



# ЭНТРОПИЯ

Авторы: Ю. В. Прохоров

ЭНТРОПИЯ в теории информации, мера неопределённости к.-л. опыта (испытания), который в зависимости от случая может заканчиваться разл. исходами. При этом предполагают, что имеются определённые вероятности появления того или иного исхода. Пусть  $x_1, x_2, \dots, x_n$  – разл. исходы опыта,  $p_1, p_2, \dots, p_n$  – соответствующие вероятности  $p_j \geq 0, \sum_{j=1}^n p_j = 1$ . Тогда Э. Н определяется выражением  $H = H(p_1, p_2, \dots, p_n) = -\sum_{j=1}^n p_j \log_2(1/p_j)$  (считается, что  $0 \log 0 = 0$ ).

Свойства Э.: Э. равна нулю в том случае, когда одно число из  $p_j$  равно единице, а остальные равны нулю, т. е. когда исход опыта достоверен; Э. достигает макс. значения при данном  $n$ , когда все исходы равновероятны; Э. объединения двух независимых опытов равна сумме их Э. Функция  $H$  от  $p_j$  является единственной, удовлетворяющей этим и ещё нескольким, столь же естественным требованиям. Однако ценность понятия Э. определяется не этим обстоятельством, а тем, что она играет важную роль в [информации теории](#).

Для теории информации особый интерес представляет случай, когда  $x_j$  суть сообщения некоторого источника [информации](#), передаваемыми по [каналу связи](#). Сообщения при этом рассматривают как временные последовательности элементов (букв), выбираемых с некоторыми вероятностями из какой-то определённой совокупности (алфавита). Выводы теории информации касаются сообщений, являющихся «достаточно длинными» (в принципе неограниченно длинными) последовательностями букв, что соответствует предположению о весьма длительной работе источников сообщений и каналов связи. Поэтому Э. источника на символ (или скорость передачи сообщений, измеряемая в двоичных единицах на символ) определяется некоторым предельным переходом. С этой целью, наряду с сообщениями, представленными в виде неограниченных последовательностей  $a_1, a_2, \dots, a_N, \dots$  букв некоторого  $s$ -буквенного алфавита, рассматривают «урезанные» сообщения длины  $N$ , т. е. цепочки  $a_1, a_2, \dots, a_N$ . Выбирая в определении Э. в качестве  $x_j$  эти  $N$ -членные цепочки и в качестве  $p_j$  – соответствующие вероятности, получают некоторую величину  $H_N$ . Отношение  $H_N/N$  даёт Э. на букву для  $N$ -членных цепочек. В теории информации устанавливается, что при очень широком допущении устойчивости вероятностных закономерностей во времени (стационарность источника) величина  $H_N/N$ , убывая, стремится к пределу  $H_\infty = \lim_{N \rightarrow \infty} H_N/N$ , называемому Э. сообщения на символ. Если все символы имеют некоторую длительность и  $t$  – их средняя длительность, то отношение  $H_\infty/t$  даёт Э. источника на единицу времени. Эти две величины  $H_\infty$  и  $H_\infty/t$  являются основными информационными характеристиками источника сообщений. Так,  $H_\infty$  позволяет оценить максимально возможную степень «сжатия» сообщения при использовании того же алфавита (см. [Избыточность сообщений](#), [Кодирование](#)). Соотношение между скоростью создания сообщений  $H_\infty/t$  и ёмкостью к.-л. канала с тем же входным алфавитом, что использован при записи сообщений, определяет возможность «почти безошибочной» передачи этих сообщений по каналу (см. [Шеннона теорема](#)).

Э. испытания с бесконечным числом исходов можно попытаться определить с помощью предельного перехода. Но этот путь приводит, как правило, к бесконечному значению Э. Поэтому задаются определённым уровнем

точности  $\epsilon$  и определяют т. н.  $\epsilon$ -энтропию как описываемого с точностью до  $\epsilon$  исхода опыта.

## Литература

Лит. см. при ст. [Информация](#).

Processing math: 0%