**Гончаренко Валерий Евстафиевич**
*Кандидат технических наук,*
*доцент Ивановского филиала*
*РЭУ им. Г.В. Плеханова, г. Иваново.*

# Путешествие по тропе первопроходцев измерения информации

## Часть 1

В данной статье автор предлагает читателям заново разработать вероятностно-статистический подход к измерению информации. Стиль изложения материала делает эту задачу непринуждённой и увлекательной. При желании читатели могут обратиться к более строгому изложению затронутых вопросов в ранее опубликованных статьях, указанных в списке использованных источников.

Современные люди часто сталкиваются с понятиями информации и её количества. Чаще всего в качестве единицы измерения используются байт (Б) и производные от него более крупные единицы, например гигабайт (ГБ), и даже такие огромные, как зеттабайт (ЗБ). В мире за двое суток производится примерно один зеттабайт данных ($1ЗБ = 2^{70}Б$).

Единицы измерения бит и байт разработаны в рамках вероятностно-статистического подхода к измерению информации, который получил широкое использование в информатике. Существует и целый ряд других подходов, от рассмотрения которых в рамках данной статьи мы воздержимся.

Предлагаю читателям стать на начало виртуальной тропы и заново прийти к открытию понятия информации и её измерения в рамках вероятностно-статистического подхода, тем более, что для некоторых понятий и сегодня используется различное толкование.

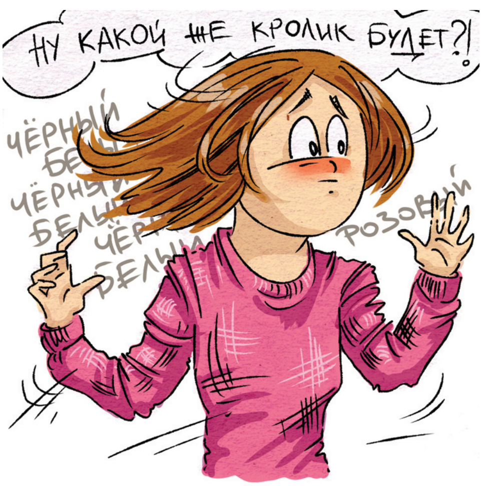
### Основные понятия и первые успехи

Итак, начинаем! Раздаётся громкий голос зазывалы: «Уважаемая публика! Кто обладает даром предвидения, может заработать деньги. Вам нужно угадать, какого из двух практически одинаковых кроликов факир наугад вытащит из мешка – белого или чёрного». Я как гид в этом виртуальном путешествии предлагаю вам воспользоваться предложением зазывалы. Ведь какое бы вы ни выбрали предвидение – белый или чёрный кролик – вероятность правильного решения 50%. А если вы в этом путешествии вдвоём с товарищем, то выигрыш вам гарантирован: достаточно одному сказать – чёрный, а другому – белый.

Давайте теперь со мной проведём логический анализ ситуации. Сделаем это сначала на таких понятиях, как случайное событие и его исходы. В данном примере случайным событием является цвет кролика, который будет наугад извлечён из мешка. И у этого события могут быть только два исхода – белый или чёрный цвет кролика. Важно подчеркнуть, что никаких других исходов быть не может, и мы это назовём полной группой исходов. При извлечении из мешка одного из кроликов нет никаких предпочтений в выборе белого или чёрного, т. е. их выбор равновероятен. Как количественно оценить эту вероятность? Существуют всего два вероятных исхода по одному из двух возможных, поэтому если возьмём отношение $p$ = 0,5 (50%), то это и будет количественной оценкой для двух исходов. Введём обозначения $p_ч$ – вероятность извлечения чёрного кролика и $p_б$ – вероятность извлечения белого кролика. В итоге можно прийти к такому формальному описанию:

$$p_ч = m_ч/N \text{ и } p_б = m_б/N,$$

где $m_ч$, $m_б$ – количество возможных исходов соответственно для чёрного и белого кроликов в полной группе исходов, а $N$ – общее количество исходов.




Вернёмся к аттракциону. Факир достаёт из мешка одного из кроликов, и вся публика восторженно ахает – кролик оказался чёрный, а вы с товарищем с улыбкой лёгкой победы забираете свой трофей. Порассуждаем, почему победа оказалась такой лёгкой. Да потому что предвидеть исход было легко, неопределённости в нём было мало. В любом случае цвет кролика оказался бы чёрным или белым. И если не сомневаться, как буриданов осёл, какую охапку сена начинать есть – справа или слева – и в результате умереть с голоду, то команда двух товарищей всегда останется в выигрыше.

Воодушевлённые, двигаемся дальше по тропе, которая выводит на площадь к другому зазывале, обращающемуся к почтенной публике: «Уважаемые граждане виртуального мира! У факира в крутящемся барабане перекатываются 64 практически одинаковых шара, которые отличаются только цветом. Таблица цветов и оттенков – у вас перед глазами, в каждый цвет окрашен только один шар. Возьмите карточку и напишите на ней цвет шара, который факир наугад вытащит из барабана. Кто угадает, получит выигрыш – тысячу виртуальных гульденов!

– Ну что, два товарища, сыграем? Я согласен быть третьим в вашей компании, но похоже, что шансов на выигрыш у нас мало. Какие бы мы ни выбрали три различных цвета, других возможных цветов остаётся намного больше.

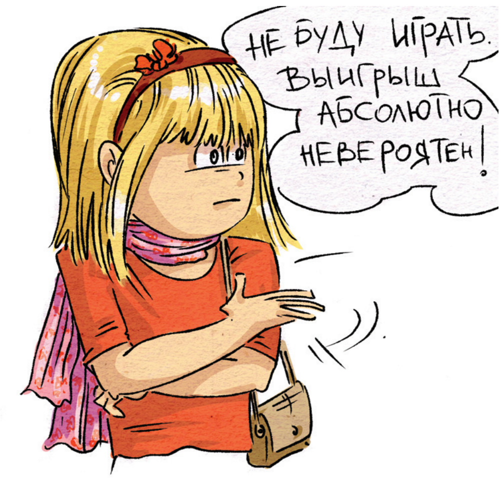
По аналогии с кроликами мы имеем случайное событие – цвет наугад извлекаемого из барабана шара, и всего возможно 64 различных исхода (полная группа исходов). Вероятность извлечения из барабана шара одного из возможных цветов будет $p_{ц} = m_{ц}/N = 1/64$, или 1,56%.

Если мы на трёх карточках напишем по одному цвету, например чёрный, белый, красный, то вероятность того, что мы угадаем, будет

$$p_{чбк} = 3/64, \text{ или } 4{,}7\%,$$

что намного меньше, чем в ситуации с кроликами, а если играть в одиночку, то и того меньше. Неопределённость в том, какой именно исход будет реализован в этом примере, больше.

Мы уже дважды упоминали неопределённость, может, её тоже оценить количественно? Для этого рассмотрим следующую игру – «фальшивая монета, или игра в поддавки».




Уважаемая публика! У меня в руках восемь одинаковых с виду золотых монет, но одна из них фальшивая. Я буду помогать вам в её определении только элементарными ответами – «да» или «нет», а вы будете задавать элементарные вопросы: фальшивая монета в правой (левой) руке? Сейчас у меня в каждой руке по 4 монеты.

Запишем последовательность вопросов и ответов.

1. Вопрос: фальшивая монета в левой руке? Ответ: нет.

Просим факира отложить в сторону монеты из левой руки (ведь там нет фальшивой), а остальные монеты взять по две в каждую руку.

2. Вопрос: фальшивая монета в левой руке? Ответ: да.

Просим факира отложить в сторону монеты из правой руки, а оставшиеся две монеты поделить по одной в каждую руку.

3. Вопрос: фальшивая монета в левой руке? Ответ: нет.

Полученный ответ позволяет сделать вывод, что фальшивая монета обнаружена и находится в правой руке факира.

Как мы видим, путём половинного деления за счёт элементарных вопросов и ответов полностью снята неопределённость в определении одной фальшивой монеты среди восьми одинаковых на вид монет. Как формально описать эти количественные пропорции с учётом деления на два? Имея небольшое представление о математических функциях, несложно догадаться и взять на вооружение функцию двоичного логарифма. В качестве аргумента функции будет общее количество равновероятных исходов случайного события, а возвращать функция будет количество элементарных ответов. Тогда получим $\log_2 8 = 3$.

Займёмся теперь названием размерности полученного значения. Обозначим ответ «да» цифрой 1, а ответ «нет» – цифрой 0. Других цифр не надо, как в двоичной системе счисления. Если воспользоваться английским языком, по праву названным языком международного общения, то по-английски двоичная цифра или двоичное число будут прописаны соответственно как *binary digit* и *binary unit*.

Это достаточно громоздко для обозначения единицы измерения, и мы поступим очень просто, взяв из этих слов только начало одного и конец другого, и получим *bit*. Ура, на нашей виртуальной тропе мы создали новую единицу измерения неопределённости – её Величество *bit* (бит). Становится очевидным, что неопределённость и количество информации должны измеряться одной и той же единицей измерения, ведь каждый ответ факира «да» или «нет» приносил нам информацию, благодаря которой мы и определили фальшивую монету, и всего в рассмотренном примере потребовалось три бита информации.

Очень важно отметить, что теперь, даже не получив ни одного ответа, мы можем оценивать неопределённость случайного события, а следовательно, и объём необходимой информации, которая полностью уничтожит эту неопределённость. Сообщение о реализации любого исхода уничтожает всю неопределённость.

Теперь легко рассчитать неопределённость с 64 шарами разных цветов. Обозначив её через $H$, получим

$$H = \log_2 64 = 6 \text{ бит.}$$

В этом случае при использовании процедуры половинного деления потребовалось бы задать шесть элементарных вопросов и получить шесть элементарных ответов, чтобы выяснить, какого цвета шар в текущем сеансе (опыте) факир достал из барабана. Если обозначить через $I$ количество информации, то получим

$$I = H = \log_2 64 = 6 \text{ бит.}$$

Оглянувшись назад, можно заключить, что тропа не такая уж тяжёлая, и, взяв на вооружение простые логические рассуждения и элементарные знания математики, мы уже научились количественно оценивать неопределённость и количество информации по формуле

$$I = H = \log_2 N. \qquad (1)$$

Теперь пришло время уточнить, для каких случаев применима наша методика:

1. Рассматривается только случайное событие с заранее известной совокупностью исходов (полная группа исходов).
2. Рассматривается только случайное событие с равновероятными исходами.

Но ведь так бывает не всегда со случайными событиями, но и наша виртуальная тропа на этом не заканчивается. Полученные результаты принесли нам мировую известность в виртуальном мире. Впереди предстоит самый трудный участок тропы, поэтому имеет смысл сделать привал и набраться сил для предстоящего крутого восхождения в следующем номере журнала.

## Литература

1. *Гончаренко В.Е.* Проблемы трактовки меры Шеннона в курсе информатики // Информатика и образование. – 2013. – №7.
2. *Гончаренко В.Е.* Сложные задачи на определение информации // Информатика. – 2011. – №10.
3. *Гончаренко В.Е.* Что измеряют по формулам Хартли и Шеннона // Информатика. – 2011. – №6.
4. *Шеннон К.* Математическая теория связи // Работы по теории информации и кибернетике / пер. с англ. ; под ред. Р.Л. Добрушина и О.Б. Лупанова. – М., 1963.