Энтропия в термодинамике и в теории информации

В.А. Зорич

11 апреля 2022 г.

1. Введение.

Существует легенда, согласно которой Ше́ннон обратился к фон Нейману за советом, каким термином назвать то, что теперь называется информационной энтропией или энтропией Шеннона. Человек широкого кругозора и глубины, фон Нейман прозорливо предложил термин *энтропия*.

Этот термин появился в класссической равновесной термодинамике. Он был введён в 1865 году Клаузиусом в связи с, так называемым, вторым началом термодинамики, началом Карно — Клаузиуса. Функция энтропии стала одной из ключевых характеристик равновесного состояния термодинамической системы.

Огромный шаг к пониманию физического смысла энтропии был сделан Больцманом, одним из творцов статистической термодинамики.

Основные функции феноменологической термодинамики (температура, давление ...), на самом деле, являются функциями огромного числа переменных (молекул, атомов). Значения таких функций, именно по этой самой причине, оказываются почти постоянными и устойчивыми с точки зрения наблюдателя, измеряющего интегральные характеристики равновесного состояния системы, например, газа, а не детали микро состояния всей системы составляющих газ молекул.

Одному наблюдаемому макросостоянию системы может отвечать огромное множество её микросостояний, а равновесному макросостоянию системы вообще отвечает подавляющее число всех её возможных микросостояний.

Принцип Больцмана или *идея Больцмана* сопоставить классической энтропии S логарифм вероятности соответствующего состояния многочастичной термодинамической системы, в контексте известных свойств классической энтропии, диктуется уже следующим.

Вероятности независимых событий перемножаются, а логарифм даст аддитивность.

Тенденция переходить от менее вероятного состояния к более вероятному соответствует тенденции роста классической энтропии состояния при эволюции изолированной термодинамической системы к равновесию.

Максимум такой энтропии соответствует равновесному состоянию системы. Около него концентрируется почти всё множество микросостояний многочастичной системы (типа молекулярного газа), отвечающих наблюдаемому равновесному макросостоянию системы!

Вот *формула Больцмана* (выгравированная на его могиле)

$$S = k \log W.$$

Здесь k— размерная постоянная (отношение R/N_A газовой постоянной и числа Авогадро), названная Планком *постоянной Больцмана*, а W — целочисленная величина, пропорциональная вероятности рассматриваемого термодинамического состояния, которую, следуя Планку, называют *статистическим весом состояния.*

Конкретнее, W — это то максимально возможное количество микросостояний, которое отвечает рассматриваемому равновесному состоянию многочастичной термодинамической системы с энтропией S.

Поясним сказанное примером. Часть выкладок этого примера мы используем ниже для демонстрации прямой связи информационной энтропии Шеннона и термодинамической энтропии Больцмана.

2. Один пример.

Следуя Шрёдингеру [1], рассмотрим ансамбль из N одинаковых, но перенумерованных систем, каждая из которых может находиться в одном из пернумерованных состояний $1, 2, \ldots, l$. Пусть $\varepsilon_1 \leq \varepsilon_2 \leq \ldots \leq \varepsilon_l$ — значения энергии индивидуальной системы в этих состояниях и $a_1, a_2, \ldots, a_l$ — количество систем ансамбля, находящихся в состояниях $1, 2, \ldots, l$ соответственно.

Такой набор $a_1, a_2, \ldots, a_l$ может реализоваться многими способами. А точнее, число способов равно $\binom{N}{a_1}\binom{N-a_1}{a_2}\cdot\ldots\cdot\binom{N-a_1-\ldots-a_{l-1}}{a_l}$, т.е.

$$G = \frac{N!}{a_1!a_2!\ldots a_l!} \ .$$

Совокупность чисел $a_1, a_2, \ldots, a_l$ должна удовлетворять условиям

$$\sum_i a_i = N \ , \qquad \sum_i \varepsilon_i a_i = E \ ,$$

где E — полная совокупная энергия систем ансамбля.

Будем искать максимальное значение $W = \max G$ (или $\ln W = \max \ln G$) при указанных ограничениях, что даст нам наиболее вероятный набор $a_1, a_2, \ldots, a_l$ чисел заполнения.

(Напомним, что в интересных для термодинамики случаях, когда число N и количество возможных уровней энергии ε_i очень велики, имеет место *явление концентрации*. Можно показать, что общее число всех возможных при наших условиях состояний ансамбля почти совпадает с максимальным значением G, которое мы намерены искать. Значит, мы действительно найдем и наиболее вероятный набор $a_1, a_2, \ldots, a_l$ чисел заполнения.)

При больших значениях n по формуле Стирлинга $n! \simeq \sqrt{2\pi n}\left(\frac{n}{e}\right)^n$. Поэтому можно считать что $\ln(n!) \approx n(\ln n - 1) \approx n \ln n$. На самом деле нам потом будет достаточно этого замечания и приведённого выше выражения для статистического веса G, но доведём этот красивый и полезный пример до конца.

а. Действуя методом Лагранжа отыскания условного экстремума, записав, что

$$\sum_i \ln a_i da_i + \lambda \sum_i da_i + \nu \sum_i \varepsilon_i da_i = 0 \ ,$$

находим, что при любом i выполняется равенство

$$\ln a_i + \lambda + \nu\varepsilon_i = 0 \quad \text{и, значит,} \quad a_i = e^{-\lambda-\nu\varepsilon_i} \ ,$$

причем λ и ν подчинены условиям

$$\sum_i e^{-\lambda-\nu\varepsilon_i} = N \ , \qquad \sum_i \varepsilon_i e^{-\lambda-\nu\varepsilon_i} = E \ .$$

b. Обозначая через $E/N = U$ среднюю энергию, приходящуюся на одну систему ансамбля, запишием полученный результат в следующем виде:

$$\frac{E}{N} = U = \frac{\sum_i \varepsilon_i e^{-\nu\varepsilon_i}}{\sum_i e^{-\nu\varepsilon_i}} = -\frac{\partial}{\partial\nu} \ln \sum_i e^{-\nu\varepsilon_i} \ ,$$

$$a_i = N \frac{e^{-\nu\varepsilon_i}}{\sum_i e^{-\nu\varepsilon_i}} = -\frac{N}{\nu}\frac{\partial}{\partial\varepsilon_i} \ln \sum_i e^{-\nu\varepsilon_i} \ .$$

Дополнительные рассмотрения, выясняющие физический смысл величины ν в термодинамической ситуации, приводят к тому, что

$$\nu = \frac{1}{kT} \ ,$$

где k — постоянная Больцмана, а T — абсолютная температура.

c. Возникает, как мы уже понимаем, важная величина, называемая *статистической суммой*:

$$Z = \sum_i e^{-\frac{\varepsilon_i}{kT}} \ .$$

Теперь можно написать, как именно числа заполнения a_i распределены по энергетическим уровням при данной абсолютной температуре:

$$a_i = N \frac{e^{-\frac{\varepsilon_i}{kT}}}{Z} \ .$$

Величина

$$Z^{-1} e^{-\frac{\varepsilon_i}{kT}}$$

это вероятность малой системе оказаться в состоянии с энергией ε_i.

Мы получили *каноническое распределение Гиббса*

$$\exp\left(\frac{\psi - \varepsilon_i}{kT}\right).$$

В рассмотренной ситуации оно показывает, как распределены по энергиям ε_i малые системы (типа молекул) в термостате (большой системе, типа газа, находящейся в равновесном термодинамическом состоянии при температуре T). Здесь введено обозначение $e^{\frac{\psi}{kT}} = Z^{-1}$.

Если найден статистический вес $W = \max G$ рассматриваемого равновесного состояния термодинамической системы, то, следуя формуле Больцмана $S = k \ln W$, можно найти энтропию S системы в этом состоянии

3. Информационная энтропия.

a. *Информация и её количественное описание.*

Если случайная величина может с равной вероятностью принимать одно из двух возможных значений, например, 0 или 1, то указание конкретного состояния такой величины есть информация, которую принято называть *битом* информации. Мы напомнили определение бита информации (и неопределённости).

Вектором, длиной n, из нулей и единиц, можно задать 2^n различных объектов.

Если имеется M равновозможных значений случайной величины, или M равновероятных событий, то указание одного такого события требует $\log_2 M$ бит информации.

Если вероятность события p, то оно встречается один раз среди $M = \frac{1}{p}$ событий. Значит, его неопределённость, измеренная в битах, есть $\log_2 M = \log_2 \frac{1}{p} = -\log_2 p$.

b. *Энтропия случайной величимы.*

Функцию со случайными значениями в математике принято называть *случайной величиной.* Пусть X случайная величина. Средняя на одно значение случайной величины X (на одно событие) неопределённость в битах — это математическое ожидание $EX = -\sum_{i=1}^{M} p_i \log_2 p_i$ случайной величины X. Это математическое ожидание, вслед за Шенноном [2], называют *энтропией случайной величины* X и обозначают символом $H(X) = -\sum_{i=1}^{M} p_i \log_2 p_i$. (Греч. $\eta\nu\tau\rho\o\pi\iota\alpha$, от др.-греч. $\eta\nu$ — в + $\tau\rho\o\pi\eta$ — превращение; обращение.) Как уже было сказано, термин *энтропия* исходно появился в термодинамике.

Если единичное событие системы X имело вероятность появления p, то его осуществление несёт $-\log_2 p$ бит информации. Но длительное наблюдение за появлениями такого события в единицу времени приносит всего $-p \log_2 p$ бит информации. Это же относится и к среднему числу бит информации $H(X) = -\sum_{i=1}^{M} p_i \log_2 p_i$, которое за единицу времени приносит появление одного из событий системы X (значений случайной величины X). В этом смысл информационной энтропии.

Укажем сразу, в чём состоит статистический характер энтропии $H(X)$. Мы этим воспользуемся ниже.

Пусть X — произвольная дискретная случайная величина, которая может принимать M различных значений x_i с вероятностями p_i соответственно. Тогда для любых положительных чисел ε, δ найдется такое число $n_{\varepsilon\delta}$, что при $n \geq n_{\varepsilon\delta}$ выполняется неравенство

$$\Pr\ \{| - \frac{1}{n} \sum_{i=1}^{n} \log_2 p_{x_i} - H(X)| < \delta\} > 1 - \varepsilon, \qquad (1)$$

где, как обычно, $\Pr\ \{\}$ — вероятность указанного в скобках события, но теперь $x_i,\ i = 1, ..., n$ — n независимых значений случайной величины X, а p_{x_i} — вероятности этих значений.

Как связана энтропия с кодированием?

Рассмотрим сообщения-слова-векторы $\bar{x} = (x_1, ..., x_n)$, образованные n последовательными независимыми значениями случайной величины X. Вероятность $p_{\bar{x}}$ появления слова $\bar{x}$ равна $p_{\bar{x}} = p_{x_1} \cdot ... \cdot p_{x_n}$. В силу соотношения (1) при $n \geq n_{\varepsilon\delta}$ с вероятностью, большей чем $1 - \varepsilon$, будем иметь

$$2^{-n(H(X)+\delta)} \leq p_{\bar{x}} \leq 2^{-n(H(X)-\delta)}. \qquad (2)$$

Слово $\bar{x}$ называют *δ-типичным*, если для него выполнены эти оценки. Ясно, что существует не более $2^{n(H(X)+\delta)}$ таких δ-типичных слов, а если $n \geq n_{\varepsilon\delta}$, то их ещё и не меньше чем $(1 - \varepsilon)2^{n(H(X)-\delta)}$, и при этом всё множество не δ-типичных слов имеет вероятность, не большую чем ε.

В принципе теперь уже можно использовать двоичные последовательности длиной $n(H(X) + \delta)$, чтобы закодировать все δ-типичные слова. Даже, если все остальные слова закодировать одним символом, вероятность ошибки при передаче слов $\bar{x}$ длины n, вызванная таким кодом, будет меньше ε.

С другой стороны (и это эффект неустойчивости экономных кодов), любой код, использующий в той же ситуации двоичные последовательности относительно чуть меньшей длины $n(H(X) - \delta)$ (например, $2\delta n$ из $n(H(X)+\delta)$ посланных символов потерялось в шуме), будет иметь асимптотически неисчезающую вероятность ошибки, стремящуюся к единице при $n \to +\infty$.

Итак, связь энтропии и кодирования информации состоит, например, в том, что эффективное кодирование асимптотически при $n \to +\infty$ требует $N \sim 2^{nH(X)}$ слов и энтропия $H(X)$ может интерпретироваться как мера количества информации в битах на передаваемый символ, т. е. на одно значение случайной величины X.

Отсюда, в частности, следует, что энтропия источника информации не должна превышать пропускную способность канала связи, если мы хотим адекватно и без задержек передавать поступающую информацию по этому каналу связи.

Если передаётся M сообщений и на передачу каждого затрачивается время T секунд, то это равносильно тому что скорость передачи в битах в секунду равна $\frac{1}{T}\log_2 M$.

Теперь посмотрим на соотношение (1) с точки зрения макро и микро состояний. Пусть $n \gg 1$. Величина $-\frac{1}{n}\sum_{i=1}^{n}\log_2 p_{x_i}$ для подавляющего числа векторов $\bar{x} = (x_1, ..., x_n)$ почти одна и та же, $H(X)$. Конкретному состоянию (значению, макросостоянию) величины $-\frac{1}{n}\sum_{i=1}^{n}\log_2 p_{x_i}$ отвечает много микросостояний $\bar{x} = (x_1, ..., x_n)$.

И не просто много, а, когда $n \gg 1$, почти каждое случайно взятое микросостояние $\bar{x} = (x_1, ..., x_n)$ реализует почти одно и то же состояние (значение) $H(X)$ наблюдаемой макро величины.

На языке термодинамики и в терминах энтропии Больцмана, это означает, что система, допускающая огромное множество микро состояний $\{\bar{x} = (x_1, ..., x_n)\}$, находится в равновесном макро состоянии.

3. Энтропия в термодинамике и в информатике.

Теперь можно конкретизировать сказанное и сопоставить энтропию равновесного состояния термодинамической системы и энтропию случайной величины.

Мы увидим, что энтропия Шеннона, с точки зрения математики, есть просто отнесённая к одной частице (удельная) энтропия Больцмана.

Что в случае случайной величины X должно отвечать равновесному состоянию, к которому со временем эволюционирует любая изолированная термодинамическая система? По-видимому, просто надо достаточно долго наблюдать значения, которые принимает случайная величина X. Длинный вектор таких значений стабилизируется в следующем смысле. В нём каждое значение x_i случайной величины X будет присутствовать в количестве a_i, пропорциональном вероятности p_i значения x_i случайной величины X и длине N вектора значений (времени наблюдения): $a_i \simeq p_i N$.

Сами векторы $\bar{x} = (x_1, ..., x_N)$ большой длины N, скорее всего, будут различными, но подавляющая часть всех таких векторов будет иметь указанную статистику $a_i \simeq p_i N$ присутствия каждого отдельного значения x_i случайной величины X, если таких значений конечное число.

Количество таких векторов находится по уже известной нам формуле:

$$G = \frac{N!}{a_1!a_2!\dots a_i!\dots}\ .$$

Берём $\log G$. Это, с точностью до размерной постоянной (постоянной Больцмана), энтропия $\log W$ равновесного состояния по Больцману. Воспользовавшись формулой Стирлинга и уже проделанными выше выкладками, найдём с учётом соотношений $a_i \simeq p_i N$:

$$\log G \simeq N(\log N - 1) - \sum_i a_i(\log a_i - 1) =$$

$$N \log N - \sum_i a_i \log a_i = N \log N - \sum_i p_i N \log(p_i N) =$$

$$N \log N - \sum_i p_i N \log p_i - \log N \sum_i p_i N =$$

$$N \log N - \sum_i p_i N \log p_i - N \log N \sum_i p_i =$$

$$N \log N - \sum_i p_i N \log p_i - N \log N = -N \sum_i p_i \log p_i.$$

Приведя здесь безразмерную энтропию Больцмана к единичному элементу (в данном случае к одному значению случайной величины, поделив на N), получаем энтропию Шеннона.

Итак, энтропия Шеннона, с точки зрения математики, есть просто отнесённая к одной частице (удельная) энтропия Больцмана.

Заметим, что изменение основания b логарифма в определении Шеннона энтропии $H(X) = -\sum_i p_i \log_b p_i$ случайной величины X, приводит лишь к появлению общего множителя — коэффициента, что равносильно переходу к соответствующему масштабу единицы измерения.

4. Небольшой комментарий.

Выражения $-\sum_i p_i \log p_i$ и $-\int p \log p$ в связи с энтропией в термодинамике появились, конечно, задолго до того, как они появились в теории информации. Сама теория информации как наука значительно моложе термодинамики.

На рубеже XIX-XX веков Гиббс [3] обогатил статистическую термодинамику замечательной общей математической моделью гамильтоновой динамической системы, наделённой инвариантной вероятностной мерой, отвечающей равновесному состоянию многочастичной термодинамической системы.

Равновесному состоянию изолированной термодинамической системы, отвечает экстремум энтропии. Если вероятностная мера в фазовом пространстве распределена с плотностью p, то поиск плотности p, отвечающей равновесию системы, приводит, например, к задаче поиска экстремума взятого по всему пространству интеграла $\int p \log p$ при двух естественных условиях: интеграл от p по всему пространству равен единице, и полная энергия E всей гамильтоновой системы постоянна. Дискретный вариант этой вариационной задачи мы рассмотрели выше и пришли к дискретному варианту канонического распределения Гиббса. Истоки такого распределения восходят к Максвеллу и Больцману.

5. Демон Максвелла и энтропия.

Мы здесь всего лишь привели математическую выкладку, демонстрирующую связь энтропии Больцмана статистической термодинамики и энтропии Шеннона информатики. Мы почти не касались сути понятия *энтропия*. Хотя бы вскользь, всё же, надо сказать, что физики довольно тщательно анализировали, обсуждали и продолжают обсуждать различные аспекты понятия энтропии. Не ушёл от их внимания и информационный аспект термодинамической энтропии.

Ещё в 1867 году Максвелл предложил свой знаменитый мысленный эксперимент, парадокс Максвелла, получивший название «Демон Максвелла». (Публикация 1871 года.) Этот мысленный эксперимент, описанный почти во всех учебниках физики, заставил физиков глубже разобраться в двух фундаментальных принципах термодинамики.

Сначала причину парадокса искали в нарушении принципа сохранения энергии (первого начала термодинамики). Но постепенно стало ясно, что дело связано со вторым началом термодинамики и проистекающим из него принципом неубывания энтропии. Когда в 1929 году, анализируя парадокс Максвелла, Лео Силард (тот самый, кто стимулировал Эйнштейна подписать письмо Рузвелту, инициировавшее разработку атомной бомбы в США) мысленно построил свой двигатель, работавший на газе из одной молекулы, стало понятно, что дело в энтропии и, более того, в спрятанной в ней информации. Считается, что к 1984 году парадокс Максвелла был разрешён. Решение пришло как раз через информационный аспект термодинамической энтропии (см. [4], [5]).

Поясним таким сравнением. Если Архимед говорил, "дайте мне точку опоры и я подниму Землю", то, например, Беннет и Фейнман могли бы сказать, "дайте мне бесконечную память (для хранения информации) и я построю вам и Демона Максвелла и вечный двигатель второго рода, нарушающий второе начало термодинамики". Ленточный двигатель Беннета, работающий, как машина Тьюринга, и, благодаря информации, без потерь перерабатывающий тепло в работу, представлен на странице 146 книги [5].

Статистическая физика, объяснила бóльшую часть феноменологической термодинамики, исходя из более фундаментальных законов, описывающих поведение многочастичных систем (например, молекулярного газа). В частности, как уже было сказано, Больцман дал статистическое описание понятия термодинамической энтропии S.

Позднее Бриллюэн (см., например, [6]) уже прямо связал термодинамическую энтропию с информационной энтропией следующим соотношением: $\Delta S \geq \Delta I$. Здесь ΔI — информация[1], полученная в результате измерения, проведённого посредством какого-то прибора, а ΔS — рост энтропии термодинамического состояния прибора в результате этого измерения. Если $T\Delta S$ через тепло характеризует соответствующие затраты энергии, то соотношение $T\Delta S \geq T\Delta I$ показывает, что чем тоньше (точнее) измерение (т. е. чем больше полученная в результате измерения информация ΔI), тем бóльших энергетических затрат оно требует. Это уже некоторая физическая предпосылка к фундаментальному принципу неопределённости, подробно рассматриваемому в квантовой механике.

И последнее.

Классическая термодинамика трудами Пуанкаре, Каратеодори, Борна приобрела вид достаточно последовательной в математическом отношении теории. Трудами Максвелла, Больцмана, Гиббса, Планка, Эйнштейна были созданы основы статистической термодинамики. Аксиома-

[1]В единицах термодинамической энтропии, т. е. I получается из информационной энтропии H умножением на постоянную Больцмана: $I = kH$.

тизация термодинамики продолжается и сейчас. (Сравните [7] и, например, работу [8], посвящённую формализации второго начала термодинамики и энтропии).

Возвращаясь к информационной энтропии, добавим, что, Шеннон уже в своей фундаментальной для теории информации работе [2] предложил набор из трёх естественных свойств — аксиом информационной энтропии, которые с неизбежностью приводят к определяющей её формуле $H(X) = -\sum_{i=1}^{M} p_i \log p_i$.

Эпилог

Открывая сборник статей о турбулентности [9], академик О.М.Белоцерковский вспоминает, что, когда он учился на физическом факультете Московского университета, лекции по электричеству им читал профессор С.Г.Калашников, который на первой же лекции рассказал следующее. Однажды на экзамене он (Калашников) спросил студента: «Что такое электричество?» Студент заёрзал, засуетился и говорит: «Вот ещё вчера знал, а забыл». На это Калашников сокрушённо заметил: «Был один человек, который знал, и тот забыл!»

Ситуация с энтропией примерно такая же.

Некоторые источники

[1] *Шрёдингер Э.* Лекции по физике. — М., Ижевск: Изд. РХД, 2001.

[2] *Shannon Claude E.* A Mathematical Theory of Communication. The Bell System Technical Journal, Vol. 27, pp. 379–423, 623–656, July, October, 1948.

Шеннон К. Работы по теории информации и кибернетике. — М.: Изд. иностр. лит., 2002

[3] *Гиббс Дж. В.* Термодинамика. Статистическая механика. — М.: Наука, 1982.

[4] *Беннет Чарлз Г.* Демоны, двигатели и второе начало термодинамики. В мире науки, 1988, 1, 52-60.

[5] *Feynman R. P.*, Feynman Lectures on Computation. Addison-Wesley Publishing Company, Inc. The Advanced Book Program, (Lectured 1983), Published 1996.

[6] *Бриллюэн Л.* Теория информации и её приложение к фундаментальным проблемам физики. «Развитие современной физики». Сб. статей. — М.: Наука, 1964. С. 324–329.

[7] *Борн М., Каратеодори К., Бриллюэн Л. / Кузнецов Б.Г. (под ред.)*, Тамм И.Е., Гинзбург В.Л. и др. Развитие современной физики. Сб. статей. — М.: Наука, 1964.

[8] *Lieb E. H., Yngvason J.* The Mathematics of the Second Law of Thermodynamics. In: Alon N., Bourgain J., Connes A., Gromov M., Milman V. (eds) Visions in Mathematics. Towards 2000. Modern Birkhauser Classics. Birkhauser Basel (2010), pp 334-358.

[9] Этюды о турбулентности. — М.: Наука, 1994.