

Формула Шеннона

Какое сообщение содержит большее количество информации?

Бабушка испекла 16 пирожков с капустой, 16 пирожков с повидлом. Маша съела один пирожок.

Бабушка испекла 8 пирожков с капустой, 24 пирожка с повидлом. Маша съела один пирожок.

Для вычисления количества информации в сообщении о неравновероятном событии используют следующую формулу:

$$I = \log_2(1/p)$$

где I – это количество информации,
 p – вероятность события.

Вероятность события выражается в долях единицы и вычисляется по формуле:

$$p = K/N,$$

где K – величина, показывающая сколько раз произошло интересующее нас событие,
 N – общее число возможных исходов какого-то процесса.

Вернемся к нашей задаче.

Пусть K_1 – это количество пирожков с повидлом, $K_1=24$

K_2 – количество пирожков с капустой, $K_2=8$

N – общее количество пирожков, $N=K_1+K_2=24+8=32$

Вычислим вероятность выбора пирожка с разной начинкой и количество информации, которое при этом было получено.

Вероятность выбора пирожка с повидлом: $p_1=24/32=3/4=0,75$.

Вероятность выбора пирожка с капустой: $p_2=8/32=1/4=0,25$.

Обращаем внимание учащихся на то, что в сумме все вероятности дают 1.

Вычислим количество информации, содержащееся в сообщении, что Маша выбрала пирожок с повидлом:

$I_1=\log_2(1/p_1)=\log_2(1/0,75)=\log_2 1,3=0,379$ бит.

Вычислим количество информации, содержащееся в сообщении, если был выбран пирожок с капустой:

$I_2=\log_2(1/p_2)=\log_2(1/0,25)=\log_2 4=2$ бит.

При сравнении результатов вычислений получается следующая ситуация: *вероятность выбора пирожка с повидлом больше, чем с капустой, а информации при этом получилось меньше. Это не случайность, а закономерность.*

Качественную связь между вероятностью события и количеством информации в сообщении об этом событии можно выразить так: ***чем меньше вероятность некоторого события, тем больше информации содержит сообщение об этом событии.***

Вернемся к нашей задаче с пирожками. *Мы еще не ответили на вопрос: сколько получим информации при выборе пирожка любого вида?*

Ответить на этот вопрос нам поможет формула вычисления количества информации для событий с различными вероятностями, которую предложил в 1948 г. американский инженер и математик **К.Шеннон**.

Если **I**-количество информации, **N**-количество возможных событий, **p_i** - вероятности отдельных событий, где **i** принимает значения от 1 до **N**, то количество информации для событий с различными вероятностями можно определить по формуле:

ФОРМУЛА ШЕННОНА

Количество информации для событий с различными вероятностями определяется по формуле:

$$I = - \sum_{i=1}^N p_i \log_2 p_i$$

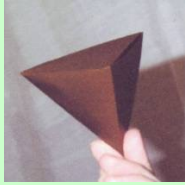
I – количество информации,
N – количество возможных событий
p_i – вероятности отдельных событий

$$I = -(p_1 \cdot \log_2 p_1 + p_2 \cdot \log_2 p_2 + p_3 \cdot \log_2 p_3 + p_4 \cdot \log_2 p_4 + \dots + p_N \cdot \log_2 p_N)$$

Рассмотрим формулу на нашем примере:

$$\begin{aligned} I &= -(p_1 \cdot \log_2 p_1 + p_2 \cdot \log_2 p_2) = -(0,75 \cdot \log_2 0,75 + 0,25 \cdot \log_2 0,25) \approx \\ &\approx -(0,75 \cdot (-0,42) + 0,25 \cdot (-2)) = 0,815 \text{ бит} \end{aligned}$$

ЗАДАНИЕ «БРОСАНИЕ ПИРАМИДКИ»



Определить количество информации, которую мы получим в результате бросания несимметричной и симметричной пирамидок.

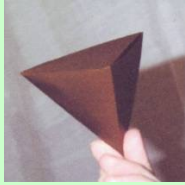
При бросании **несимметричной** четырехгранной пирамидки вероятности отдельных событий равны: $p_1=1/2$; $p_2=1/4$; $p_3=1/8$; $p_4=1/8$.

Количество информации рассчитываем по формуле:

$$I = - \sum_{i=1}^N p_i \log_2 p_i.$$

$$\begin{aligned} I &= - (1/2 \cdot \log_2 1/2 + 1/4 \cdot \log_2 1/4 + 1/8 \cdot \log_2 1/8 + 1/8 \cdot \log_2 1/8) \text{ бит} = \\ &= (1/2 \cdot \log_2 2 + 1/4 \cdot \log_2 4 + 1/8 \cdot \log_2 8 + 1/8 \cdot \log_2 8) \text{ бит} = \\ &= (1/2 + 2/4 + 3/8 + 3/8) \text{ бит} = 14/8 \text{ бит} = \mathbf{1,75 \text{ бита}}. \end{aligned}$$

ЗАДАНИЕ «БРОСАНИЕ ПИРАМИДКИ»



Определить количество информации, которую мы получим в результате бросания несимметричной и симметричной пирамидок.

При бросании **симметричной** четырехгранной пирамидки вероятности отдельных событий равны между собой: $p_1 = p_2 = p_3 = p_4 = 1/4$.

Количество информации рассчитываем по формуле: $I = \log_2 N$.

$$I = \log_2 4 = 2 \text{ бита.}$$

Количество информации, которую мы получаем, достигает **максимального значения**, если события **равновероятны**.

Если **события равновероятны** ($p_i = 1/N$) - Формула Хартли:

$$I = - \sum_{i=1}^N \frac{1}{N} \log_2 \frac{1}{N} = \log_2 N$$

ПРАКТИКУМ

Определение количества информации

Задача. В непрозрачном мешочке хранятся 10 белых, 20 красных, 30 синих и 40 зеленых шариков. Какое количество информации будет содержать зрительное сообщение о цвете вынутого шарика?

Так как количество шариков различных цветов неодинаково, то **вероятности** зрительных сообщений о цвете вынутого из мешочка шарика также различаются и **равны количеству шариков данного цвета, деленному на общее количество шариков**:

$$p_{\text{б}} = 0,1; \quad p_{\text{к}} = 0,2; \quad p_{\text{с}} = 0,3; \quad p_{\text{з}} = 0,4.$$

События **неравновероятны**, поэтому воспользуемся формулой

$$I = - \sum_{i=1}^N p_i \log_2 p_i.$$

ПРАКТИКУМ

Определение количества информации

Задача. В непрозрачном мешочке хранятся 10 белых, 20 красных, 30 синих и 40 зеленых шариков. Какое количество информации будет содержать зрительное сообщение о цвете вынутого шарика?

Так как количество шариков различных цветов неодинаково, то **вероятности** зрительных сообщений о цвете вынутого из мешочка шарика также различаются и **равны количеству шариков данного цвета, деленному на общее количество шариков**:

$$p_{\text{б}} = 0,1; \quad p_{\text{к}} = 0,2; \quad p_{\text{с}} = 0,3; \quad p_{\text{з}} = 0,4.$$

События **неравновероятны**, поэтому воспользуемся формулой

$$I = - \sum_{i=1}^N p_i \log_2 p_i.$$

$$I = - (0,1 \cdot \log_2 0,1 + 0,2 \cdot \log_2 0,2 + 0,3 \cdot \log_2 0,3 + 0,4 \cdot \log_2 0,4) \text{ бит.}$$

Для вычисления этого выражения воспользуемся компьютерным калькулятором

$$\log_a b = \frac{\log_c b}{\log_c a}$$

Таким образом, $I \approx 1,85$ бита.

ПРАКТИКУМ

Задача: В корзине лежат 32 клубка красной и черной шерсти. Среди них 4 клубка красной шерсти. Сколько информации несет сообщение, что достали клубок красной шерсти? Сколько информации несет сообщение, что достали клубок шерсти любой окраски?

ПРАКТИКУМ

Задача: В корзине лежат 32 клубка красной и черной шерсти. Среди них 4 клубка красной шерсти.

Сколько информации несет сообщение, что достали клубок красной шерсти? Сколько информации несет сообщение, что достали клубок шерсти любой окраски?

Дано: $K_K=4; N=32$

Найти: I_K, I

Решение:

1. Найдем количество клубков черной шерсти: $K_Ч=N- K_K; K_Ч=32-4=28$
2. Найдем вероятность доставания клубка каждого вида: $p_K= K_K/N=4/32=1/8; p_Ч=K_Ч/N=28/32=7/8;$
3. Найдем количество информации, которое несет сообщение, что достали клубок красной шерсти: $I_K= \log_2(1/(1/ p_K))= \log_2(1/1/8)= \log_2 8=3$ бит
4. Найдем количество информации, которое несет сообщение, что достали клубок шерсти любой окраски:

$$I = -\sum_{i=1}^N p_i \cdot \log_2 p_i = -(1/8 \cdot \log_2(1/8) + 7/8 \cdot \log_2(7/8)) = -((1/8) \cdot (-3) + (7/8) \cdot (-0,193)) =$$
$$= -(-0,375 - 0,169) = 0,547 \text{ бит}$$

Ответ: $I_K=3$ бит; $I=0,547$ бит

КОМПЬЮТЕРНЫЙ ПРАКТИКУМ

Задача №1

В озере обитает 12500 окуней, 25000 пескарей, а карасей и щук по 6250. Какое количество информации несет сообщение о ловле рыбы каждого вида. Сколько информации мы получим, когда поймем какую-нибудь рыбу?

КОМПЬЮТЕРНЫЙ ПРАКТИКУМ

Задача №1

В озере обитает 12500 окуней, 25000 пескарей, а карасей и щук по 6250. Какое количество информации несет сообщение о ловле рыбы каждого вида. Сколько информации мы получим, когда поймем какую-нибудь рыбу?

Дано: $K_o=12500$; $K_n=25000$; $K_k=K_{щ}=6250$

Найти: $I_o, I_n, I_k, I_{щ}, I$

Решение:

1. Найдем общее количество рыбы: $N = K_o + K_n + K_k + K_{щ}$.
2. Найдем вероятность ловли каждого вида рыбы: $p_o = K_o/N$; $p_n = K_n/N$; $p_k = p_{щ} = K_k/N$.
3. Найдем количество информации о ловле рыбы каждого вида: $I_o = \log_2(1/p_o)$; $I_n = \log_2(1/p_n)$; $I_k = I_{щ} = \log_2(1/p_k)$.
4. Найдем количество информации о ловле рыбы любого вида: $I = p_o \cdot \log_2 p_o + p_n \cdot \log_2 p_n + p_k \cdot \log_2 p_k + p_{щ} \cdot \log_2 p_{щ}$.

КОМПЬЮТЕРНЫЙ ПРАКТИКУМ

Мы сделаем *универсальную* таблицу, где достаточно занести данные задачи, а вычисление результатов будет происходить автоматически.

При составлении таблицы мы должны учитывать:

- 1) Ввод данных (что дано в условии).
- 2) Подсчет общего количества числа возможных исходов (формула $N = K_1 + K_2 + \dots + K_i$).
- 3) Подсчет вероятности каждого события (формула $p_i = K_i / N$).
- 4) Подсчет количества информации о каждом происходящем событии (формула $I_i = \log_2(1/p_i)$).
- 5) Подсчет количества информации для событий с различными вероятностями (формула Шеннона).

Рассмотрим заполнение таблицы на примере задачи №1.

	A	B	C	D	E	F
1	События	Число событий	Вероятность ($P=K/N$)	Кол-во информ. о происходящих событиях $I=\text{LOG}_2(1/P)$	$P \cdot \text{LOG}_2(P)$	$I = -\sum P_i \cdot \text{LOG}_2 P_i$
2	окунь	12500	=B2/\$B\$6	=LOG(1/C2;2)	=C2*LOG(C2;2)	=-СУММ(E2:E5)
3	пескарь	25000	=B3/\$B\$6	=LOG(1/C3;2)	=C3*LOG(C3;2)	
4	карась	6250	=B4/\$B\$6	=LOG(1/C4;2)	=C4*LOG(C4;2)	
5	щука	6250	=B5/\$B\$6	=LOG(1/C5;2)	=C5*LOG(C5;2)	
6	Всего (N)	=СУММ(B2:B5)				

Рис. 1. Режим отображения формул

	A	B	C	D	E	F
1	События	Число событий	Вероятность ($P=K/N$)	Кол-во информ. о происходящих событиях	$P \cdot \text{LOG}_2(P)$	$I = -\sum P_i \cdot \text{LOG}_2 P_i$
2	окунь	12500	0,2500	2,0000	-0,5000	1,7500
3	пескарь	25000	0,5000	1,0000	-0,5000	
4	карась	6250	0,1250	3,0000	-0,3750	
5	щука	6250	0,1250	3,0000	-0,3750	
6	Всего (N)	50000				

Рис. 2. Отображение результатов вычислений

Задача №2

В классе 30 человек. За контрольную работу по информатике получено 15 пятерок, 6 четверок, 8 троек и 1 двойка. Какое количество информации несет сообщение о том, что Андреев получил пятерку?

	A	B	C	D	E	F
1	События	Число событий	Вероятность ($P=K/N$)	Кол-во информ. о происходящих событиях	$P \cdot \text{LOG}_2(P)$	$I = -\sum P_i \cdot \text{LOG}_2 P_i$
2	5	15	0,5000	1,0000	-0,5000	1,6365
3	4	6	0,2000	2,3219	-0,4644	
4	3	8	0,2667	1,9069	-0,5085	
5	2	1	0,0333	4,9069	-0,1636	
6	Всего (N)	30				

КОМПЬЮТЕРНЫЙ ПРАКТИКУМ

Задача №3

В коробке лежат кубики: 10 красных, 8 зеленых, 5 желтых, 12 синих. Вычислите вероятность доставания кубика каждого цвета и количество информации, которое при этом будет получено.

Задача №4

В непрозрачном мешочке хранятся 10 белых, 20 красных, 30 синих и 40 зеленых шариков. Какое количество информации будет содержать зрительное сообщение о цвете вынутого шарика?

Задача №5

а) Решить задачу (про клубки шерсти) при помощи электронной таблицы. б) Добавить в условие задачи 14 клубков зеленой шерсти и решить при помощи электронной таблицы.