

## Основные понятия теории вероятности.

### Вероятность в математике

**ВЕРОЯТНОСТЬ** какого - либо случайного события  $P$  определяется как отношение числа случаев  $K$ , соответствующих наступлению ожидаемого события, к общему числу возможных случаев  $N$ , если случаи равновозможны (равновероятны):

$$P = K / N$$

Вероятность выражается в долях единицы. ( $0 \leq p \leq 1$ )

Вероятность достоверного события =1, вероятность невозможного события =0

**Равновероятные события** - это события, которые не имеют преимущества друг перед другом, т.е. вероятность их появления не зависит от условий проведения эксперимента (выпадение орла или решки, выпадение любой из шести граней кубика).

**Задача.**

Шарик находится в одной из трех урн: А, В или С. Какова вероятность сообщения о том, что он находится в урне В.

Решение.

$P=1/3$ , т.к. события равновероятны.

Но не все ситуации имеют одинаковые вероятности реализации. Существует много таких ситуаций, у которых вероятности реализации различаются. Например, если бросают несимметричную монету или "правило бутерброда".

**Неравновероятные события** - это события, вероятность появления которых зависит от условий проведения эксперимента (зависимость прогноза погода от времени года).

**Задача.**

На лугу растут: 2000 ромашек, 8000 васильков и 4000 колокольчиков. Определить вероятность, с которой можно встретить каждый из них на лугу?

Решение.

$$P_1=2000/14000=1/7$$

$$P_2=8000/14000=4/7$$

$$P_3=4000/14000=2/7$$

### Вероятность в информатике.

#### Количество информации.

#### Формулы Хартли и Шеннона

**Содержательный (вероятностный) подход к определению количества информации**

Если заключённые в каком-то сообщении сведения являются для человека новыми, понятными, пополняют его знания, т.е. приводят к уменьшению неопределённости знаний, то сообщение содержит информацию.

**1 бит – количество информации, которое содержится в сообщении, которое уменьшает неопределённость знаний в 2 раза.**

**Пример:**

При бросании монеты возможны 2 события (случая) – монета упадёт орлом или решкой,

причём оба события равновероятны (при большом количестве бросаний количество случаев падения монеты орлом и решкой одинаковы). После получения сообщения о результате падения монеты неопределённость знаний уменьшилась в 2 раза, и, поэтому, количество информации, полученное при этом равно 1 бит.

**!!! Содержательный (вероятностный) подход является субъективным, т.к. одну и ту же информацию разные люди могут оценивать по-разному.** Для одного человека сведения в сообщении могут быть важными и понятными, для другого бесполезными, непонятными или вредными.

### **Вычисление количества информации для равновероятных событий.**

В 1928 г. американский инженер Р. Хартли предложил научный подход к оценке сообщений. Предложенная им формула имела следующий вид:

$$I = \log_2 K$$

$K$  - количество равновероятных событий;

$I$  - количество бит в сообщении, что любое из  $K$  событий произошло.

Тогда  $K = 2^I$ .

Иногда формулу Хартли записывают так:

$$I = \log_2 K = \log_2 (1/p) = -\log_2 p,$$

т. к. каждое из  $K$  событий имеет равновероятный исход  $p = 1 / K$ , то  $K = 1 / p$ .

**Задача 1.** В коробке 32 карандаша, все карандаши разного цвета. Наугад вытащили красный. Какое количество информации при этом было получено?

**Задача 2.** Шарик находится в одной из трех урн: А, В или С. Определить сколько бит информации содержит сообщение о том, что он находится в урне В.

### **А если события происходят с различной вероятностью?**

В 1948 г. американский инженер и математик К Шеннон предложил формулу для вычисления количества информации для событий с различными вероятностями.

Если  $I$  - количество информации,

$K$  - количество возможных событий,

$p_i$  - вероятности отдельных событий,

то количество информации для событий с различными вероятностями можно определить по формуле:

$$I = - \sum p_i \log_2 p_i,$$

где  $i$  принимает значения от 1 до  $K$ .

Формулу Хартли теперь можно рассматривать как частный случай формулы Шеннона:

$$I = - \sum 1 / K \log_2 (1 / K) = I = \log_2 K.$$

**!!! При равновероятных событиях получаемое количество информации максимально.**

**Задача 3.** Определить количество информации, получаемое при реализации одного из событий, если бросают

а) несимметричную четырехгранную пирамидку;

б) симметричную и однородную четырехгранную пирамидку.

**Задача 4.** Вероятность первого события составляет 0,5, а второго и третьего 0,25. Какое количество информации мы получим после реализации одного из них?

**Задача 5.** Какое количество информации будет получено при игре в рулетку с 32-мя секторами?

**Задача 6.** Сколько различных чисел можно закодировать с помощью 8 бит?

### **Алфавитный подход к измерению информации**

При передаче и хранении информации с помощью различных технических устройств информацию следует рассматривать как последовательность знаков (цифр, букв, кодов цветов точек изображения), не рассматривая ее содержание.

Считая, что **алфавит (набор символов знаковой системы) - это событие, то появление одного из символов в сообщении можно рассматривать как одно из состояний события.** Если появление символов равновероятно, то можно рассчитать, сколько бит информации несет каждый символ.

**Информационная емкость знаков определяется их количеством в алфавите.** Чем из большего количества символов состоит алфавит, тем большее количество информации несет один знак. **Полное число символов алфавита принято называть мощностью алфавита.**

$$K=2^I$$

**I – информационная ёмкость знаков алфавита**

**K – мощность алфавита**

**Пример:** Молекулы ДНК (дезоксирибонуклеиновой кислоты) состоят из четырех различных составляющих (нуклеотидов), которые образуют генетический алфавит. Информационная емкость знака этого алфавита составляет:

$$4 = 2^I, \text{ т.е. } I = 2 \text{ бит.}$$

Каждая буква русского алфавита (если считать, что е=е) несет информацию 5 бит ( $32 = 2^5$ ).

**Количество информации, которое содержит сообщение, закодированное с помощью знаковой системы, равно количеству информации, которое несет один знак, умноженному на число знаков в сообщении.**

$$I=i*n,$$

где n – количество знаков в сообщении,

i - количеству информации, которое несет один знак