19. Кількість інформації і ентропія

Кількість інформації, що міститься у повідомленні, визначається не його змістом або сенсом. Поняттю інформації відповідає інтуїтивне уявлення про те, що кількість (міра) інформації пов'язана з степенем „несподіваності" цього повідомлення. Так, наприклад, повідомлення, відоме заздалегідь, не несе інформації. Інформація відображає стан матеріального об'єкта (джерела інформації), отже кількість інформації пов'язана з „несподіваністю" зазначеного стану серед всіх можливих станів, тобто з його ймовірностними характеристиками.

Точніше, кількість інформації I(x) у елементарному повідомленні (знаку, символі, елементі) x визначається ймовірністю p= p(x) появі цього елемента серед множини (ансамбля, абетки, алфавіта) цих елементів.

I(x) = I(p(x)) = I(p) (1)

Вид функцій I(p) може бути встановлений на підставі наступних властивостей (аксіом)

1. Невід'ємність: $I(p)\geq 0$ (2)

2. Нормованість: I(1) =0 (3)

(Вірогідне, тобто заздалегідь відоме повідомлення (p(x) =1) не містить інформації)

3. Монотонна спадність $p_1 > p_2$ ó $I(p_1) < I(p_2)$, ($p_{1,2} = p(x_{1,2})$) (4)

(Чим очікуваніше повідомлення, тим воно менш інформативне)

4. Адитивність I(x_1Ç x_2)= I(x_1)+I(x_2) (5)

(Кількість інформації у двох незалежних повідомленнях дорівнює сумі кількостей інформації у кожному з них)

Враховуючи, що за теоремою множення ймовірностей для незалежних подій

p(x_1Ç x_2)=p(x_1)·p(x_2)

з аксіоми адитивності випливає

$I(p_1 \cdot p_2) = I(p_1) + I(p_2)$ (6)

Функція (єдина), що відповідає умовам (1)+(6) - логарифмічна

$I(p) = -\log_q p$ (формула Шеннона) (7)

де $q>0$, $q\neq 1$ - довільне число, вибір якого визначає одиницю кількості інформації, якій відповідає ймовірність появи елементарного повідомлення

$$p = \frac{1}{q}$$

Якщо q= 2, $\left(p = \frac{1}{2}\right)$, то одиниця кількості інформації зветься „біт" (binary unit). Це кількість інформації, що міститься у повідомленні, ймовірність появи якого $\frac{1}{2}$

- = 0=>-log$_2$/7$_4$ = _- = iog$_2$e=> In 2

Якщо q = e, $\left(p = \frac{1}{e}\right)$, то одиниця кількості інформації зветься „нат" (natural unit).

Якщо q = 10, $\left(p = \frac{1}{10}\right)$, то одиниця кількості інформації зветься „діт" (decimal unit), або Хартлі

Найбільшого розповсюдження набула одиниця „біт" $I(x) = -\log_2 p(x)$ [Біт] (8)

і практичне її визначення - кількість інформації, що міститься у одному з двох можливих рівноймовірних повідомлень. Випадок, коли абетка містить 2 елементи, відповідає двоїстій (бінарній) комп'ютерній алгебрі.

Для абетки А обсягом N , A = { $x_1, x_2, \dots, x_N$} з рівноймовірними елементами

$p = p(x_k) = \frac{1}{N}$, (k = 1,2,...,N), кількість інформації у одному елементі

$I(x)=-\log_2 N$ (формула Хартлі (1927)), (9)
звідси випливає наведене вище визначення „Біт".
Для ансамбля повідомлень (абетки) A = { $x_1, x_2,..., x_N$} з рядом розподілу ймовірностей $p_k=p(x_k)$, (k = 1,2,...,N) кількість інформації в одному елементі

$I(x_k)=-\log_2 p_k$, (k = 1,2,...,N) (10)
є випадковою величиною, а її математичне сподівання

$$H = M(I(x_k)) = -\sum_{k=1}^{N} p_k \log_2 p_k \quad \left[\frac{Біт}{символ}\right], \qquad (11)$$

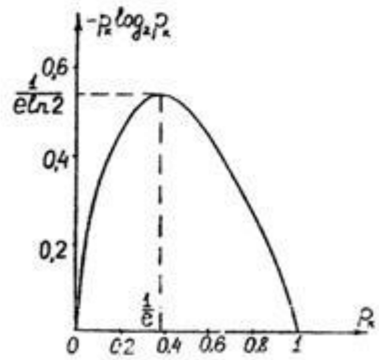
(середнє значення) зветься ентропією ансамбля (джерела повідомлень).
Ентропія являє собою міру невпорядкованості (хаотичності, невизначеності) стану джерела повідомлень.
2.2 Властивості ентропії
1. Невід' ємність і обмеженість: $H \geq 0$ (12)
(Кожний доданок $-p_k \log_2 p_k$ у проміжку $p_k \in [0;1]$ змінюється від 0 до максимума (за $p_k=\frac{1}{e}$), а потім знову спадає до нуля
Якщо $p_k \to 0$, то

$$\lim_{p_k\to 0}(-p_k \log_2 p_k) = \lim_{p_k\to 0} \frac{\log_2 \frac{1}{p_k}}{\frac{1}{p_k}} = \left(\frac{1}{p_k} = \alpha\right) = \lim_{\alpha\to\infty} \frac{\log_2 \alpha}{\alpha} = \left(\frac{\infty}{\infty}\right) = \lim_{\alpha\to\infty} \frac{\frac{1}{\alpha \cdot \ln 2}}{1} = 0$$

Якщо $p_k = 1$, то $-p_k \log_2 p_k = 0$

$$(-p_k \log_2 p_k) = \max : \frac{d}{dp_k}(-p_k \log_2 p_k) = 0 \Rightarrow -\log_2 p_k - \frac{p_k}{p_k \ln 2} = 0 \Rightarrow -\log_2 p_k = \frac{1}{\ln 2} = \log_2 e \Rightarrow$$

$$\Rightarrow p_k = \frac{1}{e}; -p_k \log_2 p_k = \frac{1}{e}\log_2 e = \frac{1}{e \ln 2} \approx 0.531$$



2. Ентропія дорівнює нулю, якщо повідомлення відомі заздалегідь:
$p_1 = p_2 = \ldots = p_N = 1 \Rightarrow H=0$
3. Ентропія максимальна, якщо всі елементи ансамбля рівноймовірні:

$$H_{max} = \log_2 N \Leftrightarrow p_k = \frac{1}{N} \qquad (13)$$

$$H-\log_2 N = \left(\sum_{i=1}^{N} p_i = 1\right) = -\sum_{i=1}^{N} p_i \log_2 p_i - \log_2 N \sum_{i=1}^{N} p_i = -\sum_{i=1}^{N} p_i (\log_2 p_i + \log_2 N) =$$

$$= \sum_{i=1}^{N} p_i \log_2 \frac{1}{Np_i} = \frac{1}{\ln 2}\sum_{i=1}^{N} p_i \ln \frac{1}{Np_i} = (\ln x \le x-1 \text{: } \ln x - x + 1 \textit{ має максимум у точці } x = 1) \le$$

$$\le \frac{1}{\ln 2}\sum_{i=1}^{N} p_i\left(\frac{1}{Np_i}-1\right) = \frac{1}{\ln 2}\sum_{i=1}^{N}\left(\frac{1}{N}-p_i\right) = \frac{1}{\ln 2}\left(\frac{N}{N}-1\right) = 0 \Rightarrow H \le \log_2 N$$

причому рівність має місце у випадку $x = \frac{1}{Np_i} = 1 \Rightarrow p_i = \frac{1}{N}$

$$H\max = \sum_{i=1}^{N} \frac{1}{N}\log_2 N = \log_2 N$$

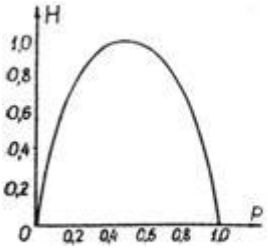
4. Ентропія бінарного джерела:
A = { x_1; x_2}; $p_1 = p_2$; $p_2 = 1-p_1 = 1 - p$
H=– p $\log_2$p-(1-p) $\log_2$(1-p) (14)
графік функції H(p) симетричний відносно прямої p = 0.5

p=0; $H(0) = \lim_{p\to 0}(-p\log_2 p) = 0$
p=1; H(1)=0

H=max: $\frac{dH}{dp} = 0 \Rightarrow$

$$\Rightarrow \log_2 p + \frac{p}{p\ln 2} - \log_2(1-p) - \frac{(1-p)}{(1-p)\ln 2} = 0;$$

$$\Rightarrow \log_2 \frac{p}{(1-p)} = 0 \Rightarrow \frac{p}{(1-p)} = 1 \Rightarrow p = \frac{1}{2}$$



5. Як зазначалось, ентропія системи - це невизначеність того, що система буде знаходитись в одному з N можливих станів і дорівнює середній кількості інформації, що міститься у одному символі. Отримана інформація пов'язана з зняттям невизначеності у знаннях таким чином, що
I = H-H апост (15)
де H - апріорна ентропія - невизначеність отримання елемента до експерименту
H апост - апостеріорна ентропія (після проведення експерименту) отримання елемента

6. У випадку взаємозалежності між символами ансамбля, коли ймовірність появи одного символу x_kзалежить від того, який символ x_i з'явився раніше, тобто є умовною ймовірністю p(x_k/ x_i), вводиться поняття умовної ентропії повідомлення x_k відносно x_i

$$H(x_k / x_i) = -\sum_{i=1}^{N} p(x_k / x_i)\log_2 p(x_k / x_i) \qquad (16)$$

Це випадкова величина, тому що випадковим є попередній елемент x_i.
Для отримання безумовної ентропії джерела необхідно виконати осереднення по всіх x_k

$$H = \sum_{i=1}^{N} H(x_k / x_i) = -\sum_{i=1}^{N} p(x_i)\left(\sum_{k=1}^{N} p(x_k / x_i)\log_2 p(x_k / x_i)\right) \qquad (17)$$

Вираз (11) є окремим випадком (17) за умови незалежності подій x_k та x_i,

$$\left(p(x_k / x_i) = p(x_k) \;\; ; \;\; \sum_{k=1}^{N} p(x_k) = 1\right)$$

7. У випадку, коли множина можливих елементів (повідомлень) -неперервна випадкова величина з щільністю φ(x), вводиться понять диференціальної ентропії, яка визначається виразом, аналогічним (11)

$$H = -\int_{-\infty}^{\infty} \varphi(x)\log_2 \varphi(x)dx \qquad \left(= -\frac{1}{\ln 2}\int_{-\infty}^{\infty} \varphi(x)\ln x dx\right) \qquad (18)$$

а кількість інформації - співвідношенням (15)

Для ентропії неперервних повідомлень чинні властивості 1)÷2), але на відміну від 3) максимум диференціальної ентропії має місце у випадку нормального розподілу

$$\varphi(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{(x-M)^2}{2\sigma^2}} \qquad (19)$$

де M і σ^2 - відповідно математичне сподівання і дисперсія.

(Максимум функціонала (18) за умов нормування і заданої дисперсії

$$\int_{-\infty}^{\infty} \varphi(x)dx = 1; \qquad \int_{-\infty}^{\infty} (x-M)^2\varphi(x)dx = \sigma^2 \qquad (20)$$

за теоремами варіаційного числення (Ейлера) відповідає максимум функціонала

$$\int_{-\infty}^{\infty}\left[\varphi(x)\ln\varphi(x) + \lambda_1\varphi(x) + \lambda_2(x-M)^2\right]dx$$

тобто умова

$$\frac{\partial}{\partial\varphi}\left[-\varphi(x)\ln\varphi(x) + \lambda_1\varphi(x) + \lambda_2(x-M)^2\right] = 0 \qquad (21)$$

де сталі λ_1, λ_2 = Const визначаються (20).

З(21)маємо $-\ln\varphi(x) - 1 + \lambda_1 + \lambda_2(x-M)^2 = 0$

$$\varphi(x) = e^{\lambda_1 - 1} \cdot e^{\lambda_1(x-M)^2}$$

За (22) інтеграли (20) збігаються, якщо $\lambda_1 = -|\lambda_2| < 0$.

Тоді $\int_{-\infty}^{\infty} e^{\lambda_1(x-M)^2}dx = \sqrt{\frac{\pi}{|\lambda_1|}}$; $\int_{-\infty}^{\infty} (x-M)^2 e^{-\lambda_1(x-M)^2}dx = \frac{1}{2|\lambda_1|}\sqrt{\frac{\pi}{|\lambda_1|}}$

рівняння (20) дають $e^{\lambda_2 - 1} = \sqrt{\frac{|\lambda_1|}{\pi}}$; $|\lambda_1| = \frac{1}{2\sigma^2}$

а (22) перетворюється у (19))

За умови (19) максимум ентропії неперервних повідомлень

$$H_{max} = \log_2\left(e\sqrt{2\pi\sigma^2}\right) \qquad (23)$$