МАТЕМАТИКА

НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ
СТРОИТЕЛЬНЫЙ УНИВЕРСИТЕТ

Кафедра прикладной математики

МАТЕМАТИЧЕСКАЯ СТАТИСТИКА. ОСНОВЫ ТЕОРИИ ВЕРОЯТНОСТЕЙ

Методические указания
к практическим занятиям и самостоятельной работе
для обучающихся бакалавриата по всем УГСН, реализуемым НИУ МГСУ

Составители:
В.Д. Петелина, Н.М. Чиганова, Е.М. Гусакова

Москва
Издательство МИСИ – МГСУ
2020

УДК 519.2
ББК 22.17
М 34

Рецензент — доктор физико-математических наук *Т.А. Мацеевич*,
заведующая кафедрой прикладной математики НИУ МГСУ

М34 **Математическая статистика. Основы теории вероятностей** [Электронный ресурс] : методические указания к практическим занятиям и самостоятельной работе для обучающихся бакалавриата по всем УГСН, реализуемым НИУ МГСУ / сост.: В.Д. Петелина, Н.М. Чиганова, Е.М. Гусакова ; Министерство науки и высшего образования Российской Федерации, Национальный исследовательский Московский государственный строительный университет, кафедра прикладной математики. — Электрон. дан. и прогр. (0,79 Мб). — Москва : Издательство МИСИ – МГСУ, 2020. — Режим доступа: http://lib.mgsu.ru. — Загл. с титул. экр.

Методические указания полностью соответствуют действующей рабочей программе по математике для бакалавров направления 08.03.01 Строительство и отражают опыт проведения практических занятий по разделам Теория вероятностей» и Математическая статистика» в МГСУ. Методические указания предназначены для самостоятельной работы студентов над практической частью курса, выполнения расчетного задания и подготовке к зачету.

Для обучающихся бакалавриата по всем УГСН, реализуемым НИУ МГСУ.

Учебное электронное издание

Редактор, корректор *Л.М. Волкова*
Компьютерная правка и верстка *Е.В. Жуковой*
Дизайн первого титульного экрана *Д.Л. Разумного*

Для создания электронного издания использовано:
Microsoft Word 2010, ПО Adobe Acrobat

Подписано к использованию 31.05.2020. Объем данных 0,79 Мб.

Федеральное государственное бюджетное
образовательное учреждение высшего образования
«Национальный исследовательский
Московский государственный строительный университет»
129337, Москва, Ярославское ш., 26

Издательство МИСИ – МГСУ
Тел.: (495) 287-49-14, вн. 13-71, (499) 188-29-75, (499) 183-97-95,
e-mail: ric@mgsu.ru, rio@mgsu.ru

ОГЛАВЛЕНИЕ

СЛУЧАЙНЫЕ СОБЫТИЯ

Основные понятия и определения

Стохастическим называется эксперимент (опыт, испытание), результат которого заранее (до его проведения) предугадать нельзя.

Случайным событием называется любое явление, которое может произойти или не произойти в результате стохастического эксперимента.

Пример 1. Проводится опыт с бросанием двух игральных костей (кубики, каждая грань которых имеет метки — очки, соответствующие цифрам 1, 2, 3, 4, 5 и 6). Результатом этого опыта — событием — может быть появление одной из пар чисел (1, 1) , (1, 2) ... (6, 5), (6, 6), где первые и вторые числа равны числу очков, выпавших, соответственно, на первой и второй костях. Можно рассматривать и другие события, заключающиеся, например, в том, что сумма выпавших очков равна пяти, четна, делится на три и т. д.

Система событий называется совокупностью элементарных событий, если в результате опыта происходит одно и только одно элементарное событие; ***каково бы ни было случайное событие A, по наступившему элементарному событию можно сказать о том, произошло или не произошло A.***

Элементарные события обозначают греческой буквой ω, снабженной при необходимости индексом (ω_i), а их совокупность (Ω) называют пространством элементарных событий.

В примере 1 в качестве элементарных событий можно рассматривать появление любой из пар чисел (a, b), где числа a и b равны числу очков, выпавших, соответственно, на первой и второй костях, причем могут принимать значения от 1 до 6. Всего в этом опыте имеется 36 элементарных событий.

Выбор элементарных событий определяется неоднозначно, что можно использовать при решении задач.

Если Ω — пространство элементарных событий рассматриваемого опыта и $\boldsymbol{A}$ — возможное событие, то ***совокупность всех элементарных событий, наступление которых необходимо влечет наступление A, называют благоприятствующими этому событию*** и обозначают тем же символом $\boldsymbol{A}$.

Событие называется достоверным (Ω)***, если оно наступает в результате появления любого элементарного события***.

Событие называется невозможным (∅)***, если оно не наступает ни при каком элементарном событии.***

Суммой (***объединением***) ***двух событий A и B называется событие A + B или*** (***A∪B***)***, происходящее тогда и только тогда, когда происходит или A, или B.***

Произведением (***пересечением***) ***двух событий A и B называется событие AB или*** (***A∩B***)***, происходящее тогда и только тогда, когда происходит и A, и B.***

Два события называются несовместными, если их одновременное появление в опыте невозможно. В этом случае $\boldsymbol{AB} = \varnothing$.

Событие $\overline{\boldsymbol{A}}$ ***называется противоположным к A, если оно происходит тогда и только тогда, когда не происходит событие A.***

Справедливы следующие свойства: $\overline{A} + A = \Omega$, $\overline{A}A = \varnothing$, $\overline{\overline{A}} = A$

Для лучшего восприятия введенных понятий и операций полезна геометрическая интерпретация — диаграммы Венна: пространство элементарных событий Ω изображается в виде квадрата, каждой точке которого соответствует элементарное событие. Тогда случайные события изображаются в виде некоторых фигур, лежащих в этом квадрате.

Пример 2. Случайные события $\boldsymbol{A}$ и $\boldsymbol{B}$ — некоторые фигуры на квадрате. Указать (заштриховать соответствующую область) события $\boldsymbol{A + B}$, $\boldsymbol{AB}$, $\boldsymbol{A\overline{B}}$, $\overline{A}$ (рис. 1).

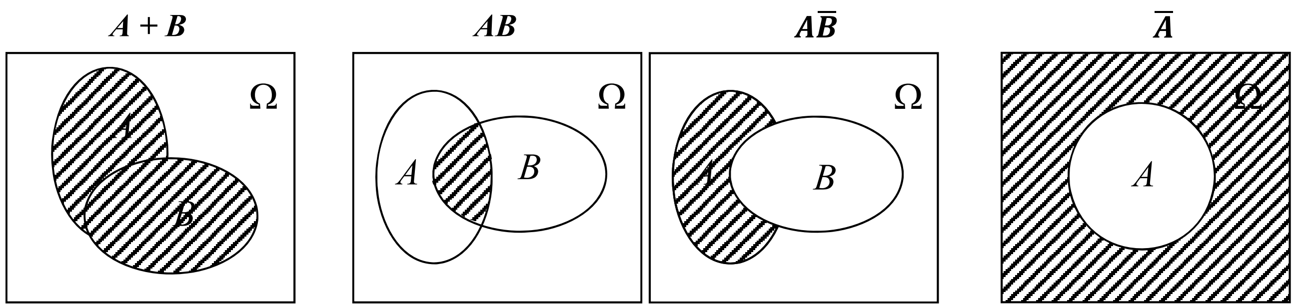

Рис. 1

Пример 3. Производятся два выстрела по цели. Пусть событие A — попадание в цель при первом выстреле, B — при втором, тогда $\overline{A}$ и $\overline{B}$ промах, соответственно, при первом и втором выстрелах. Пусть событие C — поражение цели, при условии что для этого достаточно хотя бы одного попадания. Выразить C через A и B.

Решение. Цель будет поражена в следующих случаях: попадание при первом и промах при втором, промах при первом и попадание при втором, попадание при первом и втором выстрелах. Интересующее нас событие заключается в наступлении или первого, или второго, или третьего вариантов (хотя бы одного). Используя введенные выше операции, получим: $C = A\overline{B} + \overline{A}B + AB$. С другой стороны, событие $\overline{C}$, противоположное C, есть промах при двух выстрелах, то есть $\overline{C} = \overline{A} \cdot \overline{B}$, отсюда искомое событие C можно записать в виде $C = \overline{\overline{A} \cdot \overline{B}}$. Возможность различного выражения искомого события часто оказывается полезной при решении задач.

Пусть Ω — пространство элементарных событий, соответствующих стохастическому эксперименту и пусть F — некоторая система случайных событий. ***Система событий F называется алгеброй событий, если выполняются условия***: $\Omega \in F$; ***из того, что A и B*** $\in$ ***F следует, что***: $\overline{A}$ ***∊ F, A + B*** $\in$ ***F и A · B*** $\in$ ***F.*** Следовательно, применяя любые из введенных операций к произвольной системе событий из F, получим событие, также принадлежащее F.

Определение вероятности

Если при многократном проведении одного и того же стохастического эксперимента событие A произошло $m(A)$ раз, то ***относительной частотой называется отношение*** $\nu(A) = \dfrac{m(A)}{n}$, ***где n — число проведенных опытов***.

Если при увеличении числа опытов относительная частота события $\nu(A)$ ***стремится к некоторому фиксированному числу p(A), то событие A стохастически устойчиво, а это число называют вероятностью события A.***

Если пространство элементарных событий стохастического эксперимента состоит из конечного или бесконечного (но счетного) множества элементарных событий и для каждого из них известна вероятность $P(\omega_i) = p_i$, то ***вероятностью случайного события A называется сумма вероятностей элементарных событий, благоприятствующих A***:

$$P(A) = \sum_{(\omega_i \in A)} p_i.$$

Если множество элементарных событий стохастического эксперимента конечно и они равновозможны (в этом случае их принято называть исходами), приходим к классическому определению: ***вероятностью случайного события A называется отношение числа исходов, благоприятствующих A, к общему числу исходов:*** $P(A) = \dfrac{m(A)}{n}$.

При подсчете числа исходов часто используют некоторые правила и формулы комбинаторики.

Правило произведения: ***если из некоторого множества A элемент*** a_i ***можно выбрать*** k_A ***способами, а элемент*** b_j ***из множества B —*** k_B ***способами, то совокупность (***a_i, b_j***) можно***

выбрать $k_A \cdot k_B$ способами. Правило верно и для совокупностей, состоящих из большего числа элементов.

Пример 4. Сколькими способами можно набрать семизначный номер телефона, если все его цифры различны.

Решение. Очевидно, первую цифру можно набрать 10 способами, вторую — девятью, так как одна цифра уже использована, ... седьмую — четырьмя. Согласно правилу произведения, общее число возможных номеров равно $10 \cdot 9 \cdot 8 \cdot 7 \cdot 6 \cdot 5 \cdot 4 = 604\,800$.

Если из некоторого множества, состоящего из ***n*** различимых элементов, отбираются в определенном порядке ***m*** элементов, то возможные варианты называют размещениями из ***n*** элементов по ***m,*** и их число равно

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

При ***n*** = ***m*** говорят о перестановках из ***n*** элементов, их число равно $P_n = n!$

Если порядок отбираемых ***m*** элементов из ***n*** элементов не играет роли, то говорят о сочетаниях из ***n*** элементов по ***m,*** и их число равно

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

Пример 5. Сколькими способами можно из 20 присяжных заседателей отобрать трех для участия в судебном процессе.

Решение. Поскольку не существенно, в каком порядке отобраны кандидатуры, число вариантов равно $C_{20}^3 = \dfrac{20 \cdot 19 \cdot 18}{3!} = 1140$.

Пример 6. Сколькими способами можно из 20 членов правления фирмы отобрать трех для замещения вакансий вице-президентов, отвечающих за производство, финансы и реализацию продукции.

Решение. Поскольку порядок при таком выборе играет существенную роль, то число вариантов равно $A_{20}^3 = 20 \cdot 19 \cdot 18 = 6840$.

При большом ***n*** подсчет числа вариантов по этим формулам требует громоздких вычислений $\boldsymbol{n}!$, в этом случае пользуются асимптотической формулой Стирлинга $\boldsymbol{n}! \approx \left(\dfrac{n}{e}\right)^n \cdot \sqrt{2\pi n}$.

Пример 7. К экзамену подготовлены 30 теоретических вопросов и 50 задач. Определить вероятность того, что студент получит «отлично» (для этого надо правильно ответить на два вопроса и решить три задачи, выбранные случайным образом), если студент выучил 20 вопросов и умеет решать 30 задач.

Решение. В качестве пространства элементарных событий этого опыта возьем множество всех наборов из двух вопросов и трех задач. Поскольку выбор случаен, то все исходы равновозможны и применимо классическое определение вероятности. Для подсчета ***n*** — числа исходов заметим, что два теоретических вопроса можно выбрать C_{30}^2, а три задачи C_{50}^3 способами (порядок следования здесь не важен). По правилу произведения, общее число таких наборов будет равно $n = C_{30}^2 \cdot C_{50}^3$.
Событие ***A*** — отличная оценка реализуется тогда, когда оба вопроса будут из 20 выученных и все три задачи из 30 ему известных. Число таких наборов — благоприятствующих ***A*** исходов $m(A) = C_{20}^2 \cdot C_{30}^3$ — находится аналогично. Поэтому искомая вероятность равна

$$P(A) = \frac{m(A)}{n} = \frac{C_{20}^2 \cdot C_{30}^3}{C_{30}^2 \cdot C_{50}^3} = \frac{20 \cdot 19 \cdot 30 \cdot 29 \cdot 28}{30 \cdot 29 \cdot 50 \cdot 49 \cdot 48} = 0,09.$$

Пример 8. Среди K поставленных единиц данного товара L не удовлетворяют предъявляемым условиям. Найти вероятность того, что среди $k \le K$, отобранных для выборочного контроля качества, ровно $l \le L$ не будут удовлетворять этим требованиям.

Решение. Опыт заключается в случайном отборе k образцов. Следовательно, исходы этого испытания равновозможны и их общее число равно $n = C_K^k$. Событие $\boldsymbol{A}$ состоит в том, что из k отобранных ровно l не будут удовлетворять этим требованиям. Число исходов, благоприятствующих $\boldsymbol{A}$, согласно правилу произведения, равно $m(A) = C_{K-L}^{k-l} \cdot C_L^l$, здесь первый множитель дает число вариантов отбора хороших, а второй — плохих образцов. Отсюда искомая вероятность равна

$$P(A) = \frac{m(A)}{n} = \frac{C_{K-L}^{k-l} \cdot C_L^l}{C_K^k}.$$

Аксиоматическое определение вероятности. ***Пусть Ω — пространство элементарных событий некоторого стохастического эксперимента и в Ω выделена система F событий, являющаяся алгеброй событий. Это означает, что выполняются условия***: $\Omega \in F$; ***из того, что A и $B \in F$, следует, что $\overline{A}F$, $A+B \in F$ и $A \cdot B \in F$. Пусть каждому событию $A \in F$ поставлено в соответствие число $P(A)$ и верны свойства $P(A) \ge 0$ для любого $A \in F$; $P(\Omega) = 1$; если A и B не совместны($(A \cdot B = \varnothing)$), то $P(A + B) = P(A) + P(B)$, P(A) называется вероятностью случайного события A.***

Геометрическое определение вероятности события. Рассматривается стохастический эксперимент, заключающийся в бросании случайным образом точки на некоторую фигуру (отрезок прямой или линии, плоскую область, трехмерное тело, временной интервал и т. д.), — достоверное событие Ω, причем любое положение точки на фигуре равновозможно. В этом случае вероятностью случайного события $\boldsymbol{A}$ попадания точки в область, соответствующую $\boldsymbol{A}$, называется отношение мер области $\boldsymbol{A}$ и фигуры Ω: $P(A) = \dfrac{L(A)}{L(\Omega)} = \dfrac{S(A)}{S(\Omega)} = \dfrac{V(A)}{V(\Omega)}$.

Пример 9. Два лица, M и D, договорились встретиться в определенном месте между 19 и 20 часами, причем появление любого из них равновозможно в любой момент этого часа. Найти вероятность того, что встреча состоится, если первым на место встречи приходит M, то ждет не более 20 минут; первой приходит D, то ждет не более 10 минут.

Решение. Пусть моменты прихода (отсчет времени будем проводить в минутах от начала часа) M и D на место встречи будут x и y. Интересующее нас событие A произойдет, если будет выполнена система неравенств:

$$\begin{cases} 0 \le x - y \le 20 \\ 0 \le y - x \le 10 \end{cases} \Rightarrow \begin{cases} x - 20 \le y \le x \\ x \le y \le 10 + x, \end{cases}$$

где x и y определяют координаты точки в квадрате со стороной 60. По условию, любое положение этой точки в квадрате равновозможно, поэтому вероятность события A — попадания точки в область A (рис. 2), согласно геометрической схеме, равна

$$P(A) = \frac{S(A)}{S(\Omega)} = \frac{60 \cdot 60 - 0,5 \cdot 50 \cdot 50 - 0,5 \cdot 40 \cdot 40}{60 \cdot 60} = 0,43.$$

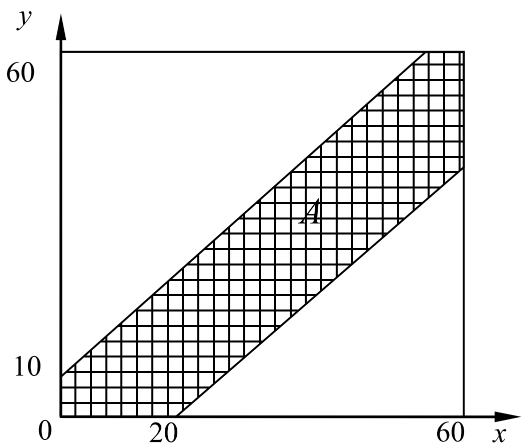

Рис. 2

Некоторые теоремы и формулы

Вероятность противоположного события $P(\overline{A}) = 1 - P(A)$. Следствие $P(\varnothing) = 1 - P(\Omega) = 0$.

Вероятность суммы двух событий $P(A + B) = P(A) + P(B) - P(AB)$.

Следствие: если события несовместны, то $P(A+B)=P(A)+P(B)$, формула справедлива для любого конечного числа несовместных событий.

Условная вероятность события В, при условии что событие А произошло, равна $P(B|A)=\dfrac{P(AB)}{P(A)}$.

Пример 10. Из 30 экзаменационных билетов студент подготовил только 25. Если он отказывается отвечать по первому взятому билету (который он не знает), то ему разрешается взять второй. Определить вероятность того, что второй билет окажется счастливым.

Решение. Пусть событие ***A*** заключается в том, что первый вытащенный билет оказался для студента «плохой», а ***B*** — «хороший». Поскольку после наступления события ***A*** один из «плохих» уже извлечен, то остается всего 29 билетов, из которых 25 студент знает. Отсюда искомая вероятность, предполагая, что появление любого билета равновозможно и они обратно не возвращаются, равна $P(B|A)=\dfrac{25}{29}$.

Вероятность произведения $P(AB)=P(A)\cdot P(B|A)$.

Следствие: формула обобщается на любое конечное число сомножителей (для трех она имеет вид $P(ABC)=P(A)\cdot P(B|A)\cdot P(C|AB)$.

Пример 11. По условиям примера 10 найти вероятность успешной сдачи экзамена, если для этого студент должен ответить на первый билет или, не ответив на первый, обязательно ответить на второй.

Решение. Пусть события A и B заключаются в том, что, соответственно, первый и второй билеты «хорошие». Тогда $\overline{A}$ — появление «плохого» билета в первый раз. Экзамен будет сдан, если произойдет событие A, или одновременно $\overline{A}$ и B. То есть искомое событие C — успешная сдача экзамена — выражается следующим образом: $C=A+\overline{A}B$. Отсюда

$$P(C)=P(A+\overline{A}B)=P(A)+P(\overline{A}B)=P(A)+P(\overline{A})\cdot P(B|\overline{A})=\frac{25}{30}+\frac{5}{30}\cdot\frac{25}{29}=0{,}997.$$

Здесь мы воспользовались несовместностью A и $\overline{A}$, а следовательно, несовместностью A и $\overline{A}B$, формулами вероятностей суммы, произведения и классическим определением вероятности.

Эту задачу можно решить и проще, если воспользоваться вероятностью противоположного события

$$P(C)=1-P(\overline{C})=1-P(\overline{A}\overline{B})=1-P(\overline{A})\cdot P(\overline{B}|\overline{A})=1-\frac{5}{30}\cdot\frac{4}{29}=0{,}997.$$

Независимость событий. ***Два события называются независимыми,*** *если* $P(AB)=P(A)\cdot P(B)$. Следствие: $P(B|A)=P(B)$ и $P(A|B)=P(A)$ — наступление одного события не изменяет вероятности появления другого, то есть условная вероятность равна безусловной. На практике пользуются правилом, согласно которому ***из физической независимости событий следует их независимость в теоретико-вероятностном смысле***.

Пример 12. Абонент забыл последние три цифры нужного ему телефонного номера, но помнит, что все они нечетные. Найти вероятность того, что ему удастся дозвониться с первого раза.

Решение. Пусть события A, B и C заключаются в том, что, соответственно, первая, вторая и третья забытые цифры будут набраны верно, тогда вероятность успеха — события $D = ABC$ (должны произойти и A, и B, и C) будет равна

$$P(D)=P(ABC)=P(A)\cdot P(B)\cdot P(C)=\frac{1}{5}\cdot\frac{1}{5}\cdot\frac{1}{5}=0{,}008.$$

Здесь мы воспользовались независимостью событий, классическим определением и тем, что из пяти нечетных цифр подходит в каждом случае только одна.

Система событий $A_1, A_2, \ldots, A_n$ ***называется независимой в совокупности, если вероятность произведения равна произведению вероятностей для любой комбинации сомножителей из этой системы.***

Формулы полной вероятности и Байеса. Пусть события $H_1, H_2, \ldots, H_n$ попарно несовместны и в сумме образуют достоверное событие (в этом случае их называют гипотезами). Эти события образуют полную группу событий. Пусть теперь интересующее нас событие A наступает после реализации одной из гипотез H_i и известны вероятности $P(H_i)$, $P(A\,|\,H_i)$. В этом случае вероятность события A вычисляется по формуле полной вероятности

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

Если дополнительно стало известно, что событие A произошло, то по формуле Байеса можно определить вероятность того, что при этом была реализована H_k гипотеза (эту вероятность гипотезы, полученную после проведения опыта, называют апостериорной)

$$P(H_k\,|\,A) = \frac{P(H_k) \cdot P(A\,|\,H_k)}{\sum\limits_{i=1}^{n} P(H_i) \cdot P(A\,|\,H_i)}.$$

Пример 13. 30 % пациентов, поступивших в больницу, принадлежат к первой социальной группе, 20 % — ко второй и 50 % — к третьей. Вероятность заболевания туберкулезом для представителей этих групп, соответственно, равны 0,02; 0,03 и 0,01. Определить вероятность того, что поступивший больной болен туберкулезом. Проведенные анализы у поступившего больного показали наличие туберкулеза. Найти вероятность того, что это представитель третьей группы.

Решение. Пусть H_1, H_2, H_3 — гипотезы, заключающиеся в том, что пациент принадлежит, соответственно, первой, второй и третьей группам. Очевидно, они образуют полную группу событий, причем $P(H_1) = 0{,}3$, $P(H_2) = 0{,}2$, $P(H_3) = 0{,}5$. Вероятность того, что пациент болен, определяем по формуле полной вероятности

$$P(A) = P(H_1) \cdot P(A\,|\,H_1) + P(H_2) \cdot P(A\,|\,H_2) + P(H_3) \cdot P(A\,|\,H_3) =$$

$$= 0{,}3 \cdot 0{,}02 + 0{,}2 \cdot 0{,}03 + 0{,}5 \cdot 0{,}01 = 0{,}017.$$

Вероятность того, что поступивший и оказавшийся больным пациент принадлежит третьей группе, определяется формулой Байеса

$$P(H_3|A) = \frac{P(H_3) \cdot P(A\,|\,H_3)}{P(A)} = \frac{0{,}5 \cdot 0{,}01}{0{,}017} = \frac{5}{17}.$$

Последовательность независимых однородных испытаний. Формула Бернулли. Формулы Муавра — Лапласа

Стохастический эксперимент заключается в проведении n независимых, однородных испытаний в каждом из которых событие A может появиться с вероятностью $p = P(A)$ и не появиться с вероятностью $q = P(\overline{A}) = 1 - P(A) = 1 - p$. Вероятность того, что событие A появится ровно m раз при n испытаниях, определяется по формуле Бернулли $P_n(m) = C_n^m \cdot p^m \cdot q^{n-m}$, а вероятность того, что событие A появится не менее m_1 и не более m_2 раз при n испытаниях, определяется по формуле $P_n(m_1 \le m \le m_2) = \sum\limits_{m=m_1}^{m_2} C_n^m \cdot p^m \cdot q^{n-m}$.

Пример 14. Вероятность того, что образец бетона выдержит нормативную нагрузку, равна 0,9, Найти вероятность того, что из семи образцов испытания выдержат: ровно пять, не менее пяти.

Решение. Очевидно, имеет место схема Бернулли, поэтому

$$P_7(5)=C_7^5\cdot 0,9^5\cdot 0,1^2=0,124,$$

$$P_7(5\le m\le 7)=P_7(5)+P_7(6)+P_7(7)=0,974.$$

При достаточно большом n и не слишком малых p и q формулой Бернулли пользоваться практически невозможно и применяются асимптотические формулы Муавра — Лапласа, которые дают хорошее приближение при $npq > 10$.

Локальная: $P_n(m)\approx\dfrac{1}{\sqrt{npq}}\cdot\varphi(x)$,

где $\varphi(x)=\dfrac{1}{\sqrt{2\pi}}\cdot e^{-\frac{x^2}{2}}$ и $x=\dfrac{m-np}{\sqrt{npq}}$, $\varphi(x)=\varphi(-x)$.

Интегральная: $P_n(m_1\le m\le m_2)\approx\Phi(x_2)-\Phi(x_1)$,

где $\Phi(x)=\dfrac{1}{\sqrt{2\pi}}\int\limits_0^x e^{-\frac{t^2}{2}}dt,\;\; x_1=\dfrac{m_1-np}{\sqrt{npq}},\;\; x_2=\dfrac{m_2-np}{\sqrt{npq}},\;\; \Phi(-x)=-\Phi(x).$

Обе функции $\varphi(x)$ и $\Phi(x)$ табулированы (приложения 1 и 2 [1]).

Пример 15. При установившемся технологическом процессе ЖБК выпускает 80 % всех изделий первым сортом. Найти вероятность, что из 100 поставленных изделий первосортных будет ровно 75, не менее 75.

Решение. Поскольку $n = 100$ велико, $p = 0,8$ и $q = 0,2$ не малы, применяем локальную и затем интегральную формулы Муавра — Лапласа

$$P_{100}(75)\approx\frac{1}{\sqrt{npq}}\cdot\varphi(x)=\frac{1}{\sqrt{100\cdot 0,8\cdot 0,2}}\cdot\varphi\left(\frac{75-100\cdot 0,8}{\sqrt{100\cdot 0,8\cdot 0,2}}\right)=$$

$$=0,25\cdot\varphi(-1,25)=0,25\cdot\varphi(1,25)=0,25\cdot 0,1826=0,0457.$$

$$P_{100}(m\ge 75)=P_{100}(75\le m\le 100)\approx\Phi\left(\frac{100-100\cdot 0,8}{\sqrt{100\cdot 0,8\cdot 0,2}}\right)-\Phi\left(\frac{75-100\cdot 0,8}{\sqrt{100\cdot 0,8\cdot 0,2}}\right)=$$

$$=\Phi(5)-\Phi(-1,2)=\Phi(5)+\Phi(1,2)=0,5+0,385=0,885.$$

Пример 16. Известно, что 30 % призывников имеют 45 размер обуви. Сколько пар обуви надо иметь на складе воинской части, чтобы с вероятностью $p_0=0,9$ были обеспечены все такие призывники, если в часть прибыло 200 новобранцев?

Решение. Очевидно, имеет место схема Бернулли: подбор пары обуви каждому призывнику — одно из 200 испытаний, причем вероятность того, что ему потребуется обувь 45 размера, равна $p = 0,3$ ($q = 0,7$). Пусть на складе имеется k пар обуви, где k пока не известно. Требуется подобрать такое k, чтобы $P_{200}\left(0\le m\le k\right)\ge p_0$. Поскольку $n = 200$ велико, а p и q не малы, применяем интегральную формул Муавра — Лапласа

$$P_{200}(0\le m\le k)\approx\Phi(x_2)-\Phi(x_1)=\Phi\left(\frac{k-np}{\sqrt{npq}}\right)-\Phi\left(\frac{0-np}{\sqrt{npq}}\right)=$$

$$=\Phi\left(\frac{k-200\cdot 0,3}{\sqrt{200\cdot 0,3\cdot 0,7}}\right)-\Phi\left(\frac{0-200\cdot 0,3}{\sqrt{200\cdot 0,3\cdot 0,7}}\right)=\Phi\left(\frac{k-60}{6,48}\right)-\Phi\left(\frac{-60}{6,48}\right)=$$

$$=\Phi\left(\frac{k-60}{6,48}\right)+\Phi(9,26)=\Phi\left(\frac{k-60}{6,48}\right)+0,5\ge 0,9.$$

Отсюда $\Phi\left(\frac{k-60}{6,48}\right)+0,5\geq 0,9\Rightarrow\frac{k-60}{6,48}\geq 1,28\Rightarrow k\geq 68,294.$

То есть на складе достаточно иметь 69 пар обуви такого размера, чтобы с вероятностью 0,9 обеспечить спрос.

Закон редких событий (формула Пуассона) и простейший стационарный (пуассоновский) поток событий

Если в схеме Бернулли число испытаний n велико, а вероятность появления события A при одном испытании $P(A)$ мала (или мало q), то применяется формула Пуассона (в этом случае говорят о законе редких событий)

$$P_n(m)\approx\frac{a^m}{m!}\cdot e^{-a},\quad P_n(m\leq k)\approx e^{-a}\cdot\sum_{m=0}^{k}\frac{a^m}{m!},\quad a=np\leq 10.$$

Если на некоторой прямой расположены точки так, что в среднем на единицу длины приходится μ точек и вероятность того, что одна точка окажется на отрезке длины l, зависит только от его длины и не зависит от его расположения на прямой. В этом случае вероятность того, что на искомом отрезке окажется ровно m точек (не более k), определяется формулой Пуассона

$$P_n(m)=\frac{a^m}{m!}\cdot e^{-a},\quad P_n(m\leq k)=e^{-a}\cdot\sum_{m=0}^{k}\frac{a^m}{m!},\quad a=\mu l.$$

Обе формулы табулированы (таблицы 1 и 2).

Пример 17. На факультете учатся 500 студентов. Найти вероятность того, что первое сентября является днем рождения: трех студентов, не менее трех.

Решение. Пусть событие A — случайно выбранный студент родился первого сентября, тогда $P(A)=p=\frac{1}{365}$. В результате пришли к схеме Бернулли, где число испытаний n = 500 велико, а p мало (события редкие) и при np = 1,37 < 10, поэтому применяем формулы Пуассона

$$P_{500}(3)\approx\frac{1,37^3}{3!}\cdot e^{-1,37}=0,11,\quad P_{500}(m\geq 3)=1-P_{500}(m\leq 2)\approx 1-0,84.$$

Здесь мы воспользовались таблицами 1 и 2 и во втором случае для этого перешли к противоположному событию.

Пример 18. Известно, что в среднем за месяц (30 суток) в районной сети водоснабжения возникает 90 ситуаций, требующих оперативного вмешательства аварийной службы. На сколько вызовов в сутки должна быть рассчитана эта служба, чтобы с вероятностью p_o = 0,9 она могла удовлетворить все поступающие за эти сутки заявки?

Решение. Предположим, что аварийная служба рассчитана на k заявок в сутки, где k пока не известно. Пусть m — число поступивших за сутки. Тогда k найдем из условия $P(m\leq k)\geq p_0$. Поскольку поток заявок представляет собой простейший стационарный (пуассоновский) поток событий, то можно применить формулу Пуассона

$$P_n(m\leq k)=e^{-a}\cdot\sum_{m=0}^{k}\frac{a^m}{m!},$$

где $a=\frac{90}{30}=3$ — среднее число заявок за сутки. Для определения k воспользуемся таблицей 2 при a = 3, подбирая k таким образом, чтобы искомая вероятность была не меньше чем p_o = 0,9. В результате получим k = 5, то есть аварийная служба должна быть рассчитана на пять заявок в сутки.

СЛУЧАЙНЫЕ ВЕЛИЧИНЫ

Основные понятия и определения

Величина называется случайной, если ее численное значение может меняться в зависимости от результата стохастического эксперимента.

Дискретной называется случайная величина, множество возможных значений которой конечно или счетно.

Законом распределения дискретной случайной величины называется правило, согласно которому каждому возможному значению ставится в соответствие вероятность, с которой случайная величина может принять это значение. Обычно закон распределения дискретной случайной величины задается в виде таблицы или некоторой формулой

ξ	x_1	x_2	…	x_n
p	p_1	p_2	…	p_n

$$p_i = P(\xi = x_i),$$

где

$$\sum_{i=1}^{n} p_i = 1 \left(\sum_{i=1}^{\infty} p_i = 1 \right).$$

Функция распределения дискретной случайной величины и ее числовые характеристики — математическое ожидание, дисперсия и среднеквадратическое отклонение — определяются по формулам

$$F(x) = P(\xi < x) = \sum_{(x_i < x)} p_i; \quad M(\xi) = a = \sum_{i=1}^{n} x_i \cdot p_i;$$

где $(x_i < x)$ означает, что суммирование распространяется на все значения p_i, для которых это условие выполняется.

Пример 19. Прибор состоит из четырех одинаковых узлов, один из которых вышел из строя. Для устранения неисправности случайно выбранный узел заменяется на имеющийся в запасе заведомо исправный и прибор испытывается. Если неисправность не устранена, то заменяется на исправный один из оставшихся, и так до тех пор, пока прибор не заработает. Составить закон распределения случайной величины ξ — числа замененных узлов, найти функцию распределения, построить ее график и найти числовые характеристики.

Решение. Очевидно, случайная величинаξ — дискретная и может принимать значения 1, 2, 3 и 4. Пусть A_1, A_2, A_3, A_4 — события, состоящие в том, что неисправный узел обнаружен, соответственно, с первой, второй, третьей, четвертой попыток. Найдем вероятности, соответствующие возможным значениям:

$$P(\xi = 1) = P(A_1) = \frac{1}{4};$$

$$P(\xi = 2) = P(\overline{A}_1 \cdot A_2) = P(\overline{A}_1) \cdot P(A_2 \mid \overline{A}_1) = \frac{3}{4} \cdot \frac{1}{3} = \frac{1}{4};$$

$$P(\xi = 3) = P()\left(\overline{A}_1 \cdot \overline{A}_2 \cdot A_3\right) = P(\overline{A}_1) \cdot P(\overline{A}_2 \mid \overline{A}_1) \cdot P(A_3 \mid \overline{A}_1 \cdot \overline{A}_2) = \frac{3}{4} \cdot \frac{2}{3} \cdot \frac{1}{2} = \frac{1}{4};$$

Аналогично находим $P(\xi = 4) = \dfrac{1}{4}$.

Закон распределения случайной величины ξ запишем в виде таблицы:

ξ	1	2	3	4
p	0,25	0,25	0,25	0,25

По полученным данным находим выражение для функции распределения, строим ее график (рис. 3) и определяем числовые характеристики:

$$M(\xi) = a = \sum_{i=1}^{4} x_i \cdot p_i = 1 \cdot 0,25 + 2 \cdot 0,25 + 3 \cdot 0,25 + 4 \cdot 0,25 = 2,5;$$

$$D(\xi) = M(\xi^2) - a^2 = \sum_{i=1}^{4} x_i^2 \cdot p_i - a^2 = 1 \cdot 0,25 + 4 \cdot 0,25 + 9 \cdot 0,25 + 16 \cdot 0,25 - 6,25 = 1,25$$

$$\sigma(\xi) = \sqrt{D(\xi)} = 1,118$$

$$F(x) = \begin{cases} 0, & x \in (-\infty, 1] \\ 0,25, & x \in (1,2] \\ 0,5, & x \in (2,3] \\ 0,75, & x \in (3,4] \\ 1, & x \in (4, \infty) \end{cases}$$

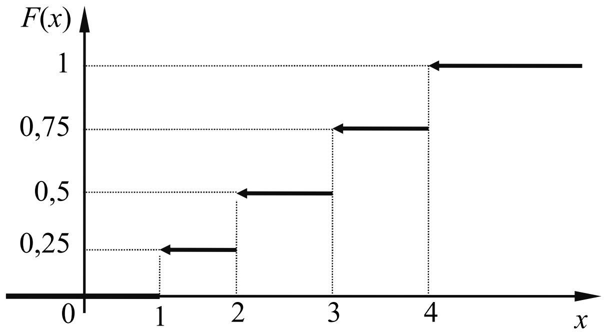

Рис. 3

Случайная величина называется непрерывной, если ее значения целиком заполняют некоторый интервал и ее функцию распределения можно представить в виде:

$$F(x) = P(\xi < x) = \int_{-\infty}^{x} f(t)dt \Rightarrow f(x) = F'(x),$$

где $f(x)$ — плотность вероятности является непрерывной или кусочно-непрерывной на всей числовой оси функцией.

Функция распределения и плотность вероятности обладают следующими свойствами:

$$0 \le F(x) \le 1; \quad F(-\infty) = 0; \quad F(\infty) = 1; \quad f(x) \ge 0;$$

$$\int_{-\infty}^{\infty} f(x)dx = F(\infty) = 1; \quad P(x_1 \le \xi \le x_2) = \int_{x_1}^{x_2} f(x)dx = F(x_2) - F(x_1).$$

Числовые характеристики непрерывной случайной величины определяются по формулам:

$$M(\xi) = a = \int_{-\infty}^{\infty} x \cdot f(x)dx; \quad D(\xi) = M\left((\xi - a)^2\right) = M(\xi^2) - a^2;$$

$$\sigma(\xi) = \sqrt{D(\xi)}.$$

Свойства математического ожидания

Математическое ожидание постоянной равно самой постоянной.

Постоянный множитель можно выносить за знак математического ожидания.

Математическое ожидание суммы конечного числа случайных величин равно сумме математических ожиданий слагаемых.

Математическое ожидание произведения конечного числа независимых случайных величин равно произведению математических ожиданий сомножителей.

Свойства дисперсии

Дисперсия постоянной равна нулю.

Постоянный множитель можно выносить за знак дисперсии, возведя его в квадрат.

Дисперсия суммы (и разности) конечного числа независимых случайных величин равна сумме дисперсией слагаемых.

Пример 20. Функция распределения непрерывной случайной величины задана следующим образом:

$$F(x)=\begin{cases} A, & x\in(-\infty,\ 0) \\ B+C\cdot\cos 3x, & x\in\left[0,\ \dfrac{\pi}{3}\right] \\ D, & x\in\left(\dfrac{\pi}{3},\ \infty\right). \end{cases}$$

Определить параметры A, B, C, D, найти выражение для плотности вероятности, числовые характеристики и вероятность того, что случайная величина примет значение в интервале $\left[-1;\ \dfrac{\pi}{9}\right]$. Построить графики $F(x)$ и $f(x)$.

Решение. Для определения параметров воспользуемся предельными значениями функции распределения и ее непрерывностью:

$$F(-\infty)=A=0;\ F(\infty)=D=1;\ F(0)=0=B+C\cdot\cos 0=B+C$$

$$F\left(\frac{\pi}{3}\right)=1=B+C\cdot\cos\pi=B-C\Rightarrow B=0,5;\ C=-0,5.$$

Теперь находим плотность вероятности, числовые характеристики и вероятность попадания в заданный интервал:

$$f(x)=F'(x)=\begin{cases} 1,5\cdot\sin 3x, & x\in\left[0,\ \dfrac{\pi}{3}\right] \\ 0, & x\notin\left[0,\ \dfrac{\pi}{3}\right] \end{cases}$$

$$M(\xi)=a=\int_{-\infty}^{\infty}x\cdot f(x)dx=1,5\cdot\int_{0}^{\frac{\pi}{3}}x\cdot\sin 3xdx=1,5\rangle\left(-\frac{1}{3}x\cos 3x+\frac{1}{9}\sin 3x\right)\Bigg|_{0}^{\frac{\pi}{3}}=\frac{\pi}{6};$$

$$D(\xi)=M(\xi^2)-a^2=1,5\cdot\int_{0}^{\frac{\pi}{3}}x^2\cdot\sin 3xdx-\left(\frac{\pi}{6}\right)^2=\frac{\pi^2-4}{18}-\frac{\pi^2}{36}=\frac{\pi^2-8}{36}.$$

$$P\left(-1\le\xi\le\frac{\pi}{9}\right)=F\left(\frac{\pi}{9}\right)-F(-1)=0,5-0,5\cdot\cos\frac{\pi}{3}-0=0,25.$$

Строим графики.

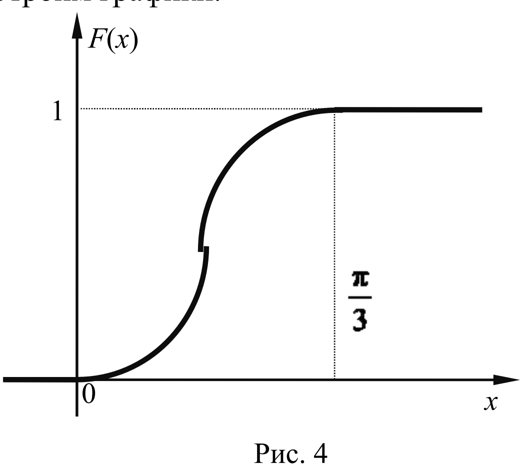

Рис. 4

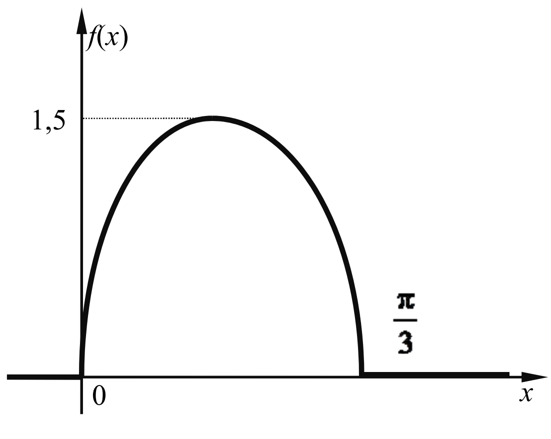

Рис. 5

При вычислении интегралов мы разбили интервал интегрирования на три, учли, что на двух из них плотность вероятности равна нулю, и воспользовались таблицей интегралов.

Пример 21. Плотность вероятности непрерывной случайной величины задана следующим образом: $f(x)=\begin{cases}0, & x\in(-\infty,1)\\ \dfrac{A}{x^4}, & x\in[1,\infty).\end{cases}$

Определить величину параметра A, найти функцию распределения, числовые характеристики и вероятность попадания случайной величины в интервал [0, 2]. Построить графики функции распределения и плотности вероятности.

Решение. Определяем величину параметра из условия:

$$\int_{-\infty}^{\infty} f(x)dx=1\Rightarrow A\cdot\int_{1}^{\infty} x^{-4}dx=A\cdot\left(-\frac{1}{3}\cdot x^{-3}\right)\Bigg|_{1}^{\infty}=\frac{A}{3}\Rightarrow A=3.$$

Здесь мы разбили интервал интегрирования на два, учтя значение плотности вероятности на каждом из них, и воспользовались правилом вычисления несобственного интеграла. Аналогично находим:

$$F(x)=\int_{-\infty}^{x} f(t)dt=\begin{cases}0,\ x\in(-\infty,1)\\ 3\cdot\int_{1}^{x} t^{-4}dt=\left(-t^{-3}\right)\Big|_{1}^{x}=1-\dfrac{1}{x^3},\ x\in[1,\infty)\end{cases}$$

$$M(\xi)=a=\int_{-\infty}^{\infty} x\cdot f(x)dx=3\cdot\int_{1}^{\infty} x\cdot x^{-4}dx=\left(-1,5\cdot x^{-2}\right)\Big|_{1}^{\infty}=1,5;$$

$$D(\xi)=M\left(\xi^2\right)-a^2=3\cdot\int_{1}^{\infty} x^2\cdot x^{-4}dx-1,5^2=3\cdot\left(-\frac{1}{x}\right)\Bigg|_{1}^{\infty}-2,25==3-2,25=0,75$$

$$P(0\le\xi\le 2)=F(2)-F(0)=\left(1-\frac{1}{8}\right)-0=0,875.$$

Строим графики:

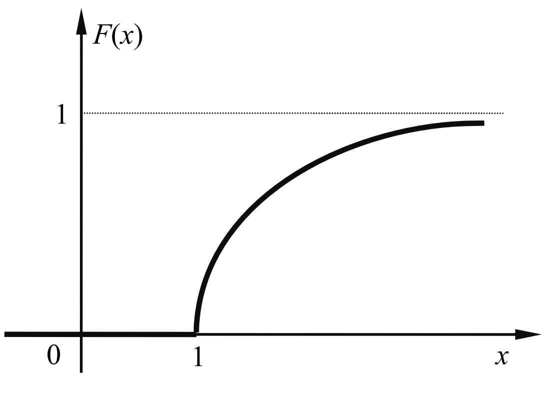

Рис. 6

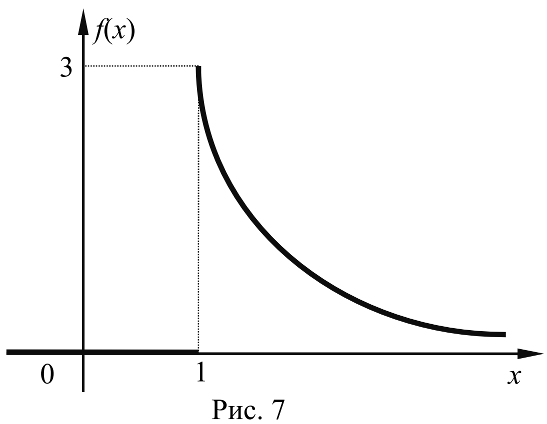

Рис. 7

Нормальное распределение

Случайная величина называется нормально распределенной, если ее плотность вероятности и функция распределения имеют вид:

$$f(x)=\frac{1}{\sigma\sqrt{2\pi}}\cdot e^{-\frac{(x-a)^2}{2\sigma^2}};\quad F(x)=0,5+\Phi\left(\frac{x-a}{\sigma}\right)$$

где $\Phi(x)$ — функция Лапласа, значения которой даны в приложении 2 [1]. Параметрами нормального распределения являются a = M(ξ) — математическое ожидание и σ = σ(ξ) — среднеквадратическое отклонение. Нормальное распределение обладает свойством устойчивости: сумма конечного числа независимых нормально распределенных величин распределена также нормально с математическим ожиданием и дисперсией, равными, соответственно, сумме математических ожиданий и дисперсий слагаемых. Вероятность попадания нормально распределенной величины в заданный интервал определяется по формуле:

$$P\left(x_1\le\xi\le x_2\right)=\Phi\left(\frac{x_2-a}{\sigma}\right)-\Phi\left(\frac{x_1-a}{\sigma}\right).$$

Пример 22. Вес цемента, упакованного автоматом в бумажный мешок, есть случайная, нормально распределенная величина с математическим ожиданием и среднеквадратическим отклонением a = 50 кг и σ = 2 кг. Найти вероятность того, что случайно выбранный мешок будет содержать не менее 48 кг цемента; партия из 100 мешков будет содержать не более 5040 кг.

Решение. Поскольку вес цемента в мешке — нормально распределенная величина, то искомая вероятность равна:

$$P\left(48\le\xi\le\infty\right)=\Phi\left(\frac{\infty-50}{2}\right)-\Phi\left(\frac{48-50}{2}\right)=\Phi(\infty)-\Phi(-1)==0,5+\Phi(1)=0,5+0,3413=0,8413.$$

Во втором случае вес всей партии (сумма 100 независимых нормально распределенных величин) является нормально распределенной случайной величиной ζ с математическим ожиданием и среднеквадратическим отклонением, соответственно, равными 100·a = 5000 и 10·σ = 20, поэтому:

$$P\left(0\le\le 5040\right)=\Phi\left(\frac{5040-5000}{20}\right)-\Phi\left(\frac{0-5000}{20}\right)=\Phi(2)+\Phi(250)==0,5+0,4772=0,9772.$$

Функция одного случайного аргумента

Пусть случайная величина η есть функция случайной величины ξ, то есть $h=j(x)$, причем закон распределения ξ известен. В случае дискретной случайной величины каждому значению $\xi=x_i$ соответствует $\eta=y_i=\varphi(x_i)$, причем в силу функциональной зависимости вероятности событий $\xi=x_i$ и $\eta=y_i$ равны. Если значения случайной величины η не упорядочены и среди них встречаются одинаковые, то их надо расположить в порядке возрастания, а одинаковые значения объединить, сложив соответствующие вероятности.

Пример 23. Случайная величина ξ — отклонение сопротивления резистора от номинала задана таблично. Составить закон распределения η — время (в минутах), необходимое на наладку прибора, если оно (в минутах) пропорционально квадрату отклонения (η = k·ξ^2).

ξ	–2	–1	0	1	2	3
p	0,1	0,2	0,3	0,2	0,1	0,1

Решение. Учитывая, что $\varphi(-2)=\varphi(2)=4\cdot k$, $\varphi(-1)=\varphi(1)=k$, $\varphi(0)=0$ и $\varphi(3)=9\cdot k$, объединяя равные значения и складывая соответствующие вероятности, получим закон распределения η в виде таблицы.

η	0	k	$4\cdot k$	$9\cdot k$
p	0,3	0,4	0,2	0,1

В случае непрерывной случайной величины по известной плотности вероятности случайной величины ξ можно найти плотность вероятности $\eta = \varphi(\xi)$. Если функция $y = \varphi(x)$ монотонна, то существует $x = \varphi(y)$ — функция, обратная к $y = \varphi(x)$ и $f(y) = f(x) \cdot |\psi'(y)|$.

По известному закону распределения дискретной случайной величины и известной плотности вероятности непрерывной случайной величины можно по известным формулам найти их числовые характеристики. Если эта зависимость линейная $\eta = A \cdot \xi + B$, то закон распределения сохраняется, а числовые характеристики случайной величины η будут равны: $M(\eta) = A \cdot M(\xi) + B$ и $s(\eta) = |A| \cdot \sigma(\xi)$.

Пример 24. Определить необходимый момент сопротивления балки, закрепленной и нагруженной, как указано на рис. 8, если нагрузка q является случайной нормально распределенной величиной с математическим ожиданием и среднеквадратическим отклонением, соответственно, равными $M(q) = 400$ кг/м и $\sigma(q) = 50$ кг/м. Предельное напряжение для материала балки принять равным $[\sigma] = 1800$ кг/см^2, а вероятность, с которой максимальное напряжение не должно превышать предельное, равной $p_о = 0{,}9$.

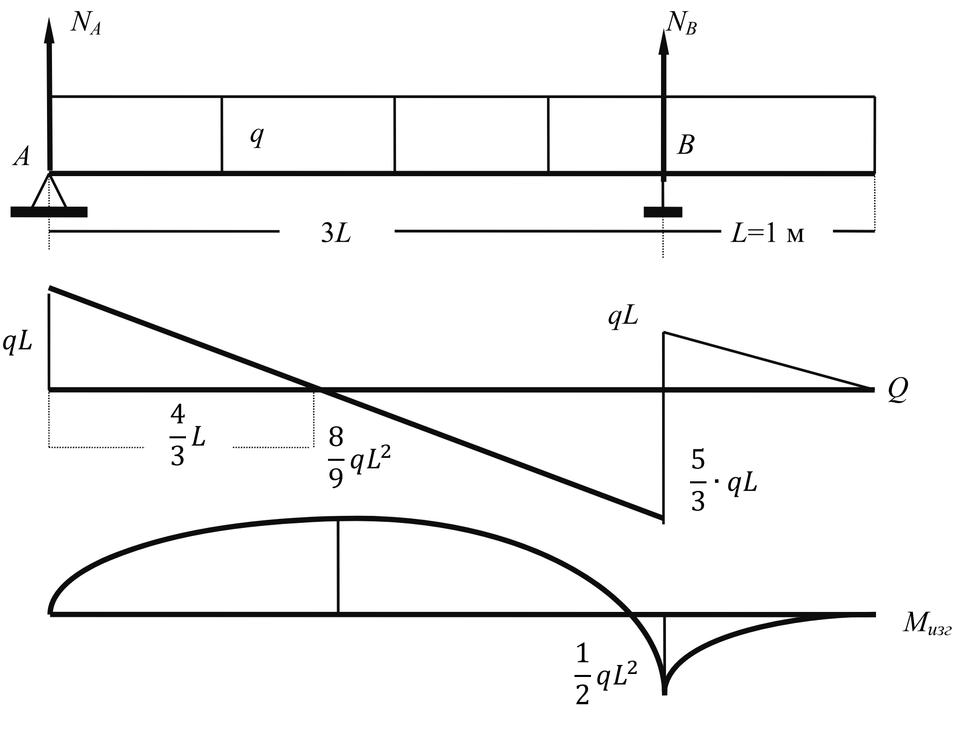

Рис. 8

Решение. Как известно из курса сопротивления материалов, необходимый момент сопротивления можно найти из условия: $\sigma_{\max} \le [\sigma]$, где $\sigma_{\max} = \dfrac{M_{\max}}{W}$, $M_{\max}$ — максимальный изгибающий момент; W — момент сопротивления.

Очевидно, $\sigma_{\max}$ зависит от размеров, характера закрепления и случайной нагрузки q и, следовательно, также является случайной величиной — функцией от q. Найдем эту функцию, решив соответствующую задачу сопротивления материалов.

Определенный таким образом максимальный изгибающий момент и соответствующее ему $\sigma_{\max}$ равны: $M_{\max} = \dfrac{8}{9} \cdot qL^2 \Rightarrow \sigma_{\max} = \dfrac{8L^2}{9W} \cdot q$.

Поскольку полученная зависимость линейная, то σ_{max}, так же как и q, распределена нормально с математическим ожиданием и среднеквадратическим отклонением, соответственно, равными:

$$M\left(\sigma_{\max}\right)=\frac{8L^2}{9W}\cdot M(q)\,;\ \sigma(\sigma_{\max})=\frac{8L^2}{9W}\cdot\sigma(q)\,.$$

Необходимый момент сопротивления балки найдем из условия

$$P\left(0\le\sigma_{max}\le[\sigma]\right)\ge p_0$$

Используя функцию Лапласа, приходим к неравенству:

$$\Phi\left(\frac{[\sigma]-M(\sigma_{\max})}{\sigma(\sigma_{\max})}\right)-\Phi\left(\frac{0-M(\sigma_{\max})}{\sigma(\sigma_{\max})}\right)\ge p_0.$$

Учитывая данные условия и свойства функции Лапласа, получим

$$\Phi\left(\frac{[\sigma]-\frac{8L^2}{9W}\cdot M(q)}{\frac{8L^2}{9W}\cdot\sigma(q)}\right)\ge 0,4\ \Rightarrow\ \frac{[\sigma]-\frac{8L^2}{9W}\cdot M(q)}{\frac{8L^2}{9W}\cdot\sigma(q)}\ge 1,28.$$

Преобразовав последнее неравенство и подставив числовые данные (переведенные в одну систему единиц), получим искомый момент сопротивления :

$$W\ge\frac{8\cdot L^2\cdot\left(M(q)+1,28\cdot\sigma(q)\right)}{9\cdot[\sigma]}=\frac{8\cdot 10\,000\cdot\left(4+1,28\cdot 0,5\right)}{9\cdot 1800}\approx 22,9\ \ \text{см}^3.$$

МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

Основные понятия и определения

Основными задачами математической статистики является разработка методов сбора, регистрации и анализа экспериментальных данных, полученных в результате наблюдений массовых явлений.

Выборочным называется метод, при котором из множества всех изучаемых объектов (их называют генеральной совокупностью, а их число N — ее объемом) выделяют ограниченное число объектов и их подвергают изучению (их называют выборочной совокупностью — кратко выборкой, а их число n — объемом выборки).

Выборка должна правильно представлять генеральную совокупность — более кратко быть репрезентативной.

Пусть в выборке объема n значения количественного признака $X = x_1$ наблюдалось n_1 раз, …, $X = x_k$ наблюдалось n_k раз. Наблюдаемые значения количественного признака x_i называются вариантами, а последовательность вариантов, записанная в порядке возрастания, — вариационным рядом. Число наблюдений значения количественного признака $X = x_i$, то есть n_i, называется частотой, а отношение n_i к объему выборки n — относительной частотой, v_i:

$$v_i = \frac{n_i}{n}, \quad \sum_{i=1}^{k} n_i = n, \quad \sum_{i=1}^{k} v_i = 1.$$

Соответствие между вариантами, записанными в порядке возрастания, и относительными частотами называется статистическим или эмпирическим распределением выборки.

Если обозначить $n(x)$ — число вариантов, меньших x, то относительная частота события $\xi < x$ называется эмпирической функцией распределения

$$F^*(x) = v(\xi < x) = \frac{n(x)}{n} .$$

Точечные и интервальные оценки неизвестных параметров

Пусть из генеральной совокупности извлечена выборка объемом n. Требуется построить формулы, зависящие от выборочных значений $x_1, ..., x_n$, для определения неизвестных параметров распределения (их принято называть статистиками).

Полученные таким образом точечные оценки должны быть несмещенными (их математическое ожидание по всевозможным выборкам объема n должно равняться истинному значению определяемого параметра) и состоятельными (при увеличении объема выборки оценка должна сходиться по вероятности к истинному значению).

Точечные и интервальные оценки параметров нормального распределения

Точечные оценки параметров нормального распределения — математического ожидания и дисперсии — определяются по формулам:

$$\tilde{a} = \overline{x} = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i; \quad \tilde{D} = s^2 = \frac{1}{n-1} \cdot \sum_{i=1}^{n} \left(x_i - \overline{x}\right)^2;$$

$$s = \sqrt{\frac{1}{n-1} \cdot \sum_{i=1}^{n} \left(x_i - \overline{x}\right)^2}.$$

Здесь $\overline{x}$ — выборочная средняя, s^2 — исправленная дисперсия, s — исправленное среднеквадратическое отклонение.

Интервальные оценки (доверительные интервалы), накрывающие неизвестные параметры распределения с заданной надежностью γ (доверительной вероятностью), определяются по формулам.

Для a:

$$\overline{x} - t_\gamma \frac{s}{\sqrt{n}} \le a \le \overline{x} + t_\gamma \frac{s}{\sqrt{n}},$$

где t_γ определяется по доверительной вероятности γ из приложения 3 [1] распределения Стьюдента с $(n-1)$ степенями свободы из условия $P\left(\left|\tau\right| \le t_\gamma\right) = \gamma$.

Для σ:

$$\sqrt{\frac{n-1}{h'_\gamma}} \cdot s \le \sigma \le \sqrt{\frac{n-1}{h''_\gamma}} \cdot s,$$

где h'_γ и h''_γ определяются из приложения 5 [1] распределения χ^2 с $(n-1)$ степенями свободы так, чтобы выполнялись соотношения:

$$P\left(\chi^2_{n-1} < h'_\gamma\right) = \frac{1-\gamma}{2} \text{ или } P\left(\chi^2_{n-1} > h'_\gamma\right) = 1 - \frac{1-\gamma}{2} = \alpha_1$$

$$\text{и} \quad P\left(\chi^2_{n-1} > h''_\gamma\right) = \frac{1-\gamma}{2} = \alpha_2.$$

Пример 25. Для проверки фасовочной установки отобраны и взвешены 20 упаковок. Получены следующие результаты (в граммах): 246; 247; 247,3; 247,4; 251,7; 252,5; 252,6; 252,8; 252,8; 252,9; 253; 253,6; 254,6; 254,7; 254,8; 256,1; 256,3; 256,8; 257,4; 259,2.

Найти доверительные интервалы для математического ожидания с надежностью 0,95 и среднеквадратического отклонения с надежностью 0,9, предполагая, что измеряемая величина распределена нормально.

Решение. Находим точечные оценки a и σ:

$$\tilde{a} = \overline{x} = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i = \frac{1}{20} \cdot \sum_{i=1}^{20} x_i = 252,98;$$

$$\tilde{\sigma}^2 = s^2 = \frac{1}{n-1} \cdot \sum_{i=1}^{n} \left(x_i - \overline{x}\right)^2 = \frac{1}{19} \cdot \sum_{i=1}^{20} \left(x_i - \overline{x}\right)^2 = 13,3;$$

$$\tilde{\sigma} = s = 3,65.$$

Определяем из приложения 3 [1] распределения Стьюдента для доверительной вероятности $\gamma = 0,95$ и числу степеней свободы $n - 1 = 19$ соответствующее значение $t_\gamma = 2,093$ и находим искомый интервал:

$$252,98 - \frac{2,093 \cdot 3,65}{\sqrt{20}} \le a \le 252,98 + \frac{2,093 \cdot 3,65}{\sqrt{20}}$$

$$251,27 \le a \le 254,69.$$

Для построения доверительного интервала для σ с надежностью $\gamma = 0,9$ находим из приложения 5 [1] распределения χ^2 с $n - 1 = 19$ степенями свободы числа h'_γ и h''_γ из условий:

$$P\left(\chi^2_{19} < h'_\gamma\right) = \frac{1-\gamma}{2} = 0,05 \text{ или } P\left(\chi^2_{19} > h'_\gamma\right) = 0,95 = \alpha_1$$

$$\text{и} \quad P\left(\chi^2_{19} > h''_\gamma\right) = \frac{1-\gamma}{2} = 0,05 = \alpha_2.$$

В результате получим $h'_\gamma = 10{,}117$ и $h''_\gamma = 30{,}144$. Отсюда искомый доверительный интервал, накрывающий σ с надежностью γ, равен 2,9 ≤σ≤ 5,0.

Применение метода наименьших квадратов к сглаживанию экспериментальных зависимостей

Пусть имеются результаты n независимых измерений — опытные точки (x_i, y_i), где $i = 1, \ldots, n$. Из теоретических или иных соображений с точностью до неизвестных параметров (для простоты мы ограничимся двумя) a и b известна функциональная зависимость y от x, то есть $y = \varphi(x, a, b)$. Экспериментальные точки уклоняются от этой зависимости вследствие неизбежных ошибок измерений. Теперь для определения параметров a и b воспользуемся методом наименьших квадратов, согласно которому неизвестные параметры находятся из условий минимума выражения:

$$\sum_{i=1}^{n}\left(y_i - \varphi(x_i, a, b)\right)^2 \tag{1}$$

которое сводится к решению двух уравнений:

$$\begin{cases} \dfrac{\partial}{\partial a}\sum_{i=1}^{n}\left(y_i - \varphi(x_i, a, b)\right)^2 = 0 \\ \dfrac{\partial}{\partial b}\sum_{i=1}^{n}\left(y_i - \varphi(x_i, a, b)\right)^2 = 0. \end{cases} \tag{2}$$

Если функциональная зависимость линейна относительно параметров a и b, то система уравнений также будет линейной и ее решение можно найти обычным способом.

Пример 26. Проведена серия опытов по определению влияния дозы внесенных удобрений на повышение урожайности пшеницы. Соответствующие данные приведены в первых трех столбцах таблицы (x — внесенная доза удобрений в центнерах на гектар, y — прирост урожайности в центнерах с гектара).

i	x_i	y_i	x_i^2	y_i^2	$x_i \cdot y_i$
1	0,342	2,10	0,1170	4,41	0,718
2	0,417	4,70	0,1739	22,09	1,960
3	0,675	6,05	0,4556	36,60	4,084
4	0,867	8,65	0,7517	74,82	7,500
5	1,000	10,00	1,0000	100,00	10,000
6	1,158	12,60	1,3410	158,76	14,591
7	1,283	12,08	1,6461	145,93	15,499
8	1,500	14,68	2,2500	215,50	22,020
9	1,733	16,65	3,0033	277,22	28,854
10	2,008	19,25	4,0321	370,56	38,654
11	2,083	19,98	4,3389	399,20	41,618
12	2,242	23,20	5,0266	538,24	52,014
13	2,508	23,93	6,2901	572,64	60,016
$\frac{1}{13}\sum_{i=1}^{13}$	1,370	13,37	2,3405	224,31	22,887

Требуется по методу наименьших квадратов подобрать линейную функцию, выражающую y через x.

Решение. Искомые величины связаны линейной зависимостью: $y = ax + b$, коэффициенты которой и требуется определить. Соотношение (1) в этом случае принимает вид:

$$\sum_{i=1}^{13}\left(y_i - ax_i - b\right)^2,$$

а система уравнений (2) записывается в виде:

$$\begin{cases} \dfrac{\partial}{\partial a}\sum_{i=1}^{13}\left(y_i - ax_i - b\right)^2 = 0 \\ \dfrac{\partial}{\partial b}\sum_{i=1}^{13}\left(y_i - ax_i - b\right)^2 = 0 \end{cases} \Rightarrow \begin{cases} -2\sum_{i=1}^{13}\left(y_i - ax_i - b\right)\cdot x_i = 0 \\ -2\sum_{i=1}^{13}\left(y_i - ax_i - b\right) = 0 \end{cases}$$

Раскрывая скобки и группируя, в результате получим следующую систему двух линейных уравнений для определения a и b:

$$\begin{cases} \left(\dfrac{1}{13}\sum_{i=1}^{13}x_i^2\right)\cdot a + \left(\dfrac{1}{13}\sum_{i=1}^{13}x_i\right)\cdot b = \dfrac{1}{13}\sum_{i=1}^{13}x_i\cdot y_i \\ \left(\dfrac{1}{13}\sum_{i=1}^{13}x_i\right)\cdot a + b = \dfrac{1}{13}\sum_{i=1}^{13}y_i. \end{cases}$$

Решая эту систему методом Гаусса (исключения), в итоге получим a = 9,86; b = –0,14 ⇒ y = 9,86x – 0,14.

Во многих приложениях часто используются зависимости вида $y = \dfrac{a}{x} + b,$ линейные относительно параметров a и b. В этом случае задача легко может быть сведена к предыдущей заменой переменной: $u = \dfrac{1}{x}$.

В случае квадратичной зависимости $y = a\cdot x^2 + b\cdot x + c$ система уравнений для определения неизвестных коэффициентов принимает вид:

$$\begin{cases} a\cdot\left(\sum_{i=1}^{n}x_i^4\right) + b\cdot\left(\sum_{i=1}^{n}x_i^3\right) + c\cdot\left(\sum_{i=1}^{n}x_i^2\right) = \sum_{i=1}^{n}x_i^2\cdot y_i \\ a\cdot\left(\sum_{i=1}^{n}x_i^4\right) + b\cdot\left(\sum_{i=1}^{n}x_i^3\right) + c\cdot\left(\sum_{i=1}^{n}x_i^2\right) = \sum_{i=1}^{n}x_i^2\cdot y_i \\ a\cdot\left(\sum_{i=1}^{n}x_i^2\right) + b\cdot\left(\sum_{i=1}^{n}x_i\right) + c\cdot n = \sum_{i=1}^{n}y_i \end{cases}$$

Эту линейную относительно неизвестных a, b и c, систему уравнений также можно решать методом Гаусса.

Замечание: использование современных инженерных калькуляторов значительно облегчает вычисления, а в научных калькуляторах даже имеются встроенные программы определения коэффициентов рассмотренных зависимостей, и требуется только ввести данные условия задачи.

Контрольные вопросы

1. Случайное, достоверное и невозможное события. Сумма и произведение событий, противоположное событие.
2. Относительная частота. Определение вероятности для дискретного (счетного) пространства элементарных событий.
3. Классическое определение вероятности. Аксиоматическое определение вероятности. Геометрическая схема теории вероятностей.
4. Совместность и несовместность событий. Вероятность противоположного события; суммы событий.
5. Зависимые и независимые события. Условная вероятность. Вероятность произведения событий.
6. Формулы полной вероятности и Байеса.
7. Последовательность независимых, однородных испытаний (схема Бернулли). Формула Бернулли.
8. Асимптотические формулы: закон редких событий — формула Пуассона; локальная и интегральная теоремы и формулы Муавра — Лапласа, функция Лапласа и ее свойства; использование таблиц.
9. Простейший, стационарный (пуассоновский) поток событий.
10. Дискретные и непрерывные случайные величины, способы их задания.
11. Функция распределения и ее свойства.
12. Плотность вероятности непрерывной случайной величины и ее свойства.
13. Числовые характеристики — математическое ожидание, дисперсия и среднеквадратическое отклонение случайной величины и их свойства.
14. Некоторые распределения и их числовые характеристики: биномиальное, пуассоновское, равномерное, показательное.
15. Нормальное распределение и его числовые характеристики. Вероятностный смысл параметров нормального распределения и их влияние на график плотности вероятности.
16. Функция распределения нормально распределенной случайной величины и ее связь с функцией Лапласа. Вероятность попадания в заданный интервал, применение таблиц, правило трех сигм.
17. Функция одного случайного аргумента, закон ее распределения и числовые характеристики.
18. Предельные теоремы: теоремы Чебышева и Ляпунова, следствия из них.
19. Генеральная и выборочная совокупности и их описание.
20. Точечные оценки неизвестных параметров и их построение по данным выборки методами наибольшего правдоподобия и моментов. Проверка несмещенности и состоятельности оценки.
21. Интервальные оценки неизвестных параметров, доверительная вероятность, построение доверительных интервалов по данным выборки.
22. Метод наименьших квадратов и его применение к сглаживанию экспериментальных зависимостей.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Гмурман В.Е. Теория вероятностей и математическая статистика : Учебное пособие для бакалавров. — 12-е изд. — Москва : Юрайт, 2013. — 479 с.
2. Гмурман В.Е. Руководство к решению задач по теории вероятностей и математической статистике : Учебное пособие для бакалавров. — 11-е изд., перераб. и доп. — Москва : Юрайт, 2013. — 404 с.

Таблицы

Таблица 1

Значения функции $p(m)=\dfrac{a^m}{m!}\cdot e^{-a}$

a\ m	0	1	2	3	4	5	6	7
0,1	**0,90484**	09048	00452	00015	00000	00000	00000	00000
0,2	81873	16375	01638	00109	00006	00000	00000	00000
0,3	74082	22225	03334	00333	00025	00002	00000	00000
0,4	67032	26813	05363	00715	00072	00006	00000	00000
0,5	69653	30327	07582	01264	00158	00016	00001	00000
0,6	54881	32929	09879	01976	00296	00036	00004	00000
0,7	49659	34761	12166	02839	00497	00070	00008	00001
0,8	44933	35946	14379	03834	00767	00123	00016	00002
0,9	40657	36591	16466	04940	01112	00200	00030	00004
1	36788	36788	18394	06131	01533	00307	00051	00007
2	13534	27067	27067	18045	09022	03609	01203	00344
3	04979	14936	22404	22404	16803	10082	05041	02160
4	01832	07326	14653	19537	19537	15629	10420	05954
5	00674	03369	08422	14037	17547	17547	14622	10445
6	00248	01487	04462	08924	13385	16062	16062	13768
7	00091	00638	02234	05213	09123	12772	14900	14900

Таблица 2

Значения функции $p(m\le k)=e^{-a}\cdot\sum_{m=0}^{k}\dfrac{a^m}{m!}$

***a*\k**	0	1	2	3	4	5	6	7
0,1	**0,90484**	99532	99985	1,0000	1,0000	1,0000	1,0000	1,0000
0,2	81873	93248	99885	99994	1,0000	1,0000	1,0000	1,0000
0,3	74082	96306	99640	99973	99998	1,0000	1,0000	1,0000
0,4	67032	93845	99207	99922	99994	1,0000	1,0000	1,0000
0,5	60653	90980	98561	99825	99983	99999	1,0000	1,0000
0,6	54881	87810	97689	99664	99961	99996	1,0000	1,0000
0,7	49659	84420	96586	99425	99921	99991	99999	1,0000
0,8	44933	80879	95258	99092	99859	99982	99998	1,0000
0,9	40657	77248	93714	98654	99766	99966	99996	1,0000
1	36788	73576	91970	98101	99634	99941	99992	99999
2	13534	40601	67668	85712	94735	98344	99547	99890
3	04979	19915	42319	64723	81526	91608	96649	98810
4	01832	09158	23810	43347	62792	81548	88876	94778
5	00674	04043	12465	26503	44049	61596	76218	86663
6	00248	01735	06197	15120	28506	44568	60630	74398
7	00091	00730	02964	08177	17299	30071	44971	59871