ISSN 1992-6138

Математическое Образование

Журнал Фонда математического образования и просвещения

Год двенадцатый

№ 2 (46)

апрель – июнь 2008 г.

Москва

МАТЕМАТИЧЕСКОЕ ОБРАЗОВАНИЕ, ISSN 1992-6138

Периодическое издание в области математического образования

Издатель и учредитель: Фонд математического образования и просвещения
117419 Москва, ул. Донская, д. 37

№ 2 (46), 2008 г.

Фонд математического образования и просвещения, Москва, 2008 г.
“Математическое образование”, периодическое издание.
Зарегистрировано в Роскомпечати РФ, лицензия № 015955 от 15.04.97.
Подписано к печати 30.06.2008 г.
Компьютерный набор и верстка, компьютерная графика: С. Кулешов.
Экспедиторы: Давыдов Д.А., Ермишкин И.А., Истомин Д.Н.
Объем 4,5 п.л. Тираж 1000 экз. Цена свободная.

Математическое образование

Журнал Фонда математического образования и просвещения

№ 2 (46), апрель – июнь 2008 г.

Содержание

К 85-летию И. Р. Шафаревича

Интервью и библиография

3 июня 2008 г. исполнилось 85 лет выдающемуся русскому математику и общественному деятелю академику Игорю Ростиславовичу Шафаревичу. Редакция поздравляет юбиляра и желает здоровья, долгих лет жизни и творческой деятельности. Публикуем выдержки из интервью, взятого у И. Р. Шафаревича Милкой Здравковской для журнала "The Mathematical Intelligencer" (том 11, №2, 1989), которое не печаталось на русском языке. Игорь Ростиславович является постоянным автором журнала "Математическое образование". Приводим список его публикаций (а также публикаций, посвященных ему) в нашем журнале.

I. Из интервью журналу "The Mathematical Intelligencer"

Об интересах в детстве

Первое мое глубокое ощущение от совершенно нового для меня феномена, не знаю, как это точно сформулировать, возможно, культурного феномена — связано с историей. Я помню, как я читал книгу, какую-то немецкую книгу о римской истории (впоследствии я даже удивлялся, что она произвела такое впечатление — это очень скучная книга). Но она открыла мои глаза на факт, что мир не кончается моими персональными ощущениями, продолжается до бесконечности по всем направлениям. Я бросился читать исторические книги и некоторое время был убежден, что, конечно, я стану профессиональным историком — всю жизнь буду работать над историей. Я пытался вспомнить, как я переключился на математику, но не могу вспомнить логики или причины для этого, хотя я думаю, что стать в то время историком было бы, вероятно, очень трудно. История считалась идеологической дисциплиной, и возможность выбора различных точек зрения была очень ограниченной. Вряд ли 10-12-летним мальчиком я мог все это понимать, но, возможно, чувствовал это подсознательно.

О поступлении в университет и учителях

Фактически я не поступал в университет. Я был самоуверенным ребенком. Еще школьником, я пришел к декану и сказал, что читаю книги по математике, но не уверен, действительно ли я их правильно понимаю. Можно ли попытаться сдать экзамен? Он ответил, что возможно сделать меня студентом-экстерном. На первый экзамен (по аналитической геометрии) он направил меня к Делоне, на второй (алгебра) — к Курошу, на третий (анализ) — к Гельфанду. Таким образом, я встретился со всеми тремя. Они были очень доброжелательны по отношению ко мне, много со мной работали, в частности, дали много литературы по тем вопросом, которые требовалось изучить на высоком уровне. Это было время перемен в средней школе: иногда она состояла из 10 классов, иногда — из 9. К этому времени я окончил 9-й класс. Я сдал почти все университетские экзамены и был зачислен студентом последнего курса, так что у меня нет аттестата о среднем образовании. По окончании университета мне исполнилось 17 лет.

Делоне и Куроша я считаю своими учителями. Они представляли собой два не просто разных, но диаметрально противоположных математических типа. Делоне был классическим представителем Петербургской школы. Он интересовался диофантовым анализом совершенно классического типа. Все более-менее абстрактные направления математики были ему полностью чужды, например, общая топология или то направление в алгебре, основоположником которого стала Эмми Нетер. Курош же был, наоборот, радикалом. Он говорил мне, что математику можно разделить на две части: философию и вычисления. Все формулы, интегралы — все это вычисления.

И все, что сделано в математике до 20-го века, уже потеряло интерес. Это была психология, аналогичная психологии футуризма или авангардизма в искусстве. Это двустороннее влияние помогло мне, в некотором смысле, найти средний путь. Вот почему, когда на Западе начали появляться новые фундаментальные направления, такие как гомологическая алгебра, я почувствовал, что это было что-то вроде синтеза абстрактных понятий с конкретными задачами теории комплексных многообразий или алгебраической геометрии. Я чувствовал, что это было нечто мне близкое.

Я бы хотел добавить, что Делоне дал мне два особенно хороших совета: первый был прочитать "Zahlbericht" (теория чисел) Гильберта, а второй — читать Гаусса. В результате этого чтения я выбрал область, над которой работал около 20 лет.

"Zahlbericht" Гильберта я читал в возрасте 15 лет. В таком возрасте книга производит особенно сильное впечатление. В 14 Делоне рекомендовал мне читать теорию Галуа. Но тогда я ее абсолютно не понял. Возможно потому, что изложение было очень тяжелым: я читал книгу Чеботарева. Это книга написано очень тяжело, запутанно. Поэтому потом я вычитал этот материал из книги Ван-дер-Вардена. В следующем году, когда мне было 15, я читал Гильберта. Я уже немного знал алгебру, чтобы понимать эту книгу. Она не использует никакого специального математического аппарата.

О Егорове

Это все произошло задолго до того, как я появился на мех-мате. Так что я знаю это, как говорится, со слов других. Я слышал это от своих старших сотрудников.

У меня сложились впечатление, что примерно одинаково важную роль в создании Московской математической школы сыграли два математика: Лузин и Егоров. Однако они были совершенно разными личностями. Лузин был очень экстравертным. Он был замечательным лектором. Некоторые даже говорили, что им не нравится излишняя театральность его лекций. Он оказывал колоссальное влияние на студентов, особенно на молодых. Они образовали вокруг него небольшую секту поклонников. Егоров же казался гораздо более замкнутым и сдержанным человеком. Однако по рассказам у меня сложилось впечатление, что его влияние на создание Московской математической школы было не меньшим. Он является основоположником школы теории функций вещественной переменной, от него исходили первые импульсы создания школы функционального анализа. Он был президентом Московского Математического Общества в 30-е годы. Я могу показать копию опубликованного Кольманом доклада о ситуации в ММО из журнала "Математические знания в массы". Он описывает, как увольняли реакционера Егорова. Были упомянуты три реакционера: Егоров, Финников и Аппельрот. О последнем я никогда не слышал. Про Финникова я знаю хорошо. Он тоже пострадал. Ничего катастрофического не произошло, но случай для него был тяжелый. Вероятно, его уволили. Однако Егоров был арестован. Некоторое время он находился в заключении, затем был сослан в Казань. Кажется, в тюрьме он заболел. Как говорил мне Чеботарев, Егоров умер в его доме, на его руках. Чеботарев уже долго жил в Казани, я думаю, с окончания гражданской войны. Он оканчивал университет вместе с Делоне. Чеботарев — математик из школы Д. А. Граве. Сам Граве не стал очень известным как математик, но он создал блестящую школу в Киеве, по крайней мере три студента которой стали знаменитыми: Делоне, Чеботарев и Шмидт.

О довоенном и послевоенном поколениях математиков

В некотором смысле, привлекательность мехмата до войны объясняется тем, что можно было оставить позади жизненные трудности. Математика была, с одной стороны, не технической специальностью, так что не возникало связи с болезненными практическими проблемами. С другой стороны, она не была гуманитарной наукой, человек не чувствовал элементов идеологического вмешательства. У меня было чувство, что мехмат был привлекателен для многих как место, где можно было заниматься наукой в ее наиболее концентрированной форме без какого-либо внешнего влияния. Во время войны отношение к науке сильно изменилось. До войны ученые скудно оплачивались и не имели особенного престижа. Уровень жизни ученых был очень низок. Во время войны все это сильно изменилось. Жалования практически внезапно увеличились в 2-3 раза. Очень усилился престиж. Об ученых начали писать. Наиболее престижно, конечно, было быть физиком. Следующее место, вероятно, занимали математики.

Следует сказать, что это не привело к немедленному увеличению количества одаренных студентов на мехмате в 1940-е годы. Но затем, неожиданно, в середине 1950-х, по причине, которой я понять не могу, это произошло. Но я должен сказать, что за время того периода, который я наблюдал, это произошло только однажды — взрыв, который буквально за 4 года дал такую волну замечательных математиков.

О выборе темы для работы, “модных” и “немодных” направлениях

Каждый математик старается найти нечто, как он полагает, интересное и неизвестное. Надо найти интересную задачу, например, что-то, на что раньше не обратили внимания. Это вопрос удачи. В моей математической жизни у меня были успехи наполовину, успехи и неудачи.

Например, я помню, как мы с Кострикиным уехали из Москвы на прогулку и возвращаясь на электричке, начали обсуждение. В это время я интересовался работой Эли Картана по “бесконечным” группам Ли, или, как их теперь называют, “псевдогруппам”. Это частью геометрическая, частью аналитическая теория. Кострикин изучал конечномерные алгебры Ли над полем конечной характеристики — чисто алгебраический вопрос. И когда мы начали сравнивать данную Картаном классификацию с известными примерами алгебр Ли, оказалось, что они совпадают. Мы заключили, что, хотя это совершенно разные области, имеется полный параллелизм и что, благодаря этому, можно перенести в другую область технику Картана. Мы сделали первые шаги в направлении доказательства. Это был успех наполовину, но не полный успех. Я очень сожалею, что моя школа не довела это дело до конца, очевидно, были люди, которые могли бы это доделать. Проблема была окончательно решена двумя американскими математиками Р. Блоком и Р. Вильсоном несколько лет назад. Очень хорошие и классические алгебраические вопросы были полностью решены.

Были случаи полных неудач, которые для всех оставались неизвестными. Например, меня всегда интересовала следующая задача. Рассмотрим все прямые, проходящие через начало координат. Они образуют проективное пространство; можно рассмотреть расслоенное пространство над проективным пространством. Далее, что представляет собой множество траекторий обыкновенного дифференциального уравнения? Абсолютно непонятно. Если решение наматывается на предельный цикл, эти две точки следует считать бесконечно близкими, неотделимыми в соответствующей топологии[1]. С другой стороны, локально, есть теорема единственности: каждое решение пересекает трансверсальную плоскость только один раз. И мне казалось, что Гротендик создал аппарат в алгебраической геометрии как раз для изучения такого типа феноменов (которые хотя и могут быть описаны локально, глобально ничего не выходит). Я имею в виду топологию Гротендика. Я заинтересовал этой проблемой своего ученика Паршина. Мы очень упорно трудились над этой задачей в течение нескольких месяцев, но абсолютно безуспешно, т.е. ничего не работало. Казалось, что мы построили некоторые топологические инварианты вроде эйлеровой характеристики или чисел Бетти... Мы надеялись, что наконец они окажутся инвариантами дифференциального уравнения, которые помогут оценить число предельных циклов или что-то в этом роде. Сейчас причина неудачи более-менее ясна. Кажется, Алан Коннэ из Парижа имеет некоторые соображения в этом направлении. Но у него появилась дополнительная идея, которая никогда не приходила нам в голову: кольца функций на этих пространствах могут оказаться некоммутативными.

В связи с этим я вспоминаю, что Меньшов, которому сейчас около 100 лет (Меньшов умер в декабре 1988 г.), давно рассказывал мне, что когда приезжал Адамар и был устроен обед в его честь, Меньшов спросил Адамара, как тот выбирает вопросы для работы. Адамар ответил такой восточной притчей:

> Одному султану надоели женщины из его гарема и он позвал своего главного евнуха и приказал ему найти новую наложницу, которая понравилась бы султану. Если в течение двух дней приказ не будет выполнен, евнуху отрубят голову. Евнух пошел

[1] Имеются в виду две точки в пространстве кривых: само решение и предельный цикл — *Прим. ред.*

> на рынок невольниц и увидел, что там очень много женщин. Но они не отличались от женщин султанского гарема. Случайно он встретил своего старого друга, купца, который только что возвратился из путешествия и выглядел помолодевшим и загорелым. Купец спросил, в чем причина печали евнуха, и тот рассказал про свое задание. "Разве это так трудно? Купи, например, эту девушку." Евнух решил рискнуть и последовал совету. На следующее утро султан позвал его и сказал: "Иди в мою сокровищницу и возьми себе столько, сколько сможешь унести." Евнух был счастлив, но опасался, что подобное задание может повториться. Поэтому он пошел к купцу и попросил: "Открой мне секрет. Как ты выбрал ее?" — "Знаешь, нет никакого секрета. Просто надо иметь глаза..."

Итак, заключил Адамар, чтобы находить хорошие математические задачи, надо иметь глаза...

О текущих математических интересах

Они связаны с некоторыми воспоминаниями моей молодости. Это воспоминания такого характера: знакомишься впервые с некоторой областью и это вызывает вопросы, многое кажется неясным. Я помню, что меня очень удивляли две вещи. В алгебре, особенно в теории Галуа, в алгебраической теории чисел, в теории полей классов, в теории алгебраических функций существует красивая теория, связанная с различными накрытиями, группами, которые являются абелевыми. Если же соответствующая группа Галуа или фундаментальная группа неабелева, теория перестает существовать. У меня сложилась идея, что к настоящему времени создана абелева математика, но математика будущего будет неабелевой. Через год, к моему великому удивлению, я обнаружил работу Андре Вейля, в которой он ранее выразил ту же самую идею. В ее развитие, он сделал кое-что очень интересное в направлении неабелевой математики. Тем не менее, я тоже продолжал работать в этом направлении и долгое время это было темой моих исследований. Уже недавно произошел такой случай: меня пригласили прочитать на собрании Московского Математического Общества для разнообразной аудитории студентов лекцию, ориентирующую на математику. Я подумал, что для них это было бы подходящей темой, потому что, хотя что-то и было сделано, в целом ситуация была такой же — все, что связано с неабелевыми группами, оставалось в большой мере таинственным. К сожалению, за неделю до лекции мне позвонили и сообщили, что "по некоторым причинам" лекция будет отменена. Я думаю, что до сих пор никто даже не очертил текущую ситуацию. Конечно, та замечательная работа Андре Вейля известна, но никто не сделал общего обзора того, что из нее вытекает. Я планирую написать об этом статью для журнала "Mathematical Intelligencer"[2].

Второе впечатление, которое остается со мной с молодости: я был очень удивлен, что в теории алгебр (существуют различные алгебры — ассоциативные и алгебры Ли; все они применяются к многообразиям) все оканчивается случаем полупростых алгебр. Алгебраисты, с которыми я об этом разговаривал, всегда по той или иной причине говорили: "Эти алгебры нильпотентны, и там нет ничего интересного". Единственный, от кого я нашел поддержку, был все тот же Андре Вейль с той же самой работой. Там он в некоторый момент доходит до нильпотентных алгебр и пишет, что о них "известно, что ничего не известно". Позже я переписывался с ним об этом. Он ответил, что до сих пор считает, что эти алгебры — нечто таинственное, но не бессмысленное. Некоторые изучают их, но это не престижное направление в математике, хотя там имеются четко сформулированные трудные проблемы, которые пытались решить, но безуспешно. Они имеют приложение, например, в классификации особенностей отображений. В своей книге Арнольд, Варченко и Гуссейн-Заде образно говорят, что некоторые объекты оказываются инфинитезимальными, "неразличимыми глазом", вот почему приходится описывать эти объекты посредством колец функций на них, а эти последние оказываются нильпотентными алгебрами. Поскольку я уже в таком возрасте, когда никто от меня ничего не ожидает, я не боюсь работать над такими вопросами. Если ничего не получится, это не будет скандальным событием. По крайней мере, я надеюсь привлечь к этому внимание. Ясно, что это имеет

[2] Abelian and Nonabelian Mathematics, The Mathematical Intelligencer, том. 13, 1991, стр. 67-75. На русском языке см. Абелева и неабелева математика, И. Р. Шафаревич, сочинения в 3 томах. Том III, Математические работы, часть I, стр. 397-415, М.: АОЗТ "Прима В", 1996 — *Прим. ред.*

взаимосвязи и с другими областями, например, векторными расслоениями. Мне кажется, это интересное направление, которое пока не привлекает должного внимания математиков.

О текущем положении в стране, пессимизме молодых и оптимизме

К сожалению, поскольку я прекратил преподавать в университете, мои контакты с молодежью сократились. Сам я оптимист. Я смотрю на происходящее иррационально оптимистически. Из чтения различных исторических книг я вынес впечатление, что человечество находит свои пути и осуществляет свои решения в результате очень сложного процесса, сложнее, чем может быть результат деятельности одного индивидуума. Поэтому, когда такое решение принято, оно непостижимо для индивидуума. Когда возникает историческая ситуация, представляющаяся индивидууму кризисом, и он просто логически видит, что выхода нет, это не значит, что выхода действительно нет, потому что выход можно найти, используя более мощное сознание человечества, а не отдельного индивидуума. Поэтому, с одной стороны, то, что сейчас — период кризиса, это, видимо, объективный факт. Я вспоминаю, что после войны одна из вещей, которую мне постоянно приходилось слышать, что мы находимся на пороге катастрофы. (Эта ситуация напоминает аналогичную в средние века, когда люди ожидали конца света.) Сначала говорили, что изобретена атомная бомба и мы наверняка обречены истребить друг друга. Затем, что демографический взрыв погубит человечество. Теперь, что будет разрушен озоновый слой, или будет недостаток чистой воды, или Земля будет отравлена химическими удобрениями. Инвариант всех этих угроз: кажется, что нарушена естественность контакта человека с природой. Возможно, сейчас действительно критическое время. Но отсюда не следует, что выхода из кризиса нет. Он, вероятно, найдется. Конечно, как при любом кризисе, это связано с потрясениями, но будут ли они большие или малые, это, конечно, неясно. Я думаю, что человек — это очень прочное создание. Мы достаточно видели, как наносится вред человеку и природе, но мы не осознаем, как много сил остается. Возьмем для примера: что пишут в газетах, как много мыслей аккумулировано в человеческих головах! Казалось бы, реакцией должно стать общее безразличие, никто ничем не будет интересоваться. Однако, наоборот. Внезапно, как только представилась возможность, возникает огромный взрыв активности. Сколько образовалось обществ! Например, общества *Доверие, Память* — они противоположны. Ясно, что все это молодые люди, которые совсем не безразличны к тому, что происходит в их стране. Они могут полностью расходиться во взглядах на то, что следует делать, но не в том, что это их касается и что они сами должны это делать. И они воспитаны на совершенно противоположных принципах.

О переменах в Советском Союзе

Это что-то совершенно непредсказуемое. Я думаю, что в итоге это приведет к чему-то, чего никто не ожидает, ни те, кто за, ни те, кто против перестройки. Кажется, что начался процесс, который никто не может контролировать. Конечно, он может окончиться чем-то печальным, например, социальным взрывом.

О музыке

Кроме математики, я больше всего интересуюсь историей (я рассматриваю ее как прикладную науку, которая дает возможность понять, что происходит сейчас) и затем музыкой. Я действительно часто посещал консерваторию. И сейчас, хотя это и стало трудней, я хожу туда регулярно. Но в то время, перед смертью Сталина, музыка играла особую роль в культуре, потому что она была, в некотором смысле, неуправляемой. Не каждый мог понять это. В прошлом году была 80-я годовщина со дня рождения Шостаковича. По такому случаю я очень много слушал его, прослушал все снова. Это удивительно. Конечно, ясно, что композитор пишет о своем времени, и что можно там увидеть? Это просто Апокалипсис. Как такое могло случиться? Это было для посвященных. Это было как язык, на котором можно общаться. Тогда еще не было ни Солженицына, никого. В целом музыка имела огромный вес в культуре.

II. Публикации И. Р. Шафаревича (и публикации о нем) в журнале "Математическое образование"

1. Избранные главы алгебры, №1, 1997 – №3-4(6-7), 1998.
2. Математическое мышление и природа, №2(5), 1998.
3. Из истории естественно-научного мировоззрения, №1(12), 2000.
4. Воспоминания об Алексее Ивановиче Кострикине, №1(16), 2001.
5. От редакции. К 80-летию И. Р. Шафаревича, №2(25), 2003.
6. Математические работы И. Р. Шафаревича, №2(25), 2003.
7. Список математических публикаций И. Р. Шафаревича, №2(25), 2003.
8. О некоторых тенденциях развития математики, №2(25), 2003.
9. Исследования Эйлера по теории чисел, №3(43), 2007.
10. К 85-летию И. Р. Шафаревича. Из интервью журналу "The Mathematical Intelligencer"; Публикации И. Р. Шафаревича (и публикации о нем) в журнале "Математическое образование", №2(46), 2008.

Сергей Львович Соболев (к столетию со дня рождения)

А. В. Фурсиков

В этой статье мы расскажем о Сергее Львовиче Соболеве — одном из крупнейших математиков XX века. В 2008 году отмечается столетие со дня его рождения. При описании жизни и деятельности С. Л. Соболева были использованы работы [1], [2]. Они могут быть рекомендованы всем желающим подробнее ознакомиться с предметом. Литературу о С. Л. Соболеве можно найти на сайте http://www.prometeus.nsc.ru/science/schools/sobolev/biblio/about.ssi

1. Блестящее начало

С. Л. Соболев родился 6 октября 1908 г. С детства он отличался разносторонней одаренностью и большой любознательностью. В 1919 году в разгар гражданской войны, спасаясь от голода, семья переезжает в Харьков. Из-за крайней бытовой неустроенности тех лет программу средней школы Сергей Львович осваивал в основном самостоятельно. Увлекался поэзией и музыкой, читал много книг по разнообразным областям знания, любил играть в шахматы.

После возвращения в Петроград в 1923г. он поступил в последний класс 190-й школы, которую окончил с отличием в 1924 году в возрасте 15 лет. Как вспоминал Сергей Львович, его учительница физики рекомендовала ему поступать на математический факультет. Однако, из-за возраста он не смог этого сделать сразу после окончания школы, и поэтому в 1924 году начал занятия в Первой государственной художественной студии по классу фортепиано, которую закончил в 1928 году.

В 1925 году Сергей Львович поступил на физико-математический факультет Ленинградского государственного университета (ЛГУ). В те годы ЛГУ был крупнейшим научным математическим центром, сохранившим замечательные традиции Петербургской математической школы, связанной с именами таких корифеев науки, как П. Л. Чебышев, А. М. Ляпунов, А. А. Марков. В ЛГУ Сергей Львович слушал лекции блестящих профессоров Н. М. Гюнтера, В. И. Смирнова, Г. М. Фихтенгольца. Дипломную работу он писал под руководством Н. М. Гюнтера. Университет закончил в 1929г.

После окончания ЛГУ С. Л. Соболев работает в теоретическом отделе Сейсмологического института АН (Академии Наук) СССР. В эти годы чрезвычайно интенсивной деятельности он весьма успешно трудится над созданием математических методов исследования прикладных задач, связанных с распространением волн в упругих средах. Именно эти исследования привели С. Л. Соболева к понятиям обобщенного решения и обобщенной функции, т.е. к замечательному открытию, приведшему к коренному преобразованию многих разделов математики. Об этих понятиях мы чуть подробнее расскажем ниже.

С 1932 г. С. Л. Соболев работает в Математическом институте им. В. А. Стеклова АН СССР. В 1934 г. в связи с переводом из Ленинграда в Москву Академии Наук и, в частности, Института, С. Л. Соболев переезжает в Москву. В 1936 году была опубликована работа С. Л. Соболева [3], в которой были заложены основы теории обобщенных функций. Переходя от динамических задач теории упругости к статическим, Сергей Львович обращается к изучению краевых задач для эллиптических уравнений высокого порядка. В результате этих исследований он ввел в математику пространства, которые сейчас называются его именем, и установил для них первые теоремы вложения, что, как выяснилось в дальнейшем, оказалось исключительно важным открытием.

Научные достижения С. Л. Соболева получили широкое признание. В 1933 году он был избран членом- корреспондентом АН СССР, а в 1939 — действительным членом Академии наук. Многие годы впоследствии он был самым молодым академиком. О нем много писали газеты, а такое внимание прессы к математику весьма необычно. “Мы будем такими полярниками как Папанин, такими летчиками как Чкалов, такими математиками как Соболев, такими шахтерами как Стаханов, такими поэтами как Маяковский” — написано в приветствии пионеров Москвы XVIII съезду ВКП(б), прошедшему в 1939 году.

Последующая жизнь молодого академика показала, что он с честью преодолел одно из труднейших жизненных испытаний — испытание успехом. Сергей Львович впоследствии достаточно скептически оценивал шумиху, поднятую вокруг его имени: "Меня хвалили безудержно. Мне это неприятно, я всё-таки считаю, что сам человек — лучший судья своих работ. Что касается моих работ, то тогда никто ещё не мог разобраться, что из этого вырастет."

2. Служение Отечеству

Вся жизнь Сергея Львовича была ярким примером беззаветного служения Родине, интересы которой для него всегда были на первом месте.

Военное время. Атомный проект

В 1941 году, вскоре после начала Великой отечественной войны Сергей Львович избирается директором Математического института им. В. А. Стеклова. На его плечи легла обязанность по организации эвакуации сотрудников института и членов их семей в Казань. В Казани С. Л. Соболев провел вместе с Институтом 2 года, много сделав для организации исследований, нужных фронту. Обладая незаурядными способностями к решению конкретных прикладных задач, он принимает непосредственное участие в решении задач по расчету артиллерийской стрельбы и бомбометания.

После того, как в декабре 1942 в Чикаго группа Э. Ферми впервые осуществила управляемую цепную ядерную реакцию, Советское правительство приняло решение о развертывании работ по созданию атомной бомбы. В 1943 году организуется так называемая лаборатория №2 во главе с академиком И. В. Курчатовым, впоследствии переименованная в Институт атомной энергии. Первым заместителем И. В. Курчатова был назначен С. Л. Соболев. И в обстановке глубочайшей секретности началась интенсивная работа по созданию ядерного щита страны. Обстановка была напряженной и во время войны, когда фашистская Германия предпринимала отчаянные усилия по созданию атомной бомбы, и по ее окончании. Ведь в сентябре 1945 года США подвергли атомной бомбардировке Японию, а Фултонская речь У. Черчиля (1946г.) положила начало холодной войне стран Запада против СССР. "Тогда было такое ощущение — вспоминал Сергей Львович, — что если не выйдет наша работа, то неизвестно, что станет со страной. Но у всех была уверенность, что выйдет, обязательно выйдет".

Сергей Львович работал над вопросами использования атомной энергии. Значительная часть этих проблем относилась к уравнениям математической физики. Нужно было глубокое понимание совершенно новых, ранее не изучавшихся физических процессов, чтобы правильно поставить задачу математической физики. Далее, для инженерно-технических воплощений необходимо было численное решение этих больших математических задач, причем при очень малых вычислительных средствах. Ведь в первые годы работы над атомным проектом в распоряжении исследователей были лишь логарифмические линейки и арифмометры. Все вычисления проводились большими коллективами вычислителей, работающих на этих примитивных устройствах. В такой ситуации требовалась исключительная математическая интуиция, изобретательность и огромный труд для разработки специальных методов счета и контроля полученных результатов. Электронные вычислительные машины (ЭВМ), появившиеся через несколько лет, облегчили вычислительную работу, но одновременно и породили ряд новых проблем . С. Л. Соболев отдавал этому делу и другим проблемам, связанным с атомным проектом, все свое время и энергию. За работы, выполненные в Институте атомной энергии, С. Л. Соболеву дважды присуждалась Государственная премия, а в январе 1952 года ему было присвоено звание Героя Социалистического Труда.

Когда Сергей Львович работал в Институте атомных исследований, он однажды сломал ногу. Наложив гипс, врачи предписали ему постельный режим на 6 недель. За это время, систематизировав результаты 15-летней работы, Сергей Львович написал книгу "Некоторые применения функционального анализа в математической физике" [4], которая была опубликована в 1950 году. Эта знаменитая книга по пространствам Соболева была в дальнейшем переведена на многие иностранные языки и сыграла важную роль в деле популяризации идей С. Л. Соболева и их внедрения в математическую физику.

Преподавательская деятельность

В течение всей своей жизни Сергей Львович много сил отдавал преподаванию математики, совмещая эту деятельность с научной работой. В 1930-31гг. он был доцентом Ленинградского электротехнического института, с 1935 года работал профессором кафедры дифференциальных уравнений Московского государственного университета (МГУ), одновременно читая лекции в ЛГУ, в 1936-37гг. он заведовал кафедрой высшей математики в Военно-технической академии. На базе лекций, прочитанных в МГУ, С. Л. Соболев написал замечательный учебник "Уравнения математической физики".

Необходимо отметить большую роль С. Л. Соболева в воспитании молодого поколения ученых-математиков в МГУ. Он был руководителем многих известных научных семинаров, например, семинара, совместного с И. Г. Петровским и А. Н. Тихоновым, который действовал с 1935г. по 1958г. с перерывом только на годы войны. На этом семинаре выросли такие выдающиеся математики, как М. И. Вишик, О. А. Ладыженская, О. А. Олейник и многие другие.

С 1952 по 1959 год Сергей Львович заведовал кафедрой вычислительной математики на механико-математическом факультет МГУ. Ректор МГУ академик И. Г. Петровский, пригласивший его на эту должность, прекрасно понимал, какую огромную пользу принесет С. Л. Соболев делу воспитания так необходимых стране специалистов по вычислениям на ЭВМ. Ведь Сергей Львович, работая в институте атомных исследований, имел серьезный опыт использования ЭВМ при решении больших прикладных задач. И действительно, специалисты, подготовленные на кафедре вычислительной математики МГУ, в скором времени обеспечили успешную работу многих организаций в СССР, занимавшихся применением вычислительной математики и вычислительной техники.

Работа над атомным проектом и использование ЭВМ привели к изменению математических интересов Сергея Львовича: начиная с 50-х годов, он много внимания уделяет вопросам вычислительной математики, используя для их решения методы задач математической физики, теории функций и функционального анализа. Отметим, что использование в методах вычислений таких абстрактных разделов математики как, например, функциональный анализ, многим в то время казалось совершенно неожиданным и странным. Тем не менее, такое применение оказалось весьма полезным и сейчас используется повсеместно. Многие работы по вычислительной математике, написанные С. Л. Соболевым, породили новые существенные направления в теории численных методов. Следует особенно отметить его многолетнюю деятельность в области кубатурных формул (т.е. формул для приближенного вычисления интегралов). Эта деятельность была подытожена в его фундаментальной монографии "Введение в теорию кубатурных формул", опубликованной в 1974 году.

Сибирский период

В середине 50-х годов в СССР проводилась большая работа по освоению Сибири и Дальнего Востока: разворачивалось строительство новых промышленных предприятий и городов. Но необходимо было развивать и науку: в то время на 30 миллионов населения, живущего восточнее Урала, приходился всего один профессор математики — в Томске. В конце 1956 года академики М. А. Лаврентьев, С. Л. Соболев и С. А. Христианович обратились к правительству с предложением разработать план создания научных центров на востоке нашей страны. ЦК КПСС и Советское правительство одобрили это предложение, а Президиум АН СССР принял постановление о создании Сибирского отделения АН СССР. С. Л. Соболев был назначен директором Института математики СО АН ССССР.

В 1958 году С. Л. Соболев переезжает на постоянную работу в г. Новосибирск. На вопрос, что побудило его покинуть сильную кафедру в ведущем университете страны, поменяв ее на научную целину, Сергей Львович со свойственной ему скромностью отвечал: "Естественное желание человека прожить несколько жизней, начать что-то новое". На самом деле это было прежде всего проявление высокого патриотизма. Он считал очень важным развитие науки в Сибири, а его личный пример в сочетании с огромным авторитетом помогли привлечь столь необходимые научные кадры. Ведь за М. А. Лаврентьевым, С. Л. Соболевым, С. А. Христиановичем в Сибирь последовали многие ученые.

Сергей Львович проводит большую организационную работу в Сибирском Отделении АН

СССР, начиная с его формирования. Он — один из основателей Новосибирского государственного университета, призванного воспитывать научные кадры для Сибири и Дальнего Востока. Трудно переоценить роль С. Л. Соболева в формировании института математики СО АН СССР. Больших успехов Институт достиг в развитии алгебры и логики, геометрии и теории вероятностей, теории функций, дифференциальных уравнений и вычислительной математики. Значительные усилия Сергей Львович приложил для развития в институте таких прикладных разделов науки, как кибернетика и математическая экономика. В 50-е годы кибернетика официально считалась в СССР буржуазной лженаукой (хотя, конечно, ЭВМ использовались при реализации атомного проекта и в других отраслях военно-промышленного комплекса). Прекрасно понимая важность кибернетики, С. Л. Соболев много сделал для ее официальной реабилитации. В частности кибернетика стала одним из ведущих направлений в Институте математики. Аналогичная ситуация была и с математической экономикой. Эту новую область науки далеко не всегда можно было легко согласовать с марксисткими догмами, и поэтому власти относились к ней весьма настороженно. Понимая всю важность математической экономики, Сергей Львович много сделал для ее развития в Институте, пригласив, в чатности, на работу в Институт основоположника этой науки в СССР академика Л. В. Канторовича. Впоследствии, за достигнутые выдающиеся успехи в этой области Л. В. Канторович был удостоен Ленинской и Нобелевской премий.

В 1984 году, оставив руководство институтом математики СО АН СССР, С. Л. Соболев переезжает в Москву. Умер Сергей Львович 3 января 1989 года.

3. Обобщенные функции

Конечно, рамки этой статьи не позволяют не только описать, но даже перечислить все результаты, полученные Сергеем Львовичем. Мы постараемся лишь чуть-чуть пояснить смыл и значение некоторых базисных определений теории обобщенных функций, — исключительно яркого и важного раздела математики, заложенного в работе С. Л. Соболева [3].

Что было до обобщенных функций

Один из крупнейших математиков XIX века Б. Риман в своей докторской диссертации (1851) при построении своей теории аналитических функций использовал следующий вариационный принцип — принцип Дирихле: Пусть Ω — ограниченная односвязная область[1] на плоскости с гладкой границей $\partial\Omega$, и на $\partial\Omega$ задана непрерывная функция f. Рассмотрим множество всех функций $u(x,y)$, $(x,y)\in\Omega$, непрерывных в Ω и равных f на границе $\partial\Omega$, для которых конечен интеграл

$$\int_\Omega \left(\frac{\partial u(x,y)}{\partial x}\right)^2 + \left(\frac{\partial u(x,y)}{\partial y}\right)^2\ dx\ dy \tag{1}$$

Тогда на некоторой функции $\widehat{u}(x,y)$из этого класса интеграл (1) достигает своего минимального значения. Принип Дирихле кажется весьма правдоподобным. Однако математического доказательства принципа Дирихле Риман не дал. Этот пробел заметил Вейерштрасс. Доказать принцип Дирихле оказалось вовсе не просто. Для этого потребовалось почти полвека и гений Д. Гильберта, который получил нужное доказательство только в 1899 году. Работая в этой области, Гильберт осознал её исключительную важность: среди 23 знаменитых проблем, поставленных Гильбертом на II-м Всемирном математическом конгрессе 1902 года, 3 проблемы (16-я, 18-я и 23-я) связаны с вариационным исчислением, т.е. с теорией минимизации интегралов, аналогичных (1). Д. Гильберт писал: “Я убежден, что будет возможно доказывать теоремы существования с помощью общего принципа... . Этот общий принцип, возможно приблизит нас к ответу на следующий вопрос: имеет ли решение каждая регулярная вариационная проблема, если самому понятию “решение” при случае придавать расширенное толкование.”

Недостаточность имеющихся математических средств ощущалась и в других областях, связанных с математической физикой. Мы здесь укажем лишь два таких случая.

[1]Т.е. область без “дырок”.

1) В газовой динамике и многих других разделах механики сплошных сред широко используется следующее уравнение в частных производных, которое называется уравнением Бюргерса:

$$\frac{\partial u(x,t)}{\partial t} + \frac{\partial}{\partial x}\left(\frac{u^2(x,t)}{2}\right) = 0, \qquad -\infty < x < \infty, \quad t > 0. \tag{2}$$

Чтобы решить это уравнение, мы должны задать начальное условие:

$$u(x,t)|_{t=0} = u_0(x). \tag{3}$$

Оказывается, что очень часто у решения задачи (2), (3) возникают разрывы[2], причем эти разрывы имеют ясный физический смысл: они соответствуют так называемым ударным волнам, иногда возникающим при движении газа. Уравнение Бюргерса как раз и было выведено физиками ради описания и исследования ударных волн[3]. Но возникновение разрывов порождает естественный математический вопрос: а как понимать дифференциальное уравнение в точках разрыва функции $u(x,t)$ — ведь в этих точках у решения $u(x,t)$ не могут существовать производные и по x, и по t?

2) В 20-х годах XX-го века бурно развивалась новая область физики: квантовая механика. В ней возникла очень странная функция — так называемая δ-функция Дирака. Она описывала вполне естественные физические объекты: например, плотность точечного заряда. С одной стороны, для любого не равного нулю числа x имеем $\delta(x) = 0$ (так как при $x \neq 0$ нет заряда). Кроме того, из самого определения плотности следует, что $\int_{-1}^{1} \delta(x)\, dx = 1$. Но эти два свойства противоречат друг другу: ведь хорошо известно, что если $\delta(x) = 0$ при $x \neq 0$, то $\int_{-1}^{1} \delta(x)\, dx = 0$ каковым бы не было значение $\delta(0)$, включая $\pm\infty$!

Трудностей подобного рода накопилось довольно много к 30-м годам прошлого века. Все они были благополучно разрешены с помощью теории обобщенных функций, основы которой были заложены С. Л. Соболевым. Постараемся дать представление об основных понятиях этой теории, рассматривая лишь функции на числовом отрезке $[-1,1]$.

Обобщенные функции: основные определения

Обозначим через $C_0^1[-1,1]$ множество (пространство) всех непрерывных функций $\varphi(x)$, заданных на отрезке $[-1,1]$, имеющих там непрерывную производную $\varphi'(x)$ и равных нулю на концах этого отрезка: $\varphi(-1) = \varphi(1) = 0$. Мы будем рассматривать функции, определенные на $C_0^1[-1,1]$, (которые, по определению, каждой функции $\varphi \in C_0^1[-1,1]$ ставят в соответствие некоторое число F). Такие "функции от функций" обычно называют функционалами. Если функционал $\langle F, \varphi\rangle$ является линейным, т.е.

$$\langle F, \alpha\varphi_1 + \beta\varphi_2\rangle = \alpha\langle F, \varphi_1\rangle + \beta\langle F, \varphi_2\rangle \qquad \forall \alpha, \beta \in \mathbb{R}, \quad \forall \varphi_1, \varphi_2 \in C_0^1[-1,1],$$

---

[2] Продемонстрируем это на примере. Возьмем в качестве начального условия

$$u_0(x) = \begin{cases} 1, & \text{при } x \le 0 \\ \max(1-x, 0), & \text{при } x > 0 \end{cases} \tag{4}$$

Рассмотрим функцию

$$u(x,t) = \begin{cases} 1, & \text{при} \quad x < t, \\ \frac{1-x}{1-t}, & \text{при} \quad t \le x \le 1 \\ 0, & \text{при} \quad x > \max(1,t) \end{cases}, \quad t \ge 0, \quad -\infty < x < \infty \tag{5}$$

Легко заметить, что у этой функции $u(x,t)$ имеется разрыв на луче $\{t = x, x \ge 1\}$. Подстановка функции $u(x,t)$ в уравнение (2) и простые вычисления производных вне линиии разрыва показывают, что она является решением уравнения (2) вне линии разрыва, а при $t = 0$ совпадает с функцией (4). (Проверьте!)

[3] Отметим, что если изменить в формуле (5) линию разрыва, на $\{t = f(x), x \ge 1\}$, где $f(x)$ — произвольная непрерывная монотонно растущая функция, а решение $u(x,t)$ оставить равным 1 левее и 0 правее разрыва, то конечно, полученная функция также будет удовлетворять уравнению (2) вне линии разрыва. Однако она уже к описываемому физическому процессу никакого отношения не имеет. "Правильную" линию разрыва физики определяют с помощью некоторого дополнительного условия: условия Гюгонио.

то такой функционал называется обобщенной функцией[4].

Приведем примеры обобщенных функций. Обозначим через $C[-1,1]$ множество всех непрерывных функций на $[-1,1]$, и по каждой функции $f(x) \in C[-1,1]$ построим обобщенную функцию F_f по формуле:

$$\langle F_f, \varphi\rangle = \int_{-1}^{1} f(x)\varphi(x)\ dx \qquad \forall \varphi \in C_0^1[-1,1] \tag{6}$$

Очевидно (6) — это линейный функционал на $C_0^1[-1,1]$. Оказывается, и это очень важно, что функция $f(x)$ определяется по значениям функционала (6) однозначно. Это означает, что

$$\text{если } f, g \in C[-1,1] \text{ и } \quad \langle F_f, \varphi\rangle = \langle F_g, \varphi\rangle\ \forall \varphi \in C_0^1[-1,1], \quad \text{то} \quad f(x) \equiv g(x). \tag{7}$$

Задача 1. Доказать утверждение (7).

Указание. Из (6), (7) следует, что

$$\langle F_f, \varphi\rangle - \langle F_g, \varphi\rangle = \int_{-1}^{1} (f(x) - g(x))\varphi(x)\ dx = 0 \quad \forall \varphi \in C_0^1[-1,1]. \tag{8}$$

Предположим, что $h(x) \equiv f(x) - g(x) \not\equiv 0$, т.е., например, $h(x_0) > 0$ для некоторого числа $x_0 \in [-1,1]$. В силу непрерывности функции $h(x)$, $h(x) > 0$ при $x \in (x_0 - \varepsilon, x_0 + \varepsilon)$ с некоторым $\varepsilon > 0$. Постройте $\tilde{\varphi} \in C_0^1[-1,1]$ так, чтобы $\tilde{\varphi} > 0$ при $x \in (x_0 - \varepsilon, x_0 + \varepsilon)$ и $\tilde{\varphi} = 0$ при $x \notin (x_0 - \varepsilon, x_0 + \varepsilon)$. Тогда $\int_{-1}^{1} h(x)\tilde{\varphi}(x)\ dx > 0$, что противоречит (8).

Обозначим множество всех обобщенных функций символом $D'(-1,1)$. Свойство (7) означает, что пространство непрерывных функций можно отождествить с некоторым подмножеством обобщенных функций, и в этом смысле справедливо включение

$$C[-1,1] \subset D'(-1,1). \tag{9}$$

Конечно, разрывные функции также могут являться обобщенными. Например, θ-функция Хевисайда

$$\theta(x) = \begin{cases} 0, & x < 0 \\ 1, & x > 0 \end{cases} \tag{10}$$

задает обобщенную функцию

$$\langle \theta, \varphi\rangle = \int_{0}^{1} \varphi(x)\ dx = 0 \quad \forall \varphi \in C_0^1[-1,1]. \tag{11}$$

Отметим, что имеются обобщенные функции, которые не являются обычными функциями. Например, упомянутая выше функция Дирака является обобщенной функцией и определяется формулой

$$\langle \delta, \varphi\rangle = \varphi(0) \quad \forall \varphi \in C_0^1[-1,1]. \tag{12}$$

Задача 2. Предполагая, что при $x \neq 0$ δ-функция Дирака (12) является обычной, причем непрерывной функцией, доказать, что $\delta(x) \equiv 0$ при $x \neq 0$.

Указание. Использовать указание к задаче 1.

Отметим, что в силу (9) для пространства непрерывно дифференцируемых функций $C^1[-1,1]$ справедливо включение $C^1[-1,1] \subset C[-1,1] \subset D'(-1,1)$. Воистину замечательным является тот факт, что операция взятия производной естественным образом распространяется с пространства $C^1[-1,1]$ на множество всех обобщенных функций. Объясним, как это делается. Хорошо известна формула интегрирования по частям:

$$\int_{-1}^{1} f'(x)g(x)\ dx = f(1)g(1) - f(-1)g(-1) - \int_{-1}^{1} f(x)g'(x)\ dx \quad f, g \in C^1[-1,1],$$

[4]Здесь, как обычно, символ $\forall$ обозначает "для любых", символ $\mathbb{R}$ означает множество вещественных чисел. В определении обобщенной функции нужно также требовать ограниченности функционала, да и сам функционал F естественнее было бы рассматривать не на $C_0^1[-1,1]$, а на пространстве бесконечно дифференцируемых функций $C_0^\infty[-1,1]$, но мы не хотим здесь вдаваться в такие подробности.

где $f'(x), g'(x)$– производные, соответственно, функций $f(x), g(x)$. Так как для $\varphi(x) \in C_0^1[-1,1]$, $\varphi(-1) = \varphi(1) = 0$, из этой формулы следует, что

$$\int_{-1}^{1} f'(x)\varphi(x)\ dx = -\int_{-1}^{1} f(x)\varphi'(x)\ dx \quad \forall f \in C^1[-1,1],\ \ \varphi \in C_0^1[-1,1]. \tag{13}$$

Основываясь на формуле (13), дадим следующее определение обобщенной производной F' для обобщенной функции F:

$$\langle F', \varphi \rangle = -\langle F, \varphi' \rangle \quad \forall \varphi \in C_0^1[-1,1]. \tag{14}$$

Если F задается функцией $f \in C^1[-1,1]$, т.е. если $F = F_f$, то формула (14) совпадает с (13), а это значит, что обобщенная производная функции $f \in C^1[-1,1]$ совпадает с классической производной. Но формула (14) справедлива и для обобщенных функций, где f — произвольная непрерывная функция. Следовательно, у любой непрерывной функции существует обобщенная производная.

Задача 3. Используя формулы (13), (14), вычислить обобщенную производную непрерывной функции $f(x) = x_+ \equiv \max(0,x)$. (Ответ: $x'_+ = \theta(x)$.)

Более того, обобщенная производная существует и у разрывных функций.

Задача 4. Используя (14), доказать, что $\theta'(x) = \delta(x)$, где $\theta(x)$ – функция Хевисайда (10), (11), а $\delta(x)$ — δ-функция Дирака (12).

Обобщенное решение

С использованием изложенных выше идей С. Л. Соболева становится ясным, как правильно определить решение уравнения в частных производных, допускающее разрывы. Покажем это на примере уравнения Бюргерса[5].

Пусть $C(\mathbb{R}^2_+)$ — множество всех непрерывных функций на полупространстве $\mathbb{R}^2_+ = \{(x,t) : -\infty < x < \infty, t > 0\}$. Рассмотрим пространство функций

$$C_0^1(\mathbb{R}^2_+) = \{\varphi(x,t) \in C(\mathbb{R}^2_+),\ \frac{\partial \varphi}{\partial x},\ \frac{\partial \varphi}{\partial t} \in C(\mathbb{R}^2_+),\ \varphi(x,t) \equiv 0 \text{ при } |x| > N(\varphi),\ \ t > N(\varphi)\}$$

где $N(\varphi) > 0$ — константа, своя у каждой функции $\varphi \in C_0^1(\mathbb{R}^2_+)$.

Пусть $u(x,t)$ — классическое решение уравнения (2), т.е. функция $u(x,t)$ непрерывна вместе с производными $\partial u/\partial x, \partial u/\partial t$, и уравнение (2) выполнено при всех $(x,t) \in \mathbb{R}^2_+$. Взяв $\varphi \in C_0^1(\mathbb{R}^2_+)$, применим к паре $\varphi(x,t)$, $\partial u(x,t)/\partial t$ при фиксированном x формулу интегрирования по частям по переменной t, а затем полученное равенство проинтегрируем по x. Учитывая, что $\varphi(x,t) = 0$ при $t > N(\varphi)$, получим в результате формулу

$$\int_{-\infty}^{\infty}\int_0^{\infty} \frac{\partial u(x,t)}{\partial t}\varphi(x,t)\ dtdx = -\int_{-\infty}^{\infty} u(x,0)\varphi(x,0)\ dx - \int_{-\infty}^{\infty}\int_0^{\infty} u(x,t)\frac{\partial \varphi(x,t)}{\partial t}\ dtdx \tag{15}$$

Аналогично, интегрируя по t формулу интегрирования по частям по переменной x получим:

$$\int_0^{\infty}\int_{-\infty}^{\infty} \frac{\partial}{\partial x}\left(\frac{u^2(x,t)}{2}\right)\varphi(x,t)\ dxdt = -\int_0^{\infty}\int_{-\infty}^{\infty} \frac{u^2(x,t)}{2}\frac{\partial \varphi(x,t)}{\partial t}\ dxdt \tag{16}$$

Итак, пусть $u(x,t)$ – классическое решение уравнения (2), удовлетворяющее начальному условию (3). Умножим обе части уравнения (2) на $\varphi(x,t) \in C_0^1(\mathbb{R}^2_+)$, проинтегрируем полученное равенство по $\mathbb{R}^2_+$ и воспользуемся формулами (15), (16). В результате получим тождество:

$$\int_0^{\infty}\int_{-\infty}^{\infty} \left(u(x,t)\frac{\partial \varphi(x,t)}{\partial t} + \frac{u^2(x,t)}{2}\frac{\partial \varphi(x,t)}{\partial x}\right)\ dxdt = \int_{-\infty}^{\infty} u(x,0)\varphi(x,0)\ dx \tag{17}$$

Но формула (17) имеет смысл и для функции $u(x,t)$, имеющей разрывы. Это позволяет дать следующее определение решения задачи Коши (2), (3), которое имеет смысл и для разрывных решений:

[5]Сам С. Л. Соболев ввел понятие обобщенного решения в случае волнового уравнения.

Функция $u(x,t)$, определенная и интегрируемая при $(x,t) \in \mathbb{R}^2_+$ называется обобщенным решением задачи Коши (2), (3), если при любой функции $\varphi(x,t) \in C^1_0(\mathbb{R}^2_+)$ она удовлетворяет интегральному тождеству (17).

Отметим, что если у обобщенного решения уравнения Бюргерса возникает разрыв, то линия разрыва однозначно определяется начальным условием $u_0(x)$ из (3), причем выделяется именно то решение, которое правильно описывает физический процесс.

Пространство Соболева

Теперь совсем нетрудно объяснить, что такое пространство Соболева. Существует много разных пространств Соболева; мы определим простейшее из них. Через $L^2(-1,1)$ обозначим множество таких функций $f(x)$, $x \in (-1,1)$ что $\int_{-1}^1 |f(x)|^2\, dx < \infty$. Из формулы (14) сразу следует, что обобщенная производная существует у любой функции $f(x) \in L^2(-1,1)$. Теперь определим пространство Соболева $H^1(-1,1)$ как множество таких функций $f(x) \in L^2(-1,1)$, у которых обобщенная производная $f'(x)$ также принадлежит $L^2(-1,1)$.

Введем класс функций $u(x) \in H^1(-1,1)$, таких что $u(-1) = u_{-1}$, $u(1) = u_1$, где u_{-1}, u_1 — заданные числа. Если на этом классе рассмотреть задачу минимизации интеграла

$$\int_{-1}^{1} (u'(x))^2\, dx \tag{18}$$

то доказательство существование функции $\widehat{u}(x) \in H^1(-1,1)$, на которой достигается минимум, оказывается совсем простым. Дело в том, что само определение пространства Соболева органично связано с интегралом (18).

Интеграл (18) является одномерным аналогом интеграла Дирихле (1). Конечно, можно определить пространство Соболева и для функций двух переменных, и доказать принцип Дирихле в этом пространстве. Именно пространства Соболева позволили дать "расширенное толкование понятия решения", которое предрекал Д. Гильберт на рубеже XIX и XX веков. Хорошо известно, что решение вариационной задачи удовлетворяет связанному с ней эллиптическому дифференциальному уравнению. Этим и объясняется исключительная важность пространств Соболева в теории уравнений в частных производных. (Сам Сергей Львович пришел к понятию пространства Соболева как раз исследуя в конце 30-х годов некоторые эллиптические дифференциальные уравнения математической физики.) Остается добавить, что процесс проникновения пространств Соболева в теорию уравнений математической физики был непростым: слишком революционным оказалось это понятие. В 40-х – 50-х годах XX века пространства Соболева использовали в своих исследованиях лишь единицы: в Советском Союзе это были в первую очередь М. И. Вишик и О. А. Ладыженская. Однако в 60-е годы произошел некоторый качественный скачок, и с тех пор пространства Соболева используются практически во всех исследованиях по уравнениям в частных производных и задачам математической физики.

Литература

[1] Н. С. Бахвалов и др. Сергей Львович Соболев (к восьмидесятилетию со дня рождения). — Успехи Матем. Наук т. 43, вып 5, (1988), с.2-15.

[2] В. М. Писаревский, В. Т. Харин. Беседы о математике и математиках. М.: Физматлит, 2006.

[3] С. Л. Соболев. Methode nouvelle á resoudre le probléme de Cauchy pour les equation linéaries hyperbolic normales, — Мат. сб. т.1(43), 1936, с.39-72. (Русск. перевод: см. Дополнение в [4]).

[4] С. Л. Соболев. Некоторые применения функционального анализа в математической физике (изд. третье), М.: Наука, 1988.

Фурсиков Андрей Владимирович,
доктор физ.-мат. наук,
профессор кафедры общих проблем управления
механико-математического факультета МГУ.

Наглядное объяснение предела

С. А. Кулешов

В статье подробно, на доступном для старшеклассников уровне разъясняются основные конструкции, ведущие к глубокому неформальному освоению понятий предела последовательности и предела функции в точке.

Введение

Понятие предела — одно из самых трудных для освоения, как в школе, так и на первом курсе ВУЗов. К сожалению, в школе зачастую «смазывают» все трудности этого определения и, когда на лекции спрашиваешь студентов об их представлении о пределе бесконечной последовательности, слышишь довольно странные высказывания. Кроме того, в связи с большим объёмом информации, который необходимо лектору изложить в течение семестрового курса математического анализа, на тщательное изучение определений в ВУЗах просто не хватает времени. Но не уяснив себе досконально, о чём, собственно, идёт речь, крайне трудно качественно усвоить предмет.

Хотелось бы как-то поправить дело не только у студентов, которым я преподаю, но и у тех, кому вопросы понимания не безразличны.

Отмечу ещё, что цель данной статьи — лишь разъяснение определений, после которого, я надеюсь, читатель сможет с пониманием читать формальное изложение теории пределов последовательностей и функций в любом стандартном учебнике.

Точная граница

На мой взгляд, следует начать не с предела, а с понятия границы множества. Здесь и далее, рассуждая о множествах, мы будем считать, что имеем дело с подмножествами числовой прямой.

Определение. *Говорят, что множество M ограничено сверху, если найдётся такое число C, что для любого элемента $m \in M$ выполнено неравенство: $m \leq C$. При этом само число C называют верхней границей множества M.*

В этом определении нет ничего загадочного. Легко представить себе ограниченные сверху множества, например, отрицательные числа, числа отрезка $[0;1]$, и т. д. А вот множество рациональных или даже целых чисел не ограничено сверху, поскольку нельзя найти ни одного числа, которое было бы больше любого целого.

Однако, несмотря на свою простоту, понятие ограниченности множества требует некоторого уточнения. Дело в том, что верхняя граница множества определена не однозначно, как, очевидно, мог заметить внимательный читатель. Действительно, рассмотрим множество $M = [0;1]$ чисел, расположенных между 0 и 1 (включая 0 и 1). Согласно определению, это множество ограничено сверху и в качестве верхней границы естественно выбрать $C = 1$. Но если мы вместо $C = 1$ возьмём $C = 2$ или 237, то всё равно будем иметь неравенство: $m \leq C$ для всех $m \in [0;1]$.

Как вы понимаете, в математике очень уважается точность. И если неопределённость куда-то закралась, с ней пытаются разобраться.

Итак, перед нами стоит задача об определении *точной* верхней границы множества, которое бы не допускало нескольких значений такой границы.

Если аналогичный вопрос поставить на лекции, то в большинстве случаев можно услышать, что как точную верхнюю границу множества нужно брать его максимальный элемент. В частности, у отрезка $[0; 1]$ таким элементом является 1. Всё было бы замечательно, если бы у каждого ограниченного множества этот максимальный элемент был. Но, как легко догадаться, существуют ограниченные множества и без оного, например интервал $(0; 1)$, который отличается от аналогичного отрезка отсутствием концов. Конечно, в этой ситуации 1 тоже хочется назвать точной верхней границей множества, однако, она не является его максимальным элементом. Поэтому надо менять определение.

Как же нащупать правильное определение? Попробуем уяснить себе, а что мы собственно хотим от этого понятия. Во-первых, ясно, что справа от точной границы элементов множества быть не должно. Во-вторых, точная граница должна быть определена *единственным* образом, в частности, никакое число меньше неё или больше, не должно подходить под определение.

Помня об этих замечаниях, можно осознать, что точная верхняя граница должна быть и просто верхней границей (справа от неё нет элементов множества), а чтобы никакое число, меньшее точной границы, не подпадало по определение, потребуем, чтобы точная верхняя граница была *минимальной* из всех верхних границ данного множества. А минимальный элемент множества (если существует, конечно) уж точно только один.

Определение. *Точной верхней границей ограниченного сверху множества M называется его минимальная верхняя граница.*

Скептик, однако, возразит, что минимальной верхней границы может и не оказаться. Ведь мы знаем примеры ограниченных снизу множеств без минимального элемента. Он безусловно прав! Существование точной верхней границы, определение которой мы только что придумали, нельзя принимать на веру, а необходимо доказывать, как, собственно, любое научное утверждение.

К сожалению, на базе школьных представлений о вещественных числах сделать этого нельзя. Все, уже начиная с 5-го класса, представляют себе, что числовая прямая непрерывна. Иными словами, любая точка числовой прямой — это какое-то вещественное число. Но для строгих рассуждений такого наивного представления о непрерывности недостаточно. Поэтому сформулируем одну из аксиом вещественных чисел.

Принцип непрерывности. *Пусть непустые подмножества A и B числовой прямой удовлетворяют следующему свойству: $a \le b$ для любых элементов $a \in A$ и $b \in B$. Тогда найдётся такое число c_0, что $a \le c_0 \le b$ для всех элементов $a \in A$ и $b \in B$.*

Теперь, вооружась подходящим инструментом, докажем существование и единственность точной верхней границы.

Предложение. *У любого ограниченного сверху непустого множества M существует точная верхняя граница, причём только одна.*

Доказательство. Рассмотрим множество M и множество C всех его верхних границ. По определению верхней границы множества, $m \le c$ для любого элемента $m \in M$ и любой верхней границы $c \in C$. Значит, выбранные множества удовлетворяют принципу непрерывности и, как следствие, существует такое число c_0, для которого $m \le c_0 \le c$ при всех $m \in M$ и $c \in C$.

Левое неравенство ($m \le c_0$) позволяет заключить, что c_0 — верхняя граница множества M, а правое ($c_0 \le c$) говорит о том, что любая другая верхняя граница не меньше c_0. Иными словами, c_0 — минимальная верхняя граница множества M, т. е. его точная верхняя граница.

Итак, мы доказали, что у любого ограниченного сверху множества M действительно существует точная верхняя граница, которая обычно обозначается через $\sup M$. Эта граница обладает очень важным и полезным свойством, которое делает её похожей на границу в обыденном понимании этого слова. А именно, справа от $\sup M$ нет ни одной точки множества M, а слева, сколь угодно близко к $\sup M$ найдутся точки этого множества. Обратите внимание, мы не можем утверждать, что любая точка слева от $\sup M$ является точкой множества.

Для пояснения этого утверждения рассмотрим пример. Пусть

$$M = \left\{ \frac{1}{2}, \frac{2}{3}, \frac{3}{4}, \ldots, \frac{n-1}{n}, \ldots \right\}.$$

Рис. 1.

Покажем, что $\sup M = 1$. Очевидно, что любой элемент $m = \frac{n-1}{n} \in M$ меньше 1. Так что 1 — верхняя граница M. Покажем, что 1 — минимальная верхняя граница. Действительно, рассмотрим число $c < 1$ и докажем, что оно не может служить верхней границей множества M. Обозначим через ε разность $1 - c$, которая положительна в силу выбора c (рис. 1). Расстояние от элемента $\frac{n-1}{n}$ до 1 равно $\frac{1}{n}$. Поэтому, когда

$$\frac{1}{n} < \varepsilon \Leftrightarrow \frac{1}{\varepsilon} < n,$$

элемент $\frac{n-1}{n}$ перейдёт выбранную «границу» c. Итак, все элементы $\frac{n-1}{n}$ с $n > \frac{1}{\varepsilon}$ будут лежать правее c (но левее 1). В частности, c не является верхней границей, а поэтому $1 = \sup M$.

На этом примере мы видим, что левее точной верхней границы точки множества безусловно есть, но далеко не любая точка левее sup будет принадлежать множеству.

Прежде чем перейти к точной формулировке упомянутого свойства точных границ, обсудим понятие «сколь угодно близко». Обиходные понятия «близко – далеко», «мало – много» весьма относительны. Действительно, два — это много или мало? Два волоса на голове — это мало, а в супе — …

Как уже отмечалось, в математике принято избегать нечётких, относительных, понятий. Когда мы говорим «сколь угодно близко», имеется ввиду — на «сколь угодно малом расстоянии». Этот термин, конечно, пока тоже довольно туманен, но его легче формализовать. Действительно, расстояние между точками на числовой прямой — это, как известно, модуль разности соответствующих чисел. Поэтому «сколь угодно малое расстояние» означает *любое* положительное число. При этом мы не можем ограничиться числами, малыми в житейском понимании, например, 10^{-10} мм. Мы должны помнить, что какое бы малое на наш взгляд число мы не взяли, всегда найдётся ещё меньшее (достаточно выбранное число поделить пополам).

Есть удобный способ в математике, позволяющий записывать термин «сколь угодно близко» на формульном уровне:

$$\forall \varepsilon > 0 \quad \exists m \in M: \qquad \sup M - \varepsilon < m < \sup M.$$

Читается эта символьная запись так: для любого положительного ε (знак $\forall$ означает «для любого», или «произвольный») существует элемент m множества M (знак $\exists$ означает «найдётся», или «существует»), такой что (двоеточие) выполнено неравенство…

Не нужно бояться символьных выражений! После некоторого привыкания формулы только помогают воспринимать информацию, что легко показать на следующем тривиальном примере. Сравните одно и то же высказывание, записанное словами и с помощью формулы:

> Два минус дробь, числитель которой равен разности десяти и квадрата трёх, а знаменатель — пять плюс четыре минус восемь, равно единице;

$$2 - \frac{10 - 3^2}{5 + 4 - 8} = 1.$$

Дальнейшие комментарии излишни…

Вернёмся к важному свойству точных верхних границ.

Определяющее свойство точной верхней границы. *Число c_0 является точной верхней границей множества M тогда и только тогда, когда*

а) $\forall m \in M \ \ m \le c_0$;

б) $\forall \varepsilon > 0 \quad \exists m \in M: \qquad c_0 - \varepsilon < m.$

Доказательство. Обратите внимание, что в утверждении, которое мы собираемся доказать, есть слова: «тогда и только тогда». Поэтому в нём фактически содержатся 2 высказывания:

1) если $c_0 = \sup M$, то выполнены свойства а) и б);

2) если для числа c_0 и множества M имеют место свойства а) и б), то $c_0 = \sup M$.

Начнём с первого. Пусть $c_0 = \sup M$. По определению это означает, что c_0 — минимальная верхняя граница множества M. Поэтому свойство а) выполнено, так как c_0 — верхняя граница множества M. Если бы свойство б) нарушалось, т. е. для некоторого $\varepsilon > 0$ все элементы $m \in M$ удовлетворяли бы неравенству $m < c_0 - \varepsilon$, то $c_0 - \varepsilon$ была бы тоже верхней границей множества M, что противоречит минимальности c_0 среди всех верхних границ. Первое утверждение доказано.

Переходим ко второму. Пусть для числа c_0 и множества M выполнены свойства а) и б). Из свойства а) следует, что c_0 — верхняя граница множества M. Свойство б), с другой стороны, говорит о том, что никакое меньшее число $c_0 - \varepsilon$ не может являться верхней границей M. Таким образом, c_0 — минимальная верхняя граница. Второе утверждение тоже доказано.

Чтобы лучше разобраться с понятием точной верхней границы и усвоить материал, очень полезно выполнить следующие простые упражнения.

1. Найдите $\sup M$, если

а) M — это множество рациональных чисел q, лежащих на отрезке $[1; 3]$;

б) M — это множество рациональных чисел q, лежащих на отрезке $[1; 3]$, знаменатель которых не превышает 10;

в) $M = \{\frac{1}{2}, 1 - \frac{1}{4}, 1 - \frac{1}{8}, \ldots, 1 - \frac{1}{2^n}, \ldots\}$.

2. Сформулируйте определение ограниченного снизу множества и его нижней границы.

3. Сформулируйте определение точной нижней границы множества M (она обозначается через $\inf M$).

4. Сформулируйте и докажите определяющее свойство точных нижних границ, как это было сделано для верхних.

5. Найдите точную нижнюю границу множества $M = \{1, \frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \ldots, \frac{1}{2^n}, \ldots\}$.

Предел последовательности

Начать, естественно, нужно с определения последовательности.

Определение. *Бесконечной последовательностью называется бесконечный упорядоченный набор чисел:* $a_1, a_2, a_3, \ldots, a_n, \ldots$

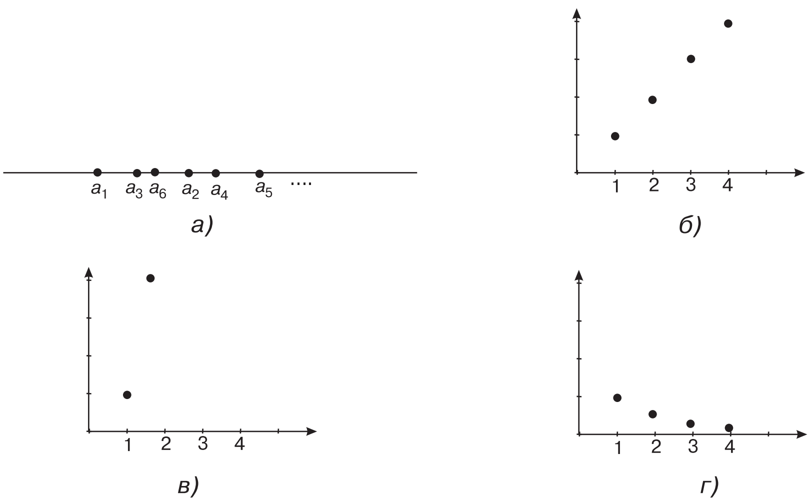

Рис. 2. Примеры последовательностей: *а)* пронумерованные точки числовой прямой; *б)* $a_n = n$; *в)* $a_n = n^2$; *г)* $a_n = 1/n$

Иногда последовательность удобно представлять себе как занумерованный набор точек на числовой прямой (рис. 2, *а*), а иногда — как функцию $f : \mathbb{N} \to \mathbb{R}$, определённую на множестве натуральных чисел $\mathbb{N}$ со значениями в множестве вещественных чисел $\mathbb{R}$. Эта функция сопоставляет каждому натуральному числу n вещественное число $f(n)$, которое мы ранее обозначали как a_n (рис. 2).

Представляя себе последовательность как пронумерованные точки на числовой прямой, легко прийти к выводу, что если все точки бесконечной последовательности лежат на каком-то отрезке, то они должны скапливаться возле какой-то точки. Если такая точка «накопления» только одна, то её и называют пределом последовательности. Конечно, это не может служить хорошим определением предела, поскольку слово «накапливаться» пока не имеет точного математического смысла. Поэтому попытаемся выработать более чёткое определение, как мы это сделали в случае точных границ.

Расхожие представления о пределе последовательности у школьников и первокурсников сводится к следующим двум высказываниям:

- Число является пределом последовательности, если последовательность *стремится* к этому числу.
- Последовательность *приближается* к пределу, но никогда его *не достигает*.

К сожалению, первое высказывание ничего не проясняет, поскольку совершенно не понятно, что означает слово «стремится». Второе же просто неверно. Действительно, если взять последовательность 0,9; 0,99; 0,999; ...; 0,99...9; ..., то к числам, к которым она приближается и никогда не достигает, можно отнести и 1 и 1,1 и вообще любое число, большее 1. Однако пределом данной последовательности может быть только одно число! Кроме того, предел последовательности 1, 1, 1, ..., 1, ..., все члены которой равны 1, очевидно равен 1. Поэтому утверждать, что последовательность «никогда не достигает предела» попросту нельзя!

С другой стороны, предел некоторых последовательностей можно назвать не задумываясь, причём абсолютно правильно. Например, предел последовательности 0,9; 0,99; 0,999; ...; 0,99...9; ... равен 1; последовательности $\frac{1}{2}$, $\frac{3}{2}$, $\frac{5}{3}$, ..., $\frac{2n-1}{n}$, ... равен 2. Такие последовательности обладают следующими полезными свойствами: во-первых, они возрастают, т. е. каждый следующий член не меньше предыдущего (на языке формул это выглядит так: $a_{n+1} \geq a_n$); во-вторых, множества элементов этих последовательностей ограничены сверху. Такие последовательности принято называть *ограниченными сверху возрастающими последовательностями.*

В обоих случаях пределом последовательности служит точная верхняя граница множества её элементов. В связи с этим разумно дать такое определение:

Определение. *Пределом ограниченной сверху возрастающей последовательности a_n называется точная верхняя граница множества её элементов.*

Хорошее, конечно, определение, но им можно пользоваться далеко не в любом случае. Поэтому на примере ограниченных возрастающих последовательностей надо исследовать дополнительные свойства предела и попытаться нащупать общее определение.

Итак, пусть a_n — возрастающая ограниченная сверху последовательность и A — точная верхняя граница её элементов, т. е. предел этой последовательности (рис. 3).

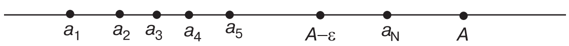

Рис. 3

Если взять какое-то число, лежащее левее A, например $A-\varepsilon$, где $\varepsilon > 0$, то по свойству точной верхней границы найдётся элемент последовательности a_N, удовлетворяющий неравенству:

$$A - \varepsilon < a_N \leq A.$$

Кроме того, мы имеем дело с возрастающей последовательностью. Поэтому при $m > N$ получаем

$$A - \varepsilon < a_N \leq a_m \leq A.$$

Неравенство $a_m \leq A$ следует из того, что A — точная верхняя граница множества всех элементов последовательности.

Обратите внимание, *все* элементы последовательности, номер которых больше или равен N (а таких бесконечно много) лежат между $A - \varepsilon$ и A, а вне этого промежутка могут находиться только те элементы, номер которых меньше N (т. е. конечное число). Кроме того, отступ ε, хоть и положителен, может быть сколь угодно малым. Таким образом, в сколь угодно малой окрестности точки A (окрестностью точки принято называть любой интервал, её содержащий) находится бесконечно много элементов последовательности, а вне её — конечное. Именно это свойство предела и можно взять в качестве определения.

Определение. *Говорят, что предел последовательности a_n равен A ($\lim\limits_{n\to\infty} a_n = A$), если в любую окрестность точки A попадает бесконечно много элементов последовательности, а вне этой окрестности остаётся лишь конечное число элементов.*

Смотрите, как замечательно получилось! Поскольку речь идёт о любой окрестности, т. е. интервале произвольной длины, мы можем выбрать такой интервал, накрывающий точку A, что визуально он просто сольётся с этой точкой (например, его длина может быть равной 0,001 мм). Тем не менее *почти вся*[1] последовательность окажется внутри этого интервала. Иными словами, если отбросить *незначительное* (по сравнению с бесконечностью) количество элементов последовательности, её с некоторой натяжкой можно просто отождествить с предельной точкой, — приём, который успешно применяют физики, отождествляя тело с материальной точкой.

Вот теперь высказывание «последовательность a_n стремится к A» (что можно обозначить как $a_n \to A$) приобретает вполне определённое толкование: $\lim\limits_{n\to\infty} a_n = A$. Кроме того, про такую последовательность принято говорить, что она *сходится* к A.

Посмотрим, а насколько удобно найденное определение, т. е. можно ли с его помощью что-нибудь доказать. Одно из первых и важных (хоть и элементарных) свойств предела — это однозначность:

Единственность предела. *Если последовательность a_n имеет предел, то этот предел определён однозначно. Иными словами, если $\lim\limits_{n\to\infty} a_n = A$ и $\lim\limits_{n\to\infty} a_n = B$, то $A = B$.*

Вообще-то такого сорта утверждения кажутся неопытному читателю казуистикой. Казалось бы, а зачем они нужны? Всё дело в том, что такая единственность служит проверкой корректности определения, поскольку, если найдётся последовательность, для которой по крайней мере две точки будут удовлетворять нашему определению предела, то таким определением будет невозможно пользоваться и придётся изобретать другое. Перейдём к доказательству утверждения.

Доказательство. Итак, предположим, что $\lim\limits_{n\to\infty} a_n = A$ и $\lim\limits_{n\to\infty} a_n = B$. Без ограничения общности можно считать, что $B > A$ (рис. 4). Выберем окрестности точек A и B так, чтобы они не пересекались, как показано на рисунке.

Тогда по определению предела в окрестности точки A должно находиться бесконечное число элементов последовательности, а вне её — конечное. В частности, в окрестность точки B может попасть только конечное число элементов последовательности. Но, с другой стороны, по тому же самому определению внутри окрестности точки B тоже должно находиться бесконечно много членов последовательности. Полученное противоречие и доказывает утверждение.

Рис. 4.

Итак, мы убедились в жизнестойкости нашего определения. К сожалению, что-то ещё интересное непосредственно с помощью этого определения доказать довольно трудно, хотя и небезнадёжно. Попытайтесь, например, доказать лемму о двух милиционерах.

[1] Термин «почти вся» на математическом жаргоне в данном контексте означает «вся за исключением конечного числа».

Лемма о двух милиционерах. *Если все члены трёх последовательностей a_n, b_n и c_n удовлетворяют неравенству*

$$a_n \le b_n \le c_n,$$

причём

$$\lim_{n\to\infty} a_n = \lim_{n\to\infty} c_n = A,$$

то $\lim\limits_{n\to\infty} b_n$ тоже существует и равен A.

Язык «эпсилон – дельта»

Как уже отмечалось, пользуясь найденным определением предела последовательности, доказать какие-то полезные свойства довольно сложно. Связано это с тем, что оперировать понятиями «конечный» и «бесконечный» без должного опыта практически невозможно. Для работы с понятием предела нужны какие-то формулы, поскольку именно они служат очень удобным инструментом в математике. В математическом анализе такие формулы называют языком ε–δ.

Попытаемся проанализировать наше определение и вывести некоторые соотношения. Итак, пусть $\lim\limits_{n\to\infty} a_n = A$. Согласно определению, в сколь угодно малом интервале, содержащем точку A, найдётся бесконечно много элементов последовательности, а вне него — конечное.

Прежде всего,

сколь угодно малый интервал, содержащий точку A,

можно записать как

$$\forall\varepsilon > 0 \qquad (A-\varepsilon, A+\varepsilon).$$

Проанализируем высказывание:

> внутри интервала $(A-\varepsilon, A+\varepsilon)$ найдётся бесконечно много элементов последовательности, а вне него — конечное.

Раз вне указанного интервала лежит конечное число элементов последовательности, то номер таких элементов не может быть очень большим. В частности, можно указать такой номер N, что любой элемент a_n с номером $n > N$ принадлежит выбранному интервалу: $a_n \in (A-\varepsilon; A+\varepsilon)$. Последнее легко записать неравенством:

$$A-\varepsilon < a_n < A+\varepsilon.$$

Собрав все формулы вместе, получим стандартное определение предела на языке ε–δ:

Определение. $\lim\limits_{n\to\infty} a_n = A$, *если* $\forall\varepsilon > 0\ \exists N\colon \forall n > N$ *выполнено неравенство:*

$$A-\varepsilon < a_n < A+\varepsilon.$$

Большинство читателей такое определение, конечно, повергнет в ступор: уж больно много здесь всяких значков и мало слов. Но я уже говорил, что после некоторой практики символы воспринимаются быстрее и точнее, чем громоздкие словесные высказывания. Поэтому очень полезно перевести новое определение на человеческий язык. И вы можете это сделать, если вспомните значение символов $\forall$ и $\exists$ (стр. 18).

Я же со своей стороны приведу краткую словесную формулировку:

Определение. *Число A называется пределом последовательности a_n, если для любого положительного числа ε все элементы последовательности, начиная с некоторого номера N попадают в интервал $(A-\varepsilon, A+\varepsilon)$.*

Пример. Посмотрим, как работает формализованное определение на простом примере. Рассмотрим последовательность $a_n = \frac{1}{n}$. Ни у кого не вызывает сомнения, что $\lim\limits_{n\to\infty} \frac{1}{n} = 0$. Докажем это, опираясь на определение. Возьмём произвольное $\varepsilon > 0$ и зафиксируем его. Если нам удастся найти такой номер N, что для всех $n > N$ выполнено неравенство

$$0 - \varepsilon < \frac{1}{n} < 0 + \varepsilon,$$

то мы докажем требуемое.

Заметим, что левая часть неравенства выполнена всегда, так как $\frac{1}{n} > 0$, а $-\varepsilon = 0 - \varepsilon < 0$. Значит, нам достаточно решить неравенство: $\frac{1}{n} < \varepsilon$, что равносильно $\frac{1}{\varepsilon} < n$. Таким образом, если мы выберем $N > \frac{1}{\varepsilon}$ (а такой номер, конечно, существует), то для всех $n > N > \frac{1}{\varepsilon}$ нужное неравенство будет выполнено, что и требовалось доказать.

Пример. Вычислим $\lim\limits_{n\to\infty}\left(1 + \frac{(-1)^n}{n}\right)$. Если взглянуть на несколько первых членов этой последовательности

$$1 - 1; 1 + \frac{1}{2}; 1 - \frac{1}{3}; 1 + \frac{1}{4}; 1 - \frac{1}{5}; \ldots,$$

то можно заметить, что они колеблются около 1, причём амплитуда колебаний неуклонно уменьшается. Поэтому естественно предположить, что искомый предел равен 1. Для доказательства этой гипотезы нужно зафиксировать произвольное $\varepsilon > 0$ и решить двойное неравенство

$$1 - \varepsilon < 1 + \frac{(-1)^n}{n} < 1 + \varepsilon \Leftrightarrow -\varepsilon < \frac{(-1)^n}{n} < +\varepsilon \qquad (1)$$

Если n — чётное, то левое неравенство всегда верно, а правое можно переписать как $\frac{1}{n} < \varepsilon$. Если же n — нечётное, то выполнено правое неравенство, а левое можно умножить на -1 и сменить знак неравенства, что опять приведёт нас к $\frac{1}{n} < \varepsilon$. Но мы уже знаем, что при $n > N > \frac{1}{\varepsilon}$ последнее неравенство верно, а значит верно и неравенство (1). Тем самым доказано, что $\lim\limits_{n\to\infty}\left(1 + \frac{(-1)^n}{n}\right) = 1$.

Очень полезно в целях освоения определения разобрать случай, когда оно не выполняется.

Пример. Докажем, что не существует предела последовательности $a_n = (-1)^n$.

Условие примера может несколько смутить, если вы плохо чувствуете разницу между множеством и последовательностью. Обратите внимание, что хотя члены последовательности и могут равняться здесь лишь ± 1, их всё равно бесконечно много: $a_1 = a_3 = a_5 = \cdots = a_{2k+1} = -1$, $a_2 = a_4 = \cdots = a_{2k} = 1$.

Значит, в любом интервале $(1 - \varepsilon; 1 + \varepsilon)$ бесконечно много членов последовательности (все с чётными номерами), но и любом интервале $(-1 - \varepsilon; -1 + \varepsilon)$ содержатся все члены с нечётными номерами, а их тоже бесконечно много.

Предположим теперь, что некоторое число A является пределом нашей последовательности и рассмотрим три случая: 1) A не равно ни -1, ни 1; 2) $A = -1$; 3) $A = 1$.

В первом случае выберем такой интервал, содержащий точку A, который не включает в себя ни -1, ни 1. Тогда внутри выбранного интервала вообще не будет элементов последовательности, а по определению предела их там должно быть бесконечно много.

Во втором случае возьмём интервал $(-2; 0)$. Внутрь попадут все элементы с нечётными номерами, коих бесконечно много, но и снаружи останется бесконечно много элементов последовательности (все с чётными номерами). Поэтому -1 не является пределом нашей последовательности.

Чтобы показать, что 1 тоже не предел, достаточно взять интервал $(0; 2)$.

Немного упражнений.

1. Докажите, что а) $\lim\limits_{n\to\infty} \frac{1}{n^2} = 0$; б) $\lim\limits_{n\to\infty} \frac{n+1}{n-1} = 1$.

2. Докажите лемму о двух милиционерах, опираясь на язык ε–δ.

3. Докажите, что

$$\lim_{n\to\infty} a_n = A \qquad \Leftrightarrow \qquad \lim_{n\to\infty}(a_n - A) = 0.$$

(Здесь утверждается, что предел последовательности a_n равен A тогда и только тогда, когда новая последовательность с элементами $a_n - A$ является бесконечно малой, т. е. её предел равен 0.)

Предел функции

Теперь можно переходить к обсуждению предела функции $\lim\limits_{x\to x_0} f(x) = A$.

Одно из наивных, но очень красочных, «определений» предела функции говорит, что *чем ближе* значения аргумента подходят к точке x_0, *тем ближе* значения функции подходят к пределу A.

Несмотря на образность, эту формулировку ни в коем случае нельзя воспринимать как математическое определение. До строгости тут далеко: что такое «ближе»; ближе на сколько? Кроме того, если о пределе последовательности были хоть какие-то приемлемые представления, то с пределом функции всё значительно сложнее. Действительно, мы прекрасно представляем себе, что означает равенство $f(x_0) = A$. Можно ли говорить, что это и есть определение предела функции? Скорее всего нет, ведь предел берётся при x, стремящемся к x_0. Если бы речь шла не о переменной x, а о последовательности x_n, стремящейся к x_0, то мы точно бы знали, о чём идёт речь: о том, что $\lim\limits_{n\to\infty} x_n = x_0$.

Поэтому стоит в качестве первого приближения к определению предела функции взять следующее:

> $\lim\limits_{x\to x_0} f(x) = A$, если для любой последовательности x_n, стремящейся к x_0 (т. е. $\lim\limits_{n\to\infty} x_n = x_0$) последовательность $f(x_n)$ будет сходиться к A ($\lim\limits_{n\to\infty} f(x_n) = A$).

Протестируем это определение на примере. Рассмотрим функцию

$$f(x) = \begin{cases} 5, & \text{если } x \neq 0, \\ 0, & \text{если } x = 0. \end{cases}$$

Казалось бы, должно быть справедливо, что $\lim\limits_{x\to 0} f(x) = 5$. Однако взяв $x_n = 0$ для всех n, мы получим $f(x_n) = 0$ при всех n и $\lim\limits_{n\to\infty} f(x_n) = 0$, а не 5. Проблема в том, что значение предела функции (при $x \to x_0$) никак не связано со значением функции в точке x_0. Более того, $f(x)$ в точке x_0 может быть вообще не определена. Поэтому мы должны рассматривать только те последовательности $x_n \to x_0$, у которых $x_n \neq x_0$ при любом n.

Учитывая это дополнение, получаем определение предела по Гейне[2]:

Определение. *Пусть функция $f(x)$ определена в окрестности точки x_0. Скажем, что $\lim\limits_{x\to x_0} f(x) = A$, если для любой последовательности x_n, такой что $x_n \neq x_0$ при любом n и $\lim\limits_{n\to\infty} x_n = x_0$ выполнено равенство $\lim\limits_{n\to\infty} f(x_n) = A$.*

Здесь может возникнуть вопрос: а зачем нам учитывать *любую* последовательность значений аргументов, сходящуюся к x_0? Казалось бы, близость, которую мы хотим формализовать, достигается и одной из них. К сожалению это не так. Даже рассуждая о близлежащих точках с интуитивной точки зрения, мы имеем ввиду *все*, расположенные не далее определённого расстояния. А одна последовательность, конечно, не исчерпывает всех точек.

Будем всё же считать, что это определение, несмотря на громоздкость и некоторой неопределённости из-за *всех* последовательностей, отражает реальность и его можно взять в качестве отправной точки. Конечно, над ним стоит ещё поработать, поскольку, опираясь на это определение, довольно трудно проверить даже такой простой факт, что $\lim\limits_{x\to 0} x^2 = 0$.

Попытаемся зайти с другой стороны, танцуя от интуитивного «определения», учитывая, что близость формализована нами как интервал, или окрестность.

Итак, $\lim\limits_{x\to x_0} f(x) = A$, если чем ближе значения аргумента к x_0, тем ближе соответствующие значения функции к A. На языке ε–δ это может звучать так:

> Если $|x - x_0| < \delta$, то $|f(x) - A| < \varepsilon$ для некоторых $\delta > 0$ и $\varepsilon > 0$.

[2] Г. Гейне — немецкий математик второй половины 19 века — *Прим. ред.*

Здесь, конечно, проблема в том, что $\delta > 0$ и $\varepsilon > 0$ никак не связаны и такое определение нам не годится. Действительно, если $f(x) \equiv 1$, то, очевидно, $\lim\limits_{x \to x_0} f(x) = 1$ для любой точки x_0 (это легко проверить, опираясь на определение предела функции по Гейне). С другой стороны, используя наше гипотетическое ε-δ-определение, можно показать, что $\lim\limits_{x \to 0} f(x) = 0$.

Действительно, если $\delta > 0$ — любое и $|x - 0| < \delta$, то $|f(x) - 0| < \varepsilon = 3$.

Загвоздка здесь в том, что в интуитивном определении есть оборот: «чем ближе…, тем ближе…» и трудность состоит именно в его формализации.

Но «нормальные герои всегда идут в обход». Если не получается решить задачу с одного конца, попробуем зайти с другого. В нашем исследовании понятия предела «концами» можно назвать части предложения:

> Чем ближе значения x к x_0, тем ближе $f(x)$ к A.

Давайте переставим их:

> Значения $f(x)$ тем ближе к A, чем ближе x к x_0.

Казалось бы хрен редьки не слаще. Смысл остался тот же. Но, поистине возможности русского языка безграничны. Переставив части предложения, мы сдвинули акцент с переменной на функцию. Теперь формализация на ε-δ-языке звучит так:

> $|f(x) - A|$ должно стать меньше ε, как только $|x - x_0| < \delta$.

Теперь, поскольку мы говорим о все более близких значениях, то мы должны это условие проверять для произвольного сколь угодно маленького $\varepsilon > 0$. Поэтому приходим к следующему определению.

> $\lim\limits_{x \to x_0} f(x) = A$, если $\forall \varepsilon > 0 \ \exists \delta > 0$: как только $|x - x_0| < \delta$, имеет место неравенство: $|f(x) - A| < \varepsilon$ (для любого положительного ε найдётся такое положительное δ, что для всех значений переменной x, удовлетворяющих условию $|x - x_0| < \delta$, будет выполнено неравенство $|f(x) - A| < \varepsilon$).

К сожалению, с этим определением тоже не всё в порядке. Действительно, если

$$f(x) = \begin{cases} 1, & x \neq 0, \\ 2, & x = 0, \end{cases}$$

то $\lim\limits_{x \to 0} f(x)$ должен быть равен 1. Однако взяв $\varepsilon = 1/2$, мы не найдём δ, подходящего к определению: какое бы $\delta > 0$ мы не выбрали, среди значений переменной x, удовлетворяющих неравенству $|x - 0| < \delta$, присутствует $x = 0$. Но $|f(0) - 1| = 1 > 1/2$.

В связи с этим точку x_0 из определения надо выкинуть. Итак, окончательно получаем следующее определение предела функции по Коши[3]:

Определение. $\lim\limits_{x \to x_0} f(x) = A$, *если* $\forall \varepsilon > 0 \ \exists \delta > 0$*: как только* $0 < |x - x_0| < \delta$*, имеет место неравенство:* $|f(x) - A| < \varepsilon$.

Пример. Вычислить $\lim\limits_{x \to 0} x^2$. Очевидно, что если x очень мало, то и x^2 будет маленьким. Поэтому естественно предположить, что искомый предел равен 0. Докажем это. Зафиксируем произвольное положительное число ε. Нам нужно найти такое $\delta > 0$, что при всех x, удовлетворяющих условию $0 < |x - 0| < \delta$, имеет место неравенство: $|x^2 - 0| < \varepsilon$.

Избавимся от модуля в последнем неравенстве и получим:

$$-\varepsilon < x^2 < \varepsilon.$$

Левая часть этого условия выполнена всегда, а правая равносильна такому:

$$-\sqrt{\varepsilon} < x < \sqrt{\varepsilon} \Leftrightarrow |x| < \sqrt{\varepsilon}.$$

---

[3] О. Коши — французский математик 19 века; внес значительный вклад в строгое обоснование математического анализа — *Прим. ред.*

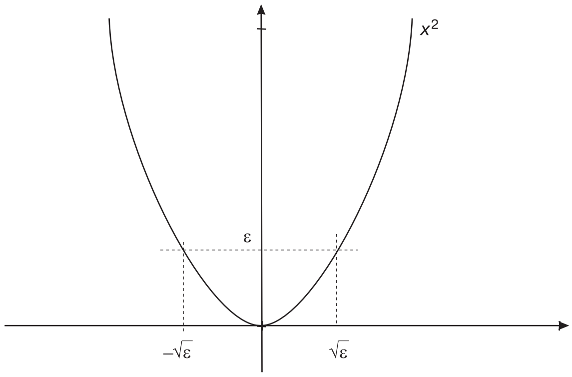

Рис. 5

Значит, если в качестве δ взять $\sqrt{\varepsilon}$ (рис. 5), мы получим требуемое.

Пример. Доказать, что не существует $\lim\limits_{x\to 0}\sin\frac{1}{x}$. Рассмотрим последовательность $a_n = \frac{1}{2\pi n-(\pi/2)}$. Тогда $\sin\frac{1}{a_n} = \sin(2\pi n - (\pi/2)) = -1$ при всех натуральных n. Кроме того, если $b_n = \frac{1}{2\pi n+(\pi/2)}$, то $\sin\frac{1}{b_n} = \sin(2\pi n + (\pi/2)) = 1$.

Возьмём произвольное положительное число δ и покажем, что среди решений неравенства $-\delta < x < \delta$ найдутся как элементы последовательности a_n, так и элементы последовательности b_n.

Ищем такое n, что $-\delta < a_n < \delta$, т. е.

$$-\delta < \frac{1}{2\pi n - (\pi/2)} < \delta$$

Так как дробь положительна, то левая часть неравенства выполнена всегда. Решаем правую часть:

$$\frac{1}{2\pi n - (\pi/2)} < \delta \Leftrightarrow \frac{1}{\delta} < 2\pi n - (\pi/2) \Leftrightarrow \left(\frac{1}{\delta} + (\pi/2)\right) : 2\pi < n.$$

Итак, если $N > \frac{1}{\delta 2\pi} + \frac{1}{4}$, то $-\delta < a_N < \delta$ (рис. 6).

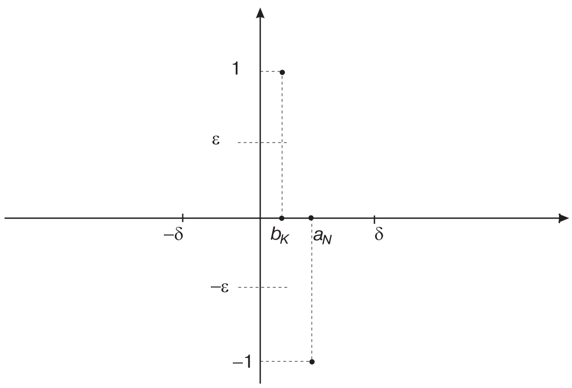

Рис. 6

Аналогично, можно показать, что если $K > \frac{1}{\delta 2\pi} - \frac{1}{4}$, то $-\delta < b_K < \delta$.

Предположим теперь, что найдётся некоторое число A, для которого $\lim\limits_{x\to 0}\sin\frac{1}{x} = A$. Согласно определению для любого положительного $\varepsilon > 0$ должно найтись такое $\delta > 0$, что для всех $x \neq 0$ из интервала $(-\delta, \delta)$ должно выполняться неравенство: $|\sin\frac{1}{x} - A| < \varepsilon$.

Тогда

$$\left|\sin\frac{1}{a_N} - \sin\frac{1}{b_K}\right| = \left|\sin\frac{1}{a_N} - A + A - \sin\frac{1}{b_K}\right| \leq \left|\sin\frac{1}{a_N} - A\right| + \left|\sin\frac{1}{b_K} - A\right| < 2\varepsilon,$$

(поскольку $a_N, b_K \in (-\delta, \delta)$). С другой стороны,

$$\left|\sin\frac{1}{a_N} - \sin\frac{1}{b_K}\right| = |-1-1| = 2.$$

Поэтому, выбрав $\varepsilon = 1/2$, получим противоречие:

$$2 = \left|\sin\frac{1}{a_N} - \sin\frac{1}{b_K}\right| = \left|\sin\frac{1}{a_N} - A + A - \sin\frac{1}{b_K}\right| \le \left|\sin\frac{1}{a_N} - A\right| + \left|\sin\frac{1}{b_K} - A\right| < 2\varepsilon = 1.$$

Значит, ни одно число A не подходит под определение $\lim\limits_{x\to 0}\sin\frac{1}{x}$, т. е. такого предела вообще не существует.

Итак, у нас есть два определения предела функции: по Гейне и по Коши. Неплохо бы показать, что они действительно определяют одно и то же, иными словами, эквивалентны.

Предположим, что для любой последовательности x_n, сходящейся к x_0, т. е. $\lim\limits_{n\to\infty} x_n = x_0$, $\lim\limits_{n\to\infty} f(x_n) = A$ (определение по Гейне). Допустим также, что определение по Коши нарушается. Иными словами, найдётся такое $\varepsilon_0 > 0$, что для любого $\delta > 0$ среди решений неравенства $x_0 - \delta < x < x_0 + \delta$ попадётся $x_\delta \ne x_0$, при котором $|f(x_\delta) - A| > \varepsilon_0$.

Опираясь на это допущение, построим последовательность x_n, сходящуюся к x_0, для которой $\lim\limits_{n\to\infty} f(x_n) \ne A$. Возьмём $\delta = \frac{1}{n}$. Тогда в интервале $(x_0 - \frac{1}{n}; x_0 + \frac{1}{n})$ находится точка $x_n \ne x_0$, для которой $|f(x_n) - A| > \varepsilon_0$. Поскольку нужное x_n можно выбрать для любого натурального n, то мы построим целую последовательность.

Покажем, что $\lim\limits_{n\to\infty} x_n = x_0$. Выберем и зафиксируем $\varepsilon > 0$. Нам нужно найти такое N, что для всех $n > N$ имеет место неравенство: $|x_n - x_0| < \varepsilon$. По построению последовательности x_n мы знаем, что $x_n \in (x_0 - \frac{1}{n}; x_0 + \frac{1}{n})$. В частности, $|x_n - x_0| < \frac{2}{n}$. Поэтому как только $n > N > \frac{2}{\varepsilon}$, $|x_n - x_0| < \varepsilon$. Значит, $\lim\limits_{n\to\infty} x_n = x_0$.

С другой стороны, мы знаем, что $|f(x_n) - A| > \varepsilon$ для всех n, откуда $\lim\limits_{n\to\infty} f(x_n) \ne A$.

Этим рассуждением мы показали, что если не выполняется определение предела по Коши, то не выполняется и определение по Гейне. Значит, если определение по Гейне для данного предела выполнено, то для него выполнено и определение по Коши.

Теперь нам нужно показать, что из определения по Коши $\lim\limits_{x\to x_0} f(x) = A$ следует определение по Гейне. Возьмём произвольную последовательность $x_n \to x_0$, у которой ни один из элементов не совпадает с x_0, и покажем, что $\lim\limits_{n\to\infty} f(x_n) = A$.

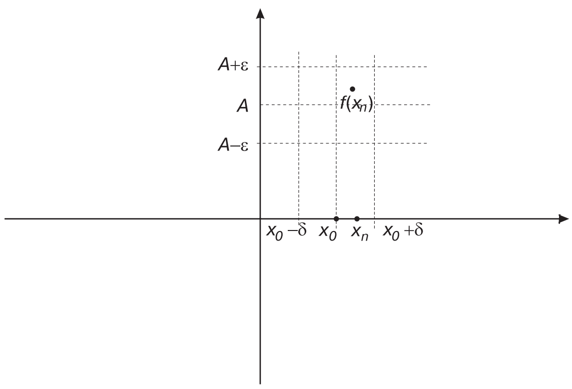

Рис. 7

Возьмём $\varepsilon > 0$. Согласно определению предела по Коши, найдётся такое $\delta > 0$, что при $x \in (x_0 - \delta, x_0 + \delta)$ и $x \ne x_0$ выполняется неравенство $|f(x) - A| < \varepsilon$. Чтобы доказать равенство $\lim\limits_{n\to\infty} f(x_n) = A$, нам достаточно предъявить такой номер N, что для всех $n > N$ верно соотношение $|f(x_n) - A| < \varepsilon$. Но мы уже знаем, что если $x_n \in (x_0 - \delta, x_0 + \delta)$, то оно выполняется. Таким образом, нам достаточно указать такое N, что для всех $n > N$ имеет место такое неравенство: $|x_n - x_0| < \delta$. А такое N, конечно же есть по определению предела последовательности $\lim\limits_{n\to\infty} x_n = x_0$.

Эквивалентность определений доказана.

В заключение предложу несколько простых упражнений.

1. Обосновать пределы функций: а) $\lim\limits_{x\to 0} x^5 = 0$; б) $\lim\limits_{x\to 0} \sin x = 0$; $\lim\limits_{x\to 0} \frac{x^2+x+1}{2x-3} = -\frac{1}{3}$.

2. Доказать, что $\lim\limits_{x\to x_0} f(x) = A$ тогда и только тогда, когда $\lim\limits_{x\to x_0} (f(x) - A) = 0$.

3. Доказать, что $\lim\limits_{x\to x_0} f(x) = 0$ тогда и только тогда, когда $\lim\limits_{x\to x_0} |f(x)| = 0$.

4. Предположим, что функция $f(x)$ определена на отрезке $[-1; 1]$, причём слева от 0 $f(x)$ возрастает (т. е. если $x_1 < x_2 < 0$, то $f(x_1) \le f(x_2)$), а справа убывает (т. е. если $x_1 > x_2 > 0$, то $f(x_1) \le f(x_2)$). Тогда $\lim\limits_{x\to 0} f(x)$ совпадает с точной верхней границей значений функции $f(x)$.

Кулешов Сергей Алексеевич,
доктор физ.-мат. наук,
профессор Военно-Воздушной Академии
имени Жуковского.

Email: KuleshovSergej@rambler.ru

Конфигурация равенства (окончание)

А. Г. Мякишев

Окончание статьи, посвященной красивым и глубоким соотношениям между замечательными точками и линиями (в том числе кривыми второго порядка), связанными с треугольником. Первая часть статьи опубликована в предыдущем номере журнала.

7. Плюс-коника конфигурации равенства

В этом разделе мы сформулируем и докажем свойство, аналогичное *коцикличности*[1] шести точек в конфигурации Конвея (см. начало статьи).

Теорема 7.1. Точки C_{a+}, B_{a+}, A_{b+}, C_{b+}, B_{c+}, A_{c+} конфигурации равенства лежат на одной конике[2].

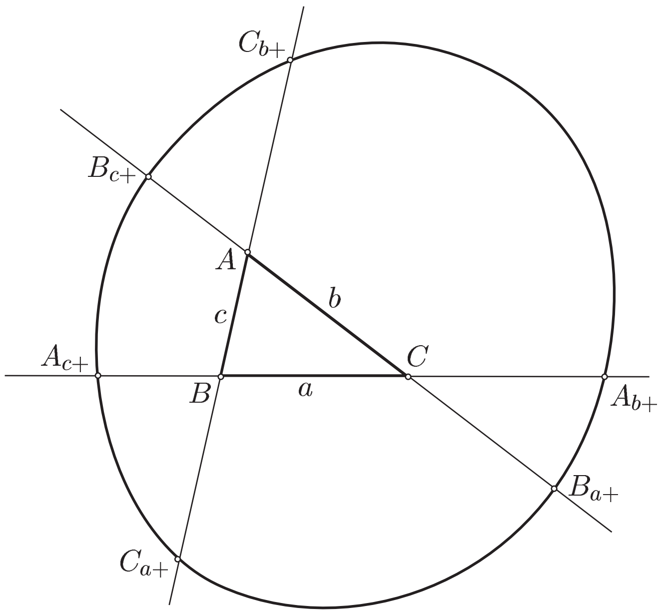

Рис. 1

Доказательство. Воспользуемся следующим утверждением:

Лемма 7.1 (теорема Карно) (см. [1], [11])[3].

Пусть точки A_1 и A_2 расположены на прямой (BC), содержащей сторону BC некоторого треугольника ABC, точки B_1, B_2 — на прямой (CA), а точки C_1, C_2 — на прямой (AB).

Тогда они принадлежат одной конике тогда и только тогда, когда выполняется равенство

$$\left(\frac{\overrightarrow{BA_1}}{\overrightarrow{CA_1}}\cdot\frac{\overrightarrow{BA_2}}{\overrightarrow{CA_2}}\right)\cdot\left(\frac{\overrightarrow{CB_1}}{\overrightarrow{AB_1}}\cdot\frac{\overrightarrow{CB_2}}{\overrightarrow{AB_2}}\right)\cdot\left(\frac{\overrightarrow{AC_1}}{\overrightarrow{BC_1}}\cdot\frac{\overrightarrow{AC_2}}{\overrightarrow{BC_2}}\right)=1^4.$$

[1] Так иногда выражаются, когда хотят сказать о какой-то совокупности точек, что они принадлежат одной окружности.

[2] Иначе говоря, эти шесть точек *конконичны*.

[3] Теорема открыта французом Лазарем Карно (1753–1823) в такой формулировке: «...если эти шесть точек принадлежат одной окружности, то...» . В этом случае утверждение почти очевидно — нужно несколько раз попользоваться теоремой о произведении хорд либо же о касательной и секущей (в зависимости от положения точек на прямых, содержащих стороны треугольника — короче говоря, теоремой о *степени точки* относительно окружности — см. [5], [7]).

Заметим еще, что всему прогрессивному человечеству более известен т. н. *цикл Карно* — но тут постарался уже не Лазарь, а его сын Сади (1796–1832).

[4] Каждая дробь в произведении берется со знаком — т.е. имеется в виду отношения длин направленных отрезков.

В случае нашего утверждения условие Карно принимает вид:

$$\left(\frac{c}{a+c}\cdot\frac{a+b}{b}\right)\cdot\left(\frac{a}{a+b}\cdot\frac{b+c}{c}\right)\cdot\left(\frac{b}{b+c}\cdot\frac{a+c}{a}\right)=1.$$

□

Доказав теорему 7.1, мы можем дать теперь

Определение 7.1. Конику, проходящую через плюс-точки, назовем плюс-коникой конфигурации равенства, или, для краткости, *коникой равенства*.

8. Уравнение коники равенства

Согласно [11], уравнение произвольной коники относительно[5] фиксированного треугольника ABC (в барицентрических координатах) имеет вид:

$$f\cdot x^2+g\cdot y^2+h\cdot z^2+2p\cdot y\cdot z+2q\cdot z\cdot x+2r\cdot x\cdot y=0.$$

Здесь $(x:y:z)$ — «текущие» координаты точки коники, а f, g, h, p, q, r — *коэффициенты уравнения* коники — некие постоянные величины, определяемые данной коникой и данным треугольником. Оказывается, в нашем случае, справедлива

Теорема 8.1. Коэффициенты коники равенства зависят от сторон треугольника ABC следующим образом:

$$\begin{aligned}
&f=2bc(a+b)(a+c), && g=2ca(b+c)(b+a),\\
&h=2ab(c+a)(c+b), && p=a(b+c)\big(bc+(b+a)(c+a)\big),\\
&q=b(c+a)\big(ca+(c+b)(a+b)\big), && r=c(a+b)\big(ab+(a+c)(b+c)\big).
\end{aligned}$$

Доказательство носит чисто технический характер: подставив в уравнение коники координаты первых пяти плюс-точек[6], и считая пока что $r=1$ (ведь уравнение *однородно*), получим систему линейных уравнений:

$$\begin{cases}
(-a)^2\cdot f+(a+c)^2\cdot g+2\cdot(-a)(a+c)=0;\\
(-a)^2\cdot f+(b+a)^2\cdot h+2\cdot(-a)(b+a)\cdot q=0;\\
(-b)^2\cdot g+(a+b)^2\cdot h+2\cdot(-b)(a+b)\cdot p=0;\\
(b+c)^2\cdot f+(-b)^2\cdot g+2(b+c)(-b)=0;\\
(b+c)^2\cdot f+(-c)^2\cdot h+2(b+c)(-c)\cdot q=0.
\end{cases}$$

По счастью, в этой системе много нулевых коэффициентов, и нет необходимости считать определители пятого порядка или обращаться за помощью к методу Гаусса. Вполне хватает лишь определителей *второго* порядка. Действительно, приглядевшись к системе внимательнее, видим, что из *первого* и *четвертого* уравнений можно найти f и g:

$$f=\frac{2b(a+c)}{ab+(a+c)(b+c)},\quad g=\frac{2a(b+c)}{ab+(a+c)(b+c)}.$$

Далее, подставив найденное значение f во *второе* и в *пятое* уравнение, найдем h и q:

$$h=\frac{2ab(c+a)(c+b)}{(a+b)c\cdot(ab+(a+c)(b+c))},\quad q=\frac{b(a+c)\cdot(ca+(c+b)(a+b))}{c(a+b)\cdot(ab+(a+c)(b+c))}.$$

После чего третье уравнение дает нам p. Вволю натешившись умножением, сложением, делением и вычитанием, наконец получим

$$p=\frac{a(b+c)\cdot(bc+(b+a)(c+a))}{c(a+b)\cdot(ab+(a+c)(b+c))}.$$

Теперь, для полного счастья, умножим все коэффициенты коники на знаменатель последней дроби. □

[5]Вообще-то, принято говорить “в базисе”.

[6]Пяти достаточно, ибо (см. свойство 6.1) пять точек как раз и определяют конику.

9. Вид коники равенства

Компьютер показывает, что при изменении длин сторон и величин углов треугольника ABC коника равенства может оказаться и эллипсом, и гиперболой, и параболой. Эллипс мы уже видели в разделе 7, полюбуемся теперь на гиперболу.

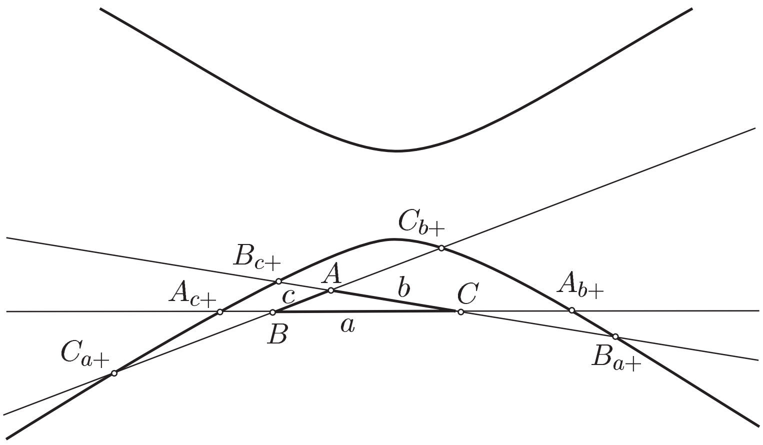

Рис. 2

Возникает вполне естественное желание выяснить, какова геометрическая подоплека происходящего — т. е. какие именно условия определяют вид коники равенства. Как, например, было бы прекрасно получить ответ вроде следующего: “если исходный треугольник остроугольный, то получим эллипс, если прямоугольный — то параболу, а если тупоугольный — то гиперболу”. Увы, все не так просто. Если верить компьютеру, то действительно, для остроугольного треугольника всегда получается эллипс, а что происходит дальше, автору понять так и не удалось.

Однако имелась надежда, что выручит алгебра.

Известно (см. [11]), что если коника задана своим уравнением (см. § 8), то вид коники зависит от знака выражения $\Phi = U + V + W + 2(F + G + H)$ (где $U = gh - p^2$, $V = hf - q^2$, $W = fg - r^2$, $F = qr - fp$, $G = rp - gq$, $H = pq - hr$).

Если $\Phi > 0$, то коника является эллипсом, если $\Phi = 0$ — параболой, а если $\Phi < 0$ — гиперболой.

Все необходимые коэффициенты имеются, казалось бы — бери и считай.

До этого, за годы многолетней практики, автору в подобных случаях всегда хватало ручки, бумаги и времени, чтобы произвести любые нужные расчеты — но здесь он впервые вынужден был отступиться. Из-под пера медленно выползал здоровенный “крокодил”, и по мере его выползания все понятнее становилось, что никакого просвета (в смысле, что удастся этого крокодила разложить на части, то бишь множители) ожидать не приходится.

Тогда ничего другого не оставалось, кроме как попробовать воспользоваться “продвинутыми счетами” — какой-нибудь вычислительной программой. И вот как отозвалась на просьбу “Expand”[7] программа «Mathematica 5.1 for Students» :

$$\begin{aligned}
&- a^6b^2 + 2a^4b^4 - a^2b^6 - 2a^6bc + 4a^4b^3c + 6a^3b^4c +\\
&2a^2b^5c - 2ab^6c - a^6c^2 + 4\ a^3b^3c^2 + 6a^2b^4c^2 + 4ab^5c^2 -\\
&b^6c^2\ -2a^4bc^3 - 2a^3b^2c^3 + 4a^2b^3c^3 + 6ab^4c^3 + 2\ b^5c^3+\\
&2a^4c^4 - 2a^3bc^4 - 4a^2b^2c^4 + 4ab^3c^4 +4b^4c^4 - a^2c^6 -\\
&2abc^6 - b^2c^6
\end{aligned}$$

На просьбу же произвести с полученным безобразием “Factor”[8] — в ответ явилось то же самое выражение — явным синонимом фразы “Ах, отстаньте, наконец”.

[7] По этой команде она производит всевозможные действия с заданными алгебраическими выражениями — например, подставляет одно выражение в другое, перемножает, складывает, раскрывает скобки и т. д.

[8] По этой команде программа как раз и должна раскладывать многочлен на множители.

В запасе, правда, было еще одно, последнее, но самое сильнодействующее средство — команда "FullSimplify"[9], которая выдала:

```
- a²b²(a² - b²)² - 2ab(a⁵ - 2a³b² - 3a²b³ - ab⁴ + b⁵)c -
(a⁶ - 4a³b³ - 6a²b⁴ - 4ab⁵ + b⁶)c² + 2b(a + b)²(- a² + ab + b²)c³ +
2(a + b)(a³ - 2a²b + 2b³)c⁴ - (a + b)²c⁶
```

Вот и все, чего удалось добиться. Как говориться, пусть тот, кто может, сделает больше.

10. Центр коники равенства

Напомним, что и эллипс, и гипербола — кривые *центрально-симметричные*, а их центры симметрии и называют просто центрами соответствующих коник. Что касается параболы, ее центром удобно считать (из предельных соображений) *бесконечно удаленную* точку *оси* параболы.

Как геометрически строить центр коники?

Оказывается, справедливо такое утверждение:

Утверждение 10.1 (см. [1], [6], [11]). Рассмотрим любое семейство параллельных друг другу хорд некоторой коники. (В это семейство, разумеется, входит и касательная, как хорда с двумя совпадающими концами.) Тогда середины этих хорд (точка касания — для касательной) лежат на прямой, проходящей через центр этой коники. (В случае параболы получим прямую, параллельную ее оси).

Кстати, дадим здесь же

Определение 10.1. Хорда, проходящая через центр коники, называется ее диаметром.

Из утверждения 10.1 вытекает простой способ построения центра коники — нужно взять две пары параллельных хорд и в каждой паре провести прямые через их середины. Эти прямые тогда пересекутся в центре коники.

Барицентрические же координаты центра коники, зная ее уравнение, можно вычислить следующим образом:

Утверждение 10.2 (см. [11]). Центр коники имеет координаты

$$(U + G + H : V + F + H : W + F + G)^{10}.$$

Без всяких колебаний запустив "Mathematica 5.1", получим основной результат этого раздела:

Теорема 10.1. Центр коники равенства имеет следующие координаты (приведена лишь первая, остальные получаются из нее циклическими перестановками):

```
- a²(b + c)(a⁴(b + c) - b(b - c)²c(b + c) + a³(b² + c²)
- a(b² + c²)² - a²(b³ + c³))
```

Или, иначе (если угодно):

```
a⁵b³ + a⁴b⁴ - a³b⁶ + a⁵b²c + a³b⁴c -
a²b⁵c - 2ab⁶c + 2a³b³c² - ab⁵c² - b⁶c² - 2a³b²c³ +
2a²b³ + ab⁴c³ - b⁵c³ + ab³c⁴ + b⁴c⁴ + ab²3c⁵ + b³c⁵
```

[9]По мнению разработчиков, "это действие приводит выражение к виду, *приятному во всех отношениях*" (вольный перевод автора статьи соответствующей инструкции из руководства пользователя).

[10]Определение соответствующих выражений см. в § 9.

11. Перспектор коники равенства

В этом разделе нам потребуются такие понятия, как *полюс* и *поляра* относительно коники.

Сначала вместо коники рассмотрим окружность с центром в точке O и радиуса R.

Определение 11.1 *(полюс и поляра относительно окружности)* (см. [1], [2], [5], [11]). Для произвольной *точки* P поставим ей в соответствие некоторую *прямую* p так, что эта прямая *перпендикулярна* прямой OP и пересекает ее в такой точке P', что выполняется равенство $OP \cdot OP' = R^2$ (т. е. P' есть *инверсный* образ точки P). Прямую p назовем *полярой* точки P относительно данной окружности, а точку P — *полюсом* прямой p. Таким образом, имеем т. н. *полярное соответствие* между точками и прямыми плоскости (относительно данной окружности). При этом еще полярой центра O считают бесконечно удаленную прямую, а полярой бесконечно удаленной точки — прямую, проходящую через центр окружности перпендикулярно *направлению*, задающему бесконечно удаленную точку.

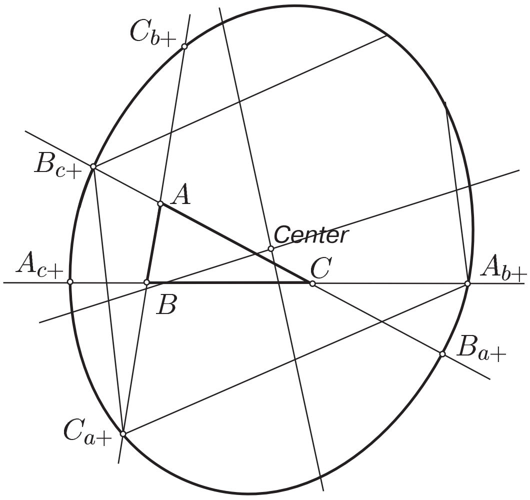

Рис. 3.

Если же вместо окружности взять конику, то поступают следующим образом:

Определение 11.2 *(полюс и поляра относительно коники)* (см. [1], [11]). Пусть имеется точка P и коника K. Переведем каким-либо *проективным* преобразованием конику в некоторую окружность ω. При этом точка P перейдет в некую точку P_1. Пусть p_1 — поляра точки P_1 относительно окружности ω. Подействовав на нее обратным преобразованием, получим прямую p. Можно показать, что результат *не будет зависеть от выбора исходного проективного преобразования.*

Прямую p назовем *полярой* точки P относительно данной коники K, а точку P — *полюсом* прямой p. При этом полярой центра коники считают бесконечно удаленную прямую, а полярой бесконечно удаленной точки — прямую, проходящую через середину любой хорды, параллельной прямым, определяющим бесконечно удаленную точку[11].

Полярные соответствия обладают множеством замечательных свойств. Сформулируем два самых основных из них.

Утверждение 11.1 (см. [1], [11]). Полюс и поляра — понятия *проективные.* То есть, если при произвольном проективном преобразовании коника K переходит в конику K$'$, полюс P и поляра p (относительно K) в какие-то соответственно точку P' и прямую p', то эти прямые и точка будут полярой и полюсом относительно коники K$'$.

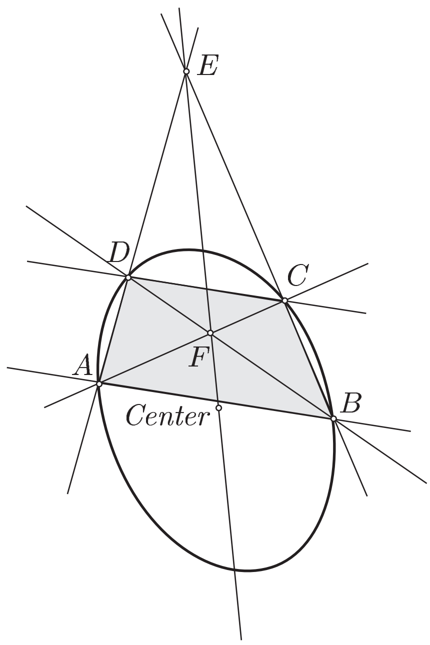

Рис. 4.

Утверждение 11.2 (см. [1], [11]). Пусть прямая p — поляра точки P относительно некоторой коники. Тогда поляра любой точки на прямой p проходит через полюс P, а полюс любой прямой, проходящей через P, лежит на поляре p.

Имеется красивый геометрический способ построения поляр и полюсов с помощью одной только линейки. Например, справедливо

[11]Проверьте, что так определенная поляра бесконечно удаленной точки в случае окружности совпадает с определением 11.1.

Утверждение 11.3 (см. [1], [11]). Проведем из данной точки P две прямые, одна из которых пересекает конику К в точках A и B, а другая — в точках D и C. Далее отметим точку F — пересечения прямых AC и BD, а также точку E, в которой пересекаются прямые AD и BC. Тогда прямая EF является полярой точки P относительно коники К.

Приведем одно, полностью в проективном духе, рассуждение, подтверждающее этот факт.

Переведем проективным преобразованием точку P в *бесконечно удаленную*. В силу утверждения 11.1, достаточно доказать справедливость нашей теоремы для этого случая.

Но тут-то и возникает *трапеция* $ABCD$, и, по известной теореме о четырех точках, (см. [5], [7]) прямая EF будет проходить через середины ее оснований, а значит (см. утверждение 10.1) и через центр коники. Взглянув еще разок на определение 11.2 (поляра бесконечно удаленной точки) мы видим, что все уже доказано.

Теперь сформулируем основную теорему и основное определение этого параграфа.

Теорема 11.1 (см. [1], [11]). Для данных треугольника и коники рассмотрим треугольник, образованный полярами вершин исходного треугольника[12]. Тогда этот *полярный* треугольник будет *перспективен* исходному[13].

Определение 11.3 (см. [1], [11]). Соответствующий перспектор называют *перспектором данной коники относительно данного треугольника.*

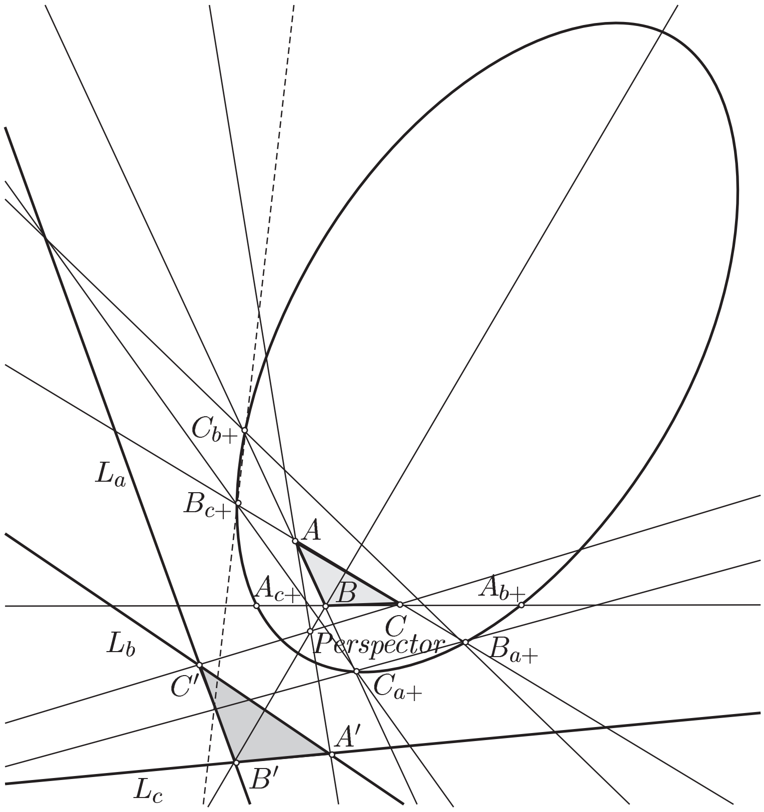

Рис. 5

Найдем в заключение этого раздела координаты перспектора коники равенства.

Утверждение 11.4 (см.[11]). Координаты перспектора коники относительно треугольника имеют вид: $(\frac{1}{F} : \frac{1}{G} : \frac{1}{H})$[14].

Теорема 11.2 (доказана программой «Mathematica 5.1»). Перспектор коники равенства имеет следующие координаты:

$$\frac{1}{\Big(b(a+b)c(a+c)\big(a^2bc - 2a^3(b+c) + bc(b+c)^2 + 2a(b^3+c^3)\big)\Big)}$$

[12] В силу утверждения 11.2 можно было сказать и так: образованного полюсами сторон исходного треугольника.

[13] Разумеется, если только он не совпадает с исходным (в этом случае треугольник называют *автополярным*).

[14] В терминах §9.

(Выписана первая координата, остальные получаются циклическими перестановками).

12. В погоне за точками

В двух предыдущих параграфах мы довольно подробно описали две точки, связанные с треугольником — центр коники равенства и ее перспектор. В этой связи не мешает, наверное, сказать несколько слов о *замечательных точках треугольника* вообще.

Строгого математического определения замечательной точки треугольника не существует.

Имеется, правда, вполне научное определение т. н. *центральной точки* (см. [4], [12]) — но, во-первых, под это определение не попадают многие точки, безусловно замечательные в интуитивном смысле, а во-вторых, попадают многие, которые скорее хотелось бы назвать *ничем не примечательными.* Говоря неформально, *«степень замечательности» той или другой точки можно оценить дробью, в числителе которой — количество нетривиальных свойств, связанных с этой точкой, а в знаменателе — «сложность» ее построения.*

Американский математик Кларк Кимберлинг в интересной книжке [12], вышедшей в 1998 году, предпринял попытку собрать под ее обложкой все известные на тот момент треугольные центры (точке присваивается порядковый номер, даются ее координаты и приводятся основные геометрические свойства).

Получилось — ни много, ни мало — 400 штук (чувствовалось, что профессор собирает их не первый год). Затем Кимберлинг разместил электронную версию своей коллекции на сайте (см. [10]) и с тех пор она непрерывно пополняется — благодаря усилиям энтузиастов, без преувеличения можно сказать, всего мира (большое спасибо современным средствам связи и современным же геометрическим программам). В настоящее время[15] количество центров достигло отметки 3292. (Забавно, но сначала я поставил сноску непосредственно за этим числом. Будем все же надеяться, что до 3292^{30} дело не дойдет). Вероятно, для некоторых поиск нового центра[16] стал своего рода охотой[17] или спортом. Оно и понятно — кому же не хотелось бы увековечить свое имя в анналах ETC[18].

Как бы оно там ни было, очевидно, что нынче открыть какую-нибудь хорошую точку, еще не попавшую в ETC, практически невозможно — все хорошие уже разобрали[19]!

И тем не менее: ни центра коники равенства, ни ее перспектора в ETC нет. Однако это не слишком удивляет (стоит лишь взглянуть на координаты), и не слишком радует — поскольку, если исходить из неформального определения, замечательность этих точек близка к нулю. Я пробовал найти связи между ними и точками из первой десятки — на предмет коллинеарности или коцикличности, но таковые не обнаружились. Возможно, среди первой тысячи чего-нибудь и нашлось, но эта гипотеза как-то не вдохновляет на дальнейшие поиски.

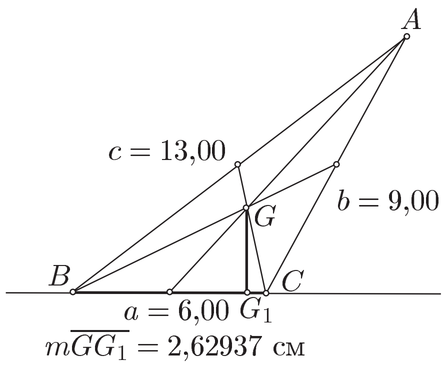

Рис. 6.

У читателя, не знакомого с ETC, мог возникнуть вполне естественный вопрос: а как все же проверить, значится ли потенциально новая точка в списках? Неужели сравнивать координаты с более чем 3200 других? Да и сами координаты, как видели, не всегда просто вычислить. Согласитесь, довольно удручающая рисуется картинка: считали-считали, сосчитали, наконец, затем начали сравнивать — и вдруг, шаге на 2000-ом, обнаружили совпадение с известной точкой.

Но на самом деле ничего такого страшного произойти не может.

[15] 24 февраля 2008 года.

[16] Конечно, для настоящего любителя геометрии это занятие не может быть самоцелью.

[17] В молодости мне попалась в руки книжка с интригующим названием "Охотники за черепами". Уж лучше за точками, чем за черепами!

[18] (См. ссылку [10] — *Прим. ред.*) Вспоминается невольно хрестоматийная надпись вроде: "Здесь был Вася" — каждому, наверное, доводилось видеть ее в самых неожиданных местах, а то и...

[19] Несколько раз я и сам был в шаге от успеха, придумав какую-нибудь более-менее симпатичную конструкцию. Но всякий раз выяснялось, что это уже было. (Причем иногда я опаздывал всего лишь года на 3–4).

Вот что придумал хитроумный Кимберлинг.

В его Энциклопедии каждой точке, помимо порядкового номера и координат, присвоено еще и некое число. Это число — расстояние от данной точки[20] до прямой, содержащей сторону BC треугольника ABC *со сторонами* $BC = 6$, $CA = 9$, $AB = 13$.

К примеру, мы открыли, что медианы пересекаются в одной точке, но не знаем, есть ли такая в ETC.

Тогда вычислим соответствующее расстояние в соответствующем треугольнике, и заглянем в Энциклопедию.

Мы увидим два столбца — в правом находится расстояние (таблица упорядочена именно по возрастанию расстояний), а в левом — порядковый номер. В нашем случае мы довольно быстро найдем нужную нам строчку:

2	2.629368792488

А затем уже открываем список точек по номерам, и видим, что точка под номером 2 — т. н. центроид. Но если Вашего числа в списках не оказалось, значит — точка новая.

Центру коники равенства соответствует число 47,86483.

Перспектору коники равенства — число −2,62937.

И этих чисел в Энциклопедии нет (пока нет).

13. Коника равенства и теорема Паскаля

Этот раздел специально предназначен фанатам новых точек и открывает перед ними широкое поле деятельности.

Вспомним вначале о теореме Паскаля.

Теорема Паскаля (см. [1], [2], [6], [7] — з. 31.52, [11]). Если шестиугольник $ABCDEF$ вписан в конику, то точки пересечения его противоположных сторон (т. е. (AB, DE), (BC, EF), (CD, FA)) лежат на одной прямой.

Замечание. Обратная теорема также справедлива — см. [1], [11].

Коника у нас есть, значит, имеем полное право применить теорему Паскаля к плюс-точкам.

Получили прямую Паскаля. Но в треугольнике каждой прямой можно поставить в соответствие точку! (т. н. *трилинейный полюс* прямой, а прямая будет называться, соответственно, *трилинейной полярой* — подробнее см. [1], [4], [11]). Делается это следующим образом: пусть наша прямая пересекает прямую (BC) в точке A_2. Эта точка делит отрезок BC в каком-то отношении внешним (внутренним) образом. Построим затем точку A_1, которая будет делить отрезок BC *в том же самом отношении* внутренним (внешним) образом.

Аналогично построим точки B_1, C_1 на двух других прямых, содержащих стороны треугольника. Прямые AA_1, BB_1, CC_1 пересекутся тогда в одной точке (что обеспечено теоремами Менелая и Чевы — см. [7], з. 5.69 и 5.85, [8], з. 342, 341), которая и называется полюсом исходной прямой.

[20] Взятое со знаком «+» , если точка находится «над» прямой BC, и со знаком «−» в противоположном случае

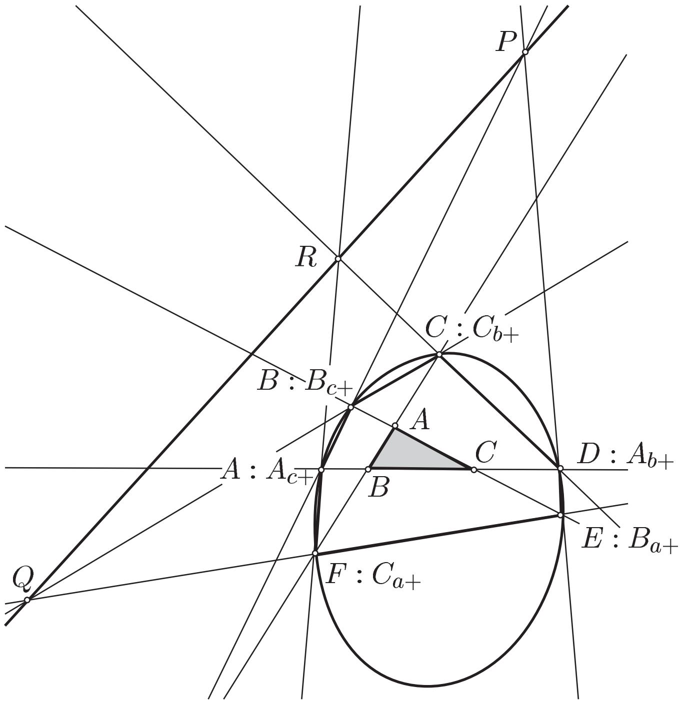

Рис. 7

Замечание. Четвертка точек B, C, A_1, A_2 при этом образует т. н. *гармоническую четверку* — см. [1], [6], [11]. На рисунке ниже указан способ построения четвертой гармонической: из точки A_2 проводим произвольную прямую, отмечаем точки пересечения ее с прямыми (CA) и (BA) — точки B_2 и C_2, потом строим точку P пересечения (BB_1) и (CC_1), и, наконец — точку A_0 пересечения (AP) и (BC).

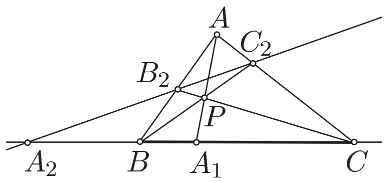

Рис. 8

Но вернемся к конике равенства и построенной для нее прямой Паскаля и рассмотрим ее полюс.

По системе «6-9-13» (см. предыдущий параграф) получаем число 2,09082. Залезаем в ETC, смотрим... увы, центр не нов, его порядковый номер — 37. (Но зато оказался нашим старым знакомым — см. теорему 2.2).

Однако стоит ли огорчаться? Дело в том, что рассматривая *всевозможные* шестиугольники с теми же вершинами (разрешаются невыпуклые и с самопересечениями) — имеем ровным счетом еще 59 прямых Паскаля, а стало быть, еще 59 точек. Уж наверное, среди них отыщется что-нибудь новенькое.

И это еще не все. Оказывается, некоторые из 60 паскалевых прямых пересекаются по три в одной точке (натурально, замечательной). А именно, справедливы теоремы:

Теорема Штейнера (см. [6], [7] — з. 31.53). Если шестиугольник $ABCDEF$ вписан в конику, то прямые Паскаля шестиугольников $ABCDEF$, $ADCFEB$, $ADEBCF$ пересекаются в одной точке.

Теорема Киркмана (см. [6], [7] — з. 31.53). Если шестиугольник $ABCDEF$ вписан в конику, то прямые Паскаля шестиугольников $ABFDCE$, $AEFBDC$, $ABDFEC$ пересекаются в одной точке.

Построенная согласно этим указаниям, точка Штейнера имеет кодовое число −49,58238 и совпадает с центром под номером 661(!), а точка Киркмана — число 7,68422 — и ни с чем не совпадает (наконец-то!)[21].

Наконец, заметим, что (в зависимости от выбора начального шестиугольника, а точнее, шестивершинника) точек Штейнера всего набирается 20 штук, а Киркмана — так и вообще 60.

Так что, работы — непочатый край, просто Клондайк какой-то!

14. О добавочных кониках равенства

Как, конечно же, помнит читатель, характерная особенность конфигурации равенства заключалась в том, что каждое ее свойство имело три с ним родственных.

Коника равенства не является исключением.

Теорема 14.1. Точки B_{a-}, C_{a-}, C_{b+}, A_{b-}, A_{c-}, B_{c+} конфигурации равенства лежат на одной конике.

Доказательство. Применим теорему Карно (см. **лемму 7.1**):

$$\left(\frac{b-a}{b}\cdot\frac{c}{c-a}\right)\cdot\left(\frac{a}{a-b}\cdot\frac{b+c}{c}\right)\cdot\left(\frac{a-c}{a}\cdot\frac{b}{b+c}\right)=1.$$

Определение 14.1. Конику, проходящую через эти точки, назовем первой добавочной коникой равенства.

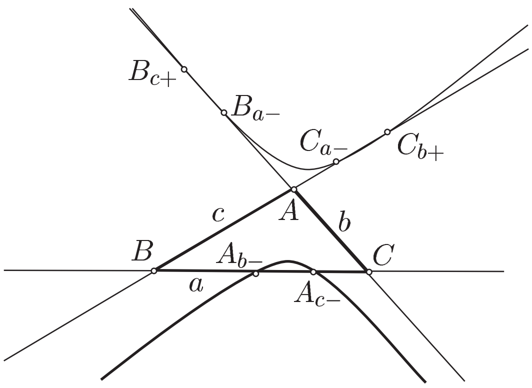

Рис. 9.

Теорема 14.2. Коэффициенты уравнения первой добавочной коники равенства зависят[22] от сторон треугольника ABC следующим образом:

$$\begin{aligned}
f &= 2bc(a-c)(b-a)(a+b),\\
g &= 2ca(b+c)(b+a)(b-a),\\
h &= 2ab(c-b)(a-c)(b-c),\\
p &= a(bc(b+c)(a+b)+(c-b)(a-c)(b-a)^2),\\
q &= b(a-c)(b-a)(ca+(c-b)(a+b)),\\
r &= c(a+b)(b-a)(ab+(a-c)(b+c)).
\end{aligned}$$

Доказательство полностью аналогично доказательству теоремы 8.1.

Замечание. Компьютер показывает, что добавочная коника *всегда* является *гиперболой*. Геометрически это связано, конечно, с тем, что шестиугольник $B_{a-}C_{a-}C_{b+}A_{b-}A_{c-}B_{c+}$ (при любой последовательности обхода вершин) *всегда* будет *невыпуклым*. Насколько очевиден этот факт? Строгое доказательство, вероятно, должно быть связано с рассмотрением всевозможных неравенств, возникающих между длинами сторон треугольника.

Можно переформулировать и так: какие бы три точки из шести мы ни выбрали, среди трех других найдется хотя бы одна, попавшая *внутрь* треугольника с вершинами в первых трех выбранных точках.

Автор будет очень признателен всякому, кто сообщит ему короткое (а пускай даже не очень) и (главное) строгое доказательство невыпуклости.

---

[21]Используя многократно волшебный определитель из доказательства теоремы 2.1, можно ведь и координаты сосчитать! А особенно, если подключить еще к этому делу не менее волшебную программу “Mathematica 5.1”.

[22]Точнее сказать, *наверное* зависят — к концу нашей сильно затянувшейся работы автор несколько подустал, и потому полученный результат (не слишком принципиальный) проверять и перепроверять (как в случае с теоремой 8.1) не захотел. Читатель волен проделать это сам, если захочет.

15. Плюс-точки: конкурентные свойства — дополнение

Формулировки и доказательства основных теорем этого параграфа принадлежат Арсению Акопяну.

Во-первых, он обнаружил вот какое интересное свойство плюс-точек.

Оказывается, не только треугольник, образованный прямыми (C_{a+}, B_{a+}), (A_{b+}, C_{b+}), (B_{c+}, A_{c+}) будет перспективен Δ_{ABC} с перспектором в точке H' — ортоцентре треугольника Жергонна (см. определение 2.1 и теорему 2.1), но и треугольник, образованный прямыми

$$(C_{a+}, A_{c+}), \quad (B_{c+}, C_{b+}), \quad (B_{a+}, A_{b+}).$$

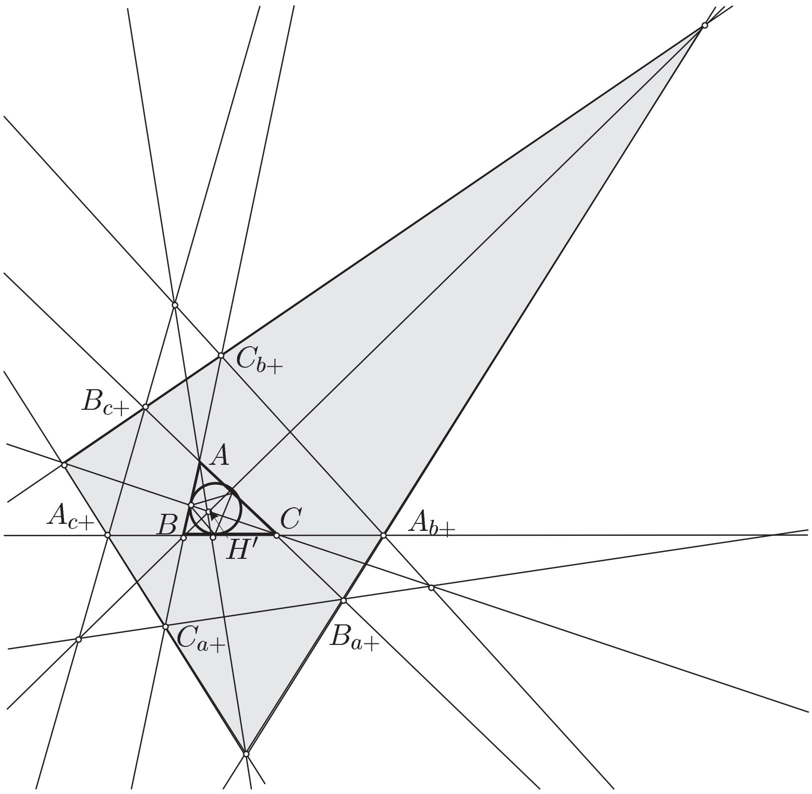

Рис. 10

Итак, справедлива

Теорема 15.1. Треугольник, образованный прямыми (C_{a+}, A_{c+}), (B_{c+}, C_{b+}), (B_{a+}, A_{b+}), перспективен исходному с перспектором в точке H' — ортоцентре треугольника Жергонна[23].

Доказательство. Понятно, что при симметрии относительно *внешней* биссектрисы $\angle ACB$, ввиду равнобедренности соответствующих треугольников, $B \to B_{a+}$, $A \to A_{b+}$, т. е. прямая (AB) переходит в прямую (B_{a+}, A_{b+}). Аналогично, при симметрии относительно *внешней* биссектрисы $\angle ABC$, $A \to A_{c+}$, $C \to C_{a+}$, т. е. прямая (AC) переходит в прямую (C_{a+}, A_{c+}). Рассмотрим теперь *вневписанную* в треугольник ABC окружность, которая касается луча AB в точке P, а луча AC — в точке R. Поскольку при симметрии относительно двух нами рассмотренных внешних биссектрис вневписанная окружность переходит сама в себя, отсюда делаем заключение, что вневписанная окружность касается еще и прямых (B_{a+}, A_{b+}) и (C_{a+}, A_{c+}), точку пересечения которых обозначим A', в каких-то точках Q и T соответственно.

Заметим далее, что при гомотетии с *вершиной* в A, переводящей *вписанную* окружность во *вневписанную*, треугольник Жергонна перейдет в некоторый треугольник, *две вершины* которого совпадают с точками P и R, причем при этой гомотетии соответственные стороны обоих треугольников будут *параллельны*. Но $(PQ)\perp(CI_a)\perp(CI)$. Последняя же прямая (внутренняя биссектриса) еще и перпендикулярна соответствующей стороне треугольника Жергонна. Из всего

[23] А значит, в силу теоремы 2.1, прямые, содержащие соответствующие вершины *трех* треугольников, пересекаются в H'.

вышесказанного следует вывод, что (PQ) перпендикулярна той стороне образа треугольника Жергонна, которая проходит через точку R. Рассуждая точно так же, доказываем, что (RT) перпендикулярна той стороне образа треугольника Жергонна, которая проходит через точку P.

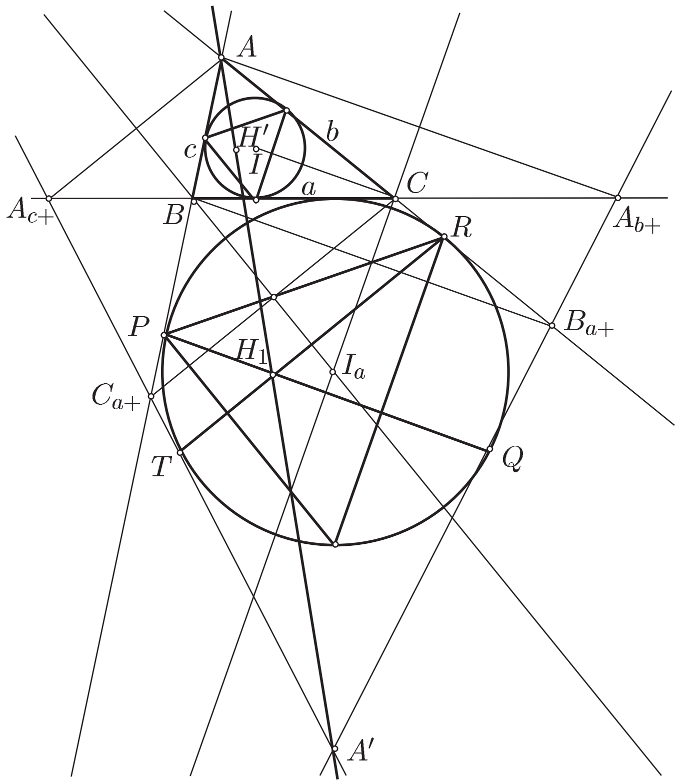

Рис. 11

То есть, прямые (PQ) и (RT) пересекаются в ортоцентре H_1 образа треугольника Жергонна при рассматриваемой гомотетии.

Поэтому точки A, H', H_1 лежат на одной прямой.

И, наконец, четырехугольник $AC_{a+}A'B_{a+}$ описан около окружности, а для таких четырехугольников, как это следует из теоремы *Брианшона* (см. [1], [2], [11]) *точка пересечения диагоналей совпадает с точкой пересечения отрезков с концами в точках касания, расположенных на противоположных сторонах.* Поэтому на одной прямой также лежат и точки A, H_1, A'. Так как через две точки можно провести ровно одну прямую, *обе прямые совпадают.*

Доказательство для двух других вершин полностью аналогично. □

Во-вторых, Арсению принадлежит еще одно красивое утверждение:

Теорема 15.2. Пусть $A_1 = (A_{b+}, C_{b+}) \cap (B_{c+}, A_{c+})$, а H_1 — ортоцентр треугольника, являющегося образом треугольника Жергонна при гомотетии с вершиной в A, переводящей вписанную окружность во вневписанную. Тогда точки A_1 и H_1[24] симметричны относительно вершины A.

Аналогичные утверждения верны и для двух других точек пересечения:

$$B_1 = (C_{a+}, B_{a+}) \cap (B_{c+}, A_{c+}) \qquad \text{и} \qquad C_1 = (A_{b+}, C_{b+}) \cap (C_{a+}, B_{a+}).$$

Так как точки A, H', H_1 лежат на одной прямой (и еще две схожие коллинеарности), как следствие мы получаем и теорему 2.1

[24]Как вытекает из доказательства **теоремы 15.1**, точка H_1 еще и лежит на стороне $C_{a+}B_{a+}$ — как точка пересечения диагоналей вписанного четырехугольника. Аналогично и для двух других ортоцентров.

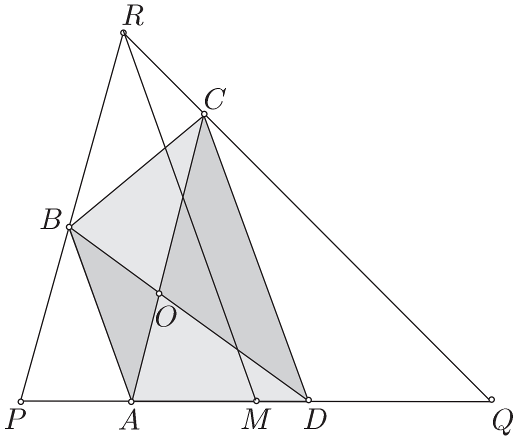

Рис. 12.

Сформулируем и докажем сначала три леммы.

Лемма 15.1 *(о равнобокой трапеции)*. Дана равнобокая трапеция $ABCD$ ($BC = AD$), O — точка пересечения ее диагоналей. Отложим на луче DA за точку A такую точку P, что $AO = AP$, а на луче AD — за точку D такую точку Q, что $DO = DQ$. Пусть R — точка пересечения прямых (PB) и (QC), а M — середина отрезка PQ. Тогда прямая (RM) параллельна основаниям трапеции.

Доказательство. Рассмотрим прямую (RM), параллельную основаниям трапеции, и докажем, что ее точка пересечения M с отрезком PQ делит этот отрезок пополам.

Заметим, что

$$\Delta PBA \sim \Delta PRM, \quad \Delta QDC \sim \Delta QMR \quad \Rightarrow \quad \frac{BA}{RM} = \frac{PA}{PM}, \quad \frac{CD}{RM} = \frac{QD}{QM}.$$

Разделив одно из этих равенств на другое, и воспользовавшись подобием треугольников BOA и DOC, получим:

$$\frac{PA}{PM} \cdot \frac{QM}{QD} = \frac{BA}{CD} = \frac{OA}{OD} = \frac{PA}{QD} \quad \Rightarrow \quad \frac{QM}{PM} = 1 \quad \Rightarrow \quad QM = PM.$$

Лемма 15.2. Пусть M_a, M_b, M_c — точки касания соответствующих вневписанных окружностей со сторонами исходного треугольника ABC. Тогда прямые A_1M_a, B_1M_b, C_1M_c будут параллельны соответствующим внутренним биссектрисам треугольника ABC[25].

Доказательство. Очевидно, четырехугольник $BB_{c+}C_{b+}C$ является равнобокой трапецией, так как прямые BB_{c+} и $C_{b+}C$ параллельны биссектрисе AI (по лемме 2.2), а $B_{c+}C_{b+} = BC$, так как треугольники ABC и $AB_{c+}C_{b+}$ равны по двум сторонам и углу между ними.

Кроме того, $CM_a = s - b$, $BM_a = s - c \Rightarrow A_{c+}M_a = A_{b+}M_a = s$ (где s — полупериметр треугольника ABC), в силу известного свойства вневписанных окружностей (см. [5], з. 3.2).

Осталось воспользоваться предыдущей леммой.

Лемма 15.3. В $+\Delta$ (треугольнике $A_1B_1C_1$) следующие пары отрезков равны между собой:

$$B_1C_{a+} = C_1B_{a+}, \quad A_1C_{b+} = C_1A_{b+}, \quad B_1A_{c+} = A_1B_{c+}.$$

Доказательство. По условию, $BC_{a+} = CB_{a+} = a$. По двум предыдущим леммам $M_cB = M_bC = s - a$ и $(M_cC_1)||(CI)||(BB_{a+})$, $(M_bB_1)||(BI)||(CC_{a+})$.

[25]И кроме того, как безапелляционно утверждают Компьютер+Кимберлинг, указанные прямые пресекаются в точке $X72$ с координатами $((b+c)\cos A : (c+a)\cos B : (a+b)\cos C)$ — т. е. плюс-треугольник перспективен еще и *треугольнику Нагеля* с перспектором в этой точке. Но довольно!.. Иначе повествование наше никогда не закончится. "The rest is silence", как метко выразился, правда, по несколько другому поводу, один литературный персонаж.

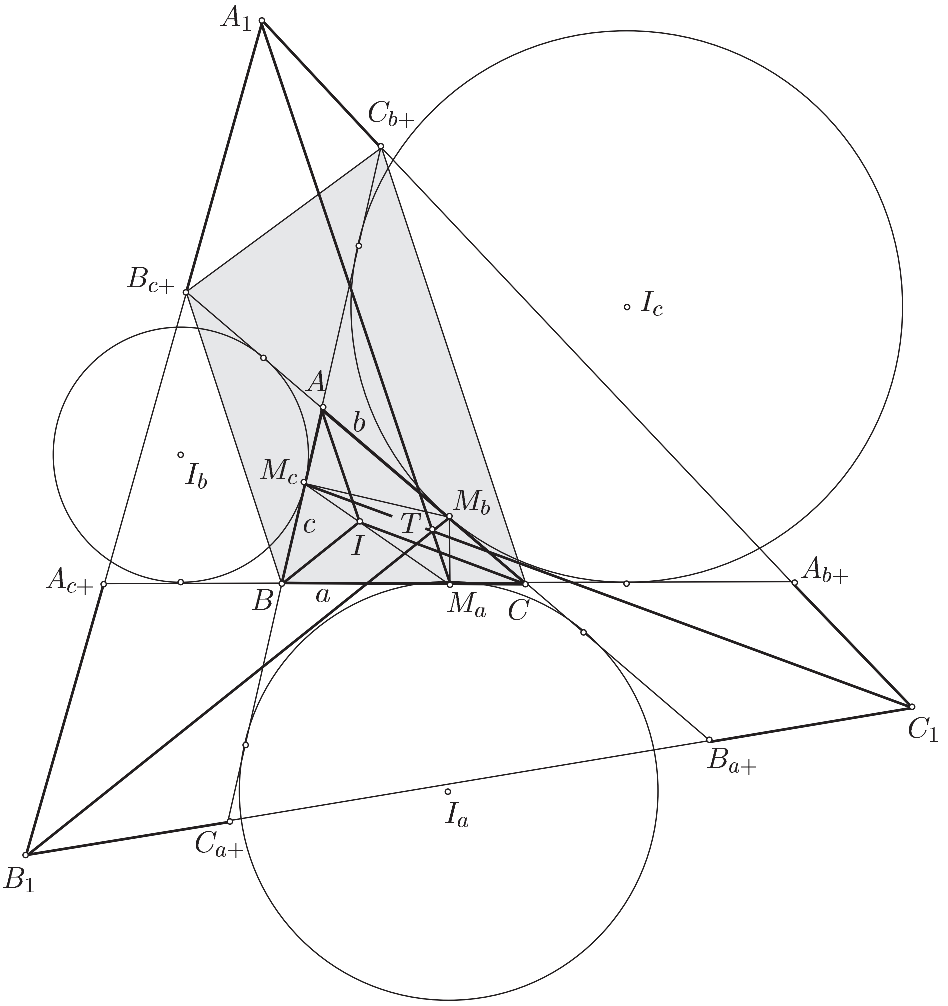

Рис. 13

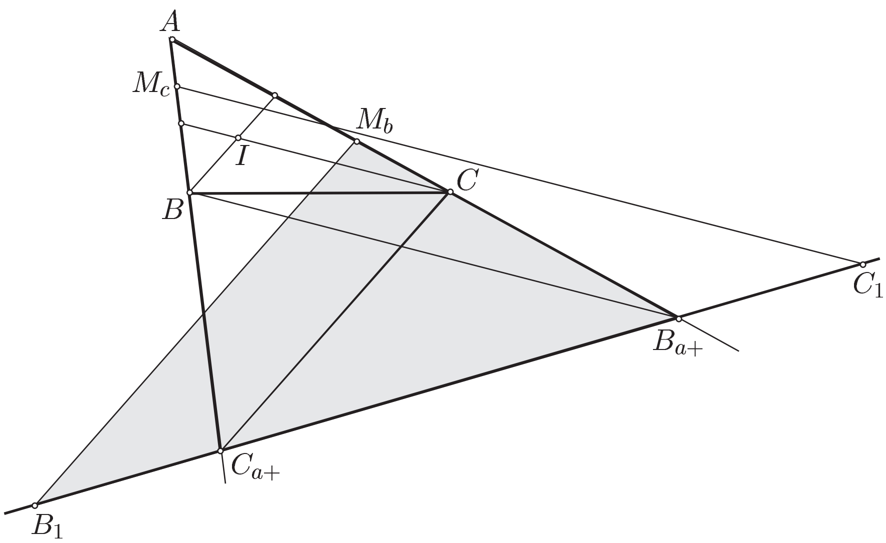

Рис. 14

Тогда из теоремы Фалеса, $\frac{C_{a+}B_{a+}}{B_1C_{a+}} = \frac{a}{s-a} = \frac{C_{a+}B_{a+}}{C_1B_{a+}} \Rightarrow B_1C_{a+} = C_1B_{a+}$. □

Доказательство теоремы 15.2. Проведем через A_1 прямую, параллельную биссектрисе (BI). Пусть N — точка пересечения этой прямой с прямой (CA). На стороне CA отметим точку M_b, так что $AM_b = s-c$, $B_{c+}M_b = s$, где s — полупериметр треугольника ABC. По лемме 15.2, $(B_1M_b)||(BI)$, и по лемме 2.2, $(A_{c+}A)||(BI)$. Следовательно, $(NA_1)||(A_{c+}A)||(B_1M_b)$. Кроме того, по лемме 15.3 $B_1A_{c+} = A_1B_{c+}$.

Покажем теперь, что $NA = s$, или, что то же самое, $NB_{c+} = AM_b$. Действительно, треугольники A_1NB_{c+}, $A_{c+}B_{c+}A$ подобны $\Rightarrow \frac{NB_{c+}}{AB_{c+}} = \frac{A_1B_{c+}}{A_{c+}B_{c+}} = \frac{B_1A_{c+}}{A_{c+}B_{c+}}$.

А так как $(A_{c+}A)||(B_1M_b)$, то, по теореме Фалеса, $\frac{B_1A_{c+}}{A_{c+}B_{c+}} = \frac{AM_b}{AB_{c+}}$. Из этих двух равенств вытекает, что $\frac{NB_{c+}}{AB_{c+}} = \frac{AM_b}{AB_{c+}} \Rightarrow NB_{c+} = AM_b$.

Остается только проверить, что $\frac{NA_1}{NA} = \frac{H'B_0}{AB_0}$.

И правда, если это равенство справедливо, то треугольник A_1NA переходит в треугольник $H'B_0A$ при гомотетии с вершиной A и коэффициентом $-\frac{NA}{AB_0} = -\frac{s}{s-c}$ (поскольку треугольники A_1NA и $H'B_0A$ будут подобны по углу и отношению прилежащих к нему сторон, ведь $\angle A_1NA = \angle AB_0H'$ как накрест лежащие), а треугольник $H'B_0A$, в свой черед, переходит в треугольник H_1RA при гомотетии с вершиной A и коэффициентом $\frac{AB_0}{AR} = \frac{s-c}{s}$ (см. [5], з. 3.2). Окончательно, стало быть, получаем, что треугольник A_1NA переходит в треугольник H_1RA при гомотетии с вершиной A и коэффициентом -1, т. е. эти треугольники симметричны относительно точки A, что нам и требовалось доказать.

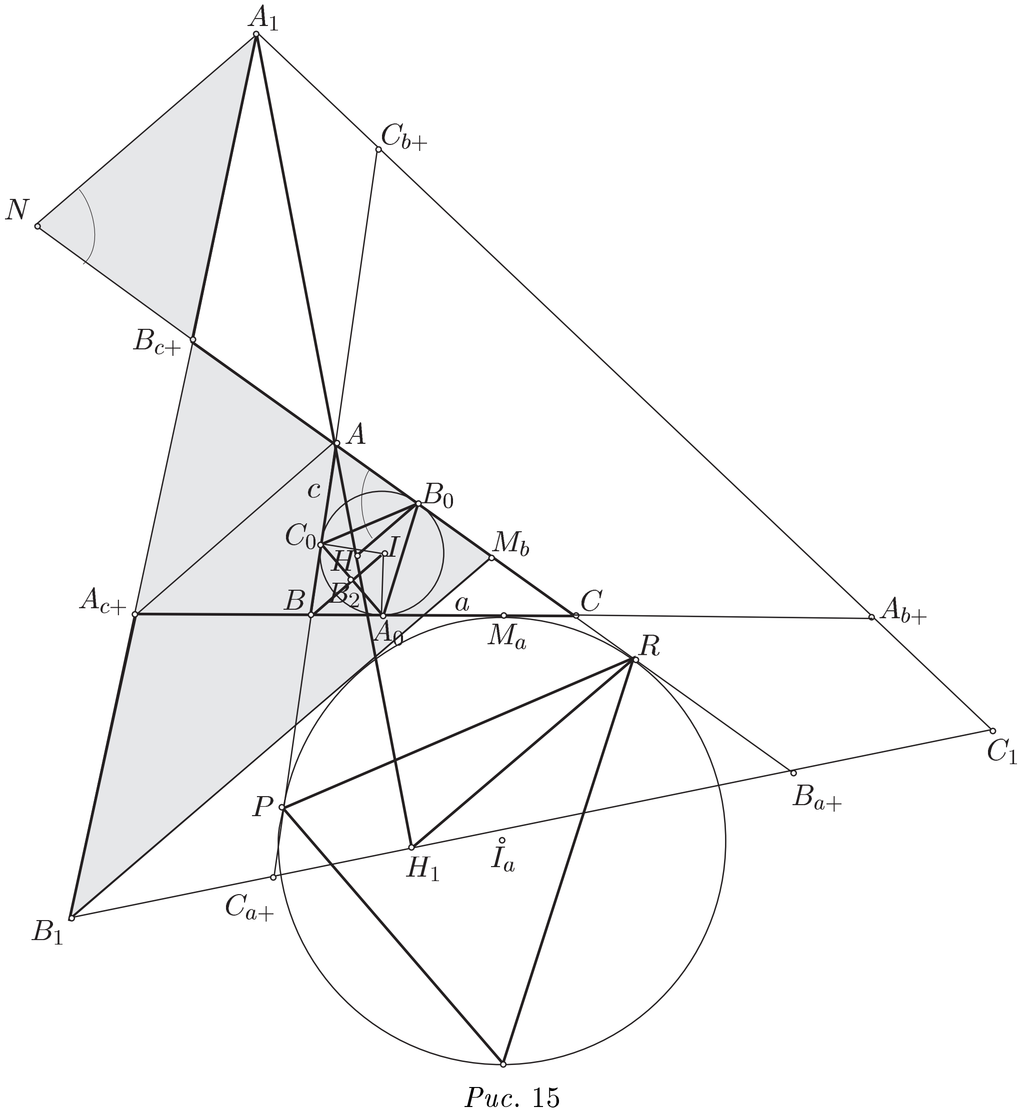

Рис. 15

Итак, покажем, что $\frac{NA_1}{NA} = \frac{H'B_0}{AB_0}$. С одной стороны, из подобия треугольников A_1NB_{c+}, $A_{c+}B_{c+}A$ имеем:

$$NA_1 = \frac{NB_{c+}}{AB_{c+}} \cdot AA_{c+} = \frac{s-c}{c} \cdot 2c\cos\frac{B}{2}$$

(треугольник $A_{c+}BA$ — равнобедренный, с углом $\pi - \angle B$ при вершине). Следовательно,

$$\frac{NA_1}{NA} = 2\cos\frac{B}{2} \cdot \frac{s-c}{s}.$$

С другой стороны, $H'B_0 = 2IB_2$ (т.к. серединный треугольник гомотетичен его порождающему треугольнику с центром в точке пересечения медиан исходного треугольника и коэффициентом $-\frac{1}{2}$). Кроме того, $\angle B_2IA_0 = \frac{1}{2}\angle C_0IA_0 = \frac{1}{2}\cdot(2\angle C_0B_0A_0)$ (по теореме о величине центрального угла) $= \pi - ((\frac{\pi}{2}-\frac{A}{2}) + (\frac{\pi}{2}-\frac{C}{2}))$ (равнобедренность треугольников AC_0B_0 и CA_0B_0) $= \frac{A+C}{2} = \frac{\pi}{2} - \frac{B}{2}$. Поэтому $IB_2 = r\sin\frac{B}{2}$ (где r — радиус вписанной в треугольник ABC окружности) и $\frac{H'B_0}{AB_0} = 2\sin\frac{B}{2}\cdot\frac{r}{s-a}$.

Но $2\cos\frac{B}{2}\cdot\frac{s-c}{s} = 2\sin\frac{B}{2}\cdot\frac{r}{s-a} \Leftrightarrow \tan\frac{B}{2}\cdot sr = (s-c)(s-a)$

Очевидно, что $\tan\frac{B}{2} = \frac{r}{s-b}$, поэтому $\tan\frac{B}{2}\cdot sr = (s-c)(s-a) \Leftrightarrow \frac{sr^2}{s-b} = (s-c)(s-a) \Leftrightarrow (sr)^2 = s(s-a)(s-b)(s-c)$.

Последнее равенство справедливо, ибо слева и справа в нем стоят квадраты площади треугольника ABC (площадь вписанного многоугольника, в частности, треугольника — есть произведение его полупериметра на радиус вписанной окружности, в правой же части возникла знаменитая *формула Герона* — см. [5], з. 12.19)[26].

Добавочные теоремы — «тройняшки» к теоремам Акопяна читателю предлагается сформулировать самостоятельно.

Литература

[1] *Акопян А., Заславский А.* Геометрические свойства кривых второго порядка. М.: МЦНМО, 2007.

[2] *Заславский А.* Геометрические преобразования. М.: МЦНМО, 2003.

[3] *Куланин Е.* Об описанных окружностях чевианных и педальных треугольников и некоторых кривых, связанных с треугольником. // Математическое просвещение. 2005. № 9.

[4] *Мякишев А.* Элементы геометрии треугольника. М.: МЦНМО, 2002.

[5] *Панарин Я.* Элементарная геометрия, том 1. М.: МЦНМО, 2004.

[6] *Прасолов В., Тихомиров В.* Геометрия. М.: МЦНМО, 2007.

[7] *Прасолов В.* Задачи по планиметрии. М.: МЦНМО, 2007.

[8] *Шарыгин И.* Геометрия. Планиметрия. (Задачник 9-11). М.: Дрофа, 2001.

[9] *Grinberg D.* Conway Circle. [http://mathworld.wolfram.com/ConwayCircle.html]

[10] *Kimberling C.* Encyclopedia of triangle centers "ETC". [http://faculty.evansville.edu/ck6/encyclopedia/].

[11] *Yiu P.* Introduction to the Geometry of the Triangle. [http://www.math.fau.edu/yiu/geometry.html]

[12] *Kimberling C.* Triangle centers and central triangles. Winnipeg: Utilitas Mathematica Publ., 1998.

Мякишев Алексей Геннадьевич,
Московский Химический Лицей.

Email: alex_geom@mtu-net.ru

[26]Это заключительное рассуждение — не слишком удачная авторская отсебятина. А можно (и нужно было!) чисто геометрически рассудить (вместо подсчета отношений), по Акопяну: мы уже доказали, что точка N при симметрии относительно A переходит в R, а прямая NA_11 в прямую RH_1. Проведем через A_1 прямую, параллельную биссектрисе угла C и пересечем с AB, получим точку K (аналогичную N). Точно так же $KA = AP$. Поэтому при симметрии относительно точки A отрезок PR перейдет в NK, прямая PH_1 в KA_1, а прямая RH_1 в NA_1. Значит, точка пересечения H_1 — в точку пересечения A_1.

Перманент матрицы и его вычисление

А. Ю. Эвнин

В статье обсуждаются различные способы вычисления перманента матрицы с оценкой их трудоёмкости. Приводится элементарный вывод формулы Райзера для квадратной матрицы.

1. Перманент матрицы

Пусть $A = (a_{ij})$ — числовая матрица размера $m \times n$, где $m \leqslant n$. *Перманент матрицы — сумма произведений её элементов, взятых по одному из каждой строки и не более чем по одному из каждого столбца:*

$$\operatorname{per} A = \sum_{(j_1, j_2, \ldots, j_m)} a_{1\,j_1} a_{2\,j_2} \ldots a_{m\,j_m}. \tag{1}$$

В формуле (1) суммирование ведётся по всем размещениям из n по m (в качестве генеральной совокупности выступает множество $\{1, 2, \ldots, n\}$).

Очевидно, что перманент квадратной матрицы равен сумме произведений её элементов, взятых по одному из каждой строки и каждого столбца. Определитель матрицы получается из перманента сменой знака слагаемых, отвечающих нечётным перестановкам $(j_1, j_2, \ldots, j_m)$.

Укажем некоторые очевидные свойства перманента:

- при перестановке строк и столбцов матрицы её перманент не меняется;
- перманент матрицы является линейной функцией её строк:

$$\operatorname{per} \begin{pmatrix} a_1 \\ \vdots \\ \alpha a_i' + \beta a_i'' \\ \vdots \\ a_m \end{pmatrix} = \alpha \operatorname{per} \begin{pmatrix} a_1 \\ \vdots \\ a_i' \\ \vdots \\ a_m \end{pmatrix} + \beta \operatorname{per} \begin{pmatrix} a_1 \\ \vdots \\ a_i'' \\ \vdots \\ a_m \end{pmatrix},$$

  где $a_1, a_2, \ldots, a_m$ — строки матрицы;
- в случае квадратной матрицы перманент не меняется при её транспонировании.

В отличие от задачи вычисления определителя, задача вычисления перманента является NP-трудной [1]. Так будет даже в случае двоичной матрицы.

Перманент можно вычислить разложением по строке. Разложение по i-й строке имеет вид:

$$\operatorname{per} A = \sum_{j=1}^{n} a_{i\,j} \operatorname{per} A_{i\,j},$$

где A_{ij} — матрица, получающаяся из матрицы A вычёркиванием i-й строки и j-го столбца. В случае квадратной матрицы можно разлагать и по столбцу.

Примеры. Пусть E — матрица размера $n \times n$, все элементы которой единицы, I — единичная матрица того же размера. Тогда

$$\operatorname{per} E = n!, \qquad \operatorname{per}(E - I) = D_n,$$

где D_n — число *беспорядков* (т. е. перестановок чисел $1, 2, \ldots, n$ без неподвижных точек[1]).

Много интересных фактов о перманентах можно найти в монографиях [2–5] и статье [6]. В данной заметке мы остановимся лишь на простейших способах вычисления перманента. Для начала заметим, что при вычислении перманента матрицы размера $m \times n$, исходя из его определения, требуется $A_n^m - 1$ сложений и $A_n^m(m-1)$ умножений.

2. Рекурсивное вычисление перманента

Рассмотрим теперь рекурсивную процедуру вычисления перманента разложением по строке.

Пусть $x_{m,n}$ и $y_{m,n}$ — количество операций соответственно умножения и сложения при указанном способе вычисления перманта матрицы размером $m \times n$.

При вычислении перманента разложением по строке складывается n перманентов матриц размером $(m-1) \times (n-1)$. Поэтому

$$y_{m,n} = n - 1 + n y_{m-1,n-1},$$

откуда

$$y_{m,n} + 1 = n(y_{m-1,n-1} + 1) = n(n-1)(y_{m-2,n-2} + 1) = \cdots = n(n-1)\ldots(n-m+2)(y_{1,n-m+1} + 1).$$

Учитывая, что $y_{1,k} = k - 1$, окончательно получаем

$$y_{m,n} + 1 = n(n-1)\ldots(n-m+2)(n-m+1) = A_n^m,$$

стало быть, $y_{m,n} = A_n^m - 1$. Как видно, количество операций сложения осталось таким же, как при непосредственном вычислении перманента.

Для количества умножений $x_{m,n}$ выполняется рекуррентное соотношение

$$x_{m,n} = n + n x_{m-1,n-1}.$$

Отсюда

$$\frac{x_{m,n}}{n!} = \frac{x_{m-1,n-1}}{(n-1)!} + \frac{1}{(n-1)!}.$$

Последовательно применяя найденное рекуррентное соотношение, получаем

$$\frac{x_{m,n}}{n!} = \frac{1}{(n-1)!} + \frac{1}{(n-2)!} + \cdots + \frac{1}{(n-m+1)!} + \frac{x_{1,n-m+1}}{(n-m+1)!}.$$

Поскольку $x_{1,k} = 0$, приходим к формуле

$$x_{m,n} = n!\left(\frac{1}{(n-1)!} + \frac{1}{(n-2)!} + \cdots + \frac{1}{(n-m+1)!}\right) = A_n^1 + A_n^2 + \cdots + A_n^{m-1}.$$

Количество умножений при рекурсивном вычислении перманента существенно уменьшилось по сравнению с вычислением, исходящим из определения перманента. Например, для квадратной матрицы размера $n \times n$

$$x_{n,n} = n!\left(\frac{1}{(n-1)!} + \frac{1}{(n-2)!} + \cdots + \frac{1}{1!}\right) \approx n!(e-1) \ll n!(n-1).$$

Имеются и более эффективные способы вычисления перманента. Наиболее известный из них открыл Г. Дж. Райзер в середине прошлого века. Этот способ использует формулу включения-исключения.

[1]Некоторые сведения, касающиеся числа беспорядков, содержатся в [7].

3. Обобщённая формула включения-исключения

Пусть A — некоторое конечное множество, $A_1, A_2, \dots, A_n$ — его подмножества. Каждому элементу a множества A поставлено в соответствие число $\mu(a)$ (*вес* элемента a). Определим *вес множества* $B \subset A$ как сумму весов его элементов:

$$\mu(B) = \sum_{b \in B} \mu(b).$$

Зададим функцию $w(n)$ следующими соотношениями:

$$w(0) = \mu(A), \qquad w(k) = \sum_{1 \leqslant i_1 < \dots < i_k \leqslant n} \mu(A_{i_1} \cap \dots \cap A_{i_k}) \qquad (k \leqslant n).$$

Пусть $M(r)$ — сумма весов всех таких элементов множества A, которые принадлежат ровно r различным множествам A_i (если $r = 0$, то — ни одному). Тогда

$$M(r) = w(r) - C_{r+1}^1 w(r+1) + C_{r+2}^2 w(r+2) - \dots + (-1)^{n-r} C_n^{n-r} w(n). \tag{2}$$

Доказательство. Возьмем произвольный элемент множества A и подсчитаем его «вклад» в правую часть доказываемого равенства. Возможны следующие 3 случая.

1) Элемент входит менее чем в r различных множеств A_i, тогда он не принадлежит никакому пересечению r и более упомянутых множеств и вклад его в правую часть (2) равен нулю.

2) Элемент принадлежит *ровно* r множествам A_i, тогда он принадлежит ровно одному пересечению r множеств и не принадлежит никакому пересечению большего числа множеств; таким образом, вклад элемента a равен $\mu(a)$.

3) Элемент принадлежит k множествам A_i, причем $k > r$. Тогда он входит в C_k^r пересечений r множеств A_i, в C_k^{r+1} пересечений $r+1$ множеств A_i, в C_k^j пересечений j множеств A_i $(j \leqslant k)$. Подсчитаем, сколько раз элемент учитывается в правой части формулы (2). Имеем[2]:

$$\sum_{j=r}^{k} (-1)^{j-r} C_j^{j-r} C_k^j = \sum_{j=r}^{k} (-1)^{j-r} C_j^r C_k^j = \sum_{j=r}^{k} (-1)^{j-r} C_k^r C_{k-r}^{j-r} =$$

$$= [i = j - r] = C_k^r \sum_{i=0}^{k-r} (-1)^i C_{k-r}^i = 0.$$

Таким образом, формула из правой части (2) дает суммарный вес тех элементов A, которые входят ровно в r подмножеств A_i. Это и требовалось доказать. □

4. Формула Райзера

Пусть $A = (a_{ij})$ — числовая матрица размера $m \times n$, где $m \leqslant n$, X — множество строк этой матрицы, а Y — множество столбцов. Рассмотрим множество функций $U = \{f : X \to Y\}$. Определим вес функции

$$\mu(f) = a_{1\,f(1)} a_{2\,f(2)} \dots a_{m\,f(m)}.$$

Несложно видеть, что перманент матрицы A равен сумме весов всех инъекций.

[2] В последующих преобразованиях используются известные комбинаторные тождества:

$$C_j^r = C_j^{j-r}; \qquad C_k^j C_j^r = C_k^r C_{k-r}^{j-r}; \qquad \sum_{i=0}^{n} (-1)^i C_n^i = 0.$$

Доказательство этих (и многих других) тождеств с помощью комбинаторных рассуждений см. в [7] и [8].

Возможность применения обобщённой формулы включения-исключения (2) для вычисления перманента определяется следующим фактом. Зафиксируем некоторое подмножество столбцов $I \subset Y$ и рассмотрим произведение строчных сумм соответствующей подматрицы матрицы A:

$$S(I) = \prod_{i=1}^{m} \sum_{j \in I} a_{ij}. \tag{3}$$

При раскрытии скобок в формуле (5) возникают веса всех функций $f : X \to I$. Поэтому $S(I)$ — сумма весов всех функций, действующих из X в I.

Введём в рассмотрение множества $A_i = \{f \in U \mid i \notin f(X)\}$. Функция f будет инъекцией тогда и только тогда, когда она принимает ровно m значений, или **не принимает** ровно $n-m$ значений, то есть входит ровно в $n-m$ множеств A_i. Значит, сумма весов всех инъекций равна $M(n-m)$ (в обозначениях предыдущего параграфа):

$$\operatorname{per} A = M(n-m) = w(n-m) - C_{n-m+1}^{1} w(n-m+1) + \dots + (-1)^m C_n^m w(n), \tag{4}$$

где

$$w(k) = \sum_{1 \leqslant i_1 < i_2 < \dots < i_k \leqslant n} \mu(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_k}) = \sum_{I \subset Y,\ |I|=k} S(Y \setminus I) = \sum_{I \subset Y,\ |I|=n-k} S(I).$$

5. Перманент квадратной матрицы

В случае квадратной матрицы (при $m = n$) формула (4) принимает вид:

$$\operatorname{per} A = \sum_{k=0}^{n} (-1)^k w(k) = \sum_{k=0}^{n} (-1)^k \sum_{|I|=n-k} S(I) = \sum_{I \subset Y} (-1)^{n-|I|} S(I).$$

Например,

$$\operatorname{per} \begin{pmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{pmatrix} = (a_1+b_1+c_1)(a_2+b_2+c_2)(a_3+b_3+c_3) -$$
$$- (a_1+b_1)(a_2+b_2)(a_3+b_3) - (a_1+c_1)(a_2+c_2)(a_3+c_3) - (b_1+c_1)(b_2+c_2)(b_3+c_3) +$$
$$+ a_1a_2a_3 + b_1b_2b_3 + c_1c_2c_3.$$

Итак, в результате довольно длинных выкладок получена элегантная формула вычисления перманента квадратной матрицы размера $n \times n$:

$$\operatorname{per} A = \sum_{I \subset Y} (-1)^{n-|I|} S(I). \tag{5}$$

Приведём *элементарное доказательство* формулы (5).

Подсчитаем, сколько раз произведение $t = a_{1\,i_1} a_{2\,i_2} \dots a_{n\,i_n}$ учитывается в правой части формулы (5). Если все индексы $i_1, i_2, \dots, i_n$ разные, произведение t появляется только при вычислении $S(Y)$ — один раз. Пусть теперь среди указанных индексов ровно k различных, где $k < n$. Тогда количество способов добавить к соответствующим k столбцам, образующим множество I, чётное число столбцов из оставшихся $n-k$ столбцов равно количеству способов добавить нечётное число столбцов (поскольку в непустом конечном множестве число подмножеств чётной мощности равно числу подмножеств нечётной мощности). Значит, сколько раз произведение t будет взято с плюсом, столько же раз и с минусом. Стало быть, вклад такого произведения в правую часть формулы (5) равен нулю. Таким образом, правая часть формулы (5) равна сумме произведений элементов матрицы, взятых по одному из каждой строки и каждого столбца, т. е. перманенту матрицы. □

6. Трудоёмкость вычисления по формуле Райзера

Подсчитаем трудоёмкость вычисления перманента матрицы $n \times n$ по формуле Райзера (5).

Всего имеется $2^n - 1$ непустых подмножеств столбцов I, а вычисление $S(I)$ требует $n-1$ умножений. Таким образом, общее количество умножений равно $(n-1)(2^n-1)$, что много меньше, чем $n!(e-1)$. В книге [9] приводится способ вычисления по формуле Райзера, позволяющий уменьшить указанное число умножений примерно вдвое.

Для минимизации количества сложений при вычислении по формуле (5) поступим следующим образом. Будем перебирать подмножества I множества столбцов Y в таком порядке, чтобы очередное множество получалось из предыдущего добавлением или удалением одного элемента[3]. Если начать с одноэлементного I, то вычисление $S(I)$ не требует сложений, а при вычислении каждого последующего $S(I)$ нужно n сложений. Учитывая, что суммируются (с разными знаками) $2^n - 1$ значений $S(I)$, окончательно получаем $(2^n - 2)(n+1)$ операций сложения (к коим мы относим и вычитание!).

Литература

[1] Гэри, М. *Вычислительные машины и труднорешаемые задачи* / М. Гэри, Д. Джонсон. — М.: Мир, 1982. – 416 с.

[2] Сачков, В. Н. *Комбинаторные методы дискретной математики* / В. Н. Сачков. — М.: Наука, 1977. – 320 с.

[3] Рыбников, К. А. *Введение в комбинаторный анализ* / К. А. Рыбников. — М.: Изд-во МГУ, 1985. – 312 с.

[4] Минк, Х. *Перманенты* / Х. Минк. — М.: Мир, 1982. – 213 с.

[5] Ловас, Л. *Прикладные задачи теории графов* / Л. Ловас, М. Пламмер. — М.: Мир, 1998. – 653 с.

[6] Robertson, N. *Permanents, Pfaffian orientations, and even directed circuits* / N. Robertson, P. D. Seymour, R. Thomas. // Annals of Mathematics, **150** (1999), pp. 929–975.

[7] Краснов, М. Л. *Вся высшая математика:* учебник. Т.7 / М. Л. Краснов, А. И. Киселёв, Г. И. Макаренко и др. — М.: КомКнига, 2006. – 208 с.

[8] Эвнин, А. Ю. *Две заметки по комбинаторике* / А. Ю. Эвнин. // Математическое образование. – 2000. – № 3(14). – С. 27–34.

[9] Nijenhuis, A. *Combinatorial algorithms* / A. Nijenhuis, H. S. Wilf. — New York: Academic press, 1978. – 302 p.

[10] Липский, В. *Комбинаторика для программистов* / В. Липский. — М.: Мир, 1988. – 213 с.

Эвнин Александр Юрьевич,
кандидат педагогических наук,
доцент кафедры прикладной математики
Южно-Уральского Государственного Университета.

Email: evnin@prima.susu.ac.ru

[3]Такую последовательность множеств называют *кодом Грея.* Алгоритм её получения можно найти в книге [10].

Выравнивание статистических рядов. Проверка правдоподобия гипотез

И. П. Костенко

Предлагаемую лекцию, посвященную введению в основные понятия и методы математической статистики, можно рассматривать как дополнение к курсу "Введение в вероятностное прогнозирование". Курс был опубликован в журнале "Математическое образование", №№ 2(21)–4(23); 2(25)–2(29) и вышел отдельным изданием в издательстве Института компьютерных исследований, Москва-Ижевск, 2004. Настоящая лекция написана автором уже после выхода книги в свет.

Наш курс назван «Введение в вероятностное прогнозирование», т. е. введение в науку, цель которой — прогнозы, т.е. объективно обоснованные предсказания о реальных массовых случайных явлениях. Само исходное понятие вероятности возникло из анализа эмпирических данных и обладает прогностической силой: если вероятность события равна 0,7, то можно быть уверенным, что при 100 повторениях опыта событие появится примерно 70 раз. Это есть вероятностный прогноз. И практика убедительно подтверждает правильность и ценность такого рода прогнозов.

Понятия и теоретические факты, с которыми вы познакомились в курсе, являются, в сущности, абстрактным математическим отображением реальных отношений и закономерностей, наблюдаемых в массовых случайных явлениях. Они служат той же конечной цели — выработке методов прогнозирования в разнообразных реальных ситуациях. К примеру, теоремы сложения и умножения и связанные с ними понятия дают метод вычисления вероятностей сложных событий (в частности, по формуле Бернулли). Центральная предельная теорема позволяет предвидеть нормальный тип распределения случайных величин, очень часто возникающих в практической работе инженера и исследователя.

На базе теории вероятностей строится научная дисциплина, которая называется «математическая статистика» и которая является прикладным разделом теории вероятностей. Обе дисциплины тесно связаны и в нашем курсе органически переплетаются. В этой последней лекции мы сделаем обзор типовых задач математической статистики и сосредоточимся на одной из важнейших — задаче проверки гипотез.

1. Начальные задачи математической статистики

Математической статистикой называют науку, которая разрабатывает методы обработки и анализа опытных (статистических) данных. Её цель — решение конкретных статистических задач, выработка общих методов решения этих задач, в сущности, опять же, методов вероятностного прогнозирования.

Исходными данными для применения методов математической статистики является совокупность $\{x^{(1)}, x^{(2)}, \ldots, x^{(k)}\}$ значений случайной величины X, которые она приняла в результате эксперимента — многократного повторения опыта (число повторений опыта k, как правило, весьма велико, порядка нескольких сотен и больше). Такие данные мы назвали *простой статистической совокупностью* (лек. 6, п. 1), — они представляют собой хаотичный набор чисел, глядя на который очень трудно заметить какие-то закономерности. Но закономерности существуют, они скрыты в данных и их надо извлечь. Как?

Первая задача. В результате эксперимента над с. в. X получена простая статистическая совокупность $\{x^{(1)}, x^{(2)}, \ldots, x^{(k)}\}$. Как следует преобразовать её, чтобы сделать более обозримой и удобной для первых выводов о характере изучаемой с. в. X? Какими формами таблиц, диаграмм лучше всего воспользоваться?

Решение. Первое представление данных даёт *статистический ряд* (лек. 6, п. 1) – это таблица 1, в которой указаны *различные* значения с. в. X, появившиеся в эксперименте и расположенные в порядке возрастания $x_1 < x_2 < \ldots < x_m$, а также указаны их частоты l_i и относительные частоты $p_i^* = l_i/k$.

$$\begin{array}{rcccc} x_i: & x_1 & x_2 & \ldots & x_m \\ l_i: & l_1 & l_2 & \ldots & l_m \\ p_i^* = l_i/k: & p_1^* & p_2^* & \ldots & p_m^* \end{array}$$

Таблица 1.

Если число различных значений x_i велико, то статистический ряд не составляют, а применяют к данной статистической совокупности $\{x^{(1)}, x^{(2)}, \ldots, x^{(k)}\}$ *метод группировки* (лек. 6, п. 2), который преобразует её в таблицу 2 — *группированный статистический ряд* (*интервальный*). Число интервалов группировки будет, конечно, значительно меньше числа значений x_i в таблице 1, но для удобства применения дальнейших формул оставим то же обозначение m.

$$\begin{array}{rcccc} x_i: & x_1 \div x_2 & x_2 \div x_3 & \ldots & x_m \div x_{m+1} \\ l_i: & l_1 & l_2 & \ldots & l_m \\ p_i^* = l_i/k: & p_1^* & p_2^* & \ldots & p_m^* \end{array}$$

Таблица 2.

Напомню последовательность действий для построения таблицы 2. Все появившиеся в опыте значения $x^{(1)}, x^{(2)}, \ldots, x^{(k)}$ заключаются в интервал $(x_1; x_{m+1})$; этот интервал разбивается на m мелких промежутков (они указаны в верхней строке таблицы 2); подсчитываются частоты l_i попадания значений с. в. в каждый промежуток (если x_i попадает на границу двух промежутков, мы условились относить его к правому); вычисляются относительные частоты $p_i^* = l_i/k$ и записываются в нижнюю строку таблицы.

Таблица 2 сжимает информацию, заключённую в статистическом материале: если число данных k — несколько сотен, то число m колонок таблицы (число промежутков разбиения) обычно 10–20. Нижняя строка таблицы показывает, в каких промежутках значения с. в. возникают часто, в каких редко, и насколько реже. Это есть первые выводы о распределении значений с. в. X.

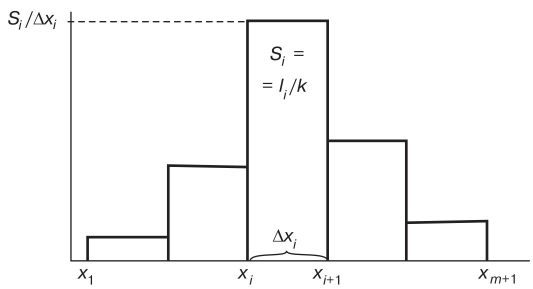

Рис. 1.

Таблицу 2 можно представить в наглядной форме, в виде *гистограммы* (рис. 1). Строится гистограмма так: на горизонтальной оси откладываются промежутки разбиения и на каждом из них строится «столбик», площадь которого равна соответствующей относительной частоте. Высоты столбиков *наглядно* показывают, насколько чаще попадают в тот или иной промежуток значения с. в. X.

Ещё одна удобная форма представления статистических данных — *дискретный группированный ряд* (таблица 3). Он получается заменой в верхней строке таблицы 2 промежутков $x_i \div x_{i+1}$ их средними значениями $\tilde{x}_i = (x_{i+1} + x_i)/2$.

$$\begin{array}{rcccc} \tilde{x}_i: & \tilde{x}_1 & \tilde{x}_2 & \ldots & \tilde{x}_m \\ l_i: & l_1 & l_2 & \ldots & l_m \\ p_i^* = l_i/k: & p_1^* & p_2^* & \ldots & p_m^* \end{array}$$

Таблица 3.

В сущности, таблица 3 есть дискретная (приближённая) модель $\tilde{X}$ данной с. в. X, причём, весьма простая, ибо $m \ll k$. Эту модель мы будем полагать в качестве исходных данных при решении задач, связанных с непрерывной с. в.

Вторая задача. По данному статистическому ряду или дискретному группированному ряду оценить (приближённо) числовые характеристики с. в. X: математическое ожидание M, дисперсию D, среднее квадратическое отклонение σ.

Решение дают формулы (1)–(3) (лек. 6, п. 4–6). Напомню их, применительно к дискретному группированному ряду (таблица 3):

Оценка *математического ожидания*:

$$\tilde{M}^* = \tilde{x}_1 \cdot l_1 + \tilde{x}_2 \cdot l_2 + \dots + \tilde{x}_m \cdot l_m. \tag{1}$$

Оценка *дисперсии*:

$$\tilde{D}^* = \frac{\left(\tilde{x}_1 - \tilde{M}^*\right)^2 + \left(\tilde{x}_2 - \tilde{M}^*\right)^2 + \ldots + \left(\tilde{x}_m - \tilde{M}^*\right)^2}{m}. \tag{2}$$

Оценка *среднего квадратического отклонения*:

$$\tilde{\sigma}* = \sqrt{\tilde{D}^*}. \tag{3}$$

Если оценки вычисляются по статистическому ряду, то формулы остаются такими же, лишь вместо усреднённых значений $\tilde{x}_i$ ставятся значения x_i, взятые из таблицы 1.

Третья задача. Оценить степень *точности* и *надёжности* оценок числовых характеристик с. в. X, т. е. построить доверительные интервалы и найти доверительные вероятности попадания истинных значений числовых характеристик в эти интервалы.

Решение этой задачи для математического ожидания M и дисперсии D проведено в предыдущей лекции (п. 5, 6). Напомню понятия и формулы.

Доверительным интервалом для оценки M^* математического ожидания M называется интервал $(M^* - \varepsilon; M^* + \varepsilon)$, который покрывает истинное значение M с вероятностью p_ε. Вероятность p_ε называется *доверительной вероятностью*, или *надёжностью* оценки, число ε — *точностью* оценки. Для дисперсии (и любой другой характеристики) определение аналогично.

Формула для расчёта надёжности оценки $\tilde{M}^*$ при требуемой точности ε:

$$p_\varepsilon(\tilde{M}^*) = P\left(\left|\tilde{M}^* - M\right| < \varepsilon\right) \approx 2\Phi\left(\frac{\varepsilon}{\sigma\left(\tilde{M}^*\right)}\right), \qquad \sigma\left(\tilde{M}^*\right) \approx \sqrt{\frac{\tilde{D}^*}{k}}. \tag{4}$$

Это, в сущности, формула (9) из лек. 12, только приспособленная для оценки $\tilde{M}^*$ по дискретному группированному ряду.

Формула для расчёта надёжности оценки дисперсии $\tilde{D}^*$ (см. формулу (9) лекции 12) аналогична предыдущей, но справедлива только для нормальных с. в. X:

$$p_\varepsilon(\tilde{D}^*) = P\left(\left|\tilde{D}^* - D\right| < \varepsilon\right) \approx 2\Phi\left(\frac{\varepsilon}{\sigma\left(\tilde{D}^*\right)}\right), \qquad \sigma\left(\tilde{D}^*\right) \approx \tilde{D}^* \cdot \sqrt{\frac{2}{k-1}}. \tag{5}$$

Примечание 1. Заметьте, — в конце обеих формул в знаменателе стоит не m (число значений дискретной модели с. в. X, по которым рассчитывались оценки $\tilde{M}^*$ и $\tilde{D}^*$), а k — число значений первичной статистической совокупности $\{x^{(1)}, x^{(2)}, \ldots, x^{(k)}\}$.

Примечание 2. При практических применениях формул (4) и (5) можно ставить в них другие виды оценок, например, оценки M^* и D^{**}, полученные из первичной статистической совокупности.

Примечание 3. С помощью формул (4) и (5) можно решать обратные задачи: а) по заданной надёжности p_ε определить точность ε (лек. 12, п. 6, пример 8); б) определить число опытов k, позволяющих при заданной точности ε добиться лучшей надёжности p_ε (лек. 12, п. 5, пример 6).

Контроль 1. Над с. в. X проведено 25 опытов и получена простая статистическая совокупность {38, 41, 41, 42, 39, 38, 41, 44, 45, 41, 44, 42, 41, 40, 40, 42, 45, 41, 44, 43, 43, 37, 40, 39, 41}. Проведите группировку, взяв за основной интервал $(x_1; x_{m+1}) = (36; 46)$ и разбив его на 5 мелких промежутков. Постройте гистограмму. Составьте дискретный группированный ряд. Вычислите оценки $\tilde{M}^*$ и $\tilde{D}^*$. Рассчитайте надёжность этих оценок при заданной точности $\varepsilon = 0,2$[1].

[1] Не забудьте, что формула (5) справедлива только для нормальных с. в. X, и применяя её к данной задаче, вы рискуете сделать ошибку. Тем не менее, применяйте, сознавая, что это учебная, тренировочная задача.

2. Подбор функции-плотности

Итак, статистический материал представлен группированным статистическим рядом (таблица 1) и гистограммой (рис. 1). С помощью этого ряда мы извлекаем из эмпирических данных первую важную информацию об исследуемой с. в. — оценки числовых характеристик, их точность и надёжность.

Следующая задача — подобрать функцию-плотность (лек. 9, п. 6, определения 1, 2) и оценить степень точности этого подбора. Разумеется, задача ставится только для непрерывных с. в., так как дискретные не имеют функции-плотности (см. лек. 9, п. 6, дополнение 2).

Задача эта возникала раньше (лек. 9, п. 4, 5, лек. 11, п. 6). Напомню кратко смысл. Если проводить новые эксперименты над с. в. (непрерывной!), увеличивая число опытов и измельчая интервалы группировки, то верхние основания гистограмм (графики статистических плотностей $f^*(x)$) будут приближаться к некоторой плавной линии — графику функции-плотности $f(x)$ исследуемой с. в. (рис. 2) и в пределе сольются с ней. Эту функцию и надо подобрать.

Она не зависит от случайностей эксперимента и концентрирует существенные черты изучаемого случайного явления, т. е. является его вероятностной моделью. Процесс подбора функции-плотности называют *выравниванием статистических рядов* (гистограмм).

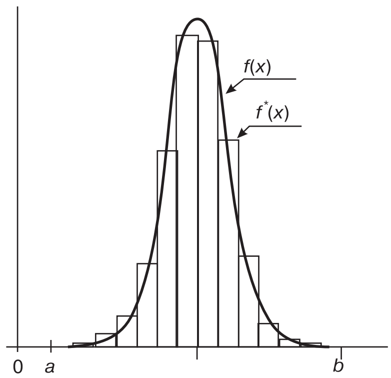

Рис. 2.

Некоторые соображения относительно подбора функции-плотности для показательных и для нормальных распределений были сделаны ранее (лек. 9, п. 5 и лек. 11, п. 6). Но теперь надо иметь в виду, что мы изучили различные распределения, знаем типы возможных предельных функций и условия их возникновения (поток событий, центральная предельная теорема и др.), а значит, понимаем, когда можно ожидать то или иное распределение. Кроме того, есть много других типов с. в., хорошо изученных (распределения Коши, Вейбулла, Эрланга, Фишера и пр., и пр.). Этот багаж знаний и помогает решать задачу.

Итак, поставим задачу чётко и опишем программу её решения. При этом разделим её на две части — подбор функции и оценку точности этого подбора. Вторую, более сложную, перенесём в последующие разделы лекции.

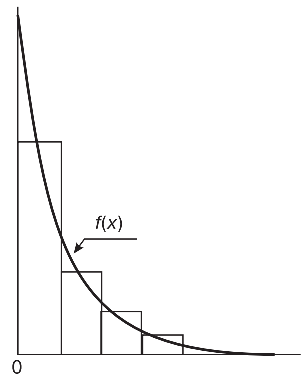

Рис. 3.

Четвёртая задача (*выравнивание статистических рядов*). Изучается непрерывная случайная величина X. Проведена серия опытов и получен статистический материал. Сделана его первичная обработка и он представлен в виде группированного статистического ряда (интервального и дискретного — таблицы 2 и 3). Построена гистограмма (рис. 1). Требуется подобрать вид функции-плотности $f(x)$ с. в. X и определить входящие в неё параметры, после чего построить график полученной функции $f(x)$ и сравнить его с гистограммой (хорошо ли, на взгляд, выравнивается гистограмма?)

Решение. 1. Как было сказано, вид функции-плотности определяется по виду гистограммы с учётом условий возникновения исследуемой с. в. и багажа теоретических знаний. Так, например, гистограмма, показанная на рис. 1 и 2, наводит на мысль от нормальном распределении, на рис. 3 — о показательном. В первом случае выберем известную Гауссовскую плотность

$$f\left(x\right) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{(x-a)^2}{2\sigma^2}} \tag{6}$$

во втором — показательную

$$f\left(x\right) = \lambda \cdot e^{-\lambda x}, \qquad x \in [0; +\infty)\,. \tag{7}$$

Если функция подбирается не из известных типовых классов, то следует проверять выполнение основных свойств функции-плотности:

$$f(x) \geq 0; \qquad \int_a^b f(x)dx = 1. \tag{8}$$

2. Для подбора значений параметров (a и σ в первом случае, λ — во втором) есть разные методы. Чаще пользуются *методом моментов*, выбирая параметры так, чтобы важнейшие моменты (математическое ожидание и дисперсия) полученной теоретической модели совпадали со статистическими моментами, т. е. $\mu_1 = M \approx M^*$ и $\mu_2 = D \approx D^*$. Иногда учитывают моменты 3-го и 4-го порядков (скошенность и эксцесс): $\mu_3 \approx \mu_3^*$ и $\mu_4 \approx \mu_4^*$[2].

В случае нормального или показательного распределений дело облегчается, поскольку первые моменты очень просто связаны с параметрами. У нормального распределения $\mu_1 = M = a$ и $\mu_2 = D = \sigma^2$ (лек. 11, п. 7, формулы (15), (16)), значит, следует положить

$$a \approx \tilde{M}^* \qquad \text{и} \qquad \sigma \approx \sqrt{\tilde{D}^*}, \tag{9}$$

где $\tilde{M}^*$ и $\tilde{D}^*$ вычисляются по формулам (1) и (2). У показательного полагаем (лек. 11, п. 5, формула (12)):

$$\lambda \approx 1/\tilde{M}^*. \tag{10}$$

3. Подставляем в выбранный тип функции-плотности найденные параметры и получаем конкретную функцию $f(x)$. Вычисляем (с помощью инженерного калькулятора) значения этой функции в граничных точках промежутков группированного ряда (таблица 2): $f(x_1)$, $f(x_2), \ldots$, $f(x_m)$, $f(x_{m+1})$. Строим график функции $f(x)$ в той же координатной системе, в которой построена гистограмма. Сравниваем гистограмму с построенной кривой и отвечаем на вопрос: достаточно ли хорошо (на взгляд!) кривая выравнивает гистограмму? Если плохо, значит, тип функции выбран неудачно (или сделана вычислительная ошибка).

Примечание 1. Если распределение нормальное, то для вычисления значений $f(x_i)$ можно воспользоваться таблицей функции Гаусса.

Примечание 2. В данной задаче предлагалось оценить удачность подбора функции-плотности субъективно. В дальнейшем (пятая задача математической статистики) сделаем это объективно, используя критерий Пирсона.

Контроль 2. С помощью нормального закона выровняйте гистограмму, полученную при выполнении контрольного упражнения 1[3]. Удачен ли оказался подбор функции-плотности?

3. Хи-квадрат распределение

Вы знаете много различных типов с. в. — биномиальные, Пуассоновские, показательные, нормальные и др. Все они возникли из реальных задач и представляют типовые вероятностные модели реальных случайных явлений. Но есть с. в., которые возникают в теоретических исследованиях как необходимый инструмент для решения статистических задач. С одним из таких типов с. в. вы сейчас познакомитесь. В следующем разделе узнаете, как он решает важнейшую статистическую задачу проверки гипотез.

Определение 1. Пусть с некоторым опытом связано r независимых[4], одинаково распределённых нормальных с. в. $X_1, X_2, \ldots, X_r$ с одинаковыми математическими ожиданиями, равными нулю, и дисперсиями, равными единице:

$$M(X_i) = 0; \qquad D(X_i) = \sigma(X_i) = 1 \quad (i = 1, 2, \ldots, r).$$

[2]Моментами выше 4-го порядка не пользуются, потому что резко падает точность их вычисления с помощью статистического ряда.

[3]Строго говоря, эта гистограмма построена для дискретной с. в., для которой не существует функции-плотности. К тому же, вид гистограммы не очень соответствует нормальному закону (лек. 6, рис. 1). Но мы используем её в качестве простого упражнения.

[4]*Независимость с. в.* означает, что независимы все связанные с ними события (лек. 7, п. 8). Это свойство необходимо для доказательств (лек. 7, п. 9).

Проделаем над данными с. в. следующие операции: возведём каждую из них в квадрат и сложим. Получим новую с. в.

$$\chi^2 = X_1^2 + X_2^2 + \ldots + X_r^2 = \sum_{i=1}^{r} X_i^2. \tag{11}$$

Эта с. в. обозначается греческой буквой χ (хи), возведённой в квадрат, и называется «χ^2-*распределением*».

Заметим, что данное определение вводит не одну с. в., а целый класс случайных величин, зависящих от одного параметра r — числа слагаемых. Параметр этот называют *числом степеней свободы* χ^2-распределения.

У вас, наверное, возникает вопрос — зачем? Зачем возводить в квадрат и складывать? Ответ найдёте в следующем разделе лекции, когда увидите, как с. в. χ^2 станет решать пятую задачу. А пока разберёмся с её конструкцией.

Смысл операций. Что значит «возвести с. в. X_i в квадрат»? Это значит, что в том же опыте, с которым связана с. в. X_i, вводится новая с. в. X_i^2 — такая, что если в результате опыта появится значение с. в. X_i, равное x_i, то в этом же опыте появится значение с. в. X_i^2, равное x_i^2 (лек. 5, п. 7).

Что значит «сложить случайные величины X_i^2 $(i = 1, 2, \ldots, m)$»? Это значит, надо сложить их появляющиеся в опыте значения x_i^2 (лек. 10, п. 4). Следовательно, значения x с. в. χ^2 в данном опыте складываются из значений с. в. X_i^2:

$$x = \sum_{i=1}^{r} x_i^2.$$

Итак, конструкция с. в. χ^2 понятна. Теперь надо понять, что же, в результате, мы получаем? Каковы характерные особенности с. в. χ^2?

Область значений. Согласно правилу трёх сигм (лек. 11, п. 8), областью практически возможных значений каждой нормально распределённой с. в. X_i является интервал $(M - 3\sigma; M + 3\sigma)$, т. е. в нашем случае интервал $(-3; 3)$. Следовательно, областью значений каждой с. в. X_i^2 и, в частности, первой X_1^2, можно считать промежуток $[0; 9)$. Второе χ^2-распределение $X_1^2 + X_2^2$ (параметр $r = 2$) расширяет область значений в два раза — до промежутка $[0; 18)$, третье $X_1^2 + X_2^2 + X_3^2$ — в три раза, до $[0; 27)$, и т. д. Поскольку область значений расширяется, то естественно предположить, что с ростом параметра r математические ожидания будут сдвигаться вправо, а дисперсии увеличиваться.

Распределение вероятностей. Задачу отыскания функции-плотности для χ^2-распределения[5] мы сами решить не сможем. Но задача эта решена и мы воспользуемся результатом [1, п. 9.7, задача 2, формула (9.7.5)], согласно которому функция-плотность с. в. χ^2 имеет вид:

$$f_r(x) = \frac{1}{2^{r/2} \cdot \Gamma\left(r/2\right)} \cdot x^{\frac{r}{2}-1} \cdot e^{-\frac{x}{2}}, \qquad x \in (0; +\infty), \tag{12}$$

где $\Gamma(\alpha) = \int\limits_0^\infty x^{\alpha-1} e^{-x} dx$ $(\alpha \geq 1)$ — известная *гамма-функция*, значения которой при целых $\alpha = r = 1, 2, 3, \ldots$ получаются очень просто: $\Gamma(r) = 1 \cdot 2 \cdot 3 \cdot \ldots \cdot (r-1) = (r-1)!$[6] (интеграл вычисляется применением $r - 1$ раз формулы интегрирования по частям).

Формула (12) , наверное, кажется вам сложной и непонятной. Но обратите внимание: дробь в этой формуле — числовой коэффициент, который обеспечивает выполнение основного свойства функции-плотности (первая формула (8)). Если обозначить этот коэффициент через A_r, то формула примет вид:

$$f_r(x) = A_r \cdot x^{\frac{r}{2}-1} \cdot e^{-\frac{x}{2}}, \quad x \in (0; +\infty), \quad A_r = 1/(2^{r/2} \cdot \Gamma(r/2)). \tag{12$'$}$$

[5]Эта задача является частным случаем общей задачи отыскания законов распределения разнообразных функций от случайных величин, которой посвящена вся 9-я глава книги [1].

[6]*Гамма-функция* $\Gamma(\alpha)$ обобщает понятие факториала на любые положительные числа так: $\alpha! = \Gamma(\alpha + 1)$. Например, $\left(\frac{1}{2}\right)! = \Gamma\left(\frac{3}{2}\right) = \int\limits_0^\infty \sqrt{x} \cdot e^{-x} dx = \frac{\sqrt{\pi}}{2}$.

Выпишем функции-плотности, придавая параметру значения $r = 1, 2, 3, 4, 5$:

$$
\begin{aligned}
&r = 1, \qquad && f_1(x) = A_1 \cdot x^{-\frac{1}{2}} \cdot e^{-\frac{x}{2}};\\
&r = 2, && f_2(x) = A_2 \cdot e^{-\frac{x}{2}};\\
&r = 3, && f_3(x) = A_3 \cdot \sqrt{x} \cdot e^{-\frac{x}{2}};\\
&r = 4, && f_4(x) = A_4 \cdot x \cdot e^{-\frac{x}{2}};\\
&r = 5, && f_5(x) = A_5 \cdot x \cdot \sqrt{x} e^{-\frac{x}{2}}.
\end{aligned}
$$

Теперь вы видите, что формулы имеют простую структуру, и все они, начиная со второй, получаются из предыдущих умножением на $\sqrt{x}$ (без учёта изменения коэффициентов). Среди них ($r = 2$, $A_2 = \frac{1}{2^{2/2} \cdot \Gamma(2/2)} = \frac{1}{2 \cdot 0!} = \frac{1}{2}$), находится показательное распределение $f(x) = \lambda \cdot e^{-\lambda x}$ (лек. 11, п. 3) с параметром $\lambda = 1/2$. Кривые распределений представлены на рис. 4 (коэффициенты $A_1 = 1/\sqrt{2\pi}$, $A_2 = 1/2$, $A_3 = 1/\sqrt{2\pi}$, … легко вычисляются с помощью рекуррентного свойства гамма-функции $\Gamma(\alpha) = (\alpha - 1) \cdot \Gamma(\alpha - 1)$ и с учётом $\Gamma(3/2) = \sqrt{\pi}/2$. Форма кривых, начиная с третьей, представляет собой несимметричную «горку», вершина которой с ростом k сдвигается вправо и теряет высоту, «горка» становится всё более и более пологой.

Числовые характеристики χ^2-распределений — математическое ожидание M, дисперсия D и мода M_x определяются очень простыми формулами ([1, (9.7.6)]):

$$M(\chi^2) = r, \qquad D(\chi^2) = 2r, \qquad M_x(\chi^2) = r - 2. \tag{13}$$

И они, действительно, растут с ростом m, причём, принимают целые значения. Напомню, мода M_x — это то значение x, в котором функция-плотность достигает максимума (лек. 10, п. 5). Знание моды помогает точнее построить график функции-плотности (рис. 4).

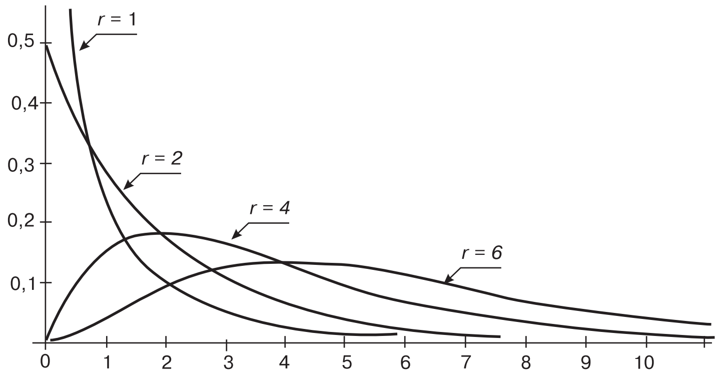

Рис. 4

Контроль 3. Запишите вид функции-плотности χ^2−распределения с параметром $r = 8$. Вычислите коэффициент A_8, числовые характеристики M, D, σ и моду M_x. Изобразите эскиз кривой распределения и укажите интервал практически возможных значений.

4. Проверка правдоподобия гипотезы о согласованности теоретического и статистического распределений. Критерий согласия Пирсона

Сосредоточимся на одной из главных задач математической статистики — проверке правдоподобия гипотез. Задача эта имеет много разновидностей, ибо гипотезы бывают разные. Мы детально вникнем в процесс и смысл решения одной из них, и вы сможете понять общую идею и схему решения всех подобных задач.

Перед тем, как формулировать задачу в общем виде, вспомним пример одной необычной с. в. — числа солдат из состава кавалерийского корпуса, убитых в течение года ударом лошадиного копыта. Статистическое распределение (число опытов $k = 200$) приведено в таблице 5 (лек. 8, п. 6). Теоретическое распределение Пуассона — в таблице 3 (там же). Эти таблицы очень похожи, — статистические вероятности (относительные частоты из второй строки таблицы 5) очень близки к теоретическим (вторая строка таблицы 3). На этом основании (субъективном) мы сделали вывод, что данную с. в. можно моделировать Пуассоновской, т. е. приняли гипотезу о Пуассоновском распределении исследуемой с. в. Там же было сделано примечание, что существуют объективные методы оценки близости теоретического и статистического распределений. Вот сейчас вы и узнаете, как это делается. В следующем разделе лекции применим то, что узнаем, к данному примеру.

Начнём с дискретных с. в.. Непрерывные с. в. сведём в дальнейшем к их дискретным аналогам.

Пятая задача. Изучается дискретная с. в. X. Над ней проведено k независимых опытов и получена простая статистическая совокупность $\{x^{(1)}, x^{(2)}, \ldots, x^{(k)}\}$, в которой значения могут повторяться. Из этой совокупности выделены все различные значения (их число m) и расположены в порядка возрастания: $x_1 < x_2 < \cdots < x_m$. Подсчитаны частоты l_i появления в эксперименте каждого из этих значений и вычислены относительные частоты $p_i^* = l_i/k$. Составлено статистическое распределение (таблица 4).

$$\begin{array}{cccccccc} X^*: & x_1 & x_2 & \ldots & x_i & \ldots & x_m \\ l: & l_1 & l_2 & \ldots & l_i & \ldots & l_m \\ p^*: & p_1^* & p_2^* & \ldots & p_i^* & \ldots & p_m^* \end{array}$$

Таблица 4.

Выдвинута *гипотеза* H о том, что с. в. X имеет распределение (теоретическое), представленное таблицей 5.

$$\begin{array}{ccccccc} X: & x_1 & x_2 & \ldots & x_i & \ldots & x_m \\ p: & p_1 & p_2 & \ldots & p_i & \ldots & p_m \end{array}$$

Таблица 5.

Гипотеза предполагает, что $p_i^* \approx p_i$, и различия между ними объясняются случайными причинами (это и означает, что гипотеза *правдоподобна*). Задача состоит в том, чтобы выработать объективную процедуру (*критерий*) проверки гипотезы, результатом которой будет её принятие или непринятие.

Примечание 1. Вы, наверное, недоумеваете, — что это такое «гипотеза H», откуда она появилась и как вычислялись вероятности p_i (таблица 5)? Опять полезно вспомнить вышеупомянутый пример с ударами лошади. Нами было замечено, что статистическое среднее $M^* = 0{,}61$ почти совпадает со статистической дисперсией $D^* = 0{,}6079$, а это признак Пуассоновской с. в. с параметром $a = 0{,}6$. Зная параметр, мы по формуле Пуассона (приложение 2) рассчитали теоретические вероятности p_i всех значений данной с. в.

Пример показывает, что для выбора теоретического распределения приходится провести «разведочный» анализ данных — найти статистические характеристики и обратить внимание на их близость. Часто помогает построение гистограммы — её вид подсказывает, какой изученный тип распределения надо выбрать. После выбора вида функции-плотности определяются её параметры и вычисляются вероятности p_i. Всё это вы увидите дальше на примерах.

Принцип сомнения. Если при выдвинутой гипотезе некоторое событие имеет очень малую вероятность (т. е. практически невозможно), но в результате одного лишь эксперимента это событие произошло, то правдоподобность гипотезы сомнительна.

Из этого принципа следует, что если удастся ввести некую обобщённую меру R, оценивающую рассогласование между статистическими p_i^* и теоретическими p_i вероятностями (таблицы 4 и 5), и после этого установить, что полученное в результате эксперимента значение этой меры R_0 очень мало вероятно (меньше некоторого порога p_α, который называют *уровнем значимости*), то рассогласование, скорее, объясняется неверным предположением, а не случайными колебаниями условий эксперимента. В этом случае правдоподобность гипотезы H следует поставить под сомнение (она неудачно моделирует исследуемую с. в.) и рекомендуется проверить какую-то другую гипотезу (её называют *альтернативной*, или *конкурирующей*).

Решение. 1. Итак, задача состоит в том, чтобы придумать *меру* R расхождения между статистическим и теоретическим распределениями и затем оценить *вероятность* наличного расхождения — $p(R_0)$.

2. *Мера расхождения.* Очевидно, чем больше разности $|p_i^* - p_i|$, тем больше расхождение между статистическим и теоретическим распределениями (таблицы 4 и 5). Поэтому в качестве меры расхождения можно было бы взять сумму $\sum\limits_{i=1}^{m} |p_i^* - p_i|$. Однако, при таком выборе меры очень трудно было бы оценить вероятности конкретных расхождений. Удобнее взять не сумму модулей, а сумму их квадратов, причём, с некоторыми «весами» c_i, т. е.

$$R = \sum_{i=1}^{m} c_i \cdot (p_i^* - p_i)^2. \tag{14}$$

3. *Пояснение о весах.* Веса c_i вводятся для того, чтобы учесть значимость каждой разности $(p_i^* - p_i)$ для меры расхождения R. Что значит «значимость»? Если вероятность p_i мала, то малое от неё отклонение статистической вероятности p_i^* может быть очень значимым для меры общего расхождения распределений R, так как может быть вызвано не случайными причинами. В то время, как то же самое отклонение p_i^* от большей вероятности p_i составит небольшой процент от неё и, значит, вполне объяснимо случайными факторами. Поэтому для малых отклонений $(p_i^* - p_i)$ следует увеличить веса c_i, а для больших — уменьшить. Но как?

4. *Мера Пирсона.* Англичанин К. Пирсон[7] предложил (1900 г.) вводить веса так:

$$c_i = k/p_i, \qquad i = 1, 2, \ldots, m, \tag{15}$$

где k — число опытов, т. е. число появившихся в эксперименте значений $x^{(i)}$ исследуемой с. в. При таком выборе получается, что веса c_i обратно пропорциональны вероятностям p_i, и, значит, увеличиваются для малых p_i и уменьшаются для больших. Коэффициент пропорциональности k — один для всех весов c_i и он не меняет этой закономерности.

Мера расхождения принимает вид:

$$R = \sum_{i=1}^{m} \frac{k}{p_i} \cdot (p_i^* - p_i)^2.$$

Для удобства расчётов преобразуем последнюю формулу, заменив $p_i^* = l_i/k$ и приведя к общему знаменателю разность, стоящую в скобке:

$$R = \sum_{i=1}^{m} \frac{k}{p_i} \cdot \left(\frac{l_i}{k} - p_i\right)^2 = \sum_{i=1}^{m} \frac{k}{p_i} \cdot \left(\frac{l_i - kp_i}{k}\right)^2 = \sum_{i=1}^{m} \frac{(l_i - kp_i)^2}{kp_i}.$$

В итоге получаем формулу для расчёта меры наличного расхождения между статистическим и теоретическим распределениями:

$$R = \sum_{i=1}^{m} \frac{(l_i - kp_i)^2}{kp_i}. \tag{16}$$

Примечание 2. В формулу (16) входят эмпирические частоты l_i, которые берутся из таблицы 4, и теоретические частоты kp_i (рассчитываются из таблицы 5). Число слагаемых m равно числу *различных* значений исследуемой с. в. X, появившихся в данном эксперименте. В сущности, формула (16) представляет собой функцию, зависящую от m аргументов (частот $l_1, l_2, \ldots, l_m$), так как в результате каждого эксперимента (при достаточно большом числе k опытов) могут меняться только частоты. Число *независимых* переменных функции обычно называют числом её *степеней свободы.* У функции (16) число степеней свободы равно $m - 1$, поскольку частоты связаны обязательным соотношением $l_1 + l_2 + \ldots + l_m = k$. Такие функции, зависящие от результатов эксперимента, называют, кратко, *статистиками.*

[7]Карл Пирсон (1857–1936) — английский математик, биолог, философ. Разрабатывал статистические методы и их приложения к вопросам наследственности и эволюции. В философии был последователем Маха.

5. *Вероятностная оценка меры R.* Если число R оказывается «большим», то расхождение между статистическим и теоретическим распределениями следует считать не случайным, а гипотезу H неправдоподобной. Если же R «мало», то гипотеза принимается, как не противоречащая опытным данным. Но что считать здесь «малым» и «большим»? Ответ должен учитывать вероятности тех или иных возможных значений меры R. Поясню.

Мера R сама является случайной величиной, связанной с экспериментом, так как при его многократном повторении она будет принимать различные значения. Одни значения меры будут возникать чаще, другие реже и надо уметь определить вероятности «больших» и «малых» значений R[8]. То есть надо знать функцию-плотность с. в. R.

Кажется, что эта задача неразрешима, ибо распределение с. в. R зависит от исходной с. в. X и потому может быть разным для разных классов X. А вот Карл Пирсон понял, что, на самом деле, распределение с. в. R слабо зависит от X и при достаточно большом числе опытов k всегда почти сопадает с χ^2-распределением (о выборе параметра r — чуть ниже). А поскольку для этих распределений функции-плотности известны (формула (12)), то можно рассчитать вероятности попадания значений с. в. R в любой промежуток $[\alpha;\beta]$ (лек. 9, п. 7, формула (7)).

И теперь, если в эксперименте мера R приняла значение R_0, мы можем найти вероятность $p(R_0) = P(R \geq R_0)$ того, что с. в. R попадёт в промежуток $[R_0;\infty)$ (или в чуть больший). Для этого составлены таблицы. Если эта вероятность оказывается очень малой (не превосходит выбранного *уровня значимости* p_α), это значит, что расхождение R_0 между статистическим и гипотетическим распределениями неправдоподобно велико и, скорее всего, объясняется ошибочностью гипотезы. В противном случае, гипотеза может быть принята, как не противоречащая опытным данным.

Примечание 3. Подчеркнём, что эти выводы не являются абсолютными и могут оказаться ошибочными. Вероятность отвергнуть правильную гипотезу (*ошибка 1-го рода*) очень мала и равна выбранному уровню значимости p_α. Этот выбор не произволен, он диктуется реальной ситуацией и стандартизируется статистической практикой. Обычно выбираются значения 0,001, 0,01, 0,05, 0,1, 0,15.

Ошибкой 2-го рода называют принятие неверной гипотезы. Если вероятность ошибки 1-го рода задаётся уровнем значимости p_α, то вероятность ошибки 2-го рода представляет новую трудную задачу, решение которой удаётся лишь в редких случаях.

Примечание 4. Большое значение вероятности $p(R_0)$ не обязательно свидетельствует о большом правдоподобии гипотезы, а может, в частности, иметь причиной не достаточное число опытов k для того, чтобы распределение меры R стало близко к χ^2.

6. *Выбор параметра r* с. в. χ^2 , близкой к с. в. R, прост:

$$r = m - s - 1^9, \tag{17}$$

где r — параметр (число степеней свободы) с. в. χ^2, m — число различных значений исследуемой с. в. X, зафиксированных в статистическом ряде (таблица 3), s — число параметров, необходимых для расчёта теоретического распределения (таблица 5) и *оценённых* по выборке (таблица 4). Заметим, что иногда какие-то параметры распределения известны точно, а другие приходится оценивать (см. п. 8, упражнение 13).

Примечание 5. Смысл формулы (17) можно пояснить так. Мера R (формула (16)) является функцией от $m-1$ независимых аргументов, так как на её m аргументов (частоты $l_1, l_2, \ldots, l_m$) накладывается дополнительная связь $l_1 + l_2 + \ldots + l_m = k$, которая делает один из аргументов зависимым от остальных. Поэтому число степеней свободы меры R равно $m-1$. Оценивание параметров теоретического распределения накладывает на частоты дополнительные связи и, следовательно, уменьшает число степеней свободы на величину s, равную числу оцениваемых параметров. В итоге, число степеней свободы становится равным $r = (m-1) - s$.

[8] Строго говоря, здесь имеются в виду не конкретные значения меры, а промежутки, в которые мера попадает с большей или меньшей вероятностью. Вероятность каждого конкретного значения меры нулевая, поскольку с. в. R непрерывная (лек. 9, п. 8, (8)).

[9] Эту формулу обосновал в 1924 г. Рональд Фишер (1890–1962) — английский математик, биолог, статистик. Сам Пирсон ввёл свой критерий в 1920 г. для случая, когда параметры распределения известны, а не оцениваются по выборке.

5. Программа применения критерия Пирсона к дискретным с. в.

В предыдущем разделе мы подробно разобрали сущность метода Пирсона. Сейчас сформулируем задачу в том виде, в котором она обычно ставится на практике, и наметим программу её решения.

Задача (для д. с. в.). Изучается дискретная с. в. X. Над ней проведено k независимых опытов, подсчитаны частоты l_i появления каждого из значений x_i и вычислены относительные частоты $p_i^* = l_i/k$. Составлено статистическое распределение с. в. X (таблица 4). На основании «разведочного» анализа высказана гипотеза H_0 о том, что данное распределение относится к одному из известных классов д. с. в. (например, Пуассоновскому, или биномиальному, или др.). Требуется проверить методом Пирсона, правдоподобна ли эта гипотеза?

Решение (программа действий).

1. Определяем необходимые статистические характеристики (M^*, D^*, σ^*) исследуемой с. в. X по формулам (1)–(3), используя таблицу 4. Причём, эти характеристики выбираются в зависимости от того, какова гипотеза H_0 (для Пуассоновский с. в. достаточно одной M^*).

2. По найденным статистическим характеристикам оцениваем параметры гипотетической с. в. (у Пуассоновской с. в. один параметр $a \approx M^*$).

3. Вычисляем теоретические (гипотетические) вероятности $p_i = P(X = x_i)$ появившихся значений с. в. X. Для Пуассоновской с. в. — по формуле

$$p_i = \frac{a^i}{i!} \cdot e^{-a}, \qquad i = 0, 1, 2, 3, \ldots . \tag{18}$$

В результате, получаем теоретическое распределение (таблица 5).

4. Вычисляем теоретические частоты $kp_i, i = 1, 2, \ldots, m$ (k — число опытов в эксперименте). Приходим к сравнительной таблице эмпирических (l_i) и теоретических (kp_i) частот:

$$\begin{array}{rcccccc} X^*: & x_1 & x_2 & \ldots & x_i & \ldots & x_m \\ l: & l_1 & l_2 & \ldots & l_i & \ldots & l_m \\ kp: & kp_1 & kp_2 & \ldots & kp_i & \ldots & kp_m \end{array}$$

Таблица 6.

Здесь надо иметь в виду следующее важное *замечание*: *если в какой-то колонке таблицы* $kp_i < 5$, *то следует объединить эту колонку с соседней* (или с соседними), сложив частоты (например, $l_i + l_{i+1}$ и $kp_i + kp_{i+1} \geq 5$) и заменив значения с. в. их средним арифметическим $(x_i + x_{i+1})/2$. Смысл этого правила будет пояснён чуть позже, в примечании.

5. Вычисляем меру Пирсона по формуле (16). Для удобства использования перепишем эту формулу в развёрнутом виде:

$$R_0 = \frac{(l_1 - kp_1)^2}{kp_1} + \frac{(l_2 - kp_2)^2}{kp_2} + \ldots + \frac{(l_m - kp_m)^2}{kp_m}. \tag{19}$$

6. По формуле (17) определяем число степеней свободы r распределения χ^2, совпадающего (почти) с распределением с. в. R. Для Пуассоновской с. в. $r = m - 2$, так как число оценённых параметров $s = 1$.

7. С помощью таблицы вычисляем вероятность $p(R_0) = P(R \geq R_0) = P(\chi^2 \geq R_0)$ наличного расхождения между эмпирическим и гипотетическим распределениями. Делается это так: в первой колонке таблицы находим значение числа степеней свободы r; в строке, в которой стоит наше r, находим значение меры R_0; выше этого числа, в первой строке таблицы находим вероятность $p(R_0)$.

Следует заметить, что в таблице может не быть нашего значения R_0, тогда заменяем его близким табличным значением R_1 и полагаем $p(R_0) \approx p(R_1)$. Можно, конечно, вычислить $p(R_0)$ более точно методом линейной интерполяции. Однако, в большинстве случаев в этом нет необходимости, — ведь надо лишь оценить $p(R_0)$ относительно выбранного уровня значимости p_α (см. п. 7 в нижеследующем примере).

8. Если $p(R_0) \le p_\alpha$, гипотеза отвергается, в противном случае — принимается, или, выражаясь более осторожно, считается не противоречащей опытным данным.

Примечание 1. Метод Пирсона эффективен только в случае, когда в формуле (16) все $kp_i \ge 5$. Это условие требуется для того, чтобы с. в. R была близка к с. в. χ^2. Необходимость этого условия указывается теорией и подтверждается практикой применения метода Пирсона. Чтобы соблюсти это условие, и было сделано замечание об укрупнении колонок таблицы 6.

Примечание 2. Существуют другие методы проверки гипотезы о законе распределения с. в. — они различаются выбором меры расхождения R между теоретическим и статистическим распределениями. Их общее название — *критерии согласия.* Изложенный выше метод называют *критерием Пирсона*, или критерием *хи-квадрат.*

Пример. Проверить методом Пирсона гипотезу о Пуассоновском распределении с. в. X — числа солдат кавалерийского корпуса, убитых лошадью в течение года. Статистический ряд известен (лекция 8, п. 4 таблица 4):

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

Таблица 7.

Решение. 1. Определяем статистическое среднее:

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

2. Оцениваем параметр гипотетической (Пуассоновской) с. в. :

$$a \approx M^* \approx 0{,}6.$$

3. Вычисляем теоретические вероятности $p_i = P(X = x_i)$ значений с. в. X по формуле (17), используя таблицу значений функции Пуассона:

X :	0	1	2	3	4	5	...
p :	0, 5488	0, 3293	0, 0988	0, 0198	0, 0030	0, 0004	...

Таблица 8.

4. Вычисляем теоретические частоты:

$$kp_1 = 200 \cdot 0{,}5488 = 109{,}76; \quad kp_2 = 200 \cdot 0{,}3293 = 65{,}86; \quad kp_3 = 200 \cdot 0{,}0988 = 19{,}76;$$
$$kp_4 = 200 \cdot 0{,}0198 = 3{,}96; \quad kp_5 = 200 \cdot 0{,}0030 = 0{,}60; \quad kp_6 = 200 \cdot 0{,}0004 = 0{,}08.$$

Приходим к сравнительной таблице эмпирических и теоретических частот:

X^* :	0	1	2	3	4	5
l_i :	109	65	22	3	1	0
kp_i :	109, 76	65, 86	19, 76	3, 96	0, 60	0, 08

Таблица 9.

В последних трёх колонках таблицы 9 суммарное значение теоретических частот меньше 5, — объединим их с 4-й колонкой и получим таблицу 10 для расчёта меры Пирсона:

X^* :	0	1	3, 5
l :	109	65	26
kp :	109, 76	65, 86	24, 40

Таблица 10.

5. Вычисляем по формуле (19) значение R_0 меры расхождения между эмпирическими и теоретическими частотами (таблица 10):

$$R_0 = \frac{(109 - 109{,}76)^2}{109{,}76} + \frac{(65 - 65{,}86)^2}{65{,}86} + \frac{(26 - 24{,}40)^2}{24{,}40} = \frac{0{,}76^2}{109{,}76} + \frac{0{,}86^2}{65{,}86} + \frac{1{,}60^2}{24{,}40} =$$
$$= \frac{0{,}5776}{109{,}76} + \frac{0{,}7396}{65{,}86} + \frac{2{,}56}{24{,}40} \approx 0{,}0053 + 0{,}0112 + 0{,}1049 = 0{,}1225.$$

6. Определяем число степеней свободы r распределения χ^2, совпадающего (почти) с распределением с. в. R. Так как в таблице 10 число значений $m = 3$, а число оценённых параметров $s = 1$, то по формуле (17) получаем:

$$r = m - s - 1 = 3 - 1 - 1 = 1.$$

7. Оцениваем вероятность $p(R_0) = P(R \geq R_0) = P(\chi^2 \geq R_0)$ наличного расхождения между эмпирическим и гипотетическим распределениями. Для этого в первой колонке таблицы находим значение $r = 1$; в этой строке (второй) находим первое значение с. в. χ^2, большее нашего $R_0 =$ 0,1225, — это будет $R_1 = 0{,}148$; над этим значением находим вероятность $p(R_1) = 0{,}7$. Из $R_0 < R_1$ следует, что $p(R_0) = P(R \geq R_0) \geq$ $\geq P(R \geq R_1) = p(R_1)$, т. е.

$$p(R_0) \geq p(R_1) > 0{,}7.$$

8. Так как $p(R_0) > 0,7 >> p_\alpha = 0,15$, гипотезу следует считать весьма правдоподобной.

Контрольные упражнения в этом и следующем разделах лекции опускаем, по причине их трудоёмкости. Отработать навык применения критерия Пирсона к дискретным и к непрерывным с. в. вы сможете в 8-м разделе упражнений.

6. Применение критерия Пирсона к непрерывным с. в.

Задача (для н.с. в.). Изучается непрерывная с. в. X. Над ней проведено k независимых опытов и получена простая статистическая совокупность $\{x^{(1)}, x^{(2)}, \ldots, x^{(k)}\}$. Произведена группировка, представленная группированным статистическим рядом (таблица 2), который затем превращён в дискретный ряд (таблица 3). На основании «разведочного» анализа высказана гипотеза H_0 о том, что данное распределение относится к одному из известных классов н. с. в. (например, к классу нормальных с. в., или показательных, или др.). Требуется проверить методом Пирсона, правдоподобна ли эта гипотеза?

Решение. идёт по той же программе, что и для д. с. в. Единственное уточнение надо сделать к 3-му этапу программы. Теоретические вероятности p_i вычисляются здесь, как вероятности попадания с. в. X в промежутки $[x_i, x_{i+1}]$. Задача эта решается общей формулой $p_i = \int\limits_{x_i}^{x_{i+1}} f(x)dx$ (лек. 9, п. 7, формула (7)), где $f(x)$ — гипотетическая функция-плотность исследуемой с. в. X.

Напомню, что функция-плотность $f(x)$ нормальной с. в. X_N имеет два параметра: a — математическое ожидание и σ — среднее квадратическое отклонение. Параметры эти связаны со статистическими характеристиками $\tilde{M}^*$ и $\tilde{\sigma}^*$, получаемыми по формулам (1)–(3), так: $a \approx \tilde{M}^*$, $\sigma \approx \tilde{\sigma}^*$. Следовательно, функция-плотность, выравнивающая наш статистический ряд, имеет вид (лек. 11, п. 6, формула (7)):

$$f(x) = \frac{1}{\tilde{\sigma}^* \sqrt{2\pi}} \cdot e^{-\frac{(x - \tilde{M}^*)^2}{2(\tilde{\sigma}^*)^2}}.$$

Интеграл от этой функции вычисляется с помощью таблицы значений функции Лапласа $\Phi(x)$ по формуле (25) (лек. 11, п. 7), которая в нашей ситуации принимает вид:

$$P(X_N \in [x_i; x_{i+1}]) = \Phi\left(\frac{x_{i+1} - \tilde{M}^*}{\tilde{\sigma}^*}\right) - \Phi\left(\frac{x_i - \tilde{M}^*}{\tilde{\sigma}^*}\right). \tag{20}$$

В нижеследующем итоговом примере сконцентрируем несколько задач математической статистики (построение гистограммы, оценка параметров распределения, выравнивание гистограммы, проверка гипотезы).

Пример. Произведено $k = 500$ опытов над непрерывной с. в. X. Все появившиеся значения с. в. заключены в промежутке $[-4; 4]$. Этот промежуток разбит на 8 меньших промежутков и подсчитаны частоты l_i попадания появившихся значений с.в. в каждый из них. Результаты сведены в группированный статистический ряд (таблица 11)[10].

I_i:	$(-4) \div (-3)$	$(-3) \div (-2)$	$(-2) \div (-1)$	$(-1) \div 0$	$0 \div 1$	$1 \div 2$	$2 \div 3$	$3 \div 4$
$\tilde{x}_i$:	$-3,5$	$-2,5$	$-1,5$	$-0,5$	$0,5$	$1,5$	$2,5$	$3,5$
l_i:	6	25	72	133	120	88	46	10

Таблица 11

Построить гистограмму и, учитывая её вид, высказать гипотезу H о том, к какому классу с.в. принадлежит исследуемая с.в X. Выровнять гистограмму (вычислить $\tilde{M}^*$ и $\tilde{\sigma}^*$, записать гипотетическую функцию-плотность $f(x)$, вычислить её значения в концевых точках промежутков группировки и построить график). С помощью критерия Пирсона проверить правдоподобие гипотезы H на уровне значимости $p_\alpha = 0{,}1$.

Решение. 1. Построение гистограммы.

Вычисляем относительные частоты:

$$p_1^* = (l_1/k) = (1/500)\cdot 6 = 0{,}002\cdot 6 = 0{,}012; \quad p_2^* = 0{,}002\cdot 25 = 0{,}050;$$
$$p_3^* = 0{,}002\cdot 72 = 0{,}144; \quad p_4^* = 0{,}002\cdot 133 = 0{,}266; \quad p_5^* = 0{,}002\cdot 120 = 0{,}240;$$
$$p_6^* = 0{,}002\cdot 88 = 0{,}176; \quad p_7^* = 0{,}002\cdot 46 = 0{,}092; \quad p_8^* = 0{,}002\cdot 10 = 0{,}020.$$

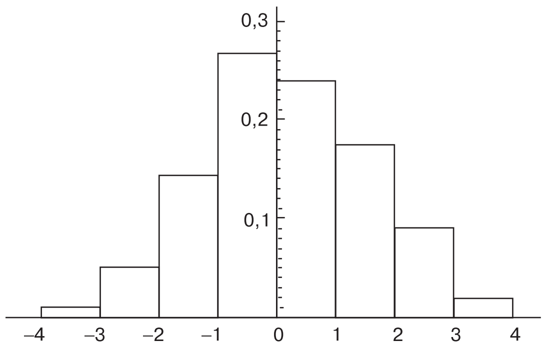

Рис. 5

Строим на промежутках группировки «столбики», площади которых (а в нашем случае и высоты) равны соответствующим относительным частотам p_i^*.

Гистограмма показана на рис. 5. Она напоминает «колокол Гаусса», поэтому высказываем гипотезу H о нормальном распределении данной с.в.

2. Оценка параметров:

$$a \approx \tilde{M}^* = 0{,}002\cdot(-3{,}5\cdot 6 - 2{,}5\cdot 25 - 1{,}5\cdot 72 - 0{,}5\cdot 133 + 0{,}5\cdot 120 + 1{,}5\cdot 88 + 2{,}5\cdot 46 + 3{,}5\cdot 10) = 0{,}168;$$

$$\tilde{D}^* = 0{,}002\cdot[(-3{,}5-0{,}168)^2 + (-2{,}5-0{,}168)^2 + (-1{,}5-0{,}168)^2 + (-0{,}5-0{,}168)^2 + \\ + (0{,}5-0{,}168)^2] + 0{,}002\cdot\left[(1{,}5-0{,}168)^2 + (2{,}5-0{,}168)^2 + (3{,}5-0{,}168)^2\right] = 2{,}098;$$

$$\sigma \approx \tilde{\sigma}^* = \sqrt{2{,}098} = 1{,}448.$$

3. Функция-плотность:

$$f(x) = \frac{1}{1,448\cdot\sqrt{2\pi}}\cdot e^{-\frac{(x-0,168)^2}{2\cdot 1,448^2}}.$$

4. График функции-плотности.

[10]Таблица 11 представляет результаты реального эксперимента по исследованию точности прибора, измеряющего угол φ, определяющий высоту объекта над горизонтом. Ошибка измерения угла вычислялась в тысячных долях радиана, она и представляет собой значения исследуемой с.в. X [1, гл. 11, п. 4].

Вычислим значения полученной функции $f(x)$ в концевых точках промежутков разбиения. Для этого используем таблицу значений функции Гаусса:

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

Найдём $f(-4)$. Для этого сначала надо вычислить соответствующий аргумент t: если $x = -4$, то $t = (-4 - 0{,}168)/1{,}448 = -4168/1{,}448 \approx -2{,}878$. Теперь найдём в таблице значение $\varphi(-2{,}878) = \varphi(2{,}878) \approx 0{,}006$ и поделим его на 1,448, получим 0,004. Итак, $\varphi(-4) \approx 0{,}004$. Аналогично вычисляем остальные значения:

$$f(-3) = \frac{1}{1{,}448} \cdot \varphi\left(\frac{-3 - 0{,}168}{1{,}448}\right) \approx 0{,}691 \cdot \varphi(-2{,}188) \approx 0{,}691 \cdot 0{,}036 \approx 0{,}025;$$
$$f(-2) = \frac{1}{1{,}448} \cdot \varphi\left(\frac{-2 - 0{,}168}{1{,}448}\right) \approx 0{,}691 \cdot \varphi(-1{,}497) \approx 0{,}691 \cdot 0{,}131 \approx 0{,}090;$$
$$f(-1) = \frac{1}{1{,}448} \cdot \varphi\left(\frac{-1 - 0{,}168}{1{,}448}\right) \approx 0{,}691 \cdot \varphi(-0{,}807) \approx 0{,}691 \cdot 0{,}288 \approx 0{,}199;$$
$$f(0) = \frac{1}{1{,}448} \cdot \varphi\left(\frac{0 - 0{,}168}{1{,}448}\right) \approx 0{,}691 \cdot \varphi(-0{,}116) \approx 0{,}691 \cdot 0{,}396 \approx 0{,}274;$$
$$f(1) = \frac{1}{1{,}448} \cdot \varphi\left(\frac{1 - 0{,}168}{1{,}448}\right) \approx 0{,}691 \cdot \varphi(0{,}575) \approx 0{,}691 \cdot 0{,}338 \approx 0{,}234;$$
$$f(2) = \frac{1}{1{,}448} \cdot \varphi\left(\frac{2 - 0{,}168}{1{,}448}\right) \approx 0{,}691 \cdot \varphi(1{,}265) \approx 0{,}691 \cdot 0{,}179 \approx 0{,}123;$$
$$f(3) = \frac{1}{1{,}448} \cdot \varphi\left(\frac{3 - 0{,}168}{1{,}448}\right) \approx 0{,}691 \cdot \varphi(1{,}956) \approx 0{,}691 \cdot 0{,}059 \approx 0{,}041;$$
$$f(4) = \frac{1}{1{,}448} \cdot \varphi\left(\frac{4 - 0{,}168}{1{,}448}\right) \approx 0{,}691 \cdot \varphi(2{,}646) \approx 0{,}691 \cdot 0{,}012 \approx 0{,}008.$$

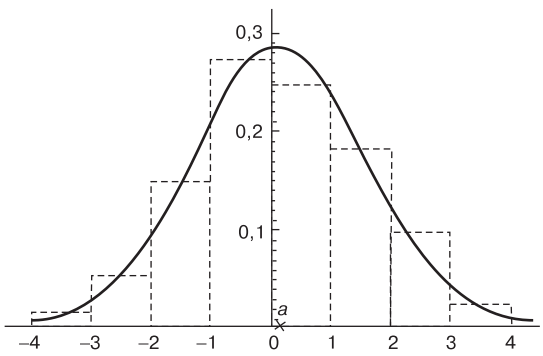

Рис. 6.

Для правильного построения графика вычислим максимальное значение

$$f(a) = f(0{,}168) = 0{,}691 \cdot \varphi(0) \approx \approx 0{,}691 \cdot 0{,}399 \approx 0{,}276.$$

Отложим на гистограмме вычисленные выше значения функции-плотности и проведём график (рис. 6).

По его виду заключаем, что он неплохо согласуется с гистограммой и, значит, наша гипотеза получает подкрепление.

5. Проверяем критерием Пирсона правдоподобие гипотезы о нормальном распределении данной с. в. X.

Нормально распределённая с. в. имеет два параметра, которые мы оценили выше: $a \approx \tilde{M}^* = 0{,}168$ и $\sigma \approx \tilde{\sigma}^* = 1{,}448$ (первые два этапа программы п. 5).

3-й этап. Вычислим по формуле (20) теоретические вероятности p_i попадания значений с. в. в промежутки группировки:

$$p_1 = P(X \in [-4; -3]) = \Phi\left(\frac{-3 - 0{,}168}{1{,}448}\right) - \Phi\left(\frac{-4 - 0{,}168}{1{,}448}\right) = \Phi(-2{,}188) - \Phi(-2{,}878) = = -0{,}486 + 0{,}498 = 0{,}012;$$

$$p_2 = P(X \in [-3; -2]) = \Phi\left(\frac{-2 - 0{,}168}{1{,}448}\right) - \Phi\left(\frac{-3 - 0{,}168}{1{,}448}\right) = \Phi(-1{,}497) - \Phi(-2{,}188) = = -0{,}433 + 0{,}498 = 0{,}065;$$

$$p_3 = P(X \in [-2; -1]) = \Phi\left(\frac{-1-0{,}168}{1{,}448}\right) - \Phi\left(\frac{-2-0{,}168}{1{,}448}\right) = \Phi(-0{,}807) - \Phi(-1{,}497) = \\ = -0{,}288 + 0{,}433 = 0{,}144;$$

$$p_4 = P(X \in [-1;\ 0]) = \Phi\left(\frac{0-0{,}168}{1{,}448}\right) - \Phi\left(\frac{-1-0{,}168}{1{,}448}\right) = \Phi(-0{,}116) - \Phi(-0{,}807) = \\ = -0{,}046 + 0{,}288 = 0{,}242;$$

$$p_5 = P(X \in [0;\ 1]) = \Phi\left(\frac{1-0{,}168}{1{,}448}\right) - \Phi\left(\frac{0-0{,}168}{1{,}448}\right) = \Phi(0{,}575) - \Phi(-0{,}116) = \\ = 0{,}217 + 0{,}046 = 0{,}263;$$

$$p_6 = P(X \in [1; 2]) = \Phi\left(\frac{2-0{,}168}{1{,}448}\right) - \Phi\left(\frac{1-0{,}168}{1{,}448}\right) = \Phi(0{,}265) - \Phi(0{,}575) = \\ = 0{,}385 - 0{,}217 = 0{,}168;$$

$$p_7 = P(X \in [2;\ 3]) = \Phi\left(\frac{3-0{,}168}{1{,}448}\right) - \Phi\left(\frac{2-0{,}168}{1{,}448}\right) = \Phi(1{,}956) - \Phi(0{,}265) = \\ = 0{,}474 - 0{,}385 = 0{,}089;$$

$$p_8 = P(X \in [3;\ 4]) = \Phi\left(\frac{4-0{,}168}{1{,}448}\right) - \Phi\left(\frac{3-0{,}168}{1{,}448}\right) = \Phi(2{,}646) - \Phi(1{,}956) = \\ = 0{,}496 - 0{,}474 = 0{,}022.$$

4-й этап. Вычисляем теоретические частоты kp_i, $i = 1, 2, \ldots, 8$ ($k = 500$ — число опытов в эксперименте): $kp_1 = 500 \cdot 0{,}012 = 6$; $kp_2 = 500 \cdot 0{,}065 = 32{,}5$;

$$kp_3 = 500 \cdot 0{,}144 = 72; \qquad kp_4 = 500 \cdot 0{,}242 = 121; \qquad kp_5 = 500 \cdot 0{,}263 = 131{,}5;$$
$$kp_6 = 500 \cdot 0{,}168 = 84; \qquad kp_7 = 500 \cdot 0{,}089 = 44{,}5; \qquad kp_8 = 500 \cdot 0{,}022 = 11.$$

Добавляем к таблице 11 строку теоретических частот и приходим к сравнительной таблице 12 эмпирических (l_i) и теоретических (kp_i) частот:

$\tilde{x}_i$:	$-3,5$	$-2,5$	$-1,5$	$-0,5$	$0,5$	$1,5$	$2,5$	$3,5$
l_i :	6	25	72	133	120	88	46	10
kp_i :	6	32, 5	72	121	131, 5	84	44, 5	11

Таблица 12

В 3-й строке таблицы все теоретические вероятности больше пяти, поэтому нет нужды в объединении колонок, как это пришлось сделать в предыдущем примере (п. 5).

5-й этап. Вычисляем меру Пирсона по формуле (19):

$$R_0 = \frac{(25-32{,}5)^2}{32{,}5} + \frac{(133-121)^2}{121} + \frac{(120-131{,}5)^2}{131{,}5} + \frac{(88-84)^2}{84} + \frac{(46-44{,}5)^2}{44{,}5} + \\ + \frac{(10-11)^2}{11} = \frac{7{,}5^2}{32{,}5} + \frac{12^2}{121} + \frac{11{,}5^2}{131{,}5} + \frac{4^2}{84} + \frac{1{,}5^2}{44{,}5} + \frac{1^2}{11} = \frac{56{,}25}{32{,}5} + \frac{144}{121} + \frac{132{,}25}{131{,}5} + \frac{16}{84} + \frac{2{,}25}{44{,}5} + \frac{1}{11} \approx \\ \approx 1{,}731 + 1{,}190 + 1{,}006 + 0{,}190 + 0{,}051 + 0{,}091 = 4{,}259.$$

6-й этап. Определяем параметр r — число степеней свободы распределения χ^2, совпадающего (почти) с распределением с. в. R. Так как в таблице 12 число значений $m = 8$, а число оценённых параметров нормальной с. в. $s = 2$, то по формуле (17) получаем:

$$r = m - s - 1 = 8 - 1 - 2 = 5.$$

7-й этап. Оцениваем вероятность $p(R_0) = P(R \geq R_0) = P(\chi^2 \geq R_0)$ наличного расхождения между эмпирическим и гипотетическим распределениями. Для этого в первой колонке таблицы находим значение $r = 5$; в этой строке (шестой) находим первое значение с. в. χ^2, большее нашего $R_0 = 4{,}259$, — это будет $R_1 = 4{,}35$; над этим значением находим вероятность $p(R_1) = 0{,}5$. Из $R_0 < R_1$ следует, что

$$p(R_0) = P(R \geq R_0) \geq P(R \geq R_1) = p(R_1), \qquad \text{т. е.} \qquad p(R_0) \geq p(R_1) > 0{,}5.$$

Заметим, что методом интерполирования искомая вероятность оценивается так: $p(R_0) \approx \approx 0{,}52$.

8-й этап. Так как $p(R_0) > 0{,}5 > p_\alpha = 0{,}1$, гипотезу следует считать весьма правдоподобной.

7. Заключение

Вот и закончен наш курс. Его задачей было введение вас в теорию вероятностей и математическую статистику. Что значит «введение»? Ознакомление с *основными*, фундаментальными понятиями, фактами, методами этих наук. Более того, ясное и глубокое *понимание сути*. Это и было моей высшей целью — помощь вам в понимании смыслов.

Только осмысленные, фундаментальные знания формируют качественного специалиста. Только такие знания открывают возможности дальнейшего *самостоятельного* их расширения и творческого применения в вашей будущей профессиональной работе.

За рамками нашего курса осталось много ценных теоретических и практических разделов науки: многомерные с. в., разнообразные классы с. в. (гамма- и бета распределения, распределения Эрланга, Фишера, Стьюдента, Вейбулла и др.), функции от случайных величин, их числовые характеристики и законы распределения, критерии согласия Фишера, Стьюдента, Колмогорова, проверка разнообразных гипотез, корреляционная зависимость между с. в. и пр., и пр. Эти разделы вы можете теперь самостоятельно изучить по книгам [1, 2]. Конечно, при условии, что материал данного курса хорошо усвоен и вы научились видеть за теоретичесими фактами их статистический смысл. Рекомендую также серьёзную, хорошо написанную и богатую практическим содержанием книгу В. Феллера «Введение в теорию вероятностей и её применение» в 2-х томах, — М., Мир, 1984.

Благодарю вас за длительную интересную и полезную совместную работу. Желаю вам успешного профессионального роста и удовлетворённости.

8. Упражнения[11]

1. Произведено 300 наблюдений над некоторой с. в. и результаты сведены в интервальный ряд (таблица 13).

Δx_i :	$(-20) \div (-10)$	$(-10) \div 0$	$0 \div 10$	$10 \div 20$	$20 \div 30$	$30 \div 40$	$40 \div 50$
l_i	20	47	80	89	40	16	8

Таблица 13

Постройте гистограмму. Выровняйте её с помощью кривой Гаусса. Используя критерий Пирсона, проверьте на уровне значимости 0,05 гипотезу о нормальном распределении данной с. в.

Ответ: $\tilde{M}^* = 10{,}4$; $\tilde{\sigma}^* = 13{,}67$; $r = 4$; $R_0 = 5{,}4$.

[11]Часть задач этого раздела и ответы к ним взяты из книги В. Е. Гмурмана «Руководство к решению задач по теории вероятностей и математической статистике», М., Высшая школа, 1979. Формулировки задач иногда изменены и дополнены.

2. Произведено 100 наблюдений над некоторой с. в. и результаты сведены в интервальный ряд (таблица 14).

Δx_i :	$6 \div 16$	$16 \div 26$	$26 \div 36$	$36 \div 46$	$46 \div 56$	$56 \div 66$	$66 \div 76$	$76 \div 86$
l_i :	8	7	16	35	15	8	6	5

Таблица 14

Используя критерий Пирсона, проверьте на уровне значимости 0,05 гипотезу о нормальном распределении данной с. в.

Ответ: $\tilde{M}^* = 42{,}5$; $\tilde{\sigma}^* = 17{,}17$; $r = 5$; $R_0 = 14$.

3. После шлифовки 200 валиков измерены их диаметры и все они оказались заключены в границах $6{,}68 \div 6{,}83$ (см.). Далее произведена группировка: за основной промежуток выбран [6,67; 6,85], он разбит на мелкие промежутки длины $\Delta x_i = 0{,}02$, подсчитаны частоты l_i попадания диаметров в каждый Δx_i. Составлен группированный ряд (таблица 15).

Δx_i :	$6,67 \div 6,69$	$6,69 \div 6,71$	$6,71 \div 6,73$	$6,73 \div 6,75$	$6,75 \div 6,77$
l_i :	2	15	17	44	52

Δx_i :	$6,77 \div 6,79$	$6,79 \div 6,81$	$6,81 \div 6,83$	$6,83 \div 6,85$
l_i :	44	14	11	1

Таблица 15

Используя критерий Пирсона, проверьте на уровне значимости 0,1 гипотезу о нормальном распределении диаметра валика.

Указание. Для оценки параметров можете воспользоваться результатами упражнения 9 (лек. 6, п. 8). Вычисляя меру R_0, не забудьте объединить первые два и последние два промежутка группировки.

Ответ: $\tilde{M}^* \approx 6{,}76$; $\tilde{\sigma}^* \approx 0{,}03$; $r = 4$; $R_0 \approx 6{,}75$.

4. Испытывались на длительность горения 450 электроламп. Эмпирическое распределение частот приведено в таблице 16 (время горения считалась в часах).

Δx_i :	$0 \div 400$	$400 \div 800$	$800 \div 1200$	$1200 \div 1600$	$1600 \div 2000$	$2000 \div 2400$	$2400 \div 2800$
l_i :	121	95	76	56	45	36	21

Таблица 16

Исходя из убывания частот, высказано предположение о показательном распределении с. в. — времени горения ламп. Проверьте эту гипотезу на уровне значимости 0,01.

Ответ: $\tilde{M}^* = 1000$; $\lambda = 0{,}001$; теор. частоты: 148,36; 99,45; 66,64; 44,68; 29,97; 20,07; 13,46; $r = 5$; $R_0 \approx 33{,}84$.

5. Исследуется с. в. X — время (в часах) безотказной работы прибора. Испытано 200 приборов и результаты сгруппированы в таблице 17.

Δx_i :	$0 \div 5$	$5 \div 10$	$10 \div 15$	$15 \div 20$	$20 \div 25$	$25 \div 30$
l_i	133	45	15	4	2	1

Таблица 17

Постройте гистограмму и выровняйте её с помощью показательного закона. Проверьте на уровне значимости 0,05 гипотезу о показательном распределении данной с. в. X.

Ответ: $\tilde{M}^* = 5$; $\lambda = 0,2$; теор. частоты: 126,42; 46,52; 17,10; 0,30; 2,32; 0,84; $r = 2$; $R_0 \approx 1,29$.

6. С целью изучения динамики посещений выставки проведена регистрация времени прихода 800 посетителей в течение 8-часового рабочего дня. Частота посещений в течение каждого часа работы приведена в таблице 18.

Δx_i :	$0 \div 1$	$1 \div 2$	$2 \div 3$	$3 \div 4$	$4 \div 5$	$5 \div 6$	$6 \div 7$	$7 \div 8$
l_i :	259	157	109	74	70	47	40	34

Таблица 18

Проверьте на уровне значимости 0,05 гипотезу о показательном распределении данной с. в. — времени прихода посетителя.

Ответ: $\tilde{M}^* = 2{,}5$; $\lambda = 0{,}4$; теор. частоты: 263,76; 176,80; 118,48; 79,44; 53,28; 35,68; 23,92; 16,00; $r = 6$; $R_0 \approx 41{,}66$.

7. На бензозаправочной станции регистрировалось время прибытия автомашин в течение дня (с 8 до 18 часов), с целью определить периоды большей и меньшей загруженности. Всего зарегистрировано 200 машин. Результаты сведены в таблицу 19 (в первой строке — интервалы времени Δt_i в часах, во второй — число машин l_i в интервале Δt_i).

Δt_i :	$8 \div 9$	$9 \div 10$	$10 \div 11$	$11 \div 12$	$12 \div 13$	$13 \div 14$	$14 \div 15$	$15 \div 16$	$16 \div 17$	$17 \div 18$
l_i :	12	40	22	16	28	6	11	33	18	14

Таблица 19

Таблица показывает, что наибольшая загруженность была утром (от 9 до 10) и во второй половине дня (от 15 до 16), а наименьшая — в середине дня (от 12 до 13). Однако, после построения гистограммы (сделайте это!) было высказано предположение, что данные колебания могут быть вызваны случайными причинами и, если значительно увеличить объём выборки, распределение частот может стать равномерным. Проверьте эту гипотезу на уровне значимости 0,01.

Указание. Равномерное распределение имеет два параметра a и b, и они связаны с числовыми характеристиками так: $M = (b - a)/2$ и $\sigma = (b - a)/2\sqrt{3}$ (лек. 11, п. 2).

Ответ: $\tilde{M}^* \approx 12{,}71$; $\tilde{\sigma}^* \approx 2{,}86$; $a^* \approx 7{,}76$; $b^* \approx 17{,}66$; $f(t) = 0{,}101$; $r = 7$; $R_0 \approx 53{,}43$.

8. На метеорологической станции ежедневно регистрировалась среднесуточная температура воздуха. Всего сделано 300 наблюдений. Результаты сведены в таблицу 20 (в первой строке — интервал температуры в градусах, во второй — количество дней).

Δt_i^0 :	$-40 \div -30$	$-30 \div -20$	$-20 \div -10$	$-10 \div 0$	$0 \div 10$	$10 \div 20$	$20 \div 30$	$30 \div 40$
l_i	25	40	30	45	40	46	48	26

Таблица 20

Постройте гистограмму и выровняйте её равномерной плотностью. Проверьте на уровне значимости 0,05 гипотезу о равномерном распределении среднесуточной температуры.

Ответ: $\tilde{M}^* \approx 1{,}5$; $\tilde{\sigma}^* \approx 21{,}31$; $a^* \approx -35{,}37$; $b^* \approx 38{,}37$; $f(t) = 0{,}014$; $r = 5$; $R_0 \approx 7{,}71$.

9. На заводе проведена выборочная проверка качества изделий: проверено 200 партий и в каждой из них зафиксировано число нестандартных изделий. В сущности, изучалась дискретная с. в. X — число нестандартных изделий в каждой выпущенной партии. Оказалось, что эта с. в. может принимать 5 значений с частотами, указанными статистическим рядом (таблица 21).

x_i :	0	1	2	3	4
l_i :	116	56	22	4	2

Таблица 21

Таблица показывает, что частоты быстро убывают. На этом не очень прочном основании была высказана гипотеза: с. в. X распределена по закону Пуассона. Подтвердите эту гипотезу геометрически, построив гистограмму и выровняв её. Проверьте гипотезу объективно методом Пирсона на уровне значимости 0,05 (не забудьте объединить две малочисленные частоты).

Ответ: $M^* = 0{,}6$; $\lambda = 0{,}6$; теор. частоты: 109,76; 65,86; 19,76; 3,96; 0,60; $r = 2$; $R_0 \approx 2{,}54$.

10. На базу прибыло 500 контейнеров со стеклянными изделиями. При вскрытии обнаружили, что больше половины контейнеров содержали от 1 до 7 изделий, повреждённых при транспортировке. С целью объективного изучения возможных потерь было проведено статистическое исследование, результаты которого приведены в таблице 22.

y_i :	0	1	2	3	4	5	6	7
l_i :	199	169	87	31	9	3	1	1

Таблица 22

Проверьте на уровне значимости 0,01 гипотезу о Пуассоновском распределении с. в. Y — числа повреждённых изделий в контейнере.

Ответ: $M^* = 1$; $\lambda = 1$; теор. частоты: 183,95; 92,00; 30,65; 7,65; 1,55; 0,25; 0,05; $r = 4$; $R_0 \approx 8,32$.

11. В автобусном парке проведено статистическое исследование с. в. Z — числа автобусов, сошедших с линии в течение рабочего дня из-за поломок в пути. Исследование проводилось в течение 200 рабочих дней. Полученный статистический ряд представлен таблицей 23.

z_i :	0	1	2	3	4	5	≥ 6
l_i :	70	78	34	13	4	1	0

Таблица 23

Постройте гистограмму. Сравните её с многоугольниками распределения Пуассоновских с. в. при $a = 0,6$ и $a = 1$ (см. лек. 8, п. 2, рис. 1). Можно ли в результате сравнения сделать предположение о Пуассоновском распределении с. в. Z и с каким параметром? Проверьте на уровне значимости 0,01 эту гипотезу.

Ответ: $M^* = 1,03$; $\lambda^* = 1$; $r = 2$; $R_0 \approx 0,90$.

12. В таблице 24 приведены результаты исследования крови, проведённого Н. Хольмбергом (в верхней строке — число красных кровяных тел в порции, в нижней — число порций с одинаковым числом красных тел, всего порций 169).

z_i :	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
l_i :	1	3	5	8	13	14	15	15	21	18	17	16	9	6	3	2	2	1

Таблица 24

Оцените («на глазок») по статистическому ряду расположение выборочного среднего M^* и затем вычислите его точно. Проверьте, выполняется ли статистический признак Пуассоновского распределения (лек. 8, п. 4). С помощью критерия χ^2 проверьте гипотезу о том, что число красных тел в порции подчиняется закону Пуассона (на каком уровне значимости?).

13[12]**.** Результаты сессии 100 студентов-заочников, сдававших 4 экзамена, приведены в таблице 25.

Число сданных экзаменов :	0	1	2	3	4
Число студентов :	1	1	3	35	60

Таблица 25

Данную ситуацию можно моделировать следующим образом: опыт состоит в сдаче студентом экзамена и повторяется 4 раза, случайная величина X — число сданных экзаменов. Эта

[12] Данное упражнение, как и упражнение 3, взято из книги В. Н. Калининой и В. Ф. Панкина «Математическая статистика», М., Высшая школа, 1998 (задачи 10.11 и 10.12). Тексты заданий изменены.

модель похожа на схему Бернулли (лек. 7, п. 2), которая дополнительно требует неизменности вероятности сдачи любого экзамена любым студентом. В реальности это, конечно, не так. Но можно предположить, что при достаточно большом числе студентов (заочников!) биномиальная модель с. в. X будет неплохо согласована с реальностью. Среднюю вероятность p сдачи экзамена произвольным студентом можно оценить статистически так:

$$p = \frac{\text{число всех сданных экзаменов}}{\text{число всех экзаменов}} = \frac{0 \cdot 1 + 1 \cdot 1 + 2 \cdot 3 + 3 \cdot 35 + 4 \cdot 60}{100 \cdot 4} = 0{,}88.$$

Вычислите теоретические вероятности значений с. в. X по формуле Бернулли (лек. 4, п. 2, формула (Б)) и составьте теоретический ряд распределения с. в. X. По таблице 25 вычислите статистические вероятности значений с. в. X и составьте статистический ряд распределения. Сравните статистический и теоретический ряды и сделайте предварительное заключение о приемлемости биномиальной модели. Постройте многоугольники распределения (лек. 5, п. 1), — близки ли полученные ломанные? На уровне значимости 0,01 проверьте приемлемость гипотезы о биномиальном распределении с. в. X.

Указание. Не забудьте объединить первые три колонки таблицы 25 и учесть, что в расчётной таблице останется $m = 3$ значения. Учтите также, что биномиальное распределение имеет два параметра: $k = 4$ — число повторений опыта и $p = 0{,}88$ — вероятность «успеха». Первый параметр известен, а второй мы оценили с помощью статистического ряда, т. е. число оценённых параметров $s = 1$. Значит, число r степеней свободы χ^2-распределения в нашем случае равно $r = m - s - 1 = 3 - 1 - 1 = 1$.

Ответ: $R_0 = 0{,}895 < 2{,}71 \Rightarrow$ гипотеза принимается.

14[13]**.** Таблица 26 показывает интервальное распределение FeO в руде горы Магнитной (в первой строке даётся содержание FeO в пробе руды, выраженное в процентах от общего веса пробы; во второй — частоты; число проб 296).

Δx_i:	$0 \div 2$	$2 \div 4$	$4 \div 6$	$6 \div 8$	$8 \div 10$	$10 \div 12$	$12 \div 14$	$14 \div 16$
l_i:	11	16	20	17	19	16	20	22

Δx_i:	$16 \div 18$	$18 \div 20$	$20 \div 22$	$22 \div 24$	$24 \div 26$	$26 \div 28$	$28 \div 30$	$30 \div 32$
l_i:	19	18	27	37	27	18	5	4

Таблица 26

Вычислите среднее содержание FeO в пробе, т. е. $\tilde{M}^*$. Рассчитайте надёжность этой оценки при требуемой точности $\varepsilon = 0{,}3$. Найдите для $\tilde{M}^*$ доверительный интервал при доверительной вероятности 0,9. Постройте гистограмму. Выровняйте её равномерной плотностью. Определите на уровне значимости 0,15, — согласуется ли равномерная плотность с эмпирическим распределением?

15. Таблица 27 показывает интервальное распределение свинца (Pb) в руде одного из Нерчинских рудников.

Δx_i:	$0 \div 5$	$5 \div 10$	$10 \div 15$	$15 \div 20$	$20 \div 25$	$25 \div 30$	$30 \div 35$	$35 \div 40$	$40 \div 45$
l_i:	122	75	43	23	13	9	7	5	1

Таблица 27

Вычислите среднее содержание Pb в пробе, т. е. $\tilde{M}^*$, статистическую дисперсию $\tilde{D}^*$ и среднее квадратическое отклонение $\tilde{\sigma}^*$. Рассчитайте надёжность оценки $\tilde{M}^*$ при требуемой точности $\varepsilon = 0{,}2$. Найдите для неё доверительный интервал при доверительной вероятности 0,9. Постройте гистограмму. Выровняйте её подходящей кривой. Определите на уровне значимости 0,01, — согласуется ли ваша выравнивающая кривая с эмпирическим распределением?

[13]Это и следующее упражнения (таблицы) взяты из книги «Сборник задач по высшей математике» под редакцией Н. М. Гюнтера и Р. О. Кузьмина, часть третья, ГОНТИ, М.-Л., 1938. Тексты (задания) изменены

Литература

1. Вентцель Е.С., Овчаров Л.А. Теория вероятностей и ее инженерные приложения. М.: Высшая школа, 2000.
2. Вентцель Е.С. Теория вероятностей. М.: Наука, 1998.

Костенко Игорь Петрович,
кандидат физ.-мат. наук,
доцент кафедры "Высшая математика-1"
Ростовского государственного университета
путей сообщения (Краснодарский филиал).

Email: kost@kubannet.ru

О Фонде математического образования и просвещения

Фонд математического образования и просвещения создан в конце 1996 г. с целью способствовать сохранению богатых традиций математического образования и науки в России. Фонд сотрудничает с организациями и гражданами, желающими участвовать в благородном деле сохранения высокого качества российского математического образования. Фонд поддерживает образовательные инициативы, направленные на достижение поставленной цели. Особое внимание оказывается образовательным инициативам в провинции, как в виде издательской поддержки, так и финансовой помощи. Фонд способствует изданию научной, учебной и методической литературы в области математики и смежных наук.

Условия подписки и приема материалов

По вопросам подписки на журнал обращайтесь по телефону: (495) 107-31-46 .

Адрес для корреспонденции Фонда: 141075 г. Королев Московской обл., пр-т Космонавтов 9-167. E-mail: matob@yandex.ru

Стоимость подписки на каждый из номеров 1-4 за 2008 год (включая стоимость пересылки) – 60 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 2008 г., количество экземпляров.

Реквизиты для перечисления:

Получатель: ИНН 7725080165 Фонд математического образования и просвещения

Расчетный счет и банк получателя:
р/с 40703810700010001416 в ОАО Банк "Развитие-Столица", г. Москва,
к/с 30101810000000000984, БИК 044525984, ОКПО 45084342

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) — 50 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