Василевский А. В.

# Теория связи

Статистическая теория связи. Кодирование. Помехозащитное кодирование.

[4] # 1

01/04/2016

# Оглавление

# От автора

Цель данной статьи — максимально кратко и быстро ввести читателя в теорию информации и кодирования (теорию связи). Сама статья есть не что иное как попытка разобраться в основах ТИиК и систематизировать их для построения базы, на которой можно будет выстраивать дальнейшие рассуждения на тему теории информации.

Автор не преследует целью полное изложение ТИиК. Напротив, многие вещи, не существенные для понимания наиболее общих результатов теории, здесь опускаются. Опускаются и доказательства многих теорем, математические выкладки. Все они могут быть найдены в одном из приведенных в списке литературы источниках.

Первое издание охватывает лишь теорию информацию как таковую, не касаясь теории кодирования. Это связано с необходимостью обработки большого количества информации из разных источников и с требованием наискорейшей публикации полученных результатов. Второе издание будет дополнено указанным недостающим разделом, в той или иной мере законченным.

# *ГЛАВА 1*

# Введение в теорию связи

## § 1.1 Системы передачи сообщений

Предметом изучения теории связи являются *системы связи* (системы передачи сообщений), т.е. системы типа *источник* $\leftrightarrow$ *приемник*. Такие системы могут иметь совершенно различную физическую природу. Однако, при всем их многообразии, во всех можно выделить базовые элементы, подведя под общую модель. Займемся ее построением.

Будем исходить из весьма общих предположений. Интуитивно можно ввести следующее нестрогое

**Определение 1.1. Сообщение** — *некоторая информация*, передаваемая от источника к приемнику.

Для передачи сообщения от источника к приемнику требуется среда передачи. В теории информации природа такой среды не имеет никакого значения. Значение имеют лишь только *статистические* характеристики среды, которые будут установлены позже, а сейчас водится

**Определение 1.2. Канал** — среда распространения (передачи) сообщений.

Канал в общем случае может иметь несколько входов и выходов. Мы же будем рассматривать исключительно каналы с одним входом и одним выходом.

Таким образом, система связи (в наиболее общем понимании) есть совокупность источника, приемника и канала.

В данной главе будут рассмотрены наиболее интуитивно понятные свойства систем передачи сообщений. В следующих главах этим свойствам будет дано математическое объяснение.

### 1.1.1 Сообщения

До сих пор мы не конкретизировали понятие сообщения, введенного выше, определив его как *некоторую информацию*. Неопределенность здесь кроется главным образом в понятии *информации*. Значит наша дальнейшая задача состоит в том, чтобы определиться, что же считать информацией.

Мы привыкли понимать под информацией некоторые *полезные* сведения (о мире). Так, фраза “Мамонты — вымершие животные” является информативной, в то время как фраза “Мамонты — это мамонты” является тавтологией и не несет в себе никакой информации. Такое понимание информации основывается на ее смысловом, т.е. *семантическом*, содержании.

Теория информации же отвлекается от всякой семантики, считая информативными *любые* сообщения.

### 1.1.2 Источники сообщений

В нестрогой формулировке источник сообщение можно ввести так:

**Определение 1.3. Источником сообщений** называется всякая система, имеющая один или несколько выходов, на которые выдаются производимые ей сообщения (выходной сигнал).

В качестве базовой характеристики источника сообщений вводят

**Определение 1.4. Алфавит** (источника) — все множество символов, которые источник способен произвести.

Источники сообщений по их алфавиту делят на *дискретные*, чьи алфавиты — конечные счетные множества, и непрерывные, алфавиты которых непрерывны (множества мощности континуума). Сообщения, производимые дискретными источниками, называют *дискретными*, а непрерывными, соответственно, — *непрерывными*.

Известно, что между дискретными и непрерывными множествами есть некоторая связь. Во-первых, непрерывные множества можно рассматривать как предельный случай дискретных, когда количество элементов множества стремится к бесконечности, а разница между ближайшими элементами — к нулю (хотя понятие *разницы* не всегда может быть четко определено для двух произвольных элементов, мы не будем заострять на этом внимание). Во-вторых, непрерывные множества можно *дискретизировать*, сделав их вполне конечными, а дискретные, при необходимости, восстановить до непрерывных (интерполировать).[1] Поэтому мы будем рассматривать исключительно дискретные системы передачи сообщений. Обобщение всех выводов на случай непрерывных систем не составляет особой сложности и может быть найдено, например, в [1] или [2].

Источники также делят на источники *с непрерывным временем*, когда символы алфавита появляются на выходе источника непрерывно, и источники *с дискретным временем*, символы на выходе которого появляются в определенные дискретные моменты времени.

Источники *без памяти* производят символы алфавита (сообщения) независимо от ранее произведенных символов. Выходной сигнал источников *с конечной памятью* зависит от некоторого конечного числа ранее произведенных символов. Предельным случаем источника с конечной памятью является источник *с бесконечной памятью*, чей выходной сигнал зависит от *всех* ранее произведенных символов.

В дальнейшем мы еще раз вернемся к обсуждению характеристик источников сообщений и рассмотрим их более подробно.

### 1.1.3 Кодирование сообщений

Естественным образом встает вопрос о способах представления информации для передачи ее по каналу, т.е. о преобразовании произвольного сообщения в сообщение, состоящее из символов алфавита источника. Вводится

**Определение 1.5. Кодирование** — процесс преобразования произвольного сообщения в сообщение, состоящее из символов алфавита источника. Устройство, производящее кодирование, называется **кодером**, *декодирование*, т.е. обратное преобразование, — **декодером**. Кодер находится на стороне источника, декодер — на стороне приемника.

[1] О дискретизации и восстановлении непрерывных сообщений (функций) можно прочитать, например, в [1, гл.2, стр.27].

Так, например, для передачи человеческого голоса по цифровому каналу его, как известно, необходимо из непрерывного *аналогового* сигнала преобразовать в дискретный *цифровой*, т.е. *дискретизировать*. Для передачи текста по двоичному каналу связи каждой букве сопоставляют некоторую последовательность двоичных символов, называемую *кодом* этой буквы.

### 1.1.4 Каналы связи

Раньше мы не строили никаких предположений о том, как символы передаются по каналу связи. На самом деле совсем не обязательно, что канал будет передавать сообщение в его первозданном виде. Более того, в любом реальном канале существуют *помехи*.[2] Они могут влиять на правильность передачи сообщений, вызывая

- замену одного символа алфавита другим;
- замену символа алфавита сторонним символом;
- удаление или добавление символа.

В дальнейшем будем считать, что помехи не сильные и не вызывают изменения количества передаваемых символов. В следующих разделах будет показано, что всегда существует способ так закодировать сообщение, чтобы свести влияние помех к минимуму и даже совсем устранить его. Коды, позволяющие *исправлять* ошибки передачи, называются *самовосстанавливающимися*. Их построение — предмет отдельного обсуждения.

Выше говорилось о том, что канал характеризуется некоторыми статистическими характеристиками. Этими характеристики как раз и будут являться вероятности неправильной передачи символов, т.е. степень подверженности канала помехам.

## § 1.2 Задачи теории информации и кодирования

Сформулируем теперь основные задачи, разрешение которых будет являться главной целью теории информации и кодирования.

1. Кодирование сообщения для передачи по каналу;
2. Декодирование сообщения, принятого источником;
3. Принятие мер для минимизации влияния помех в канале на правильность передачи сообщений.

Последняя задача, вообще говоря, уже содержится в числе первых двух, но ввиду ее высокой важности мы выделяем ее как отдельную.

Для решения поставленных задач мы должны как минимум сформировать некоторые статистические оценки, теоремы, условия, которые хотя бы принципиально разрешали бы возможность (частично или полностью) правильной передачи сообщений по каналу с шумом. Получением этих условий мы и займемся в следующих главах.

[2] Считается, что помехи возникают исключительно в канале.

# *ГЛАВА 2*

# Теория информации

## § 2.1 Общие слова

В предыдущей главе было установлено, что для построения теории информации необходимо сформулировать не только *качественные*, но и *количественные* характеристики элементов системы связи. В этой главе мы как раз и займемся введением этих характеристик.

Пусть дан источник[1] с алфавитом $Z$ и вероятностью появления $P(z_i)$ того или иного символа на выходе источника, канал с шумом и приемник. Будем как и раньше предполагать, что шум в канале не сильный и не вызывает изменение длины сообщения. На выходе канала в силу помех (см. разд. 1.1.4) приемник в общем случае примет уже не символы алфавита источника $Z$, а некоторые другие символы $V$, притом может оказаться, что $|V| \neq |Z|$, т.е. длины исходного и искаженного алфавита не совпадут.

Символы алфавита $V$ тоже можно охарактеризовать вероятностями появления $P(v_j)$.

Условную вероятность $P(z_i|v_j)$ тогда можно трактовать так: при фактически принятом (нам уже известном) $v_j$ передавался (именно) $z_i$.

Совместная вероятность $P(z_i, v_j)$ тогда будет вероятностью того, что при принятии (именно) $v_j$ передавался (именно) $z_i$.

Следует обратить внимание читателя на глубокое смысловое различие вероятностей $P(z_i|v_j)$ и $P(z_i, v_j)$. В то время, когда условная вероятность $P(z_i|v_j)$ "фиксирует" только $v_j$, совместная вероятность $P(z_i, v_j)$ фиксирует как $v_j$, так и $z_i$. Условная вероятность как бы расчитана относительно $v_j$, для конкретного $v_j$, при известном $v_j$. Или иначе. Условная вероятность отвечает на вопрос: "Какова вероятность (при принятом $v_j$) события 'передан $z_i$' ", а совместная — на вопрос: "Какова вероятность события 'передан $z_i$, принят $v_j$' ".

Введенные обозначения мы будем использовать на всем протяжении статьи.

В качестве заключительных слов к пункту дадим более строгие обозначения, которые тоже будем использовать по мере изложения материала. Итак, вводится

**Определение 2.1. Вероятностным ансамблем** (событий) $\mathcal{Z}$ назовем множество событий (символов, элементов ансамбля) $Z = \{z_1, z_2, \ldots, z_n\}$ с заданной для них вероятностной мерой $P(z_i)$: $\mathcal{Z} = (Z, P)$.

[1] Будем рассматривать источники сообщений без памяти с постоянным во времени распределением вероятности появления символов на выходе. Обобщение на случай источников с памятью не представляется сложным и предоставляется читателю. Здесь скажем лишь то, что вероятность появления символов на выходе источника будет представляться уже *условной* вероятностью $P(z_i \mid z^{(m)}, z^{(m-1)}, \ldots, z^{(1)})$, где $z^{(j)}$ — $j$-й ранее произведенный источником символ, а $m$ — объем памяти источника.

## § 2.2 Измерение информации

В своей книге [2] Р. Фано начинает с введения понятия *взаимной информации*, из которого получаются все остальные соотношения теории информации. Мы же поступим по-другому, так, как это сделано в [1, гл.3 стр.124].

Введем необходимые понятия. Практика показывает, что оперирование вероятностями в чистом виде оказывается неудобным. Поэтому поступают иначе. Дается

**Определение 2.2. Энтропией (неопределенностью) события (символа)** $z \in \mathcal{Z}$, **количеством информации**, *необходимым для идентификации* **события** $z$ вероятностного ансамбля $\mathcal{Z}$, называется логарифм[2] величины, обратной его вероятности:

$$H(z) = \log \frac{1}{P(z)} = -\log P(z). \tag{2.1}$$

Стоит отметить, почему эта величина нам *удобна*, и одновременно раскрыть ее смысловое значение. Рассмотрим сначала вероятность события $z$: $P(z)$. Для нее по определению выполняется неравенство: $0 \leq P(z) \leq 1$. Если вероятность $P(z)$ события стремится к единице, то это означает, что мы *знаем* о нем (почти) *все*. Другими словами, нам не нужна (почти) никакая дополнительная информация, чтобы идентифицировать его (указать, что следующее $n$-ное событие — скорее всего именно $z$). Если же $P(z)$ стремится к нулю, то для идентификации события нам нужно количество информации, стремящееся к бесконечности. Отсюда видно, что мера (выр. 2.1) удовлетворяет указанным требованиям: равна нулю при $P(x) = 1$ и равна бесконечности при $P(x) = 0$. Поэтому опр. 2.2 является именно определением *неопределенности выбора*.

Теперь рассмотрим процесс передачи символа по каналу связи. Пусть, как и раньше, передается символ $z_i$ с априорной вероятностью появления на входе канала (т.е. на выходе источника) $P(z_i)$, а принимается символ $v_j$. После принятия символа $v_j$ символ $z_i$ будет характеризоваться апостериорной вероятностью $P(z_i|v_j)$.

По введенным ранее обозначениям $H(z_i)$ означает неопределенность выбора элемента $z_i$ *до* принятия символа $v_j$, тогда как $H(z_i|v_j) = -\log P(z_i|v_j)$ будет характеризовать уже неопределенность выбора $z_i$ *после* принятия символа $v_j$. Отсюда можно заключить следующее: неопределенность, *снятая* в результате акта приема-передачи, будет равняться разности между априорной и апостериорной неопределенностями. Потому дается

**Определение 2.3. Взаимной информацией** двух элементов $z_i \in \mathcal{Z}$ и $v_j \in \mathcal{V}$ статистически связанных вероятностных ансамблей $\mathcal{Z}$ и $\mathcal{V}$ называется убыль неопределенности выбора элемента $z_i$ в результате того, что становится известным элемент $v_j$:

$$I(z_i; v_j) = -\Delta H = H(z_i) - H(z_i|v_j). \tag{2.2}$$

Из элементарных свойств функции логарифма и из простейших следствий теории вероятностей (теоремы Байеса, см. выр. А.5) получим соотношения:

$$I(z_i; v_j) = \log \frac{P(z_i|v_j)}{P(z_i)} = \log \frac{P(z_i, v_j)}{P(z_i)P(v_j)} = \log \frac{P(v_j|z_i)}{P(v_j)} = I(v_j; z_i). \tag{2.3}$$

Отсюда также видно, что взаимная информация симметрична относительно $z_i$ и $v_j$: сколько информации $v_j$ содержит о $z_i$, столько же и $z_i$ содержит о $v_j$.

[2]Основание логарифма влияет лишь только на размерность (масштаб) единиц измерения. Если логарифм натуральный, то информация измеряется в *натах*, если двоичный — в *битах*, троичный — в *тритах*, десятичный — в *дитах* (*Хартли*). Наибольшее распространение получили биты, поскольку современная компьютерная техника двоична. Руководствуясь сказанным, условимся здесь и далее опускать знак основания логарифма.

## § 2.3 Средняя энтропия

Понятие энтропии события часто оказывается неудобным, поскольку зависит от *конкретных элементов* ансамбля. Для характеристики больших ансамблей, однако, достаточно знать некоторое *среднее* количество информации, приходящееся на событие. Поэтому вводится

**Определение 2.4. Средней энтропией**, **энтропией ансамбля**, просто **энтропией**, **априорной**, **частной (собственной средней) энтропией**, **неопределенностью выбора** элемента из вероятностного ансамбля называется усредненная по ансамблю энтропия события, приходящаяся на одно событие:

$$H(Z) = M[H(z)], \tag{2.4}$$

где $M[I(z)]$ есть обозначение для среднего по вероятности (математического ожидания, см. выр. А.9).

Точно также вводится понятие **условной** (**апостериорной**) энтропии:

$$H(Z|V) = M[H(z|v)]. \tag{2.5}$$

Непосредственно из определения взаимной информации (выр. 2.2), среднего значения и энтропии вытекает соотношение:

$$I(Z;V) = M[I(z;v)] \overset{\text{выр. 2.2}}{=} M[H(z) - H(z|v)] = H(Z) - H(Z|V). \tag{2.6}$$

Можно также показать следующее (усреднением апостериорной энтропии события, предварительно применив к ней формулу Байеса (выр. А.5)):

$$H(Z|V) = H(Z,V) - H(V), \tag{2.7}$$

где $H(Z,V) \equiv M[H(z,v)] = M[-\log P(z,v)]$ — энтропия совместного распределения (совместная энтропия, энтропия объединения) — усредненный логарифм совместной вероятности.

Тогда после объединения выр. 2.6 с выр. 2.7 получим:

$$I(Z;V) = H(Z) - H(Z|V) = H(Z) + H(V) - H(Z,V). \tag{2.8}$$

В отличие от энтропии события, которая является мерой неопределенности конкретного события (случится оно или не случится скорее всего в данный момент), энтропия ансамбля позволяет судить о неопределенности всего ансамбля. Так, если в ансамбле есть элемент с вероятностью, равной единице (остальные элементы, соответственно, имеют нулевую вероятность), то энтропия ансамбля равна нулю — нет неопределенности в том, какое событие из ансамбля произойдет скорее всего в данный момент. Если же все события ансамбля равновероятны, то, напротив, неопределенность выбора максимальна — неизвестно, какое событие произойдет в данный момент времени скорее всего.

### 2.3.1 Свойства средней энтропии

Среди свойств энтропии можно выделить следующие:

1. $H(Z)$ — неотрицательная величина;
2. $H(Z)$ — ограниченная функция:
   - $H(Z) = 0$ тогда и только тогда, когда $P(z_k) = 1,\ P(z_{i\neq k}) = 0$: нет никакой неопределенности в выборе элемента из $Z$;
   - $H(Z) = \max$ тогда и только тогда, когда $P(z_i) = 1/N$ $(i \in [1 \dots N])$, т.е. при равномерном распределении (*энтропия Хартли*): неопределенность выбора одного элемента из множества $Z$ максимальна.

Доказательство этих утверждений не является сложной задачей — оно базируется на принципе вариационного исчисления — и предоставляется читателю.

С точки зрения передачи информации сообщение будет тем более *информативным*, чем ближе его энтропия к максимальной. Это прямо следует из свойства 2. Действительно, ситуация, когда $H(Z) \to 0$, практически означает, что появляется один выделенный символ алфавита, вероятность появления которого много выше вероятностей других символов.[3] Ясно, что передавать множество копий *одного и того же* символа не выгодно. Гораздо более выгодно передавать как можно больше *различных* символов. Поэтому вводят

**Определение 2.5. Коэффициент избыточности** $\mu$ сообщения есть величина, показывающая насколько близка энтропия сообщения к максимальной (оптимальной) энтропии:

$$\mu = \frac{H_{max} - H}{H_{max}}, \tag{2.9}$$

где под $H$ и $H_{max}$ понимаются энтропии не *одного символа сообщения*, а *всего сообщения*: $H = n \times H(Z)$, $H_{max} = n' \times H(Z')$. $Z'$ здесь — некоторый другой алфавит, более оптимальный для передачи данного сообщения.

## § 2.4 Вновь о системах передачи сообщений

Вновь вернемся к рассмотрению систем передачи сообщений. Применим к ним знания, которые были получены в предыдущем разделе. В данном разделе мы переформулируем более строго все введенные ранее нестрогие (во многом интуитивные) определения.

### 2.4.1 Источники сообщений

В современной теории информации источник сообщений определяется таким образом:

**Определение 2.6. Источник сообщений** — система с одним или несколькими выходами, выходным сигналом которой являются сообщения, полученные в результате некоторых случайных процессов и составленные из символов алфавита источника.

Поскольку рассмотрение систем с несколькими выходами ничем не отличается от систем с одним выходом, источник часто определяют как систему с одним выходом.

Будем рассматривать дискретные источники (сообщения).

Сообщение теперь вводится так:

**Определение 2.7. Сообщение** — последовательность символов $\xi^{<n>} = \xi^{(n)}\xi^{(n-1)}\dots\xi^{(1)}$ длины $n$, сгенерированная источником, либо полученная приемником. Величина $\xi^{(i)}$ — $i$-й символ сообщения.

Как уже было сказано ранее, источники бывают с памятью и без памяти. Распределение вероятности символов на выходе источника без памяти не зависит от предыдущего состояния источника (т.е. от ранее произведенных символов/сообщений): $P(z_i^{(n+1)} \mid z^{<n>}) = P(z_i^{(n+1)}) \equiv P(z_i)$ (где $z_i$ — $i$-й символ алфавита источника, $z^{<n>}$ — последовательность длины $n$ всех ранее произведенных символов). В источниках с памятью такого соотношения не выполняется, а выполняется $P(z_i^{(n+1)} \mid z^{<n>}) = P(z_i^{(n+1)} \mid z^{<n,m>})$, где $z^{<n,m>}$ — последовательность из $m$ последних символов, произведенных источником, $m$ — объем памяти источника. Зависимости могут быть и более сложными. Наша задача — рассмотреть простейшие случаи и выделить базовые закономерности, потому мы не касаемся источников с памятью вовсе. Однако для дальнейшей классификации эти сведения необходимы.

[3] Можно провести аналогию: сообщение как бы являет собой постоянный во времени "сигнал", на фоне которого изредка возникают "помехи" — другие символы алфавита.

**Определение 2.8.** Источник называется **стационарным**, если распределение вероятности его выходных символов не зависит от времени, или, что тоже самое, от количества ранее произведенных символов:

$$P(z_i^{(n_1+1)} \mid z^{<n_1,k>}) = P(z_i^{(n_2+1)} \mid z^{<n_2,k>}),$$

где $k$ — произвольное количество произведенных ранее символов (не обязательно объем памяти источника).

Свойство стационарности дополняется свойством *эргодичности*, которое в нестрогой формулировке может быть представлено через

**Определение 2.9. Эргодичность** — свойство источника, при котором *одна единственная достаточно долгая реализация* случайного выходного сигнала несет практически всю информацию о статистических характеристиках источника.

Свойство эргодичности позволяет определять вероятности появления символов на выходе источника путем достаточно долгого наблюдения и усреднения одой реализации процесса, вместо усреднения по ансамблю реализаций процесса (т.е. вместо усреднения по бесконечному числу работающих (одновременно) "копий" источника можно провести усреднение по бесконечно долгому сигналу от одного источника).

Классификация на источники дискретные и непрерывные, а также на источники с дискретным и непрерывным временем естественным образом не претерпевает никаких изменений.

### 2.4.2 Каналы связи

Классификация каналов связи, в отличие от источников, не отличается большой разнообразностью.

Существует три базовых разделения каналов: каналы *с памятью* и *без памяти*, *стационарные* и *нестационарные*, *с помехами* и *без помех*. Последний случай — идеализация предпоследнего, т.к. в реальных каналах помехи всегда существуют.

Понятия стационарности и памяти для канала даются также, как и для источника, потому их формулировка не приводится.

Среди каналов с шумом также можно выделить каналы с гауссовым шумом. Их рассмотрение выходит за рамки данной статьи.

# *ПРИЛОЖЕНИЕ А*

# Элементы теории вероятностей

## § А.1 Определения

Здесь мы будем пользоваться классической теорией вероятностей. Современная теория вероятностей формулируется более строго и громоздко. Для понимания материалов, излагаемых в данной статье, достаточно результатов классической теории вероятностей.

Пусть имеем конечное счетное множество $X = \{x_1, x_2, \ldots, x_n\}$.

Пусть в результате некоторого события (опыта) мы получили его результатом величину $\xi \in X$. После проделывания серии опытов мы сможем сказать, что, например, $\xi$ принимало значение $x_1$ больше раз, чем $x_5$.

Назовем такую величину $\xi$ *случайной величиной*. Вводится

**Определение А.1. Случайной величиной** $\xi \in X$ называется такая величина, для которой равенства $\xi = x_i \ \ (x_i \in X, \ i \in [1 \ldots n])$ выполняются с некоторыми своими вероятностями.

Мы еще не вводили понятие вероятности математически. Попробуем это сделать. Ясно, что численное значение вероятности события должно быть тем больше, чем оно вероятнее, т.е. чем чаще оно будет происходить при проведении серии опытов. В частности, вероятность невозможного события (например $\xi \notin X$) должна быть равна нулю. Напротив, вероятность события, которое происходит всегда (например $\xi \in X$), должна быть ненулевой и одинаковой для всех типов таких событий. Теперь мы уже можем ввести

**Определение А.2. Вероятностью события**[1] $\xi = x_i \ \ (x_i \in X)$ называется такая величина $P(\xi = x_i) \equiv P(x_i) \equiv p_i$, которая удовлетворяет следующим условиям:

1. $P(X) \equiv P(\xi \in X) \equiv \sum\limits_{x \in X} P(x) \equiv \sum\limits_{i=1}^{n} p_i = 1$ — условие нормированности на единицу;
2. $P(x_i) \geq 0$ — условие неотрицательности.

Численные значения вероятностей $P(x_i)$ могут быть определены из эксперимента:

$$P(x_i) = \frac{N(x_i)}{\sum\limits_{x \in X} N(x)}, \tag{A.1}$$

[1] В современной (зарубежной) литературе все чаще встречается обозначение $\Pr(x)$, вместо традиционного $P(x)$. Мы будем использовать такое обозначение лишь только в тех случаях, когда традиционное ведет к неоднозначности.

где $N(x_i)$ — количество событий, в которых оказалось $\xi = x_i$. Очевидно, что такое выражение соответствует введенному определению вероятности.

В теории вероятностей на том, откуда известны вероятности $P(x_i)$, внимание не акцентируют, а просто полагают, что они известны *откуда-нибудь*. Значения $P(x_i)$ могут быть получены из эксперимента, теоретической модели или просто заданы.

## § А.2 Сложные события

Простейшим следствием из определения вероятности события через эксперимент (см. выр. А.1) является то, что определить вероятность сложного события, состоящего из других (независимых) событий можно довольно простыми способами.

**Определение А.3.** Пусть даны две (независимые[2]) случайные величины: $\xi \in X$ и $\eta \in Y$, где $X = \{x_1, x_2, \ldots, x_n\}$, $Y = \{y_1, y_2, \ldots, y_m\}$. Пусть также имеем два события: $A = (\xi = x_i)$ и $B = (\eta = y_j)$. **Сложным событием** называется событие $C$, составленное из событий $A$ и $B$.

Все сложные события $C = C(A, B)$ можно рассматривать как события (совокупности событий) на произведении множеств $\Pi = X \times Y$. Тогда случайной величиной $\pi \in \Pi$ будет являться упорядоченная пара случайных величин $\xi$ и $\eta$: $\pi = (\xi, \eta)$. С этой позиции легко доказывать многие соотношения теории вероятностей.

Рассмотрим сначала событие $C = (A \text{ and } B)$, т.е. событие того, что $A$ и $B$ выполняются *одновременно*. Если рассмотреть такое событие на множестве $\Pi$, то легко показать, что

$$P(C) = P(A) \times P(B). \tag{A.2}$$

Точно также для события $C = (A \text{ or } B)$, т.е. событие того, что *хотя бы одно* из событий $A$ или $B$ выполняется, можно показать:

$$P(C) = P(A) + P(B). \tag{A.3}$$

Легко видеть также, что:

$$P(x_i) \equiv P(x_i, Y) = \sum_{y \in Y} P(x_i, y), \tag{A.4a}$$

$$P(y_j) \equiv P(X, y_j) = \sum_{x \in X} P(x, y_j). \tag{A.4b}$$

При этом условие нормированности вероятности, естественно, остается в силе:

$$\sum_{\substack{x \in X \\ y \in Y}} P(x, y) = 1.$$

## § А.3 Статистически связанные события

Теперь предположим, что случайные величины $\xi$ и $\eta$ связаны статистически. Это означает, что значение $\xi$ зависит от значения $\eta$ или, что тоже самое, реализация события $A = (\xi = x_i)$ зависит от реализации события $B = (\eta = y_j)$.

Практически это можно представить на следующих примерах:

[2]Важно, что события именно *независимые*. О статистически связанных событиях будет сказано в следующих разделах данной статьи.

- Две склеенные монетки. Вероятность решки на одной монетке жестко связана с вероятностью решки на другой.
- Ящик с конечным числом разноцветных шаров. При извлечении очередного шара из ящика вероятность вынуть шар того же цвета уменьшается, в то же время увеличивается вероятность вынуть шар любого другого цвета.

Для изучения связанных событий вводят

**Определение А.4. Условная вероятность** $P(x_i|y_j)$ есть вероятность того, что при реализации (выпадении) события $(\eta = y_j)$ реализуется событие $(\xi = x_i)$.

Прямо по определению видно, что $P(x_i, y_j) = P(x_i|y_j) \times P(y_j) = P(y_j|x_i) \times P(x_i)$. Действительно, вероятность того, что реализуется и $x_i$, и $y_j$ есть вероятность того, что реализуется $y_j$, и уже при заданном $y_j$ реализуется $x_i$ (и наоборот). Тогда получаем:

$$P(x_i|y_j) = \frac{P(x_i, y_j)}{P(y_j)}, \tag{A.5}$$

или

$$\frac{P(x_i|y_j)}{P(x_i)} = \frac{P(y_j|x_i)}{P(y_j)} = \frac{P(x_i, y_j)}{P(x_i)P(y_j)}. \tag{A.6}$$

Кроме того, легко видеть, что:

$$\sum_{x\in X} P(x|y_j) = \frac{1}{P(y_j)} \overbrace{\sum_{x\in X} P(x, y_j)}^{P(y_j),\ \text{выр. А.4}} = 1. \tag{A.7}$$

Формулы (выр. А.5) и (выр. А.6) называются формулами Байеса (теоремой Байеса).

## § А.4 Среднее вероятностное

Перейдем теперь к рассмотрению числовых множеств: $X, Y \subset \Re$.

Пусть проводится серия из $N$ экспериментов, в ходе которых получаются $N$ значений $x^{(k)}$. Можно посчитать среднее значение результатов всех $N$ экспериментов:

$$\frac{1}{N}\sum_{k=1}^{N} x^{(k)} = \sum_{i=1}^{n} \frac{N(x_i)}{N} x_i \overset{\text{выр. А.1}}{=} \sum_{i=1}^{n} P(x_i)x_i.$$

Здесь $n$ — количество *различных* результатов опыта, в то время как $N$ — количество *всех* опытов. $N(x_i)$, очевидно, число опытов, в которых результатом получили $x_i$.

Понятие среднего легко обобщается и на функции случайной величины $f(\xi)$: водится

**Определение А.5. Средним вероятностным**, или **математическим ожиданием**, функции $f(\xi)$, где $\xi \in X$ — случайная величина, называется число

$$\overline{f(\xi)} \equiv f(X) \equiv M[f] = \sum_{x\in X} P(x)f(x). \tag{A.8}$$

Аналогично вводится определение среднего для функции двух и более переменных. Пусть имеем случайную величину $\pi = (\xi, \eta) \in \Pi$ на произведении множеств $\Pi = X \times Y$ и функцию $f(\pi)$. Тогда

$$\overline{f(\pi)} \equiv f(\Pi) \equiv M[f] = \sum_{\substack{x\in X \\ y\in Y}} P(x, y)f(x, y). \tag{A.9}$$

Можно вводить также средние квадратичные ($M[f^2]$), средние условно-вероятностные ($M_{y_j}[f]$) и т.д. Все они в купе с математическим ожиданием играют важную роль в теории вероятностей и ее приложениях.

# Литература


1. Панин В.В., *Основы теории информации.* Учебное пособие для вузов, 4-е издание (электронное), Бином, Москва, 2012.

2. Фано Р., *Передача информации. Статистическая теория связи*, Издательство “Мир”, Москва, 1965.