МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ

РОССИЙСКИЙ ТЕХНОЛОГИЧЕСКИЙ УНИВЕРСИТЕТ
МИРЭА

И. С. Пулькин

ЗАДАЧИ ПО ТЕОРИИ ВЕРОЯТНОСТЕЙ

Учебное пособие

Москва 2022

Это вторая страница.

Она позже должна быть убрана.

А пока она для того, чтобы не сбилась нумерация страниц.

Оглавление

Предисловие

Этот задачник предназначен для студентов технических специальностей. В первую очередь автор, конечно, имел в виду студентов РТУ МИРЭА. Во время работы в этом техническом университете автор опробовал и отрабатывал различные методики преподавания теории вероятностей и математической статистики.

Цель автора — ознакомить студентов с идеями и методами курса, а также с терминологией, чтобы они могли читать техническую литературу, а также инструкции и нормативные документы.

В последнее время основной фокус внимания в РТУ МИРЭА переместился с задач радиолокации, связи и автоматики на проблемы обработки больших данных, машинного обучения и создания искусственного интеллекта. Большинству выпускников предстоит работать в сфере информационных технологий. В этой сфере значение статистики трудно переоценить. В частности, все виды машинного обучения являются, по сути своей, методами статистического обучения.

Образцом для автора всегда был задачник В. Е. Гмурмана [3]. Но со времени его выхода в свет появилось много задач и новых тем, которые должны быть рассмотрены на семинарах по теории вероятностей.

В списке литературы приведены книги, которые автор рекомендовал бы студентам для ознакомления. Этот список, разумеется, отражает личные вкусы автора и ни в коем случае не претендует на полноту.

Автор выражает благодарность И. Р. Высоцкому за многочисленные ценные советы, помогавшие в работе над книгой.

Тема 1. Случайные события

1. Классическое определение вероятности

Для подсчета шансов используется классическое определение вероятности:

$$\text{Вероятность} = \frac{\text{число благоприятных исходов}}{\text{общее число равновозможных исходов}}.$$

Задача 1. Каковы шансы (какова вероятность) того, что при бросании трех монет орел выпадет ровно один раз?

Решение

Три монеты могут упасть восемью способами:
ооо
оор
оро
орр $\surd$
роо
рор $\surd$
рро $\surd$
ррр
Благоприятных исходов среди них три — они отмечены галочками. Поэтому вероятность равна 3/8.

Следует обратить внимание на то, что исходы бросания должны быть именно равновозможными. Для каждой монеты таких исходов два, и в сочетании с каждым из них могут случиться два исхода на второй и два исхода на третьей монете. Всего получится $2^3 = 8$ исходов. Аналогично для n монет получится 2^n равновозможных исходов.

Задача 2. Сколько равновозможных исходов будет при бросании 2 игральных костей? 3 костей? n костей? Какова вероятность выпадения 7 очков при бросании 2 костей? 8 очков?

Задача 3. Игральную кость бросают два раза. Какова вероятность того, что ни разу не выпадет ни 5, ни 6?

Задача 4. Каковы вероятности получить при бросании трех игральных костей в сумме 11 очков и 12 очков? Какая из этих вероятностей больше?

Решение

Благоприятных исходов много, для подсчета их количества пальцев рук уже не хватит, поэтому следует как-то упорядочить подсчет. Сумма 11 получится, например, как $1 + 4 + 6$. Если мы будем записывать различные подобные комбинации, то стоит учитывать сразу все возможные перестановки, то есть случаи, когда на какой-то кости — 1 очко, на какой-то другой — 4, а на третьей — 6. Таких перестановок, как нетрудно убедиться, будет шесть: 146, 164, 416, 461, 614, 641. Поэтому мы будем записывать все возможные наборы, дающие нужную нам сумму, а рядом будем писать, сколькими способами такие наборы получаются. Чтобы не запутаться, все наборы будем записывать в возрастающем порядке. Для 11 и 12 очков получим:

11		12	
1 4 6	6	1 5 6	6
1 5 5	3	2 4 6	6
2 3 6	6	2 5 5	3
2 4 5	6	3 3 6	3
3 3 5	3	3 4 5	6
3 4 4	3	4 4 4	1

В результате получаем, что вероятности таковы:

$$P(11) = \frac{27}{216}; \quad P(12) = \frac{25}{216}.$$

Задача 5. Какова вероятность получить при бросании трех костей в сумме 13 очков?

Задача 6. Игральную кость бросают три раза. Какова вероятность того, что сумма очков будет не больше 5?

Задача 7. Игральную кость бросают три раза. Какова вероятность того, что каждый раз буде выпадать разное число очков? Какова вероятность того, что будет хотя бы одна пара одинаковых исходов? Какая из этих вероятностей больше?

Задача 8. Игральную кость бросают три раза. Какова вероятность того, что при каждом следующем броске число выпавших очков будет больше, чем при предыдущем?

Задача 9. Игральную кость бросают три раза. Какова вероятность того, что числа выпавших очков образуют арифметическую прогрессию а) в том порядке, в каком они выпадают? б) после какой-либо перестановки*

Задача 10. Кубик сначала окрасили, а затем распилили на тысячу одинаковых кубиков. Какова вероятность того, что наудачу выбранный кубик будет иметь как минимум две окрашенные грани?

Задача 11. Вам предлагают такую игру: вы бросаете две монеты, а ваш противник — одну. При этом вы выигрываете, если у вас выпадет больше орлов, чем у вашего противника, в противном случае — проигрываете. Какова вероятность вашего выигрыша?

Задача 12. Раньше в поездах, на вокзалах иногда предлагали сыграть в такую азартную игру. Игрок зажимает в кулаке носовой платок так, что наружу торчат только 4 угла. Вы выбираете два угла и тянете за них. Если вы ухватились за противоположные углы платка — вы выиграли, а если за смежные — проиграли. Какова вероятность вашего выигрыша?

Задача 13. Двое садятся в трамвай с прицепом. Какова вероятность что они окажутся в одном вагоне?

Задача 14. Какова вероятность того, что случайно выбранное трехзначное число содержит ровно одну цифру 5?

Задача 15. Монету бросили 5 раз. Найти вероятность того что хотя бы раз было три орла подряд.

Задача 16. Двое бросают кубики. У первого на кубике четыре двойки и две шестерки, у второго — три пятёрки и три единицы. Выигрывает тот, у кого выпадет больше очков. Какова вероятность победы первого игрока?

Задача 17(*). Существует ли три таких набора очков на трех кубиках, чтобы с вероятностью больше $1/2$ первый кубик выигрывал у второго, второй — у третьего, а третий — у первого?

Задача 18. Существует ли три таких набора очков на трех кубиках, как в предыдущей задаче, чтобы при этом сумма всех очков на каждом кубике была одинаковой?

Задача 19. Есть 4 игральные карты: 2 чёрной и 2 красной масти. Из них выбирают две случайным образом. Какова вероятность того, что масти выбранных карт окажутся одного цвета?

Задача 20. В колоде $2n$ карт: n чёрной и n красной масти. Из них выбирают две случайным образом. Какова вероятность того, что масти выбранных карт окажутся одного цвета?

Задача 21. На шахматную доску случайным образом ставят белого короля и черную ладью. Какова вероятность, что белый король окажется под шахом?

Задача 22. Секретарша раскладывает четыре письма по четырём конвертам с адресами случайным образом. 1) Какова вероятность, что ни одно письмо не попадёт по назначению? Какова вероятность, что по назначению попадёт 2) ровно одно письмо? 3) ровно два? 4) ровно три? 5) все четыре?

Задача 23. Трамвай состоит из переднего и прицепного вагонов. В трамвай случайным образом садятся 4 человека. Найти вероятности: а) все четверо окажутся в одном вагоне; б) и в переднем вагоне, и в прицепном окажутся по два человека.

Задача 24. Саша, Маша и еще 6 человек рассаживаются на стоящих в ряд 8 стульях. а) Какова вероятность того, что Саша и Маша окажутся сидящими рядом? б) Как изменится эта вероятность, если стульев будет не 8, а n?

Задача 25. Саша, Маша и еще 6 человек рассаживаются за круглый стол. Какова вероятность того, что Саша и Маша окажутся сидящими рядом? б) Как изменится эта вероятность, если мест за столом будет не 8, а n?

Задача 26. Семь человек, в том числе три женщины, садятся в случайном порядке за круглый стол. Какова вероятность того, что никакие две женщины не окажутся сидящими рядом?

Решение

Выберем из трех женщин одну Самую Главную Женщину, и будем вести отсчет от нее. Пронумеруем места за круглым столом: от 1 по правую руку от Самой Главной Женщины до 6 по левую руку от Нее. Тогда всего вариантов размещения будет $6! = 720$. Теперь подсчитаем число благоприятных вариантов. Две другие женщины могут сидеть на местах 2 и 4, или 2 и 5, или 3 и 5. Для каждого из этих случаев есть $2!$ варианта размещения женщин и $4!$ варианта размещения мужчин. Поэтому искомая вероятность равна

$$p = \frac{3 \cdot 2! \cdot 4!}{6!} = \frac{1}{5}.$$

Можете самостоятельно убедиться в том, что выбор в качестве начала отсчета Самого Главного Мужчины приводит к тому же ответу.
Ответ: 1/5.

Задача 27. Симметричную монету побрасывают 5 раз, при этом записывают 1, если выпадет орел, и 2, если выпадет решка. Записанные цифры образуют пятизначное число. Какова вероятность того, что оно делится на 18?

Решение

Всего возможно $2^5 = 32$ варианта. Для того, чтобы число делилось на 18, оно должно оканчиваться на четную цифру, т. е. на 2, а сумма его цифр должна делиться на 9. Нетрудно убедиться, что в числе должно быть ровно четыре двойки, а единственная единица не должна быть последней. Таких чисел всего 4.
Ответ: 1/8.

Задача 28. В турнире участвуют три команды. Турнир проводится в один круг, за победу команда получает 1 очко, за поражение — 0 очков, а ничьих не бывает. Найти вероятность того, что у всех трех команд очков будет поровну. А какова будет вероятность для четырех команд?

Задача 29(*). В турнире участвуют пять команд. Турнир проводится в один круг, за победу команда получает 1 очко, за поражение — 0 очков, а ничьих не бывает. Найти вероятность того, что у всех трех команд очков будет поровну.

2. Элементы комбинаторики

Комбинаторика — наука о числе возможностей выбора.

Основное правило комбинаторики. Если выбор надо произвести дважды, причем при первом выборе есть m возможностей, а при втором — n возможностей, то общее число возможностей для двойного выбора равно mn. Аналогичная формула есть и для k-кратного выбора.

Основные формулы

Число перестановок — факториал.

$$n! = 1 \cdot 2 \cdot \dots \cdot n.$$

Принято соглашение, что $0! = 1$ (пустое множество можно упорядочить единственным способом).

Число способов, которым из n предметов можно выбрать k с учетом порядка следования предметов, называется числом размещений

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

Число способов, которым из n предметов можно выбрать k без учета порядка следования предметов, называется числом сочетаний из n по k

$$C_n^k = \frac{n!}{k!(n-k)!}.$$

Задача 30. Четырёхтомное сочинение расположено на книжной полке в случайном порядке. Чему равна вероятность того, что тома стоят в должном порядке справа налево или слева направо?

Задача 31. В меню кафе четыре вида мороженого и пять видов пирожных. Какова вероятность, что Ваня закажет Мане ее любимые "лакомку" и "эклер", если он забыл поинтересоваться ее вкусами и заказывает по одному десерту каждого вида, выбирая их случайным образои?

Задача 32. Случайным образом выбирают 4 раза цифру от 0 до 9. Какова вероятность того, что все 4 выбранные цифры окажутся разными?

Задача 33. Есть пять различных предметов, например, $1, 2, 3, 4, 5$. Сколькими способами можно из них выбрать два?

Решение

Таких способов 10:

12 23 34 45
13 24 35
14 25
15
Подсчет по приведенной выше формуле даст тот же результат: $C_5^2 = 10$.
Надо обратить внимание на то, что порядок выбора несущественен, то есть выбор 1 и 2 — то же самое, что и выбор 2 и 1.

Задача 34. Доказать формулу

$$(x+y)^n = \sum_{k=0}^{n} C_n^k x^k y^{n-k}.$$

Выражение в левой части этой формулы называется бином (двучлен). Поэтому числа сочетаний называют еще биномиальными коэффициентами.

Решение

Нетрудно понять, почему так получается. Если левую часть представить как произведение

$$\underbrace{(x+y)\cdot(x+y)\cdot\dots\cdot(x+y)}_{n \text{ раз}},$$

то произведение $x^k y^{n-k}$ получается, если из k таких скобок выбрать множитель x, а из остальных $n-k$ — множитель y. Какой коэффициент будет при $x^k y^{n-k}$? Именно такой, сколькими способами можно этот выбор сделать, то есть как раз C_n^k.

Задача 35. Доказать, что лля подсчета чисел сочетаний, кроме приведенной выше формулы, можно использовать треугольник Паскаля:

$$\begin{array}{c}
1\\
1 \quad 1\\
1 \quad 2 \quad 1\\
1 \quad 3 \quad 3 \quad 1\\
1 \quad 4 \quad 6 \quad 4 \quad 1\\
1 \quad 5 \quad 10 \quad 10 \quad 5 \quad 1\\
1 \quad 6 \quad 15 \quad 20 \quad 15 \quad 6 \quad 1\\
1 \quad 7 \quad 21 \quad 35 \quad 35 \quad 21 \quad 7 \quad 1
\end{array}$$

Этот треугольник строится по простому правилу: стороны слева и справа состоят из одних единиц, а каждое число внутри равно сумме двух чисел, стоящих над ним.

Решение

Нетрудно сообразить, что каждая строка треугольника Паскаля представляет собой коэффициенты разложения бинома $(x+y)^n$. Например, в третьей строке стоят числа 1, 3, 3, 1, что соответствует хорошо известной формуле

$$(x+y)^3 = x^3 + 3x^2y + 3xy^2 + y^3.$$

Эта строка — третья, а не четвертая, как могло бы показаться, потому что номера строк начинаются с нуля: строка из одной единицы — нулевая строка. Правило построения сразу следует из того, что

$$(x+y)^n = (x+y)^{n-1} \cdot (x+y).$$

Задача 36. Доказать формулы для чисел сочетаний.
1. $C_n^k = C_n^{n-k}$.
2. $C_n^k = C_{n-1}^{k-1} + C_{n-1}^k$ (правило построения треугольника Паскаля).
3. $\sum_{k=0}^n C_n^k = 2^n$.
Привести комбинаторные доказательства.

Решение

1. Число способов отобрать k элементов из n возможных, равно, разумеется, числу способов оставить остальные $n-k$ неотобранными.

2. Для доказательства пометим один из n предметов. Сколькими способами можно выбрать k из n предметов? Если в число выбранных входит помеченный предмет, то таких способов C_{n-1}^{k-1}, а если не входит, то C_{n-1}^k.

3. Поскольку C_n^k — не что иное, как число подмножеств из k элементов у множества из n элементов, доказываемая формула утверждает, что число всех подмножеств множества из n элементов равно 2^n. Действительно, выбирая некоторое подмножество, мы для каждого из n элементов должны сделать выбор из двух вариантов: войдет он в подмножество или нет. Поскольку выбор делается n раз, в соответствии с основным правилом комбинаторики получим, что всего подмножеств (включая пустое) будет 2^n.

Задача 37. Доказать формулы из задачи 36, исходя из формулы бинома.

Задача 38. Доказать формулу

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

Эта формула легко доказывается иа основании формулы бинома. Легкое комбинаторное доказательство получается при нечетных n. Можно ли придумать комбинаторное доказательство для четных n?

Задача 39. Доказать формулу

$$\sum_{i=k}^{N} C_i^k = C_{N+1}^{k+1}.$$

Решение

Сумма в левой части — это отрезок диагонали треугольника Паскаля. Эти числа выделены жирным шрифтом на рисунке. Правая часть также выделена жирным шрифтом и обведена в рамочку.

$$\begin{array}{ccccccccccccccc}
 & & & & & & & 1 & & & & & & & \\
 & & & & & & 1 & & 1 & & & & & & \\
 & & & & & 1 & & 2 & & \mathbf{1} & & & & & \\
 & & & & 1 & & 3 & & \mathbf{3} & & 1 & & & & \\
 & & & 1 & & 4 & & \mathbf{6} & & 4 & & 1 & & & \\
 & & 1 & & 5 & & \mathbf{10} & & 10 & & 5 & & 1 & & \\
 & 1 & & 6 & & 15 & & \boxed{\mathbf{20}} & & 15 & & 6 & & 1 & \\
1 & & 7 & & 21 & & 35 & & 35 & & 21 & & 7 & & 1
\end{array}$$

Формула теперь легко доказывается индукцией по N. Действительно, при $N = k$ обе части равны 1. Пусть формула доказана для некоторого отрезка диагонали, как на рисунке. Добавляя к диагонали следующее число (на рисунке это число 15), мы получим, что сумма стала равной сумме добавленного числа и числа, взятого в рамочку. В соответствии с правилом построения треугольника Паскаля — это число, стоящее правее и ниже добавленного (на рисунке это число 35). Что и требовалось доказать.

Эта задача показывает, что треугольник Паскаля тоже может быть использован не только для иллюстрации, но и для строгих математических доказательств.

Задача 40. Написать и доказать "зеркально-симметричную"формулу к примеру 39.

Задача 41. Игральную кость бросают 8 раз. Найти вероятности событий:
— все грани выпадут хотя бы по одному разу;
— как минимум одна какая-то грань не выпадет ни разу.

Решение

Первый способ. Общее число равновозможных исходов равно, разумеется, 6^8. Займемся подсчетом числа благоприятных исходов.

Все грани хотя бы по разу могут выпасть двумя способами: либо какие-то две грани выпадут по два раза, либо какая-то одна грань выпадет трижды. В первом случае для любых двух выбранных граней число благоприятных исходов будет равно

$$N(2 \times 2) = \frac{8!}{2!2!},$$

а выбрать эту пару граней можно $C_6^2 = 15$ способами. Во втором случае выбрать грань, выпадающую трижды, можно 6 способами, и для каждого такого выбора благоприятных исходов будет

$$N(1 \times 3) = \frac{8!}{3!}.$$

Поэтому первая из искомых вероятностей равна

$$p_1 = \frac{N(2 \times 2) \cdot 15 + N(1 \times 3) \cdot 6}{6^8} = \frac{(210 + 56) \cdot 6!}{6^8} = \frac{665}{5832} \approx 0,114.$$

Вторая, соответственно

$$p_2 = 1 - p_1 = \frac{5167}{5832} \approx 0,886.$$

Второй способ отличается только тем, что подсчет числа благоприятных исходов ведется чисто комбинаторным способом. Чтобы подсчитать число исходов, когда какие-то две грани выпадут по два раза, надо перемножить следующие числа:

C_8^4 — при каких бросаниях выпадут те грани, которые выпадут по 1 разу;

C_6^4 — какие именно грани выпадут по 1 разу;

$4!$ — сколькими способами выбранные грани можно разместить по выбранным местам;

C_4^2 — сколькими способами две грани, выпавшие дважды, можно разместить на оставшихся четырех местах.

Для подсчета числа исходов, при которых какая-то грань выпадет трижды, надо, аналогичным образом, перемножить числа:

C_8^5 — при каких бросаниях выпадут те грани, которые выпадут по 1 разу;

C_6^5 — какие именно грани выпадут по 1 разу;
5! — сколькими способами выбранные грани можно разместить по местам. На оставшихся трех местах разместится единственная выпавшая трижды грань.

Можно убедиться, что подсчеты по этому способу приводят к тому же результату.

Задача 42. Найти вероятность того, что в группе из n человек хотя бы у двух совпадают даты рождения (может быть, в разные годы). Для простоты расчетов високосными годами следует пренебречь.

Решение

Вероятность того, что у n человек не совпадут дни рождения, равна

$$p = \frac{365 \cdot (365-1) \cdot (365-2) \cdot \ldots \cdot (365-n+1)}{365^n}.$$

Дальнейшие вычисления читателю предлагается провести самостоятельно. Проще всего это сделать, использовав электронные таблицы, например Excel или Libre Office Calc.

Еще можно провести эксперимент. Какова, по мнению читателя, вероятность совпадения хотя бы одной пары дней рождения для 30 человек? для 40 человек? Сколько человек надо собрать, чтобы эта вероятность стала больше 1/2?

3. Задача о выборке

Задача 43. Задача о выборке. В урне находятся n шаров, из них k белых, а остальные $n-k$ — красные. Из урны извлекают m шаров. Какова вероятность того, что будет извлечено r белых шаров и $m-r$ красных?

Решение

Выбрать m шаров из n возможных можно C_n^m способами, причем все эти способы равновозможны. Теперь подсчитаем число благоприятных исходов. Выбрать r белых шаров из k можно C_k^r способами, а выбрать $m-r$ красных шаров из $n-k$ можно C_{n-k}^{m-r} способами. В соответствии с основным правилом комбинаторики число благоприятных исходов равно $C_k^r \cdot C_{n-k}^{m-r}$, а искомая

вероятность равна
$$p = \frac{C_k^r \cdot C_{n-k}^{m-r}}{C_n^m}.$$

Задача 44. В урне 5 белых и 3 красных шара. Из урны случайным образом выбирают 2 шара. Какова вероятность того, что оба они окажутся белыми?

Задача 45. В урне 4 белых и 5 красных шаров. Из урны случайным образом выбирают 4 шара. Какова вероятность того, что среди них будут 2 белых и 2 красных?

Задача 46. В ящике 9 деталей, из которых 4 окрашены. Найти вероятность того, что из трех взятых наудачу деталей хотя бы одна окрашена.

Задача 47. Студент знает 50 из 60 вопросов программы. Каждый экзаменационный билет содержит три вопроса. Найти вероятность того, что а) студент знает все три вопроса, содержащиеся в его экзаменационном билете; б) студент знает только два вопроса своего экзаменационного билета; в) студент знает только один вопрос своего экзаменационного билета.

Задача 48. Студент знает 20 из 25 вопросов программы. Каждый экзаменационный билет содержит три вопроса. Найти вероятность того, что а) студент знает все три вопроса, содержащиеся в его экзаменационном билете; б) студент знает только два вопроса своего экзаменационного билета; в) студент знает только один вопрос своего экзаменационного билета.

Задача 49. 30% изделий данного предприятия — это продукция высшего сорта. Некто приобрел 6 изделий, изготовленных на этом предприятии. Чему равна вероятность того, что ровно 4 из них высшего сорта? Чему равна вероятность того, что не менее 4 из них высшего сорта?

Задача 50. В урне 4 занумерованных шара. Случайным образом выбирают два из них. Какова вероятность того, что будут выбраны шары № 1 и № 2?

Задача 51. Покерная колода состоит из 54 карт: 4 масти по 13 карт и два джокера. Вам при сдаче достались три туза, тройка пик и семерка треф. Какова вероятность того, что после обмена тройки и семерки у вас на руках окажется каре (4 туза)? По правилам покера джокер может быть объявлен любой картой. Как изменится вычисленная вероятность, если вам станет известно, что один из джокеров на руках у другого?

Задача 52. Покерная колода состоит из 54 карт: 4 масти по 13 карт и два джокера. Вам при сдаче достались 7, 8, 9 и 10 пик и дама треф. Какова вероятность того, что после обмена дамы у вас на руках окажется стрит (5 карт подряд)? стрит одной масти? По правилам покера джокер может быть объявлен любой картой. Как изменится вычисленная вероятность, если вам станет известно, что один из джокеров на руках у другого?

Задача 53. У игрока в покер на руках 2 туза. Какова вероятность, что после обмена трех других карт у него будет комбинация "4 туза"? По правилам покера тузом можно объявить джокера.

Задача 54. При игре в преферанс 32 карты случайным образом сдаются трем игрокам, каждый из которых получает по 10 карт. Оставшиеся 2 карты — прикуп. Допустим, что вы получили этот прикуп и обнаружили, что у вас 4 пики (всего в колоде 8 пик). Какова вероятность, что все остальные 4 пики будут у одного из ваших партнеров?

Задача 55. При игре в преферанс 32 карты случайным образом сдаются трем игрокам, каждый из которых получает по 10 карт. Оставшиеся 2 карты — прикуп. Допустим, что вы получили этот прикуп и обнаружили, что у вас 5 пик (всего в колоде 8 пик). Какова вероятность, что все остальные 3 пики будут у одного из ваших партнеров?

Задача 56. При игре в преферанс 32 карты случайным образом сдаются трем игрокам, каждый из которых получает по 10 карт. Оставшиеся 2 карты — прикуп. Какова вероятность, что в прикупе окажутся два туза? Как изменится эта вероятность, если в ваших 10 картах нет ни одного туза?

Задача 57. У одного из игроков в преферанс на руках 4 бубны. Какова вероятность, что еще ровно одна бубна в прикупе?

Задача 58. При тестировании студент отвечает на 20 вопросов. Эти вопросы разбиты на 5 дидактических единиц, по 4 вопроса в каждой. Тест считается пройденным, если студент ответил правильно не менее чем на 2 вопроса в каждой дидактической единице. Какова вероятность того, что студент прошел тест, если он ответил правильно на 17 вопросов?

Решение

Переформулируем условие задачи в виде задачи на урновую схему, часто встречающуюся в теории вероятностей. Получим такую задачу: в урне находится 20 шаров 5 различных цветов, по 4 каждого цвета. Случайным образом вынимают 3 шара (без возвращения). Какова вероятность того, что они не будут одного цвета?

Проще, разумеется, найти вероятность дополнительного события — что они будут одного цвета. Для каждого из 5 цветов это — вариант классической задачи о выборке, и вероятность этого события равна

$$\frac{C_4^3}{C_{20}^3} = \frac{4 \cdot 6}{18 \cdot 19 \cdot 20} = \frac{1}{57 \cdot 5}.$$

Для 5 цветов вероятность в 5 раз больше, поскольку соответствующие события не пересекаются.
Ответ: 56/57.

Задача 59. В чемпионате участвуют 16 команд, в том числе 4 московских. Все команды жеребьевкой разбивают на 4 группы по 4 команды в каждой. Найти вероятность того, что в каждой группе окажется ровно по одной московской команде.

Задача 60. Рыбаки выловили в озере 100 карасей, пометили их и выпустили обратно в озеро. Затем выловили 100 карасей и обнаружили среди них 10 помеченных. Найти вероятность того, что среди 100 выловленных будет 10 помеченных, если карасей в озере n. Определить, при каком значении n эта вероятность максимальна.

Решение

По формуле задачи о выборке искомая вероятность p_n равна

$$p_n = \frac{C_{n-100}^{90} C_{100}^{10}}{C_n^{100}} = \frac{((n-100)!)^2}{n!(n-190)!}.$$

Для нахождения максимума этого выражения найдем отношение p_{n+1}/p_n. Если это выражение больше 1, то p_n возрастает, а если меньше — убывает. Находим

$$\frac{p_{n+1}}{p_n} = \frac{(n-99)^2}{(n+1)(n-189)} = \frac{n^2 - 198n + 9801}{n^2 - 188n - 189}.$$

Дальше выясним, когда это выражение не превосходит 1. Сразу заметим, что $n > 189$, поэтому знаменатель положителен и при решении неравенства на него можно домножать. Получим

$$\frac{n^2 - 198n + 9801}{n^2 - 188n - 189} \leq 1;$$

$$n^2 - 198n + 9801 \leq n^2 - 188n - 189; \quad 10n \leq 9990; \quad n \leq 999.$$

Следовательно, отношение p_{n+1}/p_n возрастает до 1, пока $n < 999$, при $n = 999$ оно равно 1, а затем начинает убывать. Это означает, что вероятность p_n максимальна при двух значениях n: при $n = 999$ и при $n = 1000$.

Такой метод получения оценки (числа карасей в озере) часто используется в математической статистике. Он называется *метод максимального правдоподобия.*

4. Модели случайного выбора

Многие вероятностные задачи сводятся к задачам о случайном выборе: есть n предметов, сколькими способами их них можно выбрать k? Оказывается, выбор можно проводить по-разному. Поэтому мы рассмотрим различные модели случайного выбора.

С двумя моделями случайного выбора мы уже познакомились. Это
а). Упорядоченный выбор без возвращения. Число способов выбрать таким образом k предметов из n, как мы установили, равно A_n^k.
б). Неупорядоченный выбор без возвращения. Число способов выбрать таким образом k предметов из n, как мы установили, равно C_n^k.
в). Упорядоченный выбор с возвращением. Это новая модель случайного выбора. Можно представлять себе, что в урне лежит n различных предметов, и мы k раз извлекаем случайным образом предмет из урны, отмечаем, какой это был предмет, после чего возвращаем его в урну. Поскольку каждый раз мы выбираем один из n предметов, в соответствии с основным правилом комбинаторики число исходов равно n^k.
г). Неупорядоченный выбор с возвращением. Здесь мы не обращаем внимания на то, в каком порядке выбирались предметы, а следим только за тем, сколько раз был выбран каждый предмет. Вывод формулы для числа способов такого выбора является образцом применения комбинаторного мышления.

Задача 61. Вывести формулу для числа способов неупорядоченного выбора с возвращением.

Решение

Для каждого способа выбора составим строку из n клеточек, где крестиком будем отмечать, что соответствующий предмет был выбран, а клеточки будем разделять вертикальными черточками. Например, если мы выбираем 4 раза из 5 предметов, то строка
xx| |x| |x
означает, что первый предмет был выбран 2 раза, 3-й и 5-й — по одному разу, а 2-й и 4-й не были выбраны ни разу. Мы не потеряем в информативности, если сотрем все пробелы в этой строке:
xx||x||x
Обратно, любая строка из 4 крестиков и 4 вертикальных черточек будет определять некоторый вариант выбора. Например, строка
||||xxxx
соответствует тому варианту выбора, когда все 4 раза был выбран 5-й предмет.

Таким образом, число вариантов неупорядоченного выбора с возвращением k раз из n возможных предметов находится во взаимно однозначном соответствии с числом строк из k крестиков и $n-1$ палочек. Ясно, что таких строк ровно столько, сколькими способами можно из $n+k-1$ предметов выбрать k (крестиков, а остальные будут палочками), то есть C_{n+k-1}^{k}.

Задача 62. Пусть мы выбираем дважды из 4 предметов. Сколько будет способов такого выбора при различных моделях выбора?

Решение

а). Упорядоченный выбор без возвращения. $A_4^2 = 12$.

(1 2)	(1 3)	(1 4)	
(2 1)		(2 3)	(2 4)
(3 1)	(3 2)		(3 4)
(4 1)	(4 2)	(4 3)	

б). Неупорядоченный выбор без возвращения. $C_4^2 = 6$.

(1 2) (1 3) (1 4)
(2 3) (2 4)
(3 4)

в). Упорядоченный выбор с возвращением. $4^2 = 16$.
(1 1) (1 2) (1 3) (1 4)

(2 1)   (2 2)   (2 3)   (2 4)
(3 1)   (3 2)   (3 3)   (3 4)
(4 1)   (4 2)   (4 3)   (4 4)
г). Неупорядоченный выбор с возвращением. $C_5^2 = 10$.
(1 1)   (1 2)   (1 3)   (1 4)
(2 2)   (2 3)   (2 4)
(3 3)   (3 4)
(4 4)

Задача 63. Из n предметов выбирают k с возвращением. Найти вероятность того, что все выбранные предметы различны. При взгляде на предыдущий пример ясно, что есть два возможных решения:
1) если рассматривать упорядоченный выбор, то

$$p = \frac{A_n^k}{n^k};$$

2) если выбор неупорядоченный, то

$$p = \frac{C_n^k}{C_{n+k-1}^k}.$$

Эти формулы, однако, дают разные ответы. Какая из них верна?

Решение

При реальном выборе по одному предмету выбирают k раз, то есть выбор упорядоченный. Например, если выбор делается дважды, то выбрать 1 и 2 предметы можно двумя способами: сначала 1, затем 2, или наоборот. А дважды выбрать первый предмет можно только одним способом. То есть вероятность выбрать 1 и 2 вдвое больше, чем выбрать два раза 1. При неупорядоченном же выборе эти вероятности равны. Мы уже встречались с этим в задаче 4. Таким образом, верна первая формула.

Этот пример может создать представление, что неупорядоченный выбор в реальной жизни не встречается. Его действительно не так просто реализовать в урновой схеме, но в других случаях такой выбор вполне имеет право на существование. Например, в квантовой механике все электроны считаются одинаковыми, неразличимыми. И тогда задача о заполнении электронных уровней приводит именно к неупорядоченному выбору: важно только то, сколько электронов на каких уровнях.

Еще один пример — в следующей задаче.

Задача 64. Сколько решений имеет уравнение

$$x_1 + x_2 + x_3 + x_4 + x_5 = 24,$$

если x_1, x_2, x_3, x_4, x_5 —
a) неотрицательные целые числа?
b) положительные целые числа?

Задача 65. Сколькими способами можно разложить 5 одинаковых шаров по трем коробкам?

Задача 66. Записывают строку из нулей и единиц, причем всего символов в строке должно быть n, в том числе k нулей. Сколько всего таких строк?

Задача 67. Есто шары k различных цветов. Число шаров i-го цвета равно m_i. Всего шаров m, при этом, разумеется, $m_1 + \ldots + m_k = m$. Шары разных цветов различаются, а шары одного цвета неотличимы. Сколько всего существует способов расположить эти шары в ряд?

Решение

Если бы все шарыи был различны, всего было бы $m!$ перестановок. Но, поскольку шары одного, например, i-го, цвета считаются одинаковыми любая из $m_i!$ перестановок шаров этого цвета приведет к тому же расположению. Поэтому всего таких расположений будет

$$\frac{m!}{m_1! \cdot \cdots \cdot m_k!}.$$

Задача 68. Ребенок играет с 10 табличками, на которых написаны:
3 буквы “А”;
2 буквы “М”;
2 буквы “Т”;
1 буква “Е”;
1 буква “И”;
1 буква “К”.
Какова вероятноость того, что, выложив их в ряд, ребенок получит слово “МАТЕМАТИКА”?

Задача 69. k предметов раскладывают по n коробкам ($k < n$). Все предметы считаются эквивалентными, то есть два размещения предметов по коробкам различны, если различно число предметов в коробках. Найти число таких размещений.

Задача 70. Как и в предыдущей задаче, k предметов раскладывают по n коробкам ($k < n$), но теперь в каждую коробку можно поместить не более одного предмета. Найти число таких размещений.

5. Геометрические вероятности

Если число равновозможных исходов бесконечно, но занимает некоторую область, то используется формула геометрической вероятности

$$P = \frac{\text{площадь множества благоприятных исходов}}{\text{площадь множества всех равновозможных исходов}}.$$

Вероятность, таким образом — это мера, аналог площади.

Задача 71. (Задача о встрече). Два студента условились встретиться в определенном месте между 12 и 13 часами дня. Пришедший первым ждет второго в течение 15 минут, после чего уходит. Найти вероятность того, что встреча состоится, если каждый студент наудачу выбирает момент своего прихода в промежутке от 12 до 13 часов.

Решение

Обозначим время прихода первого студента через x, а второго — через y. Будем считать, что $0 \leqslant x \leqslant 1$, $0 \leqslant y \leqslant 1$, то есть время прихода измеряется в часах. Например, если первый студент пришел в 12-45, то $x = 3/4$.

Здесь все возможные исходы занимают единичный квадрат на плоскости, их бесконечно много. Используем формулу геометрической вероятности.

Студенты встретятся, если выполнено неравенство $|y - x| < 1/4$. Раскрывая модуль, получим двойное неравенство

$$x - \frac{1}{4} < y < x + \frac{1}{4}.$$

Соответствующие множества изображены на рисунке. При этом множество благоприятных исходов не заштриховано — оставлено белым. А заштрихованные области соответствуют тем исходам, когда студенты не встретятся.

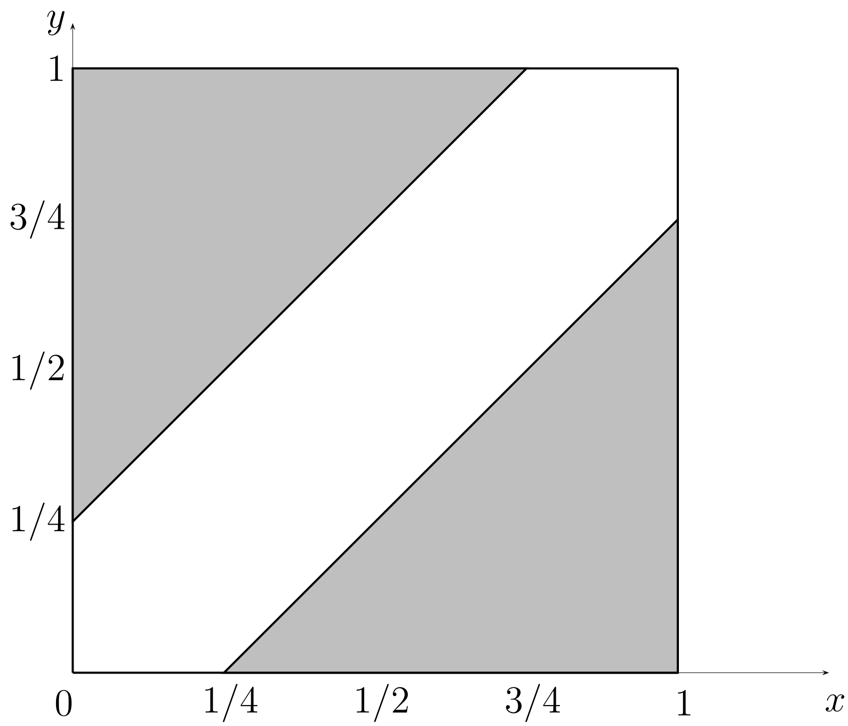


Рисунок 1. Задача о встрече.

В нашем примере площадь множества благоприятных исходов равна 7/16, а площадь множества всех возможных исходов равна 1. Поэтому вероятность равна 7/16.

Во всех задачах со случайным выбором точек подразумевается, что вероятность попадания точки в определенную область пропорциональна площади этой области и не зависти от положения этой области.

Задача 72. На стержне длины 1 случайным образом выбирают две точки, а затем стержень ломают в этих точках. Найти вероятность того, что из получившихся обломков можно составить треугольник.

Задача 73. На стержне длины 1 случайным образом выбирают точку, а затем стержень ломают в этой точке. Затем больший из получившихся обломков также ломают в случайно выбранной точке. Найти вероятность того, что из получившихся обломков можно составить треугольник.

Задача 74. На отрезке OA случайным образом выбирают точку B. Затем на отрезке BA случайным образом выбирают точку C. Найти вероятность того, что $BC < OB$.

Задача 75. Случайным образом выбраны два числа p и q не отрезке $[0, 1]$. Найти вероятность того, что квадратное урравнение $x^2 + px + q = 0$ имеет два действительных корня. Найти вероятность того, что это квадратное уравнение имеет ровно один действительный корень.

Задача 76. Случайным образом выбрирают два числа не отрезке $[0, 1]$. Найти вероятность того, что их произведение не превосходит $1/4$.

Задача 77. Случайным образом выбраны три числа не отрезке $[0, 1]$. Какова вероятность того, что их сумма также не превосходит 1?

Задача 78(*). На отрезке OA случайным образом выбирают три точки B, C, D. Найти вероятность того, что из отрезков OB, OC и OD можно сложить треугольник.

Задача 79. На поверхности шара случайным образом выбирают две точки и соединяют их меньшей дугой большого круга. Найти вероятность того, что длина этой дуги меньше одной шестой длины экватора шара.

Задача 80. На глобусе выбирают случайную точку. Какова вероятность того, что ее широта между 30° и 60°?

Задача 81. (Задача Бюффона). На плоскости проведены параллельные линии на расстоянии d друг от друга. На эту плоскость бросают иглу длиной l. Какова вероятность, что она пересечет какую-то из линий?

Задача 82. (Парадокс Бертрана). Внутри круга случайным образом выбирают некоторую хорду. Какова вероятность того, что она длиннее стороны правильного треугольника, вписанного в эту окружность?

Решение

Эта задача наглядно иллюстрирует необходимость точного выбора пространства элементарных исходов, поскольку разные выборы приводят к разным ответам.

1). Выбирается случайная точка внутри круга и объявляется серединой хорды. Тогда $p = 1/4$.

2). Некоторая точка на окружности объявляется началом отсчета, а другая выбирается на окружности случайным образом. Хорда — отрезок, соединяющий эти точки. Тогда $p = 1/3$.

3). Выбирается некоторый диаметр и на нем случайно выбирается точка, которая будет серединой хорды. Тогда $p = 1/2$.

Читателю предлагается самому восстановить подробности решения. Если не получится, то подробный разбор этой задачи приведен в [12].

Задача 83. На окружности случайным образом выбирают три точки. Найти вероятность того, что треугольник с вершинами в этих точках — остроугольный.

Решение

Примем точку A за начало отсчета и проведем диаметр AA_1, делящий нашу окружность на две полуокружности. Допустим, что точки B и C попали в одну из этих полуокружностей. Тогда треугольник — тупоугольный. Вероятность же того, что они попадут в одну полуокружность, равна $1/2$ — по аналогии с подбрасыванием двух монет: те тоже упадут одинаково с вероятностью $1/2$.

Допустим теперь, что B и C попали в разные полуокружности. Примем, что длина полуокружности равна 1, и обозначим длины дуг A_1B и A_1C через x и y соответственно. Тогда углы B и C — острые, а угол A будет острым, если выполняется условие $x + y < 1$. Нарисовав единичный квадрат $0 < x, y < 1$ и в нем треугольник $x + y < 1$, легко убедиться в том, что площадь этого треугольника равна $1/2$. Таким образом, вероятность того, что треугольник ABC остроугольный при условии, что B и C попали в разные полуокружности, равна $1/2$. Искомая же вероятность равна $1/4$.

Ответ: $1/4$.

Задача 84. На окружности наудачу ставятся три точки — A, B и C. Найти вероятность того, что в треугольнике ABC все углы меньше K°.

Решение

Пусть точка A — начало отсчета на окружности. Выберем такое направление обхода этой окружности, чтобы сначала проходилась точка B, а затем C. Будем считать, что длина окружности равна 180, то есть будем измерять дуги величиной вписанных углов.

Обозначим длину дуги AB через x, а длину дуги AC — через y. Тогда пространство элементарных исходов представляет собой прямоугольный треугольник, заданный неравенствами

$$0 < y < 180,\ 0 < x < y,$$

то есть треугольник с вершинами (0, 0), (0, 180), (180, 180). Площадь этого треугольника равна

$$S_{\text{общ}} = \frac{180^2}{2}.$$

Область благоприятных исходов будет описываться системой неравенств:

$$\begin{cases} x < K \\ y - x < K \\ 180 - y < K \end{cases}$$

Эту систему удобнее представить в виде

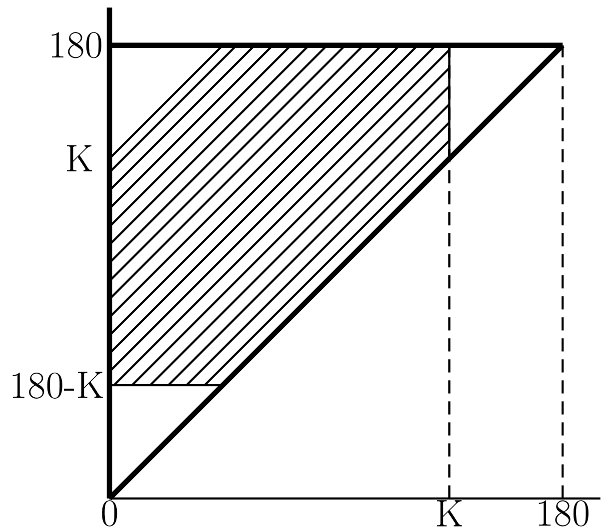
$$\begin{cases} x < K \\ y < x + K \\ y > 180 - K \end{cases}$$

Сначала рассмотрим случай $K > 90$. Тогда множество благоприятных исходов выглядит так, как показано на рисунке. Чтобы найти его площадь, проще всего вычесть из $S_{\text{общ}}$ площади трех маленьких незакрашенных треугольников. Каждая из этих площадей, как легко убедиться, равна

$$S_{\text{мал}} = \frac{(180 - K)^2}{2}.$$

Таким образом, площадь множества благоприятных исходов равна

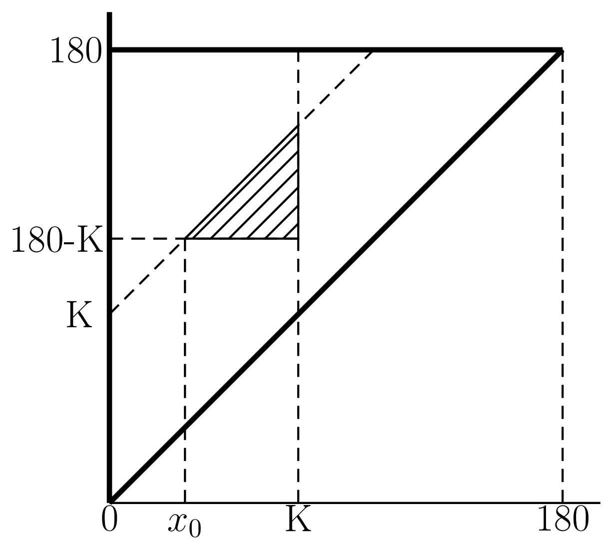
$$S_{\text{благ}} = S_{\text{общ}} - 3S_{\text{мал}} = \frac{1}{2}(180^2 - 3(180 - K)^2).$$



Поэтому искомая вероятность равна

$$P = \frac{S_{\text{благ}}}{S_{\text{общ}}} = \frac{K}{30} - \frac{3K^2}{180^2} - 2.$$

В случае $K < 90$ множество благоприятных исходов будет иметь другой вид.



Чтобы найти площадь полученного треугольника, вычислим абсциссу вершины острого угла — она обозначена на рисунке x_0. Поскольку эта вершина принадлежит прямой $y = x + K$, получаем уравнение $x_0 + K = 180 - K$, откуда $x_0 = 180 - 2K$. Поэтому катет треугольника равен $K - x_0 = 3K - 180$, а его площадь равна

$$S_{\text{благ}} = \frac{(3K-180)^2}{2}.$$

Поэтому искомая вероятность равна

$$P = \frac{S_{\text{благ}}}{S_{\text{общ}}} = \frac{K^2}{60^2} - \frac{K}{30} + 1.$$

Для проверки можно подставить $K = 90$ — это задача 83. Для обеих формул получаем значение $P = 1/4$, совпадающее с ответом той задачи.

6. Алгебра событий и аксиоматическое определение вероятности

Все возможные исходы вероятностного испытания составляют пространство элементарных исходов Ω.

Событие — подмножество пространства элементарных исходов. Для бесконечных множеств Ω событием может быть не всякое подмножество, а только такое, которое имеет площадь.

Для событий определены операции взятия дополнения $\overline{A} = \Omega \backslash A$, объединения $A \cap B$ и пересечения $A \cup B$.

Множество событий вместе с определенными на нем операциями образует *алгебру событий*.

сновные формулы алгебры событий.
Двойное отрицание: $\overline{(\overline{A})} = A$.
Коммутативность: $A \cap B = B \cap A$; $A \cup B = B \cup A$.
Ассоциативность:
$(A \cap B) \cap C = A \cap (B \cap C)$;
$(A \cup B) \cup C = A \cup (B \cup C)$.
Дистрибутивность:
$(A \cap B) \cup C = (A \cup C) \cap (B \cup C)$;
$(A \cup B) \cap C = (A \cap C) \cup (B \cap C)$.
Формулы Моргана: $\overline{(A \cap B)} = \overline{A} \cup \overline{B}$; $\overline{(A \cup B)} = \overline{A} \cap \overline{B}$.

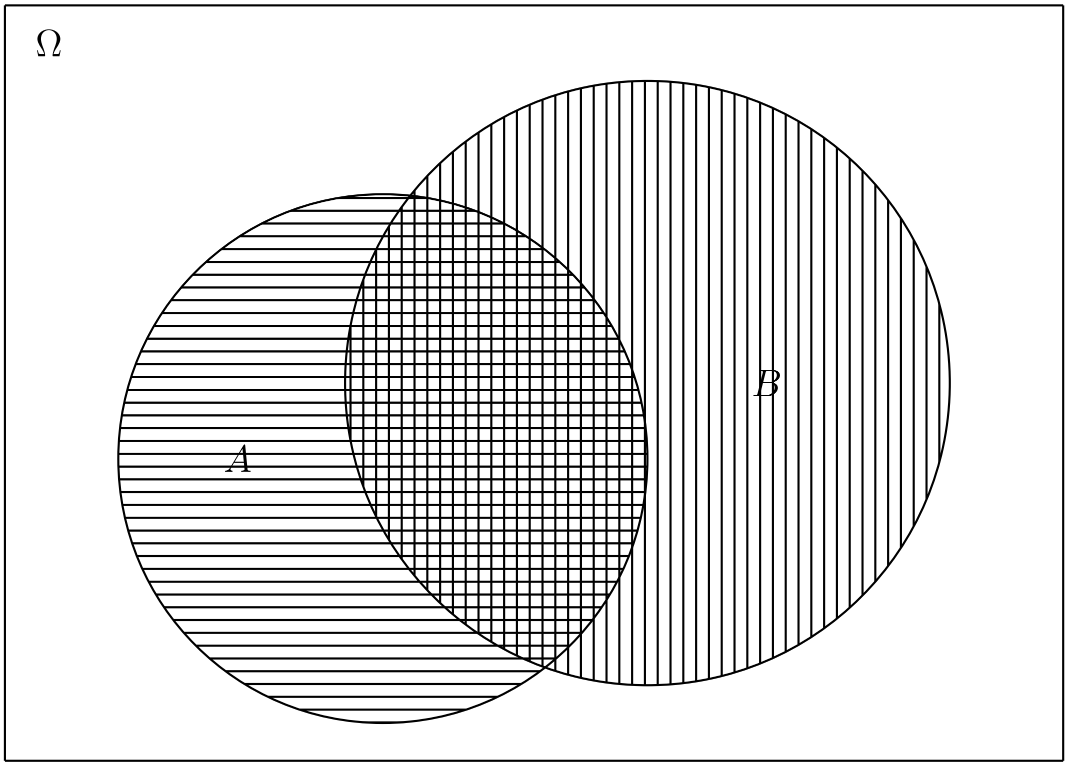


Рисунок 2. Диаграмма Венна.

Для иллюстрации (а также для доказательства) приведенных формул часто используются специальные рисунки, они называются *диаграммы Венна* (или диаграммы Эйлера — Венна).

Аксиоматическое определение вероятности было введено советским математиком, академиком А. Н. Колмогоровым в 30-е годы XX века. Приводимая ниже аксиоматика А. Н. Колмогорова лежит в основе современной теории вероятностей.

Сигма-алгеброй $\mathbb{S}$ называется такое множество подмножеств Ω, которое удовлетворяет следующим аксиомам:
1) для всякого $A \in \mathbb{S}$ его дополнение также принадлежит $\mathbb{S}$;
2) для всякого конечного или счетного набора $A_1, \ldots, A_n, \ldots$ множеств из $\mathbb{S}$ их объединение также принадлежит $\mathbb{S}$.

Вероятность (вероятностная мера) — функция, определенная для каждого из подмножества, принадлежащего сигма-алгебре и удовлетворяющая следующим аксиомам:
1) $P(A) \geqslant 0$;

2) $P(A \cup B) = P(A) + P(B)$ если $AB = \emptyset$;
3) $P(\Omega) = 1$.

Формулы вероятности суммы и произведения событий
$P(A \cup B) = P(A) + P(B) - P(AB)$;
$P(AB) = P(A) + P(B) - P(A \cup B)$.
Вероятность противоположного события.
$P((\overline{A})) = 1 - P(A)$.

Здесь использована часто встречающаяся сокращенная запись $A \cap B = AB$.

События A и B называются независимыми, если их вероятности ненулевые и $P(AB) = P(A) \cdot P(B)$.

Задача 85. Известный каждому человеку, хотя бы в интуитивной форме, *"закон всемирного свинства"* гласит: "Если для того, чтобы произошло какое-либо полезное или приятное событие, требуется выполнение двух условий, то всегда будет выполнено одно и только одно из этих условий". Выведите следствие из этого закона: если требуется выполнение n условий, то сколько условий будет выполнено?

Задача 85. Доказать, что для конечного или счетного набора $A_1, \ldots, A_n, \ldots$ множеств из сигма-алгебры $\mathbb{S}$ их пересечение также принадлежит $\mathbb{S}$.

Задача 86. Эмпирической относительной частотой события называется предел

$$p = \lim_{n \to \infty} \frac{k}{n},$$

где k — число опытов, в которых произошло событие, n — общее число опытов (при условии, что такой предел сущечтвует).
Доказать, что эмпирическая относительная частота удовлетворяет аксиоматическому определению вероятности, т. е. является вероятностной мерой.

Задача 87. Пусть элементы соединены в цепь так, как показано на рисунке 3. Цепь считается работающей, если от левого к правому концу может идти ток. Допустим, что надежность, то есть вероятность безотказной работы для элемента A равна 0,7, а для элемента B — 0,8, причем элементы выходят из строя независимо друг от друга. Какова вероятность безотказной работы цепи?

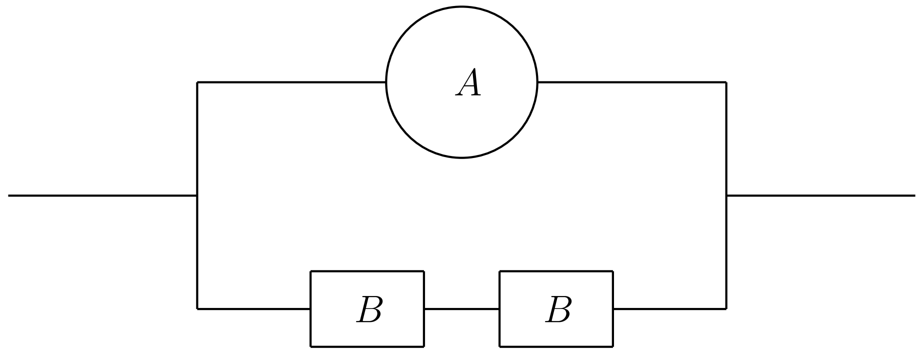


Рисунок 3. Схема цепи.

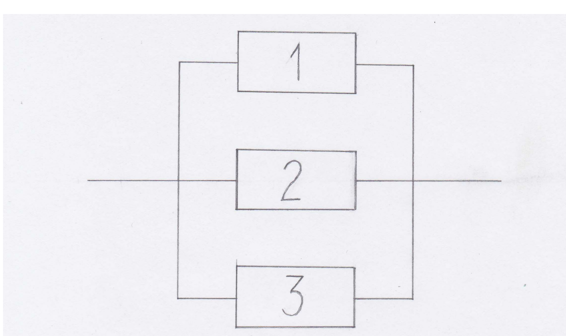
Решение

Какова вероятность выхода из строя нижнего фрагмента цепи? Для его выхода из строя нужно, чтобы отказал хотя бы один из элементов B. Поэтому вероятность безотказной работы этого фрагмента будет равна $0,8 \cdot 0,8 = 0,64$, а вероятность отказа — $0,36$.

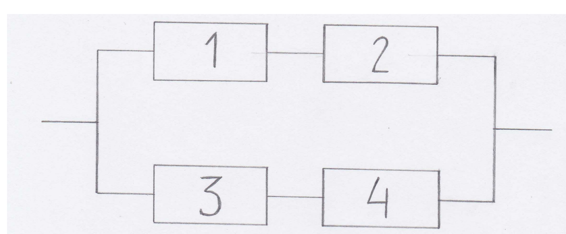
Для верхнего фрагмента вероятность отказа равна $0,3$. Чтобы цепь отказала, необходимо, чтобы отказали и верхний, и нижний фрагменты. Вероятность этого события равна $0,36 \cdot 0,3 = 0,108$. Следовательно, вероятность безотказной работы цепи равна $1 - 0,108 = 0,892$.

В задачах 88 — 93 приведены схемы соединения элементов, образующих цепь. Как и в предыдущей задавче, цепь считается работающей, если от левого к правому концу может идти ток. Допустим, что надежность, то есть вероятность безотказной работы для элемента с номером i равна p_i, причем элементы выходят из строя независимо друг от друга. Какова вероятность безотказной работы цепи?

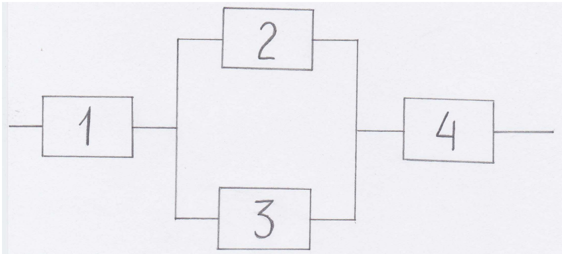
Задача 88.



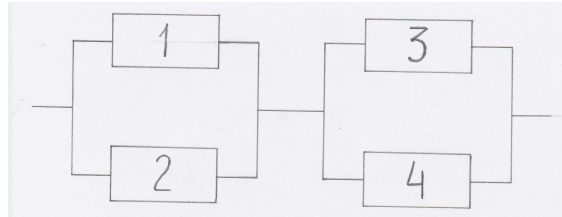
Задача 89.



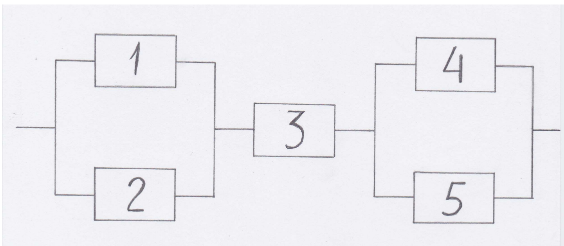
Задача 90.



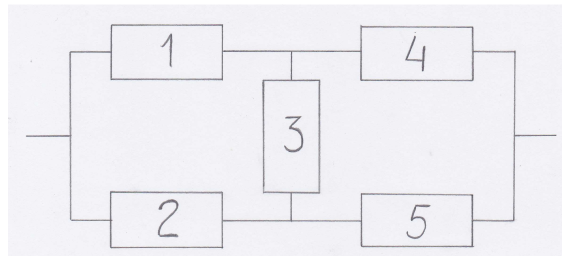
Задача 91.



Задача 92.



Задача 93.



Решение

Здесь будет изложен один из способов решения довольно трудной последней задачи 93. Этот способ достаточно общий и применимый ко многим задачам, но довольно трудоемкий. Еще один способ будет изложен в следующем разделе. Третий способ, с использованием формулы полной вероятности, будет изложен в соответствующем разделе.

Пусть событие A_i состоит в том, что i-й элемент работает, соответственно $\bar{A}_i$ — что этот элемент отказал. Тогда все пространство элементарных исходов разбивается на непересекающиеся подмножества вида $A_{00101} = \bar{A}_1\bar{A}_2A_3\bar{A}_4A_5$, иногда их называют *атомами*. Для приведенного в качестве примера атома работают только 3-й и 5-й элементы, вся цепь не работает. Для случая, когда в цепи 5 элементов, будет 32 атома. В общем случае, когда в цепи n элементов, атомов будет 2^n. Для каждого атома нетрудно определить его вероятность, например, для указанного атома

$$P(A_{00101}) = (1-p_1)(1-p_2)p_3(1-p_4)p_5.$$

Составив полный список атомов, при которых цепь работает, и сложив их вероятности, получим ответ.

Например, для изображенной на рисунке цепи полный списик квантов, когда она раьотает, выглядит так.

01001	$(1-p_1)p_2(1-p_3)(1-p_4)p_5$
01011	$(1-p_1)p_2(1-p_3)p_4p_5$
01101	$(1-p_1)p_2p_3(1-p_4)p_5$
01110	$(1-p_1)p_2p_3p_4(1-p_5)$
01111	$(1-p_1)p_2p_3p_4p_5$
10010	$p_1(1-p_2)(1-p_3)p_4(1-p_5)$
10011	$p_1(1-p_2)(1-p_3)p_4p_5$
10101	$p_1(1-p_2)p_3(1-p_4)p_5$
10110	$p_1(1-p_2)p_3p_4(1-p_5)$
10111	$p_1(1-p_2)p_3p_4p_5$
11001	$p_1p_2(1-p_3)(1-p_4)p_5$
11010	$p_1p_2(1-p_3)p_4(1-p_5)$
11011	$p_1p_2(1-p_3)p_4p_5$
11101	$p_1p_2p_3(1-p_4)p_5$
11110	$p_1p_2p_3p_4(1-p_5)$
11111	$p_1p_2p_3p_4p_5$

В левом столбце закодировано, какие элементы работают. Например, запись 10010 означает, что работают только 1-й и 4-й элементы. В правом столбце указана вероятность соответствующего кванта. Вероятность работы схемы равна сумме второго столбца.

7. Принцип включения — исключения

Задача 94. Написать формулу, выражающую вероятность объединения n событий $A_1, \ldots, A_n$ через вероятности этих событий и вероятности их пересечений. Для простоты рассмотреть сначала случай $n = 3$.

Решение

$$p(A_1 \cup A_2 \cup \ldots \cup A_n) = \sum_i p(A_i) - \sum_{i,j} p(A_i \cap A_j) + \sum_{i,j,k} p(A_i \cap A_j \cap A_k) - \ldots$$

Сначала суммируются вероятности событий, затем вычитаются вероятности попарных пересечений, затем прибавляются вероятности тройных пересечений, и так далее. Эта формула известна под названием "принцип включения — исключения" (или "формула включения — исключения").

Задача 95. Игральную кость бросают 8 раз. Найти вероятности событий:
— все грани выпадут хотя бы по одному разу;
— как минимум одна какая-то грань не выпадет ни разу.

Эта задача уже рассматривалась под номером 41. Третий способ решения этой задачи основан на "принципе включения — исключения".

Теперь мы будем подсчитывать вероятность того, что какая-то грань не выпала ни разу. Обозначим A_i событие, что i-я грань не выпала ни разу. Тогда нам нужно вычислить вероятность объединения всех A_i при $i = 1, ...6$.

В соответствии с принципом включения — исключения эта вероятность равна

$$p_2 = \sum_i p(A_i) - \sum_{i,j} p(A_i \cap A_j) + \sum_{i,j,k} p(A_i \cap A_j \cap A_k) - \dots$$

Вероятность A_i равна $5^8/6^8$, а всего таких A_i — шесть. Событие $A_i \cap A_j$ означает, что ни разу не выпадет ни i-я, ни j-я грань, поэтому его вероятность равна $4^8/6^8$. А всего возможных двукратных пересечений будет $C_6^2 = 15$. Аналогичным образом, следует учесть трехкратные, четырехкратные и пятикратные пересечения, а шестикратное пересечение уже будет невозможным событием. Таким образом

$$p_2 = \frac{6 \cdot 5^8 - 15 \cdot 4^8 + 20 \cdot 3^8 - 15 \cdot 2^8 + 6 \cdot 1^8}{6^8}.$$

Подсчет, хотя и труднее, чем для первых двух способов, но приводит к тому же результату:

$$p_2 = \frac{5167}{5832}.$$

Задача 96. Из колоды в 52 карты случайным образом выбирают 6 карт без возвращения. Какова вероятность того, что среди выбранных будут присутствовать все 4 масти?

Задача 97. n студентов пришли на семинар по борьбе с забывчивостью. После семинара никто из них не смог вспомнить, в какой шляпе он пришел, и поэтому каждый из них выбрал шляпу с вешалки случайным образом. Найти вероятность p_n того, что никто из них не выбрал свою шляпу. Найти $\lim\limits_{n\to\infty} p_n$.

Задача 98.. Из множества решений уравнения

$$x_1 + x_2 + x_3 + x_4 + x_5 = 28$$

с целыми неотрицательными x_i случайным образом выбирают одно. Найти вероятность того, что в этом выбранном решении все x_i не больше 8.

Решение

Сначала вычислим, сколько всего решений имеет это уравнение. Каждому решению,например $x_1 = 9$, $x_2 = 5$, $x_3 = 1$, $x_4 = 10$, $x_5 = 3$ соответствует строка из 28 крестиков и 4 вертикальных черточек. Для приведенного решения это строка
xxxxxxxxx|xxxxx|x|xxxxxxxxxx|xxx
Все решения этого уравнения в неотрицательных целых числах находятся во взаимно однозначном соответствии с такими строками. Следовательно, число решений равно числу способов выбрать из 32 предметов 28 крестиков (а остальные будут вертикальными черточками), то есть $C_{32}^{28} = 35960$.

Далее из этого числа следует вычесть число решений, в которых хоть одна переменная больше 8. Обозначим A_i — множество таких решений, где $x_i > 8$, а $N(A_i)$ — мощность этого множества, то есть число элементов в нем . В соответствии с принципом включения — исключения

$$N(A_1 \cup A_2 \cup \ldots \cup A_5) = \sum_i N(A_i) - \sum_{i,j} N(A_i \cap A_j) + \sum_{i,j,k} N(A_i \cap A_j \cap A_k) - \ldots$$

Сначала суммируются мощности множеств A_i, затем вычитаются мощности попарных пересечений, затем прибавляются мощности тройных пересечений, и так далее.

Вычислим мощность A_1. Сделаем замену $y_1 = x_1 - 9$, получим уравнение

$$y_1 + x_2 + x_3 + x_4 + x_5 = 19.$$

Число его решений находится аналогично тому, как это было сделано для исходного уравнения, и равно $C_{23}^{19} = 8855$. Мощности всех A_i такие же для всех i, и таких множеств $C_5^1 = 5$.

Теперь вычислим мощности попарных пересечений. Для множества $A_1 \cap A_2$ сделаем аналогичную замену: $y_1 = x_1 - 9$, $y_2 = x_2 - 9$ и получим уравнение

$$y_1 + y_2 + x_3 + x_4 + x_5 = 10.$$

Число его решений равно $C_{14}^{10} = 1001$, а число попарных пересечений равно $C_5^2 = 10$.

Для тройных пересечений получим уравнение

$$y_1 + y_2 + y_3 + x_4 + x_5 = 1,$$

имеющее $C_5^1 = 5$ решений, а число тройных пересечений равно $C_5^3 = 10$.

Четырехкратных пересечений не будет, поскольку исходное уравнение не имеет решений, где 4 неизвестных больше 8. Поэтому число решений, в которых хоть одна переменная больше 8, равно

$$C_{23}^{19} \cdot C_5^1 - C_{14}^{10} \cdot C_5^2 + C_5^1 \cdot C_5^3 = 34315,$$

а решении, где все x_i не больше 8, равно $35960 - 34315 = 1645$, Следовательно, искомая вероятность равна

$$p = \frac{1645}{35960} \approx 0,0457.$$

Задача 99. Шестизначный номер билета считается счастливым, если сумма первых трех знаков равна сумме трех последних. Определить число номеров счастливых билетов. Номером может быть любой набор из 6 цифр, в частности, 000000.
Указание. Счастливые номера находятся во взаимно однозначном соответствии с шестизначными наборами цифр, в которых сумма цифр равна 27. Для доказательства достаточно три последних цифры номера заменить на их дополнения до 9, например 026 заменить на 973.

Задача 100. Из 100 студентов, писавших контрольную, первую задачу решили 28 человек, вторую — 30, третью — 42, первую и вторую — 8, первую и третью —- 10, вторую и третью — 5; все три задачи решили 3 студента. Какова вероятность, что случайно выбранный студент не решил ни одной задачи?

Задача 101. Из набранных в первый класс 35 учеников 18 умеют писать заглавными буквами, 14 — умеют читать по складам, а 17 — умеют считать до 10. При этом умеют и писать и читать 13, читать и считать — 10, писать и считать — 9 учеников. Умеют и писать, и читать, и считать только 4 ученика. Какова вероятность того, что произвольно выбранный ученик не умеет ни писать, ни читать, ни считать?

Задача 102. Выбирается случайное натуральное число от 1 до 600. Какова вероятность, что оно не делится ни на 2, ни на 3, ни на 5?

Задача 103. Перестановка σ на множестве $\{1, 2, \dots n\}$ называется транзитивной, если для любого k $\sigma(k) \neq k$. Найти вероятность того, что случайным образом выбранная перестановка окажется транзитивной. Чему равен предел этой вероятности при $n \to \infty$?

Задача 104. Решить задачу 93 с помощью принципа включения — исключения.

Решение

Более удобными будут обозначения не A_1, A_2 и т. д., а A, B, C, D и E. Тогда нам требуется найти вероятность события $AB \cup ACE \cup DE \cup BCD$.
Вероятность этого объединения событий равна:
сумме вероятностей событий AB, ACE, DE и BCD;
минус вероятности их попарных пересечений;
плюс вероятности их тройных пересечений;
минус вероятность пересечения всех четырех.

Найдем эти пересечения.
Из шести попарных пересечений одно имеет вид $ABCDE$, а остальные пять — все возможные пересечения по четыре: $ABCD$, $ABCE$ и так далее.
Все четыре тройные пересечения имеют вид $ABCDE$.
Единственное четырехкратное пересечение также имеет вид $ABCDE$.
Поэтому искомая вероятность равна

$$p = p_1p_2 + p_1p_3p_5 + p_4p_5 + p_2p_3p_4 - p_1p_2p_3p_4 - p_1p_2p_3p_5 - p_1p_2p_4p_5 - p_1p_3p_4p_5 - p_2p_3p_4p_5 +$$
$$+2p_1p_2p_3p_4p_5.$$

8. Вероятность хотя бы одного события

Если в каждом из n независимых испытаний некоторое событие может произойти с вероятностью p, то вероятность того, что оно не произойдет ни разу, равна $(1-p)^n$, а вероятность того, что оно произойдет хотя бы один раз, равна $1-(1-p)^n$.

Задача 105. Найти вероятность того, что при четырех бросаниях игральной кости хотя бы раз выпадет шестерка.

Решение

Часто приходится встречаться с таким неправильным решением: при одном бросании шестерка выпадает с вероятностью $1/6$, поэтому при четырех будет $4/6$. Это, разумеется, неверно. Если так рассуждать, то при восьми, скажем, бросаниях вероятность будет больше единицы.

Верный ответ дает формула вероятности хотя бы одного события. При четырех бросаниях шестерка не выпадет ни разу с вероятностью $(5/6)^4 = 625/1296$. Поэтому хотя бы раз она выпадет с вероятностью $1 - 625/1296 = 671/1296$.

Задача 106. Трое стреляют в мишень. Вероятность того, что попадет A, равна $0,2$, B — $0,3$, C — $0,4$. Найти вероятность того, что мишень будет поражена.

Задача 107. Для разрушения моста достаточно попадания одной авиабомбы. Найти вероятность того, что мост будет разрушен, если на него сбросили 4 авиабомбы, вероятности попадания которых равны 0,2, 0,3, 0,4 и 0,5.

Задача 108. Встретились четыре незнакомых человека. Найти вероятность того, что хотя бы у двух из них совпадают знаки зодиака.

Задача 109. Разыскивая специальную книгу, студент решил обойти три библиотеки. Для каждой библиотеки одинаково вероятно, есть в её фондах книга или нет. Если книга есть, то одинаково вероятно, занята она другим читателем или нет. Какова вероятность того, что студент достанет книгу?

Задача 110. (Задача Льюиса Кэрролла). В ожесточенном бою не менее 70% рыцарей потеряли один глаз, не менее 75% — одно ухо, не менее 80% — одну руку, и не менее 85% — одну ногу. Каково минимальное число потерявших одновременно глаз, ухо, руку и ногу?

Задача 111. Сколько раз надо подбросить монету, чтобы с вероятностью не менее 0,9 хотя бы раз выпали два орла подряд? три орла подряд?

Задача 112. Сколько раз надо бросить игральную кость, чтобы вероятность того, что хотя бы один раз выпадет шестерка, стала больше 0,6?

Задача 113. Сколько раз нужно бросить пару игральных костей, чтобы вероятность того, что хотя бы раз выпадет пара шестерок, стала больше 0,5? больше 0,9?

Задача 114. Из n предметов выбирают n раз по одному с возвращением. Найти вероятность того, что некоторый конкретный предмет выбран не будет. Найти также предел этой вероятности при $n \to \infty$.

9. Условная вероятность. Независимость событий

Условной вероятностью события A при условии, что событие B произошло, называется

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

При этом, разумеется, $P(B) \neq 0$.

Поскольку известно, что событие B произошло, оно становится всем пространством элементарных исходов. А множеством благоприятных исходов становится пересечение AB.

Ранее было приведено определение независимых событий:

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

Тогда из определенния условной вероятности сразу следует

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

Иными словами, вероятность события A не изменилась от того, что событие B произошло.

Задача 115. Вероятность попасть в самолет равна 0,4, а вероятность его сбить равна 0,1. Найти вероятность того, что при попадании в самолет он будет сбит.

Решение

Пусть событие A — попали, B — самолет сбит. Тогда $P(A) = 0,4$, $P(B) = 0,1$ и $P(B|A) = P(AB)/P(A) = P(B)/P(A) = 0,25$.

Задача 116. Доказать, что если события A и B независимы, то независимы и события $\overline{A}$ и $\overline{B}$.

Задача 117. Жюри состоит из трех судей. Первый и второй принимают правильное решение с вероятностью p, а третий для принятия решения подбрасывает монету. Окончательное решение принимается большинством голосов. Какова вероятность того, что жюри примет правильное решение?

Задача 118. Жюри состоит из трех судей. Каждый из них принимает правильное решение с вероятностью p. Каким должно быть p, чтобы это жюри принимало правильное решение с вероятностью большей, чем жюри из предыдущей задачи?

Задача 119. Жюри состоит из трех судей. Первый и второй принимают правильное решение с вероятностью p, а третий поступает следующим образом. Если мнения первых двух совпадают, то он к ним присоединяется, а если отличаются — бросает монетку. Какова вероятность того, что это жюри примет правильное решение?

Задача 120. За некоторый промежуток времени амеба может погибнуть с вероятностью $1/4$, выжить с вероятностью $1/4$ и разделиться на две с вероятностью $1/2$. В следующий такой же промежуток с любой амебой независимо от ее "возраста" происходит то же самое. Если первоначально была одна амеба, то сколько амеб и с какими вероятностями будут существовать к концу второго промежутка времени?

Задача 121. Двое игроков поочередно бросают монету. Выигрывает тот из них, у кого первым выпадет орел. Найти вероятность выигрыша первого игрока.

Задача 122. Как и в предыдущей задаче, двое игроков бросают монету, но после каждого бросания первого игрока второй имеет право на два бросания. Как и ранее, выигрывает тот из них, у кого первым выпадет орел. Найти вероятность выигрыша первого игрока при этих условиях.

Задача 123. В семье отец, мать и сын увлекаются шахматами. Однажды отец предложил сыну сыграть матч из трех партий, чередуя партнеров (т.е. первую партию играть с отцом, вторую — с матерью, третью — опять с отцом, или наоборот). Если сын выиграет две партии подряд, отец обещал ему подарить некоторую сумму денег на карманные расходы. Отец играет лучше, чем мать. Как сыну распределить партнеров (с кем играть первую партию), чтобы увеличить свои шансы? Каковы эти шансы, если сын выигрывает у матери с вероятностью p_1, а у отца — с вероятностью p_2?

Задача 124. На одной карточке с двух сторон написана цифра 0, на другой — с двух сторон цифра 1, а на третьей — с одной строны 0, а с другой 1. На стол выложили одну из этих карточек, видна цифра 1. Какова вероятность, что на обороте тоже 1?

Задача 125. В семье двое детей, причем один из них мальчик. Какова вероятность того, что второй тоже мальчик?

Задача 126. В семье двое детей, причем старший из них — мальчик. Какова вероятность того, что второй тоже мальчик?

Задача 127. Из множества чисел $\{1, 2, \dots n\}$ выбирают три числа. Найти вероятность того, что третье меньше второго, но больше первого, при условии, что второе больше первого.

Задача 128. Доказать, что если события A и B независимы и $\bar{B}$ непусто, то

$$P(A|\bar{B}) = P(A).$$

Задача 129. Независимы ли события при бросании кости: 1) число выпавших очков делится на 2; 2) делится на 3?

Задача 130. Сумма очков при бросании нескольких игральных костей может делиться на 2 и может делиться на 3. Независимы ли эти события?

Задача 131. Какова вероятность того, что сумма очков, выпавших на двух игральных костях, четная, если
a) учитываются только те бросания, при которых хотя бы на одной кости выпала шестерка;
b) учитываются только те бросания, при которых хотя бы на одной кости выпала единица;
Какая из вероятностей больше?

Задача 132. Бросают две игральных кости. Независимы ли события "сумма выпавших очков четная" и "сумма выпавших очков делится на три", если учитываются только те бросания, при которых хотя бы на одной кости выпала шестерка.

Задача 133(*). Каков будет ответ в предыдущей задаче, если костей будет не 2, а n?

Задача 134. Из множества $\{1, 2, \dots n\}$ случайным образом выбирают два подмножества (возможно, одинаковые) так, что все подмножества выбираются с одинаковыми вероятностями. Какова вероятность того, что эти два подмножества пересекаются? Найти предел этой вероятности при $n \to \infty$.

Решение
Будем искать вероятность того, что множества не пересекаются, и обозначим эту вероятность через $P(n)$. Тогда два подмножества из набора $\{1, 2 \dots n+1\}$ не пересекаются по каким-то из первых n элементов с вероятностью $P(n)$.

Найдем вероятность того, что они не не пересекаются по элементу $n + 1$. Этот элемент входит как в первое, так и во второе подмножество с вероятностью $1/2$, входит в оба с вероятностью $1/4$, поэтому подмножества не пересекаются по этому элементу с вероятностью $3/4$.

Следовательно
$$P(n + 1) = \frac{3}{4} P(n).$$
Легко подсчитать, что $P(1) = 3/4$. Окончательно получаем
$$P(n) = \left(\frac{3}{4}\right)^n.$$

Ответ: $1 - (3/4)^n$, предел равен 1.

10. Формула полной вероятности

Пусть пространство элементарных исходов разбито на несколько непересекающихся подмножеств $H_1, \dots H_n$, объединение которых совпадает с Ω. Эти подмножества принято называть *гипотезами*. Тогда для любого события A справедлива формула полной вероятности
$$P(A) = P(A|H_1) \cdot P(H_1) + \dots + P(A|H_n) \cdot P(H_n),$$

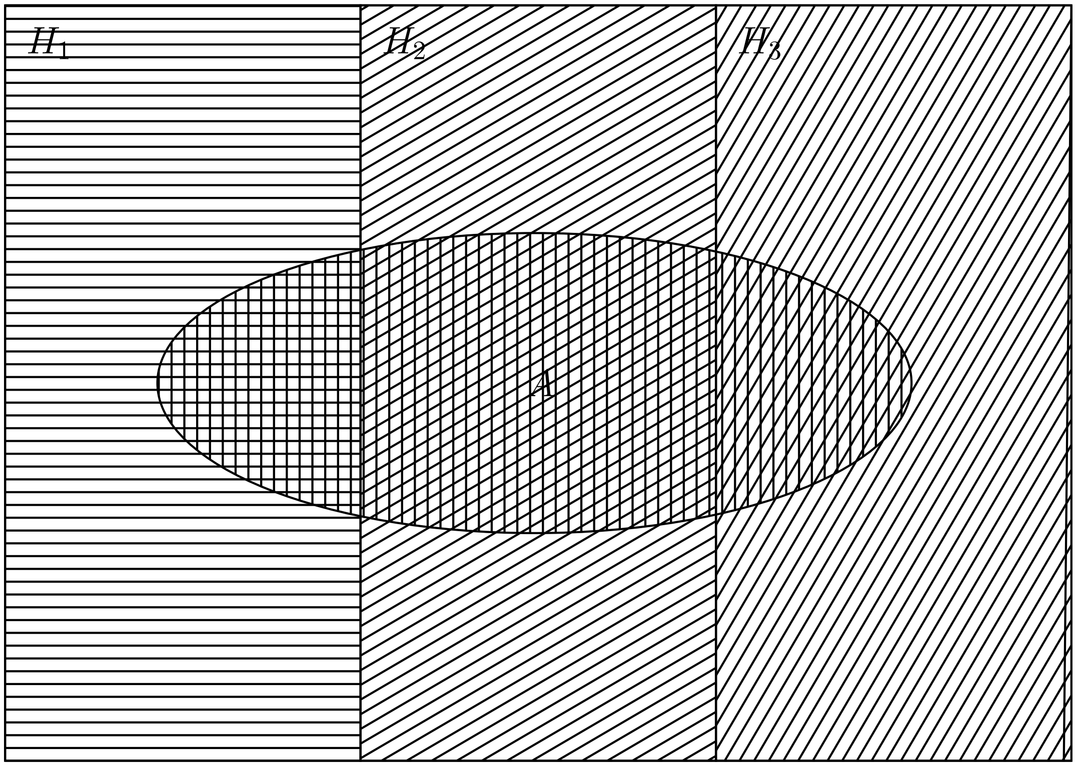


Рисунок 4. Формула полной вероятности.

Задача 135. В фирме работают 40% мужчин. В проекте заняты 25% мужчин фирмы и 15% женщин. Сколько процентов сотрудников фирмы занято в проекте?

Решение

Эта задача на формулу взвешенного среднего, но ей легко придать вероятностный смысл: какова вероятность, что наугад выбранный сотрудник занят в проекте? Запишем вероятности:
$P(H_1) = 0,4, P(H_2) = 0,6,$
$P(A|H_1) = 0,25, P(A|H_2) = 0,15.$
Поэтому по формуле полной вероятности $P(A) = 0,19$.

Задача 136. Пусть по вражескому самолету выпущено три ракеты. Вероятности попаданий равны:
ни одного попадания — 0,2;
ровно одно попадание — 0,3;
ровно два попадания — 0,3;
ровно три попадания — 0,2.

При одном попадании самолет будет сбит с вероятностью 0,2, при двух — с вероятностью 0,6, а при трех — с вероятностью 1. Найти вероятность того, что самолет будет сбит.

Решение

Обозначим H_0 событие, состоящее в том, что не было ни одного попадания, H_1 — было ровно одно попадание, и т. д. Эти события не пересекаются, и никакие другие исходы невозможны. Стало быть, мы имеем дело с набором гипотез. Применим формулу полной вероятности:

$$P(A) = P(A|H_0) \cdot P(H_0) + P(A|H_1) \cdot P(H_1) +$$
$$+ P(A|H_2) \cdot P(H_2) + P(A|H_3) \cdot P(H_3) =$$
$$= 0 \cdot 0,2 + 0,2 \cdot 0,3 + 0,6 \cdot 0,3 + 1 \cdot 0,2 = 0,44.$$

Задача 137. По вражескому самолёту было выпущено 3 ракеты. Вероятность того, что ни одна из них не попадёт в цель равна 0,1, вероятность того, что в цель попадёт ровно одна — 0,4, а того, что в цель попадут ровно две — также 0,4. Если в самолёт попадут три ракеты, то он обязательно будет сбит, если попадут две, то он будет сбит с вероятностью 0,7, а если попадёт одна — с вероятностью 0,2. Какова вероятность того, что самолёт будет сбит?

Задача 138. По вражескому самолёту было выпущено 3 ракеты. Вероятность попадания первой из них равна 0,2, второй — 0,5, а третьей — 0,8. Если в самолёт попадут три ракеты, то он обязательно будет сбит, если попадут две, то он будет сбит с вероятностью 0,6, а если попадёт одна — с вероятностью 0,2. Какова вероятность того, что самолёт будет сбит?

Задача 139. Какова надежность схемы, описанной в задаче 93?

Решение

Формула полной вероятности позволяет найти более простое решение. Если элемент 3 работает, то схема становится такой же, как в задаче 89, а если не раьотает, то как в задаче 91. Поэтому

$$P = p_3(1-(1-p_1)(1-p_4))(1-(1-p_2)(1-p_5))+(1-p_3)(1-(1-p_1p_2)(1-p_4p_5).)$$

Задача 140. За длинный стол садятся 30 человек. На одном из торцов — председатель. Аня терпеть не может Ваню и не хотела бы, чтобы он сидел рядом с ней или напротив нее. Какова вероятность, что у нее это получится, если место председателя ни Аня, ни Ваня занять не могут, а противоположный торец — могут?

Задача 141. Рабочий обслуживает 3 станка, на которых обрабатываются однотипные детали. Вероятность брака для первого станка равна 0,02, для второго — 0,03, для третьего — 0,04. Обработанные детали складываются в один ящик. Производительность первого станка в три раза больше, чем второго, а третьего в два раза меньше, чем второго. Определить вероятность того, что взятая наудачу деталь будет бракованной.

Задача 142. На овощехранилище поступает продукция от трёх хозяйств. Продукция первого хозяйства составляет 20% , второго — 46% и третьего — 34% . Известно, что средний процент нестандартных овощей для первого хозяйства равен 3% , для второго — 2% , для третьего — 1% . Найти вероятность того, что наудачу взятый овощ оказался нестандартным.

Задача 143. На заводе, производящем болты, машины A, B, C производят соответственно 25, 35 и 40% всех изделий. В их продукции брак составляет соответственно 5, 4 и 2%. Какова вероятность, что случайно выбранный из продукции болт окажется дефектным?

Задача 144. В некоторых сельских местностях России существовало когда-то следующее гадание. Девушка зажимает в руке шесть травинок так, чтобы концы травинок торчали сверху и снизу; подруга связывает эти травинки попарно между собой сверху и снизу в отдельности. Если при этом все шесть травинок оказываются связанными в кольцо, то это должно было означать, что девушка в текущем году выйдет замуж. Найти вероятность того, что травинки при завязывании наудачу образуют кольцо.

Задача 145. В каждой из трёх урн содержится 6 чёрных и 4 белых шара. Из первой урны наудачу извлечён один шар и переложен во вторую урну, после чего из второй урны наудачу извлечён один шар и переложен в третью урну. Найти вероятность того, что шар, наудачу извлечённый из третьей урны, окажется белым.

Задача 146. Из полного набора домино выбирают случайным образом две кости. Найти вероятность того, что их можно приложить друг к другу.

Задача 147. В некоторой области профессиональный и бытовой травматизм за определенный срок составляет 1 случай на 60 тыс. человек, в том числе в областном центре — 1 случай на 80 тыс. человек, а в остальной части области — 1 случай на 40 тыс. человек. Известно, что во всей области проживают 2 миллиона 400 тысяч человек. Сколько из них проживают в областном центре, а сколько — в остальной части области?

Задача 148. Отец, желая простимулировать занятия сына теорией вероятностей, предлагает ему такую игру. Сын должен десять купюр по 100$ и десять купюр по 1$ разложить по двум шляпам. Затем сын с завязанными глазами выбирает шляпу, а затем — одну купюру из выбранной шляпы. Если будет выбрана купюра в 100$, сын получит её в подарок. Купюры на ощупь не отличаются. Как сын должен распределить купюры по шляпам, чтобы увеличить вероятность успеха, и чему равна эта вероятность?

Задача 149. На шахматную доску случайным образом ставят белого короля и какую-то черную фигуру. Какова вероятность, что черная фигура окажется под боем?

Задача 150. В урне находятся 7 белых и 2 красных шара. Трое игроков по очереди извлекают по одному шару, отмечают цвет и возвращают шар обратно. Выигрывает тот, кто первым достанет красный шар. Найти вероятности выигрыша для каждого из игроков, если игра может продолжаться неограниченно.

Решение
Обозначим вероятности выигрыша первого, второго и третьего игроков через p_1, p_2 и p_3 соответственно. Если первым извлекается белый шар (а вероятность этого равна $7/9$), то второй игрок оказывается в том же положении, в котором первоначально был первый игрок. Следовательно, можно записать уравнение

$$p_2 = \frac{7}{9}p_1.$$

Аналогично

$$p_3 = \frac{7}{9}p_2.$$

Поскольку кто-то из игроков обязательно выиграет

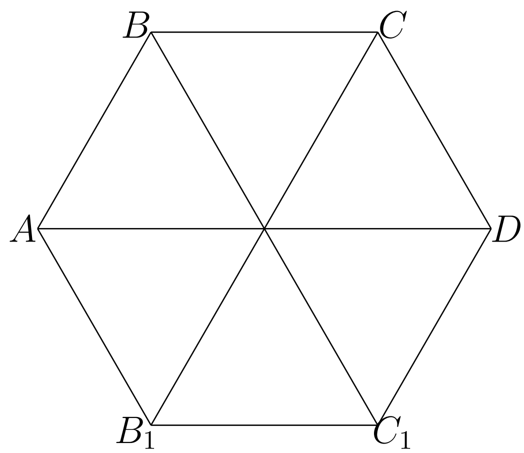
$$p_1 + p_2 + p_3 = 1.$$

Решая эту систему, получим

$$p_1 = \frac{81}{193}; \quad p_2 = \frac{63}{193}; \quad p_3 = \frac{49}{193}.$$

Ответ: 81/193; 63/193; 49/193.

Задача 151. Пьяный десантник бродит по торговому комплексу, имеющему форму правильного шестиугольника. На каждой развилке десантник выбирает дальнейший маршрут случайным образом, в частности, может повернуть назад. В центре комплекса находится фонтан. Если десантник попадает в центр, то он лезет купаться в этот фонтан, в результате чего его тут же забирают в полицию.



Какова вероятность того, что, начав свой путь в точке A, он в нее вернется?

Решение

Обозначим искомую вероятность через p_A. Вероятности возвращения в точку A из вершин B, C и D обозначим соответственно через p_B, p_C и p_D. Из симметрии сразу ясно, что вероятности возвращения в точку A из вершин B_1 и C_1 равны соответственно p_B и p_C.

Пользуясь формулой полной вероятности, можно записать систему уравнений

$$\begin{cases} p_A = \frac{2}{3}p_B \\ p_B = \frac{1}{3} + \frac{1}{3}p_C \\ p_C = \frac{1}{3}p_B + \frac{1}{3}p_D \\ p_D = \frac{2}{3}p_C \end{cases}$$

Система легко решается.
Ответ: 7/27.

11. Формула Байеса

Еще одной важной формулой, служащей для переоценки вероятностей гипотез по результатам прошедших испытаний, является формула Байеса.

Из равенств

$$P(AH_1) = P(A|H_1) \cdot P(H_1) = P(H_1|A) \cdot P(A)$$

следует

$$P(H_1|A) = \frac{P(A|H_1) \cdot P(H_1)}{P(A)}.$$

Эта формула позволяет переоценивать вероятности гипотез на основе новой информации. Эта формула часто используется, например, в задачах распознавания образов. В этих случаях гипотезы непосредственно не наблюдаемы, и о них можно судить только по косвенным данным. Пусть, например, мы пытаемся распознать сигнал (букву, команду и т. п.) из перечня (алфавита), состоящего из n сигналов (букв). Тогда $H_1, \dots H_n$ — это просто известные априорные вероятности,то есть частоты встречаемости букв. То, что мы измерили, назовем событием A. Тогда формула Байеса позволяет нам на основании известных вероятностей $P(A|H_i)$ вычислить апостериорные вероятности $P(H_i|A)$, то есть точнее распознать, какой был сигнал.

Описанное выше называют байесовским методом распознавания (или байесовским подходом, и т. д.). Такие методы используются часто, и вообще это направление в последнее время интенсивно развивается. Считается, что байесовский подход является ключевым в задачах машинного обучения — обучения на примерах.

Задача 152. Тест на туберкулез (реакция Манту) дает положительный результат для больных с вероятностью 0,95, а для здоровых — с вероятностью 0,1. Среди тестируемых 0,1 % больных. Какова вероятность, что человек здоров, если тест дал для него положительный результат?

Решение

Обозначим H_0 — событие "тестируемый здоров", H_1 — событие "тестируемый болен", A — "тест дал положительный результат". Тогда по условию задачи:
$P(H_0) = 0,999, P(H_1) = 0,001;$
$P(A|H_0) = 0,1, P(A|H_1) = 0,95.$
По формуле полной вероятности

$$P(A) = 0,999 \cdot 0,1 + 0,001 \cdot 0,95 = 0,10085.$$

По формуле Байеса
$$P(H_0|A) = \frac{0,1 \cdot 0,999}{0,10085} = 99,06\%.$$

Задача 153. Вероятность для изделий некоторого производства соответствовать стандарту равна 0,96. Предлагается упрощённая система контроля качества, дающая положительный результат с вероятностью 0,98 для изделий, удовлетворяющих стандарту, а для изделий, которые не удовлетворяют стандарту, с вероятностью 0,05. Какова вероятность, что изделие, выдержавшее испытание, удовлетворяет стандарту?

Задача 154. Передаваемое зашифрованное сообщение состоит из кодов 000, 111 или 101. Частота встречаемости первого кода равна 0,5, а второго — 0,3. Каждый символ может быть искажен при передача с вероятностью 1/5 независимо от других. Если был принят код 100, то какой код, вероятнее всего, был передан?

Задача 155. (Задача Монти Холла, из книги [8]). Участники телевикторины должны выбрать одну из трех дверей. За одной дверью находится машина, за двумя другими — по козе. Участник выбирает дверь (скажем, первую), а ведущий, которому известно, что находится за каждой из дверей, открывает одну из оставшихся (скажем, вторую), за которой коза. Затем он говорит участнику: «Итак, вы смените дверь или останетесь на месте?» Вопрос в следующем: выгодно ли участнику сменить дверь? Каковы теперь вероятности, что приз находится за первой, второй и третьей дверью?

Задача 156.При передаче сообщения сигналами азбуки Морзе сигналы “точка” и “тире” встречаются в отношении 5:3. Статистические свойства помех таковы, что 1/25 часть сигналов “точка” превращаются в “тире” и 1/30 часть сигналов “тире” превращаются в “точка”. Найти вероятность того, что принятое “тире” — действительно “тире”.

Задача 157. В прибывшей партии апельсинов качественных 90%. Сортировщик апельсинов бракует некачественный апельсин с вероятностью 0,98, а качественный — с вероятностью 0,05. Какой процент некачественных апельсинов будет среди отсортированных?

Задача 158. Когда Иван-царевич уходил из дома, царевна-лягушка всегда делала в доме уборку и готовила вкусный обед. К сожалению, Кощей похитил царевну-лягушку и унес ее на Заколдованное Болото.

Иван пока не придумал, как расправиться с Кощеем, но он сходил на Заколдованное Болото и обнаружил там 10 лягушек, неотличимых друг от друга. Одну из них он принес домой. Иван-царевич знает, что если он ошибся, и это не царевна-лягушка, то с вероятностью $1/2$ она приберется в доме и приготовит вкусный обед, а с вероятностью $1/2$ — устроит в доме дебош.

Вот уже пять дней Иван-царевич, возвращаясь домой, обнаруживает там чистоту и вкусный обед. Какова вероятность того, что он не ошибся? Сколько дней так должно продолжаться, чтобы Иван был в этом уверен не менее, чем на 99%?

Решение

Введем обозначения:
H_1 — Иван не ошибся;
H_2 — Иван ошибся;
A — пять дней в доме обед и уют.
B_n — n дней в доме обед и уют.
По условию задачи априорные вероятности равны:

$$P(H_1) = 0,1; \quad P(H_2) = 0,9 \quad P(A|H_1) = 1; \quad P(A|H_2) = \frac{1}{32}.$$

Поэтому по формуле Байеса

$$P(H_1|A) = \frac{P(A|H_1) \cdot P(H_1)}{P(A)} = \frac{32}{41}.$$

Далее

$$P(B_n|H_1) = 1; \quad P(B_n|H_2) = \frac{1}{2^n}.$$

По формуле Байеса

$$P(H_1|B_n) = \frac{P(B_n|H_1) \cdot P(H_1)}{P(B_n)} = \frac{2^n}{2^n + 9}.$$

Решая неравенство

$$\frac{2^n}{2^n + 9} \geqslant 0,99,$$

получим $2^n \geqslant 891$. Поскольку n — число целое, то $n \geq 10$.

Ответ: $32/41$; 10.

Задача 159. Профессор культурологии К., обедая в институтской столовой, обнаружил у себя в порции плова всего 1 кусочек мяса. Вызванный для объяснений шеф-повар пояснил, что мясо в котёл с пловом положено достаточно, и вероятность того, что в тарелке окажется всего 1 кусок, равна 0,01. Правда, есть ещё два котла — для студентов и для подшефной свинофермы, для которых вероятность попадания всего одного куска в порцию равны соответственно 0,2 и 0,5, но он, шеф-повар, уверен, что профессор получил свою порцию из профессорского котла. Знакомый профессора, главный инженер, рассказал затем профессору, что когда раздатчица пьяная, ей становится всё равно, из какого котла накладывать порции. Зная, что, по мнению главного инженера, вероятность пьянства раздатчицы равна $1/4$, оцените с помощью формулы Байеса справедливость высказывания шеф-повара.

Задача 160. Охранник торгового комплекса пропускает внутрь только посетителей с QR-кодом. Поскольку считывателя QR-кодов у него нет, он вынужден доверять своему представлению о том, законопослушный перед ним посетитель или нет. Но глаз у охранника наметан, и он пропускает 70% тех, у кого есть QR-код, и только 20% тех, у кого его нет. Сколько процентов вошедших посетителей имеют QR-код, если среди тех, кто пытается войти, таких 60%?

12. Разные задачи по теме 1

Задача 161. Дважды бросают две монеты. Какова вероятность того, что при обоих бросаниях выпадет одинаковое число орлов?

Задача 162. Четыре монеты весят 12, 13, 14 и 15 грамм. На каждую из двух чаш весов кладут по 2 монеты. Какова вероятность, что весы окажутся а равновесии?

Задача 163. В урне n шаров, в том числе 2 красных, остальные — белые. Случайным образом выбирают два. Найти вероятности событий:
a) оба шара белые;
b) один из шаров — белый, а другой — красный.
При каком значении n эти вероятности равны?

Задача 164. (Задача Пачоли о разделе ставки) Двое равносильных игроков играют матч до 6 побед в игру, где нет ничьих, то есть каждая партия может кончиться только победой одного из игроков (во времена Лука Пачоли, в XIV — начале XV веков, скорее всего, это была игра в орлянку). Игроки были вынуждены прервать игру при счете 5:3 в пользу одного из них. Как по справедливости должна быть разделена ставка между ними?

Задача 165. Какой ответ был бы в предыдущей задаче, если бы игра была прервана при счете 4:3?

Задача 166. Двое соперников играют матч. При этом для победы первому надо выиграть 12 партий, а второму — 6. Ничьих не бывает. Первый играет сильнее, вероятность его победы в каждой партии равна $2/3$. Матч был прерван при счете 10:5. Какова вероятность победы первого игрока?

Задача 167. Какой ответ был бы в предыдущей задаче, если бы матч был прерван при счете 8:4?

Задача 168. Ваня бросил монету 2021 раз, а Вася — 2022 раза. Найти вероятность того, что у Васи орлов выпало больше, чем у Вани.

Задача 169. На шахматную доску ставят белого короля и черную ладью. Найти вероятность того, что ладья бьет короля, а король не бьет ладью.

Задача 170. В колоде 10 красных карт. Сколько надо добавить черных карт, чтобы вероятность того, что две вытащенные карты оказались одной масти, стала больше $1/2$?

Задача 171. Из колоды карт (52 шт) выбирают две. Найти вероятность того, что вторая карта бьет первую.

Задача 172. Д'Артаньян и Портос играют в такую азартную игру. Есть три монеты, достоинством в $1/2$ луидора, 1 луидор и 2 луидора. Одна из монет принадлежит д'Артаньяну, а остальные две —Портосу. Их одновременно подбрасывают. Если выпадает решка, владелец монеты не получает ничего, а если орел — получает столько очков, какова стоимость монеты. Тот, кто набрал больше очков, забирает все три монеты. Какая монета должна принадлежать д'Артаньяну, чтобы игра была с равными шансами?

Задача 173. Двое играют в "орел или решка" несимметричной монетой, у которой вероятность выпадения орла p больше $1/2$, но меньше 1. Правила игры таковы:
1. Первый игрок бросает монету дважды и выигрывает, если выпали две раз-

ные стороны в любом порядке. Если же дважды выпала одна и та же сторона, очередь хода переходит ко второму игроку.
2. Второй игрок также бросает монету дважды. Он выигрывает в двух случаях:
— если выпали две разные стороны в любом порядке;
— если дважды выпала та же сторона, что дважды выпала у первого.
3. Если после всех этих бросаний монеты ни один игрок не выиграл, то очередь хода вновь переходит к первому игроку и игра начинается с самого начала.
Определить, при каком значении p шансы игроков на выигрыш равны.

Задача 174. Отрезок разбит на три непересекающиеся части, вероятности поподания случайной точки в них равны соответственно p_1, p_2 и p_3, причем $p_1 + p_2 + p_3 = 1$. Какова вероятностьтого, что из трех случайно брошенных на отрезок точек в каждую часть попала ровно одна?

Задача 175. Случайным образом выбирают две точки на отрезке $[0, 1]$. Найти вероятность того, что большая больше $3/4$ при условии, что меньшая меньше $1/4$.

Тема 2. Случайные величины

13. Последовательность независимых испытаний

Пусть проводится n испытаний, в каждом из которых независимо от других может быть или успех, или неудача. Обозначим вероятность успеха в одном испытании через p, а вероятность неудачи — через $q = 1 - p$. Сколько будет успехов?

Вероятность того, что будет ровно k успехов, равна

$$P_n(X = k) = C_n^k p^k q^{n-k}.$$

Такая схема называется *последовательностью независимых испытаний* или *схемой Бернулли.*

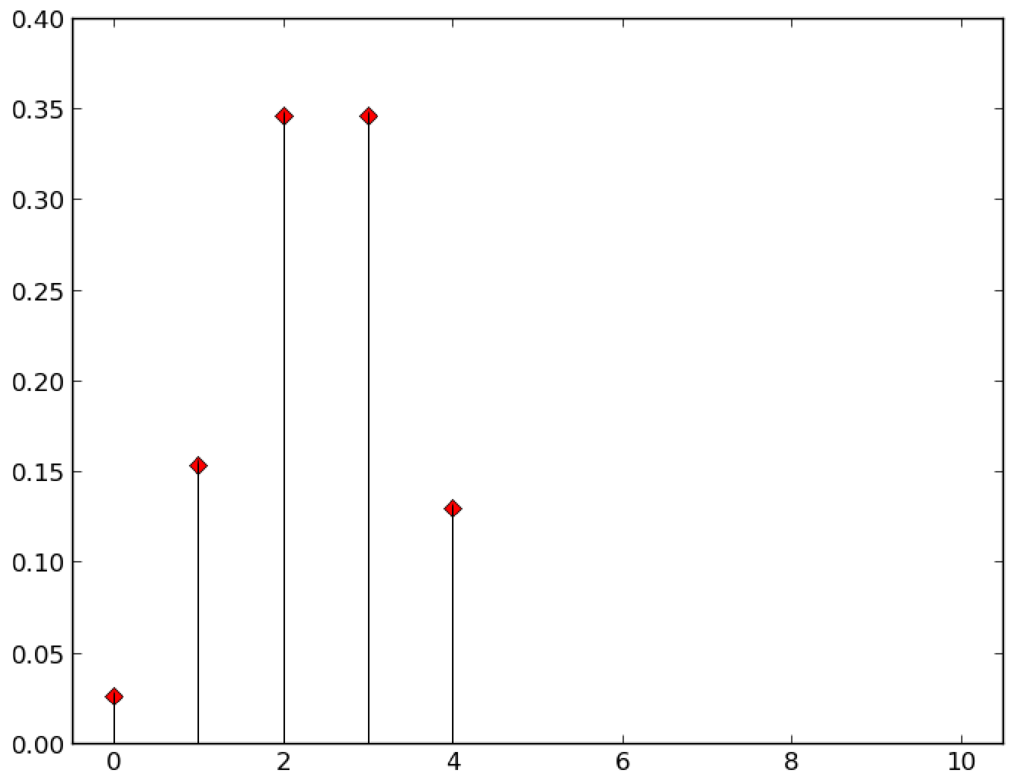


Рисунок 5. Распределение Бернулли с $n = 4, p = 0,6$.

Задача 176. Проводится матч из 4 партий между двумя игроками, причем вероятность победы первого игрока в каждой партии одинакова и равна $3/5$, а ничьх не бывает. Каков будет счет? В частности, какова вероятность того, что матч закончится вничью?

Решение. Вероятность ничьей равна

$$P(2:2) = C_4^2\left(\frac{3}{5}\right)^2\left(\frac{2}{5}\right)^2 = \frac{216}{625}.$$

Можно рассчитать вероятности для каждого возможного исхода матча. Мы получим такую таблицу, где в первой строке указано число побед первого игрока, а во второй — соответствующие вероятности.

X	0	1	2	3	4
P	0,0256	0,1536	0,3456	0,3456	0,1296

Задача 177. Двое равносильных партнеров играют в игру без ничьих матч из 4 партий, Найти вероятность того, что будет ничья 2:2.

Задача 178. Проводится матч из 4 партий между двумя игроками, причем вероятность победы первого игрока в каждой партии одинакова и равна 3/4, а ничьх не бывает. Найти закон распределения случайной величины — числа побед первого игрока. Какой счет матча наиболее вероятен? Какова вероятность того, что матч закончится вничью?

Задача 179. Устройство состоит из восьми независимо работающих элементов. Вероятность отказа одинакова для каждого из этих элементов и равна 0,2. Элементы отказывают независимо друг от друга. Найти вероятность отказа устройства, если для этого нужно, чтобы отказали не менее трех элементов.

Задача 180. В семье 5 детей. Считая, что вероятность рождения мальчика равна 0,51 и что пол каждого ребенка не зависит от пола остальных детей в семье, найти закон распределения случайной величины — числа мальчиков в семье.

Задача 181. В сентябре в Подмосковье в среднем бывает 12 дождливых дней. Найти вероятность того, что из восьми наугад взятых дней будут ровно два дождливых; ровно три дождливых. Какая из этих вероятностей больше?

Задача 182. Что более вероятно: в игре с равным противником выиграть 3 партии из 6 или 4 из 8? Вычислить обе эти вероятности. Считать, что ничьих не бывает.

Задача 183. Двое играют в игру с вероятностью успеха каждого $p = 0,5$. Ничьих не бывает (пример такой игры — подбрасывание монеты). Проводится матч из 10 партий. Назовем разгромом счет 8:2 и больше в пользу одного из участников. Какова вероятность разгрома?

Пусть состоялся не один матч из десяти партий, а десять таких матчей. Что более вероятно: что ни разу не было ни одного разгрома, или что разгром хотя бы раз все-таки был?

Задача 184. Три одинаковых игральных кости бросают 7 раз. Какова вероятность, что сумма очков, равная 7, выпадет дважды?

Задача 185. Сколько раз надо бросить игральную кость, чтобы вероятность того, что шестерка выпадет не менее двух раз, стала больше 2/3?

Решение
При n бросаниях вероятность того, что шестерка выпадет менее двух раз, равна

$$P(<2) = \left(\frac{5}{6}\right)^n + C_n^1\left(\frac{1}{6}\right)\cdot\left(\frac{5}{6}\right)^{n-1}.$$

По условию эта вероятность должна быть меньше 1/3. Получим неравенство

$$\left(\frac{5}{6}\right)^{n-1}\left(\frac{5}{6}+\frac{n}{6}\right) < \frac{1}{3}$$

$$\frac{5+n}{2} < \left(\frac{6}{5}\right)^{n-1}.$$

С помощью калькулятора легко подсчитать, что правая часть становится больше, начиная с $n = 14$.
Ответ: 14 раз.

14. Дискретные случайные величины.

Дискретной случайной величиной, или законом распределения дискретной случайной величины принято называть правило, которое позволяет определить, какие значенния и с какими вероятностями эта величина принимает. Часто такое правило можно представить в виде таблицы.

X	x_1	x_2	$\ldots$	x_n
P	p_1	p_2	$\ldots$	p_n

Здесь $x_1, \dots x_n$ — значения, которые может принимать с. в., а $p_1, \dots p_n$ — вероятности, с какими эти значения принимаются. Должны выполняться два важных свойства:
1. $p_i \geqslant 0$;
2. $p_1 + \dots + p_n = 1$.

Математическое ожидание — усредненное значение случайной величины

$$MX = x_1 p_1 + x_2 p_2 + \dots + x_n p_n.$$

Дисперсия случайной величины

$$DX = M(X - MX)^2.$$

Стандартное (среднеквадратическое) отклонение — квадратный корень из дисперсии

$$\sigma(X) = \sqrt{DX}.$$

Дисперсия и стандартное отклонение характеризуют разброс случайной величины относительно среднего.

Для распределения Бернулли

$$MX = np; \quad DX = npq.$$

Задача 186. Случайная величина X принимает значения 20 ± 5. Найти математическое ожидание и дисперсию.

Решение.

Считаем, что случайная величина X принимает значения 15 и 25 с одинаковыми вероятностями. Поэтому

$$MX = \frac{1}{2} \cdot 15 + \frac{1}{2} \cdot 25 = 20.$$

$$MX^2 = \frac{1}{2} \cdot 225 + \frac{1}{2} \cdot 625 = 425.$$

$$DX = MX^2 - (MX)^2 = 25, \quad \sigma = 5.$$

Задача 187. Случайная величина X — число очков при бросании одной игральной кости. Найти математическое ожидание и дисперсию.

Задача 188. Случайная величина принимает только значения 1 и 2, причем 1 — с вероятностью 0,3. Найти математическое ожидание и дисперсию.

Задача 189. Для приведенного закона распределения дискретной случайной ведичины найти математическое ожидание и дисперсию:

X	1	2	3	4
P	$1/2$	$1/4$	$1/8$	$1/8$

Задача 190. Для приведенного закона распределения дискретной случайной ведичины заполнить пропуск, найти математическое ожидание и дисперсию:

X	1	2	4	8
P	$1/2$	$1/4$	$1/6$	

Задача 191. Для приведенного закона распределения дискретной случайной ведичины заполнить пропуск, найти математическое ожидание и дисперсию:

X	1	2	3	4
P	$1/3$	$1/6$	$1/4$	

Задача 192. Для приведенного закона распределения дискретной случайной ведичины заполнить пропуски и найти дисперсию, если математическое ожидание равно 3.

X	1	2	3	4
P	$1/4$	$1/4$		

Задача 193. Получить прямой выкладкой формулы математического ожидания и дисперсии для биномиального распределения.

Задача 194. Вероятность успеха в каждом из независимых испытаний одинакова и равна p. Найти вероятность того, что k-й успех наступит в испытании с номером m $(k < m)$.

Решение

Искомая вероятность равна произведению вероятности того, что в предыдущих $m-1$ испытаниях был ровно $k-1$ успех, то есть

$$C_{m-1}^{k-1}p^{k-1}(1-p)^{m-k},$$

на вероятность успеха в последнем испытании, то есть на p.

Ответ:
$$C_{m-1}^{k-1}p^k(1-p)^{m-k}.$$
Полученное распределение имеет свое название: отрицательное биномиальное распределение.

Задача 195. Два стрелка независимо друг от друга делают по одному выстрелу в мишень. Первый попадает с вероятностью p_1, второй — с вероятностью p_2. Случайная величина — число попаданий. Найти ее закон распределения, математическое ожидание и дисперсию.

Задача 196. Случайная величина X принимает только целые неотрицательные значения. Доказать, что ее математическое ожидание может быть найдено по формуле
$$MX = \sum_{k=0}^{\infty} P(X > k).$$

Задача 197. Двое играют в такую игру: они одновременно выбрасывают один или два пальца, если общее число пальцев нечетно, выигрывает первый игрок, а если четно — второй. При этом проигравший выплачивает выигрыш, равный количеству пальцев. Как надо правильно играть и каковы вероятности выигрыша для игроков при правильной игре?

Решение

Это — задача из теории игр. Условия игры можно свести в таблицу — она называется матрицей платежей:

1-й игрок \ 2-й игрок	1 палец	2 пальца
1 палец	-2	3
2 пальца	3	-4

Эта таблица составлена с точки зрения первого игрока: положительные числа — это его выигрыш, а отрицательные — его проигрыш. Такие игры называют антагонистическими: то, что проигрывает первый игрок, выигрывает второй, и наоборот.

Если игра будет повторяться достаточно долго, то каждому из игроков невыгодно вести себя предсказуемо. Поэтому обоим игрокам следует выкидывать поочередно один или два пальца случайным образом с некоторыми вероятностями.

Как выбирать эти вероятности и кто выиграет при правильной игре?

Пусть первый игрок выбрасывает один палец с вероятностью x, а два — с вероятностью $1-x$, а второй соответственно с вероятностями y и $1-y$. Тогда средний выигрыш в одной партии равен

$$S(x,y) = -2xy + 3x(1-y) + 3(1-x)y - 4(1-x)(1-y) =$$
$$= -12xy + 7x + 7y - 4.$$

Находя частные производные по x и y и приравнивая их к нулю, получим оптимальные вероятности

$$x = \frac{7}{12}, \quad y = \frac{7}{12}.$$

Если каждый играет оптимально, то выигрыш в каждой партии будет представлять собой случайную величину, закон распределения которой представлен в таблице.

X	-2	3	-4
P	49/144	70/144	25/144

Математическое ожидание этой случайной величины, то есть средний выигрыш, равен

$$MX = \frac{1}{12}.$$

Таким образом, игра выгодна первому игроку. Это математическое ожидание, впрочем, можно было найти и проще: оно равно $S(x,y)$ при найденных оптимальных x и y.

Интересно отметить, что если оба игрока действуют оптимально, то второй игрок выигрывает чаще: соотношение побед — 70:74. Однако выигрыши первого игрока в среднем больше.

Задача 198. Чему равна дисперсия случайной величины — выигрыша первого игрока из предыдущей задачи?

В задачах 199 — 202 проанализировать антагонистическую игру с заданной матрицей платежей:найти закон распределения случайной величины — выигрыша первого игрока, математическое ожидание и дисперсию этой случайной величины.

Задача 199.

1-й игрок \ 2-й игрок	выбор 1	выбор 2
выбор 1	-4	2
выбор 2	5	-3

Задача 200.

1-й игрок \ 2-й игрок	выбор 1	выбор 2
выбор 1	-7	1
выбор 2	8	-4

Задача 201.

1-й игрок \ 2-й игрок	выбор 1	выбор 2
выбор 1	-5	6
выбор 2	7	-8

Задача 202.

1-й игрок \ 2-й игрок	выбор 1	выбор 2
выбор 1	-5	2
выбор 2	4	-1

Задача 203. Генерируют случайные числа, равномерно распределенные на $(0, 1)$, пока их сумма не станет больше 1. Пусть X — количество сгенерированных чисел. Найти закон распределения случайной величины X, ее математическое ожидание и дисперсию.

Решение

Пусть x_i — i-е сгенерированное число. Тогда событие $(X > n)$ в точности означает, что сумма n сгенерированных чисел меньше 1. Вероятность этого события равна n-мерному объему гиперпризмы, расположенной в n-мерном единичном кубе ниже гиперплоскости $x_1 + ... + x_n = 1$. Нетрудно доказать, например, по индукции, что этот объем равен $1/n!$. Таким образом

$$P(X > n) = \frac{1}{n!}.$$

Из выведенной формулы легко получить закон распределения вероятностей: при $n > 1$

$$P(X = n) = P(X > n-1) - P(X > n) = \frac{1}{(n-1)!} - \frac{1}{n!} = \frac{n-1}{n!}.$$

Математическое ожидание равно

$$MX = \sum_{n=2}^{\infty} \frac{n(n-1)}{n!} = \sum_{n=2}^{\infty} \frac{1}{(n-2)!} = e.$$

Математическое ожидание квадрата

$$MX^2 = \sum_{n=2}^{\infty} \frac{n^2(n-1)}{n!} = \sum_{n=2}^{\infty} \frac{n}{(n-2)!} = \sum_{n=2}^{\infty} \frac{n-2+2}{(n-2)!} =$$

$$= \sum_{n=3}^{\infty} \frac{(n-3)}{(n-3)!} + \sum_{n=2}^{\infty} \frac{2}{(n-2)!} = 3e.$$

Поэтому

$$DX = MX^2 - (MX)^2 = e(3-e).$$

Задача 204(*). В соревнованиях участвуют N спортсменов, в том числе m — из "Спартака". Пусть X — минимальный номер места, занятого спартаковцем. Найти математическое ожидание случайной величины X. Предполагается, что шансы всех спортсменов одинаковы.

15. Распределение Пуассона

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

Оно получается из распределения Бернулли при $n \to \infty$, $p \to 0$ так, что при этом $np = \lambda$. Для этого распределения

$$MX = \lambda; \quad DX = \lambda.$$

Распределение Пуассона часто используется в задачах массового обслуживания: поток вызовов часто представляет собой стационарный пуассоновский поток. Дла того, чтобы поток вызовов был Пуассоновским, достаточно выполнения следующих условий:
— среднее число событий за некоторый интервал времени пропорционально длительности интервала и не зависит от момента начала этого интервала

(стационарность);
— числа событий в непересекающиеся интервалы времени независимы (отсутствие последействия);
— события появляются поодиночке, то есть вероятность появления двух и более событий в малом интервале Δt есть бесконечно малая функция $\bar{o}(\Delta t)$ (ординарность).

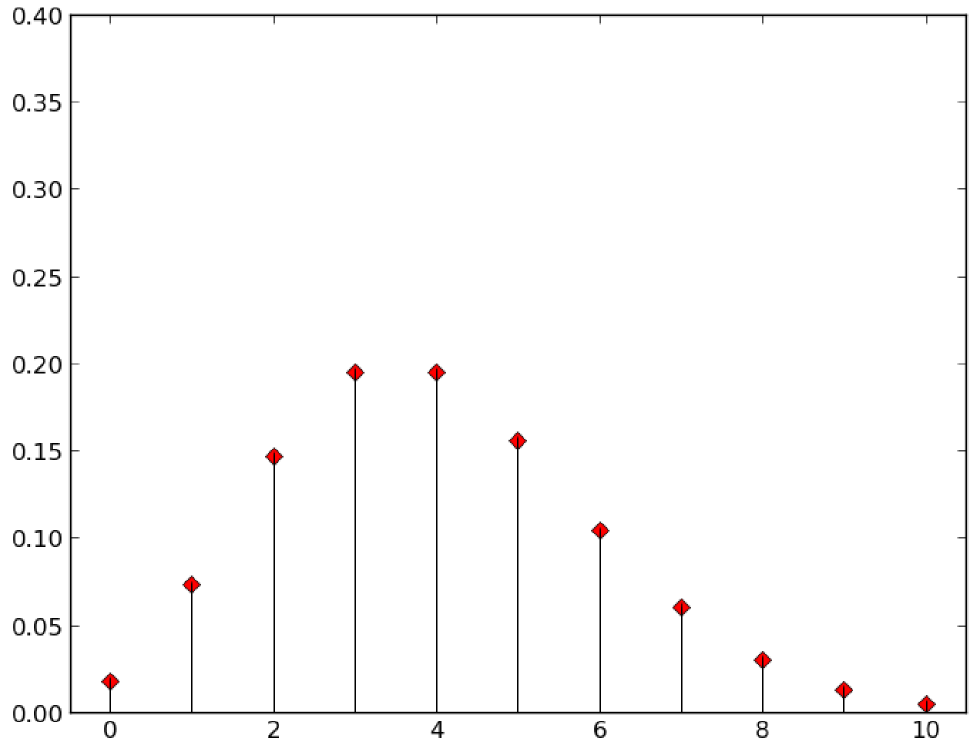


Рисунок 6. Распределение Пуассона с $\lambda = 4$.

Задача 205. Доказать, что предел распределения Бернулли при $n \to \infty$, при условии $np = \lambda$, равен распределению Пуассона.

Решение

Пусть $n \to \infty$, а $p \to 0$ так, что их произведение $np = \lambda = \text{const}$. Зафиксируем k и найдем предел вероятности события $(X = k)$. При этом

$$p = \frac{\lambda}{n}, \quad q = 1 - \frac{\lambda}{n},$$

так что

$$P(X=k)=C_n^k p^k q^{n-k}=\frac{n!}{k!(n-k)!}\cdot\left(\frac{\lambda}{n}\right)^k\cdot\left(1-\frac{\lambda}{n}\right)^{n-k}=$$

$$=\frac{\lambda^k}{k!}\cdot\frac{n!}{(n-k)!}\cdot\frac{1}{n^k}\cdot e^{(n-k)\ln(1-\frac{\lambda}{n})}.$$

Дробь $n!/(n-k)!$ является, очевидно, многочленом от n степени k. Обозначим его $M_k(n)$ и заметим, что его старший коэффициент равен 1.
Продолжим вычисления

$$P(X=k)=\frac{\lambda^k}{k!}\cdot\frac{M_k(n)}{n^k}\cdot e^{(n-k)\ln(1-\frac{\lambda}{n})}.$$

Предел отношения $M_k(n)/n^k$ очевидно, равен 1, поскольку многочлены и в числителе, и в знаменателе имеют одинаковую степень и одинаковый старший коэффициент. Осталось найти предел показателя экспоненты. Для этого преобразуем его в дробь и применим правило Лопиталя:

$$(n-k)\ln\left(1-\frac{\lambda}{n}\right)=\frac{\ln(1-\frac{\lambda}{n})}{\frac{1}{n-k}}\sim\frac{\frac{\lambda}{n^2}\cdot\frac{1}{1-\frac{\lambda}{n}}}{-\frac{1}{(n-k)^2}}=-\frac{\lambda}{1-\frac{\lambda}{n}}\cdot\frac{(n-k)^2}{n^2}\to-\lambda,$$

поскольку предел второго сомножителя в последнем выражении, очевидно, равен 1. Поэтому окончательно получаем

$$\lim_{n\to\infty}P(X=k)=\frac{\lambda^k}{k!}e^{-\lambda},$$

что и требовалось доказать.

Задача 206. В тесто для 200 сдобных булочек положили 1000 изюминок. Какова вероятность того, что в одной отдельно взятой булочке оказалось ровно 5 изюминок? не оказалось изюма вообще?

Задача 207. Сколько изюма в среднем должны содержать калорийные булочки для того, чтобы вероятность иметь в булке хотя бы одну изюминку была не менее 0,99? Считать, что число изюминок подчиняется распределению Пуассона.

Задача 208. Семена содержат 0,6% сорняков. Какова вероятность того, что при случайном отборе 1000 семян будет найдено менее 3 семян сорняков; ровно 6 семян сорняков?

Задача 209. При приемке большой партии транзисторов выбирают случайным образом 100 штук и подвергают детальной проверке. Партия бракуется, если хоть один из проверенных транзисторов бракованный. Какой должен быть процент брака в партии, чтобы с вероятностью 0,99 она не была забракована?

Задача 210. Известно, что вероятность выпуска сверла повышенной хрупкости (брак) равна 0,02. Свёрла укладывают в коробки по 100 штук. Чему равна вероятность того, что а) в коробке не окажется бракованных свёрл? б) число бракованных свёрл окажется более двух? Считать, что число бракованных свёрл подчинено распределению Пуассона.

Задача 211. Вероятность попадания в уязвимое место самолета из пехотного оружия равна 0,001. Для того, чтобы подбить самолет, необходимо как минимум два таких попадания. С какой вероятностью самолет будет подбит, если по нему произведено 4000 выстрелов?

Задача 212. На факультете 2000 студентов. Какова вероятность того, что ровно 6 из них родились 1 мая (может быть, в разные годы)? Считать, что вероятность рождения в конкретный день года равна 1/365.

Задача 213.Застраховано 10000 лиц одного возраста и одной социальной группы. Вероятность страхового случая в течение года равна $p = 0,006$. Ежегодный страховой взнос — \$120, выплата — \$10 000. Найти вероятность того, что

— страховая компания потерпит убыток;

— страховая компания получит прибыль больше \$500 000.

Задача 214. В страховом обществе застраховано 10000 лиц одного возраста и одной социальной группы. Вероятность смерти в течение года для каждого лица равна 0,006. Страховая выплата составляет \$10 000. Каким должен быть страховой взнос, чтобы с вероятностью 0,99 страховое общество получило прибыль не менее \$500 000?

Задача 215. Вероятность попадания в самолет из стрелкового оружия равна 0,001. По самолету было произведено 4000 выстрелов. Какова вероятность того, что он будет сбит, если для этого нужно 3 попадания?

Задача 216. Среднее число вызовов, поступающих на АТС в минуту, равно 120. Найти вероятности того, что

а) за две секунды не поступит ни одного вызова;
б) за две секунды поступит не менее двух вызовов;
в) за одну секунду поступит ровно три вызова;
г) за три секунды поступит не менее трех вызовов.

Задача 217. На столе площадью 0,5 квадратных метра сидит 600 мух. Какова вероятность убить 7 мух одним ударом полотенца площадью 40 квадратных сантиметров? Считать, что число убитых мух подчиняется распределению Пуассона.

Задача 218. Книга в 500 страниц содержит 250 опечаток. Какова вероятность того, что на случайно выбранной странице не менее 3 опечаток? Нет опечаток совсем? Считать, что число опечаток на странице подчиняется распределению Пуассона.

Задача 219. Получить прямой выкладкой формулы математического ожидания и дисперсии для распределения Пуассона.

Задача 220. Доказать, что сумма двух независимых случайных величин, распределенных по Пуассону с параметрами λ_1 и λ_2, также распределена по Пуассону с параметром $\lambda_1 + \lambda_2$.

Задача 221. В подготовленной к корректорской проверке книге число опечаток на странице есть случайная величина, подчиняющаяся распределению Пуассона с параметром λ. Корректор обнаруживает каждую опечатку независимо от других с вероятностью p. Определить закон распределения числа опечаток на странице после корректорской проверки.

Решение

Обозначим X — число опечаток на странице до корректуры, Y — после корректуры, $P(m,n)$ — вероятность того, что при корректуре на странице было найдено m опечаток, при условии, что их там было n. Тогда

$$P(Y=k) = P(X=k)\cdot P(0,k) + P(X=k+1)\cdot P(1,k+1) + \dots =$$

$$= \sum_{m=0}^{\infty} P(X=k+m)\cdot P(m,k+m) =$$

$$= \sum_{m=0}^{\infty} \frac{\lambda^{k+m} e^{-\lambda}}{(k+m)!}\cdot C_{k+m}^{m} p^m (1-p)^k =$$

$$= \sum_{m=0}^{\infty} \frac{\lambda^{k+m} e^{-\lambda}}{(k+m)!} \cdot \frac{(k+m)!}{k!\, m!} p^m (1-p)^k =$$

$$= \frac{\lambda^k (1-p)^k e^{-\lambda}}{k!} \sum_{m=0}^{\infty} \cdot \frac{\lambda^m p^m}{m!} =$$

$$= \frac{\lambda^k (1-p)^k e^{-\lambda}}{k!} \cdot e^{\lambda p} = \frac{(\lambda(1-p))^k e^{-\lambda(1-p)}}{k!}.$$

Ответ: Распределение Пуассона с параметром $\lambda(1-p)$.

16. Геометрическое распределение

Это — число испытаний Бернулли до первого успеха. При этом считается, что если успех наступил в первом же испытании, то случайная величина приняла значение 0.

Закон распределения:

$$P(X = k) = pq^k, \quad k = 0, 1, \ldots.$$

Как обычно, p — вероятность успеха, $q = 1 - p$.
Для этого распределения

$$MX = \frac{q}{p}; \quad DX = \frac{q}{p^2}.$$

Задача 222. Вероятность успеха в каждом из независимых испытаний равна 0,2. Сколько в среднем надо ждать первого успеха и какова вероятность того, что ждать придется меньше?

Решение

По формуле математического ожидания для геометрического распределения $MX = 4$. Это — число испытаний до первого успеха, поэтому первый успех в среднем будет в пятом испытании.

Вероятность того, что первый успех наступит раньше, равна

$$P = P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) =$$

$$= p + pq + pq^2 + pq^3 = 0,5904.$$

Этот ответ может на первый взгляд казаться парадоксальным: почему же в среднем первый успех будет в пятом испытании, если с вероятностью почти 60% успех наступит не позже четвертого? Дело тут в том, что математическое ожидание — это взвешенное среднее, а при усреднении будут вносить свой вклад длинные неудачные серии.

Задача 223. В некоторых учебниках предлагается несколько иное определение геометрического распределения: там учитывается то испытание, в котором наступил первый успех. Иными словами, не "до первого успеха", а "включая первый успех". Таким образом, возможные значения соучайной величины начинаются не с 0, а с 1. Как изменится при такой формулировке математическое ожидание и дисперсия?

Задача 224. Получить прямой выкладкой формулы математического ожидания и дисперсии для геометрического распределения.

Задача 225. Вероятность попадания в мишень постоянна и равна p. Стрелку выдают патроны до тех пор, пока он не промахнется. Пусть X — число выданных патронов. Найти математическое ожидание и дисперсию случайной величины X.

Задача 226. (О важности контрольного выстрела). Вероятность попадания постоянна и равна 0,8. Какова вероятность ни разу не попасть после одного выстрела? после двух выстрелов? после трех выстрелов?

Задача 227. Пусть случайные величины X и Y имеют геометрическое распределение. Доказать, что $\min(X, Y)$ также имеет геометрическое распределение.

Задача 228. Двое стреляют в мишень по очереди до первого попадания кого-то из них. Вероятность попадания первого при каждом выстреле равна 0,3, а второго — 0,5. Найти закон распределения числа потраченных патронов.

Задача 229. Стрелок стреляет в мишень до первого попадания. Вероятность попадания при одном выстреле равна p. У стрелка всего n патронов. Пусть X —случайная величина, означающая, сколько патронов будет потрачено. Найти закон распределения X и ее математическое ожидание.

Задача 230. Вася попадает в мишень с вероятностью 0,1. Какова вероятность того, что ему хватит 7 патронов для того, чтобы хотя бы один раз попасть? Какова вероятность, что ему хватит 10 патронов? Сколько патронов ему нужно дать, чтобы вероятность хотя бы одного попадания стала больше 0,98?

Задача 231. Коля пытается попасть снежком в окно третьего этажа. Спорщики Вася и Петя считают, что вероятность попадания при каждом броске одинакова и равна 0,2. Они поспорили, с какого раза Коля наконец попадет. По условиям спора Вася выигрывает, если Коля попадет строго с третьего раза, а Петя — если Коля попадет с шестого или седьмого раза. В остальных случаях — ничья. У кого больше шансов на победу — у Пети или Васи — и каковы эти шансы?

Задача 232. Вася стреляет в мишень до первого попадания. Вероятность попадания при одном выстреле равна 0,1. Организуется конкурс: сколько патронов потратит Вася? На что выгоднее всего поставить, если у Васи 10 патронов? Принимаются ставки также на то, что он ни разу не попадет.

Задача 233. Вася стреляет в мишень до первого попадания. Вероятность попадания при одном выстреле равна p. При каком значении p при заданном числе патронов n вероятности того, что он попадет с первого раза, и того, что он ни разу не попадет, будут равны?

Задача 234. Двое стреляют в мишень по очереди до первого попадания. Вероятность попадания при одном выстреле для первого равна p_1, а для второго p_2. Выигрывает тот, кто попадет первым. Найти вероятности победы для каждого участника.

Задача 235. Пусть в условиях предыдущей задачи p_1 известно. Чему должно быть равно p_2, чтобы шансы были равны?

17. Непрерывная случайная величина

Непрерывная случайная величина X — случайная величина, которая может принимать значения из интервала или даже на всей числовой оси. Такие величины уже нельзя охарактеризовать таблицей значений и вероятностей, поскольку вероятность каждого конкретного значения равна нулю, а смысл имеет только вероятность попадания в какой-то интервал. Для таких величин вводится функция $\rho(x)$, которая называется плотностью вероятностей. При этом должны выполняться следующие условия (аналогичные условиям для

дискретной случайной величины):
1. $\rho(x) \geqslant 0$.
2. $\int\limits_{-\infty}^{\infty} \rho(x)dx = 1$.

Вероятность попадания в интервал (a, b) при этом равна

$$P(a < X < b) = \int\limits_{a}^{b} \rho(x)dx.$$

При этом то, открытый интервал или замкнутый, то есть попадают ли в него точки a и b, — несущественно, поскольку вероятности того, что случайная величина примет значение a или b, равна нулю.

Математическое ожидание и дисперсия непрерывной случайной величины определяются следующим образом:

$$MX = \int\limits_{-\infty}^{\infty} x\rho(x)dx;$$

$$DX = M(X - MX)^2 = \int\limits_{-\infty}^{\infty} (x - MX)^2 \rho(x)dx.$$

Функция распределения случайной величины X — это функция

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

Функция рвспределения связана с плотностью вероятностей следующими формулами

$$F_X(x) = P(X < x) = \int\limits_{-\infty}^{x} \rho(t)dt;$$

$$\rho(x) = F_X'(x).$$

Задача 236. Плотность вероятности случайной величины задана формулой

$$\rho(x) = \frac{A}{1 + x^2}.$$

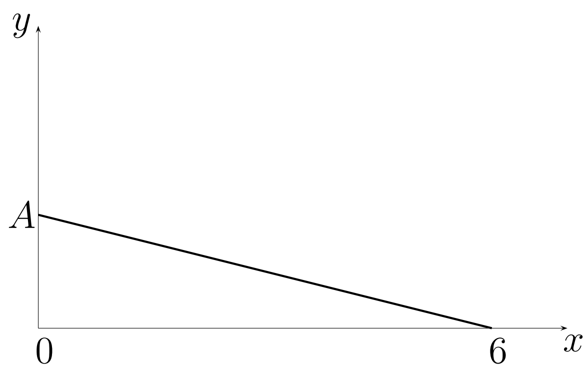
Найти константу A и вероятность того, что случайная величина примет значение больше 1.

Задача 237. Плотность вероятности случайной величины задана формулой

$$\rho(x) = \frac{A}{\sqrt{1-x^2}} \text{ при } x \in (-1, 1).$$

Найти константу A, математическое ожидание, дисперсию и вероятность того, что случайная величина примет значение, по модулю меньшее $1/2$.

Задача 238. Случайная величина отлична от нуля на интервале $(0, 6)$. График ее плотности вероятности изображен на рисунке.



Найти константу A, математическое ожидание, дисперсию и вероятность того, что случайная величина примет значение, меньшее 2.

Задача 239. Плотность вероятности случайной величины задана формулой

$$\rho(x) = A(4 - x^2) \text{ при } x \in (-2, 2).$$

Найти константу A, математическое ожидание, дисперсию и вероятность того, что случайная величина примет значение, по модулю меньшее 1.

Задача 240. Плотность вероятности случайной величины задана формулой

$$\rho(x) = A \sin x \text{ при } x \in (0, \pi).$$

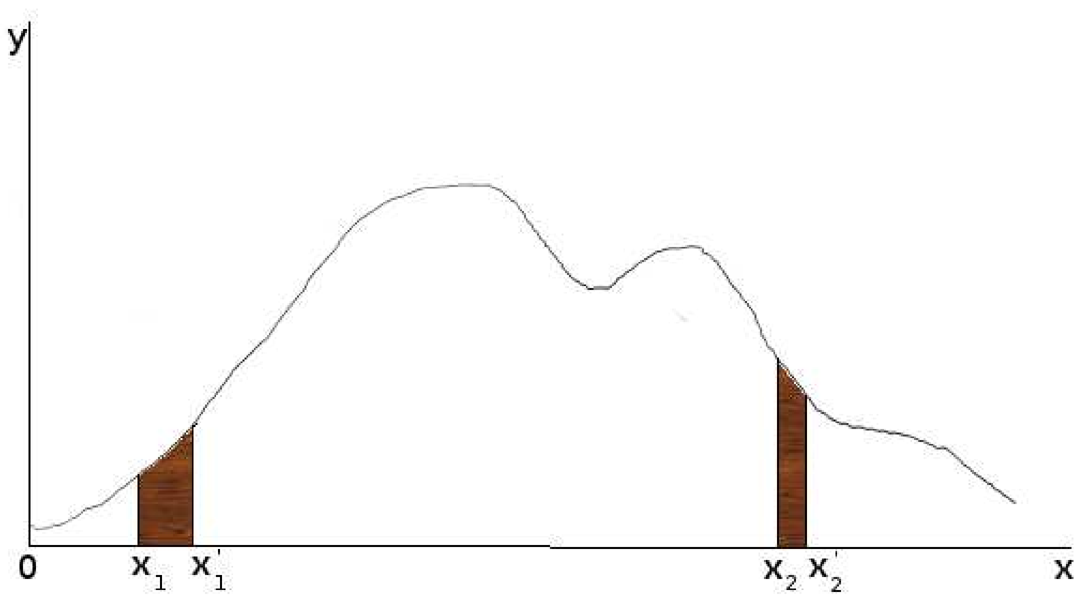
Найти константу A, математическое ожидание, дисперсию и вероятность того, что случайная величина примет значение, меньшее $\pi/2$.

Задача 241. Непрерывная функция $\rho(x)$ является плотностью вероятности случайной величины X. Число α принадлежит интервалу $(0, 1)$. Пусть x_1 и x_2

— такие числа, что $P(x_1 < X < x_2) = \alpha$, причем интервал (x_1, x_2) минимален. Верно ли, что $\rho(x_1) = \rho(x_2)$?

Решение

Равенство верно. Докажем это от противного. Пусть, для определенности $\rho(x_1) < \rho(x_2)$. Тогда, в силу непрерывности функции $\rho(x)$, можно найти такие точки x_1' и x_2', близкие к x_1 и x_2 соответственно (см. рисунок),



что выполнено следующее:

— интегралы от $\rho(x)$ по интервалам (x_1, x_1') и (x_2, x_2') равны (равны заштрихованные площади на рисунке);

— значения функции на интервале (x_1, x_1') меньше, чем на интервале (x_2, x_2') (тоже — как и изображено на рисунке).

Поскольку заштрихованные площади равны, интеграл от $\rho(x)$ по интервалу (x_1', x_2') также равен α. Но поскольку значения на левом интервале меньше, интервал (x_1, x_1') должен быть, разумеется, длиннее, чем интервал (x_2, x_2'). Следовательно, интервал (x_1', x_2') короче, чем исходный интервал (x_1, x_2). Это противоречит исходному предположению о минимальности интервала (x_1, x_2).

Задача 242. Троллейбусы одного маршрута продъезжают к остановке с интервалом ровно 5 минут, а другого — с интервалом ровно 10 минут. Пассажирам, подходящим к остановке в случайные моменты времени, подходят оба маршрута. Определить, каким может быть максимальное и минимальное среднее время ожидания.

Решение

Можно считать, что троллейбусы первого маршрута приходят в моменты времени $0, 5, 10, 15\ldots$, а троллейбусы второго маршрута — в моменты $t, t+10, t+20\ldots$, причем $0 \leqslant t \leq 5$. Тогда достаточно рассмотреть отрезок времени $[0, 10]$. Он разбивается на 3 отрезка:

1). Отрезок $[0, t]$. Вероятность попадания пассажира в этот отрезок равна $t/10$, а среднее время ожидания равно $t/2$.

2). Отрезок $[t, 5]$. Вероятность — $(5-t)/10$, а среднее время ожидания — $(5-t)/2$.

3). Отрезок $[5, 10]$. Вероятность — $1/2$, а среднее время ожидания — $5/2$.

Поэтому среднее время ожидания равно

$$t_{cp} = \frac{t}{10} \cdot \frac{t}{2} + \frac{5-t}{10} \cdot \frac{5-t}{2} + \frac{1}{2} \cdot \frac{5}{2} = \frac{1}{10}(t^2 - 5t + 25).$$

Минимальным среднее время ожидания будет при $t = 5/2$, тогда $t_{cp} = 15/8$. А максимальным среднее время ожидания будет при $t = 0$ и $t = 5$, тогда $t_{cp} = 5/2$.

Ответ: $15/8 \leq t_{cp} \leq 5/2$.

Задача 243. Пассажир подходит к остановке. Он точно знает, что троллейбус к остановке подходит с интервалом 10 минут, а автобус — с интервалом 15 минут. Пассажира устраивает любой из этих видов транспорта, но расписание движения он не знает. Найдите математическое ожидание случайной величины "время ожидания транспорта".

Решение

Вероятность того, что троллейбус придет раньше, чем через x минут, равна $x/10$ при $x < 10$. Вероятность того, что автобус придет раньше, чем через x минут, равна $x/15$ при $x < 15$. Вероятности того, что за время x троллейбус и автобус не придут равны соответственно $1 - x/10$ и $1 - x/15$. Вероятность того, что не придут оба, равна $(1 - x/10)(1 - x/15)$. Вероятность того, что хотя бы один из них все-таки придет, равна

$$F(x) = 1 - \left(1 - \frac{x}{10}\right)\left(1 - \frac{x}{15}\right) = \frac{x}{6} - \frac{x^2}{150}.$$

Это — функция распределения случайной величины "время ожидания транспорта". Плотность вероятности равна

$$\rho(x) = \frac{1}{6} - \frac{x}{75}.$$

Математическое ожидание равно

$$MX = \int\limits_0^{10} \left(\frac{x}{6} - \frac{x^2}{75}\right) dx = \frac{100}{12} - \frac{1000}{225} = \frac{35}{9}.$$

18. Равномерное распределение

Плотность вероятности для этого распределения равна константе внутри отрезка $[a, b]$ и нулю вне этого отрезка. Функция распределения линейна на отрезке, 0 — левее и 1 — правее отрезка.

Это — чисто теоретическое, модельное распределение. Хотя в теории его значение очень велико, в реальной жизни автору встречаться с ним не приходилось, только в абстрактных задачах. Программисты, однако, могут вспомнить стандартную функцию random() в C++, возвращающую случайное число, равномерно распределенное на $[0, 1]$.

Для этого распределения математическое ожидание и дисперсия равны соответственно

$$MX = \frac{a+b}{2}; \quad DX = \frac{(b-a)^2}{12}.$$

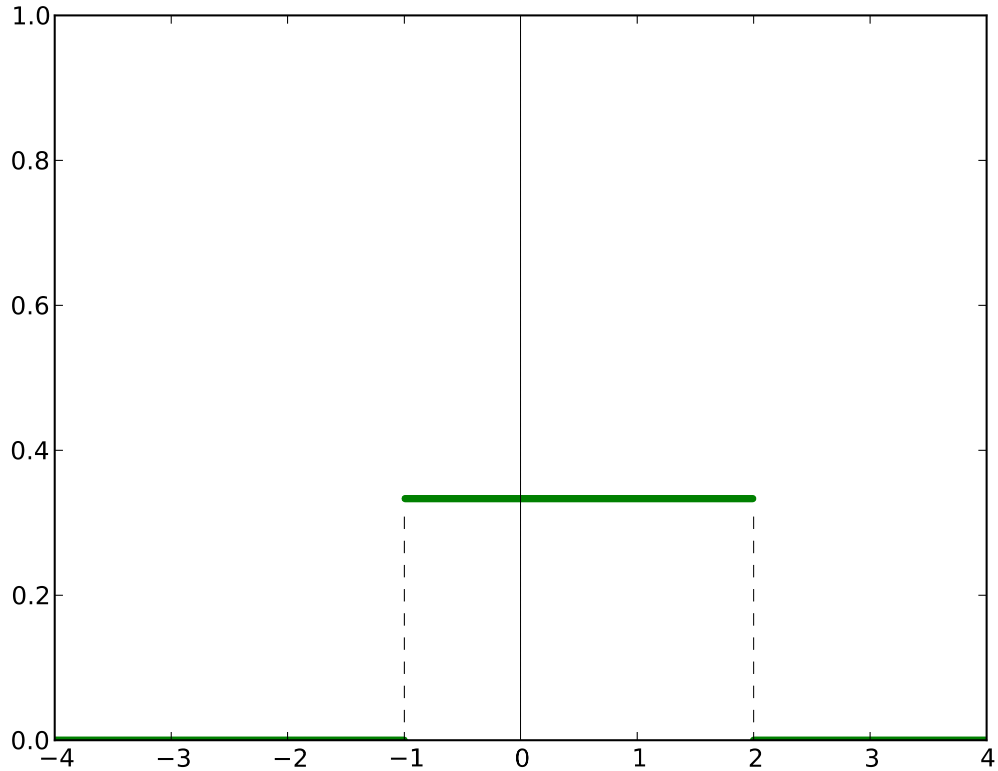


Рисунок 7. Плотность равномерного распределения на $[-1, 2]$.

Задача 244. Написать формулу плотности вероятностей для равномерного распределения.

Задача 245. Написать формулу функции распределения для равномерного распределения.

Задача 246. Доказать, что математическое ожидание равномерного распределения вычисляется по формуле

$$MX = \frac{a+b}{2}.$$

Задача 247. Доказать, что дисперсии равномерного распределения вычисляется по формуле

$$DX = \frac{(b-a)^2}{12}.$$

Задача 248. Случайные величины X и Y равномерно распределены на отрезках $[a, b]$ и $[c, d]$ соответственно, причем отрезки эти не пересекаются. Как распределена сумма $X + Y$?

Задача 249. Как и в предыдущей задаче, случайные величины X и Y равномерно распределены на отрезках $[a, b]$ и $[c, d]$ соответственно, но теперь $a < c < b < d$. Как распределена сумма $X + Y$?

Задача 250. На отрезке $[0; 1]$ случайным образом выбрали три точки, а затем расположили их по возрастанию: $x(1) < x(2) < x(3)$. Найти вероятность того, что средняя точка $x(2)$ лежит между $1/3$ и $1/2$.

Решение

Можно найти функцию распределения случайной величины $x(2)$. Событие $x(2) < x$ означает, что или

1) две точки лежат левее x, а третья — правее. Вероятность этого $C_3^2 x^2(1-x)$ (два успеха в трех испытаниях);

или

2) все три точки лежат левее x. Вероятность этого x^3. Поэтому функция распределения равна

$$F(x) = P(x(2) < x) = 3x^2(1 - x) + x^3 = 3x^2 - 2x^3.$$

А искомая вероятность равна

$$P = F(1/2) - F(1/3) = 3 \cdot \frac{1}{4} - 2 \cdot \frac{1}{8} - 3 \cdot \frac{1}{9} + 2 \cdot \frac{1}{27} = \frac{13}{54}.$$

Возможен и другой способ решения. Разделим отрезок на три: $[0; 1/3]$, $[1/3; 1/2]$, $[1/2; 1]$. Тогда благоприятные возможности таковы:

1). В каждый из отрезков попала ровно одна точка. Вероятность этого равна

$$p_1 = 3! \cdot \frac{1}{3} \cdot \frac{1}{6} \cdot \frac{1}{2} = \frac{1}{6}.$$

2). Во второй отрезок попало ровно 2 точки. Вероятность этого равна

$$p_2 = C_3^2 \cdot \left(\frac{1}{6}\right)^2 \cdot \frac{5}{6} = \frac{5}{72}.$$

3). Все три точки попали во второй отрезок. Вероятность этого равна

$$p_3 = \left(\frac{1}{6}\right)^3 = \frac{1}{216}.$$

Складывая найденные вероятности, получим тот же ответ.

Ответ: $13/54$.

19. Показательное, или экспоненциальное, распределение.

Плотность вероятностей для этого распределения равна нулю для отрицательных x, а для положительных — задается формулой

$$\rho(x) = \lambda e^{-\lambda x}$$

с некоторым положительным параметром λ. Функция распределения при $x > 0$ равна

$$F_X(x) = 1 - e^{-\lambda x}.$$

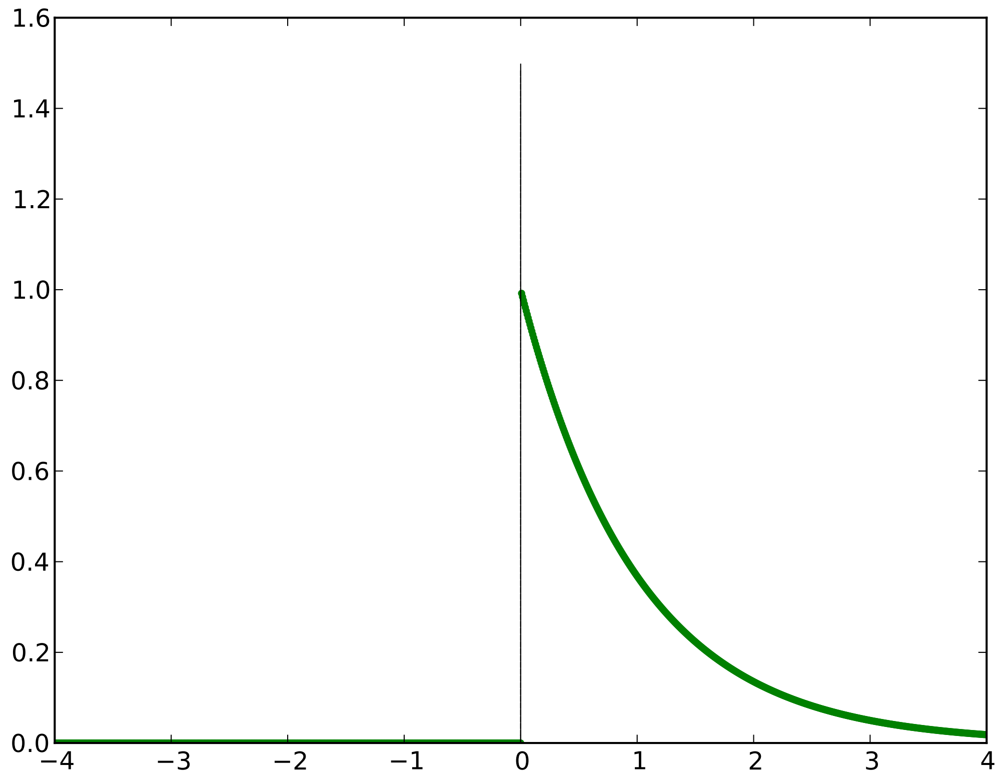


Рисунок 8. Плотность показательного распределения.

Это распределение встречается в теории массового обслуживания: для пуассоновского потока вызовов промежутки времени между последовательными вызовами имеют показательное распределение. Также оно применяется в теории надежности как распределение времен безотказной работы для каких-нибудь простых устройств или комплектующих.

Для этого распределения математическое ожидание и дисперсия равны соответственно

$$MX = \frac{1}{\lambda}; \quad DX = \frac{1}{\lambda^2}.$$

Задача 251. Средний срок службы газоразрядной лампы равен 1000 часов, причем срок службы имеет показательное распределение. Какова вероятность того, что лампа проработает менее 100 часов? более 2000 часов?

Задача 252. Агрегат состоит из трех параллельно подключенных дублирующих друг друга устройств, срок службы каждого из которых имеет показательное распределение со средним значением 1000 часов. Какова вероятность того, что агрегат проработает не менее 3000 часов?

Задача 253. Агрегат состоит из двух параллельно подключенных дублирующих друг друга устройств, срок службы каждого из которых имеет показательное распределение со средним значением 100 часов. Какова вероятность того, что агрегат проработает не менее 300 часов?

Задача 254. Время безотказной работы прибора является случайной величиной, распределенной по показательному закону со средним значением 4000 часов. Какова вероятность того, что прибор проработает менее 3000 часов? более 5000 часов?

Задача 255. Работоспособность некоего агрегата определяется работоспособностью четырех дублирующих устройств. Наработка на отказ (среднее время работы) каждого из этих трех устройств равна 100 часов. Весь агрегат выходит из строя, если отказали все четыре этих устройства. Какова вероятность того, что весь агрегат проработает не менее 400 часов? Считать, что время работы каждого из устройств подчинено показательному распределению.

Задача 256. Поток событий называется пуассрновским, если число происшедших за любой интервал времени событий имеет распределение Пуассона с параметром, пропорциональным длительности этого интервала, причем для непересекающихся интервалов эти случайные величины независимы. Доказать, что промежуток между двумя последовательными событиями является случайной величиной, имеющей показательное распределение.

Решение

Пуассоновский поток — одно из ключевых понятий теории массового обслуживания. Число событий за время t имеет распределение Пуассона со средним $\mu(t) = \lambda t$. Параметр λ называется интенсивностью потока. Корректность этого определения следует из результата решения задачи 220.

Пусть T — случайная величина, равная интервалу между последовательными событиями. Найдем ее функцию распределения. Событие $(T > t)$ означает, что за время t не произошло ни одного события. В соответствии с формулой распределения Пуассона

$$P(T > t) = \frac{(\lambda t)^0}{0!} e^{-\lambda t} = e^{-\lambda t}.$$

Следовательног, функция распределения

$$P(T < t) = 1 - e^{-\lambda t}.$$

Мы получиили функцию распределения для показательного распределения.

Задача 257. Автобусы подходят к остановке не по графику, а в соответствии с законом пуассоновского потока с интенсивностью λ. Чему равно среднее время ожидания?

Задача 258. Доказать, что показательное распределение обладает свойством “отсутствия памяти”. Пусть срок службы некоторого устройства подчиняется показательному распределению с параметром λ. Тогда срок службы этого устройства начиная с момента t_0, при условии, что оно уже проработало не менее t_0, также подчиняется показательному распределению с тем же параметром.

20. Квантили

Для различных распределений часто возникает задача нахождения такого значения x_0, что вероятность того, что случайная величина примет значение, меньшее x_0, равно заданному числу α. Такое значение называется α-квантилем распределения (ударение на “и”).

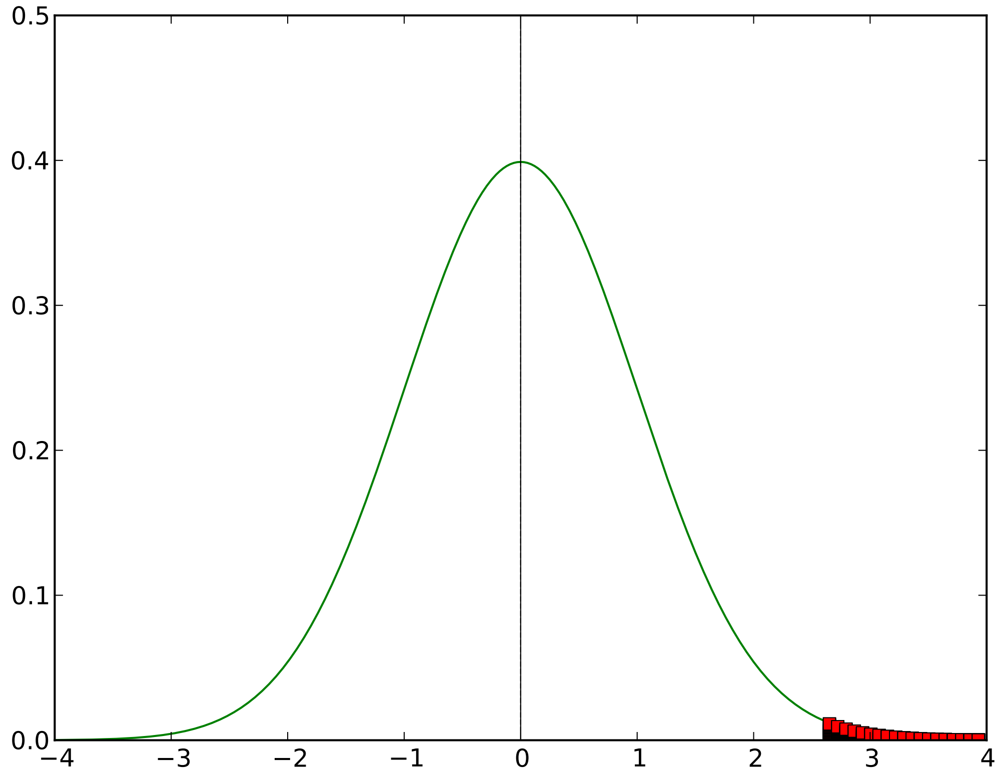


Рисунок 9. Квантиль.

Таким образом, α-квантиль распределения — это решение уравнения

$$\int_{-\infty}^{x} \rho(t)dt = \alpha.$$

Если же воспользоваться понятием функции распределения, то можно записать такое уравнение

$$F_X(x) = \alpha.$$

Слово "квантиль" обычно относят к мужскому роду: так предписывает, например, сайт *gramota.ru* или действующий в России ГОСТ Р 50779.10-2000 "Вероятность и основы статистики. Термины и определения". Однако в математической литературе это слово часто встречается в женском роде. Это — один из примеров профессионального жаргона.

Медианой называется 0,5-квантиль расспределения. Иными словами, медиана случайной величины X — это такое число m, что вероятности событий $(X < m)$ и $(X > m)$ равны 0,5. Это — одна из важных характеристик случайной величины. Ее иногда называют средним значением, или, точнее, медианным средним, поскольку она не всегда совпадает с математическим ожиданием.

Квартиль — это 0,25- или 0,75-квантиль. Разность между квартилями называется интерквартильным размахом и является, наряду со стандартным отклонением, характеристикой разброса распределения.

Используется также термин *дециль*. Это 0,1- или 0,9-квантиль. В словах квартиль и дециль также ударение на "и".

Для многих распределений для нахождения квантилей используются специальные таблицы. Стандартные функции для нахождения квантилей есть практически во всех распространенных электронных ьаблицах, таких как Microsoft Excel или Libre Office Calc.

Задача 259. Найти медиану показательного распределения с параметром λ.

Задача 260. Плотность вероятности случайной величины задана формулой

$$\rho(x) = \frac{1}{\pi(1+x^2)}.$$

Найти квартили распределения и интерквартильный размах.

Задача 261. Плотность вероятности случайной величины задана формулой

$$\rho(x) = \frac{1}{\pi\sqrt{1-x^2}} \text{ при } x \in (-1, 1).$$

Найти квартили распределения и интерквартильный размах.

Задача 262. Плотность вероятности случайной величины задана формулой

$$\rho(x) = A(4-x^2) \text{ при } x \in (-2, 2).$$

Найти квартили распределения и интерквартильный размах.

Задача 263. Плотность вероятности случайной величины задана формулой

$$\rho(x) = A \sin x \text{ при } x \in (0, \pi).$$

Найти квартили распределения и интерквартильный размах.

Задача 264. Найти медиану распределения из задачи 238.

Задача 265. Найти 0,05- и 0,95-квантили показательного распределения с параметром λ.

Задача 266. Найти квартили распределения и интерквартильный размах показательного распределения с параметром λ.

21. Нормальное распределение

Плотность вероятностей для нормального распределения задается формулой

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

Для этого распределения

$$MX = a, \quad DX = \sigma^2.$$

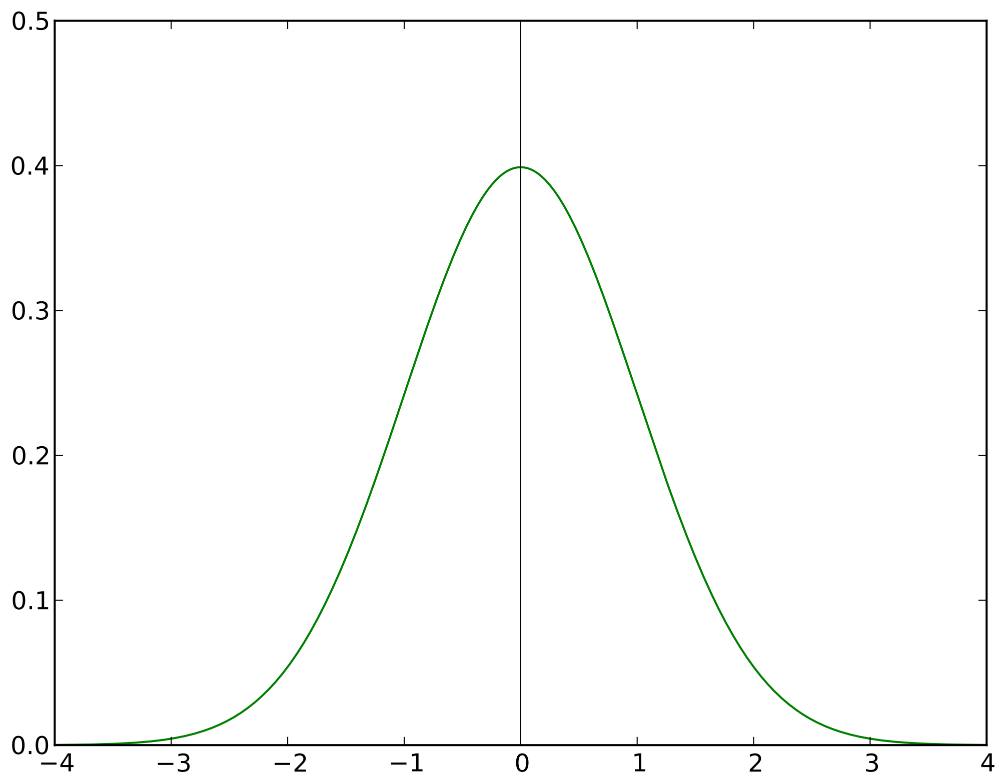


Рисунок 10. Плотность нормального распределения.

Стандартным нормальным распределением называется такое нормальное распределение, у которого среднее равно 0, а дисперсия равна 1. Плотность вероятности для него равна

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

Функция распределения равна интегралу от этой плотности вероятности. Этот интеграл не берется в элементарных функциях, но встречается часто, и поэтому удостоился специального названия. Функция

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{x} e^{\frac{-t^2}{2}} dt$$

часто называется интегралом вероятностей, или интегралом ошибок.

Значения этой функции приводятся в специальных таблицах. При работе с электронными таблицами, в частности, в Microsoft Excel, можно использовать стандартную функцию СТНОРМРАСП, а в Libre Office Calc — функцию NORMSDIST. Однако, названия этих функций могут быть разными в различных версиях.

Во многих версиях электронных таблиц используется параметр "ИНТЕГРАЛЬНЫЙ". Это логический параметр, принимающий значения 0 и 1. Для вычисления функции $\Phi(x)$ следует указать, что он равен 1, иначе будет вывчислено не значение интеграла, а значение подынтегральной функции.

В таблицах, приведенных в различных учебниках, приводятся значения интеграла ошибок только для положительных x. Для отрицательных x из того, что под интегралом стоит четная функция и из того, что $\Phi(0) = 1/2$, следует формула $\Phi(-x) = 1 - \Phi(x)$.

Во многих задачах, как мы увидим чуть позже, требуется находить квантили распределений по заданным вероятностям α. В Microsoft Excel, можно использовать стандартную функцию СТНОРМОБР, а в Libre Office Calc — функцию NORMSINV. Опять-таки, названия этих функций могут различаться в различных версиях.

Для нормального распределения верна такая теорема. Если случайная величина X имеет нормальное распределение со средним a и дисперсией σ^2, то случайная величина

$$\frac{X-a}{\sigma}$$

имеет стандартное нормальное распределение.

Эта теорема позволяет пользоваться таблицами только для стандартного нормального распределения.

Задача 267. Найти вероятность того, что случайная величина, нормально распределенная со средним $a = 8$ и дисперсией $\sigma^2 = 16$, принимает значения из интервала от 0 до 12.

Решение.

Обозначим искомую вероятность через p и запишем это формулой:

$$p = P(0 < X < 12).$$

Далее преобразуем эту формулу:

$$p = P\Big(\frac{0-8}{4} < \frac{X-a}{\sigma} < \frac{12-8}{4}\Big);$$

$$p = P\Big(-2 < \frac{X-a}{\sigma} < 1\Big).$$

Собственно, мы ничего не сделали, только вычли среднее и разделили на σ. Но теперь в центре двойного неравенства стоит случайная величина, имеющая стандартное нормальное распределение, и можно пользоваться таблицей. Окончательно получим

$$p = \Phi(1) - \Phi(-2) = \Phi(1) + \Phi(2) - 1 = 0,8413 + 0,9772 - 1 = 0,8385.$$

Задача 268. Найти вероятность того, что нормально распределенная случайная величина отклонится от среднего больше, чем на 3σ.

Решение.

Проще найти вероятность противоположного события.

$$P(|X-a| < 3\sigma) = P(-3\sigma < X - a < 3\sigma) =$$

$$= P(-3 < \frac{X-a}{\sigma} < 3) = \Phi(3) - \Phi(-3) = 2\Phi(3) - 1 = 0,9974.$$

Таким образом, вероятность отклонений, больших 3σ, составляет примерно 0,26%. На этом основано часто применяемое правило трех сигм. Если мы, скажем, проводим какие-то измерения, то мы можем столкнуться с грубыми промахами. Как узнать, является ли данный результат измерения грубым промахом? Если оно дальше от среднего, чем три сигмы, то да, а если ближе, то нет.

Задача 269. В каких пределах с заданной вероятностью $1-\alpha$ будет находиться значение нормально распределенной случайной величины с заданными средним a и дисперсией σ^2? Найти минимальный из таких интервалов.

Решение.

Будем искать симметричный интервал $(a-z, a+z)$ — он будет минимальным по длине. Напишем уравнение.

$$P(a-z<X<a+z)=1-\alpha.$$

Далее действуем уже традиционным способом:

$$P\Big(-\frac{z}{\sigma}<\frac{X-a}{\sigma}<\frac{z}{\sigma}\Big)=1-\alpha.$$

$$2\Phi\Big(\frac{z}{\sigma}\Big)-1=1-\alpha.$$

$$\Phi\Big(\frac{z}{\sigma}\Big)=1-\alpha/2.$$

Обозначим через $t_{1-\alpha/2}$ соответствующий квантиль стандартного нормального распределения, и получим выражение для искомого интервала:

$$X\in(a-\sigma\cdot t_{1-\alpha/2}, a+\sigma\cdot t_{1-\alpha/2}) \text{ с вероятностью } 1-\alpha.$$

Это выражение, а также его аналоги, мы не раз увидим при рассмотрении задач на проверку статистических гипотез. Кроме того, аналогичные выражения часто встречаются в ГОСТах и других нормативных документах, практически всегда — без всяких пояснений. Тот, кто освоил курс теории вероятностей, должен понимать, что имеется в виду.

Задача 270. Автомат выпускает детали, размер которых представляет собой нормально распределенную случайную величину со средним 81 мм и стандартным отклонением 3 мм. Деталь считается годной, если ее размер от 78 до 82 мм, в противном случае она бракуется. Каков процент брака у автомата?

Задача 271. Случайная величина X имеет нормальное распределение. Найти такое z, что $P(X<z)=p$, если известны матожидание и дисперсия.

a) MX = 7,   DX = 9;
b) MX = 3,   DX = 4;
C) MX = 1,   DX = 16.

Задача 272. Случайная величина X имеет нормальное распределение. Найти дисперсию, если известно MX и то, что с вероятностью p $X < b$.
a) $MX = 1, \;\; P(X < 3) = 0,99$;
b) $MX = 5, \;\; P(X < 21) = 0.95$;
C) $MX = 8, \;\; P(X < 2) = 0.01$.

Задача 273. Найти отношение интерквартильного размаха к стандартному отклонению для нормального распределения. в частности, доказать, что это отношение не зависит ни от a, ни от σ^2.

Задача 274. Случайная величина X имеет нормальное распределение с известными MX и DX. Найти,с какой вероятностью она примет значения из заданного интервала:
a) $MX = 1, \;\; DX = 9,$ найти $P(0 < X < 8)$;
b) $MX = 5, \;\; DX = 4,$ найти $P(3 < X < 6)$;
C) $MX = 8, \;\; DX = 16$ найти $P(5 < X < 12)$.

Задача 275. Случайная величина X имеет нормальное распределение Дана вероятность P того, что эта случайная величина отклонится от MX не более, чем на заданную величину. Найти дисперсию случайной величины X.
a) $P(|X - MX| < 10) = 0,95$;
b) $P(|X - MX| < 5) = 0,99$;
c) $P(|X - MX| < 2) = 0,98$.

Задача 276. Случайная величина X имеет нормальное распределение Дана дисперсия DX этой случайной величины. Найти, какого максимального значения не превысит отклонение $|X - MX|$ с заданной вероятностью P.
a) $DX = 1, P = 0,95$;
b) $DX = 9, P = 0,99$;
c) $DX = 25, P = 0,98$.

22. Формулы математического ожидания и дисперсии

Математическое ожидание — усредненное значение случайной величины

Для дискретной случайной величины оно равно

$$MX = x_1p_1 + x_2p_2 + \cdots + x_np_n,$$

а для непрерывной:

$$MX = \int\limits_{-\infty}^{\infty} x\rho(x)dx;$$

Дисперсия характеризуют разброс случайной величины относительно среднего. Она равна по определению

$$DX = M(X - MX)^2.$$

Для непрерывной случайной величины можно записать

$$DX = \int\limits_{-\infty}^{\infty} (x - MX)^2\rho(x)dx.$$

Стандартное (среднеквадратическое) отклонение — квадратный корень из дисперсии

$$\sigma(X) = \sqrt{DX}.$$

Для оценки масштаба уклонения от среднего надо использовать именно стандартное отклонение, потому что оно имеет ту же размерность, что и случайная величина, а размерность дисперсии — квадрат размерности случайной величины.

Задача 277. Доказать, что справедливы формулы (c — константа)
$M(c) = c$
$M(aX) = a \cdot MX$;
$D(c) = 0$;
$D(X + c) = DX$;
$D(aX) = a^2 \cdot DX$;
$DX = MX^2 - (MX)^2$.

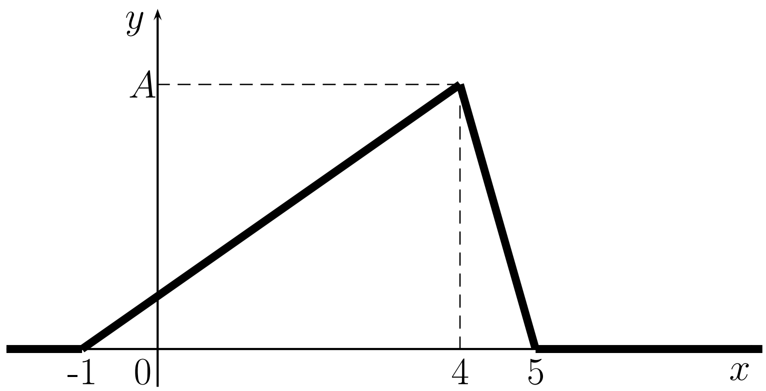
Задача 278. Доказать, что для любых случайных величин X и Y справедлива формула $M(X + Y) = MX + MY$.

Задача 279. Доказать, что для независимых случайных величин X и Y справедливы формулы $M(XY) = MX \cdot MY$;
$D(X + Y) = DX + DY$.

Задача 280. $DX = 4$, найти $D(X+3)$.

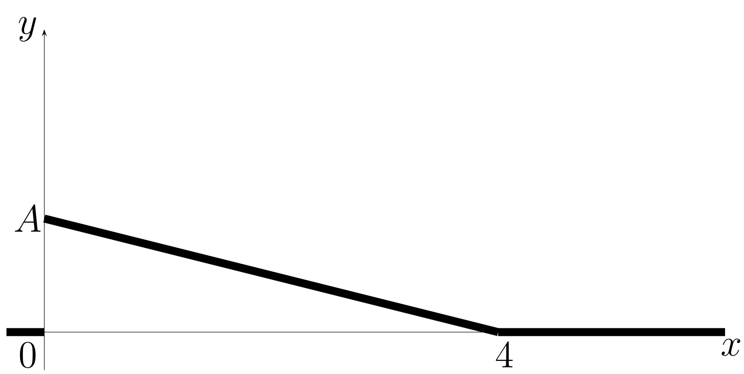
Задача 281. Случайные величины X и Y независимы, $DX = 4$, $DY = 3$. Найти $D(X - Y)$.

Задача 282. На рисунке изображен график плотности вероятностей случайной величины.



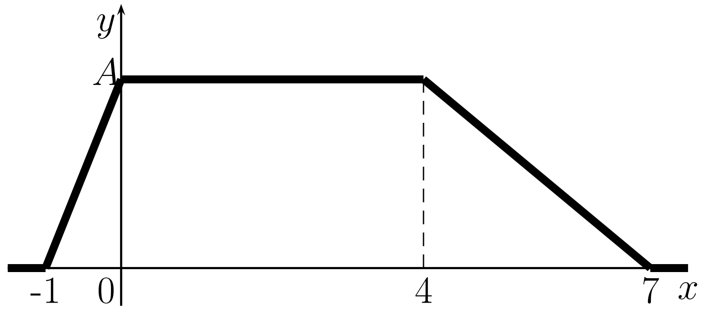
Найти константу A, математическое ожидание, дисперсию и вероятность того, что случайная величина примет значение, меньшее 2.

Задача 283. Случайная величина отлична от нуля на интервале $(0, 4)$. График ее плотности вероятности изображен на рисунке.



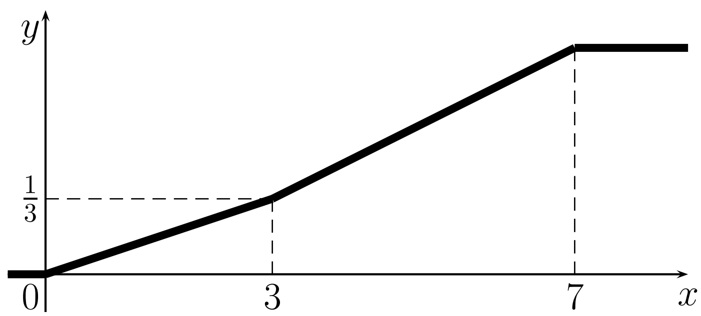
Найти константу A, математическое ожидание, дисперсию, медиану и вероятность того, что случайная величина примет значение, большее 2,4.

Задача 284. На рисунке изображен график плотности вероятностей случайной величины.



Найти константу A, математическое ожидание, дисперсию, медиану и вероятность того, что случайная величина примет значение, меньшее 2.

Задача 285. На рисунке представлен график функции распределения случайной величины.



Найти ее математическое ожидание, дисперсию, медиану и вероятность того, что случайная величина примет значение, меньшее 5.

Задача 286. Случайные величины $X_1, \ldots, X_n$ — результаты независимых измерений некоторой величины. Все измерения равноточные и характеризуются некоторой точностью (стандартным отклонением) σ. Для определения значения измеряемой величины используется среднее арифметическое

$$X = \frac{X_1 + \ldots + X_n}{n}.$$

Найти стандартное отклонение случайной величины X.

Задача 287. Случайные величины $X_1, \ldots, X_n$ — результаты независимых измерений некоторой величины. При этом измерения неравноточные — i-е измерение характеризуются точностью (стандартным отклонением) σ_i. Для определения значения измеряемой величины используется линейная комбинация данных измерений

$$X = w_1 X_1 + \ldots + w_n X_n$$

с некоторыми весами $w_1, \ldots, w_n$. Определить, при каких значениях весов точность (стандартное отклонение) величины X будет минимальным.

Задача 288. Угол A недоступен для непосредственного измерения. Чтобы измерить этот угол, его включили в пятиугольник $ABCDE$, где остальные углы можно измерить (в геодезии применяется термин *геодезический ход*). Точность измерения углов B, C, D и E составляет 1°. Какова точность определения угла A?

Задача 289. Третьим центральным моментом, или коэффициентом асимметрии непрерывного распределения с плотностью $\rho(x)$ и математическим ожиданием MX называется

$$M_3 = \int\limits_{-\infty}^{+\infty} (x - MX)^3 \rho(x) dx.$$

Чему равен коэффициент асимметрии для распределения из задачи 238?

Задача 290. Четвертым центральным моментом распределения с плотностью $\rho(x)$ и математическим ожиданием MX называется

$$M_4 = \int\limits_{-\infty}^{+\infty} (x - MX)^4 \rho(x) dx.$$

Чему равен этот коэффициент для распределения из задачи 238?

Задача 291. Коэффициентом эксцесса распределения называют исправленный четвертый центральный момент: $\varepsilon = M_4 - \mu$, где число μ подобрано так, чтобы эксцесс стандартного нормального распределения был бы равен 1. Чему равно μ?

Задача 292. Чему равны коэффициент асимметрии и коэффициент эксцесса для показательного распределения со средним $1/\lambda$?

23. Теорема Муавра — Лапласа

На “бытовом языке” теорему Муавра — Лапласа можно сформулировать так: при больших n число успехов в n испытаниях Бернулли имеет приблизительно нормальное распределение.

Приведенный ниже рисунок иллюстрирует теорему Муавра — Лапласа. Действительно, при больших n распределение числа успехов в n испытаниях похожн на нормальное.

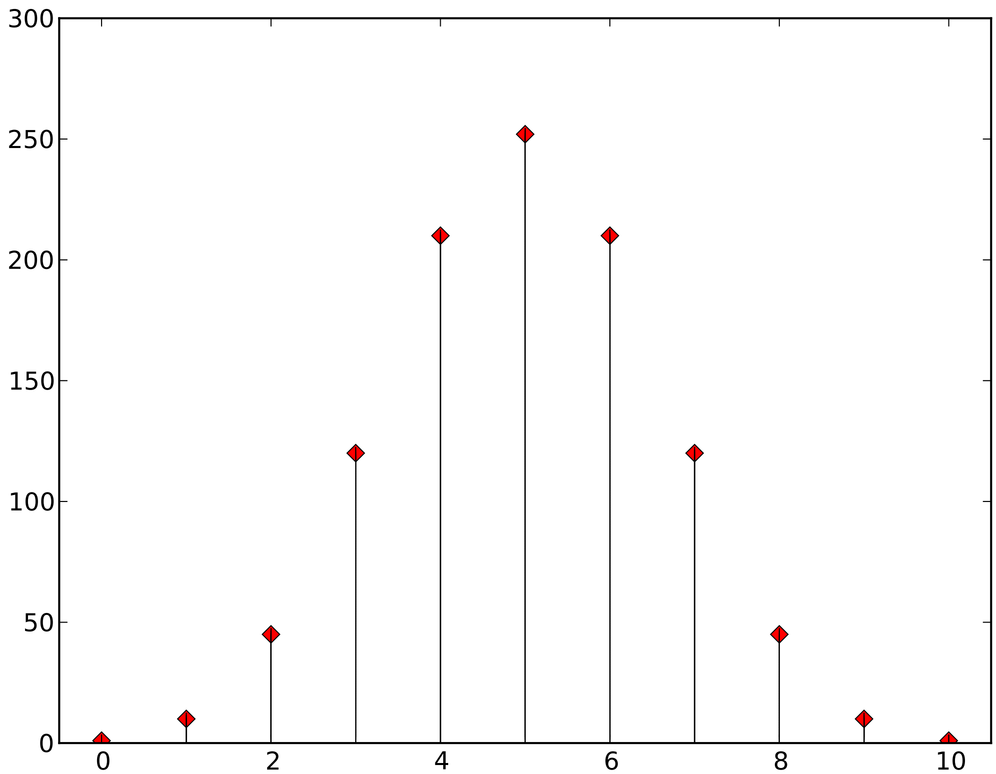


Рисунок 11. Распределение Бернулли с $n = 10$, $p = 1/2$.

В большинстве учебников приводится строгая формулировка, которую понять довольно трудно. Например, формулировка наиболее важного частного случая (называемого интегральной теоремой Лапласа) в классическом учебнике Гмурмана [4] выглядит так.

Теорема. Если вероятность p наступления события A в каждом испытании постоянна и отлична от нуля и единицы, то вероятность $P_n(k_1, k_2)$ того, что событие A появится в n независимых испытаниях от k_1 до k_2 раз, приближенно равна определенному интегралу

$$P_n(k_1, k_2) = \frac{1}{\sqrt{2\pi}} \int\limits_{x_1}^{x_2} e^{-\frac{t^2}{2}} dt,$$

где $x_1 = (k_1 - np)/\sqrt{npq}$ и $x_2 = (k_2 - np)/\sqrt{npq}$.

Чтобы понять, в чем суть дела, разберем использование этой теоремы на примере нескольких задач.

Задача 293. Вероятность заболеть гриппом во время эпидемии равна 0,4. Какова вероятность того, что заболевших на предприятии, на котором работают 600 человек, будет от 210 до 270?

Решение.

Пусть k — число “успехов”, то есть случайная величина, обозначающая число заболевших. Мы знаем, что эта случайная величина имеет распределение Бернулли. Собственно, этим можно воспользоваться: подсчитать для каждого из 61 значений (от 210 до 270) его вероятность, затем эти вероятности просуммировать... Дел хватит надолго.

Вместо этого воспользуемся теоремой Муавра — Лапласа. Нам нужно оценить вероятность

$$P = P_{600}(210, 270) = P(210 \leq k \leq 270).$$

На основании “бытовой” формулировки поступим с этой вероятностью так, как если бы k имело нормальное распределение:

$$P = P\Big(\frac{210 - np}{\sqrt{npq}} \leq \frac{k - np}{\sqrt{npq}} \leq \frac{270 - np}{\sqrt{npq}}\Big).$$

Математическое ожидание и дисперсия для распределения Бернулли нам известны, и для нашего примера легко подсчитать, что

$$np = 240; \quad npq = 144; \quad \sqrt{npq} = 12.$$

Поэтому

$$P = P\Big(\frac{210 - 240}{12} \leq \frac{k - np}{\sqrt{npq}} \leq \frac{270 - 240}{12}\Big) =$$

$$= P(-2,5 \leq \frac{k - np}{\sqrt{npq}} \leq 2,5).$$

Дробь в середине этого двойного неравенства имеет, как нетрудно догадаться, стандартное нормальное распределение. Поэтому осталось просто воспользоваться таблицей:

$$P = \Phi(2,5) - \Phi(-2,5) = 2 \cdot 0,9938 - 1 = 0,9876.$$

Воспользовавшись формулой Бернулли и электронными таблицами, например Microsoft Excel или Libre Office Calc, можно найти точное значение этой вероятности. Оно равно 0,9890. Если теорему Муавра — Лапласа использовать для оценки, а не для точного нахождения вероятностей, то точность вполне приемлема.

При рассмотрении этого примера мы не обращали внимания на такой аспект: входят ли граничные точки, то есть 210 и 270? Оказывается, это неважно. Вероятность каждого значения близка к 0, а теорема Муавра — Лапласа испоьзуется не для точного подсчета, а для приближенной оценки вероятности.

Точность на самом деле можно и повысить. Мы заменяем дискретное распределение Бернулли непрерывным нормальным распределением. Тогда получается, что мы каждое целое значение заменяем отрезком, и на каждое целое значение будет приходиться отрезок длины 1. Поэтому каждому возможному значению k из распределения Бернулли мы можем поставить в соответствие интервал $[k-1/2, k+1/2]$ для нормального распределения. Тогда для оценки надо брать не интервал $[210, 270]$, а интервал $[209,5, 270,5]$. Проведя вычисления с этим, более точным, интервалом, мы получим значение вероятности 0,9890, то есть с точностью до четырех знаков совпадающее с верным значением.

Описанный выше прием — отнесение в соответствие каждому дискретному значению интервала длины 1 — иногда называют *поправкой на дискретность.*

Задача 294. Театр вместимостью 1000 мест имеет два входа, при каждом из них есть гардероб. Какова должна быть вместимость каждого из этих гардеробов, чтобы все зрители с вероятностью 0,99 могли раздеться в гардеробе того входа, через который вошли? Предполагается, что каждый из входов зрители выбирают с вероятностью 0,5 независимо друг от друга, а театр каждый день полон.

Решение.

Мы имеем дело с испытаниями Бернулли с параметрами: $n = 1000$, $p = 1/2$. Требуется узнать, в каких пределах с вероятностью 0,99 лежит число успехов, то есть число людей, выбравших первый, скажем, вход. Можно это записатьв виде уравнения с неизвестным z:

$$P(np - z < X < np + z) = 0.99.$$

Двойное неравенство здесь потому, что нас не устраивает как слишком большое, так и слишком малое число "успехов": так как театр полон, то малое число посетителей у одного гардероба означает перегрузку второго.

Дальнейшее просто:

$$P\Big(-\frac{z}{\sqrt{npq}} < \frac{X - np}{\sqrt{npq}} < \frac{z}{\sqrt{npq}}\Big) = 0,99;$$

$$2\Phi\Big(\frac{z}{\sqrt{npq}}\Big) - 1 = 0,99;$$

$$\Phi\Big(\frac{z}{\sqrt{npq}}\Big) = 0,995;$$

$$\frac{z}{\sqrt{npq}} = 2,58; z = 2,58 \cdot \sqrt{250} \approx 41.$$

Следовательно, в гардеробе надо предусмотреть 541 место.

Задача 295. В поселке живут 180 человек. Каждый из них пять раз в месяц ездит на автобусе в город, выбирая дни поездок случайным образом независимо от других. Какой вместимости должен быть автобус, чтобы он не переполнялся в 99% случаев? Считать, что в месяце 30 дней.

Решение.

И вновь мы имеем дело с испытаниями Бернулли, на сей раз с параметрами $n = 180$, $p = 1/6$. Но, в отличие от предыдущего примера, здесь нам надо найти несимметричный интервал с заданной вероятностью: если автобус будет недогружен — ничего страшного. Поэтому можно записать такое уравнение: $P(X < z) = 0,99$. Решение его аналогично:

$$P\Big(\frac{X - np}{\sqrt{npq}} < \frac{z - np}{\sqrt{npq}}\Big) = 0,99;$$

$$\Phi\Big(\frac{z-np}{\sqrt{npq}}\Big)=0,99; \frac{z-np}{\sqrt{npq}}=2,33;$$

$$z=np+2,33\cdot\sqrt{npq}=30+2,33\cdot\sqrt{25}\approx 42.$$

Задача 296. Вероятность наступления события в каждом из одинаковых и независимых испытаний равна 0,8. Найти вероятность того, что в 225 испытаниях событие наступит не менее 170 и не более 185 раз.

Задача 297. Вероятность наступления события в каждом из независимых испытаний равна 0,2. Найти вероятность того, что в 100 испытаниях событие наступит не менее 20 раз и не более 30 раз.

Задача 298. Вероятность появления события в каждом из 2100 независимых испытаний равна 0,7. Найти математическое ожидание числа появлений события и вероятность того, что событие появится не менее 1440 и не более 1500 раз.

Задача 299. Вероятность безотказной работы некоторого устройства в течение определенного срока равна 0,8. В каких пределах лежит количество отказов в партии из 1000 штук, если эти пределы должны быть гарантированы с вероятностью 0,98?

Задача 300. Вероятность появления события в каждом из независимых испытаний равна 0,8. Сколько нужно произвести испытаний, чтобы с вероятностью 0,9 можно было ожидать, что событие появится не менее 75 раз?

Задача 301. Отдел технического контроля проверяет 475 изделий на брак. Вероятность того, что изделие бракованное, равно 0,05. Найти с вероятностью 0,9426 границы, в которых будет заключено число бракованных изделий среди проверенных.

Задача 302. Театр вместимостью 1000 мест имеет два входа, при каждом из них есть гардероб. Какова должна быть вместимость каждого из этих гардеробов, чтобы все зрители с вероятностью 0,99 могли раздеться в гардеробе того входа, через который вошли? Предполагается, что каждый из входов зрители выбирают с вероятностью 0,5 независимо друг от друга, а театр каждый день полон. В задаче 294 рассмотрен случай, когда зрители приходят поодиночке. Что изменится, если зрители будут приходить парами, и сколько тогда понадобится мест?

Задача 303. Вероятность прохождения контроля качества для некоторого экспериментального изделия равна 0,6. Сколько надо выпустить таких изделий, чтобы с вероятностью 0,99 среди них нашлось бы не менее 100 прошедших контроль качества?

Задача 304. На фирме работают 900 сотрудников. Администрация предполагает организовать на новогодние каникулы коллективный отдых сотрудников в Египте. Для этого планируется зафрахтовать авиалайнер, а билеты на оставшиеся свободными места продать на сторону. Допустим, что вероятность того, что каждый отдельно взятый сотрудник желает лететь в Египет, равна 0,2. Сколько должно быть посадочных мест в лайнере, чтобы с вероятностью 0,99 никто из желающих лететь сотрудников не остался без места?

Задача 305. Оценить вероятность $P(X = k)$ каждого отдельного значения для распределения Бернулли.

Решение

Как и в задаче 293, каждому возможному значению k из распределения Бернулли будет соответствовать интервал $[k - 1/2, k + 1/2]$ для нормального распределения. Тогда

$$P_n(k) \approx P(k - 1/2 < X < k + 1/2) =$$
$$= P\Big(\frac{k - np - 1/2}{\sqrt{npq}} < \frac{X - np}{\sqrt{npq}} < \frac{k - np + 1/2}{\sqrt{npq}}\Big).$$

Вероятность попадания в интервал, как мы знаем, равна интегралу от плотности вероятностей. Здесь же мы можем считать, что интервал маленький, и вероятность попадания в него приблизительно равна значению плотности вероятностей в середине интервала, умноженной на длину интервала. Окончательно получим:

$$P_n(k) \approx \frac{1}{\sqrt{npq}} \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \text{ где } x = \frac{k - np}{\sqrt{npq}}.$$

Этот результат носит название *локальной теоремы Муавра — Лапласа.* Как видно, вероятность каждого конкретного значения действительно стремится к 0 с ростом n.

Задача 306. Число поступивших в некий университет юношей и девушек одинаково. Какова вероятность того, что во вновь набранной группе из 30 человек юношей и девушек окажется поровну?

Задача 307. Вероятность изготовления детали высшего сорта на данном станке равна 40% . Найти вероятность того, что среди наудачу взятых 50 деталей ровно 20 окажутся высшего сорта.

Задача 308. Если в n испытаниях Бернулли было достигнуто k успехов, то относительной долей успехов (или частотой успехов) называется отношение k/n. Вычислить вероятность того, что эта частота отклонится от p не более, чем на ε.

Решение

$$P\Big(\Big|\frac{k}{n}-p\Big|<\varepsilon\Big)=P\Big(-n\varepsilon<k-np<n\varepsilon\Big)=$$
$$=P\Big(-\frac{n\varepsilon}{\sqrt{npq}}<\frac{k-np}{\sqrt{npq}}<\frac{n\varepsilon}{\sqrt{npq}}\Big)=2\Phi\Big(\varepsilon\sqrt{\frac{n}{pq}}\Big)-1.$$

Эта формула используется для решения задач, а мы обратим внимание на одно из ее следствий. Из формулы видно, что, какой бы малый ни был ε, вероятность еще меньших отклонений стремится к единице. Это важное наблюдение часто называют *законом больших чисел* для испытаний Бернулли.

Задача 309. Вероятность появления события в каждом из 10 000 независимых испытаний равна 0,75. Найти вероятность того, что относительная частота появления события отклонится от его вероятности по абсолютной величине не более чем на 0,01.

Задача 310. Сколько раз нужно бросить игральную кость, чтобы вероятность неравенства
$$\Big|\frac{k}{n}-\frac{1}{6}\Big|<0,01$$
была не меньше, чем вероятность противоположного неравенства? Здесь k — число выпавших шестерок.

Задача 311. Вероятность появления события в каждом из 10 000 независимых испытаний равна 0,75. Найти вероятность того, что относительная частота появления события отклонится от его вероятности по абсолютной величине не более чем на 0,01.

Задача 312. Вероятность появления события в каждом из 2000 независимых испытаний равна 0,75. Найти математическое ожидание числа появлений события и вероятность того, что отклонение числа событий от математического ожидания составит не более 2% .

Задача 313. За лейбористов и консерваторов голосуют примерно поровну избирателей. Проводится опрос с целью уточнить долю избирателей, голосующих за лейбористов. Сколько человек надо опросить, чтобы ошибка оценки была не более 4% с вероятностью 99%?

Решение.

Из формулы, полученной в задаче 308, следует

$$\varepsilon\sqrt{n} \geqslant 2,58\sqrt{pq}.$$

Поскольку $pq \approx 1/4$, получим

$$\sqrt{n} \approx 2,58 \cdot \frac{\sqrt{pq}}{\varepsilon} \approx 32,2; \quad n \approx 1040.$$

Задача 314. Вероятность появления события в каждом из независимых испытаний равна 0,2. Найти число испытаний, при котором с вероятностью 0,9876 можно ожидать, что относительная частота появления события отклонится от его вероятности по абсолютной величине не более чем на 0,04.

Задача 315. Вероятность появления события в каждом из независимых испытаний равна 0,75. Найти число испытаний, которые нужно провести, чтобы с вероятностью 0,99 относительная частота появления события отклонилась от значения 0,75 по абсолютной величине не более чем на 0,01.

Задача 316. Вероятность появления события в каждом из независимых испытаний равна 0,2. Найти число испытаний, при котором с вероятностью 0,9876 можно ожидать, что относительная частота появления события отклонится от его вероятности по абсолютной величине не более чем на 0,04.

Задача 317. Вероятность события A равна 0,56, а вероятность события B равна 0,42, причем эти события независимы. Проводится серия из $n = 100$ испытаний, в каждом из которых могут произойти эти события. Оценить вероятность того, что в этой серии число произошедших событий A окажется не больше, чем число произошедших событий B.

Решение.

Пусть вероятность события A равна p_1, а вероятность события B равна

p_2. Вероятности противоположных событий обозначим соответственно q_1 и q_2. Число событий A обозначим k_1, а число событий B — k_2. Это — случайные величины, имеющие биномиальное распределение. Нам следует оценить вероятность события $(k_1 \leqslant k_2)$.

В соответствии с теоремой Муавра — Лапласа, случайные величины

$$x_1 = \frac{k_1 - np_1}{\sqrt{np_1q_1}} \text{ и } x_2 = \frac{k_2 - np_2}{\sqrt{np_2q_2}}$$

имеют приблизительно стандартное нормальное распределение. Неравенство $(k_1 \leqslant k_2)$ теперь можно переписать так:

$$x_1\sqrt{p_1q_1} - x_2\sqrt{p_2q_2} \leqslant \sqrt{n}(p_2 - p_1).$$

Обозначим через L случайную величину, стоящую в левой части этого последнего неравенства. Она является разностью независимых нормально распределенных случайных величин и поэтому распределена нормально. Ее математическое ожидание равно нулю:

$$ML = \sqrt{p_1q_1}Mx_1 - \sqrt{p_2q_2}Mx_2 = \sqrt{p_1q_1} \cdot 0 - \sqrt{p_2q_2} \cdot 0 = 0.$$

А ее дисперсия равна

$$DL = p_1q_1Dx_1 + p_2q_2Dx_2 = p_1q_1 + p_2q_2.$$

Теперь можно выразить интересующую нас вероятность через функцию распределения стандартного нормального распределения:

$$P = P(k_1 \leqslant k_2) = P\Big(\frac{x_1\sqrt{p_1q_1} - x_2\sqrt{p_2q_2}}{\sqrt{p_1q_1 + p_2q_2}} \leqslant \frac{\sqrt{n}(p_2 - p_1)}{\sqrt{p_1q_1 + p_2q_2}}\Big) = \Phi\Big(\frac{\sqrt{n}(p_2 - p_1)}{\sqrt{p_1q_1 + p_2q_2}}\Big).$$

Подставляя числа из условия, находим

$$P = \Phi(-2) = 0,2275.$$

Ответ: 0,2275.

24. Центральная предельная теорема

Еще одной теоремой, объясняющей важность и широту применения нормального распределения, является центральная предельная теорема. Мы будем использовать такую формулировку этой теоремы: если случайные величины $X_1, X_2 \ldots X_n$ независимы, имеют одинаковые математические ожидания и дисперсии

$$MX_i = a; DX_i = D,$$

то при достаточно больших n сумма этих случайных величин

$$X = X_1 + X_2 + \cdots + X_n$$

является случайной величиной, имеющей приблизительно нормальное распределение.

Теорема Муавра — Лапласа, очевидно, является частным случаем центральной предельной теоремы.

На практике центральную предельную теорему используют (не для точного вычисления вероятности, а для ее оценки), когда $n \geqslant 10$.

В соответствии с правилами сложения математических ожиданий и дисперсий для независимых случайных величин получим

$$MX = na; DX = nD.$$

В частности, если обозначить среднеквадратическое отклонение (разброс) каждой из случайных величин X_i через σ, а случайной величины X — через s, то из последнего соотношения получим

$$s = \sigma\sqrt{n},$$

то есть разброс растет пропорционально квадратному корню из числа слагаемых.

Задача 318. При составлении статистического отчета надо было сложить 10 000 чисел, каждое из которых было округлено до ближайшего целого. Предполагая, что ошибки, возникшие от округления чисел, взаимно независимы и равномерно распределены на интервале (-0,5; 0,5), найти пределы, в которых с вероятностью 0,997 будет лежать суммарная ошибка.

Решение

Ошибка суммы X есть случайная величина, равная сумме случайных величин $X_i, i = 1 \ldots 10000$, равных ошибке каждого слагаемого. Поскольку

$$MX_i = 0; DX_i = 1/12,$$

получим, что

$$MX = 0; DX = s^2 = 833,33; s = 28,87.$$

Будем искать симметричный промежуток $(-z; z)$, значения в котором случайная величина X принимает с вероятностью 0,997. Запишем уравнение

$$P\{-z < X < z\} = 0,997.$$

Переходя к стандартному нормальному распределению, получим

$$P\left\{\frac{-z}{28,87} < \frac{X}{28,87} < \frac{z}{28,87}\right\} = 0,997;$$

$$2\Phi\left(\frac{z}{28,87}\right) - 1 = 0,997.$$

Из таблицы находим

$$\frac{z}{28,87} = 2,96,$$

следовательно, $z = 85,45$.
Ответ: $\pm 85,45$.

Задача 319. Игральную кость подбросили 1000 раз и просуммировали число выпавших очков. В каких пределах с вероятностью 0,99 лежит эта сумма (для "правильной" кости)?

Вася подбросил украденную из казино кость 1000 раз и насчитал в сумме 3298 очков. Правильная ли это кость?

Решение

Число очков при одном бросании — случайная величина, ее среднее значение (математическое ожидание) было подсчитано ранее, оно равно 3,5. Дисперсию ее тоже можно подсчитать — она равна $35/12$.

В соответствии с центральной предельной теоремой суммарное число выпавших очков должно быть распределено приблизительно нормально со средним

$$a = 3500$$

и дисперсией

$$\sigma^2 = 1000 \cdot \frac{35}{12}.$$

Пусть X — суммарное число выпавших очков. Составим уравнение

$$P(a - z < X < a + z) = 0.99.$$

Далее действуем уже привычным способом:

$$P(-\frac{z}{\sigma} < \frac{X - a}{\sigma} < \frac{z}{\sigma}) = 0.99.$$

Отсюда найдем по таблицам

$$\frac{z}{\sigma} = 2.58.$$

Следовательно:

$$z = 2.58 \cdot \sqrt{1000 \cdot \frac{35}{12}} = 139.$$

Поэтому сумма выпавших очков почти наверняка (с вероятностью 99 %) лежит в интервале от $a - z$ до $a + z$, то есть от 3361 до 3639. Число 3298 в этот интервал не попадает, поэтому кость на правильную не похожа.

Задача 320. Топливозаправщик заливает на нефтезаводе 10 тысяч литров солярки, после чего развозит эту солярку потребителям. Каждому потребителю следует отпускать ровно 200 литров, но топливо отпускается на глазок с точностью $\sigma = 10$ литров. Какова вероятность того, что последнему потребителю вообще не достанется топлива? достанется не более 100 литров?

Задача 321. Каков будет средний дневной заработок профессионального нищего в московском метро, если предположить, что он успевает за рабочий день обойти 80 вагонов, 10 рублей ему подадут в среднем в каждом 30-м вагоне, а полученная в каждом вагоне сумма является случайной величиной, подчиняющейся геометрическому распределению? Какова вероятность, что он заработает больше 600 рублей?

Задача 322. Импортный картофель в магазине фасуется в пакеты по 3 кг. Одна картофелина весит в среднем 120 грамм, со среднеквадратическим отклонением 20 грамм. Ленивый фасовщик просто кладет в пакет 25 случайно выбранных картофелин. Торговая инспекция взвешивает два произвольно взятых пакета и штрафует магазин, если хотя бы в одном из них меньше 2900 грамм. С какой вероятностью магазин будет оштрафован?

Задача 323. Электрики подсчитали, что лампочки в офисе фирмы должны работать 120 000 часов в течение декабря. Срок службы одной лампочки представляет собой случайную величину, распределенную по показательному закону со средним 1000 часов. На 1 декабря осталась 121 лампочка. Какова вероятность, что их не хватит до конца месяца?

Задача 324. Срок работы электрической лампочки является случайной величиной, подчиняющейся показательному распределению со средним значением 4000 часов. В каких пределах с вероятностью 0,99 будет лежать число электрических лампочек, вышедших из строя за год на предприятии, на котором используется 1000 лампочек, каждая из которых должна быть включена в среднем в течение 2500 часов в год?

Задача 325. Электрики подсчитали, что лампочки в офисе фирмы должны работать 120 000 часов в течение декабря. Срок службы одной лампочки представляет собой случайную величину, распределенную по показательному закону со средним 1000 часов. Сколько лампочек должно быть в наличии, чтобы с вероятностью 99 процентов их хватило до конца месяца?

Задача 326. Родион Раскольников забирает у каждой из зарубленных топором старушек по 20 ± 5 копеек. Сколько старушек ему надо зарубить топором, чтобы с вероятностью 99% ему хватило бы на бутылку водки, которая стоит 3 рубля 62 копейки?

Задача 327. В торгово-развлекательном центре “Мега” 170 магазинов. Эвелина Эммануиловна тратит на покупки в каждом из них 1000 ± 200 рублей. Какова вероятность, что перед последним магазином у нее не останется ни копейки, если сначала у нее было 175500 рублей?

Задача 328. Автопредприятие, вывозя грунт из котлована, выполнило 100 рейсов грузовиков. За один рейс грузовик перевозит $5 \pm 0,5$м^3 грунта. В каких пределах с вероятностью 0,98 лежит общее количество вывезенного грунта?

Задача 329. Расстояние между городама меряют стометровый рулеткой, дающей ошибку примерно ± 0.5 м. В каких пределах с вероятностью 99% будет ошибка, tсли расстояние равно 640 км?

Задача 330. В результате 16 измерений расстояния на местности получено среднее значение этого расстояния 502 метра. Выдвинута гипотеза, что расстояние на самом деле равно 500 метрам, а разница объясняется ошибками измерений. Считать, что точность одного измерения (среднеквадратическое отклонение) равно 3 метрам, а систематическая ошибка отсутствует.

Если предположить, что истинное значение расстояния равно 500 метрам, то в каких пределах с вероятностью 0,95 будут находиться средние значения результатов 16 измерений?

В задачах 331 — 334 используется следующая терминология из математической статистики. Если наблюдаемое в опыте значение лежит внутри вычисленных пределов, то говорят, что данные с доверительной вероятностью 0,95 (0,98, 0,99) согласуются с гипотезой (о случайности отклонения), в противном случае — противоречат гипотезе.

Задача 331. Покупатель коммерческого лотка в среднем тратит на покупку 100 рублей, причем эта трата подчиняется геометрическому распределению вероятностей. В каких пределах с вероятностью 0,95 лежит дневная выручка лотка, если за день приходит 100 покупателей?

В один из дней новый продавец лотка сдал выручку в размере 9400 рублей. Следует ли считать такое снижение выручки случайностью?

Задача 332. Проверка обнаружила, что у оптового строительного склада имеется в наличии 240 м^3 песка. Завоз песка осуществлялся большегрузными автомобилями, каждый из которых за один рейс перевозит $8 \pm 0,5$м^3 песка. По документам, было выполнено 36 таких рейсов. Следует ли признать факт наличия недостачи или такое отклонение вполне можно объяснить случайностью?

Задача 333. Средняя урожайность яблони в пору полного плодоношения обычно составляет от 100 до 150 килограммов. Примем, что эта урожайность является случайной величиной со средним 125 кг и среднеквадратическим отклонением 25 кг. В каких пределах тогда с вероятностью 0.98 будет находиться урожай яблок в саду, в котором 100 яблонь? Может ли урожай "случайно" составить 100 центнеров? 120 центнеров?

Задача 334. Маша выкуривает в день 30 ± 5 сигарет, причем все их она таскает из папиных запасов. В каких пределах с вероятностью 0,95 лежит число выкуренных Машей за неделю сигарет?

Как-то раз папа обнаружил, что у него в течение недели пропало всего 180 сигарет. Следует ли считать эту флуктуацию случайностью, или Маша действительно задумывается о том, чтобы бросить курить?

25. Разные задачи по теме 2

Задача 335. Распределением Коши (с нулевым сдвигом и параметром масштаба γ) называется закон распределения, описываемый плотностью вероятностей

$$\rho(x) = C\frac{\gamma}{x^2+\gamma^2}.$$

Найти константу C и доказать, что дисперсия этого распределения бесконечна, а математическое ожидание сущкствует только в смысле главного значения и в этом смысле равно нулю.

Задача 336. Найти функцию распределения для распределения Коши.

Задача 337. Доказать, что сумма независимых случайных величин, распрелеленных по Коши, также распределена по Коши.

Задача 338. Найти распределение длины отрезка, соединяющего точку $(0,1)$ на оси ординат с точкой на оси абсцисс, если угол наклона этого отрезка распределен равномерно на отрезке $[0, 2\pi]$.

Задача 339. Пусть вероятность успеха в каждом из независимых испытаний одинакова и равна p. В задаче 194 установлено, что вероятность того, что k-й успех наступит в испытании с номером m $(k < m)$, равна

$$P(X=k) = C_{m-1}^{k-1}p^k(1-p)^{m-k}.$$

Полученное распределение имеет свое название: отрицательное биномиальное распределение. Найти его математическое ожидание и дисперсию.

Задача 340. В урне находятся n шаров, из них k белых, а остальные $n-k$ — красные. Из урны извлекают m шаров без возвращения. Число белых шаров среди извлеченных — случайная величина, ее распределение называется гипергеометрическим распределением. В задаче 43 доказано, что закон распределения этой случайной величины таков:

$$P(X=r) = \frac{C_k^r \cdot C_{n-k}^{m-r}}{C_n^m}.$$

Найти математическое ожидание этой случайной величины.

Задача 341(*). Найти дисперсию гипергеометрического распределения (см. предыдущую задачу).

Задача 342. В k-мерном шаре радиуса 1 случайным образом выбрано n точек. Пусть X —расстояние от центра шара до ближайшей к нему точки. Найти медиану распределения случайной величины X. Найти также пределы этой медианы, если
1) k фиксировано, $n \to \infty$;
2) n фиксировано, $k \to \infty$.

Задача 343. В условиях предыдущей задача найти математическое ожидание X. (Достаточно выразить ответ через бета-функцию).

В задачах 344 — 346 используется следующая терминология из математической статистики. Наблюдаемое в опыте значениележит внутри вычисленных пределов, то говорят, что данные с доверительной вероятностью 0,95 согласуются с гипотезой (о случайности отклонения), в противном случае — противоречат гипотезе.

Задача 344. С 1871 по 1900 год в Швейцарии родилось 2 644 757 детей. Если принять гипотезу, что вероятность рождения мальчика равна 0,5, то в каких пределах будет с вероятностью 0,997 находиться число мальчиков?

С 1871 по 1900 год в Швейцарии родилось 1 359 671 мальчик и 1 285 086 девочек. Согласуются ли эти данные с гипотезой о том, что вероятность рождения мальчика равна 0,5? 0,515?

Задача 345. Каждый из студентов потока ходит на лекцию с вероятностью $1/5$. В каких пределах с вероятностью 0,95 будет лежать число студентов на лекции, если на потоке 100 человек?

Заместитель декана, проводя проверку посещаемости, обнаружил, что на лекцию пришло 30 человек. Следует ли считать это отклонение случайным или сведения о проверке просочились из деканата?

Задача 346. Многие ботаники делали опыты по скрещиванию жёлтого (гибридного) гороха. По известной гипотезе Менделя вероятность появления зелёного гороха в таких опытах равна 0,25. Если предположить справедливость этой гипотезы, то в каких пределах с вероятностью 0,95 будет лежать число появлений зелёного гороха при 34153 опытах скрещивания?

В указанных 34153 экспериментах в 8506 случаях был получен зелёный горох. Согласуются ли эти эксперименты с гипотезой Менделя?

Примечание. Задача (с небольшими изменениями) взята из задачника по теории вероятностей Л. Д. Мешалкина, выпущенного издательством Московского университета в 1963 году, когда генетика ещё не была общепризнанной наукой.

Задача 347. (из "Математического тривиума" В. И. Арнольда). Каждый из N участников игры в очко на пальцах, стоящих по кругу, выбрасывает несколько пальцев (от 0 до 5) правой руки, после чего для определения победителя суммарное число выкинутых пальцев отсчитывается по кругу от водящего. Водящий известен заранее. В вашей команде $N/10$ участников. Как следует расположиться членам вашей команды, чтобы увеличить шансы на выигрыш? Оценить, при каком числе участников N вероятность выигрыша хотя бы одного из членов вашей команды становится больше 0,9. Как ведет себя при $N \to \infty$ вероятность выигрыша водящего?

Тема 3. Функции случайных величин. Случайные векторы

26. Функции случайных величин

Если X — случайная величина, и f — некоторая функция, то $Y = f(X)$ также будет случайной величиной. Каково будет ее распределение?

Общую формулу можно получить для случая, когда f — монотонная функция. Тогда, если $y = f(x)$, то $x = g(y)$, где $g = f^{-1}$ обозначена обратная функция. Пусть это — монотонно возрастающая функция. Тогда

$$F_Y(y) = P(Y < y) = P(X < g(y)) = F_X(g(y)) = \int\limits_{-\infty}^{g(y)} \rho(t)dt.$$

Дифференцируя по верхнему пределу и применяя формулу дифференцирования сложной функции, получим

$$\rho_Y(y) = \rho_X(g(y))g'(y).$$

В общем случае для подобных задач следует использовать функцию распределения.

Задача 348. Плотность вероятности случайной величины X равна

$$\rho(x) = \frac{1}{\pi(1+x^2)}.$$

Найти плотность вероятности случайной величины

$$Y = |X|^3 - 4.$$

Решение.

Будем искать функцию распределения Y:

$$F_Y(y) = P(Y < y) = P(|X|^3 - 4 < y) =$$
$$= P(-\sqrt[3]{y+4} < X < \sqrt[3]{(}y+4)).$$

Последний переход справедлив для $y \geqslant -4$. Для $y < -4$ функция распределения, очевидно, равна нулю.

Продолжим вычисления:

$$F_Y(y) = \frac{1}{\pi} \int\limits_{-\sqrt[3]{y+4}}^{\sqrt[3]{y+4}} \frac{dx}{1+x^2} = \frac{1}{\pi} \operatorname{arctg} x|\,_{\sqrt[3]{y+4}}^{\sqrt[3]{y+4}} =$$

$$= \frac{2}{\pi} \operatorname{arctg} \sqrt[3]{y+4}.$$

Для того, чтобы найти плотность вероятности, надо полученное выражение продифференцировать. Получим

$$\rho_Y(y) = \frac{2}{3\pi\Big(1 + (y+4)^{2/3}\Big)(y+4)^{2/3}} \text{ при } y \geqslant -4.$$

Задача 349. Пусть X — непрерывная случайная величина с функцией распределения $F(x)$. Доказать, что случайная величина $F(X)$ распределена равномерно на $(0,1)$.

Задача 350. Распределение Максвелла — распределение скоростей молекул. Для этого распределения плотность вероятности равна

$$\rho(x) = \sqrt{\frac{2}{\pi}} \beta^{3/2} x^2 e^{-\beta x^2/2}$$

для положительных x и 0 для отрицательных. Как распределены энергии? Чему равна средняя энергия?

Задача 351. Найти функцию распределения квадрата стандартной нормальной случайной величины. Найти математическое ожидание этого квадрата.

Задача 352. Случайная величина ξ распределена равномерно на отрезке $[-3,0]$. Найти математическое ожидание, дисперсию и плотность распределения случайной величины $\eta = \ln(\xi^2 + 2\xi + 2)$.

Решение

Будем искать функцию распределения

$$F(x) = P(\eta < x) = P(\ln(\xi^2 + 2\xi + 2) < x) = P(\xi^2 + 2\xi + 2 < e^x) =$$

$$= P(\xi^2 + 2\xi + 1 < e^x - 1) = P((\xi+1)^2 < e^x - 1).$$

При $x < 0$ эта вероятность равна нулю, поэтому дальше будем считать, что $x > 0$. Тогда

$$F(x) = P(-\sqrt{e^x-1} < \xi + 1 < \sqrt{e^x-1}) = P(-\sqrt{e^x-1} - 1 < \xi < \sqrt{e^x-1} - 1)$$

Плотность вероятности случайной величины ξ равна $1/3$, поэтому

$$F(x) = \frac{1}{3}\int\limits_D dx,$$

где через D обозначено пересечение отрезков $[-3, 0]$ и $[-\sqrt{e^x-1}-1, \sqrt{e^x-1}-1]$.

Решая уравнение $\sqrt{e^x-1} - 1 = 0$, найдем $x = \ln 2$.
Решая уравнение $-\sqrt{e^x-1} - 1 = -3$, найдем $x = \ln 5$.
Поэтому длина отрезка D равна
$2\sqrt{e^x-1}$, если $x < \ln 2$;
$\sqrt{e^x-1} + 1$, если $\ln 2 < x < \ln 5$;
3 если $x > \ln 5$. Функция распределения равна

$$F(x) = \begin{cases} \frac{2}{3}\sqrt{e^x-1}, \text{ если } x < \ln 2; \\ \frac{1}{3}(\sqrt{e^x-1}+1) \text{ если } \ln 2 < x < \ln 5; \\ 1, \text{ если } x > \ln 5. \end{cases}$$

Плотность вероятностей можно найти дифференцированием

$$\rho(x) = F'(x).$$

При этом можно воспользоваться системой компьютерной алгебры. Поскольку случайная величина η, как выяснилось, принимает значения на отрезке $[0, \ln 5]$, для нахождения математического ожидания и дисперсии можно воспользоваться следующими формулами:

$$MX = \int\limits_0^{\ln 5} x\rho(x)dx;$$

$$DX = M(X - MX)^2 = \int\limits_0^{\ln 5} (x - MX)^2 \rho(x)dx.$$

Для вычисдения этих интегралов также следует воспользоваться системой компьютерной алгебры.

Задача 353. Случайная величина X распределена равномерно на интервале $(-1, 1)$. Найти распределение случайной величины

$$Y = \operatorname{tg} \frac{X}{2}$$

Задача 354. Случайная величина X распределена равномерно на интервале $(1, 2)$. Найти распределение случайной величины

$$Y = 3X - 4,$$

ее математическое ожидание и дисперсию.

Задача 355. Случайная точка равномерно распределена на единичной окружности. Как распределена ее проекция на ось абсцисс?

Задача 356. Случайные величины X и Y независимы и равномерно распределены на отрезке $[-1/2, 1/2]$. Найти плотность вероятности случайной величины $Z = XY$, ее математическое ожидание и дисперсию.

27. Средние значения функций случайных величин

Часто требуется определить не всю функцию от случайной величины, а только ее среднее значение. В этом случае можно применить формулу

$$M(f(X)) = \int\limits_{-\infty}^{\infty} f(x)\rho(x)dx.$$

Задача 357. Найти среднюю длину хорды, соединяющую случайную точку на верхней половине единичной окружности с центром в начале координат с точкой $(1, 0)$.

Задача 358. Случайная величина равномерно распределена на интервале $(-1, 1)$. Найти среднее значение ее квадрата.

Задача 359. Чему равно среднее значение квадрата случайной величины, распределенной показательно с параметром λ? Можно ли указать больше одного способа решения этой задачи?

Задача 360. Случайная величина равномерно распределена на интервале $(0, 1)$. Найти среднее значение ее квадратного корня. Совпадает ли оно с квадратным корнем из ее среднего?

Задача 361. Чему равно среднее значение квадрата стандартно нормально распределенной случайной величины?

Задача 362. Чему равно среднее значение модуля стандартно нормально распределенной случайной величины?

Задача 363. Случайная величина X равномерно распределена на интервале $(-1, 1)$. Найти среднее расстояние от начала координат до точки (X, Y), где $Y = 4 - X^2$.

Задача 364. Случайная величина X равномерно распределена на интервале $(1/2, 2)$. Найти среднее расстояние от начала координат до точки (X, Y), где $Y = 1/X$.

Задача 365. Как изменится результат предыдущей задачи, если равномерно будет распределен угол между направлением из начала координат на точку (X, Y) и осью абсцисс?

Задача 366. Пусть A — точка на отрезке, соединяющнм точки $(0, 2)$ и $(1, 1)$, а точка O — начало координат. Чему равна средняя длина отрезка AO, если
равномерно распределена точка A?
равномерно распределен угол между AO и осью абсцисс?
равномерно распределен косинус этого угла?
равномерно распределен синус этого угла?

Задача 367. Измерения угла равномерно распределены на отрезке $[\alpha - \varepsilon, \alpha + \varepsilon]$. Чему равно среднее значение синуса этого угла?

Задача 368(*). Случайная величина представляет собой результат измерения некоторого угла. Она распределена нормально со средним a и дисперсией σ^2. Чему равно среднее значение синуса этой случайной величины? Совпадает ли оно с синусом среднего значения?

28. Случайные векторы

Системы случайных величин, они же случайные векторы — это вектор $(X_1, \dots X_n)$, каждый из компонентов которого является случайной величиной.

Мы рассмотрим случайные векторы на примере двумерных.

Закон распределения дискретной двумерной случайной величины представляет собой двумерную таблицу. Для каждой пары значений (X_1, X_2) указана вероятность того,что принимается именно эта пара значений. Разумеется, все вероятности неотрицательны, а их сумма по всей таблице равна 1. Например, в приведенной ниже таблице вероятность того, что и X_1, и X_2 примут значение 1, равна 1/6.

$X_1 \setminus X_2$	1	2	3
1	1/6	0	1/12
2	0	1/4	1/4
3	1/12	1/12	1/12

Непрерывная двумерная (а также многомерная) случайная величина, задается плотностью вероятностей, как и в одномерном случае. Это функция двух переменных (или n переменных) $\rho(x_1, x_2)$, всюду неотрицательная, и при этом

$$\iint\limits_{\mathbb{R}^2} \rho(x_1, x_2) dx_1 dx_2 = 1.$$

Вероятность попадания в некоторую область Ω для непрерывной случайной величины выражается формулой:

$$P((x_1, x_2) \in \Omega) = \iint\limits_{\Omega} \rho(x_1, x_2) dx_1 dx_2.$$

Для многомерных случайных величин также определяется функция распределения:

$$F(x_1, x_2) = P(X_1 < x_1, X_2 < x_2).$$

Функция распределения и плотность вероятностей связаны следующими соотношениями:

$$F(x, y) = \int\limits_{-\infty}^{x} \int\limits_{-\infty}^{y} \rho(x, y) dx dy; \quad \rho(x, y) = \frac{\partial^2 F(x, y)}{\partial x \partial y}.$$

Из двумерных распределений первым мы встретимся с равномерным распределением. Плотность вероятности случайной величины, равномерно распределенной в некоторой ограниченной области, равна константе внутри области и нулю вне ее. Чему равна константа? Нужно, чтобы интеграл был равен единице, поэтому константа будет равна 1, деленной на площадь области.

Законы распределения компонент случайного вектора.

Для дискретных случайных величин определить закон распределения компонент довольно просто: надо просуммировать соответствующий столбец или соответствующую строку. Например, для приведенной выше таблицы закон распределения компоненты X_2 таков:

X_2	1	2	3
P	1/4	1/3	5/12

В непрерывном случае вместо суммирования надо выполнить интегрирование. Получим следующую формулу:

$$\rho_X(x) = \int\limits_{-\infty}^{\infty} \rho(x, y) dy$$

Следующим важным понятием будет понятие условного закона распределения компонента случайного вектора. И снова формулы будут различны в случае дискретных и непрерывных случайных величин. В дискретном случае

$$P(x_i | y_j) = \frac{P(x_i, y_j)}{P(y_j)}.$$

В непрерывном же случае для одного из компонент определяется плотность вероятности:

$$\rho(x|y) = \frac{\rho(x, y)}{\int\limits_{-\infty}^{\infty} \rho(x, y) dx}.$$

Две компоненты случайной величины называются независимыми, если условное распределение каждой из них не зависит от значения второго компонента. По сути, это то же самое определение независимости, что и раньше.

Коэффициентом ковариации компонент двумерного случайного вектора или пары случайных величин называется

$$\mathrm{cov}(X, Y) = M((X - MX)(Y - MY)).$$

Это — мера связи, или взаимной зависимости, между компонентами. Этот коэффициент размерный, и поэтому его значение зависит от применяемых единиц измерения. Поэтому чаще используется несколько другой показатель — *коэффициент корреляции*. Это тот же коэффициент ковариации, только нормированный. Его определение

$$\mathrm{r}(X, Y) = \frac{\mathrm{cov}(X, Y)}{\sigma_X \cdot \sigma_Y} = \frac{M((X - MX)(Y - MY))}{\sqrt{DX \cdot DY}}.$$

Коэффициент корреляции принимает значения от -1 до 1 и показывает меру линейной связи между компонентами.

Для дискретных случайных величин коэффициент ковариации равен

$$\mathrm{cov}(X, Y) = \sum_{i=1}^{n} \sum_{j=1}^{m} (x_i - MX)(y_j - MY)p(x_i, y_j),$$

где
$x_i,\ i = 1, \ldots n$ — значение компоненты X;
$y_j,\ j = 1, \ldots m$ — значение компоненты Y;
$p(x_i, y_j)$ — вероятность, с которой принимается эта пара значений.

Для непрерывных случайных величин формула аналогична:

$$\mathrm{cov}(X, Y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - MX)(y - MY)\rho(x, y)dxdy.$$

Задача 369. Доказать формулу $\mathrm{cov}(X, Y) = M(XY) - MX \cdot MY$.

Задача 370. Доказать, что компоненты случайного вектора независимы тогда и только тогда, когда функция распределения двумерной случайной величины равна произведению функций распределения компонент.

Решение

Доказательство основано на определении функции распределения. Если компоненты независимы, то независимы события $X < x$ и $Y < y$, и поэтому

$$F(x, y) = P(X < x, Y < y) = P(X < x) \cdot P(Y < y) = F_X(x) \cdot F_Y(y).$$

Обратно, если $F(x, y) = P(X < x, Y < y) = F_X(x) \cdot F_Y(y) = P(X < x) \cdot P(Y < y)$, что и требовалось.

Задача 371. Доказать следствие из предыдущей задачи.
Следствие 1. Непрерывные случайные величины независимы тогда и только тогда, когда плотность двумерного распределения равна произведению плотностей компонент.

Задача 372. Доказать еще одно следствие из задачи 370.
Следствие 2. Если двумерная случайная величина равномерно распределена в некоторой области, то компоненты независимы тогда и только тогда, когда область представляет собой прямоугольник со сторонами, параллельными осям координат.

Задача 373. Доказать, что коэффициент корреляции не зависит от применяемых систем единиц, а также от начала отсчета.

Задача 374. Доказать, что коэффициент корреляции двух случайных величин не изменяется при линейных преобразованиях этих величин.

Задача 375. Доказать, что если случайные величины независимы, то коэффициент корреляции между ними равен нулю. (Такие величины называют некоррелированными).

Задача 376. Пусть случайная величина X распределена равномерно на интервале $(-1, 1)$. Доказать, что случайные величиины X и X^2 некоррелированы, но зависимы.

Решение.

Эти случайные величины, очевидно, зависимы. Если, скажем, $X^2 > 1/4$, то ясно, что X не можкт принимать значения на интервале $(-1/2, 1/2)$.

Вычислим коэффициент ковариации. Для этого сначала найдем плотность вероятности случайной величины $Y = X^2$. Ее функция распределения равна

$$F(y) = P(Y < y) = P(X^2 < y) =$$

$$= P(-\sqrt{y} < X < \sqrt{y}) = \sqrt{y} \text{ при } 0 < y < 1.$$

Плотность вероятностей

$$\rho(y) = \frac{1}{2\sqrt{y}} \text{ при } 0 < y < 1.$$

Математическое ожидание

$$MY = \int\limits_0^1 \frac{ydy}{2\sqrt{y}} = \frac{1}{2}\int\limits_0^1 \sqrt{y}dy = \frac{1}{3}.$$

Теперь можно вычислить коэффициент ковариации

$$\mathrm{cov}(X,Y) = \int\limits_{-1}^1 \int\limits_0^1 x(y-\frac{1}{3})\cdot\frac{1}{2}\cdot\frac{1}{2}\cdot\sqrt{y}dydx =$$

$$= \int\limits_{-1}^1 \frac{1}{2}xdx \cdot \int\limits_0^1 (y-\frac{1}{3})\cdot\frac{1}{2}\cdot\sqrt{y}dy = 0,$$

потому что нулю равен первый сомножитель.

Поэтому коэффициент корреляции также равен нулю. Таким образом, X и X^2 некоррелирован

Задача 377. Доказать, что
a) $\mathrm{cov}(\alpha X_1 + \beta X_2, Y) = \alpha \mathrm{cov}(X_1, Y) + \beta \mathrm{cov}(X_2, Y)$,
b) $\mathrm{cov}(X,Y) = \mathrm{cov}(Y,X)$. Иными словами, нахождение коэффициента ковариации — операция линейная по каждому из аргументов и симметрическая.

Задача 378. Доказать, что $\mathrm{cov}(X,X) = DX$.

Задача 379. Доказать, что $|\mathrm{cov}(X,Y)| \leqslant \sqrt{DX\cdot DY}$ (неравенство Коши — Буняковского).

Задача 380. Случайные величины X и Y принимают значения -1, 0, 1 с равными вероятностями и независимы. Найти коэффициент ковариации случайных величин $U = X+Y$ и $V = XY^2$.

Задача 381. Случайные величины X и Y независимы и имеют такие законы распределения

X	1	2	3
P	0,5	0,2	0,3

Y	1	2	3
P	0,6	0,1	0,3

Найти коэффициенты ковариации и корреляции между $X+Y$ и $Y-2X$. Найти также коэффициенты ковариации и корреляции между $X-Y$ и $2X+Y$.

Задача 382. Двумерная случайная величина равномерно распределена в квадрате с вершинами $(1,0)$, $(0,1)$, $(-1,0)$, $(0,-1)$, а $F(x,y)$ — ее функция распределения. Найти $F(0,0)$.

Задача 383. Двумерная случайная величина равномерно распределена в треугольнике с вершинами $(0,0)$, $(0,2)$ и $(2,0)$, а $F(x,y)$ — ее функция распределения. Найти $F(1,1)$.

Задача 384. Дано распределение двумерного случайного вектора (X,Y).

$X \setminus Y$	1	2	3
1	1/18	1/18	1/18
2	1/9	1/9	1/9
3	1/6	1/6	1/6

Найти законы распределения компонент X и Y, их математические ожидания, дисперсии и коэффициент корреляции. Верно ли, что X и Y независимы?

Задача 385. Дано распределение двумерного случайного вектора (X,Y).

$X \setminus Y$	1	2	3
1	0	1/4	0
2	1/4	0	1/4
3	0	1/4	0

Найти законы распределения компонент X и Y, их математические ожидания, дисперсии и коэффициент корреляции. Верно ли, что X и Y независимы?

Задача 386. Дано распределение двумерного случайного вектора (X,Y).

$X \setminus Y$	1	2	3
1	1/16	1/8	1/6
2	1/3	0	1/24
3	1/12	1/12	3/16

Найти законы распределения компонент X и Y, их математические ожидания, дисперсии и коэффициент корреляции. Верно ли, что X и Y независимы?

Задача 387. Дано распределение двумерного случайного вектора (X, Y).

$X \setminus Y$	1	2	3
1	$1/15$	$1/10$	$1/6$
2	0	$1/3$	0
3	$1/6$	$1/10$	$1/15$

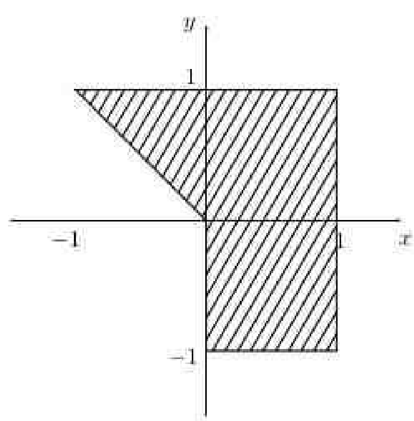
Найти законы распределения компонент X и Y, их математические ожидания, дисперсии и коэффициент корреляции. Верно ли, что X и Y независимы?

Задача 388. Случайная величина X равномерно распределена на отрезке $[-1, 1]$. Найти коэффициент корреляции между X и X^3.

Задача 389. Случайный вектор принимает значения в области, изображенной на рисунке, и имеет там плотность вероятности

$$\rho(x, y) = A(x + y + 1).$$

Найти значение нормирующей константы A, плотности вероятности, математические ожидания и дисперсии компонент X и Y. Найти коэффициент корреляции между компонентами. Верно ли, что X и Y независимы?



Решение.

Сразу, без всяких вычислений, можно дать ответ на вопрос о зависимости компонент. Действительно, при $Y = -1$ значения X находятся в пределах от 0 до 1, а при $Y = 1$ — в пределах от -1 до 1. Поэтому компоненты, разумеется, зависимы.

Обозначим нашу область Ω. Для использования любой системы компьютерной алгебры, как и для счета вручную, следует правильно расставить пределы интегрирования. Получим, что для нашей области для любой функции $f(x, y)$

$$\iint\limits_{\Omega} f(x,y)dxdy = \int\limits_{0}^{1} dy \int\limits_{-y}^{1} f(x,y)dx + \int\limits_{-1}^{0} dy \int\limits_{0}^{1} f(x,y)dx$$

Для нахождения нормирующей константы проинтегрируем плотность вероятности $\rho(x, y)$ по нашей области. Получим, что $11A/3 = 1$, следовательно $A = 3/11$.

Для нахождения плотностей вероятности компонент проинтегрируем $\rho(x, y)$ по каждой из переменных. Получим

$$\rho_X(x) = \begin{cases} \frac{3}{22}(x^2 + 4x + 3) \text{ при } x < 0 \\ \frac{6}{11}(x + 1) \text{ при } x > 0; \end{cases}$$

$$\rho_Y(y) = \begin{cases} \frac{3}{22}(2y + 3) \text{ при } y < 0 \\ \frac{3}{22}(y^2 + 4y + 3) \text{ при } y > 0. \end{cases}$$

Найдем теперь математические ожидания компонент:

$$MX = \int\limits_{-1}^{1} \rho_X(x)dx = \frac{35}{88}; \quad MY = \int\limits_{-1}^{1} \rho_Y(y)dy = \frac{27}{88}.$$

Найдем дисперсии:

$$DX = \int\limits_{-1}^{1} (x - MX)^2 \rho_X(x)dx = \frac{7251}{38720}; \quad DY = \int\limits_{-1}^{1} (y - MY)^2 \rho_Y(y)dy = \frac{10611}{38720};$$

Найдем ковариацию компонент

$$\mathrm{cov}(X, Y) = \iint\limits_{\Omega} (x - MX)(y - MY)\rho(x, y)dxdy = -\frac{2877}{38720}.$$

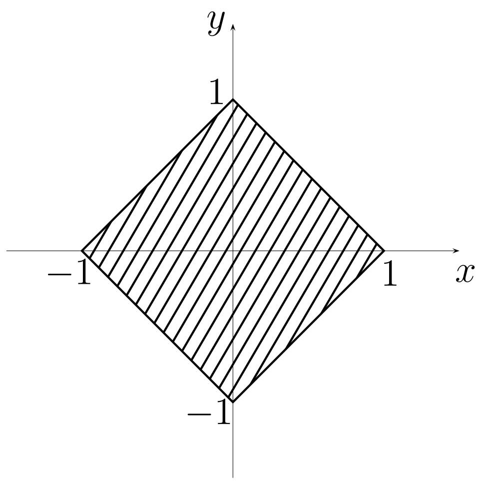
Следовательно, коэффициент корреляции равен

$$\mathrm{r}(X, Y) = \frac{\mathrm{cov}(X, Y)}{\sqrt{DX \cdot DY}} = -\frac{959\sqrt{23747025}}{14248215} \approx -0,328.$$

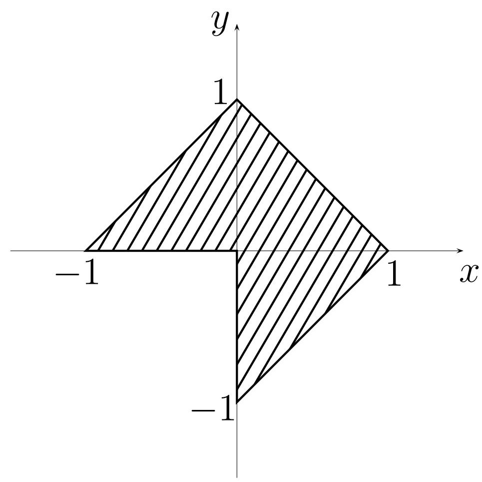
Для вычисления всех этих интегралов следует воспользоваться системой компьютерной алгебры.

В задачах 390 — 393 случайный вектор равномерно распределен в области, изображенной на рисунке. Найти плотности вероятности компонент X и Y, их математические ожидания, дисперсии и коэффициент корреляции. Верно ли, что X и Y независимы?

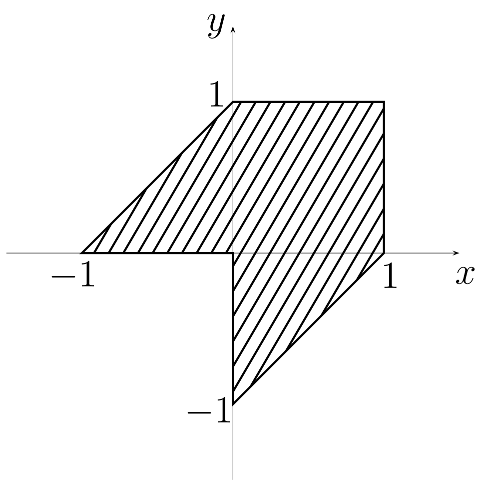
Задача 390.



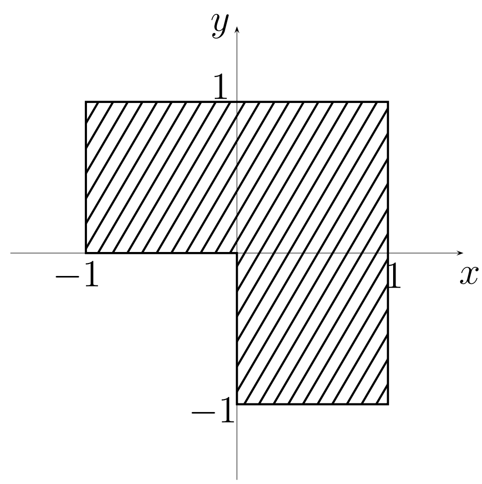
Задача 391.



Задача 392.



Задача 393.



Задача 394(*). Случайный вектор (X, Y) равномерно распределен в единичном круге. Найти плотность вероятности случайной величины $Z = X/Y$.

Задача 395(*). Случайные величины X и Y распределены стандартно нормально и независимы. Найти плотность вероятности случайной величины $Z = X/Y$.

Задача 396(*). Случайные величины X и Y распределены стандартно нормально и независимы. Найти плотность вероятности случайной величины $Z = X^2 + Y^2$.

Задача 397. Вектор (X, Y) равномерное распределен в единичном круге. Новые случайные величины U и V определяются так:

$$S = X^2 + Y^2, \quad U = X \cdot \sqrt{-\frac{2\ln S}{S}}, \quad V = Y \cdot \sqrt{-\frac{2\ln S}{S}},$$

Доказать что U и V имеют стандартное нормальное распределение и независимы.

29. Производящие функции

Производящие функции определяются для целочисленных неотрицательных дискретных случайных величин.

Пусть X — случайная величина, принимающая только целые неотрицательные значения. Пусть эта случайная величина X имеет следующий закон распределения

X	0	1	$\dots$	n	$\dots$
P	p_0	p_1	$\dots$	p_n	$\dots$

Тогда производящей функцией этой случайной величины называется сумма ряда

$$f_X(z) = p_0 + p_1 z + \dots p_n z^n + \dots$$

Здесь z — некоторая формальная переменная.

Задача 398. Доказать, что этот ряд абсолютно сходится при $|z| < 1$. Указание. Воспользоваться мажорантным признаком сходимости Вейерштрасса.

Задача 399. Доказать, что $f_X(1) = 1$.

Задача 400. Доказать, что $\frac{1}{n!} f_X^{(n)}(0) = p_n$, в частности, $f_X(0) = p_0$.

Задача 401. Доказать, что знание производящей функции позволяет восстановить закон распределения случайной величины.

Задача 402. Доказать, что $f'_X(1) = MX$.

Задача 403. Доказать, что $f''_X(1) + f'_X(1) = MX^2$.

Задача 404. Доказать, что $f''_X(1) + f'_X(1) - [f'_X(1)]^2 = DX$.

Как видно из приведенных задач, знание производящей функции позволяет восстановить закон распределения случайной величины. Кроме того, с использованием производящей функции можно вычислить математическое ожидание, дисперсию и другие числовые характеристики случайной величины, причем часто намного проще, чем напрямую.

Задача 405. Доказать, что производящую функцию для неотрицательных целочисленных случайных величин можно определить формулой

$$f_X(z) = M(z^X).$$

Задача 406. Пусть X_1 и X_2 — две независимые целые неотрицательные случайные величины и $X = X_1 + X_2$ — их сумма. Доказать, что тогда $f_X(z) = f_{X_1}(z) \cdot f_{X_2}(z)$.

Решение

Для независимых случайных величин справедлива формула

$$P(X = n) = P(X_1 = 0)P(X_2 = n) + P(X_1 = 1)P(X_2 = n - 1)+$$

$$\cdots + P(X_1 = n - 1)P(X_2 = 1) + P(X_1 = n)P(X_2 = 0),$$

Далее следует восполбзоваться формулами для коэффициентов произведения степенных рядов

Задача 407. Пусть случайная величина X принимает значение 1 с вероятностью p и значение 0 с вероятностью $q = 1 - p$ (то есть X — число успехов в единичном испытании). Найти производящую функцию. Используя полученный результат, найти производящую функцию распределения Бернулли, математическое ожидание и дисперсию.

Решение

По определению, производящая функцию X равна $f(z) = pz + q$. Пользуясь этой формулой и свойством 6, получим, что производящая функция для случайной величины, подчиняющейся распределению Бернулли (число успехов в n независимых испытаниях), вычисляется по формуле

$$f_n(z) = (pz + q)^n.$$

Используя свойства 3 и 5, получим, что математическое ожидание и дисперсию для распределения Бернулли равны соответственно

$$MX = np, \quad DX = npq.$$

Задача 408. Доказать, что производящая функция для случайной величины, подчиняющейся распределению Пуассона с параметром λ, вычисляется по формуле

$$f_{Poisson}(z) = e^{\lambda(z-1)}.$$

Вывести отсюда, что математическое ожидание и дисперсию для распределения Пуассона равны соответственно

$$MX = \lambda, \quad DX = \lambda.$$

Задача 409. Доказать, что производящая функция для случайной величины, подчиняющейся геометрическому распределению с параметром p, вычисляется по формуле

$$f_{geom}(z) = \frac{p}{1 - qz}.$$

Вывести отсюда, что математическое ожидание и дисперсия для геометрического распределения равны соответственно

$$MX = \frac{q}{p}, \quad DX = \frac{q}{p^2}.$$

Задача 410. Проводится n независимых испытаний, вероятности успеха в них равны соответственно $p_1, \dots p_n$, а вероятности неудач соответственно $q_1, \dots q_n$. Найти производящую функцию для числа успехов в этих испытаниях.

Задача 411. Орудие стреляет по цели до трех попаданий. Сначала вероятность попадания при каждом выстреле равна 0,6. После первого попадания цель начинает защищаться, и вероятность попадания при единичном выстреле падает до 0,3. Найти математическое ожидание числа потраченных снарядов.

30. Характеристические функции

Для непрерывных случайных величин вводятся *характеристические функции.* Если $\rho(x)$ — плотность вероятности случайной величины, то характеристической функцией этой случайной величины называется функция новой переменной t:

$$\phi(t) = \int\limits_{-\infty}^{\infty} e^{itx}\rho(x)dx.$$

Здесь i — мнимая единица, то есть это комплекснозначная функция. По сути дела, это преобразование Фурье плотности вероятности. Отличие только в знаке показателя.

Задача 412. Доказать, что $\phi_X(t)$ является непрерывной ограниченной функцией.

Задача 413. Доказать, что $\phi_X(0) = 1$.

Задача 414. Доказать, что характеристическую функцию случайной величины X можно определить формулой

$$\phi_X(t) = M(e^{itX}).$$

Задача 415. Доказать, что характеристическая функция суммы независимых случайных величин равна произведению характеристических функций слагаемых.

Решение

Если плотности вероятности двух случайных величин равны $\rho_1(x)$ и $\rho_2(x)$ соответственно, то можно записать, чему равна плотность вероятности их суммы:

$$\rho(x) = \int\limits_{-\infty}^{\infty} \rho_1(t)\rho_2(x-t)dt.$$

В математике это называется *сверткой* функций $\rho_1(x)$ и $\rho_2(x)$. Теперь доказываемое утверждение следует из хорошо известного факта: преобразование Фурье переводит свертку в произведение.

Задача 416. Доказать, что плотность вероятности случайной величины однозначно восстанаваливается по ее характеристической функции:

Решение

$$\rho(x) = \frac{1}{2\pi}\int\limits_{-\infty}^{\infty} e^{-itx}\phi(t)dt.$$

Это — не что иное, как формула обращения преобразования Фурье.

Задача 417. Доказать, что математическое ожидание и дисперсию можно вычислить с помощью характеристических функций по формулам:

$$MX = -i\phi'_X(0), \quad MX^2 = -\phi''_X(0),$$
$$DX = -\phi''_X(0) + (\phi'_X(0))^2.$$

Задача 418. Доказать, что, если X — случайная величина, равномерно распределенная на отрезке $[a, b]$, то ее характеристическая функция равна

$$\phi_X(t) = \frac{e^{ibt} - e^{iat}}{it(b-a)}.$$

Задача 419. Доказать, что, если X — случайная величина, имеющая показательное распределение $\rho(x) = \lambda e^{-\lambda x}$, то ее характеристическая функция равна

$$\phi_X(t) = \frac{\lambda}{\lambda - it}.$$

Задача 420. Доказать, что, если X — случайная величина, имеющая нормальное распределение со средним a и дисперсией σ^2, то ее характеристическая функция равна

$$\phi_X(t) = e^{ita - \frac{\sigma^2 t^2}{2}}.$$

Задача 421. Пользуясь результатом предыдущей задачи, докажите, что сумма двух независимых нормально распределенных случайных величин также распределена нормально.

Задача 422. Найти характеристическую функцию распределения Коши

$$\rho(x) = \frac{1}{\pi}\cdot\frac{1}{1+x^2}.$$

31. Многомерное нормальное распределение

Плотность вероятностей многомерного нормального распределения на n-мерном векторе

$$\vec{x} = \begin{pmatrix} x_1 \\ \dots \\ x_n \end{pmatrix}$$

задается формулой

$$\rho(\vec{x}) = \frac{1}{\left(\sqrt{2\pi}\right)^n \sqrt{\det(B)}} \cdot \exp\{-\frac{1}{2}(\vec{x}-\vec{a})^t B^{-1}(\vec{x}-\vec{a})\},$$

где

$$\vec{a} = \begin{pmatrix} a_1 \\ \dots \\ a_n \end{pmatrix}$$

— вектор средних значений компонент,

$$B = \begin{pmatrix} b_{11} & \dots & b_{1n} \\ \dots & \dots & \dots \\ b_{n1} & \dots & b_{nn} \end{pmatrix}$$

— симметричная положительно определенная матрица, называемая ковариационной матрицей. Как принято, верхний индекс t означает операцию транспонирования, B^{-1} — обратная матрица, $\det(B)$ — определитель. Напомним, что выражение в показателе экспоненты называется квадратичной формой.

Часто многомерное нормальное распределение возникает в задачах, связанных с измерением. Тогда компоненты случайного вектора — это результаты измерений, возможно, различных величин, причем эти результаты получены при совместных измерениях. Тогда величины могут быть связаны между собой, то есть коррелированы.

Задача 423. Доказать, что параметры $a_1, \dots a_n$ — это средние значения компонент.

Задача 424. Доказать, что коэффициент b_{ij} — это коэффициент ковариации между i-м и j-м компонентами. В частности, диагональные элементы матрицы B — дисперсии компонент.

Задача 425. Доказать, что если случайный вектор имеет многомерное нормальное распределение, то можно подобрать такую замену переменных, что в этих новых переменных компоненты случайного вектора станут независимыми. Более того, такую замену можно выбрать ортогональной.

Решение

Воспользуемся теперь известной теоремой из линейной алгебры — теоремой о приведении к главным осям. Она утверждает, что для любой квадратичной формы найдется ортонормированныый базис, в котором ее матрица диагональна.

Пусть $y_1, \dots y_n$ — координаты, в которых эта матрица диагональна. Допустим также, что все средние значения компонент y_i равны нулю — этого тоже нетрудно достичь линейным преобразованием. Тогда в этих координатах формула для плотности вероятности примет вид:

$$\rho(\vec{x}) = \frac{1}{\left(\sqrt{2\pi}\right)^n \sigma_1 \cdot \dots \cdot \sigma_n} \cdot \exp\{-\frac{1}{2}\Big(\frac{y_1^2}{\sigma_1^2} + \dots \frac{y_n^2}{\sigma_n^2}\Big)\},$$

Здесь σ_i^2 — дисперсии компонент.

Мы видим, что все ковариации между компонентами равны нулю, то есть новые переменные — некоррелированны. Но в правой части последней формулы, как легко видеть, стоит произведение плотностей вероятностей нормально распределенных величин. Таким образом, как установлено в задаче 370, компоненты независимы.

Из этого решения следует, что для многомерного нормального распределения некоррелированность и независимость — эквивалентны.

Задача 426. (Теорема об ортогональном преобразовании). Пусть случайные величины $X_1, \dots X_n$ имеют стандартное нормальное распределение и независимы. Пусть мы случайный вектор с этими компонентами подвергаем ортогональному преобразованию с матрицей Q, то есть получаем новые случайные величины $Y_1, \dots Y_n$ по формуле

$$\begin{pmatrix} Y_1 \\ \dots \\ Y_n \end{pmatrix} = Q \begin{pmatrix} X_1 \\ \dots \\ X_n \end{pmatrix}.$$

Доказать, что тогда эти случайные величины $Y_1, \dots Y_n$ также имеют стандартное нормальное распределение и независимы.

Решение

Плотность вероятности случайного вектора $(X_1, \dots X_n)^t$ в точке $\vec{x} = (x_1, \dots x_n)^t$ равна

$$\rho(\vec{x}) = \frac{1}{\left(\sqrt{2\pi}\right)^n} \cdot \exp\{-\frac{1}{2}\left(x_1^2 + \dots x_n^2\right)\}.$$

Мы можем считать, что преобразованный случайный вектор $(Y_1, \dots Y_n)$ — это тот же вектор, просто записанный в новых координатах, а матрица B — это матрица ортогональной замены базиса. Тогда для преобразованного случайного вектора $(Y_1, \dots Y_n)$ плотность вероятности в точке с координатами $\vec{y} = B\vec{x}$, разумеется, такая же (поскольку это та же самая точка). Но ортогональное преобразование сохраняет длины, то есть

$$y_1^2 + \dots + y_n^2 = x_1^2 + \dots + x_n^2.$$

Поэтому плотность вероятности случайного вектора $(Y_1, \dots Y_n)^t$ в точке $\vec{y} = (y_1, \dots y_n)^t$ также равна

$$\rho(\vec{y}) = \frac{1}{\left(\sqrt{2\pi}\right)^n} \cdot \exp\{-\frac{1}{2}\left(y_1^2 + \dots y_n^2\right)\}$$

и, как и ранее, распадается в произведение одномерных нормальных плотностей, что и требовалось доказать.

Задача 427. Плотность вероятности трехмерного нормального вектора равна $\rho(x, y, z)$. Каким условиям должны удовлетворять собственные числа ковариационной матрицы, чтобы поверхностями уровня этой функции были
— эллипсоиды;
— эллипсоиды вращения;
— сферы?

Задача 428. Случайный вектор X имеют многомерное нормальное распределение с ковариационной матрицей B. Пусть C — такая симметрическая матрица, что $B = C^2$. Доказать, что компоненты случайного вектора $Y = C^{-1}X$ имеют стандартное нормальное распределение и независимы.

Задача 429. Компоненты случайного вектора (X, Y) имеют стандартное нормальное распределение и независимы. Какова вероятность того, что этот вектор внутри единичного круга?

Задача 430. Компоненты n-мерного случайного вектора имеют стандартное нормальное распределение и независимы. Найти закон распределения проекции этого вектора на заданную прямую, проходящую через начало координат.

Задача 431. Компоненты случайного вектора (X, Y) имеют стандартное нормальное распределение и независимы. Какое распределение имеет случайная величина $\operatorname{arctg}(X/Y)$?

Задача 432. Компоненты двумерного случайного вектора $\vec{X}$ имеют стандартное нормальное распределение и независимы. Какое распределение имеет случайная величина $(\vec{X}, \vec{a})/(\vec{X}, \vec{b})$, где $\vec{a}$ и $\vec{b}$ — два различных заданных ненулевых вектора?

Задача 433. Что изменится в предыдущей задаче, если векторы $\vec{X}$, $\vec{a}$ и $\vec{b}$ будут не двумерные, а n-мерные?

Задача 434(*). Доказать, что если двумерный случайный вектор имеет независимые компоненты и его распределение изотропно, то он распределён нормально

32. Распределения хи-квадрат, Стьюдента, Фишера

Если независимые случайные величины $X_1, \ldots X_n$ независимы и имеют стандартное нормальное распределение, то можно с их помощью определить новые случайные величины, играющие важную роль в математической статистике.

Рассмотрим случайную величину

$$Z = X_1^2 + \cdots + X_n^2.$$

Распределение, которому подчиняется такая случайная величина, называется распределением χ^2 (хи-квадрат) с n степенями свободы. Математическое ожидание и дисперсия для распределения χ^2 равны соответственно:

$$MZ = n; \quad DZ = 2n.$$

Если случайная величина Z распределена по χ^2 с n степенями свободы, то обычно используется сокращенное обозначение $Z \sim \chi^2(n)$.

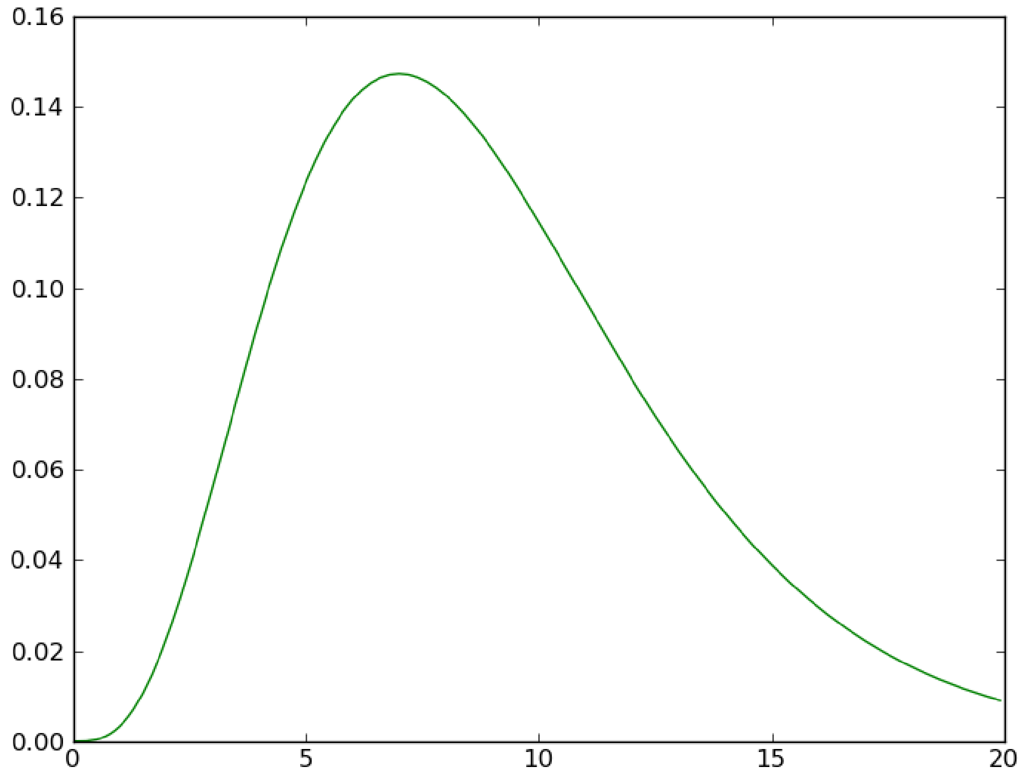


Рисунок 12. Плотность распределения χ^2 с 9 степенями свободы.

Пусть X и Z — независимые случайные величины, причем X распределена стандартно нормально, а Z распределена по χ^2 с n степенями свободы. Образуем новую случайную величину:

$$T = \frac{X}{\sqrt{Z/n}}.$$

Такую случайную величину называют распределенной по Стьюденту с n степенями свободы. Математическое ожидание и дисперсия для распределения Стьюдента равны

$$MT = 0; \quad DT = \frac{n}{n-2},$$

если же $n \leqslant 2$, то дисперсия бесконечна.

Если случайная величина T распределена по Стьюденту с n степенями свободы, то обычно используется сокращенное обозначение $T \sim t(n)$.

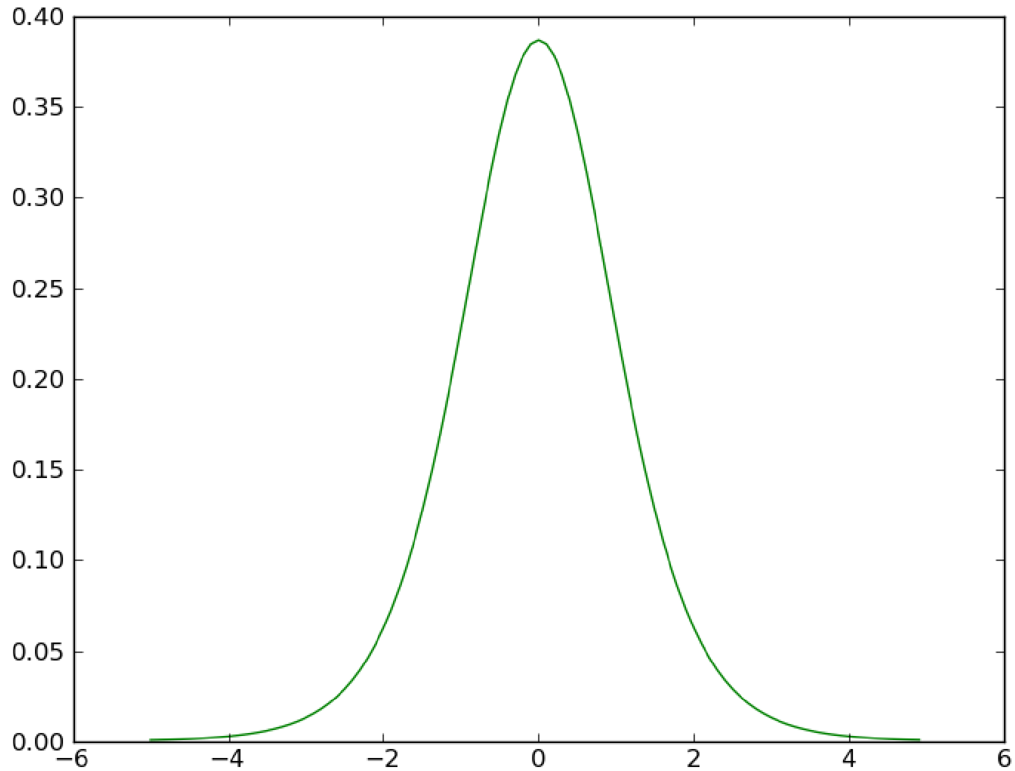


Рисунок 13. Плотность распределения Стьюдента с 8 степенями свободы.

Еще одно важное распределение называется распределением Фишера. Пусть Z_m и Z_n — две независимые случайные величины, распределенные по закону χ^2 с m и n степенями свободы соответственно. Тогда случайная величина

$$F = \frac{Z_m/m}{Z_n/n}$$

называется распределенной по Фишеру с (m, n) степенями свободы. Для этой случайной величины

$$MF = \frac{n}{n-2}; \quad DF = \frac{2n^2(m+n-2)}{m(n-2)^2(n-4)}.$$

При $n \leqslant 4$ ее дисперсия бесконечна, а при $n \leqslant 2$ бесконечно и математическое ожидание.

Если случайная величина F распределена по Фишеру с (m, n) степенями свободы, то обычно используется сокращенное обозначение $F \sim F(m, n)$.

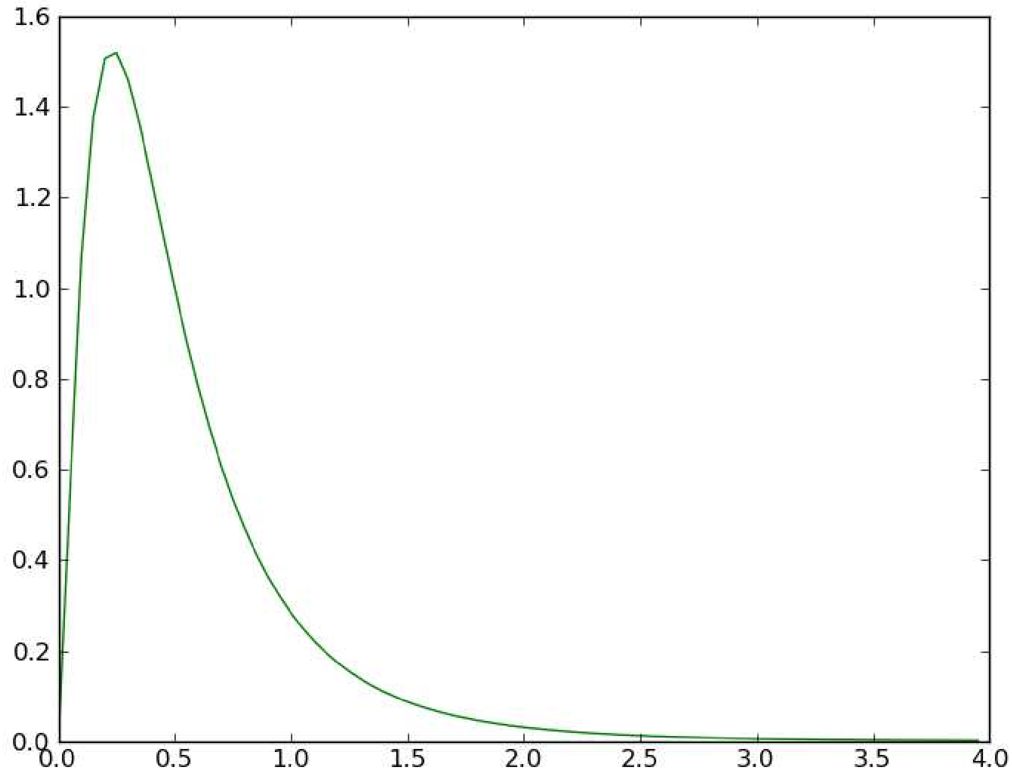


Рисунок 14. Плотность распределения Фишера с 5 степенями свободы числителя и 11 степенями свободы знаменателя.

Задача 435. Найти распределения квадрата случайной величины, распределенной по Стьюденту с n степенями свободы.

Задача 436. Доказать, что распределение хи-квадрат с 2 степенями свободы совпадает с показательным распределением.

Задача 437. Доказать, что сумма независимых случайных величин, имеющих распределение хи-квадрат, также распределена по хи-квадрат. Чему будет равно число степеней свободы у суммы?

Задача 438. Доказать, что распределения Стьюдента с 1 степенью свободы — распределение Коши с $\gamma = 1$ (задача 335).

Задача 439. Если случайная величина X имеет распределение Фишера с (m, n) степенями свободы, то как распределена случайная величина $1/X$?

Задача 440. Найти функцию распределения случайной величины, распределенной по χ^2 с одной степенью свободы.

Задача 441. Вывести формулы для математического ожидания и дисперсии распределения $\chi^2(n)$.
Указание. Воспользоваться результатом задачи 291.

Задача 442. Компоненты n-мерного случайного вектора распределены стандартно нормально и независимы. Найти закон распределения квадрата длины этого вектора.

33. Квантили. Использование электронных таблиц

Все три распределения, описанные в предыдущем разделе, широко применяются в математической статистике. Важной и часто встречающейся задачей является нахождение квантилей этих распределений.

В настоящее время для нахождения квантилей используются электронные таблицы. Встроенные функции для вычисления квантилей есть во всех достаточно распространенных офисных пакетах, как для Windows, так и для других платформ.

В таблице перечислены встроенные функции для расчета квантилей для наиболее распространенных фисных пакетов и платформ. Все эти встроенные функции находятся в разделе "статистические".

Офисный пакет	Хи-квадрат	Стьюдент	Фишер
Microsoft Office	ХИ2ОБР	СТЬЮДРАСПОБР	FРАСПОБР
Libre Office	CHIINV	TINV	FINV
Office for iPAD	ХИ2ОБР	СТЬЮДРАСПОБР	FРАСПОБР
Kingsoft Office	CHIINV	TINV	FINV

К сожалению, сразу работать с этими функциями вряд ли получится. Они, особенно в ранних версиях электронных таблиц, не только не соответствуют российскому ГОСТу, но и могут без предупреждения меняться от версии к версии. Поэтому для успешной работы с электронными таблицами следует разобраться, как именно работают встроенные статистические функции в вашей версии. Ниже приведена небольшая справочная таблица, которая позволит это понять.

Распределение	Ст. свободы	Вероятность p	p-квантиль
Стьюдента	5	0,95	2,015
Стьюдента	5	0,99	3,365
Стьюдента	10	0,98	2,359
Стьюдента	10	0,99	2,764
Хи-квадрат	5	0,95	11,070
Хи-квадрат	5	0,98	13,388
Хи-квадрат	10	0,98	21,161
Хи-квадрат	10	0,99	23,209
Фишера	5, 10	0,95	3,326
Фишера	5, 10	0,98	4,555
Фишера	12, 10	0,98	3,868
Фишера	12, 10	0,99	4,706

Задача 443. Пусть Z — α-квантиль распределения Фишера с (m, n) степенями свободы. Чему равен $(1-\alpha)$-квантиль распределения Фишера с (n, m) степенями свободы?

Задача 444. Пусть Q — p-квантиль распределения Фишера с (n, n) степенями свободы. Чему равен $(1-p)$-квантиль того же распределения?

Задача 445. Пусть Q — p-квантиль распределения Фишера с (m, n) степенями свободы. Чему равен $(1-p)$-квантиль распределения Фишера с (n, m) степенями свободы?

Задача 446. Найти вероятность того, что случайная величина, распределенная по Стьюденту с 5 степенями свободы, не превысит значения 2.

Задача 447. Найти вероятность того, что случайная величина, распределенная по хи-квадрат с 6 степенями свободы, не превысит значения 10.

Задача 448. Найти вероятность того, что случайная величина, распределенная по Фишеру с (5, 11) степенями свободы, не превысит значения 4.

Задача 449. Найти вероятность того, что случайная величина, распределенная по Стьюденту с 5 степенями свободы, не превысит значения (-2).

Задача 450. Найти вероятность того, что случайная величина, распределенная по хи-квадрат с 11 степенями свободы, не превысит значения 20.

Задача 451. Найти вероятность того, что случайная величина, распределенная по Фишеру с (12, 8) степенями свободы, не превысит значения 14.

Задача 452. Найти 96% квантиль распределения Стьюдента с 9 степенями свободы.

Задача 453. Найти 98% квантиль распределения Стьюдента с 7 степенями свободы.

Задача 454. Найти 99% квантиль распределения Стьюдента с 11 степенями свободы.

Задача 455. Найти 96% квантиль распределения хи-квадрат с 9 степенями свободы.

Задача 456. Найти 98% квантиль распределения хи-квадрат с 7 степенями свободы.

Задача 457. Найти 99% квантиль распределения хи-квадрат с 11 степенями свободы.

Задача 458. Найти 96% квантиль распределения Фишера с (4,5) степенями свободы.

Задача 459. Найти 98% квантиль распределения Фишера с (7,8) степенями свободы.

Задача 460. Найти 99% квантиль распределения Фишера с (18,12) степенями свободы.

Тема 4. Элементы математической статистики

34. Основные понятия статистики

Основные определения

Все возможные реализации случайной величины составляют *генеральную совокупность*. Генеральная совокупность может состоять из конечного или бесконечного числа возможностей. Те реализации, которые мы уже смогли наблюдать, составляют *выборку*. Основная задача математической статистики — по выборке сделать какие-то выводы о генеральной совокупности.

Выборка — всегда конечный набор значений. Обычно они обозначаются так: $x_1, \ldots x_n$. Число n называется объемом выборки.

Размахом выборки называется разность между ее максимальным и минимальным значеними.,

Видимо, самой важной характеристикой выборки является среднее значение. Его можно вычислить по формуле

$$\bar{x} = \frac{x_1 + \cdots + x_n}{n}.$$

В Microsoft Excel для вычисления среднего можно воспользоваться стандартной функцией СРЗНАЧ, а в Libre Office — функцией AVERAGE.

Эта величина называется *выборочным средним*. Также его называют выборочным средним арифметическим. По-видимому, чем больше наша выборка, тем больше это выборочное среднее похоже на истинное среднее всей генеральной совокупности.

Еще одним "средним" значением является *выборочная медиана.* все числа из выборки выстроим в порядке возрастания, медиана — то значение, которое стоит в самой середине. Скажем, если всего значений 31, то медиана — это 16-ое "по росту". А если всего значений 28, то тогда медиана — среднее между 14 и 15 значениями.

В Microsoft Excel для вычисления медианы используется стандартная функция МЕДИАНА, а в Libre Office — функция MEDIAN

Выборочная медиана может не совпадать с выборочным средним арифметическим (и чаще всего не совпадает), поэтому когда мы слышим слова “в среднем”, стоит понимать, какое именно среднее имеется в виду. Пример приведен чуть ниже, в задача 466.

Кроме среднего значения, генеральная совокупность характеризуется еще и разбросом относительно этого среднего. Для этого чаще всего используется *выборочная дисперсия.* Она вычисляется по формуле

$$s^2 = \frac{(x_1 - \bar{x})^2 + \cdots + (x_n - \bar{x})^2}{n-1}.$$

Для нахождения дисперсии в Microsoft Excel можно воспользоваться стандартной функцией ДИСП. Эта функция, впрочем, делит на n, а не на $n-1$, поэтому для получения правильного значения нужно результат умножить на дробь $n/(n-1)$.

В Libre Office для вычисления выборочной дисперсии используется стандартная функция VAR, дающая правильное значение. Есть еще функция VARP, аналогичная ДИСП.

Дисперсия — это не средний разброс, а квадрат среднего разброса. Чтобы получить разброс, нужно из дисперсии извлечь квадратный корень, и тогда получится *выборочное стандартное отклонение.* Его еще часто называют среднеквадратическим отклонением. Для его вычисления в Microsoft Office можно воспользоваться стандартной функцией СТАНДОТКЛОН, а в Libre Office — функцией STDEV.

Для характеристик совокупности используется слово “выборочная”. Это слово означает, что значение получено только по данным, представленным в выборке. Любое значение, полученное по данным выборки, любую функцию элементов выборки принято называть *статистикой.*

Задача 461. Приведите пример выборки (набора чисел), для которого среднее арифметическое не совпадает с медианой.

Задача 462. По заданной выборке:
$1, 5, 7, 12, 1, 13, 17, 4, 10, 15, 10, 16, 4, 4, 19, 19, 3, 14, 14, 18, 20, 19$
найти медиану, среднее арифметическое и несмещенную оценку дисперсии.

Задача 463. По заданной выборке:
$20, 6, 7, 8, 20, 2, 3, 17, 4, 16, 1, 10, 8, 2, 3, 4, 5, 12, 18, 15, 7$
найти медиану, среднее арифметическое и несмещенную оценку дисперсии.

Задача 464. По заданной выборке:
$4, 8, 10, 15, 4, 16, 20, 7, 13, 18, 13, 19, 7, 7, 22, 22, 6, 17, 17, 21, 23, 22$
найти медиану, среднее арифметическое и несмещенную оценку дисперсии.

Задача 465. На сайте Викиновости *http://ru.wikinews.org/* приведены средние суточные температуры в феврале 2013 года в Москве.

01	02	03	04	05	06	07
-2,3	0,3	0,6	-1,3	-4,0	0,4	-0,1
08	09	10	11	12	13	14
-1,8	-1,2	0,1	1,2	0,2	-1,8	-4,0
15	16	17	18	19	20	21
-6,3	-8,8	-7,1	-7,9	-8,6	-9,1	-10,6
22	23	24	25	26	27	28
-9,2	-7,7	-5,2	-2,1	-1,4	-0,1	1,3

Найти средние и разброс выборки.

Решение

Электронные таблицы дают такой ответ:
— средняя арифметическая суточных температур в феврале 2013 года составила -3,45 градуса;
— медиана составила -1,95 градуса.

Задача 466. В деревне Большие Ухабы проживает 20 человек. Из них 19 имеют среднемесячный доход 5 тысяч рублей, а один — 500 тысяч рублей. Каков средний доход жителей деревни Большие Ухабы? Как изменится этот средний доход, если самый богатый житель будет получать не 500, а 900 тысяч рублей?

Задача 467. Все предприятия России обязаны ежеквартально сдавать утвержденную статистическую отчетность. Форма Государственной статистической отчетности П-4 “СВЕДЕНИЯ О ЧИСЛЕННОСТИ, ЗАРАБОТНОЙ ПЛАТЕ И ДВИЖЕНИИ РАБОТНИКОВ” содержит следующие подпункты:
— Средняя численность работников за отчетный месяц;
— Фонд начисленной заработной платы работников за отчетный месяц.
Какое среднее значение можно получить на основании этих данных?

Гистограммы

Выборочные среднее и дисперсия — далеко не вся информация, которую можно извлечь из выборки. Еще одной важной вещью будет приблизительное представление о том, как выглядит распределение. Для этого используют способ наглядного представления данных — построение гистограммы.

Построение гистограммы начинается с построения группированной выборки. Все данные из выборки разбиваются на несколько интевалов, после чего подсчитывается, сколько значений лежит в каждом интервале. Таким образом будет получена группированная выборка вида

$c_0 - c_1$	$\dots$	$c_{n-1} - c_n$
k_1	$\dots$	k_n

Здесь
$c_0, \dots c_n$ — границы интервалов разбиения;
$k_1, \dots k_n$ — число наблюдений в интевалах;
$n = k_1 + \dots + k_n$ — общее число наблюдений (объем выборки).

Задача 468. Построить гистограмму распределения февральских температур, взяв интервалы от -12 до -9, от -9 до -6, и так далее.

Решение

Подсчитав, сколько раз были именно такие температуры, получим следующую группированную выборку:

$-12 - -9$	$-9 - -6$	$-6 - -3$	$-3 - 0$	$0 - 3$
3	6	3	9	7

Отметим на горизонтальной оси 5 отрезков: от (-12) до (-9), от (-9)-до (-6), и так далее до последнего интервала от 0 до 3. Над каждым интервалом изобразим столбик такой высоты, склолько наблюдений попело в этот интервал. Полученный график и называется гистограммой. Она изображена на рисунке 15.

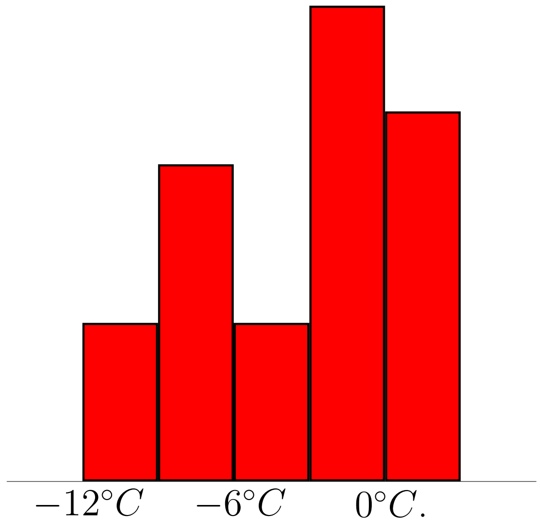


Рисунок 15. Гистограмма температур воздуха в Москве в феврале 2013 года

По группированной выборке можно построить другие оценки математического ожидания и дисперсии. Для этого мы просто считаем, что все значения выборки, попавшие в интервал, равны середине этого интервала.

Тогда мы получаем такие формулы

$$\bar{x}_{\text{групп}} = \frac{(c_0 + c_1)k_1 + \cdots + (c_{n-1} + c_n)k_n}{2n}$$

$$s^2_{\text{групп}} = \frac{(0,5(c_0 + c_1) - \bar{x}_{\text{групп}})^2 k_1 + \cdots + (0,5(c_{n-1} + c_n) - \bar{x}_{\text{групп}})^2 k_n}{n-1}.$$

Задача 469. Найти выборочные среднее и дисперсию распределения февральских температур по группированной выборке.

Решение

Для февральской погоды получим

$$\bar{x}_{\text{групп}} = 3,32; \quad s^2_{\text{групп}} = 16,89.$$

Задача 470. Пусть дана следующая выборка.

3,89	1,22	2,47	1,43	4,05	1,56
1,90	5,94	7,07	1,56	3,12	8,45
0,94	0,90	1,10	0,26	1,12	0,84
0,75	0,21	0,73	3,15	0,11	2,65
1,60	7,80	1,03	0,12	2,67	0,39

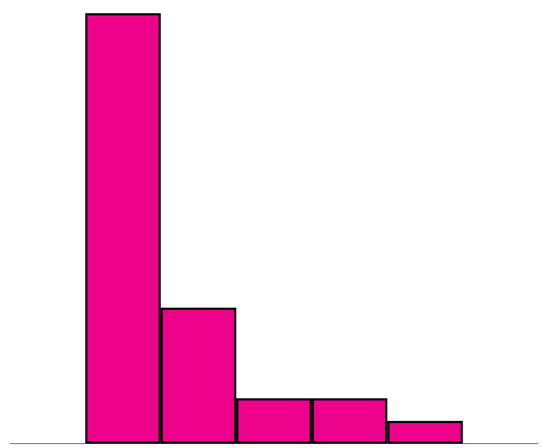
Построить гистограмму, выбрав отрезки 0 — 2, 2 — 4, и так далее до 8 — 10. Вычислить выборочные среднее и дисперсию распределения, а также выборочные среднее и дисперсию распределения по этой группированной выборке.

Рещение

Получим такую группированную выборку

0 — 2	2 — 4	4 — 6	6 — 8	8 — 10
19	6	2	2	1

Построим гистограмму по этой выборке.



Оказывается, мы имеем дело с сильно асимметричной выборкой, рисунок позволяет судить об особенностях этих данных. Вычисления показывают, что
— выборочное среднее $\bar{x} = 2,301$;
— выборочная медиана $m = 1,495$;
— выборочное среднее по группированной выборке $\bar{x}_{\text{групп}} = 2,33$;
— выборочная дисперсия $s^2 = 5,241$;
— выборочная дисперсия по группированной выборке $s^2_{\text{групп}} = 4,782$.

В задачах 471 — 474 следует по заданной выборке вычислить средние и дисперсии, построить группированную выборку, а по ней — гистограмму.

Задача 471.

9,46	4,57	11,31	10,24	11,08	5,97
13,43	9,39	15,67	13,21	6,16	5,88
16,56	12,46	10,15	8,17	7,84	9,38
11,96	8,29	6,21	8,25	6,93	3,16
9,05	7,94	8,09	9,63	11,69	3,53

Задача 472.

6,21	5,17	5,82	10,59	9,64	8,38
5,98	2,69	1,43	1,50	1,79	7,47
4,30	5,16	11,48	1,52	4,55	8,35
1,01	8,32	4,47	4,58	8,80	6,59
9,71	3,97	0,97	16,41	8,72	0,84

Задача 473.

5,78	4,37	5,23	15,07	12,56	9,68
5,45	1,99	1,24	1,27	1,42	7,89
3,38	4,36	17,67	1,28	3,65	9,62
1,04	9,57	3,56	3,67	10,60	6,38
12,75	3,05	1,02	36,82	10,43	0,97

Задача 474.

7,90	1,44	12,66	9,72	11,98	2,54
20,09	7,75	30,45	19,23	2,73	2,45
35,25	16,44	9,50	5,41	4,88	7,72
14,72	5,61	2,78	5,54	3,60	0,75
7,03	5,04	5,27	8,27	13,85	0,89

35. Метод максимального правдоподобия

Если есть основания считать, что случайная величина имеет заданное распределение, то достаточно по выборке оценить параметры этого распределения. Такой подход принято называть *параметрической статистикой.*

Для оценки параметров распределения часто используется *метод макси-*

мального правдоподобия. Он заключается в следующем. Пусть нам дана выборка $x_1, \ldots, x_n$, из n величин, подчиняющихся заданному распределению с неизвестным параметром θ. Этот параметр может быть как числом, так и вектором, то есть набором чисел. Требуется оценить этот параметр.

Первое упоминание об этом методе уже встречалось в задаче 60.

Поскольку вид распределения известен, можно записать вероятность того, что в n независимых реализациях случайной величины будут получены значения $x_1, \ldots, x_n$, как функционал от неизвестного пока параметра θ:

$$G(x_1, \ldots, x_n|\theta) = P(x_1|\theta) \cdot \dots \cdot P(x_n|\theta).$$

Этот функционал принято называть функционалом правдоподобия (likelihood). Для непрерывных случайных величин вместо вероятности надо использовать плотность вероятностей.

То значение параметра θ, которое дает максимальное значение функционала правдоподобия, называется *оценкой максимального правдоподобия*. Эту оценку часто можно получить, просто продифференцировав функционал. Сразу добавим, что часто упрощает задачу переход к логарифмическому функционалу правдоподобия по формуле

$$L(x_1, \ldots, x_n|\theta) = \ln G(x_1, \ldots, x_n|\theta),$$

поскольку у логарифма функции максимум находится там же, где и у самой функции.

Задача 475. Найти оценку максимального правдоподобия для параметра p распределения Бернулли.

Решение

Для распределения Бернулли обычно требуется оценить неизвестную вероятность p успеха в одном испытании. Пусть было проведено n испытаний, и в них было достигнуто k успехов. Вероятность этого равна

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

Это и есть функционал правдоподобия для нашего случая. Как и предлагалось ранее, для упрощения задачи перейдем к логарифмическому функционалу:

$$L(k|p) = C + k \ln p + (n-k) \ln (1-p).$$

Здесь сразу написано C вместо $\ln C_n^k$, поскольку этот член все равно пропадет при дифференцировании. Дифференцируем:

$$\frac{dL(k|p)}{dp} = \frac{k}{p} - \frac{n-k}{1-p} = \frac{k-np}{p(1-p)}.$$

Производная обращается в нуль при

$$p = \frac{k}{n}.$$

Нетрудно доказать, что это действительно максимум функционала правдоподобия. Таким образом, полученная оценка вероятности p является оценкой максимального правдоподобия.

Задача 476. Найти оценку максимального правдоподобия для параметра показательного распределения.

Решение

Для показательного распределения функцционал правдоподобия выглядит так:

$$G(x_1, \dots, x_n|\lambda) = \lambda^n \cdot e^{-\lambda(x_1+\dots+x_n)}.$$

Переходя к логарифмам и дифференцируя, получаем

$$\frac{1}{\lambda} = \frac{x_1 + \dots + x_n}{n}.$$

Задача 477. Найти оценку максимального правдоподобия для параметров нормального распределения.

Решение

Для нормального распределения нужно определить два неизвестных параметра — математическое ожидание a и стандартное отклонение σ. Функционал правдоподобия имеет вид:

$$G(x_1 \dots x_n|a,\sigma) = \frac{1}{(2\pi)^{n/2}\sigma^n} \exp\Big(-\frac{(x_1-a)^2 + \dots + (x_n-a)^2}{2\sigma^2}\Big).$$

Логарифм этого функционала равен, с точностью до слагаемого, не зависящего ни от a, ни от σ:

$$L(x_1, \dots, x_n|a,\sigma) = -n\ln\sigma - \frac{(x_1-a)^2 + \dots + (x_n-a)^2}{2\sigma^2}.$$

Частные производные равны

$$\frac{\partial L}{\partial a} = \frac{1}{\sigma^2}((x_1 - a) + \cdots + (x_n - a));$$

$$\frac{\partial L}{\partial \sigma} = -\frac{n}{\sigma} + \frac{(x_1 - a)^2 + \cdots + (x_n - a)^2}{\sigma^3}.$$

Приравнивая обе эти производные к нулю, получаем оценки максимального правдоподобия:

$$a = \frac{x_1 + \cdots + x_n}{n};$$

$$\sigma^2 = \frac{(x_1 - a)^2 + \cdots + (x_n - a)^2}{n}.$$

Задача 478. Найти оценку максимального правдоподобия для параметров равномерного распределения.

Решение

Для равномерного распределения также определить два неизвестных параметра — a и b — левую и правую границу интервала. Здесь, в отличие от предыдущих случаев, метод максимального правдоподобия приводит к не очень удачным оценкам. Поскольку плотность вероятности для равномерного распределения обратно пропорциональна длине отрезка, максимальной она будет, если взять самый маленький из возможных отрезков. Это приводит к оценкам

$$a = \min(x_1, \ldots, x_n); b = \max(x_1, \ldots, x_n).$$

Ясно, что такие оценки будут похожи на правильные только при больших n.

Задача 479. Получена выборка $k_1, \ldots k_n$, подчиняющаяся распределению Пуассона. Найти оценку максимального правдоподобия для неизвестного параметра λ.

Задача 480. Распределение Вейбулла. Функция распределения задается формулой

$$F(x) = 1 - \exp\left(-\left(\frac{x-a}{b}\right)^k\right),$$

а плотность вероятности — формулой

$$\rho(x) = \frac{k}{b^k}(x-a)^{k-1}\exp\left(-\left(\frac{x-a}{b}\right)^k\right).$$

Распределение Вейбулла зависит от трех параметров:
k — параметр формы кривой распределения;
b — параметр масштаба;
a — параметр сдвига (его, однако, часто принимают равным нулю).
Получена выборка $x_1, \ldots x_n$, подчиняющаяся распределению Вейбулла. Найти оценку максимального правдоподобия для неизвестных параметров k и b, считая, что параметр сдвига a равен нулю.

Задача 481. Распределение Парето. Оно задается плотностью вероятностей

$$\rho(x) = \frac{\alpha}{x_0}\left(\frac{x_0}{x}\right)^{\alpha+1}$$

при $x > x_0$. При $x < x_0$ плотность вероятности равна нулю.
Получена выборка $x_1, \ldots x_n$, подчиняющаяся распределению Парето. Найти оценку максимального правдоподобия для неизвестного параметра α, если параметр x_0 известен.
Доказать, что этим методом невозможно получить оценку для x_0, поскольку соответствующая частная производная в нуль не обращается.

Задача 482. Случайная величина X имеет геометрическое распределение, если $P(X = k) = pq^k$, где $q = 1 - p$, $k = 0, 1, \ldots$ Получена выборка $k_1, \ldots k_n$, подчиняющаяся геометрическому распределению. Найти оценку максимального правдоподобия для неизвестного параметра p.

Задача 483. В условиях предыдущей задачи оценка максимального правдоподобия может быть получена другим способом. Известно, что случайная величина, имеющая геометрическое распределение — не что иное, как число испытаний Бернулли до первого успеха. Таким образом, в условиях предыдущей задачи, достигнуто n успехов, при этом было проведено $k = k_1 + \cdots + k_n + n$ испытаний. Совпадают ли полученные оценки?

Задача 484. Распределение Максвелла, зависящее от параметра β, задается функцией плотности вероятностей

$$\rho(x) = C\beta^{3/2}x^2e^{-\frac{\beta x^2}{2}}.$$

Получена выборка $x_1, \ldots x_n$, подчиняющаяся этому распределению. Найти оценку максимального правдоподобия для неизвестного параметра β.

Задача 485. Логнормальное распределение, зависящее от двух параметров, a и σ, задается при $x > 0$ функцией плотности вероятностей

$$\rho(x) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{\ln x - a}{2\sigma^2}}.$$

Получена выборка $x_1, \dots x_n$, подчиняющаяся этому распределению. Найти оценки максимального правдоподобия для неизвестных параметров a и σ.

Задача 486. Распределение Рэлея, зависящее от параметра a, задается функцией плотности вероятностей

$$\rho(x) = \frac{x}{a^2} e^{-\frac{x^2}{2a^2}}.$$

Получена выборка $x_1, \dots x_n$, подчиняющаяся этому распределению. Найти оценку максимального правдоподобия для неизвестного параметра a.

Задача 487. Распределение Накагами, зависящее от двух параметров, a и ω, задается функцией плотности вероятности

$$\rho(x) = \frac{2a^a}{\Gamma(a)\omega^a} x^{2a-1} \exp\Big(-\frac{ax^2}{\omega}\Big).$$

Получена выборка $x_1, \dots x_n$, подчиняющаяся этому распределению. Считая параметр a известным, найти оценку максимального правдоподобия для неизвестного параметра ω.

36. Свойства оценок

Рассчитывая значение некоторого показателя по выборке по некоторой формуле, например, полученной методом максимального правдоподобия, мы получаем оценку этого показателя. Если же мы рассчитываем значение показателя многократно, то эта оценка становится случайной величиной.

Напомним, что любую функцию выборки принято называть статистикой. Таким образом статистика — случайная величина. Поэтому у статистики должны быть числовые характеристики: функция распределения, плотность вероятности, а также математическое ожидание и дисперсия.

Статистика, математическое ожидание которой совпадает с истинным значением оцениваемого параметра, называется *несмещенной*.

Несмещенная оценка, у которой дисперсия минимальна при заданном объеме выборки, называется *эффективной* оценкой.

Задача 488. Пусть есть выборка $x_1, \dots, x_n$ из случайных величин, имеющих одинаковое распределение со средним a и дисперсией σ^2. Доказать, что

$$\bar{x} = \frac{x_1 + \dots + x_n}{n}$$

— несмещенная оценка математического ожидания.

Задача 489. Доказать, что дисперсия оценки $\bar{x}$ из предыдущейй задачи равна

$$D\bar{x} = \frac{\sigma^2}{n}.$$

Задача 490. Доказать, что оценка дисперсии

$$\hat{s}^2 = \frac{(x_1 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n}$$

является смещенной и устранить смещение.

Решение

Найдем математическое ожидание этой статистики. Для этого преобразуем выражение

$$\hat{s}^2 = \frac{x_1^2 + \dots + x_n^2}{n} - \bar{x}^2.$$

Теперь воспользуемся формулой

$$DX = MX^2 - (MX)^2,$$

справедливой для любой случайной величины. Из нее следует, что

$$M(\bar{x}^2) = D\bar{x} + (M\bar{x})^2 = \frac{\sigma^2}{n} + a^2;$$

$$M(x_i^2) = Dx_i + (Mx_i)^2 = \sigma^2 + a^2.$$

Следовательно

$$M\hat{s}^2 = \sigma^2 + a^2 - \frac{\sigma^2}{n} - a^2 = \frac{n-1}{n}\sigma^2.$$

Таким образом, оценка $\hat{s}^2$ — смещенная, и для оценки дисперсии следует пользоваться формулой исправленной выборочной дисперсии

$$s^2 = \frac{(x_1 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n-1}.$$

Задача 491. Пусть $x_1, \ldots, x_n$ — выборка из нормального распределения с известным средним a. Доказать, что в этом случае оценка дисперсии

$$\tilde{s}^2 = \frac{(x_1 - a)^2 + \cdots + (x_n - a)^2}{n}$$

является несмещенной.

Задача 492. Пусть $x_1, \ldots, x_n$ — выборка из равномерного распределения на отрезке $[0, N]$ причем параметр N неизвестен. Для его оценки используем статистику

$$N_1 = 2 \cdot \frac{x_1 + \cdots + x_n}{n}.$$

Найти ее математическое ожидание и дисперсию и доказать, что она несмещенная.

Решение

Математическое ожидание и дисперсия для равномерного распределения нам известны: они равны

$$Mx_i = \frac{N}{2}; \quad Dx_i = \frac{N^2}{12}.$$

Следовательно, по формулам для математического ожидания и дисперсии суммы случайных величин получим:

$$MN_1 = N; \quad DN_1 = \frac{N^2}{12n}.$$

Введем обозначения $x(1), \ldots, x(n)$: это те же самые $x_1, \ldots, x_n$, только расположенные в порядке возрастания: $x(1) < \cdots < x(n)$. Вновь введенные величины $x(1), \ldots, x(n)$ называются *порядковыми статистиками.*

Задача 493. В условиях предыдущей задачи для оценки неизвестного параметра N используем статистику

$$N_2 = x(1) + x(n),$$

то есть сумму максимального и минимального чисел в выборке. Найти ее математическое ожидание и дисперсию и доказать, что она несмещенная.

Решение

Случайная величина $x(n)$ уже не будет, разумеется, распределена равномерно. Для нее, однако, можно найти функцию распределения. Действительно, событие $x(n) < x$ означает в точности то, что каждая из n случайных величин $x_1, \ldots, x_n$ меньше, чем x. Так как эти n случайных величин равномерно распределены и независимы, вероятность этого события равна $(x/N)^n$. Таким образом, функция распределения $x(n)$ равна

$$F_{x(n)}(x) = P(x(n) < x) = \left(\frac{x}{N}\right)^n;$$

плотность вероятности

$$\rho(x) = F'_{x(n)}(x) = \frac{n}{N}\left(\frac{x}{N}\right)^{n-1};$$

математическое ожидание

$$Mx(n) = \int\limits_0^N x\rho(x)dx = \frac{Nn}{n+1};$$

дисперсия

$$Dx(n) = \int\limits_0^N (x - Mx(n))^2\rho(x)dx = \frac{nN^2}{(n+1)^2(n+2)}.$$

Математическое ожидание и дисперсию первой порядковой статистики $x(1)$ можно легко найти таким образом: если на отрезке $[0; N]$ выбрано n случайных точек, то, проходя вдоль отрезка от 0 до N, мы сначала наткнемся на $x(1)$, а последней будет $x(n)$. Если же мы пойдем в обратном направлении, от N до 0, то на $x(n)$ мы наткнемся первой. Поэтому $x(1)$ распределена так же, как $N - x(n)$, а, следовательно, ее математическое ожидание и дисперсия равны:

$$Mx(1) = \frac{N}{n+1}; \quad Dx(1) = \frac{nN^2}{(n+1)^2(n+2)}.$$

Отсюда следует, что N_2 — также несмещенная оценка, а ее дисперсия равна

$$DN_2 = \frac{2nN^2}{(n+1)^2(n+2)}.$$

Задача 494. Статистика $x(n)$ является, очевидно, смещенной оценкой параметра N (см. две предыдущие задачи). Как устранить смещение и какова будет дисперсия у несмещенной оценки?

Решение

Поскольку в решении предыдущей задачи было установлено, что

$$Mx(n) = \frac{Nn}{n+1},$$

насмещенной оценкой будет :

$$N_3 = \frac{n+1}{n} x(n).$$

Ее дисперсия равна

$$DN_3 = \frac{(n+1)^2}{n^2} Dx(n) = \frac{N^2}{n(n+2)}.$$

Задача 495. Какая самая эффективная из оценок параметра N, получеeыx в трех предыдущих задачах?

Решение

Оценка N_3. Ее дисперсия минимальна. Действительно, при больших n $DN_1 \sim n^{-1}$, $DN_2 \sim n^{-2}$, а DN_3 почти в 2 раза меньше, чем DN_2.

Здесь стоит напомнить, что есть причина, почему оценка $x(n)$ с самого начала заслуживала внимательного рассмотрения: эта оценка была получена ранее в задаче 478 как оценка максимального правдоподобия.

Задача 496. Выборка $\{x_1, \ldots, x_n\}$ объема n взята из равномерного распределения на отрезке $[a-b, a+b]$, причем параметры a и b неизвестны. Для оценки середины отрезка a предлагаются такие статистики:

$$1)\frac{x_1 + \cdots + x_n}{n} \quad 2)\frac{1}{2}(\min(x_i) + \max(x_i)) \quad 3)\text{выборочная медиана.}$$

Доказать, что все оценки несмещенные и сравнить их дисперсии.

Задача 497. Выборка $\{x_1, \ldots, x_n\}$ объема n взята из равномерного распределения на отрезке $[a, a+1]$, причем параметр a неизвестен. Для его оценки предлагаются две статистики:

$$a_1 = \frac{x_1 + \cdots + x_n}{n} - \frac{1}{2} \text{ и } a_2 = \min(x_i) - \frac{n}{n+1}.$$

Доказать, что обе оценки несмещенные и сравнить их дисперсии.

Задача 498. Пусть из набора $\{1, \ldots, N\}$ взята бесповторная выборка $\{x_1, \ldots, x_n\}$ объема n. Как оценить неизвестный параметр N?

Решение

Пусть, как и ранее, $x(1) < \cdots < x(n)$ — порядковые статистики. Тогда, если $N \geqslant x(n)$, то вероятность получить из набора $\{1, \ldots, N\}$ нашу выборку одинакова для всех выборок объема n и равна

$$G = \frac{1}{C_N^n} = \frac{1}{N \cdot (N-1) \cdot \cdots \cdot (N-n+1)}.$$

Нетрудно сообразить, что, каково бы ни было n, эта функция — а это не что иное как функционал правдоподобия — является монотонно убывающей функцией от N. Поэтому оценкой максимального правдоподобия для неизвестного параметра N будет максимальная порядковая статистика $x(n)$.

Поскольку значение $x(n)$, скорее всего, будет меньше истинного значения N, это — смещенная оценка. Чтобы устранить смещение этой оценки, надо найти ее математическое ожидание. Займемся этим. Сначала вычислим вероятности:

$$P\bigl(x(n) \leqslant i\bigr) = P\bigl(x_1 \ldots x_n \leqslant i\bigr) = \frac{C_i^n}{C_N^n};$$

$$P\bigl(x(n) = i\bigr) = P\bigl(x(n) \leqslant i\bigr) - P\bigl(x(n) \leqslant i-1\bigr) = \frac{C_i^n - C_{i-1}^n}{C_N^n} = \frac{C_{i-1}^{n-1}}{C_N^n}.$$

По определению математического ожидания

$$M(x(n)) = \sum_{i=n}^{N} P(x(n) = i) \cdot i = \sum_{i=n}^{N} \frac{C_{i-1}^{n-1}}{C_N^n} \cdot i = \frac{1}{C_N^n} \sum_{i=n}^{N} C_{i-1}^{n-1} \cdot i = \frac{n}{C_N^n} \sum_{i=n}^{N} C_i^n.$$

Мы воспользовались формулой

$$C_{i-1}^{n-1} \cdot i = C_i^n \cdot n,$$

которую легко доказать. Далее мы воспользуемся формулой

$$\sum_{i=n}^{N} C_i^n = C_{N+1}^{n+1},$$

доказанной в задаче 39.

Продолжаем выкладки:

$$M(x(n)) = \frac{n}{C_N^n} \cdot C_{N+1}^{n+1} = \frac{n}{n+1} \cdot (N+1).$$

Теперь мы можем устранить смещение. Окончательно получаем, что несмещенная оценка неизвестного параметра N такова:

$$\hat{N} = \frac{n+1}{n} \cdot x(n) - 1.$$

Задача 499. Найти распределение k-й порядковой статистики из равномерного распределения на $[0,1]$ и ее математическое ожидание.

Задача 500. Пусть $x_1, \ldots, x_n$ — выборка из нормального распределения с известным средним a и неизвестной дисперсией σ^2.
a) Доказать, что оценка

$$T = \sqrt{\frac{\pi}{2}} \frac{1}{n} \sum_{i=1}^{n} |x_i - a|$$

— несмещенная оценка параметра σ.
b) (*) Найти дисперсию этой оценки.

Задача 501. Величина X оценивается по данным нескольких измерений этой величины $x_1, \ldots, x_n$, причем дисперсии x_i известны, равны σ_i^2 и, вообще говоря, могут быть различны. (Такая ситуация часто встречается в геодезии — так называемые неравноточные измерения). Строится оценка вида

$$\hat{X} = w_1 x_1 + \ldots + w_n x_n,$$

где w_i — некоторые положительные числа (веса).
a) Доказать, что условие несмещенности этой оценки:

$$w_1 + \ldots + w_n = 1.$$

b) Какие должны быть веса, чтобы эта оценка была эффективной (то есть имела наименьшую дисперсию)?

37. Проверка статистических гипотез

Статистические гипотезы проверяются на основе некоторото статистического критерия: некоторой случайной величины, имеющей, если гипотеза справедлива, некоторое заранее известное распределение. Общая схема проверки

статистических гипотез такова:
0). Допустим, что гипотеза верна.
1). Выбираем статистический критерий, имеющий, если 0) верно, заданное распределение.
2). Задаем доверительную вероятность $1 - \alpha$.
3). Выбираем область принятия гипотезы (ее вероятность $1 - \alpha$) и критическую область (ее вероятность α).
4). Вычисляем значение критерия.
5). Если критерий попадает в область принятия гипотезы — гипотеза принимается, если попадает в критическую область — отвергается.

При проверке гипотез могут появляться ошибки, причем эти ошибки могут быть разными:
1) гипотеза отвергается, хотя она верна;
2) гипотеза принимается, хотя она неверна.
Эти ошибки называются соответственно ошибками первого и второго рода.

Задача 502. Доказать (или хотя бы обосновать), что вероятность α — это вероятность ошибки первого рода.

Эта вероятность также называется уровнем значимости критерия

Проверка гипотез о значении p

В примечании перед задачей 344 упоминалось, что используется терминология из математической статистики. Напомним, как ставились эти задачи. Мы располагаем следующей информацией:
— было проведено n независимых испытаний;
— в них было достигнуто k успехов.
На основании этих данных проверяем гипотезу о том, что вероятность успеха в каждом испытаний равна p, при этом мы располагаем некоторым уровнем значимости α.

Тогда проверка заключалась в следующем.
1. Строился интервал, в который число успехов попадает с вероятностью $1-\alpha$. Во всех случаях это был центральный интервал, то есть вида $(np-z, np+z)$. Таким образом, если вероятность успеха действительно равна p, то число успехов почти наверняка, с вероятностью $1 - \alpha$, будет лежать в этом интервале.

2. Проверялось, попадает ли наблюдаемое значение k в построенный интервал. Если да, то делался вывод о том, что данные не противоречат гипотезе, если нет — гипотеза отвергалась.

Найденный интервал, в котором почти наверняка должно лежать число успехов, будет называться областью принятия гипотезы, а дополнение к этому интервалу — критической областью.

Задача 503. В $N = 450$ независимых испытаниях было получено $k = 246$ успехов. При уровне значимости $\alpha = 0.05$ проверить гипотезу о том, что вероятность успеха в каждом из испытаний $p = 0,5$.

Решение

Допустим, что гипотеза справедлива. Находим $np = 225$, $\sqrt{npq} = 10,61$. Для нахождения нужного интервала запишем уравнение с неизвестеым z:

$$P\Big(np - z < X < np + z) = 0,95.$$

Далее

$$P\Big(- \frac{z}{\sqrt{npq}} < \frac{X - np}{\sqrt{npq}} < \frac{z}{\sqrt{npq}}\Big) = 0,95;$$

$$2\Phi\Big(\frac{z}{\sqrt{npq}}\Big) - 1 = 0,95;$$

$$\Phi\Big(\frac{z}{\sqrt{npq}}\Big) = 0,975;$$

$$\frac{z}{\sqrt{npq}} = 1,96; z = 1,96 \cdot 10,61 = 20,79.$$

Таким образом, область принятия гипотезы — интервал от $225 - 20,79 = 204,21$ до $225 + 20,79 = 245,79$. Поскольку наблюдаемое значение $k = 246$ в этот интервал не попадает, на этом уровне значимости следует отвергнуть гипотезу $p = 0,5$.

В некоторых случаях ищется не центральный интервал, а какой-то другой, но имеющий ту же вероятность $1 - \alpha$. Это происходит в тех случаях, когда есть причины выдвинуть альтернативную гипотезу. Тогда интервал с той же вероятностью подбирается так, чтобы повысить вероятность этой альтернативной гипотезы и уменьшить таким образом вероятность ошибки второго рода.

Задача 504. В $N = 500$ независимых испытаниях было получено $k = 170$ успехов. При уровне значимости $\alpha = 0.05$ проверить гипотезу о том, что вероятность успеха в каждом из испытаний $p = 0,3$, если альтернативная гипотеза $H_1 : p > 0,3$.

Решение

Находим $np = 150$, $\sqrt{npq} = 10,25$. Чтобы уменьшить вероятность ошибки второго рода, ищем интервал от $-\infty$ до некоторого z, имеющий ту же вероятность $1 - \alpha$:

$$P\Big(X < z) = 0,95.$$

Далее

$$P\Big(\frac{X - np}{\sqrt{npq}} < \frac{z - np}{\sqrt{npq}}\Big) = 0,95;$$

$$\Phi\Big(\frac{z - np}{\sqrt{npq}}\Big) = 0,95;$$

$$\frac{z - np}{\sqrt{npq}} = 1,64; z = 150 + 1,64 \cdot 10,25 = 166,85.$$

Таким образом, область принятия гипотезы — интервал от $-\infty$ до $166,85$. Поскольку наблюдаемое значение $k = 170$ в этот интервал не попадает, на этом уровне значимости следует сделать выбор в пользу альтернативной гипотезы.

Задача 505. В $N = 900$ независимых испытаниях было получено $k = 300$ успехов. При уровне значимости $\alpha = 0.05$ проверить гипотезу о том, что вероятность p успеха в каждом из испытаний равна $p_0 = 0,3$, если альтернативная гипотеза $H_1 : p > p_0$.

Задача 506. В $N = 1200$ независимых испытаниях было получено $k = 450$ успехов. При уровне значимости $\alpha = 0.05$ проверить гипотезу о том, что вероятность p успеха в каждом из испытаний равна $p_0 = 0,33$, если альтернативная гипотеза $H_1 : p \neq p_0$.

Задача 507. В $N = 700$ независимых испытаниях было получено $k = 295$ успехов. При уровне значимости $\alpha = 0.02$ проверить гипотезу о том, что вероятность p успеха в каждом из испытаний равна $p_0 = 0,4$, если альтернативная гипотеза $H_1 : p > p_0$.

Задача 508. В $N = 500$ независимых испытаниях было получено $k = 289$ успехов. При уровне значимости $\alpha = 0.02$ проверить гипотезу о том, что вероятность p успеха в каждом из испытаний равна $p_0 = 0,6$, если альтернативная гипотеза $H_1 : p < p_0$.

Задача 509. В $N = 420$ независимых испытаниях было получено $k = 198$ успехов. При уровне значимости $\alpha = 0.01$ проверить гипотезу о том, что вероятность p успеха в каждом из испытаний равна $p_0 = 0,5$, если альтернативная гипотеза $H_1 : p \neq p_0$.

Задача 510. В $N = 400$ независимых испытаниях было получено $k = 103$ успеха. При уровне значимости $\alpha = 0.01$ проверить гипотезу о том, что вероятность p успеха в каждом из испытаний равна $p_0 = 0,3$, если альтернативная гипотеза $H_1 : p < p_0$.

38. Критерий согласия хи-квадрат

По заданной группированной выборке можно выдвинуть гипотезу о том, что данные подчиняются заданному распределению. Процедура проверки этой гипотезы носит названия критерия χ^2 (хи-квадрат). Она заключается в следующем.

0. Известны границы интервалов и числа n_i — число наблюдений, попавших в i–й интервал. Числа n_i принято называть экспериментальными частотами. Сумма всех n_i равна n — числу наблюдений, то есть объему выборки.

1. Рассчитываются оценки параметров распределения.

2. В случае необходимости интервалы в группированной выборке расширяются так, чтобы их объединение совпадало с носителем случайной величины. Так, если выполняется проверка на нормальность распределения, то интервалы должны охватывать всю числовую ось. Для этого крайний левый интервал продлевают до $-\infty$, а врайний правый — до $+\infty$.

3. Для каждого i-го интервала рассчитываются вероятности p_i того, что случайная величина примет значение в этом интервале. Полученные вероятности умножаются на общее число наблюдений n, таким образом рассчитываются np_i — теоретические частоты интервалов.

4. Идея критерия хи-квадрат состоит в том, что, если гипотеза справедлива,

то теоретические частоты не будут сильно оличаться от экспериментальных. Вычисляется значение выражения

$$\chi^2 = \sum_{i=1}^{m} \frac{(n_i - np_i)^2}{np_i}.$$

Можно доказать, что если гипотеза верна, то эта случайная величина при достаточно больших n приблизительно имеет распределение χ^2 с $m-1-r$ степенями свободы, где:
m — число интервалов;
r — число параметров, определяемых по выборке.

Задача 511. Дана группированная выборка.

Инт.	$0-2$	$2-4$	$4-6$	$6-8$	$8-10$	$10-12$	$12-14$
ЭЧ	12	51	179	298	170	48	9

Требуется на уровне значимости $\alpha = 0,05$ проверить гипотезу о том, соответствуют ли указанные экспериментальные частоты тому, что данные получены из нормальной генеральной совокупности.

Решение

Число наблюдение $n = 767$. Вычислим выборочные среднее и дисперсию:

$$\bar{x}_{\text{групп}} = 6,94; \quad s^2_{\text{групп}} = 4,87.$$

В нашем случае число интервалов $m = 7$. Поскольку для нормального распределения по выборке были найдены два параметра — математическое ожидание и дисперсия — то в нашем случае $r = 2$. Таким образом, число степеней свободы будет равно 4.

Для нормального распределения с вычисленными средним и дисперсией вычисляем теоретические частоты. Расширим крайние интервалы до $-\infty$ и $+\infty$ соответственно, чтобы все интервалы покрывали всю числовую ось. Продолжим таблицу, добавим в нее строку теоретических частот и строку величин $(n_i - np_i)^2/np_i$.

Инт.	$-\infty - 2$	$2 - 4$	$4 - 6$	$6 - 8$	$8 - 10$	$10 - 12$	$12 - +\infty$
ЭЧ	12	51	179	298	170	48	9
ТЧ	9,71	60.59	187.06	267.92	178.30	55.04	8.73
χ^2	0.54	1.52	0.35	3.38	0.39	0.90	0.05

Суммируя элементы последней строки, находим значение критерия

$$\chi^2 = 7,12.$$

Из таблиц находим значение 95% квантиля распределения хи-квадрат с 4 степенями свободы:

$$\chi^2_{0,95}(4) = 9,49.$$

Поскольку наблюдаемое значение меньше критического, нет оснований отвергать гипотезу.

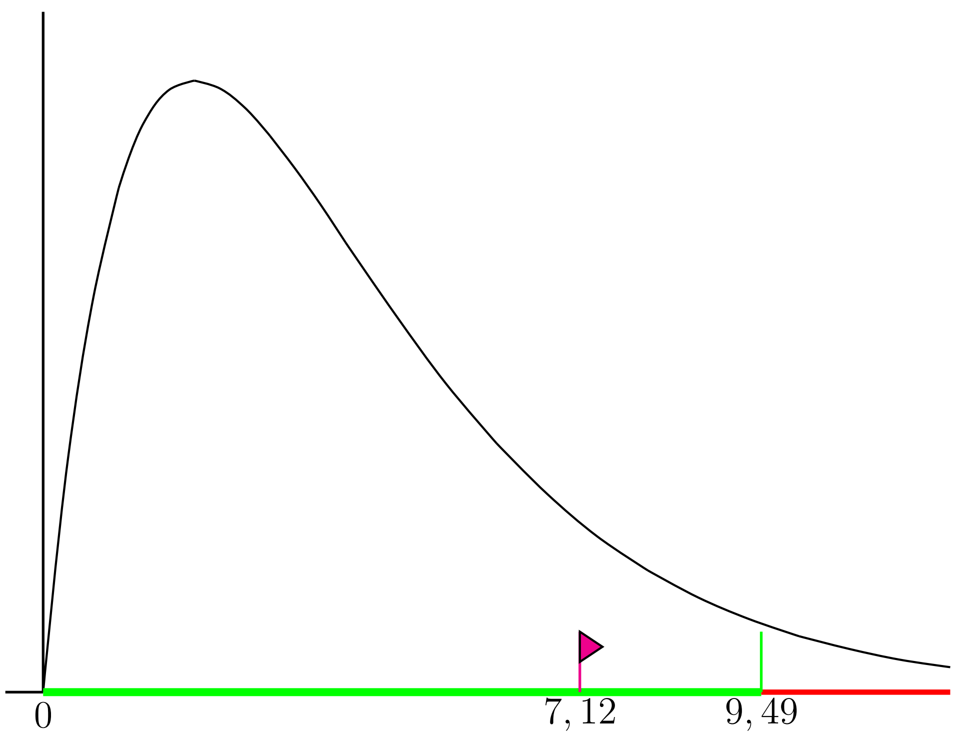


Рисунок 16. Сравнение критерия и критического значения

Процедуру проверки статистической гипотезы иллюстрирует рисунок.

Остается пояснить, почему была выбрана односторонняя критическая область. Близкое к нулю значение критерия означает, что теоретические частоты очень мало отличаются от экспериментальных. По-видимому, это не должно быть причиной отказа от гипотезы.

Задача 512. Ранее в задаче 470 была построена гистограмма, а затем было отмечено, что эти данные, похоже, подчиняются показательному распределению. Проверить эту гипотезу с помощью критерия χ^2.

Решение

Для уже построенной ранее группированной выборки вычислим теоретические частоты, предполагая, что данные подчиняются показательному распределению со средним $\bar{x} = 2,301$ (это среднее также было вычислено ранее). В нашем случае функция распределения равна

$$F(x) = 1 - e^{-\lambda x}, \text{ где } \lambda = \frac{1}{\bar{x}}.$$

Вероятность интервала, например, от 6 до 8, равна $F(8) - F(6)$, а теоретическая частота равна произведению этой вероятности на объем выборки, то есть на 30. Как и ранее, продолжим крайний интервал на всю ось, то есть вместо интервала 8 — 10 будем рассматривать интервал 8 — $+\infty$. Тогда вероятность этого интервала будет равна не $F(10) - F(8)$, а $F(+\infty) - F(8)$, то есть $1 - F(8)$. Разница, впрочем, несущественна.

Заполняем таблицу, добавив в нее строку теоретических частот и строку величин $(n_i - np_i)^2/np_i$.

Инт.	0 — 2	2 — 4	4 — 6	6 — 8	8 — 10
ЭЧ	19	6	2	2	1
ТЧ	17,42	7,30	3,066	1,28	0,93
χ^2	0.14	0,23	0,37	0,40	0,01

Суммируя элементы последней строки, находим значение критерия

$$\chi^2 = 1,15.$$

В отличие от нормального распределения, у показательного по выборке определяется только $r = 1$ параметр, поэтому для $m = 5$ интервалов число степеней свободы будет равно

$$\nu = m - 1 - r = 3.$$

Из таблиц находим значение 95% квантиля распределения хи-квадрат:

$$\chi^2_{0,95}(3) = 7,815.$$

Наблюдаемое значение меньше критического, поэтому нет причин отвергать выдвинутую ранее гипотезу.

Задача 513. На основании гистограмм, построенных в задачах 471 — 474, выдвинуть предположения о том, из какого распределения взята данная выборка. Для этого распределения получить оценку параметров и проверить с помощью критерия хи-квадрат правильность выдвинутого предположения.

В задачах 514 — 519 приведены данные группированной выборки, то есть экспериментальные частоты для заданных интервалов. Требуется:
а) построить гистограмму;
б) найти точечные оценки математического ожидания и дисперсии;
в) построить теоретическую кривую нормального распределения с найденными значениями математического ожидания и дисперсии;
г) с помощью критерия χ^2 (хи-квадрат) на уровне значимости α проверить гипотезу о нормальности распределения.
Здесь введены следующие обозначения:
α — уровень значимости;
n_1 — число наблюдений, попавших в интервал $[0, 2]$;
n_2 — число наблюдений, попавших в интервал $[2, 4]$;
n_3 — число наблюдений, попавших в интервал $[4, 6]$;
n_4 — число наблюдений, попавших в интервал $[6, 8]$;
n_5 — число наблюдений, попавших в интервал $[8, 10]$;
n_6 — число наблюдений, попавших в интервал $[10, 12]$;
n_7 — число наблюдений, попавших в интервал $[12, 14]$;
n_8 — число наблюдений, попавших в интервал $[14, 16]$.
n_9 — число наблюдений, попавших в интервал $[16, 18]$.

Задача 514.

α	n_1	n_2	n_3	n_4	n_5	n_6	n_7	n_8	n_9
0,01	11	24	85	141	183	162	60	23	13

Задача 515.

α	n_1	n_2	n_3	n_4	n_5	n_6	n_7	n_8	n_9
0,05	14	29	79	151	193	161	64	28	10

Задача 516.

α	n_1	n_2	n_3	n_4	n_5	n_6	n_7	n_8	n_9
0,05	11	24	69	176	202	156	85	22	14

Задача 517.

α	n_1	n_2	n_3	n_4	n_5	n_6	n_7	n_8	n_9
0,02	12	22	71	161	196	170	77	28	13

Задача 518.

α	n_1	n_2	n_3	n_4	n_5	n_6	n_7	n_8	n_9
0,02	12	28	60	121	210	176	80	21	14

Задача 519.

α	n_1	n_2	n_3	n_4	n_5	n_6	n_7	n_8	n_9
0,01	12	22	74	151	205	179	82	24	10

39. Гипотезы о математическом ожидании и дисперсии

Пусть имеется n независимых случайных величин $X_1, \ldots, X_n$, все они распределены нормально с одними и теми же математическими ожиданиями a и дисперсиями σ^2.

Для определения этих параметров можно использовать выведенные ранее несмещенные статистики

$$\bar{x} = \frac{x_1 + \cdots + x_n}{n}$$

для математического ожидания a, и

$$s^2 = \frac{(x_1 - \bar{x})^2 + \cdots + (x_n - \bar{x})^2}{n-1}$$

для дисперсии σ^2.

Справедлив следующий важный результат.

Теорема (Р. Фишер). Пусть $X_1, \ldots, X_n$ — независимые случайные величины, распределенные нормально с математическими ожиданиями a и дисперсиями σ^2. Тогда для ранее определенных величин $\bar{X}$ и S^2 справедливо следующее:
1) $\bar{X}$ распределена нормально со средним a и дисперсией σ^2/n;

2) $(n-1)S^2/\sigma^2$ распределено по закону χ^2 с $n-1$ степенями свободы;
3) $\bar{X}$ и S^2 независимы.

Эта теорема будет использована для проверки гипотез о значениях среднего и дисперсии, а также для проверки гипотез о равенстве средних и дисперсий для различных выборок.

Отметим, что все рассматриваемые здесь задачи играют важную роль в теории измерений. По этой причине мы и будем считать, что все выборки есть результаты измерений, то есть извлечены из нормально распределенных генеральных совокупностей.

Также во всех этих задачах мы будем предполагать, что альтернативная гипотеза — "не равно заданному" или "не равны между собой", поэтому во всех случаях будут строиться двусторонние критические области.

Задача 520. По имеющимся данным проверить на уровне значимости $\alpha = 0,05$ гипотезу о том, что дисперсия генеральной совокупности $\sigma^2 = 2$.

№	1	2	3	4	5	6	7	8	9
Данные	3,96	5,58	5,23	6,01	5,83	4,01	4,48	6,36	4,33

Решение

Пусть n — объем выборки (в нашем случае $n = 9$). Если гипотеза верна, то величина

$$\frac{(n-1)\cdot s^2}{\sigma^2},$$

как это следует из только что доказанной теоремы, имеет распределение хи-квадрат с $(n-1)$ степенью свободы.

Находим оценки среднего и дисперсии: $\bar{x} = 5,09$, $s^2 = 0,83$. Вычисляем значение критерия:

$$\frac{(n-1)\cdot s^2}{\sigma^2} = 3,33.$$

Границы области принятия гипотезы находим по таблице распределения $\chi^2(8)$, они равны $2,18$ и $17,53$. Следовательно, нет оснований отвергнуть гипотезу.

Задача 521. По имеющимся данным проверить на уровне значимости $\alpha = 0,05$гипотезу о том, что среднее генеральной совокупности равно 3. Считать известным, что дисперсия генеральной совокупности $\sigma^2 = 3$.

№	1	2	3	4	5	6	7	8	9
Данные	1,01	5,85	5,90	6,09	5,63	2,77	1,50	5,51	5,80

Решение.

Поскольку σ известно, величина

$$\frac{\bar{x} - a}{\sigma}$$

имеет нормальное распределение со средним 0 и дисперсией $1/n$, если наша гипотеза верна (n — снова объем выборки). Легко составить выражение, имеющее стандартное нормальное распределение:

$$\frac{\sqrt{n} \cdot (\bar{x} - a)}{\sigma}.$$

Подставляя значения из выборки, находим $\bar{x} = 4,45$. Поэтому значение критерия равно

$$\frac{\sqrt{n} \cdot (\bar{x} - a)}{\sigma} = 2,51.$$

Область принятия гипотезы находим по таблице нормального распределения. Это интервал $(-1,96; 1,96)$. Поэтому гипотезу следует отвергнуть.

Задача 522. По имеющимся данным проверить на уровне значимости $\alpha = 0,05$ гипотезу о том, что среднее генеральной совокупности равно 5. Дисперсию генеральной совокупности считать неизвестной.

№	1	2	3	4	5	6	7	8	9
Данные	5,47	4,63	4,63	2,64	4,36	5,15	5,94	5,99	5,45

Решение

В этом случае σ неизвестно, поэтому разброс следует оценивать по этой же выборке. В следующих формулах $n = 9$, $a = 5$, а в самих формулах, как и ранее, оставлены буквы, чтобы их можно было применять в общем случае. Величина

$$\sqrt{n} \cdot \frac{\bar{x} - a}{\sigma}$$

имеет стандартное нормальное распределение, а величина

$$\frac{(n-1)s^2}{\sigma^2}$$

имеет распределение $\chi^2(n-1)$.

Нам нужно получить такое выражение, которое имело бы заданное распределение и при этом не содержало бы неизвестных параметров. В нашем случае неизвестен один параметр — это σ^2. Чтобы σ сократилось, надо первое выражение разделить на квадратный корень из второго, и тогда получится, как легко убедиться, что величина

$$\frac{\sqrt{n}(\bar{x}-a)}{s}$$

имеет распределение Стьюдента с $(n-1)$ степенями свободы.

В нашем случае $\bar{x} = 4,92$, $s^2 = 1,06$, значение критерия равно $-0,24$, а область принятия гипотезы — это интервал $(-2,31; 2,31)$, поэтому нет оснований отвергнуть гипотезу.

Задача 523. По имеющимся данным проверить на уровне значимости $\alpha = 0,05$ гипотезу о том, что средние двух генеральных совокупностей, из которых взяты выборки, равны. Дисперсии генеральных совокупностей считать неизвестными, но равными.

№	1	2	3	4	5	6	7	8	9
Данные	5,24	3,59	4,72	6,95	6,05	5,27	5,25	3,93	5,86
Данные	4,65	5,08	4,69	4,15	5,29	4,52	3,96	5,73	5,20

Примечание. Равенство дисперсий (хотя и неизвестных) для многих практических задач можно интерпретировать как то, что измерения проводились по одинаковой методике, например, одним и тем же инструментом.

Решение

В общем случае эту задачу можно решить и для случая, когда число наблюдений в выборках не одинаково. Пусть в первой выборке m чисел, а во второй — n. Обозначим также через $\bar{x}_1$ и $\bar{x}_2$ выборочные средние по каждой из выборок, а через s_1^2 и s_2^2 — оценки дисперсии. Как и ранее, a и σ^2 — неизвестные математическое ожидание и дисперсия обеих выборок.

Если гипотеза о равенстве средних верна, то обе случайные величины $(\bar{x}_1 - a)/\sigma$ и $(\bar{x}_2 - a)/\sigma$ имеют нормальное распределение с нулевым средним и дисперсиями $1/m$ и $1/n$ соответственно. Отсюда следует, что их разность

$(\bar{x}_1 - \bar{x}_2)/\sigma$ также нормальна со средним 0 и дисперсией

$$\frac{1}{m} + \frac{1}{n} = \frac{m+n}{mn}.$$

Поэтому величина

$$\frac{\bar{x}_1 - \bar{x}_2}{\sigma} \cdot \sqrt{\frac{mn}{m+n}}$$

имеет стандартное нормальное распределение.

Случайные величины $(m-1)s_1^2/\sigma^2$ и $(n-1)s_2^2/\sigma^2$, как и ранее, имеют распределения χ^2 с $m-1$ и $n-1$ степенями свободы соответственно. Поэтому их сумма $((m-1)s_1^2 + (n-1)s_2^2)/\sigma^2$ распределена по $\chi^2(m+n-2)$.

Как и в прошлом примере, разделим первое выражение на квадратный корень из второго. Тогда получим, что величина

$$\frac{\bar{x_1} - \bar{x_2}}{\sqrt{(m-1)s_1^2 + (n-1)s_2^2}} \cdot \sqrt{\frac{mn(m+n-2)}{m+n}}$$

имеет распределение Стьюдента с $(m+n-2)$ степенями свободы.

В нашем случае $\bar{x}_1 = 5,21$, $\bar{x}_2 = 4,81$, $s_1^2 = 1,08$, $s_2^2 = 0,32$, $m = n = 9$. Следовательно, значение критерия для нашей выборки равно $1,01$. Область принятия гипотезы — интервал $(-2,12; 2,12)$, поэтому гипотеза не отвергается.

Задача 524. По имеющимся данным проверить на уровне значимости $\alpha = 0,05$ гипотезу о том, что дисперсии двух генеральных совокупностей, из которых взяты выборки, равны.

№	1	2	3	4	5	6	7	8	9
Данные	1,81	1,65	3,26	1,89	1,84	0,80	2,23	4,66	4,53
Данные	6,25	4,10	6,71	3,77	4,35	7,08	7,14	3,93	4,36

Решение

И эту задачу можно решить для случая, когда число наблюдений в выборках различно. Пусть в первой выборке m чисел, а во второй — n. Обозначим оценки дисперсий в выборках как s_1^2 и s_2^2.

Предположим, что обе генеральные совокупности имеют одинаковые дисперсии σ^2. Тогда величины

$$\frac{(m-1)s_1^2}{\sigma^2} \text{ и } \frac{(n-1)s_2^2}{\sigma^2}$$

распределены по $\chi^2(m-1)$ и $\chi^2(n-1)$ соответственно. Отсюда следует, что величина

$$\frac{s_1^2}{s_2^2}$$

подчиняется распределению Фишера с $(m-1, n-1)$ степенями свободы.

В нашем случае $m = n = 9$. Подставляя числа из наших выборок, получим, что $s_1^2 = 1,79$, $s_2^2 = 2,11$. Значение критерия равно $0,85$, а область принятия гипотезы — интервал $(0,23; 4,43)$, поэтому гипотеза не отвергается.

Задача 525. По имеющимся данным проверить при уровне значимости $\alpha = 0,05$ гипотезу о том, что дисперсия генеральной совокупности равна $\sigma^2 = 2,25$.

№	1	2	3	4	5	6	7	8	9
Данные	4,52	4,82	3,33	5,22	2,83	3,52	3,57	5,07	4,34

Задача 526. По имеющимся данным проверить при уровне значимости $\alpha = 0,02$ гипотезу о том, что дисперсия генеральной совокупности равна $\sigma^2 = 1,96$.

№	1	2	3	4	5	6	7	8	9
Данные	5,11	5,32	4,69	5,49	5,00	4,61	4,96	4,70	4,65

Задача 527. По имеющимся данным проверить при уровне значимости $\alpha = 0,05$ гипотезу о том, что среднее генеральной совокупности равна $\mu = 3$. Считать известным, что дисперсия генеральной совокупности равна $\sigma^2 = 3,24$.

№	1	2	3	4	5	6	7	8	9
Данные	4,56	1,14	6,93	5,65	3,57	1,40	2,50	1,52	5,09

Задача 528. По имеющимся данным проверить при уровне значимости $\alpha = 0,02$ гипотезу о том, что среднее генеральной совокупности равна $\mu = 2,8$. Считать известным, что дисперсия генеральной совокупности равна $\sigma^2 = 3$.

№	1	2	3	4	5	6	7	8	9
Данные	5,98	1,68	0,29	5,78	0,33	3,38	1,36	2,09	1,60

Задача 529. По имеющимся данным проверить при уровне значимости $\alpha = 0,05$ гипотезу о том, что среднее генеральной совокупности равна $\mu = 4,5$. Дисперсию генеральной совокупности считать неизвестной.

№	1	2	3	4	5	6	7	8	9
Данные	7,49	5,20	6,17	5,32	2,80	6,69	4,62	4,67	6,12

Задача 530. По имеющимся данным проверить при уровне значимости $\alpha = 0,02$ гипотезу о том, что среднее генеральной совокупности равна $\mu = 5,4$. Дисперсию генеральной совокупности считать неизвестной.

№	1	2	3	4	5	6	7	8	9
Данные	4,25	6,93	6,43	5,82	6,71	4,65	5,06	5,42	6,73

Задача 531. По имеющимся данным проверить при уровне значимости $\alpha = 0,05$ гипотезу о том, что средние двух генеральных совокупностей, из которых взяты выборки, равны. Дисперсии генеральных совокупностей считать неизвестными, но равными.

№	1	2	3	4	5	6	7	8	9
Данные	6,99	6,68	3,74	1,70	6,46	3,91	1,33	3,21	6,57
Данные	5,52	4,81	3,54	6,81	5,96	2,22	6,93	6,19	1,72

Задача 532. По имеющимся данным проверить при уровне значимости $\alpha = 0,02$ гипотезу о том, что средние двух генеральных совокупностей, из которых взяты выборки, равны. Дисперсии генеральных совокупностей считать неизвестными, но равными.

№	1	2	3	4	5	6	7	8	9
Данные	3,23	3,51	3,94	3,28	2,88	3,34	2,20	2,24	2,58
Данные	3,72	2,87	3,20	2,96	3,04	2,84	4,33	3,72	4,04

Задача 533. По имеющимся данным проверить при уровне значимости $\alpha = 0,05$ гипотезу о том, что дисперсии двух генеральных совокупностей, из которых взяты выборки, равны.

№	1	2	3	4	5	6	7	8	9
Данные	3,64	1,75	2,68	2,12	2,33	4,26	1,14	3,37	3,29
Данные	7,25	5,36	6,39	4,65	6,79	4,05	4,90	4,04	4,82

Задача 534. По имеющимся данным проверить при уровне значимости $\alpha = 0,02$ гипотезу о том, что дисперсии двух генеральных совокупностей, из

которых взяты выборки, равны.

№	1	2	3	4	5	6	7	8	9
Данные	1,65	3,73	5,51	2,37	5,87	3,75	4,95	0,72	1,04
Данные	7,85	7,87	8,44	3,67	6,20	7,32	4,20	6,73	4,66

Задача 535. По выборке объемом $n = 14$ получено, что исправленная выборочная дисперсия равна $s^2 = 1.44$. Проверить на уровне значимости $\alpha = 0.05$ гипотезу о том, что дисперсия генеральной совокупности равна 1.

Задача 536. По выборке объемом $n = 9$ получено, что выборочное среднее равно $\bar{x} = 4.45$. Проверить на уровне значимости $\alpha = 0.05$ гипотезу о том, что среднее генеральной совокупности равно 3, если известно, что дисперсия генеральной совокупности равна $\sigma^2 = 3$.

Задача 537. По выборке объемом $n = 9$ получено, что выборочное среднее равно $\bar{x} = 5.11$, а исправленная выборочная дисперсия равна $s^2 = 2.11$. Проверить на уровне значимости $\alpha = 0.05$ гипотезу о том, что среднее генеральной совокупности равно 6. Считать, что дисперсия генеральной совокупности неизвестна.

Задача 538. По двум выборкам объемом $n = 9$ получено, что их выборочные средние равны $\bar{x_1} = 11.21$ и $\bar{x_2} = 10.24$, а исправленные выборочные дисперсии равны $s_1^2 = 2.06$ и $s_2^2 = 1.94$. Проверить на уровне значимости $\alpha = 0.05$ гипотезу о том, что средние этих генеральных совокупностей равны. Считать, что дисперсии обеих совокупностей неизвестны, но равны.

Задача 539. По двум выборкам объемом $n = 12$ получено, что их исправленные выборочные дисперсии равны $s_1^2 = 4.32$ и $s_2^2 = 5.78$. Проверить на уровне значимости $\alpha = 0.05$ гипотезу о том, что дисперсии этих генеральных совокупностей равны.

40. Корреляция и регрессия

Пусть каждый элемент двумерной выборки представляет собой пару чисел (x_i, y_i), где $i = 1, \ldots, n$. Коэффициент выборочной корреляции определяется по формуле

$$\mathrm{r} = \frac{\sum\limits_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{(n-1)s_x s_y}.$$

В этой формуле использованы стандартные обозначения:
$\bar{x}$ и $\bar{y}$ — выборочные средние;
s_x и s_y — выборочные стандартные отклонения.

Для вычисления коэффициента корреляции в Microsoft Excel используется стандартная функция КОРРЕЛ. В Libre Office то же самое делает стандартная функция CORREL.

Задача 540. Доказать следующие утверждения про коэффициент выборочной корреляции:
— это число, которое может находиться в пределах от -1 до 1;
— это безразмерная величина, он не меняется при линейных преобразованиях переменных, то есть при изменении масштабов переменных, в частности:
— при изменении единиц измерения;
— при изменении начала отсчета.

Справедлива следующая теорема. Если переменные распределены нормально, то в случае правильности нулевой гипотезы, то есть при отсутствии корреляции, переменная

$$t = \frac{\mathrm{r}\sqrt{n-2}}{\sqrt{1-\mathrm{r}^2}}$$

имеет распределение Стьюдента с $(n-2)$ степенями свободы.

Задача 541. Дана двумерная выборка $(x_i, y_i), i = 1, \ldots, 9$. Найти выборочный коэффициент корреляции между x и y. При уровне значимости $\alpha = 0,05$ проверить гипотезу о значимости коэффициента корреляции.

x_i	11	14	25	27	18	10	16	20	21
y_i	7	8	14	15	10	11	9	14	18

Решение

Расчет, проведенный с помощью встроенной функции CORREL пакета Libre Office, дает результат $\mathrm{r} = 0,743$. В нашем случае $n = 9$. Подставив эти значения в формулу, найдем значение t-критерия: $t = 2,936$.

Область принятия гипотезы — область между 2,5% и 97,5% квантилями распределения Стьюдента с 7 степенями свободы. Поскольку это распределение симметричннго относительно нуля, достаточно найти 97,5% квантиль: $t_{\text{кр}} = 2,841$. Поскольку $t > t_{\text{кр}}$, на этом уровне значимости мы обязаны отклонить гипотезу о том, что коэффициент корреляции не значим.

Задача 542. Пусть дана двумерная выборка $(x_i, y_i), i = 1, \ldots, n$. Как провести прямую $y = kx + b$, оптимально аппроксимирующую данные? Назовем разность $(y_i - kx_i - b)$ между измеренным и предсказанным значениями *невязкой*. При каких значениях k и b сумма квадратов невязок будета минимальной.

Решение

Следует записать выражение для суммы квадратов невязок, продифференцировать его по k и по b, а затем приравнять к нулю получившиеся частные производные. Получатся такие уравнения

$$\begin{cases} b \cdot n + k \cdot \sum\limits_{i=1}^{n} x_i = \sum\limits_{i=1}^{n} y_i \\ b \cdot \sum\limits_{i=1}^{n} x_i + k \cdot \sum\limits_{i=1}^{n} x_i^2 = \sum\limits_{i=1}^{n} x_i y_i. \end{cases}$$

Эти уравнения называют нормальными уравнениями регрессии, коэффициенты k и b — коэффициентами регрессии, а саму найденную прямую $y = kx+b$ — линией регрессии y на x. При этом часто y называют переменной отклика, а x — факторной переменной.

Расчеты можно проводить в электронных таблицах. В пакете Libre Office коэффициент наклона k вычисляет стандартная функция SLOPE, а свободный член b — стандартная функция INTERCEPT. В пакете Microsoft Office аналогичные функции называются НАКЛОН и ОТРЕЗОК соответственно.

Задача 543. Для данных примера 541 построить линию регрессии y на x.

Решение

Электронные таблицы дают такой ответ: $y = 1,19x + 3,983$.

Задача 544. Доказать, что систему нормальных уравнений можно упростить так, чтобы матрица коэффициентов стала диагональной.

Решение

Разделив первое уравнение на n, получим

$$\bar{y} = k\bar{x} + b,$$

а это означает, что при простой линейной регрессии прямая проходит через точку $(\bar{x}, \bar{y})$, то есть через центр облака точек. Поэтому мы можем сразу записать уравнение регрессии в виде

$$y - \bar{y} = k(x - \bar{x}).$$

Тогда после преобразований получим систему нормальных уравнений с диагональной матрицей коэффициентов

$$\begin{cases} b \cdot n = \sum_{i=1}^{n} y_i \\ k \cdot \sum_{i=1}^{n} (x_i - \bar{x})^2 = \sum_{i=1}^{n} (x_i - \bar{x}) y_i. \end{cases}$$

Задача 545. Пусть двумерная генеральная совокупность конечна. Тогда, представив всю эту совокупность как выборку, можно определить выборочную корреляцию между компонентами. Доказать, что она совпадает с корреляцией между компонентами, определенной для генеральной совокупности.

Задача 546. Доказать, что коэффициент наклона линии регрессии k и коэффициент корреляции r связаны соотношением

$$k = \mathrm{r}\frac{\sigma_y}{\sigma_x}.$$

Здесь σ_x и σ_y — стандартные отклонения переменных.

Задача 547. Написать систему нормальных уравнений для определения коэффициентов квадратичной зависимости вида

$$y = ax^2 + bx + c.$$

В задачах 548 — 551 дана двумерная выборка $(x_i, y_i), i = 1, \ldots, n$.
а). Найти выборочный коэффициент корреляции между x и y.
б). Построить линию регрессии y на x.
в). При уровне значимости α проверить гипотезу о значимости коэффициента корреляции.

Задача 548. $\alpha = 0,01$.

x_i	1,37	2,00	3,01	3,85	5,07	6,16	6,87	8,39
y_i	27,28	28,13	29,18	29,87	30,91	32,50	33,21	33,57

Задача 549. $\alpha = 0,02$.

x_i	0,78	1,85	2,57	3,97	4,63	5,85	7,05	7,71
y_i	23,60	20,82	17,53	15,10	12,40	8,64	5,83	3,49

Задача 550. $\alpha = 0,02$.

x_i	1,04	2,35	3,44	4,47	5,25	5,56	6,88	7,54
y_i	26,68	26,46	24,95	23,96	22,70	21,62	20,97	20,01

Задача 551. $\alpha = 0,05$.

x_i	0,78	1,96	2,73	3,86	4,79	5,89	6,52	7,59
y_i	28,59	27,78	27,05	26,08	24,61	24,44	23,08	21,66

Список литературы

1. Вентцель Е.С. Теория вероятностей — М.: Высшая школа, 1999. — 576 с.
2. Высоцкий И. Р., Ященко И. В. Задачи заочных интернет-олимпиад по теории вероятностей и статистике. М.: МЦНМО, 2017. — 312 с.
3. Гмурман В. Е. Руководство к решению задач по теории вероятностей и математической статистике — М.: Высшая школа, 2004. — 404 с.
4. Гмурман В. Е. Теория вероятностей и математическая статистика. — М.: Высшая школа, 2003. — 479 с.
5. Ефимов А. В., Поспелов А. С. (ред.) Сборник задач по математике. Ч. 4. — М.: Издательство физико-математической литературы, 2003. — 432 с.
6. Ивченко Г. И., Медведев Ю. И. Введение в математическую статистику. М.: Издательство ЛКИ, 2010. — 600 с.
7. Лагутин М. Б. Наглядная математическая статистика. М,: БИНОМ, Лаборатория знаний, 2007. — 472 с.
8. Млодинов Л. (Не)совершенная случайность. М.: Livebook, 2013. — 352 с.
9. Пашкевич А. В. Теория вероятностей и математическая статистика для социологов и менеджеров. М.: ИЦ Академия, 2014. — 336 с.
10. Тюрин Ю. Н., Макаров А. А., Высоцкий И. Р., Ященко И. В. Теория вероятностей и статистика. М.: МЦНМО, АО Московские учебники, 2004. — 256 с.
11. Тюрин Ю. Н., Макаров А. А., Высоцкий И. Р., Ященко И. В. Теория вероятностей и статистика. Экспериментальное учебное пособие для 10 и 11 классов общеобразовательных учреждений. М.: МЦНМО, 2014. — 248 с.
12. Чжун К. Л., АитСахлиа Ф. Элементарный курс теории вероятностей. — М.: Бином, 2014 — 454 с.

Сведения об авторе

Пулькин Игорь Сергеевич — кандидат физико-математических наук, доцент кафедры высшей математики института искусственного интеллекта РТУ МИРЭА. Научные интересы: статистический анализ, статистическое моделирование.