)}{\partial x_0^2}\frac{1}{\Delta}\int_{-\infty}^{\infty} q(t_0+\Delta, y\,|\,t_0, x_0)(y-x_0)^2\,dy+$$

$$+\frac{1}{6\Delta}\int_{-\infty}^{\infty}\frac{\partial^3 q(t, x\,|\,t_0+\Delta, z)}{\partial x_0^3}q(t_0+\Delta, y\,|\,t_0, x_0)(y-x_0)^3\,dy. \qquad (74)$$

Используя условие $\sup\limits_{z\in R_1}\left|\frac{\partial^3 q(t, x\,|\,t_0+\Delta, z)}{\partial x_0^3}\right|\leqslant c$ для всех $\Delta>0$, таких, что $t_0+\Delta<t$, и условия (61)—(63), перейдем в (74) к пределу при $\Delta\to 0$, получим уравнение (72)

$$-\frac{\partial q(t, x\,|\,t_0, x_0)}{\partial t_0}=A(x_0, t_0)\frac{\partial q(t, x\,|\,t_0, x_0)}{\partial x_0}+$$

$$+\frac{B(x_0, t_0)}{2}\frac{\partial^2 q(t, x\,|\,t_0, x_0)}{\partial x_0^2}.\ \blacktriangle \qquad (75)$$

Уравнение (72) следует решать в обратную сторону по времени, для $t_0\leqslant t$, при этом решение $q(t, x|t_0, x_0)$ снова должно быть подчинено условиям (70) и (71).

Пусть в начальный момент t_0 задана плотность распределения вероятности $p(t_0, x)$ случайной величины $\xi(t_0)$. Тогда двумерная плотность вероятности для произвольного момента времени $t\geqslant t_0$ и начального момента t_0 равна

$$p_2(t, x;\ t_0, x_0)=p(t_0, x_0)\,q(t, x|t_0, x_0), \qquad (76)$$

а одномерная плотность случайного процесса $\xi(t)$ для момента времени t равна согласно (51)

$$p_1(t, x)=\int_{-\infty}^{\infty} p(t_0, x_0)\,q(t, x\,|\,t_0, x_0)\,dx_0. \qquad (77)$$

Если мы теперь уравнение (64) Эйнштейна—Фоккера—Планка умножим на $p(t_0, x_0)$ и проинтегрируем по x_0, то в силу (77), очевидно, найдем, что плотность $p_1(t, x)$ удовлетворяет тому же уравнению

$$\frac{\partial p_1(t, x)}{\partial t} = -\frac{\partial}{\partial x}(A(x, t)\, p_1(t, x)) + \frac{1}{2}\frac{\partial^2}{\partial x^2}(B(x, t)\, p_1(t, x)). \tag{78}$$

Решение уравнения (78) должно удовлетворять условиям

$$p_1(t, x) \geqslant 0, \quad \int_{-\infty}^{\infty} p_1(t, x)\, dx = 1, \quad p_1(t, x)|_{t=t_0} = p(t_0, x). \tag{79}$$

Примеры.

1. Пусть марковский процесс однороден по времени, т. е. $q(t, x|t_0, x_0) = q(x|t-t_0, x_0)$. В этом случае в условиях (61), (62) функции A и B не зависят от t. Пусть, кроме того, одномерная плотность p_1 также не зависит от времени (*стационарный* марковский процесс). Тогда уравнение (78) записывается в виде

$$\frac{d}{dx}\left[A(x)\, p_1(x) - \frac{1}{2}\frac{d}{dx}(B(x)\, p_1(x))\right] = 0. \tag{80}$$

Если на границах изменения x (т. е. области значения процесса $\xi(t)$) поток $Ap_1 - \frac{1}{2}\frac{d}{dx}(Bp_1)$ равен нулю, то в силу (80) он равен нулю всюду

$$A(x)\, p_1(x) - \frac{1}{2}\frac{d}{dx}(B(x)\, p_1(x)) = 0. \tag{81}$$

Интегрируя уравнение (81) (для этого полагаем $v = Bp_1$), получим

$$p_1(x) = \frac{c}{B(x)} e^{\int_0^x \frac{2A(y)}{B(y)} dy}, \tag{82}$$

c — постоянная, определяемая из условия нормировки (79). Физическим примером, в котором существует стационарное распределение $p_1(x)$, является броуновское движение частицы над отражающей границей при наличии силы тяжести. Здесь $A = -mg$. Ясно, что на отражающей границе выполнено условие обращения потока в нуль, так что уравнение (81) имеет место. Выражение (82) с постоянными A и B дает

$$p_1(x) = ce^{-\frac{2mqx}{B}}, \tag{83}$$

т. е. мы получили барометрическую формулу. Известно, что $B = 2kT$, k — постоянная Больцмана, T — абсолютная температура. Итак,

$$p_1(x) = ce^{-\frac{mgx}{kT}}. \tag{84}$$

2. Рассмотрим марковский процесс, однородный по координате: $q(t, x|t_0, x_0) = q(t, x-x_0|t_0)$. В этом случае в условиях (61) и (62) функции A и B зависят лишь от t и уравнение (64) имеет вид

$$\frac{\partial q(t, x-x_0|t_0)}{\partial t} = -A(t)\frac{\partial q}{\partial x} + \frac{B(t)}{2}\frac{\partial^2 q}{\partial x^2}. \tag{85}$$

С помощью замены независимых переменных $(x, t) \to (y, \tau)$ по формулам

$$y = x - x_0 - \int_{t_0}^{t} A(s)\, ds, \quad \tau = \int_{t_0}^{t} B(s)\, ds, \tag{86}$$

приходим к уравнению теплопроводности

$$\frac{\partial q}{\partial \tau} = \frac{1}{2}\frac{\partial^2 q}{\partial y^2}. \tag{87}$$

Его решение, удовлетворяющее условиям (70) и (71), имеет вид

$$q(\tau, y) = \frac{1}{\sqrt{2\pi\tau}} e^{-\frac{y^2}{2\tau}},$$

или в старых переменных

$$q(t, x - x_0|t_0) =$$
$$= \left(2\pi\int_{t_0}^{t} B(s)\, ds\right)^{-1/2} \exp\left\{-\frac{1}{2}\left(x - x_0 - \int_{t_0}^{t} A(s)ds \Big/ \int_{t_0}^{t} B(s)\, ds\right\}. \tag{88}$$

Пусть, в частности, коэффициент сноса $A(t) = 0$, а $B(t) = B$—const. Тогда

$$q(t, x - x_0|t_0) = \frac{1}{(2\pi B(t - t_0))^{1/2}} e^{-\frac{(x-x_0)^2}{2B(t-t_0)}}.$$

Отсюда среднее смещение частицы равно нулю, а средний квадрат смещения частицы равен $B(t-t_0)$, т. е. растет пропорционально времени. Этот результат впервые был получен А. Эйнштейном.

3.° Стационарные процессы

В математических моделях, описывающих ряд явлений в радиофизике, в теории квантовых генераторов, в астрономии и геофизике встречаются случайные функции, при изучении которых нельзя пренебрегать эффектами последействия. Наиболее важными среди них являются так называемые стационарные процессы. Для их описания наиболее естественно рассматривать комплексные случайные величины — напомним соответствующее определение: комплексная случайная величина $\xi=\eta+i\zeta$ — это пара (η, ζ) действительных случайных величин, относительно их совместного распределения не делается никаких предположений, $\bar{\xi}=\eta-i\zeta$, произведение $\xi\bar{\xi}=|\xi|^2$ играет ту же роль, что и ξ^2 в действительном случае, так что, например, $D\xi=M[(\xi-M\xi)(\bar{\xi}-M\bar{\xi})]$. Дадим теперь определение стационарного процесса.

Определение. Случайный процесс $\xi(t)$, $-\infty<t<\infty$, называется **стационарным**, если математическое ожидание $M\xi(t)$ и дисперсия $D\xi(t)$ не зависят от t, а корреляционная функция случайных величин $\xi(t+u)$ и $\xi(t)$ зависит лишь от u, т. е.

$$M\xi(t)=M\xi(0)=a, \tag{89}$$

$$D\xi(t)=D\xi(0)=\sigma^2, \tag{90}$$

$$R[\xi(t+u), \xi(t)]\equiv$$
$$\equiv M[(\xi(t+u)-M\xi(t+u))(\bar{\xi}(t)-M\bar{\xi}(t))]=R(u). \tag{91}$$

Не ограничивая общности, будем считать, что $a=0$, $\sigma^2=1$ (достаточно от $\xi(t)$ перейти к центрированной и нормированной случайной функции $[\xi(t)-a]/\sigma$). Тогда

$$R(u)=M\{\xi(t+u)\bar{\xi}(t)\}=M\{\xi(u)\bar{\xi}(0)\}, \tag{92}$$

$$R(0)=M\{\xi(0)\bar{\xi}(0)\}=D\xi(0)=1. \tag{93}$$

Из определения (91) следует, что если $\xi(t)$ — вещественный процесс, то $R(u)$ — четная функция.

Излагаемая ниже теория стационарных процессов называется *корреляционной теорией* ввиду того, что она выражается в терминах свойств корреляционной функции $R(u)$.

Определение. Случайный процесс называется **непрерывным в среднеквадратичном** (с. к.), если для каждого t, $-\infty<t<\infty$,

$$M|\xi(t+h)-\xi(t)|^2\to 0 \text{ при } h\to 0. \tag{94}$$

В дальнейшем будем рассматривать только с. к. непрерывные стационарные процессы. Для с. к. непрерывного стационарного процесса корреляционная функция $R(u)$ является непрерывной функцией u. В самом деле, используя неравенство Коши—Буняковского, получим

$$|R(u+\Delta u)-R(u)|=|M\{\xi(u+\Delta u)\bar{\xi}(0)\}-M\{\xi(u)\bar{\xi}(0)\}|=$$
$$=|M\{[\xi(u+\Delta u)-\xi(u)]\bar{\xi}(0)\}|\leqslant\{M|\xi(0)|^2\times$$
$$\times M|\xi(u+\Delta u)-\xi(u)|^2\}^{1/2}\to 0 \tag{95}$$

при $\Delta u\to 0$ в силу (94).

Пример стационарного процесса с дискретным спектром. Рассмотрим случайный процесс $\xi(t)$, представляющий собой сумму гармонических колебаний со случайными амплитудами и фазами:

$$\xi(t)=\sum_{k=-n}^{n}\eta_k e^{i\lambda_k t},$$

где λ_k — действительные числа, η_k — некоррелированные случайные величины (комплексные) с нулевым математическим ожиданием: $M\eta_k=0$, $M(\eta_k\bar{\eta}_j)=0$, $k\neq j$, $k, j=-n,\dots, n$. Тогда

$$M\xi(t)=\sum_{k=-n}^{n}e^{i\lambda_k t}M\eta_k=0, \tag{96}$$

$$M\{\xi(t+u)\bar{\xi}(t)\}=\sum_{k=-n}^{n}\sum_{j=-n}^{n}e^{i\lambda_k(t+u)-i\lambda_j t}M(\eta_k\bar{\eta}_j)=$$
$$=\sum_{k=-n}^{n}e^{i\lambda_k u}F_k^2\equiv R(u),$$

где $F^2_k=M|\eta_k|^2$, $k=-n,\dots,n$. Таким образом, $\xi(t)$ — стационарный случайный процесс. Он служит примером случайных колебаний (например, в электрической цепи или нагретой плазме). Его составляющие — гармонические колебания. Величину $M|\xi(t)|^2$ называют энергией стационарного процесса. Из (96) видим, что (при $u=0$)

$$M|\xi(t)|^2=\sum_{k=-n}^{n}F_k^2=\sum_{k=-n}^{n}M|\eta_k|^2, \tag{97}$$

т. е. энергия стационарного процесса $\xi(t)$ складывается из энергий его гармонических составляющих. Значения λ_k называют частотами, а их совокупность $\lambda_{-n},\dots, \lambda_n$ образуют *частотный спектр* стационарного процесса $\xi(t)$.

Для дальнейшего нам понадобятся некоторые сведения о положительно определенных функциях

Определение. Непрерывная функция $f(t)$, $-\infty<t<\infty$, называется **положительно определенной**, если для любых натурального n, вещественных чисел $t_1, t_2, \ldots, t_n$ и комплексных чисел $z_1, z_2, \ldots, z_n$ справедливо неравенство

$$\sum_{k,j=1}^{n} f(t_k - t_j)\, z_k \bar{z}_j \geqslant 0. \qquad (98)$$

Примером положительно определенной функции является характеристическая функция любой случайной величины. Действительно, в силу определения $f_\xi(t) = Me^{it\xi}$ непрерывна и

$$\sum_{k,j=1}^{n} f(t_k - t_j)\, z_k \bar{z}_j = \sum_{k,j=1}^{n} M\, e^{i(t_k - t_j)\xi} z_k \bar{z}_j =$$

$$= M \Big| \sum_{k=1}^{n} e^{it_k\xi} z_k \Big|^2 \geqslant 0. \qquad (99)$$

Мы будем опираться на следующую известную теорему анализа [4].

Лемма (Бохнер—Хинчин). Для того чтобы непрерывная функция $f(t)$, $-\infty < t < \infty$, такая, что $f(0) = 1$, являлась характеристической функцией некоторой случайной величины, необходимо и достаточно, чтобы она была положительно определенной. (Необходимость сформулированных условий очевидна в силу примера (99).)

Теперь сформулируем и докажем основную теорему теории стационарных процессов.

Теорема 5 (Хинчин). Для того чтобы абсолютно интегрируемая на всей прямой функция $R(u)$ являлась корреляционной функцией некоторого с. к. непрерывного стационарного процесса, необходимо и достаточно, чтобы она имела представление вида

$$R(u) = \int_{-\infty}^{\infty} e^{iu\lambda} p(\lambda)\, d\lambda, \qquad (100)$$

где $p(\lambda)$ — некоторая плотность вероятности

Доказательство. Необходимость. Пусть $R(u)$ — корреляционная функция с. к. непрерывного стационарного процесса $\xi(t)$. Тогда она непрерывна (см. (95)), $R(0) = 1$. Покажем, что $R(u)$ положительна определена. В самом деле, каковы бы ни были натуральное n, действительные числа $u_1, \ldots, u_n$ и комплексные $z_1, \ldots, z_n$, будем иметь

$$\sum_{k,j=1}^{n} R(u_k - u_j)\, z_k \bar{z}_j = \sum_{k,j=1}^{n} M\{\xi(u_k)\bar{\xi}(u_j)\}\, z_k \bar{z}_j =$$

$$= M\Big\{\sum_{k,j=1}^{n} \xi(u_k)\,\bar{\xi}(u_j)\, z_k \bar{z}_j\Big\} = M \Big| \sum_{k=1}^{n} \xi(u_k)\, z_k \Big|^2 \geqslant 0. \qquad (101)$$

Таким образом, согласно лемме Бохнера—Хинчина, $R(u)$ — характеристическая функция некоторой случайной величины η, а так как по условию теоремы $R(u)$ абсолютно интегрируема, $\int_{-\infty}^{\infty} |R(u)|\, du < \infty$, то в силу теоремы 4 § 11 случайная величина η непрерывна, ее плотность вероятности распределения — $p(x)$, и, стало быть,

$$R(u) = \int_{-\infty}^{\infty} e^{iu\lambda} p(\lambda)\, d\lambda, \qquad (102)$$

причем

$$p(\lambda) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-iu\lambda} R(u)\, du. \qquad (103)$$

Необходимость доказана.

Достаточность. Пусть справедливо представление (100). Покажем, что существует с. к. непрерывный стационарный процесс, для которого $R(u)$ является корреляционной функцией. Рассмотрим непрерывную случайную величину η с плотностью вероятности $p(x)$ из (100). Обозначим через $\xi_1(t)$ случайную функцию, определяемую равенством $\xi_1(t) = e^{it\eta}$, $-\infty < t < \infty$. Пусть ξ_2 — случайная величина, не зависящая от $\xi_1(t)$ при всех $t \in (-\infty, \infty)$ и принимающая два значения ± 1 с вероятностями 1/2. Положим $\xi(t) = \xi_1(t)\xi_2$. Случайный процесс $\xi(t)$ обладает следующими свойствами:

1) $M\xi(t) = M\xi_1(t) M\xi_2 = 0$, так как $M\xi_2 = 1 \cdot 1/2 - 1 \cdot 1/2 = 0$;
2) $\xi(t)$ — с. к. непрерывен, ибо

$$M|\xi(t+h) - \xi(t)|^2 = M\xi_2^2 \cdot M|e^{i(t+h)\eta} \cdot e^{it\eta}|^2 =$$

$$= M|e^{ih\eta} - 1|^2 = \int_{-\infty}^{\infty} |e^{ihx} - 1|^2 p(x)\, dx \to 0, \quad h \to 0,$$

а $M\xi_2^2 = 1 \cdot \frac{1}{2} + 1 \cdot \frac{1}{2} = 1$;

3) $R(u) = M\{\xi(t+u)\bar{\xi}(t)\} = M\xi_2^2 \cdot M\{e^{i(t+u)\eta} - e^{-it\eta}\} = Me^{iu\eta} =$

$= \int_{-\infty}^{\infty} e^{iux} p(x)\, dx$, что совпадает с (100). ▲

Заметим, что для случая вещественного стационарного процесса представление (100) имеет вид

$$R(u) = \int_{-\infty}^{\infty} \cos u\lambda\, p(\lambda)\, d\lambda. \qquad (104)$$

Стационарный процесс, корреляционная функция которого представима в виде интеграла (100) или (104), называется **стационарным процессом с непрерывным спектром**, а функция $p(\lambda)$ называется **спектральной плотностью** стационарного процесса.

Средняя энергия процесса $\xi(t)$ выражается через его спектральную плотность следующим образом:

$$M|\xi(t)|^2 = R(0) = \int_{-\infty}^{\infty} p(\lambda)\,d\lambda \tag{105}$$

(при нашем соглашении о нормировке $R(0)=1$). Эта формула вполне аналогична выражению для энергии случайного процесса с дискретным спектром, пример которого рассмотрен выше, и объясняет термин «спектральная плотность»: $p(\lambda)$ дает вклад в энергию, отвечающий спектральному интервалу $(\lambda, \lambda+d\lambda)$.

Величина

$$\tau = \int_0^{\infty} |R(t)|\,dt$$

называется **временем корреляции.** Время корреляции τ дает приближенное представление о том, на каких интервалах времени имеется корреляция между значениями (сечениями) случайного процесса. При существенно больших интервалах парными корреляциями можно пренебрегать. Ниже будут даны примеры подсчета времени корреляции.

Одним из важнейших свойств стационарных процессов является их эргодичность (см. ниже теорему 7).

Теорема 6. Для корреляционной функции $R(u)$ с. к. непрерывного стационарного процесса существует предел

$$\lim_{T\to\infty} \frac{1}{T}\int_0^T R(u)\,du = 0. \tag{106}$$

Доказательство. Ограничимся случаем вещественного стационарного процесса с непрерывным спектром. На основании представления (104) имеем

$$\frac{1}{T}\int_0^T R(u)\,du = \frac{1}{T}\int_0^T du \int_{-\infty}^{\infty} \cos u\lambda\, p(\lambda)\,d\lambda = \int_{-\infty}^{\infty} \frac{\sin T\lambda}{T\lambda} p(\lambda)\,d\lambda =$$

$$= \int_{-\delta}^{\delta} \frac{\sin T\lambda}{T\lambda} p(\lambda)\,d\lambda + \int_{|\lambda|>\delta} \frac{\sin T\lambda}{T\lambda} p(\lambda)\,d\lambda. \tag{107}$$

Пусть $\varepsilon>0$ — любое. Первый интеграл в правой части (107) оценим следующим образом, учитывая, что $|\sin T\lambda| \leqslant T|\lambda|$

для всех λ и T:

$$\left|\int_{-\delta}^{\delta} \frac{\sin T\lambda}{T\lambda} p(\lambda)\,d\lambda\right| \leqslant \int_{-\delta}^{\delta} p(\lambda)\,d\lambda < \frac{\varepsilon}{2}, \tag{108}$$

если $\delta=\delta(\varepsilon)$ достаточно мало, поскольку $p(\lambda)$ — непрерывная функция. Зафиксировав такое δ, оценим второй интеграл в правой части (107)

$$\left|\int_{|\lambda|\geqslant\delta} \frac{\sin T\lambda}{T\lambda} p(\lambda)\,d\lambda\right| \leqslant \frac{1}{T\delta}\int_{|\lambda|\geqslant\delta} p(\lambda)\,d\lambda \leqslant \frac{1}{T\delta} < \frac{\varepsilon}{2}, \tag{109}$$

если $T>0$ — достаточно велико $(T\geqslant 2/\delta(\varepsilon)\varepsilon)$. Из (108) и (109) заключаем, что

$$\lim_{T\to\infty} \frac{1}{T}\int_0^T R(u)\,du = 0. \blacktriangle \tag{110}$$

Рассмотрим ряд примеров, иллюстрирующих теорию стационарных процессов.

Примеры.

1). а) Пусть $\xi(t)$ — экспоненциально-коррелированный стационарный процесс, т. е. процесс, имеющий корреляционную функцию

$$R(u) = e^{-\beta|u|}, \quad \beta > 0.$$

Его спектральная плотность равна согласно (103)

$$p(\lambda) = \frac{\pi^{-1}\beta}{\lambda^2+\beta^2},$$

а время корреляции $\tau = 1/\beta$.

б) Если корреляционная функция $R(u)$ стационарного процесса имеет вид

$$R(u) = e^{-\beta|u|}\cos\omega u, \quad \beta, \omega > 0,$$

то его спектральная плотность равна

$$p(\lambda) = \frac{\beta}{\pi}\,\frac{\lambda^2+\omega^2+\beta^2}{(\lambda^2-\omega^2-\beta^2)^2+4\beta^2\lambda^2}.$$

При выполнении условия $\beta \ll \omega$ спектральная плотность сосредоточена в относительно узкой полосе частот $\lambda - \omega \sim \beta$ и справедливо приближенное соотношение

$$p(\lambda) \approx \frac{\beta}{\pi}\,\frac{\lambda^2+\omega^2}{(\lambda^2-\omega^2)^2+4\beta^2\lambda^2} \approx \frac{\beta}{2\pi\,[(\lambda-\omega)^2+\beta^2]},$$

а время корреляции

$$\tau \approx \frac{2}{\pi}\,\frac{1}{\beta}.$$

в) Если корреляционная функция $R(u)$ имеет вид

$$R(u) = e^{-\frac{1}{4}\beta^2 u^2},$$

то

$$p(\lambda) = \frac{1}{\sqrt{\pi}\,\beta} e^{-\lambda^2/\beta^2}, \text{ а } \tau = \frac{1}{\beta}\sqrt{\frac{\pi}{2}}.$$

2. *Влияние фильтра.*

Пусть

$$\xi(t) = \sum_{k=1}^{n} c_k \eta(t - \tau_k),$$

где c_k, τ_k — постоянные, $k = 1, \ldots, n$, причем τ_k — вещественные числа, а $\eta(t)$ — стационарный случайный процесс, с непрерывным спектром, $M\eta(t) = 0$. Тогда случайный процесс $\xi(t)$, называемый линейным преобразованием процесса $\eta(t)$, также стационарен. Действительно,

$$M\xi(t) = 0,$$

$$R_\xi(u) = M\{\xi(u)\bar{\xi}(0)\} = M\left\{\sum_{k=1}^{n} c_k \eta(u - \tau_k) \sum_{j=1}^{n} \overline{c_j \eta(-\tau_j)}\right\} =$$

$$= \sum_{k,j=1}^{n} c_k \bar{c}_j M\{\eta(u - \tau_k)\overline{\eta(-\tau_j)}\} = \sum_{k,j=1}^{n} c_k \bar{c}_j R_\eta(u - \tau_k + \tau_j).$$

Отсюда получаем выражение для спектральной плотности процесса $\xi(t)$.

$$p_\xi(\lambda) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-i\lambda u} R_\xi(u)\,du =$$

$$= \frac{1}{2\pi}\sum_{k,j=1}^{n} c_k \bar{c}_j \int_{-\infty}^{\infty} R_\eta(u - \tau_k + \tau_j) e^{-i\lambda u}\,du =$$

$$= \sum_{k,j=1}^{n} c_k \bar{c}_j e^{-i\lambda(\tau_k - \tau_j)} p_\eta(\lambda) = \left|\sum_{k=1}^{n} c_k e^{-i\lambda\tau_k}\right|^2 p_\eta(\lambda).$$

Таким образом, спектральная плотность процесса $\xi(t)$ получается из спектральной плотности процесса $\eta(t)$ умножением на квадрат модуля частотной характеристики преобразования (фильтра).

3. *Эргодичность стационарного процесса.* Оказывается, что аналог равенства (106) имеет место и для самого стационарного процесса $\xi(t)$. Однако прежде чем сформулировать и доказать соответствующую теорему, мы введем новое понятие: интеграла от случайной функции, так называемого стохастического интеграла. Случайная функция (случайный процесс) $\xi(t)$ — это функция двух переменных: $\xi(t) = \xi(t, \omega)$, $t \in T$, $\omega \in \Omega$; пусть множество T, ради определенности, интервал $[a, b]$. Даже в тех (довольно редких) случаях, когда $\xi(t, \omega)$ как функция t, $t \in [a, b]$, оказывается интегрируемой (по Риману) функцией при всех $\omega \in \Omega$, так что интеграл $\int_a^b \xi(t)\,dt \equiv \int_a^b \xi(t, \omega)\,dt$ имеет смысл при каждом фиксированном $\omega \in \Omega$, вопрос о том, является ли этот интеграл случайной величиной, определенной на Ω, нетривиален. Мы дадим определение стохастического интеграла, основанное на сходимости в среднем квадратичном.

Выше (см. (94)) было дано определение с. к. непрерывного случайного процесса. Усилим это определение: случайная функция $\xi(t)$ называется **равномерно непрерывной в среднеквадратичном,** если

$$M|\xi(t+h) - \xi(t)|^2 \to 0, \quad h \to 0, \tag{111}$$

равномерно по $t \in [a, b]$.

Переходим к определению стохастического интеграла для равномерно с. к. непрерывной случайной функции $\xi(t)$. Возьмем произвольное разбиение сегмента $[a, b]$ точками t_k:

$$a = t_0 < t_1 < \ldots < t_n = b,$$

$$\Delta t_k = t_{k+1} - t_k, \quad \Delta = \max_k \Delta t_k, \quad k = 0, 1, \ldots, n.$$

Составим интегральную сумму

$$\sum_{k=0}^{n-1} \xi(t_k)\,\Delta t_k \equiv \eta(\Delta), \tag{112}$$

$\eta(\Delta)$ — случайная величина, $M\eta(\Delta) = \sum_{k=0}^{n-1} M\xi(t_k)\,\Delta t_k$. Определим для случайной величины ξ норму по формуле (если $M\xi^2 < \infty$)

$$\|\xi\| = \sqrt{M\xi^2}. \tag{113}$$

Так определенная величина обладает всеми свойствами нормы, в частности, для нее справедливо неравенство треугольника (следствие неравенства Коши—Буняковского, см. § 9):

$$\|\xi_1 - \xi_2\| \leqslant \|\xi_1 - \xi_3\| + \|\xi_2 - \xi_3\|.$$

Покажем, что при измельчении разбиения сегмента $[a, b]$ так, что $\Delta \to 0$, последовательность интегральных сумм (112) фундаментальна в норме (113). В самом деле, пусть $\{t'_k\}$ и $\{t''_k\}$ — два произвольных разбиения сегмента $[a, b]$, $\Delta_1 = \max_k \Delta t'_k$, $\Delta_2 = \max_k \Delta t''_k$. Обозначим через $\{t_k\}$ суммарное разбиение, т. е. разбиение сегмента $[a, b]$, в которое входят все различные точки обоих разбиений $\{t'_k\}$ и $\{t''_k\}$, тогда если $\Delta = \max_k \Delta t_k$, то $\Delta \leqslant \min\{\Delta_1, \Delta_2\}$. Имеем

$$\left\| \sum_k \xi(t'_k)\,\Delta t'_k - \sum_k \xi(t''_k)\,\Delta t''_k \right\| = \left\| \sum_k [\xi(t'_k) - \xi(t''_k)]\,\Delta t_k \right\| \leqslant$$
$$\leqslant \sum_k \|\xi(t'_k) - \xi(t''_k)\|\,\Delta t_k < \varepsilon \sum_k \Delta t_k = \varepsilon(b-a) \qquad (114)$$

ввиду (111), если Δ_1 и $\Delta_2 < \delta(\varepsilon)$, предпоследнее неравенство справедливо в силу неравенства Минковского. Можно показать [3], что фундаментальная в норме (113) последовательность $\eta(\Delta)$ сходится при $\Delta \to 0$ к некоторой случайной величине. Эта случайная величина называется по определению **стохастическим интегралом** от $\xi(t)$ и обозначается $\int_a^b \xi(t)\,dt.$

Очевидно, что этот интеграл обладает свойством линейности

$$\int_a^b (c_1\xi_1(t) + c_2\xi_2(t))\,dt = c_1 \int_a^b \xi_1(t)\,dt + c_2 \int_a^b \xi_2(t)\,dt, \qquad (115)$$

и, кроме того,

$$M \int_a^b \xi(t)\,dt = \int_a^b M\,\xi(t)\,dt. \qquad (116)$$

Теорема 7. Для равномерно с. к. непрерывного стационарного процесса $\xi(t)$ случайная величина

$$\bar{\xi}(T) \equiv \frac{1}{T} \int_0^T \xi(t)\,dt \xrightarrow[p]{} 0, \quad T \to \infty \qquad (117)$$

(сходится по вероятности к нулю при $T \to \infty$).

Доказательство. Снова ограничимся случаем вещественного стационарного процесса $\xi(t)$ с непрерывным спектром. Покажем, что для любого $\varepsilon > 0$

$$\lim_{T\to\infty} P\{|\bar{\xi}(T)| > \varepsilon\} = 0. \qquad (118)$$

Поскольку

$$M\,\bar{\xi}(T) = \frac{1}{T} \int_0^T M\,\xi(t)\,dt = 0,$$

то в силу неравенства Чебышева

$$P\{|\bar{\xi}(T)| > \varepsilon\} \leqslant \frac{D\,\bar{\xi}(T)}{\varepsilon^2}, \qquad (119)$$

и нам достаточно доказать, что

$$\lim_{T\to\infty} D\,\bar{\xi}(T) = 0.$$

Так как $M\bar{\xi}(T) = 0$, то для $D\bar{\xi}(T)$ имеем выражение

$$D\bar{\xi}(T) = M(\bar{\xi}(T))^2 = \frac{1}{T^2} M\left(\int_0^T \xi(t)\,dt \int_0^T \xi(u)\,du\right) =$$
$$= \frac{1}{T^2} \int_0^T \int_0^T M\,\{\xi(t)\,\xi(u)\}\,dt\,du = \frac{1}{T^2} \int_0^T \int_0^T R(t-u)\,dt\,du. \qquad (120)$$

Отсюда, пользуясь представлением (104), получим

$$D\,\bar{\xi}(T) = \frac{1}{T^2} \int_0^T \int_0^T dt\,du \int_{-\infty}^{\infty} \cos(t-u)\lambda\,p(\lambda)\,d\lambda =$$
$$= \frac{1}{T^2} \int_{-\infty}^{\infty} p(\lambda)\,d\lambda \int_0^T \int_0^T \cos\lambda(t-u)\,dt\,du =$$
$$= 2 \int_{-\infty}^{\infty} \frac{1-\cos\lambda T}{\lambda^2 T^2}\,p(\lambda)\,d\lambda. \qquad (121)$$

Теперь интеграл в (121) оценим, как в теореме 6,

$$D\bar{\xi}(T) = 2 \int_{-\infty}^{\infty} \frac{1-\cos\lambda T}{\lambda^2 T^2}\,p(\lambda)\,d\lambda =$$
$$= 2 \int_{-\delta}^{\delta} \frac{1-\cos\lambda T}{\lambda^2 T^2}\,p(\lambda)\,d\lambda + 2 \int_{|\lambda|\geqslant\delta} \frac{1-\cos\lambda T}{\lambda^2 T^2}\,p(\lambda)\,d\lambda. \qquad (122)$$

Поскольку $0 \leqslant 1 - \cos\lambda T \leqslant \lambda^2 T^2/2$ для всех λ и T, то первый интеграл в (122) меньше любого $\varepsilon_1 > 0$, если $\delta = \delta(\varepsilon_1)$ достаточно мало. Зафиксировав это δ, второй интеграл в (122) оценим так

$$\left| 2 \int_{|\lambda|\geqslant\delta} \frac{1-\cos\lambda T}{\lambda^2 T^2}\,p(\lambda)\,d\lambda \right| \leqslant \frac{4}{\delta^2 T^2} \int_{|\lambda|\geqslant\delta} p(\lambda)\,d\lambda \leqslant \frac{4}{\delta^2 T^2} < \varepsilon_1,$$

если T достаточно велико. Тем самым равенство (120) доказано. Отсюда, как указано выше, следует (118). ▲

4. *Спектр колебания с флуктуирующей частотой.* Пусть задано колебание

$$\xi(t) = A_0 \cos\left[\omega_0 t + \int_{t_0}^{t} \eta(t)\,dt + \varphi_0\right], \quad t \geqslant t_0, \tag{123}$$

в котором девиация частоты $\eta(t)$ — вещественный стационарный равномерно с. к. непрерывный случайный процесс, а φ_0 — случайная величина, равномерно распределенная в интервале $[0, 2\pi]$. Рассмотрим задачу о том, как связан спектр колебания $\xi(t)$ со спектром девиации частоты $\eta(t)$. Естественно считать, что $M\eta(t)=0$. Согласно представлению (104) корреляционная функция $R_\eta(u)$ имеет вид

$$R_\eta(u) = M\{\eta(t+u)\,\eta(t)\} = \int_0^\infty p_\eta(\lambda)\cos\lambda u\,d\lambda, \tag{124}$$

где $p_\eta(\lambda)$ — спектральная плотность стационарного процесса $\eta(t)$ по положительным частотам. Для случайного набега фазы $\psi(\tau)$ за время от t до $t+\tau$ имеем

$$\psi(\tau) = \int_t^{t+\tau} \eta(z)\,dz. \tag{125}$$

так как $M\eta(t)=0$, то $M\psi(\tau)=0$. Обозначим через $\sigma^2(\tau)$ дисперсию $\psi(\tau)$

$$\sigma^2(\tau) = M\psi^2(\tau) = \int_t^{t+\tau}\int_t^{t+\tau} M\{\eta(z)\,\eta(u)\}\,dz\,du =$$

$$= \int_t^{t+\tau}\int_t^{t+\tau} R_\eta(z-u)\,dz\,du. \tag{126}$$

Подставляя в (126) выражение (124), получим (ср. теорему 7)

$$\sigma^2(\tau) = 2\int_0^\infty p_\eta(\lambda)\,\frac{1-\cos\tau\lambda}{\lambda^2}\,d\lambda. \tag{127}$$

Таким образом, средний квадрат набега фазы определяется спектром девиации частоты. Пусть теперь девиация частоты $\eta(t)$ распределена по нормальному закону. В силу равенства (125) это же верно и в отношении набега фазы $\psi(\tau)$, а так как $M\psi=0$, а дисперсия $\psi(\tau)$ равна $\sigma^2(\tau)$, то плотность распределения $\psi(\tau)$ имеет вид

$$p(x,\tau) = \frac{1}{\sigma(\tau)\sqrt{2\pi}}\,e^{-\frac{x^2}{2\sigma^2(\tau)}}. \tag{128}$$

Найдем корреляционную функцию $R_\xi(u)$ стационарного процесса (123), $R_\xi(u) = M\{\xi(t+u)\,\xi(t)\}$. Имеем

$$\xi(t+u)\,\xi(t) = \frac{A_0^2}{2}\left\{\cos[\omega_0 u - \psi(u)] + \cos\left[\omega_0(2t+u) - \right.\right.$$

$$\left.\left. - \int_{t_0}^{t+u}\eta(\tau)\,d\tau - \int_{t_0}^{t}\eta(\tau)\,d\tau + 2\varphi_0\right]\right\}. \tag{129}$$

Для вычисления $R_\xi(u)$ мы должны усреднить (129) при помощи распределения (128) для ψ и равномерного распределения для φ_0. Интегрирование по φ_0 от 0 до 2π обращает в нуль интеграл от второго косинуса в (129). Таким образом, получаем

$$R_\xi(u) = \frac{A_0^2}{2}\,\frac{1}{\sigma(u)\sqrt{2\pi}}\int_{-\infty}^{\infty}\cos(\omega_0 u - x)\,e^{-\frac{x^2}{2\sigma^2(u)}}\,dx =$$

$$= \frac{A_0^2}{2}\,\frac{\cos\omega_0 u}{\sigma(u)\sqrt{2\pi}}\int_{-\infty}^{\infty}\cos x\,e^{-\frac{x^2}{2\sigma^2(u)}}\,dx =$$

$$= \frac{A_0^2}{2}\,\frac{\cos\omega_0 u}{\sigma(u)\sqrt{2\pi}}\,\frac{1}{2}\int_{-\infty}^{\infty}(e^{ix}+e^{-ix})\,e^{-\frac{x^2}{2\sigma^2(u)}}\,dx =$$

$$= \frac{A_0^2}{2}\cos\omega_0 u\,e^{-\frac{\sigma^2(u)}{2}}. \tag{130}$$

(для вычисления интеграла надо выделить полный квадрат в показателе экспоненты). Поскольку по теореме Хинчина для $R_\xi(u)$ справедиво спектральное представление

$$R_\xi(u) = \int_0^\infty p_\xi(\lambda)\cos\lambda u\,d\lambda, \tag{131}$$

то согласно формуле обращения для интеграла Фурье из (130) и (131) найдем

$$p_\xi(\lambda) = \frac{A_0^2}{2\pi}\int_0^\infty e^{-\frac{\sigma^2(u)}{2}}\cos\omega_0 u\cos\lambda u\,du. \tag{132}$$

Для определения спектра девиации частоты $\eta(t)$ по спектру наблюдаемых колебаний $\xi(t)$ мы приходим к следующему нелинейному интегральному уравнению для $p_\eta(\lambda)$ (подставляем (127) в (132)):

$$p_\xi(\lambda) = \frac{A_0^2}{2\pi} \int_0^\infty \exp\left\{-\int_0^\infty p_\eta(\lambda') \frac{1-\cos u\lambda'}{\lambda'}\, d\lambda'\right\} \cos\omega_0 u \cos\lambda u\, du. \tag{133}$$

Это сложное интегральное уравнение может быть решено при различных упрощающих предположениях [10].

5. *Корреляционная функция стационарного марковского процесса.* Целью этого примера является доказательство **теоремы Дуба**: корреляционная функция $R_\xi(u)$ нормального $N(0,1)$ стационарного марковского процесса равна

$$R_\xi(u) = e^{-\alpha|u|}, \quad \alpha > 0. \tag{134}$$

В самом деле, двумерная плотность нормального $N(0,1)$ процесса $\xi(t)$ равна (см. § 9)

$$p_2(t_1, x_1;\ t_2, x_2) = \frac{1}{2\pi\sqrt{1-\sigma^2}} e^{-\frac{1}{2(1-\sigma^2)}(x_1^2+x_2^2-2\sigma x_1x_2)}, \quad t_2 > t_1, \tag{135}$$

где $\sigma = M\{\xi(t_2)\xi(t_1)\} = R(t_2-t_1)$, поскольку процесс по условию стационарный. Одномерная плотность $p_1(t_1, x_1)$, в свою очередь, равна

$$p_1(t_1, x_1) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x_1^2}{2}}, \tag{136}$$

ибо $\xi(t) \in N(0,1)$ при каждом t. Отсюда для переходной плотности вероятности находим

$$q(t_2, x_2 \mid t_1, x_1) = \frac{p_2(t_1, x_1; t_2, x_2)}{p_1(t_1, x_1)} = \frac{1}{\sqrt{2\pi(1-\sigma^2)}} e^{-\frac{1}{2}\frac{(\sigma x_1 - x_2)^2}{1-\sigma^2}}. \tag{137}$$

Переходная плотность вероятности удовлетворяет уравнению Смолуховского (56)

$$q(t_2, x_2 \mid t_1, x_1) = \int_{-\infty}^{\infty} q(t_2, x_2 \mid \tau, y)\, q(\tau, y \mid t_1, x_1)\, dy, \quad t_1 < \tau < t_2. \tag{138}$$

Подставляя в (138) выражение (137), получим $\sigma^2 = \sigma(t_2-t_1)$, $\sigma_1 = \sigma(t_2-\tau)$, $\sigma_0 = \sigma(\tau-t_1)$

$$\frac{1}{\sqrt{2\pi(1-\sigma_2^2(t_2-t_0)}} e^{-\frac{1}{2}\frac{(\sigma_2(t_2-t_0)x_1-x_2)^2}{1-\sigma_2^2(t_2-t_0)}} = \frac{1}{2\pi\sqrt{(1-\sigma_1^2)(1-\sigma_2^2)}} \times$$

$$\times \int_{-\infty}^{\infty} e^{-\frac{1}{2}\frac{(\sigma_1 y - x_2)^2}{1-\sigma_1^2} - \frac{1}{2}\frac{(\sigma_0 x_1 - y)^2}{1-\sigma_0^2}}\, dy =$$

$$= \frac{1}{\sqrt{2\pi(1-\sigma_1^2\sigma_0^2)}} e^{-\frac{1}{2}\frac{(\sigma_1\sigma_0 x_1 - x_2)^2}{1-\sigma_1^2\sigma_0^2}}. \tag{139}$$

Сравнивая левую и правую части формулы (139), заключаем, что $\sigma_2 = \sigma_1\sigma_0$, или, подробнее,

$$R(t_2-t_1) = R(t_2-\tau)R(\tau-t_1), \quad t_1 < \tau < t_2. \tag{140}$$

Решением этого функционального уравнения служит функция

$$R(u) = e^{-\alpha u}, \quad u \geqslant 0, \tag{141}$$

α — любое комплексное. Поскольку $\xi(t) \in N(0,1)$, то $R(u)$ вещественна и, кроме того, $|R(u)| \leqslant 1$, ибо $|R(u)| = |M\xi(u)\xi(0)| \leqslant \sqrt{M\xi^2(u)M\xi^2(0)} = 1$. Отсюда $\alpha \geqslant 0$, и теорема Дуба доказана.

Английский ботаник Р. Броун обнаружил (1828), что частицы пыльцы, взвешенные в воде, совершают непрерывное беспорядочное движение, названное впоследствии «броуновским». Физическая теория броуновского движения впервые была предложена А. Эйнштейном и Э. Смолуховским (1905). Л. Башелье (1900) нестрого вывел закон движения частицы, совершающей одномерное броуновское движение.

Понятие марковской зависимости было введено А. А. Марковым (1906) для случая конечных однородных цепей. А. Н. Колмогоров в работе 1931 г. дал строгое определение марковской зависимости для непрерывной схемы (время меняется непрерывно, а множество состояний конечно, счетно или непрерывно) и вывел основные дифференциальные уравнения для переходных вероятностей. Колмогоров назвал рассмотренные им процессы стохастически определенными, вскоре для них возник термин «процессы без последействия», а затем по предложению А. Я. Хинчина указанные процессы стали называть «марковскими». Дальнейшее развитие теория марковских процессов получила в работах В. Феллера (1934), П. Леви (1939) и ряда других математиков.

Основы теории стационарных процессов были заложены А. Я. Хинчиным (1934). В настоящее время теория случайных процессов — бурно развивающаяся область математики с широкими приложениями в физике и технике.

ЧАСТЬ 2

МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

Предмет теории вероятностей составляют математические модели так называемых экспериментов со случайным исходом. К математической статистике принято относить все вопросы, касающиеся сравнения этих моделей с реальностью. Сюда относится, например, важный вопрос о состоятельности математической модели реального эксперимента (явления и т. п.). В частности, речь идет о задаче проверки гипотезы о том, что результаты, получаемые на основе математической модели эксперимента, не противоречат опытным данным. Поскольку вывод о непротиворечивости должен быть основан на опытных данных, содержащих элемент случайности, принимаемое решение следует формулировать в теоретико-вероятностных терминах.

Начало периода интенсивного развития статистических методов определяется работами К. Пирсона, выполненными в конце прошлого столетия. Раздел математической статистики, в котором изучаются задачи проверки гипотез, был создан Ю. Нейманом, Э. С. Пирсоном и Р. А. Фишером в конце двадцатых — начале тридцатых годов нашего столетия и с тех пор получил значительное развитие.

Во многих задачах математической статистики опытные данные привлекаются лишь для уточнения математической модели явления. К этому кругу вопросов относятся задачи оценивания параметров, составляющие один из центральных разделов математической статистики. Первые результаты в этой области получены К. Ф. Гауссом (1809) и А. А. Марковым (1900). Методы теории статистического оценивания получили глубокое развитие в работах Р. А. Фишера.

С более общих позиций оба упомянутых раздела математической статистики объединены А. Вальдом (1939). Ему принадлежит единая исчерпывающая формулировка задач общей теории статистических решающих процедур.

§ 14. РАСПРЕДЕЛЕНИЕ ОРТОГОНАЛЬНЫХ ПРОЕКЦИЙ

Пусть ξ — случайный вектор в n-мерном евклидовом пространстве R_n. Если $e_1, \ldots, e_n$ — ортонормированный базис R_n, то

$$\xi = \sum_{j=1}^{n} \xi_j e_j,$$

где ξ_j, $j=1, \ldots, n$, координаты вектора ξ в базисе $e_1, \ldots, e_n$, причем $\xi_j = (\xi, e_j)$, $j=1, \ldots, n$. Скалярные произведения (ξ, e_j), $j=1, \ldots, n$, являются случайными величинами. Скалярное произведение двух случайных векторов ξ и η также является случайной величиной и определяется равенством

$$(\xi, \eta) = \sum_{j=1}^{n} \xi_j \eta_j,$$

где $\eta_j = (\eta, e_j)$, $j=1, \ldots, n$.

Если A — линейный оператор, действующий из R_n в R_n, то $\zeta = A\xi$ — случайный вектор, причем, как известно,

$$\zeta_i = (A\xi)_i = (A\xi, e_i) = \sum_{j=1}^{n} a_{ij} \xi_j, \quad i = 1, \ldots, n,$$

где

$$a_{ij} = (Ae_j, e_i), \quad i, j = 1, \ldots, n$$

матричные элементы матрицы оператора A в ортонормированном базисе. Матрицу оператора A будем обозначать $\breve{A}$.

Напомним, что переход к новому ортонормированному базису $e'_1, \ldots, e'_n$ может быть задан с помощью оператора ортогонального преобразования, или ортогонального оператора U:

$$e'_i = Ue_i, \quad i = 1, \ldots, n.$$

Матрица $\breve{U}$ оператора U в базисе $e_1, \ldots, e_n$ обладает свойствами

$$\det \breve{U} = \pm 1, \quad \breve{U}^{\text{тр}} = \breve{U}^{-1} \tag{1}$$

и называется ортогональной.

В новом базисе $e_1', \ldots, e'_n$ матричные элементы a'_{ij}, $i, j = 1, \ldots, n$, матрицы оператора A вычисляются по формуле

$$a'_{ij} = (Ae_j', e_i') = (AUe_j, Ue_i) = (U^* AUe_j, e_i)$$

и равны

$$a'_{ij} = \sum_{p,q=1}^{n} \overline{u}_{ip} a_{pq} u_{qj}, \quad i, j = 1, \ldots, n,$$

где u_{ij} — матричные элементы матрицы $\tilde{U}$, $\bar{u}_{ij}$ — матричные элементы матрицы $\check{U}^{\text{тр}} = \check{U}^{-1}$, $i, j = 1, \ldots, n$, в базисе $e_1, \ldots, e_n$.

Для любых двух векторов $x, y \in R_n$:

$$(x, y) = (Ux, Uy) \tag{2}$$

и, следовательно, $\|Ux\| = \|x\|$, где $\|x\|$ — обозначение для нормы (длины) вектора $x, \|x\|^2 = (x, x) = \sum_{j=1}^{n} x_j^2$.

Пусть Z — линейное подпространство R_n и $Z^\perp$ ортогональное дополнение Z в R_n, т. е. множество векторов $x \in R_n$, ортогональных всем векторам из L:

$$L^\perp = \{x \in R_n,\ (x, y) = 0,\ y \in L\}.$$

Очевидно, $L^\perp$ — также линейное пространство R_n. Как известно, всякий вектор $x \in R_n$ может быть представлен в виде суммы

$$x = x_1 + x_2,\ x_1 \in L,\ x_2 \in L^\perp. \tag{3}$$

Разложение (3) единственно. Действительно, равенство $x = x_1' + x_2'$, $x_1' \in L$, $x_2' \in L^\perp$, совместно с (3) влечет $(x_1 - x_1') + (x_2 - x_2') = 0$. Слагаемые в последнем равенстве ортогональны, так как $x_1 - x_1' \in L$, $x_2 - x_2' \in L^\perp$, поэтому $\|x_1 - x_1'\|^2 + \|x_2 - x_2'\|^2 = 0$, т. е. $x_1 = x_1'$, $x_2 = x_2'$. Следовательно каждому $x \in R_n$ разложением (3) ставится в соответствие единственный вектор $x_1 \in L$:

$$x_1 = \Pi x. \tag{4}$$

Π называется оператором ортогональнього проектирования на L, или ортогональным проектором на L. Отметим следующие свойства оператора Π.

1) Π — линейный оператор. Действительно, пусть

$$x = x_1 + x_2,\ y = y_1 + y_2;\ x_1, y_1 \in L,\ x_2, y_2 \in L^\perp. \tag{5}$$

Тогда

$$\alpha x + \beta y = (\alpha x_1 + \beta y_1) + (\alpha x_2 + \beta y_2),\ \alpha x_1 + \beta y_1 \in L,\quad \alpha x_2 + \beta y_2 \in L^\perp.$$

Следовательно, согласно определению (4) $\Pi(\alpha x + \beta y) = \alpha x_1 + \beta y_1 = \alpha \Pi x + \beta \Pi y$, так как $\Pi x = x_1$, $\Pi y = y_1$.

2) Π — самосопряженный оператор, т. е. для любых x, $y \in R_n$ $(\Pi x, y) = (x, \Pi y)$. Действительно, воспользовавшись разложением (5), найдем

$$(\Pi x, y) = (x_1, y) = (x_1, y_1) = (x, y_1) = (x, \Pi y).$$

3) Оператор Π удовлетворяет уравнению $\Pi^2 = \Pi$. Действительно, для всякого $x \in R_n: \Pi x = x_1 = \Pi x_1 = \Pi(\Pi x)$, поскольку для $x \in L$ разложение (3) имеет вид $x_1 = x_1 (= \Pi x_1)$.

На самом деле свойства 1)—3) не только необходимы, но и достаточны для того, чтобы оператор Π был ортогональным проектором. Для доказательства предположим, согласно свойству 1, что Π — линейный оператор. Обозначим через L множество решений уравнения $\Pi x = x$, N — множество решений уравнения $\Pi x = 0$. Легко убедиться, что L и N — линейные подпространства R_n, причем ортогональные, если Π удовлетворяет условию 2). В самом деле, если $x \in L$, $y \in N$, то $(x, y) = (\Pi x, y) = (x, \Pi y) = (x, 0) = 0$. Для всякого вектора $x \in R_n$ можно записать тождество

$$x = \Pi x + (I - \Pi) x. \tag{6}$$

Если Π удовлетворяет условию 3), то $\Pi(\Pi x) = \Pi x$, т. е. $\Pi x \in L$, и $\Pi(I - \Pi) x = (\Pi - \Pi^2) x = 0$, т. е. $(I - \Pi) x \in N$. Следовательно, Π — оператор ортогонального проектирования на $L = \{x \in R_n,\ \Pi x = x\}$.

Из разложения (6) следует также, что оператор $I - \Pi$ ортогонально проектирует на $N = \{x \in R_n,\ (I - \Pi) x = x\} = L^\perp$.

Отметим следующее важное свойство ортогонального проектора. Пусть Π — оператор ортогонального проектирования на линейное подпространство L и

$$\rho(x, L) = \inf\{\|x - y\| \mid y \in L\} \tag{7}$$

расстояние от x до L. Тогда

$$\rho(x, L) = \|x - \Pi x\|. \tag{8}$$

Действительно, пусть $y \in L$. Тогда

$$\Pi x - y \in L,\ x - \Pi x = (I - \Pi) x \in L^\perp,$$

и, следовательно,

$$\|x - y\|^2 = \|x - \Pi x + \Pi x - y\|^2 = \|x - \Pi x\|^2 + \\ + \|\Pi x - y\|^2 \geqslant \|x - \Pi x\|^2,$$

причем равенство здесь выполняется лишь в случае $\Pi x = y$.

Теорема 1. Пусть ξ — случайный вектор, координаты которого $\xi_1, \ldots, \xi_n$ в некотором ортонормированном базисе независимы в совокупности и нормальны $N(0, \sigma^2)$. Если U — оператор ортогонального преобразования R_n, то распределение вектора $U\xi$, совпадает с распределением вектора ξ, т. е. случайные величины $(U\xi)_1, \ldots, (U\xi)_n$ распределены так же, как $\xi_1, \ldots, \xi_n$.

Доказательство. По условию теоремы плотность совместного распределения координат $\xi_1, \ldots, \xi_n$ (плотность распределения вектора ξ) равна

$$p_\xi(x) = p_{\xi_1, \ldots, \xi_n}(x_1, \ldots, x_n) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{x_i^2}{2\sigma^2}} =$$

$$= \left(\frac{1}{\sqrt{2\pi}\,\sigma} \right)^n e^{-\frac{(x,\,x)}{2\sigma^2}}, \tag{9}$$

где $x_1, ..., x_n$ координаты вектора $x \in R_n$. Для вектора $\eta = U\xi$ согласно формуле (55) § 8 найдем

$$p_\eta(x) = p_{U\xi}(x) = p_\xi(U^{-1}x)\,|\det \breve{U}^{-1}| = p_\xi(x), \tag{10}$$

так как в равенстве (10) согласно (2) $(U^{-1}(x),\ U^{-1}(x)) = (x,\ x)$ и согласно (1) $|\det \breve{U}^{-1}| = 1$. Из (9) и (10) следует, что координаты $(U\xi)_1, ..., (U\xi)_n$ вектора $U\xi$ независимы в совокупности и нормальны $N(0, \sigma^2)$. ▲

Замечание. Поскольку

$$(U\xi)_i = (U\xi,\ e_i) = (\xi,\ U^{-1}e_i),\ i = 1, ..., n$$

и $e_i' = U^{-1}e_i$, $i = 1, ..., n$, ортонормированный базис R_n, причем в таком виде может быть представлен любой ортонормированный базис R_n, то утверждение теоремы 1 можно сформулировать еще и так: если координаты $\xi_1, .., \xi_n$ случайного вектора ξ независимы в совокупности и нормальны $N(0, \sigma^2)$ в некотором ортонормированном базисе R_n, то они независимы в совокупности и нормальны $N(0, \sigma^2)$ и в любом другом ортонормированном базисе R_n.

Напомним, что сумма m квадратов независимых нормальных $N(0, 1)$ случайных величин имеет χ_m^2-распределение *хи*-квадрат с m степенями свободы (см. § 9). В частности, сумма любых $m < n$ квадратов координат вектора ξ, удовлетворяющего условиям теоремы 1, равна $\sigma^2\chi_m^2$. Но при условиях теоремы 1 имеет место и более общий факт.

Следствие 1. Пусть R_k — k-мерное линейное подпространство R_n, и Π_k — оператор ортогонального проектирования на R_k. Если выполнены условия теоремы 1, то случайная величина

$$\|\Pi_k\xi\|^2 \equiv (\Pi_k\xi,\ \xi) \equiv (\Pi_k\xi,\ \Pi_k\xi)$$

распределена как $\sigma^2\chi_k^2$.

Доказательство. Введем в R_n ортонормированный базис $e_1', ..., e_n'$, причем так, чтобы первые k векторов $e_1', ..., e_k'$ образовывали ортонормированный базис R_k. Согласно замечанию к теореме 1 координаты $\xi_j' = (\xi,\ e_j')$, $j = 1, ..., n$, вектора ξ независимы в совокупности и нормальны $N(0, \sigma^2)$. Поэтому

$$\sum_{j=1}^{k} \xi_j'^2 = \sigma^2\chi_k^2 .$$

С другой стороны, если оператор Π_k ортогонально проектирует на R_k, то

$$\Pi_k\xi = \sum_{j=1}^{k} \xi_j' e_j',$$

так как R_k — линейная оболочка векторов $e_1', ..., e_k'$. Поэтому

$$\|\Pi_k\xi\|^2 = (\Pi_k\xi,\ \xi) = \sum_{j=1}^{k} (\xi_j')^2 = \sigma^2\chi_k^2. \quad ▲$$

Следствие 1 содержит достаточно ясную интерпретацию «числа степеней свободы» χ^2-распределения.

Как известно, если $\xi_1, ..., \xi_n$ независимы в совокупности и нормальны $N(0, 1)$, то случайная величина

$$\varphi_{s,t} = \frac{\sum_{i=1}^{s} \xi_i^2}{\sum_{j=s+1}^{m} \xi_j^2} \cdot \frac{m-s}{s}$$

имеет распределение Снедекора—Фишера со степенями свободы s и $m-s$. Результат не изменится, если $\xi_1, ..., \xi_n$ нормальны $N(0, \sigma^2)$, и, более того, при условиях теоремы 1 имеет место

Следствие 2. Пусть R_s и R_t — линейные подпространства R_n, причем R_s ортогонально R_t. Обозначим через Π_s и Π_t операторы ортогонального проектирования на R_s и R_t соответственно. Если выполнено условие теоремы 1, то

1) векторы $\Pi_s\xi$ и $\Pi_t\xi$ независимы, т. е. в любом базисе R_n совокупности случайных величин $(\Pi_s\xi)_1, ..., (\Pi_s\xi)_n$ и $(\Pi_t\xi)_1, ..., (\Pi_t\xi)_n$ (координат $\Pi_s\xi$ и $\Pi_t\xi$ соответственно) независимы;

2) случайная величина

$$\varphi_{s,t} = (\|\Pi_s\xi\|^2/s)/(\|\Pi_t\xi\|^2/t) \tag{11}$$

имеет распределение Фишера со степенями свободы s и t.

Доказательство. 1) Пусть $e_1, ..., e_n$ ортонормированный базис R_n, такой, что $e_1, ..., e_s$ — ортонормированный базис R_s и $e_{s+1}, ..., e_{s+t}$ — ортонормированный базис R_t. Тогда, как было отмечено в следствии 1,

$$\Pi_s\xi = \sum_{j=1}^{s} (\xi,\ e_j)\,e_j,\ \Pi_t\xi = \sum_{j=s+1}^{s+t} (\xi,\ e_j)\,e_j,$$

и, как следует из замечания к теореме 1, совокупности случайных величин $(\xi,\ e_1)$, ..., $(\xi,\ e_s)$ и $(\xi,\ e_{s+1})$, ..., $(\xi,\ e_{s+t})$ нормальны и независимы. Отсюда, в частности, следует независимость случайных величин

$$\|\Pi_s\xi\|^2 = \sigma^2\chi_s^2,\ \ \|\Pi_t\xi\|^2 = \sigma^2\chi_t^2, \tag{12}$$

а также тот факт, что $\|\Pi_s\xi\|^2/\sigma^2$ и $\|\Pi_t\xi\|^2/\sigma^2$ контролируются χ^2-распределениями с s и t степенями свободы соответственно. В любой другой системе координат координаты случайного вектора $\Pi_s\xi$ выражаются лишь через $(\xi, e_1), \ldots, (\xi, e_s)$, а координаты $\Pi_t\xi$ — соответственно лишь через $(\xi, e_{s+1}), \ldots, (\xi, e_{s+t})$. Отсюда следует независимость случайных векторов $\Pi_s\xi$ и $\Pi_t\xi$.

2) Так как случайные величины (12) независимы, то случайная величина (11) имеет распределение Снедекора—Фишера со степенями свободы s и t по определению. ▲

Подчеркнем, что факты, приведенные в следствиях 1, 2, носят геометрический характер и не зависят от системы координат в R_n.

В дальнейшем будут полезны следующие простые обобщения результатов, доказанных в следствиях. Очевидно, что как $\sigma^2\chi_k^2$ распределена не только случайная величина $\|\Pi_k\xi\|^2$, но также и $\|\Pi_k(\xi+a)\|^2$, где a — любой (случайный или неслучайный) вектор, ортогональный R_k. Дело в том, что в этом случае $\Pi_k a=0$ (если a случайный вектор, то $\Pi_k a=0$ с вероятностью единица). Аналогично утверждения следствия 2 справедливы не только для векторов $\Pi_s\xi$ и $\Pi_t\xi$, но также и для любых векторов $\Pi_s(\xi+a)$ и $\Pi_t(\xi+b)$, если векторы a и b ортогональны R_s и R_t соответственно.

Следствие 3. Пусть Π_1 — оператор ортогонального проектирования на одномерное подпространство R_n, натянутое на единичный вектор e, т. е. $\Pi_1\xi=(\xi, e)e$. Если случайный вектор ξ удовлетворяет условиям теоремы 1, то случайная величина

$$\tau_{n-1}=\frac{(\xi, e)}{[\|(I-\Pi_1)\xi\|^2/(n-1)]^{1/2}} \tag{13}$$

имеет распределение Стьюдента с $n-1$ степенями свободы.

Доказательство. Напомним, что дробь $\tau_k=\eta/[\chi_k^2/k]^{1/2}$ имеет распределение Стьюдента с k степенями свободы, если случайная величина η нормальна $N(0,1)$ и не зависит от χ_k^2, или, что то же самое, если η нормальна $N(0, \sigma^2)$ и $\tau_k=\eta/[\sigma^2\chi_k^2/k]^{1/2}$.

Согласно теореме 1 (ξ, e) имеет распределение $N(0, \sigma^2)$ и не зависит от $\|(I-\Pi_1)\xi\|^2$, как это показано в следствии 2. Дело в том, что оператор $I-\Pi_1$ ортогонально проектирует на $n-1$-мерное пространство векторов, ортогональных e. Отсюда же следует, что $\|(I-\Pi_1)\xi\|^2=\sigma^2\chi^2_{n-1}$, чем и завершается доказательство. ▲

Рассмотрим, в частности, случай, когда $e=\sum_{j=1}^{n} e_j/\sqrt{n}$.

Нетрудно проверить, что все координаты вектора $\Pi_1\xi$ равны

$$(\Pi_1\xi)_j=\frac{1}{n}\sum_{i=1}^{n}\xi_i, \quad j=1, \ldots, n,$$

где ξ_i, $i=1, \ldots, n$, координаты вектора ξ. Следовательно, вектор $(I-\Pi_1)\xi$ имеет координаты

$$((I-\Pi_1)\xi)_j=\xi_j-\frac{1}{n}\sum_{i=1}^{n}\xi_i, \quad j=1, \ldots, n.$$

Поэтому для данного случая τ_{n-1} в (13) принимает вид

$$\tau_{n-1}=\frac{(\xi, e)}{[\|(I-\Pi_1)\xi\|^2/(n-1)]^{1/2}}=$$

$$=\frac{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\xi_i}{\left[\frac{1}{n-1}\sum_{j=1}^{n}\left(\xi_j-\frac{1}{n}\sum_{i=1}^{n}\xi_i\right)^2\right]^{1/2}}. \tag{14}$$

Если ξ_i, $i=1, \ldots, n$, независимы в совокупности и нормальны $N(\mu, \sigma^2)$, то, подставив в (14) вместо вектора ξ вектор $\xi-m$, где m — вектор, все координаты которого равны μ, найдем, что случайная величина

$$\tau_{n-1}=\frac{(\xi-m, e)}{[\|(I-\Pi_1)\xi\|^2/(n-1)]^{1/2}}=$$

$$=\frac{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\xi_i-\mu)}{\left[\frac{1}{n-1}\sum_{j=1}^{n}\left(\xi_j-\frac{1}{n}\sum_{i=1}^{n}\xi_i\right)^2\right]^{1/2}} \tag{15}$$

имеет распределение Стьюдента с $n-1$ степенями свободы. При этом использован тот факт, что $(I-\Pi_1)m=0$.

§ 15. ИНТЕРВАЛЬНЫЕ ОЦЕНКИ ПАРАМЕТРОВ НОРМАЛЬНОГО РАСПРЕДЕЛЕНИЯ

1°. Оценка математического ожидания при известной дисперсии

Пусть ξ_j, $j=1, \ldots, n$, нормальны $N(\mu, \sigma^2)$, причем дисперсия σ^2 считается известной, но неизвестно математическое ожидание μ. Случайные величины $\xi_1, \ldots, \xi_n$ предполагаются независимыми в совокупности. В этом случае последовательность $\xi_1, \ldots, \xi_n$ называется случайной выборкой из нормального распределения $N(\mu, \sigma^2)$ объема n.

Так как случайная величина $\frac{1}{n}\sum_{j=1}^{n}\xi_j$ нормальна $N(\mu, \sigma^2/n)$, то нормальна $N(0, 1)$ случайная величина

$$\eta = \left(\frac{1}{n}\sum_{j=1}^{n}(\xi_j - \mu)\right)\frac{\sqrt{n}}{\sigma}, \tag{1}$$

и, следовательно, вероятность

$$P\{|\eta| < \varepsilon\} = \frac{1}{\sqrt{2\pi}}\int_{-\varepsilon}^{\varepsilon} e^{-x^2/2}\,dx = \Phi(\varepsilon) - \Phi(-\varepsilon),$$

где $\Phi(z)$ — функция нормального распределения

$$\Phi(z) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{z} e^{-x^2/2}\,dx.$$

Задав α, $0<\alpha<1$, и воспользовавшись таблицей нормального распределения, определим $\varepsilon>0$, для которого

$$P\left\{-\varepsilon<\left(\frac{1}{n}\sum_{j=1}^{n}\xi_j-\mu\right)\frac{\sqrt{n}}{\sigma}<\varepsilon\right\}=\Phi(\varepsilon)-\Phi(-\varepsilon)=1-\alpha. \tag{2}$$

При этом случайная величина η (1) может оказаться вне интервала $(-\varepsilon, \varepsilon)$ лишь с вероятностью α. Соотношение (2) можно записать иначе:

$$P\left\{\frac{1}{n}\sum_{j=1}^{n}\xi_j-\frac{\varepsilon\sigma}{\sqrt{n}}<\mu<\frac{1}{n}\sum_{j=1}^{n}\xi_j+\frac{\varepsilon\sigma}{\sqrt{n}}\right\}=1-\alpha. \tag{3}$$

Согласно (3) истинное значение математического ожидания μ покрывается случайным интервалом

$$\left(\frac{1}{n}\sum_{j=1}^{n}\xi_j-\frac{\varepsilon\sigma}{\sqrt{n}},\ \frac{1}{n}\sum_{j=1}^{n}\xi_j+\frac{\varepsilon\sigma}{\sqrt{n}}\right) \tag{4}$$

с вероятностью $1-\alpha$. Вне этого интервала μ может оказаться с вероятностью α. Найденный случайный интервал (4) называется интервальной оценкой параметра μ, или 100 $(1-\alpha)$ % доверительным интервалом для параметра μ.

Понятие интервальной оценки впервые встречается у Лапласа (1814) в связи с задачей определения параметра p биномиального распределения. Однако Лаплас рассматривал

p в указанных задачах как случайную величину. Правильную интерпретацию процедуры интервального оценивания, в которой речь идет о случайном доверительном интервале, предложил Е. Б. Уилсон более чем через 100 лет, в 1927 году.

2°. Интервальная оценка дисперсии при известном математическом ожидании

Пусть в аналогичной ситуации известно математическое ожидание μ, но неизвестна дисперсия σ^2. Тогда случайные величины $\xi_j-\mu$, $j=1, \ldots, n$, нормальны $N(0, \sigma^2)$ и независимы в совокупности. Следовательно, статистика *

$$\sum_{j=1}^{n}(\xi_j-\mu)^2$$

имееет распределение $\sigma^2\chi_n^2$ с n степенями свободы.

Для заданных $\varepsilon_1>0$ и $\varepsilon_2>\varepsilon_1$ по таблице χ^2-распределения можно подсчитать вероятность $P\{\varepsilon_1<\chi_n^2<\varepsilon_2\}=1-\alpha$. Однако при заданном α, $0<\alpha<1$, величины ε_1 и ε_2 вычисляются неоднозначно. Поэтому обычно полагают

$$P\{\chi_n^2\leqslant\varepsilon_1\}=P\{\chi_n^2\geqslant\varepsilon_2\}=\alpha/2$$

так, чтобы

$$P\{\varepsilon_1<\chi_n^2<\varepsilon_2\}=1-P\{\chi_n^2\geqslant\varepsilon_2\}-P\{\chi_n^2\leqslant\varepsilon_1\}=1-\alpha.$$

Подсчитав таким образом ε_1 и ε_2, получим

$$P\left\{\varepsilon_1<\frac{1}{\sigma^2}\sum_{j=1}^{n}(\xi_j-\mu)^2<\varepsilon_2\right\}=1-\alpha. \tag{5}$$

Соотношение (5) можно переписать в виде

$$P\left\{\frac{1}{\varepsilon_2}\sum_{j=1}^{n}(\xi_j-\mu)^2<\sigma^2<\frac{1}{\varepsilon_1}\sum_{j=1}^{n}(\xi_j-\mu)^2\right\}=1-\alpha,$$

или, если воспользоваться обозначениями предыдущего параграфа

$$P\left\{\frac{1}{\varepsilon_2}\|\xi-m\|^2<\sigma^2<\frac{1}{\varepsilon_1}\|\xi-m\|^2\right\}=1-\alpha. \tag{6}$$

Соотношение (6) показывает, что истинное значение дисперсии σ^2 покрывается случайным интервалом

$$\left(\frac{1}{\varepsilon_2}\|\xi-m\|^2,\ \frac{1}{\varepsilon_1}\|\xi-m\|^2\right) \tag{7}$$

* В дальнейшем случайную величину, которая является функцией выборки $\xi_1, \ldots, \xi_n$, условимся называть статистикой.

с вероятностью $1-\alpha$. Соответственно с вероятностью α дисперсия σ^2 может оказаться вне этого интервала. Найденный случайный интервал (7) называется интервальной оценкой дисперсии σ^2, или $100(1-\alpha)\%$ доверительным интервалом для σ^2.

3°. Интервальная оценка математического ожидания при неизвестной дисперсии

Как показано в следствии 3 теоремы 14.1, статистика τ_{n-1} (14.15) имеет распределение Стьюдента с $n-1$ степенями свободы. По таблице распределения Стьюдента по заданному α, $0<\alpha<1$, можно определить ε так, чтобы $P\{|\tau_{n-1}|<\varepsilon\}=1-\alpha$. Иначе это можно переписать в виде

$$P\left\{\frac{1}{n}\sum_{j=1}^{n}\xi_j-\frac{\varepsilon\|(I-\Pi_1)\xi\|}{\sqrt{n(n-1)}}<\mu<\frac{1}{n}\sum_{j=1}^{n}\xi_j+\frac{\varepsilon\|(I-\Pi_1)\xi\|}{\sqrt{n(n-1)}}\right\}=1-\alpha. \tag{8}$$

Соотношение (8) показывает, что истинное значение μ с вероятностью $1-\alpha$ покрывается случайным интервалом

$$\left(\frac{1}{n}\sum_{j=1}^{n}\xi_j-\frac{\varepsilon\|(I-\Pi_1)\xi\|}{\sqrt{n(n-1)}},\ \frac{1}{n}\sum_{j=1}^{n}\xi_j+\frac{\varepsilon\|(I-\Pi_1)\xi\|}{\sqrt{n(n-1)}}\right),$$

называется интервальной оценкой математического ожидания при неизвестной дисперсии.

4°. Интервальная оценка дисперсии при неизвестном математическом ожидании

Как отмечалось в следствии 3 теоремы 14.1

$$\|(I-\Pi_1)(\xi-m)\|^2=\|(I-\Pi_1)\xi\|^2=\sigma^2\chi^2_{n-1}. \tag{9}$$

Используя статистику (9) так же, как в пункте 2° была использована статистика $\|\xi-m\|^2=\sigma^2\chi_n^2$, по заданному α, $0<\alpha<1$, определим ε_1 и ε_2 из требований

$$P\{\chi^2_{n-1}\leqslant\varepsilon_1\}=P\{\chi^2_{n-1}\geqslant\varepsilon_2\}=\alpha/2.$$

Тогда

$$P\left\{\varepsilon_1<\frac{1}{\sigma^2}\|(I-\Pi_1)\xi\|^2<\varepsilon_2\right\}=1-\alpha,$$

или, что то же самое

$$P\left\{\frac{1}{\varepsilon_2}\|(I-\Pi_1)\xi\|^2<\sigma^2<\frac{1}{\varepsilon_1}\|(I-\Pi_1)\xi\|^2\right\}=1-\alpha. \tag{10}$$

Случайный интервал

$$\left(\frac{1}{\varepsilon_2}\|(I-\Pi_1)\xi\|^2,\ \frac{1}{\varepsilon_1}\|(I-\Pi_1)\xi\|^2\right)$$

является интервальной оценкой для дисперсии σ^2 при неизвестном математическом ожидании μ.

В дальнейшем пространство R_n с некоторым фиксированным базисом будем называть выборочным пространством. При этом векторы из R_n задаются наборами своих координат $(\xi_1,\dots,\xi_n)$, образующих выборки.

§ 16. ОБЩАЯ ЗАДАЧА ИНТЕРВАЛЬНОГО ОЦЕНИВАНИЯ

Предположим, что $\xi_1,\dots,\xi_n$ — случайная выборка из распределения $F(x,\theta)$, $x\in R_1$. Параметр θ функции распределения $F(x,\theta)$ считается неизвестным. Для простоты будем считать, что θ меняется на действительной прямой. Пусть функции $\underline{\theta}=\underline{\theta}(x_1,\dots,x_n)$ и $\overline{\theta}=\overline{\theta}(x_1,\dots,x_n)$ таковы, что

$$P\{\underline{\theta}(\xi_1,\dots,\xi_n)<\theta<\overline{\theta}(\xi_1,\dots,\xi_n)\,|\,\theta\}=\gamma,$$

где $P\{A|\theta\}$ — вероятность события $(\xi_1,\dots,\xi_n)\in A\subset R_n$, причем случайные величины $\xi_1,\dots,\xi_n$ контролируются совместной функцией распределения $F(x_1,\theta)\cdot\ldots\cdot F(x_n,\theta)$. В данном случае событие $A\in R_n$ состоит из тех $(\xi_1,\dots,\xi_n)$, для которых $\underline{\theta}(\xi_1,\dots,\xi_n)<\theta<\overline{\theta}(\xi_1,\dots,\xi_n)$. Пара случайных величин $\underline{\theta}$, $\overline{\theta}$ называется интервальной оценкой параметра θ, или $100\gamma\%$-ным доверительным интервалом для θ. Случайные величины $\underline{\theta}$ и $\overline{\theta}$ называются доверительными границами, нижней и верхней соответственно. Как и в случае нормального распределения, $\underline{\theta}$ и $\overline{\theta}$ определяют случайный интервал $(\underline{\theta},\overline{\theta})$, покрывающий истинное значение параметра θ с вероятностью γ.

Рассмотрим один из методов построения интервальных оценок.

Лемма 1. Пусть функция $g(x_1,\dots,x_n,\theta)$ определена для $-\infty<x_i<\infty$, $i=1,\dots,n$, и для любых фиксированных $x_1,\dots,x_n$ непрерывна и монотонна по θ. Пусть, кроме того, $g(\xi_1,\dots,\xi_n,\theta)$ является случайной величиной, функция распределения которой не зависит от θ, если $\xi_1,\dots,\xi_n$ контролируются функцией распределения $F(x_1,\theta)\cdot\ldots\cdot F(x_n,\theta)$. Если (g_1,g_2) — интервал, для которого

$$P\{g_1<g(\xi_1,\dots,\xi_n,\theta)<g_2|\theta\}=\gamma,$$

то

$$P\{\underline{\theta}<\theta<\overline{\theta}|\theta\}=\gamma, \tag{1}$$

где $\underline{\theta}<\overline{\theta}$ — решения относительно θ уравнений $g(\xi_1,\dots,\xi_n,\theta)=g_1,g_2$.

Доказательство. Для определенности будем считать, что $g(\cdot,\theta)$ монотонно не убывает по θ. Поскольку распределение $\zeta=g(\xi_1,\dots,\xi_n,\theta)$ от θ не зависит, то для всякого γ, $0<\gamma<1$, и θ_0, $-\infty<\theta_0<\infty$, можно указать g_1 и g_2,

$g_1 < g_2$, так чтобы

$$P\{g_1 < \zeta < g_2 \mid \theta_0\} = \gamma, \qquad (2)$$

причем g_1 и g_2 можно указать (вообще говоря не единственным способом), если задано лишь γ, а истинное значение θ_0 параметра θ неизвестно. Неравенства

$$g_1 < g(\xi_1, \ldots, \xi_n, \theta_0) < g_2 \qquad (3)$$

для каждой выборки $\xi_1, \ldots, \xi_n$ из распределения $F(x, \theta_0)$ эквивалентны неравенствам

$$\underline{\theta}(\xi_1, \ldots, \xi_n) < \theta_0 < \overline{\theta}(\xi_1, \ldots, \xi_n), \qquad (4)$$

где $\underline{\theta}(\xi_1, \ldots, \xi_n)$ и $\overline{\theta}(\xi_1, \ldots, \xi_n)$ — решения относительно θ уравнений $g(\xi_1, \ldots, \xi_n, \theta) = g_1$ и $g(\xi_1, \ldots, \xi_n, \theta) = g_2$ соответственно. Поэтому события (3) и (4) совпадают. Следовательно,

$$P\{\underline{\theta}(\xi_1, \ldots, \xi_n) < \theta_0 < \overline{\theta}(\xi_1, \ldots, \xi_n) \mid \theta_0\} = \gamma. \quad \blacktriangle$$

Однако всегда ли можно найти функцию $g(\xi_1, \ldots, \xi_n, \theta)$ с указанными в лемме свойствами? Если функция распределения $F(x, \theta)$ непрерывна по x, $-\infty < x < \infty$, для каждого θ, $-\infty < \theta < \infty$, непрерывна и монотонна по θ, $-\infty < \theta < \infty$, для каждого x, $-\infty < x < \infty$, то ответ на поставленный вопрос утвердительный. Для доказательства заметим вначале, что для всякой случайной величины ξ с непрерывной функцией распределения $F(x)$ случайная величина $\eta = F(\xi)$ равномерно распределена на $[0, 1]$. Чтобы не прерывать изложение, этот факт докажем в последнюю очередь.

Рассмотрим функцию

$$g(\xi_1, \ldots, \xi_n, \theta) = F(\xi_1, \theta) \ldots F(\xi_n, \theta).$$

Согласно замечанию ее распределение не зависит от θ, если $F(x, \theta)$ — функция распределения ξ_i, $i = 1, \ldots, n$, и случайные величины $\xi_1, \ldots, \xi_n$ независимы в совокупности. Кроме того, по условию $g(\xi_1, \ldots, \xi_n, \theta)$ монотонна и непрерывна по θ. Следовательно, $g(\cdot)$ удовлетворяет условиям леммы.

Докажем теперь, что случайная величина $\eta = F(\xi)$, где $F(\cdot)$ — непрерывная функция распределения ξ, равномерно распределена на $[0, 1]$. Действительно,

$$F(x) = P\{\xi < x\} = P\{\xi \leqslant x\}, \qquad (5)$$

поскольку функция распределения $F(x)$ непрерывна, и, следовательно, $P\{\xi = x\} = 0$. Событие $\xi \leqslant x$, очевидно, влечет событие $F(\xi) \leqslant F(x)$, поэтому в согласии с (5)

$$F(x) = P\{\xi \leqslant x\} \leqslant P\{F(\xi) \leqslant F(x)\}. \qquad (6)$$

С другой стороны,

$$\{F(\xi) \leqslant F(x)\} = \{\xi \leqslant x\} \cup \{F(\xi) = F(x)\},$$

и

$$P\{F(\xi) \leqslant F(x)\} \leqslant P\{\xi \leqslant x\} + P\{F(\xi) = F(x)\} = P\{\xi \leqslant x\}, \qquad (7)$$

так как функция распределения $F(x)$ непрерывна и, следовательно, $P\{F(\xi) = \text{const}\} = 0$. Из (6) и (7) получаем

$$P\{F(\xi) < F(x)\} = P\{F(\xi) \leqslant F(x)\} = P\{\xi < x\} = F(x),$$

что иначе можно записать следующим образом:

$$P\{\eta < z\} = \begin{cases} 0, & z < 0, \\ z, & 0 \leqslant z \leqslant 1, \\ 1, & z > 1, \end{cases}$$

$$\eta = F(\xi).$$

Равенства (5) означают, что $\eta = F(\xi)$ равномерно распределена на $[0, 1]$.

В заключение заметим, что если $g(\xi_1, \ldots, \xi_n, \theta)$ обладает всеми свойствами, перечисленными в лемме, кроме монотонности по θ, то неравенства (3), будучи разрешенными относительно θ, эквивалентны включению

$$\theta_0 \in A(\xi_1, \ldots, \xi_n),$$

где случайное множество $A(\xi_1, \ldots, \xi_n)$ определяется выборкой и в данном случае может и не быть интервалом. Вполне аналогично рассмотренному случаю равенство (2) эквивалентно

$$P\{\theta_0 \in A(\xi_1, \ldots, \xi_n) \mid \theta_0\} = \gamma.$$

Множество $A(\xi_1, \ldots, \xi_n)$ называется $100\gamma\%$ доверительным множеством для θ, или, иначе, доверительным множеством уровня γ.

Пусть $\xi_1, \ldots, \xi_n$ — выборка из распределения $N(\mu, \sigma^2)$, дисперсия σ^2 известна. Положим

$$g(x_1, \ldots, x_n, \mu) = \left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^n \exp\left\{-\frac{1}{2\sigma^2}\sum_{j=1}^{n}(x_j - \mu)^2\right\}.$$

В данном случае свойство монотонности g по μ не выполнено. Поскольку статистика

$$-2\ln g(\xi_1, \ldots, \xi_n, \mu) - 2n\ln(\sqrt{2\pi}\,\sigma) = \frac{1}{\sigma^2}\sum_{j=1}^{n}(\xi_j - \mu)^2$$

имеет распределение χ_n^2 с n степенями свободы, определим по таблице распределения χ^2 постоянную c из условия

$$P\{\chi_n^2 < \ln c\} = \gamma.$$

Тогда неравенство

$$\frac{1}{\sigma^2}\sum_{j=1}^{n}(\xi_j-\mu)^2<\ln c$$

определяет доверительное множество для μ уровня γ.

§ 17. ТОЧЕЧНЫЕ ОЦЕНКИ

Рассмотрим типичную ситуацию, в которой возникают задачи точечного оценивания. Пусть $\xi=(\xi_1,\ldots,\xi_n)$ — случайная выборка из распределения $F(x,\theta)$, зависящего от параметра $\theta=(\theta_1,\ldots,\theta_k)\in\Theta$ и $\tau(\theta)$ — известная функция, определенная на Θ. Требуется вычислить значение $\tau(\theta)$, но аргумент $\theta\in\Theta$ неизвестен. В этом случае единственная доступная наблюдению информация о $\tau(\theta)$ заключена в выборке ξ (и функции $\tau(\cdot)$), и лучшее, что можно сделать в такой ситуации, это построить так называемую оценку $\tau(\theta)$, основанную на выборке ξ. Иными словами, речь идет о построении функции $t(\cdot)$, определенной на выборочном пространстве R_n, с целью использовать статистику $t(\xi)$ вместо $\tau(\theta)$. Статистика $t(\xi)$ называется **(точечной) оценкой** $\tau(\theta)$, и при этом, естественно, предполагается что распределение $t(\xi)$ в известном смысле концентрируется около значения $\tau(\theta)$, где θ — параметр распределения выборки ξ.

Прежде чем сформулировать условия, определяющие оценку, рассмотрим несколько характерных примеров. Пусть $\xi=(\xi_1,\ldots,\xi_n)$ — выборка из нормального распределения $N(\mu,\sigma^2)$, $\theta=(\mu,\sigma^2)$, $\Theta=\{-\infty<\mu<\infty,\ 0\leqslant\sigma^2\}$. Рассмотрим статистику

$$\widehat{\mu}_n=t(\xi)=\frac{1}{n}\sum_{j=1}^{n}\xi_j. \tag{1}$$

Так как *

$$M_\theta\widehat{\mu}_n=\frac{1}{n}\sum_{j=1}^{n}M_\theta\xi_j=\mu, \tag{2}$$

то, используя $\widehat{\mu}_n$ в качестве оценки математического ожидания $\mu=\tau(\theta)$, мы не будем совершать систематической ошибки в том смысле, что $M_\theta(\widehat{\mu}_n-\mu)=0$. Оценки, обладающие свойством (2), называются **несмещенными.**

* M_θ и D_θ — обозначения математического ожидания и дисперсии, отвечающих распределению $F(\cdot,\theta)$.

Поскольку

$$M_\theta(\widehat{\mu}_n-\mu)^2=D_\theta\widehat{\mu}_n=\frac{1}{n^2}\sum_{j=1}^{n}D_\theta\xi_j=\frac{1}{n}\sigma^2, \tag{3}$$

то согласно лемме § 10 $\widehat{\mu}_n$ при $n\to\infty$ сходится по вероятности к μ. Это означает, что распределение $\widehat{\mu}_n$ при $n\to\infty$ концентрируется около μ, поскольку при $n\to\infty$ вероятность любого уклонения $\widehat{\mu}_n$ от μ стремится к нулю:

$$P\{|\widehat{\mu}_n-\mu|>\varepsilon\,|\,\theta\}<\frac{\sigma^2}{n\sigma^2}\to 0,\quad n\to\infty, \tag{4}$$

для любого $\varepsilon>0$.

Оценки, обладающие свойством (4), т. е. сходящиеся по вероятности к оцениваемому параметру (функции параметра) при объеме n выборки, стремящемся к бесконечности, называются **состоятельными.** Это, безусловно, желательное свойство оценок, но оно не определяет качества оценки при фиксированном n, если качество определить, например, как величину уклонения $\widehat{\mu}_n$ от μ и задать числом $M_\theta(\widehat{\mu}_n-\mu)^2$. Разумеется, чем меньше $M_\theta(\widehat{\mu}_n-\mu)^2$, тем лучше оценка. В данном случае речь идет о несмещенной оценке, и ее качество определяется величиной дисперсии. Поэтому наилучшую оценку μ естественно определить как оценку с минимальной дисперсией среди всех несмещенных оценок. Далее мы покажем, что в этом смысле $\widehat{\mu}_n$ действительно наилучшая.

Какова в данном случае роль несмещенности? Пусть на основании независимых выборок $\xi^1=(\xi_1^1,\ldots,\xi_{n_1}^1)$, $\xi^2=(\xi_1^2,\ldots,\xi_{n_2}^2),\ldots,\xi^k=(\xi_1^k,\ldots,\xi_{n_k}^k)$ из распределения $F(x,\theta)$ построены оценки $t_1(\xi^1),\ t_2(\xi^2),\ldots,t_k(\xi^k)$ значения $\tau(\theta)$, причем

$$M_\theta t_i(\xi^i)=\tau(\theta)+\varepsilon_i,\quad D_\theta t_i(\xi^i)=\sigma_i^2\leqslant\sigma^2,\quad i=1,2,\ldots,k.$$

Тогда для статистики

$$T_k=\frac{1}{k}\sum_{j=1}^{k}t_j(\xi^j)$$

найдем

$$M_\theta T_k=\frac{1}{k}\sum_{j=1}^{k}M_\theta t_j(\xi^j)=\tau(\theta)+\frac{1}{k}\sum_{j=1}^{k}\varepsilon_j$$

и

$$D_\theta T_k=\frac{1}{k^2}\sum_{j=1}^{k}\sigma_j^2\leqslant\sigma^2/k\to 0\quad k\to\infty.$$

Если смещение $\frac{1}{k}\sum_{j=1}^{k}\varepsilon_j$ не стремится к нулю при $k\to\infty$, то T_k при $k\to\infty$ не сходится к истинному значению $\tau(\theta)$. Например, при $\varepsilon_j=\varepsilon$, $j=1, 2, \ldots, k$, T_k сходится по вероятности к $\tau(\theta)+\varepsilon$, что, разумеется, нежелательно. С другой стороны, если $\varepsilon_j=0$, $j=1, 2, \ldots, k$, то можно говорить о сложении информации о $\tau(\theta)$, содержащейся в независимых несмещенных оценках $t_j(\xi^j)$, $j=1, 2, \ldots, k$.

Вместе с тем, очевидно, что если речь идет об оценке $\overset{\circ}{t}$ с минимальным уклонением

$$M_\theta(\overset{\circ}{t}-\tau(\theta))^2=\min_t M_\theta(t-\tau(\theta))^2, \tag{5}$$

то при прочих равных условиях дополнительное требование несмещенности может лишь увеличить уклонение (5), так как условие $M_\theta t=\tau(\theta)$ сужает класс оценок, по которым ищется минимум в (5). Более того, класс несмещенных оценок может оказаться даже пустым.

Действительно, пусть, например, ξ — число успехов в последовательности n испытаний Бернулли с неизвестной вероятностью успеха $p=\theta$, $0<\theta\leqslant 1$, и требуется несмещенно оценить $\tau(\theta)=\theta^{-1}$. Для всякой такой оценки $t(\xi)$ должно выполняться равенство

$$M_\theta t(\xi)=\sum_{i=0}^{n}C_n^i\,\theta^i(1-\theta)^{n-i}\,t(i)=\theta^{-1}$$

для всех $\theta\in(0, 1]$. Но, это, очевидно, невозможно; так, для любой функции $t(\cdot)$ при $\theta\to 0: M_\theta t(\xi)\to C_n^0 t(0)$, а $\theta^{-1}\to\infty$.

Рассмотрим статистику

$$\widehat{\sigma}_n^2=\frac{1}{n-1}\sum_{j=1}^{n}(\xi_j-\widehat{\mu}_n)^2=\frac{\|(I-\Pi_1)\xi\|^2}{n-1}, \tag{6}$$

где $\xi=(\xi_1, \ldots, \xi_n)$ — выборка из $N(\mu, \sigma^2)$, Π_1 — оператор ортогонального проектирования в R_n, введенный в § 14. Согласно следствию 1 теоремы 1 § 14 статистика (6) имеет распределение $(\sigma^2/n-1)\chi^2_{n-1}$. Поэтому согласно формуле (53) § 9

$$M_\theta\widehat{\sigma}_n^2=\frac{\sigma^2}{n-1}M\chi^2_{n-1}=\sigma^2,$$

$$D_\theta\,\widehat{\sigma}_n^2=\frac{\sigma^4}{(n-1)^2}\cdot D_\theta\chi^2_{n-1}=\frac{2\sigma^4}{n-1}\to 0,\quad n\to\infty.$$

Следовательно, статистика $\widehat{\sigma}_n^2$ является несмещенной состоя-

тельной оценкой $\sigma^2=\tau(\theta)$, $\theta=(\mu, \sigma^2)$. Однако

$$M_\theta\left(\frac{1}{n}\|(I-\Pi_1)\xi\|^2-\sigma^2\right)^2=\frac{\sigma^4}{n^2}[2(n-1)+(n-1)^2]-$$

$$-2\frac{\sigma^4}{n}(n-1)+\sigma^4=\left(\frac{2}{n}-\frac{1}{n^2}\right)\sigma^4<D_\theta\widehat{\theta}_n^2,\quad n=1,2,\ldots.$$

Поэтому для каждого $n=1, 2, \ldots$ статистика

$$\widetilde{\sigma}_n^2=\frac{1}{n}\|(I-\Pi_1)\xi\|^2$$

меньше уклоняется от σ^2, чем $\widehat{\sigma}_n^2$, хотя и имеет смещение, ибо $M_\theta\widetilde{\sigma}_n^2=\sigma^2-\sigma^2/n$. Так как $M_\theta(\widetilde{\sigma}_n^2-\sigma^2)=\sigma^2/n\to 0$, $n\to\infty$,

$$D_\theta(\widetilde{\sigma}_n^2-\sigma^2)=M_\theta(\widetilde{\sigma}_n^2-\sigma^2)^2-\frac{\sigma^4}{n^2}=\frac{2n-2}{n^2}\sigma^4\to 0,\quad n\to\infty,$$

то оценка $\widehat{\sigma}_n^2$ при $n\to\infty$ сходится по вероятности к σ^2, т. е. состоятельна. В данном случае смещенная оценка $\widetilde{\sigma}_n^2$ дисперсии σ^2 оказывается предпочтительнее несмещенной $\widehat{\sigma}_n^2$. Впрочем, заметим, что статистика

$$\breve{\sigma}_n^2=\frac{\|(I-\Pi_1)\xi\|^2}{n+1}$$

обладает еще меньшим уклонением от σ^2, более того минимальным среди всех статистик вида $\mathrm{const}\|(I-\Pi_1)\xi\|^2$,

$$M_\theta\left(\frac{1}{n+1}\|(I-\Pi_1)\xi\|^2-\sigma^2\right)^2=\frac{2\sigma^4}{n+1}.$$

Статистики $\widehat{\mu}_n$ и $\widehat{\sigma}_n^2$ как оценки для математического ожидания и дисперсии сохраняют ряд свойств и без предложения о нормальности выборки. Действительно, если $\xi_1, \ldots, \xi_n$ — выборка из произвольного распределения, причем

$$M\xi_j=\mu,\quad D\xi_j=\sigma^2,\quad j=1, 2, \ldots, n,$$

то согласно (2), (3)

$$M\widehat{\mu}_n=\mu,\quad D\widehat{\mu}_n=\sigma^2/n,$$

и тем самым $\widehat{\mu}_n$ — несмещенная и состоятельная оценка μ. Так как согласно определению оператора проектирования Π_1

$$\|(I-\Pi_1)\xi\|^2=\|(\xi-m)-\Pi_1(\xi-m)\|^2=$$

$$=\|\xi-m\|^2-\|\Pi_1(\xi-m)\|^2=\sum_{j=1}^{n}(\xi_j-\mu)^2-n\left(\frac{1}{n}\Sigma(\xi_j-\mu)\right)^2,$$

где $m=(\mu, \ldots, \mu)$, то

$$M\|(I-\Pi_1)\xi\|^2=n\sigma^2-\sigma^2.$$

Следовательно, и без предложения о нормальности статистика $\widehat{\sigma}_n^2$ является несмещенной оценкой для σ^2. Однако никакого заключения о состоятельности $\widehat{\sigma}_n^2$ без дополнительных предположений о распределении на сей раз сделать не удастся. Что касается оценок $\widetilde{\sigma}_n^2$ и $\breve{\sigma}_n^2$, то в данном случае они утрачивают преимущества перед $\widehat{\sigma}_n^2$.

Итак, для дальнейшего обсуждения мы выделяем следующие свойства оценок:

1) *несмещенность*: $M_\theta t(\xi) = \tau(\theta)$;

2) *минимальность уклонения*: $M_\theta (t(\xi) - \tau(\theta))^2 \sim \min$ или, в частности, минимальность дисперсии, если речь идет о несмещенных оценках;

3) *Состоятельность*: $t_n(\xi) \xrightarrow{P} \tau(\theta)$, $n \to \infty$.

Каждое из этих условий является ограничением на $t(\xi)$, и далеко не всегда можно гарантировать существование оценки с такими свойствами.

1°. Несмещенные оценки минимальной дисперсии

Как было отмечено, не всегда можно удовлетворить требованию несмещенности. Однако если существует несмещенная оценка минимальной дисперсии (Н. О. М. Д.), то она единственна.

Лемма 1. Пусть t_1 и t_2 — Н. О. М. Д. Тогда $t_1 = t_2$ с вероятностью единица.

Доказательство. Обозначим

$$M_\theta t_1 = M_\theta t_2 = \tau(\theta).$$

Тогда и для оценки $t_3 = 1/2(t_1 + t_2) : M_\theta t_3 = \tau(\theta)$. Кроме того, в силу неравенства Коши—Буняковского

$$\begin{aligned} D_\theta t_3 &= 1/4\{D_\theta t_1 + D_\theta t_2 + 2\operatorname{cov} t_1 t_2\} \leqslant \\ &\leqslant 1/4\{D_\theta t_1 + D_\theta t_2 + 2(D_\theta t_1 D_\theta t_2)^{1/2}\} = \\ &= 1/4\{(D_\theta t_1)^{1/2} + (D_\theta t_2)^{1/2}\}^2. \end{aligned} \tag{7}$$

Равенство в (7) выполняется тогда и только тогда, когда с вероятностью единица

$$t_1 - \tau(\theta) = k(\theta)(t_2 - \tau(\theta)),\ k(\theta) \geqslant 0.$$

Обозначим $D_\theta t_1 = D_\theta t_2 = \delta$. Из (7) следует, что $D_\theta t_3 \leqslant 1/4(\delta + \delta + 2\delta) = \delta$. Поскольку t_1 и t_2 — оценки с минимальной дисперсией, должно быть $D_\theta t_3 \geqslant \delta$, следовательно, $D_\theta t_3 = \delta$. Поэтому в (7) выполнено равенство и, следовательно, с вероятностью единица

$$t_1 - \tau(\theta) = k(\theta)(t_2 - \tau(\theta)).$$

А так как

$$M_\theta (t_1 - \tau(\theta))(t_2 - \tau(\theta)) = k(\theta) \cdot \delta = \delta,$$

то $k(\theta) = 1$. ▲

Далее статистика, имеющая конечный момент второго (а следовательно, и первого) порядка, называется **гильбертовой.**

Теорема 1. Пусть $\xi = (\xi_1, \ldots, \xi_n)$ — выборка из распределения $F(x, \theta)$, $\theta \in \Theta$. Для того чтобы гильбертова статистика $t(\xi)$ была Н. О. М. Д. для $\tau(\theta)$, необходимо и достаточно, чтобы для всякой гильбертовой статистики $\eta = s(\xi)$, такой, что $M_\theta \eta = 0$,

$$M_\theta t(\xi)\eta = 0. \tag{8}$$

Доказательство. Пусть $t(\xi)$ — гильбертова несмещенная оценка для $\tau(\theta)$. Если гильбертова статистика $\eta = s(\xi)$ такова, что $M_\theta \eta = 0$, то $t(\xi) + \lambda\eta$ также несмещенная гильбертова оценка $\tau(\theta)$ для всех λ. При этом

$$\begin{aligned} \varphi_\lambda = M_\theta (t(\xi) - \tau(\theta) + \lambda\eta)^2 &= M_\theta (t(\xi) - \tau(\theta))^2 + \\ &+ 2\lambda M_\theta (t(\xi) - \tau(\theta))\eta + \lambda^2 M_\theta \eta^2, \end{aligned}$$

и, дифференцируя по λ, найдем

$$\min_\lambda \varphi_\lambda = M_\theta (t(\xi) - \tau(\theta))^2 - \frac{[M_\theta (t(\xi) - \tau(\theta))\eta]^2}{M_\theta \eta^2}$$

при

$$\lambda = -\frac{M_\theta (t(\xi) - \tau(\theta))\eta}{M_\theta \eta^2}.$$

Пусть $t(\xi)$ — Н. О. М. Д. для $\tau(\theta)$. Тогда для всякой гильбертовой статистики $\eta = s(\xi)$, $M_\theta \eta = 0$:

$$\begin{aligned} M_\theta (t(\xi) - \tau(\theta))^2 &\leqslant \min_\lambda M_\theta (t(\xi) - \tau(\theta) + \lambda\eta)^2 = \\ &= M_\theta (t(\xi) - \tau(\theta))^2 - \frac{[M_\theta (t(\xi) - \tau(\theta))\eta]^2}{M_\theta \eta^2}. \end{aligned}$$

Отсюда следует (8):

$$0 = M_\theta (t(\xi) - \tau(\theta))\eta = M_\theta t(\xi)\eta.$$

Наоборот, если условие (8) выполнено для всякой гильбертовой статистики $\eta = s(\xi)$, $M_\theta \eta = 0$, то

$$\begin{aligned} M_\theta (t(\xi) + \lambda\eta - \tau(\theta))^2 &\geqslant \min_\lambda M_\theta (t(\xi) + \lambda\eta - \tau(\theta))^2 = \\ &= M_\theta (t(\xi) - \tau(\theta))^2. \end{aligned}$$

Следовательно, $t(\xi)$ — Н. О. М. Д. для $\tau(\theta)$, так как в виде $t(\xi) + \lambda\eta$ можно представить любую гильбертову несмещенную оценку $\tau(\theta)$. ▲

Замечание. Множество гильбертовых случайных величин, т. е. случайных величин с конечными моментами второго порядка, образует линейное пространство. Действительно, если $M\zeta^2<\infty$, $M\eta^2<\infty$, то и $M(a\zeta+b\eta)^2=a^2M\zeta^2+2abM\zeta\eta+b^2M\eta^2\leqslant a^2M\zeta^2+2|ab|(M\zeta^2M\eta^2)^{1/2}+b^2M\eta^2<\infty$. Это линейное пространство превращается в гильбертово пространство, если в нем задать скалярное произведение $(\eta, \zeta)=M\eta\zeta$, а элементами считать классы равных с вероятностью единица случайных величин. Норма случайной величины η определяется как $(\eta, \eta)^{1/2}$ и отождествляется с нормой класса случайных величин, совпадающих с η с вероятностью единица.

Возвращаясь к доказанной теореме, обозначим через H_θ гильбертово пространство гильбертовых статистик $\zeta=t(\xi)$, $M_\theta\zeta^2<\infty$. Подмножество H^0_θ статистик из H_θ, для которых $M_\theta\eta=0$, является линейным подпространством H_θ. Рассмотрим гиперплоскость $H^0_\theta+t(\xi)$, где $t(\xi)\in H_\theta$ и $M_\theta t(\xi)=\tau(\theta)$, т. е. множество всех статистик вида $t(\xi)+\eta$, $\eta\in H^0_\theta$. Н. О. М. Д. для $\tau(\theta)$ есть элемент $t(\xi)+H^0_\theta$ с минимальной нормой, т. е. ортогональная проекция $t^0(\xi)$ нуля на $t(\xi)+H^0_\theta$, определяемая условием: $(t^0(\xi), \eta)=0$ для любой статистики $\eta\in H^0_\theta$. Таков геометрический смысл теоремы 1.

2⁰. Эффективные оценки

При некоторых предположениях о функции распределения $F(x, \theta)$ может быть получена удобная априорная оценка для качества оценивания, известная как неравенство Рао—Крамера. Пусть $\xi=(\xi_1, ..., \xi_n)$ — случайная выборка объема n из распределения с плотностью вероятности $f(x, \theta)$, $x\in R_1$, $\theta\in\Theta\subset R_k$. Поскольку ξ_j, $j=1, ..., n$, независимы в совокупности, плотность совместного распределения выборки $\xi_1, ..., \xi_n$ имеет вид

$$L(x, \theta)=f(x_1, \theta)\dots f(x_n, \theta).$$

В статистических задачах $L(x, \theta)$ рассматривается как функция двух аргументов $x\in R_n$, $\theta\in\Theta$ и называется функцией правдоподобия.

Если выборку ξ_i, $i=1, ..., n$, образуют дискретные случайные величины, то функция правдоподобия определяется как произведение вероятностей

$$L(x, \theta)=P(x_1, \theta)\dots P(x_n, \theta),\ x=(x_1, ..., x_n).$$

Теорема 2 (Рао—Крамер). Пусть $\Theta=R_1$, 1) статистика $t(\xi)$, $\xi=(\xi_1, ..., \xi_n)$, является несмещенной оценкой $\tau(\theta)$: $M_\theta t(\xi)=\tau(\theta)$, $\theta\in\Theta$; 2) функции $L(x, \theta)$ и $\tau(\theta)$ дифференцируемы по $\theta\in\Theta$; 3) множество тех $x\in R_n$, для которых $L(x, \theta)>0$, не зависит от $\theta\in\Theta$ и

$$\frac{d}{d\theta}\int L(x, \theta)\,dx=\int\frac{\partial L(x, \theta)}{\partial\theta}\,dx,$$

$$\frac{d}{d\theta}\int t(x)L(x, \theta)\,dx=\int t(x)\frac{\partial L(x, \theta)}{\partial\theta}\,dx.$$

Тогда

$$D_\theta t(\xi)\geqslant\frac{[\tau'(\theta)]^2}{M_\theta\left(\frac{\partial\ln L(\xi, \theta)}{\partial\theta}\right)^2}, \tag{9}$$

причем знак равенства имеет место тогда и только тогда, когда

$$\frac{\partial\ln L(\xi, \theta)}{\partial\theta}=a(\theta)(t(\xi)-\tau(\theta)) \tag{10}$$

с вероятностью единица для некоторой функции $a(\theta)$, $\theta\in\Theta$.

Доказательство. Дифференцируя равенства

$$\int L(x, \theta)\,dx=1,\quad\int t(x)L(x, \theta)\,dx=\tau(\theta)$$

по θ, найдем

$$\int\frac{\partial\ln L(x, \theta)}{\partial\theta}L(x, \theta)=0,$$

$$\int t(x)\frac{\partial\ln L(x, \theta)}{\partial\theta}L(x, \theta)\,dx=\tau'(\theta),\quad\theta\in R_1.$$

Отсюда следует, что

$$\tau'(\theta)=\int(t(x)-\tau(\theta))\frac{\partial\ln L(x, \theta)}{\partial\theta}L(x, \theta)\,dx \tag{11}$$

и неравенство (9) получается из (11) как неравенство Коши—Буняковского

$$[\tau'(\theta)]^2\leqslant\int(t(x)-\tau(\theta))^2L(x, \theta)\,dx\int\left(\frac{\partial\ln L(x, \theta)}{\partial\theta}\right)^2L(x, \theta)\,dx. \tag{12}$$

Равенство в (12), а следовательно, и в (9), имеет место лишь тогда, когда для некоторого множителя пропорциональности $a(\theta)$

$$\sqrt{L(x, \theta)}\frac{\partial\ln L(x, \theta)}{\partial\theta}=\sqrt{L(x, \theta)}\,a(\theta)(t(x)-\tau(\theta)).$$

Это условие эквивалентно (10), верному с вероятностью единица, поскольку множество тех ξ, для которых $L(\xi, \theta)=0$, имеет вероятность, равную нулю. ▲

Определение. Несмещенная оценка $t(\xi)$, $M_\theta t(\xi)=\tau(\theta)$, называется **эффективной**, если

$$D_\theta t(\xi)=\frac{[\tau'(\theta)]^2}{M_\theta\left(\frac{\partial\ln L(\xi, \theta)}{\partial\theta}\right)^2}.$$

В условиях теоремы 2 для того, чтобы оценка $t(\xi)$ была эффективной, необходимо и достаточно, чтобы она удовлетворяла условию (10). Так как согласно (10)

$$M_\theta\left(\frac{\partial \ln L(\xi, \theta)}{\partial \theta}\right) = a^2(\theta) D_\theta t(\xi),$$

то для эффективной оценки

$$D_\theta t(\xi) = \left|\frac{\tau'(\theta)}{a(\theta)}\right|. \qquad (13)$$

Эффективная оценка является, очевидно, Н. О. М. Д., но обратное неверно. Заметим, что согласно (10) эффективная оценка может существовать лишь для вполне определенной функции $\tau(\theta)$. Если такая оценка существует для $\tau(\theta)$, то она не существует ни для какой другой функции, отличной от $a\tau(\theta)+b$, где a и b — некоторые числа.

Рассмотрим примеры эффективных оценок. Пусть $\xi = (\xi_1, ..., \xi_n)$ — выборка из распределения $N(\theta, \sigma^2)$ с известной дисперсией σ^2. В таком случае функция правдоподобия имеет вид

$$L(x, \theta) = \left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^n \exp\left\{-\frac{1}{2\sigma^2}\sum_{j=1}^{n}(x_j-\theta)^2\right\}. \qquad (14)$$

Следовательно,

$$\frac{\partial \ln L(x, \theta)}{\partial \theta} = \frac{1}{\sigma^2}\sum_{j=1}^{n}(x_j-\theta) = \frac{n}{\sigma^2}\left[\frac{1}{n}\sum_{j=1}^{n}x_j-\theta\right].$$

Поэтому согласно (10) статистика $t(\xi) = \frac{1}{n}\sum_{j=1}^{n}\xi_j$ является эффективной оценкой своего математического ожидания θ. Ее дисперсия в согласии с (13) равна σ^2/n.

Если $\xi = (\xi_1, ..., \xi_n)$ — выборка из $N(\mu, \theta^2)$ с известным математическим ожиданием μ, то функция правдоподобия будет иметь вид

$$L(x, \theta) = \left(\frac{1}{\sqrt{2\pi}\,\theta}\right)^n \exp\left\{-\frac{1}{2\theta^2}\sum_{j=1}^{n}(x_j-\mu)^2\right\},$$

и

$$\frac{\partial \ln L}{\partial \theta} = -\frac{n}{\theta} + \frac{1}{\theta^3}\sum_{j=1}^{n}(x_j-\mu)^2 = \frac{n}{\theta^3}\left\{\frac{1}{n}\sum_{j=1}^{n}(x_j-\mu)^2-\theta^2\right\}.$$

Поэтому статистика $t(\xi) = \frac{1}{n}\sum_{j=1}^{n}(\xi_j-\mu)^2$ является эффективной оценкой дисперсии θ^2. Ее дисперсия равна $|\tau'(\theta)/a(\theta)| = 2\theta^4/n$.

Теорема 2 остается верной, если плотность $f(x, \theta)$ заменить на вероятность, интегрирование — на суммирование. Рассмотрим пример эффективного оценивания параметра дискретного распределения. Пусть $\xi = (\xi_1, ..., \xi_n)$ — выборка из распределения Пуассона с неизвестным параметром θ. В таком случае логарифм функции правдоподобия будет иметь вид

$$\ln L(x, \theta) = \ln \prod_{j=1}^{n}\frac{\theta^{x_j}}{x_j!}e^{-\theta} = \sum_{j=1}^{n}(-\theta + x_j\ln\theta - \ln(x_j!)),$$

и

$$\frac{\partial \ln L(x, \theta)}{\partial \theta} = \sum_{j=1}^{n}\left(-1+\frac{x_j}{\theta}\right) = \frac{n}{\theta}\left\{\frac{1}{n}\sum_{j=1}^{n}x_j-\theta\right\}.$$

Поэтому

$$t(\xi) = \frac{1}{n}\sum_{j=1}^{n}\xi_j$$

эффективная оценка θ с дисперсией θ/n.

Рассмотрим так называемое экспоненциальное семейство распределений, для которого определяется семейство плотностей (или вероятностей) следующим образом:

$$f(x, \theta) = \exp\{a(\theta)b(x)+c(\theta)+d(x)\}, \quad x\in R_1, \ \theta\in R_1.$$

Соответственно функцию правдоподобия запишем как

$$L(x, \theta) = \exp\left\{a(\theta)\sum_{j=1}^{n}b(x_j)+nc(\theta)+\sum_{j=1}^{n}d(x_j)\right\},$$

где $x=(x_1, ..., x_n)$. Если $a(\theta)$ и $c(\theta)$ — дифференцируемые функции $\theta\in R_1$, то

$$\frac{\partial \ln L}{\partial \theta} = a'(\theta)n\left\{\frac{1}{n}\sum_{j=1}^{n}b(x_j)+\frac{c'(\theta)}{a'(\theta)}\right\}.$$

Если для $L(x, \theta)$, $\tau(\theta) = -c'(\theta)/a'(\theta)$ и $t(\xi) = \frac{1}{n}\sum_{j=1}^{n}b(\xi_j)$ выполнены условия теоремы 2, то $\frac{1}{n}\sum_{j=1}^{n}b(\xi_j)$ — эффективная оценка $\tau(\theta) = -c'(\theta)/a'(\theta)$ с дисперсией, равной $|\tau'(\theta)/(na'(\theta))|$.

Экспоненциальному семейству принадлежат многие важные для практики распределения. Примеры некоторых из них рассмотрены выше.

3°. Достаточные статистики

Изучая точечные оценки, следует обратить внимание на тот факт, что при их построении чаще всего не используется полная информация, содержащаяся в выборке. Так, для оценивания математического ожидания по выборке из нормального распределения достаточно знать лишь значение суммы $\xi_1+\ldots+\xi_n$, а не каждой случайной величины ξ_j, $j=1,\ldots,n$, в отдельности. Если μ известно, и по выборке $\xi_1,\ldots,\xi_n$ из $N(\mu,\sigma^2)$ надлежит оценить дисперсию σ^2, то для этого достаточно знать лишь $\sum_{j=1}^{n}(\xi_j-\mu)^2$ и т. п.

Оказывается, что во многих случаях для оценивания параметра θ распределения $F(x,\theta)$ по выборке $\xi_1,\ldots,\xi_n$ можно указать статистику

$$T(\xi)=(T_1(\xi),\ldots,T_m(\xi)),\quad m<n, \tag{15}$$

в известном смысле содержащую всю информацию о параметре θ.

Определение. Статистика $T(\xi)$ (15) называется **достаточной** для семейства плотностей (вероятностей) $L(x,\theta)$, $\theta\in\Theta$, $x\in R_n$, если условное распределение случайного вектора $\xi=(\xi_1,\ldots,\xi_n)$ при условии $T(\xi)=t$ не зависит от θ.

На практике пользоваться этим определением для проверки факта достаточности не очень удобно. Более удобно следующее необходимое и достаточное условие достаточности статистики $T(\xi)$.

Теорема 3 (о факторизации). $T(\xi)$ — достаточная статистика тогда и только тогда, когда функция правдоподобия может быть представлена в виде

$$L(x,\theta)=g(T(x),\theta)h(x),\quad x\in R_n,\ \theta\in\Theta. \tag{16}$$

Доказательство. Ограничимся случаем дискретного распределения. Пусть

$$p_\theta(x,t)=P\{\xi=x,\ T(\xi)=t\mid\theta\}=\begin{cases} L(x,\theta), & \text{если } T(x)=t,\\ 0, & \text{если } T(x)\neq t,\end{cases}$$

совместное распределение ξ и $T(\xi)$. Тогда распределение $T(\xi)$ определится равенством

$$p_\theta(t)=P\{T(\xi)=t\mid\theta\}=\sum_x p_\theta(x,t)=\sum_{x:T(x)=t}L(x,\theta)$$

и соответственно для условного распределения $P\{\xi=x\mid T(\xi)=t,\ \theta\}=p_\theta(x\mid t)$ найдем

$$p_\theta(x\mid t)=p_\theta(x,t)/p_\theta(t)=$$
$$=L(x,\theta)/\sum_{x:T(x)=t}L(x,\theta)=h(x)/\sum_{x:T(x)=t}h(x).$$

Следовательно, статистика $T(\xi)$ — достаточная, поскольку $p_\theta(x\mid t)$ от θ не зависит. Наоборот, пусть $p_\theta(x\mid t)=P\{\xi=x\mid T(\xi)=t,\ \theta\}=f(x,t)$ не зависит от θ. Тогда $p_\theta(x,t)=p_\theta(x\mid t)p_\theta(t)$ и для $t=T(x)$: $p_\theta(x,t)=L(x,\theta)=f(x,T(x))\times p_\theta(T(x))$, что совпадает с выражением (16), если $p_\theta(T(x))=g(T(x),\theta)$. ▲

В качестве иллюстрации заметим, что в случае нормального $N(\theta,\sigma^2)$ функция правдоподобия выборки $\xi_1,\ldots,\xi_n$ имеет вид (14), или, иначе,

$$L(x,\theta)=g(T(x),\theta)h(x),$$

где

$$g(T(x),\theta)=\exp\left\{\frac{\theta}{\sigma^2}\sum_{j=1}^{n}x_j-\frac{n}{2\sigma^2}\theta^2\right\}$$

и

$$h(x)=\left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^n\exp\left\{-\frac{1}{2\sigma^2}\sum_{j=1}^{n}x_j^2\right\}.$$

Следовательно, $T(x)=\sum_{j=1}^{n}x_j$ — достаточная статистика.

На самом деле всякая эффективная оценка при условии (10) является достаточной статистикой, так как

$$\ln L(x,\theta)=\int a(\theta)(t(x)-\tau(\theta))\,d\theta+l(x)$$

и для $L(x,\theta)$ выполнено условие факторизации (16), где

$$g(T(x),\theta)=\exp\left\{\int a(\theta)(T(x)-\tau(\theta))\,d\theta\right.$$

и

$$h(x)=\exp l(x).$$

Более того, оказывается, что если Н.О.М.Д. существует, то ее можно найти как функцию достаточной статистики. Этот факт является следствием следующей теоремы.

Теорема 4. Пусть $T(x)$ — достаточная статистика для семейства $L(x,\theta)$, $\theta\in\Theta$. Если $t(\xi)$ — гильбертова несмещенная оценка $\tau(\theta)$, то $f(T(\xi))$ также гильбертова несмещенная оценка $\tau(\theta)$, причем

$$M_\theta(t(\xi)-\tau(\theta))^2\geqslant M_\theta(f(T(\xi))-\tau(\theta))^2; \tag{17}$$

здесь $f(s)=M_\theta(t(\xi)\mid T(\xi)=s)$ — условное математическое ожидание $t(\xi)$ при условии $T(\xi)=s$.

Доказательство. Отметим вначале, что $f(s)$ не зависит от θ (как математическое ожидание по условному распределению, не зависящему от θ). Поскольку согласно свойствам условного математического ожидания, рассмотренным в § 9,

$$M_\theta t(\xi) = M_T(M(t(\xi)|T) = M_\theta f(T), \tag{18}$$

где M_T — математическое ожидание по распределению статистики $T(\xi)$ (зависящему от θ). Равенство (18) означает, что $f(T(\xi))$ — несмещенная оценка $\tau(\theta)$. Далее,

$$M_\theta f(T)(t - f(T)) = M_T\{f(T)M(t - f(T)|T)\} = 0, \tag{19}$$

так как

$$M_\theta(f(T)|T) = f(T), \quad M_\theta(t|T) = f(T).$$

Отсюда следует

$$M_\theta(t - \tau(\theta))^2 = M_\theta(t - f(T) + f(T) - \tau(\theta))^2 =$$
$$= M_\theta(t - f(T))^2 + M_\theta(f(T) - \tau(\theta))^2 \geqslant M_\theta(f(T) - \tau(\theta))^2,$$

поскольку для математического ожидания произведения согласно (19), (18)

$$M_\theta(t - f(T))(f(T) - \tau(\theta)) = -M_\theta(t - f(T))\tau(\theta) = 0.$$

Если $t(\xi)$ — Н.О.М.Д. для $\tau(\theta)$ и $T(\xi)$ — достаточная статистика, то $f(T) = M_\theta(t(\xi)|T)$ согласно (19) также несмещенная оценка для $\tau(\theta)$, а согласно (17) ее дисперсия не превосходит дисперсии $t(\xi)$. Следовательно, $f(T)$ также Н.О.М.Д. для $\tau(\theta)$. ▲

Определение. Достаточная статистика T называется **полной**, если всякая функция от нее с нулевым для всех $\theta \in \Theta$ математическим ожиданием равна нулю с вероятностью (для любого $\theta \in \Theta$) единица.

Покажем, что всякая функция $f(T)$ полной достаточной статистики является Н.О.М.Д. своего математического ожидания $\tau(\theta) = M_\theta f(T)$ для всех $\theta \in \Theta$. Действительно, как было отмечено, Н.О.М.Д. $\tau(\theta)$ следует искать в классе функций от T, а поскольку статистика T полная, то не существует двух различных функций от T, несмещенно оценивающих $\tau(\theta)$.

Если существует полная достаточная статистика T, то для получения Н.О.М.Д. $\tau(\theta)$ можно вначале построить любую несмещенную оценку $t(\xi)$, а затем взять ее условное математическое ожидание $f(s) = M_\theta(t(\xi)|T(\xi) = s)$. Тогда функция $f(s)$ не зависит от θ и $f(T(\xi))$ — Н.О.М.Д. для $\tau(\theta)$.

В качестве иллюстрации рассмотрим схему испытаний Бернулли. Пусть ξ — число успехов в серии из n испытаний, θ — вероятность успеха в отдельном испытании, $0 < \theta < 1$. Поскольку $L(x, \theta) = C_n^x \theta^x (1-\theta)^{n-x}$, $x = 0, 1, \ldots, n$, то $T(\xi) = \xi$ — достаточная статистика. Пусть $f(\xi)$ — любая функция, такая, что

$$M_\theta f(\xi) = \sum_{k=0}^{n} f(k) C_n^k \theta^k (1-\theta)^{n-k} = 0, \quad \theta \in (0, 1).$$

То же самое можно записать в виде

$$\sum_{k=0}^{n} f(k) C_n^k z^k = 0, \quad 0 < z < \infty.$$

Если полином степени n равен нулю для любого $z \in (0, \infty)$, то его коэффициенты — нули: $f(k) = 0$, $k = 0, \ldots, n$. Следовательно, ξ — полная достаточная статистика, и тем самым любая функция $f(\xi)$ является Н.О.М.Д. для $M_\theta f(\xi)$. Поскольку

$$M_\theta\left(\frac{\xi}{n}\right) = \theta, \quad M_\theta\left(\frac{\xi(\xi-1)}{n(n-1)}\right) = \theta^2, \quad M_\theta \frac{\xi(n-\xi)}{n(n-1)} = \theta(1-\theta),$$

то ξ/n, $\xi(\xi-1)/n$ и $\xi(n-\xi)/n(n-1)$ — Н.О.М.Д. для θ, θ^2 и $\theta(1-\theta)$ соответственно.

4⁰. Оценки максимального правдоподобия

Пусть $L(x, \theta)$, $x \in R_n$, $\theta \in \Theta$, — функция правдоподобия. **Оценкой максимального правдоподобия** называется статистика $\widehat{\theta} = \widehat{\theta}(\xi)$, удовлетворяющая условию: $L(\xi, \widehat{\theta}) \geqslant L(\xi, \theta)$ для всех $\theta \in \Theta$, или, что то же самое,

$$L(\xi, \widehat{\theta}) = \max_{\theta \in \Theta} L(\xi, \theta).$$

Если $L(x, \theta)$ не достигает максимума по $\theta \in \Theta$, то оценка максимального правдоподобия не существует. Название «оценка максимального правдоподобия» мы сохраним и за функцией $\widehat{\theta}(x)$ на выборочном пространстве R_n.

Пусть Θ — подмножество k-мерного евклидова пространства R_k, функция правдоподобия $L(x, \theta)$ дифференцируема по $\theta \in \Theta$ и достигает максимума по θ во внутренней точке Θ для каждого $x \in R_n$. В таком случае оценка максимального правдоподобия удовлетворяет уравнениям

$$\partial \ln L(x, \theta)/\partial \theta_j |_{\theta = \widehat{\theta}(x)} = 0, \; j = 1, \ldots, k. \tag{20}$$

Уравнения (20) являются необходимыми условиями максимума и называются уравнениями максимального правдоподобия.

Если $x = (x_1, \ldots, x_n)$ — наблюдаемое значение выборки $\xi = (\xi_1, \ldots, \xi_n)$ из дискретного распределения с параметром θ, то $L(x, \theta) = P\{\xi_1 = x_1, \ldots, \xi_n = x_n | \theta\}$ — вероятность того, что $\xi_i = x_i$, $i = 1, \ldots, n$. За оценку максимального правдоподобия $\widehat{\theta}(x)$ принимается то значение параметра θ, при котором вероятность получить наблюдаемое значение x выборки ξ максимальна. Это замечание поясняет смысл принципа максимального правдоподобия: в качестве значения неизвестного параметра предлагается принимать то, при котором вероятность наблюдаемой реализации выборки максимальна.

Принцип максимального правдоподобия не основывается на каких-либо соображениях оптимальности. Вместе с тем, если, например, существует эффективная оценка числового параметра, то согласно (10) она является оценкой максимального правдоподобия, так как $\partial \ln L(x, \theta)/\partial\theta = a(\theta)(t(x)-\theta)$. Более того, оценка максимального правдоподобия является функцией достаточной статистики, поскольку из условия $L(x, \theta)=g(T(x), \theta)h(x)$ следует, что $\widehat{\theta}=\varphi(T)$. Но действительно замечательные свойства оценок максимального правдоподобия могут быть доказаны лишь для достаточно больших объемов выборки.

Пусть $\xi=(\xi_1, \dots, \xi_n)$ — выборка из распределения $N(\mu, \sigma^2)$, $\theta=(\mu, \sigma^2)$. Поскольку

$$L(x, \theta) = \left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^n \exp\left\{-\frac{1}{2\sigma^2}\sum_{j=1}^{n}(x_j-\mu)^2\right\},$$

то уравнения (20) максимального правдоподобия принимают вид

$$\frac{\partial \ln L(x, \theta)}{\partial\mu} = \frac{1}{\sigma^2}\sum_{j=1}^{n}(x_j-\mu)=0,$$

$$\frac{\partial \ln L(x, \theta)}{\partial\sigma^2} = -\frac{n}{2\sigma^2}+\frac{1}{2\sigma^4}\sum_{j=1}^{n}(x_j-\mu)^2=0,$$

откуда следует оценка максимального правдоподобия

$$\widehat{\theta}=(\widetilde{\mu}, \widetilde{\sigma^2}),\quad \widetilde{\mu}=\frac{1}{n}\sum_{j=1}^{n}x_j,\quad \widetilde{\sigma^2}=\frac{1}{n}\sum_{j=1}^{n}(x_j-\widetilde{\mu})^2.$$

Остановимся подробнее на анализе оценки максимального правдоподобия параметров полиномиального распределения. Напомним, что полиномиальное распределение задается вектором вероятностей $\theta=(p_1, \dots, p_r)$ каждого из r возможных исходов испытаний. Если в серии из n испытаний i-й исход наблюдается n_i раз, $i=1, \dots, r$, $n_1+\dots+n_r=n$, и ν_i — частота i-го исхода, $\nu_1+\dots+\nu_r=1$, то

$$P\left\{\nu_1=\frac{n_1}{n}, \dots, \nu_r=\frac{n_r}{n}\,\Big|\,\theta\right\} =$$

$$= L(\nu, \theta) = \frac{n!}{n_1!\dots n_r!}\,p_1^{n_1}\dots p_r^{n_r}. \tag{21}$$

Здесь $n_1+\dots+n_r=n$, $\nu=(\nu_1, \dots, \nu_r)$, $\theta=(p_1, \dots, p_r)$. Пусть ν — вектор наблюдаемых частот событий в серии из n испытаний, вектор вероятностей $\theta=(p_1, \dots, p_r)$ неизвестен. Оценка $\widehat{\theta}$ максимального правдоподобия вектора θ определяется условиями

$$L(\nu, \widehat{\theta}) = \max_{\theta\in\Theta} L(\nu, \theta). \tag{22}$$

Поскольку множество

$$\Theta=\{\theta=(p_1, \dots, p_r),\ 0\leqslant p_i\leqslant 1,\ i=1, \dots, r,\ p_1+\dots+p_r=1\}$$

ограничено и замкнуто в R_r, а $L(\nu, \theta)$ при каждом $\nu\in\Theta$ как функция $\theta\in\Theta$ непрерывна, то максимум в (22) достигается в некоторой точке Θ. Следовательно, оценка максимального правдоподобия существует. Более того, множество Θ не только замкнуто и ограничено, но и выпукло в R_r. Докажем, что следующая функция строго вогнута на Θ

$$\ln L(\nu, \theta) - \ln\frac{n!}{n_1!\dots n_r!} = \sum_{i=1}^{r} n_i \ln p_i = f(\theta), \tag{23}$$

откуда будет следовать единственность оценки максимального правдоподобия. Воспользуемся строгой вогнутостью $\ln x$, $x>0$:

$$\ln(\alpha x+(1-\alpha)y) > \alpha\ln x+(1-\alpha)\ln y,\quad 0<\alpha<1,\ x, y>0.$$

Пусть

$$\alpha\theta+(1-\alpha)\bar{\theta}=(\alpha p_1+(1-\alpha)\bar{p}_1, \dots, \alpha p_r+(1-\alpha)\bar{p}_r).$$

Тогда

$$f(\alpha\theta+(1-\alpha)\bar{\theta}) = \ln L(\nu, \alpha\theta+(1-\alpha)\bar{\theta}) - \ln\frac{n!}{n_1!\dots n_r!} =$$

$$= \sum_{i=1}^{r} n_i \ln(\alpha p_i+(1-\alpha)\bar{p}_i) > \sum_{i=1}^{r} n_i(\alpha\ln p_i+(1-\alpha)\ln\bar{p}_i) =$$

$$= \alpha f(\theta)+(1-\alpha)f(\bar{\theta}),$$

что и означает строгую вогнутость функции (23).

Итак, оценка максимального правдоподобия θ существует и единственна. Пусть $\theta_0=(p_1^0, \dots, p_r^0)$ — истинные значения вероятностей исходов и $\widehat{\theta}=(\hat{p}_1, \dots, \hat{p}_r)$. Покажем, что $\widehat{\theta}\xrightarrow{P}\theta_0$ при числе испытаний n, стремящемся к бесконечности. Для этого нам потребуется

Лемма 2. Пусть $0 < p_i < 1$, $0 < q_i < 1$, $i = 1, \ldots, r$ и $\sum_{i=1}^{r} p_i = \sum_{i=1}^{r} q_i = 1$. Тогда

$$\sum_{i=1}^{r} p_i \ln \frac{p_i}{q_i} \geqslant \frac{1}{2} \sum_{i=1}^{r} p_i (q_i - p_i)^2. \qquad (24)$$

Доказательство. По формуле Тейлора

$$\ln x = \ln [1 + (x - 1)] = (x - 1) - \frac{(x-1)^2}{2z^2}$$

для некоторого $z \in [1, x]$, если $x \geqslant 1$, или $z \in [x, 1]$, если $0 < x < 1$. Следовательно,

$$\sum_{i=1}^{r} p_i \ln \frac{q_i}{p_i} = \sum_{i=1}^{r} p_i \left(\frac{q_i}{p_i} - 1\right) - \frac{1}{2} \sum_{i=1}^{r} p_i \frac{\left(\frac{q_i}{p_i} - 1\right)^2}{z_i^2} =$$

$$= -\frac{1}{2} \sum_{i=1}^{r} p_i (q_i - p_i)^2 \frac{1}{(p_i z_i)^2} \leqslant -\frac{1}{2} \sum_{i=1}^{r} p_i (q_i - p_i)^2,$$

так как $p_i z_i \in [p_i, q_i]$, если $p_i \leqslant q_i$, или $p_i z_i \in [q_i, p_i]$ в противном случае. ▲

Возвращаясь к поставленной задаче, найдем, воспользовавшись условием (22),

$$\sum_{i=1}^{r} \nu_i \ln \widehat{p}_i = \max \sum_{i=1}^{r} \nu_i \ln p_i \geqslant \sum_{i=1}^{r} \nu_i \ln p_i^0.$$

Вместе с (24) это означает, что

$$\sum_{i=1}^{r} \nu_i \ln \frac{\nu_i}{p_i^0} \geqslant \sum_{i=1}^{r} \nu_i \ln \frac{\nu_i}{\widehat{p}_i} \geqslant \frac{1}{2} \sum_{i=1}^{r} \nu_i (\nu_i - \widehat{p}_i)^2. \qquad (25)$$

Пусть теперь $n \to \infty$. Тогда в силу закона больших чисел $\nu_i \xrightarrow{P} p_i^0$, $i = 1, \ldots, 2$, а поскольку $\sum_{i=1}^{r} \nu_i \ln \frac{\nu_i}{p_i^0}$ — непрерывная функция $\nu = (\nu_1, \ldots, \nu_r)$, левая, а следовательно, и правая части неравенства (25) стремятся по вероятности к нулю. Отсюда, в свою очередь, следует, что $p_i \xrightarrow{P} p_i^0$, $i = 1, \ldots, r$.

Таким образом, оценка максимального правдоподобия параметров полиномиального распределения состоятельна.

§ 18. ЛИНЕЙНАЯ МОДЕЛЬ ИЗМЕРЕНИЙ

Рассмотрим задачу, характерную для экспериментальных исследований, когда требуется измерить вектор параметров $\alpha = (\alpha_1, \ldots, \alpha_k)$, но доступны наблюдению лишь линейные комбинации координат

$$a_{j1}\alpha_1 + \ldots + a_{jk}\alpha_k, \quad j = 1, \ldots, n,$$

причем $n \geqslant k$. Коэффициенты a_{ji}, $j = 1, \ldots, n$, $i = 1, \ldots, k$, считаются известными. Как правило, достаточно реалистичной является схема измерений следующего вида:

$$\xi_j = a_{j1}\alpha_1 + \ldots + a_{jk}\alpha_k + \nu_j, \quad j = 1, \ldots, n, \qquad (1)$$

где $\xi = (\xi_1, \ldots, \xi_n)$ — случайный вектор результатов измерений, $\nu = (\nu_1, \ldots, \nu_n)$ — случайный вектор ошибок. Поскольку систематические ошибки измерений всегда могут быть учтены, разумно считать, что $M\nu_j = 0$, $j = 1, \ldots, n$. Для простоты примем, что измерения (1) независимы в совокупности и имеют одинаковую точность, которую охарактеризуем дисперсией

$$M\nu_j^2 = \sigma^2, \quad j = 1, \ldots, n; \qquad (2)$$

σ^2, однако, не предполагается известной. Наконец, векторы $a_i = (a_{1i}, \ldots, a_{ni})$, $i = 1, \ldots, k$, в (1) будем считать линейно независимыми. Задача состоит в том, чтобы по результатам измерений $\xi_1, \ldots, \xi_n$ оценить $\alpha_1, \ldots, \alpha_k$ и σ^2.

Речь идет о типичной так называемой обратной задаче, когда по данным измерений требуется определить непосредственно не наблюдаемые параметры объекта или явления. В математической статистике рассматриваемую задачу принято называть задачей анализа линейной регрессии.

1°. Теорема Гаусса—Маркова

Запишем схему измерений (1) в векторном виде

$$\xi = \sum_{i=1}^{k} a_i \alpha_i + \nu, \quad \xi \in R_n. \qquad (3)$$

Не предполагая известным распределение вектора ξ, покажем, что существуют линейные несмещенные оценки параметров $\alpha_1, \ldots, \alpha_k$ с минимальной дисперсией, а также несмещенная оценка σ^2.

Линейную несмещенную оценку параметра α_i будем искать в виде

$$\widehat{\alpha}_i = \sum_{j=1}^{n} b_{ij}\xi_j = (b_i, \xi), \quad b_i = (b_{i1}, \ldots, b_{in}),$$

где $(\cdot, \cdot)$ — обозначение скалярного произведения в R_n. Требование несмещенности $\widehat{\alpha}_i$ приводит к условию

$$M\widehat{\alpha}_i = (b_i, M\xi) = \sum_{i'=1}^{k} (b_i, a_{i'}) \alpha_{i'} = \alpha_i.$$

Но поскольку α_i, $i=1, \ldots, k$, априори неизвестны, это означает, что

$$(b_i, a_{i'}) = \delta_{ii'} = \begin{cases} 1, & i = i', \\ 0, & i \neq i', \ i, i' = 1, \ldots, k. \end{cases} \tag{4}$$

Подсчитаем теперь дисперсию $\widehat{\alpha}_i$

$$D\widehat{\alpha}_i = \sum_{j=1}^{n} (b_{ij})^2 D\xi_j = \sigma^2 \sum_{j=1}^{n} (b_{ij})^2 = \sigma^2 (b_i, b_i). \tag{5}$$

Таким образом, задача определения несмещенной оценки минимальной дисперсии $\widehat{\alpha}_i$ сводится к задаче на минимум дисперсии (5) при ограничениях (4)

$$\min \{(b_i, b_i) \mid b_i \in R_n, (b_i, a_{i'}) = \delta_{ii'}, \ i' = 1, \ldots, k\}. \tag{6}$$

Речь идет о типичной задаче на условный экстремум, которую можно решать методом множителей Лагранжа. Введем функцию Лагранжа задачи (6)

$$L = (b_i, b_i) - 2 \sum_{i'=1}^{k} \Lambda_{ii'} (b_i, a_{i'}).$$

Здесь $\Lambda_{ii'}$, $i' = 1, \ldots, k$, множители Лагранжа. Дифференцируя L по координатам b_{ij}, $j=1, \ldots, n$, вектора $b_i = (b_{i1}, \ldots, b_{in})$, найдем

$$\frac{\partial L}{\partial b_{ij}} = 2b_{ij} - 2\sum_{i'=1}^{k} \Lambda_{ii'} a_{ji'}, \ j = 1, \ldots, n,$$

где $(a_{1i'}, \ldots, a_{ni'}) = a_{i'}$, $i' = 1, \ldots, k$. Следовательно, уравнения Лагранжа задачи (6) в векторной форме будут иметь вид

$$b_i = \sum_{i'=1}^{k} \Lambda_{ii'} a_{i'}, \ i = 1, \ldots, k,$$

и совместно с условиями (4) это дает следующие уравнения для определения множителей Лагранжа

$$\delta_{ii''} = \sum_{i'=1}^{k} \Lambda_{ii'} (a_{i'}, a_{i''}), \ i, i'' = 1, \ldots, k. \tag{7}$$

Так как векторы a_i, $i=1, \ldots, k$, линейно независимы, матрица $k \times k$ скалярных произведений $\{(a_i, a_{i'})\}$ невырождена. Поэтому из уравнений (7) следует

$$\Lambda_{ii'} = (a_i, a_{i'})^{-}, \quad i' = 1, \ldots, k,$$

где $(\cdot, \cdot)^{-}$ — обозначение для матричных элементов матрицы, обратной к $\{(a_i, a_{i'})\}$: $\{(a_i, a_{i'})^{-}\} = \{(a_i, a_{i'})\}^{-1}$. Следовательно,

$$b_i = \sum_{i'=1}^{k} (a_i, a_{i'})^{-} a_{i'},$$

и линейная несмещенная оценка минимальной дисперсии α_i имеет вид

$$\widehat{\alpha}_i = (b_i, \xi) = \sum_{i'=1}^{k} (a_i, a_{i'})^{-} (a_{i'}, \xi), \ i = 1, \ldots, k. \tag{8}$$

Определим теперь оценку $\widetilde{\alpha}_i$ параметра α_i из условия

$$\left\| \xi - \sum_{i=1}^{k} \widetilde{\alpha}_i a_i \right\| =$$

$$= \min \left\{ \left\| \xi - \sum_{i=1}^{k} \alpha_i a_i \right\| \,\middle|\, -\infty < \alpha_i < \infty, \ i = 1, \ldots, k \right\}. \tag{9}$$

Смысл задачи (9) состоит в том, что в качестве оценок $\widetilde{\alpha}_i$ выбираются те значения α_i, $i=1, \ldots, k$, при которых вектор $\sum_{i=1}^{k} \alpha_i a_i$ наименее уклоняется в R_n от результата измерений — вектора ξ. α_i, $i=1, \ldots, k$, называются оценками наименьших квадратов α_i, $i=1, \ldots, k$.

Для решения задачи (9) введем оператор Π_a ортогонального проектирования в R_n на линейную оболочку L_a векторов $a_1, \ldots, a_k$. Тогда искомые оценки $\widetilde{\alpha}_i$, $i=1, \ldots, k$, будут найдены из условия

$$\sum_{i=1}^{k} \widetilde{\alpha}_i a_i = \Pi_a \xi. \tag{10}$$

Действительно,

$$\begin{aligned} \min \{\|\xi - \eta\|^2 \mid \eta \in L_a\} &= \min \{\|\xi - \Pi_a\xi + \Pi_a\xi - \\ - \eta\|^2 \mid \eta \in L_a\} &= \|\xi - \Pi_a\xi\|^2 + \min\{\|\Pi_a\xi - \eta\|^2 \mid \eta \in L_a\}, \end{aligned} \tag{11}$$

так как векторы $\xi - \Pi_a\xi \in L_a^{\perp}$ и $\Pi_a\xi - \eta \in L_a$ ортогональны. Поскольку $\Pi_a\xi \in L_a$, из (11) следует, что $\min \|\xi - \eta\|$ достигается на $\eta = \Pi_a\xi$.

Равенство (10) выполняется тогда и только тогда, когда вектор $\xi-\sum_{i=1}^{k}\widetilde{\alpha}_i a_i=\xi-\Pi_a\xi$ ортогонален L_a, т. е.

$$\left(\xi-\sum_{i=1}^{k}\widetilde{\alpha}_i a_i,\ a_{i'}\right)=0,\quad i'=1,\ \ldots,\ k,$$

или, что то же самое,

$$\sum_{i'=1}^{k}(a_i,\ a_{i'})\widetilde{\alpha}_{i'}=(\xi,\ a_i),\quad i=1,\ \ldots,\ k.$$

Отсюда следует, что оценки наименьших квадратов

$$\widetilde{\alpha}_i=\sum_{i'=1}^{k}(a_i,\ a_{i'})^-(a_{i'},\ \xi),\quad i=1,\ \ldots,\ k, \tag{12}$$

совпадают с оценками (8), полученными из условия несмещенности и минимальности дисперсии и, следовательно, обладают этим экстремальным свойством.

Метод линейного оценивания, основанный на минимизации дисперсии, принадлежит Маркову (1900). Метод наименьших квадратов значительно раньше развит Гауссом (1809).

Сформулируем полученные результаты.

Теорема 1 (Гаусса—Маркова). Пусть случайные величины $\xi_1, \ldots, \xi_n$, независимы, имеют одинаковую дисперсию и $\xi=(\xi_1, \ldots, \xi_n)$. Если

$$M\xi=\sum_{i=1}^{k}\alpha_i a_i,$$

где векторы a_i, $i=1, \ldots, k$, известны и линейно независимы, то линейные несмещенные оценки минимальной дисперсии коэффициентов α_i, $i=1, \ldots, k$, даются равенствами (8) и совпадают с оценками наименьших квадратов (12). Матричные элементы матрицы ковариаций оценок равны

$$\operatorname{cov}\widehat{\alpha}_i\,\widehat{\alpha}_{i'}=\sigma^2(a_i,\ a_{i'})^-,\quad i,\ i'=1,\ \ldots,\ k. \tag{13}$$

Доказательство. Следует проверить лишь последние равенства. Поскольку согласно (8) и (3)

$$\widehat{\alpha}_i=\sum_{i'=1}^{k}(a_i,\ a_{i'})^-\left(a_{i'},\ \sum_{i''=1}^{k}a_{i''}\alpha_{i''}+\nu\right)=$$

$$=\alpha_i+\sum_{i'=1}^{k}(a_i,\ a_{i'})^-(a_{i'},\ \nu), \tag{14}$$

то

$$\operatorname{cov}\widehat{\alpha}_i\,\widehat{\alpha}_{i'}=M(\widehat{\alpha}_i-\alpha_i)(\widehat{\alpha}_{i'}-\alpha_{i'})=$$

$$=\sum_{i'',\,i'''=1}^{k}(a_i,\ a_{i''})^-\sigma^2(a_{i''},\ a_{i'''})(a_{i'''},\ a_{i'})^-=\sigma^2(a_i,\ a_{i'})^-.$$

Здесь использовано соотношение

$$M(a_{i''},\nu)(a_{i'''},\nu)=M\sum_{j=1}^{n}a_{ji''}\nu_j\sum_{j'=1}^{n}a_{j'i'''}\nu_{j'}=$$

$$=\sigma^2\sum_{j=1}^{n}a_{ji''}a_{ji'''}=\sigma^2(a_{i''},\ a_{i'''}).\ \blacktriangle$$

2°. Несмещенная оценка дисперсии σ^2, $n>k$

Рассмотрим статистики

$$s^2=\left\|\xi-\sum_{i=1}^{k}\alpha_i a_i\right\|^2=\|\nu\|^2, \tag{15}$$

$$s_1^2=\|\xi-\Pi_a\xi\|^2=\left\|\xi-\sum_{i=1}^{k}\widehat{\alpha}_i a_i\right\|^2=\|\nu-\Pi_a\nu\|^2,$$

$$s_2^2=\left\|\Pi_a\xi-\sum_{i=1}^{k}\alpha_i a_i\right\|^2=\left\|\sum_{i=1}^{k}(\widehat{\alpha}_i-\alpha_i)a_i\right\|^2=\|\Pi_a\nu\|^2.$$

Здесь ξ имеет выражение (3). Поскольку векторы $\xi-\Pi_a\xi\in L_a^{\perp}$ и $\Pi_a\xi-\sum_{i=1}^{k}\alpha_i a_i\in L_a$ ортогональны, то

$$s^2=\left\|\xi-\Pi_a\xi+\Pi_a\xi-\sum_{i=1}^{k}\alpha_i a_i\right\|^2=\|\xi-\Pi_a\xi\|^2+$$

$$+\left\|\Pi_a\xi-\sum_{i=1}^{k}\alpha_i a_i\right\|^2=s_1^2+s_2^2.$$

Далее:

$$Ms^2=M\|\nu\|^2=M\sum_{j=1}^{n}\nu_j^2=n\sigma^2,$$

и, следовательно,

$$Ms_2^2=M\left\|\sum_{i=1}^{k}(\widehat{\alpha}_i-\alpha_i)a_i\right\|^2=\sum_{i,i'=1}^{k}(a_i,\ a_{i'})\,M(\widehat{\alpha}_i-\alpha_i)(\widehat{\alpha}_{i'}-\alpha_{i'})=$$

$$=\sum_{i,i'=1}^{k}(a_i,\ a_{i'})(a_i,\ a_{i'})^-\sigma^2=k\sigma^2.$$

Поэтому $Ms_1^2 = M(s^2 - s_2^2) = (n-k)\sigma^2$ и, следовательно, статистика

$$\widehat{\sigma}^2 = \frac{\|\xi - \Pi_a \xi\|^2}{n-k} \qquad (16)$$

является несмещенной оценкой дисперсии σ^2.

Таким образом, составляющая $\Pi_a\xi$ вектора измерений определяет оценки $\widehat{\alpha}_i$, $i=1,\dots,k$, а оставшаяся информация в $\xi - \Pi_a\xi$ позволяет несмещенно оценить дисперсию измерений σ^2.

3⁰. Распределение оценок в случае нормального распределения ошибок измерений

Предположим дополнительно, что в равенствах (1) ν_j, $j=1,\dots,n$, нормально распределены $N(0, \sigma^2)$ и, как это оговорено вначале, независимы в совокупности. В таком случае, воспользовавшись результатами § 9 и равенствами (15), найдем

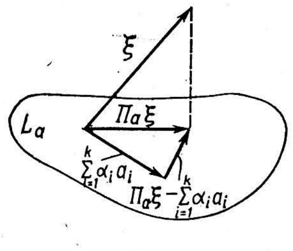


Рис. 1

$$\Big\|\xi - \sum_{i=1}^{k} \alpha_i a_i\Big\|^2 = \sigma^2 \chi_n^2, \qquad (17)$$

$$\|\xi - \Pi_a \xi\|^2 = \sigma^2 \chi_{n-k}^2,$$

$$\Big\|\Pi_a \xi - \sum_{i=1}^{k} \alpha_i a_i\Big\|^2 = \sigma^2 \chi_k^2.$$

При этом согласно следствию 2 теоремы 14.1 случайные векторы $\xi - \Pi_a\xi = \nu - \Pi_a\nu$ и $\Pi_a\xi - \sum_{i=1}^{k}\alpha_i a_i = \Pi_a\nu$ независимы как ортогональные проекции вектора ν на ортогональные подпространства $L_a^{\perp}$ и L_a соответственно. По этой причине статистика

$$\varphi_{k,n-k} = \frac{\|\Pi_a\xi - \sum_{i=1}^{k}\alpha_i a_i\|^2/k}{\|\xi - \Pi_a\xi\|^2/(n-k)} = \frac{\|\sum_{i=1}^{k}(\widehat{\alpha}_i - \alpha_i)\, a_i\|^2/k}{\|\xi - \Pi_a\xi\|^2/(n-k)}$$

контролируется распределением Фишера с k и $n-k$ степенями свободы.

Если коэффициенты $\alpha_1, \dots, \alpha_k$ рассматривать как вектор $\alpha = (\alpha_1, \dots, \alpha_k)$ в k-мерном евклидовом пространстве R_k, то статистика $\varphi_{k,n-k}$ может быть использована для получения доверительных множеств в R_k для α. Действительно, для заданного γ, $0 \leqslant \gamma \leqslant 1$, определим по таблице распределения Фишера $\varepsilon \geqslant 0$ так, чтобы

$$P\{\varphi_{k,n-k} < \varepsilon\} = \gamma.$$

Тогда с вероятностью γ вектор истинных коэффициентов $(\alpha_1, \dots, \alpha_k)$ покрывается случайным подмножеством R_k:

$$\Big\{\alpha = (\alpha_1, \dots, \alpha_k) : \Big\|\sum_{i=1}^{k}(\widehat{\alpha}_i - \alpha_i)\, a_i\Big\|^2 =$$

$$= \sum_{i,i'=1}^{k} (a_i, a_{i'})(\widehat{\alpha}_i - \alpha_i)(\widehat{\alpha}_{i'} - \alpha_{i'}) < \frac{\varepsilon k}{n-k}\|\xi - \Pi_a\xi\|^2\Big\}. \qquad (18)$$

Доверительное подмножество (18) уровня γ является эллипсоидом с центром $\widehat{\alpha} = (\widehat{\alpha}_1, \dots, \widehat{\alpha}_k)$. Понятие доверительного эллипсоида введено Хоттелингом в 1931 г.

Можно получить интервальные оценки для каждой координаты вектора $(\alpha_1, \dots, \alpha_k)$. Для этого заметим, что оценка $\widehat{\alpha}_i$, как линейная функция нормальных случайных величин, также распределена нормально. Параметры этого распределения определяются равенствами

$$M\widehat{\alpha}_i = \alpha_i, \quad D\widehat{\alpha}_i = \operatorname{cov}\widehat{\alpha}_i\,\widehat{\alpha}_i = \sigma^2 (a_i, a_i)^-, \quad i = 1, \dots, k. \qquad (19)$$

Заметим, в частности, что вектор $\widehat{\alpha} = (\widehat{\alpha}_1, \dots, \widehat{\alpha}_k)$ нормально распределен $N(\alpha, \sigma^2\{(a_i, a_j)^-\})$. Так как случайный вектор $\Pi_a\xi$ и статистика $\|\xi - \Pi_a\xi\|^2$ независимы, то соответственно независимы $\widehat{\alpha} = (\widehat{\alpha}_1, \dots, \widehat{\alpha}_k)$ и $\|\xi - \Pi_a\xi\|^2$, так как

$$\widehat{\alpha}_i = \sum_{i'=1}^{k} (a_i, a_{i'})^- (\xi, a_{i'}) =$$

$$= \sum_{i'=1}^{k} (a_i, a_{i'})^- (\Pi_a\xi, a_{i'}), \quad i = 1, \dots, k.$$

А поскольку согласно (19)

$$(\widehat{\alpha}_i - \alpha_i)/[(a_i, a_i)^-]^{1/2}$$

имеет распределение $N(0, \sigma^2)$ и $\|\xi - \Pi_a\xi\|^2 = \sigma^2\chi^2_{n-k}$, то статистика

$$\tau_{n-k} = \frac{\widehat{\alpha}_i - \alpha_i}{[(a_i, a_i)^- \|\xi - \Pi_a\xi\|^2/(n-k)]^{1/2}}$$

имеет распределение Стьюдента с $n-k$ степенями свободы. Этот факт позволяет построить доверительный интервал для α_i, $i=1,\dots,k$. Действительно, определим для заданного γ, $0 \leqslant \gamma \leqslant 1$, $\varepsilon \geqslant 0$ так, чтобы

$$P\{|\tau_{n-k}| < \varepsilon\} = \gamma.$$

Тогда случайный интервал

$$(\widehat{\alpha}_i - \varepsilon\delta_i(\xi),\ \widehat{\alpha}_i + \varepsilon\delta_i(\xi)),$$

где

$$\delta_i(\xi) = [(a_i, a_i)^- \|\xi - \Pi_a\xi\|^2/(n-k)]^{1/2} = [(a_i, a_i)^- \widehat{\sigma}^2]^{1/2}$$

является доверительным интервалом уровня γ для α_i, $i = 1, \ldots, k$.

Воспользовавшись выражением (16) для оценки дисперсии, легко получить доверительный интервал для σ^2. Действительно, в рассматриваемом случае $\widehat{\sigma}^2 = \|\xi - \Pi_a\xi\|^2/(n-k) = \sigma^2\chi^2_{n-k}$. Определим для заданного γ, $0 \leqslant \gamma \leqslant 1$, величины ε_1 и ε_2, $0 \leqslant \varepsilon_1 \leqslant \varepsilon_2$ так, чтобы

$$P\{\chi^2_{n-k} \leqslant \varepsilon_1\} = P\{\chi^2_{n-k} \geqslant \varepsilon_2\} = \frac{(1-\gamma)}{2}.$$

Тогда

$$P\{\varepsilon_1 < \widehat{\sigma}^2/\sigma^2 < \varepsilon_2\} = P\{\widehat{\sigma}^2/\varepsilon_2 < \sigma^2 < \widehat{\sigma}^2/\varepsilon_1\} = \gamma.$$

Следовательно, $(\widehat{\sigma}^2/\varepsilon_2,\ \widehat{\sigma}^2/\varepsilon_1)$ — искомый доверительный интервал уровня γ для σ^2.

Проиллюстрируем полученные результаты на примере задач о взвешивании двух предметов. Пусть α_1 и α_2 — истинные веса двух предметов. Рассмотрим следующие три схемы взвешивания:

I	II	III
$\begin{cases} \xi_1 = \alpha_1 + \nu_1, \\ \xi_2 = \alpha_2 + \nu_2; \end{cases}$	$\begin{cases} \xi_1 = \alpha_1 + \nu_1, \\ \xi_2 = \alpha_2 + \nu_2, \\ \xi_3 = \alpha_1 + \alpha_2 + \nu_3; \end{cases}$	$\begin{cases} \xi_1 = \alpha_1 + \alpha_2 + \nu_1, \\ \xi_2 = \alpha_1 - \alpha_2 + \nu_2. \end{cases}$

Согласно первой схеме каждый предмет взвешивается один раз. Во второй схеме при дополнительном взвешивании на чашку весов кладутся оба предмета. Согласно третьей схеме при первом взвешивании предметы кладутся на одну чашку весов, а при втором — на разные. Будем считать, что в любой схеме ошибки взвешивания независимы, причем $M\nu_i = 0$, $D\nu_i = \sigma^2$, $i = 1, 2, 3$.

В схеме I:

$$a_1 = (1, 0),\ a_2 = (0, 1),\ \{(a_i, a_{i'})\} = \{(a_i, a_{i'})^-\} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Следовательно,

$$\widehat{\alpha}_1 = \xi_1,\ \widehat{\alpha}_2 = \xi_2,\ \{\operatorname{cov}(\widehat{\alpha}_i\, \widehat{\alpha}_{i'})\} = \begin{pmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{pmatrix}.$$

В этой схеме $n = k$, и измерения не содержат информации, достаточной для оценивания σ^2.

В схеме II:

$$a_1 = (1, 0, 1),\ a_2 = (0, 1, 1),$$

$$\{(a_i, a_{i'})\} = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix},\ \{(a_i, a_{i'})^-\} = \begin{pmatrix} \frac{2}{3} & -\frac{1}{3} \\ -1/3 & 2/3 \end{pmatrix},$$

и по формулам (8)

$$\widehat{\alpha}_1 = \frac{2}{3}\xi_1 - \frac{1}{3}\xi_2 + \frac{1}{3}\xi_3, \qquad (20)$$

$$\widehat{\alpha}_2 = \frac{2}{3}\xi_2 - \frac{1}{3}\xi_1 + \frac{1}{3}\xi_3,$$

причем

$$\{\operatorname{cov} \widehat{\alpha}_i\, \widehat{\alpha}_{i'}\} = \sigma^2 \begin{pmatrix} 2/3 & -1/3 \\ -1/3 & 2/3 \end{pmatrix}.$$

Отсюда, в частности, следует, что дополнительное взвешивание привело к уменьшению дисперсий оценок $\widehat{\alpha}_i$, $i = 1, 2$: σ^2 в схеме I, $2\sigma^2/3$ в схеме II.

В схеме II возможно несмещенное оценивание дисперсии σ^2. Заметив, что

$$\widehat{\alpha}_1 a_1 + \widehat{\alpha}_2 a_2 =$$

$$= \frac{1}{3}((2\xi_1 - \xi_2 + \xi_3),\ (2\xi_2 - \xi_1 + \xi_3),\ (\xi_1 + \xi_2 + 2\xi_3)),$$

найдем

$$\xi - \Pi_a\xi = \xi - a_1\widehat{\alpha}_1 - a_2\widehat{\alpha}_2 =$$

$$= \frac{1}{3}((\xi_1 + \xi_2 - \xi_3),\ (\xi_2 + \xi_1 - \xi_3),\ (-\xi_1 - \xi_2 + \xi_3)).$$

Отсюда следует

$$\widehat{\sigma}^2 = \frac{1}{3}(\xi_1 + \xi_2 - \xi_3)^2.$$

Доверительный эллипсоид для этого случая определяется неравенством

$$2(\alpha_1 - \widehat{\alpha}_1)^2 + 2(\alpha_1 - \widehat{\alpha}_1)(\alpha_2 - \widehat{\alpha}_2) + 2(\alpha_2 - \widehat{\alpha}_2)^2 < 2\varepsilon\widehat{\sigma}^2.$$

Малая полуось характеризует точность измерения $\alpha_1 + \alpha_2$ и в $\sqrt{3}$ раз короче большой.

Доверительный интервал для α_i записывается в виде

$$(\widehat{\alpha}_i - \varepsilon\delta(\xi),\ \widehat{\alpha}_i + \varepsilon\delta(\xi)),$$

где

$$\delta(\xi) = \frac{\sqrt{2}}{3} | \xi_1 + \xi_2 - \xi_3 |$$

и $\widehat{\alpha}_i$ определено в (20), $i=1, 2$.

Наконец, в случае схемы III

$$\{(a_i, a_{i'})\} = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}, \quad \{(a_i, a_{i'})^-\} = \begin{pmatrix} 1/2 & 0 \\ 0 & 1/2 \end{pmatrix}$$

и

$$\widehat{\alpha}_1 = \frac{1}{2}(\xi_1 + \xi_2), \quad \widehat{\alpha}_2 = \frac{1}{2}(\xi_1 - \xi_2).$$

В этом случае разумно организованное взвешивание позволило уменьшить дисперсию оценок вдвое по сравнению с результатом схемы I и в 4/3 раза — по сравнению с результатом схемы II. В схеме III

$$\operatorname{cov} \widehat{\alpha}_1 \widehat{\alpha}_1 = \operatorname{cov} \widehat{\alpha}_2 \widehat{\alpha}_2 = \sigma^2/2.$$

Как следует изменить стратегию взвешиваний в схеме II, чтобы, не увеличивая число взвешиваний, повысить точность результатов? Для этого, очевидно, следует стремиться к тому, чтобы при каждом взвешивании на чашках весов находились оба предмета. Рассмотрим еще две схемы:

IV

$$\begin{cases} \xi_1 = \alpha_1 + \nu_1, \\ \xi_2 = \alpha_1 + \alpha_2 + \nu_2, \\ \xi_3 = \alpha_1 - \alpha_2 + \nu_3; \end{cases}$$

$a_1 = (1, 1, 1), \ a_2 = (0, 1, -1);$

$$\{(a_i, a_{i'})\} = \begin{pmatrix} 3 & 0 \\ 0 & 2 \end{pmatrix}, \quad \{(a_i, a_{i'})^-\} = \begin{pmatrix} 1/3 & 0 \\ 0 & 1/2 \end{pmatrix};$$

$$\operatorname{cov} \widehat{\alpha}_1 \widehat{\alpha}_1 = \sigma^2/3, \quad \operatorname{cov} \widehat{\alpha}_2 \widehat{\alpha}_2 = \sigma^2/2.$$

V

$$\begin{cases} \xi_1 = \alpha_1 + \alpha_2 + \nu_1, \\ \xi_2 = \alpha_1 + \alpha_2 + \nu_2, \\ \xi_3 = \alpha_1 - \alpha_2 + \nu_3; \end{cases}$$

$a_1 = (1, 1, 1), \ a_2 = (1, 1, -1);$

$$\{(a_i, a_{i'})\} = \begin{pmatrix} 3 & 1 \\ 1 & 3 \end{pmatrix}, \quad \{(a_i, a_{i'})^-\} = \begin{pmatrix} 3/8 & -1/8 \\ -1/8 & 3/8 \end{pmatrix};$$

$$\operatorname{cov} \widehat{\alpha}_1 \widehat{\alpha}_1 = 3\sigma^2/8, \quad \operatorname{cov} \widehat{\alpha}_2 \widehat{\alpha}_2 = 3\sigma^2/8.$$

Каждая из этих схем дает результаты лучшие, чем схемы I, II и III. Схема V предпочтительнее схемы IV, если в качестве критерия выбрать суммарную ошибку оценивания:

$$\frac{3}{4}\sigma^2 = \frac{3}{8}\sigma^2 + \frac{3}{8}\sigma^2 < \frac{1}{3}\sigma^2 + \frac{1}{2}\sigma^2 = \frac{5}{6}\sigma^2.$$

§ 19. ЛИНЕЙНЫЕ ЗАДАЧИ РЕДУКЦИИ ИЗМЕРЕНИЙ В ЭКСПЕРИМЕНТАЛЬНЫХ ИССЛЕДОВАНИЯХ

Рассмотрим линейную схему измерений (18.3) с несколько иной точки зрения. Для этого представим ее в виде

$$\begin{pmatrix} \xi_1 \\ \vdots \\ \xi_n \end{pmatrix} = \begin{pmatrix} a_{11} & \dots & a_{1k} \\ \vdots & & \vdots \\ a_{n1} & \dots & a_{nk} \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_k \end{pmatrix} + \begin{pmatrix} \nu_1 \\ \vdots \\ \nu_n \end{pmatrix}, \tag{1}$$

записав векторы ξ, ν, $a_i \in R_n$, $i=1, \dots, k$, и $\alpha \in R_k$ в виде столбцов и затем объединив столбцы a_i, $i=1, \dots, k$, в матрицу

$$A = \begin{pmatrix} a_{11} & \dots & a_{1k} \\ \vdots & & \vdots \\ a_{n1} & \dots & a_{nk} \end{pmatrix} \tag{2}$$

размера $n \times k$. В таких обозначениях схема (1) записывается следующим образом:

$$\xi = A\alpha + \nu \tag{3}$$

и допускает следующую интерпретацию, которой мы будем придерживаться в этом параграфе: вектор (сигнал) ξ рассматривается как результат измерения вектора (сигнала) α, полученный на линейном измерительном приборе, заданном матрицей A (на приборе A). Вектор ν определяет погрешность (шум) измерения сигнала $A\alpha$.

Итак, на вход линейного прибора A подается сигнал α, измеряется сигнал $A\alpha$ на выходе прибора, но в силу неизбежных ошибок результат измерения ξ представляется равенством (3).

Схемой (3) охватывается довольно широкий класс измерительных приборов, таких, например, как спектрометры, микроскопы оптические и электронные и т. п. Для микроскопов, например, α можно понимать как идеальное, без искажений и шума, изображение объекта, а ξ — как реальное изображение.

Современное экспериментальное оборудование, как правило, содержит вычислительную технику (ЭВМ), которую можно использовать для обработки результатов измерений. В частности, ЭВМ можно использовать для решения задачи редукции измерений, т. е. для такого преобразования сигнала $\xi \to R\xi$, в результате которого сигнал $R\xi$ можно интерпретировать так, будто он получен на выходе другого прибора, параметры которого могут быть лучше, чем параметры A. Так, например, можно решать задачу повышения разрешающей способности микроскопа, спектрометра и т. п.

В связи со сказанным возникает математическая задача описания комплекса «прибор+ЭВМ» как нового прибора с существенно более широкими возможностями. При этом, в частности, оказывается, что если в нашем распоряжении имеются два прибора, один из которых имеет «плохие» параметры, а другой — «хорошие», то, после того как к каждому будет «подключена» ЭВМ, может случиться, что комплекс «прибор+ЭВМ», соответствующий прибору с «плохими» параметрами, будет лучшим из двух.

1^0. Задача редукции к идеальному прибору

Рассмотрим линейное преобразование равенства (3)

$$R\xi = RA\alpha + R\nu. \qquad (4)$$

Вектор $R\xi$ можно интерпретировать как искаженный шумом $R\nu$ сигнал на выходе прибора RA, на вход которого подан сигнал α. Условимся считать, что идеальный прибор задается единичной матрицей

$$I = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & & & \vdots \\ 0 & \dots & 0 & 1 \end{pmatrix}$$

размера $k \times k$. Поэтому результат редукции к идеальному прибору должен иметь вид

$$\eta = R\xi = \alpha + R\nu, \qquad (5)$$

где матрица R размера $k \times n$ удовлетворяет уравнению

$$RA = I. \qquad (6)$$

Однако условием (6) матрица R, вообще говоря, не определяется однозначно. Если R — некоторое решение уравнения (6), то $R+Z$ также решение, если Z — решение однородного уравнения

$$ZA = 0. \qquad (7)$$

Обозначим $R(A)$ линейное подпространство R_n векторов вида Ax, $x \in R_k$. $R(A)$ называется **пространством значений** матрицы A и совпадает с введенной ранее линейной оболочкой L_a векторов a_i, $i=1, \dots, k$. Следовательно, оператор ортогонального проектирования на $R(A)$ определяется равенством

$$\Pi_a y = \sum_{i,i'=1}^{k} (a_i, a_{i'})^{-} (y, a_{i'}) a_i, \quad y \in R_n, \qquad (8)$$

которое можно также переписать в терминах матрицы A:

$$\Pi_a y = A(A^*A)^{-1}A^* y, \quad y \in R_n; \qquad (9)$$

здесь A^* — обозначение для транспонированной матрицы.

Впрочем, равенство (9) можно получить и непосредственно, рассмотрев вариационную задачу на минимум

$$\min \{\|y - Ax\| \mid x \in R_k\} = \|y - \Pi_a y\|. \qquad (10)$$

Действительно, вектор x, доставляющий минимум (10), удовлетворяет уравнению $A^*(y - Ax) = 0$, которое является необходимым и достаточным условием минимума квадратичной функции $f(x) = (y - Ax,\ y - Ax)$. Отсюда следует (9).

Теперь мы можем выписать общее решение уравнения (7). Умножив его справа на матрицу $(A^*A)^{-1}A^*$, найдем

$$Z\Pi_a = 0. \qquad (11)$$

С другой стороны, умножив уравнение (11) справа на матрицу A, вновь получим уравнение (7). Следовательно, уравнения (11) и (7) эквивалентны. А так как согласно (11)

$$Z = Z(I - \Pi_a),$$

то общее решение уравнения (7) можно записать в виде

$$Z = Y(I - \Pi_a),$$

где Y — произвольная матрица $k \times n$.

Поскольку $R = (A^*A)^{-1}A^*$ удовлетворяет уравнению (6), то его общее решение имеет вид

$$R = (A^*A)^{-1}A^* + Y(I - \Pi_a), \qquad (12)$$

где Y — произвольная матрица $k \times n$.

Рассмотрим задачу редукции к идеальному прибору в следующей постановке: найти матрицу R размера $k \times n$, такую, что

$$RA = I, \quad M\|R\nu\|^2 \sim \min. \qquad (13)$$

Согласно такой постановке речь идет о редукции к идеальному прибору при минимальном уровне шума $R\nu$ в (5).

Так как

$$M\|R\nu\|^2 = M(R\nu, R\nu) = \sum_{i=1}^{k} \sum_{j,s=1}^{n} R_{ij} R_{is} M\nu_j \nu_s$$

и

$$M\nu_j\nu_s = \sigma^2 \delta_{js} = \begin{cases} \sigma^2, & j = s, \\ 0, & j \neq s, \end{cases} \qquad j, s = 1, \dots, n,$$

то*

$$M\|R\nu\|^2 = \sigma^2 \operatorname{tr} RR^*.$$

* Для квадратной матрицы $S = \{s_{ij}\}_1^n$ по определению $\operatorname{tr} S = \sum_{i=1}^{n} s_{ii}$.

А поскольку $(I-\Pi_a)A=0$, то для R из (12)

$$\operatorname{tr} RR^*=\operatorname{tr}[(A^*A)^{-1}A^*+Y(I-\Pi_a)][A(A^*A)^{-1}+$$
$$+(I-\Pi_a)Y^*]=\operatorname{tr}(A^*A)^{-1}+\operatorname{tr}Y(I-\Pi_a)Y^*. \tag{14}$$

Учитывая, что для любой матрицы Y

$$0\leqslant \operatorname{tr}Y(I-\Pi_a)(I-\Pi_a)^*Y^*=\operatorname{tr}Y(I-\Pi_a)Y^*,$$

из (14) получаем, что решение задачи (13) имеет вид

$$R=(A^*A)^{-1}A^*,\quad M\|R\nu\|^2=\sigma^2\operatorname{tr}(A^*A)^{-1}, \tag{15}$$

что совпадает с результатом теоремы Гаусса—Маркова. Именно

$$R\xi=(A^*A)^{-1}A^*A\alpha+(A^*A)^{-1}A^*\nu=\alpha+(A^*A)^{-1}A^*\nu$$

есть не что иное, как вектор

$$\widehat{\alpha}=\begin{pmatrix}\widehat{\alpha}_1\\ \vdots\\ \widehat{\alpha}_k\end{pmatrix}$$

несмещенных оценок минимальной дисперсии вектора

$$\alpha=\begin{pmatrix}\alpha_1\\ \vdots\\ \alpha_k\end{pmatrix}.$$

Однако такое решение задачи редукции удовлетворительно лишь в случае достаточно низкого уровня шума $M\|R\nu\|^2=\sigma^2\operatorname{tr}(A^*A)^{-1}$. На практике обычно это не так, и требование несмещенности оказывается неоправданным.

Рассмотрим постановку задачи редукции к идеальному прибору, в которой ценой минимального смещения достигается значительное подавление шума. С этой целью определим расстояние между произвольными матрицами U и V одного размера числом

$$\rho(U,\ V)=[\operatorname{tr}(U-V)(U-V)^*]^{1/2} \tag{16}$$

и рассмотрим задачу на минимум

$$\min\{[\operatorname{tr}(RA-I)(RA-I)^*]^{1/2}\,|\,M\|R\nu\|^2\leqslant\varepsilon\}=\rho_\varepsilon. \tag{17}$$

Если R_ε — ее решение, то согласно (4) вектор $R_\varepsilon\xi$ можно рассматривать как искаженный шумом $R_\varepsilon\nu$ выходной сигнал прибора $R_\varepsilon A$, с точностью до ρ_ε совпадающего с идеальным прибором I, при заданном ограничении на уровень шума $M\|R_\varepsilon\nu\|^2\leqslant\varepsilon$.

Теорема 1. Решение задачи редукции (17) имеет вид

$$R_\varepsilon=\begin{cases}R(\omega)=A^*(AA^*+\omega\sigma^2I)^{-1},\ \omega=\omega_\varepsilon, & 0<\varepsilon<\sigma^2\operatorname{tr}(A^*A)^{-1},\\ 0,\ \varepsilon=0,\\ (A^*A)^{-1}A^*,\ \varepsilon\geqslant\sigma^2\operatorname{tr}(A^*A)^{-1},\end{cases}$$

где ω_ε — единственный корень уравнения

$$\sigma^2\operatorname{tr}A^*(AA^*+\omega\sigma^2I)^{-2}A=\varepsilon. \tag{18}$$

При $\omega\geqslant0$ имеет место закон сохранения

$$\omega\sigma^2\frac{d}{d\omega}M\|R(\omega)\nu\|^2+$$
$$+\frac{d}{d\omega}\operatorname{tr}(R(\omega)A-I)(R(\omega)A-I)^*=0. \tag{19}$$

Доказательство. Рассмотрим линейное пространство $R_{k\times n}$ матриц R размера $k\times n$ и определим в $R_{k\times n}$ расстояние по формуле (16). Поскольку для всякого $\varepsilon\geqslant0$ множество M_ε матриц R, удовлетворяющих условию

$$M\|R\nu\|^2=\sigma^2\operatorname{tr}RR^*\leqslant\varepsilon, \tag{20}$$

выпукло, ограничено и замкнуто в $R_{k\times n}$, а

$$\operatorname{tr}(RA-I)(RA-I)^*$$

как функция R на M_ε выпукла и дифференцируема (как функция матричных элементов R), то задача (17) выпукла и имеет единственное решение, которое можно найти из условий

$$(RA-I)A^*+\omega\sigma^2R=0,\ \omega\geqslant0,\ \omega(\sigma^2\operatorname{tr}RR^*-\varepsilon)=0,$$
$$\operatorname{tr}RR^*-\varepsilon\leqslant0. \tag{21}$$

Из первого уравнения (21) для $\omega\geqslant0$ найдем

$$R=R(\omega)=A^*(AA^*+\omega\sigma^2I)^{-1}. \tag{22}$$

Интенсивность шума $R\nu$

$$h(\omega)=\sigma^2\operatorname{tr}R(\omega)R^*(\omega)=\sigma^2\operatorname{tr}A^*(AA^*+\omega\sigma^2I)^{-2}A$$

является дифференцируемой функцией $\omega\geqslant0$, и ее производная равна *

$$\frac{dh(\omega)}{d\omega}=-2\sigma^4\operatorname{tr}A^*(AA^*+\omega\sigma^2I)^{-3}A. \tag{23}$$

Докажем, что производная (23) строго отрицательна для $\omega\geqslant0$. Так как

$$A^*(AA^*+\omega\sigma^2I)=(A^*A+\omega\sigma^2I)A^*, \tag{24}$$

* Так как $(AA^*+\omega\sigma^2I)(AA^*+\omega\sigma^2I)^{-1}=I$, то

$$(AA^*+\omega\sigma^2I)\frac{d}{d\omega}(AA^*+\omega\sigma^2I)^{-1}+\sigma^2I(AA^*+\omega\sigma^2I)^{-1}=0,$$

$$\frac{d}{d\omega}(AA^*+\omega\sigma^2I)^{-1}=-\sigma^2(AA^*+\omega\sigma^2I)^{-2}.$$

то, умножив (24) слева на $(A^*A+\omega\sigma^2 I)^{-1}$ и справа на $(AA^*+\omega\sigma^2 I)^{-1}$, найдем

$$A^*(AA^*+\omega\sigma^2 I)^{-1}=(A^*A+\omega\sigma^2 I)^{-1}A^*. \tag{25}$$

Поэтому

$$\operatorname{tr} A^*(AA^*+\omega\sigma^2 I)^{-2}A=\operatorname{tr}(A^*A+\omega\sigma^2 I)^{-2}A^*A.$$

Пусть $\lambda_1, \ldots, \lambda_k$ — собственные значения матрицы A^*A. Если x_p — собственный вектор, соответствующий λ_p, т. е. $AA^*x_p=\lambda_p x_p$, то

$$\lambda_p(x_p, x_p)=(x_p, A^*Ax_p)=(Ax_p, Ax_p)\geqslant 0.$$

На самом деле собственные значения λ_i, $i=1, \ldots, k$ строго положительны, поскольку матрица A^*A невырождена.

Пусть U — ортогональная матрица, такая, что

$$U^*(A^*A)U=\operatorname{diag}(\lambda_1, \ldots, \lambda_k).$$

Тогда (поскольку для ортогональной матрицы $UU^*=U^*U=I$)

$$\operatorname{tr}(A^*A+\omega\sigma^2 I)^{-3}A^*A=\operatorname{tr}U^*(A^*A+\omega\sigma^2 I)^{-3}A^*AU=$$

$$=\operatorname{tr}U^*(A^*A+\omega\sigma^2 I)^{-3}UU^*A^*AU=\sum_{i=1}^{k}\frac{\lambda_i}{(\lambda_i+\omega\sigma^2)^3}>0.$$

Поэтому

$$\frac{dh(\omega)}{d\omega}=-2\sigma^4\sum_{i=1}^{k}\frac{\lambda_i}{(\lambda_i+\omega\sigma^2)^3}<0,\ \omega\geqslant 0.$$

Отсюда следует, что $h(\omega)$ строго монотонно убывает для $\omega\geqslant 0$ и, в частности,

$$\max\{h(\omega)\,|\,\omega\geqslant 0\}=h(0)=\sigma^2\operatorname{tr}(A^*A+\omega\sigma^2 I)^{-2}A^*A|_{\omega=0}=$$
$$=\sigma^2\operatorname{tr}(A^*A)^{-1}. \tag{26}$$

Следовательно, для $\varepsilon<h(0)$ уравнение (18) $h(\omega)=\varepsilon$ однозначно разрешимо относительно ω, и если ω_ε — корень уравнения (18), то $R(\omega_\varepsilon)$ (22) удовлетворяет условиям (21) и является решением задачи (17).

Если $\varepsilon\geqslant h(0)$, то условия (21) удовлетворяются при $\omega=0$ и

$$R=R(0)=A^*(AA^*)^{-1}, \tag{27}$$

что совпадает с (15). Для $\varepsilon=0$ согласно (21) $R=0$.

Для доказательства равенства (19) найдем, воспользовавшись (22),

$$g(\omega)=\operatorname{tr}(R(\omega)A-I)(R(\omega)A-I)^*=$$
$$=\omega^2\sigma^4\operatorname{tr}(A^*A+\omega\sigma^2 I)^{-2}, \tag{28}$$

и далее

$$\frac{dg(\omega)}{d\omega}=2\omega\sigma^4\operatorname{tr}(A^*A+\omega\sigma^2 I)^{-2}-2\omega^2\sigma^6\operatorname{tr}(A^*A+\omega\sigma^2 I)^{-3}.$$

Отсюда и из (23) следует (19). ▲

Как видно из равенств (27) и (15) при $\varepsilon\geqslant h(0)$ решения задач (17) и (13) совпадают, причем в этом случае задача редукции имеет решение $R_\varepsilon\xi=R(0)\cdot\xi$ вида (4) и ему сопутствует шум $R(0)\nu$ максимальной интенсивности

$$M\|R(0)\nu\|^2=\sigma^2\operatorname{tr}(A^*A)^{-1}.$$

При этом синтезированный прибор $R(0)A$ точно совпадает с идеальным I. Если $\varepsilon<h(0)$, то согласно теореме при ограничении $M\|R\nu\|^2\leqslant\varepsilon$ на уровень шума в (4) редукция к идеальному прибору невозможна, а ближайший к идеальному прибор, редукция к которому возможна, удовлетворяет ограничению $M\|R_\varepsilon\nu\|^2=\varepsilon$ на шум и задается матрицей

$$R_\varepsilon A=A^*(AA^*+\omega_\varepsilon\sigma^2 I)^{-1}A=I-\omega_\varepsilon\sigma^2(A^*A+\omega_\varepsilon\sigma^2 I)^{-1},$$

где ω_ε — корень уравнения (18). Невязка, характеризующая отличие $R_\varepsilon A$ от I, определяется равенством (28):

$$g(\omega_\varepsilon)=\omega_\varepsilon^2\sigma^4\operatorname{tr}(A^*A+\omega\sigma^2 I)^{-2}.$$

Рассмотрим отношение

$$[h(\omega_\varepsilon)-h(0)]/[g(\omega_\varepsilon)-g(0)],\ g(0)=0, \tag{29}$$

показывающее, насколько эффективно подавляется шум h с ростом уклонения g. При $\omega_\varepsilon\to+0$ отношение (29) можно представить в виде отношения дробей:

$$\frac{h(\omega_\varepsilon)-h(0)}{\omega_\varepsilon}\ \text{и}\ \frac{g(\omega_\varepsilon)-g(0)}{\omega_\varepsilon}.$$

Первая дробь стремится к

$$h'(\omega)|_{\omega=0}=-2\sigma^4\operatorname{tr}(A^*A)^{-2},$$

вторая — к $+0$. Следовательно, при $\omega_\varepsilon\to+0$ отношение (29) стремится к $-\infty$. Впрочем, этот факт следует из закона сохранения (19) при $\omega\to 0$, если учесть, что производные $h'(\omega)$ и $g'(\omega)$ непрерывны в нуле.

Как было отмечено, при $\omega\to 0$ решение задачи редукции (17) сходится к решению задачи редукции (13). Поскольку при этом отношение (29) стремится к $-\infty$, то в задаче (17) ценой небольшого уклонения прибора $R_\varepsilon A$ от идеального I можно получить значительное подавление шума.

Связь между невязкой g и уровнем шума h

$$g(\omega)=g,\ h(\omega)=h,\ \omega\geqslant 0, \tag{30}$$

называется **оперативной характеристикой** задачи редукции (17) и может служить «паспортом» комплекса «прибор+

+ЭВМ» в задаче редукции к идеальному прибору. Оперативную характеристику можно также задать зависимостью

$$g=g(h),\ 0<h\leqslant h(0),\ g(0)=\operatorname{tr} I, \qquad (31)$$

исключив из (30) ω.

Оперативная характеристика фиксирует следующее важное свойство комплекса «прибор+ЭВМ». Для заданного уровня шума $h=\varepsilon\leqslant h(0)$ может быть синтезирован единственный ближайший к идеальному прибор $R_\varepsilon A$, причем $[\operatorname{tr}(R_\varepsilon A-I)(R_\varepsilon A-I)^*]^{1/2}=g^{1/2}(h)$. Для всякого другого прибора RA, для которого $[\operatorname{tr}(RA-I)(RA-I)^*]^{1/2}\leqslant g^{1/2}(h)$: $M\|R\nu\|^2\geqslant h=\varepsilon$.

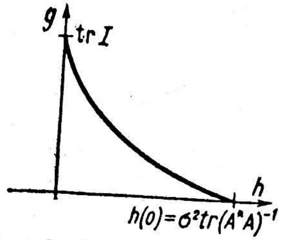


Рис. 2. Оперативная характеристика

2°. Задача редукции к заданному прибору. Понятие об общей задаче редукции.

В задаче редукции сигнала ξ к виду, какой он имел бы на выходе заданного прибора U, рассматривается линейное преобразование равенства (3)

$$R\xi=U\alpha+(RA-U)\alpha+R\nu. \qquad (32)$$

Если $RA=U$, то сигнал $R\xi$ в (32) можно интерпретировать как искаженный шумом $R\nu$ выходной сигнал прибора U, на вход которого подан сигнал α. Однако, как и в задаче редукции к идеальному прибору, определять матрицу R из условия $RA=U$ нецелесообразно по той причине, что, допуская небольшой ложный сигнал $(RA-U)\alpha$, мы получаем возможность значительно снизить уровень шума $R\nu$.

Поскольку для каждой матрицы R эффект ошибки $R\nu$ известен и определяется «энергией» $M\|R\nu\|^2$, а эффект ошибки $(RA-U)\alpha$ проявляется как ложный сигнал, неизвестный, как и сигнал α, то естественная постановка задачи редукции в данном случае сводится к минимизации ложного сигнала при заданном ограничении на уровень шума $M\|R\nu\|^2\leqslant\varepsilon$. Поскольку сигнал α априори совершенно произволен, в данном случае следует минимизировать невязку $[\operatorname{tr}(RA-U)(RA-U)^*]^{1/2}$.

Итак, рассмотрим задачу на минимум

$$\min\{[\operatorname{tr}(RA-U)(RA-U)^*]^{1/2}\,|\,M\|R\nu\|^2\leqslant\varepsilon\}=\rho_\varepsilon. \qquad (33)$$

Если R_ε — ее решение, то сигнал $R_\varepsilon\xi$ можно интерпретировать как искаженный шумом $R_\varepsilon\nu$, $M\|R_\varepsilon\nu\|^2\leqslant\varepsilon$, выходной сигнал прибора $R_\varepsilon A$, с точностью до ρ_ε совпадающего с заданным прибором U.

Решение задачи (33) вполне аналогично решению задачи редукции к идеальному прибору (17).

Теорема 2. Решение задачи редукции (33) ($U\neq 0$) имеет вид

$$R_\varepsilon=\begin{cases}0,\ \varepsilon=0,\\ R(\omega)=UA^*(AA^*+\omega\sigma^2 I)^{-1},\ \omega=\omega_\varepsilon,\\ \qquad 0<\varepsilon<\sigma^2\operatorname{tr}U(A^*A)^{-1}U^*,\\ U(A^*A)^{-1}A^*,\ \varepsilon\geqslant\sigma^2\operatorname{tr}U(A^*A)^{-1}U^*,\end{cases}$$

где ω_ε — единственный корень уравнения

$$\sigma^2\operatorname{tr}UA^*(AA^*+\omega\sigma^2 I)^{-2}AU^*=\varepsilon.$$

При $\omega\geqslant 0$ имеет место закон сохранения

$$\sigma^2\omega\frac{d}{d\omega}M\|R(\omega)\nu\|^2+\frac{d}{d\omega}\operatorname{tr}(R(\omega)A-U)(R(\omega)A-U)^*=0.$$

Доказательство может быть получено по схеме доказательства теоремы 1.

На практике часто требуется редукция не к заданному прибору, а к любому прибору, удовлетворяющему некоторым требованиям к его параметрам. Так, например, в задаче редукции данных ξ, полученных на микроскопе, к виду, какой они имели бы на приборе с более высоким разрешением, исследователю, как правило, важно контролировать лишь те параметры прибора, которые в конечном счете определят разрешение. При этом прочие параметры синтезируемого прибора могут меняться в достаточно широких пределах. Такой подход означает, что в задаче редукции следует фиксировать не прибор, а класс приборов, некоторые параметры которых подчинены заданным ограничениям. Пусть $\mathcal{U}$ — такой класс приборов. Тогда задача редукции может быть поставлена как задача на минимум

$$\min\{[\operatorname{tr}(RA-U)(RA-U)^*]^{1/2}\,|\,M\|R\nu\|^2\leqslant\varepsilon,\ U\in\mathcal{U}\}=\rho_\varepsilon. \qquad (34)$$

Если R_ε, U_ε — решение задачи (34), то сигнал

$$R_\varepsilon\xi=R_\varepsilon A\alpha+R_\varepsilon\nu=U_\varepsilon\alpha+(R_\varepsilon A-U_\varepsilon)\alpha+R_\varepsilon\nu$$

следует интерпретировать как искаженный шумом $R_\varepsilon\nu$, $M\|R_\varepsilon\nu\|^2\leqslant\varepsilon$, выходной сигнал прибора $R_\varepsilon A$, с точностью до ρ_ε совпадающего с прибором $U_\varepsilon\in\mathcal{U}$ (гарантированного качества). В задаче (34), которую будем называть общей задачей редукции, отыскивается как преобразование R_ε, подчиненное ограничению на уровень шума $M\|R_\varepsilon\nu\|^2\leqslant\varepsilon$, так и прибор U_ε из класса $\mathcal{U}$ приборов с достаточно высоким качеством, причем фактически синтезированный прибор $R_\varepsilon A$ будет ближайшим к U_ε.

Линейные приборы, используемые в физических исследованиях, часто представляются интегральными преобразованиями вида

$$f(x)=\int_a^b A(x-y)\,\alpha(y)\,dy,\ x\in[\alpha,\beta]; \tag{35}$$

здесь $A(z)$, $z\in[a-b,\ b-a]$, так называемая импульсная переходная функция прибора, $\alpha(y)$, $y\in[a,\ b]$, входной и $f(x)$, $x\in[a,\ b]$, выходной сигналы. Пусть $A(z)$ — непрерывная функция и на вход прибора поступает сигнал $\alpha(y)=$

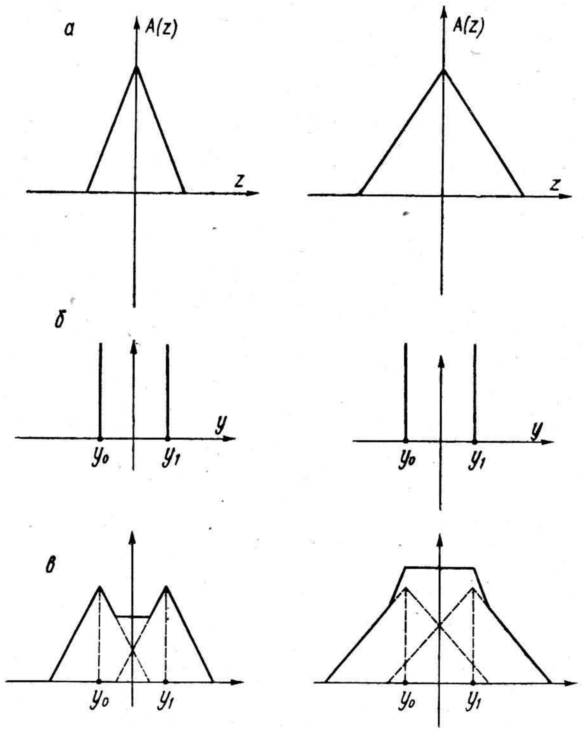


Рис. 3. *а* — импульсная переходная функция, *б* — входной сигнал, *в* — выходной сигнал

$=\delta(y-y_0)$, где $\delta(y)$ — δ-функция Дирака. Тогда для $y_0\in$ $\in[a,\ b]$ $\quad f(x)=\int_a^b A(x-y)\,\delta(y-y_0)\,dy=A(x-y_0)$ отклик прибора на импульс в точке y_0. Если на вход поступает сумма двух импульсов $\delta(y-y_0)+\delta(y-y_1)$, то отклик прибора будет равен сумме $A(x-y_0)+A(x-y_1)$, $y_0,\ y_1\in$ $\in[a,\ b]$.

На рис. 3, *а*, *б*, *в* представлены отклики на сумму импульсов приборов с «узкой» и «широкой» импульсными переходными функциями. В первом случае говорят, что прибор «разрешает» импульсы, во втором — нет. При этом качество прибора для импульсных входных сигналов можно грубо охарактеризовать «шириной» импульсной переходной функции. Чем «шире» переходная функция, тем хуже прибор.

Не вдаваясь в подробности, характеризующие связь модели (35) и ее конечномерного аналога, рассмотрим модельную задачу повышения разрешения. Зададим класс $\mathcal{U}=\mathcal{U}_s$ приборов квадратными матрицами размера $k\times k$, подчиненными условиям: $a_{ij}=0$, если $|i-j|>s$, $a_{ij}=b_{|i-j|}$, если $|i-j|\leqslant s$, где $b_0=1$, b_t, $t=1,\dots,s$, произвольны, $2s+1<k$, $i,\ j=1,\dots,k$. Будем считать, что число $s\geqslant 0$ определяет границу разрешающей способности приборов из класса $\mathcal{U}_s$. Зададим s так, чтобы разрешение приборов класса $\mathcal{U}_s$ было выше, чем у прибора A. Тогда задачу повышения разрешения можно поставить как следующую общую задачу редукции:

$$\min\{\mathrm{tr}\,(RA-U)\,(RA-U)^*\,|\,M\|R\nu\|^2\leqslant\varepsilon,\ U\in\mathcal{U}_s\}. \tag{36}$$

Если $R_{\varepsilon,s}$, $U_{\varepsilon,s}$ — решение задачи (36), то зависимости

$$g(\varepsilon,s)=\mathrm{tr}\,(R_{\varepsilon,s}A-U_{\varepsilon,s})(R_{\varepsilon,s}A-U_{\varepsilon,s})^*,$$

$$h(\varepsilon,s)=M\,\|R_{\varepsilon,s}\nu\|^2,\ \varepsilon\geqslant 0,\ s\geqslant 0,$$

определяют оперативную характеристику, связывающую уровень шума h, разрешение s и невязку g.

Если в задаче (36) s выбрано как минимальное число, при котором $A\in\mathcal{U}_s$, то ее можно рассматривать как задачу шумоподавления без ухудшения разрешения. Для этого следует выбрать ε меньшим, чем уровень шума в исходных данных $\varepsilon<M\|\nu\|^2$.

С решениями некоторых общих задач редукции можно ознакомиться в [8].

§ 20. ЗАДАЧИ ПРОВЕРКИ СТАТИСТИЧЕСКИХ ГИПОТЕЗ

В этом параграфе речь пойдет о простейших задачах проверки гипотез о параметре θ функции распределения $F(x,\ \theta)$. Как правило, гипотеза H и альтернатива K формулируются о возможных значениях неизвестного параметра $\theta\in\Theta$, а решение в пользу H или K должно быть вынесено на основании наблюдения выборки $\xi_1,\dots,\xi_n$ из распределения $F(x,\ \theta)$. Понятно, что в таком случае H и K можно просто отождествить с подмножествами Θ. Например, если параметр $\theta=\mu$ нормального распределения $N(\mu,\ \sigma^2)$ неизвестен, то можно рассматривать гипотезу H о том, что $\mu=\mu_0$, при альтернативе $K:\mu\neq\mu_0$. В этом случае $\Theta=R_1$, H можно рассматривать как подмножество прямой R_1, состоящее из одной точки $\mu_0:H=\{\mu_0\}$. Соответственно

$$K=K(\mu_0)=R_1\setminus\{\mu_0\}=\{\mu,\ -\infty<\mu<\infty,\ \mu\neq\mu_0\}.$$

Пусть D — подмножество выборочного пространства R_n, не зависящее от θ. Множество D называется критическим в задаче проверки гипотезы H при альтернативе K, если гипотеза H отвергается всякий раз, когда $(\xi_1,...,\xi_n)\in D$, и принимается в противном случае. Решение принять или отвергнуть гипотезу принимается на основе выборки $\xi_1,...,\xi_n$ из распределения $F(x,\theta)$.

Такой решающей процедуре свойственны ошибки двух типов. Говорят об **ошибке 1-го рода**, когда на основании выборки отклоняется на самом деле истинная гипотеза, и об **ошибке 2-го рода**, когда принимается на самом деле неверная гипотеза. При заданном критическом множестве D вероятности ошибок 1-го и 2-го рода равны соответственно

$$P\{(\xi_1,...,\xi_n)\in D\,|\,\theta\},\ \theta\in H,$$

$$1-P\{(\xi_1,...,\xi_n)\in D\,|\,\theta\}=P\{(\xi_1,...,\xi_n)\in R_n\setminus D\,|\,\theta\},\quad \theta\in K,$$

и зависят не только от D, H и K, но и от конкретного значения параметра $\theta\in\Theta$. Чем меньше вероятности ошибок первого и второго рода, тем лучше решающая процедура.

Вероятность отвергнуть гипотезу, когда она на самом деле не верна, равна

$$P\{(\xi_1,...,\xi_n)\in D\,|\,\theta\},\ \theta\in K,$$

и называется **мощностью решающей процедуры** (или критерия) Мощность критерия зависит от $\theta\in K$.

Для иллюстрации введенных понятий вернемся к примеру проверки гипотезы о параметре μ нормального распределения. Дисперсию σ^2 будем считать известной. Если $\mu=\mu_0$ — истинное значение математического ожидания и $(\xi_1,...,\xi_n)$ — выборка из распределения $N(\mu_0,\sigma^2)$, то согласно (15.3) μ_0 покрывается случайным доверительным интервалом (15.4) с вероятностью $1-\alpha$. Следовательно, выбирая для заданного α критическое множество в виде

$$D=\left\{(x_1,\ldots,x_n)\in R_n,\ \frac{1}{n}\sum_{j=1}^{n}x_j\geqslant\mu_0+\frac{\varepsilon\sigma}{\sqrt{n}}\right\}+$$

$$+\left\{(x_1,\ldots,x_n)\in R_n,\ \frac{1}{n}\sum_{j=1}^{n}x_j\leqslant\mu_0-\frac{\varepsilon\sigma}{\sqrt{n}}\right\},\qquad(1)$$

найдем следующее значение для ошибки 1-го рода:

$$P\{(\xi_1,\ldots,\xi_n)\in D\,|\,\mu_0\}=$$

$$=1-P\left\{-\frac{\varepsilon\sigma}{\sqrt{n}}<\frac{1}{n}\sum_{j=1}^{n}\xi_j-\mu_0<\frac{\varepsilon\sigma}{\sqrt{n}}\right\}=\alpha.$$

Величина α ошибки 1-го рода называется **уровнем критерия.**

В данном случае гипотеза H содержит единственное значение параметра $\mu=\mu_0$. Альтернатива K содержит все остальные значения μ. Поэтому ошибка второго рода

$$1-P\{(\xi_1,...,\xi_n)\in D\,|\,\mu\},\ \mu\neq\mu_0$$

зависит от неизвестного параметра μ и не может быть вычислена априори. С этим обстоятельством связаны характерные трудности, возникающие при сравнении различных решающих процедур. Действительно, если $\tilde{D}$ — другая критическая область в рассматриваемой задаче проверки гипотезы, для которой ошибка первого рода также равна α: $:P\{(\xi_1,...,\xi_n)\in\tilde{D}\,|\,\mu_0\}=\alpha$, то судить о том, какая из областей предпочтительнее, можно, лишь сравнивая вероятности ошибок второго рода для D и $\tilde{D}$ или соответственно мощности. Но последние зависят от неизвестного параметра $\mu\neq\mu_0$, и может так случиться, что для некоторых $\mu\in K$ предпочтительнее область D, а для других предпочтительнее $\tilde{D}$. В таком случае говорят, что решающие процедуры **не сравнимы.** Если оказывается, что при одном и том же уровне мощность критерия, отвечающего множеству D, не меньше, чем мощность, отвечающая $\tilde{D}$, для всех $\mu\neq\mu_0$, причем для некоторого $\mu\neq\mu_0$ неравенство строгое, то говорят, что критерий, отвечающий D, **равномерно более мощный**, чем отвечающий $\tilde{D}$. Разумеется, идеальным был бы равномерно наиболее мощный критерий, но, к сожалению, за редкими исключениями, такой критерий не существует, если K содержит более одного значения параметра.

1°. Локально наиболее мощные критерии

Если в случае $H=\{\theta_0\}$ не существует равномерно наиболее мощный критерий, т. е. не существует критическое множество, наилучшее для каждого $\theta\in K$, то можно попытаться найти критическое множество, наилучшее для значений параметра $\theta\in K$, в известном смысле близких к θ_0. Такой выбор представляет интерес, поскольку именно для близких к θ_0 значений θ велика вероятность ошибочного решения.

Рассмотрим задачу проверки гипотезы $H=\{\theta_0\}\subset\Theta$ при альтернативе $K=\{\theta\in\Theta,\ \theta>\theta_0\}$, где Θ — некоторое открытое подмножество действительной прямой R_1. Если $(\xi_1,...,\xi_n)$ — выборка из распределения с плотностью $f(x,\theta)$, $x\in R_1$, $\theta\in\Theta$, и D — критическое множество в выборочном пространстве R_n, то вероятность ошибки первого рода задается равенством

$$P\{(\xi_1,\ldots,\xi_n)\in D\mid\theta_0\}=\int\limits_D L(x,\theta_0)\,dx=\alpha, \tag{2}$$

где $L(x,\theta)=f(x_1,\theta)\ldots f(x_n,\theta)$ — функция правдоподобия, $x=(x_1,\ldots,x_n)\in R_n$, $\theta\in\Theta$. При этом мощность критерия для каждого $\theta\in K$ равна

$$\beta(\theta)=\int\limits_D L(x,\theta)\,dx. \tag{3}$$

Предположим, что для $\beta(\theta)$ имеет место формула Тейлора

$$\beta(\theta)=\alpha+(\theta-\theta_0)\beta'(\theta_0)+(\theta-\theta_0)^2\beta''(\theta_0)/2+o(\theta-\theta_0)^2,\ \alpha=\beta(\theta_0). \tag{4}$$

Для этого достаточно, чтобы функция $\beta(\theta)$ была дважды дифференцируема в некоторой окрестности $\theta=\theta_0$ и вторая производная $\beta''(\theta)$ непрерывна при $\theta=\theta_0$. Если $K=\{\theta\in\Theta,\ \theta>\theta_0\}$, то для получения локально наиболее мощного критерия следует максимизировать $\beta'(\theta_0)$ как функцию области D, или

$$\beta'(\theta_0)=\int\limits_D (\partial L(x,\theta)/\partial\theta)|_{\theta=\theta_0}\,dx, \tag{5}$$

если предположить возможность дифференцирования под знаком интеграла в (3).

Если $K=\{\theta\in\Theta,\ \theta<\theta_0\}$, то $\beta'(\theta_0)$ следует минимизировать, а в случае $K=\{\theta\in\Theta,\ \theta\neq\theta_0\}$ разумно наложить условие локальной несмещенности $\beta'(\theta_0)=0$ и максимизировать $\beta''(\theta_0)$ как функцию области D. Такой критерий естественно назвать **локально наиболее мощным, несмещенным.**

Во всех перечисленных задачах при отыскании оптимального критического множества может быть использована

Лемма (Нейман—Пирсон). Пусть $f_0(x), f_1(x),\ldots,f_m(x)$ — интегрируемые функции $x\in R_n$ и D — подмножество R_n, для которого

$$\int\limits_D f_j(x)\,dx=c_j,\ j=1,\ldots,m, \tag{6}$$

где c_j, $j=1,\ldots,m$, заданные числа. Если существуют постоянные $k_1,\ldots,k_m$, такие, что для подмножества $D_0\subset R_n$, в точках которого

$$f(x_0)\geqslant k_1f_1(x)+\ldots+k_mf_m(x)$$

и вне которого

$$f_0(x)\leqslant k_1f_1(x)+\ldots+k_mf_m(x),$$

выполнены равенства (6), то

$$\int\limits_{D_0} f_0(x)\,dx\geqslant\int\limits_D f_0(x)\,dx.$$

Доказательство. Согласно условиям леммы

$$\int\limits_{D_0} f_0(x)\,dx-\int\limits_D f_0(x)\,dx=\int\limits_{D_0\setminus(D_0\cap D)} f_0(x)\,dx-$$

$$-\int\limits_{D\setminus(D_0\cap D)} f_0(x)\,dx\geqslant\int\limits_{D_0\setminus(D_0\cap D)}\sum_{j=1}^m k_jf_j(x)\,dx-$$

$$-\int\limits_{D\setminus(D_0\cap D)}\sum_{j=1}^m k_jf_j(x)\,dx=0.$$

Равенство нулю следует из равенств

$$\int\limits_{D_0\setminus(D_0\cap D)} f_j(x)\,dx=\int\limits_{D\setminus(D_0\cap D)} f_j(x)\,dx,\ j=1,\ldots,m,$$

которые, в свою очередь, следуют из

$$\int\limits_D f_j(x)\,dx=\int\limits_{D_0} f_j(x)\,dx=c_j,\ j=1,\ldots,m.$$ ▲

Возвращаясь к задаче построения критического множества в локально наиболее мощном критерии, при оговоренных выше предположениях получаем следующий результат.

Теорема 1.

1. Если $K=\{\theta\in\Theta,\ \theta>\theta_0\}$, $H=\{\theta_0\}$, то критическое множество D локально наиболее мощного критерия уровня α имеет вид

$$D=\left\{x\in R_n,\ \frac{\partial L(x,\theta_0)}{\partial\theta_0}\geqslant k_1L(x,\theta_0)\right\}, \tag{7}$$

где k_1 определяется условием

$$\int\limits_D L(x,\theta_0)\,dx=\alpha. \tag{8}$$

Аналогично, если $K=\{\theta\in\Theta,\ \theta<\theta_0\}$, $H=\{\theta_0\}$, то

$$D=\left\{x\in R_n,\ \frac{\partial L(x,\theta_0)}{\partial\theta_0}\leqslant k_1L(x,\theta_0)\right\}$$

и k_1 определяется условием (8).

2. Если $K=\{\theta\in\Theta,\ \theta\neq\theta_0\}$, $H=\{\theta_0\}$, то критическое множество локально наиболее мощного несмещенного критерия уровня α дается равенством

$$D=\left\{x\in R_n,\ \frac{\partial^2L(x,\theta_0)}{\partial\theta_0}\geqslant k_1\frac{\partial L(x,\theta_0)}{\partial\theta_0}+k_2L(x,\theta_0)\right\},$$

где k_1 и k_2 определяются условиями

$$\beta(\theta_0) = \int_D L(x, \theta_0)\, dx = \alpha, \tag{9}$$

$$\beta'(\theta_0) = \int_D \frac{\partial L(x, \theta_0)}{\partial \theta_0}\, dx = 0.$$

В обоих случаях предполагается, что постоянные k_1 и k_1, k_2, удовлетворяющие соответственно условиям (8) и (9), существуют.

Доказательство.

1. Следует доказать, что критическое множество (7) доставляет максимум производной (5). Действительно, пусть $\tilde{D}$ — любое другое критическое множество уровня α:

$$\int_{\tilde{D}} L(x, \theta_0)\, dx = \alpha. \tag{10}$$

Выберем в лемме

$$f_1(x) = L(x, \theta_0),\ f_0(x) = \frac{\partial L(x, \theta_0)}{\partial \theta_0}.$$

Поскольку согласно (7) для $x \in D$ $f_0(x) \geqslant k_1 f_1(x)$, то на основании леммы:

$$\int_D f_0(x)\, dx \geqslant \int_{\tilde{D}} f_0(x)\, dx.$$

Следовательно, при указанных в теореме условиях производная

$$\beta'(\theta_0) = \int_D \frac{\partial L(x, \theta_0)}{\partial \theta_0}\, dx \text{ максимальна.}$$

Случай альтернативы $K = \{\theta \in \Theta,\ \theta \in \theta_0\}$ вполне аналогичен рассмотренному.

2. На этот раз выберем в лемме

$$f_0(x) = \frac{\partial^2 L(x, \theta_0)}{\partial \theta_0^2}, \quad f_1(x) = \frac{\partial L(x, \theta_0)}{\partial \theta_0},\ f_2(x) = L(x, \theta_0).$$

Тогда для любого множества $\tilde{D}$, удовлетворяющего условиям (10), согласно лемме

$$\int_D \frac{\partial^2 L(x, \theta_0)}{\partial \theta_0^2}\, dx \geqslant \int_{\tilde{D}} \frac{\partial^2 L(x, \theta_0)}{\partial \theta_0^2}\, dx,$$

т. е. D определяет несмещенный локально наиболее мощный критерий уровня α.

В качестве иллюстрации рассмотрим задачу проверки гипотезы о параметре μ нормального распределения $N(\mu, \sigma^2)$ при известной дисперсии σ^2. Пусть $H = \{\mu_0\}$, $K = \{\mu \in (-\infty, \infty),\ \mu \neq \mu_0\}$. Так как

$$L(x, \mu) = \left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^n \exp\left\{-\frac{1}{2\sigma^2}\sum_{j=1}^{n}(x_j - \mu)^2\right\}, \tag{11}$$

$$\frac{\partial \ln L(x, \mu)}{\partial \mu} = \frac{1}{\sigma^2}\sum_{j=1}^{n}(x_j - \mu),$$

$$\frac{\partial^2 \ln L(x, \mu)}{\partial \mu^2} = -\frac{n}{\sigma^2},$$

то

$$\frac{\partial L(x, \mu)}{\partial \mu} = L(x, \mu)\frac{1}{\sigma^2}\sum_{j=1}^{n}(x_j - \mu), \tag{12}$$

$$\frac{\partial^2 L(x, \mu)}{\partial \mu^2} = L(x, \mu)\left\{\left[\frac{1}{\sigma^2}\sum_{j=1}^{n}(x_j - \mu)\right]^2 - \frac{n}{\sigma^2}\right\}$$

Следовательно, критическое множество имеет вид

$$D = \left\{ x \in R_n,\ -\frac{n}{\sigma^2} + \left[\frac{1}{\sigma^2}\sum_{j=1}^{n}(x_j - \mu_0)\right]^2 \geqslant \right.$$

$$\left. \geqslant \frac{k_1}{\sigma^2}\sum_{j=1}^{n}(x_j - \mu_0) + k_2 \right\},$$

где k_1 и k_2 определяются условиями

$$\int_D L(x, \mu_0)\, dx = \alpha, \tag{13}$$

$$\int_D L(x, \mu_0)\sum_{j=1}^{n}(x_j - \mu_0)\, dx = 0.$$

Второе уравнение (13) будет удовлетворено для любого k_2, если $k_1 = 0$. Первое уравнение (13) можно записать в виде

$$P\left\{\left[\frac{1}{\sqrt{n}\,\sigma}\sum_{j=1}^{n}(\xi_j - \mu_0)\right]^2 \geqslant 1 + k_2\frac{\sigma^2}{n} \,\middle|\, \mu_0\right\} = \alpha,$$

где $(\xi_1, ..., \xi_n)$ — выборка из распределения $N(\mu_0, \sigma^2)$. Так как статистика $\frac{1}{\sqrt{n}\,\sigma}\sum_{j=1}^{n}(\xi_j-\mu_0)$ имеет распределение $N(0, 1)$, то $\left[\frac{1}{\sqrt{n}\,\sigma}\sum_{j=1}^{n}(\xi_j-\mu_0)\right]^2=\chi_1^2$ и k_2 может быть найдено по таблице распределения χ^2 из условия

$$P\left\{\chi_1^2 \geqslant 1+k_2\frac{\sigma^2}{n}\right\}=\alpha.$$

Заметим, что найденное критическое множество можно записать в виде

$$D=\left\{x\in R_n,\ \frac{1}{n}\sum_{j=1}^{n}(x_j-\mu_0)\geqslant\frac{(k_2\sigma^2+n)^{1/2}}{\sqrt{n}}\sigma\right\}+$$

$$+\left\{x\in R_n,\ \frac{1}{n}\sum_{j=1}^{n}(x_j-\mu_0)\leqslant-\frac{(k_2\sigma^2+n)^{1/2}}{\sqrt{n}}\sigma\right\},$$

совпадающем с (1).

Пусть теперь $H=\{\mu_0\}$ и $K=\{\infty>\mu>\mu_0\}$. Согласно теореме 1 критическое множество локально наиболее мощного критерия следует искать в виде

$$D=\left\{x\in R_n,\ \frac{1}{\sigma^2}\sum_{j=1}^{n}(x_j-\mu_0)\geqslant k_1\right\} \qquad (14)$$

при условии, что

$$\int_D L(x,\mu_0)\,dx=\alpha.$$

Так как статистика $\frac{1}{\sqrt{n}\,\sigma}\sum_{j=1}^{n}(\xi_j-\mu_0)=\eta_n$ имеет распределение $N(0, 1)$, а условие (15) согласно (14) можно записать в виде $P\left\{\eta_n\geqslant k_1\frac{\sigma}{\sqrt{n}}\right\}=\alpha$, то k_1 может быть найдено по таблице нормального распределения.

2⁰. Случай простой гипотезы и простой альтернативы. Наиболее мощный критерий

Если гипотеза H (альтернатива K) содержит лишь одно значение $\theta=\theta_0$ $(\theta=\theta_1)$, то гипотеза (альтернатива) называется **простой**. Выше рассматривались задачи проверки простой гипотезы при непростых альтернативах. Посмотрим, какие возможны упрощения в том случае, когда как гипотеза, так и альтернатива просты: $H=\{\theta_0\}$, $K=\{\theta_1\}$, $\theta_1\neq\theta_0$. Ограничимся случаем распределения, имеющего плотность $f(x, \theta)$, $\theta=\theta_0, \theta_1$. Рассмотрим задачу отыскания наиболее мощного критерия, отвечающего заданному уровню α. Речь идет о задаче отыскания критического множества $D\subset R_n$, для которого

$$\int_D L(x,\theta_1)\,dx\sim\max$$

при условии

$$\int_D L(x,\theta_0)\,dx=\alpha,$$

где $x=(x_1, ..., x_n)$, $L(x, \theta)=f(x_1, \theta)\ f(x_2, \theta)...f(x_n, \theta)$ — функция правдоподобия. Выберем в лемме $f_0(x)=L(x, \theta_1)$, $f_1(x)=L(x, \theta_0)$ и определим область D условием

$$D=\{x\in R_n,\ f_0(x)\geqslant k_1f_1(x)\}, \qquad (16)$$

где k_1 находится из уравнения

$$\int_D f_1(x)\,dx=\alpha. \qquad (17)$$

Если существует k_1, удовлетворяющее уравнению (16), то для любого критического множества $\tilde{D}$ уровня α, т. е. такого, что $\int_{\tilde{D}} f_1(x)\,dx=\alpha$, согласно лемме

$$\int_D f_0(x)\,dx\geqslant\int_{\tilde{D}} f_0(x)\,dx.$$

Это и означает, что критерий, отвечающий D (16), наиболее мощный среди всех критериев уровня α. Сформулируем полученный результат.

Теорема 2. В задаче проверки гипотезы $H=\{\theta_0\}$ при альтернативе $K=\{\theta_1\}$ критическая область

$$D=\{x\in R_n,\ L(x, \theta_1)\geqslant k_1L(x, \theta_0)\}$$

определяет наиболее мощный критерий уровня α, если k_1 удовлетворяет условию

$$\int_D L(x,\theta_0)\,dx=\alpha.$$

Вернемся еще раз к задаче проверки гипотезы о параметре μ нормального распределения $N(\mu, \sigma^2)$ при известной дисперсии σ^2. Пусть $H=\{\mu_0\}$, $K=\{\mu_1\}$, $\mu_1>\mu_0$. Согласно теореме 2 критическую область D, отвечающую наиболее мощному критерию, можно искать в виде

$$D=\{x\in R_n,\ L(x,\ \mu_1)\geqslant kL(x,\ \mu_0)\},\tag{18}$$

где функция правдоподобия $L(x,\ \mu)$ имеет вид (11). Семейство областей (18) можно задать иначе, воспользовавшись явным видом функции правдоподобия,

$$D=\left\{x\in R_n,\ \sum_{j=1}^{n}(x_j-\mu)^2\leqslant\sum_{j=1}^{n}(x_j=\mu_0)^2+C\right\}=\tag{19}$$

$$=\left\{x\in R_n,\ (\mu_1-\mu_0)\sum_{j=1}^{n}x_j\geqslant c_1\right\}=\left\{x\in R_n,\ \frac{1}{\sigma\sqrt{n}}\sum_{j=1}^{n}(x_j-\mu_0)\geqslant c_2\right\}.$$

Постоянную c_2 в (19) следует определить из уравнения

$$\alpha=\int_D L(x,\mu_0)\,dx=P\left\{\frac{1}{\sigma\sqrt{n}}\sum_{j=1}^{n}(\xi_j-\mu_0)\geqslant c_2\right\}.\tag{20}$$

В (19) и (20) использовано условие $\mu_1-\mu_0>0$, а также тот факт, что статистика $\frac{1}{\sigma\sqrt{n}}\sum_{j=1}^{n}(\xi_j-\mu_0)$ имеет распределение $N(0,\ 1)$, если $(\xi_1,\dots,\xi_n)$ выборка из распределения $N(\mu_0,\ \sigma^2)$. Поскольку уравнение (20) разрешимо относительно c_2 при любом $\alpha\in[0,\ 1]$, то равенство (19) при c_2, удовлетворяющем (20), определяет критическую область, отвечающую наиболее мощному критерию уровня α, $0\leqslant\alpha\leqslant1$.

В данном случае характерно, что критическая область не зависит от альтернативы μ_1 и, следовательно, определяет наиболее мощный критерий при любом $\mu_1>\mu_0$. Поэтому найденный ранее локально наиболее мощный критерий (14) в задаче проверки гипотезы $H=\{\mu_0\}$ при альтернативе $K=\{\infty>\mu>\mu_0\}$ на самом деле является равномерно наиболее мощным.

3⁰. Доверительные множества и задачи проверки гипотез

Напомним, что $100\gamma\%$-ным доверительным множеством для параметра θ распределения $F(x,\ \theta)$, или, иначе, доверительным множеством уровня γ, называется случайное подмножество $A(\xi)$ выборочного пространства R^n, такое, что

$$P\{\theta\in A(\xi)\,|\,\theta\}=\gamma,$$

где $\xi=(\xi_1,\dots,\xi_n)$ выборка из распределения $F(x,\ \theta)$. Можно сказать, что случайное множество $A(\xi)$ покрывает истинную параметрическую точку θ с вероятностью γ.

Семейство множеств $A(x)$, $x\in R_n$, зависит от вида распределения $F(x,\ \theta)$, уровня γ и называется семейством доверительных множеств уровня γ.

Связь между задачей построения доверительных множеств и задачей проверки статистических гипотез устанавливает

Теорема 3. Рассмотрим для каждого $\theta_0\in\Theta$ какой-либо критерий уровня α для проверки гипотезы $H=\{\theta_0\}$. Обозначим через $S(\theta_0)=R_n\setminus D$ область выборочного пространства R_n принятия гипотезы H (D — критическое множество). Для каждого $x\in R_n$ определим множество $A(x)$ в пространстве параметров Θ, полагая

$$A(x)=\{\theta\in\Theta,\ x\in S(\theta)\}.$$

Тогда $A(x)$, $x\in R_n$, семейство доверительных множеств для уровня $1-\alpha$, т. е.

$$P\{\theta\in A(\xi)\,|\,\theta\}=1-\alpha.$$

Если $S(\theta_0)$ определяет равномерно наиболее мощный критерий уровня α для проверки гипотезы $H=\{\theta_0\}$ при альтернативе $K(\theta_0)$, то $A(x)$ минимизирует вероятность

$$P\{\theta'\in A(\xi)\,|\,\theta\}$$

для всех $\theta\in K(\theta')$ в классе всех доверительных множеств уровня $1-\alpha$.

Доказательство. По определению множества $A(x)$: $\theta\in A(x)$ тогда и только тогда, когда $x\in S(\theta)$. Следовательно,

$$P\{\theta\in A(\xi)\,|\,\theta\}=P\{\xi\in S(\theta)\,|\,\theta\}=1-\alpha,\ \theta\in\Theta.$$

Это равенство описывает структуру доверительных множеств $A(x):A(x)$ есть совокупность тех значений θ, для которых принимается гипотеза $H=\{\theta\}$, когда наблюдается $\xi=x\in R_n$.

Пусть $\tilde{A}(x)$ — любое семейство доверительных множеств уровня $1-\alpha$ и

$$\tilde{S}'(\theta)=\{x\in R_n,\ \theta\in\tilde{A}(x)\}.$$

Тогда по определению

$$P\{\xi\in\tilde{S}(\theta)\,|\,\theta\}=P\{\theta\in\tilde{A}(\xi)\,|\,\theta\}=1-\alpha,$$

так что $\tilde{S}(\theta_0)$ является областью R_n принятия гипотезы $H=\{\theta_0\}$ уровня α. Так как $S(\theta_0)$ — область принятия гипотезы, соответствующая равномерно наиболее мощному критерию, то при любом $\theta\in K(\theta_0)$

$$P\{\xi\in\tilde{S}(\theta_0)\,|\,\theta\}\geqslant P\{\xi\in S(\theta_0)\,|\,\theta\}$$

и, следовательно,

$$P\{\theta_0\in\tilde{A}(\xi)\,|\,\theta\}\geqslant P\{\theta_0\in A(\xi)\,|\,\theta\}.\ \blacktriangle$$

В качестве простой иллюстрации рассмотрим задачу проверки гипотезы о линейной модели измерений (18.3). Будем считать, что дисперсия измерений σ^2 известна, но неизвестны векторы a_i, $i=1, ..., k$. Рассмотрим гипотезу H о том, что $a_i = \overset{\circ}{a}_i$, $i=1, ..., k$, при альтернативе $K: a_i \neq \overset{\circ}{a}_i$ хотя бы для одного $i=1, ..., k$. Если гипотеза H верна, то согласно (18.17) $\| \xi - \Pi_a^\circ \xi \|^2 = \sigma^2 \chi^2_{n-k}$, где $\Pi_a{}^0$ — оператор ортогонального проектирования на линейную оболочку векторов $\overset{\circ}{a}_1, ..., \overset{\circ}{a}_k$. Поэтому для заданного γ, $0<\gamma<1$, пользуясь таблицей распределения $\chi^2{}_{n-k}$, можно построить доверительный интервал $(\varepsilon_1, \varepsilon_2)$ уровня γ: $P\{\chi^2_{n-k} \in (\varepsilon_1 \varepsilon_2)\} = \gamma$. Соответственно критическое множество $D \subset R_n$ в задаче проверки гипотезы H при альтернативе K можно выбрать следующим образом:

$$D = \left\{ x \in R_n, \quad \frac{\| x - \Pi_a^\circ x \|^2}{\sigma^2} \in R_1 \diagdown (\varepsilon_1, \varepsilon_2) \right\}.$$

При этом вероятность ошибочно отвергнуть гипотезу H равна

$$P\{\xi \in D \,|\, a_i = \overset{\circ}{a}_i,\ i=1, ..., k\} = 1 - \gamma.$$

Как известно, ε_1 и ε_2 не определяются однозначно по заданному γ. Поэтому для определения ε_1, ε_2 можно привлечь результаты пункта 1° о локально наиболее мощных критериях.

§ 21. ЭЛЕМЕНТЫ ТЕОРИИ СТАТИСТИЧЕСКИХ РЕШЕНИЙ

Рассмотрим простейшую ситуацию, в которой возникает задача принятия решения. Будем считать известными возможные «состояния природы» $\theta \in \Theta$, например $\theta_1, \theta_2, ..., \theta_k$, и возможные «действия» $d \in D$, например, $d_1, ..., d_N$, которые связаны с состояниями природы таким образом, что действие d_i, выполненное при состоянии природы θ_j, влечет потери $l(\theta_j, d_i)$. Будем считать, что потери, сопутствующие каждой комбинации θ_j, d_i, известны, или, иначе говоря, известен риск $l(\theta, d)$, $\theta \in \Theta$, $d \in D$. Разумеется, множества Θ и D не обязательно конечны.

Если известно состояние природы θ, то вопрос о действии d естественно решается следующим образом: в каждом состоянии природы $\theta \in \Theta$ следует выполнять то или те действия, при которых риск $l(\theta, d)$ минимален. В данном случае стратегия действия состоит в наблюдении за состояниями природы и принятии определенного решения о действии $d = d_i$, если минимум $l(\theta, d)$ как функции $d \in D$ достигается на одном действии d_i. Если минимум $l(\theta, d)$, соответствующий состоянию природы θ, достигается на нескольких $d \in D$, скажем, на $d_{i_1}, \ldots, d_{i_m}$, то можно предпринять любое из

них. Но можно также воспользоваться случайным экспериментом с m исходами $\alpha_1, \alpha_2, ..., \alpha_m$, которым отвечают вероятности $p_1, p_2, ..., p_m$. В этом случае в состоянии природы θ, прежде чем принять решение о действии, можно разыграть случайный эксперимент и принять решение о действии d_{i_p}, если исходом эксперимента окажется α_p. Такая стратегия называется **рандомизированной**, в отличие от предыдущих, которые называются **чистыми**. В случае рандомизированной стратегии риск $l(\theta, d)$ при фиксированном θ является случайной величиной, причем

$$Ml(\theta, d) = l(\theta, d_i), \; i = i_1, ..., i_m.$$

На самом деле, конечно, состояние природы в момент принятия решения, как правило, неизвестно. Если, однако, о состоянии природы неизвестно ничего, то нет и задачи принятия решения: можно принимать любое решение, так как в терминах риска невозможно привести аргументов в пользу какого-либо одного из них. Поэтому будем считать, что возможны наблюдения над природой $x \in X$, например, с возможными значениями $x_1, ..., x_p$. Наблюдения должны содержать некоторую информацию о состоянии природы. Будем считать, что эта информация задается условным распределением $p(x|\theta)$ наблюдений $x \in X$ для каждого $\theta \in \Theta$. Теперь решение о действии может приниматься на основе результата наблюдения над природой.

Определим стратегию s как отображение множества наблюдений X на множество действий D. Если

$$s(x) = d, \tag{1}$$

то стратегия s при наблюдении x предписывает действие d. При этом с каждой стратегией s связано разбиение (которое мы также обозначим s) множества наблюдений X на подмножества: $X = D_1 + ... + D_N$, где

$$D_j = \{x \in X, \; s(x) = d_j\}, \; j = 1, ..., N. \tag{2}$$

Каждая стратегия действия сопряжена с риском, и, естественно, лучшей является та стратегия, которой соответствует меньший риск. Вопрос сводится к выбору лучшей стратегии.

1°. Средний риск. Рандомизация решения

Пусть s — некоторая стратегия. Тогда распределение $p(x|\theta)$ можно пересчитать в распределение $p_s(d|\theta)$ и вычислить средние потери, которые влечет применение стратегии s в состоянии природы $\theta_i \in \Theta$:

$$L_i(s) = \sum_{t=1}^{N} l(\theta, d_t) \, p_s(d_t | \theta_i), \; i = 1, \ldots, k. \tag{3}$$

Поскольку та или иная стратегия интересует нас лишь с точки зрения сопутствующих потерь, то исчерпывающей характеристикой s является точка с координатами $L_1(s), \ldots, L_k(s)$. На рис. 4 представлен случай двух состояний природы θ_1 и θ_2. Здесь изображены пять точек, характеризующие потери пяти стратегий.

Если для стратегий s_1 и s_2

$$l(s_1) = (L_1(s_1),\, L_2(s_2)) \leqslant (L_1(s_2),\, L_2(s_2)) = l(s_2),$$

что означает $L_1(s_1) \leqslant L_1(s_2)$, $L_2(s_1) \leqslant L_2(s_2)$, то говорят, что стратегия s_1 **доминирует** над s_2. При этом стратегии s_2 в любом случае будут сопутствовать потери не меньшие, чем стратегии s_1, и, следовательно, s_2 можно исключить из рассмотрения. Точки, отвечающие стратегиям s_1 и s_2, представлены на рис. 4. Из сказанного следует, что нас могут интересовать лишь те стратегии, которым на рис. 4 соответствуют такие точки, для которых нет других точек, расположенных левее и ниже.

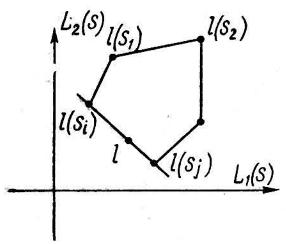


Рис. 4

Обозначим через $l(s_i)$ и $l(s_j)$ точки, соответствующие стратегиям s_i и s_j, p_i и p_j — вероятности, $p_i + p_j = 1$. Точка $l = p_i l(s_i) + p_j l(s_j)$ лежит на прямой, соединяющей $l(s_i)$ и $l(s_j)$, причем между точками $l(s_i)$ и $l(s_j)$. Рассмотрим рандомизированную стратегию s, согласно которой с вероятностью p_i применяется стратегия s_i и с вероятностью p_j — стратегия s_j. Средний риск, сопутствующий стратегии s, дается равенствами

$$L_1(s) = p_i L_1(s_i) + p_j L_1(s_j), \quad L_2(s) = p_i L_2(s_i) + p_j L_2(s_j).$$

Можно рассматривать рандомизированные стратегии, в которых используются произвольные стратегии $s_1, s_2, \ldots, s_t$. Понятно, что, например, множество точек на рис. 4, соответствующих всем рандомизированным стратегиям, является выпуклой оболочкой, построенной на $l(s_1), l(s_2), \ldots, l(s_5)$. При этом важно отметить, что множество точек, представляющих рандомизированные стратегии, выпукло.

Пусть рандомизированная стратегия s состоит из стратегий $s_1, s_2, \ldots, s_t$, применяемых с вероятностями $p_1, p_2, \ldots, p_t$. Тогда средний риск, связанный с применением s в состоянии θ_i, равен

$$\overline{L_i(s)} = \sum_{j=1}^{t} L_i(s_j)\, p_j = \sum_{j=1}^{t} \sum_{m=1}^{N} l(\theta_i, d_m)\, p_{s_j}(d_m \mid \theta_i)\, p_j. \qquad (4)$$

2⁰. Минимаксная стратегия

Минимаксная стратегия определяется условием: максимальная потеря $\max\limits_j L_j(s)$ должна быть минимальной (по s). Определим на плоскости $\{(L_1, L_2)\}$ множество точек вида

$$U(c) = \{(L_1, L_2) : \max(L_1, L_2) = c\}.$$

Очевидно, минимаксная стратегия определяется как соответствующая точке на плоскости $\{(L_1, L_2)\}$, принадлежащей как выпуклому множеству, характеризующему потери при всевозможных рандомизированных стратегиях, так и множеству $U(c)$, порожденному наименьшим c, при котором эти два множества пересекаются. Минимаксный риск равен этому минимальному значению c. Соответствующая стратегия может оказаться как рандомизированной, так и нет. Сказанное иллюстрируется на рис. 5, на котором хорошо видно преимущество рандомизации.

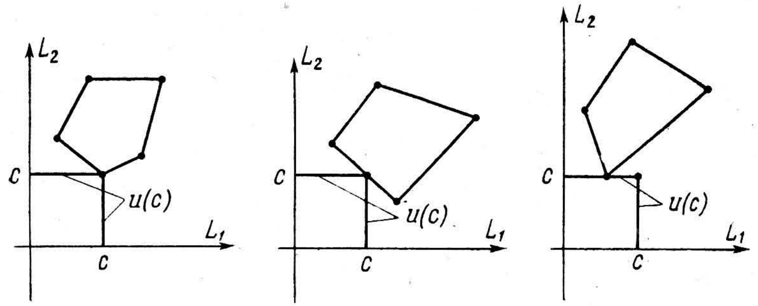


Рис. 5

3⁰. Байесовская стратегия

Байесовская стратегия применяется в случае, когда известны априорные вероятности состояния природы $p(\theta_1), \ldots, p(\theta_k)$. Байесовской называется стратегия s, минимизирующая средние потери

$$L(s) = \sum_{i=1}^{k} L_i(s)\, p(\theta_i) = \sum_{t=1}^{N} \sum_{i=1}^{k} l(\theta_i, d_t)\, p_s(d_t \mid \theta_i)\, p(\theta_i). \qquad (5)$$

Рассмотрим на плоскости $\{(L_1, L_2)\}$ семейство прямых $p_1 L_1 + p_2 L_2 = \mathrm{Const}$, $p_1 + p_2 = 1$. Очевидно, байесовская стратегия соответствует первой точке пересечения прямой при ее движении от начала координат и выпуклого множества потерь, как это показано на рис. 6. Здесь $p_1 = p(\theta_1)$, $p_2 = p(\theta_2)$, и отмечена точка на плоскости, соответствующая байесовской стратегии.

Таким образом сформулировано не только определение байесовской стратегии, но и показано, как ее найти. Однако существует другой, более эффективный способ отыскания байесовской стратегии, не требующий рассмотрения всех возможных стратегий. Мы получим этот способ несколько позже, а сейчас заметим, что минимаксная стратегия может быть получена как частный случай байесовской, если подобрать априорное распределение вероятностей состояний при-

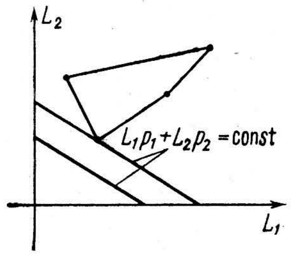


Рис. 6

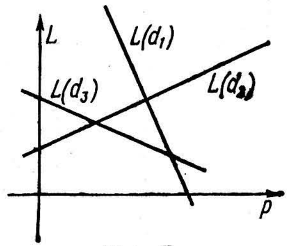


Рис. 7

роды так, чтобы соответствующие средние потери (4) оказались максимальными.

4⁰. Байесовская стратегия в случае невозможности наблюдений над природой

Пусть задано априорное распределение вероятностей состояний природы θ, $p(\theta_1),\ldots,p(\theta_k)$, но наблюдения над природой невозможны. В таком случае средние потери, связанные с действием d_j, равны

$$L(d_j)=\sum_{i=1}^{k} l(\theta_i, d_j)\,p(\theta_i), \tag{6}$$

и байесово действие определяется из условий $L(d_j)\sim\min\limits_j$.

Рассмотрим плоскость $\{(p, L)\}$ и зададим распределение состояний природы в виде $p(\theta_1)=p$, $p(\theta_2)=1-p$, где p — переменный параметр, $0\leqslant p\leqslant 1$. Тогда в зависимости от значения p байесово действие будет d_1, d_2 или d_3, как это показано на рис. 7, где представлены четыре прямые $l(\theta_1, d_j)p+ +l(\theta_2, d_j)(1-p)=L(d_j)$, $j=1, 2, 3, 4$. В точках пересечения прямых возможна рандомизация.

Заметим, что в рассмотренной ситуации определяется не стратегия, а непосредственно байесово действие. Однако оказывается, что и в общем случае байесовского решения, когда производится наблюдение над природой, задача может быть сведена к только что рассмотренной, если пересчитать априорное распределение состояний природы в апостериорное, учитывая наблюдения над природой.

5⁰. Байесово действие

Рассмотрим вместо стратегии s соответствующее разбиение s пространства наблюдений

$$X=D_1+\ldots+D_N,\ D_j=\{x\in X,\ s(x)=d_j\},\ j=1,\ldots,N.$$

Вероятность действия d_m в состоянии природы θ_j равна

$$p_s(d_m\,|\,\theta_j)=P\{x\in D_m\,|\,\theta_j\}=\sum_{x\in D_m} p(x\,|\,\theta_j), \tag{7}$$

и выражение (5) для средних потерь теперь может быть переписано в виде

$$L(s)=\sum_{i=1}^{k}\sum_{t=1}^{N} l(\theta_i, d_t)\,p_s(d_t\,|\,\theta_i)\,p(\theta_i)=$$

$$=\sum_{t=1}^{N}\sum_{i=1}^{k} l(\theta_i, d_t)\,p(\theta_i)\sum_{x\in D_t} p(x\,|\,\theta_i). \tag{8}$$

Как следует из (8), (7), для того чтобы потери (8) были минимальны, необходимо и достаточно, чтобы в множество D_t были отнесены те наблюдения $x\in X$, для которых

$$\sum_{i=1}^{k} l(\theta_i, d_t)\,p(x\,|\,\theta_i)\,p(\theta_i)\leqslant\sum_{i=1}^{k} l(\theta_i, d_j)\,p(x\,|\,\theta_i)\,p(\theta_i),\quad j=1,\ldots,N. \tag{9}$$

Иными словами, по наблюдению x следует принять решение d_t, если x удовлетворяет (9). Если с каждым наблюдением будет связано такое байесово действие, то средние потери (8) также будут байесовскими (минимальными) и стратегия s, соответствующая разбиению $X=\sum\limits_{j=1}^{N} D_j$, также будет байесовской.

Покажем, что решение этой задачи сводится к решению предыдущей с помощью байесовского пересчета априорных вероятностей состояний природы в апостериорные, индуцированные наблюдением x. Решение d_t по наблюдению x приводит к следующим средним потерям:

$$L_{d_t}(x)=\sum_{i=1}^{k} l(\theta_i, d_t)\,p(\theta_i\,|\,x)=\sum_{i=1}^{k} l(\theta_i, d_t)\times$$

$$\times\left[\frac{p(\theta_i)\,p(x\,|\,\theta_i)}{\sum\limits_{j=1}^{k} p(\theta_j)\,p(x\,|\,\theta_j)}\right]. \tag{10}$$

Сравнивая это выражение с (9), нетрудно видеть, что условие (9) эквивалентно следующему: по наблюдению x принимается то решение d_t, для которого средние потери (10) $L_{d_t}(x)$ минимальны. Но выражение (10) совпадает с (6), если в последнем априорные вероятности $p(\theta_j)$ заменить на соответствующие апостериорные $p(\theta_j|x)$, $j=1, ..., k$.

Одновременно, разумеется, определена и байесовская стратегия s. Однако s теперь определена в терминах байесовых действий: в связи с каждым наблюдением x принимается решение о байесовом действии d. Эти действия и определяют байесовскую стратегию s.

Рассмотрим произвольную стратегию s. Согласно (10) s при наблюдении x в среднем приводит к потерям

$$L_{s(x)}(x) = \sum_{i=1}^{k} l(\theta_i, s(x)) p(\theta_i|x) = \sum_{i=1}^{k} l(\theta_i, s(x)) \left[\frac{p(\theta_i)p(x|\theta_i)}{\sum_{j=1}^{k} p(\theta_j)p(x|\theta_j)} \right]. \tag{11}$$

Покажем, что средние потери $L(s)$ (5), соответствующие стратегии s, могут быть получены из (11) усреднением по всем наблюдениям, т. е. что

$$L(s) = ML_{s(x)}(x). \tag{12}$$

Действительно, математическое ожидание M можно представить в виде

$$M = \sum_{i=1}^{k} p(\theta_i) M_i,$$

где M_i — оператор условного математического ожидания при условии, что наблюдения x распределены согласно $p(x|\theta_i)$, или, иначе говоря, M_i — оператор математического ожидания при условии, что природа находится в состоянии θ_i. Пусть $D_j = \{x : s(x) = d_j\}$ и $\chi_{D_j}(x)$ — индикаторная функция D_j, так что

$$P\{s(x) = d_j | \theta_i\} = \sum_{x \in D_j} p(x|\theta_i) = \sum_{x \in X} \chi_{D_j}(x) p(x|\theta_i). \tag{13}$$

Теперь доказательство следует из цепочки равенств:

$$ML_{s(x)}(x) = \sum_{i=1}^{k} p(\theta_i) M_i \left(\sum_{j=1}^{N} \chi_{D_j}(x) L_{s(x)}(x) \right) =$$

$$= \sum_{i=1}^{k} \sum_{j=1}^{N} p(\theta_i) \sum_{x \in X} p(x|\theta_i) \chi_{D_j}(x) L_{s(x)}(x) =$$

$$= \sum_{j=1}^{N} \sum_{x \in X} \chi_{D_j}(x) \sum_{i=1}^{k} l(\theta_i, s(x)) p(\theta_i|) p(x|\theta_i) =$$

$$= \sum_{j=1}^{N} \sum_{i=1}^{k} l(\theta_i, d_j) p(d_j|\theta_i) p(\theta_i) = L(s);$$

$$l(\theta_i, s(x)) = l(\theta_i, d_j), \ x \in D_j, \ j = 1, \ldots, N.$$

Приведем теперь формальное доказательство того, что последовательность байесовых действий действительно определяет байесовскую стратегию.

Теорема 1. Пусть $\mathscr{K}$ — класс стратегий s, таких, что решение $s(x) = d_t$ принимается для x, удовлетворяющих неравенствам

$$L_{d_t}(x) \leqslant L_{d_j}(x), \ j = 1, \ldots, N,$$

где $L_{d_j}(x)$ — условные потери, определенные в (10). Тогда для всякой стратегии $s \in \mathscr{K}$ выполняется неравенство $L(s) \leqslant L(s^*)$, где s^* — произвольная стратегия. Иными словами, всякая стратегия из $\mathscr{K}$ является байесовской.

Доказательство. Обозначим $l(x) = \min_t L_{d_t}(x)$. Тогда $Ml(x) = ML_{s(x)}(x)$ при $s \in \mathscr{K}$. Действительно,

$$Ml(x) = M \sum_{j=1}^{N} \chi_{D_j}(x) \min_t L_{d_t}(x) = M \sum_{j=1}^{N} \chi_{D_j}(x) L_{d_j}(x) =$$

$$= M \sum_{j=1}^{N} \chi_{D_j}(x) L_{s(x)}(x) = ML_{s(x)}(x) = L(s),$$

где использовано, что при $x \in D_t$:

$$L_{d_t}(x) \leqslant L_{d_j}(x) \rightarrow l(x) = L_{d_t}(x), \ s(x) = d_t.$$

Отсюда следует

$$L(s) = M \min_t L_{d_t}(x) \leqslant ML_{s^*(x)}(x) = L(s^*). \qquad \blacktriangle$$

6⁰. Байесовская классификация

Рассмотрим частный случай задачи статистического решения, в котором действиями $d \in D$ являются решения о состоянии природы. Множество действий в этом случае совпадает с множеством решений. Сохраним прежние обозначения: θ — состояние природы, d — решение о состоянии природы, принятое по наблюдениям x над природой.

Полученные ранее результаты, разумеется, справедливы и в рассматриваемом случае. Байесово действие теперь является байесовским решением. Соответствующая стратегия теперь будет стратегией решения.

В задаче решения потери чаще всего оцениваются посредством количества ошибок. Рассмотрим в этой связи некоторые характерные решающие правила.

7⁰. Стратегия решения, минимизирующая среднее число ошибок

Выберем функцию потерь в виде

$$l(\theta_i, d_j) = 1 - \delta_{ij}, \quad \delta_{ij} = \begin{cases} 1, & i = j, \\ 0, & i \neq j. \end{cases}$$

Тогда средние условные потери (10) принимают вид

$$L_{d_t}(x) = \sum_{i=1}^{k} l(\theta_i, d_t)\, p(\theta_i \mid x) = 1 - p(\theta_t \mid x), \tag{14}$$

и соответственно средние потери даются равенством (5)

$$L(s) = \sum_{i,j} (1 - \delta_{ij})\, p(d_j \mid \theta_i)\, p(\theta_i) = \sum_i p(\theta_i)(1 - p(d_i \mid \theta_i)). \tag{15}$$

Но последнее выражение совпадает со средней долей ложных решений. Следовательно, байесовская стратегия решения в этом случае совпадает со стратегией, минимизирующей среднее число ошибок решения.

По данному наблюдению x байесовская стратегия решения предписывает принять решение d_t, для которого средние потери (14) минимальны. Следовательно, стратегия решения, минимизирующая среднее число ошибок, предписывает принять решение d_t, если $p(\theta_t \mid x) \geqslant p(\theta_j \mid x)$. При этом

$$D_t = \{x : s(x) = d_t\} = \{x : p(\theta_t \mid x) \geqslant p(\theta_j \mid x),\ j = 1, \dots, k\}. \tag{16}$$

Иными словами, речь идет о решении по максимуму апостериорной вероятности состояния природы θ_t; наблюдение x относится к тому состоянию природы, которое наиболее вероятно при этом наблюдении.

Согласно (16) можно также записать

$$D_t = \{x : p(x \mid \theta_t)\, p(\theta_t) \geqslant p(x \mid \theta_j)\, p(\theta_j),\ j = 1, \dots, k\}. \tag{17}$$

Заметим, что вероятность верно опознать состояние природы θ_t равна

$$P\{s(x) = d_t \mid \theta_t\} = \sum_{x \in D_t} p(x \mid \theta_t) = p(d_t \mid \theta_t),$$

и в среднем доля верных решений дается равенством

$$P = \sum_t p(d_t \mid \theta_t)\, p(\theta_t) = \sum_t \sum_{x \in D_t} p(x \mid \theta_t)\, p(\theta_t).$$

Если априорные вероятности состояний природы одинаковы, то равенство (17) переписывается в виде

$$D_t = \{x : p(x \mid \theta_t) \geqslant p(x \mid \theta_j),\ j = 1, \dots, k\},$$

и речь идет о решении по принципу максимума правдоподобия. Решение по принципу максимума правдоподобия есть частный случай байесовского при функции потерь $l(\theta_i, d_j) = 1 - \delta_{ij}$, $i, j = 1, \dots, k$, и равных априорных вероятностях состояний природы.

ЛИТЕРАТУРА

1. Боровков А. А. Теория вероятностей. — М.: Наука, 1976.
2. Ван дер Варден Б. Л. Математическая статистика. — М.: ИЛ, 1960.
3. Гихман И. И., Скороход А. В. Введение в теорию случайных процессов. — М.: Наука, 1965.
4. Гнеденко Б. В. Курс теории вероятностей. — М.: Физматгиз, 1969.
5. Ильин В. А., Позняк Э. Г. Основы математического анализа, ч. 2. — М.: Наука, 1973.
6. Колмогоров А. Н. Об аналитических методах в теории вероятностей. — УМН, 1938, вып. 5, с. 5—41.
7. Пытьев Ю. П., Шишмарев И. А. Курс теории вероятностей и математической статистики для физиков (конспект лекций). — М.: Изд-во Моск. ун-та, 1978.
8. Пытьев Ю. П. Задачи редукции в экспериментальных исследованиях. — Мат. сб., 1983, т. 120, № 2.
9. Розанов Ю. А. Случайные процессы. — М.: Наука, 1971.
10. Рытов С. М. Введение в статистическую радиофизику, ч. 1. — М.: Наука, 1976.
11. Тутубалин В. Н. Теория вероятностей. — М.: Изд-во Моск. ун-та, 1972.
12. Уилкс С. С. Математическая статистика. — М.: ИЛ, 1960.
13. Феллер В. Введение в теорию вероятностей и ее приложения. — М.: Мир, т. 1, 1964, т. 2, 1967.
14. Худсон Д. Статистика для физиков. — М.: Мир, 1970.

ЮРИЙ ПЕТРОВИЧ ПЫТЬЕВ,
ИЛЬЯ АНДРЕЕВИЧ ШИШМАРЕВ

КУРС ТЕОРИИ ВЕРОЯТНОСТЕЙ
И МАТЕМАТИЧЕСКОЙ СТАТИСТИКИ
ДЛЯ ФИЗИКОВ

Заведующий редакцией *С. И. Зеленский*
Редактор *Г. Е. Горелик*
Мл. редактор *О. Е. Силантьева*
Художник *Н. Н. Сенько*
Художественный редактор *Л. В. Мухина*
Технический редактор *Г. Д. Колоскова*
Корректоры *Л. А. Айдарбекова, Т. С. Милякова, М. К. Соболева*

Тематический план 1983 г. № 72
ИБ № 1485

Сдано в набор 10.02.83.
Подписано к печати 09.09.83.
Л-95505 Формат 60×90$^{1}/_{16}$.
Бумага тип. № 3.
Гарнитура литературная.
Высокая печать.
Усл. печ. л. 16,0 Уч.-изд. л. 16,89.
Тираж 11000 экз. Заказ 38.
Цена 75 коп. Изд. № 2353.

Ордена «Знак Почета» издательство
Московского университета.
103009, Москва, ул. Герцена, 5/7.
Типография ордена «Знак Почета»
изд-ва МГУ.
Москва, Ленинские горы