Измерение информации (вероятностный подход)

Пусть в некотором сообщении содержатся сведения о том, что произошло одно из N равновероятных событий.

> **Примечание**
>
> Равная вероятность обозначает, что ни одно событие не имеет преимуществ перед другими.

В 1928 году американский ученый-электроник Ральф Хартли предложил научный подход к оценке сообщений. Предложенная им формула имела следующий вид:

$I = \log_2 N$, или $\mathbf{N=2^{I}}$

где N – количество равновероятных событий;

I – количество информации, заключенное в этом сообщении (бит).

Задача № 1

Шарик находится в одной из восьми урн: A, B, C, D… Определить сколько бит информации содержит сообщение о том, что он находится в урне B.

Решение

Такое сообщение содержит $I = \log_2 8 = 3$ бита информации.

Задача № 2

Ученик может с равной вероятностью получить одну из оценок 5, 4, 3 или 2. Определить сколько бит информации содержит сообщение о том, что он получит одну из оценок.

Решение

Такое сообщение содержит $I = \log_2 4 = 2$ бита информации.

Разная вероятность события

Для примера возьмем школьные оценки.

Чтобы определить, какова вероятность получения каждой оценки, нужно посчитать общее количество разных оценок, полученных учеником за достаточно большой период времени, и определить, сколько из них двоек, троек, четверок и пятерок.

Допустим, что такое распределение оценок сохранится и в будущем, то можно рассчитать вероятности получения каждой из оценок. Определив, какую часть от общего числа оценок составляют двойки, найдем вероятность получения двойки. Затем, определив, какую часть составляют тройки, найдем вероятность получения тройки. Доля четверок среди всех оценок – это вероятность получения четверки, а доля пятерок – это вероятность получения пятерки.

Предположим, мы посчитали, что за год ученик получил 100 оценок. Среди них: 60 пятерок, 25 четверок, 10 троек и 5 двоек. Тогда:

Оценка	Количество	Вероятность (P)	
5	60	60/100 = 0,6	
4	25	25/100 = 0,25	
5	10	10/100 = 0,1	
2	5	5/100 = 0,05	
Итого	100	1,00	

Теперь, зная вероятности событий, можно определить количество информации в сообщении о каждом из них. Формулу предложил К.Шеннон в 1948 году.

$\mathbf{2^I = 1/P}$ или $\mathbf{I = \log_2 (l/P)}$

Определим (с точностью до трех знаков после запятой) по этой формуле количество информации, содержащееся в сообщении о получении нашим учеником каждой из оценок.

Оценка	Количество	Вероятность (P)	Количество информации, содержащееся в сообщении о получении оценки (I – бит)
5	60	60/100 = 0,6	$\log_2 (l/0,6) = \log_2 (5/3) = 0,737$
4	25	25/100 = 0,25	$\log_2 (l/0,25) = \log_2 (4) = 2$
3	10	10/100 = 0,1	$\log_2 (l/0,l) = \log_2 (10) = 3,322$
2	5	5/100 = 0,05	$\log_2 (l/0,05) = \log_2 (20) = 4,322$
Итого	100	1,00	

Количество информации в сообщении о некотором событии зависит от вероятности этого события. Чем меньше вероятность, тем больше информации.

На первый взгляд, кажется, что мы имеем две совсем разные формулы для вычисления информации. Первая – через количество событий, вторая – через вероятность

1) $\mathbf{I = \log_2 N}$

2) $\mathbf{I = \log_2 (l/P)}$

На самом деле это не разные формулы! Первая формула является частным случаем второй, когда вероятность событий одинаковая.

Представьте себе, что у нашего ученика было бы всех оценок поровну: пятерок, четверок, троек, двоек – по 25 штук.

Оценка	Количество	Вероятность (P)	Количество информации, содержащееся в сообщении о получении оценки (I – бит)
5	25	25/100 = 0,25	$\log_2 (l/0,25) = \log_2 (4) = 2$
4	25	25/100 = 0,25	$\log_2 (l/0,25) = \log_2 (4) = 2$
3	25	25/100 = 0,25	$\log_2 (l/0,25) = \log_2 (4) = 2$
2	25	25/100 = 0,25	$\log_2 (l/0,25) = \log_2 (4) = 2$
Итого	100	1,00	

Тогда вероятность каждой оценки была бы равна 25/100 = 1/4. Значит, и количество информации будет одинаковым. Посчитаем:

$\mathbf{I_5 = I_4 = I_3 = I_2 = \log_2 (1/0,25) = \log_2 (4) = 2}$ **бита.**

Но это – та же самая задача о четырех равновероятных оценках, которую мы решали раньше. И там тоже получалось 2 бита.

Задача № 1

В корзине лежат 8 черных шаров и 24 белых. Сколько информации несет сообщение о том, что достали черный шар?

Цвет шара	Количество	Вероятность (Р)	Количество информации, содержащееся в сообщении о получении оценки (I – бит)
		P = X/N	**$\log_2(I/P) = \log_2(N/X)$**
черный	8	8/32 = 0,25	$\log_2(I/0{,}25) = \log_2(4) = 2$
белый	24		
Итого	32		

Ответ: 2 бита

Задача № 2

В коробке лежат 64 цветных карандаша. Сообщение о том, что достали белый карандаш, несет 4 бита информации. Сколько не белых карандашей было в корзине?

Цвет карандаша	Количество	Вероятность (Р)	Количество информации, содержащееся в сообщении о получении оценки (I – бит)
		P = X/N	**$\log_2(I/P) = \log_2(N/X)$**
белый	$К_б$	$К_б/64$	$\log_2(64/К_б) = 4$
другие цвета	$К_{др}$		
Итого	64		

$\log_2(64/К_б) = 4$

$\log_2(64/К_б) = \log_2(16)$

$64/К_б = 16$

$К_б = 64/16 = 4$

$К_{др} = 64 – 4 = 60$

Ответ: 60 карандашей

Информационная энтропия

Мера неопределённости или непредсказуемости некоторой системы (в статистической физике или теории информации), в частности неопределённость появления какого-либо символа первичного алфавита.

Энтропия независимых случайных событий x с n возможными состояниями (от 1 до n) рассчитывается по формуле:

$$H(x) = \sum p(i) \times log_2\left(\frac{1}{p(i)}\right)$$

Эта величина также называется средней энтропией сообщения.

log2 (l/P_i) – частная энтропия, характеризующая только i-е состояние.

Задача № 3

В коробке лежат карандаши, из них 16 красных, 8 синих и 8 зеленых. Какова информационная энтропия сообщения, что выбрали карандаш любого цвета. Какова вероятность, что карандаш оказался синим? Сколько бит несет информация о том, что карандаш зеленый?

Цвет карандаша	Количество	Вероятность (P)	Количество информации, содержащееся в сообщении о получении оценки (I – бит)
		P = X/N	**log $_2$ (I/P) = log $_2$ (N/X)**
красный	16	16/32=0,5	$\log_2 (32/16) = 1$
синий	8	8/32=0,25	$\log_2 (32/8) = 2$
зеленый	8	8/32=0,25	$\log_2 (32/8) = 2$
Итого	16+8+8=32		

H = 0,5×1+0,25×2+0,25×2=2,5

Ответ: 2,5; 0,25; 2

Задача № 4

И коробке лежат кубики: 64 красных, 32 зеленых, 16 желтых, 8 синих, 8 белых. Вычислите вероятность доставания кубика каждого цвета и количество информации, которое при этом будет получено. Определите какого цвета кубик будет нести наибольшее количество информации.

Являются ли события равновероятными? Почему? (Нет, т.к. количество кубиков разное.)

Какова информационная энтропия сообщения, что выбрали кубик произвольного цвета.

Цвет	Количество	Вероятность (P)	Количество информации, содержащееся в сообщении о получении оценки (I – бит)
		P = X/N	$\log_2 (I/P) = \log_2 (N/X)$
красный	64	64/128 = 0,5	$\log_2 (l/0,5) = 1$
зеленый	32	32/128 = 0,25	$\log_2 (l/0,25) = 2$
желтый	16	16/128 = 0,125	$\log_2 (l/0,125) = 3$
синий	8	8/128 = 0,0625	$\log_2 (l/0,0625) = 4$
белые	8	8/128 = 0,0625	$\log_2 (l/0,0625) = 4$
Итого	128		

Ответ: наибольшее количество информации мы получим при доставании желтого кубика по причине качественной связи между вероятностью и количеством информации.

$H = 0,5 \times 1 + 0,25 \times 2 + 0,125 \times 3 + 0,0625 \times 4 + 0,0625 \times 4 = 1,875$

Задача № 4

В школе 32 компьютера размещены в двух кабинетах А и В. Сообщение «сломался компьютер из кабинета А несет 3 бита информации. Сколько компьютеров находится в кабинете В?

Кабинеты	Количество	Вероятность (P)	Количество информации, содержащееся в сообщении о получении оценки (I – бит)
		P = X/N	$\log_2 (I/P) = \log_2 (N/X)$
А	Ка	$P_A = Ка/N$	$\log_2 (N/Ка) = 3$
В	Кв	$P_B = Кв/N$	$\log_2 (N/Кв)$

N = Ка + Кв = 32

$\log_2 (N/Ка) = 3$

$\log_2 (32/Ка) = 3$ или $\log_2 (32/Ка) = \log_2 8$ или $32/Ка = 8$, т.е. Ка = 4

Следовательно, в кабинете В 32 – 4 = 28 компьютеров.

Ответ: 28 компьютеров

Задача № 5

В зоопарке 32 обезьяны живут в двух вольерах, А и Б. Сообщение обезьяна живет в вольере А» содержит 4 бита информации. Сколько обезьян живут в вольере Б?

Кабинеты	Количество	Вероятность (P)	Количество информации, содержащееся в сообщении о получении оценки (I – бит)
		P = X/N	**$\log_2 (I/P) = \log_2 (N/X)$**
А	Ка	P_A = Ка/N	$\log_2$(N/Ка) = 4
Б	Кб	P_B = Кб/N	$\log_2$(N/Кб)

N = Ка + Кб = 32

$\log_2$(N/Ка) = 4

$\log_2$(32/Ка) = 4 или $\log_2$(32/Ка) = $\log_2 16$ или 32/Ка = 16, т.е. Ка = 2

Следовательно, в кабинете В 32 – 2 = 30 обезьян.

Ответ: 30 обезьян

Задача № 6

Словарь людоедов племени Мумбо-Юмбо содержит слова только 3-х частей речи: существительные, глаголы и междометия. Каждый раз за обедом, по причине своей дикости, людоед произносит предложение, состоящее из одного равновероятно выбранного из словаря слова. Количество информации, содержащееся в сообщении «Предложение состоит из глагола», равно $1 + \log_2 5$ бита. Информационный объем сообщения «Предложение состоит из существительного» равен $1 + \log_2 3$ бита. В словаре только 8 слов не являются междометиями. Определить количество всех слов в словаре.

Части речи	Количество	Вероятность (P)	Количество информации, содержащееся в сообщении о получении оценки (бит)
		P = X/N	**$\log_2 (I/P) = \log_2 (N/X)$**
глагол	Г	$P_Г$ = Г/N	$\log_2$(N/Г) = $1 + \log_2 5$
существительное	С	P_C = C/N	$\log_2$(N/C) = $1 + \log_2 3$
междометие	М	P_M = M/N	

N = Г + С + М

По условию:

1) $\log_2$(N/Г) = $1 + \log_2 5$

$\log_2$(N/Г) = $\log_2 2 + \log_2 5$ или N/Г = 2×5 = 10, следовательно Г = N/10

2) $\log_2$(N/C) = $1 + \log_2 3$

$\log_2$(N/C) = $\log_2 2 + \log_2 3$ или N/C = 2×3 = 6, следовательно С = N/6

3) Г + С = 8 или N/10 + N/6 =8, следовательно N = 30

Ответ: количество всех слов в словаре равно 30

Задача № 7

У скупого рыцаря в сундуке золотые, серебряные и медные монеты. Каждый вечер он извлекает из сундука одну из лежащих в нем 96 монет, любуется ею и кладет обратно в сундук. Количество информации, содержащееся в сообщении «Из сундука извлечена серебряная монета», равно четырем битам. Информационный объем сообщения «Из сундука извлечена золотая монета» равен пяти битам. Определить количество медных монет в сундуке.

Монета	Количество	Вероятность (P)	Количество информации, содержащееся в сообщении о получении оценки (бит)
		P = X/N	**$\log_2(I/P) = \log_2(N/X)$**
золотая	Зол	$P_{Зол} = Зол/N$	$\log_2(N/Зол) = 5$
серебряная	С	$P_C = C/N$	$\log_2(N/C) = 4$
медная	М	$P_M = M/N$	

N = Зол + С + М

Зол + С + М = 96

Имеем систему из трех уравнений с тремя неизвестными.

Зол + С + М = 96

$4 = \log_2(96/C)$ $16 = 96/C$ $C = 6$

$5 = \log_2(96/Зол)$ $32 = 96/Зол$ $Зол = 3$

М = 96 – 6 – 3 = 87

Ответ: медных монет 87