Математическое Образование

Журнал Фонда математического образования и просвещения

Год седьмой

№3 (26)

Июль - сентябрь 2003 г.

Москва

МАТЕМАТИЧЕСКОЕ ОБРАЗОВАНИЕ

Периодическое издание в области математического образования

Учредитель: Фонд математического образования и просвещения

№ 3 (26), 2003 г.

Москва

Математическое образование

Журнал Фонда математического образования и просвещения

№ 3 (26), июль – сентябрь 2003 г.

Содержание

Фонд математического образования и просвещения, Москва, 2003 г.
“Математическое образование”, периодическое издание.
Зарегистрировано в Роскомпечати РФ,
лицензия №015955 от 15.04.97.
Подписано к печати 21.09.2003 г.
Компьютерный набор и верстка, компьютерная графика: С. Кулешов.
Объем 5,5 п.л. Тираж 1000 экз. Цена свободная.

К 95-летию Л. С. Понтрягина

Предисловие (или послесловие)

А. И. Понтрягина

В этом году исполнилось 95 лет со дня рождения крупнейшего русского советского математика академика Льва Семеновича Понтрягина. Предлагаем вниманию читателей воспоминания вдовы Льва Семеновича, первоначально подготовленные в качестве предисловия (послесловия) к его жизнеописанию, написанному им самим.

Лев Семенович Понтрягин своей жизнью, деятельностью - профессиональной, общественной, духовной, чисто человеческой — сумел ярко выразить нам то, на что способен Человек! Мы не используем и десятой доли того, что Богом дано Человеку.

В нем гармонично сочетались сила и кротость, обаяние и непреклонность, ангельская доброта и предельная независимость, словом все то, что можно выразить одним словом — гений. Гений — явление не простое.

Живя с ним одной жизнью, его горестями, муками, его радостью и даже ликованием, если он достигал желаемого результата, я иногда наблюдала его со стороны (не сказывалась ли у меня моя профессия врача?), я восхищалась, любовалась Львом Семеновичем, счастлива была, что я жена, подруга такого удивительного человека. Но скажу откровенно, порой мне становилось страшно: какова же мощь напора всех человеческих сил - духовных, душевных, умственных, наконец, физических, которые он бесстрашно пускал в ход, чтобы достигнуть желаемого результата в математике или общественной деятельности, или при иных жизненных ситуациях. Тут и мне уже было не до радости.

Речи не могло быть, чтобы он что-нибудь отложил "на потом" или счел, что пусть кто-то другой занимается каким-либо мало привлекательным, а то и опасным, делом. Все сам. Но крылья-то обрублены, он же абсолютно не видит окружающего мира (порой страшного). О какая трагедия! Какая ирония судьбы!

Трагедия не привела к сердечной окаменелости, к душевной слепоте. А это — главное! Круг интересов его, наоборот, расширился. Он любил жизнь, любил музыку, театр, танцы. К застольям относился равнодушно. Чаще избегал их, утомлялся. Не любил также длительный заседаний. Иногда спал на собраниях. Лев Семенович страдал жестокой бессонницей. Его бессонница доставляла нам много страданий...

Победа духа была столь велика, что мир Понтрягина — этого русского гиганта — был светлым и теплым, и многие, многие согревались этим светом и теплом. В том числе и я.

Живя математической жизнью Льва Семеновича (с медицинским миром я уже не общалась, точнее, почти не общалась), я замечала, что математики не делают никаких скидок Понтрягину из-за того, что он слепой — нет! — на равных требования, эмоции, выражения и выплескивания всех человеческих слабостей и чувств — на равных. Особенно это чувствовалось при обсуждении работ на соискании Государственных и Международных премий, наград и вообще почестей земных, которые людям очень и очень по душе. Скажу твердо, что Понтрягина почести особенно не волновали (если касались его лично). Но авторитет Понтрягина в математике и общественной жизни был непререкаем.

В тридцатилетнем возрасте он был одним из лучших кандидатов в академики, выбран же был только в члены-корреспонденты. Он не огорчился. Считал — и то здорово. А некоторые ученые при подобных обстоятельствах обижались, искали и находили своих врагов, тяжело страдали и даже получали тяжелые заболевания. Мне известен случай смерти ученого, которого не выбрали в членыкорреспонденты. Приехал к себе на дачу и в тот же вечер умер.

Причиной не избрания Понтрягина в академики в 1939 году было повышенное чувство справедливости у Льва Семеновича. Он вздумал критиковать ЦК партии за превышение своих полномочий. По наивному мнению Л. С.Понтрягина, Центральный Комитет партии не должен вмешиваться в выборы в Академию Наук СССР. Это — дело ученых. А Л.С. Понтрягину стало известно, что такое вмешательство произошло. Вот он и выступил открыто на собрании Московского математического общества. Своей краткой, четкой речью осудил ЦК и местную партийную организацию. Сделал это в период избирательной компании. Ну и получил, что называется "по заслугам". В течение 19 лет его не избирали в академики. Хотя имя Понтрягина приобрело мировую славу уже тогда, в 30-ые годы.

После вышеупомянутого собрания его спрашивали: "Ну как? Тебя еще не посадили?". Не говорит ли этот эпизод о широте его души, о бескорыстном служении науке. Лев Семенович не ожесточился и даже не чувствовал себя обделенным, продолжал упорно работать и получать блестящие результаты в области математики. Он не иссох душой, хотя общественная и бытовая атмосфера была тяжелой. 1937 год известен своими репрессиями, затем 1941—1945 годы — тяжелые испытания для России. Эвакуация в Казань, неудавшаяся женитьба, тяжелый характер матери. Все это с предельной откровенностью описано Львом Семеновичем в своем жизнеописании.

По моей неотступной, настоятельной просьбе были написаны эти страницы жизнеописания. Если их читать не предвзято, то поражаешься широте математических, общественных, бытовых и человеческих интересов. Он не вел дневников. До этого ли было?! Однако события, которые описаны, предстают перед нами, как будто автор только что пережил их — живо, красочно.

Мир Понтрягина — это Космос, где все гармонично, все осмысленно, все неспроста. Его видение жизни, предметов, видение математики, ее проблем, направлений было видением его Разума, которым он был наделен Богом. Он видел жизнь, людей не сверху, не по поверхности, не извне, а изнутри. Поэтому я никогда не чувствовала в нем беспомощности, бессилия. Я не была его поводырем. Нет! Это он вел меня по жизни, которая была насыщенна, красива, радостна и бесконечно щемяще горька... С ним никогда не было скучно. Даже в старости. Мой муж умер, не дожив 4-х месяцев до 80 лет.

С ним всегда было интересно поговорить, посоветоваться. К его слову прислушивались и принимали за истину. Сила его слова была велика и мне, чисто по-женски, иногда так хотелось бы ему возразить, но это было почти не под силу мне (такова была сила его логики), и тогда я злилась... но недолго, широко открывала глаза и смотрела на него, стараясь понять его и себя... Иногда я говорила ему в сердцах: "Ты лучший представитель homo sapiens", стараясь придать этому смысл ругательства. И он на какой-то момент становился таким кротким, задумчивым, как ребенок, который старается понять, в чем он провинился.

В тайнике души, конечно, я понимала, что я живу с человеком исключительным. Мне всегда хотелось его понять, даже на самом житейском уровне. Понять и поддержать. Последнее больше относится к его околоматематическим, академическим, общественно-политическим интересам. Об этом особый разговор.

Весь наш быт — ремонт квартиры в Москве, ремонт дачи, прокладка отопительных, водопроводных труб, их способы изоляции, строительство забора на дачном участке, устройство кухни — буквально все, все его интересовало и ничего не предпринималось без его совета, без его участия. Причем советовались с ним не только я (что вполне естественно), но и специалисты, которые были приглашены выполнять эти работы.

Вспоминается случай: на даче меняли электропроводку. Был приглашен хороший, опытный электрик. Однако, когда дело дошло до распределительного щитка, который лежал на полу и монтировался, и к которому нужно было присоединить отдельную проводку для холодильников (которые оставались бы включенными при нашем отъезде в Москву, а при этом весь свет выключался бы), а также к каждой паре из множества розеток (почему пара? — да потому, что в одной розетке ток был постоянно включен, а в другой — только при помощи отдельного тумблера, по желанию Льва Семеновича), для терморегулятора (который служил для поддержания нужной температуры в доме), для счетчика и т.д., то, конечно, электрик запутался. Я хорошо вижу, как мой муж, с янтарными четками в руках, энергично перебирая их, рассказывал электрику куда какой провод надо присоединить. Он отлично представлял в уме всю эту запутанную схему. Электрик был изумлен настолько, что и путался он больше от того, что все время, сидя на полу, снизу вверх смотрел на Льва Семеновича и старался понять и суть дела, и этого человека. Конечно, в таких случаях не обходилось дело без меня, как посредника, "третьего глаза" моего мужа, которым он широко пользовался.

Его видение было всегда точным, осмысленным, одухотворенным и возвышенным. Он всегда хотел быть полезным людям.

Вспомним его знаменитый "Принцип максимума Понтрягина" - так он назван в науке. "Принцип максимума" применяется в математике (где он открыл целое направление), в механике, аэродинамике, астронавтике, физике, химии, медицине и даже биологии.

А теперь вглядитесь в его лицо на фотографии[1] — какие сила, покой, непоколебимость, предельная независимость и доброта. От него всегда исходила высшая полнота и любовь к жизни.

Мне хотелось бы здесь привести слова Ф.М.Достоевского: "Смирись, гордый человек, и прежде всего потрудись на родной ниве. Не вне тебя правда, а в тебе самом. Найди себя в себе, подчини себя себе, овладей собой — и узришь правду. Не в вещах эта правда, не вне тебя и не за морем где-нибудь, а прежде всего в твоем собственном труде над собой. Победишь себя, усмиришь себя — и станешь свободен, как никогда и не воображал себе, и начнешь великое дело, и других свободными сделаешь, и узришь счастье, ибо наполнится жизнь твоя, и поймешь, наконец, народ свой и святую правду его. Не у цыган и нигде - мировая гармония, если ты сам не достоин, злобен и горд и требуешь жизни даром, даже и не предполагая, что за нее надобно платить". (Ф.М.Достоевский "Речь на открытии памятника А.С.Пушкину" в Москве, 6 июня 1880 г.) Чудные слова!

Большой милостью судьбы, выпавшей мне на долю, оказалось знакомство с Львом Семеновичем, перешедшее в дружбу, а затем в крепкую любовь и преданность друг другу.

А.И.Понтрягина
30 октября 1996 г.

[1] Снимок помещен на стр. 2

Учебное пособие в журнале

Введение в вероятностное прогнозирование. Курс лекций и упражнений (продолжение)

И. П. Костенко

Продолжаем публикацию учебного пособия по теории вероятностей. В данном номере публикуются лекции 8 с соответствующими упражнениями. Лекции 6, 7 опубликованы в номере 2(25) за 2003 г.

Лекция 8. Пуассоновские случайные величины

Лекция посвящена исследованию второго практически ценного класса дискретных с.в. — Пуассоновских (П.с.в.). Они похожи на биномиальные, но возникают в других ситуациях, — не только при повторении опытов. Более того, в отличие от биномиальных, для П.с.в. не удается четко охарактеризовать все условия их возникновения. Один важный класс опытов, порождающих П.с.в., — так называемые *потоки событий*, мы рассмотрим во второй части лекции.

Основные вопросы, которые ставятся при изучении с.в.: характер распределения вероятностей, его зависимость от параметров, числовые характеристики, условия возникновения. Эти вопросы направляли наше исследование биномиальных с.в. Эти же вопросы будут определять в будущем исследование основных типов непрерывных с.в. (лек. 11).

Начну лекцию не с примеров, как обычно, а с абстрактного определения и его следствий — динамики изменения распределения в зависимости от параметра, формул математического ожидания и дисперсии. В отличие от биномиальных с.в., здесь эти вопросы решаются легко. А вопрос об условиях возникновения Пуассоновских с.в. решается гораздо труднее. Вот по этой причине я и начинаю с отвлеченных рассуждений и вычислений. Примеры будут позднее.

1. Математическое задание класса Пуассоновских с.в. (П.с.в.)

Случайной величиной (с.в.) мы назвали переменную величину, связанную с определенным опытом (лек. 5, п. 2). Значения этой величины появляются в результате опыта случайным образом. Следовательно, чтобы задать конкретную с.в., надо описать условия опыта и определить, как возникают значения с.в., какие они. Чтобы задать класс с.в., надо сделать то же самое, только в общем виде. Именно так определялись биномиальные с.в. в предыдущей лекции.

Однако, подобный способ задания **класса** с.в. не всегда возможен. Не всегда удается описать единую структуру опытов, в которых появляются с.в. определенного типа. В таких случаях можно применить абстрактно-математический способ задания с.в., а именно, указать типичный ряд распределения всех с.в. данного класса (лек. 5, п. 5). Так и поступим.

Определение 1. Дискретная с.в. X_{Π} называется *Пуассоновской (П.с.в.)*, если она может принимать только целые неотрицательные значения x_l; $l = 0, 1, 2, 3, \ldots$, а вероятности p_l этих значений определяются так:

$$p_l = \frac{a^l}{l!} \cdot e^{-a}, \ l = 0, 1, 2, 3, \ldots, \tag{1}$$

где $a > 0$ имеет определенное положительное значение (свое для каждой конкретной П.с.в.) и называется *параметром* П.с.в.

Данным определением и задается типичный ряд распределения П.с.в.:

Таблица 1

X_{Π}	0	1	2	3	$\ldots$	l	$\ldots$
p_l	e^{-a}	$\frac{a}{1!} \cdot e^{-a}$	$\frac{a^2}{2!} \cdot e^{-a}$	$\frac{a^3}{3!} \cdot e^{-a}$	$\ldots$	$\frac{a^l}{l!} \cdot e^{-a}$	$\ldots$

Проверка свойства р.р. Не забыли ли вы, что вероятности p_l значений с.в. должны подчиняться требованию $p_0 + p_1 + p_2 + \ldots + p_l + \ldots = 1$? Это основное свойство любого ряда распределения (лек. 5, п. 5,). Проверим, выполняется ли оно в нашем случае. Только после этого можно будет считать определение 1 корректным (правильным).

1. В сущности, нам надо найти сумму ряда

$$\sum_l p_l = e^{-a} + \frac{a}{1!} \cdot e^{-a} + \frac{a^2}{2!} \cdot e^{-a} + \frac{a^3}{3!} \cdot e^{-a} + \ldots + \frac{a^l}{l!} \cdot e^{-a} + \ldots.$$

Этот ряд можно немного видоизменить, вынеся константу e^{-a} за скобки:

$$\sum_l p_l = e^{-a} \cdot \left(1 + \frac{a}{1!} + \frac{a^2}{2!} + \frac{a^3}{3!} + \ldots + \frac{a^l}{l!} + \ldots\right). \tag{2}$$

2. Найдем сумму ряда, стоящего в скобках. Для этого придется вспомнить формулу разложения функции e^x в степенной ряд Маклорена:

$$e^x = 1 + \frac{x}{1!} + \frac{x^2}{2!} + \frac{x^3}{3!} + \ldots + \frac{x^l}{l!} + \ldots, \ \ x \in (-\infty; +\infty) \tag{3}$$

Полагая в этой формуле $x = a$, получим:

$$e^a = 1 + \frac{a}{1!} + \frac{a^2}{2!} + \frac{a^3}{3!} + \ldots + \frac{a^l}{l!} + \ldots.$$

3. Подставляем в (2) вместо ряда, стоящего в скобках, его сумму e^a:

$$\sum_l p_l = e^{-a} \cdot e^a = 1.$$

Итак, определение 1 действительно корректно задает класс случайных величин, названных Пуассоновскими.

Примечания. Я понимаю, что в глубине души у вас возникло недоумение: откуда взялась формула (1)? Пока могу сказать, что ее открыл француз Пуассон, и ее называют *формулой Пуассона*. Она хорошо отражает многие реальные распределения вероятностей в природе. В конце лекции (п. 10) вы увидите, как она возникает из анализа некоторого класса опытов — потоков событий.

Вы, конечно, вспомнили, что формула Пуассона встречалась раньше (лек. 4, п. 5-6). Она помогала решать задачу Бернулли при очень малых вероятностях "успеха" опыта — p. В сущности, это значит, что Пуассоновское распределение X_Π порождается также биномиальным X_σ при очень малых значениях параметра p. Подробнее об этом поговорим позже.

Контроль 1. Запишите ряд распределения П.с.в. с параметром $a = 1$. Вычислите вероятности p_l, $l = 0, 1, 2, 3, \ldots$, считая $e^{-1} \approx 0,3679$. Постройте многоугольник распределения. Оцените расположение математического ожидания. Определите промежуток значений П.с.в., в который будут попадать значения П.с.в. с практически достоверной вероятностью 0,997.

2. Как распределение П.с.в. зависит от параметра

Глядя на таблицу 1, трудно представить характер изменения вероятностей p_l. Можно предположить, что они быстро убывают из-за быстрого роста факториалов, стоящих в знаменателе. Так ли это?

Давайте придадим параметру a различные числовые значения, например, $a = 0,2$, $a = 0,6$, $a = 1$, $a = 2$, $a = 5$, $a = 9$, вычислим для каждого значения a несколько вероятностей p_1 и конкретизируем таблицу 1. После этого изобразим многоугольники распределения и проанализируем характер изменения вероятностей, в зависимости от значения параметра a.

Распределение при $a = 0,2$. Подставляем во вторую строку таблицы 1 значение параметра $a = 0,2$ и получаем вероятности:

$$\begin{aligned}
&p_0 = e^{-0,2};\\
&p_1 = 0,2 \cdot e^{-0,2} = 0,2 \cdot p_0;\\
&p_2 = (0,2^2/2) \cdot e^{-0,2} = 0,02 \cdot p_0;\\
&p_3 = (0,2^3/6) \cdot e^{-0,2} \approx 0,0013 \cdot p_0.
\end{aligned}$$

Искать последующие вероятности p_4, p_5, : не имеет смысла, ибо очевидно, что они очень малы.

Вычисления сводятся к вычислению $e^{-0,2}$. Это число, приближенно равное $1/5\sqrt{2,7183}$. Его можно найти в таблице значений функции Пуассона (приложение 2) — $e^{-0,2} \approx 0,8187$. После этого искомые вероятности вычисляются так:

$$p_0 = e^{-0,2} \approx 0,8187;$$
$$p_1 = 0,2 \cdot p_0 \approx 0,2 \cdot 0,8187 \approx 0,1637;$$
$$p_2 = 0,02 \cdot p_0 \approx 0,02 \cdot 0,8187 \approx 0,0164;$$
$$p_3 = 0,0013 \cdot p_0 \approx 0,0013 \cdot 0,8187 \approx 0,0011.$$

Ряд распределения П.с.в. $X_\text{П}(a = 0,2)$ получается таким:

Таблица 2

$X_\text{П}(a = 0,2)$	0	1	2	3	...
p	0,8187	0,1637	0,0164	0,0011	...

Наше предположение оправдывается — вероятности резко убывают.

Замечание. Таблицу 2 можно было получить сразу из приложения 2. Но я хотел заставить вас произвести сам процесс вычислений. Значение $e^{-0,2}$, взятое нами из приложения, тоже рассчитывается несложно. Просмотрите, как это делается с помощью рядов, и вы с удовольствием почувствуете полезность ваших давних знаний.

Подставьте значение $x = -0,2$ в ряд Маклорена (2), получится знакочередующийся числовой ряд, сумма которого известна:

$$e^{-0,2} = 1 - 0,2 + \frac{0,2^2}{2} - \frac{0,2^3}{6} + \frac{0,2^4}{24} - \ldots \approx 1 - 0,2 + 0,02 - 0,0013 + 0,00007 - \ldots .$$

Теперь отбросьте все члены ряда, начиная с 4-го, и получите приближенное значение $e^{-0,2}$:

$$e^{-0,2} \approx 1 - 0,2 + 0,02 = 0,82.$$

Почему отброшен именно 4-й член ряда? Потому, что он очень мал, и он оценивает ошибку — $|e^{-0,2} - 0,82| < 0,0013$, т.е. найденное нами значение 0,82 отличается от истинного значения $e^{-0,2}$ менее, чем на 0,02 (сравните с более точным значением 0,8187, взятым из таблицы). Надеюсь, вы вспомнили, что такой метод оценки ошибки от замены суммы ряда частичной суммой был установлен в теории знакочередующихся рядов.

Распределения при $a = 0,6$, $a = 1$ не будем подробно рассчитывать, а возьмем из приложения 2:

Таблица 3а

$X_\text{П}(a = 0,6)$	0	1	2	3	4	5	...
p	0,5488	0,3293	0,0988	0,0198	0,0030	0,0004	...

Таблица 3б

$X_{\Pi}(a=1)$	0	1	2	3	4	5	6	7	...
p	0,3679	0,3679	0,1839	0,0613	0,0153	0,0031	0,0006	0,0001	...

Вероятности тоже быстро убывают, но скорость убывания замедляется с ростом параметра a. Это наглядно видно из рис. 1, где представлены многоугольники только что полученных распределений.

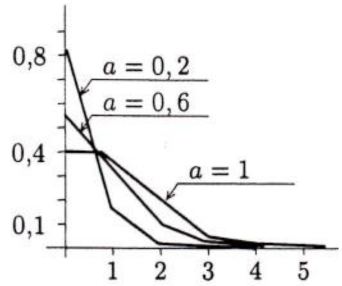


Рис. 1

Вывод 1. Если $0 < a \leq 1$, то многоугольник распределения П.с.в. $X_{\Pi}(a)$ имеет вид "обрыва". С ростом параметра a от 0 до 1 высота "обрыва" уменьшается и он становится более пологим.

Распределения для $a = 2$, $a = 5$, $a = 9$ показаны на рис. 2.

Вывод 2. Если $a > 1$, то многоугольник распределения П.с.в. $X_{\Pi}(a)$ имеет вид несимметричной "горки". С ростом a "горка" сдвигается вправо, расширяется и понижает высоту.

Сходство с биномиальными распределениями вы, наверное, уже заметили. Сравните рис. 2, лек. 7 и рис. 1 данной лекции: видно, что при очень малых p биномиальный "обрыв" похож на Пуассоновский. Сравните теперь рис. 3, лек 7 и рис. 2 данной лекции: видно, что при больших значениях другого параметра k биномиальная "горка" тоже становится похожа на Пуассоновскую. Чуть позже (в разделе 5) мы уточним условия этого сходства.

Рис. 2

Контроль 2. Нарисуйте эскиз многоугольника распределения П.с.в. для значения параметра $a = 3$. Пользуясь таблицей приложения 2, составьте ряд распределения для $a = 3$ и постройте многоугольник распределения. Согласуется ли он с вашим эскизом?

3. Математическое ожидание и дисперсия П.с.в.

Не обратили ли вы внимание на то, что вершина "горки" (рис. 2) находится над значением параметра a? Т. е. наиболее вероятные значения П.с.в. X_{Π} располагаются около a. Это признак того, что математическое ожидание X_{Π} должно быть близко к a.

Из рис. 2 можно сделать предположение и относительно дисперсии. Поскольку с ростом параметра a "горка" растягивается, то дисперсия $D(X_{\Pi})$ должна увеличиваться. Проверим эти предположения, вычислив M и D точно.

Вывод формулы $M(X_{\Pi})$. Общая формула математического ожидания д.с.в. имеет вид:

$$M = \sum_i x_i \cdot p_i.$$

где x_i — значения с.в., p_i — их вероятности.

Пуассоновская с.в. X_{Π} имеет бесконечное (счетное) множество значений — $l = 0, 1 2, 3, \ldots$, их вероятности задаются определением 1 (формула (1)), значит, математическое ожидание с.в. X_{Π} есть сумма ряда:

$$M(X_{\Pi}) = \sum_{l=0}^{\infty} l \cdot \frac{a^l}{l!} \cdot e^{-a}.$$

Первое слагаемое ряда получается при $l = 0$, значит, оно равно нулю и суммирование можно начинать с $l = 1$:

$$M(X_{\Pi}) = \sum_{l=1}^{\infty} l \cdot \frac{a^l}{l!} \cdot e^{-a}.$$

Согласно определению факториала ($l! = 1 \cdot 2 \cdot 3 \cdot \ldots \cdot (l-1) \cdot l$), в каждом l-м слагаемом ряда можно произвести сокращение на l. После этого сокращения и вынесения общего множителя ae^{-a} за знак суммы, получим:

$$M(X_{\Pi}) = \sum_{l=1}^{\infty} l \cdot \frac{a^l}{l!} \cdot e^{-a} = \sum_{l=1}^{\infty} l \cdot \frac{a^{l-1}}{1 \cdot 2 \cdot 3 \cdot \ldots \cdot (l-1) \cdot l} \cdot ae^{-a} = ae^{-a} \cdot \sum_{l=1}^{\infty} \frac{a^{l-1}}{l-1}!.$$

Развернем сумму, стоящую справа, придавая l значения $l = 0, 2, 3, \ldots$ (учтите, что $0! = 1$):

$$M(X_{\Pi}) = ae^{-a} \cdot \left(1 + \frac{a}{1!} + \frac{a^2}{2!} + \frac{a^3}{3!} + \ldots\right).$$

Ряд, стоящий в скобках, получается из ряда (3) при $x = a$, значит, его сумма равна e^a и, следовательно,

$$M(X_{\Pi}) = ae^{-a} \cdot e^a.$$

Поскольку $e^{-a} \cdot e^{a} = \frac{1}{e^a} = 1$, окончательно получаем:

$$M(X_{\Pi}) = a. \tag{4}$$

Мы предполагали $M(X_{\Pi}) \approx a$, а оказалось, что *математическое ожидание Пуассоновской с.в. в точности равно ее параметру a!*

Вывод формулы $D(X_{\Pi})$. Как и при вычислении дисперсии биномиальной с.в. (лек. 7, п. 9), используем упрощенную общую формулу дисперсии (лек. 5, п. 7, (4')), которая, с учетом (4), приобретает вид:

$$D(X_{\Pi}) = M(X_{\Pi}^2) - a^2. \tag{5}$$

Вычислим $M(X_{\Pi}^2)$. Согласно определению квадрата с.в. (X_{Π}^2) (лек. 5, п. 7), ее значения равны квадратам значений с.в. X_{Π}, т. е. равны l^2, $l = 0, 1, 2, \ldots$, а их вероятности p_l остаются теми же, что и у с.в. X_{Π} (см.). Значит, из общей формулы математического ожидания получаем:

$$M(X_{\Pi}^2) = \sum_{l=0}^{\infty} l^2 \cdot p_l = \sum_{l=0}^{\infty} l^2 \cdot \frac{a^l}{l!} \cdot e^{-a}.$$

Преобразуем ряд, стоящий справа, следующим образом. Сделаем в каждом слагаемом сокращение на l (как и раньше, при вычислении $M(X_{\Pi})$). Вынесем за знак суммы общий множитель a (а не ae^{-a}, как раньше). Разобьем каждое слагаемое на два, представив множитель l в виде суммы $l = (l-1)+1$. В результате, наш ряд распадется на сумму двух других рядов:

$$\begin{aligned} M(X_{\Pi}^2) &= \sum_{l=1}^{\infty} l \cdot l \cdot \frac{a \cdot a^{l-1}}{1 \cdot 2 \cdot 3 \cdot \ldots \cdot (l-1) \cdot l} \cdot e^{-a} = a \cdot \sum_{l=1}^{\infty} l \cdot \frac{a^{l-1}}{(l-1)l!} \cdot e^{-a} = \\ &= a \cdot \sum_{l=1}^{\infty} [(l-1)+1] \cdot \frac{a^{l-1}}{(l-1)l!} \cdot e^{-a} = a \cdot \sum_{l=1}^{\infty} \cdot \frac{a^{l-1}}{(l-1)l!} \cdot e^{-a} + a \cdot \sum_{l=1}^{\infty} \cdot \frac{a^{l-1}}{(l-1)l!} \cdot e^{-a}. \end{aligned} \tag{6}$$

Первый ряд, стоящий в правой части равенства (6), есть ни что иное, как ряд, представляющий математическое ожидание с.в. X_{Π}, и его сумма равна a. Действительно,

$$\sum_{l=1}^{\infty} (l-1) \cdot \frac{a^{l-1}}{(l-1)l!} \cdot e^{-a} = \sum_{l=1}^{\infty} l \cdot \frac{a^l}{l!} \cdot e^{-a} = M(X_{\Pi}) = a.$$

Второй ряд, стоящий в правой части равенства (6), легко суммируется после вынесения за знак суммы множителя e^{-a}:

$$\sum_{l=1}^{\infty} \frac{a^{l-1}}{(l-1)!} \cdot e^{-a} = e^{-a} \cdot \sum_{l=1}^{\infty} \frac{a^{l-1}}{(l-1)!} = e^{-a} \cdot \left(1 + \frac{a}{1!} + \frac{a^2}{2!} + \frac{a^3}{3!} + \ldots\right) = e^{-a} \cdot e^{a} = 1.$$

Подставим в правую часть равенства (6) суммы слагаемых рядов, найденные только-что, и получим значение математического ожидания квадрата П.с.в. X_{Π}^2:

$$M(X_{\Pi}^2) = a \cdot a + a \cdot 1 = a^2 + a.$$

Остается подставить $M(X_{\Pi}^2) = a^2 + a$ в равенство (5), и мы найдем дисперсию П.с.в. X_{Π}:

$$D(X_{\Pi}) = a. \tag{7}$$

Не удивительно ли, что дисперсия совпала с математическим ожиданием? Такого раньше не было. Это характерное свойство именно Пуассоновских с.в.

Вывод. *Математическое ожидание Пуассоновской случайной величины X_{Π} совпадает с ее дисперсией и совпадает со значением ее параметра a:*

$$M(X_{\Pi}) = D(X_{\Pi}) = a. \tag{8}$$

Контроль 3. Найдите среднее квадратическое отклонение $\sigma(X_{\Pi})$ П.с.в. X_{Π}, параметр которой $a = 3$. Составьте для нее трехсигмовый интервал $(M - 3\sigma;\ M + 3\sigma)$. Пользуясь рядом распределения, построенным в контроле 2, найдите вероятность попадания П.с.в. σ в этот интервал. Сделайте прогноз. Сравните эту вероятность с вероятностью, которую указывает правило трех сигм (лек. 7, п. 4, (2)), — существенно ли различие? Можно ли практически гарантировать попадание П.с.в. в трехсигмовый интервал?

4. Статистический признак П.с.в.

Итак, мы рассмотрели первые три вопроса, поставленные во введении к лекции. Как распределяются вероятности значений П.с.в. X_{Π}? Краткий ответ: "обрыв" и несимметричная "горка". Как это распределение зависит от параметра a? Ответ: с ростом a "горка" расширяется и ее высота уменьшается. Каковы числовые характеристики $M(X_{\Pi})$ и $D(X_{\Pi})$? Ответ: $M(X_{\Pi}) = D(X_{\Pi}) = a$.

Оставшаяся большая часть лекции посвящена последнему вопросу: каковы условия появления Пуассоновских с.в.? В каких опытах следует ожидать возникновение П.с.в.? Как я сказал во введении, одного простого ответа на эти вопросы нет. В данном разделе вы познакомитесь с ответом в терминах статистики, а также с удивительным примером, иллюстрирующим это ответ.

Соображения, которые приводят к нижеследующему статистическому правилу, очень просты. Они основаны на характеристическом свойстве П.с.в. X_{Π}, обнаруженном нами в предыдущем разделе, — $M(X_{\Pi}) = D(X_{\Pi}) = a$.

Признак П.с.в. Пусть изучается некоторая конкретная д.с.в. X, распределение вероятностей которой неизвестно. Проведена серия опытов над этой с.в. и по их результатам вычислены оценки математического ожидания и дисперсии с.в. X, — M^* и D^*. Если эти оценки оказались близки — $M \approx D^*$, то весьма вероятно, что рассматриваемая с.в. X Пуассоновская, параметром которой можно считать $a = M^*$. (Точнее, с.в. X можно моделировать Пуассоновской).

Данный признак, как и любое статистическое утверждение, не дает абсолютной гарантии. Но его хорошо подтверждает практика. Вот удивительный пример.

Пример 1 (удары лошади). В отчетах старых армий приводились между прочими данными данные о числе солдат, убитых ударом лошадиного копыта. В нашем распоряжении имеется достоверная сводка 200 годовых отчетов кавалерийских корпусов немецкой армии конца XIX века (таблица 4). Эта таблица взята из книги ([8] с. 170).

Таблица 4

Число смертей	0	1	2	3	4	5	6	$\sum$
Частота l_i	109	65	22	3	1	0	0	200

В вероятностно-статистической терминологии таблица 4 представляет распределение **частот** появившихся 200 значений случайной величины X — числа солдат из состава кавалерийского корпуса, убитых лошадью в течение года. Каждый отчет, в котором указано число убитых солдат, можно рассматривать, как результат опыта. Опыт повторялся 200 раз. В верхней строке таблицы 4 указаны различные значения x_i с.в. X, в нижней — частоты l_i появления этих значений в серии из $k = 200$ опытов.

Проверим выполнение признака П.с.в., формулированного выше. Для этого надо вычислить статистическое среднее M^* и статистическую дисперсию D^*. Поскольку статистический материал представлен в виде ряда частот, удобно воспользоваться формулами (1') и (2") (лек. 6, п. 4, 5). Считаем:

$$M^* = (1/200) \cdot (0 \cdot 109 + 1 \cdot 65 + 2 \cdot 22 + 3 \cdot 3 + 4 \cdot 1 + 5 \cdot 0) = 0,005 \cdot 122 = 0,61;$$

$$D^* = (1/200) \cdot [(0 - 0,61)^2 \cdot 109 + (1 - 0,61)^2 \cdot 65 + (2 - 0,61)^2 \cdot 22 + (3 - 0,61)^2 \cdot 3 +$$
$$+ (4 - 0,61)^2 \cdot 1] = (1/200) \cdot (0,61^2 \cdot 109 + 0,39^2 \cdot 65 + 1,39^2 \cdot 22 + 2,39^3 + 3,39^2 \cdot 1) =$$
$$= 0,005 \cdot (0,3721 \cdot 109 + 0,1521 \cdot 65 + 1,9321 \cdot 22 + 5,7121 \cdot 3 + 11,4921 \cdot 1) =$$
$$= 0,005 \cdot (40,5589 + 9,8865 + 42,5062 + 17,1363 + 11,4921) = 0,005 \cdot 121,58 = 0,6079.$$

Оценки M^* и D^* оказываются чрезвычайно близки, разница между ними — $|M^* - D^*| = 0,0021$. Значит, согласно сформулированному выше признаку П.с.в., весьма вероятно, что наша с.в. X Пуассоновская с параметром $a = 0,6$.

Проверим эту гипотезу. Для этого составим статистический ряд распределения данной с.в. X и сравним его с теоретическим рядом распределения Пуассоновской с.в. $X_П(a = 0,6)$, который у нас уже есть (таблица 3а).

Таблица 4 представляет статистический ряд **частот** l_i, а нам нужен ряд **относительных частот** l_i/k, $i = 0, 1, 2, 3, 4, 5$. Считаем их:

$(l_0/k)(109/200) = 0,545$; $(l_1/k)(65/200) = 0,325$; $(l_2/k)(22/200) = 0,110$;
$(l_3/k)(3/200) = 0,015$; $(l_4/k)(1/200) = 0,001$; $(l_5/k)(0/200) = 0,000$.

Статистический ряд распределения с.в. X получается таким:

Таблица 5

x_i	0	1	2	3	4	5	$\sum$
l_i/k	0,545	0,325	0,110	0,015	0,001	0,000	1

Сравнивая нижние строки таблиц 5 и 3а, убеждаемся в из близости. Значит, действительно, данную с.в. X можно моделировать Пуассоновской.

Примечание. Близость распределений с.в. X и с.в. $X_{\text{П}}$ нам приходится констатировать визуально, а значит, субъективно. В математической статистике есть методы объективной оценки этой близости, методы оценки степени правдоподобия гипотез. Это так называемые *критерии согласия* (см.). Понять эти методы можно будет после изучения непрерывных с.в.

Близость фактического и теоретического распределений с.в. X еще более разительно проявляется, если сравнить таблицы фактических и теоретических **частот**. Как вы знаете, при большом числе опытов относительные частоты близки к вероятностям, — $(l_i/k) \approx p_i$, значит, теоретические частоты можно посчитать так: $l_i \approx k \cdot p_i$. Сделаем это (теоретические вероятности p_i берем из таблицы 3а):

$$l_0 \approx 200 \cdot 0,5488 = 109,76; \quad l_1 \approx 200 \cdot 0,3293 = 65,86; \quad l_2 \approx 200 \cdot 0,0988 = 19,76;$$
$$l_3 \approx 200 \cdot 0,0198 = 3,96; \quad l_4 \approx 200 \cdot 0,0030 = 0,60; \quad l_5 \approx 200 \cdot 0,0000 = 0,00.$$

Округляем полученные частоты и записываем следующую сравнительную таблицу 6:

Таблица 6

Число смертей	0	1	2	3	4	5
Фактическая частота	109	65	22	3	1	0
Пуассоновская частота	110	66	20	4	1	0

Не поражает ли вас высокая точность совпадения частот? Получается, что такое чрезвычайно редкое и непредсказуемое явление, как гибель солдата от удара лошади, будучи рассмотрено в рамках очень большой массовости, обнаруживает таинственную закономерность, выражаемую точной формулой!! Почему так происходит? В каких других ситуациях можно ожидать проявление закона Пуассона? Как ни странно, математика может объяснить и это. Читайте следующий раздел.

Контроль 4. С.в. Z — число рождений трех и более близнецов в некотором родильном доме в течение года. Какое статистическое исследование надо провести, чтобы получить распределение вероятностей этой с.в.? Составьте четкую программу исследования.

5. Когда биномиальная с.в. превращается в Пуассоновскую?

В начале лекции (п. 2) была подмечена связь между Пуассоновскими с.в. и биномиальными: при большом числе повторений опыта k многоугольник распределения б.с.в. становится очень похожим на Пуассоновский с параметром $a = k \cdot p$ (рис. 2). Теперь, зная статистические условия возникновения П.с.в., мы сможем более точно сформулировать и обосновать эту связь.

Теорема 1 (о связи между П.с.в. и б.с.в.). Имеется эксперимент, с которым связана биномиальная с.в. X_σ, ее параметры — k и p. Если число опытов k в эксперименте достаточно велико, а вероятность "успеха" p достаточно мала и их произведение $k \cdot p = a$ небольшое ($kp < 10$), то данная с.в. X_σ близка к Пуассоновской X_Π с параметром $a = k \cdot p$.

Близость случайных величин понимается естественным образом, как близость вероятностей , соответствующих одинаковым значениям той и другой с.в..

Обоснование. Вероятность p мала $\Longrightarrow 1 - p \approx 1$. Произведение $k \cdot p$ невелико $\Longrightarrow k \cdot p \approx k \cdot p \cdot (1-p)$. Для биномиальной с.в. X_σ последнее равенство означает, что $M \approx D$, а значит, будут близки и оценки $M^* \approx D^*$. Применяя статистический признак П.с.в., заключаем, что $X_\sigma \approx X_\Pi(a = kp)$.

Добавление. Формулировка теоремы и ее обоснование, конечно, не строги, ибо остаются неопределенными термины "мало" и "близко". Хотя естественный смысл этих терминов вам понятен. Теорема утверждает, что $p_l(X_\sigma) \approx p_l(X_\Pi)$, $l = 0, 1, 2, \ldots$, или, учитывая, что левые вероятности определяются формулой Бернулли, а правые — формулой Пуассона, справедливо приближенное равенство:

$$C_k^l \cdot p^l (1-p)^{k-l} \approx \frac{a^l}{l!} \cdot e^{-a}, \; l = 0, 1, 2, \ldots .$$

Данное равенство следует из теоремы Пуассона (лек. 4, п. 6), где было строго доказано, что при $k \longrightarrow \infty$, $p \longrightarrow 0$, $kp = a = const$ левая часть имеет пределом правую, а это, в сущности, и значит, что при достаточно больших k и малых p левая часть "близка" к правой (согласно смысла понятия предела). Там же упоминалось ограничение $k \cdot p = a < 10$.

Причина. Сейчас я хочу объяснить вам причину Пуассоновского распределения с.в. X — числа солдат из состава кавалерийского корпуса, убитых лошадью в течение года (пример 1). Дело в том, что эта с.в. имеет Бернуллиевскую структуру, она биномиальная. Убедимся в этом.

Истолкуем каждую встречу солдата с лошадью, как "простой" опыт. В этом опыте может произойти или не произойти событие A — смертельный удар лошади. Вероятность p события A чрезвычайно мала, но не нулевая. Данный "простой" опыт повторяется в году очень много раз — k раз, в результате пролучается "сложный" опыт — эксперимент. С этим экспериментом и связана наша с.в. X — число l появлений события A при повторении "простого" опыта k раз.

Проверим теперь выполнение для с.в. X условий теоремы 1. Велико ли k? Да. Мало ли p? Да, очень. Достаточно ли мало их произведение $k \cdot p$? Для биномиальной с.в. X это произведение есть математическое ожидание M и в примере 1 мы оценили его статистически — $k \cdot p = M \approx M^* = 0,61$. Условие $k \cdot p < 10$ выполняется. Итак, все условия теоремы 1 соблюдены, следовательно, наша биномиальная с.в. X превращается в Пуассоновскую.

Подобным образом можно вскрыть Бернуллиевскую структуру во многих других массовых случайхных явлениях.[1] Главное, — чтобы вероятность p была **очень мала**. Такова, к примеру, с.в. Z — число рождений трех и более близнецов, с которой вы имели дело в контроле 4. Поэтому закон Пуассона называют **законом редких событий**.

А если k не велико? В этом случае теорема 1 остается справедливой, — ее обоснование не использует величину k, а использует только величину произведения $k \cdot p$, которое должно быть небольшим (оно и будет таковым при малых p). Следовательно, второй случай, когда биномиальная с.в. превращается в Пуассоновскую, — **небольшое k и очень малое p**.

Между прочим, этот факт тоже был подмечен нами ранее (п. 2): многоугольник распределения с.в. $X_\Pi(a = 0,2)$ (рис. 1) похож на распределения с.в. X_σ при малых p (лек. 7, п. 3, рис. 2).

Теперь мы можем сравнить конкретные распределения. Сравним ряд распределения Пуассоновской с.в. $X_\Pi(a = 0,2)$ (таблица 2) и биномиальной с.в. $X_\sigma(k = 2, p = 0,1)$. Именно эти с.в. и надо сравнивать, согласно теореме 1, ибо связь между П.с.в. и б.с.в. устанавливается соотношением $kp = a$. Вероятности второй с.в. вычислим по формуле Бернулли:

$$\begin{aligned} p_0 &= C_2^0 \cdot 0,1^0 \cdot 0,9^2 = 1 \cdot 1 \cdot 0,81 = 0,81; \\ p_1 &= C_2^1 \cdot 0,1^1 \cdot 0,9^1 = 2 \cdot 0,1 \cdot 0,9 = 0,18; \\ p_2 &= C_2^2 \cdot 0,1^2 \cdot 0,9^0 = 1 \cdot 0,01 \cdot 1 = 0,01. \end{aligned}$$

Ряд распределения с.в. $X_\sigma(k = 2,\ p = 0,1)$ получается таким:

Таблица 7

$X_\sigma(k = 2, p = 0,1)$	0	1	2
p	0,81	0,18	0,01

Сравнивая таблицы 7 и 2, видим, что вероятности во второй строке очень мало различаются, на одну-две сотых! Так что при малых k и p биномиальную с.в. X_σ действительно можно моделировать Пуассоновской X_Π. Остается заметить, что в

[1] Когда-то (начало лекции 4) я говорил вам о практической и теоретической ценности схемы Бернулли? Теперь вы видите, к какому широкому классу массовых случайных явлений приложима эта схема. Вы убеждаетесь также, что эта схема стимулирует развитие теории, — вслед за формулой Бернулли появляется формула Пуассона и класс Пуассоновских с.в., предельных для биномиальных. В дальнейшем, при изучении непрерывных с.в., вы увидите связь между биномиальными и Гауссовскими распределениями и познакомитесь с новыми ценными теоретическими следствиями.

отличие от случая больших k, моделирование при малых k не имеет практической ценности, ибо вероятности с.в. X_σ легко найти по формуле Бернулли.

Контроль 5. Вскройте Бернуллиевскую структуру с.в. Z — числа рождений трех и более близнецов в некотором родильном доме в течение года.

6. Поток событий

Нам осталось проанализировать один класс опытов, порождающих Пуассоновские с.в.

Пример 2 (поток вызовов на АТС). Чтобы исследовать пропускную способность автоматической телефонной станции (АТС), надо изучить динамику поступления вызовов в разные промежутки времени (в разное время суток) и за много дней. В сущности, надо провести много раз следующий опыт: выбрать какой-то отрезок времени $[T_0; T]$ (например, от 9 до 10 часов утра) и провести регистрацию последовательных моментов времени поступления вызовов — t_1, t_2, t_m. Если делать это много дней, то число вызовов m будет непредсказуемо меняться и, следовательно, является случайной величиной. Оказывается, эта с.в. распределена по закону Пуассона, что будет доказано в конце лекции (п. 10).

Практически проверить этот факт можно так же, как в примере 1. Надо зафиксировать число вызовов за много дней (несколько месяцев) — $m_1, m_2, \ldots, m_k$, вычислить их среднее арифметическое m^*, ввести П.с.в. с параметром $a = m^*$, рассчитать ее ряд распределения и сравнить его со статистическим рядом, — они окажутся очень близки.

Другие примеры. Подобные случайные величины возникают при исследовании ситуаций, связанных с появлением очередей, когда нужно выяснить эффективность обслуживания (телефонные станции, билетные кассы, ремонтные мастерские, справочные бюро, парикмахерские и пр., и пр.). И все они тоже оказываются Пуассоновскими.

Формула Пуассона является, таким образом, основой для решения задач, в которых надо определить число служащих или количество аппаратуры, необходимые для обслуживания данного вида запросов. Это может быть число телефонисток на телефонной станции или число турникетов на станции метро или еще что-то подобное. Такими задачами занимается современная ветвь теории вероятностей — *теория массового обслуживания* [3, гл. 19].[2]

Обобщение. Отвлекаясь от конкретного содержания примера 2, выделим его обобщенную структуру.

[2] В основе этой теории лежит формула и распределение Пуассона, которые, в свою очередь, возникли из схемы и формулы Бернулли. Видите, какие далекие и ценные следствия тянутся по сей день из удачно выделенного научного объекта (схемы Бернулли). Подобные научные результаты называют *фундаментальными* и *классическими*. Очень интересные реальные задачи и их решение с помощью формулы Пуассона можно найти в старой книге [8, с. 179- 188].

Определение 2. *Потоком событий* назовем опыт, который состоит в том, что в некотором временном интервале $[T_0; T]$ появляются какие-то события, следующие друг за другом в какие-то (вообще, случайные) моменты времени t_1, t_2, t_3, ...

Результатом данного опыта можно считать возрастающую последовательность чисел $\{t_1, t_2, t_3, \ldots\}$ — моментов времени, когда появляются события потока. Эта последовательность, вообще говоря, конечная, но теоретически может быть и бесконечной. *Результат опыта тоже называют потоком событий.* Сами события потока могут быть любой природы, но мы будем считать их в некотором смысле *однородными*, т. е. различающимися только моментами появления.

Моменты появления событий потока, как сказано в определении, вообще говоря, случайны. Но не исключается случай, когда события следуют друг за другом через строго определенные промежутки времени. Такой поток называют *регулярным*. Понятно, что он редко встречается в реальности. К примеру, поезда в метро иногда идут через равные временные интервалы в 2 минуты. Но, строго говоря, и здесь будут малые случайные колебания промежуточных интервалов между поездами.

Случайные величины, связанные с потоком. Введем с.в. K_τ — *число событий потока.* Если поток не регулярный и события потока появляются случайно, то их число будет меняться при повторении опыта. Значит, K_τ действительно, является случайной величиной. Ее значения — неотрицательные целые числа $x_l = l$, $l = 0, 1, 2, 3, \ldots$. Число значений может быть любым. Нам предстоит доказать, что вероятности этих значений определяются формулой Пуассона. Доказательство (п. 10) будет проведено не для любых потоков событий, а для потоков, удовлетворяющих трем условиям (стационарность, отсутствие последействия и ординарность). Поэтому следующие три раздела лекции посвятим этим трем новым понятиям.

Добавление. С нерегулярным потоком событий можно связать и другие с.в., например, с.в. T_ϕ — длительность интервала времени между двумя произвольными соседними событиями потока. Эта с.в. не дискретная, ибо длительность t может принимать любые значения из некоторого "сплошного" временного промежутка $(0; t)$. Эта с.в. появится у нас в дальнейшем, при изучении непрерывных с.в. (лек. 9, п. 2, 3, 5, лек. 11, п. 3, 5). Она принадлежит к классу *показательных* распределений, тоже связанных с формулой Пуассона. С конкретным потоком событий примера 2 можно связать еще одну непрерывную с.в. T_l — длительность произвольного разговора. Она тоже встретится вам в дальнейшем.

Контроль 6. Изучается пропускная способность эскалатора станции метро от 8 до 10 часов утра. Электроглаз фиксирует время появления на ленте эскалатора каждого пассажира. Что будет результатом этого опыта? Можно ли считать этот результат потоком событий? Что значит — "повторить опыт"? Каковы, по вашему предположению, возможные значения с.в. K_τ в этом случае? Сможете ли смоделировать эту с.в. Пуассоновской? Каков параметр a?

7. Стационарный поток

Свойство стационарности потока — это, другими словами, свойство его **почти равномерности**. Взгляните на рисунки 3а и 3б, — какой из изображенных потоков вы назовете более равномерным? Очевидно, первый, ибо у второго потока события идут вначале густо, потом реже.

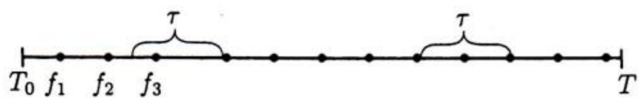


Рис. 3а

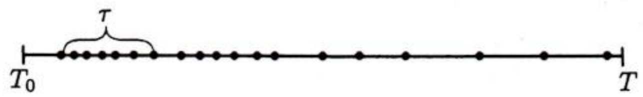


Рис. 3б

Можно сказать точнее: если передвигать небольшой интервальчик длины τ вдоль основного интервала $[T_0; T]$ первого потока, то независимо от того, где он будет находиться, в него будет попадать примерно одно и то же число точек τ_i — событий потока (рис. 3а). Для второго потока это свойство не выполняется, — если интервальчик τ находится в начале потока, в него попадет гораздо больше точек, нежели, если он находится в конце (рис. 3б).

Определение 3. Поток событий называется *стационарным*, если на любой интервал времени длины τ, $\tau \subset [T_0; T]$, в каком бы месте основного интервала $[T_0; T]$ он ни находился, попадает при повторении опыта, в среднем, одно и то же число событий потока.

Пример. Представьте, что проведен эксперимент по исследованию пропускной способности АТС в интервале между 21 часом и 21 часом 15 минутами. Это значит, что в течение многих дней (опыт повторялся k раз) фиксировался поток вызовов во временном промежутке $[T_0; T] = [21;\ 21.15]$ и подсчитывалось их число m_i. Результат: в первый день поступило m_1 вызовов, во второй — m_2 вызовов, в третий — m_3 вызовов, :, в последний k-тый день поступило m_k вызовов. Среднее число вызовов в день на данном временном промежутке определится, как среднее арифметическое: $m^* = (1/k) \cdot (m_1 + m_2 + \ldots + m_k)$.

Для наглядности давайте конкретизируем результат, считая, например, что среднее число вызовов $m^* = 150$, $m_1 = 160$, $m_2 = 157$, $m_3 = 143$, $m_k = 151$, ... (вы понимаете, конечно, что числа m_i я выбирал произвольно, но около 150). В этих условиях стационарность потока означает вот что. Если разбить основной интервал на три пятиминутных — [21; 21.05], [21.05; 21.10], [21.10; 21.15 и посчитать среднее число вызовов на каждом из них, получится примерно по 50. Если разбить на 15 одноминутных интервальчиков, то среднее число вызовов на каждом будет тоже примерно одинаково — по 10. И т. д.

Если увеличить длину основного промежутка, например, до $[21; /, 24]$, поток потеряет свойство стационарности (почему?).

Пропорциональность. Обратите внимание, — если мы уменьшаем (или увеличиваем) в несколько раз интервальчики, разбивающие $[T_0; T]$, то среднее число событий потока в них тоже уменьшается (увеличивается) **во столько же** раз. В частности, на интервал длины 2τ будет попадать, в среднем, в два раза больше событий, чем на интервал длины τ. На интервал длины $0,2\tau$ будет попадать в 5 раз меньше событий, чем на интервал длины τ.

Другими словами, в стационарном потоке среднее число событий, попадающих на интервал τ, не зависит от его положения на основном интервале $[T_0; T]$, а зависит только от его длины τ, причем, зависит **пропорционально**.

В частности, среднее число событий потока, рассчитанное для интервала τ **единичной** длины, не зависит от его положения и, следовательно, постоянно для данного потока, — его называют *интенсивностью*, или плотностью потока, — обозначим ее λ. В нашем примере интенсивность потока вызовов $\lambda = 10$ (при условии, что за единицу времени выбрана 1 минута).

Зная интенсивность потока λ, легко рассчитать среднее число событий, попадающих на любой интервал длины τ, — оно равно $\lambda \cdot \tau$. Здесь проявляет себя свойство пропорциональности, отмеченное выше: если на единичный временной интервал попадает λ событий потока, то на τ-интервал, длины, например, $\tau = 2$ попадет в два раза больше, т. е. $\lambda \cdot 2$ событий.

В литературе понятие стационарности обычно дается в несколько иной форме, использующей не статистическую, а более точную вероятностную терминологию:

Определение 3'. Поток событий называется *стационарным*, если вероятность попадания того или иного числа событий на временной интервал длины τ не зависит от его положения на промежутке $[T_0; T]$, а зависит только от его длины τ.

Контроль 7. Является ли поток событий контроля 6 стационарным (эскалатор идет вниз)? Если да, то почему? Если нет, то как следует изменить основной промежуток, чтобы поток стал таковым? Как экспериментально проверить стационарность нового потока? Как определить его интенсивность?

8. Поток без последействия

"Физическая" независимость. Второе условие, которому удовлетворяет поток вызовов на АТС (пример 2), это **независимость** вызовов друг от друга. Каждый вызов поступает, как правило, по своей собственной причине, не связанной с причинами, заставляющими других абонентов звонить по телефону в данный промежуток времени $[T_0; T]$. Если же представить, что в какой-то момент происходит нечто чрезвычайное, что заставляет многих людей звонить по телефону в этот момент, то поток теряет свойство независимости (так же, как и свойство стационарности).

Итак, свойство независимости событий потока имеет в своей основе их **“физическую”** независимость. Каждое событие потока появляется в каком-то месте промежутка $[T_0; T]$ в силу **индивидуальных** случайных причин, появляется случайно, его причина не связана с другими событиями.

Из “физической” независимости вызовов на АТС следует, что если мы знаем, когда и сколько вызовов появилось до момента t_0, то эта информация не позволяет нам предсказать, когда появится следующий вызов, много или мало вызовов поступят в ближайшее время. Все это — дело случая. Можно сказать, что “будущее” потока не зависит от его “прошлого”. Отсюда и термин — “поток без последействия”.

Определение 4. Поток событий называется *потоком без последействия*, если для любых непересекающихся временных интервалов τ' и τ'' число событий, попадающих на один из них, не зависит от того, сколько событий попало на другой.

Пример 3 (поток пассажиров метро). Рассмотрим поток пассажиров, входящих на эскалатор станции метро с 8 до 10 часов утра. Думаю, вы не затруднились ответить в контроле 6, что этот поток не стационарный, ибо в начале потока события идут очень густо (многие люди спешат на работу), а к концу — реже. По этой же причине он — с **последействием**, ибо в начале потока многие события связаны друг с другом одной причиной.

Если руководствоваться определением 4, то при выборе соседних интервалов τ' и τ'' можно предсказать, что если в первый интервал попадает много событий, то и во второй тоже (в среднем!). Это значит, что интервалы выбраны в начале потока. Если же в первый интервал попало мало событий, то и во второй тоже. Эти интервалы выбраны в конце потока. Получается, что число событий, попадающих на один интервал, зависит от того, сколько событий попадает на другой. “Прошлое” определяет “будущее”. Поток с последействием.

Ситуацию можно “исправить”, если изменить основной временной промежуток с [8;10] на [9;10]. Тем самым, мы исключим действие общей причины, заставляющей события идти густо в начале потока. Новый поток станет потоком без последействия (и стационарным, — события будут идти реже и равномернее). Если “исправить” ситуацию “в другую сторону”, изменить основной промежуток на [8;8.30], то мы не исключим общую причину, поток останется “с последействием”, но стационарность восстановится (события будут идти густо и достаточно равномерно).

Другое определение. В дальнейшем, при доказательстве теоремы о Пуассоновском распределении с.в. K_τ нам придется использовать чуть иное, более точное определение потока без последействия. Вот оно:

Определение 4′. Поток событий называется *потоком без последействия*, если **вероятность** попадания m событий потока (m — любое) на произвольный временной интервал τ' не зависит от того, сколько событий попало на любой другой интервал τ'', не пересекающийся с первым.

Пояснение. Связь определений 4 и 4′ — это связь между частотой и статистической вероятностью. Если провести много — k опытов (реализаций потока), то в интервал τ' будет попадать разное число событий потока. Подсчитаем частоту попадания на интервал τ' m событий и обозначим ее l_m. Вычислим относительную частоту — l_m/k, она будет близка к вероятности попадания m событий на интервал τ', которая определяется, как вероятностный предел этой частоты: $(p) - \lim\limits_{k\to\infty}(l_m/k) = \mathbf{P}(K_\tau = m)$. Поскольку, согласно определению 4, частота l_m не зависит от того, сколько событий потока попало на другой интервал τ'', то и вероятность $\mathbf{P}(K_\tau = m)$ не зависит от этого.

Контроль 8. Обладает ли поток пассажиров, выходящих из метро между 8 и 10 часами (эскалатор идет вверх) свойствами а) отсутствия последействия, б) стационарности? Можно ли изменить основной промежуток $[8; 10]$ так, чтобы изменились данные свойства потока?

9. Ординариный поток

Третье свойство (*ординарность*) потока вызовов на АТС (пример 2) заключается в том, что вызовы идут **по одиночке**, а не парами или тройками. Вызовы отделены друг от друга временными интервалами, которые могут быть очень малыми. Альтернатива, — когда события потока оказываются жестко сцепленными, например, в случае потока клиентов ЗАГСа. Поток клиентов парикмахерской можно считать ординарным.

Следствие ординарности. Поскольку максимально возможное число вызовов на промежутке $[T_0; T]$ практически ограничено, то, разбив его на достаточно большое количество мелких промежуточков Δt_1, $\Delta t_2, \ldots$, Δt_k, можно считать, что при любой реализации потока в каждый промежуток Δt_i попадет или ровно одно событие, или ни одного.

Два и более событий могут, конечно, попадать на промежуток Δt_i, но это будет происходить очень редко. Каждое из этих двух событий (A — на промежуток Δt_i попадает ровно одно событие потока, B — более одного) имеет свою вероятность, но при достаточно малом Δt_i вторая вероятность, очевидно, пренебрежимо мала, сравнительно с первой.[3].

Данное следствие описательного определения ординарности (события потока идут поодиночке) позволяет дать более точное определение:

Определение 5. Поток событий называется *ординарным*, если для любого достаточно **малого** временного промежутка Δt вероятность появления в нем двух и более событий потока пренебрежимо мала, сравнительно с вероятностью появления одного события.

[3]При $\Delta t_i \longrightarrow 0$ вероятности $P(A)$ и $P(B)$ становятся бесконечно малыми и вторая вероятность — бесконечно малая более высокого порядка, нежели первая. Такова точная замена фразы: "вторая вероятность пренебрежимо мала, сравнительно с первой"

Контроль 9. Рассмотрите два потока: поток покупателей, подходящих к кассе магазина, и поток отходящих от этой кассы. Выполняются ли для этих потоков свойства стационарности, отсутствия последействия и ординарности? При каких дополнительных предположениях?

10. Число событий простейшего потока подчиняется закону Пуассона

Настал момент, когда мы сможем теоретически доказать, что в примере 2 случайная величина K_τ (число вызовов на АТС) обязана быть Пуассоновской. И так будет всегда, когда поток *простейший*, т. е. когда он обладает тремя вышеопределенными свойствами. Такой поток и называют простейшим, или *Пуассоновским*.

Теорема 2 (о Пуассоновском распределении с.в. K_τ). Пусть на временном интервале $[T_0; T]$ возникает поток событий и рассматривается случайная величина K_τ — число событий этого потока. Если поток обладает свойствами стационарности, отсутствия последействия и ординарности, то с.в. K_τ — Пуассоновская с параметром a, равным среднему числу событий этого потока, появляющихся на интервале $[T_0; T]$.

Доказательство состоит: 1-2) в приближенной замене с.в. K_τ биномиальными с.в. $X_\sigma^{(k)}$, вероятности значений которых считаются по формуле Бернулли; 3) в последующем переходе в этой формуле к пределу (при $k \longrightarrow \infty$), который и будет формулой Пуассона. Кратко, — $\lim\limits_{k\to\infty} X_\sigma^{(k)}$.

1) Разобьем интервал $[T_0; T]$ на большое число k равных мелких частей Δt_1, Δt_2, $\ldots$, Δt_k так, чтобы при любой реализации потока на каждом Δt_i могло появится или нет только одно событие потока. Такое разбиение возможно, согласно свойству **ординарности**.

Возьмем произвольный интервальчик Δt_i, обозначим через A событие, состоящее в появлении на нем **ровно** одного события потока, и рассчитаем вероятность этого события — $\mathbf{P}(A) = p_k$ (она зависит от числа k, — чем больше k, тем меньше вероятность).

Для этого введем новую с.в. K_i — число событий потока, возникающих в Δt_i. Ее математическое ожидание M_i приближенно равно среднему числу событий потока в Δt_i (лек. 6, п. 4). Поскольку длина Δt_i в k раз меньше длины $[T_0; T]$, в котором, по условию теоремы, возникает, в среднем, a событий потока, то в силу **стационарности** (свойство пропорциональности), на интервальчик Δt_i попадает, в среднем, a/k событий потока и, следовательно,

$$M_i \approx a/k. \tag{9}$$

С другой стороны, математическое ожидание с.в. K_i, которая теоретически принимает любые значения $l = 0,\ 1,\ 2\ldots$, определяется по общей формуле так:

$$M_i = 0 \cdot \mathbf{P}(K_i = 0) + 1 \cdot \mathbf{P}(K_i = 1) + 2 \cdot \mathbf{P}(K_i = 2) + 3 \cdot \mathbf{P}(K_i = 3) + \ldots$$

Но интервальчики Δt_i выбраны нами столь малыми, что вероятности $\mathbf{P}(K_i = 2)$, $\mathbf{P}(K_i = 3)$, ... пренебрежимо малы, сравнительно с $\mathbf{P}(K_i = 1)$ (определение **ординарности**). Поэтому

$$M_i \approx \mathbf{P}(K_i = 1). \tag{10}$$

Обозначение $(K_i = 1)$ — иное обозначение события A. Из (9) и (10) следует

$$\mathbf{P}(A) = p_k \approx \frac{a}{k} \tag{11}$$

Остается заметить, что, поскольку в вышепроведенном рассуждении интервальчик Δt_i был произвольным, взятым из данного разбиения $[T_0; T]$ на k частей, то вероятность появления ровно одного события потока на каждом Δt_i одна и та же, и она определяется равенством (11). Т. е. вероятность p_k не меняется при переходе от одного Δt_i к другому (при фиксированном k).

2) Введем с.в. — число l интервальчиков Δt_i, на которых появилось ровно одно событие потока. Очевидно, что с.в. $X_\sigma^{(k)}$ и с.в. K_τ связаны с одним опытом (потоком на $[T_0; T]$) и принимают при каждой его реализации одинаковые значения.[4] Напомню, что, в силу ординарности, на каждом Δt_i может появиться не более одного события потока.

Убедимся в биномиальности с.в. $X_\sigma^{(k)}$.

Разбиение промежутка $[T_0; T]$ на k частей Δt_1, Δt_2, ..., Δt_k разбивает данный поток на k элементарных стационарных потоков с одинаковыми временными интервалами Δt_i, $i = 1, 2, 3, \ldots, k$. Поэтому данный опыт (поток) можно рассматривать, как повторение элементарного опыта k раз.

В каждом элементарном потоке Δt_i может появиться или нет событие A (ровно одно событие потока) с вероятностью $p_k = a/k$, приближенно вычисленной выше (формула (11)). Эта вероятность не меняется от опыта к опыту, она одинакова для всех Δt_i.

С.в. $X_\sigma^{(k)}$ мы определили как число l появлений события A при повторении опыта (элементарного потока) k раз. Все условия биномиальной с.в. (лек. 7, п. 2) выполняются.

3) Итак, имеем биномиальную с.в. $X_\sigma^{(k)}$, "близкую" к данной с.в. K_τ. Что значит "близкую"? Мы уже отмечали, что при любой реализации опыта они принимают одинаковые значения. Однако, вероятность события A мы нашли приближенно, — $p_k \approx a/k$, поэтому вероятность $\mathbf{P}(X_\sigma^{(k)} = l)$, вычисленная по формуле Бернулли, будет чуть отличаться от $\mathbf{P}(K_\tau = l)$.[5] Отличие будет тем меньше, чем меньшими взяты интервальчики Δt_i, т. е. чем больше их число k. Следовательно, точное значение этой вероятности можно получить в пределе:

$$\mathbf{P}(K_\tau = l) = \lim_{k\to\infty} \mathbf{P}(X_\sigma^{(k)} = l)$$

[4] Строго говоря, это утверждение не совсем верное, — на Δt_i могут появляться и более одного события потока, но чрезвычайно редко. Такими событиями мы условились пренебрегать, в силу их очень малой вероятности.

[5] Эти отличия будут чуть усиливаться и за счет проигнорированной нами возможности появления двух событий на Δt_i. С ростом k те и другие будут исчезать.

Займемся этим пределом.

Возможные значения с.в. $X_\sigma^{(k)}$ — $l = 0, 1, 2, 3, \ldots, k$. Вероятности этих значений определяются формулой Бернулли (учтите, $p_k \approx a/k$):

$$\mathbf{P}(X_\sigma^{(k)} = l) = C_k^l \cdot p_k^l \cdot (1-p_k)^{k-l} \approx$$
$$\approx \frac{k\cdot(k-1)\cdot(k-2)\cdot\ldots\cdot[k-(l-1)]}{l!}\cdot\left(\frac{a}{k}\right)^l\cdot\left(1-\frac{a}{k}\right)^{k-l}. \tag{12}$$

Прежде чем искать предел, сделаем в (12) несложные преобразования:

$$\mathbf{P}(X_\sigma^{(k)} = l) \approx \frac{k\cdot(k-1)\cdot(k-2)\cdot\ldots\cdot[k-(l-1)]}{k\cdot k\cdot k\cdot\ldots\cdot k}\cdot\frac{a^l}{l!}\cdot\left(1-\frac{a}{k}\right)^{-l}\cdot\left(1-\frac{a}{k}\right)^{k} =$$
$$= \frac{a^l}{l!}\cdot\frac{k-1}{k}\cdot\frac{k-2}{k}\cdot\ldots\cdot\frac{k-(l-1)}{k}\cdot\left(1-\frac{a}{k}\right)^{-l}\cdot\left(1-\frac{a}{k}\right)^{k} =$$
$$= \frac{a^l}{l!}\cdot\left(1-\frac{1}{k}\right)\cdot\left(1-\frac{2}{k}\right)\cdot\ldots\cdot\left(1-\frac{l-1}{k}\right)\cdot\frac{1}{\left(1-\frac{a}{k}\right)^l}\cdot\left(1-\frac{a}{k}\right)^{k}.$$

Теперь устремим $k \to \infty$. При этом l остается неизменным, значит, первый множитель есть константа. Следующие множители, за исключением последнего, имеют пределом единицу (именно потому, что l фиксировано). Число этих множителей тоже не меняется и равно l. Следовательно,

$$\mathbf{P}(K_\tau = l) = \lim_{k\to\infty}\frac{a^l}{l!}\cdot\left(1-\frac{a}{k}\right)^k = \frac{a^l}{l!}\cdot\lim_{-\frac{k}{a}\to\infty}\left[\left(1-\frac{1}{\left(-\frac{k}{a}\right)}\right)^{-\frac{k}{a}}\right]^{-a} = \frac{a^l}{l!}\cdot e^{-a}.$$

Вот как из формулы Бернулли появляется формула Пуассона с замечательным числом e. Между прочим, этот предельный переход мы уже делали, когда доказывали теорему Пуассона мелким шрифтом (лек. 4, п. 6).

Следствие. Если интенсивность Пуассоновского потока на $[T_0;T]$ равна λ, $[T_1;T_2] \subset [T_0;T]$ и $\tau = \frac{[T_2-T_1]}{[T-T_0]}$, то среднее число событий на $[T_1;T_2]$ есть $\lambda\tau$ и вероятность $p_\tau(l)$ того, что на $[T_1;T_2]$ появится l событий потока, можно рассчитывать по формуле Пуассона

$$p_\tau(l) = \frac{(\lambda\tau)^l}{l!}\cdot e^{-\lambda\tau} \tag{14}$$

Данное следствие и формула (14) часто используется при решении задач. Значения правой части при различных $\lambda\tau$ можно найти по таблице функции Пуассона (приложение 2).

Добавление. Не следует думать, что Пуассоновские с.в. появляются только в случае простейшего временного потока событий. Есть и другие ситуации, аналогичные по своим свойствам простейшему потоку, в которых наблюдаются Пуассоновские распределения. К примеру, представьте, что к длинному причалу последовательно причаливают лодки и закрепляются в различных точках причала.

Получается поток случайных точек, располагающихся не на временной оси, а на линейном промежутке $[0; b]$. Можно представить аналогичное "поле точек" на плоскости или в пространстве. Такое поле создают космические частицы, попадающие на поверхность спутника. Попытайтесь сами обобщить на эти ситуации свойства стационарности, ординарности и отсутствия последействия, а также понятие интенсивности потока, это не трудно. Между прочим, теорему 2 можно доказать и без требования стационарности.

Контроль 10. Докажите следствие теоремы 2.

11. Упражнения

1. Нарисуйте эскиз многоугольника распределения П.с.в. для значения параметра $a = 0,8$. Пользуясь таблицей приложения 2, составьте ряд распределения и постройте многоугольник распределения. Согласуется ли он с вашим эскизом? Найдите δ, постройте трехсигмовый интервал $(M - 3\delta; M + 3\delta)$ и с помощью ряда распределения определите вероятность попадания данной П.с.в. в этот интервал. Сравните найденную вами вероятность с той, которую указывает правило трех сигм. Какова разница?

2. Ответьте на все вопросы предыдущего упражнения при $a = 4$.

3. На телефонную станцию поступает, в среднем, 30 вызовов в течение часового промежутка времени. Найти вероятность того, что в течение минуты поступит а) ровно 2 вызова; б) не более двух; в) хотя бы один.

Ответ: 0,0758; 0,9856; 0,3935.

4. Будет ли поток заказов, поступающих на диспетчерский пункт такси простейшим? Обоснуйте. Как определить среднее число вызовов в минуту? Считая, что интенсивность потока (в минуту) равна 3, найдите вероятность того, что за 2 минуты поступит а) 4 вызова; б) менее четырех; в) не менее 4-х.

Ответ: 0,1339; 0,1512; 0,8488.

5. При работе ЭВМ время от времени возникают неисправности (сбои). Можно ли считать поток сбоев за сутки простейшим? Обоснуйте. Считая, что среднее число сбоев за сутки равно 1,5, найдите вероятности следующих событий: A — за двое суток не будет ни одного сбоя; б) B — за сутки произойдет хотя бы один сбой; в) C — за неделю (6 дней) произойдет не менее шести сбоев.

Ответ: 0,050; 0,777; 0,884.

6. Поток грузовых железнодорожных составов, прибывающих на сортировочную горку, имеет интенсивность $\lambda = 4$ состава в час. Найдите вероятность того, что а) за 0,5 часа прибудет один состав; б) хотя бы один; в) не менее трех.

Отвут: 0,2707; 0,8647; 0,3213.

7. На ткацком станке нить обрывается, в среднем, 0,375 раза за час работы станка. Найдите вероятность того, что за смену (8 часов) число обрывов нити не выйдет за пределы промежутка $[2; 4]$. Ответ: 0,6160.

8. Вероятность попадания в самолет выстрелом из ружья равна 0,001. По самолету ведет огонь подразделение, общее число выстрелов 5000. Какова вероятность а) ровно одного попадания; б) двух попаданий; в) хотя бы одного попадания; г) не менее двух?

Ответ: 0,0337; 0,0842; 0,9933; 0,9596.

9. В тесто, приготовленное для выпечки 1000 булочек, засыпают 10000 изюминок и тщательно перемешивают. Какова вероятность того, что в случайно выбранной булочке окажется а) 5 изюминок; б) не больше пяти; в) меньше десяти; г) не меньше 10-ти?

Ответ: 0,038; 0,0671; 0, 46; 0,54.

10. Искусственный спутник Земли может случайным образом сталкиваться с метеоритами. Метеориты, сталкивающиеся со спутником, образуют Пуассоновский поток с плотностью $\gamma = 0,1$ (метеоритов в сутки). Метеорит, попадающий в спутник, пробивает его оболочку с вероятностью $p_1 = 0,02$. Метеорит, пробивший оболочку, выводит из строя аппаратуру с вероятностью $p_2 = 0,07$. Найдите вероятности следующих событий: A — за месяц ($k = 30$ суток) полета спутника его оболочка будет пробита; B — будет выведена из строя аппаратура; C — будет пробита оболочка спутника, но аппаратура не пострадает.

Ответ: $1 - e^{-\gamma \cdot k \cdot p_1}$; $1 - e^{-\gamma \cdot k \cdot p_1 \cdot p_2}$; $e^{-\gamma \cdot k \cdot p_1 \cdot p_2} - e^{-\gamma \cdot k \cdot p_1}$.

11. Розничная лавка с ограниченными возможностями хранения продуктов продает, в среднем, 10 пакетов галет в неделю. Запас обычно возобновляется каждый понедельник утром. Требуется установить такой стандарт еженедельного пополнения запаса, чтобы случилось не более 1 отказа.[6]

Ответ: 16 пакетов.

Игорь Петрович Костенко,
кандидат физ.-мат. наук, доцент,
действительный член Международной
педагогической академии.
email: kost@kubannet.ru

[6]Формулировка задачи взята из старой книги [8, с. 181]. Там же проведено интересное решение с изумительным комментарием. В книге много других реальных задач, требующих применения закона Пуассона и связанных с переменной нагрузкой телефонной сети и с радиоактивным распадом вещества [8, с. 183-191].

Материалы к курсу «применимая математика» Показательная зависимость. Логарифмы. Предел $(1+1/n)^n$

А. А. Колчин, А. И. Щетников

Настоящая статья продолжает ряд публикаций, связанных с проектом учебного курса «Применимая математика» ([5]–[7]). Представленные здесь методические разработки были опробованы группой сотрудников Лаборатории теоретической и прикладной эпистемологии в учебно-познавательных семинарах для старшеклассников, проведенных в Кемерово (гимн. 42), Новосибирске (шк. 202), Мысках (шк. 8), Красноярске (МЛШР), Омске (ГМЛШ). Статья разделена на два части; материал каждой части приблизительно соответствует трехдневному семинару-погружению.

Часть I. О формировании понятий степени, показательной и логарифмической зависимости

1. Традиционный подход к формированию понятия степени и его критика

В школьной математике понятие степени формируется поэтапно: степень с натуральным показателем, степень с целым показателем, степень с рациональным показателем, степень с вещественным показателем.

На первом этапе (7 класс) определяется арифметическое действие возведения числа в степень с натуральным показателем. Здесь выражение a^n трактуется как «сокращенная запись» для произведения n одинаковых сомножителей:

$$a^n = \underbrace{a \cdot a \cdot \ldots \cdot a}_{n \text{ раз}}.$$

На основе этого определения доказываются соотношения

$$a^n \cdot a^m = a^{n+m}, \tag{1}$$

$$(a^n)^m = a^{nm}. \tag{2}$$

Обратное к (1) соотношение

$$a^n : a^m = a^{n-m} \tag{3}$$

рассматривается пока только для случая $n > m$.

На втором этапе (8 класс) вводится понятие степени с целым показателем. Соотношение (3) обобщается теперь на случай $n \leq m$, что приводит к новым формальным определениям

$$a^{-n} = \frac{1}{a^n}, \qquad a^0 = 1.$$

На третьем этапе (10 класс) вводится понятие о степени с рациональным показателем. С этой целью предварительно определяется понятие арифметического корня n-ой степени. Сначала обращают соотношение (2) на случай

$$\sqrt[n]{a^{mn}} = a^{(mn:n)} = a^m.$$

Затем уславливаются о том, что это формальное правило будет иметь место и для тех случаев, когда p не делится на q:

$$\sqrt[q]{a^p} = a^{p/q},$$

после чего выводят совокупность правил для действий с дробными показателями.

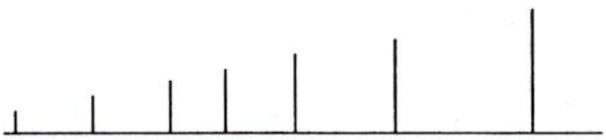
Одно из следствий такого подхода к формированию понятия степени проявляется в том, что в сознании учащихся исходное определение степени как сокращенной записи для многократного произведения продолжает конкурировать с последующими формально введенными определениями степени с целым и рациональным показателем. Это приводит к сильному недоумению: «Как можно умножить число само на себя ноль раз, минус два раза или три пятых раза?» Это недоумение не разрешается осмысленно, что усиливает отчуждение многих учащихся от предмета математики.

2. Идея показательного роста

2.1. Альтернативный подход к формированию понятия степени заключается в том, чтобы с самого начала задать его в максимально общем виде, связав с идеей непрерывного показательного роста ([1], [2]). С этой целью нужно поставить перед школьниками предметную задачу об определении значений некоторой переменной величины P (вообще говоря, непрерывной, но рассматриваемой сначала только для целочисленных значений аргумента), изменяющейся по условию задачи так, что приращению аргумента на единицу соответствует изменение P в одно и то же число раз: $P(n+1) = a \cdot P(n)$. Две задачи такого рода описаны ниже, в разделах 3 и 4.

При оформлении решения этих задач следует обратить внимание школьников на важное различение обозначения степени и способа ее вычисления. Цепочка значений $P(n)$ образует геометрическую прогрессию, уходящую бесконечно в обе стороны (рис. 1). Отношение $P(n) : P(0)$ принято обозначать с помощью символической записи «a^n». Показатель n в этой записи указывает на место, которое занимает данное отношение в рассматриваемой прогрессии. Запись «a^0» служит для обозначения нулевого члена прогрессии; по определению $a^0 = 1$. Запись «a^1» служит для обозначения первого члена прогрессии; он равен $1 \cdot a$. Запись «a^2» служит для обозначения второго члена прогрессии; он вычисляется как $1 \cdot a \cdot a$. Запись «a^{-3}» служит для обозначения члена прогрессии с номером «-3»; этот член вычисляется как $\frac{1}{a \cdot a \cdot a}$.

2.2. Проблему определения значений a^x для дробных x естественнее всего поставить и решить на языке «вставки средних» в уже имеющуюся геометрическую прогрессию. К примеру, пусть величина $P(t)$ за каждые сутки меняется вдвое. В таком случае, во сколько раз она будет меняться за каждые $\frac{1}{2}$ суток? Пусть за первые $\frac{1}{2}$ суток она выросла в $a^{1/2} = x$ раз. Но ведь и за вторые $\frac{1}{2}$ суток она выросла также в x раз. Тогда за сутки она выросла в $x \cdot x = x^2$ раз. Но мы знаем, что за полные сутки она вырастает в 2 раза. Получается, что $x^2 = 2$.



место	-3	-2	-1	0	1	2	3
обозначение	a^{-3}	a^{-2}	a^{-1}	a^0	a^1	a^2	a^3
вычисление	$\frac{1}{a \cdot a \cdot a}$	$\frac{1}{a \cdot a}$	$\frac{1}{a}$	1	$1 \cdot a$	$1 \cdot a \cdot a$	$1 \cdot a \cdot a \cdot a$

Рис. 1

Действие отыскания такого числа x, для которого $x^2 = a$, называют извлечением корня второй степени из a. Здесь мы опять сохраняем различение двух обозначений: $a^{1/2}$ есть обозначение величины, стоящей в рассматриваемом порядке величин на отметке $\frac{1}{2}$; $\sqrt{a}$ есть указание на то, что квадрат этой величины равен a.

Теперь рассмотрим более общий случай, когда отрезок между двумя целочисленными отметками делится на n равных частей (рис. 2). Аналогичным образом вводится понятие о корне n-ой степени из a и показывается, что $a^{m/n} = (\sqrt[n]{a})^m = \sqrt[n]{a^m}$.

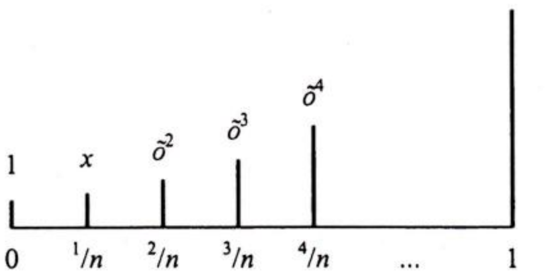


Рис. 2

2.3. Характеристическое свойство показательной зависимости, не привязанное к конкретной шкале измерения аргумента, состоит в следующем: ординаты, соответствующие произвольной последовательности абсцисс, образующих арифметическую прогрессию, сами составят геометрическую прогрессию. Иначе говоря, показательная функция изменяется (растет/убывает) равномерно, но в отличие от линейной функции, изменяющейся равномерно относительно сложения, она изменяется равномерно относительно умножения.

Отметим еще одну важную идею, которая может способствовать пониманию школьниками того, что представляет собой «показательная зависимость как таковая». Обычно мы считаем, что функции вида $y = ax^2 + bx + c$ имеют своими графиками параболы различного положения и формы. Однако можно считать, что в «идеальном мире геометрических фигур» существует одна-единственная парабола («идея параболы»), а разные конкретные формулы возникают при различном положении и выборе масштабов координатных осей. Точно так же можно считать, что существует один-единственный график показательной зависимости с приложенной к нему неразмеченной осью абсцисс («идея экспоненты»), а разные конкретные показательные функции $y = b \cdot a^x$ считываются с этого графика, когда мы задаем начало отсчета и ориентацию оси абсцисс, а также масштабы по обеим осям.

3. Задача о начислении сложных процентов

Пусть банк начисляет 100% на вклад, положенный ровно на 1 год. Банк хочет привлечь как можно больше вкладчиков — и тех, кто кладет деньги на длительные сроки, и тех, кто делает краткосрочные вложения. Поэтому он предоставляет всем вкладчикам право снять деньги со своего счета в любой момент, когда им заблагорассудится. Тогда какова должна быть месячная ставка, соответствующая данной годовой ставке? Вообще, как рассчитать причитающиеся вкладчику проценты за произвольное время?

Кажется, что для того, чтобы удовлетворить интересы различных групп вкладчиков, банк должен равномерно увеличивать начисления в течение года, — но такая схема невыгодна для банка. Упростим задачу: пусть деньги можно класть в банк минимум на полгода. Тогда вкладчику оказывается выгодным положить деньги на полгода, снять их, получив 50% прибыли, и тут же снова положить на полгода: получается рост в $(1 + 0,5)^2 = 2,25$ раза и относительный выигрыш 25%.

Банку надо подобрать такую полугодовую ставку, чтобы она давала годовой рост в 2 раза. 1,5 — слишком много. Попробуем $1,4^2 = 1,96$. Это уже лучше. $1,41^2 = 1,9881$ — это совсем хорошо; подходящая полугодовая ставка равна 41%. Точное решение получаем из уравнения $x^2 = 2$, $x = \sqrt[2]{2}$. (В этом месте ведущему группы или учителю в классе не следует форсировать события. Сначала лучше заняться подбором, и только потом перейти к формализованному составлению уравнений и извлечению корней с помощью калькулятора.)

Если банк разрешает забирать деньги минимум через квартал (3 месяца), то очевидно, что вкладчику выгодно действовать в соответствии с прежней стратегией — как можно чаще забирать деньги со счета и тут же вкладывать их обратно. Чтобы найти квартальную ставку, можно разделить год на четыре части, а можно полугодие на две. Получаем два эквивалентных решения $\sqrt[4]{2} = \sqrt[2]{\sqrt[2]{2}}$. Аналогичным образом решается вопрос о месячной ставке.

4. Задача о росте бактерий

Некая разновидность бактерий такова, что каждая бактерия делится на две дочерние бактерии ровно через 1 сутки после того, как она сама появилась на свет в результате предыдущего деления. В исследовательскую лабораторию был доставлен 1 кг бактерий. Отвлекаясь от вопросов о том, чем эти бактерии будут кормиться, где они будут содержаться и т. п., зададимся следующими вопросами: Во сколько раз число бактерий вырастет за любые двое суток? трое суток? n суток? И еще: если число бактерий за сутки выросло в 2 раза, то во сколько раз она выросло за $\frac{1}{2}$ суток? за $\frac{1}{3}$ суток? и каким оно было и будет в произвольный момент времени?

Если не привлекать каких-то специальных предположений о распределении бактерий по их возрастам, то выясняется, что последняя задача в ее общей постановке имеет не одно, а бесконечный набор решений. Каждое такое решение получается, если на отрезке $[0, 1]$ задать произвольную неубывающую функцию $F(x)$, принимающую на концах отрезка значения $F(0) = 1$ и$F(1) = 2$, а потом распространить ее на всю числовую прямую пошагово, сдвигая график на единицу и изменяя все ординаты при каждом сдвиге в 2 раза (рис. 3). Можно сказать еще, что каждое отдельное решение соответствует определенному распределению бактерий в исходной пробе по «возрастам»: если мы поделим отрезок времени $[0, 1]$ на равные интервалы, то приращение числа бактерий на каждом из этих интервалов будет в точности равно доле бактерий, деление которых произошло внутри этого интервала времени.

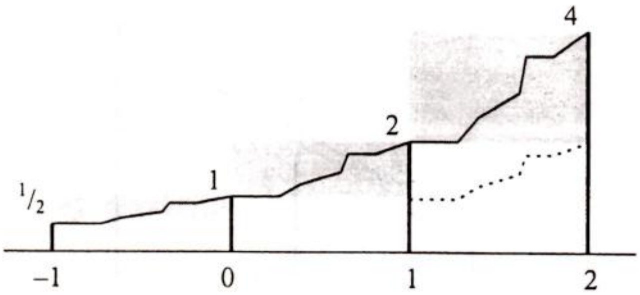


Рис. 3

Однако с некоей «натурфилософской» позиции представляется правдоподобным, что среди всей этой совокупности решений имеется одно «особенное» решение, график которого является «всюду одинаковым». Можно обсуждать его как «задачу о средних», рассматриваемую с позиции естествоиспытателя. Пусть имеется много разных популяций бактерий, никак одна с другой не связанных. Мы берем из каждой популяции по равной порции бактерий и смешиваем эти порции. При этом должна получиться усредненная (и в этом смысле самая вероятная) зависимость числа бактерий от времени.

Правдоподобно предположить, что в получившейся смеси одна половина бактерий поделится за первые $\frac{1}{2}$ суток, а вторая половина — за вторые $\frac{1}{2}$ суток. Поэтому через $\frac{1}{2}$ суток их будет в полтора раза больше, чем в начальный момент времени.

Если мы поделим первые сутки не на 2, а на 10 частей, то за каждую такую часть будет делиться 1/10 доля бактерий, и общее число бактерий будет вырастать на 1/10 от их первоначального числа. И вообще, число бактерий будет в течение первых суток нарастать линейно от первоначального числа до удвоенного. А в каждые последующие сутки вновь будет происходить линейный рост, но теперь уже от 2 до 4, от 4 до 8, от 8 до 16 и так далее.

Предыдущее рассуждение вроде бы и представляется правильным, и тем не менее в получившимся результате есть некоторая странность: мы ожидали, что рост количества бактерий будет описываться «гладкой» функцией от времени, а эта функция оказалась кусочно-линейной, с характерными изломами в целочисленных точках. За каждую секунду первых суток делилось одно и то же число бактерий; в каждую секунду вторых суток это число вновь было постоянным, но вдвое большим, и так далее. И эти «скачки» скорости роста не представляют ли собой что-то странное, противоречащее нашим ожиданиям, согласно которым «природа не делает скачков»?

Можно рассуждать об этом еще и так. Если в момент времени $t = \frac{1}{2}$ поделилась половина бактерий, то масса бактерий стала равна $1\frac{1}{2}$ кг. Точно так же в момент времени $t = 1\frac{1}{2}$ масса бактерий будет равна 3 кг. Теперь рассмотрим сутки между этими двумя моментами времени. По той же самой логике за первую половину этих суток поделится половина от $1\frac{1}{2}$ кг, то есть $\frac{3}{4}$ кг. И масса бактерий при $t = 2$ станет равной $1\frac{1}{2} + \frac{3}{4} = 2\frac{1}{4}$ кг. Но ведь по условию эта масса должна быть равной 2 кг! (рис. 4).

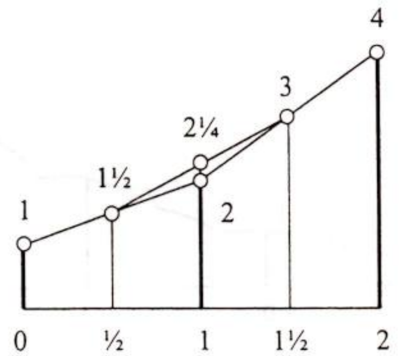


Рис. 4

Но на эти доводы имеется и сильное возражение. На самом деле никакого равноправия времен нет. В природных условиях показательный рост возможен лишь на каком-то небольшом промежутке времени. А потом пищи перестает хватать, и рост приостанавливается, выходя на насыщение (за любой промежуток времени равное число бактерий появляется и гибнет). В такой ситуации бактерии в природной популяции действительно распределятся по возрастам равномерно. Когда мы вынули часть бактерий из природной популяции и поместили их в лабораторные условия, обеспечив на некоторое время пищей и прочими условиями, необходимыми для беспрепятственного размножения, это и было настоящее начало опыта, от которого следует вести отсчет. Изломы графика — это своего рода память о мгновенном изменении жизненных условий. Наблюдатель, обнаруживший изломы, может узнать, какое время показывали часы в момент начала опыта.

5. Формирование понятия о логарифмических шкалах и логарифмах

5.1. *Установочная лекция* (идея которой заимствована из статьи [8]). Представим себе, что мы хотим изобразить на одном и том же графике следующие величины:

(1) время одного удара сердца — около 1 сек;

(2) продолжительность перемены между уроками — 10 мин = 600 сек;

(3) время, за которое можно на поезде доехать от Новосибирска до Москвы — 7 суток ≈ 200000 сек;

(4) время, за которое Земля делает оборот вокруг Солнца — 1 год ≈ 30000000 сек;

(5) время, за которое свет доходит до ближайшей к нам галактики — 4 млн. лет= = 120000000000000 сек.

Если цена миллиметрового деления будет равна 1 секунде, то отметка, изображающая продолжительность перемены, будет отстоять от нулевой отметки на 60 см, а время поездки на поезде — на 200 м (не говоря уже о двух следующих величинах). Как же нам быть?

Идея состоит в том, чтобы при переходе к каждой следующей отметке изображаемая величина возрастала не на одну и ту же добавку, но в одно и то же число раз, например, в 10 раз (рис. 5).

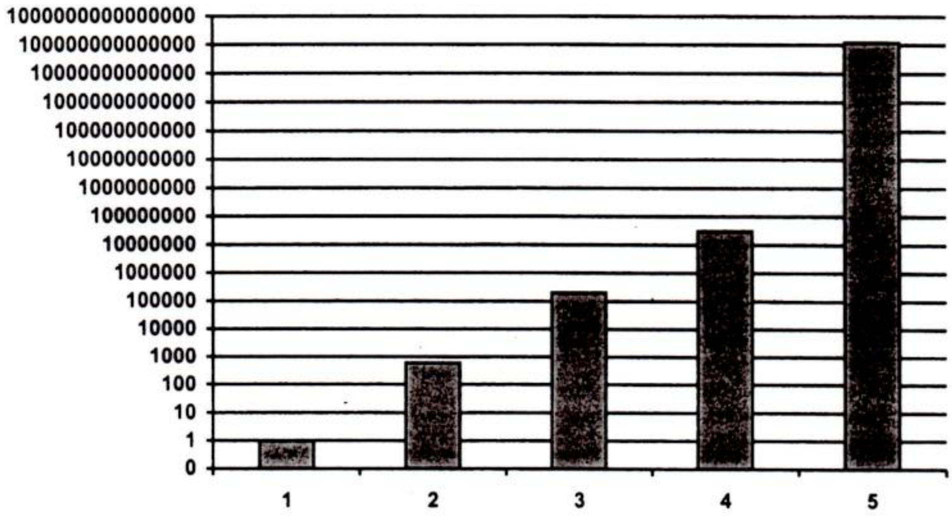


Рис. 5

Степень, в которую нужно возвести число 10, чтобы получить положительное число a, принято называть десятичным логарифмом a. И вообще, степень, в которую нужно возвести положительное число b, чтобы получить положительное число

a, принято называть логарифмомa по основанию b. Теперь можно сказать, что наша шкала тоже является равномерной, но только на ней равномерно растут не сами числа, а их десятичные логарифмы. Поэтому шкалы такого вида называются логарифмическими.

Как нам изобразить на шкале число 600? Ясно, что его отметка будет находиться между отметками 100 и 1000. И отметка будет ровно на 2 единицы выше, чем отметка числа 6. Но в какую степень нужно возвести 10, чтобы получить 6? Пусть $6 = 10^x$. Будем возводить это уравнение во 2, 3, 4 и т. д. степень, пока очередная степень шестерки не окажется достаточно близка к какой-либо степени десятки. Удобно воспользоваться тем, что $6^9 = 10077696 \approx 10^7$. Тогда $6 \approx 10^{7/9}$.

5.2. *Рабочие задания третьего дня* (группам предлагается выбрать по жребию два варианта заданий).

Вариант 1. Ось x имеет равномерную разметку; ось y имеет логарифмическую разметку. На этой шкале постройте графики следующих функций:

(а) $y = 10^x$, $y = 100^x$, $y = 0,1^x$;

(б) $y = 10^{1,5x}$, $y = 10^{-0,7x}$;

(в) $y = 2^x$, $y = 0,7^x$;

(г) $y = 10 \cdot 2^x$, $y = 0,01 \cdot 2^x$;

(д) $y = 3 \cdot 2^x$.

Вариант 2. Оси x и y имеют логарифмическую разметку. На этой шкале постройте графики следующих функций:

(а) $y = x$, $y = x^2$, $y = x^3$, $y = x^5$;

(б) $y = 1/x$, $y = 1/x^3$;

(в) $y = \sqrt{x}$, $y = x^{-1,7}$;

(г) $y = 10x$, $y = 100\sqrt{x}$, $y = 0,1/x^2$;

(д) $y = 2x^{-1,7}$.

Часть II. Задачи, связанные с замечательным пределом $(1+1/n)^n$

5. Задача о полоскании белья

5.1. После стирки хозяйка набирает полную ванну чистой воды, чтобы прополоскать белье. Но насколько эффективен такой способ полоскания? Не лучше ли делить чистую воду на части? Пусть выстиранное белье содержит 1 л мыльной воды, и пусть для его полоскания отведено 10 л чистой воды. Как лучше всего разделить эту воду на части, чтобы белье после полоскания стало максимально чистым? (После отжима в белье всегда остается ровно 1 литр воды.)

5.2. Можно начать работу с подсчета коэффициента очистки, показывающего, во сколько раз уменьшается концентрация мыла в воде, для самых простых случаев, когда вода делится на 2, 5, 10 равных частей (рис. 6).

Общая формула вычисления коэффициента очистки в случае деления воды на n равных частей имеет вид

$$K_n = \left(1 + \frac{10}{n}\right)^n.$$

Уже на первых примерах было видно, что при увеличении числа частей увеличивается (причем довольно сильно) и степень очистки. Естественно задаться вопросом, к какому результату приведет дальнейшее увеличение числа частей: будет ли коэффициент очистки возрастать неограниченно, или же его поведение с увеличением n окажется каким-то иным? При увеличении n выражение в скобках принимает все меньшие значения, сколь угодно близко подходя к единице; зато показатель степени становится все большим. «1^∞» — сколько это получится в нашем случае?

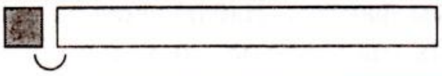
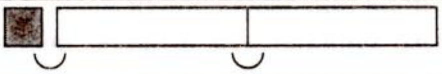
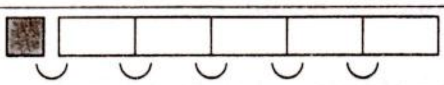
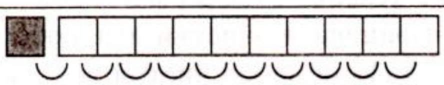
Способ очистки	*Коэффициент очистки*
	1 + 10 = 11
	(1 + 5) (1 + 5) = 36
	(1 + 2)(1 + 2)(1 + 2)(1 + 2)(1 + 2) = 243
	$(1 + 1)^{10} = 1024$

Рис. 6

5.3. Удачный ход состоит в составлении таблицы K_n для n вида 10^m. Наблюдение за таблицей показывает, что сначала коэффициент очистки растет весьма резко, но затем его рост сильно замедляется. Процесс этого замедления делается еще более заметным, если выписать разности D_n между соседними значениями K_n.

n	K_n	D_n
10^0	11	
10^1	1024	
10^2	13780,61...	
10^3	20959,15...	
10^4	21916,68...	≈ 1000
10^5	22015,45...	≈ 100
10^6	22025,36...	≈ 10
10^7	22026,35...	≈ 1
10^8	22026,45...	$\approx 0{,}1$
10^9	22026,46...	$\approx 0{,}01$

Из таблицы можно видеть, что имеет место приближенное соотношение $D_n \approx \frac{10^7}{n}$. Тем самым процесс накопления разностей оказывается схожим с процессом суммирования геометрической прогрессии, порождающим периодическую десятичную дробь $1,(1) = \frac{10}{9}$. Похоже, что величина коэффициента очистки с каждой строкой окончательно определяется в очередном десятичном знаке, и K_n никогда не превысит некоторого числа, заключенного между $22026,46$ и $22026,47$.

И все же, сам факт очистки более чем в 22 тысячи раз за счет 10-кратного отношения чистой воды к мыльному раствору — этот результат следует назвать поразительным. Теперь становится понятным, за счет чего современные стиральные машины обходятся таким малым количеством воды на полоскание!

Важной темой дальнейшего обсуждения служит различение между правдоподобной догадкой и математически доказанным фактом. Наблюдение за таблицей привело нас к выводам, в истинности которых мы уверены, поскольку нет никаких оснований предполагать, что правильный порядок, замеченный нами в первых строках таблицы, окажется затем нарушенным. Однако эти выводы нельзя считать доказанными, ибо пока не найдено никаких оснований, из которых они могли бы быть выведены.

5.4. Объем воды 10 л, о котором шла речь в условии задачи, мог быть и иным. И следует поставить вопрос о предельном коэффициенте очистки для любого объема V, а точнее — для любого отношения $\alpha = V/V_0$ объемов чистой воды и воды, остающейся в белье после выжимания. Коэффициент очистки можно представить в виде

$$K_n(\alpha) = \left(1 + \frac{\alpha}{n}\right)^n = \left[\left(1 + \frac{1}{z}\right)^z\right]^\alpha,$$

где введена новая переменная $z = n/\alpha$. При стремлении $z \to \infty$ выражение в квадратных скобках будет стремится к предельному значению, известному как число $e = 2,71...$, поэтому $K_n(\alpha) \to K(\alpha) = e^\alpha$.

Этот же результат можно получить и из менее формальных соображений. Мы можем считать, что предельная очистка α литрами чистой воды ($\alpha \in \mathbb{N}$) достигается за счет α раз повторенной предельной очистки 1 литром чистой воды, то есть

$$K(\alpha) = K(1)^{\alpha} = e^{\alpha}.$$

Когда вода делилась на «неделимые» литры, очистка шла в геометрической прогрессии с основанием 2; когда эти литры разделены на бесконечно малые порции, очистка описывается показательной зависимостью с основанием $e = 2{,}71\ldots$ (рис. 7).

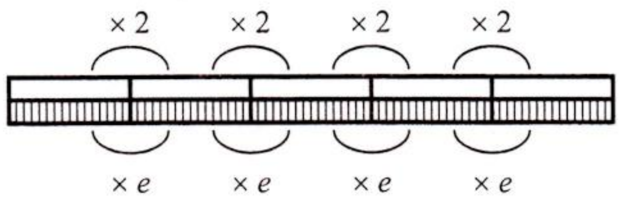


Рис. 7

5.5. Предметом параллельного теоретического рассмотрения могут стать следующие утверждения, доказательства которых основываются на простых алгебраических неравенствах:

(1) Из всех вариантов деления данной порции воды на две части наибольшую степень очистки дает тот, когда обе части равны, поскольку

$$(1+a)^2 > (1+(a-d))(1+(a+d)) = (1+a)^2 - d^2.$$

(2) Любую данную порцию воды всегда выгоднее делить на части, нежели использовать целиком, поскольку для любых a, $b > 0$ выполняется неравенство

$$(1+a)(1+b) > 1+a+b.$$

Отсюда следует вывод о том, что не существует такого деления воды на *конечное* число частей, которое приводило бы к максимальному коэффициенту очистки.

5.6. Дальнейшее обсуждение исходной задачи может быть развернуто по целому вееру направлений, от сугубо теоретической проблемы доказательства существования предела к более прикладному вопросу о том, какое деление воды на части обеспечивает максимальную эффективность, если нужно учитывать фактор времени; и, наконец, к технической задаче конструирования экономичной стиральной машины (или иного очистного устройства, теплообменника и т. п.). При правильном обустройстве пространства общей дискуссии можно не только сделать обозримым различие между способами работы «ученого» и «инженера», но также обсудить, как можно эффективно организовать кооперацию между этими позициями, и как получается, что решая идеализированные, «оторванные от практики» задачи, математик получает результаты, которые впоследствии могут найти вполне осязаемое практическое применение.

5.7. *Еще одна вариация на эту же тему.* Имеются две жидкости одинаковой теплоемкости при температуре 100° и 0°. Если привести их в тепловой контакт, между ними установится тепловое равновесие при 50°. Спрашивается, до каких температур можно охладить первую и нагреть вторую жидкость, если приводить их в контакт не целиком, а порциями (в итоге все порции каждой из жидкостей снова соединяются вместе)?

5.8. *Задача, обратная к задаче о банковской ставке.* Какую максимальную прибыль можно было бы получить, если бы (1) годовая ставка была бы равна 100%, что соответствует увеличению вклада в 2 раза; (2) банкир по ошибке ввел равномерную схему роста начислений? Если вкладчик будет через каждую $1/n$ часть года снимать все деньги с набежавшими процентами со счета и тут же снова класть эту сумму в банк, то его вклад к концу года увеличится в $(1 + 1/n)^n$ раз. Тем самым в пределе при $n \to \infty$ в конце года каждая вложенная единица увеличится в $e = 2,71...$ раз, а дополнительная прибыль составит несколько более 71%.

Чтобы при такой стратегии предприимчивого вкладчика вклад вырос бы за год только в 2 раза, сумма вклада за малую $1/n$ долю года должна увеличиваться не в $(1+1/n)$ раз, а в $(1+\alpha/n)$ раз, где $\alpha < 1$. Тогда в пределе вклад увеличится за год в $K(\alpha) = e^{\alpha}$ раз. Отсюда $\alpha = \ln 2 = 0,693...$ Этот результат можно получить проще, если сразу потребовать, чтобы рост вклада происходил по закону $2^t = (e^{ln 2})^t$.

6. Задача о реактивном движении

Какую скорость v приобретет ракета, израсходовав все запасы топлива, если ее начальная масса m_0, конечная масса m_1, скорость истечения газов из сопла u?

Эту задачу обычно решают, записывая закон сохранения импульса в виде дифференциального уравнения $m\,dv + u\,dm = 0$. Интегрируя, находят

$$v = u \int_{m_1}^{m_0} \frac{dm}{m} = u \ln \beta,$$

где $\beta = m_0/m_1$. Но нельзя ли получить решение без обращения к дифференциальному уравнению?

Заметим, что при непрерывном истечении продуктов сгорания искомая скорость зависит только от отношения начальной и конечной масс $\beta = m_0/m_1$. Это утверждение можно обосновать с помощью мысленного эксперимента в стиле Галилея: до какой скорости разгонится параллельная связка из двух одинаковых ракет, образующих одну ракету удвоенной массы с удвоенным запасом топлива?

Разобьем топливо на порции таким образом, чтобы в каждом такте сгорания отношение текущих начальной и конечной масс ракеты было равно $1+\varepsilon$, где $\varepsilon \ll 1$. При этом последовательные значения переменной массы ракеты составят геометрическую прогрессию (рис. 8).

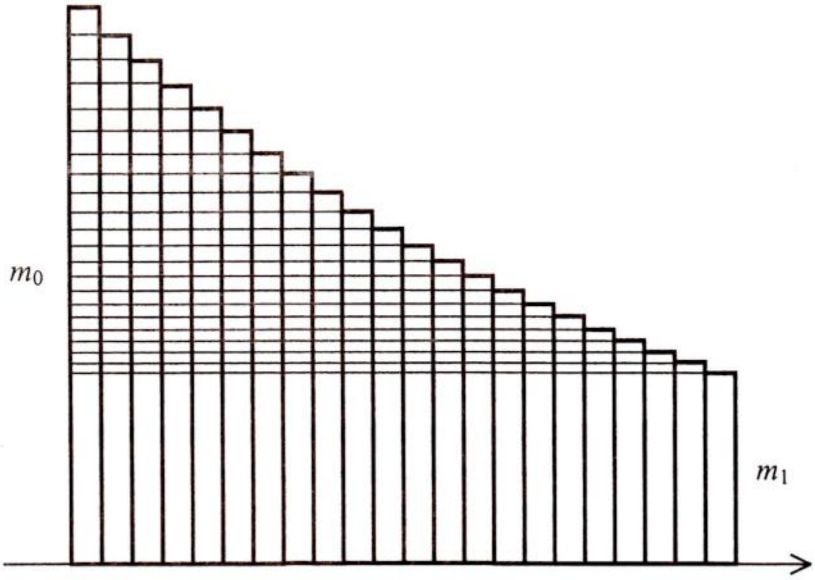


Рис. 8

Из законе сохранения импульса следует, что за один такт сгорания топлива ракета приобретает скорость $\delta v = \varepsilon u$. (Модель: в двигателе корабля происходят взрывы порций топлива, сообщающие кораблю импульсы = толчки.) Тем самым за n тактов сгорания ракета приобретет скорость $v = n\delta v = n\varepsilon u$.

Чтобы узнать, каково должно быть число тактов, чтобы масса ракеты уменьшилась в β раз, составим уравнение

$$\beta = (1+\varepsilon)^n = \left(1 + \frac{n\varepsilon}{n}\right)^n = \left(1 + \frac{(v/u)}{n}\right)^n.$$

Если устремить $n \to \infty$, получится $\beta = e^{v/u}$, что дает искомое решение $v = u \ln \beta$.

Отсюда видно, что увеличение запасов топлива является крайне неэффективным способом увеличения конечной скорости ракеты. Ведь если для разгона ракеты до скорости u должно быть $\beta = e$, то для разгона ракеты до скорости $10u$ придется взять $\beta = e^{10} \approx 22026$, что практически невозможно. (В некотором смысле этот результат обратен результату задачи об очистке белья.) Напротив, эффективное решение состоит в увеличении скорости истечения продуктов сгорания из сопла.

7. Парадокс раздачи подарков

7.1. Пусть каждый участник большой компании покупает подарок, а затем купленные подарки распределяются среди членов компании по жребию. (При работе над задачей хорошо сначала несколько раз разыграть процесс раздачи подарков «в живую», и только потом переходить к теоретическому обсуждению.) Кажется, что вероятность того, что кому-нибудь достанется собственный подарок, ничтожно мала. Однако в действительности это не так [4].

В самом деле, вероятность того, что первый участник в компании из n человек получит «свой» подарок, равна $1/n$, поэтому вероятность получения им «чужого» подарка равна $1-1/n$. При больших n можно считать, что вероятность получения «чужого» подарка вторым участником не зависит от того, чей подарок, «свой» или «чужой», получил первый участник, и поэтому она также приближенно равна $1-1/n$. Это же рассуждение справедливо и для следующих участников.

Тем самым вероятность того, что все участники получат «чужие» подарки, приближенно равна

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

Но при стремлении $n\to\infty$ будет $K(-1)=e^{-1}=0,37...$

(Конечно, формальный трюк с заменой знака подлежит какому-нибудь более аккуратному обоснованию. Но, кстати сказать: а какой смысл можно приписать отрицательным количествам чистой воды в рамках задачи о полоскании белья?)

7.2. Нетрудно понять, что при $n\gg 1$ наша задача эквивалентна следующей. В смеси из большого количества шаров двух цветов на $(n-1)$ красный шар приходится 1 синий шар. Вытащим из этой смеси ровно n шаров. Спрашивается, какова вероятность $P(k)$ того, что в этой выборке окажется $k=0,\ 1,\ 2,\ \ldots$ синих шаров?

Рассуждения, аналогичные приведенным выше, дают для $k\ll n$ значение

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

Но сумма всех этих вероятностей должна быть равна единице! Это неожиданным образом приводит нас к известному разложению

$$e=1+1+\frac{1}{2}+\frac{1}{2\cdot 3}+\frac{1}{2\cdot 3\cdot 4}+\frac{1}{2\cdot 3\cdot 4\cdot 5}+\ldots\quad .$$

Благодарности. Мы выражаем свою признательность Н. И. Кузнецовой, А. В. Любченко и А. В. Щетниковой за плодотворные дискуссии по теме настоящей статьи и за совместную работу по проведению учебно-познавательных семинаров.

Литература

[1] Абельсон И. Б. *Две прогрессии.* М.–Л., 1938.

[2] Абельсон И. Б. *Рождение логарифмов.* М.–Л., 1948.

[3] Олдридж Б. Натуральный логарифм. *Квант*, № 8, 1992.

[4] Секей Г. *Парадоксы в теории вероятностей и математической статистике.* М., 1990.

[5] Щетников А. И., Щетникова А. В. Учебный семинар «Как решать незнакомую задачу». *Труды конференции, посвященной 90-летию со дня рождения Алексея Андреевича Ляпунова.* Новосибирск, ОИИ СО РАН, 2001, с. 773–780.

[6] Щетников А. И., Щетникова А. В. Учебно-исследовательский семинар «Распределение первых значащих цифр». *Математическое образование*, № 2(21), 2002, с. 108–123.

[7] Щетников А. И., Щетникова А. В. Вероятностно-статистические закономерности: опыт обсуждения проблем и простейших моделей со старшеклассниками. *Вычислительные технологии*, 7 (2002), спец. выпуск, ч. 4, с. 336–341.

[8] Aaron K. Just what is a logarithm, anyway? http://pumas.jpl.nasa.gov/

Колчин Алексей Александрович,
учитель математики отделения "Умка"
школы № 202, г. Новосибирск.

E-mail: akolchin@ngs.ru

Щетников Андрей Иванович,
координатор Лаборатории теоретической
и экспериментальной эпистемологии.

E-mail: schetnikov@ngs.ru
http://ltpe.stsland.ru

Доказательство под микроскопом

И. Л. Тимофеева

В статье обсуждаются вопросы: что такое математическое доказательство и как оно устроено (какова его логическая структура)? Вводится на интуитивном уровне понятие математического доказательства в виде дерева. Выявлены преимущества подхода к уточнению понятия доказательства в виде дерева по сравнению с традиционным линейным доказательством.

В математике мы постоянно имеем дело с доказательствами. Однако редко задаем себе следующие вопросы (и еще реже пытаемся ответить на них): Что такое *математическое доказательство*? В чем его сущность? Как устроено математическое доказательство? Как выявить его структуру?

Представьте себе следующую ситуацию. На письменном экзамене по геометрии учащийся должен был доказать теорему Пифагора. Написав работу, ученик сдал ее преподавателю. Преподаватель, прочитав работу, возвратил ее с комментарием: «Это не доказательство!» На что ученик возразил: «Почему Вы считаете, что написанное мною — не доказательство? Что такое доказательство? Вы на уроках геометрии давали определения многих понятий, но среди них не было понятия *доказательства*. На каком основании Вы можете утверждать, будто то, что я написал, — не доказательство? Как Вы можете *доказать*, что мое доказательство не является доказательством?!». У этой истории возможны два финала. Если преподаватель обладает широтой взглядов и не лишен творческого начала, то он может повысить на балл оценку ученику, у которого возникают столь неординарные вопросы. Если преподаватель — педант, лишенный чувства юмора, то он, скорее всего, снизит оценку за непочтительность.

Эта весьма правдоподобная и поучительная история рассказана замечательным американским логиком Раймондом Смаллианом в одной из его великолепных, полных юмора популярных книг по математической логике [4]. Я прочитала ее много лет назад и с тех пор всегда пересказываю эту историю своим студентам на математическом факультете МПГУ в самом начале курса математической логики. Мне кажется, она блестяще объясняет, что будущему учителю математики изучать математическую логику совершенно необходимо хотя бы для того, чтобы прояснить для себя, что такое доказательство.

Так что же такое математическое доказательство? Может ли математика ответить на этот вопрос? Могут ли математические доказательства быть объектом изучения в математике? Можно ли изучать понятие математического доказательства с помощью математических средств и методов? На последние три вопроса ответ положительный. Математические доказательства являются главным объектом изучения *теории доказательств,* которая представляет собой основной раздел математической логики. Для того чтобы понятие математического доказательства стало объектом изучения в математике, необходимо это понятие уточнить.

Сначала надо договориться, что понимать под математическим доказательством на интуитивном уровне. Для краткости в дальнейшем будем говорить просто *доказательство* вместо слов *математическое доказательство*[1].

Прежде всего, под доказательством будем понимать не *процесс* обоснования какого-либо математического утверждения, а его *результат*, представленный в виде некоторого текста. Итак, доказательство представляет собой некоторый *текст*, состоящий из отдельных предложений. Особенность этого текста заключается в том, что составляющие его предложения логически взаимосвязаны друг с другом. Обычно эта связь выражается словами «предложение такое-то *логически следует* из предшествующих предложений таких-то». При кратком изложении доказательства используется лишь слово «следовательно» (или равнозначное слово), а из каких именно посылок следует данное предложение (делается вывод) часто явно не указано, да и сами посылки даже не всегда сформулированы. А что значит, *логически следует*? Обычно это никак не уточняется. При уточнении этого слосочетания предложение «Из $A_1, \ldots, A_n$ *логически следует* B» можно понимать следующим образом: переход от предложений $A_1, \ldots, A_n$ к предложению B происходит по некоторому *правилу вывода* (логическому правилу), разумеется, если восстановлены все пропущенные шаги. Однако на практике, в неформальном доказательстве обычно не уточняется, в соответствии с каким логическим правилом делается тот или иной вывод.

В доказательстве должны быть *исходные* предложения, которые не являются следствиями предыдущих предложений (ведь текст конечен!). Такими исходными предложениями могут служить *аксиомы* той математической теории, в рамках которой проводится доказательство. Исходными могут быть и предложения, имеющие место в силу *определения*. В некотором смысле такие предложения добавляются к аксиомам. В качестве исходных предложений могут также выступать вспомогательные *допущения*. В самом деле, в доказательствах очень часто встречаются слова «допустим, что ...».

Отношение *логического следования*, как бы мы ни уточняли это понятие, определенным образом упорядочивает члены доказательства. Более того, этот порядок древовиден (имеет вид дерева с корнем — наименьшим элементом), поскольку на каждом шаге рассуждения происходит переход от некоторых предложений $A_1, \ldots, A_n$ к некоторому единственному предложению B, непосредственно следующему из них по какому-либо правилу логики.

Итак, приходим к следующему описанию понятия доказательства в виде дерева.

Под *доказательством в виде дерева* будем понимать упорядоченную в виде дерева систему предложений, в которой каждое исходное предложение является аксиомой или допущением, или имеет место в силу определения, а каждое из остальных предложений следует из непосредственно предшествующих ему предложений по какому-либо правилу вывода (логическому правилу).

---

[1]Понятие *доказательства* (в широком смысле), используемое вне математики, мы не обсуждаем. В частности, не затрагиваем психологические и социально-исторические аспекты этого понятия [7]. Речь идет только о понятии *математического доказательства*, причем в его современном понимании. Жертвуем в дальнейшем словом «математическое» лишь для краткости.

Заметим, что на практике только для самых простых утверждений приводится доказательство в указанном (уточненном) выше смысле. Обычные математические доказательства, как правило, носят *относительный характер*, когда наряду с аксиомами используются также утверждения, доказанные ранее или про которые известно, что они доказуемы. Для каждого такого предложения теоретически можно вставить в текст его собственное доказательство, устранив тем самым относительный характер доказательства, т. е. сведя все к аксиомам.

Кроме того, на практике, в кратком доказательстве, часто пропускаются некоторые посылки в умозаключениях. Они подразумеваются, но явно не оговариваются в рассуждении. Как правило, это известные, ранее доказанные утверждения. Эти неявно используемые посылки всегда при желании можно восстановить.

Для более точного описания понятия доказательства теперь нужно выявить и описать *логические правила*, используемые в рассуждениях и обеспечивающие элементарные шаги (логические переходы) в доказательстве.

Прежде чем перейти к этому описанию, сначала рассмотрим в качестве *примера* обычное несложное доказательство следующего утверждения:

если натуральное число не делится на 6 и делится на 9, то оно нечетно.

Запишем это утверждение, используя логическую символику:

$$6 \nmid n \,\&\, 9|n \to 2 \nmid n.$$

В дальнейшем, при символической записи содержательных (неформальных) предложений, будем использовать символы & (и), $\vee$ (или), $\to$ (если ..., то), $\neg$ (не). Если m не делится на n, наряду с записью $\neg(m|n)$ будем использовать запись $m \nmid n$.

В кратком виде традиционное доказательство этого утверждения представляет собой следующее рассуждение: «Согласно условию, n не делится на 6 и делится на 9. Поскольку n делится на 9, то n делится и на 3. Кроме того, n не делится на 6, а значит, n не делится на 2 или на 3. Следовательно, n не делится на 2».

Перечислим в виде последовательности предложения — члены этого краткого рассуждения, для обозримости предварительно записав их с помощью логической символики:

$$6 \nmid n \,\&\, 9|n; \quad 9|n; \quad 3|n; \quad 6 \nmid n; \quad 2 \nmid n \vee 3 \nmid n; \quad 2 \nmid n.$$

При такой форме доказательства, т. е. в виде цепочки утверждений, абсолютно неясно, какое утверждение из какого следует. Таким образом, совершенно не просматриваются логические связи между членами этой цепочки.

Теперь то же самое краткое рассуждение представим в виде дерева, выявив все логические связи между членами рассуждения, сделав сразу наглядным и понятным, какое предложение из каких следует.

Если предложение B непосредственно следует из предложений $A_1, \ldots, A_n$ по какому-либо правилу логики, будем привычным образом записывать это так:

$$\frac{A_1 \ \ldots \ A_n}{B}.$$

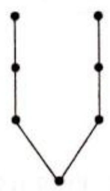
Если же на каком-то шаге рассуждения сделан сокращенный переход от предложений $A_1, \ldots, A_n$ к предложению B, который не является результатом применения какого-то логического правила, записывать это будем с помощью двойной черты следующим образом:

$$\frac{A_1 \;\ldots\; A_n}{B}.$$

Каждый такой сокращенный переход обычно возникает в результате пропуска посылки, являющейся общеизвестным утверждением, а также логических умозаключений, связанных с этой посылкой. Двойная черта означает, что эту пропущенную посылку и логические переходы можно восстановить.

Краткое рассуждение, приведенное выше, если его представить в виде дерева, имеет следующий вид (справа изображен граф, отражающий структуру дерева как частично упорядоченного множества):

$$\dfrac{\dfrac{\dfrac{6 \nmid n \,\&\, 9|n}{6 \nmid n}}{2 \nmid n \vee 3 \nmid n} \qquad \dfrac{\dfrac{6 \nmid n \,\&\, 9|n}{9|n}}{3|n}}{2 \nmid n} \tag{1}$$



Выявим теперь те логические правила, в соответствии с которыми проведены отдельные шаги рассуждения. Другими словами, выявим форму каждого логического перехода, отмеченного одной горизонтальной чертой. С этой целью обозначим буквами простые (элементарные) предложения, входящие в рассуждение.

Два первых (верхних) шага осуществлены согласно двум правилам, которые называются правилами удаления конъюнкции:

$$\frac{\mathcal{A} \,\&\, \mathcal{B}}{\mathcal{A}} \quad \text{и} \quad \frac{\mathcal{A} \,\&\, \mathcal{B}}{\mathcal{B}}.$$

Последний шаг, соответствующий самой нижней черте, осуществлен согласно следующему логическому правилу, являющемуся вариантом правила исключения возможных случаев:

$$\frac{\neg\mathcal{A} \vee \neg\mathcal{B} \quad \mathcal{B}}{\neg\mathcal{A}}.$$

Двойная черта в дереве (1), согласно договоренности, означает, что на этом шаге осуществлен сокращенный переход, который не является результатом применения какого-то логического правила. Восстановим пропущенную посылку и логические переходы в обоих случаях.

Рассмотрим сначала сокращенный переход, отраженный следующим образом:

$$\frac{6 \nmid n}{2 \nmid n \vee 3 \nmid n}.$$

Восстановив подразумеваемую посылку $2|n \,\&\, 3|n \to 6|n$, а также пропущенные логические переходы (умозаключения), получим более подробный вариант рассуждения. Представив восстановленную часть рассуждения в виде дерева (1), получим следующую картину:

$$\dfrac{\dfrac{6\nmid n \qquad \dfrac{2|n \,\&\, 3|n \to 6|n}{6\nmid n \to \neg(2|n \,\&\, 3|n)}}{\neg(2|n \,\&\, 3|n)}}{2\nmid n \vee 3\nmid n}.$$

Теперь часть дерева (1), расположенную над нижней чертой слева, можем заменить следующим деревом, отражающим более подробное рассуждение:

$$\dfrac{\dfrac{\dfrac{6\nmid n \,\&\, 9|n}{6\nmid n} \qquad \dfrac{2|n \,\&\, 3|n \to 6|n}{6\nmid n \to \neg(2|n \,\&\, 3|n)}}{\neg(2|n \,\&\, 3|n)}}{2\nmid n \vee 3\nmid n} \tag{2}$$

В этом дереве верхняя черта справа соответствует применению правила контрапозиции[2] $\dfrac{\mathcal{A} \supset \mathcal{B}}{\neg\mathcal{B} \supset \neg\mathcal{A}}$. Далее используется правило *modus ponens* $\dfrac{\mathcal{A} \quad \mathcal{A} \supset \mathcal{B}}{\mathcal{B}}$. Нижняя черта соответствует применению правила де Моргана $\dfrac{\neg(\mathcal{A}\&\mathcal{B})}{\neg\mathcal{A} \vee \neg\mathcal{B}}$.

В дереве (2) в качестве допущений использованы два предложения:

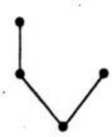
$6\nmid n \,\&\, 9|n$ (условие) и $2|n \,\&\, 3|n \to 6|n$ (известный из арифметики факт — если n делится на 2 и на 3, то n делится на 6).

Вторую часть дерева (1), содержащую сокращенный переход

$$\begin{array}{c} 9|n \\ \hline\hline 3|n \end{array},$$

можно заменить следующим деревом, отражающим более подробное рассуждение:

$$\dfrac{\dfrac{6\nmid n \,\&\, 9|n}{9|n} \qquad 9|n \to 3|n}{3|n} \tag{3}$$

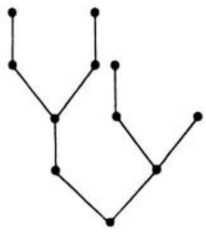


В этом дереве в качестве допущений фигурируют: условие $6\nmid n \,\&\, 9|n$ и утверждение $9|n \to 3|n$, выражающее известное из арифметики свойство делимости, которое может быть доказано отдельно. Логические переходы осуществляются в соответствии с правилами $\dfrac{\mathcal{A}\&\mathcal{B}}{\mathcal{B}}$ и $\dfrac{\mathcal{A} \quad \mathcal{A} \supset \mathcal{B}}{\mathcal{B}}$.

[2]При записи формул имело бы смысл использовать для логических связок символы, отличающиеся от символов, используемых для записи предложений. Однако мы заменим лишь символ $\to$ на символ $\supset$ (по техническим соображениям).

Если в дереве (1) два малых дерева заменить деревьями (2) и (3), то получим следующее дерево доказательства (4) с графической структурой, изображенной ниже:

$$\dfrac{\dfrac{\dfrac{\dfrac{6 \nmid n \,\&\, 9 \mid n}{6 \nmid n} \quad \dfrac{2 \mid n \,\&\, 3 \mid n \to 6 \mid n}{6 \nmid n \to \neg(2 \mid n \,\&\, 3 \mid n)}}{\neg(2 \mid n \,\&\, 3 \mid n)}}{2 \nmid n \vee 3 \nmid n} \qquad \dfrac{\dfrac{6 \nmid n \,\&\, 9 \mid n}{9 \mid n} \quad 9 \mid n \to 3 \mid n}{3 \mid n}}{2 \nmid n} \tag{4}$$



Исходными в этом дереве доказательства (4) служат следующие предложения:

$6 \nmid n \,\&\, 9|n$ — условие, которое используется дважды;

$2|n \,\&\, 3|n \to 6|n$ — утверждение из арифметики (из теории делимости), доказательство которого известно;

$9|n \to 3|n$ — утверждение из арифметики (из теории делимости), доказательство которого известно.

Чтобы указать различие между статусом исходных предложений, можно над каждым предложением, доказательство которого известно, провести двойную черту в знак того, что в этом месте можно вставить соответствующее доказательство в виде дерева. В результате получим следующее дерево:

$$\dfrac{\dfrac{\dfrac{\dfrac{6 \nmid n \,\&\, 9|n}{6 \nmid n} \quad \dfrac{\overline{\overline{2|n \,\&\, 3|n \to 6|n}}}{6 \nmid n \to \neg(2|n \,\&\, 3|n)}}{\neg(2|n \,\&\, 3|n)}}{2 \nmid n \vee 3 \nmid n} \qquad \dfrac{\dfrac{6 \nmid n \,\&\, 9|n}{9|n} \quad \overline{\overline{9|n \to 3|n}}}{3|n}}{2 \nmid n}. \tag{5}$$

Это дерево отражает то, что предложение $2 \nmid n$, фактически, выведено лишь из одного условия (допущения) $6 \nmid n \,\&\, 9|n$.

Заметим, что при необходимости справа от каждой черты можно указать символическое обозначение правила вывода, в соответствии с которым проведен шаг рассуждения.

Итак, мы упорядочили предложения, составляющие конкретное рассуждение, в виде дерева. В верхних, концевых точках «ветвей» дерева, так называемых «листьях» дерева, располагаются исходные предложения. В данном случае это допущения, одни из которых представляют собой так называемое «условие», другие —

ранее доказанные утверждения (таким образом, доказательство носит относительный характер). В каждом промежуточном узле на ветвях располагается предложение, которое получается с помощью некоторого логического правила из предложений, расположенных непосредственно над ним. Таким образом, непосредственно над каждым предложением, кроме исходных, записаны те предложения, из которых оно непосредственно следует по одному из правил логики. Корнем этого дерева служит доказываемое утверждение.

Замечание. Для того, чтобы построить доказательство в виде дерева, удобно и часто даже необходимо использовать символический язык для записи на нем содержательных предложений. Прежде всего, это позволяет выявить и сделать более наглядной логическую форму предложений, что, в свою очередь, позволяет выявить логическую форму умозаключений и используемые логические правила. Кроме того, использование символического языка позволяет компактно записывать как сами предложения, так и все доказательство в целом, делая его более наглядным и обозримым.

Если столь же подробное рассуждение представить в виде последовательности (цепочки) предложений, то получим последовательность, состоящую из тех же членов: $6 \nmid n \,\&\, 9|n$; $6 \nmid n \,\&\, 9|n \to 6 \nmid n$; $6 \nmid n$; $2|n \,\&\, 3|n \to 6|n$; $6 \nmid n \to \neg(2|n \,\&\, 3|n)$; $\neg(2|n \,\&\, 3|n)$; $2 \nmid n \vee 3 \nmid n$; $6 \nmid n \,\&\, 9|n \to 9|n$; $9|n$; $9|n \to 3|n$; $3|n$; $2 \nmid n$.

Очевидно, это традиционное *линейное доказательство*, хотя и выглядит более компактным, но, в отличие от дерева доказательства (4), лишено наглядности. Оно совершенно не отражает логические взаимосвязи между членами (логическую структуру доказательства). Если его снабдить комментарием, указывающим на логическую взаимосвязь между членами, то компактность исчезнет, а наглядности по-прежнему не будет:

(1) $6 \nmid n \,\&\, 9|n$ — допущение (условие);

(2) $6 \nmid n$ — логически следует из (1);

(3) $2|n \,\&\, 3|n \to 6|n$ — известное утверждение из теории делимости;

(4) $6 \nmid n \to \neg(2|n \,\&\, 3|n)$ — логически следует из (3);

(5) $\neg(2|n \,\&\, 3|n)$ — логически следует из (2) и (4);

(6) $2 \nmid n \vee 3 \nmid n$ — логически следует из (5);

(7) $9|n$ — логически следует из (1);

(8) $9|n \to 3|n$ — известное утверждение из теории делимости;

(9) $3|n$ — логически следует из (7) и (8);

(10) $2 \nmid n$ — логически следует из (6) и (9).

****

Теперь вернемся к обсуждению логических средств, используемых в доказательствах. Напомним, что пока не было дано описание тех логических правил, в соответствии с которыми строятся математические доказательства и которые лежат в их основе. Эти правила математики используют в своих рассуждениях явно или неявно, осознанно или не отдавая себе в этом отчета.

Итак, остаются открытыми следующие вопросы: какими логическими правилами пользуются математики в доказательствах? как описать все допустимые для использования правила? можно ли оптимальным образом организовать эти правила в некоторую систему?

Точное описание основных логических правил, используемых в математических доказательствах, предложил немецкий логик Герхард Генцен [1]. Эти правила принято называть *правилами вывода* или *правилами заключения*. Рассмотрим эти правила.

Описывая правила вывода, будем исходить из того, что читатель знаком с логическими связками: конъюнкцией (&), дизъюнкцией ($\vee$), импликацией ($\supset$) и отрицанием ($\neg$), а также имеет представление о том, что такое формула языка логики высказываний.

Правила вывода делятся на *правила введения* и *правила удаления*. Для каждой логической связки (&, $\vee$, $\supset$, $\neg$) имеется правило введения и правило удаления этой связки (иногда пара правил).

Справа от каждого правила в скобках указано его символическое обозначение (например, &*в* — обозначение правила введения конъюнкции, $\supset$*у* — обозначение правила удаления импликации):

Основные правила вывода

Правила введения	Правила удаления
$\dfrac{\mathcal{A}\quad\mathcal{B}}{\mathcal{A}\,\&\,\mathcal{B}}$ (&*в*)	$\dfrac{\mathcal{A}\,\&\,\mathcal{B}}{\mathcal{A}}\quad\dfrac{\mathcal{A}\,\&\,\mathcal{B}}{\mathcal{B}}$ (&*у*)
$\dfrac{\mathcal{A}}{\mathcal{A}\vee\mathcal{B}}\quad\dfrac{\mathcal{B}}{\mathcal{A}\vee\mathcal{B}}$ ($\vee$*в*)	$\dfrac{\mathcal{A}\vee\mathcal{B}\quad\begin{array}{c}[\mathcal{A}]\\ \mathcal{C}\end{array}\quad\begin{array}{c}[\mathcal{B}]\\ \mathcal{C}\end{array}}{\mathcal{C}}$ ($\vee$*у*)
$\dfrac{\begin{array}{c}[\mathcal{A}]\\ \mathcal{B}\end{array}}{\mathcal{A}\supset\mathcal{B}}$ ($\supset$*в*)	$\dfrac{\mathcal{A}\quad\mathcal{A}\supset\mathcal{B}}{\mathcal{B}}$ ($\supset$*у*)
$\dfrac{\begin{array}{c}[\mathcal{A}]\\ \mathcal{B}\end{array}\quad\begin{array}{c}[\mathcal{A}]\\ \neg\mathcal{B}\end{array}}{\neg\mathcal{A}}$ ($\neg$*в*)	$\dfrac{\mathcal{A}\quad\neg\mathcal{A}}{\mathcal{B}}$ ($\neg$*у*)

Приведенные в таблице правила используются как в классической, так и в неклассической (конструктивной) математике.

В *классической математике*, кроме указанных в таблице правил, используется еще одно основное правило — правило удаления двойного отрицания: $\dfrac{\neg\neg\mathcal{A}}{\mathcal{A}}$ ($\neg\neg$*у*).

Правила &*в*, &*у*, $\vee$*в*, $\supset$*у*, $\neg$*у*, $\neg\neg$*у* называются *прямыми* (или *безусловными*) правилами, а правила $\supset$*в*, $\vee$*у*, $\neg$*в* — *косвенными* (или *условными*).

Эти правила составляют полный список основных правил для логических систем *естественного* (*натурального*) вывода, построенных Генценом. Правила, приведенные в таблице, — это правила для интуиционистской системы, а если к ним добавить еще правило удаления двойного отрицания, то получим все основные правила для классической системы.

Обсудим ***содержательный смысл правил вывода.***

Правила вывода являются формализацией тех простейших правил умозаключений, в соответствии с которыми мы проводим неформальные математические рассуждения, часто не отдавая себе в этом отчета, правил перехода от одних утверждений к другим, «логически следующим» из первых.

Содержательный смысл прямых (безусловных) правил вывода достаточно прост и ясен. Так, правило $\&в$ формализует следующее рассуждение: «если обосновано утверждение A и обосновано утверждение B, то считаем обоснованным утверждение «A и B».

Правило $\supset у$ формализует рассуждение: «если обосновано утверждение A и обосновано условное утверждение «если A, то B», то мы считаем обоснованным утверждение B». Правило $\neg у$ формализует так называемый закон Дунса Скота: из противоречия A и *не*-A следует «все, что угодно», т. е. любое предложение (другими словами, исходя из противоречивых посылок, можно, рассуждая правильно, вывести как истинное, так и ложное предложение).

Условные правила являются формализацией рассуждений, которые используют вспомогательные допущения (рассуждений, исходящих из допущений).

Рассмотрим, например, условное правило введения импликации $\supset в$:

$$\frac{\begin{matrix}[\mathcal{A}]\\ \mathcal{B}\end{matrix}}{\mathcal{A} \supset \mathcal{B}} \supset в.$$

Это правило является формализацией рассуждения, широко применяемого в неформальных доказательствах. В самом деле, при доказательстве теорем, сформулированных в импликативной форме «если A, то B» («из A следует B»), обычно рассуждение ведется следующим образом. Утверждение A берется в качестве допущения (предположения), из которого выводится B. После того как, исходя из допущения A, утверждение B выведено (обосновано), делается заключение, что обосновано утверждение «если A, то B» (которому соответствует формула $A \supset B$). При этом неважно, является утверждение A доказуемым или нет, поскольку окончательное утверждение «если A, то B» не зависит от этого промежуточного допущения (как правило, оно недоказуемо).

В правиле $\supset в$, формализующем приведенное выше рассуждение, буква, заключенная в квадратные скобки, соответствует допущению A, с помощью которого обосновывается B и от которого уже не зависит $A \supset B$. Таким образом, формула B над чертой зависит от допущения A, а формула $A \supset B$ под чертой не зависит от A.

Рассмотрим в качестве примера ход содержательного доказательства известной теоремы из школьного курса геометрии: «если четырехугольник является ромбом, то его диагонали взаимно перпендикулярны». Доказательство обычно проводится следующим образом. Обозначаем через $ABCD$ произвольный четырехугольник. Делаем *допущение*, что четырехугольник $ABCD$ является ромбом. Затем, в несколько шагов, мы доказываем, что его диагонали взаимно перпендикулярны. Таким образом, мы доказываем, что $AC \perp BD$ при допущении, что $ABCD$ — ромб. Далее мы делаем следующий вывод, считая его вполне обоснованным: «если четырехугольник $ABCD$ является ромбом, то его диагонали взаимно перпендикулярны». При этом мы освобождаемся от допущения, что $ABCD$ — ромб. Это допущение было вспомогательным, промежуточным. Оно было необходимо только на стадии выведения из него утверждения $AC \perp BD$. Естественно, могли быть сделаны другие допущения, от которых наше заключение остается зависимым. В данном случае таковым можно считать допущение, что $ABCD$ — четырехугольник.

Напомним, что при доказательстве теорем вида «Если A, то B», мы часто выделяем то, что «дано», и то, что «требуется доказать». Фактически, мы декларируем, что для доказательства нашего утверждения достаточно, исходя из A как из допущения, обосновать B. Именно такой ход рассуждения выявляет и отражает правило $\supset$ *в*. Таким образом, это правило отражает правомерность сведения задачи на доказательство утверждения «Если A, то B» к задаче на доказательство утверждения B, исходя из условия (допущения) A.

Большой интерес представляют и два других условных правила: правило удаления дизъюнкции и правило введения отрицания.

Рассмотрим правило $\vee$*у*. Оно формализует так называемое доказательство *разбором случаев*: если обосновано утверждение $A \vee B$ («A или B»), то для обоснования C рассматриваем два случая. Сначала, в предположении, что имеет место A (первый случай), из допущения A выводим C. Далее, в предположении, что имеет место B (второй случай), из допущения B также выводим C. После этого считаем обоснованным утверждение C, т. е. что C имеет место независимо от обоих допущений A и B.

Наконец, рассмотрим правило $\neg$*в*. Оно формализует доказательство *приведением к нелепости* (лат. — *reductio ad absurdum*): если из допущения A мы выводим противоречащие друг другу предложения B и *не*-B ($\neg B$), то считаем, что наше допущение A неверно, а верно, обосновано *не*-A ($\neg A$).

Замечание. Не следует путать метод *доказательства приведением к нелепости* (и соответствующее ему правило $\neg$*в*, которое также называется правилом доказательства приведением к нелепости) с методом *доказательства от противного*. Правило доказательства от противного имеет вид

$$\frac{\overset{[\neg\mathcal{A}]}{\mathcal{B}} \quad \overset{[\neg\mathcal{A}]}{\neg\mathcal{B}}}{\mathcal{A}}.$$

Правило доказательства от противного в скрытой форме опирается на классическое правило снятия двойного отрицания и поэтому является источником неэффективных доказательств теорем существования. Различия между методами (и правилами) доказательства от противного и приведением к нелепости носят существенный, принципиальный характер. Эти различия достаточно подробно обсуждаются в статье [5].

Разумеется, в практике обычных математических рассуждений никто не ограничивает себя только этими *основными* правилами, хотя их вполне достаточно. Широко используются также, например, следующие известные производные (выводимые) правила:

1) $\dfrac{\mathcal{A} \supset \mathcal{B}}{\neg\mathcal{B} \supset \neg\mathcal{A}}$ — правило контрапозиции;

2) $\dfrac{\mathcal{A} \supset \mathcal{B} \quad \mathcal{B} \supset \mathcal{C}}{\mathcal{A} \supset \mathcal{C}}$ — правило силлогизма;

3) $\dfrac{\mathcal{A} \vee \mathcal{B} \quad \neg\mathcal{B}}{\mathcal{A}}$ — правило доказательства исключением случаев;

4) $\dfrac{\neg(\mathcal{A} \vee \mathcal{B})}{\neg\mathcal{A} \& \neg\mathcal{B}}, \dfrac{\neg\mathcal{A} \& \neg\mathcal{B}}{\neg(\mathcal{A} \vee \mathcal{B})}, \dfrac{\neg\mathcal{A} \vee \neg\mathcal{B}}{\neg(\mathcal{A} \& \mathcal{B})}$ — правила де Моргана;

5) $\dfrac{\neg(\mathcal{A} \& \mathcal{B})}{\neg\mathcal{A} \vee \neg\mathcal{B}}$ — правило де Моргана (классическое);

6) $\dfrac{\mathcal{A} \supset \mathcal{B}}{\neg\mathcal{A} \vee \mathcal{B}}$ — правило выражения $\supset$ через $\vee$ и $\neg$ (классическое).

Замечательным, однако, является тот факт, что основных правил достаточно, чтобы провести любое математическое доказательство (не использующее кванторы). А любое другое используемое на практике правило может быть получено точно описанным образом (выведено) из основных правил. Поэтому такие правила называют *выводимыми* или *производными*. Уточнив используемые понятия, можно доказать соответствующую теорему, выражающую полноту приведенной системы правил в указанном выше смысле. Заметим, что в рассмотренном ранее доказательстве были использованы не только основные правила, но и производные.

Следует отметить, что существуют также правила вывода, используемые в рассуждениях с кванторами («для любого» и «существует»). Однако в этой статье рассматривать их не будем. Это тема для отдельного разговора.

Логические выводы в системе натурального вывода, так называемые *деревья вывода*, представляют собой деревья формул, в которых все переходы происходят по правилам вывода (заключения) этой системы. Понятие дерева вывода имеет точное математическое определение (индуктивное определение) [6].

Понятие *дерева вывода* (вывода в виде дерева) представляет собой *математическое уточнение* понятия доказательства в виде дерева, а сами выводы в виде дерева являются *математическими моделями* неформальных математических доказательств.

Вернемся к рассмотренному ранее примеру доказательства в виде дерева.

Если вместо предложений в узлах дерева (4) поместить соответствующие формулы, являющиеся формализацией этих предложений (и отражающие логическую структуру этих предложений), то получится следующее *дерево вывода* (точнее, дерево квазивывода, поскольку кроме основных правил используются и производные). Оно полностью отражает логическую структуру нашего содержательного доказательства:

$$\dfrac{\dfrac{\dfrac{\dfrac{\neg A\&B}{\neg A} \qquad \dfrac{D\&C\supset A}{\neg A\supset\neg(D\&C)}}{\neg(D\&C)}}{\neg D\vee\neg C} \qquad \dfrac{\dfrac{\neg A\&B}{B} \qquad B\supset C}{C}}{\neg D}. \tag{4'}$$

В исходных (верхних) узлах этого дерева вывода расположены следующие формулы, являющиеся формализацией соответствующих исходных предложений в дереве (4): $\neg A\&B$, $D\&C\supset A$, $B\supset C$.

Если формализовать содержательное линейное доказательство, то получим следующую последовательность формул:

$$\neg A\&B,\ \neg A,\ D\&C\supset A,\ \neg A\supset\neg(D\&C),\ \neg(D\&C),\ \neg D\vee\neg C,\ B,\ B\supset C,\ C,\ \neg D.$$

Получили традиционный формальный так называемый *линейный вывод* (точнее квазивывод, в котором использованы производные правила). Этот линейный вывод, в отличие от дерева вывода (4'), лишен наглядности. Он не отражает логические взаимосвязи между членами (логическую структуру вывода).

****

Для логического анализа конкретного доказательства была выявлена форма каждого предложения, а также форма каждого шага рассуждения. Это позволило выявить и форму всего доказательства в целом.

Очевидно, что необходимость в столь подробном представлении доказательства, с выявлением отдельных шагов, возникает, только если нужно проанализировать его логическую структуру.

Метод, который был использован для анализа конкретного доказательства, называется методом формализации и является одним из основных методов математической логики. *Метод формализации* был предложен Д.Гильбертом (D. Hilbert, 1862–1943) как метод изучения математических доказательств и математических теорий с целью дальнейшего доказательства непротиворечивости этих теорий.

Благодаря Гильберту возник основной раздел математической логики — *теория доказательств*. Важнейшей задачей теории доказательств является математическое *уточнение понятия доказательства*.

Метод формализации является основным методом, позволяющим разобраться в том, что такое доказательство, и выработать математическое уточнение этого понятия.

В математической логике существуют *два основных типа* математического уточнения понятия доказательства. Уточнения первого типа восходят к Гильберту, и представляют собой *линейные выводы* в логических аксиоматических системах, называемых системами гильбертовского типа.

Уточнения второго типа принадлежат крупному логику, ученику Гильберта, Г. Генцену (G. Gentzen, 1909–1945) и представляют собой *выводы в виде дерева* в системах естественного (натурального) вывода [1].

Уточнение доказательства в виде дерева вывода в системах натурального вывода имеет ряд преимуществ по сравнению с линейным выводом в системах гильбертовского типа. Суть этих преимуществ заключается в том, что деревья натурального вывода представляют собой наиболее естественную и адекватную модель обычных математических доказательств.

Лирическое отступление

Не следует воспринимать рассмотренный метод анализа доказательства как универсальное средство для всеобщего и регулярного использования в математической практике доказательств. Даже довольно простое математическое доказательство становится весьма громоздким, если в нем полностью выявить все неявно используемые логические средства, все подразумеваемые, но пропущенные посылки и логические шаги. Человеческая же мысль настолько быстра, что осуществляет все логические переходы сокращенно. Обычно реальное математическое доказательство представляет собой свернутое, краткое рассуждение, в котором пропущены многие шаги, не формулируются явно некоторые посылки и, тем более, логические переходы. Главная цель такого рассуждения — добраться до конца, до слов «что и требовалось доказать».

Когда же разумно прибегать к достаточно полной формализации обычного доказательства (т. е. выявлению его формы)? Тогда, когда целью является исследование имеющегося доказательства, изучение без пропусков всех логических переходов в нем, всех используемых условий, посылок и допущений.

Можно провести следующую аналогию. Сравним неформальное доказательство с айсбергом. Реально излагаемая, «видимая» его часть, невелика, как и видимая часть айсберга. А вот «невидимая» часть огромна, как и невидимая часть айсберга.

Приходит в голову еще одно образное сравнение. В кровеносных сосудах человека течет кровь. При этом его организм функционирует, а сам человек, как правило, не думает о деталях того, как протекает этот процесс. Но в какой-то момент возникает необходимость выяснить, что происходит с его кровью. Это может быть важно как для самого человека, так и для специалиста (например, врача). В этом случае приходится прибегать к помощи микроскопа или каких-либо других специальных средств и исследовать во всех деталях, из каких частиц состоит кровь. И тогда капля крови превращается в нечто огромное, состоящее из массы частиц, ранее невидимых невооруженным глазом. Такое исследование нужно проводить не ежедневно и не каждому. Обычно его проводит специалист, и только тогда, когда это необходимо. Но в определенной степени каждый должен осознавать, в каком состоянии находится его здоровье.

Так и математические рассуждения, выстраивающиеся в доказательства, — компактны, кратки, многие их детали не видны. Человек может охватить мыслью некоторый достаточно сложный логический переход, не формулируя используемые логические правила в явной форме, и даже не отдавая себе отчет во всех деталях. Если же доказательство требуется подвергнуть исследованию, изучить все его детали, его структуру (может быть, для того чтобы исключить возможность логической ошибки или выявить такую ошибку), то для этого приходится использовать специальные средства. Эти средства и предоставляет математическая логика, теория доказательств. В частности, таким средством является формализация. При формализации доказательство, как под микроскопом, сильно увеличивается в размерах, и становятся видны все его многочисленные детали.

Каждый, кто имеет отношение к математике, *в определенной степени должен осознавать*, какие логические средства он использует в доказательствах. Однако полное исследование математических доказательств во всех деталях и подробностях может осуществить только тот специалист, который владеет методами теории доказательств. И если инженер, изучающий математику, не обязан быть таким специалистом, то учитель математики, безусловно, должен владеть соответствующими знаниями, т. е. основами теории доказательств.

****

Итак, можно сделать следующие ***выводы***.

Уточнение интуитивного понятия доказательства необходимо для понимания сущности этого понятия и для возможности проведения анализа рассуждения на предмет выявления используемых в нем логических средств, а также для выявления и устранения в нем логических ошибок. Доказательство в виде дерева более наглядно отражает логические взаимосвязи между его членами по сравнению с линейным доказательством. Уточнение понятия доказательства, как доказательства в виде дерева, лучше отражает сущность понятия доказательства.

Математическое уточнение понятия доказательства как формального вывода в той или иной логической системе необходимо для изучения этого понятия математическими средствами в теории доказательств. Только располагая математическим уточнением понятия доказательства, можно, например, доказать, что какое-то предложение не доказуемо в той или иной аксиоматической системе.

Математическое уточнение доказательства в виде дерева вывода в системах естественного (натурального) вывода представляет собой наиболее естественную и адекватную модель обычных содержательных доказательств, реально проводимых в математике.

Литература

[1] Генцен Г. Исследования логических выводов // Математическая теория логического вывода. — М.: Наука, 1967. С.9 — 74.

[2] Гладкий А.В. Математическая логика. — М.: Российский государственный гуманитарный университет, 1998. - 480 с.

[3] Гладкий А.В. Введение в современную логику. — М.: МЦНМО, 2001. - 200 с.

[4] Смаллиан Р. М. Принцесса или тигр? — М.: Мир, 1985. — 221 с.

[5] Тимофеева И.Л. Некоторые замечания о методе доказательства от противного// Математика в школе. — 1994, №3. С.36 — 38.

[6] Тимофеева И.Л. Математическая логика в вопросах и задачах. — М.: Прометей, 2002. 112 с.

[7] Успенский В.А. Семь размышлений на темы философии математики // Закономерности развития современной математики. — М.: Наука, 1987.

Тимофеева Ирина Леонидовна,
кандидат педагогических наук,
доцент кафедры математического анализа МПГУ.

Неравенство Ки Фана и его обобщения

С. И. Калинин

В данной статье мы намереваемся рассмотреть хорошо известное в тематике средних величин неравенство Ки Фана, представляющее собой своеобразный аналог классического неравенства Коши между средним арифметическим и средним геометрическим положительных чисел. В последние годы упоминаемое неравенство является объектом внимания многих авторов, чьи исследования связаны с неравенствами и средними степенными величинами. Получены всевозможные развития, обобщения и уточнения этого неравенства, описаны его геометрические иллюстрации и интерпретации, найдены некоторые приложения. По сути дела, специалисты в области неравенств его относят к разряду классических.

Из литературных источников автору данных строк известно несколько доказательств неравенства Ки Фана. Целью настоящей работы является рассмотрение еще двух новых доказательств. Кроме того, мы ставим задачу – привести некоторые обобщения интересующего нас неравенства.

Отметим, излагаемый ниже материал будет доступен школьникам старших классов, которые увлечены математикой. Мы надеемся на то, что он окажется полезным учителям математики, работающим по углубленной программе, при изучении соответствующих тем школьного курса математики, а также при организации и проведении внеклассной работы по предмету.

Обсуждаемое неравенство и связанные с ним результаты мы адресуем и студентам-математикам педвузов и университетов с целью проведения дальнейших исследований в затрагиваемой области.

§ 1. Среднее степенное и взвешенное среднее степенное положительных чисел

1. Среднее степенное n положительных чисел. Пусть $a_1, \ldots, a_n$ $(n \geq 2)$ — положительные числа. Их *средним степенным порядка* x называется величина

$$F(x) \equiv F_{a_1,\ldots,a_n}(x) = \begin{cases} \left(\frac{1}{n}\sum\limits_{i=1}^{n} a_i^x\right)^{\frac{1}{x}}, & x \neq 0, \\ \sqrt[n]{a_1 \cdot \ldots \cdot a_n}, & x = 0. \end{cases} \qquad (1)$$

Очевидно, значения F при $x = 1,\ 0,\ -1,\ 2$ есть классические средние — среднее арифметическое $A_n = \frac{a_1+...+a_n}{n}$, среднее геометрическое $G_n = \sqrt[n]{a_1 \cdot ... \cdot a_n}$, среднее гармоническое $H_n = \frac{n}{\frac{1}{a_1}+...+\frac{1}{a_n}}$ и среднее квадратичное $R_n = \sqrt{\frac{a_1^2+...+a_n^2}{n}}$ соответственно данной совокупности чисел. Таким образом, (1) обобщает приведенные широко используемые в математике и приложениях средние величины A_n, G_n, H_n, R_n, которые, как известно, друг с другом связаны соотношениями

$$H_n \le G_n \le A_n \le R_n, \tag{2}$$

где неравенство $G_n \le A_n$ есть знаменитое неравенство Коши.

Легко видеть, что если $a_1 = ... = a_n = a$, то функция $F(x)$ при любом x принимает одно и то же значение, равное a, т.е. является постоянной.

Известны (см., напр., [1]) следующие свойства среднего степенного (1). Функция $F(x)$ является строго положительной, непрерывной, неубывающей на всей числовой прямой, при этом при $x \to \pm\infty$ она стремится соответственно к $\max(a_1, ...,\ a_n)$ и $\min(a_1, ...,\ a_n)$. Таким образом, в силу отмеченных свойств, при любом x ($x \in \boldsymbol{R}$) величина (1) есть среднее значение чисел $a_1, \ldots, a_n$.

Заметим, соотношения (2) вытекают из монотонности (1).

2. Взвешенное среднее степенное n положительных чисел. *Взвешенным средним степенным порядка* x положительных чисел $a_1, \ldots, a_n$ ($n \ge 2$) называется величина

$$\tilde{F}(x) \equiv F_{a_1,...,\,a_n}^{(p_1,...,\,p_n)} = \begin{cases} \left(\frac{p_1 a_1^x+...+p_n a_n^x}{p_1+...+p_n}\right)^{\frac{1}{x}}, & x \ne 0, \\ (a_1^{p_1} \cdot ... \cdot a_n^{p_n})^{\frac{1}{p_1+...+p_n}}, & x = 0, \end{cases} \tag{3}$$

где $p_1, ..., p_n$ — положительные числа, именуемые в е с а м и. Нетрудно видеть, что среднее степенное (1) есть частный случай взвешенного среднего степенного (3): (1) получается из (3) при условии $p_1 = \ldots = p_n$. Значения $\tilde{F}$ при $x = 1,\ 0,\ -1,\ 2$ порождают соответственно классические взвешенные средние: арифметическое $\tilde{A}_n = \frac{p_1 a_1+...+p_n a_n}{p_1+...+p_n}$, геометрическое $\tilde{G}_n = (a_1^{p_1} \cdot ... \cdot a_n^{p_n})^{\frac{1}{p_1+...+p_n}}$, гармоническое $\tilde{H}_n = \frac{p_1+...+p_n}{\frac{p_1}{a_1}+...+\frac{p_n}{a_n}}$ и квадратичное $\tilde{R}_n = \sqrt{\frac{p_1 a_1^2+...+p_n a_n^2}{p_1+...+p_n}}$. Подобно (2), для взвешенных средних величин $\tilde{A}_n$, $\tilde{G}_n$, $\tilde{H}_n$, $\tilde{R}_n$ имеют место соотношения

$$\tilde{H}_n \le \tilde{G}_n \le \tilde{A}_n \le \tilde{R}_n. \tag{4}$$

В (4) неравенство $\tilde{G}_n \le \tilde{A}_n$ часто называют о б о б щ е н н ы м, или в е с о в ы м неравенством Коши.

Можно показать, что взвешенное среднее степенное (3) наследует все выше перечисленные свойства среднего степенного (1). Следовательно, справедливость соотношений (4) вытекает из монотонности функции $\tilde{F}$.

Замечание. Взвешенное среднее степенное (3) можно задать также формулой

$$\tilde{F}(x) = \begin{cases} (p_1 a_1^x + \cdots + p_n a_n^x)^{\frac{1}{x}}, & x \ne 0, \\ a_1^{p_1} \cdot \ldots \cdot a_n^{p_n}, & x = 0, \end{cases} \tag{5}$$

где весы $p_1, \ldots, p_n$ удовлетворяют условию $p_1 + \cdots + p_n = 1$. Ниже мы будем использовать обе формы задания рассматриваемой средней величины.

§2. Неравенство Ки Фана

Пусть теперь $a_1, \ldots, a_n$ — положительные числа из промежутка $\left(0; \frac{1}{2}\right]$. Тогда числа $1 - a_1, \ldots, 1 - a_n$ будут также положительными, и можно ввести в рассмотрение среднее степенное порядка x ($x \in \mathbb{R}$) для последней совокупности чисел: $\hat{F}(x) \equiv F_{1-a_1,\ldots,1-a_n}(x)$. Условимся использовать следующие общепринятые в тематике средних обозначения: $A'_n = \hat{F}(1)$, $G'_n = \hat{F}(0)$. Упоминаемое в названии параграфа неравенство Ки Фана есть неравенство

$$\frac{G_n}{G'_n} \leq \frac{A_n}{A'_n}, \tag{$*$}$$

в котором равенство достигается тогда и только тогда, когда $a_1 = \ldots = a_n$. Это неравенство имеет соответствующее сходство с классическим неравенством Коши, оно также связывает арифметико-геометрические средние положительных чисел.

По-видимому, неравенство ($*$) впервые было опубликовано в 1961 году в монографии [2]. В издании [3] цитируемой книги на русском языке авторы на с. 14 на него ссылаются как на «неопубликованное неравенство Фань Цзы, которое может быть доказано индукцией вверх и вниз». В следующем параграфе мы приведем доказательство неравенства ($*$) этим методом.

Наряду с неравенством ($*$) условимся также рассматривать обобщенное (весовое) неравенство Ки Фана

$$\frac{\tilde{G}_n}{\tilde{G}'_n} \leq \frac{\tilde{A}_n}{\tilde{A}'_n}, \tag{$**$}$$

где $\tilde{G}'_n$, $\tilde{A}'_n$ — взвешенные среднее геометрическое и среднее арифметическое соответственно чисел $1 - a_1, \ldots, 1 - a_n$ с тем же набором весов $p_1, \ldots, p_n$, что используется при описании средних $\tilde{G}_n$, $\tilde{A}_n$ совокупности чисел $a_1, \ldots, ; a_n$. Неравенство ($**$) впервые было приведено в работе [4] от 1990 года (в цитируемой статье в отношении его доказательства делается ссылка на работу [5], еще находившуюся в печати). Позднее появилось несколько различных доказательств этого неравенства. В данной работе мы рассмотрим еще два новых доказательства, использующих разные подходы. Кроме того, приведем доказательства, основанные на применении неравенства Иенсена.

Замечание. В соответствии с замечанием предыдущего параграфа подчеркнем, что неравенство ($**$) может быть записано также в форме

$$\frac{a_1^{p_1} \cdot \ldots \cdot a_n^{p_n}}{(1-a_1)^{p_1} \cdot \ldots \cdot (1-a_n)^{p_n}} \leq \frac{p_1 a_1 + \ldots + p_n a_n}{p_1(1-a_1) + \ldots + p_n(1-a_n)}, \quad p_1 + \ldots + p_n = 1. \tag{$***$}$$

§3. Доказательство неравенства Ки Фана методом прямой и обратной индукции

Установим неравенство (*) методом, который авторы монографии [3] называют индукцией вверх и вниз и который в современных руководствах по неравенствам чаще именуется методом прямой и обратной индукции.

Докажем сначала справедливость (*) для $n = 2^m, m \in$N, т.е. покажем, что является верным неравенство

$$\frac{\sqrt[2^m]{a_1 \cdot \ldots \cdot a_{2^m}}}{\sqrt[2^m]{(1-a_1)\cdot \ldots \cdot (1-a_{2^m})}} \le \frac{a_1 + \ldots + a_{2^m}}{(1-a_1) + \ldots + (1-a_{2^m})} \qquad (m \in \mathbb{N}), \tag{6}$$

при этом равенство в (6) будет достигаться лишь при условии $a_1 = \ldots = a_{2^m}$.

Для доказательства (6) применим метод математической индукции. Сначала установим базу индукции. При $m = 1$ для неравенства (6) имеем следующую цепочку равносильных соотношений

$$\frac{\sqrt{a_1 a_2}}{\sqrt{(1-a_1)(1-a_2)}} \le \frac{a_1 + a_2}{(1-a_1)+(1-a_2)} \Leftrightarrow$$
$$\Leftrightarrow \frac{a_1 a_2}{(1-a_1)(1-a_2)} \le \frac{(a_1+a_2)^2}{[(1-a_1)+(1-a_2)]^2} \Leftrightarrow$$
$$\Leftrightarrow a_1 a_2 \left[(1-a_1)^2 + 2(1-a_1)(1-a_2) + (1-a_2)^2\right] \le (1-a_1)(1-a_2)(a_1+a_2)^2 \Leftrightarrow$$
$$\Leftrightarrow a_1 a_2 \left[(1-a_1)^2 + (1-a_2)^2\right] \le (1-a_1)(1-a_2)(a_1^2+a_2^2) \Leftrightarrow$$
$$\Leftrightarrow 2a_1a_2\left[1-(a_1+a_2)\right] \le (a_1^2+a_2^2)\left[1-(a_1+a_2)\right].$$

Последнее неравенство является верным в силу неравенства Коши, причем равенство в нем будет иметь место лишь при условии $a_1 = a_2$. База индукции установлена.

Пусть (6) справедливо для $m = k$, т.е. имеет место неравенство

$$\frac{\sqrt[2^k]{a_1 \cdot \ldots \cdot a_{2^k}}}{\sqrt[2^k]{(1-a_1)\cdot \ldots \cdot (1-a_{2^k})}} \le \frac{a_1 + \ldots + a_{2^k}}{(1-a_1) + \ldots + (1-a_{2^k})}, \tag{7}$$

при этом равенство в нем может быть только в случае $a_1 = \ldots = a_{2^k}$. Установим (6) для $m = k+1$.

В силу индукционного предположения, наряду с (7) можем записать и неравенство

$$\frac{\sqrt[2^k]{a_{2^k+1} \cdot \ldots \cdot a_{2^{k+1}}}}{\sqrt[2^k]{(1-a_{2^k+1})\cdot \ldots \cdot (1-a_{2^{k+1}})}} \le \frac{a_{2^k+1} + \ldots + a_{2^{k+1}}}{(1-a_{2^k+1}) + \ldots + (1-a_{2^{k+1}})}, \tag{8}$$

в котором равенство может достигаться только при условии $a_{2^k+1} = \ldots = a_{2^{k+1}}$. Перемножая (7) и (8), будем иметь:

$$\frac{\sqrt[2^k]{a_1 \cdot \ldots \cdot a_{2^k} \cdot a_{2^k+1} \cdot \ldots \cdot a_{2^{k+1}}}}{\sqrt[2^k]{(1-a_1)\cdot \ldots \cdot (1-a_{2^k}) \cdot (1-a_{2^k+1}) \cdot \ldots \cdot (1-a_{2^{k+1}})}} \le$$
$$\le \frac{a_1 + \ldots + a_{2^k}}{(1-a_1) + \ldots + (1-a_{2^k})} \cdot \frac{a_{2^k+1} + \ldots + a_{2^{k+1}}}{(1-a_{2^k+1}) + \ldots + (1-a_{2^{k+1}})},$$

где равенство будет обеспечиваться лишь при выполнении условий $a_1 = \ldots = a_{2^k}$, $a_{2^k+1} = \ldots = a_{2^{k+1}}$. Извлекая, далее, корень квадратный из обеих частей последнего соотношения и используя базу индукции, получаем нужное неравенство:

$$\frac{\sqrt[2^{k+1}]{a_1 \cdot \ldots \cdot a_{2^{k+1}}}}{\sqrt[2^{k+1}]{(1-a_1) \cdot \ldots \cdot (1-a_{2^{k+1}})}} \leq$$

$$\leq \frac{\sqrt{a_1 + \ldots + a_{2^k}}}{\sqrt{(1-a_1) + \ldots + (1-a_{2^k})}} \cdot \frac{\sqrt{a_{2^k+1} + \ldots + a_{2^{k+1}}}}{\sqrt{(1-a_{2^k+1}) + \ldots + (1-a_{2^{k+1}})}} =$$

$$= \frac{\sqrt{\frac{a_1+\ldots+a_{2^k}}{2^k}}}{\sqrt{1-\frac{a_1+\ldots+a_{2^k}}{2^k}}} \frac{\sqrt{\frac{a_{2^k+1}+\ldots+a_{2^{k+1}}}{2^k}}}{\sqrt{1-\frac{a_{2^k+1}+\ldots+a_{2^{k+1}}}{2^k}}} \leq$$

$$\leq \frac{a_1 + \ldots + a_{2^k} + a_{2^k+1} + \ldots + a_{2^{k+1}}}{(1-a_1) + \ldots + (1-a_{2^k}) + (1-a_{2^k+1}) + \ldots + (1-a_{2^{k+1}})}.$$

В произведенных оценках равенство достигается лишь тогда, когда

$$a_1 = \ldots = a_{2^{k+1}}.$$

Неравенство (6) для $m = k+1$ полностью обосновано. Это обеспечивает индукционный переход, а значит, (6) доказано.

Проведенные рассуждения отвечают прямой индукции при доказательстве неравенства $(*)$. Проведем теперь рассуждения, соответствующие обратной индукции при обосновании этого неравенства.

Пусть неравенство $(*)$ выполняется для $n = k$ $(k > 2)$, т.е. справедливо неравенство

$$\frac{\sqrt[k]{a_1 \cdot \ldots \cdot a_k}}{\sqrt[k]{(1-a_1) \cdot \ldots \cdot (1-a_k)}} \leq \frac{a_1 + \ldots + a_k}{(1-a_1) + \ldots + (1-a_k)}, \tag{9}$$

где равенство возможно лишь при равенстве всех чисел a_i, $i = 1, \ldots, k$; докажем его для $n = k-1$. Для этого в (9) положим $a_k = A_{k-1}$. Будем иметь:

$$\frac{\sqrt[k]{a_1 \cdot \ldots \cdot a_{k-1} A_{k-1}}}{\sqrt[k]{(1-a_1) \cdot \ldots \cdot (1-a_{k-1}) A'_{k-1}}} \leq \frac{a_1 + \ldots + a_{k-1} + A_{k-1}}{(1-a_1) + \ldots + (1-a_{k-1}) + A'_{k-1}} = \frac{A_{k-1}}{A'_{k-1}}.$$

Отсюда получаем соотношение

$$\left(\frac{G_{k-1}}{G'_{k-1}}\right)^{\frac{k-1}{k}} \leq \left(\frac{A_{k-1}}{A'_{k-1}}\right)^{1-\frac{1}{k}},$$

равносильное $(*)$ при $n = k-1$, причем равенство в нем будет достигаться лишь при совпадении чисел $a_1, \ldots, a_{k-1}$. Обратная индукция реализована. Неравенство Ки Фана $(*)$ полностью обосновано.

§4. О двух новых доказательствах обобщенного неравенства Ки Фана

1. Доказательство I. Установим сначала вспомогательные утверждения, нужные нам в дальнейшем для проведения первого из упоминаемых доказательств.

Предложение 1. *Пусть $f(x)$ — положительная дважды дифференцируемая на промежутке l числовой прямой функция. Если $(\ln(f(x)))'' > 0$, $x \in l$, то f — строго выпуклая на l функция.*

Доказательство. Так как

$$(\ln(f(x)))'' = \left(\frac{f'(x)}{f(x)}\right)' = \frac{f''(x)\,f(x) - (f'(x))^2}{f^2(x)} > 0, \qquad x \in l,$$

то $f''(x)f(x) > \left(f'(x)\right)^2$, $x \in l$. Отсюда, в силу положительности f, имеем условие $f''(x) > 0$ $(x \in l)$. Последнее влечет строгую выпуклость f на l. Предложение доказано.

Предложение 2. *Пусть a и b $(a > b$ — положительные числа, не превосходящие $\frac{1}{2}$. Тогда для любых q_1, q_2 $(q_1 > 0,\ q_2 > 0)$ справедливо неравенство*

$$\left(\frac{a}{1-a}\right)^{q_1}\left(\frac{b}{1-b}\right)^{q_2} < \left(\frac{q_1 a + q_2 b}{q_1(1-a) + q_2(1-b)}\right)^{q_1+q_2}. \tag{10}$$

Доказательство. Введем в рассмотрение функцию

$$f(x) = \left(\frac{a}{1-a}\right)^{x}\left(\frac{b}{1-b}\right)^{1-x}\frac{x(1-a) + (1-x)(1-b)}{xa + (1-x)b}.$$

Очевидно, на отрезке $[0;1]$ она положительна. Покажем, что на этом отрезке функция f строго выпукла. Для этого, в силу предложения 1, достаточно установить, что $(\ln(f(x)))'' > 0$ при $x \in (0;1)$. Имеем:

$$(\ln(f(x)))'' = \left(\ln\frac{x(1-a) + (1-x)(1-b)}{xa + (1-x)b}\right)'' =$$

$$= (b-a)\left(\frac{1}{[x(1-a) + (1-x)(1-b)]\,[xa + (1-x)b]}\right)' =$$

$$= 2(b-a)^2\,\frac{\left(\frac{1}{2} - b\right) + x(b-a)}{[x(1-a) + (1-x)(1-b)]^{\,2}\,[xa + (1-x)b]^{\,2}} > 0, \quad x \in (0;1).$$

Строгая выпуклость f на отрезке $[0;1]$ показана. Так как $f(0) = f(1) = 1$, то отсюда имеем условие $f(x) < 1$, $x \in (0;1)$. В частности, при $x = \frac{q_1}{q_1+q_2}$ получаем неравенство

$$\left(\frac{a}{1-a}\right)^{\frac{q_1}{q_1+q_2}}\left(\frac{b}{1-b}\right)^{\frac{q_2}{q_1+q_2}}\frac{\frac{q_1}{q_1+q_2}(1-a) + \frac{q_2}{q_1+q_2}(1-b)}{\frac{q_1}{q_1+q_2}a + \frac{q_2}{q_1+q_2}b} < 1,$$

которое равносильно (10). Предложение доказано.

Замечание 1. Если в условиях предложения $2a = b$, то неравенство (10) перейдет в равенство.

Введем теперь в рассмотрение величины

$$y_k = \frac{\sum\limits_{i=1}^{k} p_i a_i}{\sum\limits_{i=1}^{k} p_i}, \quad k = 1, \ldots, n. \tag{11}$$

Справедливо

Предложение 3. *Если* $y_k < a_{k+1}$ $(1 \le k \le n-1)$, *то*

$$\left(\frac{y_k}{1-y_k}\right)^{\sum\limits_{i=1}^{k} p_i} \left(\frac{a_{k+1}}{1-a_{k+1}}\right)^{p_{k+1}} < \left(\frac{y_{k+1}}{1-y_{k+1}}\right)^{\sum\limits_{i=1}^{k+1} p_i}$$

Доказательство. В силу предложения 2, имеем:

$$\left(\frac{y_k}{1-y_k}\right)^{\sum\limits_{i=1}^{k} p_i} \left(\frac{a_{k+1}}{1-a_{k+1}}\right)^{p_{k+1}} < \left(\frac{\left(\sum\limits_{i=1}^{k} p_i\right) y_k + p_{k+1} a_{k+1}}{\left(\sum\limits_{i=1}^{k} p_i\right)(1-y_k) + p_{k+1}(1-a_{k+1})}\right)^{\sum\limits_{i=1}^{k+1} p_i} =$$

$$= \left(\frac{\sum\limits_{i=1}^{k+1} p_i a_i}{\sum\limits_{i=1}^{k+1} p_i (1-a_i)}\right)^{\sum\limits_{i=1}^{k+1} p_i} = \left(\frac{y_{k+1}}{1-y_{k+1}}\right)^{\sum\limits_{i=1}^{k+1} p_i}$$

Нужное неравенство установлено.

Докажем теперь обобщенное неравенство Ки Фана (**). Если $a_1 = \ldots = a_n$, то, очевидно, $\frac{G_n}{G'_n} = \frac{A_n}{A'_n}$. Предположим сейчас, что среди чисел $a_1, \ldots, a_n$ есть неравные друг другу и что эти числа перенумерованы в порядке неубывания, т.е. $a_1 \le a_2 \le \le \ldots \le a_n$ $(a_1 < a_n)$. Пусть m — наименьший из номеров k $(1 \le k \le n)$ таких, что $a_k < a_{k+1}$. Тогда

$$\frac{\tilde{G}_n}{\tilde{G}'_n} = \prod_{i=1}^{n} \left(\frac{a_i}{1-a_i}\right)^{\frac{p_i}{P_n}} = \left(\frac{y_m}{1-y_m}\right)^{\frac{1}{P_n}\sum\limits_{i=1}^{m} p_i} \left(\frac{a_{m+1}}{1-a_{m+1}}\right)^{\frac{p_{m+1}}{P_n}} \cdot \ldots \cdot \left(\frac{a_n}{1-a_n}\right)^{\frac{p_n}{P_n}}, \tag{12}$$

где $P_n = p_1 + \ldots + p_n$, а y_m определяется условием (11). Оценим выражение $\left(\frac{\tilde{G}_n}{\tilde{G}'_n}\right)^{P_n}$ сверху. Имеем:

$$\left(\frac{\tilde{G}_n}{\tilde{G}'_n}\right)^{P_n} < \left(\frac{y_{m+1}}{1-y_{m+1}}\right)^{\sum\limits_{i=1}^{m+1} p_i} \left(\frac{a_{m+2}}{1-a_{m+2}}\right)^{p_{m+2}} \cdot \ldots \cdot \left(\frac{a_n}{1-a_n}\right)^{p_n} \tag{13}$$

(мы в представлении (12) для оценки произведения $\left(\frac{y_m}{1-y_m}\right)^{\sum\limits_{i=1}^{m} p_i}\left(\frac{a_{m+1}}{1-a_{m+1}}\right)^{p_{m+1}}$ воспользовались предложением 3). Заметим, что, в силу неравенства $a_m < a_{m+1}$, будут выполнены условия $y_{m+1} < a_{m+2}$, $y_{m+2} < a_{m+3}$, $\ldots$, $y_{n-!} < a_n$. Следовательно, основываясь на предложении 3, оценку (13) можно продолжить:

$$\left(\frac{\tilde{G}_n}{\tilde{G}'_n}\right)^{P_n} < \left(\frac{y_{m+2}}{1-y_{m+2}}\right)^{\sum\limits_{i=1}^{m+2} p_i}\left(\frac{a_{m+3}}{1-a_{m+3}}\right)^{p_{m+3}}\ldots\left(\frac{a_n}{1-a_n}\right)^{p_n} < \ldots$$

$$\ldots < \left(\frac{y_{n-1}}{1-y_{n-1}}\right)^{\sum\limits_{i=1}^{n-1} p_i}\left(\frac{a_n}{1-a_n}\right)^{p_n} < \left(\frac{y_n}{1-y_n}\right)^{\sum\limits_{i=1}^{n} p_i} = \left(\frac{\tilde{A}_n}{\tilde{A}'_n}\right)^{P_n}.$$

Отсюда следует неравенство $(**)$.

Замечание 2. Рассмотренное доказательство обобщенного неравенства Ки Фана навеяно подходом к его установлению в работе [6].

2. Доказательство II. Рассмотрим еще одно новое доказательство обобщенного неравенства Ки Фана, опирающееся на свойства определенного интеграла Римана. Отметим, что оно в идейном отношении будет восходить к методам работ [7]–[8], посвященных доказательству классического неравенства Коши между средним арифметическим и средним геометрическим положительных чисел.

Прежде сформулируем одну простую лемму.

Лемма А. *Пусть* a, b $(a < b)$ — *положительные числа, не превосходящие* $\frac{1}{2}$. *Тогда функции* $f(x) = \frac{1}{x} + \frac{1}{1-x} - \frac{1}{b} - \frac{1}{1-b}$, $g(x) = \frac{1}{a} + \frac{1}{1-a} - \frac{1}{x} - \frac{1}{1-x}$ *являются неотрицательными на отрезке* $[a; b]$.

Справедливость утверждения леммы следует из того, что

$$f(b) = 0, \qquad f'(x) = -\frac{1}{x^2} + \frac{1}{(1-x)^2} = \frac{x^2-(1-x)^2}{x^2(1-x)^2} < 0 \quad \text{при} \quad x \in (a; b);$$
$$g(a) = 0, \qquad g'(x) = -f'(x) > 0 \quad \text{при} \quad x \in (a; b).$$

Перейдем теперь к доказательству неравенства $(***)$. Предположим, что числа $a_1, \ldots, a_n$ перенумерованы в порядке неубывания и пусть k $(1 \le k \le n-1)$ — такой номер, что $a_k \le \tilde{A}_n \le a_{k+1}$. Введем в рассмотрение величину

$$A = \sum_{i=1}^{k} p_i \int\limits_{a_i}^{\tilde{A}_n}\left(\frac{1}{x} + \frac{1}{1-x} - \frac{1}{\tilde{A}_n} - \frac{1}{1-\tilde{A}_n}\right)dx +$$
$$+ \sum_{i=k+1}^{n} p_i \int_{\tilde{A}_n}^{a_i}\left(\frac{1}{\tilde{A}_n} + \frac{1}{1-\tilde{A}_n} - \frac{1}{x} - \frac{1}{1-x}\right)dx.$$

Заметим, в силу леммы, слагаемые сумм, задающих A, неотрицательны, следовательно, $A \ge 0$. Очевидно, равенство $A = 0$ будет иметь место лишь тогда, когда $a_i = \tilde{A}_n$, $i = 1, \ldots, n$, т.е. лишь при совпадении чисел $a_1, \ldots, a_n$. Вычислим A, применяя формулу Ньютона-Лейбница:

$$A = \sum_{i=1}^{k} p_i \left(\ln x - \frac{1}{\tilde{A}_n} x - \ln(1-x) - \frac{1}{1-\tilde{A}_n} x \right) \Bigg|_{a_i}^{\tilde{A}_n} +$$

$$+ \sum_{i=k+1}^{n} p_i \left(\frac{1}{\tilde{A}_n} x - \ln x + \frac{1}{1-\tilde{A}_n} x + \ln(1-x) \right) \Bigg|_{\tilde{A}_n}^{a_i} =$$

$$= \sum_{i=1}^{k} p_i \ln \frac{\tilde{A}_n}{a_i} - \sum_{i=1}^{k} p_i \ln \frac{1-\tilde{A}_n}{1-a_i} - \sum_{i=1}^{k} p_i \left(\frac{1}{\tilde{A}_n} + \frac{1}{1-\tilde{A}_n} \right) \left(\tilde{A}_n - a_i \right) +$$

$$+ \sum_{i=k+1}^{n} p_i \left(\frac{1}{\tilde{A}_n} + \frac{1}{1-\tilde{A}_n} \right) \left(a_i - \tilde{A}_n \right) - \sum_{i=k+1}^{n} p_i \ln \frac{a_i}{\tilde{A}_n} + \sum_{i=k+1}^{n} p_i \ln \frac{1-a_i}{1-\tilde{A}_n} =$$

$$= \sum_{i=1}^{n} p_i \ln \frac{\tilde{A}_n}{a_i} - \sum_{i=1}^{n} p_i \ln \frac{1-\tilde{A}_n}{1-a_i} - \sum_{i=1}^{n} p_i \left(\frac{1}{\tilde{A}_n} + \frac{1}{1-\tilde{A}_n} \right) \left(\tilde{A}_n - a_i \right) =$$

$$= \ln \prod_{i=1}^{n} \frac{\tilde{A}_n^{p_i}}{a_i^{p_i}} - \ln \prod_{i=1}^{n} \frac{\left(\tilde{A}'_n\right)^{p_i}}{(1-a_i)^{p_i}} = \ln \frac{\tilde{A}_n}{\tilde{G}_n} - \ln \frac{\tilde{A}'_n}{\tilde{G}'_n}.$$

Отсюда, в силу неотрицательности A, имеем неравенство $\ln \frac{\tilde{A}_n}{\tilde{G}_n} - \ln \frac{\tilde{A}'_n}{\tilde{G}'_n} \geq 0$, эквивалентное неравенству $(***)$. Поскольку равенство $A = 0$ возможно лишь при условии $a_1 = \ldots = a_n$, то только при этом условии будет достигаться равенство в $(***)$. Неравенство $(***)$ полностью обосновано.

§5. О доказательствах неравенства Ки Фана посредством неравенства Иенсена

В данном параграфе мы рассмотрим два доказательства неравенства Ки Фана $(***)$, основывающиеся на применении неравенства Иенсена к некоторым подходяще выбранным выпуклым функциям.

1. Доказателство III. Приведем доказательство неравенства Ки Фана $(***)$, развивающее подход к доказательству неравенства $(*)$ из недавней работы [9].

Введем в рассмотрение функцию $y = \frac{e^x - K}{1+e^x}$ $(K > 0)$. Поскольку

$$y'' = (1+K)(1-e^x)\frac{e^x}{(1+e^x)^3},$$

то, очевидно, эта функция будет выпуклой на интервале $(-\infty; 0)$. Следовательно, для нее выполняется неравенство Иенсена

$$y\left(\sum_{i=1}^{n} \lambda_i x_i\right) \leq \sum_{i=1}^{n} \lambda_i y\,(x_i), \tag{14}$$

где $x_i \in (-\infty; 0)$, $\lambda_i > 0$; $i = 1, \ldots, n$; $\lambda_1 + \cdots + \lambda_n = 1$. Перепишем (14) более подробно:

$$\frac{\prod\limits_{i=1}^{n} e^{\lambda_i x_i} - K}{1 + \prod\limits_{i=1}^{n} e^{\lambda_i x_i}} \le \sum_{i=1}^{n} \lambda_i \frac{e^{x_i} - K}{1 + e^{x_i}}. \tag{15}$$

Введем теперь в рассмотрение числа $b_i = \frac{a_i}{1-a_i}$, $\bar{x}_i = \ln b_i$, $i = 1, \ldots, n$. Нетрудно видеть, что $b_i \in (0; 1]$, $\bar{x}_i \in (-\infty; 0)$. Полагая в (15) $K = b_1^{p_1} \cdot \ldots \cdot b_n^{p_n}$, $\lambda_i = p_i$, $x_i = \bar{x}_i$, $i = 1, \ldots, n$, получим неравенство

$$0 \le \sum_i p_i \frac{b_i - b_1^{p_1} \cdot \ldots \cdot b_n^{p_n}}{1 + b_i},$$

откуда следует, что

$$b_1^{p_i} \cdot \ldots \cdot b_n^{p_n} \le \frac{\sum\limits_{i=1}^{n} p_i \frac{b_i}{1+b_i}}{\sum\limits_{i=1}^{n} p_i \frac{1}{1+b_i}}. \tag{16}$$

Так как $\frac{b_i}{1+b_i} = a_i$, $\frac{1}{1+b_i} = 1 - a_i$, то из (16) следует неравенство Ки Фана $(***)$.

Поскольку функция $y = \frac{e^x - K}{1+e^x}$ не является линейной, то знак равенства в (15) достигается лишь тогда, когда $x_1 = \ldots = x_n$. Следовательно, в (16) он достигается тогда и только тогда, когда $b_1 = \ldots = b_n$, что равносильно условию $a_1 = \ldots = a_n$. Неравенство Ки Фана $(***)$ полностью обосновано.

1. Доказателство IV. Приведем еще одно доказательство неравенства Ки Фана, основанное на подходе, аналогичном рассмотренному. Введем функцию $y = \ln \frac{1-x}{x}$. На промежутке $\left(0; \frac{1}{2}\right]$ она является выпуклой, поскольку $y'' = \frac{1-2x}{x^2(1-x)^2} > 0$ при $x \in \left(0; \frac{1}{2}\right)$. Применим к ней неравенство Иенсена (14), где $x_i \in \left(0; \frac{1}{2}\right]$. Полагая в (14) $\lambda_i = p_i$, $x_i = a_i$, $i = 1, \ldots, n$, будем иметь неравенство

$$\ln \frac{1 - (p_1 a_1 + \ldots + p_n a_n)}{p_1 a_1 + \ldots + p_n a_n} \le p_1 \ln \frac{1 - a_1}{a_1} + \ldots + p_n \ln \frac{1 - a_n}{a_n},$$

которое равносильно неравенству

$$\ln \frac{\tilde{A}'_n}{\tilde{A}_n} \le \ln \frac{\tilde{G}'_n}{\tilde{G}_n}. \tag{17}$$

Отсюда следует неравенство Ки Фана.

Обоснуем условия достижения равенства в неравенстве Ки Фана. Так как функция $y = \ln \frac{1-x}{x}$ не является линейной, то равенство в (17) будет иметь место лишь при условии $a_1 = \ldots = a_n$. Значит, и в неравенстве Ки Фана $(***)$ равенство будет достигаться лишь тогда, когда $a_1 = \ldots = a_n$. Неравенство $(***)$ установлено.

§6. Обобщения неравенства Ки Фана

1. Аддитивный аналог неравенства Ки Фана и его обобщение. В теории средних степенных величин аддитивным аналогом неравенства Ки Фана (*) принято называть неравенство

$$A_n' - A_n \le G_n' - G_n, \tag{18}$$

в котором равенство достигается тогда и только тогда, когда $a_1 = \ldots = a_n$. Это неравенство впервые было установлено немецким математиком Х. Альцером в 1988 году в работе [10]. Позднее, в 1995 году в работе [11] он обобщил неравенство (18) на «весовой» случай, т.е. установил неравенство

$$\tilde{A}'_n - \tilde{A}_n \le \tilde{G}'_n - \tilde{G}_n, \tag{19}$$

в котором равенство также достигается лишь при условии $a_1 = \ldots = a_n$. Заметим, неравенства (18) и (19) являются более сильными неравенствами, чем соответственно неравенства (*) и (**). Действительно, переписав неравенство (18) в виде $A_n'\left(1 - \frac{A_n}{A_n'}\right) \le G_n'\left(1 - \frac{G_n}{G_n'}\right)$, его левую часть оценим снизу по неравенству Коши значением $G_n'\left(1 - \frac{A_n}{A_n'}\right)$. Это позволяет записать соотношение

$$1 - \frac{A_n}{A_n'} \le 1 - \frac{G_n}{G_n'},$$

равносильное, очевидно, неравенству Ки Фана (*). Аналогично получается и неравенство (**) из неравенства (19). Таким образом, аддитивный аналог неравенства Ки Фана влечет само неравенство Ки Фана.

В 1996 году в статье [12] было предложено весьма простое доказательство неравенства (18), основанное на соотношении

$$\prod_{k=1}^{n} (1 + b_k)^{\frac{1}{n}} - \prod_{k=1}^{n} (1 - b_k)^{\frac{1}{n}} \ge \frac{2}{n} \sum_{k=1}^{n} b_k, \tag{20}$$

где $b_k \in [0; 1)$, $k = 1, \ldots, n$. Ниже мы методом, отличным от подхода автора последней цитируемой работы, установим обобщение неравенства (20), которое позволяет просто доказать неравенство

$$\hat{\tilde{F}}(1) - \tilde{F}(1) \le \hat{\tilde{F}}(\alpha) - \tilde{F}(\alpha) \quad (0 < \alpha < 1), \tag{21}$$

развивающее неравенство (19). В неравенстве (21) символ $\hat{\tilde{F}}$ обозначает взвешенное среднее степенное чисел $1 - a_1, \ldots, 1 - a_n$ с тем же набором весов, что участвует в записи взвешенного среднего степенного $\tilde{F}$ чисел $a_1, \ldots, a_n$. Отметим, что равенство в (21) достигается лишь при условии $a_1 = \ldots = a_n$.

2. Вспомогательная лемма. Справедлива следующая

Лемма Б. *Пусть* $b_1, \ldots, b_n$ *— положительные числа из промежутка* $[0,1)$, $p_1, \ldots, p_n$ *— положительные числа, удовлетворяющие условию* $p_1 + \ldots + p_n = 1$. *Справедливо неравенство*

$$\left(\sum_{k=1}^{n} p_k \left(1+b_k\right)^{\alpha}\right)^{1/\alpha} - \left(\sum_{k=1}^{n} p_k \left(1-b_k\right)^{\alpha}\right)^{1/\alpha} \geq 2\sum_{k=1}^{n} p_k b_k, \quad 0<\alpha<1, \tag{22}$$

в котором равенство достигается лишь тогда, когда $b_1 = \ldots = b_n$.

Доказательство. Будем считать, что числа $b_1, \ldots, b_n$ перенумерованы в порядке неубывания. Введем в рассмотрение последовательность величин $\{f_k(x)\}$, $k = 1, \ldots, n-1$, полагая

$$\begin{aligned} f_k(x) = {} & \left(p_1(1+b_1)^{\alpha} + \ldots + p_k(1+b_k)^{\alpha} + (p_{k+1} + \ldots + p_n)(1+b_k+x)^{\alpha}\right)^{1/\alpha} - \\ & - \left(p_1(1-b_1)^{\alpha} + \ldots + p_k(1-b_k)^{\alpha} + (p_{k+1} + \ldots + p_n)(1-b_k-x)^{\alpha}\right)^{1/\alpha} - \\ & - 2\left(p_1 b_1 + \ldots + p_k b_k + (p_{k+1} + \ldots + p_n)(b_k + x)\right), \quad x \in [0, b_{k+1} - b_k]. \end{aligned}$$

Заметим, что если $b_k = b_{k+1}$, то $f_k(0) = f_{k+1}(0)$ при $k < n-1$ и

$$f_{n-1}(b_n - b_{n-1}) = \left(\sum_{i=1}^{n} p_i \left(1+b_i\right)^{\alpha}\right)^{1/\alpha} - \left(\sum_{i=1}^{n} p_i \left(1-b_i\right)^{\alpha}\right)^{1/\alpha} - 2\sum_{i=1}^{n} p_i b_i \stackrel{def}{=} B_{\alpha}.$$

Покажем, что если $b_k < b_{k+1}$ $(1 \leq k \leq n-1)$, то на отрезке $[0, b_{k+1} - b_k]$ функция $f_k(x$ будет строго возрастающей. Для доказательства последнего воспользуемся достаточным условием монотонности функции в терминах производной. Найдем $f_k'(x)$:

$$\begin{aligned} f_k'(x) = {} & \alpha^{-1}\Big(p_1(1+b_1)^{\alpha} + \ldots + p_k(1+b_k)^{\alpha} + (p_{k+1} + \ldots + p_n)(1+b_k+x)^{\alpha}\Big)^{1/\alpha - 1} \times \\ & \times \Big(p_{k+1} + \ldots + p_n\Big)\alpha\Big(1+b_k+x\Big)^{\alpha-1} + \\ & + \alpha^{-1}\Big(p_1(1-b_1)^{\alpha} + \ldots + p_k(1-b_k)^{\alpha} + (p_{k+1} + \ldots + p_n)(1-b_k-x)^{\alpha}\Big)^{1/\alpha - 1} \times \\ & \times \Big(p_{k+1} + \ldots + p_n\Big)\alpha\Big(1-b_k-x\Big)^{\alpha-1} - 2\Big(p_{k+1} + \ldots + p_n\Big) = \\ & = (p_{k+1} + \ldots + p_n) \times \\ & \times \left[\left(\frac{(1+b_k+x)^{\alpha}}{p_1(1+b_1)^{\alpha} + \ldots + p_{k+1}(1+b_k)^{\alpha} + (p_{k+1} + \ldots + p_n)(1+b_k+x)^{\alpha}}\right)^{\alpha - 1/\alpha} + \right. \\ & \left. + \left(\frac{(1-b_k-x)^{\alpha}}{p_1(1-b_1)^{\alpha} + \ldots + p_k(1-b_k)^{\alpha} + (p_{k+1} + \ldots + p_n)(1-b_k-x)^{\alpha}}\right)^{\alpha - 1/\alpha} - 2\right] = \end{aligned}$$

$$= (p_{k+1} + \ldots + p_n) \times$$
$$\times \left[\left(p_1 \left(\frac{1+b_1}{1+b_k+x} \right)^\alpha + \ldots + p_k \left(\frac{1+b_k}{1+b_k+x} \right)^\alpha + (p_{k+1} + \ldots + p_n) \right)^{1-\alpha/_\alpha} + \right.$$
$$\left. + \left(p_1 \left(\frac{1-b_1}{1-b_k-x} \right)^\alpha + \ldots + p_k \left(\frac{1-b_k}{1-b_k-x} \right)^\alpha + (p_{k+1} + \ldots + p_n) \right)^{1-\alpha/_\alpha} - 2 \right].$$

Основания степеней, стоящих в квадратных скобках, оценим снизу, используя неравенство Коши для взвешенных среднего арифметического и среднего геометрического $k+1$ чисел $\left(1+b_1/_1+b_k+x\right)^\alpha, \ldots, \left(1+b_k/_1+b_k+x\right)^\alpha, 1$ с весами $p_1, \ldots, p_k, p_{k+1} + \ldots + p_n$ и аналогично чисел $\left(1-b_1/_1-b_k-x\right)^\alpha, \ldots, \left(\frac{1-b_k}{1-b_k-x}\right)^\alpha, 1$ с такими же весами. Тогда для $f_k'(x)$ будем иметь следующую оценку:

$$f_k'(x) \geq (p_{k+1} + \ldots + p_n) \left[\left(\left(\frac{1+b_1}{1+b_k+x} \right)^{\alpha p_1} \cdot \ldots \cdot \left(\frac{1+b_k}{1+b_k+x} \right)^{\alpha p_k} \right)^{1-\alpha/_\alpha} + \right.$$
$$\left. + \left(\left(\frac{1-b_1}{1-b_k-x} \right)^{\alpha p_1} \cdot \ldots \cdot \left(\frac{1-b_k}{1-b_k-x} \right)^{\alpha p_k} \right)^{1-\alpha/_\alpha} - 2 \right].$$

Оценку снизу для $f_k'(x)$ продолжим, применяя неравенство Коши для первых двух слагаемых в последних квадратных скобках:

$$f_k'(x) \geq (p_{k+1} + \ldots + p_n) \left[2\sqrt{\left(\frac{1-b_1^2}{1-(b_k+x)^2} \right)^{p_1(1-\alpha)} \cdot \ldots \cdot \left(\frac{1-b_k^2}{1-(b_k+x)^2} \right)^{p_k(1-\alpha)}} - 2 \right] > 0.$$

Производная $f_k'(x)$ функции $f_k(x)$ на интервале $(0, b_{k+1} - b_k)$ положительна, следовательно, функция $f_k(x)$ строго возрастает на отрезке $[0, b_{k+1} - b_k]$. Таким образом, можно сделать вывод, что если $b_k < b_{k+1}$, то $f_k(0) < f_k(b_{k+1} - b_k) = f_{k+1}(0)$ при $k < n-1$ и $f_k(0) < f_k(b_k - b_{k-1}) = B_\alpha$ при $k = n-1$.

Проведенные рассуждения позволяют записать цепочку неравенств

$$0 = f_1(0) \leq f_2(0) \leq \ldots \leq f_{n-1}(0) \leq B_\alpha, \tag{23}$$

при этом всюду в (23) знак нестрогого неравенства можно заменить знаком равенства лишь в том случае, когда $b_1 = \ldots = b_n$. Если среди чисел $b_1, \ldots, b_n$ будут различные, то обязательно в (23) будут иметь место и строгие неравенства. Таким образом, неравенство $B_\alpha \geq 0$ обращается в равенство лишь при условии $b_1 = \ldots = b_n$. Так как неравенство $B_\alpha \geq 0$ равносильно неравенству (22), то лемма доказана.

Замечание. Пусть числа $b_1, \ldots, b_n$, участвующие в записи неравенства (22), перенумерованы в порядке неубывания. На основании техники доказательства этого неравенства можем заключить, что левая часть (22) тем больше правой части этого неравенства, чем больше разности $b_{k+1} - b_k$, $k = 1, \ldots, n-1$.

3. Доказательство неравенства (21). Для доказательства неравенства (21) положим в (22) $b_k = 1 - 2a_k, \quad k = 1, \ldots, n$. Так как $a_k \in \left(0; {}^1\!/_2\right]$, то, очевидно, $b_k \in [0, 1)$, и для таких чисел неравенство (22) будет иметь место. Можем записать:

$$\left(\sum_{k=1}^{n} p_k\,(2 - 2a_k)^{\alpha}\right)^{1/\alpha} - \left(\sum_{k=1}^{n} p_k\,(2a_k)^{\alpha}\right)^{1/\alpha} \geq 2\sum_{k=1}^{n} p_k\,(1 - 2a_k),$$

или

$$\left(\sum_{k=1}^{n} p_k\,(1 - a_k)^{\alpha}\right)^{1/\alpha} - \left(\sum_{k=1}^{n} p_k a_k^{\alpha}\right)^{1/\alpha} \geq \sum_{k=1}^{n} p_k\,(1 - a_k) - \sum_{k=1}^{n} p_k a_k,$$

а последнее неравенство есть неравенство (21) в случае, когда $0 < \alpha < 1$. Поскольку в (22) равенство достигается лишь при условии $b_1 = \ldots = b_n$, то в (21) оно будет иметь место лишь при условии $a_1 = \ldots = a_n$. Неравенство (21) полностью обосновано.

Замечание. Сделанное в предыдущем пункте замечание позволяет заключить: левая часть неравенства (21) тем больше отличается от его правой части, чем больше расстояния между соседними точками совокупности чисел $a_1, \ldots, a_n$ на числовой прямой.

4. Неравенство Ки Фана для средних степенных порядков $\alpha\,(0 \leq \alpha < 1)$ **и** 1. Простое неравенство Ки Фана $(*)$ или его обобщенный вариант $(**)$ — это неравенства для арифметико-геометрических средних наборов $a_1, \ldots, a_n$ и $1 - a_1, \ldots, 1 - a_n$ положительных чисел, то есть для средних порядков 0 и 1. Подобно тому, как устанавливался вывод неравенства Ки Фана $(*)$ из его аддитивного аналога (18), нетрудно показать, что неравенство (21) влечет неравенство

$$\frac{\tilde{F}\,(\alpha)}{\hat{\tilde{F}}\,(\alpha)} \leq \frac{\tilde{A}_n}{\tilde{A'}_n}, \quad 0 < \alpha < 1, \tag{24}$$

в котором равенство достигается только при условии $a_1 = \cdots = a_n$. Приведем упоминаемую процедуру вывода (24) из (21).

Неравенство (21) равносильно неравенству

$$\tilde{A'}_n \left(1 - \frac{\tilde{A}_n}{\tilde{A'}_n}\right) \leq \hat{\tilde{F}}(\alpha) \left(1 - \frac{\tilde{F}(\alpha)}{\hat{\tilde{F}}(\alpha)}\right). \tag{25}$$

Так как

$$\tilde{A'}_n \geq \hat{\tilde{F}}(\alpha) \qquad (0 \leq \alpha < 1), \tag{26}$$

то неравенство (25) влечет неравенство

$$1 - \frac{\tilde{A}_n}{\tilde{A'}_n} \leq 1 - \frac{\tilde{F}\,(\alpha)}{\hat{\tilde{F}}\,(\alpha)}, \tag{27}$$

эквивалентное (24). Поскольку равенство в неравенствах (24)–(25) и (26) достигается лишь при условии $a_1 = \ldots = a_n$, то только при этом условии оно будет иметь место и в неравенстве (27). Неравенство (24) полностью обосновано.

Ясно, что предельный переход при $\alpha \to 0$ в установленном неравенстве (24) порождает обобщенное неравенство Ки Фана. Учитывая последнее, неравенство (24) условимся называть неравенством Ки Фана для средних степенных порядков α и 1.

5. Неравенство Ки Фана для средних степенных порядков 1 **и** β $(0<\beta\leq 2)$. Целью настоящего пункта является установление неравенства типа неравенства (24). Мы покажем, что для рассматриваемых наборов положительных чисел $a_1, \ldots, a_n$ и $1 - a_1, \ldots, 1 - a_n$ будет иметь место неравенство

$$\frac{\tilde{A}_n}{\tilde{A}'_n} \leq \frac{\tilde{F}(\beta)}{\hat{\tilde{F}}(\beta)}, \qquad 1 < \beta \leq 2, \tag{28}$$

в котором равенство достигается тогда и только тогда, когда $a_1 = \ldots = a_n$. Это неравенство будем называть неравенством Ки Фана для средних степенных порядков 1 и β.

Введем в рассмотрение функцию

$$f(x) = \frac{1}{x^t} - \frac{1}{(1-x)^t} \qquad \left(x \in \left(0; \frac{1}{2}\right]\right),$$

где параметр t подчиняется условию $t \in \mathbb{R} \setminus \{-2; -1; 0\}$. Анализируя знак значений ее второй производной

$$f''(x) = t(t+1)\frac{(1-x)^{t+2} - x^{t+2}}{(x(1-x))^{t+2}},$$

заключаем, что она является выпуклой на промежутке $\left(0; 1/2\right]$ при $t \in (-2; -1)$. Применим к этой функции неравенство Иенсена (14), полагая $x_k = a_k$ и беря $\lambda_k = p_k$, будем иметь:

$$\frac{1}{\left(\sum\limits_{k=1}^{n} p_k a_k\right)^t} - \frac{1}{\left(1 - \sum\limits_{k=1}^{n} p_k a_k\right)^t} \leq \sum_{k=1}^{n} p_k \left(\frac{1}{a_k^t} - \frac{1}{(1-a_k)^t}\right), \qquad t \in (-2; -1).$$

Получившееся неравенство перепишем в более компактном виде, используя обозначения соответствующих взвешенных средних степенных:

$$\frac{1}{\tilde{A}_n^t} - \frac{1}{\left(\tilde{A}'_n\right)^t} \leq \frac{1}{\tilde{F}^t(-t)} - \frac{1}{\hat{\tilde{F}}^t(-t)}, \qquad t \in (-2; -1). \tag{29}$$

Полагая в неравенстве (29) $\beta = -t$, будем иметь неравенство

$$\hat{\tilde{F}}^\beta(\beta) - \tilde{F}^\beta(\beta) \leq \left(\tilde{A}'_n\right)^\beta - \tilde{A}_n^\beta, \qquad 1 < \beta < 2. \tag{30}$$

Кроме того, в силу монотонности функций $\tilde{F}(x)$ и $\hat{\tilde{F}}(x)$, для $1 < \beta < 2$ справедливо неравенство

$$\left(\tilde{A'}_n\right)^\beta + \tilde{A}_n^\beta \le \hat{\tilde{F}}^\beta(\beta) + \tilde{F}^\beta(\beta). \tag{31}$$

Так как левая часть (30) неотрицательна, то, перемножая (30) и (31), получаем неравенство

$$\hat{\tilde{F}}^\beta(\beta)\,\tilde{A}_n^\beta - \tilde{F}^\beta(\beta)\left(\tilde{A'}_n\right)^\beta \le \tilde{F}^\beta(\beta)\left(\tilde{A'}_n\right)^\beta - \hat{\tilde{F}}^\beta(\beta)\,\tilde{A}_n^\beta,$$

или

$$2\hat{\tilde{F}}^\beta(\beta)\,\tilde{A}_n^\beta \le 2\left(\tilde{A'}_n\right)^\beta \tilde{F}^\beta(\beta),$$

откуда следует

$$\frac{\tilde{A}_n}{\tilde{A'}_n} \le \frac{\tilde{F}(\beta)}{\hat{\tilde{F}}(\beta)}, \qquad 1 < \beta < 2. \tag{32}$$

Таким образом, неравенство (28) для $\beta \in (1;2)$ доказано.

Заметим, функция $f(x)$ при $t \in (-2;-1)$ не является линейной, поэтому равенство в (32) будет иметь место тогда и только тогда, когда $a_1 = \ldots = a_n$.

Перейдем в неравенстве (32) к пределу при $\beta \to 2-0$. Предельный переход дает неравенство

$$\frac{\tilde{A}_n}{\tilde{A'}_n} \le \frac{\tilde{R}_n}{\tilde{R'}_n}, \tag{33}$$

связывающее арифметико-квадратичные средние $\tilde{A}_n$, $\tilde{R}_n = \tilde{F}(2)$, $\tilde{A'}_n$, $\tilde{R'}_n = \hat{\tilde{F}}(2)$ чисел a_k и $1 - a_k$ $(k = 1,\ldots,n)$. Покажем, что в (33) равенство будет иметь место тогда и только тогда, когда $a_1 = \ldots = a_n$. Действительно, при $a_1 = \ldots = a_n$ выполняются соотношения

$$\tilde{A}_n = \tilde{R}_n = a_1, \tilde{A'}_n = \tilde{R'}_n = 1 - a_1,$$

значит, при таком условии (33) есть равенство. Покажем, что при невыполнении условия $a_1 = \ldots = a_n$ равенство в (33) не может иметь места. Действительно, в таком случае $\tilde{R}_n < \tilde{R'}_n$. Подберем число r $(0 < r < 1)$ такое, чтобы выполнялось неравенство $\tilde{R}_n^2 < r\left(\tilde{R'}_n\right)^2$, и введем в рассмотрение функцию $g(x) = x^2 - r(1-x)^2$. Она, очевидно, не является линейной и выпукла на промежутке $\left(0;\frac{1}{2}\right]$. Применив к ней неравенство Иенсена (14), получим неравенство

$$\left(\sum_{k=1}^n p_k a_k\right)^2 - r\left(1 - \sum_{k=1}^n p_k a_k\right)^2 < \sum_{k=1}^n p_k\left(a_k^2 - r(1-a_k)^2\right),$$

которое можно записать в виде

$$\tilde{A}_n^2 - r\left(\tilde{A'}_n\right)^2 < \tilde{R}_n^2 - r\left(\tilde{R'}_n\right)^2,$$

или

$$r\left(\tilde{R'}_n\right)^2 - \tilde{R}_n^2 < r\left(\tilde{A'}_n\right)^2 - \tilde{A}_n^2. \tag{34}$$

В соответствии с выбором числа r левая часть (34) положительна. Из неравенства (34) и очевидного неравенства

$$r\left(\tilde{A'}_n\right)^2 + \tilde{A}_n^2 < r\left(\tilde{R'}_n\right)^2 + \tilde{R}_n^2$$

посредством их перемножения стандартным образом получаем неравенство

$$\tilde{A}_n^2\left(\tilde{R'}_n\right)^2 < \tilde{R}_n^2\left(\tilde{A'}_n\right)^2,$$

откуда следует

$$\frac{\tilde{A}_n}{\tilde{A'}_n} < \frac{\tilde{R}_n}{\tilde{R'}_n}.$$

Нужное показано. Тем самым можно констатировать, что неравенство (28) полностью обосновано.

Итак, объединяя результаты пунктов 4 и 5, можно записать следующее двойное неравенство, обобщающее неравенство Ки Фана (**),

$$\frac{\tilde{F}(\alpha)}{\hat{\tilde{F}}(\alpha)} \le \frac{\tilde{A}_n}{\tilde{A'}_n} \le \frac{\tilde{F}(\beta)}{\hat{\tilde{F}}(\beta)} \qquad (0 \le \alpha < 1, \quad 1 < \beta \le 2), \tag{35}$$

в котором знаки равенства могут иметь место лишь тогда, когда $a_1 = \ldots = a_n$.

Замечание. Неравенство (35) впервые было анонсировано в работе [13] от 1997 года. В связи с этим неравенством возникают вопросы его обобщения, а также уточнения. Обозначенные вопросы мы адресуем читателю.

Литература

[1] Смышляев В.К. Практикум по решению задач школьной математики. – М.: Просвещение, 1978.

[2] Beckenbach E.F., Bellman R. Inequalities. – Berlin: Springer, 1961.

[3] Беккенбах Э., Беллман Р. Неравенства. – М.: Мир, 1965.

[4] Alzer H. On weighted arithmetic, geometric and harmonic mean values // Glasnik matematicki. – 1990. –Vol. 25 (45). – 279 – 285.

[5] Alzer H., Ando T. and Nakamura Y. The inequalities of W. Sierpinski and Ky Fan // J. Math. Anal. Appl. – 1990. – 149. – P. 497 – 512.

[6] Pečarić J.E, Alzer H. On Ky Fan's inequality // Mathematica Pannonica. – 1994. – 6/1. – P. 85 – 93.

[7] Alzer H. A proof of the arithmetic mean — geometric mean inequality // Amer. Math. Mon. – 1996. – 103, 7. – P.585.

[8] Калинин С.И. О доказательствах неравенства Коши посредством интеграла // Математическое образование. – 1999. – № 1 (8). – С. 25 – 28.

[9] Mercer A. McD. A Short Proof of Ky Fan's Arithmetic-Geometric Inequality // J. Math. Anal. and Appl. – 1996. – 204, № 3. – P. 940 – 942.

[10] Alzer H. Ungleichungen für geometrische und arithmetische Mittelwerte // Proc. Kon. Nederl. Akad. Wetensch. – 1988. – 91. – S. 365 – 374.

[11] Alzer H. The inequality of Ky Fan and related results // Acta Appl. Math. – 1995. – 38. – P. 305 – 354.

[12] McGregor Malcolm T. On an inequality of Horst Alzer // Indagat. Mathem. 1996. – 7, № 2. – P. 161 – 164.

[13] Калинин С.И. Об одном обобщении неравенства Ки Фана // Междунар. конференция по комплексному анализу и смежным вопросам. Н.Новгород, 3 – 5 июня 1997 г.: Тез. докл. – Н.Новгород: ННГУ, 1997. – С. 29 – 30.

Неопределенности функций многих переменных (часть II)

В. В. Ивлев

В работе продолжено (см. №4(23), 2002 г.) исследование неопределенностей функций многих переменных. Предложена методика построения неопределенностей с заданным пределом в заданной точке.

В предыдущей работе [1] дано обобщение известного правила Лопиталя для функций двух переменных. Получены необходимые и достаточные условия существования двойных пределов для неопределенностей различных порядков. Напомним без доказательства суть этого обобщения.

Пусть функции $f(x,y)$ и $g(x,y)$:

1) определены в некоторой окрестности точки (x_0, y_0);

2) являются бесконечно малыми при $(x,y) \longrightarrow (x_0, y_0)$;

3) дифференцируемы в точке (x_0, y_0), причем частные производные $f'(x)$, $f'(y)$, $g'(x)$ и $g'(y)$в этой точке отличны от нуля и конечны.

Тогда для существования двойного предела

$$\lim_{\substack{x \longrightarrow x_0 \\ y \longrightarrow y_0}} \frac{f(x,y)}{g(x,y)} = k, \quad k \neq 0, \;\; k \neq \pm\infty \tag{1}$$

необходимо и достаточно, чтобы имело место

$$\frac{f'_x(x_0,y_0)}{g'_x(x_0,y_0)} = \frac{f'_y(x_0,y_0)}{g'_y(x_0,y_0)} = k \tag{2}$$

Следствия:

1. Двойной предел не существует, если точно одна частная производная равна нулю или две **разноименных** частных производных равны нулю ($f'_x = g'_y = 0$ или $f'_y = g'_x = 0$).

2. Двойной предел может существовать, если точно две **одноименных** частных производных равны нулю ($f'_x = g'_x = 0$ или $f'_y = g'_y = 0$). Необходимо привлечение частных производных высших порядков.

3. Двойной предел равен нулю, если $f'_x = f'_y = 0$; предел равен $\pm\infty$, если $g'_x = g'_y = 0$.

4. Если все частные производные функций f и g первого порядка равны нулю, а все их частные производные второго порядка отличны от нуля и конечны, то для существования двойного предела необходимо и достаточно, чтобы в точке (x_0, y_0) имело место

$$\frac{f''_{x^2}}{g''_{x^2}} = \frac{f''_{xy}}{g''_{xy}} = \frac{f''_{y^2}}{g''_{y^2}} = k, \quad k \neq 0, \;\; k \neq \pm\infty$$

Рассмотрим другие обобщения правила Лопиталя.

I. Неопределенности $\frac{0}{0}$ для функций n переменных.

Пусть функция $f(x) = f(x_1, \ldots x_n)$ определена в некоторой окрестности точки $x^{(0)} = (x_1^{(0)}, \ldots x_n^{(0)})$ и m раз дифференцируема в точке $x^{(0)}$.

Определение 1. Функция $f(x)$ называется **бесконечно малой m-го порядка** при $x \longrightarrow x^{(0)}$, если

1) $f(x^{(0)}) = 0$;
2) $\mathrm{d}^i f(x^{(0)}) = 0, \;\; i = \overline{1, m-1}$,
3) $\mathrm{d}^m f(x^{(0)}) \neq 0$;

Условия 2 и 3 означают, что все частные производные функции $f(x)$ в точке $x^{(0)}$ до порядка $m-1$ включительно равны нулю, но по крайней мере одна из частных производных m-го порядка отлична от нуля.

Определение 2. Неопределенностью m-го порядка малости вида $\frac{0}{0}$ называется отношение двух бесконечно малых функций $f(x)$ и $g(x)$ m-го порядка при $x \longrightarrow x^{(0)}$.

Теорема 1. Пусть функции $f(x)$ и $g(x)$

1) определены в некоторой окрестности точки $x^{(0)}$ и дифференцируемы в точке $x^{(0)}$;

2) являются бесконечно малыми первого порядка при $x \longrightarrow x^{(0)}$, причем все первые частные производные конечны и отличны от нуля.

Тогда для существования n-кратного предела

$$\lim_{x \longrightarrow x^{(0)}} \frac{f(x)}{g(x)} = k, \quad k \neq 0, \;\; k \neq \pm\infty \tag{3}$$

необходимо и достаточно, чтобы имело место в точке $x^{(0)}$

$$\frac{f'_{x^1}}{g'_{x^1}} = \frac{f'_{x_2}}{g'_{x_2}} = \ldots = \frac{f'_{x_n}}{g'_{x_n}} = k, \quad k \neq 0, \;\; k \neq \pm\infty \tag{4}$$

▼[1] **Необходимость**. Пусть n-кратный предел (3) существует и равен k. Так как функции $f(x)$ и $g(x)$ удовлетворяют условию 2 теоремы, то в формулах Тейлора для них ограничимся дифференциалами первого порядка

$$\lim_{x \longrightarrow x_{(0)}} \frac{f(x)}{g(x)} = \lim_{x \longrightarrow x_{(0)}} \frac{\mathrm{d}f(x^{(0)}) + o_1(\rho)}{\mathrm{d}g(x^{(0)}) + o_2(\rho)} = \lim_{\Delta x \longrightarrow 0} \frac{\sum\limits_{i=1}^{n} f'_{x_i}(x^{(0)})\Delta x_i + o_1(\rho)}{\sum\limits_{i=1}^{n} g'_x(x^{(0)})\Delta x_i + o_2(\rho)} \tag{5}$$

$$\Delta x = (\Delta x_1, \ldots \Delta x_n), \qquad \rho = \sqrt{\sum_{i=1}^{n} \Delta x_i^2}$$

[1]Начало доказательства теоремы или решение примера будем обозначать символом ▼; конец доказательства теоремы или решения примера — символом ▲.

$o_1(\rho)$ и $o_2(\rho)$ — бесконечно малые функции более высокого порядка малости, чем ρ.

При переходе к пределу при $\Delta x \longrightarrow 0$ функции $o_1(\rho)$ и $o_2(\rho)$ можно в дальнейшем опустить.

Предположим, что условия (4) выполнены для всех x_i, $i = \overline{2, n}$, но для x_1 имеет место

$$\frac{f'_{x^1}(x^{(0)})}{g'_{x^1}(x^{(0)})} = k_1, \quad \text{причем } k \neq k_1.$$

Тогда (5) примет вид

$$\lim_{x \longrightarrow x^{(0)}} \frac{f(x)}{g(x)} = \lim_{\Delta x \longrightarrow 0} \frac{k_1 g'_{x_1} \Delta x_1 + k \left[\sum\limits_{i=1}^{n} g'_{x_i} \Delta x_i\right]}{g'_{x_1} \Delta x_1 + \sum\limits_{i=2}^{n} g'_{x_i} \Delta x_i}. \tag{6}$$

Так как приращения Δx_i независимы, то непосредственно из (6) следует, что при $k_1 = k$ слагаемые с Δx_i числителя и знаменателя пропорциональны k и при любых его значениях предел равен k. Если же $k_1 \neq k$, то можно, варьируя Δx_i, получить любое значение предела в интервале (k_1, k_2). Итак, необходимость условий (4) доказана.

Достаточность условия (4) очевидна и подтверждается прямой подстановкой (4) в (6). ▲

Следствия:

1. Не существует n-кратный предел, если

– точно одна частная производная равна нулю;

– точно две разноименных частных производных равны нулю ($f'_{x_i}{}' = g'_{x_i} = 0$, $i \neq j$).

2. Может существовать n-кратный предел, если $f'_{x_i} = g'_{x_i} = 0$ для одной переменной, остальные отношения одноименных частных производных равны k. Таких нулевых равенств может быть несколько, но меньше, чем n. Необходимо привлечение производных высших порядков.

3. Если $\mathrm{d}f(x^{(0)}) = 0$ или $\mathrm{d}g(x^{(0)})$, то имеет место отношение бесконечно малых функций разных порядков и? следовательно, n-кратный предел равен либо нулю, либо $\pm\infty$.

4. Если одновременно $\mathrm{d}f(x^{(0)}) = \mathrm{d}g(x^{(0)}) = 0$, то необходимо привлекать дифференциалы высших порядков.

Теорема 2. Пусть функции $f(x)$ и $g(x)$:

1) определены в некоторой окрестности точки $x^{(0)}$ и дважды дифференцируемы в точке $x^{(0)}$;

2) являются бесконечно малыми второго порядка, причем все вторые частные производные функций $f(x)$ и $g(x)$ конечны и отличны от нуля. Тогда для существования n-кратного предела необходимо и достаточно, чтобы имело место

$$\frac{f''_{x_1 x_j}(x^{(0)})}{g''_{x_1 x_j}(x^{(0)})} = k, \quad k \neq 0, \ \ k \neq \pm\infty.$$

Доказательство проводится по той же схеме, что и в теореме 1. Имеют место аналогичные следствия, что и выше.

Пусть, наконец, функции $f(x)$ и $g(x)$ являются бесконечно малыми m-го порядка при $x \longrightarrow x^{(0)}$ и m раз дифференцируемы в точке $x^{(0)}$, причем все частные производные m-го порядка конечны и отличны от нуля. Тогда для существования n-кратного предела необходимо и достаточно, чтобы имело место

$$\frac{f^{(m)}_{x_1^{m_1}\ldots x_n^{m_n}}(x^{(0)})}{g^{(m)}_{x_1^{m_1}\ldots x_n^{m_n}}(x^{(0)})} = k, \quad k \neq 0, \;\; k \neq \pm\infty, \;\; m_1 + \ldots + m_n = m.$$

Пример 1. Пусть $f(x) = x_1 + 2x_2 - 3x_3$, $g(x) = x_1^2 + 2x_2^2 - 3x_3^2$; вычислить тройной предел в точке $x^{(0)} = (1,\,1,\,1)$

$$\lim_{x \longrightarrow x(0)} \frac{f(x)}{g(x)} = \lim_{\substack{x_1 \longrightarrow 1\\ x_2 \longrightarrow 1\\ x_3 \longrightarrow 1}} \frac{x_1 + 2x_2 - 3x_3}{x_1^2 + 2x_2^2 - 3x_3^2}$$

▼ В точке $x^{(0)}$ $f(x^{(0)}) = g(x^{(0)}) = 0$. Имеем неопределенность $\frac{0}{0}$. Вычислим первые частные производные функций $f(x)$ и $g(x)$ в точке $x^{(0)}$.

$$f'_{x_1} = 1,\; f'_{x_2} = 2,\; f'_{x_3} = -3; \;\; g'_{x_1} = 2,\; g'_{x_2} = 4,\; g'_{x_3} = -6$$

Строим отношения одноименных частных производных

$$\frac{f'_{x_1}}{g'_{x_1}} = \frac{1}{2},\; \frac{f'_{x_2}}{g'_{x_2}} = \frac{1}{2},\; \frac{f'_{x_3}}{g'_{x_3}} = \frac{1}{2}.$$

Итак, тройной предел существует и равен $1/2$. ▲

В частности, тот же предел, но в точке $x^{(0)} = (0,\,0,\,0)$ равен ∞, т.к. $f'_{x_i} \neq 0$, $i = \overline{1,\,3}$, а $g'_{x_i} = 0$, $i = \overline{1,\,3}$.

II. Неопределенности других видов

Рассмотрим неопределенность вида $\frac{\infty}{\infty}$ для отношения функций двух переменных. Пусть функция двух переменных $f(x,y)$ является **бесконечно большой** при $(x,y) \longrightarrow (x_0,y_0)$, тогда функция

$$f^*(x,y) = \frac{1}{f(x,y)} = f^{-1}(x,y)$$

является бесконечно малой при $(x,y) \longrightarrow (x_0,y_0)$.

Определение 3. Функция $f(x,y)$ является **бесконечно большой m-го порядка** при $(x,y) \longrightarrow (x_0,y_0)$, если функция $f^{-1}(x,y)$ является **бесконечно малой того же порядка** при $(x,y) \longrightarrow (x_0,y_0)$.

Например, функция $f(x,y) = (2xy - x_0y - xy_0)^{-1}$ — бесконечно большая первого порядка, так как $f^*(x,y) = (2xy - x_0y - y_0x)$ — бесконечно малая функция первого порядка при $(x,y) \longrightarrow (x_0,y_0)$. Действительно, при $x_0 \neq 0$, $y_0 \neq 0$, $f^*(x_0,y_0) = 0$, а $\mathrm{d}f^*(x_0,y_0) = y_0\mathrm{d}x + x_0\mathrm{d}y \neq 0$. Функция же $f(x,y) = [1-\cos(x+y)]^{-1}$ — бесконечно большая второго порядка при $(x,y) \longrightarrow (0,0)$, так как функция $1-\cos(x+y)$ — бесконечно малая второго порядка при $(x,y) \longrightarrow (0,0)$: $f^*(0,0) = 0$, $\mathrm{d}f^*(0,0) = 0$, $\mathrm{d}^2f^*(0,0) = \mathrm{d}x^2 + 2\mathrm{d}y\mathrm{d}x + \mathrm{d}y^2 \neq 0$.

Теорема 3. Пусть функции $f(x,y)$ и $g(x,y)$:

1) определены в некоторой окрестности точки (x_0,y_0) и дифференцируемы в точке (x_0,y_0);

2) являются бесконечно большими первого порядка при $(x,y) \longrightarrow (x_0,y_0)$, причем первые частные производные функций $f^*(x,y) = f^{-1}(x,y)$ и $g^*(x,y) = g^{-1}(x,y)$ конечны и отличны от нуля в точке (x_0,y_0).

Тогда для существования двойного предела

$$\lim_{\substack{x\longrightarrow x_0\\ y\longrightarrow y_0}} \frac{f(x,y)}{g(x,y)} = k, \quad k \neq 0, \ k \neq \pm\infty \tag{7}$$

необходимо и достаточно, чтобы имело место

$$\frac{f'_x(x_0,y_0)}{g'_x(x_0,y_0)} = \frac{f'_y(x_0,y_0)}{g'_y(x_0,y_0)} = k \tag{8}$$

или

$$\frac{(g^*(x_0,y_0))'_x}{(f^*(x_0,y_0))'_x} = \frac{(g^*(x_0,y_0))'_y}{(f^*(x_0,y_0))'_y} = k. \tag{8a}$$

Доказательство теоремы сводится к неопределенности $\frac{0}{0}$ введением функций $g^*(x,y) = g^{-1}(x,y)$, $f^*(x,y) = f^{-1}(x,y)$ и далее, в соответствии с [1]. Подчеркнем, что в зависимости от вида функций $f(x,y)$ и $g(x,y)$ используются условия либо (8), либо (8a), что проще.

В случае неопределенностей вида $0\cdot\infty$, т.е. при вычислении предела

$$\lim_{\substack{x\longrightarrow x_0\\ y\longrightarrow y_0}} f(x,y)\cdot g(x,y)$$

при $f(x_0,y_0) = 0$, $g(x_0,y_0) = \infty$, задача сводится к неопределенности $\frac{0}{0}$, полагая

$$\lim_{\substack{x\longrightarrow x_0\\ y\longrightarrow y_0}} f(x,y)\cdot g(x,y) = \lim_{\substack{x\longrightarrow x_0\\ y\longrightarrow y_0}} \frac{f(x,y)}{g^{-1}(x,y)}$$

Наконец, при вычислении пределов вида 1^∞, 0^0 используется предварительное логарифмирование искомого предела. Пусть, например, требуется найти двойной предел

$$\lim_{\substack{x\longrightarrow x_0\\ y\longrightarrow y_0}} f(x,y)^{g(x,y)}, \ f(x_0,y_0) = 1, \ g(x_0,y_0) = \infty.$$

Логарифмируя, получим

$$\ln\left[\lim_{\substack{x\to x_0\\ y\to y_0}} f(x,y)^{g(x,y)}\right] = \lim_{\substack{x\to x_0\\ y\to y_0}} \ln\left[f(x,y)^{g(x,y)}\right] = \lim_{\substack{x\to x_0\\ y\to y_0}} g(x,y)\ln f(x,y).$$

Используя эквивалентность бесконечно малых

$$\ln f(x,y) \sim f(x,y) - 1 \ \text{ при } \ (x,y) \longrightarrow (x_0,y_0),$$

запишем

$$\lim_{\substack{x\to x_0\\ y\to y_0}} g(x,y)\ln f(x,y) = \lim_{\substack{x\to x_0\\ y\to y_0}} g(x,y)[f(x,y)-1].$$

Задача сведена к случаю $\infty\cdot 0$. Окончательно,

$$\lim_{\substack{x\to x_0\\ y\to y_0}} f(x,y)^{g(x,y)} = e^{\lim\limits_{\substack{x\to x_0\\ y\to y_0}} g(x,y)[f(x,y)-1]} \tag{9}$$

Пример 2. Вычислить двойной предел

$$\lim_{\substack{x\to 1\\ y\to 2}} (2x+y+1)^{\frac{1}{x^2+y+1}}.$$

▼ Положим $f(x,y) = 2x+y+1$, $g(x,y) = (x^2+y+1)^{-1}$. Так как $f(1,-2) = 1$, $g(1,-2) = \infty$ то имеем неопределенность 1^∞. В соответствии с (9) вычислим логарифм искомого предела, равный

$$\lim_{\substack{x\to 1\\ y\to 2}} \frac{2x+y}{x^2+y+1}.$$

Находим частные производные числителя и знаменателя в точке $(1,-2)$:

$$\begin{aligned} (2x+y)'_x &= 2, & (2x+y)'_y &= 1,\\ (x^2+y+1)'_x &= 2, & (x^2+y+1)'_y &= 1. \end{aligned}$$

Строим отношения одноименных частных производных

$$\frac{(2x+y)'_x}{(x^2+y+1)'_x} = 1, \quad \frac{(2x+y)'_y}{(x^2+y+1)'_y} = 1.$$

Предел логарифма существует и равен единице. Искомый предел равен

$$\lim_{\substack{x\to 1\\ y\to 2}} (x,y)^{g(x,y)} = e^1 = e. \ \blacktriangle$$

III. Построение неопределенностей

В методических пособиях и задачниках-практикумах по математическому анализу недостаточно примеров на "раскрытие неопределенностей" для функций двух и более переменных. Это связано с отсутствием способов построения таких пределов. Обобщение правила Лопиталя для функций многих переменных позволяет строить неопределенности различных видов и порядков с заранее заданным значением предела в заданной точке.

Рассмотрим один из способов формирования заданных неопределенностей на примере функций двух переменных. Пусть $f(x,y) = f(M)$ — m раз дифференцируемая функция в точке $M_0 = (x_0, y_0)$. Формула Тейлора данной функции в достаточно малой окрестности точки M_0 с точностью до членов m-го порядка имеет вид

$$f(M) = f(M_0) + \sum_{i=1}^{m} \frac{\mathrm{d}^i f(M_0)}{i!} + o(\rho^m) \tag{10}$$

В связи с предельным переходом при $\rho \longrightarrow 0$ остаточным членом в (10) можно пренебречь.

Очевидно, что функция

$$f_1(M) = f(M) - f(M_0) = f(x,y) - f(x_0, y_0)$$

при $\mathrm{d}f(M_0) \neq 0$ является бесконечно малой первого порядка при $M \longrightarrow M_0$.

Далее, функция $f_2(M)$,

$$f_2(M) = f(M) - f(M_0) - \mathrm{d}f(M_0) = f(x,y) - f(x_0,y_0) - f'_x(x - x_0) - f'_y(y - y_0)$$

при $\mathrm{d}^2 f(M_0) \neq 0$ является бесконечно малой второго порядка при $M \longrightarrow M_0$, и т.д.

Таким образом, используя (10) и перенося последовательно слагаемые правой части (10) влево, можно получать бесконечно малые функции нужного порядка в заданной точке M_0.

Пусть, с учетом сказанного, даны две функции $f(M)$ и $g(M)$, дифференцируемые в точке M_0. Тогда неопределенность первого порядка вида $\frac{0}{0}$ имеет вид

$$\lim_{M \longrightarrow M_0} \frac{f_1(M)}{g_1(M)} = \lim_{M \longrightarrow M_0} \frac{f(M) - f(M_0)}{g(M) - g(M_0)}. \tag{11}$$

Для дважды дифференцируемых в точке M_0 функций $f(M)$ и $g(M)$ неопределенность второго порядка есть

$$\lim_{M \longrightarrow M_0} \frac{f_2(M)}{g_2(M)} = \lim_{M \longrightarrow M_0} \frac{f(M) - f(M_0) - f'_x(M_0)(x - x_0) - f'_y(M_0)(y - y_0)}{g(M) - g(M_0) - g'_x(M_0)(x - x_0) - g'_y(M_0)(y - y_0)}. \tag{12}$$

Итак, построены неопределенности (11) и (12) в заданной точке $M_0 = (x_0, y_0)$. Однако, пока неясно, существует ли двойной предел в точке M_0, и если да, то чему он равен. Рассмотрим неопределенность первого порядка (11). Дополним функцию $f_1(M)$ слагаемыми

$$f_1^*(M) = f_1(M) + C_1(x - x_0) + C_2(y - y_0)$$

Зададим значение двойного предела, равное k ($k \neq 0$, $k \neq \infty$). Очевидно, что функция $f_1^*(M)$ также первого порядка малости, что и $f_1(M)$ при $M \longrightarrow M_0$. Используя условия существования двойного предела (2), имеем

$$\lim_{M \longrightarrow M_0} \frac{f_1^*(M)}{g_1(M)} = \frac{(f_1^*)'_x}{(g_1)'_x} = \frac{(f_1^*)'_y}{(g_1)'_y} = \frac{f'_x(M_0) + C_1}{(g'_x(M_0)} = \frac{f'_y(M_0) + C_2}{(g'_y(M_0)} = k. \tag{13}$$

Из (13) находим требуемые значения C_1 и C_2.

$$C_1 = g'_x(M_0) \cdot k - f'_x(M_0), \;\; C_2 = g'_y(M_0) \cdot k - f'_y(M_0). \tag{14}$$

Таким образом, построен двойной предел с неопределенностью первого порядка

$$\lim_{M \longrightarrow M_0} \frac{f_1^*(M)}{g_1(M)} = \lim_{M \longrightarrow M_0} \frac{f(M) - f(M_0) + C_1(x - x_0) + C_2(x - x_0)}{g(M) - g(M_0)},$$

где C_1 и C_2 подчинены условиям (14).

Для неопределенности второго порядка при тех же M_0 и k, используя (12), легко получить

$$\lim_{M \longrightarrow M_0} \frac{f_2^*(M)}{g_2(M)} =$$

$$= \lim_{M \longrightarrow M_0} \frac{f(M) - f(M_0) - \mathrm{d}f(M_0) + C_1(x - x_0)^2 + C_2(x - x_0)(y - y_0) + C_3(y - y_0)^2}{g(M) - g(M_0) - \mathrm{d}g(M_0)}, \tag{15}$$

где

$$\begin{gathered} C_1 = \tfrac{1}{2}\left[k \cdot g''_{x^2}(M_0) - f''_{x^2}(M_0)\right], \; C_2 = k \cdot g''_{xy}(M_0) - f''_{xy}(M_0), \\ C_3 = \tfrac{1}{2}\left[k \cdot g''_{y^2}(M_0) - f''_{y^2}(M_0)\right]. \end{gathered} \tag{16}$$

Заметим, что при одних и тех же исходных функция $f(x, y)$ и $g(x, y)$, варьируя точки M_0 и значения k, можно получить множество различных двойных пределов.

Пример 3. Построить предел с неопределенностью второго порядка в точке $(0, 0)$ и значением предела, равным 3, если $f(M) = f(x, y) = x^3 + 2xy + y^3 + y^2$, $g(M) = g(x, y) = f(x, y) = x^2 + y^2 + yx$, т.е.

$$\lim_{M \longrightarrow M_0} \frac{f_2^*(M)}{g_2(M)} = 3.$$

▼ Схема построения предела такова.

а) Формирует функции $f_2^*(M)$ и $g_2(M)$:

$$f_2(M) = f(M) - f(M_0) - \mathrm{d}f(M_0).$$

Так как $f(M_0) = f(0,0) = 0$, $\mathrm{d}f(M_0) = 0$, то $f_2^*(M_0) = f(M) + C_1x^2 + C_2xy + C_3y^2$. Далее: $g_2(M) = g(M) - g(M_0) - \mathrm{d}g(M_0) = g(M)$, т.к. $g(M_0) = 0$, $\mathrm{d}g(M_0) = 0$.

б) В соответствии с (15) имеем

$$\lim_{\substack{x\to 0\\ y\to 0}} \frac{x^3 + 2xy^2 + y^3 + y^2 + C_1x^2 + C_2xy + C_3y^2}{x^2 + y^2 + yx}.$$

с) Вычислим коэффициенты C_1, C_2 и C_3 по формулам (16):

$$\begin{array}{lll} f''_{x^2}(M_0) = 0, & f''_{y^2}(M_0) = 0, & f''_{xy}(M_0) = 0, \\ g''_{x^2} = 2, & g''_{xy}(M_0) = 1, & g''_{y^2}(M_0) = 2, \\ C_1 = 3, & C_2 = 3, & C_3 = 3. \end{array}$$

Окончательно получим

$$\lim_{\substack{x\to 0\\ y\to 0}} \frac{x^3 + 2xy^2 + y^3 + 3(y^2 + x^2 + xy)}{x^2 + y^2 + yx} = 3. \ \blacktriangle$$

В работе не рассматриваются вопросы вычисления повторных пределов и их связи с двойными пределами. С помощью обобщенного правила Лопиталя эти вопросы решаются без существенных затруднений.

Литература

1. Ивлев В. В. "Неопределенности функций многих переменных" (часть I). "Математическое образование", №4(23), 2002г.

2. Фихтенгольц Г. М. "Курс дифференциального и интегрального исчисления", том I, Санкт-Петербург, изд "Лань", 1997г.

Валерий Васильевич Ивлев,
доктор технических наук,
профессор Московского государственного
открытого университета им. Шолохова.

Экспериментальный учебник для общеобразовательной школы "Геометрия 7-11"

Щетников А. И.

Автор излагает свое понимание целей и методов преподавания геометрии в общеобразовательной средней школе. В соответствии с этими целями и методами им написан экспериментальный учебник геометрии, который к настоящему времени несколько раз издавался пробными тиражами и по которому работали учителя в нескольких школах Кемеровской области и Новосибирска.

Краткая история проекта

К работе над экспериментальным учебником геометрии для 7-11 классов я приступил в 1997 году. В настоящее время один класс проучился по пилотной версии учебника от 7 класса до окончания школы; с другой стороны, главы, относящиеся к 7—8 классу, были существенно переработаны несколько раз с учетом опыта преподавания разных учителей.

Описание исходного замысла этого проекта было опубликовано в номере 3(14) "Математического образования" за 2000 г. В настоящем сообщении речь идет о проекте, который можно считать уже реализованным в "лабораторном" масштабе, и результатами которого я хотел бы заинтересовать более широкий круг лиц.

Апробация курса велась по личной договоренности с директорами и учителями четырех школ в Кемеровской области и одной школы в Новосибирске. Почти все эти школы с начала 90-х гг. участвовали в проекте "Развивающее обучение по системе Эльконина-Давыдова". Должен сказать, что я вовсе не стремился разработать учебный курс геометрии "в духе развивающего обучения"; мне всего лишь хотелось, чтобы на уроках геометрии в школе действительно можно было учиться видеть, думать и рассуждать. Мне казалось, что в лице Ассоциации развивающего обучения я нашел единомышленников в этом деле; впрочем, с тех пор многое переменилось.

Я хочу поблагодарить всех, кто вложил свои усилия в реализацию проекта, и прежде всего — Татьяну Павловну Сушину, вместе с которой мы, как нам кажется, выработали стиль ведения урока, соответствующий содержанию курса и тем целям, которых мы стремились достичь.

Общие принципы

1. Приоритетная цель курса геометрии, как я ее для себя сформулировал, состояла в том, чтобы все ученики в классе поняли, что такое геометрическая теорема и в чем состоит смысл доказательства, научились самостоятельно отыскивать доказательства относительно простых геометрических теорем и решать относительно несложные задачи на построение. Эта цель, какой бы простой она ни выглядела на первый взгляд, — в действительности (я имею в виду массовую общеобразовательную школу, а не специализированный физматкласс) достаточно сложна; и я не уверен в том, что самые распространенные в нынешней российской школе учебники геометрии ориентированы на ее достижение.

А еще я думаю, что самой важной частью курса геометрии является традиционная планиметрия, излагаемая в духе Евклида. Ведь эта древняя наука содержит в себе огромный заряд математических и эпистемологических идей; и не зря же Платон написал некогда над входом в Академию: "Да не войдет сюда не знающий геометрии"! Если школьники поймут, как устроена эта система знаний, как даются и проверяются определения, как доказательство собирает разрозненные факты в одну логически выстроенную цепочку, зачем нужны аксиомы, что мы делаем, когда ведем доказательство "от противного", — то общеобразовательная цель курса геометрии уже будет достигнута. Дело, как мне кажется, прежде всего в этих идейных основах курса, — а совсем не в том, чтобы закачать в головы школьникам массу "геометрических знаний".

Этой установкой обусловлена одна важная особенность учебника: значительная часть теорем приведена здесь без доказательств. Я рассчитывал на то, что учебный процесс будет строиться так, чтобы все эти теоремы доказывались самими учащимися (в коллективном обсуждении либо индивидуально). Тем, кто способен доказать теорему, будет полезнее сделать это самостоятельно, нежели узнать доказательство от учителя или из учебника; а тем, у кого это пока не получается, опять-таки будет лучше увидеть результат, полученный товарищем по классу.

2. Курс начинается не с традиционного перечисления аксиом, назначение которых остается по сути дела непонятным ученику, а с примера геометрической теоремы, которую предлагается доказать самим ученикам. Опыт показывает, что в классе всегда найдется кто-нибудь, способный это сделать.

Начиная с первого урока, мы стремимся к тому, чтобы у учеников сформировался образ *геометрии как деятельности*: на уроках геометрии мы в первую очередь доказываем теоремы; а для того, чтобы доказать теорему, нужно суметь разложить ее чертеж на отдельные существенные для дела факты и связать эти факты между собой в единое рассуждение.

Надо сказать, что первый год систематического изучения геометрии уходит на привыкание к этому образу предмета ("что значит доказывать?"; "а зачем вообще нужно доказывать?"; "а зачем мне нужно доказывать?"). Если учитель это понимает и ставит на первое место не передачу знаний, а формирование навыков геометрического мышления, то его усилия рано или поздно принесут ощутимые результаты.

3. Аксиома трактуется в нашем учебнике как такой геометрический факт, в который упирается доказательство некоторых теорем и который в силу его "элементарности" мы решаем принять без доказательства. Принятый в учебнике набор аксиом является одновременно и неполным, и избыточным.

С одной стороны, я не стремился устранить все неявные леммы. В частности, в учебнике нет явно выделенных аксиом о взаимном расположении точек и прямых на плоскости. Осмысленное выделение и использование таких аксиом требует способности работать на достаточно высоком уровне логической абстракции, чего мы никак не можем ожидать от учащихся, только приступивших к изучению систематического курса геометрии. А если мы не будем ими пользоваться, то зачем они нужны?

С другой стороны, принятые аксиомы не являются независимыми. В частности, за аксиомы приняты три признака равенства треугольников, истинность которых обосновывается работой с механическими моделями, демонстрирующими свойство "жесткости" треугольника. Если кому-то из читателей этой статьи данная идея покажется непривычной и странной, пусть он заглянет в любой из действующих школьных учебников геометрии, прочитает в нем доказательства этих признаков (строятся ли они "наложением" или "от противного"), а затем представит себя на месте учителя, объясняющего эти доказательства семиклассникам, только что начавшим изучать геометрию.

4. При проектировании многих разделов учебника я старался руководствоваться следующими словами Д. Д. МОРДУХАЙ-БОЛТОВСКОГО: "Следует рассматривать всякое школьное доказательство как наложение двух доказательств — интуитивного и логического, взаимно усиливающих убедительность друг друга. Первое осуществляется *подвижной* моделью и ей соответствующим процессом вооображения, второе развертывает *силлогизмы*, приводящие к обоснованию этих операций. То доказательство лучше, где ясно выявляются оба этих слоя — *интуитивный и логический*".

5. Мне думается, что когда ясно определена цель, отдельные простые детали также приобретают свой смысл в контексте целого. Для примера приведу здесь извлечение из методического руководства для 7 класса, часть "общие методические принципы": "Для того, чтобы геометрия была одним из привлекательных школьных предметов, нам нужно обустроить пространство урока так, чтобы работа в нем доставляла учащимся *эстетическое удовольствие.*

• Самое пристальное внимание следует уделить качеству доски и мела. Если линии на доске будут нечеткими и бледными, то усилия учителя по созданию в классе рабочей атмосферы во многом могут оказаться тщетными.

• Важнейшая задача 7 класса — сформировать у учащихся навык зрительного выделения разных фигур на одном и том же чертеже. Выделение элементов чертежа цветом (обведение контура, заливка площади) служит важным вспомогательным средством для этого. Пользуясь цветными мелками, мы можем в процессе решения задач говорить не о "треугольниках ABC и DEF", но о красном и синем треугольниках.

• Некоторые чертежи имеет смысл делать фломастерами на листе ватмана. Ведь в таком виде их можно сохранить до следующего урока, чтобы вернуться к ним еще раз.

• Следует ли чертить прямые линии на доске по линейке или от руки? Может быть, важнее научить учащихся проводить прямую от руки, изображать параллельные так, чтобы они выглядели параллельными, и т. п.

• В индивидуальной работе следует различать *наброски*, которые делаются в черновике от руки, и *чертежи*, выполняемые в основной рабочей тетради с помощью чертежных инструментов. Геометрическое мышление осуществляется не "в голове", а на листе бумаги, в набросках, когда рука и глаз вместе ищут, где и какую линию следует провести. Поэтому специальные черновые тетради представляют собой важнейшее средство для решения задач по геометрии. Надо завести их и приучить учеников к тому, что *поиск доказательства ведется в черновиках*. Но такое название не означает, что наброски делаются здесь небрежно. Может быть, лучше говорить о "тетрадях для решения задач"? Аккуратный чертеж иногда оказывается ключом к решению, потому что на нем лучше видно, что следует делать."

Некоторые частные детали

Перечисленные ниже детали сами по себе не образуют системы, однако мне думается, что все они проистекают из заявленных выше общих целей.

1. Такая важная теорема, как теорема Пифагора, рассматривается в учебнике несколько раз. На первом проходе она доказывается несколькими способами, основанными на идее равносоставленных фигур; затем мы возвращаемся к ней еще раз и доказываем ее исходя из идеи геометрического подобия.

2. Переход от сравнения величин к их измерению отложен (в духе Евклида) до изучения темы "геометрическое подобие". До этого момента мы довольствуемся аксиомой "целое больше части", ее нам вполне хватает. Точно также мы сначала доказываем теорему "параллелограмм равновелик прямоугольнику с таким же основанием и высотой" (эта формулировка уже схватывает преобразование чертежа, которым теорема доказывается), а потом только превращаем ее в формулу "площадь параллелограмма равна произведению его основания на высоту".

3. Школьный учебник алгебры не доказывает, что $\sqrt{2} \cdot \sqrt{3} = \sqrt{6}$; кажется, что и учебнику геометрии нет нужды гнаться за строгим обоснованием всех теорем, в которых явно или неявно содержится понятие действительного числа и предела.

• Для теоремы Фалеса мы приводим доказательство, основанное на применении десятичных дробей; но перестановка членов пропорции производится без доказательства, со ссылкой на курс алгебры (где тоже ничего не доказывается, и для 8 класса это правильно).

• Тот факт, что площадь прямоугольника равна произведению длин его сторон, для общего случая также не доказывается строго (тем более что школьное определение произведения двух чисел само опирается на наглядную схему прямоугольника).

• Наличие у окружности такой характеристики, как "длина", предполагается интуитивно очевидным: длина окружности — это длина нитки, которую можно распрямить и сравнить с диаметром. Мы воздерживаемся от того, чтобы *определять* длину окружности как предел последовательности вписанных и описанных многоугольников. Построения такой последовательности рассматривается как основа для вычислительной процедуры, позволяющей установить значение числа π с любой степенью точности.

• Теоремы об объемах пирамиды, конуса и шара в 11 классе доказываются на основе принципа Кавальери, а не с использованием формул интегрального исчисления. Ранее в 8 классе некий аналог принципа Кавальери (с разворачиванием концентрических окружностей в прямые) используется при установлении теоремы о площади круга.

4. При изучении координатного метода в 9 классе сначала вводится общая косоугольная система координат, с ее помощью решается ряд задач аффинного характера. Декартова система координат (известная школьникам из курса алгебры) вводится как удобное средство для решения задач метрического характера.

5. Вся "геометрическая" (т. е. связанная с рассмотрением чертежей, а не с формальными алгебраическими преобразованиями) тригонометрия собрана в один блок, изучаемый в конце 9 класса. К этому блоку относятся: (1) "решение прямоугольных треугольников"; (2) определения тригонометрических функций для произвольного значения аргумента как проекций точки на единичной окружности; (3) теоремы сложения; (4) понятие скалярного произведения векторов. На наш взгляд, весь этот материал имеет смысл излагать в одной главе, а не вперемешку с другими разделами.

6. При изучении главы "взаимное расположение прямых и плоскостей в пространстве" основной упор делается на построение сечений многогранников. Причем мы сначала учимся строить сечения (для этого даже сделаны специальные рабочие тетрадки), а потом только выделяем аксиомы и доказываем теоремы.

7. Тема "векторы и координаты в пространстве" из учебника исключена совсем. По большей части она дублирует своим содержанием соответствующие темы 9 класса; а то новое, что может быть в нее внесено (уравнение плоскости и параметрическое уравнение прямой) не находит себе использования в адекватных задачах. Остается определение компланарных векторов — но настолько ли оно важно? И вообще мы смотрим на эту тему как на искусственно перенесенную в школу из вузовского курса.

Щетников Андрей Иванович,
координатор Лаборатории теоретической
и экспериментальной эпистемологии.

E-mail: schetnikov@ngs.ru
http://ltpe.stsland.ru

О Фонде математического образования и просвещения

Фонд математического образования и просвещения создан в конце 1996 г. с целью обеспечения условий, способствующих сохранению богатых традиций математического образования и науки в России. Фонд сотрудничает с организациями и гражданами, желающими участвовать в благородном деле сохранения лучших традиций и высокого качества математического образования в России. Фонд поддерживает образовательные инициативы, способствующие поставленной цели. Особое внимание оказывает образовательным инициативам в провинции, как в виде издательской поддержки, так и финансовой помощи. Фонд издает научную, учебную и методическую литературу в области математики и смежных наук.

Условия подписки и приема материалов

По вопросам подписки на журнал обращайтесь по адресу: 111250, Москва, пр-д завода "Серп и Молот", д.3а.

Контактные телефоны: (095) 362-91-70, (095) 362-91-02.

Этот же адрес и телефоны для корреспонденции Фонда.

E-mail: fmop@dnttm.ru

Стоимость подписки на каждый из номеров 1-4 за 2003 год (включая стоимость пересылки) – 45 рублей.

Для получения номеров журнала необходимо выслать в адрес редакции копию платежного документа, подтверждающего оплату подписки. Сообщите адрес, по которому вы хотели бы получать журнал. В платежном документе укажите, что перевод делается для журнала "Математическое образование", номер журнала за 2003 г., количество экземпляров.

Реквизиты для перечисления:

Получатель: ИНН 7725080165 Фонд математического образования и просвещения

Расчетный счет и банк получателя:
р/с 40703810138120100114 в Московском банке СБ РФ, Лефортовском отделении №6901/019 г. Москвы, к/с 30101810400000000225, БИК 044525225

С сентября 2000 выходит "Обозрение Z" — научно-популярное приложение к журналу "Математическое образование". Условия подписки (адрес, реквизиты, стоимость одного номера) — те же, что и для журнала.

Вы также можете заказать необходимое вам количество отдельных номеров журнала за предыдущие годы. В этом случае пересылка осуществляется наложенным платежом или на основании платежного документа (реквизиты те же). В заказе (в платежном документе) укажите, за какие номера и в каком количестве экземпляров за номер, делается перечисление.

Стоимость одного экземпляра журнала (с учетом пересылки) — 35 руб., сдвоенных номеров 3-4 (6-7) за 1998 г. и 2-3 (9-10) за 1999 г. — 45 руб.

Редакция принимает рукописи материалов с четко прорисованными формулами. По согласованию с редакцией принимаются материалы в электронном виде, обязательно прилагать распечатку.

Рукописи не возвращаются и не рецензируются. После публикации авторские права сохраняются за авторами материалов. Авторы опубликованных материалов получают бесплатно по 10 экз. соответствующего выпуска журнала.

Мнение редакции не всегда совпадает с мнением авторов.

Contents