Содержательный подход и вероятность

Содержательный подход и вероятность

До сих пор речь шла о равновероятных событиях. Но в реальности существует множество ситуаций, когда возможные события имеют различные вероятности реализации. Рассмотрим примеры таких событий:

1. В коробке 20 карандашей, из них 15 красных и 5 чёрных. Вероятность вытащить наугад красный карандаш больше, чем чёрный.
2. Сообщения об осадках в зимнем прогнозе погоды. Зимой бывает снег, бывает без осадков и, очень редко, бывает дождь (во время сильной оттепели). Дождь — событие маловероятное. Поэтому зимой сообщение о дожде несет самую большую информацию.
3. В пруду живут 8000 карасей, 2000 щук и 40000 пескарей. Самая большая вероятность для рыбака – поймать в этом пруду пескаря, на втором месте – карася, на третьем – щуку.
4. Интуитивно понятно, например, что для ученика-отличника получение пятерки и получение двойки — события не равновероятные. Для такого ученика получить пятерку —очень вероятное событие, а получение двойки — маловероятно. Для двоечника же — все наоборот.

Вычисление вероятности

Вероятность показывает **<u>какую часть от общего числа</u>** событий составляет данное событие

$$p_i = \frac{n_i}{N}$$

n_i число данных событий

N общее число событий

Вероятность события (p) – число от 0 до 1, показывающее, как часто случается это событие в большой серии одинаковых опытов.

p = 0 событие **никогда** не происходит (нет неопределенности)

p = 0,5 событие происходит в половине случаев (есть неопределенность)

p = 1 событие происходит **всегда** нет неопределенности)

Полная система событий: одно из N событий обязательно произойдет (и только одно!).

p_i **– вероятность выбора** i**-ого варианта (**$i=1,\ldots,N$**)**

p= 1/N ⇒ $$i = \log_2 \frac{1}{p}$$

Вероятностный подход

Для оценки средней информативности событий с учетом разной вероятности их благоприятных исходов формула **К. Шеннона:**

$$\boxed{i=\sum_{n=1}^{N} p_n\cdot\log_2\frac{1}{p_n}=p_1\cdot\log_2\frac{1}{p_1}+p_2\cdot\log_2\frac{1}{p_2}+\ldots+p_N\log_2\frac{1}{p_N}}$$

i	средняя информативность в сообщении о том, что произошло одно из N событий.
p_n	вероятность (частота) наступления событий
N	количество возможных событий (мощность алфавита)

Если: $p_1=p_2=\ldots=p_n=p$ (<u>события равновероятны</u>), тогда:

$$i=p\cdot\log_2\frac{1}{p}+p\cdot\log_2\frac{1}{p}+\ldots+p\cdot\log_2\frac{1}{p}=\underbrace{p\cdot N}_{p=\frac{1}{N}}\cdot\log_2\frac{1}{p}\ \Rightarrow\ \boxed{\begin{matrix}i=\log_2 N\\ N=2^i\end{matrix}}$$

Формула Хартли

Содержательный подход и вероятность

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. Если допустить, что такое же 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.

Задача

За год ученик получил 100 оценок. Среди них: 60 пятерок, 25 четверок, 10 троек и 5 двоек. Определить количество информации в сообщениях о получении каждой из оценок

Решение

P_5=60/100 =0,6(60%);
P_4=25/100 =0,25(25%);
P_3=10/100 =0,1(10%);
P_2=5/100 =0,05(5%)

$$2^i=1/P \qquad i = \log_2(l/P)$$

$I_5 = \log_2(l/0,6) = \log_2(5/3) = 0,737$ бит,
$I_4 = \log_2(l/0,25) = \log_2(4) = 2$ бита,
$I_3 = \log_2(l/0,l) == \log_2(10) = 3,322$ бита,
$I_2 = \log_2(l/0,05) = \log_2(20) = 4,322$ бита.

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

Содержательный подход и вероятность

Задача

В пруду живут 100 рыб, из них 20 карасей, 30 пескарей, а остальные – окуни. *Сколько информации* несет сообщение о том, что рыбак поймал ***карася*** (***пескаря, окуня***), если все рыбы одинаково голодны?

Решение:

$$I_i = -\log_2 p_i = \log_2 \frac{1}{p_i}$$

карась $p_1 = 20/100 = 0{,}2$ $\quad I_1 = -\log_2 0{,}2 = \log_2 5 \approx 2{,}32$ бита

пескарь $p_2 = 30/100 = 0{,}3$ $\quad I_2 = -\log_2 0{,}3 \approx \log_2 3{,}33 \approx 1{,}74$ бита

окунь $p_3 = 50/100 = 0{,}5$ $\quad I_3 = -\log_2 0{,}5 = \log_2 2 = 1$ бит

Содержательный подход и вероятность

Задача.
В коробке 50 шаров, из них 40 белых и 10 чёрных. Определить количество информации в сообщении о вытаскивании наугад белого шара и чёрного шара.

Решение.
Вероятность вытаскивания белого шара P1 = 40/50 = 0,8
Вероятность вытаскивания чёрного шара P2 = 10/50 = 0,2

Количество информации о вытаскивании белого шара
I1 = log2(1/0,8) = log21,25 = log1,25/log2 ≈ 0,32 бит

Количество информации о вытаскивании чёрного шара
I2 = log2(1/0,2) = log25 = log5/log2 ≈ 2,32 бит