Министерство образования и науки
Российской Федерации
Тверской государственный университет

Ю.С. ХОХЛОВ , О.И. СИДОРОВА, И.В. ЗАХАРОВА

ТЕОРИЯ ВЕРОЯТНОСТЕЙ
И МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

Учебное пособие

Тверь 2014

УДК 519.2

Хохлов Ю.С., Сидорова О.И., Захарова И.В. Теория вероятностей и математическая статистика: Учебное пособие. – Изд. 2-е, перераб. и доп. / ТвГУ.- Тверь, 2014.- 227 с.

Настоящее пособие состоит из двух частей. В первой части излагаются основные разделы теории вероятностей: вероятностное пространство, события, вероятности событий и их свойства, случайные величины и векторы и их распределения, числовые характеристики случайных величин, условные распределения, предельные теоремы. Во второй части излагаются основные понятия математической статистики: статистическая структура, точечные и интервальные оценки и их свойства, основы теории проверки статистических гипотез.

Рекомендуется студентам математических специальностей, а также экономистам.

Библиогр. 14.

Рецензенты: Кафедра математической статистики Московского госуниверситета им. М.В. Ломоносова;

доктор физико-математических наук
В.В. Сенатов

ISBN 5 - 7609 - 0101 - X

Оглавление

I Теория вероятностей **7**

1 Предмет теории вероятностей **8**

2 Дискретное вероятностное пространство **11**

2.1 Определение вероятностного пространства 11

2.2 События и операции над ними. 13

2.3 Вероятности событий и их свойства. 17

3 Элементы комбинаторики. Модели случайного выбора. **22**

4 Условная вероятность. Независимость событий. **27**

4.1 Условная вероятность 27

4.2 Независимость событий 29

4.3 Формулы полной вероятности и Байеса 31

5 Последовательности испытаний **35**

5.1 Определение последовательности испытаний 35

5.2 Схема Бернулли. Биномиальное распределение 37

5.3 Предельные теоремы в схеме Бернулли 40

5.4 Полиномиальное распределение 44

6 Определение вероятностного пространства в общем случае **47**

6.1 Постановка задачи 47

6.2 Определение σ–алгебры 48

6.3 Свойства непрерывности вероятности 51

7 Случайная величина и ее распределение **55**
7.1 Определение случайной величины и ее распределения . 55
7.2 Классификация распределений. 59
7.3 Примеры стандартных распределений 63
7.4 Функциональные преобразования случайной величины. 67

8 Случайный вектор **71**
8.1 Распределение случайного вектора 71
8.2 Классификация распределений 73
8.3 Независимые случайные величины 77
8.4 Функциональные преобразования случайных векторов 79

9 Математическое ожидание случайной величины **84**
9.1 Математическое ожидание дискретной случайной величины. 84
9.2 Определение математического ожидания в общем случае 87
9.3 Математическое ожидание как интеграл Римана–Стилтьеса 89
9.4 Свойства математических ожиданий 92

10 Числовые характеристики случайных величин и векторов **96**
10.1 Моменты 96
10.2 Дисперсия и ее свойства 100
10.3 Коэффициент корреляции и его свойства 104
10.4 Характеристики расположения и формы распределения 106

11 Гильбертово пространство случайных величин **109**

12 Условные распределения и условные математические ожидания **113**

13 Многомерное нормальное распределение **122**

14 Закон больших чисел **124**

15 Центральная предельная теорема **129**
15.1 Постановка задачи 129
15.2 Характеристические и производящие функции 132
15.3 Центральная предельная теорема 139

16 Цепи Маркова **145**
16.1 Определение цепи Маркова 145
16.2 Классификация состояний 149
16.3 Предельные теоремы для цепей Маркова 151
16.4 Примеры 156

17 Задачи для самостоятельного решения **160**

II Математическая статистика **166**

1 Основная задача математической статистики. Статистическая структура **167**

2 Выборка и выборочные характеристики **171**
2.1 Выборка 171
2.2 Эмпирическое распределение 174
2.3 Выборочные характеристики 177

3 Точечные оценки параметров **180**
3.1 Параметрические статистические структуры 180
3.2 Точечные оценки параметров и их свойства 184
3.3 Неравенство Рао–Крамера 188
3.4 Достаточные статистики 191
3.5 Полные статистики 195
3.6 Построение оптимальных оценок 198
3.7 Методы построения оценок 200
3.7.1 Метод моментов 200
3.7.2 Метод наибольшего правдоподобия 201

4 Интервальные оценки параметров **203**

4.1 Определение доверительного интервала 203

4.2 Некоторые распределения, связанные с нормальным . . 205

4.3 Построение доверительных интервалов для параметров нормального распределения 206

4.4 Построение доверительных интервалов с помощью центральных статистик 208

4.5 Построение доверительных интервалов с помощью асимптотически нормальных оценок 210

5 Критерии согласия **213**

5.1 Общий метод построения критериев согласия 213

5.2 Критерий согласия Колмогорова 214

5.3 χ^2–критерий согласия Пирсона 215

5.4 Проверка однородности двух выборок 217

5.5 Проверка гипотезы о независимости 218

6 Проверка статистических гипотез **220**

6.1 Что такое статистическая гипотеза? 220

6.2 Основные понятия теории проверки гипотез 221

6.3 Проверка простой гипотезы против простой альтернативы 224

6.4 Проверка сложных гипотез. Критерий отношения правдоподобия 228

6.5 Проверка сложных гипотез для распределений с монотонным отношением правдоподобия 231

Часть I

Теория вероятностей

Глава 1

Предмет теории вероятностей

Теория вероятностей является математической наукой. Как и другие разделы математики, она имеет дело не напрямую с объектами и явлениями реального мира, а с их математическими моделями. Процесс построения математической модели, ее исследования и применения — довольно сложный многоступенчатый процесс. Вначале, наблюдая за реальным объектом, проводя эксперименты и накапливая факты, строят модели конкретных наук, таких, как физика, химия, биология, экономика и т.д. Далее, сравнивая различные модели, выделяя их существенные черты, отвлекаются от конкретных объектов и исследуют только структуру моделей. Это и приводит нас к математической модели. Изучая свойства математической модели, приходят к открытию новых эффектов и предсказанию поведения этого объекта в новых условиях, где он ранее не наблюдался. Таким образом, мы вновь возвращаемся к реальному объекту.

До возникновения теории вероятностей предметом математики были только модели таких явлений, в которых исход тех или иных экспериментов однозначно определялся заданием некоторого комплекса начальных условий. Классическим примером являются механические модели. Эксперименты такого типа будем называть **детерминированными**. Но нетрудно привести примеры и таких экспериментов, когда при возможной точности фиксации начальных условий исход эксперимента однозначно не определен, т.е. некоторое событие $\mathbf{A}$ при заданном комплексе условий $\mathbf{K}$ иногда происходит, а иногда нет. Такие эксперименты мы будем называть **экспериментами с неопределенным исходом** (термин «случайный» мы сохраним для более точных формулировок). Классическим примером является подбрасывание монеты. Какого типа закономерности можно изучать для таких неопределенных ситуаций? Сразу же оговоримся, что теория вероятностей имеет дело не с любыми экспериментами с неопределенным исходом, а только с так называемыми **массовыми явлениями**, когда эксперимент проводится большое число раз в одинаковых условиях или рассматривается большая совокупность однородных объектов. Типичный вопрос теории вероятностей (с практической точки зрения) — насколько часто происходит данное событие **A** при заданном комплексе условий **K** в длинной серии испытаний. Чтобы перейти к более аккуратным формулировкам, уточним некоторые понятия. Пусть мы N раз провели некоторый эксперимент, в котором событие A произошло в $N(A)$ испытаниях. **Относительной частотой** появления события A в N испытаниях называется число

$$h_N(A) \;=\; \frac{N(A)}{N}\;.$$

Чтобы понять, модели каких экспериментов изучаются в теории вероятностей, рассмотрим простейший пример — подбрасывание монеты. Выполним 6 серий по 100 подбрасываний и в каждой серии подсчитаем число $N(A)$ появлений события A, когда монета падала гербом вверх.

$$N = 100.$$

Серия	N(A)	$h_N(A)$
1	56	0,56
2	48	0,48
3	52	0,52
4	50	0,50
5	47	0,47
6	46	0,46

‘

Сгруппируем эти результаты в три серии по 200 подбрасываний.

Серия	N(A)	$h_N(A)$
1	104	0,520
2	102	0,510
3	93	0,465

Можно увеличивать число испытаний в одной серии и следить за поведением относительной частоты $h_N(A)$. В результате мы получаем последовательно: 0,56; 0,52; 0,52; 0,515; 0,506; 0,498. На этом примере мы видим, что выполнены следующие свойства:

1) относительная частота $h_N(A)$ события A в длинной серии испытаний «тяготеет» к некоторому постоянному неслучайному числу;

2) в разных сериях испытаний, но проводимых в одинаковых условиях, относительные частоты приблизительно равны;

3) если мы из данной серии испытаний выберем некоторую подсерию, не используя информацию о результатах эксперимента, то новая относительная частота тяготеет к тому же числу.

Будем говорить, что для некоторого эксперимента выполнено свойство **статистической устойчивости частот**, если выполнены свойства 1-3. **Случайным экспериментом** будем называть такой, в котором выполнено свойство устойчивости частот.

Теория вероятностей — это раздел математики, где изучаются модели массовых случайных явлений, для которых выполняется свойство устойчивости частот.

Число, к которому тяготеет относительная частота $h_n(A)$, будем называть **вероятностью** этого события. Вероятность события A измеряет меру возможности его появления в случайном эксперименте.

В теории вероятностей изучаются свойства вероятностей различных событий. *Основная задача теории вероятностей как математической науки состоит в том, чтобы, зная вероятности одних событий, вычислить вероятности других, так или иначе связанных с первыми.*

Глава 2

Дискретное вероятностное пространство

2.1 Определение вероятностного пространства

При построении математической модели мы должны найти компромисс между двумя обстоятельствами. С одной стороны, она должна быть достаточно подробной, чтобы учесть все существенные черты изучаемого явления. С другой стороны, необходимо отбросить все несущественные детали, затемняющие суть дела. Излишняя подробность затрудняет изучение свойств модели, а чрезмерное упрощение может привести к неправильным выводам относительно поведения реальной системы.

Мы начинаем изучение курса теории вероятностей с исследования свойств моделей таких случайных экспериментов, которые имеют конечное или счетное число исходов. **Элементарным исходом** мы будем называть такое событие, которое однозначно (с определенной точки зрения) говорит о том, чем закончился эксперимент. Это сразу же накладывает на множество элементарных исходов следующее важное ограничение: в каждом испытании происходит один и только один элементарный исход.

Чтобы понять, как должна выглядеть наша модель, рассмотрим пример. Однородный игральный кубик в одинаковых условиях подбрасывают много раз и отмечают число очков, выпавших на верхней грани. Ясно, что в этом эксперименте есть 6 элементарных исходов, которые мы обозначим $\omega_1, \omega_2, \ldots, \omega_6$ (ω_k означает, что выпало k очков). Пусть $h_N(\omega_k)$ — относительная частота появления исхода ω_k. Тогда эти частоты обладают следующими свойствами:

1) $\quad h_N(\omega_k) \geqslant 0, \ \forall k,$

2) $\quad \sum\limits_{k=1}^{6} h_N(\omega_k) = 1.$

Как отмечалось выше, частоты тяготеют к некоторым числам, которые мы будем называть вероятностями этих исходов. Ясно, что они должны наследовать свойства частот. Эти предварительные рассмотрения приводят нас к следующему определению.

Определение 1 . Дискретным вероятностным пространством *называется пара* (Ω, P)*, где* Ω *- конечное или счетное множество,* P *— вещественная функция, заданная на* Ω*, такая, что*

1) $\quad P(\omega) \geqslant 0, \ \forall \omega \in \Omega,$

2) $\quad \sum\limits_{\omega \in \Omega} P(\omega) = 1.$

Множество Ω называется **пространством элементарных исходов**, его элементы ω — **элементарными исходами**, а число $P(\omega)$ — **вероятностью появления** элементарного исхода ω.

Пример 1. Симметричную монету подбрасывают один раз. Здесь два элементарных исхода: выпал герб — Г, выпала цифра — Ц. Таким образом, $\Omega = \{,\}$. В силу симметрии естественно положить $P() = 1/2, \ P() = 1/2$.

Пример 2. Однородный симметричный игральный кубик подбрасывают один раз. В этом случае

$$\Omega = \{1, 2, 3, 4, 5, 6\}, \ P(\omega) = 1/6.$$

Другие примеры будут приведены на практических занятиях. Важную роль играет следующий частный случай дискретного вероятностного пространства.

Определение 2 . *Говорят, что мы имеем задачу на классическое определение вероятности, если $\Omega = \{\omega_1, \ldots, \omega_n\}$ — конечное множество и для всех ω_i, $P(\omega_i) = 1/n$, т.е. все исходы* **равновозможны**.

Обычно предположение о равновозможности исходов делается из соображений симметрии задачи. Но так ли это на самом деле (т.е. верна ли модель), можно установить только из сравнения с экспериментальными данными.

2.2 События и операции над ними.

До сих пор мы рассматривали только элементарные исходы, т.е. в некотором смысле простейшие события. Но кроме них нас могут интересовать и другие, более сложные события. В примере 2 мы можем рассмотреть событие A, состоящее в том, что выпало четное число очков. В теории вероятностей о каждом событии мы хотим знать только одно: произошло оно или нет в данном испытании. Каждое испытание (т.е. однократное проведение эксперимента) заканчивается появлением одного из элементарных исходов, которые однозначно описывают то, чем закончился эксперимент. В частности, по элементарному исходу ω можно определить, произошло событие A или нет. Поэтому все элементарные исходы делятся на две группы: те ω, которые приводят к появлению события A (назовем их **благоприятными** этому событию), и все остальные. С точки зрения их появления в рассматриваемом эксперименте событие A и множество благоприятных для него исходов являются для нас эквивалентными. Таким образом, мы приходим к следующему определению.

Определение 3 . **Случайным событием** *назовем произвольное подмножество A пространства элементарных исходов Ω.*

Будем говорить, что событие произошло, если появился элементарный исход, ему принадлежащий, т.е. благоприятный.

Пример 3. Подбрасывают игральный кубик, A — выпала четная цифра. Тогда $A = \{2, 4, 6\}$.

В силу того, что каждое случайное событие отождествляется с некоторым подмножеством A пространства элементарных исходов Ω, различные операции над множествами позволяют определить некоторые операции над событиями. С точки зрения теории вероятностей каждое событие характеризуется только тем, когда оно происходит, а когда нет. Поэтому определения операций над событиями даются именно в этих терминах. С другой стороны, они соответствуют определенным операциям над множествами. Отсюда появляется определенная двойственность терминологии.

Определение 4 . *1) Событие называется* **достоверным***, если оно происходит всегда, и* **невозможным***, если оно никогда не происходит.*

Этим событиям соответствуют все пространство Ω и пустое множество $\emptyset$.

2) **Объединением** *двух событий A и B называется такое событие C, которое происходит тогда и только тогда, когда происходит хотя бы одно из этих двух событий.*

На языке теории множеств это соответствует операции объединения множеств и обозначается как $C = A \cup B$.

3) **Пересечением** *двух событий A и B называется такое событие C, которое происходит тогда и только тогда, когда происходят оба эти события одновременно.*

Это соответствует операции пересечения множеств и обозначается $C = A \cap B$ или $C = AB$.

4) События A и B называются **несовместными (непересекающимися)***, если они не могут происходить одновременно.*

Это соответствует непересекающимся множествам и обозначается $AB = \emptyset$.

5) **Суммой** *событий A и B называется их объединение в случае, когда они несовместны.*

Это не новая операция, а частный случай определения 2 и обозначается $A + B$.

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

На языке теории множеств это соответствует переходу к дополнению множества A.

7) **Разностью** *двух событий A и B называется такое событие C, которое происходит тогда и только тогда, когда происходит A и не происходит B.*

Это соответствует операции разности множеств и обозначается $C = A \setminus B$.

8) Говорят, что событие A **влечет** *событие B, если при появлении события A, обязательно происходит и событие B.*

Это означает, что множество A есть часть (подмножество) множества B и обозначается $A \subset B$.

Чтобы наглядно представлять себе операции над событиями, полезно рисовать их в виде некоторых фигур на плоскости, например кругов. Картинки такого рода называются **диаграммами Венна**.

Конкретные примеры событий и операций над ними будут рассмотрены далее , а также на практических занятиях. Некоторые свойства операций над событиями собраны в следующем предложении.

Предложение 1

1) $A \cup A = A$, 2) $A \cap A = A$, 3) $A \cup \overline{A} = \Omega$,
4) $A \cap \overline{A} = \emptyset$, 5) $A \cup \Omega = \Omega$, 6) $A \cap \Omega = A$,
7) $A \cup \emptyset = A$, 8) $A \cap \emptyset = \emptyset$, 9) $A \subset B, B \subset C,\ A \subset C$,
10) $A \subset B,\ B = A + (B \setminus A)$, 11) $A \setminus B = A \cap \overline{B}$,
12) $A \cup B = A + (B \setminus A) = A + \overline{A}B$, 13) $A \cup (B \cup C) = (A \cup B) \cup C$,
14) $A \cap (B \cap C) = (A \cap B) \cap C$,
15) $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$,
16) $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$,
17) $\bigcup\limits_{\alpha \in I} \overline{A_\alpha} = \overline{\bigcap\limits_{\alpha \in I} A_\alpha}$,
18) $\bigcap\limits_{\alpha \in I} \overline{A_\alpha} = \overline{\bigcup\limits_{\alpha \in I} A_\alpha}$,
19) $A_1 \cup A_2 \cup \ldots \cup A_n = A_1 + \overline{A_1}A_2 + \overline{A_1}\,\overline{A_2}A_3 + \ldots + \overline{A_1}\ldots\overline{A_n}A_n$.

Задача 1. Доказать предложение 1.

Во многих задачах нас интересует не все множество событий, связанных с данным экспериментом, а только некоторые из них. Но всегда нам хотелось бы, чтобы определенные выше операции над событиями не выводили нас за пределы рассматриваемого множества событий. В связи с этим полезно следующее понятие.

Определение 5 . *Некоторый класс* $\mathcal{A}$ *событий называется* **алгеброй событий**, *если*
1) $\Omega \in \mathcal{A}$,
2) если $A \in \mathcal{A}$, *то* $\overline{A} \in \mathcal{A}$,
3) если $A, B \in \mathcal{A}$, *то* $A \cup B \in \mathcal{A}$.

Задача 2. Доказать, что все определенные выше операции не выводят нас за пределы алгебры $\mathcal{A}$.

С практической точки зрения выбор некоторой алгебры событий соответствует определенному взгляду на случайный эксперимент. Алгебра событий — это только те события, которые нас интересуют с этой точки зрения (например, те, которые доступны наблюдению).

Пример 4. Монету подбросили 2 раза. Пусть $\omega_i = 1$, если выпал герб и $\omega_i = 0$, если выпала решка, $i = 1, 2$.

Пространство элементарных исходов в этом эксперименте будет иметь вид
$$\Omega = \Big\{(0,0), (0,1), (1,0), (1,1)\Big\}$$

Рассмотрим для примера 4 класса событий:

$\mathcal{A}_0 = \Big\{\emptyset, \Omega\Big\}$ — тривиальная алгебра, содержащая информацию об эксперименте до момента его проведения (самая бедная).

$\mathcal{A}_1 = \Big\{\emptyset, \Omega, \{(0,0), (0,1)\}, \{(1,0), (1,1)\}\Big\}$ — алгебра, содержащая информацию о первом броске монеты.

$\mathcal{A}_2 = \Big\{\emptyset, \Omega, \{(0,0), (1,0)\}, \{(0,1), (1,1)\}\Big\}$ — алгебра, содержащая информацию о втором броске монеты.

$\mathcal{A}_3 = \Big\{\emptyset, \Omega, \text{все подмножества пространства}\, \Omega\Big\}$ — алгебра, содержащая полную информацию об эксперименте.

Имеем $\mathcal{A}_0 \subset \mathcal{A}_1 \subset \mathcal{A}_3$ и $\mathcal{A}_0 \subset \mathcal{A}_2 \subset \mathcal{A}_3$.

2.3 Вероятности событий и их свойства.

До сих пор мы рассматривали только вероятности элементарных исходов. Теперь мы определим вероятности событий и исследуем некоторые их свойства.

Определение 6 . **Вероятностью** *события A называется число*

$$P(A) := \sum_{\omega \in A} P(\omega). \tag{2.1}$$

Пример 5. Симметричный игральный кубик подбрасывают один раз. Найти вероятность события A, состоящего в том, что выпала четная цифра.

В этом случае $A = \{2, 4, 6\}$, $P(A) = P(2) + P(4) + P(6) = 3/6 = 1/2$.

Пример 6. Пусть мы имеем задачу на классическое определение вероятности. Если $\Omega = \{\omega_1, \ldots, \omega_n\}$, где n - общее число элементарных исходов, а m — число благоприятных исходов для события A, то

$$P(A) = \underbrace{\frac{1}{n} + \ldots + \frac{1}{n}}_{m} = \frac{m}{n}.$$

Именно этот результат обычно приводят в качестве определения в элементарных учебниках по теории вероятностей.

Соберем некоторые простейшие свойства вероятностей в виде следующего предложения.

Предложение 2 . *Пусть выделена некоторая алгебра $\mathcal{A}$ событий, для которых определены вероятности по формуле (1). Тогда справедливы следующие свойства:*

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

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

3. Если $A_1,\ A_2,\ \ldots,\ A_n$ — попарно несовместны, то

$$P\Big(\sum_{k=1}^{n} A_k\Big) = \sum_{k=1}^{n} P(A_k).$$

Это свойство называется **теоремой сложения** *или свойством* **конечной аддитивности** *вероятности.*

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

5. $P(A \cup B) = P(A) + P(B) - P(AB)$.

6. $A \subset B,\quad P(B \setminus A) = P(B) - P(A)$.

7. $A \subset B,\quad P(A) \leq P(B)$.

8. $P\Big(\bigcup\limits_{k=1}^{n}\Big) \leqslant \sum\limits_{k=1}^{n} P(A_k)$ — *свойство полуаддитивности вероятности.*

9. $P(A) \leqslant 1$.

10. $P\Big(\bigcup\limits_{k=1}^{n} A_k\Big) = \sum\limits_{k=1}^{n} P(A_k) - \sum\limits_{i<j} P(A_iA_j) + \sum\limits_{i<j<k} P(A_iA_jA_k) - \ldots (-1)^n P(A_1 \ldots A_n)$

— *формула включения и исключения.*

Доказательство. Основными являются свойства 1-3. Только здесь мы будем использовать в явном виде то, что мы работаем в рамках дискретного вероятностного пространства. Все остальные свойства будут выведены из этих трех.

Свойства 1 и 2 вытекают непосредственно из определений дискретного вероятностного пространства и вероятности события.

Докажем свойство 3. Пусть, вначале, $n = 2$. Тогда мы имеем два события A_1 и A_2. Так как они несовместны, то исходы $\omega \in A_1 + A_2$ распадаются на два непересекающихся класса: те, что принадлежат A_1, и те, что принадлежат A_2. В силу свойств рядов с неотрицательными членами имеем

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

Для произвольного n доказательство проводится по индукции. Предлагается это сделать самостоятельно.

Свойство 4 очевидно следует из определения вероятности события и того, что сумма вероятностей всех элементарных исходов равна 1.

События A и $\overline{A}$ несовместны, и $A + \overline{A} = \Omega$. Используя свойства 2 и 3, имеем:

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

Представим события A, B и $A \cup B$ в следующем виде:

$$A = AB + (A \setminus B),\ B = AB + (B \setminus A),\ A \cup B = AB + (A \setminus B) + (B \setminus A).$$

Далее используем свойство 3:

$$P(A) = P(AB) + P(A \setminus B),\ P(B) = P(AB) + P(B \setminus A),$$

$$P(A \cup B) = P(A) + P(B) - P(AB).$$

Этим доказано свойство 5.

Свойства 6,7,8,9 можно доказать аналогично (сделать самостоятельно!).

Докажем свойство 10. Для упрощения доказательства введем одно новое понятие, которое будет полезным и в других вопросах.

Индикатором события A называется функция $I(A, \omega)$, заданная на пространстве элементарных исходов Ω по правилу:

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

Легко доказать следующие свойства индикаторов событий:

1) $I(\Omega, \omega) \equiv 1,\ I(\emptyset, \omega) \equiv 0,$

2) $I(\overline{A}, \omega) = 1 - I(A, \omega),$

3) $I\left(\bigcap\limits_{k=1}^{n} A_k, \omega\right) = \prod\limits_{k=1}^{n} I(A_k, \omega).$

Используя понятие индикатора события, можно в следующем виде записать определение вероятности события:

$$P(A) = \sum_{\omega \in \Omega} P(\omega) I(A, \omega). \tag{2.2}$$

Применим этот результат к доказательству свойства 10. В силу предложения 2.1 мы имеем $\overline{\bigcup\limits_{k=1}^{n} A_k} = \bigcap\limits_{k=1}^{n} \overline{A_k}$. Тогда для индикаторов получаем

$$\begin{aligned}
1 - I\Big(\bigcup_{k=1}^{n} A_k, \omega\Big) &= I\Big(\overline{\bigcup_{k=1}^{n} A_k}, \omega\Big) = I\Big(\bigcap_{k=1}^{n} \overline{A_k}, \omega\Big) \\
&= \prod_{k=1}^{n} I(\overline{A_k}, \omega) = \prod_{k=1}^{n} \Big(1 - I(\overline{A_k}, \omega)\Big) = \\
&= 1 - \sum_{k=1}^{n} I(A_k, \omega) + \sum_{i<j} I(A_i A_j, \omega) - \\
&\quad - \sum_{i<j<k} I(A_i A_j A_k, \omega) + \ldots
\end{aligned}$$

Дальше воспользуемся формулой (2), свойством аддитивности суммы ряда и перейдем к противоположному событию.

Из вышеизложенного следует, что основой нашей модели является некоторое множество (алгебра) событий и вероятности этих событий, обладающие свойствами 1–3 из предложения 2.2. Поэтому мы можем дать новое определение вероятностного пространства.

Определение 7 . Вероятностным пространством *(в слабом смысле) называется тройка* $(\Omega, \mathcal{A}, P)$*, где* Ω *— произвольное множество,* $\mathcal{A}$ *— некоторая алгебра его подмножеств, а* P *— вещественная функция, заданная на* $\mathcal{A}$ *и обладающая свойствами:*
1) $\forall A \in \mathcal{A}, P(A) \geqslant 0$,
2) $P(\Omega) = 1$,
3) если $A_1, \ldots, A_n$ *— попарно несовместны , то*

$$P\Big(\sum_{k=1}^{n} A_k\Big) = \sum_{k=1}^{n} P(A_k).$$

Множество Ω *называется* **пространством элементарных исходов***, элементы алгебры* $\mathcal{A}$ *называется* **событиями***, число* $P(A)$ *—* **вероятностью** *события* A.

Преимущество такого определения в том, что оно применимо и к некоторым ситуациям, в которых мы имеем несчетное множество элементарных исходов. Одним из первым был рассмотрен следующий частный случай.

Определение 8 . *Говорят, что мы имеем задачу на* **геометрическое определение вероятности**, *если* Ω *есть ограниченное борелевское подмножество в* R^d, $\mathcal{A}$ *— алгебра всех его борелевских подмножеств, а вероятность событий задается по правилу*

$$P(A) = \frac{L(A)}{L(\Omega)},$$

где $L(A)$ *— мера Лебега(длина, площадь, объем, ...) множества* A.

Пример 7. Из отрезка [0,1] случайным образом выбирают точку. Найти вероятность того, что она лежит на расстоянии не более $1/4$ от точки 0.

В результате такого эксперимента мы получаем некоторую точку $\omega \in [0,1]$. Поэтому естественно положить $\Omega = [0,1]$. Если воспользоваться геометрическим определением вероятности,то получаем $A = [0,1/4]$ и

$$P(A) = \frac{L(A)}{L(\Omega)} = \frac{1/4}{1} = 1/4.$$

Глава 3

Элементы комбинаторики. Модели случайного выбора.

Вопросы, которые рассматриваются в этом параграфе, не имеют особого значения с точки зрения общей теории, так как здесь изучается некоторая частная модель, связанная с классическим определением вероятности. Но модели случайного выбора широко используются в экономике, социологии, демографии и других науках. В силу этого они очень важны с практической точки зрения.

Как мы видели выше, при классическом определении вероятности нам нужно решить две задачи: вычислить число всех элементарных исходов n и число m благоприятных исходов для события A. Обычно это сводится к подсчету того, сколькими способами может быть выполнено некоторое действие, т. е. сколько существует вариантов. Задачами такого типа занимается специальный раздел математики, называемый **комбинаторикой**. Мы начнем изучение основ комбинаторики с одной элементарной, но очень полезной леммы.

Лемма 1 (**умножения**). *Пусть некоторое действие осуществляется в два этапа. На первом этапе мы имеем n_1 вариантов его осуществления, а на втором, независимо от того, что произошло на первом этапе, n_2 - вариантов. Тогда общее число вариантов осуществления такого действия равно $n_1 n_2$.*

Доказательство этой леммы сводится к установлению взаимно однозначного соответствия между двухэтапными вариантами и клетками таблицы размера $n_1 \times n_2$.

Нас в основном будут интересовать задачи комбинаторики, которые можно сформулировать в терминах выбора некоторого количества предметов из заданной совокупности $U = \{a_1, \dots, a_k\}$. При этом необходимо учитывать два обстоятельства: как производился выбор и как фиксировался результат выбора. В первом случае мы имеем два варианта: **выбор с повторением**, когда на каждом этапе выбранный предмет возвращается назад и выбор производится из одной и той же совокупности; **выбор без повторения**, когда выбранный на некотором шаге предмет уже не используется в дальнейшем. При фиксации результата мы также имеем два варианта: **с учетом порядка** (в этом случае мы знаем, какие выбраны предметы и в каком порядке) и **без учета порядка** (здесь мы знаем только, какие были выбраны предметы, но не имеем информации о порядке их появления). Имея в виду вероятностные и статистические применения рассматриваемых задач, мы будем называть результат выбора элементарным исходом или **выборкой**. Число элементов в выборке называется ее **объемом**. Рассмотрим несколько стандартных примеров.

Определение 9 . Размещением с повторением *из k элементов по l называется произвольная упорядоченная выборка с возвращением, объема l из совокупности U.*

В этом случае $\omega = (\omega_1, \dots, \omega_l)$, где $\omega_i \in U$. Используя лемму умножения, легко подсчитать ,что число различных элементарных исходов будет равно $n = k^l$.

Определение 10 . Размещением *(без повторения) из k элементов по l называется произвольная упорядоченная выборка без возвращения объема l из совокупности U.*

Здесь $\omega = (\omega_1, \dots, \omega_l),\ \omega_i \in U,\ \omega_i \neq \omega_j,\ i \neq j$. Вновь по лемме умножения получаем ,что число различных размещений равно

$$A_k^l = k(k-1)\dots[k-(l-1)].$$

Если $l = k$, то такое размещение называется **перестановкой**. Число различных перестановок k элементов равно

$$P_k = k(k-1)\dots 1 = k!$$

Определение 11 . Сочетанием *(без повторений) из k элементов по l называется произвольная неупорядоченная выборка без возвращения объема l из совокупности U.*

В этом случае мы отмечаем только, какие именно элементы вошли в эту выборку, и не учитываем порядок их появления. Поэтому для каждого элемента достаточно указать, входит он в выборку или нет.Таким образом, $\omega = (\omega_1, \dots, \omega_k)$, где $\omega_i = 1$, если a_i входит в выборку ω, и равно 0 в противном случае. Из каждого сочетания, переставляя его элементы, мы получаем $l!$ различных размещений. В силу этого число сочетаний в $l!$ раз меньше числа размещений, т. е.

$$C_k^l = \frac{A_k^l}{l!} = \frac{k(k-1)\dots[k-(l-1)]}{l!} = \frac{k!}{l!(k-l)!}.$$

Определение 12 . Сочетанием с повторением *из k элементов по l называется произвольная неупорядоченная выборка с возвращением объема l из совокупности U.*

Так как теперь один элемент может входить в выборку несколько раз, то элементарный исход имеет вид $\omega = (\omega_1, \dots, \omega_k)$, где $\omega_i = q$ означает, что элемент a_i входит в выборку q раз. Для подсчета числа таких выборок применим следующий вспомогательный прием. Выпишем сначала все элементы a_1, затем a_2 и так далее. Между элементами разных типов поставим перегородки. Таким образом, мы имеем l элементов и $k-1$ перегородку. Если каких–то элементов нет, то перегородки стоят рядом. Для задания сочетания с повторением достаточно расставить $k-1$ перегородку на $k-1+l$ возможных мест без учета порядка. Это можно сделать C_{k-1+l}^{k-1} способами.

Рассмотрим два важных с практической точки зрения примера.

Пример 1. Пусть мы имеем некоторую совокупность, содержащую N предметов, из которых N_1 — одного типа и N_2 — другого типа.

С возвращением выбираем k предметов из этой совокупности. Найти вероятность события того,что среди них будет l предметов первого типа.

Так как выбор осуществляется с возвращением, то мы имеем всего N^k различных элементарных исходов в этой задаче. Благоприятные исходы легче пересчитать, используя следующее замечание. Сначала мы должны выбрать, на каких шагах мы будем отбирать предметы первого типа. Это можно сделать C_k^l способами. Затем на каждом из этих шагов мы имеем N_1^l вариантов, а на остальных- N_2^{k-l} вариантов. По лемме умножения число благоприятных исходов будет равно $m = C_k^l N_1^l N_2^{k-l}$. Если обозначить через $p = N_1/N$ долю предметов первого типа во всей совокупности, то получим

$$P(A) = C_k^l p^l (1-p)^{k-l}.$$

Пример 2. Пусть мы имеем ту же задачу, что и в примере 1, но выбор осуществляется без возвращения.

В этом примере удобнее фиксировать результат без учета порядка следования элементов. Тогда общее число элементарных исходов $n = C_N^k$. Для получения благоприятного исхода нужно в два этапа набрать l элементов первого типа (из N_1) и $k-l$ элементов второго типа (из N_2), т. е. по лемме умножения $m = C_{N_1}^l C_{N_2}^{k-l}$. Тогда по классическому определению вероятности получаем

$$P(A) = \frac{C_{N_1}^l C_{N_2}^{k-l}}{C_N^k}.$$

Задача 1. Пусть в примере $N, N_1, N_2 \to \infty$ так, что $N_1/N \to p$, где $p \in (0,1)$, а k и l — фиксированы. Тогда

$$P(A) \to C_k^l p^l (1-p)^{k-l}.$$

Таким образом, в этом случае выбор с возвращением и выбор без возвращения эквивалентны. Легко понять, почему так получилось. Дело в том, что для больших совокупностей удаление небольшого числа элементов не меняет практически пропорций, т. е. мы имеем на каждом шаге выбор из той же совокупности.

Пример 3. В некотором городе живет 100 000 человек, среди которых 60 000 женщин. Для проведения социологического обследования производят выборку объема 500. Найти вероятность того, что среди них будет 300 женщин.

Так как отбор 500 человек не изменит существенно пропорций, то можно считать, что мы имеем выбор с возвращением. Здесь $N = 100000$, $N_1 = 60000$, $p = N_1/N = 0.6$, $k = 500$, $l = 300$. Тогда

$$P(A) = C_{500}^{300}(0.6)^{300}(0.4)^{200}.$$

Примеры 1 и 2 можно обобщить на случай выбора из совокупностей, где есть предметы r типов ($r \geqslant 2$).

Примеры других задач будут приведены на практических занятиях.

Глава 4

Условная вероятность. Независимость событий.

4.1 Условная вероятность

Как отмечалось в начале нашего курса, мы подразумеваем, что эксперимент проводится при некотором фиксированном комплексе условий **K**. Если эти условия изменились, то изменяется и вероятность событий, относящихся к этому эксперименту. Такое изменение всегда можно понимать как появление некоторого события (кроме исходного комплекса условий **K**). Чтобы понять, как определить в этом случае новую (условную) вероятность, рассмотрим соответствующие частоты. Пусть эксперимент проведен N раз, событие B произошло $N(B)$ раз, а события A и B вместе $N(AB)$ раз. Тогда «условная» частота события A среди тех экспериментов, где произошло событие B, равна

$$h_N(A|B) = \frac{N(AB)}{N(B)} = \frac{N(AB)/N}{N(B)/N} = \frac{h_N(AB)}{h_N(B)}.$$

Имея в виду, что вероятность наследует свойства частот, можно дать следующее

Определение 13 . Условной вероятностью *события A при усло-*

вии, что произошло событие B ($P(B) \neq 0$), называется число

$$P(A|B) := \frac{P(AB)}{P(B)}.$$

Иногда применяют и другое обозначение $P_B(A)$.

Пример 1 . *Симметричную монету подбрасывают два раза. Известно, что выпал один герб (событие B). Найти вероятность события A, состоящего в том, что герб выпал при первом бросании.*

Легко вычислить, что $P(B) = 1/2$, а $P(AB) = 1/4$. Отсюда следует, что $P(A|B) = (1/4)/(1/2) = 1/2$.

Нетрудно проверить, что при фиксированном B условная вероятность обладает следующими свойствами:

1. $P(A|B) \geqslant 0$,
2. $P(\Omega|B) = 1$,
3. Если $A_1, \ldots, A_n$ — попарно несовместны, то

$$P\Big(\sum_{k=1}^{n} A_k|B\Big) = \sum_{k=1}^{n} P(A_k|B).$$

Таким образом, условная вероятность обладает всеми основными свойствами вероятности.

Очень важную роль играет следующая

Теорема умножения. *Пусть A и B — два события и $P(B) > 0$. Тогда*

$$P(AB) = P(B)P(A|B).$$

Ее доказательство следует из определения условной вероятности. Польза этой теоремы состоит в том, что иногда мы можем вычислить условную вероятность непосредственно и затем использовать это для вычисления $P(AB)$.

Пример 2 . *В урне 5 шаров — 3 белых и 2 черных. Без возвращения выбираем два шара. Найти вероятность того, что оба шара белые.*

Пусть событие A_1 состоит в том, что первый шар белый, а событие A_2 — второй шар белый. Легко вычислить, что $P(A_1) = 3/5$. После того, как мы вынули один шар и знаем, что он белый, мы имеем 4 шара и среди них 2 белых. Тогда $P(A_2|A_1) = 2/4$. По теореме умножения

$$P(A_1 A_2) = P(A_1)P(A_2|A_1) = 3/5 \cdot 2/4 = 3/10.$$

Теорему умножения легко распространить на любое конечное число событий.

Следствие 1 . *Пусть $A_1, \ldots, A_n$ — случайные события, тогда*

$$P(A_1 \ldots A_n) = P(A_1)P(A_2|A_1)P(A_3|A_1 A_2) \ldots P(A_n|A_1 \ldots A_{n-1}).$$

4.2 Независимость событий

Если появление события B не меняет вероятности события A, т. е. $P(A|B) = P(A)$, то такие события естественно назвать независимыми. В этом случае по теореме умножения мы получаем $P(AB) = P(A)P(B)$. Последнее соотношение симметрично относительно A и B и имеет смысл при $P(B) = 0$. Поэтому мы возьмем его в качестве определения.

Определение 14 . *События A и B называются* **независимыми**, *если*

$$P(AB) = P(A)P(B).$$

Пример 3 . *Подбрасывают две симметричных монеты. Событие A состоит в том, что на первой монете выпал герб, а событие B — на второй монете выпал герб.*

Интуитивно ясно, что такие события должны быть независимыми. Действительно, $P(A) = 1/2$, $P(B) = 1/2$, $P(AB) = 1/4 = 1/2 \cdot 1/2 = P(A)P(B)$. Таким образом A и B — независимы в смысле определения. Менее очевидно, что независимы события A и C, где C означает, что выпал только один герб (доказать !).

Сложнее определяется независимость более двух событий.

Определение 15 . *События $A_1, \ldots, A_n$ называем* **независимыми в совокупности**, *если для любого $2 \leqslant k \leqslant n$ и любых событий $A_{i_1}, \ldots, A_{i_k}$ из рассматриваемых справедливо*

$$P(A_{i_1} \ldots A_{i_k}) = P(A_{i_1}) \ldots P(A_{i_k}).$$

Покажем на примерах, что попарной независимости и выполнения последнего равенства для перечня всех событий недостаточно для независимости в совокупности.

Пример 4 . *Правильный тетраэдр окрашен тремя цветами: одна грань — в синий цвет, вторая — в красный, третья — в зеленый, а на четвертой присутствуют все три цвета. Этот тетраэдр подбрасывают и отмечают, какой гранью он выпал.*

Пусть A_1 означает появление синего цвета, A_2 — красного, A_3 — зеленого. Тогда $P(A_1) = 2/4 = 1/2$, $P(A_2) = 1/2$, $P(A_3) = 1/2$, $P(A_1A_2) = P(A_2A_3) = P(A_1A_3) = P(A_1A_2A_3) = 1/4$. Отсюда мы получаем, что $P(A_1A_2) = 1/4 = 1/2 \cdot 1/2 = P(A_1)P(A_2)$. Аналогично для других пар. Таким образом, мы имеем попарную независимость. Но $P(A_1A_2A_3) = 1/4 \neq 1/2 \cdot 1/2 \cdot 1/2 = P(A_1)P(A_2)P(A_3)$.

Задача 1 . *Придумать пример эксперимента и трех событий A_1, A_2, A_3, для которых $P(A_1A_2A_3) = P(A_1)P(A_2)P(A_3)$, но которые не являются попарно независимыми.*

Можно дать следующее более общее

Определение 16 . *Пусть $\mathcal{M}, \ldots, \mathcal{M}_n$ — некоторые классы событий. Они называются* **независимыми**, *если любые события $A_1 \in \mathcal{M}_1, \ldots, A_n \in \mathcal{M}_n$ — независимы в совокупности.*

Типичная ситуация описана в следующем примере.

Пример 5 . *Симметричный игральный кубик подбрасывают два раза. $\mathcal{M}_1$ обозначает набор событий, связанных с результатом первого бросания. $\mathcal{M}_2$ определяется аналогично для результата второго бросания. Тогда $\mathcal{M}_1$ и $\mathcal{M}_2$ — независимы.*

Во многих задачах является полезным следующий результат.

Предложение 3 . *Если события A и B независимы, то независимы и любые два из следующих: A и $\overline{B}$, $\overline{A}$ и B, $\overline{A}$ и $\overline{B}$.*

Доказательство. Докажем независимость A и $\overline{B}$.

$$\begin{aligned} P(A\overline{B}) &= P(A) - P(AB) = P(A) - P(A)P(B) \\ &= P(A)(1 - P(B)) = P(A)P(\overline{B}). \end{aligned}$$

Независимость остальных пар событий предлагается доказать самостоятельно.

4.3 Формулы полной вероятности и Байеса

Во многих ситуациях мы встречаемся с такими экспериментами, которые можно разложить на два (или более) этапов. На первом этапе мы имеем несколько вариантов, а спрашивается что-либо о том, что произошло в конце — на втором этапе. В этом случае чрезвычайно полезен приводимый ниже результат. Начнем со следующего определения.

Определение 17 . *События $H_1, \ldots, H_n$ образуют* **полную группу событий** *(***разбиение пространства** Ω*), если*

1. $H_iH_j = \emptyset, \ i \neq j,$

2. $H_1 + \ldots + H_n = \Omega.$

Теорема 1 . *Пусть события $H_1, \ldots, H_n$ образуют полную группу событий, $P(H_k) > 0$ для всех k и A — произвольное событие. Тогда*

$$P(A) = \sum_{k=1}^{n} P(H_k)P(A|H_k)$$ — **формула полной вероятности**.

Доказательство. Так как события $H_1, \ldots, H_n$ образуют полную группу, то мы имеем

$$(AH_i) \cap (AH_j) = AH_iH_j = \emptyset$$

и

$$A = A \cap \Omega = A(H_1 + \ldots + H_n) = AH_1 + \ldots + AH_n.$$

Отсюда получаем

$$P(A) = \sum_{k=1}^{n} P(AH_k) = \sum_{k=1}^{n} P(H_k)P(A|H_k),$$

где мы использовали теорему умножения.

Пример 6 . *На некоторой фабрике 30% продукции производится машиной A, 25% продукции — машиной B, а остальная продукция — машиной C. У машины A в брак идет 1% производимой ей продукции, у машины — 1,2% , а у машины C — 2%. Из всей произведенной продукции случайно выбрано одно изделие. Какова вероятность того, что оно бракованное?*

Пусть H_1 обозначает событие, состоящее в том, что выбранная деталь изготовлена на машине A, H_2 — на машине B, H_3 — на машине C. Обозначим через D событие, состоящее в том, что выбранная деталь бракованная. События H_1, H_2, H_3 образуют полную группу событий. По условию задачи

$$P(H_1) = 0.3,\ P(H_2) = 0.25,\ P(H_3) = 0.45,$$

$$P(D|H_1) = 0.01,\ P(D|H_2) = 0.012,\ P(D|H_3) = 0.02.$$

По формуле полной вероятности получаем

$$\begin{aligned} P(D) &= P(H_1)P(D|H_1) + P(H_2)P(D|H_2) + P(H_3)P(D|H_3) = \\ &= 0.3 \cdot 0.01 + 0.25 \cdot 0.012 + 0.45 \cdot 0.02 = 0.016. \end{aligned}$$

В статистических приложениях обычно приходится решать обратную задачу. Мы имеем несколько вариантов $H_1, \ldots, H_n$ условий проведения эксперимента. Для каждого из этих вариантов на основе прошлой информации нам известна вероятность $P(H_k)$ его реализации в данном испытании. В результате проведения эксперимента мы получили некоторую дополнительную информацию (произошло событие A). Теперь мы хотим оценить вероятность того, что реализовался вариант H_k. Это можно сделать с помощью следующей теоремы.

Теорема 2 . *Если $H_1, \ldots, H_n$ образуют полную группу событий, A — произвольное событие и $P(H_k) > 0,\ k = \overline{1,n},\ P(A) > 0$, то*

$$P(H_k|A) = \frac{P(H_k)P(A|H_k)}{P(A)} = \frac{P(H_k)P(A|H_k)}{\sum\limits_{j=1}^{n} P(H_j)P(A|H_j)}$$

— формула **Байеса**.

Доказательство. По определению условной вероятности и теореме умножения

$$P(H_k|A) = \frac{P(AH_k)}{P(A)} = \frac{P(H_k)P(A|H_k)}{P(A)}.$$

Для вычисления $P(A)$ используем формулу полной вероятности.

Вероятности $P(H_k)$ называются **априорными вероятностями гипотез** H_k, а $P(H_k|A)$ — **апостериорными вероятностями**.

Пример 7 . *В урне находятся три шара, которые могут быть либо черными, либо белыми, но конкретный состав урны не известен. С возвращением выбираем 5 шаров. Среди них оказалось 3 белых (событие A). Какой состав урны наиболее вероятен?*

Мы имеем четыре варианта состава урны:
H_0 — нет белых шаров,
H_1 — один белый шар,
H_2 — два белых шара,
H_3 — три белых шара.

Мы предполагаем, что априорные вероятности $P(H_k) = 1/4,\ k = 0, 1, 2, 3$. Выбор производится с возвращением. Используя результат примера 3.1, получаем
$P(A|H_0) = 0$,
$P(A|H_1) = C_5^3(1/3)^3(2/3)^2 = 40/243$,
$P(A|H_2) = C_5^3(2/3)^3(1/3)^2 = 80/243$,
$P(A|H_3) = 0$.

По формуле полной вероятности

$$P(A) = 1/4 \cdot 0 + 1/4 \cdot 40/243 + 1/4 \cdot 80/243 + 1/4 \cdot 0 = 30/243.$$

Окончательно по формуле Байеса получаем

$$P(H_k|A) = \frac{P(H_k)P(A|H_k)}{P(A)},$$

$$P(H_0|A) = 0,\ P(H_1|A) = 1/3,\ P(H_2|A) = 2/3,\ P(H_3|A) = 0.$$

Таким образом, наиболее вероятно, что в урне было два белых шара.

Глава 5

Последовательности испытаний

5.1 Определение последовательности испытаний

До сих пор мы в основном рассматривали случайные эксперименты, состоящие как бы из одного этапа. Хотя встречались примеры и «многоступенчатых» экспериментов. В этом разделе мы дадим описание математической модели эксперимента, который состоит из нескольких шагов или этапов.

Пусть $X_1, \ldots, X_n$ — некоторые конечные множества. Элементы множества X_k — это возможные результаты эксперимента на k–м шаге. Тогда в качестве элементарного исхода такого эксперимента, состоящего из n шагов, естественно взять $\omega = (\omega_1, \ldots, \omega_n)$, где $\omega_k \in X_k,\ k = \overline{1,n}$. В этом случае $\Omega = X_1 \times \ldots \times X_n$. Чтобы полностью задать вероятностью модель, необходимо задать вероятности

$$P(\omega) = P(\omega_1, \ldots, \omega_n)$$

элементарных исходов.

Определение 18 . *Вероятностное пространство* (Ω, P), *где* $\Omega = X_1 \times \ldots \times X_n$, *называется* **последовательностью испытаний** *с множеством значений* X_k *на* k*-м шаге* $(k = \overline{1,n})$.

В этом определении вероятность задана сразу для всей последовательности $\omega = (\omega_1, \ldots, \omega_n)$. Но в реальных задачах результаты эксперимента появляются последовательно шаг за шагом. Поэтому и вероятностную структуру естественно задавать с помощью вероятностей появления ω_k, если мы знаем, что на предыдущих шагах уже появились $\omega_1, \dots, \omega_{k-1}$. Используя теорему умножения, мы можем записать

$$\begin{aligned} P(\omega_1, \dots, \omega_n) &= P(\omega_1) \cdot P(\omega_2|\omega_1) \cdot \ldots \cdot P(\omega_n|\omega_1, \dots, \omega_{n-1}) \\ &= P^{(1)}(\omega_1) \cdot P^{(2)}(\omega_2|\omega_1) \cdot \ldots \cdot P^{(n)}(\omega_n|\omega_1, \dots, \omega_{n-1}) \ . \end{aligned}$$

$P^{(1)}$ называется **начальным распределением**, а $P^{(2)}, \dots, P^{(n)}$ — **вероятностями переходов** на соответствующих шагах для последовательности из n испытаний.

Пример (модель Эренфестов). Пусть мы имеем две урны. В начальный момент в первой урне два белых шара, во второй — два черных. Сначала мы выбираем случайно один шар из первой урны, отмечаем его цвет и кладем его во вторую урну. Затем выбираем случайно один шар из второй урны, отмечаем его цвет и кладем в первую урну и т.д. Пусть проводим n испытаний. На каждом шаге $X_k = \{,\}$. На первом шаге с вероятностью 1 выбирают белый шар. После $(k-1)$ шагов мы знаем состав очередной урны и можем найти вероятность появления белого или черного шара, используя классическое определение вероятностей.

Если в этом примере отмечать не цвет появившегося шара, а вновь полученный состав урн (что эквивалентно), то вероятности перехода $P^{(k)}$ будут зависеть от результата эксперимента только на последнем шаге, а не на всех предыдущих шагах. Это свойство имеет место для многих реальных задач. Поэтому целесообразно выделить этот случай отдельно.

Определение 19 . *Последовательность испытаний называется* **цепью Маркова**, *если для всех* $k = \overline{2, n}$ *и всех* $\omega_1, \dots, \omega_k$

$$P^{(k)}(\omega_k|\omega_1, \dots, \omega_{k-1}) = P^{(k)}(\omega_k|\omega_{k-1}) \ . \qquad (5.1)$$

Еще более простую ситуацию мы получаем, если $P^{(k)}$ вообще не зависят от условий $\omega_1, \dots, \omega_{k-1}$.

Определение 20 . *Говорят, что мы имеем* **последовательность n независимых испытаний**, *если* $\forall \omega = (\omega_1, \ldots, \omega_n)$

$$P(\omega) = P^{(1)}(\omega_1) \cdot \ldots \cdot P^{(n)}(\omega_n) \ . \tag{5.2}$$

Нетрудно проверить, что в этом случае для любого $k = \overline{1, n}$ имеет место

1) $P^{(k)}(\omega_k) \geqslant 0, \ \omega_k \in X_k$,

2) $\sum\limits_{\omega_k \in X_k} P^{(k)}(\omega_k) = 1$.

Определение 21 . *Вероятностное пространство* (Ω, P) *называется* **последовательностью n независимых и одинаковых испытаний**, *если это последовательность независимых испытаний,* $X_1 = \ldots = X_n = X$ *и* $P^{(1)} = \ldots = P^{(n)}$.

Можно показать, что события, связанные с разными испытаниями, являются независимыми.

Простейшим является случай, когда $X = \{0, 1\}$, т.е. множество исходов на каждом шаге состоит из двух элементов. Модель такого экперимента называется **схемой Бернулли** и изучается подробно в следующем разделе.

5.2 Схема Бернулли. Биномиальное распределение

Начнем с неформального определения. **Схемой Бернулли или последовательностью n независимых одинаковых испытаний с двумя исходами** называется случайный эксперимент, в котором:

1) проводится n независимых испытаний;

2) каждое испытание кончается одним из двух исходов (один исход называется «успех» и обозначается 1, а второй — «неуспех» и обозначается 0);

3) вероятность появления «успеха» одна и та же в каждом испытании и равна p.

Числа n и p называются **параметрами** схемы Бернулли. Формальное описание модели такого эксперимента дано в следующем определении.

Определение 22 . **Схемой Бернулли** *с параметрами n и p называется дискретное вероятностное пространство (Ω, P), где Ω состоит из элементарных исходов вида $\omega = (\omega_1, \ldots, \omega_n)$, $\omega_k = 0, 1$, $k = \overline{1, n}$, а вероятности элементарных исходов ω задаются по правилу*

$$P(\omega) = p^m (1-p)^{n-m} \ , \tag{5.3}$$

где m — число единиц в исходе ω.

Пример 8 . *Симметричную монету подбрасывают 5 раз и на каждом шаге отмечают какой стороной выпала монета.*

Выпадение герба будем считать «успехом». Это схема Бернулли с параметрами $n = 5$ и $p = 1/2$.

Обычно в рамках схемы Бернулли мы хотим вычислить вероятность не отдельного элементарного исхода, а некоторого более сложного события. Например, в предыдущем примере нас может интересовать вероятность того, что выпало ровно 3 герба. Такой вопрос является наиболее типичным для схемы Бернулли. Пусть A_m есть событие, состоящее в том, что в n независимых испытаниях мы получили ровно m успехов. По определению вероятности события

$$P(A_m) = \sum_{\omega \in A_m} P(\omega) = C_n^m p^m (1-p)^{n-m} \ ,$$

где мы воспользовались тем, что вероятности элементарных исходов в A_m все одинаковые и вычисляются по формуле (3), а число различных исходов равно размещению m единиц по n мест без учета порядка.

Если нас интересует число успехов, а не когда именно они появились, то мы можем построить менее подробную модель.

Определение 23 . **Биномиальной моделью** *с параметрами n и p называется вероятностное пространство (Ω, P), где $\Omega = \{\omega_0, \ldots, \omega_n\}$ и*

$$P(\omega_m) = b(n,p,m) = C_n^m p^m (1-p)^{n-m} \ , \tag{5.4}$$

$m = \overline{0,n}$, $0 < p < 1$.

Вероятности $b(n,p,m)$, вычисляемые по формуле (4), образуют так называемое **биномиальное распределение**. Нетрудно проверить, что они обладают обычными свойствами вероятностей:

1) $b(n,p,m) \geqslant 0$,

2) $\sum\limits_{m=0}^{n} b(n,p,m) = 1$,

и, кроме того,

3) $b(n,p,m) = b(n,1-p,n-m)$.

Часто в прикладных задачах нас интересует, какое число успехов m_0 имеет наибольшую вероятность. Для этого сравним вероятность двух соседних значений.

$$\frac{b(n,p,m+1)}{b(n,p,m)} = \frac{C_n^{m+1} p^{m+1} (1-p)^{n-m-1}}{C_n^m p^m (1-p)^{n-m}} = \frac{n-m}{m+1} \frac{p}{1-p} \ .$$

Последнее выражение больше 1 при $m < (n+1)p + 1$ и меньше 1 при $m > (n+1)p - 1$. Это приводит нас к следующим свойствам биномиальных вероятностей:

4) если $m < (n+1)p - 1$, то при переходе от m к $m+1$ вероятность возрастает,

5) если $m > (n+1)p - 1$, то при переходе от m к $m+1$ вероятность убывает,

6) если число $(n+1)p$ — целое, то имеем два наиболее вероятных значения для числа успехов: $m_0^{(1)} = (n+1)p - 1$ и $m_0^{(2)} = p(n+1)$,

если число $(n+1)p$ — дробное, то имеем одно наиболее вероятное значение — $m_0 = \big[(n+1)p\big]$, здесь $[x]$ означает целую часть числа x.

Полезно нарисовать графики изменения вероятностей $b(n,p,m)$ с изменением m для разных n и p.

Пример 9 . *Симметричную игральную кость подбрасывают 6 раз. Найти вероятность того, что выпадут ровно два герба и наиболее вероятное число появлений шестерки.*

В этой задаче $n = 6,\ p = 1/6$. Тогда

$$b(6, 1/2, 2) = C_6^2 0.5^2 0.5^4 = 0.235.$$

Число $(n+1)p = 7/6$ является дробным, поэтому наиболее вероятное число появлений шестерки есть $m_0 = \big[(n+1)p\big] = 1$.

5.3 Предельные теоремы в схеме Бернулли

Выше мы получили формулу, по которой можно рассчитать вероятность того, что в серии из n испытаний Бернулли с вероятностью успеха p будет получено ровно m успехов. А именно

$$b(n,p,m) = C_n^m p^m (1-p)^{n-m}\ .$$

В реальных задачах число испытаний бывает достаточно большим, и производить расчеты по этой формуле становится затруднительным. В этих случаях обычно стараются найти более простые выражения, которые асимптотически эквивалентны точным формулам, когда те или иные параметры меняются определенным образом. Для нашей модели существуют две аппроксимации, которые находят широкие приложения в практических задачах и, как мы увидим позднее, имеют и самостоятельное значение.

Теорема Пуассона. *Пусть мы имеем схему Бернулли с параметрами n и p. Пусть $n \to \infty,\ p \to 0$ так, что $np \to \lambda,\ 0 < \lambda < \infty$. Тогда для любого фиксированного m*

$$b(n,p,m) \to \pi_m(\lambda) = \frac{\lambda^m}{m!} e^{-\lambda}\ .$$

Доказательство. Зафиксируем некоторое целое неотрицательное m. Тогда

$$b(n,p,m) = C_n^m p^m (1-p)^{n-m} = \frac{n!}{m!(n-m)!} \cdot p^m (1-p)^{n-m} =$$

$$\frac{1}{m!} \cdot n(n-1) \cdot \ldots \cdot \big[n-(m-1)\big] \cdot p^m (1-p)^{-m} (1-p)^n =$$

$$= \frac{1}{m!} \cdot n(n-1) \cdot \ldots \cdot \big[n-(m-1)\big] \cdot \left[\frac{\lambda}{n} + o\left(\frac{1}{n}\right)\right]^m \times$$

$$\times \left[1 - \frac{\lambda}{n} + o\left(\frac{1}{n}\right)\right]^{-m} \cdot \left[1 - \frac{\lambda}{n} + o\left(\frac{1}{n}\right)\right]^n$$

$$\to \frac{1}{m!} \cdot \lambda^m \cdot e^{-\lambda} .$$

Более аккуратный анализ позволяет доказать, что

$$\sum_{m=0}^{\infty} |b(n,p,m) - \pi_m(\lambda)| \leqslant np^2,$$

если $np = \lambda$ и $b(n,p,m) = 0$ для $m \geqslant n$.

Пример 10 . *На некоторой телефонной станции 10 000 номеров. В день через станцию поступает в среднем 30 000 вызовов. Найти вероятность того, что по некоторому конкретному номеру будет ровно два звонка.*

Предположим, что вызов по любому номеру является равновероятным и при каждом вызове номер выбирается независимо от других вызовов. Тогда мы имеем схему Бернулли с параметрами $p = 10^{-4},\ n = 3 \cdot 10^4$. Используя теорему Пуассона, получаем ($\lambda = 3$)

$$b(n,p,2) = C_{30000}^2 p^2 (1-p)^{29998} \approx \frac{3^2}{2!} e^{-3} = 0.1804 .$$

Для распределения Пуассона составлены таблицы.

Другой асимптотический результат получается, если $n \to \infty$, а m выбрано так, что $m \to \infty$, но величина $x_{n,m} = \frac{m-np}{\sqrt{np(1-p)}}$ лежит в некотором фиксированном интервале (a, b), где $-\infty < a < b < \infty$.

Локальная теорема Муавра–Лапласа. *Пусть мы имеем схему Бернулли с параметрами n и p. Тогда при $n \to \infty$ и фиксированном p*

$$b(n, p, m) \sim \frac{1}{\sqrt{np(1-p)}}\varphi(x_{n,m})$$

равномерно по всем m, для которых $-\infty < a < x_{n,m} < b < \infty$. Здесь

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

Для функции $\varphi(x)$ составлены таблицы. Отметим, что $\varphi(x) = \varphi(-x)$.

Пример 11 . *Симметричную монету подбрасывают 100 раз. Найти вероятность того, что герб выпадет ровно 50 раз.*

В этой задаче $n = 100,\ p = 0.5,\ m = 50$. Тогда

$$x_{n,m} = \frac{50 - 100 \cdot 0.5}{\sqrt{100 \cdot 0.5 \cdot 0.5}} = 0.$$

Используя локальную теорему Муавра–Лапласа, получаем

$$b(100, 0.5, 50) \approx \varphi(0) \cdot \frac{1}{\sqrt{100 \cdot 0.5 \cdot 0.5}} = \frac{0.3989}{5} \approx 0.08.$$

Как видно из последнего примера, при больших n и фиксированном p вероятности $b(n, p, m)$ для всех значений m очень малы. Поэтому обычно интересуются не тем, какое конкретное число успехов будет в нашем эксперименте, а в каких пределах оно окажется. Например, мы может вычислить вероятность того, что число появлений герба при 100 подбрасываниях будет лежать в пределах от 40 до 60. В таких задачах оказывается полезной следующая

Интегральная теорема Муавра–Лапласа. *Пусть мы имеем схему Бернулли с параметрами n и p. Если $n \to \infty$, а p — фиксировано, то равномерно по всем $m_1 < m_2$*

$$P(m_1 \leqslant S_n < m_2) \sim \Phi(x_{n,m_2}) - \Phi(x_{n,m_1}),$$

где S_n — число успехов в n испытаниях, а

$$\Phi(x) = \int\limits_{-\infty}^{x} \varphi(y)\, dy \ .$$

Более того, для любых $m_1 < m_2$ имеет место оценка

$$|P(m_1 \leqslant S_n < m_2) - \Phi(x_{n,m_2}) + \Phi(x_{n,m_1})| \leqslant \frac{p^2 + (1-p)^2}{\sqrt{np(1-p)}}.$$

Для функции $\Phi(x)$, называемой **функцией распределения стандартного нормального закона**, составлены таблицы. Она обладает свойствами:

1) $\Phi(-\infty) = 0,\ \Phi(\infty) = 1,\ \Phi(0) = 0.5,$

2) $\Phi(-x) = 1 - \Phi(x).$

В силу свойства 2 таблицы обычно составляют только для положительных или только для отрицательных x. Так как $\Phi(3.9) = 0.999$, то для $x \geqslant 4$ можно считать, что $\Phi(x) \approx 1$. Очень часто вместо $\Phi(x)$ используют функцию

$$\Phi_0(x) = \int\limits_{0}^{x} \varphi(y) dy \ ,$$

которая называется **функцией Лапласа** или **интегралом вероятностей**. Она обладает следующими свойствами:

1) $\Phi_0(0) = 0,\ \Phi_0(\infty) = 0.5,$

2) $\Phi_0(-x) = -\Phi_0(x),$

3) $\Phi(x) = \Phi_0(x) + 0.5.$

Для нее также составлены таблицы (для $x > 0$). В силу свойства 3 ее можно использовать в интегральной теореме Муавра — Лапласа вместо функции $\Phi(x)$.

Пример 12 . *Симметричную монету подбрасывают 100 раз. Найти вероятность того, что число появившихся гербов будет лежать в пределах от 40 до 60.*

В этой задаче $n = 100$, $p = 0.5$, $m_1 = 40$, $m_2 = 60$. В силу интегральной предельной теоремы Муавра-Лапласа

$$P(m_1 \leqslant S_n < m_2) \approx \Phi_0(x_{n,m_2}) - \Phi_0(x_{n,m_1}).$$

$$x_{n,m_2} = \frac{60 - 100 \cdot 0.5}{\sqrt{100 \cdot 0.5 \cdot 0.5}} = \frac{10}{5} = 2,$$

$$x_{n,m_1} = \frac{40 - 100 \cdot 0.5}{\sqrt{100 \cdot 0.5 \cdot 0.5}} = -\frac{10}{5} = -2,$$

$$P(40 \leqslant S_n < 60) \approx \Phi_0(2) - \Phi_0(-2) = 2\Phi_0(2) = 2 \cdot 0.4772 = 0.9544.$$

Применяя эту аппроксимацию, мы допускаем ошибку, которая не превышает величины

$$\frac{0.5^2 + 0.5^2}{\sqrt{100 \cdot 0.5 \cdot 0.5}} = \frac{0.5}{5} = 0.1.$$

Более точный анализ показывает, что эта ошибка гораздо меньше.

Интегральная предельная теорема Муавра–Лапласа является частным случаем более общего результата, называемого **центральной предельной теоремой**, доказательство которого будет приведено позднее.

5.4 Полиномиальное распределение

Последняя модель, а именно биномиальное распределение, имеет очевидное обобщение на случай, когда число исходов в каждом испытании одинаково, но, возможно, отлично от двух.

Пусть мы имеем последовательность независимых испытаний, каждое из которых кончается одним из r исходов, вероятности которых равны соответственно $p_1, \dots, p_r$, и они одни и те же во всех испытаниях. Пусть, далее, m_k есть число появлений k-го исхода в этих n

испытаниях. Тогда вектор $m = (m_1, \ldots, m_r)$ дает нам описание того, чем закончился такой эксперимент. Используя те же рассуждения, что и в биномиальной модели, нетрудно доказать, что вероятность такого исхода вычисляется по формуле

$$P(m) = P(m_1, \ldots, m_r) = \frac{n!}{m_1! \ldots m_r!} p_1^{m_1} \ldots p_r^{m_r} \,. \tag{5.5}$$

Суммируя все вышеизложенное, приходим к определению.

Определение 24 . *Вероятностное пространство* (Ω, P), *в котором элементарные исходы* ω *имеют вид* $\omega = (m_1, \ldots, m_r)$, *где* $(m_1, \ldots, m_r)$ — *целые неотрицательные числа такие, что* $m_1 + \ldots + m_r = n$, *а их вероятности вычисляются по формуле*

$$P(\omega) = P(m_1, \ldots, m_r) = \frac{n!}{m_1! \ldots m_r!} p_1^{m_1} \ldots p_r^{m_r},$$

называется **полиномиальным распределением**.

Пример 13 . *Симметричный игральный кубик подбрасывают 10 раз. Найти вероятность события* A, *состоящего в том, что выпадут 2 шестерки и одна пятерка.*

В этом эксперименте проводится 10 испытаний, в которых естественно фиксировать три различных исхода: выпали шестерка, пятерка и другая цифра, вероятности которых равны 1/6, 1/6 и 4/6 соответственно. Тогда

$$P(A) = \frac{10!}{2!1!7!} \left(\frac{1}{6}\right)^2 \left(\frac{1}{6}\right)^1 \left(\frac{4}{6}\right)^7 = \frac{5}{3} \left(\frac{2}{3}\right)^7 \approx 0.098.$$

При больших n расчеты по формуле (5) становятся затруднительными. В этом случае применяются асимптотические формулы, аналогичные тем, что мы рассматривали для биномиального распределения.

Задача 2 . *Пусть мы провели* n *испытаний с тремя исходами, вероятности которых равны* p_1, p_2 *и* p_3 *соответственно. Предположим, что* $n \to \infty$, $p_1 \to 0$, $p_2 \to 0$ *таким образом, что* $np_1 \to$

$\lambda_1,\ np_2 \to \lambda_2$. *Тогда при фиксированных* m_1 *и* m_2

$$P(m_1, m_2, n - m_1 - m_2) \to \frac{\lambda^{m_1}}{m_1!} e^{-\lambda_1} \frac{\lambda^{m_2}}{m_2!} e^{-\lambda_2}\ .$$

Глава 6

Определение вероятностного пространства в общем случае

6.1 Постановка задачи

До сих пор мы работали только с простейшей моделью случайного эксперимента, а именно с дискретным вероятностным пространством. Но многие реальные задачи невозможно описать в рамках этой модели, так как в них число элементарных исходов несчетно. Одним из таких примеров была задача на геометрическую вероятность, в которой рассматривается модель случайного выбора точки из некоторой области. Необходимо дать общее определение вероятностного пространства, которое охватывало бы и такие ситуации. Мы начнем с того, что напомним определение вероятностного пространства в слабом смысле.

Определение 25 . Вероятностным пространством (в слабом смысле) *называется тройка* $(\Omega, \mathcal{A}, P)$*, где* Ω *— произвольное множество,* $\mathcal{A}$ *— алгебра подмножеств пространства* Ω*,* P *— вещественная функция, заданная на* $\mathcal{A}$ *и обладающая свойствами:*

1) $P(A) \geqslant 0, \ \forall A \in \mathcal{A}$*,*

2) $P(\Omega) = 1$*,*

3) если $A_1, \ldots, A_n \in \mathcal{A}$ *— попарно несовместны, то*

$$P\Big(\sum_{k=1}^{n} A_k\Big) = \sum_{k=1}^{n} P(A_k) \ .$$

Мы рассматриваем Ω как пространство элементарных исходов некоторого эксперимента, $\mathcal{A}$ — это алгебра интересующих нас событий, $P(A)$ — вероятность события. Чтобы понять, почему такого определения недостаточно для наших целей, рассмотрим простой пример.

Пример 14 . *Из единичного квадрата случайным образом выбирают точку. Событие K состоит в том, что выбранная точка принадлежит кругу с центром $(1/2, 1/2)$ радиуса 1/4.*

Это типичная задача на геометрическое определение вероятностей. Здесь $\omega = (x_1, x_2)$, $x_1, x_2 \in [0,1]$, $\Omega = [0,1] \times [0,1] = [0,1]^2$. Интересующее нас событие K естественно отождествить с указанным кругом. Тогда по геометрическому определению вероятности

$$P(K) = \frac{S(K)}{S(\Omega)} \ ,$$

где $S(K)$ — площадь круга. Но круг — это довольно сложная фигура, и непросто определить, что такое его площадь. Из школьного курса геометрии мы знаем, как вычислить площадь прямоугольника. Затем мы исчерпываем круг такими прямоугольниками и определяем его площадь как сумму площадей этих прямоугольников. Но таких прямоугольников будет бесконечное число.

Приведенный выше анализ показывает, что для таких событий, как попадание случайной точки в круг K, необходимо применять бесконечное число операций над более простыми событиями (попадание в прямоугольник), а для определения его вероятности нужно более сильное свойство, чем конечная аддитивность, так как мы складываем площади бесконечного числа прямоугольников.

6.2 Определение σ–алгебры

Определение 26 . *Система $\mathcal{A}$ подмножеств пространства Ω называется σ–**алгеброй**, если выполнены следующие свойства:*

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

2) если $A \in \mathcal{A}$, то $\bar{A} \in \mathcal{A}$, *3) если* $A_1, A_2, \ldots \in \mathcal{A}$, *то* $\bigcup\limits_{k=1}^{\infty} A_k \in \mathcal{A}$.

Замечание. Можно показать, что применение любых ранее определенных операций над событиями, выполненными не более чем в счетном числе, не выводит нас за пределы σ–алгебры.

Примеры. 1. $\mathcal{A} = \{\emptyset, \Omega\}$ — тривиальная σ-алгебра, соответствует случаю, когда мы ничего не знаем о рассматриваемом эксперименте.

2. $\mathcal{A}$ — σ–алгебра всех подмножеств пространства Ω соответствует полной информации об эксперименте.

3. Любая конечная алгебра $\mathcal{A}$ (т. е. содержащая конечное число событий) является σ–алгеброй (задача!).

Обычно в реальной задаче мы начинаем с некоторого класса событий $\mathcal{M}$, который, возможно, не является σ–алгеброй. Например, в рассмотренном выше примере это был класс прямоугольников. Хотелось бы дополнить его так, чтобы получить σ–алгебру.

Определение 27 . *Пусть* $\mathcal{M}$ *— некоторая система подмножеств пространства* Ω. *Класс* $\mathcal{A} = \sigma(\mathcal{M})$ *подмножеств пространства* Ω *называется* **σ–алгеброй, порожденной системой** $\mathcal{M}$, *если выполнены следующие свойства:*

1) $\mathcal{M} \subset \mathcal{A}$,

2) $\mathcal{A}$ *—* σ*–алгебра* ,

3) если $\mathcal{A}_1$ *— некоторая* σ*–алгебра, содержащая* $\mathcal{M}$, *то* $\mathcal{A} \subset \mathcal{A}_1$.

Замечания. 1. Свойство 3 означает, что $\sigma(\mathcal{M})$ в определенном смысле самая маленькая σ–алгебра, содержащая систему $\mathcal{M}$.

2. Порожденная σ–алгебра всегда существует. Для доказательства нужно рассмотреть семейство всех σ–алгебр, содержащих $\mathcal{M}$ (такие существуют), образовать их пересечение и доказать,что это σ-алгебра (провести подробное доказательство самостоятельно!).

Пример. Пусть 1 раз подбрасывается симметричный кубик, на гранях которого проставлены цифры 1, 2, 3, 4, 5, 6. Результаты этого опыта фиксируются следующим образом: исходам 1 и 2 в соответствие ставим ноль, исходам 3 и 4 — единицу, а исходам 5 и 6 — двойку.

Класс множеств $\mathcal{M} = \Big\{\{1,2\},\{3,4\},\{5,6\}\Big\}$ не является σ–алгеброй. Порожденной этим классом σ–алгеброй будет

$$\sigma(\mathcal{M}) = \Big\{\Omega, \emptyset, \{1,2\}, \{3,4\}, \{5,6\}, \{1,2,3,4\}, \{1,2,5,6\}, \{3,4,5,6\}\Big\}.$$

Это действительно самая маленькая система множеств, удовлетворяющая определению σ–алгебры и содержащая в себе исходный класс множеств $\mathcal{M}$.

Задача. Построить с σ–алгебру, соответствующую случаю, когда исходы 5 и 6 различаются и сравните с $\sigma(\mathcal{M})$.

Приведём примеры σ–алгебр, которые нам понадобятся в дальнейшем при работе со случайными величинами и случайными векторами.

Пример. *Пусть* $\Omega = R^1$, $\mathcal{M}$ *— класс всех интервалов.* $\mathcal{A} = \sigma(\mathcal{M})$ *называется* **борелевской** *σ–алгеброй, а элементы A из $\mathcal{A}$ называются* **борелевскими** *множествами.*

Если $\Omega = R^n$, а $\mathcal{M}$ — класс всех открытых шаров в R^n, то $\mathcal{A} = \sigma(\mathcal{M})$ — **борелевская** *σ–алгебра подмножеств в R^n. Элементы этой σ–алгебры называются* **борелевскими подмножествами** *в R^n.*

Замечание. В борелевской σ–алгебре содержатся все привычные нам множества на прямой, включая одноточечные множества вида $\{x\}$, где $x \in R^1$ и их счётные объединения; интервалы вида (a,b), $[a,b)$, $(a,b]$ или $[a,b]$ и их счётные объединения; множества N — натуральных чисел, Q — рациональных чисел, R^1 — вещественных чисел и т.д. Для определения множества, не являющегося борелевским, требуются некоторые специальные построения.

Если у нас есть вероятностное пространство в слабом смысле, то мы можем расширить алгебру событий до порожденной σ–алгебры. Теперь необходимо продолжить вероятность P на эту более широкую систему. Это можно сделать только в том случае, когда P обладает определенными свойствами непрерывности.

6.3 Свойства непрерывности вероятности

Лемма 2 . *Если* $(\Omega, \mathcal{A}, P)$ *— вероятностное пространство в слабом смысле, то следующие свойства вероятности* P *эквивалентны:*

1) если $A_1, A_2, \ldots \in \mathcal{A}$ *— попарно несовместны и* $A = \bigcup\limits_{k=1}^{\infty} A_k \in \mathcal{A}$*, то*

$$P\Big(\sum_{k=1}^{\infty} A_k\Big) = \sum_{k=1}^{\infty} P(A_k)$$

— **счетная аддитивность вероятности***;*

2) если $A_1, A_2, \ldots \in \mathcal{A}$*,* $A_1 \subset A_2 \subset \ldots$ *и* $A = \bigcup\limits_{k=1}^{\infty} A_k \in \mathcal{A}$*, то*

$$P(A_k) \nearrow P(A)$$

— **непрерывность снизу***;*

3) если $A_1, A_2, \ldots \in \mathcal{A}$*,* $A_1 \supset A_2 \supset \ldots$ *и* $A = \bigcap\limits_{k=1}^{\infty} A_k \in \mathcal{A}$*, то*

$$P(A_k) \searrow P(A)$$

— **непрерывность сверху***,*

4) если $A_1, A_2, \ldots \in \mathcal{A}$*,* $A_1 \supset A_2 \supset \ldots$ *и* $A = \bigcap\limits_{k=1}^{\infty} A_k = \emptyset$*, то*

$$P(A_k) \searrow 0$$

— **непрерывность сверху на** $\emptyset$*.*

Доказательство. 1. $\boxed{1} \Rightarrow \boxed{2}$.

Пусть $A_1, A_2, \ldots \in \mathcal{A}$, $A_1 \subset A_2 \subset \ldots$ и $A = \bigcup\limits_{k=1}^{\infty} A_k \in \mathcal{A}$. Определим $B_1 = A_1,\ B_2 = A_2 \backslash A_1,\ \ldots\ ,\ B_k = A_k \backslash A_{k-1},\ \ldots$. События $B_1, B_2, \ \ldots$ — попарно несовместны, $A_k = B_1 + \ldots + B_k$ и $A = \sum\limits_{k=1}^{\infty} B_k$. В силу свойства 1

$$P(A_k) = \sum_{k=1}^{n} P(B_k) \nearrow \sum_{k=1}^{\infty} P(B_k) = P(A)\ .$$

2. $\boxed{2} \Rightarrow \boxed{3}$.

Пусть $A_1, A_2, \ldots \in \mathcal{A}$, $A_1 \supset A_2 \supset \ldots$ и $A = \bigcap\limits_{k=1}^{\infty} A_k \in \mathcal{A}$. Определим $B_k = \overline{A_k} \in \mathcal{A}$. Новая последовательность событий обладает свойствами: $B_1 \subset B_2 \subset \ldots$ и $B = \bigcup\limits_{k=1}^{\infty} B_k = \bigcup\limits_{k=1}^{\infty} \overline{A_k} = \overline{\bigcap\limits_{k=1}^{\infty} A_k} = \overline{A}$. Исходя из свойства 2 и используя свойства вероятностей событий, получаем

$$1 - P(A_k) = P(\overline{A_k}) = P(B_k) \nearrow P(B) = P(\overline{A}) = 1 - P(A)\ .$$

Отсюда следует, что $P(A_k) \searrow P(A)$.

3. $\boxed{3} \Rightarrow \boxed{4}$. Эта импликация тривиальна, так как свойство 4 есть частный случай свойства 3.

4. $\boxed{4} \Rightarrow \boxed{1}$.

Пусть $A_1, A_2, \ldots \in \mathcal{A}$ — попарно несовместны и $A = \sum\limits_{k=1}^{\infty} A_k \in \mathcal{A}$. Обозначим $B_n = A_1 + \ldots + A_n,\ C_n = \sum\limits_{k=n+1}^{\infty} A_k$. Последовательность событий C_n обладает свойствами: $C_1 \supset C_2 \supset \ldots$ и $\bigcap\limits_{n=1}^{\infty} C_n = \emptyset$. Кроме того, $A = B_n + C_n$. Используя конечную аддитивность вероятности и свойство 4, получаем $P(C_n) \searrow 0$ и, так как $B_n \cap C_n = \emptyset$,

$$P(A) = P(B_n) + P(C_n) = \sum_{k=1}^{n} P(A_k) + P(C_n) = \sum_{k=1}^{\infty} P(A_k) + 0\ .$$

Как мы отмечали выше, нам необходимо продолжить вероятность, заданную на некоторой алгебре событий $\mathcal{A}_0$, на σ–алгебру $\mathcal{A} = \sigma(\mathcal{A}_0)$, порожденную этой алгеброй. Немецкий математик Каратеодори доказал, что если вероятность P обладает свойством счетной аддитивности, то ее можно продолжить с алгебры $\mathcal{A}_0$ на σ–алгебру $\mathcal{A} = \sigma(\mathcal{A}_0)$, причем единственным образом.

Теорема о продолжении вероятности (Каратеодори). Пусть $(\Omega, \mathcal{A}_0, P)$ — вероятностное пространство в слабом смысле и вероятность P обладает свойством счетной аддитивности. Тогда существует единственная счетно–аддитивная вероятность Q на $\mathcal{A} = \sigma(\mathcal{A}_0)$:

$$P(A) = Q(A) \ \ \forall A \in \mathcal{A}_0 \ .$$

Приведенные выше рассуждения и примеры приводят нас к следующему определению.

Определение 28 . **Вероятностным пространством** *называется тройка* $(\Omega, \mathcal{A}, P)$*, где* Ω *— произвольное множество,* $\mathcal{A}$ *— некоторая* σ*–алгебра его подмножеств,* P *— вещественная функция на* $\mathcal{A}$*:*

1) $P(A) \geqslant 0, \ \forall A \in \mathcal{A}$*,*

2) $P(\Omega) = 1$*,*

3) если $A_1, A_2, \dots$ *— попарно несовместны, то*

$$P\Big(\sum_{k=1}^{\infty} A_k\Big) = \sum_{k=1}^{\infty} P(A_k).$$

Такое определение впервые было предложено А.Н. Колмогоровым в книге «Основные понятия теории вероятностей», опубликованной в 1933 г. на немецком языке (русский перевод — 1936 г.).

Замечание. Отметим, что все свойства операций над событиями и свойства вероятностей, доказанные ранее, остаются справедливыми.

Примеры. 1. **Дискретное вероятностное пространство**.

Пусть мы имеем пару (Ω, P), где Ω — конечное или счетное множество, P — вещественная функция на Ω:

1) $P(\omega) \geqslant 0, \ \forall \omega \in \Omega$,

2) $\sum\limits_{\omega \in \Omega} P(\omega) = 1$.

Пусть $\mathcal{A}$ — σ–алгебра всех подмножеств пространства Ω, а P задана на $\mathcal{A}$ по правилу

$$P(A) := \sum_{\omega \in A} P(\omega).$$

Задача. *Доказать, что заданная таким образом вероятность* P *на* $\mathcal{A}$ *обладает свойством счетной аддитивности.*

Таким образом, $(\Omega, \mathcal{A}, P)$ удовлетворяет определению 4.

2. Пусть $\Omega \subset R^n$ — ограниченное борелевское множество, $f(x)$ — неотрицательная вещественная функция, заданная на Ω:

$$\int\limits_{\Omega} f(x)dx = 1.$$

Пусть, далее, $\mathcal{A}$ — σ–алгебра борелевских подмножеств пространства Ω, а вероятность P задается на $\mathcal{A}$ по правилу:

$$P(A) = \int\limits_{A} f(x)dx.$$

Тройка $(\Omega, \mathcal{A}, P)$ определяет вероятностное пространство.

Если $f(x) \equiv C = const$, то приходим вновь к геометрическому определению вероятности.

Глава 7

Случайная величина и ее распределение

7.1 Определение случайной величины и ее распределения

Во многих практических задачах мы имеем дело с такими экспериментами, в которых мы изучаем некоторые числовые характеристики. Приведем несколько примеров из тех, что встречались нам ранее.

Примеры. 1. Симметричную монету подбрасываем три раза и отмечаем число выпавших гербов.

2. Симметричную кость подбрасываем два раза и отмечаем сумму выпавших очков.

3. Пусть мы имеем схему Бернулли с n испытаниями и подсчитываем число успехов.

Во всех этих примерах мы видим, что в результате эксперимента мы получаем некоторое число, которое однозначно определяется элементарным исходом. Это приводит нас к определению случайной величины как функции на пространстве элементарных исходов. Элементарные соображения, связанные с решением практических задач, показывают, что эта функция не может быть произвольной, а должна удовлетворять определенным ограничениям. Действительно, как отмечалось выше, вероятностное пространство есть тройка $(\Omega, \mathcal{A}, P)$.

В качестве событий рассматриваются только те подмножества пространства Ω, которые принадлежат σ–алгебре $\mathcal{A}$. Только им мы можем приписать некоторую вероятность. С практической точки зрения хотелось бы, чтобы все множества вида $\Big\{\omega:\ a<\xi(\omega)<b\Big\}$, где ξ — случайная величина, были событиями и им можно было приписать вероятность. Это приводит нас к следующему определению.

Определение 29 . *Пусть* $(\Omega,\mathcal{A},P)$ *— вероятностное пространство, а* $(R^1,\mathcal{B})$ *— вещественная прямая с выделенной на ней борелевской σ–алгеброй подмножеств.* **Случайной величиной** *называется функция* $\xi:\ \Omega\to R^1$*, которая обладает следующим свойством:* $\forall B\in\mathcal{B}$

$$\xi^{-1}(B):=\Big\{\omega\in\Omega:\ \xi(\omega)\in B\Big\}\in\mathcal{A}\ . \tag{7.1}$$

Такая функция ξ называется **измеримой**. Таким образом, случайными величинами мы будем называть вещественные измеримые функции на пространстве Ω. Случайные величины мы будем обозначать в дальнейшем греческими буквами $\xi,\ \eta,\ \zeta$ и т. д.

Пример. Монету подбросили 2 раза. Случайная величина ξ равна числу выпавших гербов.

Пусть $\omega_i=1$, если выпал герб и $\omega_i=0$, если выпала решка, $i=1,2$. Тогда пространство элементарных исходов в этом эксперименте будет имеет вид

$$\Omega=\Big\{(0,0),(0,1),(1,0),(1,1)\Big\}$$

Обозначим через $\mathcal{A}$ — σ–алгебру всех подмножеств пространства Ω, $\mathcal{A}_1$ — σ–алгебру, содержащую информацию о первом броске монеты.

Имеем

$$\mathcal{A}_3=\Big\{\emptyset,\Omega,\{(0,0)\},\{(0,1)\},\{(1,0)\},\{(1,1)\},\{(0,0),(1,0)\},\{(0,1),(1,1)\},$$
$$\{(1,0),(1,1)\},\{(0,0),(1,0)\},\{(0,1),(1,0)\},\{(0,0),(1,1)\},\{(0,0),(0,1),(1,0)\},$$
$$\{(0,0),(1,0),(1,1)\},\{(0,0),(0,1),(1,1)\},\{(1,0),(0,1),(1,1)\}\Big\}$$

и

$$\mathcal{A}_1 = \Big\{\emptyset, \Omega, \{(0,0),(0,1)\}, \{(1,0),(1,1)\}\Big\}$$

Изучаемая случайная величина ξ принимает значения 0, 1 или 2. Рассмотрим событие $\Big\{\xi(\omega) < a\Big\}$ при разных a. Имеем

$$\begin{cases} \emptyset, & a \leqslant 0 \\ \{(0,0)\}, & 0 < a \leqslant 1 \\ \{(0,0),(0,1),(1,0)\}, & 1 < a \leqslant 2 \\ \Omega, & a > 2. \end{cases}$$

Множество $\{(0,0),(0,1),(1,0)\} \in \mathcal{A}$ и $\{(0,0),(0,1),(1,0)\} \notin \mathcal{A}_1$. Следовательно, функция ξ измерима относительно σ–алгебры $\mathcal{A}$ и неизмерима относительно σ–алгебры $\mathcal{A}_1$. Иными словами, ξ является случайной величиной на пространстве $(\Omega, \mathcal{A}, P)$ и не является случайной величиной на пространстве $(\Omega, \mathcal{A}_1, P)$.

Замечание. С практической точки зрения достаточно было бы потребовать выполнения свойства 1 для интервалов, т. е. когда $B = (a, b)$. Но нетрудно доказать, что тогда оно справедливо и для любых борелевских множеств (задача!).

Определение 30 . Распределением случайной величины ξ *называется функция P_ξ, заданная на борелевской σ–алгебре $\mathcal{B}$ по правилу: $\forall B \in \mathcal{B}$*

$$P_\xi(B) := P\Big(\xi \in B\Big)\ . \tag{7.2}$$

Распределение вероятностей случайной величины ξ показывает, какова вероятность попадания случайной величины в то или иное множество. Необходимо отметить, что наша модель случайной величины как функции на пространстве элементарных исходов — это некоторая абстракция. В реальном эксперименте мы производим измерение и получаем конкретное число. По большому числу независимых измерений мы можем вычислить частоты, а значит, и вероятности попадания в различные множества и больше ничего. Таким образом, объективной характеристикой случайной величины является ее распределение, так как только его мы можем восстановить на основе результатов эксперимента.

Но распределение случайной величины — это довольно сложный объект, так как надо задать вероятность $P_\xi(B)$ для всех борелевских множеств $B \in \mathcal{B}$, которых достаточно много. Для более компактного описания распределения вводится понятие функции распределения.

Определение 31 . **Функция распределения** $F_\xi(x)$ *случайной величины* ξ *определяется по правилу:* $\forall x \in R^1$

$$F_\xi(x) = P(\xi < x). \tag{7.3}$$

Используя свойства вероятностей событий, нетрудно доказать следующее

Предложение 4 . *Если* $F_\xi(x)$ *— функция распределения случайной величины* ξ, *то*

1) $\forall x \in R^1,\ 0 \leqslant F_\xi(x) \leqslant 1,$

2) если $x_1 \leqslant x_2$, *то* $F_\xi(x_1) \leqslant F_\xi(x_2)$,

3) $F_\xi(x)$ *— непрерывна слева,*

4) $\lim\limits_{x\to-\infty} F_\xi(x) = 0,\ \lim\limits_{x\to\infty} F_\xi(x) = 1.$

5) $P(a \leqslant \xi < b) = F_\xi(b) - F_\xi(a).$

Доказательство. Свойство 1 следует из свойств вероятностей событий. Определим событие $A(x) := (\xi < x)$. Если $x_1 \leqslant x_2$, то $A(x_1) \subset A(x_2)$ и

$$F_\xi(x_1) = P\Big(A(x_1)\Big) \leqslant P\Big(A(x_2)\Big) = F_\xi(x_2).$$

Пусть последовательность $\{x_n\}$ монотонно возрастает и $\lim\limits_n x_n = x$. Тогда последовательность событий $\Big\{A(x_n)\Big\}$ также монотонно возрастает и $\bigcup\limits_{n=1}^{\infty} A(x_n) = A(x)$. Используя свойства непрерывности вероятности, получаем

$$F_\xi(x_n) = P\Big(A(x_n)\Big) \nearrow P\Big(A(x)\Big) = F_\xi(x).$$

Аналогично доказывается свойство 4.

Нетрудно заметить, что

$$(a \leqslant \xi < b) = (\xi < b) - (\xi < a).$$

Тогда

$$P(a \leqslant \xi < b) = P(\xi < b) - P(\xi < a) = F_\xi(b) - F_\xi(a).$$

Замечание. Зная функцию распределения $F_\xi(x)$ случайной величины ξ, мы можем восстановить и все распределение. Наметим схему доказательства.

1. Для интервалов вида $[a, b)$ вероятность находится из свойства 5 функции распределения.

2. Если борелевское множество B есть сумма конечного или счетного числа непересекающихся интервалов, то вероятность попадания в такое множество равна сумме вероятностей попадания в составляющие его интервалы.

3. Произвольное борелевское множество можно аппроксимировать (в определенном смысле) множествами из пункта 2 так, что вероятность попадания в это борелевское множество является пределом вероятностей попадания в аппроксимирующие множества (пример — площадь круга).

7.2 Классификация распределений.

В реальных задачах нам редко приходится работать с распределениями общего типа. Чаще всего мы имеем дело с так называемыми дискретными и абсолютно непрерывными распределениями и их смесями. Иже мы приводим соответствующие определения и примеры.

Определение 32 . *Случайная величина ξ имеет* **дискретное распределение**, *если существует такое конечное или счетное множество $X = \{x_1, x_2, \ldots\}$, что $P(\xi \in X) = 1$. Числа $x_1, x_2, \ldots$ называются* **значениями** *случайной величины ξ, а $p_k = P(\xi = x_k)$ — вероятностями этих значений.*

Предложение 5 . *Пусть случайная величина ξ имеет дискретное распределение с множеством значений $X = \{x_1, x_2, \dots\}$ и вероятностями этих значений $\{p_k\}$. Тогда*

1. $p_k \geqslant 0$.

2. $\sum\limits_k p_k = 1$.

3. $\forall B \in \mathcal{B},\ P_\xi(B) = \sum\limits_{x_k \in B} p_k$.

В частности,

4. $\forall x \in R^1,\ F_\xi(x) = \sum\limits_{x_k < x} p_k$.

Обратно

5. $\forall x_n \in X,\ p_n = P(\xi = x_n) = F_\xi(x_n + 0) - F_\xi(x_n)$.

6. $P(a < \xi < b) = \sum\limits_{a < x_k < b} p_k$.

Задача 3 *Доказать предложение 2.*

Из свойства 3 мы видим, что, зная множество значений X и вероятности $\{p_k\}$ всех возможных значений, можно восстановить и все распределение P_ξ. Поэтому пару $\Big(X, \{p_k\}\Big)$ называют распределением дискретной случайной величины (что, строго говоря, не совсем верно) и записывают ее в виде **таблицы распределения**

x	x_1	x_2	$\dots$	x_n	$\dots$
p	p_1	p_2	$\dots$	p_n	$\dots$

Пример 15 . *Симметричную монету подбрасывают три раза, случайная величина ξ есть число выпавших гербов. Это дискретная величина с таблицей распределения*

x	*0*	*1*	*2*	*3*
p	*1/8*	*3/8*	*3/8*	*1/8*

Определение 33 . *Распределение случайной величины ξ называется* **абсолютно непрерывным**, *если существует такая вещественная функция $\rho_\xi(x)$, что $\forall B \in \mathcal{B}$*

$$P_\xi(B) = P(\xi \in B) = \int\limits_B \rho_\xi(x)\, dx. \tag{7.4}$$

Функция $\rho_\xi(x)$ называется **плотностью распределения** *случайной величины ξ.*

В дальнейшем абсолютно непрерывные распределения будут называться просто непрерывными.

Предложение 6 . *Пусть случайная величина ξ имеет абсолютно непрерывное распределение с плотностью распределения $\rho_\xi(x)$. Тогда справедливы следующие свойства.*

1. $\forall x \in R^1,\ \rho_\xi(x) \geqslant 0.$

2. $\displaystyle\int\limits_{-\infty}^{\infty} \rho_\xi(x)\, dx = 1.$

3. $\displaystyle\forall B \in \mathcal{B},\ P_\xi(B) = P(\xi \in B) = \int\limits_B \rho_\xi(x)\, dx.$

В частности,

4. $\displaystyle\forall x \in R^1,\ F_\xi(x) = \int\limits_{-\infty}^{x} \rho_\xi(y)\, dy.$

5. $\displaystyle P(a \leqslant \xi < b) = F_\xi(b) - F_\xi(a) = \int\limits_a^b \rho_\xi(x)\, dx.$

Обратно

6. $\forall x \in R^1$, *где* $\rho_\xi(x)$ *непрерывна* $\rho_\xi(x) = \dfrac{d}{dx} F_\xi(x).$

7. $\forall x \in R^1\ P(\xi = x) = 0.$

Задача 4 . *Доказать предложение 3.*

Свойства 1 и 2 являются характеристическими свойствами плотности распределения. Любая функция $\rho(x)$, обладающая свойствами 1 и 2, является плотностью распределения некоторой случайной величины ξ. Из свойства 3 мы видим, что, зная плотность распределения, мы можем восстановить и все распределение.

Пример 16 . *Из отрезка $[0, 1]$ случайным образом выбирают точку, ξ — координата выбранной точки.*

Используя геометрическое определение вероятности, получаем

$$F_\xi(x) = P(\xi < x) = \begin{cases} 0 & , \quad x \leqslant 0, \\ x & , \quad 0 < x \leqslant 1, \\ 1 & , \quad x > 1, \end{cases}$$

и

$$\rho_\xi(x) = \begin{cases} 1 & , \quad x \in [0,1], \\ 0 & , \quad x \notin [0,1]. \end{cases}$$

Это равномерное распределение на отрезке $[0,1]$.

Кроме дискретных и абсолютно непрерывных, существуют еще так называемые **сингулярные** распределения. В нашем курсе мы не будем рассматривать распределения такого типа.

Определение 34 . *Пусть $X = \{x_1, x_2, \cdots\}$ — конечное или счетное множество, $\{p_k\}$ — некоторый набор положительных чисел, а $\rho_\xi(x)$ — некоторая неотрицательная функция. Распределение случайной величины называется* **смешанным**, *если*

$$P_\xi(B) = P(\xi \in B) := \sum_{x_k \in B} p_k + \int\limits_B \rho_\xi(x)\,dx.$$

$\Big(X, \{p_k\}\Big)$ *называется* **дискретной компонентой**, *а плотность $\rho_\xi(x)$ —* **непрерывной компонентой** *распределения с.в. ξ. Числа*

$$\alpha_d := \sum_{x_k \in B} p_k, \;\; \alpha_c := \int\limits_B \rho_\xi(x)\,dx$$

называются **весами** *соответствующих компонент. Ясно, что $\alpha_d, \alpha_c \geqslant 0,\;\; \alpha_d + \alpha_c = 1$. Если распределение содержит только одну компоненту, то оно называется* **чистым**.

Пример 17 . *Из отрезка $[0,1]$ случайным образом выбираем точку ω.*

$$\xi = \begin{cases} 0 & , \quad \omega < 1/4, \\ \omega - 1/4 & , \quad 1/4 \leqslant \omega < 3/4, \\ 1/2 & , \quad \omega \geqslant 3/4. \end{cases}$$

Такого типа преобразования часто применяются в теории страхования. Случайная величина ξ имеет две дискретные точки $x_1 = 0$ и $x_2 = 1/2$ с вероятностями $p_1 = P(\xi = 0) = 1/4,\ p_2 = P(\xi = 1/2) = 1/4$ и равномерное распределение на отрезке $[1/4,\ \ 3/4]$, т.е. $\rho_\xi(x) = 1$ при $x \in [1/4,\, 3/4]$. Веса компонент равны $\alpha_d = 1/2$ и $\alpha_c = 1/2$.

7.3 Примеры стандартных распределений

В этом разделе мы приведем несколько примеров дискретных и непрерывных распределений, которые применяются как при исследовании теоретических вопросов, так и при решении практических задач.

1. Случайная величина ξ имеет **вырожденное в точке x распределение**, если $P(\xi = x) = 1$.

Фактически мы имеем не случайную величину, а константу.

2. Случайная величина ξ называется **индикатором события** A, если

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

Это дискретная случайная величина с множеством значений $X = \{0, 1\}$ и вероятностями значений $p = P(\xi = 1)$ и $q = 1 - p = P(\xi = 0)$. Ее распределение называется **распределением Бернулли** с параметром p. Обозначение: $\xi \in Bi(1, p)$.

Индикатор события является в некотором смысле «элементарным кирпичиком» при построении произвольных дискретных случайных величин.

Задача 5 . *Пусть ξ есть дискретная случайная величина с множеством значений $X = \{x_1, x_2, \cdots\}$ и вероятностями значений $p_k = P(\xi = x_k)$. Обозначим $A_k = \{\xi = x_k\}$. Тогда*

1) $A_i A_j = \emptyset,\ i \neq j$,

2) $A_1 + A_2 + \cdots = \Omega$,

3) $\xi(\omega) = \sum\limits_k x_k I_{A_k}(\omega)$.

3. Дискретная случайная величина ξ имеет **биномиальное распределение** с параметрами n и p, если $X = \{0, 1, \cdots, n\}$ и

$$P(\xi = m) = C_n^m p^m (1-p)^{n-m}.$$

В этом случае будем писать $\xi \in Bi(n, p)$. Здесь n — целое число, $n \geqslant 1$, $0 < p < 1$. Эта случайная величина есть число успехов в n независимых испытаниях схемы Бернулли. Такие величины часто появляются в теории страхования, социологии, экономике, физике и других науках.

4. Дискретная случайная величина ξ имеет **геометрическое распределение** с параметром p, если $X = \{0, 1, 2, \ldots\}$ и

$$P(\xi = m) = p(1-p)^m, \ m = 0, 1, \ldots.$$

Здесь $0 < p < 1$. Эта случайная величина равна числу испытаний в схеме Бернулли, предшествующих появлению первого успеха.

5. Дискретная случайная величина ξ имеет **отрицательное биномиальное распределение** с параметрами r и p, где $r \geqslant 1$ — целое, $0 < p < 1$, если $X = \{0, 1, 2, \ldots\}$ и

$$P(\xi = m) = C_{r+m-1}^m p^r (1-p)^m.$$

При $r = 1$ получаем геометрическое распределение. Эти случайные величины равны числу испытаний в схеме Бернулли, предшествующих появлению r–го успеха. Это распределение часто используется в теории страхования при описании числа исков, поступивших в страховую компанию за определенный промежуток времени.

6. Дискретная случайная величина ξ имеет **распределение Пуассона** с параметром λ, если $X = \{0, 1, 2, \ldots\}$ и

$$P(\xi = m) = \frac{\lambda^m}{m!} e^{-\lambda}, \ m = 0, 1, \ldots.$$

Появляется как предельный случай для биномиального распределения, если $p \to 0$, $n \to \infty$, $np = \lambda$. Часто используется в теории страхования, теории массового обслуживания, теории надежности и других прикладных разделах теории вероятностей. Описывает, как правило, число исков, заявок, отказов, поступивших за определенный промежуток времени.

7. Случайная величина ξ имеет **равномерное распределение** на отрезке $[a, b]$, где $-\infty < a < b < \infty$, если у нее существует плотность

$$\rho_\xi(x) = \begin{cases} \dfrac{1}{b-a} & , \quad x \in (a, b)\ , \\ 0 & , \quad x \notin (a, b)\ . \end{cases}$$

Эта модель часто используется для описания распределения случайного момента времени, если известно, что он меняется в ограниченном интервале.

8. Случайная величина ξ имеет **показательное распределение** с параметром $\lambda > 0$, если она обладает плотностью

$$\rho_\xi(x) = \begin{cases} \lambda e^{-\lambda x} & , \quad x > 0, \\ 0 & , \quad x \leqslant 0. \end{cases}$$

Это распределение обладает целым рядом замечательных свойств и часто используется при описании времени между поступлениями двух последовательных заявок, исков, отказов и т.п.

9. Случайная величина ξ имеет **гамма–распределение** с параметрами (α, β), $\alpha, \beta > 0$, если оно обладает плотностью

$$\rho_\xi(x) = \begin{cases} \dfrac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} & , \quad x > 0, \\ 0 & , \quad x \leqslant 0, \end{cases}$$

где

$$\Gamma(\alpha) = \int\limits_0^\infty x^{\alpha-1} e^{-x} dx$$

есть гамма–функция Эйлера. Напомним, что гамма–функция обладает следующими свойствами:

1) $\Gamma(\alpha+1) = \alpha\Gamma(\alpha)$,

2) $\Gamma(1/2) = \sqrt{\pi}$, $\Gamma(1) = 1$,

3) $\Gamma(n+1) = n!$.

Это распределение находит применение в теории массового обслуживания, теории надежности, теории страхования и риска и других прикладных разделах теории вероятностей. При $\alpha = 1,\ \beta = \lambda$ получаем показательное распределение с параметром λ.

10. Случайная величина ξ имеет **нормальное (гауссовское)** распределение с параметрами $(a, \sigma^2),\ a \in R^1, \sigma^2 > 0$, если оно обладает плотностью

$$\rho_\xi(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-a)^2}{2\sigma^2}},\ x \in R^1.$$

Мы будем использовать обозначение $\xi \in N(a, \sigma^2)$.

Если $a = 0,\ \sigma^2 = 1$, то мы имеем **стандартное нормальное распределение**. В этом случае

$$\rho_\xi(x) = \varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2},\ x \in R^1.$$

Функция

$$F_\xi(x) = \Phi(x) = \int\limits_{-\infty}^{x} \varphi(y)\,dy$$

есть **функция распределения** стандартного нормального закона. Часто используется другая функция

$$\Phi_0(x) = \int\limits_{0}^{x} \varphi(y)\,dy,$$

которая называется **функцией Лапласа** или **интегралом вероятностей**. Справедливы следующие соотношения:

1) $\Phi(x) = \Phi_0(x) + 1/2$,

2) $\Phi_0(+\infty) = 1/2$,

3) $\Phi_0(-x) = -\Phi_0(x)$.

Для функцией $\varphi(x),\ \Phi(x)$ и $\Phi_0(x)$ составлены подробные таблицы (смотри, например, Большев Л.Н., Смирнов Н.И. «Таблицы математической статистики»). Ниже будет показано: если $\xi \in N(a, \sigma^2)$, то

случайная величина
$$\xi_0 = \frac{\xi - a}{\sigma}$$
имеет стандартное нормальное распределение. Таким образом, зная $\Phi(x)$ или $\Phi_0(x)$, мы можем вычислять вероятности для случайных величин с произвольным нормальным распределением. Например,
$$P(|\xi - a| \leqslant 3\sigma) = \Phi(3) - \Phi(-3) = 2\Phi_0(3) \approx 0.9973,$$
т. е. практически достоверно, что случайная величина ξ отклоняется от точки a на расстояние не более 3σ. Этот результат часто применяется на практике и носит название «правило трех σ».

7.4 Функциональные преобразования случайной величины.

Одной из наиболее важных прикладных задач, связанных со случайными величинами, является задача нахождения распределения функции от случайной величины. При этом в качестве функционального преобразования нежно брать такое, чтобы мы получили вновь случайную величину.

Определение 35 . *Функция $g : R^1 \to R^1$ называется* **борелевской**, *если $\forall B \in \mathcal{B}$ множество*
$$g^{-1}(B) := \{x \in R^1 : g(x) \in B\}$$
также является борелевским.

Если $\xi : \Omega \to R^1$ есть случайная величина, а $g : R^1 \to R^1$ - борелевская функция, то $\eta = f(\xi)$ также является случайной величиной. Действительно,
$$\eta : \Omega \xrightarrow{\xi} R^1 \xrightarrow{g} R^1 .$$
Тогда $\eta^{-1}(B) = \xi^{-1}(g^{-1}(B)) \in \mathcal{A}$, т.е. является случайным событием для любого $B \in \mathcal{B}$.

Найдем распределение с.в. η. По определению

$$P_\eta(B) = P(\eta \in B) = P\Big(g(\xi) \in B\Big) = P\Big(\xi \in g^{-1}(B)\Big) = P_\xi\Big(g^{-1}(B)\Big)$$

т. е.

$$P_\eta(B) = P_\xi\Big(g^{-1}(B)\Big) \ .$$

Покажем, как это можно записать в терминах функций распределения. В общем случае это сделать не удается, но в некоторых специальных случаях можно получить удовлетворительные результаты.

Предложение 7 . *Пусть $y = g(x)$ — строго монотонная функция. Тогда для $\eta = g(\xi)$ мы имеем*

$$F_\eta(y) = F_\xi\Big(g^{-1}(y)\Big) \ ,$$

если $g(x)$ строго возрастает, и

$$F_\eta(y) = 1 - F_\xi\Big(g^{-1}(y) + 0\Big) \ ,$$

если $g(x)$ строго убывает.

Доказательство. Докажем результат для сличая строго возрастающей функции $g(x)$. Второй случай доказывается аналогично. По определению функции распределения

$$F_\eta(y) = P(\eta < y) = P\Big(g(\xi) < y\Big) = P\Big(\xi < g^{-1}(y)\Big) = F_\xi\Big(g^{-1}(y)\Big) \ .$$

Примеры. 1. $y = g(x) = \sigma x + a, \ \sigma > 0, \ a \in R^1$. Тогда для $\eta = \sigma\xi + a$ получаем

$$F_\eta(y) = P(\sigma\xi + a < y) = P\left(\xi < \frac{y-a}{\sigma}\right) = F_\xi\left(\frac{y-a}{\sigma}\right) \ .$$

В частности, если ξ имеет плотность $\rho_\xi(x)$, то существует

$$\rho_\eta(y) = \frac{d}{dy}F_\eta(y) = \frac{d}{dy}F_\xi\left(\frac{y-a}{\sigma}\right) = \frac{1}{\sigma}\rho_\xi\left(\frac{y-a}{\sigma}\right) \ .$$

2. Если $\xi \in N(0,1)$, то $\eta = \sigma\xi + a \in N(a, \sigma^2)$.

Рассмотрим теперь подробнее случаи дискретного и абсолютно непрерывного распределений.

Если ξ имеет дискретное распределение с множеством значений $X = \{x_1, x_2, \ldots\}$ и вероятностями появления значений $p_n = P(\xi = x_n)$, то случайная величина $\eta = g(\xi)$ также будет иметь дискретное распределение с множеством значений $Y = \{y_1, y_2, \ldots\}$, где $y_k = g(x_m)$ для некоторого x_m, и вероятностями значений

$$q_k = P(\eta = y_k) = \sum_{x_n : g(x_n) = y_k} p_n \ .$$

Пример. Случайная величина ξ имеет дискретное распределение следующего вида:

x	-2	-1	0	1	2
p	0.3	0.1	0.1	0.3	0.2

Тогда случайная величина $\eta = \xi^2$ является дискретной и принимает значения 0,1 и 4 с вероятностями

$$P(\eta = 0) = P(\xi = 0) = 0.1 \ ,$$
$$P(\eta = 1) = P(\xi = -1) + P(\xi = 1) = 0.4 \ ,$$
$$P(\eta = 4) = P(\xi = -2) + P(\xi = 2) = 0.5 \ .$$

Если ξ имеет непрерывное распределение с плотностью $\rho_\xi(x)$, то распределение случайной величины $\eta = g(\xi)$ может и не являться непрерывным. Более того, оно может быть даже дискретным.

Пример. ξ имеет равномерное на $[0,1]$ распределение,

$$g(x) = \begin{cases} 0 \quad , \quad x \in [0, 0.5] \ , \\ 1 \quad , \quad x \in [0.5, 1] \ . \end{cases}$$

Случайная величина $\eta = g(\xi)$ принимает два значения 0 и 1 с равными вероятностями $1/2$.

Таким образом, на функцию $g(x)$ необходимо наложить некоторые дополнительные ограничения. Один частный, но практически важный случай рассматривается в следующей теореме.

Теорема 3 . *Пусть с.в. ξ имеет непрерывное распределение с плотностью $\rho_\xi(x)$, $y = g(x)$ — строго монотонная дифференцируемая функция. Тогда с.в. $\eta = g(\xi)$ также имеет непрерывное распределение и ее плотность имеет вид*

$$\rho_\eta(y) = \rho_\xi\Big(g^{-1}(y)\Big)\left|\frac{d}{dx}g(x)_{x=g^{-1}(y)}\right|^{-1} .$$

Доказательство. Проведем доказательство для сличая, когда $y = g(x)$ строго возрастает. Пусть $y \in R^1$ такая точка, что
$0 < P(\eta < y) < 1$. Тогда

$$F_\eta(y) = P(\eta < y) = P\Big(g(\xi) < y\Big) = P\Big(\xi < g^{-1}(y)\Big) = F_\xi\Big(g^{-1}(y)\Big) .$$

Используя правила дифференцирования сложной и обратной функций, получаем

$$\rho_\eta(y) = \frac{d}{dy}F_\eta(y) = \rho_\xi\Big(g^{-1}(y)\Big)\left(\frac{d}{dx}g\Big(g^{-1}(y)\Big)\right)^{-1} .$$

Доказательство для строго бывающей функции аналогично.

Аналогичною теорем можно сформулировать и для сличая немонотонных функций $y = g(x)$. Но такого рода результат редко используется в реальных задачах. Обычно легче провести заново все расчеты в каждом конкретном случае.

Пример. *С.в. $\xi \in N(0,1)$, найти распределение с.в.$\eta = \xi^2$.*

Новая случайная величина $\eta = \xi^2$ принимает только положительные значения. Пусть $y > 0$.

$$\begin{aligned} F_\eta &= P(\eta < y) = P\big(\xi^2 < y\big) = P(|\xi| < \sqrt{y}) \\ &= P\big(-\sqrt{y} < \xi < \sqrt{y}\big) = F_\xi\big(\sqrt{y}\big) - F_\xi\big(-\sqrt{y}\big). \end{aligned}$$

$$\begin{aligned} \rho_\eta(y) &= \frac{d}{dy}F_\eta(y) = \rho_\xi\big(\sqrt{y}\big)\frac{1}{2\sqrt{y}} + \rho\big(-\sqrt{y}\big)\frac{1}{2\sqrt{y}} = \\ &= \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}y}\frac{1}{2\sqrt{y}} + \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}y}\frac{1}{2\sqrt{y}} = \frac{1}{\sqrt{2\pi}}y^{-\frac{1}{2}}e^{-\frac{1}{2}y} . \end{aligned}$$

Глава 8

Случайный вектор

8.1 Распределение случайного вектора

Во многих реальных задачах мы имеем не одну, а несколько случайных величин в одном и том же эксперименте. Иногда их удобно рассматривать как единый объект. Это приводит нас к следующему определению.

Определение 36 . *n–мерным случайным вектором называется набор $\xi = (\xi_1, \ldots, \xi_n)$ случайных величин, заданных на одном и том же вероятностном пространстве $(\Omega, \mathcal{A}, P)$.*

Фактически случайный вектор ξ есть отображение $\xi : \Omega \to R^n$. Нетрудно показать (задача !), что это отображение является **борелевским**, т.е. для любого борелевского подмножества $B \subset R^n$ (σ–алгебру всех борелевских подмножеств в R^n мы будем обозначать $\mathcal{B}_n$) мы имеем $\xi^{-1}(B) \in \mathcal{A}$. Как и для случайных величин, можно дать следующее

Определение 37 . *Распределением случайного вектора ξ называется функция P_ξ , заданная на σ–алгебре $\mathcal{B}_n$ по правилу*

$$P_\xi(B) = P(\xi \in B).$$

Распределение является объективной характеристикой случайного вектора, которую можно однозначно восстановить из эксперимента.

Но распределение, будучи удобной характеристикой в теоретических исследованиях, является довольно сложным для реальных задач. Как и в одномерном случае, используют понятие функции распределения.

Определение 38 . **Функцией распределения** *случайного вектора* $\xi = (\xi_1, \ldots, \xi_n)$ *называется функция* $F_\xi(x),\ x \in R^n$, *такая, что* $\forall x = (x_1, \ldots, x_n) \in R^n$,

$$F_\xi(x_1, \ldots, x_n) = P(\xi_1 < x_1, \ldots, \xi_n < x_n).$$

Основные свойства функции распределения случайного вектора собраны в следующем предложении.

Предложение 8 . *Функция распределения* $F_\xi(x) = F_\xi(x_1, \ldots, x_n)$ *случайного вектора* $\xi = (\xi_1, \ldots, \xi_n)$ *обладает следующими свойствами:*

1. $\forall x \in R^n,\ 0 \leqslant F_\xi(x) \leqslant 1$.

2. $F_\xi(x)$ *не убывает по каждому аргументу* $x_i,\ i = \overline{1, n}$.

3. $F_\xi(x)$ — *непрерывна слева по каждому аргументу* $x_i,\ i = \overline{1, n}$.

4. $F_\xi(x) \to 0$, *если* **некоторое** $x_i \to -\infty,\ i = \overline{1, n}$.
$F_\xi(x) \to 1$, *если* **все** $x_i \to \infty,\ i = \overline{1, n}$.

5. $\forall x = (x_1, \ldots, x_n) \in R^n$

$$P(x_1 \leqslant \xi_1 < x_1 + h_1, \ldots, x_n \leqslant \xi_n < x_n + h_n) =$$
$$\Delta h_1 \ldots \Delta h_n F_\xi(x_1, \ldots, x_n)\ ,$$

где

$$\Delta h_i F_\xi(x_1, \ldots, x_n) =$$
$$F_\xi(x_1, \ldots x_i + h_i, \ldots, x_n) - F_\xi(x_1, \ldots x_i, \ldots, x_n)\ .$$

6. $F_\xi(x_1, \ldots x_{i-1}, \infty, x_{i+1}, \ldots, x_n)$ *есть функция распределения случайного вектора* $(\xi_1, \ldots \xi_{i-1}, \xi_{i+1}, \ldots, \xi_n)$.

Задача 6 . *Доказать предложение 1.*

Замечание. В силу свойства 5 по функции распределения $F_\xi(x)$ можно найти вероятности попадания в множества $B = [a_1, b_1) \times \ldots \times [a_n, b_n)$. Далее, так же как и в одномерном случае, можно восстановить распределение ξ для любых борелевских множеств B, аппроксимируя их параллелограммами.

8.2 Классификация распределений

Как и в одномерном случае, мы выделим два важных частных случая распределений, которые наиболее часто используются на практике. Конечно, бывают и более общие примеры, но мы не будем их подробно рассматривать в нашем курсе.

Определение 39 . *Случайный вектор $\xi = (\xi_1, \ldots, \xi_n)$ имеет* **дискретное распределение***, если существует конечное или счетное множество $X \subset R^n$, такое, что $P(\xi \in X) = 1$.*

Если $x = (x_1, \ldots, x_n) \in X$ — одно из возможных значений случайного вектора ξ, то $p(x) = P(\xi = x)$ называется **вероятностью появления значения** *x.*

Обычно используют следующую стандартную форму описания распределения дискретного случайного вектора. Ясно, что каждая координата ξ_k случайного вектора ξ имеет дискретное распределение. Пусть $X^{(k)} = \left\{x_1^{(k)}, x_2^{(k)}, \ldots, x_{i_k}^{(k)}, \ldots\right\}$ есть множество значений случайной величины ξ_k. Образуем множество

$$X = X^{(1)} \times \ldots \times X^{(n)} = \left\{x = (x_1, \ldots, x_n) : x_k \in X^{(k)},\ k = \overline{1, n}\right\}$$

в R^n.

Задача. Доказать, что $P(\xi \in X) = 1$, т.е. X можно взять в качестве множества значений случайного вектора ξ.

Для произвольного вектора $x = (x_{i_1}^{(1)}, \ldots, x_{i_n}^{(n)})$, где $x_{i_k}^{(k)} \in X^{(k)}$, обозначим через

$$P_{i_1 \ldots i_n} = P(\xi = x) = P\Big(\xi_1 = x_{i_1}^{(1)}, \ldots, \xi_n = x_{i_n}^{(n)}\Big) \tag{8.1}$$

вероятность появления значения x случайного вектора ξ. При таком выборе множества X некоторые его элементы будут появляться с вероятностью 0.

Пример. Случайный вектор $\xi = (\xi_1, \xi_2)$ имеет два значения $(1, 1)$ и $(2, 2)$, которые появляются с вероятностями $1/2$.

Имеем $X^{(1)} = \{1, 2\}$ и $X^{(2)} = \{1, 2\}$ и, следовательно,

$$X = X^{(1)} \times X^{(2)} = \Big\{(1,1), (1,2), (2,1), (2,2)\Big\}.$$

Точка $x = (1, 2)$ входит в построенное выше множество X, но $P(\xi = x) = 0$.

Пару $\Big(X, \{P_{i_1 \dots i_n}\}\Big)$ будем называть **распределением дискретного случайного вектора** ξ, хотя, строго говоря, это не совсем точно. Для $n = 2$ распределение дискретного случайного вектора обычно задают в виде следующей таблицы, называемой **таблицей распределения**:

$\xi_2 \setminus \xi_1$	x_1	x_2	$\dots$	x_i	$\dots$
y_1	P_{11}	P_{21}	$\dots$	P_{i1}	$\dots$
y_2	P_{12}	P_{22}	$\dots$	P_{i2}	$\dots$
$\vdots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
y_j	P_{1j}	P_{2j}	$\dots$	P_{ij}	$\dots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$

Здесь $\{x_1, x_2, \dots, x_i, \dots\}$ — множество значений для ξ_1, $\{y_1, y_2, \dots, y_j, \dots\}$ — множество значений для ξ_2, а $P_{ij} = P(\xi_1 = x_i, \xi_2 = y_j)$.

Предложение 9 . *Распределение* $\Big(X, \{P_{i_1,\dots,i_n}\}\Big)$ *дискретного случайного вектора* $\xi = (\xi_1, \dots, \xi_n)$ *обладает следующими свойствами:*

1) $\forall x = \big(x^{(1)}_{i_1}, \dots, x^{(n)}_{i_n}\big) \in X \quad P_{i_1,\dots,i_n} \geqslant 0$,

2) $\sum\limits_{i_1,\dots,i_n} P_{i_1,\dots,i_n} = 1$,

3) $\forall B \in \mathcal{B}_n$

$$P_\xi(B) = P(\xi \in B) = \sum_{(x^{(1)}_{i_1},\dots,x^{(n)}_{i_n}) \in B} P_{i_1,\dots,i_n} \ ,$$

4) $P\Big(\xi_1 = x^{(1)}_{i_1}, \dots \xi_{k-1} = x^{(k-1)}_{i_{k-1}}, \xi_{k+1} = x^{(k+1)}_{i_{k+1}}, \dots, \xi_n = x^{(n)}_{i_n}\Big) =$

$$= \sum_{i_k} P_{i_1,\dots,i_{k-1},i_k,i_{k+1},\dots,i_n} \ .$$

Все эти свойства легко следуют из приведенных выше определений и свойств вероятностей. Поэтому доказательство этого предложения предлагается в виде задачи.

Пример. Пусть мы приводим n независимых испытаний, каждое из которых может закончиться одним из r исходов ($r \geqslant 2$) и вероятности появления этих исходов одни и те же в каждом испытании и равны $p_1, \ldots, p_r$. Пусть ξ_k есть число появлений k–го исхода в этих n испытаниях. Тогда $\xi = (\xi_1, \ldots, \xi_r)$ есть дискретный случайный вектор. Его значениями являются векторы $m = (m_1, \ldots, m_r)$, такие, что m_k — целые неотрицательные числа и $m_1 + \ldots + m_r = n$. Как было показано выше, при изучении последовательностей независимых испытаний

$$P(\xi = m) = \frac{n!}{m_1! \ldots m_r!} p_1^{m_1} \ldots p_r^{m_r} \ . \qquad (8.2)$$

Такое распределение называется **полиномиальным распределением** с параметрами $(n; p_1, \ldots, p_r)$.

Определение 40 . *Случайный вектор* $\xi = (\xi_1, \ldots, \xi_n)$ *имеет* **абсолютно непрерывное распределение**, *если существует вещественная функция* $\rho_\xi(x),\ x \in R^n$, *такая, что* $\forall B \in \mathcal{B}_n$

$$P_\xi(B) = P(\xi \in B) = \int\limits_B \rho_\xi(x)\, dx \ .$$

Функция $\rho_\xi(x)$ *называется* **плотностью распределения** *случайного вектора* ξ.

Нетрудно доказать следующее утверждение, доказательство которого предлагается в качестве задачи.

Предложение 10 . *Случайный вектор* $\xi = (\xi_1, \ldots, \xi_n)$ *имеет абсолютно непрерывное распределение с плотностью* $\rho_\xi(x),\ x \in R^n$. *Тогда справедливы следующие свойства:*

1) $\forall \in R^n\ \rho_\xi(x) \geqslant 0$ *;*

2) $\int\limits_{R^n} \rho_\xi(x) dx = 1$ *;*

3) $\forall B \in \mathcal{B}_n \ P(\xi \in B) = \int\limits_B \rho_\xi(x)\,dx$ *;*

4) $\forall x = (x_1, \ldots, x_n) \in R^n$

$$F_\xi(x_1, \ldots, x_n) = \int\limits_{-\infty}^{x_1} \cdots \int\limits_{-\infty}^{x_n} \rho_\xi(y_1, \ldots, y_n)\,dy_n \ldots dy_1 \ ;$$

5) если $(x_1, \ldots, x_n)$ *— точка непрерывности плотности* $\rho_\xi(x)$ *, то*

$$\rho_\xi(x_1, \ldots, x_n) = \frac{\partial^n F_\xi(x_1, \ldots, x_n)}{\partial x_1 \ldots \partial x_n} \ ;$$

6) плотность случайного вектора $\tilde{\xi} = (\xi_1, \ldots, \xi_{k-1}, \xi_{k+1}, \ldots, \xi_n)$ *можно вычислить по формуле*

$$\rho_{\tilde{\xi}}(x_1, \ldots, x_{k-1}, x_{k+1}, \ldots, x_n) =$$

$$\int\limits_{-\infty}^{\infty} \rho_\xi(x_1, \ldots, x_{k-1}, x_k, x_{k+1}, \ldots, x_n)\,dx_k \ .$$

Замечание. Если мы имеем некоторый случайный вектор $\xi = (\xi_1, \ldots, \xi_n)$, то, выбирая некоторые из его координат, например первые m, мы получаем новый случайный вектор $\tilde{\xi} = (\xi_1, \ldots, \xi_m)$, который называют **подвектором** вектора ξ. Выше было показано, как найти распределение подвектора, когда убирают одну из координат. Применяя эту процедуру несколько раз, мы сможем найти распределение произвольного подвектора. Распределение отдельно взятой координаты ξ_i вектора ξ называется **одномерным** или **маргинальным** распределением.

Как и в одномерном случае, можно ввести понятие смеси распределений, но мы не будем его рассматривать подробно так как здесь не возникает ничего нового.

Примеры. 1. Случайный вектор $\xi = (\xi_1, \ldots, \xi_n)$ имеет равномерное распределение в области D, если он обладает плотностью распределения следующего вида:

$$\rho_\xi(x) = \begin{cases} \frac{1}{L(D)} & , \quad x \in D \ , \\ 0 & , \quad x \notin D \ , \end{cases}$$

где $L(D)$ — мера Лебега области D. Фактически мы имеем дело с геометрическим определением вероятности.

2. Случайный вектор $\xi = (\xi_1, \xi_2)$ имеет **двумерное нормальное распределение**, если он обладает плотностью распределения следующего вида:

$$\rho_\xi(x_1, x_2) = \frac{1}{2\pi\sqrt{\sigma_1^2\sigma_2^2(1-\rho^2)}} \times$$

$$\exp\left\{ -\frac{1}{2(1-\rho^2)} \Big[\frac{(x_2-a_2)^2}{\sigma_2^2} - 2\rho\frac{(x_1-a_1)}{\sigma_1}\frac{(x_2-a_2)}{\sigma_2} + \frac{(x_1-a_1)^2}{\sigma_1^2} \Big] \right\} \ .$$

Числа $a_1, a_2 \in R^1$, $\sigma_1^2, \sigma_2^2 > 0$, $\rho \in (-1, 1)$ называются **параметрами** двумерного нормального распределения. Их вероятностный смысл будет выяснен позднее.

8.3 Независимые случайные величины

При изучении свойств вероятностей случайных событий мы видели, что понятие независимости событий играет важную роль при вычислении вероятностей сложных событий. Аналогично понятие независимости является центральным понятием в теории случайных величин, их функциональных преобразований и других вопросах.

Определение 41 . *Случайные величины* $\xi_1, \ldots, \xi_n$ *называются* **независимыми***, если для любых борелевских* $B_1, \ldots, B_n \in \mathcal{B}_1$

$$P(\xi_1 \in B_1, \ldots, \xi_n \in B_n) = P(\xi_1 \in B_1) \cdot \ldots \cdot P(\xi_n \in B_n) \ . \qquad (8.3)$$

Дадим эквивалентные формулировки понятия независимости случайных величин в терминах функций распределения, а также для случаев дискретных и непрерывных распределений.

Предложение 11 . *Пусть мы имеем случайный вектор* $\xi = (\xi_1, \ldots, \xi_n)$. *Его компоненты* $\xi_1, \ldots, \xi_n$ *независимы тогда и только тогда, когда*

$$F_\xi(x_1, \ldots, x_n) = F_{\xi_1}(x_1) \cdot \ldots \cdot F_{\xi_n}(x_n) \ . \tag{8.4}$$

В случае дискретных распределений условие независимости эквивалентно условию

$$P\Big(\xi_1 = x_{i_1}^{(1)}, \ldots, \xi_n = x_{i_n}^{(n)}\Big) = P\Big(\xi_1 = x_{i_1}^{(1)}\Big) \cdot \ldots \cdot P\Big(\xi_n = x_{i_n}^{(n)}\Big) \ , \tag{8.5}$$

а в случае непрерывных - условию

$$\rho_\xi(x_1, \ldots, x_n) = \rho_{\xi_1}(x_1) \cdot \ldots \cdot \rho_{\xi_n}(x_n) \ . \tag{8.6}$$

Доказательство. Рассмотрим множества

$$B_i = \{x \in R^1 : \ x < x_i\}, \ i = \overline{1, n} \ .$$

Для них из (3) следует (4). Обратно, из (4) легко получить (3) для параллелепипедов, а затем аппроксимировать произвольные B_i с помощью сумм отрезков. Свойство (6) получается из (4) дифференцированием. Свойство (5) следует непосредственно из определения независимости.

Пример 1. Пусть мы имеем схему Бернулли с параметрами n и p. Пусть $\varepsilon_i = 1$, если в i-м испытании был «успех», и равно 0 в противном случае. Тогда случайные величины $\varepsilon_1, \ldots, \varepsilon_n$ — независимы. Кстати, число успехов S_n в этих n испытаниях представимо в виде $S_n = \varepsilon_1 + \ldots + \varepsilon_n$.

Пример 2. Пусть $\xi = (\xi_1, \xi_2)$ имеет двумерное нормальное распределение. ξ_1 ξ_2 будут независимы тогда и только тогда, когда $\rho = 0$ (задача!).

Нетрудно доказать следующий полезный результат (задача!).

Предложение 12 . *Пусть случайные величины* $\xi_1, \ldots, \xi_n, \xi_{n+1}, \ldots, \xi_{n+m}$ — *независимы, а* $y = f(x_1, \ldots, x_n)$ *и* $y = g(x_1, \ldots, x_m)$ — *борелевские функции. Тогда случайные величины* $\eta_1 = f(\xi_1, \ldots, \xi_n)$ *и* $\eta_2 = g(\xi_{n+1}, \ldots, \xi_{n+m})$ *также являются независимыми.*

Пример 3. Пусть мы имеем схему Бернулли с $n_1 + n_2$ испытаниями. Тогда число успехов S_{n_1} в первых n_1 испытаниях и число успехов S_{n_2} в последующих n_2 испытаниях — независимые случайные величины.

8.4 Функциональные преобразования случайных векторов

Как и в одномерном случае, важной с практической точки зрения является задача о вычислении распределения функционального преобразования случайного вектора.

Определение 42 . *Отображение $g : R^n \to R^m$ называется* **борелевским**, *если $\forall B \in \mathcal{B}_m$ мы имеем $g^{-1}(B) \in \mathcal{B}_n$.*

Если g — борелевское отображение, $\xi = (\xi_1, \dots, \xi_n)$ — случайный вектор, то $\eta = (\eta_1, \dots, \eta_m) = g(\xi)$ вновь является случайным вектором. Действительно, если $B \in \mathcal{B}_m$, то $g^{-1}(B) \in \mathcal{B}_n$, а $\eta^{-1}(B) = \xi^{-1} \circ g^{-1}(B) \in \mathcal{A}$. Отсюда нетрудно получить выражение для распределения вектора η, если мы знаем распределение вектора $\xi :\ \forall B \in \mathcal{B}_m$

$$P_\eta(B) = P(\eta \in B) = P\Big(g(\xi) \in B\Big) = \tag{8.7}$$
$$P\Big(\xi \in g^{-1}(B)\Big) = P_\xi\Big(g^{-1}(B)\Big)\ .$$

Рассмотрим теперь отдельно случаи дискретного и непрерывного распределений.

Если ξ имеет дискретное распределение с множеством значений $X = \{x_1, x_2, \dots\}$ и вероятностями $p_n = P(\xi = x_n)$ появления этим значений, то ясно, что случайный вектор $\eta = g(\xi)$ также имеет дискретное распределение с множеством значений $Y = \{y_1, y_2, \dots, y_m, \dots\}$, где каждое $y_m = g(x_n)$ для некоторого x_n, а вероятности $q_m = P(\eta = y_m)$ появления значения можно вычислить по формуле

$$q_m = \sum_{x_n : g(x_n) = y_m} p_n\ . \tag{8.8}$$

Пример. Пусть $\xi = (\xi_1, \xi_2)$ — двумерный случайный вектор с дискретным распределением, $p(x, y) = P(\xi_1 = x, \xi_2 = y)$ — вероятность появления некоторого значения (x, y). Рассмотрим функцию $z = g(x, y) = x + y$. Тогда $\eta = \xi_1 + \xi_2$ есть дискретная случайная величина и вероятность того, что $\eta = z$, где z — одно из возможных значений суммы $\xi_1 + \xi_2$, можно рассчитать по формуле

$$P(\eta = z) = \sum_{(x,y):x+y=z} p(x, y) = \sum_x p(x, z - x) \ . \tag{8.9}$$

Если ξ_1 и ξ_2 независимы, то

$$p(x, y) = P(\xi_1 = x, \xi_2 = y) = P(\xi_1 = x) \cdot P(\xi_2 = y) = P_1(x) P_2(y) \ .$$

В этом случае мы получаем

$$P(\xi_1 + \xi_2 = z) = \sum_x P_1(x) P_2(z - x) \ . \tag{8.10}$$

Это **формула свертки** для дискретных распределений.

Пусть теперь $\xi = (\xi_1, \ldots, \xi_n)$ — случайный вектор, который имеет плотность распределения $\rho_\xi(x), \ x \in R^n$. Как и в одномерном случае, распределение случайного вектора $\eta = g(\xi)$ может не иметь плотности и даже быть дискретным. Необходимы некоторые дополнительные ограничения на функцию $y = g(x)$. Рассмотрим один частный, но практически важный случай.

Предложение 13 . *Пусть $\xi = (\xi_1, \ldots, \xi_n)$ — случайный вектор, который имеет плотность распределения $\rho_\xi(x), \ g : R^n \to R^n$ - взаимно однозначное и непрерывно дифференцируемое отображение. Тогда распределение случайного вектора $\eta = g(\xi)$ является абсолютно непрерывным и его плотность $\rho_\eta(y)$ можно вычислить по формуле*

$$\rho_\eta(y) = \rho_\xi\Big(g^{-1}(y)\Big) \left| J\Big(g^{-1}(y)\Big) \right|^{-1} \ , \tag{8.11}$$

где $J(x)$ - якобиан отображения $y = g(x)$.

Доказательство этого предложения дословно повторяет доказательство в одномерном случае, но теперь мы должны сделать замену переменных в R^n.

Пример. Пусть $y = Ax$, где A — невырожденная квадратная матрица размера n, т.е. мы имеем линейное отображение в R^n. В этом случае $J(x) = detA$ и

$$\rho_\eta(y) = \rho_\xi\Big(A^{-1}y\Big) \cdot \Big|detA\Big|^{-1} .$$

В тех случаях, когда $n \neq m$, последнее предложение не применимо. Но часто можно дополнить отображение $g : R^n \to R^m$ еще одним отображением $f : R^n \to R^{n-m}$ так, чтобы отображение $(g, f) : R^n \to R^n$ уже обладало нужными свойствами.

Пример. Пусть случайный вектор $\xi = (\xi_1, \xi_2)$ имеет плотность распределения $\rho_\xi(x_1, x_2)$. Найдем плотность распределения случайной величины $\eta_1 = \xi_1 + \xi_2$. Здесь $n = 2, m = 1$. Рассмотрим еще одну случайную величину $\eta_2 = \xi_2$. Тогда в целом мы имеем следующее линейное отображение $y_1 = x_1 + x_2,\ y_2 = x_2$. Матрица A этого отображения имеет вид

$$A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} ,$$

$detA = 1$, а обратная матрица A^{-1} равна

$$A^{-1} = \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix} .$$

В предыдущем примере мы получили, что

$$\rho_\eta(y_1, y_2) = \rho_\xi\Big(A^{-1}y\Big) \cdot \Big|detA\Big|^{-1},$$

откуда следует, что

$$\rho_\eta(y_1, y_2) = \rho_\xi(y_1 - y_2, y_2).$$

Чтобы найти плотность распределения для η_1, достаточно проинтегрировать по координате y_2, т.е.

$$\rho_{\eta_1}(y_1) = \int\limits_{-\infty}^{\infty} \rho_\xi(y_1 - y_2, y_2)\, dy_2 \ . \tag{8.12}$$

Если ξ_1 и ξ_2 — независимы, то $\rho_\xi(x_1, x_2) = \rho_{\xi_1}(x_1)\rho_{\xi_2}(x_2)$. Заменяя y_1 на y и y_2 на x, получаем

$$\rho_{\xi_1+\xi_2}(y) = \int\limits_{-\infty}^{\infty} \rho_{\xi_1}(y - x)\rho_{\xi_2}(x)\, dx \ . \tag{8.13}$$

Это **формула свертки** для непрерывных распределений.

В более сложных ситуациях, когда не удается свести задачу к предложению 6, необходимо провести прямые расчеты, вычисляя распределение P_η (например, функцию распределения), а затем находя плотность. Технически это сводится к нахождению множества $g^{-1}(B)$ и вычислению интеграла от $\rho_\xi(x)$ по этому множеству. Чтобы продемонстрировать, как работает этот метод, рассмотрим тот же самый пример: $\eta = \xi_1 + \xi_2$. Вычислим для η функцию распределения:

$$F_\eta(y) = P(\eta < y) = P(\xi_1 + \xi_2 < y) \ .$$

Фактически нам нужно найти вероятность попадания случайного вектора $\xi = (\xi_1, \xi_2)$ в множество

$$B(y) = \Big\{(x_1, x_2) : \ x_1 + x_2 < y\Big\} \subset R^2 \ .$$

Тогда мы имеем

$$F_\eta(y) = P\Big((\xi_1, \xi_2) \in B(y)\Big) = \iint\limits_{B(y)} \rho_\xi(x_1, x_2)\, dx_1\, dx_2 \ .$$

Далее, расписывая двойной интеграл в виде повторного, получаем

$$F_\eta(y) \ = \int\limits_{-\infty}^{\infty} \int\limits_{-\infty}^{y-x_2} \rho_\xi(x_1, x_2)\, dx_1\, dx_2 = \left| \begin{array}{l} x = x_1 + x_2 \\ z = x_2 \end{array} \right| =$$

$$= \int\limits_{-\infty}^{\infty} \int\limits_{-\infty}^{y} \rho_\xi(x-z, z)\, dx\, dz = \int\limits_{-\infty}^{y} \Big[\int\limits_{-\infty}^{\infty} \rho_\xi(x-z, z)\, dz \Big]\, dx \ .$$

Дифференцируя по y, окончательно получаем

$$\rho_\eta(y) = \frac{d}{dy} F_\eta(y) = \int\limits_{-\infty}^{\infty} \rho_\xi(y-z, z)\, dz \ .$$

Таким образом, мы имеем тот же результат, что и ранее.

Глава 9

Математическое ожидание случайной величины

Как отмечалось ранее, наиболее полной характеристикой случайной величины ξ, которую можно получить из результатов эксперимента, является ее распределение P_ξ. Но сделать это реально довольно трудно, а часто и ненужно. Во многих задачах мы интересуемся только тем, около какой точки «в среднем» меняются значения случайной величины ξ и как далеко они (опять «в среднем») могут отклоняться от этого значения. Для этого в теории вероятностей используются понятия математического ожидания (как центра распределения) и дисперсии (для оценки отклонения от этого центра). Мы начнем изложение нашей теории с простейшего случая, когда ξ имеет дискретное распределение.

9.1 Математическое ожидание дискретной случайной величины.

Прежде чем дать абстрактное определение, мы рассмотрим простой пример, который объяснит нам его разумность.

Пусть с.в. ξ принимает конечное множество значений $x_1, \ldots, x_n$ с вероятностями появления этих значений $p_k = P(\xi = x_k)$, $k = 1, \ldots, n$. Предположим, что мы провели N независимых измерений этой случайной величины ξ в одинаковых условиях и получили результаты

$y_1, \ldots, y_N$. Пусть N_k есть число появлений значения x_k в этой серии измерений. Интуитивно мы представляем себе среднее в длинной серии измерений как среднее арифметическое. Тогда

$$\frac{y_1 + \ldots + y_N}{N} = x_1 \frac{N_1}{N} + \ldots + x_n \frac{N_n}{N} \approx x_1 p_1 + \ldots + x_n p_n \ ,$$

т.к. в длинной серии испытаний частоты «тяготеют» к вероятностям. Последнюю сумму естественно считать некоторым теорети- ческим средним или центром распределения случайной величины ξ.

Эти рассуждения приводят нас к следующему определению.

Определение 43 . **Математическим ожиданием (средним)** *дискретной случайной величины* ξ *называется число*

$$M\xi = \sum_n x_n p_n \ ,$$

где $\{x_1, x_2, \ldots\}$ *— множество значений с.в.* ξ, *а* $p_n = P(\xi = x_n)$, *при условии, что последний ряд сходится абсолютно.*

Иногда используют и другое обозначение для математического ожидания — $E\xi$. Абсолютная сходимость нужна для того, чтобы величина $M\xi$ не зависела от порядка нумерации значений с.в. ξ.

Посмотрим, какими свойствами обладают математические ожидания дискретных случайных величин. Для удобства сведем все эти свойства в следующее

Предложение 14 . *Математические ожидания дискретных случайных величин (если они существуют) обладают следующими свойствами.*

1. *Если* $P(\xi = C) = 1$, *то* $M\xi = C$.
2. *Если* $C = const$, *то* $M(C\xi) = CM\xi$.
3. $M(\xi_1 + \xi_2) = M\xi_1 + M\xi_2$.
4. *Если* $\xi \geqslant 0$, *то* $M\xi \geqslant 0$.
5. *Если* $\xi_1 \geqslant \xi_2$, *то* $M\xi_1 \geqslant M\xi_2$.
6. *Если* ξ_1 *и* ξ_2 *— независимы, то* $M(\xi_1 \xi_2) = M\xi_1 \cdot M\xi_2$.

Доказательство. 1. Если $P(\xi = C) = 1$, то $M\xi = C \cdot 1 = C$.

2. Случайная величина $\eta = C\xi$ имеет множество значений

$$Y = \{Cx_1, Cx_2, \dots\}$$

и

$$P(\eta = Cx_n) = p_n \ .$$

Тогда

$$M\eta = \sum_n (Cx_n)p_n = C\sum_n x_n p_n = CM\xi \ .$$

3. Пусть $X = \{x_1, \dots, x_n, \dots\}$ — множество значений с.в. ξ_1, $Y = \{y_1, \dots, y_m, \dots,\}$ — множество значений с.в. ξ_2. Тогда множеством значений с.в. $\eta = \xi_1 + \xi_2$ будет $Z = \{z_1, \dots, z_k, \dots\}$, где $z_k = x_n + y_m$ для некоторых x_n и y_m. Если $p_{nm} = P(\xi_1 = x_n, \xi_2 = y_m)$, то

$$q_k = P(\xi_1 + \xi_2 = z_k) = \sum_{x_n + y_m = z_k} p_{nm} \ .$$

Напомним, что

$$P(\xi_1 = x_n) = \sum_m p_{nm} \ , \ P(\xi_2 = y_m) = \sum_n p_{nm} \ .$$

Отсюда получаем

$$M(\xi_1 + \xi_2) = \sum_k z_k q_k = \sum_k z_k \sum_{x_n + y_m = z_k} p_{nm} =$$

$$= \sum_{n,m} (x_n + y_m) p_{nm} = \sum_n x_n (\sum_m p_{nm}) + \sum_m y_m (\sum_n p_{nm}) =$$

$$\sum_n x_n P(\xi_1 = x_n) + \sum_m y_m P(\xi_2 = y_m) = M\xi_1 + M\xi_2 \ .$$

4. $\xi \geqslant 0$ означает, что все значения $x_n \geqslant 0$. Тогда

$$M\xi = \sum_n x_n p_n \geqslant 0 \ ,$$

т.к. это сумма неотрицательных слагаемых.

5. Это свойство есть прямое следствие свойств 2,3 и 4. Действительно, $\eta = \xi_1 - \xi_2 \geqslant 0$. Тогда

$$M\eta = M(\xi_1 - \xi_2) = M\xi_1 + M(-\xi_2) = M\xi_1 - M\xi_2 \geqslant 0 \ .$$

6. Это свойство доказывается так же, как и свойство 3. Необходимо только отметить, что в случае независимых с.в.

$$p_{nm} = P(\xi_1 = x_n, \xi_2 = y_m) = P(\xi_1 = x_n) \cdot P(\xi_2 = y_m) \ .$$

9.2 Определение математического ожидания в общем случае

Теперь мы хотим определить математическое ожидание с.в. ξ, распределение которой не обязательно дискретно. Начнем со случая неотрицательных случайных величин. Наш план состоит в том, чтобы аппроксимировать такие случайные величины с помощью дискретных, для которых математическое ожидание уже определено, а математическое ожидание ξ положить равным пределу математических ожиданий приближающих ее дискретных с.в. Кстати, это очень полезная общая идея, состоящая в том, что некоторая характеристика сначала определяется для простых объектов, а затем для более сложных объектов она определяется с помощью аппроксимации их более простыми (смотри, например, определение площадей, исследование функций путем разложения их в ряды и т.п.).

Лемма 3 . *Пусть ξ есть произвольная неотрицательная случайная величина. Тогда существует последовательность $\{\xi_n\}$ дискретных случайных величин, таких, что*

1) $\xi_n(\omega) \geqslant 0$, $\forall n$,

2) $\xi_n(\omega) \leqslant \xi_{n+1}(\omega)$,

3) $\xi_n(\omega) \to \xi(\omega)$ равномерно по $\omega \in \Omega$ при $n \to \infty$.

Доказательство. Разобьем полуось $[0, \infty)$ на равные отрезки длины 2^{-n} и определим

$$\xi_n(\omega) = \frac{k}{2^n} \ , \quad k2^{-n} \leqslant \xi(\omega) < (k+1)2^{-n}.$$

Тогда свойства 1 и 2 легко следуют из определения с.в. ξ_n, и

$$\left|\xi_n(\omega) - \xi(\omega)\right| \leqslant 2^{-n} \to 0,\ n \to \infty,$$

$\forall \omega \in \Omega$.

Лемма 4 . *Пусть ξ — неотрицательная случайная величина и $\{\xi_n\}$ и $\{\eta_n\}$ — две последовательности дискретных с.в., обладающих свойствами 1 — 3 из леммы 1. Тогда*

$$\lim_n M\xi_n = \lim_n M\eta_n \ .$$

Доказательство. Отметим, что для неотрицательных случайных величин мы допускаем $M\xi = \infty$.

В силу свойства 3 легко видеть, что существует последовательность $\{\varepsilon_n\}$ положительных чисел, такая, что

$$\left|\xi_n(\omega) - \eta_n(\omega)\right| \leqslant \varepsilon_n \to 0,\ n \to \infty \ .$$

Отсюда следует, что

$$\xi_n(\omega) - \varepsilon_n \leqslant \eta_n(\omega) \leqslant \xi_n(\omega) + \varepsilon_n$$

$\forall \omega \in \Omega$. Используя свойства математических ожиданий для дискретных случайных величин, получаем

$$M\xi_n - \varepsilon_n \leqslant M\eta_n \leqslant M\xi_n + \varepsilon_n \ .$$

Переходя к пределу при $n \to \infty$, получаем утверждение леммы 2.

Определение 44 . *Пусть ξ — неотрицательная случайная величина, $\{\xi_n\}$ — последовательность дискретных случайных величин, обладающих свойствами 1 — 3 из леммы 1. Математическим ожиданием с.в. ξ называется число*

$$M\xi = \lim_n M\xi_n \ .$$

Лемма 2 гарантирует, что $M\xi$ не зависит от выбора аппроксимирующей последовательности $\{\xi_n\}$.

Пусть теперь ξ — произвольная случайная величина. Определим

$$\xi^+(\omega) = \begin{cases} \xi(\omega) & , \quad \xi(\omega) \geqslant 0 \ , \\ 0 & , \quad \xi(\omega) < 0 \ , \end{cases}$$

$$\xi^-(\omega) = \begin{cases} 0 & , \quad \xi(\omega) \geqslant 0 \ , \\ -\xi(\omega) & , \quad \xi(\omega) < 0 \ . \end{cases}$$

Из определения ξ^+ и ξ^- легко следует, что

1) $\xi^+ \geqslant 0, \ \xi^- \geqslant 0$,

2) $\xi(\omega) = \xi^+(\omega) - \xi^-(\omega)$.

Определение 45 . **Математическим ожиданием** *произвольной случайной величины ξ называется число*

$$M\xi = M\xi^+ - M\xi^- \ ,$$

если хотя бы одно из чисел в правой части этого равенства конечно.

Случайная величина ξ называется **интегрируемой**, *если $M\xi^+ < \infty$ и $M\xi^- < \infty$.*

Всюду далее мы при вычислении математических ожиданий будем предполагать, что соответствующие с.в. ξ являются интегрируемыми.

9.3 Математическое ожидание как интеграл Римана–Стилтьеса

Выше мы уже отмечали, что единственной объективной характеристикой с.в. ξ, которую можно восстановить однозначно по результатам эксперимента (по–крайней мере, в принципе), является ее распределение. Отсюда следует, что все другие характеристики должны выписываться с помощью распределения. Ниже мы покажем, как это сделать для математического ожидания. Для этого нам потребуется одно понятие из математического анализа, а именно интеграл Римана — Стилтьеса.

Пусть $F(x)$ — неубывающая ограниченная функция, $f(x)$ — произвольная борелевская функция, которые заданы на интервале $[a, b)$. Рассмотрим разбиение $a = t_0 < t_1 < \ldots < t_k < t_{k+1} < \ldots < t_n = b$ и выберем по точке $s_k \in [t_k, t_{k+1})$. Образуем интегральную сумму

$$S_n = \sum_{k=0}^{n-1} f(s_k)[F(t_{k+1}) - F(t_k)] \ .$$

Определение 46 . *Функция $f(x)$ называется* **интегрируемой в смысле Римана–Стилтьеса** *относительно функции $F(x)$ на $[a, b)$, если существует предел интегральных сумм S_n при*
$\max\limits_k |t_{k+1} - t_k| \to 0$ *и этот предел не зависит от способа измельчения интервала $[a, b)$ и выбора точек s_k.*

Полученный предел называется **интегралом Римана–Стилтьеса** *функции $f(x)$ относительно функции $F(x)$ и обозначается*

$$\int\limits_a^b f(x)\, dF(x).$$

Замечания. 1. Если $F(x) = x$, то мы получаем определение обычного интеграла Римана.

2. Если интервал бесконечный, то сначала интеграл определяют на конечном интервале, а затем переходят к пределу, когда один или оба конца интервала уходят в бесконечность.

3. Достаточным условием интегрируемости по Риману–Стилтьесу на конечном интервале является непрерывность функции $f(x)$ на замкнутом отрезке $[a, b]$.

4. Если существует $\rho_\xi(x) = F^{'}(x)$, то $\int\limits_a^b f(x)\, dF(x) = \int\limits_a^b f(x)\rho_\xi(x)\, dx$, где последний интеграл нужно понимать как интеграл Римана.

Применим эту конструкцию к нашей ситуации, когда мы вычисляем математическое ожидание. Пусть ξ — неотрицательная случайная величина. Напомним, что аппроксимирующая ее последовательность дискретных с.в. ξ_n строится по правилу:

$$\xi_n(\omega) = \frac{k}{2^n}\ , \quad \frac{k}{2^n} \leqslant \xi(\omega) < \frac{k+1}{2^n}.$$

Тогда

$$M\xi = \lim_n M\xi_n = \lim_n \sum_{k=0}^{\infty} \frac{k}{2^n}\left[F_\xi(\frac{k+1}{2^n}) - F_\xi(\frac{k}{2^n})\right].$$

Сравнивая последнее равенство с определением интеграла Римана–Стилтьеса для $f(x) = x$ и $F(x) = F_\xi(x)$, получаем

$$M\xi = \int\limits_0^\infty x\, dF_\xi(x).$$

Разлагая ξ в разность положительной и отрицательной частей, аналогично получаем в общем случае

$$M\xi = \int\limits_{-\infty}^{\infty} x\, dF_\xi(x). \tag{9.1}$$

Приведем без доказательства следующий полезный результат.

Предложение 15 . *Если с.в.* ξ *имеет функцию распределения* $F_\xi(x)$, $y = g(x)$ *— борелевская функция такая, что с.в.* $\eta = g(\xi)$ *— интегрируема, то*

$$M\eta = \int\limits_{-\infty}^{\infty} g(x)\, dF_\xi(x). \tag{9.2}$$

Если ξ — дискретная с.в. с множеством значений $X = \{x_1, x_2, \ldots\}$ и $p_n = P(\xi = x_n)$, то

$$M\xi = \sum_n x_n p_n, \tag{9.3}$$

и

$$M\eta = \sum_n g(x_n) p_n. \tag{9.4}$$

Если ξ имеет плотность распределения $\rho_\xi(x)$, то

$$M\xi = \int\limits_{-\infty}^{\infty} x\rho_\xi(x)\,dx. \tag{9.5}$$

и

$$M\eta = \int\limits_{-\infty}^{\infty} g(x)\rho_\xi(x)\,dx. \tag{9.6}$$

9.4 Свойства математических ожиданий

Всюду в этом разделе предполагается, что все написанные математические ожидания конечны.

1. Если $P(\xi = C) = 1$, то $M\xi = C$.

2. $M(C\xi) = C \cdot M\xi$.

3. $M(\xi_1 + \xi_2) = M\xi_1 + M\xi_2$.

4. Если $\xi \geqslant 0$, то $M\xi \geqslant 0$, причем $M\xi = 0$ тогда и только тогда, когда $P(\xi = 0) = 1$.

Эти четыре свойства получаются предельным переходом из аналогичных свойств математических ожиданий дискретных случайных величин.

5. Если $\xi_1 \geqslant \xi_2$, то $M\xi_1 \geqslant M\xi_2$.

Это свойство есть прямое следствие свойств 2 — 4.

6. $|M\xi| \leqslant M|\xi|$.

Доказательство. По определению абсолютной величины вещественного числа имеем

$$-|\xi| \leqslant \xi \leqslant |\xi| \; .$$

Далее применяем свойство 5.

7. Если ξ — индикатор некоторого случайного события A, то $M\xi = P(A)$.

Доказательство. По определению индикатора события с.в. ξ принимает два значения: 1 — с вероятностью $P(A)$ и 0 — с вероятностью

$1-P(A)$. Тогда по определению математического ожидания для дискретной случайной величины имеем

$$M\xi = 1 \cdot P(A) + 0 \cdot (1 - P(A)) = P(A) \ .$$

8. Если ξ_1 и ξ_2 — независимы, то $M(\xi_1 \xi_2) = M\xi_1 \cdot M\xi_2$.

Доказательство будет разбито на несколько шагов.

a) Если ξ_1 и ξ_2 — дискретные случайные величины, то это свойство уже доказано ранее.

b) Пусть ξ_1 и ξ_2 — неотрицательные интегрируемые случайные величины. Построим две последовательности $\left\{\xi_1^{(n)}\right\}$ и $\left\{\xi_2^{(n)}\right\}$ дискретных случайных величин, обладающих свойствами из леммы 1. Из доказательства этой леммы мы знаем, что $\xi_1^{(n)}$ и $\xi_2^{(n)}$ можно построить как функции от ξ_1 и ξ_2. Отсюда следует, что они независимы. Предположим, что

$$|\xi_1^{(n)} - \xi_1| \leqslant \frac{1}{n} \ , \ |\xi_2^{(n)} - \xi_2| \leqslant \frac{1}{n} \ .$$

По определению математического ожидания неотрицательной случайной величины

$$M\xi_1 = \lim_n M\xi_1^{(n)} \ , \ M\xi_2 = \lim_n M\xi_2^{(n)} \ .$$

Используя все указанное выше, получаем

$$\begin{gathered} \bigl|M(\xi_1\xi_2) - M\xi_1 \cdot M\xi_2\bigr| = \\ = \bigl|M(\xi_1\xi_2) - M(\xi_1^{(n)}\xi_2) + M(\xi_1^{(n)}\xi_2) - \\ - M(\xi_1^{(n)}\xi_2^{(n)}) + M(\xi_1^{(n)}\xi_2^{(n)}) - M\xi_1 \cdot M\xi_2\bigr| \\ \leqslant M\left(\bigl|\xi_1 - \xi_1^{(n)}\bigr|\bigl|\xi_2\bigr|\right) + M\left(\bigl|\xi_1^{(n)}\bigr|\bigl|\xi_2 - \xi_2^{(n)}\bigr|\right) + \\ + \bigl|M\xi_1^{(n)} \cdot M\xi_2^{(n)} - M(\xi_1) \cdot M(\xi_2)\bigr| \\ \leqslant \frac{1}{n} M\left(\bigl|\xi_2\bigr|\right) + \frac{1}{n} M\left(\bigl|\xi_1\bigr|\right) + \bigl|M\xi_1^{(n)} \cdot M\xi_2^{(n)} - M(\xi_1) \cdot M(\xi_2)\bigr| \to 0 \ , \end{gathered}$$

при $n \to \infty$.

c) Произвольные случайные величины ξ_1 и ξ_2 представим в виде разностей $\xi_1 = \xi_1^+ - \xi_1^-$ и $\xi_2 = \xi_2^+ - \xi_2^-$ их положительных и отрицательных частей. В силу независимости ξ_1 и ξ_2 пары (ξ_1^+, ξ_1^-) и (ξ_2^+, ξ_2^-) также будут независимы. Используя пункт b), получаем

$$M(\xi_1\xi_2) = M\left[(\xi_1^+ - \xi_1^-)(\xi_2^+ - \xi_2^-)\right] =$$
$$= M\big(\xi_1^+\xi_2^+ - \xi_1^+\xi_2^- - \xi_1^-\xi_2^+ + \xi_1^-\xi_2^-\big) =$$
$$= M\big(\xi_1^+\big)M(\xi_2^+\big) - M\big(\xi_1^+\big)M(\xi_2^-\big) - M\big(\xi_1^-\big)M\big(\xi_2^+\big) + M(\xi_1^-)M\big(\xi_2^-\big) =$$
$$= M\big(\xi_1^+ - \xi_1^-\big)M\big(\xi_2^+ - \xi_2^-\big) = M\xi_1 \cdot M\xi_2 \ .$$

Далее мы докажем несколько неравенств, которые будут часто использоваться в нашем курсе.

9. **Неравенство Маркова**. Если $\xi \geqslant 0$, то $\forall \varepsilon > 0$

$$P(\xi \geqslant \varepsilon) \leqslant \frac{M\xi}{\varepsilon} \ .$$

Доказательство. Рассмотрим случайное событие $A = (\xi \geqslant \varepsilon)$ и его индикатор I_A. Нетрудно проверить (задача!), что имеет место следующее неравенство:

$$\varepsilon \cdot I_A(\omega) \leqslant \xi(\omega) \ .$$

Тогда, применяя свойства 2,5 и 7, получаем

$$M(\varepsilon I_A) = \varepsilon P(A) \leqslant M\xi \ .$$

10. **Неравенство Коши–Буняковского**. Если ξ_1 и ξ_2 — случайные величины, то

$$\big(M(\xi_1\xi_2)\big)^2 \leqslant M\big(\xi_1^2\big) \cdot M\big(\xi_2^2\big),$$

причем знак равенства достигается тогда и только тогда, когда существует некоторая константа C, такая, что $\xi_1 = C\xi_2$ или $\xi_2 = C\xi_1$.

Доказательство. Возьмем $\lambda \in R^1$ и образуем случайную величину $\eta = (\xi_1 + \lambda\xi_2)^2$. Так как $\eta \geqslant 0$, то

$$M\eta = M\big(\xi_1^2\big) + 2M\big(\xi_1\xi_2\big)\lambda + M\big(\xi_2^2\big)\lambda^2 \geqslant 0.$$

Мы имеем полином второго порядка от λ, который принимает только неотрицательные значения. В этом случае его дискриминант должен быть неположительным:

$$D = \big(M(\xi_1\xi_2)\big)^2 - M\big(\xi_1^2\big)M\big(\xi_2^2\big) \leqslant 0 \ .$$

Если в последнем неравенстве мы имеем знак равенства, т. е. $D = 0$, то существует некоторое $\lambda_0 \in R^1$, такое, что

$$M\eta = M(\xi_1 + \lambda_0\xi_2)^2 = 0 \ .$$

Но тогда (смотри свойство 4) мы имеем

$$\xi_1 + \lambda_0\xi_2 = 0 \ \ \xi_1 = -\lambda_0\xi_2 \ .$$

11. **Неравенство Иенсена**. Пусть $y = g(x)$ — выпуклая вниз борелевская функция, ξ — некоторая случайная величина. Тогда

$$g(M\xi) \leqslant Mg(\xi).$$

Доказательство. Из определения выпуклости следует, что $\forall x_0 \in R^1$ существует $\lambda(x_0) \in R^1$, такое, что $\forall x \in R^1$

$$g(x) - g(x_0) \geqslant \lambda(x_0)(x - x_0).$$

Мы построили так называемую опорную прямую. Подставим в это неравенство ξ и $M\xi$ вместо x и x_0:

$$g(\xi) - g(M\xi) \geqslant \lambda(M\xi)(\xi - M\xi) \ .$$

В силу свойства (5) получаем

$$Mg(\xi) - g(M\xi) \geqslant \lambda(M\xi)(M\xi - M\xi) = 0 \ .$$

12. **Неравенство Ляпунова**. Если ξ — некоторая случайная величина, $0 < s < t < \infty$, то

$$(M|\xi|^s)^{1/s} \leqslant (M|\xi|^t)^{1/t} \ .$$

Доказательство. Пусть $r = t/s > 1, \ g(x) = |x|^r, \ \eta = |\xi|^s$. Тогда $g(x)$ — выпуклая вниз функция. В силу неравенства Иенсена

$$g(M\eta) = \big|M|\xi|^s\big|^{t/s} \leqslant M\big(g(\eta)\big) = M\big(|\xi|^{s\cdot\frac{t}{s}}\big) = M\big(|\xi|^t\big) \ .$$

Отсюда получаем

$$\big(M|\xi|^s\big)^{1/s} \leqslant \big(M|\xi|^t\big)^{1/t} \ .$$

Глава 10

Числовые характеристики случайных величин и векторов

Как мы отмечали ранее, распределение случайной величины является ее полной характеристикой. Но эта характеристика является достаточно сложной и в реальных задачах редко бывает известна достоверно. Поэтому пытаются описать распределение с помощью конечного числа сравнительно простых числовых характеристик. Одной из них является математическое ожидание. В этом параграфе мы вводим некоторые новые характеристики.

10.1 Моменты

Определение 47 . Моментом порядка k относительно точки *a случайной величины ξ называется число*

$$M(\xi - a)^k \ .$$

Если число $a = 0$, то момент $\alpha_k = M(\xi^k)$ называется **начальным**.

Если число $a = M\xi$, то момент $\mu_k = M(\xi - M\xi)^k$ называется **центральным**.

Центральный момент второго порядка

$$\mu_2 = D(\xi) := M(\xi - M\xi)^2$$

называется **дисперсией**. Эта характеристика будет изучаться подробнее ниже.

Между центральными и начальными моментами существует простая связь:

$$\mu_k = \sum_{m=0}^{k} (-1)^m C_k^m \alpha_1^m \alpha_{k-m} \ .$$

Определяются также **абсолютные** моменты:

$$\beta_k = M\big(|\xi|^k\big) \ .$$

Предполагается, что задание достаточного числа моментов определяет распределение с нужной степенью точности. В связи с этой идеей важное значение имеет следующая

Проблема моментов. Пусть мы имеем некоторую последовательность чисел $\alpha_0, \alpha_1, \alpha_2, \ldots$.

1. При каких условиях последовательность $\{\alpha_k\}$ является последовательностью моментов какой-либо случайной величины?

2. При каких условиях последовательность моментов $\{\alpha_k\}$ однозначно определяет распределение?

На первый вопрос отвечает следующая

Теорема 4 . *Последовательность чисел $\{\alpha_k\}$ является последовательностью моментов некоторой случайной величины тогда и только тогда, когда $\alpha_0 = 1$ и для любого натурального $N \geqslant 1$ и любых комплекных чисел $c_0, \ldots, c_N$*

$$\sum_{m,n=0}^{N} \alpha_{m+n} c_m \overline{c_n} \geqslant 0 \ ,$$

т.е. матрица $(b_{mn} = \alpha_{m+n})$ является положительно определенной.

Второй вопрос является более сложным. Мы приводим только одно достаточное условие. Но прежде нам необходимо ввести одно вспомогательное понятие.

Определение 48 . **Производящей функцией моментов** *с.в. ξ (или ее распределения) называется функция*

$$\psi_\xi(t) = M\big(e^{t\xi}\big) \ .$$

Свойства производящих функций.

1. $\psi_\xi(0) = 1$.

2. Если с.в. ξ_1 и ξ_2 независимы, то

$$\psi_{\xi_1+\xi_2}(t) = \psi_{\xi_1}(t) \cdot \psi_{\xi_2}(t)\ .$$

3. Если для некоторого $t_0 > 0$ производящая функция моментов $\psi(t)$ существует для всех $t:\ |t| < t_0$, то $\forall k \geqslant 1$

$$\exists \frac{d^k}{dt^k}\psi(t)\,\Big|_{t=0} = M(\xi^k)\ \ .$$

Эти и другие свойства производящих функций моментов будут доказаны позднее.

Пример. Пусть $S_n = \varepsilon_1 + \ldots + \varepsilon_n$ — число успехов в схеме Бернулли с параметрами n и p, где $\varepsilon_1, \ldots, \varepsilon_n$ — н.о.р. и имеют распределение Бернулли с параметром p. Нетрудно вычислить, что

$$\psi_{\varepsilon_k}(t) = e^{t\cdot 1}p + e^{t\cdot 0}(1-p) = 1 + p(e^t - 1)\ ,$$

$$\psi(t) = \psi_{S_n}(t) = \prod_{k=1}^{n} \psi_{\varepsilon_k}(t) = \big[1 + p(e^t - 1)\big]^n\ .$$

Дифференцируя, получаем

$$\frac{d}{dt}\psi(t) = n\big[1 + p(e^t-1)\big]^{n-1} pe^t\ ,$$

$$\frac{d^2}{dt^2}\psi(t) = n(n-1)\big[1 + p(e^t-1)\big]^{n-2} p^2 e^{2t} + n\big[1 + p(e^t-1)\big]^{n-1} pe^t\ .$$

Применяя свойство 3, получаем

$$M(S_n) = np,\ M(S_n^2) = n(n-1)p^2 + np\ .$$

Теорема 5 . *Пусть $\{\alpha_n\}$ — последовательность моментов с.в. ξ. Если существует $r > 0$, такое, что ряд*

$$\sum_{n=0}^{\infty} \frac{\alpha_n}{n!} r^n$$

сходится абсолютно (что эквивалентно условию $\psi_\xi(r) = M\big(e^{r\xi}\big) < \infty$, и в этом случае обе величины совпадают), то распределение с.в. ξ однозначно определяется своими моментами.

Смысл этого утверждения в том, что если нам удалось каким-либо образом вычислить моменты распределения, то мы можем утверждать, что это то или иное из известных нам распределений.

Пример. Пусть ξ имеет стандартное нормальное распределение. Тогда

$$\psi_\xi(r) = M(e^{r\xi}) = \int\limits_{-\infty}^{\infty} e^{rx}\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}\,dx = \int\limits_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}}e^{-\frac{(x-r)^2}{2}}\,dx e^{\frac{r^2}{2}} = e^{\frac{r^2}{2}} < \infty\,.$$

Таким образом, нормальное распределение однозначно определяется своими моментами.

Пусть мы теперь имеем случайный вектор $\xi = (\xi_1, \ldots, \xi_d)$.

Определение 49 . **Моментом порядка** k *относительно точки* $a \in R^d$ *случайного вектора* ξ *называется число*

$$M\Big[(\xi_1 - a_1)^{k_1} \cdot \ldots \cdot (\xi_d - a_d)^{k_d}\Big]\ ,$$

где $k_1 + \ldots + k_d = k$.

Если $a_1 = \ldots = a_d = 0$, *то момент называется* **начальным**. *Если* $a_1 = M\xi_1, \ldots, a_d = M\xi_d$, *то момент называется* **центральным**.

В отличие от одномерного случая, теперь могут существовать несколько разных моментов одного порядка.

Для $k = 1$ начальные моменты совпадают с математическими ожиданиями координат: $M\xi_1, \ldots, M\xi_d$, а все центральные моменты равны нулю.

Для $k = 2$ имеем следующие начальные моменты:

$$M\big(\xi_1^2\big), \ldots, M\big(\xi_d^2\big),\ M\big(\xi_i\xi_j\big)\ ,\ i, j = \overline{1, d},\ i \neq j.$$

Чаще используют центральные моменты:

$$\sigma_{11} = D(\xi_1) = M\big[(\xi_1 - M\xi_1)^2\big], \ldots, \sigma_{dd} = D(\xi_d) = M\big[(\xi_d - M\xi_d)^2\big]$$

— **дисперсии** и

$$\sigma_{ij} = cov(\xi_i, \xi_j) = M\big[(\xi_i - M\xi_i)(\xi_j - M\xi_j)\big]$$

— **ковариации** координат случайного вектора ξ.

Обычно все центральные моменты второго порядка собирают в одну матрицу

$$\Sigma = (\sigma_{ij}) \ ,$$

которая называется **матрицей ковариаций** или **ковариационной матрицей** случайного вектора ξ.

10.2 Дисперсия и ее свойства

Определение 50 . **Дисперсией** *случайной величины* ξ *называется число*

$$D(\xi) = M\big[(\xi - M\xi)^2\big] \ .$$

Если $F(x)$ — функция распределения с.в. ξ, то

$$D(\xi) = \int\limits_{-\infty}^{\infty} (x - M\xi)^2 \, dF(x) \ .$$

Для дискретной случайной величины

$$D(\xi) = \sum_n (x_n - M\xi)^2 \cdot p_n \ .$$

В случае абсолютно непрерывного распределения с плотностью $\rho(x)$ имеем

$$D(\xi) = \int\limits_{-\infty}^{\infty} (x - M\xi)^2 \rho(x) \, dx \ .$$

Перечислим и докажем основные свойства дисперсии.

1. $D(\xi) \geqslant 0$, $D(\xi) = 0$ тогда и только тогда, когда $\xi = C$ п.н.

Доказательство. С.в. $(\xi - M\xi)^2 \geq 0$. Отсюда в силу свойства 4 для математических ожиданий случайных величин получаем нужное утверждение.

2. $D(\xi + C) = D(\xi)$.

3. $D(C\xi) = C^2 D(\xi)$.

Доказательство. По определению дисперсии

$$D(\xi + C) = M\Big[\big((\xi + C) - M(\xi + C)\big)^2\Big] = M\Big[(\xi + C - M\xi - C)^2\Big] =$$

$$M\Big[(\xi - M\xi)^2\Big] = D(\xi) \ .$$

Свойство 3 доказывается аналогично.

4. $D(\xi) = M(\xi^2) - (M\xi)^2$.

Доказательство. Вновь по определению дисперсии и используя свойства математических ожиданий, получаем

$$D(\xi) = M\Big[(\xi - M\xi)^2\Big] = M\Big[\xi^2 - 2\xi \cdot M\xi + (M\xi)^2\Big] = M(\xi^2) - (M\xi)^2 \ .$$

Далее нам потребуются некоторые свойства ковариации случайных величин. Сначала напомним ее определение.

Определение 51 . *Ковариацией случайных величин ξ_1 и ξ_2 называется число*

$$cov(\xi_1, \xi_2) := M\Big[(\xi_1 - M(\xi_1))(\xi_2 - M(\xi_2))\Big] \ .$$

Нетрудно доказать (задача!), что справедливы следующие свойства ковариаций.

1) $cov(\xi, \xi) = D(\xi) \geqslant 0$,

2) $cov(\xi_1, \xi_2) = cov(\xi_2, \xi_1)$,

3) $cov\big(c_1\xi_1 + c_2\xi_2, \xi_3\big) = c_1 \cdot cov(\xi_1, \xi_3) + c_2 \cdot cov(\xi_2, \xi_3)$,

4) $\big|cov(\xi_1, \xi_2)\big|^2 \leqslant D(\xi_1) \cdot D(\xi_2)$ (неравенство Коши — Буняковского!),

5) $cov(\xi_1, \xi_2) = M(\xi_1 \cdot \xi_2) - M(\xi_1) \cdot M(\xi_2)$,

6) для $\varepsilon > 0$

$$P\big(|\xi - M(\xi)\big| > \varepsilon) \leqslant \frac{D(\xi)}{\varepsilon^2} \ .$$

Последнее свойство будет часто использоваться далее и носит название **неравенство Чебышева**. Его легко вывести из неравенства Маркова (смотри выше).

Определение 52 . *С.в. ξ_1 и ξ_2 называются* **некоррелированными**, *если*

$$cov(\xi_1, \xi_2) = M\big[(\xi_1 - M\xi_1)(\xi_2 - M\xi_2)\big] = 0 \ .$$

Лемма 5 . *Если с.в. ξ_1 и ξ_2 независимы, то они и некоррелированы.*

Доказательство. Если ξ_1 и ξ_2 независимы, то независимы и с.в. $\xi_1 - M\xi_1$ и $\xi_2 - M\xi_2$. Тогда

$$M\big[(\xi_1 - M\xi_1)(\xi_2 - M\xi_2)\big] = M(\xi_1 - M\xi_1) \cdot M(\xi_2 - M\xi_2) =$$

$$(M\xi_1 - M\xi_1) \cdot M(\xi_2 - M\xi_2) = 0 \ .$$

Приведем пример, который показывает, что обратное неверно.

Пусть случайная величина ξ имеет равномерное распределение на $[-1, 1]$. Положим $\xi_1 = \xi$, $\xi_2 = \xi^2$. В этом случае $\xi_2 = \xi_1^2$, т.е. эти величины функционально связаны и поэтому зависимы. Но

$$cov(\xi_1, \xi_2) = M\big[(\xi - M\xi)(\xi^2 - M(\xi^2))\big] = M(\xi^3) - M(\xi)M(\xi^2) = 0 \ ,$$

т.к. в силу симметрии распределения $M(\xi) = M(\xi^3) = 0$.

5. $D(\xi_1 + \xi_2) = D(\xi_1) + D(\xi_2) + 2cov(\xi_1, \xi_2)$.

Если ξ_1 и ξ_2 некоррелированы (в частности, независимы), то

$$D(\xi_1 + \xi_2) = D(\xi_1) + D(\xi_2) \ .$$

Предлагается доказать это свойство самостоятельно.

Замечания. 1. $D(\xi)$ оценивает степень рассеивания с.в. ξ около своего «центра» $M\xi$.

2. Часто полезно (вспомним случай нормального распределения) перейти к так называемой нормированной случайной величине

$$\xi_0 = \frac{\xi - M\xi}{\sqrt{D(\xi)}} .$$

Используя свойства математических ожиданий и дисперсий, легко показать (задача!), что $M\xi_0 = 0,\ D(\xi_0) = 1$.

3. Некоррелированность является ослабленным вариантом понятия независимости. Геометрическая интерпретация некоррелированности будет дана ниже.

В заключение этого раздела выпишем значения математических ожиданий и дисперсий для стандартных распределений.

1. Вырожденное распределение в точке a.

$$M\xi = a,\ D(\xi) = 0 .$$

2. Распределение Бернулли с параметром p.

$$M\xi = p\ ,\ D(\xi) = p(1-p) .$$

3. Биномиальное распределение с параметрами n и p.

$$M\xi = np\ ,\ D(\xi) = np(1-p) .$$

4. Геометрическое распределение с параметром p.

$$M\xi = \frac{1-p}{p},\ D(\xi) = \frac{1-p}{p^2} .$$

5. Отрицательное биномиальное распределение с параметрами . r и p.

$$M\xi = \frac{1-p}{p}r\ ,\ D(\xi) = \frac{1-p}{p^2}r .$$

6. Распределение Пуассона с параметром λ.

$$M\xi = \lambda,\ D(\xi) = \lambda .$$

7. Равномерное распределение на отрезке $[a,b]$.

$$M\xi = \frac{b-a}{2},\ D(\xi) = \frac{(b-a)^2}{12}\ .$$

8. Показательное распределение с параметром λ.

$$M\xi = \frac{1}{\lambda}\ ,\ D(\xi) = \frac{1}{\lambda^2}\ .$$

9. Гамма-распределение с параметрами α и β.

$$M\xi = \frac{\alpha}{\beta}\ ,\ D(\xi) = \frac{\alpha}{\beta^2}\ .$$

10. Нормальное распределение с параметрами a и σ^2.

$$M\xi = a\ ,\ D(\xi) = \sigma^2\ .$$

Эти результаты будут получены на практических занятиях.

10.3 Коэффициент корреляции и его свойства

Выше мы уже отмечали, что некоррелированность, т.е. равенство нулю ковариации случайных величин, является ослабленным вариантом независимости. Интуитивно понятно, что если изменить единицу измерения и начало отсчета, то степень зависимости между случайными величинами не должна измениться. Таким образом, для описания силы связи между случайными величинами естественно использовать некоторый коэффициент, который был бы пропорционален ковариации, но не менялся бы при линейных преобразованиях.

Определение 53 . **Коэффициентом корреляции** *случайных величин* ξ_1 *и* ξ_2 *называется число*

$$\rho(\xi_1,\xi_2) = \frac{cov(\xi_1,\xi_2)}{\sqrt{D(\xi_1)D(\xi_2)}}\ .$$

Это определение имеет смысл, если случайные величины ξ_1 и ξ_2 невырождены. Основные свойства коэффициента корреляции собраны в следующей теореме.

Теорема 6 . *Пусть ξ_1 и ξ_2 — невырожденные случайные величины. Тогда справедливы следующие свойства:*

1. $|\rho(\xi_1, \xi_2)| \leqslant 1$.

2. Если ξ_1 и ξ_2 некоррелированы, то $\rho(\xi_1, \xi_2) = 0$.

3. Пусть $\eta_1 = a_1\xi_1 + b_1$, $\eta_2 = a_2\xi_2 + b_2$, $a_1 \cdot a_2 \neq 0$. Тогда

$$\rho(\eta_1, \eta_2) = \rho(\xi_1, \xi_2) \cdot sign(a_1 a_2) \ .$$

4. Если $\rho(\xi_1, \xi_2) = 1$, то существуют $a > 0$, $b \in R^1$, такие, что $\xi_1 = a\xi_2 + b$.

Если $\rho(\xi_1, \xi_2) = -1$, то существуют $a < 0$, $b \in R^1$, такие, что $\xi_1 = a\xi_2 + b$.

Доказательство. Начнем с доказательства свойства 3. Из определения с.в. η_1 и η_2, используя свойства математических ожиданий и дисперсий, легко получаем

$$M\eta_1 = a_1 M\xi_1 + b_1 \ , \ M\eta_2 = a_2 M\xi_2 + b_2 \ , D(\eta_1) = a_1^2 D(\xi_1) \ ,$$

$$D(\eta_2) = a_2^2 D(\xi_2) \ , \ cov(\eta_1, \eta_2) = a_1 a_2 cov(\xi_1, \xi_2) \ .$$

Отсюда следует, что

$$\rho(\eta_1, \eta_2) := \frac{cov(\eta_1, \eta_2)}{\sqrt{D(\eta_1) D(\eta_2)}} =$$

$$\frac{a_1 a_2 cov(\xi_1, \xi_2)}{|a_1 a_2| \sqrt{D(\xi_1) D(\xi_2)}} = \rho(\xi_1, \xi_2) \cdot sign(a_1 a_2) \ .$$

Для доказательства свойства 1 введем с.в.

$$\eta_1 = \frac{\xi_1 - M\xi_1}{\sqrt{D(\xi_1)}} \ , \ \eta_2 = \frac{\xi_2 - M\xi_2}{\sqrt{D(\xi_2)}} \ .$$

В силу свойства 3 получаем

$$\left|\rho(\xi_1, \xi_2)\right|^2 = \left|\rho(\eta_1, \eta_2)\right|^2 = \left|M(\eta_1 \cdot \eta_2)\right|^2 \leqslant$$

$$M(\eta_1^2) \cdot M(\eta_2^2) = D(\eta_1)D(\eta_2) = 1 \ .$$

На последнем шаге мы использовали неравенство Коши–Буняковского и замечание 2 о нормированной с.в.

Свойство 2 есть тривиальное следствие из определений некоррелированности и коэффициента корреляции.

Для доказательства свойства 4 вновь введем нормированные с.в. η_1 и η_2. В неравенстве Коши–Буняковского достигается знак равенства тогда и только тогда, когда существует такая константа c, что $\eta_1 = c\eta_2$. Возвращаясь к исходным величинам ξ_1 и ξ_2, мы получаем $\xi_1 = a\xi_2 + b$, где

$$a = c\sqrt{\frac{D(\xi_1)}{D(\xi_2}} \quad b = M\xi_1 - c\sqrt{\frac{D(\xi_1)}{D(\xi_2}}M\xi_2 \ .$$

Используя свойство 3, легко показать, что $a > 0$ для $\rho = 1$ и $a < 0$ для $\rho = -1$.

Замечания. 1. Как мы увидим позднее, коэффициент корреляции является мерой **линейной** связи между случайными величинами.

2. На практических занятиях будет показано, что для с.в. ξ_1 и ξ_2, которые имеют двумерное нормальное распределение, параметр ρ равен коэффициенту корреляции для этих величин.

10.4 Характеристики расположения и формы распределения

В реальных задачах бывает полезно разметить прямую точками, которые делят ее на несколько отрезков, вероятности попадания в которые имеют определенные значения.

Определение 54 . **Квантилью порядка** *p, $0 < p < 1$, случайной величины ξ (или ее распределения) называется число $x_p \in R^1$:*

$$P(\xi \leqslant x_p) \geqslant p \ , \ P(\xi \geqslant x_p) \geqslant 1 - p \ .$$

Замечание. Если $P(\xi = x) = 0$ для всех x, то

$$P(\xi < x_p) = F_\xi(x_p) = p\ .$$

Примеры. 1. Величины $x_{1/4}, x_{1/2}, x_{3/4}$ называются **квартилями**. Число $x_{1/2}$ называется **медианой**. Оно определяет «центр» распределения. Разность $x_{3/4} - x_{1/4}$ называется **семиинтерквартильной широтой**. Она характеризует степень рассеивания распределения около центра.

2. Числа $x_{0,1}, x_{0,2}, \ldots, x_{0,9}$ называются **децилями**.

3. Наиболее часто используются на практике **процентили**, т.е. величины вида $x_{0,01}, \ldots, x_{0,99}$.

Часто хотелось бы знать, какие значения более вероятны, а какие менее. В связи с этим является полезным следующее понятие.

Определение 55 . **Модой дискретного распределения** $(X, \{p_n\})$ $(x_1 < x_2 < \ldots)$ *называется такое значение* x_n, *что* $p_n \geqslant p_{n-1}$, $p_n \geqslant p_{n+1}$.

Модой абсолютно непрерывного распределения *называется любая точка максимума плотности* $\rho_\xi(x)$.

Распределение называется **унимодальным**, *если оно имеет одну моду, и* **полимодальным** *— в противном случае.*

Примеры. 1. Нормальное распределение с параметрами a и σ^2 является унимодальным, мода которого совпадает с математическим ожиданием и равна a.

2. Биномиальное распределение с параметрами n и p является унимодальным, если число np — дробное, и **бимодальным**, если это число — целое (смотри выше).

Определение 56 . **Коэффициентом асимметрии** *называется число*

$$\frac{M(\xi - M\xi)^3}{\sigma^3}\ ,$$

где $\sigma^2 = D(\xi)$.

Коэффициентом эксцесса *называется число*

$$\frac{M(\xi - M\xi)^4}{\sigma^4} - 3 \ .$$

Эти характеристики применяются для грубого сравнения данного распределения с нормальным, для которого оба коэффициента равны нулю. Асимметрия измеряет величину «скошенности» распределения в ту или иную сторону. Эксцесс измеряет степень «островершинности» распределения. Если эксцесс положительный, то распределение является более плоским (размазанным) по сравнению с нормальным, в противном случае оно является более островершинным.

Глава 11

Гильбертово пространство случайных величин

В этом разделе мы выделим некоторое подмножество случайных величин и покажем, как в нем можно определить геометрическую структуру.

Рассмотрим множество L_2, состоящее из случайных величин ξ, таких, что $M\big(|\xi|^2\big) < \infty$. Используя неравенство Коши — Буняковского, нетрудно показать, что L_2 — это линейное пространство (задача!). Для каждой пары $\xi, \eta \in L_2$ определим число

$$(\xi, \eta) := M(\xi \cdot \eta) \ .$$

Будем говорить, что случайные величины ξ и η совпадают **почти наверное** (п.н.), если $P(\xi \neq \eta) = 0$. В дальнейшем мы не будем различать такие случайные величины и объединим их в один класс. Таким образом, мы рассматриваем L_2 как совокупность классов эквивалентности.

Задача. Введенный выше функционал обладает следующими свойствами:

a) $(\xi, \xi) \geqslant 0, \ (\xi, \xi) = 0 \Leftrightarrow \xi = 0$ п.н.,

b) $\forall \xi, \eta \in L_2, \ (\xi, \eta) = (\eta, \xi)$,

c) $\forall \xi_1, \xi_2, \eta \in L_2, \ c_1, \ c_2 \in R^1 \ (c_1\xi_1 + c_2\xi_2, \eta) = c_1(\xi_1, \eta) + c_2(\xi_2, \eta)$.

Функционал, который приписывает каждой паре элементов из линейного пространства некоторое число так, что выполнены свойства а) — с), называется **скалярным произведением**. Используя скалярное произведение, можно определить **норму** произвольного элемента $\xi \in L_2$:

$$\|\xi\| := \sqrt{(\xi, \xi)} = \left(M(|\xi|^2)\right)^{1/2} .$$

Норма элемента ξ определяет его длину. Будем говорить, что последовательность с.в. $\{\xi_n\}$ **сходится в среднем квадратическом** (с.к.) к с.в. ξ, если

$$\|\xi_n - \xi\|^2 = M\left(|\xi_n - \xi|^2\right) \to 0,\ n \to \infty .$$

В этом случае мы будем писать $\xi_n \overset{..}{\to} \xi$. Можно показать, что пространство L_2 полно относительно такой сходимости, т.е. всякая фундаментальная последовательность в L_2 имеет предел. Линейное пространство, в котором определено скалярное произведение и которое полно относительно порожденной этим скалярным произведением сходимости, называется **гильбертовым пространством**. Таким образом, мы показали, что L_2 — гильбертово пространство.

Выше мы отмечали, что норма вектора определяет его длину. Но в геометрии нам необходимо измерять не только длины, но и углы. Используя аналогию с понятием скалярного произведения в планиметрии, определим косинус угла α между элементами $\xi, \eta \in L_2$ по правилу

$$\cos\alpha = \frac{M(\xi\eta)}{\sqrt{M|\xi|^2 \cdot M|\eta|^2}} = \frac{(\xi, \eta)}{\|\xi\| \cdot \|\eta\|} .$$

Назовем ξ и η **ортогональными**, если $(\xi, \eta) = M(\xi\eta) = 0$.

Замечание. Пусть $M\xi = M\eta = 0$. Тогда имеет место следующее:

1) $M(\xi\eta) = cov(\xi, \eta) = 0$, т.е. понятие ортогональности совпадает с понятием некоррелированности;

2) $\cos\alpha = \rho(\xi, \eta)$, т.е. косинус угла между случайными величинами совпадает с коэффициентом корреляции.

Используя введенные выше понятия, мы решим очень важную с практической точки зрения задачу. Пусть $\mathcal{L} \subset L_2$ — некоторое линейное подпространство, которое замкнуто относительно сходимости в среднем квадратическом, а η — произвольный элемент из L_2.

Определение 57 . *Случайная величина* $\hat\eta \in L_2$ *называется* **наилучшим приближением с.в.** η **в пространстве** $\mathcal{L}$, *если*

1) $\hat\eta \in \mathcal{L}$,

2) $\|\eta-\hat\eta\|^2 = M|\eta-\hat\eta|^2 \leqslant M|\eta-\xi|^2 = \|\eta-\xi\|^2$, $\forall \xi \in \mathcal{L}$.

Лемма о перпендикуляре. *С.в.* $\hat\eta$ *является наилучшим приближением с.в.* η *в линейном пространстве* $\mathcal{L}$ *тогда и только тогда, когда*

1) $\hat\eta \in \mathcal{L}$,

2) $(\eta-\hat\eta,\xi) = M\big[(\eta-\hat\eta)\xi\big] = 0, \ \forall \xi \in \mathcal{L}$.

Доказательство. Пусть выполнено свойство 2 и $\xi \in \mathcal{L}$. Тогда

$$\|\eta-\xi\|^2 = M|\eta-\xi|^2 = M\big|(\eta-\hat\eta)+(\hat\eta-\xi)\big|^2 =$$
$$= M|\eta-\hat\eta|^2 + M|\hat\eta-\xi|^2 + 2M\big[(\eta-\hat\eta)(\hat\eta-\xi)\big] =$$
$$= M|\eta-\hat\eta|^2 + M|\hat\eta-\xi|^2 + 0 \geqslant M|\eta-\hat\eta|^2 \ .$$

Здесь мы использовали тот факт, что $\hat\eta-\xi \in \mathcal{L}$.

Обратно, пусть мы знаем, что $\hat\eta$ — наилучшее приближение для η в $\mathcal{L}$. Возьмем произвольные $\xi \in \mathcal{L}$ и $\lambda \in R^1$. В силу оптимальности $\hat\eta$ получаем

$$M|\eta-\hat\eta|^2 \le M|\eta-\hat\eta+\lambda\xi|^2 =$$
$$M|\eta-\hat\eta|^2 + 2M\big[(\eta-\hat\eta)\xi\big]\cdot\lambda + M|\xi|^2\cdot\lambda^2 \ .$$

Последнее выражение есть полином второго порядка от λ, у которого коэффициент при λ^2 положительный и который принимает минимальное значение при $\lambda = 0$. Из курса математического анализа мы знаем, что в этом случае коэффициент при λ, т.е. $M\big[(\eta-\hat\eta)\xi\big]$, равен нулю. А это и есть условие 2.

Дадим геометрическую интерпретацию полученному результату. Свойство 2 означает, что $\eta-\hat\eta$ является **перпендикуляром** к подпространству $\mathcal{L}$ (эквивалентно: $\hat\eta$ — **проекция** η на $\mathcal{L}$).

Иногда с.в. $\hat\eta$ называют **условным математическим ожиданием в широком смысле** с.в. η относительно подпространства $\mathcal{L}$. Смысл такого названия мы объясним позднее.

Рассмотрим несколько более общую ситуацию. Пусть $\mathcal{L} = \xi_0 + \mathcal{L}_0$, где $\xi_0 \in L_2$ — фиксированный элемент, а $\mathcal{L}_0$ — замкнутое линейное подпространство в L_2. Если $\xi_0 \neq 0$, то $\mathcal{L}$ не является линейным пространством и называется **гиперплоскостью**. Пусть далее, $\eta \in L_2$ — некоторая с.в. Определение наилучшего приближения $\hat{\eta}$ для с.в. η в гиперплоскости $\mathcal{L}$ дословно повторяет определение 1. Нетрудно доказать следующий результат.

Задача. $\hat{\eta}$ является наилучшим приближением для η в гиперплоскости $\mathcal{L}$, если

1) $\hat{\eta} \in \mathcal{L}$,

2) $(\eta - \hat{\eta}, \xi - \hat{\eta}) = M[(\eta - \hat{\eta}, \xi - \hat{\eta})] = 0 \ , \ \forall \xi \in \mathcal{L}$.

Применим эти результаты к следующей задаче. Пусть $\xi, \eta \in L_2$ — две случайные величины. Необходимо найти наилучшее линейное приближение с.в. η с помощью случайной величины ξ. В этом случае $\mathcal{L} = \{c_1 + c_2\xi \ , \ c_1, c_2 \in R^1\}$. Это двумерное линейное подпространство в L_2, базисом которого являются с.в. $\xi_1 = 1$ и $\xi_2 = \xi$. Наилучшее приближение должно иметь вид $\hat{\eta} = \hat{c}_1 + \hat{c}_2\xi$, т.е. мы должны найти $\hat{c}_1$ и $\hat{c}_2$. По лемме о перпендикуляре $(\eta - \hat{\eta}, \theta) = 0$, $\forall \theta \in \mathcal{L}$. В частности,

$$\big(\eta - \hat{\eta}, \xi_1\big) = M\big[(\eta - \hat{\eta}) \cdot 1\big] = M\eta - \hat{c}_1 - \hat{c}_2 M\xi = 0 \ ,$$

$$\big(\eta - \hat{\eta}, \xi_2\big) = M\big[(\eta - \hat{\eta}) \cdot \xi\big] = M(\eta\xi) - \hat{c}_1 M\xi - \hat{c}_2 M(\xi^2) = 0 \ .$$

Решая эти уравнения относительно $\hat{c}_1$ и $\hat{c}_2$, получим:

$$\hat{c}_2 = \frac{cov(\xi, \eta)}{D(\xi)} = \rho(\xi, \eta) \cdot \frac{\sigma_\eta}{\sigma_\xi} \ ,$$

$$\hat{c}_1 = M\eta - \rho(\xi, \eta) \cdot \frac{\sigma_\eta}{\sigma_\xi} \cdot M\xi \ ,$$

где $\sigma_\eta^2 = D(\eta)$, $\sigma_\eta^2 = D(\eta)$. Таким образом, окончательно получаем

$$\hat{\eta} = M\eta + \rho(\xi, \eta) \cdot \frac{\sigma_\eta}{\sigma_\xi} \cdot (\xi - M\xi) \ .$$

Глава 12

Условные распределения и условные математические ожидания

Во многих прикладных задачах мы встречаемся с такой ситуацией, когда часть компонент случайного вектора фиксирована и необходимо найти распределение остальных компонент. Например, мы фиксируем цену некоторого товара и интересуемся распределением величины спроса на этот товар при заданной цене. Это приводит нас к понятию условного распределения. Общая теория условных распределений и условных математических ожиданий является одной из наиболее сложных тем в теории вероятностей. Поэтому мы ограничимся рассмотрением двух частных, но наиболее важных с практической точки зрения случаев.

Рассмотрим дискретный случайный вектор размерности $m+n$ — $\xi = (\xi_1, \ldots, \xi_m,\, \xi_{m+1}, \ldots, \xi_{m+n})$. Предположим, что мы фиксировали значения первых m координат. Тогда условное распределение последних n координат определяется по формуле

$$P\big(x_{m+1}, \ldots, x_{m+n} | x_1, \ldots, x_m\big) =$$
$$\frac{P(\xi_1 = x_1, \ldots, \xi_m = x_m, \xi_{m+1} = x_{m+1}, \ldots, \xi_{m+n} = x_{m+n})}{P(\xi_1 = x_1, \ldots, \xi_m = x_m)}\ . \qquad (12.1)$$

Можно проверить, что условные вероятности в (1) (как функции от $x_{m+1}, \ldots, x_{m+n}$) обладают всеми свойствами дискретного распределения. Это распределение называют **условным распределением** $\xi_{m+1}, \ldots, \xi_{m+n}$ при условии, что $\xi_1 = x_1, \ldots, \xi_m = x_m$. Ясно, что, меняя условия $x_1, \ldots, x_m$ мы получим разные условные распределения.

Пример. Дискретный случайный вектор (ξ_1, ξ_2) имеет следующее распределение:

$\xi_1 \backslash \xi_2$	- 1	0	1
0	0.1	0.3	0.1
1	0.2	0.2	0.1

Найти условное распределение ξ_2 при условии $\xi_1 = 1$. Вероятность того, что $\xi_1 = 1$, равна $0.2 + 0.2 + 0.1 = 0.5$. Тогда по формуле (1) находим, что условное распределение ξ_2 задается таблицей

- 1	0	1
0.4	0.4	0.2

Выпишем основные свойства условного распределения. Для упрощения обозначений ограничимся двумерным случаем.

1) $P\big(\xi_2 = y | \xi_1 = x\big) \geqslant 0$.
2) $\sum\limits_y P\big(\xi_2 = y | \xi_1 = x\big) = 1$.
3) $P\big(\xi_1 = x, \xi_2 = y\big) = P(\xi_1 = x) P\big(\xi_2 = y | \xi_1 = x\big)$.
4) $P(\xi_2 \in B) = \sum\limits_x P\big(\xi_2 \in B | \xi_1 = x\big) P(\xi_1 = x)$.

Как и в случае безусловных распределений, можно определить понятие математического ожидания. Пусть $\xi = (\xi_1, \dots, \xi_m, \xi_{m+1})$ – дискретный случайный вектор. **Условное математическое ожидание** с.в. ξ_{m+1} при условии, что $\xi_1 = x_1, \dots, \xi_m = x_m$, есть число

$$M\big(\xi_{m+1}|\xi_1 = x_1, \dots, \xi_m = x_m\big) =$$

$$\sum_y yP\big(\xi_{m+1} = y|\xi_1 = x_1, \dots, \xi_m = x_m\big) \ . \tag{12.2}$$

Ясно, что при различных условиях мы будем получать разные условные математические ожидания, т.е. условное математическое ожидание ξ_{m+1} при условии $\xi_1 = x_1, \dots, \xi_m = x_m$ есть некоторая функция $g(x_1, \dots, x_m)$. Иногда нам будет удобно подставлять в эту функцию сами случайные величины $\xi_1, \dots, \xi_m$ и рассматривать условное математическое ожидание как случайную величину. В этом случае мы будем использовать обозначение $M\big(\xi_{m+1}|\xi_1, \dots, \xi_m\big)$. На самом деле последнее выражение просто равно $g(\xi_1, \dots, \xi_m)$. Функция $y = g(x_1, \dots, x_m)$ называется **функцией регрессии** случайной величины ξ_{m+1} на случайные величины $\xi_1, \dots, \xi_m$. Условное математическое ожидание, как функция от ξ_{m+1}, обладает всеми свойствами обычного математического ожидания. Но, кроме того, справедлив следующий аналог формулы полной вероятности:

$$M(\xi_{m+1}) = M\Big[M\big(\xi_{m+1}|\xi_1, \dots, \xi_m\big)\Big] \ . \tag{12.3}$$

Этот результат называется **формулой полного математического ожидания**.

Задача. Доказать формулу (3).

В качестве задачи предлагается доказать следующие свойства условного математического ожидания:

1) $M\big(\varphi(\xi)|\xi\big) = \varphi(\xi)$.

2) $M\big(\varphi(\xi_1)\xi_2|\xi_1\big) = \varphi(\xi_1)M\big(\xi_2|\xi_1\big)$.

3) Если ξ_1 и ξ_2 — независимы, то для любой борелевской функции $y = \psi(x_1, x_2)$

$$M\big[\psi(\xi_1, \xi_2)|\xi_1 = x\big] = M\big[\psi(x, \xi_2)\big] \ .$$

В частности,

$$M\big(\xi_2|\xi_1\big) = M(\xi_2) \ .$$

Замечания. 1. Можно рассмотреть и более общую задачу: найти условное распределение случайного вектора (ξ_1, ξ_2) при некотором условии на обе координаты. Например, в качестве такого условия можно взять $\varphi(\xi_1, \xi_2) = y$ или $\varphi(\xi_1, \xi_2) > y$. Если распределение (ξ_1, ξ_2) задано в виде таблицы, то для того, чтобы найти условное распределение (ξ_1, ξ_2) при некотором условии, нужно выбрать те клетки таблицы, где это условие выполнено, и разделить соответствующие вероятности на сумму всех вероятностей в этих клетках, а в остальных клетках нужно поставить нули.

2. В прикладных задачах часто вначале заданы условное распределение и вероятности условий. Тогда, используя формулу полной вероятности, можно найти и безусловное распределение.

Пусть теперь случайный вектор $\xi = (\xi_1, \ldots, \xi_m, \xi_{m+1}, \ldots, \xi_{m+n}) = (\xi', \xi'')$ имеет абсолютно непрерывное распределение с плотностью $\rho_\xi\big(x_1, \ldots, x_m, x_{m+1}, \ldots, x_{m+n}\big)$. По аналогии с формулой (1) определим условную плотность подвектора $(\xi_{m+1}, \ldots, \xi_{m+n})$ при условии $\xi_1 = x_1, \ldots, \xi_m = x_m$ по формуле

$$\rho_{\xi''|\xi'}\big(x_{m+1}, \ldots, x_{m+n}|x_1, \ldots, x_m\big) = \frac{\rho_\xi(x_1, \ldots, x_m, x_{m+1}, \ldots, x_{m+n})}{\rho_{\xi'}(x_1, \ldots, x_m)} \ . \tag{12.4}$$

При таком определении условной плотности естественно определить условное распределение по формуле

$$P_{\xi''|\xi'}\big(B|x_1, \ldots, x_m\big) = \int\limits_B \rho_{\xi''|\xi'}\big(x_{m+1}, \ldots, x_{m+n}|x_1, \ldots, x_m\big)\, dx_{m+1} \ldots dx_{m+n} \ , \tag{12.5}$$

где B — борелевское подмножество в R^n. В частности, мы можем определить условную функцию распределения

$$F_{\xi''|\xi'}\big(x_{m+1}, \ldots, x_{m+n}|x_1, \ldots, x_m\big) = P\big(\xi_{m+1} < x_{m+1}, \ldots, \xi_{m+n} < x_{m+n}|x_1, \ldots, x_m\big) \ . \tag{12.6}$$

Нетрудно проверить, что справедливы следующие свойства:

1) $\rho_{\xi''|\xi'}\bigl(x''|x'\bigr) \geqslant 0$,

2) $\int\limits_{R^n} \rho_{\xi''|\xi'}\bigl(x''|x'\bigr)\,dx'' = 1$,

3) $\rho_\xi(x', x'') = \rho_{\xi'}(x')\rho_{\xi''|\xi'}\bigl(x''|x'\bigr)$,

4) $\rho_{\xi''}(x'') = \int\limits_{R^m} \rho_{\xi'}(x')\rho_{\xi''|\xi'}\bigl(x''|x'\bigr)\,dx'$ — формула полной вероятности для плотностей.

Т.к. условная плотность $\rho_{\xi''|\xi'}\bigl(x''|x'\bigr)$ зависит от x', то все вычисленные по ней характеристики будут функциями от x'. В частности, мы можем определить **условное математическое ожидание** ξ_{m+1} при условии $\xi' = x'$:

$$M\bigl[\xi_{m+1}|\xi' = x'\bigr] := \int\limits_{-\infty}^{\infty} y\rho_{\xi_{m+1}|\xi'}\bigl(y|x'\bigr)\,dy = g(x') \ . \qquad (12.7)$$

Функция $y = g(x')$ называется **функцией регрессии** с.в. ξ_{m+1} на случайный вектор ξ'. Если в этом определении сделать условие случайным, т.е. рассмотреть $g(\xi')$, то мы получим новую случайную величину, которую вновь будем называть условным математическим ожиданием ξ_{m+1} относительно случайного вектора ξ' и использовать для него обозначение $M\bigl[\xi_{m+1}|\xi_1, \ldots, \xi_m\bigr]$. Оно обладает всеми обычными свойствами математического ожидания по аргументу ξ_{m+1}. Но, кроме того, справедливы следующие свойства, которые предлагается доказать самостоятельно.

Задача. 1) $M\bigl[\varphi(\xi')|\xi'\bigr] = \varphi(\xi')$.

2) $M\bigl[\xi_{m+1}\varphi(\xi')|\xi'\bigr] = \varphi(\xi')M\bigl[\xi_{m+1}|\xi'\bigr]$.

3) $M(\xi_{m+1}) = M\Bigl\{M\bigl[\xi_{m+1}|\xi'\bigr]\Bigr\}$ — формула полного математического ожидания.

4) Если ξ_{m+1} и ξ' — независимы, то

$$M\bigl[\psi(\xi', \xi_{m+1})|\xi' = x'\bigr] = M\bigl[\psi(x', \xi_{m+1})\bigr] \ ,$$

в частности,

$$M\bigl[\xi_{m+1}|\xi'\bigr] = M(\xi_{m+1}) \ .$$

Без доказательства рассмотрим несколько более общую ситуацию, которая иллюстрирует смысл условного математического ожидания и формулу полного математического ожидания. Для простоты рассмотрим только двумерный случай. Пусть (ξ_1, ξ_2) — двумерный случайный вектор и $\eta_1 = \psi_1(\xi_1, \xi_2)$ — некоторая борелевская функция от этого случайного вектора. Это новая случайная величина. Уравнение $y = \psi_1(x_1, x_2)$ выделяет некоторую линию уровня этой функции. Вся плоскость будет разбита на такие линии. Рассмотрим еще одну случайную величину $\eta_2 = \psi_2(\xi_1, \xi_2)$. В силу формулы полного математического ожидания мы имеем

$$M\eta_2 = M\big[\psi_2(\xi_1, \xi_2)\big] = M\Big\{M\big[\psi_2(\xi_1, \xi_2)|\psi_1(\xi_1, \xi_2)\big]\Big\} .$$

Смысл этого результата состоит в следующем. Сначала мы находим условное математическое ожидание вдоль каждой линии, а затем усредняем полученные результаты по всем линиям. Частный случай этого результата известен из курса математического анализа — это известная формула Фубини для вычисления кратных интегралов.

Используя сформулированные выше свойства, можно доказать следующие полезные для практических применений результаты.

Задача. Пусть (ξ_1, ξ_2) — случайный вектор. Тогда

1) $M(\xi_2) = M\big[M(\xi_2|\xi_1)\big]$;

2) $D(\xi_2) = M\big[D(\xi_2|\xi_1)\big] + D\big[M(\xi_2|\xi_1)\big]$.

Чаще всего этот результат применяется к случайным суммам, например в теории страхования. Пусть $X_1, X_2, \ldots$ — последовательность независимых и одинаково распределенных с.в., а N — независимая от них с.в., принимающая целые неотрицательные значения. Тогда величина $S_N = X_1 + \cdots + X_N$ $(S_0 = 0)$ называется случайной суммой. Используя предыдущую задачу, можно доказать следующее.

Задача. 1) $M(S_N) = M(X_1) \cdot M(N)$.

2) $D(S_N) = D(X_1) \cdot M(N) + \big(M(X_1)\big)^2 \cdot D(N)$.

В заключение мы рассмотрим одно экстремальное свойство функции регрессии, которое будет использоваться в математической статистике. В качестве разминки предлагается решить следующую задачу.

Задача. Пусть ξ — с.в. с конечным вторым моментом. Тогда $\forall a \in R^1$

$$D(\xi) = M(\xi - M\xi)^2 \leqslant M(\xi - a)^2 = f(a) \ ,$$

т.е. функция $f(a)$ принимает наименьшее значение при $a = M\xi$.

Аналогичный результат справедлив и для условных математических ожиданий.

Теорема 7 (**экстремальное свойство функции регрессии**). *Пусть ξ и η — две случайные величины, $y = f(x)$ — некоторая борелевская функция, причем $M(\eta^2) < \infty$ и $M\big([f(\xi)]^2\big) < \infty$. Если $y = g(x) = M\big[\eta|\xi = x\big]$ есть функция регрессии с.в. η на с.в. ξ, то*

$$M\big|\eta - g(\xi)\big|^2 \leqslant M\big|\eta - f(\xi)\big|^2 \ .$$

Доказательство. В силу свойств условных математических ожиданий и сформулированной выше задачи при каждом x мы имеем

$$M\big[(\eta - g(\xi))^2|\xi = x\big] = M\big[(\eta - g(x))^2|\xi = x\big] \leqslant$$

$$M\big[(\eta - f(x))^2|\xi = x\big] = M\big[(\eta - f(\xi))^2|\xi = x\big] \ .$$

Усредняя по условию, из формулы полного математического ожидания мы получаем нужный результат.

На самом деле мы доказали даже больше. А именно, что нужное неравенство справедливо для условных математических ожиданий. Объясним смысл полученного результата. По условию с.в. $\eta \in L_2$. Рассмотрим множество $\mathcal{L}$ с.в., которые можно представить в виде $f(\xi)$, причем $M(|f(\xi)|^2) < \infty$. Можно проверить, что $\mathcal{L}$ — замкнутое линейное подпространство в L_2. Теорема утверждает, что наилучшее приближение η в пространстве $\mathcal{L}$ (т.е. с помощью функций от ξ) дает нам условное математическое ожидание $g(\xi)$. Ранее мы получили аналогичный результат для линейных функций от ξ. Если рассмотреть произвольное линейное подпространство $\mathcal{L}$ в L_2, то естественно назвать наилучшее приближение для с.в. η в пространстве $\mathcal{L}$ **условным математическим ожиданием в широком смысле**.

Наиболее важным примером в статистике является многомерное нормальное распределение. Для простоты вычислений мы рассмотрим только двумерный случай.

Пример (двумерное нормальное распределение). Напомним, что случайный вектор $\xi = (\xi_1, \xi_2)$ имеет двумерное нормальное распределение, если он имеет плотность распределения следующего вида:

$$\begin{aligned}\rho_\xi(x_1,x_2) &= \frac{1}{2\pi\sqrt{\sigma_1^2\sigma_2^2(1-\rho^2)}}\exp\left\{-\frac{1}{2(1-\rho^2)}\Big[\frac{(x_1-a_1)^2}{\sigma_1^2}-\right.\\ &\left.-2\rho\frac{(x_1-a_1)}{\sigma_1}\cdot\frac{(x_2-a_2)}{\sigma_2}+\frac{(x_2-a_2)^2}{\sigma_2^2}\Big]\right\}\ ,(x_1,x_2)\in R^2\ .\end{aligned}$$

Выделяя полный квадрат по x_2 и интегрируя по этой переменной, мы получаем, что $\xi_1 \in N(a_1, \sigma_1^2)$. Аналогичный результат справедлив и для ξ_2. Отсюда мы получаем, что $a_1 = M\xi_1$, $a_2 = M\xi_2$, $\sigma_1^2 = D(\xi_1)$, $\sigma_2^2 = D(\xi_2)$. Вычисляя соответствующий двумерный интеграл, можно показать, что $\rho = \rho(\xi_1, \xi_2)$. Вычислим теперь условное распределение ξ_2 при условии $\xi_1 = x_1$. По формуле (4) после стандартных вычислений получаем

$$\begin{aligned}\rho_{\xi_2|\xi_1}\big(x_2|x_1\big) &= \frac{\rho_{\xi_1}(x_1,x_2)}{\rho_{\xi_1}(x_1)} =\\ &= \frac{1}{\sqrt{2\pi\sigma_2^2(1-\rho^2)}}\cdot\exp\left\{-\frac{\big[x_2-(a_2+\rho\frac{\sigma_2}{\sigma_1}(x_1-a_1))\big]^2}{2\sigma_2^2(1-\rho^2)}\right\}\ .\end{aligned}$$

Таким образом, мы видим, что условное распределение ξ_2 при условии $\xi_1 = x_1$ является нормальным со средним

$$M\big(\xi_2|\xi_1 = x_1\big) = g(x_1) = a_2 + \rho\frac{\sigma_2}{\sigma_1}(x_1 - a_1)$$

и дисперсией

$$D\big(\xi_2|\xi_1 = x_1\big) = \sigma_2^2(1-\rho^2)\ .$$

Условное среднее является линейной функцией от x_1, а условная дисперсия не зависит от условия. Кроме того, в силу доказанной выше теоремы мы видим, что наилучшее приближение ξ_2 с помощью функций от ξ_1 является линейным в случае нормального распределения. Теперь мы можем записать с.в. ξ_2 в виде

$$\xi_2 = a_2 + \rho \cdot \frac{\sigma_2}{\sigma_1}(\xi_1 - a_1) + \varepsilon \ , \tag{12.8}$$

где

$$\varepsilon = \xi_2 - \left[a_2 + \rho \cdot \frac{\sigma_2}{\sigma_1}(\xi_1 - a_1)\right] \ .$$

Оба слагаемых в представлении (8) имеют нормальное распределение и некоррелированы (а значит, и независимы). Первое слагаемое описывает влияние на ξ_2 с.в. ξ_1, а второе — влияние каких-то других, неучтенных факторов.

Глава 13

Многомерное нормальное распределение

Нормальное распределение играет важную роль в теории вероятностей и математической статистике. В этом параграфе мы дадим определение и перечислим без доказательства основные свойства многомерного нормального распределения.

Определение 58 . *Случайный вектор* $\xi = (\xi_1, \ldots, \xi_n)$ *имеет* **многомерное нормальное распределение**, *если он обладает плотностью распределения следующего вида:*

$$\rho_\xi(x_1, \ldots, x_n) = (2\pi)^{-\frac{n}{2}} det(A) \exp\Big\{-\frac{1}{2}(A(x-m), x-m)\Big\} =$$

$$(2\pi)^{-\frac{n}{2}} det(A) \exp\Big\{-\frac{1}{2}\sum_{i,j=1}^{n} a_{ij}(x_i - m_i)(x_j - m_j)\Big\} ,$$

где $m = (m_1, \ldots, m_n) \in R^n$, $A = (a_{ij})$ *— симметричная положительно определенная* $n \times n$*–матрица.*

Выясним смысл параметров многомерного нормального распределения. Пусть $\Sigma = (\sigma_{ij}) = A^{-1}$. Тогда $m_j = M\xi_j$,

$$\sigma_{ij} = cov(\xi_i, \xi_j) = M(\xi_i - m_i)(\xi_j - m_j) .$$

m называется **вектором средних**, а Σ — **матрицей ковариаций**. Если с.в. ξ имеет многомерное нормальное распределение с вектором средних m и матрицей ковариаций Σ, то мы будем использовать обозначение $\xi \in N(m, \Sigma)$.

Теорема 8 . *Если* $C : R^n \to R^k$ *— линейное отображение* R^n *на* R^k, $\xi \in N(m, \Sigma)$, *то* $C\xi \in N(Cm, C\Sigma C^T)$.

Разобьем случайный вектор ξ на два подвектора размерности n_1 и n_2: $n_1 + n_2 = n$, т.е. $\xi = (\xi', \xi'')$. Это порождает разложения вектора средних $m = (m', m'')$ и матрицы ковариаций

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} .$$

Теорема 9 . *Если* $\xi = (\xi_1, \ldots, \xi_n) \in N(m, \Sigma)$, *то*

1) $\xi' \in N(m', \Sigma_{11})$,

2) ξ' *и* ξ'' *— независимы тогда и только тогда, когда*

$$\Sigma_{12} = \Sigma_{21}^T = 0.$$

Теорема 10 . *Если* $\xi = (\xi_1, \ldots, \xi_n) \in N(m, \Sigma)$, *то условное распределение* ξ'' *при условии* $\xi' = x'$ *является многомерным нормальным распределением с вектором средних*

$$m'' + \Sigma_{21}\Sigma_{11}^{-1}(x' - m')$$

и матрицей ковариаций

$$\Sigma_{22.1} = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12} .$$

Глава 14

Закон больших чисел

В самом начале нашего курса мы отметили следующий экспериментальный факт. В длинной серии независимых и одинаковых испытаний частота появления некоторого события, которая является случайной величиной, тяготеет к некоторому постоянному числу, которое мы назвали вероятностью этого события. Далее мы фиксировали в виде аксиом основные свойства вероятностей и в качестве следствия доказали более сложные свойства вероятностей. Но хорошая математическая теория должна не только аккумулировать в своих определениях те или иные экспериментальные факты, но и содержать некоторые теоремы, которые дают описание наиболее важных результатов, отмеченных первоначально эмпирически. Свойство устойчивости частот говорит об определенном типе сходимости случайных величин. Но это не есть сходимость в привычном для нас смысле, как это принято в математическом анализе. Если в некоторой серии испытаний событие A не произойдет ни разу, то его частота равна нулю. Наоборот, если оно будет происходить всегда, то его частота равна 1. Но таких серий будет немного. Эксперимент показывает, что *большинство* серий испытаний будут такими, что частота появления события A в каждой из них будет примерно одинаковой. Таким образом, наше утверждение об устойчивости частот само носит вероятностный характер. Эти предварительные рассуждения приводят нас к следующему определению.

Определение 59 . *Последовательность с.в.* $\{\xi_n\}$ **сходится по ве-**

роятности *к с.в.* ξ, *если* $\forall\varepsilon > 0$

$$P\big(|\xi_n - \xi| > \varepsilon\big) \to 0 \ ,$$

или

$$P\big(|\xi_n - \xi| \le \varepsilon\big) \to 1.$$

при $n \to \infty$.

Такая сходимость обозначается $\xi_n \xrightarrow{P} \xi$, $n \to \infty$.

В этом определении утверждается, что событие, состоящее в том, что ξ_n и ξ близки $\big(|\xi_n - \xi| \leqslant \varepsilon\big)$, имеет вероятность, близкую к 1, т.е. это *практически достоверное* событие.

Вернемся теперь к свойству устойчивости частот. Пусть ε_k равно 1, если в k–м испытании событие A произошло, и равно 0 — в противном случае. Тогда $S_n = \varepsilon_1 + \ldots + \varepsilon_n$ есть число появлений события A в n испытаниях, а S_n/n есть относительная частота появления события A. Легко вычислить, что

$$M\left(\frac{S_n}{n}\right) = \frac{M\varepsilon_1 + \ldots + M\varepsilon_n}{n} = p \ \big(= P(A)\big) \ .$$

Тогда свойство устойчивости относительных частот означает, что

$$\frac{S_n}{n} - \frac{MS_n}{n} = \frac{S_n}{n} - p \to 0 \ , \ n \to \infty \ ,$$

в некотором смысле. Используя эти соображения, дадим следующее

Определение 60 . *Говорят, что к последовательности с.в.* $\{\xi_n\}$ *применим* **закон больших чисел** *(ЗБЧ), если*

$$\frac{\xi_1 + \ldots + \xi_n}{n} - \frac{M\xi_1 + \ldots + M\xi_n}{n} \xrightarrow{P} 0 \ , \ n \to \infty \ .$$

В определении дано только описание некоторого свойства. Ниже мы приведем условия, при которых это свойство имеет место. По определению нам нужно доказать сходимость по вероятности. Полезным техническим средством при этом является неравенство Чебышева

$$P\big(|\xi - M\xi| > \varepsilon\big) \leqslant \frac{D(\xi)}{\varepsilon^2} \ ,$$

которое было доказано ранее.

Лемма 6 . *Пусть с.в.* ξ_n *имеют конечные математические ожидания* $M\xi_n$ *и дисперсии* $D(\xi_n)$. *Если* $D(\xi_n) \to 0,\ n \to \infty$, *то* $\xi_n - M\xi_n \xrightarrow{P} 0$.

Доказательство следует непосредственно из определения и неравенства Чебышева.

Теорема 11 . *Пусть* $\{\xi_n\}$ *— последовательность случайных величин, таких, что*

1) $\{\xi_n\}$ *— независимы,*

2) существуют конечные $M\xi_n$ *и* $D(\xi_n)$,

3) $n^{-2} \sum\limits_{k=1}^{n} D(\xi_k) \to 0\ ,\ n \to \infty$.

Тогда применим ЗБЧ, т.е.

$$\frac{\xi_1 + \ldots + \xi_n}{n} - \frac{M\xi_1 + \ldots + M\xi_n}{n} \xrightarrow{P} 0\ ,\ n \to \infty\ .$$

Доказательство. Положим, $S_n = \xi_1 + \ldots + \xi_n$ и

$$\overline{S}_n = \frac{\xi_1 + \ldots + \xi_n}{n}\ .$$

Тогда

$$M(\overline{S}_n) = \frac{M\xi_1 + \ldots + M\xi_n}{n}\ ,\ D(\overline{S}_n) = \frac{1}{n^2}\sum_{k=1}^{n} D(\xi_k) \to 0\ .$$

Утверждение теоремы следует теперь из сформулированной выше леммы.

Замечание. Условие 1 нашей теоремы можно значительно ослабить, потребовав только попарной *некоррелированности*.

Теорема 12 . *Пусть* $\{\xi_n\}$ *— последовательность случайных величин:*

1) $\{\xi_n\}$ *— независимы,*

2) существуют конечные $M\xi_n$ *и* $D(\xi_n)$,

3) $\forall n\ D(\xi_n) \leqslant C < \infty$.

Тогда применим ЗБЧ.

Доказательство. Достаточно проверить условие 3 теоремы 1.

$$n^{-2}\sum_{k=1}^{n} D(\xi_k) \le n^{-2}\cdot n\cdot C = C/n \to 0\ ,\ n\to\infty\ .$$

Теорема 13 . *Пусть $\{\xi_n\}$ — последовательность случайных величин:*

1) $\{\xi_n\}$ — независимы,

2) $\{\xi_n\}$ — одинаково распределены,

3) $\exists M\xi_n = a,\ D(\xi_n) = \sigma^2 < \infty$.

Тогда применим ЗБЧ, т.е.

$$\frac{\xi_1+\ldots+\xi_n}{n} \xrightarrow{P} a\ ,\ n\to\infty\ .$$

Это частный случай теоремы 2.

Пример. Для измерения некоторой постоянной величины произвели n независимых измерений в одинаковых условиях. Каждое измерение ξ_n можно представить в виде $\xi_n = \varepsilon_n + a$, где $\{\varepsilon_n\}$ – последовательность н.о.р.с.в.: $M(\varepsilon_n) = 0,\ D(\varepsilon_n) = \sigma^2$. Тогда, в силу ЗБЧ,

$$\frac{\xi_1+\ldots+\xi_n}{n} = \frac{\varepsilon_1+\ldots+\varepsilon_n}{n} + a \xrightarrow{P} a\ ,\ n\to\infty\ ,$$

т.е. среднее арифметическое большого числа измерений сходится к истинному значению.

Замечание. А.Я. Хинчин получил более сильный результат: ЗБЧ справедлив для последовательности н.о.р.с.в. с конечным математическим ожиданием, т.е. условие конечности дисперсии является излишним.

В заключение приведем два исторически первых результата из числа ЗБЧ.

Теорема 14 *(**Я. Бернулли, 1713**). Пусть S_n/n — относительная частота появления события A в n независимых испытаниях, вероятность которого равна p. Тогда*

$$\frac{S_n}{n} \xrightarrow{P} p\ .$$

Эта теорема есть частный случай предыдущей теоремы. Действительно, выше мы показали, что $S_n = \varepsilon_1 + \cdots + \varepsilon_n$, где $\{\varepsilon_n\}$ — н.о.р. и $M\varepsilon_n) = p,\ D(\varepsilon_n) = p(1-p) \leqslant 1/4$.

В некоторых ситуациях вероятность появления события A меняется от испытания к испытанию. Тогда имеет место

Теорема 15 (**Пуассон, 1837**). *Пусть мы имеем последовательность независимых испытаний, где p_k — вероятность появления события A в k - м испытании. Если S_n — число появлений события A в n испытаниях, то*

$$\frac{S_n}{n} - \frac{p_1 + \ldots + p_n}{n} \xrightarrow{P} 0\ .$$

Глава 15

Центральная предельная теорема

15.1 Постановка задачи

Во многих теоретических и прикладных задачах теории вероятностей мы встречаемся с ситуацией, когда нам нужно найти распределение суммы независимых величин. Формально мы знаем ее решение. Нужно последовательно несколько раз применить формулу свертки. Но здесь мы встречаем два препятствия. Во-первых, для вычисления свертки нам нужно точно знать аналитический вид каждого слагаемого. Во-вторых, вычисление свертки довольно сложная операция, не всегда реализуемая в конкретной задаче. Во многих науках эти трудности преодолеваются путем построения тех или иных простых аппроксимаций для более сложных моделей. То же самое делают и в теории вероятностей. Можно даже сказать, что основная часть теории вероятностей посвящена описанию таких аппроксимаций. Одной из наиболее известных является так называемая *центральная предельная теорема* (ЦПТ), к изучению которой мы и приступаем.

Чтобы понять суть проблемы, вернемся к простой модели последовательности независимых испытаний, которую мы изучали ранее. Пусть мы имеем последовательность независимых испытаний с двумя исходами и постоянной вероятностью успеха p. Обозначим через ε_k случайную величину, равную 1, если в k-м испытании был успех, и 0 – в противном случае. Тогда число успехов в n испытаниях есть случайная величина $S_n = \varepsilon_1 + \ldots + \varepsilon_n$. Она имеет биномиальное распределение с параметрами n и p. При больших n расчеты по формулам биномиального распределения становятся довольно сложными. Поэтому применяют те или иные приближенные формулы. Одну из таких формул дает теорема Пуассона, другую — теорема Муавра–Лапласа. Рассмотрим нормированное число успехов

$$S_n^* = \frac{S_n - np}{\sqrt{np(1-p)}} ,$$

где $np = M(S_n),\ np(1-p) = D(S_n)$.

Теорема Муавра — Лапласа. *При $n \to \infty$ для $x \in R^1$ имеет место сходимость*

$$P(S_n^* < x) = F_{S_n^*}(x) \to \Phi(x) ,$$

где

$$\Phi(x) = \int\limits_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2} dy$$

— функция распределения стандартного нормального закона.

Таким образом, эта теорема утверждает, что при большом n и фиксированном p распределение соответствующим образом нормированной суммы случайных величин аппроксимируется нормальным распределением. Нормальное распределение хорошо изучено, обладает целым рядом привлекательных свойств, для него составлены таблицы. Все это позволяет надеяться, что такая аппроксимация будет удобной в практических задачах. Хотелось бы получить подобный результат и в более общей ситуации. Для этого отметим существенные черты полученного результата. Мы имеем последовательность независимых с.в. $\{\xi_n\}$, из которых образована последовательность сумм $S_n = \xi_1 + \ldots + \xi_n,\ n \geqslant 1$. Если эти суммы соответствующим образом нормировать и центрировать, т. е. рассмотреть

$$S_n^* = \frac{S_n - A_n}{B_n} ,$$

где $A_n \in R^1$, $B_n > 0$, то функции распределения нормированных сумм хорошо аппроксимируются функцией распределения стандартного нормального закона, а именно

$$F_{S_n^*}(x) \to \Phi(x)\ ,\ n \to \infty\ .$$

Это приводит нас к следующему определению.

Определение 61 . *Говорят, что к последовательности случайных величин $\{\xi_n\}$ применима* **центральная предельная теорема** *(ЦПТ), если для любого $n \geqslant 1$ существуют центрирующие и нормирующие константы $A_n \in R^1$ и $B_n > 0$, такие, что для*

$$S_n^* = \frac{\xi_1 + \ldots + \xi_n - A_n}{B_n} \tag{15.1}$$

имеет место сходимость

$$P(S_n^* < x) = F_{S_n^*}(x) \to \Phi(x) \tag{15.2}$$

при $n \to \infty$ для любого $x \in R^1$.

Замечания. 1. Обычно в реальных задачах $A_n = M(S_n)$, $B_n = \sqrt{D(S_n)}$. При такой нормировке мы получаем $M(S_n^*) = 0$, $D(S_n^*) = 1$.

2. В этом определении сформулировано некоторое свойство последовательности с.в. Но нас, конечно, интересует, при каких условиях оно имеет место. Утверждения, которые обычно называют ЦПТ, как раз и содержат те условия, при которых это свойство выполнено.

В ЦПТ мы изучаем сходимость не самих случайных величин а соответствующих им функций распределения. Поэтому в дальнейшем будет полезно следующее определение.

Определение 62 . *Последовательность функций распределения $\{F_n\}$* **слабо сходится** *к функции распределения F_0, если*

$$F_n(x) \to F_0(x)$$

при $n \to \infty$ для всех точек x, где предельная ф.р. F_0 непрерывна. Для соответствующих с.в. ξ_n будем говорить, что они сходятся к с.в. ξ_0 **по распределению**, *и писать*

$$\xi_n \xrightarrow{d} \xi_0 \ .$$

Так как функция распределения нормального закона всюду непрерывна, то мы изучаем обычную сходимость функций распределений.

В ЦПТ рассматривается сходимость распределений для сумм. Даже если слагаемые независимы, то нужно применять сложную операцию — свертку распределений. Поэтому при доказательстве ЦПТ применяют особый метод — **метод интегральных преобразований**. Он оказывается полезным и в других задачах.

15.2 Характеристические и производящие функции

Характеристические функции в теории вероятностей применяли уже Лаплас и Коши. Но систематическое их использование началось только после того, как А.М. Ляпунов (ученик П.Л. Чебышева) применил их в 1901 году для доказательства очень сильного варианта ЦПТ. Сегодня метод характеристических функций один из основных инструментов теории вероятностей.

Определение 63 . **Характеристической функцией** *с.в. ξ (или соответствующей ф.р. F) называется комплекснозначная функция $\varphi_\xi(t)$, $t \in R^1$, определяемая по правилу*

$$\varphi_\xi(t) = M(e^{it\xi}) = \int\limits_{-\infty}^{\infty} e^{itx}\, dF(x) \ . \tag{15.3}$$

Для дискретной с.в. ξ мы имеем формулу

$$\varphi_\xi(t) = \sum_n e^{itx_n} \cdot p_n \ , \tag{15.4}$$

а для непрерывной с.в. ξ с плотностью распределения $\rho_\xi(x)$ —

$$\varphi_\xi(t) = \int\limits_{-\infty}^{\infty} e^{itx} \rho_\xi(x)\, dx \ . \tag{15.5}$$

Напомним, что формула (5) есть обычное преобразование Фурье для плотности $\rho_\xi(x)$.

Начнем с того, что перечислим некоторые простейшие свойства характеристических функций.

Теорема 16 . *1.* $\forall \in R^1 \ \left|\varphi_\xi(t)\right| \leqslant 1 = \varphi_\xi(0)$.

2. $\varphi_{-\xi}(t) = \varphi_\xi(-t) = \overline{\varphi_\xi(t)}$.

3. $\varphi_\xi(t)$ *— равномерно непрерывная функция.*

4. Функция $\varphi_\xi(t)$ *положительно определена, т.е.* $\forall t_1, \ldots, t_n \in R^1$, $\forall c_1, \ldots, c_n \in C$

$$\sum_{k,m=1}^{n} \varphi_\xi(t_k - t_m) \cdot c_k \cdot \overline{c_m} \geqslant 0 \ .$$

5. Если $\eta = a\xi + b$, *то* $\varphi_\eta(t) = \varphi_\xi(at) \cdot e^{itb}$.

6. Если ξ_1 *и* ξ_2 *независимы, то*

$$\varphi_{\xi_1+\xi_2}(t) = \varphi_{\xi_1}(t) \cdot \varphi_{\xi_2}(t) \ .$$

7. Если $M\big(|\xi|^k\big) < \infty$ *для некоторого натурального* k, *то существует* k*–я производная для функции* $\varphi_\xi(t)$ *и*

$$\frac{d^k}{dt^k}\varphi_\xi(t)\Big|_{t=0} = (i)^k \cdot M(\xi^k) \ .$$

Доказательство. 1. Для любого $t \in R^1$

$$\begin{aligned} |\varphi_\xi(t)| &= \left|M(e^{it\xi})\right| \leqslant M\left|e^{it\xi}\right| = M(1) = 1 = \\ &= M\big(e^{i\cdot 0\cdot \xi}\big) = \varphi_\xi(0) \ . \end{aligned}$$

2. $\varphi_{-\xi}(t) = M\big(e^{it(-\xi)}\big) = M\big(e^{-it\xi}\big) = M\big(\overline{e^{it\xi}}\big) = \overline{M\big(e^{it\xi}\big)} = \overline{\varphi_\xi(t)}$.

3. Пусть $t \in R^1, \ h > 0$. Тогда

$$\varphi_\xi(t+h) - \varphi_\xi(t) = M\big[e^{i(t+h)\xi} - e^{it\xi}\big] =$$

$$M\big[e^{it\xi}(e^{ih\xi}-1)\big]=\int\limits_{-\infty}^{\infty}e^{itx}\big(e^{ihx}-1\big)\,dF_\xi(x)\ .$$

Так как $F_\xi(x)\to 0$ при $x\to-\infty$ и $F_\xi(x)\to 1$ при $x\to\infty$, то для любого $\varepsilon>0$ можно найти $A>0$, такое, что

$$F_\xi(-A)<\varepsilon/6\ ,\ 1-F_\xi(A)<\varepsilon/6\ .$$

Тогда мы получаем

$$\big|\varphi_\xi(t+h)-\varphi_\xi(t)\big|\leqslant\int\limits_{-\infty}^{\infty}\big|e^{ihx}-1\big|\,dF_\xi(x)=$$

$$\int\limits_{|x|\leqslant A}\big|e^{ihx}-1\big|\,dF_\xi(x)+\int\limits_{|x|>A}\big|e^{ihx}-1\big|\,dF_\xi(x)\leqslant$$

$$\varepsilon/3\cdot P(|\xi|\leqslant A)+2P(|\xi|>A)=\varepsilon/3+2\varepsilon/3=\varepsilon\ ,$$

если $h<\delta$. На последнем шаге мы использовали свойство непрерывности в нуле функции e^{iy} и то, что hx можно сделать равномерно малым при $|x|\leqslant A$ за счет выбора малого h.

4. Пусть $t_1,\dots,t_n\in R^1,\ c_1,\dots,c_n\in C$. Тогда

$$\sum_{k,m=1}^{n}\varphi_\xi(t_k-t_m)\cdot c_k\cdot\overline{c_m}=\sum_{k,m=1}^{n}M\big(e^{i(t_k-t_m)\xi}\big)c_k\overline{c_m}=$$

$$M\Big(\sum_{k=1}^{n}\sum_{m=1}^{n}e^{it_k\xi}c_k\overline{e^{it_m\xi}c_m}\Big)=M\Big(\sum_{k=1}^{n}e^{it_k\xi}c_k\Big)\overline{\Big(\sum_{m=1}^{n}e^{it_k\xi}c_m\Big)}=$$

$$M\Big|\sum_{k=1}^{n}e^{it_k\xi}c_k\Big|^2\geqslant 0\ .$$

5. Пусть $\eta=a\xi+b$, где $a,b\in R^1$. Тогда

$$\varphi_\eta(t)=M\big(e^{it\eta}\big)=M(e^{i(at)\xi}e^{itb})=e^{itb}\varphi_\xi(at)\ .$$

6. Если ξ_1 и ξ_2 — независимы, то с.в. $\eta_1 = e^{it\xi_1}$ и $\eta_2 = e^{it\xi_2}$ также будут независимы. Отсюда мы получаем

$$\varphi_{\xi_1+\xi_2}(t) = M\Big(e^{it(\xi_1+\xi_2)}\Big) = M\Big(e^{it\xi_1}e^{it\xi_2}\Big) =$$

$$M\Big(e^{it\xi_1}\Big)M\Big(e^{it\xi_2}\Big) = \varphi_{\xi_1}(t)\varphi_{\xi_2}(t)\ .$$

Последние два свойства показывают, что характеристические функции для линейных преобразований и для сумм независимых случайных величин легко вычисляются.

7. Продифференцируем формально выражение для характеристической функции:

$$\begin{aligned}\frac{d^k}{dt^k}\varphi_\xi(t) &= \frac{d^k}{dt^k}\left(\int\limits_{-\infty}^{\infty} e^{itx}dF_\xi(x)\right) = \\ &= \int\limits_{-\infty}^{\infty}(ix)^k e^{itx}dF_\xi(x)\ .\end{aligned}$$

Последний интеграл сходится абсолютно, если $M(|\xi|^k) < \infty$, так как

$$\Big|\int\limits_{-\infty}^{\infty}(ix)^k e^{itx}dF_\xi(x)\Big| \leqslant \int\limits_{-\infty}^{\infty}|x|^k\,dF_\xi(x) = M\big(|\xi|^k\big) < \infty\ .$$

Известный результат из математического анализа гарантирует нам тогда возможность дифференцирования по параметру под знаком интеграла. При $t = 0$ получаем

$$\frac{d^k}{dt^k}\varphi_\xi(t)\Big|_{t=0} = i^k\int\limits_{-\infty}^{\infty}x^k dF_\xi(x) = i^k M(\xi^k)\ . \tag{15.6}$$

В качестве следствия свойства 7 получаем следующее разложение характеристической функции:

$$\varphi_\xi(t) = 1 + M\xi\cdot\frac{it}{1!} + \ldots + M(\xi^k)\cdot\frac{(it)^k}{k!} + o(|t|^k) \tag{15.7}$$

при $t \to 0$.

Следующие свойства характеристических функций доказываются более громоздко, поэтому они будут приведены без доказательства. В силу своей важности они выделены в отдельные теоремы.

Теорема 17 (**единственности**). *Соответствие между функциями распределениями F и характеристическими функциями φ является взаимно однозначным. Более того, если x_1 и x_2 — точки непрерывности ф.р. F, то*

$$F(x_2) - F(x_1) = \lim_{A\to\infty} \frac{1}{2\pi} \int\limits_{-A}^{A} \frac{e^{-itx_2} - e^{-itx_1}}{-it} \varphi(t)dt \ . \tag{15.8}$$

В частности, если существует плотность $\rho(x)$, то

$$\rho(x) = \frac{1}{2\pi} \int\limits_{-\infty}^{\infty} e^{-itx} \varphi(t)dt \ . \tag{15.9}$$

Соотношения (8) и (9) называются **формулами обращения**. Эта теорема показывает, что по характеристической функции можно узнать, с каким распределением мы имеем дело.

Теорема 18 (**непрерывности**). *Пусть $F_0, F_1, F_2, \ldots$ — последовательность ф.р., а $\varphi_0, \varphi_1, \varphi_2, \ldots$ — соответствующая ей последовательность х.ф. Тогда следующие утверждения эквивалентны:*

1) F_n слабо сходится к F_0,

2) $\varphi_n(t) \to \varphi_0(t)$ при $n \to \infty$ для любого $t \in R^1$.

Эта теорема позволяет сводить задачу о слабой сходимости функций распределения к эквивалентной ей (а технически более простой) задаче о сходимости характеристических функций.

На практических занятиях будут вычислены характеристические функции некоторых стандартных распределений. Приведем без доказательства окончательные результаты.

Примеры. 1) Если $P(\xi = a) = 1$ для некоторого $a \in R^1$, то

$$\varphi_\xi(t) = e^{ita} \ .$$

2) Если ξ имеет распределение Бернулли с параметром p, т.е.

$$\xi = \begin{cases} 1 & , \quad p \ , \\ 0 & , \quad 1-p \ , \end{cases}$$

то

$$\varphi_\xi(t) = 1 + p(e^{it} - 1) \ .$$

3) Если ξ имеет биномиальное распределение с параметрами n и p, то

$$\varphi_\xi(t) = [1 + p(e^{it} - 1)]^n \ .$$

4) Если ξ имеет распределение Пуассона с параметром λ, то

$$\varphi_\xi(t) = e^{\lambda(e^{it}-1)} \ .$$

5) Если ξ имеет геометрическое распределение с параметром p, то

$$\varphi_\xi(t) = \frac{p}{1-(1-p)e^{it}} \ .$$

6) Если ξ имеет равномерное распределение на $[a, b]$, то

$$\varphi_\xi(t) = \frac{e^{itb} - e^{ita}}{it(b-a)} \ .$$

В частности, для равномерного распределения на $[-a, a]$ имеем

$$\varphi_\xi(t) = \frac{\sin at}{at} \ .$$

7) Если ξ имеет показательное распределение с параметром λ, то

$$\varphi_\xi(t) = \frac{\lambda}{\lambda - it} \ .$$

8) Если ξ имеет нормальное распределение с параметрами a и σ^2, то

$$\varphi_\xi(t) = e^{ita - \frac{\sigma^2 t^2}{2}} .$$

В частности, для стандартного нормального распределения ($a = 0$, $\sigma^2 = 1$) имеем

$$\varphi_\xi(t) = e^{-\frac{1}{2}t^2} .$$

Для дискретных случайных величин, принимающих значения вида $a + h \cdot n$, $n \in N$, (в частности, целочисленных) чаще используют так называемые производящие функции.

Определение 64 . *Пусть с.в.* ξ *принимает значения* $0, 1, 2, \ldots$ *с вероятностями* $p_k = P(\xi = k)$. **Производящая функция** *с.в.* ξ *(дискретного распределения* $\{p_k\}$*) определяется по формуле*

$$P_\xi(z) = \sum_{k=0}^{\infty} p_k z^k = M\bigl(z^\xi\bigr) . \tag{15.10}$$

Здесь z — комплексное число. Ряд (10) сходится, по крайней мере, для $|z| \le 1$. Производящие функции обладают свойствами, аналогичными свойствам характеристических функций, которые легко получить, используя следующее соответствие между этими функциями: если $z = e^{it}$, то

$$\varphi_\xi(t) = P_\xi(z) = P_\xi\bigl(e^{it}\bigr) .$$

Название «производящая функция» связано со следующим ее свойством:

$$p_k = P(\xi = k) = \frac{1}{k!} \frac{d^k}{dz^k} P_\xi(z)\Big|_{z=0} . \tag{15.11}$$

Функция $P_\xi(z)$ **производит** вероятности p_k по формуле (11).

Если мы работаем с моментами, то более удобно использовать **производящую функцию моментов**:

$$M_\xi(t) = M(e^{t\xi}) . \tag{15.12}$$

Может случиться, что $M_\xi(t)$ существует только при $t = 0$. Если $M_\xi(t) < \infty$ для всех $|t| \le r$, где $r > 0$, то существуют все моменты, они однозначно определяют распределение, и имеет место следующий аналог формул (6) и (11):

$$\frac{d^k}{dt^k} M_\xi(t)\Big|_{t=o} = M\big(\xi^k\big) \ . \tag{15.13}$$

15.3 Центральная предельная теорема

В этом разделе мы рассмотрим несколько вариантов ЦПТ. Доказательство будет приведено только в простейшей ситуации, когда мы имеем последовательность независимых одинаково распределенных величин. В более сложных случаях доказательство, по-существу, то же, но технически труднее.

Теорема 19 . *Пусть последовательность с.в.* $\{\xi_n\}$ *удовлетворяет условиям:*

1) $\{\xi_n\}$ *— независимы,*

2) $\{\xi_n\}$ *— одинаково распределены,*

3) существуют конечные $M(\xi_1) = a$ *и* $D(\xi_1) = \sigma^2$.

Тогда к этой последовательности применима ЦПТ. Более точно, если

$$S_n^* = \frac{\xi_1 + \ldots + \xi_n - na}{\sigma\sqrt{n}} \ ,$$

то

$$F_{S_n^*}(x) \to \Phi(x) \ , \ n \to \infty \ .$$

Доказательство. Для доказательства будем использовать характеристические функции. В силу теоремы непрерывности достаточно доказать, что

$$\varphi_{S_n^*}(t) \to e^{-\frac{1}{2}t^2} \ , \ n \to \infty \ .$$

Запишем S_n^* в виде

$$S_n^* = \sum_{k=1}^{n} \frac{\xi_k - a}{\sigma\sqrt{n}} = \frac{1}{\sqrt{n}} \sum_{k=1}^{n} \xi_k^* \ .$$

По условию теоремы с.в. $\{\xi_k^*\}$ — н.о.р. и $M(\xi_k^*)=0$, $D(\xi_k^*)=1$. Обозначим через φ характеристическую функцию с.в. ξ_1^*. Тогда в силу формулы (7) имеем

$$\varphi(t)=1-\frac{1}{2}t^2+o(t^2)$$

при $t\to 0$. Характеристическая функция с.в. S_n^* имеет вид (смотри свойства 5 и 6)

$$\varphi_{S_n^*}(t)=\left[\varphi\left(\frac{t}{\sqrt{n}}\right)\right]^n .$$

Используя разложение для φ, получаем при любом фиксированном $t\in R^1$

$$\varphi_{S_n^*}(t)=\left[1-\frac{1}{2}t^2\cdot\frac{1}{n}+o\left(\frac{t^2}{n}\right)\right]^n\to e^{-\frac{1}{2}t^2} .$$

Теорема доказана.

Из последней теоремы мы получаем как следствие интегральную теорему Муавра – Лапласа. Действительно, пусть $\varepsilon_k=1$, если в k-м испытании был успех (это происходит с вероятностью p), и $\varepsilon_k=0$ – в противном случае. Тогда $S_n=\varepsilon_1+\ldots+\varepsilon_n$ есть число успехов в n независимых испытаниях. С.в. $\{\varepsilon_k\}$ — н.о.р., $M(\varepsilon_k)=p$, $D(\varepsilon_k)=p(1-p)$. Отсюда следует, что $M(S_n)=np$, $D(S_n)=np(1-p)$. Окончательно получаем следующее выражение для нормированной суммы:

$$S_n^*=\frac{S_n-np}{\sqrt{np(1-p)}} .$$

Применяя только что доказанную теорему, получаем, что S_n^* имеет асимптотически стандартное нормальное распределение.

Выше мы доказали ЦПТ при довольно жестких условиях. Далее мы попытаемся без доказательства **объяснить**, в каких именно ситуациях следует ожидать нормальное предельное распределение для нормированных сумм случайных величин.

Пусть $\{\xi_k\}$ — последовательность независимых случайных величин, для которых существуют конечные $M(\xi_k)=a_k$ и $D(\xi_k)=\sigma_k^2$. Как и ранее, образуем суммы $S_n=\xi_1+\ldots+\xi_n$. Тогда

$M(S_n) = a_1 + \ldots + a_n = A_n$ и (в силу независимости) $D(S_n) = \sigma_1^2 + \ldots + \sigma_n^2 = B_n^2$. Определим нормированную сумму

$$S_n^* = \frac{S_n - A_n}{B_n} = \sum_{k=1}^{n} \xi_{kn} \ , \tag{15.14}$$

где $\xi_{kn} = (\xi_k - a_k)/B_n,\ k = \overline{1,n}$, есть независимые случайные величины, для которых $M(\xi_{kn}) = 0,\ D(\xi_{kn}) = \sigma_{kn}^2$ и $\sum\limits_{k=1}^{n} \sigma_{kn}^2 = 1$.

Теорема 20 . *Пусть для любого n мы имеем набор $\{\xi_{nk}\}$ независимых случайных величин, таких, что $M(\xi_{kn}) = 0$, $D(\xi_{kn}) = \sigma_{kn}^2$ и $\sum\limits_{k=1}^{n} \sigma_{kn}^2 = 1$. Если выполнено условие*

$$P\Big(\max_{1\leqslant k\leqslant n} |\xi_{kn}| > \varepsilon\Big) \to 0,\ n \to \infty \ , \tag{15.15}$$

или, эквивалентно,

$$\sum_{k=1}^{n} P\big(|\xi_{kn}| > \varepsilon\big) \to 0,\ n \to \infty \ , \tag{15.16}$$

то применима ЦПТ, т.е.

$$P\big(\xi_{1n} + \ldots + \xi_{nn} < x\big) \to \Phi(x)\ ,\ n \to \infty \ .$$

Это одна из наиболее общих форм ЦПТ. Смысл условия (15) состоит в том, что каждое из слагаемых ξ_{kn} суммы $S_n^* = \xi_{1n} + \ldots + \xi_{nn}$ вносит малый вклад по сравнению со всей суммой S_n^*. С математической точки зрения это наиболее ясная и простая форма ЦПТ. Но в реальных задачах проверить условие (15) бывает довольно трудно. Ниже мы приводим другие формулировки, которые содержат более легко проверяемые условия. Кстати, эти теоремы появились исторически раньше и помогли прояснить суть ЦПТ.

Теорема (Линдеберг, 1924). *Пусть введены те же обозначения, что и раньше. Если $\{\xi_{kn}\}$ — независимые случайные величины и* *выполнено* **условие Линдеберга***:* $\forall\varepsilon > 0$

$$\sum_{k=1}^{n} \int\limits_{|x|>\varepsilon} x^2\, dF_{kn}(x) \to 0\ ,\ n \to \infty\ . \qquad (15.17)$$

где $F_{kn}(x) = P(\xi_{kn} < x)$*, то применима ЦПТ, т.е.*

$$P\big(\xi_{1n} + \ldots + \xi_{nn} < x\big) \to \Phi(x)\ ,\ n \to \infty\ .$$

Условие (17) легко следует из условия (16) с помощью аналога неравенства Чебышева.

Теорема (Ляпунов, 1900). *Пусть* $\{\xi_k\}$ *— последовательность независимых случайных величин, для которых существуют конечные* $M(\xi_k) = a_k$*,* $D(\xi_k) = \sigma_k^2$ *и* $\beta_k = M(|\xi_k - a_k|^3)$*. Обозначим* $S_n = \xi_1 + \ldots + \xi_n$*,* $A_n = M(S_n) = a_1 + \ldots + a_n$*,* $B_n^2 = D(S_n) = \sigma_1^2 + \ldots + \sigma_n^2$ *и* $C_n = \beta_1 + \ldots + \beta_n$*. Если выполнено* **условие Ляпунова**

$$\frac{C_n}{B_n^{3/2}} \to 0\ ,\ n \to \infty\ , \qquad (15.18)$$

то применима ЦПТ, т.е. для $S_n^* = (S_n - A_n)/B_n$

$$P\big(S_n^* < x\big) \to \Phi(x)\ .$$

Преимущество последней теоремы состоит в том, что в условие (18) входят только моменты отдельных слагаемых, вычислять которые в практических задачах обычно довольно легко.

Во всех сформулированных выше теоремах присутствует условие независимости случайных величин. На самом деле это условие можно значительно ослабить и потребовать, чтобы слагаемые в сумме были зависимы в определенном смысле слабо.

Наряду с ЗБЧ и теоремой Пуассона о сходимости к пуассоновскому распределению в схеме Бернулли ЦПТ есть наиболее часто применяемый результат в практических задачах. Мы проиллюстрируем, как это делается, на примере одной задачи из теории страхования.

Пример. Некоторая страховая компания продала 100 страховых полисов, каждый из которых представляет собой договор страхования от несчастного случая и смерти. При наступлении несчастного случая лицу, указанному в договоре, выплачивается сумма 100 000 рублей, а в случае смерти застрахованного выплачивается сумма 10 000 000 рублей. Какова должна быть стоимость такого полиса, чтобы вероятность безубыточной деятельности компании была не менее 0.95? Собранный статистический материал показывает, что для той группы лиц, которой предлагается данная страховка, вероятность несчастного случая равна 0.001, а вероятность смерти равна 0.0001.

Решение. В данном случае выплаты X_k по k–му полису представляют собой с.в. со следующим распределением:

$$X_k = \begin{cases} 10^7 & , \quad 10^{-4} \ , \\ 10^5 & , \quad 10^{-3} \ , \\ \ \ 0 & , \quad 0.9989 \ . \end{cases}$$

Если пренебречь случаями каких-либо катастроф, эпидемий и т.п., то можно предположить, что $\{X_k\}$ — н.о.р.с.в. Легко вычислить, что $a = M(X_k) = 1010$, а $\sigma^2 = D(X_k) \approx 10^{10}$. Пусть стоимость страховки есть величина b, которую необходимо определить. Тогда общая собранная с застрахованных сумма равна $nb = 100b$, а суммарные выплаты по проданным полисам равны $S_n = X_1 + \ldots + X_n$. Компания понесет убытки, если $(S_n > nb)$. Вероятность этого события должна быть не более 0.05. Вычислим эту вероятность в нашей задаче:

$$P(S_n > nb) = P\left(\frac{S_n - na}{\sigma\sqrt{n}} > \frac{n(b-a)}{\sigma\sqrt{n}}\right) \approx 1 - \Phi\left(\left(\frac{b-a}{\sigma}\right)\sqrt{n}\right) \ .$$

Приравнивая

$$1 - \Phi\left(\left(\frac{b-a}{\sigma}\right)\sqrt{n}\right) = 0.05 \ ,$$

находим по таблицам нормального распределения, что

$$\frac{(b-a)}{\sigma}\sqrt{n} = 1.64 \ .$$

Отсюда получаем

$$b = a + 1.64\frac{\sigma}{\sqrt{n}} = 1010 + 1.64\frac{10^5}{10} = 17410 \ .$$

Глава 16

Цепи Маркова

До сих пор мы изучали последовательности независимых случайных величин. Но поведение реальных систем в каждый момент времени зависит, вообще говоря, от предыстории процесса, в простейшем случае — от состояния системы в данный момент времени. Это приводит нас к понятию цепи Маркова. Впервые модели такого типа начал изучать известный российский математик академик Андрей Андреевич Марков, в честь которого они и получили свое название.

16.1 Определение цепи Маркова

Определение цепи Маркова уже было дано при изучении последовательности независимых испытаний. Для удобства мы повторим его, но уже в терминах случайных величин.

Определение 65 . Цепью Маркова *называется последовательность с.в.* $\{X_n, n \geqslant 0\}$, *принимающих значения в конечном или счетном множестве* $X = \big\{x_1, x_2, \ldots\big\}$, *такая, что* $\forall n \geqslant 1$, $\forall x_{i_0}, \ldots, x_{i_n} \in X$

$$P(X_n = x_{i_n} | X_0 = x_{i_0}, \ldots, X_{n-1} = x_{i_{n-1}}) = P(X_n = x_{i_n} | X_{n-1} = x_{i_{n-1}}) \ . \tag{16.1}$$

Множество X называется **пространством состояний** *цепи Маркова, а вероятности*

$$P_{ij}^{(n)} := P(X_n = x_j | X_{n-1} = x_i) \tag{16.2}$$

называются **вероятностями перехода** *на n–м шаге.*

Набор вероятностей

$$P_i^{(0)} = P(X_0 = x_i)\ ,\ x_i \in X\ , \tag{16.3}$$

определяет **начальное распределение** *цепи Маркова.*

Во многих задачах необходимо знать также **распределение цепи на n–м шаге**, т.е. вероятности

$$P_i^{(n)} = P(X_n = x_i)\ ,\ x_i \in X\ , \tag{16.4}$$

и **вероятности перехода за m шагов**

$$P_{ij}^{(n)}(m) = P(X_{n-1+m} = x_j | X_{n-1} = x_i)\ . \tag{16.5}$$

Для упрощения обозначений мы будем писать как и раньше $P_{ij}^{(n)}(1) = P_{ij}^{(n)}$.

Легко показать (задача!), что вероятности перехода обладают следующими свойствами: для любых натуральных m, n, i, j

1) $P_{ij}^{(n)}(m) \geqslant 0$,

2) $\sum\limits_j P_{ij}^{(n)}(m) = 1$.

Используя теорему умножения и формулу полной вероятности, нетрудно доказать (задача!) следующие соотношения: $\forall n \geqslant 1,\ \forall x_{i_0}, \ldots, x_{i_n} \in X$

$$P(X_0 = x_{i_0}, X_1 = x_{i_1}, \ldots, X_n = x_{i_n}) = P_{i_0}^{(0)} \cdot P_{i_0 i_1}^{(1)} \cdot \ldots \cdot P_{i_{n-1} i_n}^{(n)} \tag{16.6}$$

и

$$P_j^{(n)} = \sum_i P_i^{(0)} P_{ij}^{(1)}(n)\ . \tag{16.7}$$

Таким образом, зная начальное распределение и матрицы переходных вероятностей, можно вычислить вероятности любых событий, связанных с поведением цепи Маркова. На самом деле достаточно знать матрицы перехода за один шаг. Действительно, имеет место следующее соотношение (задача!): $\forall n, m_1, m_2, i, j \geqslant 1$

$$P_{ij}^{(n)}(m_1+m_2) = \sum_k P_{ik}^{(n)}(m_1) \cdot P_{kj}^{(n+m_1)}(m_2) \ , \tag{16.8}$$

откуда следует

$$P_{ij}^{(n)}(m) = \sum_{k_1} \cdots \sum_{k_{m-1}} P_{ik_1}^{(n)} \cdots P_{k_{m-1}j}^{(n+m-1)} \ . \tag{16.9}$$

Цепь Маркова называется **однородной** (по времени), если вероятности перехода одни и те же на каждом шаге, т.е. $\forall n \geqslant 1$ $P_{ij}^{(n)} = P_{ij}$. Всюду далее мы будем рассматривать только такие цепи Маркова. В этом случае формулы (6), (7) и (8) можно переписать соответственно в следующем виде:

$$P(X_0 = x_{i_0}, X_1 = x_{i_1}, \ldots, X_n = x_{i_n}) = P_{i_0}^{(0)} \cdot P_{i_0 i_1} \cdot \ldots \cdot P_{i_{n-1} i_n} \ ; \tag{16.10}$$

$$P_j^{(n)} = \sum_i P_i^{(0)} P_{ij}(n) \ ; \tag{16.11}$$

$$P_{ij}(m_1+m_2) = \sum_k P_{ik}(m_1) \cdot P_{kj}(m_2) \ . \tag{16.12}$$

Последнее соотношение называется **уравнением Маркова**.

Используя матричные обозначения

$$P^{(n)} = \big(P_1^{(n)}, P_2^{(n)}, \ldots\big), \ P(n) = \big(P_{ij}(n)\big), \ P := P(1),$$

можно переписать соотношения (11) и (12) в следующем виде:

$$P^{(n)} = P^{(0)} \cdot P(n) \ , \tag{16.13}$$

$$P(m+n) = P(m) \cdot P(n) \ . \tag{16.14}$$

Из последнего соотношения получаем, что

$$P(n) = P^n \ . \tag{16.15}$$

Отсюда следует, что для полного описания однородной цепи Маркова достаточно знать начальное распределение и матрицу перехода за один шаг.

Для анализа динамики цепи Маркова полезно использовать граф, построенный по матрице P переходных вероятностей за один шаг. Его вершинами являются элементы пространства состояний X. Если $P_{ij} > 0$, то проводим **ориентированное** ребро, ведущее из i в j. Примеры применения этого понятия для анализа поведения цепи Маркова будут приведены ниже.

В заключение этого раздела рассмотрим классическую **задачу о разорении игрока**, которая впервые была сформулирована еще Гюйгенсом в 1656 году и которой интересовались многие известные математики в прошлом. Два игрока играют в некоторую игру, в которой первый игрок выигрывает с вероятностью p и проигрывает с вероятностью $q = 1 - p,\ 0 < p < 1$. Тот, кто выиграл, получает от проигравшего одну единицу его капитала. Игроки вступают в игру, имея на руках суммарный капитал величины a: z — у первого игрока и $a - z$ — у второго. Игра заканчивается, когда один из игроков полностью разорится. Нетрудно показать (задача!), что игра закончится с вероятностью 1 за конечное (но случайное) время. Необходимо вычислить вероятности разорения для каждого из игроков. Многие экономические задачи можно свести к задаче подобного типа.

Обозначим через q_z вероятность разорения первого игрока, если он начинает игру, имея на руках капитал $z,\ 0 < z < a$. По смыслу задачи естественно доопределить $q_0 = 1$ и $q_a = 0$. Это типичная задача о цепи Маркова. Выделим одного из игроков, например первого. Он может находиться в одном из состояний $z,\ 0 \le z \le a$. После каждой игры он переходит либо в состояние $z - 1$ с вероятностью q, либо в состояние $z + 1$ с вероятностью p, если $0 < z < a$. Попав в состояние 0 или a, он навсегда в нем остается. Необходимо найти вероятность попадания в состояние 0, если мы начинаем движение из состояния z.

Т.к. на каждом шаге мы имеем ту же задачу (однородная цепь Маркова!), но с разными z, то в силу формулы полной вероятности мы имеем следующее соотношение:

$$q_z = pq_{z+1} + qq_{z-1} \ , \ 0 < z < a - 1 \ . \tag{16.16}$$

Кроме того, в граничных точках мы имеем

$$q_0 = 1 \ , \ q_a = 0 \ . \tag{16.17}$$

Соотношение (16) есть однородное разностное уравнение второго порядка. При $p \neq q$ оно имеет два независимых частных решения: $q_z \equiv 1$ и $q_z = (q/p)^z$ (проверить!). Тогда общее решение имеет вид

$$q_z = A + B(q/p)^z \ .$$

Используя граничные условия (17), окончательно получим

$$q_z = \frac{(q/p)^a - (q/p)^z}{(q/p)^a - 1} \ . \tag{16.18}$$

Для $p = q$ уравнение (16) имеет независимые частные решения $q_z \equiv 1$ и $q_z = z$ (проверить!). Это приводит нас к решению

$$q_z = 1 - \frac{z}{a} \ . \tag{16.19}$$

16.2 Классификация состояний

Далее в этом параграфе мы будем рассматривать только конечные однородные цепи Маркова, т.е. такие, у которых множество $X = \{x_1, x_2, \ldots, x_r\}$ конечно.

Вероятностные свойства систем проявляются асимптотически, когда мы наблюдаем за ними достаточно долго и следим за тем, как часто происходят те или иные события. В силу этого нас будут интересовать асимптотические свойства цепей Маркова. С этой точки зрения разные состояния цепи обладают неодинаковыми свойствами. Поэтому мы начнем изложение нашей теории с классификации состояний. Для этой цели оказывается очень полезным введенное нами ранее понятие графа цепи Маркова, т.к. оно позволяет представить наглядно определяемые понятия.

Будем говорить, что состояние j **достижимо** из состояния i, если существует $n \geqslant 1$, такое, что $P_{ij}(n) > 0$. Состояния i и j называются **сообщающимися**, если i достижимо из j, а j достижимо из i. Состояние i называется **несущественным**, если можно найти состояние j и число $n_0 \geqslant 1$, такие, что $P_{ij}(n_0) > 0$, но $P_{ji}(n) = 0,\ \forall n \geqslant 1$. В противном случае состояние называется **существенным**.

Все состояния объединяются в некоторое число **замкнутых** классов. Внутри одного класса все состояния сообщаются, но это не так для состояний из разных классов. Доказательство этого проводится дословно так же, как и при определении классов смежности относительно линейного подпространства. Цепь Маркова называется **неразложимой**, если она имеет один замкнутый класс существенных состояний, и **разложимой** — в противном случае. Состояние i называется **поглощающим**, если $P_{ii} = 1$. Попав в такое состояние, мы остаемся в нем навсегда. Поглощающее состояние образует отдельный замкнутый класс.

Несущественное состояние обладает тем свойством, что с вероятностью 1 мы когда-то в очередной раз выйдем из этого состояния и больше в него не вернемся (задача!). Поэтому оно и является **несущественным** с точки зрения асимптотических свойств цепи.

Попав в один из замкнутых существенных классов, мы далее продолжаем движение только в нем. Это означает, что цепь Маркова распадается на некоторое число в определенном смысле «независимых» цепей. Поэтому достаточно изучить только цепи Маркова с одним замкнутым классом.

Периодом состояния i называется наибольший общий делитель d_i чисел n, для которых $P_{ii}(n) > 0$. Можно показать, что все состояния внутри одного замкнутого класса имеют один и тот же период d, который называют периодом данного класса. Класс называют **апериодическим**, если для него $d = 1$. **Периодический** класс разбивается на $d \geqslant 2$ подклассов, которые можно упорядочить так, что за один шаг возможны переходы только из одного подкласса в соседний.

16.3 Предельные теоремы для цепей Маркова

Большинство реальных систем ведут себя следующим образом. Некоторое время в системе происходят так называемые переходные явления, но через достаточно большой промежуток времени система выходит на стационарный режим и ее характеристики стабилизируются. Если динамика такой системы исследуется в рамках теории цепей Маркова, то нам необходимо ответить на следующие вопросы: что такое стационарный режим для цепи Маркова и что означает выход цепи в пределе на стационарный режим?

Определение 66 . *Распределение вероятностей* $Q = (q_1, q_2, \ldots, q_r)$ *на пространстве состояний* X *называется* **стационарным** *для однородной цепи Маркова с матрицей* $P = (P_{ij})$ *вероятностей перехода за один шаг, если имеет место соотношение*

$$Q = Q \cdot P \ , \tag{16.20}$$

или, более подробно,

$$q_j = \sum_i q_i \cdot P_{ij} \ . \tag{16.21}$$

Если взять такое распределение в качестве начального, т.е. $P^{(0)} = Q$, то в силу формул (13) и (15) имеем

$$P^{(n)} = P^{(0)} \cdot P(n) = Q \cdot P^n = Q \ .$$

Распределение цепи Маркова на n–м шаге будет совпадать с ее начальным распределением, т.е. не будет меняться со временем. Но именно это и означает стационарность системы.

Определение 67 . *Цепь Маркова называется* **эргодической**, *если для любого начального распределения* $P^{(0)}$ *существуют пределы*

$$\pi_j = \lim_{n\to\infty} P_j^{(n)} \tag{16.22}$$

и предельное распределение $\pi = (\pi_1, \ldots, \pi_r)$ не зависит от начального распределения. В частности, для любых i, j существуют **финальные вероятности**

$$\pi_j = \lim_{n\to\infty} P_{ij}(n) \ . \tag{16.23}$$

В последнем случае мы начинаем движение из фиксированного состояния x_i, т.е. $P(X_0 = x_i) = P_i^{(0)} = 1$.

Напомним, что мы рассматриваем **конечные** цепи Маркова. Тогда существование пределов (22) и (23) гарантирует, что финальные вероятности образуют распределение вероятностей π на X, т.е.

$$\pi_j \geqslant 0, \ \sum_j \pi_j = 1 \ . \tag{16.24}$$

Когда число состояний бесконечно, условие (24) необходимо явно вписать в определение эргодичности.

Нетрудно доказать (задача!), что финальные вероятности, если они существуют, образуют стационарное распределение, т.е.

$$\pi = \pi \cdot P \ . \tag{16.25}$$

Из определения эргодичности цепи Маркова следует, что в этом случае существует только одно стационарное распределение, называемое **эргодическим**, и оно совпадает с набором π финальных вероятностей (почему?). В общем случае это не так.

Пример. Пространство состояний X состоит из четырех элементов, которые мы обозначим их номерами 1,2,3 и 4. $P_{12} = P_{21} = P_{34} = P_{43} = 1/2$. Все остальные вероятности перехода равны нулю. Пусть, например, $Q_1 = (1/4, 1/4, 1/4, 1/4)$ и $Q_1 = (1/6, 1/6, 1/3, 1/3)$. Непосредственная проверка показывает, что Q_1 и Q_2 — стационарные распределения для нашей цепи Маркова. Построив граф этой цепи, мы видим, что она является разложимой. Это и есть причина неединственности стационарного распределения.

Задача. Пусть мы имеем разложимую цепь Маркова и $Q_1, \ldots, Q_k$ — стационарные распределения для каждой компоненты, т.е. стационарное распределение Q_l сосредоточено на l-й компоненте. Доказать, что распределение $Q = \alpha_1 Q_1 + \ldots + \alpha_r Q_r$, $\alpha \geqslant 0$, $\sum\limits_l \alpha_l = 1$, является стационарным распределением для всей цепи Маркова. Показать, что любое стационарное распределение представимо в таком виде.

Таким образом, одной из причин неэргодичности цепи Маркова является ее разложимость. Другой причиной является периодичность класса состояний. Пусть мы имеем один замкнутый класс с периодом d. Разобьем его на d подклассов, как это описано выше. Тогда $P_{ij}(nd) > 0$, если i и j лежат в одном подклассе, и $P_{ij}(nd) = 0$, если i и j лежат в разных подклассах. Отсюда видно, что, двигаясь по подпоследовательности $\{nd\}$, мы получим разные пределы для переходных вероятностей в зависимости от начального состояния. Для конечных цепей Маркова этим и исчерпывается список причин, приводящих к неэргодичности.

Теорема 21 . *Пусть мы имеем конечную однородную цепь Маркова, все состояния которой сообщаются. Тогда эта цепь Маркова является эргодической, т.е. имеют место соотношения (22). Более того, если число $\lambda > 0$ таково, что $\forall i$ и некоторого j_0*

$$P_{ij_0} \geqslant \lambda > 0 \ , \tag{16.26}$$

то для любых двух начальных распределений $Q_1 = (q_1, \ldots, q_r)$ и $Q_2 = (p_1, \ldots, p_r)$ соответствующие распределения $Q_1^{(n)} = \left(q_1^{(n)}, \ldots, q_r^{(n)}\right)$ и $Q_2^{(n)} = \left(p_1^{(n)}, \ldots, p_r^{(n)}\right)$ на n–м шаге удовлетворяют соотношению

$$\sum_{j=1}^{r} \left| q_j^{(n)} - p_j^{(n)} \right| \leqslant 2(1-\lambda)^n \ ,$$

т.е. эти вероятности сближаются при $n \to \infty$. Эргодическое распределение π является единственным решением уравнения (20), удовлетворяющим условиям (24).

Отметим, что рассматриваемая цепь Маркова является неразложимой и апериодической без несущественных состояний. Можно показать, что для любой цепи с такими свойствами все состояния сообщаются (задача!). Такая цепь Маркова называется **регулярной**.

Доказательство. Во-первых, заметим, что в нашем случае условие (26) всегда выполнено, т.к. $P_{ij} > 0 \ \forall i, j$ и этих вероятностей конечное число.

Начнем с доказательства неравенства (27), которое называется **неравенством Бернштейна**. Т.к. $Q_1^{(n)}$ и $Q_2^{(n)}$ являются распределениями вероятностей, то

$$\sum_j q_j^{(n)} = \sum_j p_j^{(n)} = 1 \ .$$

Отсюда следует, что

$$0 = \sum_j \bigl(q_j^{(n)} - p_j^{(n)}\bigr) = \sum_j{}^{+} \bigl(q_j^{(n)} - p_j^{(n)}\bigr) + \sum_j{}^{-} \bigl(q_j^{(n)} - p_j^{(n)}\bigr) \ .$$

Σ^+ содержит положительные слагаемые, а Σ^- – неположительные. Из последнего соотношения следует, что

$$\sum_j{}^{+} \bigl(q_j^{(n)} - p_j^{(n)}\bigr) = -\sum_j{}^{-} \bigl(q_j^{(n)} - p_j^{(n)}\bigr)$$

и

$$L_n := \sum_j \bigl|q_j^{(n)} - p_j^{(n)}\bigr| = 2\sum_j{}^{+} \bigl(q_j^{(n)} - p_j^{(n)}\bigr) \ .$$

На $(n-1)$–м шаге мы имеем распределения $Q_1^{(n-1)}$ и $Q_2^{(n-1)}$. Тогда на следующем шаге получаем

$$q_j^{(n)} = \sum_i q_i^{(n-1)} P_{ij} \ , \ p_j^{(n)} = \sum_i p_i^{(n-1)} P_{ij}$$

и

$$L_n = 2\sum_j{}^{+} \sum_i \bigl(q_i^{(n-1)} - p_i^{(n-1)}\bigr) \cdot P_{ij} \ .$$

Предположим, что $q_{j_0}^{(n)} - p_{j_0}^{(n)} \leqslant 0$ (если имеет место противоположное неравенство, то доказательство проводится аналогично). Тогда в $\sum_j^+$ все $j \neq j_0$ и $\sum_j^+ P_{ij} \leqslant (1-\lambda)$ для любого i. Оставляя в сумме по i только положительные слагаемые, получаем

$$L_n \leqslant 2\sum_j^+ \sum_i^+ \left(q_i^{(n-1)} - p_i^{(n-1)}\right) \cdot P_{ij} = 2\sum_i^+ \left(q_i^{(n-1)} - p_i^{(n-1)}\right) \sum_j^+ P_{ij}$$

$$\leqslant 2(1-\lambda)\sum_i^+ \left(q_i^{(n-1)} - p_i^{(n-1)}\right) = (1-\lambda)L_{n-1}\ .$$

Повторив это рассуждение несколько раз, приходим к неравенству

$$\begin{aligned} L_n \ &\leqslant (1-\lambda)^n L_0 = (1-\lambda)^n \sum_j \left|q_j^{(0)} - p_j^{(0)}\right| \\ &\leqslant (1-\lambda)^n \left(\sum_j q_j^{(0)} + \sum_j p_j^{(0)}\right) = 2(1-\lambda)^n\ . \end{aligned}$$

Это доказательство было предложено еще А.А. Марковым и усовершенствовано С.Н. Бернштейном.

Покажем теперь, что финальные вероятности π_j существуют. Пусть мы начинаем движение из состояния i, т.е. $P_i^{(0)} = 1$. Тогда $P_j^{(n)} = P_{ij}(n)$. Положим $m_j(n) = \min\limits_i P_{ij}(n)$. Последовательность $\{m_j(n)\}$ монотонно не убывает по n. Действительно

$$m_j(n) = \min_i \sum_k P_{ik} P_{kj}(n-1) \geqslant \min_i \sum_k P_{ik} m_j(n-1) = m_j(n-1)\ .$$

В силу монотонности существует

$$\pi_j = \lim_{n\to\infty} m_j(n)\ .$$

Обозначим $M_j(n) = \max_i P_{ij}$. По определению величин $m_j(n)$ и $M_j(n)$ имеем

$$m_j(n) \leqslant P_{ij}(n) \leqslant M_j(n)\ .$$

Из неравенства Бернштейна можно вывести (задача!), что

$$M_j(n) - m_j(n) \leqslant 2(1-\lambda)^n \to 0\ ,\ n \to \infty\ .$$

Отсюда следует, что существует

$$\lim_{n\to\infty} P_{ij}(n) = \pi_j \ .$$

Утверждения о стационарности финального распределения и единственности стационарного распределения сформулированы выше в виде задачи.

Сформулируем без доказательства аналог закона больших чисел для цепей Маркова.

Теорема 22 . *Пусть* $\big(X_n,\ n \geqslant 1\big)$ *— эргодическая цепь Маркова с финальным распределением* π. *Если* f *— вещественная функция на пространстве состояний* X, *то*

$$\frac{f(X_0) + \ldots f(X_{n-1})}{n} \xrightarrow{P} \sum_j f(x_j)\pi_j \ .$$

В частности, если $f(x_j) = 1$ *для* $j = i_0$ *и* $f(x_j) = 0$ *для* $j \neq i_0$, *то*

$$\frac{\nu_{i_0}(n)}{n} \xrightarrow{P} \pi_{i_0} \ , \ n \to \infty \ ,$$

где $\nu_{i_0}(n)$ *— число посещений состояния* i_0 *за первые* n *шагов.*

Замечание. Обозначим через τ_j момент первого возвращения в состояние j (после выхода из него же). Более тонкий анализ позволяет доказать, что $\pi_j = 1/M(\tau_j)$.

Всюду в этом параграфе мы рассматривали только конечные цепи Маркова с дискретным временем. Свойства цепей Маркова со счетным множеством состояний, а также с непрерывным временем будут рассматриваться позднее в курсе теории случайных процессов.

16.4 Примеры

Приведем несколько примеров, которые описывают ситуации, близкие к реальным.

Пример 1. Пусть мы имеем схему Бернулли с вероятностью успеха p в одном испытании. Это простейшая однородная цепь Маркова с пространством состояний $X = \{0, 1\}$ и матрицей перехода за один шаг

$$P = \begin{pmatrix} 1-p & p \\ 1-p & p \end{pmatrix} .$$

Пример 2. В предыдущем примере нас может интересовать появление различных сочетаний символов на последовательных шагах. Например, для двух последовательных шагов мы имеем четыре различные комбинации: $x_1 = (00)$, $x_2 = (01)$, $x_3 = (10)$, $x_4 = (11)$. В этом случае мы имеем дело с однородной цепью Маркова с матрицей перехода

$$P = \begin{pmatrix} 1-p & p & 0 & 0 \\ 0 & 0 & 1-p & p \\ 1-p & p & 0 & 0 \\ 0 & 0 & 1-p & p \end{pmatrix}$$

Пример 3. Компания по страхованию автомобилей провела обследование среди потенциальных клиентов и получила следующие результаты. Из водителей, попавших в аварию в течение года, 20% попали в аварию и в следующем году. Из тех, кто не попал в аварию в течение года, только 10% попали в аварию в следующем. Данную ситуацию можно описать с помощью однородной цепи Маркова с двумя состояниями: $x_1 = A$, $x_2 = \overline{A}$, где A обозначает событие, состоящее в том, что водитель попал в аварию, $\overline{A}$ — противоположное событие. По имеющимся данным получаем следующую матрицу перехода:

$$P = \begin{pmatrix} 0.2 & 0.8 \\ 0.1 & 0.9 \end{pmatrix}$$

Предположим, что в этом году среди застрахованных водителей 8% попали в аварию. Какова вероятность того, что случайно выбранный водитель попадет в аварию в следующем году?

Мы начинаем с начального распределения $P^{(0)} = (0.08, 0.92)$. Тогда

$$P^{(1)} = P^{(0)} \cdot P = (0.108, 0.892) ,$$

т.е. искомая вероятность равна 0.108.

Найдем стационарное распределение в этой задаче. Стационарное распределение $\pi = (\pi_1, \pi_2)$ удовлетворяет уравнению

$$\pi P = \pi$$

и условию нормировки

$$\pi_1 + \pi_2 = 1 \ .$$

Это приводит к следующей системе уравнений:

$$\begin{cases} 0.2\pi_1 + 0.1\pi_2 & = & \pi_1 \ , \\ 0.8\pi_1 + 0.9\pi_2 & = & \pi_2 \ , \\ \quad \pi_1 + \pi_2 & = & 1 \ . \end{cases}$$

Она имеет единственное решение $\pi_1 = 1/9, \ \pi_2 = 8/9$.

Пример 4. В стране Оз никогда не бывает подряд двух ясных дней. Если сегодня ясно, то завтра будет плохая погода (снег или дождь с равной вероятностью). Если сегодня снег (или дождь), то погода на следующий день не изменится с вероятностью $1/2$. Если она все–таки изменится, то лишь в половине случаев будет ясно. Построить модель изменения погоды в виде цепи Маркова с тремя состояниями, вычислить матрицу перехода и найти стационарное распределение.

Пример 5. Обучение в некотором колледже продолжается 4 года. Многолетняя статистика показывает, что студент каждого курса выбывает из колледжа с вероятностью p, остается на том же курсе с вероятностью q и переходит на следующий курс с вероятностью r. Образуем цепь Маркова, введя состояния: x_0 — выбыл, x_k — студент k-го курса, $k = \overline{1,4}$, x_5 — окончил колледж. Выписать матрицу переходных вероятностей.

Следующий пример имеет иллюстративный характер.

Пример 6. Рассмотрим цепь Маркова с семью состояниями, которая имеет следующую матрицу переходных вероятностей

$$P = \begin{pmatrix} 0 & 0.2 & 0.8 & 0 & 0 & 0 & 0 \\ 0.6 & 0 & 0.4 & 0 & 0 & 0 & 0 \\ 0 & 0.7 & 0 & 0.3 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.5 & 0.5 \\ 0 & 0 & 0 & 0 & 0 & 0.3 & .07 \\ 0 & 0 & 0 & 0.4 & 0.6 & 0 & 0 \\ 0 & 0 & 0 & 0.2 & 0.8 & 0 & 0 \end{pmatrix}$$

Нарисовать граф этой цепи и провести классификацию состояний.

Глава 17

Задачи для самостоятельного решения

В дополнение к тем задачам, что были приведены в основном тексте, мы предлагаем несколько задач стандартного типа, которые помогут студентам подготовиться к экзаменам.

1. Для социологического обследования из группы в 100 человек, среди которых 40 мужчин и 60 женщин, случайно отбирают 10 человек. Какова вероятность того, что среди них будет четверо мужчин?

2. Из отрезка [0,1] случайно и независимо выбирают две точки. Какова вероятность того, что расстояние между ними будет больше $1/2$?

3. Два человека играют в следующую игру: независимо друг от друга они выбирают одну из монет в 10 или 20 рублей и отмечают полученную сумму. Если эта сумма равна 20, то выигрывает первый, если она равна 40, то выигрывает второй, в противном случае фиксируется ничья. Какова вероятность того, что первый игрок не проиграет, если они выбирают монеты с одинаковыми вероятностями?

4. Есть две монеты: одна обыкновенная, а другая — с двумя гербами. Случайно выбрали одну из монет и подбросили два раза. Выпало два герба. Какова вероятность того, что это обыкновенная монета?

5. Две трети секретарей большого стенографического бюро имеют водительские права. Для участия в некоторой поездке случайно отобраны 4 секретаря. Какова вероятность того, что по крайней мере двое из них имеют водительские права?

6. Вероятность выигрыша в некоторой лотерее равна 0.3. Вы купили 100 билетов. Не менее какого числа выигрышей мы можем гарантировать с вероятностью не менее 0.9?

7. По некоторому каналу связи передаются сообщения. Вероятность ошибки при передаче одного символа равна 0.01. При каком количестве символов вероятность того, что мы имеем в сообщении не менее двух ошибок, не менее 0.95?

8. Симметричную монету подбрасывают 4 раза. ξ — число выпавших гербов. Найти распределение вероятностей с.в. ξ. Вычислить ее математическое ожидание, дисперсию и $P(\xi \geqslant 2)$.

9. Для случайной величины ξ из задачи 8 определим $\eta = (\xi - 2)^2$. Найти:

1) распределение η,
2) $M\eta,\ D(\eta)$.

10. Есть три монеты. Вероятность появления герба равна 0.3 для первой монеты, 0.5 — для второй и 0.6 — для третьей. Пусть ξ равна числу выпавших гербов. Найти распределение ξ и вычислить $M\xi$, $D(\xi)$.

11. Из отрезка [0,1] случайно и независимо выбирают две точки. Найти плотность распределения разности координат этих точек. Найти математическое ожидание и дисперсию этой случайной величины.

12. Для случайной величины ξ из задачи 11 определим $\eta = \xi^2$. Найти:

1) распределение η,
2) $M\eta,\ D(\eta)$.

13. Распределение дискретного случайного вектора $\xi = (\xi_1, \xi_2)$ за-

дано таблицей

$\xi_2 \setminus \xi_1$	-1	0	1
0	0.05	0.1	0.15
1	0.2	0.05	0.05
2	0.1	0.2	0.1

Найти:

1) распределение с.в. ξ_1 и ξ_2,

2) проверить независимость ξ_1 и ξ_2,

3) вероятность того, что $\xi_1\xi_2 = 0$,

4) коэффициент корреляции для ξ_1 и ξ_2,

5) совместное распределение случайного вектора $\eta = (\eta_1, \eta_2)$, где $\eta_1 = \xi_1,\ \eta_2 = \xi_1\xi_2$.

14. ξ_1 и ξ_2 независимы и имеют показательное распределение с параметром λ. Найти плотность распределения случайной величины $\eta = \xi_1 - \xi_2$.

15. Случайный вектор $\xi = (\xi_1, \xi_2)$ имеет распределение с плотностью

$$\rho_\xi(x_1, x_2) = \begin{cases} C(x_1 + 2x_2) & , \quad x_1, x_2 \in [0, 1], \\ 0 & , \quad \text{иначе}. \end{cases}$$

Найти: 1) константу C,

2) $\rho_{\xi_1}(x_1)$,

3) проверить независимость ξ_1 и ξ_2,

4) $P(\xi_1 > \xi_2)$.

5) $M(\xi_1), M(\xi_2), cov(\xi_1, \xi_2), \rho(\xi_1, \xi_2)$.

16. ξ_1 и ξ_2 независимы и одинаково распределены,

$$\rho_{\xi_1}(x) = \rho_{\xi_2}(x) = \begin{cases} 2x & , \quad x \in [0, 1]\ , \\ 0 & , \quad \text{иначе}. \end{cases}$$

Случайная величина $\eta = \xi_1 + 2\xi_2$. Найти $\rho_\eta(y)$.

17. Дискретная случайная величина ξ имеет распределение

x_n	-2	-1	0	1	2
p_n	0.2	0.1	0.2	0.4	0.1

С.в. $\eta = \xi^2$. Вычислить $M\xi, M\eta, D(\xi), D(\eta)$.

18. ξ_1 и ξ_2 – независимы, $M(\xi_1) = 2$, $M(\xi_2) = 3$, $D(\xi_1) = 4$, $D(\xi_2) = 9$. Случайная величина $\eta = 2\xi_1 + 4\xi_2 + 5$. Вычислить $M(\eta)$ и $D(\eta)$.

19. Вычислить математические ожидания и дисперсии для всех стандартных распределений.

20. С.в. ξ имеет $M(\xi) = a$ и $D(\xi) = \sigma^2 > 0$. Положим $\xi_0 = (\xi - a)/\sigma$. Доказать, что $M(\xi_0) = 0$, $D(\xi_0) = 1$.

21. С.в. ξ имеет нормальное распределение с параметрами $a = 2$, $\sigma^2 = 9$. Вычислить: $P(\xi \geqslant 5)$, $P(\xi < 8)$, $P(-7 < \xi < 11)$.

22. С.в. ξ имеет плотность распределения

$$\rho_\xi(x) = \begin{cases} 2x & , \quad x \in [0,1] \ , \\ 0 & , \quad \text{иначе.} \end{cases}$$

Вычислить $M(\xi^2)$ и $D(\xi^2)$.

23. С.в. ξ имеет показательное распределение с параметром $\lambda = 2$. Найти медиану и квантиль порядка 0.95 с.в. ξ.

24. Случайный вектор $\xi = (\xi_1, \xi_2)$ имеет распределение

$\xi_2 \backslash \xi_1$	-1	0	1	2
-1	0.05	0.15	0.05	0.05
0	0.1	0.1	0.1	0.1
1	0.05	0.1	0.15	0

Найти: 1) условное распределение вектора ξ при условии $\xi_1\xi_2 = 0$,

2) условное распределение ξ_1 при условии $\xi_2 = -1$,

3) условное математическое ожидание ξ_1 при условии $\xi_2 = -1$,

4) $\rho(\xi_1, \xi_2)$.

25. Случайный вектор $\xi = (\xi_1, \xi_2)$ имеет плотность распределения

$$\rho_\xi(x_1, x_2) = \begin{cases} C(2x_1 + 3x_2) & , \quad x_1, x_2 \in [0,1] \ , \\ 0 & , \quad \text{иначе.} \end{cases}$$

Найти: 1) константу C,

2) условную плотность ξ_1 при условии $\xi_2 = 1/2$,

3) условное математическое ожидание ξ_1 при условии $\xi_2 = 1/2$.

4) $\rho(\xi_1, \xi_2)$.

26. Вычислить характеристические и производящие функции для стандартных распределений.

27. Доказать, что сумма двух независимых с.в. ξ_1 и ξ_2, имеющих распределения Пуассона с параметрами λ_1 и λ_2, также имеет распределение Пуассона, но с параметром $\lambda_1 + \lambda_2$.

28. С.в. ξ имеет геометрическое распределение с параметром p. Положим $\eta_p = p \cdot \xi$. Найти предельное распределение для η_p при $p \to 0$.

29. Используя характеристическую функцию, вычислить математическое ожидание и дисперсию распределения Пуассона с параметром λ.

30. Случайные величины ξ_1, ξ_2, $\ldots$ — н.о.р., $M(\xi_1) = 1$, $D(\xi_1) = 4$. Положим $S_n = \xi_1 + \ldots \xi_n$. Найти минимальное n, при котором $P(S_n > 120) \geqslant 0.95$.

31. Ресторан под открытым небом работает только в погожие дни. Если в данной местности сегодня идет дождь, то с вероятностью 0.4 он будет и завтра, если же сегодня дождя нет, то завтра он будет с вероятностью 0.06. Описать эту ситуацию с помощью однородной цепи Маркова и найти матрицу перехода за один шаг. Найти стационарное распределение. Сегодня ресторан был закрыт. Найти вероятность того, что он будет закрыт еще два дня.

32. Однородная цепь Маркова имеет следующую матрицу перехода за один шаг

$$P = \begin{pmatrix} 1/3 & 1/3 & 0 & 1/3 & 0 & 0 \\ 0 & 1/2 & 1/2 & 0 & 0 & 0 \\ 0 & 1/2 & 1/2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1/3 & 2/3 \\ 0 & 0 & 0 & 1/3 & 0 & 2/3 \\ 0 & 0 & 0 & 2/3 & 1/3 & 0 \end{pmatrix}$$

Нарисовать соответствующий граф и провести классификацию состояний.

Часть II

Математическая статистика

Глава 1

Основная задача математической статистики. Статистическая структура

В курсе теории вероятностей мы всегда предполагали, что мы имеем некоторое фиксированное пространство $(\Omega, \mathcal{A}, P)$, т.е. исходная модель изучаемого явления задана однозначно. Задача состояла в том, чтобы изучить свойства этой модели. Основной проблемой являлось разработка методов вычисления вероятностей более сложных событий , исходя из вероятностей других, более простых событий, в рамках данной модели. В частности, при изучении случайных величин мы интересовались их распределением. Но мы оставляли в стороне вопрос о том, как была построена данная конкретная модель, как она была выбрана из множества других. Задачи такого типа относятся к другому разделу нашей науки о случайных явлениях – математической статистике. Эта наука имеет тот же предмет, что и теория вероятностей. Она изучает случайные эксперименты с неопределенным исходом, в которых выполнено свойство устойчивости частот, но интересуется несколько иными вопросами. В определенном смысле она дополняет теорию вероятностей, а та, в свою очередь, дает теоретическую основу для методов математической статистики.

Чтобы лучше понять соотношение этих двух дисциплин рассмотрим простой пример. Предположим, что мы проводим эксперимент, в котором независимо n раз подбрасываем некоторую монету. Если выпадет герб, то будем ставить 1, если цифра — то 0. Тогда исход такого эксперимента ω можно записать в виде $\omega = (\omega_1, \ldots, \omega_n)$, где ω_k — это результат k–ого подбрасывания. Множество Ω всех таких исходов определяет нам пространство элементарных событий, а класс $\mathcal{A}$ всех подмножеств Ω это класс событий, связанных с таким экспериментом. Если мы имеем основания предполагать, что монета симметричная, то, в сочетании с предположением о независимости испытаний, мы получаем, что $p(\omega) = 1/2^n$. Далее мы можем попытаться вычислить вероятности более сложных событий. Например, если A_m есть событие, состоящее в том, что в таком эксперименте появилось ровно m гербов, то, как мы уже знаем, вероятность этого события равна

$$(A_m) = C_n^m \cdot \left(\frac{1}{2}\right)^m \cdot \left(\frac{1}{2}\right)^{n-m} = C_n^m \cdot \frac{1}{2^n} \ .$$

Это типичная вероятностная задача: модель точно задана, необходимо что–либо вычислить в рамках этой фиксированной модели.

Посмотрим на этот эксперимент с другой стороны. Обычно мы не знаем точно свойств нашей монеты и не можем предполагать, что она симметричная. Тогда мы имеем те же Ω и $\mathcal{A}$, но для вероятностей элементарных исходов получаем только

$$P(\omega) = p^m(1-p)^{n-m} \ , \tag{1.1}$$

где m–число единиц в исходе ω, а вероятность $0 \leqslant p \leqslant 1$ появления единицы нам неизвестна. Таким образом, мы имеем целый набор $\mathcal{P}$ вероятностных мер на $(\Omega, \mathcal{A})$, которые могут реализоваться в нашем эксперименте. В этой ситуации мы можем поставить следующие вопросы. Пусть, для определенности, мы производим 100 подбрасываний нашей монеты и получаем 47 гербов.

1) Как на основе этой информации оценить неизвестную вероятность p?

2) Какова точность полученной оценки?

3) Сколько нужно произвести испытаний, чтобы добиться нужной точности оценки?

4) Можно ли считать, что $p = 1/2$?

Что общего во всех перечисленных вопросах? У нас есть некоторый априорно заданный класс возможных вероятностных моделей для нашего эксперимента, т.е. мы имеем не одну, а целый набор $\mathcal{P}$ вероятностных мер на $(\Omega, \mathcal{A})$. Мы хотели бы **на основе экспериментальных данных** уточнить наши знания об истинной вероятностной мере $P \in \mathcal{P}$, которая дает описание вероятностного механизма рассматриваемого эксперимента.

Рассмотрение этого примера приводит нас к определению **статистической структуры** и основной задаче математической статистики.

Определение 1 . Статистической структурой *называется тройка* $(\Omega, \mathcal{A}, \mathcal{P})$, *где*

1) Ω *— произвольное множество = пространство элементарных исходов,*

2) $\mathcal{A}$ *—* σ*–алгебра подмножеств* Ω *= события, доступные наблюдению,*

3) $\mathcal{P}$ *— некоторый набор вероятностных мер на* $(\Omega, \mathcal{A})$.

Примеры. 1) Рассмотренный выше пример приводит нас к модели $(\Omega, \mathcal{A}, \mathcal{P})$, где ω — набор длины n, состоящий из нулей и единиц, $\mathcal{A}$ — σ–алгебра всех подмножеств пространства Ω, $\mathcal{P}$ — набор вероятностных мер, которые задаются на элементарных исходах по формуле (1), $0 \le p \le 1$.

2) Пусть мы имеем одно измерение случайной величины ξ. В этом случае естественно взять $\Omega = R^1$, $\mathcal{A} = \mathcal{B}$ — σ–алгебра всех борелевских подмножеств R^1, а $\mathcal{P}$ — класс всевозможных распределений вероятностей для ξ на R^1 (заданных, например, с помощью функций распределения).

3) Иногда мы имеем какую–либо дополнительную информацию о распределении X. Например, если мы знаем, что X есть сумма большого числа маленьких слагаемых, то естественно предположить, что распределение X будет нормальным. В этом случае $\mathcal{P}$ есть класс всех нормальных распределений, вид которых известен, но есть два неизвестных параметра $a \in R^1,\ \sigma^2 > 0$.

4) Если мы производим n независимых измерений некоторой количественной характеристики в одинаковых условиях, то мы приходим к случайному вектору $X = \big(X_1, \dots, X_n\big)$, где с.в. $X_1, \dots, X_n$ —

н.о.р. В этом случае распределение X задается, например, совместной функцией распределения. Тогда $\Omega = R^n$, $\mathcal{A} = \mathcal{B}_n$ — класс всех борелевских подмножеств в R^n, $\mathcal{P}$ — класс всевозможных n–мерных функций распределения, соответствующих случаю н.о.р.с.в.

Если существует конечное число числовых параметров $(\theta_1, \ldots, \theta_m) = \theta$, $\theta \in \Theta \subset R^m$, с помощью которых удается занумеровать все распределения P из класса $\mathcal{P}$, то статистическая структура $(\Omega, \mathcal{A}, \mathcal{P})$ (класс распределений $\mathcal{P}$) называется **параметрической**. В противном случае мы имеем **непараметрическую** структуру. В рассмотренных выше примерах в случаях 1 и 3 соответствующие структуры — параметрические, а в случаях 2 и 4 — непараметрические.

Теперь мы готовы сформулировать **основную задачу математической статистики**: *на основе экспериментальных данных сузить класс $\mathcal{P}$ априорно заданных вероятностных мер до некоторого более узкого подкласса $\mathcal{P}_0 \subset \mathcal{P}$ (в идеале выбрать одно распределение).* Например, в случае 1 выше мы хотели бы, зная число m (экспериментальные данные!), оценить неизвестную вероятность p выпадения герба, т.е. выделить одно распределение.

Глава 2

Выборка и выборочные характеристики

2.1 Выборка

Напомним еще раз основную задачу: на основе экспериментальных данных уточнить наши знания о модели, т.е. о распределении вероятностей. В этом параграфе мы уточним, что понимается в нашей теории под экспериментальными данными. Всюду далее мы будем рассматривать в качестве основного примера с.в. ξ, распределение вероятностей P_ξ которой (или ее функция распределения $F_\xi(y)$) неизвестно или известно не полностью. Для того, чтобы получить информацию об этой с.в. ξ мы организуем случайный эксперимент, в котором реализуется при определенных условиях одно или несколько измерений этой случайной величины. В результате мы получаем несколько чисел. Набор $x = (x_1, \dots, x_N)$ этих чисел называется **выборкой** . Происхождение этого термина связана с тем, что в простейших ситуациях проведение эксперимента связано с реальным выбором из конечной совокупности (выбор шара в лотерее, выбор объекта для обследования и т.п.). Число N измерений называют **объемом выборки**. Способы получения выборок из конечных совокупностей для изучения статистических свойств этих совокупностей представляют из себя отдельную и очень непростую задачу. Они изучаются в курсе так называемой *общей статистики* и не будут рассматриваться в нашем курсе.

Мы уже обсуждали в курсе теории вероятностей, что нас интересует, обычно, не результат отдельного эксперимента, а что мы получаем **в среднем**, в длинной серии экспериментов. Точно также и в математической статистике нас будут интересовать свойства предложенных нам статистических процедур не при однократном их применении, а в среднем, когда они применяются много раз. Другой вариант — мы показываем, в большинстве реализаций нашего эксперимента предложенная статистическая процедура дает хороший результат. В обоих случаях утверждение носит вероятностный характер. Поэтому далее мы часто будем рассматривать отдельное измерение как случайную величину X_k, а выборку — как случайный вектор $X = (X_1, \ldots, X_N)$. Чтобы отличать выборку, рассматриваемую как набор конкретных чисел, от выборки — случайного вектора, в первом случае мы будем писать $x = (x_1, \ldots, x_N)$, а во втором — $X = (X_1, \ldots, X_N)$. Множество $\mathcal{X}$ всех возможных значений выборки X называется *выборочным пространством.*

Как мы говорили выше, выборка это результат измерений случайной величины ξ, которые производились в определенных условиях. Если в разных измерениях условия сильно меняются, то это означает, что мы имеем дело, вообще говоря, с разными объектами. Если условия эксперимента фиксированы столь сильно, что ничего не меняется, то имеем, по сути дела, одно измерение. Поэтому для «хорошего» эксперимента естественно предположить, что случайные величины $X_1, \ldots, X_N$ **одинаково распределены** (**однородная** выборка) и **независимы**. Когда выполнены оба свойства выборка X называется **повторной**. Чтобы уточнить дальнейшую терминологию рассмотрим следующий

Пример. Из *большой* совокупности людей (например, жителей некоторого города) отбирают некоторое количество их представителей для статистического обследования. Нас интересует некоторая количественная характеристика, измеряемая для каждого отобранного человека (например, возраст). В пределах данной совокупности эту характеристику мы рассматриваем как случайную величину ξ с некоторым распределением $P = P_\xi$. Отобрав N человек, мы получим N измерений $x_1, \ldots, x_N$ интересующей нас характеристики. Вся совокупность людей называется **генеральной совокупностью**. Поэтому мы говорим, что имеем выборку объема N из генеральной совокупности с распределением P некоторой случайной величины ξ или, более кратко, **выборку объема N из генеральной совокупности с распределением P**.

Эта терминология будет применяться и тогда, когда никакой реальной генеральной совокупности нет, а мы имеем случайную величину ξ с распределением P_ξ.

Если бы мы перебрали все элементы из генеральной совокупности, т.е. произвели бы так называемое *сплошное обследование*, то мы бы полностью восстановили распределение интересующей нас величины ξ. Но, обычно, это невозможно в силу большого размера генеральной совокупности.

В реальных задачах выборка содержит очень большое число измерений. Хотелось бы как–то их упорядочить, чтобы лучше понять, что же мы имеем. Первое, что обычно делают, это располагают все наблюдения в порядке возрастания

$$x_{(1)} \leqslant x_{(2)} \leqslant \cdots \leqslant x_{(N)} \tag{2.1}$$

или

$$\begin{array}{cccc} x_{(1)} < & x_{(2)} < & \cdots < & x_{(n)} \\ m_1 & m_2 & \cdots & m_n \end{array}$$

где во втором случае приведены только различные измерения, а m_k есть число появлений значения $x_{(k)}$. Полученный ряд неубывающих чисел называют *вариационным рядом.* Это те же самые числа, что и в выборке, но расположенные в порядке возрастания. В дальнейшем любую функцию от выборочных значений мы будем называть *статистикой.* Таким образом, статистика в этом специальном значении термина есть любая функция от наблюдений. Число $x_{(k)}$ называется *k–ой порядковой статистикой.* Статистики $x_{(1)}$ и $x_{(N)}$ — определяют *экстремальные значения* выборки. Они позволяют оценить интервал возможных значений с.в. ξ. *Размах* выборки $x_{(N)} - x_{(1)}$ оценивает величину разброса значений с.в. ξ. Но число различных значений в вариационном ряду может быть все еще достаточно большим. Чтобы сократить объем хранимой информации применяют *группировку* элементов выборки. Множество возможных значений с. в. ξ делят на несколько интервалов

$$-\infty = a_0 < a_1 < \ldots < a_k < a_{k+1} < \ldots < a_r < a_{r+1} = \infty$$

и подсчитывают сколько измерений попало в каждый интервал. Результаты записывают в виде таблицы ($n_0 + n_1 + \ldots + n_r = N$)

$a_0 \div a_1$	$a_1 \div a_2$	$\ldots$	$a_k \div a_{k+1}$	$\ldots$	$a_r \div a_{r+1}$
n_0	n_1	$\ldots$	n_k	$\ldots$	n_r

2.2 Эмпирическое распределение

Основной задачей математической статистики является оценка неизвестного распределения P_ξ случайной величины ξ на основе экспериментальных данных, т. е. используя выборку $x = (x_1, \ldots, x_N)$. Рассмотрим новую случайную величину ξ^*, множеством значений которой являются числа $x_1, \ldots, x_N$, каждому из которых приписывается вероятность $1/N$ (если некоторое значение появляется несколько раз, то его вероятность увеличивается в то же число раз). Случайная величина ξ^* является дискретной и ее распределение называется *эмпирическим распределением*, построенным по выборке x. По определению с. в. ξ^* для любого борелевского множества $B \subset R^1$

$$P(\xi^* \in B) = P_N^*(B) = \frac{\nu_N(B)}{N} \ , \tag{2.2}$$

где $\nu_N(B)$ — число элементов выборки, которые попали во множество B. Таким образом, P_N^* есть, фактически, относительная частота появления события ($\xi \in B$) в серии из N независимых и одинаковых испытаний.

Отметим, что при таком подходе мы рассматриваем полученную выборку как *модель* исходной генеральной совокупности. Для этой новой совокупности мы проводим уже *сплошное обследование*.

Напомним, что выборку можно рассматривать и как случайный вектор $X = (X_1, \ldots, X_N)$. В этом случае эмпирическое распределение P_N^* становится случайной величиной. В силу закона больших чисел (теорема Бернулли) для каждого фиксированного B мы имеем

$$P_N^*(B) \xrightarrow{P} P_\xi(B) = P(\xi \in B) \ , N \to \infty \ . \tag{2.3}$$

Распределение случайной величины часто задают с помощью функции распределения. Для ее оценки мы будем использовать *эмпирическую функцию распределения*. По определению

$$F_\xi(y) = P(\xi < y) \ .$$

Тогда, в силу формулы (2), для эмпирической функции распределения получаем

$$F_N^*(y) = \frac{N(y)}{N} \ , \tag{2.4}$$

где $N(y)$ — число элементов x_k выборки x, для которых $x_k < y$. Вновь рассматривая выборку как случайный вектор, получаем, что

$$F_N^*(y) \xrightarrow{P} F_\xi(y) \ , \ N \to \infty \ , \tag{2.5}$$

для каждого фиксированного y. На самом деле справедлив гораздо более сильный результат, известный как *теорема Гливенко-Кантелли*:

$$P\left\{\lim_{N\to\infty} \sup_y \left|F_N^*(y) - F_\xi(y)\right| = 0\right\} = 1 \ .$$

В тех случаях, когда априори известно, что распределение является непрерывным, хотелось бы и в качестве оценки получить непрерывное распределение. Тогда применяют так называемые *ядерные оценки*. Пусть Q есть некоторое фиксированное распределение вероятностей, обладающее нужными свойствами, например, абсолютно непрерывное. Тогда в качестве оценки неизвестного распределения P_ξ мы берем

$$P^*(B) = \frac{1}{N} \sum_{k=1}^{N} Q(B - x_k) \ .$$

Для оценки плотности распределения используют *гистограмму*, которая определятся следующим образом. Пусть мы сгруппировали все элементы x_i выборки x в r интервалов, длина k–го интервала равна Δ_k, а N_k есть число элементов выборки, попавших в k–тый интервал. По определению гистограмма есть функция $\rho_N^*(y)$, определяемая по правилу

$$\rho_N^*(y) := \frac{N_k}{N \cdot \Delta_k} \ , \tag{2.6}$$

если y принадлежит k–му интервалу, и равная нулю в противном случае. Легко проверить, $\rho_N^*(y)$ обладает следующими свойствами:

1) $\rho_N^*(y) \geqslant 0$,
2) $\int\limits_{-\infty}^{\infty} \rho_N^*(y)dy = 1$,
3) $\int\limits_a^b \rho_N^*(y)dy = P(a \leqslant \xi^* < b)$.

Таким образом, $\rho_N^*(y)$ обладает всеми основными свойствами плотности. Более того, если при $N \to \infty$ мы имеем $r \to \infty$, $\max\limits_k \Delta_k \to 0$, но некоторым согласованным образом, то

$$P(\lim_{N\to\infty} \max_{-\infty<y<\infty} \left|\rho_N^*(y) - \rho_\xi(y)\right| = 0) = 1 \ . \tag{2.7}$$

В некоторых задачах мы знаем, что плотность является достаточно гладкой функцией, и хотим, чтобы оценка плотности обладала такими же свойствами. В этом случае вновь используют ядерные оценки. Пусть $\rho(y)$ — некоторая заданная функция («ядро»), которая является плотностью. По выборке $x = (x_1, \ldots, x_N)$ построим функцию

$$\rho_N(y) = \frac{1}{N} \sum_{k=1}^{N} \frac{1}{b_N} \rho\left(\frac{y - x_k}{b_N}\right) \ , \tag{2.8}$$

где $\{b_n\}$ — некоторая последовательность положительных чисел. Если ядро $\rho(y)$ и последовательность $\{b_n\}$ обладают некоторыми свойствами, то мы вновь можем гарантировать сходимость $\rho_N(y)$ к $\rho(y)$ при $N \to \infty$.

2.3 Выборочные характеристики

Ранее, при изучении случайных величин, мы отмечали, что распределение является наиболее полной вероятностной характеристикой случайной величины. Но во многих задачах бывает затруднительно (а иногда и не нужно) дать полное описание распределения. В этом случае при описании распределения используют небольшое число числовых характеристик, отражающих те или иные стороны распределения. Обычно для этой цели используют моменты (наиболее часто математическое ожидание и дисперсию) и квантили. Эти характеристики определяются через распределение случайной величины и представляют собой некоторые функционалы от этого распределения. Подставляя в них эмпирическое распределение, можно получить их эмпирические аналоги, которые называются *выборочными* или *эмпирическими характеристиками*.

Рассмотрим несколько примеров. Пусть мы имеем случайную величину ξ. Напомним, что ее момент порядка m относительно точки a определяется по правилу

$$\nu_m(a) = M\Big[(\xi - a)^m\Big] = \int\limits_{-\infty}^{\infty} (y-a)^m \, dF_\xi(y) \ .$$

В частности, мы можем определить математическое ожидание $M\xi$, характеризующее «центр» распределения, и дисперсию

$$D(\xi) = M\Big[(\xi - M\xi)^2\Big] = M(\xi^2) - (M\xi)^2 \ ,$$

характеризующую «разброс» или «рассеяние» распределения относительно этого центра. Их выборочными аналогами будут: *выборочный момент* (при $a = 0$)

$$\hat{\nu}_m = \frac{1}{N} \sum_{k=1}^{N} x_k^m \ ,$$

выборочное среднее

$$\bar{x} = \frac{1}{N} \sum_{k=1}^{N} x_k$$

и *выборочная дисперсия*

$$S^2 = \frac{1}{N}\sum_{k=1}^{N} x_k^2 - \bar{x}^2 = \frac{1}{N}\sum_{k=1}^{N}(x_k - \bar{x})^2 \ .$$

Следующая теорема дает описание некоторых свойств выборочных моментов. Она есть непосредственное следствие закона больших чисел и центральной предельной теоремы.

Теорема 1 . *Если $M\big(|\xi|^{2m}\big) < \infty$ и $\hat{\nu}_m$ — выборочный момент порядка m, построенный по повторной выборке объема N, то*

1) $M\big(\hat{\nu}_m\big) = \nu_m$;

2) $\hat{\nu}_m \xrightarrow{P} \nu_m = M\big(\xi^m\big),\ N \to \infty$,

в частности,

$$\bar{X} \xrightarrow{P} a = M\xi \ , \ S^2 \xrightarrow{P} \sigma^2 = D(\xi) \ ;$$

3) $\hat{\nu}_m$ имеет асимптотически нормальное распределение со средним ν_m и дисперсией $\big(\nu_{2m} - \nu_m^2\big)/N$.

Доказательство. Напомним еще раз, что повторная выборка есть случайный вектор $(X_1, \dots, X_N)$, состоящий из независимых и одинаково распределенных случайных величин. Учитывая это и условия теоремы, имеем

1) $\big(X_1^m, \dots, X_N^m\big)$ — независимы,

2) $\big(X_1^m, \dots, X_N^m\big)$ — одинаково распределены,

3) существуют конечные математическое ожидание $M\big(X_1^m\big) = \nu_m$ и дисперсия $D\big(X_1^m\big) = M\big(X_1^{2m}\big) - \Big(M\big(X_1^m\big)\Big)^2$.

Тогда

$$M(\hat{\nu}_m) = M\left(\frac{1}{N}\sum_{k=1}^{N} X_k^m\right) = \frac{1}{N}\sum_{k=1}^{N} M(X_k^m) = \frac{1}{N}\cdot N \cdot \nu_m = \nu_m \ .$$

В силу закона больших чисел для н.о.р.с.в. получаем

$$\frac{1}{N}\sum_{k=1}^{N} X_k^m \xrightarrow{P} M(X_1^m) = \nu_m \ .$$

Далее,

$$D\left(\frac{1}{N}\sum_{k=1}^{N}X_k^m\right) = D(X_1^m)/N = \left(\nu_{2m}-\nu_m^2\right)/N \ .$$

В силу ЦПТ для н.о.р.с.в. случайная величина

$$\frac{\hat{\nu}_m - M(\hat{\nu}_m)}{\sqrt{D(\hat{\nu}_m)}} = \frac{\hat{\nu}_m - \nu_m}{\sqrt{(\nu_{2m}-\nu_m^2)/N}}$$

имеет асимптотически при $N \to \infty$ стандартное нормальное распределение

Квантиль порядка p, $0 < p < 1$, определяется как такое число y_p, для которого

$$P(\xi \leqslant y_p) \geqslant p \ , \ P(\xi \geqslant y_p) \geqslant 1-p \ .$$

Если распределение не имеет дискретных точек, то это сводится к уравнению

$$P(\xi < y_p) = F_\xi(y_p) = p \ .$$

В частности, *медиана* $y_{1/2}$ распределения с.в. ξ определяется из соотношения

$$P(\xi < y_{1/2}) = \frac{1}{2} \ ,$$

где мы предположили, для простоты, что распределение не имеет дискретных точек. Медиана является еще одной характеристикой, оценивающей центр распределения. *Выборочная квантиль* порядка p определяется по правилу

$$\hat{y}_p = x_{\left([Np]\right)} \ .$$

Чаще других квантилей используется медиана $y_{1/2}$. Здесь для улучшения свойств оценки используется следующее определение:

$$\hat{y}_{1/2} = \begin{cases} x_{(n+1)} & , \quad N = 2n+1 \ , \\ \frac{1}{2}(x_{(n)} + x_{(n+1)}) & , \quad N = 2n \ . \end{cases}$$

Задача. Если y_p есть точка непрерывности и точка роста ф.р. $F_\xi(y)$, то

$$\hat{y}_p \xrightarrow{P} y_p \ , \ N \to \infty \ .$$

Глава 3

Точечные оценки параметров

3.1 Параметрические статистические структуры

Как уже отмечалось выше, всюду далее мы рассматриваем случай, когда имеем некоторую случайную величину ξ, распределение которой P_ξ (или ее ф.р. F_ξ) неизвестно или известно не полностью. В этом и нескольких последующих параграфах будут рассматриваться так называемые *параметрические структуры* $\big(\Omega, \mathcal{A}, \mathcal{P}\big)$, для которых семейство распределений $\mathcal{P}$ является *параметрическим*, т. е. существует конечное число вещественных параметров $\theta_1, \dots, \theta_m$, с помощью которых можно занумеровать все распределения P из нашего семейства $\mathcal{P}$. В дальнейшем мы будем записывать параметры в виде одного многомерного параметра $\theta = (\theta_1, \dots, \theta_m)$, который пробегает некоторое множество $\Theta \subset R^m$. Тогда мы можем записать $\mathcal{P} = \big\{P_\theta\ ,\ \theta \in \Theta\big\}$. Для нашей задачи, когда мы изучаем одну случайную величину ξ, в качестве Ω мы выбираем R^1, $\mathcal{A} = \mathcal{B}$ есть σ–алгебра борелевских подмножеств. Тогда $\mathcal{P}$ есть некоторое семейство возможных распределений случайной величины ξ. Ниже мы рассматриваем несколько примеров параметрических семейств распределений.

Примеры. 1. **Распределение Бернулли** — это распределение случайной величины ξ, которая принимает два значения: 1 — с вероятностью p и 0 — с вероятностью $1 - p$. Число $p \in (0, 1)$ является параметром распределения.

Для этого распределения

$$M(\xi) = p\ ,\ D(\xi) = p(1 - p)\ .$$

2. **Биномиальное распределение** — это распределение случайной величины ξ, которая принимает значения $0, 1, \ldots, n$ с вероятностями

$$P(\xi = m) = C_n^m p^m (1-p)^{n-m} \ .$$

Целое число $n \geqslant 1$ и $p \in (0,1)$ являются параметрами распределения.

Для этого распределения

$$M(\xi) = np \ , \ D(\xi) = np(1-p) \ .$$

3. **Распределение Пуассона** — это распределение случайной величины ξ, которая принимает значения $0, 1, \ldots$ с вероятностями

$$P(\xi = m) = \frac{\lambda^m}{m!} e^{-\lambda} \ .$$

Число $\lambda > 0$ является параметром распределения.

Для этого распределения

$$M(\xi) = \lambda \ , \ D(\xi) = \lambda \ .$$

4. **Геометрическое распределение** — это распределение случайной величины ξ, которая принимает значения $1, 2 \ldots$ с вероятностями

$$P(\xi = m) = p(1-p)^{m-1} \ .$$

Число $p \in (0,1)$ является параметром распределения.

Для этого распределения

$$M(\xi) = \frac{1}{p} \ , \ D(\xi) = \frac{1-p}{p^2} \ .$$

5. **Отрицательное биномиальное распределение** — это распределение случайной величины ξ, которая принимает значения $0, 1, 2 \ldots$ с вероятностями

$$P(\xi = m) = C_m^{m+r-1} p^r (1-p)^m \ .$$

Целое число $r \geqslant 1$ и $p \in (0,1)$ являются параметрами распределения. При $r = 1$ получаем геометрическое распределение.

Для этого распределения

$$M(\xi) = \frac{1-p}{p} r \ , \ D(\xi) = r\frac{1-p}{p^2} \ .$$

6. **Равномерное распределение** на отрезке (a, b) — это распределение случайной величины ξ, которая имеет плотность распределения следующего вида:

$$\rho_\xi(y) = \begin{cases} \frac{1}{b-a} & , \quad y \in (a,b) \ , \\ 0 & , \quad \text{иначе.} \end{cases}$$

Числа a и b, $a < b$, являются параметрами распределения.

Для этого распределения

$$M(\xi) = \frac{a+b}{2} \ , \ D(\xi) = \frac{(b-a)^2}{12} \ .$$

7. **Показательное распределение** — это распределение случайной величины ξ, которая имеет плотность распределения следующего вида:

$$\rho_\xi(y) = \begin{cases} \lambda e^{-\lambda y} & , \quad y > 0 \ , \\ 0 & , \quad \text{иначе.} \end{cases}$$

Число $\lambda > 0$ является параметром распределения.

Для этого распределения

$$M(\xi) = \frac{1}{\lambda} \ , \ D(\xi) = \frac{1}{\lambda^2} \ .$$

8. **Гамма–распределение** — это распределение случайной величины ξ, которая имеет плотность распределения следующего вида:

$$\rho_\xi(y) = \begin{cases} \frac{\beta^\alpha}{\Gamma(\alpha)} y^{\alpha-1} e^{-\beta y} & , \quad y > 0 \ , \\ 0 & , \quad \text{иначе.} \end{cases}$$

Числа $\alpha > 0$ и $\beta > 0$ являются параметрами распределения. При $\alpha = 1$ и $\beta = \lambda$ получаем показательное распределение.

Для этого распределения

$$M(\xi) = \frac{\alpha}{\beta}\ ,\ D(\xi) = \frac{\alpha}{\beta^2}\ .$$

9. **Бета–распределение** — это распределение случайной величины ξ, которая имеет плотность распределения следующего вида:

$$\rho_\xi(y) = \begin{cases} \frac{\Gamma(r+s)}{\Gamma(r)\Gamma(s)} y^{r-1}(1-y)^{s-1} & ,\quad 0 < y < 1\ , \\ 0 & ,\quad \text{иначе.} \end{cases}$$

Числа $r > 0$ и $s > 0$ являются параметрами распределения. При $r = s = 1$ получаем равномерное на [0,1] распределение.

Для этого распределения

$$M(\xi) = \frac{r}{r+s}\ ,\ D(\xi) = \frac{rs}{(r+s)^2(r+s+1)}\ .$$

10. **Нормальное распределение** — это распределение случайной величины ξ, которая имеет плотность распределения следующего вида:

$$\rho_\xi(y) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y-a)^2}{2\sigma^2}}\ ,\ y \in R^1\ .$$

Числа $a \in R^1$ и $\sigma^2 > 0$ являются параметрами распределения.

Для этого распределения

$$M(\xi) = a\ ,\ D(\xi) = \sigma^2\ .$$

11. **Распределение Коши** — это распределение случайной величины ξ, которая имеет плотность распределения следующего вида:

$$\rho_\xi(y) = \frac{1}{\pi}\ \frac{1}{1 + \left(\frac{y-a}{b}\right)^2}\ .$$

Числа $a \in R^1$ и $b > 0$ являются параметрами распределения.

Это распределение не имеет математического ожидания и дисперсии.

12. **Сдвиг–масштабное семейство распределений**.

Пусть мы имеем фиксированную функцию распределения $F(y)$. Семейство распределений

$$\mathcal{P} = \left\{ F(y, a, b) := F\left(\frac{y-a}{b}\right) \right\}$$

называется **сдвиг–масштабным**. Числа $a \in R^1$ и $b > 0$ называются *параметрами сдвига и масштаба* соответственно.

Если существует плотность $\rho(y) = \frac{d}{dy}F(y)$, то

$$\rho(y, a, b) = \frac{1}{b}\,\rho\left(\frac{y-a}{b}\right)\ .$$

Для этого распределения

$$M(\xi) = b \cdot a_0 + a\ ,\ D(\xi) = b^2\sigma^2\ ,$$

где a_0 и σ^2 есть математическое ожидание и дисперсия для распределения из нашего семейства с параметрами $a = 0$, $b = 1$ (если они существуют).

3.2 Точечные оценки параметров и их свойства

Основной задачей математической статистики является выбор неизвестного распределения, используя имеющиеся экспериментальные данные. Если мы имеем случайную величину ξ с неизвестным распределением из некоторого параметрического семейства $\mathcal{P} = \left\{P_\theta,\ \theta \in \Theta \subset \mathbf{R}^{\mathbf{m}}\right\}$, то основная задача сводится к оценке неизвестных параметров. Экспериментальными данными в нашей задаче является повторная выборка $x = (x_1, \ldots, x_N)$. Что значит построить оценку? В конкретной ситуации мы должны для заданного набора чисел $x_1, \ldots, x_N$ указать соответствующее им значение параметра θ. Рассматривая задачу более общо, с теоретической точки зрения, мы должны указать правило, по которому каждой выборке $x = (x_1, \ldots, x_N)$ сопоставляется некоторое значение параметра θ. Это приводит нас к следующему определению.

Определение. Оценкой параметра θ называется произвольная функция $\hat{\theta}$ из выборочного пространства $\mathcal{X}$ в пространство параметров $\Theta \subset \mathbf{R^m}$.

Выше мы любую функцию от выборки назвали статистикой. Если она используется в целях оценивания неизвестного параметра, то она называется **оценкой**. В этом определении в качестве оценки предлагается любая функция от выборки. Ясно, что не любая такая функция будет «хорошей» оценкой. Далее мы рассмотрим некоторые свойства, которыми должны обладать «хорошие» оценки. Как уже отмечалось ранее, свойства оценок будут изучаться «в среднем», когда предложенная процедура применяется многократно в длинной серии испытаний. В этом случае мы рассматриваем выборку как случайный вектор $X = (X_1, \ldots, X_N)$. Но тогда оценка $\hat{\theta}$ является случайной величиной $\hat{\theta} = \hat{\theta}_N(X) = \hat{\theta}_N(X_1, \ldots, X_N)$.

При вычислении математического ожидания или вероятности мы должны указать какое именно распределение P из нашего семейства $\mathcal{P}$ мы используем. Для этого мы будем использовать обозначения M_θ и P_θ, которые означают, что мы вычисляем математическое ожидание и вероятность в предположении, что истинным значением параметра является θ.

Ниже мы перечислим некоторые свойства, которые обычно требуют от «хороших» оценок. Для простоты обозначений мы предполагаем, что мы имеем один неизвестный параметр θ, то есть $m = 1$.

Определение. Оценка $\hat{\theta} = \hat{\theta}_N(X_1, \ldots, X_N)$ параметра θ называется **несмещенной**, если

$$M_\theta\Big(\hat{\theta}_N(X)\Big) = \theta \qquad \forall \theta \in \Theta.$$

В силу сформулированной в §2 теоремы мы видим, что выборочные моменты являются несмещенными оценками истинных моментов. Например, $M(\bar{x}) = a = M(\xi)$. То же самое справедливо относительно линейных функций от моментов. Но в общем случае, если параметр есть некоторая функция от начального момента, то, подставляя в нее выборочный момент, мы получим смещенную оценку.

На практических занятиях мы выясним, что выборочная дисперсия

$$S^2 = \frac{1}{N}\sum_{k=1}^{N}(X_k - \bar{X})^2$$

является смещенной оценкой для $\sigma^2 = D(\xi)$, так как

$$M(S^2) = \frac{N-1}{N}\sigma^2.$$

Несмещенной оценкой для σ^2 является так называемая **исправленная выборочная дисперсия**

$$S_1^2 = \frac{1}{N-1}\sum_{k=1}^{N}(X_k - \bar{X})^2.$$

Определение. Последовательность оценок $\left\{\hat{\theta}_N(X),\ \ N \geq N_0\right\}$ параметра θ называется **состоятельной**, если

$$\hat{\theta}_N(X) \xrightarrow{P_\theta} \theta \qquad \forall\theta \in \Theta.$$

Иногда слово последовательность опускают и говорят, что оценки $\hat{\theta}_N$ состоятельны, хотя это и не точно. Вновь обращаясь к теореме из §2, мы видим, что выборочные моменты являются состоятельными оценками истинных моментов. Например, выборочное среднее $\bar{X}$ является состоятельной оценкой математического ожидания a, а оба варианта выборочной дисперсии S^2 и S_1^2 являются состоятельными оценками для дисперсии σ^2. Приведем более сложный пример.

Задача. Пусть случайная величина ξ имеет непрерывную и строго монотонную функцию распределения. Тогда выборочная медиана $\hat{y}_{1/2}$ является несмещенной и состоятельной оценкой для истинной медианы.

Сформулированные выше два свойства оценок ничего не говорят о точности полученных оценок.

Определение. Оценка $\hat{\theta} = \hat{\theta}(X)$ называется **эффективной** или **несмещенной оценкой с наименьшей дисперсией**, если

1. $M_\theta(\hat{\theta}) \equiv \theta \qquad \forall\theta$, то есть это несмещенная оценка;

2. $D_\theta(\hat{\theta}) = M_\theta(\hat{\theta} - \theta)^2 \leqslant M_\theta(\tilde{\theta} - \theta)^2 = D_\theta(\tilde{\theta}) \qquad \forall\theta,$

где $\tilde{\theta}$ — любая другая несмещенная оценка для параметра θ.

Эффективная оценка является **оптимальной в среднем квадратическом** оценкой в классе несмещенных оценок (см. тему «Гильбертово пространство случайных величин»).

Позже мы покажем, что выборочное среднее $\bar{X}$ является эффективной оценкой параметра a в классе нормальных распределений.

В реальных задачах нам необходимо не просто указать оценку неизвестного параметра, но и уметь оценивать вероятности различных отклонений этой оценки от истинного параметра. В этом случае оказывается полезным следующее понятие.

Определение. Последовательность оценок $\left\{\hat{\theta}_N,\ N \geqslant N_0\right\}$ называется **асимптотически нормальной**, если $\forall N \geqslant N_0$ существуют константы $A_N(\theta) \in \mathbf{R^1}$ и $B_N(\theta) > 0$ такие, что случайная величина

$$\frac{\hat{\theta}_N - A_N(\theta)}{B_N(\theta)}$$

имеет асимптотически стандартное нормальное распределение, то есть ее функция распределения сходится к функции распределения стандартного нормального закона.

Теорема из §2 говорит нам, что выборочные моменты являются асимптотически нормальными оценками истинных моментов. В частности, выборочное среднее $\bar{X}$ является асимптотически нормальной оценкой для математического ожидания.

Замечание. Иногда складывается ложное впечатление, что выборочное среднее $\bar{X}$ является некоторой универсальной оценкой центра распределения, которая хороша во всех случаях. Рассмотрим семейство распределений Коши с плотностями

$$\rho(y, \theta) = \frac{1}{\pi} \cdot \frac{1}{1 + (y - \theta)^2}, \quad y \in \mathbf{R^1}.$$

Случайная величина ξ с таким распределением не имеет математического ожидания. Если $X = (X_1, \ldots, X_N)$ есть повторная выборка из такого распределения, то выборочное среднее

$$\bar{X} = \frac{1}{N} \sum_{k=1}^{N} X_k$$

одинаково распределено с ξ. Отсюда мы видим, что $\bar{X}$ не является несмещенной, а также состоятельной и асимптотически нормальной. В этом случае хорошей оценкой для центра распределения θ является выборочная медиана $\hat{y}_{1/2}$.

3.3 Неравенство Рао–Крамера

Выше мы дали определение эффективной оценки и отметили, что она является оптимальной в среднем квадратическом в классе всех несмещенных оценок. Но остался открытым вопрос о том, как в конкретной задаче найти такую оптимальную оценку. Из-за недостатка времени мы не будем рассматривать методы построения эффективных оценок. Вместо этого мы докажем некоторый результат, который позволит для конкретной оценки доказать, что она является эффективной.

Напомним определение эффективной оценки: $\hat{\theta}_N = \hat{\theta}_N(X)$ является **эффективной**, если она несмещенная и для любой другой несмещенной оценки $\tilde{\theta}_N = \tilde{\theta}_N(X)$ мы имеем

$$D_\theta(\hat{\theta}_N) \leqslant D_\theta(\tilde{\theta}_N).$$

Предположим, что мы нашли некоторую нижнюю границу для дисперсий всех несмещенных оценок. Если эта граница достигается на некоторой конкретной несмещенной оценке, то она и будет эффективной. Ниже мы получим такую нижнюю границу.

Введем вначале некоторую вспомогательную величину. Пусть случайная величина ξ имеет плотность $\rho(y, \theta)$, где $\theta \in \Theta \subset \mathbf{R^1}$ — скалярный параметр распределения.

Определение. Величина

$$I(\theta) = \int\limits_{-\infty}^{\infty} \left(\frac{\partial \ln \rho(y,\theta)}{\partial \theta} \right)^2 \rho(y,\theta)\, dy = M_\theta \left(\frac{\partial \ln \rho(\xi,\theta)}{\partial \theta} \right)^2$$

называется **информацией по Фишеру** о параметре θ, содержащейся в одном измерении случайной величины ξ.

Если мы имеем повторную выборку $X = (X_1, \ldots, X_N)$ из генеральной совокупности с плотностью распределения $\rho(y,\theta)$, то ее совместная плотность распределения имеет вид

$$\mathcal{L}(x,\theta) = \rho(x_1,\theta) \cdot \ldots \cdot \rho(x_N,\theta).$$

Функция $\mathcal{L}(x,\theta)$ как функция от θ при фиксированном x называется **функцией правдоподобия**. Информацией по Фишеру о параметре θ, содержащейся в выборке $X = (X_1, \ldots, X_N)$, называется величина

$$I_N(\theta) = \int\limits_{\mathbf{R^N}} \left(\frac{\partial \ln \mathcal{L}(x,\theta)}{\partial \theta} \right)^2 \mathcal{L}(x,\theta)\, dx_1 \ldots dx_N = M_\theta \left(\frac{\partial \ln \mathcal{L}(X,\theta)}{\partial \theta} \right)^2.$$

Нетрудно показать, что для повторной выборки имеет место соотношение

$$I_N(\theta) = N \cdot I(\theta).$$

Если ξ имеет дискретное распределение, то в определении информации по Фишеру плотность нужно заменить на вероятность, а интеграл на сумму.

Сформулированная ниже теорема справедлива, если распределение вероятностей случайной величины ξ удовлетворяет некоторым условиям регулярности. В силу громоздкости этих условий мы не будем выписывать их в явном виде. Отметим только, что эти условия гарантируют нам возможность дифференцирования по параметру под знаком интеграла (математического ожидания).

Мы будем рассматривать несколько более общую задачу, когда оценивается некоторая вещественная функция $g(\theta)$ от скалярного параметра $\theta \in \Theta \subset \mathbf{R^1}$.

Теорема 2 *Пусть $\hat{g}_N = \hat{g}_N(X)$ — некоторая несмещенная оценка для вещественной функции $g(\theta)$ от параметра θ, построенная по повторной выборке $X = (X_1, \ldots, X_N)$ из генеральной совокупности с функцией распределения $F(y, \theta)$, удовлетворяющей условиям регулярности. Тогда имеет место неравенство*

$$D_\theta(\hat{g}_N) \geqslant \frac{\left[g'(\theta)\right]^2}{N \cdot I(\theta)}. \tag{1}$$

В частности, если $g(\theta) \equiv \theta$, то

$$D_\theta \geqslant \frac{1}{I_N(\theta)}\ ,\ \forall\theta.$$

Последние неравенства называются **неравенствами Рао–Крамера**.

Доказательство. Из свойств плотности и определения несмещенной оценки получаем:

$$M_\theta(1) = \int\limits_{\mathbf{R^N}} \mathcal{L}(x, \theta)\, dx_1 \ldots dx_N \equiv 1, \tag{2}$$

$$M_\theta(\hat{g}_N) = \int\limits_{\mathbf{R^N}} \hat{g}_N(x) \cdot \mathcal{L}(x, \theta)\, dx = g(\theta). \tag{3}$$

Дифференцируя (2) и (3) по θ (здесь нужны условия регулярности), получаем

$$0 = \int\limits_{\mathbf{R^N}} \frac{\partial}{\partial\theta}\mathcal{L}(x, \theta)\, dx, \tag{4}$$

$$g'(\theta) = \int\limits_{\mathbf{R^N}} \hat{g}_N(x) \frac{\partial}{\partial\theta}\mathcal{L}(x, \theta)\, dx. \tag{5}$$

Умножим (4) на $g(\theta)$ и вычтем почленно из (5):

$$\begin{aligned} g'(\theta) &= \int\limits_{\mathbf{R^N}} \big(\hat{g}_N(x) - g(\theta)\big) \cdot \tfrac{\partial}{\partial\theta}\mathcal{L}(x, \theta)\, dx = \\ &= \int\limits_{\mathbf{R^N}} \big(\hat{g}_N(x) - g(\theta)\big) \cdot \tfrac{\frac{\partial}{\partial\theta}\mathcal{L}(x,\theta)}{\mathcal{L}(x,\theta)} \cdot \mathcal{L}(x, \theta)\, dx = \\ &= \int\limits_{\mathbf{R^N}} \big(\hat{g}_N(x) - g(\theta)\big) \cdot \tfrac{\partial}{\partial\theta} \ln \mathcal{L}(x, \theta) \cdot \mathcal{L}(x, \theta)\, dx = \\ &= M_\theta \left[\big(\hat{g}_N(X) - g(\theta)\big) \cdot \tfrac{\partial}{\partial\theta} \ln \mathcal{L}(X, \theta)\right]. \end{aligned} \tag{6}$$

Применим к (6) неравенство Коши–Буняковского:

$$\begin{aligned}
&\big(g'(\theta)\big)^2 = \left[M_\theta\big(\hat g_N(X) - g(\theta)\big)\left(\tfrac{\partial}{\partial\theta}\ln\mathcal{L}(X,\theta)\right)\right]^2 \leqslant \\
&\leqslant M_\theta\big(\hat g_N(X) - g(\theta)\big)^2 \cdot M_\theta\left(\tfrac{\partial}{\partial\theta}\ln\mathcal{L}(X,\theta)\right)^2 = \\
&= D_\theta(\hat g_N)\cdot I_N(\theta).
\end{aligned}$$

Что и требовалось доказать.

Пример. ξ имеет нормальное распределение с параметрами a и σ^2. Нас интересует параметр $\theta = a$. В качестве оценки a предлагается выборочное среднее

$$\bar X = \frac{1}{N}\sum_{k=1}^{N} X_k.$$

Покажем, что это эффективная оценка. В нашем случае $g(\theta) = \theta$, $g'(\theta) \equiv 1$, $I(\theta) = 1/\sigma^2$. Тогда

$$D_\theta(\bar X) = \frac{\sigma^2}{N} = \frac{1}{N/\sigma^2} = \frac{1}{I_N(\theta)}.$$

Таким образом, нижняя граница для дисперсий несмещенных оценок достигается.

3.4 Достаточные статистики

Обычно, выборка содержит достаточно большое число измерений и хранить всю имеющуюся информацию сложно. Но во многих задачах это и не нужно. Например, пусть мы имеем повторную выборку $X = (X_1, \dots, X_N)$ из генеральной совокупности, параметры которой есть a и σ^2, где $a = M(X_j) = \nu_1$, $\sigma^2 = D(X_j) = \nu_2 - (\nu_1)^2$. Оценки этих параметров имеют вид

$$\hat a = \bar X = \frac{1}{N}\sum_{j=1}^{N} X_j\ ,$$

$$\hat{\sigma^2} = S_1^2 = \frac{1}{N-1}\sum_{j=1}^{N}(X_j - \bar X)^2 = \frac{1}{N-1}\sum_{j=1}^{N} X_j^2 - \frac{1}{N(N-1)}\sum_{j=1}^{N} X_j\ .$$

Для вычисления этих оценок не нужно знать (и, следовательно, хранить) весь объем наблюдений. Достаточно знать значения двух статистик

$$T_1(X) = \sum_{j=1}^{N} X_j \ , \ T_2(X) = \sum_{j=1}^{N} X_j^2 \ .$$

Этих двух статистик *достаточно* для вычисления указанных выше оценок параметров a и σ^2. Оказывается, что их достаточно и для решения практически всех статистических задач, связанных с параметрами a и σ^2.

Подобная ситуация встречается и во многих других задачах. Как выяснится позднее, оптимальные оценки, если они существуют, вычисляются через такие статистики. В силу этого, полезно изучить статистики такого типа подробнее.

Пусть мы имеем параметрическое семейство распределений $\mathcal{P} = \{P_\theta, \theta \in \Theta \subset R^m\}$. Всюду далее мы будем предполагать, что P_θ есть распределение случайной величины ξ, которая имеет либо дискретное распределение и в этом случае вероятности появления ее значений y задаются с помощью функции $p(y, \theta)$, либо непрерывное распределение и в этом случае ее плотность распределения в точке y задается с помощью функции $\rho(y, \theta)$.

Пусть, далее, $X = (X_1, \ldots, X_N)$ есть повторная выборка из генеральной совокупности с распределением P_θ. Для определенности далее мы рассматриваем случай дискретного распределения. Тогда для конкретной реализации $x = (x_1, \ldots, x_N)$ выборки X мы имеем

$$P_\theta(X = x) = \mathcal{L}(x, \theta) = p(x_1, \theta) \cdot \ldots \cdot p(x_N, \theta) \ .$$

Напомним, что $\mathcal{L}(x, \theta)$ называется функцией правдоподобия.

Определение 2 . *Некоторая статистика $T(X)$ называется достаточной для параметра θ (более точно, семейства распределений $\mathcal{P} = \{P_\theta, \theta \in \Theta\}$), если $\forall x \in \mathcal{X}$ и любого возможного значения t статистики $T(X)$ условная вероятность*

$$P_\theta\big(X = x | T(X) = t\big)$$

не зависит от параметра θ.

Замечания. 1. На интуитивном уровне это означает, что статистика $T(X)$ содержит в себе всю информацию о параметре θ, т.е. дополнительная информацию, содержащаяся в выборке $X = (X_1, \ldots, X_N)$ (кроме той, что есть в $T(X)$), не говорит нам ничего нового о параметре θ.

2. Если мы рассматриваем семейство непрерывных распределений, то в определении достаточной статистики вероятности нужно заменить на плотности.

Пример. Пусть случайная величина ξ принимает значения 1 и 0 с вероятностями p и $1-p$ соответственно, где $0 < p < 1$, а $X = (X_1, \ldots, X_N)$ есть повторная выборка из генеральной совокупности с таким распределением. Из курса теории вероятностей мы знаем, что это схема Бернулли с параметрами N и p. Рассмотрим статистику $S_N = X_1 + \ldots + X_N$, т.е. число успехов в испытаниях Бернулли. Как известно, случайная величина S_N имеет биномиальное распределение с параметрами N и p, т.е. она принимает значения $m = 0, 1, \ldots, N$ с вероятностями

$$P(S_N = m) = C_N^m p^m (1-p)^{N-m} \ .$$

Пусть $x = (x_1, \ldots, x_N)$ есть конкретная реализация выборки X и $S_N(x) = x_1 + \ldots + x_N = m$.Тогда для любого $0 < p < 1$ имеем:

$$P\big(X = x | S_N(X) = m\big) = \frac{P\big((X = x) \cap (S_N(X) = m)\big)}{P\big(S_N(X) = m\big)} =$$

$$= \frac{P(X = x)}{P(S_N(X) = m)} = \frac{p^m(1-p)^{N-m}}{C_N^m p^m (1-p)^{N-m}} = \frac{1}{C_n^m} \ ,$$

т.е. получаем результат, который не зависит от конкретного значения параметра $0 < p < 1$.

Вычислять условные вероятности, тем более, условные плотности, достаточно трудно. Поэтому, хотелось бы иметь более простой метод проверки достаточности статистики $T(X)$ чем это предлагает ее определение.

Теорема 3 *(Критерий факторизации). Пусть $X = (X_1, \dots, X_N)$ есть повторная выборка из генеральной совокупности с дискретным распределением $p(y, \theta)$, $\theta \in \Theta$. Статистика $T = T(X) = \big(T_1(X), \dots, T_r(X)\big)$ будет достаточной для параметра θ тогда и только тогда,когда существуют две функции $g(t, \theta)$ и $h(x)$, такие, что*

$$P_\theta(X = x) = \mathcal{L}(x, \theta) = g(T(x), \theta) \cdot h(x) \ .$$

Доказательство. Пусть $x = (x_1, \dots, x_N)$ есть конкретная реализация выборки X. Легко видеть, что $(X = x) \cap \big(T(X) = t\big) = (X = x)$. Тогда

$$\begin{aligned} P_\theta\big(X = x | T(X) = t\big) &= \frac{P_\theta\big((X = x) \cap (T(X) = t\big)}{P_\theta(T(X) = t)} = \\ &= \frac{P_\theta(X = x)}{P_\theta(T(X) = t)} = h(x) \ , \end{aligned}$$

где $h(x)$ не зависит от θ по определению достаточной статистики. Обозначим $g(t, \theta) = P_\theta(T(X) = t)$. Тогда из последнего соотношения получаем

$$P_\theta(X = x) = P_\theta\big(T(X) = t\big) \cdot h(x) = g\big(T(x), \theta\big) \cdot h(x) \ .$$

Замечание. Аналогичный результат справедлив для непрерывных распределений, если вероятности заменить на плотности.

Пример. Пусть $X = (X_1, \dots, X_N)$ есть повторная выборка из генеральной совокупности с нормальным распределением, которое имеет параметры $\theta_1 = a \in R^1$, $\theta_2 = \sigma^2 > 0$. Для нормального распределения его плотность в точке y записывается в виде

$$\begin{aligned} \rho(y, a, \sigma^2) = \rho(y, \theta_1, \theta_2) &= \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{(y-a)^2}{2\sigma^2}} = \\ &= (2\pi\theta_2)^{-\frac{1}{2}} e^{-\frac{(y-\theta_1)^2}{2\theta_2}} \ . \end{aligned}$$

Запишем функцию правдоподобия для $x = (x_1, \ldots, x_N)$:

$$
\begin{aligned}
\mathcal{L}(x, \theta_1, \theta_2) &= \prod_{j=1}^{N} \rho(x_j, \theta_1, \theta_2) = \prod_{j=1}^{N} (2\pi\theta_2)^{-1/2} \cdot \exp\left\{ -\frac{1}{2\theta_2}(x_j - \theta_1)^2 \right\} = \\
&= (2\pi\theta_2)^{-\frac{N}{2}} \exp\left\{ -\frac{1}{2\theta_2} \cdot \sum_{j=1}^{N} (x_j - \theta_1)^2 \right\} = \\
&= (2\pi\theta_2)^{-\frac{N}{2}} \exp\left\{ -\frac{1}{2\theta_2} \cdot \sum_{j=1}^{N} x_j^2 + \frac{\theta_1}{\theta_2} \sum_{j=1}^{N} x_j - \frac{N\theta_1^2}{2\theta_2} \right\} = \\
&= (2\pi\theta_2)^{-\frac{N}{2}} \exp\left\{ -\frac{1}{2\theta_2} \cdot T_2(x) + \frac{\theta_1}{\theta_2} T_1(x) - \frac{N\theta_1^2}{2\theta_2} \right\} \cdot 1 = \\
&= g(T_1(x), T_2(x), \theta_1, \theta_2) \cdot h(x) \ .
\end{aligned}
$$

По критерию факторизации векторная статистика $T(x) = (T_1(x), T_2(x))$, где

$$
T_1(x) = \sum_{j=1}^{N} x_j \ , \ T_2(x) = \sum_{j=1}^{N} x_j^2 \ ,
$$

является достаточной для параметра $\theta = (\theta_1, \theta_2)$.

3.5 Полные статистики

Наша основная цель есть построение оптимальных оценок. Есть один случай, когда достаточно легко найти оптимальную оценку. Предположим, что оптимальная оценка имеет некоторый специальный вид. Далее мы доказываем, что оценка такого типа только одна. Если мы найдем какую-либо оценку такого типа, то она и будет оптимальной. В этом разделе мы реализуем эту программу.

Пусть мы имеем некоторую векторную статистику $T(x) = \big(T_1(x), \ldots, T_r(x)\big)$, т.е. отображение $T : \mathcal{X} \to R^r$. Пусть $\mathcal{P} = \big\{P_\theta, \theta \in \Theta\big\}$ есть семейство распределений вероятностей с.в. ξ. $X = (X_1, \ldots X_N)$ есть повторная выборка и ее совместное распределение задается с помощью функции правдоподобия $\mathcal{L}(x, \theta)$. Если рассмотреть $T(X) = \big(T_1(X), \ldots, T_r(X)\big)$, то мы имеем случайный вектор со значениями в R^r. Он имеет некоторое распределение P_θ^T в пространстве R^r, если выборка X имеет компоненты X_k с распределением P_θ. Таким образом, мы имеем новое семейство распределений $\mathcal{F} = \{P_\theta^T, \theta \in \Theta\}$.

Определение 3 . *Статистика* $T(x) = \big(T_1(x), \ldots, T_r(x)\big)$ *называется* **полной**, *если для любой функции* $g(t_1, \ldots, t_r)$, *для которой с.в.* $g\big(T_1(X), \ldots, T_r(X)\big)$ *имеет конечное математическое ожидание, для любого распределения* $P_\theta, \theta \in \Theta$, *из условия*

$$M_\theta\Big[g\big(T_1(X), \ldots, T_r(X)\big)\Big] = 0 \ \forall \theta \in \Theta$$

следует, что

$$g(t_1, \ldots, t_r) = 0 \ ,$$

для почти всех точек $t = (t_1, \ldots, t_r)$ *из множества значений статистики* T.

Замечания. 1. Свойство полноты статистики T есть, на самом деле, свойство семейства распределений $\mathcal{F}$, введенного выше. Оно говорит о том, что, зная средние значения функции $g(t_1, \ldots, t_r)$ относительно всех распределений из семейства $\mathcal{F}$, мы однозначно восстанавливаем эту функцию.

2. Чтобы проверить полноту конкретной статистики, мы должны решить некоторую аналитическую задачу из математического анализа.

Пример. Пусть случайная величина ξ принимает значения 1 и 0 с вероятностями p и $1-p$ соответственно, где $p = \theta \in (0,1)$ есть неизвестный параметр. Пусть, далее, $X = (X_1, \ldots X_N)$ есть повторная выборка из генеральной совокупности с таким распределением. Рассмотрим статистику $S_N(X) = X_1 + \ldots + X_N$. Она имеет биномиальное распределение с параметрами N и $p = \theta$, т.е. принимает значения $m = 0, 1, \ldots, N$ с вероятностями

$$P(S_N = m) = C_N^m \cdot \theta^m \cdot (1-\theta)^{N-m} \ .$$

Пусть на множестве $\{0, 1, \ldots, N\}$ значений статистики S_N задана функция g, т.е. задана конечная последовательность чисел $g_0, g_1, \ldots, g_N$. Пусть, далее,

$$\begin{aligned} M_\theta[g(S_N(X))] &= \sum_{m=0}^{N} g_m \cdot P_\theta(S_N = m) \\ &= \sum_{m=0}^{N} g_m \cdot C_N^m \cdot \theta^m \cdot (1-\theta)^{N-m} \\ &= (1-\theta)^N \cdot \sum_{m=0}^{N} g_m \cdot C_N^m \cdot \left(\frac{\theta}{1-\theta}\right)^m = 0 \end{aligned}$$

для всех $\theta \in (0, 1)$. Отсюда следует, что

$$\sum_{m=0}^{N} g_m C_N^m \lambda^m \equiv 0,$$

где $\lambda = \theta/(1-\theta) \in (0, \infty)$. Мы имеем полином от λ, где λ пробегает все положительные числа. Такой полином равен тождественно нулю тогда и только тогда, когда равны нулю все его коэффициенты. Отсюда получаем, что $g_m \cdot C_N^m = 0$ или $g_m = 0$ для всех $m = 0, 1, \ldots, N$, т.е. статистика S_N является полной.

Теперь мы покажем, как применяются полные статистики в теории оптимального оценивания. Пусть $g(\theta)$ есть некоторая параметрическая функция, которую мы хотим оценить, используя повторную выборку $X = (X_1, \ldots X_N)$ из генеральной совокупности с распределением $P_\theta, \theta \in \Theta$. Предположим, что в рассматриваемой задаче существует полная статистика $T(X)$.

Теорема 4 . *Если для функции $g(\theta)$ существует несмещенная оценка, которая есть функция от полной статистики $T(X)$, то она определяется единственным образом, т.е. если $g_1\big(T(X)\big)$ и $g_2\big(T(X)\big)$ — две несмещенные оценки для $g(\theta)$, то $g_1(t) = g_2(t)$ п.н. на множестве значений статистики $T(X)$.*

Доказательство. Рассмотрим функцию $\tilde{g}(t) = g_1(t) - g_2(t)$. Тогда, в силу несмещенности,

$$\begin{aligned} M_\theta\Big[g\big(T(X)\big)\Big] &= M_\theta\Big[g_1\big(T(X)\big) - g_2\big(T(X)\big)\Big] = \\ &= g(\theta) - g(\theta) = 0, \ \forall\theta \ . \end{aligned}$$

и, в силу полноты статистики $T(X)$, мы получаем

$$g(t) = g_1(t) - g_2(t) = 0$$

п.н. на множестве значений статистики $T(X)$.

3.6 Построение оптимальных оценок

Выше мы уже обсуждали вопрос об оптимальности в среднем квадратическом для несмещенных оценок. В частности, неравенство Рао — Крамера позволяет в некоторых случаях для конкретных оценок доказать их оптимальность. Но если для конкретной оценки в неравенстве Рао-Крамера не достигается знак равенства, то это еще не означает, что она не является оптимальной. Но главная проблема состоит в том, что остался открытым вопрос о том, как *построить* оптимальную оценку в конкретной задаче. Этот вопрос рассматривается в этом разделе нашего курса.

Теорема 5 *(Рао–Блекуэлла–Колмогорова). Пусть $\hat{g}(X)$ есть несмещенная оценка для параметрической функции $g(\theta)$ и $T(X)$ — достаточная статистика. Определим*

$$f(t) := M_\theta\big[\hat{g}(X)|T(X) = t\big].$$

тогда справедливы следующие свойства:
1) $\hat{g}_1(X) := f(T(X))$ — несмещенная оценка для $g(\theta)$,
2) $D_\theta(\hat{g}_1(X)) \leqslant D_\theta(\hat{g}(X))$.

Замечание. Смысл этой теоремы в том, что имея какую-либо несмещенную оценку, мы можем построить новую оценку, которая также является несмещенной, имеет вид $f\big(T(X)\big)$ и ее дисперсия будет не больше дисперсии исходной оценки. Таким образом, оптимальную оценку нужна искать среди оценок вида $f\big(T(X)\big)$.

Далее, как показано в предыдущем пункте, если $T(X)$ — полная статистика, то несмещенная оценка вида $f\big(T(X)\big)$ единственна.

Пусть мы нашли полную и достаточную статистику $T(X)$. Тогда оптимальную оценку строят следующим образом.

1. Находим какую-либо несмещенную оценку $\hat{g}(X)$ для $g(\theta)$ (если она существует),

2. Вычисляем функцию

$$f(t) = M_\theta\big[\hat{g}(X)|T(X) = t\big].$$

С.в. $\hat{g}_1(X) := f\big(T(X)\big)$ является несмещенной оценкой для $g(\theta)$, которая не хуже $\hat{g}(X)$.

3. В силу полноты $T(X)$ оценка такого типа является единственной и, поэтому, оптимальна.

Пример. Пусть ξ принимает два значения 1 и 0 с вероятностями θ и $1-\theta$, где $\theta \in (0,1)$ есть неизвестный параметр. Мы хотим оценить $g(\theta) = \theta$. Пусть $X = (X_1, \ldots, X_N)$ есть повторная выборка из генеральной совокупности с таким распределением. Ранее мы уже показали, что статистика $T(X) = X_1 + \ldots + X_N$ является полной и достаточной. Рассмотрим оценку $\tilde{\theta}(X) = X_1$ для параметра θ. Тогда

$$M_\theta(\tilde{\theta}) = M_\theta(X_1) = 1 \cdot \theta + 0 \cdot (1-\theta) = \theta \ ,$$

т.е. это несмещенная оценка параметра θ. Далее,

$$M_\theta\big(T(X)|T(X) = t\big) = t = M_\theta\big(X_1 + \ldots + X_N|T(X) = t\big) =$$

$$= M_\theta\big(X_1|T(X) = t\big) + \ldots + M_\theta\big(X_N|T(X) = t\big) = N \cdot M_\theta\big(X_1|T(X) = t\big).$$

Отсюда следует, что

$$M_\theta\big(X_1|T(X) = t\big) = \frac{t}{N} =: f(t),$$

и оценка

$$\hat{\theta} = f\big(T(X)\big) = \frac{T(X)}{N} = \bar{X}$$

является оптимальной оценкой для параметра θ.

3.7 Методы построения оценок

В этом параграфе мы рассмотрим два наиболее популярных метода построения оценок.

3.7.1 Метод моментов

Пусть случайная величина ξ имеет функцию распределения $F_\xi(y, \theta_1, \ldots, \theta_m)$, которая содержит несколько неизвестных параметров, а в остальном вид этой функции известен. Тогда мы можем вычислить несколько первых моментов

$$\begin{cases} M\xi = f_1(\theta_1, \ldots, \theta_m) = \nu_1, \\ M\xi^2 = f_2(\theta_1, \ldots, \theta_m) = \nu_2, \\ \ldots, \\ M\xi^m = f_m(\theta_1, \ldots, \theta_m) = \nu_m, \end{cases}$$

которые являются функциями от неизвестных параметров $\theta_1, \ldots,\ \theta_m$. Предположим, что мы решили эти уравнения относительно параметров:

$$\begin{cases} \theta_1 = g_1(\nu_1, \ldots, \nu_m), \\ \ldots, \\ \theta_m = g_m(\nu_1, \ldots, \nu_m). \end{cases} \tag{7}$$

В начале этого параграфа мы выяснили, что хорошими оценками для моментов являются их выборочные аналоги $\hat{\nu}_1,\ \ldots,\ \hat{\nu}_m$. Подставляя их в уравнения (7) мы получим оценки $\hat{\theta}_1,\ \ldots,\ \hat{\theta}_m$, построенные по **методу моментов**.

Пример. ξ имеет нормальное распределение с параметрами $a = \theta_1$ и $\sigma^2 = \theta_2$. Мы знаем, что $a = M\xi = \nu_1$, $\sigma^2 = D(\xi) = M\xi^2 - (M\xi)^2 = \nu_2 - \nu_1^2$. Отсюда получаем

$$\hat{a} = \bar{X} = \tfrac{1}{N}\sum_{k=1}^{N} x_k,$$
$$\sigma^2 = \hat{\nu}_2 - (\hat{\nu}_1)^2 = \tfrac{1}{N}\sum_{k=1}^{N} x_k^2 - (\bar{X})^2 = \tfrac{1}{N}\sum_{k=1}^{N}(x_k - \bar{X})^2 = S^2.$$

Замечание. Метод моментов как правило дает состоятельные оценки, но они могут быть смещенными и не эффективными. Этот метод является довольно простым с вычислительной точки зрения.

3.7.2 Метод наибольшего правдоподобия

Этот метод основан на интуитивном представлении о том, что реализуются в случайном эксперименте те события, которые имеют большие вероятности. Обращая это утверждение, мы приходим к следующему **принципу наибольшего правдоподобия**: при заданном наборе экспериментальных данных нужно выбирать ту модель, для которой эти данные наиболее вероятны. Применим этот принцип к оценке параметров.

Пусть ξ — дискретная случайная величина, вероятности значений y которой вычисляются с помощью некоторой функции $p(y, \theta_1, \ldots, \theta_m) = p(y, \theta)$. Пусть мы имеем конкретную повторную выборку $x = (x_1, \ldots, x_N)$. Вычислим вероятность появления именно таких значений:

$$\mathcal{L}(x, \theta) = p(x_1, \theta) \cdot \ldots \cdot p(x_N, \theta).$$

Выше мы назвали $\mathcal{L}(x, \theta)$ функцией правдоподобия. При фиксированном x найдем такое θ, которое есть решение экстремальной задачи:

$$\mathcal{L}(x, \theta) \longrightarrow \max_{\theta}.$$

Ясно, что при разных x мы будем получать разные решения. Таким образом мы получим оценку $\hat{\theta}_N = \hat{\theta}_N(x)$ — **оценку по методу наибольшего правдоподобия** (ОМНП).

Если ξ имеет плотность распределения, то в функции правдоподобия нужно вероятности заменить на плотности.

Примеры. 1. ξ имеет нормальное распределение с параметрами $a = \theta_1$ и $\sigma^2 = \theta_2$. ОМНП в этом случае вновь будут $\bar{X}$ и S^2.

2. ξ имеет распределение Пуассона с параметром λ. Если $x = (x_1, \ldots, x_N)$ — повторная выборка, то

$$\mathcal{L}(x, \lambda) = \frac{\lambda^{x_1}}{x_1!} e^{-\lambda} \cdot \ldots \cdot \frac{\lambda^{x_N}}{x_N!} e^{-\lambda} = \frac{\lambda^{x_1 + \cdots + x_N}}{x_1! \cdot \ldots \cdot x_N!} e^{-N\lambda}.$$

Нам нужно найти точку, где достигается экстремум, а не величину экстремума. Поэтому любая монотонная функция от функции правдоподобия дает тот же результат. В нашем случае удобно перейти к логарифмам.

$$L(x,\lambda) = \ln \mathcal{L}(x,\lambda) = (x_1 + \cdots + x_N) \cdot \ln \lambda - N \cdot \lambda - \ln(x_1! \cdot \cdots \cdot x_N!).$$

$$\frac{\partial L(x,\lambda)}{\partial \lambda} = \frac{(x_1+\cdots+x_N)}{\lambda} - N = 0$$
$$\Longrightarrow \hat{\lambda} = \frac{1}{N}(x_1 + \cdots + x_N) = \bar{X}.$$

Несложно показать, что это точка наибольшего значения для $\mathcal{L}(x,\lambda)$.

Перечислим без доказательства некоторые свойства ОМНП. Если выполнены некоторые условия регулярности для функции распределения $F(y,\theta)$ случайной величины ξ, то ОМНП $\hat{\theta}_N$ обладают свойствами:

1. $\hat{\theta}_N$ — состоятельны,
2. $\hat{\theta}_N$ — асимптотически нормальны,
3. $\hat{\theta}_N$ — асимптотически эффективны.

Глава 4

Интервальные оценки параметров

4.1 Определение доверительного интервала

До сих пор в качестве оценки мы указывали некоторое конкретное значение $\hat{\theta}_N$ параметра. Но в реальной задаче даже самая хорошая оценка, которая возможна в рассматриваемой ситуации, может оказаться неудовлетворительной, так как не обеспечивает нужной точности. Таким образом, мы хотели бы дополнить $\hat{\theta}_N$ оценкой точности, то есть неравенством вида

$$\left|\hat{\theta}_N - \theta\right| < \delta.$$

Покажем, что в нашей задаче это, вообще говоря, невозможно, если не сделать некоторых оговорок. Чтобы прояснить ситуацию рассмотрим

Пример. $X = (X_1, \ldots, X_N)$ — повторная выборка из генеральной совокупности с нормальным распределением, имеющим параметры a и σ^2. Интересующий нас параметр $\theta = a$. Мы уже знаем, что оптимальной оценкой является $\hat{a} = \bar{X}$. Но это случайная величина, имеющая нормальное распределение со средним a и дисперсией σ^2/N. Такая случайная величина принимает любые значения от $-\infty$ до $+\infty$. Поэтому для любого $\delta > 0$ неравенство

$$\left|\bar{X} - a\right| < \delta$$

наверняка выполняться не может. Всегда есть положительная вероятность того, что оно будет не выполнено. Но при достаточно большом δ эта вероятность будет достаточно мала. Поэтому, практически достоверно, что оно выполнено. Решая неравенство относительно a, получаем

$$\bar{X} - \delta < a < \bar{X} + \delta.$$

Таким образом, этот подход дает нам в качестве оценки некоторый интервал, который покрывает истинное значение параметра с большой вероятностью.

Определение. Пусть мы имеем повторную выборку $X = (X_1, \dots, X_N)$ из генеральной совокупности с функцией распределения $F(y, \theta)$, где $\theta \in \Theta \subset \mathbf{R^1}$ — скалярный параметр. **Доверительным интервалом** уровня γ для параметра θ называется интервал $\big(\hat{\theta}^{(1)}(X), \hat{\theta}^{(2)}(X)\big)$ со случайными концами такой, что

$$P_\theta\big(\hat{\theta}^{(1)}(X) < \theta < \hat{\theta}^{(2)}(X)\big) \geqslant \gamma \qquad \forall \theta \in \Theta. \tag{1}$$

Число γ называется доверительным уровнем интервала.

Чем больше число γ, тем выше вероятность того, что интервал покрывает истинное значение параметра. В этом смысле число γ характеризует **надежность** этого интервала. Увеличивая длину интервала, мы, естественно, увеличиваем и его надежность. Но при этом уменьшается **точность** нашей оценки, которую естественно характеризовать средней длиной нашего интервала:

$$l = M_\theta\big(\hat{\theta}^{(2)} - \hat{\theta}^{(1)}\big).$$

γ и l — это два «конфликтующих» показателя. Улучшая один, мы ухудшаем другой и наоборот. Это типичная ситуация в так называемых многокритериальных задачах. В этих задачах нам нужно выбрать некоторый объект, который был бы хорошим сразу по нескольким показателям. Как правило, невозможно найти объект, который был бы наилучшим по всем показателям сразу. В этом случае применяют следующий подход. Мы отбираем все объекты, которые обладают достаточно хорошими (пусть и не наилучшими) свойствами по некоторым наиболее важным для нас критериям. Затем среди отобранных объектов стараются найти такой, который оптимизирует остальные критерии. В нашей задаче мы фиксируем доверительный уровень γ. Обычно это число вида 0.9, 0.95, 0.99, то есть достаточно близкое к 1. Затем среди всех доверительных интервалов уровня γ пытаются найти такой, у которого средняя длина l будет наименьшей, то есть наиболее точный.

Единого метода построения доверительных интервалов, который срабатывал бы во всех случаях жизни, не существует. Но тем не менее есть достаточно общие методы, которые мы изложим ниже. Начнем мы с наиболее важного примера — нормального распределения.

4.2 Некоторые распределения, связанные с нормальным

При построении доверительных интервалов и проверке гипотез о параметрах нормального распределения нам потребуются некоторые специальные распределения вероятностей.

Определение 4 . *Пусть $\xi_1, \xi_2, \ldots, \xi_n$ есть независимые случайные величины, имеющие стандартное нормальное распределение. Распределение случайной величины*

$$\chi_n^2 := \xi_1^2 + \ldots + \xi_n^2$$

называется χ^2–распределением с n степенями свободы.

Это распределение имеет плотность распределения вида

$$\rho_{\chi_n^2}(y) = \frac{1}{\Gamma(\frac{n}{2}) \cdot 2^{n/2}} y^{\frac{n}{2}-1} e^{-\frac{1}{2}y} \ , \ y > 0 \ .$$

Таким образом, это частный случай гамма-распределения с $\alpha = \frac{n}{2}$, $\beta = \frac{1}{2}$.

Определение 5 . *Пусть $\xi_1, \xi_2, \ldots, \xi_n, \xi_{n+1}$ есть независимые случайные величины, имеющие стандартное нормальное распределение. Распределение случайной величины*

$$t_n := \frac{\xi_{n+1}}{\sqrt{\frac{1}{n}(\xi_1^2 + \ldots + \xi_n^2)}}$$

называется распределением Стьюдента с n степенями свободы.

Происхождение такого названия мы объясним позднее. Это распределение имеет плотность распределения вида

$$\rho_{t_n}(y) = \frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n}{2})\sqrt{\pi n}}(1+\frac{y^2}{n})^{-\frac{n+1}{2}} \ , \ y \in R^1 \ .$$

Заметим, что при $n = 1$ получаем *распределение Коши.*

Определение 6 . *Пусть $\xi_1, \xi_2, \ldots, \xi_n, \xi_{n+1}, \ldots, \xi_{n+m}$ есть независимые случайные величины, имеющие стандартное нормальное распределение. Распределение случайной величины*

$$F_{n,m} := \frac{\frac{1}{n}(\xi_1^2 + \ldots + \xi_n^2)}{\frac{1}{m}(\xi_{n+1}^2 + \ldots + \xi_{n+m}^2)}$$

называется распределением Снедекора или распределением дисперсионного отношения Фишера с n и m степенями свободы.

В дальнейшем мы будем называть его *распределением Фишера–Снедекора.* Оно имеет плотность распределения следующего вида

$$\rho_{F_{n,m}}(y) = \frac{(\frac{n}{m})^{n/2} \cdot \Gamma(\frac{n+m}{2})}{\Gamma(\frac{n}{2}) \cdot \Gamma(\frac{m}{2})} y^{\frac{n}{2}} \cdot \left(1 + \frac{n}{m} \cdot y\right)^{-\frac{(n+m)}{2}} \ , \ y > 0 \ .$$

Заметим, что при $n = 1$ получаем $t_m^2 \overset{d}{=} F_{1,m}$, т. е. квадрат с. в. t_m, имеющей распределение Стьюдента с m степенями свободы, имеет распределение Фишера–Снедекора с 1 и m степенями свободы.

Для всех определенных выше распределений составлены таблицы (см., например, Большев Л. Н., Смирнов Н. И. «Таблицы математической статистики»).

4.3 Построение доверительных интервалов для параметров нормального распределения

Нам потребуется один результат, который мы приведем без доказательства.

Теорема 6 . *Определим*

$$\bar{X} = \frac{1}{N}\sum_{k=1}^{N} X_k, \quad S_1^2 = \frac{1}{N-1}\sum_{k=1}^{N}(X_k - \bar{X})^2$$

— выборочное среднее и исправленную выборочную дисперсию. Тогда:

1. *$\bar{X}$ и S_1^2 — независимы,*
2. *$\bar{X}$ имеет нормальное распределение*
3. *$(N-1)\cdot S_1/\sigma^2$ имеет χ^2–распределение с $N-1$ степенями свободы,*
4. *$\frac{(\bar{X}-a)}{S_1}\cdot\sqrt{N}$ имеет распределение Стьюдента с $N-1$ степенями свободы.*

Мы разберем отдельно несколько случаев.

a) $\theta = a$, σ^2 — известно.

Рассмотрим случайную величину

$$Y = \frac{\bar{X} - a}{\sigma/\sqrt{N}} \ . \tag{4.1}$$

В силу теоремы 6 с. в. Y имеет стандартное нормальное распределение. По таблицам для заданного γ найдем константу $C(\gamma)$, для которой

$$P\big(|Y| < C(\gamma)\big) = 2\Phi_0\big(C_\gamma\big) = \gamma \ . \tag{4.2}$$

Тогда из соотношений (1) и (2) получаем

$$\bar{X} - \frac{C(\gamma)\cdot\sigma}{\sqrt{N}} < a < \bar{X} + \frac{C(\gamma)\cdot\sigma}{\sqrt{N}}$$

b) $\theta = a$, σ^2 — неизвестно.

Рассмотрим случайную величину

$$t_{N-1} = \frac{\bar{X} - a}{S_1}\cdot\sqrt{N} \ . \tag{4.3}$$

В силу теоремы 6 с. в. t_{N-1} имеет распределение Стьюдента с $N-1$ степенями свободы. По таблицам для заданного γ найдем константу $t_{N-1}(\gamma)$, для которой

$$P\big(|t_{N-1}| < t_{N-1}(\gamma)\big) = \gamma \ . \tag{4.4}$$

Тогда из соотношений (3) и (4) получаем

$$\bar{X} - \frac{t_{N-1}(\gamma) \cdot S_1}{\sqrt{N}} < a < \bar{X} + \frac{t_{N-1}(\gamma) \cdot S_1}{\sqrt{N}} \ .$$

c) $\theta = \sigma^2$, a — неизвестно.

Рассмотрим случайную величину

$$\chi^2_{N-1} = \frac{(N-1) \cdot S_1^2}{\sigma^2} \ . \tag{4.5}$$

В силу теоремы 6 с. в. χ^2_{N-1} имеет χ^2–распределение с $N-1$ степенями свободы. По таблицам для заданного γ найдем константы ${}_1(\gamma)$ и $C_2(\gamma)$ для которых

$$P\big(\chi^2_{N-1} < C_1(\gamma)\big) = \frac{1-\gamma}{2} \ , P\big(\chi^2_{N-1} > C_2(\gamma)\big) = \frac{1-\gamma}{2} \ . \tag{4.6}$$

Тогда из соотношений (5) и (6) получаем

$$\frac{(N-1)S_1^2}{C_2(\gamma)} < \sigma^2 < \frac{(N-1)S_1^2}{C_1(\gamma)} \ .$$

Замечание. Можно показать, что для случаев a) и b) построенные интервалы являются наиболее точными среди всех доверительных интервалов уровня γ, основанных на статистиках Y и t_{N-1}. Для случая c) это не так. Чтобы построить наиболее точный интервал в этой ситуации, нужно решить некоторую оптимизационную задачу. Мы используем соотношения (6), чтобы упростить задачу.

4.4 Построение доверительных интервалов с помощью центральных статистик

Все примеры, рассмотренные выше, имеют одну и ту же особенность. А именно, нам удалось найти такую функцию $T(X, \theta)$ от выборки X и оцениваемого параметра θ, распределение которой не зависит от θ, и это распределение удалось вычислить в явном виде. Ниже мы опишем общую процедуру построения доверительного интервала, применимую в подобных ситуациях. Пусть случайная величина ξ имеет функцию распределения $F_\xi(y) = F(y, \theta, \theta')$, где θ — неизвестный скалярный параметр, подлежащий оцениванию, а $\theta' = (\theta_1, \ldots, \theta_m) \in \Theta \subset \mathbf{R}^m$ — некоторое число так называемых «мешающих параметров». Мы не интересуемся значением θ', но оно влияет на распределение случайной величины ξ. Пусть $X = (X_1, \ldots, X_N)$ есть повторная выборка из генеральной совокупности с функцией распределения $F(y, \theta, \theta')$.

Определение. Функция $T(X, \theta)$ от выборки X и основного параметра θ называется **центральной статистикой**, если

1. случайная величина $T(X, \theta)$ имеет распределение, не зависящее от параметров θ и θ';
2. $T(X, \theta)$ есть монотонная функция от параметра θ.

Если мы имеем некоторую центральную статистику $T(X, \theta)$, то легко построить доверительный интервал для θ. Для заданного γ найдём две константы C_1 и C_2, для которых

$$P\big(C_1 < T(X, \theta) < C_2\big) = \gamma.$$

В силу монотонности по θ неравенство

$$C_1 < T(X, \theta) < C_2$$

эквивалентно неравенству

$$T_1(X, C_1, C_2) < \theta < T_2(X, C_1, C_2) \tag{8}$$

для некоторых статистик $T_1(X, C_1, C_2)$ и $T_2(X, C_1, C_2)$. Формула задаёт некоторый доверительный интервал уровня γ для параметра θ.

Заметим, что этот метод срабатывает, если распределение $T(X, \theta)$ найдено в явном виде. Например, для него составлены таблицы.

Пример. Случайная величина ξ имеет показательное распределение с параметром λ. Пусть $X = (X_1, \dots, X_N)$ есть повторная выборка. Необходимо построить доверительный интервал для параметра $\theta = \lambda$. Параметр λ является масштабным, поэтому случайная величина X_k/λ имеет показательное распределение с параметром $\lambda' = 1$. Показательное распределение есть частный случай гамма-распределения. В частности, X_k/λ имеет гамма-распределение с параметрами $\alpha = 1$ и $\beta = 1$. На практических занятиях было показано, что случайная величина

$$T(X, \lambda) = \lambda \cdot (X_1 + \dots + X_N)$$

имеет гамма–распределение с параметрами $\alpha = N$ и $\beta = 1$. Легко видеть, что $T(X, \lambda)$ — центральная статистика. Удобнее рассмотреть случайную величину

$$T_1(X, \lambda) = 2 \cdot T(X, \lambda) = 2 \cdot \lambda \cdot (X_1 + \dots + X_N)$$

так как она имеет гамма–распределение с параметрами $\alpha = N$ и $\beta = 1/2$, а это есть χ^2–распределение с $2N$ степенями свободы. Далее, для заданного γ по таблицам χ^2–распределения находим две константы $C_1(\gamma)$ и $C_2(\gamma)$, для которых

$$P\big(T_1(X, \lambda) < C_1(\gamma)\big) = \frac{1-\gamma}{2}, \quad P\big(T_1(X, \lambda) > C_2(\gamma)\big) = \frac{1-\gamma}{2}.$$

Тогда событие, состоящее в том, что

$$C_1(\gamma) < T_1(X, \lambda) < C_2(\gamma),$$

имеет вероятность, равную γ. Из последнего неравенства и определения $T_1(X, \lambda)$ получаем

$$\frac{C_1(\gamma)}{2(X_1 + \dots + X_N)} < \lambda < \frac{C_2(\gamma)}{2(X_1 + \dots + X_N)}.$$

4.5 Построение доверительных интервалов с помощью асимптотически нормальных оценок

Центральные статистики существуют только для довольно узкого класса статистических моделей. Поэтому чаще строят приближённые доверительные интервалы, используя асимптотическую нормальность точечных оценок. Пусть $\hat{\theta}_N = \hat{\theta}_N(X)$ есть асимптотически нормальная оценка параметра θ, т.е. случайная величина

$$\frac{\hat{\theta}_N - \theta}{B_N(\theta)}$$

имеет асимптотически стандартное нормальное распределение, где $B_N(\theta) > 0$ есть некоторая нормирующая константа. Тогда для заданного γ по таблицам нормального распределения найдём константу $C(\gamma)$, для которой событие, состоящее в том, что

$$\left|\frac{\hat{\theta}_N - \theta}{B_N(\theta)}\right| < C(\gamma) \tag{9}$$

приближённо (при $N \to \infty$) имеет вероятность, равную γ. Отсюда мы получаем, что

$$\hat{\theta}_N - C(\gamma) \cdot B_N(\theta) < \theta < \hat{\theta}_N + C(\gamma) \cdot B_N(\theta). \tag{10}$$

Но интервал (10), вообще говоря, не является доверительным, так как его границы зависят от оцениваемого параметра θ. Ситуацию можно исправить следующим образом.

a) Если $B_N(\theta) \equiv B_N$ (т.е. нет зависимости от θ), то (10) задаёт настоящий доверительный интервал.

b) Если $B_N(\theta) \leqslant B_N$, где B_N не зависит от θ, то доверительный интервал вида

$$\hat{\theta}_N - C(\gamma) \cdot B_N < \theta < \hat{\theta}_N + C(\gamma) \cdot B_N,$$

как более широкий, имеет доверительный уровень не менее γ.

c) Иногда удаётся разрешить непосредственно неравенство (9) и получить решение в виде интервала.

Пример. Случайная величина ξ принимает значения 1 и 0 с вероятностями p и $1 - p$ соответственно, т.е. p есть вероятность появления некоторого события A, которую необходимо оценить. Пусть $X = (X_1, \ldots, X_N)$ есть повторная выборка. Тогда случайная величина $S_N = X_1 + \cdots + X_N$ есть число успехов в схеме Бернулли и имеет биномиальное распределение с параметрами N и p. В силу интегральной теоремы Муавра-Лапласа случайная величина

$$S_N^* = \frac{S_N - N \cdot p}{\sqrt{Np(1-p)}}$$

имеет асимптотически стандартное нормальное распределение. По заданному γ по таблицам нормального распределения найдём такую константу $C(\gamma) > 0$, для которой

$$P\big(|S_N^*| < C(\gamma)\big) \approx 2\Phi_0\big(C(\gamma)\big) = \gamma.$$

Для построения доверительного интервала для p нам необходимо решить неравенство

$$\left| \frac{S_N - N \cdot p}{\sqrt{Np(1-p)}} \right| = \left| \frac{\bar{X} - p}{\sqrt{p(1-p)/N}} \right| < C(\gamma). \tag{11}$$

Это квадратичное неравенство можно решить и получить некоторый интервал. Простой (но более широкий!) интервал можно получить, если заметить, что $p(1-p) \leqslant \frac{1}{4}$. Тогда

$$\bar{X} - \frac{C(\gamma)}{2\sqrt{N}} < p < \bar{X} + \frac{C(\gamma)}{2\sqrt{N}}.$$

Глава 5

Критерии согласия

Формулировки задач и связанных с ними результатов, которые мы рассматривали выше, всегда были связаны с предположением, что мы имеем повторную выборку $X = (X_1, \ldots, X_N)$ из генеральной совокупности с функцией распределения $F(y)$, которая принадлежит некоторому конкретному (как правило, параметрическому) семейству $\mathcal{F}$. В этом параграфе мы рассмотрим статистические процедуры, которые позволяют проверить справедливость этого предположения. Такие процедуры называются **критериями согласия**.

5.1 Общий метод построения критериев согласия

Сформулируем более аккуратно стоящую перед нами задачу. Пусть $\mathcal{F}$ есть некоторое семейство функций распределения. ξ есть случайная величина с неизвестной функцией распределения $F_\xi(y)$. **Гипотезой о виде распределения** называется предположение о том, что $F_\xi \in \mathcal{F}$. Кратко это будет записываться в виде:

$$H: \ F_\xi \in \mathcal{F}.$$

Мы не знаем истинного распределения F_ξ, но на основе повторной выборки $X = (X_1, \ldots, X_N)$ можем построить его оценку, например, взять эмпирическую функцию распределения F_N^*. Выберем в пространстве всех функций распределения некоторую метрику (расстояние) ρ и оценим согласие эмпирических данных с нашей гипотезой по величине расстояния F_N^* от класса $\mathcal{F}$, которое определим по формуле

$$\Delta_N := \inf_{F \in \mathcal{F}} \rho(F_N^*, F).$$

Так как F_N^* построена по выборке, то величина Δ_N является случайной. Предположим, что нам удалось найти её распределение при условии, что гипотеза H верна. Для некоторого малого положительного числа α найдём положительную константу C, для которой

$$P\big(\Delta_N > C|H\big) \leqslant \alpha.$$

Таким образом, если гипотеза H верна, то вероятность того, что Δ_N принимает большие значения ($> C$), мала и можно считать, что такое событие «практически невозможно».

Теперь правило проверки H можно сформулировать следующим образом: если реально полученное значение Δ_N больше C, то мы **отвергаем** H, в противном случае говорим, что H **не противоречит экспериментальным данным**.

Далее мы рассматриваем несколько конкретных примеров применения этого общего подхода.

5.2 Критерий согласия Колмогорова

Мы рассматриваем гипотезу

$$H: \; F_\xi(y) \equiv F_0(y),$$

где $F_0(y)$ — точно заданная непрерывная функция распределения.

А. Н. Колмогоров предложил рассматривать следующее расстояние между функциями распределения:

$$\Delta_N := \sup_y \big|F_N^*(y) - F_0(y)\big|.$$

Теорема 7 (А. Н. Колмогоров, 1933.) *Если гипотеза H верна, то при $N \to \infty$*

$$P\big(D_N = \sqrt{N} \cdot \Delta_N < y\big) \longrightarrow K(y),$$

где $K(y)$ — некоторая явно вычисляемая функция распределения.

Для функции распределения $K(y)$ составлены таблицы (см. Большев А. Н., Смирнов Н. И., 1972).

Далее по таблицам для заданного α находим $K_\alpha > 0$:

$$P(D_N > K_\alpha) \approx \alpha.$$

Если реально полученное D_N будет больше K_α, то мы отвергаем гипотезу H, в противном случае говорим, что H не противоречит экспериментальным данным.

Главный недостаток этого критерия в том, что он не применим к ситуации, когда семейство $\mathcal{F}$ содержит более одного распределения.

5.3 χ^2–критерий согласия Пирсона

Мы вновь начинаем с простейшей ситуации

$$H: \ F_\xi(y) \equiv F_0(y),$$

где $F_0(y)$ — точно заданная функция распределения. Пусть $X = (X_1, \ldots, X_N)$ есть повторная выборка. Рассмотрим разбиение

$$-\infty = a_0 < a_1 < \cdots < a_k < a_{k+1} < \cdots < a_r < a_{r+1} = +\infty$$

вещественной прямой $\mathbf{R}^1$ (множество значений случайной величины ξ) на $r+1$ интервал. Пусть $p_k := F_0(a_{k+1}) - F_0(a_k)$ есть гипотетическая вероятность попадания случайной величины ξ в k–ый интервал, n_k — число элементов выборки, попавших в этот интервал. Для проверки гипотезы естественно сравнить гипотетические вероятности p_k и их оценки n_k/N, то есть частоты. В качестве меры сравнения Карл Пирсон предложил использовать следующую величину:

$$\Delta_N := \sum_{k=0}^{r} \frac{(n_k - N \cdot p_k)^2}{N \cdot p_k} = N \cdot \sum_{k=0}^{r} \left(\frac{n_k}{N} - p_k\right)^2 / (p_k).$$

Теорема 8 (К. Пирсон, 1900.) *Если гипотеза H верна, то при $N \to \infty$ случайная величина ΔN имеет асимптотически χ^2–распределение с $r = (r+1) - 1$ степенями свободы.*

Далее по таблицам для заданного α находим константу $\chi^2_r(\alpha) > 0$:

$$P\big(\Delta N > \chi^2_r(\alpha)\big) \approx \alpha.$$

Если реально полученное Δ_N будет больше $\chi^2_r(\alpha)$, то отвергаем H, в противном случае говорим, что H не противоречит экспериментальным данным.

Критерий Пирсона слабее критерия Колмогорова, но зато он применим в более сложной ситуации. Пусть мы рассматриваем гипотезу

$$H: \ F_\xi(y) = F(y, \theta_1, \ldots, \theta_m),$$

где вид функции распределения $F(y, \theta_1, \ldots, \theta_m)$ полностью задан, но параметры $(\theta_1, \ldots, \theta_m) = \theta \in \Theta \subset \mathbf{R}^m$ неизвестны. В этом случае гипотетические вероятности

$$P_k(\theta) = F(a_{k+1}, \theta) - F(a_k, \theta)$$

есть функции от неизвестных параметров и, потому, также неизвестны.

В этом случае предлагается рассмотреть величину

$$\Delta_N := \inf_{\theta \in \Theta} \Delta_N(\theta) = \inf_{\theta \in \Theta} \sum_{k=0}^{r} \frac{\big(n_k - N \cdot p_k(\theta)\big)^2}{N \cdot p_k(\theta)},$$

т.е. сравнивать экспериментальные данные с наиболее подходящим к ним распределением из нашего семейства $\mathcal{F} = \big\{F(y, \theta), \quad \theta \in \Theta\big\}$.

Теорема 9 (Р. Фишер, 1922.) *Если гипотеза H верна, то при $N \to \infty$ случайная величина ΔN имеет асимптотически χ^2–распределение с $r - m = (r+1) - 1 - m$ степенями свободы.*

Далее критерий согласия строится также, как и раньше.

Замечание. Можно показать, что наименьшее значение $\Delta_N(\theta)$ достигается при том же значении θ^* параметра $\theta \in \Theta$, при котором достигается наибольшее значение величины

$$L(X, \theta) = \frac{N!}{n_0! \cdot \cdots \cdot n_r!} \big[p_0(\theta)\big]^{n_0} \cdot \cdots \cdot \big[p_r(\theta)\big]^{n_r},$$

т.е. θ^* есть полиномиальная оценка наибольшего правдоподобия.

5.4 Проверка однородности двух выборок

Во многих прикладных задачах мы сталкиваемся с ситуацией, когда получены две повторные выборки $X = (X_1, \dots, X_{N_1})$, $Y = (Y_1, \dots, Y_{N_2})$. Важным является вопрос о том, будут ли они выборками из одной и той же генеральной совокупности. Если это так, то данные можно объединить и обрабатывать как единую выборку большего объёма. Пусть F_1 и F_2 есть функции распределения, отвечающие выборкам X и Y соответственно.

Гипотеза однородности имеет вид:

$$H: \; F_1(y) \equiv F_2(y),$$

причём F_1 и F_2 неизвестны. Существуют различные процедуры для проверки такой гипотезы. Мы рассмотрим только два критерия.

a) Критерий однородности Смирнова.

Пусть $F_{1,N_1}(y)$ и $F_{2,N_2}(y)$ — эмпирические функции распределения, построенные по выборкам X и Y соответственно. Определим величину

$$D_{N_1,N_2} := \sqrt{\frac{N_1 \cdot N_2}{N_1 + N_2}} \cdot \sup_y \left| F_{1,N_1}(y) - F_{2,N_2}(y) \right|.$$

Теорема 10 (Н. В. Смирнов, 1939.) *Если верна гипотеза H, то при $N_1, N_2 \to \infty$*

$$P(D_{N_1,N_2} < y) \to K(y), \quad y > 0,$$

где $K(y)$ — функция распределения Колмогорова.

Далее описание процедуры проверки гипотезы H такое же, как и раньше.

b) χ^2–критерий однородности.

Пусть как и ранее выбрано некоторое разбиение

$$-\infty = a_0 < a_1 < \dots < a_k < a_{k+1} < \dots < a_r < a_{r+1} = +\infty$$

пространства $\mathbf{R}^1$ (множества значений). Обозначим через n_{k1} и n_{k2} число элементов выборок X и Y соответственно, попавших в k-ый интервал. Для проверки гипотезы H естественно сравнить частоты n_{k1}/N_1 и n_{k2}/N_2. Сравнение производится с помощью величины

$$\Delta_{N_1,N_2} := \sum_{k=0}^{r}\sum_{j=1}^{2}\frac{(n_{kj}-N_j\cdot n_{k.}/N)^2}{N_j\cdot n_{k.}} = N\left(\sum_{k=0}^{r}\sum_{j=1}^{2}\frac{n_{kj}^2}{N_j\cdot n_{k.}}-1\right),$$

где

$$n_{k.} = n_{k1}+n_{k2}, \quad N = N_1+N_2.$$

Теорема 11 *Если верна гипотеза H, то при $N_1, N_2 \to \infty$ случайная величина Δ_{N_1,N_2} имеет асимптотически χ^2–распределение с $r = \big[(r+1)-1\big](2-1)$ степенями свободы.*

Далее описание процедуры проверки H, аналогично приведённым выше.

Заметим, что в отличие от критерия Смирнова, этот критерий легко обобщается на случай $s \geqslant 2$ выборок.

5.5 Проверка гипотезы о независимости

Пусть $\xi = (\xi_1, \xi_2)$ — двумерный случайный вектор с функцией распределения $F(z_1, z_2)$, $F_1(z_1)$ — функция распределения для ξ_1, $F_2(z_2)$ — функция распределения для ξ_2. **Гипотеза о независимости** имеет вид:

$$H:\ F(z_1, z_2) = F_1(z_1)\cdot F_2(z_2).$$

Рассмотрим два разбиения

$$-\infty = a_0 < a_1 < \cdots < a_k < a_{k+1} < \cdots < a_r < a_{r+1} = +\infty,$$
$$-\infty = b_0 < b_1 < \cdots < b_j < b_{j+1} < \cdots < b_s < b_{s+1} = +\infty.$$

Пусть мы имеем повторную выборку $(X_1, Y_1), \ldots, (X_N, Y_N)$ измерений случайного вектора (ξ_1, ξ_2). Обозначим через n_{kj} число элементов выборки, попавших в прямоугольник $[a_k, a_{k+1}) \times [b_j, b_{j+1})$, $n_{k.} = \sum\limits_j n_{kj}$, $n_{.j} = \sum\limits_k n_{kj}$, $N = \sum\limits_k n_{k.} = \sum\limits_j n_{.j}$. Если верна гипотеза H, то вероятность попадания в $[a_k, a_{k+1}) \times [b_j, b_{j+1})$ равна произведению вероятностей попадания в интервалы $[a_k, a_{k+1})$ и $[b_j, b_{j+1})$. Поэтому для проверки гипотезы H предлагается использовать величину

$$\Delta_N := N \cdot \sum_{k=0}^{r} \sum_{j=0}^{s} \frac{(n_{kj} - n_{k\cdot} \cdot n_{\cdot j}/N)^2}{n_{k\cdot} \cdot n_{\cdot j}} = N \left(\sum_{k=0}^{r} \sum_{j=0}^{s} \frac{n_{kj}^2}{n_{k\cdot} \cdot n_{\cdot j}} - 1 \right).$$

Теорема 12 *Если верна гипотеза* H, *то при* $N \to \infty$ *случайная величина* Δ_N *имеет асимптотически* χ^2*–распределение с* $r \cdot s$ *степенями свободы.*

Далее описание процедуры проверки аналогично прежнему.

Обычно экспериментальные данные представляют в виде следующей таблицы:

$Y \qquad X$	$a_0 \div a_1$	$\dots$	$a_k \div a_{k+1}$	$\dots$	$a_r \div a_{r+1}$
$b_0 \div b_1$	n_{00}	$\dots$	n_{k0}	$\dots$	n_{r0}
$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
$b_j \div b_{j+1}$	n_{0j}	$\dots$	n_{kj}	$\dots$	n_{rj}
$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
$b_s \div b_{s+1}$	n_{0s}	$\dots$	n_{ks}	$\dots$	n_{rs}

называемой **таблицей сопряжённости**. Поэтому соответствующий критерий называется **анализом таблиц сопряжённости**.

Глава 6

Проверка статистических гипотез

В этом разделе будут изложены основные понятия теории проверки статистических гипотез.

6.1 Что такое статистическая гипотеза?

Математической моделью статистического эксперимента является статистическая структура $(\Omega, \mathcal{A}, \mathcal{P})$. Пытаясь уточнить нашу модель, мы делаем те или иные предположения о неизвестном распределении вероятностей $P \in \mathcal{P}$.

Примеры. 1. В эксперименте изучается случайная величина ξ с неизвестной функцией распределения F_ξ. Мы предполагаем, что $F_\xi \in \mathcal{F}$, где $\mathcal{F}$ — некоторое заданное семейство функций распределения, например, нормальные функции распределения.

2. Известно, что случайная величина ξ имеет нормальное распределение. Если ξ является ошибкой измерения некоторого прибора, то для хорошо отлаженного прибора естественно предположить, что $M(\xi) = a = 0$.

3. В некотором эксперименте изучаются две числовые характеристики ξ_1 и ξ_2, которые являются случайными величинами. Иногда можно предположить, что ξ_1 и ξ_2 независимы.

Все рассмотренные выше примеры имеют одну и ту же особенность, мы делаем то или иное предположение о неизвестном распределении вероятностей. Таким образом, неформально под **статистической гипотезой** понимают любое предположение о распределении вероятностей в рассматриваемом эксперименте. С формальной точки зрения, если мы что-то предположили об этом распределении, то этим выделили некоторый подкласс $\mathcal{P}_0$ в классе всех априорных возможных распределений $\mathcal{P}$.

Определение. Статистической гипотезой называется подкласс $\mathcal{P}_0 \subset \mathcal{P}$. Условно это записывается в виде

$$H: \quad P \in \mathcal{P}_0.$$

Статистические гипотезы будут обозначаться H, H_0, $H_1, \ldots$.

6.2 Основные понятия теории проверки гипотез

Пусть мы имеем статистическую структуру $\big(\Omega, \mathcal{A}, \mathcal{P}\big)$ и сформулировали некоторую статистическую гипотезу $H : P \in \mathcal{P}_0$. Если $\mathcal{P}_0$ содержит только один элемент, т.е. мы полностью фиксируем распределение, то гипотеза H называется **простой**. В противном случае гипотеза называется **сложной**. Если семейство $\mathcal{P}$ является параметрическим, т.е. $\mathcal{P} = \big\{P_\theta, \theta \in \Theta\big\}$, а гипотеза H имеет вид $\theta \in \Theta_0 \subset \Theta$, то она называется **параметрической**.

Обычно формулируют несколько гипотез. Одну из них выделяют в качестве основной и называют ее **нулевой**. Как правило, ее обозначают через H_0. Остальные гипотезы называют **альтернативами** и обозначают H_1, $H_2, \ldots$.

Всюду далее мы рассматриваем случай, когда нулевая гипотеза H_0 проверяется против одной альтернативы H_1. Для простоты обозначений мы предполагаем, что эти гипотезы являются параметрическими и записываем их в виде:

$$\begin{aligned} H_0: &\quad \theta \in \Theta_0, \\ H_1: &\quad \theta \in \Theta_1, \end{aligned}$$

где Θ_0, Θ_1 есть подмножества основного пространства параметров Θ, такие, что $\Theta_0 \cup \Theta_1 = \Theta$ и $\Theta_0 \cap \Theta_1 = \emptyset$.

Основная задача состоит в том, чтобы на основе экспериментальных данных выбрать одну из предложенных нам гипотез H_0 и H_1. Это делается с помощью так называемых **критериев** или **решающих правил**. В качестве экспериментальных данных мы используем, как и раньше, повторную выборку $X = (X_1, \dots, X_N)$. Обозначим через $\mathcal{X}$ выборочное пространство, т.е. множество возможных значений x выборки X.

Определение. Статистическим критерием (тестом, решающим правилом) для проверки гипотезы H_0 против альтернативы H_1 называется произвольное отображение $\varphi : \mathcal{X} \to \{0, 1\}$. Если $\varphi(x) = 0$, то принимаем гипотезу H_0. Если $\varphi(x) = 1$, то принимаем гипотезу H_1.

Обозначим через K подмножество в $\mathcal{X}$, где мы принимаем H_1, т.е.

$$K = \Big\{x \in \mathcal{X} : \quad \varphi(x) = 1\Big\}.$$

Множество K называется **критической зоной** теста φ. Задание критической зоны K однозначно определяет тест φ. Действительно, $\varphi(x) = 1$ тогда и только тогда, когда $x \in K$. Очень часто критическая зона теста задается по правилу

$$K = \Big\{x \in \mathcal{X} : \quad T(x) > c\Big\},$$

или

$$K = \Big\{x \in \mathcal{X} : \quad T(x) < c\Big\},$$

где $T(x)$ — некоторая статистика, называемая **статистикой критерия** φ, а c — некоторая константа, называемая **критической константой**.

Иногда возникает такая ситуация, когда для полученной выборки x мы не можем однозначно принять решение в пользу гипотезы H_0 или H_1. В этом случае приписывают некоторую вероятность тому, что верна H_0, и дополнительную вероятность тому, что верна H_1. Такой вариант принятия решения фиксируется в следующем понятии.

Определение. Отображение $\varphi : \mathcal{X} \to [0, 1]$ называется **рандомизированным** критерием (тестом, решающим правилом). Значение $\varphi(x)$ интерпретируется как **вероятность принять гипотезу** H_1, если получена выборка x. Если φ принимает только значения 0 и 1, то критерий называется **нерандомизированным**.

Обычно необходимость в рандомизации возникает на границе критической области.

Для проверки гипотезы H_0 против альтернативы H_1 можно придумать много различных критериев. Хотелось бы выбрать среди них в определенном смысле наилучший. Интуитивно мы стремимся построить такое правило, при котором ошибки возникают редко. В нашей задаче есть два типа ошибок. **Ошибки первого рода:** мы принимаем гипотезу H_1, когда верна H_0. **Ошибки второго рода:** мы принимаем H_0, когда верна H_1. Так как выборка x получается в результате случайного эксперимента, но мы не знаем заранее, к какому решению мы придем, применяя правило φ. Поэтому при оценке качества теста φ естественно исходить из величины **вероятностей ошибок** первого и второго рода. Если θ есть истинное значение неизвестного параметра, то

$$\alpha(\theta) := P_\theta(X \in K), \quad \theta \in \Theta_0, \tag{1}$$

определяет вероятность ошибки первого рода, а

$$1 - \beta(\theta) := P_\theta(X \notin K), \quad \theta \in \Theta_1, \tag{2}$$

определяет вероятность ошибки второго рода. Нам бы хотелось минимизировать величины $\alpha(\theta)$ для всех $\theta \in \Theta_0$ и $1-\beta(\theta)$ для всех $\theta \in \Theta_1$. Но из выражений (1) и (2) мы видим, что для уменьшения $\alpha(\theta)$ нужно сужать критическую область K, а для уменьшения $1 - \beta(\theta)$ нужно, наоборот, расширять критическую область K. Т.е. мы приходим к двум «конфликтующим» критериям качества (вспомните построение доверительных интервалов!). Улучшая один, мы ухудшаем другой, и наоборот. В качестве H_0 обычно выбирают более важную для нас гипотезу, которую мы не хотим отвергать (альтернативно, не хотим принимать H_1) без достаточных на то оснований. В силу этого мы выбираем некоторое малое α и рассматриваем только такие критерии φ, для которых

$$\sup_{\theta \in \Theta_0} \alpha(\theta) \leqslant \alpha. \tag{3}$$

Такие критерии называются критериями **уровня значимости** α. Точное значение верхней грани в (3) называется **размером** критерия φ. Затем среди всех критериев уровня α мы ищем такой, для которого величина $1-\beta(\theta)$, $\theta \in \Theta_1$ будет наименьшей. Эквивалентно, можно максимизировать величину $\beta(\theta)$. Функция

$$\beta(\theta) := P_\theta(X \in K), \quad \theta \in \Theta_1, \tag{4}$$

называется **функцией мощности**. Она показывает, какова вероятность принять гипотезу H_1, когда она верна, для различных распределений, входящих в H_1. Хотелось бы найти такой критерий φ_0, у которого функция мощности $\beta(\theta)$ была наибольшей среди всех критериев φ уровня α при любых $\theta \in \Theta_1$.

Определение. Критерий φ_0 уровня α называется **равномерно наиболее мощным критерием уровня** α, если для любого другого критерия φ уровня α имеет место

$$\beta_{\varphi_0}(\theta) \geqslant \beta_\varphi(\theta), \quad \forall\theta \in \Theta_1. \tag{5}$$

Заметим, что нам нужно максимизировать мощность при всех значениях $\theta \in \Theta_1$. Это удается сделать довольно редко.

6.3 Проверка простой гипотезы против простой альтернативы

Как мы отмечали выше, равномерно наиболее мощные тесты существуют не всегда. Но если мы проверяем простую гипотезу H_0 против простой альтернативы H_1, т.е.

$$\begin{aligned} H_0: &\quad \theta = \Theta_0, \\ H_1: &\quad \theta = \Theta_1, \end{aligned}$$

то наилучший (оптимальный) критерий существует. Чтобы понять, как он должен выглядеть, рассмотрим дискретную случайную величину ξ. Обозначим через $P_0(x)$ и $P_1(x)$ вероятности события $(X = x)$ в случае выполнения гипотез H_0 и H_1 соответственно. Для построения критерия нам необходимо описать критическую зону K. Следуя принципу наибольшего правдоподобия естественно помещать в K те точки x, для которых больше отношение $P_1(x)/P_0(x)$ (т.е. те, которые больше говорят в пользу H_1).

Лемма (Е. Нейман, Е. Пирсон, 1933). *Пусть случайная величина ξ имеет дискретное распределение и проверяется простая гипотеза H_0 против простой альтернативы H_1. Тогда среди всех тестов φ уровня α существует наиболее мощный тест φ_0, и он имеет следующий вид:*

$$\varphi_0(x) = \begin{cases} 1, & \text{если } P_1(x) > c \cdot P_0(x), \\ 0, & \text{если } P_1(x) < c \cdot P_0(x), \\ \gamma, & \text{если } P_1(x) = c \cdot P_0(x), \end{cases}$$

где константы $c > 0$ и $0 \leq \gamma \leq 1$ определяются из условия

$$P_0(X \in K) = \alpha.$$

Доказательство. Заметим, что для любого критерия φ его уровень и мощность вычисляются по формулам

$$\alpha_\varphi = M_0\varphi(X) = \sum_x \varphi(x)P_0(x), \quad \beta_\varphi = M_1\varphi(X) = \sum_x \varphi(x)P_1(x).$$

Пусть критерий φ (вообще говоря, рандомизированный) имеет уровень α, т.е. $\alpha_\varphi \leqslant \alpha$. Предположим, что $0 < \alpha < 1$, и определим множества

$$S^+ = \big\{x : \varphi_0(x) - \varphi(x) > 0\big\},$$
$$S^- = \big\{x : \varphi_0(x) - \varphi(x) < 0\big\}.$$

Если $x \in S^+$, то

$$\varphi_0(x) > 0 \quad \text{и} \quad P_1(x) \geqslant c \cdot P_0(x).$$

Если $x \in S^-$, то

$$\varphi_0(x) < 0 \quad \text{и} \quad P_1(x) \leqslant c \cdot P_0(x).$$

Используя эти соотношения, получим

$$\overbrace{\sum_x [\varphi_0(x)-\varphi(x)]P_1(x)}^{\beta_{\varphi_0}-\beta_{\varphi_1}} - c\cdot \overbrace{\sum_x [\varphi_0(x)-\varphi(x)]P_0(x)}^{\alpha_{\varphi_0}-\alpha_{\varphi_1}} =$$
$$= \sum_x [\varphi_0(x)-\varphi(x)][P_1(x)-c\cdot P_0(x)] =$$
$$= \sum_{x\in S^+\cup S^-} [\varphi_0(x)-\varphi(x)][P_1(x)-c\cdot P_0(x)] =$$
$$= \sum_{x\in S^+} [\varphi_0(x)-\varphi(x)][P_1(x)-c\cdot P_0(x)] +$$
$$+ \sum_{x\in S^-} [\varphi_0(x)-\varphi(x)][P_1(x)-c\cdot P_0(x)] \geqslant 0.$$

Отсюда следует, что

$$\beta_{\varphi_0}-\beta_{\varphi} = \sum_x \varphi_0(x)\cdot P_1(x) - \sum_x \varphi(x)\cdot P_1(x) =$$
$$= \sum_x [\varphi_0(x)-\varphi(x)]\cdot P_1(x) \geqslant c\cdot \sum_x [\varphi_0(x)-\varphi(x)]\cdot P_0(x) =$$
$$= c\cdot(\alpha-\alpha_\varphi) \geqslant 0.$$

Замечание. Доказанный выше результат справедлив и для абсолютно непрерывных распределений. В этом случае в формулировке леммы вероятности нужно заменить на плотности, а при доказательстве заменить суммы на интегралы.

Пример. Случайная величина ξ имеет нормальное распределение с параметрами (a,σ^2). Предполагая, что дисперсия $\sigma^2=\sigma_0^2$ известна, необходимо проверить гипотезу

$$H_0: \quad a=a_0$$

против альтернативы

$$H_1: \quad a=a_1.$$

Пусть для определенности $a_1>a_0$. Если $X=(X_1,\ldots,X_N)$ есть повторная выборка, то плотность распределения случайного вектора X в точке $x=(x_1,\ldots,x_N)$ имеет вид

$$\rho(x)=(2\pi\sigma_0^2)^{-\frac{N}{2}}\exp\left\{-\frac{1}{2\sigma_0^2}\sum_{j=1}^N (x_j-a)^2\right\}.$$

По лемме Неймана–Пирсона нам нужно рассмотреть отношение

$$\frac{\rho_1(x)}{\rho_0(x)} = \frac{\exp\left\{-\frac{1}{2\sigma_0^2}\sum\limits_{j=1}^{N}(x_j-a_1)^2\right\}}{\exp\left\{-\frac{1}{2\sigma_0^2}\sum\limits_{j=1}^{N}(x_j-a_0)^2\right\}} = \frac{\exp\left\{-\frac{1}{2\sigma_0^2}\left[\sum\limits_{j=1}^{N}x_j^2-2a_1\sum\limits_{j=1}^{N}x_j+Na_1^2\right]\right\}}{\exp\left\{-\frac{1}{2\sigma_0^2}\left[\sum\limits_{j=1}^{N}x_j^2-2a_0\sum\limits_{j=1}^{N}x_j+Na_0^2\right]\right\}} =$$
$$= \exp\left\{-\tfrac{1}{2\sigma_0^2}\left[-2N(a_1-a_0)\cdot\bar{x}+N(a_1^2-a_0^2)\right]\right\}.$$

Критическая область оптимального критерия φ_0 имеет вид

$$\frac{\rho_1(x)}{\rho_0(x)} = \exp\left\{\frac{N}{\sigma_0^2}(a_1-a_0)\cdot\bar{x}\right\}\cdot\exp\left\{-\frac{N}{2\sigma_0^2}(a_1^2-a_0^2)\right\} > c.$$

Произведем последовательно несколько эквивалентных преобразований, которые дают различные варианты записи одной и той же области K.

$$\begin{array}{lll} \rho_1(x)/\rho_0(x) > c & \iff & \frac{N}{\sigma_0^2}(a_1-a_0)\cdot\bar{x} \quad \iff \\ \bar{x} > c_2 & \iff & \frac{\bar{x}-a_0}{\sigma_0}\cdot\sqrt{N} > c_3. \end{array}$$

Случайная величина
$$y = \frac{\bar{x}-a_0}{\sigma_0}\cdot\sqrt{N}$$
в случае гипотезы H_0 имеет стандартное нормальное распределение. Таким образом, константу c_3 можно найти по таблицам нормального распределения из условия

$$P_0(y > c_3) = 1-\Phi(c_3) = \frac{1}{2}-\Phi_0(c_3) = \alpha.$$

Окончательно оптимальный критерий φ имеет вид

$$\varphi_0(x) = \begin{cases} 1, & \text{если } \bar{x} > a_0 + \frac{c_3\cdot\sigma_0}{\sqrt{N}} = c_4, \\ 0, & \text{если } \bar{x} \leqslant c_4. \end{cases}$$

Полезно нарисовать следующую картинку, которая наглядно иллюстрирует полученный результат.

6.4 Проверка сложных гипотез. Критерий отношения правдоподобия

Конечно, случай, когда мы проверяем простую гипотезу H_0 против простой альтернативы H_1, является исключительным и редко встречается в реальных задачах. Обычно нужно проверять сложные гипотезы. Но лемма Неймана–Пирсона подсказывает метод построения критериев для сложных гипотез. Этот метод был предложен Нейманом и Пирсоном в 1928 году.

Пусть случайная величина ξ имеет функцию распределения $F(y, \theta)$, где $\theta \in \Theta$ — неизвестный параметр. Пусть Θ_0 и Θ_1 — подмножества Θ, для которых $\Theta_0 \cap \Theta_1 = \emptyset$, $\Theta_0 \cup \Theta_1 = \Theta$. Проверяется гипотеза

$$H_0 : \quad \theta \in \Theta_0$$

против альтернативы

$$H_1 : \quad \theta \in \Theta_1.$$

Для проверки таких гипотез предлагается использовать статистику

$$\lambda(x) = \frac{\sup\limits_{\theta \in \Theta_1} \mathcal{L}(x, \theta)}{\sup\limits_{\theta \in \Theta_0} \mathcal{L}(x, \theta)}$$

или, эквивалентно,

$$\lambda_1(x) = \frac{\sup\limits_{\theta \in \Theta} \mathcal{L}(x, \theta)}{\sup\limits_{\theta \in \Theta_0} \mathcal{L}(x, \theta)} = \max\bigl(\lambda(x), 1\bigr)$$

где $\mathcal{L}(x, \theta)$ есть функция правдоподобия, т.е. распределение вероятности (плотность) повторной выборки $X = (X_1, \ldots, X_N)$. Это естественное обобщение той статистики, что используется в лемме Неймана–Пирсона. Если удается найти распределение этих статистик, соответствующий критерий, называемый **критерием отношения правдоподобия**, задается с помощью критической области следующего вида

$$K = \bigl\{x : \quad \lambda_1(x) > c\bigr\},$$

где критическая константа c находится из условия

$$P_\theta(\lambda_1(x) > c) \leqslant \alpha, \quad \theta \in \Theta.$$

Пример. Случайная величина ξ имеет нормальное распределение с параметрами (a, σ^2), причем σ^2 неизвестно. Проверяется гипотеза

$$H_0: \quad a = a_0$$

против альтернативы

$$H_1: \quad a \neq a_0.$$

Заметим, что обе гипотезы H_0 и H_1 являются сложными, т.к.

$$\Theta_0 = \big\{(a, \sigma^2) : a = a_0, \sigma^2 > 0\big\}, \quad \Theta_1 = \big\{(a, \sigma^2) : a \neq a_0, \sigma^2 > 0\big\}.$$

Пусть $X = (X_1, \ldots, X_N)$ есть повторная выборка. Тогда функция правдоподобия (плотности распределения выборки X) имеет вид

$$\mathcal{L}(x, a, \sigma^2) = (2\pi\sigma^2)^{-\frac{N}{2}} \exp\left\{-\frac{1}{2\sigma^2}\sum_{j=1}^{N}(x_j - a)^2\right\}.$$

При оценке параметров по методу наибольшего правдоподобия мы уже нашли, что

$$\sup_{a,\sigma^2} \mathcal{L}(x, a, \sigma^2) = \mathcal{L}(x, \bar{x}, S^2),$$

где $\bar{x}$ и S^2 есть выборочное среднее и выборочная дисперсия соответственно.

Аналогично можно показать, что

$$\sup_{a=a_0,\sigma^2} \mathcal{L}(x, a_0, \sigma^2) = \mathcal{L}(x, a_0, S_0^2),$$

где

$$S_0^2 = \frac{1}{N}\sum_j (x_j - a_0)^2 = S^2 + (\bar{x} - a_0)^2.$$

В этом случае

$$\lambda_1(x) = \frac{(2\pi S^2)^{-\frac{N}{2}} \cdot e^{-\frac{N}{2}}}{\big(2\pi[S^2 + (\bar{x} - a_0)^2]\big)^{-\frac{N}{2}} \exp\left\{-\frac{N}{2}\right\}} = \left(\left|\frac{\bar{x} - a_0}{S}\right|^2 + 1\right)^{\frac{N}{2}}.$$

Критическая область имеет вид

$$\lambda_1(x) > c$$

или, эквивалентно,

$$|t_{N-1}| = \left|\frac{\bar{x} - a_0}{S} \cdot \sqrt{N-1}\right| > C_1.$$

Как отмечалось ранее (при построении доверительных интервалов), если верна гипотеза H_0, то случайная величина t_{N-1} имеет распределение Стьюдента с $N-1$ степенями свободы. Тогда для заданного α константа c_1 находится по таблицам распределения Стьюдента из соотношения

$$P_0\big(|t_{N-1}| > c_1\big) = \alpha.$$

Задача. Дать описание критериев отношения правдоподобия для проверки следующих гипотез.

1) $X = (X_1, \ldots, X_N)$ — повторная выборка из генеральной совокупности с нормальным распределением с параметрами (a, σ^2). Проверяется гипотеза

$$H_0: \quad \sigma^2 = \sigma_0^2$$

против альтернативы

$$H_1: \quad \sigma^2 \neq \sigma_0^2.$$

2) $X = (X_1, \ldots, X_N)$ и $Y = (Y_1, \ldots, Y_N)$ — независимые повторные выборки из генеральной совокупности с нормальными распределениями с параметрами (a_1, σ^2) и (a_2, σ^2) соответственно. Проверяется гипотеза

$$H_0: \quad a_1 = a_2$$

против альтернативы

$$H_1: \quad a_1 \neq a_2.$$

6.5 Проверка сложных гипотез для распределений с монотонным отношением правдоподобия

В этом случае рассматривается семейство распределений с одним специальным свойством, для которого существуют равномерно наиболее мощные тесты для некоторых сложных гипотез.

Пусть случайная величина ξ имеет плотность распределения $\rho(y,\theta)$, где $\theta \in \Theta \subset \mathbf{R}^1$ есть скалярный параметр. Пусть, далее, $X = (X_1, \ldots, X_N)$ есть повторная выборка из генеральной совокупности с таким распределением. Предположим, что функция правдоподобия

$$\mathcal{L}(x,\theta) = \rho(x_1,\theta) \cdot \cdots \cdot \rho(x_N,\theta)$$

допускает представление

$$\mathcal{L}(x,\theta) = g\bigl(T(x),\theta\bigr) \cdot h(x), \tag{6}$$

где $g(t,\theta)$ — некоторая функция от двух скалярных аргументов. Статистика $T(x)$ называется **достаточной** статистикой для параметра θ.

Определение. Семейство распределений $\mathcal{P} = \bigl\{\rho(y,\theta), \theta \in \Theta \subset \mathbf{R}^1\bigr\}$ имеет **монотонное отношение правдоподобия**, если для любых фиксированных $\theta_0 < \theta_1$ из Θ функция $g(t,\theta_1)/g(t,\theta_0)$ монотонно возрастает (убывает) по t.

Проверяется гипотеза

$$H_0: \quad \theta = \theta_0$$

против альтернативы

$$H_1: \quad \theta \geqslant \theta_1,$$

где $\theta_1 > \theta_0$. Выберем некоторое фиксированное $\theta' \geqslant \theta_1$ и построим критерий Неймана–Пирсона для проверки простой гипотезы

$$H_0: \quad \theta = \theta_0$$

против простой альтернативы

$$H_0': \quad \theta = \theta'.$$

Критическая зона этого критерия выделяется условием

$$\mathcal{L}(x,\theta')/\mathcal{L}(x,\theta_0) = g\big(T(x),\theta'\big)/g\big(T(x),\theta_0\big) > c.$$

В силу монотонности отношения правдоподобия это эквивалентно условию

$$T(x) > c_1. \tag{7}$$

Константа c_1 находится из условия

$$P_{\theta_0}\big(T(x) > c_1\big) = \alpha,$$

т.е. полностью определяется значением θ_0 и не зависит от частного выбора $\theta' \geqslant \theta_1$.

В силу леммы Неймана–Пирсона построенный критерий является наиболее мощным для проверки H_0 против H_1' для любого $\theta' \geqslant \theta_1$, т.е. является равномерно наиболее мощным критерием уровня α для проверки H_0 против сложной альтернативы H_1.

Пример. ξ имеет нормальное распределение с параметрами (a,σ^2). Пусть $\sigma^2 = \sigma_0^2$ известно и проверяется гипотеза

$$H_0: \quad a = a_0$$

против альтернативы

$$H_1: \quad a \geqslant a_1,$$

где $a_1 > a_0$. В этой задаче функция правдоподобия равна

$$\begin{aligned}
&\mathcal{L}(x,a) = (2\pi\sigma_0^2)^{-\frac{N}{2}} \exp\left\{-\tfrac{1}{2\sigma_0^2}\sum_{j=1}^{N}(x_j-a)^2\right\} = \\
&= \exp\left\{\tfrac{1}{\sigma_0^2}\cdot a \cdot \sum_{j=1}^{N} x_j - \tfrac{N\cdot a^2}{2\sigma_0^2}\right\}\cdot(2\pi\sigma_0^2)^{-\frac{N}{2}} \exp\left\{-\tfrac{1}{2\sigma_0^2}\sum_{j=1}^{N} x_j^2\right\} = \\
&= g\big(T(x),a\big)\cdot h(x),
\end{aligned}$$

где $T(x) = \sum\limits_j x_j$, т.е. имеет вид (6). Это семейство с монотонным отношением правдоподобия, т.к. $g(t,a')/g(t,a_0)$ строго возрастает по

t при любых $a' > a_0$. В силу сказанного выше существует равномерно наиболее мощный тест для проверки H_0 против альтернативы H_1, критическая область которого выделяется условием

$$T(x) = \sum_{j=1}^{N} x_j > c.$$

После эквивалентных преобразований ее можно записать в виде

$$\frac{\bar{x} - a_0}{\sigma_0} \cdot \sqrt{N} > c_1.$$

Критическая константа c_1 находится из условия, что

$$P_0\left(\frac{\bar{x} - a_0}{\sigma_0} \cdot \sqrt{N} > c_1\right) = 1 - \Phi(c_1) = \frac{1}{2} - \Phi_0(c_1) = \alpha.$$

Здесь мы использовали тот факт, что при гипотезе H_0 статистика

$$Y = \frac{\bar{x} - a_0}{\sigma_0} \cdot \sqrt{N}$$

имеет стандартное нормальное распределение.

Замечание. Аналогичные результаты справедливы и для дискретных распределений, если во всех формулировках плотности заменить на вероятности.

Список литературы

а) Основной

1. Гнеденко Б.В. Курс теории вероятностей: Учебник – Изд. 6-е, перераб. и доп. – М.: Наука, 1988

2. Севастьянов Б.А. Курс теории вероятностей и математической статистики. - Москва-Ижевск, Иститут компьютерных исследований, 2004.

3. Чистяков В.П. Курс теории вероятностей. - М.: Наука, 1982.

4. Ширяев А.Н. Вероятность. - М.: МЦНМО, 2004.

5. Агапов Г.И. Задачник по теории вероятностей: Учебное пособие для студентов втузов. - М.: Высш. шк., 1986.

6. Зубков А.М., Севастьянов Б.А., Чистяков В.П. Сборник задач по теории вероятностей. - М.: Наука, 1989.

7. Ивченко Г.И., Медведев Ю.И. Математическая статистика: – М.: URSS, 2014.

b) Дополнительный

8. Колмогоров А.Н. Основные понятия теории вероятностей. - М.: Наука, 1974.

9. Феллер В. Введение в теорию вероятностей: В 2 т.- М.: Мир, 1984.

10. Вероятность и математическая статистика: Энциклопедия/ Гл. ред. Ю.В. Прохоров. – М.: Большая Российская энциклопедия, 1999. – 910 с.

11. Кендалл М., Стьюарт А. Статистические выводы и связи. – М.: Наука, 1973. – 899 с.

12. Крамер Г. Математические методы статистики. – М.: Мир, 1975. – 648 с.

13. Хургин Я.И. Да, нет или может быть. - М.: Наука, 1977.

14. Хургин Я.И. Как объять необъятное. - М.: Знание, 1979.

Технический редактор А.В. Жильцов
Подписано в печать 15.04.2014. Формат 60 84 1/16.
Усл. печ. л. 14,75. Тираж 10 экз. Заказ № 164.
Тверской государственный университет
Редакционно-издательское управление
Адрес: Россия, 170100, г. Тверь, ул. Желябова, 33.
Тел. РИУ: (4822) 35-60-63.