Тема 1: Элементарная теория вероятностей случайных событий

1.1 **Исходные положения**

Теория вероятностей (ТВ) – раздел математики, в котором изучаются методы исследования количественных закономерностей, характерных для событий, происходящих в многократно повторяющихся ситуациях, в которых появление таких событий невозможно заранее предсказать.

Определим и поясним смысл основных терминов, используемых в теории вероятностей событий.

1 Испытания

Упорядоченная совокупность действий с идеальными объектами некоторой рассматриваемой математической модели, многократно повторяющаяся в определенных условиях, в ТВ называется *испытанием (мысленным экпериментом)*.

Испытаниями, например, являются подбрасывания монеты или игрального кубика, извлечение шаров из урны и т. п. При решении вероятностных задач реальные ситуации необходимо предварительно формализовать (см. Практикум, часть 1, Приложение) в виде определенной схемы испытаний с идеальными объектами в идеальных условиях.

В реальных условиях всегда имеют место отклонения от идеальности объектов испытаний, а также действий и условий их выполнения при испытаниях. В теории вероятностей совокупное влияние указанных отклонений рассматривается как общая причина проявления случайности в результатах испытаний.

Законченную упорядоченную последовательность всех действий с объектами в определенных условиях, после чего эти действия при испытаниях повторяются, будем называть *схемой испытаний*. Однократное применение схемы испытаний будем считать *однократным испытанием*, а повторяющиеся конечное количество раз (более одного) однократные испытания будем называть *многократным испытанием*.

Однократные испытания могут производиться одновременно с несколькими объектами (параллельные испытания) или поочередно (последовательные испытания).

Пример 1. *Урновая схема испытаний* (используется как расчетная схема при формулировке и решении задач статистического контроля качества продукции методом математического моделирования).

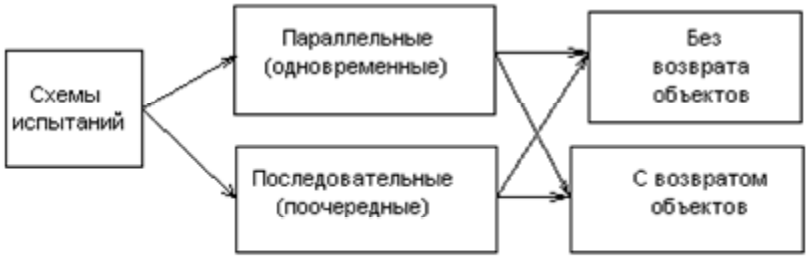


Рис. 25.1 Варианты схем испытаний

В урне имеется некоторое количество пронумерованных или разноцветных шаров, неразличимых на ощупь. Из урны случайным образом вынимаются шары.

Однократное испытание с возвратом (с возвращением) – из урны вынимают сразу несколько шаров, регистрируют их номера (цвет) и возвращают в урну. Затем можно провести следующее однократное испытание с тем же набором шаров в урне.

Однократное испытание без возврата - из урны вынимают сразу несколько шаров, регистрируют их номера (цвет) и не возвращают в урну. Затем можно провести следующее однократное испытание, но уже с оставшимся набором шаров в урне .

Единичное испытание – из урны вынимают один шар, регистрируют и возвращают или не возвращают в урну.

Основные варианты урновой схемы испытаний представлены на Рис.25.1.

Замечание. При последовательной схеме испытаний однократное испытание с несколькими объектами при необходимости представляют как последовательность единичных испытаний.

2 События

Любой результат испытаний в ТВ называют *событием*.

Исходной классификацией событий является деление их на *достоверные, случайные* и *невозможные*.

Определение. Событие называется *достоверным*, если оно обязательно должно произойти в каждом однократном испытании при любом многократном их повторении в неизменных условиях. Событие называется *невозможным*, если оно не может произойти при любом числе повторений однократных испытаний в неизменных условиях. Все остальные события называются *случайными*.

Замечание. Событие является случайным, если оно обязательно происходит при многократных испытаниях, но заранее невозможно предсказать, сколько раз и в каких именно однократных испытаниях.

Пример 2. В урне находятся 10 шаров, неразличимых на ощупь, среди которых 6 белых шаров и 4 черных.

Следующие события являются достоверными:

- наугад вынутый из урны шар не является красным;
- количество наугад вынутых из урны шаров не превышает 10;
- среди пяти наугад вынутых из урны шаров есть хотя бы один белый;

Следующие события являются невозможными:

- наугад вынутый из урны шар – желтый;
- среди семи вынутых из урны шаров нет черных;

Следующие события являются вероятными:

- два наугад вынутых из урны шара – белые;
- из урны последовательно вынимаются белый шар, затем черный, затем – снова белый;

Замечание. Деление событий на достоверные, случайные и невозможные происходит в рамках составленной расчетной схемы реальной ситуации. Например, в задачах о подбрасывании идеальной монеты, как правило, считается, что возможны ровно два случайных события – выпадение «орла» или «решки». В реальных условиях монета может докатиться до стены и стать возле нее на «ребро». Поверхность, на которую падает монета, может не быть идеально горизонтальной и т.п. При обработке результатов реальных экспериментов методами ТВ подобные «нерасчетные» реальные ситуации относят к невозможным и либо не рассматривают, либо усложняют расчетную схему с расширением множества видов случайных событий.

В ТВ события обычно обозначаются заглавными буквами, как правило, латинскими ($A, B, C, D, \ldots$ или $A_1, A_2, A_3, A_4, \ldots$) и описываются в виде высказываний о том, что произошло или

не произошло в результате испытаний. Описание событий можно приводить в кавычках или фигурных скобках, используют также символьные описания событий.

Пример 3. Схема испытания – многократное подбрасывание игрального кубика

$A = \{выпало\ четное\ количество\ очков\}$ - текстовое описание события;

$A = \{A(2), A(4), A(6)\} = \{2;4;6\}$ - символьные описания события.

Каждое однократное испытание заканчивается одним исходным событием (*исходом*)

Замечание. События, допускающие несколько различных вариантов исхода однократного испытания, называют *составными событиями*. Их основное отличие заключается в том, что в любом однократном испытании может реализоваться только один из возможных вариантов составного события. В примере 3 событие A является составным, оно включает три исходных события (событие $A(2) = \{выпадение\ 2 - х\ очков\}$, событие $A(4) = \{выпадение\ 4 - х\ очков\}$, $A(6) = \{выпадение\ 6 - и\ очков\}$).

События, которые не могут произойти вместе в одном и том же однократном испытании, называются *несовместными событиями*. Если различные события могут произойти в одном и том же однократном испытании, то их называют *совместными событиями*. Событие $C = \{выпадение\ не\ менее\ 4 - х\ очков\}$ совместно с событием A из примера 3.

3 Вероятность события

Вероятностью события называется количественная характеристика (уровень, степень) возможности появления события в однократном испытании при многократном повторении испытаний в неизменных условиях, по одной и той же схеме.

Определение. Вероятность достоверного события полагается равной 1; вероятность невозможного события принимают равной 0; вероятности случайных событий находятся в интервале $(0;1)$.

При математическом моделировании реальных вероятностных ситуаций используют три вида объективных вероятностей:

1) классическая вероятность;
2) статистическая вероятность;
3) геометрическая вероятность;

Кроме того, при строгом математическом обосновании ТВ вводят *аксиоматическую вероятность*, которая является обобщением указанных объективных вероятностей.

Замечание. В социальных и экономических науках, а также в психологии при принятии управленческих решений в условиях неопределенности или условиях рисков используют также *субъективные* (экспертные) вероятности, которые в классической ТВ не рассматриваются.

4 Классическая вероятность

При многократных испытаниях по одной и той же схеме с конечным количеством объектов можно выделить конечное множество различных исходных событий $S = \{S_1, S_2, ..., S_n\}$, объединенных по некоторому общему отличительному признаку.

Определение. Конечное множество исходных событий $S_1, S_2, ...S_n$ образуют *полную группу несовместных событий*, если при каждом однократном испытании обязательно происходит только одно из этих событий.

Определение. События называются *равновозможными* по некоторому общему признаку, если нет оснований полагать, что какое-либо из этих событий более возможно, чем какое-либо другое.

Замечание 1. Равновозможность событий нельзя доказать математическими методами. Она либо постулируется в явной форме, либо предполагается по умолчанию. При решении прикладных задач равновозможность событий обосновывается из соображений симметрии и однородности при формализации задачи.

Замечание 2. События полной группы могут быть равновозможными или не равновозможными.

Пример 1. Классическая схема подбрасывания идеальной монеты

События $S_1 = \{выпадение\ "орла"\}$ и $S_2 = \{выпадение\ "решки"\}$ составляют полную группу $S = \{S_1, S_2\}$.

Пример 2. Подбрасывание «толстой» монеты (допускается падение на «ребро»)

События $S_1 = \{выпадение\ "орла"\}$, $S_2 = \{выпадение\ "решки"\}$ и $S_3 = \{падение\ на\ "ребро"\}$ составляют полную группу $S = \{S_1, S_2, S_3\}$

Пример 3. Подбрасывание игрального кубика (вариант первый)

События $S_1 = \{выпадение\ "1"\}$, $S_2 = \{выпадение\ "2"\}$, $S_3 = \{выпадение\ "3"\}$, , $S_4 = \{выпадение\ "4"\}$, $S_5 = \{выпадение\ "5"\}$ и $S_6 = \{выпадение\ "6"\}$ составляют полную группу $S = \{S_1, S_2, ..., S_6\}$.

В примерах 1,3 события полной группы равновозможны если монета и кубик обладают геометрической и материальной cbмметрией (центры масс и центры симметрии совпадают). В примере 2 события полной группы не равновозможны.

Пример 4 . Подбрасывание игрального кубика (вариант второй)

События $D_1 = \{выпадение\ "1"\}$, $D_2 = \{выпадение\ четного\ числа\ очков\}$, $D_3 = \{выпадение\ \ числа\ очков, делящегося\ на\ 3\}$, , $D_4 = \{выпадение\ "5"\}$ не составляют полную группу несовместных событий, так как события D_2 и D_3 совместны.

Пусть по условию задачи определена полная группа несовместных равновозможных (по некоторому общему признаку) событий. Из этих событий выделим события $A_1, A_2, ...A_k$ по другому общему (особому, дополнительному) признаку. Эти события можно рассматривать как возможные варианты составного события $A = \{A_1, A_2, ...A_k\}$. События $A_1, A_2, ...A_k$ называют событиями, благоприятными для события A.

Определение. Событие A_i называется *благоприятным* для события A, если наступление события A_i влечет за собой наступление события A.

Пример 5. В рассмотренных выше примерах 3 и 4 событие $S_2 = \{выпадение\ "2"\}$ благоприятно для события $D_2 = \{выпадение\ четного\ числа\ очков\}$; каждое из событий $S_4 = \{выпадение "4"\}$ и $S_6 = \{выпадение\ "6"\}$ также благоприятно для события $D_2 = \{выпадение\ четного\ числа\ очков\}$. Каждое из событий $S_3 = \{выпадение\ "3"\}$ и $S_6 = \{выпадение\ "6"\}$ благоприятно для события $D_3 = \{выпадение\ числа\ очков, делящегося\ на\ 3\}$.

Пусть по условию задачи определена полная группа несовместных равновозможных событий S, среди которых события $A_1, A_2, ...A_k$ являются благоприятными для некоторого события A, вероятность которого нас интересует.

Тогда классическая вероятность события A определяется по формуле

$$P(A) = \frac{k}{n}, \qquad (1)$$

где n - количество событий в рассматриваемой полной группе, k - количество событий из полной группы, благоприятных для события A.

Замечание 1. Очевидно, что $0 \le \frac{k}{n} \le 1$, что соответствует данному выше определению вероятности.

Замечание 2. Формула классической вероятности применяется только при наличии полной группы несовместных равновозможных исходных событий, которые называют также *элементарными событиями* (т.е. неделимыми в рамках данной схемы испытаний).

Определение. Пусть дано некоторое событие A. Событие $\overline{A}$, состоящее в том, что событие A не произошло, называется событием, *противоположным* к событию A.

Полную группу несовместных событий можно представить в виде совокупности двух непересекающихся подмножеств (составных событий) $\{A, \overline{A}\}$, где событие A называют прямым событием, а событие $\overline{A}$ является противоположным событием. Если в подмножество A входят все события $A_{i_1}, A_{i_2}, ...A_{i_k}$, благоприятные для события B и никакие другие, то, очевидно, что $A = B$. В этом случае общее количество событий в полной группе (n) будет равно сумме вариантов прямого события (k) и вариантов противоположного события $(m = n - k)$: $n = k + m$.

Тогда вероятность события A можно определить так:

$$P(A) = \frac{k}{n} = \frac{n-m}{n} = 1 - \frac{m}{n} = 1 - P(\overline{A}), \qquad (2)$$

где $P(\overline{A})$ - вероятность появления противоположного события $\overline{A}$. При этом из (2) следует $P(A)+P(\overline{A})=1$.

1.2 **Вычисление классической вероятности с непосредственным подсчетом исходных событий**

Простейшим (но нередко – очень громоздким) способом вычисления вероятности некоторого события A по формуле (1) является непосредственный подсчет всех возможных различных исходных событий, образующих полную группу и всех исходных событий из полной группы, благоприятных для события A.

Решение задач с вычислением классической вероятности целесообразно проводить по следующей схеме:

1. Определить в чем состоит однократное испытание и что является исходом (исходным событием) однократного испытания;
2. Определить равновозможные несовместные события, образующие полную группу по общему отличительному признаку, выбранному по условиям задачи;
3. Отделить в полной группе подмножество событий, благоприятных для заданного события A (обладающих общим дополнительным признаком);
4. Определить общее количество событий в полной группе (n) и количество событий k, благоприятных для заданного события A;
5. Найти вероятность события A как отношение $\frac{k}{n}$;

Замечание 1. Некоторые этапы решения могут выполняться в уме («по умолчанию»), если они следуют из условий задачи очевидным образом. В текстовое описание события необходимо включать общий отличительный (ограничительный) признак, по которому события объединены в полную группу S или являются событиями, благоприятными для заданного события A.

При непосредственном подсчете исходные события считают *элементарными* в том смысле, что их принимают за единицы счета (элементы) при однократном испытании.

Пример 1. Производится однократное подбрасывание четырех монет. Найти вероятность того, что из четырех монет ровно на трех монетах выпадет «орел».

Решение. Пусть A - событие, состоящее в том, что ровно на трех монетах из четырех выпадет «орел».

Однократное испытание – одновременное подбрасывание четырех монет. Исходом однократного испытания (элементарным событием) можно считать упорядоченный набор вида ООРР, ОРРР, ОООО и т.п., где первая буква означает выпадение «орла» (О) или «решки» (Р) на

первой монете, вторая буква – выпадение «орла» или «решки» на второй монете, третья – на третьей монете, четвертая – на четвертой.

Равновозможные несовместные события, образующие полную группу – всевозможные упорядоченные наборы вида ООРР, ОРРР, ОООО и т.п.(общий признак событий полной группы),

События, благоприятные для события A - это упорядоченные наборы букв «О» и «Р», в которых ровно три буквы «О» (дополнительный общий признак для события A).

Для подсчета общего количества событий в полной группе и благоприятных событий составим таблицу.

4 «орла»	3 «орла»	2 «орла»	1 «орел»	0 «орлов»
ОООО	ОООР ООРО ОРОО РООО	ООРР ОРРО РРОО РОРО ОРОР РООР	РРРО РРОР РОРР ОРРР	РРРР
$n_4 = 1$	$n_3 = 4$	$n_2 = 6$	$n_1 = 4$	$n_0 = 1$

Общее количество всех элементарных событий в полной группе равно $n = 1+4+6+4+1 = 16$, количество событий, благоприятных событию A равно $n_3 = 4$.

Вероятность события A равна $P(A) = \frac{4}{16} = \frac{1}{4} = 0{,}25$.

Пример 2. Найти вероятность того, что при одновременном подбрасывании двух игральных кубиков на их верхних гранях выпадет:

- в сумме не менее 9 очков (событие A);
- одинаковое количество очков (событие B).

Решение. Однократное испытание – одновременное подбрасывание двух кубиков. Исходом однократного испытания можно считать двузначное число, первая цифра которого (от 1 до 6) равна числу очков, выпавших на первом кубике, вторая - числу очков, выпавших на втором кубике.

Равновозможные несовместные события, образующие полную группу – всевозможные двузначные числа, цифрами которых являются 1, 2, 3, 4, 5, 6 (общий признак событий полной группы).

События, благоприятные событию A - двузначные числа, цифрами которых являются 1, 2, 3, 4, 5, 6 и при этом сумма цифр не менее 9 (дополнительный общий признак для составного события A); события, благоприятные событию B - двузначные числа, цифрами которых являются 1, 2, 3, 4, 5, 6 и при этом обе цифры одинаковы.

Для подсчета общего количества событий в полной группе и благоприятных событий составим таблицу.

11	12	13	14	15	16
21	22	23	24	25	26
31	32	33	34	35	36

41	42	43	44	45	46
51	52	53	54	55	56
61	62	63	64	65	66

Из таблицы следует, что общее число событий в полной группе равно $n = 6 \cdot 6 = 36$. Число событий, благоприятных событию A равно $k_A = 10$ (выделены серым цветом), число событий, благоприятных событию B равно $k_B = 6$ (числа, расположенные в диагональных клетках таблицы). $P(A) = \frac{10}{36} = \frac{5}{18}$, $P(B) = \frac{6}{36} = \frac{1}{6}$.

Рассмотрим решение задачи с использованием противоположных событий.

Пример 3. Найти вероятность того, что из 100 жетонов с номерами от 1 до 100 случайно взят жетон с номером, не содержащим цифру 5.

Решение. Пусть A - событие, состоящее в том, что случайно взят жетон с номером, не содержащим цифру 5. Тогда противоположное событие $\overline{A}$ состоит в том, что случайно взят жетон с номером, содержащим цифру 5.

Однократное испытание – случайный выбор одного жетона из ста. Отличительным признаком естественно считать номер выбранного жетона, исходным событием – выбор одного жетона с каким-либо номером.

Равновозможные несовместные события, образующие полную группу – целые положительные числа от 1 до 100.

Номеров жетонов, содержащих цифру 5 меньше, чем номеров, не содержащих эту цифру; поэтому посчитаем количество номеров, содержащих цифру 5: 5, 15, 25, 35, 45, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 65, 75, 85, 95 – всего 19 номеров.

$$P(A) = 1 - P(\overline{A}) = 1 - \frac{19}{100} = \frac{81}{100} = 0{,}81$$

1.3 **Способ вычисления вероятностей событий с помощью подсчета объектов задачи (урновая схема испытаний)**

Конечное множество всех объектов задачи, обладающих некоторым общим отличительным признаком, будем называть *генеральной совокупностью* объектов. Пусть среди объектов генеральной совокупности имеются объекты, обладающие некоторым интересующим нас особым признаком.

Пусть условия задачи позволяют считать, что между объектами генеральной совокупности и полной группой несовместных равновозможных событий, соответствующих извлечению этих объектов из генеральной совокупности по одному, установлено взаимно однозначное соответствие. Тогда вероятность извлечения одного объекта с особым признаком из генеральной

совокупности можно определить по формуле $P(A)=\frac{M}{N}$, где N - общее количество объектов в генеральной совокупности, M - общее количество объектов с особым признаком в той же генеральной совокупности.

Пример 1. В урне находятся a зеленых и b красных шаров, неразличимых на ощупь. Из урны извлекается красный шар и не возвращается в урну. Затем вынимают второй шар. Найти вероятность того, что этот шар красный.

Решение. Пусть A - событие, состоящее в том, что второй вынутый шар – красный.

После извлечения первого (красного) шара количество элементов генеральной совокупности составит $N=a+b-1$, а количество элементов с особым признаком (красный цвет шара) составит $M=a-1$. Тогда $P(A)=\frac{M}{N}=\frac{a-1}{a+b-1}$.

Пример 2. Из букв слова ДИФФЕРЕНЦИАЛ наугад выбрана одна буква. Найти вероятности следующих событий:

- событие A: выбрана буква Е;

- событие B: выбрана согласная буква;

- событие C: выбрана буква Ы;

Решение: В генеральной совокупности (в слове ДИФФЕРЕНЦИАЛ) всего 12 букв. Среди них: букв Е – две $(M_A=2)$, согласных букв – 7 $(M_B=7)$, букв Ы – 0 $(M_C=0)$.

$P(A)=\frac{M_A}{N}=\frac{2}{12}=\frac{1}{6}$; $P(B)=\frac{M_B}{N}=\frac{7}{12}$; $P(C)=\frac{M_C}{N}=\frac{0}{12}=0$.

1.4 **Определение геометрической вероятности**

При использовании геометрической вероятности предполагается следующая схема мысленных экспериментов (испытаний): в некоторую ограниченную область G пространства (поверхности или линии) случайным образом бросается точка, обязательно попадающая в область G. При этом попадание точки в любую часть $G_0 \subset G$ зависит только от объема (площади или длины) области G_0 и не зависит от того, в какой части области G расположена область G_0. Область G в рассматриваемых задачах является аналогом полной группы несовместных

равновозможных событий. Внутри области G выделяется некоторая подобласть $G_0(A) \subset G$, точки которой являются аналогом событий, благоприятных для рассматриваемого события A.

Геометрической вероятностью события A называется отношение меры (объема, площади, длины) области $G_0(A)$ к мере (объему, площади, длине) области G (если эта мера существует и конечна): $P(A) = \dfrac{V(G_0)}{V(G)}$ или $P(A) = \dfrac{S(G_0)}{S(G)}$ или $P(A) = \dfrac{l(G_0)}{l(G)}$.

При вычислении геометрических вероятностей случайным событиям ставят во взаимно однозначное соответствие случайные величины, которые рассматривают как случайные координаты точек в областях G и $G_0(A)$.

Если исходному случайному событию соответствует одна случайная координата, выражаемая одним действительным числом, то используют одномерную схему испытаний со случайным бросанием точки на отрезок прямой. Если случайному событию соответствуют две случайные координаты, выражаемые парой действительных чисел, то принимают двумерную схему испытаний со случайным бросанием двух точек на отрезок прямой или случайным бросанием точки на плоскую фигуру.

Пример 1. На участке AB длиной L электрический кабель имеет точечный дефект. Найти вероятность того, что дефект удален от точки A на расстояние не меньше l $(l < L)$.

Решение. Пусть событие D состоит в том, что дефект удален от точки A на расстояние не меньше l. Используем одномерную схему испытаний.

«Областью» G является отрезок AB длины L, а «благоприятной» областью $G_0(D)$ - отрезок CB (Рис. 25.3).

Вероятность события D равна $P(D) = \dfrac{L-l}{L} = 1 - \dfrac{l}{L}$.

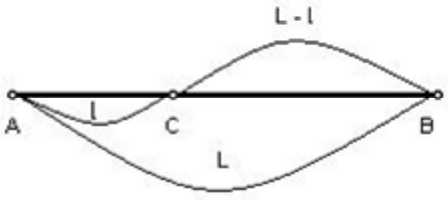


Рис 25.3 Расчетная схема к Примеру 1

Пример 2. (Задача о случайной встрече) Два теплохода должны подойти к одному и тому же причалу. Время прибытия каждого из теплоходов независимо и равновозможно в течение суток. Время стоянки первого теплохода у причала составляет 1 час, а второго – 2 часа. Определить вероятность того, что одному из теплоходов придется ждать освобождения причала.

Решение. Пусть A - событие, состоящее в том, что одному из теплоходов придется ждать освобождения причала. Используем двумерную схему испытаний.

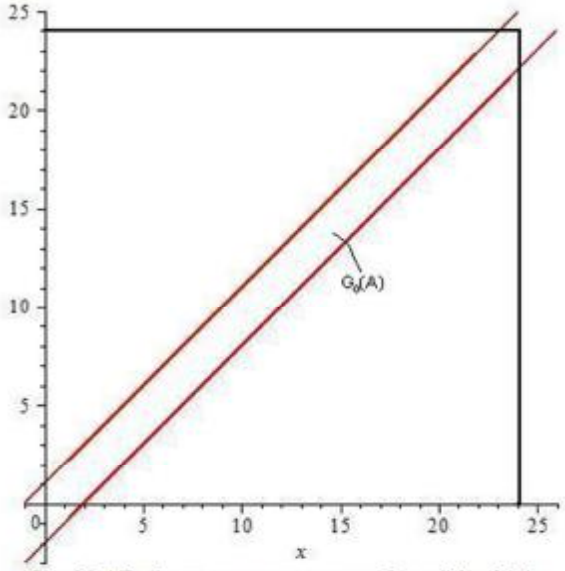


Рис. 25. 4 Графическое представление областей G и G(A) к Примеру 2

Событиям полной группы соответствует множество случайных точек (x, y) на плоскости, где x - координата, соответствующая времени подхода к причалу первого теплохода, y - координата, соответствующая времени подхода к причалу второго теплохода. По условию задачи $x \in [0; 24]$, $y \in [0; 24]$, поэтому область G (полная группа случайных точек на плоскости) будет представлять собой квадрат $0 \le x \le 24$, $0 \le y \le 24$ со стороной 24 (Рис. 25.4)

Событие A разобьем на два подмножества событий $A = \{A_1, A_2\}$, где событие A_1 состоит в том, что первый теплоход стоит у причала, а второй ожидает на рейде, а событие A_2 состоит в том, что второй теплоход стоит у причала, а первый ожидает на рейде.

По условию задачи событие A_1 произойдет, если $y - x \le 1$ (час.), а событие A_2 произойдет, если $x - y \le 2$. Границы области $G_0(A)$, соответствующей событию A, находим переходом от неравенств к равенствам, тогда, перейдя в неравенстве $y - x \le 1$ к равенству $y - x = 1$ получим уравнение прямой. Точки, лежащие ниже этой прямой, удовлетворяют условию $y - x \le 1$. Аналогично, точки, лежащие выше прямой $x - y = 2$ удовлетворяют условию $x - y \le 2$. Эти прямые и область $G_0(A)$ показаны на Рис.25.4.

Вероятность того, что один из теплоходов будет ожидать на рейде, определится так:

$$P(A) = \frac{S(G_0(A))}{S(G)} = \frac{24 \cdot 24 - 0{,}5 \cdot (23 \cdot 23 + 22 \cdot 22)}{24 \cdot 24} = 0{,}12 .$$

Пример 3. Найти вероятность того, что из трех отрезков, длины которых не превышают 1 см. можно составить треугольник.

Решение. Пусть отрезки имеют длины $x,\ y,\ z$, тогда $x \in (0;\ 1),\ y \in (0;\ 1),\ z \in (0;\ 1)$. Событиям будут соответствовать точки $(x,\ y,\ z)$ в пространстве, при этом область G будет представлять собой куб со стороной 1 (Рис 25.5). Для того, чтобы из трех отрезков $x,\ y,\ z$ можно было составить треугольник, необходимо и достаточно выполнение трех неравенств: $x + y > z$, $x + z > y$, $z + y > x$. Уравнение $x + y - z = 0$, соответствующее неравенству $x + y > z$ задает в пространстве плоскость, проходящую через начало координат и точки $A'(1;\ 0,\ 1)$, $C'(0;\ 1,\ 1)$ (Рис. 25.5). Для того, чтобы понять какое полупространство, ограниченное этой плоскостью, удовлетворяет неравенству $x + y > z$, достаточно подставить в это неравенство координаты любой точки, не лежащей на этой плоскости. Подставим в неравенство координаты точки $B(1;\ 1,\ 0)$, расположенной ниже плоскости $x + y - z = 0$: $1 + 1 > 0$. Получили верное числовое неравенство, значит, неравенству $x + y > z$ соответствуют точки, лежащие ниже плоскости $x + y - z = 0$. Аналогично, рассматривая уравнение $z + y - x = 0$, соответствующее неравенству $z + y > x$, получим плоскость, OBA' (Рис. 25.5).Подставляя в неравенство $z + y > x$ координаты точки $C'(0;\ 1,\ 1)$, лежащей выше рассматриваемой плоскости, получим верное числовое неравенство $1 + 1 > 0$. И, наконец, рассматривая уравнение $x + z - y = 0$, соответствующее неравенству $x + z > y$, получим плоскость OBC'. Подставляя в неравенство $x + z > y$ координаты точки $A'(1;\ 0,\ 1)$, получим верное числовое неравенство.

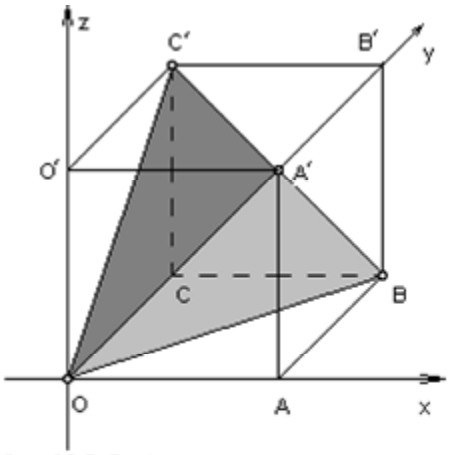


Рис. 25.5. Графическое представление областей к Примеру 3

Все три неравенства $x + y > z$, $x + z > y$ и $z + y > x$ выполняются в той части куба, которая ограничена плоскостями $x + y - z = 0$, $z + y - x = 0$ и $x + z - y = 0$, то есть в объединении трех треугольных пирамид $OBA'C'$ и $B'BA'C'$.

Искомая вероятность будет равна отношению суммы объемов этих пирамид к объему куба. Для вычисления объема пирамиды $OBA'C'$ найдем координаты векторов, $\vec{a} = \overrightarrow{OA'}$, $\vec{b} = \overrightarrow{OB}$, $\vec{c} = \overrightarrow{OC'}$: $\vec{a}(1;\ 0;\ 1)$, $\vec{b}(1;\ 1;\ 0)$, $\vec{c}(0;\ 1;\ 1)$. Объем пирамиды, ограниченной тремя заданными

векторами равен одной шестой модуля их смешанного произведения (Практикум, часть 1, с. 133, задача 3.6.16). $V_{OBA'C'} = \frac{1}{6}\begin{vmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \end{vmatrix} = \frac{1}{3}$. Объем пирамиды $B'BA'C'$ можно найти методами элементарной геометрии: $V_{B'BA'C'} = \frac{1}{3} \cdot S_{A'B'C'} \cdot \left|BB'\right| = \frac{1}{3} \cdot \frac{1}{2} \cdot 1 \cdot 1 \cdot 1 = \frac{1}{6}$.

$$P(A) = \frac{V_{OBA'C'} + V_{B'BA'C'}}{1 \cdot 1 \cdot 1} = \frac{1}{2}.$$

Тема 2: Вычисление вероятностей событий с помощью методов комбинаторики

Возможности непосредственного подсчета при вычислении вероятностей событий ограничены и во многих случаях практически неосуществимы при большом количестве вариантов событий. При решении подобных задач используют методы комбинаторики.

Комбинаторика – раздел дискретной математики, в котором изучаются методы решения задач с использованием различных соединений (наборов, комбинаций) элементов конечных множеств.

При формализации прикладных задач ТВ принимают, что между событиями или объектами задачи и элементами конечного множества имеет место взаимно однозначное соответствие. При этом различные варианты выбора любого события или объекта из генеральной совокупности объектов и соответствующих элементов из конечного множества считаются равновозможными.

2.1 Правила комбинаторики и их применение при вычислении вероятностей событий

Пусть из условий задачи следует, что имеет место взаимно однозначное соответствие между числом исходных событий и числом способов выбора объектов, соответствующих этим событиям. Тогда число исходных событий можно определить с помощью комбинаторных *правила сложения* и *правила умножения*.

Пусть даны два непересекающихся конечных множества E_1 и E_2. Тогда, если выбор элемента $A \in E_1$ можно осуществить m_A способами, а выбор элемента $B \in E_2$ - m_B способами, то:

- выбор элементов $A \in E_1$ и $B \in E_2$ можно осуществить $m_A \cdot m_B$ способами (*правило умножения*);

- выбор элемента $A \in E_1$ или $B \in E_2$ можно осуществить $m_A + m_B$ способами (*правило сложения*);

Замечание 1. Правила сложения и умножения можно распространить на выбор элементов из любого количества множеств. Пусть даны непересекающиеся конечные множества $E_1, E_2, ..., E_k$. Если выбор элемента $A_i \in E_i$, $(i = 1, 2, ..., k)$ можно осуществить m_i способами, то:

- выбор элементов $A_1 \in E_1$ и $A_2 \in E_2$ и… $A_k \in E_k$ можно осуществить $m_1 \cdot m_2 \cdot ... \cdot m_k$ способами (*правило умножения*);

- выбор элемента $A_1 \in E_1$ или $A_2 \in E_2$ или… $A_k \in E_k$ можно осуществить $m_1 + m_2 + ... + m_k$ способами (*правило сложения*);

Замечание 2. Правило умножения можно также распространить на случай одного множества следующим образом: после выбора одного элемента из множества E_1 одним из n_1 способов, его можно считать другим множеством E_2, элементы которого можно выбрать $n_2 = n_1 - 1$ способами и т.д.

Пример 1. Сколько трехзначных чисел можно составить из цифр 1, 2, 3, 4, 5, если цифры не повторяются?

Решение. Первую цифру можно выбрать пятью способами, вторую – четырьмя, третью – тремя. По правилу умножения общее количество трехзначных чисел с разными цифрами равно $5 \cdot 4 \cdot 3 = 60$.

Пример 2. Код замка состоит из трех символов: одной буквы из 26 и двух цифр от 0 до 9. Найти вероятность открыть замок при случайном наборе кода.

Решение. Пусть событие A - открытие замка при случайном наборе кода.

Правильный набор каждого из трех символов кода – единственное благоприятное событие, поэтому $k = 1$. Символы кода выбирают по одному из трех множеств $E_1 = \{a, b, c, ...x, y, z\}$, $E_2 = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$, $E_3 = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$. Количество различных кодов равно количеству различных способов выбора по одному элементу из указанных трех множеств. По правилу произведения получаем $n = 26 \cdot 10 \cdot 10 = 2600$.

Вероятность открыть замок при случайном наборе кода равна $P(A) = \dfrac{1}{2600} \approx 3{,}85 \cdot 10^{-4}$

Пример 3. Имеется 20 изделий первого сорта и 30 изделий второго сорта. Определить число способов выбора изделий одного сорта, если учитывается порядок выбора изделий.

Решение. Имеем два множества: $E_1 = \{20\ \textit{изделий первого сорта}\}$ и $E_2 = \{30\ \textit{изделий второго сорта}\}$.

Число способов выбора одного изделия первого сорта равно 20. Если одно изделие первого сорта уже выбрано, то число способов выбора второго изделия первого сорта равно 19. Число способов упорядоченного выбора двух изделий первого сорта находим по правилу умножения: $n_1 = 20 \cdot 19 = 380$.

Аналогично, число способов выбора одного изделия второго сорта равно 30. Если одно изделие второго сорта уже выбрано, то число способов выбора второго изделия второго сорта равно 29. Число способов упорядоченного выбора двух изделий второго сорта находим по правилу умножения: $n_2 = 30 \cdot 29 = 870$.

Число способов выбора изделий одного сорта (первого или второго) находим по правилу сложения: $n = n_1 + n_2 = 380 + 870 = 1250$.

Пример 4. На первом этаже семиэтажного дома в лифт вошли три человека. Вероятность выхода на любом этаже одинакова (кроме первого, где никто не будет выходить). Найти вероятности событий:

- событие A - все вышли на четвертом этаже;

- событие B - все вышли на одном и том же этаже;

- событие C - все вышли на разных этажах;

Решение. Полная группа несовместных равновозможных событий состоит из событий вида S_i - «выход каждого из трех человек независимо друг от друга на любом из этажей со второго по седьмой». Будем учитывать не число событий, а число способов осуществить события, так как между этими событиями и способами имеет место взаимно однозначное соответствие.

Учитывая, что каждый человек может выйти на любом из шести этажей, находим, используя правило умножения, общее число способов выхода трех пассажиров: $n = 6 \cdot 6 \cdot 6 = 216$.

Для события A количество благоприятных способов также находим по правилу умножения: $k_A = 1 \cdot 1 \cdot 1 = 1$. Тогда $P(A) = \dfrac{1}{216} \approx 0{,}0046$.

Для события B количество благоприятных способов также находим по правилу умножения: $k_B = 6 \cdot 1 \cdot 1 = 6$. Тогда $P(A) = \frac{6}{216} = \frac{1}{36} \approx 0{,}028$.

Для события C количество благоприятных способов также находим по правилу умножения: $k_c = 6 \cdot 5 \cdot 4 = 120$. Тогда $P(A) = \frac{120}{216} = \frac{5}{9} \approx 0{,}56$.

Замечание 3. Исходные правила сложения и умножения сформулированы для выбора отдельных элементов из непересекающихся множеств. При вычислении вероятности наборов объектов из разных подмножеств используются обобщенные правила сложения и умножения.

Пусть все элементы (объекты испытаний конечного множества) можно разбить на k :непересекающихся групп (подмножеств). При этом наборы элементов из первой группы можно осуществить m_1 способами, наборы элементов из второй группы - m_2 способами и так далее для всех k групп. Тогда:

- общее количество наборов элементов из первой, второй и т. д. до k-ой группы можно осуществить $m_1 \cdot m_2 \cdot ... \cdot m_k$ способами (правило умножения);

- общее количество наборов элементов из объединения всех групп можно осуществить $m_1 + m_2 + ... + m_k$ способами (правило сложения).

Обобщенные правила умножения и сложения обычно используют совместно с формулами комбинаторики. Поэтому примеры использования указанных правил приведены в разделах 26.3 и 26.4.

2.1 **Основные формулы комбинаторики**

Кроме правил комбинаторики в ТВ широко используют формулы комбинаторики, которые позволяют определить количество различных видов соединений (наборов, комбинаций) из элементов конечного множества, объединенных по некоторому общему отличительному признаку.

Различают следующие основные виды соединений: сочетания, размещения и перестановки. Вычисления количества соединений, содержащих повторяющиеся и неповторяющиеся элементы, производят по разным формулам.

2.1.1 **Соединения с различными элементами без повторений**

1. Сочетания

Сочетаниями без повторений называют неупорядоченные наборы различных элементов из конечного множества.

Если конечное множество содержит n различных элементов, то число различных сочетаний, каждое из которых содержит k элементов $(k \le n)$ определяют по формуле

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

Символ C_n^k читается «це из эн (элементов) по ка (элементов)».

Замечание. Факториалом $n!$ натурального числа n называется натуральное число, полученное перемножением последовательных натуральных чисел от 1 до n: $n! = 1 \cdot 2 \cdot \ldots \cdot n$. Отдельно определяется факториал ноля (ноль не является натуральным числом): $0! = 1$.

Примеры: $5! = 1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 = 120$; $10! = 1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 \cdot 6 \cdot 7 \cdot 8 \cdot 9 \cdot 10 = 3628800$;

Свойства факториала: $1! = 1$; $(n+1)! = (n+1) \cdot n!$; $\frac{n!}{k!} = n \cdot (n-1) \cdot \ldots \cdot (n-k+1)$

Свойства сочетаний: $C_n^0 = C_n^n = 1$, $C_n^1 = n$, $C_n^k = C_n^{n-k}$.

2. Размещения

Размещениями без повторений называют упорядоченные наборы различных элементов из конечного множества.

Если конечное множество содержит n различных элементов, то число различных размещений, каждое из которых содержит k элементов $(k \le n)$ определяют по формуле

$$A_n^k = \frac{n!}{(n-k)!} \qquad (2)$$

Символ C_n^k читается «а из эн (элементов)по ка (элементов)».

Свойства размещений: $A_n^0 = 1$, $A_n^1 = n$, $A_n^n = n!$.

3. Перестановки

Перестановками без повторений называют упорядоченные наборы всех элементов некоторого конечного множества.

Если конечное множество содержит n различных элементов, то число его различных перестановок вычисляют по формуле

$$P_n = n! \qquad (3)$$

Замечание. Перестановки являются, по существу, размещениями без повторений при $n = k$: $P_n = A_n^n$.

Рассмотрим как формируются различные виды соединений из конечного множества с различными элементами и как определяется общее число таких соединений на конкретных примерах.

Пример 1. Дано множество из четырех элементов, пронумерованных от 1 до 4. Определить:

- число размещений по два элемента;

- число сочетаний по два элемента;

- число перестановок элементов множества.

Выписать все наборы, соответствующие размещениям, сочетаниям и перестановкам.

Решение. Общее количество размещений из четырех элементов по два находим по формуле $A_4^2 = \frac{4!}{(4-2)!} = \frac{2!\cdot 3\cdot 4}{2!} = 12$. Выпишем все размещения: 12, 13, 14, 21, 23, 24, 31, 32, 34, 41, 42, 43.

Общее количество сочетаний: $C_4^2 = \frac{4!}{2!\cdot(4-2)!} = \frac{2!\cdot 3\cdot 4}{2!\cdot 2!} = 6$. Выпишем все сочетания: 12, 13, 14, 23, 24, 34.

Общее количество перестановок: $4! = 1\cdot 2\cdot 3\cdot 4 = 24$. Выпишем все перестановки: 1234, 1243, 1324, 1342, 1423, 1432, 2134, 2143, 2314, 2341, 2413, 2431, 3124, 3142, 3214, 3241, 3412, 3421, 4123, 4132, 4213, 4231, 4312, 4321.

Замечание: чтобы не пропустить ни одну перестановку, они выписывались по возрастанию как четырехзначные числа с разными цифрами; по тому же принципу выписывались размещения и сочетания.

Пример 2. Определить количество способов выбора трех инженеров из десяти:

- на три различные должности;

- на три одинаковые должности.

Решение. При выборе на три различные должности речь идет об упорядоченной выборке без возвращений, т. е о размещении (важен состав и распределение по должностям и нельзя одного инженера назначить на несколько должностей). $A_{10}^3 = \frac{10!}{(10-3)!} = \frac{7!\cdot 8\cdot 9\cdot 10}{7!} = 720$.

При выборе на три одинаковые должности речь идет о неупорядоченной выборке без возвращений, т. е о сочетании (важен состав и не важно распределение по должностям, при этом

нельзя одного инженера назначить на несколько должностей).

$$C_{10}^{3}=\frac{10!}{3!\cdot(10-3)!}=\frac{7!\cdot 8\cdot 9\cdot 10}{7!\cdot 3!}=\frac{720}{6}=120.$$

2.1.2 **Соединения с повторениями элементов**

1. Сочетания с повторениями

Сочетание из n элементов по k элементов с повторениями – это неупорядоченный набор k элементов из множества, содержащего n элементов, причем, среди элементов этого набора могут встречаться и одинаковые (повторяющиеся) элементы.

Если конечное множество содержит n различных элементов, то число различных сочетаний с повторениями, каждое из которых содержит k элементов определяют по формуле

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

Замечание. В сочетаниях с повторениями число k может быть больше числа n.

Пример 5. Найти общее количество способов выбора шести тетрадей, если в киоске продаются четыре различных вида тетрадей.

Решение. Проведем анализ и формализацию прикладной задачи, чтобы привести ее к комбинаторной математической задаче.

Покупатель может выбрать и одинаковые виды тетрадей, при этом порядок выбора (какую тетрадь он выберет первой, какую второй и т.д.) неважен. Таким образом, прикладная задача приводится к математической задаче определения количества сочетаний с повторениями из $n=4$ элементов по $k=6$ элементов.

Воспользуемся формулой $C_{(n)}^{k}=C_{n+k-1}^{k}=\frac{(n+k-1)!}{k!(n-1)!}$ при $n=4$ и $k=6$:

$$C_{(4)}^{6}=\frac{(4+6-1)!}{6!(4-1)!}=\frac{9!}{6!\cdot 3!}=\frac{6!\cdot 7\cdot 8\cdot 9}{6!\cdot 2\cdot 3}=7\cdot 4\cdot 3=84.$$

2. Размещения с повторениями

Размещениями с повторениями называют упорядоченные наборы элементов из конечного множества, причем, среди элементов этого набора могут встречаться и одинаковые.

Если конечное множество содержит n различных элементов, то число различных размещений, с повторениями каждое из которых содержит k элементов определяют по формуле

$$A_{(n)}^{k} = n^{k}$$

Пример 4. Автомобильный номер содержит шесть символов: три буквы из 30 и три цифры из 10. Определить общее количество различных номеров.

Решение. При формализации прикладной задачи предполагаем, что между общим количеством различных наборов (номеров) и общим количеством способов выбора набора из трех букв из 30 и трех цифр из 10 имеет место взаимно однозначное соответствие.

Указанные наборы из букв и цифр являются упорядоченными и содержат элементы (цифры и буквы), которые выбираются из двух различных конечных множеств (множества букв и множества цифр) и могут повторяться. Отсюда следует, что каждый набор является размещением с повторением и общее количество таких наборов необходимо находить с использованием обобщенного правила умножения.

Общее количество указанных наборов (автомобильных номеров) определяется так:

$m = m_{Б} \cdot m_{Ц} = A_{(30)}^{3} \cdot A_{(10)}^{3} = 30^{3} \cdot 10^{3} = 27 \cdot 10^{6} - 27000000$ (двадцать семь миллионов).

3. Перестановки с повторениями

Пусть конечное множество содержит n_1 неразличимых между собой элементов первого типа, n_2 неразличимых между собой элементов второго типа и т. д., n_k неразличимых между собой элементов k-го типа. Тогда общее число перестановок такого множества равно

$$P(n_1, n_2, ..., n_k) = \frac{(n_1 + n_2 + ... + n_k)!}{n_1! \cdot n_2! \cdot ... n_k!}$$

Пример 5. Сколько различных семизначных чисел можно составить из цифр 2, 2, 2, 4, 4, 6, 6 (каждая из написанных цифр берется ровно по одному разу)?

Решение. По условиям задачи все семь заданных цифр присутствуют в каждом семизначном числе, которое можно интерпретировать как упорядоченный набор из семи заданных цифр (перестановка с повторениями).

По условиям задачи количество «двоек» равно $n_1 = 3$, количество «четверок» - $n_2 = 2$ и количество «шестерок» - $n_3 = 2$. Тогда общее количество семизначных чисел равно

$$P(n_1, n_2, n_3) = \frac{(3+2+2)!}{3! \cdot 2! \cdot ... 2!} = \frac{7!}{3! \cdot 2! \cdot 2!} = \frac{4 \cdot 5 \cdot 6 \cdot 7}{2 \cdot 2} = 210$$

Пример 6. Определить количество способов разбиения группы из 26 студентов на три подгруппы по 7, 9 и 10 человек.

Решение. Общее количество разбиений равно $P(26;7,9,10)=\frac{26!}{7!\cdot 9!\cdot 10!}=\frac{11\cdot 12\cdot 13\cdot ...25\cdot 26}{(2\cdot 3\cdot 4\cdot 5\cdot 6\cdot 7)\cdot(2\cdot 3\cdot ...\cdot 9)}=486129987200$ (486 миллиардов 129 миллионов 987 тысяч двести).

При решении задач по теории вероятностей с помощью формул комбинаторики удобно использовать таблицу

Виды соединений	Количество соединений (наборов) объектов задачи	
	Без повторений	С повторениями
Сочетания (неупорядоченные наборы)	$C_n^k=\frac{n!}{k!(n-k)!}$	$C_{(n)}^k=\frac{(n+k-1)!}{k!(n-1)!}$
Размещения (упорядоченные наборы)	$A_n^k=\frac{n!}{(n-k)!}$	$A_{(n)}^k=n^k$
Перестановки	$P_n=n!$	$P(n_1,n_2,...,n_k)=\frac{(n_1+n_2+...+n_k)!}{n_1!\cdot n_2!\cdot ...n_k!}$

2.3 **Вычисление вероятностей событий с помощью правил и формул комбинаторики**

При решении вероятностных задач с помощью методов комбинаторики, ее правила и формулы часто используют совместно.

Комбинаторика оперирует элементами из конечного множества и числом способов выбора наборов элементов из указанного множества (см. 26.1). Поэтому при формализации прикладных задач используются обобщенные правила сложения и умножения (см. 26.1., Замечание 3) и способы вычисления упорядоченных и неупорядоченных наборов объектов задачи (см. 26.2).

Пример 1. На полке в один ряд расставлены восемь книг, среди которых имеется ровно три книги по математике. Найти вероятность того, что книги по математике стоят на полке рядом.

Решение. Обозначим $A=\{книги\ по\ математике\ стоят\ на\ полке\ рядом\}$ - событие, вероятность которого нужно найти. Событиями полной группы будут различные перестановки из 8 книг. Их количество равно $n_S=n!=8!$. По условию задачи книги по математике должны стоять рядом, но могут быть расположены в любом месте среди пяти остальных книг. Чтобы обеспечить

это условие, соединим книги по математике в один «трехтомник», который будем условно считать одной «книгой». В результате получим 6 «книг», которые можно переставить $k_1 = 6!$ различными способами. Кроме того, книги «трехтомника» можно переставлять между собой $k_2 = 3!$ способами. Общее количество способов перестановки книг с трехтомником» находим с помощью обобщенного правила умножения $k_A = k_1 k_2 = 6! \cdot 3!$

Общее количество способов перестановки книг с трехтомником» совпадает с количеством всех перестановок, соответствующих благоприятному событию A. Вероятность этого события определяем так: $P(A) = \frac{k_A}{n_S} = \frac{6! \cdot 3!}{8!} = \frac{6}{56} = \frac{3}{28}$.

Пример 2. Из четырех одинаковых карточек с буквами К, Р, И, Т после перемешивания выбирают поочередно случайным образом три карточки и располагают в ряд. Найти вероятность того, что получится слово «ТИР».

Решение. Пусть событие $A = \{получилось\ слово\ "ТИР"\}$.В конечном множестве $\{К, Р, И, Т\}$ все буквы различны, поэтому число исходов, благоприятствующих событию A, равно $k_A = 1$. Общее число всех исходных событий в полной группе – это количество всевозможных наборов (трехбуквенных «слов») из четырех букв множества. Такие наборы являются размещениями, так как в «слове» важен порядок букв. Значит, при $k = 3$ и $n = 4$ число событий в полной группе определяется так: $n_S = A_n^k = A_4^3 = \frac{4!}{(4-3)!} = 24$. Тогда вероятность события A равна $P(A) = \frac{k_A}{n_S} = \frac{1}{24}$.

Пример 3. Буквы А, Н, А, Н, А, С написаны на отдельных карточках. Ребенок в случайном порядке раскладывает эти карточки в один ряд. Какова вероятность, что получится слово «АНАНАС»?

Решение. Обозначим $A = \{получилось\ слово\ "АНАНАС"\}$ - событие, вероятность которого нужно найти.

События полной группы – это перестановки с повторениями (т.к. есть одинаковые буквы). Количество таких перестановок найдем по формуле $P(n_1, n_2, ..., n_k) = \frac{n!}{n_1! \cdot n_2! \cdot ... \cdot n_k!}$, где $n_1 = 3$

(количество букв «А»), $n_2 = 2$ (количество букв «Н»), $n_3 = 1$ (количество букв «С»), $n = n_1 + n_2 + n_3 = 3 + 2 + 1 = 6$. $P(n_1, n_2, ..., n_k) = \frac{6!}{3! \cdot 2! \cdot 1!} = \frac{3! \cdot 4 \cdot 5 \cdot 6}{3! \cdot 2} = 60$. $P(A) = \frac{1}{60}$.

2.4 **Определение вероятностей событий в задачах с однородной или неоднородной выборкой**

Многие прикладные задачи, связанные со статистическим контролем качества продукции, приводятся к модельным задачам об однородных и неоднородных выборках из генеральной совокупности объектов, которые являются обобщением задач, рассмотренных в 25.3. на конечное число объектов в выборке $(n \geq 1)$.

Определение. Однородной выборкой называется упорядоченный или неупорядоченный набор (подмножество) объектов из генеральной совокупности, содержащих только равновозможные объекты по одному особому признаку.

Задачи о вероятности появления случайной однородной выборки приводятся к следующей модельной задаче: Генеральная совокупность содержит N объектов, среди которых ровно M объектов обладают особым признаком. Найти вероятность извлечения выборки, содержащей n объектов с указанным особым признаком (событие A). Вероятность извлечения однородной выборки находят по формуле

$$P(A) = \frac{C_M^n}{C_N^n}. \qquad (1)$$

Замечание 1. Число объектов в выборке называют объемом выборки.

Пример 1. Имеется пять приборов, среди них два прибора с дефектом. Найти вероятность того, что при случайном выборе будут включены два годных прибора (приборы без дефектов).

Решение. Обозначим $A = \{включены\ два\ годных прибора\}$ - событие, вероятность которого нужно найти. По условиям задачи имеем: $N = 5$ (генеральная совокупность объектов), $M = 3$ (количество годных приборов – объектов с особым признаком), выборка $n = 2$ (число включенных приборов). Тогда вероятность события A равна

$$P(A) = \frac{C_M^m}{C_N^m} = \frac{C_3^2}{C_5^2} = \frac{3!}{2! \cdot (3-2)!} \cdot \frac{2! \cdot (5-2)!}{5!} = 3 \cdot \frac{2}{4 \cdot 5} = \frac{3}{10}.$$

Определение. *Неоднородной выборкой* называется подмножество (возможно – упорядоченное) объектов задачи, выделенных из генеральной совокупности и содержащих объекты с различной равновозможностью. Задачи о вероятности появления случайной неоднородной выборки приводятся к следующей простейшей модельной задаче:

Генеральная совокупность содержит N объектов, среди которых ровно M объектов обладают особым признаком (соответственно, $N-M$ объектов этим признаком не обладают). Из генеральной совокупности взята случайная выборка, содержащая n объектов. Найти вероятность того, что в этой выборке окажется ровно m объектов, обладающих особым признаком.

Для наглядности решения указанной задачи целесообразно использовать табличную форму представления данных.

Вид совокупности	**Всего объектов**	**Число объектов с особым признаком**	**Остальные объекты**
Генеральная совокупность	$N=\ldots$	$M=\ldots$	$N-M=\ldots$
Выборка	$n=\ldots$	$m=\ldots$	$n-m=\ldots$

Используя табличные данные, вероятность появления выборки из n объектов, среди которых находятся ровно m объектов с особым признаком, вычисляют по формуле

$$P(A)=\frac{C_M^m \cdot C_{N-M}^{n-m}}{C_N^n} \qquad (2).$$

Замечание 2. Из одной и той же генеральной совокупности, содержащей M объектов с особым признаком, могут быть извлечены однородные $(n=m)$ или неоднородные $(n\neq m)$ выборки. В случае однородной выборки $(n=m, n-m=0)$ из формулы (2) следует формула (1).

Пример 1. Среди 100 деталей имеются ровно 10 деталей с дефектом. Для контроля выбраны 5 деталей. Найти вероятность того, что среди них ровно две дефектные детали.

Решение. Событие A -«среди выбранных пяти деталей ровно две дефектные» является событием, вероятность которого нужно найти.

Вид совокупности	**Всего объектов**	**Число объектов с дефектом**	**Число объектов без дефекта**

Генеральная совокупность	$N=100$	$M=10$	$N-M=90$
Выборка	$n=5$	$m=2$	$n-m=3$

Имеем неоднородную выборку $(n \neq m,\ 5 \neq 2)$, поэтому используем формулу (2)

$$P(A)=\frac{C_{10}^{2}\cdot C_{90}^{3}}{C_{100}^{10}}=\frac{10!}{2!\cdot 8!}\cdot\frac{90!}{3!\cdot 87!}\cdot\frac{5!\cdot 95!}{100!}=\frac{90}{2}\cdot\frac{88\cdot 89\cdot 90}{6}\cdot\frac{2\cdot 3\cdot\ldots\cdot 10}{91\cdot 92\cdot\ldots\cdot 100}\approx 0{,}056\,.$$

Решение некоторых задач о неоднородных выборках можно упростить, используя противоположные события, что иногда позволяет перейти к однородным выборкам.

Пример 2. В лотерее 100 билетов, из них 10 выигрышных. Куплено три билета. Какова вероятность того, что выиграет хотя бы один билет?

Решение. Пусть событие A - «выиграет хотя бы один билет», тогда противоположное событие $\overline{A}$ формулируется так: «не выиграет ни один билет». Найдем вероятность противоположного события. Количество событий в полной группе равно C_{100}^{3}, а количество событий из полной группы, благоприятных событию $\overline{A}$ равно C_{90}^{3}. Поэтому, по формуле классической вероятности получаем $P(\overline{A})=\frac{C_{90}^{3}}{C_{100}^{3}}=\frac{90!}{3!\cdot 87!}\cdot\frac{3!\cdot 97!}{100!}=\frac{88\cdot 89\cdot 90}{98\cdot 99\cdot 100}\approx 0{,}727$. Вероятность события A находится по формуле $P(A)=1-P(\overline{A})\approx 1-0{,}727=0{,}273$.

Если неоднородная выборка содержит объекты с несколькими различными особыми признаками, причем каждый объект обладает ровно одним из этих признаков, то вероятность появления такой выборки определяют по формуле

$$P(A)=\frac{C_{M_1}^{m_1}\cdot C_{M_2}^{m_2}\cdot\ldots\cdot C_{M_k}^{m_k}}{C_{N}^{n}}, \qquad (3)$$

где $M_1, M_2, \ldots, M_k$ - количества объектов с особыми признаками $1, 2, \ldots, k$ в генеральной совокупности, $m_1, m_2, \ldots, m_k$ - количества объектов с особыми признаками $1, 2, \ldots, k$ в выборке, $M_1+M_2+\ldots+M_k=N$ и $m_1+m_2+\ldots+m_k=n$.

Пример 3. В урне 15 красных, 9 синих и 6 зеленых шаров, неразличимых на ощупь. Случайно вынули 6 шаров. Найти вероятность того, что среди них 3 красных, 2 синих и 1 зеленый шар.

Решение. Событие A - «из урны вынуты 3 красных, 2 синих и 1 зеленый шар» является событием, вероятность которого нужно найти.

$M_1 = 15,\ M_2 = 9,\ M_3 = 6,\ N = M_1 + M_2 + M_3 = 15 + 9 + 6 = 30;$

$m_1 = 3,\quad m_2 = 2,\quad m_3 = 1,\quad n = 3 + 2 + 1 = 6.$ По формуле (3) получаем:

$$P(A) = \frac{C_{15}^3 \cdot C_9^2 \cdot C_6^1}{C_{30}^6} = \frac{15!}{3! \cdot 12!} \cdot \frac{9!}{2! \cdot 7!} \cdot \frac{6!}{1! \cdot 5!} \cdot \frac{6! \cdot 24!}{30!} \approx 0{,}17.$$

Тема 3: Основные формулы и теоремы теории вероятностей событий

3.1 **Алгебраические операции с событиями**

Алгебраические операции с событиями используются в вероятностных задачах, в которых необходимо одни события представить составленными из других, вероятности которых известны или легко определяются из условий задачи.

Наглядное представление об алгебраических операциях с событиями дают диаграммы Эйлера-Венна. В этих диаграммах полную группу (полное множество) исходных событий S обычно представляют в виде прямоугольника, множество точек которого соответствует различным

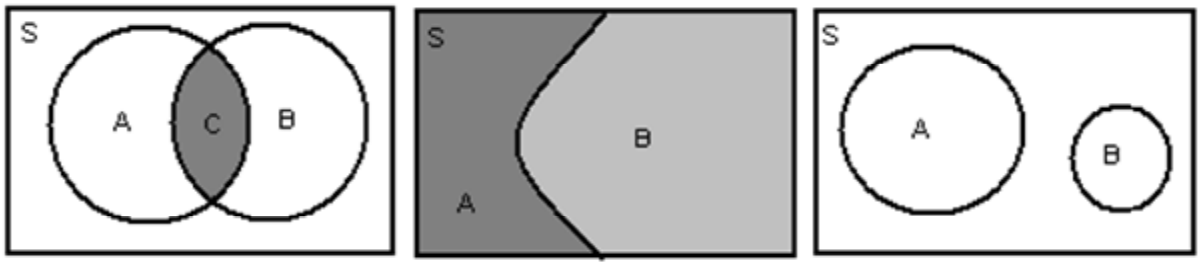


Рис. 27.1 Диаграммы Эйлера-Венна для произведения событий
а) события A и B совместны (C = AB), б), в) события A и B несовместны

возможным вариантам исходных событий. Подмножества полной группы событий S представляются в виде областей, лежащих в прямоугольнике (Рис. 27.1). Обычно эти области представляют в виде кругов.

Определение. *Произведением (логическим произведением)* двух событий A и B называется новое событие $C = A \cdot B$, которое считается наступившим, если произошли события A и B в одном и том же однократном испытании.

На Рис. 27.1 а) показана диаграмма Эйлера-Венна для произведения совместных событий A и B. В задачах теории вероятностей исходные события при необходимости представляются как

сложные в виде произведения простых событий, которые не раскладываются на другие, более простые, события на смысловом уровне.

Замечание. Наглядным примером является представление слов как сложных событий с помощью букв, которые соответствуют простым событиям.

Пример 1. Пусть $A = \{попадание\ в\ мишень\ при\ первом\ выстреле\}$,

$B = \{попадание\ в\ мишень\ при\ втором\ выстреле\}$ - простые события. Если однократное испытание состоит из двух выстрелов, то исходное событие

$C = AB = \{два\ попадания\ в\ мишень\ при\ двух\ выстрелах\}$ является сложным событием.

Понятие произведения событий можно обобщить на любое количество событий:

Определение. *Произведением (логическим произведением)* событий A_1, A_2, ... , A_n называется новое событие $C = A_1 \cdot A_2 \cdot ... \cdot A_n$, которое считается наступившим в однократном испытании, если произошли все события A_1, A_2, ... , A_n в этом испытании.

Определение. *Суммой (логической суммой)* двух событий A и B называется новое событие $C = A + B$, которое считается наступившим, если произошло хотя бы одно из событий A или B, или произошли оба события A и B вместе (Рис. 27.2).

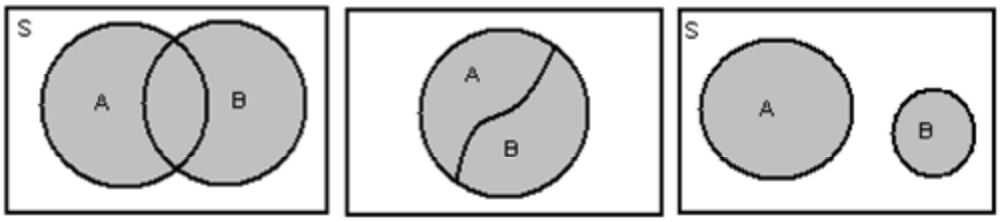


Рис. 27.2 Диаграммы Эйлера-Венна для суммы событий A+B
а) События A и B совместны б), в) События A и B несовместны

Определение. События A и B называются *совместными*, если они могут произойти вместе (совместно) в одном и том же однократном испытании.

Определение. События A и B называются *несовместными*, если они не могут произойти вместе в одном однократном испытании.

Пример 2. События $A = \{выпадение\ "орла"\ при\ одном\ подбрасывании\ одной\ монеты\}$ и

$B = \{выпадение\ "решки"\ при\ таком\ же\ подбрасывании\ монеты\}$ - несовместны.

Пример 3. Пусть событие $A = \{выигрыш\ в\ лотерее\ по\ одному\ билету\}$, событие $B = \{выигрыш\ в\ той\ же\ лотерее\ по\ второму\ билету\}$. События A и B совместны, так как возможно событие $C = A + B = \{выигрыш\ в\ той\ же\ лотерее\ по\ обоим\ билетам\}$

Понятие суммы событий можно обобщить на любое количество событий

Определение. *Суммой (логической суммой)* событий $A_1, A_2, ..., A_n$ называется новое событие $C = A_1 + A_2 + ... + A_n$, которое считается наступившим, если произошло хотя бы одно из событий $A_1, A_2, ..., A_n$ (или несколько этих событий, или все эти событий)

Если событие A можно представить в виде суммы несовместных событий, то это событие называют *составным*, а слагаемые интерпретируют при необходимости как варианты события A или события, благоприятные для события A.

Пример 4. Пусть $A = \{выпадение\ четного\ количества\ очков\ при\ подбрасывании\ кубика\}$,

$A_1 = \{выпадение\ "двойки"\ при\ подбрасывании\ кубика\}$,

$A_2 = \{выпадение\ "четверки"\ при\ подбрасывании\ кубика\}$,

$A_3 = \{выпадение\ "шестерки"\ при\ подбрасывании\ кубика\}$, тогда $A = A_1 + A_2 + A_3$.

События A_1, A_2, A_3 - варианты события A (события, благоприятные для события A).

Событие A считается наступившим, если наступил один из вариантов этого события - A_1, A_2 или A_3.

Определение. *Разностью (логической разностью)* событий A и B называют событие $C = A - B$, которое наступает тогда, когда наступает событие A и при этом не наступает событие B.

Замечание. Событие C включает в себя все варианты события A, которые неблагоприятны событию B.

Определение. *Дополнением* $\overline{A}$ события A (или *событием, противоположным к событию* A) называется событие, которое состоит в том, что событие A не наступило.

Событие A и противоположное событие $\overline{A}$ образуют полную группу несовместных событий: $A + \overline{A} = U$ (U - достоверное событие).

Операции сложения и умножения событий обладают следующими свойствами:

1. Ассоциативность: $(A+B)+C=A+(B+C)$, $(A\cdot B)\cdot C=A\cdot(B\cdot C)$;

   Замечание. Из свойства ассоциативности следует что в суммах и произведениях событий скобки можно расставлять произвольным образом, то есть, по сути, оперировать с событиями как с числами. Однако, для разностей событий это не верно. Например, $(A-B)+(B-A)\neq\varnothing$

2. Коммутативность: $A+B=B+A$, $A\cdot B=B\cdot A$;
3. Дистрибутивность: $A\cdot(B+C)=A\cdot B+A\cdot C$
4. Правила де Моргана $\overline{A+B}=\overline{A}\cdot\overline{B}$ $\Leftrightarrow$ $\overline{A\cdot B}=\overline{A}+\overline{B}$.

Кроме того, справедливы соотношения: $A+\overline{A}=U$, $A\cdot\overline{A}=V$ (невозможное событие), $A+V=A$, $A+U=U$, $A\cdot V=V$, $A\cdot U=A$, $A+A=A$, $A\cdot A=A$.

Между операциями с событиями и операциями с множествами можно установить взаимно-однозначное соответствие и представлять одни события через другие используя теоретико-множественную символику: $A+B=A\cup B$ (объединение двух событий), $A\cdot B=A\cap B$ (пересечение двух событий), $A-B=A\setminus B$.

Если $B\subset A$, то $A\cdot B=B$, $A+B=A$ и, следовательно, $A\cdot B+A=A$ - *формула поглощения*.

Любое событие можно представить в виде суммы несовместных событий: $A=A\cdot U=A\cdot(B+\overline{B})=AB+A\overline{B}$ (1) .Если $B\subset A$, то $AB=B$ (2). Из (1) и (2) следует $A=B+A\overline{B}$.

Пример 5. Три выстрела в мишень составляют однократное испытание. С помощью логических операций над событиями $A_i,\ (i=1,\ 2,\ 3)$ - попадание в мишень при i-ом выстреле описать события:

а) два попадания после одного промаха (событие A);

б) два промаха при трех выстрелах (событие B);

в) не менее двух промахов (событие C);

Решение. Поскольку $A_i,\ (i=1,\ 2,\ 3)$ - попадание в мишень при i-ом выстреле, то $\overline{A}_i,\ (i=1,\ 2,\ 3)$ - промах при i-ом выстреле.

а) Событие A рассматриваем как сложное и представляем в виде произведения трех событий: $A = \overline{A}_1 A_2 A_3$;

б) По условию задачи для события B не указано, при каких выстрелах могут быть промахи. Поэтому событие B имеет несколько вариантов и его следует рассматривать как составное событие. С другой стороны, каждый из вариантов является сложным событием, состоящим из трех событий: $B = \overline{A}_1 \overline{A}_2 A_3 + \overline{A}_1 A_2 \overline{A}_3 + A_1 \overline{A}_2 \overline{A}_3$

.в) Аналогично рассматривается событие $C = \overline{A}_1 \overline{A}_2 A_3 + \overline{A}_1 A_2 \overline{A}_3 + A_1 \overline{A}_2 \overline{A}_3 + \overline{A}_1 \overline{A}_2 \overline{A}_3$

Замечание. В дальнейшем события, являющиеся суммами произведений событий, будем, при необходимости, называть *сложносоставными событиями*.

Пример 6. Взятая со склада деталь может оказаться деталью первого сорта (событие A), второго сорта (событие B) или третьего сорта (событие C). Необходимо вербально определить события $A + B$, $\overline{A + C}$, $AB + C$.

Замечание. Определить «вербально» - значит определить словами на смысловом уровне.

Решение. $D = A + B = \{взята\ деталь\ первого\ или\ второго\ сорта\}$,

$\overline{A + C} = \{взята\ деталь\ не\ первого\ и\ не\ третьего\ сорта\}$ - взята деталь второго сорта,

$AB = V$ - невозможное событие, поэтому

$AB + C = C$

Пример 7. Определить условия работоспособности электрической цепи с помощью алгебры событий.

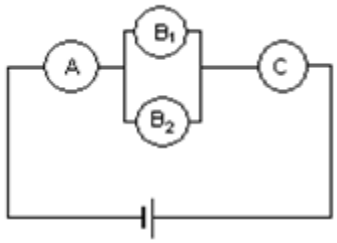


Рис. 27.3 Электрическая цепь к Примеру 6

Решение. Вводим события:

$A = \{элемент\ A\ \ работоспособен\}$,

$B_1 = \{элемент\ B_1\ \ работоспособен\}$, $B_2 = \{элемент\ B_2\ \ работоспособен\}$,

$C = \{элемент\ C\ \ работоспособен\}$. Для того чтобы ток шел по цепи (событие D), очевидно, что элементы A и C должны быть работоспособны и, кроме того, должен быть работоспособен хотя бы один из элементов B_1, B_2: $D = AC\left(B_1 B_2 + B_1 \overline{B}_2 + \overline{B}_1 B_2\right)$.

Пример 8. Доказать, что событие $(A + B)(A + \overline{B}) + (\overline{A} + B)(\overline{A} + \overline{B})$ является достоверным

Решение. Раскрываем скобки: $AA + BA + A\overline{B} + B\overline{B} + \overline{A}\overline{A} + B\overline{A} + \overline{A}\overline{B} + B\overline{B}$. Учитывая, что $AA = A$, $\overline{A}\overline{A} = \overline{A}$ и $B\overline{B} = V$ - невозможное событие, получим

$A + BA + A\overline{B} + \overline{A} + B\overline{A} + \overline{A}\,\overline{B} = A + A\left(B + \overline{B}\right) + \overline{A} + \overline{A}\left(B + \overline{B}\right)$. Учитывая, что $A + \overline{A} = U$ и $B + \overline{B} = U$ - достоверные события, получим $U + AU + \overline{A}\,U = U + A + \overline{A} = U + U = U$.

3.2 **Теоремы умножения и сложения вероятностей событий**

Теоремы умножения и сложения вероятностей используют для определения вероятностей сложных и сложносоставных событий, которые можно представить в виде сумм и произведений других событий, вероятности которых известны или легко определяются непосредственным подсчетом по условию задачи.

Замечание. По существу, математическая теория вероятностей определяется как «раздел математики, в котором по вероятностям одних случайных событий находят вероятности других событий, связанных как-либо с первыми» (А.Н. Колмогоров, БСЭ). С этой точки зрения темы 25 и 26 являются полуэмпирической базой, на основе которой развита аксиоматическая теория вероятностей и некоторые теоремы которой рассматриваются далее в содержательных терминах.

3.2.1 **Теоремы умножения вероятностей событий**

При использовании теорем умножения события делят на зависимые и независимые.

Определение. Событие B называется *зависимым* от события A, если вероятность события B зависит от того, произошло событие A или нет. В противном случае событие B называется *независимым* от события A.

Замечание. Ниже будет показано, что, если событие B зависит от события A, то и событие A зависит от события B.Поэтому можно говорить, что события A и B *зависимы* или *независимы* (без уточнения того, какое событие от какого зависит).

Пример 1. Из урны, содержащей 4 белых и 6 черных шаров, наудачу один за другим вынуты два шара. Проверить зависимость (или независимость) событий $A = \{первый\ шар - белый\}$ $B = \{второй\ шар - белый\}$

Решение. По определению зависимости (независимости) событий нужно посчитать вероятность $P(B)$ при условии, что событие A произошло и ту же вероятность при условии, что событие A не произошло. Если эти вероятности совпадут – события A и B независимы, если не совпадут – зависимы.

Пусть событие A произошло, т.е. первый вынутый шар был белым. Тогда в урне осталось 9 шаров, среди которых 3 белых и $P(B)=\frac{3}{9}$.

Пусть событие A не произошло, т.е. первый вынутый шар не был белым (был черным). Тогда в урне осталось 9 шаров, среди которых 4 белых и . $P(B)=\frac{4}{9}$.

Вероятности не совпали, значит, события A и B зависимы.

Определение. *Условной вероятностью* события B называют вероятность этого события, вычисленную при условии, что предварительно произошло некоторое событие A, от которого может зависеть вероятность события B.

В математической форме условная вероятность вводится формулой $P(B \mid A)=\frac{P(AB)}{P(A)}$, откуда получается теорема умножения вероятностей двух произвольных событий: $P(AB)=P(A)P(B \mid A)$

.

Используя свойство коммутативности логической операции умножения событий, получаем $P(AB)=P(BA) \Rightarrow P(A)P(B \mid A)=P(B)P(A \mid B)$.

Отсюда и следует (см. *Замечание*), что зависимость и независимость событий всегда взаимны.

Если события A и B независимы, то . $P(A \mid B)=P(A)$, $P(B \mid A)=P(B)$, т.е. условные вероятности независимых событий равны безусловным вероятностям этих событий. Поэтому формула вычисления вероятности произведения двух независимых событий будет иметь вид $P(AB)=P(A)P(B)$.

Независимость событий обычно определяется по условиям прикладной задачи, исходя из следующего принципа: *если события причинно независимы, то они независимы и в вероятностном смысле*. Одни и те же события могу быть зависимыми или независимыми в разных схемах испытаний (с возвращениями или без возвращений).

Пример 2. В урне 6 шаров, среди них 3 белых и три красных шара. Найти вероятность последовательного извлечения двух красных шаров при двух различных схемах испытаний:

а) *каждый шар после извлечения не возвращают в урну*;

б) *каждый шар после извлечения возвращают в урну*

Решение. Обозначим $A_1 = \{первый\ шар - красный\}$ $A_2 = \{второй\ шар - красный\}$, $A = \{оба\ шара - красные\}$, тогда $A = A_1 A_2$

а) События A_1, A_2 - зависимые (см. Пример 1), поэтому

$$P(A) = P(A_1 A_2) = P(A_1)P(A_2 \mid A_1) = \frac{3}{6} \cdot \frac{2}{5} = 0{,}2;$$

б) События A_1, A_2 - независимые. После возвращения в урну первого шара в урне восстанавливаются те же условия, что были до извлечения первого шара, поэтому $P(A) = P(A_1 A_2) = P(A_1)P(A_2) = \frac{3}{6} \cdot \frac{3}{6} = 0{,}25$ (впрочем, можно было и не устанавливать независимость событий A_1, A_2, а пользоваться формулой $P(A) = P(A_1)P(A_2 \mid A_1)$, что привело бы к тому же ответу).

Теорему умножения вероятностей событий можно обобщить с двух событий на любое их конечное количество, например, $P(ABC) = P(A)P(B \mid A)P(C \mid AB)$, $P(ABCD) = P(A)P(B \mid A)P(C \mid AB)P(D \mid ABC)$ и так далее.

Пример 3. Студент сдает экзамен, зная 25 вопросов из 30. В билете 3 вопроса. Какова вероятность тог, что студент знает все три вопроса (событие A)?

Решение. Событие A является сложным и содержит три простых зависимых события $A_i = \{студент\ знает\ i - ый\ вопрос\}$, $i = 1,2,3$: $A = A_1 A_2 A_3$.

$$P(A_1 A_2 A_3) = P(A_1)P(A_2 \mid A_1)P(A_3 \mid A_1 A_2) = \frac{25}{30} \cdot \frac{24}{29} \cdot \frac{23}{28} \approx 0{,}57$$

Пример 4. Из карточек, на каждой из которых написано по одной букве, составлено слово «ПАПАХА». Затем карточки перевернули и перемешали. Найти вероятность того, что при последовательном случайном извлечении четырех карточек без возвращения будет составлено слово «ПАПА» (событие E).

Решение. Всего имеется 6 карточек, из них две с буквой «П», три – с буквой «А» и одна с буквой «Х». Исходное событие E является сложным и включает в себя 4 простых события по случайному поочередному выбору четырех карточек по одной без возвращений. вероятности простых событий определим непосредственным подсчетом.

Событие $A = \{первой\ случайно\ извлечена\ карточка\ "П"\}$; $P(A) = \frac{2}{6} = \frac{1}{3}$;

Событие $B = \{второй\ случайно\ извлечена\ карточка\ "А"\};\ P(B \mid A) = \frac{3}{6-1} = \frac{3}{5}$

Событие $C = \{третьей\ случайно\ извлечена\ карточка\ "П"\};\ P(C \mid AB) = \frac{2-1}{6-2} = \frac{1}{4}$

Событие $B = \{четвертой\ случайно\ извлечена\ карточка\ "А"\};\ P(D \mid ABC) = \frac{3-1}{6-3} = \frac{2}{3}$

$$P(E) = P(A)P(B \mid A)P(C \mid AB)P(D \mid ABC) = \frac{1}{3} \cdot \frac{3}{5} \cdot \frac{1}{4} \cdot \frac{2}{3} = \frac{1}{30}$$

Теорему умножения вероятностей независимых событий также можно обобщить с двух событий на любое их конечное количество: $P(A_1 A_2 .. A_n) = P(A_1)P(A_2)...P(A_n)$.

3.2.2 **Теоремы сложения вероятностей событий**

При использовании теорем сложения вероятностей целесообразно предварительно проверить их совместность (несовместность). Если события совместны, то целесообразно также проверить их зависимость (независимость).

Пример 6. Контролируются детали по двум размерам.

Событие $A = \{брак\ детали\ по\ первому\ размеру\}$,

Событие $B = \{брак\ детали\ по\ второму\ размеру\}$.

События A и B - совместные, так как возможен брак по обоим размерам.

Замечание. Совместность и зависимость одних и тех же событий являются различными свойствами, не имеющими в общем случае взаимных связей между собой.

Теоремы сложения вероятностей двух событий в математической форме представляются так:

1) Общий случай (произвольные события A и B): $P(A+B) = P(A) + P(B) - P(AB)$;
2) Случай совместных и зависимых событий $P(A+B) = P(A) + P(B) - P(A)P(B \mid A)$;
3) Случай совместных и независимых событий $P(A+B) = P(A) + P(B) - P(A)P(B)$;
4) Случай несовместных событий $P(A+B) = P(A) + P(B)$.

Пример 7. Из 100 лотерейных билетов 5 билетов являются выигрышными. Какова вероятность выиграть ровно по одному билету из двух купленных (событие A)?

Решение. Событие $A = \{выигрывает\ ровно\ один\ билет\}$ может быть реализовано в двух вариантах: $A_1 = \{выигрывает\ только\ первый\ билет\}$ или

$A_2 = \{выигрывает\ только\ второй\ билет\}$ то есть $A = A_1 + A_2$. События A_1, A_2 - совместные (могут выиграть оба билета) и зависимые (после выигрыша по первому билету шансы на выигрыш по второму меняются). Поэтому теорему сложения вероятностей используем в форме $P(A+B) = P(A) + P(B) - P(A)P(B \mid A)$:

$$P(A) = P(A_1) + P(A_2) - P(A_2)P(A_2 \mid A_1) = \frac{5}{100} + \frac{5}{100} - \frac{5}{100} \cdot \frac{4}{99} = \approx 0{,}1$$

Теорему сложения вероятностей (общий случай) можно обобщить с двух событий на любое их количество, например, $P(A+B+C) = P(A) + P(B) + P(C) - P(AB) - P(AC) - P(BC) + P(ABC)$.

Если события $A_1, A_2, ..., A_n$ несовместны, то $P(A_1 + A_2 + ... + A_n) = P(A_1) + P(A_2) + ... + P(A_n)$.

3.2.3 **Совместное применение теорем умножения и сложения вероятностей событий**

При решении вероятностных задач теоремы умножения и сложения вероятностей нередко используют совместно.

Пример 8. Для сигнализации об аварии установлены два независимо работающих сигнализатора. вероятность срабатывания первого сигнализатора равна $p_1 = 0{,}95$, второго - $p_2 = 0{,}9$. Найти вероятность того, что при аварии сработает ровно один сигнализатор (событие A).

Решение. Обозначим $A_1 = \{сработает\ только\ первый\ сигнализатор\}$, $A_2 = \{сработает\ только\ второй\ сигнализатор\}$, тогда $A = A_1\overline{A}_2 + \overline{A}_1 A_2$.. События A_1, A_2 - независимые, значит, события $\overline{A}_1$, $\overline{A}_2$ - тоже независимые. События $A_{(1)} = A_1\overline{A}_2$ и $A_{(2)} = \overline{A}_1 A_2$ - несовместные. значит,

$$P(A) = P(A_1\overline{A}_2 + \overline{A}_1 A_2) = P(A_1)P(\overline{A}_2) + P(\overline{A}_1)P(A_2) = 0{,}95 \cdot 0{,}1 + 0{,}9 \cdot 0{,}05 = 0{,}14.$$

Пример 9. Вероятности того, что каждая из трех машин доедет до финиша, равны соответственно 0,8; 0,4; 0,7. найти вероятность того, что к финишу доедут ровно две машины (событие A).

Решение. Событие A содержит варианты, каждый вариант является сложным событием. Вводим единичные события $A_i = \{до\ финиша\ доедет\ только\ одна\ машина\}$, $i = 1, 2, 3$. Из условия задачи $P(A_1) = p_1 = 0{,}8$, $P(A_2) = p_2 = 0{,}4$, $P(A_{31}) = p_3 = 0{,}7$ определяем вероятности событий, противоположных к единичным: $P(\overline{A}_1) = q_1 = 1 - 0{,}8 = 0{,}2$, $P(\overline{A}_2) = q_2 = 1 - 0{,}4 = 0{,}6$, $P(\overline{A}_3) = q_3 = 1 - 0{,}7 = 0{,}3$. $A = \overline{A}_1 A_2 A_3 + A_1 \overline{A}_2 A_3 + A_1 A_2 \overline{A}_3$. События A_1, A_2, A_3 - независимы,

события $\overline{A}_1$, $\overline{A}_2$, $\overline{A}_3$ тоже независимы. События $A_{(1)} = \overline{A}_1 A_2 A_3$, $A_{(2)} = A_1 \overline{A}_2 A_3$, $A_{(3)} = A_1 A_2 \overline{A}_3$ - несовместны, значит, $P(A) = P(\overline{A}_1)P(A_2)P(A_3) + P(A_1)P(\overline{A}_2)P(A_3) + P(A_1)P(A_2)P(\overline{A}_3) =$

$= q_1 p_2 p_3 + p_1 q_2 p_3 + p_1 p_2 q_3 = 0{,}2 \cdot 0{,}4 \cdot 0{,}7 + 0{,}8 \cdot 0{,}6 \cdot 0{,}7 + 0{,}8 \cdot 0{,}4 \cdot 0{,}3 = 0{,}712$.

3.3 **Вычисление вероятности появления хотя бы одного события**

Определение вероятности хотя бы одного из событий $A_1, A_2, ..., A_n$ сводится к задаче вычисления вероятности произведения противоположных событий по формуле $P(A) = P(A_1 + A_2 + ... + A_n) = 1 - P(\overline{A}_1 \overline{A}_2 ... \overline{A}_n)$.

Далее для краткости и при необходимости вероятности прямых событий $A_1, A_2, ..., A_n$ будем обозначать $p_1, p_2, ..., p_n$, а вероятности противоположных событий $\overline{A}_1, \overline{A}_2, ..., \overline{A}_n$ - $q_1, q_2, ..., q_n$.

Пример 1. Из 100 лотерейных билетов 5 билетов являются выигрышными. Приобретено 4 билета. Найти вероятность выигрыша хотя бы по одному билету (событие A).

Решение. Вводим единичные события $A_i = \{i - ый\ билет\ выиграет\}$, $i = 1, 2, 3, 4$. тогда $A = A_1 + A_2 + A_3 + A_4$. При этом события A_1, A_2, A_3, A_4 являются зависимыми в совокупности (после выигрыша или не выигрыша каждого билета изменяется число выигрышных билетов и общее число билетов).

$P(A) = 1 - P(\overline{A}_1 \overline{A}_2 \overline{A}_3 \overline{A}_4) = 1 - P(A_1)P(A_2 \mid A_1)P(A_3 \mid A_1 A_2)P(A_4 \mid A_1 A_2 A_3) =$

$= 1 - \frac{95}{100} \cdot \frac{94}{99} \cdot \frac{93}{98} \cdot \frac{92}{97} \approx 0{,}19$.

Пример 2. В подъезде установлены три видеокамеры, которые включаются и выключаются независимо друг от друга. При этом вероятность включения каждой видеокамеры равна 0,6. Какова вероятность того, что включена хотя бы одна видеокамера (событие A)?

Решение. Переходим к противоположному событию $\overline{A} = \{выключены\ все\ три\ видеокамеры\}$ и введем простые (единичные) события $E_i = \{включена\ i - ая\ камера\}$, $\overline{E}_i = \{выключена\ i - ая\ камера\}, i = 1, 2, 3$. По условию задачи $P(E_i) = 0{,}6$, значит, $P(\overline{E}_i) = 1 - 0{,}6 = 0{,}4$.

Тогда $P(A) = 1 - P(\overline{A}) = 1 - P(\overline{E}_1 \overline{E}_2 \overline{E}_3) = 1 - P(\overline{E}_1)P(\overline{E}_2)P(\overline{E}_3) = 1 - 0{,}4^3 = 0{,}936$.

Используя формулу $P(A) = 1 - q^n$ можно найти минимальное количество одинаковых независимых событий (или объектов, соответствующих этим событиям) для обеспечения заданной вероятности события A: $P(A) = 1 - q^n \Rightarrow q^n = 1 - P(A) \Rightarrow n \ln q = \ln(1 - P(A))$.

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

Замечание. Поскольку $n \in N$, то, более точно, $n = \left[\frac{\ln(1-P(A))}{\ln q}\right] + 1$, где $\left[\frac{\ln(1-P(A))}{\ln q}\right]$ - целая часть числа $\frac{\ln(1-P(A))}{\ln q}$.

Пример 3. Вероятность безотказной работы прибора равна $p = 0{,}7$. Для повышения надежности работы технической системы он дублируется другими такими же приборами. Сколько приборов нужно взять, чтобы обеспечить вероятность 0,95 безотказной работы технической системы (событие A)?

Решение. Введем простые (единичные) события $E_i = \{работает\ i-ый\ прибор\}$, $i = 1, 2, ..., n$; $P(E_i) = p = 0{,}7$, $P(\overline{E}_i) = q = 1 - 0{,}7 = 0{,}3$.

$$0{,}95 = P(A) = 1 - q^n = 1 - 0{,}3^n \Rightarrow 0{,}3^n = 1 - 0{,}95 = 0{,}05 \Rightarrow n \ln 0{,}3 = \ln 0{,}05 \Rightarrow n = \left[\frac{\ln 0{,}05}{\ln 0{,}3}\right] + 1$$

$$\Rightarrow n = \left[\frac{\ln\frac{5}{100}}{\ln\frac{3}{10}}\right] + 1 = \left[\frac{\ln 5 - \ln 100}{\ln 3 - \ln 10}\right] + 1 = [2{,}488] + 1 = 2 + 1 = 3$$

3.4 **Полная вероятность события. Формула Байеса**

Пусть дана полная группа несовместных событий $H_1, H_2, ..., H_n$ (не обязательно равновозможных), которые будем назвать гипотезами. Вероятности гипотез будем считать известными или определяемыми из условия задачи. при этом, в силу определения полной группы несовместных событий, $\sum_{i=1}^{n} p(H_i) = 1$. В силу того же определения, любое событие F происходит совместно (после или одновременно) с одной из гипотез $H_1, H_2, ..., H_n$. При указанных условиях вероятность события F вычисляется по формуле (*формула полной вероятности*)

$$P(F) = \sum_{i=1}^{n} P(H_i) P(F \mid H_i) \qquad (1)$$

Замечание. По существу, используя формулу (1), находим некоторую «среднюю» вероятность появления события F, при условии, что произошло одно из событий – гипотез.

В задачах, предполагающих использование формулы полной вероятности, можно выделить три последовательных этапа:

Первый этап. Определение полной группы событий – гипотез и априорных (безусловных) вероятностей $P(H_i)$ каждой из гипотез. Проверка того, что гипотезы образуют полную группу, состоит в проверке равенства $\sum_{i=1}^{k} P(H_i) = 1$, где k - общее количество гипотез.

Второй этап: определение условных вероятностей $P(F \mid H_i)$ появления события F при условии, что гипотезы H_i при $i = 1, 2, ..., k$ реализуются поочередно.

Третий этап: Вычисление полной вероятности события F по формуле (1), используя результаты первых двух этапов.

Пример 1. В тире пять винтовок, из них три – с оптическим прицелом. Вероятность поражения мишени из винтовки с оптическим прицелом равна 0,95, а из винтовки без оптического прицела – 0,7. Найти вероятность поражения мишени при одном выстреле из наугад взятой винтовки (событие F).

Решение. Первый этап. Вероятность поражения мишени зависит от того, какая винтовка выбрана – с оптическим прицелом или без него. Поэтому естественно в полную группу несовместных событий (гипотез) включить $H_1 = \{выбрана\ винтовка\ с\ прицелом\}$ и $H_2 = \{выбрана\ винтовка\ без\ прицела\}$. По формуле классической вероятности $P(H_1) = \frac{3}{5} = 0{,}6$, $P(H_2) = \frac{2}{5} = 0{,}4$. Проверка: $P(H_1) + P(H_2) = 0{,}4 + 0{,}6 = 1$.

Второй этап. Вероятности $P(F \mid H_1)$ и $P(F \mid H_2)$ даны по условию задачи: $P(F \mid H_1) = 0{,}95$, $P(F \mid H_2) = 0{,}7$.

Третий этап. $P(F) = P(H_1)P(F \mid H_1) + P(H_2)P(F \mid H_2) = 0{,}6 \cdot 0{,}95 + 0{,}4 \cdot 0{,}7 = 0{,}85$.

Если безусловная вероятность события F известна, то можно уточнить (переоценить) априорные вероятности гипотез, используя *формулу Байеса*:

$$P(H_i \mid F) = \frac{P(H_i)P(F \mid H_i)}{P(F)} = \frac{P(H_i)P(F \mid H_i)}{\sum_{j=1}^{n} P(H_j)P(F \mid H_j)}$$

Полученные вероятности называют апостериорными (послеопытными).

Пример 2. В тире пять винтовок, из них три – с оптическим прицелом. Вероятность поражения мишени из винтовки с оптическим прицелом равна 0,95, а из винтовки без оптического прицела – 0,7. Мишень была поражена из наугад взятой винтовки. Найти вероятность того, что эта винтовка была без оптического прицела.

Решение. Будем придерживаться обозначений Примера 1. Вероятность поражения мишени из наугад взятой винтовки найдена в Примере 1: $P(F) = 0{,}85$. По формуле Байеса получаем: $P(H_2 \mid F) = \frac{P(H_2)P(F \mid H_2)}{P(F)} = \frac{0{,}4 \cdot 0{,}7}{0{,}85} \approx 0{,}33$.

Корректировка вероятностей событий по мере поступления новой информации с использованием формулы Байеса применяется в технике и экономике и называется Байесовским подходом.

Пример 3. В первой урне – 10 шаров, среди которых 8 белых. Во второй урне 20 шаров, из них 4 белых. Из каждой урны случайным образом извлечено по одному шару, затем из этих двух шаров наугад выбран один шар. Найти вероятность того, что он белый (событие F).

Решение. Первый этап. Рассмотрим события $W_1 = \{из\ первой\ урны\ вынут\ белый\ шар\}$ и $W_2 = \{из\ второй\ урны\ вынут\ белый\ шар\}$, тогда $\overline{W_1} = \{из\ первой\ урны\ вынут\ не\ белый\ шар\}$ и $\overline{W_2} = \{из\ второй\ урны\ вынут\ не\ белый\ шар\}$. Множество гипотез будет выглядеть так: $H_1 = W_1W_2$, $H_2 = W_1\overline{W_2}$, $H_3 = \overline{W_1}W_2$, $H_4 = \overline{W_1}\overline{W_2}$.

$$P(H_1) = P(W_1W_2) = P(W_1)P(W_2) = \frac{8}{10}\cdot\frac{4}{20} = 0{,}16;$$

$$P(H_2) = P(W_1\overline{W_2}) = P(W_1)P(\overline{W_2}) = \frac{8}{10}\cdot\frac{16}{20} = 0{,}64;$$

$$P(H_3) = P(\overline{W_1}W_2) = P(\overline{W_1})P(W_2) = \frac{2}{10}\cdot\frac{4}{20} = 0{,}04;$$

$$P(H_4) = P(\overline{W_1}\overline{W_2}) = P(\overline{W_1})P(\overline{W_2}) = \frac{2}{10}\cdot\frac{16}{20} = 0{,}16.$$

Проверка правильности вычислений: $P(H_1)+P(H_2)+P(H_3)+P(H_4) = 0{,}16+0{,}64+0{,}04+0{,}16 = 1$ - верно.

Второй этап. $P(F \mid H_1) = 1$, $P(F \mid H_2) = \frac{1}{2}$, $P(F \mid H_3) = \frac{1}{2}$, $P(F \mid H_4) = 0$.

Третий этап. $P(F) = 0{,}16\cdot 1 + 0{,}64\cdot\frac{1}{2} + 0{,}04\cdot\frac{1}{2} + 0{,}16\cdot 0 = 0{,}5$

3.5 **Повторные независимые испытания. Формула Бернулли**

Прикладные задачи во многих случаях приводятся к схеме многократного повторения серий из n независимых испытаний. Схема испытаний, удовлетворяющая следующим условиям, называется *схемой Бернулли*:

1. Каждая серия состоит из одного и того же количества n единичных испытаний и считается однократным испытанием;
2. В каждом единичном испытании может произойти только одно из двух простых взаимоисключающих событий (успех - неуспех, да - нет, выигрыш - проигрыш и т.п.) с постоянной вероятностью, не зависящей от вероятностей появления любых предшествующих и (или) последующих событий;

3. Последовательность (порядок) появления успехов в любой серии единичных испытаний является несущественным. Существенным является только количество успехов в серии испытаний.

Для решения задач по определению вероятности появления k успехов в серии из n испытаний по схеме Бернулли при вероятности p успеха в единичном испытании используется формула Бернулли $P_n(k) = C_n^k p^k (1-p)^{n-k}$.

Замечание. Часто вместо разности $1-p$ в формуле Бернулли используется обозначение q: $P_n(k) = C_n^k p^k q^{n-k}$.

Пример 1. Брак при изготовлении деталей на станке составляет 4%. Найти вероятность того, что среди 30 изготовленных на этом станке деталей ровно 2 окажутся браком (событие A).

Решение. Обозначим простое единичное событие $E = \{случайно\ выбранная\ деталь – брак\}$. По условию задачи $P(E) = 0{,}04$, значит, $p = 0{,}04$, $q = 1 - 0{,}04 = 0{,}96$ и, кроме того, $n = 30$, $k = 2$.

$$P(A) = P_{30}(2) = C_{30}^2 (0{,}04)^2 (0{,}96)^{30-2} = \frac{30!}{2!\cdot 28!} \cdot (0{,}04)^2 \cdot (0.96)^{28} \approx 0{,}222.$$

Пример 2. Расход электроэнергии в течение суток не превышает норму с вероятностью $p = \frac{3}{4}$.

Найти вероятность того, что расход энергии не будет превышать норму в течение четырех суток из шести.

Решение. Обозначим простое единичное событие $E = \{не\ превышена\ норма\ расхода\ за\ сутки\}$. По условию задачи $P(E) = \frac{3}{4}$, значит, $p = \frac{3}{4}$, $q = 1 - \frac{3}{4} = \frac{1}{4}$ и, кроме того, $n = 6$, $k = 4$.

. $P(A) = P_6(4) = C_6^4 \left(\frac{3}{4}\right)^4 \left(\frac{1}{4}\right)^{6-4} = \frac{6!}{4!\cdot 2!} \cdot \frac{81}{256} \cdot \frac{1}{16} = \frac{1215}{4096} \approx 0{,}297$.

В многих задачах формулу Бернулли используют многократно. Важнейшие частные случаи:

1. Вероятность того, что в схеме Бернулли успех наступит не менее k_1 раз и не более k_2 раз определяют по формуле $P_n(k_1, k_2) = \sum_{k=k_1}^{k_2} C_n^k p^k q^{n-k} = P_n(k_1) + P_n(k_1+1) + \ldots + P_n(k_2)$.

2. Вероятность того, что в схеме Бернулли успех наступит не более k_0 раз определяют по формуле $P_n(0,k_0)=\sum_{k=0}^{k_0} C_n^k p^k q^{n-k} = P_n(0)+P_n(1)+...+P_n(k_0)$.

3. Вероятность того, что в схеме Бернулли успех наступит не менее k_0 раз определяют по формуле $P_n(k_0,n)=\sum_{k=k_0}^{n} C_n^k p^k q^{n-k} = P_n(k_0)+P_n(k_0+1)+...+P_n(n)$.

4. Вероятность того, что в схеме Бернулли успех наступит хотя бы один раз можно определить по формуле $P_n(1,n)=1-P_n(0)=1-q^n$.

5. Количество единичных испытаний в серии, которое нужно провести, чтобы с вероятностью не меньше γ можно было бы ожидать хотя бы один успех, определяют по формуле $n \geq \frac{\ln(1-\gamma)}{\ln q}$ (см. замечание в разделе 27.3).

Пример 3. На автобазе 10 машин. При нормальной работе автобазы в смену должно работать не менее 8 машин. Вероятность нормальной работы машины в течение смены равна 0,9. Найти вероятность нормальной работы автобазы (событие A).

Решение. Событие A - составное: $A = A_8 + A_9 + A_{10}$, где $A_8 = \{работают\ ровно\ 8\ машин\}$, $A_9 = \{работают\ ровно\ 9\ машин\}$, $A_{10} = \{работают\ ровно\ 10\ машин\}$- несовместные события.

$$P(A)=P(A_8)+P(A_9)+P(A_{10})=C_{10}^8(0{,}9)^8(0{,}1)^2+C_{10}^9(0{,}9)^9(0{,}1)+C_{10}^{10}(0{,}9)^{10}(0{,}1)^0 \approx 0{,}93.$$

Пример 4. Вероятность попадания в воздушную цель при одном выстреле равна $p=0{,}9$. вероятность поражения цели при $k \geq 1$ равна $1-q^k$, где k - число попаданий в цель, $q=1-p$. Найти вероятность того, что для поражения цели хватит не более трех выстрелов (событие A).

Решение. Постановка задачи допускает поражение цели и при одном или двух попаданиях. Поэтому придется рассматривать различные случаи и кроме формулы Бернулли применять формулу полной вероятности.

Первый этап. Рассмотрим гипотезы: $H_0 = \{ни\ одного\ попадания\ при\ трех\ выстрелах\}$ $H_1 = \{ровно\ одно\ попадание\ при\ трех\ выстрелах\}$,

$H_2 = \{\textit{ровно два попадания при трех выстрелах}\}$,

$H_3 = \{\textit{ровно три попадание при трех выстрелах}\}$.

По формуле Бернулли получаем: $P(H_0) = P_3(0) = C_3^0 (0{,}9)^{01} (0{,}1)^3 = 0{,}001$,

$P(H_1) = P_3(1) = C_3^1 (0{,}9)^1 (0{,}1)^2 = 0{,}027$, $P(H_2) = P_3(2) = C_3^2 (0{,}9)^2 (0{,}1)^1 = 0{,}243$,

$P(H_3) = P_3(3) = C_3^3 (0{,}9)^3 (0{,}1)^0 = 0{,}729$. Проверка: $P(H_0) + P(H_1) + P(H_2) + P(H_3) = 1$.

Второй этап. $P(A \mid H_0) = 1 - q^0 = 1 - (0{,}1)^0 = 0$, $P(A \mid H_1) = 1 - q^1 = 1 - 0{,}1 = 0{,}9$,

$P(A \mid H_2) = 1 - q^2 = 1 - (0{,}1)^2 = 0{,}99$, $P(A \mid H_3) = 1 - q^3 = 1 - (0{,}1)^3 = 0{,}999$.

Третий этап. По формуле полной вероятности получаем:

$P(A) = 0{,}001 \cdot 0 + 0{,}027 \cdot 0{,}9 + 0{,}243 \cdot 0{,}99 + 0{,}729 \cdot 0{,}999 = 0{,}993$.

Тема 4 Дискретные случайные величины

4.1 **Исходные положения**

Случайной величиной (СВ) называется переменная величина, которая может принимать значения из некоторого заданного числового множества с различной вероятностью при многократных испытаниях по одной и той же схеме.

Случайные величины, рассматриваемые как множества в целом, будем обозначать прописными латинскими буквами $X, Y, Z, \ldots$ Строчными буквами $x, y, z, \ldots$ или $x_1, x_2, x_3, \ldots$ обозначают возможные значения СВ (как элементы множеств, которыми являются СВ в целом). Конкретное числовое значение x, которое принимает СВ X в каком-либо конкретном испытании, называется *реализацией* этой СВ в данном испытании.

Пример 1.

1. Случайная величина X - количество выпавших гербов при однократном подбрасывании пяти - монет может принимать значения из конечного множества $X = \{0, 1, 2, 3, 4, 5\}$ (может не выпасть ни один герб, может выпасть ровно один, могут выпасть два, три, четыре или пять гербов);
2. Случайная величина Y - количество подбрасываний игрального кубика до первого появления «шестерки» может принимать значения из бесконечного множества $Y = \{1, 2, 3, 4, 5, 6, 7 \ldots\}$ («шестерка» может выпасть при первом же подбрасывании, а может не выпадать как угодно долго);
3. Поезда метро идут с интервалом в 5 минут. Случайная величина Z - время ожидания пассажиром поезда метро может принимать значения из конечного интервала $Z = [0;\ 5]$, содержащего бесконечное множество действительных чисел.

Замечание. Если случайным событиям поставлены в соответствие числа, то между ними и соответствующими значениями СВ можно установить взаимно однозначное соответствие

Пример 2. В схеме испытаний с однократным подбрасыванием игрального кубика обозначим событие $A_i = \{выпадение\ i\ очков\ на\ верхней\ грани\ кубика\}$, $i \in \{1,2,3,4,5.6\}$. Пусть случайная величина X - количество очков, выпавших на верхней грани кубика. Каждому случайному событию A_i, $i \in \{1,2,3,4,5.6\}$ сопоставим значение случайной величины $X = i$ (событию A_1 .соответствует 1, событию A_2 соответствует 2 и т.д.).

Определение. Множество значений, которые может принимать СВ X (не обязательно дискретная), называется *множеством возможных значений* этой СВ.

Различают дискретные и непрерывные СВ.

Случайная величина называется *дискретной случайной величиной (ДСВ)*, если множество ее возможных значений состоит из отдельных изолированных действительных чисел.

В приложениях и справочниках множество изолированных возможных значений ДСВ часто представляют в виде натурального ряда с нулем $\{0,1,2,3,...\}$. Это обусловлено тем, что между указанными множествами имеет место взаимно однозначное соответствие.

Всюду в дальнейшем запись $X = x$ будет означать, что СВ X принимает конкретное числовое значение x, а запись $X < x$ будет означать, что СВ принимает любое значение из множества возможных значений, меньшее, чем выбранное конкретное числовое значение x. Запись $p(x)$ будет обозначать вероятность того, что СВ X принимает конкретное числовое значение x, а запись $P(X < x)$ будет означать, что СВ принимает любое значение из множества возможных значений, меньшее, чем конкретное числовое значение x.

4.2 **Способы задания дискретной случайной величины**

Наиболее полную информацию о случайной величине дает закон распределения.

Определение. *Законом распределения* называется зависимость между возможными значениями СВ и вероятностями, соответствующими этим значениям.

Для ДСВ закон распределения может быть задан:

- в виде ряда распределения (в табличной форме);
- в виде графика зависимости между возможными значениями СВ и вероятностями, соответствующими этим значениям;
- в аналитической форме (в виде формулы, связывающей возможные значения СВ с вероятностями, соответствующими этим значениям);
- в виде функции распределения.

Ряд распределения устанавливает связь между каждым возможным значением x_i ДСВ и вероятностью его реализации:

$X = x_i$	x_1	x_2	$\cdots$	x_n
$P(X = x_i)$	p_1	p_2	$\cdots$	p_n

Для вероятностей, входящих в ряд распределения, должно выполняться *условие нормировки*

$$\sum_{i=1}^{n} p_i = 1.$$

Замечание. Термин «распределение» введен в связи с тем, что суммарная вероятность $\sum_{i=1}^{n} p_i = 1$ по тому или иному закону распределена между всеми возможными значениями $x_1, x_2, ...x_n$ случайной величины X.

При графическом задании ДСВ X на координатной плоскости отмечаются точки $M_i(x_i, p_i)$, которые последовательно соединяются отрезками прямых (Рис. 28.1). Полученная ломаная называется многоугольником (полигоном) распределения ДСВ X.

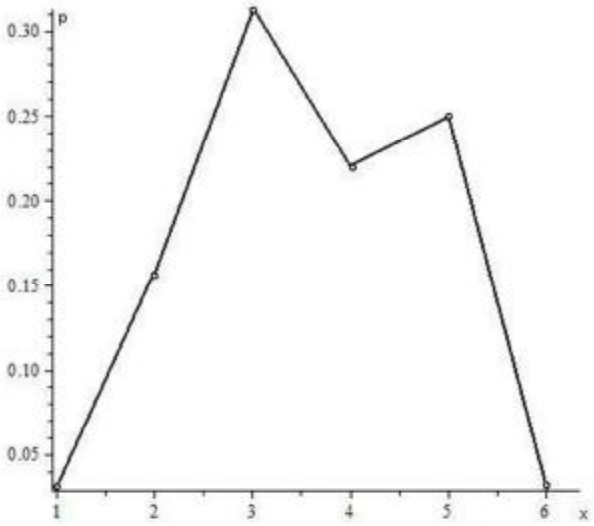


Рис. 28.1 Пример многоугольника распределения

При аналитическом способе задания ДСВ ее закон распределения представляется в виде некоторой математической формулы, позволяющей вычислить вероятности любых значений ДСВ.

Пример 3. Формула Бернулли $P_n(x) = C_n^x p^x (1-p)^{n-x}$ позволяет при заданном количестве независимых единичных испытаний n и заданной вероятности простого события («успеха») p в единичном испытании найти вероятность любого количества «успехов» x, $x \in \{0,1,2,...n\}$, то есть полностью задать закон распределения случайной величины X - количества «успехов» в серии из n независимых единичных испытаний.

Дискретная случайная величина X может быть задана в виде функции распределения $F_X(x) = P(X < x) = \sum_{x_i < x} p(x_i)$.

Вероятностный смысл функции распределения заключается в том, что при любом известном (выбранном, заданном) числовом значении x случайной величины X $(x \in X)$ функция распределения равна вероятности того, что СВ X принимает значение $x_i < x$.При этом заданная числовая величина x может не принадлежать ДСВ X .

Функция распределения дискретной случайной величины обладает свойствами:

1. $0 \le F_X(x) \le 1$;
2. $F_X(x)$ - неубывающая функция, то есть из неравенства $x_2 > x_1$ следует $F_X(x_2) \ge F_X(x_1)$;
3. $\lim_{x \to -\infty} F_X(x) = 0$ и $\lim_{x \to \infty} F_X(x) = 1$;
4. В окрестности каждой точки разрыва x_i (при их наличии) функция распределения непрерывна слева $F_X(x_i) = \lim_{x \to x_0 - 0} F(x) = F(x_i - 0)$
5. Функция распределения является кусочно-постоянной (ступенчатой) на интервалах, в которых возможные значения ДСВ отсутствуют..

Замечание. Величина скачка (разности между правым и левым пределами) функции $F_X(x)$ в точке разрыва x_i равна вероятности того, что СВ примет значение x_i: $\lim_{x \to x_i + 0} F_X(x) - \lim_{x \to x_i - 0} F_X(x) = P(X = x_i)$.Общее число скачков равно числу возможных значений ДСВ.

Пусть ДСВ задана своим рядом распределения:

X	x_1	x_2	x_3	$\cdots$	x_n
$P(X = x_i)$	p_1	p_2	p_3	$\cdots$	p_n

Тогда, используя свойства функции распределения, ее можно в аналитической форме представить так:

$$F_X(x)=\begin{cases} 0 \ при \ x \le x_1 \\ p_1, \ при \ x \in (x_1; \ x_2] \\ p_1+p_2, \ при \ x \in (x_2, \ x_3] \\ p_1+p_2+p_3, \ при \ x \in (x_3; \ x_4] \\ \ldots \\ p_1+p_2+p_3+\ldots+p_n=1, \ при \ x > x_n \end{cases}$$

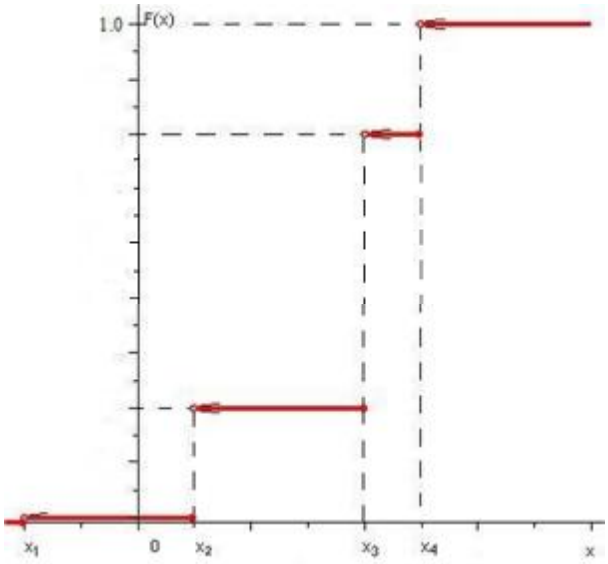


Рис. 28. 2 Пример графика функции распределения ДСВ

График функции распределения будет иметь вид (Рис. 28.2):

Замечание. Функция распределения сохраняет свое значение при подходе к точке разрыва слева и в самой точке разрыва (это отмечается на графике точкой). Затем функция $F(x)$ скачком изменяет свое значение и сохраняет новое значение при подходе к точке разрыва справа (это отмечается на графике стрелкой).

Все способы задания закона распределения ДСВ взаимосвязаны, что позволяет, при необходимости, переходить от одного способа задания к другому. Рассмотрим способы задания ДСВ на примерах.

Пример 4. ДСВ X задана в аналитическом виде $p(X=x)=\dfrac{C_5^x C_3^{4-x}}{C_8^4}$. Требуется:

- составить ряд распределения ДСВ X ;

- построить многоугольник распределения;

- найти функцию распределения и построить ее график;

Решение. Находим конечное множество значений ДСВ X: Из выражения C_5^x следует, что $x \in \{0,1,2.3.4\}$. Подставляя полученные значения в выражение C_3^{4-x}, видим, что при $x=0$ выражение $C_3^{4-x}=C_3^4$ не имеет смысла. Значит, $X \in \{1,2.3.4\}$.

а) Находим вероятности $p(X=1)$, $p(X=2)$, $p(X=3)$ и $p(X=4)$:

$$p(X=1)=\frac{C_5^1C_3^3}{C_8^4}=5\cdot1\cdot\frac{4!\cdot4!}{8!}=\frac{1}{14};\qquad p(X=2)=\frac{C_5^2C_3^2}{C_8^4}=\frac{5!}{2!\cdot3!}\cdot3\frac{4!\cdot4!}{8!}=\frac{3}{7},$$

$$p(X=3)=\frac{C_5^3C_3^1}{C_8^4}=\frac{5!}{3!\cdot2!}\cdot3\frac{4!\cdot4!}{8!}=\frac{3}{7};$$

$p(X=4)=\frac{C_5^4C_3^0}{C_8^4}=5\cdot1\cdot\frac{4!\cdot4!}{8!}=\frac{1}{14}$. Запишем ряд распределения:

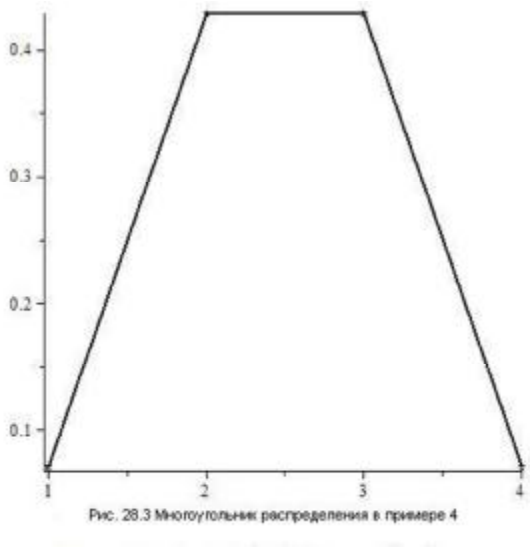


Рис. 28.3 Многоугольник распределения в примере 4

$X=x_i$	1	2	3	4
$P(X=x_i)$	$\frac{1}{14}$	$\frac{3}{7}$	$\frac{3}{7}$	$\frac{1}{14}$

Проверим условие нормировки: $\frac{1}{14}+\frac{3}{7}+\frac{3}{7}+\frac{1}{14}=\frac{1+6+6+1}{14}=\frac{14}{14}=1$. Условие выполнено.

б) по данным ряда распределения строим многоугольник распределения (Рис. 28.3).

в) Находим функцию распределения и строим ее график (Рис. 28.4)

$$F_X(x)=\begin{cases}0,\ при\ x\le1\\ \frac{1}{14},\ при\ x\in(1;\,2]\\ \frac{1}{14}+\frac{3}{7}=\frac{1}{2},\ при\ x\in(2;\,3]\\ \frac{1}{2}+\frac{3}{7}=\frac{13}{14},\ при\ x\in(3;\,4]\\ 1,\ при\ x>4\end{cases}$$

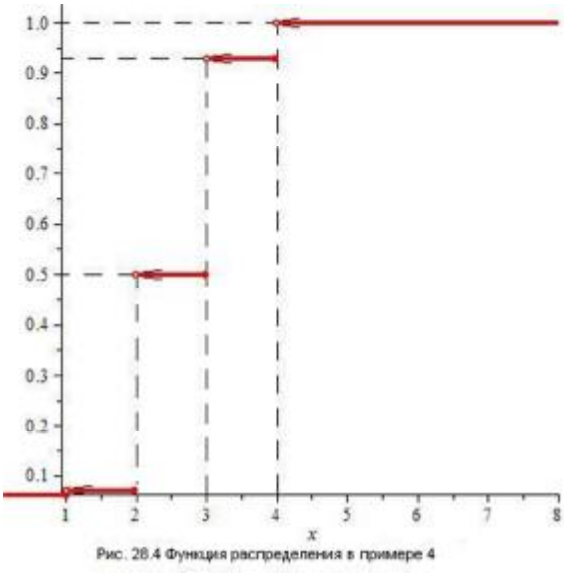


Рис. 28.4 Функция распределения в примере 4

Пример 5. Прибор состоит из трех блоков, которые работают и отказывают независимо друг от друга. Вероятности отказов за время T равны соответственно $p_1=0{,}1$, $p_2=0{,}2$ и $p_3=0{,}25$. Найти ряд распределения СВ, характеризующей число отказавших блоков.

Решение. Обозначим через X рассматриваемую СВ, тогда $X=\{0,1,2,3\}$.

Возможные значения СВ X соответствуют сложным событиям (не отказал ни один блок, отказал ровно один блок и т.д.), Поэтому введем простые события $A_i=\{отказал\ i-й\ блок\}$, $i=1,2,3$ и получим $P(A_1)=0{,}1$, $P(A_2)=0{,}2$, $P(A_3)=0{,}25$, $P(\overline{A}_1)=0{,}9$, $P(\overline{A}_2)=0{,}8$, $P(\overline{A}_3)=0{,}75$.

События A_1, A_2, A_3 независимы по условию задачи, значит,

$p_0 = P(X=0) = P(\overline{A}_1\overline{A}_2\overline{A}_3) = P(\overline{A}_1)P(\overline{A}_2)P(\overline{A}_3) = 0{,}9 \cdot 0{,}8 \cdot 0{,}75 = 0{,}54$;

$p_1 = P(X=1) = P(A_1\overline{A}_2\overline{A}_3 + \overline{A}_1A_2\overline{A}_3 + \overline{A}_1\overline{A}_2A_3) = 0{,}1 \cdot 0{,}8 \cdot 0{,}75 + 0{,}9 \cdot 0{,}2 \cdot 0{,}75 +$

$+ 0{,}9 \cdot 0{,}8 \cdot 0{,}25 = 0{,}375$;

$p_2 = P(X=2) = P(A_1A_2\overline{A}_3 + \overline{A}_1A_2A_3 + A_1\overline{A}_2A_3) = 0{,}1 \cdot 0{,}2 \cdot 0{,}75 + 0{,}9 \cdot 0{,}2 \cdot 0{,}25 +$

$+ 0{,}1 \cdot 0{,}8 \cdot 0{,}25 = 0{,}08$;

$p_3 = P(X=3) = P(A_1A_2A_3) = 0{,}1 \cdot 0{,}2 \cdot 0{,}25 = 0{,}005$.

Рассмотренные сложные события являются исходными и образуют полную группу, так как сумма вероятностей этих событий равна 1:: $p_0 + p_1 + p_2 + p_3 = 0{,}54 + 0{,}375 + 0{,}08 + 0{,}005 = 1$.

Цифровые индексы вероятностей указанных событий, примем за возможные значения ДСВ, характеризующей эти события и составим ряд распределения:

$X = x_i$	0	1	2	3
p_i	0,54	0,375	0,08	0,005

4.3 **Определение вероятностей ДСВ**

Вероятности реализации любых возможных значений ДСВ содержатся в ряде распределения этой ДСВ. Используя эти данные и свойства функции распределения, вычисляют вероятности ДСВ в различных частных случаях:

1. Из определения функции распределения $F_X(x)$ следует, что

$$P(X < x) = F_X(x) \qquad (1)$$

2. Вероятность попадания значения ДСВ в полуинтервал $[\alpha;\ \beta)$ равна разности значений функции распределения на концах этого полуинтервала $[\alpha;\ \beta)$:

$$P(\alpha \le X < \beta) = F_X(\beta) - F_X(\alpha) \qquad (2)$$

3. Вероятность того, что ДСВ примет определенное возможное значение $X = x_i$, определяют так:

$$P(X = x_i) = \begin{cases} p_i, \text{ если } x_i \in X \\ 0, \text{ если } x_i \notin X \end{cases} \qquad (3)$$

4. Значение, соответствующее неравенствам, определяют по формулам:

$$P(\alpha \le X \le \beta) = F_X(\beta) - F_X(\alpha) + P(X = \beta) \quad (4)$$

$$P(X \le x_i) = F_X(x_i) + p_i \quad (5)$$

$$P(X \ge x_i) = 1 - F_X(x_i) \quad (6)$$

$$P(X > x_i) = 1 - F_X(x_i) - p_i \quad (7)$$

$$P(\alpha < X < \beta) = F_X(\beta) - F_X(\alpha) - P(X = \alpha) \quad (8)$$

$$P(\alpha < X \le \beta) = F_X(\beta) - F_X(\alpha) + P(X = \beta) - P(X = \alpha) \quad (9)$$

Пример 6. Используя результаты Примера 4 (28.2), найти вероятности следующих реализаций ДСВ X: а) $X \in \left(\frac{1}{2}; \frac{5}{2}\right)$; б) $X < 1{,}5$; в) $X \le 2$; г) $X \ge 2{,}5$; д) $X > 1{,}5$;

Решение. а) По формуле (8) получаем:

$$P\left(X \in \left(\frac{1}{2}; \frac{5}{2}\right)\right) = F_X\left(\frac{5}{2}\right) - F_X\left(\frac{1}{2}\right) - P\left(X = \frac{1}{2}\right) = \frac{1}{2} - 0 - 0 = \frac{1}{2};$$

б) По формуле (1) получаем $P(X < 1{,}5) = F_X(1{,}5) = \frac{1}{14}$;

в) По формуле (5) получаем: $P(X \le 2) = F_X(2) + P(X = 2) = \frac{1}{14} + \frac{3}{7} = \frac{1}{2}$;

г) По формуле (6) получаем: $P(X \ge 2{,}5) = 1 - F_X(2{,}5) = 1 - \frac{1}{2} = \frac{1}{2}$;

д) По формуле (7) получаем $P(X > 1{,}5) = 1 - F_X(1{,}5) - P(X = 1{,}5) = 1 - \frac{1}{14} - 0 = \frac{13}{14}$;

4.4 **Арифметические операции над дискретными независимыми СВ**

Определение. Две случайные величины X и Y называются независимыми, если вероятности реализаций значений одной из них не зависят от того, какие реализации принимает другая.

Независимость случайных величин следует из независимости событий, которые характеризуют эти случайные величины.

Пусть независимые ДСВ X и Y заданы своими рядами распределения:

$X = x_i$	x_1	x_2	$\cdots$	x_n

$P(X=x_i)$	p_1	p_2	$\cdots$	p_n

$Y=y_j$	y_1	y_2	$\cdots$	y_m
$P(Y=y_j)$	q_1	q_2	$\cdots$	q_m

Рассмотрим основные способы выполнения математических операций над ДСВ.

1. Произведением СВ X на постоянный коэффициент k называется новая СВ Z, которая принимает возможные значения $z_i = kx_i$ с теми же вероятностями p_i, что и СВ X.
2. Для нахождения ряда распределения новой случайной величины $Z = X \pm Y$, $Z = X \cdot Y$, необходимо:

2.1 Вычислить значения $z_{1j} = f(x_1, y_j)$ при каждом $j = 1,...,m$; затем вычислить значения $z_{2j} = f(x_2, y_j)$ при каждом $j = 1,...,m$; и т.д. до всех значений $z_{nj} = f(x_n, y_j)$ при каждом $j = 1,...,m$;

2.2 Вероятность каждого значения $z_{1j} = f(x_1, y_j)$ вычисляется как произведение $p_i \cdot q_j$, $i = 1,...,n$, $j = 1,...,m$;

2.3. При наличии одинаковых значений $z_{1j} = f(x_1, y_j)$ они заменяются одним значением, с вероятностью, равной сумме вероятностей этих значений ;

Пример 7. Независимые случайные величины X и Y заданы своими рядами распределения:

$X=x_i$	10	15	20	25
$P(X=x_i)$	0,1	0,2	0,4	0,3

$Y=y_j$	5	10	15
$P(Y=y_j)$	0,2	0,3	0,5

Составить ряд распределения случайной величины $Z = 2X + Y$

Решение. Вычисления ведем в табличной форме, используя равенства $z_{ij} = 2x_i + y_j$, $p_{ij} = p(x_i) \cdot p(y_j)$

Для первого значения СВ $X = 10$ найдем суммы $2X + Y$ с каждым значением СВ Y и перемножим соответствующие вероятности:

Z	25	30	35
$p(z_{ij})$	0,02	0,03	0,05

Для второго значения СВ $X = 15$ найдем суммы $2X + Y$ с каждым значением СВ Y, перемножим соответствующие вероятности и добавим в таблицу:

Z	25	30	35	35	40	45
$p(z_{ij})$	0,02	0,03	0,05	0,04	0,06	0,1

Для третьего значения СВ $X = 20$ найдем суммы $2X + Y$ с каждым значением СВ Y, перемножим соответствующие вероятности и добавим в таблицу:

Z	25	30	35	35	40	45	45	50	55
$p(z_{ij})$	0,02	0,03	0,05	0,04	0,06	0,1	0,08	0,12	0,2

Для четвертого значения СВ $X = 25$ найдем суммы $2X + Y$ с каждым значением СВ Y, перемножим соответствующие вероятности и добавим в таблицу:

Z	25	30	35	35	40	45	45	50	55	55	60	65
$p(z_{ij})$	0,02	0,03	0,05	0,04	0,06	0,1	0,08	0,12	0,2	0,06	0,09	0,15

Оставляем в верхней строке указанного ряда только различные значения СВ $Z = 2X + Y$, и, складывая соответствующие вероятности, получаем дискретный ряд распределения:

Z	25	30	35	40	45	50	55	60	65
$p(z_{ij})$	0,02	0,03	0,09	0,06	0,18	0,12	0,26	0,09	0,15

Проверим условие нормировки: 0,02+0,03+0,09+0,06+0,18+0,12+0,26+0,09+0,15=1.

Пример 8. Независимые случайные величины X и Y заданы своими рядами распределения:

$X = x_i$	0	2	4
$P(X = x_i)$	0,5	0,2	0,3

$Y = y_j$	-2	0	2
$P(Y = y_j)$	0,1	0,6	0,3

Составить ряд распределения случайной величины $Z = X \cdot Y$

Решение. Для первого значения СВ $X = 0$ найдем произведения XY с каждым значением СВ Y, затем для второго значения СВ $X = 2$ найдем произведения XY с каждым значением СВ Y и, наконец, для третьего значения СВ $X = 2$ найдем произведения XY с каждым значением СВ Y.

$Z = X \cdot Y$	0	0	0	- 4	0	4	-8	0	8
$p(z_{ij})$	0,05	0,3	0,15	0,02	0,12	0,06	0,03	0,18	0,09

Оставляем в верхней строке таблицы только различные значения СВ $Z = X \cdot Y$, складывая соответствующие вероятности:

$Z = z_k$	- 8	- 4	0	4	8
$p(z_k)$	0,03	0,02	0,8	0,06	0,09

Проверим условие нормировки: 0,03+0,02+0,8+0,06+0,09=1.

При возведении СВ X в степень m, получаем новую СВ $Z = X^m$, которая принимает возможные значения $z_i = x_i^m$ с теми же вероятностями p_i, что и возможные значения $x_i \in X$. Если при возведении в четную степень получают несколько одинаковых возможных значений новой СВ Z, то оставляют одно такое значение с вероятностью, равной сумме соответствующих вероятностей.

Пример 9. Составить ряд распределения для СВ $Z = X^2$, если известен ряд распределения для СВ X:

$X = x_i$	-3	2	3
$P(X = x_i)$	0,5	0,3	0,2

Решение. Возможные значения СВ X после возведения в квадрат будут равны $x_1^2 = (-3)^2 = 9$, $x_2^2 = 2^2 = 4$, $x_3^2 = 3^2 = 9 = x_1^2$. Поэтому для новой СВ остаются два возможных значения $z_1 = 4$ с вероятностью 0,3 и $z_2 = 9$ с вероятностью $0{,}5 + 0{,}2 = 0{,}7$. В результате получаем

$Z = z_k$	Z	4	9
$p(z_k)$	$P(Z = z_i)$	0,3	0,7

4.5 **Числовые характеристики ДСВ**

Ряд распределения, многоугольник распределения или функция распределения дают наиболее полную информацию о случайной величине. Однако сравнивать и оценивать СВ лучше по ее отдельным, наиболее характерным, свойствам. Например, стрелков оценивают в первом приближении по среднему количеству выбитых очков в одинаковых сериях выстрелов. При этом законы распределения у них обычно существенно различны.

Для количественного описания различных существенных свойств закона распределения случайных величин используют числовые характеристики, вычисляемые по данным ряда распределения.

Числовые характеристики СВ – это постоянные числовые величины, характеризующие отдельные свойства закона распределения СВ.

Формулы для вычисления одной и той же числовой характеристики случайных величин могут быть различными, а вероятностный смысл – один и тот же. Именно вероятностный смысл обеспечивает понимание сущности результатов вычислений.

Основными числовыми характеристиками СВ являются:

- математическое ожидание;

-дисперсия (или среднее квадратичное отклонение – СКО);

Вероятностный смысл математического ожидания СВ – это ожидаемое среднее значение случайной величины, определенное с учетом вероятностей появлений (реализаций) возможных значений этой величины

Определение. Пусть ДСВ задана рядом распределения

$X = x_i$	x_1	x_2	$\cdots$	x_n

$P(X=x_i)$	p_1	p_2	$\cdots$	p_n

тогда *математическое ожидание* ДСВ X равно $M(X)=x_1p_1+x_2p_2+...+x_np_n$ (1)

Математическое ожидание обладает следующими важными для вычислений свойствами,

1. $M(X)$ - неслучайная величина;
2. Математическое ожидание постоянной величины равно этой постоянной величине: $M(C)=C$;
3. Постоянный множитель можно выносить за знак математического ожидания: $M(CX)=CM(X)$;
4. Математическое ожидание суммы конечного числа случайных величин равно сумме математических ожиданий этих величин: $M(X_1\pm X_2\pm...\pm X_n)=M(X_1)\pm M(X_2)\pm...\pm M(X_n)$;
5. Математическое ожидание произведения конечного числа независимых случайных величин равно произведению математических ожиданий этих величин: $M(X_1\cdot X_2\cdot...\cdot X_n)=M(X_1)\cdot M(X_2)\cdot...\cdot M(X_n)$;
6. Если все значения СВ уменьшить или увеличить на одно и то же число $C=const$ то математическое ожидание этой СВ уменьшится (увеличится) на то же число: $M(X\pm C)=M(X)\pm C$

Второй основной числовой характеристикой СВ является дисперсия. Вероятностный смысл дисперсии заключается в том, что она характеризует в количественной усредненной форме степень рассеивания возможных значений СВ относительно ее математического ожидания (центра рассеивания).

Для практических вычислений дисперсии ДСВ обычно используют формулу

$D(X)=M(X^2)-(M(X))^2$ (3)

Замечание 1. Величина $M(X^2)$, называемая моментом второго порядка СВ X вычисляется как математическое ожидание случайной величины $X^2=\{x_1^2,x_2^2,...,x_n^2\}$: $M(X^2)=x_1^2p_1+x_2^2p_2+...+x_n^2p_n$.

Дисперсия обладает следующими важными для вычислений свойствами,

1. $D(X)$ - неслучайная величина;
2. Дисперсия любой СВ неотрицательна: $D(C) \geq 0$
3. Дисперсия постоянной величины равна нулю: $D(C) = 0$;
4. Постоянный множитель можно выносить за знак дисперсии, возводя его в квадрат: $D(CX) = C^2 D(X)$;
5. Дисперсия суммы конечного числа независимых случайных величин равна сумме дисперсий этих величин: $D(X_1 + X_2 + ... + X_n) = D(X_1) + D(X_2) + ... + D(X_n)$;
6. Если все значения СВ уменьшить или увеличить на одно и то же число $C = const$ то дисперсия этой СВ не изменится: $D(X \pm C) = D(X)$;
7. Дисперсия разности СВ равна сумме дисперсий этих СВ: $D(X_1 - X_2) = D(X_1) + D(X_2)$;

В случае СВ, имеющих размерность, широко используемых в прикладных задачах, применение дисперсии как характеристики рассеивания СВ неудобно, так как размерности СВ и ее дисперсии различны (размерность дисперсии в два раза больше размерности СВ). Поэтому в приложениях вместо дисперсии часто используют *среднее квадратичное отклонение (СКО)*, вычисляемое как арифметический квадратный корень из дисперсии: $\sigma(X) = \sqrt{D(X)}$..

Рассмотрим примеры вычисления числовых характеристик ДСВ:

Пример 10. Дан ряд распределения ДСВ:

$X = x_i$	1	2	3	4
$P(X = x_i)$	$\frac{1}{14}$	$\frac{3}{7}$	$\frac{3}{7}$	$\frac{1}{14}$

Найти $M(X)$, $D(X)$, $\sigma(X)$.

Решение. Математическое ожидание: $M(X) = 1 \cdot \frac{1}{14} + 2 \cdot \frac{3}{7} + 3 \cdot \frac{3}{7} + 4 \cdot \frac{1}{14} = 2{,}5$.

Дисперсия: $M(X^2) = 1^2 \cdot \frac{1}{14} + 2^2 \cdot \frac{3}{7} + 3^2 \cdot \frac{3}{7} + 4^2 \cdot \frac{1}{14} = \frac{1 + 24 + 54 + 16}{14} = \frac{95}{14} \approx 6{,}7857$;

$D(X) = M(X^2) - (M(X))^2 = \frac{95}{14} - \left(\frac{5}{2}\right)^2 = \frac{190 - 175}{28} = \frac{15}{28} \approx 0{,}5357$;

СКО: $\sigma(X)=\sqrt{D(X)}=\sqrt{\frac{15}{28}}=\frac{1}{2}\sqrt{\frac{15}{7}}\approx 0{,}7319$.

Пример 11. Известен закон распределения вагонов по массе (в тоннах) в составе:

$X=x_i$	40	50	60	70	80
$P(X=x_i)$	0,15	0,1	0,35	0,3	0,1

Найти среднюю массу отцепа из трех вагонов и СКО в тоннах

Решение. В прикладных вероятностных задачах средняя величина интерпретируется как математическое ожидание. В данном примере средняя масса одного вагона равна $M(X)=40\cdot 0{,}15+50\cdot 0{,}1+60\cdot 0{,}35+70\cdot 0{,}3+80\cdot 0{,}1=61m$. Средняя масса отцепа из трех вагонов равна $M(3X)=3M(X)=3\cdot 61=183m$.

$M(X^2)=40^2\cdot 0{,}15+50^2\cdot 0{,}1+60^2\cdot 0{,}35+70^2\cdot 0{,}3+80^2\cdot 0{,}1=3860$;

Дисперсия: $D(X)=M(X^2)-(M(X))^2=3860-61^2=139m^2$;

СКО: $\sigma(3X)=\sqrt{D(3X)}=\sqrt{3^2D(X)}=3\sqrt{D(X)}=3\sqrt{139}\approx 35{,}37m$.

Пример 12. Найти математическое ожидание произведения числа очков на верхних гранях при бросании двух игральных кубиков.

Решение. Пусть СВ X - количество очков на верхней грани первого кубика, Y - количество очков на верхней грани второго кубика. Очевидно, что случайные величины X и Y - независимы и имеют один и тот же ряд распределения:

$X=x_i,\ Y=y_j$	1	2	3	4	5	6
$P(X=x_i)=P(Y=y_j)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

$M(X)=M(Y)=\frac{1}{6}(1+2+3+4+5+6)=\frac{21}{6}=3{,}5$; $\quad M(XY)=M(X)\cdot M(Y)=(3{,}5)^2=12{,}25$.

Кроме математического ожидания, дисперсии и СКО рассматриваются (хотя и значительно реже) другие числовые характеристики, которые выражаются через моменты различных порядков..

Определение. *Моментом порядка* k ДСВ X называется $M(X^k)=x_1^k p_1+x_2^k p_2+\ldots+x_n^k p_n$.

Замечание 2. Момент первого порядка совпадает с математическим ожиданием СВ, момент второго порядка использовался выше для вычисления дисперсии. Моменты третьего и четвертого порядков можно использовать для вычисления асимметрии и эксцесса СВ.

Определение. *Асимметрией* (скошенностью распределения) ДСВ X называется неслучайная величина $A(X)=\dfrac{M(X-M(X))^3}{\sigma^3(X)}$.

Пользуясь свойствами математического ожидания, можно вывести более удобную формулу для вычисления асимметрии - $A(X)=\dfrac{M(X^3)-3M(X^2)M(X)+2(M(X))^3}{\sigma^3(X)}$.

Вероятностный смысл асимметрии понятен из ее названия: если многоугольник распределения ДСВ X имеет вертикальную ось симметрии (и только в этом случае) асимметрия равна нулю.

Определение. *Эксцессом* ДСВ X называется неслучайная величина $E(X)=\dfrac{M(X-M(X))^4}{\sigma^4(X)}-3$;

. Пользуясь свойствами математического ожидания, можно вывести более удобную формулу для вычисления эксцесса - $E(X)=\dfrac{M(X^4)-4M(X^3)M(X)+6M(X^2)(M(X))^2-3(M(X))^4}{\sigma^4(X)}-3$.

Замечание 3. Эксцесс («островершинность» или «плосковершинность» соответствующих графиков законов распределения) применительно к ДСВ практически не используется и применяется в учебных задачах для освоения техники вычислений.

Пример 13. Дан ряд распределения ДСВ:

$X=x_i$	1	2	3	4
$P(X=x_i)$	$\frac{1}{14}$	$\frac{3}{7}$	$\frac{3}{7}$	$\frac{1}{14}$

Найти асимметрию и эксцесс ДСВ X.

Решение. Математическое ожидание, момент второго порядка и СКО СВ X найдены в Примере 1: $M(X)=2{,}5$, $M(X^2)=\dfrac{95}{14}$, $\sigma(X)=\dfrac{1}{2}\sqrt{\dfrac{15}{7}}$. Найдем моменты третьего и четвертого порядков, необходимые для вычисления асимметрии и эксцесса:

$$M\left(X^3\right)=1^3\cdot\frac{1}{14}+2^3\cdot\frac{3}{7}+3^3\cdot\frac{3}{7}+4^3\cdot\frac{1}{14}=\frac{1+48+162+64}{14}=\frac{275}{14};$$

$$M\left(X^4\right)=1^4\cdot\frac{1}{14}+2^4\cdot\frac{3}{7}+3^4\cdot\frac{3}{7}+4^4\cdot\frac{1}{14}=\frac{1+96+486+256}{14}=\frac{839}{14}.$$

$$A(X)=\frac{\frac{275}{14}-3\cdot\frac{95}{14}\cdot\frac{5}{2}+2\cdot\left(\frac{5}{2}\right)^3}{\left(\frac{1}{2}\sqrt{\frac{15}{7}}\right)^3}=\frac{275\cdot2-15\cdot95+125\cdot7}{28}\cdot8\left(\sqrt{\frac{7}{15}}\right)^3=0,$$ что и следовало ожидать, так как многоугольник распределения данной СВ имеет вертикальную ось симметрии $x=2{,}5$ (Рис. 28.5).

$$E(X)=\frac{\frac{839}{14}-4\cdot\frac{275}{14}\cdot\frac{5}{2}+6\cdot\frac{95}{14}\cdot\left(\frac{5}{2}\right)^2-3\cdot\left(\frac{5}{2}\right)^4}{\left(\frac{15}{28}\right)^2}-3=\frac{87}{112}\cdot\frac{784}{225}-3=\frac{609-675}{225}\approx0{,}2933.$$

4.6 Биномиальный закон распределения

Пусть производятся многократные испытания в виде серий из n единичных независимых испытаний; при этом вероятность «успеха» в единичном испытании равна p. По формуле Бернулли (см. 27.5) вероятность ровно k успехов в серии из n единичных независимых испытаний равна $P_k=C_n^k p^k(1-p)^{n-k}$.

Рассмотрим случайную величину X - количество «успехов» в серии из n единичных независимых испытаний. Ее ряд распределения будет иметь вид

$X=k$	0	1	2	…	k	…	n
$P(X=k)$	P_0	P_1	P_2	…	P_k	…	P_n

где P_k - вероятность k-ого возможного значения СВ (вычисляется по формуле Бернулли).

Определение Дискретная СВ имеет биномиальное распределение с параметрами n и p, если вероятности ее возможных значений описываются формулой $P_k=C_n^k p^k(1-p)^{n-k}$.

На Рис. 28.5. представлены многоугольники распределения вероятностей для биномиальных распределений с одним для всех параметром n и тремя различными параметрами $p=0{,}2$,

$p = 0{,}5$ и $p = 0{,}8$. Из графиков видно, что при биномиальном распределении либо существует ровно одно наиболее вероятное событие (верхняя точка многоугольника распределения), либо существуют ровно два наиболее вероятных события с одинаковой вероятностью.

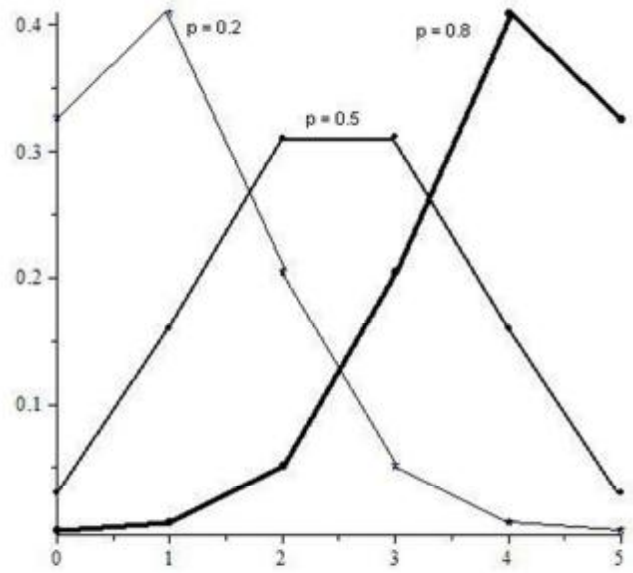


Рис. 28.5 Многоугольники биномиальных распределений вероятностей

Аналитически наиболее вероятное число появления «успехов» k_0 определяется из неравенств $np + p - 1 \le k_0 \le np + p$.

Пример 14. Найти наиболее вероятное число выпадений «гербов» в серии из 10 подбрасываний монеты.

Решение. $n = 10$, $p = 0{,}5$, Из формулы $np + p - 1 \le k_0 \le np + p$ получаем $10 \cdot 0{,}5 + 0{,}5 - 1 \le k_0 \le 10 \cdot 0{,}5 + 0{,}5$; $4{,}5 \le k_0 \le 5{,}5$. Поскольку k_0 - целое число, то $k_0 = 5$.

Пример 15. Вероятность появления события A в каждом единичном испытании в некоторой серии испытаний равна $p = 0{,}7$. Сколько таких испытаний должно быть в серии, чтобы наиболее вероятное число появлений события A было равно 20?

Решение. $p = 0{,}7$, $k_0 = 20$: $p(n+1) - 1 \le k_0 \le p(n+1)$; $0{,}7(n+1) - 1 \le 20 \le 0{,}7(n+1)$;

$\begin{cases} 0{,}7(n+1) - 1 \le 20 \\ 0{,}7(n+1) \ge 20 \end{cases}$; $\begin{cases} 0{,}7(n+1) \le 21 \\ 0{,}7(n+1) \ge 20 \end{cases}$; $\begin{cases} n+1 \le \dfrac{210}{7} = 30 \\ n+1 \ge \dfrac{200}{7} \approx 28{,}57 \end{cases}$; $\begin{cases} n \le 29 \\ n \ge 27{,}57 \end{cases}$, откуда получаем $n = 28$

.

Математическое ожидание и дисперсия ДСВ X, подчиненной биномиальному закону распределения, определяются по формулам $M(X) = np$, $D(X) = np(1-p)$.

Пример 16. На проверку перед сборкой поступили три прибора. вероятность того, что прибор пройдет проверку, равна $p = 0{,}3$. Составить ряд распределения вероятностей СВ X, характеризующей количество приборов, прошедших проверку. Найти среднее число приборов, прошедших проверку и СКО случайной величины X. Найти вероятность наиболее вероятного числа приборов, прошедших проверку.

Решение. Очевидно, что $X \in \{0,1,2,3\}$. Используя формулу Бернулли, составляем ряд распределения: $p_0 = C_3^0 p^0 (1-p)^3 = 0{,}343$, $p_1 = C_3^1 p^1 (1-p)^2 = 0{,}441$, $p_2 = C_3^2 p^2 (1-p)^1 = 0{,}189$, $p_3 = C_3^3 p^3 (1-p)^0 = 0{,}027$. Проверка условия нормировки: $0{,}343 + 0{,}441 + 0{,}189 + 0{,}027 = 1$.

$X = x_i$	0	1	2	3
$P(X = x_i)$	0,343	0,441	0,189	0,027

Среднее число приборов $n_{ср}$, прошедших проверку, определяется по математическому ожиданию СВ X: $M(X) = np = 3 \cdot 0{,}3 = 0{,}9$. После округления получаем . $n_{ср} = 1$

Проверка: $M(X) = \sum_{i=0}^{3} x_i p_i = 0 \cdot 0{,}343 + 1 \cdot 0{,}441 + 2 \cdot 0{,}189 + 3 \cdot 0{,}027 = 0{,}9$

Дисперсия СВ X равна $D(X) = np(1-p) = 0{,}9 \cdot 0{,}7 = 0{,}63$. СКО равно $\sigma(X) = \sqrt{D(X)} = \sqrt{0{,}63} \approx 0{,}8$.

Наиболее вероятное число приборов, прошедших проверку находим из ряда распределения: $k_0 = 1$, вероятность этого события равна 0,441. Проверка: $np + p - 1 \le k_0 \le np + p$; $0{,}9 + 0{,}3 - 1 \le k_0 \le 0{,}9 + 0{,}3$; $0{,}2 \le k_0 \le 1{,}2$; $k_0 = 1$.

Биномиальный закон распределения широко используется при статистическом контроле качества продукции и при математическом описании функционирования систем массового обслуживания (производство изделий на определенном оборудовании при постоянных технологических и организационных условиях, автозаправочные станции и другие системы обслуживания).

Дополнительные материалы о биномиальном распределении и его приложениях приведены в теме 30.

Тема 5 Непрерывные случайные величины

5.1 Исходные положения

Случайную величину X будем называть *непрерывной случайной величиной (НСВ)*, если она может принимать любые значения из конечного или бесконечного числового интервала.

Непрерывная случайная величина X может быть задана с помощью функции распределения ее возможных значений $F_X(x) = P(X < x)$.

Как и в случае ДСВ, вероятностный смысл функции распределения $F_X(x)$ заключается в том, что НСВ в результате испытаний может принять любые возможные значения меньше некоторого выбранного значения $x \in X$ ..

Замечание 1. Непрерывная случайная величина, как правило, может принимать любые числовые значения из действительного отрезка, полуинтервала или интервала (конечного или бесконечного), или из объединения таких интервалов.

Для большинства НСВ, имеющих важное прикладное значение, функция распределения является дифференцируемой на всей числовой прямой (возможно, за исключением конечного количества точек). Такие функции распределения, обычно называемые *интегральными функциями распределения*, будем рассматривать ниже.

Основные свойства интегральной функции распределения заключаются в следующем:

- все значения функции $F_X(x)$ принадлежат отрезку $[0; 1]$ (свойство ограниченности функции распределения); Если, кроме того, все возможные значения НСВ принадлежат отрезку $[a; b]$ или интервалу $(a; b)$, то $F_X(a) = 0$ и $F_X(b) = 1$;

- при $x_2 > x_1$ имеет место неравенство $F_X(x_2) \geq F_X(x_1)$, то есть функция $F_X(x)$ является неубывающей (свойство монотонности);

- вероятность попадания возможных значений НСВ в интервал $(x_1; x_1)$, полуинтервал $[x_1, x_2)$ или $(x_1, x_2]$ или отрезок $[x_1, x_2]$ равна приращению функции распределения на этом интервале (полуинтервале,отрезке):

$$P(x_1 < X < x_2) = P(x_1 < X \leq x_2) = P(x_1 \leq X < x_2) = P(x_1 \leq X \leq x_2) = F_X(x_2) - F_X(x_1)$$

(1)

Если функция распределения $F_X(x)$ дифференцируема, то это позволяет ввести *дифференциальную функцию распределения* (*плотность*

распределения вероятности, плотность вероятности) по формуле

$$f_X(x) = F_X'(x) \qquad (1)$$

Плотность вероятности обладает двумя характеристическими свойствами (первое свойство следует из неубывания функции распределения, второе – из предельных значений функции распределения):

1. $\forall x \in R \;\; f_X(x) \geq 0$;
2. $\int\limits_{-\infty}^{\infty} f_X(x)dx = 1$ (условие нормировки) (2)

Любая кусочно-непрерывная функция $f(x)$, удовлетворяющая этим двум свойствам является плотностью распределения некоторой непрерывной случайной величины.

Если все возможные значения СВ X принадлежат конечному интервалу $(a; b)$ или полуинтервалу $[a; b)$ $((a; b])$ или отрезку $[a; b]$, то формула (2) примет вид

$$\int\limits_{a}^{b} f_X(x)dx = 1 \qquad (3)$$

Вероятности нахождения возможных значений НСВ внутри промежутка (отрезка, полуинтервала, интервала) определяется по формуле

$$P(x_1 < X < x_2) = P(x_1 < X \leq x_2) = P(x_1 \leq X < x_2) = P(x_1 \leq X \leq x_2) = \int\limits_{x_1}^{x_2} f(x)dx \qquad (4)$$

Для нахождения функции распределения $F_X(x)$ по известной плотности $f_X(x)$ используется определенный интеграл с переменным верхним пределом

$$F_X(x) = \int_{-\infty}^{x} f_X(t)dt$$

(5)

Известно, что для неотрицательной функции $f_X(x)$, определенный интеграл с постоянными пределами $\int_{\alpha}^{\beta} f_X(x)dx$ численно равен площади плоской фигуры (криволинейной трапеции), ограниченной графиком функции $f_X(x)$, осью абсцисс и вертикальными прямыми $x = \alpha$, $x = \beta$.

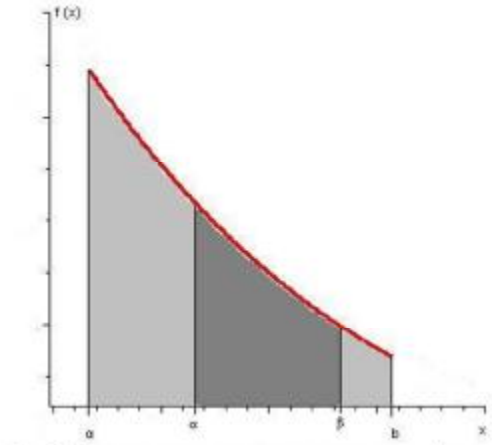


Рис. 29.1 Геометрическая интерпретация вероятности по графику плотности распределения вероятностей

На Рис. 29.1 показана общая площадь криволинейной трапеции (закрашена), равная 1 и площадь трапеции (закрашена более темным цветом), равная вероятности попадания случайной величины в интервал $(\alpha; \beta)$.

Пример 1. Известна плотность вероятности СВ X: $f_X(x) = \begin{cases} 0, \text{ при } x \notin \left[-\frac{\pi}{2}; \frac{\pi}{2}\right] \\ A\cos x, \text{ при } x \in \left[-\frac{\pi}{2}; \frac{\pi}{2}\right] \end{cases}$. Найти: а) значение коэффициента A; б) функцию распределения $F_X(x)$; в) вероятность попадания значения СВ X в интервал $\left(-\frac{\pi}{2}; 0\right)$. Построить графики плотности и функции распределения.

Решение. а) коэффициент A находим из условия нормировки $\int_{a}^{b} f_X(x)dx = 1$:

$A\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \cos x dx = 1$; $A \cdot \left(\sin x \Big|_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \right) = A(1-(-1)) = 2A = 1., \Rightarrow A = \frac{1}{2}$. В результате плотность

вероятностей задается формулой $f_X(x)=\begin{cases}\frac{1}{2}\cos x,\ при\ x\in\left[-\frac{\pi}{2};\frac{\pi}{2}\right]\\ 0,\ при\ x\notin\left[-\frac{\pi}{2};\frac{\pi}{2}\right]\end{cases}$ или

$$f_X(x)=\begin{cases}0,\ при\ x<-\frac{\pi}{2}\\ \frac{1}{2}\cos x,\ при\ x\in\left[-\frac{\pi}{2};\frac{\pi}{2}\right]\\ 0,\ при\ x>\frac{\pi}{2}\end{cases}.$$

б) Для определения функции распределения воспользуемся формулой (5). Поскольку плотность задана по-разному на интервалах $\left(-\infty;-\frac{\pi}{2}\right)$, $\left[-\frac{\pi}{2};\frac{\pi}{2}\right]$ и $\left(\frac{\pi}{2};\infty\right)$, то будем вычислять значения функции распределения в зависимости от того, какому из трех интервалов принадлежит значение переменной x. Пусть $x\in\left(-\infty;-\frac{\pi}{2}\right)$, тогда $F_X(x)=\int\limits_{-\infty}^{x}f_X(t)dt=\int\limits_{-\infty}^{x}0\cdot dt=0$. Пусть $x\in\left[-\frac{\pi}{2};\frac{\pi}{2}\right]$, тогда

$$F_X(x)=\int\limits_{-\infty}^{x}f_X(t)dt=\int\limits_{-\infty}^{-\frac{\pi}{2}}0\cdot dt+\int\limits_{-\frac{\pi}{2}}^{x}\frac{1}{2}\cos t dt=0+\frac{1}{2}\sin t\Big|_{-\frac{\pi}{2}}^{x}=\frac{1}{2}(\sin x+1).$$ Пусть $x\in\left(\frac{\pi}{2};\infty\right)$,

тогда $$F_X(x)=\int\limits_{-\infty}^{x}f_X(t)dt=\int\limits_{-\infty}^{-\frac{\pi}{2}}0\cdot dt+\int\limits_{-\frac{\pi}{2}}^{\frac{\pi}{2}}\frac{1}{2}\cos t dt+\int\limits_{\frac{\pi}{2}}^{\infty}0\cdot dt=0+1+0=1.$$ Объединяя результаты проведенных вычислений, получим

$$F_X(x)=\begin{cases}0,\ при\ x<-\frac{\pi}{2}\\ \frac{1}{2}(\sin x+1),\ при\ x\in\left[-\frac{\pi}{2};\frac{\pi}{2}\right]\\ 1,\ при\ x>\frac{\pi}{2}\end{cases}.$$

в) В силу четности функции $f_X(x)=\frac{1}{2}\cos x$, симметричности относительно начала координат отрезка $\left[-\frac{\pi}{2};\frac{\pi}{2}\right]$ и равенства единице площади,

заключенной под графиком плотности , очевидно имеем $P\left(-\frac{\pi}{2} < X < 0\right) = \frac{1}{2}$.

Проверка: $P\left(-\frac{\pi}{2} < X < 0\right) = F_X(0) - F_X\left(-\frac{\pi}{2}\right) = \frac{1}{2}(\sin 0 + 1) - 0 = \frac{1}{2}$.

График плотности распределения вероятностей $f_X(x) = \frac{1}{2}\cos x$ на отрезке $\left[-\frac{\pi}{2}; \frac{\pi}{2}\right]$ получается из известного графика $y = \cos x$ (см. Практикум Часть II, Приложение 1) сжатием в два раза по оси OY. На рис. 29.2 график плотности выделен жирной линией. Кроме того, на этом рисунке пунктирной линией показан известный график функции $y = \sin x$, который понадобится для построения графика функции распределения.

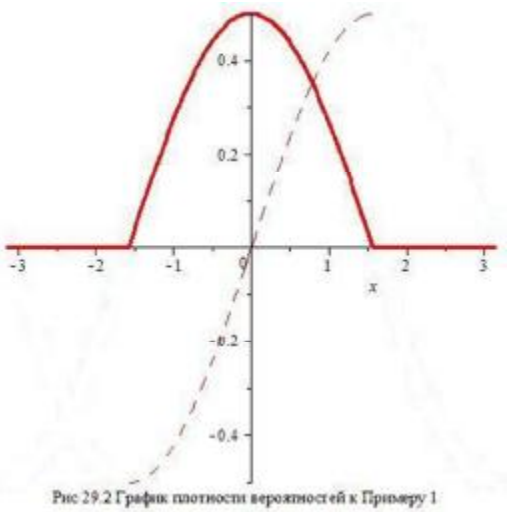


Рис 29.2 График плотности вероятностей к Примеру 1

График функции распределения $F_X(x) = \frac{1}{2}(\sin x + 1)$ на отрезке $\left[-\frac{\pi}{2}; \frac{\pi}{2}\right]$ получается из графика функции $y = \sin x$ сдвигом на 1 вверх по оси OY и последующим сжатием в два раза по той же оси (Рис. 29.3).

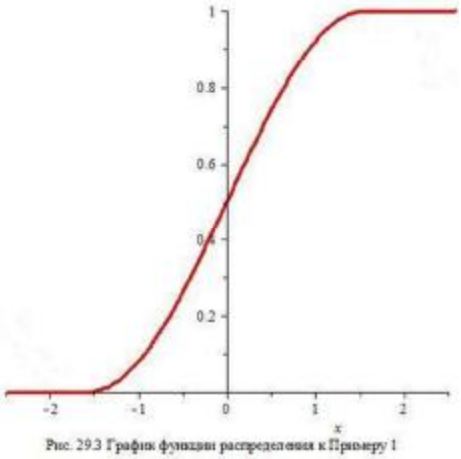


Рис. 29.3 График функции распределения к Примеру 1

5.2 Основные числовые характеристики непрерывных случайных величин

Для количественного описания существенных свойств законов распределения НСВ используют числовые характеристики, аналогичные характеристикам ДСВ.

Основными числовыми характеристиками являются:

- математическое ожидание;

- дисперсия (или СКО);

Математическое ожидание непрерывной случайной величины X с плотностью вероятности $f_X(x)$ определяется как неслучайная постоянная величина $M(X)=\int\limits_{-\infty}^{\infty} x\cdot f_X(x)dx$

Замечание 1. Если плотность вероятности $f_X(x)$ равна нулю всюду, кроме точек отрезка $[a;\,b]$, то $\int\limits_{-\infty}^{\infty} x\cdot f_X(x)dx=\int\limits_{a}^{b} x\cdot f_X(x)dx$;

Замечание 2. Вероятностный смысл математического ожидания НСВ аналогичен вероятностному смыслу математического ожидания ДСВ: оно равно среднему значению НСВ, определенному с учетом вероятностей реализации возможных значений этой случайной величины.

Вторая числовая характеристика НСВ – дисперсия – постоянная положительная величина, характеризующая в количественной усредненной форме степень рассеивания возможных значений СВ относительно ее математического ожидания (центра рассеивания).

Для практического вычисления дисперсии вместо формулы (2) обычно используют формулу: $D(X)=\int\limits_{-\infty}^{\infty} x^2\cdot f_X(x)dx-(M(X))^2$. Если плотность вероятностей $f_X(x)$ равна нулю всюду, кроме точек отрезка $[a;\,b]$, то дисперсию определяют так: $D(X)=\int\limits_{a}^{b} x^2\cdot f_X(x)dx-(M(X))^2$.

Все основные свойства математического ожидания и дисперсии, рассмотренные для ДСВ в разделе 28.5., справедливы и для непрерывных случайных величин. Среднее квадратичное отклонение определяется также как арифметический квадратный корень из дисперсии: $\sigma(X)=\sqrt{D(X)}$.

Дополнительные числовые характеристики, используемые для описания существенных особенностей законов распределения НСВ рассмотрены ниже в разделе 30.3..

Пример 1. Известна плотность вероятности СВ X: $f_X(x)=\begin{cases} 0, \text{ при } x\notin\left[-\frac{\pi}{2};\frac{\pi}{2}\right] \\ \frac{1}{2}\cos x, \text{ при } x\in\left[-\frac{\pi}{2};\frac{\pi}{2}\right] \end{cases}$. Найти: а) математическое ожидание; б) дисперсию и среднее квадратичное отклонение (СКВ);

Решение. а) В силу того, что площадь фигуры, заключенной между графиком плотности и осью абсцисс конечна (значит, математическое ожидание существует), а эта фигура имеет ось симметрии $x=0$ (пример 1, рис.29.2), математическое ожидание равно нулю. Проверка:

$$M(X)=\frac{1}{2}\int\limits_{-\frac{\pi}{2}}^{\frac{\pi}{2}} x\cdot\cos x dx=\frac{1}{2}\left(x\sin x+\cos x\right)\Big|_{-\frac{\pi}{2}}^{\frac{\pi}{2}}=0;$$

б) Момент второго порядка $M(X^2)=\frac{1}{2}\int\limits_{-\frac{\pi}{2}}^{\frac{\pi}{2}} x^2\cdot\cos x dx$ также вычислим интегрированием по частям:

$$M(X^2)=\frac{1}{2}\int\limits_{-\frac{\pi}{2}}^{\frac{\pi}{2}} x^2\cdot\cos x dx=\frac{1}{2}\left(x^2\sin x\Big|_{-\frac{\pi}{2}}^{\frac{\pi}{2}}-2\int\limits_{-\frac{\pi}{2}}^{\frac{\pi}{2}} x\sin x dx\right)=\ldots=\frac{\pi^2}{4}-2\approx 0{,}467;$$

Дисперсия: $D(X)=M(X^2)-(M(X))^2=\frac{\pi^2}{4}-2-0^2\approx 0{,}467$;

СКО: $\sigma(X)=\sqrt{D(X)}=\sqrt{\frac{\pi^2}{4}-2}\approx 0{,}683$;

Законы распределения НСВ

НСВ обычно задаются дифференциальным законом распределения (плотностью распределения вероятности).

В приложениях наиболее широко используются нижеследующие законы распределения НСВ:

- равномерное распределение;
- нормальное распределение;

- показательное распределение;

5.3 **Равномерное распределение**

Определение. НСВ подчинена *равномерному закону распределения* на отрезке $[a; b]$, если плотность распределения вероятности постоянна на этом отрезке и равна нулю вне этого отрезка.

Рис.29.4 График плотности равномерного закона распределения

Постоянная на отрезке $[a; b]$ плотность распределения $f(x) = C = const$ находится из условия нормировки:

Рис.29.5 График функции распределения равномерного закона распределения

$\int_a^b C dx = 1 \Rightarrow C = \frac{1}{b-a}$ и, следовательно, плотность равномерно распределенной СВ X задается формулой $f_X(x) = \begin{cases} 0, \ при \ x < a \\ \frac{1}{b-a}, \ при \ x \in [a; b] \\ 0, \ при \ x > b \end{cases}$.

Отсюда, в частности, следует, что закон равномерного распределения является двухпараметрическим (параметры a и b).

По формуле (5) раздела 29. 1 получаем аналитическое выражение для функции равномерного распределения на отрезке $[a; b]$: $F_X(x) = \int_a^x f_X(x)dx = \frac{x-a}{b-a}$. Затем представляем функцию распределения в полном виде с использованием общих свойств $F(x)$ (раздел 29.1):

$$F_X(x) = \begin{cases} 0, \ при \ x < a \\ \frac{x-a}{b-a}, \ при \ x \in [a; b] \\ 1, \ при \ x > b \end{cases}.$$

Графики плотности и функции распределения приведены, соответственно, на Рис.29.4 и Рис. 29.5.

Вероятность нахождения возможных значений НСВ с равномерным распределением в интервале (α, β) или на отрезке $[\alpha, \beta]$ при условии, что $[\alpha, \beta] \subseteq [a, b]$ определяют по формуле

$$P(\alpha < X < \beta) = P(\alpha \le X \le \beta) = F(\beta) - F(\alpha) = \frac{\beta - \alpha}{b - a}.$$

Находим основные числовые характеристики НСВ в случае равномерного распределения.

$$M(X) = \int_a^b xf(x)dx = \frac{1}{b-a}\int_a^b xdx = \frac{x^2}{2(b-a)}\Bigg|_a^b = \frac{(b^2 - a^2)}{2(b-a)} = \frac{a+b}{2}$$

Дисперсия равна $D(X) = \int_a^b x^2 \cdot \frac{1}{b-a}dx - \frac{(a+b)^2}{4} = ... = \frac{(b-a)^2}{12}$, и, соответственно,

$$\sigma(X) = \frac{b-a}{2\sqrt{3}}.$$

Равномерное распределение в прикладных задачах используется тогда, когда по условиям задачи можно считать, что вероятность попадания случайной точки на любой отрезок $[\alpha; \beta] \subseteq [a; b]$ зависит только от длины этого отрезка и не зависит от его расположения на отрезке $[a; b]$.

В частности, равномерный закон распределения используют при имитационном моделировании производственных систем (управление запасами, снабжение материалами и т.п.) Этот закон также лежит в основе методов генерирования случайных чисел.

Замечание 1. С законом равномерного распределения мы уже имели дело при изучении геометрических вероятностей (раздел 25.5, Замечание 2 и Пример 1).

Рассмотрим вероятностные задачи, в которых используется равномерный закон распределения.

Пример 1. В электрических часах минутная стрелка меняет положение скачком в конце каждой минуты. Найти вероятность того, что в данное мгновение время на часах будет отличаться от истинного времени не более, чем на 20 секунд.

Решение. Ошибку определения времени по часам считаем непрерывной СВ X, подчиненной равномерному закону распределения с параметрами $a = 0$, $b = 60$ (секунд). Отсюда следует, что плотность распределения задается формулой $f_X(x) = \begin{cases} 0, \text{ при } x < 0 \\ \dfrac{1}{60}, \text{при } x \in [0;60] \\ 0, \text{ при } x > 60 \end{cases}$. Пусть событие

$A = \{\text{в данный момент времени ошибка показаний часов не превышает } 20 \text{ секунд}\}$.

Событие A означает принадлежность случайной величины X интервалу $[0;\ 20]$, поэтому $P(A) = P(X \in [0;\ 20]) = \dfrac{1}{60}\int\limits_0^{20} dx = \dfrac{1}{60}(20-0) = \dfrac{1}{3}$.

Пример 2. На перекрестке движение автомобилей регулируется светофором. Красный свет включается на 2 минуты после зеленого и желтого. Время горения желтого света считаем пренебрежимо малым. Определите среднее время простоя автомобиля на перекрестке. Найдите вероятность того, что подъехавший в случайное время автомобиль будет ожидать на перекрестке не более 0,4 мин.

Решение. Принимаем, что время простоя на светофоре является непрерывной случайной величиной, подчиненной равномерному закону распределения с параметрами $a = 0$, $b = 2$. Среднее время простоя отождествляем с математическим ожиданием $M(X) = \dfrac{a+b}{2} = 1$. Вероятность ожидания на перекрестке не более 0,4 минут равно $P(X \in [1{,}6;\ 2]) = \dfrac{1}{2-0}\int\limits_{1{,}6}^{2} dx = \dfrac{1}{2}(2-1{,}6) = 0{,}8$.

5.4 **Нормальное распределение**

Нормальное распределение наиболее часто встречается в приложениях.

Такое распределение обычно формируется при суммарном влиянии на некоторую величину многих независимых случайных причин одного порядка.

Пример 1. Отклонения фактических размеров деталей от проектных при механической обработке на станках-автоматах обычно имеют нормальное распределение. Эти отклонения зависят от неточности установки детали,

неоднородности обрабатываемого материала, износа режущего инструмента, упруго-пластических деформаций в системе «станок-приспособление-инструмент-деталь» при действии сил резания, и т.д.

Нормальный закон распределения нередко называют законом ошибок и широко используют при расчетах случайных ошибок измерений.

Определение. Непрерывная случайная величина X имеет нормальное распределение (закон Гаусса), если плотность ее вероятности имеет вид

$$f_X(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-a)^2}{2\sigma^2}} \quad \text{при} \quad x\in(-\infty;\infty), \quad a,\sigma\in R, \quad \sigma>0. \tag{1}$$

Нормальное распределение является двухпараметрическим распределением с параметрами a и σ. При этом основные характеристики СВ совпадают с параметрами $M(X)=a$, $D(X)=\sigma^2$, $\sigma(X)=\sigma$. График функции $f_X(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-a)^2}{2\sigma^2}}$ (*кривой Гаусса*) называется нормальной кривой, которая обладает следующими свойствами:

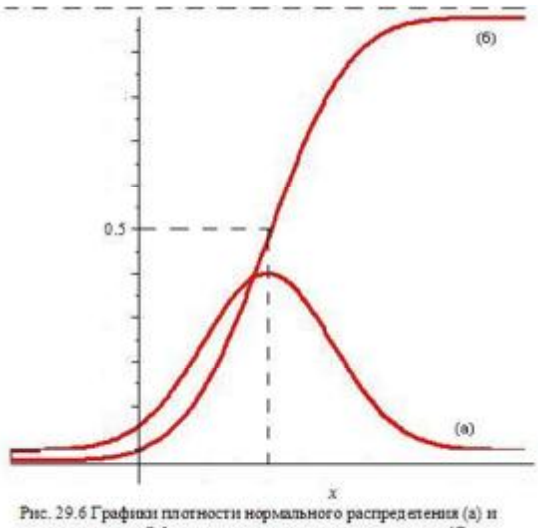


Рис. 29.6 Графики плотности нормального распределения (а) и интегральной функции нормального распределения (б)

- при всех возможных значениях НСВ кривая расположена выше оси Ox, являющейся горизонтальной асимптотой $(y=0)$;

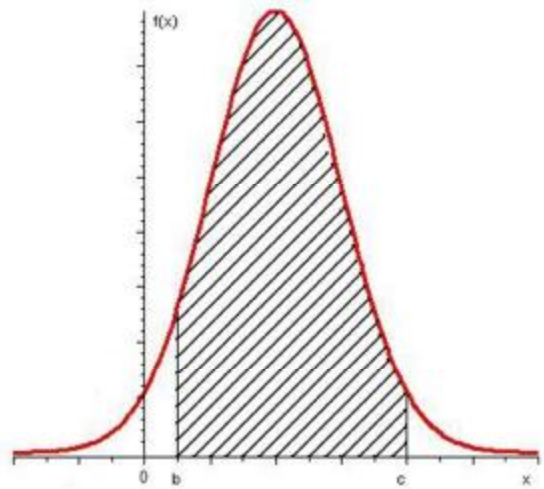


Рис.29.7 Геометрический смысл вероятности попадания нормальной случайной величины в интервал [b, c]

- кривая имеет единственный максимум, равный $\frac{1}{\sigma\sqrt{2\pi}}$ и лежащий на вертикальной оси симметрии $x=a$;

- кривая имеет две точки перегиба $x_{1,2} = a \pm \sigma$; в этих точках $f_X(x_1) = f_X(x_2) = \dfrac{1}{\sigma\sqrt{2\pi e}}$, где $e \approx 2{,}71828$;

- площадь бесконечной области, заключенной между осью Ox и нормальной кривой, равна единице (условие нормировки);

- площадь фигуры, ограниченной осью Ox, нормальной кривой и вертикальными прямыми $x = b$ и $x = c$ (Рис. 29.7) равна вероятности нахождения нормальной СВ внутри отрезка $[b, c]$.

В соответствии с формулой (5) раздела 29.1, интегральная функция распределения нормальной случайной величины имеет вид

$$F_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{(t-a)^2}{2\sigma^2}}\, dt\,. \quad (4)$$

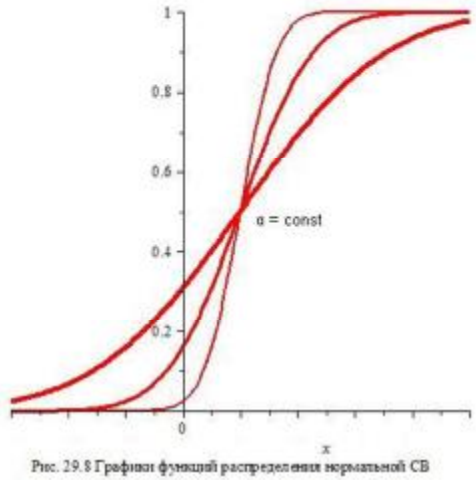


Рис. 29.8 Графики функций распределения нормальной СВ

На Рис. 29.8 показан график этой функции. Ее основные свойства соответствуют свойствам интегральной функции распределения НСВ, приведенным в разделе 29.1.

На рис. 29.9 приведены нормальные кривые для трех значений параметра σ $\left(\sigma_1 = \dfrac{1}{2}, \sigma_2 = 1, \sigma_3 = 2\right)$ и фиксированном значении параметра $a = 1$, совпадающим по величине с математическим ожиданием $(a = M(X))$.

Параметр σ совпадает с СКО нормального распределения НСВ. Увеличение СКО характеризует увеличение рассеяния нормальной СВ относительно центра рассеивания $M(X)$. У нормальной кривой это отражается в уменьшении максимума и расширении нормальной кривой (удалении точек перегиба относительно вертикальной

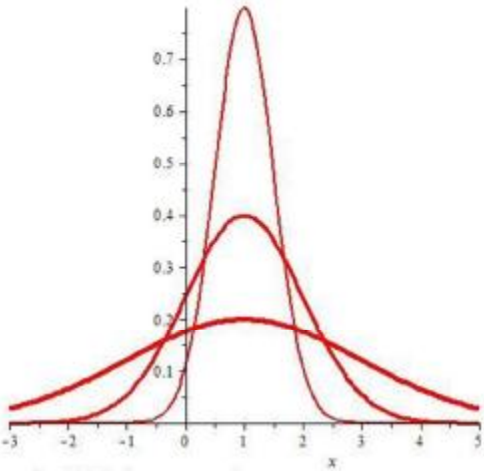


Рис. 29.9 Графики плотностей нормального распределения

оси симметрии $x = a$).

При изменении параметра a и постоянном СКО $(\sigma = const)$ нормальные кривые сдвигаются вдоль оси Ox влево или вправо без изменения формы. Таким образом, параметр $a = M(X)$ определяет положение нормальной кривой, а параметр $\sigma = \sqrt{D(X)}$ характеризует форму нормальной кривой.

В соответствии с формулой (4) раздела 29.1, вероятность попадания нормальной СВ в интервал $(\alpha; \beta)$ равна $P(X \in (\alpha; \beta)) = \frac{1}{\sigma\sqrt{2\pi}} \int_{\alpha}^{\beta} e^{-\frac{(x-a)^2}{2\sigma^2}} dx$. При практическом применении данной формулы возникают существенные вычислительные сложности, так как этот интеграл является «неберущимся» (см. Практикум, Часть II, раздел 13.1).

Чтобы упростить вычисления основные формулы общего нормального распределения (при $a \neq 0$, $\sigma \neq 1$) выражаются через стандартное (нормированное) распределение ($a = 0$, $\sigma = 1$).

Плотность вероятности стандартной нормальной случайной величины имеет вид

$$\varphi(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}$$

(6)

где $t = \frac{x-a}{\sigma}$ - возможное значение стандартной нормальной СВ $T = \frac{X-a}{\sigma}$. Для функции $\varphi(t)$ составлена специальная таблица 1 (см. Приложение 1). При использовании этой таблицы необходимо учитывать, что функция $\varphi(t)$ - четная, поэтому $\varphi(-t) = \varphi(t)$. График стандартной (центрированной нормальной кривой, соответствующей функции (6) показан на рис. 29.10.

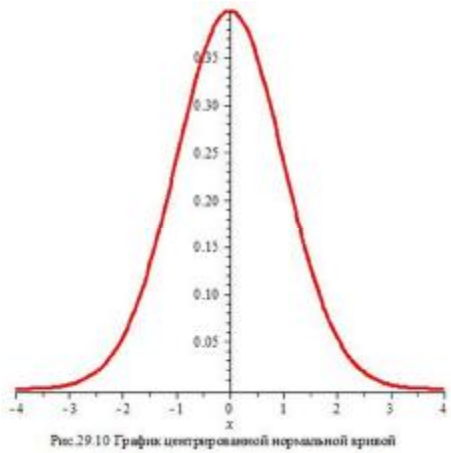


Рис.29.10 График центрированной нормальной кривой

Интегральную функцию распределения для нормальной СВ общего вида (2) после перехода к центрированной СВ представляют так:

$$F(x)=\Phi(t)=0{,}5+\Phi\left(\frac{x-a}{\sigma}\right),$$

(7)

где $\Phi(t)=\frac{1}{\sqrt{2\pi}}\int_0^t e^{-\frac{t^2}{2}}dt$ - функция Лапласа, для которой также составлена таблица 2 (см. Приложение 1). График функции Лапласа (7) показан на рис. 29.11.

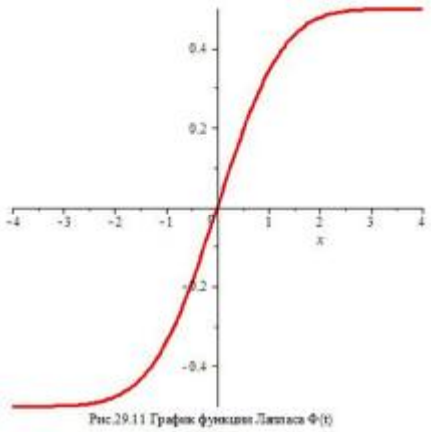


Рис.29.11 График функции Лапласа Ф(t)

Функция Лапласа обладает следующими основными свойствами, которые используют при расчетах:

1) функция $\Phi(t)$ является нечетной $(\Phi-(t)=-\Phi(t))$

   Замечание. Таблица для функции $\Phi(t)$ составлена только для неотрицательных значений аргумента $(t\geq 0)$;

2) При $t=0$ имеем $\Phi(t)=0$;

3) Функция $\Phi(t)$ монотонно возрастает, имеет горизонтальную асимптоту $y=-0{,}5$ при $t\to-\infty$ и горизонтальную асимптоту $y=0{,}5$ при $t\to\infty$. Обычно при расчетах принимают $\Phi(t)\approx 0{,}5$ при $t\geq 4$.

Определим вероятности нормальной СВ:

1. Вероятности того, что СВ примет любое значение меньше x $(X<x)$ или больше x $(X>x)$ определяют по формулам

$$P(X<x)=\frac{1}{2}+\Phi\left(\frac{x-a}{\sigma}\right)$$

(8)

$$P(X>x)=\frac{1}{2}-\Phi\left(\frac{x-a}{\sigma}\right)$$

(9)

2. При определении вероятности нахождения возможных значений нормальной СВ в отрезке $[b, c]$, расположенном несимметрично относительно вертикальной прямой $x = a$ (Рис. 29.10) используют формулу $P(b < X < c) = \Phi\left(\frac{c-a}{\sigma}\right) - \Phi\left(\frac{b-a}{\sigma}\right)$ (10)

3. Если интервал (x_1, x_2) или отрезок $[x_1, x_2]$ симметричен относительно прямой $x = a$, то вероятность нахождения возможных значений нормальной СВ определяют по формуле

$$P(x_1 < X < x_2) = P(|X - a| < \Delta) = P(a - \Delta < X < a + \Delta) = 2\Phi\left(\frac{\Delta}{\sigma}\right)$$

(11)

где Δ - абсолютное значение отклонения нормальной СВ от ее математического ожидания $M(X) = a$ (рис. 29.12).

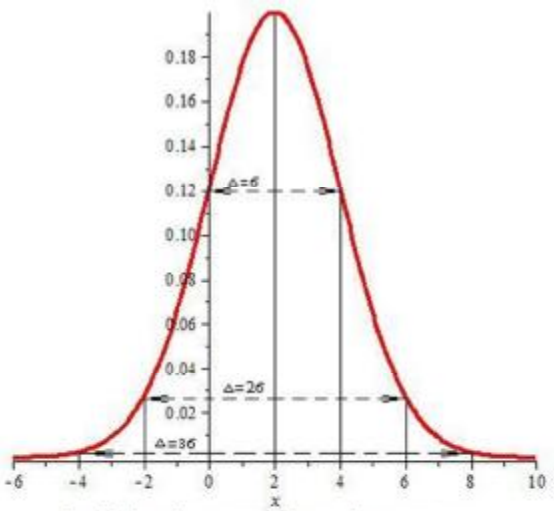


Рис. 29.12 график нормальной кривой с различными симметричными интервалами относительно M(X)

Все возможные значения нормальной СВ находятся в неограниченном (бесконечном) интервале $(-\infty, \infty)$. Для большинства прикладных задач можно использовать ограниченный симметричный относительно прямой $x = a$ интервал $(x_{\min}, x_{\max})$ в соответствии с «правилом трех сигм». Согласно этому правилу, с погрешностью, меньшей $0{,}2\%$ (две десятых процента или два промилле) можно принять, что все возможные значений нормальной СВ принадлежат ограниченному отрезку $[a - 3\sigma;\ a + 3\sigma]$ (Рис 29.12).

Рассмотрим примеры вычисления вероятностей нормальной СВ.

Пример 1. На грузовой станции находятся нагруженные контейнеры. Масса контейнера является нормальной СВ с параметрами $M(X) = 950$ кг., $\sigma(X) = 150$ кг. Найти вероятность того, что масса случайно выбранного контейнера : а) больше 1250 кг.; б) менее 850 кг.; в) будет находиться в интервале $(800; 1300)$;

г) отклонится от математического ожидания меньше, чем на 50 кг; д) отклонится от математического ожидания больше, чем на 50 кг.

е) С вероятностью 0,899 определить границы, в которых будет находиться масса случайно отобранного контейнера.

Решение. Обозначим СВ (массу контейнера) через X. Используем правило трех сигм: $a-3\sigma=950-450=500$, $a+3\sigma=1400$. Значит, все возможные значения СВ X с вероятностью, практически равной 1, принадлежат отрезку $[500;\,1400]$.

а) $P(X>1250)=P(X\in(1250;\,1400))=\Phi\left(\frac{1400-950}{150}\right)-\Phi\left(\frac{1250-950}{150}\right)\approx$

$\approx 0{,}4987-0{,}4772=0{,}0215$.

б) $P(X<850)=P(X\in(500;\,850))=\Phi\left(\frac{850-950}{150}\right)-\Phi\left(\frac{500-950}{150}\right)=$

$=\Phi\left(-\frac{2}{3}\right)-\Phi(-3)=-\Phi\left(\frac{2}{3}\right)+\Phi(3)\approx -0{,}2486+0{,}4987=0{,}2501$.

в) $P(X\in(800;\,1300))=\Phi\left(\frac{1300-950}{150}\right)-\Phi\left(\frac{800-950}{150}\right)=\Phi\left(\frac{7}{3}\right)-\Phi(-1)\approx$

$\approx 0{,}4904+0{,}3413=0{,}8317$.

г) $P(|X-950|<50)=P(-50<X-950<50)=P(X\in(900;\,1000))=$

$=\Phi\left(\frac{1000-950}{150}\right)-\Phi\left(\frac{900-950}{150}\right)=2\Phi\left(\frac{1}{3}\right)\approx 2\Phi(0{,}333)\approx 2\cdot 0{,}1293=0{,}2586$.

д) $P(|X-950|>50)=1-P(|X-950|<50)\approx 1-0{,}2586=0{,}7414$;

е) $P(|X-950|<\Delta)=2\Phi\left(\frac{\Delta}{150}\right)=0{,}8999\Rightarrow\Phi\left(\frac{\Delta}{150}\right)\approx 0{,}4499\Rightarrow$

$\Rightarrow\frac{\Delta}{150}\approx 1{,}64$ (по таблице 2 Приложения 1), $\Delta\approx 1{,}64\cdot 150=246$. Значит, с вероятностью 0,8999 масса контейнера не меньше, чем $950-246=704$ и не больше, чем $950+246=1196$, т.е. масса контейнера с вероятностью 0,8999 принадлежит отрезку $[704;\,1196]$.

Пример 2. Ошибки измерения подчиняются нормальному закону распределения с параметрами $a=0$, $\sigma=20$. Найти вероятность того, при трех

независимых измерениях абсолютное значение случайной ошибки хотя бы в одном измерении не превзойдет 4.

Решение. Вводим событие $A = \{при\ трех\ измерениях\ хотя\ бы\ один\ раз\ модуль\ ошибки \le 4\ мм\}$. По формуле вероятности хотя бы одного события имеем $P(A) = 1 - P(\overline{A})$, где $\overline{A} = \{при\ всех\ трех\ измерениях\ \ модуль\ ошибки > 4\ мм\}$. Событие $\overline{A}$ является сложным и равно $\overline{A} = \overline{A}_1 \overline{A}_2 \overline{A}_3$, где $\overline{A}_i = \{при\ i-ом\ измерении\ \ модуль\ ошибки > 4\ мм\}$. $\overline{A}_1, \overline{A}_2, \overline{A}_3$ - независимые события, поэтому $P(\overline{A}) = P(\overline{A}_1)P(\overline{A}_2)P(\overline{A}_3)$. Вероятность одного события $\overline{A}_i$ равна $P(\overline{A}_i) = 1 - P(A_i)$. По условию задачи событиям A_1, A_2, A_3 соответствует одна и та же СВ с параметрами $a = 0$, $\sigma = 0$. Тогда

$P(A_1) = P(A_2) = P(A_3) = P(|X| < 4) = P(X \in (-4; 4)) =$

$= 2\Phi\left(\frac{4}{20}\right) = 2\Phi(0{,}2) = 2 \cdot 0{,}0793 = 0{,}1586$; $P(\overline{A}_1) = P(\overline{A}_2) = P(\overline{A}_3) = 1 - 0{,}1586 = 0{,}8414$;

$P(A) = 1 - P(\overline{A}) = 1 - (0{,}8414)^3 \approx 0{,}41$.

5.5 **Показательное (экспоненциальное) распределение и его приложения**

Определение. *Показательным распределением* называется распределение НСВ, плотность которой задается формулой $f_X(x) = \begin{cases} 0, при\ x < 0 \\ \lambda e^{-\lambda x}, при\ x \ge 0 \end{cases}$

Показательное распределение является однопараметрическим (с параметром λ)

Интегральная функция распределения задается формулой $F_X(x) = \begin{cases} 0, при\ x < 0 \\ 1 - e^{-\lambda x}, при\ x \ge 0 \end{cases}$.

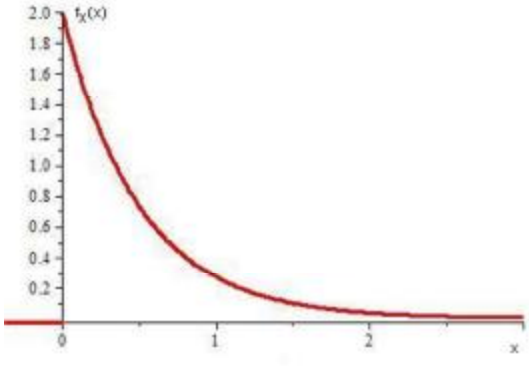


Рис. 29.13.График плотности показательно распределения вероятностей

Графики плотности и интегральной функции показательного распределения при $\lambda = 2$ приведены на Рис. 29.13 и Рис.29.14.

Вероятность, того, что значение случайной величины, подчиненной показательному закону распределения, принадлежит данному интервалу $(\alpha;\beta)$, находится по формуле $P(X\in(\alpha;\beta))=e^{-\lambda\alpha}-e^{-\lambda\beta}$.

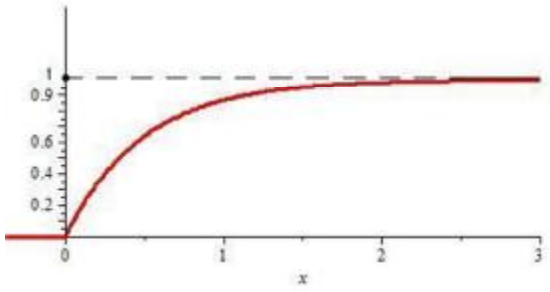


Рис. 29.14. График функции показательно распределения вероятностей

При нахождении значений показательной функции можно использовать специальные таблицы (таблица 3 Приложения 1) или раскладывать экспоненту в ряд Маклорена.

Основные числовые характеристики определяются по формулам $M(X)=\frac{1}{\lambda}$, $D(X)=\frac{1}{\lambda^2}$, $\sigma(X)=\frac{1}{\lambda}$.

Рассмотрим прикладную задачу с использованием показательного распределения.

Пример 1. По статистическим данным установлено, что среднее время ремонта автобуса является НСВ с показательным законом распределения. Среднее время ремонта автобуса составляет 15 дней. Найти вероятности того, что: а) ремонт случайно выбранного автобуса продлится не менее 20 дней; б) ремонт случайно выбранного автобуса продлится не более 15 дней;

Решение. По условию задачи $M(X)=15$ дней. Значит, $M(X)=\frac{1}{\lambda}=15\Rightarrow\lambda=\frac{1}{15}$.

а) $P(X\geq 20)=1-P(X<20)=1-F_X(20)=1-\left(1-e^{-\frac{20}{15}}\right)=e^{-\frac{4}{3}}\approx 0{,}2745$;

б) $P(X\leq 15)=P(X<15)=F_X(15)=1-e^{-\frac{15}{15}}=1-e^{-1}\approx 1-0{,}3679\approx 0{,}6321$;

Показательное распределение используется в теории надежности технических систем. В частности, с помощью такого распределения описывают внезапные отказы системы в случайные моменты времени.

Пусть T - непрерывная СВ, характеризующая длительность времени безотказной работы технического элемента. Функция распределения СВ T, по определению равная, равна вероятности безотказной работы элемента в течение времени, не превышающего $t > 0$. Для практики наибольший интерес обычно представляет вероятность противоположного события $R(t) = P(T \geq t) = 1 - F_T(t) = e^{-\lambda t}$, где $R(t)$ - функция надежности, определяющая вероятность безотказной работы элемента за время, большее, чем $t > 0$. Параметр λ - интенсивность отказов (среднее число отказов в единицу времени).

Показательное распределение обладает важным свойством «*отсутствия последействия*»: вероятность появления событий в любом интервале времени не зависит от вероятности появления событий в предшествующие интервалы времени.

Пример 2. Испытывают прибор с тремя независимо работающими элементами, отказы которых описываются при $t > 0$ показательными распределениями: для первого элемента - $F_1(t) = 1 - e^{-0,1t}$, для второго элемента - $F_2(t) = 1 - e^{-0,2t}$, для третьего элемента - $F_3(t) = 1 - e^{-0,3t}$. Найти вероятность того, что в течение 5 часов работы прибора

а) откажет ровно один элемент; б) откажут ровно два элемента; в) откажут три элемента;

Решение. Вычислим вероятности отказов каждого из элементов в течение 5 часов:

для первого элемента $p_1 = F_1(5) - F_1(0) = e^{-0,1 \cdot 0} - e^{-0,1 \cdot 5} \approx 1 - 0,606 = 0,394$;

для второго элемента $p_2 = F_2(5) - F_{21}(0) = e^{-0,2 \cdot 0} - e^{-0,2 \cdot 5} \approx 1 - 0,368 = 0,632$;

для третьего элемента $p_3 = F_3(5) - F_{31}(0) = e^{-0,3 \cdot 0} - e^{-0,3 \cdot 5} \approx 1 - 0,223 = 0,777$.

Тогда вероятности безотказной работы элементов в течение 5 часов будут равны: $q_1 = 1 - p_1 \approx 1 - 0,394 = 0,606$, $q_2 = 1 - p_2 \approx 1 - 0,632 = 0,368$, $q_3 = 1 - p_3 \approx 1 - 0,777 = 0,223$.

а) Вероятность того, что откажет ровно один элемент равна $p(1) = p_1 q_2 q_3 + q_1 p_2 q_3 + q_1 q_2 p_3 \approx 0{,}291$;

б) Вероятность того, что откажут ровно два элемента равна $p(2) = p_1 p_2 q_3 + p_1 q_2 p_3 + q_1 p_2 p_3 \approx 0{,}466$;

в) Вероятность того, что откажут три элемента равна $p(3) = p_1 p_2 p_3 \approx 0{,}194$;

Пример 3. На базу, работающую с 8 часов утра до 8 часов вечера, в среднем в течение часа, под погрузку товаров приезжают 5 автомобилей. Найти вероятность того, что с 8^{15} утра до 8^{30} утра прибудет один автомобиль.

Решение. Считаем распределение длительности времени прибытия автомобилей случайной величиной X с показательным распределением. Средний промежуток времени между моментами прибытия автомобилей, равное $x_{ср} = \frac{60}{5} = 12$ минут принимаем за математическое ожидание. Тогда интенсивность прибытия автомобилей равна $\lambda = \frac{1}{M(X)} = \frac{1}{12}$.

Если принять промежуток времени в 15 минут за единицу длительности времени, то один час представляется в виде четырех таких единиц.

Тогда вероятность того, что с 8^{15} утра до 8^{30} утра прибудет один автомобиль, определяется так: $P(X \in (1;\ 2)) = e^{-\frac{1}{12}} - e^{-\frac{2}{12}} \approx 0{,}9231 - 0{,}8437 = 0{,}0764$

Тема 6 Двумерные случайные величины

6.1 Исходные положения

Совокупность двух случайных величин (X, Y), рассматриваемых совместно при условии, что определена вероятность $P(X < x, Y < y)$ совместного появления событий $X < x$ и $Y < y$ для любых возможных значений $x \in R$ и $y \in R$, называют системой двух СВ или двумерной СВ (X, Y).

Двумерную СВ можно с геометрической точки зрения интерпретировать как множество случайных точек с координатами (x_i, y_j) на плоскости или как множество случайных радиус-векторов.

Двумерные случайные величины, так же как и одномерные могут быть дискретными или непрерывными.

6.2 **Двумерные дискретные случайные величины**

Двумерную случайную величину (X,Y) называют *дискретной двумерной случайной величиной*, если множество ее возможных значений, интерпретируемых как точки на координатной плоскости, является конечной или счетной совокупностью изолированных пар действительных чисел.

Далее будем рассматривать двумерные случайные величины (X,Y), составляющие (компоненты) которых X и Y имеют конечные числа возможных значений $X \in \{x_1, x_2, ..., x_k\}$, $Y \in \{y_1, y_2, ..., y_m\}$. Вероятность того, что СВ X примет значение x_i, а СВ Y - значение y_j будем обозначать p_{ij}, где первый индекс i соответствует (одномерной) СВ X, а второй индекс j - (одномерной) СВ Y.

Двумерная дискретная СВ обычно задается таблицей (матрицей) совместного распределения (Таблица 1), и ее называют *законом распределения двумерной случайной величины*.

Таблица 1

$X \quad Y \rightarrow$ $\downarrow$	y_1	y_2	…	y_j	…	y_m
x_1	p_{11}	p_{12}	…	p_{1j}	…	p_{1m}
x_2	p_{21}	p_{22}	…	p_{2j}	…	p_{2m}
…	…	…	…	…	…	…
x_i	p_{i1}	p_{i2}	…	p_{ij}	…	p_{im}
…	…	…	…	…	…	…
x_k	p_{k1}	p_{k2}	…	p_{kj}	…	p_{km}

В прикладных задачах заполнение таблицы 1 начинают с определения несовместных событий и их вероятностей, используя способы, рассмотренные в темах 25-27. Каждому такому событию сопоставляют возможные значения СВ X или Y двумерной СВ (X,Y). В результате возможным значениям (x_i, y_j) двумерной СВ (X,Y) будут соответствовать сложные события $\{X = x_i, Y = y_j\}$ с вероятностями $p_{ij} = P(X = x_i, Y = y_j)$. Эти события в совокупности составляют полную группу событий, поэтому сумма их вероятностей равна 1: $\sum_{i=1}^{k}\sum_{j=1}^{m} p_{ij} = 1$.

Рассмотрим пример формирования таблицы распределения двумерной СВ при переходе от случайных событий к двумерной СВ.

Пример 1. В двух ящиках находятся белые, синие и красные шары, неразличимые на ощупь. В первом ящике находятся: один белый шар, два синих шара и три красных. Во втором – два белых, три синих и один красный. Из каждого ящика случайным образом вынули по одному шару. Составить таблицу совместного распределения вероятностей различных цветов двух вынутых шаров.

Решение. Пусть простым событиям по извлечению шара из первого ящика соответствуют возможные значения СВ X, из второго ящика - возможные значения СВ Y.

Перейдем к точечным представлениям возможных значений двумерной СВ возможные значения СВ (X,Y). Для этого присвоим белым шарам номер1, синим – номер 2 и красным – номер 3.

Тогда получим точки со следующими различными координатами (x_i, y_j):

$$(1,1),\ (1,2),\ (1,3),\ (2,1),\ (2,2),\ (2,3),\ (3,1),\ (3,2),\ (3,3). \qquad (1)$$

Так как цвета шаров повторяются (в каждом ящике имеются шары одинакового цвета), то необходимо определить общее количество точек с учетом их кратности. Кратные точки имеют одинаковые координаты.

Для определения количества таких точек множества X и Y представим в виде трех подмножеств $X = \{Б(1),\ C(2),\ K(3)\},\quad Y = \{Б(2),\ C(3),\ K(1)\}$, где $Б,\ C,\ K$ - соответственно подмножества белых, синих и красных шаров в ящиках. В круглых скобках указано число шаров одного цвета в каждом ящике.

Количество кратных точек определим с помощью комбинаторного правила умножения $m_{(1)} \cdot m_{(2)}$, где $m_{(1)}$, $m_{(2)}$ - соответственно количества шаров какого-либо одного цвета в первом и втором ящиках.

Результаты определения количества кратных точек представим в табличной форме.

Координаты точки (x_i, y_j)	Количество шаров одинакового цвета в ящиках				$(m_{(1)} \cdot m_{(2)})_{ij}$
	цвет	$m_{(1)}$	цвет	$m_{(2)}$	
$(1,1)$	белый	1	белый	2	2
$(1,2)$	белый	1	синий	3	3

$(1,3)$	белый	1	красный	1	1
$(2,1)$	синий	2	белый	2	4
$(2,2)$	синий	2	синий	3	6
$(2,3)$	синий	2	красный	1	2
$(3,1)$	красный	3	белый	2	6
$(3,2)$	красный	3	синий	3	9
$(3,3)$	красный	3	красный	1	3

Всего случайных точек с учетом их кратности: $n_s = \sum_{i=1}^{3}\sum_{j=1}^{3}\left(m_{(1)} \cdot m_{(2)}\right)_{ij} = 2+3+1+4+...+3 = 36$.

Вероятности указанных точек определяются по формуле $p_{ij} = \dfrac{\left(m_{(1)} \cdot m_{(2)}\right)_{ij}}{n_s}$.

Полученные результаты позволяют заполнить таблицу совместного распределения двумерной СВ (X,Y).

Таблица 2

$X \downarrow Y \rightarrow$	1	2	3
1	$p_{11} = \frac{2}{36} = \frac{1}{18}$	$p_{12} = \frac{1}{12}$	$p_{13} = \frac{1}{36}$
2	$p_{21} = \frac{4}{36} = \frac{1}{9}$	$p_{22} = \frac{1}{6}$	$p_{23} = \frac{1}{18}$
3	$p_{31} = \frac{6}{36} = \frac{1}{6}$	$p_{32} = \frac{1}{4}$	$p_{33} = \frac{1}{12}$

Проверяем условие нормировки: $\sum_{i=1}^{3}\sum_{j=1}^{3} p_{ij} = p_{11} + p_{12} + ... + p_{33} = \frac{1}{18} + \frac{1}{12} + ... + \frac{1}{12} = 1$

Если известен закон распределения двумерной СВ, то ее функция распределения может быть задана формулой $F(x,y) = P(X < x, Y < y) = \sum_{x_i < x}\sum_{y_j < y} p_{ij}$.

Функция распределения обладает следующими свойствами:

1. Все значения функции $F(x,y)$ принадлежат отрезку $[0;1]$;
2. Функция $F(x,y)$ является неубывающей по каждому аргументу при фиксированном значении другого аргумента: $F(x_2,y_0) \geq F(x_1,y_0)$ при $x_2 > x_1$ и $F(x_0,y_2) \geq F(x_0,y_1)$ при $y_2 > y_1$;
3. Функция $F(x,y)$ непрерывна слева по каждому из своих аргументов: $\lim\limits_{x \to x_0 - 0} F(x,y) = F(x_0,y)$ и $\lim\limits_{y \to y_0 - 0} F(x,y) = F(x,y_0)$;

Функцию $F(x,y)$ удобно находить по закону распределения соответствующей СВ способом верхнего левого угла. Рассмотрим этот способ на конкретном примере.

Пример 2. Дана таблица распределения двумерной СВ (Таблица 3). Найти функцию распределения заданной СВ.

Таблица 3

$X \downarrow Y \rightarrow$	0	1	2
0	0,0324	0,0432	0,0144
1	0,1512	0,2016	0,0672
2	0,1764	0,2352	0,0784

Решение. Функция распределения $F(x,y)$ представлена в виде таблицы 4:

Таблица 4

$X \downarrow Y \rightarrow$	$y \leq 0$	$0 < y \leq 1$	$1 < y \leq 2$	$y > 2$
$x \leq 0$	0	0	0	0
$0 < x \leq 1$	0	0,0324	0,0756	0,0900
$1 < x \leq 2$	0	0,1836	0,4284	0,5100
$x > 2$	0	0,3600	0,8400	1,0000

Первая строка и первый столбец таблицы значений функции $F(x,y)$ для любой дискретной СВ с конечным множеством значений будут состоять из нолей. Количество строк и столбцов оставшейся части таблицы (в таблице 4 закрашены серым цветом) обязательно совпадет с количеством строк и столбцов таблицы распределения СВ (Таблица 3). В нашем примере эти

количества равны 3 и 3. Для заполнения соответствующей клетки таблицы значений функции $F(x,y)$ находим в таблице распределения «прямоугольник», левый верхний угол которого совпадает с левой верхней клеткой таблицы, а правый нижний – с соответствующей клеткой таблицы распределения, и считаем сумму всех вероятностей в этом прямоугольнике. Например, число 0,4284 в таблице значений функции $F(x,y)$ получилось при суммировании вероятностей в клетках таблицы 3, закрашенных серым цветом.

Двумерная случайная величина, по определению, является упорядоченной совокупностью двух одномерных величин. Вероятность $P(X=x_i)$ можно найти, если просуммировать все вероятности $p_{i1}, p_{i2}, ..., p_{im}$ (в наших обозначениях это сумма всех вероятностей в соответствующей строке ряда распределения): $P(X=x_i)=\sum_{j=1}^{m} p_{ij}$. Аналогично, $P(Y=y_j)=\sum_{i=1}^{k} p_{ij}$ (в наших обозначениях это сумма всех вероятностей в соответствующем столбце таблицы распределения). Значит, если известен закон распределения двумерной дискретной СВ, то по нему можно составить ряды распределения соответствующих одномерных величин. Из этого следует, в частности, что $\lim_{x\to\infty} F(x,y)=F_Y(y)$ и $\lim_{y\to\infty} F(x,y)=F_X(x)$.

Случайные величины X и Y, составляющие двумерную СВ (X,Y), могут быть зависимыми или независимыми. Необходимым и достаточным условием независимости случайных величин X и Y является выполнение равенств $p_{ij}=p(x_i)\cdot p(y_j)$ при всех значениях индексов $i=1,2,...,k$, $j=1,2,...,m$.

Пример 3. В группе из 20 студентов двое пропустили более половины занятий и получили оценку «неудовлетворительно» на экзамене, пять студентов получили «отлично», десять – «хорошо», трое – «удовлетворительно». а) Составить ряд распределения двумерной СВ (X,Y), где X - оценка на экзамене, Y - пропуск более половины занятий. б) составить ряды распределения одномерных СВ X и Y. в) Проверить, являются ли величины X и Y зависимыми.

Решение. а) Случайная величина X может принимать значения $\{2;3;4;5\}$. Для СВ Y положим $Y=0$, если число пропусков занятий не более 50% и $Y=1$, если число пропусков занятий более 50%. Таким образом, $Y\in\{0;1\}$.

Таблица 5

$Y \;\; X \rightarrow$ $\downarrow$	$x_1 = 2$	$x_2 = 3$	$x_3 = 4$	$x_4 = 5$	$p(y_j)$
$y_1 = 0$	0	$\frac{3}{20}$	$\frac{10}{20}$	$\frac{5}{20}$	$\frac{18}{20}$
$y_2 = 1$	$\frac{2}{20}$	0	0	0	$\frac{2}{20}$
$p(x_i)$	$\frac{2}{20}$	$\frac{3}{20}$	$\frac{10}{20}$	$\frac{5}{20}$	1

б) Составляем ряды распределения одномерных СВ X и Y (данные выписаны из таблицы 5):

Таблица 6

X	2	3	4	5
$p(x_i)$	$\frac{2}{20}$	$\frac{3}{20}$	$\frac{10}{20}$	$\frac{5}{20}$

Таблица 7

Y	0	1
$p(y_j)$	$\frac{18}{20}$	$\frac{2}{20}$

.

Замечание 3. Ряды распределения одномерных СВ X и Y можно совмещать с таблицей распределения двумерной СВ (X,Y), как и сделано в таблице 5.

в) С помощью условия $p_{ij} = p(x_i) \cdot p(y_j)$ проверяем зависимость СВ X и Y .Для $i = 1, j = 1$ из Таблиц 6 и 7 находим $p(x_1) = \frac{2}{20}$, $p(y_1) = \frac{18}{20}$, а из таблицы 5 имеем $p_{11} = 0$. Очевидно, что $0 = p_{11} \neq p(x_1) \cdot p(y_2) = \frac{2}{20} \cdot \frac{18}{20}$. Это значит, что условие независимости не выполнилось при первых же значениях индексов $i,\ j$. Величины X и Y являются зависимыми.

Замечание 4. Случайные величины, принимающие ровно два значения, нередко называют *индикаторами событий*. В примере 3 СВ Y является индикатором 50-процентного пропуска занятий.

Основными числовыми характеристиками двумерной СВ являются:

- математические ожидания $M(X),\ M(Y)$;

- дисперсии $D(X),\ D(Y)$;

- ковариация (корреляционный момент);

С геометрической точки зрения упорядоченная совокупность математических ожиданий $(M(X), M(Y))$ - это координаты точки, являющейся центром рассеяния значений двумерной СВ (X, Y).

Если задана таблица распределения двумерной СВ, то, после составления рядов распределения одномерных СВ X и Y можно вычислить $M(X)$, $M(Y)$, $D(X)$, $D(Y)$, $\sigma(X)$, $\sigma(Y)$ по известным формулам (см. 28.5).

Корреляционный момент (ковариация) характеризует в количественной форме взаимную связь СВ X и Y, образующих двумерную СВ. Кроме того, он является и дополнительной мерой рассеивания двумерной СВ относительно центра рассеивания.

Корреляционный момент можно представить в виде формулы $K_{xy} = \operatorname{cov}(X, Y) - M(XY) - M(X)M(Y)$, которую обычно используют для практических вычислений.

Из свойств математического ожидания (см. 28.5) следует, что для независимых СВ X и Y корреляционный момент равен нулю. Обратное утверждение, в общем случае, неверно. Если СВ X и Y не коррелированы, то может существовать нелинейная вероятностная связь.

Свойства ковариации:

1. Ковариация симметрична $K_{xy} = K_{yx}$;
2. Постоянный множитель можно выносить за знак ковариации и переносить от одной СВ к другой: $K(CX, Y) = CK(X, Y) = K(X, CY)$
3. Ковариация не изменится, если к одной или двум СВ прибавить постоянный множитель: $K(X + C_1, Y) = K(X, Y + C_2) = K(X + C_1, Y + C_2)$;
4. Ковариация двумерной СВ по модулю не превосходит произведения их СКО: $|K_{xy}| \le \sigma(X)\sigma(Y)$;

Связь между математическими ожиданиями, дисперсиями и ковариацией

1. $M(XY) = M(X)M(Y) + K_{xy}$;
2. $D(X \pm Y) = D(X) + D(Y) \pm 2K_{xy}$;
3. $K_{xx} = D(X)$;

Последнее свойство позволяет связать с каждой двумерной СВ ковариационную матрицу $\begin{pmatrix} K_{xx} & K_{xy} \\ K_{yx} & K_{yy} \end{pmatrix}$. Кориационные матрицы по аналогии можно ввести и для произвольных n-мерных СВ.

Чтобы выделить и количественно определить степень линейной зависимости между СВ X и Y (если зависимость нелинейна, то речь идет о линейном приближении зависимости), вводят дополнительную безразмерную числовую характеристику – *коэффициент корреляции* $r_{xy} = \dfrac{K_{xy}}{\sigma(X)\sigma(Y)}$. Этот коэффициент является наилучшей количественной оценкой степени влияния одной СВ на другую в линейном приближении.

Свойства коэффициента корреляции:

1. $|r_{xy}| \le 1$;
2. При $r_{xy} > 0$ имеем положительную корреляционную связь между X и Y (нестрого говоря – чем больше X, тем больше Y), а при $r_{xy} < 0$ - отрицательную (чем больше X, тем меньше Y);
3. Если $r_{xy} = \pm 1$, то между СВ X и Y имеет место линейная функциональная связь, т.е. СВ Y является функцией (причем, линейной) от СВ X (см. 28.5).

Пример 3. Найти числовые характеристики двумерной СВ из примера 2.

Решение. Используя данные таблиц 5 и 6, находим математические ожидания и дисперсии СВ X и Y:

$$M(X) = \sum_{i=1}^{4} x_i p(x_i) = 2 \cdot \frac{2}{20} + 3 \cdot \frac{3}{20} + 4 \cdot \frac{10}{20} + 5 \cdot \frac{5}{20} = 3{,}9;$$

$$M(Y) = \sum_{j=1}^{2} y_j p(y_j) = 0 \cdot \frac{18}{20} + 1 \cdot \frac{2}{20} = 0{,}1;$$

$$D(X) = \sum_{i=1}^{4} x_i^2 p(x_i) - (M(X))^2 = 2^2 \cdot \frac{2}{20} + 3^2 \cdot \frac{3}{20} + 4^2 \cdot \frac{10}{20} + 5^2 \cdot \frac{5}{20} - (3{,}9)^2 = 0{,}79;$$

$$D(Y) = \sum_{j=1}^{2} y_j^2 p(y_j) - (M(Y))^2 = 0^2 \cdot \frac{18}{20} + 1^2 \cdot \frac{2}{20} - (0{,}1)^2 = 0{,}09;$$

Определяем корреляционный момент (ковариацию) и коэффициент корреляции:

$$M(XY) = \sum_{i=1}^{4} \sum_{j=1}^{2} p_{ij} x_i y_j = 2 \cdot 1 \cdot \frac{2}{20} = 0{,}2;\ K_{xy} = M(XY) - M(X)M(Y) = 0{,}2 - 3{,}9 \cdot 0{,}1 = -0{,}19;$$

$$r_{xy} = \frac{K_{xy}}{\sigma(X)\sigma(Y)} = \frac{-0{,}19}{\sqrt{0{,}79}\sqrt{0{,}09}} \approx -0{,}713 .$$

Коэффициент корреляции получился отрицательным, это значит, что при увеличении СВ X другая СВ (Y) имеет тенденцию к уменьшению. Для нашего примера это значит, что оценка объективно повышается при уменьшении количества пропусков занятий.

6.3 **Условные законы распределения двумерной дискретной СВ. Функция регрессии**

Взаимозависимость СВ X и Y в двумерной СВ (X, Y) описывают с помощью условных законов распределения.

Вероятность того, что СВ Y примет значение y_j при условии, что $X = x_i$, называется *условной вероятностью* и определяется равенством $P(y_j \mid x_i) = \frac{p_{ij}}{p(x_i)}$. Вычисляя вероятности $p(y_1 \mid x_i)$, $p(y_2 \mid x_i)$, ..., $p(y_m \mid x_i)$, получим *условный закон распределения* СВ Y при фиксированном значении СВ $X = x_i$. Ряд условного распределения приведен в таблице:

Таблица 10

$Y_{X=x_i}$	y_1	y_2	⋯	y_m
$p(y_j \mid x_i)$	$p(y_1 \mid x_i)$	$p(y_2 \mid x_i)$	⋯	$p(y_m \mid x_i)$

Если случайная величина X имеет k различных возможных значений $X \in \{x_1, x_2, ..., x_k\}$, то можно составить k различных условных законов распределения СВ Y.

Замечание 1. Сумма условных вероятностей в ряде распределения (таблица 10) равна 1.

Замечание 2. Аналогично можно определить условные вероятности $P(x_i \mid y_j) = \frac{p_{ij}}{p(y_j)}$ и составить условные законы распределения для СВ X при каждом значении СВ $Y = y_j$.

Если известны условные законы распределения, то можно найти и условные числовые характеристики, в частности, условные математические ожидания $M(X \mid y_j)$ и $M(Y \mid x_j)$. Используя m условных законов распределения СВ Y при различных фиксированных значениях

СВ X находим k условных математических ожиданий $M(Y \mid x_1)$, $M(Y \mid x_2)$, ..., $M(Y \mid x_k)$. В результате получаем дискретную функцию условного математического ожидания одной СВ Y относительно возможных значений другой СВ $X \in \{x_1, x_2, ..., x_k\}$. Такая функциональная зависимость называется *функцией регрессии* Y *на* X или *регрессией* Y *на* X.

Функцию регрессии в табличной форме можно представить так:

Таблица 11

$X = x_i$	x_1	x_2	⋯	x_k
$M(Y \mid x_i)$	$M(Y \mid x_1)$	$M(Y \mid x_2)$	⋯	$M(Y \mid x_k)$

Аналогично можно представить регрессию X на Y.

Рассмотрим в плоской системе координат точки с координатами $(x_i, M(Y \mid x_i))$. Последовательно соединяя эти точки отрезками прямых, получим ломаную линию, называемую *многоугольником (полигоном) регрессии*. Многоугольник регрессии позволяет дать качественную оценку тенденции изменения среднего значения одной СВ от значений другой СВ.

Для приближенной количественной оценки используют *среднюю квадратичную линейную регрессию* (или просто *линейную регрессию*), которая является линейной функцией.

Уравнение линейной регрессии Y на X имеет вид

$$y = m_y + \rho_y(x - m_x) \qquad (1)$$

где $m_x = M(X), m_y = M(Y)$ - математические ожидания СВ X и Y соответственно, $\rho_y = r\dfrac{\sigma_y}{\sigma_x}$ - коэффициент регрессии СВ Y *на* X.

Аналогично, уравнение линейной регрессии X на Y представляется так:

$$x = m_x + \rho_x(y - m_y) \qquad (2)$$

где $\rho_x = r\dfrac{\sigma_x}{\sigma_y}$ - коэффициент регрессии СВ X *на*. Y.

Прямые линии (1) и (2) пересекаются в точке (m_x, m_y) - центре рассеивания. Если коэффициент корреляции $r = 0$, то линии регрессии параллельны координатным осям и, соответственно, взаимно перпендикулярны.

Выражая в уравнении (2) переменную y через переменную x, вообще говоря, не получим уравнение (1). Это произойдет лишь в том случае, когда $r = \pm 1$ и, следовательно, СВ X и Y связаны линейной функциональной зависимостью.

Пример 1. Найти условные законы распределения СВ Y при фиксированных значениях СВ X. Найти линейные регрессии Y *на* X и X *на*. Y для двумерной СВ, заданной таблицей распределения. Построить графики регрессий.

Таблица 12

$Y \downarrow X \rightarrow$	$x_1 = 2$	$x_2 = 5$	$x_3 = 8$
$y_1 = 4$	p_{11}	0,3	0,35
$y_2 = 8$	0,05	0,12	0,03

Решение. 1. Находим p_{11}, используя условие нормировки $\sum_{i=1}^{k}\sum_{j=1}^{m} p_{ij} = 1$:

$p_{11} + 0{,}3 + 0{,}35 + 0{,}05 + 0{,}12 + 0{,}03 = 1 \Rightarrow p_{11} = 0{,}15$.

2. Определяем законы распределения одномерных СВ X и. Y:

Таблица 13

X	2	5	8
$p(x_i)$	0,2	0,42	0,38

Условие нормировки $\sum_{i=1}^{k} p(x_i) = 0{,}2 + 0{,}42 + 0{,}38 = 1$ выполнено.

Таблица 14

Y	4	8
$p(y_j)$	0,8	0,2

Условие нормировки $\sum_{ji=1}^{k} p(y_j) = 0{,}8 + 0{,}2 = 1$ выполнено.

3. Находим координаты центра рассеивания (m_x, m_y):

$$m_x = \sum_{i=1}^{3} x_i p(x_i) = 2 \cdot 0{,}2 + 5 \cdot 0{,}42 + 8 \cdot 0{,}38 = 5{,}54;\ m_y = \sum_{j=1}^{2} y_j p(y_j) = 4 \cdot 0{,}8 + 8 \cdot 0{,}2 = 4{,}8.$$

4. Определяем условные законы распределения для СВ Y при $x_1 = 2$, $x_2 = 5$, $x_3 = 8$, используя формулу $P(y_j \mid x_i) = \dfrac{p_{ij}}{p(x_i)}$.

При $X = x_1 = 2$ получаем $P(y_1 \mid x_1) = \dfrac{p_{11}}{p(x_1)} = \dfrac{0{,}15}{0{,}2} = 0{,}75$, $P(y_2 \mid x_1) = \dfrac{p_{12}}{p(x_1)} = \dfrac{0{,}05}{0{,}2} = 0{,}25$.

Таблица 15

$Y_{X=x_1}$	4	8
$p(y_j \mid x_1)$	0,75	0,25

При $X = x_2 = 5$ получаем $p(y_1 \mid x_2) = \dfrac{p_{21}}{p(x_2)} = \dfrac{0{,}3}{0{,}42} = 0{,}714$, $p(y_2 \mid x_2) = \dfrac{p_{22}}{p(x_2)} = \dfrac{0{,}12}{0{,}42} = 0{,}286$

.

Таблица 16

$Y_{X=x_2}$	4	8
$p(y_j \mid x_2)$	0,714	0,286

При $X = x_3 = 8$ получаем $p(y_1 \mid x_3) = \dfrac{p_{31}}{p(x_3)} = \dfrac{0{,}35}{0{,}38} = 0{,}92$, $p(y_2 \mid x_3) = \dfrac{p_{32}}{p(x_3)} = \dfrac{0{,}03}{0{,}38} = 0{,}08$.

Таблица 17

$Y_{X=x_3}$	4	8
$p(y_j \mid x_3)$	0,92	0,08

5. Находим функцию регрессии Y *на* X по данным условных законов распределения и формуле $M(Y \mid x_i) = \sum_{j=1}^{m} y_j \cdot p(y_j \mid x_i)$:

$$M(Y \mid x_1) = \sum_{j=1}^{2} y_j \cdot p(y_j \mid x_1) = 4 \cdot 0{,}75 + 8 \cdot 0{,}25 = 5;$$

$$M(Y \mid x_2) = \sum_{j=1}^{2} y_j \cdot p(y_j \mid x_2) = 4 \cdot 0{,}714 + 8 \cdot 0{,}286 = 5{,}144;$$

$$M(Y \mid x_3) = \sum_{j=1}^{2} y_j \cdot p(y_j \mid x_3) = 4 \cdot 0{,}92 + 8 \cdot 0{,}08 = 4{,}064.$$

Таким образом, получаем функцию регрессии в табличной форме:

Таблица 18

x_i	2	5	8
$M(Y \mid x_i)$	5	5,144	4,064

6.Находим условные законы распределения для СВ X при фиксированных значениях СВ Y, используя формулу $p(x_i \mid y_j) = \dfrac{p_{ij}}{p(y_j)}$.

При $Y = y_1 = 4$ получаем $p(x_1 \mid y_1) = \dfrac{p_{11}}{p(y_1)} = \dfrac{0{,}15}{0{,}8} = 0{,}1875$, $p(x_2 \mid y_1) = \dfrac{p_{21}}{p(y_1)} = \dfrac{0{,}3}{0{,}8} = 0{,}375$,

$$p(x_3 \mid y_1) = \frac{p_{31}}{p(y_1)} = \frac{0{,}35}{0{,}8} = 0{,}4375.$$

Таблица 19

$X_{Y=y_1}$	2	5	8
$p(x_i \mid y_1)$	0,1875	0,375	0,4375

При $Y = y_2 = 8$ получаем $p(x_1 \mid y_2) = \dfrac{p_{12}}{p(y_2)} = \dfrac{0{,}05}{0{,}2} = 0{,}25$, $p(x_2 \mid y_2) = \dfrac{p_{22}}{p(y_2)} = \dfrac{0{,}12}{0{,}2} = 0{,}6$,

$$p(x_3 \mid y_2) = \frac{p_{32}}{p(y_2)} = \frac{0{,}03}{0{,}2} = 0{,}15.$$

Таблица 20

$X_{Y=y_2}$	2	5	8
$p(x_i \mid y_2)$	0,25	0,6	0,15

7. Находим функцию регрессии X *на* Y по данным условных законов распределения и формуле $M(X \mid y_j) = \sum_{j=1}^{m} x_i \cdot p(x_i \mid y_j)$:

$$M(X \mid y_1) = \sum_{i=1}^{3} x_i \cdot p(x_i \mid y_1) = 2 \cdot 0{,}1875 + 5 \cdot 0{,}375 + 8 \cdot 0{,}4375 = 5{,}75;$$

$$M(X \mid y_2) = \sum_{i=1}^{3} x_i \cdot p(x_i \mid y_2) = 2 \cdot 0{,}25 + 5 \cdot 0{,}6 + 8 \cdot 0{,}15 = 4{,}7.$$

Таким образом, получаем функцию регрессии в табличной форме:

Таблица 21

y_j	4	8
$M(X \mid y_j)$	5,75	4,7

8. Находим СКО одномерных СВ X и Y:

$$M(X^2) = \sum_{i=1}^{3} x_i^2 \cdot p(x_i) = 2^2 \cdot 0{,}2 + 5^2 \cdot 0{,}42 + 8^2 \cdot 0{,}38 = 35{,}62;$$

$$D(X) = M(X^2) - (M(X))^2 = 35{,}62 - (5{,}54)^2 = 4{,}928;\ \sigma_x = \sqrt{D(X)} = \sqrt{4{,}928} \approx 2{,}22;$$

$$M(Y^2) = \sum_{j=1}^{2} y_j^2 \cdot p(y_j) = 4^2 \cdot 0{,}8 + 8^2 \cdot 0{,}2 = 25{,}6;\ D(Y) = M(Y^2) - (M(Y))^2 = 25{,}6 - (4{,}8)^2 = 2{,}56$$

; $\sigma_y = \sqrt{D(Y)} = \sqrt{2{,}56} = 1{,}6$.

9. Определяем корреляционный момент и коэффициент корреляции.

$$M(XY) = \sum_{i=1}^{3} \sum_{j=1}^{2} x_i y_j p_{ij} = 2 \cdot 4 \cdot 0{,}15 + 2 \cdot 8 \cdot 0{,}05 + 5 \cdot 4 \cdot 0{,}3 + 5 \cdot 8 \cdot 0{,}12 +$$

$$+ 8 \cdot 4 \cdot 0{,}35 + 8 \cdot 8 \cdot 0{,}03 = 25{,}92;$$

$$K_{xy} = M(XY) - m_x \cdot m_y = 25{,}92 - 5{,}54 \cdot 4{,}8 = -0{,}672$$

;

$$r_{xy} = \frac{K_{xy}}{\sigma_x \sigma_y} = \frac{-0{,}672}{2{,}22 \cdot 1{,}6} \approx -0{,}186.$$

10. Составляем уравнения линейных регрессий.

Регрессия Y на X: $y = m_y + \rho_y (x - m_x)$;

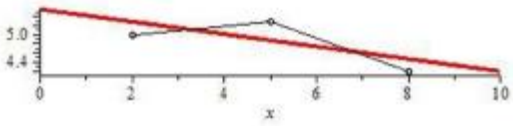


Рис.31. 1 График функции регрессии Y на X в примере 1

$$y = m_y + r_{xy}\frac{\sigma_y}{\sigma_x}(x - m_x);\ y = 4{,}8 - 0{,}186 \cdot \frac{1{,}6}{2{,}22}(x - 5{,}54);\ y = -0{,}136x + 5{,}55.$$

Регрессия X на Y: $x = m_x + \rho_x(y - m_y);$ $x = m_x + r_{xy}\frac{\sigma_x}{\sigma_y}(y - m_y);$

$$x = 5{,}54 - 0{,}186 \cdot \frac{2{,}22}{1{,}6}(y - 4{,}8);\ x = -0{,}258y + 6{,}78.$$

Рис 32.2 График функции регрессии X на Y в примере 1

Проверка. Прямые, задаваемые уравнениями регрессии должны проходить через центр рассеивания $(5{,}54;\ 4{,}8)$. Подставляем координаты центра рассеивания в уравнение регрессии Y на X: $-0{,}136 \cdot 5{,}54 + 5{,}55 \approx 4{,}8$ - верно. Подставляем координаты центра рассеивания в уравнение регрессии X на Y: $-0{,}258 \cdot 4{,}8 + 6{,}78 \approx 5{,}54$ - тоже верно.

11. Строим в системе координат линии регрессии и, для наглядности, многоугольники распределения (линия регрессии должна не сильно отклоняться от вершин многоугольника распределения (Рис. 31.1 и 31.2).

Замечание. В случае регрессии X на Y: многоугольник распределения имеет всего две вершины и поэтому состоит из единственного отрезка, лежащего на линии регрессии.

Тема 7 Способы представления выборочных данных и приближенных вычислений параметров генеральной совокупности

7.1 Исходные положения математической статистики

Математическая статистика – это раздел математики, изучающий методы получения научно обоснованных выводов о случайных событиях и случайных величинах на основе результатов реальных экспериментов или наблюдений.

В математической статистике, по существу, исследуются и решаются задачи, обратные к задачам теории вероятностей. Основные задачи математической статистики в практических приложениях сводятся к следующему: по результатам наблюдений за неизвестной (или недостаточно известной) случайной величиной (одномерной или многомерной) делается научно обоснованный вывод о законе распределения этой СВ или о тех или иных ее параметрах. Разумеется, такой вывод можно сделать только с некоторой вероятностью (желательно, достаточно высокой).

Пример. Необходимо определить вероятность брака в партии деталей по результатам контроля случайным образом взятой выборки из указанной партии.

Реальные объекты, с которыми проводятся эксперименты или наблюдения, называются *объектами наблюдения*. Обычно исследуются признаки (свойства) объектов наблюдения с их количественной стороны.

Объект наблюдения состоит из единиц наблюдения (единиц отбора), которые обладают в той или иной степени исследуемым признаком.

Полная совокупность всех единиц наблюдения, над которыми могли бы быть проведены наблюдения (эксперименты), называется *генеральной совокупностью*. Генеральной совокупностью называют также множество всех возможных значений исследуемого признака, которые можно получить при обследовании всех единиц наблюдения (как правило, в реальных задачах практики это множество неизвестно). Иными словами, генеральная совокупность – это то множество числовых значений, относящихся к объекту наблюдения, о котором нужно сделать те или иные выводы по результатам исследования доступной части этого множества – *выборки*.

Общее количество единиц наблюдения (элементов) в генеральной совокупности и выборке называют соответственно, *объемом генеральной совокупности* и *объемом выборки*. Объем выборки – всегда конечное число $n \in N$. Объем генеральной совокупности может быть как конечным, так и бесконечным.

В математической статистике выборочные данные используют как исходную экспериментальную основу для исследования генеральной совокупности *выборочным методом*, сущность которого заключается в том, чтобы установить по выборочным данным с определенной достоверностью закон распределения и (или) числовые характеристики генеральной совокупности как случайной величины, подчиняющейся вероятностным закономерностям.

Основным достоинством выборочного метода является экономия материальных, временных и трудовых ресурсов. Кроме того, во многих случаях этот метод является единственно возможным, например, если после испытаний объекты выборки исчерпывают свой ресурс. Основной недостаток выборочного метода – возможность значительных ошибок, особенно при малых $(n < 30)$ объемах выборок или при искажениях, связанных с недостатками методов наблюдений (экспериментов), не обеспечивающих репрезентативность выборок.

Пример. Пусть производятся наблюдения за СВ Y, которая связана с равномерно распределенной на отрезке $[0; 1000]$ случайной величиной X, формулой $Y = \sin \pi x$. При этом

наблюдателю известен закон распределения СВ X, но не известна формула, связывающая случайные величины X и Y. Наблюдатель задает тысячу значений $x_1 = 0$, $x_2 = 1$, $x_3 = 2$, , $x_{1000} = 999$, равномерно отстоящих друг от друга. Естественно, получается $y_1 = \sin 0 = 0$, $y_2 = \sin \pi = 0$, $y_3 = \sin 2\pi = 0$, ..., $y_{1000} = \sin 999\pi = 0$. Откуда напрашивается вывод, что СВ Y тождественно равна нулю, что, очевидно, неверно.

При получении исходных выборочных данных применяются специальные методы обеспечения представительности (*репрезентативности)* выборок. Эти специальные методы являются предметом исследования *общей статистики*. В рамках математической статистики будем предполагать, что все выборки являются репрезентативными.

7.2 **Способы представления исходных выборочных данных**

Исходные выборочные данные – первичные результаты выборочного обследования (измерений, опросов и т.п.)

Эти данные упорядочивают и группируют обычно в виде статистического ряда распределения. Такой ряд представляет собой таблицу, содержащую в первой строке возможные числовые значения элементов выборки, а во второй строке – соответствующие *частоты* наблюдения (количество *повторяющихся элементов выборки*). Вместо частот (или совместно с ними) используются *относительные частоты* – количества повторяющихся элементов выборки, деленные на объем выборки. Статистический ряд (*дискретный статистический ряд*) в общем виде представляется так:

x_i	x_1	x_2	$\cdots$	x_k
n_i	n_1	n_2	$\cdots$	n_k
$w_i = \frac{n_i}{n}$	w_1	w_2	$\cdots$	w_k

Замечание 1. Относительные частоты $w_i = \frac{n_i}{n}$ являются положительными числами, не превосходящими единицу и при этом сумма всех относительных частот равна единице (докажите). Поэтому дискретный статистический ряд (*простой статистический ряд относительных*

частот) является аналогом ряда распределения ДСВ. Это позволяет применять к такому ряду методы теории вероятностей.

Замечание 2. Не равные между собой значения $x_1, x_2, ..., x_k$ называют также ***вариантами*** выборки (в единственном числе – ***варианта***).

Пример 1. Исходные выборочные данные равны -5; 3; 0; -1; -3; 0; 1; 3; 0; 0; 1; -3; 1;-1; 0; -5; -1. Составить дискретный статистический ряд частот и дискретный статистический ряд относительных частот.

Решение. Дискретный статистический ряд частот имеет вид

x_i	-5	-3	-1	0	1	3
n_i	2	2	3	5	3	2

Поскольку объем выборки равен $k = 17$, то дискретный статистический ряд относительных частот имеет вид

x_i	-5	-3	-1	0	1	3
w_i	$\frac{2}{17}$	$\frac{2}{17}$	$\frac{3}{17}$	$\frac{5}{17}$	$\frac{3}{17}$	$\frac{2}{17}$

Если имеется большое количество исходных выборочных данных, близких по величине (например, при наблюдении значений непрерывной СВ), то исходные выборочные данные группируют в частичные полуинтервалы и определяют частоты (относительные частоты) попадания вариант в эти полуинтервалы, В результате получаем *интервальный статистический ряд*

Частичные полуинтервалы	$[x_1; x_2)$	$[x_2; x_3)$	…	$[x_{k-1}; x_k]$
n_i	n_1	n_2	…	n_k
$w_i = \frac{n_i}{n}$	w_1	w_2	…	w_k

Пример 2. При наблюдении за случайной величиной получены следующие числовые данные:

0.126, 0.215, 0.301, 0.525, 0.781, 1.034, 1.099, 1.137, 1.222, 1.288, 1.327, 1.472, 1.569, 1.620, 1.774, 1.835, 2.012, 2.074, 2.134, 2.154, 2.231, 2.279, 2.333, 2.367, 2.456, 2.543, 2.576, 2.667, 2.753, 2.826, 2.893, 2.965, 3.012, 3.095, 3.123, 3.145, 3.234, 3.278, 3,321, 3.356, 3.456, 3.492, 3.539, 3.548, 3.648, 3.684, 3.734, 3.771, 3.801, 3.874, 3.921, 3.989, 4.002, 4.015, 4.031, 4.063, 4.093, 4.115, 4.137, 4.155,

4.178, 4.187, 4.217, 4.222, 4.251, 4.276, 4.287, 4.302, 4.327, 4.365, 4.376, 4.388, 4.421, 4.432, 4.458, 4.469, 4.475, 4.531, 4.548, 4.558, 4.579, 4.581, 4.625, 4.634, 4.643, 4.684, 4.687, 4.732, 4.734, 4.752, 4.765, 4.781, 4.823, 4.805, 4.844, 4.850, 4.912, 4.942, 4.955, 4.983, 4.991, 5.001, 5.045, 5.056, 5.081, 5.099, 5.111, 5.123, 5.135, 5.159, 5.167, 5.204, 5.226, 5.234, 5.269, 5.271, 5.319, 5.323, 5.347, 5.352, 5.384, 5.418, 5.439, 5.452, 5.469, 5.478, 5.512, 5.523, 5.544, 5.575, 5.578, 5.621, 5.633, 5.644, 5.673, 5.687, 5.713, 5.723, 5.749, 5.756, 5.787, 5.821, 5.848, 5.867, 5.876, 5.921, 5.923, 5.945, 5.987, 6.052, 6.079, 6.123, 6.173, 6.234, 6.271, 6.329, 6.395, 6.452, 6.464, 6.511, 6.585, 6.673, 6.685, 6.754, 6.761, 6.823, 6.867, 6.898, 6.945, 7.099, 7.134, 7.156, 7.224, 7.267, 7.342, 7.389, 7.447, 7.478, 7.549, 7.676, 7.765, 7.834, 7.888, 7.984, 8.012, 8.134, 8.236, 8.341, 8.436, 8.576, 8.659, 8.753, 8.888, 8.903, 9.129, 9.267, 9.332, 9.567, 9.789, 9.879.

Разбивая интервал наблюдения на пять равных частей, составить интервальный ряд относительных частот.

Решение. Все наблюдаемые значения содержатся в отрезке $[0; 10]$. Поскольку, по условию, этот отрезок надо разбить на пять равных частей, то этими частями будут полуинтервалы длины 2: $[0; 2)$, $[2; 4)$, $[4; 6)$, $[6; 8)$, $[8; 10]$. Сначала составим интервальный ряд частот, подсчитывая количество значений СВ на каждом из пяти полуинтервалов:

$[x_i, x_{i+1})$	$[0; 2)$	$[2; 4)$	$[4; 6)$	$[6; 8)$	$[8; 10]$
n_i	16	36	97	35	16

Объем выборки равен $n = 16 + 36 + 97 + 35 + 16 = 200$. Разделив каждую частоту n_i на $n = 200$, получим интервальный ряд относительных частот

$[x_i, x_{i+1})$	$[0; 2)$	$[2; 4)$	$[4; 6)$	$[6; 8)$	$[8; 10]$
w_i	0,08	0,18	0,485	0,175	0,08

Дискретные статистические ряды можно представить в графической форме в виде *полигона частот* или *полигона относительных частот*. Полигон относительных частот является аналогом многоугольника распределения ДСВ и строится так же.

Пример 3. Задан дискретный ряд частот. Построить полигон относительных частот.

x_i	10	12	14	16	18	20	22
n_i	8	20	36	20	40	20	6

Решение. Объем выборки равен $n = 8 + 20 + 36 + 20 + 40 + 20 + 6 = 200$. Преобразуем дискретный ряд частот в простой ряд относительных частот

x_i	10	12	14	16	18	20	22

w_i	0,04	0,1	0,18	0,35	0,2	0,1	0,03

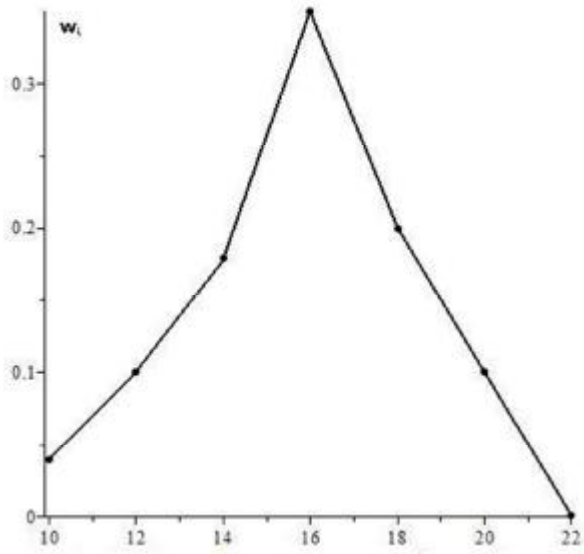


Рис.33.1 Полигон относительных частот к Примеру 3

Проверка условия нормировки: $0,04+0,1+0,18+0,35+0,2+0,1+0,03=1$.

Строим полигон относительных частот (Рис. 33.1).

Интервальные статистические ряды обычно представляют в виде *гистограммы*.

Гистограмма представляет собой ступенчатую фигуру, образованную прямоугольниками, основаниями которых служат частичные полуинтервалы интервального статистического ряда, а высоты равны плотностям частот (или плотностям относительных частот). Плотности частот (относительных частот) вычисляются как результаты деления частот (относительных частот) на длины оснований прямоугольников.

Число частичных полуинтервалов при составлении интервального ряда при необходимости рекомендуется определять по формуле Стерджеса $k \approx 1+3,322\ln n$, где n - объем выборки. Длина отрезка основания каждого прямоугольника в гистограмме определяется по формуле $h=\dfrac{x_{\max}-x_{\min}}{k}$, где $x_{\max}$ - наибольшее значение варианты, $x_{\min}$ - ее наименьшее значение.

Пример 4. Задан интервальный статистический ряд. Построить гистограмму относительных частот.

$[x_i,\ x_{i+1})$	$[4;\ 6)$	$[6;\ 8)$	$[8;\ 10)$	$[10;\ 12)$	$[12;\ 14)$	$[14;\ 16)$	$[16;\ 18]$
n_i	3	12	21	45	24	12	4

Объем выборки равен $n=3+12+21+45+24+12+4=121$, длина частичного полуинтервала равна $h=\dfrac{x_{\max}-x_{\min}}{k}=\dfrac{18-4}{7}=2$. Вычислим относительные частоты и плотности относительных частот.

$[x_i,\ x_{i+1})$	$[4;\ 6)$	$[6;\ 8)$	$[8;\ 10)$	$[10;\ 12)$	$[12;\ 14)$	$[14;\ 16)$	$[16;\ 18]$
n_i	3	12	21	45	24	12	4
$w_i=\dfrac{n_i}{n}$	0,0248	0,0991	0,1736	0,3719	0,1983	0,0991	0,0331
$\dfrac{w_i}{h}$	0,0124	0,0496	0,0868	0,1860	0,0991	0,0496	0,0166

Строим гистограмму (Рис.33.2).

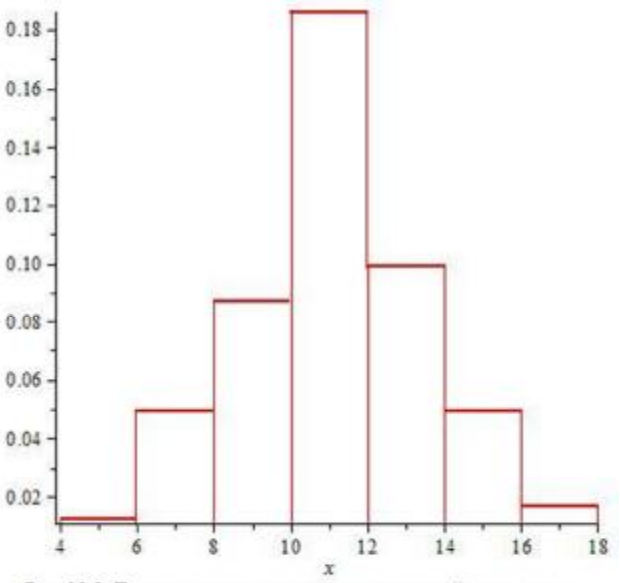


Рис. 33.2 Гистограмма плотности относительной частоты к Примеру 4

Замечание. Если рассматривать гистограмму как график кусочно-постоянной функции $f^*(x)$, равной нулю при $x < x_{\min}$ и при $x > x_{\max}$, то эта функция будет обладать свойствами $f^*(x) \geq 0$ и $\int_{-\infty}^{\infty} f^*(x)dx = 1$ (интеграл в данном случае равен сумме площадей прямоугольников, составляющих гистограмму). Значит, гистограмма является дискретным аналогом плотности распределения непрерывной СВ.

Наиболее полное представление о генеральной совокупности как о случайной величине могла бы дать теоретическая функция распределения $F(x)$, но ее невозможно определить, так как известна только выборка. По этим выборочным данным определяют *эмпирическую функцию распределения* $F^*(x)$, которая является приближенным представлением теоретической функции распределения генеральной совокупности.

Эмпирическая функция распределения для дискретной СВ определяется по выборке с использованием формулы

$$F^*(x) = \frac{n_x}{n} = \frac{1}{n}\sum_{x_i \leq x} n_i = \sum_{x_i \leq x} w_i ,$$

где n_x - количество вариант выборки, меньших данного числа x. Из того, что дискретный статистический ряд является аналогом ряда распределения и из формулы $F^*(x) = \sum_{x_i \leq x} w_i$ следует, что эмпирическая функция распределения обладает всеми свойствами функции распределения СВ и метод ее нахождения идентичен методу нахождения теоретической функции распределения СВ. Кроме того, при выполнении ряда естественных условий (в первую очередь – условия репрезентативности выборок), можно утверждать, что $\lim_{n\to\infty} F^*(x) = F(x)$.

Пример 5. Задан дискретный статистический ряд относительных частот. Найти эмпирическую функцию распределения и построить ее график.

x_i	10	12	14	16	18	20	22
w_i	0,04	0,1	0,18	0,35	0,2	0,1	0,03

Решение. При определении эмпирической функции распределения ДСВ будем пользоваться формулой $F^*(x)=\sum_{x_i<x} w_i$. При $x\le 10$ имеем $F^*(x)=0$, При $x\in(10;\,12]$ имеем $F^*(x)=0{,}04$; при $x\in(12;\,14]$ - $F^*(x)=0{,}04+0{,}1=0{,}14$ и т.д., при $x>22$ $F^*(x)=1$.

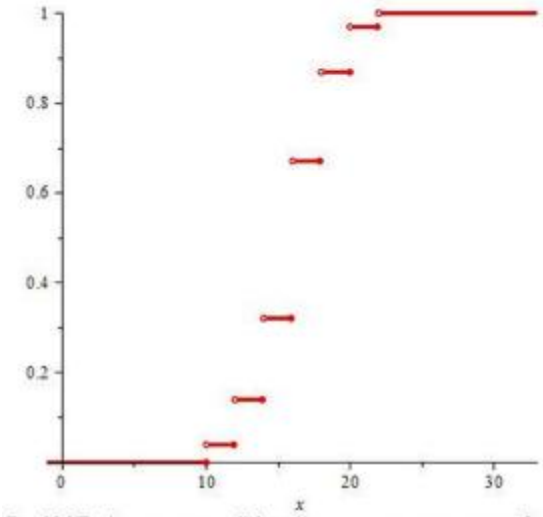


Рис. 33.3 График эмпирической функции распределения к примеру 5

$$F^*(x)=\begin{cases} 0,\ при\ x\le 10\\ 0{,}04,\ при\ x\in(10;\,12]\\ 0{,}14,\ при\ x\in(12;\,14]\\ 0{,}32,\ при\ x\in(14;\,16]\\ 0{,}67,\ при\ x\in(16;\,18]\\ 0{,}87,\ при\ x\in(18;\,20]\\ 0{,}97,\ при\ x\in(20;\,22]\\ 1,\ при\ x>22 \end{cases}$$

По результатам вычислений строим график эмпирической функции распределения (Рис. 33.3).

Если известно, что генеральная совокупность является непрерывной СВ, то, по определению, функция распределения этой СВ непрерывна. Поэтому вместо ступенчатой эмпирической функции распределения необходимо иметь непрерывную функцию, приближенно представляющую функцию распределения генеральной совокупности.

В этом случае сначала составляется ступенчатая выборочная функция относительных частот $F_n^*(x)$ по методу, рассмотренному в Примере 5. Затем в той же системе координат строят эмпирическую функцию распределения $F^*(x)$ как непрерывную ломаную линию из отрезков прямых, которые лежат в пределах частичных полуинтервалов. В каждом таком полуинтервале $[x_i;\,x_{i+1})$ отрезок прямой проходит через точки $(x_i;\,F_n^*(x_i))$ и $(x_{i+1};\,F_n^*(x_{i+1}))$.

Для полноты картины полученная ломаная дополнятся лучами $\begin{cases} y=0\\ x<x_{\min}\end{cases}$ и $\begin{cases} y=1\\ x>x_{\max}\end{cases}$. Данная ломаная называется эмпирической функцией распределения или *кумулятивной кривой (кумулятой).*

Пример 6. Задан интервальный ряд относительных частот. Построить кумуляту.

$[x_i,\,x_{i+1})$	$[4;\,6)$	$[6;\,8)$	$[8;\,10)$	$[10;\,12)$	$[12;\,14)$	$[14;\,16)$	$[16;\,18]$
n_i	3	12	21	45	24	12	4
$w_i=\frac{n_i}{n}$	0,0248	0,0991	0,1736	0,3719	0,1983	0,0991	0,0331

Решение. $F_n^*(4)=0$, $F_n^*(6)=0{,}0248$, $F_n^*(8)=0{,}1239$, $F_n^*(10)=0{,}2975$, $F_n^*(12)=0{,}6694$, $F_n^*(14)=0{,}8677$, $F_n^*(16)=0{,}9668$, $F_n^*(18)=1$.

Графики выборочной функции относительных частот $F_n^*(x)$ и кумуляты $F^*(x)$ приведены на рис.33.4.

Кумулята обладает всеми свойствами функции распределения СВ.

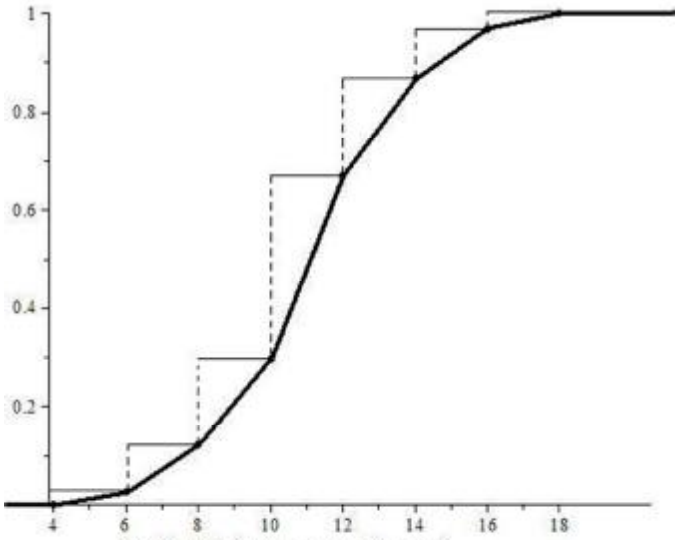


Рис. 33.4. График кумуляты к Примеру 6

7.3 **Числовые характеристики (параметры) генеральной совокупности и их точечные оценки**

В теории вероятностей различают числовые характеристики СВ ($M(X), D(X)$, и т.д.) и параметры закона распределения. При этом числовые характеристики выражаются через параметры с помощью определенных функциональных зависимостей (Таблица 1)

Таблица 1

Закон распределения	Параметры	Числовые характеристики
Биномиальный	p, n	$M(X)=np, D(X)=np(1-p), \sigma(X)=\sqrt{np(1-p)}$
Пуассоновский	λ	$M(X)=\lambda, D(X)=\lambda, \sigma(X)=\sqrt{\lambda}$
Равномерный	a, b	$M(X)=\frac{a+b}{2}, D(X)=\frac{(b-a)^2}{12}, \sigma(X)=\frac{b-a}{2\sqrt{3}}$
Нормальный	a, σ	$M(X)=a, D(X)=\sigma^2, \sigma(X)=\sigma$
Показательный	λ	$M(X)=\frac{1}{\lambda}, D(X)=\frac{1}{\lambda^2}, \sigma(X)=\frac{1}{\lambda}$

В математической статистике термин «параметр» используется в обобщенном смысле – под параметрами подразумевают как числовые характеристики, так и собственно параметры генеральной совокупности как случайной величины. Основными параметрами (числовыми характеристиками) генеральной совокупности являются:

- *генеральная средняя* $\overline{x}_Г$, равная математическому ожиданию генеральной совокупности как случайной величины;

- *генеральная дисперсия* $\overline{D}_Г$, равная дисперсии генеральной совокупности как случайной величины.

Замечание. Черта над параметрами $\overline{x}_Г$ и $\overline{D}_Г$ является условным знаком средней величины.

Генеральная совокупность в задачах математической статистики, как правило, неизвестна. Поэтому неизвестны и параметры (числовые характеристики) генеральной совокупности. Обычно известна выборка, на основе которой необходимо приближенно определить (оценить) параметры генеральной совокупности.

Приближенное числовое значение параметра генеральной совокупности, найденное по имеющейся конкретной выборке, называют *точечной оценкой параметра*.

Точечной оценкой математического ожидания (генеральной средней) для генеральной совокупности служит ***выборочная средняя*** $\overline{x}_в$, правила вычисления которой аналогичны правилам вычисления математического ожидания ДСВ.

- если выборочные данные $x_1, x_2, ..., x_m$ никак не упорядочены (например, как в условии задач 33.2.2 – 33.2.8), то выборочная средняя равна среднему арифметическому всех выборочных данных: $\overline{x}_в = \dfrac{x_1, + x_2 + ... + x_m}{m}$;

- если выборка представлена дискретным статистическим рядом частот, то $\overline{x}_в = \dfrac{\sum\limits_{i=1}^{k} x_i n_i}{\sum\limits_{i=1}^{k} n_i}$;

- если выборка представлена дискретным статистическим рядом относительных частот, то $\overline{x}_в = \sum\limits_{i=1}^{k} x_i w_i$;

- если выборка представлена интервальным статистическим рядом (с частичными интервалами одинаковой длины), то предварительно находят середины частичных интервалов x_{icp} и вычисляют выборочную среднюю по формуле $\overline{x}_в = \dfrac{\sum\limits_{i=1}^{k} x_{icp} n_i}{\sum\limits_{i=1}^{k} n_i}$ или $\overline{x}_в = \sum\limits_{i=1}^{k} x_i w_i$;

Замечание. Выборочная средняя, найденная по одной выборке из генеральной совокупности может отличаться от выборочной средней, найденной по другой выборке. Однако, при условии

репрезентативности выборок значения этих выборочных средних будут достаточно близки. Такое свойство выборок называется устойчивостью выборочных средних. Далее, как и раньше, будем считать все выборки репрезентативными и, значит, будем предполагать устойчивость выборочных средних.

Точечной оценкой генеральной дисперсии служит ***выборочная дисперсия*** $\overline{D}_в$, правила вычисления которой аналогичны правилам вычисления дисперсии ДСВ.

- если выборочные данные $x_1, x_2, ..., x_m$ никак не упорядочены (например, как в условии задач 33.2.2 – 33.2.8), то выборочная дисперсия равна среднему арифметическому квадратов отклонений выборочных значений от выборочной средней: $\overline{D}_в = \frac{1}{m}\sum_{i=1}^{m}\left(x_i - \overline{x}_в\right)^2$;

- если выборка представлена дискретным статистическим рядом частот, то $\overline{D}_в = \frac{\sum_{i=1}^{k} n_i\left(x_i - \overline{x}_в\right)^2}{\sum_{i=1}^{k} n_i}$;

- если выборка представлена дискретным статистическим рядом относительных частот, то $\overline{D}_в = \sum_{i=1}^{k} \omega_i\left(x_i - \overline{x}_в\right)^2$;

- если выборка представлена интервальным статистическим рядом (с частичными интервалами одинаковой длины), то предварительно находят середины частичных интервалов x_{icp} и вычисляют выборочную среднюю по формуле $\overline{D}_в = \frac{\sum_{i=1}^{k} n_i\left(x_{icp} - \overline{x}_в\right)^2}{\sum_{i=1}^{k} n_i}$ или $\overline{D}_в = \sum_{i=1}^{k} w_i\left(x_{icp} - \overline{x}_в\right)^2$;

Замечание 3. Вычисление $\overline{D}_в$ можно существенно упростить, если ее вычислять по формуле $\overline{D}_в = \frac{\sum_{i=1}^{k} n_i x_i^2}{\sum_{i=1}^{k} n_i} - \left(\overline{x}_в\right)^2$ (для дискретного статистического ряда) или $\overline{D}_в = \frac{\sum_{i=1}^{k} n_i x_{icp}^2}{\sum_{i=1}^{k} n_i} - \left(\overline{x}_в\right)^2$ (для интервального статистического ряда).

Замечание 4. При малом объеме выборки $\left(n < 30 \div 50\right)$ используют *исправленную выборочную дисперсию,* которую вычисляют по формуле $s_в^2 = \frac{n}{n-1}\overline{D}_в$. Наиболее удобной для вычисления

исправленной точечной оценки дисперсии является формула $s_в^2 = \frac{n}{n-1} \cdot \overline{D}_в$ или аналогичная формула для интервального статистического ряда с использованием средних значений вариант для частичных полуинтервалов.

Точечная оценка среднего квадратичного отклонения, соответственно, находится по формуле $s_в = \sqrt{\frac{n}{n-1} \cdot \overline{D}_в}$

Рассмотрим примеры вычисления точечных оценок числовых характеристик дискретных и непрерывных генеральных совокупностей по известным выборкам.

Пример 7. Найти точечные оценки основных параметров генеральной совокупности по выборке, представленной дискретным статистическим рядом

x_i	10	12	14	16	18	20	22
n_i	8	20	36	70	40	20	6

Решение. $\overline{x}_в = \frac{\sum_{i=1}^{k} x_i n_i}{\sum_{i=1}^{k} n_i} = \frac{1}{200}\left(10 \cdot 8 + 12 \cdot 20 + 14 \cdot 36 + 16 \cdot 70 + 18 \cdot 40 + 20 \cdot 20 + 22 \cdot 6\right) \approx 15{,}98;$

Определяем выборочную дисперсию и выборочное СКО.

$$\overline{D}_в = \frac{\sum_{i=1}^{k} n_i x_i^2}{\sum_{i=1}^{k} n_i} - \left(\overline{x}_в\right)^2 = \left(10^2 \cdot 8 + 12^2 \cdot 20 + 14^2 \cdot 36 + 16^2 \cdot 70 + 18^2 \cdot 40 + 20^2 \cdot 20 + 22^2 \cdot 6\right) -$$

$$-\left(15{,}98\right)^2 \approx 7{,}24;\ s_в = \sqrt{\frac{n}{n-1} \cdot \overline{D}_в} = \sqrt{\frac{200}{199} \cdot 7{,}24} \approx 2{,}7$$

При больших объемах выборок $\left(n > 100\right)$ вычисления выборочных дисперсий и СКО обычно проводят без учета поправки $\frac{n}{n-1}$. Для сравнения приводим результаты вычислений с учетом и без учета поправки с точностью до второго знака после запятой: $\overline{D}_в \approx 7{,}24$, $s_в^2 \approx \frac{200}{199} \cdot 7{,}24 = 7{,}28$.

Пример 8. Дана выборка, представленная интервальным статистическим рядом. Найти точечные оценки основных параметров генеральной совокупности по выборочным данным.

$[x_i, x_{i+1})$	$[4; 6)$	$[6; 8)$	$[8; 10)$	$[10; 12)$	$[12; 14)$	$[14; 16)$	$[16; 18]$
n_i	3	12	21	45	24	12	4

Решение. Преобразуем интервальный статистический ряд в дискретный статистический ряд путем замены частичных полуинтервалов их средними значениями:

x_i	5	7	9	11	13	15	17
n_i	3	12	21	45	24	12	4

Находим выборочную среднюю по данным преобразованного ряда:

$$\bar{x}_в = \frac{\sum_{i=1}^{k} x_{icp} n_i}{\sum_{i=1}^{k} n_i} = \frac{1}{121}\left(5 \cdot 3 + 7 \cdot 12 + 9 \cdot 21 + 11 \cdot 45 + 13 \cdot 24 + 15 \cdot 12 + 17 \cdot 4\right) = \frac{1343}{121} \approx 11{,}1 .$$

Определяем выборочную дисперсию:

$$\overline{D}_в = \frac{\sum_{i=1}^{k} n_i x_{icp}^2}{\sum_{i=1}^{k} n_i} - \left(\bar{x}_в\right)^2 = \frac{1}{121}\left(5^2 \cdot 3 + 7^2 \cdot 12 + ... + 17^2 \cdot 4\right) - \left(11{,}1\right)^2 \approx 6{,}72 .$$

Вычисляем исправленное СКО: $s_в = \sqrt{\frac{n}{n-1} \cdot \overline{D}_в} = \sqrt{\frac{121}{120} \cdot 6{,}72} \approx 2{,}6$

Тема 8 Проверка статистических гипотез

8.1 **Исходные положения**

Статистической гипотезой называется любое утверждение о законах или параметрах распределения генеральных совокупностей, которое проверяется по выборочным данным с той или иной вероятностью.

Совместно с выдвигаемой гипотезой H_0 (*основной или нулевой гипотезой*) рассматривается и *альтернативная* или *конкурирующая гипотеза* H_1, состоящая в отрицании гипотезы H_0.

Для приложений наиболее важными являются следующие виды гипотез:

1. Гипотеза о законе распределения генеральной совокупности как СВ.

Такая гипотеза выдвигается на основе анализа результатов экспериментов, представленных в виде гистограммы или многоугольника распределения, на основе оценок асимметрии и эксцесса генеральной совокупности, а также на основе теоретических рассуждений и опыте аналогичных исследований.

2. Гипотеза о равенстве математических ожиданий двух генеральных совокупностей.

Такая гипотеза используется, например, если необходимо проверить, одинаковы ли основные параметры изделий, производимых на разных станках, в разных цехах, на разных предприятиях.

3. Гипотеза о равенстве дисперсий двух или более генеральных совокупностей.

Такую гипотезу используют, например, если необходимо сравнить случайные погрешности различных измерительных приборов, применяемых для контроля одной и той же величины.

4. Гипотеза об однородности выборки

Однородность выборки позволяет дать оценку того, что различные измерения проводились в одних и тех же условиях и дает ответ на вопрос, можно ли объединить несколько малых выборок в одну большую.

5. Гипотеза о наличии аномальных выбросов в результатах эксперимента.

Аномальные результаты желательно исключать из выборочных данных, но для того, чтобы считать данные аномальными, нужны достаточные статистические основания. По существу, эта гипотеза является частным случаем гипотезы об однородности выборки.

При статистической проверке гипотез могут быть допущены два вида ошибок.

При *ошибках первого рода* отвергается правильная гипотеза. Вероятность ошибки первого рода характеризуется *уровнем значимости* α, который выбирается из внешних соображений об опасности допущения указанной ошибки. Например, уровень значимости $\alpha = 0{,}01$ означает, что в среднем в одном случае из 100 возможна ошибка первого рода.

При *ошибках второго рода* принимается неправильная гипотеза как верная. Вероятность β ошибки второго рода также выбирается из внешних соображений с учетом уровня значимости.

При контроле качества продукции ошибки первого рода называют риском производителя (годная продукция признается негодной), а ошибки второго рода – риском потребителя (негодная продукция признается годной).

Для проверки правильности выдвинутой основной гипотезы H_0 используют специальную непрерывную случайную величину, называемую *статистическим критерием*, значение которого вычисляется по выборочным данным. При проверке гипотезы о виде закона распределения

генеральной совокупности статистический критерий называют также *критерием согласия*. При проверке всех остальных перечисленных выше видов гипотез статистический критерий называется *критерием значимости*.

Замечание. Любое утверждение о нулевой гипотезе H_0, полученное с помощью статистического критерия, является *правдоподобным*, то есть верным лишь с определенной вероятностью $\gamma = 1 - \alpha$, называемой *доверительной вероятностью*.

Рис. 34.1 Взаимное расположение критической области и области принятия гипотезы: а) правосторонняя критическая область б) левосторонняя критическая область в) двусторонняя критическая область

Общую область возможных значений статистического критерия как случайной величины можно разделить на непересекающиеся частные области: *область принятия гипотезы* и *критическую область*, в которой нулевая гипотеза отвергается. указанные области разделяются *критическими точками* $k_{кр}$, которые соответствуют критическим значениям критерия K. Возможные значения статистического критерия, найденные по выборочным данным, называются *наблюдаемыми (эмпирическими) значениями критерия* $(k_{набл} \in K)$.

Возможны три варианта расположения критической области относительно области принятия нулевой гипотезы на числовой оси возможных значений критерия (Рис.34.1).

1. Правосторонняя критическая область $(k_{кр}; \infty)$ определяется из условия $P(K > k_{кр}) = \alpha$ (Рис. 34.1 а)). Гипотеза H_0 принимается при $k_{набл} < k_{кр}$.
2. Левосторонняя критическая область $(-\infty; k_{кр})$ определяется из условия $P(K < k_{кр}) = \alpha$ (Рис. 34.1.б)). Гипотеза H_0 принимается при $k_{набл} > k_{кр}$.
3. Двусторонняя критическая область $(-\infty; k_{1кр}) \cup (k_{2кр}; \infty)$ определяется из условий $P(K < k_{1кр}) = P(K > k_{2кр}) = \frac{\alpha}{2}$. Гипотеза H_0 принимается при $k_{1кр} < k_{набл} < k_{2кр}$.

Для проверки разных гипотез используют разные статистические критерии. Важнейшими из них являются критерий χ^2 («хи - квадрат»), критерий Стьюдента и критерий Фишера. Для указанных

критериев составлены специальные таблицы (См. Приложение 1) критических значений, используемые в прикладных задачах.

8.2 **Проверка статистической гипотезы о законе распределения с помощью критерия Пирсона**

Общая схема проверки статистической гипотезы о законе распределения генеральной совокупности включает следующие этапы:

Первый этап. На основе выборочных данных определяют (при необходимости) точечные оценки параметров, строят гистограмму и на основе ее анализа выдвигают гипотезу H_0 о законе распределения генеральной совокупности. Если на основе гистограммы однозначное выдвижение гипотезы затруднительно (например, гистограмма визуально может напоминать как график плотности показательного распределения, так и часть графика плотности нормального распределения), то дополнительно проводят оценки асимметрии и эксцесса генеральной совокупности.

Второй этап. Из внешних соображений (соображений, не связанных с математической моделью) об опасности невыполнения выдвинутой гипотезы H_0 выбирают стандартный (согласованный с таблицами) уровень значимости α (0,1; 0,05; 0,01; 0,001).

Третий этап. Для выдвинутой гипотезы H_0 выбирают статистический критерий согласия с учетом соответствия выборочных данных условиям корректного применения выбранного статистического критерия.

Четвертый этап. Используя выборочные данные (при необходимости – после их преобразования для удовлетворения условиям применения критерия), вычисляют наблюдаемое значение критерия $k_{набл}$.

Пятый этап. По таблице критических значений выбранного критерия находят критическое значение $k_{кр}$, соответствующее выбранному уровню значимости. На основе сравнения $k_{набл}$ и $k_{кр}$ делают заключение о принятии гипотезы H_0 или ее отклонении.

Методику проверки статистических гипотез рассмотрим на конкретных примерах с использованием статистического критерия Пирсона, который задается формулой $k_{набл} = \chi^2 = \sum \frac{\left(n_i - n_i^{'}\right)^2}{n_i^{'}}$, где n_i - эмпирические частоты наблюдения вариант в i-ом частичном интервале, $n_i^{'}$ - теоретические частоты. Этот критерий годится для проверки гипотез о любом дискретном или непрерывном распределении генеральной совокупности. Различие, в зависимости от проверяемого вида распределения, будет заключаться только в вычислении теоретических частот, входящих в критерий Пирсона.

Критическая область для критерия Пирсона является правосторонней, поэтому основная гипотеза H_0 принимается при условии $k_{набл} < k_{кр}$.

Рассмотрим методику применения данного критерия для проверки вида распределения непрерывной СВ, заданной интервальным статистическим рядом.

Для корректного применения критерия Пирсона необходимо выполнение следующих условий:

1. Объем выборки более 50;
2. Количество эмпирических частот в каждом частичном интервале должно быть не меньше 5. Если это условие не выполняется, то объединяют соседние интервалы и их частоты суммируют. Далее используют преобразованный статистический ряд

При нахождении табличного критического значения критерия используется постоянная величина, называемая *количеством степеней свободы,* которая вычисляется по формуле $m = n - (s+1)$, где n - количество частичных интервалов в преобразованном статистическом ряду, а s - количество параметров гипотетического распределения.

Например, количество параметров в нормальном законе распределения равно двум (a и σ), поэтому при проверке гипотезы о нормальном законе распределения число степеней свободы определяется по формуле $m = n - 3$. Количество параметров в показательном законе распределения равно одному (λ), поэтому при проверке гипотезы о показательном законе распределения число степеней свободы определяется по формуле $m = n - 2$. Количество параметров в равномерном законе распределения равно двум (a и b), поэтому при проверке гипотезы о равномерном законе распределения число степеней свободы определяется по формуле $m = n - 3$.

Замечание 1. Термин «число степеней свободы» заимствован из аналитической механики, в которой его используют, в частности, для описания общего количества кинематически независимых движений (обобщенных координат) несвободного твердого тела, часть движений которого ограничена контактными соединениями с другими телами.

В математической статистике «число степеней свободы» - это общее количество независимых частичных интервалов, которое находят как разность между количеством частичных интервалов в преобразованном статистическом ряду и количеством функциональных соотношений $(s+1)$, характеризующих объем выборки и точечные оценки параметров гипотетической генеральной совокупности (s).

Рассмотрим конкретные примеры

Пример 1. Выборочные данные представлены интервальным статистическим рядом

$[x_i, x_{i+1})$	$[1; 2)$	$[2; 3)$	$[3; 4)$	$[4; 5)$	$[5; 6)$	$[6; 7)$	$[7; 8)$	$[8 : 9]$
n_i	14	15	16	15	14	17	15	14

При уровне значимости $\alpha = 0{,}05$ проверить гипотезу о равномерном распределении генеральной совокупности по критерию Пирсона.

Решение. Объем выборки равен $n = 14+15+16+15+14+17+15+14 = 120 > 50$, количество всех эмпирических частот в каждом частичном полуинтервале $n_i > 5$.

Первый этап. Гипотезу о равномерном распределении можно выдвинуть как на основании гистограммы, так и просто заметив, что все эмпирические частоты не сильно отличаются друг от друга.

Второй этап. Уровень значимости $\alpha = 0{,}05$ задан в условии задачи.

Третий этап. Статистический критерий также задан по условию задачи.

Четвертый этап. Для применения критерия Пирсона предстоит определить теоретические частоты. Равномерное распределение обладает тем свойством, что вероятность попадания в интервал зависит только от длины интервала, и не зависит от расположения этого интервала на числовой прямой. Все частичные интервалы в заданном статистическом ряду имеют одну и ту же длину. Значит, теоретически на каждый интервал должно попасть одинаковое количество наблюдений (значений выборки). Поскольку объем выборки равен 120, а количество частичных интервалов равно 8, то все теоретические частоты равны $n_i^{'} = \frac{120}{8} = 15$. Находим наблюдаемое значение критерия:

$$k_{набл} = \sum \frac{\left(n_i - n_i^{'}\right)^2}{n_i^{'}} = \frac{(14-15)^2}{15} + \frac{(15-15)^2}{15} + \frac{(15-16)^2}{15} + \frac{(15-15)^2}{15} + \frac{(14-15)^2}{15} +$$

$$+ \frac{(17-15)^2}{15} + \frac{(15-15)^2}{15} + \frac{(14-15)^2}{15} = \frac{1+1+1+4+1}{15} = \frac{8}{15} \approx 0{,}5.$$

Пятый этап. Число степеней свободы равно $m = n - 3 = 8 - 3 = 5$, уровень значимости по условию равен $\alpha = 0{,}05$. По заданному уровню значимости и числу степеней свободы по таблице 8 (Приложение 1) находим $k_{кр} = 11{,}1$. Поскольку $k_{набл} = 0{,}5 < k_{кр} = 11{,}1$, гипотеза о равномерном распределении генеральной совокупности принимается с вероятностью 95% (на самом деле, из таблицы критических значений видно, что эту гипотезу можно принять и с вероятностью 99%).

Пример 2. Выборочные данные представлены интервальным статистическим рядом

$[x_i, x_{i+1})$	$[4; 6)$	$[6; 8)$	$[8; 10)$	$[10; 12)$	$[12; 14)$	$[14; 16)$	$[16; 18]$
n_i	3	12	21	45	24	12	4

При уровне значимости $\alpha = 0{,}01$ проверить гипотезу о нормальном распределении генеральной совокупности по критерию Пирсона.

Решение. Объем выборки равен $n = 3+12+21+45+24+12+4 = 121 > 50$, однако объемы эмпирических частот не во всех частичных полуинтервалах больше 5. Поэтому объединяем соседние интервалы слева и справа и суммируем их эмпирические частоты. Кроме того, учитывая, что проверяется гипотеза о нормальном распределении генеральной совокупности, крайние значения вариант справа и слева считаем неограниченными и заменяем бесконечностями:

x_i	$(-\infty; 8)$	$[8; 10)$	$[10; 12)$	$[12; 14)$	$[14; \infty)$
n_i	15	21	45	24	16

Первый этап. Гистограмма, построенная по заданному интервальному статистическому ряду (пример 4, раздел 33.2), позволяет выдвинуть гипотезу о том, что выборка взята из генеральной совокупности с нормальным распределением. Определяем точечные оценки параметров методом, рассмотренным в 33.3.2: $\bar{x}_в = 11{,}1$, $s_в = 2{,}6$.

Второй этап. Уровень значимости $\alpha = 0{,}05$ задан в условии задачи.

Третий этап. Статистический критерий также задан по условию задачи.

Четвертый этап. Теоретические частоты нормального закона распределения определяются по формулам: $n_i' = np_i$, где n - объем выборки, $p_i = P(x_i < X < x_{i+1}) = \Phi(u_{i+1}) - \Phi(u_i)$, где Φ - интегральная функция Лапласа, $u_i = \frac{x_i - \bar{x}_в}{s_в}$, $u_{i+1} = \frac{x_{i+1} - \bar{x}_в}{s_в}$.

Для удобства вычислений p_i обычно составляют таблицу

Таблица 1

$[x_i, x_{i+1})$	u_i	u_{i+1}	$\Phi(u_i)$	$\Phi(u_{i+1})$	p_i
$(-\infty; 8]$	$-\infty$	-1,19	-0,5	-0,3830	0,1170
$(8; 10]$	-1,19	-0,42	-0,3830	-0,1628	0,2202
$(10; 12]$	-0,42	0,35	-0,1628	0,1368	0,2996
$(12; 14]$	0,35	1,12	0,1368	0,3686	0,2318
$(14; \infty)$	1,12	∞	0,3686	0,5	0,1314
Проверка : $\sum p_i = 1{,}000$, $k_{набл} = \sum \frac{(n_i - n_i')^2}{n_i'} =$					3,9418

Определяем наблюдаемое значение критерия Пирсона $k_{набл} = \chi^2_{набл} = \sum_{i=1}^{5} \frac{(n_i - np_i)^2}{np_i}$ в табличной форме.

Таблица 2

$[x_i, x_{i+1})$	n_i	p_i	np_i	$\frac{(n_i - np_i)^2}{np_i}$	Для проверки	
					n_i^2	$\frac{n_i^2}{np_i}$
$(-\infty; 8]$	15	0,1170	14,1570	0,0502	225	15,8932
$(8; 10]$	21	0,2202	26,6442	1,1956	441	16,5514
$(10; 12]$	45	0,2996	36,2516	2,1112	2025	55,8596
$(12; 14]$	24	0,2318	28,0478	0,5842	576	20,5364

$(14;\infty)$	16	0,1314	15,8994	0,00063	256	16,1012
$\sum$	121	1,0000	-	$k_{набл}=3{,}9418$	-	124, 9418

Для проверки используем равенство $\sum_{i=1}^{5}\frac{n_i^2}{np_i}-n=k_{набл}\Rightarrow 124{,}9418-121=3{,}9418=k_{набл}$.

По результатам вычислений принимаем $k_{набл}=3{,}9$.

Замечание 2 округление до одного знака после запятой проводим потому, что с такой точностью составляем таблицы для χ^2 - распределения (таблица 5, приложение 1).

Пятый этап. Число степеней свободы равно $m=n-3=5-3=2$, уровень значимости по условию равен $\alpha=0{,}01$. По заданному уровню значимости и числу степеней свободы по таблице 8 (Приложение 1) находим $k_{кр}=9{,}2$. Поскольку $k_{набл}=3{,}9<k_{кр}=9{,}2$, гипотеза о нормальном распределении генеральной совокупности принимается с уровнем доверия 99%.

Замечание 3. При большом количестве частичных интервалов велика вероятность допущения ошибки в вычислениях, поэтому рекомендуется на каждом этапе вычисления проводить хотя бы минимальную проверку. Так, во втором и третьем столбцах таблицы 1 количество положительных и отрицательных чисел должно либо совпадать, либо отличаться не больше, чем на 1 (в нашем примере это 2 и 3) В четвертом и пятом столбцах отрицательным значениям u_i и u_{i+1} должны соответствовать отрицательные значения функции Лапласа, а положительным – положительные значения. Сумма чисел шестого столбца (вероятностей) должна равняться единице или незначительно (за счет погрешностей вычисления) от нее отличаться. В нашем примере $\sum_{i=1}^{5}p_i=0{,}117+0{,}2202+0{,}2996+0{,}2318+0{,}1314=1{,}0000$.

Замечание 4. Проверку гипотезы по критерию Пирсона целесообразно дополнить проверкой с помощью критерия Романовского, содержащего в своем составе χ^2 -распределение.

Наблюдаемое значение критерия Романовского определяется так: $k_{r\ набл}=\frac{|k_{набл}-m|}{\sqrt{2m}}$. Критерий Романовского является правосторонним, его критическое значение постоянно $k_{r\ кр}=3$.

Для рассмотренного примера имеем: $k_{r\ набл}=\frac{|3{,}9-2|}{\sqrt{2\cdot 2}}=0{,}95<3$.

Таким образом, с помощью критерия Романовского дополнительно подтверждена справедливость гипотезы о нормальном распределении генеральной совокупности при уровне значимости $\alpha=0{,}01$.

Тема 9 Основы корреляционного и регрессионного анализа выборочных данных

9.1 **Исходные положения**

В предыдущих темах, посвященных математической статистике, рассматривались методы исследования по выборочным данным генеральных совокупностей, представляющих множества возможных значений одной СВ. Теперь будем рассматривать выборки из двух генеральных совокупностей или – из двумерной генеральной совокупности, рассматриваемой как двумерная СВ. Составляющие такой двумерной СВ в прикладной статистике обычно называют *признаками* или *показателями*.

Если изменение одной СВ X (одномерной составляющей) связано с изменением другой СВ Y, то такая взаимосвязь случайных величин X и Y называется *статистической (стохастической) зависимостью*. Частным случаем статистической зависимости является корреляция (Тема 31). Корреляционная зависимость характеризует взаимосвязь между возможными значениями одной СВ и средними значениями другой СВ. Такие зависимости изучаются с помощью корреляционно-регрессионного анализа.

Корреляционный анализ выборочных данных включает в себя следующие основные задачи:

- построение поля корреляции и (или) составление корреляционной таблицы;
- вычисление выборочного коэффициента корреляции и выборочных корреляционных отношений;
- проверка статистических гипотез о значимости корреляционной связи.

Для установления по выборочным данным конкретного вида зависимости выборочной условной средней какой-либо СВ от возможных значений (вариант) другой СВ используется *регрессионный анализ*.

Такие зависимости называются выборочными уравнениями регрессии Y на X $\left(\bar{y}_x^* = \varphi(x)\right)$ и X на Y $\left(\bar{x}_y^* = \psi(y)\right)$, где $\bar{y}_x^*$, $\bar{x}_y^*$ - точечные оценки условных математических ожиданий $M_X(Y)$ и $M_Y(X)$ генеральных совокупностей как случайных величин.

Кроме того, в регрессионном анализе разрабатываются методы выбора моделей регрессии и методы проверки статистических гипотез о регрессии.

9.2 **Качественные оценки корреляционной зависимости**

Исходные выборочные данные, как правило, являются не сгруппированными. При большом числе наблюдений с повторяющимися значениями вариант выборочные данные из двумерной

генеральной совокупности (X, Y) группируются и представляются в виде выборочной корреляционной таблицы с двумя входами (Таблица 1).

Таблица 1

$x_i \rightarrow$ $y_j \downarrow$	x_1	x_2	…	x_i	…	x_k	n_{yj}
y_1	n_{11}	n_{21}	…	n_{i1}	…	n_{k1}	n_{y1}
y_2	n_{12}	n_{22}	…	n_{i2}	…	n_{k2}	n_{y2}
…	…	…	…	…	…	…	…
y_j	n_{1j}	n_{2j}	…	n_{ij}	…	n_{kj}	n_{yj}
…	…	…	…	…	…	…	…
y_m	n_{1m}	n_{2m}	…	n_{im}	…	n_{km}	n_{ym}
n_{xi}	n_{x1}	n_{x2}	…	n_{xi}	…	n_{xk}	n

Замечание. Серым цветом выделены дополнительная строка и дополнительный столбец, представляющие собой дискретные статистические ряды частот случайных величин X и Y. При этом каждое число в выделенной строке равно сумме чисел вышерасположенного столбца, а каждое число в выделенном столбце равно сумме чисел расположенной слева строки. Невыделенные строки и столбцы представляют собой собственно корреляционную таблицу, являющуюся аналогом ряда распределения двумерной дискретной СВ (Тема 31).

Общее количество пар наблюдаемых значений (x_i, y_j) в двумерной выборке $n = \sum_{i=1}^{k} \sum_{j=1}^{m} n_{ij}$ называется *объемом выборки*.

Предварительный качественный анализ связи между одномерными составляющими X и Y двумерной СВ (X, Y) по выборочным данным проводят с помощью поля корреляции (диаграммы рассеивания). Полем корреляции называют графическое представление выборочных данных в виде множества точек (x_i, y_j) на плоскости в прямоугольной декартовой системе координат.

Поле корреляции позволяет визуально оценить наличие значимой или слабой корреляционной связи, определить тенденцию (положительная, отрицательная) и вид (линейная, нелинейная) связи (Рис. 35.1)

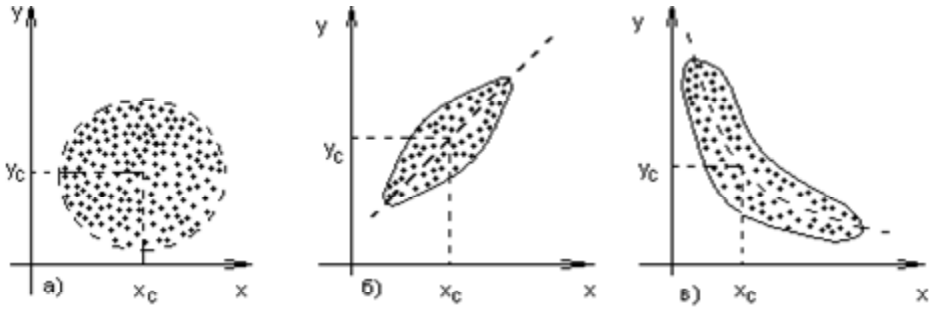


Рис. 35.1. Примеры полей корреляции: а) связь отсутствует; б) связь умеренная, положительная, близкая к линейной; в) связь умеренная, отрицательная, нелинейная

Пример 1. Дана несгруппированная выборка в виде пар измерений веса и роста десяти студентов:

x , кг.	67,5	68,2	70,8	70,2	65,8	68,3	72,4	73,0	71,8	69,1
y , м.	1,72	1,70	1,81	1,78	1,61	1,84	1,94	1,90	1,80	1,73

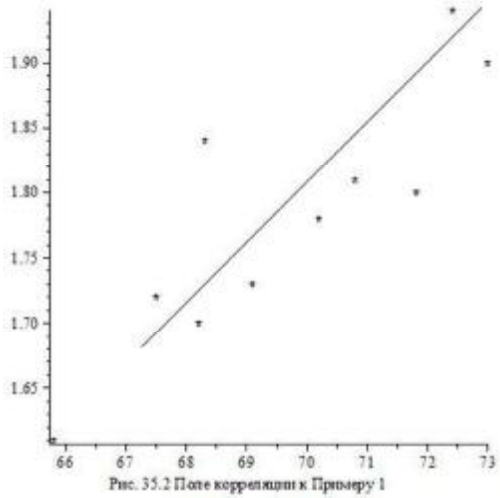


Рис. 35.2 Поле корреляции к Примеру 1

Построить поле корреляции и оценить визуально наличие связи, ее тенденцию и вид.

Решение. Из рисунка 35.2 следует, что имеет место значимая положительная корреляционная связь между весом и ростом студентов, которая близка к линейной связи.

Пример 2. Дана выборка в виде корреляционной таблицы 3. Построить поле корреляции и оценить визуально наличие корреляционной связи, ее тенденцию и вид.

Таблица 3

$x_i \rightarrow$ $y_j \downarrow$	10	15	20	25	30	35	n_{yj}
15	6	4	-	-	-	-	10
25	-	6	8	-	-	-	14
35	-	-	-	21	2	5	28
45	-	-	-	4	12	6	22
55	-	-	-	-	1	5	6
n_{xi}	6	10	8	25	15	16	80

Решение. При построении поля корреляции по корреляционной таблице возникает техническая сложность, связанная с тем, что у нескольких точек могут быть одинаковые координаты. Например, в приведенной таблице дана 21 точка с координатами (25, 35). Поскольку при построении поля корреляции речь идет лишь, по сути, о выдвижении некоторой гипотезы на основе визуальной информации и не проводятся никакие вычисления, то вместо нескольких точек с одинаковыми координатами можно строить точки с разными, но достаточно близкими координатами.

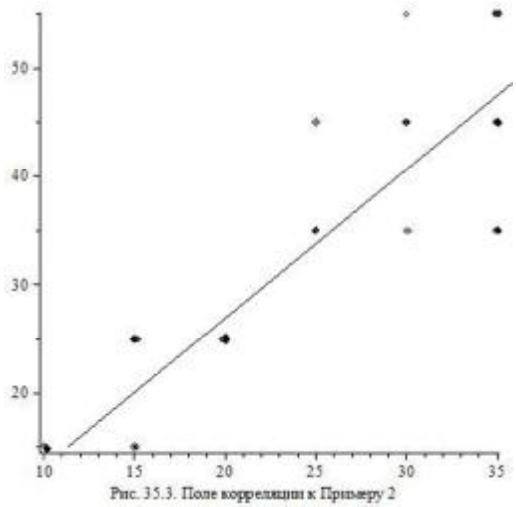


Рис. 35.3. Поле корреляции к Примеру 2

Строим поле корреляции (Рис. 35.3). На рисунке за счет наложения большого количества точек с близкими координатами появились «жирные» точки.

Из рисунка 35.3 следует, что имеет место значимая положительная, близкая к линейной корреляционная связь.

9.3 **Точечные оценки числовых характеристик одномерных составляющих двумерной генеральной совокупности (X, Y) и их корреляционной связи**

Для определения точечных оценок числовых характеристик случайных величин X и Y предварительно по корреляционной таблице находят выборочные одномерные ряды распределения (см. Замечание в 35.2)

X	x_i	x_1	x_2	$\cdots$	x_i	$\cdots$	x_k
	n_{xi}	n_{x1}	n_{x2}	$\cdots$	n_{xi}	$\cdots$	n_{xk}

Y	y_j	y_1	y_2	$\cdots$	y_j	$\cdots$	y_m
	n_{yj}	n_{y1}	n_{y2}	$\cdots$	n_{yj}	$\cdots$	n_{ym}

Пример 1. Дана выборка в виде корреляционной таблицы 3. Составить выборочные одномерные ряды распределения случайных величин X и Y

$x_i \rightarrow$ $y_j \downarrow$	10	15	20	25	30	35
15	6	4	-	-	-	-
25	-	6	8	-	-	-
35	-	-	-	21	2	5
45	-	-	-	4	12	6
55	-	-	-	-	1	5

Решение. $n_{x1} = n_{11} = 6$, $n_{x2} = n_{21} + n_{22} = 4 + 6 = 10$, $n_{x3} = 8$, $n_{x4} = 21 + 4 = 25$, $n_{x5} = 2 + 12 + 1 = 15$, $n_{x6} = 5 + 6 + 5 = 16$:

X	x_i	x_1	x_2	x_3	x_4	x_5	x_6
	n_{xi}	6	10	8	25	15	16

Аналогично, складывая числа по строкам, получим

Y	y_j	y_1	y_2	y_3	y_4	y_5
	n_{yj}	10	14	28	22	6

Проверка: $6+10+8+25+15+16=10+14+28+22+6=80$.

По данным выборочных одномерных рядов можно найти точечные оценки числовых характеристик случайных величин X и Y.

Оценки математического ожидания: $\bar{x}_в = \dfrac{\sum x_i n_{xi}}{n}$, $\bar{y}_в = \dfrac{\sum y_j n_{yj}}{n}$.

Оценки дисперсии: $D_x^* = \dfrac{\sum x_i^2 n_{xi}}{n} - (\bar{x}_в)^2$, $D_y^* = \dfrac{\sum y_i^2 n_{yj}}{n} - (\bar{y}_в)^2$.

Несмещенные оценки СКО: $s_x^* = \sqrt{\dfrac{n}{n-1} D_x^*}$, $s_y^* = \sqrt{\dfrac{n}{n-1} D_y^*}$

Оценку значимости и направления линейной связи между признаками X и Y производят по значению выборочного коэффициента корреляции $r_b = \dfrac{\sum_{i=1}^{k}\sum_{j=1}^{m} n_{ij} x_i y_j - n\bar{x}_в \bar{y}_в}{n\sigma_x^* \sigma_y^*}$. Можно доказать, что $|r_b| \le 1$ (см., например, 31.2).

Оценку тесноты линейной корреляционной связи проводят по абсолютному значению выборочного коэффициента корреляции.

При $|r_b| \le 0{,}1$ линейная связь между величинами X и Y отсутствует (либо связь нелинейна, либо нет вообще никакой связи).

При $0{,}1 < |r_b| \le 0{,}3$ связь слабая.

При $0{,}3 < |r_b| \le 0{,}65$ связь умеренная.

При $0{,}65 < |r_b| \le 0{,}8$ связь тесная

При $0{,}8 < |r_b| \le 0{,}95$ связь очень тесная.

При $|r_b| \ge 0{,}96$ связь принято считать функциональной, т.е. каждому возможному значению одной СВ соответствует ровно одно значение другой СВ.

Направление корреляционной связи определяется по знаку выборочного коэффициента корреляции:

- при $r_b > 0$ связь положительная (прямая);

- при $r_b < 0$ связь отрицательная (обратная).

Пример 2. Дана корреляционная таблица. Оценить значимость и направление корреляционной связи между признаками X и Y

$y_j \rightarrow$ $x_i \downarrow$	$y_1 = 110$	$y_2 = 120$	$y_3 = 130$	$y_4 = 140$	$y_5 = 150$
$x_1 = 2$	2	-	-	-	-
$x_2 = 7$	4	6	-	-	-

$x_3 = 12$	-	2	3	1	-
$x_4 = 17$	-	-	50	10	4
$x_5 = 22$	-	-	2	6	7
$x_{61} = 27$	-	-	-	-	3

Решение. Определяем объем выборки:

$$n = \sum_{i=1}^{k}\sum_{j=1}^{m} n_{ij} = 2+4+6+2+3+1+50+10+4+2+6+7+3+ = 100.$$

Находим выборочные одномерные законы распределения признаков X и Y:

x_i	2	7	12	17	22	27
n_{xi}	2	10	6	64	15	3

y_j	110	120	130	140	150
n_{yj}	6	8	55	17	14

Находим выборочные средние признаков X и Y:

$$\bar{x}_в = \frac{\sum x_i n_{xi}}{n} = \frac{1}{100}\left(2\cdot 2 + 7\cdot 10 + ... + 27\cdot 3\right) = 16{,}45;$$

$$\bar{y}_в = \frac{\sum y_j n_{yj}}{n} = \frac{1}{100}\left(110\cdot 6 + 120\cdot 8 + ... + 150\cdot 14\right) = 132{,}5.$$

Находим выборочные дисперсии и СКО. Так как $n = 100 > 50$, то поправку на смещенность оценки можно не учитывать:

$$D_x^* = \frac{\sum x_i^2 n_{xi}}{n} - \left(\bar{x}_в\right)^2 = \frac{1}{100}\left(2^2\cdot 2 + 7^2\cdot 10 + ... + 27^2\cdot 3\right) - \left(16{,}45\right)^2 \approx 22{,}4475$$

$$\sigma_x^* = \sqrt{D_x^*} \approx \sqrt{22{,}4475} \approx 4{,}7379$$

$$D_y^* = \frac{\sum y_j^2 n_{yj}}{n} - \left(\bar{y}_в\right)^2 = \frac{1}{100}\left(110^2\cdot 2 + 120^2\cdot 10 + ... + 150^2\cdot 3\right) - \left(132{,}5\right)^2 \approx 98{,}75;$$

$$\sigma_y^* = \sqrt{D_y^*} \approx \sqrt{98{,}75} \approx 9{,}9379$$

Находим величину $\sum_{i=1}^{k}\sum_{j=1}^{m} n_{ij} x_i y_j$:

$$\sum_{i=1}^{k}\sum_{j=1}^{m} n_{ij} x_i y_j = 2\cdot 2\cdot 110 + 4\cdot 7\cdot 110 + 6\cdot 7\cdot 120 + ... + 3\cdot 27\cdot 150 = 221750$$

Находим выборочный коэффициент корреляции:

$$r_b = \frac{\sum_{i=1}^{k}\sum_{j=1}^{m} n_{ij} x_i y_j - n\bar{x}_в \bar{y}_в}{n\sigma_x^* \sigma_y^*} = \frac{221750 - 100 \cdot 16{,}45 \cdot 132{,}5}{100 \cdot 4{,}7379 \cdot 9{,}9373} \approx 0{,}8045 .$$

Так как $r_b > 0$ и $0{,}8 < |r_b| \le 0{,}95$, имеем очень тесную прямую корреляционную связь между признаками X и Y.

Для оценки значимости корреляционной связи между признаками используют метод статистических гипотез. Обычно проверяют основную гипотезу $H_0 : r_Г = 0$ об отсутствии линейной корреляционной связи между признаками X и Y. Альтернативная гипотеза - $H_1 : r_Г \neq 0$. В качестве статистического критерия используется случайная величина T с распределением Стьюдента. Ее возможные значения имеют вид $t = \frac{r_в \sqrt{n-2}}{\sqrt{1-r_в^2}}$. Гипотеза об отсутствии корреляционной связи при известном уровне значимости α отвергается, если $t_{набл} > t_{кр} = t(\alpha, k)$, где $k = n - 2$ - число степеней свободы. Наблюдаемое значение статистического критерия $t_{набл} = \frac{r_в \sqrt{n-2}}{\sqrt{1-r_в^2}}$ определяется по данным конкретной выборки, а критическое значение $t_{кр} = t(\alpha, k)$ - по специальной таблице 7 (см. Приложение 1)

Для грубой оценки наличия корреляционной связи между признаками используют неравенство

$$|r_в| \cdot \sqrt{n-1} > 3 .$$

При выполнении этого неравенства имеет место корреляционная связь с уровнем значимости $\alpha = 0{,}01$

Если линейная корреляционная связь является значимой, то интервальную оценку генерального коэффициента корреляции проводят по формуле

$$r_в - 3 \cdot \frac{1-r_в^2}{\sqrt{n}} < r_Г < r_в + 3 \cdot \frac{1-r_в^2}{\sqrt{n}} .$$

Пример 3. По данным примера 2 проверить значимость выборочного коэффициента корреляции $r_в$ при уровне значимости $\alpha = 0{,}01$ и определить доверительный интервал для $r_Г$.

Решение. В примере 2 имеем $n = 100$ и получили $r_в \approx 0{,}8045$. Выдвигаем основную гипотезу о незначимости корреляционной связи $H_0 : r_Г = 0$. Находим $t_{набл} = \frac{r_в \sqrt{n-2}}{\sqrt{1-r_в^2}} = \frac{0{,}8045 \cdot \sqrt{98}}{\sqrt{1-0{,}8045^2}} \approx 13{,}408$. При заданном уровне значимости $\alpha = 0{,}01$ и числе степеней свободы $k = n - 2 = 98$ находим по таблице 7 приложения 1 критическое значение статистического критерия $2{,}62 < t_{кр} < 2{,}66$. Поскольку $t_{набл} > t_{кр}$, основная гипотеза о незначимости корреляционной связи отвергается. Корреляционная связь признаков X и Y считается значимой.

Грубая проверка нулевой гипотезы без таблицы дает $|r_в| \cdot \sqrt{n-1} = 0{,}8045 \cdot \sqrt{99} \approx 8 > 3$. При грубой проверке нулевая гипотеза также не подтверждается.

Определяем доверительный интервал для генерального коэффициента корреляции:

$$3 \cdot \frac{1 - r_в^2}{\sqrt{n}} = 3 \cdot \frac{1 - 0{,}8045^2}{\sqrt{100}} \approx 0{,}1058.$$

$$0{,}8045 - 0{,}1058 < r_Г < 0{,}8045 + 0{,}1058;\ r_Г \in (0{,}7;\ 0{,}91).$$

Задачи к разделу 35.3

35.3.1.- 35.3.6. В условиях задач **35.2.1.- 35.2.6.** оценить значимость и направление корреляционной связи между признаками X и Y, найти доверительный интервал для генерального коэффициента корреляции при уровне значимости $\alpha = 0{,}05$.

9.4 **Основы регрессионного анализа**

9.4.1 **Эмпирические графики регрессий двух признаков**

Основная задача регрессионного анализа заключается в установлении по выборочным данным приближенных аналитических зависимостей между признаками генеральной совокупности в виде уравнений регрессии.

В общем случае регрессией называют зависимость среднего значения одной случайной величины от значений другой СВ.

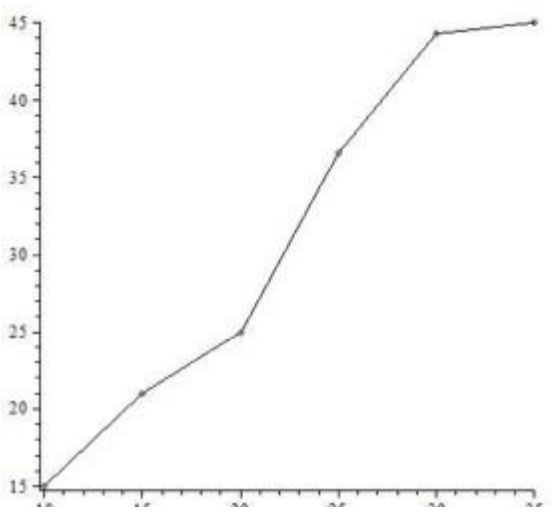
Если бы случайные величины X и Y были заданы (например, совместным рядом распределения), то методами теории вероятностей можно было бы найти уравнения регрессий: $M(Y \mid x) = \varphi_1(x)$, $M(X \mid y) = \varphi_2(y)$, где $M(Y \mid x)$, $M(X \mid y)$ - условные математические ожидания (см. 31.3).

В математической статистике, поскольку закон совместного распределения СВ X и Y неизвестен, то по выборочным данным находят оценки условных математических ожиданий. «Точечной оценкой» в этом случае является выборочное уравнение регрессии, которое каждому возможному значению одной случайной величины ставит в соответствие выборочное условное математическое ожидание (*выборочное условное среднее*) другой СВ.

Если задана выборочная корреляционная таблица (см. 35.2, Таблица 1), то выборочные условные средние вычисляют как групповые средние.

Групповой средней называется среднее арифметическое числовых значений признака, принадлежащих одной группе. В корреляционной таблице можно принять, что признак Y разбит на k групп, а признак X - на m групп. При этом каждая группа соответствует фиксированному значению соответствующей СВ.

В общем виде групповую среднюю для признака Y при фиксированном значении признака $X = x_i$



можно представить так: $\bar{y}^*_{x_i} = \dfrac{\sum_{j=1}^{m} y_j n_{ij}}{n_{xi}}$. Аналогично групповую среднюю для признака X при фиксированном значении признака $Y = y_j$ определяем формулой $\bar{x}^*_{y_j} = \dfrac{\sum_{i=1}^{k} x_i n_{ij}}{n_{yj}}$.

После вычисления всех групповых средних по корреляционной таблице можно построить эмпирический график регрессии $\bar{y}_{x_i} = f(x)$ в виде ломаной, состоящей из отрезков прямых с концами в точках $M_1(x_1, \bar{y}_{x1})$, $M_2(x_2, \bar{y}_{x2}),\ldots,$ $M_k(x_k, \bar{y}_{xk})$. Аналогично строится график регрессии $\bar{x}_{y_j} = \varphi(y)$ как ломаная с концами в точках $N_1(\bar{x}_{y1}, y_1)$, $N_2(\bar{x}_{y2}, y_2),\ldots,$ $N_m(\bar{x}_{ym}, y_m)$.

Пример 1. Найти групповые средние $\bar{y}^*_{x_i}$ и $\bar{x}^*_{y_j}$ по данным корреляционной таблицы 3 Примера 2 (см. 35.2) и построить эмпирические графики регрессии.

Решение. Групповые средние $\bar{y}^*_{x_i} = \dfrac{\sum_{j=1}^{m} y_j n_{ij}}{n_{xi}}$:

При $x_1 = 10$ получим $\bar{y}^*_{x1} = \dfrac{15 \cdot 6}{6} = 15$. При $x_2 = 15$ получим $\bar{y}^*_{x2} = \dfrac{15 \cdot 4 + 25 \cdot 6}{10} = 21$.

При $x_3 = 20$ получим $\bar{y}^*_{x3} = \dfrac{25 \cdot 8}{8} = 25$. При $x_4 = 25$ получим $\bar{y}^*_{x4} = \dfrac{35 \cdot 21 + 45 \cdot 4}{25} = 36{,}6$.

При $x_5 = 30$ получим $\bar{y}^*_{x5} = \dfrac{35 \cdot 2 + 45 \cdot 12 + 55 \cdot 1}{15} = 44{,}3$.

При $x_6 = 35$ получим $\bar{y}^*_{x6} = \dfrac{35 \cdot 5 + 45 \cdot 6 + 55 \cdot 5}{16} = 45$.

Аналогично находим групповые средние $\bar{x}^*_{y_j} = \dfrac{\sum_{i=1}^{k} x_i n_{ij}}{n_{yj}}$: $\bar{x}^*_{y1} = \dfrac{10 \cdot 6 + 15 \cdot 4}{10} = 12$,

$\bar{x}^*_{y2} = \dfrac{15 \cdot 6 + 20 \cdot 8}{14} = 17{,}8$,

$\bar{x}^*_{y3} = \dfrac{21 \cdot 25 + 2 \cdot 30 + 5 \cdot 35}{28} = 27{,}1$,

$\bar{x}^*_{y4} = \dfrac{4 \cdot 25 + 12 \cdot 30 + 6 \cdot 35}{22} = 30{,}4$,

$\bar{x}^*_{y5} = \dfrac{30 \cdot 1 + 35 \cdot 5}{6} = 34{,}2$.

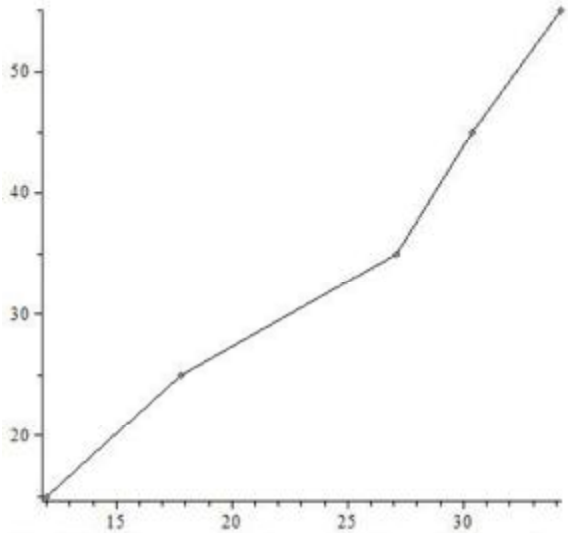


Рис. 35.5 График эмпирической регрессии $x^* = \varphi(y)$ в *примере* 1

По результатам проведенных вычислений строим графики эмпирических регрессий $\bar{y}_{x_i} = f(x)$ и $\bar{x}_{y_j} = \varphi(y)$ (Рис.35.4 и Рис 35.5).

Из визуального анализа графиков следует, что между признаками X и Y имеет место прямая корреляционная связь, близкая к линейной.

9.4.2 **Выборочные линейные среднеквадратические регрессии**

Эмпирические линии регрессий $\bar{y}_{x_i} = f(x)$ и $\bar{x}_{y_j} = \varphi(y)$ представляют собой линии, отражающие общую качественную тенденцию изменения групповой средней одного признака от другого.

Для количественной оценки тенденции изменения групповой средней одного признака от другого в первом (линейном) приближении используют линейные среднеквадратические регрессии

$$\bar{y}^* = \bar{y}_в + r_в \cdot \frac{\sigma_y^*}{\sigma_x^*}(x - \bar{x}_в) = \rho x + b, \qquad (1)$$

$$\bar{x}^* = \bar{x}_в + r_в \cdot \frac{\sigma_x^*}{\sigma_y^*}(y - \bar{y}_в) = \rho y + b \qquad (2)$$

где $\bar{x}_в$, $\bar{y}_в$ - общие выборочные средние, σ_x^*, σ_y^* - точечные оценки СКО, $r_в$ - выборочный коэффициент корреляции, являющийся точечной оценкой тесноты линейной связи между признаками X и Y гипотетической генеральной совокупности, $\rho = r_в \cdot \frac{\sigma_y^*}{\sigma_x^*}$ - коэффициент регрессии. Прямые (1) и (2) пересекаются в центре рассеяния $(\bar{x}_в, \bar{y}_в)$ в общем случае под острым углом $tg\varphi = \frac{1 - r_в^2}{|r_в|}$.

При $r_в = 0$ линейная корреляционная связь отсутствует и прямые линейных регрессий параллельны осям координат. При $r_в = \pm 1$ корреляционная связь переходит в линейную функциональную связь; при этом прямые (1) и (2) совпадают.

Из уравнений (1) и (2) следует, что регрессионный анализ производится на основе результатов корреляционного анализа. Поэтому в прикладных задачах корреляционный и регрессионный анализ, как правило, проводят совместно.

Если количество выборочных данных мало, то наличие линейной корреляционной связи и линейной среднеквадратичной регрессии можно определить по первичным, не сгруппированным, данным (Пример 1). Если выборка достаточно большая, то ее предварительно группируют в виде корреляционной таблицы.

Пример 1. На основании результатов измерений признаков X и Y генеральной совокупности

x_i	4	6	8	10	12
y_i	5	8	7	9	14

найти выборочную линейную регрессию Y на X. Найти выборочный коэффициент корреляции. Проверить значимость корреляционной связи при уровне значимости $\alpha = 0{,}05$. Определить приближенно какие значения признака Y должны соответствовать значениям $X = 5$ (линейная интерполяция) и $X = 13$ (линейная экстраполяция).

Решение. Коэффициенты ρ и b выборочного уравнения линейной регрессии удовлетворяют системе уравнений $\begin{cases} \rho\sum_{i=1}^{m} x_i^2 + b\sum_{i=1}^{m} x_i = \sum_{i=1}^{m} x_i y_i \\ \rho\sum_{i=1}^{m} x_i + mb = \sum_{i=1}^{m} y_i \end{cases}$, где m - количество результатов измерений (Практикум, Часть III, 17.5). составим вспомогательную таблицу:

№ измерения	x_i	y_i	$x_i y_i$	x_i^2	y_i^2
1	4	5	20	16	25
2	6	8	48	36	64
3	8	7	56	64	49
4	10	9	90	100	81
5	12	14	168	144	196
Сумма	40	43	382	360	416
Среднее	8	8,6	Не нужно	72	83,2

Подставляем данные предпоследней строки таблицы в систему уравнений: $\begin{cases} 360\rho + 40b = 382 \\ 40\rho + 5b = 43 \end{cases}$.

Решив систему, получим $\rho = 0{,}95$, $b = 1$. Таким образом, уравнение выборочной линейной регрессии Y на X имеет вид $\bar{y}^*_{x_i} = 0{,}95x + 1$.

Поскольку $\rho = r_в \cdot \frac{\sigma^*_y}{\sigma^*_x}$, то выборочный коэффициент корреляции равен $r_в = \rho \frac{\sigma^*_x}{\sigma^*_y}$.

Используя данные последней строки таблицы, получим $\sigma^*_x = \sqrt{\frac{5}{5-1}}\sqrt{72 - 8^2} \approx 3{,}164$, $\sigma^*_y = \sqrt{\frac{5}{5-1}}\sqrt{83{,}2 - 8{,}6^2} \approx 3{,}354$. Получаем: $r_в = \rho \frac{\sigma^*_x}{\sigma^*_y} = 0{,}95 \cdot \frac{3{,}164}{3{,}354} \approx 0{,}896$.

Проверяем значимость корреляционной связи. Выдвигаем гипотезу об отсутствии корреляционной связи $H_0 : r_Г = 0$ при уровне значимости $\alpha = 0{,}05$. Наблюдаемое значение t-критерия

Стьюдента равно $t_{набл} = \dfrac{r_в\sqrt{m-2}}{\sqrt{1-r_в^2}} = \dfrac{0{,}896\cdot\sqrt{5-2}}{\sqrt{1-0{,}896^2}} \approx 3{,}475$. Критическое значение t-критерия при $\alpha = 0{,}05$ и $k = 3$ равно (см. Приложение 1, таблица 7) $t_{rh} = t(\alpha,\, k) = t(0{,}05,\, 3) = 2{,}35$. Поскольку $t_{набл} > t_{кр}$, гипотеза об отсутствии (незначимости) корреляционной связи отвергается. Связь является значимой, прямой и очень тесной.

Подставив в уравнение регрессии значение $x = 5$, получим $\bar{y}^*_{x_i}(5) = 5{,}75$; аналогично получим $\bar{y}^*_{x_i}(13) = 13{,}35$.

Пример 2. Найти линейные регрессии по выборочным данным, представленным корреляционной таблицей. Построить графики найденных регрессий.

$x_i \rightarrow$ $y_j \downarrow$	10	15	20	25	30	35	n_{yj}
15	6	4	-	-	-	-	10
25	-	6	8	-	-	-	14
35	-	-	-	21	2	5	28
45	-	-	-	4	12	6	22
55	-	-	-	-	1	5	6
n_{xi}	6	10	8	25	15	16	80

Решение. В примере 2 (35.2) для этой корреляционной таблицы построено поле корреляции и сделан качественный вывод, что имеет место значимая положительная близкая к линейной корреляционная связь.

Для нахождения уравнений регрессий $\bar{y}^* = \bar{y}_в + r_в \cdot \dfrac{\sigma^*_y}{\sigma^*_x}(x - \bar{x}_в)$ и $\bar{x}^* = \bar{x}_в + r_в \cdot \dfrac{\sigma^*_x}{\sigma^*_y}(y - \bar{y}_в)$ понадобятся выборочные средние, оценки СКО и выборочный коэффициент корреляции.

Выборочные средние находим из одномерных рядов распределения (последняя строка и последний столбец корреляционной таблицы):

$$\bar{x}_в = \frac{10\cdot 6 + 15\cdot 10 + \ldots + 35\cdot 16}{80} = 25{,}06,$$

$$\bar{y}_в = \frac{15\cdot 10 + 25\cdot 14 + \ldots + 55\cdot 6}{80} = 35.$$

$$\bar{x}^2_в = \frac{10^2\cdot 6 + 15^2\cdot 10 + \ldots + 35^2\cdot 16}{80} = 684{,}6875,$$

$$\bar{y}^2_в = \frac{15^2\cdot 10 + 25^2\cdot 14 + \ldots + 55^2\cdot 6}{80} = 1350;$$

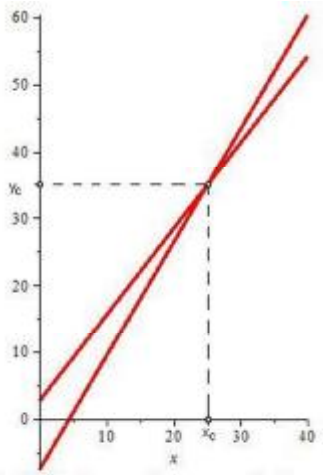


Рис. 35.6 Графики линейных регрессий к Примеру 2

$D_x^* = 684{,}6875 - (25{,}06)^2 = 56{,}6839$, $D_y^* = 1350 - 35^2 = 125$;

$\sigma_x^* = \sqrt{56{,}6839} \approx 7{,}53$, $\sigma_y^* = \sqrt{125} \approx 11{,}18$;

$$\sum_{i=1}^{6}\sum_{j=1}^{5} n_{ij} x_i y_j = 6 \cdot 10 \cdot 15 + 4 \cdot 15 \cdot 15 + 6 \cdot 15 \cdot 25 + \ldots + 5 \cdot 35 \cdot 55 = 76075;$$

$$r_b = \frac{\sum_{i=1}^{6}\sum_{j=1}^{5} n_{ij} x_i y_j - n\bar{x}_в \bar{y}_в}{n\sigma_x^* \sigma_y^*} = \frac{76075 - 80 \cdot 25{,}06 \cdot 35}{80 \cdot 7{,}53 \cdot 11{,}18} = 0{,}877;$$

$$r_в \cdot \frac{\sigma_y^*}{\sigma_x^*} = 0{,}877 \cdot \frac{11{,}18}{7{,}53} = 1.29;\ r_в \cdot \frac{\sigma_x^*}{\sigma_y^*} = 0{,}877 \cdot \frac{7{,}53}{11{,}18} = 0{,}59;$$

Уравнение линейной регрессии Y на X: $\bar{y}^* = 35 + 1.3(x - 25.06);\ \bar{y}^* = 1.29x + 2{,}64$.

Уравнение линейной регрессии X на Y: $\bar{x}^* = 25{,}06 + 0{,}59(y - 35);\ \bar{x}^* = 0{,}59y + 4{,}41$

Графики линейных регрессий представлены на Рис. 35.6.

Коэффициент корреляции определяет тесноту только линейной связи между признаками. Чтобы оценить в какой степени изменение независимого признака объясняет общую изменчивость зависимого признака, используют коэффициент детерминации, вычисляемый по формуле $R^2 = \frac{D_{регр}^{(y)}}{D_{общ}^{(y)}} = 1 - \frac{D_{ост}^{(y)}}{D_{общ}^{(y)}}$, где $D_{общ}^{(y)}$ - общая дисперсия, характеризующая общее рассеяние возможных значений зависимого признака Y относительно общей выборочной средней $\bar{y}_в$, $D_{регр}^{(y)}$ - дисперсия, обусловленная регрессией (наличием линейной корреляционной связи), $D_{ост}^{(y)}$ - остаточная дисперсия, обусловленная действием прочих факторов, вызывающих рассеивание зависимого признака относительно линейной регрессии. Значение коэффициента детерминации изменяется в пределах от 0 до 1; это значение показывает какая доля дисперсии объясняется изменением независимого признака. Если зависимый признак Y зависит только от одного независимого признака X (случай зависимости от нескольких признаков в рамках этой темы нами не рассматривался), то $R^2 = r_в^2$.

В примере 2 $r_b = 0{,}877$, значит, $R^2 = (0{,}877)^2 \approx 0{,}77$, значит, 77% общего случайного изменения признака Y объясняется изменением независимого признака X, а оставшиеся 23% связаны с влиянием других (неучтенных) факторов.