

ЭНТРОПИЯ И ИНФОРМАЦИЯ – УНИВЕРСАЛЬНЫЕ ЕСТЕСТВЕННОНАУЧНЫЕ ПОНЯТИЯ*

С.М. Коротаев

Содержание

1. Введение
2. Основные сведения из теории информации
3. Различные формы определения энтропии
4. Второе начало термодинамики и 7-я теорема Шеннона
5. Расслоенные пространства определения энтропии
6. Статистический ансамбль
7. Термодинамическое соответствие
8. Заключение

1. Введение

Трудно найти понятия более общие для всех наук (не только естественных) и, вместе с тем, иногда носящих оттенок загадочности, чем энтропия и информация. Отчасти это связано с самими названиями. Если бы не звучное название «энтропия» осталась бы с момента первого рождения всего лишь «интегралом Клаузиуса», вряд ли она бы не рождалась вновь и вновь в разных областях науки под одним именем. Кроме того, ее первооткрыватель Клаузиус, первым же положил начало применению введенного им для, казалось, бы узкоспециальных термодинамических целей понятия к глобальным космологическим проблемам (тепловая смерть Вселенной). С тех пор энтропия многократно фигурировала в оставшихся навсегда знаменитыми спорах [1]. В настоящее время универсальный характер этого понятия общепризнан и она плодотворно используется во многих областях.

Термин «информация» замечателен тем, что, существующему с давних пор бытовому понятию, К.Шенноном [2] был придан математически точный смысл. Неопределенно-бытовой смысл этого термина уже научного. Это приводило и приводит ко многим недоразумениям и спекуляциям. Интересно и то, что К.Шеннон как создатель теории информации – по существу, раздела математики, был не чистым математиком, а инженером-теоретиком. Поэтому его работы написаны языком ясным для понимания инженеров, естественников и даже сведущих в математике гуманитариев. Профессиональные

* Работа поддержана Российским гуманитарным научным фондом, грант № 00-03-00360а.

математики проявили активность в этой области позднее, но их капитальный подход (см., например [3]) не востребован пока в приложениях, за исключением, возможно, работ А.Н. Колмогорова [4,5].

Наш подход по степени формализованности будет близок к шенноновскому, доступным для понимания широкому кругу читателей и, при этом, вполне строгим. Основная цель обзора – так описать вопросы связанные с энтропией, чтобы без изложения многочисленных примеров стала ясна общность этого понятия и читатель был бы готов, прочитав его, самостоятельно применять энтропийный подход в своей области. Круг читателей, на которых рассчитан обзор – это как раз не специалисты в термодинамике и в теории информации, а физики всех других направлений, астрономы, геофизики, геологи, океанологи, биофизики и т.д. В чем-то обзор может быть полезен также экономистам и социологам. Объединяющее начало здесь – необходимость изучения естественных процессов, разворачивающихся во времени или пространстве (вариаций). Под естественными процессами мы понимаем такие, параметры которых (величина, длительность) заранее неизвестны. Это близко к понятию случайного процесса, но не совпадает с ним, поскольку естественный процесс может быть детерминированным. Мы используем язык классической теории информации, в связи с чем в п.1 приводятся без доказательств основные положения этой теории. Для углубления понимания рекомендуется прочесть книги [5-10] (пронумерованные в порядке нарастания строгости, от совершенно популярной до фундаментальной научной литературы. Следует также отметить, что мы не будем следовать историческому подходу, обычно принятому в курсах физики и начинать с термодинамики, где впервые появилось понятие энтропии. Желающим глубоко изучить термодинамические и теплотехнические применения энтропии стоит прочитать прекрасную книгу П. Шамбадала [1]. Мы же, напротив, будем скорее исходить из теории информации и с этих позиций обосновывать общность термодинамического подхода.

1. Основные сведения из теории информации

Базисным понятием всей теории информации является понятие энтропии. Энтропия – мера неопределенности некоторой ситуации. Можно также назвать ее мерой рассеяния и в этом смысле она подобна дисперсии. Но если дисперсия является адекватной мерой рассеяния лишь для специальных распределений вероятностей случайных величин (а именно – для двухмоментных распределений, в частности, для гауссова распределения), то энтропия не зависит от типа распределения. С другой стороны, энтропия вводится так, что-

бы обладать, кроме универсальности и другими желательными свойствами. Так, если некий опыт имеет n равновероятных исходов, а другой опыт m равновероятных исходов, то составной опыт имеет nm таких исходов. Если мы вводим меру неопределенности f , то естественно потребовать, чтобы она была такова, чтобы во-первых, неопределенность росла с ростом числа возможных исходов, а во-вторых, неопределенность составного опыта была равна просто сумме неопределенности отдельных опытов, иначе говоря, мера неопределенности была аддитивной: $f(nm)=f(n)+f(m)$. Именно такая удобная мера неопределенности была введена К. Шенноном:

$$H(X) = - \sum_{i=1}^N P(X_i) \log P(X_i),$$

где X – дискретная случайная величина с диапазоном изменчивости N , $P(X_i)$ – вероятность i – го уровня X .

В дальнейшем мы будем рассматривать X как некоторую физическую величину, меняющуюся во времени или пространстве. Временной или пространственный ряд X_j (j – индекс временной или пространственной координаты r) будем называть, как это принято в ряде естественных наук, «вариацией». В самой теории информации такое пространственно-временное упорядочение совершенно не обязательно, но, во-первых, анализ именно таких вариаций составляет суть всех естественных наук, во-вторых, это с первых шагов позволяет лучше ощутить смысл новых понятий. Заметим также, что если даже пространственная или временная упорядоченность величины X в явном виде отсутствует, она неизбежно существует неявно. Например, положим, что j – номер различимой частицы, а X_j – ее импульс. X – неупорядоченная случайная величина (ее номер j присваивается произвольно), но все эти частицы неизбежно разнесены в пространстве (раз мы можем их различить) и, при необходимости, мы можем их соединить некоторой (ломаной) осью и восстановить упорядоченность. Но для понимания проще представлять X как сигнал, который может быть записан самописцем, как рельеф местности вдоль некоторого профиля, как пространственное распределение плотности энергии поля и т.п.

Таким образом, чтобы рассчитать $H(X)$, берется запись вариации X_j , разность между максимальным и минимальными значениями X_j разбивается на N квантов (обычно равных разрешающей способности прибора) и подсчитывается число m_i заполнения каждого i – го уровня (число благоприятных случаев). Общее число случаев M – это число пространственных или временных ячеек, опять-таки обычно определяемых разрешением прибора. В результате мы получаем

распределение вероятностей $P(X_i)=m_i/M$, которое подставляем в формулу $H(x)$.

В теории информации в формуле для энтропии обычно используют двоичные логарифмы, тогда (энтропия и информация) измеряется в битах. Это удобно тем, что выбор между двумя равновероятными уровнями X_i (как в двоичном) сигнале характеризуется неопределенностью 1 *бит*. В популярной литературе иногда пользуются десятичными логарифмами и единицей энтропии является *дит*. В физике удобнее пользоваться натуральными логарифмами и единицей энтропии является *нат* (поскольку в дальнейшем наш подход существенно физический, мы также используем натуральные логарифмы). Выбор основания – лишь вопрос масштаба, в любом случае энтропия безразмерна. Возможная величина энтропии заключена в пределах:

$$0 \leq H(X) \leq \log N.$$

Нижняя грань соответствует вырожденному распределению. Неопределенность величин X отсутствует. В вариационном ряду это соответствует $X_j=const$. Верхняя грань соответствует равномерному распределению. Все N значений X_i встречаются с равной вероятностью. В вариационном ряду это может соответствовать, в частности, линейному тренду $X_j=ar_j$.

Если две случайные величины X и Y , каким-то образом связанные друг с другом (например на входе и выходе какой-то системы), то знание одной из них, очевидно уменьшает неопределенность значений другой. Остающаяся неопределенность оценивается условной энтропией. Так, условная энтропия X при условии знания Y определяется как:

$$H(X|Y) = \sum_{k=1}^K P(Y_k) \sum_{i=1}^N P(X_i|Y_k) \log P(X_i|Y_k),$$

где $P(X_i|Y_k)$ – условные вероятности (вероятность i -го значения X при условии $Y=Y_k$), диапазоны изменчивости X и Y (соответственно N и K) не обязательно совпадают.

Чтобы рассчитать $H(X|Y)$, рассчитывают K энтропий X , соответствующих фиксированному Y_k и затем суммируют результаты с весами $P(Y_k)$. Очевидно, условная энтропия меньше безусловной, точнее:

$$0 \leq H(X|Y) \leq H(X).$$

Нижняя грань соответствует однозначной зависимости X от Y , верхняя – полной независимости.

Информация определяется разностью между безусловной и условной энтропиями. Это уменьшение неопределенности «знания чего-то

за счет того, что известно что-то». При этом замечательно, что информация I симметрична, т.е. $I_{YX}=I_{XY}$:

$$I_{XY}=H(X)-H(X/Y)=H(Y)-H(Y/X)= I_{YX}.$$

Информация всегда неотрицательна; она равна нулю, когда X и Y независимы; информация максимальна и равна безусловной энтропии, когда между X и Y имеется однозначная зависимость. Таким образом, безусловная энтропия – это максимальная информация, потенциально содержащаяся в системе (вариационном ряде). Заметим, что мы сказали однозначная, но не взаимно-однозначная зависимость. Это значит, что несмотря на симметрию, верхние грани I_{XY} и I_{YX} отличаются:

$$0 \leq I_{XY} \leq H(X), \quad 0 \leq I_{YX} \leq H(Y).$$

Как это может быть? Положим, $X \Rightarrow Y$ (но обратное неверно). Тогда $H(Y/X)=0$, $H(X/Y) \neq 0$, $I_{YX}=H(Y)= I_{XY}$. Очевидно, это возможно только когда $H(X) > H(Y)$.

Информация – это всего лишь характеристика степени зависимости некоторых переменных, ничего более загадочного в ней нет. Зато это предельно общая характеристика. Ее можно сравнить с корреляцией, но если корреляция характеризует лишь *линейную* связь переменных, информация характеризует *любую* связь. Тип связи может быть совершенно любым и, более того, неизвестным нам. Это не мешает рассчитать информацию, количественно сравнивать между собой разнотипные зависимости и т.д. Платой за общность является лишь невозможность, зная количество информации написать уравнение связи переменных (в отличие от того, как корреляция позволяет легко переходить к регрессии). Можно определить и совместную энтропию X и Y по их двумерному распределению. При этом:

$$H(XY)=H(X)+H(Y/X)=H(Y)+H(X/Y).$$

Иначе говоря, энтропия субаддитивна, аддитивность ($H(XY)=H(X)+H(Y)$) достигается только при полной независимости X и Y .

С помощью совместной энтропии можно написать выражение для информации $I_{XY}= I_{YX}=I$ в симметричном виде:

$$I=H(X)+H(Y)-H(XY).$$

Интуитивно ясно, что включение в рассмотрение третьей переменной может лишь увеличить информацию. Это действительно так:

$$I_{(YX)Z} \geq I_{XZ}, \quad I_{(XY)Z} \geq I_{YZ}.$$

Симметрия случая трех переменного описывается формулой тройной информации:

$$I_{(XY)Z} + I_{XY} = I_{(YZ)X} + I_{YZ} = I_{(XZ)Y} + I_{XZ}.$$

Важную роль в теории информации играет представление о максимальной скорости передачи сообщения. При этом имеется в виду скорость при которой еще возможен безошибочное получение информации на приемном конце канала связи. «Канал связи» при создании теории информации рассматривался действительно как технический канал (и задача состояла в изучении его потенциальных возможностей, тех теоретических пределов его использования, которые нельзя превзойти в принципе). В действительности «каналом связи» является любая естественная или искусственная система в которой можно выделить начальное состояние (входной сигнал) X и конечное (выходной сигнал) Y . Законы передачи информации по каналу связи универсальны, поэтому, например, любое физическое взаимодействие в реальных системах подчиняется не только соответствующему физическому закону, но и им также. Несмотря на сложившуюся терминологию, лучше говорить не о скорости передачи, а о скорости приема информации, которая ограничена неравенством

$$v \leq \frac{C}{H(Y)},$$

где C – пропускная способность канала (иногда ее называют также информационной емкостью, иногда емкостью называется произведение $\delta t \cdot C$, где δt – длительность элементарного сигнала). При отсутствии шумов:

$$C = \frac{1}{\delta t} \log N,$$

при наличии шумов:

$$C = \frac{1}{\delta t} \max I_{XY},$$

максимум отыскивается при возможном варьировании входного сигнала X . Таким образом, в общем случае:

$$v \leq \frac{1}{\delta t} \frac{\max(H(Y) - H(Y|X))}{H(Y)}.$$

Чем больше шумов или, на физическом языке, чем более открыта система, тем медленнее выход (конечное состояние) может воспроизвести изменения входа (начального состояния). Важную роль играет также понятие избыточности R :

$$R = 1 - \frac{H(X)}{\log N},$$

$$0 \leq R \leq 1.$$

Нижняя грань достигается при равномерном распределении, верхняя – при вырожденном ($X_j=const$). Чем больше избыточность поля, тем, очевидно, меньше скорость передачи информации. Но тем меньше вероятность ошибки при приеме! Например, чрезмерно избыточный текст скучно читать (избыток «воды»). Но малоизбыточный читать трудно, он требует крайнего внимания, в пределе (недостижимом в обычном языке) ни одну утерянную букву невозможно восстановить по смыслу. Таким образом, избыточный сигнал содержит внутреннее корреляции, которые используются при восстановлении (декодировании) сообщения. Увеличивая избыточность и, соответственно, уменьшая скорость приема-передачи можно воспроизвести на выходе информацию, поданную на вход при любом уровне шумов.

Канал, для которого $H(Y/X)=0$ называют детерминированным, для которого $H(X/Y)=0$ – бесшумовым. Последний термин неудачен, поскольку зашумленной может быть не только величина X , но и Y . Канал, для которого $H(Y/X)=H(X/Y)=0$, называют идеальным. В идеальном канале X и Y являются взаимно – однозначными функциями. Среди нескольких теорем Шеннона особую роль в естественно-научных приложениях играет 7-я теорема, которая гласит, что в замкнутой системе энтропия при любом преобразовании не увеличивается:

$$H(Y) \leq H(X).$$

На менее строгом языке это можно перефразировать так, что информацию на пути от входа к выходу можно только потерять, но не увеличить. Интуитивно это очевидно. Но если X – начальное состояние замкнутой физической системы, а Y – конечное, то эта теорема означает необратимое уменьшение энтропии, что диаметрально противоположно общеизвестной формулировке второго начала термодинамики. Мы рассмотрим этот вопрос в п.4.

Следует отметить одну исторически обусловленную особенность приложений теории информации. С момента зарождения [11] и после наиболее мощного импульса в ее развитии приданного работами К. Шеннона [2], несмотря на общность математического аппарата и быстро замеченную ассоциированность с термодинамикой [1, 2], теория информации развивалась как раздел теории связи, Приложения быстро вышли за пределы теории связи как таковой [12]. Но и в расширенных приложениях от теории связи в неявном виде унаследовано представление об одушевленном потребителе информации. Более того, на начальном этапе авторы специально приносили своего рода извинения за то, что термин «информация» в теории имеет более широкий смысл, чем его бытовое употребление, подразумеваю-

щее некоторую ценность для потребителя. В дальнейшем оказалось, что и понятие ценности информации может быть формализовано [9].

Существует однако более общая концепция приложения понятия информации как универсальной меры физического взаимодействия. Этому посвящен ряд работ с достаточно специальным подходом, например, [13-17], наиболее общая трактовка дана Х. Хармутом [18]. Наша трактовка опирается именно на эту концепцию. Но прежде чем применить ее, необходимо подробно разобрать такое базисное понятие как энтропия, поскольку, с одной стороны возможны различные формулировки, а с другой, наоборот – одинаковые математические формы имеют разный смысл (и приводят к различным результатам) в зависимости от наполнения.

3. Различные формы энтропии

Энтропия (S или H) независимо вводилась Клазиусом:

$$S = \int \frac{dQ}{T}, \quad (1)$$

где Q – энергия, T – температура; Л. Больцманом:

$$S = H = \ln W, \quad (2)$$

где W – вероятность состояния; К. Шенноном:

$$H = -\sum_{i=1}^N P(X_i) \ln P(X_i), \quad (3)$$

где P – вероятность i -го уровня переменной (сигнала) X с диапазоном изменчивости N ; фон Нейманом:

$$H = -tr \rho \ln \rho, \quad (4)$$

где плотность ρ определяется через волновую функцию ψ : $\rho = |\psi\rangle\langle\psi|$.

Оказалось, что (1) и (3) являются следствиями (1) [1,2,10], хотя для их применения, соответственно в термодинамике и в теории информации, это не имеет большого значения. (4) является квантовомеханическим обобщением (3), при этом (4) \equiv (3) для чистых состояний. Классический уровень предоставляет слишком широкие возможности для обобщений, чтобы затрагивать в данном обзоре и квантовый, поэтому мы не будем переходить на последний и лишь слегка коснемся удивительных особенностей квантовой энтропии в Заключение.

Можно заметить, что в отличие от (2) – (4) согласно определению (1) энтропии размерна (Дж/К). Здесь мы просто сохранили оригинальные определения. Безразмерная энтропия, конечно удобнее и, чтобы привести (1) к безразмерному виду, надо просто разделить правую часть на постоянную Больцмана k . Эта постоянная не имеет большего смысла, чем коэффициент связи между единицами измерения энергии и темпе-

ратуры. Если бы мы стали измерять температуру в джоулях (что неудобно, но законно), то надобность в этой константе отпала бы. Однако здесь есть чисто количественный нюанс. Величина k весьма мала: $1,38 \cdot 10^{-23}$ Дж/К. Разделив (1) на k мы сразу получаем представление о том, насколько велики изменения энтропии в самых заурядных термодинамических процессах по сравнению с теми, которыми мы оперируем в информатике. Это дает представление о том, насколько велика недоступная на макроуровне информация о микросостоянии вещества. Определение (1) – практически самое важное для теплофизике [1], но, пользуясь им, трудно увидеть универсальность понятия энтропии. Принципиальным недостатком (1) является также то, что это формула верна только для квазиравновесных состояний. Поэтому мы сосредоточимся на определениях (2) и (3).

Вероятность состояния W определяется как отношение числа микросостояний K , благоприятных данному макросостоянию, к полному числу возможных микросостояний L :

$$W = K/L. \quad (5)$$

Пусть рассматриваются пространственные или временные вариации некоторого параметра X (в дискретном виде). Тогда K – это число вариаций, удовлетворяющих данному распределению вероятностей $P(X_i)$, L – полное число возможных вариаций.

Пусть M – число членов статистического ансамбля (в нашем контексте это объем, а для одномерного случая – длина ряда, выраженная числом искусственных квантов пространства, т.е. отсчетов), m_i – число заполнения i -го уровня (т.е. $P(X_i)=m_i/M$). Тогда

$$K = \frac{M!}{\prod_{i=1}^N m_i!}. \quad (6)$$

Число L однозначно определить обычно сложнее. Поэтому вместо W пользуются K (так называемая абсолютная или термодинамическая вероятность) и определяют не нормированную энтропию (2), а абсолютную H_a :

$$H_a = \ln K = M \ln M - \sum_{i=1}^N m_i \ln m_i = -M \sum_{i=1}^N P(X_i) \ln P(X_i). \quad (7)$$

Можно показать [10], что энтропия Клазиуса (1) является частным случаем (7). Нормируя на число членов ансамбля (длину ряда), получаем абсолютную удельную энтропию

$$H_{ay} = - \sum_{i=1}^N P(X_i) \ln P(X_i), \quad (8)$$

$$0 \leq H_{ay} \leq \ln N, \quad (9)$$

совпадающую с (3), которая была введена К.Шенноном просто как удобный функционал и была названа им «энтропия» именно из-за обнаруженного родства с термодинамической энтропией. Верхняя грань достигается при равномерном распределении, нижняя – при вырожденном (X не зависит от координаты r : $X_j = \text{const}$).

Микросостояния в нашем случае различимы, следовательно ансамбль удовлетворяет статистике Максвелла-Больцмана. Роль различимой частицы в геофизической интерпретации играет квант пространства (или времени). Для этой статистики в термодинамике число L определяется как [10]:

$$L_l = N^M \quad (10)$$

и соответствует в теории информации числу возможных сообщений [8]. На нашем языке это можно перефразировать как число различных вариаций в диапазоне N при объеме (или длине ряда) M . Тогда согласно (5) нормированная удельная энтропия равна:

$$H_{ly} = -\sum_{i=1}^N P(X_i) \ln P(X_i) - \ln N. \quad (11)$$

Вполне естественное для термодинамики (M – число частиц в фиксированном объеме N) и теории технического канала связи (M – длина сообщения, N – фиксированный диапазон уровней) определение l_l трудно считать адекватным в контексте естественных процессов. Здесь M может быть естественным образом ограничена, и даже искусственное ограничение не вызывает особых затруднений (подобно спектральному анализу и вообще выборочной статистике), но диапазон N трудно считать фиксированным, в особенности для естественных полей. Так при гауссовом распределении $N \sim \sqrt{M}$.

В статистике Бозе-Эйнштейна принято другое определение L [10], которое в нашем контексте может быть интерпретировано как число различных вариаций с данным интегралом (точнее, суммой) U при объеме M :

$$L_2 = \left(\frac{M+U-1}{U} \right) = \frac{(M+U-1)!}{(M-1)!U!}, \quad (12)$$

где $U = \sum_{i=1}^N m_i X_i = \sum_{j=1}^M X_j$, причем все $X_j \geq 0$ (практически это легко достигается рассмотрением вместо уровней X «энергетических» уровней X^2). Тогда

$$H_{2y} = -\sum_{i=1}^N P(X_i) \ln P(X_i) - \frac{M+U-1}{M} \ln(M+U-1) + \frac{U}{M} \ln U + \frac{M-1}{M} \ln(M-1). \quad (13)$$

При $M \gg 1$, $U \gg 1$ (естественное условие применения любой статистики):

$$H_{2y} = -\sum_{i=1}^N P(X_i) \ln P(X_i) - (1+u) \ln(M+U) + u \ln U + \ln M, \quad (13')$$

где $u = U/M$.

Таким образом, L – это, в итоге, не вообще полное число различных вариаций (которое бесконечно), а полное число при некоторых ограничениях – при фиксированном диапазоне N или при фиксированном интеграле U , который может в ряде случаев интерпретироваться как энергия вариации. Возможны и другие варианты ограничений.

Нормированные энтропии (11) и (13) имели бы наиболее ясный физический смысл, однако в случае изучения вариаций естественных полей возникают дополнительные трудности.

Так из (9) и (11) следует

$$-\ln N \leq H_{1y} \leq 0.$$

Но для геофизических вариационных рядов N определяется естественно только из самой реализации. Нижняя грань соответствует $X_j = \text{const}$, т.е. $N = 1$, и оба противоположных случая равномерного и вырожденного распределений смыкаются.

Нижняя грань H_{2y} для вырожденного распределения ($X_j = \text{const}$) определяется из (13') как

$$\inf H_{2y} = -2 \ln 2.$$

Заметим, что это выражение совпадает с классическим пределом энтропии квантовомеханического ансамбля, выведенным из других соображений [15].

Верхняя грань H_{2y} соответствует, однако, не равномерному распределению, а определяется из уравнения

$$\frac{M!}{\prod_{i=1}^N m_i!} = \frac{(M+U-1)!}{(M-1)!U!}.$$

Единственное решение этого уравнения $U=1$ соответствует импульсной функции в вариационном ряду X_j . Для геофизического вариационного ряда естественная длина для финитной функции X_j в данном случае получается $M=1$. Это вырожденный случай для самого функционала энтропии. Таким образом, формальное соотношение

$$-2 \ln 2 \leq H_{2y} \leq 0$$

не имеет такого же ясного смысла как (9).

По этим причинам мы в дальнейшем в основном будем пользоваться абсолютной удельной энтропией (8), опуская для краткости нижние индексы подобно (3).

4. Второе начало термодинамики и 7-я теорема Шеннона

Рассмотрим замкнутую систему, в которой X_j представляют плотности энергии (например, энергии электромагнитного поля) в пространственных ячейках r_j . Самопроизвольная эволюция системы сводится к конечному выравниванию плотности энергии по всем ячейкам $X(r_j) = X_j = const$. Второе начало термодинамики отражает этот факт через увеличение энтропии S до максимального значения:

$$S = -\sum_{j=1}^M P(X_j) \ln P(X_j) \rightarrow \ln M, \quad (14)$$

где M – число пространственных ячеек, $P(X_j)$ определяются здесь как

$$P(X_j) = \frac{X_j}{\sum_{j=1}^M X_j} = \frac{X_j}{U}. \quad (15)$$

В тоже время, воспользовавшись определением вероятности через частоту заполнения m_i i -го энергетического уровня

$$P(X_i) = m_i / M, \quad (16)$$

получим стремление шенноновской энтропии к минимальному значению

$$H = -\sum_{i=1}^N P(X_i) \ln P(X_i) \rightarrow 0. \quad (17)$$

Этот факт отражен в 7-й теореме Шеннона [2], утверждающей невозрастание энтропии выходного сигнала при любом преобразовании.

Это простое явление противоположного поведения термодинамической и информационной энтропий при полной идентичности математических форм (14) и (17), насколько известно автору, нигде в литературе специально не отмечалось. В то же время при применении теории информации к процессам физического взаимодействия оно способно приводить к известной путанице.

Противоположные значения $\inf S$ и $\sup H$ отвечают различным функциям $X(r_j)$

$$\inf S = 0$$

соответствует импульсной функции: $X = X_j$ при $r=r_j$, $X=0$ при $r \neq r_j$.

При этом легко показать, что

$$H = \left(\frac{1}{M} - 1 \right) \ln(M-1) + \ln M.$$

С другой стороны,

$$\sup H = \ln N.$$

В частности, это соответствует линейной функции $X_j = ar_j$. Пользуясь (15) после ряда преобразований получаем для этой функции

$$S = \ln \frac{M}{2} + \frac{1}{2}.$$

Итак, конечное состояние $X(r_j) = \text{const}$ характеризуется взаимнооднозначным соответствием S и H ($\sup S$ соответствует $\inf H$). Для иных состояний такого соответствия вообще говоря нет. Качественно это легко понять, т.к. H есть логарифм числа наблюдаемых микросостояний (числа вариаций), удовлетворяющих данному распределению вероятностей уровней $P(X_i)$, в то время как S есть логарифм числа ненаблюдаемых микросостояний, удовлетворяющих данному пространственному распределению $P(X_j)$ (т.е. сама пространственная вариация рассматривается как распределение вероятностей). Очевидно, что функционал H определен на меньшем множестве функций чем функционал S , поэтому и соответствия между ними в общем случае нет. Одному значению H может соответствовать множество значений S .

Ясное понимание различий S и H помогает легче установить математический смысл негэнтропийного принципа деградации Л. Бриллюэна [8] – получение некоторой информации о системе требует не меньшего увеличения энтропии системы. Негэнтропия Бриллюэна определяется как $\Delta N = -\Delta S$. Т.к. $\Delta N \leq 0$, то $\Delta H \leq 0$, то $\Delta N + \Delta H \leq 0$ или

$$\Delta (-S + H) \leq 0. \quad (18)$$

Соотношение (18) названо Л. Бриллюэном принципом деградации, причем если ΔS понимается им как изменение термодинамической энтропии, то ΔH – как некоторая не ясно определенная им свободная информация [8]. В нашем понимании (18) соответствует большей «чувствительности» S к изменениям $X(r_j)$ по сравнению с H .

5. Расслоенные пространства определения энтропии

Продемонстрированное в предыдущем разделе отличие энтропий S и H обязывает сделать некоторое обобщение. Все различие между S и H связано с различием пространств в которых действует оператор вероятности P .

Вообще оператор плотности вероятности p может быть определен исходя из процедуры ее нахождения по функции $X(r)$ [19]:

$$p = \sum_{k=1}^J \partial \Lambda_k, \quad (19)$$

где ∂ – оператор дифференцирования, Λ – оператор обращения (функции $X(r)$ в $r(X)$), k – символ интервала монотонности, J – общее число таких интервалов. Оператор вероятности является дискретным аналогом (19). Но для нас важен не сам вид оператора, а

пространство в котором он действует. Для энтропии S – это сам положительно определенный вариационный ряд (15), для энтропии H – распределение частот заполнения уровней (16). Можно заметить, что интервал изменчивости функции N в первом случае (S) переходит в интервал изменчивости аргумента во втором случае (H). Но ничто не мешает построенное распределение вероятностей уровней рассматривать как «вариацию» и с помощью той же процедуры построить новое распределение вероятностей. Другими словами оператор (19) может быть применен многократно. Для каждого нового распределения может быть определена энтропия.

Обобщение легко достигается в терминах теории расслоенных пространств [20]. Полное пространство в нашем случае – это множество вариаций, база – множество распределений. Характеристикой расслоенности является число K (6). Первым слоем является наблюдаемый вариационный ряд, вторым – распределение вероятностей его уровней и т.д. Число заполнения отождествляется с функцией X : $m^{(n)} = X^{(n)}$. Тогда между объемами M и интервалами изменчивости N справедливо соотношение

$$N^{(n)} = M^{(n+1)},$$

где n – номер слоя. Сумма U («полная энергия») может быть определена двояко:

$$U_1^{(n)} = \sum_{j=1}^{M^{(n)}} X_j^{(n)},$$

$$U_2^{(n+1)} = \sum_{i=1}^{N^{(n)}} r_i^{(n+1)} X_i^{(n+1)},$$

$$U_1^{(n)} = U_2^{(n+1)}.$$

С другой стороны:

$$M^{(n)} = \sum_{j=1}^{M^{(n+1)}} X_j^{(n+1)} = \sum_{i=1}^{N^{(n)}} r_i^{(n+2)} X_i^{(n+2)},$$

$$N^{(n)} = \sum_{j=1}^{M^{(n+2)}} X_j^{(n+2)} = \sum_{i=1}^{N^{(n+1)}} r_i^{(n+3)} X_i^{(n+3)},$$

где $r^{(n)}$ – аргумент (координата в n – ом слое. Следовательно,

$$U_1^{(n)} = M^{(n-1)} = N^{(n-2)}, \quad (20)$$

$$U_2^{(n)} = M^{(n-2)} = N^{(n-3)}. \quad (21)$$

Таким образом, разница между M , N , U относительна и определяется слоем, к которому они относятся. Вместо (6), использующего параметры первого и второго слоев, можно написать общее выражение:

$$K^{(n)} = \frac{M^{(n-1)}}{\prod_{j=1}^{M^{(n)}} X_j^{(n)}!} = \frac{N^{(n-2)}}{\prod_{i=1}^{N^{(n-1)}} X_i^{(n)}!}. \quad (22)$$

Абсолютная удельная энтропия определенная по n – ому слою в общем виде:

$$H^{(n)} = \frac{\ln K^{(n)}}{M^{(n-1)}} = \frac{\ln K^{(n)}}{N^{(n-2)}}. \quad (23)$$

Числитель (22) и знаменатель (23) могут быть заменены также на $U_1^{(n)}$ $U_2^{(n+1)}$ в соответствии с (20) и (21).

Аналогично могут быть представлены обобщения чисел возможных сообщений L_1 (10) и L_2 (12) для произвольного слоя:

$$L_1^{(n)} = [M^{(n+1)}]^{M^{(n)}} = [N^{(n)}]^{N^{(n-1)}}, \quad (24)$$

$$L^{(n)} = \frac{[M^{(n)} + M^{(n-1)} - 1]!}{[M^{(n)} - 1]! M^{(n-1)}!} = \frac{[N^{(n-1)} + N^{(n-2)} - 1]!}{[N^{(n-1)} - 1]! N^{(n-2)}!}. \quad (25)$$

Наконец, определение вероятности состояния (5):

$$W^{(n)} = \frac{K^{(n)}}{L^{(n-1)}} \quad (26)$$

отражает тот факт, что $K^{(n)}$ есть подмножество множеств $L^{(n-1)}$ предыдущего слоя.

В предыдущем разделе обсуждалось различие энтропий S и H . В терминологии расслоенных пространств им соответствуют $H^{(1)}$ и $H^{(2)}$. Отмечалось снижение «чувствительности» к изменениям в вариационном ряду в энтропии второго слоя. Интуитивно ясно, что при многократном применении оператора вероятности распределения стремится к вырождению и, соответственно, энтропия должна уменьшаться. Это явление вырождения нетрудно доказать.

Из (26)

$$L^{(n-1)} = \sup K^{(n)},$$

следовательно, согласно (7)

$$\ln L^{(n-1)} = \sup H_a^{(n)}, \quad (27)$$

$K_i^{(n)}$ – число вариаций $(n-1)$ -го слоя, удовлетворяющих одной i – й вариации n – го слоя, очевидным образом связано с числами возможных сообщений:

$$L^{(n)} \cdot \sum_{i=1}^{L^{(n)}} K_i^{(n)} = L^{(n-1)}. \quad (28)$$

Логарифмируя (28) с учетом (27) и (7), имеем

$$\sup H_a^{(n+1)} + \ln \sum_{i=1}^{L^{(n)}} K_i^{(n)} = \sup H_a^{(n)}.$$

Следовательно,

$$0 \leq \sup H_a^{(n+1)} < \sup H_a^{(n)}.$$

Асимптотическое вырождение доказано.

6. Статистический ансамбль

Под статистическим ансамблем в математическом смысле понимается полная группа событий, а в физическом – некоторая счетная совокупность объектов, параметры которых в принципе возможно измерить.

Членами ансамбля, как мы увидели, могут пространственные или временные кванты (дискреты) вариационных рядов. Измеряемыми параметрами, подлежащими статистическому исследованию, при этом могут быть компоненты электромагнитного поля, скорости течения и т.п. Несмотря на то, что члены ансамбля в нашем случае макроскопичны, такой ансамбль формально неотличим от ансамбля молекул Максвелла – Больцмана. Роль «частицы» играет (искусственный) квант пространства или времени, роль ячейки фазового пространства – квант напряженности, скорости и т.п.

В случае статистики Бозе – Эйнштейна «частицей» может быть квант уровня энергии или подобной положительно определенной величины, ячейка фазового пространства тождественна пространственному или временному кванту.

Ансамбль Гиббса ниже применяется при установлении термодинамического соответствия.

Эргодическая гипотеза имеет существенное значение только в смысле физической тождественности результата, к которому приводит второе начало термодинамики ($H^{(1)} \rightarrow \max$, $H^{(1)}$ определена по пространственному ансамблю) и 7-я теорема Шеннона ($H^{(2)} \rightarrow \min$, $H^{(2)}$ определена в теории информации обычно по временному ансамблю).

В зависимости от числа включаемых в рассмотрение пространственных координат (x, y, z, t) , $P(X_i)$ в (8) является скалярной функцией от 1 до 4 координат. В соответствии с (8) размерность пространства определения энтропии меньше минимум на единицу. Иначе говоря, размерности D связаны соотношением

$$D(H) + D(X) \leq 4.$$

Это, в частности, означает, что по временным ансамблям можно построить трехмерное энтропийное поле, а для изучения временной эволюции энтропии системы необходим, по крайней мере, одномерный пространственный ансамбль. Впрочем, возможен и иной подход [21], когда используются две временных масштаба – микромасштаб для определения энтропии и макромасштаб для изучения эволюции.

В практических оценках, учитывая реальные возможности натурального эксперимента, необходимая статистика может быть получена, в основном, по временным рядам. В соответствии с этим, для оценок энтропии практически обычно могут быть использованы только одномерные ансамбли.

7. Термодинамическое соответствие

Вовлечение в обсуждение конкретных естественнонаучных проблем подходов из казалось бы удаленных областей физики (второе начало термодинамики и т.д.) может вызвать определенную настороженность. Целью данного раздела является демонстрация правомерности использования понятий термодинамики в более общем случае, нежели это обычно делается.

В статистической физике важную роль играет каноническое распределение Гиббса. Членами ансамбля Гиббса могут быть как материальные частицы, так и ячейки (кванты) пространства, заполненные энергией любой природы. Это может быть, например, энергия электромагнитного поля. Каждая ячейка обменивается энергией со всей системой (термостатом), суммарная энергия которой велика по сравнению с любой X_j . Тогда распределение Гиббса, выводимое из статистических рассуждений таково:

$$P_i = \frac{\exp(-X_i / \theta)}{\left[\sum_{i=0}^N \exp(-X_i / \theta) \right]^f}, \quad (29)$$

где f – число степеней свободы [22]. Для трехкомпонентного электромагнитного поля $f=3$. В (22) также предполагается $K(X_i) \equiv 1$. Величина θ в (29) – статистическая температура. Она измеряется в энергетических единицах (или после деления на постоянную Больцмана – в кельвинах. Как известно, шкалы статистической, термодинамической (используемой в (1)) и кинетической температур совпадают. Если, как и ранее, рассматривать энергии X_i как безразмерные величины, то θ также безразмерна. Статистическая температура интересна тем, что является некоторой функцией средней энергии (например энергии электромагнитного поля). Иначе говоря, температура не является чисто молекулярным понятием, как это часто считается. Можно показать [22], что

$$\frac{1}{\theta} = f \ln \left(\frac{M + U - X_i - 2}{U - X_i} \right),$$

а так как $U \gg X_i$, то

$$\frac{1}{\theta} = f \ln \left(\frac{1+u}{u} \right). \quad (30)$$

Из (30) сразу видно, что статистическая температура является функцией средней энергии (любого происхождения). При этом

$$u \rightarrow 0 \Leftrightarrow \theta \rightarrow 0, u \rightarrow \infty \Leftrightarrow \theta \rightarrow \infty,$$

т.е. свойства этой функции совершенно естественны. (30) можно упростить, если средняя энергия велика по сравнению с порогом разрешения (искусственным квантом), $u \gg 1$. Разлагая в (30) логарифм по степеням $1/u$, имеем в первом приближении

$$\theta = \frac{u}{f}. \quad (31)$$

(31) можно назвать высокотемпературным приближением.

Важность распределения (29) заключается в многочисленных следствиях. Например, из него можно получить что энергия электромагнитного поля описывается распределением Бозе–Эйнштейна, при этом любая компонента – распределением Гаусса, а модуль – распределением Максвелла – Больцмана. Но все это верно при условиях при которых в статистике выводится (29): замкнутость и квазиравновесность системы, причем первое четко оговаривается, второе обычно завуалировано и проявляется в том, что в разложении функции энергии термостата в ряд Тейлора можно ограничиться первым членом.

Покажем теперь, что (29) можно вывести из теории энтропии, а затем выразить саму энтропию через макрохарактеристики и далее, в весьма общем виде, получить основное уравнение термодинамики. Таким образом, мы покажем, в главных чертах, переход с языка теории информации на язык термодинамики. При этом общность результата ясно покажет, что термодинамические соотношения применимы к широкому кругу естественнонаучных задач, совсем не обязательно связанных с тепловыми процессами.

Пусть система замкнута:

$$U = \sum_{i=1}^N m_i X_i = \text{const}$$

и имеет неизменный объем (точнее размер по каждой координате):

$$M = \sum_{i=1}^N m_i = \text{const}.$$

Если система находится вблизи равновесного состояния, то применяя вариационный метод Лагранжа имеем из (7):

$$\delta H_a = \delta \left(M \ln M - \sum_{i=1}^N m_i \ln m_i \right) = - \sum_{i=1}^N (\ln m_i + 1) \delta m_i = 0, \quad (32)$$

$$- a \delta M = - a \sum_{i=1}^N \delta m_i = 0, \quad (33)$$

$$-b\delta U = -b \sum_{i=1}^N X_i \delta m_i = 0. \quad (34)$$

Складывая (32)-(34) при условии $\ln m_i \gg 1$, получаем:

$$\sum_{i=1}^N (\ln m_i + a + bX_i) \delta m_i = 0.$$

Следовательно,

$$\ln m_i + a + bX_i = 0,$$

т.е.

$$m_i = \exp(-a - bX_i),$$

откуда

$$P_i = \frac{m_i}{M^f} = \frac{\exp(-a - bX_i)}{M^f}. \quad (35)$$

Можно заметить, что (35) является распределением Максвелла – Больцмана. Из условия нормировки

$$\sum_{i=1}^N P_i = 1$$

следует

$$M = \exp(-a) \sum_{i=1}^N \exp(-bX_i).$$

Поэтому (35) представимо в виде

$$P_i = \frac{\exp(-bX_i)}{\left(\sum_{i=1}^N \exp(-bX_i) \right)^f}, \quad (36)$$

(36) тождественно (29), при этом

$$b = 1/\theta. \quad (37)$$

Соответствие установлено.

Можно пойти по этому пути дальше. Положив

$$c = f \ln \sum_{i=1}^N \exp(-bX_i), \quad (38)$$

представим (36) в виде

$$P_i = \exp(-c - bX_i). \quad (39)$$

Подставляя (39) в (7) получаем выражения энтропии через макрохарактеристики:

$$H_a = c + bu, \quad (40)$$

где u – по прежнему средняя плотность энергии.

Согласно [24] c – функция Массье, b – характеристика способности к диффузии между частями системы, у которых b различна. Последнее имеет простейший термодинамический смысл согласно (37).

Пользуясь функцией Массье, покажем справедливость основного уравнения термодинамики в контексте любого поля. Из (38) следует, что

$$\frac{\partial c}{\partial b} = fu. \quad (41)$$

Введем градиент энергии \vec{g}_i (в смысле обобщенной функции):

$$\vec{g}_i = \nabla X_i. \quad (42)$$

Средний градиент:

$$\bar{\vec{g}} = \sum_{i=1}^N P_i \vec{g}_i. \quad (43)$$

Подставляя в (43) выражения (39) и (42), получаем

$$\bar{\vec{g}} = \frac{e^{-c}}{b} \nabla \sum_{i=1}^N \exp(-bX_i).$$

Это выражение можно упростить с учетом (38):

$$\bar{\vec{g}} = \frac{1}{fb} \nabla c, \quad (44)$$

где c – функция b и пространственного положения \vec{r} . Итак,

$$dc = \frac{\partial c}{\partial b} db + \nabla c d\vec{r}. \quad (45)$$

Из (40), (45), (41) и (44) получаем

$$dH_a = bd\bar{X} + b\bar{\vec{g}}d\vec{r}$$

или

$$\theta dH_a = du + \bar{\vec{g}}d\vec{r}. \quad (46)$$

(46) – основное уравнение термодинамики (соотношение Гиббса), полученное нами любой по природе вариации со средней по пространству плотностью энергии u .

8. Заключение

Традиционно в курсах физики энтропия вводится в разделе термодинамики как интеграл Клаузиуса, затем вводится формула Больцмана и, наконец, упоминается о связи энтропии и информации. Современный подход, которому следовали мы – обратный. При этом в результате мы не только получаем известные термодинамические соотношения, но и показываем их универсальность для всех естественных наук, где предметом изучения является временная или пространственная изменчивость любых измеряемых величин. Для любой вариации (процесса) может быть вычислена энтропия как по первому слою (S), так и второму (H), а при необходимости и последующих. Для двух и более процессов могут быть вычислены условные энтропии и информация. В частности, это могут быть не только действительно разные процессы, но и разные отрезки реализации

одного процесса. Правда, для естественных процессов (которые нельзя «включить» или «выключить», или более формально – которые не могут быть заданы как существующие в единственном кванте временной или пространственной оси) условные энтропии и информацию можно рассчитывать только по второму слою (и далее), например, $H(X/Y)$, но не $S(X/Y)$ [25]. Вычислительные алгоритмы реализуются достаточно просто. Единственным условием является достаточно богатая экспериментальная статистика. Из практики расчетов следует, что для приемлемой оценки S вариационный ряд должен содержать, грубо говоря, не менее 100 отсчетов, для H – 1000 отсчетов (имеются в виду условные энтропии, для безусловных можно иметь отсчетов на порядок меньше).

За исключением упоминания энтропии фон Неймана (4), мы оставались всецело в рамках классической информации. Теория квантовой информации чрезвычайно интенсивно развивается в последнее десятилетие в связи с проблемами квантовой нелокальности. Квантовая нелокальность – удивительное явление, которое можно кратко определить так: существуют системы, находящиеся в чистом состоянии, в то время, как их подсистемы – в смешанном. Не менее удивительны и свойства квантовой информации. Так, условная энтропия здесь может быть отрицательна, а информация – больше безусловной энтропии, именно:

$$0 \leq I_{xy} \leq 2 \min (H(X), H(Y)).$$

Отсюда возникают совершенно необычные неклассические связи между процессами. Мы ограничимся лишь упоминанием этой области. Наиболее последовательное изменение темы можно найти в [26].

Но и в классических рамках многие вопросы развития энтропийного подхода мы не могли здесь затронуть, например, формальное определение причинности и практический причинный анализ [25, 27 – 30].

Остается надеяться, что читатели не только увидят возможности применения энтропийного подхода в своей области, но и разовьют его далее.

Список литературы

1. Шамбадаль П. Развитие и приложения понятия энтропии . М.: Наука, 1967 -290.
2. Шеннон К.Э. Работы по теории информации и кибернетике. М.: Ил., 1963 – 829с.
3. Мартин Н., Ингленд Дж. Математическая теория энтропий. М.: Мир. 1988 –251с.

4. Колмогоров А.Н. Три подхода к определению понятия «количество информации» // Проблемы передачи информации. 1965. Т.1. №1. С.3-11.
5. Колмогоров А.Н. К логическим основам теории информации и теории вероятностей // Проблемы передачи информации и теории вероятностей. 1969. Т.5.№3.С.3-7.
6. Тростников В.Н. Человек и информация. М.: Наука, 1970 – 188 с.
7. Яглом А.М., Яглом И.М. Вероятность и информация. М.: Наука, 1973 – 512 с.
8. Брюллюэн Л. Наука и теория информации. М.: Физматгиз, 1960 – 392с.
9. Коган И.М. Прикладная теория информации. М.: Радио и связь. 1981 – 216с.
10. Рейхенбах Г. Направление времени. М.: ИЛ, 1962 -316 с.
11. Хартли Р. Передача информации// Теория информации и ее приложения. М.: Физматгиз. 1959. – С. 5-35.
12. Темников Ф.Е., Афонин В.А., Дмитриев В.И. Теоретические основы информационной техники. М.: Энергия. 1971 – 424 с.
13. Поплавский Р.П. Дефон Максвелла и соотношения между информацией и энтропией // УФН. 1979. Т. 128. Вып. 1. С.165-176.
14. Partovi M.H. Entropic Formulation of uncertainty for quantum measurements // Phys. Rev. Lett. 1983. V. 50. N24. P. 1883-1885.
15. Sharma C.S. The entropy of an ensemble of quantum mechanical systems // Phys. Rev. Lett. 1986. V. 114A. N7. P. 344-352.
16. Николис Дж. Динамика иерархических систем. М.: Мир.1989 – 488с.
17. Рудаков Е.С. Теорема об энтропии взаимодействия как дополнение к третьему закону термодинамики // ДАН. 1991. Т.318. №4. С. 897- 900.
18. Хармут Х. Применение методов теории информации в физике. М.: Мир. 1989 – 334 с.
19. Абезгауз Г.Г., Тронь А.П., Копейкин Ю.М., Коровина И.А. Справочник по вероятностным расчетам. М.: Изд-во МО СССР. 1970 – 536с.
20. Бернштейн Г., Филлипс Э. Расслоения и квантовая теория // УФН. 1982. Т. 136. Вып. 4. С.665-692.
21. Пригожин И. От существующего к возникающему. М.: Наука. 1985 – 328 с.
22. Рейф Ф. Статистическая физика. М.: Наука. 1986 – 336 с.
23. Коротяев С.М. Интерпретация распределений вероятностей напряженности электромагнитного поля в океане. // Геомагнетизм и аэрономия. 1995. Т. 35. №4. С. 119-127.

24. Физика океана //Ред. *Доронин Ю.П.* Л.: Гидрометеиздат. 1978 – 296 с.
25. *Коротаев С.М.* Роль различных определений энтропии в причинном анализе геофизических процессов и их приложении к электромагнитной индукции в морских течениях // *Геомагнетизм и аэрономия.* 1995. Т. 35. №3. С. 116-125.
26. *Cerf N.J.* Entropic bounds on coding for noisy quantum channels // *Phys. Rev. A.* 1998. V. 57. N5. P.3330-3347.
27. *Коротаев С.М., Шабелянский С.В., Сердюк В.О.* Обобщенный причинный анализ и его применение для изучения электромагнитного поля в море // *Изв. АН Физика Земли.* 1992. №6. С. 77-86.
28. *Коротаев С.М., Хачай О.А., Шабелянский С.В.* Причинный анализ процесса горизонтальной информационной диффузии электромагнитного поля в океане // *Геомагнетизм и аэрономия.* 1993. Т. 33. №2. С. 128-133.
29. *Korotaev S.M.* Formal definition of causality and Kozyrev's axioms // *Galilean Electrodynamics.* 1993. V.4. N5. P. 86-89.
30. *Арушанов М.Л., Коротаев С.М.* Причинный анализ и его применение для изучения физических процессов в атмосфере // *Метеорология и гидрология.* 1994. №6. С. 15-22.