Министерство образования РФ
Вятский государственный университет
Факультет Прикладной Математики и Телекоммуникаций
Кафедра высшей математики

А.И.Карпей, А.Н.Рапопорт

Сборник задач по теории вероятностей и математической статистике

Утверждено
Учёным советом университета
в качестве учебного пособия

Киров 2003

Печатается по решению редакционно-издательского совета Вятского государственного университета

УДК 519.21/.22(07)
К261

Рецензент: заведующий кафедрой высшей математики и физики ВГСХА, доктор технических наук, профессор С.М.Решетников

Карпей А.И., Рапопорт А.Н. Сборник задач по теории вероятностей и математической статистике. Учебное пособие/ Под редакцией А.Н.Рапопорта.- Киров: Издательство ВятГУ, 2003. - 122с.

Редактор Е.Г. Козвонина
Компьютерная верстка М.Е. Родионовой

Подписано в печать Усл.печ.л. 7,6.
Бумага офсетная. Печать копир Aficio 1022
Заказ № 189. Тираж 52
Текст напечатан с оригинала - макета, представленного авторами.

610000, г. Киров, ул. Московская, 36.
Оформление обложки, изготовление – ПРИП ВятГУ.

Учебное пособие составлено в соответствии с программой по высшей математике. Содержит основные разделы курса теории вероятностей и математической статистики. Каждый раздел содержит теоретические сведения, снабжённые большим количеством разобранных примеров, и необходимое количество заданий.

Данное учебное пособие представляет собой вторую часть сборника задач по математике, изданного в 1997г.

Учебное издание

Оглавление

Предисловие

Настоящий сборник задач написан на основе многолетнего опыта преподавания теории вероятностей и математической статистики, а также опыта применения вероятностных методов для решения практических задач.

Сборник состоит из пяти глав. В начале каждой главы приведена сводка теоретических сведений и формул, необходимых для решения задач, помещённых в главе. Все теоретические положения проиллюстрированы разобранными примерами.

Задачи, имеющиеся в сборнике, весьма различны по трудности. Среди них есть как задачи, предназначенные для простого приобретения навыков применения готовых формул и теорем, так и более сложные задачи, решение которых требует изобретательности. Все задачи снабжены ответами. В интересах удобства пользования авторы отступили от традиционного разделения текстов сборника на «задачи» и «ответы» к ним. Авторы предпочли давать ответы или решения каждой задачи непосредственно вслед за её формулировкой.

Глава 1. Случайные события

1.1 Элементы комбинаторики

При подсчете вероятностей большую пользу оказывают комбинаторные формулы. Приведем наиболее важные из них.

Комбинации элементов, выбираемых из разных групп. Пусть имеется k групп элементов, причем i-я группа состоит из n_i элементов. Выберем по одному элементу из каждой группы. Две комбинации по k элементов считаются различными, если они отличаются или хотя бы одним элементом или их порядком. Тогда общее число способов, которыми можно произвести такой выбор, определяется соотношением

$$N=n_1 n_2 \ldots n_k,$$

которое называется ***основной формулой комбинаторики***.

Пример. Из трех студенческих групп отбирают по одному студенту для участия в олимпиадах по математике, физике и иностранному языку. Сколько различных троек можно выбрать, если в первой группе 18 студентов, во второй- 20, в третьей- 25 студентов ?

Решение: Число таких троек N= 18 $\cdot$20 $\cdot$25= 9000.

В том случае, когда все группы состоят из одинакового числа элементов, т.е. $n_1=n_2=\ldots=n_k = n$, можно считать, что каждый раз выбор производится из одной и той же группы, причем элемент после *выбора* снова возвращается в группу. Тогда число всех способов выбора равно n^k. Такой способ выбора носит название ***выборки с возвращением.***

Пример. Имеется кодовый замок, содержащий 4 диска. Каждый диск разделен на 8 секторов, на которых нанесены различные символы (цифры, буквы) Сколько различных комбинаций можно установить на этом замке?

Решение. На каждом диске устанавливается какой-то один из восьми возможных символов и таких дисков 4, поэтому число различных комбинаций $N=8^4$.

Пример. Группа студентов из r человек садится в электричку, насчитывающую $n \geq r$ вагонов. Предположим, что каждый студент выбирает свой вагон случайно и равновозможно оказывается в любом вагоне. Сколько различных способов рассадки студентов по вагонам возможно?

Решение. Каждый студент может выбрать один из n вагонов, так что всевозможных комбинаций $(i_1,i_2,\ldots,i_r)$, где i_k - номер вагона, выбранный k- м студентом, насчитывается $N=n^r$.

***Перестановкой из n* элементов** называется любой *упорядоченный набор* этих элементов. Так, всевозможными перестановками чисел 1, 2, 3 являются: (1,2,3), (1,3,2), (2,1,3), (2,3,1), (3,1,2), (3,2,1). Для определения числа различных перестановок из n элементов, которое обозначается через P_n, заметим, что на первом месте перестановки может стоять любой из n элементов, на втором- любой из $n-1$ оставшихся, на третьем любой из остальных $n-2$ и т.д. В силу основной формулы комбинаторики (в данном случае мы имеем n групп элементов размеров $n, n-1,\ldots 2,1)$ получаем $P_n=n(n-1)(n-2)\ldots 2\ 1=n!$.

Пример. Сколькими способами можно рассадить 5 человек на пяти стульях?

Решение. Число различных способов рассадки людей равно $P_5=5!=120$.

Размещением* из *n* элементов по *m называется любой *упорядоченный* набор из *n* различных элементов, выбранных из общей совокупности в n элементов. Выпишем для примера все размещения из четырех чисел 1,2,3,4 по два: (1,2), (1,3), (1,4), (2,1), (2,3), (2,4), (3,1), (3,2), (3,4), (4,1), (4,2), (4,3). Число размещений A_n^m (используется также запись $(n)_m$) подсчитывается точно так же, как и число перестановок: на первом месте может находиться любой из n элементов, на втором- любой из $n-1$ оставшихся, …, на m-м месте- любой из $n-m+1$ элементов. Снова воспользовавшись основной формулой комбинаторики (выбор осуществляется из групп размеров $n, n-1, \ldots, n-m+1$), имеем

$$A_n^m=n(n-1)\ldots(n-m+1)=\frac{n!}{(n-m)!}.$$

Способ выбора, приводящий к перестановкам и размещениям, носит название ***выборки без возвращения***.

Пример. В группе 20 студентов. Нужно выбрать старосту, зам. старосты и профорга. Сколькими способами это можно сделать?

Решение. Актив группы можно выбрать A_{20}^3 способами: $A_{20}^3=20(20-1)(20-2)=71640.$

Сочетанием* из *n* элементов по *m называется любой *неупорядоченный набор* из m различных элементов, выбранных из общей совокупности в n элементов. Сочетаниями из четырех чисел 1, 2, 3, 4 по два являются: (1,2), (1,3), (1,4), (2,3), (2,4) и (3,4). Для определения числа сочетаний C_n^m (употребляется также запись $\binom{n}{m}$) заметим, что сочетание от размещения отличается только тем, что входящие в него элементы неупорядочены. Но, как известно, m элементов можно упорядочить $m!$ способами. Значит, каждое сочетание соответствует $m!$ размещениям. Поэтому $C_n^m=\frac{A_n^m}{m!}=\frac{n!}{(n-m)!m!}$.

Число различных сочетаний из n элементов по n_1, n_2, …, n_k элементов (перестановки с повторением). Обобщением числа сочетаний является случай, когда совокупность из n элементов разбивается на фиксированное число k групп, причем фиксировано и число в каждой группе (в группе с номером i ровно n_i, элементов, $\sum_{i=1}^{k} n_i = n$). В остальном, элементы группируются совершенно произвольно. Число таких сочетаний подсчитаем следующим образом: в первую группу могут попасть любые n_1 элементов из n имеющихся первоначально, поскольку порядок выбора элементов несущественен, то это можно сделать C_n^m способами. Как только первая группа заполнена, у нас остается $n-n_1$ элементов, и вторую группу мы можем заполнить C_n^m способами. Продолжая эту процедуру и используя основную формулу комбинаторики, получаем, что *число различных* *сочетаний из n элементов по $n_1, n_2, ..., n_k$ элементов* есть: $N = C_n^{n_1} C_{n-n_1}^{n_2} ... C_{n-n_1-...-n_{k-2}}^{n_{k-1}} = \frac{n!}{n_1! n_2! ... n_k!}$, где $\sum_{i=1}^{k} n_i = n$.

Пример. Десять мужчин размещаются в гостинице в два трехместных и один четырехместный номер. Сколько существует способов их расселения?

Решение. Число различных способов их размещения $N = \frac{10!}{3!3!4!} = 4200$.

Формула Стирлинга. Во всех приведенных формулах встречается выражение *n!=n(n-1)(n-2)…2 1*. При больших *n* справедливо: $n! \approx \sqrt{2\pi} \cdot n^n \cdot e^{-n}$.

1.2. Понятие случайного события

Предметом теории вероятностей являются модели экспериментов со случайными исходами. Любой наблюдаемый результат интерпретируется как случайный исход опыта. ***Случайным*** будем называть *событие*, которое в результате эксперимента может произойти или не произойти.

Математическая формализация модели случайного эксперимента включает в себя: 1) построение множества элементарных исходов Ω; 2) описание поля событий для данного эксперимента.

Под *множеством элементарных исходов* понимают такое множество взаимоисключающих исходов, что в результате эксперимента всегда является один и только один исход. Любое подмножество данного множества Ω интерпретируется как *случайное событие* (обозначается буквами A,B,C,D,..). Совокупность всех наблюдаемых событий составляет *поле событий* для данного эксперимента. Говорят, что событие A произошло, если результатом эксперимента является элементарный исход ω, принадлежащий A ($\omega \in A$).

Достоверным событием назовем событие, которое всегда происходит (обозначается Ω или D). ***Невозможным*** событием назовем событие, которое никогда не происходит (обозначается $\varnothing$ или N).

Пример. Эксперимент состоит в подбрасывании один раз правильной игральной кости. Обозначим *X* число очков, выпавших на верхней грани кости. Описать множество элементарных исходов Ω и указать состав подмножеств, соответствующих следующим событиям: A={*X* кратно 3}, B={*X* нечетно}, C={*X*>3}, D={*X*<7}, N={*X* дробно}.

Решение. Введем обозначения для следующих наблюдаемых в данном эксперименте событий $\omega_к$={*X*=*k*}, *k=1,2,…,6.* На базе данных исходов сконструируем множество элементарных исходов $\Omega = \{\omega_1, \omega_2, ..., \omega_6\}$.

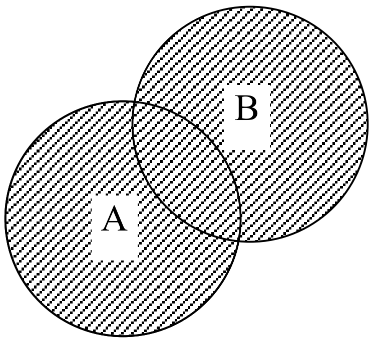
Все перечисленные события могут быть описаны как подмножества множества Ω. Действительно, $A = \{\omega_3, \omega_6\}, B = \{\omega_1, \omega_3, \omega_5\}, C = \{\omega_4, \omega_5, \omega 6\}$, $D = \{\omega_1, \omega_2, ..., \omega_6\}$, *N*=$\varnothing$.

Событие $\overline{A}$ назовем событием, ***противоположным*** событию A (не A), если оно происходит тогда и только тогда, когда не происходит событие A.

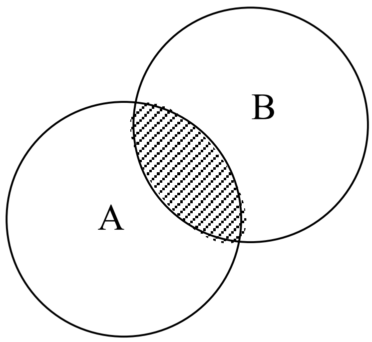
Суммой* или *объединением событий A и B назовем событие, обозначаемое A∪B или A+B, которое происходит тогда и только тогда, когда происходят A, или B, или оба вместе.

Произведением* или *пересечением событии A и B назовем событие, обозначаемое A∩B или AB, которое происходит тогда и только тогда, когда происходят события и A и B вместе.

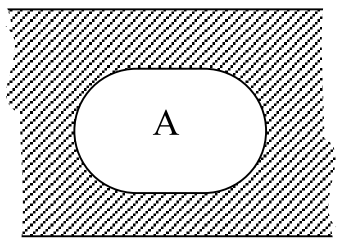
События A и B назовем ***несовместными* или *несовместимыми***, если AB=∅, т.е. такие события, которые не могут произойти одновременно.



A∪B



A∩B



Система F подмножеств множества Ω такая, что в результате применения любой из описанных операций к любым двум элементам системы снова получается элемент данной системы, называется *булевой алгеброй.* Под *наблюдаемым событием* будем понимать такое подмножество множества Ω, которое одновременно принадлежит и булевой алгебре F.

Пример. Произведено три выстрела по цели. События $A_к$={ попадание при к - м выстреле} (к=1,2,3). Записать следующие события: B={только одно попадание}, C={два попадания}, D={три попадания}, E={хотя бы одно попадание}, G={хотя бы один промах}, F={не меньше двух попадании}.

Решение. $B = A_1\overline{A_2}\,\overline{A_3} + \overline{A_1}A_2\overline{A_3} + \overline{A_1}\,\overline{A_2}A_3$;

$C = A_1A_2\overline{A_3} + A_1\overline{A_2}A_3 + \overline{A_1}A_2A_3$; D$=A_1A_2A_3$;

E$=A_1 + A_2 + A_3$; G$=\overline{A_1} + \overline{A_2} + \overline{A_3}$;

F=C+D+$A_1A_2\overline{A_3} + A_1\overline{A_2}A_3 + \overline{A_1}A_2A_3 + A_1A_2A_3$.

ЗАДАНИЯ

1.1. E_1 = {из трех событий A, B, C произойдет ровно одно}, F_1 = {из трех событий произойдет ровно два}.

Ответ: $E_1 = A\bar{B}\bar{C} + \bar{A}B\bar{C} + \bar{A}\bar{B}C, F_1 = AB\bar{C} + A\bar{B}C + \bar{A}BC$.

1.2. E_2 = {из трех событий A, B, C произойдет хотя бы одно}, F_2 = {из трех событий произойдет не меньше двух}.

Ответ: $E_2 = A + B + C, F_2 = F_1 + ABC$.

1.3. E_3 = {из трех событий A, B, C не произойдет ни одного}, F_3 = {из трех событий произойдет хотя бы два}, G = {из трех событий A, B, C не произойдет хотя бы одно}.

Ответ: $E_3 = \bar{A}\bar{B}\bar{C}, F_3 = F_1 + ABC = F_2, G = \bar{A} + \bar{B} + \bar{C} = \overline{ABC}$.

1.4. На отрезке [a, b] наудачу ставятся две точки. Пусть x и y - координаты этих точек. Изобразить на плоскости Oxy области, соответствующие событиям Ω, A, B, AB, A–B, A+B, где A = {вторая точка ближе к левому концу отрезка, чем первая точка к правому концу}, B = {расстояние между точками меньше половины длины отрезка}.

1.5. Произведено три выстрела из орудия по цели. Событие A_k = {попадание при k-м выстреле} (k = 1, 2, 3).

а) Выяснить состав множества Ω, выразив каждый элементарный исход w_i через события A_k.

б) Записать в алгебре событий следующие события: A = {ровно попадание}; B = {хотя бы одно попадание}; C = {хотя бы один промах}; D = {не меньше двух попаданий}; E = {попадание не раньше, чем при третьем выстреле}.

Ответ: а)

$$\Omega = \{\omega_1, \omega_2, ..., \omega_8\} = \{\bar{A}_1\bar{A}_2\bar{A}_3, A_1\bar{A}_2\bar{A}_3, \bar{A}_1A_2\bar{A}_3, \bar{A}_1\bar{A}_2A_3, A_1A_2\bar{A}_3, A_1\bar{A}_2A_3, \bar{A}_1A_2A_3, A_1A_2A_3\}$$

б) $A = \omega_2 + \omega_3 + \omega_4 = A_1\bar{A}_2\bar{A}_3 + \bar{A}_1A_2\bar{A}_3 + \bar{A}_1\bar{A}_2A_3$,

$B = \omega_2 + ... + \omega_8 = \bar{\omega}_1 = \overline{\bar{A}_1\bar{A}_2\bar{A}_3} = A_1 + A_2 + A_3$,

$C = \omega_1 + ... + \omega_7 = \overline{\omega_8} = \overline{A_1A_2A_3} = \bar{A}_1 + \bar{A}_2 + \bar{A}_3$,

$D = \omega_5 + \omega_6 + \omega_7 + \omega_8 = A_1A_2\bar{A}_3 + A_1\bar{A}_2A_3 + \bar{A}_1A_2A_3 + A_1A_2A_3$,

$E = \omega_1 + \omega_4 = \bar{A}_1\bar{A}_2\bar{A}_3 + \bar{A}_1\bar{A}_2A_3 = \bar{A}_1\bar{A}_2$.

1.6. События: A - хотя бы один из трех проверяемых приборов бракованный, B - все приборы доброкачественные. Что означают события: а) A + B; б) AB; в) $\bar{A}$ и $\bar{B}$? Ответ: а) достоверное событие U; б) невозможное событие V.

1.7. Событие A – хотя бы одно из имеющихся четырех изделий бракованное, событие B – бракованных изделий среди них не менее двух. Что означают противоположные события $\bar{A}$ и $\bar{B}$?

Ответ: $\bar{A}$ - все изделия доброкачественные; $\bar{B}$ - бракованных изделий одно или нет ни одного.

1.8. Доказать, что $\bar{A}B + A\bar{B} + \overline{AB} = \overline{AB}$.

Ответ: Воспользоваться равенствами $\overline{A} = \overline{A}B + \overline{AB}, \overline{B} = A\overline{B} + \overline{AB}$.

1.9. Машинно-котельная установка состоит из двух котлов и одной машины. Событие А – исправна машина, событие B_k (k=1,2) – исправен k-й котел. Событие С означает работоспособность машинно-котельной установки, что будет в том случае, если исправны машина и хотя бы один котел. Выразить событие С и $\overline{C}$ через А и B_k. Ответ: $C = A(B_1 + B_2)$, $\overline{C} = \overline{A} + \overline{B}_1\overline{B}_2$.

1.10. Найти случайное событие X из равенства $\overline{X+A}+\overline{X+\overline{A}}=B$. Ответ: $X = \overline{B}$.

1.3 Аксиоматическое определение вероятности события

Для количественного описания степени объективной возможности наступления того или иного события вводится специальная числовая функция Р(А)*, называемая* вероятностью события А.

Пусть F - поле событий для данного эксперимента. *Вероятностью Р(А)* называется числовая функция, определенная для всех $A \subset F$ и удовлетворяющая трем *аксиомам вероятностей*:

1 .$P(A) \geq 0$;

2. $P(\Omega)=1$;

3. Для любой конечной или бесконечной совокупности наблюдаемых событий $A_1, A_2, \ldots, A_n, \ldots$, таких, что $A_iA_j=\varnothing$ при $i \neq j$, ($P(\sum\limits_k A_k) = \sum\limits_k P(A_k)$).

Смысл этих аксиом: *вероятность есть неотрицательная, нормированная и аддитивная функция множеств, принадлежащих алгебре событии F*.

Для любого случайного эксперимента можно построить *вероятностное пространство тройку* $\{\Omega, F, P\}$, где F - алгебра подмножеств множества Ω - множества элементарных исходов, P - числовая функция, удовлетворяющая аксиомам вероятностей.

В дальнейшем рассмотрим различные определения вероятности, удовлетворяющие этим аксиомам (классическое, геометрическое и статистическое (частотное)).

1.3.1. Классическое определение вероятности

Предположим, что результатом некоторого испытания может быть *конечное число n несовместных и равновозможных* исходов (классическая, урновая модель) и, что в m исходах, называемых благоприятными, осуществляется событие A.

Тогда *вероятность события А равна отношению числа благоприятных исходов к общему числу исходов испытания:* $P(A) = \dfrac{m}{n}$.

Вероятность осуществления события *A* или вероятность события *A* обозначим *P(A)*. Вероятность **достоверного** события равна **единице**, **невозможного** события - **нулю**. **Для случайного события 0≤*P(A)*≤1.**

Пример. В урне находятся 10 шаров, три из которых белых. Найти вероятность, что наудачу вытащенный шар - белый.

Решение. Пусть случайное событие А - наудачу вынутый из урны шар - белый. Тогда Р(А)=$\frac{3}{10}$, т.к. всего различных исходов 10 (n=10), а благоприятствуют всего три исхода (m=3).

Пример. Группа студентов из пяти человек садится в электричку, насчитывающую 7 вагонов. Предположим, что каждый студент выбирает свой вагон случайно и равновозможно оказывается в любом вагоне. Какова вероятность того, что все они попадут в разные вагоны?

Решение. Каждый студент может выбрать один из семи вагонов, так что всевозможных комбинаций ($i_1, i_2, \ldots, i_k$), где i_k - номер вагона, выбранный *k*- *м* студентом, насчитывается $n=7^5$. Случайное событие А - ни в какой вагон не попало больше одного из рассматриваемых студентов, наступает тогда и только тогда, когда все элементы $i_1, i_2, \ldots, i_5$ различны между собой. Можно считать, что совершается выбор (без возвращения) пяти элементов из семи Следовательно, число комбинаций $i_1, i_2, \ldots, i_5$ приводящих к наступлению события А, составляет $m = A_7^5 = 7 \cdot 6 \cdot 5 \cdot 4 \cdot 3 = 2520$ и искомая вероятность $P(A) = \frac{2570}{7^5} = \frac{360}{2401}$.

1.3.2. Геометрические вероятности

Геометрическое определение вероятности обобщает классическое определение вероятности на случай бесконечного числа исходов (на случай непрерывных пространств). Пусть Ω - некоторая область, имеющая меру $\mu(\Omega)$ (длину, площадь, объем и т.д.), такую, что $0 < \mu(\Omega) < \infty$. Скажем, что точка *равномерным образом* попадает в Ω, если вероятность Р(А) попадания ее в область А, являющуюся подобластью Ω, *пропорциональна мере этой области* $\mu(A)$.В силу аксиомы нормированности геометрическая вероятность события А определяется по формуле Р(А)=$\frac{\mu(A)}{\mu(\Omega)}$.

Пример. В окружность радиуса R вписан квадрат. В круг наудачу бросается точка. Какова вероятность того, что эта точка попадет в квадрат?

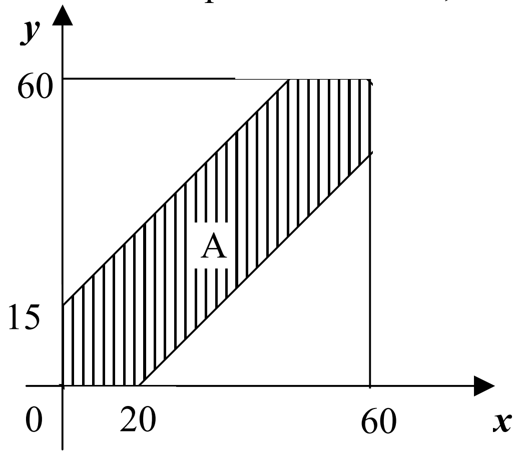


Рис. 1

Решение. Пусть случайное событие A - наудачу брошенная в круг точка попадет в квадрат. Тогда вероятность события A равна отношению площади квадрата к площади круга, т.е. $P(A)=\dfrac{(R\sqrt{2})^2}{\pi R^2}=\dfrac{2}{\pi}$.

Пример. Два друга договорились о встрече в определенном месте в промежутке времени между 7-ю и 8-ю часами. Найти вероятность того, что встреча состоится, если время ожидания первым лицом равно 15 минутам (после чего он уходит, не дождавшись друга), а вторым лицом - 20 минутам.

Решение. Событие A-встреча состоится. Пусть время прихода первым x (минут после семи часов), а вторым - y. Очевидно, что $0 \le y \le 60$ и $0 \le y \le 60$, т.е. время прихода друзей к месту встречи есть точка, принадлежащая квадрату с площадью S=3600. Множество точек, благоприятствующих событию A, удовлетворяют системе: $y - x \le 15$ и $x - y \le 20$. Площадь этой области *S(A)* = 3600 - 0,5(2025 + 1600) = 1787,5. Вероятность встречи *P(A)=S(A)/S*=1787,5/3600 =0,496.

К недостаткам классического и геометрического определений вероятности следует отнести необходимость обоснования равновозможности исходов испытания (опыта, эксперимента).

1.3.3. Условные вероятности. Независимость событий

Пусть A и B наблюдаемые события в эксперименте.

Условной вероятностью $P(B|A)$ осуществления события B, при условии, что событие A произошло в результате данного эксперимента, называется величина, определяемая равенством
$$P(B|A)=\frac{P(AB)}{P(A)},$$
$$P(A)\neq 0$$

Нетрудно проверить, что в случае произвольного вероятностного пространства вероятность $P(B|A)$, рассматриваемая как функция наблюдаемых событий $B \in F$ при фиксированном событии A, удовлетворяет трем основным аксиомам и всем их следствиям.

Событие A называется *независимым от события B* если выполняется равенство $P(A|B)=P(A)$ или события A и B называются независимыми, если $P(AB)=P(A)P(B)$.

Во многих задачах условную вероятность можно подсчитать непосредственно из условия.

Пример. Из урны содержащей 3 белых и 7 черных шаров, наудачу последовательно и без возвращений извлекаются 2 шара. События A={первый шар белый}, B={второй шар белый}. Определить вероятности P(A) и P(B/A).

*Решение.*По формуле классической вероятности P(A)=3/10. После того как извлекли первый белый шар, в урне осталось 9 шаров, из них 2-белых, и в этом случае P(B/A)=2/9.

Пример. Из колоды карт, содержащей 36 карт, вынули карту. Пусть событие A- вынули даму, событие B- вынутая карта пиковой масти. Выяснить являются ли события A и B зависимыми.

Решение. Т.к. $P(A)=\frac{1}{9}$ и $PB=\frac{1}{4}$, то $P(A)P(B)=\frac{1}{36}$. Событие AB- вынутая карта - дама пик и $P(AB)=\frac{1}{36}$. Мы получили, что выполняется теорема умножения для независимых событий ($P(AB) = P(A)P(B) = \frac{1}{36}$) , следовательно события A и B - независимы.

Пример. Из урны содержащей 3 белых и 7 черных шаров, наудачу последовательно и без возвращений извлекаются 2 шара. События A={первый шар белый}, B={второй шар белый}, C={хотя бы один шар белый}. Вычислить вероятности $P(B|A)$, $P(A|B)$ и $P(A|C)$. Установить, являются ли события A и B; A и C; B и C и события A, B и C в совокупности независимыми.

Решение. Занумеруем белые шары цифрами 1, 2, 3, а черные – цифрами 4, 5, …, 10. Согласно описанию эксперимента имеем следующую схему: выбор наудачу без возвращения пары чисел из множества {1, 2, …, 10} с упорядочиванием, поэтому множество элементарных исходов можно записать в виде $n = A_{10}^2 = 10*9 = 90$. Число исходов благоприятствующих событию A: $m_A=3\cdot9$, событию B – $m_B=9\cdot3$. $P(A)=\frac{3}{10}; P(B)=\frac{3}{10}$.

Событию AB благоприятствует $m(AB)=3\cdot2=6$ исходов, следовательно, $P(A)=\frac{1}{15}$. значит

$$P(B|A) = \frac{P(AB)}{P(A)} = \frac{2}{9}; P(A|B) = \frac{P(AB)}{P(B)} = \frac{2}{9};$$

$$P(C) = 1 - P(\overline{C}) = 1 - \frac{C_7^2}{C_{10}^2} = \frac{8}{15}.$$

$$AC = A(A+B) = A + AB = A; P(AC) = P(A) = \frac{3}{10} \Rightarrow P(A|C) = \frac{P(AC)}{P(C)} = \frac{9}{16}.$$

Итак, имеем

$$P(AB) = \frac{1}{15}; P(A)P(B) = \frac{9}{100} \neq P(AB),$$

т.е. события A и B не являются независимыми. Аналогично для других пар событий.

ЗАДАНИЯ

1.11. Игральная кость подбрасывается один раз. Найти вероятности следующих событий: A={число очков равно 6}; B={число очков кратно трём}; C={число очков четно}; D={число очков меньше пяти}; E={число очков больше двух}. Ответ: P(A)=1/6; P(B)=1/3; P(C)=1/2; P(D)=2/3; P(E)=2/3.

1.12. Подбрасываются две игральные кости. Найти вероятность указанных событий: A={числа очков на обеих костях совпадают}, B={число очков на первой кости больше, чем на второй}. Ответ: P(A)=1/6; P(B)=5/12.

1.13. Наудачу выбирается пятизначное число. Какова вероятность следующих событий: A={число одинаково читается как слева направо, так и справа налево (как, например, 13531)}, B={число кратно пяти}, C={число состоит из нечетных цифр}. Ответ: P(A)=0,01; P(B)=0,2; P(C)=5/144.

1.14. На шахматную доску случайным образом ставят две ладьи – белую и черную. Какова вероятность того, что ладьи не побьют друг друга? Ответ: 7/9.

1.15. У человека имеется N ключей, из которых только один подходит к его двери. Он последовательно испытывает их, выбирая случайным образом (без возвращения). Найдите вероятность того, что этот процесс закончится на k-м испытании $k \leq N$. Ответ: $P= 1/N$.

1.16. Из десяти первых букв русского алфавита выбирают наудачу без возвращения четыре буквы и записывают в порядке поступления слева направо. .Какова вероятность того, что составленное „слово" будет оканчиваться на букву „*А*"? Ответ: $P= 1/10$.

1.17. Из шести карточек с буквами „***Л***", „***И***", „***Т***", „***Е***", „***Р***", „***А***" выбирают наугад в определенном порядке четыре. Найдите вероятность того, что при этом получится слово „***ТИРЕ***". Ответ: $P = A_6^4 \approx 0{,}0028$.

1.18 Набирая номер телефона, абонент забыл две последние цифры и, помня лишь то, что эти цифры различны, набрал их наугад. Определите вероятность того, что набраны нужные цифры. Ответ: $P = 1 / A_{10}^2 \approx 0.011$.

1.19 При наборе телефонного номера абонент забыл две последние цифры и набрал их наудачу, помня только, что эти цифры нечетные и разные. Найдите вероятность того, что номер набран правильно. Ответ: $P = 1/A_5^2 = 0{,}05$.

1.20 Среди 25 экзаменационных билетов пять „хороших". Три студента по очереди берут по одному билету. Найдите вероятности следующих событий: A – третий студент взял „хороший" билет; B – все три студента взяли „хороший" билет. Ответ: $P(A) = 5/25$; $P(B) = 1/230 \approx 0{,}0044$.

1.21 В урне пять белых и четыре черных шара. Из урны в случайном порядке извлекают все находящиеся в ней шары. Найдите вероятность того, что вторым по порядку будет вынут белый шар. Ответ: Р= 5/9 ≈ 0,56.

1.22 Кодовые комбинации содержат пять различных цифр от 1 до 5. Какова вероятность того, что цифры в случайным образом выбранной кодовой комбинации образуют последовательность 1,2,3,4,5? Ответ: Р= 1/5! ≈ 0,0083.

1.23 Из урны, содержащей 10 перенумерованных шаров, наугад выбирают один за другим все находящиеся в ней шары. Найдите вероятность того, что все номера вынутых шаров будут идти по порядку. Ответ: Р =1/10! ≈ $2{,}8\cdot10^{-7}$.

1.24 Из урны, содержащей шары с номерами 1,2,...,9, пять раз наугад вынимают шар и каждый раз возвращают обратно. Найдите вероятность того, что из номеров шаров можно составить возрастающую последовательность. Ответ: $P = C_9^5 / 9^5 \approx 0{,}0021$.

1.25 В лифт семиэтажного дома на первом этаже вошли три человека. Каждый из них случайным образом может выйти на любом из этажей, начиная со второго. Найдите вероятности следующих событий: А – все пассажиры выйдут на четвертом этаже; В – все пассажиры выйдут на одном и том же этаже; С – все пассажиры выйдут на разных этажах. Ответ: Р(А) = =$1/6^3 \approx 0{,}0046$; Р(В) = $6/6^3 \approx 0{,}028$; $P(C) = A_6^3 / 6^3 \approx 0{,}56$.

1.26 Какова вероятность того, что в группе из n ($n \le 365$) случайно отобранных студентов хотя бы у двоих окажется один и тот же день рождения? Ответ: $P = 1 - A_{365}^n / 365^n$.

1.27 Найдите вероятность того, что дни рождения 12 случайным образом выбранных человек придутся на разные месяцы года. Ответ: $P = 12!/12^{12} \approx 5{,}4\cdot10^{-6}$.

1.28 Десять студентов договорились о поездке за город, но не договорились о вагоне. Любой из студентов наугад может сесть в любой из десяти вагонов поезда. Какова вероятность того, что они все попадут в разные вагоны? Ответ: $P=10!/10^{10} \approx 0{,}00036$.

1.29 Колоду из 52 карт случайным образом делят пополам. Найдите вероятность того, что в каждой половине будет по два „туза". Ответ: $P = C_4^2 C_{48}^{24} / C_{52}^{26} \approx 0{,}39$.

1.30 Некто купил карточку „Спортлото 6 из 49" и отметил в ней шесть из имеющихся 49 номеров. В тираже разыгрываются шесть „выигрышных" номеров. Найдите вероятности следующих событий: A_3 – угадано три номера; A_4 – угадано четыре номера; A_5 – угадано пять номеров; A_6 – угадано шесть номеров. Ответ: $P(A_3) = C_6^3 C_{43}^3 / C_{49}^6 \approx 0{,}018$;.

$P(A_4) = C_6^4 C_{43}^2 / C_{49}^6 \approx 0{,}00097$; $P(A_5) = C_6^5 C_{43}^1 / C_{49}^6 \approx 1{,}8 \cdot 10^{-5}$;

$P(A_6) = C_6^6 C_{43}^0 / C_{49}^6 \approx 7{,}2 \cdot 10^{-8}$.

1.31 На плоскость с нанесенной сеткой квадратов со стороной a наудачу брошена монета радиуса r<a/2. Найти вероятность того, что монета не пересечет ни одной из сторон квадрата. Предполагается, что вероятность падения точки в плоскую фигуру пропорциональна площади фигуры и не зависит от её расположения. Ответ: $P=(a-2r)^2/a^2$.

1.32 На отрезке OA длины L числовой оси Ox наудачу поставлены две точки: B(x) И C(y). Найти вероятность того, что из трех получившихся отрезков можно построить треугольник. Ответ: P=1/4.

1.33 Два приятеля условились встретиться в определенном месте между двумя и тремя часами дня. Пришедший первым ждет другого в течение 10 мин, после чего уходит. Чему равна вероятность встречи приятелей, если приход каждого из них в течение указанного часа может произойти в любое время? Ответ: $P = 11/36 \approx 0{,}31$.

1.34 Два парохода должны подойти к одному и тому же причалу. Время прихода обоих пароходов независимо и равновозможно в течение данных суток. Определить вероятность того, что одному из пароходов придется ожидать освобождения причала, если время стоянки первого парохода – один час, а второго – два часа. Ответ: $P = 139/1152$.

1.35 Значение a и b равновозможны в квадрате $|a| \le 1, |b| \le 1$. Найти вероятности следующих событий: A={корни квадратного трехчлена $x^2+2ax+b$ действительны}, B={корни квадратного трехчлена $x^2+2ax+b$ положительны}. Ответ: $P(A)=2/3$, $p(B)=1/12$.

1.36 Наудачу взяты два положительных числа x и y, каждое из которых не превышает двух. Найти вероятность того, что произведение xy будет не больше единицы, а частное y/x - не больше двух.
Ответ: $P=(1+3\ln 2)/8 \approx 0{,}38$.

1.37 Один раз подбрасывается игральная кость. События: A={выпало простое число очков}, B={выпало четное число очков}. Вычислить вероятность P(A/B). Ответ: P=1/3.

1.38 Вероятность попасть в самолет равна 0,4, а вероятность его сбить равна 0,1. Найти вероятности того, что при попадании в самолет он будет сбит.
Ответ: P=1/4.

1.39 Подбрасывают наудачу три игральные кости. Наблюдаемые события: A – на трех костях выпадут разные числа очков, B – хотя бы на одной из костей выпадет „шестерка". Вычислите P(A|B) и P(B|A).
Ответ: $P(A|B) = 60/91$; $P(B|A) = 1/2$.

1.40 Из колоды в 36 карт наудачу извлекается одна карта. События: A={вынутая карта - туз}, B={вынутая карта черной масти}, F={вынутая карта – фигура, т.е. является валетом, дамой, королем или тузом}. Установить, зависимы или независимы следующие три пары событий: A и B, A и F, F и B.

Ответ: A и B, F и B независимы, A и F зависимы.

1.4 Вероятности сложных событий

Сложным событием называется наблюдаемое событие, выраженное через другие наблюдаемые в том же эксперименте события с помощью допустимых алгебраических операций.

Вероятность осуществления того или иного сложного события вычисляется по правилам, основу которых составляют:
формула умножения вероятностей $P(AB)=P(B)\,P(A|B)= P(A)\,P(B|A)$,
и *формула сложения вероятностей* совместных событий

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

Пример. В колоде 36 карт. Наудачу извлечена карта. Найти вероятность, что извлекли даму или туза.

Решение. Пусть случайные события: A - извлеченная карта дама, B - извлеченная карта туз, C - извлеченная карта дама или туз. C=A+B. Тогда P(C)=P(A)+P(B), т.к. события A и B несовместные. P(A)=P(B)=4/36=1/9 Следовательно P(C)= 1/9+1 /9=2/9.

Пример. Два стрелка независимо друг от друга стреляют по мишени. Найти вероятность попадания в мишень двумя стрелками, если вероятности попадания для них соответственно равны 0,8 и 0,9.

Решение. Пусть случайные события: A - первый стрелок попал в мишень, B - второй попал, C - оба попали. Очевидно, что C=AB, где события A и B независимы по условию. Тогда $P(C)=P(AB)=P(A)P(B)=0,8\cdot 0,9=0,72$.

Пример. Два стрелка независимо друг от друга стреляют по мишени. Найти вероятность попадания в мишень хотя бы одним стрелком, если вероятности попадания для них соответственно равны 0,8 и 0,9.

Решение. Пусть случайное событие D - хотя бы один стрелок попал в мишень. Очевидно, что D=A+B и тогда P(D) = P(A+B) = P(A) + P(B)-P(AB)=0,8 + 0,9 - 0,72 = 0,98.

Пример. Определить вероятность того, что выбранное наудачу изделие является первосортным, если 4% всей продукции составляет брак, а 75% не бракованных изделий являются первосортными.

Решение. Пусть событие A - выбранное изделие не бракованное, событие B - выбранное изделие первосортное, событие C - наудачу выбранное изделие является первосортным. C=AB и $P(C)=P(A)P(B|A)$. Т.к. P(A)=1 - 0,04=0,96 и $P(B|A)=0,75$, то $P(C)=0,96\cdot 0,75=0,72$.

Замечание. Во многих задачах нас будет интересовать вероятность события *«хотя бы один»*. В этом случае проще найти вероятность противоположного события «ни разу не наступит», а затем искомого события.

Пример. Три стрелка стреляют по мишени независимо друг от друга. Найти вероятность хотя бы одного попадания в мишень, если вероятность поражения мишени для стрелков соответственно равны 0,5; 0,8 и 0,6.

Решение. Введем следующие события: A - первый стрелок поразил мишень B – второй; C - третий и D - хотя бы один.

1-й способ. Т.к. D=A+B+C, имеем P(D)=P(A) + P(B) + P(C) - [P(AB) + P(AC) + P(BC)] + P(ABC) = 0,5 + 0,8 + 0,6 - (0,40 + 0,30 + 0,48)+0,24=0,96.

2-й способ. Событие $\overline{D}$ - нет попаданий в мишень. Тогда $\overline{D} = \overline{A}\ \overline{B}\ \overline{C}$ и вероятность $P(D)=P(A\ \overline{B}\ \overline{C})$, $P(\overline{D})=(1-0,5)(1-0,8)(1-0,6)-0,04$. Тогда P(D)=1- P(D) = 1 - 0,04 = 0,96.

ЗАДАНИЯ

1.41 На карточках разрезной азбуки написаны 32 буквы русского алфавита. Пять карточек вынимают наудачу одна за другой и укладывают на стол в порядке появления. Найдите вероятность того, что получится слово „КОНЕЦ".

Ответ: $P = 1/32 \cdot 1/31 \cdot 1/30 \cdot 1/29 \cdot 1/28 \approx 4,14 \cdot 10^{-8}$.

1.42 Два стрелка стреляют в цель, причем каждый делает по одному выстрелу. Для первого стрелка вероятность попадания в цель равна 0,7, для второго – 0,8. Какова вероятность попасть в цель хотя бы одному стрелку?

Ответ: *P = 0,94*.

1.43 Система состоит из четырех узлов (см. рис.2). Вероятности P_i безотказной работы узлов равны соответственно P_1, P_2, P_3 и P_4. Вычислите вероятность безотказной работы всей системы.

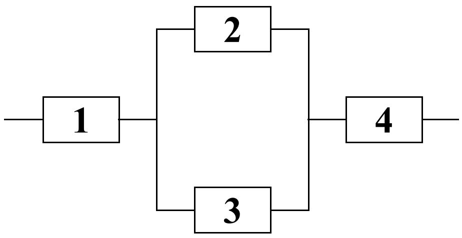


Рис.2

Ответ: $P=P_1[1-(1-P_2)(1-P_3)]P_4$.

1.44 Для сигнализации об аварии установлены два независимо работающих сигнализатора. Вероятность того, что при аварии сигнализатор сработает, равна 0,95 для первого сигнализатора и 0,9 для второго. Найти вероятность того, что при аварии сработает только один сигнализатор. Ответ: P=0,14.

1.45 Вероятность одного попадания в цель при одном залпе из двух орудий равна 0,38. Найти вероятность поражения цели при одном выстреле первым из орудий, если известно, что для второго орудия эта вероятность равна 0,8.

Ответ: P=0,7.

1.46 Устройство состоит из трех элементов, работающих независимо. Вероятности безотказной работы (за время t) первого, второго и третьего элементов соответственно равны 0,6; 0,7; 0,8. Найти вероятности того, что за время t безотказно будут работать: а) только один элемент; б) только два элемента; в) все три элемента. Ответ: а) P=0,188; б) P=0,0452; в) P=0,336.

1.47 Вероятность попадания в мишень стрелком при одном выстреле равна 0,8. Сколько выстрелов должен произвести стрелок, чтобы с вероятностью, меньшей 0,4, можно было ожидать, что не будет ни одного промаха? Ответ: n≥5.

1.48 В ящике 10 деталей, среди которых шесть окрашенных. Сборщик наудачу извлекает четыре детали. Найти вероятность того, что все извлеченные детали окажутся окрашенными. Ответ: P=1/14.

1.49 Студент знает 20 из 25 вопросов и программы. Найти вероятность того, что студент знает предложенные ему экзаменатором три вопроса. Ответ: P=57/115

1.50 В мешочке содержаться 10 одинаковых кубиков с номерами от 1 до 10. Наудачу извлекаются по одному три кубика. Найти вероятность того, что последовательно появятся кубики с номерами 1, 2, 3, если кубики извлекаются: а) без возвращения; б) с возвращением (извлеченный кубик возвращается в мешочек). Ответ: P=1/720.

1.5. Формула полной вероятности. Формула Байеса

Пусть событие A может произойти только совместно с одним из событий $H_1, H_2, \dots, H_n$, образующих полную группу несовместных событий ($H_iH_j=\varnothing$ $(i\neq j)$; $\sum_{i=1}^{n} H_i = \Omega$), которые часто называют гипотезами. Т.е. система множеств $H_1, H_2, \dots, H_n$ образует разбиение множества Ω.

Имеет место *формула полной вероятности*: $P(A) = \sum_{i=1}^{n} P(H_i)P(A/H_i)$. Безусловные вероятности P(H_i) трактуются как *доопытные (априорные)* вероятности гипотез.

Пусть событие A осуществилось. Какова *послеопытная (апостериорная)* вероятность гипотез $H_k (k = 1,2,\dots,n)$ при условии, что событие A имело место?

Ответ дается *формулой Байеса* $P(H_k / A) = \dfrac{P(H_k)P(A/H_k)}{P(A)}$, где $P(A) = \sum_{i=1}^{n} P(H_i)P(A/H_i)$ - полная вероятность события A.

Формула Байеса позволяет «переоценить» вероятность каждой из гипотез после поступления новой информации относительно наблюдаемых событий.

Пример. Литье в болванках поступает из двух подготовительных цехов: 70% из первого и 30% из второго. При этом заготовки первого цеха имеют 10% брака, а второго - 20%. Найти вероятность того, что взятая наугад болванка имеет брак и что эта бракованная болванка изготовлена в первом цехе.

Решение. Случайное событие A - взятая наугад болванка бракованная, гипотезы: H_1 - взятая наугад болванка из первого цеха, H_2 - из второго. По формуле полной вероятности $P(A)=P(H_1)P(A/H_1) + P(H_2)P(A/H_2)$, где $P(H_1)=0{,}7$ и $P(H_2)=0{,}3$, а $P(A/H_1)=0{,}1$ и $P(A/H_2)=0{,}2$, получим $P(A)=0{,}7\cdot 0{,}1+0{,}3\cdot 0{,}2=0{,}13$; вероятность, что наудачу взятая бракованная болванка из первого цеха $P(H_1/A)$ найдем по формуле Байеса: $P(H_1 / A) = \dfrac{P(H_1)P(A/H_1)}{P(A)} = \dfrac{0{,}7\cdot 0{,}1}{0{,}13} = \dfrac{7}{13}$.

Замечание. При составлении гипотез, которые поясняют все возможные ситуации для наступления интересующего нас события, нужно убедиться, что они *несовместны и образуют полную группу*, т.е. $\sum_{i=1}^{n} P(H_i) = 1$ и $H_iH_j=\varnothing$ ($i\neq j$).

ЗАДАНИЯ

1.51 В первой урне лежат 10 шаров, из них восемь белых, во второй – 20 шаров, из них четыре белых. Из каждой урны наудачу извлекают по одному шару, а затем из этих двух наудачу берется один шар. Найдите вероятность того, что это будет белый шар. Ответ: P = 0,5.

1.52 На заводе, изготавливающем болты, первый станок производит 25%, второй 35% и третий 40% всех изделий. В их продукции брак составляет 5%, 4% и 2% соответственно.

а) Какова вероятность того, что случайно выбранный болт будет дефектным?

б) Случайно выбранный болт оказался дефектным. Найдите вероятности P_1, P_2 и P_3 того, что он был произведен первым, вторым, третьим станком?

Ответ: а) P = 0,0345, б) P_1 = 125/345≈0,36; P_2 = 140/345≈0,406; P_3 = 80/345≈0,23.

1.53 Два стрелка стреляют по одной мишени, делая каждый по одному выстрелу. Вероятность попадания в мишень для первого стрелка равна 0,8, для второго – 0,4. После стрельбы в мишени обнаружена одна пробоина. Определите вероятность того, что в цель попал первый стрелок.

Ответ: P = 6/7 ≈ 0,857.

1.54 В продажу поступают телевизоры трех заводов. Продукция первого завода содержит 20% телевизоров со скрытым дефектом, второго – 10% и третьего – 5%. Какова вероятность приобрести исправный телевизор, если в магазин поступило 30% телевизоров с первого завода, 20% - со второго и 50% - с третьего? Ответ: P=0,895.

1.55 В ящике лежат 20 теннисных мячей, в том числе 15 новых и 5 играных. Для игры наудачу выбирается два мяча, и после игры возвращаются обратно. Затем для второй игры также наудачу извлекаются еще два мяча. Какова вероятность того, что вторая игра будет проводиться новыми мячами?

Ответ: P=0,445.

1.56 Три стрелка, вероятности попадания которых при одном выстреле в мишень в неименных условиях постоянны и соответственно равны p_1=0,8; p_2=0,7; p_3=0,6 делают по одному выстрелу в одну и ту же мишень. Вычислить вероятность события A={в мишени окажется ровно две пробоины}, приняв в качестве гипотез элементарные исходы данного эксперимента.
Ответ: P=0,0565.

1.57 Число грузовых автомашин, проезжающих по шоссе, на котором стоит бензоколонка, относится к числу легковых машин, проезжающих по тому же шоссе как 3:2/ Вероятность того, что будет заправляться грузовая машина, равна 0,1; для легковой машины эта вероятность равна 0,2. К бензоколонке для заправки подъехала машина. Найти вероятность того, что это грузовая машина. Ответ: P=3/7.

1.58 В специализированную больницу поступают в среднем 50% больных заболевание K, 30% - с заболеванием L, 20% - с заболеванием M/ Вероятность полного излечения болезни K равна 0,7; для болезней L и M эти вероятности соответственно равны 0,8 и 0,9. Больной, поступивший в больницу, был выписан здоровым. Найти вероятность того, что этот больной страдал заболеванием K Ответ: P=5/11.

1.59 Батарея из трех орудий произвела залп, причем два снаряда попали в цель. Найти вероятность того, что первое орудие дало попадание, если вероятность попадания в цель первым, вторым и третьим орудиями соответственно равны p_1=0,4; p_2=0,3; p_3=0,5. Ответ: $P(B_1|A)$=20/29.

1.6. Схема независимых испытаний. Формула Бернулли. Предельные теоремы Муавра - Лапласа и Пуассона

Пусть производится n независимых опытов (испытаний), в каждом из которых вероятность наступления события A одинакова и равна p. Требуется определить вероятность того, что событие A появится ровно k раз. Эта вероятность определяется формулой Бернулли:

$$P_n(k) = C_n^k p^k (1-p)^{n-k} \ (\mathrm{k} = 0,1,2,...,\mathrm{n}).$$

Пример. Баскетболист бросает пять раз по кольцу. Вероятность попадания при каждом броске равна 0,9. Найти вероятности того, что: 1) он попадет ровно три раза; 2) он попадет не менее трех раз; 3) он попадет хотя бы один раз.

Решение. Введем следующие события: A - баскетболист попадет ровно 3 раза; B - попадет не менее трех раз; C - попадет хотя бы один раз Тогда:

$$\mathrm{P(A)} = P_5(3) = C_5^3 0{,}9^3(1-0{,}9)^2 = \frac{5\cdot 4\cdot 3}{3\cdot 2\cdot 1}\cdot 0{,}729\cdot 0{,}01 = 0{,}0729;$$

$$\mathrm{P(B)} = P_5(3)+P_5(4)+P_5(5) = C_5^3 0{,}9^3(1-0{,}9)^2 + C_5^4 0{,}9^4(1-0{,}9)^1 + 0{,}9^5 \cong 0{,}991;$$

$P(C)=1-P(\bar{C})$; $P(\bar{C})=P_5(0)=(1-0{,}9)^5=0{,}00001$ и P(C)=0,99999.

При больших n пользоваться формулой Бернулли становится неудобно Справедливы следующие асимптотические теоремы.

Локальная теорема Муавра-Лапласа: $P_n(k) \cong \frac{1}{\sqrt{np(1-p)}}\varphi(x)$, где $x = \frac{k-np}{\sqrt{np(1-p)}}$ и $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$. Значения функции *φ(x)* даны в табл. 1 (*φ(x)* - четная).

Интегральная теорема Муавра – Лапласа: $P_n(k_1 \le k \le k_2)=\Phi(x_2)-\Phi(x_1)$, где $x_1 = \frac{k-np}{\sqrt{np(1-p)}}$ (i=1,2), а $\Phi(x) = \frac{1}{2\pi}\int_0^x e^{-\frac{x^2}{2}}dt$ - функция Лапласа (она нечетная). Значения функции $\Phi(x)$ даны в табл. 2.

Теорема Пуассона. Если $np=\lambda$, а $\lambda \le 10$, то справедлива *формула Пуассона*: $P_n(k) \cong \frac{\lambda^k}{k!}e^{-\lambda}$.

Пример. Производится 400 независимых испытаний, в каждом из которых событие А может произойти с вероятностью p=0,2. Определить вероятность того, что: событие А наступит ровно 80 раз; событие А наступит от 60 до 96 раз включительно.

Решение. Введем события: В - событие А наступит ровно 80 раз; С событие А наступит от 60 до 96 раз. Имеем, что Р(В)=$P_{400}(80) = C_{400}^{80}0{,}2^{80}0{,}8^{320} \cong \frac{1}{\sqrt{400\cdot 0{,}2(1-0{,}2)}}\varphi(x)$, где $x = \frac{80-400\cdot 0{,}2}{\sqrt{400\cdot 0{,}2\cdot 0{,}8}} = 0$ и по таблице $\varphi(0)$=0,3989, тогда Р(В)=0,3989/8=0,0499. Р(С)=$P_{400}(60 \le k \le 96)$=$\Phi(x_1)$-$\Phi(x_2)$, где $x_1 = \frac{60-400\cdot 0{,}2}{\sqrt{400\cdot 0{,}2\cdot 0{,}8}} = -2{,}5$ и $x_2 = \frac{96-400\cdot 0{,}2}{\sqrt{400\cdot 0{,}2\cdot 0{,}8}} = 2$, по таблице функции Лапласа находим Ф(2)=0,4772 и Ф(-2,5)=-Ф(2,5)=-0,4938. Окончательно Р(С)=0,9700.

Пример. Если брак составляет 1% от всей продукции, то какова вероятность того, что среди 200 изделий окажется: ровно 4 бракованных изделия; не менее четырех изделий?

Решение. В данной задаче n=200, а np=200·0,01=2. Применим для расчетов формулу Пуассона. Вероятность, что среди 200-х изделий ровно 4 бракованных изделия $P_{200}(4) \cong \frac{2^4}{4!}e^{-2} \cong 0{,}09$.

Для определения вероятности, что среди 200-х изделий не менее 4-х бракованных, найдем вероятность противоположного события - среди 200-х изделий меньше 4-х бракованных.

$$P_{200}(0 \le k \le 4) = P_{200}(0) + P_{200}(1) + P_{200}(2) + P_{200}(3) \cong \frac{2^0}{0!}e^{-2} + \frac{2^1}{1!}e^{-2} + \frac{2^2}{2!}e^{-2} + \frac{2^3}{3!}e^{-2} \cong .0{,}85$$

Тогда искомая вероятность $P(k \ge 4) = 1 - 0{,}85 = 0{,}15$.

Пример. Вероятность наступления события А в каждом испытании равна p=0,2. Проведено 1600 (n=1600) испытаний. Найти:

1) вероятность того, что частота наступления события А отличается от р по абсолютной величине на 0,02;

2) в каких границах находится относительная частота, если вероятность отклонения (по абсолютной величине) относительной частоты от вероятности *p* равна 0,985?

3) сколько необходимо произвести испытаний, чтобы абсолютное отклонение относительной частоты от вероятности, не превосходящее 0,01, было равно 0,995?

Решение. 1) *Для оценки. отклонения относительной частоты от вероятности при независимых испытаниях применяют формулу*

$$P(\left|\frac{m}{n}-p\right| \le \varepsilon) = 2\Phi(\varepsilon\sqrt{\frac{n}{p(1-p)}})\text{; тогда}$$

$$P(\left|\frac{m}{n}-0{,}2\right| \le 0{,}02) \cong 2\Phi(0{,}02\sqrt{\frac{1600}{0{,}2(1-0{,}2)}}) = 2\Phi(2) = 0{,}9544;$$

2) нам нужно определить ε, если известно, что 2Φ(t)=0,985. По таблице функции Лапласа находим t=2,43, т.е. $\varepsilon\sqrt{\frac{1600}{0{,}2(1-0{,}2)}} = 2{,}43 \Rightarrow$ ε=0,0243.

3) по условию ε=0,01 и $P(\left|\frac{m}{n}-0{,}2\right| \le 0{,}01) \cong 2\Phi(0{,}01\sqrt{\frac{n}{0{,}2(1-0{,}2)}}) = 0{,}995 \Rightarrow$

$t = 0{,}01\sqrt{\frac{n}{0{,}2(1-0{,}2)}} = 2{,}8$ (по таблице функции Лапласа находим *t*=2,8). Определяем, что *n*=12544.

ЗАДАНИЯ

1.60 Бросают пять игральных костей. Вычислите вероятность того, что на трех из них выпадет пятерка. Ответ: $P = C_5^3(1/6)^3(5/6)^2 \approx 0{,}032$

1.61 Для стрелка, выполняющего упражнения в тире, вероятность попасть в «яблочко» при одном выстреле не зависит от результатов предшествующих выстрелов и равна p=1/4. Спортсмен сделал 5 выстрелов. Найти вероятности событий: A={хотя бы одно попадание}, B={ровно одно попадание}, C= {ровно два попадания}, D={не менее трех попаданий }.
Ответ: P(A)$\approx$0,7627, P(B)$\approx$0,3955, P(C)$\approx$0,2637,P(D)$\approx$0,1035.

1.62 Отрезок АВ разделен точкой С в отношении 2:1. На этот отрезок наудачу брошены четыре точки. Найти вероятность того , что две из них окажутся левее точки С и две правее. Ответ: Р=8/27.

1.63 Оптовая база снабжает 10 магазинов, от каждого из которых может поступить заявка на очередной день с вероятностью 0,4 независимо от заявок других магазинов. Найдите вероятность того, что в день поступит четыре заявки. Ответ: Р $\approx$ 0,251.

1.64 Событие В наступает в том случае, если событие А появится не менее трех раз. Определить вероятность появления события В, если вероятность появления события А при одном опыте равна 0.3 и произведено:

а) пять независимых опытов ; в) семь независимых опытов.

Ответ: а)Р= 0,163; б) Р= 0,353.

1.65 Вероятность того,что лампа останется исправной после 1000 часов работы, равна 0,2. Какова вероятность того, что хотя бы одна из трех ламп останется исправной после 1000 часов работы? Ответ: Р=0,992.

1.66 Найти вероятность того, что событие А наступит 1400 раз в 2400 испытаниях, если вероятность появления этого события в каждом испытании равна 0.6. Ответ: Р=0,0041.

1.67 Вероятность поражения мишени при одном выстреле равна 0,8.Найти вероятность того, что при 100 выстрелах мишень будет поражена ровно 75 раз. Ответ: Р=0,04565.

1.68 Вероятность появления события в каждом из 100 независимых испытаний постоянна и равна р=0,8. Найти вероятность того, что событие появится: а) не менее 75 раз и не более 90 раз; б) не менее 75 раз; в) не более 74 раз

Ответ: а) Р=0,8882; б) Р=0,8944; в) Р=0,1056.

1.69 Вероятность появления события в каждом из 21 независимых испытаний равна 0,7. Найти вероятность того, что событие появится в большинстве испытаний. Ответ: Р=0,95945.

1.70 Вероятность появления события в каждом из 2100 независимых испытаний равна 0,7. Найти вероятность того, чтособытие появится: а) не менее 1470 и не более 1500 раз; б) не менее 1470 раз; в) не более 1469 раз.

Ответ: а) Р=0,4236; б) Р= 0,5; в) Р=0,5.

1.71 Вероятность появления события в каждом из независимых испытаний равна 0.8. Сколько нужно произвести испытаний, чтобы с вероятностью 0.9 можно было ожидать, что событие появится не менее 75 раз? Ответ: n=100.

1.72 Известно, что на выпечку 1000 булочек с изюмом нужно израсходовать 10000 изюмин. Найдите вероятность того, что:

а) наудачу выбранная булочка не будет содержать изюма;

б) среди пяти выбранных наудачу булочек две не будут содержать изюма, а в остальных будет хотя бы по одной изюмине.

Ответ: а) $P = e^{-10} \approx 0{,}0000468$; б) $P \approx C_5^2 (e^{-10})^2 (1 - e^{-10})^3 \approx 2{,}19 \cdot 10^{-8}$.

1.73 Устройство состоит из 1000 элементов, работающих независимо один от другого. Вероятность отказа любого элемента в течении времени Т равна 0,002. Найти вероятность того, что за время Т откажут ровно три элемента.

Ответ: Р=0,18.

1.74 Завод отправил на базу 500 изделий. Вероятность повреждения в пути равна 0.002. Найти вероятности того, что в пути будет повреждено изделий: а) ровно три; б) менее трех; в)более трех; г) хотя бы одно.

Ответ: а) Р=0,0613; б) Р=0,9197; в) Р=0,019; г) Р=0,632.

1.75 Вероятность появления события в каждом из 10000 независимых испытаний равна 0,75. Найти вероятность того, что относительная частота появления события отклонится от его вероятности по абсолютной величине не более чем на 0,01. Ответ: Р=0,979.

1.76 Французский ученый Бюффон бросил монету 4040 раз, причем «герб» появился 2048 раз. Найти вероятность того, что при повторении опыта Бюффона относительная частота появления «герба» отклонится от вероятности появления «герба» не более чем в опыте Бюффона.

Ответ: Р=0,6196.

1.77 Симметричную монету подбрасывают 10000 раз. Найдите вероятность того, что наблюденная частота выпадения „герба" будет отличаться от 1/2 не более чем на 2%. Ответ: $P \approx \Phi_0(2) - \Phi_0(-2) = 0{,}9545$.

1.78 Вероятность появления события в каждом из независимых испытаний равна 0.5. Найти число испытаний n, при котором с вероятностью 0,7698 можно ожидать, что относительная частота появления события отклонится от его вероятности по абсолютной величине не более чем на 0,02.

Ответ: n=900.

Глава 2. Случайные величины

2.1. Дискретные и непрерывные случайные величины. Способы их задания

Случайной величиной X называется действительная функция $X=X(\varpi)$, определенная на множестве элементарных исходов Ω. Случайные величины принято обозначать большими буквами латинского алфавита, а их возможные значения - соответствующими малыми буквами. Для задания (описания) случайных величин используют закон (плотность) распределения и функцию (интегральную функцию) распределения случайной величины. Существуют дискретные и непрерывные случайные величины. Универсальной характеристикой, пригодной для описания дискретных и непрерывных случайных величин, является функция *распределения вероятностей $F(x)$*, определяющая вероятность того, что случайная величина X примет значение меньше некоторого числа x, т.е. $\boldsymbol{F(x)=P(X<x)}$.

Отметим основные свойства функции $F(x)$:

1. $0 \le F(x) \le 1$, $-\infty < x < \infty$;
2. $F(-\infty)=0$, $F(\infty)=1$;
3. Функция $F(x)$ непрерывна слева;
4. Функция $F(x)$ - неубывающая функция, т.е. $F(x_1) \le F(x_2)$ если $x_1 < x_2$.

Вероятность попадания случайной величины X на произвольный интервал равна $P(x_1 \le X < x_2) = F(x_2) - F(x_1)$.

Случайная величина X называется *дискретной случайной величиной* (Д.С.В), если множество ее возможных значений конечно или счетно.

Перечень всех возможных значений Д.С.В. и соответствующих этим значениям вероятностей называется законом распределения дискретной случайной величины. Закон распределения можно задать в аналитической форме: $P(X=x_k)=p_k$, где $p_k>0$ и $\sum\limits_k p_k=1$; в табличной - *ряд распределения*; графической – *многоугольник распределения*.

Зная закон распределения Д.С.В., можно найти функцию распределения $F(x)=\sum\limits_{x_i<x} P(x_i)$, где суммирование распространяется на все значения индекса i, для которых $x_1<x_2$. Из этой формулы, в частности, следует, что

$$F(x_k+0) - F(x_k) = P(X = x_k), \qquad (*)$$

т.е. функция распределения Д.С.В. испытывает разрывы в точках x, для которых существует положительная вероятность события $\{X=x_k\}$.

Для Д.С.В. функция распределения кусочно-постоянная. Зная функцию распределения Д.С.В., можно найти по формуле (*) закон распределения.

Случайная величина X называется непрерывной случайной величиной (Н.С.В.), если множество ее значении представляет собой интервал.

Н.С.В. задается *плотностью вероятности $f(x)$ (дифференциальным законом распределения).* Вероятностный смысл плотности вероятностей состоит в следующем: $f(x)\,\Delta x \cong P(x \le X < x + \Delta x)$ при $\Delta x \to 0$. Основные свойства функции $f(x)$:

1) $f(x) \ge 0$; $-\infty < x < \infty$; 2) $\int\limits_{-\infty}^{+\infty} f(x)\,dx = 1$ условие нормировки;

3) $P(x_1 \le X < x_2) = \int\limits_{x_1}^{x_2} f(x)\,dx$.

Связь между плотностью вероятностей и функцией распределения вероятностей (интегральной функцией) следующая:

$$f(x) = F'(x) \text{ и } F(x) = \int\limits_{-\infty}^{x} f(x)\,dx \quad .$$

Замечание. Для Н.С.В. функция распределения непрерывная. Это означает, что вероятность "попасть в точку" для Н.С.В. равна нулю.

2.2. Числовые характеристики случайных величин

Случайные величины, помимо законов распределения (полная характеристика), могут описываться *числовыми характеристиками*. Рассмотрим наиболее употребительные (основные).

Математическое ожидание (среднее значение по распределению) - это число, которое определяется формулой

$$M(x) = \sum_k x_k p_k \text{ для Д.С.В. и } M(x) = \int\limits_{-\infty}^{\infty} x f(x)\,dx \text{ для Н.С.В.}$$

Механический смысл математического ожидания - это центр тяжести.

Математическое ожидание (операция) обладает следующими свойствами:

1) математическое ожидание постоянной величины равно самой постоянной $M(C)=C$;

2) постоянный множитель можно выносить за знак математического ожидания $M(CX)=CM(X)$;

3) математическое ожидание суммы случайных величин равно сумме математических ожидании слагаемых $M(\sum\limits_{k=1}^{n} X_k) = \sum\limits_{k=1}^{n} M(X_k)$;

4) математическое ожидание произведения взаимно независимых случайных величин равно произведению математических ожиданий сомножителей $M(X_1 X_2 \ldots X_n) = M(X_1) M(X_2) \ldots M(X_n)$.

Основными характеристиками рассеяния возможных значений случайной величины около математического ожидания служат *дисперсия* и *среднее квадратическое отклонение.*

Дисперсией случайной величины X называют математическое ожидание квадрата отклонения случайной величины от ее математического ожидания:

$D(x) = M[X - M(X)]^2$.

Дисперсию удобно вычислять по формуле $D(x) = M(X^2) - [M(X)]^2$.

Приведем формулы вычисления дисперсии для Д.С.В. и Н.С.В.: $D(x)=\sum_k (x_k - M(x))^2 p_k$ или $D(x)=\sum_k x_k^2 p_k - [M(x)]^2$ для Д.С.В.;

$D(x)=\int_{-\infty}^{\infty}(x-M(x))^2 f(x)dx$ или $D(x)=\int_{-\infty}^{\infty}x^2 f(x)dx-[M(x)]^2$ для Н.С.В.

Дисперсия обладает следующими свойствами:

1) дисперсия постоянной величины равна нулю $D(C)=0$;
2) $D(CX)=C^2D(X)$;
3) дисперсия суммы независимых случайных величин равна сумме

дисперсий слагаемых $D(\sum_{k=1}^{n} X_k)=\sum_{k=1}^{n} D(X_k)$.

Так как дисперсия имеет размерность квадратичную, то вводится среднее квадратическое отклонение $\sigma(x)=\sqrt{D(x)}$.

Пример. Из десяти лотерейных билетов два выигрышных. Куплено два билета. Для случайного числа X выигрышных билетов среди купленных найти:

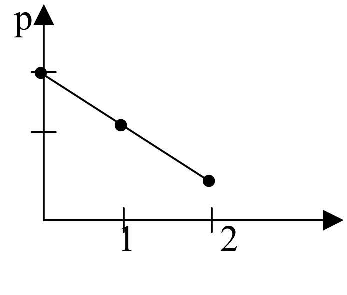
1) закон распределения;
2) функцию распределения;
3) основные числовые характеристики $M(x)$, $D(x)$ и $\sigma(x)$;
4) вероятность попадания случайной величины X в интервал (-1; 1,5).

Решение. 1) Возможные значения случайной величины X: 0, 1 и 2. Найдем соответствующие им вероятности (закон распределения): $P(X=0)=\frac{C_8^2}{C_{10}^2}=\frac{28}{45}$;

$P(X=1)=\frac{C_8^1 C_2^1}{C_{10}^2}=\frac{16}{45}$ и $P(X=2)=\frac{C_2^2}{C_{10}^2}=\frac{1}{45}$ (проверяем, что $P(X=0)+P(X=1)+P(X=2)=1$).

Ряд распределения

x_i	0	1	2
p_i	28/45	16/45	1/45



Многоугольник распределения

2) Построим функцию распределения:

если $x \le 0$, то $F(x)=P(X<x)=0$;

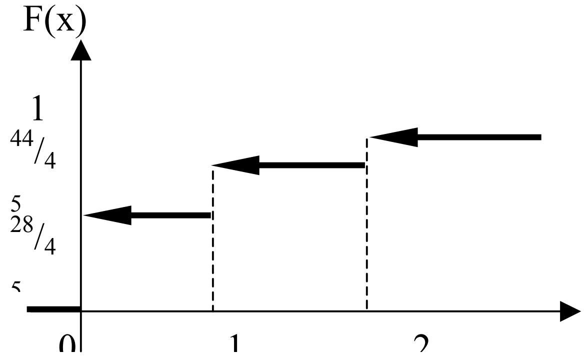
если $0 < x \le 1$, то $F(x)=P(X<x)=P(X=0)=28/45$;

если $1 < x \le 2$, то $F(x)=P(X<x)=P(X=0)+P(X=1)=28/45+16/45=44/45$;

если $x > 2$, то $F(x)=P(X<x)=P(X=0)+P(X=1)+P(X=2)=28/45+16/45+1/45=1$.

Функция распределения и ее график имеют вид

$$F(x)=\begin{cases}0\ \textit{при}\ x\le 0;\\ \dfrac{28}{45}\ \text{при}\ 0<x\le 1;\\ \dfrac{44}{45}\ \text{при}\ 1<x\le 2;\\ 1\ \text{при}\ x>2.\end{cases}$$



4) Вычислим математическое ожидание $M(x)=0\cdot\dfrac{28}{45}+1\cdot\dfrac{16}{45}+2\cdot\dfrac{1}{45}=\dfrac{2}{5}$;

найдем дисперсию по определению

$D(x)=\left(0-\dfrac{2}{5}\right)^2\cdot\dfrac{28}{45}+\left(1-\dfrac{2}{5}\right)^2\cdot\dfrac{16}{45}+\left(2-\dfrac{2}{5}\right)^2\cdot\dfrac{1}{45}=\dfrac{64}{225}$, (сравните с вычислением по другой формуле: $D(x)=0^2\cdot\dfrac{28}{45}+1^2\cdot\dfrac{16}{45}+2^2\cdot\dfrac{1}{45}-\left(\dfrac{2}{5}\right)^2=\dfrac{64}{225}$); вычислим среднее квадратическое отклонение $\sigma(x)=\sqrt{\dfrac{64}{225}}=\dfrac{8}{15}$.

4) Вероятность попадания случайной величины X в интервал (-1; 1,5) P(-1<X<1,5)=P(X=0)+P(X=1)=44/45 или P(-1<X<1,5)=F(1,5)-F(-1)=44/45.

Пример. Задана функция распределения $F(x)=\begin{cases}a\ \text{при}\ x\le 0;\\ bx^2+c\ \text{при}\ 0<x\le 3;\\ d\ \text{при}\ x>3.\end{cases}$

Найти:

1) неизвестные параметры;
2) плотность распределения вероятностей *f(x)*;
3) числовые характеристики - *M(x), D(x)* и $\sigma(x)$;
4) вероятность попасть в интервал (-1; 1,5).

Решение. 1) Для определения неизвестных параметров воспользуемся свойствами функции распределения. Так как. $F(-\infty)=0$, $F(\infty)=1$ $\Rightarrow$ $a=0$, $d=1$. Из свойства непрерывности функции распределения следует, что $\lim\limits_{x\to 0-0}F(x)=\lim\limits_{x\to 0+0}F(x)\Rightarrow \mathrm{a}=\mathrm{c}=0$ и $\lim\limits_{x\to 3-0}F(x)=\lim\limits_{x\to 3+0}F(x)\Rightarrow 9b=d=1\Rightarrow b=\dfrac{1}{9}$.

Следовательно, $F(x) = \begin{cases} 0 \text{ при } x \le 0; \\ \frac{1}{9}x^2 \text{ при } 0 < x \le 3; \\ 1 \text{ при } x > 3. \end{cases}$

2) Плотность распределения вероятностей

$$f(x) = F'(x) = \begin{cases} 0 \text{ при } x < 0; \\ \frac{2}{9}x \text{ при } 0 < x < 3; \\ 0 \text{ при } x > 3. \end{cases}$$

3) $M(x) = \int_{-\infty}^{+\infty} xf(x)dx = \int_{-\infty}^{0} x \cdot 0dx + \int_{0}^{3} x\frac{2}{9}xdx + \int_{3}^{+\infty} x \cdot 0dx = \frac{2}{27}x^3 |_0^3 = 2$;

$$D(x) = \int_{-\infty}^{+\infty} x^2 f(x)dx - [M(x)]^2 = \int_{-\infty}^{\infty} x^2 \cdot 0dx + \int_{0}^{3} x^2\frac{2}{9}xdx + \int_{3}^{+\infty} x^2 \cdot 0dx - 2^2 =$$

$$= \frac{x^4}{18}|_0^3 - 4 = \frac{1}{2}; \qquad \sigma(x) = \sqrt{D(x)} = \sqrt{0{,}5}.$$

4) Вероятность случайной величине X попасть в интервал (-1; 1,5)

$P(-1 < X < 1,5) = F(1,5) - F(-1) = 0,25 - 0 = 0,25$ или

$$P(-1 < X < 1,5) = \int_{-1}^{1,5} f(x)dx = \int_{-1}^{0} 0dx + \int_{0}^{1,5}\frac{2}{9}xdx = \frac{1}{4}.$$

Пример. Задана плотность распределения вероятностей $f(x) = \frac{a}{1+x^2}$.

Найти: 1) параметр a; 2) функцию распределения F(x); 3) вероятность попасть в интервал (-π/4; π/4).

Решение. 1) Для определения неизвестного параметра воспользуемся свойством плотности распределения вероятностей, что $\int_{-\infty}^{\infty} f(x)dx = 1$ (условие нормировки), $\int_{-\infty}^{\infty}\frac{a}{1+x^2}dx = 1 \Rightarrow a \cdot \text{arctg x}|_{-\infty}^{\infty} = a \cdot \pi = 1 \Rightarrow a = \frac{1}{\pi}$.

2) $F(x) = \int_{-\infty}^{x}\frac{a}{1+t^2}dt = a \cdot \text{arctg t}|_{-\infty}^{x} = \frac{1}{\pi}(\frac{\pi}{2} + arctg\text{ x}) = \frac{1}{2} + \frac{1}{\pi}arctg\text{ x}.$

3) Вероятность попасть в интервал (-π/4; π/4):

$P(-\pi/4 < x < -\pi/4) = \int_{-\pi/4}^{\pi/4}\frac{a}{1+x^2}dx = \frac{1}{\pi}\text{arctg x}|_{-\pi/4}^{\pi/4} = \frac{2}{\pi}arctg\frac{\pi}{4} \approx 0{,}4$ или

$P(-\pi/4 < x < -\pi/4) = F(\frac{\pi}{4}) - F(-\frac{\pi}{4}) = \frac{2}{\pi}arctg\frac{\pi}{4}.$

Начальным моментом m-го порядка (m =0,1.2....) *случайной величины X называется число* ν_m, *определяемое по формуле*

$$\nu_m = M(X^m) = \begin{cases} \sum\limits_k x_k^m p \text{ , если X - Д.С.В.,} \\ \int\limits_{-\infty}^{\infty} x^m f(x)dx, \text{если X - Н.С.В.} \end{cases}$$

Центральным моментом m-го порядка *случайной величины X называется число μ_m, определяемое по формуле*

$$\mu_m = M[(X - M(x))^m] = \begin{cases} \sum\limits_k (x_k - M(x))^m p_k \text{, если X - Д.С.В.,} \\ \int\limits_{-\infty}^{\infty} (x - M(x))^m f(x)dx, \text{если X - Н.С.В.} \end{cases}$$

Из определения моментов, в частности, следует, что $\nu_0=\mu_0=1$, $\nu_1=M(x)$, $\mu_2=\nu_2-[M(x)]^2=D(x)=\sigma^2(x)$.

Также часто используются коэффициент асимметрии $a(x)=\dfrac{\mu_3}{\sigma^3}$ и коэффициент эксцесса $e(x)=\dfrac{\mu_4}{\sigma^4}-3$.

Квантилью порядка p Н.С.В. X называется число t_p, удовлетворяющее уравнению $P\{X< t_p\}=p$ или $F(t_p)=p$, в частности медиана $h(x)=t_{0,5}$.

2.3. Основные законы распределения вероятностей

Среди всех вероятностных распределений есть такие, которые на практике используются наиболее часто. Многие из них лежат в основе целых областей знаний. Эти распределения хорошо изучены и свойства их известны.

Биномиальное распределение (Б.Р.)*. Это одно из самых распространенных дискретных распределений, которое служит вероятностной моделью для многих явлений. Б.Р. возникает при исследовании последовательности (схемы) независимых испытаний, в которых результатом каждого опыта испытания может быть один из двух исходов (например, успех и неудача), и вероятности их в каждом из опытов неизменны (схема Бернулли).*
Случайная величина (Д.С.В.) X имеет биномиальное распределение с параметрами *n* и *p*, если закон распределения имеет вид

$$P(X = k) = C_n^k p^k (1-p)^{n-k} \text{ (k = 0,1,2,...,n).}$$

Математическое ожидание и дисперсия для Б.Р.: $M(x)=np$, $D(x)=np(1-p)$.
Б.Р. тесно связано с другими распределениями:

1.Б.Р. с параметрами *n* и *p* может быть аппроксимировано нормальным распределением со средним a=*np* и стандартным отклонением $\sigma = \sqrt{np(1-p)}$, $np(1-p)>5$ и $0{,}1 \le p \le 0{,}9$ или $np(1-p)>25$.

2.Б.Р. с параметрами *n* и *p* может быть аппроксимировано распределением Пуассона со средним $\lambda=np$, если *n* достаточно велико и $np<10$.

Пример. Из орудия произведено пять выстрелов по цели, вероятность попадания в которую при каждом выстреле равна 0,6. Для случайной величины

X - количества попаданий в цель, записать закон распределения, найти *M(x)* и *D(x)*; определить вероятность уничтожения цели, если для этого требуется не менее четырех попаданий.

Решение. Д.С.В. X распределена по биномиальному закону (имеем независимые испытания) с параметрами n=5 и p=0,6. Закон распределения имеет вид $P(X=k)=C_5^k\,0{,}6^k\,(1-0{,}6)^{5-k}$ $(\mathrm{k}=0,1,2,...,5)$; *M(x)*=5·0,6=3 и *D(x)*=5·0,6·(1-0,6)=1,2. Пусть событие А - цель уничтожена, тогда *P(A)=P(X=4)+P(X=5)=*$C_5^4\,0{,}6^4\,0{,}4+0{,}6^5=0{,}33696$.

Распределение Пуассона (Р.П.). Это *дискретное* распределение играет важную роль в ряде вопросов физики (радиоактивный распад), теории связи, теории надежности, теории массового обслуживания (телефонных вызовов, отказов оборудования, несчастных случаев, эпидемий и т.п). Р.П. *имеет дело с простейшими потоками.*

Случайная величина (Д.С.В.) X распределена по закону Пуассона с параметром λ>0, если закон распределения имеет вид

$$P(X=k)=\frac{\lambda^k}{k!}e^{-\lambda}\ (\mathrm{k}=0,1,2,...).$$

Математическое ожидание и дисперсия для Р.П.: M(x)=λ*, D(x)=*λ*.*

Замечание. Простейшим (пуассоновским) потоком называют поток событий, обладающий тремя свойствами: стационарностью, отсутствием последействия и ординарностью.

Свойство стационарности состоит в том, что вероятность появления k событий в любом промежутке времени зависит только от числа k и от длительности t промежутка времени и не зависит от начала его отсчета.

Поток без последействия, т.е. вероятность появления k событий в любом промежутке времени не зависит от того, появлялись или не появлялись события в моменты времени, предшествующие началу рассматриваемого промежутка. Другими словами, предыстория потока не влияет на вероятность появления событий в ближайшем будущем.

Поток ординарен, т.е. вероятность появления двух и более событий за малый промежуток времени практически невозможна. Другими словами, вероятность появления более одного события за малый промежуток времени существенно меньше вероятности появления только одного события.

Параметр λ характеризует *интенсивность простейшего потока событии* (среднее число событий в единицу времени).

Р.П. тесно связано с другими распределениями:

1) Р.П. с параметром λ может быть аппроксимировано нормальным распределением со средним а=λ и дисперсией λ, если λ>9.

2) Б.Р. с параметрами n и p может быть аппроксимировано распределением Пуассона со средним $\lambda=np$, если n достаточно велико и np<10.

Сумма n независимых случайных величин, имеющих пуассоновское распределение с параметрами $\lambda_1, \lambda_2, \ldots, \lambda_n$ *соответственно, имеет также Р.П. с параметром* $\lambda=\lambda_1+\lambda_2+\ldots+\lambda_n$.

Пример. Среднее число заказов такси, поступающих на диспетчерский пункт в одну минуту равно трем. Для случайной величины X - числа заказов такси за одну минуту - записать закон распределения и найти дисперсию. Поток заказов считать простейшим.

Решение. Д.С.В. X имеет пуассоновское распределение с параметром λ=3, т.е. закон распределения $P(X=k)=\dfrac{3^k}{k!}e^{-3}$ (k = 0,1,2,...) и $D(x)$=3.

Пример. На базу поступили 100 телевизоров, в среднем 2% среди них бракованных. Найти вероятность, что ровно три телевизора неисправны.

Решение. Случайная величина X - количество бракованных телевизоров из 100, имеет Б.Р. с параметрами n=100 и p=0,02. Тогда $P(X=3)=C_{100}^3\cdot 0{,}02^3\cdot 0{,}98^{97}$. Так как вычислить эту величину трудно, то будем считать, что случайная величина X - количество бракованных телевизоров из 100 имеет Р.П. с параметром $\lambda= np$=2. Тогда $P(X=3)=\dfrac{2^3}{3!}e^{-2}\cong 0{,}182$.

Геометрическое распределение (Г.Р.). Это еще одно распределение, связанное с повторными независимыми испытаниями (схемой Бернулли). Геометрическое распределение используется в задачах контроля продукции и других практических задачах. *Пусть производятся испытания по схеме Бернулли с вероятностью успеха в каждом из испытании, равной p. Испытания продолжаются до тех пор, пока не появится успех, после чего прекращаются. Тогда Д.С.В. X - число проведенных испытании до первого успеха включительно, имеет геометрическое распределение.*

Случайная величина (Д.С.В.) X имеет геометрическое распределение с параметром p, если закон распределения имеет вид

$$P(X=k)=p(1-p)^{k-1}\ (\mathrm{k}=1,2,3,...).$$

Математическое ожидание и дисперсия для Г.Р.: $M(x)=\dfrac{1}{p},\ \mathrm{D(x)}=\dfrac{1-\mathrm{p}}{\mathrm{p}^2}$.

Пример. Бросают игральный кубик до тех пор, пока не выпадет шестерка. Для случайной величины X - числа произведенных бросаний кубика, найти $M(x)$, $D(x)$ и $P(x\le 3)$.

Решение. Случайная величина X имеет Г.Р. с параметром p=1/6.

Следовательно, $M(x)=\dfrac{1}{1/6}=6,\ \mathrm{D(x)}=\dfrac{1-\frac{1}{6}}{\frac{1}{36}}=30;$

P(x≤3)=P(x=1)+P(x=2)+P(x=3) = $1/6+1/6\cdot 5/6+1/6(5/6)^2=91/216$.

Закон равномерной плотности (равномерное распределение). Этому закону может подчиняться, например, погрешность при измерениях с округлением или положение объекта в некоторой области, если ни одному из возможных положений нельзя отдать предпочтения. *Пусть возможные значения Н.С.В. X лежат в интервале (a;b) и нет оснований для того, чтобы отдать* *предпочтение какому-либо из этих значений. Тогда Н.С.В. X имеет равномерное распределение (Р.Р.).*

Случайная величина (Н.С.В.) X имеет равномерное распределение с параметрами a и b, если плотность распределения вероятностей

$$f(x) = \begin{cases} \dfrac{1}{b-a}, \text{если } \mathrm{x} \in (\mathrm{a};\mathrm{b}), \\ 0, \text{если } \mathrm{x} \notin (\mathrm{a};\mathrm{b}). \end{cases}$$

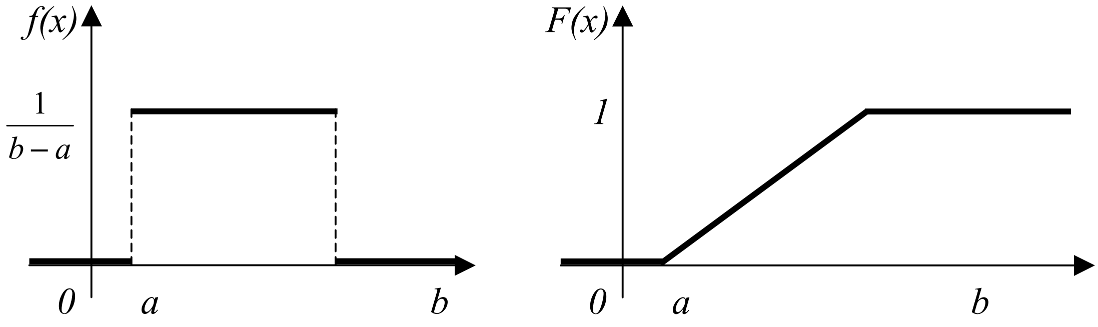


Рис. 3

Математическое ожидание, дисперсия и среднее квадратическое отклонение для Р.Р.: $M(x) = \dfrac{a+b}{2}$; $D(x) = \dfrac{(b-a)^2}{12}$; $\sigma(x) = \dfrac{b-a}{2\sqrt{3}}$.

Пример. Автобусы некоторого маршрута идут строго по расписанию. Интервал движения 5 мин. Найти среднее время ожидания пассажиром автобуса и вероятность того, что пассажир, подошедший к остановке, будет ожидать очередной автобус меньше 3 мин.

Решение. Н.С.В. *X* - время ожидания пассажиром очередного автобуса, имеет равномерное распределение на интервале (0; 5). Следовательно, $M(x) = (0+5)/2 = 2{,}5$ (мин), и вероятность того, что пассажир, подошедший к остановке, будет ожидать очередной автобус меньше 3 мин $P(0 < x < 3) = \int\limits_0^3 \dfrac{1}{5-0}dx = \dfrac{3}{5}$.

Показательное (экспоненциальное) распределение (Э.Р.) применяется в теории надежности, задачах массового обслуживания и при решении многих других практических задач. Пусть рассматривается простейший непрерывный поток однородных событии и произвольно назначен момент начала отсчета текущего времени.

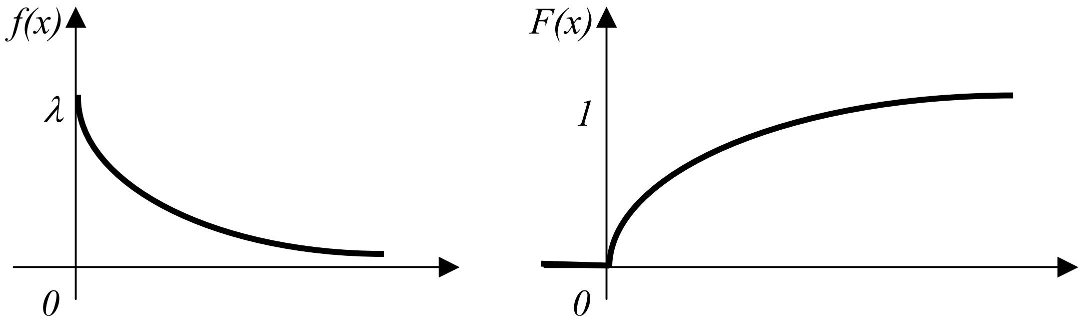


Рис. 4

Тогда Н.С.В. X время появления очередного события, подчиняется Э.Р.

В частности, этому распределению подчиняется величина промежутка времени между двумя смежными событиями простейшего потока.

Случайная величина (Н.С.В.) X имеет показательное распределение с параметром λ (λ>0), если плотность распределения вероятностей

$$f(x)=\begin{cases}0 \text{ при } x \le 0;\\ \lambda e^{-\lambda x} \text{ при } x>0.\end{cases}$$

Математическое ожидание, дисперсия и среднее квадратические отклонение для Э.Р.: $M(x)=\dfrac{1}{\lambda}$; $D(x)=\dfrac{1}{\lambda^2}$; $\sigma(x)=\dfrac{1}{\lambda}$.

Замечание. Параметр $x_0=\dfrac{1}{\lambda}$ есть среднее время между моментами наступления двух смежных событии.

Пример. Время ожидания у бензоколонки автозаправочной станции является случайной величиной *X*, распределенной по Э.Р. со средним временем ожидания, равным 2 мин. Записать плотность распределения случайной величины *X* найти вероятность $P(1<X<3)$.

Решение. Плотность распределения вероятности времени ожидания у бензоколонки имеет вид $f(x)=\begin{cases}0 \text{ при } x \le 0;\\ 0{,}5e^{-0{,}5x} \text{ при } x>0,\end{cases}$ т.к. $x_0=\dfrac{1}{\lambda}=2$. Вероятность

$$P(1<X<3)=\int_1^3 0{,}5e^{-0{,}5x}dx=e^{-0{,}5x}\Big|_1^3\cong 0{,}3834.$$

Нормальное распределение (Н.Р.). Это самое распространенное и часто используемое распределение. Причиной этого является то, что распределение суммы большого числа независимых или слабо зависимых случайных величин. распределение которых, как правило неизвестно, при весьма общих дополнительных условиях хорошо аппроксимируется нормальным распределением (центральная предельная теорема). Нормальное распределение применяется и в тех случаях, когда истинное распределение известно, но вычисления по этому закону затруднительно, а аппроксимация его Н.Р. допустима (для Б.Р., для Р.П. и многих других).

Н.С.В. X имеет нормальное распределение с параметрами a и σ, если плотность распределения вероятности $f(x)=\dfrac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-a)^2}{2\sigma^2}}$. $-\infty<x<\infty$.

Математическое ожидание, дисперсия и среднее квадратическое отклонение для Н.Р.: $\mathrm{M(x)=a; D(x)}=\sigma^2; \sigma(x)=\sigma.$ Т.е. в формуле плотности распределения вероятности *параметр a - математическое ожидание,* σ *- среднее квадратическое отклонение*.

Асимметрия *a(x)*=0 и эксцесс *e(x)*=0.

Н.Р. с параметрами *a* и σ обозначается N(*a;σ*).

Для Н.Р. вероятность того, что X примет значение, принадлежащие интервалу ($\alpha;\beta$), определяется по формуле $P(\alpha \le X < \beta) = \Phi(\frac{\beta\text{ - a}}{\sigma}) - \Phi(\frac{\alpha\text{ - a}}{\sigma})$, где

$\Phi(\mathrm{t}) = \frac{1}{\sqrt{2\pi}}\int_0^t e^{-x^2/2}dx$ - функция Лапласа.

Справедливо «правило трех сигм» (правило практической достоверности)*, состоящее в том, что* $\mathrm{P}(|\mathrm{X} - \mathrm{a}| < 3\sigma) = 0{,}9973 \cong 1$, т.е. практически все значения случайной величины X, имеющей Н.Р., расположены на интервале ($\mathrm{a}-3\sigma$; $\mathrm{a}+3\sigma$).

Пример. Случайные ошибки измерения X подчинены нормальному закону со средним квадратическим отклонением σ=20 мм и математическим ожиданием a=0. Записать плотность распределения X. Найти вероятность того, что ошибка измерения не превзойдет по абсолютной величине 4 мм.

Решение. Случайные ошибки измерения X подчинены нормальному закону N(0;20). Плотность распределения имеет вид $f(x) = \frac{1}{20\sqrt{2\pi}}e^{-\frac{x^2}{2\cdot 20^2}}$.

Вероятность того, что ошибка измерения не превзойдет по абсолютной величине 4мм $P(|X| < 4) = 2\Phi(\frac{4\text{ - }0}{20}) = 2\Phi(0{,}2) = 0{,}0796$.

Пример. Автомат штампует детали. Контролируется длина детали X, которая распределена нормально с M(x)=50 мм (проектная длина). Фактически длина изготовленных деталей не менее 32 мм и не более 68мм. Записать плотность распределения X. Найти вероятность того, что длина наудачу взятой детали: а) меньше 40мм; б) больше 55мм.

Решение. В соответствии с «правилом трех сигм» P(32<X<68)=1 или P(|X-54|<18)=1, т.е 3σ=18 $\Rightarrow$ σ=6.

Запишем плотность распределения вероятности $f(x) = \frac{1}{6\sqrt{2\pi}}e^{-\frac{(x-50)^2}{2\cdot 6^2}}$.

Найдем вероятность того, что длина наудачу взятой детали меньше 40 мм $P(X < 40) = \int_{-\infty}^{40} f(x)dx = \Phi(\frac{40\text{ - }50}{6}) - \Phi(-\infty) = \Phi(-1{,}67) + 0{,}5 = 0{,}0027$. Вероятность того, что длина наудачу взятой детали больше 55 мм $P(X > 55) = \int_{55}^{\infty} f(x)dx = \Phi(\infty) - \Phi(\frac{55-50}{6}) = 0{,}5 - \Phi(0{,}83) = 0{,}0823$.

ЗАДАНИЯ

2.1. Дайте определение случайной величины.

2.2. Что называют законом распределения (вероятностей) случайной величины?

2.3. Как, зная функцию распределения, найти вероятность попадания случайной величины в заданный интервал?

2.4. Какую случайную величину называют дискретной? Приведите примеры дискретных случайных величин.

2.5. Какой вид имеет функция распределения дискретной случайной величины?

2.6. Какую случайную величину называют непрерывной? Приведите примеры непрерывных случайных величин.

2.7. Как, зная плотность распределения, найти вероятность попадания случайной величины в заданный интервал?

2.8. Чем различаются графики функций распределения дискретной и непрерывной случайных величин?

2.9. Из партии в 10 деталей, среди которых две бракованные, наудачу выбирают три детали. Найдите закон распределения числа бракованных деталей среди выбранных. Постройте функцию распределения.

Ответ: $P\{X=i\}=\dfrac{C_2^i C_8^{3-i}}{C_{10}^3}, \quad i=0,1,2.$.

$$F(x)=\begin{cases} 0, & x\le 0;\\ 7/15, & x\in(0,1];\\ 14/15 & x\in(1,2];\\ 1 & x\rangle 2.\end{cases}$$

2.10. Вероятность приема самолетом радиосигнала при каждой передаче равна 0,7. Найдите ряд распределения и функцию распределения числа X принятых сигналов при шестикратной передаче.

Ответ: Ряд распределения и функцию распределения случайной величины X легко построить, зная, что $P\{X=i\}=C_6^i(0{,}7)^i(0{,}3)^{15-i}$, $i=\overline{0{,}6}$.

2.11. В течение часа на станцию скорой помощи поступает случайное число X вызовов, распределенное по закону Пуассона с параметром $\lambda = 5$. Найдите вероятность того, что в течение часа поступит:

а) ровно два вызова;

б) не более двух вызовов;

в) не менее двух вызовов.

Ответ: а) $P\{X = 2\} = 5^2e^{-5}/2! \approx 0{,}086$; б) $P\{X \le 2\} =$
$=(5^0/0! + 5^1/1! + 5^2/2!)e^{-5} \approx 0{,}127$;
г) $P\{X \ge 2\} = 1 - P\{X < 2\} = 1 - (5^0/0! + 5^1/1!)e^{-5} \approx 0{,}041$.

2.12. Число вызовов, поступающих на АТС (автоматическая телефонная станция) каждую минуту, распределено по закону Пуассона с параметром $\lambda = 1{,}5$. Найдите вероятность того, что за минуту поступит: а) ровно три вызова; б) хотя бы один вызов; в) менее пяти вызовов.

Ответ: а) 0,12551; 6)0,77687; в) 0,98143.

2.13. По цели производят серию независимых выстрелов до первого попадания. Даны вероятность *p* попадания в цель при одном выстреле и запас патронов *n*. Найдите ряд распределения и функцию распределения числа X израсходованных патронов.

Ответ: $P\{X=i\}=\begin{cases} pq^{i-1}, & i=\overline{0,n-1} \quad (q=1-p). \\ q^{n-1}, & i=n \end{cases}$

2.14. Непрерывная случайная величина X распределена по экспоненциальному закону с параметром $\lambda = 0,2$. Найдите вероятность попадания этой случайной величины в интервал (0; 2). Ответ: $1 - e^{-0,4} \approx 0,33$.

2.15 Длительность времени X безотказной работы элемента имеет экспоненциальное распределение с параметром $\lambda = 0,02$ ч$^{-1}$. Вычислите вероятность того, что за время t = 100 ч элемент: а) выйдет из строя; б) будет исправно работать. Ответ: а) $1 - e^{-2} \approx 0,865$; б) $e^{-2} \approx 0,135$.

2.16. Найти математическое ожидание случайной величины X, заданной рядом распределения

а)

X	-4	6	10 ;
P	0.2	0.3	0.5.

б)

X	0.21	0.54	0.61;
p	0.1	0.5	0.4 .

Ответ: а) M(X)= 6; б) M(X)= 0.535.

2.17. Дан перечень возможных значений дискретной случайной величины X : $x_1=-1$, $x_2=0$, $x_3=1$, а также известны математические ожидания этой величины и ее квадрата: M(X)=0.1, M(X^2)=0.9. Найти вероятности p_1, p_2, p_3, соответствующие возможным значениям x_1, x_2, x_3. Ответ: $p_1=0,4$; $p_2=0,1$; $p_3=0,5$.

2.18. Производятся независимые испытания трех приборов. Вероятность отказа каждого прибора соответственно равна 0,1; 0,2; 0,3. Для случайного числа отказавших приборов найти:

а) ряд распределения;

б) числовые характеристики M(X) и D(X);

в) P(0,3< X< 2,1). Ответ: M(X)=0,6; D(X)= 0,46; P=0,49.

2.19. Из урны, содержащей пять белых и три черных шара, последовательно вынимают шары с возвращением в урну до появления белого шара. Составить ряд распределения числа извлеченных черных шаров и вычислить M(X) и D(X). Ответ: M(X)=3/5; D(X)=24/25.

2.20. Функция распределения случайной величины X имеет следующий вид:

$$F(x)=\begin{cases} 0; \text{если } x \le 2 \\ 0,3; \text{если } 2 < x \le 3 \\ 0,5; \text{если } 3 < x \le 4 \\ 1, \text{если } x > 4. \end{cases}$$

Описать закон распределения случайной величины X и найти M(X),D(X) и $P\{X \ge 3,5\}$. Ответ: M(X)=3,2; D(X)=0,76;P=0,5.

2.21. Из урны,содержащей четыре белых и шесть черных шаров, случайным образом и без возвращения извлекается три шара. Случайная величина X - число белых шаров в выборке. Найти M(X) и D(X).

Ответ: M(X)=6/5;D(X)=14/25.

2.22. Случайная величина X задана на всей оси O*x* функцией распределения $F(x)=1/2+(1/\pi)arctg(x/2)$. Найти возможное значение x_1, удовлетворяющее условию: с вероятностью ¼ случайная величина X в результате испытания примет значение, большее x_1. Ответ: $x_1=2$.

2.23. Плотность вероятности непрерывной случайной величины X задана на всей оси Ox равенством $f(x)=2c/(1+x^2)$. Найти постоянный параметр с.
Ответ: $c=1/2\pi$.

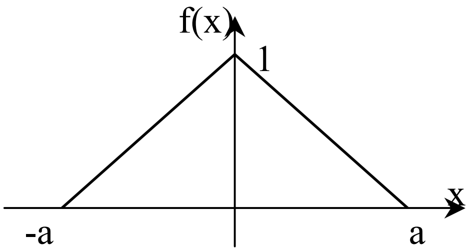
2.24. Случайная величина X распределена по закону, определяемому плотностью вероятности $f(x)=c\cos(x)$ в интервале $(-\pi/2, \pi/2)$; вне этого интервала $f(x)=0$. Найти константу c, M(X), D(X) и вычислить $P\{|x|<\pi/4\}$.

Ответ: $c=1/2$, $M(X)=0$, $D(X)=\pi^2/4-2$, $P=\sqrt{2}/2$.

2.25. Случайная величина X подчиняется закону *арксинуса* с плотностью распределения вероятностей $f(x)=\begin{cases} 0, & \text{если } |x|\geq a \\ \dfrac{1}{\pi\sqrt{a^2-x^2}}, & \text{если } |x|<a \end{cases}$.

Найти функцию распределения и вычислить M(X) и D(X).
Ответ: $M(X)=0$, $D(X)=a^2/2$.

2.26. Случайная величина X распределена по закону равнобедренного треугольника в интервале (-a,a) (закон Симпсона), если она непрерывного типа и ее плотность распределения вероятностей имеет вид , изображенный на рисунке. Найти а, M(X), D(X).
Ответ: a=1; M(X)=0; D(X)=1/6.



2.27. Азимутальный лимб имеет цену делений один градус. Какова вероятность при считывании азимута угла сделать ошибку в пределах ± 10 мин, если отсчет округляется до ближайшего целого числа градусов? Ответ: P=1/3.

2.28. Случайная величина X имеет нормальное распределение с параметрами m = 2 и σ = 1. Определите вероятность попадания случайной величины в интервал (1;5). Ответ: 0,83999.

2.29. Измерительный прибор имеет систематическую погрешность 5 м. Случайные погрешности подчиняются нормальному закону со средним квадратическим отклонением, равным 10 м. Какова вероятность того, что погрешность измерения не превзойдет по абсолютному значению 5 м?
Ответ: 0,3413.

2.30. Измерение дальности до объекта сопровождается случайными погрешностями, подчиняющимися нормальному закону со средним квадратичным отклонением, равным 50 м. Систематическая погрешность отсутствует. Найдите:

а) вероятность измерения дальности с погрешностью, не превосходящей по абсолютному значению 100 м;

б) вероятность того, что измеренная дальность не превзойдет истинной.

Ответ: а) 0,9545; б) 0,5.

2.31. Высотомер имеет случайную и систематическую погрешности. Систематическая погрешность равна 20 м. Случайная погрешность распределена по нормальному закону. Какую среднюю квадратичную погрешность должен иметь прибор, чтобы с вероятностью 0,9452 погрешность измерения высоты была меньше 10 м? Ответ: 50 м.

2.4. Закон больших чисел. Предельные теоремы

Теоремы, относящиеся к закону больших чисел, устанавливают условия, при которых среднее арифметическое случайных величин обладает свойством устойчивости. Следующие утверждения составляют содержание закона больших чисел.

Если случайная величина X имеет конечный первый абсолютный момент *M(X)*, то для любого $\varepsilon>0$ $P(|X| \geq \varepsilon) \leq \dfrac{M(|X|)}{\varepsilon}$.

В частности, если X≥0 и существует M(x), то $P(X \geq \varepsilon) \leq \dfrac{M(x)}{\varepsilon}$ *- это первое неравенство Чебышева (Маркова), или* $P(X < \varepsilon) > 1 - \dfrac{M(x)}{\varepsilon}$.

Если существует $M(X^2)$, то при любом $\varepsilon>0$ справедливы *второе неравенство Чебышева в нецентрированной форме*: $P(|X| \geq \varepsilon) \leq \dfrac{M(X^2)}{\varepsilon^2}$ и *неравенство Чебышева в центрированной форме* $P(|X - M(x)| \geq \varepsilon) \leq \dfrac{D(x)}{\varepsilon^2}$ или $P(|X - M(x)| < \varepsilon) > 1 - \dfrac{D(x)}{\varepsilon^2}$.

Замечание. Следует учесть, что приведенные неравенства дают очень грубые оценки.

Пример. Для некоторого автопарка среднее число автобусов, отправляемых на ремонт, равно пяти. Определить вероятность того, что на ремонт будет отправлено меньше 15 автобусов, если: а) информация о дисперсии отсутствует; б) дисперсия равна четырём.

Решение. Для определения вероятности того, что на ремонт будет отправлено меньше 15 автобусов, если информация о дисперсии отсутствует, воспользуемся неравенством Маркова $P(X<15)>1-5/15=2/3$. Если дополнительно известна дисперсия, то воспользуемся вторым неравенством Чебышева $P(X<15)>1-M(X^2)/225=196/225$, так как $D(x)=4=M(X^2)-25 \Rightarrow M(X^2)=29$. Дополнительная информация уточнила оценку!

Пример. Брак продукции цеха составляет в среднем 2,5%. С помощью неравенства Чебышева оценить вероятность того, что из 8000 изделии число бракованных будет от 160 до 240 включительно.

Уточнить найденную оценку вероятности, используя интегральную теорему Муавра-Лапласа.

Решение. В данной задаче мы имеем схему независимых испытаний, так как вероятность события (брак) для каждого изделия равна 0,025 и проведено n=8000 испытаний. Случайная величина X - количество бракованных изделий из 8000, имеет Б.Р. с параметрами n=8000 и p=0,025. Поэтому $M(x)=np=200$ и $D(x)=np(1-p)=195$.

С помощью неравенства Чебышева оценим вероятность того, что из 8000 изделий число бракованных будет от 160 до 240:
$P(|X-200|<40)>1-\dfrac{195}{1600}=0{,}878.$

Уточним найденную оценку вероятности, используя интегральную теорему Муавра-Лапласа: $P(|X-200|\le 40)=2\Phi(\dfrac{40}{\sqrt{195}})\ \ 2\Phi(2{,}86)=0{,}9964.$

Теорема Маркова *(закон больших чисел в общей формулировке)*. Если дисперсии произвольных случайных величин X_1, X_2, ..., X_n, удовлетворяют условию $\lim\limits_{n\to\infty}\dfrac{1}{n^2}D(\sum\limits_{k=1}^{n}X_k)=0$, то для любого ε>0

$$\lim_{n\to\infty}P\{|\frac{1}{n}\sum_{k-1}^{n}X_k-\frac{1}{n}\sum_{k-1}^{n}M(X_k)|\ge\varepsilon\}=0,$$

т.е. при выполнении сформулированных условий последовательность средних арифметических n случайных величин сходится по вероятности к среднему арифметическому их математических ожидании.

Центральная предельная теорема (в упрощенной формулировке Ляпунова). Если случайные величины X_1, X_2, ..., X_n, … независимы, одинаково распределены и имеют конечные математическое ожидание $M(x)$ и дисперсию $\sigma^2(x)$, то распределение случайной величины $Y_n=\dfrac{\frac{1}{n}\sum\limits_{k=1}^{n}X_k-M(x)}{\frac{\sigma(x)}{\sqrt{n}}}$ при $n\to\infty$ стремится к нормальному закону $N(0;1)$.

Следствием центральной предельной теоремы являются теоремы Муавра-Лапласа и многие другие утверждения, которые применяются на практике.

ЗАДАНИЯ

2.32. Напишите первое неравенство Чебышева.

2.33. Напишите второе неравенство Чебышева.

2.34. Средний ежедневный расход воды в некотором населенном пункте составляет 50.000 л. Оцените с помощью первого неравенства Чебышева вероятность того, что в произвольно выбранный день расход воды в этом пункте превысит 150.000 л. Ответ: P{X>150.000} < 1/3.

2.35. Среднее потребление электроэнергии в мае в некотором населенном пункте составляет 360.000 кВт*ч.

а) Оцените с помощью первого неравенства Чебышева вероятность того, что потребление электроэнергии в мае текущего года в этом населенном пункте превысит 1.000.000 кВт*ч.

б) Оцените с помощью второго неравенства Чебышева ту же вероятность, если известно, что среднее квадратичное отклонение потребления электроэнергии в мае равно 40.000 кВт*ч.

Ответ: а) $P\{X>1.000.000\}<0,36$; б) $P\{X>1.000.000\}<1/256$.

2.36. Вероятность появления некоторого события в каждом из 800 независимых испытаний равна ¼. Воспользовавшись вторым неравенством Чебышева, оцените вероятность того, что число X появлений этого события заключено в пределах от 150 до 250. Ответ: $P\{150 \le X \le 250\} \ge 0,94$.

2.36. Пусть дана последовательность $X_1, X_2 \dots, X_n \dots$ независимых случайных величин, причем ряд распределения случайной величины X_n представлен в таблице.

Таблица

X_{ni}	$-\sqrt{n}$	0	$\sqrt{n}$
$P(X_n=x_{ni})$	$1/(2n)$	$1-1/n$	$1/(2n)$

Проверьте, применим ли к этой последовательности закон больших чисел в форме Чебышева. Ответ: да, применим.

2.38. Решите задачу 2.36., воспользовавшись для приближенной оценки искомой вероятности интегральной теоремой Муавра – Лапласа. Сравните полученные результаты.

Ответ: $P\{150 \le X \le 250\} \approx 0,9999366$. Сравнивая полученные результаты, видим, что интегральная теорема Муавра – Лапласа дает гораздо более точный ответ.

Глава 3. Случайные векторы (многомерные случайные величины)

3.1. Законы распределения многомерных случайных величин

3.1.1. Функция распределения *n*-мерного случайного вектора

Функцией распределения *n*-мерного случайного вектора или функцией совместного распределения системы случайных величин $X_1, X_2, ..., X_n$ называется неслучайная функция *n* переменных, определяемая как вероятность выполнения системы неравенств $F(x_1, x_2, ..., x_n) = P\{X_1 < x_1, ..., X_n < x_n\}$.

В частном случае, для случайного двумерного вектора (*X,У*), $F(x, y) = P\{X < x, Y < y\}$. На плоскости (*X,У*) *F*(*x,y*) есть вероятность попасть случайной точке в квадрат с вершиной в точке (*x,y*).

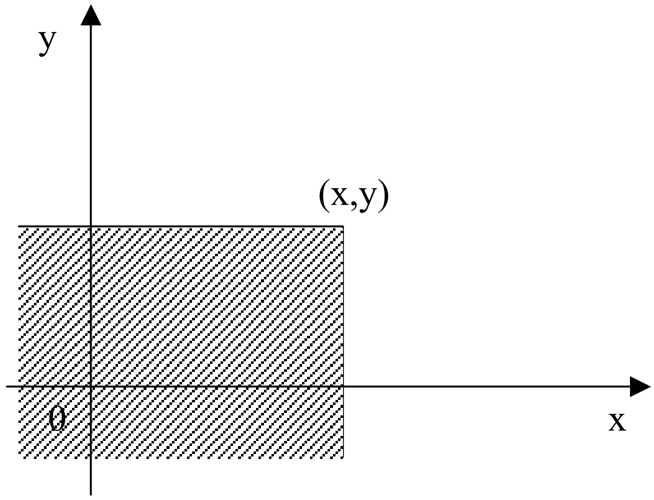


Рис. 5

Функция распределения *F*(*x,y*) обладает следующими свойствами:

1. $\lim\limits_{x\to-\infty} F(x, y) = \lim\limits_{y\to-\infty} F(x, y) = 0$;
2. $\lim\limits_{x\to\infty} F(x, y) = F(y)$; $\lim\limits_{y\to\infty} F(x, y) = F(x)$ - условие (свойство) согласованности,

которое означает, что функции распределения отдельных компонент двумерного случайного вектора могут быть найдены предельным переходом из функции распределения *F*(*x,y*).

3. $\lim\limits_{x\to\infty, y\to\infty} F(x, y) = 1$.
4. Функция непрерывная слева по каждому из аргументов.
5. Неубывающая функция своих аргументов.

С помощью функции распределения можно вычислить вероятность попадания случайной точки на плоскости (*X,У*) в прямоугольник по формуле

$$P\{x_1 \le X < x_2, y_1 \le Y < y_2\} = F(x_1, y_1) + F(x_2, y_2) - F(x_1, y_2) - F(x_2, y_1).$$

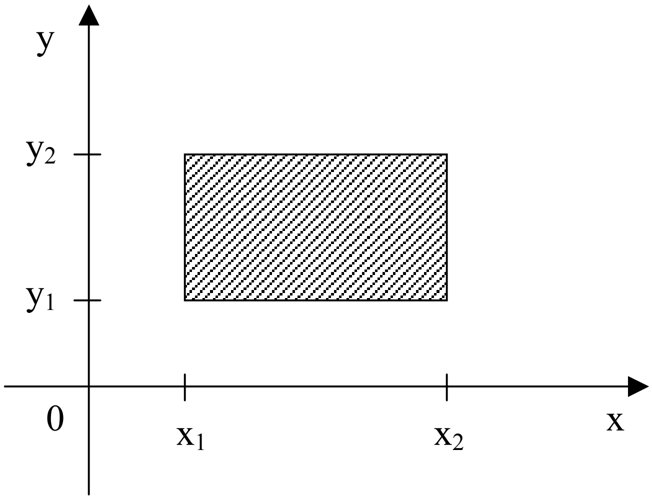


Рис. 6

3.1.2. Дискретные случайные векторы. Закон распределения многомерной случайной величины

Если X и $У$ – дискретные случайные величины, то закон распределения вероятностей случайного дискретного вектора $(X,У)$: $p_{ij} = P\{X = x_i, Y = y_j\}$, где $p_{ij} \ge 0$ и $\sum_{i,j} p_{ij} = 1$. Закон распределения может быть задан в виде таблицы.

Одномерные законы распределения отдельных компонент определяются по формулам $p_i = P\{X = x_i\} = \sum_j p_{ij}$, $p_j = P\{Y = y_j\} = \sum_i p_{ij}$, где суммирование распространяется на всевозможные значения индексов.

Условные законы распределения могут быть найдены по формулам

$$P\{X = x_i / Y = y_j\} = \frac{P\{X = x_i, Y = y_j\}}{P\{Y = y_j\}} = \frac{p_{ij}}{p_j},$$

$$P\{Y = y_j / X = x_i\} = \frac{P\{X = x_i, Y = y_j\}}{P\{X = x_i\}} = \frac{p_{ij}}{p_i}.$$

Пусть G – произвольная область на плоскости, тогда $P\{(X,Y) \in G\} = \sum_{i,j} p_{ij}$, где суммирование распространяется на те значения индексов, для которых $(x_i, y_j) \in G$.

3.1.3. Непрерывные случайные векторы. Плотность распределения вероятностей многомерной случайной величины

Пусть случайные величины *X* и *У* являются непрерывными случайными величинами и функция распределения *F*(*x,y*) непрерывна на всей плоскости, и если существует такая неотрицательная функция *f*(*x,y*), называемая *плотностью распределения* многомерной случайной величины (*X,У*), то случайный вектор (*X,У*) называется непрерывным случайным вектором (многомерной непрерывной случайной величиной). Плотность распределения вероятностей обладает следующими свойствами:

1) $f(x,y) \geq 0$;

2) $\int\limits_{-\infty}^{\infty} dx \int\limits_{-\infty}^{\infty} f(x,y)dy = 1$ (условие нормировки).

Функция распределения и плотность распределения вероятностей для случайных векторов связаны между собой следующими соотношениями:

$$F(x,y) = \int\limits_{-\infty}^{x} ds \int\limits_{-\infty}^{y} f(s,t)dt \qquad \text{и} \qquad f(x,y) = \frac{\partial^2 F(x,y)}{\partial x \partial y}.$$

Плотности распределения отдельных компонент случайного вектора находятся по формулам

$$f_x(x) = \int\limits_{-\infty}^{\infty} f(x,y)dy \qquad \text{и} \qquad f_y(y) = \int\limits_{-\infty}^{\infty} f(x,y)dx .$$

Условные плотности распределения вероятностей можно определить по следующим формулам $f(x/y) = \dfrac{f(x,y)}{f_y(y)}$ и $f(y/x) = \dfrac{f(x,y)}{f_x(x)}$.

Вероятность попадания случайной точки в произвольную область G на плоскости определяется по формуле $P\{(X,Y) \in G\} = \iint\limits_G f(x,y)dxdy$.

3.2. Числовые характеристики случайных векторов

Для двумерного случайного вектора вводятся следующие числовые характеристики.

Начальным моментом порядка *k*+*s* случайного вектора называется число $\nu_{k,s}$, которое определяется формулой

$$\nu_{k,s} = M[X^k Y^s] = \begin{cases} \int\limits_{-\infty}^{\infty} \int\limits_{-\infty}^{\infty} x^k y^s dxdy, & \text{если} \quad (X,Y) - \text{непрерывный}, \\ \sum\limits_i \sum\limits_j x_i^k y_j^s p_{ij}, & \text{если} \quad (X,Y) - \text{дискретный}. \end{cases}$$

В частности, $\nu_{k,0} = M[X^k]$, $\nu_{0,s} = M[Y^s]$ - соответствующие начальные моменты отдельных компонент.

Вектор с неслучайными координатами $\{m_x, m_y\} = \{\nu_{1,0}, \nu_{0,1}\}$ называется *математическим ожиданием* случайного вектора (*X,У*) или *центром рассеивания*.

Центральным моментом порядка k+s случайного вектора называется число $\mu_{k,s}$, которое определяется формулой

$$\mu_{k,s}=M[(X-m_x)^k(Y-m_y)^s]=$$

$$=\begin{cases}\int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty}(x-m_x)^k(y-m_y)^s f(x,y)dxdy, & \textit{если} \quad (X,Y)-\textit{непрерывный},\\ \sum\limits_{i}\sum\limits_{j}(x-m_x)^k(y-m_y)^s p_{ij}, & \textit{если} \quad (X,Y)-\textit{дискретный}.\end{cases}$$

В частности, $\mu_{2,0}=M[(X-m_x)^2]=D(x)$, $\mu_{0,2}=M[(Y-m_y)^2]=D(y)$.

Центральный момент $\mu_{1,1}$ называется *ковариацией* (*корреляционным моментом*) и обозначается K_{xy}. Таким образом, по определению

$$K_{xy}=M[(X-m_x)(Y-m_y)]=K_{yx}=M[XY]-m_x m_y.$$

Нормированная ковариация $\rho_{xy}=\dfrac{K_{xy}}{\sqrt{D(x)D(y)}}=\dfrac{K_{xy}}{\sigma(x)\sigma(y)}$ называется *коэффициентом корреляции* двух случайных компонент X и $У$ случайного вектора.

Коэффициент корреляции ρ_{xy} удовлетворяет условию $|\rho_{xy}|\le 1$ и определяет степень линейной зависимости между X и $У$ (при $|\rho_{xy}|=1$ они линейно зависимы). Случайные величины, для которых $\rho_{xy}=0$, называются *некоррелированными*.

Для n-мерного случайного вектора $\{X_1,X_2,...,X_n\}$ без труда обобщаются формулы для числовых характеристик. Например, вектор с неслучайными координатами $\{m_1,m_2,...,m_n\}$, где m_i - математическое ожидание случайной величины X_i, определяемое по формуле $m_i=M[X_i]$, называется *центром рассеивания* n-мерного случайного вектора $\{X_1,X_2,...,X_n\}$.

Ковариационной матрицей n-мерного случайного вектора $\{X_1,X_2,...,X_n\}$ называется симметрическая матрица, элементы которой представляют собой ковариации соответствующих пар компонент:

$$\begin{pmatrix}K_{11} & K_{12} & \dots & K_{1n}\\ & K_{22} & \dots & K_{2n}\\ & & \dots & \dots\\ & & & K_{nn}\end{pmatrix},$$

где $K_{ij}=M[(X_i-m_i)(X_j-m_j)]=K_{ji}$ - ковариация компонент X_i и X_j. Очевидно, что $K_{ii}=M[(X_i-m_i)^2]=D(x_i)$ - дисперсия компоненты X_i.

Корреляционной матрицей n-мерного случайного вектора $\{X_1, X_2, ..., X_n\}$ называется нормированная ковариационная матрица

$$R=\begin{pmatrix} 1 & \rho_{12} & ... & \rho_{1n} \\ & 1 & ... & \rho_{2n} \\ & & ... & ... \\ & & & 1 \end{pmatrix},$$

где $\rho_{ij} = \dfrac{K_{ij}}{\sigma_i \sigma_j}$ - коэффициент корреляции компонент X_i и X_j; $\sigma_i = \sqrt{D(x_i)}$ - среднее квадратическое отклонение компоненты X_i.

3.3. Независимость случайных величин

Понятие *независимости* является одним из важнейших в теории вероятностей. Именно это понятие выделяет теорию вероятностей из общей теории измеримых пространств.

Случайные величины $X_1, X_2, ..., X_n$ называются *независимыми* (*в совокупности*), если для любого набора событий $\{X_i \in A_i\}$, где A_i - подмножество числовой прямой (i=1,2,…,n), выполняется равенство

$$P\{X_1 \in A_1, ..., X_n \in A_n\} = P\{X_1 \in A_1\} \cdot ... \cdot P\{X_n \in A_n\}.$$

Теорема. Случайные величины $X_1, X_2, ..., X_n$ являются *независимыми* (*в совокупности*) тогда и только тогда, когда

$F(x_1, x_2, ..., x_n) = F_{x_1}(x_1) F_{x_2}(x_2) ... F_{x_n}(x_n)$.

Из этой теоремы следует, что для независимости компонент n-мерного случайного вектора $\{X_1, X_2, ..., X_n\}$ необходимо и достаточно, чтобы

$$f(x_1, x_2, ..., x_n) = f_{x_1}(x_1) f_{x_2}(x_2) ... f_{x_n}(x_n) \text{ - для непрерывного,}$$

$$P\{X_1 = x_1, ..., X_n = x_n\} = P\{X_1 = x_1\} \cdot ... \cdot P\{X_n = x_n\} \text{ - для дискретного.}$$

Ковариация и коэффициент корреляции независимых компонент X_i и X_j равны нулю, т.е. из *независимости случайных величин следует их некоррелированность* (*обратное неверно*).

3.4. Примеры

Пример 1. Случайная точка на плоскости распределена по закону, приведенному в таблице.

Таблица

Y \ X	$x_1 = 0$	$x_2 = 1$
$y_1 = -1$	0,10	0,15
$y_2 = 0$	0,15	0,25
$y_3 = 1$	0,20	0,15

Найти:

1) одномерные законы распределения;
2) условные законы распределения;
3) установить, зависимы или независимы компоненты X и Y;
4) вычислить числовые характеристики (центр рассеивания, ковариационную и корреляционную матрицы);
5) найти вероятность случайной точке попасть в область G: $(x-0{,}5)^2 + (y-0{,}5)^2 \le 1$.

Решение

1. Сложив вероятности “по столбцам”, получим закон распределения компоненты X:

x_i	0	1
p_i	0,45	0,55

Сложив вероятности “по строкам”, получим закон распределения компоненты Y:

y_j	-1	0	1
p_j	0,25	0,40	0,35

2. Составим некоторые условные законы:

x_i / y=-1	0	1
p_i	$\frac{0{,}10}{0{,}25} = 0{,}4$	$\frac{0{,}15}{0{,}25} = 0{,}6$

x_i / y=0	0	1
p_i	$\frac{0{,}15}{0{,}40} = 0{,}375$	$\frac{0{,}25}{0{,}40} = 0{,}675$

3. Самостоятельно составьте условный закон распределения компоненты X при условии, что y=1.

Таблица

y_j / x=0	-1	0	1
p_j	$\frac{0{,}10}{0{,}45}=2/9$	$\frac{0{,}15}{0{,}45}=1/3$	$\frac{0{,}20}{0{,}45}=4/9$

Самостоятельно составьте условный закон распределения компоненты Y при условии, что x=1.

4. Очевидно, что компоненты X и Y зависимы, так как $p_{ij} \neq p_i p_j$ (например, 0,10≠0,45·0,25).

5. Найдем $m_x = \sum_i \sum_j x_i p_{ij} = \sum_{i=1}^{2} x_i p_i = 0 \cdot 0{,}45 + 1 \cdot 0{,}55 = 0{,}55$;

$m_y = \sum_i \sum_j y_j p_{ij} = \sum_{j=1}^{3} y_j p_j = -1 \cdot 0{,}25 + 0 \cdot 0{,}40 + 1 \cdot 0{,}35 = 0{,}10$, т.е. центр рассеивания - вектор {0,55; 0,10}.

6. Вычислим
$$K_{xx} = D(x) = \sum_i \sum_j (x_i - m_x)^2 p_{ij} = M(x^2) - (m_x)^2 = $$
$$= 0^2 \cdot 0{,}45 + 1^2 \cdot 0{,}55 - 0{,}55^2 \approx 0{,}248;$$

$$K_{yy} = D(y) = \sum_i \sum_j (y_j - m_y)^2 p_{ij} = M(y^2) - (m_y)^2 =$$
$$= (-1)^2 \cdot 0{,}25 + 0^2 \cdot 0{,}40 + 1^2 \cdot 0{,}35 - 0{,}10^2 = 0{,}59;$$

$$K_{xy} = M[(X - m_x)(Y - m_y)] = \sum_{i=1}^{2} \sum_{j=1}^{3} (x_i - 0{,}55)(y_j - 0{,}10) p_{ij} = -0{,}055 =$$
$$= (0-0{,}55)(-1-0{,}10)0{,}10 + (1-0{,}55)(-1-0{,}10)0{,}15 + (0-0{,}55)(0-0{,}10)0{,}15 +$$
$$+ (1-0{,}55)(0-0{,}10)0{,}25 + (0-0{,}55)(1-0{,}10)0{,}20 + (1-0{,}55)(1-0{,}10)0{,}15.$$

Так как корреляционный момент отличен от нуля, то это подтверждает, что компоненты X и Y зависимы. Ковариационная матрица

$$\begin{pmatrix} 0{,}248 & -0{,}055 \\ -0{,}055 & 0{,}59 \end{pmatrix}.$$

Средние квадратические отклонения $\sigma(x) = \sqrt{0{,}248} \approx 0{,}5$; $\sigma(y) = \sqrt{0{,}59} \approx 0{,}77$. Корреляционная матрица

$$\begin{pmatrix} 1 & -0{,}14 \\ -0{,}14 & 1 \end{pmatrix}.$$

7. Вероятность случайной точке попасть в область G: $(x - 0{,}5)^2 + (y - 0{,}5)^2 \le 1$ определим следующим образом: P{(X,Y)∈G}=

=P{X=0,Y=0}+P{X=0,Y=1}+P{X=1,Y=0}+P{X=1,Y=1}=0,15+0,20+0,25+ 0,15 = =0,75.

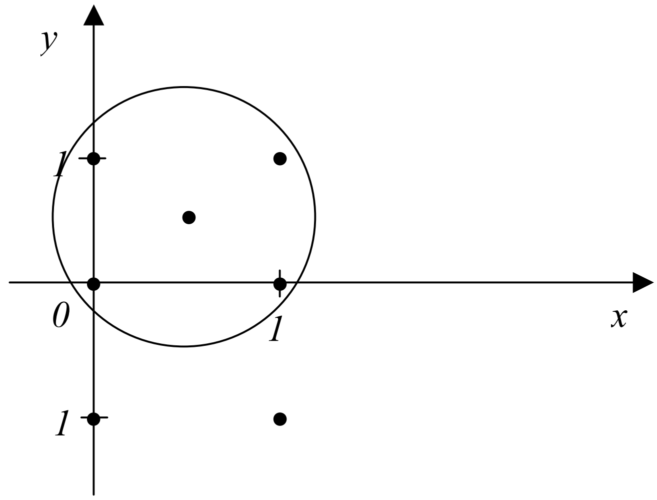


Рис. 7

Пример 2. В продукции завода брак вследствие дефекта А составляет 3%, а вследствие дефекта Б – 4,5% . Годная продукция составляет 95% . Пусть случайная величина X равна 1, если в случайно выбранной детали дефект А есть, и 0, если его нет. Аналогично случайная величина Y равна 1, если в случайно выбранной детали дефект Б есть, и 0, если его нет. Составить закон распределения двумерной случайной величины (X,Y); найти функцию распределения; вычислить коффициент корреляции.

Решение. Из условий задачи легко записать одномерные законы распределения X и Y:

x_i	0	1
p_i	0,97	0,03

y_j	0	1
p_j	0,955	0,045

Кроме того, известна вероятность P{X=0,Y=0}=0,95.

Так как P{X=0}=P{X=0,Y=0}+P{X=0,Y=1}, то P{X=0,Y=1}=0,97-0,95=0,02.

Аналогично P{X=1,Y=0}=0,955-0,95=0,005. Из условия нормировки находим, что P{X=1,Y=1}=1-0,95-0.02-0,005=0,025. Окончательно, распределение (X,Y) имеет вид

X \ Y	0	1
0	0,95	0,02
1	0,005	0,025

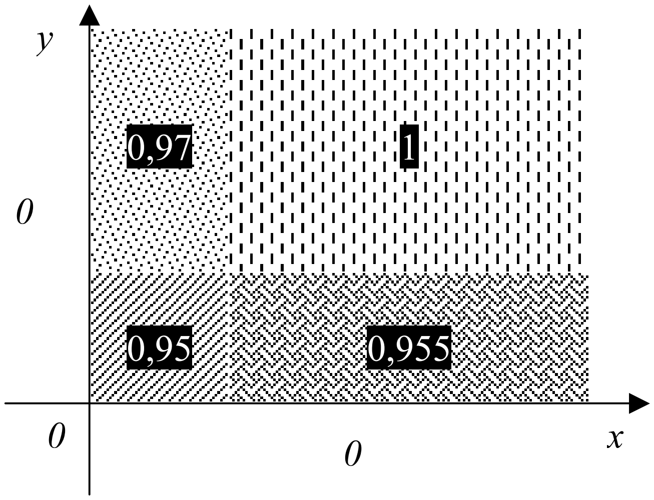


Рис. 8

Построим функцию распределения вероятностей двумерной случайной величины (X,Y): при X≤0 или Y≤0 - F(x,y)=0; при 0<x≤1 и 0<y≤1 - F(x,y)=0,95; при 0<x≤1 и y>1 - F(x,y)=0,97; при x>1 и 0<y≤1 - F(x,y)=0,955; при x>1 и

y>1 – F(x,y)=1. Для определения коэффициента корреляции найдем необходимые числовые характеристики: $m_x = 0{,}03;\quad m_y = 0{,}045;$

$D(x)=0{,}03-(0{,}03)^2=0{,}0201;$

$D(y)=0{,}045-(0{,}045)=0{,}043.$

$$K_{xy} = M(xy) - m_x m_y = 0{,}025 - 0{,}03 \cdot 0{,}045 \approx 0{,}0236;$$

$$\rho = \frac{K_{xy}}{\sqrt{D(x)D(y)}} = 0{,}784.$$

Пример 3. Система случайных величин (X,Y) распределена по закону

$$f(x,y) = \frac{a}{1 + x^2 + x^2y^2 + y^2}.$$

Найти:

1) коэффициент a;
2) одномерные плотности распределения вероятностей;
3) установить, являются ли величины X и Y зависимыми;
4) функцию распределения;
5) центр рассеивания;
6) коэффициент корреляции;
7) вероятность попадания в квадрат К: $\begin{cases} |x| \le 1, \\ |y| \le 1. \end{cases}$

Решение. 1. Из условия нормировки находим коэффициент а:

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{a dx dy}{1 + x^2 + x^2y^2 + y^2} = 1 \quad \Rightarrow \quad a = 1/\pi^2.$$

2. $f_x(x) = \int_{-\infty}^{\infty} \frac{1/\pi^2 dy}{(1+x^2))(1+y^2)} = \frac{1}{\pi(1+x^2)};$

$$f_y(y)=\int_{-\infty}^{\infty}\frac{1/\pi^2 dx}{(1+x^2))(1+y^2)}=\frac{1}{\pi(1+y^2)}.$$

3. Случайные величины X и Y независимы, так как $f(x,y)=f_x(x)f_y(y)$.

4. Так как случайные величины X и Y независимы, то $F(x,y)=F_x(x)F_y(y)$.

Найдем одномерные функции распределения:

$$F_x(x)=\int_{-\infty}^{x}\frac{dt}{\pi(1+t^2)}=\frac{1}{\pi}arc\,tgx+1/2,$$

$$F_y(y)=\int_{-\infty}^{y}\frac{dt}{\pi(1+t^2)}=\frac{1}{\pi}arc\,tgy+1/2.$$

Тогда $F(x,y)=(\frac{1}{\pi}arc\,tgx+1/2)\,(\frac{1}{\pi}arc\,tgy+1/2)$.

5. Так как $m_x=M(x)=\int_{-\infty}^{\infty}x\frac{dx}{\pi(1+x^2)}=0$ и аналогично $m_y=0$, то вектор {0;0} – центр рассеивания.

6. Коэффициент корреляции равен нулю, так как случайные величины X и Y независимы.

7. Вероятность попадания в квадрат К можно определить двумя способами: $P\{(X,Y)\in K\}=F(1;1)+F(-1;-1)-F(-1;1)-F(1,-1)=1/4$; или

$$P\{(X,Y)\in K\}=\iint_K\frac{dxdy}{\pi^2(1+x^2)(1+y^2)}=\int_{-1}^{1}\int_{-1}^{1}\frac{dxdy}{\pi^2(1+x^2)(1+y^2)}=1/4.$$

Пример 4. Поверхность распределения f(x,y) случайного вектора (X,Y) представляет прямой круговой цилиндр радиуса r, центр основания которого совпадает с началом координат. Определить высоту цилиндра h; найти одномерные плотности распределения; определить, являются ли случайные величины X и Y зависимыми; найти условные законы (плотности) распределения; определить центр рассеивания и записать ковариационную матрицу.

Решение. Плотность распределения f(x,y) случайного вектор (X,Y) имеет вид $f(x,y)=\begin{cases}h, & если\quad x^2+y^2<r^2,\\ 0, & если\quad x^2+y^2>r^2.\end{cases}$

Высоту h цилиндра определим из условия нормировки. Условие $\int_{-\infty}^{\infty}dx\int_{-\infty}^{\infty}f(x,y)dy=1$ означает, что объем цилиндра равен единице, откуда получим $h=1/\pi r^2$.

Найдем одномерные плотности распределения:

$$f_x(x)=\int_{-\infty}^{\infty}f(x,y)dy=\begin{cases}2h\sqrt{r^2-x^2}, & если\quad |x|<r,\\ 0, & если\quad |x|>r.\end{cases}$$

$$f_y(y)=\int\limits_{-\infty}^{\infty} f(x,y)dx=\begin{cases}2h\sqrt{r^2-y^2}, & если \quad |y|<r,\\ 0, & если \quad |y|>r.\end{cases}$$

Очевидно, что случайные величины X и Y зависимы.

Запишем условные законы (плотности) распределения. При |y|<r

$$f(x/y)=\frac{f(x,y)}{f_y(y)}=\begin{cases}\dfrac{1}{2\sqrt{r^2-y^2}} & при \quad |x|<\sqrt{r^2-y^2},\\ 0 & при \quad |x|>\sqrt{r^2-y^2}.\end{cases}$$

Аналогично при |x|<r $f(y/x)=\begin{cases}\dfrac{1}{2\sqrt{r^2-x^2}} & при \quad |y|<\sqrt{r^2-x^2},\\ 0 & при \quad |y|>\sqrt{r^2-x^2}.\end{cases}$

Математические ожидания равны нулю (симметричность плотностей распределения): $m_x=m_y=0$ - центр рассеивания вектор {0;0}.

$D(x)=K_{xx}=2h\int\limits_{-r}^{r}x^2\sqrt{r^2-x^2}dx=r^2/4$; аналогично $D(y)=K_{yy}=r^2/4$.

$K_{xy}=h\iint\limits_{x^2+y^2<r^2}xydxdy=0.$ Т.е. случайные величины X и Y зависимые, но не коррелированные. Ковариационная матрица имеет вид $\begin{pmatrix}r^2/4 & 0\\ 0 & r^2/4\end{pmatrix}$.

ЗАДАНИЯ

3.1. Как найти вероятность попадания двумерной случайной величины (X,Y) в прямоугольник ($a_1 \le x < a_2$, $b_1 \le y < b_2$) с помощью совместной функции распределения $F_{XY}(x,y)$?

3.2. Какую двумерную случайную величину называют дискретной?

3.3. Какую двумерную случайную величину называют непрерывной?

3.4. Как найти одномерные плотности распределения по двумерной плотности распределения?

3.5. Как найти вероятность попадания в некоторую область с помощью плотности распределения?

3.6. Какие случайные величины называют независимыми?

3.7. Как проверить независимость двух дискретных случайных величин?

3.8. Как проверить независимость двух непрерывных случайных величин?

3.9. Двумерная случайная величина (X,Y) имеет совместную функцию распределения $F(x,y)=\frac{1}{4\pi^2}(4 \operatorname{arctg} x \operatorname{arctg} y + 2\pi \operatorname{arctg} x + 2\pi \operatorname{arctg} y)$.

Найдите:

а) вероятности событий $\{-1 \le X < 1,\ 1 \le Y < \sqrt{3}\}$ и $\{X \ge 1, Y \ge \sqrt{3}\}$;

б) частные функции распределения случайных величин X и Y.
в) проверьте, являются ли случайные величины X и Y независимыми.
Ответ:
а) **P**$\{-1 \le X < 1,\ 1 \le Y < \sqrt{3}\} = 1/24$, **P**$\{X \ge 1, Y \ge \sqrt{3}\} = 1/24$;
б) $F_X(x) = \dfrac{1}{2\pi^2}(2\ \mathrm{arctg}\ x + \pi)$, $F_Y(y) = \dfrac{1}{2\pi^2}(2\ \mathrm{arctg}\ x + \pi)$;
в) да, являются.

3.10. Распределение вероятностей дискретной двумерной случайной величины (X,Y) задано таблицей

Таблица

X	Y			
	10	20	30	40
0,5	0,05	0,12	0,08	0,04
2,5	0,09	0,30	0,11	0,21

Найдите:
а) ряды распределения случайных величин X и Y;
б) значения совместной функции распределения F(x, y) в точках (2,5;25) и (9;11), а также вероятность события $\{2 \le X < 9,\ 10 \le Y \le 30\}$.
Проверьте, являются ли случайные величины X и Y независимыми.
Ответ:
а) Ряды распределения случайных величин X и Y приведены в таблицах:

X	0,5	2,5
P	0,29	0,71

Y	10	20	30	40
P	0,14	0,42	0,19	0,25

б) $F(2{,}5, 25) = 0{,}17$, $F(9{,}11) = 0{,}14$, $P\{2 \le X < 9,\ 10 \le Y \le 30\} = 0{,}50$;
в) нет, не являются.

3.11. Совместная плотность распределения непрерывной двумерной случайной величины (X,Y) имеет вид

$$p(x,y) = \begin{cases} 0, & (x-3)^2 + (y-2)^2 > 4; \\ C(2 - \sqrt{(x-3)^2 + (y-2)^2}) & (x-3)^2 + (y-2)^2 \le 4. \end{cases}$$

Найдите:
а) постоянную C;
б) частные плотности распределения случайных величин X и Y;
в) вероятность попадания случайного вектора (X,Y) в круг
$(x-3)^2 + (y-2)^2 \le 1$;
Проверьте, являются ли случайные величины X и Y независимыми.
Ответ: а) $C = 3/(8\pi)$.

б) $$p_X(x) = \begin{cases} 0, & x \notin [1,5]; \\ \dfrac{3}{8\pi}(2\sqrt{4-(x-3)^2} - (x-3)^2 \ln \dfrac{2+\sqrt{4-(x-3)^2}}{|x-3|}, & x \in [1,5]; \end{cases}$$

$$p_Y(y) = \begin{cases} 0, & y \notin [0,4]; \\ \dfrac{3}{8\pi}(2\sqrt{4-(y-3)^2} - (y-3)^2 \ln \dfrac{2+\sqrt{4-(y-3)^2}}{|y-3|}, & y \in [0,4]; \end{cases}$$

в) P = 1/2;
г) нет, не являются.

3.12. Совместная плотность распределения непрерывной двумерной случайной величины (X,Y) имеет вид

$$p(x,y)=\begin{cases}0, & x\le 0 \text{ или } y\le 0;\\ Ce^{-4x-2y} & x>0 \text{ и } y>0.\end{cases}$$

Найдите:

а) постоянную C;

б) совместную функцию распределения;

в) частные плотности распределения случайных величин X и Y;

г) вероятность попадания случайного вектора (X, Y) в область, ограниченную прямыми y = x, x + y = 2 и x = 0.

Проверьте, являются ли случайные величины X и Y независимыми.

Ответ: а) C = 8;

б) $F(x,y)=\begin{cases}0, & x\le 0 \text{ или } y\le 0;\\ (1-e^{-4x})(1-e^{-2y}), & x>0 \text{ и } y>0.\end{cases}$

в) $pX(x)=\begin{cases}0, & x\le 0\\ 4e^{-4x}, & x>0\end{cases}$ $\quad pY(y)=\begin{cases}0, & y\le 0;\\ 2e^{-2y} & y>0;\end{cases}$

г) P = $2(1-3e^{-4}+2e^{-6})/3$.

д) да, являются.

3.13. Непрерывная двумерная случайная величина (X,Y) распределена равномерно в квадрате с вершинами (0,0), (0,1), (1,0) и (1,1).

Найдите:

а) совместную плотность распределения;

б) совместную функцию распределения;

в) частные плотности распределения случайных величин X и Y;

г) вероятность попадания случайного вектора (X,Y) в круг $(x-1)^2+(y-1)^2\le 1/2$.

Проверьте, являются ли случайные величины X и Y независимыми.

Ответ:

а) $p(x,y)=\begin{cases}0, & x\notin[0.1] \text{ или } y\notin[0,1];\\ 1, & x\in[0.1] \text{ или } y\in[0,1];\end{cases}$

б) $F(x,y)=\begin{cases}0, & x\le 0 & \text{или} & y\le 0;\\ xy, & 0<x\le 1 & и & 0<y\le 1;\\ x, & 0<x\le 1 & и & y>1;\\ y, & x>1 & и & 0<y\le 1;\\ 1, & x>1 & и & y>1;\end{cases}$

в) $px(x)=\begin{cases}0, & x\notin[0,1];\\ 1, & x\in[0,1];\end{cases}$ $pY(y)=\begin{cases}0, & y\notin[0,1];\\ 1, & y\in[0,1];\end{cases}$

г) P = π/8;

д) да, являются.

3.14. Непрерывная двумерная случайная величина (X,Y) имеет совместную плотность распределения

$$p(x,y)=\frac{C}{1+x^2+y^2+x^2y^2}.$$

Найдите:

а) постоянную C;

б) совместную функцию распределения;

в) частные плотности распределения случайных величин X и Y;

г) вероятность попадания случайного вектора (X, Y) в треугольник с вершинами в точках (–1;1), (1;1) и (0;0).

Проверьте, являются ли случайные величины X и Y независимыми.

Ответ: а) C= 1/π;

б) $F(x,y)=\left(\frac{1}{\pi}arctgx+\frac{1}{2}\right)\left(\frac{1}{\pi}arctgy+\frac{1}{2}\right)$;

в) $p_x(x)=\frac{1}{\pi\left(1+x^2\right)}$, $p_y(y)=\frac{1}{\pi\left(1+y^2\right)}$;

г) P = 1/16;

д) да, являются.

3.15. Распределение вероятностей дискретной двумерной случайной величины (X, Y) задано в таблице. Проверьте, являются ли случайные величины X и Y независимыми.

X	Y			
	-3	-1	1	3
-1	0,06	0,02	0,04	0,08
o	0,15	0,005	0,10	0,20
1	0,09	0,03	0,06	0,12

Ответ: да, являются.

Глава 4. Функции случайных аргументов

4.1. Законы распределения функций случайных аргументов

1. Если X – Д.С.В. с известным законом распределения и U=g(X), где g – неслучайная функция, то U также Д.С.В., причем ее возможные значения $u_k = g(x_k)$. Если все u_k различны (например, функция g(x) строго монотонна), то $P\{U = u_k\} = P\{X = x_k\}$. Если же среди u_k имеются одинаковые значения, то $P\{U = u_k\} = \sum_{i:g(x_i)=u_k} P\{X = x_i\}$, т.е. необходимо сложить вероятности тех значений x_i, для которых $g(x_i) = u_k$.

Пример. Случайная величина X имеет следующий закон распределения вероятностей:

x_k	-1	0	1	3
p_k	0,2	0,1	0,4	0,3

Найти закон распределения вероятностей случайной величины $U = X^2 - 5$.

Решение. Возможные значения случайной величины $U = X^2 - 5$: это –4, -5, -4, 4. Так как значение –4 повторяется дважды, то соответствующие этим значениям вероятности нужно сложить. Получим закон распределения вероятностей случайной величины $U = X^2 - 5$ вида

u_k	-5	-4	4
p_k	0,1	0,2+0,4=0,6	0,3

2. Если X – Н.С.В. и U=g(X), где g – неслучайная функция, причем функция g(x) строго монотонная непрерывно дифференцируемая, то U также Н.С.В., функция распределения которой может быть найдена по формуле

$$F_u(u) = P\{U < u\} = \int_{-\infty}^{g^{-1}(u)} f_x(x)dx,$$

где функция $g^{-1}(u)$ - обратная функция к g(x). Плотность распределения вероятностей $f_u(u) = F_u'(u) = f_x(g^{-1}(u)) \left| \frac{dg^{-1}(u)}{du} \right|$.

Если функция g(x) немонотонная, то функция распределения может быть найдена по формуле $F_u(u) = \sum_i \int_{\Delta_i(u)} f_x(x)dx$, где $\Delta_i(u)$ означает интервал на оси Ox, на котором g(x)<u. Плотность получим дифференцированием по u:

$$f_u(u) = F_u'(u) = \sum_i f_x(g_i^{-1}(u)) \left| \frac{dg_i^{-1}(u)}{du} \right|,$$

где $g_i^{-1}(u)$ - значения обратной функции на интервалах $\Delta_i(u)$ оси Ох, на которых g(x)<u.

Пример. Случайная величина X распределена равномерно на интервале (-π/2; π/2). Найти плотность распределения случайной величины U=cos X.

Решение. Плотность распределения случайной величины X имеет вид $f_x(x) = 0$ при $x \notin$(-π/2; π/2) и $f_x(x) = 1/\pi$ при $x \in$(-π/2; π/2).

Функция U=cos x немонотонная на интервале (-π/2; π/2). Обратные функции и модуль их производной соответствено равны $g_1^{-1}(u)$=-arccosu и $g_2^{-1}(u)$=arccos u; $|(g_i^{-1}(u))'| = \frac{1}{\sqrt{1-u^2}}$ $(i = 1{,}2)$ при $u \in (0;1)$. Тогда плотность распределения случайной величины U имеет вид $f_u(u) = \frac{1}{\pi} \cdot \frac{1}{\sqrt{1-u^2}} + \frac{1}{\pi} \cdot \frac{1}{\sqrt{1-u^2}} = \frac{2}{\pi\sqrt{1-u^2}}$ при $u \in (0;1)$.

3. Если (X,Y) случайный вектор с заданным законом распределения и U=g(X,Y), где g(x,y) – неслучайная функция, то

$$F_u(u) = P\{U < u\} = \begin{cases} \sum\limits_i \sum\limits_j p_{ij} \ (\textit{для тех } i \textit{ и } j: g(x_i, y_j) < u), & \textit{если дискретный}, \\ \iint\limits_{g(x,y)<u} f(x,y)dxdy, & \textit{если непрерывный}. \end{cases}$$

Найдя функцию распределения, по известным правилам определим закон распределения случайной величины U.

Замечание. Для многомерных дискретных случайных величин закон распределения может быть найден так же, как и для одномерных.

Пример. Случайный вектор (X,Y) распределен по закону, заданному таблицей

X / Y	0	2	3
1	0,25	0,15	0,2
4	0,18	0,12	0,1

Составить закон распределения вероятностей случайной величины $U = X - 3Y + 7$.

Решение. Для каждой пары возможных значений (x_i, y_j) вычислим соответствующие значения $u(x_i, y_j) = x_i - 3y_j + 7$ и результат оформим в виде таблицы.

Таблица

(x_i, y_j)	(1;0)	(1;2)	(1;3)	(4;0)	(4;2)	(4;3)
$\mathrm{U}(x_i, y_j)$	8	2	-1	11	5	2
$P\{X = x_i, Y = y_j\}$	0,25	0,15	0,2	0,18	0,12	0,1

Анализируя таблицу, замечаем, что значение u=2 встречается дважды, и по правилу сложения вероятностей получаем окончательный результат: закон распределения вероятностей случайной величины $U = X - 3Y + 7$ имеет вид

u_k	-1	2	5	8	11
p_k	0,2	0,25	0,12	0,25	0,18

Пример. Случайная величина X распределена равномерно на интервале (0;2), случайная величина Y распределена равномерно на интервале (-1;1); случайные величины X и Y независимы. Найти плотность распределения случайной величины U=X+Y.

Решение. Так как случайные величины X и Y независимы, то двумерная плотность распределения равна произведению одномерных плотностей. Тогда

$$f_x(x) = \begin{cases} 1/2, & \text{если} \quad x \in (0;2), \\ 0, & \text{если} \quad x \notin (0;2). \end{cases} ; f_y(y) = \begin{cases} 1/2, & \text{если} \quad y \in (-1;1), \\ 0, & \text{если} \quad x \notin (-1;1). \end{cases} ;$$

$$f(x,y) = \begin{cases} 1/4, & \text{если} \quad x \in (0;2) \cap y \in (-1;1), \\ 0, & \text{если} \quad x \notin (0;2) \cup y \notin (-1;1). \end{cases}$$

Сначала составим функцию распределения $F_u(u)$ случайной величины U

$F_u(u) = P\{U < u\} = P\{X + Y < u\}$. Рассмотрим плоскость (x,y) и укажем, где выполняется неравенство x+y<u при различных значениях u. Если $u \le -1$, неравенство выполняется в области D_1 (на рисунке) и тогда $P\{X + Y < u\} = 0$, так как f(x,y)=0 в этой области.

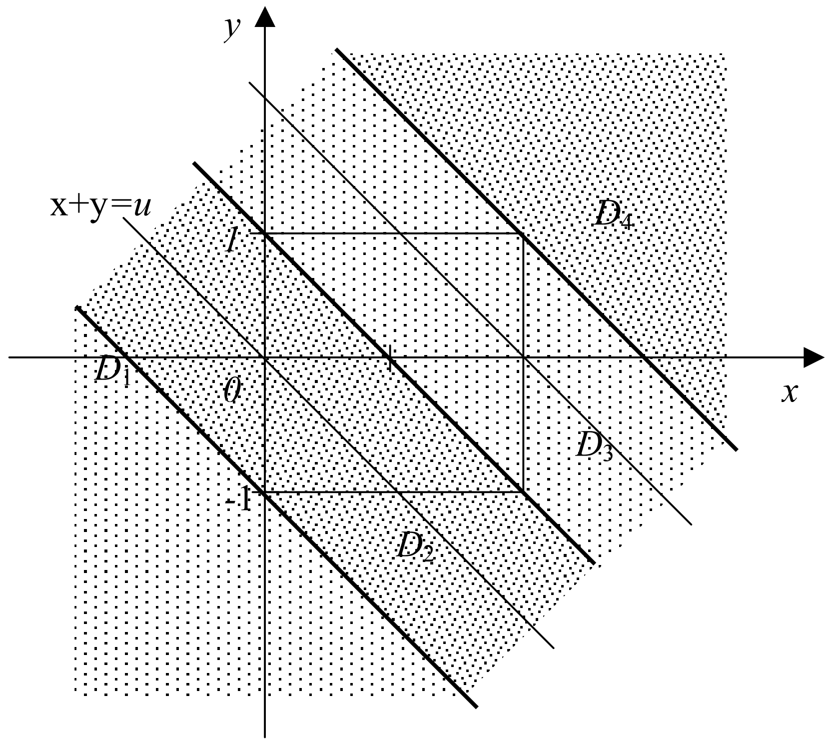


Рис. 9

При $-1 < u \le 1$ неравенство x+y<u выполняется в области D_2 и $P\{X+Y<u\} = \int\limits_{-1}^{u} dy \int\limits_{0}^{u-y} \frac{1}{4} dx = (u+1)^2/8$.

При $1 < u \le 3$ неравенство x+y<u выполняется в области D_3 и $P\{X+Y<u\} = \int\limits_{-1}^{1} dy \int\limits_{0}^{u-1} \frac{1}{4} dx + \int\limits_{u-1}^{2} dx \int\limits_{-1}^{u-x} \frac{1}{4} dy = (-u^2+6u-1)/8$.

При $u > 3$ неравенство x+y<u выполняется в области D_4 и $P\{X+Y<u\} = 1$.

Итак,

$$F_u(u) = \begin{cases} 0, & u \le 1; \\ (u+1)^2/8, & -1 < u \le 1; \\ (-u^2+6u-1)/8, & 1 < u \le 3; \\ 1, & u > 3. \end{cases} \Rightarrow f_u(u) = \begin{cases} 0, & u \le 1; \\ (u+1)/4, & -1 < u \le 1; \\ (-u+3)/4, & 1 < u \le 3; \\ 0, & u > 3. \end{cases}$$

Замечание. В одном из важных частных случаев функциональной зависимости U=X+Y решение задачи о законе распределения случайной величины U, по известной плотности совместного распределения f(x,y) может быть получено по формуле $f_u(u) = \int_{-\infty}^{\infty} f(x, u-x)dx = \int_{-\infty}^{\infty} f(u-y, y)dy$.

Если случайные величины X и Y независимы, то формула сводится к свертке двух одномерных плотностей $f_u(u) = \int_{-\infty}^{\infty} f_x(x) f_y(u-x)dx = \int_{-\infty}^{\infty} f_x(u-y) f_y(y)dy$. Самостоятельно решите предыдущий пример с использованием этой формулы.

Задача об определении закона распределения суммы независимых случайных величин носит название задачи композиции.

ЗАДАНИЯ

4.1. Как найти ряд распределения функции от дискретной случайной величины?

4.2. Как найти функцию распределения функции от непрерывной случайной величины?

4.3. Как найти плотность распределения монотонной функции от непрерывной случайной величины?

4.4. Как найти плотность распределения кусочно-монотонной функции от непрерывной случайной величины?

4.5. Как найти ряд распределения скалярной функции от дискретного случайного вектора?

4.6. Дискретная случайная величина X имеет ряд распределения, представленный в таблице. Найдите ряд распределения случайной величины, если:

а) Y = 10X – 1; б) $Y = -X^2$; в) $У = 2^X$.

X	-0,5	0	0,5	1	1,5
P	0,1	0,4	0,1	0,3	0,1

Ответ:

а) ряд распределения случайной величины Y представлен в таблице

Y	-6	-1	4	9	14
P	0,1	0,4	0,1	0,3	0,1

б) ряд распределения случайной величины Y представлен в таблице

Y	-2,25	-1	-0,25	0
P	0,1	0,3	0,2	0,4

в) ряд распределения случайной величины Y представлен в таблице

Y	$1/\sqrt{2}$	1	$\sqrt{2}$	2	$2\sqrt{2}$
P	0,1	0,4	0,1	0,3	0,1

4.7. Случайная величина X распределена равномерно в интервале (0,3). Найдите функцию распределения случайной величины $Y=X^2+1$.

Ответ: $F_Y(y)=\begin{cases}0, & y\le 1;\\ \dfrac{\sqrt{y-1}}{3}, & 1\le y\le 10;\\ 1, & y>10.\end{cases}$

4.8. Распределение случайной величины (X_1,X_2) задается таблицей.

X_2	X_1		
	10	12	14
1	0,08	0,02	0,10
2	0,32	0,08	0,40

Найдите ряд распределения случайной величины Y, если:

а) $Y = X_1 - 2X_2 - 8$;

б) $Y = (X_1 - 12)^2 + X^2_2 - 1$;

в) $Y = (X_1 - 12)/X_2$.

Ответ:

а) ряд распределения случайной величины Y представлен в таблице

Y	-2	0	2	4
P	0,32	0,16	0,42	0,10

б) ряд распределения случайной величины Y представлен в таблице

Y	0	3	4	7
P	0,02	0,08	0,18	0,72

в) ряд распределения случайной величины Y представлен в таблице

Y	-2	-1	0	1	2
P	0,08	0,32	0,10	0,40	0,10

4.9. Двумерная случайная величина (X_1,X_2) распределена равномерно в прямоугольнике с вершинами в точках $A_1(0;0)$, $A_2(0;2)$, $A_3(3;2)$, $A_4(3;0)$. Найдите функцию распределения случайной величины Y, если:

а) $Y = X_1 + X_2$;

б) $Y = X_1/X_2$.

Ответ:

а) $F_Y(y)=\begin{cases}0, & y\le 0;\\ y^2/12 & 0<y\le 2;\\ (y-1)/3, & 2<y\le 3;\\ \left[12-(5-y)^2\right]/12, & 3<y\le 5;\\ 1, & y>5.\end{cases}$

б) $F_Y(y)=\begin{cases}0, & y\le 0;\\ y/3, & 0<y\le 3/2;\\ 1-3/(4y), & y>3/2.\end{cases}$

4.10. Независимые случайные величины X_1 и X_2 имеют экспоненциальное распределение с параметрами $\lambda_1=1$ и $\lambda_2=2$ соответственно. Найдите плотность распределения случайной величины Y = X_1+X_2. Проверить результат по формуле свертки.

Ответ: $p_Y(y)=\begin{cases} 0, & y\le 0;\\ 2\left(e^{-y}-e^{-2y}\right), & y>0.\end{cases}$

4.11. Независимые случайные величины X_1 и X_2 имеют равномерное распределение на отрезках [0,1] и [0,2] соответственно. Воспользовавшись формулой свертки, найдите плотность распределения случайной величины Y = X_1 + X_2.

Ответ: $p_Y(y)=\begin{cases} 0, & y\notin(0,3);\\ y/2, & 0<y\le 1;\\ 1/2, & 1<y\le 2;\\ (3-y)/2, & 2<y<3.\end{cases}$

4.2. Числовые характеристики функций случайных величин

Если X – случайная величина с известным законом распределения и U=g(X), где g – неслучайная функция, то математическое ожидание и дисперсия случайной величины U могут быть найдены по формулам

$$m_u=M(u)=\begin{cases}\sum\limits_k g(x_k)P\{X=x_k\}, & \text{если}\quad X-ДСВ.,\\ \int\limits_{-\infty}^{\infty} g(x)f_x(x)dx, & \text{если}\quad X-Н.С.В.,\end{cases}$$

$$D_u=M(u-m_u)^2=\begin{cases}\sum\limits_k (g(x_k)-m_u)^2P\{X=x_k\}, & \text{если}\quad X-ДСВ.,\\ \int\limits_{-\infty}^{\infty} (g(x)-m_u)^2 f_x(x)dx, & \text{если}\quad X-Н.С.В.\end{cases}$$

Аналогичные формулы имеют место и для других начальных и центральных моментов случайной величины U.

Следовательно, для вычисления числовых характеристик функций случайных величин не нужно знать закона распределения случайной величины U=g(X), а достаточно знать только закон распределения аргумента X.

Сформулированное правило естественно обобщается на функции от многомерных случайных величин. Например, если (X,Y) случайный вектор с заданным законом распределения и U=g(X,Y), где g(x,y) – неслучайная функция, то

$$m_u=M(u)=\begin{cases}\sum\limits_k\sum\limits_j g(x_k,y_j)P\{X=x_k,Y=y_j\}, & \text{если}\quad (X,Y)-Д.С.В.,\\ \int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty} g(x,y)f_{xy}(x,y)dxdy, & \text{если}\quad (X,Y)-Н.С.В.\end{cases}$$

Пример. Случайная величина X имеет следующий закон распределения вероятностей:

x_k	-1	0	1	3
p_k	0,2	0,1	0,4	0,3

Найти математическое ожидание и дисперсию случайной величины $U = X^2 - 5$.

Решение. Математическое ожидание и дисперсию случайной величины U определяем по соответствующим формулам $M(u) = ((-1)^2 - 5)0{,}2 + (0^2 - 5)0{,}1 + (1^2 - 5)\ 0{,}4 + (3^2 - 5)0{,}3$=-1,7.

$$D(u) = [((-1)^2 - 5) + 1{,}7]^2 0{,}2 + [(0^2 - 5) + 1{,}7]^2 0{,}1 + [(1^2 - 5) + 1{,}7]^2 0{,}4 +$$
$$+ [(3^2 - 5) + 1{,}7]^2 0{,}3 = 14{,}01.$$

Проверьте результаты, используя решение соответствующего примера из предыдущего параграфа.

Пример. Случайный вектор (X,Y) распределен по закону, заданному таблицей

X / Y	0	2	3
1	0,25	0,15	0,2
4	0,18	0,12	0,1

Найти математическое ожидание случайной величины $U = X - 3Y + 7$.

Решение. Математическое ожидание случайной величины U определяем по соответствующей формуле M(u)=(1-3·0+7)·0,25+(1-3·2+7)·0,15+(1-3·3+ +7)·0,2+(4- -3·0+7)·0,18+(4-3·2+7)·0,12+(4-3·3+7)·0,1=4,88.

Проверьте результат, используя решение соответствующего примера из предыдущего параграфа.

Пример. Случайная величина X распределена равномерно на интервале (-π/2; π/2). Найти математическое ожидание и дисперсию случайной величины U=cos x.

Решение. Плотность распределения случайной величины X имеет вид $f_x(x) = 0$ при $x \notin$(-π/2; π/2) и $f_x(x) = 1/\pi$ при $x \in$(-π/2; π/2). Математическое ожидание и дисперсию случайной величины U определяем по соответствующим формулам

$$M(u) = \int_{-\pi/2}^{\pi/2} \cos x \,\frac{1}{\pi} dx = 2/\pi; \quad D(u) = \int_{-\pi/2}^{\pi/2} (\cos x - \frac{2}{\pi})^2 \frac{1}{\pi} dx = \frac{\pi^2 - 8}{2\pi^2}.$$

Пример. Случайная величина X распределена равномерно на интервале (0;2), случайная величина Y распределена равномерно на интервале (-1;1); случайные величины X и Y независимы. Найти математическое ожидание случайной величины U=X+Y.

Решение. Так как случайные величины X и Y независимы, то двумерная плотность распределения равна произведению одномерных плотностей. Тогда

$$f_x(x)=\begin{cases}1/2, & если \quad x\in(0;2),\\ 0, & если \quad x\notin(0;2).\end{cases};\quad f_y(y)=\begin{cases}1/2, & если \quad y\in(-1;1),\\ 0, & если \quad x\notin(-1;1).\end{cases}$$

$$f(x,y)=\begin{cases}1/4, & если \quad x\in(0;2)\cap y\in(-1;1),\\ 0, & если \quad x\notin(0;2)\cup y\notin(-1;1).\end{cases}$$

Тогда математическое ожидание случайной величины U=X+Y равно

$$M(u)=\int_0^2 dx\int_{-1}^1 (x+y)\frac{1}{4}dy=1.$$

Проверьте результат, используя решение соответствующего примера из предыдущего параграфа.

Отметим основные свойства математического ожидания и дисперсии.

1. Для любых случайных величин $X_k \quad (k=1,2,...,n)$

$M[\sum_{k=1}^{n} a_k X_k + b]=\sum_{k=1}^{n} a_k M(X_k)+b$ (свойство линейности).

$$D[\sum_{k=1}^{n} a_k X_k + b]=\sum_{k=1}^{n} a^2{}_k D(X_k)+2\sum_{i,j=1}^{n}\sum_{j>i} a_i a_j K_{ij},$$

где $K_{ij}=M[(X_i-m_i)(X_j-m_j)]=K_{ji}$ - ковариация X_i и X_j.

В частности, $D[aX+bY+c]=a^2D(x)+b^2D(y)+2abK_{xy}$.

2. $M[XY]=M(X)M(Y)+K_{xy}$.

3. Если X и Y независимы, то

$$M[XY]=M(X)M(Y);\quad D(X\pm Y)=D(X)+D(Y);$$

$$D[XY]=D(X)D(Y)+m^2{}_x D(Y)+m_y^2 D(X).$$

Пример. Найти математическое ожидание и дисперсию по условию предыдущего примера, используя свойства математического ожидания и дисперсии.

Решение. Вычислим математические ожидания и дисперсии независимых случайных величин X и Y, одномерные плотности которых известны:

$$f_x(x)=\begin{cases}1/2, & если \quad x\in(0;2),\\ 0, & если \quad x\notin(0;2).\end{cases}\quad и \quad f_y(y)=\begin{cases}1/2, & если \quad y\in(-1;1),\\ 0, & если \quad x\notin(-1;1).\end{cases}$$

$$M(x)=\int_0^2 x\frac{1}{2}dx=1;\quad D(x)=\int_0^2 (x-1)^2\frac{1}{2}dx=1/3;$$

$$M(y)=\int_{-1}^1 y\frac{1}{2}dy=0;\quad D(y)=\int_{-1}^1 (y-0)^2\frac{1}{2}dy=1/3.$$

По свойствам математического ожидания и дисперсии независимых случайных величин X и Y,

$$M(u) = M(X+Y) = 1+0 = 1;\ \ D(u) = D(X+Y) = 1/3 + 1/3 = 2/3.$$

Пример. На вход некоторого устройства поступает случайный вектор (X,У) со следующими числовыми характеристиками:
$m_x = -1, m_y = 1, D(x) = 4, D(y) = 9, \rho_{xy} = 0{,}5$. На выходе измеряется величина $U = (X+Y)^2$. Определить математическое ожидание случайной величины U.

Решение. Воспользуемся соответствующими свойствами математического ожидания

$$M(u) = M[X^2 + Y^2 + 2XY] = M(X^2) + M(Y^2) + 2M(XY) = D(x) +$$
$$+ m_x^2 + D(y) + m_y^2 + 2(m_x m_y + K_{xy}) = 4 + 9 + 2 + 2(-1 + 0{,}5 \cdot 2 \cdot 3) = 19.$$

ЗАДАНИЯ

4.12. Перечислите свойства математического ожидания случайной величины.

4.13. В каком случае математическое ожидания суммы случайных величин равно сумме математических ожиданий слагаемых?

4.14. Что называют дисперсией случайной величины?

4.15. Что называют средним квадратичным отклонением случайной величины?

4.16. Перечислите свойства дисперсии случайной величины.

4.17. Что называют начальным моментом k-го порядка случайной величины?

4.18. Что называют центральным моментом k-го порядка случайной величины?

4.19. Как называют случайные величины, ковариация которых равна нулю?

4.20. Что называют коэффициентом корреляции случайных величин?

4.21. Перечислите свойства коэффициента корреляции случайных величин.

4.22. Найдите математическое ожидание, дисперсию и среднее квадратичное отклонение дискретной случайной величины X, ряд распределения которой представлен в таблице

X	1	2	3
P	0.30	0.21	0.49

Ответ: MX = 2,19, DX = 5,55, $\sigma \approx 2{,}35$.

4.23. Вероятность того, что при трех выстрелах стрелок попадет в цель хотя бы один раз, равна 0,992. Найдите математическое ожидание и дисперсию числа X попаданий при двадцати выстрелах.

Ответ: MX = 16, DX = 3,2

4.24. Время X безотказной работы станка имеет экспоненциальное распределение. Известно, что вероятность отказа станка за 5 ч равна 0,39347. Найдите математическое ожидание, дисперсию и среднее квадратичное отклонение времени безотказной работы станка.

Ответ: MX = 10 ч, DX = 100 ч^2, σ = 10 ч.

4.25. Независимые случайные величины X_1 и X_2 имеют экспоненциальное распределение с параметрами λ_1 и λ_2 соответственно. Найдите математическое ожидание и дисперсию случайной величины $Y = X^2 + 1$.

Ответ: $MY = 1/(\lambda_1\lambda_2)$.

4.26. Дискретная случайная величина X имеет ряд распределения, представленный в таблице. Найдите математическое ожидание и дисперсию случайной величины $Y = X^2 + 1$.

X	*1*	*2*	*3*	*4*
P	0.1	0.4	0.3	0.2

Ответ: MY = 2,6, DY = 0,84.

4.27. Двумерная случайная величина (X_1,X_2) распределена равномерно в квадрате $\{0 < x_1 < 1,\ 0 < x_2 < 1\}$. Найдите математическое ожидание и дисперсию площади Y прямоугольника со сторонами X_1 и X_2.

Ответ: MY = 1/4, DY = 7/144.

4.28. Плотность распределения непрерывной случайной величины X имеет вид

$$p(x) = \begin{cases} 0, & x \notin (-1;1); \\ 3(1-x^2)/4, & x \in (-1;1). \end{cases}$$

Найдите начальные и центральные моменты первого, второго, третьего и четвертого порядка, а также асимметрию и эксцесс случайной величины X.

Ответ:
$$\begin{gathered} m_1 = \overset{0}{m}_1 = 0, \\ m_2 = \overset{0}{m}_2 = DX = 1/5 \\ m_3 = \overset{0}{m}_3 = 0, \\ m_4 = \overset{0}{m}_4 = 3/35, \\ \gamma_1 = 0, \\ \gamma_2 = -6/7. \end{gathered}$$

4.29. Совместная плотность распределения двумерной непрерывной случайной величины (X_1,X_2) имеет вид

$$p(x_1,x_2) = \begin{cases} 0, & \textit{при} \quad x_1 \le 0 \text{ или } x_2 \le 0; \\ 4x_1x_2e^{-(x_1^2+x_2^2)}, & \textit{при} \quad x_1 > 0 \text{ и } x_2 > 0. \end{cases}$$

Найдите математические ожидания, дисперсии, ковариацию, коэффициент корреляции, а также ковариационную и корреляционную матрицы случайных величин X_1 и X_2.

Ответ:

$MX_1 = MX_2 = \sqrt{\pi}/2$, $DX_1 = DX_2 = (4-\pi)/4$, $\mathrm{cov}(X_1,X_2) = 0$, $\rho = 0$,

$$\sum \begin{pmatrix} (4-\pi)/4 & 0 \\ 0 & (4-\pi)/4 \end{pmatrix}, \qquad P = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

4.30. Случайные величины X_1 и X_2 имеют математические ожидания $MX_1 = -5$, $MX_2 = 2$, дисперсии $DX_1 = 0{,}5$, $DX_2 = 0{,}4$ и ковариацию $\mathrm{cov}(X_1,X_2) = 0{,}2$. Найдите математическое ожидание и дисперсию случайной величины $Y = 4X_1 - 5X_2 + 25$.

Ответ: MY = -5, DY = 10.

Глава 5. Основные понятия математической статистики

5.1. Предмет и задачи математической статистики

Теория вероятностей и математическая статистика занимаются количественным и качественным анализом закономерностей случайных массовых явлений. При изучении курса теории вероятностей предполагалось, что вероятности наступления отдельных событий известны. Считались известными законы распределения случайных величин или их числовые характеристики. Мы оперировали этими понятиями, находили вероятности, законы распределения и числовые характеристики других более сложных событий и случайных величин. Как правило, на практике вероятности наступления событий, законы распределения случайных величин или параметры этих законов распределения неизвестны. Для их определения (оценивания) необходимо производить эксперимент, специальные испытания.

Математическая статистика разрабатывает методы математической обработки результатов испытания с целью получения сведений о вероятностях наступления отдельных событий, о законах распределения случайных величин или параметрах этих законов.

При обработке результатов эксперимента статистическими методами основные понятия теории вероятностей – вероятность наступления случайного события, законы распределения случайных величин, параметры законов распределения случайных величин и т.д. - выступают как некоторые математические модели реальных закономерностей. Таким образом, теория вероятностей разрабатывает математические модели для описания реальных закономерностей случайных массовых явлений, формирует систему взглядов на статистическую обработку результатов эксперимента.

Основой статистических методов являются экспериментальные данные, часто называемые *статистическими данными*.

Статистическими данными называют сведения о числе объектов, обладающих теми или иными признаками. Например, статистическими данными являются данные отклонений размеров деталей от номинального размера; данные о прочностных признаках образцов некоторого сорта стали; данные о числе вызовов на телефонной станции между восьмью и девятью часами утра; данные о производительности труда рабочих автотранспортного предприятия за отчетный период и т.д. Перечисленные данные являются числовыми характеристиками массовых случайных явлений (сортности деталей, нагрузки АТС, производительности труда), поэтому предметом математической статистики являются случайные явления, а ее основной задачей – количественный и качественный анализы этих явлений.

Основные задачи математической статистики состоят в разработке методов:

1) организации и планирования статистических наблюдений;
2) сбора статистических данных;
3) «свертки информации», т.е. методов группировки и сокращения статистических данных с целью сведения большого числа таких данных к

небольшому числу параметров, которые в сжатом виде характеризуют всю исследуемую совокупность;
4) анализа статистических данных;
5) принятия решений, рекомендаций и выводов на основе анализа статистических данных;
6) прогнозирования случайных явлений.

Одним из основных методом статистического наблюдения является выборочный метод. Рассмотрим основные понятия этого метода.

5.2. Методы статистического описания результатов наблюдения

5.2.1. Выборка и способы её представления

Математическая статистика позволяет получать обоснованные выводы о параметрах или виде распределения случайных величин по совокупности наблюдений над ними.

Генеральной совокупностью X называют множество результатов всех мыслимых наблюдений, которые могут быть сделаны при данном комплексе условий.

Выборочной совокупностью (выборкой) называют множество результатов, случайно отобранных из генеральной совокупности. Выборка должна быть репрезентативной (представительной), т.е. правильно отражать пропорции генеральной совокупности. Это достигается случайностью отбора, когда все объекты генеральной совокупности имеют одинаковую вероятность быть отобранными.

Задачи математической практически сводятся к обоснованному суждению об объективных свойствах генеральной совокупности по результатам случайной выборки.

Выборкой объема n из генеральной совокупности с функцией распределения F(x) называется последовательность $x_1, x_2, \ldots, x_n$ наблюдаемых значений случайной величины X, соответствующих n независимым повторениям опыта.

Выборка может быть записана ввиде вариационного ряда или в виде статистического ряда. ***Вариационным рядом*** выборки $x_1, x_2, \ldots, x_n$ называется способ её записи, при котором элементы упорядочиваются по величине. Разность между максимальным и минимальным элементами выборки w называется размахом выборки. Пусть в выборке объема n элемент x_i встречается n_i раз, тогда число n_i называется *частотой* элемента x_i. Очевидно, что $\sum_{i=1}^{n} n_i = n$.

Статистическим рядом называется последовательность пар (x_i, n_i). Обычно статистический ряд записывается в виде таблицы, первая строка которой содержит элементы x_i, а вторая их частоты.

При большом объеме выборки её элементы объединяются в группы (разряды), представляя результаты опытов в виде группированного статистического ряда. Для этого интервал, содержащий все элементы выборки, разбивается на k частичных непересекающихся интервалов (лучше одинаковой длины). После того как частичные интервалы выбраны, определяются частоты – количество n_i^* элементов выборки, попавших в i-й интервал (элемент совпадающий с верхней границей интервала, относительно к последующему интервалу). Получающийся статистический ряд в верхней строке содержит середины x_i^* интервалов группировки, а в нижней – частоты n_i^*

В зависимости от объемов выборки число интервалов группировки k берётся от 6 до 20.

Гистограммой частот группированной выборки называется кусочно постоянная функция, постоянная на интервалах группировки и принимающая на каждом из них значения n_i/b соответственно. Площадь ступенчатой фигуры под графиком гистограммы равна объему выборки n. Если на гистограмме частот соединить середины верхних сторон элементарных прямоугольников, то полученная замкнутая ломаная образует *полигон частот.*

Аналогично можно построить полигон и гистограмму относительных частот. *В теории вероятностей гистограмме относительных частот соответствует график плотности распределения. Для дискретных случайных величин полигону относительных частот соответствует многоугольник распределения.*

Пример 1. Представить выборку 55 наблюдений в виде таблицы частот, используя семь равных интервалов группировки. Выборка

17	19	23	18	21	15	16	13	20	18	15
20	14	20	16	14	20	19	15	19	16	19
15	22	21	12	10	21	18	14	14	17	16
13	19	18	20	24	16	20	19	17	18	18
21	17	19	17	13	17	11	18	19	19	17

Решение. Размах выборки w=24-10=14 длина интервала b=14/7-2.

Наряду с частотами одновременно подсчитываются также накопленные частоты $\sum_{j=1}^{i} n_j$, относительные частоты n_i^*/n и накопленные относительные частоты $\sum_{j=1}^{i} n_j^*/n$, *(i=1,2,..,k)*. Полученные результаты сводятся в таблицу частот группированной выборки:

Таблица

Номер интервала i	Границы интервала	Середина интервала x_i^*	Частота n_i^*	Накопленная частота $\sum_{j=1}^{i} n_i^*$	Относительная частота n_i^*/n	Накопленная относительная частота $\sum_{j=1}^{i} n_i^*/n$

1	10-12	11	2	2	0,0364	0,0364
2	12-14	13	4	6	0,0727	0,1091
3	14-16	15	8	14	0,1455	0,2546
4	16-18	17	12	26	0,2182	0,4728
5	18-20	19	16	42	0,2909	0,7637
6	20-22	21	10	52	0,1818	0,9455
7	22-24	23	3	55	0,0545	1,0000

Полученные результаты можно представить графически (рис. 9,10).

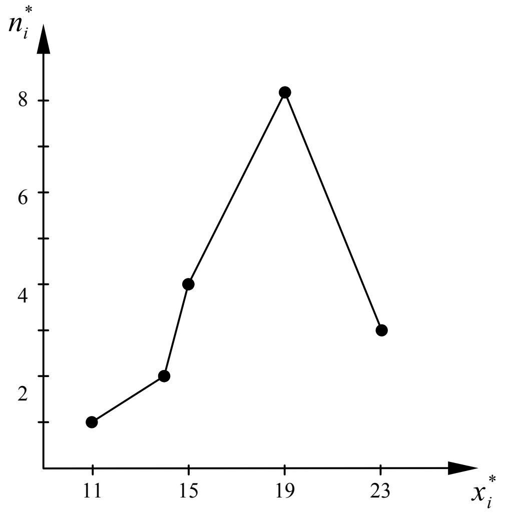


Рис. 9. Полигон частот

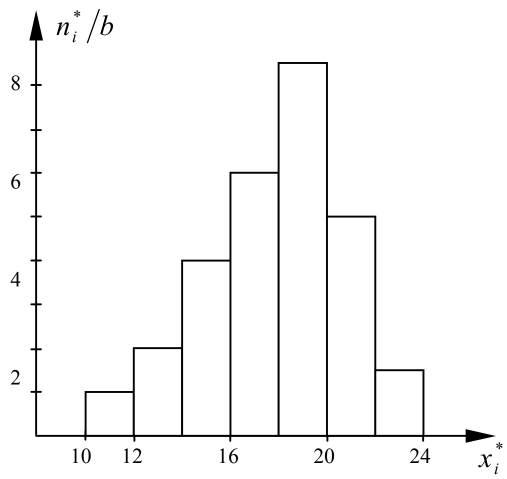


Рис. 10. Гистограмма частот

Эмпирическая функция распределения $F^*(x)$ определяется по значениям накопленных относительных частот соотношением $F^*(x) = \frac{1}{n}\sum_{x_i^* < x} n_i$, где суммируются относительные частоты тех элементов выборки, для которых выполняется неравенство $x_i^* < x$.

Для приведенного примера

$$F^*(x) = \begin{cases} 0 & x \le 11; \\ 0{,}0364 & 11 < x \le 13; \\ 0{,}1091 & 13 < x \le 15; \\ 0{,}2546 & 15 < x \le 17; \\ 0{,}4728 & 17 < x \le 19; \\ 0{,}7637 & 19 < x \le 21; \\ 0{,}9455 & 21 < x \le 23; \\ 1 & 23 < x. \end{cases}$$

5.2.2. Числовые характеристики выборочного распределения

Пусть $x_1, x_2, ..., x_n$ - выборка объема *n* из генеральной совокупности с функцией распределения $F_x(x)$. Рассмотрим выборочное распределение, как распределение ДСВ, принимающей значения $x_1, x_2, ..., x_n$ с вероятностями, равными относительным частотам n_i^* / n, и определим соответственно числовые характеристики этого эмпирического распределения.

Следует отметить, что выборочные числовые характеристики являются характеристиками данной выборки, но не являются характеристиками рассматриваемого распределения генеральной совокупности.

Для негруппированной выборки *n* выборочное среднее (математическое ожидание) $\overline{x_в}$ и выборочная дисперсия $D_в$, определяются выражениями

$$m_x^* = \overline{x_в} = \frac{1}{n}\sum_{j=1}^{n} x_j, \quad D_x^* = D_в = \frac{1}{n}\sum_{j=1}^{n}\left(x_j - \overline{x_в}\right)^2 = \frac{1}{n}\left(\sum_{j=1}^{n} x_j^2 - n\overline{x}^2\right).$$

Для сгруппированной выборки

$$\overline{x_в} = \frac{1}{n}\left(\sum_{i=1}^{n} n_i x_i\right), \quad D_в = \frac{1}{n}\left(\sum_{i=1}^{n} n_i \left(x_i - \overline{x_в}\right)^2\right) = \overline{x^2} - \left[\overline{x}\right]^2 = \frac{\sum_{i=1}^{n} n_i x_i^2}{n} - \left[\frac{\sum_{i=1}^{n} n_i x_i}{n}\right]^2.$$

Аналогично можно подсчитать начальные и центральные моменты выборочного распределения любых порядков.

Пример. В течение недели регистрировались разладки в работе 25 однотипных станков, потребовавшие кратковременной остановки их для регулирования. В результате регистрации числа разладок каждого из станков получили статистические данные: 4, 1, 3, 2, 0, 0, 0, 1, 3, 2, 1, 2, 2, 2, 3, 5, 3, 4, 2, 1, 2, 1, 3, 1, 2.

а) Составить статистический ряд распределения относительных частот наблюдаемых значений Д.С.В. x - числа разладок станков.

б) Построить полигон относительных частот.

в) Найти эмпирическую функцию распределения и построить её график.

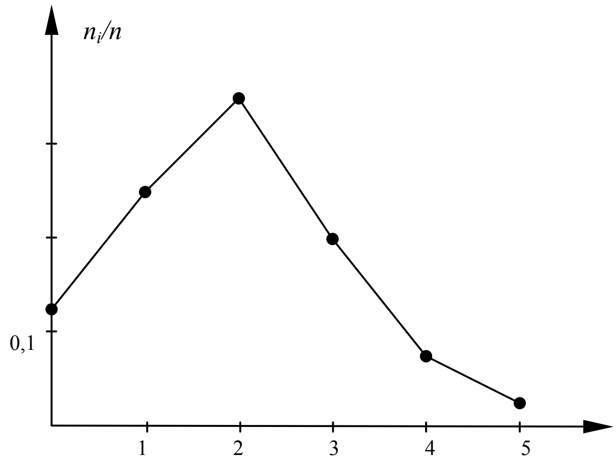
г) Вычислить числовые характеристики выборки: выборочное среднее, выборочное среднеквадратическое отклонение, выборочные коэффициенты асимметрии и эксцесса.

Решение.

Составим статистический рад распределения относительных частот

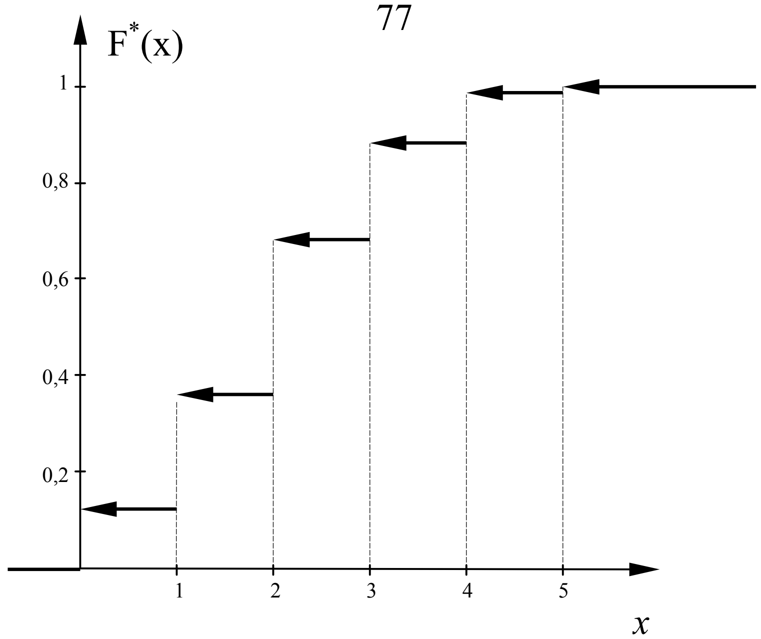
X	0	1	2	3	4	5
n_i^*	3	6	8	5	2	1
n_i^* / n	0,12	0,24	0,32	0,20	0,08	0,04
$\sum_{j=1}^{n} n_i^* / n$	0,12	0,36	0,68	0,88	0,96	1,00

а)



б) эмпирическая функция распределения $F^*(x)$ являет x накопленной относительной частотой

$$F^*(x) = \begin{cases} 0 & x \le 0; \\ 0{,}12 & 0 < x \le 1; \\ 0{,}36 & 1 < x \le 2; \\ 0{,}68 & 2 < x \le 3; \\ 0{,}88 & 3 < x \le 4; \\ 0{,}96 & 4 < x \le 5; \\ 1 & x > 5. \end{cases}$$



Вычисления числовых характеристик выборки сведены в таблицу

x_i	n_i^*	$x_i n_i^*$	$x_i^2 n_i^*$	$x_i^3 n_i^*$	$x_i^4 n_i^*$
0	3	0	0	0	0
1	6	6	6	6	6
2	8	16	32	64	128
3	5	15	45	135	405
4	2	8	32	128	512
5	1	5	25	125	625
$\sum$	25	50	140	458	1676

Вычислим четыре первых начальных эмпирических момента:

$$\nu_{1в} = \frac{\sum x_i n_i^*}{n} = \frac{50}{25} = 2; \quad \nu_{2в} = \frac{\sum x_i^2 n_i^*}{n} = \frac{140}{25} = 5{,}6; \quad \nu_{3в} = \frac{\sum x_i^3 n_i^*}{n} = \frac{458}{25} = 18{,}32;$$

$$\nu_{4в} = \frac{\sum x_i^4 n_i^*}{n} = \frac{1676}{25} = 67{,}04.$$

Вычислим второй, третий и четвертый эмпирические центральные моменты выборки

$$\mu_{2в} = \nu_{2в} - \nu_{1в}^2 = 5{,}6 - 4 = 1{,}6 \quad \mu_{3в} = \nu_{3в} - 3\nu_{1в}\nu_{2в} + 2\nu_{1в}^3 = 18{,}32 - 3*2*5{,}6 + 2*8 = 0{,}72,$$

$$\mu_{4в} = \nu_{4в} - 4\nu_{3в}\nu_{1в} + 6\nu_{2в}\nu_{1в}^2 - 3\nu_{1в}^4 = 67{,}04 - 4*18{,}32*2 + 65{,}64 - 3*16 = 6{,}88.$$

Вычислим числовые характеристики выборки: выборочные средние числа разладок станков $\overline{x} = \nu_{1в} = 2$; выборочную дисперсия числа разладок $D_в = \mu_{2в} = 1{,}6$; выборочное средне квадратическое отклонение числа разладок $\sigma_в = \sqrt{D_в} = 1{,}26$;

выборочный коэффициент эксцесса $Э_в = \dfrac{\mu_{4в}}{\sigma^4} - 4 = \dfrac{6{,}88}{(1{,}26)^4} - 3 = -0{,}31$.

5.2.3. Статистическое описание и выборочные характеристики двумерного случайного вектора

Пусть (x_i,y_i), $i = 1,2,\dots,n$, - выборка объема n из наблюдений случайного двумерного вектора (x,y). Предварительное представление о двумерной генеральной совокупности можно получить, изображая элементы выборки точками на плоскости с выбранной декартовой прямоугольной системой координат. Это представление выборки называется диаграммой рассеивания.

Распределением двумерной выборки называется распределение двумерного дискретного случайного вектора, принимающего значения (x_i,y_i), $i = 1,2,\dots,n$ с вероятностями, равными $1/n$. Выборочные числовые характеристики вычисляются, как соответствующие числовые характеристики двумерного случайного вектора дискретного типа.

Вычисление указанных выборочных характеристик удобно выполнять в следующей последовательности: сначала вычисляют суммы

$$\sum x_i, \sum y_i, \sum x_i^2, \sum y_i^2, \sum x_i y_i, \sum (x_i + y_i)^2 .$$

Для контроля правильности вычисления используется тождество

$$\sum (x_i + y_i)^2 \equiv \sum x_i^2 + 2\sum x_i y_i + \sum y_i^2 .$$

Выборочные средства находятся по формулам

$$\overline{x_b} = 1/n\sum x_i , \qquad \overline{y_b} = 1/n\sum y_i .$$

Затем вычисляются суммы квадратов отклонений от средних

$$Q_x = \sum (x_i - \overline{x})^2 = \sum x_i^2 - \frac{\left(\sum x_i\right)^2}{n}. \quad Q_y = \sum (y_i - \overline{y})^2 = \sum y_i^2 - \frac{\left(\sum y_i\right)^2}{n},$$

$$Q_{xy} = \sum (x_i - \overline{x})(y_i - \overline{y}) = \sum x_i y_i - \frac{\left(\sum x_i\right)\left(\sum y_i\right)}{n}.$$

Отсюда $\quad D_{xв} = \frac{1}{n}Q_x, \quad D_{yв} = \frac{1}{n}Q_y, \quad r_в = \frac{Q_{xy}}{\sqrt{Q_x Q_y}}.$

Выборочная линейная регрессия y на x по выборке (x_i,y_i), $i = 1,2,\dots,n$ определяется уравнением $\quad y = \beta_0^* + \beta_1^* x = \overline{y}_в + r_в \frac{D_{yв}}{D_{xв}}\left(x - \overline{x}_в\right)$

Коэффициенты называются выборочными коэффициентами регрессии. Они определяются формулами $\quad \beta_1^* = \frac{n\sum x_i y_i - \left(\sum x_i \sum y_i\right)}{n\sum x_i^2 - \left(\sum x_i\right)^2} = \frac{Q_{xy}}{Q_x}, \quad \beta_0^* = \overline{y}_в - \beta_1 \overline{x}_в .$

Аналогично определяется выборочная линейная регрессия x на y:

$$x = \beta_0^{'*} + \beta_1^{'*} y = \overline{x} + r_в \frac{D_{xв}}{D_{yв}}\left(y - \overline{y}_в\right),$$

где $\quad \beta_1^* = \frac{n\sum x_i y_i - \left(\sum x_i \sum y_i\right)}{n\sum y_i^2 - \left(\sum y_i\right)^2} = \frac{Q_{xy}}{Q_y}, \quad \beta_0^* = \overline{x}_в - \beta_1^{'*} \overline{y}_в .$

Для контроля правильности расчетов используют соотношение

$$\sqrt{\beta_1^* \beta_1^{'*}} = |r_в| .$$

Прямые $y = \beta_0^* + \beta_1^* x$ и $x = \beta_0^{'*} + \beta_1^{'*} y$ пересекаются в точке с координатами $\left(\overline{x}, \overline{y}\right)$.

ЗАДАНИЯ

5.1. Чем занимается математическая статистика, и каковы ее основные задачи?

5.2. Что называется генеральной совокупностью? выборочной совокупностью? Укажите применимость выборочного метода исследования.

5.3. Как связан объем выборки с надежностью выводов об изучаемом явлении?

5.4. Укажите способы группировки статистических данных. Как устанавливаются число, величина и положение разрядов интервальной группировки?

5.5. Что называется эмпирической функцией распределения? Перечислите ее свойства.

5.6. Какие типы графиков применяются для изображения дискретных статистических рядов? интервальных статистических рядов?

В задачах 5.7 – 5.10 построить графики эмпирических функций распределения, гистограммы и полигоны частот для выборок, представленных статистическими рядами.

5.7.

z_i	15	16	17	18	19
n_i	1	4	5	4	2

5.8.

z_i	2	3	4	5	6	7	8
n_i	1	3	4	6	5	2	1

5.9.

Границы интервалов	10-20	20-30	30-40	40-50	50-60	60-70	70-80
Частоты	1	2	7	18	12	8	2

5.10.

Границы интервалов	18-20	20-22	22-24	24-26	26-28	28-30	30-32	32-34
Частоты	4	3	3	2	4	7	12	5

В задачах 5.11 – 5.13 представить графические данные статистического ряда, найти эмпирическую функцию распределения и построить её график, вычислить основные числовые характеристики выборки.

5.11. Ряд распределения прочности при разрыве осевой стали имеет вид

Предел прочности (кГ/мм²)	48	50	52	54	56	58	60	62	64
Частота	3	17	75	194	228	155	91	23	5

Ответ: $\bar{x}$=56,048; σ=2,7546; A=+0,131; Э=-0,097; v=4,91%.

5.12. Ряд распределения ударной вязкости стали

Варианты (в кГ/мм²)	4,5	5,5	6,5	7,5	8,5	9,5	10,5	11,5	12,5
Частота	2	10	92	168	214	132	103	53	16

Ответ: $\bar{x}$=8,715; σ=1,5496; A=+0,297; Э=-0,336; v=17,78%.

5.13. Измерялась степень осаждения гидроксида железа (в %) в присутствии хрома. Результаты эксперимента следующие:

94,89; 95,69; 98,18; 98,54; 97,08; 93,80; 94,89; 95,62.

Ответ: $\bar{x}$=96,086; σ=1,6842; A=0,287; Э=-1,42; v=1,8%.

5.14. Каждым сварщиком было изготовлено по 100 опытных образцов из одних и тех же исходных материалов. Эти образцы испытывались на растяжение с целью выяснения механической прочности выполненного шва. Результаты испытаний приведены в таблице:

Предел прочности (кГ/мм²)	28	30	32	34	36	38	40	42	44
Частота									
I	-	-	-	2	3	30	40	20	5
II	-	3	5	30	35	20	5	2	-
III	8	15	15	12	15	20	10	5	-

Оценить квалификацию каждого специалиста и выяснить способность сохранять постоянство качества своей работы.

Ответ: Наилучшей квалификацией обладает первый сварщик. У этого же сварщика наблюдается и наибольшее постоянство качества работы.

В задачах 5.15 – 5.17 вычислить коэффициенты корреляции, определить и нанести на диаграмму рассеивания прямые регрессии Y на X и X на Y по данным выборкам.

5.15.

x	8	10	5	8	9
y	1	3	1	2	3

Ответ: r≈0,535, y=-1,43+0,43x, x=5+1,5y.

5.16.

x	9	10	12	5
y	6	4	7	3

Ответ: r≈0,806, y=0,5+0,5x, x=2,5+1,3y.

5.17.

x	10	2	7	5
y	8	2	6	4

Ответ: r≈0,983, y=0,5+0,77x, x=-0,5+1,3y.

5.18. Предел выносливости стали при изгибе Y (Н/мм²) оценивается на основании другой её характеристики – предел упругости при кручении X (Н/мм²). По опытным данным для 12 марок стали найти уравнения линейной регрессии Y на X и X на Y и вычислить коэффициент корреляции между этими характеристиками. Результаты измерений

x	51	67	84	81	101	109	71	97	109	51	105	89
y	25	30	43	44	57	58	43	46	62	45	55	45

$\sum x_i = 1015, \sum y_i = 553, \sum x_i^2 = 90667, \sum y_i^2 = 26807, \sum x_i y_i = 48888,$

Ответ: r≈0,837, y=8,96+0,44x, x=10,96+1,6y.

5.19. По данным измерения двух переменных

x	66	70	75	80	82	85	90	92	95	98
y	60	78	65	87	74	70	78	95	88	90

вычислить коэффициент корреляции и найти уравнение линейной регрессии Y на X.

Ответ: r≈0,743, y=12,25+0,8x.

5.3. Статистическая оценка неизвестных параметров распределения

Предположим, что для оценки распределения исследуемой СВ *X* из генеральной совокупности с неизвестной функцией распределения *F(x)* извлечена выборка $x_1, x_2, ..., x_n$. Предположим, что экспериментатор визуально по виду гистограммы или полигона частостей или исходя из каких-либо других соображений выбрал класс Ω функций определенного вида (нормальных, показательных, биноминальных и т.д.), к которому может принадлежать функция распределения вероятностей исследуемой СВ *X*.

Наиболее часто при исследовании непрерывных случайных величин экспериментатор старается выбрать класс нормальных функций, т.е. построить нормальную модель генеральной совокупности, так как эта модель наиболее разработана. Кроме того, поскольку нормальный закон является предельным законом многих законов распределения, то класс нормальных функций часто можно принимать за приближенную грубую модель генеральной совокупности.

После того, как класс функций Ω выбран, производится оценка (подгонка) параметров внутри выбранного класса функций. Например, если выбран нормальный класс функций для описания исследуемой случайной величины *X*, то по выборке $x_1, x_2, ..., x_n$ требуется оценить два параметра – математическое ожидание a и среднее квадратическое отклонение σ, от которого зависит нормальное распределение.

Если выбран класс функций Пуассона, то на основании выборки $x_1, x_2, ..., x_n$ требуется оценить только один параметр λ, которым определяется закон Пуассона: $P_n(x) = \dfrac{\lambda^x e^{-\lambda}}{x!}$.

Пусть из генеральной совокупности с функцией распределения $F(x,\theta)$, где θ - неизвестный параметр, произведена выборка объема n и получены результаты $x_1, x_2, ..., x_n$. Вообще говоря, по результатам выборки, какого бы большого объема она ни была, нельзя определить точное значение неизвестного параметра θ, а можно найти его приближенное значение θ, которое и называется оценкой.

Для нахождения приближенных значений (оценки) неизвестного параметра θ будем рассматривать функции вида $\hat{\theta}=u\ (x_1, x_2, ..., x_n)$, которые называются выборочными функциями или статистиками. Задача оценки неизвестного параметра θ сводится к нахождению таких выборочных функций $\hat{\theta}=u\ (x_1, x_2, ..., x_n)$, которые можно использовать в качестве оценки неизвестного параметра θ.

Любая выборка является конечной и случайной. Следовательно, все выборочные функции $\hat{\theta}=u\ (x_1, x_2, ..., x_n)$ также являются случайными.

Таким образом, будем рассматривать оценку $\hat{\theta}=u(x_1, x_2, ..., x_n)$ неизвестного параметра θ как случайную величину, а ее значение, вычисленное по данной выборке объема n, как одну реализацию случайной величины, т.е. как одно из множества возможных значений этой случайной величины.

Оценки параметров подразделяются на точечные и интервальные. ***Точечная оценка*** параметра $\hat{\theta}$ определяется одним числом $\theta = u\ (x_1, x_2, ..., x_n)$.

Интервальной оценкой называют оценку, которая определяется двумя числами $\hat{\theta}_1, \hat{\theta}_2$ - концами интервала, накрывающего оцениваемый параметр θ.

Сформулируем основные свойства, которые должны иметь «хорошие» оценки неизвестного параметра $\hat{\theta}=u\ (x_1, x_2, ..., x_n)$.

Прежде всего с точки зрения точности и надежности оценок желательно, чтобы найденные на основании выборочных функций $\hat{\theta}=u\ (x_1, x_2, ..., x_n)$ оценки неизвестных параметров по возможности были тесно сконцентрированы около значений оцениваемых параметров, другими словами, чтобы рассеивание случайной величины $\hat{\theta}$ около θ было по возможности наименьшим.

Определение 1. *Оценка $\hat{\theta}=u\ (x_1, x_2, ..., x_n)$ называется состоятельной, если при увеличении числа измерений оценка сходится по вероятности к оцениваемому параметру, т.е. если* $\lim\limits_{n\to\infty} P(|\theta - \widehat{\theta}| < \varepsilon) = 1$.

Требование состоятельности гарантирует от грубых ошибок в определении θ при достаточно больших *n*.

Определение 2. *Оценка* $\hat{\theta}=u\ (x_1, x_2, ..., x_n)$ *называется несмещенной (оценкой без систематической ошибки), если ее математическое ожидание равно оцениваемому параметру, т.е. если* $M(\theta)=\theta.$

Если условие не выполняется, то оценка называется смещенной (содержащей систематическую ошибку).

Определение 3. *Оценка* $\hat{\theta}=u(x_1, x_2, ..., x_n)$*, которая обладает наименьшей дисперсией, называется эффективной* ($\lim\limits_{n\to\infty} D(\hat{\theta})=0$).

Пример. Покажем, что относительная частота *m/n* появления события *A* при *n* испытаниях Бернулли является состоятельной, несмещенной и эффективной оценкой вероятности *p*.

Состоятельность оценки следует из теоремы Бернулли.

Определим математическое ожидание относительной частоты, полагая *m=X*, т.е. появления события *A* в *n* испытаниях Бернулли рассматриваем как случайную величину *X*, распределенную по биноминальному закону

$$M\left(\frac{X}{n}\right)=M(X)/n=np/n=p.$$

Так как $M\left(\frac{m}{n}\right)=p$, то относительная частота является несмещенной оценкой вероятности *p*. Определим дисперсию относительной частоты:

$$D\left(\frac{X}{n}\right)=D(X)/n^2=np(1-p)/n^2=p(1-p)/n.\ \text{Тогда}\ \lim_{n\to\infty} D\left(\frac{X}{n}\right)=0.$$

Ниже приведем основные методы нахождения точечных оценок неизвестных параметров и укажем свойства таких оценок.

5.3.1. Метод моментов

Пусть по результатам выборки $x_1, x_2, ..., x_n$ объема n, извлеченной из генеральной совокупности с функцией распределения *F(x,θ),* требуется оценить неизвестный параметр θ этого распределения.

По аналогии с моментами случайной величины *X* введем понятие эмпирических моментов. Эмпирические начальные моменты порядка *k* определяются следующей формулой: $\nu_k^*=\frac{1}{n}\sum\limits_i x_i^k$.

В частности заметим, что начальный теоретический момент первого порядка является математическим ожиданием случайной величины *X*, а эмпирический начальный момент первого порядка является средним арифметическим значением наблюденных значений случайной величины *X*, т.е. $M(x) \cong \overline{x} = \frac{1}{n}\sum_i x_i$.

Аналогично эмпирические центральные моменты порядка k определяются формулой $\mu_k^* = \sum_i (x_i - \overline{x})^k / n$.

Теоретический центральный момент второго порядка является дисперсией случайной величины *X*, т.е. $D(x) \approx \sum_i (x_i - \overline{x})^2 / n$.

Следует иметь в виду, что *эмпирические моменты являются случайными величинами, в то время как теоретические моменты являются фиксированными постоянными величинами*.

Наиболее простым является метод оценивания, называемый *методом моментов*, предложенный английским статистиком Карлом Пирсоном в 1894г. Метод моментов основывается на том, что *эмпирические моменты (или их функции) принимаются за оценки соответствующих теоретических моментов (или их функций) и параметры выражаются через эти моменты*.

Таким образом, приравнивая теоретический начальный момент первого порядка к эмпирическому начальному моменту первого порядка, приходим к выводу, что оценкой математического ожидания случайной величины *X*, распределенной по любому закону, является средняя арифметическая наблюденных значений случайной величины *X*, т.е.

$$M(x) \cong \overline{x} = \frac{1}{n}\sum_i x_i .$$

Приравнивая теоретический и эмпирический центральные моменты второго порядка, приходим к выводу, что оценка дисперсии случайной величины *X*, распределенной по любому закону, находится по формуле

$$D(x) \approx \sum_i (x_i - \overline{x})^2 / n .$$

Поступая аналогичным образом, можно найти оценки теоретических моментов любого порядка.

Пример. Пусть случайная величина *X* распределена по нормальному закону с параметрами $M(X)=a$ и $\sigma = \sqrt{D(X)}$ или, более кратко, пусть $X \to N(a,\sigma)$. Требуется по результатам наблюдений $(x_1, x_2, ..., x_n)$ оценить параметры a, σ и найти оценки асимметрии и эксцесса.

Решение. Применяя метод моментов, получим

$$\begin{cases} \nu_1 = \nu_1^* \Rightarrow \hat{a} = \sum_i x_i / n; \\ \mu_2 = \mu_2^* \Rightarrow \hat{\sigma}^2 = s^2 = \sum_i (x_i - \overline{x})^2 / n. \end{cases}$$

Среднее арифметическое наблюдаемых значений является состоятельной, несмещенной и эффективной оценкой математического ожидания.

Оценка дисперсии $D(x) \approx \sum_i (x_i - \bar{x})^2 / n = s^2$ - является смещенной, а несмещенной оценкой (при малом объеме выборки n<30) - $s^{*2} = \dfrac{\sum_i (x_i - \bar{x})^2}{n-1}$.

Оценки асимметрии и эксцесса, являющиеся функциями центральных моментов, находятся по формулам

$$\begin{cases} \dfrac{\mu_3}{\sigma^3} = (\dfrac{\mu_3}{\sigma^3})^* \Rightarrow \widehat{A} = \dfrac{\sum_i (x_i - \bar{x})^3}{ns^3}; \\ \dfrac{\mu_4}{\sigma^4} - 3 = (\dfrac{\mu_4}{\sigma^4})^* - 3 \Rightarrow \widehat{E} = \dfrac{\sum_i (x_i - \bar{x})^4}{ns^4} - 3. \end{cases}$$

Пример. Случайная величина X распределена по закону Пуассона $P(X = k) = \dfrac{\lambda^k e^{-\lambda}}{k!}$, где $\theta = \lambda$ - неизвестный параметр, который надо оценить по данным выборки. Требуется по результатам наблюдений $x_1, x_2, ..., x_n$ оценить параметр θ.

Решение. Начальный теоретический момент первого порядка $\nu_1 = \lambda$.

Эмпирический начальный момент первого порядка равен $\bar{x} = \sum_i x_i / n$. Следовательно, $\widehat{\lambda} = \sum_i x_i / n$.

*Замечание***.** Метод моментов отличается простотой, однако, оценки, найденные этим методом, как правило, являются смещенными и малоэффективными, т.е. не являются наилучшими из возможных.

Исключение представляет лишь нормальное распределение, при котором метод моментов дает эффективные и состоятельные оценки $\bar{x}$ и s параметров a и σ.

5.3.2. Метод наибольшего правдоподобия

Метод наибольшего правдоподобия является широко распространенным методом получения точечной оценки. Он предложен в 1912 г. английским статистиком Р.Фишером.

Пусть из генеральной совокупности с плотностью распределения вероятностей *f(x, θ)* произведена выборка объема n и получены результаты $x_1, x_2, ..., x_n$. Предположим вначале, что Х –дискретная случайная величина, закон распределения которой зависит от неизвестного параметра θ. Например, можно предположить, что случайная величина X распределена по закону Пуассона $P(X = k) = \dfrac{\lambda^k e^{-\lambda}}{k!}$, где $\theta = \lambda$ - неизвестный параметр, который надо оценить по данным выборки. Будем рассматривать результаты выборки как реализацию

n-мерной случайной величины $(X_1, X_2, ..., X_n)$. Предположим далее, что составляющие этой случайной величины независимы. В этом случае вероятность того, что составляющие примут значения, равные наблюденным значениям, называемая функцией правдоподобия, равна

$$L = P\{X_1 = x_1, X_2 = x_2, ..., X_n = x_n\} = \prod_{i=1}^{n} P(x_i, \theta).$$

В случае непрерывной случайной величины функция правдоподобия имеет вид $L = f(x_1, x_2, ..., x_n) = \prod_{i=1}^{n} f(x_i, \theta)$. Формула определяет плотность распределения вероятностей непрерывной случайной величины $(X_1, X_2, ..., X_n)$ или плотность распределения выборки.

В качестве оценки неизвестного параметра θ, найденной по методу наибольшего правдоподобия, *выбирается такая функция* $\hat{\theta} = u\ (x_1, x_2, ..., x_n)$ *, которая максимизирует функцию правдоподобия*. Следовательно, для нахождения оценок наибольшего правдоподобия составляется система m уравнений (m – число оцениваемых параметров)

$$\frac{\partial L}{\partial \theta_i} = 0, \quad (i = 1, 2, ..., m)$$

и выбирается то решение, которое обращает функцию правдоподобия в максимум. Поскольку экстремум функции L и $\ln L$ достигается при одних и тех же значениях $\hat{\theta} = u\ (x_1, x_2, ..., x_n)$, то иногда для упрощения расчетов пользуются логарифмической функцией правдоподобия. В этом случае оценки наибольшего правдоподобия находятся из системы уравнений

$$\frac{\partial \ln L}{\partial \theta_i} = 0, \quad (i = 1, 2, ..., m).$$

Пример. Пусть случайная величина X распределена по закону Пуассона $P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}$, где k=0, 1, 2,…. Требуется по результатам наблюденных значений $x_1 = k_1, x_2 = k_2, ..., x_n = k_n$ оценить неизвестный параметр λ этого распределения.

Решение. Запишем функцию правдоподобия

$$L = \frac{\lambda^{x_1}}{x_1!} e^{-\lambda} \frac{\lambda^{x_2}}{x_2!} e^{-\lambda} ... \frac{\lambda^{x_n}}{x_n!} e^{-\lambda} = \prod_{i=1}^{n} \frac{\lambda^{x_i}}{x_i!} e^{-\lambda}.$$

Для нахождения оценки λ, т.е. такого значения $\lambda = \lambda(x_1, x_2, ..., x_n)$, при котором функция правдоподобия обращается в максимум, удобно перейти к логарифмической функции правдоподобия

$$\ln L = -n\lambda + \ln \lambda \cdot \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} \ln x_i!$$

Следовательно, $\dfrac{\partial \ln L}{\partial \lambda} = -n + \sum_{i=1}^{n} x_i / \lambda$. Приравнивая производную к нулю, имеем $\hat{\lambda} = \sum_{i=1}^{n} x_i / n = \bar{x}$.

Пример. Случайная величина X распределена по нормальному закону. Требуется по результатам наблюденных значений $x_1, x_2, ..., x_n$ этой случайной величины оценить параметры a и σ нормального закона.

Решение. Плотность распределения вероятностей нормально распределенной случайной величины равна $f(x, a, \sigma) = \dfrac{1}{\sigma\sqrt{2\pi}} \exp(-\dfrac{(x-a)^2}{2\sigma^2})$.

Следовательно, функция правдоподобия имеет вид

$$L = \frac{1}{(\sigma\sqrt{2\pi})^n} \exp(-\frac{\sum_{i=1}^{n}(x_i - a)^2}{2\sigma^2}).$$

Записывая логарифмическую функцию правдоподобия и дифференцируя ее по a и σ, получим $\hat{a} = \bar{x} = \sum_{i=1}^{n} x_i / n$; $\hat{\sigma}^2 = s^2 = \sum (x_i - \bar{x})^2 / n$.

Эти оценки совпали с оценками, полученными методом моментов.

Метод наибольшего правдоподобия *обладает рядом преимуществ по сравнению с методом моментов*. Укажем некоторые важные свойства оценок наибольшего правдоподобия.

1. Метод наибольшего правдоподобия дает состоятельные оценки.
2. Если существует эффективная оценка, то метод наибольшего правдоподобия дает эту оценку.
3. Оценки наибольшего правдоподобия асимптотически эффективны.

5.4. Метод наименьших квадратов

Метод наименьших квадратов (МНК) нашел широкое применение как для оценки неизвестных параметров, так и для обработки экспериментальных данных (их аппроксимации).

Рассмотрим следующую задачу. Пусть в распоряжении экспериментатора имеются только наблюденные значения – точки (x_i, y_i) $(i = 1, 2, \cdots, n)$ или, более кратко, имеется выборка объема *n*, по которой нужно оценить зависимость между *x* и *y* (записать в виде формулы). Нужно найти такую функцию $f(x, \theta)$, которая наилучшим образом (в смысле метода наименьших квадратов) приближает наши экспериментальные данные. Общий вид функции $f(x, \theta)$ может быть известен из сущности исходной задачи либо подобран из каких-либо других соображений (графических, визуальных, умозрительных, опыта предыдущих исследований и т.п.). Оценка неизвестного параметра θ будет наилучшей в смысле МНК в том смысле, что сумма квадратов отклонений наблюденных значений переменной *y* от соответствующих значений функции $f(x,\theta)$ будет минимальной, т.е. параметр находится из условия

$$\sum_{i=1}^{n}[y_i - f(x_i,\theta)]^2 - \min .$$

Рассмотрим методику МНК в простейшем случае – линейной функции, т.е. $f(x,\theta) = ax + b$, где a и b неизвестны. Найдем эти коэффициенты методом наименьших квадратов, т.е. составим функцию $F(x,a,b) = \sum_{i=1}^{n}[y_i - (ax_i + b)]^2$ и найдем ее минимум. Для этого находим частные производные $\frac{\partial F}{\partial a}$ *и* $\frac{\partial F}{\partial b}$, приравняем их к нулю. Проведя элементарные преобразования, получим систему нормальных линейных уравнений

$$\begin{aligned} &a\sum_{i=1}^{n}x_i^2 + b\sum_{i=1}^{n}x_i = \sum_{i=1}^{n}x_i y_i, \\ &a\sum_{i=1}^{n}x_i + bn = \sum_{i=1}^{n}y_i. \end{aligned} \qquad (*)$$

Решая эту систему, находим a и b.

Приведем пример, характеризующий и вычислительную сторону МНК.

Пример. Найти зависимость между x и y (линейную) по четырем парам наблюденных данных

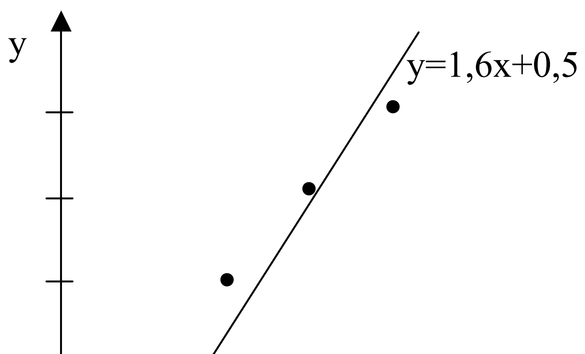
x_i	1	2	3	4
y_i	2	4	5	7

Решение. Если на график нанести точки (x_i, y_i), то видно, что они группируются вокруг прямой линии. Поэтому будем подбирать функцию линейного вида $f(x,\theta) = ax + b$. Вычислим необходимые суммы, входящие в систему (*). Для удобства все вычисления расположим в таблице:

x_i	y_i	$x_i\, y_i$	$x^2{}_i$
1	2	2	1
2	4	8	4
3	5	15	9
4	7	28	16
$\sum$ 10	18	53	30

Подставляя найденные суммы в систему (*), получаем 30*а*+10*в*=53, 10a+4в=18.

Решая систему, находим *а*=1,6 и *в*=0,5, т.е. *у*=1,6*х*+0,5.



Полученное уравнение y=ax+b можно интерпретировать как уравнение прямой регрессии Y на X:

$$y - \overline{y} = \frac{s_y}{s_x} r_в (x - \overline{x}),$$

что удобней с вычислительной точки зрения. Действительно для того чтобы записать уравнение регрессии Y на X, достаточно найти точечные оценки параметров двумерной случайной величины (X,Y): $\overline{x}, \overline{y}, s_x, s_y, r_в$.

5.5. Интервальные оценки параметров

При статистической обработке результатов наблюдений часто необходимо не только найти оценку $\hat{\theta}$ неизвестного параметра θ, но и охарактеризовать точность этой оценки. В связи с этим во многих случаях более выгодно пользоваться интервальной оценкой, основанной на определении некоторого интервала, внутри которого с определенной вероятностью находится неизвестное значение параметра θ.

Доверительным интервалом для параметра θ будем называть интервал (θ_1, θ_2) накрывающий истинное значение θ с заданной вероятностью $P = 1 - \alpha$, т.е.

$$P[\theta_1 < \theta < \theta_2] = 1 - \alpha .$$

Число называется $P = 1 - \alpha$ ***доверительной вероятностью***, а значение α - ***уровнем значимости***.

Статистики $\theta_1 = \theta_1(x_1, x_2, ..., x_n)$ *и* $\theta_2 = \theta_2(x_1, x_2, ..., x_n)$*,* определяемые по выборке $x_1, x_2, ..., x_n$ из генеральной совокупности с неизвестным параметром θ, *называются соответственно нижней и верхней границами доверительного интервала.*

Пусть
$$P\left[\left|\theta - \hat{\theta}\right| < \mathrm{E}\right] = P\left[\hat{\theta} - \mathrm{E} < \theta < \hat{\theta} + \mathrm{E}\right] = 1 - \alpha ,$$

где E - характеризует точность оценки.

Длина доверительного интервала, характеризующая точность интервального оценивания, зависит от объема выборки n и доверительной вероятности 1-α; при увеличении объема выборки длина доверительного интервала уменьшается, а с приближением доверительной вероятности
к единице - увеличивается. Выбор доверительной вероятности определяется конкретными условиями, обычно используются значения 1- α, равные 0,90; 0,95; 0,99.

Чтобы найти доверительный интервал для параметра θ, необходимо знать закон распределения статистики $\hat{\theta} = \hat{\theta}(x_1, x_2, ..., x_n)$, значение которой является оценкой параметра θ. При этом для получения доверительного интервала наименьшей длины при данном объеме выборки и заданной доверительной вероятности (1-α) в качестве оценки $\hat{\theta}$ параметра θ следует брать эффективную либо асимптотически эффективную оценку.

Один из методов построения доверительных интервалов состоит в следующем: пусть существует статистика $Y = Y(\hat{\theta}, \theta)$ такая, что:

а) закон распределения Y известен и не зависит от θ;

б) функция $Y(\hat{\theta}, \theta)$ непрерывна и строго монотонна по θ.

Пусть далее (1-α) – заданная доверительная вероятность, а $Y_{\alpha/2}$ и $Y_{1-\alpha/2}$ - квантили распределения статистики Y порядков $\alpha/2$ и $(1-\alpha/2)$) соответственно. Тогда с вероятностью (1-α) выполняется неравенство

$$Y_{\alpha/2} < Y(\hat{\theta}, \theta) < Y_{1-J/2}$$

Решая это неравенство относительно θ, найдем границы θ_1 и θ_2 доверительного интервала для θ.

Общая схема построения доверительных интервалов состоит в следующем.

1. Из генеральной совокупности с функцией распределения *F(x,θ)* извлекается выборка объема *n*. По результатам этой выборки методом моментов или методом наибольшего правдоподобия находится точечная оценка оцениваемого параметра *θ*.

2. Составляется случайная величина $Y(\theta)$ связанная с параметром θ и имеющая известную плотность распределения вероятностей $f(y, \theta)$.

3. Задается доверительная вероятность (1-α).

4. Используя плотность распределения вероятностей случайной величины *Y*, находят такие два числа θ_1 и θ_2, что

$$P\left(\hat{\theta}_1 < Y(\theta) < \hat{\theta}_2\right) = \int_{\theta_1}^{\theta_2} \alpha(y, \theta) ay = 1 - \alpha .$$

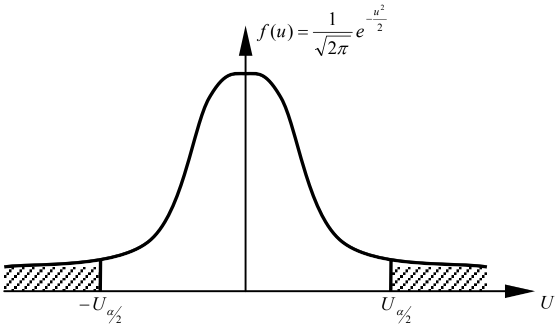
Значения $\hat{\theta}_1$ и $\hat{\theta}_2$ выбираются при условиях

$$P\left[Y(\theta) < \hat{\theta}_1\right] = \frac{\alpha}{2} \quad , \quad P\left[Y(\theta) < \hat{\theta}_2\right] = \frac{\alpha}{2} .$$

Используем указанную схему для нахождения доверительных интервалов параметров нормального закона a и σ.

Пусть $x_1, x_2, \ldots x_n$ выборка из нормально распределенной генеральной совокупности. Найдем доверительный интервал для математического ожидания *a* при условии, что дисперсия генеральной совокупности известна и равна δ_2, а доверительная вероятность равна *1-α*.

В качестве оценки математического ожидания *a* возьмем выборочное среднее $\bar{x} = \frac{1}{n}\sum x_i$. Выборочное среднее $\bar{x}$ в данном случае имеет нормальное распределение $N\left(a, \frac{\delta}{\sqrt{n}}\right)$.



Рассмотрим статистику $u = \frac{\bar{x} - a}{\sigma / \sqrt{n}}$, имеющую стандартизованное нормальное распределение *N(0,1)*.

Решая неравенство $-u_{\alpha/2} < \frac{\bar{x} - a}{\sigma / \sqrt{n}} < u_{\alpha/2}$ относительно *a*, получим, что с вероятностью *1-α* выполняется следующее условие:

$\bar{x}$ $< \bar{x} + u_{\alpha/2} \frac{\sigma}{\sqrt{n}}$.

Обычно квантили нормального распределения $u_{\alpha/2}$ находят по таблицам функции Лапласа из условия $\Phi(u_{\alpha/2}) = 1\text{-}\alpha$.

Для наиболее употребительных значений вероятности *(1-α)* квантили стандартизованного нормального распределения приведены в сокращенной таблице:

Доверительная вероятность *(1-α)*	Квантили $u_{\alpha/2}$
0,90	1,64
0,95	1,96
0,99	2,58
0,9973	3,00

Точность оценки математического ожидания (предельная погрешность)
$\Sigma = u_{\alpha/2} \frac{\sigma}{\sqrt{n}}$.

Пример 1. Пусть случайная величина *x* имеет нормальное распределение с известным средним квадратическим отклонением *δ=2*. Требуется найти доверительный интервал для математического ожидания *a*, если среднее арифметическое значение результатов *n=16* прямых равноточных измерений $\bar{x} = 20{,}09$. Доверительную вероятность *(1-α)* принять равной 0,90.

Решение. Пользуясь таблицами функции Лапласа по заданной вероятности *(1-α)=0,90* находим $u_{0,05}=1,64$. Найдем точность оценки $\Sigma = 1,64\frac{2}{\sqrt{16}} = 0,82$.

Следовательно, искомый доверительный интервал (20,09-0,82; 20,09+0,82) или (19,27; 20,91). Смысл полученного результата: если будет произведено достаточно большое число выборок данного объема, то в 90% из них доверительные интервалы накроют математическое ожидание.

Если дисперсия генеральной совокупности, имеющей нормальное распределение, неизвестна, а в качестве точечных оценок используются несмещенные оценки неизвестных параметров,

$$\hat{a} = \bar{x},$$

$$\hat{\sigma} = s = \sqrt{\frac{\sum\left(x_i - \bar{x}\right)^2}{n-1}},$$

то при доверительной вероятности *(1-α)* доверительный интервал для математического ожидания a имеет вид $\bar{x} - t_{\alpha/2;n-1}\frac{s}{\sqrt{n}} < a < \bar{x} + t_{\alpha/2;n-1}\frac{s}{\sqrt{n}}$, где $t_{\alpha/2;n-1}$ - квантиль распределения Стьюдента с (n-1) степенью свободы.

Пусть случайная величина x распределена по нормальному закону, причем математическое ожидание a и среднеквадратичное отклонение σ не известны. По выборке объема n найдем точечные несмещенные оценки неизвестных параметров:

$$\hat{a} = \bar{x};$$

$$\hat{\sigma} = s = \sqrt{\frac{\sum\left(x_i - \bar{x}\right)^2}{n-1}}.$$

Тогда доверительный интервал для дисперсии при доверительной вероятности *(1-α)* имеет вид $s\gamma_1 < \sigma < s\gamma_2$,

где $\gamma_1 = \sqrt{\frac{n-1}{x^2_{\alpha/2;n-1}}}$, $\gamma_2 = \sqrt{\frac{n-1}{x^2_{1-\alpha/2;n-1}}}$;

$x^2_{p;n}$ - квантиль распределения x^2 с n степенями свободы. Коэффициенты γ_1 и γ_2, соответствующие доверительной вероятности (1-α) и числу степеней свободы (n-1), помещены в табл. 4(приложения).

Пример 2. Пусть x_1=2?015; x_2=2,020; x_3=2,025; x_4=2,020; x_5=2,015 – результаты независимых равноточных измерений толщины пластинки. Требуется:

1) оценить с помощью доверительного интервала истинную толщину (математическое ожидание) пластинки *(1-α)=0,95*;

2) найти минимальное число измерений, которые надо выполнить, чтобы с надежностью *1-α=0,95* можно было утверждать, что предельная погрешность точной оценки истинной толщины пластинки не превышала *0,003*.

Решение. Будем считать результаты измерения наблюденными значениями случайной величины *x*, распределенной по нормальному закону с неизвестными параметрами α и δ.

Найдем точечные оценки этих параметров в таблице

Номер наблюдений	x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
1	2,015	-0,004	0,000016
2	2,020	+0,001	0,000001
3	2,025	+0,006	0,000036
4	2,020	+0,001	0,000001
5	2,015	-0,004	0,000016
Σ	10,095	0	0,000070

Следовательно,

$$\hat{a} = \bar{x} = \frac{10{,}095}{5} = 2{,}019;$$

$$\hat{\delta} = s_{несм.} = \sqrt{\frac{0{,}00007}{4}} = 0.0042.$$

По таблицам распределения Стьюдента (прил. 5) по доверительной вероятности *1-α=0,95* и числу степени свободы n-1=4 находим квантиль распределения $t_{0,025;4}=2{,}776$. Следовательно, предельная погрешность точечной оценки

$$\Sigma = t_{\alpha/2;n-1}\frac{s}{\sqrt{n}} = 2{,}776\frac{0{,}0042}{\sqrt{5}}0{,}0048.$$

Искомый доверительный интервал для математического ожидания

$$(\bar{x} - \Sigma < a < \bar{x} + \Sigma) = (2{,}0142 < a < 2{,}0238).$$

Для расчета минимального числа измерений, необходимых для определения истинной толщины пластинки с погрешностью, не превышающей 0,003, применим формулу

$$\Sigma = t^2_{\alpha/2;n-1}\frac{s}{\sqrt{n}}.$$

Следовательно,

$$n \ge \frac{t^2_{\alpha/2;n-1}s^2}{\Sigma^2} = \frac{(2{,}776)^2\,0{,}000018}{0{,}000009} \approx 16.$$

Таким образом, для того чтобы предельная погрешность *Σ* истинного значения толщины пластинки не превышала *Σ=0,003*, следует выполнить не менее
16 измерений.

Пример 3. Построить доверительный интервал, накрывающий среднее квадратическое отклонение δ с надежностью *1-α=0,98* по данным примера 2.

Решение. n=5; $\bar{x}=2{,}019$; $\bar{\delta}=s=0{,}0042$ По заданной доверительной вероятности *1-α=0,98* и числу степеней свободы n-1=4, пользуясь таблицами (прил.4), находим γ_1=0,55 и γ_2=3,67. Следовательно, искомый доверительный интервал равен $0{,}55\cdot 0{,}0042<\delta<3{,}67\cdot 0{,}0042$ или $0{,}0023<\delta<0{,}0154$.

ЗАДАНИЯ

5.20. Что называется оценкой неизвестного параметра распределения?

5.21. Какая оценка называется состоятельной? несмещенной? эффективной?

5.22. В чем сущность метода моментов нахождения точечных оценок? метода наибольшего правдоподобия?

5.23. По каким формулам находятся точечные оценки параметров нормального закона?

5.24. Что называется доверительным интервалом, доверительной вероятностью (надежностью)?

5.25. По каким формулам находятся доверительные интервалы для параметров нормального закона распределения?

5.26. Оценка величины сопротивления большой партии однотипных резисторов, определенная по результатам измерений 100 случайно отобранных экземпляров, равна $\bar{x}=10\text{кОм}$: а) считая, что дисперсия измерений известна: $\sigma^2=1\text{кОм}^2$, найти вероятность того, что для резисторов всей партии величина сопротивления лежит в пределах 10±0,1кОм; б) сколько измерений нужно произвести, чтобы с вероятностью 0,95 утверждать, что для всей партии резисторов величина сопротивления лежит в пределах 10±0,1кОм?

Ответ: а) 0,68; б) n≥385.

5.27. Для определения вертикального угла ориентира используют среднее арифметическое нескольких замеров угла при помощи секстанта. Для углов, измеряемых секстантом, среднеквадратическое отклонение принимается равным σ=1,5'. Найти количество замеров, которое нужно произвести, чтобы: а) погрешность результата с вероятностью 0,99 не превосходила 1'; б) погрешность результата с вероятностью 0,95 не превосходила $1{,}5^0$.

Ответ: а) n≥15; б) n≥4.Результаты 10 измерений емкости конденсатора прибором, не имеющим систематической ошибки, дали такие отклонения от номинала (пкФ): 5,4; -13,9; -11; 7,2; -15,6; 29,2; 1,4; -0,3; 6,6; -9,9. Найти 90%-ный доверительный интервал для дисперсии и среднего квадратичного отклонения.

5.28. Результаты 10 измерений ёмкости конденсатора прибором, не имеющим систематической ошибки, дали такие отклонения от номинала (пкФ): 5,4;-13,9; -11; 7,2; -15,6; 29,2; 1,4; -0,3; 6,6; -9,9. Найти 90%-ный доверительный интервал для дисперсии и среднего квадратичного отклонения.

Ответ: (96,81; 491,34), (9,84; 22,17).

5.29. Даны результаты пяти независимых равноточных измерений длины детали (в миллиметрах): 18,306, 18,305, 18,311, 18,309, 18,304. Предполагая, что результаты измерений распределены по нормальному закону, требуется: а) найти точечные оценки математического ожидания и среднего квадратического отклонения; б) найти доверительный интервал, покрывающий истинную длину детали с надежностью (1-α)=0,95, считая среднее квадратическое отклонение σ известным и равным несмещенной оценке S; в) найти доверительный интервал, покрывающий истинную длину детали с заданной надежностью (1-α)=0,95, считая σ неизвестным; г) найти доверительный интервал, покрывающий неизвестное квадратическое отклонение σ с заданной надежностью (1-α)=0,95; д) принимая доверительную вероятность (1-α)=0,99, найти предельную погрешность, с которой среднее арифметическое оценивает истинную длину детали; е) найти минимальное число измерений, которое надо произвести, чтобы с доверительной вероятностью (1-α)=0,98 можно было бы утверждать, что принимая среднее арифметическое за истинную длину стержня, мы совершаем погрешность, не превышающую 0,002мм, считая σ=0,0029.

Ответ: а) $\bar{x}$=18,307; s=0.0029; б) (18,3045;18,3095); в) (18,3024;18,3106); г) (0,0017;0,0084); д) ε=0,006; е) n$\approx$11.

5.30. Шестикратное взвешивание изделия из ценного материала дало следующие результаты (в граммах): 5,850, 5,861, 5,844, 5,857, 5,846, 5,825. Предполагая, что результаты измерений распределены по номальному закону, требуется: а) найти точечные оценки математического ожидания и среднего квадратического отклонения; б) найти 99%-ный доверительный интервал, покрывающий истинный вес изделия; в) найти 95%-ный доверительный интервал, покрывающий среднее квадратическое отклонение; г) найти предельную погрешность, которую мы допускаем, считая истинный вес изделия равным средней арифметической (доверительную вероятность принять равной 0,98).

Ответ: $\bar{x}$=5,850; s=0,008; 5,837<a<5,863; 0,005<σ<0,020; ε=0,12.

5.31. По результатам испытаний на разрыв 20 образцов дюралюминия вычислены выборочные характеристики $\bar{x}$=45,3кГ/мм² и S=1,13кГ/мм². Предполагая, что результаты испытания на разрыв подчиняются нормальному закону распределения, требуется: а) найти 90%-ный доверительный интервал, покрывающий неизвестное математическое ожидание предела прочности дюралюминия; б) найти 90%-ный доверительный интервал, покрывающий неизвестное средне квадратическое отклонение.

Ответ: 44,9<α<45,7; 0,87<σ<1,51.

5.32. Определение скорости автомобиля с прицепом было определено на мерном участке за пять испытаний, в результате которых вычислена оценка $\bar{x}$=52,22км/ч. Найти доверительный интервал, покрывающий значение скорости с надежностью 95%, если известно, что рассеивание скорости подчинено нормальному закону со средним квадратическим отклонением σ=0,126км/ч.

Ответ: (52,05, 52,38).

5.33. В результате 10 измерений индуктивности катушки получены следующие значения, мг:

8,345;	8,346;	8,348;	8,342;	8,343;
8,345;	8,343;	8,347;	8,344;	8,347.

1. С практической достоверностью, отвечающей надежности 0,98, указать границы, между которыми лежит истинное значение индуктивности.
2. С практической достоверностью, отвечающей надежности 0,94, указать границы, между которыми находится среднее квадратичское отклонение возможного результата измерений.

Ответ: 1) (8,3442, 8,3468); 2) (0,0008, 0,0032).

5.6. Проверка статистических гипотез

В экономике, технике, естествознании и т.д. часто для выяснения того или иного случайного явления прибегают к высказыванию гипотез (предположений), которые можно проверить статистически, т.е. опираясь на результаты наблюдений в случайной выборке.

Статистической гипотезой называют любое предположение о виде неизвестного закона распределения, случайной величины или значений его параметров.

Статистическая гипотеза называется ***непараметрической***, если в ней сформулированы предположения относительно вида функции распределения.

Статистическая гипотеза называется ***параметрической***, если в ней сформулированы предположения относительно значений параметров функции распределения известного вида.

Процесс использования выборки для проверки истинности (ложности) статистических гипотез называется ***статистическим доказательством истинности (ложности) выдвинутой гипотезы***.

5.6.1. Статистическая проверка параметрических гипотез

Основные принципы проверки статистических гипотез состоят в следующем. Пусть $f(x,\theta)$ - плотность распределения случайной величины x, зависящей от одного параметра θ. Предположим, что необходимо проверить гипотезу о том, что $\theta=\theta_0$, где θ_0 – определенное число. Назовем эту гипотезу нулевой (основной) и обозначим ее через H_0.

Нулевой гипотезой H_0 называют выдвинутую гипотезу, которую необходимо проверить. Обычно нулевые гипотезы утверждают, что различие между сравниваемыми величинами отсутствует, а наблюдаемые отклонения объясняются лишь случайными колебаниями выборки.

Альтернативной гипотезой H_1 называется гипотеза, конкурирующая с нулевой гипотезой в том смысле, что, если нулевая гипотеза отвергается, то принимается альтернативная.

Предположим, что по выборке $x_1, x_2, ..., x_n$ построена нормальная модель с параметрами α и δ. Тогда относительно параметров генеральной совокупности α и δ можно выдвинуть следующие гипотезы

Нулевые:	Альтернативные:
1) $H_0 : a = a_0$;	1) $H_1 : \alpha \neq \alpha_0; \alpha < \alpha_0; \alpha > \alpha_0$;
2) $H_0 : \sigma = \sigma_0$;	2) $H_1 : \sigma \neq \sigma_0; \sigma < \sigma_0; \sigma > \sigma_0$;
3) $H_0 : \begin{cases} a = a_0 \\ \sigma = \sigma_0 \end{cases}$.	3) $H_1 : \begin{cases} a \neq a_0 \\ \sigma \neq \sigma_0 \end{cases}$.

Содержательная сущность этих гипотез может принимать различный вид в зависимости от конкретной задачи исследования. Рассмотрим, например, следующую задачу: будет ли новый способ изготовления электроламп увеличивать срок их службы по сравнению с существующим способом, при котором средний срок службы $\alpha_0 = 4500 час.$? Испытания небольшой партии электроламп дали $\overline{x} = 4800 ч.$ Можно ли на основании этих данных считать, что новый способ производства электроламп лучше старого?

Можно выдвинуть нулевую и альтернативную гипотезы:

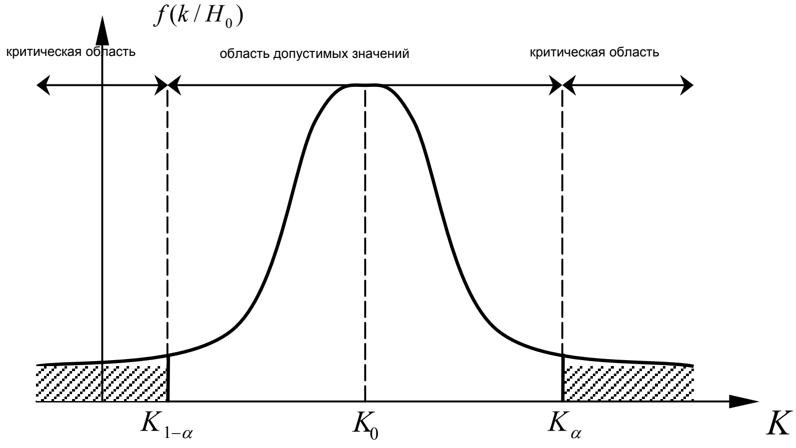
$H_0 : a = a_0$ - при новом способе производства срок службы электроламп остался прежним;

$H_1 : a \rangle a_0$ - при новом способе производства срок службы электроламп остался прежним.

Проверка статистических гипотез осуществляется на основе данных выборки. Для этого используют специальным образом подобранную случайную величину (выборочную статистику), являющуюся функцией ранее наблюдаемых значений, точное или приближенное распределение которой известно.

Статистическим критерием К будем называть случайную величину *К*, с помощью которой принимаются решения о принятии или отвержении выдвинутой нулевой гипотезы.

Для проверки нулевых гипотез по выборочным данным вычисляют частные значения входящих в критерий величин и, таким образом, получают наблюдаемое значение критерия.



Пусть для проверки некоторой нулевой гипотезы H_0 относительно параметров распределения служит выборочная статистика (критерий) *К*.

Предположим, что плотность распределения вероятностей выборочной статистики *К* при условии справедливости проверяемой гипотезы H_0 равна $f(K/H_0)$, а математическое ожидание статистики K равно K_0. Тогда вероятность того, что случайная величина K попадет в произвольный интервал $(K_{1-\alpha/2};K_{\alpha/2})$, можно найти по формуле

$$P(K_{1-\alpha/2} < K < K_{\alpha/2}) = \int\limits_{K_{1-\alpha/2}}^{K_{\alpha/2}} f(K/H_0)dK .$$

Зададим эту вероятность равной *(1-α)* и определим квантили K–распределения $K_{1-\alpha/2}$ и $K_{\alpha/2}$ при условиях

$$P(K \le K_{1-\alpha/2}) = \int\limits_{-\infty}^{K_{1-\alpha/2}} f(K/H_0)dK = \frac{\alpha}{2},$$

$$P(K \ge K_{\alpha/2}) = \int\limits_{K_{\alpha/2}}^{\infty} f(K/H_0)dK = \frac{\alpha}{2}.$$

Следовательно, вероятность того, что СВ K будет находиться внутри интервала $(K_{1-\alpha/2};K_{\alpha/2})$, равна *1-α*.

Зададим вероятность α настолько малой, чтобы попадание СВ K за пределы интервала $(K_{1-\alpha/2};K_{\alpha/2})$ можно было считать маловероятным событием. Тогда, исходя из принципа практической невозможности маловероятных событий, можно считать, что если нулевая гипотеза справедлива, то при ее проверке наблюдаемое (частное) значение критерия K должно обязательно попасть в интервал $(K_{1-\alpha/2};K_{\alpha/2})$. Поэтому область $(K_{1-\alpha/2};K_{\alpha/2})$ называют областью допустимых значений СВ К, при которых нулевая гипотеза не отклоняется, а области $(-\infty;K_{1-\alpha/2})$ и $(K_{\alpha/2};\infty)$ - областями отклонения проверяемой нулевой гипотезы или критической областью критерия K.

Если критические области располагаются справа и слева от математического ожидания СВ K, то критическая область называется двусторонней.

В некоторых случаях экспериментатор бывает твердо убежден, что $K > K_0$ или $K < K_0$.

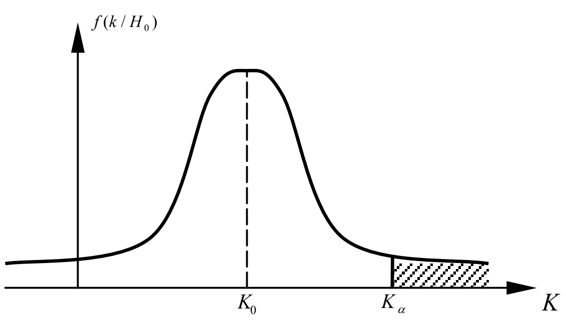


Рис.10

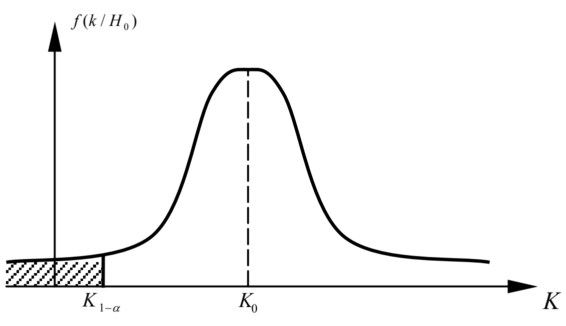


Рис.11

В этом случае критические области являются односторонними. На этих рисунках изображены правосторонняя и левосторонняя критические области.

При использовании этого принципа возможны четыре случая:

1) гипотеза H_0 верна, и ее принимают согласно критерию;
2) гипотеза H_0 неверна, и ее отвергают согласно критерию;

3) гипотеза H_0 верна, но ее отвергают согласно критерию, т.е. допускается ошибка, которую принято называть ***ошибкой первого рода***;
4) гипотеза H_0 неверна, но ее принимают согласно критерию, т.е. допускается ***ошибка второго рода***.

Уровнем значимости α называют вероятность совершить ошибку первого рода, т.е. вероятность отвергнуть гипотезу H_0, когда она верна. С уменьшением α возрастает вероятность ошибки второго рода.

Таким образом, проверка параметрической гипотезы при помощи статистического критерия может быть разбита на следующие этапы:

1) сформулировать основную H_0 и альтернативную H_1 гипотезы;
2) назначить уровень значимости α;
3) выбрать выборочную статистику *K* для проверки гипотезы H_0;
4) определить выборочное распределение статистики *K* при условии, что верна гипотеза H_0;
5) в зависимости от формулировки альтернативной гипотезы определить критическую область одним из неравенств $K < K_{1-\alpha}$, $K > K_\alpha$ или совокупностью неравенств $K < K_{1-\alpha/2}$ и $K > K_{\alpha/2}$;
6) получить выборку наблюдений и вычислить выборочное значение статистики критерия $K_{набл}$;
7) принять статистическое решение: если $K_{набл}$ принадлежит области допустимых значений СВ *K*, то принять гипотезу H_0, т.е. считать что она не противоречит результатам наблюдений; если $K_{набл}$ принадлежит критической области, то отклонить гипотезу H_0 как не согласующуюся с результатами наблюдений.

Замечание. Обычно на этапах 4 - 7 используют статистику, квантили которой табулировали: статистику u или z с нормальным распределением *N(0,1)*; статистику *t*, имеющую распределение Стьюдента; статистику x^2 *Пирсона* или статистику *F*, имеющую распределение Фишера.

5.6.1.1. Проверка гипотез о математическом ожидании случайной величины, распределенной по нормальному закону

Модель 1. Дисперсия генеральной совокупности известна.

Правило 1. *Для того чтобы при заданном уровне значимости α проверить нулевую гипотезу $H_0 : a = a_0$ о равенстве генеральной средней a нормальной совокупности с известной дисперсией σ^2 гипотетическому (предполагаемому) значению a_0 при конкурирующей гипотезе $H_1 : a \neq a_0$, надо вычислить наблюдаемое значение критерия* $U_{набл} = \dfrac{(\bar{x} - a_0)\sqrt{n}}{\sigma}$

и по таблице функции Лапласа найти критическую точку $u_{кр}$ *двусторонней критической области из равенства* $\Phi(u_{кр}) = \frac{(1-\alpha)}{2}$.

Если $|U_{набл}| < u_{кр}$, *то нет оснований отвергнуть нулевую гипотезу. Если* $|U_{набл}| > u_{кр}$, *то нулевую гипотезу отвергают.*

Правило 2. *При конкурирующей гипотезе* $H_1 : a > a_0$ *критическую точку правосторонней критической области находят из равенства* $\Phi(u_{кр}) = \frac{(1-2\alpha)}{2}$. *Если* $U_{набл} < u_{кр}$, *то нет оснований отвергнуть нулевую гипотезу. Если* $U_{набл} > u_{кр}$, *то нулевую гипотезу отвергают.*

Правило 3. *При конкурирующей гипотезе* $H_1 : a < a_0$ *сначала находят вспомогательную критическую точку* $u_{кр}$ *по правилу 2, а затем полагают границу левосторонней критической области* $u'_{кр} = -u_{кр}$.

Если $U_{набл} > -u_{кр}$, *то нет оснований отвергнуть нулевую гипотезу. Если* $U_{набл} < -u_{кр}$, *то нулевую гипотезу отвергают.*

Мощность критерия проверки нулевой гипотезы $H_0 : a = a_0$ о равенстве генеральной средней гипотетическому значению a_0 при известном среднем квадратическом отклонении σ находят в зависимости от вида конкурирующей гипотезы.

При конкурирующей гипотезе $H_1 : a > a_0$ для гипотетического значения генеральной средней $a = a_1 > a_0$ мощность правостороннего критерия

$1 - \beta = 0{,}5 - \Phi(u_{кр} - \lambda)$, (*), где $u_{кр}$ находят из равенства

$$\Phi(u_{кр}) = (1-2\alpha)/2,$$

$$\lambda = \frac{(a_1 - a_0)\sqrt{n}}{\sigma}.$$

При различных значениях a_1 функция мощности одностороннего критерия

$$\pi_1(a_1) = 0{,}5 - \Phi(u_{кр} - \lambda).$$

При конкурирующей гипотезе $H_1 : a \neq a_0$ для гипотетического значения генеральной средней $a = a_1$ мощность двустороннего критерия

$1 - \beta = 1 - [\Phi(u_{кр} - \lambda) + \Phi(u_{кр} + \lambda)]$, (**), где $u_{кр}$ находят из равенства

$$\Phi(u_{кр}) = (1-\alpha)/2,$$

$$\lambda = \frac{(a_1 - a_0)\sqrt{n}}{\sigma}.$$

При различных значениях a_1 функция мощности двустороннего критерия

$$\pi_1(a_1) = 1 - [\Phi(u_{кр} - \lambda) + \Phi(u_{кр} + \lambda)].$$

В формулах (*) и (**) $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-z^2/2} dz$ - функция Лапласа.

Пример. Из нормальной генеральной совокупности с известным средним квадратическим отклонением $\sigma = 5{,}2$ извлечена выборка объема n=100, и по ней найдена выборочная средняя $\bar{x} = 27{,}56$. Требуется при уровне значимости 0,05 проверить нулевую гипотезу $H_0: a = a_0 = 26$ при конкурирующей гипотезе $H_1: a \neq 26$.

Решение. Найдем наблюдаемое значение критерия:

$$U_{набл} = \frac{(\overline{x} - a_0)\sqrt{n}}{\sigma} = \quad 27{,}56 \qquad .$$

По условию, конкурирующая гипотеза имеет вид $a \neq a_0$, поэтому критическая область - двусторонняя.

Найдем критическую точку из равенства $\Phi(u_{кр}) = \frac{(1-\alpha)}{2} = \frac{(1-0{,}05)}{2} = 0{,}475$.

По таблице функции Лапласа (см. табл.2 прил.) находим $u_{кр}$=1,96.

Так как $U_{набл} > u_{кр}$ - нулевую гипотезу отвергаем. Другими словами, выборочная и гипотетическая генеральная средние *различаются значимо*.

Модель 2. Дисперсия генеральной совокупности неизвестна.

Если дисперсия генеральной совокупности неизвестна (например, в случае малых выборок), то в качестве критерия проверки нулевой гипотезы принимают случайную величину $T = (\overline{X} - a_0)\sqrt{n} / S$, где $S = \sqrt{\frac{\sum n_i x_i^2 - \left[\sum n_i x_i\right]^2 / n}{n-1}}$ - исправленное среднее квадратическое отклонение. Величина *T* имеет распределение Стьюдента с *k=n—1* степенями свободы.

Правило 1. *Для того чтобы при заданном уровне значимости α, проверить нулевую гипотезу $H_0: a = a_0$ о равенстве неизвестной генеральной средней а (нормальной совокупности с неизвестной дисперсией) гипотетическому значению a_0 при конкурирующей гипотезе $H_1: a \neq a_0$, надо вычислить наблюдаемое значение критерия $T_{набл} = (\overline{x} - a_0)\sqrt{n} / s$ и по таблице критических точек распределения Стьюдента, по заданному уровню значимости а, помещенному в верхней строке таблицы, и числу степеней свободы k = n— 1 найти критическую точку $t_{двуст.\ кр}(\alpha, k)$.*

Если $|T_{набл}| < t_{двуст.кр}$, то нет оснований отвергнуть нулевую гипотезу. Если

Правило 2. *При конкурирующей гипотезе $H_1: a > a_0$ по уровню значимости а, помещенному в нижней строке таблицы приложения* 6, *и числу степеней свободы k=n—1 находят критическую точку $t_{правост.кр}(\alpha; k)$ правосторонней критической области. Если $T_{набл} < t_{правост.кр}$, то нет оснований отвергнуть нулевую гипотезу. Если $T_{набл} > t_{правост.кр}$, то нулевую гипотезу отвергают.*

Правило 3. *При конкурирующей гипотезе $H_1: a < a_0$ сначала находят «вспомогательную» критическую точку (по правилу 2) $t_{правост.кр}(\alpha; k)$ и полагают границу левосторонней критической области $t_{левост.кр} = -t_{правост.кр}$ Если $T_{набл} > t_{правост.кр}$*

нет оснований отвергнуть нулевую гипотезу. Если $T_{набл} < -t_{правост.кр}$, *то нулевую гипотезу отвергают.*

Пример. По выборке объема n =16, извлеченной из нормальной генеральной совокупности, найдены выборочная средняя $\bar{x} = 118$ и «исправленное» среднее квадратическое отклонение s=3,6. Требуется при уровне значимости 0,05 проверить нулевую гипотезу $H_0 : a = a_0 = 120$ при конкурирующей гипотезе $H_1 : a \neq 120$.

Решение. Найдем наблюдаемое значение критерия:

$$T_{набл} = \left(\bar{x} - a_0\right)\sqrt{n} / s = \frac{(118{,}2 - 120)\sqrt{16}}{3{,}6} = -2 .$$

По условию, конкурирующая гипотеза имеет вид $a \neq a_0$, поэтому критическая область — двусторонняя.

По таблице критических точек распределения Стьюдента (см. табл. 4 прил.), по уровню значимости a=0,05, помещенному в верхней строке таблицы, и по числу степеней свободы k=n—1=16—1=15 находим критическую точку $t_{двуст.кр}(0{,}05;15) = 2{,}13$. Так как $|T_{набл}| < t_{двуст.кр}$, нет оснований отвергнуть нулевую гипотезу. Другими словами, выборочная средняя незначимо отличается от гипотетической генеральной средней a_0=120.

5.6.1.2. Проверка гипотез о дисперсии случайной величины х, распределенной по нормальному закону

Обозначим через n объем выборки, по которой найдена исправленная дисперсия s^2.

Правило 1. *Для того чтобы при заданном уровне значимости* α *проверить нулевую гипотезу* $H_0 : \sigma^2 = \sigma_0^2$ *о равенстве неизвестной генеральной дисперсии* σ^2 *гипотетическому (предполагаемому) значению* σ_0^2 *при конкурирующей гипотезе* $H_1 : \sigma^2 > \sigma_0^2$, *надо вычислить наблюдаемое значение критерия* $\chi^2_{набл} = \dfrac{(n-1)s^2}{\sigma_0^2}$ *и по таблице критических точек распределения* χ^2, *по заданному уровню значимости* α *и числу степеней свободы* k=n—1 *найти критическую точку* $\chi^2_{кр}(\alpha;k)$. *Если* $\chi^2_{набл} < \chi^2_{кр}$, *то нет оснований отвергнуть нулевую гипотезу. Если* $\chi^2_{набл} > \chi^2_{кр}$, *то нулевую гипотезу отвергают.*

Правило 2. *При конкурирующей гипотезе* $H_1 : \sigma^2 \neq \sigma_0^2$ *находят левую* $\chi^2_{лев.кр}(1 - \alpha/2;k)$ *и правую* $\chi^2_{прав.кр}(\alpha/2;k)$ *критические точки. Если* $\chi^2_{лев.кр} < \chi^2_{набл} < \chi^2_{прав.кр}$ *- нет оснований отвергнуть нулевую гипотезу. Если* $\chi^2_{набл} < \chi^2_{лев.кр}$ *или* $\chi^2_{набл} > \chi^2_{прав.кр}$ *- нулевую гипотезу отвергают.*

Правило 3. *При конкурирующей гипотезе* $H_1 : \sigma^2 < \sigma_0^2$ *находят критическую точку* $\chi^2_{кр}(1 - \alpha;k)$. *Если* $\chi^2_{набл} > \chi^2_{кр}(1 - \alpha;k)$, *то нет оснований* *отвергнуть нулевую гипотезу. Если* $\chi^2_{набл} < \chi^2_{кр}(1-\alpha;k)$, *то нулевую гипотезу отвергают.*

Замечание. Если число степеней свободы k > 30, то критическую точку $\chi^2_{кр}(\alpha;k)$ можно найти из равенства Уилсона-Гильферти:

$$\chi^2_{кр}(\alpha;k) = k\,1 - 2/9k + z_\alpha\sqrt{2/\quad} \quad ,$$

где z_α находят, используя функцию Лапласа (см. табл.2 прил.), из равенства $\Phi(z_\alpha) = (1-2\alpha)/2$.

Пример. Из нормальной генеральной совокупности извлечена выборка объема n = 21, и по ней найдена исправленная выборочная дисперсия s^2=16,2. Требуется при уровне значимости 0,01 проверить нулевую гипотезу $H_0 : \sigma^2 = \sigma_0^2 = 15$, приняв в качестве конкурирующей гипотезы $H_1 : \sigma_0^2 > 15$.

Решение. Найдем наблюдаемое значение критерия:

$$\chi^2_{набл} = \frac{(n-1)s^2}{\sigma_0^2} = \frac{(21-1)16{,}2}{15} \qquad .$$

По условию, конкурирующая гипотеза имеет вид $\sigma^2 > 15$, поэтому критическая область-правосторонняя (правило 1). По таблице приложения 5, по уровню значимости 0,01 и числу степеней свободы k=n—1=21-1=20 находим критическую точку $\chi^2_{кр}(0{,}01;20) = 37{,}6$.

Так как $\chi^2_{набл} < \chi^2_{кр}$, то нет оснований отвергнуть нулевую гипотезу о равенстве генеральной дисперсии гипотетическому значению $\sigma_0^2 = 15$. Другими словами, различие между исправленной дисперсией (16,2) и гипотетической генеральной дисперсией (15) *незначимо*.

5.6.1.3. Сравнение двух дисперсий нормальных генеральных совокупностей

По независимым выборкам, объемы которых n_1, n_2, извлеченным из нормальных генеральных совокупностей, найдены исправленные выборочные дисперсии s_X^2 и s_Y^2. Требуется сравнить эти дисперсии.

Правило 1. *Для того чтобы при заданном уровне значимости* α *проверить нулевую гипотезу H_0: D(X)=D(Y) о равенстве генеральных дисперсий нормальных совокупностей при конкурирующей гипотезе H_1: D(X)>D(Y), надо вычислить наблюдаемое значение критерия (отношение большей исправленной дисперсии к меньшей)* $F_{набл} = s_Б^2 / s_М^2$ *и по таблице критических точек распределения Фишера - Снедькора, по заданному уровню значимости* α *и числам степеней свободы* $k_1 = n_1 - 1$, $k_2 = n_2 - 1$ *(*k_1 *- число степеней свободы большей исправленной дисперсии) найти критическую точку* . *Если* $F_{набл} < F_{кр}$, *то нет оснований отвергнуть нулевую гипотезу. Если* $> F_{кр}$, *то нулевую гипотезу*

Правило 2. *При конкурирующей гипотезе* $H_1 : D(X) \neq D(Y)$ *критическую точку* $F_{кр}(\alpha/2; k_1; k_2)$ *находят по уровню значимости* $\alpha/2$ *(вдвое меньшему заданного) и числам степеней свободы* k_1 *и* k_2 *(*k_1 *- число степеней свободы большей дисперсии). Если* $F_{набл} < F_{кр}$, *то нет оснований отвергнуть нулевую гипотезу. Если* $F_{набл} > F_{кр}$, *то нулевую гипотезу отвергают.*

Пример. По двум независимым выборкам, объемы которых n_1=11 и n_2=14, извлеченным из нормальных генеральных совокупностей X и Y, найдены исправленные выборочные дисперсии s_X^2=0,76 и s_Y^2=0,38. При уровне значимости α=0,05 проверить нулевую гипотезу $H_0 : D(X) = D(Y)$ о равенстве генеральных дисперсий при конкурирующей гипотезе $H_1 : D(X) > D(Y)$.

Решение. Найдем отношение большей исправленной дисперсии к меньшей: $F_{\text{набл}}$=0,76/0,38=2. По условию конкурирующая гипотеза имеет вид $D(X) > D(Y)$, поэтому критическая область – правосторонняя.

По табл.7приложения, по уровню значимости α=0,05 и числам степеней свободы $k_1 = n_1 - 1 = 11 - 1 = 10$ и $k_2 = n_2 - 1 = 14 - 1 = 13$ находят критическую точку

$$F_{кр}(0{,}05;10;13) = 2{,}67 .$$

Так как $F_{набл} < F_{кр}$ - нет оснований отвергнуть гипотезу о равенстве генеральных дисперсий. Другими словами, выборочные исправленные дисперсии различаются *незначимо*.

5.6.1.4. Сравнение двух средних нормальных генеральных совокупностей, дисперсии которых известны

Обозначим через n и m объемы больших (n>30, m>30) независимых выборок, по которым найдены соответствующие выборочные средние $\overline{x}$ и $\overline{y}$. Генеральные дисперсии $D(X)$ и $D(Y)$ известны.

Правило 1. *Для того чтобы при заданном уровне значимости проверить нулевую гипотезу* H_0: $M(X)=M(Y)$ *о равенстве математических ожиданий (генеральных средних) двух нормальных генеральных совокупностей с известными дисперсиями (в случае больших выборок) при конкурирующей гипотезе* $H_1 : M(X) \neq M(Y)$, *надо вычислить наблюдаемое значение критерия*

$$Z_{набл} = \frac{\overline{x} - \overline{y}}{\sqrt{D(X)/n + D(Y)/m}}$$

и по таблице функции Лапласа найти критическую точку $z_{\text{кр}}$ *из равенства*

$$\Phi(z_{кр}) = (1 - \alpha)/2.$$

Если $|Z_{набл}| < z_{кр}$, *то нет оснований отвергнуть нулевую гипотезу. Если* $|Z_{набл}| > z_{кр}$, *то нулевую гипотезу отвергают.*

Правило 2. *При конкурирующей гипотезе* $H_1 : M(X) > M(Y)$ *находят критическую точку* $z_{кр}$ *по таблице функции Лапласа из равенства*

$$\Phi(z_{кр}) = (1 - 2\alpha)/2..$$

Если $Z_{набл} < z_{кр}$*, то нет оснований отвергнуть нулевую гипотезу. Если* $Z_{набл} > z_{кр}$*, то нулевую гипотезу отвергают.*

Правило 3. *При конкурирующей гипотезе* $H_1 : M(X) < M(Y)$ *находят «вспомогательную точку»* $z_{кр}$ *по правилу 2. Если* $Z_{набл} > -z_{кр}$*, то нет оснований отвергнуть нулевую гипотезу. Если* $Z_{набл} < -z_{кр}$*, то нулевую гипотезу отвергают.*

Пример. По двум независимым выборкам, объемы которых *n*=40 и *m*=50, извлеченным из нормальных генеральных совокупностей, найдены выборочные средние: $\overline{x}$=130 и $\overline{y}$=140. Генеральные дисперсии известны: *D(X)*=80, *D(Y)*=100. Требуется при уровне значимости 0,01 проверить нулевую гипотезу $H_0 : M(X) = M(Y)$ при конкурирующей гипотезе $H_1 : M(X) \neq M(Y)$.

Решение. Найдем наблюдаемое значение критерия:

$$Z_{набл} = \frac{\overline{x} - \overline{y}}{\sqrt{D(X)/n + D(Y)/m}} = \frac{130 - 140}{\sqrt{80/40 + 100/50}} = -5 .$$

По условию, конкурирующая гипотеза имеет вид $M(X) \neq M(Y)$, поэтому критическая область - двусторонняя. Найдем правую критическую точку из равенства $\Phi(z_{кр}) = (1-\alpha)/2. = (1-0{,}01)/2 = 0{,}495.$

По таблице функции Лапласа (см. табл. 2 прил.) находим $z_{кр}$=2,58.

Так как $|Z_{набл}| > z_{кр}$, то в соответствии с правилом 1 нулевую гипотезу отвергаем. Другими словами, выборочные средние различаются значимо.

5.6.1.5. Сравнение двух средних нормальных генеральных совокупностей, дисперсии которых неизвестны и одинаковы

Обозначим через *n* и *m* объемы малых независимых выборок (*n*>30, *m*>30), по которым найдены соответствующие выборочные средние $\overline{x}$ и $\overline{y}$ и исправленные выборочные дисперсии s_X^2 и s_Y^2. Генеральные дисперсии, хотя и неизвестны, но предполагаются одинаковыми.

Правило 1. *Для того чтобы при заданном уровне значимости* α *проверить нулевую гипотезу* H_0*: M(X)=M(Y) о равенстве математических ожиданий (генеральных средних) двух нормальных совокупностей с неизвестными, но одинаковыми дисперсиями (в случае малых независимых выборок) при конкурирующей гипотезе* $H_1 : M(X) \neq M(Y)$*, надо вычислить наблюдаемое значение критерия* $T_{набл} = \dfrac{\overline{x} - \overline{y}}{(n-1)s_X^2 + (m-1)s_Y^2}\sqrt{\dfrac{nm(n+m-2)}{n+m}}$ *и по таблице критических точек распределения Стьюдента, по заданному уровню значимости* α*, помещенному в верхней строке таблицы приложения 6, и числу степеней свободы* $k = n + m - 2$ *найти критическую точку* $t_{двуст.кр}(\alpha;k)$*. Если* $|T_{набл}| < t_{двуст.кр}(\alpha;k)$*, то нет оснований отвергнуть нулевую гипотезу. Если* $|T_{набл}| > t_{двуст.кр}(\alpha;k)$*, то нулевую гипотезу отвергают.*

Правило 2. *При конкурирующей гипотезе M(X)>M(Y) находят критическую точку* $t_{правост.кр}(\alpha;k)$ *по таблице приложения 6 по уровню значимости* α, *помещенному в нижней строке таблицы, и числу степеней свободы* $k = n + m - 2$. *Если* $T_{набл} < t_{правост.кр}$, *то нет оснований отвергнуть нулевую гипотезу. Если* $T_{набл} > t_{правост.кр}$, *то нулевую гипотезу отвергают.*

Правило 3. *При конкурирующей гипотезе M(X)<M(Y) находят сначала критическую точку* $t_{правост.кр}$ *по правилу 2 и полагают* $t_{левост.кр} = -t_{правост.кр}$. *Если* $T_{набл} > -t_{правост.кр}$, *то нет оснований отвергнуть нулевую гипотезу. Если* $T_{набл} < -t_{правост.кр}$, *то нулевую гипотезу отвергают.*

Пример. По двум независимым малым выборкам, объемы которых n=12 и m=18, извлеченным из нормальных генеральных совокупностей X и Y, найдены выборочные средние: $\overline{x}$=31,2, $\overline{y}$=29,2 и исправленные дисперсии: s_X^2=0,84 и s_Y^2=0,40. Требуется при уровне значимости 0,05 проверить нулевую гипотезу H_0: *M(X)=M(Y)* при конкурирующей гипотезе $H_1 : M(X) \neq M(Y)$.

Решение. Исправленные дисперсии различны, поэтому проверим предварительно гипотезу о равенстве генеральных дисперсий, используя критерий Фишера-Снедокора.

Найдем отношение большей дисперсии к меньшей: $F_{набл} = 0{,}84 / 0{,}40 = 2{,}1$. Дисперсия s_X^2 значительно больше дисперсии s_Y^2, поэтому в качестве конкурирующей примем гипотезу $H_1 : M(X) > M(Y)$. В этом случае критическая область - правосторонняя. По таблице приложения 7, по уровню значимости α=0,05 и числам степеней свободы $k_1 = n - 1 = 12 - 1 = 11$ и $k_2 = m - 1 = 18 - 1 = 17$ находим критическую точку $F_{кр}(0{,}05;11;17) = 2{,}41$.

Так как $F_{набл} < F_{кр}$ - нет оснований отвергнуть нулевую гипотезу о равенстве генеральных дисперсий. Предположение о равенстве генеральных дисперсии выполняется, поэтому сравним средние.

Вычислим наблюдаемое значение критерия Стьюдента:

$$T_{набл} = \frac{\overline{x} - \overline{y}}{(n-1)s_X^2 + (m-1)s_Y^2} \sqrt{\frac{nm(n+m-2)}{n+m}}.$$

Подставив числовые значения входящих в эту формулу величин, получим $T_{набл} = 7{,}1$.

По условию, конкурирующая гипотеза имеет вид $M(X) \neq M(Y)$, поэтому критическая область - двусторонняя. По уровню значимости 0,05 и числу степеней свободы $k = n + m - 2 = 12 + 18 - 2 = 28$ находим по таблице приложения 4 критическую точку $t_{двуст.кр}(0{,}05;28) = 2{,}05$.

Так как $T_{набл} > t_{двуст.кр}$, то нулевую гипотезу о равенстве средних отвергаем. Другими словами, выборочные средние *различаются значимо*.

5.6.2. Статистическая проверка непараметрических гипотез

При обработке статистических данных для характеристики частотных свойств ряда наблюдений x_1, x_2, ..., x_n экспериментатор подбирает теоретико-вероятностную модель (нормальную, показательную, биномиальную и т.д.) этого ряда. Предположим, что экспериментатор визуально по виду гистограммы или полигона частостей или из каких-либо других соображений выдвинул гипотезу о множестве функций Ω определенного вида (нормальных, показательных, биномиальных и т. д.), к которому может принадлежать функция распределения исследуемой СВ X. Предположения такого рода называются *непараметрическими гипотезами.*

Предположим, что класс таких функций выбран и произведена точечная оценка параметров этих функций внутри выбранного класса. Дальнейшая задача экспериментатора состоит в проверке выдвинутой гипотезы о классе функций Ω, т.е. в выяснении, насколько хорошо подобрана вероятностная модель ряда наблюдений.

Проверка гипотезы о предполагаемом законе распределения производится с помощью непараметрических критериев значимости. Проверка непараметрических гипотез производится на основании вычисления некоторой выборочной статистики (критерия), закон распределения которой получен в предположении истинности нулевой гипотезы и сравнения наблюдаемого значения этой выборочной статистики критическим значением. Рассмотрим наиболее распространенный критерий согласия χ^2 Пирсона.

Критерий χ^2 Пирсона позволяет производить проверку согласия эмпирической функции распределения с гипотетической функцией $F(x)$, принадлежащей к некоторому множеству Ω функций определенного вида (нормальных, показательных, биномиальных и т.д.). Сформулируем основные вероятностные предпосылки и ограничения, которые должны быть выполнены при применении критерия χ^2 в виде модели.

Модель. Пусть генеральная совокупность имеет функцию распределения $F(x)$, принадлежащую некоторому классу функций Ω. Из генеральной совокупности извлечена выборка объема n(n≥50).

Разобьем весь диапазон полученных результатов на k частичных интервалов равной длины, и пусть в каждом частичном интервале оказалось m_i, измерений, причем $\sum_{i=1}^{k} m_i = n$. Составим сгруппированный статистический ряд распределения частот.

Интервалы наблюденных значений СВ X	$(x_0; x_1)$	$(x_1; x_2)$	…	$(x_{i-1}; x_i)$	…	$(x_{k-1}; x_k)$
Частоты m_i	m_1	m_2	…	m_i	…	M_k

$$\sum_{i=1}^{k} m_i = n.$$

Требуется на основе имеющейся информации проверить нулевую гипотезу о том, что гипотетическая функция распределения $F(x)$ значимо представляет данную выборку, т.е. H_0: $F(x)\in\Omega$. При проверке нулевой гипотезы с помощью критерия согласия χ^2 придерживаются следующей последовательности действий:

1) на основании гипотетической функции $F(x)$ вычисляют вероятности попадания СВ X в частичные интервалы (разряды) $[x_{i-1}; x_i)$:

$$p_i = P(x_{i-1} \le X < x_i) = \int_{x_{i-1}}^{x_i} f(x)dx = F(x_i) - F(x_{i-1}); (i = 1,2,...,k);$$

2) умножая полученные вероятности рi на объем выборки n, получают теоретические частоты np_i частичных интервалов $(x_{i-1}; x_i)$, т.е. частоты, которые следует ожидать, если нулевая гипотеза справедлива;

3) вычисляют выборочную статистику (критерий) χ^2: $\chi^2 = \sum_{i=1}^{k} \frac{(m_i - np_i)^2}{np_i}$ (*).

Критерий χ^2 сконструирован таким образом, что, чем ближе к нулю наблюдаемое значение критерия χ^2, тем вероятнее, что нулевая гипотеза справедлива. Поэтому для проверки нулевой гипотезы применяется критерий χ^2 с правосторонней критической областью. Следовательно, для того чтобы проверить нулевую гипотезу, необходимо найти по таблицам квантилей χ^2 - распределения по заданному уровню значимости α и числу степеней свободы $\nu = k - r - 1$ критическое значение $\chi^2_{\alpha;\upsilon}$, удовлетворяющее условию $P(\chi^2 \ge \chi^2_{\alpha;\upsilon}) = \alpha$. Сравнивая наблюдаемое значение выборочной статистики χ^2, вычисленное по формуле (*), с критическим значением $\chi^2_{\alpha;\upsilon}$, принимают одно из двух решений:

1) если $\chi^2_{набл} \ge \chi^2_{\alpha;\upsilon}$, то нулевая гипотеза H_0: $F(x)\in\Omega$ отвергается в пользу альтернативной H_a: $F(x) \in \Omega$, т.е. считается, что гипотетическая функция не согласуется с опытными данными;
2) если $\chi^2_{набл} < \chi^2_{\alpha;\upsilon}$, то считается, что нет оснований для отклонения нулевой гипотезы, т.е. гипотетическая функция $F(x)$ согласуется с опытными данными.

Замечание. При применении критерия χ^2 необходимо, чтобы в каждом частичном интервале было *не менее пяти элементов*. Если число элементов (частота) меньше пяти, то рекомендуется объединять такие частичные интервалы с соседними.

Пример. Результаты исследования прочности на сжатие (СВ X) 200 образцов бетона представлены в виде сгруппированного статистического ряда ($\sum m_i = n = 200$). Требуется проверить нулевую гипотезу о нормальном законе распределения прочности на сжатие. Уровень значимости принять α=0,05.

Интервалы прочности, кг/см^2	Частоты m_i
193—200	10

200—210	26
210—220	56
220—230	64
230—240	30
240—250	14

Решение. Из условия следует, что точные параметры гипотетического нормального закона нам неизвестны, поэтому нулевую гипотезу можно сформулировать следующим образом: H_0: *F(x)* является функцией нормального распределения с параметрами $M(x)=\hat{a}=\bar{x}$ и $D(x)=\hat{\sigma}^2=s^2$.

Для проверки этой нулевой гипотезы определим значение x_i^*, середин интервалов и найдем точечные оценки математического ожидания и среднего квадратического отклонения нормально распределенной случайной величины по формулам

$$\hat{a}=\bar{x}=\frac{\sum x_i^* m_i}{n}=\frac{195\cdot 10+205\cdot 26+215\cdot 56+225\cdot 64+235\cdot 30+245\cdot 14}{200}=221\ \text{кг/см}^2,$$

$$\hat{\sigma}^2=s^2=\frac{1}{n}\sum (x_i^*-\bar{x})^2 m_i=\frac{1}{200}((-26)^2\cdot 10+(-16)^2\cdot 26+(-6)^2 56+4^2\cdot 64+14^2\cdot 30+24^2\cdot 14)=$$

$$=152; \qquad \hat{\sigma}=s=\sqrt{152}=12{,}33\ \text{кГ/см}^2.$$

Вычислим теоретические вероятности p_i попадания СВ X в частичные интервалы $[x_{i-i}; x_i)$ по формуле $p_i=P(x_{i-1}\le X<x_i)=\frac{1}{2}(\Phi(u_i)-\Phi(u_{i-1}))\ (i=1,2,\dots,k)$,

где $u_i=\dfrac{x_i^*-\bar{x}}{s}$; $\Phi(\mathrm{u_i})=\dfrac{2}{\sqrt{2\pi}}\int\limits_0^{u_i} e^{-\frac{t^2}{2}}dt$.

Интервалы изменения наблюдаемых значений СВ *X*	Частоты m_i	Нормированные интервалы $[u_i; u_{i+1})$	p_i=½[Φ(u_{i+1}) - -Φ(u_i)]	np_i	$(m_i-np_i)^2$	$\dfrac{(m_i-np_i)^2}{np_i}$
190—200	10	(-∞;-1,70)	0,045	9	1	0,11
200—210	26	(-1,70; -0,89)	0,142	28,4	5,76	0,20
210—220	56	(-0,89; -0,08)	0,281	56,2	0,04	0,00
220—230	64	(-0,08; +0,73)	0.299	59,8	17,6i	0,29
230—240	30	(+0,73;+1,54)	0,171	34,2	17,64	0,52
240—250	14	(+1,54;+∞)	0,062	12,4	2,56	0,23
Суммы:		*N*=200	1,000	200,0		$\chi^2_{\text{набл}}=1{,}35$

Вычисления, необходимые для определения наблюдаемого значения выборочной статистики χ^2, проведены в табл.3 (приложения).

Замечание. Так как СВ *X*, распределенная по нормальному закону, определена на (-∞;+∞), то наименьшее значение стандартизованной переменной

$\frac{190-22}{12,33}=-2,51$ заменено на $-\infty$, наибольшее значение $\frac{250-221}{12,33}=2,35$ заменено на $+\infty$.

В результате вычислений получили $\chi^2_{\text{набл}}=1,35$. Найдем по табл.3 (см. приложение) квантилей χ^2 - распределения по заданному уровню значимости α=0,05 и числу степеней свободы $\nu=k-r-1=6-2-1=3$ критическое значение $\chi^2_{0,05;3}=7,815$. Так как $\chi^2_{\text{набл}}=1,35<7,815$, то нет оснований для отклонения нулевой гипотезы о нормальном законе распределения предела прочности на сжатие с параметрами α=221 и σ^2=152.

Пример. В течение 100 дней фиксировалось количество аварий водопроводно-канализационной сети в некотором районе города М. Получены следующие числовые данные:

Число аварий (СВ X)	Частоты m_i
0	8
1	28
2	31
3	18
4	9
5	6
	$n=\Sigma m_i=100$

Проверить гипотезу о том, что распределение числа аварий водопроводно-канализационной сети города М. подчиняется закону Пуассона. Уровень значимости принять α=0,05.

Решение. Согласно условию, необходимо проверить нулевую гипотезу H_0: $F(x)$ - функция распределения числа аварии имеет вид $F(x,\lambda)=\sum\limits_{i=1}^{x}\frac{\lambda^i e^{-\lambda}}{i!}$ с параметром $\hat{\lambda}=\bar{x}=\frac{\sum m_i x_i}{n}=\frac{8\cdot 0+28\cdot 1+31\cdot 2+18\cdot 3+9\cdot 4+6\cdot 5}{100}=2,1$.

Вычислим теоретические вероятности p_i появления ровно x_i аварий в течение n дней по формуле Пуассона: $p_i=P_n(x_i)=\frac{\lambda^{x_i}e^{-\lambda}}{x_i!}\ (\mathrm{x_i}=0,1,2,3,4,5)$.

Дальнейшие вычисления сводим в таблицу:

Число аварий x_i	m_i	$p_i=\frac{(2,1)^{x_i}e^{-2,1}}{x_i!}$	np_i	$(m_i-np_i)^2$	$\frac{(m_i-np_i)^2}{np_i}$

0	8	0,122	12,2	17,64	1,45
1	28	0,257	25,7	5,29	0,21
2	31	0,270	27,0	16,00	0,59
3	18	0,189	18,9	0,81	0,04
4	9	0,099	9,9	0,81	0,08
5	6	0,063	6,3	0,03	0,01
Суммы:	100	1,000	100,0		$\chi^2_{\text{набл}} = 2{,}38$

По табл.3 (см. приложение) квантилей χ^2 - распределения по заданному уровню значимости α=0,05 и числу степеней свободы ν=k-r-1=6-1-1=4 найдем критическое значение $\chi^2_{0,05;4} = 9{,}488$. Так как $\chi^2_{\text{набл}} = 2{,}38 < 9{,}488$, то нет основании для отклонения гипотезы о том, что закон распределения числа аварий водопроводно-канализационной сети является законом Пуассона с параметром λ=2,1.

ЗАДАНИЯ

5.34 а) Из нормальной генеральной совокупности с известным средним квад-ратическим отклонением $\sigma = 40$ извлечена выборка объема n=64 и по ней найдена выборочная средняя $\bar{x} = 136{,}5$. Требуется при уровне значимости 0,01 проверить нулевую гипотезу H_0: a=a_0=130 при конкурирующей гипотезе H_1:a$\neq$130. б) Решить эту задачу при конкурирующей гипотезе H_1: a>130.

Ответ: а) $U_{\text{набл}}$=1,3; $U_{\text{кр}}$=2,57. Нет оснований отвергнуть нулевую гипотезу. б) $U_{\text{кр}}$= 2,33. Нет оснований отвергнуть нулевую гипотезу.

5.35. Предприятие, производящее электрические лампочки, гарантирует, что среднее время безоткатной работы лампочек мощностью 60 Вт равно 800 ч часов со стандартным отклонением 120 ч. Из некоторой партии производится случайная выборка 25 лампочек, для которых выборочное среднее время работы оказалось равным 750 ч. Можно ли на основании этого сказать, что исследуемая выборка лампочек не удовлетворяет гарантии?

Ответ: исследуемая партия лампочек должна быть забракована.

5.36. Проектный контролируемый размер изделий, изготовляемых станком автоматом, a=a_0=35мм. Измерения 20 случайно отобранных изделий дали следующие результаты:

контролируемый размер	x_i	34,8	34,9	35,0	35,1	35,3;
частота (число изделий)	n_i	2	3	4	6	5.

Требуется при уровне значимости 0.05 проверить нулевую гипотезу H_0: a=a_0=35 при конкурирующей гипотезе H_1: a$\neq$ 35.

Ответ: нет оснований отвергнуть нулевую гипотезу.

5.37. Техническая норма предусматривает в среднем 40 сек на выполнение определенной технической операции. От работниц, работающих на этой операции, поступили жалобы, что они в действительности затрачивают на эту операцию больше времени. Для проверки данной жалобы произведены хронометрические измерения времени выполнения этой операции у 16 работниц, занятых на этой операции, и получены следующие результаты: $\bar{x}$=42 сек, s=3,5 сек. Можно ли по имеющимся хронометрическим данным на уровне значимости 0,01 отклонить гипотезу о том, что действительное среднее время исполнения этой технической операции соответствует норме.

Ответ: нет оснований для отклонения гипотезы H_0.

5.38. Из нормальной генеральной совокупности извлечена выборка объема n=17 и по ней найдена исправленная выборочная дисперсия s^2=0,24. Требуется при уровне значимости 0.05 проверить нулевую гипотезу H_0: $\sigma^2 = \sigma_0^2 = 0{,}18$, приняв в качестве альтернативной гипотезы H_1: $\sigma^2 > 0{,}18$.

Ответ: нет оснований отвергнуть нулевую гипотезу.

5.39. Точность работы станка-автомата проверяется по дисперсии контролируемого размера деталей, которая не должна превышать σ_0^2=0,04. Взята проба из n=11 случайно отобранных деталей и получены следующие результаты, мм: 100,6; 99,6; 100,0; 100,1; 100,3; 100,0; 99,9; 100,2; 100,4; 100,6; 100,5. На основании имеющихся данных проверить, обеспечивает ли станок заданную точность. Уровень значимости принять равным 0.05.

Ответ: станок не обеспечивает заданной точности и требует переналадки.

5.40. По двум независимым выборкам, объемы которых n_1=9 b и n_2=16, извлеченным из нормальных генеральных совокупностей X и Y, найдены исправленные выборочные дисперсии $s_x^2 = 34{,}02$ и $s_y^2 = 12{,}15$. При уровне значимости 0.01 проверить нулевую гипотезу H_0: D(X)=D(Y) о равенстве генеральных дисперсий при конкурирующей гипотезе H_1: D(X)>D(Y).

Ответ: $F_{набл}$=2,8; $F_{кр}$(0,01; 8; 15)=4. Нет оснований отвергнуть нулевую гипотезу.

5.41. Двумя методами проведены измерения одной и той же величины. Получены следующие результаты :

а) в первом случае x_1=9,6; x_2=10,0; x_3=9,8; x_4=10,2 x_5=10,6;

б) во втором случае y_1=10,4; y_2=9,7; y_3=10,0; y_4=10,3.

Можно ли считать, что оба метода обеспечивают одинаковую точность измерений, если принять уровень значимости равным 0,1? Предполагается, что результаты измерений распределены нормально и выборки независимы.

Ответ: оба метода обеспечивают одинаковую точность измерений.

5.42. По выборке объема n=30 найден средний вес $\bar{x} = 130$ г изделий, изготовленных на первом станке; по выборке объема m=40 найден средний вес $\bar{y} = 125$ г изделий, изготовленных на втором станке. Генеральные дисперсии известны: D(X)=60 г^2, D(Y)=80 г^2. Требуется при уровне значимости 0,05 проверить нулевую гипотезу H_0: M(X)=M(Y) при конкурирующей гипотезе M(X)$\neq$M(Y). Предполагается, что случайные величины X и Y распределены нормально и выборки независимы.

Ответ: $z_{набл}$=2,5, $z_{кр}$=1,96. Нулевая гипотеза отвергается.

5.43. Выборка в 50 электроламп завода А показала среднюю продолжительность работы $\bar{x} = 1282$ ч со средним квадратическим отклонением s_1=80 ч, а такая же по объему выборка того же типа ламп с завода Б - $\bar{y} = 1208$ ч со средним квадратическим отклонением s_2=94 ч. Требуется, при уровне значимости

0,05 проверить нулевую гипотезу H_0: M(X)=M(Y) (средний срок службы ламп обоих заводов одинаков) при конкурирующей гипотезе M(X)>M(Y).

Ответ: нулевая гипотеза отклоняется.

5.44. По двум независимым малым выборкам, объемы которых n=10 и m=8, извлеченным из нормальных генеральных совокупностей, найдены выборочные средние: $\bar{x}$=142,3, $\bar{y}$=145,3 и исправленные дисперсии $s_x^2 = 2{,}7$ и $s_y^2 = 3{,}2$. При уровне значимости 0,01 проверить нулевую гипотезу H_0: M(X)=M(Y) при конкурирующей гипотезе H_1: M(X)$\neq$M(Y).

Ответ: $|T_{набл}|$=3,7; $t_{двуст.\ кр}$(0,01; 16)=2,92. Нулевая гипотеза отвергается.

5.45. Выдвинута гипотеза, что применение нового типа резца сокращает время обработки некоторой детали. Проведено 10 измерений времени, затрачиваемого на обработку этой детали старым и новым резцом. Получены следующие результаты (в минутах): старый тип резца - 58, 58, 56, 38, 70, 38, 42, 75, 68, 67; новый тип резца - 57, 55, 63, 24, 67, 43, 33, 68, 56, 54. Проверить гипотезу равенства среднего времени, затрачиваемого на изготовление этой детали с помощью двух типов резцов. Уровень значимости принять равным 0,05.

Ответ: нет оснований для отклонения нулевой гипотезы равенства среднего времени, затрачиваемого на изготовление детали двумя типами резцов.

5.46. При 50 подбрасываниях монеты герб появился 20 раз. Можно ли считать монету симметричной? Уровень значимости принять равным 0,10.

Ответ: да.

5.47. Используя критерий Пирсона, при уровне значимости 0,05 установить, случайно или значимо расхождение между эмпирическими частотами n_i и теоретическими частотами n_i', которые вычислены исходя из гипотезы о нормальном распределении генеральной совокупности X_i:

а)

n_i	5	10	20	8	7
n_i'	6	14	18	7	5;

б)

n_i	6	8	13	15	20	16	10	7	5
n_i'	5	9	14	16	18	16	9	9	7;

в)

n_i	14	18	32	70	20	36	10
n_i'	10	24	34	80	18	22	12.

Ответ: а) случайно; k=2, $\chi^2_{набл}$ =2,47, $\chi^2_{кр}$(0,05;2)=6,0; б) случайно; в) значимо; k=4, $\chi^2_{набл}$=13,93, $\chi^2_{кр}$(0,05; 4)=9,5.

5.48. Используя критерий Пирсона, при уровне значимости 0,05 проверить, согласуется ли гипотеза о нормальном распределении генеральной совокупности X с эмпирическим распределением выборки объема n=200:

x_i	5	7	9	11	13	15	17	19	21
n_i	15	26	25	30	26	21	24	20	13.

Ответ: гипотеза о нормальном распределении отвергается.

5.49. На экзамене студент отвечает только на один вопрос по одной из трех частей курса. Анализ вопросов, заданных 60 студентам, показал, что 23 студента получили вопросы из первой, 15 - из второй и 22 - из третьей части курса. Можно ли считать, что студент, идущий на экзамен, с равной вероятностью получит вопрос по любой из трех частей курса?

Принять α =0,1.

Ответ: да.

5.50. При испытании радиоэлектронной аппаратуры фиксировалось число отказов. Результаты 59 испытаний приводятся ниже.

Число отказов	0	1	2	3
Число испытаний	42	10	4	3

Проверить гипотезу H_0 о том, что число отказов имеет распределение Пуассона, при α=0,10.

Ответ: гипотеза H_0 отклоняется.