В 1948 г. Клод Шеннон в своих работах по теории связи выписывает формулы для вычисления количества *информация* и энтропии.

Термин *энтропия* используется Шенноном по совету патриарха компьютерной эры фон Неймана, отметившего, что полученные Шенноном для теории связи формулы для ее расчета совпали с соответствующими формулами статистической физики, а также то, что "точно никто не знает" что же такое *энтропия*.

Энтропия

Энтропия как мера неопределенности

То, что событие случайно, означает отсутствие полной уверенности в его наступлении, что, в свою очередь, создает ***неопределенность*** в исходах опытов, связанных с данным событием. Безусловно, ***степень неопределенности*** различна для разных ситуаций.

Для практики важно иметь возможность произвести ***численную оценку неопределенности разных опытов***. Попробуем ввести такую количественную меру неопределенности

Рассмотри опыт с ***n*** **равновероятных** исходов. Очевидно, что неопределенность каждого из них зависит от n, т.е. *мера неопределенности является функцией числа исходов f(n).*

Можно указать некоторые свойства этой функции:

1. $f(1) = 0$, поскольку при $n = 1$ исход опыта не является случайным и, следовательно, неопределенность отсутствует;
2. $f(n)$ возрастает с ростом n, поскольку чем больше число возможных исходов, тем более затруднительным становится предсказание результата опыта.

Для определения явного вида функции ***f(n)*** рассмотрим два *независимых* опыта α и β с количествами равновероятных исходов, соответственно n_α **и** n_β.

Пусть имеет место сложный опыт, который состоит в одновременном выполнении опытов **α** и **β**; число возможных его исходов равно $n_\alpha \cdot n_\beta$, причем, все они равновероятны.

Очевидно, неопределенность исхода такого сложного опыта **α ^ β** будет больше неопределенности опыта α, поскольку к ней добавляется неопределенность β;

мера неопределенности сложного опыта равна $f(n_\alpha \cdot n_\beta)$. С другой стороны, меры неопределенности отдельных α и β составляют, соответственно, $f(n_\alpha)$ и $f(n_\beta)$.

В первом случае (сложный опыт) проявляется общая (суммарная) неопределенность совместных событий, во втором - неопределенность каждого из событий в отдельности. Однако из независимости α и β следует, что в сложном опыте они никак не могут повлиять друг на друга и, в частности, α не может оказать воздействия на неопределенность β, и наоборот.

Следовательно, мера суммарной неопределенности должна быть равна сумме мер неопределенности каждого из опытов, т.е. мера неопределенности аддитивна:

$f(n_\alpha \cdot n_\beta). = f(n_\alpha) + \mathrm{f}(n_\beta)$

За меру неопределенности опыта с *n* равновероятными исходами можно принять число *log(n)*.

То есть

$$f(n) = \log(n) \qquad (1)$$

Следует заметить, что выбор основания логарифма в данном случае значения не имеет, поскольку в силу известной формулы преобразования логарифма от одного основания к другому ($\log_b n=\log_b a*\log_a n$)

*Единица измерения неопределенности при двух возможных равновероятных исходах опыта называется **бит.***

Эта величина получила название ***энтропия***. В дальнейшем будем обозначать ее ***H*.**

Вновь рассмотрим опыт с ***n*** равновероятными исходами. Поскольку каждый исход случаен, он вносит свой вклад в неопределенность всего опыта, но так как все ***n*** исходов равнозначны, разумно допустить, что и их неопределённости одинаковы.

Из свойства аддитивности неопределенности, а также того, что согласно (1) общая неопределенность равна $log_2 n$, следует, что неопределенность, вносимая одним исходом составляет:

$$\frac{1}{n} * \log_2 n = -\frac{1}{n} * \log_2 n = -p * \log_2 p$$

Таким образом, неопределенность, вносимая каждым из равновероятных исходов, равна:

$$H = -p * \log_2 p \qquad (2)$$

Теперь попробуем обобщить формулу (2) на ситуацию, когда исходы опытов *неравновероятны,* например, $p(A_1)$ и $p(A_2)$. Тогда:

$$H_1 = -p(A_1) * \log_2 p(A_1) \text{ и } H_2 = -p(A_2) * \log_2 p(A_2)$$

$$H = H_1 + H_2 = -p(A_1) * \log_2 p(A_1) - p(A_2) * \log_2 p(A_2)$$

Обобщая это выражение на ситуацию, когда опыт α имеет n неравновероятных исходов $A_1, A_2, \dots, A_n$, получим :

$$H(\alpha) = -\sum_{i=1}^{n} p(A_i) * \log_2 p(A_i) \qquad (3)$$

Введенная таким образом величина называется *энтропией опыта* α.

Энтропия *является мерой неопределенности опыта, в котором проявляются случайные события, и равна средней неопределенности всех возможных его исходов.*

Для практики формула (3) важна тем, что позволяет сравнить неопределенности различных опытов со случайными исходами.

Свойства энтропии

1) Как следует из (3), $H = 0$ только в двух случаях:

   (a) какая-либо из $p(A_j) = 1$; однако, при этом из свойства суммы вероятностей следует, что все остальные $p(A_i) = 0$ $(i \neq j)$, т.е. реализуется ситуация, когда один из исходов является *достоверным* (и общий итог опыта перестает быть случайным);

   (b) все $p(A_i) = 0$, т.е. никакие из рассматриваемых исходов опыта невозможны, поскольку нетрудно показать, что $\lim\limits_{p \to 0} (\mathrm{p} \cdot \log \mathrm{p}) = 0$ во всех остальных случаях, очевидно, что $\mathrm{H} > 0$.

2) Для двух независимых опытов α и β $H(\alpha \wedge \beta) = \mathrm{H}(\alpha) + \mathrm{H}(\beta)$ (4)

Энтропия сложного опыта, состоящего из нескольких независимых, равна сумме энтропии отдельных опытов.

В справедливости (4) можно убедиться непосредственно: Пусть опыт α имеет n исходов $A_1, A_2, \ldots A_n$, которые реализуются с вероятностями $p(A_1), p(A_2), \ldots p(A_n)$, а событие β - m исходов $B_1, B_2, \ldots B_m$ с вероятностями $p(B_1), p(B_2), \ldots p(B_m)$. Сложный опыт α ^ β имеет $n \cdot m$ исходов типа $A_i B_j$ $(i = 1 \ldots n, j = 1 \ldots m)$. Следовательно:

$$H(\alpha \wedge \beta) = -\sum_{i=1}^{n} \sum_{j=1}^{m} p(A_i \wedge B_j) * \log_2 p(A_i \wedge B_j) \qquad (5)$$

Поскольку α и β - независимы, то независимыми окажутся события в любой паре $A_i \wedge B_j$. Тогда, согласно Формулы условной вероятности,

$$p(A_i \wedge B_j) = p(A_i) \cdot p(B_j) \quad \text{и} \quad \log_2 p(A_i \wedge B_j) = \log_2 p(A_i) + \log_2 p(B_j)$$

$$H(\alpha \wedge \beta) = -\sum_{i=1}^{n}\sum_{j=1}^{m} p(A_i)p(B_j) \cdot \{\log_2 p(A_i) + \log_2 p(B_j)\} =$$

$$= -\sum_{i=1}^{n}\sum_{j=1}^{m} p(A_i)p(B_j) \cdot \log_2 p(A_i) - \sum_{i=1}^{n}\sum_{j=1}^{m} p(A_i)p(B_j) \cdot \log_2 p(B_j) =$$

$$= -\sum_{i=1}^{n} p(A_i) \cdot \log_2 p(A_i) \sum_{j=1}^{m} p(B_j) - \sum_{j=1}^{m} p(B_j) \cdot \log_2 p(B_j) \sum_{i=1}^{n} p(A_i).$$

В слагаемых произведено изменение порядка суммирования в соответствии со значениями индексов. Далее, по условию нормировки $\sum_{i=1}^{n} p(A_i) = 1 \quad \sum_{i=1}^{m} p(B_j) = 1$

$$-\sum_{i=1}^{n} p(A_i) * \log_2 p(A_i) = H(\alpha) \qquad -\sum_{j=1}^{m} p(B_j) * \log_2 p(B_j) = H(\beta)$$

$$\boxed{H(\alpha \wedge \beta) = H(\alpha) + H(\beta)}$$

Теперь рассмотрим ситуацию, когда имеются два опыта с одинаковым числом исходов *n,* но в одном случае они равновероятны, а в другом - нет. Каково соотношение энтропии опытов? Примем без доказательства следующее утверждение:

$$-\sum_{i=1}^{n} p(A_i) log_2 p(A_i) \leq \log_2 n$$

Условная энтропия

Найдем энтропию сложного опыта α ^ β в том случае, если опыты не являются независимыми, т.е. если на исход β оказывает влияние результат опыта α. Например, если в ящике всего два разноцветных шара и α состоит в извлечении первого, а β - второго из них, то a полностью снимает неопределенность сложного опыта α ^ β, т.е. оказывается *H* (α ^ β) = *H(α),* а не сумме энтропии, как следует из (4).

Связь между α и β состоит в том, что какие-то из исходов $A^{(\alpha)}$ могут оказывать влияние на исходы из $B^{(\beta)}$, т.е. некоторые пары событий $A_i \wedge B_j$ не являются независимыми. Но тогда в (5) $p(A_i \wedge B_j)$ следует заменять не произведением вероятностей, а:

$$p\left(A_i \wedge B_j\right) = p(A_i) * P_{A_i}(B_j)$$

где - $P_{A_i}(B_j)$вероятность наступления исхода B, при условии, что в первом опыте имел место исход A_i. Тогда $\log_2 p\left(A_i \wedge B_j\right) = \log_2 p(A_i) + \log_2 p_{A_i}(B_j)$

При подстановке в (5) получаем:

$$H(\alpha \wedge \beta) = -\sum_{i=1}^{n} \sum_{j=1}^{m} p(A_i) * p\left({}^{B_j}/_{A_i}\right) * \{\log_2 p\ (A_i) + \log_2 p\ \left({}^{B_j}/_{A_i}\right)\} =$$

$$= -\sum_{i=1}^{n} \sum_{j=1}^{m} p(A_i) * p\left({}^{B_j}\Big/_{A_i}\right) * \log_2 p\ (A_i) - \sum_{i=1}^{n} \sum_{j=1}^{m} p(A_i) * p\left({}^{B_j}\Big/_{A_i}\right) * \log_2 p\left({}^{B_j}\Big/_{A_i}\right)$$

Свойства условной энтропии

1. $H_\alpha(\beta) = 0 \iff \forall$ исход α полностью определяет исход β, т. е.
$H_{A_1}(\beta) = H_{A_2}(\beta) = \cdots = H_{A_n}(\beta) = 0$
$H(\alpha \wedge \beta) = H(\alpha)$

2. Если α и β независимы, то $H_\alpha(\beta) = H(\beta)$, причём это оказывается наибольшим значением условной энтропии.

$$0 \leq H_\alpha(\beta) \leq H(\beta)$$

3. $H(\alpha \wedge \beta) \leq H(\alpha) + H(\beta)$

Энтропия и информация

- Предшествующий опыт α может уменьшить количество исходов и как следствие неопределённость последующего опыта.

$\boldsymbol{H}(\boldsymbol{\beta}) - \boldsymbol{H}_{\boldsymbol{\alpha}}(\boldsymbol{\beta})$ – показывает какие новые сведения относительно β получаем произведя опыт α.

$$\boldsymbol{I}(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \boldsymbol{H}_{\boldsymbol{\alpha}}(\boldsymbol{\beta}) \qquad (1)$$

$$\text{численное измерение количества информации}$$

1) Бит энтропии – бит информации
2) Энтропия равна информации относительно опыта, которая содержится в нём самом $\boldsymbol{I}(\boldsymbol{\beta}, \boldsymbol{\beta}) = \boldsymbol{H}(\boldsymbol{\beta})$

Энтропия опыта равна той информации, которую получаем в результате его осуществления.

Свойства:

1. $I(\alpha, \beta) \geq 0$

$I(\alpha, \beta) = 0 \Leftrightarrow \alpha$ и β −независимы

2. $I(\alpha, \beta) = I(\beta, \alpha)$

3. На основании следствия 2)$\mathrm{I} = -\sum_{i=1}^{n} p(A_i) \times \log_2 p\,(A_i)$ (2)

$I(\alpha, \beta)$ - среднее значениес случайной величины

$H(\beta) - H_{A_i}(\beta)$, связанной с отдельными исходами A_i опыта α

Задача 1

Пусть опыт β состоит в извлечении одного шара из урны, содержащей 5 черных и 10 белых шаров, опыт α_k - в предварительном извлечении из той же урны (без возвращения обратно) k шаров. Чему равна энтропия опыта β и информация об этом опыте, содержащаяся в опытах α_1, α_2, α_3 и α_4 ?

Задача 2.

Какое количество информации требуется, что-бы узнать исход броска монеты?

Решение: $n = 2, \quad p_1 = p_2 = 0{,}5$

$I = -0{,}5 \times \log_2 0{,}5 - 0{,}5 \times \log_2 0{,}5 = 1$ (бит)

Легко получить следствия формулы (2) для случая, когда все n исходов равновероятны.

$$\boldsymbol{p(A_i) = \frac{1}{n} \ \forall\, i = 1 \ldots n => I = \sum_{i=1}^{n}\left(\frac{1}{n} \times log_2\, n\right) = log_2\, n} \qquad \textbf{(3)}$$

формула Хартли выведена им в 1928 г., связывает количество равновероятных состояний (n) и количество информации в сообщении (*I*), что ∀ из этих состояний реализовалось.

$$\mathbf{Если\ n = 2^k, то\ I = k\ (бит)} \qquad \textbf{(4)}$$

- ***Задача 3***

Случайным образом вынимается карта из колоды в 32 карты. Какое количество информации требуется, что-бы угадать что это за карта?

Решение: $n = 2^5$, значит $k = 5$ и $=> I = 5$ (бит)

Задача 4

Некто задумал целое число в интервале от 0 до 3. Наш опыт состоит в угадывании этого числа. На наши вопросы некто может отвечать либо «да», либо «нет». Какое количество информации должны получить, что-бы узнать задуманное число, т.е. полностью снять начальную неопределённость?

Исходы: A_1 – "задуман 0"

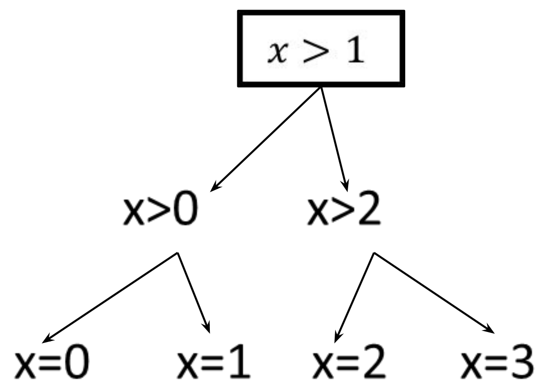
A_2 – "задумана 1" $P_1 = P_2 = P_3 = P_4$

A_3 – "задумана 2" $n = 4$

A_4 – "задумана 3" $\mathrm{p}(A_i) = \frac{1}{4}$

$\log_2 p(A_i) = -2, \qquad I = 2$ (бит)

Т.е. для конечного снятия неопределённости необходимо 2 бита информации, т.е. 2 вопроса.



Подобная процедура позволяет определить количество информации в ∀ задаче, интерпретация которой может быть сведена к парному выбору.

Замечание: Количество информации численно равно числу вопросов с равновероятными бинарными вариантами ответов, которые необходимо задать, чтобы полностью снять неопределенность задачи.

Свойства количества информации и энтропии

- Выражение (1) можно интерпретировать: если начальная энтропия опыта H_1, а в результате сообщения информации I энтропия становится равной H_2, то

$I = H_1 - H_2$, т.е. информация равна убыли энтропии.

<u>Информация – это содержание сообщения, понижающего неопределённость некоторого опыта с неоднозначным исходом, убыль связанной с ним энтропии является количественной мерой информации.</u>

$I = \log_2 n_1 - \log_2 n_2 = \log_2 \frac{n_1}{n_2}$ – в случае равновероятных исходов.

2. Из аддитивности энтропии независимых опытов следует аддитивность информации.

$$I = \log_2 n = \log_2(n_A \times n_B) = \log_2 n_A + \log_2 n_B = I_A + I_B$$

- **3.** Количество информации может быть измерено числом вопросов с двумя равновероятными ответами.

4. Формула (2) приводит к следующему выводу. Пусть некоторый опыт имеет 2 исхода А и В., $p_A = 0{,}99, \qquad p_B = 0{,}01.$

$I_A = -\log_2 0{,}99 = 0{,}0145 \text{ бит}$

$I_B = -\log_2 0{,}01 = 6{,}644 \text{ бит}$

Больше информации связано с теми исходами у которых вероятность меньше.

$I = 0{,}99 \times I_A + 0{,}01 \times I_B \approx 0{,}081 \text{ бит}$ –среднее количество информации.

Задача 1.4. В алфавите три буквы A, B, C. а) Составить максимальное количество сообщений, комбинируя по три буквы в сообщении. б) Какое количество информации приходится на одно такое сообщение? в) Чему равно количество информации на символ первичного алфавита?

Решение. а) $m_2 = 3$; $n = 3$; $N = m_2^n = 3^3 = 27$

AAA	BAA	CAA
AAB	BAB	CAB
AAC	BAC	CAC
ABA	BBA	CBA
ABB	BBB	CBB
ABC	BBC	CBC
ACA	BCA	CCA
ACB	BCB	CCB
ACC	BCC	CCC

б) $I = \log_2 N = \log_2 27 = 4{,}75489$ *бит*;

в) $H = \log_2 m_1 = \log_2 N = \log_2 m_2^n$.

Задача 1.10. Алфавит состоит из букв A, B, C, D. Вероятности появления букв равны соответственно $p_A = p_B = 0{,}25$; $p_C = 0{,}34$; $p_D = 0{,}16$. Определить количество информации на символ сообщения, составленного из такого алфавита.

Решение. Количество информации на символ алфавита есть энтропия данного алфавита. Так как символы алфавита неравновероятны, то энтропия равна

$$H = -\sum_{i=1}^{m} p_i \log_2 p_i = -(2\cdot 0{,}25 \log_2 0{,}25 + 0{,}34 \log_2 0{,}34 +$$
$$+ 0{,}16 \log_2 0{,}16) = 2 \cdot 0{,}5 + 0{,}529174 + 0{,}423017 =$$
$$= 1{,}952191 \textit{ бит/символ.}$$

Задача 1.16. Чему равно количество информации при получении 8 сообщений равномерного четырехзначного троичного кода?

Решение. Число качественных признаков $m = 3$. В коде они комбинируются по 4, т. е. $n = 4$. Число сообщений такого кода $N = m^n = 3^4$. Энтропия на одно сообщение $H = \log_2 N = 4 \log_2 3$. Количество информации в 8 сообщениях

$$I = 8 \cdot 4 \cdot \log_2 3 = 50{,}72 \textit{ бит.}$$

Примечание. Можно считать количество информации, определив энтропию на букву, блок, страницу и т. д. Количество информации в определенном объеме определяется умножением полученного значения энтропии соответственно на число букв, блоков, страниц.

Задача 1.28. Определить объем и количество информации в тексте «Широка страна моя родная», переданном стандартным телеграфным кодом № 3 (см. приложение 4).

Решение. Число принятых символов, включая пробел,

$$k = 24.$$

Объем информации

$$Q = kl_{\text{ср}} = 24 \cdot 7 = 168 \text{ } \textit{бит}.$$

Количество информации: а) для равновероятного алфавита

$$H_1 = \log_2 m = \log_2 32 = 5 \text{ } \textit{бит/символ},$$

$$I_1 = kH_1 = 24 \cdot 5 = 120 \text{ } \textit{бит};$$

б) для неравновероятного алфавита (в этом и подобных случаях энтропия первичного алфавита не высчитывается каждый раз, а берется энтропия русского алфавита)

$$H_2 = -\sum_{i=1}^{32} p_i \log_2 p_i = -(p_а \log_2 p_а + p_б \log_2 p_б + \ldots + p_я \log_2 p_я) \approx$$

$$\approx 4{,}36 \text{ } \textit{бит/символ},$$

$$I_2 = kH_2 \approx 24 \cdot 4{,}36 \approx 104{,}64 \text{ } \textit{бит}.$$