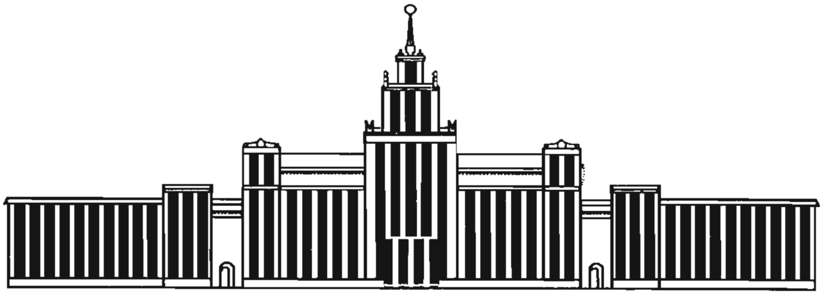
МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ



ЮЖНО-УРАЛЬСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

519.2(07)
А458

Алеев Р.Ж., Молодорич М.И.

ТЕОРИЯ ВЕРОЯТНОСТЕЙ И МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

Учебное пособие

Челябинск
2015

Министерство образования и науки Российской Федерации
Южно-Уральский государственный университет
Факультет вычислительной математики и информатики

519.2(07)
А458

Алеев Р.Ж., Молодорич М.И.

ТЕОРИЯ ВЕРОЯТНОСТЕЙ И МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

Учебное пособие

Челябинск
Издательский центр ЮУрГУ
2015

УДК 519.21(075.8) + 519.22(075.8)
А458

Одобрено
учебно-методической комиссией факультета
Вычислительной математики и информатики

Рецензенты:
кафедра математического анализа Челябинского государственного
университета (зав. кафедрой В.Е. Фёдоров),
доктор физ.-мат. наук А.С. Кондратьев

А458 **Алеев, Р.Ж.**
Теория вероятностей и математическая статистика: учебное пособие/ Р.Ж. Алеев, М.И. Молодорич. – Челябинск: Издательский центр ЮУрГУ, 2015. – 200 с.

В учебном пособии рассмотрены основные результаты теории вероятностей и математической статистики. Изложение сопровождается примерами. Также предлагаются вопросы и задачи для тестов. В дополнениях приведены результаты теории вероятностей, которые не излагаются на лекциях. Пособие рассчитано на читателей, знакомых с основами математического анализа. Оно основана на лекциях и практических занятиях по дисциплине "Теория вероятностей и математическая статистика", которые проводились авторами на факультете вычислительной математики и информатики.

УДК 519.21(075.8) + 519.22(075.8)

Оглавление

Введение

Это пособие возникло, как переработка и расширение лекций, которые читались по курсу «Теория вероятностей и математическая статистика». Автору пришлось это делать сначала эпизодически для химиков и инженеров, а потом пришлось читать данный курс регулярно в следующих вариантах.

1. Сначала для экономистов с раскладом часов в неделю: 2 часа лекций и 2 часа практики. Итог — зачёт и экзамен.
2. Потом для прикладных математиков и программистов (в одном потоке) с раскладом часов в неделю в первой половине семестра: 3 часа лекций и 2 часа практики, а во второй: 2 часа лекций и 3 часа практики. Итог — зачёт и экзамен.
3. Потом для программистов с раскладом часов в неделю: 2 часа лекций и 1 час практики. Итог — зачёт.

Понятно, что количество часов и контингент сильно влияли на отбор материала для лекций. Также чисто по причинам отсутствия других задачников, в основном, использовался задачник Гмурмана. Попытки заставить эти контингенты к самостоятельному изучению были заранее обречены на провал (хотя были энтузиасты). Поэтому нужен был компактный замкнутый курс лекций.

Отметим, особенности изложения.

1. Для случайных величин использовались латинские буквы, а не греческие. Так как эти контингенты упорно не различали ξ и ζ, σ и δ, особенно при написании на доске.
2. Часто вместо принятого $g \in \Phi_{0,1}$ для стандартного нормального распределения пишем $\Phi_{0,1}$ — стандартное нормальное распределение. Того же придерживаемся и для других распределений.
3. Большой раздел «Дополнения» читался только при варианте 2.

Следует отметить, что курс (что касается теории вероятностей) находился и находится под сильнейшим влиянием замечательной книги:

> Боровков, А.А. Теория вероятностей: учеб. пособие для вузов / А.А. Боровков — 2-е изд., перераб. и доп. — М.: Наука. Гл. ред. физ.- мат. лит., 1986. — 432 с.

Что же касается математической статистики, то здесь всё сложнее. Фактически, получилась «сборная солянка» из различных книг Колемаева, Гмурмана, Глаголева и Солнцевой и др.

1. Дискретное пространство элементарных событий

Определение. Множество называется *не более чем счётным*, если оно либо конечно, либо счётно, то есть, его можно занумеровать натуральными числами.

В случае конечного множества из n элементов имеем

$$\Omega = \{\omega_1, \ldots, \omega_n\},$$

а в случае бесконечного множества имеем

$$\Omega = \{\omega_1, \ldots, \omega_n, \ldots\},$$

здесь n может принимать любое натуральное значение.

Определения. Зафиксируем не более чем счётное непустое множество Ω, элементы которого будем называть *элементарными событиями*, или *исходами*. *Любое* подмножество из Ω будем называть *событием*. Также будем называть:

Ω — *достоверным* событием,

$\varnothing$ — *невозможным* событием

и для события A назовём

$$\overline{A} = -A = \Omega \setminus A = \Omega - A = \{\omega \in \Omega \mid \omega \notin A\}$$

дополнительным (или *противоположным*) к A событием.

под *суммой* событий понимаем их объединение как множеств, то есть

$$A + B = A \cup B = \{\omega \in \Omega \mid \omega \in A \text{ или } \omega \in B\},$$

под *произведением* событий понимаем их пересечение как множеств, то есть

$$AB = A \cdot B = A \cap B = \{\omega \in \Omega \mid \omega \in A \text{ и } \omega \in B\},$$

и под *разностью* событий понимаем их теоретико-множественную разность, то есть

$$A - B = A \setminus B = \{\omega \in \Omega \mid \omega \in A \text{ и } \omega \notin B\}.$$

Определение. События A и B будем называть *несовместными*, если их произведение невозможное событие, то есть,

$$AB = A \cap B = \varnothing.$$

Определение. Будем говорить, что на Ω задана *вероятность*, если определена функция

$$\mathsf{P} : \Omega \to [0,1] \subseteq \mathbf{R}$$

со свойством

$$\sum_{\omega \in \Omega} \mathsf{P}(\omega) = 1.$$

Определение. Вероятность события A определим как число

$$\mathsf{P}(A) = \sum_{\omega \in A} \mathsf{P}(\omega) \geqslant 0.$$

В частности $\mathsf{P}(\Omega) = 1$ (вероятность достоверного события).

Замечание. В силу неотрицательности $\mathsf{P}(\omega)$ для любого $\omega \in \Omega$ замечательная теорема Римана о рядах с неотрицательными членами обеспечивает, что $\sum_{\omega \in \Omega} \mathsf{P}(\omega)$ *не зависит* от того в каком порядке она суммируется, и гарантирует корректность корректность определения вероятности любого события.

Ещё одно полезное и важное

Замечание. Возникает вполне разумный вопрос. Чему равна вероятность невозможного события? В самом деле, мы имеем

$$\mathsf{P}(\varnothing) = \sum_{\omega \in \varnothing} \mathsf{P}(\omega).$$

Однако любой исход $\omega \in \Omega$ *не* принадлежит $\varnothing$. Поэтому в приведённой формуле справа стоит сумма пустого множества чисел, или, по-иному говоря, суммируется 0 слагаемых. В математике принято считать, что в таких случаях сумма равна 0, то есть мы имеем

$$\mathsf{P}(\varnothing) = 0.$$

Аналогично произведение пустого множества сомножителей считается равным 1, Это неудивительно, если принять во внимание замечательную пару функций $\exp(x) = e^x$ и $\ln x$, ибо $e^0 = 1$, а $\ln 1 = 0$, также $e^{\sum_i x_i} = \prod_i e^{x_i}$, а $\ln \prod_i x_i = \sum_i \ln x_i$ (здесь $x_i > 0$ для любого i).

Определение. Пара $\langle\Omega, \mathsf{P}\rangle$, удовлетворяющая вышеприведённым свойствам, называется *дискретным пространством элементарных событий.*

Теорема (о дискретной вероятности).
Пусть $\langle\Omega, \mathsf{P}\rangle$ *— дискретное пространством элементарных событий. Тогда*

1) $\mathsf{P}(\Omega) = 1$ *и* $\mathsf{P}(\varnothing) = 0$;
2) *для любых событий* A *и* B

$$\mathsf{P}(A + B) = \mathsf{P}(A) + \mathsf{P}(B) - \mathsf{P}(AB)$$

в частности, $\mathsf{P}(A + B) \leqslant \mathsf{P}(A) + \mathsf{P}(B)$,
и для несовместных событий $\mathsf{P}(A + B) = \mathsf{P}(A) + \mathsf{P}(B)$;

3) *для любого события* A *вероятность дополнительного события*

$$\mathsf{P}(\overline{A}) = 1 - \mathsf{P}(A);$$

4) *если* $A \subseteq B$, *то*

$$\mathsf{P}(A) \leqslant \mathsf{P}(B);$$

5) *если* $\{A_i\}_i$ *— не более чем счётный набор попарно несовместных событий* ($A_iA_j = \varnothing$ *для всех* $i \neq j$), *то*

$$\mathsf{P}\Big(\sum_i A_i\Big) = \sum_i \mathsf{P}(A_i),$$

если же не требовать попарной несовместности, то

$$\mathsf{P}\Big(\sum_i A_i\Big) \leqslant \sum_i \mathsf{P}(A_i).$$

Замечание. В последнем случае справа может стоять ∞.

Доказательство. 1. Было ранее.

2. В самом деле

$$\mathsf{P}(A + B) = \sum_{\omega \in A+B} \mathsf{P}(\omega) = \sum_{\omega \in A} \mathsf{P}(\omega) + \sum_{\omega \in B} \mathsf{P}(\omega) - \sum_{\omega \in AB} \mathsf{P}(\omega),$$

так как элементы AB учтены дважды: один раз в A, второй — в B. Поэтому

$$\mathsf{P}(A + B) = \mathsf{P}(A) + \mathsf{P}(B) - \mathsf{P}(AB).$$

Свойство

$$\mathsf{P}(A + B) \leqslant \mathsf{P}(A) + \mathsf{P}(B)$$

очевидно следует из неотрицательности $\mathsf{P}(AB)$. Для несовместных событий
$$\mathsf{P}(A+B) = \mathsf{P}(A) + \mathsf{P}(B),$$
потому что $\mathsf{P}(AB) = \mathsf{P}(\varnothing) = 0$ по свойству 1.

3. Имеем $A + \overline{A} = \Omega$ и $A\overline{A} = \varnothing$. Поэтому по свойствам 1 и 2
$$1 = \mathsf{P}(\Omega) = \mathsf{P}(A) + \mathsf{P}(\overline{A}).$$
Отсюда
$$\mathsf{P}(\overline{A}) = 1 - \mathsf{P}(A).$$

4. Имеем
$$\mathsf{P}(A) = \sum_{\omega\in A}\mathsf{P}(\omega) \leqslant \sum_{\omega\in B}\mathsf{P}(\omega) = \mathsf{P}(B),$$
поскольку $\mathsf{P}(\omega) \geqslant 0$ для любого $\omega \in \Omega$.

5. В силу упоминавшейся ранее теоремы Римана
$$\mathsf{P}(\sum_i A_i) = \sum_{\omega\in\sum_i A_i}\mathsf{P}(\omega) = \sum_i\sum_{\omega\in A_i}\mathsf{P}(\omega),$$
так как в силу попарной несовместности событий каждый исход $\omega \in \sum_i A_i$ принадлежит точно одному событию A_i. Следовательно
$$\mathsf{P}(\sum_i A_i) = \sum_i\sum_{\omega\in A_i}\mathsf{P}(\omega) = \sum_i\mathsf{P}(A_i).$$
Если же отказаться от попарной несовместности событий $\{A_i\}_i$, то имеем
$$\mathsf{P}(\sum_i A_i) = \sum_{\omega\in\sum_i A_i}\mathsf{P}(\omega) \leqslant \sum_i\sum_{\omega\in A_i}\mathsf{P}(\omega) = \sum_i\mathsf{P}(A_i),$$
так как некоторые исходы могут учитываться несколько раз.

□

Замечание. Утверждение 2 называют *Теоремой сложения вероятностей.*

1.1. Классическая схема

Приводимая ниже схема определения вероятности исторически была первой, в которой появилось, окрепло и развилось понятие вероятности.

Пусть множество Ω состоит из n элементов, и все исходы равновероятны. Иными словами, пусть для каждого $\omega \in \Omega$
$$\mathsf{P}(\omega) = \frac{1}{n}.$$

Такой способ определения вероятности называется *классической схемой* или *равномерным дискретным распределением*. В данном случае для любого события A имеем

$$\mathsf{P}(A) = \sum_{\omega\in A}\mathsf{P}(\omega) = \sum_{\omega\in A}\frac{1}{n} = \frac{\text{число элементов в } A}{n}.$$

Поэтому обычно говорят, что в классической схеме вероятность события равна

отношению *числа благоприятных исходов* к *числу всех исходов*.

Понятно, что под благоприятными (для данного события) исходами понимаются исходы, принадлежащие данному событию.

Классическое определение вероятности имеет важные обобщения такие, как *равномерное распределение на отрезке* или, более общо, *геометрическую вероятность*.

1.2. Выборки

Изложенное здесь — пример применения классической схемы. Одновременно с этим показывается насколько важно определить *множество всех возможных исходов*. Забегая вперёд, ещё также отметим, что понятие выборки в более обобщённом варианте является одним из основных понятий математической статистики.

Понятие выборки встречается в разных вариантах, но исходным для всех является понятие *генеральной совокупности* — множества объектов, из которых производится выборка. Сразу же отметим, что генеральная совокупность *НЕ* является множеством всех исходов, а лишь определяет его в зависимости от типа выборки.

Пусть генеральная совокупность

$$\Gamma = \{\gamma_1, \ldots, \gamma_k\}$$

состоит из k элементов. Обычно в качестве модели выступает урна с шарами, которые могут быть либо занумерованы, либо разных цветов, либо ещё что-нибудь другое.

Под *выборкой объёма* s будем понимать произвольный набор из s элементов генеральной совокупности. Отметим, что среди элементов выборки могут встречаться совпадающие. Теперь рассмотрим различные типы выборок.

1.2.1. Последовательные выборки без возвращения

Последовательность понимается как то, что указывается какой элемент выборки первый, какой — второй и т. д. Также о последовательной выборке говорят как об *упорядоченной* выборке, когда в модели урны с шарами достаётся первый шар выборки, затем второй и т.д., но шары в урну *не возвращаются.* Следует отметить, что в рассматриваемом случае мы должны различать шары, то есть они либо пронумерованы, либо покрашены все в разные цвета.

В данном случае выборка объёма s представляет собой упорядоченный набор

$$(\gamma_{j_1}, \ldots, \gamma_{j_s})$$

из s различных элементов.

Заметим сразу, что

$$s \leqslant k$$

— объём выборки не может быть больше числа элементов генеральной совокупности. Поэтому имеем

1) первый элемент $\gamma_{j_1} \in \Gamma$ произволен, и число возможностей для него равно k;
2) второй элемент $\gamma_{j_1} \in \Gamma \setminus \{\gamma_{j_1}\}$, и число возможностей для него равно $k-1$ и т.д.

Таким образом число последовательных выборок без возвращения объёма s равно

$$A_k^s = (k)_s = k(k-1)\ldots(k-(s-1)) = \frac{k!}{(k-s)!}.$$

Согласно классической схеме любой последовательной выборке без возвращения из k элементов объёма s приписывается вероятность

$$\frac{1}{A_k^s} = \frac{(k-s)!}{k!}.$$

1.2.2. Выборки с возвращением

Снова рассмотрим модель урны с пронумерованными шарами. Теперь каждый вынутый из урны шар возвращается обратно. Поэтому, независимо от того каким по счёту вынимается шар, число возможностей для него равно k, и число выборок объёма s равно (здесь может быть $s > k$)

$$k^s.$$

Согласно классической схеме любой выборке с возвращением из k элементов объёма s приписывается вероятность

$$\frac{1}{k^s}.$$

Задача. *Найти вероятность того, что выборка с возвращением объёма s из k элементов будет состоять из* различных *шаров.*

Решение. Эта задача, как легко понять, является комбинацией выборок с возвращением и без возвращения. Общее число исходов равно k^s, а число выборок из различных элементов A_k^s. Поэтому искомая вероятность равна

$$\frac{A_k^s}{k^s} = \frac{k!}{(k-s)!k^s}.$$

$\square$

1.2.3. Одновременные выборки (без возвращения)

Опять обратимся к модели урны с пронумерованными шарами. На этот раз будем вынимать шары одновременно, по-иному говоря, мы не будем интересоваться какой шар вынут первым, а какой — вторым и т. д. Другими словами ещё можно говорить о *неупорядоченной* выборке, так как нас в данном случае не интересует *порядок* элементов в выборке. Число перестановок множества из s элементов равно $s!$. Тогда число одновременных выборок (без возвращения) объёма s из k элементов равно

$$\frac{A_k^s}{s!} = \frac{k!}{(k-s)!s!} = C_k^s = \binom{k}{s}.$$

Отметим, что C_k^s — это в точности биномиальный коэффициент с номером s в формуле бинома Ньютона

$$(a+b)^k = \sum_{s=0}^{k} C_k^s a^{k-s} b^s.$$

Согласно классической схеме любой одновременной выборке (без возвращения) объёма s из k элементов приписывается вероятность

$$\frac{1}{C_k^s} = \frac{(k-s)!s!}{k!}.$$

1.3. Гипергеометрическое распределение

У нас уже встречалось слово "распределение", и теперь снова оно появляется. Пока мы будем пользоваться им в обыкновенном, житейском смысле, понимая, как набор каких-то объектов распределённых (= расположенных) по какому-то признаку. Впоследствии мы для понятия распределения придадим точный смысл.

Гипергеометрическое определение будем излагать на примере урны с шарами. Пусть в урне k шаров, среди которых k_1 чёрных и $k_2 = k - k_1$ белых. Отметим, что количество цветов может больше 2, и возникают другие распределения, но подход к ним будет очень похожим на подход в рассматриваемом случае.

Произведём одновременную выборку объёма s, и будем интересоваться вопросом какова вероятность события, что в выборке окажется точно s_1 чёрных шаров.

Число благоприятных для на исходов равно

$$C_{k_1}^{s_1} C_{k_2}^{s_2}.$$

Здесь мы учитываем, что оставшиеся $s_2 = s - s_1$ шаров должны быть *белыми.* Поэтому искомая вероятность равна

$$P_{k_1,k}(s_1, s) = \frac{C_{k_1}^{s_1} C_{k_2}^{s_2}}{C_k^s} = \frac{s!(k-s)!)}{k!} \cdot \frac{k_1!}{s_1!(k_1 - s_1)!} \cdot \frac{k_2!}{s_2!(k_2 - s_2)!}.$$

Снова отметим, что здесь $k_2 = k - k_1$ и $s_2 = s - s_1$. Также необходимо отметить, что приведённое выражение для $\mathsf{P}_{k_1,k}(s_1, s)$ может оказаться бессмысленным, если нарушаются явные ограничения (имеется в виду $s > k$, $s_1 > s$ и $s_1 > k_1$) и неявное ограничение при $s_2 > k_2$.

Пример. Пусть $k = 10$, $k_1 = 7$, $s = 6$ и $s_1 = 2$. Тогда $s_2 = 6 - 2 = 4 > 3 = k_2$, что невозможно.

В таких случаях можно либо считать, что $P_{k_1,k}(s_1, s) = 0$, либо исключать эти случаи из рассмотрения.

Определение. Набор

$$\{P_{k_1,k}(0, s), P_{k_1,k}(1, s), \ldots, P_{k_1,k}(s, s)\}$$

называется *гипергеометрическим распределением* для выборки объёма s.

Ещё раз напомним, если приведённое ранее выражение для $P_{k_1,k}(s_1, s)$ бессмысленно, то в этом распределении либо такие числа равны 0, либо исключены вовсе.

Игра Спортлото

Напомним, в чём она состоит. Нужно угадать не менее *трёх* номеров из 5 или 6 номеров, выбираемых случайно из 49, 45 или 36 чисел

$$\{1, 2, 3, \dots\}.$$

Проинтерпретируем Спортлото на модели урны с шарами. Среди данных шаров есть 5 или 6 чёрных, и нужно вытянуть менее *трёх* чёрных шаров. Теперь рассмотрим более подробно каждый из вариантов игры.

1. Игра 6 из 49

Эта игра была самой первой, но сразу стало очевидным, что почти невозможно получить максимальный выигрыш, угадав 6 номеров, то есть это говорило об очень маленькой вероятности такого выигрыша. В самом деле, общее число исходов равно

$$C_{49}^{6} = \frac{49!}{43!6!} = \frac{49 \cdot 48 \cdot 47 \cdot 46 \cdot 45 \cdot 44}{1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 \cdot 6} = 49 \cdot 47 \cdot 46 \cdot 3 \cdot 44 =$$
$$= 2^3 \cdot 3 \cdot 7^2 \cdot 11 \cdot 23 \cdot 47 = 13983816.$$

6 номеров. Вероятность угадывания 6 номеров равна

$$P_{6,49}(6,6) = \frac{C_6^6 C_{43}^0}{C_{49}^6} = \frac{1}{C_{49}^6} = \alpha = 7,151123842018516 \cdot 10^{-8}.$$

5 номеров. Вероятность угадывания 5 номеров равна

$$P_{6,49}(5,6) = \frac{C_6^5 C_{43}^1}{C_{49}^6} = \frac{6 \cdot 43}{C_{49}^6} = 258\alpha = 1,844989951240777 \cdot 10^{-5}.$$

4 номера. Вероятность угадывания 4 номеров равна

$$P_{6,49}(4,6) = \frac{C_6^4 C_{43}^2}{C_{49}^6} = \frac{15 \cdot 21 \cdot 43}{C_{49}^6} = 13545\alpha = 9,68619724401408 \cdot 10^{-4}.$$

3 номера. Вероятность угадывания 3 номеров равна

$$P_{6,49}(3,6) = \frac{C_6^3 C_{43}^3}{C_{49}^6} = \frac{20 \cdot 41 \cdot 7 \cdot 43}{C_{49}^6} = 246820\alpha = 0,0176504038668701.$$

Такие маленькие вероятности выигрышей привели к тому, что популярность этой разновидности Спортлото всё время падала, и она была заменена другой — 6 из 45.

2. Игра 6 из 45

В данном случае общее число исходов равно

$$C_{45}^6 = \frac{45!}{39!6!} = \frac{45 \cdot 44 \cdot 43 \cdot 42 \cdot 41 \cdot 40}{1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 \cdot 6} = 3 \cdot 11 \cdot 43 \cdot 7 \cdot 41 \cdot 20 = \\ = 2^2 \cdot 3 \cdot 5 \cdot 7^2 \cdot 11 \cdot 41 \cdot 43 = 8145060.$$

6 номеров. Вероятность угадывания 6 номеров равна

$$P_{6,45}(6,6) = \frac{C_6^6 C_{39}^0}{C_{45}^6} = \frac{1}{C_{45}^6} = \beta = 1,227738039989883 \cdot 10^{-7}.$$

5 номеров. Вероятность угадывания 5 номеров равна

$$P_{6,45}(5,6) = \frac{C_6^5 C_{39}^1}{C_{45}^6} = \frac{6 \cdot 39}{C_{45}^6} = 234\beta = 2,872907013576327 \cdot 10^{-5}.$$

4 номера. Вероятность угадывания 4 номеров равна

$$P_{6,45}(4,6) = \frac{C_6^4 C_{39}^2}{C_{45}^6} = \frac{15 \cdot 19 \cdot 39}{C_{45}^6} = 11115\beta = 0,0013646308314488.$$

3 номера. Вероятность угадывания 3 номеров равна

$$P_{6,45}(3,6) = \frac{C_6^3 C_{39}^3}{C_{45}^6} = \frac{20 \cdot 37 \cdot 19 \cdot 13}{C_{45}^6} = 182780\beta = 0,0224405958949351.$$

Снова, как легко заметить, вероятности больших выигрышей были очень маленькими, и поэтому наибольшей популярностью пользовался третий вид Спортлото.

3. Игра 5 из 36

Теперь общее число исходов равно

$$C_{36}^5 = \frac{36!}{31!5!} = \frac{36 \cdot 35 \cdot 34 \cdot 33 \cdot 32}{1 \cdot 2 \cdot 3 \cdot 4 \cdot 5} = 3 \cdot 7 \cdot 17 \cdot 33 \cdot 32 = \\ = 2^5 \cdot 3^2 \cdot 7 \cdot 11 \cdot 17 = 376992.$$

5 номеров. Вероятность угадывания 5 номеров равна

$$P_{5,36}(5,5) = \frac{C_5^5 C_{31}^0}{C_{36}^5} = \frac{1}{C_{36}^5} = \gamma = 2,652576181987947 \cdot 10^{-6}.$$

4 номера. Вероятность угадывания 4 номеров равна

$$P_{5,36}(4,5) = \frac{C_5^4 C_{31}^1}{C_{36}^5} = \frac{5 \cdot 31}{C_{36}^5} = 155\gamma = 4,111493082081317 \cdot 10^{-4}.$$

3 номера. Вероятность угадывания 3 номеров равна

$$P_{5,36}(3,5) = \frac{C_5^3 C_{31}^2}{C_{36}^5} = \frac{10 \cdot 15 \cdot 31}{C_{36}^5} = 4650\gamma = 0,012334479246244.$$

1.4. Схема Бернулли. Биномиальное распределение

Пусть генеральная совокупность — $\{0,1\}$, и будем интерпретировать 1 как *успех*, 0 как *неудачу*. Предположим также, что вероятность появления 1 равна $p \in [0,1]$, а вероятность появления 0 равна $q = 1-p \in [0,1]$. Такое распределение вероятностей назовём *схемой Бернулли*. Оно встречается при стрельбе (биатлон, стендовая стрельба), если вероятность попадания — p, а промаха — $q = 1-p$, также при бросании монеты, обычно $p = q = \frac{1}{2}$.

Будем производить выборку с возвращением объёма r, таких выборок будет 2^r. Если ω — такая выборка, то вероятность появления точно k единиц в выборке ω равна

$$p^k(1-p)^{r-k} = p^k q^{r-k},$$

Пусть A_k — событие, состоящее из всех выборок, в которых точно k единиц. Так как число способов размещения k единиц в наборе длины r равно C_r^k, то вероятность

$$\mathsf{P}(A_k) = C_r^k p^k (1-p)^{r-k} = P(k,r).$$

Определение. Набор

$$\{P(0,r), P(1,r), \ldots, P(r,r)\}$$

называется *биномиальным распределением* для выборки объёма r, или по-другому, для r независимых друг от друга испытаний.

2. Аксиоматический подход к теории вероятностей

Изложенное в первом параграфе понятие вероятности нельзя признать всеобъемлющим. Например, в него совершенно не укладывается изучение вероятности попадания точки в данной подмножество при бросании точки в заданный отрезок. Дело в том, что, совершенно очевидно, вероятность попадания точки в фиксированную точку отрезка равна 0, но вероятность попадания точки в подотрезок ненулевой длины из заданного отрезка будет ненулевой, а потому получать вероятность сложением вероятностей элементарных событий невозможно, и нужны новые подходы. Конечно же есть и другие примеры, что изложенный ранее подход не охватывает всего многообразия процессов, требующих для своего описания вероятностных методов.

Всё это привело к тому, что нужно было строго определить понятие вероятности так, чтобы его использование давало возможность исследовать вероятностными методами многие процессы, возникающие в науке, производстве и т. д. Наиболее известен аксиоматический подход к вероятности, который был предложен нашим великим математиком Андреем Николаевичем Колмогоровым. При этом подходе теорию вероятностей можно рассматривать как специфическую часть теории меры и интеграла по мере. К сожалению, подобные вопросы либо вообще не изучаются, либо изучаются позже, что приводит к тому, что не для всех результатов мы сможем дать строгие доказательства.

2.1. Аксиомы теории вероятностей

Зафиксируем непустое множество Ω.

Определение. Набор $\mathcal{A}$ подмножеств из Ω назовём *алгеброй событий* на Ω, если выполняются следующие условия:

1) $\Omega \in \mathcal{A}$,
2) если $A \in \mathcal{A}$, то $\Omega \setminus A = \overline{A} \in \mathcal{A}$,
3) если $\{A_n\}_n \subseteq \mathcal{A}$ — не более чем счётный набор множеств из $\mathcal{A}$ ($\forall n$ $A_n \in \mathcal{A}$), то
$$\bigcap_n A_n \in \mathcal{A} \quad \text{и} \quad \bigcup_n A_n \in \mathcal{A}$$

Элементы из $\mathcal{A}$ будем называть *событиями*.

Замечания. Мы будем придерживаться терминологии из раздела 1. Будем называть:

- Ω — *достоверным* событием,

- $\varnothing$ — *невозможным* событием
- и для события A назовём
$$\overline{A} = -A = \Omega \setminus A = \Omega - A = \{\omega \in \Omega \mid \omega \notin A\}$$
дополнительным (или *противоположным*) к A событием.

Замечания. Если A и B — события, то:

- под *суммой* событий понимаем их объединение как множеств, то есть
$$A + B = A \cup B = \{\omega \in \Omega \mid \omega \in A \text{ или } \omega \in B\},$$
- под *произведением* событий понимаем их пересечение как множеств, то есть
$$AB = A \cdot B = A \cap B = \{\omega \in \Omega \mid \omega \in A \text{ и } \omega \in B\},$$
- и под *разностью* событий понимаем их теоретико-множественную разность, то есть
$$A - B = A \setminus B = \{\omega \in \Omega \mid \omega \in A \text{ и } \omega \notin B\}.$$
- События A и B будем называть *несовместными*, если их произведение невозможное событие, то есть,
$$AB = A \cap B = \varnothing.$$
- Если $A, B \in \mathfrak{A}$, то $A - B = A\overline{B} \in \mathfrak{A}$ и $\varnothing = \Omega - \Omega \in \mathfrak{A}$.

Примеры. Заметим, что с каждым непустым множеством Ω можно связать по крайней мере две алгебры событий.

1. $\mathfrak{A} = \{\varnothing, \Omega\}$ — это самая “тощая” алгебра событий (она содержит наименьшее возможное количество событий).
2. Пусть 2^Ω — совокупность *всех* подмножеств из Ω. Тогда $\mathfrak{A} = 2^\Omega$ — это самая “жирная” алгебра событий (она содержит наибольшее возможное количество событий).

Определение. Будем называть *вероятностным пространством* тройку
$$\langle \Omega, \mathfrak{A}, \mathsf{P} \rangle,$$
где Ω — непустое множество, $\mathfrak{A}$ — алгебра событий на Ω и
$$\mathsf{P} : \mathfrak{A} \to [0,1] \subseteq \mathbf{R}$$
— функция, называемая *вероятностью* на данном вероятностном пространстве, имеющая следующие свойства:

1) $\mathsf{P}(\Omega) = 1$,
2) если $\{A_n\}_n \subseteq \mathcal{A}$ — не более чем счётный набор попарно несовместных событий из $\mathcal{A}$ ($\forall n \neq m \quad A_n A_m = \varnothing$), то

$$\mathsf{P}\Big(\sum_n A_n\Big) = \sum_n \mathsf{P}(A_n).$$

Замечания.

1. *Крайне важное!* Вероятность определяется *не* для всех подмножеств из Ω, а только для подмножеств из $\mathcal{A}$.
2. Так как вероятность неотрицательна, то теорема Римана обеспечивает корректность суммы в свойстве 2 вероятности.

Примеры.

1. Дискретное пространство элементарных событий является вероятностным пространством.
2. *Бросание точки в отрезок* $[0, 1]$. Пусть $\Omega = [0, 1]$, $\mathcal{A}$ — набор всех подмножеств из $[0, 1]$, которые имеют длину (отметим, что не все подмножества из $[0, 1]$ имеют длину). Можно показать, что $\mathcal{A}$ образует алгебру событий. Определим для $A \in \mathcal{A}$ вероятность $\mathsf{P}(A) = \ell(A)$ (здесь $\ell(A)$ — длина A). Тогда получим вероятностное пространство. Проверку всех аксиом мы опустим, поскольку она носит рутинный и абстрактный характер. Отметим, что все "хорошие" подмножества отрезка $[0, 1]$ (отрезки $\langle a, b\rangle \subseteq [0, 1]$ (с концами и без), их не более чем счётные объединения и пересечения) всегда будут иметь длину.
3. *Геометрическая вероятность.* Это обобщение предыдущего примера. Допустим, что D — множество на прямой, плоскости, в обычном трёхмерном пространстве, которое может быть "измерено", то есть имеет соответственно длину, площадь, объём. Мы будем говорить, что D измеримо, и обозначать его длину, площадь, объём через $m(D)$. Допустим, что $m(D) < \infty$. Можно показать, что подмножества из D, которые измеримы, будут образовывать алгебру событий $\mathcal{A}(D)$. Положим для $C \in \mathcal{A}(D)$ вероятность

   $$\mathsf{P}(C) = \frac{m(C)}{m(D)}.$$

   Это можно истолковывать, как попадание произвольно выбранной точки в C (конечно, с условием, что она берётся из D). Такое определение вероятности *не* зависит от расположения C, а зависит только от $m(C)$. Подобные задачи называются задачами на *геометрическую вероятность*.

Лемма (законы де Моргана).
Пусть Ω — множество, его подмножества A, B и A_i для любого i. Тогда справедливы следующие утверждения:

1) $\overline{\overline{A}} = A$ *(инволютивность),*
2) $A \subseteq B \longleftrightarrow \overline{A} \supseteq \overline{B}$,
3) $\overline{\bigcup_i A_i} = \bigcap_i \overline{A_i}$,
4) $\overline{\bigcap_i A_i} = \bigcup_i \overline{A_i}$.

Доказательство.

1. $\overline{\overline{A}} = \{\omega \in \Omega \mid \omega \notin \overline{A}\} = \{\omega \in \Omega \mid \omega \in A\} = A$.
2. Пусть $A \subseteq B$. Если $\omega \in \overline{B}$, то $\omega \notin B \supseteq A$. Поэтому $\omega \notin A$, то есть $\omega \in \overline{A}$. Откуда $\overline{A} \supseteq \overline{B}$.
Пусть $\overline{A} \supseteq \overline{B}$. Если $\omega \in A$, то $\omega \notin \overline{A} \supseteq \overline{B}$. Поэтому $\omega \notin \overline{B}$, то есть $\omega \in B$. Откуда $A \subseteq B$.
3. Имеем
$$\omega \in \overline{\bigcup_i A_i} \longleftrightarrow \forall i \quad \omega \notin A_i \longleftrightarrow \forall i \quad \omega \in \overline{A_i} \longleftrightarrow \forall i \quad \omega \in \bigcap_i \overline{A_i}.$$
4. Пользуясь инволютивностью дополнения (утверждение 1) этой леммы, получим
$$\overline{\bigcap_i A_i} = \overline{\bigcap_i \overline{\overline{A_i}}} = \overline{\overline{\bigcup_i \overline{A_i}}} = \bigcup_i \overline{A_i},$$
что завершает доказательство теоремы.

□

Теорема (о свойствах вероятности).
Пусть $\langle \Omega, \mathfrak{A}, \mathsf{P} \rangle$ — вероятностное пространство. Тогда

1) $\mathsf{P}(\Omega) = 1$ *и* $\mathsf{P}(\varnothing) = 0$;
2) *для любых событий A и B из $\mathfrak{A}$*
$$\mathsf{P}(A + B) = \mathsf{P}(A) + \mathsf{P}(B) - \mathsf{P}(AB)$$
в частности,
$$\mathsf{P}(A + B) \leqslant \mathsf{P}(A) + \mathsf{P}(B)$$
и для несовместных событий
$$\mathsf{P}(A + B) = \mathsf{P}(A) + \mathsf{P}(B);$$
3) *для любого события A вероятность дополнительного события*
$$\mathsf{P}(\overline{A}) = 1 - \mathsf{P}(A);$$

4) *если* $A \subseteq B$, *то*
$$\mathsf{P}(A) \leqslant \mathsf{P}(B);$$

5) а) *если* $\{B_n\}_{n=1}^{\infty} \subseteq \mathfrak{A}$ *— последовательность вложенных по возрастанию событий (то есть, для любого* n *имеем* $B_n \subseteq B_{n+1}$*) и* $B = \sum_n B_n$, *то*
$$\lim_{n\to\infty} \mathsf{P}(B_n) = \mathsf{P}(B);$$

б) *если* $\{C_n\}_{n=1}^{\infty} \subseteq \mathfrak{A}$ *— последовательность вложенных по убыванию событий (то есть, для любого* n *имеем* $C_n \supseteq C_{n+1}$*) и* $C = \sum_n C_n$, *то*
$$\lim_{n\to\infty} \mathsf{P}(C_n) = \mathsf{P}(C).$$

Доказательство.

1. Так как $\Omega = \Omega + \varnothing$ и $\Omega \cdot \varnothing = \varnothing$, по свойствам 1–3 из определения вероятности
$$1 = \mathsf{P}(\Omega) = \mathsf{P}(\Omega + \varnothing) = \mathsf{P}(\Omega) + \mathsf{P}(\varnothing) = 1 + \mathsf{P}(\varnothing).$$
Отсюда $\mathsf{P}(\varnothing) = 0$.

2. Так как $B = AB + (B - AB)$ и $AB \cdot (B - AB) = \varnothing$, по свойству 2 из определения вероятности
$$\mathsf{P}(B) = \mathsf{P}(AB) + \mathsf{P}(B - AB).$$
Отсюда $\mathsf{P}(B - AB) = \mathsf{P}(B) - \mathsf{P}(AB)$. Так как $A + B = A + (B - AB)$ и $A \cdot (B - AB) = \varnothing$, по свойству 2 из определения вероятности
$$\mathsf{P}(A + B) = \mathsf{P}(A) + \mathsf{P}(B - AB) = \mathsf{P}(A) + \mathsf{P}(B) - \mathsf{P}(AB).$$
Оставшаяся часть утверждения 2 и утверждения 3 и 4 доказываются точно также, как соответствующие утверждения в теореме о дискретной вероятности.

5. а) Положим $A_1 = B_1$, а для любого $k > 1$ положим
$$A_k = B_k \overline{B}_{k-1} = B_k - B_{k-1}.$$
Тогда $\{A_k\}_k$ попарно несовместны. В самом деле, пусть $i < j$. Сначала допустим, что $i = 1$. В этом случае
$$A_j = B_j - B_{j-1} \supseteq B_j - B_1 = B_j - A_1.$$
Отсюда $A_1 A_j = \varnothing$. Теперь $i > 1$, и
$$A_i = B_i - B_{i-1} \subseteq B_i, \quad \text{а} \quad A_j = B_j - B_{j-1} \subseteq \overline{B}_{j-1} \subseteq \overline{B}_i.$$

Таким образом для всех $i \neq j$ имеем $A_iA_j = \varnothing$. Далее, так как $A_n \subseteq B_n$ для любого n, то $\sum_n A_n \subseteq \sum_n B_n = B$. С другой стороны, пусть $\omega \in B = \sum_n B_n$. Выберем такое наименьшее n, что $\omega \in B_n$. Если $n = 1$, то $\omega \in A_1 = B_1$, а если $n > 1$, то $\omega \in B_n$ и $\omega \notin B_{n-1}$, то есть, $\omega \in B_n\overline{B}_{n-1} = A_n$. Таким образом, $\sum_n A_n \supseteq \sum_n B_n = B$, и потому $\sum_n A_n = \sum_n B_n = B$.

Аналогично имеем, что $\sum_{n=1}^k A_n = B_k$. В самом деле, для любого $n \leqslant k$ имеем $A_n \subseteq B_n \subseteq B_k$, и потому $\sum_{n=1}^k A_n \subseteq B_k$. С другой стороны, пусть $\omega \in B_k$. Выберем такое наименьшее n, что $\omega \in B_n$. Если $n = 1$, то $\omega \in A_1 = B_1$, а если $n > 1$, то $\omega \in B_n$ и $\omega \notin B_{n-1}$, то есть, $\omega \in B_n\overline{B}_{n-1} = A_n$. Таким образом, $\sum_{n=1}^k A_n \supseteq B_k$, и потому $\sum_n A_n = B_k$.

По свойству 2 вероятности

$$\begin{aligned}\mathsf{P}(B) &= \sum_{n=1}^{\infty} \mathsf{P}(A_n) = \lim_{k\to\infty} \sum_{n=1}^{k} \mathsf{P}(A_n) = \lim_{k\to\infty} \mathsf{P}(\sum_{n=1}^{k} A_n) = \\ &= \lim_{k\to\infty} \mathsf{P}(B_k),\end{aligned}$$

что и было нужно.

б) По законам де Моргана для любого n имеем $\overline{C}_n \subseteq \overline{C}_{n+1}$. Поэтому по предыдущему утверждению

$$\mathsf{P}(\sum_{n=1}^{\infty} \overline{C}_n) = \lim_{n\to\infty} \mathsf{P}(\overline{C}_n),$$

Снова применим закон де Моргана и получим

$$\sum_{n=1}^{\infty} \overline{C}_n = \overline{\prod_{n=1}^{\infty} C_n} = \overline{C}.$$

Теперь по утверждению 3

$$\begin{aligned}\mathsf{P}(C) &= 1 - \mathsf{P}(\overline{C}) = 1 - \mathsf{P}(\overline{\prod_{n=1}^{\infty} C_n}) = 1 - \mathsf{P}(\sum_{n=1}^{\infty} \overline{C_n}) = \\ &= 1 - \lim_{n\to\infty} \mathsf{P}(\overline{C_n}) = 1 - \lim_{n\to\infty} (1 - \mathsf{P}(C_n)) = \lim_{n\to\infty} \mathsf{P}(C_n).\end{aligned}$$

□

2.2. Условная вероятность. Независимость событий

Наличие дополнительной информации может очень сильно повлиять на вероятность события.

Примеры.

1. Вероятность вытащить бубновую даму из колоды игральных карт равна
$$\frac{1}{n},$$
где n — число карт в колоде (обычно $n \in \{32, 36, 52.54\}$). Если же нам известно, что *вытащенная карта — дама*, то вероятность становится равной
$$\frac{1}{4}.$$
2. Три раз бросаем симметричную монету, число исходов равно $2^3 = 8$. Событие
$$A = \{\text{герб выпадет точно 1 раз}\},$$
благоприятные для A исходы — (*грр*), (*ргр*) и (*ррг*). Поэтому вероятность события A
$$\frac{3}{8}.$$
Событие
$$B = \{\text{герб выпадет } \textit{нечётное} \text{ число раз}\},$$
благоприятные для B исходы — (*грр*), (*ргр*), (*ррг*) и (*ггг*). Поэтому вероятность события B
$$\frac{4}{8}.$$
Зададимся вопросом:

> Какова вероятность, что произойдёт A, если известно, что произошло B?

Общее число нужных исходов — 4, а благоприятных — 3. Поэтому вероятность события A при условии B, равна
$$\frac{3}{4} \quad \text{или} \quad \frac{3/8}{4/8}.$$

Определение. Пусть A и B — события и вероятность $\mathsf{P}(B) > 0$. *Условная) вероятность события A при условии B* по определению равна
$$\mathsf{P}_B(A) = \mathsf{P}(A|B) = \mathsf{P}(A/B) = \frac{\mathsf{P}(AB)}{\mathsf{P}(B)}.$$

Замечание. Из этого определения следует, что
$$\mathsf{P}(AB) = \mathsf{P}(B)\,\mathsf{P}(A/B).$$

Это утверждение называют *Теоремой умножения вероятностей.*

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

$$\mathsf{P}(AB) = \mathsf{P}(A)\,\mathsf{P}(B).$$

Теорема (о свойствах независимых событий).
Пусть A, B и C — события.

1. *Предположим, что $\mathsf{P}(B) > 0$. События A и B независимы тогда и только тогда, когда*
$$\mathsf{P}(A/B) = \mathsf{P}(A).$$
2. *Если A и B независимы, то $\overline{A}$ и B независимы.*
3. *Если A и B независимы, A и C независимы, а B и C* несовместны, *то A и $B+C$ независимы.*

Доказательство. 1. В самом деле

$$\mathsf{P}(A) = \mathsf{P}(A/B) = \frac{\mathsf{P}(AB)}{\mathsf{P}(B)} \longleftrightarrow \mathsf{P}(A)\,\mathsf{P}(B) = \mathsf{P}(AB).$$

2. Как в доказательстве утверждения 2 теоремы о свойствах вероятности

$$\mathsf{P}(B - AB) = \mathsf{P}(B) - \mathsf{P}(AB).$$

Далее заметим, что $\overline{A}B = B - AB$. Теперь, используя независимость A и B ($\mathsf{P}(AB) = \mathsf{P}(A)\,\mathsf{P}(B)$) и то, что $\mathsf{P}(\overline{A}) = 1 - \mathsf{P}(A)$, имеем,

$$\mathsf{P}(\overline{A}B) = \mathsf{P}(B - AB) = \mathsf{P}(B) - \mathsf{P}(AB) = \mathsf{P}(B) - \mathsf{P}(A)\,\mathsf{P}(B) = \\ = \mathsf{P}(B)(1 - \mathsf{P}(A)) = \mathsf{P}(\overline{A})\,\mathsf{P}(B).$$

3. Теперь

$$\begin{aligned}\mathsf{P}(A(B+C)) &= \mathsf{P}(AB + AC) = \mathsf{P}(AB) + \mathsf{P}(AC) = \\ &= \mathsf{P}(A)\,\mathsf{P}(B) + \mathsf{P}(A)\,\mathsf{P}(C) = \mathsf{P}(A)(\mathsf{P}(B) + \mathsf{P}(C)) = \\ &= \mathsf{P}(A)(\mathsf{P}(B+C)).\end{aligned}$$

□

Пример Бернштейна

Пусть грани правильного тетраэдра окрашены следующим образом. Одна — красная, одна — синяя, одна — зелёная, а четвёртая покрашена во все три цвета — красный, синий и зелёный. Тетраэдр бросается на плоскость. Пусть события таковы:

$$R = \text{падение на грань с красным цветом},$$

$$B = \text{падение на грань с синим цветом,}$$
$$G = \text{падение на грань с зелёным цветом.}$$

Тогда ясно, что

$$\mathsf{P}(R) = \mathsf{P}(B) = \mathsf{P}(G) = \frac{2}{4} = \frac{1}{2}.$$

Так как пара цветов встречается только на одной грани, то

$$\mathsf{P}(RB) = \mathsf{P}(RG) = \mathsf{P}(BG) = \frac{1}{4} = \frac{1}{2} \cdot \frac{1}{2}.$$

Следовательно пары событий R и B, R и G, B и G независимы.

Однако, заметим, что

$$\mathsf{P}(RBG) = \frac{1}{4} \neq \mathsf{P}(R)\,\mathsf{P}(B)\,\mathsf{P}(G) = \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{8}.$$

Поэтому независимость по парам *не переносится* на независимость по совокупности.

2.3. Формулы полной вероятности и Байеса

Теорема (формулы полной вероятности и Байеса).
Пусть A, $B_1, \ldots, B_k$ — события, причём $B_1, \ldots, B_k$ попарно несовместны и каждое из них имеет ненулевую вероятность. *Допустим, что $A \subseteq \sum_{j=1}^{k} B_j$. Тогда выполняются следующие утверждения.*

1. ***Формула полной вероятности***

$$\mathsf{P}(A) = \sum_{j=1}^{k} \mathsf{P}(B_j)\,\mathsf{P}(A/B_j).$$

2. ***Формула Байеса.***
Если $\mathsf{P}(A) > 0$, то для любого $s \in \{1, \ldots, k\}$

$$\mathsf{P}(B_s/A) = \frac{\mathsf{P}(B_s)\,\mathsf{P}(A/B_s)}{\mathsf{P}(A)} = \frac{\mathsf{P}(B_s)\,\mathsf{P}(A/B_s)}{\sum_{j=1}^{k} \mathsf{P}(B_j)\,\mathsf{P}(A/B_j)}.$$

Доказательство.

1. Ясно, что $A = A(B_1 + \cdots + B_k) = AB_1 + \cdots + AB_k$ и $AB_1, \ldots, AB_k$ попарно несовместны. Поэтому

$$\mathsf{P}(A) = \mathsf{P}\Big(\sum_{j=1}^{k} AB_j\Big) = \sum_{j=1}^{k} \mathsf{P}(AB_j) = \sum_{j=1}^{k} \mathsf{P}(B_j)\,\mathsf{P}(A/B_j).$$

2. Имеем

$$\mathsf{P}(A)\,\mathsf{P}(B_s/A) = \mathsf{P}(AB_s) = \mathsf{P}(B_s)\,\mathsf{P}(A/B_s).$$

□

Полная группа событий

Обычно для формул полной вероятности и Байеса рассматривается чуть менее общая ситуация.

Определение. *Полной группой событий* называется такой набор событий $H_1, \ldots, H_k$, что

1) $\Omega = H_1 + \cdots + H_k$,
2) $H_1, \ldots, H_k$ попарно несовместны
3) и $\mathsf{P}(H_i) > 0$ для любого $i \in \{1, \ldots, k\}$.

Понятно, что, если есть полная группа событий, то можно применять формулы полной вероятности и Байеса для любого события A.

Также употребляется следующая терминология. События $H_1, \ldots, H_k$ называются *гипотезами*, вероятности $\mathsf{P}(H_1), \ldots, \mathsf{P}(H_k)$ — *априорными* вероятностями гипотез (вероятности до испытания), вероятности $\mathsf{P}(H_1/A), \ldots, \mathsf{P}(H_k/A)$ — *апостериорными* вероятностями гипотез (вероятности после испытания).

Замечание. Часто встречается следующая ситуация. Известны вероятности события при выполнении и невыполнении некоторого другого события H. Если H и $\overline{H}$ имеют *ненулевые* вероятности, то H и $\overline{H}$ образуют полную группу событий, и можно применять формулы полной вероятности и Байеса.

Задача. *Пусть имеются приборы двух типов — отечественные и неотечественные, причём доля последних в $m(= 3) \neq 0$ больше. Для отечественного прибора вероятность ошибки измерения равна $p_1(= 0,05)$, а для неотечественного — $p_2(= 0,02)$. Какова вероятность, что прибором неизвестного типа будете измерять отечественным прибором и измерите неверно?*

Решение. Пусть событие

$$H = \text{выбран отечественный прибор.}$$

Тогда дополнительное событие

$$\overline{H} = \text{выбран неотечественный прибор.}$$

В этом случае имеем полную группу событий H и $\overline{H}$, так как

$$\mathsf{P}(H) = \frac{1}{1+m} \neq 0 \quad \text{и} \quad \mathsf{P}(\overline{H}) = \frac{m}{1+m} \neq 0.$$

Для предложенных начальных данных имеем

$$\mathsf{P}(H) = \frac{1}{1+3} = \frac{1}{4} \quad \text{и} \quad \mathsf{P}(\overline{H}) = \frac{3}{1+3} = \frac{3}{4}.$$

Пусть событие

$$A = \text{измерено неверно.}$$

Тогда

$$\mathsf{P}(A/H) = p_1 \quad \text{и} \quad \mathsf{P}(A/\overline{H}) = p_2.$$

По формуле полной вероятности

$$\mathsf{P}(A) = \mathsf{P}(H)\,\mathsf{P}(A/H) + \mathsf{P}(\overline{H})\,\mathsf{P}(A/\overline{H}) = \frac{1}{1+m} \cdot p_1 + \frac{m}{1+m} \cdot p_2 = \\ = \frac{p_1 + mp_2}{1+m}.$$

Для предложенных начальных данных имеем

$$\mathsf{P}(A) = \frac{p_1 + mp_2}{1+m} = \frac{0,05 + 3 \cdot 0,02}{1+3} = \frac{0,11}{4} = \frac{11}{400} = 0,275.$$

Нам нужно найти $\mathsf{P}(H/A)$. По формуле Байеса

$$\mathsf{P}(H/A) = \frac{\frac{p_1}{1+m}}{\frac{p_1+mp_2}{1+m}} = \frac{p_1}{p_1 + mp_2}$$

и аналогично

$$\mathsf{P}(\overline{H}/A) = \frac{\frac{mp_2}{1+m}}{\frac{p_1+mp_2}{1+m}} = \frac{mp_2}{p_1 + mp_2}.$$

Для предложенных начальных данных имеем

$$\mathsf{P}(H/A) = \frac{p_1}{p_1 + mp_2} = \frac{0,05}{0,11} = \frac{5}{11} = 0,4545454545454545\ldots,$$
$$\mathsf{P}(\overline{H}/A) = \frac{mp_2}{p_1 + mp_2} = \frac{0,06}{0,11} = \frac{6}{11} = 0,5454545454545454\ldots.$$

Отметим, что $\mathsf{P}(H/A) + \mathsf{P}(\overline{H}/A) = 1$. Это не спроста! □

3. Случайные величины

3.1. Основные понятия

Определение. Пусть $\langle \Omega, \mathfrak{A}, \mathsf{P} \rangle$ — вероятностное пространство. Функция

$$g : \Omega \to \mathbf{R}$$

называется *случайной величиной*, если для каждого $x \in \mathbf{R}$ множество

$$\{g < x\} = \{\omega \in \Omega \mid g(\omega) < x\} \text{ — событие, то есть } \{g < x\} \in \mathfrak{A}.$$

Функция

$$F_g(x) = \mathsf{P}(g < x) : \mathbf{R} \to [0, 1]$$

называется *функцией распределения* случайной величины g.

Коль скоро функция распределения действительнозначная функция одного действительного аргумента, то можно с ней работать с использованием средств математического анализа. Знание функции распределения даёт очень много информации о вероятностях попадания случайной величины в различные подмножества действительной прямой $\mathbf{R}$. Следующая лемма даёт примеры таких вероятностей, сразу обратим внимание, что особенно первое утверждение леммы о вероятности попадания в полуинтервал, с чем мы будем встречаться неоднократно.

Лемма. *Если g — случайная величина, то*

1) *для любых $x < y$*

$$\{g \in [x, y)\} \in \mathfrak{A} \text{ и } \mathsf{P}\left(g \in [x, y)\right) = F_g(y) - F_g(x),$$

2) *для любого x*

$$\{g = x\} \in \mathfrak{A} \text{ и } \mathsf{P}\left(g = x\right) = \lim_{n\to\infty} F_g\left(x + \tfrac{1}{n}\right) - F_g(x).$$

Доказательство. 1. Ясно, что $\{g \geqslant x\} = \overline{\{g < x\}}$, потому $\{g \geqslant x\} \in \mathfrak{A}$ и

$$\{g \in [x, y)\} = \{g \geqslant x\} \cdot \{g < y\} \in \mathfrak{A}.$$

Теперь

$$F_g(y) = \mathsf{P}(g < y) = \mathsf{P}\left(\{g < x\} + \{g \in [x, y)\}\right) = F_g(x) + \mathsf{P}\left(g \in [x, y)\right),$$

что даёт требуемое.

2. Заметим, что

$$\{x\} = \prod_{n=1}^{\infty}[x+\tfrac{1}{y}), \text{ а потому } \{g=x\} = \prod_{n=1}^{\infty}\{g \in [x+\tfrac{1}{y})\} \in \mathfrak{A}.$$

Получаем вложенные по убыванию события, и по теореме о свойствах вероятности

$$\mathsf{P}(g=x) = \lim_{n\to\infty} \mathsf{P}\left(g \in [x+\tfrac{1}{n})\right) = \lim_{n\to\infty} F_g\left(x+\tfrac{1}{n}\right) - F_g(x)$$

□

Замечание. Из леммы следует, что с помощью функции распределения можно найти вероятность попадания в любой отрезок $\langle a, b\rangle$ (концы могут входить, могут не входить). Более общо, можно найти вероятность попадания в любое множество, полученное из отрезков применением не более чем счётного числа пересечений и объединений, конечного числа разностей, такие множества называются *борелевскими* и образуют алгебру событий $\mathfrak{B}$, которая носит название *борелевской* алгебры (в честь Эмиля Бореля — французского математика, внёсшего большой вклад в теорию вероятностей). Набор

$$\{\mathsf{P}(g \in B \mid B \in \mathfrak{B}\}$$

называется *распределением случайной величины g.*

Теорема (о свойствах функции распределения).

1. *Пусть F — функция распределения некоторой случайной величины. Тогда F имеет следующие свойства.*

   а) **Монотонность (неубываемость).**
   Для любых $x_1 \leqslant x_2$ имеем $F(x_1) \leqslant F(x_2)$.

   б) **Поведение на бесконечностях**

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

   в) **Непрерывность слева.**
   Для любого $x_0 \in \mathbf{R}$

   $$\lim_{x\to x_0-0} F(x) = \lim_{\substack{x_0\to x_0\\ x<x_0}} F(x) = F(x_0).$$

2. *Если функция F имеет три приведённых выше свойства* а), б) *и* в), *то существует такая случайная величина, функцией распределения которой является данная функция F.*

Доказательство.

1. Пусть F — функция распределения случайной величины g.

а) Так как $\{g < x_1\} \subseteq \{g\}$, то из свойства вероятности

$$F(x_1) = \mathsf{P}(g < x_1) \leqslant \mathsf{P}(g < x_2) = F(x_2).$$

Итак, F монотонна, точнее, не убывает.

б) Так как $F(x) = \mathsf{P}(g < x) \in [0,1]$ для любого $x \in \mathbf{R}$, то в силу существования предела ограниченной монотонной функции существуют пределы

$$\lim_{x\to-\infty} F(x) \quad \text{и} \quad \lim_{x\to+\infty} F(x),$$

и надо лишь найти их конкретные значения. Пусть для любого натурального n

$$B_n = \{g < n\} \quad \text{и} \quad C_n = \{g < -n\},$$

и ясно, что

$$B_{n+1} \supseteq B_n \quad \text{и} \quad C_{n+1} \subseteq C_n.$$

Так как $\sum_{n=1}^{\infty} B_n = \Omega$, а произведение $\prod_{n=1}^{\infty} C_n = \varnothing$, то по теореме о свойствах вероятности

$$\lim_{n\to\infty} \mathsf{P}(B_n) = \mathsf{P}(\sum_{n=1}^{\infty} B_n) = \mathsf{P}(\Omega) = 1 \quad \text{и}$$

$$\lim_{n\to\infty} \mathsf{P}(C_n) = \mathsf{P}(\prod_{n=1}^{\infty} C_n) = \mathsf{P}(\varnothing) = 0,$$

что было нужно.

в) Снова, благодаря существованию предела ограниченной монотонной функции, существует предел

$$\lim_{x\to x_0-0} F(x),$$

и надо лишь найти его конкретное значение.

Пусть для любого натурального n

$$D_n = \{g < x_0 - \tfrac{1}{n}\} \text{ и ясно, что } D_{n+1} \supseteq D_n.$$

Заметим, что $\sum_{n=1}^{\infty} D_n = D = \{g < x_0\}$. В самом деле, для всякого n понятно, что

$$D_n = \{g < x_0 - \tfrac{1}{n}\} \subseteq \{g < x_0\}, \text{ откуда } \sum_{n=1}^{\infty} D_n = D \subseteq \{g < x_0\}.$$

С другой стороны, если $\omega \in \{g < x_0\}$, то $g(\omega) < x_0$, и найдётся такое натуральное n, что

$$\frac{1}{n} < x_0 - g(\omega) \quad \text{или} \quad g(\omega) < x_0 - \frac{1}{n}, \text{ то есть,}$$

$$\omega \in D_n \text{ и потому } \{g < x_0\} \subseteq \sum_{n=1}^{\infty} D_n = D.$$

Итак, имеем требуемое $\sum_{n=1}^{\infty} D_n = D = \{g < x_0\}$.
По теореме о свойствах вероятности

$$\lim_{n\to\infty} F(x_0 - \tfrac{1}{n}) = \lim_{n\to\infty} \mathsf{P}(D_n) = \mathsf{P}(\sum_{n=1}^{\infty} D_n) = \mathsf{P}(D) =$$
$$= \mathsf{P}(g < x_0) = F(x_0),$$

что было нужно.

2. Это утверждение мы не будем доказывать, только отметим, что в качестве Ω выступает $\mathbf{R}$, алгебры событий — борелевская алгебра $\mathfrak{B}$.

□

3.2. Основные типы случайных величин

3.2.1. Дискретные случайные величины

Определение. *Дискретной случайной величиной* называется случайная величина, принимающая не более чем счётное число значений.

Пусть g — дискретная случайная величина. Заметим, что для её задания достаточно указать её значения и их вероятности. В самом деле, пусть $\{x_n\}_n$ — все значения случайной величины g, и пусть $p_n = \mathsf{P}(g = x_n)$ для всех n. Тогда

$$1 = \mathsf{P}(\Omega) = \sum_n \mathsf{P}(g = x_n) = \sum_n p_n.$$

По теореме Римана о неотрицательных рядах получаем, что корректно задана функция распределения

$$F_g(x) = \sum_{x_n < x} \mathsf{P}(g = x_n) = \sum_{x_n < x} p_n,$$

и в силу теоремы о свойствах функции распределения это полностью определяет (в определённом смысле) случайную величину g.

Замечание. **Очень важное!** В силу леммы в начале этого параграфа и теоремы о свойствах функции распределения, используя существование предела у ограниченной монотонной функции, мы имеем для любой случайной случайной величины g и любого $x \in \mathbf{R}$

$$\mathsf{P}(g = x) = \lim_{n \to x+0} F_g(x) - F_g(x).$$

В случае дискретной случайной величины это даёт, что для любого n вероятность

$$p_n = \mathsf{P}(g = x_n) = \lim_{n \to x_n+0} F_g(x) - F_g(x_n)$$

— "скачок" функции распределения в точке x_n.

3.2.2. Непрерывные случайные величины

Определение. *Непрерывной случайной величиной* называется случайная величина g, у которой функция распределения равна

$$F_g(x) = \int\limits_{-\infty}^{x} dF_g(t).$$

Если F_g — дифференцируемая функция, то $dF_g(t) = F_g'(t)dt$ и, положив $F_g'(t) = \mathrm{pd}_g(t)$, получим

$$F_g(x) = \int\limits_{-\infty}^{x} \mathrm{pd}_g(t)\, dt.$$

Определение. Полученная функция pd_g называется *плотностью* случайной величины g.

Замечания.

1. В силу приведённых выше формул называют функцию распределения непрерывной случайной величины *интегральной* функцией распределения, плотность *дифференциальной* функцией распределения.
2. Формула $F_g(x) = \int_{-\infty}^{x} dF_g(t)$ является более общей по сравнению с формулой $F_g(x) = \int_{-\infty}^{x} \mathrm{pd}_g(t)\, dt$, поскольку она может выражать свойство, что интеграл берётся в смысле Лебега–Стильтьеса. Кроме того, отметим, что *необязательно,* чтобы производная $F_g'(t)$ была определена для любого $t \in \mathbf{R}$. Как мы увидим, достаточно, чтобы функция распределения была *кусочно дифференцируемой,* что приведёт к тому, что плотность необязательно непрерывна.

3. Теорема Лебега утверждает, что любая функция распределения есть сумма дискретной, непрерывной и сингулярной. Не вдаваясь в подробности, можем заключить, что рассмотрение дискретных и непрерывных функций распределения охватывает *очень* широкий класс функций распределения.

3.3. Числовые характеристики случайных величин

При изучении случайных величин важно уметь указывать какие-то их числовые характеристики, чтобы уметь как-то сравнивать различные случайные величины. Такой подход типичен для математики. В качестве, примеров укажем следующие:

1) геометрические фигуры сравниваются по площади, тела — по объёму;
2) квадратные матрицы характеризуются своими определителями.

Надо отметить, что часто случайные величины зависят от параметров, которые могут часто определять и определяться числовыми характеристиками. В заключение отметим, что иногда числовые характеристики называют *интегральными*, что связано не только с тем, что для вычисления этих характеристик используются интегралы, но скорее с одним значений слова *интегральный* (общий, объемлющий).

3.3.1. Математическое ожидание

Математическое ожидание — самая важная числовая характеристика случайной величины. Сразу же отметим, что математическое ожидание может не существовать дл некоторых случайных величин. Мы определим математическое ожидание в двух важнейших случаях — дискретном и непрерывном, а в общем случае только укажем как его находить, не вдаваясь в подробности, которые требуют достаточно глубоких знаний по теории меры и интеграла.

Определение. Пусть g — дискретная случайная величина со множеством значений $\{x_k\}_k$. Тогда её *математическим ожиданием* называется число

$$\mathsf{M}\, g = \mathsf{E}\, g = \sum_k x_k \,\mathsf{P}(g = x_k)$$

при условии, что в случае бесконечного числа значений данный ряд сходится.

Определение. Если g — непрерывная случайная величина, то её *математическим ожиданием* называется число (если данный интеграл сходится)

$$\mathsf{M}\, g = \mathsf{E}\, g = \int_{-\infty}^{\infty} x\, dF_g(x) \quad (\text{если есть плотность } = \int_{-\infty}^{\infty} x\, \mathrm{pd}_g(x)\, dx).$$

Замечание. Мы увидим из примеров, что случайные величины и дискретные (бесконечные), и непрерывные могут не иметь математических ожиданий.

Математическое ожидание произвольной случайно величины

Для определения математического ожидания произвольной случайной величины надо ввести понятие интеграла по вероятности, что является частным случаем более общего понятия интеграла по мере (интеграла Лебега). Изложим вкратце это важнейшее понятие, говоря только о математических ожиданиях.

Пусть $\langle \omega, \mathcal{A}, \mathsf{P} \rangle$ — вероятностное пространство. Случайная величина $g : \Omega \to \mathbf{R}$ называется *простой*, если множество её значений конечно, и $x_1, \ldots, x_k$ — все её различные значения. Тогда, как было замечено ранее, для любого $i \in \{1, \ldots, k\}$ множество

$$A_i = \{g = x_i\} = \{\omega \in \Omega \mid g(\omega) = x_i\} \text{ — событие, то есть } A_i \in \mathcal{A}.$$

Математическое ожидание простой случайной величины равно

$$\mathsf{M}\, g = \sum_{i=1}^{k} x_i \, \mathsf{P}(A_i).$$

Отметим, что простая случайная величина — синоним конечной дискретной случайной величины.

Далее без доказательства приведём факты, необходимые для определения математического ожидания произвольной случайной величины.

1. Пусть $g : \Omega \to \mathbf{R}$ – неотрицательная случайная величина ($g(\omega) \geqslant 0$) для всех $\omega \in \Omega$). Тогда существует такая *возрастающая* последовательность $\{g_n\}_{n=1}^{\infty}$ простых случайных величин, что
   а) для всех $\omega \in \Omega$ и натуральных n
   $$g_n(\omega) \leqslant g(\omega)$$
   б) и $\{g_n\}_{n=1}^{\infty}$ сходится к g поточечно на Ω, то есть, для всех $\omega \in \Omega$
   $$\lim_{n \to \infty} g_n(\omega) = g(\omega).$$
2. Пусть $g : \Omega \to \mathbf{R}$ — неотрицательная случайная величина, и последовательность $\{g_n\}_{n=1}^{\infty}$ удовлетворяет всем условиям, изложенным в утверждении 1, тогда по определению, полагаем, что математическое ожидание (если этот предел конечен)
   $$\mathsf{M}\, g = \lim_{n \to \infty} \mathsf{M}\, g_n.$$

Такое определение математического ожидания корректно, то есть, если $\{h_n\}_{n=1}^{\infty}$ — другая последовательность со свойствами из 1 (возрастание, ограниченность случайной величиной g, поточечная сходимость), тогда

$$\mathsf{M}\, g = \lim_{n\to\infty} \mathsf{M}\, g_n = \lim_{n\to\infty} \mathsf{M}\, h_n.$$

3. Для введения математического ожидания произвольной случайной величины нам потребуется очень полезное вспомогательное понятие.
Пусть $g : \Omega \to \mathbf{R}$ — произвольная случайная величина. Тогда определим для каждого $\omega \in \Omega$

$$g_+(\omega) = \begin{cases} g(\omega), & \text{если } g(\omega) \leqslant 0, \\ 0, & \text{если } g(\omega) < 0, \end{cases} \quad \text{и}$$

$$g_-(\omega) = \begin{cases} 0, & \text{если } g(\omega) \leqslant 0, \\ -g(\omega), & \text{если } g(\omega) < 0. \end{cases}$$

Нетрудно понять, я что g_+ и g_- — неотрицательные случайные величны, причём

$$g = g_+ - g_- \quad \text{и} \quad |g| = g_+ + g_-.$$

По определению полагаем, что математическое ожидание

$$\mathsf{M}\, g = \mathsf{M}\, g_+ - \mathsf{M}\, g_-,$$

если существуют $\mathsf{M}\, g_+$ и $\mathsf{M}\, g_-$.
Заметим сразу, что

$$\mathsf{M}\, |g| = \mathsf{M}\, g_+ + \mathsf{M}\, g_-,$$

если существуют $\mathsf{M}\, g_+$ и $\mathsf{M}\, g_-$, то есть, g и $|g|$ *одновременно* либо имеют, либо не имеют математическое ожидание.

Определение. Случайные величины g и h, заданные на одном вероятностном пространстве, называются *независимыми*, если для любых действительных чисел x и y имеем

$$\mathsf{P}(g < x, h < y) = \mathsf{P}(g < x)\, \mathsf{P}(h < y) = F_g(x) F_h(y) \quad \text{и}$$
$$\mathsf{P}(g = x, h = y) = \mathsf{P}(g = x)\, \mathsf{P}(h = y).$$

Замечание. Достаточно, чтобы выполнялось только первое свойство (с неравенствами), однако второе свойство особенно удобно для дискретных случайных величин.

Теорема (о свойствах математического ожидания).
Пусть случайные величины g и h заданы на одном вероятностном пространстве и имеют математическое ожидания. Тогда выполняются следующие утверждения.

1. *Для любых действительных чисел a и b*

$$\mathsf{M}(a + bg) = a + b\,\mathsf{M}\,g.$$

2. $\mathsf{M}(g + h) = \mathsf{M}\,g + \mathsf{M}\,h$.
3. *Для любых действительных чисел a и b, если $g \geqslant a$, или $g \leqslant b$, то*

$$\mathsf{M}\,g \geqslant a, \quad \text{или соответственно} \quad \mathsf{M}\,g \leqslant b.$$

В частности, $\mathsf{M}\,g \leqslant \mathsf{M}\,|g|$.
4. *Если $g \geqslant 0$ и $\mathsf{M}\,g = 0$, то $\mathsf{P}(g = 0) = 1$.*
5. *Если g и h независимы, то*

$$\mathsf{M}(gh) = \mathsf{M}\,g\,\mathsf{M}\,h.$$

Замечание. Нам достаточно часто будет встречаться требование выполнения условия $\mathsf{M}(gh) = \mathsf{M}\,g\,\mathsf{M}\,h$. Поэтому, допуская некоторую вольность речи, мы будем говорить об этом свойстве, как о *свойстве произведения математических ожиданий.*

К сожалению, мы не в состоянии дать полное доказательство этой теоремы, поскольку, как уже было отмечено, у нас нет техники для работы с интегралом Лебега. Поэтому будем доказывать теорему для дискретных и непрерывных случайных величин (причём иногда не совсем строго), а в общем случае дадим иногда только эскиз доказательства.

Доказательство теоремы в дискретном случае. В этом случае мы сможем дать полное доказательство теоремы.

1. Пусть дискретная случайная величина g имеет в качестве множества значений $\{x_i\}_i$. Если $b = 0$, то

$$\mathsf{M}(a + bg) = \mathsf{M}\,a = \sum_i a\,\mathsf{P}(g = x_i) = a \sum_i \mathsf{P}(g = x_i) = a,$$

так как $\sum_i \mathsf{P}(g = x_i) = 1$.
Если же $b \neq 0$, то для любого i

$$\mathsf{P}(a + bg = a + bx_i) = \mathsf{P}(g = x_i).$$

2. Теперь

$$\begin{aligned}\mathsf{M}(a+bg) &= \sum_i (a+bx_i)\,\mathsf{P}(a+bg=a+bx_i) = \\ &= \sum_i (a+bx_i)\,\mathsf{P}(g=x_i) = \\ &= \sum_i a\,\mathsf{P}(g=x_i) + \sum_i bx_i\,\mathsf{P}(g=x_i) = \\ &= a\sum_i \mathsf{P}(g=x_i) + b\sum_i x_i\,\mathsf{P}(g=x_i) = a + b\,\mathsf{M}\,g.\end{aligned}$$

3. Пусть дискретная случайная величина g имеет в качестве множества значений $\{x_i\}_i$, а дискретная случайная величина h имеет в качестве множества значений $\{y_j\}_j$. Тогда множество всех значений случайной величины $g+h$ равно $\{x_i+y_j\}$, и имеем для любых i и j по формуле полной вероятности

$$\begin{aligned}\mathsf{P}(g=x_i) &= \sum_j \mathsf{P}(h=y_j)\,\mathsf{P}(g=x_i/h=y_j) \quad \text{и} \\ \mathsf{P}(h=y_j) &= \sum_i \mathsf{P}(g=x_i)\,\mathsf{P}(h=y_i/g=y_j).\end{aligned}$$

Поэтому

$$\begin{aligned}\mathsf{M}(g+h) &= \sum_{i,j}(x_i+y_j)\,\mathsf{P}(g=x_i,h=y_j) = \\ &= \sum_{i,j} x_i\,\mathsf{P}(g=x_i,h=y_j) + \sum_{i,j} y_j\,\mathsf{P}(g=x_i,h=y_j) = \\ &= \sum_{i,j} x_i\,\mathsf{P}(g=h_j)\,\mathsf{P}(g=x_i/h=y_j) + \\ &\qquad + \sum_{i,j} y_j\,\mathsf{P}(g=x_i)\,\mathsf{P}(h=y_j/g=x_i) = \\ &= \sum_i x_i\left(\sum_j \mathsf{P}(g=h_j)\,\mathsf{P}(g=x_i/h=y_j)\right) + \\ &\qquad + \sum_j y_j\,(\mathsf{P}(g=x_i)\,\mathsf{P}(h=y_j/g=x_i)) = \\ &= \sum_i x_i\,\mathsf{P}(g=x_i) + \sum_j y_j\,\mathsf{P}(h=y_j) = \mathsf{M}\,g + \mathsf{M}\,h.\end{aligned}$$

4. Докажем только для $\leqslant$, а для $\geqslant$ всё совершенно аналогично. Сохраняя предыдущие обозначения, имеем

$$a = a \cdot 1 = a \sum_i \mathsf{P}(g = x_i) = \sum_i a\,\mathsf{P}(g = x_i) \leqslant \sum_i x_i\,\mathsf{P}(g = x_i) = \mathsf{M}\,g.$$

5. Если бы $\mathsf{P}(g \neq 0) > 0$, то для некоторого $x_i > 0$ было бы $\mathsf{P}(g = x_i) > 0$ и потому

$$0 = \mathsf{M}\,g = \sum_i x_i\,\mathsf{P}(g = x_i) = \sum_{i, x_i > 0} x_i\,\mathsf{P}(g = x_i) > 0.$$

Противоречие.

6. Сохраним прежние обозначения, и имеем

$$\mathsf{M}\,gh = \sum_{i,j} x_i y_j\,\mathsf{P}(g = x_i, h = y_j) = \sum_{i,j} x_i y_j\,\mathsf{P}(g = x_i)\,\mathsf{P}(h = y_j) =$$
$$= \left(\sum_i x_i\,\mathsf{P}(g = x_i)\right)\left(\sum_j y_j\,\mathsf{P}(h = y_j)\right) = \mathsf{M}\,g \cdot \mathsf{M}\,h.$$

□

Доказательство теоремы в непрерывном случае. В этом случае мы сможем дать доказательство теоремы, которое иногда будет не совсем строго.

1. Если $b = 0$, то случайная величина $a + bg$ постоянна и равна a, то есть, на деле, получилась дискретная случайная величина, имеющая всего одно значение, а это случай рассмотрен ранее.

   Пусть $b \neq 0$. Тогда для любого действительного x

$$g = x \iff a + bg = a + bx.$$

   Так как можно истолковывать $dF_g(x)$, как вероятность принятия случайной величиной g значения, равного x („мгновенное" приращение вероятности), то можно считать, что $dF_g(x) = dF_{a+bg}(x)$. Отсюда

$$\mathsf{M}(a + bg) = \int_{-\infty}^{\infty} (a + bx)\,dF_g(x) = \int_{-\infty}^{\infty} a\,dF_g(x) + \int_{-\infty}^{\infty} bx\,dF_g(x) =$$
$$= a \int_{-\infty}^{\infty} dF_g(x) + b \int_{-\infty}^{\infty} x\,dF_g(x) =$$
$$= a(F_g(\infty) - F_g(-\infty)) + b\,\mathsf{M}\,g = a(1 - 0) + b\,\mathsf{M}\,g =$$
$$= a + b\,\mathsf{M}\,g.$$

2. Имеем

$$\mathsf{M}(g+h) = \int\limits_{-\infty}^{\infty} x\,dF_{g+h}(x) = \int\limits_{-\infty}^{\infty} x\,(dF_g(x) + dF_h(x)) = \\ = \int\limits_{-\infty}^{\infty} x\,dF_g(x) + \int\limits_{-\infty}^{\infty} x\,dF_h(x) = \mathsf{M}\,g + \mathsf{M}\,h.$$

3. Имеем

$$a = a \cdot 1 = a \int\limits_{-\infty}^{\infty} dF_g(x) = \int\limits_{-\infty}^{\infty} a\,dF_g(x) \leqslant \int\limits_{-\infty}^{\infty} x\,dF_g(x) = \mathsf{M}\,g,$$

поскольку $a \leqslant x$ для всех "существенных" значений вероятности. Дело в том, что условие $a \leqslant g$ влечёт, что при $x < a$ дифференциал $dF_g(x) = 0$, поскольку таких значений g нет, и потому нулевой рост, который характеризуется дифференциалом.

4. В этом случае утверждение весьма сложно. Поэтому мы рассмотрим лишь один важный случай, когда есть плотность случайной величины и она кусочно непрерывна. В данном случае число точек разрыва не более чем счётно, и потому интеграл по ним будет равен 0. Поймём, что в любой точке непрерывности плотность равна 0. В самом деле, пусть, от противного, плотность $\mathrm{pd}_g \neq 0$. Тогда в силу свойства непрерывных функций (точнее, это свойство предела функции в точке) существует такая окрестность $[c,d)$ (может быть односторонняя) точки x_0, в которой функция pd_g сохраняет знак и

$$\mathsf{P}(g \in [c,d)) = F_g(d) - F_g(c) = \int\limits_c^d \mathrm{pd}_g(t)\,dt.$$

Если бы $\mathrm{pd}_g(x_0) < 0$, то имели бы $\mathsf{P}(g \in [c,d)) < 0$, что невозможно. Кроме того, если $x_0 < 0$, то получили, бы что случайная величина принимает отрицательные значения, что противоречит условию. Итак,$x_0 \leqslant 0$ и $\mathrm{pd}_g(x) > 0$, но тогда

$$0 = \mathsf{M}\,g = \int\limits_{-\infty}^{\infty} x\,\mathrm{pd}_g(x)\,dx \leqslant \int\limits_{0}^{\infty} x\,\mathrm{pd}_g(x)\,dx \leqslant \int\limits_{-\infty}^{\infty} x\,\mathrm{pd}_g(x)\,dx > 0,$$

что не так по условию.

5. Так как можно истолковывать $dF_g(x)$, как вероятность принятия случайной величиной g значения, равного x ("мгновенное" приращение вероятности) и $dF_h(x)$ и $dF_{gh}(x,y)$ имеют аналогичное понимание, то в силу независимости случайных величин g и h

$$dF_{g,h}(xy) = dF_g(x) \cdot dF_h(y).$$

Отсюда

$$\mathsf{M}\, gh = \iint\limits_{\mathbf{R}^2} xy\, dF_{gh}(x,y) = \iint\limits_{\mathbf{R}^2} xy\, dF_g(x) dF_h(y) =$$
$$= \left(\int\limits_{-\infty}^{\infty} x\, dF_g(x) \right) \left(\int\limits_{-\infty}^{\infty} y\, dF_h(y) \right) = \mathsf{M}\, g \cdot \mathsf{M}\, h.$$

□

Доказательство теоремы в общем случае.
К сожалению, мы не в состоянии дать полное доказательство этой теоремы, поскольку, как уже было отмечено, у нас нет техники для работы с интегралом Лебега. Поэтому в общем случае дадим только эскиз доказательства.

В этом случае надо сначала рассмотреть простые случайные величины, и утверждение получается как для дискретных случайных величин, а затем нужно сделать предельный переход. □

3.3.2. Дисперсия

Определение. Пусть g — случайная величина, имеющая математическое ожидание. Тогда *дисперсией* случайной величины g называется

$$\mathsf{D}\, g = \mathsf{M}(g - \mathsf{M}\, g)^2$$

при условии, что такое математическое ожидание существует.

$\sqrt{\mathsf{D}\, g}$ называется *стандартным (средним, среднеквадратическим) уклонением (отклонением)* случайной величины g.

Замечание. Утверждение 3 теоремы о свойствах математического ожидания обеспечивает, что всегда дисперсия $\mathsf{D}\, g \geqslant 0$, поскольку $(g - \mathsf{M}\, g)^2 \geqslant 0$. Поэтому, в частности, понятие стандартного уклонения имеет вполне определённый смысл.

Определение. Пусть случайные величины g и h заданы на одном вероятностном пространстве и имеют математическое ожидания. Тогда *ковариацией* случайных величины g и h называется

$$\mathsf{cov}(g, h) = \mathsf{M}\left((g - \mathsf{M}\, g)(h - \mathsf{M}\, h)\right)$$

при условии, что такое математическое ожидание существует.

Замечание. Ясно, что

$$\mathsf{D}\, g = \mathsf{cov}(g, g).$$

Теорема (о свойствах дисперсии).
Пусть случайные величины g и h заданы на одном вероятностном пространстве и имеют математическое ожидания и дисперсии. Тогда выполняются следующие утверждения.

1. $\mathsf{D}\, g = \mathsf{M}\, g^2 - (\mathsf{M}\, g)^2$ *и* $\mathsf{cov}(g, h) = \mathsf{M}(gh) - (\mathsf{M}\, g)(\mathsf{M}\, h)$.
2. ***Положительная определённость дисперсии.***
   $\mathsf{D}\, g \geqslant 0$ *и* $\mathsf{D}\, g = 0$ *тогда и только тогда, когда существует такое действительное число c, что* $\mathsf{P}(g = c) = 1$.
3. *Для любых действительных чисел a и b*

   $$\mathsf{D}(a + bg) = b^2\, \mathsf{D}\, g.$$

   В частности, $\mathsf{D}(-g) = \mathsf{D}\, g$.
4. $\mathsf{D}(g + h) = \mathsf{D}\, g + \mathsf{D}\, h + 2\,\mathsf{cov}(g, h)$.
5. *Если g и h независимы, или, более общо, имеют свойство произведения математических ожиданий, то*

   $$\mathsf{D}(g + h) = \mathsf{D}\, g + \mathsf{D}\, h \quad \text{и} \quad \mathsf{cov}(g, h) = 0.$$

   Кроме того, $\mathsf{D}(g - h) = \mathsf{D}\, g + \mathsf{D}\, h$.

Доказательство. 1. Достаточно доказать для ковариации

$$\mathsf{cov}(g, h) = \mathsf{M}\left((g - \mathsf{M}\, g)(h - \mathsf{M}\, h)\right) = \\ = \mathsf{M}(gh - g\,\mathsf{M}\, h - h\,\mathsf{M}\, g + (\mathsf{M}\, g)(\mathsf{M}\, h)).$$

Так как $\mathsf{M}\, g$ и $\mathsf{M}\, h$ — постоянные величины, то по теореме о свойствах математических ожиданий имеем

$$\mathsf{cov}(g, h) = \mathsf{M}(gh) - (\mathsf{M}\, g)(\mathsf{M}\, h) - (\mathsf{M}\, h)(\mathsf{M}\, g) + (\mathsf{M}\, g)(\mathsf{M}\, h)) = \\ = \mathsf{M}(gh) - (\mathsf{M}\, g)(\mathsf{M}\, h).$$

2. То, что $\mathsf{D}\,g \geqslant 0$, уже было отмечено. Далее

$$\mathsf{D}\,g = 0 \iff \mathsf{M}(g - \mathsf{M}\,g)^2 = 0 \iff$$

по теореме о свойствах математических ожиданий

$$\iff \mathsf{P}\left((g - \mathsf{M}\,g)^2 = 0\right) = 1 \iff \mathsf{P}\,(g - \mathsf{M}\,g = 0) = 1 \iff$$
$$\iff \mathsf{P}\,(g = \mathsf{M}\,g) = 1.$$

3. По теореме о свойствах математических ожиданий

$$\mathsf{M}(a + bg) = a + b\,\mathsf{M}\,g$$

и теперь

$$\mathsf{D}(a + bg) = \mathsf{M}\,(a + bg - \mathsf{M}(a + bg))^2 = \mathsf{M}\,(a + bg - a - b\,\mathsf{M}\,g))^2 =$$
$$= \mathsf{M}\,(bg - b\,\mathsf{M}\,g))^2 = \mathsf{M}\,b^2\,(g - \mathsf{M}\,g))^2 = b^2\,\mathsf{M}\,(g - \mathsf{M}\,g))^2 = b^2\,\mathsf{D}\,g.$$

В частности,

$$\mathsf{D}(-g) = (-1)^2\,\mathsf{D}\,g = \mathsf{D}\,g.$$

4. В самом деле

$$\mathsf{D}(g + h) = \mathsf{M}\,(g + h - \mathsf{M}(g + h))^2 = \mathsf{M}\,(g + h - \mathsf{M}\,g - \mathsf{M}\,h))^2 =$$
$$= \mathsf{M}\,\left((g - \mathsf{M}\,g)^2 + (h - \mathsf{M}\,h)^2 + 2(g - \mathsf{M}\,g)(h - \mathsf{M}\,h)\right) =$$
$$= \mathsf{M}\,(g - \mathsf{M}\,g))^2 + \mathsf{M}\,(h - \mathsf{M}\,h))^2 + 2\,\mathsf{M}\,((g - \mathsf{M}\,g)(h - \mathsf{M}\,h)) =$$
$$= \mathsf{D}\,g + \mathsf{D}\,h + 2\,\mathsf{cov}(g, h).$$

5. В данном случае $\mathsf{M}(gh) = (\mathsf{M}\,g)(\mathsf{M}\,h)$ и потому по утверждению 1

$$\mathsf{cov}(g, h) = \mathsf{M}(gh) - (\mathsf{M}\,g)(\mathsf{M}\,h) = 0.$$

Откуда по утверждению 4 $\mathsf{D}(g + h) = \mathsf{D}\,g + \mathsf{D}\,h$ и по утверждению 3

$$\mathsf{D}(g - h) = \mathsf{D}\,g + \mathsf{D}(-h) = \mathsf{D}\,g + \mathsf{D}\,h.$$

$\square$

3.3.3. Коэффициент корреляции

Определение. Случайная величина g называется *центрированной*, если она имеет математическое ожидание и $\mathsf{M}\,g = 0$ и называется *нормированной*, если она центрирована, имеет дисперсию и $\mathsf{D}\,g = 1$.

Лемма. *Пусть g — случайная величина, имеющая математическое ожидание. Тогда*

1) *случайная величина $g_C = g - \mathsf{M}\, g$ центрирована,*
2) *если случайная величина g имеет* ненулевую *дисперсию, то случайная величина*
$$g_N = \frac{g_C}{\sqrt{\mathsf{D}\, g}} = \frac{g - \mathsf{M}\, g}{\sqrt{\mathsf{D}\, g}}$$
нормирована.

Доказательство. 1. В самом деле
$$\mathsf{M}\, g_C = \mathsf{M}(g - \mathsf{M}\, g) = \mathsf{M}\, g - \mathsf{M}\, g = 0.$$

2. По теореме о свойствах дисперсии
$$\mathsf{M}\, g_N = \mathsf{M}\left(\frac{g_C}{\sqrt{\mathsf{D}\, g}}\right) = \frac{\mathsf{M}\, g_C}{\sqrt{\mathsf{D}\, g}} = \frac{0}{\sqrt{\mathsf{D}\, g}} = 0 \quad \text{и}$$
$$\mathsf{D}\, g_N = \mathsf{D}\left(\frac{g_C}{\sqrt{\mathsf{D}\, g}}\right) = \frac{1}{\left(\sqrt{\mathsf{D}\, g}\right)^2} \cdot \mathsf{D}(g - \mathsf{M}\, g) = \frac{1}{\mathsf{D}\, g} \cdot \mathsf{D}\, g = 1.$$

□

Определение. Пусть случайные величины g и h заданы на одном вероятностном пространстве и имеют математическое ожидания и ненулевые дисперсии. Тогда *коэффициентом корреляции* случайных величины g и h называется
$$\begin{aligned}\rho(g, h) = \mathsf{M}(g_N h_N) &= \mathsf{M}\left(\frac{g_C}{\sqrt{\mathsf{D}\, g}} \cdot \frac{h_C}{\sqrt{\mathsf{D}\, h}}\right) = \\ &= \mathsf{M}\left(\frac{g - \mathsf{M}\, g}{\sqrt{\mathsf{D}\, g}} \cdot \frac{h - \mathsf{M}\, h}{\sqrt{\mathsf{D}\, h}}\right) = \\ &= \frac{\mathsf{cov}(g, h)}{\sqrt{\mathsf{D}\, g \cdot \mathsf{D}\, h}}\end{aligned}$$

Теорема (о свойствах коэффициента корреляции).
Пусть случайные величины g и h заданы на одном вероятностном пространстве и имеют математическое ожидания и ненулевые дисперсии. Тогда выполняются следующие утверждения.

1. $|\rho(g, h)| \leqslant 1$.
2. *Если g и h независимы, или, более общо, имеют свойство произведения математических ожиданий, то $\rho(g, h) = 0$.*

3. $|\rho(g,h)| = 1$ *тогда и только тогда, когда существуют такие действительные числа a и b, что вероятность*

$$\mathsf{P}(h = a + bg) = 1.$$

оказательство.

1. Пусть $\delta \in \{-1, 1\}$. По теоремам о свойствах математического ожидания и дисперсии имеем

$$0 \leqslant \mathsf{D}\,(\delta g_N - h_N) = \mathsf{D}\,\delta g_N + \mathsf{D}\,h_N - 2\,\mathsf{cov}(\delta g_N, h_N) = \\ = \delta^2\,\mathsf{D}\,g_N + 1 - 2\,\mathsf{cov}(\delta g_N, h_N) = 2 - 2\,\mathsf{cov}(\delta g_N, h_N).$$

Далее, так как $\mathsf{M}\,g_N = \mathsf{M}\,h_N = 0$ и $\mathsf{D}\,g_N = \mathsf{D}\,h_N = 1$, то

$$\mathsf{cov}(\delta g_N, h_N) = \delta\,\mathsf{M}(g_N h_N) = \delta\rho(g,h).$$

Откуда

$$0 \leqslant 1 - \delta\rho(g,h) \iff \delta\rho(g,h) \leqslant 1.$$

Теперь

$$\begin{cases} \rho(g,h) \leqslant 1 & \text{при } \delta = 1, \\ \rho(g,h) \geqslant -1 & \text{при } \delta = -1. \end{cases}$$

Поэтому $|\rho(g,h)| \leqslant 1$.

2. Так как

$$\rho(g,h) = \frac{\mathsf{cov}(g,h)}{\sqrt{\mathsf{D}\,g \cdot \mathsf{D}\,h}},$$

то всё следует утверждения 5 теоремы о свойствах дисперсии.

3. **Необходимость** ($\longrightarrow$).

Пусть $\rho(g,h) = \delta \in \{-1, 1\}$. Тогда как при доказательстве утверждения 1 этой теоремы

$$0 = 2\,(1 - \delta\rho(g,h)) = \mathsf{D}\,(\delta g_N - h_N)\,.$$

В силу положительной определённости дисперсии получаем

$$\mathsf{P}\,(\delta g_N - h_N = c) = 1$$

для некоторого c, но, так как

$$c = \mathsf{M}\,(\delta g_N - h_N) = \delta\,\mathsf{M}\,g_N - \mathsf{M}\,h_N = 0 - 0 = 0,$$

то имеем

$$\mathsf{P}\,(h_N = \delta g_N) = 1 \iff \mathsf{P}\left(\frac{h - \mathsf{M}\,h}{\sqrt{\mathsf{D}\,h}} = \delta\frac{g - \mathsf{M}\,g}{\sqrt{\mathsf{D}\,g}}\right) = 1 \iff$$

$$\longleftrightarrow \ \mathsf{P}\left(h = \mathsf{M}\,h + \sqrt{\mathsf{D}\,h}\cdot\delta\frac{g - \mathsf{M}\,g}{\sqrt{\mathsf{D}\,g}}\right) = 1 \ \longleftrightarrow$$
$$\longleftrightarrow \ \mathsf{P}\,(h = a + bg) = 1,$$

где

$$a = \mathsf{M}\,h - \sqrt{\mathsf{D}\,h}\cdot\delta\frac{\mathsf{M}\,g}{\sqrt{\mathsf{D}\,g}} = \mathsf{M}\,h - \delta\frac{\sqrt{\mathsf{D}\,h}}{\sqrt{\mathsf{D}\,g}}\mathsf{M}\,g \quad \text{и}$$
$$b = \sqrt{\mathsf{D}\,h}\cdot\delta\frac{1}{\sqrt{\mathsf{D}\,g}} = \delta\frac{\sqrt{\mathsf{D}\,h}}{\sqrt{\mathsf{D}\,g}}.$$

Достаточность ($\longleftarrow$).
Положим $h = h_0 + h_1$, где $h_1 = a + bg$. Тогда $\mathsf{P}(h = h_1) = 1$ и $\mathsf{P}(h_0 = 0) = 1$. Теперь вполне понятно, что

$$\mathsf{M}\,h = \mathsf{M}\,h_1 \quad \text{и} \quad \mathsf{M}\,h_0 = 0.$$

Найдём ковариацию

$$\mathsf{cov}(g, h) = \mathsf{M}(gh) - (\mathsf{M}\,g)(\mathsf{M}\,h) = \mathsf{M}\left(g(h_0 + h_1)\right) - (\mathsf{M}\,g)(\mathsf{M}\,h_1) =$$
$$= \mathsf{M}\,(gh_0) + \mathsf{M}\,(gh_1)) - (\mathsf{M}\,g)(\mathsf{M}\,h_1) = \mathsf{cov}(g, h_1),$$

поскольку $\mathsf{M}\,(gh_0))$, ибо опять $\mathsf{P}(gh_0 = 0) = 1$. Далее

$$\begin{aligned}\mathsf{cov}(g, h) = \mathsf{cov}(g, h_1) &= \mathsf{M}(gh_1) - (\mathsf{M}\,g)(\mathsf{M}\,h_1) = \\ &= \mathsf{M}\left(g(a + bg)\right) - (\mathsf{M}\,g)\left(\mathsf{M}(a + bg)\right) = \\ &= a\,\mathsf{M}\,g + b\,\mathsf{M}\,g^2 - (\mathsf{M}\,g)\left(a + b\,\mathsf{M}\,g)\right) = \\ &= a\,\mathsf{M}\,g + b\,\mathsf{M}\,g^2 - a\,\mathsf{M}\,g - b(\mathsf{M}\,g)^2 = \\ &= b(\mathsf{M}\,g^2 - \left(\mathsf{M}\,g)^2\right) = b\,\mathsf{D}\,g.\end{aligned}$$

Снова понятно, что

$$\mathsf{D}\,h = \mathsf{D}\,h_1 = \mathsf{D}(a + bg) = b^2\,\mathsf{D}\,g \quad \text{и}$$
$$\sqrt{\mathsf{D}\,h} = \sqrt{\mathsf{D}\,h_1} = \sqrt{b^2\,\mathsf{D}\,g} = |b|\sqrt{\mathsf{D}\,g}.$$

Отсюда

$$\begin{aligned}\rho(g, h) &= \frac{\mathsf{cov}(g, h)}{\sqrt{\mathsf{D}\,g\cdot\mathsf{D}\,h}} = \frac{\mathsf{cov}(g, h_1)}{\sqrt{\mathsf{D}\,g}\cdot\sqrt{\mathsf{D}\,h_1}} = \\ &= \frac{b\,\mathsf{D}\,g}{\sqrt{\mathsf{D}\,g}\cdot|b|\sqrt{\mathsf{D}\,g}} = \frac{b\,\mathsf{D}\,g}{|b|\,\mathsf{D}\,g} = \\ &= \frac{b}{|b|} \in \{-1, 1\},\end{aligned}$$

что и требовалось.

$\square$

3.4. Конечные дискретные распределения

3.4.1. Общие соображения

Рассмотрим некий процесс, в котором задействованы какие-то участники, разбитые на группы, и каждой группе приписано какое-то число, причём разным группам соответствуют разные числа. К примеру, рассматривается технологический процесс, который может быть проделан на установках разных типов и в каждой группе группе все установки имеют одинаковую производительность, а в разных группах группах – производительности разные.

Итак, формально мы имеем некоторый набор попарно различных чисел $x_1, \ldots, x_m$. Далее, пусть каждому числу x_i приписана некоторая числовая характеристика $k_i \geqslant 0$, $i \in \{1, \ldots, m\}$. Например, это может быть вероятность попадания в данную группу, число элементов в группе и т.д. и т.п. Важно только то, что над этими характеристиками можно производить арифметические операции (например, сложение, деление). Развивая приведённый выше пример с установками,можем считать, что k_i — либо число установок с данной производительностью, либо процент процент таких установок от общего числа установок, либо вероятность проведения данного технологического процесса на данном типе установок (вероятность может определяться какими-то внешними факторами, например, возможностью перемещения материалов к данному типу установок т.д. и т.п.).

Пусть
$$k = k_1 + \cdots + k_m > 0,$$
для любого $i \in \{1, \ldots, m\}$
$$p_i = \frac{k_i}{k},$$
и мы получаем вероятность того, что в *данной* группе будет происходить процесс.

Распределение в этом случае полностью описывается таблицей

x	x_1	x_2	$\ldots$	x_m
$\mathsf{P}(g = x)$	p_1	p_2	$\ldots$	p_m

.

Часто для удобства значения $x_1, \ldots, x_m$ упорядочивают, например так, чтобы
$$x_1 < \cdots < x_m.$$

Также используется геометрический подход к описанию таких распределений, когда на координатной плоскости располагаются точки с координатами (x_i, p_i), $i \in \{1, \ldots, m\}$, которые последовательно соединяются отрезками. Полученную геометрическую фигуру называют *многоугольником (полигоном)* распределения.

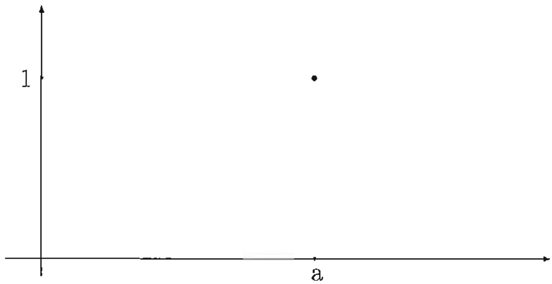


Рис. 1. Многоугольник распределения для вырожденного распределения

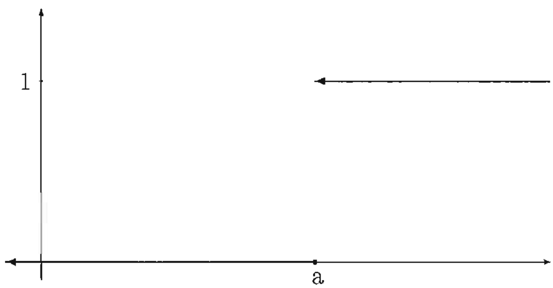


Рис. 2. График функции распределения для вырожденного распределения

3.4.2. Вырожденное распределение

Рассмотрим вырожденный случай, когда $\mathsf{P}(g = a) = 1$ для некоторого $a \in \mathbf{R}$.

Многоугольник распределения и график функции распределения изображены на рисунках 1 и 2.

Ясно, что
$$\mathsf{M}\, g = a \quad \text{и} \quad \mathsf{D}\, g = 0.$$

3.4.3. Пример конечного дискретного распределения

Пусть распределение задаётся таблицей

x	1	2	3	4
$\mathsf{P}(g = x)$	0.1	0.4	0.3	0.2

.

Тогда многоугольник распределения выглядит как на рисунке 3.

Посчитаем математическое ожидание и дисперсию.

$$\mathsf{M}\, g = 1 \cdot 0.1 + 2 \cdot 0.4 + 3 \cdot 0.3 + 4 \cdot 0.2 = 0.1 + 0.8 + 0.9 + 0.8 = 2.6,$$

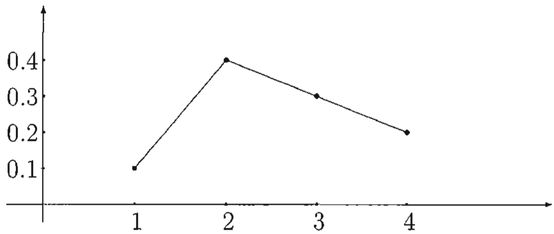


Рис. 3. Многоугольник распределения

$$\begin{aligned}
&\mathsf{M}\, g^2 = 1 \cdot 0.1 + 4 \cdot 0.4 + 9 \cdot 0.3 + 16 \cdot 0.2 = 0.1 + 1.6 + 2.7 + 3.2 = 7.6,\\
&\mathsf{D}\, g = \mathsf{M}\, g^2 - (\mathsf{M}\, g)^2 = 7.6 - (2.6)^2 = 7.6 - 6.76 = 0.84,\\
&\sqrt{\mathsf{D}\, g} = \sqrt{0.84} = 0.916515138991168\ldots
\end{aligned}$$

Ещё посчитаем дисперсию по определению.

$x - \mathsf{M}\, g$	-1.6	-0.6	0.4	1.4
$\mathsf{P}(g - \mathsf{M}\, g = x - \mathsf{M}\, g)$	0.1	0.4	0.3	0.2

и

$(x - \mathsf{M}\, g)^2$	2.56	0.36	0.16	1.96
$\mathsf{P}\left((g - \mathsf{M}\, g)^2 = (x - \mathsf{M}\, g)^2\right)$	0.1	0.4	0.3	0.2

.

Тогда

$$\begin{aligned}
\mathsf{D}\, g = \mathsf{M}\, (g - \mathsf{M}\, g)^2 = 2.56 \cdot 0.1 + 0.36 \cdot 0.4 + 0.16 \cdot 0.3 + 1.96 \cdot 0.2 = \\
= 0.256 + 0.144 + 0.048 + 0.392 = 0.84.
\end{aligned}$$

Далее заметим, что функция распределения

$$F(x) = \begin{cases} 0 & \text{при } x \leqslant 1, \text{ поскольку нет значений (строго) меньших } 1;\\ 0.1 & \text{при } x \in (1, 2], \text{ поскольку есть только одно значение меньшее } 2, \text{ а именно, при } x = 1;\\ 0.5 & \text{при } x \in (2, 3], \text{ поскольку добавляется ещё одно значение меньшее } 3, \text{ а именно, при } x = 2;\\ 0.8 & \text{при } x \in (3, 4], \text{ поскольку добавляется ещё одно значение меньшее } 4, \text{ а именно, при } x = 3;\\ 1 & \text{при } x > 4, \text{ поскольку добавляется последнее возможное значение, а именно, при } x = 4. \end{cases}$$

График функции распределения изображён на рисунке 4.

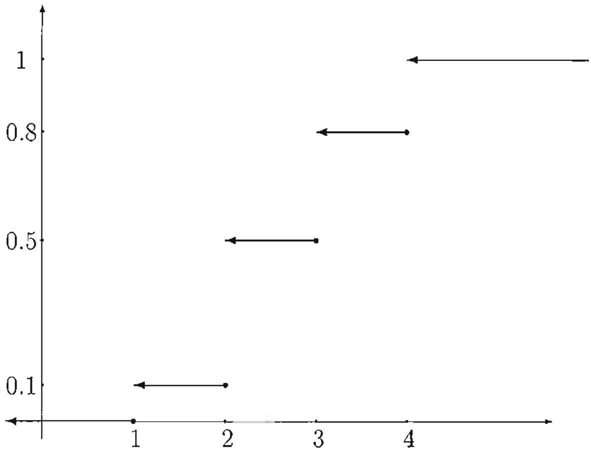


Рис. 4. График функции распределения

3.4.4. Схема Бернулли $\mathbf{B}(p)$

Если случайная величина распределена по схеме Бернулли, то её обозначим $\mathbf{B}(p)$ и имеем

x	0	1
$\mathsf{P}(\mathbf{B}(p)=x)$	$q=1-p$	p

.

Заметим, что $\mathbf{B}(p)=\mathbf{B}(p)^2$, и имеем

$$\mathsf{M}\,\mathbf{B}(p)=0\cdot q+1\cdot p=p,\quad \mathsf{M}\,\mathbf{B}(p)^2=\mathsf{M}\,\mathbf{B}(p)=p \quad \text{и}$$
$$\mathsf{D}\,\mathbf{B}(p)=\mathsf{M}\,\mathbf{B}(p)^2-(\mathsf{M}\,\mathbf{B}(p))^2=p-p^2=p(1-p)=pq.$$

3.4.5. Биномиальное распределение

Биномиальное распределение для r испытаний с вероятностью успеха p будет давать случайную величину (в дальнейшем будем обозначать её $\mathbf{B}_r(p)$).

x	0	1	...	k	...	$r-1$	r
$\mathsf{P}(\mathbf{B}_r(p)=x)$	q^r	rpq^r	...	$C_r^k p^k q^{r-k}$	...	$rp^{r-1}q$	p^r

.

Заметим, что по формуле бинома Ньютона

$$\sum_{x=0}^{r}\mathsf{P}(\mathbf{B}_r(p)=x)=\sum_{x=0}^{r}C_r^x p^x q^{r-x}=(p+q)^r=1,$$

и мы получаем действительно распределение.

Найдём математическое ожидание. Сначала проведём вспомогательное вычисление для любого $x \in \{1, 2, \ldots, r\}$

$$xC_r^x = x \cdot \frac{r!}{x!}(r-x)! = r \cdot \frac{(r-1)!}{(x-1)!}(r-x)! = rC_{r-1}^{x-1}.$$

Теперь имеем

$$\begin{aligned}\mathsf{M}\,\mathbf{B}_r(p) &= \sum_{x=0}^{r} x\,\mathsf{P}(\mathbf{B}_r(p) = x) = \sum_{x=1}^{r} xC_r^x p^x q^{r-x} = \\ &= \sum_{x=1}^{r} rC_{r-1}^{x-1} p^x q^{r-x} = rp\sum_{x=1}^{r} C_{r-1}^{x-1} p^{x-1} q^{r-x} = \\ &= rp\sum_{k=0}^{r-1} C_{r-1}^{k} p^k q^{r-k} = rp(p+q)^{r-1} = rp\cdot 1 = \\ &= rp.\end{aligned}$$

Поэтому говорят

математическое ожидание биномиального распределения равно произведению числа испытаний на вероятность успеха.

Для вычисления дисперсии мы также сначала проведём вспомогательное вычисление для любого $x \in \{2, \ldots, r\}$

$$\begin{aligned}x^2C_r^x &= x^2C_r^x - xC_r^x + xC_r^x = x(x-1)\cdot\frac{r!}{x!}(r-x)! + xC_r^x = \\ &= r(r-1)\cdot\frac{(r-2)!}{(x-2)!}(r-x)! + rC_{r-1}^{x-1} = r(r-1)C_{r-2}^{x-2} + rC_{r-1}^{x-1}.\end{aligned}$$

Имеем

$$\begin{aligned}\mathsf{M}\,\mathbf{B}_r(p)^2 &= \sum_{x=0}^{r} x^2\,\mathsf{P}(\mathbf{B}_r(p) = x) = \sum_{x=1}^{r} x^2C_r^x p^x q^{r-x} = \\ &= rpq^{r-1} + \sum_{x=2}^{r} x^2C_{r-1}^{x-1} p^x q^{r-x} = \\ &= rpq^{r-1} + \sum_{x=2}^{r}\left(r(r-1)C_{r-2}^{x-2} + rC_{r-1}^{x-1}\right) p^x q^{r-x} = \\ &= rpq^{r-1} + \sum_{x=2}^{r} r(r-1)C_{r-2}^{x-2} p^x q^{r-x} + \sum_{x=2}^{r} rC_{r-1}^{x-1} p^x q^{r-x} =\end{aligned}$$

$$= rpq^{r-1} + r(r-1)p^2 \sum_{x=2}^{r} C_{r-2}^{x-2} p^{x-2} q^{r-x} + rp \sum_{x=2}^{r} C_{r-1}^{x-1} p^{x-1} q^{r-x} =$$
$$= r(r-1)p^2 \sum_{k=0}^{r-2} C_{r-2}^{k} p^{k} q^{r-x} + rp \left(q^{r-1} + \sum_{l=1}^{r-1} C_{r-1}^{l} p^{l} q^{r-x} \right) =$$
$$= r(r-1)p^2 (p+q)^{r-2} + rp(p+q)^{r-1} = r(r-1)p^2 \cdot 1 + rp \cdot 1 =$$
$$= r^2p^2 - rp^2 + rp = r^2p^2 + rp(1-p) = r^2p^2 + rpq.$$

Отсюда

$$\mathsf{D}\,\mathbf{B}_r(p) = \mathsf{M}\,\mathbf{B}_r(p)^2 - (\mathsf{M}\,\mathbf{B}_r(p))^2 = r^2p^2 + rpq - (rp)^2 = rpq.$$

Поэтому говорят

> *дисперсия биномиального распределения равно произведению числа испытаний на вероятность успеха и на вероятность неудачи.*

Замечание. **Очень важное!**
Можно истолковывать случайную величину $\mathbf{B}_r(p)$ как сумму r независимых случайных величин с одинаковым распределением $\mathbf{B}(p)$. Тогда по теоремам о свойствах математического ожидания и дисперсии

$$\mathsf{M}\,\mathbf{B}_r(p) = r \cdot \mathsf{M}\,\mathbf{B}(p) = rp \quad \text{и} \quad \mathsf{D}\,\mathbf{B}_r(p) = r \cdot \mathsf{D}\,\mathbf{B}(p) = r(pq).$$

Пример. Рассмотрим биномиальное распределение с $r = 3$ и $p = \frac{1}{4}$. В наших обозначениях это $\mathbf{B}_3(\frac{1}{4})$. В этом случае $q = 1 - p = \frac{3}{4}$ и

$$\mathsf{P}\left(\mathbf{B}_3(\tfrac{1}{4}) = 0\right) = P(0,3) = C_3^0 p^0 q^3 = \left(\frac{3}{4}\right)^3 = \frac{27}{64} = 0.421875,$$
$$\mathsf{P}\left(\mathbf{B}_3(\tfrac{1}{4}) = 1\right) = P(1,3) = C_3^1 p^1 q^2 = 3 \cdot \frac{1}{4} \cdot \left(\frac{3}{4}\right)^2 = \frac{27}{64} = 0.421875,$$
$$\mathsf{P}\left(\mathbf{B}_3(\tfrac{1}{4}) = 2\right) = P(2,3) = C_3^2 p^2 q^1 = 3 \cdot \left(\frac{1}{4}\right)^2 \cdot \frac{3}{4} = \frac{9}{64} = 0.140625,$$
$$\mathsf{P}\left(\mathbf{B}_3(\tfrac{1}{4}) = 3\right) = P(3,3) = C_3^3 p^3 q^0 = \left(\frac{1}{4}\right)^3 = \frac{1}{64} = 0.015625.$$

Получаем

x	0	1	2	3
$\mathsf{P}(\mathbf{B}_3(\frac{1}{4}) = x)$	$\frac{27}{64}$	$\frac{27}{64}$	$\frac{9}{64}$	$\frac{1}{64}$

Тогда многоугольник распределения выглядит как на рисунке 5.

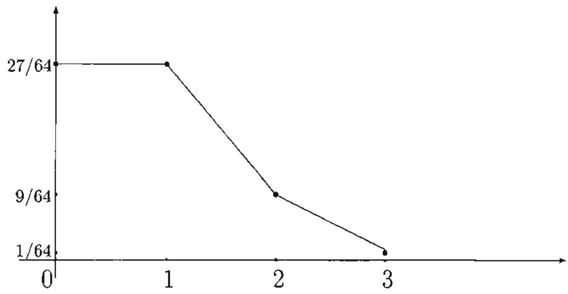


Рис. 5. Многоугольник распределения

Далее заметим, что функция распределения

$$F(x) = \begin{cases} 0 & \text{при } x \leqslant 0, \text{ поскольку нет значений (строго) меньших 0;} \\ \frac{27}{64} & \text{при } x \in (0,1], \text{ поскольку есть только одно значение меньшее 1, а именно, при } x = 0; \\ \frac{27}{32} & \text{при } x \in (1,2], \text{ поскольку добавляется ещё одно значение меньшее 2, а именно, при } x = 1; \\ \frac{63}{64} & \text{при } x \in (2,3], \text{ поскольку добавляется ещё одно значение меньшее 3, а именно, при } x = 2; \\ 1 & \text{при } x > 3, \text{ поскольку добавляется последнее возможное значение, а именно, при } x = 3. \end{cases}$$

График функции распределения изображён на рисунке 6.

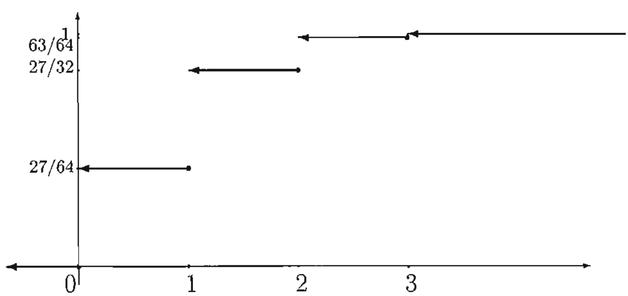


Рис. 6. График функции распределения

Найдём математическое ожидание и дисперсию. Непосредственно имеем

$$\mathsf{M}\,\mathrm{B}_3(\tfrac{1}{4}) = \sum_{x=0}^{3} x\,\mathsf{P}(\mathrm{B}_3(\tfrac{1}{4}) = x) =$$

$$
\begin{aligned}
&= 0 \cdot \frac{27}{64} + 1 \cdot \frac{27}{64} + 2 \cdot \frac{9}{64} + 3 \cdot \frac{1}{64} = \frac{27}{64} + \frac{18}{64} + \frac{3}{64} = \\
&= \frac{27+18+3}{64} = \frac{48}{64} = \frac{3}{4},\\
\mathsf{M}\,\mathbf{B}_3(\tfrac{1}{4})^2 &= \sum_{x=0}^{3} x^2\, \mathsf{P}(\mathbf{B}_3(\tfrac{1}{4}) = x) = \\
&= 0 \cdot \frac{27}{64} + 1 \cdot \frac{27}{64} + 4 \cdot \frac{9}{64} + 9 \cdot \frac{1}{64} = \frac{27}{64} + \frac{36}{64} + \frac{9}{64} = \\
&= \frac{27+36+9}{64} = \frac{72}{64} = \frac{9}{8},\\
\mathsf{D}\,\mathbf{B}_3(\tfrac{1}{4}) &= \mathsf{M}\,\mathbf{B}_3(\tfrac{1}{4})^2 - \left(\mathsf{M}\,\mathbf{B}_3(\tfrac{1}{4})\right)^2 = \frac{9}{8} - \left(\frac{3}{4}\right)^2 = \frac{9}{8} - \frac{9}{16} = \\
&= \frac{9}{16}(2-1) = \frac{9}{16},\\
\sqrt{\mathsf{D}\,g} &= \sqrt{\frac{9}{16}} = \frac{3}{4} = 0.75.
\end{aligned}
$$

Поскольку

$$
\mathsf{M}\,\mathbf{B}_3(\tfrac{1}{4}) = \frac{3}{4} = 3 \cdot \frac{1}{4} = rp \quad \text{и} \quad \mathsf{D}\,\mathbf{B}_3(\tfrac{1}{4}) = \frac{9}{16} = 3 \cdot \frac{1}{4} \cdot \frac{3}{4} = rpq,
$$

то всё согласуется с полученными ранее результатами.

3.4.6. Гипергеометрическое распределение

Как было отмечено в первом разделе, набор

$$
\{P_{k_1,k}(0,s), P_{k_1,k}(1,s), \ldots, P_{k_1,k}(s,s)\}
$$

называется гипергеометрическим распределением для выборки объёма s, и

$$
P_{k_1,k}(s_1,s) = \frac{C_{k_1}^{s_1} C_{k_2}^{s_2}}{C_k^s} = \frac{s!(k-s)!)}{k!} \cdot \frac{k_1!}{s_1!(k_1-s_1)!} \cdot \frac{k_2!}{s_2!(k_2-s_2)!}.
$$

Отметим, что здесь $k_2 = k - k_1$ и $s_2 = s - s_1$. Также необходимо отметить, что приведённое выражение для $\mathsf{P}_{k_1,k}(s_1,s)$ может оказаться бессмысленным, если нарушаются явные ограничения (имеется в виду $s > k$, $s_1 > s$ и $s_1 > k_1$) и неявное оганичение при $s_2 > k_2$. В таких случаях можно либо считать, что $P_{k_1,k}(s_1,s) = 0$, либо исключать эти случаи из рассмотрения.

Поэтому пусть $S = \{\text{все допустимые значения } s_1\}$. Тогда случайная величина g задаётся для любого $s_1 \in S$ так

$$
\mathsf{P}(g = s_1) = P_{k_1,k}(s_1,s) = \frac{C_{k_1}^{s_1} C_{k_2}^{s_2}}{C_k^s} =
$$

$$= \frac{s!(k-s)!)}{k!} \cdot \frac{k_1!}{s_1!(k_1-s_1)!} \cdot \frac{k_2!}{s_2!(k_2-s_2)!}.$$

Поэтому имеем

$$\mathsf{M}\, g = \sum_{s_1 \in S} s_1 \frac{C_{k_1}^{s_1} C_{k_2}^{s_2}}{C_k^s} \quad \text{и} \quad \mathsf{M}\, g^2 = \sum_{s_1 \in S} s_1^2 \frac{C_{k_1}^{s_1} C_{k_2}^{s_2}}{C_k^s}.$$

Представляется весьма сложным нахождение в общем виде выражений для математического ожидания и дисперсии гипергеометрического распределения.

Пример. Рассмотрим гипергеометрическое распределение с $k = 5$, $k_1 = 2$ и $s = 3$. Число исходов равно

$$C_5^3 = \frac{5!}{3!2!} = \frac{5 \cdot 4}{2} = 10.$$

При выборке объёма $s = 3$ *не* можем выбрать, более 2 чёрных шаров, ибо общее число чёрных шаров $k_1 = 2$. Поэтому

$$s_1 \in \{0, 1, 2\},$$

и имеем распределение

$$\mathsf{P}\,(g = 0) = P_{2,5}(0,3) = \frac{C_2^0 C_3^3}{C_5^3} = \frac{1 \cdot 1}{10} = \frac{1}{10} = 0.1,$$

Пример.

$$\mathsf{P}\,(g = 1) = P_{2,5}(1,3) = \frac{C_2^1 C_3^2}{C_5^3} = \frac{2 \cdot 3}{10} = \frac{6}{10} = \frac{3}{5} = 0.6,$$

$$\mathsf{P}\,(g = 2) = P_{2,5}(2,3) = \frac{C_2^2 C_3^1}{C_5^3} = \frac{1 \cdot 3}{10} = \frac{3}{10} = 0.3.$$

Получаем

x	0	1	2
$\mathsf{P}(g = x)$	0.1	0.6	0.3

Так как

$$\sum_{x=0}^{2} \mathsf{P}(g = x) = 0.1 + 0.6 + 03 = 1,$$

то получаем действительно распределение.

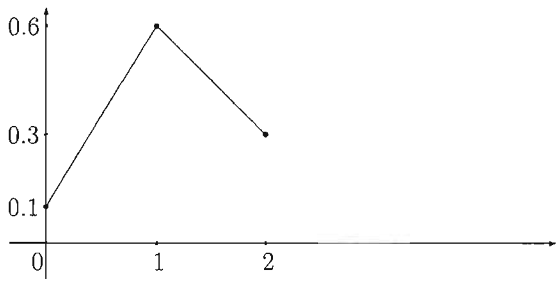


Рис. 7. Многоугольник распределения

Для функции распределения имеем

$$F_g(x) = \begin{cases} 0 & \text{при } x \leqslant 0, \text{ поскольку нет значений (строго) меньших } 0; \\ 0.1 & \text{при } x \in (0,1], \text{ поскольку есть только одно значение меньшее } 1, \text{ а именно, при } x = 0; \\ 0.7 & \text{при } x \in (1,2], \text{ поскольку добавляется ещё одно значение меньшее } 2, \text{ а именно, при } x = 1; \\ 1 & \text{при } x > 2, \text{ поскольку добавляется последнее возможное значение, а именно, при } x = 2. \end{cases}$$

График функции распределения изображён на рисунке 8.

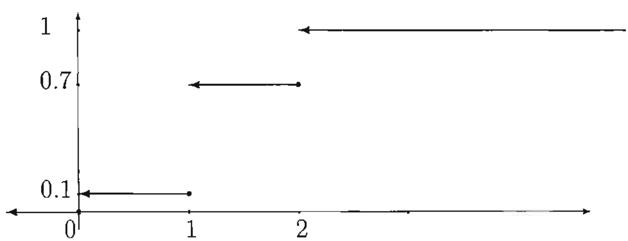


Рис. 8. График функции распределения

Найдём математическое ожидание и дисперсию. Непосредственно имеем

$$\mathsf{M}\, g = \sum_{x=0}^{2} x\, \mathsf{P}(g = x) = 0 \cdot 0.1 + 1 \cdot 0.6 + 2 \cdot 0.3 = 0 + 0.6 + 0.6 = 1.2,$$

$$\begin{aligned}
\mathsf{M}\, g^2 &= \sum_{x=0}^{2} x^2\, \mathsf{P}(g) = x) = 0 \cdot 0.1 + 1 \cdot 0.6 + 4 \cdot 0.3 = 0 + 0.6 + 1.2 = \\
&= 1.8, \\
\mathsf{D}\, g &= \mathsf{M}\, g^2 - (\mathsf{M}\, g)^2 = 1.8 - 1.44 = 0.36, \\
\sqrt{\mathsf{D}\, g} &= \sqrt{0.36} = 0.6.
\end{aligned}$$

3.5. Бесконечные (счётные) дискретные распределения

Опишем этот случай чисто формально, а содержательно его можно описать подобно конечному случаю.

Пусть задана последовательность попарно различных чисел $\{x_i\}_{i=1}^{\infty}$, каждому из которых приписано число $k_i \geqslant 0$. Допустим, что ряд $\sum_{i=1}^{\infty} k_i$ сходится, и $K > 0$ — его сумма. Положим для каждого натурального i

$$p_i = \frac{k_i}{K} \quad \text{и} \quad \sum_{i=1}^{\infty} p_i = \sum_{i=1}^{\infty} \frac{k_i}{K} = \frac{1}{K} \sum_{i=1}^{\infty} k_i = \frac{K}{K} = 1.$$

Можно распределение изобразить в виде полубесконечной таблицы

x_1	x_2	$\ldots$	x_i	$\ldots$
p_1	p_2	$\ldots$	p_i	$\ldots$

.

3.5.1. Счётные распределения, связанные с гармоническим рядом с показателем 2

Известно, что

$$\sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{\pi^2}{6}.$$

Поэтому для любой последовательность попарно различных чисел $\{x_i\}_{i=1}^{\infty}$, если положить

$$\mathsf{P}(g = x_i) = \frac{6}{\pi^2} \cdot \frac{1}{i^2},$$

то получим дискретную бесконечную (счётную) случайную величину g. В частности, это верно, когда для всех натуральных i

$$x_i \in \left\{i, (-1)^i i, \tfrac{1}{i}\right\}.$$

Примеры. Существование математического ожидания и дисперсии очень зависят от того, чему равны x_i, $i \in \{1, 2, \ldots, n, \ldots\}$

$x_i = i$. В этом случае

$$\sum_{i=1}^{\infty} i \cdot \frac{6}{\pi^2} \cdot \frac{1}{i^2} = \frac{6}{\pi^2} \sum_{i=1}^{\infty} \frac{1}{i} = \infty,$$

поскольку получилось ненулевое кратное гармонического ряда, который расходится. Поэтому нет математического ожидания и дисперсии.

$x_i = (-1)^i i$. Теперь

$$\sum_{i=1}^{\infty} (-1)^i i \cdot \frac{6}{\pi^2} \cdot \frac{1}{i^2} = \frac{6}{\pi^2} \sum_{i=1}^{\infty} \frac{(-1)^i}{i},$$

и мы получаем знакочередующийся лейбницевский ряд, которых сходится. Его сумма равна

$$\begin{aligned} \mathsf{M}\, g &= \frac{6}{\pi^2} \cdot (-\ln 2) = -\frac{6 \ln 2}{\pi^2} = \\ &= -\frac{4,1588830833596718565033927287491}{9,8696044010893586188344909998762} = \\ &= -0,42138295663609725845363159698198. \end{aligned}$$

Для дисперсии имеем

$$\sum_{i=1}^{\infty} i^2 \cdot \frac{6}{\pi^2} \cdot \frac{1}{i^2} = \frac{6}{\pi^2} \sum_{i=1}^{\infty} 1 = \infty.$$

Поэтому дисперсии нет!

$x_i = \frac{1}{i}$. Имеем

$$\sum_{i=1}^{\infty} \frac{1}{i} \cdot \frac{6}{\pi^2} \cdot \frac{1}{i^2} = \frac{6}{\pi^2} \sum_{i=1}^{\infty} \frac{1}{i^3}.$$

Сумма ряда

$$\sum_{i=1}^{\infty} \frac{1}{i^3} = \zeta(3) = 1,2020569031595942853997381615114499907649862 92.$$

Замечание. Для любопытных отметим, что $\zeta(3)$ — значение знаменитой дзета функции Римана $\zeta(s)$ при $s = 3$. Про это число мало что известно. Только в 1978 году французский математик Роже́ Апери́ (Roger Apéry) доказал, что оно иррационально, но его трансцендентность неизвестна. Называют его теперь также *постоянной Апери*.

Итак
$$\mathsf{M}\,g = \frac{6\zeta(3)}{\pi^2} = 0,7307629694014384987260367313076 8.$$

Найдём дисперсию. Известно, что
$$\sum_{i=1}^{\infty} \frac{1}{i^4} = \frac{\pi^4}{90}.$$

Поэтому
$$\mathsf{M}\,g^2 = \frac{6}{\pi^2} \cdot \frac{\pi^4}{90} = \frac{\pi^2}{15} = 0,6579736267392905745889660666584 1.$$

Теперь
$$\mathsf{D}\,g = \mathsf{M}\,g^2 - (\mathsf{M}\,g)^2 = 0,1239591092908828350279305596520 8,$$
$$\sqrt{\mathsf{D}\,g} = 0,3520782715404102227823757842249 6.$$

3.5.2. Распределение Пуассона

Распределение Пуассона обозначается Π_λ и задаётся для любых $\lambda > 0$ и целого неотрицательного m
$$\mathsf{P}(\Pi_\lambda = m) = e^{-\lambda}\frac{\lambda^m}{m!},$$
в частности, $\mathsf{P}(\Pi_\lambda = 0) = e^{-\lambda}$, $\mathsf{P}(\Pi_\lambda = 1) = e^{-\lambda}\lambda$ и $\mathsf{P}(\Pi_\lambda = 2) = e^{-\lambda}\frac{\lambda^2}{2}$.

Так как
$$\sum_{m=0}^{\infty} \mathsf{P}(\Pi_\lambda = m) = \sum_{m=0}^{\infty} e^{-\lambda}\frac{\lambda^m}{m!} = e^{-\lambda}\sum_{m=0}^{\infty}\frac{\lambda^m}{m!} = e^{-\lambda}e^{\lambda} = 1,$$
то мы действительно получаем распределение.

Найдём математическое ожидание
$$\mathsf{M}\,\Pi_\lambda = \sum_{m=0}^{\infty} m \cdot e^{-\lambda}\frac{\lambda^m}{m!} = e^{-\lambda}\sum_{m=1}^{\infty} m\frac{\lambda^m}{m!} = e^{-\lambda}\lambda\sum_{m=1}^{\infty}\frac{\lambda^{m-1}}{(m-1)!} =$$
$$= e^{-\lambda}\lambda\sum_{l=0}^{\infty}\frac{\lambda^l}{l!} = e^{-\lambda}\lambda \cdot e^{\lambda} = \lambda.$$

Далее

$$\mathsf{M}(\Pi_\lambda)^2 = \sum_{m=0}^{\infty} m^2 \cdot e^{-\lambda}\frac{\lambda^m}{m!} = e^{-\lambda}\sum_{m=1}^{\infty} m^2\frac{\lambda^m}{m!} = e^{-\lambda}\lambda\sum_{m=1}^{\infty} m\frac{\lambda^{m-1}}{(m-1)!} =$$

$$= e^{-\lambda}\lambda\sum_{m=1}^{\infty}(m-1+1)\frac{\lambda^{m-1}}{(m-1)!} =$$

$$= e^{-\lambda}\lambda\sum_{m=1}^{\infty}(m-1)\frac{\lambda^{m-1}}{(m-1)!} + e^{-\lambda}\lambda\sum_{m=1}^{\infty}\frac{\lambda^{m-1}}{(m-1)!} =$$

$$= e^{-\lambda}\lambda\sum_{m=2}^{\infty}\frac{\lambda^{m-1}}{(m-2)!} + e^{-\lambda}\lambda\sum_{l=0}^{\infty}\frac{\lambda^l}{l!} =$$

$$= e^{-\lambda}\lambda^2\sum_{m=2}^{\infty}\frac{\lambda^{m-2}}{(m-2)!} + e^{-\lambda}\lambda\cdot e^{\lambda} =$$

$$= e^{-\lambda}\lambda^2\sum_{k=0}^{\infty}\frac{\lambda^k}{k!} + \lambda = e^{-\lambda}\lambda^2\cdot e^{\lambda} + \lambda = \lambda^2 + \lambda.$$

Найдём дисперсию

$$\mathsf{D}\,\Pi_\lambda = \mathsf{M}(\Pi_\lambda)^2 - (\mathsf{M}\,\Pi_\lambda)^2 = \lambda^2 + \lambda - (\lambda)^2 = \lambda.$$

Итак

$$\mathsf{M}\,\Pi_\lambda = \mathsf{D}\,\Pi_\lambda = \lambda,$$

и говорят, что

математическое ожидание и дисперсия распределение Пуассона совпадают с его параметром.

3.5.3. Геометрическое распределение

Для случайной величины g положим для любых $p \in [0,1]$ и целого неотрицательного m

$$\mathsf{P}(g_p = m) = (1-p)p^m.$$

Это можно истолковывать как стрельбу до первого промаха, если вероятность попадания при одном выстреле равна p. Такого сорта вероятности возникают при перестрелках при определении победителей в стендовой стрельбе. Как обычно положим $q = 1 - p$. Очевидно, что случаи $p \in \{0,1\}$ приводят к вырожденным распределениям, и потому не интересны. В последующем всегда считаем, что $p \in (0,1)$. В этом случае получаем распределение, т.к.

$$\sum_{m=0}^{\infty}\mathsf{P}(g_p = m) = \sum_{m=0}^{\infty}(1-p)p^m = \frac{1-p}{1-p} = 1.$$

Ряд

$$\sum_{m=0}^{\infty} p^m = \frac{1}{1-p}$$

продифференцируем по p и получим

$$\sum_{m=1}^{\infty} mp^{m-1} = \frac{1}{(1-p)^2} = \frac{1}{q^2}.$$

Используем этот факт для нахождения математического ожидания

$$\mathsf{M}\, g_p = \sum_{m=0}^{\infty} m \cdot qp^m = q \sum_{m=1}^{\infty} mp^m = qp \sum_{m=1}^{\infty} mp^{m-1} = qp \cdot \frac{1}{q^2} = \frac{p}{q}.$$

Для нахождения дисперсии ряд

$$\sum_{m=0}^{\infty} p^m = \frac{1}{1-p}$$

продифференцируем по p дважды и получим

$$\sum_{m=2}^{\infty} m(m-1)p^{m-2} = \left(\frac{1}{(1-p)^2}\right)' = \frac{-2(1-p)(-1)}{(1-p)^4} = \frac{2}{(1-p)^3} = \frac{2}{q^3}.$$

Теперь

$$\begin{aligned}
\mathsf{M}\, g_p^2 &= \sum_{m=0}^{\infty} m^2 \cdot qp^m = \sum_{m=1}^{\infty} (m^2 - m + m)qp^m = \\
&= \sum_{m=1}^{\infty} mqp^m + \sum_{m=1}^{\infty} (m^2 - m)qp^m = \\
&= \mathsf{M}\, g_p + qp^2 \sum_{m=2}^{\infty} m(m-1)p^{m-2} = \\
&= \frac{p}{q} + qp^2 \cdot \frac{2}{q^3} = \frac{p}{q} + \frac{2p^2}{q^2}.
\end{aligned}$$

Найдём дисперсию

$$\begin{aligned}
\mathsf{D}\, g_p &= \mathsf{M}\, g_p^2 - (\mathsf{M}\, g_p)^2 = \frac{p}{q} + \frac{2p^2}{q^2} - \frac{p^2}{q^2} = \frac{p}{q} + \frac{p^2}{q^2} = \frac{pq + p^2}{q^2} = \\
&= \frac{p(q+p)}{q^2} = \frac{p}{q^2} = \frac{1-q}{q^2}.
\end{aligned}$$

Замечание. В задачнике Гмурмана (задача **222***) рассматривается случайная величина X, которая определена для всех натуральных k с вероятностями

$$\mathsf{P}(X = k) = q^{k-1}p.$$

Ясно, что $X = g_q + 1$ и

$$\mathsf{M}\,X = \frac{q}{p} + 1 = \frac{q+p}{p} = \frac{1}{p} \quad \text{и} \quad \mathsf{D}\,X = \mathsf{D}\,g_q = \frac{1-p}{p^2}.$$

3.6. Непрерывные распределения

3.6.1. Свойства плотности

Пусть $p : \mathbf{R} \to \mathbf{R}$ — плотность некоторого непрерывного распределения случайной величины g. Так как для любых действительных a и b со свойством $a < b$ имеем

$$\mathsf{P}(g \in [a,b)) = F_g(b) - F_g(a) = \int\limits_a^b p(t)\,dt \geqslant 0,$$

то ясно, что функция p не может быть произвольной. А именно, если она непрерывна на отрезке $[a,b)$, то она должна быть неотрицательна. В самом деле, пусть $x_0 \in [a,b)$ и $p(x_0) < 0$. Тогда в силу свойства непрерывных функций (точнее, это свойство предела функции в точке) существует такая окрестность $[c,d)$ (может быть односторонняя) точки x_0, в которой функция p отрицательна, но тогда

$$\mathsf{P}(g \in [c,d)) = F_g(d) - F_g(c) = \int\limits_c^d p(t)\,dt < 0,$$

что невозможно. Все приведённые далее примеры (равномерное, нормальное, показательное и Коши распределения) имели неотрицательную, кусочно непрерывную (а нормальное и Коши даже всюду непрерывную) плотность.

Из сказанного выше вытекает, что можно строить распределение так. Выбрать кусочно непрерывную или кусочно монотонную неотрицательную, кроме не более чем счётного числа точек, такую функцию $f : \mathbf{R} \to \mathbf{R}$, что $\int_{-\infty}^{\infty} f(t)dt = S > 0$ сходится в $-\infty$ и в $+\infty$. Тогда в качестве плотности можно взять функцию

$$p(t) = \frac{f(t)}{S}, \text{ так как } \int_{-\infty}^{\infty} p(t)dt = \frac{1}{S}\int_{-\infty}^{\infty} f(t)dt = \frac{S}{S} = 1,$$

а остальные свойства очевидны из свойств интеграла, и мы получаем, что $F(x) = \int_{-\infty}^{x} p(t)dt$ будет являться функцией распределения некоторой случайной величины.

3.6.2. Равномерное распределение

Это распределение, с которым мы уже встречались. Зафиксируем некоторый (конечный) отрезок

$$[a, b] \quad (-\infty < a < b < \infty).$$

Рассмотрим случайную величину $g = \mathbf{U}_{a,b}$, которая состоит в бросании (можно наоборот в выборе из отрезка) точки в отрезок $[a, b]$.

Замечание. Строго говоря, положим $[a, b] = \Omega$, $\mathcal{A} = \mathcal{B} \cap [a, b]$ (борелевские подмножества в $[a, b]$) и пусть $g(\omega) = \omega \in [a, b]$. Тогда

$$g : \Omega \to [a, b] \subset \mathbf{R}$$

— случайная величина.

1) Функция распределения и плотность

Для $x \in \mathbf{R}$ определим $\mathsf{P}(g < x) = F_g(x)$, как вероятность попадания попадания точки в отрезок $[a, x)$. Возможны следующие случаи.

$x \leqslant a$. Так как $[a, x) = \varnothing$ в этом случае, то $F_g(x) = 0$.

$x \in (a, b]$. Будем считать, что $\mathsf{P}(g < x) = F_g(x)$ пропорциональна длине отрезка и положим

$$\mathsf{P}(g < x) = F_g(x) = \frac{x-a}{b-a},$$

то есть, мы считаем, что данная вероятность равна

отношению длины $x - a$ полуинтервала $[a, x)$ к длине $b - a$ отрезка $[a, b]$.

$x > b$. Теперь $[a, x) \supseteq [a, b]$, и потому $\mathsf{P}(g < x) = F_g(x) = 1$.

Таким образом получаем функцию распределения

$$F_g(x) = \begin{cases} 0 & \text{при } x \leqslant a, \\ \frac{x-a}{b-a} & \text{при } x \in (a, b], \\ 1 & \text{при } x > b. \end{cases}$$

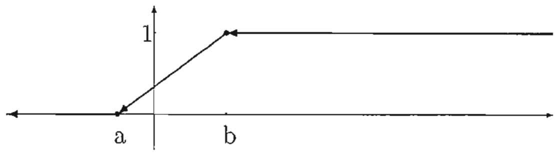


Рис. 9. График функции распределения

Плотность распределения равна

$$\mathrm{pd}_g(x) = F_g(x)' = \begin{cases} 0 & \text{при } x < a, \\ \frac{1}{b-a} & \text{при } x \in (a,b), \\ 0 & \text{при } x > b. \end{cases}$$

и её график на рисунке 10.

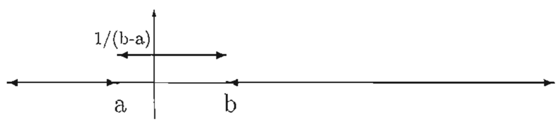


Рис. 10. График плотности распределения

Заметим, что в точках a и b производная F_g' не определена, и можно плотность доопределить произвольно, ибо для любого $x \in \mathbf{R}$ на *интеграл*

$$\int_{-\infty}^{x} \mathrm{pd}(t)\, dt = F_g(x)$$

это *никак* не повлияет!

2. Математическое ожидание

Найдём математическое ожидание

$$\mathsf{M}\, g = \int_{-\infty}^{\infty} x\, \mathrm{pd}_g(x)\, dx = \int_a^b \frac{x}{b-a} dx = \frac{1}{b-a} \cdot \frac{x^2}{2}\bigg|_a^b = \\ = \frac{b^2 - a^2}{2(b-a)} = \frac{(b+a)(b-a)}{2(b-a)} = \frac{b+a}{2}.$$

Поэтому говорят, что

математическое ожидание равномерного распределения на отрезке $[a, b]$ *совпадает с серединой этого отрезка.*

3. Дисперсия

Найдём дисперсию

$$
\begin{aligned}
\mathsf{M}\, g^2 &= \int_{-\infty}^{\infty} x^2\, \mathrm{pd}_g(x)\, dx = \int_a^b \frac{x^2}{b-a} dx = \frac{1}{b-a} \cdot \frac{x^3}{3}\Big|_a^b = \\
&= \frac{b^3 - a^3}{3(b-a)} = \frac{(b-a)(b^2+ab+a^2)}{3(b-a)} = \frac{b^2+ab+a^2}{3}, \\
\mathsf{D}\, g &= \mathsf{M}\, g^2 - (\mathsf{M}\, g)^2 = \frac{b^2+ab+a^2}{3} - \left(\frac{b+a}{2}\right)^2 = \\
&= \frac{b^2+ab+a^2}{3} - \frac{b^2+2ab+a^2}{4} = \\
&= \frac{4b^2+4ab+4a^2-3b^2-6ab-3a^2}{12} = \\
&= \frac{b^2-2ab+a^2}{12} = \frac{(b-a)^2}{12}.
\end{aligned}
$$

3.6.3. Нормальное распределение (закон Гаусса)

Это распределение — самое основное в теории вероятностей и математической статистике, как мы увидим позже. Можно даже сказать:

“Нормальный закон правит миром!”

Определение. Будем говорить, что случайная величина g имеет *нормальное распределение с параметрами* (a, σ^2) (здесь $a, \sigma \in \mathbf{R}$ и $\sigma > 0$), если для любого $x \in \mathbf{R}$ функция распределения

$$F_g(x) = \Phi_{a,\sigma^2}(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{(t-a)^2}{2\sigma^2}} dt.$$

Замечание. Иногда вместо выражения

> случайная величина g имеет нормальное распределение с параметрами (a, σ^2)

будем писать (если это не вызывает путаницы)

$$g = \Phi_{a,\sigma^2}.$$

Исследуем более тщательно определение нормального распределения.

1. Корректность определения

Интеграл

$$\frac{1}{\sigma\sqrt{2\pi}}\int e^{-\frac{(t-a)^2}{2\sigma^2}}dt$$

сходится в $-\infty$ и $+\infty$, поскольку e^{-u^2} убывает на бесконечности быстрее любой степени $\frac{1}{u^n}$, $n \geqslant 2$. В силу этого заданная так функция определена корректно.

2. Переход к стандартным параметрам

Отметим что, интеграл $\frac{1}{\sigma\sqrt{2\pi}}\int e^{-\frac{(t-a)^2}{2\sigma^2}}dt$ не берётся в элементарных функциях, поэтому подсчёт его значений может быть только приближённым. Отметим сразу, что разложением в ряд Тейлора функции $e^{-\frac{(t-a)^2}{2\sigma^2}}$ в точке a получаем ряд, который будет очень быстро сходиться. Интегрируя его почленно, получим, в принципе, очень быстро довольно точные значения. Однако наличие двух параметров (a, σ^2) привело бы к тому, что нужно было бы *необыкновенно огромное* количество таблиц значений $\Phi_{a,\sigma^2}(x)$. На самом деле можно обойтись всего одной парой $(0,1)$ параметров, эти параметры называются *стандартными*, а нормальное распределение с параметрами $(0,1)$ назовём *стандартным.*

Чтобы понять это, сделаем замену

$$u = \frac{t-a}{\sigma} \quad \text{и} \quad du = \frac{dt}{\sigma}.$$

Откуда по формуле замены переменной в определённом (несобственном) интеграле получим

$$\Phi_{a,\sigma^2}(x) = \frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{(t-a)^2}{2\sigma^2}}dt = \frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\frac{x-a}{\sigma}} e^{-\frac{u^2}{2}}\sigma\,dt =$$
$$= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\frac{x-a}{\sigma}} e^{-\frac{u^2}{2}}dt = \Phi_{0,1}\left(\frac{x-a}{\sigma}\right).$$

Замечание. Таким образом, можно сказать, что

> в нормальном распределении с параметрами (a, σ^2) параметр a играет роль *"сдвига"*, а параметр σ — роль *"масштаба"*. Геометрически, как можно увидеть на приведённых далее графиках, параметр a сдвигает график, параметр σ поджимает график.

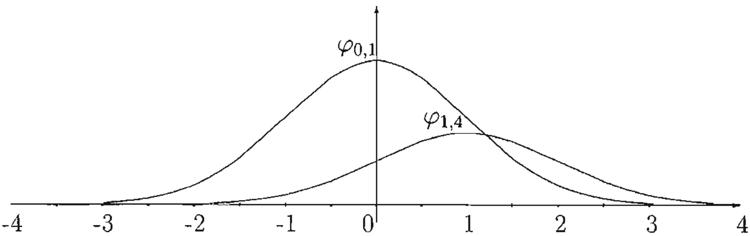


Рис. 11. Плотность нормального распределения φ_{a,σ^2}

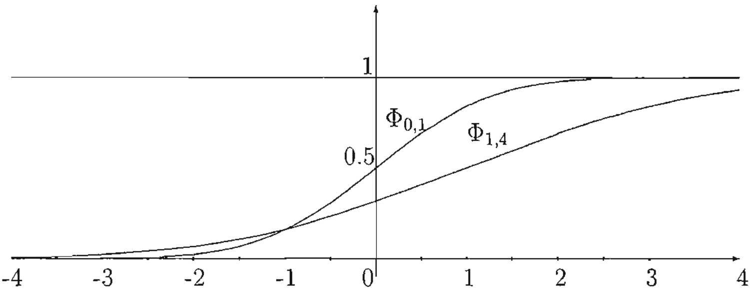


Рис. 12. Функция нормального распределения Φ_{a,σ^2}

3. Основной интеграл теории вероятностей

Так как функция $\frac{1}{\sqrt{2\pi}}e^{-\frac{u^2}{2}} > 0$ и непрерывна, то ясно, что все утверждения из теоремы о свойствах функции распределения очевидно выполняются кроме, быть может, свойства

$$1 = \lim_{x\to+\infty} F_g(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{u^2}{2}} du = \Phi_{0,1}(+\infty).$$

Этот интеграл (и равносильные ему) часто называют *основным интегралом теории вероятностей*, или *интегралом Пуассона*, или *интегралом Эйлера–Пуассона*, или *интегралом Гаусса*.

Так как $e^{-\frac{u^2}{2}}$ — чётная функция и интеграл $\int_{-\infty}^{\infty} e^{-\frac{u^2}{2}} du$ сходится в $-\infty$ и $+\infty$ (отмечено ранее), то

$$\int_{-\infty}^{\infty} e^{-\frac{u^2}{2}} du = 2\int_0^{\infty} e^{-\frac{u^2}{2}} du.$$

Сделаем замену $u = \sqrt{2}v$ и $du = \sqrt{2}dv$, получим

$$\int_0^{\infty} e^{-\frac{u^2}{2}} du = \sqrt{2}\int_0^{\infty} e^{-v^2} dv \quad \text{и} \quad \int_{-\infty}^{\infty} e^{-\frac{u^2}{2}} du = 2\sqrt{2}\int_0^{\infty} e^{-v^2} dv.$$

Для нахождения интеграла $I = \int_0^\infty e^{-v^2} dv$ используется следующий остроумный приём. Подсчитаем квадрат данного интеграла.

$$I^2 = \left(\int_0^\infty e^{-v^2} dv\right)^2 = \left(\int_0^\infty e^{-v^2} dv\right) \cdot \left(\int_0^\infty e^{-w^2} dw\right) =$$

перейдём к двойному интегралу

$$= \iint\limits_{\substack{0 \leqslant v < \infty \\ 0 \leqslant w < \infty}} e^{-(v^2+w^2)} dv dw.$$

Теперь перейдём к полярным координатам $v = r\cos\varphi$ и $w = r\sin\varphi$. Для замены переменных в интеграле надо подсчитать якобиан

$$J = \det\begin{pmatrix} \frac{\partial v}{\partial r} & \frac{\partial w}{\partial r} \\ \frac{\partial v}{\partial \varphi} & \frac{\partial w}{\partial \varphi} \end{pmatrix} = \det\begin{pmatrix} \cos\varphi & \sin\varphi \\ -r\sin\varphi & r\cos\varphi \end{pmatrix} = r(\cos^2\varphi + \sin^2\varphi) = r.$$

Далее заметим, что пределы интегрирования в полярных координатах таковы:

$$r \in [0, \infty) \quad \text{и} \quad \varphi \in [0, \tfrac{\pi}{2}].$$

Получаем, что

$$I^2 = \iint\limits_{\substack{0 \leqslant v < \infty \\ 0 \leqslant w < \infty}} e^{-(v^2+w^2)} dv dw = \int\limits_0^{\frac{\pi}{2}} d\varphi \int\limits_0^\infty e^{-r^2} r\, dr = \frac{\pi}{2} \cdot \frac{1}{2} \int\limits_0^\infty e^{-r^2}\, dr^2 =$$
$$= \frac{\pi}{4}\left(-e^{-r^2}\right)\Big|_0^\infty = \frac{\pi}{4}(-0+1) = \frac{\pi}{4}.$$

Отсюда

$$\int_0^\infty e^{-v^2} dv = I = \frac{\sqrt{\pi}}{2},$$
$$\int_{-\infty}^\infty e^{-\frac{u^2}{2}} du = 2\sqrt{2}\int_0^\infty e^{-v^2} dv = 2\sqrt{2} \cdot I = 2\sqrt{2} \cdot \frac{\sqrt{\pi}}{2} = \sqrt{2\pi} \quad \text{и}$$
$$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^\infty e^{-\frac{u^2}{2}} du = \Phi_{0,1}(+\infty) = 1,$$

что и надо! В частности мы имеем

$$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^0 e^{-\frac{u^2}{2}} du = \frac{1}{\sqrt{2\pi}}\int_0^\infty e^{-\frac{u^2}{2}} du = \frac{1}{2}.$$

4. Попадание в интервал для различных вариантов функции Φ

Важной задачей является нахождение вероятности попадания случайной величины g, распределённой по стандартному нормальному закону в конечный полуинтервал $[a, b)$, где $-\infty < a < b < \infty$. Как понимали раньше

$$\mathsf{P}(g \in [a, b)) = \Phi_{0,1}(b) - \Phi_{0,1}(a).$$

Замечание. Необычайно важно указать на то, что часто функцию $\Phi_{0,1}$ обозначают через Φ, но (!!!) в разных книгах под функцией Φ понимаются ***разные*** функции.

Рассмотрим нахождение вероятности попадания случайной величины g, распределённой по стандартному нормальному закону в конечный полуинтервал $[a, b)$, где $-\infty < a < b < \infty$, для различных заданий функции $\Phi_{0,1}$.

Задачник и учебник Гмурмана

В книгах Гмурмана протабулирована функция

$$\Phi_{\text{Гм}}(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{u^2}{2}} du \quad \text{для } x \leqslant 0.$$

$x \geqslant 0$. Имеем

$$\Phi_{0,1}(x) = \int_{-\infty}^x e^{-\frac{u^2}{2}} du = \int_{-\infty}^0 e^{-\frac{u^2}{2}} du + \int_0^x e^{-\frac{u^2}{2}} du = \\ = \frac{1}{2} + \Phi_{\text{Гм}}(x).$$

$x \leqslant 0$. В этом случае в силу чётности функции $e^{-\frac{u^2}{2}}$ получаем

$$\begin{aligned} \Phi_{0,1}(x) &= \int_{-\infty}^x e^{-\frac{u^2}{2}} du = \int_{-\infty}^0 e^{-\frac{u^2}{2}} du + \int_0^x e^{-\frac{u^2}{2}} du = \\ &= \frac{1}{2} - \int_x^0 e^{-\frac{u^2}{2}} du = \frac{1}{2} + \int_{-x}^0 e^{-\frac{u^2}{2}} du = \\ &= \frac{1}{2} - \int_0^{-x} e^{-\frac{u^2}{2}} du = \\ &= \frac{1}{2} - \Phi_{\text{Гм}}(-x) = \frac{1}{2} - \Phi_{\text{Гм}}(|x|). \end{aligned}$$

Вероятность попадания случайной величины g, распределённой по стандартному нормальному закону в конечный полуинтервал $[a, b)$, где $-\infty < a < b < \infty$, для функции $\Phi_{\text{Гм}}$.

$0 \leqslant a < b$. В этом случае

$$\begin{aligned} \mathsf{P}(g \in [a,b)) &= \Phi_{0,1}(b) - \Phi_{0,1}(a) = \\ &= \left(\tfrac{1}{2} + \Phi_{\text{Гм}}(b)\right) - \left(\tfrac{1}{2} + \Phi_{\text{Гм}}(a)\right) = \\ &= \Phi_{\text{Гм}}(b) - \Phi_{\text{Гм}}(a). \end{aligned}$$

$a < 0 \leqslant b$. Теперь

$$\begin{aligned} \mathsf{P}(g \in [a,b)) &= \Phi_{0,1}(b) - \Phi_{0,1}(a) = \\ &= \left(\tfrac{1}{2} + \Phi_{\text{Гм}}(b)\right) - \left(\tfrac{1}{2} - \Phi_{\text{Гм}}(|a|)\right) = \\ &= \Phi_{\text{Гм}}(b) + \Phi_{\text{Гм}}(|a|). \end{aligned}$$

$a < b \leqslant 0$. Наконец

$$\begin{aligned} \mathsf{P}(g \in [a,b)) &= \Phi_{0,1}(b) - \Phi_{0,1}(a) = \\ &\left(\tfrac{1}{2} - \Phi_{\text{Гм}}(|b|)\right) - \left(\tfrac{1}{2} - \Phi_{\text{Гм}}(|a|)\right) = \\ &= \Phi_{\text{Гм}}(|a|) - \Phi_{\text{Гм}}(|b|). \end{aligned}$$

Учебник Боровкова

В данной книге протабулирована функция

$$\Phi_{\text{Б}}(x) = \Phi_0(x) = \frac{1}{\sqrt{2\pi}} \int_x^{\infty} e^{-\frac{u^2}{2}} du \quad \text{для } x \leqslant 0.$$

$x \geqslant 0$. Имеем

$$\begin{aligned} \Phi_{0,1}(x) &= \int_{-\infty}^{x} e^{-\frac{u^2}{2}} du = \\ &= \int_{-\infty}^{\infty} e^{-\frac{u^2}{2}} du - \int_x^{\infty} e^{-\frac{u^2}{2}} du = \\ &= 1 - \Phi_{\text{Б}}(x). \end{aligned}$$

$x \leqslant 0$. В этом случае получаем

$$\begin{aligned} \Phi_{0,1}(x) &= \int_{-\infty}^{x} e^{-\frac{u^2}{2}} du = \int_{-\infty}^{0} e^{-\frac{u^2}{2}} du - \int_x^{0} e^{-\frac{u^2}{2}} du = \\ &= \frac{1}{2} + \int_{-x}^{0} e^{-\frac{u^2}{2}} du = \frac{1}{2} - \int_0^{-x} e^{-\frac{u^2}{2}} du = \\ &= \frac{1}{2} - \left(\int_0^{\infty} e^{-\frac{u^2}{2}} d(-u) - \int_{-x}^{\infty} e^{-\frac{u^2}{2}} du\right) = \\ &= \frac{1}{2} - \left(\frac{1}{2} - \Phi_{\text{Б}}(-x)\right) = \Phi_{\text{Б}}(|x|). \end{aligned}$$

Вероятность попадания случайной величины g, распределённой по стандартному нормальному закону в конечный полуинтервал $[a,b)$, где $-\infty < a < b < \infty$, для функции $\Phi_Б$.

$0 \leqslant a < b$. В этом случае

$$\begin{aligned}\mathsf{P}(g \in [a,b)) &= \Phi_{0,1}(b) - \Phi_{0,1}(a) = \\ &= (1 - \Phi_Б(b)) - (1 - \Phi_Б(a)) = \\ &= \Phi_Б(a) - \Phi_Б(b).\end{aligned}$$

$a < 0 \leqslant b$. Теперь

$$\begin{aligned}\mathsf{P}(g \in [a,b)) &= \Phi_{0,1}(b) - \Phi_{0,1}(a) = \\ &= (1 - \Phi_Б(b)) - \Phi_Б(|a|) = \\ &= 1 - \Phi_Б(b) - \Phi_Б(|a|).\end{aligned}$$

$a < b \leqslant 0$. Наконец

$$\mathsf{P}(g \in [a,b)) = \Phi_{0,1}(b) - \Phi_{0,1}(a) = \Phi_Б(|b|) - \Phi_Б(|a|).$$

4. Математическое ожидание

Имеем

$$\begin{aligned}\mathsf{M}\,\Phi_{a,\sigma^2} &= \frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\infty} t \cdot e^{-\frac{(t-a)^2}{2\sigma^2}}dt = \\ &= \frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\infty} (t-a+a)e^{-\frac{(t-a)^2}{2\sigma^2}}dt = \\ &= \frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\infty} (t-a)e^{-\frac{(t-a)^2}{2\sigma^2}}dt + \frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\infty} a \cdot e^{-\frac{(t-a)^2}{2\sigma^2}}dt.\end{aligned}$$

Сделаем в обоих интегралах замену $\frac{t-a}{\sigma} = u$, потому $dt = \sigma\,du$ и получим

$$\mathsf{M}\,\Phi_{a,\sigma^2} = \frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\infty} \sigma u \cdot e^{-\frac{u^2}{2}}du + a \cdot \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{u^2}{2}}du = 0 + a = a,$$

потому что первый интеграл от нечётной функции по симметричному промежутку, а сходимость его обеспечивается тем, что в $\pm\infty$ функция $e^{-\frac{u^2}{2}}$ убывает быстрее любой степени $\frac{1}{u^n}$, $n \geqslant 3$; второй интеграл $\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{u^2}{2}}du = 1$ — основной интеграл теории вероятностей (интеграл Пуассона).

Замечание. Таким образом, можно сказать, что

> *в нормальном распределении с параметрами (a, σ^2) первый параметр a равен его математическому ожиданию.*

5. Дисперсия

Теперь знаем, что $\mathsf{M}\,\Phi_{a,\sigma^2} = a$. Поэтому

$$\mathsf{D}\,\Phi_{a,\sigma^2} = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} (t-a)^2 \cdot e^{-\frac{(t-a)^2}{2\sigma^2}} dt =$$

сделаем замену $\frac{t-a}{\sigma} = u$, потому $dt = \sigma\, du$ и получим

$$= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} \sigma^2 u^2 \cdot e^{-\frac{u^2}{2}} \sigma du = \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} u^2 e^{-\frac{u^2}{2}} du =$$
$$= \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} u d\left(-e^{-\frac{u^2}{2}}\right).$$

Теперь по частям

$$\mathsf{D}\,\Phi_{a,\sigma^2} = \frac{\sigma^2}{\sqrt{2\pi}} \left(-ue^{-\frac{u^2}{2}}\Big|_{-\infty}^{\infty} + \int_{-\infty}^{\infty} e^{-\frac{u^2}{2}} du \right) = \frac{\sigma^2}{\sqrt{2\pi}} \left(-0 + 0 + \sqrt{2\pi}\right) = \sigma^2,$$

поскольку $\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{u^2}{2}} du = 1$ — основной интеграл теории вероятностей (интеграл Пуассона).

Замечания.

1. Таким образом, можно сказать, что

   в нормальном распределении с параметрами (a, σ^2) *второй параметр* σ^2 *равен его дисперсии.*

2. Теперь понятно, что переход к стандартным параметрам есть ничто иное как *нормировка.*

3.6.4. Показательное распределение

1. Плотность и функция распределения

Рассмотрим непрерывное распределение, имеющее плотность

$$p(t) = \begin{cases} 0 & \text{при } t < 0; \\ Ae^{-\lambda t}, & \text{при } t \geqslant 0 \end{cases}$$

для $\lambda > 0$. Сначала найдём неизвестное нам число A, подобные вопросы типичны для задач на случайные величины, чем и вызвано рассмотрение этой задачи. В самом деле, мы должны иметь

$$1 = \int_{-\infty}^{\infty} p(t)dt = \int_0^{\infty} Ae^{-\lambda t} dt = A \cdot \frac{e^{-\lambda t}}{-\lambda}\Big|_0^{\infty} = A\left(\frac{0}{-\lambda} - \frac{1}{-\lambda}\right) = \frac{A}{\lambda}.$$

Откуда

$$A = \lambda \quad \text{и} \quad p(t) = \begin{cases} 0, & t < 0; \\ \lambda e^{-\lambda t}, & t \geqslant 0 \end{cases}$$

Полученное распределение называется *показательным с параметром* λ и обозначается Γ_λ.

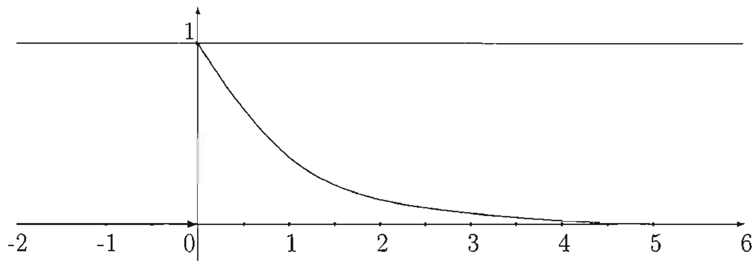


Рис. 13. Плотность показательного распределения

В данном случае функция распределения

при $x < 0$ получаем, что $F(x) = 0$;

а при $x \geqslant 0$ имеем

$$\begin{aligned} F(x) &= \int_{-\infty}^{x} p(t)\,dt = \int_{0}^{x} p(t)\,dt = \int_{-\infty}^{x} \lambda e^{-\lambda t}\,dt = \\ &= \lambda \cdot \left.\frac{e^{-\lambda t}}{-\lambda}\right|_0^x = -e^{-\lambda t}\Big|_0^x = -\left(e^{-\lambda x} - e^0\right) = \\ &= 1 - e^{-\lambda x} \end{aligned}$$

Таким образом получаем функцию распределения

$$F(x) = \begin{cases} 0 & \text{при } x < 0, \\ 1 - e^{-\lambda x} & \text{при } x \geqslant 0. \end{cases}$$

Функцию показательного распределения называют иногда *функцией надёжности,* так как с её помощью описывают вероятность отказа большого числа однородных устройств, к примеру, систем реле в электрических цепях, наборов транзисторов в радиосхемах, интегральных схемах и т.д.

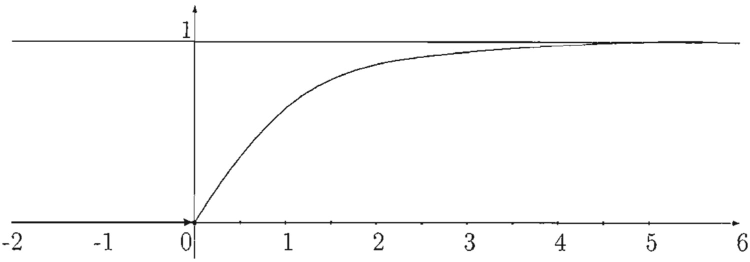


Рис. 14. Функция распределения для показательного распределения

2. Математическое ожидание

Найдём математическое ожидание

$$\mathsf{M}\,\Gamma_\lambda = \int_{\infty}^{\infty} t\cdot p(t)\,dt = \int_0^{\infty} t\cdot\lambda e^{-\lambda t}dt = \lambda\int_0^{\infty} te^{-\lambda t}dt =$$
$$= \lambda\int_0^{\infty} td\left(\frac{e^{-\lambda t}}{-\lambda}\right) = |\text{теперь по частям}| =$$
$$= \lambda\left(t\cdot\frac{e^{-\lambda t}}{-\lambda}\bigg|_0^{\infty} + \int_0^{\infty} 1\cdot\frac{e^{-\lambda t}}{\lambda}dt\right).$$

Так как $\lambda > 0$, то $e^{-\lambda t}$ убывает на ∞ быстрее любой степени $\frac{1}{t^n}$, $n \geqslant 2$, и потому при $t \to \infty$ получаем $t\cdot e^{-\lambda t} \to 0$, итак $t\cdot\frac{e^{-\lambda t}}{-\lambda}\Big|_0^{\infty} = 0$ и

$$\mathsf{M}\,\Gamma_\lambda = \lambda\left(\frac{1}{\lambda}\cdot\frac{e^{-\lambda t}}{-\lambda}\bigg|_0^{\infty}\right) = \frac{1}{-\lambda}(0-1) = \frac{1}{\lambda}.$$

Замечание. Таким образом, можно сказать, что

> *для показательного распределения с параметром λ математическое ожидание — число, обратное параметру.*

3. Дисперсия

Имеем

$$\mathsf{M}\,(\Gamma_\lambda)^2 = \int_{\infty}^{\infty} t^2\cdot p(t)\,dt = \int_0^{\infty} t^2\cdot\lambda e^{-\lambda t}dt = \lambda\int_0^{\infty} t^2d\left(\frac{e^{-\lambda t}}{-\lambda}\right) =$$

теперь по частям

$$= \lambda\left(t^2\cdot\frac{e^{-\lambda t}}{-\lambda}\bigg|_0^{\infty} + \int_0^{\infty} 2t\cdot\frac{e^{-\lambda t}}{\lambda}dt\right).$$

Так как $\lambda > 0$, то $e^{-\lambda t}$ убывает на ∞ быстрее любой степени $\frac{1}{t^n}$, $n \geqslant 3$, и потому при $t \to \infty$ получаем $t^2 \cdot e^{-\lambda t} \to 0$, итак $t^2 \cdot \frac{e^{-\lambda t}}{-\lambda}\Big|_0^\infty = 0$ и

$$\mathsf{M}\,(\Gamma_\lambda)^2 = 2\int_0^\infty t \cdot e^{-\lambda t}dt = \frac{2}{\lambda}\int_0^\infty \lambda t \cdot e^{-\lambda t}dt = \frac{2}{\lambda}\,\mathsf{M}\,\Gamma_\lambda = \frac{2}{\lambda^2} \quad \text{и}$$

$$\mathsf{D}\,\Gamma_\lambda = \mathsf{M}\,(\Gamma_\lambda)^2 - (\mathsf{M}\,\Gamma_\lambda)^2 = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}.$$

Замечание. Таким образом, можно сказать, что

> *для показательного распределения с параметром λ дисперсия — число, обратное квадрату параметра.*

3.6.5. Распределение Коши

1. Плотность и функция распределения

Рассмотрим непрерывное распределение, имеющее плотность

$$p(t) = \frac{A}{1+t^2}.$$

Снова найдём неизвестное нам число A. В самом деле, мы должны иметь

$$1 = \int_{-\infty}^\infty p(t)dt = \int_0^\infty \frac{A}{1+t^2}dt = A \cdot \operatorname{arctg} t\big|_{-\infty}^\infty = A\left(\frac{\pi}{2} - \left(-\frac{\pi}{2}\right)\right) =$$
$$= A \cdot \pi.$$

Откуда

$$A = \frac{1}{\pi} \quad \text{и} \quad p(t) = \frac{1}{\pi} \cdot \frac{1}{1+t^2}.$$

Полученное распределение называется *распределением Коши*, будем обозначать его K.

Из графика плотности распределения Коши (Рис. 14) видно, что этот график похож на график плотности нормального распределения, но в данном случае "плечики более пологие".

В данном случае функция распределения

$$F(x) = \int_{-\infty}^x p(t)\,dt = \frac{1}{\pi}\int_{-\infty}^x \frac{1}{1+t^2}\,dt = \frac{1}{\pi} \cdot \operatorname{arctg} t\Big|_{-\infty}^x =$$
$$= \frac{1}{\pi}\left(\operatorname{arctg} x - \left(-\frac{\pi}{2}\right)\right) = \frac{\operatorname{arctg} x}{\pi} + \frac{\pi}{2}.$$

Функция распределения изображена на рисунке 16.

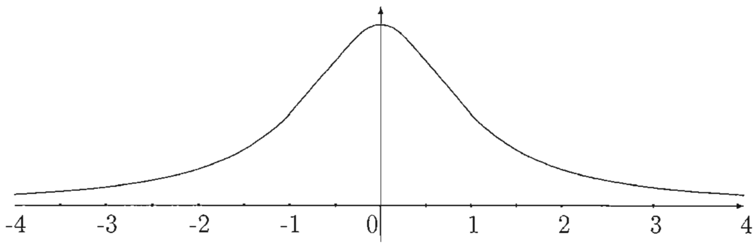


Рис. 15. Плотность распределения Коши

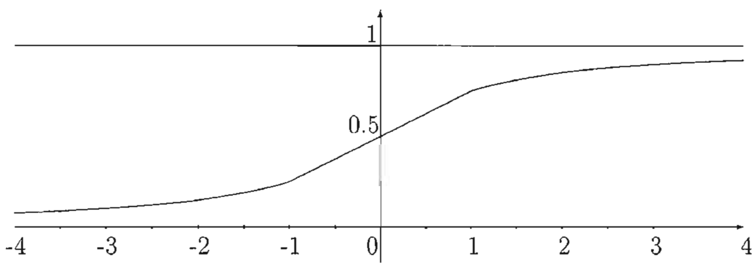


Рис. 16. Функция распределения Коши

2. Математическое ожидание

Найдём математическое ожидание

$$\mathsf{M}\,K = \int_{\infty}^{\infty} t\cdot p(t)\,dt = \int_{-\infty}^{\infty} t\cdot\frac{1}{\pi}\cdot\frac{1}{1+t^2}\,dt = \frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{d(1+t^2)}{1+t^2}\,dt = \\ = \frac{1}{2\pi}\cdot\ln\left(1+t^2\right)\Big|_{-\infty}^{\infty}.$$

Таким образом получаем, что интеграл для математического ожидания распределения Коши расходится и в $-\infty$, и в ∞ (заметим, что здесь из-за расходимости *нельзя* применять рассуждения об интеграле от нечётной функции по симметричному промежутку). Поэтому распределение Коши *не имеет* математического ожидания.

3.6.6. δ-функция Дирака

Закончим рассмотрение одним важным примером, который используется часто в физике.

Пусть $\{f_n\}_{n=1}^{\infty}$ — последовательность плотностей нормальных случайных величин с математическим ожиданием δ и дисперсиями $\frac{1}{n^2}$.

Тогда для любого натурального n

$$f_n(t) = \frac{n}{\sqrt{2\pi}} e^{-\frac{(t-\delta)^2 n^2}{2}} \quad \forall t \in \mathbf{R},$$

и в точке δ

$$f_n(\delta) = \frac{n}{\sqrt{2\pi}} \longrightarrow \infty \quad \text{при } n \to \infty.$$

В пределе возникает функция f со следующими свойствами:

1) $\int_{-\infty}^{\infty} f(t)dt = 1$,

2) $f(t) = 0$ для всех $t \neq \delta$.

Такая функция и называется *δ-функцией Дирака,* она используется при описании различных типов мгновенных процессов — взрывы, удары, выстрелы и т.п.

4. Предельные теоремы

4.1. Неравенство Чебышёва

Этот пункт — вспомогательный, но очень важный, поскольку позволяет оценивать вероятность уклонения случайной величины от её математического ожидания, кроме того, он будет полезен в доказательстве дальнейших результатов.

Теорема (Неравенство Чебышёва).
Пусть g — случайная величина, имеющая математическое ожидание и дисперсию. Пусть $\varepsilon > 0$ — произвольное действительное число. Тогда

$$\mathsf{P}(|g - \mathsf{M}\, g| \geqslant \varepsilon) \leqslant \frac{\mathsf{D}\, g}{\varepsilon^2}, \quad \text{\textit{или равносильно}}$$
$$\mathsf{P}(|g - \mathsf{M}\, g| < \varepsilon) \geqslant 1 - \frac{\mathsf{D}\, g}{\varepsilon^2}.$$

Доказательство. Пусть $\langle \Omega, \mathcal{A}, \mathsf{P} \rangle$ — вероятностное пространство, на котором задана случайная величина g. Определим случайную величину h_ε следующим образом:

$$h_\varepsilon(\omega) = \begin{cases} \varepsilon^2, & \text{если } (g(\omega) - \mathsf{M}\, g)^2 \geqslant \varepsilon^2, \\ 0, & \text{если } (g(\omega) - \mathsf{M}\, g)^2 < \varepsilon^2. \end{cases}$$

Тогда для всех $\omega \in \Omega$

$$(g(\omega) - \mathsf{M}\, g)^2 \geqslant h_\varepsilon(\omega).$$

Теперь по теореме о свойствах математического ожидания

$$\begin{aligned} \mathsf{D}\, g &= \mathsf{M}(g(\omega) - \mathsf{M}\, g)^2 \geqslant \mathsf{M}\, h_\varepsilon = \\ &= \varepsilon^2 \cdot \mathsf{P}\left((g(\omega) - \mathsf{M}\, g)^2 \geqslant \varepsilon^2\right) + 0 \cdot \mathsf{P}\left((g(\omega) - \mathsf{M}\, g)^2 < \varepsilon^2\right) = \\ &= \varepsilon^2 \cdot \mathsf{P}\left(|g(\omega) - \mathsf{M}\, g| \geqslant \varepsilon\right). \end{aligned}$$

Таким образом

$$\mathsf{D}\, g \geqslant \varepsilon^2 \cdot \mathsf{P}\left(|g(\omega) - \mathsf{M}\, g| \geqslant \varepsilon\right),$$

что и надо. □

Следствие (Правило 3σ).
Пусть g — случайная величина, имеющая математическое ожидание и дисперсию σ^2, $\sigma > 0$. Тогда выполняются следующие условия.

1. $\mathsf{P}(|g - \mathsf{M}\, g| < 3\sigma) \geqslant \frac{8}{9} = 0.88888\ldots$.

2 Если $g = \Phi_{a,\sigma^2}$, то в приведённых ранее обозначениях

$$\mathsf{P}(|g - \mathsf{M}\,g| < 3\sigma) = \frac{1}{\sqrt{2\pi}}\int_{-3}^{3} e^{-\frac{t^2}{2}}dt = \Phi_{0,1}(3) - \Phi_{0,1}(-3) = \\ = 2\Phi_{\text{Гм}}(3) = 2\cdot 0.49865\ldots = 0.9973\ldots.$$

Доказательство.

1. По неравенству Чебышёва

$$\mathsf{P}(|g - \mathsf{M}\,g| < 3\sigma) \geqslant 1 - \frac{\mathsf{D}\,g}{(3\sigma)^2} = 1 - \frac{\sigma^2}{9\sigma^2} = \frac{8}{9}.$$

2. В данном случае, используя переход к стандартным параметрам для нормального распределения, имеем

$$\mathsf{P}(|g - \mathsf{M}\,g| < 3\sigma) = \mathsf{P}\left(\frac{|g-a|}{\sigma} < 3\right) = \mathsf{P}\left(\frac{|g-a|}{\sigma} \in (-3,3)\right) = \\ = \Phi_{0,1}(3) - \Phi_{0,1}(-3).$$

□

4.2. Законы больших чисел (ЗБЧ)

Теорема (Законы больших чисел).

1. ***Закон больших чисел в форме Чебышёва для произвольно распределённых случайных величин***

Пусть $\{g_n\}_{n=1}^{\infty}$ — последовательность независимых случайных величин на одном вероятностном пространстве, которые имеют математическое ожидание и ограниченную в совокупности дисперсию, то есть существует такое действительное число C, не зависящее от n, что $\mathsf{D}\,g_n \leqslant C$ для всех натуральных n. Тогда для любого $\varepsilon > 0$

$$\lim_{n\to\infty} \mathsf{P}\left(\left|\frac{\sum_{k=1}^{n} g_k}{n} - \frac{\sum_{k=1}^{n} \mathsf{M}\,g_k}{n}\right| < \varepsilon\right) = 1.$$

2. ***Закон больших чисел в форме Чебышёва для одинаково распределённых случайных величин***

Пусть $\{g_n\}_{n=1}^{\infty}$ — последовательность независимых одинаково распределённых случайных величин на одном вероятностном пространстве, имеющих математическое ожидание, равное a, и дисперсию. Тогда для любого $\varepsilon > 0$

$$\lim_{n\to\infty} \mathsf{P}\left(\left|\frac{\sum_{k=1}^{n} g_k}{n} - a\right| < \varepsilon\right) = 1.$$

3. ***Закон больших чисел для биномиальных распределений с одинаковой вероятностью успеха (Закон Бернулли)***
Пусть $\{\mathbf{B}_n(p)\}_{n=1}^{\infty}$ *— последовательность случайных величин с биномиальным распределением с одинаковой вероятностью успеха, равной* p*. Тогда для любого* $\varepsilon > 0$

$$\lim_{n\to\infty} \mathsf{P}\left(\left|\frac{\mathbf{B}_n(p)}{n} - p\right| < \varepsilon\right) = 1.$$

Доказательство. 1. Для любого натурального n положим

$$G_n = \frac{\sum_{k=1}^n g_k}{n}.$$

По свойствам математического ожидания и дисперсии для каждого натурального n имееи

$$\mathsf{M}\, G_n = \frac{\sum_{k=1}^n \mathsf{M}\, g_k}{n} \quad \text{и}$$

$$\mathsf{D}\, G_n = \mathsf{D}\, \frac{\sum_{k=1}^n g_k}{n} = \frac{1}{n^2}\sum_{k=1}^n \mathsf{D}\, g_k \leqslant \frac{nC}{n^2} = \frac{C}{n}.$$

По неравенству Чебышёва

$$1 \geqslant \mathsf{P}\left(\left|\frac{\sum_{k=1}^n g_k}{n} - \frac{\sum_{k=1}^n \mathsf{M}\, g_k}{n}\right| < \varepsilon\right) = \mathsf{P}\left(|G_n - \mathsf{M}\, G_n| < \varepsilon\right) \geqslant$$
$$\geqslant 1 - \frac{\mathsf{D}\, G_n}{\varepsilon^2} \geqslant 1 - \frac{C}{n\varepsilon^2} \longrightarrow 1 \quad \text{при} \quad n \to \infty.$$

Всё доказано, по теореме “о двух милиционерах (о зажатой переменной)”.

2. В этом случае для каждого натурального n имеем

$$\mathsf{M}\, G_n = \frac{\sum_{k=1}^n \mathsf{M}\, g_k}{n} = \frac{na}{n} = a,$$

и всё следует из утверждения 1.

3. Пусть $\{g_k\}_{k=1}^{\infty}$ — последовательность независимых случайных величин на одном вероятностном пространстве, одинаково распределённых по схеме Бернулли, то есть для каждого натурального k имеем $\mathsf{P}(g_k = 1) = p$ и $\mathsf{P}(g_k = 0) = q = 1 - p$. Тогда ясно, что

$$\mathbf{B}_n(p) = g_1 + \cdots + g_n,$$

и всё следует из утверждения 2.

□

К ЗБЧ

Здесь будут приведены без доказательства более глубокие факты, связанные с законами больших чисел.

1. **Закон больших чисел в форме Хинчина для одинаково распределённых случайных величин.**

   Пусть $\{g_n\}_{n=1}^{\infty}$ *— последовательность независимых одинаково распределённых случайных величин на одном вероятностном пространстве, имеющих математическое ожидание, равное* a.

   Тогда для любого $\varepsilon > 0$

$$\lim_{n\to\infty} \mathsf{P}\left(\left|\frac{\sum_{k=1}^{n} g_k}{n} - a\right| < \varepsilon\right) = 1.$$

   Таким образом, теорема Хинчина убирает требование наличия дисперсии в утверждении 2) теоремы (закон больших чисел в форме Чебышёва для одинаково распределённых случайных величин).

2. **Усиленный закон больших чисел в форме Колмогорова для одинаково распределённых случайных величин.**

   Пусть $\{g_n\}_{n=1}^{\infty}$ *— последовательность независимых одинаково распределённых случайных величин на одном вероятностном пространстве.*

   Тогда для любого $\varepsilon > 0$

$$\lim_{n\to\infty} \mathsf{P}\left(\left|\frac{\sum_{k=1}^{n} g_k}{n} - a\right| < \varepsilon\right) = 1$$

   тогда и только тогда, когда для любого натурального k

$$\mathsf{M}\, g_k = a.$$

   Таким образом, теорема Колмогорова показывает, что среднее арифметическое одинаково распределённых случайных величин может сходится *только* к их математическому ожиданию.

3. **Закон 0 и 1 Колмогорова.**

   Прежде чем формулировать упомянутое утверждение, нам необходимо ввести ряд понятий.

   Пусть $\{g_n\}_{n=1}^{\infty}$ — последовательность независимых одинаково распределённых случайных величин на одном вероятностном пространстве $\langle\Omega, \mathcal{A}, \mathsf{P}\rangle$. Для каждого натурального n обозначим через

$$\mathcal{A}_n = \sigma(g_n, g_{n+1}, \dots)$$

   алгебру событий, порождённую всеми событиями $\{g_k < x\}$ для всех натуральных $k \geqslant n$ и действительных x. Порождение событиями означает пересечение всех алгебр событий, которые содержат данные события. Одна такая алгебра событий всегда есть — это сама $\mathfrak{A}$ на Ω, поэтому порождение — *наименьшая* алгебра событий, содержащая данные события.

Ясно, что

$$\mathfrak{A}_1 = \sigma(g_1, g_2, \dots) \supseteq \dots \supseteq \mathfrak{A}_n = \sigma(g_n, g_{n+1}, \dots) \supseteq \cdots,$$

и получается убывающая цепочка вложенных алгебр событий.

Событие называется *остаточным*, если оно принадлежит

$$\mathfrak{A}_n = \sigma(g_n, g_{n+1}, \dots) \quad \text{для всех натуральных } n.$$

Теперь заявленный выше результат.

Вероятность остаточного события равна 0 *или* 1.

Отметим, что доказательство состоит в том, что остаточное событие A оказывается независимым от себя самого, и потому

$$\mathsf{P}(A) = \mathsf{P}(AA) = \mathsf{P}(A)\,\mathsf{P}(A).$$

Откуда $\mathsf{P}(A) = 0$ или $\mathsf{P}(A) = 1$.

4.3. Предельная теорема для гипергеометрического распределения

Напомним задание гипергеометрического распределения. Пусть k, k_1 и s — натуральные числа, причём $k \geqslant k_1$ и $k \geqslant s$. Положим $k_2 = k - k_1$, и для любого неотрицательного целого $s_1 \leqslant s$ пусть $s_2 = s - s_1$. Тогда гипергеометрическое распределение задаётся так:

$$\left\{ P_{k_1,k}(s_1, s) = \frac{C_{k_1}^{s_1} C_{k_2}^{s_2}}{C_k^s} \,\middle|\, \text{для всех возможных } s_1 \right\}.$$

Замечали ранее, что s_1 может *не* принимать все значения от 0 до k_1.

Теорема. *Предположим, что*

$$\lim_{k\to\infty} \frac{k_1}{k} = p.$$

Тогда для каждого допустимого значения s_1

$$\lim_{k\to\infty} P_{k_1,k}(s_1, s) = C_s^{s_1} p^{s_1} q^{s-s_1} \quad (q = 1 - p).$$

Доказательство. В самом деле

$$P_{k_1,k}(s_1,s) = \frac{s!(k-s)!}{k!} \cdot \frac{k_1!}{s_1!(k_1-s_1)!} \cdot \frac{k_2!}{s_2!(k_2-s_2)!} =$$
$$= \frac{s!}{s_1!s_2!} \cdot \frac{(k-s)!}{k!} \cdot \frac{k_1!}{(k_1-s_1)!} \cdot \frac{k_2!}{(k_2-s_2)!}.$$

С учётом того, что $s_2 = s - s_1$

$$P_{k_1,k}(s_1,s) = C_s^{s_1} \cdot \frac{1}{k\cdots(k-s+1)} \cdot k_1\cdots(k_1-s_1+1)\times$$
$$\times k_2\cdots(k_2-s_2+1) =$$
$$= C_s^{s_1} \cdot \frac{(k_1\cdots(k_1-s_1+1))\cdot(k_2\cdots(k_2-s_2+1))}{k\cdots(k-s+1)}.$$

Поделим каждый сомножитель числителя и знаменателя на k (сверху и снизу по $s_2+s_1=s$ сомножителей, поэтому значение не изменится) и получим

$$P_{k_1,k}(s_1,s) = C_s^{s_1} \cdot \frac{\left(\frac{k_1}{k}\cdots\left(\frac{k_1}{k}-\frac{s_1-1}{k}\right)\right)\cdot\left(\frac{k_2}{k}\cdots\left(\frac{k_2}{k}-\frac{s_2-1}{k}\right)\right)}{\left(\frac{k_1}{k}\cdots\left(\frac{k}{k}-\frac{s-1}{k}\right)\right)}.$$

Теперь

$$\lim_{k\to\infty}\left(\frac{k_1}{k}-\frac{i}{k}\right) = p - 0 = p \quad \text{для любого } i\in\{0,\dots,s_1\},$$
$$\lim_{k\to\infty}\left(\frac{k_2}{k}-\frac{j}{k}\right) = \lim_{k\to\infty}\left(\frac{k-k_1}{k}-\frac{j}{k}\right) = \lim_{k\to\infty}\left(1-\frac{k_1}{k}-\frac{j}{k}\right) =$$
$$= 1-p-0 = q \quad \text{для любого } j\in\{0,\dots,s_2\} \quad \text{и}$$
$$\lim_{k\to\infty}\left(\frac{k}{k}-\frac{l}{k}\right) = 1-0 = 1 \quad \text{для любого } l\in\{0,\dots,s\}.$$

Следовательно

$$\lim_{k\to\infty} P_{k_1,k}(s_1,s) = C_s^{s_1}p^{s_1}q^{s-s_1},$$

что и было нужно. □

Сходимость гипергеометрического распределения к биномиальному

Рассмотрим гипергеометрическое распределение с $k = 2k_1 \geqslant 50$, $s = 50$ и $s_1 = 25$. Тогда

$$P_{k_1} = P_{k_1,2k_1}(25,50) = \frac{C_{k_1}^{25}\cdot C_{k_1}^{25}}{C_{2k_1}^{50}} = \frac{\left(C_{k_1}^{25}\right)^2}{C_{2k_1}^{50}} =$$

$$
\begin{aligned}
&= \frac{(k_1!)^2}{(25!)^2 \cdot ((k_1-25)!)^2} \cdot \frac{50! \cdot (2k_1-50)!}{(2k_1)!} = \\
&= \frac{50!}{(25!)^2} \cdot \frac{(k_1!)^2 \cdot (2k_1-50)!}{((k_1-25)!)^2 \cdot (2k_1)!} = \\
&= C_{50}^{25} \cdot \left(\frac{k_1!}{(k_1-25)!}\right)^2 \cdot \frac{(2k_1-50)!}{(2k_1)!} = C_{50}^{25} \cdot \frac{(a_{k_1})^2}{b_{k_1}} = \\
&= C_{50}^{25} \cdot \frac{c_{k_1}}{d_{k_1}} = C_{50}^{25} \cdot f_{k_1},
\end{aligned}
$$

где для любого натурального $n \geqslant 25$

$$
a_n = \frac{n!}{(n-25)!} \quad \text{и} \quad b_n = \frac{(2n)!}{(2n-50)!},
$$
$$
c_n = (n!)^2 \cdot (2n-50)! \quad \text{и} \quad d_n = (2n)! \cdot ((n-25)!)^2,
$$
$$
f_n = \frac{c_n}{d_n}.
$$

С помощью GAP

$$
C_{50}^{25} = 126410606437752 = 2^3 \cdot 3^2 \cdot 7^2 \cdot 13 \cdot 29 \cdot 31 \cdot 37 \cdot 41 \cdot 43 \cdot 47.
$$

Теперь применим предельную теорему для гипергеометрического распределения

$$
P_{k_1} = P_{k_1,2k_1}(25,50) \sim C_{50}^{25} \left(\frac{1}{2}\right)^{25} \cdot \left(\frac{1}{2}\right)^{50-25} = C_{50}^{25} \cdot \frac{1}{2^{50}} = \frac{C_{50}^{25}}{2^{50}}.
$$

С помощью GAP и калькулятора.

Калькулятор

$$
\frac{C_{50}^{25}}{2^{50}} = 0,11227517265921704847642104141414.
$$

GAP

$$
\begin{aligned}
\frac{C_{50}^{25}}{2^{50}} &= \frac{2^3 \cdot 3^2 \cdot 7^2 \cdot 13 \cdot 29 \cdot 31 \cdot 37 \cdot 41 \cdot 43 \cdot 47}{2^{50}} = \\
&= \frac{3^2 \cdot 7^2 \cdot 13 \cdot 29 \cdot 31 \cdot 37 \cdot 41 \cdot 43 \cdot 47}{2^{47}} = \\
&= \frac{15801325804719}{140737488355328} = \\
&= 0,11227517265921704847642104141414.
\end{aligned}
$$

$$k_1 = 25.$$

В этом случае ясно, что

$$P_{25} = 1.$$

$$k_1 = 50.$$

С помощью GAP и калькулятора.

Калькулятор

$$\begin{aligned}
a_{50} &= \frac{50!}{25!} = 1,96078146816081941570317208046 8 \cdot 10^{39},\\
(a_{50})^2 &= 3,8446639658828984840365988928603 \cdot 10^{78},\\
b_{50} &= 3,0685187562549660372027304595295 \cdot 10^{93},\\
P_{50} &= 0,1583846611613304257693501737109 1.
\end{aligned}$$

GAP

$$\begin{aligned}
a_{50} &= 1960781468160819415703172080467968000000 =\\
&= 2^{25} \cdot 3^{12} \cdot 5^{6} \cdot 7^{5} \cdot 11^{2} \cdot 13^{2} \cdot 17 \cdot 19 \cdot 23 \cdot 29 \cdot 31 \times\\
&\quad \times 37 \cdot 41 \cdot 43 \cdot 47,\\
(a_{50})^2 &= 3844663965882898484036598892860302514778 88\\
&\quad 205474709046408587404902400000000000000,\\
b_{50} &= 3068518756254966037202730459529469739228459\\
&\quad 7216846889594477869869821589587723550720000\\
&\quad 00000000 =\\
&= 2^{50} \cdot 3^{26} \cdot 5^{12} \cdot 7^{8} \cdot 11^{5} \cdot 13^{4} \cdot 17^{3} \cdot 19^{3} \cdot 23^{2} \cdot 29^{2} \cdot 31^{2} \times\\
&\quad \times 37 \cdot 41 \cdot 43 \cdot 47 \cdot 53 \cdot 59 \cdot 61 \cdot 67 \cdot 71 \cdot 73 \cdot 79 \cdot 83 \times\\
&\quad \times 89 \cdot 97,\\
f_{50} &= \frac{(a_{50})^2}{b_{50}} = \frac{150226993}{119899775326768814036979} =\\
&= \frac{7^2 \cdot 37 \cdot 41 \cdot 43 \cdot 47}{3^2 \cdot 11 \cdot 17 \cdot 19 \cdot 53 \cdot 59 \cdot 61 \cdot 67 \cdot 71 \cdot 73 \cdot 79 \cdot 83 \cdot 89 \cdot 97} =\\
&= 1,2529380692380691544524905229157 \cdot 10^{-15},\\
P_{50} &= \frac{21100316987166582933 04}{13322197258529868226331} =\\
&= \frac{2^3 \cdot 7^4 \cdot 13 \cdot 29 \cdot 31 \cdot 37^2 \cdot 41^2 \cdot 43^2 \cdot 47^2}{11 \cdot 17 \cdot 19 \cdot 53 \cdot 59 \cdot 61 \cdot 67 \cdot 71 \cdot 73 \cdot 79 \cdot 83 \cdot 89 \cdot 97} =\\
&= 0,1583846611613304257693501737109 1.
\end{aligned}$$

$$k_1 = 100.$$

Так как калькулятор даёт приемлемые результаты и GAP будет выдавать увеличении k_1 необозримые данные, будем делать необходимые вычисления здесь и далее с помощью калькулятора, вычисляя только c_{k_1}, d_{k_1} и f_{k_1}.

$$\begin{aligned}
c_{100} &= (100!)^2 \cdot 150! = 4,9762331537296877038953922768903 \cdot 10^{578},\\
d_{100} &= (75!)^2 \cdot 200! = 4,8541376569182715366023348479905 \cdot 10^{593},\\
f_{100} &= \frac{c_{100}}{d_{100}} = 1,0251528706931913597284396344516 \cdot 10^{-15},\\
P_{100} &= 0,12959019607572867930892195372929.
\end{aligned}$$

$$k_1 = 200.$$

Получаем

$$\begin{aligned}
c_{200} &= (200!)^2 \cdot 350! = 7,6869046905572420752626426856337 \cdot 10^{1489},\\
d_{200} &= (175!)^2 \cdot 400! = 8,0964396338857707132520116107856 \cdot 10^{1504},\\
f_{200} &= \frac{c_{200}}{d_{200}} = 9,4941789702049835655763123099093 \cdot 10^{-16},\\
P_{200} &= 0,12001649212521637493097143441391.
\end{aligned}$$

$$k_1 = 400.$$

Получаем

$$\begin{aligned}
c_{400} &= (400!)^2 \cdot 750! = 1,0582378541770389910385173461899 \cdot 10^{3570},\\
d_{400} &= (375!)^2 \cdot 800! = 1,1536598057504167828958660318687 \cdot 10^{3585},\\
f_{400} &= \frac{c_{400}}{d_{400}} = 9,1728761711403386235996070828908 \cdot 10^{-16},\\
P_{400} &= 0,11595488395722548061234577458681.
\end{aligned}$$

$$k_1 = 800.$$

Получаем

$$\begin{aligned}
c_{800} &= (800!)^2 \cdot 1550! = 4,2279790007012244086075803129864 \cdot 10^{8227},\\
d_{800} &= (775!)^2 \cdot 1600! = 4,6853350355539437919471889921054 \cdot 10^{8242},\\
f_{800} &= \frac{c_{800}}{d_{800}} = 9,0238562848075036148792850483703 \cdot 10^{-16},\\
P_{800} &= 0,11407111453696362616911353622319.
\end{aligned}$$

$$k_1 = 1600.$$

В этом случае уже не удаётся просчитать из-за переполнения, хотя если действовать более разумно, видимо, можно.

4.4. Теорема Пуассона (Закон редких событий)

Теорема. *Пусть* $\{\mathbf{B}_n(p_n)\}_{n=1}^{\infty}$ *— последовательность случайных величин с биномиальными законами распределения. Предположим, что*

$$\lim_{n\to\infty} np_n = \lambda > 0.$$

Тогда для любого $k \in \{0, \dots, n\}$

$$\lim_{n\to\infty} \mathsf{P}\left(\mathbf{B}_n(p_n) = k\right) = e^{-\lambda}\frac{\lambda^k}{k!} = \mathsf{P}(\Pi_\lambda = k) \text{ (распределение Пуассона).}$$

Замечание. Так как $np_n \to \lambda > 0$ при $n \to \infty$, то вероятности успеха $p_n \to 0$ при $n \to \infty$. Поэтому теорему Пуассона и называют *законом редких событий.*

Доказательство. В самом деле

$$\mathsf{P}\left(\mathbf{B}_n(p_n) = k\right) = C_n^k p_n^k q_n^{n-k} = \frac{n!}{k!(n-k)!} \cdot p_n^k q_n^{n-k} =$$
$$= \frac{1}{k!} \cdot (np_n)^k \cdot \frac{n!}{n^k(n-k)!} \cdot q_n^{-k} \cdot q_n^n.$$

Рассмотрим сомножители по отдельности. Так как

$$\lim_{n\to\infty} np_n = \lambda, \quad \text{то} \quad \lim_{n\to\infty} (np_n)^k = \lambda^k.$$

Далее

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

Поэтому

$$\lim_{n\to\infty} \frac{n!}{n^k(n-k)!} = 1.$$

Так как

$$\lim_{n\to\infty} p_n = 0, \quad \text{то} \quad \lim_{n\to\infty} (q_n) = \lim_{n\to\infty} (1-p_n) = 1 - 0 = 1.$$

Отсюда

$$\lim_{n\to\infty} (q_n)^{-k} = 1.$$

Теперь

$$\lim_{n\to\infty} (q_n)^n = \lim_{n\to\infty} (1-p_n)^n = \lim_{n\to\infty} \left((1-p_n)^{-\frac{1}{p_n}}\right)^{-np_n}.$$

Так как $p_n \to 0$ при $n \to \infty$, можно применить второй замечательный предел и то, что $np_n \to \lambda > 0$ при $n \to \infty$, нам даёт

$$\lim_{n\to\infty} (q_n)^n = \lim_{n\to\infty} \left((1-p_n)^{-\frac{1}{p_n}}\right)^{-np_n} = e^{-\lambda}.$$

Таким образом получаем

$$\lim_{n\to\infty} \mathsf{P}\left(\mathbf{B}_n(p_n) = k\right) = \frac{1}{k!} \cdot \lambda^k \cdot 1 \cdot 1 \cdot e^{-\lambda} = e^{-\lambda}\frac{\lambda^k}{k!} = \mathsf{P}(\Pi_\lambda = k).$$

□

4.5. Локальная теорема Муавра–Лапласа

Обозначение. Положим

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

— это плотность стандартного нормального распределения $\Phi_{0,1}$.

Определения. Пусть $\{a_n\}_{n=1}^\infty$ и $\{b_n\}_{n=1}^\infty$ — две числовые последовательности. Напомним некоторые важные понятия о связях последовательностей (аналогичные понятия можно ввести для функций).

Эквивалентность. Последовательности $\{a_n\}_{n=1}^\infty$ и $\{b_n\}_{n=1}^\infty$ будем называть *эквивалентными при* $n \to \infty$ и обозначать $\{a_n\}_{n=1}^\infty \sim \{b_n\}_{n=1}^\infty$ при $n \to \infty$, если

$$\lim_{n\to\infty} \frac{a_n}{b_n} = 1.$$

o **(о малое).** Будем писать $a_n = o(b_n)$ *при* $n \to \infty$, если

$$\lim_{n\to\infty} \frac{a_n}{b_n} = 0.$$

Отметим, что иногда удобно пользоваться равносильной формой

$$a_n = o(b_n) \text{ при } n \to \infty \quad \longleftrightarrow \quad a_n = \alpha_n b_n, \text{ где } \lim_{n\to\infty} \alpha_n = 0.$$

Кроме того нам потребуется следующий важный результат, который приведём без доказательства.

Теорема (Формула Стирлинга).
Для любого натурального m факториал

$$m! = \sqrt{2\pi m}\, m^m e^{-m+\theta(m)}, \quad \text{где} \quad \frac{1}{12m+1} < \theta(m) < \frac{1}{12m}.$$

В частности получаем, что

$$m! \sim \sqrt{2\pi m}\, m^m e^{-m} = \sqrt{2\pi m} \left(\frac{m}{e}\right)^m \quad \text{при } m \to \infty.$$

Примеры на формулу Стирлинга

Прежде всего отметим, что формула Стирлинга является первым приближением при разложении факториала в ряд Стирлинга:

$$n! = \sqrt{2\pi n}\left(\frac{n}{e}\right)^n \left(1 + \frac{1}{12n} + \frac{1}{288n^2} - \frac{139}{51840n^3} - \frac{571}{2488320n^4} + \right.$$
$$\left. + \frac{163879}{209018880n^5} + \frac{5246819}{75246796800n^6} + O\left(n^{-7}\right)\right).$$

Теперь сами примеры на формулу Стирлинга.

$$n = 25.$$

Точно

$$25! = 15511210043330985984000000.$$

Приближённо по формуле Стирлинга

$$\sqrt{2\pi \cdot 25} = 12,533141373155002512078826424055.$$
$$\frac{25}{e} = 9,1969860292860580398880942540365.$$
$$\left(\frac{25}{e}\right)^{25} = 1233497203486779322881773,5014934.$$
$$25! \approx 15459594834691148955313139,186652.$$

Отношение точного к приближённому и оценки

$$e^{\theta(25)} = \frac{15511210043330985984000000}{15459594834691148955313139,186652} =$$
$$= 1,0033387167769761409048818350452.$$
$$\theta(25) = 0,0033331556367280928758070191173б.$$
$$\frac{1}{12 \cdot 25} = 0,0033333333333333333333333333333333.$$
$$\frac{1}{12 \cdot 25 + 1} = 0,0033222591362126245847176079734 2.$$
$$\theta(25) \in \left(\frac{1}{12 \cdot 25 + 1}, \frac{1}{12 \cdot 25}\right).$$

$$n = 50.$$

Точно

$$50! = 30414093201713378043612608166064768844377641568960512 \cdot 10^{12} =$$
$$= 3,0414093201713378043612608166065 \cdot 10^{64}.$$

Приближённо по формуле Стирлинга

$$\sqrt{2\pi \cdot 50} = 17,724538509055160272981674833411.$$
$$\frac{50}{e} = 18,393972058572116079776188508073.$$
$$\left(\frac{50}{e}\right)^{50} = 1,7130739919614493009956948701662 \cdot 10^{63}.$$
$$50! \approx 3,0363445939381558207983726752111 \cdot 10^{64}.$$

Отношение точного к приближённому и оценки

$$e^{\theta(50)} = \frac{3,0414093201713378043612608166065 \cdot 10^{64}}{3,0363445939381558207983726752111 \cdot 10^{64}} =$$
$$= 1,0016680340707353658689745934915.$$
$$\theta(50) = 0,0016666444469833655099493249910 7.$$
$$\frac{1}{12 \cdot 50} = 0,0016666666666666666666666666666667.$$
$$\frac{1}{12 \cdot 50 + 1} = 0,0016638935108153078202995008319 5.$$
$$\theta(50) \in \left(\frac{1}{12 \cdot 50 + 1}, \frac{1}{12 \cdot 50}\right).$$

$$n = 100.$$

Точно

$$100! = 93326215443944152681699238856266700490715968264381621468592963$$
$$8952175999932299156089414639761565182862536979208272237582511$$
$$85210916864 \cdot 10^{24} =$$
$$= 9,3326215443944152681699238856267 \cdot 10^{157}.$$

Приближённо по формуле Стирлинга

$$\sqrt{2\pi \cdot 100} = 25,066282746310005024157652848111.$$

$$\frac{100}{e} = 36,787944117144232159552377016146.$$
$$\left(\frac{100}{e}\right)^{100} = 3,7200759760208359629596958038631 \cdot 10^{156}.$$
$$100! \approx 9,3248476252693432477647561271785 \cdot 10^{157}.$$

Отношение точного к приближённому и оценки

$$e^{\theta(100)} = \frac{9,3326215443944152681699238856267 \cdot 10^{157}}{9,3248476252693432477647561271785 \cdot 10^{157}} =$$
$$= 1,0008336778720121418498278568362.$$
$$\theta(100) = 8,3330555634914683381241692836336 \cdot 10^{-4}.$$
$$\frac{1}{12 \cdot 100} = 8,3333333333333333333333333333333 \cdot 10^{-4}.$$
$$\frac{1}{12 \cdot 100 + 1} = 8,3263946711074104912572855953372 \cdot 10^{-4}.$$
$$\theta(100) \in \left(\frac{1}{12 \cdot 100 + 1}, \frac{1}{12 \cdot 100}\right).$$

Лемма (о формуле Тейлора).
Пусть $p \in (0,1)$ и $q = 1 - p$. Положим для любого $t \in (0,1)$

$$H(t) = t \ln \frac{t}{p} + (1-t) \ln \frac{1-t}{1-p}.$$

Тогда в окрестности точки p

$$H(t) = \frac{1}{2pq}(t-p)^2 + o(|t-p|^2).$$

Доказательство. Хотим применить формулу Тейлора для $H(t)$ в окрестности точки p. Для этого посчитаем производные

$$H'(t) = \ln \frac{t}{p} + t \cdot \frac{1}{\frac{t}{p}} \cdot \frac{1}{p} + (-1) \ln \frac{1-t}{1-p} + (1-t) \cdot \frac{1}{\frac{1-t}{1-p}} \cdot \frac{-1}{1-p} =$$
$$= \ln \frac{t}{p} + 1 - \ln \frac{1-t}{1-p} - 1 = \ln \frac{t}{p} - \ln \frac{1-t}{1-p},$$
$$H''(t) = \frac{1}{\frac{t}{p}} \cdot \frac{1}{p} - \frac{1}{\frac{1-t}{1-p}} \cdot \frac{-1}{1-p} = \frac{1}{t} + \frac{1}{1-t} = \frac{1}{t(1-t)}.$$

Найдём значения функции H и её производных в точке p

$$H(p) = p \ln 1 + (1-p) \ln 1 = 0 + 0,$$

$$H'(p) = \ln 1 - \ln 1 = 0,$$
$$H''(p) = \frac{1}{p(1-p)} = \frac{1}{pq}.$$

Таким образом по формуле Тейлора для $H(t)$ с остаточным членом в форме Пеано в окрестности точки p получаем

$$H(t) = 0 + 0 \cdot (t-p) + \frac{1}{pq} \cdot \frac{(t-p)^2}{2!} + o((t-p)^2) = \frac{1}{2pq}(t-p)^2 + o((t-p)^2).$$

□

Замечание. Отметим, что при больших n вероятности $\mathsf{P}(\mathbf{B}_n(p) = 0) = q^n = (1-p)^n$ и $\mathsf{P}(\mathbf{B}_n(p) = n) = p^n$ либо тривиальны при $p \in \{0, 1\}$, либо должны рассматриваться отдельно и, как правило, очень малы.

Лемма (о приближённом значении).
Пусть n — натуральное число и $k \in \{1, \ldots, n-1\}$. Положим

$$p^* = \frac{k}{n}.$$

Пусть $p \in (0,1)$ и $q = 1 - p$. Тогда выполняются следующие утверждения.

1. *Биномиальный коэффициент*

$$C_n^k = \frac{1}{\sqrt{2\pi}} \cdot \frac{1}{\sqrt{np^*(1-p^*)}} \cdot e^{-n(p^* \ln p^* + (1-p^*)\ln(1-p^*))} \cdot e^{\theta(k,n)},$$

где для $\theta(n)$, $\theta(k)$, $\theta(n-k)$ из формулы Стирлинга

$$\theta(k,n) = \theta(n) - \theta(k) - \theta(n-k) \quad \text{и}$$
$$\frac{1}{12n+1} - \frac{1}{12k} - \frac{1}{12(n-k)} < \theta(k,n) <$$
$$< \frac{1}{12n} - \frac{1}{12k+1} - \frac{1}{12(n-k)+1}.$$

2. *Значение*

$$\mathsf{P}(\mathbf{B}_n(p) = k) = C_n^k p^k q^{n-k} =$$
$$= \frac{1}{\sqrt{2\pi}} \cdot \frac{1}{\sqrt{np^*(1-p^*)}} \cdot e^{-n\left(p^* \ln \frac{p^*}{p} + (1-p^*)\ln \frac{1-p^*}{1-p}\right)} \cdot e^{\theta(k,n)} =$$

где $\theta(k,n)$ из 1

$$= \frac{1}{\sqrt{2\pi}} \cdot \frac{1}{\sqrt{np^*(1-p^*)}} \cdot e^{-nH(p^*)} \cdot e^{\theta(k,n)},$$

где функция H из леммы о формуле Тейлора.

3. В частности при $k, n-k \to \infty$

$$\mathsf{P}(\mathrm{B}_n(p) = k) = C_n^k p^k q^{n-k} \sim \frac{1}{\sqrt{2\pi}} \cdot \frac{1}{\sqrt{np^*(1-p^*)}} \cdot e^{-nH(p^*)}$$

Доказательство.

1. Применим формулу Стирлинга

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

$$= \frac{\sqrt{2\pi n}\, n^n e^{-n+\theta(n)}}{\sqrt{2\pi k}\, k^k e^{-k+\theta(k)} \sqrt{2\pi(n-k)}\, (n-k)^{n-k} e^{-(n-k)+\theta(n-k)}} =$$

$$= \frac{1}{\sqrt{2\pi}} \cdot \frac{\sqrt{n}}{\sqrt{k(n-k)}} \cdot \frac{n^n}{k^k (n-k)^{n-k}} \cdot 1 \cdot e^{\theta(n)-\theta(k)-\theta(n-k)} =$$

$$= \frac{1}{\sqrt{2\pi}} \cdot \frac{1}{\sqrt{n \cdot \frac{k}{n} \cdot \left(1 - \frac{k}{n}\right)}} \cdot \frac{e^{n \ln n}}{e^{k \ln k} \cdot e^{(n-k)\ln(n-k)}} \cdot e^{\theta(k,n)} =$$

$$= \frac{1}{\sqrt{2\pi}} \cdot \frac{1}{\sqrt{n \cdot p^* \cdot (1-p^*)}} \cdot e^{n \ln n - k \ln k - (n-k)\ln(n-k)} \cdot e^{\theta(k,n)} =$$

$$= \frac{1}{\sqrt{2\pi}} \cdot \frac{1}{\sqrt{np^*\,(1-p^*)}} \cdot e^{(k+n-k)\ln n - k \ln k - (n-k)\ln(n-k)} \cdot e^{\theta(k,n)} =$$

$$= \frac{1}{\sqrt{2\pi}} \cdot \frac{1}{\sqrt{np^*\,(1-p^*)}} \cdot e^{-k \ln \frac{k}{n} - (n-k) \ln \frac{n-k}{n}} \cdot e^{\theta(k,n)} =$$

$$= \frac{1}{\sqrt{2\pi}} \cdot \frac{1}{\sqrt{np^*\,(1-p^*)}} \cdot e^{-n\left(\frac{k}{n} \ln \frac{k}{n} - \left(1-\frac{k}{n}\right)\ln\left(1-\frac{k}{n}\right)\right)} \cdot e^{\theta(k,n)} =$$

$$= \frac{1}{\sqrt{2\pi}} \cdot \frac{1}{\sqrt{np^*\,(1-p^*)}} \cdot e^{-n(p^* \ln p^* - (1-p^*)\ln(1-p^*))} \cdot e^{\theta(k,n)}.$$

Теперь займёмся $\theta(k,n)$. Из формулы Стирлинга имеем

$$\frac{1}{12n+1} < \theta(n) < \frac{1}{12n}, \quad -\frac{1}{12k} < -\theta(k) < -\frac{1}{12k+1},$$
$$-\frac{1}{12(n-k)} < -\theta(n-k) < -\frac{1}{12(n-k)+1}.$$

Складывая эти три неравенства, получаем

$$\frac{1}{12n+1} - \frac{1}{12k} - \frac{1}{12(n-k)} < \theta(n) - \theta(k) - \theta(n-k) = \theta(k,n) <$$
$$< \frac{1}{12n} - \frac{1}{12k+1} - \frac{1}{12(n-k)+1}.$$

2. Имеем

$$p^k q^{n-k} = e^{k\ln p+(n-k)\ln q} = e^{n\left(\frac{k}{n}\ln p+\left(1-\frac{k}{n}\right)\ln(1-p)\right)} = e^{n(p^*\ln p+(1-p^*)\ln(1-p))}.$$

Это выражение и 1 дают

$$\begin{aligned}
&\mathsf{P}(\mathbf{B}_n(p)=k) = C_n^k p^k q^{n-k} = \\
&= \frac{1}{\sqrt{2\pi n p^*(1-p^*)}} e^{-n(p^*\ln p^*+(1-p^*)\ln(1-p^*))}\cdot e^{\theta(k,n)} e^{n(p^*\ln p+(1-p^*)\ln(1-p))} = \\
&= \frac{1}{\sqrt{2\pi n p^*(1-p^*)}} e^{-n(p^*(\ln p^*-\ln p)+(1-p^*)(\ln(1-p^*)-\ln(1-p)))}\cdot e^{\theta(k,n)} = \\
&= \frac{1}{\sqrt{2\pi n p^*(1-p^*)}} e^{-n\left(p^*\ln\frac{p^*}{p}+(1-p^*)\ln\frac{1-p^*}{1-p}\right)}\cdot e^{\theta(k,n)} = \\
&= \frac{1}{\sqrt{2\pi n p^*(1-p^*)}} e^{-nH(p^*)}\cdot e^{\theta(k,n)},
\end{aligned}$$

что было нужно.

3. Следует непосредственно из утверждений 1 и 2.

□

Теорема (Локальная теорема Муавра–Лапласа).
Пусть n — натуральное число и $k\in\{1,\ldots,n-1\}$. Пусть $p\in(0,1)$ и $q=1-p$. Положим

$$x_{k,n} = \frac{k-np}{\sqrt{npq}}.$$

Тогда при $n\to\infty$, если $k-np=o(\sqrt{n})$, то

$$\mathsf{P}(\mathbf{B}_n(p)=k) = C_n^k p^k q^{n-k} \sim \frac{\varphi(x_{k,n})}{\sqrt{npq}}.$$

Замечание. Закон больших чисел для биномиальных распределений с одинаковой вероятностью успеха (Закон Бернулли) показывает, что при $n\to\infty$ вероятность $\mathsf{P}(\mathbf{B}_n(p)=k)$ возрастает при стремлении $\dfrac{k}{n}$ к p. Поэтому условие

$$k-np=o(n) \quad \text{при} \quad n\to\infty \longleftrightarrow \lim_{n\to\infty}\frac{k}{n}=p$$

становится вполне понятным.

Доказательство. Уже было отмечено, что условие $k-np=o(\sqrt{n})$ при $n\to\infty$ даёт, что $k-np=\sqrt{n}\alpha_n$, где $\alpha_n\to 0$ при $n\to\infty$, а это влечёт, что

$$\frac{k-np}{n} = \frac{\sqrt{n}\alpha_n}{n} = \frac{\alpha_n}{\sqrt{n}} = 0, \quad \text{то есть} \quad k-np=o(n).$$

Отсюда ясно, что

$$\frac{k}{n} - p \to 0 \quad \text{при} \quad n \to \infty \longleftrightarrow \lim_{n\to\infty} \frac{k}{n} = p > 0,$$

и

$$k \to \infty \quad \text{при} \quad n \to \infty,$$

Так как

$$\lim_{n\to\infty} \frac{k}{n} = p < 1,$$

то

$$n - k \to \infty \quad \text{при} \quad n \to \infty.$$

Таким образом по лемме о приближённом значении при $p^* = \frac{k}{n}$.

$$\mathsf{P}(\mathrm{B}_n(p) = k) = C_n^k p^k q^{n-k} \sim \frac{1}{\sqrt{2\pi}} \cdot \frac{1}{\sqrt{np^*(1-p^*)}} \cdot e^{-nH(p^*)}.$$

Уже отмечали, что $\lim\limits_{n\to\infty} \frac{k}{n} = p$, поэтому

$$\frac{1}{\sqrt{np^*(1-p^*)}} \sim \frac{1}{\sqrt{np(1-p)}} = \frac{1}{\sqrt{npq}} \quad \text{при} \quad n \to \infty.$$

По лемме о формуле Тейлора

$$H(p^*) = \frac{1}{2pq}(p^* - p)^2 + o((p^* - p)^2).$$

Так как $p^* = \frac{k}{n}$, то

$$(p^* - p)^2 = \left(\frac{k}{n} - p\right)^2 = \frac{(k - np)^2}{n^2}$$

и представим

$$o((p^* - p)^2) = o\left(\left(\frac{k}{n} - p\right)^2\right) = o\left(\frac{(k-np)^2}{n^2}\right) = \beta_n \cdot \frac{\alpha_n^2 n}{n^2} = \frac{\beta_n \alpha_n^2}{n}, \quad \text{где}$$

$$\beta_n \to 0 \text{ при } n \to \infty.$$

Теперь

$$n \cdot o\left((p^* - p)^2\right) = n \cdot \frac{\beta_n \alpha_n^2}{n} = \beta_n \alpha_n^2 \to 0 \quad \text{при} \quad n \to \infty.$$

Поэтому

$$nH(p^*) = n\left(\frac{1}{2pq}(p^*-p)^2 + o((p^*-p)^2)\right) =$$
$$= \frac{1}{2}\cdot\frac{(k-np)^2}{npq} + n\cdot o((p^*-p)^2) \sim \frac{1}{2}\cdot\left(\frac{k-np}{\sqrt{npq}}\right)^2 = \frac{x_{k,n}^2}{2}.$$

Отсюда

$$e^{-nH(p^*)} \sim e^{-\frac{x_{k,n}^2}{2}}.$$

Окончательно получаем

$$\mathsf{P}(\mathbf{B}_n(p) = k) = C_n^k p^k q^{n-k} \sim \frac{1}{\sqrt{2\pi}}\cdot\frac{1}{\sqrt{npq}}\cdot e^{-\frac{x_{k,n}^2}{2}} = \frac{\varphi(x_{k,n})}{\sqrt{npq}}.$$

$\square$

Замечания.

1. Теорема Пуассона (закон редких событий) и локальная теорема Муавра–Лапласа имеют общие области применения. А именно, теорема Пуассона применима, когда np $(p < q)$ мало, то есть при $n \approx 1, 2, 3, \ldots$ или np меньше 1. Локальная теорема Муавра–Лапласа хорошо применима, если $\left|\frac{k}{n} - p\right|$ достаточна мала, а именно, когда $\left|\frac{k}{n} - p\right| < \frac{1}{2}\min\{p, q\}$.
2 Приведём сведения из учебника:

   Кремер, Н.Ш. Теория вероятностей и математическая статистика: Учебник для вузов / Н.Ш. Кремер — 2-е изд., перераб. и доп. – М.: ЮНИТИ–ДАНА, 2004. – 573 с.

   Там на стр. 73 написано следующее.

   Приближённые значения вероятности, даваемые локальной формулой, на практике используются как точные при npq порядка двух и более десятков, т.е. при условии $npq \geqslant 20$.

Примеры применения локальной теоремы Муавра–Лапласа

Пусть

$$P_n = \mathsf{P}(\mathbf{B}_{2n^2}(\tfrac{1}{2}) = n^2)) = C_{2n^2}^{n^2}\frac{1}{2^{2n^2}} = \frac{(2n^2)!}{((n^2)!)^2}\cdot\frac{1}{2^{2n^2}} =$$
$$= \frac{(2n^2)!}{((n^2)!)^2 2^{2n^2}} = \frac{(2n^2)!}{((n^2)!2^{n^2})^2}.$$

У нас было

$$P_5 = \frac{C_{50}^{25}}{2^{50}} = 0,11227517265921704847642104141414.$$

Теперь

$$x_{n^2,2n^2} = \frac{n^2 - 2n^2 \cdot \frac{1}{2}}{\sqrt{2n^2pq}} = 0, \quad 2n^2pq = 2n^2 \cdot \frac{1}{4} = \frac{n^2}{2} \quad \text{и} \quad \sqrt{2n^2pq} = \frac{n}{\sqrt{2}}.$$

Локальная теорема Муавра–Лапласа даёт, что при $n \to \infty$

$$\mathsf{P}(\mathbf{B}_{2n^2}(\tfrac{1}{2}) = 2n^2)) = C_{2n^2}^{n^2} \frac{1}{2^{2n^2}} \sim \frac{\varphi(0)}{\sqrt{2n^2pq}} = \frac{\frac{1}{\sqrt{2\pi}}}{\frac{n}{\sqrt{2}}} =$$
$$= \frac{\sqrt{2}}{\sqrt{2\pi}\,n} = \frac{1}{n\sqrt{\pi}} = f_n.$$

$n = 1$. Имеем

$$P_1 = 2 \cdot \frac{1}{4} = \frac{1}{2},$$
$$f_1 = 0,56418958354775628694807945156077\ldots.$$

$n = 2$. Имеем

$$P_2 = C_8^4 \cdot \frac{1}{2^8} = \frac{70}{256} = \cdot 35128 = 0,2734375,$$
$$f_2 = 0,28209479177387814347403972578039\ldots.$$

$n = 4$. Имеем

$$P_4 = C_{32}^{16} \cdot \frac{1}{2^{32}} = \frac{601080390}{4294967296} = \frac{300540195}{2147483648} =$$
$$= 0,13994993409141898155212402343 75,$$
$$f_4 = 0,14104739588693907173701986289019\ldots.$$

$n = 8$. Имеем

$$P_8 = C_{128}^{64} \cdot \frac{1}{2^{128}} =$$
$$= 0,07038609217001513176821625194378\ldots,$$
$$f_8 = 0,07052369794346953586850993144 51\ldots.$$

$n = 16$. Имеем

$$P_{16} = C_{512}^{256} \cdot \frac{1}{2^{512}} =$$
$$= 0,035244635485838740627778947931 27\ldots,$$
$$f_{16} = 0,0352618489717347679342549657 2255\ldots.$$

$n = 32$. Имеем

$$P_{32} = C_{2048}^{1024} \cdot \frac{1}{2^{2048}} =$$
$$= 0,017628772404846519975685035029 77\ldots,$$
$$f_{32} = 0,017630924485867383967127482861 27\ldots.$$

Наша следующая цель — интегральная теорема Муавра–Лапласа, которая оценивает вероятность $\mathsf{P}(\mathbf{B}_n(p) \in [a, b))$.

4.6. Интегральная теорема Муавра–Лапласа

Теорема. *Пусть $\{\mathbf{B}_n(p)\}_{n=0}^{\infty}$ — последовательность биномиальных распределений с одинаковой вероятностью успеха $p \in (0, 1)$. Пусть*

$$-\infty \leqslant a < b \leqslant \infty.$$

Тогда при $n \to \infty$

$$\mathsf{P}(\mathbf{B}_n(p) \in [a, b)) \sim \Phi_{0,1}\left(\frac{b - np}{\sqrt{npq}}\right) - \Phi_{0,1}\left(\frac{a - np}{\sqrt{npq}}\right) =$$
$$= \frac{1}{\sqrt{2\pi}} \int_{\frac{a-np}{\sqrt{npq}}}^{\frac{b-np}{\sqrt{npq}}} e^{-\frac{x^2}{2}}\, dx.$$

Псевдодоказательство. Нормируем $\mathbf{B}_n(p)$ и получаем

$$h_n = \frac{\mathbf{B}_n(p) - np}{\sqrt{npq}}.$$

По локальной теореме Муавра–Лапласа

$$\mathsf{P}(\mathbf{B}_n(p) \in [a, b)) = \mathsf{P}\left(h_n \in \left[\frac{b - np}{\sqrt{npq}}, \frac{a - np}{\sqrt{npq}}\right)\right) =$$
$$= \sum_{k \in [a,b) \cap \mathbf{Z}} \varphi\left(x_{k,n}\right) \cdot \frac{1}{\sqrt{npq}} \left(1 + \varepsilon\left(x_{k,n}, n\right)\right),$$

здесь имеем $\frac{1}{\sqrt{npq}} \to 0$ и $\varepsilon\left(x_{k,n}, n\right) \to 0$ при $n \to \infty$.

Получается интегральная сумма для

$$\int_{\frac{a-np}{\sqrt{npq}}}^{\frac{b-np}{\sqrt{npq}}} \varphi(x)\, dx = \Phi\left(\frac{b-np}{\sqrt{npq}}\right) - \Phi\left(\frac{a-np}{\sqrt{npq}}\right).$$

Поэтому

$$\mathsf{P}(\mathbf{B}_n(p) \in [a, b)) \longrightarrow \Phi\left(\frac{b-np}{\sqrt{npq}}\right) - \Phi\left(\frac{a-np}{\sqrt{npq}}\right) \quad \text{при } n \to \infty.$$

□

Замечание. Нужно объяснить, почему "псевдодоказательство", то есть, что в этом рассуждении не полностью обосновано. Всё дело во фразе "получается интегральная сумма". Тут загвоздка в том, что требуется более чёткое и точное обоснование для того насколько, мал поправочный член $\varepsilon\left(x_{k,n}, n\right)$, а тут нужны гораздо более изощрённые рассуждения.

4.7. Центральная предельная теорема (ЦПТ)

Теорема (Центральная предельная теорема для одинаково распределённых случайных величин).
Пусть $\{g_n\}_{n=0}^{\infty}$ *— последовательность независимых одинаково распределённых случайных величин на одном вероятностном пространстве с математическим ожиданием* a *и дисперсией* $\sigma^2 > 0$ $(\sigma > 0)$. *Положим для любого натурального* n

$$S_n = g_1 + \dots + g_n \quad \text{и} \quad h_n = \frac{S_n - na}{\sigma\sqrt{n}}.$$

Тогда для любого действительного x

$$\lim_{n\to\infty} F_{h_n}(x) = \lim_{n\to\infty} \mathsf{P}(h_n < x) = \Phi_{0,1}(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{x^2}{2}} dx$$

равномерно по x.

Соображения о доказательстве. Мы не будем доказывать эту теорему, только укажем некоторые моменты.

1. Заметим, что

$$\mathsf{M}\, S_n = \mathsf{M}\, g_1 + \dots + \mathsf{M}\, g_n = na, \quad \mathsf{D}\, S_n = \mathsf{D}\, g_1 + \dots + \mathsf{D}\, g_n = n\sigma^2 \quad \text{и}$$
$$\sqrt{\mathsf{D}\, S_n} = \sigma\sqrt{n}.$$

Поэтому h_n получена нормировкой S_n.

2. Напомним, что такое *равномерно по* x в нашем случае. Последовательность функций $f_n : \mathbf{R} \to \mathbf{R}$ сходится равномерно по x к функции $f : \mathbf{R} \to \mathbf{R}$, если для любого $\varepsilon > 0$ существует такое $n_0 = n_0(\varepsilon)$ (зависящее только от $\varepsilon > 0$ и не зависящее от $x \in \mathbf{R}$), что для любого $x \in \mathbf{R}$ при $n > n_0$
$$|f_n(x) - f(x)| < \varepsilon.$$
3. Равносильная формулировка —
$$\lim_{n\to\infty} \sup_{x\in\mathbf{R}} |f_n(x) - f(x)| = 0.$$

□

Замечание. Приведём сведения из учебника:

> Общий курс высшей математики для экономистов: учебник / Под общей редакцией В.И. Ермакова — М.: ИНФРА-М, 2007. – 656 с. (100 лет РЭА им. Г.В. Плеханова).

Там на стр. 452 написано следующее.

> **Теорема 7.2.** *Если случайная величина X представляет собой сумму очень большого числа взаимно независимых случайных величин, влияние каждой из которых на всю сумму ничтожно мало, то X имеет распределение близкое к нормальному.*

Оценка погрешности

Сохраним обозначения из ЦПТ для одинаково распределённых случайных величин. Допустим, что существует $\mu = \mathsf{M}\,|g_1|^3$. Тогда для любого действительного x
$$|\mathsf{P}(h_n < x) - \Phi_{0,1}(x)| \leqslant c\frac{\mu}{3\sigma^3\sqrt{n}},$$
где c — абсолютная константа, удовлетворяющая двойному неравенству
$$\frac{1}{\sqrt{2\pi}} = 0,3989422804014326779399460599343 8 < c < 0,8.$$

Эти сведения приведены в книге Боровкова на стр. 197. Это неравенство называют *неравенством Берри–Эссеена.*

В статье

> Григорьева, М.Е. Уточнение неравенства Каца–Берри–Эссеена / М.Е. Григорьева, И.Г. Шевцова // Информатика и её применения — 2010 — Т. 4. — Вып. 2. — С. 75-82.

на стр. 76 приведены оценки

$$\frac{\sqrt{10}+3}{6\sqrt{2\pi}} = \frac{6,16227766}{15,03976965} = 0,40973218 < c < 0,4784,$$

и сказано, что нижняя оценка получена в

> Esseen C.-G. A moment inequality with an application to the central limit theorem
> Skand. Aktuarietidskr., 39 (1956), 160–170,

а верхняя в

> Королев, В.Ю. Уточнение неравенства Берри–Эссеена с приложениями к пуассоновским и смешанным пуассоновским случайным суммам / В.Ю. Королев, И.Г. Шевцова // Обозрение прикладной и промышленной математики, 17:1 (2010), 25–56.

Общий случай

Теперь рассмотрим общий случай, когда случайные величины не обязательно одинаково распределены. На самом деле деле мы рассмотрим ещё более общую ситуацию, когда для каждого натурального n на n-ом шаге происходит n испытаний, которые мы будем трактовать как n случайных величин. Более формально, имеем последовательность серий

$$\{(g_{1,n},\dots,g_{n,n})\}_{n=1}^{\infty}$$

на одном вероятностном пространстве со следующими свойствами.

ЦПТ 1. В каждой серии случайные величины $g_{1,n},\dots,g_{n,n}$ независимы.

ЦПТ 2. Для любых i и n случайная величина $g_{i,n}$ центрирована, то есть $\mathsf{M}\, g_{i,n} = 0$.

ЦПТ 3. Для каждого натурального n положим

$$S_n = \sum_{i=1}^{n} g_{i,n},$$

и допустим, что S_n нормирована, то есть

$$\mathsf{M}\, S_n = \sum_{i=1}^{n} \mathsf{M}\, g_{i,n} = 0 \quad \text{и} \quad \mathsf{D}\, S_n = 1.$$

Замечание. Условия ЦПТ 2 и ЦПТ 3 легко получить стандартными приёмами центрирования и нормирования как в разделе 3.

ЦПТ 4. Условие Линдеберга.

Для любого действительного $\tau > 0$

$$\lim_{n\to\infty} \sum_{i=1}^{n} \mathsf{M}(g_{i,n}^2 \mid |g_{i,n}| > \tau) = 0.$$

Замечание. Поясним, что под $\mathsf{M}(g_{i,n}^2 \mid |g_{i,n}| > \tau)$ понимаем математическое ожидание случайной величины

$$h_{i,n}(\omega) = \begin{cases} g_{i,n}^2(\omega) & \text{при } g_{i,n}^2(\omega) > \tau, \\ 0 & \text{при } g_{i,n}^2(\omega) \leqslant \tau. \end{cases}$$

Подобный приём использовался при доказательстве неравенства Чебышёва.

Теорема (Центральная предельная теорема).
Пусть для последовательности серий

$$\{(g_{1,n}, \dots, g_{n,n})\}_{n=1}^{\infty}$$

выполнены условия **ЦПТ 1–ЦПТ 4**. *Тогда для любого действительного x*

$$\lim_{n\to\infty} \mathsf{P}(S_n < x) = \Phi_{0,1}(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt$$

равномерно по x.

Доказывать не будем.

ЦПТ 5. Условие Ляпунова.

Вместо условия Линдеберга ЦПТ 4 можно рассматривать более сильное условие — условие Ляпунова. Пусть $s > 2$ и для любого натурального n положим

$$L_s(n) = \sum_{i=1}^{n} \mathsf{M}\, |g_{i,n}|^s$$

и допустим, что

$$\lim_{n\to\infty} L_s(n) = 0.$$

Следствие (Интегральная теорема Муавра–Лапласа). *Пусть* $\{\mathbf{B}_n(p)\}_{n=0}^{\infty}$ — *последовательность биномиальных распределений с одинаковой вероятностью успеха* $p \in (0,1)$. *Пусть*

$$-\infty \leqslant a < b \leqslant \infty.$$

Тогда при $n \to \infty$

$$\mathsf{P}(\mathbf{B}_n(p) \in [a,b)) \sim \Phi_{0,1}\left(\frac{b-np}{\sqrt{npq}}\right) - \Phi_{0,1}\left(\frac{a-np}{\sqrt{npq}}\right) =$$
$$= \frac{1}{\sqrt{2\pi}} \int_{\frac{a-np}{\sqrt{npq}}}^{\frac{b-np}{\sqrt{npq}}} e^{-\frac{x^2}{2}}\,dx.$$

Доказательство. В самом деле, рассмотрим биномиальное распределение $\mathbf{B}_n(p)$, как сумму независимых случайных величин $g_1, \ldots, g_n$ на одном вероятностном пространстве, распределённых по схеме Бернулли с одной и той же вероятностью успеха p. Пронормируем и получим

$$h_n = \frac{\mathbf{B}_n(p) - np}{\sqrt{npq}}.$$

По свойству функции распределения

$$\mathsf{P}(h_n \in [\alpha, \beta)) = \mathsf{P}(h_n < \beta) - \mathsf{P}(h_n < \alpha).$$

По ЦПТ для одинаково распределённых случайных величин

$$\lim_{n\to\infty} \mathsf{P}(h_n \in [\alpha, \beta)) = \Phi_{0,1}(\beta) - \Phi(\alpha).$$

Заметим, что

$$\mathsf{P}(\mathbf{B}_n(p) \in [a,b)) = \mathsf{P}\left(h_n = \frac{\mathbf{B}_n(p) - np}{\sqrt{npq}} \in \left[\frac{b-np}{\sqrt{npq}}, \frac{a-np}{\sqrt{npq}}\right)\right).$$

Поэтому при

$$\alpha = \frac{a-np}{\sqrt{npq}} \quad \text{и} \quad \beta = \frac{b-np}{\sqrt{npq}}$$

получаем требуемое. □

Замечание. Приведём сведения из учебника:

Кремер, Н.Ш. Теория вероятностей и математическая статистика: учебник для вузов / Н.Ш. Кремер — 2-е изд., перераб. и доп. — М.: ЮНИТИ–ДАНА, 2004. — 573 с.

Там на стр. 75 написано следующее.

> При выполнении условия $npq \geqslant 20$ интегральная формула, так же как и локальная, даст, как правило, удовлетворительную для практики погрешность вычисления вероятностей.

Примеры применения интегральной теоремы Муавра–Лапласа

Пусть

$$\mathsf{P}(\mathbf{B}_{2m}(\tfrac{1}{2}) = m)) = C_{2m}^{m}\frac{1}{2^{2m}} = \frac{(2m)!}{(m!)^2}\cdot\frac{1}{2^{2m}} = \frac{(2m)!}{(m!)^2 2^{2m}} = \frac{(2m)!}{(m!2^m)^2}.$$

Заметим, что

$$\mathsf{P}(\mathbf{B}_{2m}(\tfrac{1}{2}) = m)) = \mathsf{P}(\mathbf{B}_{2m}(\tfrac{1}{2}) \in [m-0.5, m+0.5))$$

и можно применять интегральную теорему Муавра–Лапласа

$$\mathsf{P}(\mathbf{B}_{2m}(\tfrac{1}{2}) = m)) \sim \Phi_{0,1}\left(\frac{0.5}{\sqrt{2m\cdot 0,25}}\right) - \Phi_{0,1}\left(\frac{-0.5}{\sqrt{2m\cdot 0,25}}\right) =$$
$$= 2\Phi_{\text{Гм}}\left(\frac{0.5}{0,5\sqrt{2m}}\right) = 2\Phi_{\text{Гм}}\left(\frac{1}{\sqrt{2m}}\right) =$$
$$= \frac{2}{\sqrt{2\pi}}\int_0^{\frac{1}{\sqrt{2m}}} e^{-\frac{x^2}{2}}\,dx.$$

Рассмотрим $m = 2n^2$. Пусть

$$b_n = \mathsf{P}(\mathbf{B}_{4n^2}(\tfrac{1}{2}) = 2n^2)) = C_{4n^2}^{2n^2}\frac{1}{2^{4n^2}} = \frac{(4n^2)!}{((2n^2)!)^2}\cdot\frac{1}{2^{4n^2}} = \frac{(4n^2)!}{((2n^2)!)^2 2^{4n^2}} =$$
$$= \frac{(4n^2)!}{((2n^2)!2^{2n^2})^2} \quad \text{и}$$
$$c_n = \frac{2}{\sqrt{2\pi}}\int_0^{\frac{1}{2n}} e^{-\frac{x^2}{2}}\,dx.$$

$n = 1$. Имеем

$$b_1 = 6\cdot\frac{1}{16} = \frac{3}{8} = 0,375,$$
$$c_1 = 0,382924922548026\ldots.$$

$n = 2$. Имеем

$$b_2 = C_{16}^{8}\cdot\frac{1}{2^{16}} = \frac{12870}{65536} = \frac{6435}{32768} = 0,196380615234375,$$
$$c_2 = 0,197412651365847\ldots.$$

$n = 4$. Имеем

$$\begin{aligned} b_4 &= C_{64}^{32} \cdot \frac{1}{2^{64}} = \frac{1832624140942590534}{18446744073709551616} = \\ &= \frac{916312070471295267}{9223372036854775808} = \\ &= 0,0993467537479668965283083337 2477\ldots, \\ c_4 &= 0,099476449660226\ldots. \end{aligned}$$

$n = 8$. Имеем

$$\begin{aligned} b_8 &= C_{256}^{128} \cdot \frac{1}{2^{256}} = \\ &= 0,04981910993614015123827971711748\ldots, \\ c_8 &= 0,049835338058494\ldots. \end{aligned}$$

$n = 16$. Имеем

$$\begin{aligned} b_{16} &= C_{1024}^{512} \cdot \frac{1}{2^{1024}} = \\ &= 0,02492780589297954371223463220026\ldots, \\ c_{16} &= 0,024929834868754\ldots. \end{aligned}$$

5. Двумерные случайные величины

5.1. Понятие о многомерных случайных величинах

Пусть заданы n случайных величин $g_1, \ldots, g_n$, вообще говоря, на разных вероятностных пространствах. Тогда упорядоченный набор

$$(g_1, \ldots, g_n)$$

назовём *n-мерной случайной величиной.* Функцию

$$F_{(g_1,\ldots,g_n)} : \mathbf{R}^n \to \mathbf{R},$$

где для любой точки $(x_1, \ldots, x_n) \in \mathbf{R}^n$,

$$F_{(g_1,\ldots,g_n)}((x_1, \ldots, x_n)) = \mathsf{P}(g_1 < x_1, \ldots, g_n < x_n))$$

назовём *функцией распределения* n-мерной случайной величины $(g_1, \ldots, g_n)$.

Далее более подробно рассмотрим двумерные случайные величины.

5.2. Двумерные дискретные случайные величины

Как и в одномерном случае дискретную двумерную случайную величину удобнее задавать набором её значений и работать с этим набором. Более точно, пусть $\{(x_i, y_j)\}_{i,j}$ — набор всех значений двумерной дискретной случайной величины (g, h). Тогда нам нужно задать вероятности

$$\mathsf{P}(g = x_i.h = y_j) = p_{ij}$$

для всех пар (i, j). таким образом возникает матрица $(p_{ij})_{i,j}$, которая может быть бесконечной как по строкам, так и по столбцам. Заметим, что

$$\sum_{j,j} \mathsf{P}(g = x_i, h = y_j) = \sum_{j,j} p_{ij} = 1.$$

Теперь можно достаточно просто определить вероятности значений случайной величины g, а именно, для каждого i

$$\mathsf{P}(g = x_i) = \sum_j \mathsf{P}(g = x_i.h = y_j) = \sum_j p_{ij} = p_i,$$

сумма по строчке матрицы верятностей $(p_{ij})_{i,j}$.

Аналогично, для каждого j

$$\mathsf{P}(h = y_j) = \sum_i \mathsf{P}(g = x_i, h = y_j) = \sum_i p_{ij} = q_j,$$

сумма по столбцу матрицы вероятностей $(p_{ij})_{i,j}$.

Напомним, что для независимости дискретных случайных величин g и h нужно проверить равенства

$$\mathsf{P}(g = x_i, h = y_j) = \mathsf{P}(g = x_i)\,\mathsf{P}(h = y_j)$$

для любых пар (i, j). В наших обозначениях это записывается следующим образом. Для любых i и j

$$p_{ij} = p_i q_j.$$

Нарушение этого равенства хотя бы для одной пары (i, j) влечёт *зависимость* случайных величин g и h.

5.3. Двумерные непрерывные случайные величины

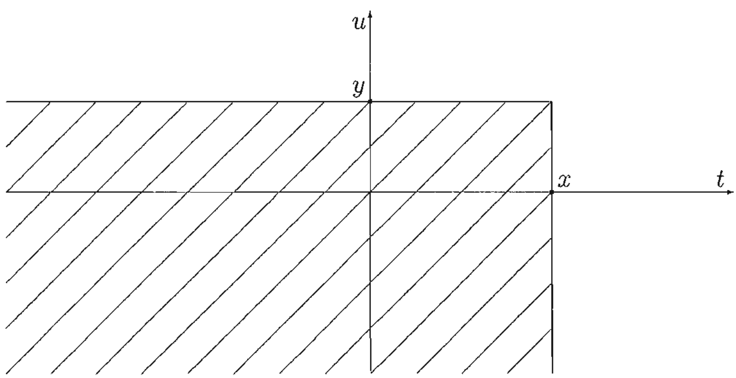
Функция распределения непрерывной случайной величины (g, h) задаётся при любых действительных x и y как

$$F_{(g,h)}(x, y) = \mathsf{P}(g < x, h < y) = \int_{-\infty}^{x}\int_{-\infty}^{y} dF_{(g,h)}(t, u),$$

или при наличии двумерной плотности $\mathrm{pd}_{(g,h)}(t, u)$ имеем

$$F_{(g,h)}(x, y) = \int_{-\infty}^{x}\int_{-\infty}^{y} \mathrm{pd}_{(g,h)}\,dt\,du.$$

Геометрически значение $F_{(g,h)}(x, y)$ — это вероятность попадания в “угол”



Заметим, что можно весьма просто определить функции распределения случайных величин g и h, а именно, для любых действительных x и y

$$F_g(x) = \mathsf{P}(g < x) = \int_{-\infty}^{x}\int_{-\infty}^{\infty} dF_{(g,h)}(t,u),$$

$$F_h(y) = \mathsf{P}(h < y) = \int_{-\infty}^{\infty}\int_{-\infty}^{y} dF_{(g,h)}(t,u),$$

или при наличии двумерной плотности $\mathrm{pd}_{(g,h)}(t,u)$ имеем

$$F_g(x) = \int_{-\infty}^{x}\int_{-\infty}^{\infty} \mathrm{pd}_{(g,h)}\,dt\,du,$$

$$F_h(y) = \int_{-\infty}^{\infty}\int_{-\infty}^{y} \mathrm{pd}_{(g,h)}\,dt\,du.$$

Напомним, что независимость случайных величин g и h по определению означает, что для любых действительных x и y

$$F_{(g,h)}(x,y) = F_g(x)F_h(y).$$

Поэтому, на деле, в рассматриваемом случае независимость проверяется вычислением трёх двойных интегралов.

Теперь о плотностях. Плотность $\mathrm{pd}_{(g,h)}(t,u)$ можно найти, дифференцируя двойной интеграл для функции распределения (конечно, если это можно делать), а именно,

$$\mathrm{pd}_{(g,h)}(t,u) = \frac{\partial^2 F}{\partial x \partial y}(t,u).$$

Плотность распределения g

$$\mathrm{pd}_g(t) = \int_{-\infty}^{\infty} \mathrm{pd}_{(g,h)}(t,u)du.$$

Аналогично для распределения h

$$\mathrm{pd}_h(u) = \int_{-\infty}^{\infty} \mathrm{pd}_{(g,h)}(t,u)dt.$$

6. Математическая статистика. Основные подходы

6.1. Постановка основных задач математической статистики

Пусть имеется некоторый набор чисел, которые (известно или предположительно) имеют вероятностный характер, например получены в результате некоторых наблюдений, на которые влияют многие факторы, которые не всегда и не все можно и нужно учитывать.

Классический пример такого случая — бросание монеты. Если писать уравнение движения брошенной монеты, то нужно учитывать огромное множество факторов: начальные положение, сила и направление бросания, характеристики воздуха (температура, влажность, давление, наличие воздушных потоков), характеристики пола (ровность, упругость (может монета подскакивать)) и т.д., и т.п. Ясно, что перечислить все факторы влияющие на движение монеты иногда и невозможно, к примеру, неожиданно открытые дверь, форточка. А самое главное состоит в том, что пока решается это уравнение движения, монета уже давно упала, и решение становится бесполезным.

Можно ставить две основных задачи.

Задача о прошлом.

Что же лежит в основе полученных данных?

Задача о будущем.

Если мы ещё сделаем наблюдения, то что можно будет о них сказать?

На деле это всё таки одна задача — получение как можно более полной информации из имеющихся данных. Конечно может оказаться, что некоторые данные ошибочны или очень не точны, или существенно зависят от других данных и т.п. Конечно, было бы желательно иметь методы и подходы на все случаи жизни, но вряд ли это возможно. Поэтому мы выделим только один основной класс задач, который мы будем изучать, впрочем иногда сворачивая и немного в сторону.

Основные предположения

Есть некоторая случайная величина, о которой либо совсем ничего не знаем, либо знания неполны, например, есть нормальное распределение с неизвестными параметрами. Во всяком случае, наши знания об этой случайной величины не позволяют полностью описать её. Конечно, в идеале надо бы восстановить функцию распределения этой случайной величины и ясно, что это не всегда возможно, но с другой стороны можно пытаться получить какую-то достаточно полную информацию о важнейших характеристиках случайной величины таких как: математическое ожидание, дисперсия и т.п. Итак всегда будем подразумевать, что выполнены следующие предположения.

1. Генеральная совокупность

Множество всех значений неизвестной случайной величины.

2. Выборка

Конечный набор (возможны повторения) чисел из указанной генеральной совокупности, то есть конечный набор значений неизвестной случайной величины. Отметим, что можно рассматривать и бесконечные выборки, но мы этого не будем делать.

3. Объём выборки

Число элементов выборки. Сразу же отметим, что выборка должна быть *репрезентативной*, то есть, объём выборки должен быть достаточно велик, конкретное значение для объёма очень зависит от задач, которые поставлены. Для экспериментов, которые легко могут быть проведены, объём может быть очень большим, а для трудно повторяемых приходится довольствоваться малыми объёмами. Часто вполне может быть достаточно объёма в 20 или более элементов. Отметим также, что когда объём выборки очень велик, то разумно использовать подходы для бесконечных выборок, но, как уже было сказано, мы этим заниматься не будем.

4. Вариационный ряд

Вариационный ряд — это выборка упорядоченная по возрастанию, то есть представленная в виде

$$x_1 \leqslant x_2 \leqslant \cdots \leqslant x_{n-1} \leqslant x_n.$$

5. Характеристики выборки

Кроме объёма используют часто другие характеристики выборки, например, точность имеющихся данных. Кроме того, могут быть использованы ещё следующие.

Мода — наибольшее значение объёма среди различных элементов выборки, то есть, тот элемент который встречается чаще всего.

Медиана — среднее значение среди элементов выборки, более точно, для вариационного ряда при нечётном объёме $n = 2m-1$ медианой будет элемент x_m, а при чётном объёме $n = 2m$ медианой будет $\frac{x_m+x_{m+1}}{2}$.

6. Элементы выборки

Будем считать что элементы выборки независимы друг от друга, поэтому их можно рассматривать как *независимые случайные величины с тем же законом распределения что и неизвестная случайная величина.*

7. Оценка

Оценка — всякая "неплохая" функция от элементов выборки, и более точно, согласно нашему подходу — это всякая функция от случайных величин с распределением как у нашей неизвестной случайной величины, которая сама является случайной величиной. Обычно оценки для математического ожидания, дисперсии и т.п.

8. Свойства оценок

Конечно, если не требовать от оценки никаких связей с тем, что она оценивает, то оценка бесполезна. Надо наложить такие требования на оценку, чтобы она давала, грубо говоря, похожее на оцениваемое. Пусть θ — некоторый параметр неизвестного распределения и θ^* — оценка для этого параметра, которую мы будем рассматривать как случайную величину.

Несмещённость оценки. Разумно требовать, чтобы в среднем в пределе оценка совпадала с параметром. Более точно, θ^* называется *несмещённой* оценкой параметра θ, если её математическое ожидание совпадает с θ, то есть

$$\mathsf{M}\,\theta^* = \theta.$$

Обычно говорят, что несмещённость обеспечивает отсутствие систематических ошибок.

Состоятельность оценки. Разумно также требовать, чтобы оценка при увеличении объёма выборки становилась точнее, поэтому мы для выборки объёма n будем записывать оценку как θ_n^*. Итак, θ_n^* называется *состоятельной* оценкой (при $n \to \infty$), если для любого $\varepsilon > 0$

$$\lim_{n\to\infty} \mathsf{P}(|\theta_n^* - \theta| < \varepsilon) = 1.$$

Эффективность оценки. Если рассматривается целый класс Θ оценок, то θ^* называется *эффективной* оценкой в классе Θ, если

$$\mathsf{D}\,\theta^* = \min\{\mathsf{D}\,\eta \mid \eta \in \Theta\}.$$

Сразу же отметим, что мы не будем рассматривать эффективность оценок.

Вышеизложенные свойства (несмещённость, состоятельность, эффективность) относят к классу *точечных* оценок, когда значение оценки задаёт приближённое значение оцениваемого параметра.

Другой класс — вероятностные или интервальные оценки.

9. Доверительная вероятность = надёжность оценки

Пусть $\varepsilon > 0$. *Доверительная вероятность* с точностью ε равна $\mathsf{P}(|\theta^* - \theta| < \varepsilon) = \gamma$. Чуть более общо. Зададим интервал (α, β). Пусть

$$\mathsf{P}(|\theta^* - \theta| \in (\alpha, \beta)) = \mathsf{P}(\theta^* - \alpha < \theta < \theta^* + \beta) = \\ = \mathsf{P}\left(\theta \in (\theta^* - \alpha, \theta^* + \beta)\right) = \gamma.$$

Тогда будем говорить, что параметр θ находится в интервале

$$(\theta^* - \alpha, \theta^* + \beta)$$

с доверительной вероятностью γ.

10. Доверительный интервал

Можно исходить от обратной задачи. Зададим γ и найдём такое $\varepsilon > 0$, что $\mathsf{P}(|\theta^* - \theta| < \varepsilon) = \gamma$. Тогда назовём $(\theta^* - \varepsilon, \theta^* + \varepsilon)$ *доверительным интервалом* для доверительной вероятности γ. Более общо, если

$$\mathsf{P}\left(\theta \in (\theta^* - \alpha, \theta^* + \beta)\right) = \gamma,$$

то $(\theta^* - \alpha, \theta^* + \beta)$ — *доверительным интервалом* для доверительной вероятности γ.

6.2. Эмпирические вероятности

Рассмотрим вариационный ряд

$$x_1 \leqslant x_2 \leqslant \cdots \leqslant x_{n-1} \leqslant x_n.$$

Часто нужно рассматривать группы значений. Для этого есть по крайней мере две причины.

1. Среди значений много одинаковых.
2. Среди значений немного одинаковых, но известно что значения либо не совсем точны, либо заранее известно, что значения допускают некоторый разброс, к примеру, рост одного и того же человека довольно сильно разнится утром и вечером. Поэтому разумно определить в одну группу близкие значения.

Оставим в стороне такой важный вопрос как группировать значения и будем считать, что выделены значения

$$\tilde{x}_1 < \tilde{x}_2 < \dots < \tilde{x}_k.$$

Пусть для каждого $i \in \{1, \dots, k\}$ через n_i обозначено число значений, которые равны $\tilde{x}_i$, более точно, соответствуют $\tilde{x}_i$. Тогда

$$n_1 + n_2 + \dots + n_k = n,$$

положим

$$\nu_1 = \frac{n_1}{n}, \ \nu_2 = \frac{n_2}{n}, \dots, \nu_k = \frac{n_k}{n}$$

и назовём их *эмпирическими* вероятностями.

Тогда вариационный ряд можно представить следующими способами.

Значение	$\tilde{x}_1$	$\tilde{x}_2$	$\cdots$	$\tilde{x}_k$
Частота	n_1	n_2	$\cdots$	n_k

или

Значение	$\tilde{x}_1$	$\tilde{x}_2$	$\cdots$	$\tilde{x}_k$
Эмпирическая вероятность	ν_1	ν_2	$\cdots$	ν_k

и объём выборки равен n.

Отметим, что добавку “и объём выборки равен n” нельзя отбросить без ущерба. Мы увидим это неоднократно впоследствии, ну а сейчас уже можно сказать, что хотя

$$0.1 = \frac{1}{10} = \frac{10}{100},$$

третье выражение несёт больше информации, разумеется, при условии, что $n = 100$.

С эмпирическими вероятностями можно связать *эмпирическую* функцию распределения F_n^*, положив для любого действительного x

$$n(x) = \text{число элементов выборки} < x,$$

$$F_n^*(x) = \sum_{\tilde{x}_i < x} \nu_i = \sum_{\tilde{x}_i < x} \frac{n_i}{n} = \frac{\sum_{\tilde{x}_i < x} n_i}{n} = \frac{\sum_{x_j < x} 1}{n} =$$

$$= \frac{n(x)}{n} = \frac{m}{n}, \quad \text{если } x \in (\tilde{x}_m, \tilde{x}_{m+1}],$$

где $m \in \{0, \dots, k\}$ и $\tilde{x}_0 = -\infty$, $\tilde{x}_{k+1} = \infty$.

Теорема (Гливенко–Кантелли).
Пусть $F(x)$ — функция распределения случайной величины, а $F_n^*(x)$ — эмпирическая функция распределения, построенная по выборке объёма n из значений этой случайной величины. Тогда для любых действительных x и $\varepsilon > 0$

$$\lim_{n\to\infty} \mathsf{P}(|F_n^*(x) - F(x)| < \varepsilon) = 1.$$

Доказательство. Зафиксируем $x \in \mathbf{R}$ и положим для любого $i \in \{1, \ldots, n\}$

$$g_i(x) = \begin{cases} 0, & x_i \geqslant x, \\ 1, & x_i < x. \end{cases}$$

Тогда все $g_1, \ldots, g_n$ — случайные величины, распределённые по схеме Бернулли

$$\mathsf{P}(g_i = 0) = \mathsf{P}(x_i \geqslant x) = 1 - F(x) \quad \text{и} \quad \mathsf{P}(g_i = 1) = \mathsf{P}(x_i < 0) = F(x).$$

и для любого $i \in \{1, \ldots, n\}$

$$\mathsf{M}\, g_i = 0 \cdot \mathsf{P}(g_i = 0) + 1 \cdot \mathsf{P}(g_i = 1) = F(x).$$

Заметим, что

$$\frac{g_1 + \cdots + g_n}{n} = F_n^*(x).$$

По закону больших чисел в форме Бернулли для любого $\varepsilon > 0$

$$\lim_{n\to\infty} \mathsf{P}\left(\left|\frac{g_1 + \cdots + g_n}{n} - F(x)\right| < \varepsilon\right) = \lim_{n\to\infty} \mathsf{P}\left(|F_n^*(x) - F(x)| < \varepsilon\right) = 1,$$

что и нужно. □

Замечания. Сделаем несколько полезных замечаний.

1. Если неизвестная случайная величина дискретна, то со значениями

$$\tilde{x}_1, \ldots, \tilde{x}_k$$

   или близкими к ним связаны некоторые вероятности $p_1, \ldots, p_k$. Наша цель состоит в том, чтобы стараться оценить $|\nu_1 - p_1|, \ldots, |\nu_k - p_k|$, то есть разности между эмпирическими и настоящими вероятностями.
2. Если неизвестная случайная величина непрерывна, то обычно, используют группы результатов. Разбивают действительную прямую $\mathbf{R}$ на конечное число интервалов, причём крайние интервалы бесконечны, в

каждом из них выбирают точку, в конечных интервалах часто это середина интервала. На основе этого получают эмпирические вероятности и функцию распределения. Опишем это более формально. Пусть

$$-\infty = x_0 < t_1 < t_2 < \cdots < t_{k-1} < t_k = \infty.$$

Для каждого $i \in \{1, \ldots, k-1\}$ выберем

$$\tilde{x}_i \in (t_{i-1}, t_i] \quad \text{и} \quad \tilde{x}_k \in (t_{k-1}, t_k = \infty).$$

Тогда частоты определяются как число значений попавших данный интервал, то есть для каждого $i \in \{1, \ldots, k-1\}$

$$n_i = \text{ число значений, попавших в интервал } (t_{i-1}, t_i] \quad \text{и}$$
$$n_k = \text{ число значений, попавших в интервал } (t_{k-1}, t_k = \infty).$$

Далее действуем как раньше.

3. Если посмотреть решения в задачнике Гмурмана, относящиеся к математической статистике, как, впрочем, и к математическим ожиданиям, и дисперсиям, то можно заметить, что очень часто предлагается перейти к другим параметрам при вычислениях. Исторически это было предопределено тем, что фактически не было средств для вычислений, поэтому надо было вычислять очень экономно. С другой стороны, такие подходы часто дают результаты, которые выглядят более ясно и часто позволяют быстро находить заведомо неверные вычисления. Опишем один из подобных подходов. Зафиксируем некоторое число c, часто его берут равным медиане, и рассматривают

$$y_1 = \tilde{x}_1 - c, \ldots, y_k = \tilde{x}_k - c.$$

Происходит как бы центрирование, часть чисел становится отрицательной, часть — положительной. Поэтому при сложении или возведении в квадрат могут уменьшиться вычисления.

Далее предположим, что была группировка в выборке, либо из-за неточности данных, либо при предположении непрерывности случайной величины, и допустим, что все интервалы конечной длины это может быть разумно даже в непрерывном случае, к примеру, рост человека меняется в очень ограниченных пределах.

Пусть $h_1, \ldots, h_k$ — длины интервалов, тогда рассматривают

$$z_1 = \frac{y_1}{h_1} = \frac{\tilde{x}_1 - c}{h_1}, \ldots, z_k = \frac{y_k}{h_k} = \frac{\tilde{x}_k - c}{h_k}.$$

Если изучается непрерывное распределение, то можно оценить плотность распределения следующим образом. Пусть границы интервалов

$$t_0 < t_1 < t_2 < \cdots < t_{k-1} < t_k.$$

Для любого $i \in \{1, \ldots, k\}$

$$h_i = t_i - t_{i-1}.$$

Для любого $i \in \{1, \ldots, k\}$ приращение эмпирической функции распределения на i-том интервале равно

$$F^*(t_i) - F^*(t_{i-1}) = \frac{n(t_i)}{n} - \frac{n(t_{i-1})}{n} = \frac{n(t_i) - n(t_{i-1})}{n} = \frac{n_i}{n} = \nu_i.$$

Поэтому относительное приращение эмпирической функции распределения на i-том интервале равно

$$\frac{F^*(t_i) - F^*(t_{i-1})}{h_i} = \frac{\frac{n_i}{n}}{h_i} = \frac{\nu_i}{h_i}.$$

Поскольку плотность — это производная функции распределения, то можно считать, что тем самым оценили и плотность.

Удобнее всего работать, когда

$$h_i = h = \text{const} \quad \text{для всех } i \in \{1, \ldots, k\}.$$

6.3. Графические изображения

Часто удобно и важно изображать выборки в графическом виде.

6.3.1. Полигон (многоугольник) частот

Этот способ очень похож на график конечных дискретных распределений, то что было названо “многоугольник распределения”. Поэтому мы ограничимся иллюстрацией (Рис. 17).

Можно вместо частот откладывать эмпирические вероятности.

6.3.2. Гистограмма (столбчатая диаграмма) частот

Если использовали группировку или интервальное задание в непрерывном случае, то удобно использовать гистограммы (Рис. 18).

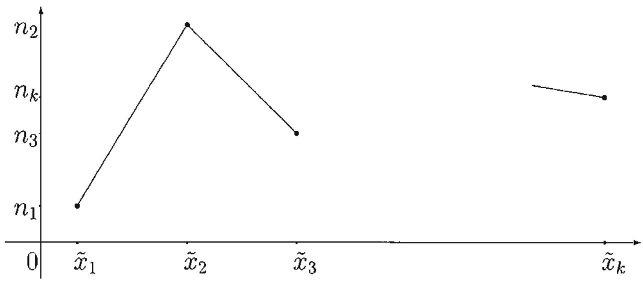


Рис. 17. Полигон (многоугольник) частот

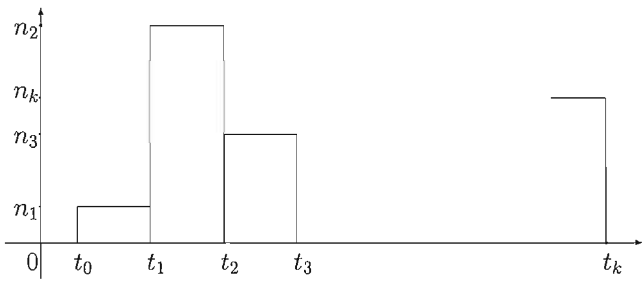
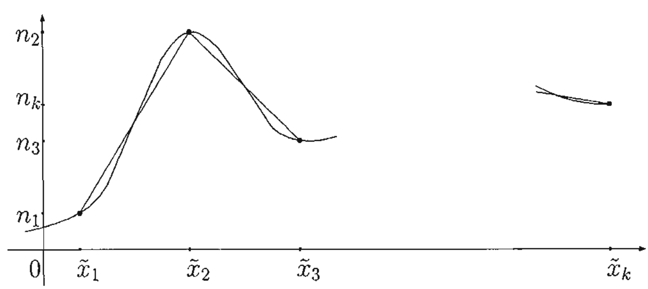


Рис. 18. Гистограмма частот

6.3.3. Непрерывный случай

Если знаем, что неизвестная случайная величина непрерывна, то можно соединить кривой точки, соответствующие данным (интерполяция), или указать кривую данного вида, которая проходит близко к данным точкам (аппроксимация).



7. Точечное и интервальное оценивание

7.1. Точечные оценки

Основной задачей точечного оценивания является построение таких функций от выборок, которые могли бы дать представление о параметрах распределения. Мы сосредоточимся только на двух параметрах — математическое ожидание и дисперсия. Конечно, есть и другие параметры, но эти самые важные, а другие либо оцениваются сходными методами, либо встречаются весьма редко, и потому не стоит тратить наше очень ограниченное время.

7.1.1. Математическое ожидание

Математическое ожидание можно оценивать по-разному. Например — медианой, модой, полусуммой наибольшего и наименьшего значений и т.п. Однако самой распространённой и, во многих смыслах, самой лучшей оценкой является среднее арифметическое (мы будем часто говорить просто — среднее)

$$\overline{x} = \overline{x}_n = \frac{x_1 + \cdots + x_n}{n} = \frac{n_1\tilde{x}_1 + \cdots + n_k\tilde{x}_k}{n} = \tilde{x}_1\nu_1 + \cdots + \tilde{x}_k\nu_k.$$

Теорема (Об оценивании математического ожидания).
$\overline{x}$ — несмещённая, состоятельная оценка математического ожидания.

Доказательство. Будем пользоваться основными предположениями, сделанными ранее в пункте 6.1. По нашим предположениям любой элемент x_i выборки имеет то же распределение, что и неизвестная случайная величина g, в частности, $\mathsf{M}\, x_i = \mathsf{M}\, g$.

Несмещённость.

Имеем

$$\mathsf{M}\,\overline{x} = \frac{\mathsf{M}\, x_1 + \cdots + \mathsf{M}\, x_n}{n} = \frac{n\,\mathsf{M}\, g}{n} = \mathsf{M}\, g,$$

что и нужно.

Состоятельность.

По закону больших чисел для одинаково распределённых случайных величин для любого $\varepsilon > 0$

$$\mathsf{P}\left(|\overline{x} - \mathsf{M}\, g| < \varepsilon\right) = \mathsf{P}\left(\left|\frac{x_1 + \cdots + x_n}{n} - \mathsf{M}\, g\right| < \varepsilon\right) \to 1 \quad \text{при } n \to \infty,$$

что хотели.

□

7.1.2. Дисперсия

Вполне допустима ситуация, когда нам заранее известно математическое ожидание, равное a. Можно указать по крайней мере две причины — это просто задано заранее, или до этого удалось, используя предшествующие испытания, найти его. Тогда в качестве оценки дисперсии укажем

$$\sigma_n^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - a)^2 .$$

Теорема (Об оценке дисперсии при известном математическом ожидании). *σ_n^2 — несмещённая, состоятельная оценка дисперсии.*

Доказательство. Будем пользоваться основными предположениями, сделанными ранее в пункте 6.1. По нашим предположениям любой элемент x_i выборки имеет то же распределение, что и неизвестная случайная величина g.

Несмещённость.

Имеем

$$\mathsf{M}\,\sigma_n^2 = \frac{1}{n}\sum_{i=1}^{n}\mathsf{M}\,(x_i - a)^2 = \frac{n\,\mathsf{D}\,g}{n} = \mathsf{D}\,g,$$

что и нужно.

Состоятельность.

По закону больших чисел для одинаково распределённых случайных величин для любого $\varepsilon > 0$

$$\mathsf{P}\left(\left|\sigma_n^2 - \mathsf{D}\,g\right| < \varepsilon\right) = \\ = \mathsf{P}\left(\left|\frac{\sum_{i=1}^{n}(x_i - a)^2}{n} - \mathsf{M}(g-a)^2\right| < \varepsilon\right) \to 1$$

при $n \to \infty$,

что хотели.

□

Если неизвестно математическое ожидание, в качестве оценки дисперсии укажем

$$s_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2 .$$

Часто рассматривают в качестве оценки дисперсии

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

Последняя оценка состоятельна, но смещена.

Теорема (Об оценивании дисперсии при неизвестном математическом ожидании).
s_n^2 — *несмещённая, состоятельная оценка дисперсии.*

Доказательство. Будем пользоваться основными предположениями, сделанными ранее в пункте 6.1. По нашим предположениям любой элемент x_i выборки имеет то же распределение, что и неизвестная случайная величина g.

Несмещённость.
Сначала преобразуем

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

Также

$$n\overline{x}^2=n\cdot\frac{1}{n^2}\left(\sum_{i=1}^{n}x_i\right)^2=\frac{1}{n}\left(\sum_{i=1}^{n}x_i^2+2\sum_{1\leqslant i<j\leqslant n}x_ix_j\right)$$

Поэтому

$$s_n^2=\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\overline{x})^2=\frac{1}{n-1}\sum_{i=1}^{n}x_i^2-\frac{n}{n-1}\overline{x}^2=$$
$$=\frac{1}{n-1}\sum_{i=1}^{n}x_i^2-\frac{1}{n-1}\cdot\frac{1}{n}\left(\sum_{i=1}^{n}x_i^2+2\sum_{1\leqslant i<g\leqslant n}x_ix_j\right)=$$
$$=\frac{1}{n-1}\left(1-\frac{1}{n}\right)\sum_{i=1}^{n}x_i^2-\frac{2}{n(n-1)}\sum_{1\leqslant i<g\leqslant n}x_ix_j=$$
$$=\frac{1}{n}\sum_{i=1}^{n}x_i^2-\frac{2}{n(n-1)}\sum_{1\leqslant i<g\leqslant n}x_ix_j.$$

Отсюда математическое ожидание

$$\begin{aligned}
\mathsf{M}\, n\overline{x}^2 &= \frac{1}{n}\left(\sum_{i=1}^{n} \mathsf{M}\, x_i^2 + 2\sum_{1\leqslant i<g\leqslant n} \mathsf{M}(x_i x_j)\right) = \\
&= \frac{1}{n}\left(n\,\mathsf{M}\, g^2 + 2\sum_{1\leqslant i<j\leqslant n} \mathsf{M}\, x_i \cdot \mathsf{M}\, x_j\right) = \\
&= \mathsf{M}\, g^2 + \frac{2}{n}\cdot\frac{n(n-1)}{2}(\mathsf{M}\, g)^2 = \\
&= \mathsf{M}\, g^2 + (n-1)(\mathsf{M}\, g)^2.
\end{aligned}$$

Теперь

$$\begin{aligned}
\mathsf{M}\, s_n^2 &= \frac{1}{n-1}\sum_{i=1}^{n} \mathsf{M}\, x_i^2 - \frac{1}{n-1}\,\mathsf{M}\, n\overline{x}^2 = \\
&= \frac{1}{n-1}\cdot n\,\mathsf{M}\, g^2 - \frac{1}{n-1}\left(\mathsf{M}\, g^2 + (n-1)(\mathsf{M}\, g)^2\right) = \\
&= \frac{n}{n-1}\cdot \mathsf{M}\, g^2 - \frac{1}{n-1}\,\mathsf{M}\, g^2 - (\mathsf{M}\, g)^2 = \\
&= \mathsf{M}\, g^2 - (\mathsf{M}\, g)^2 = \mathsf{D}\, g,
\end{aligned}$$

что и нужно.

Состоятельность.

Имеем

$$s_n^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \frac{2}{n(n-1)}\sum_{1\leqslant i<g\leqslant n} x_i x_j.$$

По закону больших чисел для одинаково распределённых случайных величин для любого $\varepsilon > 0$

$$\mathsf{P}\left(\left|\frac{\sum_{i=1}^{n} x_i^2}{n} - \mathsf{M}\, g^2\right| < \frac{\varepsilon}{2}\right) \to 1 \quad \text{при } n \to \infty,$$

также

$$\mathsf{P}\left(\left|\frac{\sum_{1\leqslant i<g\leqslant n} x_i x_j}{\frac{n(n-1)}{2}} - (\mathsf{M}\, g)^2\right| < \frac{\varepsilon}{2}\right) \to 1 \quad \text{при } n \to \infty.$$

Далее

$$\left|s_n^2 - \mathsf{D}\, g\right| =$$

$$= \left| \left(\frac{1}{n} \sum_{i=1}^{n} x_i^2 - \frac{2}{n(n-1)} \sum_{1 \leqslant i < j \leqslant n} x_i x_j \right) - (\mathsf{M}\, g^2 - (\mathsf{M}\, g)^2) \right| =$$

$$= \left| \left(\frac{1}{n} \sum_{i=1}^{n} x_i^2 - \mathsf{M}\, g^2 \right) - \left(\frac{2}{n(n-1)} \sum_{1 \leqslant i < g \leqslant n} x_i x_j - (\mathsf{M}\, g)^2) \right) \right| \leqslant$$

$$\leqslant \left| \frac{1}{n} \sum_{i=1}^{n} x_i^2 - \mathsf{M}\, g^2 \right| + \left| \frac{2}{n(n-1)} \sum_{1 \leqslant i < g \leqslant n} x_i x_j - (\mathsf{M}\, g)^2) \right|.$$

Поэтому

$$\mathsf{P}\left(\left|s_n^2 - \mathsf{D}\, g\right| < \varepsilon\right) \leqslant \mathsf{P}\left(\left|\frac{\sum_{i=1}^{n} x_i^2}{n} - \mathsf{M}\, g^2\right| < \frac{\varepsilon}{2}\right) +$$

$$+ \mathsf{P}\left(\left|\frac{\sum_{1 \leqslant i < g \leqslant n} x_i x_j}{\frac{n(n-1)}{2}} - (\mathsf{M}\, g)^2\right| < \frac{\varepsilon}{2}\right) \to 1 \quad \text{при } n \to \infty.$$

□

7.2. Интервальное оценивание

7.2.1. Предварительные сведения

При интервальном оценивании определяются некоторые вероятности, поэтому мы должны иметь такие случайные величины, которые могли бы дать возможность это сделать. Мы сосредоточимся на оценивании нормального распределения по следующим причинам:

1) хотя бы в силу центральной предельной теоремы ЦПТ) утверждение, что “нормальный закон правит миром”, не является пустым звуком и показывает распространённость нормального распределения;
2) при проведении реальных экспериментов при обработке ошибок измерений считается, что ошибки удовлетворяют нормальным распределениям и крайне важно важно понимать вероятность влияния ошибки на результат;
3) не последнюю роль играет то обстоятельство, что для этого случая разработана хорошая и практичная теория, имеются разнообразные таблицы и т.п.

Мы знаем, что всякое нормальное распределение определяется двумя параметрами — математическим ожиданием и дисперсией. Нам нужно понять, как оцениваются эти параметры теми оценками, что были предложены ранее. Для этого применяются случайные величины, которые пока нам не встречались.

Определение. *Гамма-функцией* называется функция для положительного $x \in \mathbf{R}$, определяемая равенством

$$\Gamma(x) = \int_0^\infty t^{x-1}e^{-t}dt.$$

Известно, что $\Gamma(n) = (n-1)!$ для любого натурального n и вообще справедливо для $x > 0$

$$x\Gamma(x) = \Gamma(x+1). \qquad \text{(уравнение Эйлера)}$$

Также известно, что для любого натурального n

$$\Gamma\left(\tfrac{1}{2}+n\right) = \frac{(2n)!}{4^n n!}\sqrt{\pi}, \quad \text{в частности } \Gamma\left(\tfrac{1}{2}\right) = \sqrt{\pi} \text{ и } \Gamma\left(\tfrac{3}{2}\right) = \frac{\sqrt{\pi}}{2}.$$

Определение. Пусть $g_1,\ldots, g_n$ — независимые случайные величины со стандартным нормальным распределением $\Phi_{0,1}$. Тогда говорят, что случайная величина

$$g_1^2 + \cdots + g_n^2$$

имеет *распределение* $\chi^2 = \chi_n^2$ *с* n *степенями свободы*.

Известно, что для распределения χ_n^2

$$\begin{cases} \dfrac{1}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})} \cdot t^{\frac{n}{2}-1}e^{-\frac{t}{2}} & \text{при } x > 0, \\ 0 & \text{при } x \leqslant 0, \end{cases} \qquad \text{(плотность)}$$

$$\mathsf{M}\,\chi_n^2 = n, \qquad \text{(математическое ожидание)}$$

$$\mathsf{D}\,\chi_n^2 = 2n. \qquad \text{(дисперсия)}$$

Определение. Пусть g_0, $g_1,\ldots, g_n$ — независимые случайные величины со стандартным нормальным распределением $\Phi_{0,1}$. Тогда говорят, что случайная величина

$$\frac{g_0}{\sqrt{\frac{1}{n}(g_1^2 + \cdots + g_n^2)}} = \frac{g_0}{\sqrt{\chi_n^2}} \cdot \sqrt{n}$$

имеет *распределение Стьюдента* $t = t_n$ *с* n *степенями свободы*.

Известно, что для распределения t_n

$$\frac{1}{\sqrt{\pi n}} \cdot \frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n}{2})} \cdot \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}} \quad \text{при любом } x, \qquad \text{(плотность)}$$

$$\mathsf{M}\,t_n^2 = 0, \qquad \text{(математическое ожидание)}$$

$$\mathsf{D}\,t_n^2 = \frac{n}{n-2} \quad \text{при } n > 2. \qquad \text{(дисперсия)}$$

При больших $n \leqslant 30(20)$ распределение t_n близко к стандартному нормальному.

7.2.2. Математическое ожидание

Теорема (интервальное оценивание математического ожидания нормального распределения при известной дисперсии).
Допустим, что известна дисперсия σ^2 нормального распределения. Тогда для любого $\gamma \in (0,1)$ для математического ожидания a

$$\mathsf{P}\left(\overline{x} - \frac{\sigma\varepsilon}{\sqrt{n}} < a < \overline{x} + \frac{\sigma\varepsilon}{\sqrt{n}}\right) = \gamma,$$

где ε находится как решение уравнения

$$\Phi_{0,1}(\varepsilon) = \frac{1+\alpha}{2} \quad \textit{или} \quad \Phi_{\text{Гм}}(\varepsilon) = \frac{\gamma}{2}.$$

Доказательство. Случайная величина $\overline{x}$ имеет нормальное распределение с параметрами $(a, \frac{\sigma^2}{n})$. Поэтому нормированная случайная величина

$$y = \frac{\overline{x} - a}{\sigma}\sqrt{n} = \Phi_{0,1}.$$

Поэтому

$$\mathsf{P}(-\varepsilon < y < \varepsilon) = \Phi_{0,1}(\varepsilon) - \Phi_{0,1}(-\varepsilon) = 2\Phi_{\text{Гм}}(\varepsilon) = \gamma.$$

Так как

$$-\varepsilon < y = \frac{\overline{x} - a}{\sigma}\sqrt{n} < \varepsilon \longleftrightarrow \overline{x} - \frac{\sigma\varepsilon}{\sqrt{n}} < a < \overline{x} + \frac{\sigma\varepsilon}{\sqrt{n}},$$

то получаем требуемое. □

Теорема (интервальное оценивание математического ожидания нормального распределения при неизвестной дисперсии).
Допустим, что неизвестна дисперсия σ^2 нормального распределения. Положим $s = \sqrt{s_n^2}$, где s_n^2 — оценка дисперсии. Тогда для любого $\gamma \in (0,1)$ для математического ожидания a

$$\mathsf{P}\left(\overline{x} - \frac{st_\gamma}{\sqrt{n}} < a < \overline{x} + \frac{st_\gamma}{\sqrt{n}}\right) = \gamma,$$

где t_γ находится как решение уравнения

$$t_{n-1}(t_\gamma) = \gamma.$$

Здесь t_{n-1} — распределение Стьюдента.

Доказательство. Без доказательства примем следующий результат.

Лемма (Фишера).
Пусть $g_1, \dots, g_n$ — независимые случайные величины со стандартным нормальным распределением $\Phi_{0,1}$ и

$$g_0 = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} g_i.$$

Тогда случайные величины g_0 и $\left(\sum_{i=1}^{n} g_i^2\right) - g_0^2$ независимы, и вторая из них имеет распределение χ^2_{n-1}.

Как в доказательстве теоремы об оценивании дисперсии при неизвестном математическом ожидании, имеем

$$\sum_{i=1}^{n}(x_i - \overline{x})^2 = \sum_{i=1}^{n} x_i^2 - n\overline{x}^2 =$$
$$= \sum_{i=1}^{n}(x_i - a)^2 - n(\overline{x} - a)^2 - \left(-2a\sum_{i=1}^{n} x_i + na^2 - (-2n\overline{x} + na^2)\right) =$$
$$= \sum_{i=1}^{n}(x_i - a)^2 - n(\overline{x} - a)^2.$$

Отсюда

$$\frac{(n-1)s_n^2}{\sigma^2} = \sum_{i=1}^{n}\left(\frac{x_i - \overline{x}}{\sigma}\right)^2 = \sum_{i=1}^{n}\left(\frac{x_i - a}{\sigma}\right)^2 - n\left(\frac{\overline{x} - a}{\sigma}\right)^2.$$

Таким образом мы пронормировали и можем применять лемму Фишера для

$$g_0 = \frac{\overline{x} - a}{\sigma},\ g_1 = \frac{x_1 - a}{\sigma}, \dots, g_n = \frac{x_n - a}{\sigma}.$$

Тогда

$$\frac{(n-1)s_n^2}{\sigma^2} = \sum_{i=1}^{n}\left(\frac{x_i - a}{\sigma}\right)^2 - n\left(\frac{\overline{x} - a}{\sigma}\right)^2 = \chi^2_{n-1}$$

и получим, что

$$\frac{\overline{x} - a}{s_n}\sqrt{n} = \frac{\sigma g_0}{s_n} = \frac{g_0}{\sqrt{\frac{(n-1)}{\sigma^2}}} = t_{n-1}$$

— распределение Стьюдента. И всё доказано. □

Замечание. При нахождении t_γ нужно учесть следующее. У Гмурмана в таблицах надо смотреть значения для $1 - \gamma$.

7.2.3. Дисперсия

Теорема (Об интервальном оценивании дисперсии нормального распределения при известном математическом ожидании).
Допустим, что известно математическое ожидание a нормального распределения. Тогда для любого $\gamma \in (0,1)$ для дисперсии σ^2

$$\mathsf{P}\left(\frac{n\sigma_n^2}{y^+} < \sigma^2 < \frac{n\sigma_n^2}{y^-}\right) = \gamma,$$

где y^+ и y^- находятся как решения уравнения

$$\mathsf{P}\left(y^- < \chi_n^2 < y^+\right) = \gamma.$$

Доказательство. Оценкой дисперсии в данном случае служит

$$\sigma_n^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - a)^2.$$

Пронормируем и получим

$$\frac{n\sigma_n^2}{\sigma^2} = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i - a}{\sigma}\right)^2 = \chi_n^2.$$

Далее

$$y^- < \frac{n\sigma_n^2}{\sigma^2} < y^+ \longleftrightarrow \frac{y^-}{n\sigma_n^2} < \frac{1}{\sigma^2} < \frac{y^+}{n\sigma_n^2} \longleftrightarrow \frac{n\sigma_n^2}{y^+} < \sigma^2 < \frac{n\sigma_n^2}{y^-},$$

что и требовалось. □

Замечание. В книгах Гмурмана нет таблицы значений распределений χ_n^2 в виде, который бы похож на таблицы для $\Phi_{\text{ГМ}}$, есть только таблицы так называемых критических точек. Позднее мы обсудим понятие критической точки, а сейчас воспользуемся разъяснениями Колемаева по использованию этих таблиц для наших целей. y^+ можно определить как решение уравнения уравнения

$$F_{\chi_n^2}(y^+) = \delta > \frac{1}{2},$$

а y^- можно определить как решение уравнения уравнения

$$F_{\chi_n^2}(y^-) = 1 - \beta > \frac{1}{2}.$$

Тогда
$$\mathsf{P}\left(y^- < \chi_n^2 < y^+\right) = \gamma = \delta - \beta.$$
Итак надо просто подобрать δ и β так, чтобы $\gamma = \delta - \beta$. Часто берут
$$\delta = \frac{\gamma}{2} \quad \text{и} \quad \beta = 1 - \frac{\gamma}{2}.$$
Следует учесть, что в таблицах Гмурмана приведены значения для
$$\alpha = 1 - \gamma.$$

Теорема (Об интервальном оценивании дисперсии нормального распределения при неизвестном математическом ожидании).
Допустим, что неизвестно математическое ожидание нормального распределения. Тогда для любого $\gamma \in (0,1)$ для дисперсии σ^2
$$\mathsf{P}\left(\max\{s_n(1-q), 0\} < \sigma < s_n(1+q)\right) = \gamma,$$
где q находится как решение уравнения

при $q < 1$
$$\mathsf{P}\left(\frac{\sqrt{n-1}}{1+q} < \chi_{n-1} < \frac{\sqrt{n-1}}{1-q}\right) = \gamma,$$
где $\chi_{n-1} = \sqrt{\chi_{n-1}^2}$;

при $q \geqslant 1$
$$\mathsf{P}\left(\chi_{n-1} > \frac{\sqrt{n-1}}{1+q}\right) = \gamma.$$

Доказательство. Так как ранее отмечено, что
$$\frac{(n-1)s_n^2}{\sigma^2} = \chi_{n-1}^2.$$
Поэтому
$$\frac{s_n}{\sigma}\sqrt{n-1} = \chi_{n-1}.$$
Далее
$$\frac{s_n}{\sigma}\sqrt{n-1} > \frac{\sqrt{n-1}}{1+q} \longleftrightarrow \frac{s_n}{\sigma} > \frac{1}{1+q} \longleftrightarrow \frac{s_n(1+q)}{\sigma} > 1 \longleftrightarrow s_n(1+q) > \sigma.$$
При $q < 1$ неравенство
$$\frac{s_n}{\sigma}\sqrt{n-1} < \frac{\sqrt{n-1}}{1-q}$$
рассматривается аналогично, а при $q \geqslant 1$ оценка становится отрицательной, что не имеет смысла. □

8. Гипотезы. Регрессия

8.1. Гипотезы и их проверка

8.1.1. Выдвижение гипотез

Имея выборку достаточно большого объёма, можно выдвинуть гипотезу о возможном характере неизвестного распределения — нормальное, показательное, равномерное, Пуассона, биномиальное и т.п. Если удаётся получить достаточно убедительное подтверждение гипотезы, то в дальнейшем новые выборки будут уточнять гипотезу и из них будет возможно сразу отсеивать заведомо неверные данные.

Чтобы принять или отвергнуть выдвинутую гипотезу используют следующий подход.

1. Сначала задаётся так называемый уровень значимости $\alpha \in (0,1)$ – это просто значение вероятности, при превышении которого может быть отброшена верная гипотеза, а $1-\alpha$ — вероятность того, что верная гипотеза будет рассматриваться, как не противоречащая данной выборке. Поэтому обычно α берётся маленьким.
2. На основе имеющейся выборки находится некоторое число β.
3. По α находится, так называемое, критическое значение δ. Это число определяется как наибольшее значение, при котором вероятность отвержения нашей гипотезы не превосходит α.
4. Сравниваем β и δ. Если $\beta > \delta$, то это будет обозначать, что произошло событие, вероятность которого меньше α, поэтому у нас недостаточно оснований для принятия гипотезы.

8.1.2. Проверка гипотез по критерию Пирсона χ^2

Рассмотрим проверку гипотезы о нормальном распределении.

1. Пусть выборка задана так:

Значение	$\tilde{x}_1$	$\tilde{x}_2$	$\cdots$	$\tilde{x}_k$
Частота	n_1	n_2	$\cdots$	n_k

и $n = n_1 + \cdots + n_k$.

Если интервальное задание выборки, то в качестве $\tilde{x}_1, \ldots, \tilde{x}_k$ часто берут середины соответствующих интервалов.

2. Находим оценки математического ожидания и дисперсии

$$\overline{x} = \frac{\sum_{i=1}^{k} n_i \tilde{x}_i}{n} \quad \text{и} \quad s_n^2 = \frac{\sum_{i=1}^{k} n_i (\tilde{x}_i - \overline{x})^2}{n-1},$$

и вычислим

$$s = \sqrt{s_n^2}.$$

3. Задаём границы интервалов

$$y_0 = -\infty,\ y_1 = \frac{\tilde{x}_1 + \tilde{x}_2}{2}, \ldots, y_{k-1} = \frac{\tilde{x}_{k-1} + \tilde{x}_k}{2},\ y_k = \infty.$$

При интервальном задании выборки крайние значения "утягиваем" в бесконечности.

4. Нормируем

$$z_0 = -\infty,\ z_1 = \frac{y_1 - \overline{x}}{s}, \ldots, z_{k-1} = \frac{y_{k-1} - \overline{x}}{s},\ z_k = \infty.$$

5. Находим значения

$$\Phi_{0,1}(z_0) = 0,\ \Phi_{0,1}(z_1), \ldots, \Phi_{0,1}(z_{k-1}),\ \Phi_{0,1}(z_k) = 1.$$

Ясно, что если выдвинута гипотеза о другом характере рапределения, то надо считать по-другому.

6. Вычисляем теоретические вероятности для любого $i \in \{1, \ldots, k\}$

$$p_i = \Phi_{0,1}(z_i) - \Phi_{0,1}(z_{i-1}).$$

7. Вычисляем теоретические частоты для любого $i \in \{1, \ldots, k\}$

$$n_i^T = np_i.$$

Предостережение! Ни в коем случае не округляйте их!

8. Находим заключительное число

$$\beta = \chi^2_{\text{набл}} = \sum_{i=1}^{k} \frac{(n_i - n_i^T)^2}{n_i^T}.$$

9. Определим число степеней свободы

$$r = (k-1) - 2.$$

Отнимается от $k-1$ число оцениваемых параметров, для нормального распределения их 2 — математическое ожидание и дисперсия, а для рапределения Пуассона и показательного 1, обозначали λ.

10. Ищем критическое значение по таблице

$$\delta = \chi^2_{\text{кр}}(\alpha, r).$$

11. Если $\chi^2_{\text{набл}} > \chi^2_{\text{кр}}(\alpha, r)$, то гипотеза о нормальном распределении отвергается, иначе считаем, что она не противоречит нашим данным.
Этот вывод основан на том, что

$$\sum_{i=1}^{k} \frac{(n_i - n_i^T)^2}{n_i^T} \quad \text{средняя сумма квадратов ошибок}$$

должно вести себя как распределение χ^2_r.

Замечание. Часто в статистике (например, в изложенном критерии χ^2) используются слова — квантиль и квартиль (они женского рода). Эти слова означают следующее. Пусть есть F — функция распределения некоторой случайной величины и $p \in [0, 1]$. Тогда *p-квантиль* равна

$$x_p = \sup\{x \in \mathbf{R} \mid F(x) \leqslant p\}.$$

Далее

$$\begin{aligned} p &= 0.5 \text{ — медиана},\\ p &= 0.25 \text{ — нижняя квартиль},\\ p &= 0.75 \text{ — верхняя квартиль}. \end{aligned}$$

8.2. Регрессия

Здесь познакомимся с той частью математической статистики (впрочем, во многом, уже покинувшей математическую статистику), которая может иметь в зависимости от обстоятельств (и авторов) различные названия: регрессионный анализ, теория планирования эксперимента, корреляционный анализ и т.д. Не вдаваясь в дискуссию о том, что и как правильно называть мы отметим, что во всех этих науках есть нечто общее. А именно, есть несколько групп данных и ставится задача о том, чтобы понять, как связаны между собой эти группы данные при условии, что заведомо известно, что данные могут быть не совсем точными, и/или не учитывающими некоторые другие факторы. Что значит понять? Надо построить какую функцию, которая связывала бы эти группы данных и давала бы картину поведения в тех случаях, когда мы не можем измерить или хотим знать предполагаемый ответ для новых данных.

Регрессия означает движение вспять, обращение (антоним прогрессии). Для нас это означает, что мы делаем выводы на основе *уже имеющихся* данных, то есть на основе прошлого. Но вот парадокс! Чаще всего регрессия применяется для предсказаний, то есть для будущего.

Впервые слово регрессия употребил Фрэнсис Гальтон в статье *"Регрессия в наследовании роста"* в Журнале антропологического института. Он исследовал влияние роста родителей на рост ребёнка. Пусть x — средний рост двух родителей, $\overline{x}$ — средний рост всех исследованных родителей, $\overline{y}$ — средний рост исследованных детей. Тогда предсказываемый рост ребёнка

$$\hat{y} = \overline{y} + \frac{2}{3}(x - \overline{x}).$$

Этот вывод был сделан на основе исследования роста 930 (по другим данным 928) детей.

Менделеев в своих исследованиях тоже применял регрессию.

8.2.1. Основные понятия

Ограничимся двумя параметрами x и y и допустим, что имеем результаты наблюдений

$$\begin{array}{c|c|c|c} x_1 & x_2 & \cdots & x_n \\ \hline y_1 & y_2 & \cdots & y_n \end{array}.$$

Например, есть некоторый технологический процесс (плавка стали, сборка машин и т.п.). Хотим понять, как один из параметров процесса влияет на определённую характеристику конечного продукта. Пусть

$$\begin{aligned} x &— \text{изучаемый параметр,} \\ y &— \text{характеристика конечного продукта,} \\ \hat{y} = f(x) &— \text{прямая регрессия,} \\ \hat{x} = \varphi(y) &— \text{обратная регрессия.} \end{aligned}$$

С помощью прямой регрессии предсказывается поведение характеристики конечного продукта в зависимости от изучаемого параметра, а с помощью обратной можно указывать каким должен быть этот параметр при заданной характеристике.

Функции f и φ могут быть самыми разнообразными. Наиболее популярны *многочлены*, но это может быть опасно, так как при высоких степенях они ведут себя непредсказуемо, сильно искажают результаты, неустойчивы (малая подвижка аргумента приводит к очень сильному сдвигу значения функции). Можно также использовать рациональные функции (частное двух многочленов), многочлены Лорана (допускаются отрицательные степени аргумента) и ещё очень много хороших функций.

Самая популярная — линейная регрессия, когда f и φ берутся линейными, то есть

$$\hat{y} = a + bx \quad \text{и} \quad \hat{x} = a_1 + b_1 y.$$

Скачков не будет, хотя есть недостатки, например, слабо отражает ситуацию в случае сильного разброса значений.

Более точно считаем (почему отдельный вопрос), что наша модель (почти) линейна

$$y = \alpha + \beta x + \varepsilon,$$

где ε – некоторая добавка либо из-за неточности наблюдений, либо из-за отбрасывания других факторов.

Основная цель состоит в том, чтобы регрессия как можно меньше отличалась от предложенной модели.

Аналогичные подходы применяются и для обратной регрессии.

8.2.2. Построение линейной регрессии

Для построения линейной регрессии будет использоваться — Метод Наименьших Квадратов (МНК). Историческая справка: Гаусс использовал в 1803 году (сам говорил, что знал с 1795 года), в 1805 году было сообщение Лежандра (в 1809 году Гаусс сказал, что знал раньше), обычно считается, что термин и употребление МНК началось с 1808 года с Эдрейна (Adrain). Надо сказать, что МНК применим очень широко ко многим моделям.

Обозначения. Используя введённые ранее обозначения, положим

$$S_{xy} = \sum_{i=1}^{n} (x_i - \overline{x})(y_i - \overline{y}).$$

Для $y = x$

$$S_{xx} = \sum_{i=1}^{n} (x_i - \overline{x})^2,$$

для $x = y$

$$S_{yy} = \sum_{i=1}^{n} (y_i - \overline{y})^2.$$

Теорема (о линейной регрессии).
При введённых ранее обозначениях имеем для прямой регрессии

$$\hat{y} = \overline{y} + b(x - \overline{x}), \text{ где } b = \frac{S_{xy}}{S_{xx}},$$

для обратной регрессии

$$\hat{x} = \overline{x} + b_1(y - \overline{y}), \text{ где } b_1 = \frac{S_{yx}}{S_{yy}}.$$

Доказательство. Для любого $i \in \{1, \dots, n\}$ имеем

$$y_i = \alpha + \beta x_i + \varepsilon_i.$$

Пусть

$$f(\alpha, \beta) = \sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} (y_i - \alpha - \beta x_i)^2 \geqslant 0.$$

Так как функция $f(\alpha, \beta)$ непрерывна и ограничена снизу (нулём), то она имеет наименьшее значение, которое будем искать (теперь становиться понятным термин — метод наименьших квадратов).

Для этого по необходимому условию экстремальности надо найти частные производные по α и β и приравнять их 0.

$$\begin{cases} \dfrac{\partial f}{\partial \alpha} = -2 \displaystyle\sum_{i=1}^{n} (y_i - \alpha - \beta x_i) = 0, \\ \dfrac{\partial f}{\partial \beta} = -2 \displaystyle\sum_{i=1}^{n} x_i (y_i - \alpha - \beta x_i) = 0. \end{cases}$$

Решим эту систему

$$\begin{cases} \displaystyle\sum_{i=1}^{n} (y_i - \alpha - \beta x_i) = 0, \\ \displaystyle\sum_{i=1}^{n} x_i (y_i - \alpha - \beta x_i) = 0, \end{cases} \longleftrightarrow \begin{cases} \displaystyle\sum_{i=1}^{n} y_i = n\alpha + \beta \sum_{i=1}^{n} x_i, \\ \displaystyle\sum_{i=1}^{n} x_i y_i = \alpha \sum_{i=1}^{n} x_i + \beta \sum_{i=1}^{n} x_i^2. \end{cases}$$

Полученные уравнения называют *нормальными уравнениями регрессии*. Теперь разделим оба уравнения на n и, вспомнив определения $\overline{x}$) и $\overline{y}$, получим

$$\begin{cases} \overline{y} = \alpha + \beta \overline{x}, \\ \dfrac{\sum_{i=1}^{n} x_i y_i}{n} = \alpha \overline{x} + \beta \dfrac{\sum_{i=1}^{n} x_i^2}{n}, \end{cases} \longleftrightarrow$$

$$\longleftrightarrow \begin{cases} \alpha = \overline{y} - \beta \overline{x}, \\ \dfrac{\sum_{i=1}^{n} x_i y_i}{n} = \overline{x}(\overline{y} - \beta \overline{x}) + \beta \dfrac{\sum_{i=1}^{n} x_i^2}{n}, \end{cases} \longleftrightarrow$$

$$\longleftrightarrow \begin{cases} \alpha = \overline{y} - \beta \overline{x}, \\ \dfrac{\sum_{i=1}^{n} x_i y_i}{n} = \overline{x} \cdot \overline{y} + \beta \left(\dfrac{\sum_{i=1}^{n} x_i^2}{n} - \overline{x}^2 \right), \end{cases} \longleftrightarrow$$

$$\begin{cases} \alpha = \overline{y} - \beta\overline{x}, \\ \beta = \dfrac{\frac{\sum_{i=1}^n x_i y_i}{n} - \overline{x}\cdot\overline{y}}{\frac{\sum_{i=1}^n x_i^2}{n} - \overline{x}^2} = \dfrac{\sum_{i=1}^n x_i y_i - n\overline{x}\cdot\overline{y}}{\sum_{i=1}^n x_i^2 - n\overline{x}^2}. \end{cases}$$

Преобразуем последнюю дробь. Для числителя

$$\begin{aligned} \sum_{i=1}^n x_i y_i - n\overline{x}\cdot\overline{y} &= \sum_{i=1}^n x_i y_i - (n\overline{x})\cdot\overline{y} - \overline{x}(n\overline{y}) + n\overline{x}\cdot\overline{y} = \\ &= \sum_{i=1}^n x_i y_i - (\sum_{i=1}^n x_i)\cdot\overline{y} - \overline{x}(\sum_{i=1}^n y_i) + n\overline{x}\cdot\overline{y} = \\ &= \sum_{i=1}^n x_i(y_i - \overline{y}) - \overline{x}\sum_{i=1}^n (y_i - \overline{y}) = \sum_{i=1}^n (x_i - \overline{x})(y_i - \overline{y}) = S_{xy}. \end{aligned}$$

Положим $y_i = x_i$ и получим знаменатель

$$\sum_{i=1}^n x_i^2 i - n\overline{x}^2 \sum_{i=1}^n (x_i - \overline{x})^2 = S_{xx}.$$

Таким образом получаем для наименьшего значения

$$\begin{aligned} \beta &= \frac{S_{xy}}{S_{xx}} = b, \\ \alpha &= \overline{y} - b\overline{x} = a. \end{aligned}$$

Теперь для уравнения прямой регрессии имеем

$$\hat{y} = a + bx = (\overline{y} - b\overline{x}) + bx = \overline{y} + b(\overline{y} - \overline{x}).$$

Чтобы получить уравнение обратной регрессии нужно только поменять x и y. □

Следствие. *В обозначениях теоремы*

$$bb_1 = \frac{S_{xy}^2}{S_{xx}S_{yy}}$$

при больших n может служить оценкой квадрата коэффициента корреляции между x и y.

Доказательство. Не будем вдаваться в полное обоснование высказанного утверждения, а укажем только основные моменты. Мы знаем, что при больших n

$$\frac{S_{xx}}{n} = \frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n} = \frac{n-1}{n}s_n^2 \approx s_n^2.$$

Поэтому можно считать, что $\frac{S_{xx}}{n}$ оценивает дисперсию x.

Аналогично $\frac{S_{yy}}{n}$ оценивает дисперсию y, и $\frac{S_{xy}}{n}$ оценивает ковариацию x и y. Поэтому

$$bb_1 = \frac{S_{xy}^2}{S_{xx}S_{yy}} = \frac{\left(\frac{S_{xy}}{n}\right)^2}{\frac{S_{xx}^2}{n} \cdot \frac{S_{yy}^2}{n}}$$

можно считать, что оценивается

$$\frac{\operatorname{cov}(x,y)^2}{\mathsf{D}\,x \cdot \mathsf{D}\,y} = \left(\frac{\operatorname{cov}(x,y)}{\sqrt{\mathsf{D}\,x} \cdot \sqrt{\mathsf{D}\,y}}\right)^2 = \rho(x,y)^2$$

— квадрат коэффициента корреляции. □

8.2.3. Точность регрессии

Здесь мы покажем, что отклонение регрессии от истинных значений ограничено в терминах только последних.

Сначала заметим, что для любого $i \in \{1, \ldots, n\}$

$$\begin{aligned}
&\hat{y}_i - \overline{y} = \overline{y} + b(x_i - \overline{x}) - \overline{y} = b(x_i - \overline{x}),\\
&y_i - \hat{y}_i = y_i - \overline{y} - b(x_i - \overline{x}) = (y_i - \overline{y}) - b(x_i - \overline{x}),\\
&y_i - \overline{y} = (\hat{y}_i - \overline{y}) + (y_i - \hat{y}_i).
\end{aligned}$$

Теперь

$$\begin{aligned}
\sum_{i=1}^{n}(y_i - \overline{y})^2 &= \sum_{i=1}^{n}\left((\hat{y}_i - \overline{y}) + (y_i - \hat{y}_i)\right)^2 = \\
&= \sum_{i=1}^{n}(\hat{y}_i - \overline{y})^2 + \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 + 2\sum_{i=1}^{n}(\hat{y}_i - \overline{y})(y_i - \hat{y}_i).
\end{aligned}$$

К последней сумме применим полученные ранее соотношения

$$\begin{aligned}
&\sum_{i=1}^{n}(\hat{y}_i - \overline{y})(y_i - \hat{y}_i) = \sum_{i=1}^{n} b(x_i - \overline{x})\left((y_i - \overline{y}) - b(x_i - \overline{x})\right) = \\
&= b\left(\sum_{i=1}^{n}\left((x_i - \overline{x})(y_i - \overline{y}\right) - b\sum_{i=1}^{n}(x_i - \overline{x})^2\right) = b\left(S_{xy} - \frac{S_{xy}}{S_{xx}} \cdot S_{xx}\right) = 0.
\end{aligned}$$

Таким образом
$$\sum_{i=1}^{n}(y_i-\overline{y})^2=\sum_{i=1}^{n}(\hat{y}_i-\overline{y})^2+\sum_{i=1}^{n}(y_i-\hat{y}_i)^2.$$

Следовательно отклонение регрессии от истинных значений равно
$$\sum_{i=1}^{n}(y_i-\hat{y}_i)^2=\sum_{i=1}^{n}(y_i-\overline{y})^2-\sum_{i=1}^{n}(\hat{y}_i-\overline{y})^2\leqslant\sum_{i=1}^{n}(y_i-\overline{y})^2,$$
что и хотели.

8.2.4. Дисперсии коэффициентов регрессии

Будем искать $\mathsf{D}\,b$, считая $b=\frac{S_{xy}}{S_{xx}}$, зависящей только от y, что естественно, ибо оцениваем y через x. Сначала заметим, что
$$S_{xy}=\sum_{i=1}^{n}((x_i-\overline{x})(y_i-\overline{y})=\sum_{i=1}^{n}(x_i-\overline{x})y_i-\overline{y}\sum_{i=1}^{n}(x_i-\overline{x})=\sum_{i=1}^{n}(x_i-\overline{x})y_i,$$
поскольку
$$\sum_{i=1}^{n}(x_i-\overline{x})=\sum_{i=1}^{n}x_i-n\overline{x}=n\overline{x}-n\overline{x}=0.$$
Отсюда
$$\mathsf{D}\,S_{xy}=\mathsf{D}\sum_{i=1}^{n}(x_i-\overline{x})y_i=\sum_{i=1}^{n}(x_i-\overline{x})^2\,\mathsf{D}\,y=S_{xx}\,\mathsf{D}\,y.$$
и
$$\mathsf{D}\,b=\mathsf{D}\,\frac{S_{xy}}{S_{xx}}=\frac{1}{S_{xx}^2}\,\mathsf{D}\,S_{xy}=\frac{S_{xx}\,\mathsf{D}\,y}{S_{xx}^2}=\frac{\mathsf{D}\,y}{S_{xx}}.$$

Далее
$$\mathsf{D}\,\overline{y}=\mathsf{D}\left(\frac{\sum_{i=1}^{n}y_i}{n}\right)=\frac{1}{n^2}\cdot n\,\mathsf{D}\,y=\frac{\mathsf{D}\,y}{n}.$$

Теперь
$$\mathsf{D}\,a=\mathsf{D}(\overline{y}-b\overline{x})=\mathsf{D}\,\overline{y}+\overline{x}^2\,\mathsf{D}\,b=\frac{\mathsf{D}\,y}{n}+\overline{x}^2\frac{\mathsf{D}\,y}{S_{xx}}=\mathsf{D}\,y\left(\frac{\overline{x}^2}{S_{xx}}+\frac{1}{n}\right).$$

Для выражения в скобках числитель
$$S_{xx}+n\overline{x}^2=\sum_{i=1}^{n}(x_i-\overline{x})^2+n\overline{x}^2=\sum_{i=1}^{n}x_i(x_i-\overline{x})-\overline{x}\sum_{i=1}^{n}(x_i-\overline{x})+n\overline{x}^2=$$

$$= \sum_{i=1}^{n} x_i(x_i - \overline{x}) + n\overline{x}^2 = \sum_{i=1}^{n} x_i^2 - \overline{x}\sum_{i=1}^{n} x_i + n\overline{x}^2 = \sum_{i=1}^{n} x_i^2.$$

Поэтому

$$\mathsf{D}\,a = \mathsf{D}\,y \cdot \frac{\sum_{i=1}^{n} x_i^2}{nS_{xx}}.$$

8.2.5. Корреляция коэффициентов регрессии

Найдём коэффициент корреляции между b и $\overline{y}$. Известно, что

$$\rho(b,\overline{y}) = \frac{\mathsf{cov}(b,\overline{y})}{\sqrt{\mathsf{D}\,b}\cdot\sqrt{\mathsf{D}\,\overline{y}}} \quad \text{и} \quad \mathsf{cov}(b,\overline{y}) = \mathsf{M}(b\overline{y}) - \mathsf{M}\,b\cdot\mathsf{M}\,\overline{y}.$$

Имеем

$$\mathsf{M}\,\overline{y} = \mathsf{M}\,\frac{\sum_{i=1}^{n} y_i}{n} = \frac{1}{n}\cdot\mathsf{M}\,y = \mathsf{M}\,y.$$

и

$$\mathsf{M}\,b = \mathsf{M}\,\frac{S_{xy}}{S_{xx}} = \frac{1}{S_{xx}}\,\mathsf{M}\,S_{xy} = \frac{1}{S_{xx}}\,\mathsf{M}\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y}) =$$

$$= \frac{1}{S_{xx}}\sum_{i=1}^{n}(x_i - \overline{x})\,\mathsf{M}(y_i - \overline{y}) = \frac{1}{S_{xx}}\sum_{i=1}^{n}(x_i - \overline{x})(\mathsf{M}\,y_i - \mathsf{M}\,\overline{y}) =$$

$$= \frac{1}{S_{xx}}\sum_{i=1}^{n}(x_i - \overline{x})(\mathsf{M}\,y - \mathsf{M}\,y) = 0.$$

Следовательно

$$\mathsf{cov}(b,\overline{y}) = \mathsf{M}(b\overline{y})$$

Теперь

$$b\overline{y} = \frac{1}{S_{xx}}\left(\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})\right)\overline{y} = \frac{1}{S_{xx}}\sum_{i=1}^{n}(x_i - \overline{x})(y_i\overline{y} - \overline{y}^2) =$$

$$= \frac{1}{S_{xx}}\left(\sum_{i=1}^{n}(x_i - \overline{x})y_i\overline{y} - \overline{y}^2\sum_{i=1}^{n}(x_i - \overline{x})\right) = \frac{1}{S_{xx}}\sum_{i=1}^{n}(x_i - \overline{x})y_i\overline{y}.$$

и

$$\mathsf{M}(b\overline{y}) = \frac{1}{S_{xx}}\sum_{i=1}^{n}(x_i - \overline{x})(\mathsf{M}(y_i\overline{y}).$$

Далее для любого $i \in \{1,\ldots,n\}$ $\mathsf{M}(y_i\overline{y}) = c$ не зависит от i.

Отсюда

$$\mathsf{M}(b\overline{y}) = \frac{1}{S_{xx}} \sum_{i=1}^{n} (x_i - \overline{x})c = 0.$$

Таким образом

$$\mathsf{cov}(b, \overline{y}) = \mathsf{M}(b\overline{y}) = 0 \quad \text{и} \quad \rho(b, \overline{y}) = 0.$$

8.3. Дисперсионный анализ

Допустим, что есть p неизвестных нормальных величин с *одинаковой* (может быть неизвестной) дисперсией. Спрашивается равны ли их математические ожидания? Это задача однофакторного дисперсионного анализа.

Пусть проводится q серий испытаний по p штук в каждом (каждая случайная величина испытывается ровно 1 раз в каждом испытании), всего $n = pq$ данных. Пусть x_{ij} — результат для j-той случайной величины при i-том испытании. Пусть

$$\overline{x} = \frac{\sum_{i,j} x_{ij}}{n} \text{ — общее среднее,}$$

$$\overline{x}_j = \frac{\sum_{i=1}^{q} x_{ij}}{q} \text{ — среднее для } j\text{-той случайной величины, } j \in \{1, \ldots, p\}.$$

Вычисляются следующие три числа:

$$S_{\text{общ}} = \sum_{j=1}^{p} \sum_{i=1}^{q} (x_{ij} - \overline{x})^2 \ (S \text{ общее})$$

характеризует общий разброс,

$$S_{\text{факт}} = q \sum_{j=1}^{p} (\overline{x}_j - \overline{x})^2 \ (S \text{ факторное})$$

характеризует разброс между случайными величинами,

$$S_{\text{ост}} = \sum_{j=1}^{p} \sum_{i=1}^{q} (x_{ij} - \overline{x}_j)^2 \ (S \text{ остаточное})$$

характеризует сумму разбросов между значениями случайной величины.

Заметим, что

$$S_{\text{общ}} = \sum_{j=1}^{p} \sum_{i=1}^{q} (x_{ij} - \overline{x})^2 = \sum_{j=1}^{p} \sum_{i=1}^{q} \left((x_{ij} - \overline{x}_j) + (\overline{x}_j - \overline{x})\right)^2 =$$

$$= \sum_{j=1}^{p}\sum_{i=1}^{q}(x_{ij}-\overline{x}_j)^2 + \sum_{j=1}^{p}\sum_{i=1}^{q}(\overline{x}_j-\overline{x})^2 + 2\sum_{j=1}^{p}\sum_{i=1}^{q}(x_{ij}-\overline{x}_j)(\overline{x}_j-\overline{x}) =$$

$$= S_{\text{ост}} + S_{\text{факт}} + 2\sum_{j=1}^{p}(\overline{x}_j-\overline{x})\left(\sum_{i=1}^{q}(x_{ij}-\overline{x}_j)\right) = S_{\text{ост}} + S_{\text{факт}},$$

поскольку

$$\sum_{i=1}^{q}(x_{ij}-\overline{x}_j) = \sum_{i=1}^{q}x_{ij} - q\overline{x}_j = q\overline{x}_j - q\overline{x}_j = 0.$$

Далее вычисляются:

$s^2_{\text{общ}} = \dfrac{S_{\text{общ}}}{pq-1}$ — оценка общей дисперсии,

$s^2_{\text{факт}} = \dfrac{S_{\text{факт}}}{p-1}$ — характеризует средний разброс между случ. величинами,

$s^2_{\text{ост}} = \dfrac{S_{\text{ост}}}{p(q-1)}$— оценка суммы дисперсий случайных величин.

Показывается, что

$$F_{\text{набл}} = \frac{s^2_{\text{факт}}}{s^2_{\text{ост}}}$$

распределено как

$$\frac{\frac{\chi^2_{p-1}}{p-1}}{\frac{\chi^2_{p(q-1)}}{p(q-1)}} = Z[p-1, p(q-1)].$$

Распределение

$$Z[n,m] = \frac{\frac{\chi^2_n}{n}}{\frac{\chi^2_m}{m}}$$

называется распределением *Фишера-Снедекора*. Известно, что для распределения $Z[n,m]$

$$\frac{\Gamma(\frac{n+m}{2})\cdot}{\Gamma(\frac{n}{2})\cdot\Gamma(\frac{m}{2})}\left(\frac{n}{m}\right)^{\frac{n}{2}\cdot\frac{m}{2}-1} x\left(1+\frac{n}{m}x\right)^{-\frac{n+m}{2}}, \ x>0, \qquad \text{(плотность)}$$

$$\mathsf{M}\,Z[n,m] = \frac{m}{m-2} \text{ при } m>2, \qquad \text{(математическое ожидание)}$$

$$\mathsf{D}\,Z[n,m] = \frac{2m^2(n+m-1)}{n(m-2)^2(m-4)} \text{ при } m>4. \qquad \text{(дисперсия)}$$

Далее рассуждения как в случае χ^2-критерия Пирсона. Находим критическую точку и если $F_{\text{набл}} > F_{\text{крит}}$, то гипотеза о равенстве математических ожиданий отвергается.

8.4. Контрольное задание по математической статистике

В этом задании студенту предлагается выполнить следующее. Заданы опытные данные.

Часть 1. Гипотеза и её проверка

- найти оценки для математического ожидания и дисперсии;
- выдвинуть гипотезу о законе распределения случайной величины;
- вычислить гипотетические частоты;
- пользуясь критерием согласия χ^2, установить, соответствуют ли опытные данные предположению о распределении случайной величины по избранному гипотетическому закону с уровнем значимости равным $0,05$.

Часть 2. Регрессия

- построить прямую и обратную линейные регрессии,
- оценить коэффициент корреляции,
- найти предсказываемые значения,
- отклонение предсказываемых значения от заданных.

Пример выполнения контрольного задания по математической статистике

Задача (Данные задачи № 459 из Гмурмана).
Результаты измерения роста 100 *подростков:*

Рост	154–158	158–162	162–166	166–170	170-174	174-178	178-182
Число	10	14	26	28	12	8	2

Решение 1 части (Гипотеза и её проверка).

1. Так как данные заданы интервально, то надо перейти к точечному заданию, взяв середины интервалов и получим:

Рост	156	160	164	168	172	176	180
Число	10	14	26	28	12	8	2

2. Находим оценки для математического ожидания и дисперсии.

Оценка для математического ожидания.

$$\overline{x} = \frac{156\cdot10+160\cdot14+164\cdot26+168\cdot28+172\cdot12+176\cdot8+180\cdot2}{100} =$$
$$= \frac{1560 + 2240 + 4264 + 4704 + 2064 + 1408 + 360}{100} =$$
$$= \frac{16600}{100} = 166.$$

Оценка для дисперсии. Удобно и полезно отнять от данных роста полученную оценку математического ожидания и получим:

Рост $-\ \overline{x}$	-10	-6	-2	2	6	10	14
Число	10	14	26	28	12	8	2

Тогда оценка для дисперсии

$$s^2_{100} = \frac{(-10)^2\cdot10+(-6)^2\cdot14+(-2)^2\cdot26+2^2\cdot28+6^2\cdot12+10^2\cdot8+14^2\cdot2}{100-1} =$$
$$= \frac{100(10+8)+36(14+12)+4(26+28)+196\cdot 2}{99} =$$
$$= \frac{1800+936+216+392}{99} = \frac{3344}{99} = \frac{304}{9} = \frac{16}{9}\cdot 19 =$$
$$= 33,7777777777777777777777777778.$$

Замечание. Для получения правдоподобных результатов при вычислениях надо *всегда* оставлять не менее 4 знаков после запятой.

Мы оставим 6 знаков. Итак

$$s^2_{100} = 33,777778; \quad s = \sqrt{s^2_{100}} = \frac{4}{3}\sqrt{19} = 5,811865.$$

Замечание. Посмотрим, что получиться, если использовать смещённую оценку дисперсии.

$$\tilde{s}^2_{100} = \frac{3344}{100} = \frac{16\cdot 11\cdot 19}{100} = 33,44; \quad \sqrt{\tilde{s}^2_{100}} = \frac{2}{5}\sqrt{209} = 5,782733.$$

Разница есть!

3. **Выдвижение гипотезы** Поскольку значения частот на промежутках одинаковой длины существенно различаются, это распределение НЕ будет равномерным.

   Так как $\overline{x}$ и s^2_{100} существенно различаются, это распределение НЕ будет распределением Пуассона.

   Наконец, существенное различие $\overline{x}$ и s влечёт, что это распределение НЕ будет показательным.

   Поэтому выдвигаем следующую гипотезу.

Гипотеза о нормальном распределении.

Более точно, о нормальном распределении с параметрами

$$a = \overline{x} = 166 \quad \text{и} \quad \sigma^2 = s_{100}^2 = \frac{16}{9} \cdot 19 = 33,777778,$$

то есть о распределении

$$\Phi_{166, \frac{16}{9} \cdot 19}.$$

4. Нужно задать границы интервалов. Так как в начале имели интервальное задание, то нужно только “утянуть”крайние значения в бесконечность. Таким образом,

$$y_0 = -\infty,\ y_1 = 158,\ y_2 = 162,\ y_3 = 166,\ y_4 = 170,$$
$$y_5 = 174,\ y_6 = 178,\ y_7 = \infty.$$

5. Нормируем по правилу $y \longrightarrow z = \dfrac{y - \overline{x}}{s}$:

$$\begin{aligned}
z_0 &= -\infty, \\
z_1 &= \frac{158 - 166}{\frac{4}{3}\sqrt{19}} = -\frac{3 \cdot 8}{4\sqrt{19}} = -\frac{6}{\sqrt{19}} = -1,376494, \\
z_2 &= \frac{162 - 166}{\frac{4}{3}\sqrt{19}} = -\frac{3 \cdot 4}{4\sqrt{19}} = -\frac{3}{\sqrt{19}} = -0,688247, \\
z_3 &= 0, \\
z_4 &= \frac{170 - 166}{\frac{4}{3}\sqrt{19}} = \frac{3 \cdot 4}{4\sqrt{19}} = \frac{3}{\sqrt{19}} = 0,688247, \\
z_5 &= \frac{174 - 166}{\frac{4}{3}\sqrt{19}} = \frac{3 \cdot 8}{4\sqrt{19}} = \frac{6}{\sqrt{19}} = 1,376494, \\
z_6 &= \frac{178 - 166}{\frac{4}{3}\sqrt{19}} = \frac{3 \cdot 12}{4\sqrt{19}} = \frac{9}{\sqrt{19}} = 2,064742, \\
z_7 &= \infty.
\end{aligned}$$

6. Можно находить значения $\Phi_{0,1}$ по таблице из Гмурмана (ПРИЛОЖЕНИЕ 3) по правилу

$$\Phi_{0,1}(z) = \begin{cases} 0,5 - \Phi_{\text{Гм}} & \text{при } z < 0; \\ 0,5 + \Phi_{\text{Гм}} & \text{при } z \geqslant 0; \end{cases}$$

округляя и беря средние значения. Мы же воспользуемся **Excel** — статистическая функция НОРМ.СТ.РАСП.

$$\Phi_{0,1}(z_0) = 0; \qquad \Phi_{0,1}(z_1) = 0,084334;$$

$$\Phi_{0,1}(z_2) = 0,245649; \qquad \Phi_{0,1}(z_3) = 0,5;$$
$$\Phi_{0,1}(z_4) = 0,754351; \qquad \Phi_{0,1}(z_5) = 0,915666;$$
$$\Phi_{0,1}(z_6) = 0,980526; \qquad \Phi_{0,1}(z_7) = 1.$$

7. Вычисляем теоретические вероятности

$$\begin{aligned}
p_1 &= \Phi_{0,1}(z_1) - \Phi_{0,1}(z_0) = 0,084334 - 0 = 0,084334;\\
p_2 &= \Phi_{0,1}(z_2) - \Phi_{0,1}(z_1) = 0,245649 - 0,084334 = 0,161315;\\
p_3 &= \Phi_{0,1}(z_3) - \Phi_{0,1}(z_2) = 0,5 - 0,245649 = 0,254351;\\
p_4 &= \Phi_{0,1}(z_4) - \Phi_{0,1}(z_3) = 0,754351 - 0,5 = 0,254351;\\
p_5 &= \Phi_{0,1}(z_5) - \Phi_{0,1}(z_4) = 0,915666 - 0,754351 = 0,161315;\\
p_6 &= \Phi_{0,1}(z_6) - \Phi_{0,1}(z_5) = 0,980526 - 0,915666 = 0,06486;\\
p_7 &= \Phi_{0,1}(z_7) - \Phi_{0,1}(z_6) = 1 - 0,980526 = 0,019474.
\end{aligned}$$

8. Вычисляем теоретические частоты по формуле $n_i = 100 \cdot p_i$

$$\begin{aligned}
n_1^T &= 100 \cdot 0,084334 = 8,4334;\\
n_2^T &= 100 \cdot 0,161315 = 16,1315;\\
n_3^T &= 100 \cdot 0,254351 = 25,4351;\\
n_4^T &= 100 \cdot 0,254351 = 25,4351;\\
n_5^T &= 100 \cdot 0,161315 = 16,1315;\\
n_6^T &= 100 \cdot 0,06486 = 6,486;\\
n_7^T &= 100 \cdot 0,019474 = 1,9474.
\end{aligned}$$

Контроль

$$\begin{aligned}
&n_1^T + n_2^T + n_3^T + n_4^T + n_5^T + n_6^T + n_7^T =\\
&\quad = 8,4334 + 2 \cdot 16,1315 + 2 \cdot 25,4351 + 6,486 + 1,9474 = 100.
\end{aligned}$$

Предупреждение! *Ни в коем случае не округляйте до целых теоретические частоты!*

9. Вычисляем $\chi^2_{\text{набл}}$ (хи квадрат наблюдаемое)

$$\begin{aligned}
\chi^2_{\text{набл}} &= \sum_{i=1}^{7} \frac{(n_i - n_i^T)^2}{n_i^T} = \frac{(10 - 8,4334)^2}{8,4334} + \frac{(14 - 16,1315)^2}{16,1315} +\\
&+ \frac{(26 - 25,4351)^2}{25,4351} + \frac{(28 - 25,4351)^2}{25,4351} + \frac{(12 - 16,1315)^2}{16,1315} +\\
&+ \frac{(8 - 6,486)^2}{6,486} + \frac{(2 - 1,9474)^2}{1,9474} =
\end{aligned}$$

$$= \frac{2,454236}{8,4334} + \frac{4,543292 + 17,069292}{16,1315} +$$
$$+ \frac{0,319112 + 6,578712}{25,4351} =$$
$$= \frac{2,454236}{8,4334} + \frac{21,612584}{16,1315} + \frac{6,897824}{25,4351} =$$
$$= 0,291014 + 1,339775 + 0,271193 = 1,901982.$$

10. Число степеней свободы

$$r = (7-1) - 2 = 4.$$

11. Ищем критическое значение по таблице (Гмурман, ПРИЛОЖЕНИЕ 4)

$$\chi^2_{\text{кр}}(0,05;4) = 9,5.$$

12. Так как

$$\chi^2_{\text{набл}} = 1,901982 < 9,5 = \chi^2_{\text{кр}}(0,05;4),$$

то гипотеза о нормальном распределении с параметрами

$$a = \overline{x} = 166 \quad \text{и} \quad \sigma^2 = s^2_{100} = \frac{16}{9} \cdot 19 = 33,777778,$$

то есть о распределении

$$\Phi_{166,\frac{16}{9}\cdot 19},$$

НЕ противоречит исходным данным.

$\square$

Решение 2 части (Регрессия).

Нужно пользоваться таблицей, полученной при переходе к точечным данным:

x	156	160	164	168	172	176	180
y	10	14	26	28	12	8	2

1. Ищем средние

$$\overline{x} = \frac{156 + 160 + 164 + 168 + 172 + 176 + 180}{7} = \frac{1176}{7} = 168,$$
$$\overline{y} = \frac{10 + 14 + 26 + 28 + 12 + 8 + 2}{7} = \frac{100}{7} = 14,285714.$$

Предупреждение! *Ни в коем случае не путайте со средним $\overline{x}$ из Части 1*

2. Удобно составить таблицу, в которой от исходных данных отняты средние

$x-\overline{x}$	-12	-8	-4	0	4	8	12
$y-\overline{y}$	$-\frac{30}{7}$	$-\frac{2}{7}$	$\frac{82}{7}$	$\frac{96}{7}$	$-\frac{16}{7}$	$-\frac{44}{7}$	$-\frac{86}{7}$

Контроль

$$\begin{aligned}\sum(x_i-\overline{x}) &= -12-8-4+0+4+8+12=0,\\ \sum(y_i-\overline{y}) &= -\frac{30}{7}-\frac{2}{7}+\frac{82}{7}+\frac{96}{7}-\frac{16}{7}-\frac{44}{7}-\frac{86}{7}=\\ &=\frac{-30-2+82+96-16-44-86}{7}=\\ &=\frac{(82+96)-(30+2+16+44+86)}{7}=\\ &=\frac{178-178}{7}=0.\end{aligned}$$

3. Считаем суммы

$$\begin{aligned}S_{xx} &= \sum(x_i-\overline{x})^2 = 12^2+8^2+4^2+0+4^4+8^2+12^2=\\ &=2(144+64+16)=2\cdot 224=448,\\ S_{yy} &= \sum(y_i-\overline{y})^2=\frac{30^2}{7^2}+\frac{2^2}{7^2}+\frac{82^2}{7^2}+\frac{96^2}{7^2}+\frac{16^2}{7^2}+\frac{44^2}{7^2}+\frac{86^2}{7^2}=\\ &=\frac{30^2+2^2+82^2+96^2+16^2+44^2+86^2}{49}=\\ &=\frac{900+4+6724+9216+256+1936+7396}{49}=\frac{26432}{49}=\\ &=\frac{3776}{7}=539,428571,\\ S_{xy} &= \sum(x_i-\overline{x})(y_i-\overline{y})=-12\cdot\left(-\frac{30}{7}\right)-8\cdot\left(-\frac{2}{7}\right)-4\cdot\frac{82}{7}+\\ &\quad+0\cdot\frac{96}{7}+4\cdot\left(-\frac{16}{7}\right)+8\cdot\left(-\frac{44}{7}\right)+12\cdot\left(-\frac{86}{7}\right)=\\ &=\frac{360+16+328+0-64-352-1032}{7}=\\ &=-\frac{744}{7}=-106,285714.\end{aligned}$$

4. Подсчитаем коэффициенты

$$b=\frac{S_{xy}}{S_{xx}}=\frac{-\frac{744}{7}}{448}=-\frac{744}{7\cdot 448}=-\frac{93}{7\cdot 56}=\frac{93}{392}=-0,2372449,$$

$$b_1 = \frac{S_{xy}}{S_{yy}} = \frac{-\frac{744}{7}}{\frac{3776}{7}} = -\frac{744}{3776} = -\frac{93}{472} = -0,197034.$$

5. Уравнение прямой регрессии

$$\begin{aligned}\hat{y} &= \overline{y} + b(x - \overline{x}) = \frac{100}{7} - \frac{93}{392}(x - 168) = \\ &= -\frac{93}{392}x + \frac{100}{7} + \frac{93}{392} \cdot 168 = -\frac{93}{392}x + \frac{100}{7} + \frac{93 \cdot 3}{7} = \\ &= -\frac{93}{392}x + \frac{100 + 279}{7} = -\frac{93}{392}x + \frac{379}{7} = \frac{1}{392}(-93x + 21224) = \\ &= -0,2372449x + 54,142857.\end{aligned}$$

Уравнение обратной регрессии

$$\begin{aligned}\hat{x} &= \overline{x} + b_1(y - \overline{y}) = 168 - \frac{93}{472}\left(y - \frac{100}{7}\right) = \\ &= -\frac{93}{472}y + 168 + \frac{93}{472} \cdot \frac{100}{7} = -\frac{93}{472}x + 168 + \frac{93 \cdot 25}{7 \cdot 118} = \\ &= -\frac{93}{472}y + \frac{138768 + 2325}{826} = -\frac{93}{472}y + \frac{141093}{826} = \\ &= \frac{3}{3304}(-31 \cdot 7y + 47031 \cdot 4) = \frac{3}{3304}(-217y + 188124) = \\ &= -0,197034y + 170,81477.\end{aligned}$$

6. Оценка квадрата коэффициента корреляции

$$\begin{aligned}bb_1 &= \left(-\frac{93}{392}\right) \cdot \left(-\frac{93}{472}\right) = \frac{93^2}{392 \cdot 472} = \frac{93^2}{185024} = \frac{93^2}{2^6 \cdot 7^2 \cdot 59} = \\ &= \frac{1}{59} \cdot \left(\frac{93}{2^3 \cdot 7}\right)^2 = \frac{1}{59} \cdot \left(\frac{93}{56}\right)^2 = 0,046745.\end{aligned}$$

Оценка коэффициента корреляции

$$-\sqrt{bb_1} = -\frac{1}{\sqrt{59}} \cdot \frac{93}{56} = -0,216207.$$

7. Используем уравнение прямой регрессии

$$\hat{y} = \frac{100}{7} - \frac{93}{392}(x - 168)$$

и получаем с использованием

$x - \overline{x}$	-12	-8	-4	0	4	8	12

предсказываемые по уравнению прямой регрессии значения

$$\hat{y}_1 = \frac{100}{7} - \frac{93}{392}(-12) = \frac{100}{7} + \frac{93}{392}\cdot 12 = \frac{100}{7} + \frac{93\cdot 3}{98} = $$
$$= \frac{100\cdot 14 + 279}{98} = \frac{1679}{98} = 17,132653;$$
$$\hat{y}_2 = \frac{100}{7} - \frac{93}{392}(-8) = \frac{100}{7} + \frac{93}{392}\cdot 8 = \frac{100}{7} + \frac{93}{49} = $$
$$= \frac{100\cdot 7 + 93}{49} = \frac{793}{49} = 16,183673;$$
$$\hat{y}_3 = \frac{100}{7} - \frac{93}{392}(-4) = \frac{100}{7} + \frac{93}{392}\cdot 4 = \frac{100}{7} + \frac{93}{98} = $$
$$= \frac{100\cdot 14 + 93}{98} = \frac{1493}{98} = 15,234694;$$
$$\hat{y}_4 = \frac{100}{7} - \frac{93}{392}\cdot 0 = \frac{100}{7} = 14,285714;$$
$$\hat{y}_5 = \frac{100}{7} - \frac{93}{392}\cdot 4 = \frac{100}{7} - \frac{93}{98} = \frac{100\cdot 14 - 93}{98} = \frac{1307}{98} = 13,336735;$$
$$\hat{y}_6 = \frac{100}{7} - \frac{93}{392}\cdot 8 = \frac{100}{7} + \frac{93}{49} = \frac{100\cdot 7 - 93}{49} = \frac{607}{49} = 12,387755;$$
$$\hat{y}_7 = \frac{100}{7} - \frac{93}{392}\cdot 12 = \frac{100}{7} - \frac{93\cdot 3}{98} = \frac{100\cdot 14 - 279}{98} = \frac{1121}{98} = $$
$$= 11,438776.$$

8. Используем уравнение обратной регрессии

$$\hat{x} = 168 - \frac{93}{472}\left(y - \frac{100}{7}\right)$$

и получаем с использованием

$y - \bar{y}$	$-\frac{30}{7}$	$-\frac{2}{7}$	$\frac{82}{7}$	$\frac{96}{7}$	$-\frac{16}{7}$	$-\frac{44}{7}$	$-\frac{86}{7}$

предсказываемые по уравнению обратной регрессии значения

$$\hat{x}_1 = 168 - \frac{93}{472}\left(-\frac{30}{7}\right) = 168 + \frac{93\cdot 15}{7\cdot 236} = 168 + \frac{1395}{1652} = $$
$$= 168 + 0,844431 = 168,844431;$$
$$\hat{x}_2 = 168 - \frac{93}{472}\left(-\frac{2}{7}\right) = 168 + \frac{93}{7\cdot 236} = 168 + \frac{93}{1652} = $$
$$= 168 + 0,056295 = 168,056295;$$
$$\hat{x}_3 = 168 - \frac{93}{472}\cdot\frac{82}{7} = 168 - \frac{93\cdot 41}{7\cdot 236} = 168 - \frac{3813}{1652} = $$

$$= 168 - 2,308111 = 165,691889;$$

$$\hat{x}_4 = 168 - \frac{93}{472} \cdot \frac{96}{7} = 168 + \frac{93 \cdot 12}{7 \cdot 59} = 168 - \frac{1116}{413} =$$

$$= 168 - 2,702179 = 165,297821;$$

$$\hat{x}_5 = 168 - \frac{93}{472} \left(-\frac{16}{7}\right) = 168 + \frac{93 \cdot 2}{7 \cdot 59} = 168 + \frac{186}{413} =$$

$$= 168 + 0,450363 = 168,450363;$$

$$\hat{x}_6 = 168 - \frac{93}{472} \left(-\frac{44}{7}\right) = 168 + \frac{93 \cdot 11}{7 \cdot 118} = 168 + \frac{1023}{826} =$$

$$= 168 + 1,238499 = 169,238499;$$

$$\hat{x}_7 = 168 - \frac{93}{472} \left(-\frac{86}{7}\right) = 168 + \frac{93 \cdot 43}{7 \cdot 236} = 168 + \frac{3999}{1652} =$$

$$= 168 + 2,420702 = 170,420702.$$

9. Отклонение прямой регрессии от истинных значений равно

$$\sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} (y_i - \overline{y})^2 - \sum_{i=1}^{n} (\hat{y}_i - \overline{y})^2 =$$

$$= S_{yy} - \sum_{i=1}^{n} \left(b(\hat{x}_i - \overline{x})\right)^2 = S_{yy} - b^2 S_{xx} =$$

$$= S_{yy} - \frac{S_{xy}^2}{S_{xx}^2} \cdot S_{xx} = S_{yy} - \frac{S_{xy}^2}{S_{xx}} = \frac{S_{yy}S_{xx} - S_{xy}^2}{S_{xx}},$$

и аналогично отклонение обратной регрессии от истинных значений равно

$$\sum_{i=1}^{n} (x_i - \hat{x}_i)^2 = \frac{S_{yy}S_{xx} - S_{xy}^2}{S_{yy}}.$$

Подсчитаем

$$S_{yy}S_{xx} - S_{xy}^2 = 448 \cdot \frac{3776}{7} - \left(-\frac{744}{7}\right)^2 = 64 \cdot 3776 - \frac{744^2}{49} =$$

$$= 64 \left(3776 - \frac{93^2}{49}\right) = 64 \cdot \frac{3776 \cdot 49 - 93^2}{49} =$$

$$= 64 \cdot \frac{185204 - 8649}{49} = 64 \cdot \frac{176375}{49} =$$

$$= \frac{11288000}{49} = 230367,346939.$$

Таким образом, отклонение прямой регрессии от истинных значений равно

$$\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \frac{S_{yy}S_{xx} - S_{xy}^2}{S_{xx}} = \frac{\frac{11288000}{49}}{448} = \frac{176375}{7 \cdot 49} = \\ = \frac{176375}{343} = 514,212828;$$

и отклонение обратной регрессии от истинных значений равно

$$\sum_{i=1}^{n}(x_i - \hat{x}_i)^2 = \frac{S_{yy}S_{xx} - S_{xy}^2}{S_{yy}} = \frac{\frac{11288000}{49}}{\frac{3776}{7}} = \frac{11288000}{7 \cdot 3776} = \frac{176375}{7 \cdot 59} = \\ = \frac{176375}{413} = 427,058111.$$

□

9. Дополнения

9.1. Доказательство предельной теоремы для гипергеометрического распределения с помощью формулы Стирлинга

Доказательство. Имеем

$$\begin{aligned}
P_{k_1,k}(s_1,s) = \frac{C_{k_1}^{s_1} C_{k_2}^{s_2}}{C_k^s} &= \frac{s!(k-s)!}{k!} \cdot \frac{k_1!}{s_1!(k_1-s_1)!} \cdot \frac{k_2!}{s_2!(k_2-s_2)!} = \\
&= \frac{s!}{s_1! s_2!} \cdot \frac{(k-s)!}{k!} \cdot \frac{k_1!}{(k_1-s_1)!} \cdot \frac{k_2!}{(k_2-s_2)!} = \\
&= C_s^{s_1} \cdot \frac{(k-s)! \cdot k_1! \cdot k_2!}{k!(k_1-s_1)! \cdot (k_2-s_2)!}.
\end{aligned}$$

По формуле Стирлинга

$$\begin{aligned}
(k-s)! &= \sqrt{2\pi(k-s)}\,(k-s)^{k-s} e^{-(k-s)+\theta(k-s)}, \\
k_1! &= \sqrt{2\pi k_1}\, k_1^{k_1} e^{-k_1+\theta(k_1)}, \\
k_2! &= \sqrt{2\pi k_2}\, k_2^{k_2} e^{-k_2+\theta(k_2)}, \\
k! &= \sqrt{2\pi k}\, k^k e^{-k+\theta(k)}, \\
(k_1-s_1)! &= \sqrt{2\pi(k_1-s_1)}\,(k_1-s_1)^{k_1-s_1} e^{-(k_1-s_1)+\theta(k_1-s_1)}, \\
(k_2-s_2)! &= \sqrt{2\pi(k_2-s_2)}\,(k_2-s_2)^{k_2-s_2} e^{-(k_2-s_2)+\theta(k_2-s_2)}.
\end{aligned}$$

с учётом того, что $k_2 = k - k_1$ и $s_2 = s - s_1$ получаем — числитель

$$\begin{aligned}
(k-s)! \cdot k_1! \cdot k_2! &= \sqrt{2\pi(k-s) \cdot 2\pi k_1 \cdot 2\pi k_2}\,(k-s)^{k-s} \cdot k_1^{k_1} \cdot k_2^{k_2} \times \\
&\quad \times e^{-(k-s)-k_1-k_2} \cdot e^{\theta(k-s)+\theta(k_1)+\theta(k_2)} = \\
&= \sqrt{(2\pi)^3 (k-s) k_1 k_2}\,(k-s)^{k-s} k_1^{k_1} k_2^{k_2} \cdot e^s \times \\
&\quad \times e^{\theta(k-s)+\theta(k_1)+\theta(k_2)}
\end{aligned}$$

и знаменатель

$$\begin{aligned}
k! \cdot (k_1-s_1)! \cdot (k_2-s_2)! &= \sqrt{2\pi k \cdot 2\pi(k_1-s_1) \cdot 2\pi(k_2-s_2)} \times \\
&\quad \times k^k \cdot (k_1-s_1)^{k_1-s_1} \cdot (k_2-s_2)^{k_2-s_2} \times \\
&\quad \times e^{-k-k_1+s_1-k_2+s_2} \cdot e^{\theta(k)+\theta(k_1-s_1)+\theta(k_2-s_2)} = \\
&= \sqrt{(2\pi)^3 k (k_1-s_1)(k_2-s_2)} \times e^{\theta(k)+\theta(k_1-s_1)+\theta(k_2-s_2)} \times \\
&\quad \times k^k \cdot (k_1-s_1)^{k_1-s_1} \cdot (k_2-s_2)^{k_2-s_2} \cdot e^s.
\end{aligned}$$

Теперь

$$
\begin{aligned}
\frac{(k-s)!\cdot k_1!\cdot k_2!}{k!(k_1-s_1)!\cdot(k_2-s_2)!} &= \frac{\sqrt{(2\pi)^3(k-s)k_1k_2}}{\sqrt{(2\pi)^3k(k_1-s_1)(k_2-s_2)}}\times\\
&\times\frac{(k-s)^{k-s}k_1^{k_1}k_2^{k_2}}{k^k\cdot(k_1-s_1)^{k_1-s_1}\cdot(k_2-s_2)^{k_2-s_2}}\cdot\frac{e^s}{e^s}\times\\
&\times e^{\theta(k-s)+\theta(k_1)+\theta(k_2)-\theta(k)-\theta(k_1-s_1)-\theta(k_2-s_2)}=\\
&=\sqrt{\frac{(k-s)k_1k_2}{k(k_1-s_1)(k_2-s_2)}}\times\\
&\times\frac{(k-s)^k k_1^{k_1}k_2^{k_2}}{k^k\cdot(k_1-s_1)^{k_1}\cdot(k_2-s_2)^{k_2}}\times\\
&\times\frac{(k-s)^{-s}}{(k_1-s_1)^{-s_1}\cdot(k_2-s_2)^{-s_2}}\times\\
&\times e^{\theta(k-s)+\theta(k_1)+\theta(k_2)-\theta(k)-\theta(k_1-s_1)-\theta(k_2-s_2)}=\\
&=\sqrt{\frac{k-s}{k}\cdot\frac{k_1}{k_1-s_1}\cdot\frac{k_2}{k_2-s_2}}\times\\
&\times\left(\frac{k-s}{k}\right)^k\cdot\left(\frac{k_1}{k_1-s_1}\right)^{k_1}\cdot\left(\frac{k_2}{k_2-s_2}\right)^{k_2}\times\\
&\times\frac{(k_1-s_1)^{s_1}\cdot(k_2-s_2)^{s_2}}{(k-s)^s}\times\\
&\times e^{\theta(k-s)+\theta(k_1)+\theta(k_2)-\theta(k)-\theta(k_1-s_1)-\theta(k_2-s_2)}.
\end{aligned}
$$

Так как s, s_1 и $s_2=s-s_1$ фиксированы, а $\frac{k_1}{k}$ стремится к p при $k\to\infty$, то при $k\to\infty$

$$
\begin{aligned}
&\frac{k-s}{k}=1-\frac{s}{k}\longrightarrow 1,\\
&\frac{k_1}{k_1-s_1}=\frac{\frac{k_1}{k}}{\frac{k_1}{k}-\frac{s_1}{k}}\longrightarrow\frac{p}{p}=1,\\
&\frac{k_2}{k_2-s_2}=\frac{k-k_1}{k-k_1-s_2}=\frac{1-\frac{k_1}{k}}{1-\frac{k_1}{k}-\frac{s_2}{k}}\longrightarrow\frac{1-p}{1-p}=1,
\end{aligned}
$$

Аналогично при $k\to\infty$

$$
\frac{(k_1-s_1)^{s_1}\cdot(k_2-s_2)^{s_2}}{(k-s)^s}=\frac{\left(\frac{k_1-s_1}{k}\right)^{s_1}\cdot\left(\frac{k_2-s_2}{k}\right)^{s_2}}{\left(\frac{k-s}{k}\right)^s}\longrightarrow\frac{p^{s_1}(1-p)^{s_2}}{1}=p^{s_1}q^{s-s_1}.
$$

Далее при $k \to \infty$

$$\left(\frac{k-s}{k}\right)^k = \left(1-\frac{s}{k}\right)^{-\frac{k}{s}(-s)} \longrightarrow \left(\lim_{-\frac{s}{k}=\alpha\to 0}(1+\alpha)^{\alpha}\right)^{-s} = e^{-s}$$

по второму замечательному пределу. Также при $k \to \infty$ имеем $k_1 \to \infty$ (ибо $\frac{k_1}{k}$ стремится к p при $k \to \infty$)

$$\left(\frac{k_1-s_1}{k_1}\right)^{k_1} = \left(1-\frac{s_1}{k_1}\right)^{-\frac{k_1}{s_1}(-s_1)} \longrightarrow \left(\lim_{-\frac{s_1}{k_1}=\alpha\to 0}(1+\alpha)^{\alpha}\right)^{-s_1} = e^{-s_1}.$$

Аналогично

$$\left(\frac{k_1-s_1}{k_1}\right)^{k_1} \longrightarrow e^{-s_2}.$$

Таким образом при $k \to \infty$

$$\left(\frac{k-s}{k}\right)^k \left(\frac{k_1}{k_1-s_1}\right)^{k_1} \left(\frac{k_2}{k_2-s_2}\right)^{k_2} = \frac{\left(\frac{k-s}{k}\right)^k}{\left(\frac{k_1-s_1}{k_1}\right)^{k_1}\left(\frac{k_2-s_2}{k_2}\right)^{k_2}} \longrightarrow \frac{e^{-s}}{e^{-s_1}e^{-s_2}} = 1.$$

□

9.2. Различные типы сходимости в теории вероятностей

Соглашение. Если не будет оговорено противное, считаем, что все упоминаемые в этом пункте случайные величины заданы на одном вероятностном пространстве $\langle \Omega, \mathcal{A}, \mathsf{P} \rangle$.

Определение. Последовательность случайных величин $\{g_n\}_{n=1}^{\infty}$ сходится к случайной величине g *по вероятности*, если для всякого действительного числа $\varepsilon > 0$

$$\lim_{n\to\infty} \mathsf{P}(|g_n - g| < \varepsilon) = 1.$$

Обозначается сходимость по вероятности так

$$g_n \underset{\mathsf{P}}{\to} g \quad \text{при } n \to \infty.$$

Замечание. Утверждения из законов больших чисел говорят, что там имеется сходимость по вероятности.

Определение. Последовательность случайных величин $\{g_n\}_{n=1}^{\infty}$ сходится к случайной величине g *с вероятностью* 1, *или почти наверное, или почти всюду*, если существует такое событие $A \in \mathfrak{A}$, что

$$\mathsf{P}(A) = 1 \quad \text{и} \quad \lim_{n\to\infty} g_n(\omega) = g(\omega) \quad \text{для всех } \omega \in A.$$

Обозначается сходимость с вероятностью 1 так

$$g_n \xrightarrow[\text{п. н.}]{} g \quad \text{при } n \to \infty.$$

Замечание. Легко понять, что сходимость с вероятностью 1 влечёт сходимость по вероятности. Наоборот, вообще говоря, неверно. Однако в случае монотонности последовательности $\{g_n\}_{n=1}^{\infty}$ это верно. Не ограничивая общности, можем считать, что $g \equiv 0$на Ω, $g_n \geqslant 0$ для всех n и последовательность $\{g_n\}_{n=1}^{\infty}$ убывает. Допустим, что последовательность $\{g_n\}_{n=1}^{\infty}$ не сходится к 0 с вероятностью 1. Это означает, что для некоторого $\varepsilon > 0$ существуют такие натуральное m, $\delta > 0$ и $A \in \mathfrak{A}$, что $\mathsf{P}(A) > \delta > 0$ и $g_k(\omega) > \varepsilon$ для всех $\omega \in A$ и $k \geqslant m$. Теперь в силу убывания последовательности $\{g_n\}_{n=1}^{\infty}$ получаем, что

$$\mathsf{P}(g_n > \varepsilon) \geqslant \mathsf{P}(A) > \delta > 0 \quad \text{для всех } n,$$

что противоречит сходимости по вероятности.

В общем случае, если последовательность $\{g_n\}_{n=1}^{\infty}$ сходится к случайной величине g по вероятности, то существует подпоследовательность $\{g_{n_k}\}_{k=1}^{\infty}$, которая сходится к g по вероятностью 1.

Определение. Пусть $r > 0$. Тогда последовательность случайных величин $\{g_n\}_{n=1}^{\infty}$ сходится к случайной величине g *в среднем r-го порядка (при $r = 2$ среднеквадратически)*, если

$$\lim_{n\to\infty} \mathsf{M}\left(|g_n - g|^r\right) = 1.$$

Обозначается сходимость в среднем r-го так

$$g_n \xrightarrow[(r)]{} g \quad \text{при } n \to \infty.$$

Замечание. Сходимость в среднем r-го порядка влечёт сходимость по вероятности.

Определение. Последовательность случайных величин $\{g_n\}_{n=1}^{\infty}$ сходится к случайной величине g *по распределению*, если для любой непрерывной и ограниченной на $\mathbf{R}$ функции f

$$\lim_{n\to\infty} \int_{-\infty}^{\infty} f(x)\, dF_{g_n}(x) = \int_{-\infty}^{\infty} f(x)\, dF_g(x).$$

Обозначается сходимость по распределению r-го так

$$g_n \Longrightarrow g \quad \text{при } n \to \infty.$$

Замечания.
Здесь будут приведены факты, связанные со сходимостью по распределению.

1. Вообще говоря, при сходимости по распределению не требуется, чтобы случайные величины были определены на одном вероятностном пространстве.
2. Сходимость по распределению — это частный случай слабой сходимости функций. Говорят, что последовательность функций $\{F_n : \mathbf{R} \to \mathbf{R}\}_{n=1}^{\infty}$ *слабо сходится* к функции $F : \mathbf{R} \to \mathbf{R}$, если для любой непрерывной и ограниченной на $\mathbf{R}$ функции f
$$\lim_{n\to\infty} \int_{-\infty}^{\infty} f(x)\, dF_n(x) = \int_{-\infty}^{\infty} f(x)\, dF(x)$$
при условии, что такие интегралы определены.
3. Упомянем без доказательства два важных факта относительно слабой сходимости (значит и относительно сходимости по распределению).
Теорема. *Последовательность функций $\{F_n : \mathbf{R} \to \mathbf{R}\}_{n=1}^{\infty}$ слабо сходится к функции $F : \mathbf{R} \to \mathbf{R}$ тогда и только тогда, когда*
$$\lim_{n\to\infty} F_n(x) = F(x)$$
в любой точке непрерывности x функции F.
Следствие. *Пусть последовательность функций $\{F_n : \mathbf{R} \to \mathbf{R}\}_{n=1}^{\infty}$ слабо сходится к функции $F : \mathbf{R} \to \mathbf{R}$. Допустим, что F непрерывна на $\mathbf{R}$. Тогда $\{F_n\}_{n=1}^{\infty}$ равномерно сходится к функции $F : \mathbf{R} \to \mathbf{R}$, то есть*
$$\lim_{n\to\infty} \sup_{x\in\mathbf{R}} |F_n(x) - F(x)| = 0.$$

9.3. Характеристические функции

Пусть g_1 и g_2 — случайные величины на одном вероятностном пространстве. Можно рассматривать комплекснозначную случайную величину $g_1 + ig_2$ и изучать её свойства. Впервые полезность таких случайных величин понял (насколько мне известно) А.М. Ляпунов, и он применил такие случайные величины к предельным теоремам. Более точно, Ляпуновым были введены характеристические функции случайных величин, к изучению которых мы приступаем.

Определение. Пусть g — случайная величина. Тогда её *характеристической функцией* называется функция

$$\varphi_g : \mathbf{R} \to \mathbf{C}, \quad \text{для любого } t \in \mathbf{R}$$
$$\varphi_g(t) = \mathsf{M}\, e^{itg} = \mathsf{M}(\cos tg + i \sin tg) = \mathsf{M} \cos tg + i\, \mathsf{M} \sin tg.$$

Замечания.

1. Если случайная величина g дискретна и имеет значения $\{x_k\}_k$, то мы получаем
$$\varphi_g(t) = \mathsf{M}\, e^{itg} = \sum_k e^{itx_k}\, \mathsf{P}(g = x_k).$$

2. Если же случайная величина непрерывна (или если мы стоим на общей точке зрения на математическое ожидание, используя понятие интеграла Лебега), то мы имеем
$$\varphi_g(t) = \mathsf{M}\, e^{itg} = \int_{-\infty}^{\infty} e^{itx} dF_g(x).$$
Если плотность непрерывной случайной величины g равна pd_g, то
$$\varphi_g(t) = \mathsf{M}\, e^{itg} = \int_{-\infty}^{\infty} e^{itx}\, \mathrm{pd}_g(x)\, dx.$$
С абстрактной точки зрения характеристическая функция является ничем иным, как преобразованием Фурье(-Стильтьеса) плотности pd_g.

Предложение (о свойствах характеристических функций).
Пусть g — случайная величина и φ_g — её характеристическая функция. Тогда выполняются следующие утверждения.

1. $\varphi_g(0) = 1$ *и* $\varphi_g(t) \leqslant 1$ *для любого* $t \in \mathbf{R}$.

2. Для любых действительных чисел a и b
$$\varphi_{ag+b}(t) = e^{itb}\varphi_g(at).$$

3. Допустим, что $g_1, \ldots, g_n$ — независимые случайные величины на одном вероятностном пространстве и
$$g = g_1 + \cdots + g_n.$$
Тогда
$$\varphi_g(t) = \varphi_{g_1}(t) \ldots \varphi_{g_n}(t) \quad \text{для любого } t \in \mathbf{R}.$$

4. Характеристическая функция φ_g равномерно непрерывна на $\mathbf{R}$.

5. *Пусть n — неотрицательное целое число. Допустим, что существует конечное математическое ожидание $\mathsf{M}\,|g|^n < \infty$. Тогда существует n-я непрерывная производная характерирристической функции φ_g, причём $\varphi_g^{(n)}(0) = i^n \,\mathsf{M}\, g^n$.*

Доказательство.

1. В самом деле
$$\varphi_g(0) = \mathsf{M}\, e^{i0g} = \mathsf{M}\, e^0 = \mathsf{M}\, 1 = 1.$$
Для дискретной случайной величины со значениями $\{x_k\}_k$
$$|\varphi_g(t)| = \left|\sum_k e^{itx_k}\,\mathsf{P}(g = x_k)\right| \leqslant \sum_k \left|e^{itx_k}\right| \mathsf{P}(g = x_k) = \sum_k \mathsf{P}(g = x_k) = 1.$$
Для непрерывной случайной величины
$$|\varphi_g(t)| = \left|\int_{-\infty}^{\infty} e^{itx} dF_g(x)\right| \leqslant \int_{-\infty}^{\infty} \left|e^{itx}\right|\, dF_g(x) = \int_{-\infty}^{\infty} \left|e^{itx}\right|\, dF_g(x) = 1.$$
2. Имеем
$$\varphi_{ag+b}(t) = \mathsf{M}\, e^{it(ag+b)} = \mathsf{M}\, e^{itb} e^{itag} = e^{itb}\,\mathsf{M}\, e^{itag} = e^{itb}\varphi_g(at),$$
поскольку e^{itb} — константа и её можно, как легко понять, выносить.
3. Получаем
$$\varphi_g(t) = \mathsf{M}\, e^{itg} = \mathsf{M}\, e^{it(g_1+\dots+g_n} = \mathsf{M}\left(e^{itg_1} \cdot \ldots \cdot e^{itg_n}\right) = \mathsf{M}\, e^{itg_1} \cdot \ldots \cdot \mathsf{M}\, e^{itg_n} = \\ = \varphi_{g_1}(t) \cdot \ldots \cdot \varphi_{g_n}(t),$$
по свойствам математического ожидания в силу независимости случайных величин.
4. Для любого действительного h имеем
$$|\varphi_g(t+h) - \varphi_g(t)| = \left|\mathsf{M}\, e^{i(t+h)g} - \mathsf{M}\, e^{itg}\right| = \left|\mathsf{M}\, e^{itg}(e^{ihg} - 1)\right| \leqslant \\ \leqslant \mathsf{M}\left(\left|e^{itg}\right| \left|e^{igh} - 1\right|\right) \leqslant \mathsf{M}\left|e^{ihg} - 1\right| \to 1 \quad \text{при } h \to 0,$$
поскольку $e^{ihg} \to 1$ при $h \to 0$.
5. Рассмотрим случай $n = 1$. Для дискретной случайной величины со значениями $\{x_k\}_k$
$$\left|\sum_k ix_k e^{itx_k}\,\mathsf{P}(g = x_k)\right| \leqslant \sum_k |x_k|\,\mathsf{P}(g = x_k) = \mathsf{M}\,|g| < \infty.$$

Для непрерывной случайной величины
$$\left|\int_{-\infty}^{\infty} ixe^{itx}dF_g(x)\right| \leqslant \int_{-\infty}^{\infty} |x|\; dF_g(x) = \mathsf{M}\,|g| < \infty.$$
Из этих соотношений следует, что сумма и интеграл равномерно сходятся относительно t, поэтому можно дифференцировать под знаками суммы и интеграла. Откуда получаем для дискретной случайной величины
$$\begin{aligned}\varphi_g'(t) &= \left(\sum_k e^{itx_k}\,\mathsf{P}(g = x_k)\right)' = \sum_k ix_k e^{itx_k}\,\mathsf{P}(g = x_k) = \\ &= i\sum_k x_k e^{itx_k}\,\mathsf{P}(g = x_k) \quad \text{и при } t = 0 \\ \varphi_g'(0) &= i\sum_k x_k\,\mathsf{P}(g = x_k) = i\,\mathsf{M}\,g.\end{aligned}$$
Для непрерывной случайной величины
$$\begin{aligned}\varphi_g'(t) &= \left(\int_{-\infty}^{\infty} e^{itx}dF_g(x)\right)' = \int_{-\infty}^{\infty} ixe^{itx}dF_g(x) = \\ &= i\int_{-\infty}^{\infty} xe^{itx}dF_g(x) \quad \text{и при } t = 0 \\ \varphi_g'(0) &= ii\int_{-\infty}^{\infty} xdF_g(x) = i\,\mathsf{M}\,g.\end{aligned}$$
Теперь, беря производную n раз, можно получить, что снова можно дифференцировать под знаками суммы и интеграла, при каждом дифференцировании будет выноситься i, и всё получается по индукции.

□

9.3.1. Примеры характеристических функций

1. Пусть вероятность $\mathsf{P}(g = a) = 1$. Тогда
$$\varphi_g(t) = e^{ita}\cdot 1 = e^{ita}.$$
2. Пусть g распределена по схеме Бернулли с вероятностью успеха p, то есть
$$g = \begin{cases} 1 & \text{с вероятностью } p, \\ 0 & \text{с вероятностью } q = 1 - p. \end{cases}$$
Тогда
$$\varphi_g(t) = e^{it\cdot 1}\cdot p + e^{it\cdot 0}\cdot q = pe^{it} + 1 - p = 1 + p(e^{it} - 1).$$

3. Пусть g — стандартное нормальное распределение. Тогда

$$\varphi_g(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{itx} e^{-\frac{x^2}{2}} dx.$$

Ясно, что этот интеграл сходится равномерно по t. Поэтому дифференцируем его по t, и получаем

$$\begin{aligned}
\varphi_g'(t) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} ixe^{itx} e^{-\frac{x^2}{2}} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} -ie^{itx} d(e^{-\frac{x^2}{2}}) \quad \text{(по частям)} = \\
&= -ie^{itx} \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \Big|_{-\infty}^{\infty} - \left(\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (-i)ite^{itx} e^{-\frac{x^2}{2}} dx \right) = \\
&= -t \cdot \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{itx} e^{-\frac{x^2}{2}} dx = -t\varphi_g(t).
\end{aligned}$$

Таким образом получаем дифференциальное уравнение

$$\varphi_g'(t) = -t\varphi_g(t).$$

Отсюда

$$\ln \varphi_g'(t) = -t \quad \text{и потому} \quad \ln \varphi_g(t) = -\frac{t^2}{2} + C.$$

Так как $\varphi_g(0) = 1$, то $C = 0$ и $\ln \varphi_g(t) = -\frac{t^2}{2}$, а

$$\varphi_g(t) = e^{-\frac{t^2}{2}}.$$

4. Распределение Пуассона Π_λ с параметром $\lambda > 0$. В этом случае

$$\begin{aligned}
\varphi_{\Pi_\lambda}(t) &= \sum_{k=0}^{\infty} e^{itk}\, \mathsf{P}(\Pi_\lambda = k) = \sum_{k=0}^{\infty} e^{itk} e^{-\lambda} \frac{\lambda^k}{k!} = e^{-\lambda} \sum_{k=0}^{\infty} e^{itk} \frac{\lambda^k}{k!} = \\
&= e^{-\lambda} \sum_{k=0}^{\infty} \frac{(e^{it}\lambda)^k}{k!} = e^{-\lambda} e^{e^{it}\lambda} = e^{\lambda(e^{it}-1)}.
\end{aligned}$$

5. Показательное распределение i_λ с параметром $\lambda > 0$. В этом случае плотность

$$\mathrm{pd}_{P_\lambda}(x) = \begin{cases} \lambda e^{-\lambda x} & \text{при } x \leqslant 0, \\ 0 & \text{при } x < 0. \end{cases}$$

Отсюда

$$\varphi_{Pi_\lambda}(t) = \int_{-\infty}^{\infty} e^{itx}\, \mathrm{pd}_{P_\lambda}(x) dx = \int_0^{\infty} e^{itx} \lambda e^{-\lambda x} dx = \lambda \int_0^{\infty} e^{(it-\lambda)x} dx =$$

$$= \lambda \frac{e^{(it-\lambda)x}}{it-\lambda}\Big|_0^\infty = 0 - \lambda \frac{e^{(it-\lambda)0}}{it-\lambda} = \frac{\lambda}{\lambda - it}$$

Итак
$$\varphi_{Pi_\lambda}(t) = \frac{\lambda}{\lambda - it}.$$

6. Равномерное распределение на отрезке $\langle a, b\rangle$ для $-\infty < a < b < \infty$. В этом случае плотность
$$\mathrm{pd}_g(x) = \begin{cases} \frac{1}{b-a} & \text{при } x \in \langle a, b\rangle, \\ 0 & \text{при } x \notin \langle a, b\rangle. \end{cases}$$

Следовательно
$$\varphi_g(t) = \int_{-\infty}^{\infty} e^{itx}\,\mathrm{pd}_g(x)dx = \int_a^b \frac{e^{itx}}{b-a}dx = \frac{1}{b-a}\frac{e^{itx}}{it}\Big|_a^b =$$
$$= \frac{1}{b-a}\frac{e^{itb}}{it} - \frac{1}{b-a}\frac{e^{ita}}{it} = \frac{e^{itb}-e^{ita}}{it(b-a)}.$$

Таким образом
$$\varphi_g(t) = \frac{e^{itb}-e^{ita}}{it(b-a)},$$
в частности, для отрезка $\langle 0, 1\rangle$
$$\varphi_g(t) = \frac{e^{it}-1}{it}.$$

Следующие два предложения — без доказательства.

Предложение.

Формула обращения.

Пусть g — случайная величина, F_g — её функция распределения и φ_g — её характеристическая функция. Тогда для любых точек непрерывности x и y функции F_g имеем
$$F_g(y) - F_g(x) = \frac{1}{2\pi}\lim_{\sigma\to 0}\int_{-\infty}^{\infty} \frac{e^{-itx}-e^{-ity}}{it}\varphi_g(t)e^{-t^2\sigma^2}dt.$$

Если же, вдобавок, функция $\frac{\varphi_g(t)}{t}$ интегрируема при $t \to \pm\infty$, то
$$F_g(y) - F_g(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{e^{-itx}-e^{-ity}}{it}\varphi_g(t)\,dt.$$

Теорема единственности.

Характеристическая функция случайной величины однозначно определяет её функцию распределения.

Предложение (Теорема непрерывности (сходимости)).
Пусть $\{g_n\}_{n=1}^{\infty}$ *и* g *— случайные величины. Последовательность функций распределения* $\{F_{g_n}\}_{n=1}^{\infty}$ *слабо сходится к функции распределения* F_g *тогда и только тогда, когда последовательность характеристических функций* $\{\varphi_{g_n}(t)\}_{n=1}^{\infty}$ *сходится к характеристической функции* $\varphi_g(t)$ *для любого* $t \in \mathbf{R}$.

9.4. Доказательство центральной предельной теоремы (ЦПТ) для одинаково распределённых случайных величин

Теорема (Центральная предельная теорема для одинаково распределённых случайных величин).
Пусть $\{g_n\}_{n=1}^{\infty}$ *— последовательность независимых одинаково распределённых случайных величин на одном вероятностном пространстве с математическим ожиданием* $\mathsf{M}\, g_n = a$ *и дисперсией* $\mathsf{D}\, g_n = \sigma^2 \neq 0$ *для любого натурального* n. *Положим*

$$S_n = g_1 + \cdots + g_n \quad \text{и} \quad h_n = \frac{S_n - na}{\sigma\sqrt{n}}$$

(так как $na = \mathsf{M}\, S_n$ *и* $n\sigma^2 = \mathsf{D}\, S_n$*, то* h_n *— нормировка* S_n*). Тогда для любого действительного* x

$$\lim_{n\to\infty} \mathsf{P}(h_n < x) = \Phi_{0,1}(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt$$

равномерно по x.

Доказательство. Свойства сходимости, отмеченные в конце раздела 9.2, позволяют заключить, что равномерная сходимость — следствие слабой сходимости и непрерывности $\Phi_{0,1}(x)$. Поэтому, благодаря теореме непрерывности (сходимости) (конец предшествующего пункта), достаточно доказать, что для любого $t \in \mathbf{R}$ последовательность характеристических функций $\{\varphi_{h_n}(t)\}_{n=1}^{\infty}$ сходится к $e^{-\frac{t^2}{2}}$ — характеристической функции для $\Phi_{0,1}$.

Для удобства можем считать, что $a = 0$, то есть $\mathsf{M}\, g_n = 0$ для любого натурального n. В самом деле, если $a \neq 0$, то рассмотрим последовательность $\{g'_n = g_n - a\}_{n=1}^{\infty}$. Тогда

$$S'_n = g'_1 + \cdots + g'_n = g_1 + \cdots + g_n - na = S_n - na.$$

Поэтому нормированная последовательность $\{h_n\}_{n=1}^{\infty}$ тчно такой же, как для исходной последовательности $\{g_n\}_{n=1}^{\infty}$.

По утверждениям 2 и 3 из предложения о свойствах характеристических функций (предыдущий пункт), так как

$$h_n = \frac{g_1}{\sigma\sqrt{n}} + \cdots + \frac{g_n}{\sigma\sqrt{n}},$$

то

$$\varphi_{h_n}(t) = \varphi_{g_1}\left(\frac{t}{\sigma\sqrt{n}}\right) \cdot \cdots \cdot \varphi_{g_n}\left(\frac{t}{\sigma\sqrt{n}}\right) = \varphi_{g_1}^n\left(\frac{t}{\sigma\sqrt{n}}\right)$$

в силу одинаковой распределённости случайных величин $g_1, \ldots, g_n$.

Далее, так как g_1 имеет дисперсию, то существует $\mathsf{M}\, g_1^2 = \mathsf{D}\, g_1 = \sigma^2$ по утверждению 5 из предложения о свойствах характеристических функций существует непрерывная вторая производная $\varphi_{g_1}''(t)$. Поэтому можно применить к функции $\varphi_{g_1}(t)$ формулу Тейлора с остаточным членом в форме Пеано при $t \to 0$

$$\varphi_{g_1}(t) = \varphi_{g_1}(0) + t\varphi_{g_1}'(0) + \frac{t^2}{2}\varphi_{g_1}''(0) + o(t^2).$$

Теперь по утверждениям 1 и 5 из предложения о свойствах характеристических функций имеем

$$\varphi_{g_1}(t) = 1 - \frac{t^2\sigma^2}{2} + o(t^2).$$

Отсюда

$$\varphi_{h_n}(t) = \left(1 - \frac{\sigma^2\left(\frac{t}{\sigma\sqrt{n}}\right)^2}{2} + o\left(\frac{t}{\sigma\sqrt{n}}\right)^2\right)^n = \left(1 - \frac{t^2}{2n} + o\left(\frac{t^2}{\sigma^2 n}\right)\right)^n.$$

Следовательно

$$\ln(\varphi_{h_n}(t)) = n\ln\left(1 - \frac{t^2}{2n} + o\left(\frac{t^2}{\sigma^2 n}\right)\right) =$$

по формуле Тейлора для функции $\ln(1+x)$

$$= n\left(-\frac{t^2}{2n} + t^2 \cdot o\left(\frac{t^2}{\sigma^2 n}\right)\right) = -\frac{t^2}{2} + o(1) \to -\frac{t^2}{2} \text{ при } n \to \infty,$$

что нам нужно. □

9.5. Цепи Маркова

9.5.1. Основные понятия

Пусть G — опыт с не более чем счётным числом исходов $\{E_i\}_i$. Сопоставим ему последовательность целочисленных случайных величин $\{g_n\}_{n=0}^{\infty}$ по следующему правилу. Исходное состояние будем считать 0-ым испытанием. Если в n-ом испытании получен исход E_j, то положим $g_n = j$.

Определение. Последовательность случайных величин $\{g_n\}_{n=0}^{\infty}$ назовём *цепью Маркова,* если для любого натурального n условная вероятность

$$\mathsf{P}(g_n = j/g_0 = k_0, \ldots, g_{n-1} = k_{n-1} = i) = \mathsf{P}(g_n = j/g_{n-1} = i) = p_{i,j}^{(n)}.$$

Замечания.

1. Число $p_{i,j}^{(n)}$ — *вероятность* перехода из состояния E_i в состояние E_j на шаге n.
2. Должны иметь $\sum_j p_{i,j}^{(n)} = 1$.
3. Для полного определения состояния опыта G нужно задать начальные вероятности

$$\mathsf{P}(g_0 = j) = p_j^{(0)} \quad \text{и нужно, чтобы } \sum_j p_j^{(0)} = 1.$$

. Таким образом для построения цепи Маркова надо задать

$$\left\{ \{p_{i,j}^{(n)}\}_{i,j} \middle| \, p_{i,j}^{(n)} \geqslant 0, \sum_j p_{i,j}^{(n)} = 1 \right\} \text{ — вероятности переходов и}$$

$$\left\{ \{p_j^{(0)}\}_j \middle| \, p_j^{(0)} \geqslant 0, \sum_j p_j^{(0)} = 1 \right\} \text{ — начальное распределение.}$$

5. Свойство, определяющее цепь Маркова, ещё формулируют так:

   При фиксированном настоящем будущее не зависит от прошлого.

   Также можно это свойство сформулировать так:

   Вероятность любого исхода в очередном испытании однозначно определяется результатом непосредственно предшествующего испытания.

Теперь найдём вероятность нахождения опыта в данной последовательности исходов $E_{k_0}, \ldots, E_{k_n}$. Это равносильно нахождению вероятности

$$\mathsf{P}(g_0 = k_0, \ldots, g_n = k_n).$$

Теорема (о марковости).

$$\mathsf{P}(g_0 = k_0, \ldots, g_n = k_n) = p_{k_0}^{(0)} p_{k_0,k_1}^{(1)} \cdots p_{k_{n-1},k_n}^{(n)}.$$

Это свойство называют марковостью.

Доказательство. В самом деле,

$$\begin{aligned}&\mathsf{P}(g_0 = k_0, \ldots, g_n = k_n) = \\ &\quad = \mathsf{P}(g_0 = k_0, \ldots, g_{n-1} = k_{n-1})\, \mathsf{P}(g_n = k_n / g_0 = k_0, \ldots, g_{n-1} = k_{n-1})\end{aligned}$$

(по формуле умножения вероятностей)

$$= \mathsf{P}(g_0 = k_0, \ldots, g_{n-1} = k_{n-1}) p_{k_{n-1},k_n}^{(n)} =$$

по индукции получаем

$$= p_{k_0}^{(0)} p_{k_0,k_1}^{(1)} \cdots p_{k_{n-1},k_n}^{(n)},$$

что и нужно. □

В дальнейшем, в основном буем рассматривать частный случай цепей Маркова — однородные цепи Маркова.

Определение. Цепь Маркова $\{g_n\}_{n=0}^{\infty}$ назовём *однородной,* если для любых i и j вероятность перехода из состояния E_i в состояние E_j не зависит от номера испытания n, то есть для i и j

$$\mathsf{P}(g_n = j / g_{n-1} = i) = p_{i,j}^{(n)} = p_{ij} = \text{const по } n.$$

Замечания.

1. Таким образом для построения однородной цепи Маркова надо задать

$$\begin{gathered}(p_{ij})_{i,j} \text{ — матрицу вероятностей переходов со свойством} \\ p_{i,j} \geqslant 0 \text{ для всех } i \text{ и } j \text{ и } \sum_j p_{i,j} = 1 \text{для всех } i \text{ и } j\end{gathered}$$

— это свойство называется *стохастичностью* и матрица $(p_{ij})_{i,j}$ называется *стохастической,*

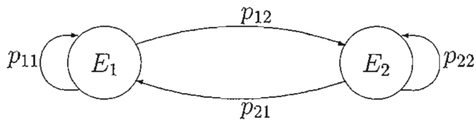
$$\begin{gathered}\left(p_j^{(0)}\right)_j \text{ — вектор начальных состояний со свойством} \\ p_j^{(0)} \geqslant 0 \text{ для всех } i \text{ и } j \text{ и } \sum_j p_j^{(0)} = 1.\end{gathered}$$

2. Свойство марковости выглядит в случае однородной цепи так:

$$\mathsf{P}(g_0 = k_0, \ldots, g_n = k_n) = p_{k_0}^{(0)} p_{k_0,k_1} \cdots p_{k_{n-1},k_n}.$$

В случае однородной цепи Маркова, часто матрицу вероятностей переходов изображают в виде графа с вершинами $\{E_i\}_i$ следующим образом. Состояние E_i соединяют дугой (ребром с направлением) с состоянием E_j, помеченной числом p_{ij}, при $p_{ij} = 0$ дугу обычно не рисуют. Заметим, что в этом графе могут быть петли, если $p_{ii} \neq 0$ для некоторого i. Впоследствии в задачах мы покажем такие графы. Сейчас ограничимся одним примером.

Пример. Пусть у нас всего два состояния — E_1 и E_2. Тогда в общем виде граф такой:



$$p_{11} + p_{12} = 1, \quad p_{21} + p_{12} = 1 \quad \text{и} \quad p_{ij} \geqslant 0 \text{ для всех } i, j = 1, 2.$$

В заключение этого пункта приведём важный результат, но поскольку не будем его использовать, то дадим его в виде упражнения.

Упражнение. *Пусть $\{h_n\}_{n=0}^{\infty}$ — последовательность независимых случайных величин, принимающих целые значения. Для любого натурального n положим $S_n = \sum_{k=0}^{n} h_n$. Тогда $\{S_n\}_{n=0}^{\infty}$ образует цепь Маркова.*

Так как последовательности сумм независимых случайных величин у нас встречались ранее (например, в предельных теоремах), то становится понятной и важной роль цепей Маркова в теории вероятностей.

9.5.2. Классификация состояний

Соглашение. Пусть задана однородная цепь Маркова $\{g_n\}_{n=0}^{\infty}$ со множеством состояний $\{E_i\}_i$.

Обозначение. Для всякого натурального k положим

$$p_{ij}(k) = \mathsf{P}(g_k = j / g_0 = i) \quad \text{и} \quad P(k) = (p_{ij}(k))_{ij}.$$

Определение. Состояние E_i называется *несущественным*, если найдутся такие состояние E_j и натуральное число t_0, что

$$p_{ij}(t_0) > 0 \quad \text{и} \quad p_{ji}(t) = 0 \text{ для любого натурального } t.$$

(Можем уйти из E_j в E_i, но не можем вернуться обратно!)
В противном случае состояние E_i называется *существенным*.

Определение. Существенные состояния E_j и E_i называются *сообщающимися*, если найдутся такие натуральные числа t и s, что

$$p_{ij}(t) > 0 \quad \text{и} \quad p_{ji}(s) > 0.$$

(Можем уйти из E_j в E_i, и можем вернуться обратно!)

Пример. Пусть система может находиться в одном из четырёх состояний $\{E_1, E_2, E_3, E_4\}$ с матрицей вероятностей переходов:

$$\begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & 0 & 0 & \frac{1}{2} \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \end{pmatrix}$$

Тогда в этой цепи состояния E_1 и E_2 несущественны, а E_3 и E_4 являются существенными сообщающимися состояниями.

Обозначения. Обозначим через S^0 — множество всех несущественных состояний. Для существенного состояния E положим

$$S_E = \{E\} \cup \{\text{все состояния, сообщающиеся с } E\}.$$

Если $E_j \in S_{E_i}$, то E_j существенно, сообщается с E_i и $E_i \in S_{E_j}$. Следовательно, $S_{E_i} = S_{E_j}$. Таким образом возникает разбиение множества состояний на классы эквивалентности

$$\{E_i\}_i = S^0 \cup S_{E_{i_1}} \cup \dots \cup S_{E_{i_k}} \cup \cdots$$

Определение. Состояние E называется *поглощающим*, если

$$S_E = \{E\}.$$

(Можем попасть в E, а выйти не можем!)

Определение. Цепь Маркова называется *неразложимой*, если в ней всего один класс существенных состояний (несущественных нет), если же более одного класса, то цепь называется *разложимой*.

Обозначения. Для натурального n обозначим через

$$f_j(n) = \mathsf{P}(g_n = j, g_{n-1} \neq j, \dots, g_1 \neq j / g_0 = j)$$

— вероятность возвращения в исходное состояние E_j *точно через n шагов*. Также положим

$$F_j = \sum_{n=0}^{\infty} f_j(n)$$

— вероятность возвращения в исходное состояние E_j *когда-нибудь*.

Определение. Состояние E_j называется *возвратным*, если $F_j = 1$, и *невозвратным*, если $F_j < 1$.

Определение. Состояние E_j называется *нулевым*, если $p_{jj} \to 0$ при $n \to \infty$, иначе E_j называется *ненулевым*.

Определение. Состояние E_j называется *периодическим* с периодом $d_j > 1$, если возвращение в E_j с положительной вероятностью возможно за число шагов кратное d_j и d_j — наибольшее число с таким свойством. Иначе говоря

$$d_j = \text{НОД}\{n \mid f_j(n) > 0\}.$$

Здесь НОД — наибольший общий делитель.

Определение. Пусть задана произвольная числовая (из комплексных чисел) последовательность $\{a_n\}_{n=0}^{\infty}$. Тогда *производящей функцией* последовательности $\{a_n\}_{n=0}^{\infty}$ называется (формальный) степенной ряд

$$a(z) = \sum_{n=0}^{\infty} a_n z^n, \ z \in \mathbf{C}.$$

Замечание. Очевидно, что указанный ряд сходится при $z = 0$, поэтому возникает функция

$$a(z) : O \to \mathbf{C},$$

где O — область сходимости данного ряда. Однако для производящих функций это не суть важно (поэтому и употребляется термин *формальный* ряд), и часто допускается случай, когда $O = \{0\}$.

Теорема (Критерий возвратности).
Для состояния E_j положим

$$P_j = \sum_{n=1}^{\infty} p_{jj}(n).$$

Состояние E_j возвратно тогда и только тогда, когда $P_j = \infty$ (то есть приведённый выше ряд расходится). Если состояние E_j невозвратно, то

$$F_j = \frac{P_j}{1 + P_j}.$$

Доказательство. Для последовательностей $\{p_{jj}(n)\}_{n=1}^{\infty}$ и $\{f_j(n)\}_{n=1}^{\infty}$ производящими функциями являются, соответственно

$$P_j(z) = \sum_{n=1}^{\infty} p_{jj}(n) z^n \quad \text{и} \quad F_j(z) = \sum_{n=1}^{\infty} f_j(n) z^n,$$

которые сходятся внутри круга $|z| < 1$, т.к. для любого натурального n

$$0 \leqslant p_{jj}(n), f_j(n) \leqslant 1.$$

Следовательно внутри круга $|z| < 1$ возникают аналитические функции $P_j(z)$ и $F_j(z)$.

Из того, что если в E_j можно вернуться не более чем за n шагов, то в E_j можно вернуться точно за 1, 2, ..., n шагов, по формуле полной вероятности имеем

$$p_{jj}(n) = f_j(1)p_{jj}(n-1) + f_j(2)p_{jj}(n-2) + \dots + f_j(n-1)p_{jj}(1) + f_j(n)\cdot 1.$$

Умножим это равенство на z^n и получим

$$\begin{aligned} p_{jj}(n)z^n &= f_j(1)p_{jj}(n-1)z^n + f_j(2)p_{jj}(n-2)z^n + \cdots \\ &\quad \cdots + f_j(n-1)p_{jj}(1)z^n + f_j(n)z^n = \\ &= f_j(n)z^n + zf_j(1)\cdot p_{jj}(n-1)z^{n-1} + z^2 f_j(2)\cdot p_{jj}(n-2)z^{n-2} + \cdots \\ &\quad \cdots + z^{n-1}f_j(n-1)\cdot p_{jj}(1)z. \end{aligned}$$

Теперь полученное равенство просуммируем по n

$$\begin{aligned} P_j(z) &= \sum_{n=1}^{\infty} p_{jj}(n)z^n = \sum_{n=1}^{\infty} f_j(n)z^n + \\ &\quad + zf_j(1)\cdot\sum_{n=2}^{\infty} p_{jj}(n-1)z^{n-1} + z^2 f_j(2)\cdot\sum_{n=3}^{\infty} p_{jj}(n-2)z^{n-2} + \cdots \\ &= F_j(z) + P_j(z)\sum_{k=1}^{\infty} z^k f_j(k) = F_j(z) + P_j(z)F_j(z) = F_j(z)\left(1 + P_j(z)\right). \end{aligned}$$

Отсюда

$$F_j(z) = \frac{P_j(z)}{1 + P_j(z)} \quad \text{и} \quad P_j(z) = \frac{F_j(z)}{1 - F_j(z)}.$$

Перейдём к доказательству теоремы. Сначала докажем утверждение о возвратности.

Необходимость. Если состояние E_j возвратно, то по определению $F_j = F_j(1) = 1$, то есть ряд $F_j(z) = \sum_{n=1}^{\infty} f_j(n)z^n$ сходится в точке $z = 1$. Поэтому по второй теореме Абеля для степенных рядов

$$\lim_{\substack{z\to 1\\ z\in(0,1)}} F_j(z) = 1 \quad \text{и} \quad \lim_{\substack{z\to 1\\ z\in(0,1)}} P_j(z) = \lim_{\substack{z\to 1\\ z\in(0,1)}} \frac{F_j(z)}{1 - F_j(z)} = \infty,$$

то есть $P_j = P_j(1) = \infty$.

Достаточность. Допустим, что $P_j = P_j(1) = \infty$. Тогда по второй теореме Абеля для степенных рядов

$$\lim_{\substack{z\to 1\\ z\in(0,1)}} P_j(z) = \infty \quad \text{и} \quad \lim_{\substack{z\to 1\\ z\in(0,1)}} F_j(z) = \lim_{\substack{z\to 1\\ z\in(0,1)}} \frac{P_j(z)}{1+P_j(z)} = 1,$$

то есть $F_j = F_j(1) = 1$, и состояниеE_j возвратно.

Теперь предположим, что состояние E_j невозвратно.
Тогда $P_j = P_j(1) < \infty$, то есть при $z = 1$ ряд

$$P_j(z) = \sum_{n=1}^{\infty} p_{jj}(n) z^n$$

сходится, и из формулы $F_j(z) = \frac{P_j(z)}{1+P_j(z)}$ получаем при $z = 1$

$$F_j = \frac{P_j}{1+P_j}.$$

□

Следствие. *Невозвратное состояние всегда нулевое.*

Доказательство. Для невозвратного состояния E_j по теореме

$$P_j = \sum_{n=1}^{\infty} p_{jj}(n) < \infty.$$

Необходимый признак сходимости рядов даёт, что

$$p_{jj}(n) \to 0 \text{ при } n \to \infty,$$

что означает, что состояние E_j — нулевое. □

Теорема (Теорема солидарности).
В неразложимой цепи Маркова все состояния принадлежат одному типу:
1) *если хоть одно состояние возвратно, то все состояния возвратны;*
2) *если хоть одно состояние нулевое, то все состояния нулевые;*
3) *если хоть одно состояние периодично с периодом d, то все состояния периодичны с периодом d.*

Доказательство. Пусть E_k и E_j — два различных состояния. Тогда они сообщающиеся, и по определению сообщающихся состояний существуют такие натуральные числа t и s, что

$$\alpha = p_{kj}(t) > 0 \quad \text{и} \quad \beta = p_{jk}(s) > 0.$$

По формуле полной вероятности для любого натурального n имеем

$$p_{kk}(t+n+s) = \sum_{l,m} p_{kl}(t)p_{lm}(n)p_{mk}(s).$$

В частности получаем

$$p_{kk}(t+n+s) \geqslant p_{kj}(t)p_{jj}(n)p_{jk}(s) = \alpha\beta p_{jj}(n).$$

Аналогично получаем

$$p_{jj}(t+n+s) \geqslant \alpha\beta p_{kk}(n).$$

Из этих полученных двух неравенств имеем при $n > t+s$

$$\frac{1}{\alpha\beta}p_{kk}(t+n+s) \geqslant p_{jj}(n) \geqslant \alpha\beta p_{kk}(n-t-s).$$

Таким образом асимптотические свойства при $n \to \infty$ последовательностей $\{p_{kk}(n)\}_{n=1}^{\infty}$ и $\{p_{jj}(n)\}_{n=1}^{\infty}$ одинаковы.

Теперь перейдём к доказательству однотипности.

1. Если состояние E_k возвратно, то по критерию возвратности

$$P_k = \sum_{n=1}^{\infty} p_{kk}(n) = \infty.$$

Отсюда

$$P_j = \sum_{n=1}^{\infty} p_{jj}(n) \geqslant \sum_{n=t+s+1}^{\infty} p_{jj}(n) \geqslant \alpha\beta \sum_{n=t+s+1}^{\infty} p_{kk}(n-t-s) = \infty,$$

что по критерию возвратности даёт, что состояние E_j возвратно.

2. Если состояние E_k — нулевое, то по определению $p_{kk}(n) \to 0$ при $n \to \infty$. Откуда $p_{jj}(n) \to 0$ при $n \to \infty$, и по определению E_j — нулевое.
3. Пусть состояние E_k периодично с периодом d_k. Если $p_{kk}(n) > 0$, то d_k делит n. Так как по формуле полной вероятности

$$p_{kk}(t+s) = \sum_{l} p_{kl}(t)p_{lk}(s) \geqslant p_{kj}(t)p_{jk}(s) = \alpha\beta > 0.$$

Значит d_k делит $t + s$.

Покажем, что состояние E_j тоже периодично с периодом d_j и $d_j = d_k$. Если $p_{jj}(m) > 0$, то, как показали ранее

$$p_{kk}(t + s + m) \geqslant \alpha\beta p_{jj}(m) > 0.$$

А отсюда d_k делит $t+s+m$ и потому d_k делит m. В частности получим, что d_k делит d_j, но аналогично d_j делит d_k. Следовательно $d_j = d_k$.

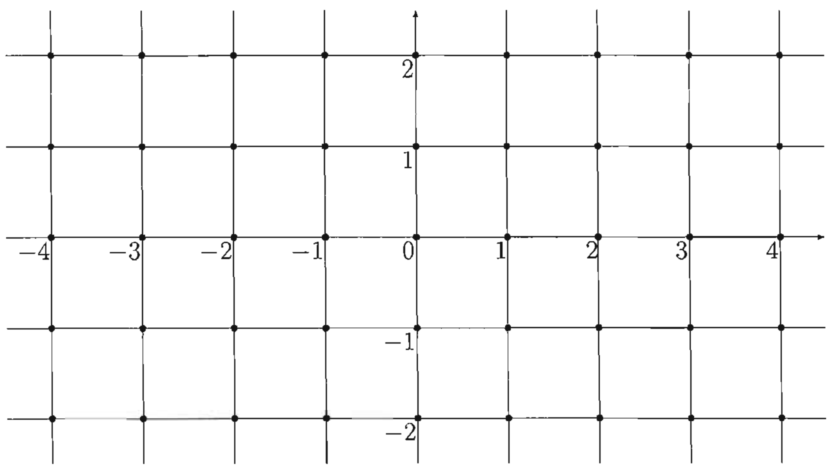
□

9.5.3. Симметричные случайные блуждания

Пусть k — натуральное число. Рассмотрим решётку

$$\mathbf{Z}^k = \{(a_1, \dots, a_k)|\, a_i \in \mathbf{Z}\ \forall i \in \{1, \dots, k\}\}$$

в пространстве $\mathbf{R}^k$. Такое название, отчасти, объясняется следующим рисунком для $k = 2$



Рассмотрим следующий процесс. Если в некоторый момент времени частица находится точке $(m_1, \dots, m_k)$ с целыми координатами (это некоторый узел решётки $\mathbf{Z}_k$), то в следующий момент времени она перемещается в точку $(n_1, \dots, n_k)$ с целыми координатами, причём $|n_j - m_j| = 1$ для каждого $j \in \{1, \dots, k\}$ (снова в узел решётки $\mathbf{Z}^k$), иными словами в точку

$$(m_1 \pm 1, m_2 \pm 1, \dots, m_k \pm 1)$$

с произвольным расположением знаков $+$ и $-$. Таких точек будет 2^k, и мы допускаем, что в каждую из них равновероятно, поэтому вероятность перехода в каждую из них должна быть равна $\frac{1}{2^k}$. Описанный процесс будем называть *симметричным случайным блужданием.*

Так как вероятность перехода в данную точку зависит только от того, в какой точке частица находилась в предыдущий момент времени, то понятно, что у нас возникает однородная цепь Маркова.

Теорема (о блужданиях).
Симметричное случайное блуждание возвратно при $k \leqslant 2$, а при $k \geqslant 3$ невозвратно.

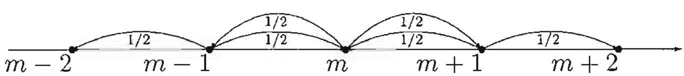
Замечание. Известный специалист по теории вероятностей Феллер (William Feller) выразил утверждение этой теоремы словами:

Все дороги ведут в Рим только при $k \leqslant 2$.

Доказательство теоремы о блужданиях. Разобьём доказательство на случаи:
1) блуждание на прямой $k = 1$;
2) блуждание на плоскости $k = 2$;
3) блуждание в пространствах размерности $k \geqslant 3$.

Блуждания на прямой

Имеем следующую картинку Поэтому ясно, что возникает неразложимая



периодическая цепь Маркова периода 2. Вероятность возвращения

$$f_n(2) = 2 \cdot \frac{1}{2} \cdot \frac{1}{2}.$$

В силу теоремы солидарности достаточно тип только одной точки 0 (нуля).

В силу периодичности нельзя вернуться за нечётное число шагов, то есть $p_{00}(2i-1) = 0$ для любого натурального i, и потому

$$P_0 = \sum_{n=1}^{\infty} p_{00}(n) = \sum_{i=1}^{\infty} p_{00}(2i).$$

Пусть S_{2i} — сумма координат блуждающей частицы за $2i$ блужданий. Ясно, что $p_{00}(2i) = \mathsf{P}(S_{2i} = 0)$, а $S_{2i} = 0$ тогда и только тогда, когда i раз была 1 и i раз была -1. Можно интерпретировать рассматриваемое блуждание как биномиальное распределение $\mathbf{B}_{2i}(\frac{1}{2})$ (успех — прибавление $(+1)$, неудача — отнимание (-1)). Применим локальную теорему Муавра–Лапласа. Имеем

$$x_{i,2i} = \frac{i - 2i \cdot \frac{1}{2}}{\sqrt{2i \cdot \frac{1}{2} \cdot \frac{1}{2}}} = 0.$$

Поэтому по локальной теореме Муавра–Лапласа вероятность

$$\mathsf{P}(S_{2i} = 0) = \mathsf{P}(\mathbf{B}_{2i}(\tfrac{1}{2}) = i)$$

асимптотически совпадает при $i \to \infty$ с

$$\frac{1}{\sqrt{2\pi \cdot 2i \cdot \frac{1}{2} \cdot \frac{1}{2}}} e^{\frac{x_{i,2i}^2}{2}} = \frac{1}{\sqrt{\pi \cdot i}} e^0 = \frac{1}{\sqrt{\pi \cdot i}},$$

Следовательно ряд

$$P_0 = \sum_{i=1}^{\infty} p_{00}(2i) \text{ асимптотически эквивалентен ряду } \sum_{i=1}^{\infty} \frac{1}{\sqrt{\pi \cdot i}}.$$

Поэтому $P_0 = \infty$, и по критерию возвратности получаем, что 0 — точка возврата.

Блуждания на плоскости

Пусть положение частицы в n-ый момент времени $g_n = (x_n, y_n)$. Тогда

$$g_n = (x_n, y_n) = (x_n, 0) + (0, y_n),$$

и мы получаем совместное блуждание по оси x и оси y. Отсюда

$$g_{n+1} = g_n + h_n, \text{ где } h_n \in \{(1,1), (1,-1), (-1,1), (-1,-1)\} = \{(\pm 1, \pm 1)\}.$$

Отметим, что

$$\mathsf{P}(h_n = (1,1)) = \mathsf{P}(h_n = (1,-1)) = \mathsf{P}(h_n = (-1,1)) = \mathsf{P}(h_n = (-1,-1)) = \frac{1}{4}.$$

Снова надо рассмотреть тип точки $O = (0,0)$, и опять ясно, что возврат возможен *только* за чётное число шагов. Имеем

$$p_{OO}(2n) = \mathsf{P}(g_{2n} = (0,0)/g_0 = (0,0)) = \mathsf{P}(x_{2n} = 0/x_0 = 0) \cdot \mathsf{P}(y_{2n} = 0/g_0 = 0).$$

При рассмотрении блуждания на прямой показано, что

$$\mathsf{P}(x_{2n} = 0/x_0 = 0) \quad \text{и} \quad \mathsf{P}(y_{2n} = 0/g_0 = 0)$$

асимптотически эквивалентны $\frac{1}{\sqrt{\pi n}}$ при $n \to \infty$. Поэтому вероятность $p_{OO}(2n)$ асимптотически эквивалентна $\frac{1}{\pi n}$ при $n \to \infty$. Следовательно ряд

$$P_0 = \sum_{n=1}^{\infty} p_{OO}(2n) \text{ асимптотически эквивалентен ряду } \sum_{n=1}^{\infty} \frac{1}{\pi n},$$

который расходится, что опять по критерию возвратности говорит о том, что точка $O = (0, 0)$ возвратна.

Блуждания в пространствах размерности $k \geqslant 3$

Будем рассуждать как в случае блуждания на плоскости, и опять разложим блуждание на блуждания по k координатным осям

$$g_n = (x_1, \ldots, x_k) = (x_1, 0, \ldots, 0) + (0, x_2, \ldots, 0) + \cdots + (0, \ldots, x_k).$$

Рассуждая, как в случае плоскости, получим, что тип точки $O = (0, \ldots, 0) \in \mathbf{Z}^k$ определяется рядом асимптотически эквивалентным при $n \to \infty$ ряду с общим членом

$$\left(\frac{1}{\sqrt{\pi n}}\right)^k = \frac{1}{(\pi n)^{k/2}},$$

который сходится при $k \geqslant 3$, что по критерию возвратности говорит о том, что точка $O = (0, \ldots, 0)$ не возвратна.

Все случаи разобраны, и теорема доказана. □

Замечание. Пусть блуждания состоят в перемещениях с равными вероятностями за один шаг на ± 1 по одной из осей координат. Тогда в случае прямой это в точности рассмотренный нами случай. При $k \geqslant 2$ блуждание невозвратно (теорема Пойа).

9.5.4. Предельные вероятности

Пусть задана однородная цепь Маркова $\{g_n\}_{n=0}^{\infty}$ с состояниями $\{E_i\}_i$. Напомним, что для любого натурального k

$$p_{ij}(k) = \mathsf{P}(g_k = j/g_0 = i) \quad \text{и} \quad P(k) = (p_{ij}(k))_{i,j}.$$

Лемма (уравнения Колмогорова–Чэпмена). *Для однородной цепи Маркова и любых натуральных k и l*

$$p_{ij}(k+l) = \sum_n p_{in}(k)p_{nj}(l),$$

или в матричной форме

$$P(k+l) = P(k)\cdot P(l) \quad \textit{и} \quad P(k) = (P(1))^k.$$

В частности,

$$p_{ij}(k+1) = \sum_n p_{in}(k)p_{nj} - \textit{прямое уравнение Колмогорова–Чэпмена},$$
$$p_{ij}(1+l) = \sum_n p_{in}p_{nj}(l) - \textit{обратное уравнение Колмогорова–Чэпмена}.$$

Доказательство. В самом деле

$$p_{ij}(k+l) = \mathsf{P}(g_{k+l} = j/g_0 = i) =$$

по формуле полной вероятности

$$= \sum_n \mathsf{P}(g_{k+l} = j, g_k = n/g_0 = i) =$$
$$= \sum_n \mathsf{P}(g_{k+l} = j/g_k = n)\,\mathsf{P}(g_k = n/g_0 = i) = \sum_n p_{nj}(l)p_{in}(k) =$$
$$= \sum_n p_{in}(k)p_{nj}(l).$$

□

Следующая теорема — важнейший результат из теории цепей Маркова. Название этой теоремы, которое мы приведём, кажется, не встречается в литературе, обычно же довольствуются несколькими словами (не всеми) из данного названия.

Теорема (Эргодическая теорема Маркова о предельных вероятностях). *Пусть $P = (p_{ij}))_{i,j=1}^n$ — матрица вероятностей переходов однородной цепи Маркова с конечным множеством состояний $\{E_1, \ldots, E_n\}$.*

1. *Если существует такое натуральное число n_0, что $\min_{ij} p_{ij}(n_0) > 0$ (то есть все элементы матрицы $P(n_0) = P^{n_0}$ строго положительны, пишут кратко $P^{n_0} > 0$), то существуют числа $\pi_1, \ldots, \pi_n$, называемые* предельными вероятностями *со следующими свойствами:*

а) *для любого* $j \in \{1, \dots, n\}$

$$\pi_j > 0 \quad и \quad \sum_{j=1}^{n} \pi_j = 1$$

(эти свойства называют эргодичностью*) и*

б) *для любых* $i, j \in \{1, \dots, n\}$

$$\lim_{s \to \infty} p_{ij}(s) = \pi_j.$$

2. *Обратно, если существуют числа* $\pi_1, \dots, \pi_n$*, удовлетворяющие вышеприведённым свойствам* а) *и* б)*, то существует такое натуральное* n_0*, что выполнено условие* $\min_{ij} p_{ij}(n_0) > 0$.
3. *Числа* $\pi_1, \dots, \pi_n$ *удовлетворяют системе уравнений*

$$\begin{cases} \pi_j = \sum\limits_{k=1}^{n} \pi_k p_{kj}, \ \forall j \in \{1, \dots, n\}, \\ \sum\limits_{j=1}^{n} \pi_j = 1; \end{cases}$$

или в матричной форме

$$\begin{cases} (\pi_1, \dots, \pi_n) = (\pi_1, \dots, \pi_n) P, \\ \sum\limits_{j=1}^{n} \pi_j = 1. \end{cases}$$

Доказательство.

1. Положим

$$m_j(s) = \min_i p_{ij}(s) \quad и \quad M_j(s) = \max_i p_{ij}(s).$$

Поскольку по лемме (уравнения Колмогорова–Чэпмена)

$$p_{ij}(s+1) = \sum_k p_{ik} p_{kj}(s),$$

то

$$m_j(s+1) = \min_i p_{ij}(s+1) = \min_i \sum_k p_{ik} p_{kj}(s) \geqslant$$
$$\geqslant \min_i \sum_k p_{ik} \left(\min_k p_{kj}(s) \right) = m_j(s),$$

так как $\sum_k p_{ik} = 1$. Следовательно,

$$m_j(s) \leqslant m_j(s+1)$$

и аналогично

$$M_j(s) \geqslant M_j(s+1).$$

Для доказательства б) достаточно доказать, что

$$M_j(s) - m_j(s) \to 0 \text{ при } s \to \infty \text{ для всех } j \in \{1, \ldots, n\}.$$

Пусть $\delta = \min_{ij} p_{ij}(n_0) > 0$. Тогда по лемме (уравнения Колмогорова–Чэпмена)

$$\begin{aligned}
p_{ij}(n_0 + s) &= \sum_k p_{ik}(n_0) p_{kj}(s) = \\
&= \sum_k \left(p_{ik}(n_0) - \delta p_{jk}(s)\right) p_{kj}(s) + \delta \sum_k p_{jk}(s) p_{kj}(s) = \\
&= \sum_k \left(p_{ik}(n_0) - \delta jk p_{jk}(s)\right) p_{kj}(s) + \delta p_{jj}(2s) \geqslant \\
&\geqslant m_j(s) \sum_k \left(p_{ik}(n_0) - \delta p_{jk}(s)\right) + \delta p_{jj}(2s) = \\
&= m_j(s)\,(1-\delta) + \delta p_{jj}(2s).
\end{aligned}$$

Следовательно

$$m_j(n_0 + s) \geqslant m_j(s)\,(1-\delta) + \delta p_{jj}(2s).$$

Аналогично получим

$$M_j(n_0 + s) \leqslant M_j(s)\,(1-\delta) + \delta p_{jj}(2s).$$

Вычитая из этого неравенства предыдущее, получаем

$$M_j(n_0 + s) - m_j(n_0 + s) \leqslant \left(M_j(s) - m_j(s)\right)(1-\delta).$$

Далее, по индукции для любого натурального k имеем

$$0 \leqslant M_j(kn_0 + s) - m_j(kn_0 + s) \leqslant \left(M_j(s) - m_j(s)\right)(1-\delta)^k.$$

Так как $\delta > 0$, то при $k \to \infty$

$$\left(M_j(s) - m_j(s)\right)(1-\delta)^k \to 0.$$

Поэтому при $k \to \infty$

$$M_j(kn_0 + s) - m_j(kn_0 + s) \to 0.$$

Так как последовательность $\{M_j(s) - m_j(s)\}_{s=1}^{\infty}$ ограничена и убывает, то она стремится к 0 при $s \to \infty$.

Последовательность $\{m_j(s)\}_{s=1}^{\infty}$ ограничена и возрастает. Поэтому существует предел

$$\lim_{s\to\infty} m_j(s) = \pi_j,$$

а так как при $s \geqslant n_0$

$$m_j(s) \geqslant m_j(n_0) = \delta > 0,$$

то $\pi_j > 0$. Аналогично существует предел

$$\lim_{s\to\infty} M_j(s),$$

а так как

$$\lim_{s\to\infty} (M_j(s) - m_j(s)) = 0,$$

то

$$\lim_{s\to\infty} M_j(s) = \lim_{s\to\infty} m_j(s) = \pi_j.$$

Также для любого i

$$m_j(s) \leqslant p_{ij}(s) \leqslant M_j(s).$$

Поэтому по теореме о зажатой переменной (теорема о двух милиционерах) для любого i

$$\lim_{s\to\infty} m_j(s) = \lim_{s\to\infty} p_{ij}(s) = \lim_{s\to\infty} M_j(s) = \pi_j.$$

2. Так как для любого $i \in \{1, \ldots, n\}$

$$\lim_{s\to\infty} p_{ij}(s) = \pi_j > 0,$$

то существует такое s_j, что

$$m_j(s_j) = \min_i p_{ij}(s_j) > 0.$$

Пусть

$$n_0 = s_1 \cdot s_2 \cdots s_n.$$

Тогда для любых i и j по лемме (уравнения Колмогорова–Чэпмена)

$$p_{ij}(n_0) = \sum_k p_{ik}\left(\frac{n_0}{s_j}\right) p_{kj}(s_j) \geqslant m_j(s_j) \sum_k p_{ik}\left(\frac{n_0}{s_j}\right) = m_(s_j) > 0.$$

3. По лемме (уравнения Колмогорова–Чэпмена) и определению цепи Маркова для любых i и j

$$\begin{cases} p_{ij}(s+1) = \sum\limits_{k=1}^{n} p_{ik}(s)p_{kj}, \ \forall i \in \{1, \ldots, n\}, \\ \sum\limits_{j=1}^{n} p_{ij}(s) = 1. \end{cases}$$

Переходя к пределу при $s \to \infty$, получаем

$$\begin{cases} \pi_j = \sum\limits_{k=1}^{n} \pi_k p_{kj}, \ \forall j \in \{1, \ldots, n\}, \\ \sum\limits_{j=1}^{n} \pi_j = 1. \end{cases}$$

Замечание. Легко заметить, что свойство

$$\begin{cases} (\pi_1, \ldots, \pi_n) = (\pi_1, \ldots, \pi_n)P, \\ \sum\limits_{j=1}^{n} \pi_j = 1. \end{cases}$$

обеспечивает, что 1 — собственное значение матрицы P (корень её характеристического многочлена), а $(\pi_1, \ldots, \pi_n)$ — соответствующий собственный вектор (при строчечной записи).

□

Выдержки из книги Боровкова (с небольшими изменениями)

Стр. 284–285. Рассмотрим следующие два условия:

(А) Цепь неразложима и непериодична.

(Б) Существует такое состояние, что случайная величина g, равная вероятности возвращения в E_0, то есть $\mathsf{P}(g = n) = f_0(n)$, имеет конечное математическое ожидание $\mathsf{M}\, g$.

Теорема (эргодическая).
Условия (А) *и* (Б) *необходимы и достаточны для того, чтобы при любых i и j существовали не зависящие от i положительные пределы*

$$\lim_{n\to\infty} p_{ij}(n) = \pi_j, \ \forall i, j \in \{0, 1, 2, \ldots, \}.$$

Числа $\{\pi_j\}_j$ являются единственным в классе последовательностей, образующих абсолютно сходящиеся ряды, решением системы уравнений

$$\begin{cases} \pi_j = \sum\limits_{k=0}^{\infty} \pi_k p_{kj}, \ \forall j \in \{0, 1, 2, \dots\}, \\ \sum\limits_{j=0}^{\infty} \pi_j = 1. \end{cases} \tag{e}$$

Цепь, обладающая свойством (e), называется *эргодической*.

Числа π_j есть по существу вероятности попадания системы в состояние E_j через большой интервал времени. При этом оказывается, что эти вероятности не зависят от начального состояния системы. Система «забывает», откуда началось движение. Распределение $\{\pi_j\}_j$ называют *стационарным* или *финальным*. Свойство (e) выражает инвариантность распределения относительно переходных вероятностей p_{ij}. Другими словами, если $\mathsf{P}(g_n = k) = \pi_k$, то $\mathsf{P}(g_{n+1} = k) = \sum \pi_j p_{jk}$ также равна π_k.

Далее обсудим случай периодических цепей Маркова.

Стр. 278. Покажем теперь, что изучение периодических цепей в значительной мере может быть сведено к изучению непериодических. Для этого нам понадобится

Теорема (о периодических цепях).
Если цепь Маркова — периодическая с периодом d, то множество состояний разбивается на d таких классов Ψ_0, $\Psi_1, \dots, \Psi_{d-1}$, что с вероятностью 1 за один шаг система из класса Ψ_k переходит в класс Ψ_{k+1}, из класса Ψ_{d-1} система переходит в класс Ψ_0.

Стр. 287. Если цепь Маркова — периодическая с периодом d, то $p_{ij}(t) = 0$, еси $t \neq kd$ для любой пары состояний E_i и E_j, принадлежащих одному классу (в смысле теоремы о периодических цепях). Если же $t = kd$, то из эргодической теоремы и теоремы о периодических цепях следует, что существует предел

$$\lim_{k\to\infty} p_{ij}(kd) = \pi_j > 0$$

и не зависит от i.

9.5.5. Примеры цепей Маркова

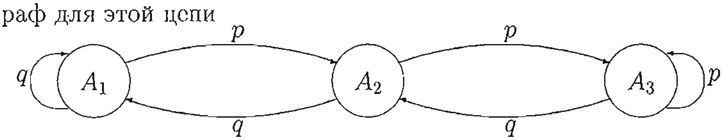
Задача. *Пусть частица движется по трём точкам A_1, A_2 и A_3 следующим образом. Если в какой-то момент времени $t = n$ ($n \in \{0, 1, 2, \dots\}$) частица находится в точке A_2, то в следующий момент времени $t = n+1$ она переходит в A_1 с вероятностью q ($0 < q < 1$), или в A_3 с вероятностью $p = 1 - q$ ($0 < p < 1$). Если же частица оказалась в точке в A_1, то в следующий момент времени она останется там же с вероятностью q или перейдёт в A_2 с вероятностью p. Если же частица оказалась в точке в A_3, то в следующий момент времени она останется там же с вероятностью p или перейдёт в A_2 с вероятностью q. Нужно:*

1) *найти матрицу вероятностей переходов P,*
2) *построить граф для этой цепи,*
3) *определить применима ли теорема Маркова к этой цепи,*
4) *если да, то найти предельные вероятности.*

Решение. Будем отвечать по пунктам.

1. Матрица вероятностей переходов
$$P = \begin{pmatrix} q & p & 0 \\ q & 0 & p \\ 0 & q & p \end{pmatrix}.$$

2. Граф для этой цепи



3. Матрица переходов за два шага
$$P(2) = P^2 = \begin{pmatrix} q & p & 0 \\ q & 0 & p \\ 0 & q & p \end{pmatrix}^2 = \begin{pmatrix} q^2 + qp & qp & p^2 \\ q^2 & qp + qp & p^2 \\ q^2 & qp & p^2 + pq \end{pmatrix} =$$
$$= \begin{pmatrix} q(q+p) & qp & p^2 \\ q^2 & 2qp & p^2 \\ q^2 & qp & p(p+q) \end{pmatrix} = \begin{pmatrix} q & qp & p^2 \\ q^2 & 2qp & p^2 \\ q^2 & qp & p \end{pmatrix} > 0.$$

   Таким образом, все элементы $P(2)$ строго положительны. Поэтому можно применять терему Маркова, и предельные вероятности существуют.

4. Для нахождения предельных вероятностей решим систему
$$\begin{cases} (\pi_1, \pi_2, \pi_3) = (\pi_1, \pi_2, \pi_3) \begin{pmatrix} q & p & 0 \\ q & 0 & p \\ 0 & q & p \end{pmatrix} \\ \pi_1 + \pi_2 + \pi_3 = 1 \end{cases} \longleftrightarrow \begin{cases} \pi_1 = q\pi_1 + q\pi_2 \\ \pi_2 = p\pi_1 + q\pi_3 \\ \pi_3 = p\pi_2 + p\pi_3 \\ \pi_1 + \pi_2 + \pi_3 = 1 \end{cases} \longleftrightarrow$$

$$\longleftrightarrow \begin{cases} (1-q)\pi_1 = q\pi_2 \\ \pi_2 = p\pi_1 + q\pi_3 \\ (1-p)\pi_3 = p\pi_2 \\ \pi_1 + \pi_2 + \pi_3 = 1 \end{cases} \longleftrightarrow \begin{cases} p\pi_1 = q\pi_2 \\ \pi_2 = p\pi_1 + q\pi_3 \\ q\pi_3 = p\pi_2 \\ \pi_1 + \pi_2 + \pi_3 = 1 \end{cases} \longleftrightarrow$$

$$\longleftrightarrow \begin{cases} \pi_1 = \dfrac{q}{p}\pi_2 \\ \pi_3 = \dfrac{p}{q}\pi_2 \\ \pi_2 = (q+p)\pi_2 = \pi_2 \\ \pi_2 = \dfrac{1}{\frac{q}{p} + 1 + \frac{p}{q}} = \dfrac{pq}{q^2 + pq + p^2} = \dfrac{pq}{1 - pq} \end{cases} \longleftrightarrow$$

$$\longleftrightarrow \begin{cases} \pi_1 = \dfrac{q^2}{1-pq} \\ \pi_2 = \dfrac{pq}{1-pq} \\ \pi_3 = \dfrac{p^2}{1-pq} \end{cases}.$$

В частном случае, когда

$$p = q = \frac{1}{2},$$

получим

$$\pi_1 = \pi_2 = \pi_3 = \frac{1}{3}.$$

□

Задача. *Изменим условие примера 1. А именно, пусть точки A_1 и A_3 — поглощающие экраны, то есть частица, попадая в эти точки, остаётся в них с вероятностью 1. Нужно:*

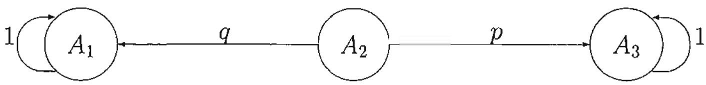
1) *найти матрицу вероятностей переходов P,*
2) *построить граф для этой цепи,*
3) *определить применима ли теорема Маркова к этой цепи.*

Решение. Будем отвечать по пунктам.

1. Матрица вероятностей переходов

$$P = \begin{pmatrix} 1 & 0 & 0 \\ q & 0 & p \\ 0 & 0 & 1 \end{pmatrix}.$$

2. Граф для этой цепи



3. Матрица переходов за два шага

$$P(2) = P^2 = \begin{pmatrix} 1 & 0 & 0 \\ q & 0 & p \\ 0 & 0 & 1 \end{pmatrix}^2 = \begin{pmatrix} 1 & 0 & 0 \\ q & 0 & p \\ 0 & 0 & 1 \end{pmatrix} = P.$$

Таким образом, для любого натурального n имеем $P(n) = P^n = P$. Поэтому; всегда есть нулевой столбец, и нельзя применять терему Маркова.

□

Задача. *Пусть частица движется по трём точкам A_1, A_2 и A_3 следующим образом. Если в какой-то момент времени частица находится в точке A_2, то в следующий момент времени она с вероятностями, равными $\frac{1}{3}$, или остаётся в A_2, или переходит в A_1, или в A_3. Если же частица оказалась в точках в A_1 или A_3, то в следующий момент времени она с вероятностями, равными $\frac{1}{1}$, или остаётся на месте, или переходит в A_2. Нужно:*

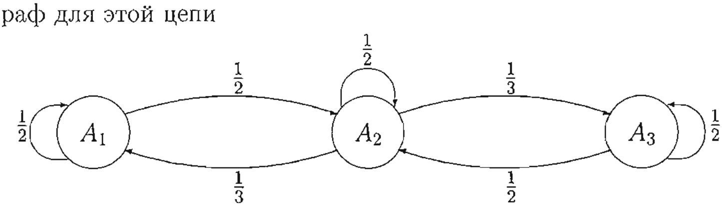
1) *найти матрицу вероятностей переходов P,*
2) *построить граф для этой цепи,*
3) *определить применима ли теорема Маркова к этой цепи,*
4) *если да, то найти предельные вероятности.*

Решение. Будем отвечать по пунктам.

1. Матрица вероятностей переходов

$$P = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ 0 & \frac{1}{2} & \frac{1}{2} \end{pmatrix}.$$

2. Граф для этой цепи



3. Матрица переходов за два шага

$$P(2) = P^2 = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ 0 & \frac{1}{2} & \frac{1}{2} \end{pmatrix}^2 = \begin{pmatrix} \frac{1}{4}+\frac{1}{6} & \frac{1}{4}+\frac{1}{6} & \frac{1}{6} \\ \frac{1}{9}+\frac{1}{6} & \frac{1}{6}+\frac{1}{9}+\frac{1}{6} & \frac{1}{9}+\frac{1}{6} \\ \frac{1}{6} & \frac{1}{4}+\frac{1}{6} & \frac{1}{4}+\frac{1}{6} \end{pmatrix} =$$

$$= \begin{pmatrix} \frac{5}{12} & \frac{5}{12} & \frac{1}{6} \\ \frac{5}{18} & \frac{4}{9} & \frac{5}{18} \\ \frac{1}{6} & \frac{5}{12} & \frac{5}{12} \end{pmatrix} > 0.$$

Таким образом, все элементы $P(2)$ строго положительны. Поэтому можно применять теорему Маркова, и предельные вероятности существуют.

4. Для нахождения предельных вероятностей решим систему

$$\begin{cases} (\pi_1, \pi_2, \pi_3) = (\pi_1, \pi_2, \pi_3) \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ 0 & \frac{1}{2} & \frac{1}{2} \end{pmatrix} \\ \pi_1 + \pi_2 + \pi_3 = 1 \end{cases} \longleftrightarrow$$

$$\longleftrightarrow \begin{cases} \pi_1 = \frac{1}{2}\pi_1 + \frac{1}{3}\pi_2 \\ \pi_2 = \frac{1}{2}\pi_1 + \frac{1}{3}\pi_2 + \frac{1}{2}\pi_3 \\ \pi_3 = \frac{1}{3}\pi_2 + \frac{1}{2}\pi_3 \\ \pi_1 + \pi_2 + \pi_3 = 1 \end{cases} \longleftrightarrow \begin{cases} \frac{1}{2}\pi_1 = \frac{1}{3}\pi_2 \\ \frac{2}{3}\pi_2 = \frac{1}{2}\pi_1 + \frac{1}{2}\pi_3 \\ \frac{1}{2}\pi_3 = \frac{1}{3}\pi_2 \\ \pi_1 + \pi_2 + \pi_3 = 1 \end{cases} \longleftrightarrow$$

$$\longleftrightarrow \begin{cases} \pi_1 = \frac{2}{3}\pi_2 \\ \pi_2 = \frac{3}{4}(\pi_1 + \pi_3) \\ \pi_3 = \frac{2}{3}\pi_2 \\ \pi_1 + \pi_2 + \pi_3 = 1 \end{cases} \longleftrightarrow \begin{cases} \pi_1 = \pi_3 = \frac{2}{3}\pi_2 \\ \pi_2 = \frac{3}{4} \cdot 2 \cdot \frac{2}{3}\pi_2 = \pi_2 \\ \left(1 + \frac{4}{3}\right)\pi_2 = 1 \end{cases} \longleftrightarrow$$

$$\longleftrightarrow \begin{cases} \pi_2 = \frac{3}{7} \\ \pi_1 = \pi_3 = \frac{2}{7} \end{cases}.$$

□

9.6. Информация и энтропия

9.6.1. Количество информации

Определения. Пусть A и B — события из одного вероятностного пространства, имеющие *ненулевые* вероятности. *Количеством информации,* заключённым в событии B относительно события A, называется число

$$\mathrm{I}(A/B) = \log_2 \frac{\mathsf{P}(A/B)}{\mathsf{P}(A)} = \log_2 \frac{\frac{\mathsf{P}(AB)}{\mathsf{P}(B)}}{\mathsf{P}(A)} = \log_2 \frac{\mathsf{P}(AB)}{\mathsf{P}(A)\,\mathsf{P}(B)}.$$

При $B = A$

$$\mathrm{I}(A/A) = \mathrm{I}(A) = \log_2 \frac{\mathsf{P}(A)}{\mathsf{P}(A)\,\mathsf{P}(A)} = -\log_2 \mathsf{P}(A).$$

называется *количеством информации* в событии A.

Теорема (о количестве информации).

1. *Если события A и B независимы, то* $\mathrm{I}(A/B) = 0$.
2. $\mathrm{I}(A/B) = \mathrm{I}(B/A)$.

Доказательство.

1. В самом деле, для независимых событий $\mathsf{P}(AB) = \mathsf{P}(A)\,\mathsf{P}(B)$. Поэтому

$$\mathrm{I}(A/B) = \log_2 \frac{\mathsf{P}(AB)}{\mathsf{P}(A)\,\mathsf{P}(B)} = \log_2 \frac{\mathsf{P}(A)\,\mathsf{P}(B)}{\mathsf{P}(A)\,\mathsf{P}(B)} = \log_2 1 = 0.$$

2. Очевидно из определения.

$\square$

Замечание. Использование 2 в качестве основания логарифма не является критически важным. Можно использовать любое основание большее 1. Однако широкое использование в компьютерных теориях двоичного представления информации делает использование 2 весьма обоснованным и удобным.

9.6.2. Энтропия

Определения. Пусть проводится некоторый опыт G с исходами $E_1, \ldots, E_n$, которые осуществляются с вероятностями $p_1, \ldots, p_n$, соответственно. Если все эти вероятности ненулевые, то возникает случайная величина

$$\mathrm{I}_G(E_i) = -\log_2 p_i \text{ для каждого } i \in \{1, \ldots, n\}.$$

Математическое ожидание этой случайной величины

$$\mathsf{M}\,\mathrm{I}_G = -\sum_{i=1}^{n} p_i \log_2 p_i = \mathrm{H}(G)$$

назовём *энтропией опыта* G. Так как $x\log_2 x \to 0$ при $x \to +0$, то, вполне разумно, для $p_i = 0$ положить $p_i \log_2 p_i = 0$, и тогда понятие энтропии распространяется на *любые* вероятности.

Теорема (Свойства энтропии).

1. *Энтропия* $\mathrm{H}(G) = 0$ *тогда и только тогда, когда существует такое* $j \in \{1,\dots,n\}$, *что* $p_j = 1$.
2. *Энтропия* $\mathrm{H}(G)$ *максимальна, если* $p_j = \frac{1}{n}$ *для любого* $j \in \{1,\dots,n\}$. *В этом случае*
$$\mathrm{H}(G) = \log_2 n.$$
3. *Если опыты* G_1 *и* G_2 *независимы, то для совместного эксперимента* $G = G_1 \times G_2$
$$\mathrm{H}(G) = \mathrm{H}(G_1) + \mathrm{H}(G_2).$$

Доказательство.

1. В самом деле
$$\mathrm{H}(G) = -\sum_{i=1}^{n} p_i \log_2 p_i = 0 \iff p_i \log_2 p_i = 0\ \forall i \iff$$
$$\iff p_i = 0 \text{ или } i \log_2 p_i = 0\ \forall i.$$
Так как $\sum_j p_j = 1$, то должен быть исход E_j с ненулевой вероятностью, но тогда должно быть $\log_2 p_j = 0$, а это выполняется тогда и только тогда, когда $p_j = 1$.
2. Рассмотрим функцию $f(x) = x \ln x$ на $(0, +\infty)$. Найдём производные
$$f'(x) = \ln x + x \cdot \frac{1}{x} = \ln x + 1,$$
$$f''(x) = \frac{1}{x} > 0 \text{ на } (0, +\infty).$$
Положительность второй производной влечёт, что $f(x)$ выпукла (вниз) на $(0, +\infty)$. Поэтому для $q_1 \geqslant 0, \dots, q_n \geqslant 0$ и $\sum_{i=1}^{n} q_i = 1$
$$f\Big(\sum_{i=1}^{n} q_i x_i\Big) \leqslant \sum_{i=1}^{n} q_i f(x_i) \text{ для всех } \{x_1,\dots,x_n\} \subseteq (0, +\infty).$$

Далее, т.к. $\log_2 x = \frac{\ln x}{\ln 2}$, то $f_1(x) = x\log_2 x$ на $(0, +\infty)$ также выпукла и

$$f_1\Big(\sum_{i=1}^{n} q_i x_i\Big) \leqslant \sum_{i=1}^{n} q_i f_1(x_i) \text{ для всех } \{x_1, \ldots, x_n\} \subseteq (0, +\infty).$$

Пусть для каждого $i \in \{1, \ldots, n\}$

$$q_i = \frac{1}{n} \quad \text{и} \quad x_i = p_i.$$

Придерживаясь прежнего соглашения, что $x\log_2 x = 0$ при $x = 0$ с учётом, что $\sum_i p_i = 1$, получаем

$$\frac{1}{n}\log_2\left(\frac{1}{n}\right) = \left(\frac{1}{n}\sum_i p_i\right)\log_2\left(\frac{1}{n}\sum_i p_i\right) = f_1\left(\sum_i q_i x_i\right) \leqslant$$
$$\leqslant \sum_{i=1}^{n} q_i f_1(x_i) = \sum_{i=1}^{n} \frac{1}{n}\log_2 p_i = \frac{1}{n}(-\operatorname{H}(G)).$$

Таким образом,

$$-\log_2 n = \log_2\left(\frac{1}{n}\right) \leqslant -\operatorname{H}(G) \iff \operatorname{H}(G) \leqslant \log_2 n.$$

Если $p_j = \frac{1}{n}$ для любого $j \in \{1, \ldots, n\}$, то

$$\operatorname{H}(G) = -\sum_{j=1}^{n} p_j \log_2 p_j = -\sum_{j=1}^{n} \frac{1}{n}\log_2\left(\frac{1}{n}\right) = -n \cdot \frac{1}{n}(-\log_2 n) = \log_2 n.$$

3. Опыты G_1 и G_2 определяются матрицами

$$G_1 = \begin{pmatrix} E_1 & \ldots & E_n \\ p_1 & \ldots & p_n \end{pmatrix} \quad \text{и} \quad G_2 = \begin{pmatrix} F_1 & \ldots & F_m \\ q_1 & \ldots & q_m \end{pmatrix}.$$

Тогда в силу независимости G_1 и G_2

$$G = G_1 \times G_2 = \begin{pmatrix} E_1F_1 & E_1F_2 & \ldots & E_nF_m \\ p_1q_1 & p_1q_2 & \ldots & p_nq_m \end{pmatrix}.$$

Придерживаясь прежнего соглашения, что $x\log_2 x = 0$ при $x = 0$ с учётом, что $\sum_i p_i = 1$, получаем

$$\operatorname{H}(G) = -\sum_{i,j} p_i q_j \log_2(p_i q_j) = -\sum_{i,j} p_i q_j \left(\log_2 p_i + \log_2 q_j\right) =$$

$$= -\sum_{i,j} p_i q_j \log_2 p_i - \sum_{i,j} p_i q_j \log_2 q_j =$$

$$= \left(\sum_j q_j\right)\left(-\sum_i p_i \log_2 p_i\right) + \left(\sum_i p_i\right)\left(-\sum_j q_j \log_2 q_j\right) =$$

$$= -\sum_i p_i \log_2 p_i - \sum_j q_j \log_2 q_j = \mathrm{H}(G_1) + \mathrm{H}(G_2).$$

$\square$

9.6.3. Энтропия цепи Маркова

Рассмотрим однородную цепь Маркова $C = \{g_k\}_{k=0}^{\infty}$ со следующими свойствами.

1. Цепь Маркова C имеет конечное множество состояний $\{E_1, \dots, E_n\}$.
2. Цепь Маркова C неразложима и непериодична.
3. Цепь Маркова C стационарна, что означает для любого $k \in \{1, 2, \dots\}$

$$\mathsf{P}(g_k = j) = \pi_j \text{ (не зависит от } k).$$

Совпадение с финальными вероятностями.

Определение. *Энтропией стационарной цепи Маркова* с матрицей переходных вероятностей $(p_{ij})_{i,j}$ называется число

$$\mathrm{H}(C) = -\sum_i \pi_i \sum_{ij} p_{ij} \log_2 p_{ij}.$$

Можно рассматривать число

$$-\frac{\log_2 \mathsf{P}(g = i_1, \dots, g_n = i_n)}{n}$$

как значение некоторой случайной величины I_n. Справедлива следующая теорема, которую мы не будем доказывать.

Теорема (об асимптотике).
При $n \to \infty$

$$I_n \underset{\mathsf{P}}{\to} \mathrm{H}(C).$$

Напомним, что $\underset{\mathsf{P}}{\to}$ означает сходимость по вероятности, то есть для всякого действительного числа $\varepsilon > 0$

$$\lim_{n\to\infty} \mathsf{P}\left(|I_n - \mathrm{H}(C)| < \varepsilon\right) = 1.$$

Замечание. Приведённая теорема имеет приложения в теории кодирования.

10. Задачи

10.1. Особые задачи

Приведённые ниже задачи предлагались для самостоятельного решения. За их решение полагались различные "блага" такие, как повышение оценки на экзамене, и даже просто "5" за экзамен. Все задачи удалось решить только 1 студенту.

Подобранные задачи возникали в процессе чтения учебника Боровкова и других книг. Не стоит удивляться тому, что среди задач встречаются такие, решения которых есть в вышеизложенном материале. Просто не всё читалось студентам на лекциях.

1. Построить все алгебры событий для множеств из 1, 2, 3 и 4 элементов.
2. Оценить число алгебр событий для множества из конечного числа элементов, используя свойства алгебр событий.
3. Доказать, что по всякому сходящемуся ряду с неотрицательными членами можно построить функцию распределения некоторой случайной величины.
4. Для гипергеометрического и биномиального распределения доказать, что они действительно являются распределениями некоторых случайных величин.
5. Найти непосредственно дисперсию в биномиальном распределении.
6. Построить пример распределения, имеющего математическое ожидание, но не имеющего дисперсии.
7. Функция распределения
$$\frac{1}{a}\int_{-\infty}^{x}\frac{dx}{x^4+1}$$
Найти число a, математическое ожидание и дисперсию.
8. Построить пример двух случайных величин, имеющих нулевой коэффициент корреляции, но зависимых.
9. Пусть $F_n(x) = \mathsf{P}(S_n/\sqrt{n} < x)$ — функция распределения нормированных сумм S_n, образованных независимыми случайными величинами g_k, равномерно распределённых на $[-\sqrt{3}, \sqrt{3}]$. Найти
$$\max_x |F_n(x) - \Phi(x)|$$
в зависимости от n и указать это значение для $n = 5$.
10. Вероятность появления события $p = 0.7$. Проводится $2n + 1$ независимых испытаний. Найти непосредственно (через биномиальное распределение) и через локальную теорему Муавра–Лапласа такое n, чтобы вероятность появления не менее $n + 1$ события была $0,9$.

11. В водохранилище объема 1 млн. кубометров находится 1 млрд. опасных бактерий. Берется проба объемом 10 литров. Какова вероятность, что найдем в пробе хоть 1 бактерию?
12. Решить упражнение из лекции про цепи Маркова.
13. Доказать состоятельность оценки для дисперсии.
14. Найти доверительный интервал для оценки вероятности уклонения биномиального распределения от его математического ожидания с надежностью $0,9$ в зависимости от числа опытов.
15. Доказать минимальность оценок для линейной регрессии. Будут ли эти оценки состоятельными и несмещенными?

10.2. Тесты на лекциях

10.2.1. Тест по первым двум разделам (Дискретное пространство элементарных событий. Аксиоматический подход к теории вероятностей)

Дать определение:

1) достоверного, невозможного и противоположного событий;
2) суммы, произведения и несовместных событий;
3) дискретного пространства элементарных событий;
4) суммы и произведения событий;
5) классической схемы и вероятности в ней;
6) последовательной выборки без возвращения и вероятности для неё;
7) выборки с возвращением и вероятности для неё;
8) одновременной выборки и вероятности для неё;
9) гипергеометрического распределения;
10) схемы Бернулли и биномиального распределения;
11) алгебры событий;
12) вероятностного пространства;
13) условной вероятности и независимости событий;
14) полной группы событий.

Сформулировать теорему:

1) о дискретной вероятности;
2) о свойствах вероятности;
3) о свойствах независимых событий;
4) о формулах полной вероятности и Байеса.

Решить задачу

1. В урне 5 чёрных, 3 красных и 2 белых шара. Какова вероятность, что в одновременной выборке 5 шаров ровно 3 чёрных шара и по одному красному и белому?
2. Для двух машин вероятность ложного срабатывания сигнализации $0,1$ и $0,05$, соответственно. Какова вероятность ложного срабатывания сигнализации только одной машины?
3. Вероятность разбития окна при игре во дворе футбол за один удар по мячу в сторону дома равна $0,1$. Было сделано 5 ударов. Какова вероятность, что окно останется целым?
4. В компьютерном классе 10 новых и 5 старых компьютеров. Отказали 2 компьютера. Какова вероятность, что среди них есть старый?

10.2.2. Тест по третьему разделу (Случайные величины)

Дать определение:

1) случайной величины;
2) функции распределения случайной величины;
3) дискретной случайной величины;
4) непрерывной случайной величины;
5) математического ожидания дискретной случайной величины;
6) математического ожидания непрерывной случайной величины;
7) дисперсии случайной величины;
8) ковариации двух случайных величин;
9) центрированной случайной величины;
10) нормированной случайной величины;
11) коэффициента корреляции двух случайных величин;
12) распределения Пуассона;
13) равномерного распределения;
14) нормального распределения;
15) показательного распределения.

Сформулировать теорему или привести:

1) о свойствах функции распределения;
2) о свойствах математического ожидания;
3) о свойствах дисперсии;
4) о свойствах коэффициента корреляции;
5) мат.ожидание и дисперсию для биномиального распределения;
6) мат.ожидание и дисперсию для распределения Пуассона;
7) мат.ожидание и дисперсию для равномерного распределения;

8) мат.ожидание и дисперсию для нормального распределения;
9) мат.ожидание и дисперсию для показательного распределения.

Решить задачу

1. Найти математическое ожидание и дисперсию числа вынутых чёрных шаров из урны с 2 чёрными и 3 белыми шарами при одновременной выборке 3 шаров.
2. Определить тип и найти математическое ожидание и дисперсию распределения с плотностью $\begin{cases} 1 & \text{при } x \in (1,2), \\ 0 & \text{при } x \notin (1,2). \end{cases}$
3. Найти математическое ожидание и дисперсию числа вынутых чёрных шаров из урны с 2 чёрными и 3 белыми шарами при выборке объёма 5 с возвращением.
4. Определить тип и найти математическое ожидание и дисперсию распределения с плотностью $\frac{1}{\sqrt{10\pi}} e^{-\frac{(t-5)^2}{10}}$.

10.2.3. Тест по четвёртому разделу (Предельные теоремы)

Написать формулу для:

1) гипергеометрического распределения;
2) схемы Бернулли;
3) биномиального распределения;
4) распределения Пуассона;
5) равномерного распределения;
6) нормального распределения;
7) показательного распределения.

Сформулировать теорему:

1) неравенство Чебышёва;
2) правило 3σ;
3) закон больших чисел в форме Чебышёва для произвольно распределённых случайных величин;
4) закон больших чисел в форме Чебышёва для одинаково распределённых случайных величин;
5) закон больших чисел для биномиальных распределений с одинаковой вероятностью успеха (закон Бернулли);
6) предельная теорема для гипергеометрического распределения;
7) теорема Пуассона (закон редких событий);
8) локальная теорема Муавра–Лапласа;

9) интегральная теорема Муавра–Лапласа;
10) центральная предельная теорема для одинаково распределённых случайных величин.

Решить задачу

1. Может ли в теореме Пуассона вероятности $p_n \to 10^{-100}$ при $n \to \infty$? Ответ обосновать.
2. Верно ли, что локальную теорему Муавра–Лапласа нельзя применить при вероятности успеха равной 1? Ответ обосновать.
3. Объяснить, почему нельзя применить центральную предельную теорему при нулевой дисперсии случайных величин последовательности и что тогда произойдёт?
4. Верно ли, что правило 3σ выполняется при $\sigma = 0$? Ответ обосновать.

10.3. Задачи для тестов на практических занятиях

Данный набор задач возник как некая кумуляция задач из задачника Гмурмана и других задачников, а также из электронных ресурсов.

10.3.1. Определение вероятности

1. Цифры 1, 2, 3, ..., 9, выписанные на отдельные карточки складывают в ящик и тщательно перемешивают. Наугад вынимают одну карточку. Найти вероятность того, что число, написанное на этой карточке чётное.
   а) $\frac{1}{2}$; б) $\frac{4}{5}$; в) $\frac{4}{9}$; г) $\frac{5}{9}$.
2. В круг радиусом $R = 3$ равномерно бросается точка. Найти вероятность события , заключающегося в попадании этой точки в круг радиусом $r = 1$ с тем же центром.
   а) $\frac{1}{3}$; б) $\frac{3}{\pi}$; в) $\frac{1}{9}$; г) $\frac{\pi}{9}$.
3. В урне 15 шаров: 5 белых и 10 черных. Какова вероятность вынуть из урны белый шар?
   а) $\frac{1}{2}$; б) $\frac{1}{3}$; в) $\frac{2}{3}$; г) $\frac{3}{10}$.
4. Из букв слова *УРАВНЕНИЕ* выбирается наугад одна буква. Какова вероятность, что эта буква будет согласной.
   а) $\frac{1}{2}$; б) $\frac{4}{5}$; в) $\frac{4}{9}$; г) $\frac{21}{33}$.
5. В семье трое детей. Какова вероятность того, что все они мальчики?
   а) $\frac{1}{2}$; б) $\frac{1}{3}$; в) $\frac{1}{4}$; г) $\frac{1}{8}$.
6. В ящике в 5 раз больше красных шаров, чем черных (шары одинаковы во всем за исключением цвета). Наугад выбирается один шар. Найти вероятность того, что он будет красным.
   а) $\frac{1}{5}$; б) $\frac{1}{6}$; в) $\frac{5}{6}$; г) $\frac{6}{11}$.

7. Абонент забыл последние 2 цифры телефонного номера, но помнит, что они различны и образуют двузначное число, меньшее 30. С учетом этого он набирает наугад 2 цифры. Найти вероятность того, что это будут нужные цифры.
   а) $\frac{1}{18}$; б) $\frac{2}{30}$; в) $\frac{1}{90}$; г) $\frac{1}{100}$; д) $\frac{27}{100}$.
8. В урне 10 шаров: 6 белых и 4 черных. Вынули два шара. Какова вероятность, что оба шара белые?
   а) $\frac{1}{3}$; б) $\frac{2}{3}$; в) $\frac{2}{10}$; г) $\frac{6}{10}$.
9. Трехзначное число образовано случайным выбором трех неповторяющихся цифр из набора 1, 2, 3, 4, 5, 6. Какова вероятность того, что это число делится на 5?
   а) $\frac{1}{2}$; б) $\frac{1}{6}$; в) $\frac{3}{5}$; г) $\frac{2}{3}$.
10. Трехзначное число образовано случайным выбором трех неповторяющихся цифр из набора 1, 2, 3, 4, 5. Какова вероятность того, что это число нечётное?
    а) $\frac{1}{2}$; б) $\frac{1}{5}$; в) $\frac{3}{5}$; г) $\frac{4}{10}$.
11. Из 10 теннисных мячей, среди которых 4 мяча новые, для очередной игры случайным образом берут 3. Какова вероятность того, что среди взятых мячей 2 мяча будут новыми?
    а) $\frac{1}{10}$; б) $\frac{3}{10}$; в) $\frac{4}{10}$; г) $\frac{1}{20}$.
12. В сумке лежат 10 мячей, пронумерованных от 1 до 10. Наугад вынимают два мяча. Какова вероятность, что это мячи с номерами 3 и 7?
    а) $\frac{2}{10}$; б) $\frac{1}{45}$; в) $\frac{1}{90}$; г) $\frac{1}{100}$.
13. Бросаются одна за другой две игральные кости. Какова вероятность того, что сумма выпавших очков равна 5?
    а) $\frac{1}{5}$; б) $\frac{1}{9}$; в) $\frac{1}{10}$; г) $\frac{1}{18}$.
14. Бросаются две игральные кости. Чему равна вероятность того, что число очков на одной кости в два раза больше числа очков на другой кости?
    а) $\frac{1}{2}$; б) $\frac{1}{6}$; в) $\frac{1}{12}$; г) $\frac{1}{36}$.
15. Колода из 36 карт хорошо перемешана, то есть все возможные распределения карт равновероятны. Найти вероятность события, что все четыре короля расположены рядом.
    а) $\frac{1}{9}$; б) $\frac{11}{595}$; в) $\frac{1}{1785}$; г) $\frac{1}{42840}$.

10.3.2. Теоремы о сложении и умножении вероятностей

1. В магазин поступила новая продукция с трех предприятий. Процентный состав этой продукции следующий: 20% — продукция первого предприятия, 30% — продукция второго предприятия, 50% — продукция третьего предприятия; далее, 10% продукции первого предприятия высшего сорта, на втором предприятии —5% и на третьем — 20% продукции

высшего сорта. Найти вероятность того, что случайно купленная новая продукция окажется высшего сорта.
а) $0,135$; б) $0,2$; в) $0,35$; г) $0,65$.

2. В группе спортсменов лыжников в 2 раза больше, чем бегунов, а бегунов в 3 раза больше, чем велосипедистов. Вероятность выполнить норму для лыжника $0,9$, для бегуна — $0,75$, для велосипедиста — $0,8$. Найти вероятность того, что спортсмен, выбранный наугад, выполнит норму.
а) $0,45$; б) $0,9$; в) $0,8$; г) $0,845$.
3. Из 1000 ламп 380 принадлежат к 1 партии, 270 — ко второй партии, остальные к третьей. В первой партии 4% брака, во второй — 3%, в третьей — 6%. Наудачу выбирается одна лампа. Определить вероятность того, что выбранная лампа — бракованная.
а) $0,0233$; б) $0,04$; в) $0,04434$; г) $0,13$.
4. В первой урне 4 белых и 6 чёрных шаров, а во второй — 5 белых и 3 чёрных. Из каждой урны наудачу извлекается один шар, а затем из этих двух наудачу берется один. Какова вероятность, что это будет белый шар?
а) $0,3125$; б) $0,5$; в) $0,5125$; г) $0,625$.
5. Из 5 стрелков 2 попадают в цель с вероятностью $0,6$ и 3 — с вероятностью $0,4$. С какой вероятностью наудачу выбранный стрелок попадет в цель?
а) $0,12$; б) $0,24$; в) $0,4$; г) $0,48$.
6. Три экзаменатора принимают экзамен по некоторому предмету у группы в 30 человек, причем первый опрашивает 6 студентов, второй — 3 студентов, а третий – 21 студента (выбор студентов производится случайным образом из списка). Отношение трёх экзаменаторов к слабо подготовившимся различное: шансы таких студентов сдать экзамен у первого преподавателя равны 40%, у второго — только 10%, у третьего — 70%. Найти вероятность того, что слабо подготовившийся студент сдаст экзамен.
а) $0,21$; б) $0,49$; в) $0,58$; г) $0,7$.
7. В первом ящике 1 белый и 5 черных шаров, во втором 8 белых и 4 чёрных шара. Из каждого ящика вынули по шару. Найти вероятность того, что один из вынутых шаров белый, а другой — чёрный.
а) $\frac{5}{36}$; б) $\frac{1}{4}$; в) $\frac{10}{324}$; г) $\frac{11}{18}$.
8. Фирма имеет три источника поставки комплектующих — фирмы A, B и C. На долю фирмы A приходится 50% общего объема поставок, B — 30% и C — 20%. Из практики известно, что среди поставляемых фирмой A деталей 10% бракованных, фирмой B — 5% и фирмой C — 6%. Какова вероятность, что взятая наугад деталь окажется бракованной?
а) $0,077$; б) $0,1$; в) $0,105$; г) $0,21$.

9. В магазин поступили электрические лампочки одного типа, изготовленные на четырех ламповых заводах: с 1-го завода 250 шт., со 2-го — 525 шт., с 3-го — 275 шт. и с 4-го — 950 шт. Вероятность того, что лампочка прогорит более 1500 часов, для 1-го завода равна $0,15$, для 2-го — $0,30$, для 3-го — $0,20$, для 4-го — $0,10$. При раскладке по полкам магазина лампочки были перемешаны. Какова вероятность того, что купленная лампочка прогорит более 1500 часов?
   а) $0,08625$; б) $0,1$; в) $0,1725$; г) $0,75$.
10. Студент разыскивает нужную ему формулу в трех справочниках. Вероятности того, что формула содержится в первом справочнике — $0,6$, во втором — $0,7$, в третьем — $0,8$. Найти вероятности того, что формула содержится только в двух справочниках.
    а) $0,487$; б) $0,9$; в) $0,452$; г) $0,667$.
11. Студент из 20 билетов подготовил к экзамену 12. Студент взял билет, к которому он не подготовился. Преподаватель в виде исключения разрешил взять второй билет. Какова вероятность того, что студенту во второй попытке достанется один из подготовленных билетов.
    а) $\frac{1}{12}$; б) $\frac{96}{380}$; в) $\frac{12}{19}$; г) $\frac{12}{20}$.
12. Студент подготовил к экзамену 20 билетов из 25. Какова вероятность взять известный билет, если студент пришел на экзамен вторым?
    а) $\frac{19}{25}$; б) $\frac{76}{125}$; в) $\frac{8}{10}$; г) $\frac{20}{25}$.
13. Имеются три колоды в 32 карты, две — в 36 и одна — в 52. Выбираем колоду и из нее одну карту. Какова вероятность того, что вынутая карта окажется тузом?
    а) $0,112$; б) $0,125$; в) $0,32$; г) $0,5$.
14. В ящике лежат 20 теннисных мячей, в том числе 15 новых и 5 играных. Для игры выбираются 2 мяча и после игры возвращаются обратно. Затем для второй игры также наудачу извлекаются ещё два мяча. Найти вероятность того, что вторая игра будет проводиться новыми мячами.
    а) $0,036$; б) $0,053$; в) $0,14$; г)$0,25$.
15. В урну, содержащую 3 шара, опущен белый шар, после чего наудачу извлечён один шар. Найти вероятность того, что он белый, если равновозможны все возможные предположения о первоначальном составе шаров.
    а) $0,125$; б) $0,25$; в) $0,5$; г) $0,625$.

10.3.3. Формула Байеса

1. Один из трех стрелков вызывается на линию огня и производит два выстрела. Вероятность попадания в мишень при одном выстреле для первого стрелка равна $0,3$, для второго — $0,5$; для третьего — $0,8$. Ми-

шень не поражена. Найти вероятность того, что выстрелы произведены первым стрелком.

а) 0,05625; б) 0,1875; в) 0,5; д) 0,63.

2. Из 30 стрелков 12 попадают в цель с вероятностью 0,6, 8 — с вероятностью 0,5 и 10 — с вероятностью 0,7. Наудачу выбранный стрелок произвел выстрел, поразив цель. Какова вероятность, что это стрелок из третей группы?

   а) 0,14; б) 0,333; в) 0,385; г) 0,7.

3. Предположим, что 5 мужчин из 100 и 25 женщин из 10000 являются дальтониками. Наугад выбранное лицо страдает дальтонизмом. Какова вероятность, что это мужчина?

   а) 0,05; б) 0,5; в) 0,0525; г(0,95.

4. Когда возникает аварийная ситуация звуковой сигнал срабатывает с вероятностью 0.95, звуковой сигнал может срабатывать случайно и без аварийной ситуации с вероятностью 0.05, реальная вероятность аварийной ситуации равна 0.004. Предположим, звуковой сигнал сработал. Чему равна вероятность реальной аварийной ситуации ?

   а) 0,0038; б) 0,004; в) 0,07; г) 0,08.

5. В одном ящике 3 белых и 5 черных шаров, в другом ящике — 6 белых и 4 черных шара. Из каждого ящика вынуто по одному шару. Были вынуты белый и чёрный шары. Какова вероятность, что белый шар вынут из первого ящика?

   а) 0,286; б) 0,375; в) 0,385; г) 0,985.

6. В трех ящиках содержатся, соответственно, две золотые, одна золотая и одна серебряная и две серебряные монеты. Случайным образом выбирается ящик и из него произвольно вынимается монета. Монета оказалась золотой. Какова вероятность, что вторая монета в этом ящике также золотая?

   а) $\frac{1}{6}$; б) $\frac{1}{3}$; в) $\frac{1}{2}$; г) $\frac{2}{3}$.

10.3.4. Дискретные случайные величины

1. Найти математическое ожидание случайной величины g.

g	-10	-1	8
P	$\frac{1}{6}$	$\frac{1}{3}$	$\frac{1}{2}$

   а) -3; б) 1; в) 2; г) 8.

2. Производят 4 независимых опыта, в каждом из которых некоторое событие появляется с вероятностью $=0,8$. Найти математическое ожидание случайной величины g — числа появлений события в 4 опытах.

   а) 0,2; б) 0,3072; в) 1,9712; г) 3,2.

3. Дан ряд распределения случайной величины g

g	-10	-1	8
P	$\frac{1}{6}$	$\frac{1}{3}$	$\frac{1}{2}$

Дисперсия заданной величины равна:
а) 9; б) 18; в) 45; д) 55.

4. Дан ряд распределения случайной величины g

g	1	3	4	7	9
P	0,1	0,2	0,1	0,3	0,3

Математическое ожидание заданной величины равно:
а) 4,8; б) 5,9; в) 7; г) 9.

5. Дан ряд распределения случайной величины g

g	1	3	4	7	9
P	0,1	0,2	0,1	0,3	0,3

Чему равна дисперсия заданной величины?
а) 7,69; б) 31,2; в) 42,5; г) 64.

6. Найти математическое ожидание дискретной случайной величины g, заданной законом распределения:

g	-4	6	10
P	0,2	0,3	0,5

а) 2; б) 4; в) 6; г) 10.

7. Найти дисперсию дискретной случайной величины g, заданной законом распределения:

g	-4	6	10
P	0,2	0,3	0,5

а) 14; б) 28; в) 50,67; г) 64.

8. Найти математическое ожидание дискретной случайной величины g, заданной законом распределения:

g	0,21	0,54	0,61
P	0,1	0,5	0,4

а) 0,27; б) 0,453; в) 0,535; г) 0,54.

9. Найти дисперсию дискретной случайной величины g, заданной законом распределения:

g	0,21	0,54	0,61
P	0,1	0,5	0,4

а) 0,012825; б) 0,2359; в) 0,29905; г) 0,4.

10. Дан ряд распределения случайной величины g

g	4	6	x_3
P	0,5	0,3	p_3

Найти x_3, зная, что $\mathsf{M}(g) = 8$.
а) 3,8; б) 10; в) 16; г) 21.

11. В партии из 10 деталей содержится три нестандартных. Наудачу отобраны две детали. Найти математическое ожидание дискретной случайной величины g — числа нестандартных деталей среди двух отобранных.
а) 0,3; б) 0,47; в) 0,6; г) 1.

12. Дан ряд распределения случайной величины g

g	4	5	6	8
P	0,5	0,1	0,2	0,2

Найти математическое ожидание.
а) 2; б) 4; в) 5,3; г) 5,75.

13. Дан ряд распределения случайной величины g

g	4	5	6	8
P	0,5	0,1	0,2	0,2

Найти дисперсию.
а) 2,41; б) 4; в) 30,5; г) 35,25.

10.3.5. Непрерывные случайные величины

1. Плотность распределения непрерывной случайной величины g задана следующим образом:

$$f(x) = \begin{cases} 1 + \frac{x}{2} & x \in [-2; 0] \\ 0 & x \notin [-2; 0] \end{cases}$$

Найти вероятность, что величина принадлежит интервалу $(-1; 1)$.
а) $\frac{1}{4}$; б) -1; в) $\frac{1}{2}$; г) $-\frac{3}{4}$.

2. Плотность распределения непрерывной случайной величины g задана следующим образом:

$$f(x) = \begin{cases} 1 + \frac{x}{2} & x \in [-2; 0] \\ 0 & x \notin [-2; 0] \end{cases}$$

Найти математическое ожидание.
а) $-\infty$; б) $-\frac{5}{4}$; в) $-\frac{2}{3}$; г) 1.

3. Функция распределения непрерывной случайной величины g задана следующим образом:

$$F(x) = \begin{cases} 0 & x \leqslant 0 \\ x^2 & 0 < x \leqslant 1 \\ 1 & x > 1 \end{cases}$$

Найти вероятность попадания случайной величины g в интервал $(0,25; 0,5)$.
а) $0,0365$; б) $0,1875$; в) $0,25$; г) $0,5$.

4. Непрерывная случайная величина g задана плотностью распределения вероятностей

$$f(x) = \begin{cases} 0 & |x| > 1 \\ 1 + x & -1 < x \leqslant 0 \\ 1 - x & 0 < x \leqslant 1 \end{cases}$$

Найти математическое ожидание.
а) -1; б) $-\frac{1}{2}$; в) 0; г) $\frac{1}{2}$; д) 1.

5. Непрерывная случайная величина g задана плотностью распределения вероятностей

$$f(x) = \begin{cases} 0 & |x| > 1 \\ 1 + x & -1 < x \leqslant 0 \\ 1 - x & 0 < x \leqslant 1 \end{cases}$$

Найти вероятность, что g будет больше $0,5$.
а) $\frac{1}{8}$; б) $\frac{1}{4}$; в) $\frac{1}{2}$; г) $\frac{2}{3}$.

6. Непрерывная случайная величина g задана функцией распределения

$$F(x) = \begin{cases} 0 & x \leqslant 0 \\ \frac{x^2}{4} & 0 < x \leqslant 2 \\ 1 & x > 2 \end{cases}$$

Найти вероятность, что величина будет меньше 1.
а) $-\frac{1}{4}$; б) $-\frac{1}{12}$; в) 0; г) $\frac{1}{12}$; д) $\frac{1}{4}$.

7. Непрерывная случайная величина g задана функцией распределения

$$F(x) = \begin{cases} 0 & x \leqslant 0 \\ \frac{x^2}{4} & 0 < x \leqslant 2 \\ 1 & x > 2 \end{cases}$$

Найти математическое ожидание.
а) -1; б) $-\frac{1}{4}$; в) 0; г) $\frac{1}{4}$; д) 1.

8. Непрерывная случайная величина g задана функцией распределения

$$F(x) = \begin{cases} 0 & x \leqslant 1 \\ \frac{x^2-1}{2} & 1 < x \leqslant 2 \\ 1 & x > 2 \end{cases}$$

Найти дисперсию.
а) $\frac{13}{6}$; б) $-\frac{49}{144}$; в) $\frac{17}{169}$; г) $-\frac{169}{36}$.

9. Непрерывная случайная величина g задана функцией распределения

$$F(x) = \begin{cases} 0 & x \leqslant 1 \\ \frac{x^2-1}{2} & 1 < x \leqslant 2 \\ 1 & x > 2 \end{cases}$$

Найти математическое ожидание.
а) $\frac{5}{12}$; б) $\frac{7}{14}$; в) 1; г) $\frac{19}{12}$.

10. Непрерывная случайная величина g задана функцией распределения

$$F(x) = \begin{cases} 0 & x \leqslant 1 \\ x-1 & 1 < x \leqslant 2 \\ 1 & x > 2 \end{cases}$$

Найти математическое ожидание.
а) $\frac{1}{8}$, б) $\frac{1}{2}$, в) $\frac{2}{3}$, г) $\frac{3}{2}$.

11. Непрерывная случайная величина g задана функцией распределения

$$F(x) = \begin{cases} 0 & x \leqslant 1 \\ x-1 & 1 < x \leqslant 2 \\ 1 & x > 2 \end{cases}$$

Найти вероятность, что величина принадлежит интервалу $(-\frac{3}{2}, \frac{3}{2})$.
а) $\frac{1}{8}$; б) $\frac{1}{2}$; в) $\frac{5}{4}$; г) 3.

12. Непрерывная случайная величина g задана плотностью распределения вероятностей

$$f(x) = \begin{cases} 0 & x \leqslant 1 \\ \frac{a}{x^2} & x > 1 \end{cases}$$

Найти коэффициент a.
а) $-\frac{1}{4}$; б) $\frac{1}{3}$; в) 1; г) 2.

13. Непрерывная случайная величина g задана плотностью распределения вероятностей

$$f(x) = \begin{cases} 0 & x \leqslant 1 \\ \frac{1}{x^2} & x > 1 \end{cases}$$

Найти функцию распределения при $x > 1$.
а) $1 - \frac{1}{x}$; б) $\frac{1}{x}$; в) $-\frac{2}{x}$; г) 1.

14. Плотность распределения вероятностей случайной величины g

$$f(x) = \begin{cases} 0 & x \leqslant 1, x > 3 \\ a(x-1) & 1 < x \leqslant 3 \end{cases}$$

Найти коэффициент a.
а) $\frac{1}{2}$; б) $\frac{2}{3}$; в) 1; г) 2.

15. Плотность распределения вероятностей случайной величины g

$$f(x) = \begin{cases} 0 & x \leqslant 1, x > 3 \\ \frac{x-1}{2} & 1 < x \leqslant 3 \end{cases}$$

Найти вероятность того, что величина примет значение из интервала $(-1; 2)$.
а) $\frac{1}{4}$; б) $\frac{1}{3}$; в) $\frac{1}{2}$; г) $\frac{3}{4}$.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1.* Боровков, А.А. Математическая статистика: учеб. пособие для мат. и физ. специальностей вузов / А.А. Боровков. – 4-е изд.,стер. – СПб.: Лань, 2010. – 703 с.

2.* Боровков, А.А. Теория вероятностей: учеб. пособие для вузов по направлениям 010100 "Математика" / А.А. Боровков. – 5-е изд.,сущ. перераб. и доп. – М.: URSS: Либроком, 2009. – 652 с. : ил.

3.* Вентцель, Е.С. Теория вероятностей: учеб. для высш. техн. учеб. заведений / Е.С. Вентцель. – 11-е изд., стер. – М.: КноРус., 2010. – 658 с.: ил.

4.* Глаголев, А.А. Курс высшей математики: для экон. специальностей вузов / А.А. Глаголев, Т.В. Солнцева. – 2-е изд., перараб. и доп. – М.: Высш. шк., 1971. – 654 с.: черт.

5.* Гмурман, В.Е. Руководство к решению задач по теории вероятностей и математической статистике: учеб. пособие для студентов вузов / В.Е. Гмурман. – 11-е изд., стер. – М.: Юрайт, 2010. – 403 с.: ил.

6.* Гмурман, В.Е. Теория вероятностей и математическая статистика: учеб. пособие для вузов / В.Е. Гмурман. – 12-е изд., стер. – М.: Юрайт., 2014. – 478 с.: ил.

7. Дрейпер, Н. Прикладной регрессионный анализ Кн. 1: В 2 кн. / Н. Дрейпер, Г. Смит. Пер. с англ. Ю.П. Адлера, В.Г. Горского – 2-е изд., пераpаб. и доп. – М.: Финансы и статистика., 1986. – 365 с.: ил.

8. Заляпин, В. И. Математическая статистика: учеб. пособие / В.И. Заляпин, Е.В. Харитонова; Юж.-Урал. гос. ун-т, Каф. Мат. анализ; ЮУрГУ. – Челябинск: Издательство ЮУрГУ, 2008. – 146 с.

9. Колемаев, В.А. Теория вероятностей и математическая статистика: учеб. пособие для вузов по направлению и специальности "Менеджмент" / В.А. Колемаев. – М.: ИНФРА-М, 1999. – 300 с.

10.* Колемаев, В.А. Теория вероятностей и математическая статистика: учеб. пособие для вузов по экон. специальностям / В.А. Колемаев, В.Н. Калинина. – 3-е изд., перараб. и доп. – М.: КноРус, 2009. – 375 с.: ил.

11.* Сборник задач по теории вероятностей, математической статистике и теории случайных функций: учеб. пособие для вузов / Б.Г. Володин и др.; под общ. ред. А.А. Свешникова. – 5-е изд.,стер. – СПб.: Лань, 2013. – 445 с.: ил.

12.# Электронные ресурсы:
`http://www.matburo.ru/tv_book.php`
`http://ido.aspu.ru/COURSES/course113/mods/1675/prak6.htm`
`http://cito-web.yspu.org/link1/metod/theory/node10.html`

* – можно использовать другие издания.

\# – можно использовать другие ресурсы.

Учебное издание

Алеев Рифкат Жалялович,
Молодорич Маргарита Ивановна

ТЕОРИЯ ВЕРОЯТНОСТЕЙ
И МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

Учебное пособие

Техн. редактор *А.В. Миних*

Издательский центр Южно-Уральского государственного университета

Подписано в печать 10.12.2015. Формат 60×84 1/16. Печать цифровая.
Усл. печ. л. 11,86. Тираж 60 экз. Заказ 709/35.

Отпечатано в типографии Издательского центра ЮУрГУ.
454080, г. Челябинск, пр. им. В.И. Ленина, 76.