

УРДЕНА ЛЕНИНА ИНСТИТУТ ПРИКЛАДНОЙ МАТЕМАТИКИ
АКАДЕМИИ НАУК СССР

Молчанов А.М.

ОБ ОДНОЙ ТЕОРЕМЕ А.Я. ХИНЧИНА

(Препринт)

г. Москва, 1968 г.

Об одной теореме А.Я.Хинчина

Сформулирована и доказана теорема об аппроксимации функции одной переменной любой сумматорной функцией большого числа переменных.

Аппроксимация (в смысле интеграла от квадрата разности) становится все более точной по мере увеличения числа переменных.

ВВЕДЕНИЕ

А.Я.Хинчин в своих исследованиях по математическим основаниям статистической механики систематически использовал репрезентативность среднего значения сумматорных величин.

Этот факт является, по существу, теоремой анализа и в своих лекциях, к сожалению не опубликованных, А.Я.Хинчин особенно подчеркивал желательность освобождения от вероятностной терминологии.

Задача заметки—формулировка теоремы анализа, равносильной факту репрезентативности (1) среднего и установления связи этой теоремы с обобщенной (2) формулой Больцмана.

п. I. Сумматорные функции. Разбиение на доминанту и флуктуанту

В книге Хинчина (1) все термины (кроме термина "сумматорная функция") взяты из термодинамики и статистической физики. Однако, за четверть века, прошедшие со времени выхода этой книги, стало ясно, что значение ее основных идей существенно

выходит за рамки узкой задачи обоснования статистической механики. Это самостоятельная область математического анализа — функции большого числа переменных со своими задачами и методами, но сохранившая, к сожалению, чужую терминологию, неминуемо приводящую к путанице и противоречиям. Полная смена всей терминологии — это, конечно, крайность. Будет трудно понимать даже простые вещи только потому, что они сказаны непривычными словами. С другой стороны такие слова, как "вероятность" и "энтропия" настолько перегружены историей науки, случайными $\times/$, превходящими ассоциациями, что должны быть решительно изгнаны из употребления. Новые же термины желательно вводить только в случае крайней необходимости. Автору кажется, что введение терминов "доминанта" и "флуктуанта" не слишком дорогая цена за простое объяснение глубокой идеи А.Я.Хинчина.

Обычно считается (и, вообще говоря, это правильно), что сложность строения функции растет по мере увеличения числа ее аргументов.

х/ Лебег ⁽³⁾ пишет "Если бы предел P_n был назван "тарарабумбией" круга, то вряд ли кто-нибудь позволил бы себе вывести из нее величину тарарабумбий сектора и сегмента, но делать это разрешают себе, когда вместо слова тарарабумбия употребили слово площадь! Это тяжчайшее преступление против здравого смысла".

Стоит добавить, что это детские игрушки по сравнению с тем, как изменяется над замечательным понятием энтропии.

Тем более неожиданна идея Хинчина.

Для многих классов функций, важных для математического естествознания, имеет место прямо противоположная ситуация - они сводятся к функциям одной переменной E .

Суть дела, разумеется, именно в выделении правильного класса функций. Этот класс должен быть достаточно узким, чтобы можно было надеяться на существование нетривиальных закономерностей. Но, вместе с тем, он должен быть настолько широким, чтобы эти закономерности имели общую значимость.

Наша ближайшая задача - точная формулировка и доказательство принципа Хинчина для самого простого класса - сумматорных функций.

Определение

Функция $A(X)$,
$$A(X) = a(x_1) + a(x_2) + \dots + a(x_n) \quad (1)$$

называется сумматорной, если функция $a(x)$ одна и та же во всех слагаемых, а ее аргументы x, x_1, \dots, x_n задаются проекциями точки X ,

$$X = (x, x_1, \dots, x_n) \quad (2)$$

на все компоненты.

Теорема (Хинчин)

Любая сумматорная функция представима в виде суммы двух функций

$$A(X) = A(H) + F(X) \quad (3)$$

Первое слагаемое - доминанта $A(H)$ имеет величину порядка n и сохраняет постоянное значение на каждой поверхности уровня H - функции.

Второе слагаемое — флуктуанта $\mathcal{F}(X)$ может меняться вдоль поверхности уровня, но зато относительно мала в том смысле, что ее среднее квадратичное значение имеет порядок \sqrt{m} . Флуктуанта, конечно, не всюду (равномерно) мала. Обязательно найдутся точки, в которых она сравнима с доминантой. Однако в тех задачах, где множества малой меры не имеют значения, функция огромного числа переменных может быть заменена функцией $A(E)$ одной существенной переменной и притом тем точнее, чем больше переменных у функции $\mathcal{A}(X)$.

Суть идеи можно пояснить еще одним способом. Введем в пространстве X новую систему координат (E, Y) , используя функцию H , $E = H(X)$, в качестве одной из новых независимых переменных. Формула (3) в новых координатах имеет вид:

$$\mathcal{A}(E, Y) = A(E) \cdot \mathcal{F}(E, Y) \quad (4)$$

Точка зрения замены переменных полезна тем, что подчеркивает главную мысль. Равноправные и симметричные переменные x_1, x_2, \dots, x_n неудобны при рассмотрении вопросов, относящихся к системе в целом. Нужно ввести иерархию независимых переменных, выбрав $E = H(X)$ в качестве главного переменного (аналогичного радиусу в полярной системе координат). Тогда оказывается, что все сумматорные функции почти не зависят от огромного числа "угловых" переменных Y , а зависят только от "радиуса" E . Для выделения доминанта из сумматорной функции несущественно как именно выбраны координаты Y на поверхности уровня H -функции, (5)

$$E = H(X)$$

Употребление буквы E всюду в дальнейшем означает, что рассмотрение ведется в координатах E, Y . Если же выкладки

удобнее вести в исходных переменных x_1, x_2, \dots, x_n , то естественно сохранить для E обозначение $H(X)$ и термин H — функция.

п.2. Усреднение, как способ вычисления доминанты

Основная мысль состоит, конечно, в том, что доминанта вообще существует X' . После того как вопрос о доминанте поставлен и угадана основная формула разложения сумматорной функции, само стыскание составляющих труда не представляет.

Определяющим свойством доминанты является ее постоянство вдоль поверхностей уровня $H(X) = E$, поэтому она не меняется при осреднении по этим поверхностям.

Напротив того флуктуанту можно задать как раз условием равенства нулю ее среднего значения по каждой поверхности уровня. Покажем, что это наилучший выбор флуктуанты.

Пусть есть два разложения сумматорной функции. Одно разложение,

$$A(X) = A(H) + S(X), \quad (6)$$

построено так, что среднее от флуктуанты равно нулю. Второе разложение произвольно

$$A(X) = A'(H) + F(X) \quad (7)$$

Выразим среднее значение квадрата второй флуктуанты через первую:

X/ Несомненно, что замысел навеян законом больших чисел. Но когда шпленок идеи выдулился из яйца догадки нужно немедленно освободиться от скорлупы вероятностной терминологии, заслоняющей глаза новску существу и путающейся у него в ногах.

$$\begin{aligned} \overline{\mathcal{F}(X)} &= \overline{[A(H) - A'(H) + \mathcal{F}(X)]^2} = \\ \overline{[A(H) - A'(H)]^2 + 2[A(H) - A'(H)] \mathcal{F}(X) + \mathcal{F}(X)^2} &= \\ \overline{[A(H) - A'(H)]^2 + 2[A(H) - A'(H)] \overline{\mathcal{F}(X)} + \overline{[\mathcal{F}(X)]^2}} & \end{aligned}$$

Итак

$$\overline{[\mathcal{F}(X)]^2} = [A(H) - A'(H)]^2 + \overline{[\mathcal{F}(X)]^2} \quad (8)$$

В этой традиционной выкладке, идейно восходящей к тригонометрическим рядам, использованы только два свойства постоянство любой доминанты вдоль поверхностей уровня и равенство нулю среднего от "правильной" флуктуанты. Полученное равенство доказывает экстремальное свойство правильного разложения.

Среди всех возможных разложений наименьшую квадратичную погрешность обеспечивает доминанта, получаемая осреднением сумматорной функции вдоль поверхностей уровня $H(X) = E$

Стоит подчеркнуть два обстоятельства. Во-первых, экстремальное свойство является точным, а не асимптотическим. Во-вторых, идея осреднения вторична и вытекает из идеи доминанты. Физики нередко (4) пишут о "гипотезе молекулярного хаоса", как о самостоятельной физической идее. Математик обязан уточнить, что осреднением доказывается вовсе не "доминантность" доминанты, а всего лишь единственность наилучшего разложения.

Это уточнение тем более необходимо, что доказательство сохраняет силу для произвольной функции. Для любой функции, ее наилучшую аппроксимацию при помощи функций одной X/ Как-будто хаос становится лучше от того, что он "молекулярный".

переменной, всегда дает метод осреднения. Не это вовсе не означает, что он всегда дает хорошую аппроксимацию. Это особенно ясно в случае среднего, равного нулю, когда разложение оказывается полностью бессодержательным.

В книгах [5] по физике (явно или неявно) ссылаются обычно на практический успех такой процедуры при решении некоторых конкретных задач. Однако для новых ситуаций важно знать причину успеха и предвидеть возможность неудачи. Подробный анализ привел Хинчина к выводу, что существо дела в свойствах тех конкретных функций, которые реально интересны в статистической физике — их сумматорности. Но и в этом случае необходима осторожность. Даже для сумматорных функций аппроксимация становится хорошей только асимптотически, при $n \gg 1$.

п.3. Алгоритм вычисления дивергенции

Основная идея выкладки, приводящей к экстремальному свойству, допускает модификацию, удобную для дальнейшего. Умножим обе части равенства (6) на произвольную (бесконечно-дифференцируемую) функцию $f(E)$ и проинтегрируем по всему пространству X .

$$\int_X f(H) A(X) dX = \int_X f(H) A(H) dX \quad (9)$$

Второй член справа равен нулю, так как обращая в нуль, его интеграл по каждой поверхности уровня $H=E$. Преобразуем полученное равенство, учитывая, что f — сумматорная функция, а справа под знаком интеграла стоит функция только H .

$$\sum_X \int f(H) a(x_i) dX = \int f(E) A(E) \Omega(E) dE \quad (10)$$

Свойства и асимптотика структурных функций подробно разобраны в (2). Все интегралы слева одинаковы и равны, поэтому достаточно вычислить один из них. Обозначим через Y прямое произведение всех компонент кроме первой

$$Y = (x_1, \dots, x_n) \quad (2)$$

Тогда X можно представить в виде прямого произведения двух компонент (опуская индекс y малой компоненты)

$$X = (x, Y) \quad (3)$$

Поэтому

$$\int_X f(h) a(x) dx = \int_Y \int_X f(h) a(x) dx dY =$$

$$\int_Y a(x) \left[\int f(E+h) \Omega(E) dE \right] dY \quad (4)$$

Сдвинем переменную интегрирования во внутреннем интеграле и переставим порядок интегрирования. Получится следующее равенство, справедливое для любой функции $f(E)$:

$$\int f(E) \left[\int n a(x) \Omega(E-h(x)) dx \right] dE = \int f(E) A(E) \Omega(E) dE \quad (5)$$

Из него вытекает равенство для обобщенных функций,

$$A(E) \Omega(E) = n \int a(x) \Omega(E-h(x)) dx \quad (6)$$

которое служит для вычисления $A(l)$. Следует иметь в виду, что все функции кроме $a(x)$ и $h(x)$ зависят от числа n и интерес представляет именно их асимптотическое поведение при $n \gg 1$.

Полученное равенство можно записать в виде

$$\frac{A(E)}{n} = \int a(x) K_n(E, h(x)) dx \quad (7)$$

где здесь $K_n(E, h)$ зависит только от скалярных аргументов E и h и является решением уравнения

$$K_n(E, h) \Omega_n(L) = \Omega_n(E-h) \quad (8)$$

Если бы была определена операция деления на обобщенную функцию, то уравнение для ядра имело бы очевидное решение:

$$K_0(E, b) = \frac{0 \cdot (F-1)}{1 \cdot (F)} \quad (16)$$

Однако обобщенного определения не существует, поэтому полученная формула имеет точный смысл, если (F) является обычной функцией в асимптотическом смысле в случае обобщенной функции. В этом последнем случае смысл формулы (16) следующий - разложение ядра по обратным степеням получается формальным делением разложений числителя и знаменателя.

Любопытно, что уравнение (17) не только более осмысленно, но и более удобно для фактического вычисления асимптотики ядра.

Несложные, но довольно громоздкие K_1 выкладки, основанные на обобщенной формуле Больцмана, приводят к следующему результату

$$K_1(E, b) = \frac{\sum_{i=1}^n \rho^i}{\rho^i} \left[1 + \frac{1}{\rho^i} \right] \quad (17)$$

где величина ρ (интенсивность) связана с E "уравнением состояния"

$$\frac{E}{n} = - \frac{d \ln \rho}{d \rho}$$

Главное отличие этого вывода (по существу, конечно, совпадающее со выводом Хинчина в его книге) состоит в установлении прямой связи с обобщенной формулой Больцмана.

Полезно собрать вместе формулы, дающие метод вычисления доминанты. выкладки проводятся в перемете на одну компоненту и в главном члене асимптотической теории. Во всех формулах поправки имеют порядок $\frac{1}{n}$, так как все они

X/ Очень интересно, что эти выкладки содержат "вычитание бесконечностей". Главные члены в $\frac{1}{n}$ и $\frac{1}{n^2}$ сокращаются, что в терминах "энтропии" (то есть логарифма структурной функции) означает вычитание членов, пропорциональных $\frac{1}{n}$.

являются следствием поправки к ядру, имеющей именно такую величину.

По заданной функции $h(x)$ строится функция $q(\beta)$

$$q(\beta) = \int_x e^{-\beta h(x)} dx, \quad (21)$$

после чего решается (относительно интенсивности β) уравнение

$$\frac{E}{n} = \frac{1}{q(\beta)} \int h(x) e^{-\beta h(x)} dx = -\frac{d \ln q}{d \beta}. \quad (22)$$

Впрочем, как уже неоднократно говорилось, удобнее считать β независимым переменным, а формулу (22) рассматривать, как определение зависимости E от n и β .

Величина доминанты находится по формуле, вполне аналогичной формуле для E

$$\frac{A}{n} = \frac{1}{q(\beta)} \int a(x) e^{-\beta h(x)} dx + O\left(\frac{1}{n}\right) \quad (23)$$

Отличие состоит в том, что формула для E является точной, а формула для A асимптотической. Второе важное отличие заключается в необходимости интегрирования по пространству x , в то время как E можно найти (вторая часть формулы (22)) дифференцированием по переменной β .

Найденная формула для среднего значения допускает очень важное для дальнейшего обобщения. Выкладки, вполне аналогичные проведенным дадут формулу среднего для функции $a(x; y)$

зависящей от двух компонент

$$O_{\epsilon}(\beta) = \frac{1}{\sqrt{\pi}} \int \int \alpha(x, y) e^{-\beta h(x) - \beta h(y)} dx dy + O\left(\frac{1}{n}\right) \quad (24)$$

Обобщение на случай большого числа компонент очевидно. Однако нужно иметь в виду, что формула

$$O_{\epsilon}(\beta) = \frac{1}{\sqrt{\pi}} \int \dots \int \alpha(x_1, \dots, x_n) e^{-\beta h(x_1) - \dots - \beta h(x_n)} dx_1 \dots dx_n + O\left(\frac{1}{n}\right) \quad (25)$$

является асимптотически правильной только в том случае, если число n фиксировано или, по крайней мере, растет медленнее n^2 , так как формула для ядра имеет поправку порядка $\frac{1}{n^2}$.

Впрочем для наших ближайших целей вполне достаточно формулы (24), в которой важную роль играет поправка $O\left(\frac{1}{n}\right)$.

п.4. Оценка флуктуанты

Дальнейшие выкладки текстуально совпадают с аналогичными выкладками А.Я.Хинчина. Тем не менее их повторение целесообразно хотя бы потому, что книга Хинчина является сейчас библиографической редкостью.

флуктуанта $\mathcal{F}(X)$, так же как и основная функция $\mathcal{H}(X)$, относится к классу сумматорных функций:

$$\mathcal{F}(X) = f(x_1) + f(x_2) + \dots + f(x_n) \quad (26)$$

Однако слагаемые флуктуанты

$$f(x) = \alpha(x) - \bar{\alpha} \quad (27)$$

имеют среднее, равное нулю, так как вычитаемое является как раз средним значением уменьшаемого

$$\overline{f(x)} = 0 \quad (28)$$

Квадрат флуктуанты содержит n существенно положительных слагаемых и $n^2 - n$ величин, средние от которых необходимо оценить:

$$\mathcal{F}^2(X) = \sum_{i=1}^n f^2(x_i) + \sum_{i \neq k} f(x_i) f(x_k) \quad (29)$$

Так как средние не зависят от индексов, то для среднего от \mathcal{F}^2 получаем простое выражение

$$\overline{\mathcal{F}^2(X)} = n \overline{f^2(x)} + (n^2 - n) \overline{f(x)} \overline{f(y)} \quad (30)$$

Из формулы (24) вытекает, что в главном члене среднее от произведения равно произведению средних. Эта асимптотическая независимость разных слагаемых и приводит к тому, что главный член среднего от квадрата флуктуанты оказывается величиной порядка всего лишь h , а не n^2 . Действительно:

$$\frac{1}{\sqrt{(\beta)}} \iint_{x,y} f(x)f(y) e^{-\beta h(x) - \beta h(y)} dx dy = \frac{1}{\sqrt{(\beta)}} \int_x f(x) e^{-\beta h(x)} dx \frac{1}{\sqrt{(\beta)}} \int_y f(y) e^{-\beta h(y)} dy = 0 \quad (31)$$

Поэтому поправка к ядру в формуле (24) дает оценку

$$\overline{f(x)f(y)} = O\left(\frac{1}{n}\right) \quad (32)$$

которая приводит к оценке флуктуанты,

$$\overline{\mathcal{F}^2(X)} = O(n), \quad (33)$$

завершающей доказательство репрезентативности доминанты.

Заключение:

Наглядно говоря, теорема Пифагора является основной причиной улучшения аппроксимации суммарной функции ее доминантой.

Конец ряда отрезков, укладываемых вдоль прямой, удаляется со скоростью $1/n$. Отрезки, используемые в качестве катетов дают рост гипотенузы всего лишь со скоростью \sqrt{n} . Поэтому прямолинейная составляющая рано или поздно становится доминирующей.

Литература

- 1 Хинчин А.Д. Математические основания статистической механики, Гостехиздат Москва 1943г.
- 2 Молчанов А.М. Теория выравнивания и формула Больцмана. Преприят ИЛМ 1968г.
- 3 Лебег, Анри Об измерении величин. Москва, Учпедгиз 1938, стр.80
- 4 Пригожин И. Введение в термодинамику необратимых процессов. М.ИИИ 1960г.
- 5 Хилл Т. Статистическая механика. М. ИИИ 1960г.

Т 17657 ст 30 XII 1908 г. Заказ № 1 1908 г.

Институт прикладной математики
Москва, Мясницкая ст., 4