УЧЕБНИК

В.П. Омельченко
А.А. Демидова

ИНФОРМАТИКА

МЕДИЦИНСКАЯ ИНФОРМАТИКА

СТАТИСТИКА

Рекомендовано УМО РАЕ по классическому университетскому и техническому образованию в качестве учебника для студентов высших учебных заведений, обучающихся по специальности 32.05.01 «Медико-профилактическое дело»
Протокол № 842 от 26 августа 2020 года

Москва
ИЗДАТЕЛЬСКАЯ ГРУППА
«ГЭОТАР-Медиа»
2021

ОГЛАВЛЕНИЕ

Предисловие 9
Список сокращений и условных обозначений 12

ЧАСТЬ I. Статистика 13

Глава 1. Основы теории вероятностей 14
1.1. Случайные события и операции над ними 14
1.1.1. Случайные события 14
1.1.2. Операции над событиями 17
1.1.3. Определение вероятности события 19
1.1.4. Теорема сложения вероятностей 21
1.1.5. Теорема умножения вероятностей 23
1.1.6. Формула полной вероятности 25
1.1.7. Формула Байеса 26
1.2. Случайные величины 28
1.2.1. Распределение дискретных и непрерывных случайных величин 29
1.2.2. Числовые характеристики случайных величин 34
1.2.3. Законы распределения непрерывных случайных величин 39
1.2.4. Закон больших чисел 45
Контрольные вопросы 46

Глава 2. Описательная статистика 47
2.1. Типы данных 48
2.2. Генеральная совокупность и выборка 49
2.3. Статистическое распределение (вариационный ряд). Гистограмма. Полигон 50
2.4. Характеристики положения и рассеяния статистического распределения 53
2.5. Показатели, характеризующие форму распределения 57
2.6. Оценка параметров генеральной совокупности по ее выборке 58
2.7. Интервальная оценка. Доверительный интервал и доверительная вероятность 62
Контрольные вопросы 64

Глава 3. Статистическая проверка гипотез 65
3.1. Общая постановка задачи проверки гипотез 65
3.2. Алгоритм проверки статистических гипотез 67
3.3. Проверка гипотез относительно средних 68
3.4. Проверка гипотез для дисперсий 69
3.5. Проверка гипотез о законах распределения 70

3.6. Непараметрические критерии . . . 72
3.6.1. Критерий знаков . . . 73
3.6.2. Критерий Уилкоксона–Манна–Уитни (U-критерий) . . . 74
Контрольные вопросы . . . 76

Глава 4. Корреляционный и регрессионный анализ . . . 77
4.1. Корреляционный анализ . . . 77
4.1.1. Функциональная и корреляционная зависимости . . . 78
4.1.2. Коэффициент линейной корреляции и его свойства . . . 79
4.1.3. Проверка гипотезы о значимости выборочного коэффициента линейной корреляции . . . 82
4.1.4. Выборочный коэффициент ранговой корреляции Спирмена . . . 84
4.2. Выборочное уравнение линейной регрессии . . . 86
4.2.1. Метод наименьших квадратов . . . 86
4.2.2. Нелинейная регрессия . . . 91
Контрольные вопросы . . . 92

Глава 5. Основы дисперсионного анализа . . . 93
5.1. Основные понятия и символы . . . 93
5.2. Схема однофакторного дисперсионного анализа . . . 94
5.3. Двухфакторный дисперсионный анализ . . . 101
Контрольные вопросы . . . 107

Глава 6. Анализ временных рядов . . . 108
6.1. Понятие временного ряда . . . 108
6.2. Определение тренда . . . 111
6.3. Анализ периодических временных рядов . . . 118
6.4. Анализ случайных (стохастических) временных рядов . . . 121
Контрольные вопросы . . . 124

ЧАСТЬ II. Информатика . . . 125

Глава 7. Автоматизированная обработка информации . . . 126
7.1. Информация и ее свойства . . . 126
7.2. Кодирование информации . . . 129
7.2.1. Кодирование чисел . . . 131
7.2.2. Кодирование текста . . . 131
7.2.3. Кодирование графической информации . . . 133
7.2.4. Кодирование звуковой информации . . . 135
7.2.5. Кодирование видеоинформации . . . 135
7.3. Измерение информации . . . 136
7.4. Предмет и задачи информатики . . . 138

7.5. Информационные технологии и их применение в медицине и здравоохранении . . . 139
7.5.1. Понятие информационной технологии . . . 139
7.5.2. Применение информационных технологий в медицине и здравоохранении . . . 142
Контрольные вопросы . . . 156

Глава 8. Техническая и программная база информатики . . . 158
8.1. Аппаратное обеспечение персональных компьютеров . . . 158
8.1.1. Принципы работы ЭВМ . . . 158
8.1.2. Классификация ЭВМ . . . 163
8.1.3. Структурная схема персонального компьютера . . . 169
8.1.4. Состав персонального компьютера . . . 171
8.1.5. Периферийные устройства персонального компьютера . . . 185
8.2. Программное обеспечение персональных компьютеров . . . 208
8.2.1. Защита информации . . . 208
8.2.2. Классификация программных средств . . . 216
8.2.3. Операционные системы и оболочки операционных систем . . . 219
8.2.4. Программы архивации данных . . . 238
Контрольные вопросы . . . 245

Глава 9. Организация профессиональной деятельности с помощью средств Microsoft Office . . . 247
9.1. Обработка информации средствами Microsoft Word . . . 247
9.1.1. Понятие текстового процессора и его основные функции . . . 247
9.1.2. Возможности текстового редактора Microsoft Word . . . 248
9.1.3. Настройка интерфейса Microsoft Word . . . 249
9.1.4. Создание и редактирование текстового документа . . . 258
9.1.5. Настройка интервалов. Абзацные отступы . . . 261
9.1.6. Работа со списками . . . 263
9.1.7. Работа с окнами . . . 264
9.1.8. Принципы создания таблицы . . . 265
9.1.9. Стили и темы в документе. Использование гиперссылок . . . 270
9.1.10. Создание титульного листа . . . 273
9.1.11. Вставка графических изображений в документ. Объекты WordArt . . . 274
9.1.12. Оформление страниц . . . 285
9.1.13. Вид документа . . . 290
9.1.14. Печать документов . . . 293
9.1.15. Сохранение документов . . . 294
9.2. Обработка информации средствами Microsoft Excel . . . 294
9.2.1. Назначение электронных таблиц . . . 294
9.2.2. Ввод данных в ячейки Microsoft Excel . . . 299
9.2.3. Выполнение операций перемещения, копирования и заполнения ячеек. Автозаполнение . . . 302
9.2.4. Создание и редактирование табличного документа . . . 309

9.2.5. Работа с диаграммами 311
9.2.6. Ссылки. Встроенные функции. Статистические и логические функции 315
9.2.7. Вычисления в электронных таблицах 318
9.2.8. Фильтрация (выборка) данных из списка 321
9.2.9. Сортировка данных. 325
9.3. Обработка информации средствами Microsoft Access 328
9.3.1. Назначение Microsoft Access 328
9.3.2. Интерфейс Microsoft Access 330
9.3.3. Создание таблиц 336
9.3.4. Ввод и редактирование данных таблицы 339
9.3.5. Создание связей между таблицами 341
9.3.6. Работа с базой данных 343
9.3.7. Создание запросов 356
9.3.8. Составление отчетов 365
9.4. Создание презентаций средствами Microsoft PowerPoint 368
9.4.1. Возможности технологии компьютерной презентации 368
9.4.2. Основные элементы Microsoft PowerPoint 369
9.4.3. Общая схема создания первой презентации. 374
9.4.4. Изменение презентации 376
9.4.5. Добавление фигур, схем, значков и изображений на слайд . . . 378
9.4.6. Создание таблиц и диаграмм 384
9.4.7. Анимация объектов и переходы между слайдами 388
9.4.8. Основные правила создания презентации 390
9.5. Компьютерная графика 391
9.5.1. Растровая графика. 393
9.5.2. Векторная графика 401
9.5.3. Фрактальная графика 407
9.5.4. Трехмерная графика 409
Контрольные вопросы 410

Глава 10. Использование информационных технологий статистической обработки данных в медицине и биологии 412
10.1. STATISTICA 415
10.2. GraphPad Prism 419
10.3. Advanced Grapher 421
10.4. Пример использования Advanced Grapher, STATISTICA и GraphPad Prism для расчетов 422
Контрольные вопросы 428

ЧАСТЬ III. Медицинская информатика 429

Глава 11. Локальные и глобальные компьютерные сети 430
11.1. Сетевые технологии обработки информации 430
11.1.1. Топология локальных сетей 434
11.1.2. Протоколы 443

11.1.3. Прикладные протоколы . . . 449
11.1.4. Общие сведения о подключении локальных сетей к Интернету . . . 451
11.2. Глобальная сеть Интернет . . . 453
11.2.1. Структура и адресация в Интернете . . . 453
11.2.2. Подключение к Интернету . . . 456
11.2.3. Информационные ресурсы Интернета . . . 457
11.2.4. Работа с поисковыми системами . . . 459
11.2.5. Язык HTML. Создание веб-страниц . . . 463
11.2.6. Медицинские ресурсы в Интернете . . . 464
Контрольные вопросы . . . 466

Глава 12. Деятельность санитарной эпидемиологической службы территориального и федерального уровней и подходы к ее автоматизации . . . 467
12.1. Цель и задачи информатизации Роспотребнадзора . . . 469
12.2. Приоритетные направления информатизации Роспотребнадзора . . . 470
12.3. Развитие официального сайта Роспотребнадзора, сайтов его территориальных органов и организаций . . . 472
12.4. Практические результаты информатизации Роспотребнадзора . . . 473
12.5. Современное состояние уровня информатизации Роспотребнадзора . . . 477
Контрольные вопросы . . . 479

Глава 13. Медицинские информационные системы общего клинического направления и санитарно-эпидемиологической службы . . . 480
13.1. Понятие информационной системы и медицинской автоматизированной информационной системы . . . 480
13.1.1. Цель, задачи и функции медицинской информационной системы . . . 481
13.1.2. Классификация медицинских информационных систем . . . 482
13.1.3. Структура медицинской информационной системы . . . 489
13.1.4. Автоматизированное рабочее место медицинского персонала . . . 492
13.2. Основы функционирования медицинской информационной системы на примере комплексной автоматизированной информационной системы по соблюдению санитарных правил, контролю качества товаров, правил торговли, обслуживания населения и защиты прав потребителей («КАИС-Комплекс») . . . 495
13.2.1. Организационная структура системы «КАИС-Комплекс» . . . 497
13.2.2. Основные комплексы задач, решаемые системой «КАИС-Комплекс» . . . 497
Контрольные вопросы . . . 515

Глава 14. Автоматизированные медико-технологические системы лабораторных исследований . . . 516
14.1. Актуальность автоматизации лабораторной деятельности . . . 516
14.2. Структура лабораторных информационных систем . . . 517
14.3. Функции лабораторных информационных систем . . . 523
14.4. Организация технологического процесса в медицинской лаборатории . . . 524
14.5. Обзор современных лабораторных информационных систем . . . 527
14.5.1. ALTEY Laboratory . . . 528
14.5.2. ILIMS . . . 529
14.5.3. LabTrak . . . 529
14.5.4. LabSystem . . . 530
14.5.5. Medap-LIS . . . 530
14.5.6. PSM-АКЛ . . . 531
14.5.7. ЛИС «АЛИСА» . . . 532
14.6. Понятие лабораторной информатики . . . 533
14.7. Информативность диагностических исследований . . . 534
14.8. Показатели информативности диагностических методов . . . 537
14.8.1. Определение диагностической чувствительности . . . 537
14.8.2. Диагностическая специфичность . . . 538
14.8.3. Диагностическая точность . . . 540
14.8.4. Прогностическая ценность метода . . . 542
14.8.5. Варианты сочетанного применения лабораторных диагностических исследований . . . 544
14.9. Понятие ROC-анализа . . . 546
14.9.1. Этапы ROC-анализа . . . 547
Контрольные вопросы . . . 549

Глава 15. Телекоммуникационные технологии в медицине . . . 550
15.1. Телемедицина, ее цель и направления . . . 550
15.2. Телемедицинская сеть как элемент единого информационного пространства системы здравоохранения . . . 551
15.3. Направления работы телемедицинских центров . . . 551
15.4. Основные инструменты телемедицины . . . 552
15.5. Этапы развития телемедицины . . . 554
15.6. Нормативно-правовая база развития телемедицины в Российской Федерации . . . 562
15.7. Разделы телемедицины . . . 566
Контрольные вопросы . . . 567

Глоссарий . . . 568

Список литературы . . . 583

Приложение. Таблицы критериев . . . 586

Предметный указатель . . . 598

СПИСОК СОКРАЩЕНИЙ И УСЛОВНЫХ ОБОЗНАЧЕНИЙ

АРМ — автоматизированное рабочее место
АЦП — аналого-цифровой преобразователь
БД — база данных
ЕГИС — Единая государственная информационная система
ЛВС — локальная вычислительная сеть
ЛИС — лабораторная информационная система
ЛПУ — лечебно-профилактическое учреждение
МИС — медицинская информационная система
МП — микропроцессор
НЖМД — накопитель на жестких магнитных дисках
ОЗУ — оперативное запоминающее устройство
ОМС — обязательное медицинское страхование
ПК — персональный компьютер
СУБД — система управления базами данных
ЦАП — цифроаналоговый преобразователь
ЦВМ — цифровая вычислительная машина
ЭВМ — электронная вычислительная машина
ЭКГ — электрокардиограмма, электрокардиография, электрокардиографический

TCP/IP — протокол управления передачей/интернет-протокол (от англ. Transmission Control Protocol/Internet Protocol)
URL — универсальный указатель ресурсов (от англ. Uniform Resource Locator)
WWW — Всемирная паутина (от англ. World Wide Web)
XML — расширяемый язык разметки (от англ. eXtensible Markup Language)

◁ — конец примера

Часть I

СТАТИСТИКА

Глава 1

ОСНОВЫ ТЕОРИИ ВЕРОЯТНОСТЕЙ

Определения

Математическая статистика — это раздел математики, изучающий методы сбора, анализа и обработки статистических данных для научных и практических целей.

Статистические данные — события и величины, которые могут иметь несколько исходов или принимать набор значений, при этом каждый исход и каждое значение характеризуются определенной вероятностью наблюдений.

Изучение методов математической статистики базируется на знании основ теории вероятностей.

1.1. СЛУЧАЙНЫЕ СОБЫТИЯ И ОПЕРАЦИИ НАД НИМИ

Определение

Теория вероятностей — это раздел математики, изучающий закономерности массовых случайных событий.

1.1.1. Случайные события

Определения

Случайным называется событие, наступление которого нельзя гарантировать.

Случайность того или иного события определяется множеством причин, которые существуют объективно, но учесть их все, а также степень их влияния на изучаемое событие, невозможно. К таким случайным событиям относят выпадение того или иного числа при бросании игральной кости, выигрыш в лотерее, количество больных, записавшихся на прием к врачу, и т.п.

И хотя в каждом конкретном случае трудно предсказать исход испытания, при достаточно большом числе наблюдений можно установить наличие некоторой закономерности. Подбрасывая монету, можно заметить, что количества выпадений орла и решки примерно равны, а при бросании игральной кости различные грани также появляются примерно одинаково. Это говорит о том, что случайным явлениям присущи свои закономерности, но они проявляются лишь при большом количестве испытаний. Правильность этого подтверждает закон больших чисел, который лежит в основе теории вероятностей.

Рассмотрим основные термины и понятия теории вероятностей.

Определения

Испытанием называется совокупность условий, при которых может произойти данное случайное событие.

Событие — это факт, который при осуществлении определенных условий может произойти или не произойти.

События обозначают большими буквами латинского алфавита: A, B, C, ...

Например: событие A — рождение мальчика, событие B — выигрыш в лотерее, событие C — выпадение цифры 4 при бросании игральной кости.

События бывают достоверные, невозможные и случайные.

Определение

Достоверное событие — это событие, которое в результате испытания непременно должно произойти.

Например, если на игральной кости на всех шести гранях нанести цифру 1, тогда выпадение цифры 1 при бросании кости есть событие достоверное.

Определение

Невозможное событие — это событие, которое в результате испытания не может произойти.

Например, невозможным событием будет выпадение любой цифры, кроме 1, если пометить все грани игральной кости цифрой 1.

Определение

Случайное событие — это событие, которое при испытаниях может произойти или не произойти. Те или иные события реализуются с различной возможностью.

Например, завтра днем ожидается дождь. Здесь наступление дня является испытанием, а выпадение дождя — случайное событие.

Определение

События называются **несовместными**, если в результате данного испытания появление одного из них исключает появление другого.

Например, при бросании монеты выпадение одновременно орла и решки есть событие несовместное.

Определение

События называются **совместными**, если в результате данного испытания появление одного из них не исключает появление другого.

Например, при игре в карты появление валета и масти пик — события совместные.

Определение

События называются **равновозможными**, если нет оснований считать, что одно из них происходит чаще, чем другое.

Например, выпадение любой грани игрального кубика есть равновозможное событие.

События образуют *полную группу событий*, если в результате испытания обязательно произойдет хотя бы одно из них и любые два из них несовместны.

Например, при 10 выстрелах в мишень возможно от 0 до 10 попаданий. При бросании игрального кубика может выпасть цифра от 1 до 6. Эти события образуют полную группу.

Определение

События, входящие в полную группу попарно несовместных и равновозможных событий, называются **исходами**, или **элементарными событиями**.

Согласно определению достоверного события можно считать, что событие, состоящее в появлении одного, неважно какого, из событий полной группы есть событие достоверное.

Например, при бросании одного игрального кубика выпадает число меньше 7. Это пример достоверного события.

Частным случаем событий, образующих полную группу, являются противоположные события.

Определение

Два несовместных события A и $\overline{A}$ (читается «не A») называются **противоположными**, если в результате испытания одно из них должно обязательно произойти.

Например, при бросании монеты могут выпасть только орел или решка. Следовательно, это два противоположных события.

1.1.2. Операции над событиями

Определение

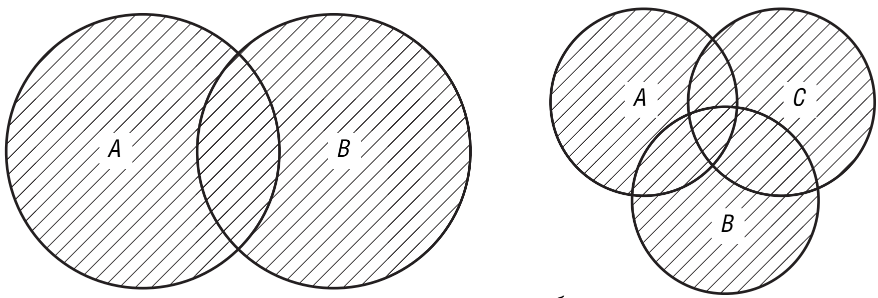
Суммой нескольких событий называется событие, состоящее в наступлении хотя бы одного из них в результате испытания.

Сумма событий может быть обозначена знаками «+», «∨», «или».

На рис. 1.1 представлена геометрическая интерпретация с помощью диаграмм Эйлера–Венна. Сумме событий $A + B$ будет соответствовать вся заштрихованная область.

Область пересечения событий A и B соответствует совместным событиям, которые могут произойти одновременно. Аналогично для событий A, B и C имеются совместные события A и B; A и C; B и C; A и B и C, которые могут произойти одновременно.

Например, в урне находятся белые, красные и синие шары. Возможны следующие события: A — вынут белый шар; B — вынут красный



а *б*

Рис. 1.1. Диаграмма Эйлера–Венна для суммы событий: *а* — $A + B$; *б* — $A + B + C$

шар; C — вынут синий шар. Событие $B + C$ означает, что произошло событие — вынут цветной шар или вынут не белый шар.

Определение

Произведением нескольких событий называется событие, которое состоит в совместном наступлении всех этих событий в результате испытания.

Произведение событий может быть обозначено знаками «×», «∧», «и».

Геометрическая интерпретация произведения событий представлена на рис. 1.2. Произведением событий A и B будет заштрихованная область пересечения площадей A и B, а для трех событий A и B и C — общая площадь, одновременно входящая во все три события.

Например, пусть из колоды карт наугад извлекается карта. Событие A — вынута карта пиковой масти; B — вынут валет. Тогда событие $A \times B$ означает событие — вынут валет пик.

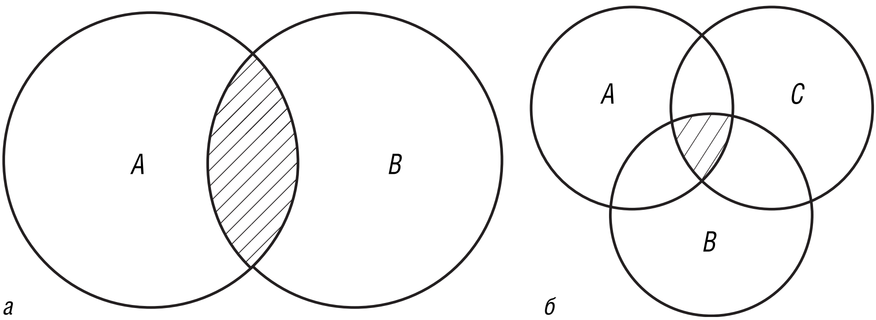


Рис. 1.2. Диаграмма Эйлера–Венна для произведения событий: *а* — $A \times B$; *б* — $A \times B \times C$

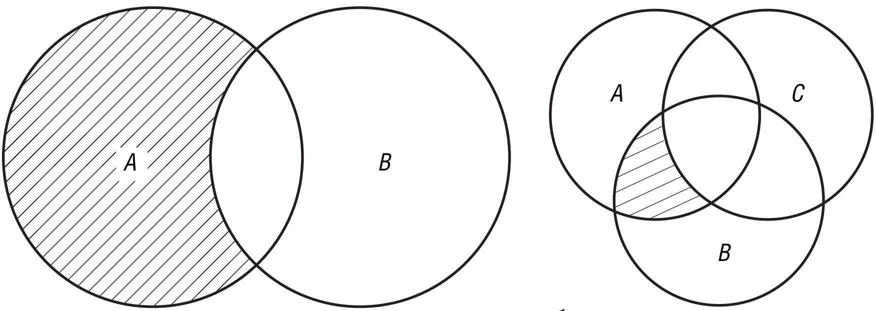


Рис. 1.3. Диаграмма Эйлера–Венна для разности событий: *а* — $A - B$; *б* — $A \times B - C$

Определение

Разностью двух событий $A - B$ называется событие, состоящее из исходов, входящих в A, но не входящих в B.

На рис. 1.3 представлена иллюстрация разности событий с помощью диаграмм Эйлера—Венна. Разностью двух событий $A - B$ является заштрихованная область A без той части, которая входит в событие B. Разностью между произведением событий A и B и событием C будет совместная площадь события A и события B без совместной с нею площади события C.

Например, пусть при бросании игрального кубика событие A — появление четных чисел (2, 4, 6), а событие B — чисел, кратных 3, то есть (3, 6). Тогда событие $A - B$ — это появление чисел (2, 4).

1.1.3. Определение вероятности события

Случайные события реализуются с различной возможностью. Одни происходят чаще, другие реже. Для количественной оценки возможностей реализации события вводится понятие вероятности события.

Определение

Вероятность события — это число, характеризующее степень возможности появления событий при многократном повторении испытаний.

Вероятность обозначают буквой P (от англ. probability — вероятность). Вероятность является одним из основных понятий теории вероятностей. Существует несколько определений этого понятия.

Классическое определение вероятности заключается в следующем: если известны все возможные исходы испытания и нет оснований считать, что одно случайное событие появлялось бы чаще других, то есть события равновозможны и несовместны, то имеется возможность аналитического определения вероятности события.

Вероятностью $P(A)$ события A называется отношение числа благоприятствующих исходов m к общему числу равновозможных несовместных исходов n:

$$P(A) = \frac{m}{n}. \quad (1.1)$$

Свойства вероятности

1. Вероятность случайного события A находится между 0 и 1.

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

2. Вероятность достоверного события равна 1.

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

3. Вероятность невозможного события равна 0.

$$P(A)=\frac{m}{n}=\frac{0}{n}=0.$$

Пример 1.1. Найти вероятность выпадения числа, кратного 3, при одном бросании игрального кубика.

Решение. Событие A — выпадение числа, кратного 3. Этому событию благоприятствуют два исхода: числа 3 и 6, то есть $m = 2$. Общее число исходов состоит в выпадении чисел: 1, 2, 3, 4, 5, 6, то есть $n = 6$. Очевидно, что эти события равновозможны и образуют полную группу. Тогда искомая вероятность, по определению, равна отношению числа благоприятствующих исходов к числу всех исходов.

$$P(A)=\frac{m}{n}=\frac{2}{6}=\frac{1}{3}. \triangleleft$$

Пример 1.2. В урне 10 белых, 5 красных и 5 зеленых шаров. Найти вероятность того, что вынутый наугад шар будет цветным (не белым).

Решение. Число исходов, благоприятствующих событию A, равно сумме красных и зеленых шаров: $m = 10$. Общее число равновозможных несовместных исходов равно общему числу шаров в урне: $n = 20$. Тогда:

$$P(A)=\frac{m}{n}=\frac{10}{20}=0{,}5. \triangleleft$$

При расчете вероятности события по ее классическому определению требуется выполнение определенных условий. Эти условия заключаются в равновозможности и несовместности событий, входящих в полную группу событий, вероятность которых надо определить. На практике не всегда можно определить все возможные варианты исходов, а тем более обосновать их равновозможность. Именно поэтому при невозможности удовлетворения требованиям классического определения вероятности используют статистическую оценку вероятности события. При этом вводится понятие *относительной частоты* появления события A, равной отношению m/n, где m — число испытаний, в которых произошло событие A; n — общее число испытаний.

Якоб Бернулли доказал, что при неограниченном увеличении числа испытаний относительная частота события A будет сколь угодно мало отличаться от вероятности события A.

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

Это равенство справедливо при неизменности условий, при которых проводится эксперимент.

Справедливость теоремы Бернулли была доказана и в многочисленных опытах по сравнению вероятностей, вычисленных классическим и статистическим методами. Так, в опытах Пирсона по определению вероятности выпадения «герба» при 12 000 бросков статистическая вероятность была равна 0,5016, а при 24 000 бросков — 0,5005, что показывает приближение к значению вероятности 0,5 по мере увеличения числа опытов. Близость значений вероятности, определенных различными способами, указывает на объективность возможности наступления этого события.

1.1.4. Теорема сложения вероятностей

Зная вероятности одних событий, можно вычислить вероятности других событий, если они связаны между собой. Теорема сложения вероятностей позволяет определить вероятность появления одного из нескольких случайных событий.

Теорема 1.1 (сложение вероятностей)

Вероятность суммы двух несовместных событий A и B равна сумме вероятностей этих событий.

$$P(A+B)=P(A)+P(B). \tag{1.2}$$

Доказательство. Пусть n — общее число равновозможных несовместных элементарных исходов; m_1 — число исходов, благоприятствующих событию A; m_2 — число исходов, благоприятствующих событию B. Поскольку A и B — несовместные события, то событию $A+B$ будет благоприятствовать m_1+m_2 исходов. Тогда согласно классическому определению вероятности

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

Расширяя это доказательство на n событий, можно доказать следующую теорему.

Теорема 1.2

Вероятность суммы конечного числа попарно несовместных событий A_1, A_2, ..., A_n равна сумме вероятностей этих событий, то есть

$$P(A_1 + A_2 + \ldots + A_n) = P(A_1) + P(A_2) + \ldots + P(A_n). \tag{1.3}$$

Из этой теоремы можно вывести два следствия.

Следствие 1

Если события A_1, A_2, ..., A_n образуют полную группу, то сумма их вероятностей равна единице, то есть

$$P(A_1) + P(A_2) + \ldots + P(A_n) = 1. \tag{1.4}$$

Доказательство. Если события A_1, A_2, ..., A_n образуют полную группу, то наступление хотя бы одного из них есть событие достоверное. Следовательно,

$$P(A_1 + A_2 + \ldots + A_n) = 1$$

и

$$P(A_1 + A_2 + \ldots + A_n) = P(A_1) + P(A_2) + \ldots + P(A_n) = 1.$$

Следствие 2

Сумма вероятностей противоположных событий равна единице, то есть

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

Доказательство. Противоположные события несовместны и образуют полную группу, а сумма вероятностей таких событий равна 1.

Пример 1.3. Найти вероятность выпадения цифры 2 или 3 при бросании игральной кости.

Решение. Событие A — выпадение цифры 2, вероятность этого события $P(A) = 1/6$. Событие B — выпадение цифры 3, вероятность этого события $P(B) = 1/6$. События несовместные, поэтому

$$P(A+B) = P(A) + P(B) = \frac{1}{6} + \frac{1}{6} = \frac{2}{6} = \frac{1}{3}. \triangleleft$$

В том случае если события A и B являются совместными, справедлива следующая теорема.

Теорема 1.3

Вероятность появления хотя бы одного из двух совместных событий равна сумме вероятностей этих событий без вероятности их совместного наступления, то есть

$$P(A + B) = P(A) + P(B) - P(A \times B). \tag{1.5}$$

Доказательство. Пусть для полной группы событий, имеющих n исходов, m_1 исходов благоприятствуют событию A, m_2 — событию B, а l исходов благоприятствуют как событию A, так и событию B. Тогда

$$P(A) = \frac{m_1}{n}, \quad P(B) = \frac{m_2}{n}, \quad P(A \times B) = \frac{l}{n}.$$

Поскольку событие $A + B$ состоит в том, что произошло событие A, либо событие B, либо событие A и B, то ему будет благоприятствовать $m_1 + m_2 - l$ исходов.

Следовательно,

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

Пример 1.4. Вероятность попадания в мишень одного стрелка равна 0,65, а второго 0,6. Определить вероятность поражения мишени при одновременных выстрелах двух стрелков.

Решение. Поскольку при стрельбе возможно попадание в мишень двумя стрелками, то эти события совместные, следовательно,

$$P(A + B) = P(A) + P(B) - P(A \times B) = 0{,}65 + 0{,}6 - 0{,}39 = 0{,}86. \triangleleft$$

1.1.5. Теорема умножения вероятностей

Определение

Событие A называется **независимым** от события B, если вероятность осуществления события A не зависит от того, произошло событие B или нет.

Например, при повторении бросания игральной кости вероятность выпадения цифры 1 (событие A) не зависит от появления или не появления цифры 1 при первом бросании кости (событие B).

Определение

Событие A называется **зависимым** от события B, если его вероятность меняется в зависимости от того, произошло событие B или нет.

Например, если в урне находятся черные и белые шары, то вероятность повторного появления черного шара (событие A) будет зависеть от того, какой шар вынули первый раз.

В случае зависимых событий A и B вводится понятие *условной вероятности*, под которой понимается вероятность события A при условии, что событие B произошло. Обозначается $P(A/B)$.

Пример 1.5. В урне находится 10 шаров: 3 белых и 7 черных. Первым был вынут черный шар. Найти вероятность того, что второй шар будет черным.

Решение. Вероятность появления черного шара первый раз (событие B) равна $P(B) = 3/10$; вероятность появления его второй раз (событие A), при условии, что событие B произошло, равна $P(A/B) = 2/9$, так как в урне осталось 9 шаров, из них 2 черных. ◁

Рассмотрим закон умножения вероятностей для независимых событий.

Повторим, что *произведением* двух событий A и B называют событие $C = A \times B$, состоящее в совместном осуществлении этих событий.

Теорема 1.4

Вероятность произведения двух независимых событий A и B равна произведению вероятностей этих событий:

$$P(A \text{ и } B) = P(A \times B) = P(A) \cdot P(B).$$

Этот закон справедлив и для n независимых событий.

$$P(A_1 \times A_2 \times \ldots \times A_n) = P(A_1) \cdot P(A_2) \cdots P(A_n). \tag{1.6}$$

Пример 1.6. В билете 3 раздела. Из 40 вопросов первого раздела студент знает 30 вопросов, из 30 вопросов второго — 15, из 30 вопросов третьего — 10. Определить вероятность правильного ответа студента по билету.

Решение. Учитывая, что ответы на каждый раздел есть независимые события A_1, A_2 и A_3, а их вероятности соответственно равны, получаем:

$$P(A_1) = \frac{30}{40} = \frac{3}{4}, \quad P(A_2) = \frac{15}{30} = \frac{1}{2}, \quad P(A_3) = \frac{10}{30} = \frac{1}{3}.$$

Тогда вероятность правильного ответа на билет $P(B)$ можно найти по формуле (1.6):

$$P(B) = P(A_1) \cdot P(A_2) \cdot P(A_3) = \frac{3}{4} \cdot \frac{1}{2} \cdot \frac{1}{3} = \frac{1}{8} = 0{,}125. \ ◁$$

Теорема 1.5

Вероятность произведения двух зависимых событий A и B равна произведению одного из них на условную вероятность второго, вычисленную при условии, что первое событие осуществилось.

$$P(A \text{ и } B) = P(A \times B) = P(A) \cdot P(B/A). \tag{1.7}$$

Формула умножения вероятностей может быть обобщена на случай событий $A_1, A_2, \ldots, A_n$:

$$P(A_1 \times A_2 \times \ldots \times A_n) =$$
$$= P(A_1)P(A_2/A_3)P(A_3/(A_1A_2)) \cdots P(A_n/(A_1A_2A_{n-1})),$$

причем вероятность каждого последующего события вычисляется в предположении, что все предыдущие произошли.

Пример 1.7. В группе из 20 человек 5 студентов не подготовили задание. Какова вероятность того, что два первых студента, вызванные наугад, будут не готовы к ответу?

Решение. Вероятность того, что первый студент не готов к ответу $P(A) = 5/20$. Вероятность того, что и второй студент так же не подготовлен, как и первый, $P(B/A) = 4/19$. Тогда для ответа на вопрос воспользуемся формулой (1.7):

$$P(A \text{ и } B) = P(A) \cdot P(B/A) = \frac{5}{20} \cdot \frac{4}{19} = \frac{20}{380} \approx 0{,}05. \ \triangleleft$$

1.1.6. Формула полной вероятности

Формула полной вероятности является обобщением теорем умножения и сложения вероятностей. Формула применяется для решения задач на определение вероятности события, которое может произойти с одним из несовместных событий, образующих полную группу.

Пусть некоторое событие A может произойти лишь вместе с одним из несовместных событий $H_1, H_2, \ldots, H_n$, имеющих вероятности $P(H_1)$, $P(H_1), \ldots, P(H_n)$, и $\sum_{i=1}^{n} P(H_i) = 1$. Такие события H_i называются *гипотезами*. Известны также и условные вероятности события A с каждой из гипотез: $P(A/H_1), P(A/H_2), \ldots, P(A/H_n)$. Тогда вероятность наступления события A определяется *формулой полной вероятности:*

$$P(A) = \sum_{i=1}^{n} P(H_i)P(A/H_i). \tag{1.8}$$

Теорема 1.6

Вероятность события A, которое может наступить только при условии появления одного из событий H_1, H_2, ..., H_n, образующих полную группу попарно несовместных событий, равна сумме произведений вероятностей каждого из событий H_1, H_2, ..., H_n на соответствующую условную вероятность события A.

Пример 1.8. На аптечный склад поступили лекарственные средства от трех фирм-производителей. Первый производитель поставил 200 ампул, второй — 200 и третий — 400. Известно, что процент бракованной продукции составит: у первого поставщика 2%, второго — 3% и третьего 1%. Для контроля была взята 1 ампула. Какова вероятность, что она окажется бракованной?

Решение. Пусть событие A — взятая ампула оказалась бракованной. Тогда возможны следующие три гипотезы:

- гипотеза H_1: ампула изготовлена первым производителем и
$$P(H_1)=\frac{200}{800}=0,25;$$
- гипотеза H_2: ампула изготовлена вторым производителем и
$$P(H_2)=\frac{200}{800}=0,25;$$
- гипотеза H_3: ампула изготовлена третьим производителем и
$$P(H_3)=\frac{400}{800}=0,5;$$

Условные вероятности события A при этих гипотезах, согласно условию задачи, равны соответственно: $P(A/H_1) = 0,02$; $P(A/H_2) = 0,03$; $P(A/H_3) = 0,01$. Тогда по формуле (1.8) имеем:

$$P(A) = P(H_1)P(A/H_1) + P(H_2)P(A/H_2) + P(H_3)P(A/H_3) =$$
$$= 0,25\cdot 0,02 + 0,25\cdot 0,03 + 0,5\cdot 0,01 = 0,0175. \triangleleft$$

1.1.7. Формула Байеса

Она применяется, когда событие A, которое может появиться только с одной из гипотез H_1, H_2, ..., H_n, произошло, и необходимо произвести количественную переоценку *априорных* вероятностей этих гипотез

$P(H_1)$, $P(H_2)$, ..., $P(H_n)$, известных до испытания, то есть найти *апостериорные* (получаемые после проведения испытания) условные вероятности гипотез $P(H_1/A)$, $P(H_2/A)$, ..., $P(H_n/A)$:

$$P(H_i/A) = \frac{P(H_i)\cdot P(A/H_i)}{P(A)}.$$

Или вместо $P(A)$ используем ее значение, вычисленное по формуле полной вероятности:

$$P(H_i/A) = \frac{P(H_i)\cdot P(A/H_i)}{P(H_i)\cdot P(A/H_i) + ... + P(H_n)P(A/H_n)}.$$

Итак, пусть до опыта имеется гипотеза H_1, H_2, ..., H_n. После опыта становится известной информация о результатах опыта, но не полная, а именно: результаты наблюдений показывают, что наступило некоторое событие A.

Считается, что до опыта были известны (*априорные*) вероятности гипотез $P(H_1)$, $P(H_2)$, ..., $P(H_n)$ и *условные* вероятности $P(A/H_1)$, $P(A/H_2)$, ..., $P(A/H_n)$. Необходимо определить *апостериорные* вероятности гипотез $P(H_1/A)$, $P(H_2/A)$, ..., $P(H_n/A)$.

Значение формулы Байеса состоит в том, что при наступлении события A, то есть по мере получения новой информации, мы можем проверять и корректировать выдвинутые до испытания гипотезы. Такой подход называется *байесовским*.

Пример 1.9. Два охотника одновременно стреляют одинаковыми пулями в медведя. В результате медведь был убит одной пулей (событие A). Как охотники должны поделить шкуру убитого медведя, если известно, что вероятность попадания у первого охотника 0,3, а у второго 0,6?

Решение. Воспользуемся формулой Байеса. Определим предварительно гипотезы.

Гипотеза H_1: попал первый охотник, второй промахнулся.

Гипотеза H_2: попал второй, первый промахнулся.

Гипотеза H_3: попали оба охотника.

Гипотеза H_4: оба промахнулись.

Событие A может произойти только тогда, когда подтвердилась либо гипотеза H_1, либо гипотеза H_2, то есть:

$$P(A/H_1) = 1,\ P(A/H_3) = 0;$$

$$P(A/H_2) = 1,\ P(A/H_4) = 0.$$

Предполагаем, естественно, что попадания охотников в медведя не зависят друг от друга. И получаем:

$$P(H_1) = 0,3 \cdot (1 - 0,6) = 0,12;$$

$$P(H_2) = 0,6 \cdot (1 - 0,3) = 0,42;$$

$$P(H_3) = 0,3 \cdot 0,6 = 0,18;$$

$$P(H_4) = (1 - 0,3)(1 - 0,6) = 0,28.$$

Применяем формулу Байеса:

$$P(H_1/A) =$$

$$= \frac{P(H_1) \cdot P(A/H_1)}{P(H_1) \cdot P(A/H_1) + P(H_2) \cdot P(A/H_2) + P(H_3) \cdot P(A/H_3) + P(H_4) \cdot P(A/H_4)} =$$

$$= \frac{0,12 \cdot 1}{0,12 \cdot 1 + 0,42 \cdot 1 + 0,18 \cdot 0 + 0,28 \cdot 0} = \frac{2}{9};$$

$$P(H_2/A) =$$

$$= \frac{P(H_2) \cdot P(A/H_2)}{P(H_1) \cdot P(A/H_1) + P(H_2) \cdot P(A/H_2) + P(H_3) \cdot P(A/H_3) + P(H_4) \cdot P(A/H_4)} =$$

$$= \frac{0,42 \cdot 1}{0,12 \cdot 1 + 0,42 \cdot 1 + 0,18 \cdot 0 + 0,28 \cdot 0} = \frac{7}{9}.$$

Таким образом, при справедливом делении первый охотник должен получить 2/9 шкуры, то есть меньше четвертой части шкуры, в то время как на первый взгляд казалось, что ему причитается 1/3 шкуры (0,3). ◁

1.2. СЛУЧАЙНЫЕ ВЕЛИЧИНЫ

Определение

Случайной называют величину, которая принимает в результате испытания то или иное возможное значение, заранее неизвестное, меняющееся от испытания к испытанию и зависящее от случайных обстоятельств.

Случайные величины могут быть дискретными и непрерывными. *Дискретной* называют такую случайную величину, которая принимает счетное множество значений, то есть такое множество, элементы которого можно подсчитать. Примером дискретной величины является количество студентов на лекции, число бракованных изделий в поставленной продукции, число новорожденных за сутки.

Непрерывной называют такую случайную величину, которая может принимать любые значения в определенном интервале. Пронумеровать все значения величины, попадающие даже в узкий интервал, принципиально невозможно. Эти значения образуют несчетное бесконечное множество.

Например, температура тела пациента за определенный промежуток времени; дальность полета футбольного мяча; объем утечки воды из городского водопровода.

Случайные величины обозначают прописными буквами латинского алфавита X, Y, Z, а их возможное значение — соответствующими строчными буквами x, y, z.

При многократных испытаниях определенные значения случайной величины могут встречаться несколько раз. Именно поэтому для задания случайной величины недостаточно перечислить лишь все ее возможные значения. Необходимо также знать, как часто могут появляться те или иные значения в результате испытания при одних и тех же условиях, то есть нужно задать вероятности их появления.

1.2.1. Распределение дискретных и непрерывных случайных величин

Случайная величина считается заданной, если известен закон распределения случайной величины.

Определение

Распределением (законом) случайной величины называется всякое соотношение между возможными значениями случайной величины и соответствующими им вероятностями.

Распределение дискретной случайной величины может быть задано в виде *таблицы*, в *графическом* и *аналитическом виде.*

Пусть дискретная случайная величина X принимает значения $X = x_1$, $X = x_2$, ..., $X = x_n$. Обозначим вероятности этих событий соответственно: $P(X = x_1) = p_1$, $P(X = x_2) = p_2$, ..., $P(X = x_n) = p_n$.

Таблица, содержащая возможные значения случайной величины и соответствующие вероятности, является простейшей формой задания распределения дискретной случайной величины (табл. 1.1).

Таблица 1.1. Распределение дискретной случайной величины в табличном виде

Значения случайной величины x_i	x_1	x_2	...	x_n
Вероятности значений p_i	p_1	p_2	...	p_n

Поскольку в результате случайная величина X всегда примет одно из своих возможных значений x_1, x_2, ... x_n, то эти случайные события образуют полную группу событий и

$$p_1 + p_2 + \ldots + p_n = \sum_{i=1}^{n} p_i = 1 .$$

Табличную формулу задания называют также *рядом распределения*. Для наглядности ряд распределения можно представить в *графическом виде*, где по оси абсцисс откладывают значения случайной величины, а по оси ординат вероятности этих значений.

Пример 1.10. Построить график ряда распределения значений частоты пульса в гипотетической группе из 47 человек (табл. 1.2).

Решение. По данным таблицы построен график (рис. 1.4), который называется *многоугольником распределения вероятностей*. ◁

Таблица 1.2. Распределение дискретной случайной величины для примера 1.10

Значения случайной величины, уд./мин	65	66	67	68	69	70	71	72	73	74	75
Значения вероятности $p(x_i)$	$\frac{2}{47}$	$\frac{2}{47}$	$\frac{4}{47}$	$\frac{5}{47}$	$\frac{7}{47}$	$\frac{7}{47}$	$\frac{6}{47}$	$\frac{4}{47}$	$\frac{4}{47}$	$\frac{3}{47}$	$\frac{3}{47}$

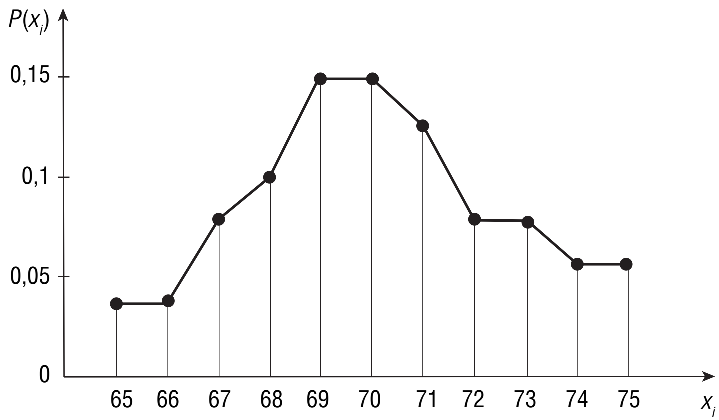


Рис. 1.4. Многоугольник распределения вероятностей

В ряде практических случаев вместо вероятности того, что случайная величина X принимает некоторое определенное значение x_i, необходимо знать, что случайная величина X меньше x_i. Эта вероятность задается интегральной функцией распределения.

Определение

Функция распределения определяет вероятность того, что случайная величина X принимает значение, меньшее фиксированного действительного числа x, то есть

$$F(x) = P(X < x).$$

Функцию распределения $F(x)$ иногда называют *интегральной функцией распределения* или *интегральным законом распределения*.

Функцию $F(x)$ можно получить, суммируя значения вероятностей по тем значениям случайной величины, которые меньше x_i, то есть

$$F(x_i) = P(X < x_i) = \sum_{x<x_i} P(x_i),$$

где неравенство $X < x_i$ под знаком суммы показывает, что суммирование распространяется на все значения x меньше x_i.

Пример 1.11. Используя данные табл. 1.2, получить интегральную функцию распределения частоты пульса.

Решение. Заполним табл. 1.3.

Значения $F(x_i)$ в табл. 1.3 получены следующим образом. Вероятность того, что $P(X < 65) = 0$, так как значений меньше 65 нет.

Тогда:

- при $x \le 65$ $F(x) = P(x < 65) = 0$ (в том числе и при $x = 65$);
- при $65 \le x \le 66$ $F(x) = P(x < 66) = P(x = 65) = 2/47$ (в том числе и при $x = 66$);
- при $66 < x \le 67$ $F(x) = P(x < 67) = P(x = 65) + P(x = 66) = 4/47$ (в том числе и при $x = 67$);
- . . . .
- при $x > 75$ $F(x) = P(x < 76) = P(x = 65) + P(x = 66) + ... + P(x = 75) = 1$.

Таблица 1.3. Интегральная функция распределения частоты пульса

Значения случайной величины x_i, уд./мин	65	66	67	68	69	70	71	72	73	74	75	76
$F(x_i)$	0	$\frac{2}{47}$	$\frac{4}{47}$	$\frac{8}{47}$	$\frac{13}{47}$	$\frac{20}{47}$	$\frac{27}{47}$	$\frac{33}{47}$	$\frac{37}{47}$	$\frac{41}{47}$	$\frac{44}{47}$	1

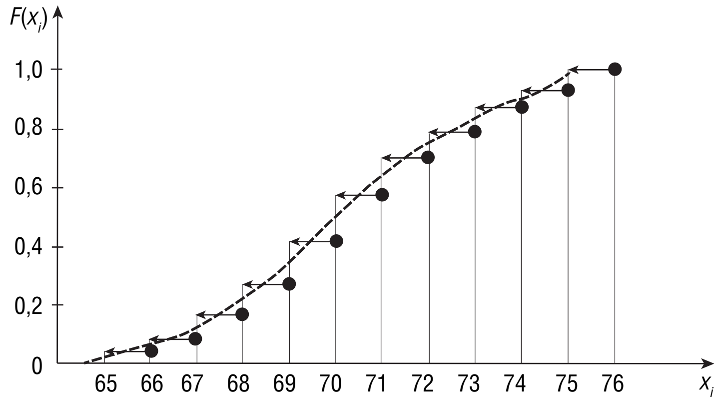


Рис. 1.5. График интегральной функции для примера 1.11

График интегральной функции по данным табл. 1.3 приведен на рис. 1.5.

Для дискретной случайной величины график функции распределения представляет собой разрывную ступенчатую функцию. Когда переменная x принимает какое-нибудь из своих возможных значений, функция распределения увеличивается скачкообразно на величину вероятности этого значения. Причем при подходе слева к точкам разрыва функция сохраняет свое значение. На графике это отмечено черной точкой. Сумма величин всех скачков функции $F(x)$ равна 1. $\triangleleft$

Свойства функции распределения

1. Функция распределения случайной величины есть неотрицательная функция, заключенная между нулем и единицей.

$$0 \le F(x) \le 1.$$

2. Функция распределения случайной величины есть неубывающая функция на всей числовой оси, и для любых $\alpha < \beta$ выполняется равенство

$$P(\alpha \le X < \beta) = F(\beta) - F(\alpha).$$

3. На минус бесконечности функция распределения равна нулю, на плюс бесконечности равна единице, то есть

$$\lim_{x \to -\infty} F(x) = 0 \quad \text{и} \quad \lim_{x \to \infty} F(x) = 1.$$

При возрастании числа значений случайной величины ($n \to \infty$ и увеличении количества интервалов на графике уменьшается их ширина

($\Delta x \to 0$), и функция распределения вместо ступенчатого принимает плавный характер (пунктирная линия на рис. 1.5).

Таким образом, интегральная функция распределения применяется для описания всех случайных величин, как дискретных, так и непрерывных.

Согласно свойствам функции распределения, зная непрерывную функцию распределения случайной величины $F(x)$, можно определить вероятность попадания случайной величины X в некоторый интервал $(x, x + \Delta x)$ — рис. 1.6.

$$P(x < X < x + \Delta x) = F(x + \Delta x) - F(x).$$

Разделим левую и правую части этого выражения на Δx и найдем предел при $\Delta x \to 0$:

$$\lim_{\Delta x \to 0} \frac{P(x < X < x + \Delta x)}{\Delta x} = \lim_{\Delta x \to 0} \frac{F(x + \Delta x) - F(x)}{\Delta x} = F'(x) = f(x).$$

Определение

Функцию $f(x)$ называют **дифференциальной функцией распределения** или **плотностью распределения (плотностью вероятности)** непрерывной случайной величины X.

Плотность распределения непрерывной случайной величины есть предел отношения вероятности $P(\Delta x)$ попадания случайной величины X в интервал Δx к величине этого интервала.

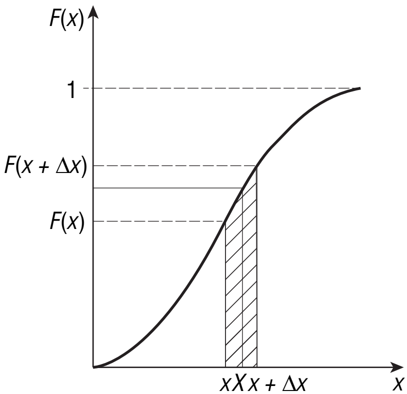


Рис. 1.6. Вероятность попадания случайной величины X в некоторый интервал

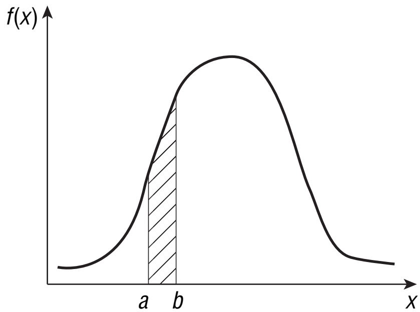


Рис. 1.7. Геометрический смысл плотности распределения вероятностей $f(x)$

Геометрический смысл плотности распределения вероятностей $f(x)$ заключается в следующем (рис. 1.7). Зная $f(x)$, можно вычислить вероятность того, что случайная величина примет значение, принадлежащее заданному интервалу (a, b):

$$P(a < X < b) = \int_a^b f(x)dx = F(b) - F(a). \qquad (1.9)$$

Основные свойства дифференциальной функции распределения

1. Для любых x дифференциальная функция распределения неотрицательна, то есть $f(x) \leq 0$.

2. Для дифференциальной функции распределения имеет место равенство:

$$P(\alpha < X < \beta) = \int_\alpha^\beta f(x)dx.$$

3. Для дифференциальной функции распределения имеет место равенство:

$$\int_{-\infty}^{\infty} f(x)dx = 1.$$

Это свойство называется *условием нормировки плотности вероятностей*.

1.2.2. Числовые характеристики случайных величин

Закон распределения полности характеризует случайную величину. Однако при решении ряда практических задач нет необходимости знать все возможные значения случайной величины и соответствующие им вероятности, а удобнее пользоваться некоторыми количественными показателями, которые в сжатой форме дают достаточную информацию о случайной величине. Такие показатели называются *числовыми характеристиками случайной величины*. Основными из них являются: *математическое ожидание, дисперсия* и *среднеквадратическое отклонение*.

Математическое ожидание характеризует положение случайной величины на числовой оси, определяя некоторое среднее значение, около которого сосредоточены все возможные значения случайной величины.

Определение

Математическое ожидание дискретной случайной величины равно сумме произведений всех возможных ее значений на соответствующие вероятности:

$$M(X) = x_1 p_1 + x_2 p_2 + \ldots + x_n p_n = \sum_{i=1}^{n} x_i p_i. \tag{1.10}$$

Для непрерывных случайных величин с плотностью распределения $f(x)$ математическое ожидание равно определенному интегралу:

$$M(X) = \int_a^b x f(x) dx, \tag{1.11}$$

где a и b — пределы интегрирования, которые соответствуют крайним границам возможных значений X.

В общем случае

$$M(X) = \int_{-\infty}^{\infty} x f(x) dx. \tag{1.12}$$

Формула (1.11) получается из формулы (1.10), если в ней заменить отдельные значения x_i на непрерывно изменяющийся параметр x, соответствующие вероятности p_i — на элемент вероятности $f(x)dx$, конечную сумму — на интеграл.

Математическое ожидание является центром распределения вероятностей случайной величины X. На рис. 1.8 приведены графики распределения случайной величины, описанные одинаковым законом, но имеющие различные значения математического ожидания.

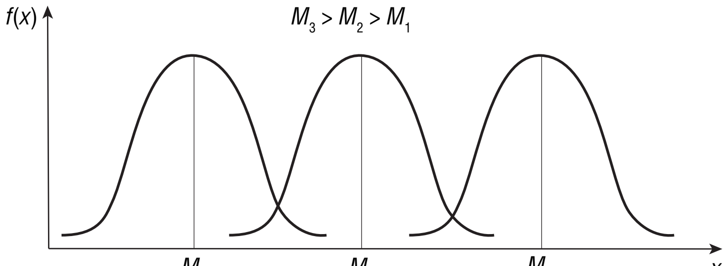


Рис. 1.8. Графики распределения случайной величины с одинаковым законом распределения, но разными математическими ожиданиями

Пример 1.12. Найти математическое ожидание дискретной случайной величины X, зная закон ее распределения (табл. 1.4).

Таблица 1.4. Распределение дискретной случайной величины для примера 1.12

X	–1	0	1	2	3
p	0,05	0,2	0,4	0,3	0,05

Решение. По формуле (1.10) находим:

$$M(X) = -1 \cdot 0{,}05 + 0 \cdot 0{,}2 + 1 \cdot 0{,}4 + 2 \cdot 0{,}3 + 3 \cdot 0{,}05 = 1{,}1. \lhd$$

Пример 1.13. Найти математическое ожидание непрерывной случайной величины X, зная закон ее распределения:

$$f(x) = \begin{cases} 0 & \text{при } x < 0; \\ x & \text{при } 0 \le x < 1; \\ -x + 2 & \text{при } 1 \le x < 2; \\ 0 & \text{при } x \ge 2. \end{cases}$$

Решение. По формуле (1.11) находим:

$$M(X) = \int_0^2 xf(x)dx = \int_0^1 x^2 dx + \int_1^2 x(2-x)dx = \left.\frac{x^3}{3}\right|_0^1 + \left.x^2\right|_1^2 - \left.\frac{x^3}{3}\right|_1^2 = 1. \lhd$$

Свойства математического ожидания

1. Математическое ожидание постоянной величины равно этой постоянной.

$$M(C) = C.$$

2. Постоянный множитель можно выносить за знак математического ожидания.

$$M(CX) = CM(X).$$

3. Математическое ожидание алгебраической суммы случайных величин равно алгебраической сумме их математических ожиданий.

$$M(X \pm Y) = M(X) \pm M(Y).$$

4. Две случайные величины X и Y называются независимыми, если распределение одной из них не зависит от того, какое значение приняла другая величина.

Математическое ожидание произведения независимых случайных величин равно произведению их математических ожиданий.

$$M(X \cdot Y) = M(X) \cdot M(Y).$$

5. Математическое ожидание отклонения случайной величины от ее математического ожидания всегда равно нулю:

$$M(X - M(X)) = 0.$$

Дисперсия характеризует рассеяние (отклонение) случайной величины относительно математического ожидания.

Определение

Математическое ожидание квадрата отклонения случайной величины X от ее математического ожидания $M(X)$ называют **дисперсией случайной величины *X*** и обозначают $D(X)$, то есть

$$D(X) = M(X - M(X))^2. \tag{1.13}$$

Для *дискретных случайных величин* эту формулу можно записать в следующем виде:

$$D(X) = \sum_{i=1}^{n} \left[x_i - M(X)\right]^2 p_i. \tag{1.14}$$

Для *непрерывных случайных величин* с плотностью вероятности $f(x)$:

$$D(X) = \int_{-\infty}^{\infty} \left(x - M(X)\right)^2 f(x)dx. \tag{1.15}$$

На рис. 1.9 приведены графики плотности распределения случайных величин с одинаковыми значениями математического ожидания, но различными дисперсиями. Следует отметить, что

$$\int_{-\infty}^{\infty} f_1(x)dx = \int_{-\infty}^{\infty} f_2(x)dx = 1.$$

Размерность дисперсии равна квадрату случайной величины, и ее неудобно использовать для характеристики разброса, поэтому удобнее применять корень квадратный из дисперсии — *среднеквадратическое отклонение*. Эта величина дает представление о размахе колебаний случайной величины около математического ожидания.

$$\sigma = \sqrt{D(X)}. \tag{1.16}$$

Пример 1.14. Случайная величина задана рядом распределения, в табличном виде (табл. 1.5). Найти математическое ожидание и дисперсию этой величины.

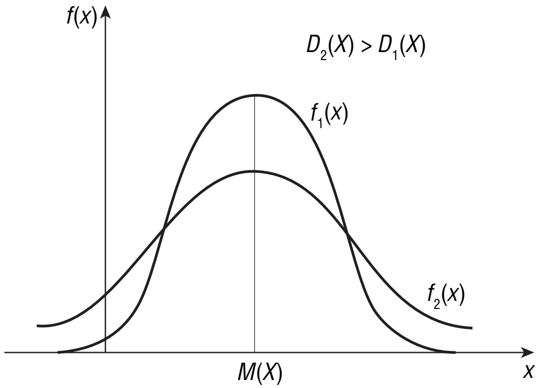


Рис. 1.9. Графики плотности распределения случайных величин с одинаковыми значениями математического ожидания, но различными дисперсиями

Таблица 1.5. Ряд распределения дискретной случайной величины для примера 1.14

X	–1	0	1	2
p	0,1	0,3	0,4	0,2

Таблица 1.6. Результаты вычислений для примера 1.14

x	p_i	x_ip_i	$x_i - M(X)$	$(x_i - M(X))^2$	$(x_i - M(X))^2p_i$
–1	0,1	–0,1	–1,7	2,89	0,289
0	0,3	0	–0,7	0,49	0,147
1	0,4	0,4	0,3	0,09	0,036
2	0,2	0,4	1,3	1,69	0,338
Σ	1	0,7			0,81

Решение. Для нахождения математического ожидания воспользуемся формулой (1.10), а для дисперсии (1.14). Результаты вычисления сведем в табл. 1.6.

Из табл. 1.6 следует, что $M(X) = 0{,}7$; $D(X) = 0{,}81$. ◁

Пример 1.15. Случайная величина задана плотностью вероятности:

$$f(x) = \begin{cases} 0 & \text{при } x < 0; \\ 2x & \text{при } 0 \le x \le 1; \\ 0 & \text{при } x > 0. \end{cases}$$

Найти математическое ожидание и дисперсию этой случайной величины.

Решение. Математическое ожидание найдем по формуле (1.11).

$$M(X)=\int_{-\infty}^{\infty} x f(x)dx=\int_{0}^{1} 2x^2 dx=\left.\frac{2x^3}{3}\right|_0^1=\frac{2}{3}.$$

Далее по формуле (1.15) получим:

$$D(X)=\int_{-\infty}^{\infty}\left(x-M(x)\right)^2 f(x)dx=\int_0^1\left(x-\frac{2}{3}\right)^2 2xdx=\int_0^1\left(x^2-\frac{4x}{3}+\frac{4}{9}\right)2xdx=$$

$$=\int_0^1\left(2x^3-\frac{8x^2}{3}+\frac{8x}{9}\right)dx=2\left.\left(\frac{x^4}{4}-\frac{4x^3}{9}+\frac{2x^2}{9}\right)\right|_0^1=\frac{1}{18}. \triangleleft$$

Основные свойства дисперсии

1. Дисперсия алгебраической суммы двух независимых случайных величин X и Y равна сумме дисперсий этих величин.

$$D(X \pm Y) = D(X) \pm D(Y).$$

2. Дисперсия постоянной величины C равна нулю, то есть $D(C) = 0$.
3. Постоянный множитель C случайной величины X можно выносить за знак дисперсии, предварительно возведя его в квадрат.

$$D(CX) = C^2D(X).$$

4. Дисперсия случайной величины X равна разности между математическим ожиданием квадрата случайной величины и квадратом ее математического ожидания.

$$D(X) = M(X^2) - [M(X)]^2. \qquad (1.17)$$

Последняя формула более удобна для вычислений, чем формула (1.14).

1.2.3. Законы распределения непрерывных случайных величин

Распределение вероятностей носит название теоретического распределения случайной величины. Теоретические распределения получают исходя из некоторых предположений относительно простейших закономерностей данного явления. Знание теоретических законов распределения изучаемых случайных величин позволяет оценивать их параметры, определять допустимые отклонения от истинных значений, проверять гипотезы и т.д.

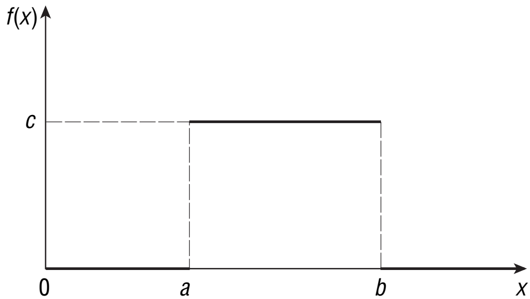


Рис. 1.10. График равномерного распределения

Рассмотрим наиболее распространенные законы распределения непрерывных случайных величин.

Равномерное распределение

Равномерным называется распределение непрерывных случайных величин, все значения которых лежат на отрезке [a; b] и имеют постоянную плотность вероятности на этом отрезке (рис. 1.10).

График равномерно распределенной случайной величины может быть описан следующей функцией:

$$f(x)=\begin{cases}0 & \text{при } x<a;\\ c & \text{при } a\le x\le b;\\ 0 & \text{при } x>b.\end{cases}$$

Учитывая, что площадь под кривой распределения равна 1, получим:

$$\int_{-\infty}^{\infty} f(x)dx=\int_a^b cdx=cx\int_a^b c(b-a)=1\text{, отсюда } c=\frac{1}{b-a}.$$

Тогда плотность вероятности случайной величины, равномерно распределенной на отрезке [a, b], имеет вид:

$$f(x)=\begin{cases}0 & \text{при } x<a;\\ \dfrac{1}{b-a} & \text{при } a\le x\le b;\\ 0 & \text{при } x>b.\end{cases}$$

Нормальный закон распределения

Одним из наиболее важных и часто используемых распределений непрерывных случайных величин является *нормальное распределение (закон Гаусса).*

Теоретическим обоснованием частого использования нормального распределения является одна из *центральных предельных теорем*. Согласно этой теореме распределение среднего n независимых случайных величин, распределенных по любому закону или даже имеющих n различных распределений с конечным математическим ожиданием и дисперсией, при увеличении числа наблюдений в выборе приближается к нормальному. Таким образом, если случайная величина подвержена достаточно большому числу независимых «небольших» воздействий, то можно ожидать, что эта случайная величина будет иметь распределение, близкое к нормальному. Этим объясняется широкое использование нормального закона распределения в теории вероятностей.

Однако в тех случаях, когда имеет место воздействие одного фактора и значительное его превышение над другими, более мелкими факторами, нельзя использовать нормальный закон распределения. Необходимо соответствующее теоретическое обоснование выбора закона распределения или экспериментальная проверка соответствия тому или иному закону распределения.

Непрерывная случайная величина X имеет *нормальное распределение вероятностей* с параметрами a и σ, если ее плотность распределения задается формулой:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}},$$

где a — математическое ожидание случайной величины; σ^2 — дисперсия случайной величины; σ — среднее квадратическое отклонение.

Кривая нормального закона распределения имеет колоколообразную форму, симметричную относительно математического ожидания (рис. 1.11).

Основные свойства нормального распределения

1. Областью распределения функции $f(x)$ является вся числовая ось.

2. Функция $f(x)$ может принимать только положительные значения, то есть $f(x) > 0$.

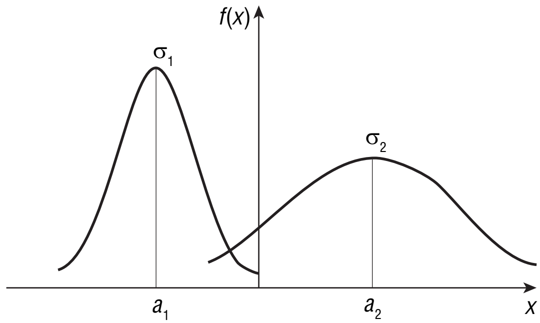


Рис. 1.11. Графики плотностей нормального распределения с разными *a* и σ

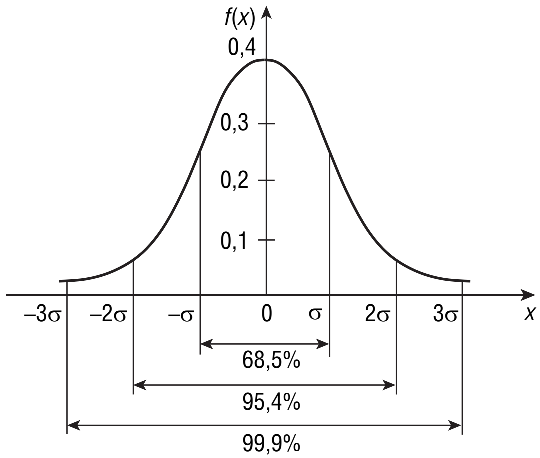


Рис. 1.12. Распределение площадей под кривой плотности стандартного нормального распределения

3. Предел функции $f(x)$ при неограниченном возрастании $|x|$ равен нулю, то есть ось $0x$ является горизонтальной асимптотой графика функции.

4. Функция $f(x)$ имеет в точке $x = a$ максимум, равный $\frac{1}{\sigma\sqrt{2\pi}}$.

5. График функции $f(x)$ симметричен относительно прямой $x = a$.

6. Кривая нормального распределения в точках $x = a \pm \sigma$ имеет перегиб.

Параметр a характеризует положение графика на числовой оси (параметр положения), а σ — степень сжатия или растяжения графика относительно оси a (параметр сжатия). Площадь под кривой во всех случаях должна быть одинаковой и равной 1 (условие нормировки).

Полученные данные показывают, что на расстоянии σ по обе стороны от математического ожидания находится 68,5% возможных значений случайной величины, в диапазон $\pm 2\sigma$ попадает 95,4%, а в диапазоне $\pm 3\sigma$ (трех сигм) находится 99,86%, то есть практически все возможные значения случайной величины X (рис. 1.12).

Отсюда вытекает правило «трех сигм»: *если случайная величина X имеет нормальный закон распределения с параметрами a и σ^2, то ее значения практически не выходят за интервал* $(a - 3\sigma;\ a + 3\sigma)$.

Распределение χ^2 (хи-квадрат)

Распределением χ^2 с k степенями свободы называется распределение суммы квадратов k независимых случайных величин, распределенных по стандартному нормальному закону, то есть

$$\chi^2 = \sum_{i=1}^{k} Z_i^2 ,$$

где Z_i ($i = 1, 2, \ldots, k$) имеет нормальное распределение $N(0; 1)$.

Кривые χ^2-распределения для различных значений числа степеней свободы k приведены на рис. 1.13. Видно, что с увеличением k распределение медленно приближается к нормальному.

В приложении (см. табл. П7) приведены числовые значения распределения χ^2 при некоторых k.

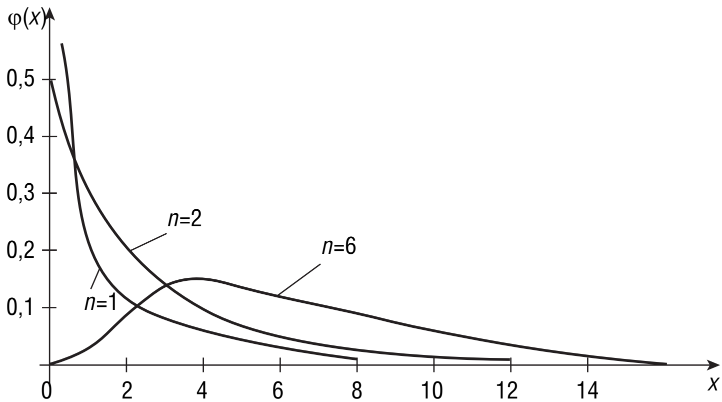


Рис. 1.13. Распределение χ^2