Лекции по теории вероятностей и математической статистике.

Ю. В. Прохоров Л. С. Пономаренко

2004 г.

Оглавление

Введение **v**

1 Вероятностное пространство **2**
1.1 Конечное вероятностное пространство 2
1.2 Классическая вероятность 3
1.2.1 Генуэзская лотерея 3
1.2.2 Игральные кости 5
1.2.3 Случайные перестановки 7
1.2.4 Игра в бридж 8
1.2.5 Абсолютно случайные последовательности 9

2 Случайные величины и случайные события **11**
2.1 Случайные величины 11
2.2 Операции над случайными событиями 12
2.3 Операции над индикаторами 14

3 Свойства вероятности и математического ожидания **15**
3.1 Свойства вероятности 15
3.2 Свойства математического ожидания случайных величин . 18
3.3 Вероятность появления хотя бы одного из n событий 20

4 Независимость случайных событий и случайных величин **24**
4.1 Условная вероятность. Независимость двух случайных событий 24
4.2 Независимость случайных величин. Взаимная независимость нескольких случайных событий 25
4.3 Свойства независимых случайных величин и взаимно независимых случайных событий 27
4.4 Критерий независимости случайных величин 29
4.5 Мультипликативное свойство математического ожидания независимых случайных величин 30

5 Суммирование независимых случайных величин 32

5.1 Производящая функция целочисленной случайной величины 33

5.2 Производящая функция моментов 34

5.3 Свойства числовых характеристик распределений сумм независимых случайных величин 36

6 Неравенства Чебышева. Отклонения сумм независимых случайных величин 40

6.1 Схемы Бернулли и Пуассона 40

6.2 Неравенства Чебышева 41

6.3 Отклонения сумм независимых случайных величин 43

7 Закон больших чисел 47

7.1 Закон больших чисел в форме Чебышева 47

7.2 Теорема Бернулли. Отклонение частоты наступления события от его вероятности 49

7.3 Вероятностное доказательство теоремы Вейерштрасса . . . 51

8 Неравенства для максимума сумм независимых случайных величин 54

8.1 Неравенство А. Н. Колмогорова. 55

8.2 Неравенство Поля Леви 57

9 Математические основы теории вероятностей 62

9.1 Общее определение вероятностного пространства 62

9.1.1 Порожденные алгебры и σ - алгебры 66

9.1.2 Борелевские σ - алгебры множеств 68

9.1.3 Вероятностные меры или распределения вероятностей 69

9.2 Вероятностные меры в евклидовых пространствах 74

9.2.1 Вероятностные распределения на прямой 74

9.2.2 Вероятностные распределения на плоскости и в пространстве 77

9.2.3 Два основных типа распределений в евклидовых пространствах 79

9.3 Случайные величины 80

9.3.1 σ- алгебра, порожденная случайной величиной . . . 84

9.3.2 Распределения случайных величин 84

9.4 Математические ожидания случайных величин (Общий случай) 85

9.4.1 Основные свойства математического ожидания . . . 86

9.5 Независимые случайные величины 90
9.6 Мультипликативное свойство математического ожидания . 92

10 Усиленный закон больших чисел 94
10.1 Лемма Бореля – Кантелли 94
10.2 Сходимость с вероятностью 1 96
10.3 Усиленный закон больших чисел 97

11 Предельные теоремы и метод характеристических функций 100
11.1 Обозначения и формулировки предельных теорем 100
11.2 Характеристические функции. Определение и свойства . . 103
11.3 Формулы обращения для характеристических функций . . 106
11.4 Свойство непрерывности соответствия характеристических функций и функций распределения 107
11.5 Примеры слабой сходимости последовательностей характеристических функций 114
11.6 Доказательство центральной предельной теоремы 116
11.7 Теорема Пуассона 118

12 Задачи математической статистики. Основные понятия 127
12.1 Сходимость по вероятности 129
12.2 Асимптотическая нормальность 131
12.3 Некоторые важные преобразования случайных величин . . 133
12.4 Эмпирическая функция распределения 136

13 Проверка гипотезы о виде распределения 140
13.1 Критерий согласия А. Н. Колмогорова. 140
13.2 Критерий согласия Пирсона "хи – квадрат" 142

14 Проверка параметрических гипотез. Фундаментальная лемма Неймана – Пирсона 147
14.1 Квантили и процентные точки нормального распределения 147
14.2 Постановка задачи. Ошибки первого и второго рода. 149
14.3 Лемма Неймана – Пирсона 151
14.4 Проверка гипотез о параметрах нормального распределения 153
14.5 Проверка гипотез о параметре биномиального распределения 157

15 Доверительные интервалы 158
15.1 Постановка задачи и основные определения 158

15.2 Доверительный интервал для математического ожидания нормального распределения при известной дисперсии . . . 159
15.3 Построение доверительного интервала для дисперсии нормального распределения 161
15.3.1 Совместное распределение статистик $\overline{X}$ и S^2 162
15.4 Доверительный интервал для неизвестного математического ожидания нормальной выборки при неизвестной дисперсии. 165
15.4.1 Распределение Стьюдента. 165
15.5 Асимптотический доверительный интервал для параметра p биномиального распределения 167

16 Точечные оценки для неизвестных параметров 169
16.1 Сравнение свойств несмещенных оценок 169
16.2 Семейства распределений 171
16.3 Метод максимального правдоподобия. 172
16.4 Неравенство Рао - Крамера 172

Приложение 1. Основные распределения и их свойства 178

Приложение 2. Экзаменационные вопросы по курсу "Теория вероятностей и математическая статистика" 185

Введение

Настоящее учебное пособие основано на материале лекций, которые много лет читались студентам второго курса факультета вычислительной математики и кибернетики.

Вашему вниманию предлагается самый краткий вариант. В отдельные годы он дополнялся и другими вопросами, и представление о полной программе этого двухсеместрового курса можно получить из приведенного в приложении 2 списка экзаменационных вопросов.

Особое внимание обращается на оценки вероятностей либо в виде приближенных формул, либо в виде неравенств. Это вполне соответствует классической традиции, когда в названии книг не было слов "теория вероятностей", а употреблялись выражения "исчисление вероятностей". В качестве примеров можно привести учебники Пуанкаре, Маркова.

Ряд теорем приводится без доказательства. Мы считаем, что студентам нужно уметь пользоваться этими теоремами. Источники, в которых при желании можно найти эти доказательства, обычно указываются.

При изложении вопросов математической статистики произведен очень жесткий отбор материала. Часто из учебников большого объема математическая статистика представляется как цепочка определений, лемм и теорем без применений этих теорем к обработке статистического материала. Примером удачного соединения теории и практики может служить книга Г. Крамера "Математические методы статистики"([6]).

Мы надеемся в будущем дополнить эту книгу новыми главами, которые могут быть полезны не только студентам, интересующимся теорией вероятностей, но и начинающим преподавателям этой дисциплины.

Мы начинаем изложение со случая конечных вероятностных пространств, поскольку к началу третьего семестра студенты не имеют еще достаточных знаний по математическому анализу. С другой стороны, это позволяет иметь дело только с конечными суммами, все функции от элементарных исходов являются случайными величинами, математические ожидания и дисперсии существуют и можно продвинуться с этими средствами достаточно далеко, например, доказать закон больших чисел и даже центральную предельную теорему.

Мы будем благодарны всем читателям, которые пожелают сообщить нам свои замечания.

Лаплас: "...теория вероятностей есть в сущности не что иное, как здравый смысл, сведенный к исчислению."

Глава 1

Вероятностное пространство

"Теория вероятностей — математическая наука, позволяющая по *вероятностям* одних случайных событий находить вероятности других случайных событий, связанных каким-либо способом с первыми",— такое определение приводится в математической энциклопедии.

Знакомство с этой математической наукой начнем с основного понятия теории вероятностей — понятия вероятностного пространства.

Вероятностным пространством называется тройка элементов

$$(\Omega, \mathcal{A}, \mathbf{P}),$$

в которой Ω – множество элементарных исходов ω, $\mathcal{A}$ – некоторый класс подмножеств Ω, называемых случайными событиями, $\mathbf{P}$ – распределение вероятностей или вероятность случайных событий.

Разберем все эти понятия сначала для конечного вероятностного пространства.

1.1 Конечное вероятностное пространство

Пусть Ω– некоторое конечное множество, состоящее из s элементов $\omega_1, \ldots, \omega_s$, называемых элементарными исходами или элементарными событиями.

Обычно это множество выбирается в соответствии с проводимым случайным экспериментом таким образом, чтобы всякий мыслимый неразложимый исход эксперимента описывался единственным элементарным исходом.

Событием в этом случае будем называть любое подмножество множества Ω. Таким образом $\mathcal{A}$ будет состоять из 2^s событий (подмножеств), включая подмножество, совпадающее с Ω, и пустое подмножество $\varnothing$. Событие Ω называют достоверным, а $\varnothing$ – невозможным событием.

Распределение $\mathbf{P}$ для конечных вероятностных пространств задается следующим образом: каждому элементарному исходу ω_i ставится в соответствие число $\mathbf{P}(\omega_i) = p_i \geq 0$ так, чтобы $\sum\limits_{i=1}^{s} p_i = 1$.

Тогда вероятность $\mathbf{P}(A)$ произвольного случайного события $A = \{\omega_{i_1}, \dots \omega_{i_r}\}$ определяется следующим образом:

$$\mathbf{P}(A) = \sum_{k=1}^{r} p(\omega_{i_k}) = \sum_{\omega \in A} p(\omega).$$

Исходы $\omega \in A$ называют *благоприятствующими* для события A, так как наступление любого такого исхода повлечет наступление события A.

Подробнее рассмотрим частный случай конечного вероятностного пространства, который будем называть классической вероятностной моделью.

1.2 Классическая вероятность

Пусть все s возможных исходов случайного эксперимента равновероятны, тогда в соответствии с изложенным выше $p(\omega_i) = \frac{1}{s}$ для любого i и

$$\mathbf{P}(A) = \frac{r_A}{s},$$

где s– по-прежнему общее число возможных исходов, а r_A– число исходов, благоприятствующих A.

1.2.1 Генуэзская лотерея

Первое упоминание об этой лотерее в литературе относится к 1620 г. Условия лотереи таковы: производится розыгрыш 5 выигрышных билетов из 90 имеющихся. Все билеты пронумерованы числами от 1 до 90. Участники розыгрыша могут делать ставки: на 1 номер (простая единица), на 2 номера (амбо), на 3 номера (терно), на 4 номера (кватерно), на 5 номеров (квинтерно).

При этом, если игрок угадывал, то он получал свою ставку, умноженную на 15 для простой единицы, на 270 для амбо, на 5500 для терно, на 75000 для кватерно, на 1000000 для квинтерно. Пользуясь моделью классической вероятности, найдем вероятности таких выигрышей.

Для этого введем элементарные исходы как упорядоченные наборы из 5 различных между собой чисел :

$$\omega = (i_1, i_2, i_3, i_4, i_5),$$

где i_j – номер числа, выпавшего в j- ом розыгрыше, $1 \le i_j \le 90$.

Подсчет общего числа исходов s проводится следующим образом: для i_1 имеется всего 90 различных вариантов, для i_2 уже на единицу меньше – 89, для i_3 – 88, для i_4 – 87, для i_5 – 86. Перемножив эти числа, получим общее число исходов

$$s = 90 \cdot 89 \cdot 88 \cdot 87 \cdot 86 = 5273912160.$$

Тогда вероятность любого элементарного исхода ω равна $\mathbf{P}(\omega) = \frac{1}{s}$. Введем события A_i – выиграть при ставке на i чисел. Проведем для примера вычисление $\mathbf{P}(A_2)$.

Пусть 7 и 11 — два числа, на которые делается ставка амбо. Тогда любой благоприятный исход для события A_2 будет иметь вид

$$(*, 7, *, 11, *),$$

где * помечены различные между собой числа, не равные 7 и 11, причем 7 и 11 могут появиться в этой цепочке на любых местах. Таких благоприятствующих исходов будет

$$C_5^2 \cdot 2 \cdot 88 \cdot 87 \cdot 86,$$

где $C_5^2 \cdot 2$ – это число способов выбора двух мест, на которых можно в любом порядке расположить числа 7 и 11, 88 – возможные варианты выбора числа на первое из оставшихся мест, 87 – на второе оставшееся место, 86 – на третье свободное место. Получаем, что

$$\mathbf{P}(A_2) = \frac{C_5^2 \cdot 2 \cdot 88 \cdot 87 \cdot 86}{90 \cdot 89 \cdot 88 \cdot 87 \cdot 86} = \frac{2}{801}.$$

Задача 1.1. Убедитесь, что это число можно получить и другим способом, вводя в качестве элементарных исходов неупорядоченные наборы различных между собой 5 чисел. Тогда общее число исходов $s' = C_{90}^5$, а

$$\mathbf{P}(A_2) = \frac{C_2^2 \cdot C_{88}^3}{C_{90}^5}.$$

Аналогичным образом можно вычислить все вероятности $\mathbf{P}(A_i)$. Приведем результаты вычислений в таблице

Таблица 1.1: вероятности выигрышей в генуэзской лотерее

i	увеличение ставки в m_i раз	$\mathbf{P}(A_i)$
1	15	$\frac{1}{18} \approx 0.055555$
2	270	$\frac{2}{801} \approx 0.002497$
3	5500	$\frac{1}{11748} \approx 0.0000851$
4	75000	$\frac{1}{511038} \approx 0.0000019$
5	1000000	$\frac{1}{43949268} \approx 10^{-8} \cdot 2.275$

Задача 1.2. Какова вероятность того, что при розыгрыше лотереи номера выйдут в возрастающем или убывающем порядке? (Ответ: 1/60.)

Приведем имеющиеся статистические данные по розыгрышам этой лотереи в Праге. За 133 года с 1754г. по 1886г. было проведено 2854 тиража, при этом номера выходили в возрастающем или убывающем порядке с частотой 0.01612, тогда как точное значение вероятности такого события 0.01667. Можно ли считать классическую модель, в рамках которой была вычислена эта вероятность удовлетворительной? Ответ на этот вопрос мы сможем дать, когда познакомимся с основными понятиями теории проверки статистических гипотез во второй части нашего курса.

1.2.2 Игральные кости

Игроки бросают по 6 игральных костей. Ставка игрока составляет 10 копеек. Если сумма выпавших очков 6 или 36, то игрок получает премию 78 рублей, 7 или 35 очков – 24 рубля, 8 или 34 – 4 рубля, в остальных случаях не получает ничего.[1])

Проведем вычисление вероятностей данных премий. Для этого занумеруем кости от 1 до 6 и введем элементарный исход как

$$\omega = (i_1, i_2, i_3, i_4, i_5, i_6),$$

где $1 \leq i_j \leq 6$ – результат бросания $j-$ой кости. Всего элементарных исходов $s = 6^6 = 46656$.

[1]Игра относится ко второй половине 19 века и описана в книге П. Л. Чебышева "Курс лекций по теории вероятностей"(курс читался в С.– Петербургском университете в 1879/1880гг.[11]

Обозначим $S(\omega) = i_1 + i_2 + i_3 + i_4 + i_5 + i_6$ – сумму выпавших очков. Тогда

$$\mathbf{P}\{S(\omega) = 6\} = \mathbf{P}\{S(\omega) = 36\} = \frac{1}{s},$$

так как единственными благоприятствующими исходами для этих событий являются $(1, 1, 1, 1, 1, 1)$ и $(6, 6, 6, 6, 6, 6)$ сответственно. Следовательно, вероятность выиграть 78 рублей равна

$$p_1 = \frac{2}{s} = \approx 0.000042866941.$$

Аналогичными рассуждениями можно получить

$$\mathbf{P}\{S(\omega) = 7\} = \mathbf{P}\{S(\omega) = 35\} = \frac{6}{s},$$

так как для каждого из этих событий имется уже по 6 благоприятствующих исходов, например событие $S(\omega) = 7$ произойдет в любом из 6 случаев

$$\begin{array}{lll} (2,1,1,1,1,1) & (1,2,1,1,1,1) & (1,1,2,1,1,1) \\ (1,1,1,2,1,1) & (1,1,1,1,2,1) & (1,1,1,1,1,2). \end{array}$$

Значит, вероятность выиграть 24 рубля равна

$$p_2 = \frac{2 \cdot 6}{s} \approx 0.000257202646.$$

Подобными рассуждениями, представив всеми возможными способами число $8 = 1 + 1 + 1 + 1 + 2 + 2 = \ldots = 1 + 1 + 1 + 1 + 1 + 3$, получим, что

$$\mathbf{P}\{S(\omega) = 8\} = \mathbf{P}\{S(\omega) = 34\} = \frac{C_6^2 + C_6^1}{6^6} = \frac{15 + 6}{6^6},$$

и вероятность выиграть 4 рубля равна

$$p_3 = \frac{2 \cdot 21}{6^6} \approx 0.000900206761.$$

Таким образом вероятность выиграть хоть что-нибудь в этой игре равна

$$p_1 + p_2 + p_3 = 0.001200274,$$

а средний выигрыш равен

$$(78 \cdot p_1 + 24 \cdot p_2 + 4 \cdot p_3) \times 100 = 1.3117284\,(\text{коп.}),$$

что составляет около 1 копейки и почти в 10 раз меньше ставки.[2] Как видим, условия этой игры весьма невыгодны для игроков, ставки должны быть повышены.

Вычисление вероятностей $\mathbf{P}\{S(\omega) = k\}$ было проведено в простых случаях, при приближении k к среднему значению 19 вычисления становятся все более громоздкими и требуется уже некий алгоритм, заменяющий простой перебор.

Разберем способ, использующий бином Ньютона и разложения в ряды для простых функций и применимый при любом числе игральных костей.

Рассмотрим многочлен 36 степени

$$Q(x) = (x + x^2 + x^3 + x^4 + x^5 + x^6)^6 = a_{36}x^{36} + a_{35}x^{35} + \cdots + a_1x + a_0,$$

($a_0 = 0$). Коэффициент a_8, к примеру, в этом разложении равен числу способов, которыми можно получить 8, т.е. 21. Это мы вычислили выше. Но можно этот коэффициент получить и по-другому. Используя формулу для суммы геометрической прогрессии, получим, что

$$Q(x) = \left(\frac{x(1-x^6)}{1-x}\right)^6 = \frac{x^6(1-x^6)^6}{(1-x)^6}.$$

Поскольку

$$\frac{1}{(1-x)^6} = (1-x)^{-6} = 1+\frac{(-6)}{1}(-x)+\frac{(-6)(-7)}{1\cdot 2}(-x)^2+\frac{(-6)(-7)(-8)}{1\cdot 2\cdot 3}(-x)^3+\ldots,$$

то

$$Q(x) = x^6(1 - C_6^1x^6 + C_6^2x^{12} - \cdots + x^{36}) \cdot (1 + 6x + 21x^2 + \cdots) =$$
$$= x^6(1 - 6x^6 + 15x^{12} + \cdots)(1 + 6x + 21x^2 + \cdots).$$

Следовательно, $a_8 = 21$, а $a_{15} = 1666$.

1.2.3 Случайные перестановки

Рассмотрим в качестве элементарных исходов ω различные перестановки чисел $1, 2, \ldots, n$, их будет $n!$, всем исходам приписываем вероятность $\frac{1}{n!}$.

Пусть A_2 – событие, состоящее в том, что число 2 появится на втором месте, тогда

$$\mathbf{P}(A_2) = \frac{(n-1)!}{n!} = \frac{1}{n}.$$

[2]Чебышев [11, стр.213] пишет: "...так что 9/10 ставки шло в пользу откупщика; поэтому такие лотереи во всех государствах упразднены.

Какова вероятность того, что хотя бы одно число встретится на своем месте? Обозначим это событие A.

Для небольших n можно выписать все возможные исходы и подсчитать среди них число благоприятствующих. Так для $n = 3$ имеем 6 исходов:

$$\begin{array}{lll} \omega_1 = (1,2,3) & \omega_2 = (1,3,2) & \omega_3 = (2,1,3) \\ \omega_4 = (2,3,1) & \omega_5 = (3,1,2) & \omega_6 = (3,2,1). \end{array}$$

Для события A благоприятствующими являются исходы $\omega_1, \omega_2, \omega_3, \omega_6$. Следовательно,

$$\mathbf{P}(A) = \frac{4}{6} = \frac{2}{3}.$$

Такой способ вычисления вероятности события A, конечно, не годится при больших n. В дальнейшем мы сможем показать, что $\mathbf{P}(A) \approx 1 - \frac{1}{e}$, но для этого потребуется формула для вычисления вероятности объединения конечного числа случайных событий.

1.2.4 Игра в бридж

Рассмотрим колоду, состоящую из 52 карт. Четырем игрокам раздают по 13 карт. Какова вероятность того, что у одного из них, скажем у первого, на руках окажется полная масть?

Обозначим это событие B. Занумеруем карты следующим образом: номерами от 1 до 13 червы, от 14 до 26 – бубны, от 27 до 39 – трефы и от 40 до 52 – пики.

Здесь элементарный исход ω — это перестановка из 52 чисел, описывающая начальное расположение карт перед их раздачей. При этом первому игроку достаются первая, пятая, девятая и т.д. карты из колоды, второму – вторая, шестая и т.д.

Первому игроку достанутся все карты червовой масти, если на местах $1, 5, 9, \ldots$ стоят числа $1, 2, \ldots, 13$ в том или ином порядке. Поскольку всего исходов $s = 52!$, благоприятствующих – $4 \cdot 13! \cdot 39!$, то

$$\mathbf{P}(B) = \frac{4 \cdot 13! \cdot 39!}{52!} \approx 6.3 \cdot 10^{-12}.$$

Хотя это событие и маловероятно, но возможно, что игроки в бридж могли наблюдать его.[3]

[3] Подробнее о вычислении вероятностей при игре в бридж можно прочитать в [9]

1.2.5 Абсолютно случайные последовательности

Рассмотрим в качестве элементарных исходов $\omega = (\delta_1, \delta_2, \dots \delta_n)$ –цепочки длины n, где δ_i – это либо 0, либо 1, причем появление 0 или 1 на каждом месте равновозможно.[4] Всего таких исходов будет $s = 2^n$. Вероятность каждого отдельного исхода равна $\mathbf{P}(\omega) = \frac{1}{2^n}$. При больших n эти вероятности очень малы, так при $n = 100 \quad \mathbf{P}(\omega) \approx 10^{-30}$, а при $n = 1000 \quad \mathbf{P}(\omega) \approx 10^{-300}$.

Такие последовательности играют важную роль в теории информации, в частности используются при кодировании и защите информации. Сформулируем ряд свойств таких последовательностей в виде задач.

Задача 1.3. Пусть $\omega = (\delta_1, \dots, \delta_n)$ – абсолютно случайная последовательность длины n, и $1 \le m_1 < m_2 < \dots < m_k \le n$ – некоторая последовательность длины k. Тогда новая подпоследовательность $\omega' = (\delta_{m_1}, \delta_{m_2}, \dots, \delta_{m_k})$ – также абсолютно случайная последовательность, но длины k.

Поясним утверждение на примере. Пусть $n = 5, \quad m_1 = 1 \quad m_2 = 2$ и $\omega' = (0, 0)$. Тогда $\mathbf{P}(\omega') = \frac{1}{2^2}$. Действительно, ω' может быть получена из любой из 8 последовательностей:

$$\begin{array}{cccc} (00000) & (00001) & (00010) & (00011) \\ (00100) & (00101) & (00110) & (00111). \end{array}$$

Вероятность события A, для которого эти исходы являются благоприятствующими, равна

$$\mathbf{P}(A) = \frac{2^3}{2^5} = \frac{1}{2^2}.$$

Для реализации абсолютно случайных последовательностей используют либо физические датчики, либо последовательности так называемых псевдослучайных чисел, полученных по определенному вычислительному алгоритму и потому не являющихся случайными в прямом смысле, но обладающих схожими свойствами.[5] Существуют таблицы случайных чисел, самая значительная из них "Миллион случайных чисел".[6]

Заметим, что если бросать монету и отмечать выпадение герба цифрой 1, а решки – цифрой 0, то получится последовательность нулей и

[4] В дальнейшем такие последовательности из 0 и 1 будем называть *абсолютно случайными* последовательностями.

[5] См. подробнее [3].

[6] A million random digits with 100000 normal deviates, Rand Corporation, Glencoe, Illinois, 1955.

единиц, но обычно полученные таким образом результаты не удовлетворяют критериям случайности. Это связано либо с тем, что монета не является идеально симметричной, либо с особенностями бросающего монету.

Задача 1.4. Пусть новая последовательность ω' длины $n-1$ получена из ω следующим образом: смены 0 и 1 соответствуют в новой последовательности 1, а отсутствие перемены – 0. Скажем, если $\omega = (01101)$, то $\omega' = (1011)$. Тогда ω' также абсолютно случайная последовательность длины $n-1$.

Задача 1.5. Рассмотрим фиксированную последовательность $\varepsilon_0 = (0,0,1,0,1)$ и случайную последовательность $\varepsilon_1 = (\delta_1, \ldots, \delta_5)$. Сложив их почленно по $mod\,2$, получим последовательность $\varepsilon_2 = ({\delta'}_1, \ldots, {\delta'}_5)$. Тогда эта последовательность также абсолютно случайна.

Задача 1.6. Рассмотрим буквы русского алфавита, отождествив предварительно е с ё. Их окажется 32. Закодируем их, ставя в соответствие каждой букве её порядковый номер минус 1, записанный в двоичной системе и дополненный нулями, чтобы получилась цепочка длины 5: А(00000), Б(00001), В(00010),...,Я(11111). Если взять произвольный текст, то из страницы текста получится некоторая последовательность ε_0. Сложим теперь ее по $mod2$ со случайной последовательностью такой же длины ε_1. Тогда в силу утверждения предыдущей задачи полученная последовательность ε_2 будет абсолютно случайной. Расшифровать основной текст можно, сложив $\varepsilon_2 + \varepsilon_1 = \varepsilon_0 (mod\,2)$.

Это один из способов кодирования секретной информации.[7]

[7]Использовался в горячей линии Москва – Вашингтон.

Глава 2

Случайные величины и случайные события

В предыдущей главе мы познакомились с примерами, в которых вероятности вычислялись прямым подсчетом. Для дальнейшего развития теории нам понадобятся новые понятия, новые обозначения.

В теории вероятностей принята своя терминология, отличная от используемой, скажем, в математическом анализе. В учебнике А. Н. Ширяева "Вероятность"([10]стр.149-150, второе издание) можно найти таблицу соответствия некоторых терминов.

2.1 Случайные величины

Рассмотрим конечное вероятностное пространство $(\Omega, \mathcal{A}, \mathbf{P})$. Напомним, что

Ω — это конечное множество, состоящее из s элементарных исходов $\omega_1, \dots, \omega_s$,

$\mathcal{A}$ — класс всех подмножеств, называемых *случайными событиями,*

$\mathbf{P}$ — вероятность или распределение вероятностей на классе $\mathcal{A}$.

Определение 2.1. Случайной величиной, заданной на конечном вероятностном пространстве, называется любая функция, зависящая от элементарного исхода ω и принимающая числовые значения.

Обозначать случайные величины будем заглавными буквами конца латинского алфавита $X = X(\omega), Y = Y(\omega), Z = Z(\omega), \dots$ или греческими буквами $\xi = \xi(\omega), \eta = \eta(\omega), \zeta = \zeta(\omega), \dots$ Аргумент ω часто опускают.

Замечание. В более сложных случаях, когда пространство ω состоит из бесконечного числа элементарных исходов, все значительно усложняется. Общее определение вероятностного пространства и случайной величины появится позже (см. главу 9).

Определение 2.2. Математическим ожиданием случайной величины $X(\omega)$, определенной на конечном вероятностном пространстве, называется число

$$\mathbf{E}\,X = \sum_{\omega \in A} X(\omega) \cdot p(\omega).$$

Обозначение $\mathbf{E}\,X$ происходит от первой буквы английского слова expectation – ожидание. В отечественной литературе для обозначения математического ожидания также широко используется $\mathbf{M}X$.

Определение 2.3. Индикатором случайного события A называется случайная величина

$$I_A(\omega) = \begin{cases} 1, & \text{если} \quad \omega \in A \\ 0, & \text{если} \quad \omega \notin A. \end{cases}$$

Для индикатора события A

$$\mathbf{E}\,I_A(\omega) = \sum_{\omega} I_A(\omega) p(\omega) = \sum_{\omega} 1 \cdot p(\omega) = \mathbf{P}(A).$$

Таким образом **математическим ожиданием индикатора случайного событ ия является его вероятность.**

Случайные события обычно обозначают первыми заглавными буквами латинского алфавита: $A, B, C, \ldots$. Наряду с отдельными случайными событиями мы будем рассматривать классы (множества) случайных событий, для обозначения которых будем использовать рукописные заглавные буквы: $\mathcal{A}, \mathcal{B}, \mathcal{C} \ldots$

2.2 Операции над случайными событиями

1. **Включение событий:**

$A \subset B\,(A \subseteq B)$, если из условия $\omega \in A$ следует, что $\omega \in B$. В теории вероятностей в этом случае говорят, что событие A влечет наступление события B. Если $A \subset B$, а $B \subset A$, то $A = B$.

2. **Объединение событий:**

$A \cup B = \{\omega : \quad \omega \in A \quad \text{или} \quad \omega \in B\}$ – это событие, состоящее из исходов, которые принадлежат или A, или B, или одновременно двум этим событиям.

3. **Пересечение событий:**

$A \cap B = \{\omega : \quad \omega \in A \text{ и } \omega \in B\}$ – это событие, состоящее из тех элементарных исходов, которые одновременно принадлежат A и B. Эту операцию также обозначают как $A \cdot B = AB$.

4. **Разность событий:**

$A \setminus B = \{\omega : \quad \omega \in A, \text{ но } \omega \notin B\}$ – событие, состоящее из элементарных исходов, принадлежащих A, но не принадлежащих B.

5. **Отрицание:**

$\overline{A} = \Omega \setminus A$ – это событие состоит из всех элементарных исходов, которые не принадлежат A, $\overline{A}$ также называют дополнением к A или событием, противоположным A.

5. **Симметрическая разность:**

$A \triangle B = (A \setminus B) \cup (B \setminus A)$ – событие, состоящее из исходов, которые содержатся или в A, или в B, но не принадлежат обоим событиям одновременно.

События называются *несовместимыми,* если у них нет общих исходов, т.е. $A \cap B = \varnothing$.

Для несовместимых событий симметрическая разность этих событий совпадает с их объединением.

Задача 2.1. Докажите следующие свойства операций над случайными событиями:

1) коммутативность:

$$A \cap B = B \cap A, \quad A \cup B = B \cup A;$$

2) ассоциативность:

$$A \cup (B \cup C) = (A \cup B) \cup C = A \cup B \cup C,$$

$$A \cap (B \cap C) = (A \cap B) \cap C = A \cap B \cap C;$$

3) дистрибутивность:

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

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

4) принцип двойственности:

$$\overline{A \cup B} = \overline{A} \cap \overline{B}, \qquad \overline{A \cap B} = \overline{A} \cup \overline{B}.$$

2.3 Операции над индикаторами

Операциям над случайными событиями соответствуют операции над индикаторами этих событий.

Задача 2.2. Докажите, что
1) $I_{A\cap B}(\omega) = I_A(\omega)\cdot I_B(\omega)$,
2) $I_{\overline{A}}(\omega) = 1 - I_A(\omega)$,
3) $I_{A\cup B}(\omega) = I_A(\omega) + I_B(\omega) - I_{A\cap B}(\omega)$,
4) $I_{A\triangle B}(\omega) = |I_A(\omega) - I_B(\omega)| = I_A(\omega) + I_B(\omega)(mod\, 2)$.

Приведем для примера доказательство свойства 3 индикаторов случайных событий:

$$I_{A\cup B}(\omega) = 1 - I_{\overline{A\cup B}}(\omega) = 1 - I_{\overline{A}\cap\overline{B}}(\omega) =$$

$$= 1 - I_{\overline{A}}(\omega)\cdot I_{\overline{B}}(\omega) = 1-(1-I_A(\omega))\cdot(1-I_B(\omega)) = I_A(\omega)+I_B(\omega)-I_{A\cap B}(\omega).$$

Задача 2.3. Докажите обобщение этого утверждения для n событий, а именно

$$I_{\bigcup\limits_{i=1}^{n} A_i}(\omega) = \sum_{i=1}^{n} I_{A_i}(\omega) - \sum_{1\le i<j\le n} I_{A_iA_j}(\omega) + \dots + (-1)^{n+1} I_{A_1A_2\dots A_n}(\omega).$$

Задача 2.4. Докажите свойства симметрической разности:

$$1) A\triangle(B\triangle C) = (A\triangle B)\triangle C,$$

$$2) A\cup B = (A\triangle B)\triangle(A\cap B).$$

Пусть $A_1,\dots,A_n$ – произвольные n случайных событий.

Определение 2.4. Событие F определяется по событиям $A_1,\dots,A_n$, если индикатор события F есть функция от индикаторов этих событий, т.е.

$$I_F = H(I_{A_1}(\omega),\dots,I_{A_n}(\omega)),$$

где $H(\cdot)$ – функция, определенная в вершинах единичного $n-$ мерного куба.

В этом случае, зная, какие из событий $A_1,\dots,A_n$ наступили и какие не наступили, можно однозначно судить о том, наступило или нет событие F.

Глава 3

Свойства вероятности и математического ожидания

3.1 Свойства вероятности

Из определения вероятности $\mathbf{P}(A), \quad A \in \mathcal{A}$ для конечных вероятностных пространств следует, что вероятность обладает следующими свойствами

P1. $\mathbf{P}(A) \geq 0$;

P2. $\mathbf{P}(\Omega) = 1, \quad \mathbf{P}(\varnothing) = 0$;

P3. Для попарно несовместимых случайных событий $A_1, \ldots, A_n$

$$\mathbf{P}(\bigcup_{i=1}^{n} A_i) = \sum_{i=1}^{n} \mathbf{P}(A_i).$$

Последнее свойство называют свойством *конечной аддитивности* вероятности. Напомним, что попарная несовместимость событий означает, что $A_i \cap A_j = \varnothing$, если $i \neq j$.

Первые два свойства вероятности очевидны. Докажем свойство P3.

Доказательство проведем по индукции по числу событий n.

Пусть $n = 2$. Тогда

$$\mathbf{P}(A_1 \cup A_2) = \sum_{\omega \in A_1 \cup A_2} p(\omega) = \sum_{\omega \in A_1} p(\omega) + \sum_{\omega \in A_2} p(\omega) = \mathbf{P}(A_1) + \mathbf{P}(A_2).$$

Заметим, что возможность изменения порядка суммирования в данном случае сомнений не вызывает, поскольку все суммы содержат конечное число слагаемых.

Пусть свойство выполняется для $n-1$ события $A_1, \dots, A_{n-1}$, тогда

$$\mathbf{P}\left(\bigcup_{i=1}^{n} A_i\right) = \mathbf{P}\left\{\left(\bigcup_{i=1}^{n-1} A_i\right) \bigcup A_n\right\} = \mathbf{P}\left(\bigcup_{i=1}^{n-1} A_i\right) + \mathbf{P}(A_n) = \sum_{i=1}^{n} \mathbf{P}(A_i),$$

что и доказывает свойство P3.

Все остальные свойства, приведенные ниже, могут быть получены из свойств P1, P2, P3.[1]

4. Если $A \subset B$, то $\mathbf{P}(A) \le \mathbf{P}(B)$.

Доказательство. Поскольку событие $B = A \cup (B \setminus A)$ можно представить как объединение несовместимых событий, то

$$\mathbf{P}(B) = \mathbf{P}(A) + \mathbf{P}(B \setminus A) \ge \mathbf{P}(A).$$

□

5. Для любого события A $\quad 0 \le \mathbf{P}(A) \le 1$.

6. $\mathbf{P}(\overline{A}) = 1 - \mathbf{P}(A)$.

Задача 3.1. Докажите свойства вероятности 4, 5.

7. Если $A_1, \dots, A_n$ – произвольные случайные события, то

$$\mathbf{P}(\bigcup_{i=1}^{n} A_i) \le \sum_{i=1}^{n} \mathbf{P}(A_i). \tag{3.1}$$

Это свойство называется свойством *полуаддитивности* вероятности.

Доказательство. Введем события

$$\begin{aligned}
&B_1 = A_1,\\
&B_2 = A_2 \setminus A_1 = A_2 \cap \overline{A_1},\\
&B_3 = A_3 \setminus (A_1 \cup A_2) = A_3 \cap \overline{A_1} \cap \overline{A_2}\\
&\vdots\\
&B_n = A_n \setminus (A_1 \cup \dots \cup A_{n-1}) = A_n \cap \overline{A_1} \cap \dots \cap \overline{A_{n-1}}.
\end{aligned}$$

1) Заметим, что события $B_1, \dots, B_n$ попарно несовместимы. Действительно, пусть $i < j$, тогда по построению $B_i \subset A_i$, $\quad B_j \subset \overline{A_i}$. но это означает, что $B_i \cap B_j = \varnothing$.

[1] Это замечание выполняется также для общей теории, где свойства P1–P3 будут служить определением вероятности (с заменой конечной аддитивности на счетную).

2) Рассмотрим события

$$F = \bigcup_{i=1}^{n} B_i, \qquad C = \bigcup_{i=1}^{n} A_i.$$

Покажем,что $F = C$.

Пусть $\omega \in C$, тогда существует такое i, что $\omega \in A_i$. Возьмем наименьшее такое i. Обозначим его i_0. Тогда $\omega \in A_{i_0}$, но $\omega \notin A_1, \ldots, A_{i_0-1}$, а значит, $\omega \in B_{i_0}$, $\omega \in \bigcup\limits_{i=1}^{n} B_i = F$. Следовательно, $C \subset F$.

Докажем теперь обратное включение. Пусть $\omega' \in F$. Тогда найдется такое j_0, что $\omega' \in B_{j_0} \subset A_{j_0}$. Но это означает, что $\omega' \in \bigcup\limits_{j=1}^{n} A_j = C$. Следовательно, $F \subset C$, и стало быть, $F = C$.

Из доказанного вытекает, что

$$\mathbf{P}(\bigcup_{i=1}^{n} A_i) = \mathbf{P}(\bigcup_{j=1}^{n} B_j) = \sum_{j=1}^{n} \mathbf{P}(B_j) \le \sum_{j=1}^{n} \mathbf{P}(A_j).$$

$\square$

Замечание. В общей теории сходное доказательство будет применено и к бесконечным последовательностям случайных событий $A_1, A_2 \ldots$. В этом случае полуаддитивность вероятности можно сформулировать следующим образом.

7. Пусть $A_1, A_2 \ldots$ — произвольная последовательность случайных событий. Тогда

$$\mathbf{P}(\bigcup_{i=1}^{\infty} A_i) \le \sum_{i=1}^{\infty} \mathbf{P}(A_i).$$

Задача 3.2. Докажите для произвольных событий $A_1, \ldots, A_n$, что

$$\mathbf{P}(A_1 \cup \cdots \cup A_n) = \sum_{i=1}^{n} \mathbf{P}(A_i) - \sum_{1 \le i < j \le n} \mathbf{P}(A_i \cap A_j) +$$

$$+ \sum_{1 \le i < j < k \le n} \mathbf{P}(A_i \cap A_j \cap A_k) - \cdots + (-1)^{n+1}\, \mathbf{P}(A_1 \cap \cdots \cap A_n).$$

3.2 Свойства математического ожидания случайных величин

В рассматриваемой схеме конечных вероятностных пространств случайной величиной называется произвольная числовая функция $X = X(\omega)$, зависящая от элементарного исхода. Математическое ожидание мы определили как

$$EX = \sum_{\omega} X(\omega)p(\omega). \tag{3.2}$$

Сформулируем основные свойства математического ожидания.

1. $\mathbf{E}\, I_A = \mathbf{P}(A)$.
2. $\mathbf{E}(X+Y) = \mathbf{E}\, X + \mathbf{E}\, Y$.
3. $\mathbf{E}(c \cdot X) = c \cdot \mathbf{E}\, X$, где c – произвольная константа.
4. Если $X(\omega) \geq 0$ для всех ω, то $\mathbf{E}\, X \geq 0$.

Проведем для примера доказательство свойства 2.

Доказательство. Согласно определению

$$EX = \sum_{\omega} X(\omega)p(\omega), \quad \mathbf{E}\, Y = \sum_{\omega} Y(\omega)p(\omega). \tag{3.3}$$

Складывая математические ожидания в (3.3), получим

$$\mathbf{E}\, X + \mathbf{E}\, Y = \sum_{\omega} (X(\omega) + Y(\omega))p(\omega) = \mathbf{E}(X+Y).$$

□

Замечание. Математическое ожидание можно рассматривать как *линейный функционал,* определенный на множестве случайных величин, заданных на одном вероятностном пространстве.[2] Линейный функционал, для которого выполняется свойство 4, называют неотрицательным. В качестве линейного неотрицательного функционала, можно привести следующий пример.

Пусть $X(\omega)$ — непрерывная функция на $[0, 1]$. Тогда $\int\limits_0^1 X(\omega)d\omega$ обладает всеми перечисленными свойствами математического ожидания.

В действительности связь между понятиями математического ожидания и интеграла более глубокая. Общее понятие математического ожидания в теории вероятностей соответствует понятию интеграла Лебега в

[2]В общей теории к этим свойствам будет добавлено некоторое свойство "непрерывности"математического ожидания.

функциональном анализе. Но пока мы ограничиваемся простым случаем конечного вероятностного пространства.

Как правило, в теории вероятностей важно знать не зависимость случайной величины $X(\omega)$ от $\omega \in \Omega$, а важно знать ее *распределение.* Для случайных величин с конечным числом различных между собой значений (а мы имеем дело пока именно с такими случайными величинами), распределением называют набор значений $x_1, \ldots, x_n$ и соответствующих этим значениям вероятностей $p_1, \ldots, p_n$, где

$$p_i = \mathbf{P}\{X(\omega) = x_i\}, \quad i = 1, ..., n.$$

Иногда мы будем задавать распределение случайной величины X таблицей

x_1	x_2	$\ldots$	x_n
p_1	p_2	$\ldots$	p_n.

Обратите внимание на то, что в этой таблице указываются все различные между собой значения случайной величины X, а сумма вероятностей всегда равна 1.

Приведем несколько утверждений, позволяющих вычислять $\mathbf{E}\,X$, используя распределение случайной величины X.

Определение 3.1. Класс $\mathcal{B} = \{B_1, \ldots, B_q\}$ называется конечным разбиением Ω, если при всех $i \neq j \quad B_i \cap B_j = \emptyset$ и $\Omega = \bigcup\limits_{i=1}^{q} B_i$.

Пусть есть два разбиения:

$$\mathcal{B} = \{B_1, \ldots, B_q\},$$

$$\mathcal{C} = \{C_1, \ldots, C_r\}.$$

Составим всевозможные пересечения $B_i \cap C_j, \quad 1 \le i \le q, \quad 1 \le j \le r$. Таких пересечений $q\dot{r}$. Получили новое разбиение

$$\mathcal{A} = \{B_i \cap C_j\},$$

которое называют пересечением разбиений и обозначают

$$\mathcal{A} = \mathcal{B} \times \mathcal{C}.$$

Лемма 3.1. Пусть $X(\omega)$ – случайная величина, заданная на конечном вероятностном пространстве, $\mathcal{B} = \{B_1, \ldots, B_q\}$ – некоторое разбиение Ω, причем $X(\omega)$ постоянна на каждом из B_j и $X(\omega) = x'_j$, если $\omega \in B_j$. Тогда

$$\mathbf{E}\,X = \sum_{j=1}^{q} x'_j\,\mathbf{P}(B_j).$$

Доказательство. Поскольку случайная величина $X(\omega)$ постоянна на каждом B_j разбиения , то ее можно представить в виде

$$X(\omega) = \sum_{j=1}^{q} x'_j \cdot I_{B_j}(\omega).$$

Возьмем математические ожидания от обеих частей этого равенства и воспользуемся свойствами математического ожидания. Получим

$$\mathbf{E}\, X(\omega) = \sum_{j=1}^{q} x'_j \cdot \mathbf{E}\, I_{B_j}(\omega) = \sum_{j=1}^{q} x'_j \cdot \mathbf{P}(B_j).$$

$\square$

С каждой случайной величиной можно связать естественное разбиение. Пусть распределение $X(\omega)$ задается таблицей

$$\begin{array}{c|c|c|c} x_1 & x_2 & \dots & x_n \\ \hline p_1 & p_2 & \dots & p_n. \end{array} \tag{3.4}$$

Обозначим $A_j = \{\omega : X(\omega) = x_j\}, \quad p_j = \mathbf{P}(A_j), \quad j = 1, \dots, n$. События A_j образуют разбиение множества элементарных исходов Ω и являются аналогом линий уровня в математическом анализе.

Следствие 3.2.1. *1.* $\mathbf{E}\, X = \sum\limits_{j=1}^{n} x_j p_j$.

2. Если случайная величина $Y = h(X)$, *то* $\mathbf{E}\, Y = \sum\limits_{j=1}^{n} h(x_j) \cdot p_j$.

Следовательно, для вычисления математического ожидания случайной величины совсем необязательно знать зависимость $X(\omega)$, а достаточно знать ее распределение.

3.3 Вероятность появления хотя бы одного из n событий

Пусть $A_1, \dots, A_n$ – произвольные случайные события. Обозначим $N(\omega)$ – случайную величину, равную числу наступающих событий среди $A_1, \dots, A_n$. Для $0 \le k \le n$ событие $\{N(\omega) = k\}$ эквивалентно тому, что найдутся $1 \le i_1 < i_2 < \dots < i_k \le n$ такие, что

$$\omega \in A_{i_1}, A_{i_2} \cdots A_{i_k}, \text{ но } \omega \notin A_j, \text{ если } j \ne i_1, i_2, \dots i_k.$$

Задача 3.3. Для любого n вычислите вероятность $\mathbf{P}\{\omega : N(\omega) = k\}$.

Событие $C = \bigcup\limits_{i=1}^{n} A_i$ означает, что наступит хотя бы одно из событий $A_1, \ldots, A_n$, т.е.

$$\bigcup_{i=1}^{n} A_i = \{\omega : N(\omega) \geq 1\}.$$

Для вывода формулы для вероятности события C воспользуемся свойствами индикаторов случайных событий и свойствами математических ожиданий. Введем

$$\begin{aligned}
&S_1 = \sum_{i=1}^{n} \mathbf{P}(A_i),\\
&S_2 = \sum_{1 \leq i < j \leq n} \mathbf{P}(A_i A_j),\\
&S_3 = \sum_{1 \leq i < j < k \leq n} \mathbf{P}(A_i A_j A_k),\\
&\vdots\\
&S_n = \mathbf{P}(A_1 A_2 \cdots A_n).
\end{aligned}$$

Вероятности событий, связанных со случайной величиной $N(\omega)$, выражаются через величины $S_1, \ldots, S_n$. Для индикатора события C имеем

$$I_C(\omega) = 1 - I_{\overline{C}}(\omega) = 1 - I_{\overline{A_1} \cap \cdots \cap \overline{A_n}}(\omega) = 1 - \prod_{j=1}^{n} (1 - I_{A_j}(\omega)) = 1 -$$

$$- \left(1 - \sum_{j=1}^{n} I_{A_j}(\omega) + \sum_{1 \leq i < j \leq n} I_{A_i}(\omega) \cdot I_{A_j}(\omega) - \cdots + (-1)^n I_{A_1}(\omega) \cdot \ldots \cdot I_{A_n}(\omega) \right).$$

Возьмем математическое ожидание от обеих частей этого равенства. Из свойств аддитивности математического ожидания следует

$$\mathbf{P}(C) = S_1 - S_2 + S_3 - \ldots + (-1)^{n+1} S_n. \tag{3.5}$$

Мы получили формулу для вероятности объединения n случайных событий.

Задача 3.4. Докажите, что обрывая сумму в (3.5) на каком-либо члене, получим неравенства сверху и снизу:

$$S_1 - S_2 \leq \mathbf{P}(C) \leq S_1,$$

$$S_1 - S_2 + S_3 - S_4 \leq \mathbf{P}(C) \leq S_1 - S_2 + S_3$$

и т.д.

Замечание. Подробнее о вычислении вероятностей событий $\{N(\omega) = k\}$ можно прочитать в учебнике В. Феллера, том 1.[3]

Вернемся к примеру 3 лекции 1, в котором рассматривались случайные перестановки. Всего существует $n!$ перестановок n различных чисел. Припишем каждой из них вероятность $\frac{1}{n!}$. Обозначим A_k – событие, состоящее в том, что число k в перестановке встречается на k-ом месте, в этом случае будем говорить, что k – неподвижная точка.

Используя полученные формулы, вычислим вероятность того, что есть хотя бы одна неподвижная точка.

Поскольку

$$\mathbf{P}(A_i) = \frac{(n-1)!}{n!}, \quad 1 \leq i \leq n,$$

$$\mathbf{P}(A_iA_j) = \frac{(n-2)!}{n!}, \quad i \neq j, \dots$$

то

$$S_1 = n{\cdot}\mathbf{P}(A_i) = 1, \quad S_2 = C_n^2{\cdot}\mathbf{P}(A_iA_j) = \frac{1}{2!}, \quad S_3 = C_n^3{\cdot}\mathbf{P}(A_iA_jA_k) = \frac{1}{3!}, \dots$$

Таким образом

$$\mathbf{P}(\bigcup_{i=1}^{n} A_i) = 1 - \frac{1}{2!} + \frac{1}{3!} - \ldots + (-1)^{n+1} \cdot \frac{1}{n!}. \tag{3.6}$$

Сравнивая это выражение с разложением для

$$e^{-1} = 1 - 1 + \frac{1}{2!} - \frac{1}{3!} + \cdots ,$$

получим, что выражение (3.6) – это сумма первых n слагаемых в разложении для $1 - e^{-1}$. Поскольку ряд сходится быстро, то получим хорошее приближение

$$\mathbf{P}(\bigcup_{i=1}^{n} A_i) \approx 1 - e^{-1} \approx 1 - 0.36788 = 0.63212.$$

Уже при $n = 7$ точное значение вероятности совпадает с приведенным приближенным значением в 5 знаках после запятой.

В учебнике В.Феллера приводится способ вычисления вероятностей того, что имеет место в точности k совпадений и показано, что

$$\mathbf{P}\{N(\omega) = k\} \approx \frac{e^{-1}}{k!}.$$

Так для $n = 10$ результаты приведены в таблице

[3] Эта книга выдержала несколько изданий, в том числе и на русском языке. Первый том посвящен дискретным вероятностным моделям, а второй – непрерывным.

$N = k$	$\mathbf{P}\{N = k\}$
0	0.36788
1	0.36788
2	0.13394
3	0.06131
4	0.01534
5	0.00306
6	0.00052
7	0.00007
8	0.00001
9	0.00000

Эти результаты можно использовать в задачах угадывания: пусть имеется несколько наборов из 10 карт (Т,К,Д,В,10,9,8,7,6,5). Тогда вероятности угадать определенное число карт вычисляются по приведенным формулам и указаны в таблице.

Глава 4

Независимость случайных событий и случайных величин

4.1 Условная вероятность. Независимость двух случайных событий

Понятие независимости встречается во многих разделах математики: это линейная независимость векторов в линейной алгебре, функциональная независимость систем функций в математическом анализе и т.д. Чтобы подчеркнуть различие понятия независимости в теории вероятностей ее называют статистической или стохастической независимостью. Но поскольку в нашем курсе будет рассматриваться только такая независимость, то в дальнейшем мы будем говорить просто о *независимости* случайных величин или о независимости случайных событий.

С понятием независимости в теории вероятностей тесно связано понятие *условной вероятности.* Приведем для начала некоторые наглядные соображения, поясняющие эти понятия.

Пусть n раз подбрасывают 2 игральных кости, событие A происходит, если на первой кости выпала 1, событие B, если на второй выпала 5. Результаты опыта будем отмечать в таблице, ставя знак $+$, если в соответствующем опыте наблюдалось наступление события, и знак $-$, если это не так. К примеру, после 6 проведенных испытаний мы могли получить следующую таблицу

номер испытания	1	2	3	4	5	6
A	$+$	$+$	$+$	$-$	$-$	$+$
B	$+$	$-$	$-$	$+$	$+$	$-$

Пусть n –это число проведенных испытаний (в примере $n = 6$),

n_A – число испытаний, в которых наблюдалось событие $A(n_A = 4)$, n_B – число испытаний, в которых произошло событие $B(n_B = 3)$, n_{AB} – число испытаний, в которых наступили оба события ($n_{AB} = 1$).

Тогда $\frac{n_{AB}}{n_A}$ - относительная частота наступления события B в ряду тех испытаний, в которых произошло событие A. Поскольку при большом числе испытаний относительная частота наступления события близка к вероятности этого события, то мы можем записать

$$\frac{n_{AB}}{n_A} = \frac{\frac{n_{AB}}{n}}{\frac{n_A}{n}} \approx \frac{\mathbf{P}(AB)}{\mathbf{P}(A)}.$$

Определение 4.1. Отношение $\frac{\mathbf{P}(AB)}{\mathbf{P}(A)}$ называют *условной вероятностью* события B при условия, что наступило событие A, и обозначают $\mathbf{P}(B|A)$.

В случае, когда $\mathbf{P}(B) = \mathbf{P}(B|A)$, естественно говорить, что событие B не зависит от события A. Но в этом случае $\mathbf{P}(B) \cdot \mathbf{P}(A) = \mathbf{P}(AB)$, откуда следует, что $\mathbf{P}(A) = \mathbf{P}(A|B)$, если только $\mathbf{P}(B) \neq 0$. То есть в том случае, когда A не зависит от B, событие B также не зависит от A. Поэтому естественнее дать определение независимости двух событий в симметричной форме, распространяющееся и на события с нулевой вероятностью.

Определение 4.2. События A и B называются *независимыми,* если

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

Задача 4.1. Проведем опыт, состоящий в подбрасывании двух костей. Покажите, что события A и B, введенные в рассмотренном выше примере, независимы.

В математической статистике мы рассмотрим более содержательные примеры: независимость пола ребенка от различия в возрасте родителей, независимость пола второго ребенка от пола старшего ребенка в двухдетной семье, независимость длительности продолжения телефонного разговора от того, сколько он уже длится и т.д.

4.2 Независимость случайных величин. Взаимная независимость нескольких случайных событий

Рассмотрим несколько случайных величин $X_1(\omega), X_2(\omega), \ldots, X_N(\omega)$ с возможными значениями x_{i,j_i}, где первый индекс соответствует номеру случайной величины, а второй номеру значения ($1 \leq j_i \leq m_i$).

Определение 4.3. Случайные величины $X_1, \ldots, X_N$ называются *взаимно независимыми* (или просто *независимыми*), если выполнены $m_1 \times m_2 \times \ldots \times m_N$ соотношений:

$$\mathbf{P}\{X_1(\omega) = x_{1,j_1}, \ldots, X_N(\omega) = x_{N,j_N}\} = \prod_{k=1}^{N} \mathbf{P}\{X_k(\omega) = x_{k,j_k}\},$$

где $1 \le j_k \le m_k, \quad k = 1, \cdots, N$.

Замечание. Отметим, что существуют различные критерии независимости случайных величин, с которыми познакомимся в дальнейшем. В практических задачах обычно на основании каких-либо физических (или других) соображений предполагается независимость некоторых величин, а из этого выводится независимость (или зависимость) других случайных величин.

Понятие *взаимной независимости* случайных событий можно ввести, исходя из определения независимости случайных величин.

Определение 4.4. События $A_1, \ldots, A_N$ *взаимно независимы,* если взаимно независимы индикаторы этих событий $I_{A_1}(\omega), \ldots, I_{A_N}(\omega)$.

Это означает, что

$$\mathbf{P}\{\omega : I_{A_1}(\omega) = \delta_1, \ldots, I_{A_N}(\omega) = \delta_N\} = \prod_{k=1}^{N} \mathbf{P}\{\omega : I_{A_k}(\omega) = \delta_k\},$$

где

$$\delta_k = \begin{cases} 1, \\ 0, \end{cases} \quad i = 1, \cdots, N.$$

Всего требуется выполнения 2^N таких соотношений.

Поясним определение взаимной независимости случайных событий для частных случаев $N = 2, 3$.

Пусть имеется два события A_1, A_2. Тогда необходимо выполнение 2^2 соотношений:

$$\mathbf{P}(A_1 A_2) = \mathbf{P}(A_1) \cdot \mathbf{P}(A_2), \quad \mathbf{P}(\overline{A_1} A_2) = \mathbf{P}(\overline{A_1}) \cdot \mathbf{P}(A_2),$$

$$\mathbf{P}(A_1 \overline{A_2}) = \mathbf{P}(A_1) \cdot \mathbf{P}(\overline{A_2}), \quad \mathbf{P}(\overline{A_1} \overline{A_2}) = \mathbf{P}(\overline{A_1})\, \mathbf{P}(\overline{A_2}).$$

Задача 4.2. Покажите, что в этом случае все соотношения вытекают из какого-либо одного, например,из первого.

Таким образом взаимная независимость двух событий эквивалентна определенной выше независимости событий.

Аналогичным образом можно выписать $2^3 = 8$ соотношений, необходимых для взаимной независимости событий A_1, A_2, A_3 :

$$\begin{aligned} \mathbf{P}(A_1A_2A_3) &= \mathbf{P}(A_1)\,\mathbf{P}(A_2)\,\mathbf{P}(A_3), \\ &\vdots \\ \mathbf{P}(\overline{A_1}\;\overline{A_2}\;\overline{A_3}) &= \mathbf{P}(\overline{A_1})\,\mathbf{P}(\overline{A_2})\,\mathbf{P}(\overline{A_3}). \end{aligned} \tag{4.1}$$

Задача 4.3. Докажите, что в этом случае выполнение условий 4.1 взаимной независимости трех событий эквивалентно выполнению следующих равенств:

$$\begin{aligned} \mathbf{P}(A_1A_2) &= \mathbf{P}(A_1)\,\mathbf{P}(A_2), \\ \mathbf{P}(A_1A_3) &= \mathbf{P}(A_1)\,\mathbf{P}(A_3), \\ \mathbf{P}(A_2A_3) &= \mathbf{P}(A_2)\,\mathbf{P}(A_3), \\ \mathbf{P}(A_1A_2A_3) &= \mathbf{P}(A_1)\,\mathbf{P}(A_2)\,\mathbf{P}(A_3). \end{aligned} \tag{4.2}$$

Задача 4.4. Покажите, что в общем случае для взаимной независимости N случайных событий необходимо и достаточно выполнение $2^N - N - 1$ условий: для любого $2 \le k \le N$, для любых наборов различных между собой индексов $i_1 < i_2 < \ldots < i_k$

$$\mathbf{P}(A_{i_1} \ldots A_{i_k}) = \mathbf{P}(A_{i_1}) \cdot \ldots \cdot \mathbf{P}(A_{i_k}). \tag{4.3}$$

Определение 4.5. События $A_1, \ldots, A_N$ называются *попарно независимыми,* если в (4.3) выполнены соотношения для $k = 2$.

В частности, три события A_1, A_2, A_3 называются попарно независимыми, если в (4.2) выполнены первые три равенства.

Задача 4.5. Пусть события A, B, C, D– взаимно независимы,
$E = A \cup B, \quad F = C \cup D$. Докажите, что события E и F независимы.

4.3 Свойства независимых случайных величин и взаимно независимых случайных событий

Сформулируем сначала общие свойства независимых случайных величин.

Теорема 4.1. *Пусть случайные величины $X_1, \ldots, X_N$ – независимы. Тогда для любых наборов $1 < j_1 < j_2 < \ldots j_k \le N$ случайные величины $X_{j_1}, X_{j_2}, \ldots, X_{j_k}$ – независимы.*

Доказательство. Проведем сначала доказательство для трех случайных величин. Чтобы показать, как это делается, и не загромождать записи индексами, рассмотрим независимые случайные величины U, V, W соответственно со значениями

$$\begin{aligned} U:&\;\; u_1, \ldots, u_l, \\ V:&\;\; v_1, \ldots, v_m, \\ W:&\;\; w_1, \ldots, w_n. \end{aligned}$$

Покажем, что из независимости трех случайных величин следует, что любые две, например, U и V, также независимы. Для любых значений этих случайных величин из определения независимости случайных величин следует

$$\mathbf{P}\{U = u_i, V = v_j, W = w_k\} = \mathbf{P}\{U = u_i\}\,\mathbf{P}\{V = v_j\}\,\mathbf{P}\{W = w_k\}.$$

Просуммируем эти равенства по $1 \le k \le n$, получим

$$\sum_{k=1}^{n} \mathbf{P}\{U = u_i, V = v_j, W = w_k\} = \mathbf{P}\{U = u_i\}\,\mathbf{P}\{V = v_j\} \sum_{k=1}^{n} \mathbf{P}\{W = w_k\},$$

откуда получаем

$$\mathbf{P}\{U = u_i, V = v_j\} = \mathbf{P}\{U = u_i\}\,\mathbf{P}\{V = v_j\}.$$

Аналогичными рассуждениями в общем случае можно показать, что если независимы любые N случайных величин, то независимы и произвольные $N - 1$ из этого набора случайных величин.

$\square$

Теорема 4.2. *(*Без доказательства.*) Пусть случайные величины $X_1, \ldots, X_N$ – независимы и разбиты на непересекающиеся группы:*

$$\begin{array}{lll} X_{1,1}, & \ldots, & X_{1,n_1}, \\ X_{2,1}, & \cdots, & X_{2,n_2}, \\ \vdots & & \\ X_{k,1}, & \ldots, & X_{k,n_k}, \end{array}$$

$N = n_1 + n_2 + \cdots + n_k, \quad k \ge 2.$ *Определим новые случайные величины $Y_1 = g_1(X_{1,1}, \ldots, X_{1,n_1}), \ldots, Y_k = g_k(X_{k,1}, \ldots, X_{k,n_k})$. Тогда $Y_1, \ldots, Y_k$ также независимы.*

Эти теоремы вполне соответствуют нашим представлениям о независимости и аналогичны теоремам линейной алгебры (если $g_1, \ldots, g_k$-невырожденные линейные преобразования векторов).

Задача 4.6. Докажите утверждение теоремы 4.2 для случая $n_1 = n_2 = \dots = n_k = 1$.

Для случайных событий из этих теорем вытекают следующие утверждения.

Теорема 4.3. *Если события $A_1, \dots, A_N$ взаимно независимы, то и любые k из них также взаимно независимы.*

Для формулировки следующей теоремы напомним, что мы говорим, что событие A определяется по событиям $A_1, \dots, A_l$, если

$$I_A(\omega) = f(I_{A_1}(\omega), \dots, I_{A_l}(\omega)).$$

Определение 4.6. Событие A определяется по случайным величинам $X_1, \dots, X_l$, если

$$I_A(\omega) = g(X_1(\omega), \dots, X_l(\omega)),$$

т.е. если по известным значениям случайных величин можно установить, наступило событие A или нет.

Теорема 4.4. *Если взаимно независимые события $A_1, \dots, A_N$ разбиты на непересекающиеся группы событий*

$$\begin{array}{lll} A_{1,1}, & \cdots, & A_{1,n_1}, \\ \vdots & & \\ A_{k,1}, & \dots, & A_{k,n_k}, \end{array}$$

а события $B_1, \dots, B_k$ определяются каждое по своей группе событий, то $B_1, \dots, B_k$ также взаимно независимы.

4.4 Критерий независимости случайных величин

Теорема 4.5. *Для независимости случайных величин $X_1, \dots, X_n$ необходимо и достаточно, чтобы при любом выборе полуинтервалов $[a_1, b_1), \dots, [a_N, b_N)$ на числовой прямой*

$$\mathbf{P}\{\omega : a_1 \le X_1(\omega) < b_1, \dots, a_N \le X_N(\omega) < b_N\} = \prod_{k=1}^{N} \mathbf{P}\{\omega : a_k \le X_k < b_k\}. \tag{4.4}$$

Замечание. В общем случае выполнение условий (4.4) берется в качестве определения независимости случайных величин.

Доказательство. 1. Доказательство теоремы проведем для двух случайных величин U и V со значениями $u_1, \dots, u_l$ и $v_1, \dots, v_m$ соответственно. Докажем сначала достаточность условия (4.4). Пусть u_j, v_k – произвольные значения этих случайных величин. Выберем интервалы $[a, b), [c, d)$ таким образом, чтобы $u_j \in [a, b), v_k \in [c, d)$, но других значений случайных величин в них не попадало. Тогда

$$\mathbf{P}\{U(\omega) = u_j, V(\omega) = v_k\} = \mathbf{P}\{a \le U(\omega) < b, c \le V(\omega) < d\} =$$

$$= \mathbf{P}\{a \le U(\omega) < b\}\, \mathbf{P}\{c \le V(\omega) < d\} = \mathbf{P}\{U(\omega) = u_j\}\, \mathbf{P}\{V(\omega) = v_k\}.$$

2. Для доказательства необходимости (4.4) рассмотрим событие

$$C = \{a \le U(\omega) < b, c \le V(\omega) < d\} = \bigcup_{u_j \in [a,b), v_k \in [c,d)} \{U(\omega) = u_j, V(\omega) = v_k\}.$$

Поскольку объединяемые события попарно не пересекаются, то

$$\mathbf{P}(C) = \sum_{u_j \in [a,b), v_k \in [c,d)} \mathbf{P}\{U(\omega) = u_j, V(\omega) = v_k\}.$$

Воспользовавшись независимостью случайных величин, получаем

$$\mathbf{P}(C) = \sum_{u_j \in [a,b), v_k \in [c,d)} \mathbf{P}\{U(\omega) = u_j\}\, \mathbf{P}\{V(\omega) = v_k\} =$$

$$= \left(\sum_{u_j \in [a,b)} \mathbf{P}\{U(\omega) = u_j\} \right) \cdot \left(\sum_{v_k \in [c,d)} \mathbf{P}\{V(\omega) = v_k\} \right) =$$

$$= \mathbf{P}\{a \le U(\omega) < b\}\, \mathbf{P}\{c \le V(\omega) < d\}.$$

□

4.5 Мультипликативное свойство математического ожидания независимых случайных величин

Теорема 4.6. *Пусть $X_1, \dots, X_N$ – независимые случайные величины. Тогда*

$$\mathbf{E}\left(\prod_{i=1}^{N} X_i \right) = \prod_{i=1}^{N} \mathbf{E}\, X_i.$$

Доказательство. 1. Докажем утверждение для двух случайных величин U и V. Обозначим $A_j = \{U(\omega) = u_j\}, B_k = \{V(\omega) = v_k\}$.

События $A_j \cap B_k, \quad (j = \overline{1,l}, k = \overline{1,m})$ образуют разбиение Ω, причем если $\omega \in A_j \cap B_k$, то $U(\omega) \cdot V(\omega) = u_j \cdot v_k$. Из леммы 3.1

$$\mathbf{E}(UV) = \sum_{j=1}^{l} \sum_{k=1}^{m} u_j v_k \, \mathbf{P}\{A_j \cap B_k\}.$$

Используя независимость случайных величин, получим, что

$$\mathbf{E}(UV) = \sum_{j=1}^{l} \sum_{k=1}^{m} u_j v_k \, \mathbf{P}(A_j) \, \mathbf{P}(B_k) =$$

$$= \left(\sum_{j=1}^{l} u_j \, \mathbf{P}(A_j) \right) \cdot \left(\sum_{k=1}^{m} v_k \, \mathbf{P}(B_k) \right) = \mathbf{E}\, U \cdot \mathbf{E}\, V.$$

2. Пусть теперь $N \geq 3$ и для любого числа сомножителей, не превосходящего $N-1$, теорема верна. Представим

$$X_1 \cdot \ldots \cdot X_N = U \cdot V,$$

где $U = X_1 \cdot \ldots \cdot X_{N-1}, \quad V = X_N$. В силу теоремы 4.2 случайные величины U и V независимы, поэтому

$$\mathbf{E}(UV) = \mathbf{E}\, U \cdot \mathbf{E}\, V = \mathbf{E}(X_1 \cdot \ldots \cdot X_{N-1}) \cdot \mathbf{E}\, X_N.$$

Но по предположению индукции первый множитель в правой части равен произведению математических ожиданий случайных величин, следовательно,

$$\mathbf{E}(X_1 \cdot \ldots \cdot X_N) = \mathbf{E}\, X_1 \cdot \ldots \cdot \mathbf{E}\, X_N.$$

□

Глава 5

Суммирование независимых случайных величин

Начиная с 1712 г., когда Я. Бернулли открыл закон больших чисел, на протяжении двух с половиной столетий основные теоремы теории вероятностей касались поведения сумм независимых случайных величин, играющих весьма важную роль как в теории вероятностей и математической статистике, так и в их приложениях. Около 50 лет назад вышла книга Б. В. Гнеденко и А. Н. Колмогорова "Предельные распределения для сумм независимых случайных величин", в которую были включены как результаты предшественников, так и результаты авторов, внесших значительный вклад в развитие теории и методов суммирования случайных величин. Появление этой книги вызвало новый интерес к данному вопросу и послужило толчком в развитии методов исследования распределений сумм случайных величин.

Находить точные распределения сумм большого числа случайных величин, как правило, весьма сложно даже для простых распределений. Так на первой лекции мы искали точное распределение суммы очков, выпавших на 6 костях, которая представляет собой сумму $X_1+X_2+\ldots+X_6$ всего лишь шести слагаемых. Точные расчеты оказались весьма непростыми, хотя распределения самих X_i несложные: каждая из этих случайных величин принимает значения $1, 2, \ldots, 6$ с равными вероятностями: $\mathbf{P}\{X_i = k\} = 1/6$.

Всюду в дальнейшем для суммы случайных величин $X_1, \ldots, X_N$ будем использовать обозначение

$$S_N = X_1 + \cdots + X_N.$$

Нам понадобится некоторый аналитический аппарат, связанный с математическими ожиданиями случайных величин.

5.1 Производящая функция целочисленной случайной величины

Пусть случайная величина X принимает целые неотрицательные значения: $0, 1, 2, \dots, m$ с вероятностями $p_0, p_1, \dots, p_m$. Рассмотрим функцию действительной переменной z

$$G(z, X) = \mathbf{E}(z^X) = \sum_{k=1}^{m} z^k \cdot p_k. \tag{5.1}$$

Эта функция называется *производящей функцией* случайной величины X.

Замечание. Часть слагаемых в сумме (5.1) может равняться 0, если соответствующая вероятность $p_k = 0$. Производящая функция может быть определена и для целочисленных случайных величин, принимающих бесконечное число значений:

$$G(z, X) = \sum_{k=1}^{\infty} z^k p_k,$$

но тогда степенной ряд может сходиться не при всех z. Однако при $|z| \le 1$ такие ряды сходятся, причем равномерно, так как $\sum\limits_{k=1}^{\infty} p_k = 1$.

Если случайные величины $X_1, \dots, X_N$ езависимы, то для производящей функции выполняется свойство *мультипликативности:*

$$G(z, S_N) = \prod_{k=1}^{N} G(z, X_k).$$

По производящей функции распределение случайной величины восстанавливается однозначно, поскольку вероятности p_k есть не что иное, как коэффициенты разложения в ряд Тейлора для этой функции в окрестности 0:

$$p_0 = G(0, X), \quad p_1 = G'(0, X), \dots p_k = \frac{G^{(k)}(0, X)}{k!}.$$

Пример 5.1. Пусть производящая функция случайной величины X равна

$$G(z, X) = (1 - p + pz)^2.$$

Найти распределение X.

Очевидно, что

$$G(z,X) = (1-p)^2 + 2p(1-p)z + p^2 z^2.$$

Поскольку вероятности $p_k = \mathbf{P}\{X = k\}$ — это коэффициенты при z^k, то получаем

$$p_0 = (1-p)^2, \quad p_1 = 2p(1-p), \quad p_2 = p^2.$$

Таким образом случайная величина X, имеющая заданную производящую функцию распределения принимает значения $0, 1, 2$ с вычисленными выше вероятностями.

Задача 5.1. Найдите производящую функцию случайной величины X, принимающей значения $0, 1, \ldots, n$ с вероятностями

$$p_k = C_n^k p^k (1-p)^{n-k}.$$

Такое распределение вероятностей называют *биномиальным* с параметрами $n \in \mathcal{N}$ и $0 < p < 1$. В этом случае будем писать $X \sim B(n,p)$.

Задача 5.2. Найдите производящую функцию случайной величины X, принимающей значения $0, 1, 2, \ldots$ с вероятностями $p_k = \frac{\lambda^k \cdot e^{-\lambda}}{k!}$, где $\lambda > 0$ – некоторый параметр. Это распределение называют *распределением Пуассона* с параметром λ и обозначают $X \sim \Pi(\lambda)$.

Задача 5.3. Используя свойство мультипликативности производящих функций, покажите, что сумма независимых случайных величин X_1 и X_2, имеющих распределение Пуассона с параметрами λ_1 и λ_2 соответственно , также распределена по закону Пуассона,
но с параметром $\lambda_1 + \lambda_2$.

Задача 5.4. Пусть $X \geq 0$ – целочисленная случайная величина и известна ее производящая функция $G(z,X)$. Найдите $\mathbf{E}\,X, \quad \mathbf{D}\,X$.

5.2 Производящая функция моментов

Рассмотрим случайную величину X, распределение которой задается таблицей

$$\begin{matrix} x_1, & x_2, & \ldots, & x_m \\ p_1, & p_2, & \ldots, & p_m, \end{matrix} \tag{5.2}$$

в которой указаны все различные между собой значения случайной величины и соответствующие им вероятности.

Производящей функцией моментов случайной величины X называется функция действительной переменной h, определенная следующим образом:

$$g(h,X) = \mathbf{E}\, e^{hX} = \sum_{k=1}^{m} e^{hx_k} p_k.$$

Замечание. Поясним происхождение названия этой функции. Рассмотрим разложение в ряд Тейлора по степеням h функции e^{hX} :

$$e^{hX} = 1 + \frac{hX}{1!} + \frac{h^2X^2}{2!} + \ldots + \frac{h^nX^n}{n!} + \ldots .$$

если формально взять математические ожидания от обеих частей этого равенства, то получим

$$\mathbf{E}\, e^{hX} = 1 + \frac{\mathbf{E}\, X}{1!}h + \frac{\mathbf{E}\, X^2}{2!}h^2 + \ldots + \frac{\mathbf{E}\, X^n}{n!}h^n + \ldots .$$

Таким образом по функции $g(h,X)$, разложив ее в ряд Тейлора, можно найти $m_n = \mathbf{E}\, X^n$, называемые моментами случайной величины X.

Задача 5.5. Пусть $X_1, \ldots, X_N$ — независимые случайные величины,

$$S_n = X_1 + X_2 + \cdots + X_N$$

— сумма независимых случайных величин. Докажите свойство мультипликативности для производящей функции моментов:

$$g(h,S_N) = g(h,X_1) \cdot g(h,X_2) \cdot \ldots \cdot g(h,X_N).$$

Сформулируем следующее свойство показательных функций.

Лемма 5.1. Пусть $-\infty < \alpha_1 < \alpha_2 < \alpha_3 < \ldots \infty$ — произвольные действительные числа. Если для всех $-\infty < h < \infty$ выполняется

$$\lambda_1 e^{\alpha_1 h} + \lambda_2 e^{\alpha_2 h} + \ldots + \lambda_m e^{\alpha_m h} = 0,$$

то $\lambda_1 = \lambda_2 = \cdots = \lambda_m = 0$.

Доказательство. Доказательство проведем для случая $m = 3$. Итак, пусть

$$\lambda_1 e^{\alpha_1 h} + \lambda_2 e^{\alpha_2 h} + \lambda_3 e^{\alpha_3 h} = 0 \tag{5.3}$$

для всех h. Продифференцируем (5.3) дважды по h :

$$\begin{array}{lllllll} \alpha_1\lambda_1 e^{\alpha_1 h} & + & \alpha_2\lambda_2 e^{\alpha_2 h} & + & \alpha_3\lambda_3 e^{\alpha_3 h} & = 0, \\ \alpha_1^2\lambda_1 e^{\alpha_1 h} & + & \alpha_2^2\lambda_2 e^{\alpha_2 h} & + & \alpha_3^2\lambda_3 e^{\alpha_3 h} & = 0. \end{array} \tag{5.4}$$

Подставляя $h = 0$ в (5.3),(5.4) получим систему линейных уравнений относительно переменных $\lambda_1, \lambda_2, \lambda_3$:

$$\begin{array}{ccccccc} \lambda_1 & + & \lambda_2 & + & \lambda_3 & = 0, \\ \alpha_1\lambda_1 & + & \alpha_2\lambda_2 & + & \alpha_3\lambda_3 & = 0, \\ \alpha_1^2\lambda_1 & + & \alpha_2^2\lambda_2 & + & \alpha_3^2\lambda_3 & = 0. \end{array}$$

Определитель системы

$$\begin{vmatrix} 1 & 1 & 1 \\ \alpha_1 & \alpha_2 & \alpha_3 \\ \alpha_1^2 & \alpha_2^2 & \alpha_3^2 \end{vmatrix}$$

— это определитель Вандермонда. Из курса линейной алгебры известно, что при $\alpha_1 < \alpha_2 < \alpha_3$ этот определитель не равен 0, и значит, однородная система линейных уравнений имеет единственное решение

$$\lambda_1 = \lambda_2 = \lambda_3 = 0.$$

$\square$

Замечание. Свойство показательных функций, сформулированное в лемме, в математическом анализе называется линейной независимостью этих функций.

Задача 5.6. Используя это свойство показательных функций, докажите, что если производящие функции моментов двух случайных величин X и Y совпадают при всех h, то эти случайные величины имеют одинаковые распределения, т.е. принимают одинаковые значения с равными вероятностями.

Таким образом, производящая функция моментов случайной величины однозначно определяет распределение. Более того, соответствие между производящими функциями моментов (а также производящими функциями целочисленных случайных величин) и распределениями случайных величин оказывается непрерывным в том смысле, что если распределения близки между собой, то близки и производящие функции, и обратно.

5.3 Свойства числовых характеристик распределений сумм независимых случайных величин

Из свойств математического ожидания следует, что

$$\mathbf{E}\, S_N = \mathbf{E}\, X_1 + \mathbf{E}\, X_2 + \ldots + \mathbf{E}\, X_N.$$

Существуют ли другие числовые характеристики, которые для независимых случайных величин обладали столь же простым свойством?

Ответ на этот вопрос положителен и одной из таких характеристик является *дисперсия.*

Определение 5.1. Дисперсией случайной величины X называется число

$$\mathbf{D}\,X = \mathbf{E}(X - \mathbf{E}\,X)^2.$$

Эта числовая характеристика распределения случайной величины является мерой разброса (рассеивания) значений случайной величины от ее среднего значения $\mathbf{E}\,X$. Если распределение задано таблицей (5.2), то

$$\mathbf{D}\,X = \sum_{k=1}^{m} (x_k - \mathbf{E}\,X)^2 p_k.$$

Иногда дисперсию удобно вычислять по - другому, используя тот факт, что

$$\mathbf{D}\,X = \mathbf{E}\,X^2 - (\mathbf{E}\,X)^2 = m_2 - (m_1)^2,$$

где $m_1,\ m_2$ — обозначения для моментов первого и второго порядков. Для момента первого порядка (математического ожидания) мы также будем использовать обозначение $m = m_1$.

Действительно, используя простые преобразования и свойства математического ожидания, имеем

$$\mathbf{D}\,X = \mathbf{E}(X - \mathbf{E}\,X)^2 = \mathbf{E}\left(X^2 - 2X\,\mathbf{E}\,X + (\mathbf{E}\,X)^2\right) =$$

$$= \mathbf{E}\,X^2 - 2\,\mathbf{E}\,X \cdot \mathbf{E}\,X + (\mathbf{E}\,X)^2 = \mathbf{E}\,X^2 - (\mathbf{E}\,X)^2.$$

Перечислим свойства дисперсии.

1. Для любой случайной величины $\mathbf{D}\,X \geq 0$. Отсюда в частности следует неравенство Коши-Буняковского:

$$\mathbf{E}\,X^2 \geq (\mathbf{E}\,X)^2.$$

2. Дисперсия не изменяется при добавлении к случайной величине произвольной константы $c:\ \mathbf{D}(X + c) = \mathbf{D}\,X$.

3. $\mathbf{D}(cX) = c^2\,\mathbf{D}\,X$ – при умножении случайной величины на константу c ее дисперсия изменяется в c^2 раз.

4. Если $X_1, \ldots, X_N$ – независимые случайные величины, то

$$\mathbf{D}\,S_N = \mathbf{D}\,X_1 + \ldots + \mathbf{D}\,X_N. \tag{5.5}$$

Доказательство. Доказательство этого свойства сначала проведем для двух случайных величин X_1 и X_2. Введем новые случайные величины

$$Y_1 = X_1 - \mathbf{E}\, X_1, \quad Y_2 = X_2 - \mathbf{E}\, X_2,$$

тогда $\mathbf{E}\, Y_1 = \mathbf{E}\, Y_2 = 0$. Применяя свойства математического ожидания, получим

$$\mathbf{D}(X_1 + X_2) = \mathbf{E}(X_1 + X_2 - (\mathbf{E}\, X_1 + \mathbf{E}\, X_2))^2 = \mathbf{E}(Y_1 + Y_2)^2 =$$

$$\mathbf{E}(Y_1)^2 + 2\,\mathbf{E}(Y_1 Y_2) + \mathbf{E}(Y_2)^2 = \mathbf{D}\, X_1 + \mathbf{D}\, X_2 + 2\,\mathbf{E}\, Y_1 \cdot \mathbf{E}\, Y_2 = \mathbf{D}\, X_1 + \mathbf{D}\, X_2.$$

Далее по индукции можно получить, что утверждение верно для произвольного числа случайных величин N. □

Можно рассматривать и другие числовые характеристики распределений, обладающих свойством, подобным свойству (5.5) дисперсий. Их можно получать следующим образом. Рассмотрим производящую функцию моментов

$$\mathbf{E}\, e^{hS_N} = \prod_{k=1}^{N} \mathbf{E}\, e^{hX_k},$$

ее логарифм равен

$$\ln(\mathbf{E}\, e^{hS_N}) = \sum_{j=1}^{N} \ln(\mathbf{E}\, e^{hX_j}),$$

а любые производные логарифмов будут складываться.

Задача 5.7. Проверить, что

$$\frac{d}{dh} \ln(\mathbf{E}\, e^{hX})|_{h=0} = \mathbf{E}\, X,$$

$$\frac{d^2}{dh^2} \ln(\mathbf{E}\, e^{hX})|_{h=0} = \mathbf{D}\, X,$$

$$\frac{d^3}{dh^3} \ln(\mathbf{E}\, e^{hX})|_{h=0} = \mathbf{E}(X - \mathbf{E}\, X)^3.$$

Замечание. Коэффициенты в разложении функции $\ln(\mathbf{E}\, e^{hX})$ при $h^k/k!$ называют *семиинвариантами* случайной величины X. Обычно семиинвариант k – ого порядка обозначают $\varkappa_k$. Так $\varkappa_1 = \mathbf{E}\, X$, $\varkappa_2 = \mathbf{E}(X - \mathbf{E}\, X)^2$, $\varkappa_3 = \mathbf{E}(X - \mathbf{E}\, X)^3$.

Задача 5.8. Выразите через моменты семиинварианты четвертого и пятого порядков.

Задача 5.9. Докажите следующее свойство семиинвариантов для независимых случайных величин:

$$\varkappa_k(S_N) = \varkappa_k(X_1) + \ldots + \varkappa_k(X_N).$$

Числовые характеристики $\mathbf{E}(X - \mathbf{E}\,X)^k$ называются *центральными моментами порядка k*.

Задача 5.10. Выразите центральные моменты до пятого порядка включительно через моменты и через семиинварианты.

Задача 5.11. Рассмотрим функцию $f(c) = \mathbf{E}(X - c)^2$. При каком значении c функция $f(c)$ достигает минимального значения?

Глава 6

Неравенства Чебышева. Отклонения сумм независимых случайных величин

6.1 Схемы Бернулли и Пуассона

Пусть $A_1, \ldots, A_N$— независимые случайные события. Обозначим

$$p_k = \mathbf{P}(A_k), \quad q_k = \mathbf{P}(\overline{A_k}) = 1 - p_k.$$

Рассмотрим случайную величину $\mu_N(\omega)$ – число наступивших событий среди $A_1, \ldots, A_N$, если осуществился элементарный исход ω. Другими словами, $\mu_N(\omega)$ – это число событий A_k, которым одновременно принадлежит исход ω. Эту случайную величину удобно выразить через индикаторы случайных событий, а именно

$$\mu_N(\omega) = I_1(\omega) + \ldots + I_N(\omega), \quad I_k(\omega) = I_{A_k}(\omega).$$

Из независимости событий следует независимость их индикаторов, поэтому, учитывая свойства математического ожидания и дисперсии для независимых случайных величин, получим

$$\mathbf{E}\,\mu_N = \mathbf{E}\,I_1 + \ldots + \mathbf{E}\,I_N = p_1 + \ldots + p_N,$$

$$\mathbf{D}\,\mu_N = \mathbf{D}\,I_1 + \ldots + \mathbf{D}\,I_N = p_1q_1 + \ldots + p_Nq_N.$$

Производящая функция случайной величины μ_N равна

$$\mathbf{E}(z^{\mu_N}) = \prod_{j=1}^{N} \mathbf{E}\,z^{I_{A_j}} = \prod_{j=1}^{N} (q_j + zp_j).$$

Коэффициенты разложения производящей функции по степеням z — это вероятности $\mathbf{P}\{\mu_N(\omega) = k\}$.

В частности, если $p_1 = p_2 = \ldots = p_N = p, \quad q = 1 - p$, то

$$\mathbf{E}(z^{\mu_N}) = (q + zp)^N = \sum_{k=0}^{N} C_N^k p^k q^{N-k} z^k.$$

Следовательно, в этом случае

$$\mathbf{P}\{\omega : \mu_N(\omega) = k\} = C_N^k p^k q^{N-k}, \quad 0 \leq k \leq N.$$

Такое распределение вероятностей называют *биномиальным* с параметрами (N, p), так как вероятности в этом случае определяются по формулам бинома Ньютона для $(p + q)^N$.

В общем случае, когда вероятности событий A_k не все равны между собой, формулы получаются громоздкими.

Описанная выше модель называется схемой Бернулли для равных вероятностей и схемой Пуассона в общем случае.

6.2 Неравенства Чебышева

В дальнейшем нам понадобятся две теоремы, носящие имя известного русского математика П.Л.Чебышева (1821-1894 гг.), выпускника Московского университета, академика, члена многих академий мира. П.Л.Чебышев создал метод математических ожиданий, благодаря которому упростились и стали возможными доказательства многих теорем.

Лемма 6.1. Пусть случайная величина $Y \geq 0, \quad \varepsilon > 0$ – произвольное положительное число. Тогда

$$\mathbf{P}\{\omega : Y(\omega) \geq \varepsilon\} \leq \frac{\mathbf{E}\,Y}{\varepsilon}. \tag{6.1}$$

Доказательство. Обозначим $A = \{\omega : Y(\omega) \geq \varepsilon\}$. Тогда справедливо равенство

$$I_A(\omega) + I_{\overline{A}}(\omega) = 1.$$

Умножим обе части этого равенства на Y, получим

$$Y(\omega) = Y(\omega)I_A(\omega) + Y(\omega)I_{\overline{A}}(\omega).$$

Второе слагаемое неотрицательно, а первое можно оценить следующим образом:

$$Y(\omega)I_A(\omega) \geq \varepsilon I_A(\omega).$$

Но тогда

$$\mathbf{E}\,Y(\omega) \geq \mathbf{E}(Y(\omega)I_A(\omega)) \geq \varepsilon \cdot \mathbf{P}(A),$$

откуда и вытекает утверждение леммы. □

Теорема 6.1. *(Неравенство Чебышева)*

Пусть X – случайная величина, $\mathbf{E}\,X = a, \quad \varepsilon > 0$. *Тогда*

$$\mathbf{P}\{\omega : |X(\omega) - a| \geq \varepsilon\} \leq \frac{\mathbf{D}\,X}{\varepsilon^2}.$$

Доказательство. Обозначим $Y = (X - a)^2 \geq 0$. Проводя простые преобразования и используя утверждение леммы, получим

$$\mathbf{P}\{|X - a| \geq \varepsilon\} = \mathbf{P}\{Y \geq \varepsilon^2\} \leq \frac{\mathbf{E}\,Y}{\varepsilon^2} = \frac{\mathbf{D}\,X}{\varepsilon^2}.$$

□

Замечание. Введем еще одну числовую характеристику распределения случайной величины X — *стандартное отклонение* $\sigma_X = \sqrt{\mathbf{D}\,X}$. Стандартное отклонение также как и дисперсия является характеристикой рассеяния значений случайной величины, но имеет ту же размерность, что и сама случайная величина X. Рассмотрим $\varepsilon = t\sigma_X, \quad t > 0$. Тогда из неравенства Чебышева следует, что вероятность отклонений,[1] превышающих по абсолютной величине стандартное в t раз, не превосходит $\frac{1}{t^2}$:

$$\mathbf{P}\{|X - a| \geq t\sigma_X\} \leq \frac{\mathbf{D}\,X}{t^2\sigma_X^2} = \frac{1}{t^2},$$

а вероятность отклонений, в t раз меньших стандартного, наоборот, больше $1 - \frac{1}{t^2}$:

$$\mathbf{P}\{|X - a| < t\sigma_X\} \geq 1 - \frac{1}{t^2}.$$

Теорема 6.2. *(Экспоненциальное неравенство Чебышева)*

Для любого $h > 0$

$$\mathbf{P}\{\omega : X(\omega) \geq \varepsilon\} \leq \frac{\mathbf{E}\,e^{hX}}{e^{h\varepsilon}}.$$

Доказательство. Действительно,

$$\mathbf{P}\{X(\omega) \geq \varepsilon\} = \mathbf{P}\{e^{hX} \geq e^{h\varepsilon}\}.$$

Применяя теперь лемму 6.1 к случайной величине $Y = e^{hX}$, получим нужное неравенство. □

[1] Отклонением X (может быть неудачный, но принятый термин) называют разность $X - \mathbf{E}\,X$.

Замечание. В обоих случаях неравенства удобны для изучения распределений сумм независимых случайных величин, так как дисперсии в этом случае складываются, а производящие функции моментов перемножаются.

6.3 Отклонения сумм независимых случайных величин

Применим неравенство Чебышева к суммам независимых случайных величин.

Пусть $X_1, X_2 \ldots$ - последовательность независимых случайных величин, для которых $\mathbf{E}\, X_k = a_k, \quad \mathbf{D}\, X_k = b_k^2$. Введем

$$S_N = X_1 + \ldots + X_N,$$

$$A_N = \mathbf{E}\, S_N = a_1 + \ldots + a_N,$$
$$B_N = \mathbf{D}\, S_N = b_1^2 + \ldots + b_N^2.$$

Тогда

$$\mathbf{P}\{\omega : |S_N - A_N| \geq t\sqrt{B_N}\} \leq \frac{1}{t^2} \tag{6.2}$$

В действительности во многих случаях вероятность (6.2) при больших значениях t можно оценить гораздо точнее.

Пример 6.1. Рассмотрим независимые случайные величины $X_1, \ldots, X_N$, каждая из которых принимает значения 1 и -1 с вероятностями

$$\mathbf{P}\{X_k = 1\} = \mathbf{P}\{X_k = -1\} = \frac{1}{2}.$$

Для этих случайных величин $A_N = \mathbf{E}\, S_N = 0, \quad B_N = \mathbf{D}\, S_N = N$. Но тогда

$$\mathbf{P}\{\omega : |S_N| \geq t\sqrt{N}\} \leq \frac{1}{t^2}.$$

В данном случае, как это будет показано ниже, вероятность можно оценить сверху величиной $2e^{-t^2/2}$, что при больших и даже не слишком больших t гораздо меньше. Так для $t = 4 \quad \frac{1}{t^2} = 0.0625$, а $2e^{-t^2/2} = 0.000671$, при $t = 5 \quad \frac{1}{t^2} = 0.04$, $e^{-t^2/2} = 7.453 \cdot 10^{-6}$.

Для получения более точной оценки понадобится экспоненциальное неравенство Чебышева. Заметим также, что в силу симметричности распределений отдельных слагаемых X_k сумма S_N также имеет симметричное распределение (вероятности противоположных по знаку значений равны), но тогда

$$\mathbf{P}\{\omega : |S_N| \geq k\} = 2\,\mathbf{P}\{\omega : S_N \geq k\}$$

для любого положительного числа k. Следовательно, при всех $h > 0$

$$\mathbf{P}\{\omega : |S_N| \geq t\sqrt{N}\} \leq 2\frac{\mathbf{E}\, e^{hS_N}}{e^{ht\sqrt{N}}}. \tag{6.3}$$

Вычислим производящую функцию моментов:

$$\mathbf{E}\, e^{hS_N} = \prod_{j=1}^{N}(\frac{1}{2}e^h + \frac{1}{2}e^{-h}) = \prod_{j=1}^{N} ch\, h = (ch\, h)^N,$$

где $ch\, h$ обозначает косинус гиперболический числа h. Перепишем (6.3) в виде

$$\mathbf{P}\{|S_N| \geq t\sqrt{N}\} \leq 2e^{-ht\sqrt{N}} \cdot (ch\, h)^N = 2f(h).$$

Выберем теперь h таким образом, чтобы минимизировать $f(h)$. Поскольку минимумы функций $f(h)$ и $\ln f(h)$ достигаются одновременно, то достаточно это сделать для

$$g(h) = \ln f(h) = -ht\sqrt{N} + N \ln ch\, h = N(-h\frac{t}{\sqrt{N}} + \ln ch\, h).$$

Точное решение этой задачи легко выписать, но вид этого решения не удобен для дальнейшего анализа. Можно поступить следующим образом: оценить $g(h)$ сверху более простой функцией и выбрать h, минимизирующее эту более простую функцию.

Обозначим $\varepsilon = \frac{t}{\sqrt{N}}$, тогда функцию $g(h)$ можно переписать следующим образом:

$$g(h) = N(-\varepsilon h + \ln ch\, h).$$

Вычислим первые две производные функции $\ln ch\, h = u(h)$:

$$u'(h) = \frac{sh\, h}{ch\, h}, \quad u''(h) = \frac{ch^2\, h - sh^2\, h}{ch^2\, h} = \frac{1}{ch^2\, h}.$$

Поскольку при всех h $\;ch\, h \geq 1$, то

$$u(h) = u(0) + hu'(0) + \frac{h^2}{2}u''(\theta h) \leq \frac{h^2}{2}, \quad (\theta \in (0,1))\,.$$

Следовательно,

$$g(h) \leq N(-\varepsilon h + \frac{h^2}{2}).$$

Выберем теперь $h = \varepsilon$, чтобы минимизировать правую часть неравенства. Получим
$$\mathbf{P}\{|S_N| \geq t\sqrt{N}\} \leq 2e^{-N\varepsilon^2/2}.$$
Вспоминая теперь, что $\varepsilon = \frac{t}{\sqrt{N}}$, получаем необходимое неравенство
$$\mathbf{P}\{|S_N| \geq t\sqrt{N}\} \leq 2e^{-\frac{t^2}{2}}.$$

Замечание. Доказательство последнего неравенства более полно использовало независимость случайных величин, а именно свойство производящей функции моментов, а также более сильное экспоненциальное неравенство Чебышева. Это позволило в конечном итоге получить гораздо более сильный результат, чем в неравенстве (6.2).

Пример 6.2. Рассмотрим генуэзскую лотерею с точки зрения игрока. Допустим, что игрок делает ставку на пару номеров (i, j). Выигрыш игрока в k- ом тираже составляет
$$X_k = \begin{cases} 270M & \text{с вероятностью} \quad p = \frac{2}{801}, \\ 0 & \text{с вероятностью} \quad 1 - p = \frac{799}{801}, \end{cases}$$
где M- ставка игрока. Обозначим $X'_k = X_k - M$ – изменение капитала игрока после k -ого тиража, тогда суммарное изменение капитала после N тиражей запишется как
$$S'_N = X'_1 + X'_2 + \ldots + X'_N.$$
Поскольку
$$\mathbf{E}\, X'_k = \mathbf{E}\, X_k - M = \frac{540}{801}M - M = -\frac{29}{89}M \approx 0.3258M,$$
$$\mathbf{D}\, X'_k = \mathbf{D}\, X_k = 270^2 \cdot M^2 pq,$$
$$\sigma_{X'_k} = 270 \cdot M\sqrt{pq} \approx 13.4747 \cdot M,$$
то
$$\mathbf{E}\, S_{N'} \approx -0.3258 \cdot MN,$$
$$\sigma'_N = \sqrt{\mathbf{D}\, S'_N} \approx 13.4747 \cdot M\sqrt{N}.$$
Выберем $t > 0$. Тогда из неравенства Чебышева имеем
$$\mathbf{P}\{\omega : |S'_N - \mathbf{E}\, S'_N| < t\sigma'_{S'_N}\} \geq 1 - \frac{1}{t^2}.$$

Раскрыв модуль, получаем, что для любого $t > 0$

$$\mathbf{P}\{\mathbf{E}\, S'_N - t\sigma'_N < S'_N < \mathbf{E}\, S'_N + t\sigma'_N\} \geq 1 - \frac{1}{t^2}.$$

Для игрока важен ответ на вопрос: при каком N вероятность отрицательных значений изменения капитала (т.е. проигрыша) близка к 1? Например, если $t = 5$, то $1 - \frac{1}{t^2} = 0.96$. Но тогда с вероятностью 0.96 и большей

$$-0.3258MN - 5 \cdot 13.4747 \cdot M\sqrt{N} < S'_N < -0.3258MN + 5 \cdot 13.4747M\sqrt{N}.$$

Невыгодность игры для игрока становится заметной не сразу, хотя ясно, что начиная с некоторого $N(N > (\frac{5 \cdot 13.4747}{0.3258})^2)$, правая часть неравенства становится отрицательной. На самом деле до разорения понадобится N значительно меньше, как это следует из уточнения неравенства Чебышева. Для устроителей лотереи игра выгодна, поскольку участвуют в каждом розыгрыше много участников и в этом случае можно считать N изначально большим.

Изменение капитала игрока можно изобразить графически. На рис.6.1

Рис. 6.1: График изменения капиталла игрока с ростом числа тиражей.

случайная ломаная показывает значение капиталла игрока после каждого тиража. Если ломаная достигает в какой-либо момент оси абсцисс, то это означает, что в данный момент произошло разорение игрока.

Аналогичный вид имеют траектории, описывающие деятельность страховых компаний. Несколько позже мы рассмотрим подробно задачу о разорении игрока при игре в орлянку, где также появятся похожего вида траектории.

Глава 7

Закон больших чисел

В этой главе мы применим неравенство Чебышева к доказательству предельных теорем, касающихся поведения арифметического среднего случайных величин.

Функции от большого числа случайных переменных часто бывают "почти постоянными". В теории вероятностей утверждения о предельном постоянстве каких-либо случайных величин относят к законам больших чисел. В нашей книге (см. также главу 10) мы познакомимся с несколькими такими утверждениями. Теоремы подобного рода очень важны для приложений теории вероятностей, поскольку позволяют установить достаточные условия, при которых можно считать соответствующие случайные величины практически неизменными и на этом основании строить расчеты и прогнозы. С другой стороны, многие предельные теоремы позволяют глубже понять содержательную сторону теории вероятностей, объясняют смысл таких важных понятий как вероятность случайного события, математическое ожидание случайной величины.

В дальнейшем мы будем рассматривать последовательности независимых случайных величин. Сразу же заметим, что такие объекты существуют. Всегда можно указать вероятностное пространство, на котором определены независимые случайные величины с заданными законами распределения.[1]

7.1 Закон больших чисел в форме Чебышева

Пусть $X_1, X_2 \dots$ — последовательность независимых случайных величин.

[1]О последовательностях независимых случайных величин и событий см. также главу 10.

Обозначим
$$\overline{S_N} = \frac{S_N}{N} = \frac{X_1 + \ldots + X_N}{N}$$
– арифметическое среднее случайных величин. Тогда
$$\mathbf{E}\,\overline{S_N} = \frac{\mathbf{E}\,X_1 + \ldots + \mathbf{E}\,X_N}{N}, \quad \mathbf{D}\,\overline{S_N} = \frac{\mathbf{D}\,X_1 + \ldots + \mathbf{D}\,X_N}{N^2}.$$
Оказывается, что при достаточно общих предположениях о распределениях отдельных слагаемых X_i, их арифметическое среднее $\overline{S_N}$ "почти постоянно"в том смысле, что если N достаточно велико, то значения $\overline{S_N}$ мало отличаются от константы $\mathbf{E}\,\overline{S_N}$.

Теорема 7.1. *Пусть $X_1, X_2 \ldots$ — последовательность независимых случайных величин, дисперсии которых равномерно ограничены некоторой константой $c > 0$: $\mathbf{D}\,X_k \le c\,(k = 1, 2, \ldots)$. Тогда для любого положительного $\varepsilon > 0$ при $N \to \infty$*
$$\mathbf{P}\left\{|\overline{S_N} - \mathbf{E}\,\overline{S_N}| \ge \varepsilon\right\} \longrightarrow 0. \tag{7.1}$$
Доказательство. В силу неравенства Чебышева
$$\mathbf{P}\{|\overline{S_N} - \mathbf{E}\,\overline{S_N}| \ge \varepsilon\} \le \frac{\mathbf{D}\,\overline{S_N}}{\varepsilon^2} = \frac{\mathbf{D}\,X_1 + \ldots + \mathbf{D}\,X_N}{\varepsilon^2 N^2} \le \frac{cN}{\varepsilon^2 N^2} \longrightarrow 0$$
□

Утверждение (7.1) теоремы можно переписать иначе:
$$\mathbf{P}\{|\overline{S_N} - \mathbf{E}\,\overline{S_N}| \le \varepsilon\} \longrightarrow 1 \tag{7.2}$$
при $N \to \infty$.

Если для последовательности случайных величин выполнены (7.1) или (7.2), то говорят, что для нее выполняется *закон больших чисел.*

Следствие 7.1.1. *Пусть случайные величины $X_1, X_2 \ldots$ независимы, имеют одинаковое распределение с математическим ожиданием $\mathbf{E}\,X_k = \alpha$ и $\mathbf{D}\,X_k = \sigma^2$. Тогда для всех $\varepsilon > 0$*
$$\mathbf{P}\{|\overline{S_N} - \alpha\,| \le \varepsilon\} \longrightarrow 1.$$
Из последнего следствия вытекает, что усреднив значения случайной величины, наблюдаемые в независимых и проводимых в одинаковых условиях опытах, мы получим число, близкое к математическому ожиданию. Это является способом практического оценивания математического ожидания случайной величины.

Замечание. Утверждение теоремы и способ доказательства остаются справедливыми и для произвольных случайных величин (необязательно с конечным числом значений), если потребовать от них наличие конечных моментов первого и второго порядков. Из доказательства теоремы также видно, что условия на случайные величины можно несколько ослабить.

Введем числовые характеристики, отражающие взаимное влияние случайных величин друг на друга. Это ковариация и коэффициент корреляции случайных величин. Их обозначают соответственно $cov\,(X, Y)$ и $\rho(X, Y)$.

Определение 7.1. Ковариацией двух случайных величин X, Y называется число $cov\,(X, Y) = \mathbf{E}(X - \mathbf{E}\,X)(Y - \mathbf{E}\,Y)$.

Определение 7.2. Коэффициент корреляции $\rho(X, Y) = \frac{cov(X,Y)}{\sigma_X \sigma_Y}$.

Для независимых случайных величин ковариация и коэффициент корреляции равны 0. Если же $cov\,(X, Y) = 0$, то такие случайные величины называют некоррелированными.

Задача 7.1. Приведите пример некоррелированных случайных величин, не являющихся независимыми.

Легко показать, что некоррелированность случайных величин означает, что $\mathbf{E}\,XY = \mathbf{E}\,X \cdot \mathbf{E}\,Y$.

Задача 7.2. Докажите, что если случайные величины $X_1, X_2 \ldots$ попарно не коррелированы, имеют равномерно ограниченные дисперсии, то для них также справедливо предельное соотношение (7.1).

Задача 7.3. Покажите, что утверждение теоремы 7.1 останется справедливым, если требование независимости заменить условием

$$cov\,(X_i, X_j) \longrightarrow 0$$

равномерно при $|i - j| \to \infty$.

7.2 Теорема Бернулли. Отклонение частоты наступления события от его вероятности

Частным случаем закона больших чисел в форме Чебышева является закон больших чисел Я. Бернулли, доказанный намного раньше теоремы Чебышева совершенно другим методом.

Теорема 7.2. *Пусть $A_1, A_2 \ldots$— последовательность независимых событий с одинаковыми вероятностями $\mathbf{P}(A_k) = p$, $k = 1, 2, \ldots$. Тогда при любом $\varepsilon > 0$ и $N \to \infty$*

$$\mathbf{P}\{\omega : |\frac{\mu_N(\omega)}{N} - p| \geq \varepsilon\} \longrightarrow 0.$$

Таким образом, если число испытаний велико, то относительная частота наступления случайных событий становится близка к вероятности этих событий. Этот факт позволяет в практических задачах использовать в качестве оценки неизвестной вероятности ее приближенное значение: $p \approx \frac{\mu_N}{N}$.

Доказательство. Поскольку μ_N представляет собой сумму независимых случайных величин X_k, для которых $\mathbf{E}\,X_k = p$, $\mathbf{D}\,X_k = pq$, то применима теорема **7.1** и утверждение теоремы Бернулли является следствием более общей теоремы Чебышева. □

Неравенство Чебышева позволяет также оценить отклонения относительной частоты от вероятности, а именно

$$\mathbf{P}\{\omega : |\frac{\mu_N}{N} - p| \geq \varepsilon\} \leq \frac{\mathbf{D}(\mu_N/N)}{\varepsilon^2} = \frac{Npq}{N^2\varepsilon^2} \leq \frac{1}{4N^2\varepsilon^2}. \tag{7.3}$$

При выводе неравенства мы воспользовались тем, что $\mathbf{D}\,\mu_N = Npq$, а величина $pq = p(1-p)$ при $0 < p < 1$ достигает максимального значения $\frac{1}{4}$ в точке $p = \frac{1}{2}$.

Замечание. В действительности правую часть неравенства (7.3) можно заметно усилить. Так в учебнике А.Н.Ширяева (второе издание, стр.81) приводится следующая оценка:

$$\mathbf{P}\{\omega : |\frac{\mu_N}{N} - p| \geq \varepsilon\} \leq 2e^{-2N\varepsilon^2}.$$

Как мы уже упоминали, Я. Бернулли доказал свою теорему[2] много раньше теоремы Н.Чебышева. Выписывая точные значения вероятностей

$$\mathbf{P}\{|\mu_N - Np| < N\varepsilon\} = \sum_{Np - N\varepsilon < k < Np + N\varepsilon} C_N^k p^k q^{N-k} \tag{7.4}$$

при различных значениях p, он обнаружил, что биномиальные вероятности быстро убывают при удалении от Np. Так при $p = q = 1/2$ величина биномиальных вероятностей равна $\frac{1}{2^N} C_N^k$, поведение вероятностей определяется числами C_N^k. Приведем для иллюстрации таблицу биномиальных коэффициентов до $N \leq 10$, известную как треугольник Паскаля:

[2]Результат был опубликован в 1713г.

k	0	1	2	3	4	5	6	7	8	9	10
N=1	1	1									
2	1	2	1								
3	1	3	3	1							
4	1	4	6	4	1						
5	1	5	10	10	5	1					
6	1	6	15	20	15	6	1				
7	1	7	21	35	35	21	7	1			
8	1	8	28	56	70	56	28	8	1		
9	1	9	36	84	126	126	84	36	9	1	
10	1	10	45	120	210	252	210	120	45	10	1

Мы видим, что поведение биномиальных коэффициентов всегда одинаково: сначала они возрастают, достигают максимального значения, а затем убывают, при этом средние коэффициенты значительно больше остальных. Еще более это заметно при больших N. Так при $N = 20$

k	C_N^k
0	1
1	20
2	190
3	1140
4	4845
5	15504
6	38760
7	77520
8	125970
9	167960
10	184756

средние 6 коэффициентов (мы учли их симметричность) гораздо больше всех остальных. Таким образом, возвращаясь к вероятностям в распределении частоты успехов, замечаем, что значительная часть всей вероятности сосредоточена вокруг значения $k = 10$. То же самое верно и в общем случае. Заметив этот факт, Бернулли смог показать, что вероятность в (7.4) стремится к 1 при $N \to \infty$.

7.3 Вероятностное доказательство теоремы Вейерштрасса

Применим закон больших чисел для доказательства теоремы Вейерштрасса о равномерном приближении полиномами функции, непрерывной на отрезке.

Пусть $f(x)$ – непрерывная функция на $[0,1]$. Составим многочлены Бернштейна

$$B_n(x;f) = \sum_{k=0}^{n} C_n^k x^k (1-x)^{n-k} f(\frac{k}{n}).$$

Обозначим μ_n – число успехов в n испытаниях Бернулли с вероятностью успеха x, в отдельном испытании. Тогда

$$B_n(x;f) = \mathbf{E}\, f(\frac{\mu_n}{n}).$$

Задача 7.4. Покажите, что 1) если $f(x) = x$, то $B_n(X;f) = x$; 2) если $f(x) = x^2$, то $B_n(x;f) = \frac{n(n-1)x^2}{n^2} + \frac{x}{n}$, 3) если $f(x) = e^x$, то $B_n(x;f) = (1 + x(e^{1/n} - 1))^n$.

Докажем, что полиномы Бернштейна равномерно сходятся к $f(x)$ при $n \to \infty$. Из свойств непрерывной на отрезке функции следует, что

1) $f(x)$ ограничена и, значит, существует константа $M_f = \max\limits_{0 \le x \le 1} |f(x)|$,

2) $f(x)$ равномерно непрерывна на $[0,1]$.

Рассмотрим величину

$$\omega(\delta; f) = \sup_{|x'-x''|<\delta} |f(x') - f(x'')|,$$

называемую модулем непрерывности функции $f(x)$. Равномерная непрерывность функции на отрезке $[0,1]$ эквивалентна тому, что $\omega(\delta; f)$ монотонно стремится к 0 при $\delta \to 0$.

Поскольку $\sum\limits_{k=0}^{n} C_n^k x^k (1-x)^{n-k} = 1$ при всех $x \in [0.1]$, то

$$f(x) \equiv \sum_{k=0}^{n} C_n^k x^k (1-x)^{n-k} f(x).$$

Но тогда

$$f(x) - B_n(x;f) = \sum_{k=0}^{n} C_n^k x^k (1-x)^{n-k} \left(f(x) - f(k/n)\right).$$

Следовательно,

$$|f(x) - B_n(x;f)| \le \Sigma_1 + \Sigma_2,$$

где

$$\Sigma_1 = \sum_{k:|k-nx|<n\delta} C_n^k x^k (1-x)^{n-k} |f(x) - f(k/n)|,$$

$$\Sigma_2 = \sum_{k:|k-nx|\geq n\delta} C_n^k x^k (1-x)^{n-k} |f(x) - f(k/n)|.$$

Оценим каждую сумму по отдельности.

$$\Sigma_2 \leq 2M_f \sum_{k:|\frac{k}{n}-x|\geq\delta} C_n^k x^k (1-x)^{n-k} = 2M_f\,\mathbf{P}\{|\frac{\mu_n}{n}-x| \geq \delta\} \leq 2M_f \frac{\mathbf{D}\,\mu_n}{n^2\delta^2} \leq \frac{2M_f}{4n\delta^2},$$

$$\Sigma_1 \leq \left(\sum_{k:|\frac{k}{n}-x|<\delta} C_n^k x^k (1-x)^{n-k} \right) \omega(\delta; f) \leq \omega(\delta; f).$$

Таким образом,

$$|f(x) - B_n(x; f)| \leq \omega(\delta; f) + \frac{M_f}{2n\delta^2}.$$

Пусть $\varepsilon > 0$ – произвольное положительное число. Из равномерной непрерывности функции $f(x)$ следует, что существует δ_0 такое, что $\omega(\delta_0; f) < \frac{\varepsilon}{2}$. Выберем теперь n_0 так, чтобы при всех $n \geq n_0$ выполнялось неравенство $\frac{M_f}{2n\delta_0^2} < \frac{\varepsilon}{2}$. Но тогда, начиная с номера n_0 одновременно для всех $x \in [0, 1]$

$$|f(x) - B_n(x; f)| < \varepsilon,$$

т.е. многочлены Бернштейна равномерно сходятся к функции $f(x)$.

Замечание. Из этой теоремы можно получить вторую теорему Вейерштрасса о равномерном приближении непрерывной периодической функции тригонометрическими полиномами.

Глава 8

Неравенства для максимума сумм независимых случайных величин

До сих пор мы рассматривали вероятности, связанные с отдельными суммами случайных величин. Но иногда приходится иметь дело с последовательностями частичных сумм.

Пусть $X_1, X_2 \ldots$—последовательность независимых случайных величин. Введем новую последовательность случайных величин

$$\begin{aligned}
&S_0 = 0,\\
&S_1 = X_1,\\
&S_2 = X_1 + X_2,\\
&\ldots\\
&S_N = X_1 + \cdots + X_N, \ldots
\end{aligned}$$

Поведение сумм удобно изображать графически (См. рис. 8.1). Ломаная с вершинами (k, S_k) наглядно изображает изменения сумм с накоплением

Рис. 8.1: Поведение частичных сумм.

Рис. 8.2: Изменение капитала игрока.

слагаемых. Практический интерес представляют задачи о вероятностях нахождения сумм в заданных границах, о вероятностях выхода за какую - либо границу.

Вернемся к генуэзской лотерее. Обозначим A – начальный капитал игрока, который делает ставку M на один номер и в случае выигрыша получачает $15M$.

Изменение капитала игрока также удобно изображать графически (рис. 8.2).

Проигрыш игрока в каком - либо тираже означает уменьшение капитала на M единиц, а выигрыш – увеличение на $14M$ единиц. Если график достигает в какой - то момент оси абсцисс, то происходит разорение игрока. Аналогичным образом выглядят траектории изменения капитала страховых компаний.

8.1 Неравенство А. Н. Колмогорова.

Теорема 8.1. *(Неравенство Колмогорова.) Пусть $X_1, X_2 \dots$ —последовательность независимых случайных величин. Тогда*

$$\mathbf{P}\{\omega : \max_{1\le k\le n} |S_k - \mathbf{E}\, S_k| \ge \varepsilon\} \le \frac{\mathbf{D}\, S_n}{\varepsilon^2} \tag{8.1}$$

Доказательство. 1. Не ограничивая общности рассуждений, можно считать $\mathbf{E}\, X_k = 0$ для любого k. Если это не так, то можно ввести

$$X'_k = X_k - \mathbf{E}\, X_k, \quad S'_k = X'_1 + \dots + X'_k,$$

для которых это условие выполнено. Тогда $\mathbf{D}\,X_k = \mathbf{D}\,X'_k = \mathbf{E}(X'_k)^2$ и утверждение теоремы (8.1) примет вид

$$\mathbf{P}\{\omega : \max_{1\le k\le n} |S'_k| \ge \varepsilon\} \le \frac{\mathbf{E}(S'_k)^2}{\varepsilon^2}. \tag{8.2}$$

2. Введем события

$$\begin{aligned}
&F_1 = \{\omega : |S_1| \ge \varepsilon\},\\
&F_2 = \{\omega : |S_1| < \varepsilon, |S_2| \ge \varepsilon\},\\
&\vdots\\
&F_n = \{\omega : |S_1| < \varepsilon, |S_2| < \varepsilon, \ldots, |S_{n-1} < \varepsilon, |S_n| \ge \varepsilon\},\\
&F_0 = \{\omega : |S_1| < \varepsilon, \ldots, |S_n| < \varepsilon\}.
\end{aligned}$$

Эти события попарно не пересекаются, т.е. $F_i \cap F_j = \varnothing$, если $i \ne j$. Кроме того, $\bigcup\limits_{k=0}^{n} F_k = \Omega$. Обозначим $I_k(\omega) = I_{F_k}(\omega)$, тогда $\sum\limits_{k=0}^{n} I_k(\omega) = 1$, а для события

$$F = \bigcup_{k=1}^{n} F_k = \{\omega : \max_{1\le k\le n} |S_k| \ge \varepsilon\}$$

$I_F(\omega) = \sum\limits_{k=1}^{n} I_k(\omega)$.

Заметим, что $I_k(\omega)$ – это функция от частичных сумм $S_1, S_2, \ldots, S_k$, т.е. это функция, зависящая от $X_1, X_2, \ldots, X_k$ и не зависящая от $X_{k+1}, \ldots, X_n$.

Лемма 8.1. При $1 \le k \le n$ и всех ω

$$S_k^2(\omega) \cdot I_k(\omega) \ge \varepsilon^2 \cdot I_k(\omega).$$

Утверждение почти очевидно. Действительно, если $\omega \in \overline{F_k}$, то это верно, поскольку $I_k(\omega) = 0$. Если же $\omega \in F_k$. то $|S_k(\omega)| \ge \varepsilon,\quad I_k(\omega) = 1$, и неравенство также выполняется.

Вернемся к доказательству теоремы. Нетрудно видеть, что

$$S_n^2 = (S_k + (S_n - S_k))^2 = S_k^2 + 2S_k(S_n - S_k) + (S_n - S_k)^2 \ge$$

$$\ge S_k^2 + 2S_k(S_n - S_k).$$

Но тогда

$$S_n^2 = S_n^2 \cdot \sum_{k=0}^{n} I_k(\omega) \ge \sum_{k=1}^{n-1} S_n^2 I_k(\omega) + S_n^2 I_n(\omega) \ge$$

$$\geq \sum_{k=1}^{n-1} S_k^2 I_k(\omega) + S_n^2(\omega) I_n(\omega) + 2\sum_{k=1}^{n-1} S_k I_k (S_n - S_k),$$

и в силу леммы

$$S_n^2(\omega) \geq \varepsilon^2 \sum_{k=1}^{n} I_k(\omega) + 2\sum_{k=1}^{n-1} S_k I_k (S_n - S_k).$$

Переходя к математическим ожиданиям, получим

$$\mathbf{E}\, S_n^2 \geq \varepsilon^2 \sum_{k=1}^{n} \mathbf{E}\, I_k(\omega) + 2\sum_{k=1}^{n} \mathbf{E}\left((S_k I_k)(S_n - S_k)\right).$$

Покажем, что вторая сумма равна 0. Поскольку случайная величина $S_k I_k$ зависит от $X_1, \ldots, X_k$, а $S_n - S_k$ от $X_{k+1}, \ldots, X_n$, то эти случайные величины независимы и для всех k

$$\mathbf{E}\left((S_k I_k)(S_n - S_k)\right) = \mathbf{E}(S_k I_k)\, \mathbf{E}(S_n - S_k) = 0.$$

Но тогда

$$\mathbf{E}\, S_n^2 \geq \varepsilon^2\, \mathbf{P}(F),$$

что и доказывает теорему. □

Замечание. В этом доказательстве существенно использовалась взаимная независимость случайных величин, что позволило получить результат, много сильнее, чем в неравенстве Чебышева. Сравнивая эти два неравенства, видим, что второе можно получить как следствие из неравенства Колмогорова.

8.2 Неравенство Поля Леви

Для симметрично распределенных величин вместо неравенства Колмогорова можно использовать неравенство П. Леви.

Определение 8.1. Случайная величина распределена симметрично, если случайные величины X и $-X$ имеют одинаковое распределение.

Пример 8.1. Рассмотрим случайную величину X, принимающую два значения 1 и -1 с вероятностью $\frac{1}{2}$. Не вызывает сомнения, что X распределена симметрично.

Сформулируем два свойства симметрично распределенных случайных величин.

Лемма 8.2. Если случайная величина X имеет симметричное распределение, то

$$\mathbf{P}\{X \geq 0\} = \mathbf{P}\{X \leq 0\} \geq \frac{1}{2}.$$

Действительно, поскольку $\{X \geq 0\} \cup \{X \leq 0\} = \Omega$, то

$$\mathbf{P}\{X \geq 0\} + \mathbf{P}\{X \leq 0\} \geq 1.$$

Учитывая, что слагаемые в этой сумме равны в силу симметричности распределения, получим утверждение леммы.

Лемма 8.3. Если X, Y –независимые случайные величины, то случайная величина $Z = X + Y$ также имеет симметричное распределение.

Доказательство. Действительно,

$$\mathbf{P}\{Z = z\} = \sum_x \mathbf{P}\{X = x, Y = z - x\} = \sum_x \mathbf{P}\{X = x\}\,\mathbf{P}\{Y = z - x\} =$$

$$\sum_x \mathbf{P}\{-X = x\}\,\mathbf{P}\{-Y = z - x\} = \sum_x \mathbf{P}\{-X - Y = z\} = \mathbf{P}\{-Z = z\}.$$

$\square$

Теорема 8.2. *(Неравенство Поля Леви.) Пусть $X_1, \ldots, X_n$ – независимые симметрично распределенные случайные величины, тогда для всех a выполняется неравенство*

$$\mathbf{P}\{S_n \geq a\} \leq \mathbf{P}\{\omega : \max_{1 \leq k \leq n} S_k \geq a\} \leq 2\,\mathbf{P}\{S_n \geq a\}. \tag{8.3}$$

Доказательство. В идейном плане доказательство неравенства Леви похоже на доказательство неравенства Колмогорова. Введем события

$$\begin{aligned}
&D_1 = \{\omega : S_1(\omega) \geq a\},\\
&D_2 = \{\omega : S_1(\omega) < a, S_2(\omega) \geq a\},\\
&\vdots\\
&D_n = \{\omega : S_1(\omega) < a, \ldots, S_{n-1}(\omega) < a, S_n(\omega) \geq a\},\\
&D = \bigcup_{k=1}^{n} D_k = \{\omega : \max_{1 \leq k \leq n} S_k \geq a\}.
\end{aligned}$$

События $D_1, \ldots, D_n$ попарно не пересекаются. Введем также события

$$\begin{aligned}
&E_1 = \{\omega : S_n(\omega) - S_1(\omega) \geq 0\},\\
&E_2 = \{\omega : S_n(\omega) - S_2(\omega) \geq 0\},\\
&\vdots\\
&E_{n-1} = \{\omega : S_n(\omega) - S_{n-1}(\omega) \geq 0\},\\
&E_n = D_n.
\end{aligned}$$

Для произвольного $k \le n-1$ события D_k определяются по случайным величинам $X_1, \dots, X_k$, а события E_k – по случайным величинам $X_{k+1}, \dots, X_n$. Следовательно, они независимы. Рассмотрим новые события

$$C_k = D_k \cap E_k \subset D_k, \quad 1 \le k \le n-1,$$

$$C_n = D_n \cap E_n = D_n.$$

Заметим, что как и события D_k, события C_k попарно не пересекаются.

Если $\omega \in C_k \, (k \le n-1)$, то $S_k(\omega) \ge a, \quad S_n - S_k \ge 0$, но это означает, что $S_n(\omega) \ge a$, и следовательно, $C_k \subset \{\omega : S_n(\omega) \ge a\}$. Но тогда

$$\bigcup_{k=1}^{n} C_k \subset \{\omega : S_n(\omega) \ge a\}.$$

Отсюда следует, что

$$\sum_{k=1}^{n} \mathbf{P}(C_k) \le \mathbf{P}\{\omega : S_n(\omega) \ge a\}. \tag{8.4}$$

Рассмотрим левую часть неравенства (8.4). Из независимости D_k и E_k, получим

$$\sum_{k=1}^{n} \mathbf{P}(C_k) = \mathbf{P}(C_n) + \sum_{k=1}^{n-1} \mathbf{P}(C_k) = \mathbf{P}(C_n) + \sum_{k=1}^{n-1} \mathbf{P}(D_k)\,\mathbf{P}(E_k).$$

Поскольку из лемм следует, что $\mathbf{P}(E_k) \ge \frac{1}{2}$, то

$$\sum_{k=1}^{n} \mathbf{P}(C_k) \ge \frac{1}{2}(\sum_{k=1}^{n-1} \mathbf{P}(D_k) + \mathbf{P}(D_n)) = \frac{1}{2}\,\mathbf{P}(D).$$

Учитывая это, из (8.4) получаем

$$\frac{1}{2}\,\mathbf{P}(D) \le \mathbf{P}\{S_n \ge a\}.$$

Умножив полученное неравенство на 2, приходим к (8.3), что и доказывает теорему. □

Следствие 8.2.1.

$$\mathbf{P}\{\max_{1 \le k \le n} |S_k| \ge a\} \le 2\,\mathbf{P}\{|S_n| \ge a\}. \tag{8.5}$$

Доказательство. Поскольку

$$\{\omega : \max_{1\le k\le n} |S_k| \ge a\} = \{\omega : \max_k S_k \ge a\} \cup \{\omega : \min_k S_k \le -a\},$$

то

$$\mathbf{P}\{\max_k |S_k| \ge a\} \le \mathbf{P}\{\max_k S_k \ge a\} + \mathbf{P}\{\min_k S_k \le -a\}.$$

К первому слагаемому этой суммы можно непосредственно применить утверждение теоремы, второе же придется оценивать.

Наряду со случайными величинами X_k рассмотрим также $X'_k = -X_k$ и их суммы $S'_k = -S_k, \quad 1 \le k \le n$. Тогда, используя симметричность распределений этих случайных величин, получим

$$\mathbf{P}\{\min_k S_k \le -a\} = \mathbf{P}\{\max_k(-S_k) \ge a\} = \mathbf{P}\{\max_k S'_k \ge a\} \le 2\,\mathbf{P}\{S'_n \ge a\} =$$
$$= 2\,\mathbf{P}\{S_n \le -a\}.$$

Следовательно,

$$\mathbf{P}\{\max_k |S_k| \ge a\} \le 2\,\mathbf{P}\{S_n \ge a\} + 2\,\mathbf{P}\{S_n \le -a\} = 2\,\mathbf{P}\{|S_n| \ge a\}.$$

□

Замечание. Следует обратить внимание на то, что неравенство Леви (8.3) дает оценку, имеющую правильный порядок. В этом легко убедиться, сравнивая левые и правые части этого неравенства. Проиллюстрируем использование неравенства Леви для оценки вероятности разорения одного из игроков в "орлянку".

Пример 8.2. Пусть А и В – два игрока, начальные капиталы которых соответственно составляют a и b рублей. Пусть игрок А играет с бесконечно богатым противником, т.е. $b = \infty$. С вероятностью $p = \frac{1}{2}$ игрок А в каждой игре может выиграть 1 рубль и с вероятностью $q = \frac{1}{2}$ проиграть такую же сумму.

Обозначим $S_k = X_1 + \ldots + X_k$ – суммарный выигрыш игрока А после N игр.

Изменение выигрыша удобно изображать графически с помощью траекторий, которые выходят из 0 и в каждый момент времени могут смещаться на единицу вверх или вниз. Пример такой траектории приводится на рисунке 8.3. Разорение игрока наступает в тот момент, когда траектория достигает уровня $-a$, т.е. когда впервые $S_k = -a$.

Вероятность разорения до момента N – это $\mathbf{P}\{\min\limits_{1\le k\le N} S_k \le -a\}$. Применяя неравенство Леви, получим

$$\mathbf{P}\{\min_{1\le k\le N} S_k \le -a\} \le 2\,\mathbf{P}\{S_N \ge a\}.$$

Рис. 8.3: К задаче о разорении игрока.

Такие вероятности мы умеем оценивать (см. главу 6), а именно при $a = t\sqrt{N}$

$$\mathbf{P}\{\min_{1\le k\le N} S_k \le -t\sqrt{N}\} \le 2\,\mathbf{P}\{S_N \ge t\sqrt{N}\} \le 2e^{-t^2/2}.$$

Для иллюстрации возьмем конкретные числа: $t = 4$, $N = 625$. Тогда $2e^{-t^2/2} \approx 0.0007$, и значит, вероятность разорения за 625 игр при начальном капитале всего лишь в 100 рублей, очень невелика.

Ситуации, когда происходит игра с бесконечно богатым противником, встречаются довольно часто. Так играя в казино, каждый из игроков играет против казино, капиталы которого много больше капитала отдельного игрока. С другой стороны, устроители казино играют с большим числом игроков, суммарный капитал которых тоже можно считать бесконечно большим. Аналогичное замечание касается также деятельности страховых компаний.

Глава 9

Математические основы теории вероятностей

9.1 Общее определение вероятностного пространства

В самом начале курса было введено понятие вероятностного пространства:
$$(\Omega, \mathcal{A}, \mathbf{P})$$
как тройки элементов, в которой

- Ω — некоторое множество элементарных исходов ω;
- $\mathcal{A}$ — класс подмножеств Ω, которые в дальнейшем назывались случайными событиями;
- $\mathbf{P}$ — вероятность или распределение вероятностей — неотрицательная функция, определенная на $\mathcal{A}$.

До сих пор мы рассматривали дискретные вероятностные пространства, в которых множество элементарных исходов конечно:
$$\Omega = \{\omega_1, \dots, \omega_s\},$$
класс $\mathcal{A}$ — множество *всех подмножеств* Ω, а функция $\mathbf{P}$ определялась как
$$\mathbf{P}(A) = \sum_{\omega \in A} p(\omega),$$
где $p(\omega) \geqslant 0$ для всех ω и $\mathbf{P}(\Omega) = 1$.

Для таких вероятностных пространств случайные величины — это любые функции, зависящие от элементарного исхода, в этом случае свойства математических ожиданий выводились без особого труда. Доказательства и формулировки остальных теорем (неравенства Чебышева, Колмогорова, Леви, предельные теоремы) приведены в форме, справедливой в общем случае.

Перейдем к общему определению вероятностного пространства.
По-прежнему Ω — это множество элементарных исходов ω, но теперь уже произвольной мощности.

Подмножества Ω по-прежнему будем обозначать заглавными буквами начала латинского алфавита: $A, B, C, D, \ldots$.

Классы подмножеств — рукописными заглавными буквами $\mathcal{A}, \mathcal{B}, \mathcal{C}, \mathcal{D} \ldots$

В общей теории нам встретятся следующие классы подмножеств Ω:

1) разбиения (конечные или счетные);

2) полукольца и полуалгебры;

3) алгебры;

4) $\sigma-$ алгебры ("сигма-алгебры").

Определение 9.1. Подмножества $A_1, A_2 \ldots$ образуют **разбиение** множества A, если они попарно не пересекаются ($A_i \bigcap A_j = \varnothing$, если $i \neq j$) и $A = \bigcup\limits_i A_i$.

Разбиение называется *конечным,* если оно состоит из конечного набора подмножеств $A_1, \ldots, A_n$ и *счетным,* если элементы этого разбиения $A_1, A_2 \ldots$ можно поставить во взаимно однозначное соответствие с натуральным рядом.

Определение 9.2. Класс $\mathcal{A}$ подмножеств Ω называется **алгеброй,** если

1) $\Omega \in \mathcal{A}$,

2) из условия $A \in \mathcal{A}$ следует, что $\overline{A} \in \mathcal{A}$,

3) если A, $B \in \mathcal{A}$, то $A \bigcup B \in \mathcal{A}$.

Другими словами, класс подмножеств множества Ω образует алгебру, если он содержит достоверное событие, и операции взятия дополнения и объединения событий не выводят из этого класса.

Задача 9.1. Докажите, что 1) невозможное событие $\varnothing \in \mathcal{A}$;
2)если $A, B \in \mathcal{A}$, то $A \bigcap B$, $A \triangle B$, $A \setminus B \in \mathcal{A}$;
3) если $A_1, \ldots, A_n \in \mathcal{A}$, то $\bigcup\limits_{i=1}^{n} A_i$, $\bigcap\limits_{i=1}^{n} A_i \in \mathcal{A}$ для любого натурального n.

Определение 9.3. Класс $\mathcal{A}$ называется $\sigma-$ **алгеброй,** если выполняются свойства

1) $\Omega \in \mathcal{A}$;
2) если $A \in \mathcal{A}$, то $\overline{A} \in \mathcal{A}$;
3) для любой последовательности $A_1, A_2 \ldots \in \mathcal{A}$ следует, что

$$\bigcup_{i=1}^{\infty} A_i \in \mathcal{A}.$$

Задача 9.2. Докажите, что 1) всякая σ-алгебра является алгеброй,
2) $\bigcap\limits_{i=1}^{\infty} A_i \in \mathcal{A}$.

Определение 9.4. Класс $\mathcal{S}$ подмножеств Ω образует **полукольцо,** если

1) $\varnothing \in \mathcal{S}$;
2) если $A, B \in \mathcal{S}$, то $A \bigcap B \in \mathcal{S}$;
3) для любых $A, A_1 \in \mathcal{S}$ таких, что $A_1 \subseteq A$, существует конечное разбиение A, все элементы которого входят в $\mathcal{S}$ и одним из элементов разбиения является A_1.

Замечание.

Последнее свойство в определении полукольца можно сформулировать по-другому: разность $A \setminus A_1$ любых двух элементов полукольца может быть представлена в виде объединения конечного числа попарно не пересекающихся $A_i \in \mathcal{S}$. Заметим однако, что сама эта разность полукольцу может и не принадлежать. Кроме того, это представление может оказаться не единственным.

Полукольцо подмножеств является, как мы это увидим ниже, очень полезным техническим средством при построении теории меры.

Рассмотрим несколько примеров наиболее часто используемых полуколец.

Пример 9.1. Класс $\mathcal{S}$ полуинтервалов $[a; b)$ на числовой прямой
$(a \leqslant b)$ образует полукольцо.

Рис. 9.1: Полукольцо прямоугольников на плоскости.

Действительно, если $a = b$, то $[a; b) = \varnothing$, и, следовательно, невозможное событие является элементом этого класса. Легко видеть, что непустое пересечение двух полуинтервалов данного вида также является полуинтервалом из $\mathcal{S}$, а разность двух полуинтервалов, один из которых содержится в другом, представима и виде объединения не более чем двух непересекающихся полуинтервалов.

Пример 9.2. Класс $\mathcal{S}$ прямоугольников $[a; b) \times [c; d)$ на плоскости ($a \leqslant b$, $c \leqslant d$) образует полукольцо.

Проверьте это самостоятельно.

Этот пример легко обобщается на случай произвольного евклидова пространства $\mathbb{R}^n$, где в качестве полукольца можно рассмотреть множество n – мерных параллелепипедов.

Понятие полуинтервалов можно распространить на случай, когда элементами Ω являются функции.

Пример 9.3. Полукольцо полуинтервалов в $\mathbb{C}[0; 1]$.

Выберем некоторое множество точек: $0 \leqslant t_1 < t_2 < \ldots < t_n \leqslant 1$, и столько же произвольных полуинтервалов: $[a_i; b_i)$, $\quad (1 \leqslant i \leqslant n)$. Подмножества в $\mathbb{C}[0; 1]$ вида

$$\{f(t) :\ a_1 \leqslant f(t_1) < b_1;\ a_2 \leqslant f(t_2) < b_2; \ldots a_n \leqslant f(t_n) < b_n\} \tag{9.1}$$

образуют полукольцо. Эти подмножества $\mathbb{C}[0; 1]$ также называют полуинтервалами.

Рис. 9.2: Полуинтервалы функций.

Определение 9.5. Полукольцо $\mathcal{S}$ называется полукольцом с единицей если $\Omega \in \mathcal{S}$.

Замечание. В примере 1, чтобы получить полукольцо с единицей, надо добавить полуинтервалы вида

$$(-\infty; b), \quad [a; +\infty), \quad (-\infty; +\infty),$$

т.е. разрешить a, b принимать все значения, удовлетворяющие неравенству $-\infty \leqslant a \leqslant b \leqslant +\infty$.

Полукольцо с единицей также называют *полуалгеброй.*

9.1.1 Порожденные алгебры и σ - алгебры

Лемма 9.1. Для любого класса непустого $\mathcal{E}$ существует и только одна алгебра $\mathcal{A}_0$, такая, что $\mathcal{E} \subset \mathcal{A}_0$ и для любой алгебры $\mathcal{A}_1$, содержащей $\mathcal{E}$, $\quad \mathcal{A}_0 \subset \mathcal{A}_1$.

Определение 9.6. Такую алгебру будем обозначать $\mathcal{A}(\mathcal{E})$ и называть алгеброй, порожденной классом $\mathcal{E}$.

Аналогичная лемма верна и для $\sigma-$ алгебр.

Лемма 9.2. Для любого непустого класса $\mathcal{E}$ существует единственная $\sigma-$алгебра $\mathcal{B}_0$,такая, что $\mathcal{E} \subset \mathcal{B}_0$ и для любой $\sigma-$алгебры $\mathcal{B}_1 \supset \mathcal{E}$ следует, что
$\mathcal{B}_0 \subset \mathcal{B}_1$.

$\mathcal{B}_0$ называют $\sigma-$алгеброй, порожденной классом $\mathcal{E}$, это наименьшая $\sigma-$ алгебра, содержащая класс $\mathcal{E}$.

Для примера приведем доказательство первой леммы, вторая доказывается аналогично.

Доказательство. 1). Покажем сначала, что пересечение алгебр является алгеброй.

Пусть $\mathcal{A}_1$, $\mathcal{A}_2$ — две алгебры. Обозначим $\mathcal{A} = \mathcal{A}_1 \bigcap \mathcal{A}_2$.

Пусть $A, B \in \mathcal{A}$, но тогда $A, B \in \mathcal{A}_1, \quad A, B \in \mathcal{A}_2$ и в силу определения алгебры события Ω, $\overline{A}, A \bigcup B$ принадлежат одновременно обеим алгебрам, и, следовательно, принадлежат их пересечению $\mathcal{A}$, значит, $\mathcal{A}$ является алгеброй.

Замечание. Данное утверждение справедливо в более общем случае, когда пересечение $\bigcap\limits_{t \in T} \mathcal{A}_t$ берется по всем значениям t из *непустого множества* T, (которое может содержать бесконечно много точек и даже быть более мощным, чем счетное множество.)

2). Определим теперь $\mathcal{A}_0$ как пересечение всех алгебр, содержащих $\mathcal{E}$. Поскольку для любого класса подмножеств $\mathcal{E}$ всегда найдется хотя бы одна алгебра, содержащая $\mathcal{E}$ (например, алгебра всех подмножеств Ω), то пересечение берется по непустой системе алгебр, и, следовательно,$\mathcal{A}_0$ – минимальная алгебра, содержащая $\mathcal{E}$.

□

Замечание. Все вышесказанное справедливо и для σ - алгебр.

Приведенное доказательство ничего не говорит о конструкции алгебр, порожденных каким-либо классом подмножеств. Чтобы разобраться в этом вопросе, приведем ряд примеров.

Пример 9.4. Алгебра, порожденная конечным разбиением Ω.

Пусть $\Omega = A_1 \bigcup A_2 \bigcup \ldots \bigcup A_n, \quad A_i \bigcap A_j = \varnothing \, (i \neq j)$. Тогда наименьшая алгебра, содержащая данное разбиение, должна содержать

1) два обязательных для каждой алгебры элемента Ω, $\varnothing$,
2) каждый из элементов разбиения: $A_1, A_2, \ldots A_n$,
3) объединения любых двух элементов разбиения – $A_i \bigcup A_j$
4) объединения любых трех – $A_i \bigcup A_j \bigcup A_k, \ldots$

Таким образом, всего элементов должно быть

$$1 + C_n^1 + C_n^2 + C_n^3 + \ldots + C_n^{n-1} + 1 = 2^n.$$

Так например, разбиение, состоящее из четырех элементов, порождает алгебру из 16 элементов.

В учебнике Б.А.Севастьянова приводится лемма, утверждающая, что всякая конечная алгебра порождается некоторым разбиением, и, значит, число ее элементов может быть равным только степени числа 2.

Пример 9.5. Пусть A и B – произвольные два подмножества Ω. Разберем, как устроена минимальная алгебра, содержащая A и B. Рассмотрим разбиение Ω :

$$\mathcal{E} = \left\{AB;\ A\overline{B};\ \overline{A}B;\ \overline{AB}\right\}.$$

Порожденная этим разбиением алгебра совпадает с минимальной алгеброй, содержащей A и B, следовательно, $\mathcal{A}(\mathcal{E})$ состоит из не более чем $2^4 = 16$ элементов.

Столь же просто описываются алгебры, порожденные полукольцом с единицей.

Задача 9.3. Покажите, что алгебра, порожденная полукольцом с единицей, состоит из всех конечных объединений попарно не пересекающихся элементов полукольца.

9.1.2 Борелевские σ - алгебры множеств

Пусть теперь множество элементарных исходов совпадает с s -мерным евклидовым пространством: $\Omega = \mathbb{R}^s$. В этом случае алгебры (σ - алгебры) событий связываются с наиболее типичными подмножествами $\mathbb{R}^s$, встречающимися в математическом анализе.

Пример 9.6. Рассмотрим на числовой прямой ($s = 1$) полукольцо полуинтервалов с **1**.

Наименьшая σ -алгебра, содержащая все полуинтервалы вида $[a;b)$ $-\infty \leqslant a \leqslant b \leqslant +\infty$, называется *борелевской* и обозначается $\mathcal{B}^1$. Нетрудно убедиться в том, что произвольные интервалы и отрезки числовой прямой принадлежат борелевской σ - алгебре. Действительно,

$$(a;b) = \bigcup_{n=1}^{\infty}[a - \frac{1}{n};b)$$

$$[a;b] = \bigcap_{n=1}^{\infty}[a;b + \frac{1}{n}).$$

Поскольку всякое открытое множество на прямой можно представить в виде объединения не более чем счетного множества непересекающихся интервалов, то все открытые множества также являются борелевскими, а значит, и их дополнения — замкнутые множества — тоже борелевские.

Аналогично определяется класс борелевских множеств $\mathcal{B}^2$ как наименьшая σ - алгебра, содержащая прямоугольники в $\mathbb{R}^2$.

Борелевские алгебры в $\mathbb{R}^s$ являются естественной областью определения вероятностных мер в евклидовых пространствах.

9.1.3 Вероятностные меры или распределения вероятностей

Этот раздел посвящен третьей составляющей, входящей в определение вероятностного пространства — вероятностной мере или распределению вероятностей $\mathbf{P}$. Первое название связано с общим понятием меры в функциональном анализе. И тот, и другой термин общеприняты в вероятностной литературе.

Определение 9.7. Функция множеств $\mathbf{P}(A)$, определенная на $\mathcal{A}$, называется вероятностной мерой (вероятностью), если

- $\mathbf{P}(A) \geqslant 0 \, (\forall A \in \mathcal{A})$;
- $\mathbf{P}(\Omega) = 1$;
- выполнено свойство **счетной аддитивности:** для любой последовательности попарно несовместимых событий $A_1, A_2 \dots$ $(\bigcup\limits_{i=1}^{\infty} A_i \in \mathcal{A})$

$$\mathbf{P}\left(\bigcup_{i=1}^{\infty} A_n\right) = \sum_{i=1}^{\infty} \mathbf{P}(A_n).$$

Замечание. В этом определении класс $\mathcal{A}$ может быть полуалгеброй (кольцом с единицей), алгеброй, σ -алгеброй.

Задача 9.4. Покажите, что

1) $\mathbf{P}(\varnothing) = 0$,

2) для вероятности выполняется свойство конечной аддитивности,

3) выполняются свойства вероятности 4 – 7, приведенные в главе 3.

Приведем без доказательства теорему, очень важную для построения всей теории вероятностей.

Теорема 9.1. (О продолжении вероятностной меры.) *Пусть* $\mathbf{P}(A)$— *вероятностная мера, заданная на полукольце с единицей* $\mathcal{S}$. *Тогда существует и при том только одна вероятностная мера* $\mathbf{P}_1$, *определенная на* $\sigma-$ *алгебре* $\mathcal{B}(\mathcal{S})$, *порожденной* $\mathcal{S}$, *такая, что для* $\forall A \in \mathcal{S} \qquad \mathbf{P}(A) = \mathbf{P}_1(A)$.

Другими словами, всякая счетно-аддитивная мера может быть продолжена с полукольца с единицей на наименьшую $\sigma-$ алгебру, содержащую это полукольцо. Подробнее об этом можно прочитать в учебниках ([10],[8]).

Из аксиом в определении вероятности можно вывести другие важные свойства вероятностной меры. Для этого нам понадобятся некоторые вспомогательные утверждения.

Лемма 9.3. Пусть $\mathcal{S}-$ полукольцо и $A, A_1, \ldots, A_m \in \mathcal{S}$,
причем все $A_i \subset A, \quad A_i \bigcap A_j = \varnothing, \quad$ если $\quad i \neq j$. Тогда найдутся $A_{m+1}, \ldots A_n \in \mathcal{S}$ такие, что события

$$A_1, \ldots A_m, A_{m+1}, \ldots A_n$$

образуют конечное разбиение A.

Это утверждение почти очевидным образом вытекает из определения полукольца.

Лемма 9.4. Пусть $\mathbf{P}(A)$— вероятностная мера, определенная на полукольце $\mathcal{S}, \quad A, A_1, \ldots, A_m-$ события, удовлетворяющие условию предыдущей леммы, тогда

$$\mathbf{P}(A) \geqslant \sum_{j=1}^{m} \mathbf{P}(A_j).$$

Доказательство. Дополним $A_1, \ldots, A_m$ до разбиения
$A: \ A_1, \ldots, A_m, A_{m+1}, \ldots A_n$. Тогда из свойства счетной аддитивности следует, что

$$\mathbf{P}(A) = \sum_{j=1}^{n} \mathbf{P}(A_j) \geqslant \sum_{j=1}^{m} \mathbf{P}(A_j).$$

□

Лемма 9.5. Пусть $A \subset \bigcup\limits_{j=1}^{n} A_j, \quad A, A_1, \ldots, A_n-$ элементы полукольца с $\mathbf{1}$. Тогда

$$\mathbf{P}(A) \leqslant \sum_{j=1}^{n} \mathbf{P}(A_j).$$

Доказательство. Представим событие A в следующем виде:

$$A = \bigcup_{j=1}^{n} (A \bigcap A_j).$$

Из определения полукольца следует, что события $B_j = A \bigcap A_j \in \mathcal{S}$, при этом B_j попарно несовместимы.

Рассмотрим для начала случай $n = 2$:

$$A = B_1 \bigcup B_2 = B_1 \bigcup (B_2 \bigcap \overline{B_1}).$$

Если $\mathcal{S}$— полукольцо с **1** , то, воспользовавшись утверждением леммы 3, получим, что для $\overline{B_1} = \Omega \setminus B_1$ существует конечное разбиение:

$$\overline{B_1} = C_1 \bigcup \ldots \bigcup C_l,$$

такое, что все $C_i \in \mathcal{S}$. Но тогда

$$B_2 \bigcap \overline{B_1} = \bigcup_{i=1}^{l} (B_2 \bigcap C_i).$$

Следовательно, для A существует конечное разбиение

$$A = B_1 \bigcup (B_2 \bigcap C_1) \bigcup \ldots (B_2 \bigcap C_l),$$

все элементы которого принадлежат полукольцу $\mathcal{S}$. Но тогда в силу леммы 9.5

$$\mathbf{P}(A) = \mathbf{P}(B_1) + \sum_{i=1}^{l} \mathbf{P}(B_2 \bigcap C_i) \leqslant \mathbf{P}(B_1) + \mathbf{P}(B_2).$$

По индукции можно доказать справедливость этого неравенства для произвольного n.

$\square$

Поскольку вероятностная мера, определенная на полукольце (алгебре), может быть единственным образом продолжена на порожденную этим полукольцом (алгеброй) σ - алгебру, то в дальнейшем будем считать, что **областью определения вероятности является σ - алгебра.**

Приведем несколько важных свойств вероятностной меры, определенной на σ - алгебре $\mathcal{A}$.

- **Свойство 1.** Пусть $A_1 \subset A_2 \subset \ldots \subset A_n \subset A_{n+1} \subset \ldots \in \mathcal{A}$— неубывающая последовательность случайных событий, $A = \bigcup\limits_{n=1}^{\infty} A_n$, тогда $\mathbf{P}(A) = \lim\limits_{n\to\infty} \mathbf{P}(A_n)$.
- **Свойство 2.** Пусть $B_1 \supset B_2 \supset \ldots \supset B_n \supset B_{n+1} \supset \ldots \in \mathcal{A}$— невозрастающая последовательность случайных событий, $B = \bigcap\limits_{n=1}^{\infty} B_n$. Тогда $\mathbf{P}(B) = \lim\limits_{n\to\infty} \mathbf{P}(B_n)$
- **Свойство 3.** Если $A = \bigcup\limits_{n=1}^{\infty} A_n, \quad A_n \in \mathcal{A}$, то $\mathbf{P}(A) \leqslant \sum\limits_{n=1}^{\infty} \mathbf{P}(A_n)$.

Замечание. Эти свойства вытекают из свойства счетной аддитивности вероятностной меры, определенной на σ - алгебре. Первые два называются свойствами непрерывности по монотонным последовательностям событий, так как в этих случаях можно считать

$$A = \lim_{n\to\infty} A_n, \quad B = \lim_{n\to\infty} B_n$$

и сформулировать их в другом виде:

$$\mathbf{P}(\lim_{n\to\infty} A_n) = \lim_{n\to\infty} \mathbf{P}(A_n) \qquad \mathbf{P}(\lim_{n\to\infty} B_n) = \lim_{n\to} \mathbf{P}(B_n).$$

Иногда в определении вероятности свойство счетной аддитивности заменяют двумя свойствами: конечной аддитивностью и непрерывностью вероятности по убывающей к $\varnothing$ последовательности событий. Свойство 3 называют свойством полуаддитивности вероятности.

Докажем свойство 1.

Доказательство. Представим $A = \bigcup\limits_{n=1}^{\infty} A_n$ в виде объединения попарно не пересекающихся событий B_n :

$$B_1 = A_1, \quad B_2 = A_2 \setminus A_1, \quad B_3 = A_3 \setminus A_2, \ldots$$

$$B_n = A_n \setminus A_{n-1} \ldots .$$

Покажем, что $A = \bigcup\limits_{n=1}^{\infty} B_n \qquad A_N = \bigcup\limits_{n=1}^{N} B_n$. Действительно,

$$A_2 = B_1 \bigcup B_2,$$

$$A_3 = A_2 \bigcup B_3 = B_1 \bigcup B_2 \bigcup B_3, \ldots$$

Следовательно,

$$\mathbf{P}(A)=\sum_{n=1}^{\infty}\mathbf{P}(B_n)=\lim_{N\to\infty}\sum_{n=1}^{N}\mathbf{P}(B_n)=\lim_{N\to\infty}\mathbf{P}(\bigcup_{n=1}^{N}B_n)=\lim_{N\to\infty}\mathbf{P}(A_N).$$

□

Свойство 2 выводится из свойства 1, если ввести неубывающую последовательность событий

$$A_1=\overline{B_1},\quad A_2=\overline{B_2},\quad \ldots A_n=\overline{B_n},\ldots.$$

Докажем свойство 3.[1]

Доказательство. Снова представим событие A в виде объединения попарно несовместимых событий. Для этого введем

$$B_1=A_1,\quad B_2=A_2\setminus A_1,\quad B_3=A_3\setminus(A_1\cup A_2),\ldots,B_n=A_n\setminus\bigcup_{i=1}^{n-1}A_i.$$

Покажем, что

$$\bigcup_{n=1}^{\infty}A_n=\bigcup_{n=1}^{\infty}B_n. \tag{9.2}$$

Если $\omega_0\in\bigcup\limits_{n=1}^{\infty}A_n$, то найдется номер n такой, что $\omega_0\in A_n$. Среди таких номеров выберем наименьший — n_0. Тогда при всех $n<n_0\quad \omega\notin A_n$. Но тогда $\omega_{n_0}\in B_{n_0}$(по определению B_{n_0}.) Следовательно, $\omega_{n_0}\in\bigcup\limits_{n=1}^{\infty}B_n$.

Покажем обратное включение. Пусть $\omega_1\in\bigcup\limits_{n=1}^{\infty}B_n$. Тогда найдется номер n_1 такой, что $\omega_1\in B_{n_1}$. Поскольку $B_{n_1}\subset A_{n_1}$ по построению, то $\omega_{n_1}\in A_{n_1}\subset\bigcup\limits_{n=1}^{\infty}A_n$, что и доказывает (9.2), следовательно,

$$\mathbf{P}(\bigcup_{n=1}^{\infty}A_n)=\mathbf{P}(\bigcup_{n=1}^{\infty}B_n)=\sum_{n=1}^{\infty}\mathbf{P}(B_n)\leqslant\sum_{n=1}^{\infty}\mathbf{P}(A_n).$$

□

[1] Сравни с доказательством свойства вероятности 7, приведенном в главе 3.

9.2 Вероятностные меры в евклидовых пространствах

Остановимся подробнее именно на этом случае, поскольку именно он чаще всего используется в приложениях теории вероятностей и в математической статистике.

Пусть $\Omega = \mathbb{R}^d$, $\mathcal{A} = \mathcal{B}^d$ — наименьшая σ -алгебра, содержащая полуинтервалы (параллелепипеды).

9.2.1 Вероятностные распределения на прямой

Рассмотрим сначала задание вероятностной меры $\mathbf{P}$ на числовой прямой ($d = 1$).

Определение 9.8. Функцией распределения, соответствующей мере $\mathbf{P}$, называется функция $F_P(x)$ действительной переменной x, которая в каждой точке определяется как

$$F_P(x) = \mathbf{P}\{(-\infty; x)\}.$$

Всякая функция распределения обладает следующими свойствами:

1) $F_P(x)$— неубывающая функция x;

2) в любой точке x $F_P(x)$ непрерывна слева, т.е. $F_P(x_n) \to F_P(x)$, если $x_n \uparrow x$;

3) $0 \leqslant F_P(x) \leqslant 1$, причем $\lim\limits_{x\to-\infty} F_P(x) = 0$, $\lim\limits_{x\to+\infty} F_P(x) = 1$.

При доказательстве этих свойств функций распределения используются свойства вероятностных мер и равенство

$$F_P(x+h) - F_P(x) = \mathbf{P}([x; x+h)), \tag{9.3}$$

справедливое при всех $h > 0$. Приведем доказательство свойства 3.

Доказательство. Пусть $x_n \uparrow +\infty$. Обозначим $A_n = (-\infty; x_n)$. Легко видеть, что

$$A_1 \subset A_2 \subset \ldots; \quad \bigcup_n A_n = (-\infty; +\infty).$$

Применив одно из свойств непрерывности вероятности, имеем

$$\mathbf{P}(A_n) = F_P(x_n) \to \mathbf{P}(-\infty; +\infty) = 1.$$

□

Задача 9.5. Проведите самостоятельно доказательство остальных свойств вероятности.

Замечание. Понятие функции распределения используется также в курсе функционального анализа для неубывающих функций с ограниченной вариацией, непрерывных слева.

Определение 9.9. Функцией распределения называют всякую функцию, обладающую свойствами 1) – 3) на стр. 74.

Оказывается, что каждая функция распределения порождает вероятностную меру и при том только одну.

Между классом функций распределения и множеством вероятностных мер существует взаимно однозначное соответствие. Если $F(x)$ — функция распределения, то соответствующую ей меру $\mathbf{P}_F$ определим согласно равенствам:

$$\begin{aligned} \mathbf{P}_F([a;b)) &= F(b) - F(a), \qquad & \mathbf{P}_F((-\infty;b)) &= F(b), \\ \mathbf{P}_F([a;+\infty)) &= 1 - F(a), \qquad & \mathbf{P}_F((-\infty;+\infty)) &= 1. \end{aligned} \tag{9.4}$$

Поскольку мера $\mathbf{P}_F$ определена на полуинтервалах, образующих полукольцо с единицей, то доказав счетную аддитивность, можно утверждать, что она допускает единственное продолжение на борелевскую алгебру множеств на прямой.

Замечание. В некоторых учебниках (например, [10],[8]) функция распределения определяется как $F_P(x) = \mathbf{P}((-\infty;x])$, Отличие в свойствах этих функций только в том, что таким образом определенная функция оказывается непрерывной справа. Все остальные свойства остаются без изменений. Функция $\mathbf{P}_F$, определяемая соотношениями (9.4), обладает также свойствами:

1) $\forall\triangle \in \mathcal{S} \quad \mathbf{P}_F(\triangle) \geqslant 0$;
2) $\mathbf{P}_F(\mathbb{R}^1) = 1$;
3) счетно - аддитивна на $\mathcal{S}$.

Доказательство. Первые два свойства очевидным образом следуют из определения меры, легко видеть также, что выполняется свойство конечной аддитивности. Докажем счетную аддитивность.

Пусть $\triangle \supset \bigcup\limits_{n=1}^{N} \triangle_n, \quad \triangle_i \bigcap \triangle_j \neq \varnothing$. Тогда согласно лемме 9.4

$$\mathbf{P}_F(\triangle) \geqslant \sum_{n=1}^{N} \mathbf{P}_F(\triangle_n). \tag{9.5}$$

Рассмотрим полуинтервал $\triangle = \bigcup\limits_{n=1}^{\infty} \triangle_n$, представляющий объединение попарно не пересекающихся интервалов $\triangle_n = [a_n; b_n)$, тогда для любого N выполняется (9.5). Переходя к пределу при $N \to \infty$,

$$\mathbf{P}_F(\triangle) \geqslant \sum_{n=1}^{\infty} \mathbf{P}_F(\triangle_n). \tag{9.6}$$

Покажем теперь, что справедливо также противоположное неравенство.

Рассмотрим только случай конечного полуинтервала $\triangle = [a; b)$. В остальных случаях рассуждения аналогичны, можете провести их самостоятельно.

Пусть $\lambda, \lambda_1, \dots, \lambda_n > 0$ — пока произвольные достаточно малые числа, для которых существуют определяемые ниже промежутки. Рассмотрим $\triangle' = [a; b-\lambda] \subset [a; b)$, $\quad \triangle'_n = (a_n - \lambda_n; b_n) \supset [a_n; b_n)$. Система открытых интервалов $\triangle'_n$ образует покрытие отрезка $\triangle'$. По лемме Гейне – Бореля из этой системы интервалов можно выделить конечную подсистему $\triangle'_{n_1}, \dots, \triangle'_{n_k}$, также покрывающую $\triangle'$, т.е.

$$\triangle' \subset \bigcup_{j=1}^{k} \triangle'_{n_j}.$$

В силу леммы (9.5)

$$\mathbf{P}_F(\triangle') \leqslant \sum_{j=1}^{k} \mathbf{P}_F(\triangle'_{n_j}) \leqslant \sum_{n=1}^{\infty} \mathbf{P}_F(\triangle'_n).$$

Из определения меры $\mathbf{P}_F$ на полуинтервалах получаем

$$F(b-\lambda) - F(a) \leqslant \sum_{n=1}^{\infty} (F(b_n) - F(a_n - \lambda_n)) =$$

$$= \sum_{n=1}^{\infty} (F(b_n) - F(a_n)) + \sum_{n=1}^{\infty} (F(a_n) - F(a_n - \lambda_n)).$$

Поскольку функция распределения $F(x)$ непрерывна слева во всех точках x, то для $\forall \varepsilon > 0$ и $\forall n$ можно выбрать λ_n таким образом, чтобы

$$0 \leqslant F(a_n) - F(a_n - \lambda_n) < \frac{\varepsilon}{2^n}.$$

Но тогда

$$F(b-\lambda) - F(a) < \sum_{n=1}^{\infty} \mathbf{P}_F(\triangle'_n) + \sum_{n=1}^{\infty} \frac{\varepsilon}{2^n} = \sum_{n=1}^{\infty} \mathbf{P}_F(\triangle'_n) + \varepsilon.$$

Устремив $\varepsilon \to 0+$, получим

$$F(b-\lambda) - F(a) \leqslant \sum_{n=1}^{\infty} \mathbf{P}_F(\triangle_n).$$

Поскольку $\lambda > 0$ – произвольно, то можно также перейти к пределу при $\lambda \to 0+$. Из непрерывности функции распределения слева в каждой точке следует, что

$$\mathbf{P}_F([a;b)) = F(b) - F(a) \leqslant \sum_{n=1}^{\infty} \mathbf{P}_F(\triangle_n) \tag{9.7}$$

Учитывая одновременное выполнение неравенств (9.5), (9.7), получим

$$\mathbf{P}_F(\triangle) = \sum_{n=1}^{\infty} \mathbf{P}_F(\triangle_n),$$

т.е. мера $\mathbf{P}_F$, определенная на полукольце интервалов, счетно - аддитивна, и следовательно, допускает единственное продолжение на σ – алгебру $\mathcal{B}^1$, порожденную этим полукольцом. □

Таким образом доказана следующая

Теорема 9.2. *Пусть $F(x)$ — функция распределения. Тогда существует единственная вероятностная мера $\mathbf{P}_F$, определенная на борелевской σ – алгебре $\mathcal{B}'$, для которой $F(x)$ является ее функцией распределения.*

9.2.2 Вероятностные распределения на плоскости и в пространстве

Пусть теперь $\Omega = \mathbb{R}^2$, σ-алгебра случайных событий — борелевские множества из $\mathcal{B}^2$, $\mathbf{P}$ — вероятностная мера на $\mathcal{B}^2$.

Определение 9.10. Функцией распределения вероятностной меры $\mathbf{P}$ называется функция $F_P(x,y)$ двух действительных переменных x, y, равная

$$F_P(x,y) = \mathbf{P}\left\{(u,v) : u < x,\ v < y\right\}.$$

Из свойств вероятностной меры вытекают следующие свойства функции распределения $F_P(x,y)$:

1) $F_P(x,y)$ не убывает по каждой из переменных;

Рис. 9.3: Распределение вероятностей на плоскости.

2) $F_P(x, -\infty) = F_P(-\infty, y) = 0 \quad \forall x, y$;

3) $F_P(+\infty, +\infty) = 1$,

4) $\forall h, k > 0$, $\forall x, y$

$$F_P(x+h, y+k) - F_P(x+h, y) - F_P(x, y+k) + F_P(x, y) \geqslant 0. \quad (9.8)$$

5) Свойство непрерывности слева по совокупности переменных: $F(x-h, y-k) \to F(x, y)$ при $h, k \to 0+$.

Выражение в левой части (9.8) есть не что иное как вероятность попадания в прямоугольник $\triangle = [x; x+h) \times [y; y+k)$ (см. рис.9.2.2)

Замечание. Это выражение, если его поделить на hk , представляет собой вторую смешанную разность для $F_P(x, y)$. Если перейти к пределу при $h, k \to 0+$, то получим, что должно выполняться неравенство $\frac{\partial^2 F_P}{\partial x \partial y} \geqslant 0$ для тех точек, где эта производная существует.

Определение 9.11. Функция от двух переменных, для которой выполнены свойства 1) —5) на стр. 77, называется двухмерной функцией распределения.

Теорема 9.3. *Для любой функции распределения $F(x, y)$ существует единственная вероятностная мера $\mathbf{P}_F$, определенная на $\mathcal{B}^2$ и такая, что $F(x, y)$ является ее функцией распределения.*

Доказательство аналогично тому, что проводилось в одномерном случае.

Аналогичным образом можно определять вероятностные меры в евклидовых пространствах большей размерности, определяя ее сначала на прямоугольных параллелепипедах, а затем продолжая на борелевские σ– алгебры. Отметим одно важное свойство вероятностных мер на борелевских множествах.

Пусть $(\mathbb{R}^d, \mathcal{B}^d, \mathbf{P})$ — вероятностное пространство, $\mathcal{G}$— класс открытых подмножеств $\mathbb{R}^d$, $\mathcal{F}$ — класс замкнутых множеств. Если $B \in \mathcal{B}^d$,

$$\mathbf{P}(B) = \inf_{B \subset G \in \mathcal{G}} \mathbf{P}(G) = \sup_{B \supset F \in \mathcal{F}} \tag{9.9}$$

Замечание. Вероятностную меру в произвольных метрических пространствах стремятся выбирать таким образом, чтобы выполнялось (9.9).

9.2.3 Два основных типа распределений в евклидовых пространствах

1.**Дискретный тип распределений.**

Определение 9.12. Распределение вероятностей $\mathbf{P}$ на $(\mathbb{R}^1, \mathcal{B}^1)$ называется дискретным, если существует последовательность $x_1, x_2 \ldots$ различных между собой точек, таких, что

$$\mathbf{P}(x_k) = p_k \geqslant 0, \quad \sum_k p_k = 1.$$

В этом случае вся вероятность сосредоточена на счетном или конечном множестве точек. Вероятности других событий определяются как

$$\mathbf{P}(B) = \sum_{x_k \in B} p_k.$$

В качестве примеров распределений дискретного типа могут служить биномиальное и пуассоновское распределения.

2.**Непрерывный тип распределений.**[2] Рассмотрим сначала распределения на прямой. Для непрерывного (более точно — абсолютно непрерывного) распределения функция распределения задается интегралом от некоторой неотрицательной функции $p(x)$, называемой плотностью распределения:

$$F(x) = \int\limits_{-\infty}^{x} p(u)du, \quad \int_{-\infty}^{\infty} p(u)du = 1.$$

[2]Более точное название: абсолютно непрерывные распределения по отношению к мере Лебега.

В известной монографии Г. Крамера ([6])говорится, что в приложениях можно ограничиться случаями, когда плотность непрерывна всюду, кроме конечного числа точек, или когда множество точек разрыва бесконечно, но не имеет конечной предельной точки. Интеграл понимается в этих случаях как интеграл Римана.

Подобным же образом можно определить непрерывный тип распределения в двухмерном случае:

$$F(x, y) = \int\limits_{-\infty}^{x} \int\limits_{-\infty}^{y} p(u, v) du dv,$$

$p(u, v) \geqslant 0, \quad \int\int\limits_{\mathbb{R}^2} p(u, v) du dv = 1$. Функцию $p(x, y)$ также называют плотностью распределения вероятностей.

9.3 Случайные величины

Пусть $(\Omega, \mathcal{A}, \mathbf{P})$ — фиксированное вероятностное пространство.

Определение 9.13. Случайной величиной, заданной на вероятностном пространстве $(\Omega, \mathcal{A}, \mathbf{P})$, называется функция $X(\omega)$ такая, что для любых a и b $(a \leqslant b)$

$$\{\omega : a \leqslant X(\omega) < b\} \in \mathcal{A}. \tag{9.10}$$

Задача 9.6. Покажите, что условие (9.10) в определении случайной величины эквивалентно каждому из следующих условий:

1) $\forall x \in \mathbb{R}^1 \quad \{\omega : X(\omega) < x\} \in \mathcal{A}$;

2) $\forall B \in \mathcal{B}^1$ полный прообраз B при отображении X является элементом $\mathcal{A}$: $\quad \{\omega : X(\omega) \in B\} \in \mathcal{A}$.[3]

В теории функций понятие случайной величины эквивалентно понятию измеримой функции.

Пара $(\Omega, \mathcal{A})$ называется измеримым пространством. Рассмотрим еще одно измеримое пространство $(\mathbb{R}^1, \mathcal{B}^1)$.

Определение 9.14. Функция $X : (\Omega, \mathcal{A}) \to (\mathbb{R}^1, \mathcal{B}^1)$ называется $\mathcal{A}$ - измеримой, если $\forall B \in \mathcal{B}^1 \quad X^{-1}(B) \in \mathcal{A}$.

Теорема 9.4. *Пусть последовательность случайных величин* $X_1, X_2 \ldots$ *определена на* $(\Omega, \mathcal{A}, \mathbf{P})$ *и для* $\forall \omega \quad \exists \lim\limits_{n\to\infty} X_n(\omega) = X(\omega)$. *Тогда* $X(\omega)$ *также является случайной величиной.*

[3]Полным прообразом B при отображении X называют множество $X^{-1}(B) = \{\omega : X(\omega) \in B\}$

Другими словами, класс случайных величин, определенных на фиксированном вероятностном пространстве, замкнут относительно операции предельного перехода.

Доказательство. Покажем сначала, что для $\forall c \in \mathbb{R}$ справедливо равенство

$$\{\omega : X(\omega) < c\} = \bigcup_{k=1}^{\infty} \bigcup_{m=1}^{\infty} \bigcap_{n \geqslant m} \{\omega : X_n(\omega) < c - \frac{1}{k}\}. \tag{9.11}$$

1). Пусть $\omega_0 \in \{\omega : X(\omega) < c\}$. Тогда $X(\omega_0) < c$ и, следовательно, существует такое натуральное число k, что $X(\omega_0) < c - \frac{1}{k}$. Из определения предела вытекает, что найдется номер m, начиная с которого
$\forall n \geqslant m \quad X_n(\omega) < c - \frac{1}{k}$. Это означает, что

$$\omega_0 \in \bigcup_{k=1}^{\infty} \bigcup_{m=1}^{\infty} \bigcap_{n \geqslant m} \{\omega : X_n(\omega) < c - \frac{1}{k}\},$$

т.е.

$$\{\omega : X(\omega) < c\} \subset \bigcup_{k=1}^{\infty} \bigcup_{m=1}^{\infty} \bigcap_{n \geqslant m} \{\omega : X_n(\omega) < c - \frac{1}{k}\}. \tag{9.12}$$

2). Аналогичным образом доказывается обратное вложение

$$\{\omega : X(\omega) < c\} \supset \bigcup_{k=1}^{\infty} \bigcup_{m=1}^{\infty} \bigcap_{n \geqslant m} \{\omega : X_n(\omega) < c - \frac{1}{k}\}. \tag{9.13}$$

Из (9.12),(9.13) вытекает справедливость равенства (9.11). Поскольку из определения случайной величины для $X_n(\omega)$ следует выполнение

$$\{\omega : X_n(\omega) < c - \frac{1}{k}\} \in \mathcal{A}$$

для всех $n, k,$ то событие в правой части (9.11) принадлежит σ – алгебре $\mathcal{A}$, а значит, и $\{\omega : X(\omega) < c\} \in \mathcal{A}$, т.е. предел последовательности случайных величин является случайной величиной. □

Определение 9.15. Случайная величина $X(\omega)$ называется элементарной, если она принимает конечное или счетное число попарно различных значений $x_1, x_2 \ldots$.

Рассмотрим события

$$A_j = \{\omega : X(\omega) = x_j\} \in \mathcal{A}.$$

Элементарную случайную величину $X(\omega)$ можно представить в виде

$$X(\omega) = \sum_j x_j I_{A_j}(\omega). \tag{9.14}$$

Поскольку события $A_1, A_2 \ldots$ образуют разбиение Ω, то в сумме (9.14) для любого элементарного исхода только одно слагаемое отлично от 0, и, следовательно, ряд сходится в каждой точке.

Теорема 9.5. *Функция $X(\omega)$ является случайной величиной $\Leftrightarrow$ существует последовательность элементарных случайных величин $X_1(\omega), X_2(\omega), \ldots,$ равномерно сходящаяся к $X(\omega)$.*

Доказательство. Достаточность следует из предыдущей теоремы, поэтому докажем только необходимость. Доказательство будет конструктивным. Приведем два примера таких последовательностей.

1). Пусть

$$X_n(\omega) = \sum_{k \in \mathbb{Z}} \frac{k}{n} \cdot I_{\{k/n \leqslant X(\omega) < (k+1)/n\}},$$

при этом полагаем, что $I_{\varnothing} = 0$.

2).

$$X_n^*(\omega) = \sum_{l \in \mathbb{Z}} \frac{l}{2^n} I_{\{l/2^n \leqslant X(\omega) < (l+1)/2^n\}}. \tag{9.15}$$

Задача 9.7. Покажите, что каждая из этих последовательностей равномерно сходится к случайной величине $X(\omega)$, причем последовательность X_n^* не убывает.

$\square$

Теорема 9.6. *Пусть $f(x_1, \ldots, x_d)$ — непрерывная функция на $\mathbb{R}^d$ и $X_1(\omega), \ldots, X_d(\omega)$— случайные величины на $(\Omega, \mathcal{A}, \mathbf{P})$.*
Тогда $Y(\omega) = f(X_1(\omega), \ldots, X_d(\omega))$ также является случайной величиной на данном вероятностном пространстве.

Доказательство. Достаточно доказать, что существует последовательность случайных величин, в каждой точке ω сходящаяся к $Y(\omega)$. Мы укажем последовательность элементарных случайных величин, сходящуюся к $Y(\omega)$. Из предыдущей теоремы следует, что для каждой из $X_i(\omega) \quad (1 \leqslant i \leqslant d)$ существует последовательность элементарных случайных величин, которая сходится к ней равномерно на Ω. Пусть при

$n \to \infty$

$$\begin{array}{lcl}
X_{1,n}(\omega) & \text{сходится равномерно к} & X_1(\omega) \\
X_{2,n}(\omega) & \text{сходится равномерно к} & X_2(\omega) \\
\vdots & \dots & \vdots \\
X_{d,n}(\omega) & \text{сходится равномерно к} & X_d(\omega)
\end{array}$$

Покажем, что последовательность элементарных случайных величин

$$Y_n(\omega) = f(X_{1,n}(\omega), \dots, X_{d,n}(\omega))$$

сходится к $Y(\omega)$. Выберем произвольное малое $\delta > 0$.

Поскольку равномерно сходящихся последовательностей $\{X_{in}(\omega)\}$ конечное число, то для $\forall \delta > 0$ существует такое $N = N(\delta)$, что одновременно для всех i и ω выполняются неравенства

$$|X_{i,n}(\omega) - X_i(\omega)| < \frac{\delta}{\sqrt{d}},$$

если $n \geqslant N$. Для таких n

$$\sqrt{(X_{1,n}(\omega) - X_1(\omega))^2 + \dots (X_{d,n}(\omega) - X_d(\omega))^2} < \delta$$

Рассмотрим фиксированное $\omega_0 \in \Omega$. Функция $f(x_1, \dots, x_d)$ непрерывна в точке $\mathbf{x_0} = (X_1(\omega_0), \dots X_d(\omega_0))$. Следовательно, для $\forall \varepsilon > 0$
$\exists \delta_0 = \delta(\varepsilon, \omega_0) > 0$ такое, что если

$$\sqrt{(X_{1,n}(\omega_0) - X_1(\omega_0))^2 + \dots (X_{d,n}(\omega_0) - X_d(\omega_0))^2} < \delta_0, \tag{9.16}$$

то

$$|f(X_1(\omega_0), \dots, X_d(\omega_0)) - f(X_{1,n}(\omega_0), \dots, X_{d,n}(\omega_0))| < \varepsilon.$$

Выберем теперь $N_0 = N(\delta_0)$, тогда при $n \geqslant N_0$, выполняется (9.16), а значит,

$$|Y_n(\omega_0) - Y(\omega_0)| < \varepsilon,$$

но это и означает, что последовательность $Y_n(\omega_0)$ сходится к $Y(\omega_0)$. Поскольку сходимость выполняется для $\forall \omega_0$, то в силу теоремы 9.4
$Y(\omega) = f(X_1(\omega), \dots, X_d(\omega))$ является случайной величиной. $\square$

Замечание. Заметим, что последовательность $Y_n(\omega)$ не обязательно сходится к $Y(\omega)$ равномерно. Приведите пример, когда это не так.

Как следствие из этой теоремы получаем, что сумма, разность, произведение и другие непрерывные функции от случайных величин являются случайными величинами. Предельным переходом, используя теорему 9.4, можно класс функций $f(x_1, \dots, x_d)$ значительно расширить.

Задача 9.8. Покажите, что $Y(\omega) = f(X_1, \dots, X_d)$ является случайной величиной для любой борелевской функции от d переменных.

9.3.1 σ- алгебра, порожденная случайной величиной

Пусть $X(\omega)$ — случайная величина на $(\Omega, \mathcal{A}, \mathbf{P})$. Остановимся подробнее на понятии полного прообраза некоторого борелевского множества B при отображении X.

Пример 9.7. Пусть $\Omega = \mathbb{R}^2, \quad \omega = (x_1, x_2), \quad X(\omega) = \sqrt{x_1^2 + x_2^2}$. Тогда полными прообразами борелевских множеств

$$B_1 = (-\infty; b);\ B_2 = [a; b),\, B_3 = \{b\}\, (0 < a < b)$$

будут соответственно круг, кольцо, окружность.

Пример 9.8. По – прежнему элементарные исходы $\omega = (x_1, x_2)$ — точки координатной плоскости, $X(\omega) = x_1$, тогда полными прообразами множеств из предыдущего примера служат полуплоскость, полоса, прямая.

Задача 9.9. Докажите справедливость следующих утверждений:
1)$X^{-1}(B_1 \bigcup B_2) = X^{-1}(B_1) \bigcup X^{-1}(B_2)$;
2)$X^{-1}(B_1 \bigcap B_2) = X_{-1}(B_1) \bigcap X_{-1}(B_2)$;
3)$X^{-1}(\overline{B}) = \overline{X^{-1}(B)}$.

Из этих соотношений следует, что класс подмножеств Ω, являющихся полными прообразами борелевских множеств при отображении $X(\omega)$, образует σ-алгебру. Будем обозначать ее $\mathcal{A}_X$.

σ-алгебра $\mathcal{A}_X \subseteq \mathcal{A}$ и называется σ-алгеброй, порожденной случайной величиной X.

Задача 9.10. Выясните, какие σ-алгебры порождаются случайными величинами, принимающими конечное число значений, в том числе константами.

9.3.2 Распределения случайных величин

С каждой случайной величиной можно связать вероятностное пространство $(\mathbb{R}, \mathcal{B}, \mathbf{P}_F)$, где вероятностная мера $\mathbf{P}_F$— это мера, функция распределения которой равна

$$F(x) = \mathbf{P}\{\omega :\ X(\omega) < x\}. \tag{9.17}$$

Задача 9.11. Покажите, что функция, определенная в (9.17), действительно является функцией распределения.

Определение 9.16. Функция $F(x)$, определенная равенством (9.17), называется функцией распределения случайной величины $X(\omega)$.

В нашем курсе мы будем иметь дело только с двумя основными типами распределений: дискретным и абсолютно непрерывным (см. раздел 9.2.3). Однако заметим, что в общем случае функция распределения — это смесь функций распределений разных типов – дискретного, абсолютно непрерывного, сингулярного. Прочитать об этом можно, например, в учебнике по теории функций ([5]).

О свойствах наиболее часто встречающихся распределений можно посмотреть в приложении 1.

9.4 Математические ожидания случайных величин (Общий случай)

Пусть $(\Omega, \mathcal{A}, \mathbf{P})$ — некоторое вероятностное пространство, $X(\omega)$ — элементарная случайная величина, принимающая значения $x_1, x_2 \ldots$ $(x_i \neq x_j \quad \forall i \neq j)$. Обозначим $A_k = \{\omega : X(\omega) = x_k\}$. Тогда $A_k \in \mathcal{A}$, образуют разбиение Ω и

$$X(\omega) = \sum_k x_k I_{A_k}(\omega).$$

Определение 9.17. Математическим ожиданием элементарной случайной величины $X(\omega)$ называется сумма

$$\mathbf{E}\, X = \sum_k x_k\, \mathbf{P}(A_k), \tag{9.18}$$

причем, для случайных величин, имеющих счетное число значений, ряд в (9.18) должен сходиться абсолютно.

Замечание. Требование абсолютной сходимости в определении математического ожидания существенно. Напомним, что сумма абсолютно сходящегося ряда не зависит от порядка суммирования, тогда как условно сходящиеся ряды таким свойством не обладают. Поэтому, если в определении математического ожидания не потребовать абсолютную сходимость, то эта числовая характеристика может оказаться зависещей от того, каким образом мы перенумеровали значения случайной величины.

Далее мы докажем лемму, аналогичную лемме 3.1.

Лемма 9.6. Пусть $B_1, B_2 \ldots$— конечное или счетное разбиение Ω, $B_l \in \mathcal{A}$ и случайная величина $X(\omega)$ постоянна на каждом из B_l, т.е. ее можно представить в виде

$$X(\omega) = \sum_l x_l' \cdot I_{B_l}(\omega).$$

Тогда
$$\mathbf{E}\, X(\omega) = \sum_l x'_l\, \mathbf{P}(B_l), \tag{9.19}$$
если этот ряд сходится абсолютно.

Замечание. В определении математического ожидания (9.18) все значения случайной величины различны между собой, в (9.19) могут встречаться одинаковые, т.е. разбиение $B_1, B_2 \ldots$ более "мелкое", элементы первого разбиения можно представить в виде объединения событий B_l.

Доказательство. 1). Пусть ряд (9.19) сходится абсолютно. Тогда его сумма не зависит от порядка суммирования и ряд можно переписать в другом виде:

$$\sum_l x'_l\, \mathbf{P}(B_l) = \sum_k \sum_{l:\, x'_l = x_k} x'_l\, \mathbf{P}(B_l) = \sum_k x_k \sum_{l:\, x'_l = x_k} \mathbf{P}(B_l) = \sum_k x_k\, \mathbf{P}(A_k).$$

Замечание. Поскольку с абсолютно сходящимися рядами можно поступать также, как с конечными суммами, то при доказательстве безразлично, с какой суммой (конечной или нет) мы имеем дело.

2). Пусть математическое ожидание случайной величины $X(\omega)$ существует, тогда его можно переписать в виде

$$\mathbf{E}\, X = \sum_k x_k\, \mathbf{P}(A_k) = \sum_k x_k \sum_{l:\, x_k = x'_l} \mathbf{P}(B_l) = \sum_l x'_l\, \mathbf{P}(B_l).$$

□

9.4.1 Основные свойства математического ожидания

Сформулируем основные свойства математического ожидания

1). Математические ожидания $\mathbf{E}\, X$ и $\mathbf{E}\, |X|$ существуют или не существуют одновременно и $|\,\mathbf{E}\, X| \leqslant \mathbf{E}\, |X|$.

2). Если $X \geqslant 0$, то $\mathbf{E}\, X \geqslant 0$, причем в этом случае
$$\mathbf{E}\, X = 0 \Leftrightarrow \mathbf{P}\{\omega :\ X(\omega) = 0\} = 1.$$

3). Если существует $\mathbf{E}\, X$, то для любой константы c существует
$$\mathbf{E}(cX) = c \cdot \mathbf{E}\, X.$$

4). Если существуют $\mathbf{E}\,X$, $\mathbf{E}\,Y$, то существует математическое ожидание суммы этих случайных величин и

$$\mathbf{E}(X+Y)=\mathbf{E}\,X+\mathbf{E}\,Y.$$

5). Пусть $Y, X, X_1, X_2\dots$ — случайные величины такие, что $|X_n(\omega)| \leqslant Y(\omega)$, $X_n(\omega)\to X(\omega)$, существует $\mathbf{E}\,Y$. Тогда существует $\mathbf{E}\,X$ и

$$\lim_{n\to\infty}\mathbf{E}\,X_n=\mathbf{E}\,X.$$

Последнее утверждение носит название теоремы Лебега о мажорируемой сходимости.

Замечание. Математическое ожидание можно рассматривать как функцию, определенную на множестве случайных величин, заданных на некотором вероятностном пространстве $(\Omega,\mathcal{A},\mathbf{P})$. Такого рода функции называются функционалами. Данный функционал обладает свойствами линейности 3), 4), свойством неотрицательности 2). Аналогичными свойствами обладает интеграл.

Покажем справедливость свойства 4).

Доказательство. Пусть

$$X=\sum_k x_k I_{A_k}(\omega),\quad Y(\omega)=\sum_l y_l I_{B_l}(\omega),$$

где $x_1, x_2\dots,\quad y_1, y_2\dots$ — все различные между собой значения случайных величин X и Y. Рассмотрим события $C_{kl}=A_k\bigcap B_l$, на которых случайная величина $(X+Y)(\omega)$ постоянна и равна x_k+y_l, но тогда

$$X+Y=\sum_{k,l}(x_k+y_l)I_{C_{kl}}.$$

Следовательно, ее математическое ожидание равно сумме ряда

$$\mathbf{E}(X+Y)=\sum_{k,l}(x_k+y_l)\,\mathbf{P}(C_{kl}),$$

если этот ряд сходится абсолютно. С другой стороны, в силу утверждения леммы

$$\mathbf{E}\,X=\sum_{k,l}x_k\,\mathbf{P}(C_{kl})\quad \mathbf{E}\,Y=\sum_{k,l}y_l(C_{kl}),$$

начит,

$$\mathbf{E}\,X+\mathbf{E}\,Y=\sum_{k,l}(x_k+y_l)\,\mathbf{P}(C_{kl}).$$

Из абсолютной сходимости первых двух рядов вытекает абсолютная сходимость последнего ряда.

□

Доказательство свойства 5) опирается на глубокие факты сходимостей случайных величин, приводить которые в данном курсе мы не будем.

Введем обозначение $L_1 = L_1(\Omega, \mathcal{A}, \mathbf{P})$ для класса случайных величин, определенных на вероятностном пространстве $(\Omega, \mathcal{A}, \mathbf{P})$ и таких, что существуют элементарная случайная величина $X'(\omega)$ с конечным математическим ожиданием и константа k, для которой при всех ω

$$|X(\omega) - X'(\omega)| \leqslant k.$$

В частности, в класс L_1 входят все элементарные случайные величины, имеющие конечное математическое ожидание, все ограниченные случайные величины (в этом случае можно взять $X'(\omega) \equiv 0$). Так же как в теореме о продолжении меры можно доказать, что функционал $\mathbf{E}\,X$, определенный для элементарных случайных величин и обладающий свойствами 1)–5), может быть продолжен на L_1 .

Теорема 9.7. *На классе случайных величин $L_1(\Omega, \mathcal{A}, \mathbf{P})$ можно и при том единственным образом определить математическое ожидание $\mathbf{E}\,X$, так, что для элементарных случайных величин этот функционал совпадает с определенным ранее и для него выполнены свойства 1)–5).*

Замечание. Обычно в учебниках описывают конструкцию, как строится этот функционал с той или иной степенью строгости. Наиболее полное изложение содержится в учебниках ([10],[8]). Мы также приведем этапы построения этого продолжения.

1). Определение и доказательство свойств математического ожидания для элементарных случайных величин.

2). Согласно теореме 9.5 для любой случайной величины $X(\omega)$ существует неубывающая последовательность элементарных случайных величин, равномерно сходящаяся к $X(\omega)$. В качестве таковой можно к примеру взять $X_n^*(\omega)$ определенную в (9.15):

$$X_n^*(\omega) = \sum_{l \in \mathbb{Z}} \frac{l}{2^n} I_{\{l/2^n \leqslant X(\omega) < (l+1)/2^n\}}.$$

3). Поскольку $X_n^*(\omega) \leqslant X(\omega)$, то для того, чтобы выполнялось свойство 5) математического ожидания, необходимо определить $\mathbf{E}\,X(\omega)$ следующим образом:

$$\mathbf{E}\,X = \lim_{n\to\infty} \mathbf{E}\,X_n^*(\omega).$$

4). Доказательство свойств математического ожидания для общего случая.

Из теорем о математическом ожидании приведем следующие.

Теорема 9.8. *Если существует* $\mathbf{E}\,Y(\omega)$ *и для всех* ω *выполняется неравенство* $|X(\omega)| \leqslant Y(\omega)$, *то существует и математическое ожидание случайной величины* $X(\omega)$ *и*

$$\mathbf{E}\,|X(\omega)| \leqslant \mathbf{E}\,Y(\omega).$$

Доказательство. Доказательство проведем для элементарных случайных величин. Поскольку

$$\sum_{k,l} |x_k|\,\mathbf{P}\{X(\omega) = x_k, Y(\omega) = y_l\} \leqslant \sum_{k,l} y_l\,\mathbf{P}\{X(\omega) = x_k, Y(\omega) = y_l\} =$$

$$= \sum_{l} y_l\,\mathbf{P}\{y(\omega) = y_l\} = \mathbf{E}\,Y(\omega),$$

то ряд в левой части неравенства сходится, причем абсолютно, можно изменять порядок суммирования. Но тогда

$$\sum_{k,l} |x_k|\,\mathbf{P}\{X(\omega) = x_k, Y(\omega) = y_l\} = \sum_{k} |x_k|\,\mathbf{P}\{X(\omega) = x_k\} = \mathbf{E}\,|X(\omega)|.$$

□

Схема построения математического ожидания и свойства этого линейного функционала совпадают с построением и свойствами интеграла Лебега. Поэтому часто математическое ожидание записывают как интеграл Лебега:

$$\mathbf{E}\,X = \int\limits_{\Omega} X(\omega) P(d\omega).$$

Если Ω – числовая прямая (плоскость,евклидово пространство $\mathbb{R}^d$), то этот интеграл может превратиться в обычный интеграл Римана. Так если распределение случайной величины $X(\omega)$ абсолютно непрерывно и $p(x)$ – плотность этого распределения, то

$$\mathbf{E}\,X = \int\limits_{-\infty}^{+\infty} x p(x) dx, \tag{9.20}$$

при условии, что интеграл сходится абсолютно.

Доказательство.

$$\mathbf{E}\,X = \lim_{n\to\infty}\mathbf{E}\,X_n(\omega) = \lim_{n\to\infty}\sum_{k=-\infty}^{\infty}\frac{k}{n}\int\limits_{\frac{k}{n}}^{\frac{k+1}{n}} p(x)dx =$$

$$= \int\limits_{-\infty}^{\infty} xp(x)dx - \left(\sum_{k=-\infty}^{\infty}(\int\limits_{\frac{k}{n}}^{\frac{k+1}{n}} xp(x)dx - \int\limits_{\frac{k}{n}}^{\frac{k+1}{n}}\frac{k}{n}p(x)dx)\right) =$$

$$= \int\limits_{-\infty}^{\infty} xp(x)dx - \sum_{k=-\infty}^{\infty}\int\limits_{\frac{k}{n}}^{\frac{k+1}{n}}(x-\frac{k}{n})p(x)dx.$$

Откуда получаем, что

$$|\,\mathbf{E}\,X - \int\limits_{-\infty}^{\infty} xp(x)dx| \leqslant \frac{1}{n}\sum_{k=-\infty}^{\infty}\int\limits_{\frac{k}{n}}^{\frac{k+1}{n}} p(x)dx = \frac{1}{n}.$$

Устремив $n\to\infty$, получим равенство (9.20). □

Аналогичным образом для произвольной измеримой функции $f(x)$

$$\mathbf{E}\,f(X) = \int\limits_{-\infty}^{+\infty} f(x)p(x)dx,$$

если интеграл сходится абсолютно.

9.5 Независимые случайные величины

Определение 9.18. Случайные величины $X_1,\ldots,X_n$ называются **независимыми (взаимно независимыми)**, если для любых полуинтервалов $\triangle_1=[a_1,b_1),\triangle_2=[a_2,b_2),\ldots\triangle_n=[a_n,b_n)$ выполняется равенство

$$\mathbf{P}\{\omega:\,X_1(\omega)\in\triangle_1, X_2(\omega)\in\triangle_2,\ldots,X_n(\omega)\in\triangle_n\} = \tag{9.21}$$

$$= \mathbf{P}\{\omega:\,X_1(\omega)\in\triangle_1\}\,\mathbf{P}\{\omega:\,X_2(\omega)\in\triangle_2\}\cdots\mathbf{P}\{\omega:\,X_n(\omega)\in\triangle_n\}.$$

Определение независимости можно сформулировать и в терминах функций распределений, а для случайных величин с абсолютно непрерывными распределениями — в терминах плотностей.

Определение 9.19. Случайные величины $X_1, \ldots, X_n$ называются **независимыми,** если совместная функция распределения этих величин во всех точках равна произведению частных функций распределения, т.е.

$$F(x_1, \ldots, x_n) = F_1(x_1)F_2(x_2)\cdots F_n(x_n),$$

где

$$F(x_1, \ldots, x_n) = \mathbf{P}\{\omega : X_1(\omega) < x_1, \ldots, X_n(\omega) < x_n\},$$
$$F_i(x_i) = \mathbf{P}\{\omega : X_i(\omega) < x_i\}.$$

Задача 9.12. Докажите эквивалентность этих определений независимости случайных величин.

Пусть совместное распределение случайных величин X и Y абсолютно непрерывно, т.е. существует такая неотрицательная функция $p(x, y)$, что для $\forall B \in \mathcal{B}^2$

$$\mathbf{P}\{(X, Y) \in B\} = \iint\limits_B p(x, y) dx\, dy.$$

Функция $p(x, y)$ называется совместной плотностью распределения, с функцией распределения $F(x, y) \quad \forall x, y$ она связана соотношением

$$F(x, y) = \int\limits_{-\infty}^{x}\int\limits_{-\infty}^{y} p(u, v) du\, dv.$$

Задача 9.13. Докажите, что абсолютно непрерывные случайные величины X и Y независимы $\Longleftrightarrow$ почти всюду по мере Лебега

$$p(x, y) = p_X(x)p_Y(y),$$

где $p_X(x)$, $p_Y(y)$— частные плотности распределений.

Для независимых случайных величин сохраняются все утверждения глав 4 – 5, доказанные в предположении, что X и Y принимают конечное число значений. В частности, борелевские функции от независимых случайных величин также являются независимыми случайными величинами.

9.6 Мультипликативное свойство математического ожидания

Теорема 9.9. *Пусть X и Y – независимые случайные величины, для которых существуют математические ожидания. Тогда существует* $\mathbf{E}(X \cdot Y)$ *и*

$$\mathbf{E}\, XY = \mathbf{E}\, X\, \mathbf{E}\, Y.$$

Доказательство. При доказательстве этого свойства продемонстрируем прием перехода от элементарных случайных величин к произвольным.

1). Пусть X и Y — элементарные случайные величины:

$$X(\omega) = \sum_k x_k I_{A_k}(\omega), \quad Y(\omega) = \sum_l y_l I_{B_l}(\omega),$$

причем в этих представлениях $x_i \neq x_j$, $y_i \neq y_j$, если $i \neq j$,

$$A_k = \{\omega : \, X(\omega) = x_k\}, \quad B_l = \{\omega : \, Y(\omega) = y_l\}.$$

Обозначим $C_{kl} = A_k \bigcap B_l$. Тогда случайная величина $X(\omega)Y(\omega)$ постоянна на C_{kl} и

$$X(\omega)Y(\omega) = \sum_{k,l} x_k y_l I_{C_{kl}}.$$

По свойству математического ожидания, сформулированному в лемме 9.6

$$\mathbf{E}\, X(\omega)Y(\omega) = \sum_{k,l} x_k y_l\, \mathbf{P}(C_{kl}), \tag{9.22}$$

если ряд (9.22) сходится абсолютно. Покажем это. По условию теоремы существуют

$$\mathbf{E}\, X(\omega) = \sum_k x_k\, \mathbf{P}(A_k), \quad \mathbf{E}\, Y(\omega) = \sum_l y_l\, \mathbf{P}(B_l). \tag{9.23}$$

Абсолютно сходящиеся ряды можно перемножать и менять произвольным образом порядки суммирования, при этом будем получать также абсолютно сходящиеся ряды. Поэтому

$$\mathbf{E}\, X\, \mathbf{E}\, Y = \sum_k x_k\, \mathbf{P}(A_k) \sum_l y_l\, \mathbf{P}(B_l) = \sum_{k,l} x_k y_l\, \mathbf{P}(A_k)\, \mathbf{P}(B_l) =$$

$$= \sum_{k,l} x_k y_l\, \mathbf{P}(A_k \bigcap B_l) = \sum_{k,l} x_k y_l\, \mathbf{P}(C_{kl}).$$

Сравнивая последнее выражение с (9.23), видим, что математическое ожидание произведения случайных величин существует и
$\mathbf{E}(XY) = \mathbf{E}\,X\,\mathbf{E}\,Y$, что доказывает утверждение теоремы для элементарных случайных величин.

2). Пусть теперь $X(\omega)$ и $Y(\omega)$ — произвольные случайные величины, имеющие математические ожидания. Рассмотрим последовательности элементарных случайных величин

$$X_n^*(\omega) = \sum_k \frac{k}{n} I_{\{\frac{k}{n} \leqslant X(\omega) < \frac{k+1}{n}\}}(\omega), \quad \mathbf{E}\,X(\omega) = \lim_{n\to\infty} \mathbf{E}\,X_n^*$$

$$Y_n^*(\omega) = \sum_l \frac{l}{n} I_{\{\frac{l}{n} \leqslant Y(\omega) < \frac{l+1}{n}\}}(\omega), \quad \mathbf{E}\,Y(\omega) = \lim_{n\to\infty} \mathbf{E}\,Y_n^*.$$

По построению $X_n^*(\omega) \rightrightarrows X(\omega) \quad Y_n^*(\omega) \rightrightarrows Y(\omega)$ и для $\forall\omega$

$$0 \leqslant X(\omega) - X_n^*(\omega) \leqslant \frac{1}{n}, \quad 0 \leqslant Y(\omega) - Y_n^*(\omega) \leqslant \frac{1}{n},$$

откуда следует, что

$$|X_n^*(\omega)| \leqslant |X(\omega)| + 1, \quad |Y_n^*(\omega)| \leqslant |Y(\omega)| + 1.$$

Из конечности математических ожиданий для $X(\omega)$ и $Y(\omega)$ получаем, что для $\forall n$ существуют $\mathbf{E}\,X_n^*(\omega)$, $\mathbf{E}\,Y_n^*(\omega)$. Поскольку элементарные случайные величины $X_n^*(\omega)$, $Y_n^*(\omega)$ являются независимыми, то в силу доказанного выше имеем

$$\mathbf{E}(X_n^*(\omega)Y_n^*(\omega)) = \mathbf{E}\,X_n^*(\omega)\,\mathbf{E}\,Y_n^*(\omega).$$

Осталось только перейти к пределу при $n \to \infty$. Выпишем несколько простых соотношений:

$$XY - X_n^*Y_n^* = (X - X_n^*)Y + X_n^*(Y - Y_n^*),$$

$$|XY - X_n^*Y_n^*| \leqslant |X - X_n^*| \cdot |Y| + |X_n^*| \cdot |Y - Y_n^*| \leqslant \frac{1}{n}(|X| + |Y| + 1).$$

Но тогда

$$|\,\mathbf{E}(XY) - \mathbf{E}(X_n^*Y_n^*)| \leqslant \mathbf{E}\,|XY - X_n^*Y_n^*| \leqslant \frac{1}{n}\,\mathbf{E}(|X| + |Y| + 1) \to 0$$

при $n \to \infty$, значит,

$$\mathbf{E}(XY) = \lim_{n\to\infty} \mathbf{E}(X_n^*Y_n^*) = \lim_{n\to\infty} \mathbf{E}\,X_n^* \cdot \lim_{n\to\infty} \mathbf{E}\,Y_n^* = \mathbf{E}\,X \cdot \mathbf{E}\,Y.$$

□

Замечание. В других учебниках (см.[10],[8]) схема построения может быть несколько иной: сначала определяется математическое ожидание простых случайных величин, затем для неотрицательных, затем уже общий случай.

Глава 10

Усиленный закон больших чисел

10.1 Лемма Бореля – Кантелли

Пусть $(\Omega, \mathcal{A}, \mathbf{P})$ – некоторое вероятностное пространство и
$A_1, A_2 \ldots \in \mathcal{A}$ – последовательность случайных событий,
$X_1(\omega), X_2(\omega), X_3(\omega), \ldots$ – последовательность случайных величин, определенных на этом вероятностном пространстве.

Определение 10.1. События $A_1, A_2 \ldots$ независимы, если $\forall m \in \mathbb{N}$, $(m \geqslant 2)$ для любых наборов натуральных чисел $j_1 < j_2 < \ldots < j_m$ независимы события $A_{j_1}, A_{j_2}, \ldots, A_{j_m}$.

Определение 10.2. Случайные величины $X_1(\omega), X_2(\omega), X_3(\omega), \ldots$ называются независимыми, если для $\forall m \in \mathbb{N}$, $(m \geqslant 2)$ для любых наборов натуральных чисел $j_1 < j_2 < \ldots < j_m$ независимы случайные величины $X_{j_1}(\omega), X_{j_2}(\omega), \ldots, X_{j_m}(\omega)$.

Заметим, что независимость последовательности случайных событий эквивалентна независимости последовательности $I_{A_1}(\omega), I_{A_2}(\omega), I_{A_3}(\omega), ...$ – индикаторов этих событий.

Введем следующие события:

$$C = \{\omega : \sum_{k=1}^{\infty} I_{A_k}(\omega) < \infty\}$$

– происходит лишь конечное число событий A_k,

$$D = \{\omega : \sum_{k=1}^{\infty} I_{A_k}(\omega) = \infty\}$$

– одновременно происходит бесконечно много событий A_k.

Задача 10.1. Покажите, что

$$C = \bigcup_{k=1}^{\infty} \bigcap_{l=k}^{\infty} \overline{A_l}, \quad D = \bigcap_{k=1}^{\infty} \bigcup_{l=k}^{\infty} A_l, \quad C = \overline{D}. \tag{10.1}$$

Лемма 10.1. (Бореля – Кантелли)

1. Если последовательность событий $A_1, A_2 \ldots$ такова, что ряд $\sum\limits_{n=1}^{\infty} \mathbf{P}(A_n) < \infty$ – сходится, то $\mathbf{P}(D) = 0, \quad \mathbf{P}(C) = 1$.

2. Если для последовательности независимых событий $A_1, A_2 \ldots$ ряд $\sum\limits_{n=1}^{\infty} \mathbf{P}(A_n) = \infty$ – расходится, то $\mathbf{P}(D) = 1, \quad \mathbf{P}(C) = 0$.

Доказательство. 1). Оценим вероятность события D сверху, для этого воспользуемся представлением (10.1):

$$\mathbf{P}(D) \leqslant \mathbf{P}\left\{\bigcup_{l=k}^{\infty} A_l\right\} \leqslant \sum_{l=k}^{\infty} \mathbf{P}(A_l), \tag{10.2}$$

— это неравенство выполняется для любого k. Из сходимости ряда $\sum\limits_{n=1}^{\infty} \mathbf{P}(A_n)$ следует, что правую часть неравенства (10.2) можно сделать сколь - угодно малой, выбрав должным образом k. Следовательно,

$$\mathbf{P}(D) = 0, \quad \mathbf{P}(C) = 1 - \mathbf{P}(D) = 1.$$

2). Воспользовавшись представлением (10.1) для события C, получаем:

$$\mathbf{P}(C) \leqslant \sum_{k=1}^{\infty} \mathbf{P}\left\{\bigcap_{l=k}^{\infty} \overline{A_l}\right\}.$$

Покажем, что каждое слагаемое в этой сумме равно 0. Действительно, в силу независимости событий A_l и справедливости почти очевидного неравенства: $1 - z \leqslant e^{-z}$, имеем

$$\mathbf{P}\left\{\bigcap_{l=k}^{\infty} \overline{A_l}\right\} \leqslant \mathbf{P}\left\{\bigcap_{l=k}^{k+m} \overline{A_l}\right\} = \prod_{l=k}^{k+m} (1 - \mathbf{P}(A_l)) \leqslant e^{-\sum\limits_{l=k}^{k+m} \mathbf{P}(A_l)}.$$

Поскольку ряд $\sum\limits_{n=1}^{\infty} \mathbf{P}(A_n) = \infty$ – расходится, то, устремив $m \to \infty$, получим

$$\mathbf{P}\left\{\bigcap_{l=k}^{\infty} \overline{A_l}\right\} = 0$$

для всех k и, следовательно, $\mathbf{P}(C) = 0$. $\square$

10.2 Сходимость с вероятностью 1

Лемма Бореля – Кантелли является удобным инструментом для установления сходимости последовательностей случайных величин с вероятностью 1.[1]

Определение 10.3. Последовательность случайных величин $X_1(\omega), X_2(\omega), X_3(\omega), \ldots$ сходится к 0 с вероятностью 1, если

$$\mathbf{P}\{\omega : \lim_{n\to\infty} X_n(\omega) = 0\} = 1. \tag{10.3}$$

Равенство (10.3) означает, что сходимость к 0 имеет место для всех элементарных исходов, за исключением, быть может, точек, общая вероятность которых равна 0. Поэтому условие (10.3) можно переписать иначе:

$$\mathbf{P}\{\omega : X_n(\omega) \nrightarrow 0\} = 0. \tag{10.4}$$

Сходимость с вероятностью 1 называют также сходимостью почти всюду или сходимостью почти наверное и обозначают

$$X_n(\omega) \xrightarrow{\text{п.в.}} 0 \quad \text{или} \quad X_n(\omega) \xrightarrow{\text{п.н.}} 0.$$

Задача 10.2. Докажите, что для независимых случайных величин $X_1, X_2 \ldots$

$$X_n(\omega) \xrightarrow{\text{п.н.}} 0 \Leftrightarrow \forall\varepsilon > 0 \quad \sum_{n}^{\infty} \mathbf{P}\{\omega : |X_n(\omega)| > \varepsilon\} < \infty. \tag{10.5}$$

Определение 10.4. Последовательность случайных величин $X_1(\omega), X_2(\omega), X_3(\omega), \ldots$ сходится к случайной величине $X(\omega)$ с вероятностью 1 ($X_n(\omega) \xrightarrow{\text{п.н.}} X(\omega)$), если $X_n(\omega) - X(\omega) \xrightarrow{\text{п.н.}} 0$.

Приведем без доказательства несколько утверждений о сходимости почти наверное. Желающие могут посмотреть доказательство, например, в учебнике А.Н.Ширяева или провести его самостоятельно.

Теорема 10.1.

$$X_n(\omega) \xrightarrow{n.н.} 0 \Leftrightarrow \lim_{n\to\infty} \mathbf{P}\{\omega : \sup_{k\geqslant n} |X_k(\omega)| \geqslant \varepsilon\} = 0.$$

[1] О различных видах сходимости для последовательностей случайных величин и соотношениях между ними можно прочитать в учебнике А.Н.Ширяева "Вероятность", глава 2, §10.

Теорема 10.2. *Последовательность случайных величин* $X_1(\omega), X_2(\omega), X_3(\omega), \ldots$ *сходится с вероятностью 1 (фундаментальна с вероятностью 1)* $\Leftrightarrow \forall \varepsilon > 0$

$$\mathbf{P}\{\sup_{k,l \geqslant n} |X_k(\omega) - X_l(\omega)| \geqslant \varepsilon\} \to 0$$

при $n \to \infty$.

10.3 Усиленный закон больших чисел

Теорема 10.3. *(А. Н. Колмогоров) Пусть* $X_1(\omega), X_2(\omega), X_3(\omega), \ldots$ *– последовательность независимых случайных величин, для которых существуют математические ожидания и дисперсии, причем* $\mathbf{E}\,X_k = m_k$, $\mathbf{D}\,X_k = \sigma_k^2$, *и сходится ряд* $\sum\limits_{k=1}^{\infty} \frac{\sigma_k^2}{k^2} < \infty$. *Тогда*

$$\mathbf{P}\left\{\omega : \frac{X_1(\omega) + X_2(\omega) + \ldots + X_n(\omega)}{n} - \frac{m_1 + m_2 + \ldots + m_n}{n} \to 0\right\} = 1 \tag{10.6}$$

Мы будем говорить, что для последовательности случайных величин выполняется **усиленный закон больших чисел,** если выполняется (10.6), другими словами, если

$$\frac{X_1(\omega) + X_2(\omega) + \ldots + X_n(\omega)}{n} - \frac{m_1 + m_2 + \ldots + m_n}{n} \xrightarrow{\text{п.н.}} 0. \tag{10.7}$$

Доказательство. Рассмотрим случайные величины $X'_k(\omega) = X'_k(\omega) - m_k$. Для них $\mathbf{E}\,X'_k = 0$, $\mathbf{D}\,X'_k = \sigma_k^2$, а выполнение утверждения теоремы (10.6) или (10.7) означает одновременное выполнение усиленного закона больших чисел и для новых случайных величин. Поэтому в дальнейшем при доказательстве теоремы будем считать, что $m_k = 0$.

Обозначим $S_n = X_1 + \ldots + X_n$ и перепишем S_n в следующем виде, учитывая, что всегда найдется такое m, что $2^m \leqslant n < 2^{m+1}$:

$$S_n = X_1 + (X_2 + X_3) + (X_4 + X_5 + X_6 + X_7) + \ldots + (X_{2^m} + \ldots + X_n) =$$

$$= \sum_{i=0}^{m} \left(\sum_{k=2^i}^{2^{i+1}-1} X_k \right) + \sum_{k=2^m}^{n} X_k.$$

Тогда

$$|S_n| \leqslant \sum_{i=0}^{m} \max_{2^i \leqslant j < 2^{i+1}} \left| \sum_{k=2^i}^{j} X_k \right| + \max_{2^m \leqslant j \leqslant n} \left| \sum_{k=2^m}^{j} X_k \right|.$$

Введем обозначения для случайных величин

$$\eta_i = \frac{1}{2^i}\left(\max_{2^i \leqslant j < 2^{i+1}} \left|\sum_{k=2^i}^{j} X_k\right|\right), \tag{10.8}$$

тогда

$$|S_n| \leqslant \eta_0 + 2\eta_1 + 2^2\eta_2 + \ldots + 2^m\eta_m,$$

и значит,

$$\frac{|S_n|}{n} \leqslant \frac{\eta_0 + 2\eta_1 + \ldots + 2^m\eta_m}{2^m}. \tag{10.9}$$

Лемма 10.2. Пусть $\{a_n\}_0^\infty, \{b_n\}_0^\infty$ – две последовательности чисел, связанных соотношением

$$b_n = \frac{a_0 + 2a_1 + \ldots + 2^n a_n}{2^n}.$$

Тогда

$$a_n \to 0 \Leftrightarrow b_n \to 0,$$

т. е. последовательности чисел a_n и b_n стремятся (или не стремятся) к 0 одновременно.

Доказательство леммы.

1). Пусть $b_n \to 0$. Тогда $2^n b_n = 2^{n-1}b_{n-1} + 2^n a_n$. Откуда следует

$$a_n = b_n - \frac{1}{2}b_{n-1} \to 0.$$

2). Пусть теперь $a_n \to 0$. Тогда для $\forall \varepsilon > 0\ \exists n_1$ такое, что при всех $n \geqslant n_1 \quad |a_n| < \frac{\varepsilon}{4}$, а также существует константа $c > 0$ такая, что $|a_n| < c$ для всех n. Проведя несложные выкладки, получим

$$|b_n| \leqslant \frac{|a_0| + 2|a_1| + \ldots + 2^{n_1-1}|a_{n_1-1}|}{2^n} + \frac{2^{n_1}|a_{n_1}| + \ldots + 2^n|a_n|}{2^n} \leqslant$$

$$\leqslant \frac{c \cdot 2^{n_1}}{2^n} + \frac{\varepsilon(2^{n_1} + 2^{n_1+1} + \ldots + 2^n)}{4 \cdot 2^n} = \frac{c \cdot 2^{n_1}}{2^n} + \frac{\varepsilon \cdot 2^{n_1} \cdot (2^{n-n_1+1} - 1)}{4 \cdot 2^n}.$$

Последнее выражение можно сделать меньше произвольного ε, если выбрать n достаточно большим. Следовательно, $b_n \to 0$. Лемма доказана.

Возвращаясь к (10.9), видим, что для доказательства теоремы осталось показать, что $\eta_n(\omega) \xrightarrow{\text{п.н.}} 0$.

Действительно, применяя неравенство Колмогорова для максимума сумм независимых случайных величин, и проводя несложные выкладки, получим для $\forall\varepsilon > 0$

$$\mathbf{P}\{|\eta_n| \geqslant \varepsilon\} = \mathbf{P}\{\max_{2^n \leqslant j < 2^{n+1}} |\sum_{k=2^n}^{j} X_k| \geqslant \varepsilon \cdot 2^n\} \leqslant \frac{1}{2^{2n}\varepsilon^2} \sum_{k=2^n}^{2^{n+1}-1} \sigma_k^2 \leqslant$$

$$\leqslant \frac{2^{2(n+1)}}{\varepsilon^2 2^{2n}} \sum_{k=2^n}^{2^{n+1}-1} \frac{\sigma_k^2}{k^2} = \frac{4}{\varepsilon^2} \sum_{k=2^n}^{2^{n+1}-1} \frac{\sigma_k^2}{k^2} \to 0,$$

что следует из сходимости соответствующего ряда по условию теоремы. Но это лишь доказывает сходимость $\eta_n \to 0$ по вероятности.[2] Для доказательства $\eta_n(\omega) \xrightarrow{\text{п.н.}} 0$ применим лемму Бореля – Кантелли. Поскольку из условия теоремы и выше доказанного неравенства следует, что

$$\sum_{n=0}^{\infty} \mathbf{P}\{|\eta_n(\omega)| \geqslant \varepsilon\} \leqslant \frac{4}{\varepsilon^2} \sum_{n=0}^{\infty} \sum_{k=2^n}^{2^{n+1}-1} \frac{\sigma_k^2}{k^2} = \frac{4}{\varepsilon^2} \sum_{k=1}^{\infty} \frac{\sigma_k^2}{k^2} < \infty,$$

— ряд из вероятностей сходится, то с вероятностью 1 осуществится лишь конечное число событий $|\eta_n(\omega)| \geqslant \varepsilon$, и значит, начиная с некоторого номера $N(\omega)$ $\quad |\eta_n(\omega)| < \varepsilon$. Тем самым

$$\mathbf{P}\{\omega : \eta_n(\omega) \to 0\} = 1.$$

$\square$

[2]См. раздел "Сходимость по вероятности"главы 12.

Глава 11

Предельные теоремы и метод характеристических функций

11.1 Обозначения и формулировки предельных теорем

Определение 11.1. Случайная величина Z имеет **стандартное нормальное распределение,** если ее плотность распределения $\varphi(x)$ имеет вид
$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$
при всех значениях x.

Если плотность распределения случайной величины Y равна
$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}},$$
то распределение с такой плотностью называется **нормальным с параметрами** a **и** σ^2**.** В дальнейшем будем в этом случае писать $Y \sim N(a, \sigma^2)$. Заметим, что $\mathbf{E}Y = a$, $\mathbf{D}Y = \sigma^2$. Для стандартного нормального распределения $a = 0$, $\sigma = 1$.

Пусть $X_1, X_2, \ldots$ — независимые случайные величины, для которых существуют $\mathbf{E}X_k = a_k$, $\mathbf{D}X_k = b_k^2$, $c_k = \mathbf{E} \mid X_k - a_k \mid^3$.

Обозначим
$$S_n = X_1 + X_2 + \cdots + X_n,$$
$$A_n = \mathbf{E}S_n = a_1 + a_2 + \cdots + a_n,$$
$$B_n^2 = \mathbf{D}S_n = b_1^2 + b_2^2 + \cdots + b_n^2,$$

$$C_n = c_1 + c_2 + \cdots + c_n.$$

Замечание 1. Если существует такая константа L, что $\mid X_k - a_k \mid \leq L$ при любом $k \geq 1$, то $\mid X_k - a_k \mid^3 \leq L(X_k - a_k)^2$, а значит, $c_k \leq Lb_k^2$ и, следовательно, $C_n \leq LB_n^2$.

Замечание 2. Если случайные величины X_k равномерно ограничены, т.е. существует константа L' такая, что для всех k $|X_k| \leqslant L'$, то $|X_k - a_k| \leq 2L'$, и тогда $C_n \leq 2L'B_n^2$.

Обозначим $S_n^* = \frac{S_n - A_n}{B_n}$ - нормированную сумму. После такой нормировки $\mathbf{E}S_n^* = 0, \mathbf{D}S_n^* = 1$.

Теорема 11.1. *(А.М. Ляпунов).*[1]

$$\Delta_n = \sup_{-\infty < a < b < +\infty} |\mathbf{P}\left\{a \leq S_n^* \leq b\right\} - \frac{1}{\sqrt{2\pi}} \int_a^b e^{-x^2/2} dx| \leq 2\frac{C_n}{B_n^3}.$$

Несколько комментариев к этой теореме.

Если величина, стоящая в правой части неравенства мала, то это означает, что распределение S_n^* близко к стандартному нормальному распределению и теорема позволяет оценить точность данного приближения.

Приведем несколько примеров.

Пример 11.1. Пусть для всех k выполнено неравенство $|X_k - a_k| \leq L$, то, воспользовавшись замечанием 1, получим $\Delta_n \leq \frac{2L}{B_n}$. Таким образом, если L мало по сравнению со стандартным отклонением всей суммы S_n, то можно считать распределение S_n^* "почти нормальным".

Пример 11.2. Пусть $A_1, A_2, \ldots$ – последовательность независимых событий, причем для любого k

$$\mathbf{P}(A_k) = p,\ 0 < p < 1,\ q = 1 - p.$$

Обозначим $\mu_n(\omega) = \sum_{i=1}^n I_{A_i}(\omega)$ – число произошедших событий среди первых n. По теореме Бернулли $\frac{\mu_n}{n} \to p$ по вероятности. Так как случайная величина μ_n имеет биномиальное распределение с параметрами (n, p), то $\mathbf{E}\mu_n = np, \mathbf{D}\mu_n = npq$, и в этом случае

$$S_n^* = \frac{\mu_n - np}{\sqrt{npq}}.$$

[1]Заметим, что правая часть неравенства в теореме 11.1 меньше, чем в оригинальном неравенстве Ляпунова.

Поскольку для индикатора любого случайного события A выполняется неравенство $|I_A(\omega) - p| \leq 1$, то в нашем случае можно применить вывод предыдущего примера и получить оценку

$$\Delta_n \leq \frac{2}{\sqrt{npq}}.$$

Нормальной аппроксимацией для биномиального распределения можно пользоваться, если значение величины $\frac{2}{\sqrt{npq}}$ мало.

Отсюда следует

Теорема 11.2. *(Интегральная теорема А.Муавра – П.Лапласа). Если $n \to \infty$, то*

$$P\left\{m_1 \leq \mu_n(\omega) \leq m_2\right\} \to \int_a^b \varphi(x)dx,$$

где $a = \frac{m_1 - np}{\sqrt{npq}}, b = \frac{m_2 - np}{\sqrt{npq}}$.

Прямой вывод интегральной теоремы Муавра - Лапласа состоит в непосредственном подсчете вероятностей биномиального распределения

$$\mathbf{P}\{m_1 \leq \mu_n \leq m_2\} = \sum_{m=m_1}^{m_2} C_n^m p^m q^{n-m}$$

и применении формулы Стирлинга к отдельным слагаемым данной суммы.

Теорема Муавра - Лапласа позволяет также вычислять вероятности отклонений относительной частоты наступления события от его вероятности. Действительно,

$$\mathbf{P}\{a \leq \frac{\mu_n - np}{\sqrt{npq}} \leq b\} = \mathbf{P}\{a\frac{\sqrt{pq}}{\sqrt{n}} \leq \frac{\mu_n}{n} - p \leq b\frac{\sqrt{pq}}{\sqrt{n}}\}.$$

При этом видно, что эти отклонения имеют порядок $\sqrt{pq}/\sqrt{n}$.

Замечание 3. В практических задачах обычно заменяют

$$\mathbf{P}\{a \leq S_n^* \leq b\} = \int_a^b \varphi(x)dx,$$

не учитывая при этом, что в случаях, когда значения этого интеграла малы, применение теоремы неоправдано, так как погрешность аппроксимации может оказаться больше, чем значения этого интеграла.

Математики затратили немало усилий, чтобы справиться с этими трудностями. Соответствующие теоремы носят название теорем о больших уклонениях. Более подробно об этом можно прочитать в учебниках В.Феллера и А.А.Боровкова.

Замечание 4. Попытки получить обобщение теоремы Муавра - Лапласа привели к созданию метода характеристических функций (1901 г., Ляпунов).

11.2 Характеристические функции. Определение и свойства

Определение 11.2. Характеристической функцией случайной величины X называется функция $f(t; X)$ действительной переменной $t \in (-\infty; +\infty)$, определяемая как

$$f(t; X) = \mathbf{E}e^{itX},$$

где i–чисто мнимая единица.

Напомним, что по известной формуле Эйлера $e^{itX} = \cos(tX) + i\sin(tX)$. Математическое ожидание этой комплекснозначной величины определим как

$$\mathbf{E}e^{itX} = \mathbf{E}\cos(tX) + i\mathbf{E}\sin(tX).$$

Поскольку всякая ограниченная случайная величина имеет конечное математическое ожидание, то для всякой случайной величины существует характеристическая функция. Иногда для краткости, когда понятно, о какой случайной величине идет речь, будем обозначать характеристическую функцию $f(t)$.

Рассмотрим два частных случая.

I. Распределение дискретной случайной величины X задано таблицей

$$\begin{array}{ccccc} x_1, & x_2, & \dots & x_n, & \dots \\ p_1, & p_2, & \dots & p_n, & \dots \end{array}$$

Тогда характеристическая функция этой случайной величины равна

$$f(t; X) = \sum_{k=1}^{\infty} e^{itx_k} p_k,$$

причем если случайная величина X принимает конечное число значений, то и в сумме будет такое же число слагаемых.

II. Распределение абсолютно непрерывной случайной величины X задано плотностью распределения вероятностей $p(x)$. Тогда ее характеристическая функция равна

$$f(t;X) = \int_{-\infty}^{+\infty} e^{itx} p(x) dx.$$

Пример 11.3. Рассмотрим в качестве примера вычисление характеристической функции индикатора случайного события A

$$I_A(\omega) = \begin{cases} 1, & \omega \in A, \quad \text{с вероятностью} \quad p = P(A) \\ 0, & \omega \notin A, \quad \text{с вероятностью} \quad q = P(\bar{A}) \end{cases}$$

Характеристическая функция равна

$$f(t; I_A(\omega)) = pe^{it} + q,$$

если $p = q = \frac{1}{2}$, то $f(t) = e^{\frac{it}{2}} \cos(\frac{t}{2})$.

Пример 11.4. Вычислим характеристическую функцию случайной величины X с симметричным распределением

$$X = \begin{cases} +1, & \frac{1}{2} \\ -1, & \frac{1}{2} \end{cases}$$

Нетрудно показать, что в этом случае характеристическая функция равна $f(t) = \cos t$.

Пример 11.5. Пусть $X \sim N(0,1)$, тогда можно показать, что характеристическая функция равна

$$f(t) = e^{-\frac{t^2}{2}}.$$

Во многих учебниках по теории вероятностей имеются таблицы характеристических функций для наиболее часто встречающихся распределений, но существуют также более полные таблицы преобразований Фурье (а характеристическая функция есть не что иное как преобразование Фурье плотности или функции распределения).

1. Мультипликативность: если случайные величины X и Y независимы, то

$$f(t; X+Y) = f(t;X) \cdot f(t;Y)$$

–характеристическая функция суммы равна произведению характеристических функций слагаемых.

Доказательство: действительно, пользуясь известными формулами тригонометрии и независимостью случайных величин, получаем

$$f(t; X+Y) = \mathbf{E}\cos t(X+Y) + i\mathbf{E}\sin t(X+Y) =$$

$$= \mathbf{E}\cos tX\mathbf{E}\cos tY - \mathbf{E}\sin tX\mathbf{E}\sin tY +$$

$$+ i\mathbf{E}\sin tX\mathbf{E}\cos tY + i\mathbf{E}\cos tX\mathbf{E}\sin tY = f(t;X)f(t;Y).$$

2. Для любой характеристической функции

$$|f(t;X)| \leq 1, \ f(0;X) = 1.$$

Задача 11.1. Покажите, что если характеристическая функция

$$f(2\pi; X) = 1,$$

то она является периодической и соответствует целочисленной случайной величине.

3. При линейном преобразовании случайной величины X характеристическая функция изменяется следующим образом:

$$f(t; aX+b) = e^{itb} f(at; X).$$

Пример 11.6. Пусть $X \sim N(0,1)$, $\quad Y = \sigma X + a$. Тогда $Y \sim N(a, \sigma^2)$ и характеристическая функция случайной величины Y равна

$$f(t;Y) = e^{ita} e^{-\frac{t^2}{2}}.$$

4. Взаимная однозначность соответствия между функциями распределения и характеристическими функциями. Если случайная величина X имеет функцию распределения $F(x)$, то

$$f(t;X) = \int_{-\infty}^{+\infty} e^{itx}\, \mathbf{P}_F(dx),$$

где интеграл Лебега-Стилтьеса берется по мере $\mathbf{P}_F$, порожденной на числовой прямой функцией распределения $F(x)$. Следовательно, всякая функция распределения однозначно определяет характеристическую функцию. Но верно и обратное утверждение: функция распределения однозначно восстанавливается по характеристической функции. Доказывается это утверждение с помощью формул обращения.

11.3 Формулы обращения для характеристических функций

Прежде чем привести примеры различных формул обращения, вспомним некоторые свойства функций распределения. Мы определили ранее функцию распределения случайной величины X как непрерывную слева неубывающую функцию соотношением

$$F(x) = \mathbf{P}(X < x).$$

Такое определение приводится, например, в учебниках Б.В.Гнеденко, А.А.Боровкова. В книгах В.Феллера, А.Н.Ширяева, Б.А.Севастьянова функция распределения определяется несколько иначе:

$$F(x) = \mathbf{P}(X \leq x).$$

Отличие только в том, что эта функция непрерывна справа.

Всякая монотонно неубывающая ограниченная функция имеет не более чем счетное множество точек разрыва. Действительно, для функции распределения разрывов, по величине превосходящих $1/2$, не более одного, разрывов, превышающих или равных $1/3$, не более двух и т.д. Таким образом все точки разрыва можно перенумеровать, и, значит, их не более чем счетное число.

Точек же непрерывности – континуум, причем множество точек непрерывности всякой функции распределения всюду плотно на числовой прямой.

Если $F_1(x)$ и $F_2(x)$ – две функции распределения, то множество общих точек непрерывности также всюду плотно.

Приведем в качестве одной из формул обращения следующую теорему.

Теорема 11.3. *Если $F(x)$ и $f(t)$ – функция распределения и характеристическая функция случайной величины X, $x_1 < x_2$ –точки непрерывности $F(x)$, то*

$$F(x_2) - F(x_1) = \lim_{T\to\infty} \frac{1}{2\pi} \int\limits_{-T}^{T} \frac{e^{-itx_2} - e^{-itx_1}}{-it} f(t)dt.$$

Замечание 1. Если в этой формуле устремить $x_1 \to -\infty$ по точкам непрерывности, то получим $F(x_2)$. Поскольку x_2 – произвольная точка непрерывности, то функцию распределения можно однозначно восстановить по характеристической функции.

Существуют и другие варианты формул обращения. Так если характеристическая функция абсолютно интегрируема, то у функции распределения существует ограниченная производная $F'(x) = p(x)$ и

$$p(x) = \frac{1}{2\pi} \int\limits_{-\infty}^{\infty} e^{-itx} f(t) dt. \tag{11.1}$$

Формулу (11.1) называют формулой обращения для плотности.

Замечание 2. Формулы

$$f(t) = \int\limits_{-\infty}^{\infty} e^{itx} p(x) dx,$$

$$p(x) = \frac{1}{2\pi} \int\limits_{-\infty}^{\infty} e^{-itx} f(t) dt$$

очень похожи, за исключением знака в показателях экспонент и множителя перед вторым интегралом. Первую из них называют преобразованием Фурье функции $p(x)$, вторую – обратным преобразованием Фурье. Подробно с этой темой вы ознакомитесь в курсе функционального анализа.

Приведем еще один вариант формулы обращения для целочисленной случайной величины X :

$$\mathbf{P}\{\omega :\ X(\omega) = m\} = \frac{1}{2\pi} \int\limits_{-\pi}^{\pi} e^{-itm} f(t) dt$$

Итак, между функциями распределения и характеристическими функциями существует взаимно однозначное соответствие. В этом состояло **свойство 4** характеристических функций.

11.4 Свойство непрерывности соответствия характеристических функций и функций распределения

Для того, чтобы сформулировать точно это свойство, введем обозначения.

Пусть $\{X_n\}_1^\infty$ — последовательность случайных величин. Обозначим $F_n(x)$ и $f(t; X_n)$ – соответственно функцию распределения и характеристическую функцию случайной величины $X_n(n = 1, 2, \ldots)$.

Определение 11.3. Последовательность случайных величин X_n сходится по распределению к некоторой случайной величине X ($X_n \xrightarrow{d} X$) с функцией распределения $F(x)$, если $F_n(x) \to F(x)$ в каждой точке непрерывности $F(x)$. Эту сходимость функций распределения называют **слабой сходимостью** и обозначают

$$F_n(x) \Rightarrow F(x).$$

Сформулируем теорему о непрерывном соответствии между характеристическими функциями и функциями распределения.

Теорема 11.4. *Для сходимости $F_n(x) \Rightarrow F(x)$ необходимо и достаточно, чтобы $f(t; X_n) \to f(t; X)$ равномерно по t в каждом конечном интервале.*

Такой вид сходимости характеристических функций также называют слабой сходимостью и обозначают

$$f(t; X_n) \Rightarrow f(t; X).$$

Задача 11.2. Привести пример последовательности характеристических функций, которые сходятся равномерно на каждом конечном интервале, но не сходятся равномерно на всей прямой.

Мы не будем приводить доказательство теоремы 11.4 в общем случае, а ограничимся частным случаем, а именно докажем, что из слабой сходимости характеристических функций нормированных сумм к характеристической функции стандартного нормального распределения

$$f(t; S_n^*) \Rightarrow e^{-t^2/2} \quad (\mathbf{E}e^{itS_n^*} \Rightarrow \mathbf{E}e^{itZ}) \tag{11.2}$$

следует сходимость математических ожиданий

$$\mathbf{E}H(S_n^*) \to \mathbf{E}H(Z) \tag{11.3}$$

для достаточно широкого класса функций. С помощью (11.3) можно будет доказать, что для произвольных $-\infty < c < d < \infty$ выполняется

$$\mathbf{E}I_{[c,d]}(S_n^*) \to \mathbf{E}I_{[c,d]}(Z),$$

что и означает утверждение центральной предельной теоремы.

Пусть выполняется (11.2).

Рассмотрим для любого натурального n и произвольных комплексных чисел $c_1, c_2, \dots c_n$ и действительных $t_1 < t_2 < \dots < t_n$ функцию

$$Q(x) = \sum_{k=1}^{n} c_k e^{it_k x}. \tag{11.4}$$

Из (11.2) следует, что для функций указанного вида

$$\mathbf{E}Q(S_n^*) \to \mathbf{E}Q(Z).$$

Возникает вопрос, насколько богат "запас"функций действительной переменной, которые можно приближать функциями вида (11.4) или функциями более общего вида

$$Q(x) = \sum_{k=1}^{n} \left(\alpha_k \cos(t_k x) + i\beta_k \sin(t_k x)\right), \tag{11.5}$$

где $\alpha_k,\ \beta_k,\ 1 \le k \le n$ – произвольные комплексные числа.

Пример 11.7. Рассмотрим функцию

$$H(x) = \begin{cases} 1 - \frac{|x|}{a} & , \quad |x| \le a \\ 0 & , \quad |x| > a. \end{cases}$$

В книге В.Феллера показывается, что

$$H(x) = \frac{1}{\pi} \int_{-\infty}^{\infty} e^{itx}\, \frac{1 - \cos at}{at^2} dt = \int_{-\infty}^{\infty} \cos tx\, \frac{1 - \cos at}{at^2} dt.$$

Пример 11.8. Рассмотрим периодическую с периодом 2 функцию

$$H(x) = \frac{1}{2} + \frac{4}{\pi^2} \left(\frac{\cos \pi x}{1^2} + \frac{\cos 3\pi x}{3^2} + \frac{\cos 5\pi x}{5^2} + \dots \right),$$

которая при $x \in [-1, 1]$ совпадает с функцией предыдущего примера, если $a = 1$. Эта функция также представляет собой предел функций вида (11.5).

Пример 11.9. Следующая функция такого же вида "похожа"на индикатор отрезка $[-l, l]$, если $0 < \delta < l$ мало:

$$H(x) = \frac{1}{\pi} \int_{-\infty}^{\infty} e^{itx}\, \frac{\sin lt}{t}\, \frac{\sin \delta t}{\delta t} dt.$$

Пример 11.10. (Распределение Коши.) Этот пример отличается от предыдущих тем, что приведенная ниже функция не является функцией вида (11.5):

$$H(x) = e^{-\lambda|x|} = \int\limits_{-\infty}^{\infty} e^{itx} \frac{\lambda}{\pi(\lambda^2 + t^2)} dt.$$

Тем не менее общее есть: в примерах 11.7, 11.9, 11.10 функции можно представить в виде

$$H(x) = \int\limits_{-\infty}^{\infty} e^{itx} h(t) dt, \tag{11.6}$$

где $h(t)$ — непрерывная ограниченная функция, интегрируемая на всей прямой.

Обозначим $\mathcal{H}$ — класс функций вида (11.6).

Лемма 11.1. Пусть функция $H(x)$ равна 0 вне конечного интервала, непрерывна и имеет две непрерывные производные на всей прямой. Тогда $H(x) \in \mathcal{H}$.

Замечание. Функция примера 3 входит в этот класс, хотя и не удовлетворяет условиям леммы, то есть условие является достаточным, но не является необходимым.

На самом деле, утверждения, которые последуют дальше, верны и для более широкого класса функций (от условия равенства 0 вне конечного интервала можно отказаться, но тогда для доказательства наших теорем понадобилась бы теорема Фубини для интеграла Лебега, которая в этом курсе не рассматривалась).

Лемма 11.2. Пусть $H(x) \in \mathcal{H}$. Тогда для произвольной случайной величины Y

$$\mathbf{E}H(Y) = \int\limits_{-\infty}^{\infty} f(t; Y) h(t) dt\ .$$

Доказательство. Доказательство проведем для дискретной случайной величины, принимающей конечное число различных между собой значений $y_1, y_2, \ldots, y_s$ с вероятностями $p_1, p_2, \ldots, p_s$:

$$\mathbf{E}H(Y) = \sum_{j=1}^{s} H(y_j) p_j = \sum_{j=1}^{s} \left(\int\limits_{-\infty}^{\infty} e^{ity_j} h(t) dt \right) p_j = \int\limits_{-\infty}^{\infty} \left(\sum_{j=1}^{s} e^{ity_j} p_j \right) h(t) dt =$$

$$= \int\limits_{-\infty}^{\infty} f(t;Y)h(t)dt.$$

$\square$

Заметим, что в рассмотренном случае правомерность замены порядков суммирования и интегрирования сомнений не вызывала. В общем случае надо использовать теорему Фубини для интеграла Лебега.

Повторим рассуждения для случайной величины $Z \sim N(0,1)$ с плотностью $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$:

$$\mathbf{E}H(Z) = \int\limits_{-\infty}^{\infty} H(x)\varphi(x)dx = \int\limits_{-\infty}^{\infty} \left(\int\limits_{-\infty}^{\infty} e^{itx} h(t)dt \right) \varphi(x)dx =$$

$$= \int\limits_{-\infty}^{\infty} h(t) \left(\int\limits_{-\infty}^{\infty} e^{itx}\varphi(x)dx \right) dt = \int\limits_{-\infty}^{\infty} h(t)e^{-t^2/2}\, dt\ .$$

В этом случае замена порядков интегрирования, хотя и верна, но уже не столь очевидна. Впрочем, для интегралов Римана теорема Фубини в курсе математического анализа была, и вы можете убедиться в правильности проведенных выкладок.

Теорема 11.5. *Пусть $H(x) \in \mathcal{H}$, $\{Y_n\}$ – последовательность случайных величин, для которой характеристические функции сходятся к характеристической функции случайной величины Y*

$$f(t;Y_n) \to f(t;Y)$$

равномерно в каждом конечном интервале. Тогда

$$\mathbf{E}H(Y_n) \to \mathbf{E}H(Y)\ .$$

Замечание 2. На самом деле утверждение теоремы выполняется для более широкого класса функций $H(x)$ – непрерывных и ограниченных. Нам же эта теорема понадобится только для случайной величины $Z \sim N(0,1)$, поэтому доказательство проведем только для $H(x) \in \mathcal{H}$.

Доказательство.

$$\mathbf{E}H(Y_n) - \mathbf{E}H(Z) = \int\limits_{-\infty}^{\infty} h(t) \left(f(t;Y_n) - e^{-t^2/2} \right) dt = I_1 + I_2 + I_3\ ,$$

где интегралы I_1, I_2, I_3 берутся соответственно по промежуткам $[-T,T], (-\infty,-T], [T,\infty)$. Оценим каждый из них по отдельности. Выбор константы T произведем позднее.

$$|I_1| \leq 2TA_h \sup_{|t|\leq T} \left| f(t;Y_n) - e^{-t^2/2} \right| ,$$

где $A_h = \sup_t h(t)$,

$$|I_2| \leq \int\limits_{-\infty}^{-T} h(t)dt \ , \qquad |I_2| \leq \int\limits_{T}^{\infty} h(t)dt \ .$$

Таким образом получаем

$$|\mathbf{E}H(Y_n) - \mathbf{E}H(Z)| \leq |I_1| + |I_2| + |I_3| \ .$$

Поскольку $h(t)$ — абсолютно интегрируемая на всей прямой функция, то для $\forall\varepsilon > 0$ существует $T = T(\varepsilon, h)$ такое, что

$$|I_2| + |I_3| < 2\varepsilon/3 \ .$$

Оставшееся слагаемое

$$|I_1| \leq 2T(\varepsilon,h)A_h \sup_{|t|\leq T} \left| f(t,Y_n) - e^{-t^2/2} \right|$$

и его можно сделать меньше, чем $\varepsilon/3$, выбрав n достаточно большим, так как существует n_0 такое, что для всех $n \geq n_0$

$$\sup_{|t|\leq T} \left| f(t,Y_n) - e^{-t^2/2} \right| < \frac{\varepsilon}{6TA_h} \ .$$

Следовательно, $\mathbf{E}H(Y_n) \to \mathbf{E}H(Z)$ для всех функций $H(x) \in \mathcal{H}$. □

Покажем теперь, что аналогичное утверждение верно не только для функций из $\mathcal{H}$. Рассмотрим функцию $I_{[c,d]}(x)$ — индикатор произвольного отрезка $[-1,1]$. Такие функции можно приближать функциями из класса $\mathcal{H}$.

А именно для малых $\delta > 0$ (см. рис.11.1) существует $H_\delta^+(x) \in \mathcal{H}$ такая, что

$$H_\delta^+(x) \geq I_{[c,d]}, \qquad H_\delta^+(x) = 0 \quad \text{для всех} \quad x \notin [c-\delta, d+\delta] \ .$$

Тогда

$$\mathbf{P}\{c \leq Y_n \leq d\} = \mathbf{E}I_{[c,d]}(Y_n) \leq \mathbf{E}H_\delta^+(Y_n) \ .$$

Рис. 11.1:

Переходя к пределу при $n \to \infty$, получим

$$\limsup_{n\to\infty} \mathbf{P}\{c \le Y_n \le d\} \le \limsup_{n\to\infty} \mathbf{E} H_\delta^+(Y_n) = \lim_{n\to\infty} \mathbf{E} H_\delta^+(Y_n) =$$

$$= \mathbf{E} H_\delta^+(Z) \le \mathbf{E} I_{[c-\delta, d+\delta]}(Z) \ .$$

Следовательно,

$$\limsup_{n\to\infty} \mathbf{P}\{c \le Y_n \le d\} \le \int\limits_{c-\delta}^{d+\delta} \varphi(x)dx \le \int\limits_{c}^{d} \varphi(x)dx + \int\limits_{c-\delta}^{c} \varphi(x)dx + \int\limits_{d}^{d+\delta} \varphi(x)dx \le$$

$$\le \int\limits_{c}^{d} \varphi(x)dx + \frac{1}{\sqrt{2\pi}} 2\delta \ .$$

Устремив теперь $\delta \to 0+$, получим

$$\limsup_{n\to\infty} \mathbf{P}\{c \le Y_n \le d\} \le \int\limits_{c}^{d} \varphi(x)dx = \mathbf{P}\{c \le Z \le d\} \ .$$

Аналогичным образом, выбирая $H_\delta^-(x) \in \mathcal{H}$

Рис. 11.2:

(см. рис.11.2) такую, что

$$H_\delta^-(x) = 0 \quad \text{для всех} \quad x \notin [c, d] \ , \quad H_\delta^-(x) \le I_{[c,d]} \ ,$$

получим, что

$$\limsup_{n\to\infty} \mathbf{P}\{c \le Y_n \le d\} \ge \mathbf{P}\{c \le Z \le d\} \ .$$

Итак, доказана следующая

Теорема 11.6. *Если* $f(t, Y_n) \Rightarrow e^{-t^2/2}$, *то для любого отрезка* $[c, d]$

$$\lim_{n\to\infty} \mathbf{P}\{c \le Y_n \le d\} = \mathbf{P}\{c \le Z \le d\} \ .$$

11.5 Примеры слабой сходимости последовательностей характеристических функций

Рассмотрим сначала на примерах, как можно установить сходимость характеристических функций.

Пример 11.11. Пусть $X_1, X_2, \dots$ — независимые случайные величины, принимающие значения 1 и (-1) с вероятностями $\frac{1}{2}$ с характеристической функцией $f(t, X_k) = \cos t$, $\mathbf{E}X_k = 0$. $\mathbf{D}X_k = 1$. Тогда для нормированной суммы $S_n^* = \frac{S_n}{\sqrt{n}}$ имеем

$$f(t, S_n^*) = f(\frac{t}{\sqrt{n}}, S_n) = \cos^n(\frac{t}{\sqrt{n}}) \ .$$

При оценке близости характеристических функций воспользуемся простой, но полезной леммой.

Лемма 11.3. Если $|\alpha| \le 1, \quad |\beta| \le 1$, то

$$|\alpha^n - \beta^n| \le |\alpha - \beta| n$$

для любого натурального n.

Воспользовавшись неравенством леммы 3, получим

$$\left| f(t; S_n^*) - e^{-t^2/2} \right| = \left| \cos^n(\frac{t}{\sqrt{n}}) - (e^{-\frac{t^2}{2n}})^n \right| \le n \left| \cos(\frac{t}{\sqrt{n}}) - e^{-\frac{t^2}{2n}} \right| \ .$$

Поскольку $|t| \le T, \ n \to \infty$, то необходимо уметь оценивать эту разность лишь в малой окрестности 0. Воспользуемся конкретными свойствами функций:

$$\cos u = 1 - \frac{u^2}{2} + \frac{u^4}{4!} - \dots + (-1)^n \frac{u^{2n}}{(2n)!} + R_n(u),$$

причем остаток $R_n(u)$ не превосходит по модулю первого отброшенного члена в разложении Тейлора и имеет знак первого отброшенного члена;

$$e^{-u} = 1 + u + \frac{u^2}{2!} - \frac{u^3}{3!} + \cdots + (-1)^n \frac{u^n}{n!} + Q_n(u) \ ,$$

где остаток $Q_n(u)$ обладает аналогичным свойством для всех $u > 0$. Такое свойство рядов Тейлора называют свойством "обертывания". Следовательно,

$$0 \leq \cos u - (1 - \frac{u^2}{2}) \leq \frac{u^4}{4!} \ , \qquad 0 \leq e^{-u} - (1 - u) \leq \frac{u^2}{2} \ ,$$

и, значит,

$$0 \leq \cos(\frac{t}{\sqrt{n}}) - (1 - \frac{t^2}{2n}) \leq \left(\frac{t^2}{n}\right)^2 \frac{1}{24},$$

$$0 \leq e^{-\frac{t^2}{2n}} - (1 - \frac{t^2}{2n}) \leq \left(\frac{t^2}{2n}\right)^2 \frac{1}{2}.$$

Таким образом для всех t имеем

$$\left|\cos(\frac{t}{\sqrt{n}}) - e^{-\frac{t^2}{2n}}\right| \leq \frac{t^4}{n^2}(\frac{1}{24} + \frac{1}{8}) = \frac{t^4}{6n^2}.$$

Следовательно, для всех $|t| \leq T$

$$|f(t; S_n^*) - e^{-t^2/2}| \leq n \left|\cos(\frac{t}{\sqrt{n}}) - e^{-\frac{t^2}{2n}}\right| \leq \frac{T^4}{6n} \to 0$$

равномерно в каждом конечном интервале.

Пример 11.12. Пусть $A_1, A_2, \ldots$ — последовательность независимых событий, каждое из которых имеет вероятность p. Будем говорить, что в испытании с номером k произошел успех, если произошло событие A_k. Тогда $S_n = \sum\limits_{k=1}^{n} I_{A_k}(\omega)$ – это число успехов в n испытаниях. Нетрудно убедиться в том, что

$$f(t; S_n^*) = \left(qe^{-\frac{itp}{\sqrt{npq}}} + pe^{\frac{itq}{\sqrt{npq}}}\right)^n .$$

Эта функция является комплекснозначной. Но и в этом случае, проводя непосредственые вычисления, можно убедиться в том. что $f(t; S_n^*) \Rightarrow e^{-t^2/2}$.

11.6 Доказательство центральной предельной теоремы

Метод характеристических функций — основной при доказательстве **центральной предельной теоремы**. Ранее была сформулирована теорема Ляпунова. Приведем более простой вариант центральной предельной теоремы для одинаково распределенных случайных величин. Сформулируем центральную предельную теорему в форме Линдеберга – Леви.

Теорема 11.7. *Пусть $X_1, X_2, \ldots$ — независимые одинаково распределенные случайные величины, имеющие конечные моменты первого и второго порядков: $a = \mathbf{E}X_k,\ b^2 = \mathbf{D}X_k,\ k = 1, 2, \ldots$. Тогда*

$$\sup_{-\infty \le c < d \le \infty} \left| \mathbf{P}\left\{c \le S_n^* \le d\right\} - \frac{1}{\sqrt{2\pi}} \int\limits_c^d e^{-x^2/2} dx \right| \to 0 \qquad (11.7)$$

при $n \to \infty$.

Примеры предыдущей главы показывали, каким образом можно установить, что

$$f(t; S_n^*) \Rightarrow e^{-t^2/2}.$$

Рассмотрим

$$S_n^* = \frac{S_n - an}{b\sqrt{n}} = \sum_{k=1}^n \frac{X_k - a}{b\sqrt{n}} = \frac{1}{\sqrt{n}} \sum_{k=1}^n X_k^*,$$

где $X_k^* = \frac{X_k - a}{b}, \quad \mathbf{E}X_k^* = 0, \quad \mathbf{D}X_k^* = 1.$

Пользуясь свойствами характеристических функций, получаем

$$f(t; S_n^*) = f^n\left(t; \frac{X_1 - a}{b\sqrt{n}}\right) = f^n\left(\frac{t}{\sqrt{n}}; X_k^*\right).$$

Подобно тому, как мы поступали в примере 1 предыдущего раздела, запишем

$$\left|f(t; S_n^*) - e^{-t^2/2}\right| = \left|f^n(\frac{t}{\sqrt{n}}; X_1^*) - (e^{-t^2/2n})^n\right| \le n \left|f(\frac{t}{\sqrt{n}}; X_1^*) - e^{-t^2/2n}\right| =$$

$$= n \left|f(\frac{t}{\sqrt{n}}; X_1^*) - (1 - \frac{t^2}{2n}) - (e^{-t^2/2n} - (1 - \frac{t^2}{2n}))\right| \le$$

$$\le n \left|f(\frac{t}{\sqrt{n}}; X_1^*) - (1 - \frac{t^2}{2n})\right| + n\frac{t^4}{8n^2}.$$

Если $|t| \le T$, а $n \to \infty$, то $\frac{t}{\sqrt{n}} \to 0$, и значит, необходимо изучить поведение характеристических функций в окрестности 0.

Пример 11.13. Пусть случайная величина Y принимает конечное число различных значений $y_1, y_2, \dots, y_s$ с вероятностями $q_1, q_2, \dots, q_s$. В этом случае

$$f(t; Y) = \sum_{l=1}^{s} e^{ity_l} q_l.$$

У этой функции существуют производные всех порядков и

$$\frac{d^k f(t; Y)}{dt^k} = \sum_{l=1}^{s} (iy_l)^k e^{ity_l} q_l \ ,$$

и если теперь возьмем эти производные в точке $t = 0$, то получим

$$\left.\frac{d^k f(t; Y)}{dt^k}\right|_{t=0} = i^k \sum_{l=1}^{s} y_l^s q_l = i^k \mathbf{E} Y^k \ .$$

Таким образом, производные характеристической функции в $t = 0$ связаны с моментами случайной величины.

Применим к характеристической функции разложение в ряд Тейлора в окрестности 0:

$$f(t; Y) = 1 + it\mathbf{E}Y + \frac{1}{2} t^2 i^2 \mathbf{E} Y^2 + o(t^2).$$

Если $\mathbf{E}Y = 0, \quad \mathbf{D}Y = 1$, то

$$f(t; Y) = 1 - \frac{t^2}{2} + o(t^2).$$

Если вместо t подставить $\frac{t}{\sqrt{n}}$, то получим

$$f(\frac{t}{\sqrt{n}}; Y) = 1 - \frac{t^2}{2n} + o(\frac{t^2}{2n}).$$

Доказательство этого факта в общем случае требует применения теорем о предельном переходе под знаком интеграла. Поэтому мы только сформулируем общее утверждение.

Лемма 11.4. Если существуют $\mathbf{E}Y$, $\mathbf{D}Y$, то

$$f(t) = 1 + it\mathbf{E}Y - \frac{t^2}{2} \mathbf{E} Y^2 + o(t^2).$$

Для доказательства следующего утверждения потребуется лемма из анализа.

Лемма 11.5. При любом действительном z и любом натуральном n

$$\left| e^{iz} - \left(1 + iz + \frac{(iz)^2}{2!} + \dots + \frac{(iz)^n}{n!} \right) \right| \le \frac{|z|^{n+1}}{(n+1)!}.$$

Доказательство этой леммы можно прочитать в учебниках А. Н. Ширяева, Б. А. Севастьянова.

Лемма 11.6. Пусть в дополнение к условиям леммы 1 случайная величина Y имеет третий конечный момент. Тогда

$$\left| f(t;Y) - \left(1 + it\mathbf{E}Y - \frac{1}{2}t^2\mathbf{E}Y^2 \right) \right| \le \frac{1}{6}|t|^3\mathbf{E}|Y|^3 .$$

Доказательство следует из леммы 2 при $n = 2$, если от обеих частей неравенства взять математические ожидания.

Возвращаясь к доказательству центральной предельной теоремы, получаем

$$\left| f(t;S_n^*) - e^{-t^2/2} \right| \le n \left| f(\frac{t}{\sqrt{n}};X_1^*) - (1 - \frac{t^2}{2n}) \right| + \frac{t^4}{8n} \le \frac{n|t|^3\mathbf{E}|X_1^*|^3}{6(\sqrt{n})^3} + \frac{t^4}{8n}.$$

Поскольку $|t| \le T$, то последнее выражение не превосходит

$$\frac{T^3\mathbf{E}|X_1 - a|^3}{6\sqrt{n}b^3} + \frac{T^4}{8n} \to 0.$$

Следовательно, $f(t;S_n^*) \Rightarrow f(t;Z)$, причем скорость сходимости имеет порядок $O(\frac{1}{\sqrt{n}})$.

Итак, доказана центральная предельная теорема в предположении существования конечного третьего момента. На самом деле для ее выполнения достаточно наличие первых двух моментов.

Замечание. Центральной предельной теоремой называют все теоремы, в которых устанавливается сходимость к нормальному распределению.

11.7 Теорема Пуассона

В теореме Пуассона устанавливается сходимость к другому распределению — распределению Пуассона. С.Пуассон — известный французский математик первой половины XIX века, доказавший, что в схеме испытаний Бернулли, когда $n \to \infty$, распределение числа успехов при определенных условиях сходится не к нормальному распределению, а к распределению Пуассона.

Рассмотрим схему серий испытаний Бернулли: в серии с номером n проводится n независимых испытаний, каждое из которых имеет вероятность p_n, зависящую от номера серии.

Теорема 11.8. *Пусть $n \to \infty$, $\quad p_n \to 0$ таким образом, что $np_n \to \lambda > 0$. Тогда*

$$P_n(m) = C_n^m p_n^m (1-p_n)^m \to \frac{e^{-\lambda} \cdot \lambda^m}{m!}.$$

Доказательство теоремы в этой форме чрезвычайно просто. Легко показать,

$$P_n(0) = (1-p_n)^n = e^{n\ln(1-p_n)} = e^{-np_n \cdot \frac{\ln(1-p_n)}{-p_n}} \to e^{-\lambda},$$

$$\frac{P_n(k)}{P_n(k-1)} = \frac{n! p_n^k (1-p_n)^{n-k}}{k!(n-k)!} \cdot \frac{(k-1)!(n-k+1)!}{n! p_n^{k-1}(1-p_n)^{n-k+1}} = \frac{n-k+1}{k} \cdot \frac{p_n}{1-p_n} =$$

$$= \frac{np_n(1-\frac{k-1}{n})}{k(1-p_n)} \to \frac{\lambda}{k}$$

для всех $k \geq 1$.

Тогда

$$P_n(m) = \frac{P_n(m)}{P_n(m-1)} \cdot \frac{P_n(m-1)}{P_n(m-2)} \cdots \frac{P_n(1)}{P_n(0)} \cdot P_n(0) \to \frac{\lambda}{m} \cdot \frac{\lambda}{m-1} \cdots \frac{\lambda}{1} \cdot e^{-\lambda} = \frac{\lambda^m e^{-\lambda}}{m!}.$$

Мы доказали очень простым способом теорему Пуассона для схемы испытаний Бернулли. Эта теорема имеет обобщения, в том числе позволяющие получить оценку скорости сходимости к распределению Пуассона.

Рассмотрим следующую схему, которая называется **схемой Пуассона**.

Пусть $A_1, A_2, A_3, \ldots$ — независимые случайные события, для которых $\mathbf{P}(A_k) = p_k, \quad k = 1, 2, \ldots$. Обозначим

$$\mu_n(\omega) = \sum_{k=1}^{n} I_{A_k}(\omega)$$

— число наступивших событий A_k среди первых n событий нашей последовательности.

Замечание. Если $\mathbf{P}(A_k) = p$ для всех k, то получим как частный случай схему Бернулли, в которой случайную величину $\mu_n(\omega)$ называли числом успехов в n испытаниях.

Случайная величина $\mu_n(\omega)$ принимает целые значения от 0 до n, но выписать распределение в явном виде для различных между собой p_k и тем более работать с ним — задача достаточно сложная. Проиллюстрируем это на примере.

Пример 11.14. Рассмотрим схему Пуассона при $n = 5,\ m = 2$. Тогда вероятность

$$\mathbf{P}\{\mu_n = 2\} = p_1p_2q_3q_4q_5 + p_1q_2p_3q_4q_5 + \cdots + q_1q_2q_3p_4p_5$$

выражается через сумму $C_5^2 = 10$ различных слагаемых.

Как видим, даже при малых n формулы получаются громоздкими и вряд ли их можно использовать для доказательства предельных теорем. Поэтому воспользуемся методом характеристических функций.

Приведем несколько утверждений для характеристических функций целочисленных случайных величин.

Пусть случайная величина X принимает целые значения $m \quad (m = 0, \pm 1, \pm 2, \ldots)$ с вероятностями $p_m, \quad \sum\limits_m p_m = 1$. Тогда характеристическая функция имеет вид

$$f(t; X) = \sum_{m=-\infty}^{\infty} e^{itm} p_m, \tag{11.8}$$

причем ряд сходится абсолютно и равномерно на всей прямой.

Замечание. Ряд вида (11.8) называется тригонометрическим. Мы будем рассматривать тригонометрические ряды и более общего вида:

$$\sum_{k=-\infty}^{\infty} c_k e^{itk},$$

где $c_k(-\infty < k < \infty)$ — комплексные числа, такие, что $\sum\limits_{k=-\infty}^{\infty} |c_k| < \infty$.

Функции $e^{itk}, \quad k = 0, \pm 1, \pm 2, \ldots$ обладают следующими свойствами:

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} e^{itk} dt = \begin{cases} 1, & k = 0 \\ 0, & k \neq 0 \end{cases},$$

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} e^{int} e^{-imt} dt = \begin{cases} 1, & n = m \\ 0, & n \neq m \end{cases}. \tag{11.9}$$

Замечание. На множестве комплекснозначных функций, определенных на $[-\pi; \pi]$, для которых

$$\int_{-\pi}^{\pi} |f(t)|^2 dt < \infty,$$

можно ввести скалярное произведение

$$(f,g)=\int\limits_{-\pi}^{\pi} f(t)\overline{g}(t)dt.$$

Тогда свойство (11.9) означает ортогональность функций e^{itn} и e^{itm} при $m\neq n$.

Пусть

$$f(t)=\sum_{k=-\infty}^{\infty} c_k e^{itk}, \quad \sum_{k-\infty}^{\infty}|c_k|<\infty. \tag{11.10}$$

Умножим (11.10) на e^{-imt} и проинтегрируем по отрезку $[-\pi,\pi]$. Тогда получим, используя свойства абсолютно сходящихся рядов, что

$$\frac{1}{2\pi}\int\limits_{-\pi}^{\pi} e^{-imt}f(t)dt=\sum_{k=-\infty}^{\infty} c_k\frac{1}{2\pi}\int\limits_{-\pi}^{\pi} e^{-imt}\cdot e^{ikt}dt=c_m.$$

Следовательно, мы можем, зная характеристическую функцию, однозначно восстановить распределение целочисленной случайной величины. Тем самым доказана следующая

Лемма 11.7. Если $f(t;x)$ — характеристическая функция случайной величины X, принимающей значение m с вероятностью $p_m \quad (m=\pm1,\pm2,\cdots)$, то

$$p_m=\frac{1}{2\pi}\int\limits_{-\pi}^{\pi} e^{-itm}f(t;X)dt.$$

Эта формула является аналогом формулы обращения для целочисленных случайных величин.

Задача 11.3. Предложите аналог формулы обращения для дискретных случайных величин.

Лемма 11.8. Пусть выполнены условия леммы 11.7. Тогда для произвольных целых чисел $m_1\leq m_2$

$$\mathbf{P}\{m_1\leq X\leq m_2\}=\frac{1}{2\pi}\int\limits_{-\pi}^{\pi} S_{m_1,m_2}(t)f(t)dt, \tag{11.11}$$

где

$$S_{m_1,m_2}=\frac{e^{-it(m_2+\frac{1}{2})}-e^{-it(m_1-\frac{1}{2})}}{-2i\sin\frac{t}{2}},$$

Доказательство.

$$\mathbf{P}\{m_1 \le X \le m_2\} = \sum_{m=m_1}^{m_2} p_m = \sum_{m=m_1}^{m_2} \frac{1}{2\pi}\int\limits_{-\pi}^{\pi} e^{-itm} f(t)dt =$$

$$= \frac{1}{2\pi}\int\limits_{-\pi}^{\pi} f(t)\cdot \sum_{m=m_1}^{m_2} e^{-itm}dt = \frac{1}{2\pi}\int\limits_{-\pi}^{\pi} f(t)\cdot \frac{e^{-itm_1} - e^{-it(m_2+1)}}{1-e^{-it}}dt.$$

Домножив числитель и знаменатель дроби, стоящей под интегралом, на $e^{\frac{it}{2}}$, получим

$$\mathbf{P}\{m_1 \le X \le m_2\} = \frac{1}{2\pi}\int\limits_{-\pi}^{\pi} \frac{e^{-it(m_1-1/2)} - e^{-it(m_2+1/2)}}{e^{\frac{it}{2}} - e^{-\frac{it}{2}}} f(t)dt =$$

$$= \frac{1}{2\pi}\int\limits_{-\pi}^{\pi} \frac{e^{-it(m_1-1/2)} - e^{-it(m_2+1/2)}}{2i\sin\frac{t}{2}} f(t)dt,$$

откуда и следует утверждение леммы 11.8. $\square$

Формула (11.11) является аналогом формулы обращения для функций распределения.

Следствие 11.7.1. *Для всех* $m_1 \le m_2$

$$|S_{m_1,m_2}(t)| \le \frac{2}{|\sin t/2|}.$$

Лемма 11.9. Если

$$f(t) = \sum_k c_k e^{itk}, \quad \sum_k |c_k| < \infty,$$

$$g(t) = \sum_k d_k e^{itk}, \quad \sum |d_k| < \infty,$$

то

$$\sup_m |c_m - d_m| \le \frac{1}{2\pi}\int\limits_{-\pi}^{\pi} |f(t) - g(t)|dt, \tag{11.12}$$

$$\sup_{m_1\le m_2} \left|\sum_{m=m_1}^{m_2} c_m - \sum_{m=m_1}^{m_2} d_m\right| \le \frac{1}{2\pi}\int\limits_{-\pi}^{\pi} \frac{|f(t)-g(t)|}{|\sin t/2|}dt. \tag{11.13}$$

Эта лемма позволяет оценить близость функций распределения, если известны соответствующие характеристические функции, и является аналогом формулы обращения, приведенной без доказательства в теореме 11.3. Из леммы следует (для целочисленных случайных величин), что из сходимости характеристических функций $f_n(t) \Rightarrow g(t)$ вытекает сходимость функций распределения $F_n(x) \to F(x)$ в каждой точке распределения $F(x)$.

Доказательство. Из леммы 11.7 следует, что для всех целых m

$$c_m - d_m = \frac{1}{2\pi} \int\limits_{-\pi}^{\pi} e^{-itx} \left(f(t) - g(t)\right) dt,$$

и, значит,

$$\sup_m |c_m - d_m| \leq \frac{1}{2\pi} \int\limits_{-\pi}^{\pi} |f(t) - g(t)| dt.$$

Аналогично, применяя лемму 11.8 для $m_1 \leq m_2$, получаем

$$\left| \sum_{m=m_1}^{m_2} c_m - \sum_{m=m_1}^{m_2} d_m \right| \leq \frac{1}{2\pi} \int\limits_{-\pi}^{\pi} \frac{|f(t) - g(t)|}{|\sin t/2|} dt.$$

□

Вернемся к схеме Пуассона: $A_1, A_2, \ldots$ — независимые случайные события, $\mathbf{P}(A_k) = p_k$, и

$$\mu_n(\omega) = \sum_{k=1}^{n} I_{A_k}(\omega).$$

Характеристическая функция случайной величины $\mu_n(\omega)$ равна

$$f(t; \mu_n) = \prod_{k=1}^{n} f(t; I_{A_k}) = \prod_{k=1}^{n} (1 - p_k + p_k e^{it}) = \prod_{k=1}^{n} \left(1 + p_k(e^{it} - 1)\right).$$

Обозначим $\lambda = \mathbf{E}\mu_n = \sum\limits_{k=1}^{n} p_k$ и рассмотрим случайную величину Y, имеющую распределение Пуассона с параметром λ :

$$\mathbf{P}\{Y = m\} = \frac{e^{-\lambda}\lambda^m}{m!}, \quad m = 0, 1, 2, \ldots$$

Нетрудно показать, что

$$f(t;Y) = e^{\lambda(e^{it}-1)}, \quad \mathbf{E}Y = \lambda, \quad \mathbf{D}Y = \lambda \,.$$

Обозначим

$$\Delta_m = |\mathbf{P}\{\mu_n = m\} - \mathbf{P}\{Y = m\}| \,.$$

Теорема 11.9. *Для схемы Бернулли, если* $p_k = p \leq \frac{1}{2}$ *для всех* k, *и* $\lambda = np$, *то*

$$\Delta_m \leq \frac{3}{2} \cdot \frac{\lambda^2}{n} \,. \tag{11.14}$$

Доказательство. В силу леммы 11.9 и неравенства леммы 11.5 имеем

$$\Delta_m \leq \frac{1}{2\pi} \int\limits_{-\pi}^{\pi} \left| \big(1 + p(e^{it} - 1)\big)^n - \left(e^{\frac{\lambda}{n}(e^{it}-1)}\right)^n \right| dt \leq$$

$$\leq \frac{1}{2\pi} \int\limits_{-\pi}^{\pi} n \left| 1 + p(e^{it} - 1) - e^{-p(e^{it}-1)} \right| dt \,.$$

Обозначим $z = e^{it} - 1$, тогда необходимо уметь оценивать сверху величину $|e^{pz} - (1 + pz)|$. Учитывая, что при $n \to \infty$ и фиксированном $\lambda > 0$, $p \to 0$, то надо уметь это делать в малой окрестности 0.

Лемма 11.10. Если z - комплексное число, такое что $|z| \leq 1$, то

$$|e^z - (1 + z)| \leq (e - 2)|z|^2.$$

Действительно,

$$|e^z - (1 + z)| = \left| \frac{z^2}{2!} + \frac{z^3}{3!} + \cdots \right| \leq |z|^2 \left(\frac{1}{2!} + \frac{1}{3!} + \cdots \right) = |z|^2 \cdot (e - 2).$$

Применим эту лемму к доказательству теоремы. Так как

$$z = e^{it} - 1 = e^{\frac{it}{2}} \left(e^{\frac{it}{2}} - e^{\frac{-it}{2}} \right) = 2i \sin \frac{t}{2} e^{\frac{it}{2}},$$

то $|z| \leq 2|\sin \frac{t}{2}|$ и при $p \leq \frac{1}{2}$ имеем $|pz| \leq 2p \leq 1$.

Следовательно, можно применить лемму 11.10:

$$|e^{pz} - (1 + pz)| \leq |pz|^2(e - 2) \leq p^2 \cdot 4 \sin^2 \frac{t}{2} \cdot (e - 2).$$

Таким образом получаем

$$\Delta_m \leq 2(e - 2)p^2 n \cdot \frac{1}{2\pi} \int\limits_{-\pi}^{\pi} 2 \sin^2 \frac{t}{2} dt = 2(e - 2)p^2 n \leq \frac{3}{2} \cdot \frac{\lambda^2}{n}.$$

Теорема 11.9 доказана. □

Рассмотрим теперь общий случай схемы Пуассона.

Теорема 11.10. *Пусть $A_1, A_2, \dots$ — независимые случайные события, такие, что $p_k = \mathbf{P}(A_k) \le \frac{1}{2}$. Тогда*

$$\Delta_m = |\mathbf{P}\{\mu_n = m\} - \mathbf{P}\{Y = m\}| \le \frac{3}{2} \cdot \sum_{k=1}^{n} p_k^2.$$

Доказательство. Из леммы 11.9 следует, что

$$\Delta_m \le \frac{1}{2\pi} \int_{-\pi}^{\pi} \left| \prod_{k=1}^{n} \big(1 + p_k(e^{it} - 1)\big) - \prod_{k=1}^{n} e^{p_k(e^{it}-1)} \right| dt. \tag{11.15}$$

Для оценки подынтегральной функции воспользуемся следующей леммой.

Лемма 11.11. Пусть α_k, β_k $(k = 1, 2, \dots, n)$ — комплексные числа, такие что $|\alpha_k| \le 1,\ |\beta_k| \le 1$. Тогда

$$\left| \prod_{k=1}^{n} \alpha_k - \prod_{k=1}^{n} \beta_k \right| \le \sum_{k=1}^{n} |\alpha_k - \beta_k|.$$

Доказательство леммы проведем по индукции. При $n = 1$ утверждение леммы выполняется тривиальным образом. Допустив верность утверждения для некоторого n, для $n+1$ получаем

$$|\alpha_1 \cdot \ldots \cdot \alpha_n \alpha_{n+1} - \beta_1 \cdot \ldots \cdot \beta_n \beta_{n+1}| \le |\alpha_1 \ldots \alpha_n \alpha_{n+1} - \alpha_1 \cdots \alpha_n \beta_{n+1}| +$$

$$+ |\alpha_1 \cdots \alpha_n \beta_{n+1} - \beta_1 \cdots \beta_n \beta_{n+1}| \le |\alpha_{n+1} - \beta_{n+1}| \cdot \left| \prod_{k=1}^{n} \alpha_k \right| + |\beta_{n+1}| \cdot \left| \prod_{k=1}^{n} \alpha_k - \prod_{k=1}^{n} \beta_k \right| \le$$

$$\le |\alpha_{n+1} - \beta_{n+1}| + \sum_{k=1}^{n} |\alpha_k - \beta_k| = \sum_{k=1}^{n+1} |\alpha_k - \beta_k|.$$

Воспользовавшись этой леммой, из (11.15) получаем

$$\Delta_m \le \frac{1}{2\pi} \int_{-\pi}^{\pi} \sum_{k=1}^{n} \left| 1 + p_k(e^{it} - 1) - e^{p_k(e^{it}-1)} \right| dt.$$

Если все $p_k \le \frac{1}{2}$, то снова можно применять лемму 11.10, получим

$$\Delta_m \le \frac{1}{2\pi} \int_{-\pi}^{\pi} \sum_{k=1}^{n} p_k^2 \cdot 4 \sin^2 \frac{t}{2} (e-2) dt \le 2(e-2) \sum_{k=1}^{n} p_k^2 \frac{1}{2\pi} \int_{-\pi}^{\pi} 2 \sin^2 \frac{t}{2} dt \le \frac{3}{2} \cdot \sum_{k=1}^{n} p_k^2.$$

□

Замечание. Пуассоновское приближение имеет смысл применять, когда величина $\lambda = \sum\limits_{k=1}^{n} p_k^2$ мала.

Приведем несколько примеров.

Пример 11.15. Пусть в схеме Бернулли $n = 1000$; $p = 0,01$. Тогда $\lambda = 10$. Вычисляя по формулам биномиального распределения с точностью до пятого знака после запятой, получим

$$\mathbf{P}\{\mu_n = 10\} = 0.12574,$$

тогда как пуассоновская вероятность равна 0.12511, $\Delta_{10} = 0.00063$.

Неравенство теоремы 11.9 для этой величины дает оценку

$$\Delta_{10} \le \frac{3}{2}\frac{\lambda^2}{n} = 0,15$$

Мы видим,что оценка довольно грубая.

Пример 11.16. Пусть $n = 100$, $p = 0.01$, $\lambda = 1$. Тогда $\mathbf{P}\{\mu_n = 3\} = 0.066999$, а пуассоновская аппроксимация дает 0.061313. Здесь расхождение $\Delta_3 = 0.005686$ сравнимо с оценкой теоремы, согласно которой $\Delta_m \le 0.01$.

Глава 12

Задачи математической статистики. Основные понятия

В качестве вступительного слова к новой части нашего курса — математической статистике,— приведем цитату из статьи академика
А. Н. Колмогорова из БСЭ:

"Математическая статистика — раздел математики, посвященный математическим методам систематизации, обработки и использования статистических данных для научных и практических выводов. При этом статистическими данными называют сведения о числе объектов в какой-либо более или менее обширной совокупности, обладающих теми или иными признаками."

Ставший классическим учебник Г. Крамера "Методы математической статистики"содержит множество различных примеров статистических данных. Упомянем несколько из них:

данные о возрасте родителей для проверки гипотезы о влиянии возраста родителей на пол ребенка;

данные об уровнях воды в озере Вэнерн за 1807-1930г.г., которые использовались при проектировании плотины и т.д.

В "Таблицах математической статистики"авторов Л. Н. Большева и Н. В. Смирнова половину объема всей книги составляет текст, в котором в рецептурной форме даются рекомендации по решению прикладных задач.

Теория вероятностей дает математический аппарат для решения задач математической статистики.

Результаты наблюдений в математической статистике трактуются как значения независимых и одинаково распределенных случайных величин $X_1, X_2, \ldots, X_n$. В некоторых случаях это могут быть и случайные векторы, и зависимые случайные величины с некоторым совместным распределением вероятностей.

Как правило, мы будем отдельно разбирать следующие случаи:

- распределение случайных величин $X_1, X_2, \ldots, X_n$ непрерывно и имеет плотность распределения вероятностей;
- распределение случайных величин дискретное.

Часто доказательство теорем мы будем проводить на примере одного из этих случаев, хотя утверждения теорем могут быть верны и в более общих ситуациях.

Что касается числа наблюдений n, то различают следующие ситуации:

1) большие выборки, когда число наблюдений n велико (порядка нескольких сотен), тогда используют асимптотические теоремы теории вероятностей;
2) умеренные выборки ($30 \leq n \leq 100$);
3) малые выборки ($n \leq 30$).

В математической статистике независимые одинаково распределенные случайные величины $X_1, X_2, \ldots, X_n$ с общей функцией распределения $F(x)$, называют случайной *выборкой* объема n из генеральной совокупности с распределением $F(x)$.

Приведем также классификацию основных задач математической статистики по характеру статистических выводов.

1) Проверка статистических гипотез. Во многих задачах формулируется некоторая гипотеза (например, гипотеза о влиянии возраста родителей на пол ребенка), которая может быть принята или отвергнута на основании статистических данных.
2) Оценивание неизвестных параметров распределения. Различают оценки точечные и интервальные.
3) Определение видов зависимостей между различными объектами.

В математической статистике вместо вероятностного пространства $(\Omega, \mathbb{F}, \mathbf{P})$ рассматривают семейства вероятностных пространств $\{(\Omega, \mathbb{F}, \mathbf{P}), \mathbf{P} \in \mathbb{P}\}$, где $\mathbb{P}$ — семейство вероятностных распределений на $(\Omega, \mathbb{F})$. Это семейство распределений может быть параметрическим, например, семейство нормальных распределений с двумя параметрами (a, σ^2). В этом случае мы будем писать $\mathbb{P} = \{\mathbf{P}_\theta, \theta \in \Theta\}$, где Θ — множество допустимых значений параметра.

12.1 Сходимость по вероятности

Основным видом сходимости в математической статистике является сходимость по вероятности. Напомним определение.

Определение 12.1. Последовательность случайных величин $\{Y_n\}_{n=1}^{\infty}$ сходится к случайной величине Y по вероятности ($Y_n \xrightarrow{p} Y$), если для любого $\varepsilon > 0$

$$\mathbf{P}\{\omega : |Y_n(\omega) - Y(\omega)| \geq \epsilon\} \to 0 \tag{12.1}$$

при $n \to \infty$.

Заметим, что выполнение (12.1) равносильно тому, что

$$\mathbf{P}\{\omega : |Y_n(\omega) - Y(\omega)| < \varepsilon\} \to 1 \tag{12.2}$$

Определение 12.2. Последовательность случайных величин $\{W_n\}_{n=1}^{\infty}$ называется *ограниченной по вероятности,* если для любого $\varepsilon > 0$ существует $K = K(\varepsilon)$ такое, что при всех n

$$\mathbf{P}\{\omega : |W_n(\omega)| \leq K\} \geq 1 - \varepsilon. \tag{12.3}$$

В качестве самостоятельных упражнений докажите следующие свойства сходимости по вероятности.

Задача 12.1. Если последовательность случайных величин $\{a_n\}$ ограничена по вероятности, а $b_n \xrightarrow{p} 0$, то $a_n b_n \xrightarrow{p} 0$.

Задача 12.2. Если последовательности случайных величин $\{a_n\}$, $\{b_n\}$ ограничены по вероятности, то последовательность $\{c_n\}$ $(c_n = a_n b_n)$ также является ограниченной по вероятности.

Сформулируем несколько утверждений, касающихся этого вида сходимости, которые нам понадобятся в дальнейшем.

Пусть $F_U(x)$ обозначает функцию распределения случайной величины U, а соотношение $F_{Y_{n(x)}} \Rightarrow F_Y(x)$ – слабую сходимость функций распределения.

Теорема 12.1. *Если $Y_n \xrightarrow{p} Y$, то $F_{Y_n}(x) \Rightarrow F_Y(x)$, то есть из сходимости по вероятности следует слабая сходимость функций распределения.*

Теорема 12.2. *Пусть $F_{W_n}(x) \Rightarrow F_W(x)$, $\quad Y_n \xrightarrow{p} 0$. Тогда*

$$F_{W_n+Y_n}(x) \Rightarrow F_W(x).$$

Докажите эти утверждения самостоятельно.

Теорема 12.3. *Пусть $h(x)$ — всюду непрерывная функция, а $Y_n \xrightarrow{p} 0$. Тогда для всех a* $\quad h(a+Y_n) \xrightarrow{p} h(a)$.

Доказательство. В соответствии с определением сходимости по вероятности необходимо показать, что для любого $\varepsilon > 0$ при $n \to \infty$ выполняется

$$\mathbf{P}\{|h(a+Y_n) - h(a)| \geq \varepsilon\} \to 0. \tag{12.4}$$

Из непрерывности функции $h(x)$ в точке $x = a$ следует, что $\forall\varepsilon > 0$ существует $\lambda = \lambda(\varepsilon)$ такое, что если $|x-a| < \lambda$, то $|h(x) - h(a)| < \varepsilon$.

Представим вероятность в (12.4) следующим образом:

$$\mathbf{P}\{\omega: \ |h(a+Y_n) - h(a)| \geq \varepsilon\} = \mathbf{P}\{\omega: \ |h(a+Y_n) - h(a)| \geq \varepsilon, \ |Y_n| < \lambda\} +$$

$$+\mathbf{P}\{\omega: \ |h(a+Y_n) - h(a)| \geq \varepsilon, \ |Y_n| \geq \lambda\} = \mathbf{P}(A_n(\varepsilon,\lambda)) + \mathbf{P}(B_n(\varepsilon,\lambda)).$$

Поскольку $Y_n \xrightarrow{p} 0$, то для $\forall\lambda$, а значит, и для выбранного $\lambda = \lambda(\varepsilon)$, выполняется

$$\mathbf{P}\{\omega: |Y_n| \geq \lambda\} \to 0$$

при $n \to \infty$. Следовательно,

$$\mathbf{P}(B_n(\varepsilon,\lambda)) \leq \mathbf{P}\{|Y_n| \geq \lambda\} \to 0.$$

Рассмотрим другое слагаемое. Поскольку если $\omega \in A_n(\varepsilon,\lambda)$, то в этом случае $|Y_n(\omega)| < \lambda$, и значит, $|(Y_n + a) - a| < \lambda$. Но тогда в соответствии с выбором λ

$$|h(a+Y_n) - h(a)| < \varepsilon.$$

Следовательно, $\mathbf{P}(A_n(\varepsilon,\lambda)) = 0$, что и завершает доказательство теоремы. $\square$

Теорема 12.4. *Если $Y_n \xrightarrow{p} a$ при $n \to \infty$, то последовательность случайных величин$\{Y_n\}$ ограничена по вероятности.*

Доказательство. Пусть $Y_n \xrightarrow{p} 0$, тогда из (12.2) следует, что
$\forall\varepsilon > 0, \qquad \forall\delta > 0 \quad \exists N = N(\varepsilon,\delta) \quad$ такой, что $\quad \forall n > N$

$$\mathbf{P}\{\omega: \ |Y_n - a| \leq \delta\} > 1 - \frac{1}{2}\varepsilon.$$

Выберем $M^* = \max(|a+\delta|, |a-\delta|)$, тогда для всех $n > N$

$$\mathbf{P}\{\omega: \ |Y_n| \leq M^*\} > 1 - \frac{1}{2}\varepsilon.$$

Задача 12.3. Покажите, что для произвольной случайной величины X и любого $\varepsilon > 0$ существует $M_X = M_X(\varepsilon)$ такое, что

$$\mathbf{P}\{|X| \le M_X\} > 1 - \varepsilon.$$

Пользуясь этим фактом, получим, что существуют константы $M_1, M_2, \dots, M_N$ такие, что

$$\mathbf{P}\{|Y_1| \le M_1\} > 1 - \frac{\varepsilon}{2N}, \cdots, \mathbf{P}\{|Y_N| \le M_N\} > 1 - \frac{\varepsilon}{2N},$$

Но тогда для $M_0 = \max(M_1, \cdots, M_N)$ выполняется

$$\mathbf{P}\{|Y_1| \le M_0, \cdots, |Y_n| \le M_0\} = 1 - \mathbf{P}\left\{\bigcup_{i=1}^{N}\{|Y_i| > M_0\}\right\} \ge$$

$$\ge 1 - \sum_{i=1}^{N}\mathbf{P}\{|Y_i| > M_0\} > 1 - \frac{\varepsilon}{2N} \cdot N = 1 - \frac{\varepsilon}{2}.$$

Следовательно, если выбрать $M = \max(M^*, M_0)$, то для всех n

$$\mathbf{P}\{|Y_n| \le M\} > 1 - \varepsilon,$$

то есть последовательность $\{Y_n\}$ ограничена по вероятности. □

12.2 Асимптотическая нормальность

Определение 12.3. Последовательность случайных величин Y_n асимптотически нормальна с параметрами (A_n, B_n), если для всех x при $n \to \infty$

$$\mathbf{P}\left\{\frac{Y_n - A_n}{B_N} < x\right\} \to \Phi(x). \tag{12.5}$$

В определении предполагается, что $-\infty < A_n < \infty, \quad B_n > 0$.

Пример 12.1. Пусть μ_n - число успехов в n испытаниях Бернулли с вероятностью успеха в одном испытании p $(0 < p < 1)$. Из теоремы Муавра - Лапласа следует, что μ_n асимптотически нормальна с параметрами $(np, \sqrt{npq})$, а относительная частота успеха $\frac{\mu_n}{n}$ асимптотически нормальна с параметрами $(p, \sqrt{\frac{pq}{n}})$.

Задача 12.4. Докажите, что из асимптотической нормальности последовательности случайных величин Y_n с параметрами (A_n, B_n) вытекает асимптотическая нормальность $U_n = \alpha Y_n + \beta$ с параметрами $(\alpha A_n + \beta, \alpha B_n)$.

Теорема 12.5. *Пусть*

1) последовательность случайных величин Y_n асимптотически нормальна с параметрами (a, b_n), причем $b_n \to 0$ при $n \to \infty$,

2) функция $g(x)$ дважды непрерывно дифференцируема на всей прямой.

Тогда последовательность случайных величин $W_n = g(Y_n)$ асимптотически нормальна с параметрами $(g(a); b_n|g'(a)|)$.

Доказательство. Доказательство теоремы довольно громоздко, его можно посмотреть в учебнике Б.А.Севастьянова. Мы же разберем простой случай, когда случайные величины Y_n нормально распределены.

Пусть $Z_n \sim N(0,1), \quad Y_n = a + b_n Z_n, \quad b_n > 0, \quad b_n \to 0$. Покажем в этих дополнительных предположениях справедливость теоремы 5.

Для дважды непрерывно дифференцируемой функции $g(x)$ можно написать разложение по формуле Тейлора в окрестности точки a :

$$g(Y_n) = g(a) + b_n Z_n g'(a) + \frac{b_n Z_n^2}{2} \cdot g''(a + \theta b_n Z_n),$$

где $|\theta| \le 1$.

Для определенности будем считать, что $g'(a) > 0$. Тогда

$$\frac{g(Y_n) - g(a)}{b_n g'(a)} = Z_n + \frac{b_n^2 Z_n}{2} \cdot \frac{1}{g'(a)} \cdot g''(a + \theta b_n Z_n). \tag{12.6}$$

Для того,чтобы доказать утверждение теоремы, достаточно показать, что последнее слагаемое в (12.6) сходится по вероятности к 0.

Так как для всех $n \quad F_{Z_n}(x) = \Phi(x)$, то для любого $\varepsilon > 0 \quad \exists M = M(\varepsilon)$ такое, что

$$\mathbf{P}\{|Z_n| \le M\} > 1 - \varepsilon$$

для всех n одновременно. Значит, последовательность $\{Z_n\}$ ограничена по вероятности. Но тогда (утверждение задачи 12.1) $b_n Z_n \xrightarrow{p} 0$, и,стало быть,

$$\frac{b_n Z_n^2}{2g'(a)} \xrightarrow{p} 0.$$

Кроме того, из непрерывности второй производной функции $g(x)$ следует сходимость

$$g''(a + \theta b_n Z_n) \xrightarrow{p} g''(a).$$

Следовательно, по теореме 12.4 последовательность $g''(a + \theta b_n Z_n)$ ограничена по вероятности. Учитывая все, что было сказано выше, получим, что второе слагаемое в (12.6) сходится по вероятности к 0. Но тогда в силу утверждения теоремы 2

$$\frac{g(Y_n) - g(a)}{b_n g'(a)} \xrightarrow{d} Z,$$

или, что то же самое, $g(Y_n)$ асимптотически нормальна с параметрами $(g(a), b_n g'(a))$, что и требовалось доказать. □

12.3 Некоторые важные преобразования случайных величин

Пусть U — случайная величина, равномерно распределенная на $[0,1]$ с плотностью

$$p_U(x) = \begin{cases} 1, & x \in [0,1] \\ 0, & x \notin [0,1] \end{cases}$$

и функцией распределения

$$F_U(x) = \begin{cases} 0, & x \le 0 \\ x, & 0 < x \le 1 \\ 1, & x > 1. \end{cases} \tag{12.7}$$

Лемма 12.1. Пусть X — случайная величина с непрерывной функцией распределения $F(x)$, строго возрастающей в окрестности любой точки x_0, где $0 < F(x_0) < 1$, $F^{-1}(x)$ — функция, обратная к $F(x)$. Тогда

1) случайная величина $Y = F(X)$ имеет равномерное распределение на $[0,1]$,

2) случайная величина $V = F^{-1}(U)$ имеет функцию распределения $F(x)$.

Доказательство. 1). Поскольку $0 \le F(x) \le 1$, то $F_Y(x) = 0$ при $x \le 0$ и $F_Y(x) = 1$ при $x > 1$. Для $0 < x < 1$ имеем

$$F_Y(x) = \mathbf{P}\{Y < x\} = \mathbf{P}\{F(X) < x\} = \mathbf{P}\{X < F^{-1}(x)\} = F\left(F^{-1}(x)\right) = x,$$

то есть функция распределения случайной величины Y совпадает с функцией равномерного распределения на $[0,1]$.

2). Функция распределения случайной величины V равна

$$F_V(x) = \mathbf{P}\{V < x\} = \mathbf{P}\{F^{-1}(U) < x\} = \mathbf{P}\{U < F(x)\} = F(x),$$

что и требовалось доказать. □

Задача 12.5. Докажите, что утверждение 1) леммы верно для всякой непрерывной функции распределения $F(x)$.

Замечание. Утверждение 2) леммы используется при моделировании случайных величин с заданной непрерывной функцией распределения $F(x)$. Для этого надо уметь моделировать равномерно распределенные случайные величины и применить к ним обратную функцию $F^{-1}(x)$.

Определение 12.4. Пусть $0 < \gamma < 1$. Решение уравнения $F(x_\gamma) = \gamma$ называется **квантилью уровня** γ.

Если $\gamma = 1/2$, то $x_{1/2}$ называется *медианой* распределения, для нее

$$\mathbf{P}\{X \geq x_{1/2}\} = \mathbf{P}\{X \leq x_{1/2}\} = 1/2. \tag{12.8}$$

Если $\gamma = 1/4,\ 2/4,\ 3/4,$ то $x_{1/4},\ x_{2/4},\ x_{3/4}$ называются квартилями распределения. Если $\gamma = k/10$ $(k = 1, 2, \ldots, 9)$, то $x_{k/10}$ называют децилями распределения.

Медиана является центром распределения в смысле (12.8), а квартили и децили служат характеристиками разброса относительно медианы, так например,

$$\mathbf{P}\{x_{1/4} \leq X \leq x_{3/4}\} = 1/2.$$

Эти характеристики широко используются в экономико – статистических методах изучения свойств различных совокупностей. Скажем, если рассматривается распределение населения по уровню годового дохода, то нижняя дециль $x_{1/10}$ показывает, какая часть населения имеет наиболее низкий доход (доля бедных людей), а верхняя квантиль $x_{9/10}$ отделяет наиболее богатых.

На рисунках 1 и 2 изображены распределения по доходам в двух различных совокупностях. В первой из них больше половины населения имеют крайне низкий доход и очень мало богатых. Во второй — картина не такая "удручающая": гораздо меньше доля очень бедных и больше богатых.

Рис. 12.1: Распределение населения по уровню доходов.

К наиболее часто используемым квантилям относятся квантили уровней

$$\gamma = 0.0001; \quad 0.005; \quad 0.01; \quad 0.025; \quad 0.05;$$
$$0,9999; \quad 0.995; \quad 0.99; \quad 0.975; \quad 0.95.$$

В учебнике Б.А.Севастьянова именно для этих значений γ приводятся значения квантилей основных распределений. Более полные таблицы можно найти в книге Л.Н.Большева и Н.В.Смирнова "Таблицы математической ствтистики".

Лемма 12.2. Если функция распределения $F(x)$ непрерывна во всех точках x, то она равномерно непрерывна на всей числовой прямой.

Проведите самостоятельно доказательство этого утверждения.

Лемма 12.3. Пусть случайные величины $X_1, X_2, \ldots, X_n$ — независимы, одинаково распределены с непрерывной функцией распределения $F(x)$. Тогда вероятность того, что какие – либо две случайные величины примут одинаковые значения, равна 0.

Доказательство. Обозначим $B = \{\omega : \ \exists\ i \neq j$ такие, что $X_i(\omega) = X_j(\omega)\}$. Тогда

$$B \subseteq \bigcup_{1 \le i < j \le n} \{\omega : X_i(\omega) = X_j(\omega)\},$$

и, значит,

$$\mathbf{P}(B) \le \sum_{1 \le i < j \le n} \mathbf{P}\{\omega : X_i(\omega) = X_j(\omega)\}.$$

Покажем, что каждое слагаемое в этой сумме равно 0. Пусть $h > 0$. Тогда

$$\{\omega : \ X_i(\omega) = X_j(\omega)\} \subseteq \bigcup_{k=-\infty}^{\infty} \{\omega : \ kh \le X_i(\omega) < (k+1)h, \ kh \le X_j(\omega) < (k+1)h\}.$$

Следовательно,

$$\mathbf{P}\{X_i = X_j\} \le \sum_{k=-\infty}^{\infty} \mathbf{P}\{kh \le X_i < (k+1)h\} \cdot \mathbf{P}\{kh \le X_j < (k+1)h\} =$$

$$= \sum_{k=-\infty}^{\infty} (F(h(k+1)) - F(hk))^2 \le \sup_k (F((k+1)h) - F(kh)) \cdot 1.$$

В силу леммы 12.2, выбирая h, мы можем сделать эту величину сколь – угодно малой: $\forall \epsilon > 0 \quad \exists\, h(\epsilon)$ такое, что $\mathbf{P}\{X_i \ne X_j\} < \epsilon$. Устремив $\epsilon \to 0$, получим нужное утверждение. $\square$

Из леммы 12.3 следует, что если мы наблюдаем независимые случайные величины $X_1.X_2, \dots, X_n$ с одинаковой **непрерывной** функцией распределения $F(x)$, то с вероятностью 1 все наблюдения различны и их можно расположить в порядка возрастания. Полученную упорядоченную выборку будем обозначать

$$X_{(1)} < X_{(2)} < \ldots < X_{(n)}$$

и называть вариационным рядом, члены вариационного ряда $X_{(i)}$ — вариантами. Заметим, что $X_{(1)} = \min\limits_{1 \le i \le n} X_i, \quad X_{(n)} = \max\limits_{1 \le i \le n} X_{(i)}$.

Из леммы 12.1 вытекает, что

$$F(X_1), F(X_2), \dots, F(X_n)$$

— независимы и равномерно распределены на $[0, 1]$ случайные величины, причем

$$F(X_{(1)}) < F(X_{(2)}) < \ldots F(X_{(n)}).$$

12.4 Эмпирическая функция распределения

Определение 12.5. Эмпирической (или **выборочной**) функцией распределения называется функция действительной переменной x

$$F_n(x; \omega) = F_n(x) = \frac{1}{n} \sum_{i=1}^{n} I_{\{X_i < x\}},$$

равная числу наблюдений, меньших x, деленному на общее количество наблюдений.

Эмпирическая функция распределения при любом фиксированном $x \in \mathbb{R}$ является случайной величиной, так как это среднее арифметическое случайных величин

$$I_{\{X_i<x\}} = \begin{cases} 1, & \text{если} \quad X_i(\omega) < x, \\ 0, & \text{если} \quad X_i(\omega) \geq x. \end{cases}$$

Чтобы подчеркнуть это, мы будем также использовать запись $F_n(x) = F_n(x, \omega)$.

С другой стороны, если в результате наблюдений реализуется конкретный элементарный исход ω, случайные величины $X_1(\omega), X_2(\omega), \ldots, X_n(\omega)$ принимают конкретные значения

$$x_1, x_2, \ldots, x_n.$$

По этим значениям можно построить **вариационный ряд**

$$x_{(1)} < x_{(2)} < \ldots < x_{(n)},$$

и тогда эмпирическая функция $F_n(x, \omega)$ представляет собой ступенчатую функцию, изменяющуюся скачками величины $1/n$ в точках $x_{(i)}$.

Эмпирическая функция распределения является примером случайной функции или, как говорят еще по – другому, случайного процесса, когда элементарный исход ω определяет не число, а функцию от x .

Замечание. Поскольку обычно результаты наблюдений записываются с округлением до какого - то знака (часто определяемого точностью измерительного прибора), то в выборках встречаются одинаковые значения. В этом случае в соответствующей точке $x_{(i)}$ изменение эмпирической функции будет равно не $1/n$, а k/n, где k — кратность данного значения.

Для эмпирической функции распределения выполнены соотношения:

$$\mathbf{E}F_n(x) = \frac{1}{n}\sum_{i=1}^{n}\mathbf{E}I\{X_i < x\} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{P}\{X_i < x\} = F(x),$$

$$\mathbf{D}F_n(x) = \frac{1}{n^2}\sum_{i=1}^{n}\mathbf{D}I\{X_i < x\} = \frac{1}{n}F(x)(1 - F(x)),$$

и при любых x

$$F_n(x) \xrightarrow{p} F(x).$$

Отметим, что перечисленные свойства справедливы для случайной выборки с любой функцией распределения $F(x)$, если же распределение наблюдений непрерывно, то можно сформулировать более сильные утверждения.

Теорема 12.6. *Пусть $X_1, X_2, \dots, X_n$ - независимые одинаково распределенные случайные величины с непрерывной функцией распределения $F(x)$, $\quad F_n(x)$ — эмпирическая функция распределения. Обозначим*

$$D_n(\omega) = \sup_{-\infty<x<\infty} |F_n(x,\omega) - F(x)|.$$

Тогда распределение $D_n(\omega)$ не зависит от $F(x)$.

Ведем специальное обозначение для функции распределения случайной величины $\sqrt{n}D_n(\omega)$. Пусть

$$K_n(\lambda) = \mathbf{P}\{\omega : \ \sqrt{n}D_n(\omega) < \lambda\}.$$

Из теоремы 12.6 следует, что функция $K_n(\lambda)$ при всех n и λ не зависит от $F(x)$.

Теорема 12.7. *(Колмогоров А.Н.) Для $\forall\lambda > 0$ при $n \to \infty$*

$$K_n(\lambda) \to K(\lambda) = \sum_{j=-\infty}^{\infty} (-1)^j e^{-2j^2\lambda^2}.$$

Функция $K(\lambda)$ называется функцией распределения Колмогорова. Перепишем ее несколько в ином виде:

$$K(\lambda) = 1 + 2\sum_{j=1}^{\infty}(-1)^j e^{-2j^2\lambda^2}. \tag{12.9}$$

Для ее вычисления можно использовать "Таблицы математической статистики"Л.Н.Большева, Н.В.Смирнова. Кстати, в этой же книге имеются таблицы функции распределения $K_n(\lambda)$ статистики $\sqrt{n}D_n$ и при малых n.

Отметим, что при $\lambda \geq 1$ знакопеременный ряд (12.9) сходится очень быстро, ошибка приближения конечной суммой для $K(\lambda)$ не превосходит по абсолютной величине первого отброшенного члена и совпадает с ним по знаку.

Например, при $k \geq 1.5$

$$K(\lambda) \approx 1 - 2e^{-2\lambda^2}$$

и погрешность $0 < K(\lambda) - (1 - 2e^{-2\lambda^2}) \leq 2e^{-8\lambda^2} < 10^{-7}$.

Доказательство теоремы Колмогорова достаточно сложно, поэтому приводить его не будем.

Доказательство теоремы 12.6 проведем при дополнительном предположении, что функция $F(x)$ строго возрастает в окрестности любой точки x_0, где $0 < F(x_0) < 1$.

Доказательство. Пусть $0 < y < 1$ и $F(x) = y$, тогда $x = F^{-1}(y)$. В этом случае

$$I\{X_j < x\} = I\{X_j < F^{-1}(y)\} = I\{F(X_j) < y\} = \begin{cases} 1, & \text{если } F(X_j) < y, \\ 0, & \text{если } F(X_j) \geq y. \end{cases}$$

Введем новые случайные величины $U_j = F(X_j), \quad j = 1, 2, \ldots, n$. Эти случайные величины также независимы и одинаково распределены, но каждая из них имеет функцию распределения

$$G(y) = \begin{cases} 0, & y \leq 0, \\ y, & 0 < y \leq 1, \\ 1, & y > 1. \end{cases}$$

Чтобы подчеркнуть, что выборочная функция $F_n(x)$ является случайной величиной, зависящей от случайной выборки $X_1, \ldots, X_n$ с распределением $F(x)$, будем использовать запись

$$F_n(x) = F_n(x; X_1, \ldots, X_n; F).$$

Тогда

$$F_n(x) = F_n(x; X_1, \ldots, X_n; F) = \frac{1}{n}\sum_{j=1}^{n} I\{X_j < x\} = \frac{1}{n}\sum_{j=1}^{n} I\{U_j < y\} =$$

$$= F_n(y; U_1, \ldots, U_n; G),$$

и, значит,

$$D_n = \sup_{-\infty < x < \infty} |F_n(x) - F(x)| = \sup_{-\infty < y < \infty} |F_n(y; U_1, \ldots, U_n; G) - G(y)| =$$

$$= \sup_{0 \leq y \leq 1} |F_n(y; U_1, \ldots, U_n; G) - y|.$$

Но последняя величина определяется только значениями равномерно распределенных случайных величин U_j и не зависит от $F(x)$. Теорема 12.6 доказана. □

С помощью теоремы Колмогорова можно строить доверительные интервалы для неизвестной функции распределения случайной выборки, проверять гипотезы о виде распределения.

Глава 13

Проверка гипотезы о виде распределения

13.1 Критерий согласия А. Н. Колмогорова.

Пусть по независимым наблюдениям $X_1, X_2, \ldots, X_n$ с одинаковой, но неизвестной непрерывной функцией распределения $F(x)$ проверяется гипотеза

$$H_0: \ F(x) = F_0(x),$$

согласно которой неизвестная функция распределения $F(x)$ совпадает с некоторой известной непрерывной функцией распределения $F_0(x)$.

В качестве меры рассхождения наблюдаемых данных с выдвинутой гипотезой выберем статистику

$$D_n = \sup_{-\infty < x < \infty} |F_n(x) - F_0(x)|.$$

Поскольку $F_0(x)$ непрерывная функция, а $F_n(x)$ – ступенчатая, то максимальное различие между ними будет в одной из точек роста последней, т.е. будет достигаться в одном или нескольких наблюдаемых значениях X_i.

Если проверяемая гипотеза верна, то вычисленное по выборке значение D_n не может быть слишком большим, в противном случае гипотезу H_0 следует отвергнуть или хотя бы подвергнуть сомнению.

Пример 13.1. Допустим, что вычисления по результатам наблюдений дали

$$\sqrt{n} D_n = 1.73.$$

Если гипотезе H_0 верна, то при достаточно больших n по теореме Колмогорова (обычно ее используют уже при $n \geq 30$)

$$\mathbf{P_{H_0}}\{\sqrt{n}D_n \geq \lambda\} = 1 - K(\lambda)$$

и, следовательно,

$$\mathbf{P_{H_0}}\{\sqrt{n}D_n \geq 1.73\} = 1 - K(1.73) = 1 - 0.995 = 0.005.$$

Мы видим, что вероятность наблюдать такие или большие расхождения с проверяемой гипотезой мала, следовательно, вряд ли мы могли наблюдать столь маловероятное событие, и тогда проверяемая гипотеза должна быть отвергнута. При этом, отвергая гипотезу, мы могли совершить ошибку, но вероятность этой ошибки невелика, она равна 0.005.

Аналогичным образом поступают и в общем случае, только обычно вероятность ошибки неправильного решения задают заранее. А именно выбирают малое положительное число α и считают, что ошибка отвергнуть правильную гипотезу не должна превышать α. Число α называется уровнем значимости критерия. Соображения, сколь малым должно быть это число, лежат вне теории вероятностей, они определяются теми потерями, которые могут произойти при неправильном решении отвергнуть верную гипотезу. В каждой конкретной области руководствуются своими допустимыми нормами вероятностей ошибок.

Итак, уровень значимости α выбран. По таблицам значений функции $K(\lambda)$ можно найти λ_α – решение уравнения

$$1 - K(\lambda_\alpha) = \alpha.$$

Если число наблюдений n достаточно велико ($n \geqslant 30$), то

$$\mathbf{P}\{\sqrt{n}D_n \geqslant \lambda_\alpha\} \approx \alpha,$$

и далее действовать как в примере: если в результате проведенных наблюдений осуществилось событие

$$S = \{\sqrt{n}D_n > x_{1-\lambda}\}, \tag{13.1}$$

то гипотезу H_0 отвергают, в противном случае принимают. При этом вероятность отвергнуть верную гипотезу равна α.

Множество (13.1) называют критическим множеством, а λ_α – критическим значением уровня α.

Такой способ проверки гипотезы о виде распределения называют **критерием согласия Колмогорова.**

Замечание. При небольшом числе наблюдений действуют аналогичным образом с той лишь разницей, что для отыскания критического значения используют таблицы значений функции $K_n(\lambda)$.

Сформулированный критерий проверки гипотезы

$$H_0: \ F(x) = F_0(x)$$

о виде распределения называют критерием согласия Колмогорова.

Существуют и другие критерии согласия.

13.2 Критерий согласия Пирсона "хи – квадрат"

Этот критерий был предложен Карлом Пирсоном.

Пусть по – прежнему есть случайная выборка $X_1, X_2, \ldots, X_n$ с неизвестной функцией распределения $F(x)$.

Разобьем числовую прямую точками

$$-\infty = z_0 < z_1 < z_2 < \ldots < z_{r-1} < z_r = +\infty$$

на r непересекающихся интервалов. Обозначим

ν_1- число наблюдений X_i, попавших на $A_1 = (-\infty; z_1)$,
ν_2- число наблюдений X_i, попавших на $A_2 = [z_1, z_2)$,
$\ldots\ldots$
ν_r- число наблюдений X_i, попавших на $A_r = [z_{r-1}, \infty)$.

Тогда $\nu_1 + \nu_2 + \cdots + \nu_r = n$.

Пусть снова проверяется гипотеза

$$H_0: \ F(x) = F_0(x).$$

Заметим, что теперь $F_0(x)$ может быть любой функцией распределения, в том числе и дискретной.

Вычислим $p_j = P_{H_o}\{X_i \in A_j\}, \ j = 1, 2, \ldots, r$. А именно

$$p_j = F_0(z_j) - F_0(z_{j-1}).$$

Если гипотеза H_0 верна, то для $1 \le j \le r \quad \nu_j \sim B(n; p_j)$. При этом случайные величины $\frac{\nu_j - np_j}{\sqrt{np_j(1-p_J)}}$ асимптотически нормальны
с параметрами $(0, 1)$, и $\frac{\nu_j}{n} \xrightarrow{p} p_j$.

Основываясь на этих свойствах частот ν_j, Карл Пирсон предложил ввести следующую меру расхождения гипотезы H_0 с имеющимися данными:

$$\chi^2 = \sum_{j=1}^{r} \frac{(\nu_j - np_j)^2}{np_j} = \sum_{j=1}^{r} \left(\frac{\nu_j - np_j}{\sqrt{np_j q_j}} \right)^2 q_j, \tag{13.2}$$

где $q_j = 1 - p_j$.

К.Пирсон назвал эту статистику "хи – квадрат"по названию греческой буквы, которой он обозначил статистику критерия.

Теорема 13.1. *(К.Пирсон.) Если верна гипотеза H_0, то для всех $x > 0$ при $n \to \infty$*

$$\mathbf{P}\{\chi^2 < x\} \to \int_0^x p_{r-1}(y)dy,$$

где

$$p_{r-1} = \frac{y^{\frac{r-3}{2}} e^{-y/2}}{2^{\frac{r-1}{2}} \Gamma(\frac{r-1}{2})}. \tag{13.3}$$

Замечание. Отметим, что при конечных n распределение статистики χ^2 зависит от $F_0(x)$, но предельное распределение при $n \to \infty$ зависит только от r.

Замечание 1. Рассмотрим частный случай $r = 2$. Тогда

$$\chi^2 = \frac{(\nu_1 - np_1)^2}{n} + \frac{(n - \nu_1 - n(1 - p_1))^2}{n(1 - p_1)} = \frac{(\nu_1 - np_1)^2}{n} \cdot \left(\frac{1 - p_1 + p_1}{p_1(1 - p_1)} \right) =$$

$$= \left(\frac{\nu_1 - np_1}{\sqrt{np_1(1 - p_1)}} \right)^2 \xrightarrow{d} Z_1^2,$$

где Z_1 обозначает случайную величину со стандартным нормальным законом распределения. Плотность случайной величины Z_1^2 равна

$$p_1(y) = \begin{cases} \frac{y^{-1/2} e^{-y/2}}{\sqrt{2\pi}}, & \text{если } y > 0, \\ 0, & \text{если } y \le 0, \end{cases}$$

то есть плотность имеет вид (13.3) с $r = 2$. Подобное обстоятельство имеет место и при $r > 2$.

Определение 13.1. Пусть $Z_1, Z_2, \ldots, Z_k$ — независимые стандартные нормальные величины. Распределение случайной величины

$$\chi_k^2 = Z_1^2 + Z_2^2 + \cdots + Z_k^2$$

называют распределением хи-квадрат с k степенями свободы.

Задача 13.1. Покажите, что плотность распределения χ_k^2 равна

$$p_k(x) = \begin{cases} \frac{x^{k/2-1}e^{-x/2}}{2^{k/2}\Gamma(k/2)}, & \text{если } x > 0, \\ 0, & \text{если } x \le 0. \end{cases}$$

Таким образом теорема Пирсона утверждает, что если справедлива гипотеза H_0, то статистика χ^2, определенная в (13.2), сходится по распределению к χ_{r-1}^2.

Если число наблюдений n велико ($n \ge 30$), то теорема Пирсона позволяет проверять гипотезу

$$H_0: \ F(x) = F_0(x)$$

следующим образом:

1) по наблюдениям вычисляется значение статистики χ^2;

2) гипотезу H_0 отвергают, если расхождение с гипотезой χ^2 окажется велико, и принимают, если это расхождение невелико.

Критическое множество имеет вид

$$S = \{\chi^2 > x_\alpha\},$$

где критическое значение x_α находится по таблицам распределения χ_{r-1}^2 из условия

$$\mathbf{P}\{\chi_{r-1}^2 > x_\alpha\} = \alpha.$$

При таком выборе критического значения вероятность отклонить верную гипотезу равна

$$\mathbf{P}_{H_0}(S) = \mathbf{P}\{\chi_{r-1}^2 > x_\alpha\} = \alpha.$$

Замечание 2. В случае $r = 2$ критическое множество

$$S = \left\{ \left(\frac{\nu_1 - np_1}{\sqrt{np_1(1-p_1)}} \right)^2 \ge x_\alpha \right\} = \left\{ \frac{|\nu_1 - np_1|}{\sqrt{np_1(1-p_1)}} \ge \sqrt{x_\alpha} \right\}.$$

Следовательно, мы должны принимать решение в зависимости от значения статистики

$$\frac{\nu_1 - np_1}{\sqrt{np_1(1-p_1)}},$$

которая при $n \to \infty$ асимптотически нормальна с параметрами $(0, 1)$.

Пример 13.2. В учебнике Г.Крамера ([6])приведены данные о рождении мальчиков и девочек в 1935 году в Швеции. Всего в октябре родилось 6903 ребенка, среди которых было 3512 мальчика и 3391девочка. Если считать рождение девочки успехом и обозначить вероятность этого события p, то для проверки гипотезы

$$H_0: \ p = 1/2$$

можно воспользоваться критерием χ^2. Вычисления дают

$$\frac{|\nu_1 - np_1|}{\sqrt{np_1(1-p_1)}} = \frac{|3391 - 0.5 \cdot 6903|}{\sqrt{6903 \cdot 0.5 \cdot 0.5}} = 1.456.$$

Если гипотеза H_0 верна, то вероятность наблюдать данное отклонение от гипотезы равна

$$\mathbf{P}\{|Z| \geq 1.456\} = 0.1454.$$

Оснований сомневаться в справедливости выдвинутой гипотезы нет.

С другой стороны по данным Г.Крамера в апреле из общего числа 7884 родившихся детей мальчиков было 4173, девочек — 3711. В этом случае вычисления дают

$$\frac{|3711 - 0.5 \cdot 7884|}{\sqrt{7884 \cdot 0.5 \cdot 0.5}} = 5.203,$$

вероятность таких и больших отклонений оказывается равной

$$\mathbf{P}\{|Z| \geq 5.203\} = 2 \cdot 10^{-7}$$

слишком малой, поэтому гипотеза H_0 доджна быть отвергнута. К такому же выводу мы прийдем, если проанализируем данные за весь год: среди общего числа 88273 родившихся детей мальчиков было 45682, а девочек — 42591. Значение статистики критерия в этом случае оказалось равным 10.4. Вероятность наблюдать такие отклонения равна $2 \cdot 10^{-23}$, что можно считать практически невозможным.

Пример 13.3. Некоторый любознательный статистик, рассматривая витрины часовщиков, записывал время на часах, причем минуты не учитывались. Полученные данные приведены в таблице

0	1	2	3	4	5	6	7	8	9	10	11
41	34	54	39	49	45	41	33	37	41	47	39

Можно ли считать, что верна гипотеза

$$H_0: \ p_0 = p_1 = \ldots = p_{12} = \frac{1}{12}$$

о равномерности распределения. Поскольку

$$\chi^2 = \sum_{j=1}^{12} \frac{(\nu_j - 500 \cdot 1/12)^2}{41.67} = 10.00,$$

то, вычислив по таблицам

$$\mathbf{P}\{\chi^2 \geq 10.00\} = 0.53039,$$

приходим к выводу, что если гипотеза верна, то это событие происходит в среднем в 53 случаях из 100. Гипотезу следует принять.

Глава 14

Проверка параметрических гипотез. Фундаментальная лемма Неймана – Пирсона

Лаплас: "...теория вероятностей есть в сущности не что иное, как здравый смысл, сведенный к исчислению."

Мы не случайно начали эту главу с высказывания выдающегося математика, поскольку, как вы уже могли заметить, при принятии решений, выработке статистических критериев для проверки гипотез математики во многом руководствуются здравым смыслом. Нижеследующий материал станет тому дальнейшим подтверждением.

14.1 Квантили и процентные точки нормального распределения

Пусть

$$\Phi(x) = \int\limits_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-u^2/2} du$$

— функция распределения стандартного нормального закона, соответствующая случайной величине Z.

Обозначим p_α, λ_α, u_α — решения уравнений (14.1),(14.2),(14.3) соответственно:

$$\mathbf{P}\{Z \le p_\alpha\} = \alpha \Longleftrightarrow \Phi(p_\alpha) = \alpha; \tag{14.1}$$

$$\mathbf{P}\{|Z| \ge \lambda_\alpha\} = \alpha \Longleftrightarrow 2(1 - \Phi(\lambda_\alpha)) = \alpha; \tag{14.2}$$

$$\mathbf{P}\{Z \ge u_\alpha\} = \alpha \Longleftrightarrow 1 - \Phi(u_\alpha) = \alpha. \tag{14.3}$$

Поскольку между $p_\alpha,\ \lambda_\alpha,\ u_\alpha$ выполняются равенства

$$p_\alpha = u_{1-\alpha}; \quad \lambda_\alpha = u_{\alpha/2},$$

то достаточно по значениям α уметь вычислять u_α. Иногда u_α называют $\alpha \cdot 100\%$ – процентной точкой распределения.

Приведем таблицу некоторых значений u_α, к которым мы будем обращаться на протяжении последующих лекций.

α	0.0001	0.001	0.005	0.01	0.025	0.05
u_α	3,7190	3,0902	2.5758	2,3263	1.9600	1.6449

С ростом u_α значения α очень быстро убывают.

Для λ_α приведем лишь несколько значений. Если $\alpha = 0.5$, то $\lambda_{0.5} = 0.6745$. Эта точка называется *вероятным отклонением* стандартного нормального распределения, для нее

$$\mathbf{P}\{|Z| \le \lambda_{0.5}\} = \mathbf{P}\{|Z| \ge \lambda_{0.5}\} = 0.5.$$

Если $\alpha = 0.9973$, то $\lambda_{0.9973} = 3$, то есть

$$\mathbf{P}\{|Z| \le 3\} = 0.9973,$$

и почти все распределение сосредоточено на отрезке $[-3, 3]$. Для случайной величины $V \sim N(a, \sigma^2)$

$$\mathbf{P}\{|V - a| \ge 3\sigma\} = 0.0027.$$

Во многих приложениях теории вероятностей событие с такой малой вероятностью считают практически невозможным.Это служит обоснованием эмпирического правила "трех σ": если наблюдаются отклонения от среднего значения,

превышающие 3 стандартных отклонения, то гипотеза о нормальном распределении для такой случайной величины должна быть отклонена или хотя бы подвергнута сомнению.

Критерии Колмогорова и хи – квадрат К.Пирсона относятся к критериям согласия,

когда по результатам наблюдений $X_1, X_2, \ldots, X_n$ с общей функцией распределения $F(x)$ проверяется гипотеза

$$H_0: \ F(x) = F_0(x),$$

где $F_0(x)$ - некоторая фиксированная функция распределения.

Пример 14.1. Пусть проводится только одно наблюдение $X \sim N(a, 1)$. По значению X требуется проверить гипотезу

$$H_0: \ F(x) = \Phi(x) \text{ или, что тоже самое } \quad a = 0.$$

Зададим $\alpha = 0.05$ – вероятность, которой можно пренебречь. Тогда $\lambda_{0.05} = 1.96$. Если гипотеза H_0 верна, то

$$\mathbf{P}\{|X| > 1.96\} = 0.05,$$

то есть вероятность наблюдать такие значения пренебрежимо мала. Следовательно, если мы получим в ходе наблюдений, что $|X| > 1.96$, то гипотеза должна быть отвергнута. Но в такой постановке задачи непонятно, что же делать дальше.

14.2 Постановка задачи. Ошибки первого и второго рода.

В тридцатых годах прошлого века Ю.Нейман и Э.Пирсон предложили другую постановку задач. Их идеи послужили основой современной теории проверки статистических гипотез.

Проиллюстрируем разницу в подходах на предыдущем примере. Сформулируем **основную гипотезу**

$$H_0: \ a = a_0 \ (a_0 = 0)$$

и **альтернативную или конкурирующую гипотезу**

$$H_1: \ a = a_1 \ (a_1 = 2.3).$$

Обозначим $p_0(x) = \varphi(x)$ — плотность распределения X при основной гипотезе H_0 и $p_1(x) = \varphi(x - a_1)$ — плотность распределения X при конкурирующей гипотезе H_1.

Выберем некоторую границу x_0 и будем поступать следующим образом:

если $X < x_0$, то будем принимать H_0, отвергая H_1,
если $X \geq x_0$, то отвергнем H_0 и примем H_1.

При этом мы можем совершить ошибку. Ошибки бывают двух типов:
1) ошибка первого рода, когда отвергаем верную гипотезу H_0,
2) ошибка второго рода, когда принимаем H_0, а верна H_1.

Рис. 14.1: Проверка гипотез о параметре сдвига нормального распределения.

Этим ошибкам соответствуют вероятности

$$\alpha = \mathbf{P}_{a_0}\{X \geq x_0\} = \int\limits_{x_0}^{\infty} p_0(x)dx,$$

$$\beta = \mathbf{P}_{a_1}\{X < x_0\} = \int\limits_{-\infty}^{x_0} p_1(x)dx.$$

На рисунке (14.2) ошибке первого рода соответствует заштрихованная площадь справа от точки x_0, ошибке второго рода — заштрихованная площадь слева от точки x_0. Из рисунка видно, что уменьшая одну вероятность, мы увеличиваем другую.

При заданном числе наблюдений нельзя одновременно сделать обе ошибки сколь – угодно малыми. Обычно задают величину вероятности первого рода α (ее называют уровнем значимости критерия) и ищут такое x_0, чтобы минимизировать вероятность ошибки второго рода β.

Поскольку выполняется равенство $\mathbf{P}\{X \geq x_0\} = \alpha$, то $x_0 = u_\alpha$, но тогда

$$\beta = \mathbf{P}_{a_1}\{X < u_\alpha\} = \mathbf{P}_{a_1}\{X - a_1 < u_\alpha - a_1\} = \Phi(u_\alpha - a_1) = -u_\beta.$$

При проведении выкладок мы использовали соотношения $a_0 = 0 < a_1 = 2.3$. Таким образом, если α и β заданы заранее, то должны выполняться неравенства

$$u_\alpha - a_1 \leq -u_\beta \Longleftrightarrow u_\alpha + u_\beta \leq a_1,$$

то есть такие вероятности ошибок α и β возможны лишь в том случае, когда гипотеза и альтернатива достаточно далеки друг от друга.

14.3 Лемма Неймана – Пирсона

Эта лемма может быть сформулирована несколькими способами. Приведем наиболее удобную с нашей точки зрения.

Мы начнем со случая конечного вероятностного пространства, когда $\Omega = \{\omega_1, \omega_2, \dots, \omega_s\}$ состоит из конечного числа элементарных исходов, алгебра $\mathcal{A}$ содержит все подмножества и два возможных распределения вероятностей:

$$\mathbf{P}_0(A) = \sum_{\omega \in A} p_0(\omega),$$

$$\mathbf{P}_1(A) = \sum_{\omega \in A} p_1(A)$$

на измеримом пространстве $(\Omega, \mathcal{A})$.

Для проверки основной гипотезы

$$H_0 : \ \mathbf{P} = \mathbf{P}_0$$

против конкурирующей

$$H_1 : \ \mathbf{P} = \mathbf{P}_1$$

строят критическое множество S и далее действуют по следующему правилу: если результат наблюдений $\omega \in S$, то гипотезу H_0 отвергают и принимают H_1, если же $\omega \notin S$, то принимают H_0 и отвергают H_1. Такой способ действий называется критерием S. Каждому выбору S соответствуют вероятности ошибок первого и второго рода:

$$\alpha(S) = \mathbf{P}_0(S) \text{ и } \ \beta(S) = \mathbf{P}_1(\overline{S}).$$

Теорема 14.1. *Пусть* $0 < \alpha < 1$ *и существует* C_α *такое, что* $\mathbf{P}_0\{\omega : \ \frac{p_1(\omega)}{p_0(\omega)} \geq c_\alpha\} = \alpha$. *Выберем*

$$S^* = \{\omega : \ \frac{p_1(\omega)}{p_0(\omega)} \geq c_\alpha\}. \tag{14.4}$$

Тогда для любого критерия S *уровня* α ($\mathbf{P_0}(S) = \alpha$)

$$\mathbf{P}_1(\overline{S^*}) \leq \mathbf{P}_1(\overline{S}).$$

Доказательство. Пусть S – произвольный критерий уровня α, а S^* определяется в (14.4), тогда

$$\alpha = \mathbf{P}_0(S^*) = \mathbf{P}_0(S^* \cap \overline{S}) + \mathbf{P}_0(S^* \cap S),$$

$$\alpha = \mathbf{P}_0(S) = \mathbf{P}_0(\overline{S^*} \cap S) + \mathbf{P}_0(S^* \cap S).$$

Откуда получаем

$$\mathbf{P}_0(S^* \cap \overline{S}) = \mathbf{P}_0(\overline{S^*} \cap S).$$

Далее

$$1 - \beta(S^*) = \mathbf{P}_1(S^*) = \P_1(S^* \cap \overline{S}) + \mathbf{P}_1(S^* \cap S),$$

$$1 - \beta(S) = \mathbf{P}_1(S) = \mathbf{P}_1(S \cap \overline{S^*}) + \mathbf{P}_1(S^* \cap S).$$

мы желаем доказать, что $\beta(S^*) \leq \beta(S)$ или, что то же самое, $\mathbf{P}_1(S^* \cap \overline{S}) \geq \mathbf{P}_1(S \cap \overline{S^*})$.

Действительно

$$\mathbf{P}_1(S^* \cap \overline{S}) = \sum_{\omega \in S^* \cap \overline{S}} p_1(\omega) \geq c_\alpha \sum_{\omega \in S^* \cap \overline{S}} p_0(\omega) = c_\alpha \mathbf{P}_0(S^* \cap \overline{S}) = c_\alpha \mathbf{P}_0(\overline{S^*} \cap S) =$$

$$= c_\alpha \sum_{\omega \in \overline{S^*} \cap S} p_0(\omega) \geq \sum_{\omega \in \overline{S^*} \cap S} p_1(\omega) = \mathbf{P}_1(\overline{S^*} \cap S).$$

Теорема доказана. □

Определение 14.1. Число $1 - \beta(S)$ называют **мощностью критерия** S.

Критерий S^* в лемме Неймана – Пирсона минимизирует вероятность ошибки второго рода среди всех критериев уровня α. А это означает, что в данном классе критериев S^* максимизирует мощность, поэтому его называют **наиболее мощным критерием уровня** α.

Лемма Неймана-Пирсона показывает, как устроено наиболее мощное критическое множество S^*:

$$S^* = S^*(c_\alpha) = \left\{ \omega : \frac{p_1(\omega)}{p_0(\omega)} \geq c_\alpha \right\}. \tag{14.5}$$

В критическое множество S^* включаются те точки, вероятность которых при H_1 существенно больше, чем при H_0. Заметим, что константа c_α определяется из условия

$$\mathbf{P_0}(S^*(c_\alpha)) = \alpha,$$

и поэтому для дискретных распределений c_α существует не для всех α.

Лемма Неймана-Пирсона была доказана нами для дискретных случайных величин, но ее формулировка и доказательство легко переносятся на непрерывный случай.

В классических задачах математической статистики как правило множество элементарных исходов $\Omega = \mathbb{R}^n$. В качестве класса случайных событий обычно выбирают борелевскую $\sigma-$алгебру в $\mathbb{R}^n$.

Пусть на $\mathbb{R}^n$ заданы два распределения вероятностей $\mathbf{P}_0$ и $\mathbf{P}_1$.

В этих случаях результат наблюдения $\omega = (x_1, x_2, \ldots, x_n)$- точка $n-$мерного пространства, а $X_i(\omega)$ - кординатные случайные величины, то есть $X_1(\omega) = x_1, \ldots, X_n(\omega) = x_n$.

Распределения будем рассматривать или дискретные, или непрерывные, имеющие плотность распределения вероятностей.

Для $k = 0, 1$ $\hat{p}_k(x_1, x_2, \ldots, x_n)$ будет обозначать либо совместную вероятность $\mathbf{P}_k\{X_1 = x_1, \ldots, X_n = x_n\}$, либо совместную плотность распределения компонент случайного вектора $\vec{X} = (X_1, \ldots, X_n)$. Если $X_1, \ldots, X_n$ - независимые, одинаково распределенные случайные величины, то

$$\hat{p}_k(x_1, x_2, \ldots, x_n) = p_k(x_1) \cdot p_k(x_2) \cdots p_k(x_n),$$

где $p_k(x_i)$ обозначают или $\mathbf{P}_k\{X_i = x_i\}$, или плотность распределения случайной величины X_i при гипотезе H_k. Тогда наиболее мощное критическое множество (14.5) перепишется в виде

$$S^* = S^*(c_\alpha) = \left\{ \frac{\widehat{p_1}(X_1, \ldots, X_n)}{\widehat{p_0}(X_1, \ldots, X_n)} \geq c_\alpha \right\}. \tag{14.6}$$

14.4 Проверка гипотез о параметрах нормального распределения

Пусть $X_1, \ldots X_n$ - независимые случайные величины, распределенные одинаково по нормальному закону $N(a, \sigma^2)$. Рассмотрим задачу о различении двух простых гипотез:

$$\begin{array}{ll} H_0: & a = a_0, \\ H_1: & a = a_1 \ (a_1 > a_0). \end{array}$$

Отметим, что

$$a_0 = \mathbf{E}_0 X_i, \quad a_1 = \mathbf{E}_1 X_i, \quad \sigma^2 = \mathbf{D}_0 X_i = \mathbf{D}_1 X_i.$$

Индексы у математического ожидания и дисперсии показывают, по какому из распределений $\mathbf{P}_0$ или $\mathbf{P}_1$ вычислялись эти числовые характеристики.

Согласно лемме Неймана-Пирсона наиболее мощный критерий в этом случае имеет вид

$$S^* = \left\{ \omega : \frac{p_1(X_1)\cdots p_1(X_n)}{p_0(X_1)\cdots p_0(X_n)} \geq c \right\} = \left\{ \omega : \sum_{j=1}^{n} \ln \frac{p_1(X_j)}{p_0(X_j)} \geq c' \right\},$$

где $c' = \ln c$.

Обозначим

$$W_j = \ln \frac{p_1(X_j)}{p_0(X_j)},$$

при этом случайные величины $W_1, \ldots, W_n$ независимы и одинаково распределены. Критическое множество можно переписать следующим образом

$$S^* = \{W_1 + W_2 + \cdots + W_n \geq c'\}.$$

Вероятности ошибок первого и второго рода запишутся теперь как

$$\mathbf{P}_0\{W_1 + \cdots + W_n \geq c'\} = \alpha, \quad \mathbf{P}_1\{W_1 + \cdots + W_n < c'\} = \beta.$$

Для нормального распределения все вычисления можно провести точно, что мы продемонстрируем позже. Для других распределений будем вычислять искомые вероятности при $n \to \infty$, пользуясь центральной предельной теоремой.

Пусть по – прежнему u_α — корень уравнения

$$1 - \Phi(u_\alpha) = \alpha,$$

тогда при $n \to \infty$

$$\mathbf{P}_0\{W_1 + \cdots + W_n > c'\} = \mathbf{P}_0 \left\{ \frac{W_1 + \cdots + W_n - n\mathbf{E}_0 W_1}{\sqrt{n\mathbf{D}_0 W_1}} > \frac{c' - n\mathbf{E}_0 W_1}{\sqrt{n\mathbf{D}_0 W_1}} \right\} \approx$$

$$\approx 1 - \Phi\left(\frac{c' - n\mathbf{E}_0 W_1}{\sqrt{n\mathbf{D}_0 W_1}} \right) = \alpha.$$

Из уравнения

$$u_\alpha = \frac{c' - n\mathbf{E}_0 W_1}{\sqrt{n\mathbf{D}_0 W_1}}$$

находим

$$c' = n\mathbf{E}_0 W_1 + u_\alpha \sqrt{n\mathbf{D}_0 W_1}.$$

Критерий S^* в этом случае называется асимптотическим, поскольку

$$\mathbf{P}_0\{W_1 + \cdots + W_n\} \to \alpha$$

при $n \to \infty$.

Для нормального распределения

$$p_0(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-a_0)^2}{2\sigma^2}},$$

$$p_1(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-a_1)^2}{2\sigma^2}}.$$

Поэтому

$$W_j = \ln \frac{p_1(X_j)}{p_0(X_j)} = \frac{-(X_j - a_1)^2 + (X_j - a_0)^2}{2\sigma^2} = \frac{a_1 - a_0}{\sigma^2} X_j - \frac{a_1^2 - a_0^2}{2\sigma^2}.$$

Но тогда

$$W_1 + \dots + W_n = \frac{a_1 - a_0}{\sigma^2} \sum_{j=1}^{n} X_j - \frac{n(a_1^2 - a_0^2)}{2\sigma^2}.$$

Следовательно, критическое множество можно переписать в другом виде:

$$S^* = \left\{ \sum_{j=1}^{n} W_j \geq c' \right\} = \left\{ \sum_{j=1}^{n} X_j \geq c'' \right\},$$

где $c'' = \left(c' + \frac{n(a_1^2 - a_0^2)}{2\sigma^2} \right) \cdot \frac{\sigma^2}{a_1 - a_0}$.

Воспользуемся тем, что $\sum\limits_{j=1}^{n} X_j \sim N(an; \sigma^2 n)$. Если верна гипотеза H_0, то $\mathbf{E}_0(\sum\limits_{j=1}^{n} X_j) = a_0 n,\ \mathbf{D}_0(\sum\limits_{j=1}^{n} X_j) = n\sigma^2$. Если же верна H_1, то

$$\mathbf{E}_1(\sum_{j=1}^{n} X_j) = a_1 n,\ \mathbf{D}_1(\sum_{j=1}^{n} X_j) = n\sigma^2.$$

Выберем $c'' = na_0 + u_\alpha \sigma \sqrt{n}$, тогда вероятность ошибки первого рода равна

$$\mathbf{P}_0 \left\{ \sum_{j=1}^{n} X_j \geq na_0 + u_\alpha \sigma \sqrt{n} \right\} = \mathbf{P}_0 \left\{ \frac{\sum\limits_{j=1}^{n} X_j - na_0}{\sigma\sqrt{n}} \geq u_\alpha \right\} = 1 - \Phi(u_\alpha) = \alpha.$$

Вероятность ошибки второго рода равна

$$\beta = \mathbf{P}_1 \left\{ \sum_{j=1}^{n} X_j < na_0 + u_\alpha \sigma \sqrt{n} \right\} =$$

$$\mathbf{P}_1\left\{\frac{\sum\limits_{j=1}^{n} X_j - na_1}{\sigma\sqrt{n}} < u_\alpha - \frac{n(a_1-a_0)}{\sigma\sqrt{n}}\right\} = \Phi(u_\alpha - \frac{(a_1-a_0)\sqrt{n}}{\sigma}).$$

Воспользовавшись тем, что $\Phi(-x) = 1 - \Phi(x)$, получим

$$\beta = 1 - \Phi(\sqrt{n}\frac{(a_1-a_0)}{\sigma} - u_\alpha) \to 0$$

при $n \to \infty$.

Отметим также, что β зависит от величины $\frac{a_1-a_0}{\sigma}$, характеризующей "расстояние"между гипотезами, измеряемого в масштабе σ.

Лемма 14.1. 1). При $x > 0$

$$1 - \Phi(x) \leq \frac{1}{x\sqrt{2\pi}} e^{-x^2/2}.$$

2). При $x > 1$

$$1 - \Phi(x) \geq \frac{1}{x\sqrt{2\pi}}(1 - \frac{1}{x^2})e^{-x^2/2}.$$

Доказательство.

$$1 - \Phi(x) = \frac{1}{\sqrt{2\pi}}\int\limits_x^\infty e^{-u^2/2}du \leq \frac{1}{\sqrt{2\pi}}\int\limits_x^\infty \frac{u}{x}e^{-u^2/2}du = \frac{1}{x\sqrt{2\pi}}e^{-x^2/2}.$$

Доказательство второй части утверждения леммы можно посмотреть в учебнике Феллера, том 1. $\square$

Воспользовавшись леммой, получим оценку для вероятности ошибки второго рода:

$$\beta \leq ce^{-n/2},$$

то есть эта вероятность убывает с ростом числа наблюдений n как геометрическая прогрессия.

Выясним, сколько нужно иметь наблюдений, чтобы вероятность ошибки второго рода была меньше заданного числа β. Для этого решим неравенство

$$\Phi(-\sqrt{n}\frac{a_1-a_0}{\sigma} + u_\alpha) \leq \beta \Longleftrightarrow -\sqrt{n}\frac{a_1-a_0}{\sigma} + u_\alpha \leq -u_\beta \Longleftrightarrow$$

$$\sqrt{n}\frac{a_1-a_0}{\sigma} \geq u_\alpha + u_\beta \Longleftrightarrow \sqrt{n} \geq (u_\alpha + u_\beta)\frac{\sigma}{a_1-a_0} \Longleftrightarrow$$

$$n \geq \frac{\sigma^2(u_\alpha + u_\beta)^2}{(a_1 - a_0)^2} = \frac{(u_\alpha + u_\beta)^2}{(\frac{a_1 - a_0}{\sigma})^2}. \tag{14.7}$$

Величину $\frac{a_1-a_0}{\sigma}$ называют информационным расстоянием между гипотезами. Чтобы различать близкие между собой гипотезы с малыми вероятностями ошибок первого и второго рода необходимо проводить достаточно много наблюдений.

В учебнике Б.А.Севастьянова приводится оценка для числа наблюдений n в задаче проверки гипотез о параметре биномиального распределения.

14.5 Проверка гипотез о параметре биномиального распределения

Пусть $A_1, A_2, \ldots, A_n$ — взаимно независимые случайные события, имеющие одинаковую вероятность $p = \mathbf{P}(A_i)$. Рассмотрим следующую задачу:

$$\begin{array}{ll} H_0: & p = p_0, \\ H_1: & p = p_1 \ (p_0 < p_1). \end{array}$$

Тогда при фиксированных значениях α, β ошибок первого и второго рода необходимо иметь число наблюдений n порядка

$$\frac{(u_\alpha\sqrt{p_0q_0} + u_\beta\sqrt{p_1q_1})^2}{(p_1 - p_0)^2}. \tag{14.8}$$

Так при $p_0 = 0.5$, $p_1 = 0.5 + 0.01$, $\alpha = 0.01$, $\beta = 0.05$ получим необходимое число наблюдений $n \approx 38000$.

Различать близкие между собой гипотезы не так просто. Оценки необходимого числа наблюдений в (14.7),(14.8) показывают, что числители дробей мало чувствительны к изменениям параметров, порядок роста n определяют знаменатели этих дробей.

Глава 15

Доверительные интервалы

15.1 Постановка задачи и основные определения

Пусть случайные величины $X_1, X_2, \dots, X_n$ — независимы и одинаково распределены с функцией распределения $F(x) = F(x, \theta)$, зависящей от неизвестного параметра $\theta \in \Theta$.

Важной задачей математической статистики, как это уже отмечалось ранее, является задача оценивания неизвестных параметров.

Всякую измеримую функцию от наблюдений $T(X_1, \dots, X_n)$ называют *статистикой*, а статистику,которую используют в качестве приближенного значения неизвестного параметра θ, называют *оценкой* этого параметра. Оценку обычно обозначают той же буквой, что и сам параметр, только с каким-либо значком сверху, например $\hat{\theta}, \theta^*$ и т.д.

Определение 15.1. Оценка $\hat{\theta} = \hat{\theta}(X_1, \dots, X_n)$ называется **несмещенной** оценкой параметра θ, если

$$\mathbf{E}_\theta \hat{\theta} = \theta \quad \text{для } \forall \theta \in \Theta.$$

Прикладники в этом случае говорят, что оценка лишена систематической ошибки.

Определение 15.2. Оценка $\hat{\theta}$ называется **состоятельной** для параметра θ, если для $\forall \varepsilon > 0$

$$\mathbf{P}_\theta\{|\hat{\theta} - \theta| > \varepsilon\} \to 0 \ \text{ при } \ n \to \infty$$

для всех $\theta \in \Theta$.

Другими словами, оценка называется состоятельной для θ, если она сходится по веростности к неизвестному параметру θ.

Подробно о свойствах этих (так называемых точечных) оценок и методах их получения мы будем говорить в следующей главе, а пока остановимся на другом способе оценивания — интервальных оценках (доверительных интервалах), когда по результатам наблюдений указывается не одно приближенное значение параметра, а целое множество приближенных значений.

Пусть $\mathbf{P}_\theta\{T_1(X_1,\dots,X_n) < T_2(X_1,\dots,X_n)\} = 1$ для всех $\theta \in \Theta$.

Определение 15.3. Случайный интервал $(T_1(X_1,\dots,X_n), T_2(X_1,\dots,X_n))$ называется **доверительным интервалом** для неизвестного параметра θ с надежностью $1-\alpha$, если для всех допустимых значений параметра

$$\mathbf{P}_\theta\{T_1(X_1,\dots,X_n) < \theta < T_2(X_1,\dots,X_n)\} = 1-\alpha. \qquad (15.1)$$

Число $1-\alpha$ называют также **доверительной вероятностью** интервала, обычно его выбирают близким к 1. Таким образом, каково бы ни было значение параметра, интервал (T_1, T_2) накрывает параметр с заданной вероятностью.

Для дискретных распределений не всегда удается построить доверительный интервал так, чтобы в (15.1) выполнялось равенство. Поэтому для дискретных распределений будем строить доверительные интервалы, удовлетворяющие неравенству

$$\mathbf{P}_\theta\{T_1(X_1,\dots,X_n) < \theta < T_2(X_1,\dots,X_n)\} \geq 1-\alpha, \qquad (15.2)$$

добиваясь при этом, чтобы вероятность в (15.2) была как можно ближе к $1-\alpha$.

Существует несколько стандартных приемов построения доверительных интервалов. Продемонстрируем их при построении доверительных интервалов для параметров нормального и биномиального распределений.

15.2 Доверительный интервал для математического ожидания нормального распределения при известной дисперсии

Пусть случайные величины $X_1,\dots,X_n$ независимы и имеют одинаковое нормальное распределение $N(a,\sigma_0^2)$. Математическое ожидание неизвестно и для него требуется построить доверительный интервал надежности $1-\alpha$, дисперсию считаем известной и равной σ_0^2.

Для построения доверительного интервала воспользуемся статистикой $\overline{X}$. Известно, что $\overline{X} \sim N(a, \frac{\sigma_0^2}{n})$. С помощью невырожденного линейного преобразования получим новую статистику

$$Z = \frac{\overline{X} - a}{\sigma_0/\sqrt{n}} = \frac{\overline{X} - a}{\sigma_0}\sqrt{n},$$

также имеющую нормальное распределение, но с параметрами $(0, 1)$. Это распределение не зависит от неизвестного параметра. Поэтому для заданного числа $1-\alpha$ можно указать неслучайный интервал $(c_1(\alpha), c_2(\alpha))$ такой, что

$$\mathbf{P}_a\{c_1(\alpha) < Z < c_2(\alpha)\} = 1 - \alpha$$

одновременно для всех$a \in (-\infty, \infty)$.

Такой интервал определен неоднозначно, и обычно при выборе чисел c_1, c_2 используют некоторые дополнительные соображения. Например, можно выбирать интервал наименьшей длины. Тогда

$$-c_1(\alpha) = c_2(\alpha) = u_{\alpha/2},$$

где по – прежнему $u_{\alpha/2}$ — это решение уравнения $\mathbf{P}\{|Z| \geq u_{\alpha/2}\} = \alpha$.

Поскольку

$$|Z| < u_{\alpha/2} \Longleftrightarrow \frac{|\overline{X} - a|}{\sigma_0}\sqrt{n} < u_{\alpha/2} \Longleftrightarrow |\overline{X} - a| < u_{\alpha/2}\frac{\sigma_0}{\sqrt{n}} \Longleftrightarrow$$

$$\overline{X} - u_{\alpha/2}\frac{\sigma_0}{\sqrt{n}} < a < \overline{X} + u_{\alpha/2}\frac{\sigma_0}{\sqrt{n}}, \tag{15.3}$$

то полученный интервал (15.3) является искомым доверительным интервалом.

Разобранный способ построения доверительного интервала использовал свойства статистики $Z = \frac{\overline{X}-a}{\sigma_0}\sqrt{n}$:

1) распределение статистики не зависит от неизвестных параметров;

2) статистика Z является монотонной функцией неизвестного параметра a, что позволило решить неравенство $|Z| < u_{\alpha/2}$ относительно неизвестной величины a.

Статистику, обладающую перечисленными свойствами, называют **центральной статистикой,** а описанный выше метод построения доверительного интервала, называют методом, использующим центральную статистику.

15.3 Построение доверительного интервала для дисперсии нормального распределения

Пусть случайные наблюдения $X_1, \dots, X_n$ независимы, одинаково нормально распределены с параметрами (a_0, σ^2), но теперь неизвестным параметром является дисперсия σ^2, а математическое ожидание известно и равно a_0.

Для построения доверительного интервала воспользуемся статистикой

$$S_0^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - a_0)^2,$$

которую можно использовать в качестве оценки для неизвестной дисперсии σ^2, так как

$$S_0^2 = \frac{\sigma^2}{n}\sum_{i=1}^{n}\left(\frac{X_i - a_0}{\sigma}\right)^2 = \chi_n^2, \tag{15.4}$$

то

$$\mathbf{E}S_0^2 = \frac{\sigma^2}{n}\mathbf{E}\chi_n^2 = \frac{\sigma^2}{n}\cdot n = \sigma^2,$$

$$\mathbf{D}S_0^2 = \frac{\sigma^4}{n^2}\mathbf{D}\chi_n^2 = \frac{\sigma^4}{n^2}\cdot 2n = \frac{2\sigma^4}{n} \to 0,$$

и, следовательно, эта оценка является несмещенной, состоятельной оценкой для неизвестной дисперсии. Кроме того, из нее легко получить центральную статистику

$$G = \frac{nS_0^2}{\sigma^2} = \chi_n^2.$$

Поскольку эта статистика имеет распределение хи-квадрат с n степенями свободы, не зависящее от неизвестного параметра, то по таблицам этого распределения для заданной надежности $1-\alpha$ можно определить числа $c_1 = c_1(\alpha),\ c_2 = c_2(\alpha)$, такие, что

$$\mathbf{P}_\sigma\{c_1 < G < c_2\} = 1 - \alpha.$$

Выбор чисел c_1 и c_2 неоднозначен. Можно, например, выбрать их из условий

$$\mathbf{P}\{\chi_n^2 \le c_1\} = \alpha/2, \quad \mathbf{P}\{\chi_n^2 \ge c_2\} = \alpha/2.$$

Такой интервал называют центральным.

Обозначим $g_n(\alpha)$ – решение уравнения $\mathbf{P}\{\chi_n^2 \le g(n;\alpha)\} = \alpha$. Такие числа, как известно, называют квантилями данного распределения. Тогда $c_1 = g(n;\alpha/2), \quad c_2 = g(n; 1-\alpha/2)$. Решая неравенство относительно неизвестного параметра,

$$c_1 < G < c_2 \Longleftrightarrow c_1 < \frac{nS_0^2}{\sigma^2} < c_2,$$

находим искомый доверительный интервал:

$$\frac{nS_0^2}{g(n;1-\alpha/2)} < \sigma^2 < \frac{nS_0^2}{g(n;\alpha/2)}. \tag{15.5}$$

Замечание. Если использовать как и в предыдущем примере соображения о кратчайшей длине интервала, то получим другие значения констант c_1, c_2. Покажите, что в этом случае они должны удовлетворять уравнению

$$c_2 p_n(c_1) = c_1 p_n(c_2),$$

где $p_n(x)$ — плотность распределения хи-квадрат с n степенями свободы.

15.3.1 Совместное распределение статистик $\overline{X}$ и S^2

Пусть по-прежнему $X_1, \dots, X_n$ — независимые случайные величины с одинаковым нормальным распределением $N(a, \sigma^2)$, но теперь оба параметра этого распределения неизвестны.

Обозначим

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i, \quad S^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$$

– выборочное среднее и выборочную дисперсию. Докажем следующую теорему о совместном распределении статистик $\overline{X}$ и S^2 в нормальных выборках.

Теорема 15.1. *Если $X_1, X_2, \dots, X_n$ - независимая выборка из нормального распределения с параметрами (a, σ^2), то*

1) $\overline{X}$ и S^2 независимы,

2) $\overline{X} \sim N(a, \frac{\sigma^2}{n})$,

3) $S^2 = \frac{\sigma^2}{n}\chi^2_{n-1}$.

Доказательство теоремы использует свойства многомерного нормального распределения.

Определение 15.4. Случайный вектор $\vec{Y} = (Y_1, Y_2, \dots, Y_n)^T$ имеет n - мерное стандартное нормальное распределение, если совместная плотность распределения равна

$$p_Y(y_1, \dots, y_n) = \left(\frac{1}{\sqrt{2\pi}}\right)^n e^{-\frac{1}{2}\sum_{i=1}^{n} y_i^2}.$$

Нетрудно проверить, что компоненты Y_i случайного вектора, имеющего стандартное нормальное распределение, независимы и $Y_i \sim N(0,1)$.

Лемма 15.1. Пусть случайный вектор $\vec{Z}$ получен из вектора $\vec{Y}$ со стандартным нормальным распределением с помощью ортогонального преобразования $\vec{Z} = \mathbf{C}\vec{Y}$. Тогда $\vec{Z}$ также имеет стандартное нормальное распределение.

Доказательство. Напомним, что строки матрицы ортогонального преобразования образуют ортонормированный базис в $\mathbb{R}^n$,

$$\mathbf{C}^{-1} = \mathbf{C}^T, \quad \det \mathbf{C} = 1$$

и что ортогональное преобразование сумму квадратов переменных переводит в сумму квадратов новых переменных.

Пусть A - событие из борелевской σ- алгебры в $\mathbb{R}^n$, $B = \mathbf{C}^{-1}(A)-$ прообраз события A при преобразовании $\mathbf{C}$. Тогда

$$\mathbf{P}\{\vec{Z} \in A\} = \mathbf{P}\{\mathbf{C}\vec{Y} \in A\} = \int\limits_B p_Y(y_1, \ldots, y_n) dy_1 \cdots dy_n =$$

$$\int\limits_B \left(\frac{1}{2\pi}\right)^{n/2} \exp\left(-\frac{1}{2}\sum_{i=1}^n y_i^2\right) dy_1 \cdots dy_n..$$

В последнем интеграле сделаем замену переменных $\vec{z} = \mathbf{C}\vec{y}$, тогда

$$\mathbf{P}\{\vec{Z} \in A\} = \int\limits_A \left(\frac{1}{2\pi}\right)^{n/2} \exp\left(-\frac{1}{2}\sum_{i=1}^n z_i^2\right) dz_1 \cdots dz_n.$$

Подинтегральная функция – по определению является плотностью распределения случайного вектора $\vec{Z}$, следовательно, распределение $\vec{Z}$ является стандартным нормальным. □

Возвратимся к доказательству теоремы.

Доказательство. Утверждение 2) теоремы почти очевидно, поэтому докажем только утверждения 1), 3). Наряду с исходными случайными величинами X_i рассмотрим

$$Y_i = \frac{X_i - a}{\sigma} \; (i = 1, \ldots, n).$$

Отметим, что случайный вектор $\vec{Y} = (Y_1 \ldots, Y_n)^T$ имеет стандартное нормальное распределение. Рассмотрим ортогональную матрицу

$$\mathbf{C} = \begin{pmatrix} \frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} & \ldots & \frac{1}{\sqrt{n}} \\ c_{21} & c_{22} & \ldots & c_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ c_{n1} & c_{n2} & \ldots & c_{nn} \end{pmatrix},$$

у которой первая строка выбрана специальным образом, а остальные произвольным образом дополняют первую строку до ортогонального базиса в $\mathbb{R}^n$.

В силу леммы $\vec{Z} = \mathbf{C}\vec{Y}$ также имеет стандартное нормальное распределение, поэтому компоненты вектора $Z_i \sim N(0,1)$ и независимы. При этом

$$\overline{X} = a + \sigma\overline{Y} = a + \frac{\sigma}{\sqrt{n}} Z_1,$$

$$S^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2 = \frac{\sigma^2}{n}\sum_{i=1}^{n}(Y_i - \overline{Y})^2 = \sigma^2\left(\frac{1}{n}\sum_{i=1}^{n} Y_i^2 - \overline{Y}^2\right) =$$

$$= \sigma^2\left(\frac{1}{n}\sum_{i=1}^{n} Z_i^2 - (\frac{1}{\sqrt{n}}Z_1)^2\right) = \frac{\sigma^2}{n}\sum_{i=2}^{n} Z_i^2.$$

Мы видим, что выборочное среднее $\overline{X}$ зависит лишь от Z_1, а выборочная дисперсия S^2 от не зависящих от Z_1 случайных величин $Z_2, \ldots, Z_n$, следовательно, $\overline{X}$ и S^2 — независимые случайные величины.

Кроме того,

$$S^2 = \frac{\sigma^2}{n} \cdot \chi^2_{n-1},$$

так как по определению случайная величина, представленная в виде суммы квадратов независимых случайных величин со стандартным нормальным распределением, имеет распределение хи – квадрат с числом степеней свободы, равным числу независимых слагаемых.

Тем самым теорема доказана. □

Воспользуемся доказанными свойствами статистик $\overline{X}$ и S^2 при построении доверительных интервалов для неизвестных математического ожидания a и дисперсии σ^2.

Доверительный интервал для σ^2 строится с помощью центральной статистики

$$G = \frac{nS^2}{\sigma^2} = \chi^2_{n-1},$$

которая имеет распределение хи – квадрат с $n-1$ степенью свободы. Построение полностью повторяет рассуждения примера 2, поэтому сразу выпишем интервал с надежностью $1-\alpha$:

$$\frac{nS^2}{g(n-1;1-\alpha/2)} < \sigma^2 < \frac{nS^2}{g(n-1;\alpha/2)}, \tag{15.6}$$

где $g_{n-1}(\alpha)$ обозначает α- квантиль распределения хи-квадрат.

Замечание. Проводя простые вычисления, можно получить, что

$$\mathbf{E}S^2 = \frac{\sigma^2}{n} \cdot \mathbf{E}\chi^2_{n-1} = \frac{n-1}{n} \cdot \sigma^2,$$

то есть выборочная дисперсия не является несмещенной оценкой для неизвестной дисперсии. Используя эту оценку при небольшом количестве проведенных наблюдений n, мы будем получать в среднем числа, несколько меньшие действительного значения σ^2.

Но это легко исправить, если рассмотреть другую оценку

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2,$$

которую часто так и называют исправленной выборочной дисперсией. Впрочем, асимптотические свойства этих двух оценок совпадают.

Задача 15.1. Докажите, что S^2 и s^2 являются состоятельными оценками для неизвестной дисперсии для любой независимой выборки из распределения с конечной дисперсией.

Доверительный интервал (15.6) можно переписать в виде

$$\frac{(n-1)s^2}{g(n-1;1-\alpha/2)} < \sigma^2 < \frac{(n-1)s^2}{g(n-1;\alpha/2)}.$$

15.4 Доверительный интервал для неизвестного математического ожидания нормальной выборки при неизвестной дисперсии.

15.4.1 Распределение Стьюдента.

При построении доверительного интервала для a мы не можем как ранее, в примере 1, использовать статистику $Z = \frac{(\overline{X}-a)\sqrt{n}}{\sigma}$, поскольку интервал будет зависеть от второго неизвестного параметра σ. Американский математик У.С.Госсет, публиковавший свои труды под псевдонимом Student, предложил центральную статистику

$$t_{n-1} = \frac{(\overline{X} - a)\sqrt{n}}{s}.$$

Покажем, что ее распределение не зависит от неизвестных параметров. Действительно,

$$t_{n-1} = \frac{\frac{(\overline{X}-a)\sqrt{n}}{\sigma}}{\frac{s}{\sigma}} = \frac{Z}{\sqrt{\frac{1}{n-1}\chi^2_{n-1}}}$$

представляется в виде дроби, числитель и знаменатель которой — независимые случайные величины, причем числитель имеет стандартное нормальное распределение, а знаменатель представим в виде $\sqrt{\frac{1}{n-1}\chi^2_{n-1}}$. Статистику t_{n-1} называют статистикой **Стьюдента,** а ее распределение – **распределением Стьюдента** с $n-1$ степенью свободы.

Замечание. Приведем формулу плотности для распределения Стьюдента с m степенями свободы:

$$p(x) = \frac{\Gamma\left((m+1)/2\right)}{\sqrt{\pi m}\Gamma(m/2)}(1 + x^2/m)^{-(m+1)/2}.$$

В частности, если $m = 1$, то это распределение известно как распределение Коши.

Математическое ожидание существует при $m > 1$ и равно 0, дисперсия существует при $m > 2$ и равна $m/(m-2)$.

По таблицам распределения Стьюдента для заданной надежности $1-\alpha$ определим $t(n-1;\alpha/2)$, являющееся решением уравнения

$$\mathbf{P}\{|t_{n-1}| > t(n-1;\alpha/2)\} = \alpha, \text{ или } \mathbf{P}\{t_{n-1} > t(n-1;\alpha/2)\} = \alpha/2.$$

Поскольку

$$|t_{n-1}| > t(n-1;\alpha/2) \Longleftrightarrow \left|\frac{(\overline{X} - a)\sqrt{n}}{s}\right| > t(n-1;\alpha/2) \Longleftrightarrow$$

$$\overline{X} - \frac{t(n-1;\alpha/2)s}{\sqrt{n}} < a < \overline{X} + \frac{t(n-1;\alpha/2)s}{\sqrt{n}}, \qquad (15.7)$$

то (15.7) представляет собой доверительный интервал для a с надежностью $1-\alpha$.

15.5 Асимптотический доверительный интервал для параметра p биномиального распределения

На следующем примере проиллюстрируем способ построения асимптотического доверительного интервала, который можно использовать только для достаточно больших выборок.

Пусть проведено n независимых испытаний Бернулли с вероятностью успеха p в отдельном испытании,

$$X_i = \begin{cases} 1, & \text{если в испытании с номером } i \text{ произошел успех} \\ 0, & \text{если в испытании с номером } i \text{ произошел неуспех,} \end{cases}$$

$(i = 1, 2, \ldots, n)$. Построим асимптотический доверительный интервал для p.

Пусть $\mu_n = \sum\limits_{i=1}^{n} X_i$ - число успехов в n испытаниях. Известно, что относительная частота наступления успеха $\frac{\mu_n}{n}$ асимптотически нормальна с параметрами $(p, \frac{p(1-p)}{n})$. Это означает, что для $\forall x$

$$\mathbf{P}\left\{ \frac{\frac{\mu_n}{n} - p}{\sqrt{\frac{p(1-p)}{n}}} < x \right\} \to \Phi(x).$$

Далее существует несколько способов построения доверительного интервала.

Мы сначала рассмотрим способ, основанный на преобразовании, стабилизирующем дисперсию, то есть ищется преобразование относительной частоты, в результате которого также получается асимптотически нормальная слдучайная величина, но с дисперсией, не зависящей от параметра p.

Для этого воспользуемся теоремой 12.5:

если последовательность случайных величин Y_n асимптотически нормальна с параметрами (a, b_n^2), причем $b_n \to 0$, функция $g(x)$ дважды непрерывно дифференцируема, то последовательность $W_n = g(Y_n)$ также асимптотически нормальна, но с параметрами $(g(a), b_n^2(g'(a))^2)$.

Рассмотрим функцию $g(x) = 2\arcsin\sqrt{x}$, $g'(x) = \frac{1}{\sqrt{x(1-x)}}$. Применяя теорему, получим, что последовательность

$$\eta_n = 2\arcsin\frac{\mu_n}{n}$$

асимптотически нормальна с параметрами $(2\arcsin\sqrt{p}; \frac{1}{n})$. Тогда при достаточно больших n (хотя бы несколько десятков) выполняется приближенное равенство

$$\mathbf{P}\left\{\left|2\arcsin\sqrt{\frac{\mu_n}{n}} - 2\arcsin\sqrt{p}\right| < \frac{u_{\alpha/2}}{\sqrt{n}}\right\} = 1-\alpha.$$

Решив неравенство в фигурных скобках относительно параметра p, получим доверительный интервал

$$\sin^2\left(\arcsin\sqrt{\frac{\mu_n}{n}} - \frac{u_{\alpha/2}}{2\sqrt{n}}\right) < p < \sin^2\left(\arcsin\sqrt{\frac{\mu_n}{n}} + \frac{u_{\alpha/2}}{2\sqrt{n}}\right) \qquad (15.8)$$

Задача 15.2. Докажите, что последовательность случайных величин

$$V_n = \frac{(\mu_n/n - p)\sqrt{n}}{\sqrt{\mu_n/n(1-\mu_n/n)}}$$

асимптотически нормальна с параметрами (0,1).

Используя этот факт, можно построить еще один асимптотический доверительный интервал с надежностью $1-\alpha$

$$\overline{X} - u_{\alpha/2}\frac{\sqrt{\frac{\mu_n}{n}(1-\frac{\mu_n}{n})}}{\sqrt{n}} < p < \overline{X} + u_{\alpha/2}\frac{\sqrt{\frac{\mu_n}{n}(1-\frac{\mu_n}{n})}}{\sqrt{n}}. \qquad (15.9)$$

Доверительные интервалы (15.8), (15.9) при достаточно больших n не слишкои сильно отличаются один от другого. Так при $n = 65$, $\mu_n = 30$, $1-\alpha = 0.95$ и тот, и другой дают при подстановке данных результат

$$0.340 < p < 0.583,$$

тогда как точный доверительный интервал, вычисленный по таблицам Л.Н.Большева, Н.В.Смирнова,

$$0.337 < p < 0.590.$$

Глава 16

Точечные оценки для неизвестных параметров

16.1 Сравнение свойств несмещенных оценок

Пример 16.1. По регистрационным номерам танков во время второй мировой войны оценивался объем производства военной техники. Задача сводится к оценке параметра равномерного распределения.

Пусть $X_1, X_2, \ldots, X_n$ - независимые случайные величины с плотностью распределения

$$p(x;\theta) = \begin{cases} 1/\theta, & 0 < x < \theta, \\ 0, & \text{иначе.} \end{cases}$$

Введем новые случайные величины $X'_j = \frac{X_j}{\theta}$, которые распределены равномерно на отрезке $[0,1]$ и имеют плотность распределения

$$p_1(x) = \begin{cases} 1, & 0 < x < 1, \\ 0, & \text{иначе.} \end{cases}$$

Для этих случайных величин

$$\mathbf{E}X'_j = \int\limits_0^1 x dx = \frac{1}{2}; \quad \mathbf{E}(X'_j)^2 = \int\limits_0^1 x^2 dx = \frac{1}{3}; \quad \mathbf{D}X'_j = \frac{1}{12}.$$

Пусть

$$Y = \max_{1 \le j \le n} X'_j.$$

Для этой случайной величины при $x \in [0,1]$ функция распределения и плотность соответсвенно равны

$$F_Y(x) = \mathbf{P}\{Y < x\} = \mathbf{P}^n(X'_j < x) = x^n, \quad p_Y(x) = nx^{n-1}.$$

Следовательно,

$$\mathbf{E}Y = \int\limits_0^1 nx^n dx = \frac{n}{n+1}$$

$$\mathbf{D}Y = \mathbf{E}Y^2 - (\mathbf{E}Y)^2 = \frac{n}{(n+2)(n+1)^2}.$$

В качестве оценок для параметра θ равномерного распределения рассмотрим две статистики

$$\theta_1^* = 2\overline{X}, \qquad \theta_2^* = \frac{n+1}{n} \max_{1 \le j \le n} X_j$$

и сравним их свойства.

Обе оценки являются несмещенными, так как для всех $\theta > 0$

$$\mathbf{E}_\theta \theta_1^* = 2\theta \mathbf{E} \frac{X'_1 + \cdots + X'_n}{n} = \theta,$$

$$\mathbf{E}_\theta \theta_2^* = \frac{n+1}{n} \mathbf{E}_\theta Y = \theta.$$

Обе оценки являются состоятельными, поскольку их дисперсии стремятся к 0 при $n \to \infty$. Действительно,

$$\mathbf{D}_\theta \theta_1^* = \frac{\theta^2}{n^2} \cdot \frac{n}{3} = \frac{\theta^2}{3n},$$

$$\mathbf{D}_\theta \theta_2^* = \theta^2 \mathbf{D}_\theta (\frac{n+1}{n} Y) = \frac{\theta^2}{n(n+2)}.$$

Однако при всех $n \ge 2$ и при всех θ одновременно дисперсия второй оценки меньше дисперсии первой оценки, при $n = 1$ эти оценки просто совпадают. В таких ситуациях будем говорить, что θ_2^* имеет дисперсию *равномерно меньшую,* чем θ_1^*. А это означает, что вторая оценка предпочтительнее первой (особенно при больших n), так как меньше отклоняется от неизвестного параметра.

16.2 Семейства распределений

Рассмотрим два основных типа параметрических семейств распределений.

I тип. Все распределения семейства получаются из некоторого фиксированного распределения путем преобразований, зависящих от неизвестного параметра θ:

а) θ — параметр сдвига, в этом случае

$$p(x;\theta) = p(x-\theta), \quad \theta \in \Theta = \mathbb{R};$$

б) θ — параметр масштаба, в этом случае

$$p(x;\theta) = \frac{1}{\theta} p\left(\frac{x}{\theta}\right), \quad \theta > 0;$$

в) присутствуют и параметр сдвига, и параметр масштаба, $\theta = (\theta_1, \theta_2)$:

$$p(x;\theta_1,\theta_2) = \frac{1}{\theta_2} p\left(\frac{x-\theta_1}{\theta_2}\right), \quad \theta_1 \in \mathbb{R}, \quad \theta_2 > 0.$$

Так получено семейство распределений примера 1, так получается семейство нормальных распределений на прямой.

II тип. Экспоненциальные семейства распределений. Плотность распределения имеет вид:

$$p(x;\theta_1,\ldots,\theta_k) = h(x)\cdot c(\theta_1,\ldots,\theta_k)\cdot e^{\sum\limits_{i=1}^{k}\theta_i T_i(x)},$$

где $\theta = (\theta_1,\ldots,\theta_k)$ — $k-$ мерный параметр.

Примерами экспоненциальных семейств распределений являются распределение Пуассона, биномиальное распределение, нормальное распределение.

Пусть $X_1,\ldots,X_n$ — независимые одинаково распределенные случайные величины с плотностью распределения вероятностей $p(x;\theta)$. Совместная плотность распределения равна

$$\hat{p}(x_1,\ldots,x_n;\theta) = p(x_1;\theta)\cdots p(x_n;\theta).$$

Для дискретных случайных величин $\hat{o}(x_1,\ldots,x_n)$ будет обозначать совместную вероятность

$$\hat{p}(x_1,\ldots,x_n;\theta) = \mathbf{P}_\theta\{X_1 = x_1,\ldots,X_n = x_n\}.$$

16.3 Метод максимального правдоподобия.

Наиболее обоснованным и распространенным методом отыскания оценок является метод максимального правдоподобия.

Функция $\hat{p}(x_1, \ldots, x_n; \theta)$, которую рассматривают как функцию параметра θ, считая при этом остальные переменные фиксированными, называется функцией правдоподобия.

Метод максимального правдоподобия состоит в следующем. Если в результате наблюдений $X_1 = x_1, \ldots, X_n = x_n$, то в качестве оценки неизвестного параметра выбирается θ^*, при котором

$$\hat{p}(x_1, \ldots, x_n; \theta^*) = \max_\theta \hat{p}(x_1, \ldots, x_n; \theta),$$

т.е. ищем такое значение θ, при котором наблюдаемый результат наиболее вероятен.

В учебнике Б.А.Севастьянова есть теорема, в которой доказывается асимптотическая эффективность оценок максимального правдоподобия.

16.4 Неравенство Рао - Крамера

Будем рассматривать только случай одномерного параметра θ. Обозначим $\vec{X} = (X_1, \ldots, X_n)^T$ — n- мерный случайный вектор, а полученное в результате наблюдений значение этого вектора — $\vec{x} = (x_1, \ldots, x_n)^T$.

В асимптотической теории, когда $n \to \infty$, используют такие понятия и свойства оценок, как состоятельность, асимптотическая нормальность, асимптотическая эффективность.

Для небольших значений n действуют понятия несмещенности и эффективности.

Если оценивается функция неизвестного параметра $\tau(\theta)$, то оценка $\tau^*(\vec{X})$ называется **несмещенной,** если

$$\mathbf{E}_\theta \tau^*(\vec{X}) = \tau(\theta)$$

при всех допустимых значениях параметра θ.

Для определенных семейств распределений выполняется неравенство Рао — Крамера

$$\mathbf{D}_\theta \theta^* \geq \frac{1}{n\mathbf{E}_\theta \left(\dfrac{\partial \ln p(\vec{x}; \theta)}{\partial \theta} \right)^2}, \tag{16.1}$$

согласно которому дисперсия несмещенной оценки параметра θ не может быть сколь угодно малой. Неравенство (16.1) указывает нижнюю границу для дисперсии несмещенных оценок. Это неравенство часто называют также неравенством информации.

В примере 16.1 нашлась оценка, для которой дисперсия имела порядок $\frac{1}{n^2}$, В рассматриваемом случае это невозможно.

Определение 16.1. Если дисперсия несмещенно оценки параметра θ равна при всех θ нижней границе в неравенстве Рао – Крамера (16.1), то оценка называется **эффективной.**

Поясним понятие эффективности. Мы сравниваем точность несмещенных оценок по их дисперсиям. Пусть θ_1^* и θ_2^* — две несмещенные оценки. Поскольку их дисперсии являются функциями, зависящими от параметра, то может оказаться, что для одних значений параметра предпочтительнее одна оценка, а для других – другая. Такие две оценки будем считать несравнимыми.

Эффективная оценка, если существует, имеет дисперсию, которая одновременно для всех значений параметра меньше (во всяком случае не больше) дисперсии любой другой оценки.

Формальное доказательство неравенства информации достаточно просто, оно использует неравенство Коши — Буняковского, которое в дискретном случае сводится к тому, что скалярное произведение двух векторов по абсолютной величине не превосходит произведения длин этих векторов:

$$\left(\sum_{j=1}^{m} a_j b_j\right)^2 \leq \left(\sum_{j=1}^{m} a_j^2\right) \cdot \left(\sum_{j=1}^{n} b_j^2\right),$$

причем неравенство обращается в равенство тогда и только тогда, когда существуют $\lambda,\ \mu\ (\lambda^2 + \mu^2 > 0)$ такие, что

$$\lambda a_j + \mu b_j = 0,\ j = 1, 2, \ldots, m.$$

В непрерывном случае неравенство Коши — Буняковского принимает вид

$$\left(\int f(x)g(x)dx\right)^2 \leq \left(\int f^2(x)dx\right)\left(\int g^2(x)dx\right).$$

Для выполнения равенства необходимо и достаточно, чтобы существовали $\lambda,\ \mu\ (\lambda^2 + \mu^2 > 0)$ такие, что

$$\lambda f(x) + \mu g(x) = 0 \quad \text{почти всюду по } x.$$

При доказательстве неравенства Рао - Крамера придется дифференцировать под знаком интеграла или дифференцировать ряды, поэтому необходимы условия, обеспечивающие возможность такого действия. В различных учебниках приводятся различные достаточные условия. Но одно условие крайне существенно:

$$\{x:\ \hat{p}(x;\theta)>0\}\ \text{ не зависит от } \theta.$$

В примере 16.1, когда нашлась "сверхэффективная"оценка, это условие не выполнялось.

Доказательство. неравенства информации. Поскольку тождественно по всем θ выполняются равенства

$$\int\limits_{\mathbb{R}^n}\hat{p}(\vec{x};\theta)d\vec{x}=1,$$

$$\int\limits_{\mathbb{R}^n}\theta^*(\vec{x})\hat{p}(\vec{x};\theta)d\vec{x}=\theta,$$

то продифференцировав их по θ, получим

$$\int\limits_{\mathbb{R}^n}\frac{\partial\ln\hat{p}(\vec{x};\theta)}{\partial\theta}\hat{p}(\vec{x};\theta)d\vec{x}=0, \tag{16.2}$$

$$\int\limits_{\mathbb{R}^n}\theta^*(\vec{x})\frac{\partial\ln\hat{p}(\vec{x};\theta)}{\partial\theta}\hat{p}(\vec{x};\theta)d\vec{x}=1. \tag{16.3}$$

При дифференцировании мы воспользовались тем, что $\frac{df}{dx}=\frac{d\ln f(x)}{dx}f(x)$.

Умножим (16.2) на θ и вычтем из (16.3), получим

$$\int\limits_{\mathbb{R}^n}(\theta^*(\vec{x})-\theta)\frac{\partial\ln\hat{p}(\vec{x};\theta)}{\partial\theta}\hat{p}(\vec{x};\theta)d\vec{x}=1. \tag{16.4}$$

Обозначим

$$f(\vec{x})=(\theta^*(\vec{x})-\theta)\sqrt{\hat{p}(\vec{x};\theta)},$$
$$g(\vec{x})=\frac{\partial\ln\hat{p}(\vec{x};\theta)}{\partial\theta}\sqrt{\hat{p}(\vec{x};\theta)}.$$

Тогда, применяя к (16.4) неравенство Коши - Буняковского, имеем

$$1=\left(\int\limits_{\mathbb{R}^n}f(\vec{x})g(\vec{x})d\vec{x}\right)^2\le\int\limits_{\mathbb{R}^n}f^2(\vec{x})d\vec{x}\cdot\int\limits_{\mathbb{R}^n}g^2(\vec{x})d\vec{x}=$$

$$= \int\limits_{\mathbb{R}^n} (\theta^* - \theta)^2 \hat{p}(\vec{x};\theta) d\vec{x} \cdot \int\limits_{\mathbb{R}^n} \left(\frac{\partial \hat{p}(\vec{x};\theta)}{\partial \theta} \right)^2 \hat{p}(\vec{x};\theta) d\vec{x}.$$

Следовательно,

$$1 \leq \mathbf{D}_\theta \theta^*(\vec{X}) \cdot \mathbf{E}_\theta \left(\frac{\partial \ln \hat{p}(\vec{X};\theta)}{\partial \theta} \right)^2 .$$

Откуда получаем

$$\mathbf{D}_\theta \theta^* \geq \frac{1}{\mathbf{E}_\theta \left(\frac{\partial \hat{p}(\vec{X};\theta)}{\partial \theta} \right)^2}. \tag{16.5}$$

Заметим, что пока наши рассуждения нигде не использовали независимость и одинаковую распределенность наблюдений, и поэтому неравенство (16.5) справедливо для произвольного случайного вектора $\vec{X}$, совместное распределение компонент которого задается плотностью $\hat{p}(\vec{x};\theta)$.

Пусть теперь $\hat{p}(\vec{x};\theta) = p(x_1;\theta)\cdots p(x_n;\theta)$. Тогда

$$\frac{\partial \ln \hat{p}(\vec{X};\theta)}{\partial \theta} = \sum_{i=1}^{n} W_i,$$

где $W_i = \frac{\partial \ln p(X_i;\theta)}{\theta}$ - независимые одинаково распределенные случайные величины, для которых из (16.2) следует, что $\mathbf{E}_\theta W_i = 0$. Но тогда

$$\mathbf{E}_\theta \left(\frac{\partial \ln \hat{p}(\vec{X};\theta)}{\partial \theta} \right)^2 = \mathbf{D}_\theta \sum_{i=1}^{n} W_i = n \mathbf{E}_\theta W_1^2 \tag{16.6}$$

и мы приходим к неравенству Рао - Крамера (16.1). □

Величину $\mathbf{E}_\theta W_1^2 = \mathbf{E}_\theta \left(\frac{\partial p(X_1;\theta)}{\partial \theta} \right)^2 = I(\theta)$ называют *количеством информации* о параметре θ, содержащейся в одном наблюдении. Неравенство (16.1) можно переписать в другом виде

$$\mathbf{D}_\theta \theta^* \geq \frac{1}{nI(\theta)},$$

известном как неравенство информации.

Необходимым и достаточным условием достижения равенства в неравенстве информации является существование λ и μ, не зависящих от $\vec{x}$, но, вообще говоря, зависящих от θ, таких, что с вероятностью 1 для всех θ выполняется соотношение

$$\lambda(\theta^* - \theta) = \mu \frac{\partial \ln \hat{p}(\vec{x}; \theta)}{\partial \theta}.$$

Обычно его переписывают несколько в другом виде

$$\frac{\partial \ln \hat{p}(\vec{x}; \theta)}{\partial \theta} = k(\theta)(\theta^* - \theta), \tag{16.7}$$

где $k(\theta)$ не зависит от $\vec{X}$.

Пусть существует эффективная оценка параметра θ^*. Вспомним метод максимального правдоподобия. Если выполнены условия гладкости и максимум достигается во внутренней точке, то оценка максимального правдоподобия может быть найдена из уравнения

$$\frac{\partial \ln \hat{p}(\vec{X}; \theta)}{\theta} = 0,$$

и, следовательно, будет совпадать с эффективной оценкой θ^* параметра θ.

Из условия существования эффективной оценки (16.7) можно получить выражение плотности (совместной вероятности) параметрического семейства:

$$\ln \hat{p}(\vec{x}; \theta) = \int k(\theta)(\theta^* - \theta) d\theta + c(\vec{x})$$

или

$$\hat{p}(\vec{x}; \theta) = h(\vec{x}) \cdot e^{\theta^* \int k(\theta) d\theta} \cdot u(\theta).$$

Отметим, что первый множитель $h(\vec{x})$ зависит только от наблюдений и не зависит от неизвестного параметра.

Таким образом, если есть эффективная оценка, то семейство распределений должно быть экспоненциальным.

Пример 16.2. Пусть $X_1, \ldots, X_n$ — независимые случайные величины, распределенные по закону Пуассона с параметром $\theta > 0$. Тогда их совместная вероятность определяется по формуле

$$\mathbf{P}_\theta\{X_1 = m_1, \ldots, X_n = m_n\} = \left(\prod_{i=1}^{n} m_i!\right)^{-1} \theta^{\sum\limits_{i=1}^{n} m_i} e^{-n\theta},$$

следовательно, функция правдоподобия имеет вид

$$\hat{p}(\vec{X}; \theta) = \left(\prod_{i=1}^{n} X_i!\right)^{-1} \theta^{\sum\limits_{i=1}^{n} X_i} e^{-n\theta}.$$

Но тогда

$$\frac{\partial \ln \hat{p}(\vec{X};\theta)}{\partial \theta} = \frac{1}{\theta}\sum_{i=1}^{n} X_i - n = \frac{n}{\theta}(\overline{X} - \theta).$$

Откуда находим эффективную оценку неизвестного параметра $\theta^* = \overline{X}$.

Приложение 1. Основные распределения и их свойства

- **Распределение Бернулли** $B(1;p)$ с параметром $p \in (0,1)$.
  Случайная величина $X \sim B(1;p)$, если

  $$\mathbf{P}\{X = 1\} = p, \quad \mathbf{P}\{X = 0\} = 1 - p.$$

  $$\mathbf{E}\, x = p, \quad \mathbf{D}\, x = p(1-p),$$

  характеристическая функция $f(t) = 1 - p + pe^{it}$.

  Примером случайной величины с таким распределением может служить индикатор случайного события A : $I_A(\omega) = 1$, если $\omega \in A$, и
  $I_A(\omega) = 0$, если $\omega \in \overline{A}$, $p = \mathbf{P}(A)$.

- **Биномиальное распределение** $B(n;p)$ с параметрами $n \in \mathbb{N}$, $p \in (0;1)$.
  Случайная величина $X \sim B(n;p)$, если

  $$\mathbf{P}\{X = k\} = C_n^k p^k (1-p)^{n-k}, \quad k = 0, 1, \ldots, n.$$

  $$\mathbf{E}\, X = np, \quad \mathbf{D}\, X = np(1-p), \quad f(t) = (1 - p + pe^{it})^n$$

  Примером случайной величины с таким распределением может служить число успехов в n независимых испытаниях Бернулли с вероятностью успеха p в каждом из них.

- **Распределение Пуассона**$\Pi(\lambda)$ с параметром $\lambda > 0$.
  Случайная величина $X \sim \Pi(\lambda)$, если

  $$\mathbf{P}\{X = k\} = \frac{\lambda^k e^{-\lambda}}{k!}, \quad k = 0, 1, 2, \ldots$$

  $$\mathbf{E}\, X = \lambda, \quad \mathbf{D}\, X = \lambda, \quad f(t) = \exp(\lambda(e^{it} - 1)).$$

Одно из основных свойств этого распределения: сумма независимых случайных величин с пуассоновскими распределениями также имеет распределение Пуассона с параметром, равным сумме параметров слагаемых.

- **Геометрическое распределение** $G(p)$ с параметром $p \in (0,1)$. Случайная величина $X \sim G(p)$, если

$$\mathbf{P}\{X = k\} = p(1-p)^k, \quad k = 0, 1, 2, \ldots$$

$$\mathbf{E}\, X = \frac{1-p}{p}; \quad \mathbf{D}\, X = \frac{1-p}{p^2}; \quad f(t) = p(1-(1-p)e^{it})^{-1}.$$

Такое распределение имеет случайная величина, равная числу неудач, предшествующих первому успеху в испытаниях Бернулли.

- **Отрицательное биномиальное распределение** $Bi^-(m;p)$ с параметрами $m \in \mathbb{N}$, $p \in (0,1)$. Случайная величина $X \sim Bi^-(m;p)$, если

$$\mathbf{P}\{X = k\} = C_{k+m-1}^{k} p^m (1-p)^k, \quad k = 0, 1, 2, \ldots$$

$$\mathbf{E}\, X = \frac{m(1-p)}{p}; \quad \mathbf{D}\, X = \frac{m(1-p)}{p^2}, \quad f(t) = p^m(1-(1-(1-p)e^{it})^{-m}.$$

Примером случайной величины с отрицательным биномиальным распределением может служить число неудач, предшествующих наступлению m- ого успеха в испытаниях Бернулли. Заметим, что $G(p) = Bi^-(1;p)$.

- **Гипергеометрическое распределение** $HG(N, M, n)$. Случайная величина $X \sim HG(N, M, n)$, если

$$\mathbf{P}\{X = k\} = \frac{C_m^k C_{N-M}^{n-k}}{C_N^n},$$

где $\max(0; n - N + M) \leqslant k \leqslant \min(n; M)$, $k \in \mathbb{Z}$,

$$\mathbf{E}\, X = \frac{nM}{N}, \quad \mathbf{D}\, X = \frac{nM}{N}\left(1 - \frac{M}{N}\right)\frac{N-n}{N-1}.$$

Примером случайной величины с таким распределением является случайная величина, равная числу красных шаров среди n отобранных, если производится случайный выбор без возвращения из совокупности, содержащей M красных и $N - M$ белых шаров.

- **Полиномиальное распределение** $B(n;p_1,\ldots,p_m)$ с параметрами $n\in\mathbb{N}$, $p_i>0$, $\sum_i^m p_i=1$.
Случайный вектор $(X_1,\ldots,X_m)\sim B(n;p_1,\ldots,p_m)$, если

$$\mathbf{P}\{X_1=k_1,\ldots,X_m=k_m\}=\frac{n!}{k_1!\ldots k_m!}p_1^{k_1}\ldots p_m^{k_m},$$

где $k_1+\ldots+k_m=n$, k_m- целые неотрицательные числа. Среди свойств этого распределения отметим, что

$$X_i\sim B(n;p_i),\quad cov(X_i,X_j)=-np_ip_j,\ \text{если } i\neq j.$$

Полиномиальное распределение применяется в независимых повторных испытаниях с m различными исходами для совместного распределения случайных величин X_i, равных числу наступлений исходов с соответствующими номерами в n испытаниях (p_k – вероятность k - ого исхода в отдельном испытании.) .

- **Нормальное распределение** $N(a,\sigma^2)$ с параметрами $-\infty<a<+\infty,\quad \sigma>0$.
Случайная величина $X\sim N(a,\sigma^2)$, если ее плотность распределения имеет вид

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

$$\mathbf{E}\,X=a,\quad \mathbf{D}\,X=\sigma^2,\quad f(t)=e^{iat-\frac{\sigma^2t^2}{2}}.$$

Для плотности и функции распределения стандартного нормального закона $N(0,1)$ приняты специальные обозначения:

$$\varphi(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}},\quad \Phi(x)=\int\limits_{-\infty}^{x}\frac{1}{\sqrt{2\pi}}e^{-\frac{u^2}{2}}du$$

соответственно. Для семейства нормальных распределений параметры a и σ являются параметрами сдвига и масштаба, поскольку

$$p(x;a,\sigma)=\frac{1}{\sigma}\varphi(\frac{x-a}{\sigma}),\quad F(x;a,\sigma)=\Phi(\frac{x-a}{\sigma}),\quad \frac{X-a}{\sigma}\sim N(0,1),$$

где $F(x;a,\sigma)$ обозначает функцию распределения случайной величины X.

Любая нетривиальная линейная комбинация независимых нормально распределенных случайных величин снова имеет нормальное распределение.

- **Равномерное распределение**$U(a,b)$ с параметрами $-\infty < a < b < +\infty$.
  Случайная величина $X \sim U(a,b)$, если плотность распределения определяется формулой
  $$p(x;a,b) = \begin{cases} \frac{1}{b-a} & \text{при} \quad x \in [a;b], \\ 0 & \text{при} \quad x \notin [a;b]. \end{cases}$$
  $$\mathbf{E}\, X = \frac{a+b}{2}, \quad \mathbf{D}\, X = \frac{(b-a)^2}{12}, f(t) = \frac{e^{itb} - e^{ita}}{it(b-a)}.$$
  Всякое невырожденное линейное преобразование $Y = cX + d$ сноваимеет равномерное распределение, в том числе $X^* = \frac{X-a}{b-a} \sim U(0,1)$.

- **Показательное (экспоненциальное) распределение**$E(\alpha,\beta)$ с параметрами $\alpha > 0$, $\infty < -\beta < +\infty$.
  Случайная величина $X \sim E(\alpha,\beta)$, если ее плотность распределения равна
  $$p(x;\alpha,\beta) = \begin{cases} \alpha e^{-\alpha(x-\beta)} & \text{при} \quad x > \beta, \\ 0 & \text{при} \quad x \leqslant \beta. \end{cases}$$
  $$\mathbf{E}\, X = \beta + \frac{1}{\alpha}, \quad \mathbf{D}\, X = \frac{1}{\alpha^2}, \quad f(t) = \frac{\alpha e^{it\beta}}{\alpha - it}.$$
  Параметры β и α^{-1} являются соответственно параметрами сдвига и масштаба, $Y = \alpha(X-\beta) \sim E(1,0)$.

- **Гамма – распределение** $\Gamma(\alpha,\lambda)$ с параметрами $\alpha > 0$, $\lambda > 0$.
  Случайная величина $X \sim \Gamma(\alpha,\lambda)$, если она имеет плотность следующего вида
  $$p(x;\alpha,\lambda) = \begin{cases} \frac{\alpha^\lambda}{\Gamma(\lambda)} x^{\lambda-1} e^{-\alpha x} & \text{при} \quad x > 0, \\ 0 & \text{при} \quad x \leqslant 0. \end{cases}$$
  Параметр λ называют параметром формы, а параметр $\frac{1}{\alpha}$ является параметром масштаба, поскольку $\alpha X \sim \Gamma(\alpha,\lambda)$.
  $$\mathbf{E}\, X = \frac{\lambda}{\alpha}, \quad \mathbf{D}\, X = \frac{\lambda}{\alpha^2}, \quad f(t) = \left(1 - \frac{it}{\alpha}\right)^{-\lambda}.$$
  Отметим, что распределение $\Gamma(\alpha,1)$ совпадает с распределением $E(\alpha,0)$.

- **Распределение хи – квадрат** $\chi^2(k)$ с k степенями свободы, $k \in \mathbb{N}$.
Случайная величина $X \sim \chi^2(k)$, если ее плотность

$$p(x;k) = \begin{cases} \frac{x^{\frac{k}{2}-1}e^{-\frac{x}{2}}}{\Gamma(\frac{k}{2})2^{\frac{k}{2}}} & \text{при} \quad x > 0, \\ 0 & \text{при} \quad x \leqslant 0. \end{cases}$$

$$\mathbf{E}\, X = k, \quad \mathbf{D}\, X = 2k, \quad \mathbf{E}(X-k)^3 = 8k, \quad f(t) = (1-2it)^{-\frac{k}{2}}.$$

Данное распределение имеет случайная величина

$$X = \sum_{j=1}^{k} \xi_j^2,$$

где $\xi_1, \ldots, \xi_k$– независимые случайные величины с распределением $N(0,1)$. Это распределение совпадает с распределением $\Gamma(\frac{1}{2}, \frac{k}{2})$.

- **Бета – распределение**$\beta(r,s)$ с параметрами $r, s > 0$.
Случайная величина $X \sim \beta(r,s)$, если плотность распределения равна

$$p(x;r,s) = \begin{cases} \frac{\Gamma(r+s)}{\Gamma(r)\Gamma(s)} x^{r-1}(1-x)^{s-1} & \text{при} \quad x \in (0,1) \\ 0 & \text{при} \quad x \leqslant 0, x \geqslant 1. \end{cases}$$

$$\mathbf{E}\, X = \frac{r}{r+s}, \quad \mathbf{D}\, X = \frac{rs}{(r+s)^2(r+s+1)}.$$

Отметим, что распределение $\beta(1,1)$ совпадает с $U(0,1)$.

- **Распределение Стьюдента** t(n) с n степенями свободы, $n \in \mathbb{N}$.
Случайная величина $X \sim t(n)$, если она по распределению совпадает со случайной величиной

$$\frac{\xi_0}{\sqrt{\frac{1}{n}\sum\limits_{j=1}^{n} \xi_j^2}},$$

где $\xi_0, \xi_1, \ldots, \xi_n$ — независимые случайные величины с распределением $N(0;1)$. Плотность распределения X равна

$$p(x;n) = \frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi}\Gamma(\frac{n}{2})} \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}}.$$

$$\mathbf{E}\, X = 0\, (n > 1), \quad \mathbf{D}\, X = \frac{n}{n-2}\, (n > 2).$$

- **Распределение Коши**$K(a,b)$ с параметрами $-\infty < a < +\infty$, $b > 0$.
Случайная величина $X \sim K(a,b)$, если плотность распределения задается формулой

$$p(x;a,b) = \frac{b}{\pi(b^2 + (x-a)^2)}.$$

Математическое ожидание и дисперсия не существуют, $f(t) = e^{ita - b|t|}$. Параметры a и b являются параметрами сдвига и масштаба соответственно. Если $a = 0$, $b = 1$, то распределение Коши совпадает с распределением Стьюдента с одной степенью свободы.

- **Распределение Фишера**$F(m,n)$ с параметрами $n, m \in \mathbb{N}$.
Случайная величина $X \sim F(m,n)$, если она совпадает по распределению с величиной

$$\frac{\frac{1}{m}\xi}{\frac{1}{n}\eta}, \quad \xi \sim \chi^2(m),\ \eta \sim \chi^2(n),$$

причем ξ и η — независимые случайные величины.

- **Многомерное нормальное распределение**$N(\mathbf{a}, \mathbf{\Sigma})$,
$\mathbf{a} = (a_1, \ldots, a_n)^T$— действительный n–мерный вектор, $\mathbf{\Sigma}$— действительная неотрицательно определенная симметричная матрица порядка $n \times n$ с элементами σ_{ij}.
Случайный вектор $\mathbf{X} = (X_1, \ldots, X_n)^T$ имеет невырожденное (собственное) n-мерное распределение $N(\mathbf{a}, \mathbf{\Sigma})$, если $\det \mathbf{\Sigma} > 0$ и плотность распределения имеет вид

$$p(\mathbf{x}) = \frac{1}{(2\pi)^{n/2}\sqrt{\det \mathbf{\Sigma}}} \exp\left(-\frac{1}{2}(\mathbf{x}-\mathbf{a})^T \mathbf{\Sigma}^{-1}(\mathbf{x}-\mathbf{a})\right).$$

Характеристическая функция этого распределения равна

$$f(\mathbf{t}) = \mathbf{E}\, e^{i(\mathbf{t},\mathbf{X})} = \exp\left(i(\mathbf{t},\mathbf{a}) - \frac{1}{2}\mathbf{t}^T \mathbf{\Sigma} \mathbf{t}\right),$$

эта формула имеет смысл и в том случае, когда плотность распределения не существует ($\det \mathbf{\Sigma} = 0$.) Тогда распределение называют несобственным нормальным распределением. Несобственное нормальное распределение сосредоточено на некотором линейном подпространстве размерности $m = rg\mathbf{\Sigma} < n$.

Пусть $\mathbf{Y}$ —проекция случайного вектора на это подпространство и $m > 0$. Тогда $\mathbf{Y}$ имеет невырожденное m– мерное нормальное распределение.
Компоненты X_k случайного вектора $\mathbf{X}$ — это проекции на координатные оси,

$$X_k \sim N(a_k, \sigma_{kk}), \quad cov\,(X_k, X_l) = \sigma_{kl}.$$

В частном случае при $n = 2$ и $\mathbf{a} = \mathbf{0}$:

$$p(x_1, x_2) = \frac{1}{2\pi\sqrt{1-\varrho^2}} \exp\left(-\frac{1}{2(1-\varrho^2)}\left(\frac{x_1^2}{\sigma_1^2} - 2\varrho\frac{x_1 x_2}{\sigma_1\sigma_2} + \frac{x_2^2}{\sigma_2^2}\right)\right)$$

В этом случае пишут $(X_1, X_2) \sim N(a_1, a_2, \sigma_1^2, \sigma_2^2, \varrho)$, $\sigma_1 > 0, \sigma_2 > 0, |\varrho| < 1$.

$$\mathbf{E}\,X_k = a_k, \quad \mathbf{D}\,X_k = \sigma_k^2, \quad \varrho = cor(X_1, X_2).$$

Приложение 2. Экзаменационные вопросы по курсу "Теория вероятностей и математическая статистика"

1. Вероятностное пространство. Счетная аддитивность, монотонность вероятностной меры. Вероятность объединения событий. Лемма Бореля – Кантелли.

2. Независимость событий, случайных величин.

3. Случайные величины и их распределения вероятностей.

4. Биномиальное распределение. Пуассоновская аппроксимация (предельная теорема и неравенство).

5. Биномиальное распределение. Интегральная теорема Муавра – Лапласа (вывод ее из локальной теоремы Муавра – Лапласа или из центральной предельной теоремы).

6. Неравенство Чебышева и его уточнения. Закон больших чисел. Теорема Бернулли.

7. Вероятностное доказательство теоремы Вейерштрасса о равномерном приближении непрерывной на отрезке функции полиномами.

8. Классическое и геометрическое определение вероятности. Свойства вероятностей.

9. Формула композиции распределений и ее применение. Распределение суммы независимых нормально распределенных случайных величин. Распределение суммы независимых случайных величин, равномерно распределенных на отрезке $[0;1]$.

10. Неравенства для распределений максимума сумм независимых случайных величин.

11. Задача о разорении игрока.

12. Математические ожидания и их свойства.

13. Характеристические функции: формула обращения, теорема единственности (план доказательства), теорема непрерывности (без доказательства).

14. Центральная предельная теорема для одинаково распределенных случайных величин. Теорема Ляпунова (без доказательства).

15. Задача выбора одной из двух простых гипотез. Оценка снизу необходимого числа независимых наблюдений с помощью неравенства Йенсена.

16. Задача выбора одной из двух простых гипотез. Лемма Неймана – Пирсона и ее применение к проверке гипотез о математическом ожидании нормального распределения.

17. Задача выбора одной из двух простых гипотез. Лемма Неймана – Пирсона и ее применение к проверке гипотез о вероятности успеха в схеме Бернулли.

18. Несмещенные оценки. Неравенство Рао – Крамера. Эффективные оценки.

19. Эффективные оценки: метод максимального правдоподобия, оценки с дисперсией, меньшей, чем граница Рао – Крамера.

20. Асимптотическое распределение выборочной медианы в выборке из нормального распределения. Выборочная медиана как оценка неизвестного математического ожидания в этом случае.

21. Определение доверительного интервала. Доверительные интервалы для неизвестного математического ожидания нормального распределения (при известной и неизвестной дисперсии).

22. Доверительные интервалы для вероятности успеха в схеме Бернулли.

23. Различные виды сходимости последовательностей случайных величин.

24. Теорема о предельном распределении статистики критерия согласия Пирсона хи - квадрат.

25. Усиленный закон больших чисел.

26. Сходимость рядов из независимых случайных величин.

27. Эргодическая теорема для однородных цепей Маркова с конечным числом состояний.

28. Сложение независимых случайных величин по $mod\,1$. Сложение целочисленных случайных величин по $mod\,1$. Сходимость к равномерному распределению на соответствующей группе.

Литература

[1] А. А. Боровков, Теория вероятностей, Москва, 1986 г.

[2] Б. В. Гнеденко, Курс теории вероятностей, 1986 г.

[3] С. М. Ермаков, Метод Монте – Карло и смежные вопросы, Москва, изд -во "Наука", 1971г.

[4] Г. И. Ивченко, Ю. И. Медведев, Математическая статистика. Москва, 1984 г.

[5] А. Н. Колмогоров, С. В. Фомин, Элементы теории функций и функционального анализа, Москва, изд-во "Наука", 1968г.

[6] Г. Крамер, Математические методы статистики. Москва, 1975 г.

[7] Марков, Теория вероятностей.

[8] Б. А. Севастьянов, Курс теории вероятностей и математической статистики, Москва, 1982 г.

[9] В. Феллер, Введение в теорию вероятностей и ее приложения, т.1,2. Москва, 1984 г.

[10] А. Н. Ширяев, Вероятность, Москва, 1989 г.(1980 г.)

[11] П. Л. Чебышев, Курс лекций по теории вероятностей, изд - во АН СССР, 1936г.