МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное
образовательное учреждение высшего образования
«Юго–Западный государственный университет»
(ЮЗГУ)

А. В. Хмелевская, А. М. Потапенко

ОСНОВЫ ТЕОРИИ ИНФОРМАЦИИ И КОДИРОВАНИЯ

Учебное пособие

*Утверждено учебно–методическим советом университета
Юго–Западного государственного университета*

Курск 2016

УДК 621.391
ББК32.811.4я73
Х651

Рецензенты:

Доктор технических наук, зам. генерального конструктора по инвестиционным проектам АО «Авиаавтоматика» имени В.В. Тарасова», старший научный сотрудник И. Е. Мухин

Доктор технических наук, главный научный сотрудник НИИЦ (г. Курск) 18 ЦНИИ МО РФ В. Г. Довбня

Хмелевская А. В.
Основы теории информации и кодирования: Учеб. пособие. / А. В. Хмелевская, А. М. Потапенко – Курск: Юго–Зап. гос. ун–т, 2016. 241 с.
ISBN

Содержание учебного пособия соответствует требованиям ФГОС ВО по техническим специальностям и направлениям подготовки, в том числе по специальностям 10.05.02 и 10.05.03 и направлениям подготовки 10.03.01 и 11.03.02

Учебное пособие содержит краткие теоретические положения основ теории информации и кодирования, а именно: основы теории информации, основы теории кодирования и передачи информации по каналам связи. Для удобства пользования пособие содержит краткие справочные сведения по теории вероятностей, которые крайне необходимы при рассмотрении основных положений теории информации, теории кодирования и передачи информации по каналам связи.

Предназначено для студентов высших учебных заведений.

УДК 621.391
ББК 32.811.4я73

ISBN

© Юго–западный государственный университет, 2016
© Хмелевская А. В., Потапенко А. М., 2016

ОГЛАВЛЕНИЕ

Предисловие.....6
Введение.....7
Глава 1. Введение в теорию информации 11
1.1. Информация. Общие понятия.....11
1.1.1. Информация и материя.....11
1.1.2. Информация в адаптивной системе 12
1.2. Методологическая схема формирования и материализации информации 13
1.3. Этапы обращения информации 14
1.4. Виды информации.....16
1.5. Структура информации 20
Контрольные вопросы и задания.....23
Глава 2. Измерение информации.....24
2.1 Структурные меры количества информации 24
2.2. Объем информации.....33
2.3. Семантические меры информации.....35
2.3.1. Содержательность информации 36
2.3.2. Целесообразность информации.....37
2.4. Вероятность и информация.....40
2.5 Количественная оценка информации 42
2.5.1 Понятие энтропии 42
2.5.2 Энтропия ансамбля 45
2.5.3. Свойства энтропии.....48
2.5.4. Избыточность источника.....51
2.5.5. Производительность источника 52
2.6 Сложная энтропия.....54
2.6.1 Энтропия объединения.....54
2.6.2. Условная энтропия.....60
2.6.3. Основные свойства сложной энтропии 66
2.6.4. Совместная энтропия двух источников.....67
2.7. Энтропия непрерывных сообщений.....70
2.7.1. Экстремальные свойства энтропии непрерывных сообщений 75
2.7.2. Эпсилон–энтропия источника непрерывных сообщений.....80

2.7.3. Энтропия дискретного сообщения, получаемого из непрерывного путем его квантования по уровню 88
Контрольные вопросы 89
Контрольные задания.....91
Глава 3. Дискретизация и квантование сигналов 94
3.1. Виды дискретизации (квантования).....94
3.1.1. Критерии точности представления квантованного сигнала...97
3.1.2. Элементы обобщенной спектральной теории сигналов 99
3.2. Дискретизация по времени.....102
3.2.1. Разложение в ряд Котельникова (Теорема Котельникова) ..102
3.2.2. Выбор периода дискретизации (квантования по времени) по критерию наибольшего отклонения.....106
3.2.3. Выбор интервала дискретизации по критерию среднеквадратического отклонения.....116
3.3. Квантование по уровню.....118
3.3.1. Оптимальное квантование по уровню 118
3.3.2. Дисперсия ошибки в случае использования равномерной шкалы квантования по уровню.....123
3.3.3. Расчет неравномерной оптимальной в смысле максимума количества информации в квантованном сигнале шкалы квантования.....129
3.3.4. Закон компандирования при условии равномерного закона распределения квантуемого сигнала.....130
Контрольные вопросы 131
Глава 4. Кодирование информации.....133
4.1. Элементы теории кодирования.....136
4.1.1. Неравенство Крафта.....138
4.1.2. Теорема об обобщении некоторых результатов, полученных для префиксных кодов, на все однозначно декодируемые коды...140
4.1.3. Основная теорема кодирования для канала связи без шума 141
4.1.4. Теорема о минимальной средней длине кодового слова при поблочном кодировании.....144
4.2. Оптимальные неравномерные коды.....145
4.2.1. Коды Хаффмена 151
4.2.2. Коды Шеннона − Фано 153
4.2.3. Параметры эффективности оптимальных кодов 157
4.2.4. Особенности эффективных кодов. 159

4.3. Помехоустойчивое кодирование 160
4.3.1. Классификация кодов 161
4.3.2. Простейшие модели цифровых каналов связи с помехами .163
4.3.3. Расчет вероятности искажения кодового слова в ДСМК 164
4.3.4. Общие принципы использования избыточности.....165
4.3.5. Граница Хэмминга 170
4.3.6. Избыточность помехоустойчивых кодов 171
4.3.7. Математическое введение к алгебраическим кодам 171
4.3.8. Линейные коды.....174
4.3.9. Упрощённый способ построения линейного кода 175
4.3.10. Порядок декодирования 179
4.3.11. Двоичные циклические коды.....184
4.3.12. Некоторые свойства циклических кодов.....189
4.3.13. Построение кода с заданной корректирующей способностью 192
4.3.14. Матричное описание циклических кодов 193
4.3.15. Выбор образующего полинома.....194
Контрольные вопросы 195
Контрольные задания.....197
Глава 5. Передача информации 198
5.1. Виды каналов передачи информации 198
5.2. Разделение каналов связи.....200
5.3. Пропускная способность каналов связи 206
5.3.1. Пропускная способность дискретного канала связи с шумом 209
5.3.2. Типичные последовательности и их свойства 214
5.3.3. Основная теорема Шеннона для дискретного канала с шумом 215
5.3.4. Пропускная способность непрерывного канала при наличии аддитивного шума 220
Контрольные вопросы 225
Контрольные задания.....226
Заключение 227
Рекомендательный список литературы 228
Приложение 230

Предисловие

Целью данного учебного пособия является изучение теоретических основ дисциплин «Теория информации», «Основы теории информации», «Теория информации и кодирования».

Содержание учебного пособия соответствует требованиям Федерального государственного образовательного стандарта высшего образования для студентов специальностей 10.05.02 и 10.05.03 и направления подготовки 10.03.01 и 11.03.02, рабочим программам соответствующих дисциплин и основным задачам их освоения.

Читательский адрес издания – для студентов высших учебных заведений, обучающихся по направлениям подготовки 11.03.02 «Инфокоммуникационные технологии и системы связи» и 10.03.01 «Информационная безопасность», и по специальностям 10.05.02 «Информационная безопасность телекоммуникационных систем» и 10.05.03 «Информационная безопасность автоматизированных систем» очной и заочной форм обучения.

Тип учебного издания и его место в системе других изданий – учебное пособие (теоретическое).

Пособие состоит из пяти теоретических глав и приложения. В первой главе содержатся общие сведения по основам теории информации. Во второй главе рассматриваются вопросы измерения информации, третья глава посвящена краткому изложению вопросов дискретизации и квантования сигналов, в четвёртой главе приведены основы теории кодирования, в пятой главе кратко изложена теория передачи информации по каналам связи.

Изложение теоретических глав написано в предположении, что студенты знакомы с основными понятиями из курса высшей математики.

Приложение содержит краткий справочный материал по основам теории вероятностей, знание которого необходимо при изучении основного теоретического материала, представленного в пособии.

При подготовке пособия использованы теоретические положения, отраженные в научных статьях, в учебной литературе и монографиях.

Введение

Понятие информации – это одно из основных понятий кибернетики. Существует много определений информации:

– в переводе с греческого языка informatikoa (лат. Informatio) означает сообщение, разъяснение, однако в настоящее время показано, что не всякое сообщение имеет информацию;

– философское определение информации – это отражение окружающего мира в человеческом сознании;

– энциклопедическое определение информации – это совокупность сведений, определяющих меру наших знаний об объектах, процессах, событиях, явлениях и т. д.;

– техническое определение информации – это совокупность данных, являющихся объектом получения, преобразования, воспроизведения (отображения) и использования, т. е. информация отождествляется с материей (одной из её форм) и является объектом исследования в научных и практических целях.

Появление теории информации связывают с публикацией книги американского математика Клода Шеннона «Математическая теория связи» в 1948 г. Именно К. Шеннон ввёл понятие информации для описания процессов в каналах связи. Однако ещё в 1928 г. была опубликована статья английского инженера Ричарда Хартли «Передача информации», а в 1946 г. опубликована книга советского учёного академика В. А. Котельникова «Теория потенциальной помехоустойчивости». Тем не менее основоположником теории информации считают К. Шеннона.

К теории информации (в ее узкой классической постановке) относят результаты решения ряда фундаментальных теоретических вопросов, касающихся повышения эффективности функционирования систем связи. Это в первую очередь:

– анализ сигналов как средства передачи сообщений, включающий вопросы оценки переносимого ими «количества информации»;

– анализ информационных характеристик источников сообщений и каналов связи и обоснование принципиальной возможности кодирования и декодирования сообщений, обеспечивающих

предельно допустимую скорость передачи сообщений по каналу связи, как при отсутствии, так и при наличии помех.

Прикладные результаты приводятся здесь только для пояснения основ теории. При более широкой трактовке теории информации результаты рассмотрения указанных вопросов составляют ее основу.

Если расширение связано с приложением теории в технике связи, а именно рассмотрением проблемы разработки конкретных методов и средств кодирования сообщений, то совокупность излагаемых вопросов называют «Теорией информации и кодирования» или «Прикладной теорией информации».

Другая точка зрения состоит в том, что глобальной проблемой теории информации следует считать разработку принципов оптимизации системы связи в целом. В этом случае к ней относят все локальные проблемы систем связи, включая, например, проблему оптимального приема и др.

В соответствии с третьей крайней точкой зрения к компетенции теории информации относят все проблемы и задачи, в формулировку которых входит понятие информации. Ее предметом считают изучение процессов, связанных с получением, передачей, хранением, обработкой и использованием информации. В такой постановке она затрагивает проблемы многих наук (в частности, кибернетики, биологии, психологии, лингвистики, педагогики) на всех трех уровнях (синтаксическом, семантическом и прагматическом).

Попытки широкого использования идей теории информации в различных областях науки связаны с тем, что в основе своей эта теория математическая. Основные ее понятия (энтропия, количество информации, пропускная способность) определяются только через вероятности событий, которым может быть приписано самое различное физическое содержание. Подход к исследованиям в других областях науки с позиций использования основных идей теории информации получил название теоретико–информационного подхода. Его применение в ряде случаев позволило получить новые теоретические результаты и ценные практические рекомендации. Однако нередко такой подход приводит к созданию моделей процессов, далеко не адекватных реальной действительности. Поэтому в любых исследованиях, выходящих за рамки чисто технических

проблем передачи и хранения сообщений, теорией информации следует пользоваться с большой осторожностью. Особенно это касается моделирования умственной деятельности человека, процессов восприятия и обработки им информации.

В настоящем пособии термин «Информация» понимается в узком смысле, принятом при описании так называемых информационных систем.

К ним относятся телекоммуникационные и вычислительные сети, автоматизированные системы управления и контроля и т. п. В данном случае понятие количества информации определяется как частота употребления знаков. Количество информации в указанном смысле не отражает ни семантики, ни прагматической ценности информации.

Информационные системы – это класс технических систем, предназначенных для хранения, передачи и преобразования информации. Соответственно информация – это сведения, являющиеся объектом хранения, передачи и преобразования, а теория информации – раздел кибернетики, занимающийся математическим описанием методов передачи, хранения, извлечения (обработки) и классификации информации. Заметим, что сама информация, как правило, используется для осуществления каких–либо управляющих воздействий.

Таким образом, предметом нашего рассмотрения является теория информации в классическом смысле – решение теоретических вопросов, касающихся повышения эффективности и функционирования информационных систем, в частности систем связи. Она включает в себя:

1) анализ сигналов как средства передачи информации;

2) анализ информационных характеристик источников сообщения и каналов связи;

3) теорию кодирования;

4) методы приема и обработки информации.

Каждый из указанных разделов может быть (и, как правило, является) предметом самостоятельного глубокого изучения в соответствующих дисциплинах различных специальностей информационного направления. В настоящем курсе мы стремились акцентировать внимание на наиболее общих фундаментальных законах, име-

ющих существенное значение для восприятия указанных разделов как единого целого. На наш взгляд, таким общим фундаментом являются теоремы К. Шеннона о кодировании.

Глава 1. Введение в теорию информации

При исследовании или наблюдении любого физического процесса, объекта или явления можно ввести понятие информации. Существует достаточно много определений понятия информация, существенно различающихся в зависимости от сферы деятельности, в применении к которой даётся конкретное определение этого понятия.

Такое положение, при котором существует множество различных определений и отсутствует единое общепринятое и точное определение этого понятия, вытекает из того, что понятие информации, наряду с такими фундаментальными понятиями, как материя и энергия в физике или как множество или точка в математике, является первичным понятием, и поэтому не может иметь четкого формализованного определения, т. е. не может быть определено через более простые известные объекты, имеющие четкие определения.

Поэтому при определении основных фундаментальных первичных понятий используют интуитивный подход, определяя понятие через совокупность присущих ему свойств. Действительно, исходя из практического опыта, интуитивно, можно представить информацию как совокупность содержательных сведений, заключенных в том или ином объекте или явлении, и перечислить ее основные свойства, тем самым интуитивно, исходя из этих свойств, её определить.

1.1. Информация. Общие понятия

1.1.1. Информация и материя

С точки зрения философии *информация* – наиболее общее понятие наряду с материей, можно сказать, что она *отражает организацию материи*. При этом информация не только пассивно характеризует структуру материи, но и *способна активно создавать и воспроизводить эту структуру.* Например, любой созданный человеком объект первоначально существует в виде идеи (образа) в голове его творца, а свойства самого человека в значительной степени запрограммированы информацией, хранящейся в его геноме.

«Антиподом» *информации*, характеризующей стуктурированность материи является *энтропия,* которая *отражает ее неупорядоченность* («хаос»). Соотношение материи, информации и энтропии показано на рисунке 1.1.

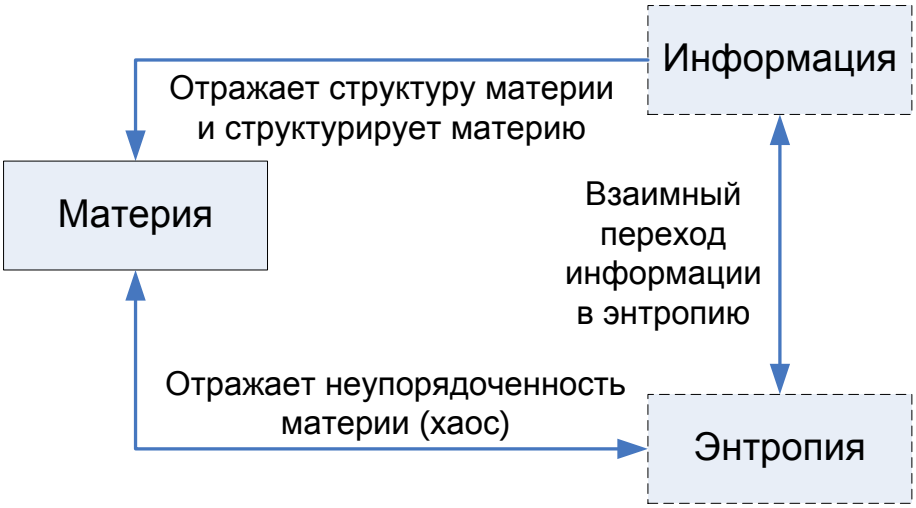


Рис. 1.1. Соотношение материи, информации и энтропии

Упорядоченность и хаос (и соответственно, энтропия и информация) в видимой вселенной непрерывно перетекают друг в друга: например, строить означает упорядочивать, а разрушать – вносить беспорядок. Исходя из этого существует предположение о «законе сохранения» количества информации во Вселенной. Однако есть и прямо противоположная точка зрения: след всего, что происходит, неуничтожим на «тонких уровнях» организации материи, так что информация постоянно накапливается.

1.1.2. Информация в адаптивной системе

Упомянутый выше подход к информации можно считать наиболее общим, однако он мало что дает в практическом плане. *Понятия количества, значения и ценности информации приобретают смысл, если в рассмотрении появляется субъект – система, которая эту информацию использует.* Такой системой может быть, например, живой организм, сообщество людей или компьютер, управляющий некоторым агрегатом.

Система всегда существует в определенной среде. Чтобы решать свои задачи, она должна иметь собственную *модель среды* и постоянно корректировать ее на основе получаемых сведений (адаптировать). В рамках такого подхода *информация есть «представление субъекта об окружающей среде»* (рис. 1.2). Напротив, энтропия – неопределенность в таком представлении. Появление новых сведений *снимает часть неопределенности* и энтропия («незнание») заменяется информацией («знанием»).

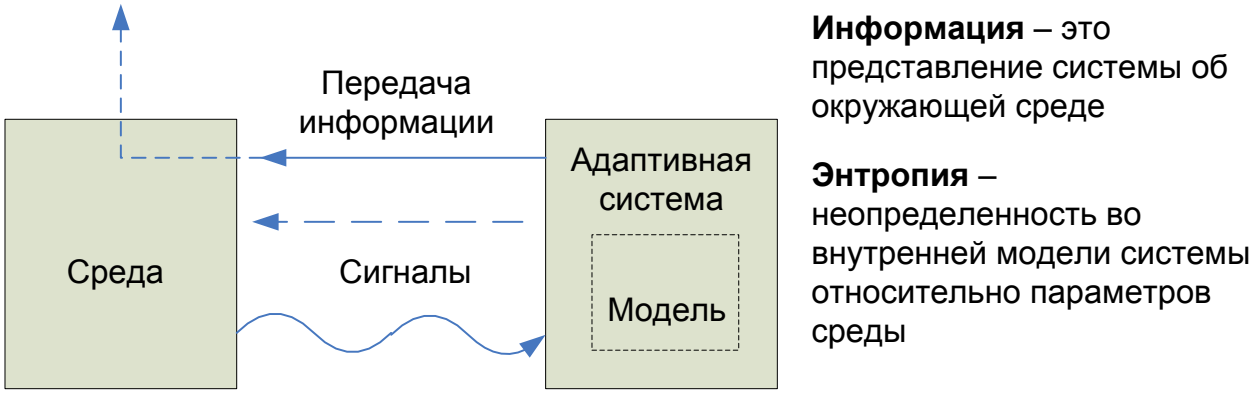


Рис. 1.2. Понятие информации при наличии субъекта.

Обратите внимание, что *информация (как «представления субъекта») сама по себе не материальна, однако она всегда имеет материальные носители: сигналы, которые ее переносят, или параметры элементов, с помощью которых она хранится*. Сигналы служат носителями информации при *восприятии* ее из среды и при передаче от одного субъекта другому (при этом «другой субъект» представляется во «внутренней модели», как элемент среды). Обратите внимание, что из многообразных физических воздействий среды на субъекта *сигналами* будут лишь те, из которых субъект получает информацию.

1.2. Методологическая схема формирования и материализации информации

Сама по себе информация может быть отнесена к области абстрактных категорий, подобных, например, математическим формулам. Однако проявляется она всегда в материаль-

но–энергетической форме в виде сигналов. Методологическая схема образования сигнала показана на рисунке 1.3.

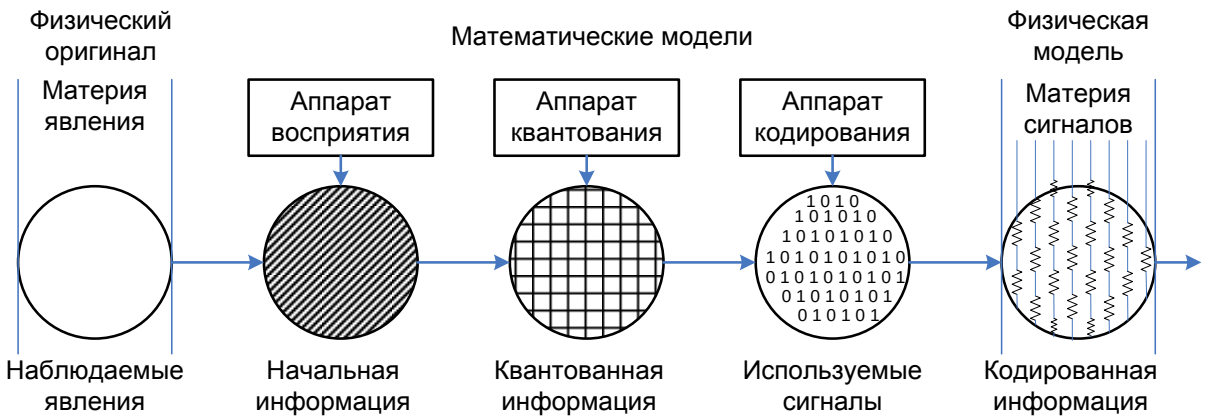


Рис. 1.3. Методологическая схема формирования и материализации информации

С передачей и обработкой информации связаны действия любого автоматического устройства, поведение живого существа, творческая деятельность человека, развитие науки и техники, экономические и социальные преобразования в обществе и сама жизнь.

Как понятие энергии привело к единой и прогрессивной точке зрения на физические явления природы и техники, так и понятие информации объединило инженерно–психологические аспекты современной техники.

Если материал (вещество) и энергия сравнительно полно изучены, то законы получения, преобразования и использования информации еще являются неизведанной областью, таящей в себе много неожиданных проявлений.

1.3. Этапы обращения информации

Так как материальным носителем информации является сигнал, то можно следующим образом схематично представить этапы обращения и преобразования сигналов (рис. 1.4).

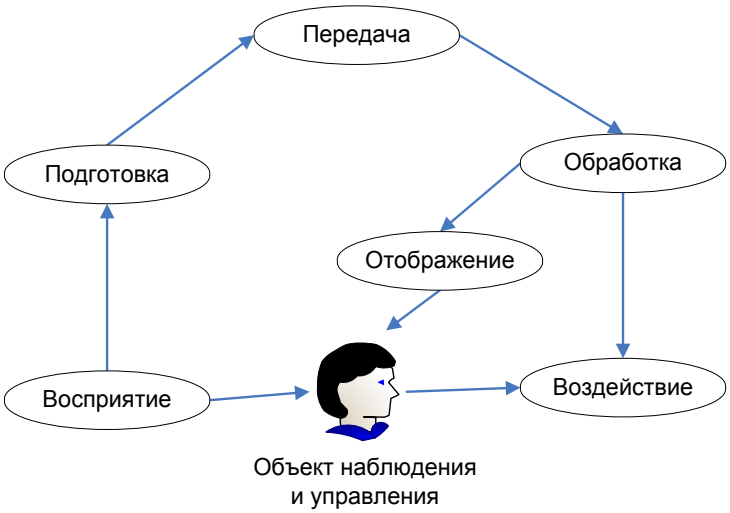


Рис. 1.4. Этапы обращения информации

Восприятие состоит в том, что формируется образ объекта, производятся его опознание и оценка. При восприятии нужно отделить полезную информацию от шумов, что в некоторых случаях (тонкие биологические, физико–химические эксперименты; сложные производственные условия; радиосвязь, локация, астрономия и др.) связано со значительными трудностями. В результате восприятия получается сигнал в форме, удобной для передачи или обработки.

Подготовка информации включает в себя операции ее нормализации, квантования, кодирования, модуляции сигналов и построения моделей.

Передача информации состоит в переносе ее на расстояние посредством сигналов различной физической природы соответственно по механическим, гидравлическим, пневматическим, акустическим, оптическим, электрическим или электромагнитным каналам. Прием информации на другой стороне канала имеет характер вторичного восприятия со свойственными ему операциями борьбы с шумами.

Обработка информации заключается в решении задач, связанных с преобразованием информации, независимо от их функционального назначения. Применение ЭВМ обобщает и централизует функции обработки, имеющие отношение главным образом к моделям ситуаций и принятию решений при управлении. Обработка производится при помощи устройств или машин, осуществляющих аналоговые или цифровые преобразования поступающих величин и

функций. Промежуточным этапом обработки может быть хранение в запоминающих устройствах. Извлечение информации из запоминающих устройств также имеет характер восприятия и связано с борьбой с помехами.

Отображение информации требуется тогда, когда в цикле обращения информации принимает участие человек. Оно заключается в демонстрации перед человеком условных изображений, содержащих качественные и количественные характеристики выходной информации. Для этого используются устройства, способные воздействовать на органы чувств человека.

Из устройства обработки информация может выводиться не только оператору, но и непосредственно воздействовать на объект управления.

Воздействие состоит в том, что сигналы, несущие информацию, производят регулирующие или защитные действия, вызывая изменения в самом объекте.

1.4. Виды информации

Информацию можно различать по областям знаний (биологическая, техническая, экономическая и др.), по физической природе восприятия (зрительная, слуховая, вкусовая и др.), а также по структурно–метрическим свойствам.

Остановимся на последней классификации (табл. 1.1), наиболее пригодной для технических приложений.

Табл. 1.1.

Виды и формы предоставления информации

Вид информации	Обозначение	Формы представления информации		
		топологическая	абстрактная	лингвистическая
Событие	$Ф^0$	Точка	Суждение	Знак
Величина	$Ф^1$	Линия	Понятие	Буква
Функция	$Ф^2$	Поверхность	Образ	Слово
Комплекс	$Ф^3$	Объем	Система	Предложение
. . . . .	. .	. . . . .	. . . . .	. . . . .
Поле	$Ф^n$	Пространство	Универсум	Фонд

К *топологической информации* относятся геометрические образы, карты местности, различные плоские изображения и объемные объекты. *Топологической информацией* удобно выражать образы и ситуации, подлежащие распознаванию.

Абстрактную информацию применяют в исследованиях на высоком теоретическом уровне, когда нужны отвлечения, обобщения и символизация.

В таблице 1.1 виды информации можно разделить по размерности информационных множеств. Назовем:

– *информацией нулевого порядка* (нуль–мерной информацией) такую, которая соответствует мощности точки,

– *первого порядка* (одномерная информация) – линии,

– *второго порядка* (двумерная информация) – поверхности,

– *третьего порядка* (трехмерная информация) – объема, …

– *n–го порядка* (*n*–мерная информация) – *n*–мерного пространства.

Таким образом, структуру информации можно изменять, переходя от одного вида информации к другому. Все виды информации можно интерпретировать геометрическими образами, что бывает удобно на практике.

В инженерной практике широкое распространение имеет параметрическая информация, которую можно свести к следующим четырем основным видам: событию, величине, функции и комплексу. Им соответствуют введенные ниже дополнительные обозначения для физической реализации множеств.

1. Событие. Первичным и неделимым элементом информации является элементарное двоичное событие А (рис. 1.5, а) — выбор из утверждения или отрицания, истины или лжи, согласия или несогласия, наличия или отсутствия какого–либо явления. Примером могут служить сведения об импульсе или паузе в электрической цепи, выпуске годного или негодного изделия, достижении или недостижении измеряемой величиной одного определенного значения, черном или белом элементах телевизионного изображения, попадании или непопадании в цель, наличии или отсутствии команды и т. д.

Событие является категорией нулевой меры, т. е. не имеет геометрических измерений, поэтому оно и представимо точкой.

2. Величина. Величина *X* (рис. 1.5, б) есть упорядоченное в одном измерении (по шкале значений) множество событий, причем каждое из них отвечает принятию величиной какого–либо одного значения. Величина может быть или дискретной, или непрерывной; в первом случае множество событий счетно, во втором — несчетно. Геометрически величину можно представить линией.

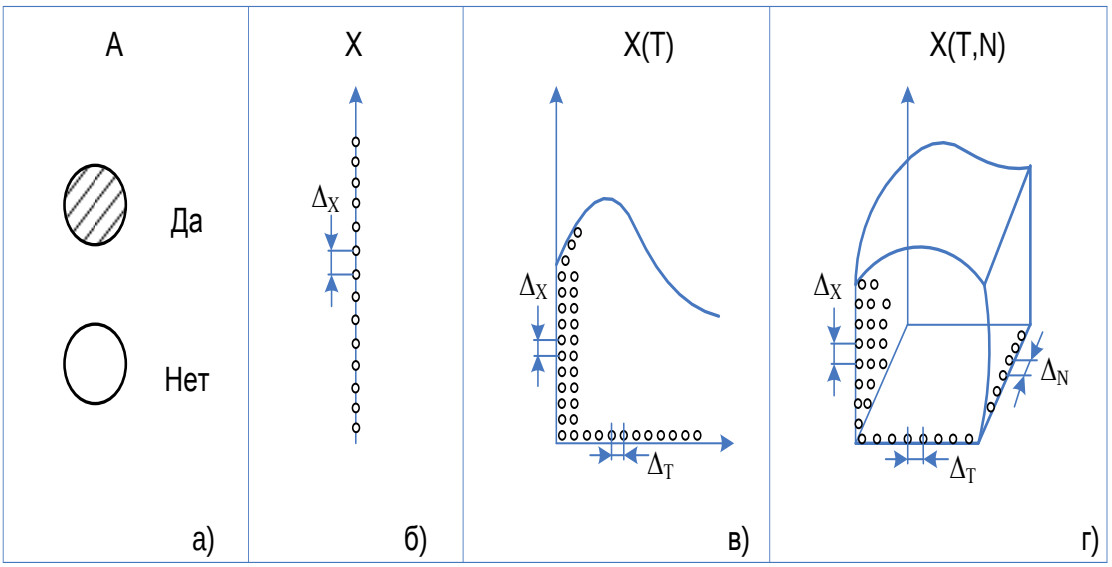


Рис. 1.5. Виды параметрической информации
а – событие; б – величина; в – функция; г – комплекс.

3. Функция. Функция $X(T)$ (рис. 1.5, в), $X(N)$ или $X_2(X_1)$ есть соотношение между величиной и временем, пространством или другой величиной. В этом смысле функцию можно трактовать как двумерное поле событий.

4. Комплекс. Полный комплекс информации $X(T, N)$ (рис. 1.5, г) есть соответствие между величиной, с одной стороны, и временем и пространством — с другой. Таким образом, полный комплекс информации есть трехмерное поле событий.

Как указывалось, информация может быть представлена моделями с различной размерностью. Отвлекаясь от конкретного вида координат (параметр *X*, время *T*, пространство *N*) и введя обобщенную координату информации Ф (согласно таблице 1.2), получаем следующую классификацию:

$$\Phi^0, \Phi^1, \Phi^2, \Phi^3, ..., \Phi^n,$$

где Φ° – нуль–мерная информация (событие); Φ^1 – одномерная информация (величина); Φ^2 – двумерная информация (функция); Φ^3 – трехмерная информация (комплекс); Φ^n– n–мерная информация (n–мерное пространство).

Табл. 1.2.

Структурные информационные формулы

События Φ^0 (A) Φ^1 (A_1, A_2, A_3, …, A_n)	**Пространство событий** Φ^1(N) Φ^2(N_1, N_2) Φ^3(N_1, N_2, N_3)
События во времени Φ^1 (A, T) Φ^2 (A_1, A_2, A_3, …, A_n, T)	**Пространство событий во времени** Φ^2 (N, T) Φ^3 (N_1, N_2, T) Φ^4(N_1, N_2, N_3, T)
Параметры Φ^1 (X) Φ^2 (X_1, X_2) Φ^3 (X_1, X_2, X_3) . . . . . . Φ^n (X_1, X_2, X_3, … , X_n)	**Параметрические пространства** Φ^2 (X, N) Φ^3 (X, N_1, N_2) Φ^4(X, N_1, N_2, N_3)
Параметры во времени Φ^2 (X, T) Φ^3 (X_1, X_2, T) . . . . . . Φ^n (X_1, X_2, X_3, … , X_{n-1}, T)	**Параметрические пространства во времени** Φ^3 (X, N, T) Φ^4 (X, N_1, N_2, T) Φ^4 (X, N_1, N_2, N_3, T)

Верхний индекс указывает размерность или порядок информации. Наиболее часто встречающиеся на практике разновидности научной и технической информации могут быть выражены теперь структурными информационными формулами (см. табл. 1.2), которые отличаются тем, что в них указываются только размерность и компоненты информации, но не функциональные зависимости между компонентами.

Одно событие есть нуль–мерная категория, так что совокупность пронумерованных подряд событий занимает одно измерение N.

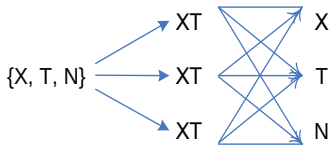
Время T само по себе не содержит информации. Множество событий во времени можно упорядочить относительно координат N и T в виде функции $N\ (\ T\ )$.

1.5. Структура информации

Информация может претерпеть различные структурные преобразования, показанные в таблице 1.3. Последовательность этих преобразований может быть различной в различных информационных системах. Получаемые в процессе преобразований структуры имеют абстрактный характер и не соответствуют строго тем или иным этапам обработки информации в информационных системах.

Табл. 1.3.

Структура информации

Условное обозначение	Структура	Характеристика структуры
{X}, {T}, {N}	Натуральная	Первоначальная структура информации
M, D, L{X}, {T}, {N}	Нормализованная	Приведена к единому масштабу, диапазону и началу отсчета
{X, T, N}	Комплексированная	Приведена к комплексу с обобщенными координатами X, T
	Декомпонированная	Преобразованы число измерений, структура и расположение
G_A {X, T, N}	Генерализованная	Устранена избыточность, выделена существенная часть по условию А
X*, T*, N*	Дискретная (квантованная)	Отсчеты в непрерывной форме
q_X, q_T, q_N	Безразмерная	Дискретные отсчеты приведены к безразмерной форме
	Кодированная	Цифровая форма информации

Натуральная информация отражает реальное существование объектов. Она имеет аналоговую форму, засорена шумами, неоптимальна по диапазонам и началам отсчетов значений параметров. Все эти ограничения обусловлены непосредственно физическими свойствами наблюдаемого объекта. Натуральную информацию можно условно представить как совокупности величин *X*,

моментов времени *T* и точек пространства *N* в виде множеств *{ X }*, *{ T }* и *{ N }*.

Нормализованная информация отличается от натуральной тем, что в ней каждое множество *{ X }*, *{ T }*, { *N* } уже приведено к одному масштабу, диапазону, началу отсчета и другим общим унифицированным характеристикам. Нормализованную информацию можно трактовать как результат воздействия на натуральную информацию операторов: масштабного *M,* диапазонного *D* и локализованного *L*.

Комплексированная информация образуется в результате приведения всей информации к полному комплексу, т. е. к трехмерной системе *XTN,* где *X* — обобщенная координата значений параметров или унифицированная шкала каких–либо оценок; T — обобщенная координата времени; *N* — обобщенная координата пространства источников информации. Комплексированная информация представляет собой связанное и координированное множество *X, T, N.* Изменение количества измерений структуры и расположения элементов в информационных комплексах приводит к форме *декомпонированной информации*. Особенно часто применяют следующие два вида декомпозиции:

– приведение физического пространства трех измерений (объема) физических полей, объемных объектов;

– многомерных систем датчиков, векторных и комплексных величин к пространствам двух и одного измерений;

– приведение полного комплекса информации *XTN* к любой плоскости *XT, XN, NT* или оси *X, T, N* координат измерений.

Декомпонированная информация декоррелирована, в ней нарушены или удлинены связи между отдельными элементами информации.

Структурная формула декомпозиции имеет вид, показанный на рисунке 1.6.

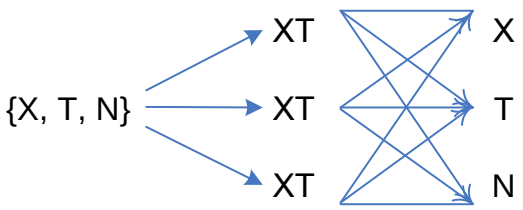


Рис. 1.6. Структурная формула декомпозиции

В *генерализованной информации* исключены второстепенные ее части, данные обобщены и укрупнены. Генерализация может охватывать как номенклатуру параметров, так и моменты времени, диапазоны измерения и степень подробности их отображения. Формула $G_A\{X, T, N\}$ показывает, что производится генерализация G по алгоритму A комплекса $\{X, T, N\}$.

Дискретная (квантованная) информация совпадает с исходной непрерывной информацией по физической размерности, отличаясь от нее лишь прерывным характером. Дискретизация может быть осуществлена по осям X, T и N параметрического комплекса. Дискретная информация удобна для расчетов и экономична в реализациях. Дискретизация может быть равномерной или неравномерной, производиться по постоянному или изменяющемуся во времени закону.

Безразмерная информация отличается универсальной безразмерной числовой формой. Число, отображающее безразмерную информацию, соответствует количеству информационных элементов (квантов) и получается в результате дискретизации информационного комплекса, т. е. равно отношению любой координаты к ее интервалу дискретности:

$$q_X = \frac{X}{\Delta_X};\ q_T = \frac{T}{\Delta_T};\ q_N = \frac{N}{\Delta_N} \qquad (1.1)$$

Кодированная информация имеет форму совокупности чисел, или цифровую форму, основанную на применении какой–либо системы счисления или кодирования.

Контрольные вопросы и задания

1. Что такое информация?
2. Как соотносятся информация и материя, информация и энтропия?
3. Охарактеризуйте каждый из этапов обращения информации.
4. Как классифицируется информация по видам с точки зрения теории информации?
5. Какие формы представления информации Вам известны? Охарактеризуйте каждый из них.
6. Приведите структурные информационные формулы наиболее часто встречающихся на практике разновидностей научной и технической информации.
7. Какие структурные преобразования может претерпевать информация? Охарактеризуйте каждое из этих преобразований.
8. В каких случаях говорят о декомпонированной информации? Охарактеризуйте ее и приведите структурную формулу.
9. Охарактеризуйте все виды параметрической информации и приведите схематичные изображения.
10. Постройте и охарактеризуйте методологическую схему формирования и материализации информации.

Глава 2. Измерение информации

Важнейшим вопросом теории информации является установление меры количества и качества информации.

Информационные меры отвечают трем основным направлениям в теории информации: структурному, статистическому и семантическому.

Структурная теория рассматривает дискретное строение массивов информации и их измерение простым подсчетом информационных элементов (квантов) или комбинаторным методом, предполагающим простейшее кодирование массивов информации.

Статистическая теория оперирует понятием энтропии как меры неопределенности, учитывающей вероятность появления, а следовательно, и информативность тех или иных сообщений.

Семантическая теория учитывает целесообразность, ценность, полезность или существенность информации.

Указанные три направления (структурное, семантическое и статистическое) имеют свои области применения, каждое из них имеет право на существование и развитие. *Структурная теория* применяется для оценки возможностей аппаратуры информационных систем (каналов связи, запоминающих и регистрирующих устройств) вне зависимости от условий их применения. *Статистическая теория* дает оценки информационных систем в конкретных применениях, например при передаче по системе связи информации с определенными статистическими характеристиками. Наконец, *семантическая теория* прилагается к оценке эффективности логического опыта.

2.1 Структурные меры количества информации

В структурных мерах во внимание принимается только дискретное строение носителя информации – количество содержащихся в нем информационных элементов, связей между ними или комбинаций из них.

Под информационными элементами понимаются неделимые части – кванты – информации в дискретных моделях реальных носителей информации, а также элементы алфавитов в числовых си-

стемах. В структурной теории различают геометрическую, комбинаторную и аддитивную меры информации.

Наибольшее распространение получила аддитивная мера Хартли – двоичная аддитивная мера, измеряющая количество информации в двоичных единицах – битах (binary digit).

2.1.1 Геометрическая мера

Определение количества информации геометрическим методом сводится к измерению длины линии, площади или объема геометрической модели данного носителя информации или сообщения. По геометрическим размерам носителя геометрическая мера определяет потенциальное, т. е. максимально возможное количество информации в данных структурных габаритах. Это количество называется информационной емкостью носителя. Информационная емкость – число, показывающее, какое количество квантов содержится на носителе.

Геометрическую меру можно применить не только для оценки информационной емкости, но и для оценки количества информации, содержащегося в отдельном сообщении. Если о величине, отображаемой сообщением, известно, что она имеет максимальное значение из того ряда значений, которые она уже принимала ранее, то можно считать, что количество информации, содержащееся как в этом, так и в любых более ранних сообщениях, определяется числом квантов, содержащихся в максимальном значении.

Пусть информация отражается полным комплексом *XTN*.

Если дискретные отсчеты осуществляются по осям *X, T* и *N* соответственно через интервалы Δ_X, Δ_T и Δ_N, то непрерывные координаты распадаются на элементы (кванты), количество которых составляет:

$$m_X = \frac{X}{\Delta_X};\ m_T = \frac{T}{\Delta_T};\ m_N = \frac{N}{\Delta_N}.$$

Тогда количество информации (в квантах) в полном комплексе *XTN*, определенное геометрическим методом, равно

$$M = m_X \cdot m_T \cdot m_N .$$

Может иметь место неравномерная (по осям) и нестационарная (изменяющая свой характер во времени) дискретизация. Тогда количество информации определяется по более сложным формулам, вытекающим из переменных характеристик дискретизации.

В случае сообщения его можно записать (если это текст) в строчку и длину получившейся строки разделить на длину одной буквы. Таким образом, геометрическая мера здесь сводится к определению числа букв в сообщении.

Другой пример: имеется две фотографии поверхности земли разной площади, но с одинаковой разрешающей способностью. Количество информации в смысле геометрической меры будет пропорционально площади фотографии.

2.1.2 Комбинаторная мера

Этот вид меры информации используют тогда, когда требуется оценить возможность передачи информации при помощи различных комбинаций букв. Образование комбинаций – одна из форм кодирования информации.

Количество информации в комбинаторной мере равно количеству разрешенных комбинаций букв.

Можно договориться (условиться) о различных правилах образования комбинаций. Различные комбинации могут, например, отличаться буквами или порядком их следования или и тем и другим одновременно.

В комбинаторике рассматриваются различные виды соединения элементов.

Сочетания из h элементов по l различаются составом элементов. Их возможное число равно

$$Q = \binom{l}{h} = \frac{h!}{l! \cdot (h-l)!} = \frac{h \cdot (h-1) \cdot \ldots \cdot (h-l+1)}{1 \cdot 2 \cdot 3 \cdot \ldots \cdot l} .$$

Сочетание с повторениями также различаются составом элементов, но элементы в них могут повторяться до l раз. Число различных сочетаний с повторениями из h элементов по l равно:

$$Q = \binom{l}{h}_{ПОВТ} = \frac{(h+l-1)!}{l! \cdot (h-1)!} = \binom{l}{h+l-1}.$$

Перестановки h элементов различаются их порядком. Число возможных перестановок h элементов

$$Q = 1 \cdot 2 \cdot 3 \cdot .... \cdot h = h!$$

Перестановка с повторениями элементов, причем один из элементов повторяется α, другой – β, наконец, последний — γ раз, характеризуется возможным числом

$$Q = \frac{(\alpha + \beta + ... + \gamma)!}{\alpha! \cdot \beta! \cdot ... \cdot \gamma!}.$$

Размещения из h элементов по l элементов различаются и составом элементов и их порядком. Возможное число размещений из h элементов по l

$$Q = \binom{l}{h} = h \cdot (h-1) \cdot (h-2) \cdot ... \cdot (h-l+1) = \frac{h}{(h-l)!}.$$

Возможное число размещений с повторениями из h элементов по l

$$Q\binom{h}{l}_{ПОВТ} = h^{\,l}.$$

При применении комбинаторной меры возможное количество информации Q совпадает с числом возможных соединений. Таким образом, определение количества информации Q в комбина-

торной мере заключается не в простом подсчете квантов, как это было при оценке в геометрической мере, а в определении количества возможных или действительно осуществленных комбинаций, т. е. в оценке структурного разнообразия.

Количество информации при том же количестве элементов теперь многократно увеличивается. *Так, например, в случае сочетаний из* 10 *элементов по* 0, 1, 2, 3, 4, 5, 6, 7, 8 *и* 9 *элементов получается следующее число образований, являющееся мерой количества информации:*

$$Q=\frac{10!}{0!\cdot(10-0)!}+\frac{10!}{1!\cdot(10-1)!}+\ldots+\frac{10!}{9!\cdot(10-9!)}+\frac{10!}{10!\cdot(10-10)!}=1+10+45+$$

$$+120+210+\ldots+252+210+120+45+10+1=$$
$$=1024 \text{ преобразования.}$$

Перестановки тех же 10 элементов дают:

$$Q=h!=1\cdot2\cdot3\cdot4\cdot5\cdot6\cdot7\cdot8\cdot9\cdot10=3628800 \text{ преобразований.}$$

Размещения 10 различных элементов по 10 различным позициям приводят к еще большему потенциальному количеству информации

$$Q=h^l=10^{10} \text{ образований.}$$

2.1.3. Аддитивная мера (мера Хартли)

Аддитивную меру можно рассматривать как более удобную для ряда применений комбинаторную меру. Наши интуитивные представления об информации предполагают, чтобы количество информации увеличивалось приблизительно пропорционально объему сообщения и объему используемого при этом алфавита. Ни геометрическая, ни комбинаторная меры этим свойством не обладают. Геометрическая мера вообще не учитывает свойства алфавита, объем же информации в случае комбинаторной меры растет не

прямо пропорционально размеру сообщения, а по закону геометрической прогрессии. *Например, если сообщение представлено не одной, а двумя перфокартами, число возможных комбинаций возрастет в* 2^{800} *раз и станет равным* 2^{1600}.

В 1929 г. Хартли была предложена аддитивная мера. Словом «Аддитивность» (суммируемость) здесь подчеркивается важное свойство этой меры, заключающееся в том, что измеренное при ее помощи количество информации в совокупности нескольких сообщений равно сумме количеств информации в каждом из сообщений в отдельности.

Предположим, что производится многократное измерение какой–либо величины в одинаковых условиях и каждый раз мы получаем одни и те же значения этой величины. Следовательно, никакой новой информации каждое последующее измерение не даст, однако, если измеряемая величина от опыта к опыту изменяется и чем больше различных значений она принимает, тем больше мы получаем информации об этой величине. Следовательно, чем ближе неопределённость величины до опыта, тем больше информации мы получим после опыта, т. е. количество комбинаций различных значений измеряемой величины косвенно характеризует количество информации, содержащейся в измеренных её значениях после опытов.

Это количество комбинаций можно определить из следующей зависимости:

$$Q = \frac{x_{\max} - x_{\min}}{\Delta x} + 1 \qquad (2.1)$$

где Q – количество комбинаций измеренных (определенных в результате измерений) значений величины;
$x_{\max}$ – максимальное значение измеренной величины;
$x_{\min}$ – минимальное значение измеренной величины
Δx – интервал квантования измеряемой величины.

Пример. Производится измерение температуры в аудитории. В результате 5–и измерений получены значения $T_{\max}$ = 20 °С, $T_{\min}$ =

19,5 °С, шаг квантования $\Delta T = 0{,}1$ *°С*. Количество комбинаций по формуле (2.1) будет:

$$Q = \frac{20° - 19{,}5°}{0{,}1°} + 1 = 6$$

Введем понятия глубины h и длины l числа.

Глубиной h числа называется количество различных элементов (знаков), содержащееся в принятом алфавите. Глубина числа соответствует основанию системы счисления и кодирования. Один полный алфавит занимает одно числовое гнездо, глубина которого также равна h.

В каждый данный момент реализуется только один какой-либо знак из h *возможных. Технически это может быть выполнено, например, путем смещения знаков вперед или назад до появления нужного знака.*

Длиной l числа называется количество числовых гнезд, т. е. количество повторений алфавита, необходимых и достаточных для представления чисел нужной величины. Длина числа соответствует разрядности системы счисления и кодирования. Один набор из l гнезд-алфавитов составляет одну числовую гряду, способную представлять и хранить одно полное число длиной l.

Некоторое количество чисел Q представляется *числовым полем*.

При глубине h и длине l количество чисел, которое можно представить с помощью числовой гряды, выразится формулой

$$Q = h^l, \tag{2.2}$$

т. е. емкость гряды экспоненциально зависит от длины числа l.

Пример. Допустим необходимо сформулировать сообщения по две буквы из заданных трёх букв А, В, С. Обозначим через $h = 3$ число символов в алфавите, а через $l = 2$ длину сообщения (количество символов в сообщении). Тогда число Q возможных сообщений будет равно

$$Q = h^l = 3^2 \rightarrow Q = \begin{vmatrix} AA & AB & AC \\ BA & BB & BC \\ CA & CB & CC \end{vmatrix}$$

Т. е. $Q = 9$ – это возможное количество комбинаций символов в сообщении.

Укажем на возможные конструкции подобных систем:

а) роликовый счетчик с l роликами и h цифрами на ободке каждого ролика;

б) комбинированный коммутатор с l переключателями, из которых каждый переключает h цепей;

в) запоминающее устройство с l ячейками, каждая емкостью h единиц;

г) изображение, состоящее из l дискретных элементов, причем каждый элемент изображения характеризуется h градациями цвета и тона;

д) страница печатного литературного текста, в котором содержится l_1 строк и l_2 букв в каждой строке (в среднем), т. е. всего $l = l_1 \times l_2$ числовых или буквенных гнезд; каждое гнездо имеет глубину h, т. е. способно выставить одну из h букв.

Во всех указанных случаях общее количество возможных состояний определяется выражением $Q = h^l$.

Вследствие показательного закона зависимости Q от l число Q не является удобной мерой для оценки информационной емкости. Поэтому Хартли ввел аддитивную двоичную логарифмическую меру, позволяющую вычислять количество информации в двоичных единицах – битах, сокращенно обозначаемых «бит».

Для этого берется не само число Q, а его двоичный логарифм

$$I = \log_2 Q = \log_2 h^l = l \cdot \log_2 h, \qquad (2.3)$$

где l обозначает количество информации по Хартли.

Если количество разрядов (длина l числа) равно единице, принята двоичная система счисления (глубина h числа равна двум)

и используется двоичный логарифм, то потенциальное количество информации равно одному биту:

$$\log_2 2 = 1.$$

Это и есть единица информации в принятой системе оценки. Она соответствует одному элементарному событию, которое может произойти или не произойти.

Аддитивная мера удобна тем, что она обеспечивает возможность сложения и пропорциональность количества информации длине числа l, т. е. количеству числовых гнезд.

Введенное количество информации эквивалентно количеству двоичных знаков 0 и 1 при кодировании сообщений по двоичной системе счисления. Одному биту соответствует одна двоичная единица.

Пусть, например, определяется потенциальное количество информации, содержащееся в системе, информационная емкость которой характеризуется десятичным числом:

$$Q = 1000 \approx 999; \quad h = 10; \quad l = 3;$$
$$I = \log_2 Q = \log_2 h^l = \log_2 10^3 = 3 \cdot \log_2 10 \approx 10.$$

Закодируем это число по двоичной системе $h = 2$ и найдем l:

$$Q = 2^l \quad \log_2 Q = l \cdot \log_2 2 = l,$$

тогда

$$l = \log_2 1024 = 10 \approx I.$$

Это означает, что по двоичной системе данное число запишется (закодируется) десятью единицами:

$$Q = 1024 \approx 1023 = 2^9 + 2^8 + 2^7 + 2^6 + 2^5 + 2^4 + 2^3 + 2^2 + 2^1 + 2^0 =$$
$$= 1\ 111\ 111\ 111.$$

Точное значение $Q = 1024 = 2^{10}$ нужно записать одной единицей в позиции, соответствующей 2^{10}, и десятью нулями 10 000 000 000. Приближенные равенства показывают, что количество знаков приходится брать равным либо I, если I – целое число, либо дополнять его до ближайшего целого числа, если I – дробное число.

При наличии нескольких источников информации общее количество информации, которое можно получить от всех источников, вместе взятых

$$I\left(Q_1, Q_2, ..., Q_R\right) = I\left(Q_1\right) + I\left(Q_2\right) + ... + I\left(Q_R\right).$$

Таким образом, более совершенная, отвечающая нашим интуитивным представлениям об информации мера информации должна быть связана с вероятностью появления сообщения, что и имеет место в статистических мерах информации.

2.2. Объем информации

Так как информация есть неопределенность, снимаемая при получении сообщения, то количество информации может быть представлено как произведение общего числа сообщений *к* на среднюю энтропию *H*, приходящуюся на одно сообщение:

$$I = k \cdot H \text{ бит.}$$

Количество информации определяется исключительно характеристиками первичного алфавита, объем – характеристиками вторичного алфавита. Объем[1] информации

$$Q = k \cdot l_{cp} \ ,$$

[1] Строго говоря, объема информации не существует. Мы вкладываем в этот термин то, что привыкли под этим подразумевать, а именно количество элементарных символов в принятом (вторичном) сообщении.

где $l_{ср}$ – средняя длина кодовых слов вторичного алфавита. Для равномерных кодов (все комбинации кода содержат одинаковое количество разрядов)

$$Q = k \cdot n \ ,$$

где n – длина кода (число элементарных посылок в коде).

Объем равен количеству информации, если $l_{ср} = H$, т. е. в случае максимальной информационной нагрузки на символ сообщения. Во всех остальных случаях $I < Q$.

Например, если кодировать в коде Бодо некоторый равновероятный алфавит, состоящий из 32 символов, то

$$I = k \cdot H = k \cdot \log_2 m = k \cdot \log_2 32 = k \cdot 5 \text{ бит;}$$
$$Q = k \cdot l_{ср} = k \cdot 5 \text{ бит.}$$

Если закодировать в коде Бодо русский 32-буквенный алфавит, то без учета корреляции между буквами количество информации

$$I = k \cdot H = k \cdot 4.358 \text{ бит, } Q = k \cdot 5; \ Q > I,$$

т. е. если в коде существует избыточность и $H \neq H_{\max}$, то объем в битах всегда больше количества информации в тех же единицах.

Пример. Символы алфавита обладают двумя качественными признаками (m = 2, символов первичного алфавита).

1. Какое количество сообщений можно получить, комбинируя по 3, 4, 5 и 6 элементов в сообщениях?

2. Какое количество информации приходится на один элемент таких сообщений?

Решение.

1. $N = m^n; m = 2;$ $N_1 = 2^3 = 8;$ $N_2 = 2^4 = 16;$ $N_3 = 2^5 = 32;$ $N_4 = 2^6 = 64.$

2. $I = \log_2 N;$ $I_1 = \log_2 8 = 3$ бит; $I_2 = \log_2 16 = 4$ бит;

$I_3 = \log_2 32 = 5$ бит; $I_4 = \log_2 64 = 6$ бит.

2.3. Семантические меры информации

Под «семантикой» понимается смысл, содержание информации. Место семантики в системе понятий семиотики (наука о знаках, словах и языках) указано на рисунках 2.1 и 2.2.

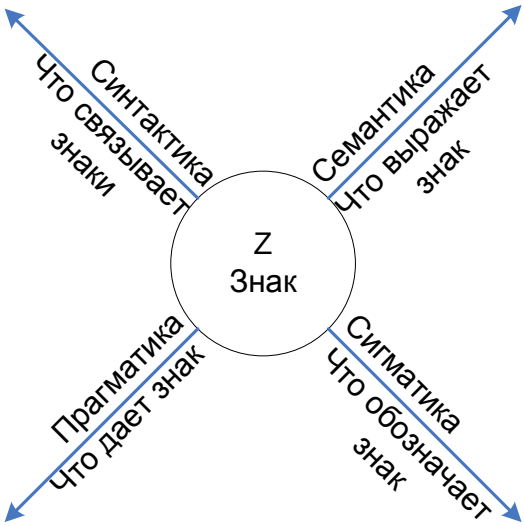


Рис. 2.1. К основным понятиям семиотики

Рис. 2.2. К основным понятиям семиотики:
Z - знак; Z' - соотносящийся знак; C - смысл; O - объект; Ч - человек

Знаком называется условное изображение элементов сообщения, *словом* — совокупность знаков, имеющая смысловое (предметное) значение, *языком* — словарь и правила пользования им.

Соответственно приведенной структуре в семиотике различаются синтактический, семантический, сигматический и прагматический аспекты теории информации.

Структурная и статистическая оценки информации относятся к *синтактическому аспекту*. *Сигматический аспект* отображается теорией сигналов и кодов, рассматривающей условные обозначения элементов информации. Сигналы являются физическими носителями обозначенных элементов, а коды — обозначениями этих элементов. Сигматические оценки не имеют прямого отношения к мерам информации. Поэтому остается рассмотреть семантические и прагматические оценки информации.

Рассмотрим оценки, отвечающие как семантическому, так и прагматическому аспектам теории информации. В инженерных применениях *прагматические оценки* сливаются с семантическими, поскольку не имеющие смысл сведения бесполезны, а бесполезные сведения бессмысленны.

Оценка эффективности логического вывода, степени приближения к истине требует некоторой формализации, в данном случае – формализации смысла.

Один из путей такой формализации предлагается семантической теорией информации.

Карнап и Бар-Хиллел предложили использовать для целей измерения смысла функции истинности и ложности логических высказываний (предложений). За основу дискретного описания объекта берется атомарное (неделимое) предложение, подобное элементарному событию теории вероятностей и соответствующее неделимому кванту сообщения.

Полученная таким образом оценка получила название *содержательности информации*.

2.3.1. Содержательность информации

Мера содержательности обозначается *cont* (от английского content — содержание).

Содержательность события i выражается через функцию меры $m(i)$ — содержательности его отрицания — как

$$cont(\,i\,) = m(\,{\sim}i\,) = 1 - m(\,i\,),$$

где i – рассматриваемое событие;

m – функция меры;

~ – знак отрицания.

Оценка содержательности основана на математической логике, в которой логические функции истинности $m(\,i\,)$ и ложности $m(\,{\sim}i\,)$ имеют формальное сходство с функциями вероятностей события $p(\,i\,)$ и антисобытия $q(\,i\,)$ в теории вероятностей.

В обоих случаях имеют место сходные условия

$$m(i)+ m({\sim}i)=1; \quad p(i)+q(i)=1.$$

Как и вероятность, содержательность изменяется в пределах

$$0 \le m(\,i\,) \le 1$$

Логическая оценка количества информации, получившая обозначение *Inf*, определяется следующим выражением:

$$Inf = \log_2\left(\frac{1}{1-cont(i)}\right) = \log_2\frac{1}{m(i)} = -\log_2 m({\sim}i).$$

Суть логической оценки состоит в том, что учитываются меры истинности или ложности событий, что приближает к оценке смысла информации.

2.3.2. Целесообразность информации

Если информация используется в системах управления, то ее полезность разумно оценивать по тому эффекту, который она оказывает на результат управления. В связи с этим А. А. Харкевичем была предложена мера целесообразности информации, которая определяется как изменение вероятности достижения цели при получении дополнительной информации.

Полученная информация может быть пустой, т. е. не изменять вероятности достижения цели, и в этом случае ее мера равна

нулю. В других случаях полученная информация может изменять положение дела в худшую сторону, т. е. уменьшить вероятность достижения цели, и тогда она будет дезинформацией, которая измеряется отрицательным значением количества информации. Наконец, в третьем, благоприятном случае получается добротная информация, которая увеличивает вероятность достижения цели и измеряется положительной величиной количества информации.

На рисунке 2.3, а показано исходное состояние (точка 1), из которого возможны два пути: путь 1–2 и путь 1–3. Пусть точка 3 представляет собой цель, а точка 2 – некоторый промежуточный пункт в стороне от цели.

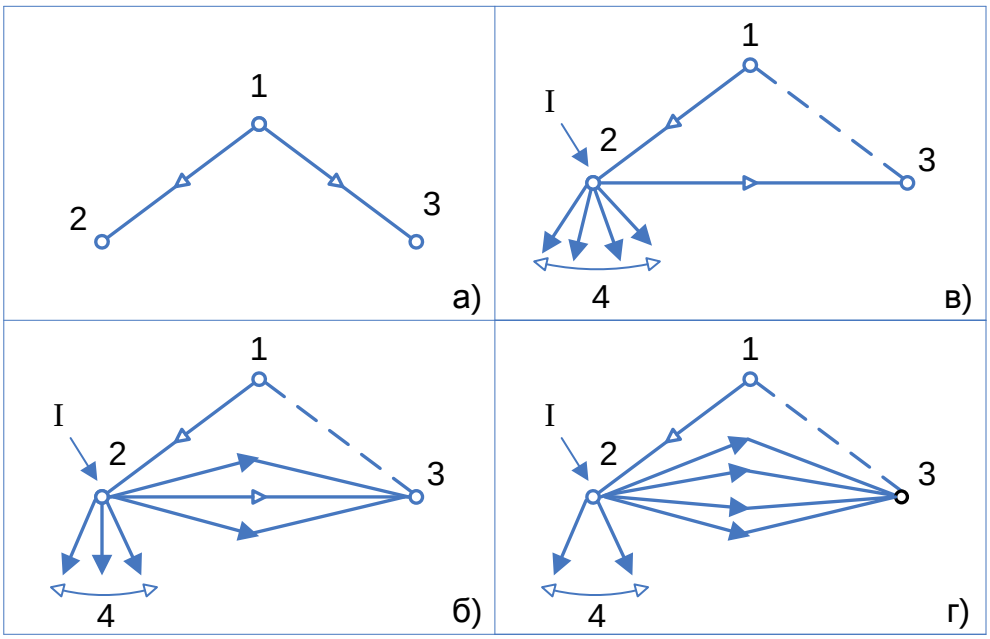


Рис. 2.3. К применению меры Харкевича для оценки информации, получаемой на пути движения к цели.

Если неизвестны пути к цели, то вероятности достижения цели по путям 1–2 и 1–3 будут одинаковы, а так как схема опыта содержит два исхода, то можно записать условие

$$p\,(1-2)=p\,(1-3)=\frac{1}{2}.$$

Предположим, что достигнута точка 2 (рис. 2.3, б) и в этой точке 2 получается информация, однако эта информация оказалась нейтральной, так как она оставила неизменной вероятность достижения цели.

Тогда

$$I = \log_2 \frac{p(2-3)}{p(1-2)} = \log_2 \frac{3/6}{3/6} = 0.$$

Второй (неблагоприятный) случай показан на рисунке 2.3, в. Здесь в точке *2* получена ложная информация, которая уменьшила вероятность достижения цели ввиду преобладания ложных направлений *4*. Количество информации

$$I = \log_2 \frac{p(2-3)}{p(1-2)} = \log_2 \frac{1/6}{1/2} = -1{,}58.$$

Остановимся также на третьем (благоприятном) случае (рисунок 2.3, г), когда вероятность достижения цели *3* увеличивается. В этом случае также в точку *2* поступает информация, которая только два исхода из шести оставляет ложными (направление *4),* но четыре исхода из шести являются благоприятными и ведут к цели *3*. Количество информации

$$I = \log_2 \frac{p(2-3)}{p(1-2)} = \log_2 \frac{4/6}{1/2} = -0{,}42,$$

Мера целесообразности в общем виде может быть аналитически выражена в виде соотношения

$$I_{цел} = \log_2 p_1 - \log_2 p_0 = \log_2 \frac{p_1}{p_0}, \qquad (2.4)$$

где p_0 и p_1 — начальная (до получения информации) и конечная (после получения информации) вероятности достижения цели.

Пример. Пусть вероятность p_2 сдачи экзамена по теории информации до получения сообщения (подсказки от соседа) оценивается студентом со значением 0,2. После того, как ему удалось получить подсказку, вероятность сдачи увеличилась: p_1 = 0,8. Определить количество информации, содержащейся в подсказке, с точки зрения ее целесообразности.

В соответствии с формулой (2.4) имеем

$$I = \log_2 (0{,}8/0{,}2) = \log_2 4 = 2 \text{ бит.}$$

Пример. Пусть положение студента до получения подсказки оценивается аналогично предыдущему примеру. После получения подсказки, вопреки ожиданиям, вероятность сдачи еще уменьшилась, поскольку подсказка содержала неверную информацию: p_1 = 0,1. Определить количество информации, содержащейся в подсказке, с точки зрения ее целесообразности.

В соответствии с формулой (2.4) имеем

$$\mathrm{I} = \log_2 (0{,}1/0{,}2) = \log_2 0{,}5 = -1.$$

Таким образом, полученная информация является дезинформацией, поскольку имеет отрицательный знак при измерении.

2.4. Вероятность и информация

При вероятностном подходе информация рассматривается как сообщение об исходе случайных событий, реализации случайных величин и функций, а количество информации ставится в зависимость от априорных вероятностей этих событий, величин, функций.

Когда получается сообщение о часто встречающихся событиях, вероятность появления которых стремится к единице, т. е. к показателю полной достоверности, то такое сообщение мало информативно. Столь же мало информативны сообщения о противоположных событиях (антисобытиях), вероятности которых стремятся к нулю и которые, следовательно, почти невозможны. Например,

событие «часы идут» имеет вероятность $p = 1$, тогда как антисобытие «часы не идут» имеет вероятность $q=1 – p = 0$.

Событие и антисобытие составляют одно двоичное однопредметное явление. Большинство видов информации можно свести к двоичным явлениям «да – нет» и к паре «событие – антисобытие». Именно эта пара явлений является простейшим и неделимым элементом (квантом) информации (таблица 2.1).

Табл. 2.1.

Двоичные однопредметные явления

Событие	***Антисобытие***
Дождь идет	Дождь не идет
Снег идет	Снег не идет
Машина работает	Машина не работает
Лампа горит	Лампа погашена
Уровень достигнут	Уровень не достигнут

Может быть также двоичное двухпредметное событие, заключающееся в выборе одного из двух возможных предметов, например, черного или белого шара из урны, герба или решки на монете.

Другой вид сообщений составляют двухпредметные двоичные явления, которые распадаются на четыре элементарных акта. Например: 1) вынут белый шар; 2) не вынут белый шар; 3) не вынут черный шар; 4) вынут черный шар (рис. 2.4).

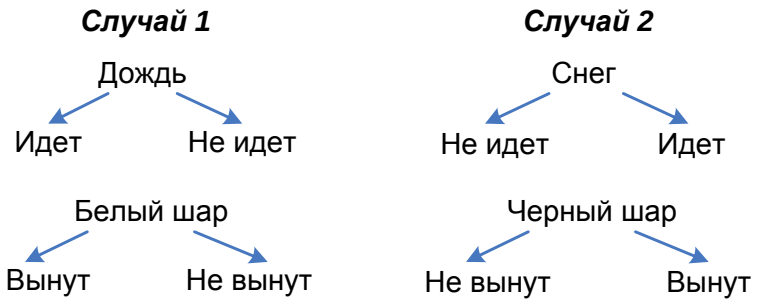


Рисунок 2.4 – Двоичные двухпредметные явления.

Могут быть также исключающие друг друга события. Например, если выпал герб, то не может одновременно выпасть

решка. Если в цепи установилось напряжение 100 В, то не может в той же цепи одновременно существовать напряжение 12 В.

Будем под событием далее понимать элементарное однопредметное явление, которое может быть с вероятностью от $p = 0$ до $p = 1$ или не быть с вероятностью от $q = 1 - p = 1$ до $q = 1 - p = 0$.

Когда $p = 0.5$ и $q = 0.5$, будет иметь место наибольшая неопределенность в ситуации.

2.5 Количественная оценка информации

2.5.1 Понятие энтропии

Сведения являющиеся объектом хранения, передачи и преобразования, называются *информацией*. Мера измерения количества информации основана на понятии энтропии. *Энтропия* – это мера степени неопределенности состояния системы X (случайной величины) с конечным или счетным числом исходов.

Что означает неопределенность и как ее измерить?

Пример. Пусть имеются две системы: первая система – игральная кость (имеет 6 состояний), вторая система – монета (имеет 2 состояния).

Спрашивается: состояние какой системы труднее предугадать (неопределенность какой системы больше)? Естественно неопределенность первой системы выше. Значит, степень неопределенности системы зависит от числа возможных ее состояний. Однако число состояний не является исчерпывающей характеристикой степени неопределенности.

Покажем это на примере для систем с двумя устойчивыми состояниями. Пусть имеются 2 монеты M_1 и M_2 (у монеты имеется два возможных состояния: орёл О и решка Р):

M_1	О	Р
p_i	0,5	0,5

M_2	О	р
p_i	0,999	0,001

где p_i – вероятности нахождения монет в состоянии орел или решка.

Нетрудно заметить, что неопределённость этих двух систем будет различной. Неопределенность первой системы больше, так

как из таблицы видно, что вторая монета практически постоянно находится в состоянии орел. Первая же монета неизвестно в каком находится состоянии, она с равной вероятностью может находиться или в состоянии орел, или в состоянии решка.

Таким образом, мы видим, что степень неопределенности определяется также и вероятностями состояний системы. В качестве меры априорной неопределенности теория информации предлагает энтропию.

Количественная мера информации тесно связана с понятием неопределенности ситуации – энтропии и равна уменьшению этой неопределенности после получения сообщения, т. е. разности априорной (до получения сообщения) и апостериорной (после получения сообщения) энтропии.

Если произошло событие (например, вы получили сообщение), в результате чего неопределенность была полностью снята, априорная вероятность возникновения которого p, то количество информации по Хартли, которое при этом будет получено, равно

$$I = \log_2 p.$$

Если же событие не произошло, однако известна вероятность его появления, то имеем дело с неопределенностью в отношении этого события – энтропией, которая вычисляется по аналогичной формуле:

$$H = \log_2 p.$$

В зависимости от происхождения и, соответственно, смысла, который вкладывается в вероятность, используемую для вычисления количества информации и энтропии, меняется смысл получаемой информации и энтропии. Если же после получения сообщения неопределенность снимается не полностью, то количество информации измеряется как разность между априорной энтропией (неопределенностью до получения сообщения) и апостериорной энтропией (неопределенностью после получения сообщения):

$$I = H_{\text{апр}} - H_{\text{апст}}.$$

Надо при этом иметь в виду, что как информация, так и энтропия – понятия относительные и имеют смысл по отношению к некоторым определенным событиям, вероятности которых используются в соответствующих формулах.

Понятие энтропии (от греческого эн-тропе — обращение) распространилось на ряд областей знания.

Энтропия в термодинамике означает вероятность теплового состояния вещества, в математике – степень неопределенности ситуации или задачи, в информатике она характеризует способность источника отдавать информацию.

Все эти понятия родственны между собой и в общем отображают степени богатства и неожиданности состояний

Согласно второму закону термодинамики (Больцмана) энтропия замкнутого пространства выражается как

$$H = -\frac{1}{N} \cdot \sum_{i=1}^{k} n_i \cdot \ln \frac{n_i}{N},$$

где N – количество молекул в данном пространстве; n_i – количество молекул скорости $v_i + \Delta v$

Но n_i /N есть частоты событий и, следовательно, вероятности того, что молекулы имеют скорость $v_i + \Delta v$ будут равны

$$n_i / N = p_i$$

Тогда

$$H = -\cdot \sum_{i=1}^{k} p_i \cdot \ln p_i$$

Можно заменить ln на $\log_2$, для того, чтобы измерять энтропию в двоичных единицах – битах. Тогда получим

$$H = -\cdot \sum_{i=1}^{k} p_i \cdot \log_2 p_i. \qquad (2.5)$$

Все другие виды энтропии выражаются аналогичными формулами.

Здесь следует только заметить, что с вероятностными оценками нужно поступать осмотрительно. Не всегда можно полагаться на идеализированные характеристики, начальные условия и неизменность (стационарность) параметров. На практике всегда что-либо меняется. Если не меняется параметр состояния, то меняются время и место, и уже это создает новую информацию. Кроме того, сам факт неизменности является несокращаемой информацией, так как в любой момент времени фактически может начаться или произойти изменение. Может, наконец, меняться сам субъект – приемник информации при неизменном состоянии источника. Практически приходится учитывать всевозможные обстоятельства: что, когда, кто, кому, зачем, как, т. е. конкретизировать вид информации, время, отправителя и получателя, назначение и способ реализации, все изменения, происходящие с ними.

2.5.2 Энтропия ансамбля

Ансамблем называется полная группа событий или, иначе, поле совместных событий с известным распределением вероятностей, составляющих в сумме единицу. Здесь имеется в виду конечное множество событий и, следовательно, дискретная система состояний, значений, положений и т. д.

Энтропия ансамбля есть количественная мера его неопределенности, а следовательно, и информативности.

Опытом может быть и измерение случайной величины X, принимающей различные значения. Тогда каждое определяемое значение имеет смысл исхода, или элементарного события.

Вообще событиями $x_1, x_2, \ldots, x_k$ могут быть k возможных дискретных состояний какой–либо физической системы, например: k значений измеряемой величины, k положений регулирующего органа, состояние k элементов производственного оборудования и т. д.

Этому соответствует таблица 2.2, называемая схемой ансамбля.

Табл. 2.2.

Схема ансамбля

Исходы опыта	A_1	A_2	…	A_i	…	A_k
Значения измеряемой величины	x_1	x_2	…	x_i	…	x_k
Вероятности исхода или значения	p_1	p_2	…	p_i	…	p_k

В простейшем случае эти события несовместимы. Они составляют полную группу, в которой обязательно реализуется одно из событий и имеет место условие

$$\sum_{i=1}^{k} p(x_i) = p(x_1) + p(x_2) + ... + p(x_k) = 1.$$

В общем случае вероятности не остаются постоянными. Они могут изменяться во времени, в зависимости от условий и обстоятельств. Тогда и статистические характеристики (среднее значение и дисперсия) становятся переменными величинами. Процессы, описываемые этими величинами, называются *нестационарными* в статистическом смысле.

В статистической теории информации (теории связи), предложенной Шенноном в 1948 г., энтропия количественно выражается как средняя функция множества вероятностей каждого из возможных исходов опыта. Пусть имеется всего N возможных исходов опыта, из них k разных, и i–й исход ($i = 1, ..., k$) повторяется n_i раз и вносит информацию, количество которой оценивается как I_i. Тогда средняя информация, доставляемая одним опытом,

$$I_{ср} = \frac{n_1 \cdot I_1 + n_2 \cdot I_2 + ... + n_k \cdot I_k}{N}.$$

Но количество информации в каждом исходе связано с его вероятностью p_i и выражается в двоичных единицах (битах) по формуле Хартли

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

тогда

$$I_{cp} = \frac{n_1 \cdot \left(-\log_2 p_1\right) + n_2 \cdot \left(-\log_2 p_2\right) + \ldots + n_k \cdot \left(-\log_2 p_k\right)}{N}.$$

Последнее выражение можно записать также в виде

$$I_{cp} = \frac{n_1}{N} \cdot \left(-\log_2 p_1\right) + \frac{n_2}{N} \cdot \left(-\log_2 p_2\right) + \ldots + \frac{n_k}{N} \cdot \left(-\log_2 p_k\right).$$

Но отношения n_i / N представляют собой частоты повторения исходов, а следовательно, могут быть заменены их вероятностями:

$$n_i / N = p_i,$$

поэтому средняя информация в битах может быть выражена следующим образом:

$$I_{cp} = p_1 \cdot \left(-\log_2 p_1\right) + p_2 \cdot \left(-\log_2 p_2\right) + \ldots + p_k \cdot \left(-\log_2 p_k\right),$$

или

$$I_{cp} = -\sum_{i=1}^{k} p_i \cdot \log_2 p_i.$$

Полученную величину Шеннон назвал *энтропией* и обозначил буквой H:

$$H = I_{cp} = -\sum_{i=1}^{k} p_i \cdot \log_2 p_i. \qquad (2.6)$$

2.5.3. Свойства энтропии

Энтропия может быть определена так же, как среднее количество информации на одно сообщение или математическое ожидание количества информации *I* для измеряемой величины *X:*

$$H(X) = M\left[I(X)\right] = -\sum_{i=1}^{k} p_i(x_i) \cdot \log_2 p_i(x_i).$$

Функция $H(p)$, где $p = (p_1, \ldots, p_k)$ — вектор вероятности исходов, была выбрана Шенноном так, чтобы она удовлетворяла следующим требованиям:

1) $H(p)$ непрерывна на интервале $0 \le p \le 1$, т. е. при малых изменениях p величина H изменяется мало;

2) $H(p)$ симметрична относительно p, т. е. не изменяется при любой перемене мест аргументов p_i;

3) $H(p_1, p_2, \ldots, p_{k-1}, q_1, q_2) = H(p_1, p_2, \ldots, p_k) + p_k \cdot H(q_1|p_k, q_2|p_k)$, т. е. если событие x_k состоит из двух событий x'_k и x''_k с вероятностями q_1 и q_2, $q_1 + q_2 = p_k$, то общая энтропия .будет равна сумме энтропии – неразветвленной системы и разветвленной части с весом p_k при условных вероятностях $q_1| p_k$ и $q_2| p_k$ (рис. 2.5).

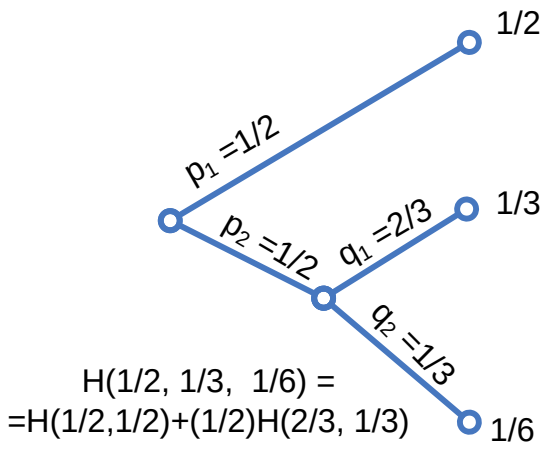


Рис. 2.5. Энтропия разветвленной системы

Кроме того, энтропия H характеризуется следующими свойствами:

1) энтропия всегда неотрицательна, т. к. значения вероятностей выражаются дробными величинами, а их логарифмы — отрицательными величинами, так что члены $\log_2 p_i = -(-a)$ неотрицательны;

2) энтропия равна нулю в том крайнем случае, когда одно событие равно единице, а все остальные – нулю. Это тот случай, когда об опыте или величине все известно заранее и результат не приносит никакой новой информации;

3) энтропия имеет наибольшее значение при условии, когда все вероятности равны между собой:

$$p_1 = p_2 = ... = p_i = ... = p_k = \frac{1}{k}.$$

При этом

$$H = -\log_2 \frac{1}{k} = \log_2 k\,.$$

Логарифмическая статистическая мера информации связана с аддитивной логарифмической мерой Хартли:

$$I' = \log_2 h.$$

Совпадение оценок количества информации по Шеннону и по Хартли свидетельствует о полном использовании информационной емкости системы. В случае неравных вероятностей количество информации по Шеннону меньше информационной емкости системы. Так, энтропия для двух неравновероятных состояний одного элемента (h =2) равна:

$$H = -\left(p_1 \cdot \log_2 p_1 + p_2 \cdot \log_2 p_2\right). \qquad (2.7)$$

Она меньше информационной емкости двоичной ячейки, составляющей 1 бит, как это видно из приведенных ниже примеров.

Пример. Требуется определить энтропию для равновероятных, неравновероятных и детерменированных состояний.

Решение.

а) равновероятные состояния

$p_1 = p_2 = 0.5$; $p_1 + p_2 = 1$;

$$H = -(0{,}5 \cdot \log_2 0{,}5 + 0{,}5 \cdot \log_2 0{,}5) = -[0{,}5 \cdot (-1) + 0{,}5 \cdot (-1)] = 1 \text{ бит};$$

б) неравновероятные состояния

$p_1 = 0.9$; $p_2 = 0.1$; $p_1 + p_2 = 1$;

$$H = -(0{,}9 \cdot \log_2 0{,}9 + 0{,}1 \cdot \log_2 0{,}1) = -[0{,}9 \cdot (-0.1520) + 0{,}1 \cdot (-3.3219)] = 0{,}46; \text{ бит.}$$

в). Детерминированные состояния

$p_1 = 1$; $p_2 = 0$; $p_1 + p_2 = 1$;

$$H = -(1 \cdot \log_2 1 + 0 \cdot \log_2 0) = 0 \text{ бит.}$$

Изменение энтропии H в зависимости от вероятности p однопредметного события показано на рис. 2.6. Максимум $H = 1$ достигается при $p = 0{,}5$, когда два состояния равновероятны. При вероятностях $p = 0$ или $p = 1$, что соответствует полной невозможности или полной достоверности события, энтропия равна нулю.

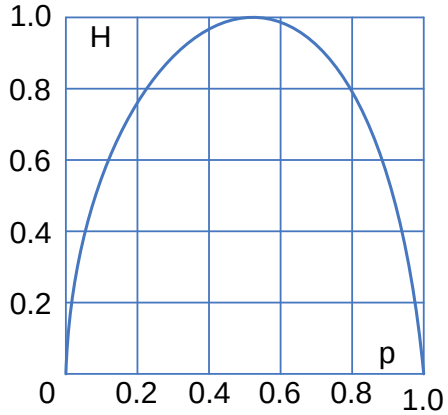


Рис. 2.6. Изменение энтропии зависимости от вероятности

элементарного (двоичного) события

2.5.4. Избыточность источника

Если энтропия источника сообщений не равна максимальной энтропии для алфавита с данным количеством качественных признаков (имеются в виду качественные признаки алфавита, при помощи которых составляются сообщения), то это, прежде всего, означает, что сообщения данного источника могли бы нести большее количество информации. Абсолютная избыточность на символ сообщений такого источника

$$\Delta D = (H_{\text{макс}} - H) \text{ бит/символ.} \qquad (2.8)$$

Для определения количества «лишней» информации, которая заложена в структуре алфавита либо в природе кода, вводится понятие избыточности. Избыточность, с которой мы имеем дело в теории информации, не зависит от содержания сообщения и обычно заранее известна из статистических данных[1].

Информационная избыточность показывает относительную избыточность на символ алфавита и является безразмерной величиной:

$$D = \frac{H_{макс} - H}{H_{макс}} = 1 - \frac{H}{H_{макс}}, \qquad (2.9)$$

где $\frac{H}{H_{макс}} = \mu$ — коэффициент сжатия (относительная энтропия); H и $H_{макс}$ берутся относительно одного и того же алфавита.

Наличие избыточности приводит к загрузке канала связи передачей лишних букв сообщений, которые не несут информации (их можно угадать и не передавая).

Однако преднамеренная избыточность в сообщениях иногда используется для повышения достоверности передачи информации

[1] Рассмотрение семантической избыточности не входит в задачи теории информации.

– например, при помехоустойчивом кодировании в системах передачи информации с исправлением ошибок. Большую избыточность имеет любой разговорный язык. Например, избыточность русского языка (как и других) около 50%. Благодаря избыточности облегчается понимание речи при наличии дефектов в произношении или при искажениях речевых сигналов в каналах связи.

Пример. Сообщение составляется из алфавита *a*, *b*, *c*, *d*. Вероятность появления букв алфавита в текстах равна соответственно: $p_a = 0{,}2$; $p_b = 0{,}3$; $p_c = 0{,}4$; $p_d = 0{,}1$. Найти избыточность сообщений, составленных из данного алфавита.

Решение: Избыточность, определим по формуле (2.9). Для алфавита из четырех букв максимальная энтропия

$$H_{макс} = \log_2 m = \log_2 4 = 2 \text{ бит/символ;}$$

средняя энтропия на символ сообщения

$$H = -\sum_i p_i \log p_i = -(0{,}2\log_2 0{,}2 + 0{,}3\log_2 0{,}3 + 0{,}4\log_2 0{,}4 +$$
$$+ 0{,}1\log_2 0{,}1) = 0{,}4644 + 0{,}5211 + 0{,}5288 + 0{,}3322 =$$
$$= 1{,}8465 \text{ бит/символ,}$$

тогда избыточность

$$D = 1 - \frac{1{,}8465}{2} = 1 - 0{,}9232 = 0{,}07688.$$

2.5.5. Производительность источника

Производительность источника определяется количеством информации, передаваемой в единицу времени. Измеряется производительность количеством двоичных единиц информации (бит) в секунду. Если все элементы сообщения имеют одинаковую длительность τ, то производительность

$$H'(x) = \frac{H(x)}{\tau} \ . \qquad (2.10)$$

Если же различные элементы сообщения имеют разную длительность, то в формуле (2.10) надо учитывать вместо длительности τ среднюю длительность $\bar{\tau}$, равную математическому ожиданию величины τ:

$$\bar{\tau} = \sum_{i=1}^{K} \tau_i p(\tau_i).$$

Однако в последней формуле $p(\tau_i)$ можно заменить на $p(x_i)$ (вероятность i–го сообщения), так как эти вероятности равны. В результате получаем

$$\bar{\tau} = \sum_{i=1}^{K} \tau_i p(x_i),$$

а производительность источника будет равна

$$H'(x) = \frac{H(x)}{\bar{\tau}}.$$

Максимально возможная производительность дискретного источника равна

$$H'_{\max}(x) = \frac{H_{\max}(x)}{\bar{\tau}} = \frac{\log k}{\bar{\tau}}.$$

Для двоичного источника, имеющего одинаковую длительность элементов сообщения ($k = 2$, $\bar{\tau} = \tau$) имеем

$$H'_{\max}(x) = \frac{\log_2 2}{\tau} = \frac{1}{\tau} \text{ бит/с.}$$

При укрупнении алфавита в слова по n букв, когда $k = 2^n$, $\bar{\tau} = n\tau$, имеем

$$H'_{\max}(x) = \frac{\log_2 2^n}{n\tau} = \frac{1}{\tau} \text{ бит/с.}$$

Таким образом, путём укрепления алфавита увеличить производительность источника нельзя, так как в этом случае и энтропия, и длительность сообщения одновременно возрастают в одинаковое число раз (n).

Увеличить производительность можно путём уменьшения длительности элементов сообщения, однако возможность эта ограничивается полосой пропускания канала связи. Поэтому производительность источника можно увеличить за счет более экономного использования полосы пропускания, например, путем применения сложных многоуровневых сигналов.

2.6 Сложная энтропия

2.6.1 Энтропия объединения

Объединением называется совокупность двух и более взаимозависимых ансамблей дискретных случайных переменных.

Энтропия объединения используется для вычисления энтропии совместного появления статистических зависимых сообщений. Например, передав сто раз цифру 5 по каналу связи с помехами, заметим, что цифра 5 была принята 90 раз, цифра 6 – 8 раз и цифра 4 – 2 раза. Неопределенность возникновения комбинаций вида 5 – 4, 5 – 5, 5 – 6 при передаче цифры 5 может быть описана при помощи энтропии объединения. $H(A, B)$ – неопределенность того, что будет послано A, а принято B. Для ансамблей переданных сообщений A и принятых сообщений B энтропия объединения представляет собой сумму вида

$$H(X,Y) = -\sum_i \sum_j p(x_i, y_j) \log p(x_i, y_j) \quad \text{бит/два символа.}$$

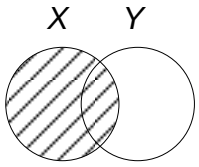
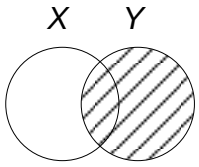
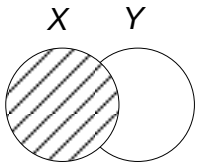
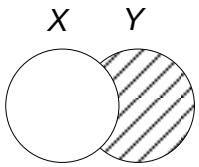
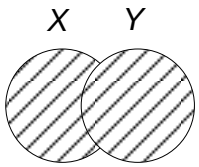
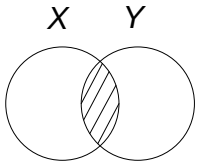
Энтропия объединения и условная энтропия связаны между собой следующими соотношениями:

$$H(X,Y) = H(X) + H(Y/X) = H(Y) + H(X/Y);$$
$$H(Y/X) = H(X,Y) - H(X); \quad H(X/Y) = H(X,Y) - H(Y).$$

В таблице 2.3 схематично представлена энтропия объединения.

Табл. 2.3.

Энтропия объединения

Наименование	Обозначение	Соотношение	Диаграмма
Безусловная энтропия	$H(X)$	$H(X) \geq H(X \mid Y)$ $H(X) = H(X \mid Y) + H(X \cdot Y)$	
	$H(Y)$	$H(Y) \geq H(Y \mid X)$ $H(Y) = H(Y \mid X) + H(X \cdot Y)$	
Условная энтропия	$H(X \mid Y)$	$H(X \mid Y) = H(X) - H(X \cdot Y)$	
	$H(Y \mid X)$	$H(Y \mid X) = H(Y) - H(X \cdot Y)$	
Совместная энтропия	$H(X,Y) = H(Y,X)$	$H(X,Y) = H(X) + H(Y \mid X) =$ $= H(Y) + H(X \mid Y) =$ $= H(X) + H(Y) - H(X \cdot Y)$	
Взаимная энтропия	$H(X \cdot Y) = H(Y \cdot X)$	$H(X \cdot Y) = H(X) - H(X \mid Y) =$ $= H(Y) - H(Y \mid X) = H(X,Y) -$ $- H(X \mid Y) - H(Y \mid X)$	

Энтропия объединения может быть подсчитана при помощи матрицы вида

$$p(x_i, y_j) = \begin{vmatrix} p(x_1, y_1) & p(x_1, y_2) & \dots & p(x_1, y_m) \\ p(x_2, y_1) & p(x_2, y_2) & \dots & p(x_2, y_m) \\ \dots\dots\dots & \dots\dots\dots & \dots & \dots\dots\dots \\ p(x_m, y_1) & p(x_m, y_2) & \dots & p(x_m, y_m) \end{vmatrix}.$$

Такая матрица обладает замечательным свойством:

$$\sum_i p(x_i, y_j) = p(y_j); \ \sum_j p(x_i, y_j) = p(x_i),$$

при этом

$$\sum_i p(x_i) = \sum_j p(y_j) = 1.$$

Это свойство, в свою очередь, позволяет вычислять энтропию как источника, так и приемника сообщений непосредственно по канальной матрице

$$H(X) = -\sum_i \sum_j p(x_i, y_j) \log \sum_j p(x_i, y_j);$$
$$H(Y) = -\sum_i \sum_j p(y_j, x_i) \log \sum_i p(y_j, x_i).$$

Суммирование производится по i и j, т. к. для того, чтобы найти безусловные вероятности, необходимо суммировать их по одной координате (имея в виду матричное представление вероятностей), а для нахождения H суммирование производится по другой координате.

Условные вероятности вида $p(x_i / y_j)$ и $p(y_j / x_i)$ вычисляются как

$$p(x_i, y_j) = \frac{p(x_i, y_j)}{\sum_i p(x_i, y_j)}; \quad p(y_j / x_i) = \frac{p(x_i, y_j)}{\sum_j p(x_i, y_j)}. \qquad (2.11)$$

Количество информации на символ сообщения, переданного по каналу связи, в котором влияние помех описывается при помощи энтропии объединения, подсчитывается следующим образом

$$I(X,Y) = H(X) + H(Y) - H(Y,X). \qquad (2.12)$$

Пример. В результате статистических испытаний установлено, что при передаче каждых 100 сообщений длиной по 5 символов в сообщении символ *К* встречается 50 раз, а символ *Т* – 30 раз. Вместе с символом *К* символ *Т* встречается 10 раз. Определить условные энтропии *H(К/Т)* и *H(Т/К).*

Решение.

Общее количество переданных символов $n = 100 \cdot 5 = 500$; Вероятность появления символа *К* составляет

$$p(K) = \frac{50}{500} = 0{,}1,$$

вероятность появления символа *Т* –

$$p(T) = \frac{30}{500} = 0{,}06,$$

вероятность совместного появления символа *К* и *Т* –

$$p(KT) = \frac{10}{500} = 0{,}02.$$

Так как

$$p(KT) = p(T)\,p(K/T) = p(K)\,p(T/K),$$

то условная вероятность появления символа *К* относительно *Т*

$$p(K/T)=\frac{p(KT)}{p(T)}=\frac{0,02}{0,06}=0,33.$$

Условная вероятность появления символа *Т* относительно символа *К* –

$$p(T/K)=\frac{p(KT)}{p(K)}=\frac{0,02}{0,1}=0,2.$$

Условная энтропия символа *К* относительно *Т* –

$$H(K/T)=-\sum p(y_j/x_i)\log_2 p(y_j/x_i)=-\{p(K/T)\log_2 p(K/T)+$$

$$+[1-p(K/T)]\log_2[1-p(K/T)]\}=-(0,33\log_2 0,33+0,67\log_2 0,67)=$$

$$=0,9149 \text{ бит/символ.}$$

Условная энтропия появления символа *Т* относительно *К*

$$H(T/K)=-(0,2\log_2 0,2+0,8\log_2 0,8)=0,7219 \text{ бит/символ.}$$

Введенные понятия можно проиллюстрировать примером передачи информации по схеме (рис. 2.7).

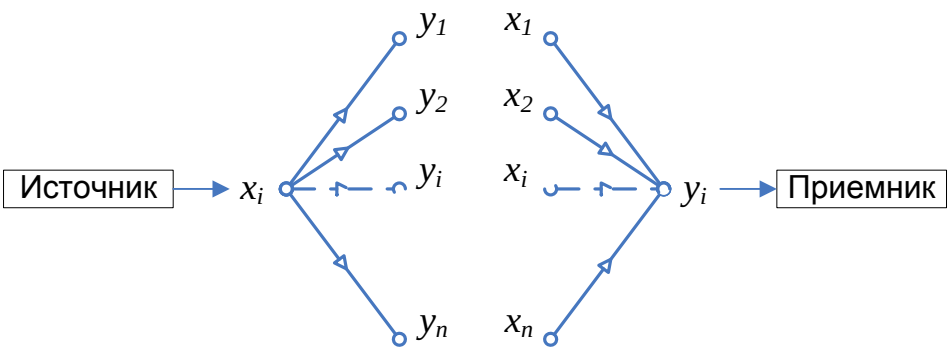


Рис. 2.7. Схема передачи дискретной информации по каналу связи

Последовательность символов $x_1, x_2, ..., x_i, ..., x_n$, создаваемая источником, может претерпевать искажения по пути к приемнику.

Символ x_i может быть принят не только как однозначно ему соответствующий символ y_i, но и как любой из возможных символов $y_1, y_2, ..., y_i, ..., y_m$ с соответствующими вероятностями.

Различные виды энтропии в данной схеме имеют следующий смысл:

$H(x)$ — *безусловная энтропия источника*, или среднее количество информации на символ, выдаваемое источником;

$H(y)$ — *безусловная энтропия приемника*, или среднее количество информации на символ, получаемое приемником;

$H(x, y)$ — *взаимная энтропия* системы передачи — приема в целом, или средняя информация на пару (переданного и принятого) символов;

$H(Y|X)$ — *условная энтропия* Y относительно X, или мера количества информации в приемнике, когда известно, что передается X;

$H(X|Y)$ — *условная энтропия* X относительно Y, или мера количества информации об источнике, когда известно, что принимается Y.

Если в системе нет потерь и искажений, то условные энтропии равны нулю:

$$H(Y|X) = 0, \quad H(X|Y) = 0,$$

а количество взаимной информации равно энтропии либо источника, либо приемника:

$$I(X,Y) = H(X,Y) = H(X) = H(Y).$$

На основании статистических данных могут быть установлены вероятности событий $y_1, y_2, ..., y_m$ при условии, что имели место события x_i, а именно $p(y_1|x_i), p(y_2|x_i), ..., p(y_m|x_i)$. Тогда частная энтропия будет равна:

$$H(Y|x_i) = -\sum_{i=1}^{n} p(y_i|x_i) \cdot \log_2 p(y_i|x_i) \qquad (2.13)$$

Далее нужно подсчитать среднее значение $H(Y|x_i)$ для всех переданных символов *x*. Это будет условная энтропия канала

$$H(Y|X) = -\sum_{i=1}^{n} p(x_i) \cdot H(Y|x_i), \tag{2.14}$$

или в развернутом виде

$$H(Y|X) = -\sum_{i=1}^{n} \sum_{j=1}^{m} p(x_i, y_j) \cdot \log_2(y_j|x_i). \tag{2.15}$$

Аналогично получается условная энтропия $H(X|Y)$, учитывающая условные вероятности $p(x_i|y_i)$

$$H(X|Y) = -\sum_{j=1}^{m} p(y_j) \cdot H(X|y_j).$$

или

$$H(X|Y) = -\sum_{i=1}^{n} \sum_{j=1}^{m} p(x_i, y_j) \cdot \log_2 p(x_i|y_j).$$

Безусловная энтропия ансамбля *X*

$$H(X) = -\sum_{i=1}^{n} p(x_i) \cdot \log_2 p(x_i). \tag{2.16}$$

Безусловная энтропия ансамбля *Y*

$$H(Y) = -\sum_{j=1}^{m} p(y_j) \cdot \log_2 p(y_j). \tag{2.17}$$

2.6.2. Условная энтропия

Понятие условной энтропии в теории информации используется при определении взаимозависимости[1] между символами коди-

---

[1] Суть взаимозависимости символов букв алфавита заключается в том, что вероятность появления *i*-й буквы в любом месте сообщения зависит от того, какие буквы стоят перед ней и после нее, и будет отличаться от безусловной вероятности p_i, известной из статистических свойств данного алфавита.

руемого алфавита, для определения потерь при передаче информации по каналам связи, при вычислении энтропии объединения.

Во всех случаях при вычислении условной энтропии в том или ином виде используются условные вероятности.

Если при передаче n сообщений символ X появился m раз, символ Y появился l раз, а символ X вместе с символом Y – $к$ раз, то вероятность появления символа X составляет $p(X)=\frac{m}{n}$; вероятность появления символа Y $p(Y)=\frac{l}{n}$; вероятность совместного появления символов X и Y $p(XY)=\frac{k}{n}$; условная вероятность появления символа X относительно символа Y и условная вероятность появления символа Y относительно символа X составляет

$$p(X/Y)=\frac{p(XY)}{p(Y)}=\frac{k}{l};\quad p(Y/X)=\frac{p(XY)}{p(Y)}=\frac{k/n}{m/n}=\frac{k}{m}.$$

Если известна условная вероятность, то можно легко определить и вероятность совместного появления символов X и Y, используя приведенное выше выражение:

$$p(XY)=p(Y)p(X/Y)=p(X)p(Y/X).$$

От шенноновского определения энтропии формула условной энтропии отличается тем, что в ней вероятности условные:

$$H(y_j/x_i)=-\sum_j p(y_j/x_i)\log p(y_j/x_i);$$
$$H(x_i/y_j)=-\sum_i p(x_i/y_j)\log p(x_i/y_j),$$

где индекс i выбран для характеристики произвольного состояния источника сообщения X, индекс j выбран для характеристики произвольного состояния адресата Y.

Различают понятия частной и общей условной энтропии. Приведенные выше выражения представляют собой *частные условные энтропии*.

Общая условная энтропия сообщения *Y* относительно сообщения *X* характеризует количество информации, содержащейся в любом символе алфавита, и определяется усреднением по всем символам, т. е. по всем состояниям с учетом вероятности появления каждого из состояний, и равна сумме вероятностей появления символов алфавита на неопределенность, которая остается после того, как адресат принял сигнал

$$H(Y/X) = -\sum_i p(x_i) H(y_j / x_i) = -\sum_i \sum_j p(x_i) p(y_j / x_i) \log p(y_j / x_i)$$

Такое выражение является общим выражением для определения количества информации на один символ сообщения для случая неравномерных и взаимонезависимых символов.

Так как $p(x_i) p(y_j / x_i)$ представляет собой вероятность совместного появления двух событий $p(x_i, y_j)$, то предыдущую формулу можно записать следующим образом:

$$H(Y/X) = -\sum_i \sum_j p(x_i, y_j) \log p(y_j / x_i).$$

Понятие общей и частной условной энтропии широко используется при вычислении информационных потерь в каналах связи с шумами.

В общем случае, если мы передаем *m* сигналов *X* и ожидаем получить *m* сигналов *Y*, влияние помех в канале связи полностью описывается *канальной матрицей*, которая приведена в таблице 2.4.

Вероятности, которые расположены по диагонали, определяют правильный прием, остальные – ложный. Значение цифр, заполняющих колонки канальной матрицы, обычно уменьшаются по мере удаления от главной диагонали и при полном отсутствии помех всех, кроме цифр, расположенных на главной диагонали, равны нулю.

Табл. 2.4.

Канальная матрица

Y / X	y_1	y_2	…	y_j	…	y_m
x_1	$p(y_1/x_2)$,	$p(y_2/x_1)$, …,		$p(y_j/x_1)$,	…,	$p(y_m/x_1)$
x_2	$p(y_1/x_2)$,	$p(y_2/x_2)$, …,		$p(y_j/x_2)$,	…,	$p(y_m/x_2)$
…				…		
x_i	$p(y_1/x_i)$,	$p(y_2/x_i)$, …,		$p(y_j/x_i)$,	…,	$p(y_m/x_2)$
…				…		
x_m	$p(y_1/x_m)$,	$p(y_2/x_m)$,	…,	$p(y_j/x_m)$, …,		$p(y_m/x_m)$

Если описывать канал связи со стороны источника сообщений, то прохождение данного вида сигнала в данном канале связи описывается распределением условных вероятностей вида $p(y_j/x_i)$. Так для сигнала x_i распределение имеет вид

$$p(y_1/x_1)+p(y_2/x_1)+\ldots+p(y_j/x_i)+\ldots+p(y_m/x_1).$$

Сумма вероятностей приведенного распределения всегда равно 1. Потери информации, которые приходятся на долю сигнала x_1, описываются при помощи *частной условной энтропии* вида

$$H(y_j/x_1)=-\sum_j p(y_j/x_1)\log p(y_j/x_1).$$

Суммирование производится по j, т. к. i–е состояние (в данном случаи первое) остается постоянным.

Чтобы учесть потери при передаче всех сигналов по данному каналу связи, следует просуммировать все частные условные энтропии, т. е. произвести двойное суммирование по i и j. При этом в случае равновероятностных появлений сигналов на выходе источника сообщений

$$H(Y/X)=-\frac{1}{m}\sum_j p(y_j/x_i)\log p(y_j/x_i),$$

на m делим, так как энтропия есть неопределенность на один символ).

В случае неравновероятного появления символов источника сообщений следует учесть вероятность появления каждого символа, умножив на нее соответствующую частную условную энтропию. При этом общая условная энтропия

$$H(Y/X) = -\sum_i p(x_i)\sum_j p(y_j/x_i)\log p(y_j/x_i).$$

Если мы исследуем канал связи со стороны приемника сообщений, то с получением сигнала y_j предполагаем, что был послан какой–то из сигналов $x_1, x_2, ..., x_i, ..., x_m$. При этом канальная матрица будет иметь вид (табл. 2.4).

Табл. 2.5.

Канальная матрица

Y / X	y_1	y_2	...	y_j	...	y_m
x_1	$p(x_1/y_2)$,	$p(x_2/y_1)$, ...,		$p(x_2/y_j)$,	...,	$p(x_1/y_m)$
x_2	$p(x_1/y_2)$,	$p(x_2/y_2)$, ...,		$p(x_2/y_j)$,	...,	$p(x_1/y_m)$
...	.....					
x_i	$p(x_i/y_1)$,	$p(x_i/y_2)$, ...,		$p(x_i/y_j)$,	...,	$p(x_i/y_m)$
...	.....					
x_m	$p(x_m/y_1)$,	$p(x_m/y_2)$, ...,		$p(x_m/y_j)$,	...,	$p(x_m/y_m)$

В этом случае единице должны равняться суммы условных вероятностей не по строкам, а по столбцам канальной матрицы

$$p(x_1/y_j) + p(x_2/y_j) + ... + p(x_i/y_j) + ... + p(x_m/y_j) = 1.$$

Частная условная энтропия

$$H = (x_i / y_j) = -\sum_{i=1}^{m} p(x_i / y_j) \log p(x_i / y_j),$$

а общая условная энтропия

$$H(X/Y) = -\sum_j p\left(y_j\right) \sum_i p\left(x_i / y_j\right) \log p\left(x_i / y_j\right)$$

Так как

$$p(x_i) p(y_j / x_i) = p(x_i, y_j),$$

то для вычисления общей условной энтропии может быть использовано следующее выражение:

$$H(Y/X) = -\sum_i \sum_j p(x_i, y_j) \log p(y_j / x_i).$$

Если заданы канальная матрица вида $p(y_j / x_i)$ и безусловные вероятности вида $p(x_i)$, то безусловные вероятности приемника $p(y_j)$ находим как $\sum_i p(x_i) p(y_j / x_i)$, т. е. если заданы безусловные вероятности источника и канальная матрица, то может быть вычислена энтропия приемника

$$H(Y) = -\sum_j j p(y_j) \log p(y_j),$$

и наоборот, если заданы вероятности вида $p(y_j)$ и канальная матрица, описывающая канал связи со стороны приемника сообщений, то $p(x_i) = \sum_j p(y_j) p(x_i / y_j),$ а значит может быть определена энтропия источника сообщений

$$H(X) = -\sum_i p(x_i) \log p(x_i).$$

Если взаимозависимость связывает 3 элемента x_i, x_j, x_k, то условная энтропия вычисляется по формуле

$$H(X/Y,K) = -\sum_i \sum_j \sum_k p(x_i, x_j, x_k) \log p(x_i, x_j, x_k),$$

аналогично и для 4, 5, …, n элементов.

2.6.3. Основные свойства сложной энтропии

1. Если сообщения независимы, то их совместная энтропия равна сумме энтропий каждого сообщения:

$$H(x, y) = H(x) + H(y);$$
$$H(y, x) = H(y) + H(x).$$

Это обусловлено тем, что для независимых сообщений условная вероятность равна безусловной:

$$p(x, y) = p(x) \cdot p(y).$$

2. Если сообщения x и y зависимы, то их совместная энтропия равна безусловной энтропии одного из сообщений, т. к. условные вероятности зависимых событий могут быть равны 0 или 1, то

$$p(y_j/x_i) \log p(y_j/x_i) = 0|1 \quad \text{и} \quad p(x_i/y_j) \log p(x_i/y_j) = 0|1,$$
$$H(x, y) = H(x) = H(y).$$

3. Условная энтропия может изменяться в пределах:

$$0 \le H(y/x) \le H(y)$$

Это свойство определяется тем, что условная энтропия неотрицательна, равна нуля для зависимых сообщений, максимальна для независимых сообщений и равна безусловной энтропии.

4. Для совместной энтропии всегда справедливо соотношение: совместная энтропия не может быть больше суммы энтропий x и y

$$H(x, y) \le H(x) + H(y).$$

2.6.4. Совместная энтропия двух источников

Пусть имеется два дискретных источника с энтропиями $H(x)$ и $H(y)$ и объёмами алфавитов k и l (рис. 2.8).

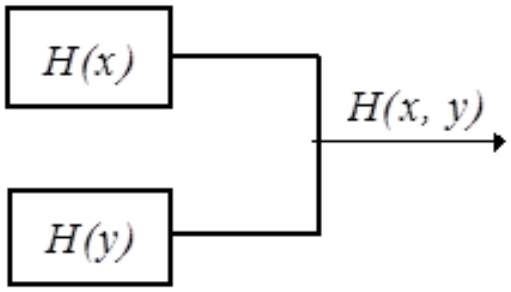


Рис. 2.8. Совместная энтропия двух источников

Объединим оба эти источника в один сложный источник и определим совместную энтропию. Элементарное сообщение на выходе системы содержит элементарное сообщение x_i и сообщение y_j. Алфавит сложной системы будет иметь объём $k \cdot l$, а энтропия будет равна

$$H(x, y) = m\{-\log p(x, y)\}, \qquad (2.14)$$

или

$$H(x, y) = \sum_{i=1}^{K} \sum_{j=1}^{l} p(x_i y_j) \log p(x_i y_j).$$

По теореме умножения вероятностей

$$p(x, y) = p(x) \cdot p(y/x) = p(y) \cdot p(x/y),$$

Подставляя эти соотношения в (2.14), получим

$$H(x,y)=m\{-\log p(x)-\log p(y/x)\}=H(x)+H(y/x).$$

Аналогично можно получить

$$H(x,y)=H(y)+H(x/y),$$

где $H(x)$ и $H(y)$ – собственная энтропия источников x и y;

$H(y/x)$ – условная энтропия источника y относительно источника x,

$$H(y/x)=-\sum_i\sum_j p(x_i y_j)\log p(y_j/x_i).$$

Она показывает, какую энтропию имеют сообщения y, когда уже известно сообщение x.

Если источники независимы, то

$$p(y/x) = p(y) \text{ и } H(y/x)=H(y).$$

В этом случае

$$H(x,y) = H(x) + H(y).$$

Если источники частично зависимы, то

$$H(x,y) < H(x)+H(y).$$

Если источники полностью зависимы (x и y содержат одну и ту же информацию), то

$$H(y/x) = 0 \text{ и } H(x,y) = H(x)=H(y).$$

2.6.5. Взаимная информация источников сообщений

На рисунке 2.9 показана условно собственная энтропия $H(x)$ и $H(y)$, условные энтропии $H(x/y)$ и $H(y/x)$ и совместная энтропия $H(x, y)$.

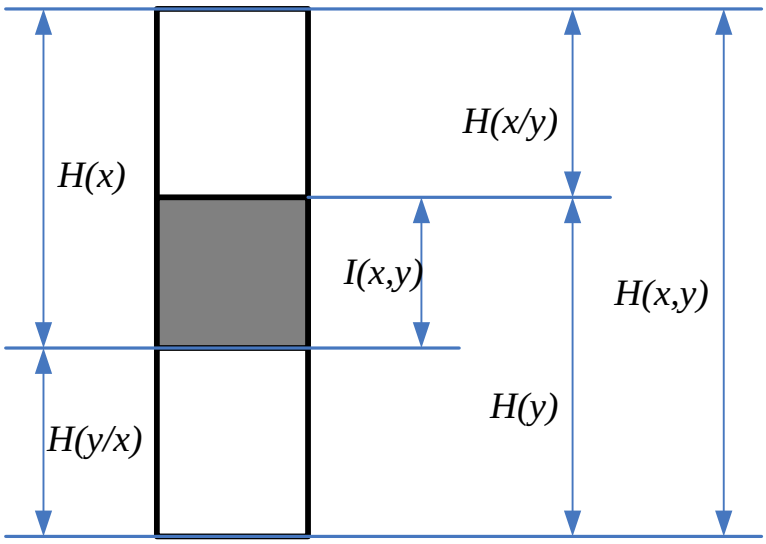


Рис. 2.9. Понятие сложной энтропии

Часть рисунка 2.9, отмеченная штриховкой, называется взаимной информацией $I(x, y)$. Она показывает, какое (в среднем) количество информации содержит сообщение x о сообщении y (или наоборот, сообщение y о сообщении x).

Как следует из рисунка 2.9,

$$I(x,y)=H(x)-H(x/y)=H(y)-H(y/x).$$

Если сообщение x и y полностью независимы, то взаимная информация отсутствует и $I(x, y) = 0$.

Если x и y полностью зависимы (x и y – содержат одну и ту же информацию), то $I(x, y) = H(x)=H(y)$.

Понятие взаимной информации очень широко используется в теории передачи информации. Требования к взаимной информации различны в зависимости от того, с какой информацией мы имеем дело. Например, если x и y – это сообщения, публикуемые различными газетами, то для получения возможно большей суммарной

(совместной) информации взаимная (т. е. одинаковая в данном случае) информация должна быть минимальной.

Если же *x* и *y* – это сообщения на входе и на выходе канала связи с помехами, то для получения возможно большей информации её получателем необходимо, чтобы взаимная информация была наибольшей. Тогда условная энтропия *H*(*x/y*) – это потери информации в канале связи (*ненадежность канала*), *H*(*y/x*) – это информация о помехах (энтропия источника помех – *H*(*n*)), поступающая в канал извне или создаваемая внутренними помехами в канале (схематически этот процесс показан на рисунке 2.10).

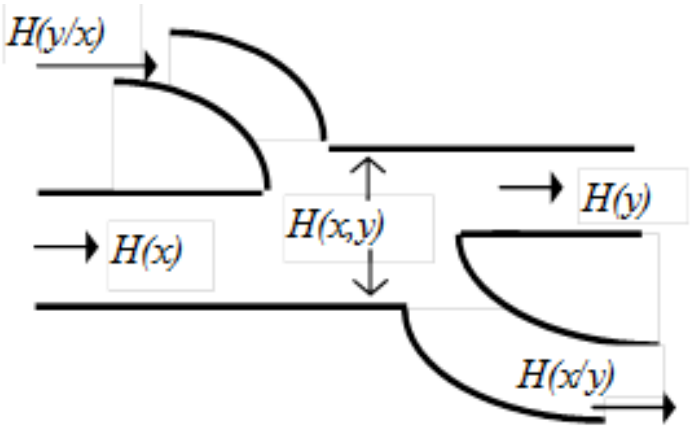


Рис. 2.10. Условная энтропия в канале связи с помехами

2.7. Энтропия непрерывных сообщений

Исходные данные часто представляются в виде непрерывных величин, например, температура воздуха или морской воды. Поэтому представляет интерес измерение количества содержащейся в таких сообщениях информации.

Как было сообщено выше, средняя энтропия дискретного сообщения, входящего в полную группу из *m* независимых сообщений, находится по формуле

$$H = -\sum_{i=1}^{m} P(a_i)\log_2 P(a_i).$$

Эта формула справедлива, в частности, для сообщений, состоящих из конечного числа букв (в нашем случае *m* штук).

Если рассматривать попадание значения непрерывного сообщения в определенный интервал как появление определенной буквы, то непрерывное сообщение можно рассматривать как предельный случай дискретного, если число букв устремить к бесконечности (рис. 2.11).

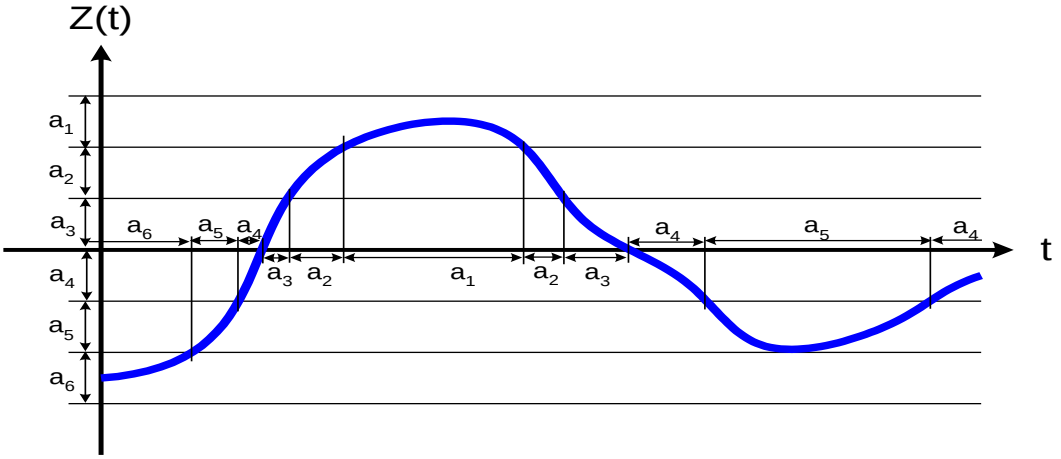


Рис. 2.11. Представление непрерывного процесса в виде дискретной последовательности букв.

Для вычисления энтропии дискретного сообщения необходимо было знать множество букв и вероятности их появления. Аналогом набора вероятностей букв в непрерывном сообщении является плотность вероятностей $f(a)$ распределения значений непрерывного сообщения (рис. 2.12).

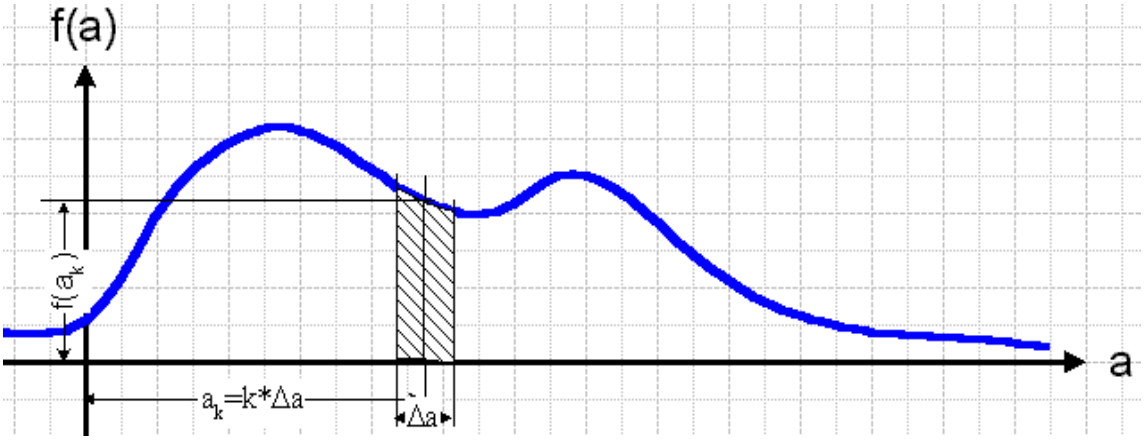


Рис. 2.12. Вычисление вероятности попадания значений непрерывного процесса в некоторый интервал шириной Δa

Непрерывную величину можно считать предельным случаем квантованной по уровню, если размер шага квантования Δa устремить к нулю.

Зная плотность распределения вероятности (дифференциальный закон распределения) и предполагая независимость соседних значений случайной величины, можно найти, используя известную формулу, среднюю энтропию одного ее значения:

$$H(\Delta a) = -\sum_{k=1}^{\infty} P(a_k)\log_2 P(a_k),$$

где $a_k = k \cdot \Delta a$

Из теории вероятностей известно, что вероятность попадания случайной величины в некоторый интервал находится через дифференциальный закон ее распределения по формуле

$$P(a_k) = \int_{a_k - da/2}^{a_k + da/2} f(a)da$$

При достаточно гладкой функции $P(a)$ и малом Δa можно считать, что

$$\int_{a_k - da/2}^{a_k + da/2} f(a)da \approx f(a_k)\Delta a,$$

поэтому

$$H_{непр} = \lim_{\Delta a \to \infty} H(\Delta a) = \lim_{\Delta a \to \infty}\left\{-\sum_k f(a_k)\Delta a \log_2\left[f(a_k)\Delta a\right]\right\} =$$

$$= -\lim_{\Delta a \to 0}\left\{\Delta a \sum f(a_k)\log_2\left[f(a_k)\right]\right\} - \lim_{\Delta a \to 0}\left\{\Delta a \log_2\left[\Delta a\right]\sum_k f(a_k)\right\} .$$

Первое слагаемое в пределе превращается в интеграл:

$$\lim_{\Delta a \to 0}\left\{-\sum_{k} f(a_k)\log_2\left[f(a_k)\right]\Delta a\right\} = \int_{-\infty}^{\infty} f(a)\log_2 f(a)da\,.$$

Во втором слагаемом тоже образуется интеграл. Это определенный интеграл от дифференциального закона распределения с бесконечными пределами, который, как известно, равен 1. В результате получаем:

$$\lim_{\Delta a \to 0}\left\{\log_2\left[\Delta a\right]\sum_{k} f(a_k)\Delta a\right\} = \lim_{\Delta a \to 0}\log_2\left[\Delta a\right]\int f(a)da = \lim_{\Delta a \to 0}\log_2\left[\Delta a\right].$$

Ясно теперь, что первое слагаемое – конечное число, а второе – бесконечное. Как и следовало ожидать, энтропия непрерывного сообщения в математическом смысле бесконечно велика, что означает, что непрерывный сигнал несет в себе бесконечно большое количество шенноновской информации. Однако в реальной жизни это не так. Любое непрерывное сообщение (например, показания термометра за окном или пружинных весов), человек так или иначе подвергает квантованию, что приводит к уменьшению количества информации по Шеннону.

Таким образом, в математическом смысле непрерывные сигналы–сообщения не имеют абсолютной меры энтропии. Поэтому для них вводится понятие *относительной энтропии*. Это понятие подразумевает наличие эталона, с которым сравнивается энтропия непрерывного сообщения. В качестве такого эталона выбирается также непрерывное сообщение–сигнал, имеющий равномерный закон распределения в интервале ε (эпсилон). График равномерного закона приведен на рисунке 2.13.

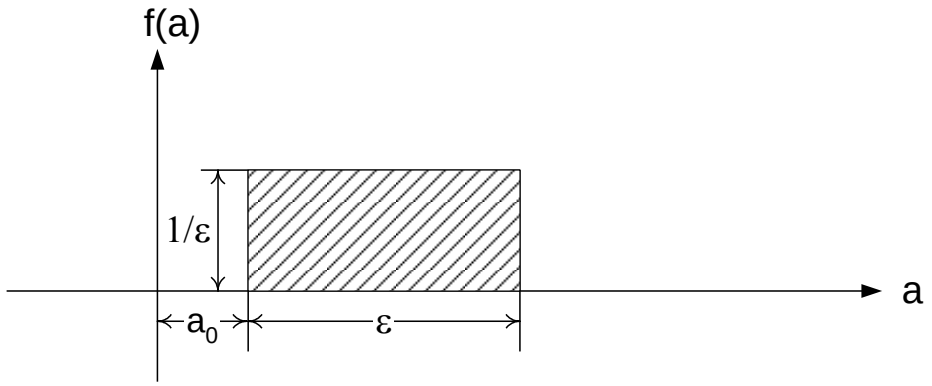


Рис. 2.13. График равномерного закона распределения

Абсолютная энтропия такого эталонного сообщения равна

$$H(\varepsilon) = \int_{a_0}^{a_0+\varepsilon} \frac{1}{\varepsilon} \log_2 \varepsilon \, da - \lim_{\Delta a \to 0} \log_2 \Delta a = \log_2 \varepsilon - \lim_{\Delta a \to 0} \log_2 \Delta a \ .$$

Из–за второго члена $H(\varepsilon)$ также бесконечно велика.

Относительная энтропия непрерывного сообщения в данном случае определяется как разность между абсолютной энтропией $H_{\text{непр}}$ этого сообщения и абсолютной энтропией $H(\varepsilon)$ эталонного сообщения. Ее принято называть *дифференциальной эпсилон–энтропией* непрерывного сообщения. Она обозначается H_ε и вычисляется по формуле

$$H_\varepsilon = \lim_{\Delta a \to 0} [H_{непр} - H(\varepsilon)] = -\int_{-\infty}^{\infty} f(a) \log_2 f(a) da - \log_2 \varepsilon =$$

$$= -\int_{-\infty}^{\infty} f(a) \log_2 [\varepsilon \cdot f(a)] da.$$

Если положить $\varepsilon = 1$, то дифференциальная энтропия (слово *эпсилон* часто не произносится) запишется в виде:

$$H_{\varepsilon=1} = -\int_{-\infty}^{\infty} f(a) \log_2 [f(a)] da \ .$$

Видим, что формула дифференциальной энтропии похожа на формулу энтропии дискретного источника, но следует помнить, что она описывает не абсолютное, а относительное значение энтропии, когда за стандарт взята случайная величина, равномерно распределенная на единичном интервале.

Дифференциальная энтропия позволяет сравнивать различные непрерывные источники сообщений по степени информативности.

2.7.1. Экстремальные свойства энтропии непрерывных сообщений

Представляет интерес решение следующей задачи. Заданы некоторые ограничения на параметры закона распределения непрерывного сообщения. Требуется с учетом этих ограничений найти такой закон распределения, при котором дифференциальная энтропия максимальна.

Решение подобных задач имеет практическое значение. Знание таких законов полезно при исследовании космоса на предмет поиска в нем сигналов, имеющих искусственное происхождения, при постановке максимально эффективных помех в заданной полосе частот с помощью передатчика с ограниченной мощностью (например, при подавлении сотовой связи) и т. д.

В первом случае значения случайной величины a ограничены интервалом $[b,c]$. В этом случае определенный интеграл ее плотности распределения вероятностей (дифференциального закона распределения вероятностей) на этом интервале равен 1:

$$\int_b^c f(a)da = 1 \ .$$

Найдем аналитическое выражение для закона $f(a)$, который дает максимум дифференциальной энтропии

$$H_{\varepsilon=1} = -\int_b^c f(a)\log_2[f(a)]da \, .$$

Эту задачу можно интерпретировать как предельный случай задачи на поиск условного экстремума функции бесконечного количества аргументов. Эти аргументы – точки искомой плотности распределения вероятностей $f(a)$. Тогда аналитическое выражение для закона $f(a)$ является условием, которое нужно соблюсти при поиске экстремума.

Для решения этой задачи воспользуемся вышеописанным методом множителей Лагранжа, который для непрерывного случая переходит в вариационное исчисление.

Вспомогательная функция Φ в нашем случае будет иметь вид:

$$\Phi = -\int_b^c f(a)\log_2[f(a)]da + \lambda\left[\int_b^c f(a)da - 1\right].$$

Дифференцируя по $f(a)$, получаем систему уравнений:

$$\begin{cases} -\left[\dfrac{1}{\ln 2} + \dfrac{\ln f(a)}{\ln 2}\right]\cdot da + \lambda\cdot da = 0 \\ \int_b^c f(a)da = 1 \end{cases}.$$

Из первого уравнения получим

$$f(a) = \exp\left[\lambda\cdot\ln(2) - 1\right].$$

Подставим $f(a)$ во второе уравнение, учитывая, что $f(a)$ от a не зависит:

$$\int_b^c f(a)da = f(a)\int_b^c da = f(a)(c-b) = 1.$$

Отсюда

$$f(a)=\begin{cases}\dfrac{1}{c-b} & при \quad b\le a\le c\\ 0 & в \quad остальных \quad случаях\end{cases}.$$

В итоге мы получили уже известный для дискретных сообщений результат, указывающий на то, что максимум энтропии достигается, если все значения сигнала равновероятны, т. е. такой сигнал имеет *равномерный закон распределения*.

Во втором случае предположим, что область определения значений случайной величины не ограничена, но задана ее дисперсия D и математическое ожидание M. Заметим, что дисперсия прямо пропорциональна мощности сигнала.

Из этого вытекают следующие условия, накладывающие ограничения на искомый закон распределения:

$$\int_{-\infty}^{\infty}(a-M)^2 f(a)da=D;$$

$$\int_{-\infty}^{\infty} af(a)da=M;$$

$$\int_{-\infty}^{\infty} f(a)da=1.$$

Опять будем искать функцию $f(a)$, при которой дифференциальная энтропия принимает максимальное значение. Еще раз используем метод множителей Лагранжа. Вспомогательная функция Φ в этом случае имеет вид

$$\Phi=-\int_{-\infty}^{\infty} f(a)\log_2[f(a)]da+\lambda_1\left[\int_{-\infty}^{\infty}(a-M)^2 f(a)da-D\right]+$$
$$+\lambda_2\left[\int_{-\infty}^{\infty} af(a)da-M\right]+\lambda_3\left[\int_{-\infty}^{\infty} f(a)da-1\right].$$

Дифференцируя по $f(a)$, получим систему уравнений:

$$\begin{cases} -\left(\dfrac{1}{\ln 2}-\dfrac{\ln P(a)}{\ln 2}\right)da+\lambda_1(a-M)^2da+\lambda_2 ada+\lambda_3 da=0 \\ \int\limits_{-\infty}^{\infty}(a-M)^2 f(a)da=D \\ \int\limits_{-\infty}^{\infty} af(a)da=M \\ \int\limits_{-\infty}^{\infty} f(a)da=1 \end{cases}$$

Используя первое уравнение, найдем

$$f(a)=\exp\left[\ln 2\cdot(\lambda_1(a-M)^2+\lambda_2 a+\lambda_3)-1\right].$$

Подставим $f(a)$ в третье уравнение системы уравнений:

$$\int\limits_{\infty}^{-\infty} f(a)da=\exp(\ln 2\cdot\lambda_3-1)\int\limits_{-\infty}^{\infty}\exp\left[\ln 2\cdot(\lambda_1(a-M)^2+\lambda_2 a\right]da=1.$$

По таблицам находим значение интеграла:

$$\int\limits_{-\infty}^{\infty}\exp\left[\ln 2\cdot(\lambda_1(a-M)^2+\lambda_2 a\right]da=\exp\left(\lambda_2 M\right)\cdot\exp\left(-\frac{\lambda_2}{4\lambda_1}\sqrt{\frac{\pi}{-\lambda_1}}\right).$$

Подставив найденный интеграл в предыдущее выражение, получим

$$\exp(\ln 2\cdot\lambda_3-1)\exp\left(\lambda_2 M\right)\cdot\exp\left(-\frac{\lambda_2}{4\lambda_1}\sqrt{\frac{\pi}{-\lambda_1}}\right)=$$

$$=\exp\left[\ln 2\cdot\lambda_3-1+\lambda_2 M-\frac{\lambda_2}{4\lambda_1}\sqrt{\frac{\pi}{-\lambda_1}}\right]=1.$$

Значит,

$$\ln 2 \cdot \lambda_3 - 1 + \lambda_2 M - \frac{\lambda_2}{4\lambda_1}\sqrt{\frac{\pi}{-\lambda_1}} = 0;$$

$$\text{и } \lambda_3 = \frac{1}{\ln 2}\left(1 - \lambda_2 M + \frac{\lambda_2}{4\lambda_1}\sqrt{\frac{\pi}{-\lambda_1}}\right).$$

Теперь полученное выражение нужно подставить в $f(a)$, полученный результат подставить во второе условие и найти λ_2. Далее λ_2 подставляется в $f(a)$, а полученный результат – в первое уравнение. Таким образом находится λ_1. Далее λ_1 подставляется в $f(a)$, в результате чего находится искомый закон распределения

$$f(a) = \frac{1}{D\sqrt{2\pi}}\exp\left[-\frac{(a-M)^2}{2D}\right].$$

Видим, что это часто встречающийся в природе нормальный закон распределения. Его вид зависит от 2 параметров – математического ожидания (среднего значения) M и дисперсии D, описывающей степень разброса случайной величины.

На рисунке 2.14 изображены графики этого закона при разных значениях математического ожидания и дисперсии.

Аналогичным образом можно показать, что если случайная величина может принимать только положительные значения, т. е. $0 \le a \le \infty$, то максимальная дифференциальная энтропия получается при распределении сигнала по экспоненциальному закону:

$$f(a) = \begin{cases} \frac{1}{M}\exp\left(-\frac{a}{M}\right) & при \quad a \ge 0 \\ 0 \quad при \quad a < 0 \end{cases},$$

где M – математическое ожидание.

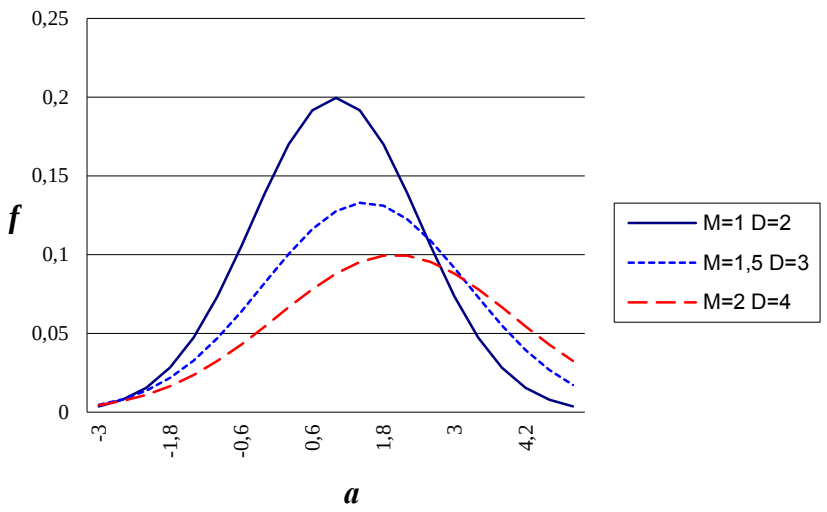


Рис. 2.14. Графики нормального закона распределения

На рисунке 2.15 изображены графики экспоненциального закона распределения при $M = 1$ и $M = 2$.

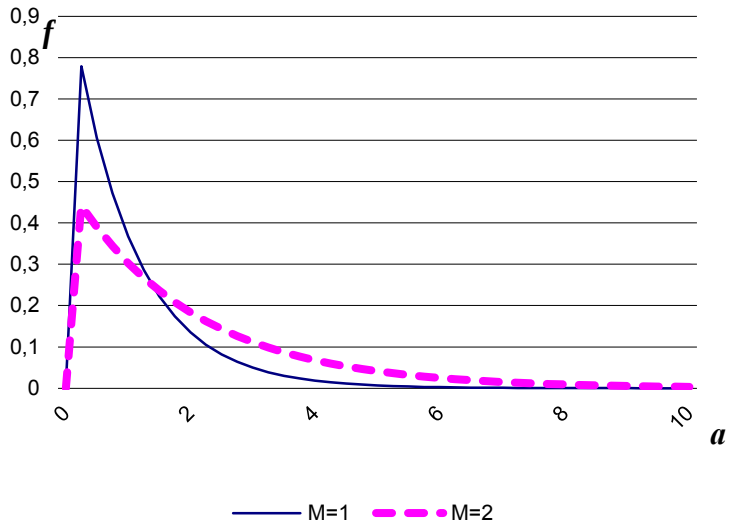


Рис. 2.15. Графики экспоненциального закона распределения

2.7.2. Эпсилон–энтропия источника непрерывных сообщений

Сигнал на выходе источника (рис. 2.16) непрерывных сообщений (например, микрофона, телефона, датчика температуры и пр.) представляет собой непрерывный случайный процесс, энтропия которого в любом из сечений в общем случае равна бесконечности.

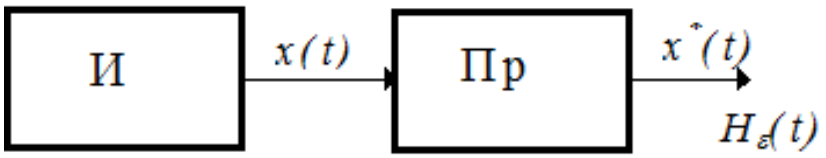


Рис. 2.16. Эпсилон–энтропия непрерывных сообщений

Такое количество информации не может быть передано по реальному каналу связи, пропускная способность которого всегда ограничена. Это и не нужно, т. к. скорость восприятия информации любым потребителем на выходе канала всегда ограничена его физическими возможностями. Поэтому непрерывный сигнал на выходе канала связи даже без помех отличается от сигнала на входе, так как содержит не всю информацию о нем (причем под каналом связи можно понимать любое преобразование одного ансамбля сигналов в другое: модуляцию, усиление, дискретизацию и пр.).

Уже преобразование непрерывного сообщения в сигнал соответствующим устройством (микрофоном или др. датчиком сигнала) связано с потерей части информации, а сигнал отображает сообщение лишь с некоторой точностью

$$\varepsilon(t) = x(t) - x^*(t),$$

где $x(t)$– сигнал на входе преобразователя;

$x^*(t)$– сигнал на выходе преобразователя (Пр) или оценка входного сигнала преобразователем, который всегда представляет собой некоторое решающее устройство, работающее по определенному правилу и заданному критерию качества.

Критерий качества, как известно, определяется потребителем информации, например, среднеквадратическое отклонение

$$\varepsilon_{cp} = m\{[\, x(t) - x^*(t)]^2\} \,,$$

или дисперсия ошибки

$$\sigma_\varepsilon^2 = m\{\varepsilon^2(t)\} \,.$$

Эпсилон–энтропией $H_\varepsilon(x)$ (ε–энтропией) называется минимальное количество информации, которое должно содержаться в выходном сигнале $x^*(t)$ о входном сигнале $x(t)$, чтобы этот сигнал можно было восстановить с заданной точностью $\varepsilon_{ср}$.

$$H_\varepsilon(x) = min\ I(x,x^*) = H(x) - max\ H(x/x^*),$$

где $I(x,x^*)$ – взаимная информация x и x^*;

$H(x)$ и $H(x/x^*)$ – соответственно, дифференциальная энтропия сигнала $x(t)$ и условная энтропия $x(t)$, когда $x^*(t)$ известно;

min и *max* берутся по всевозможным условным распределениям $w(x/x^*)$.

В общем случае, когда сигнал (или сообщение) $x(t)$ является гауссовским с дисперсией σ_x^2, ошибка $\varepsilon(t)$ также является гауссовской с дисперсией σ_ε^2, а с учетом аддитивного характера ошибки $\varepsilon(t)$ условная энтропия $H(x/x^*)$ полностью определяется дифференциальной энтропией $H(\varepsilon)$. Соответственно,

$$\max H\left(x/x^*\right) = \max \mathrm{H}\left(\varepsilon\right) = \log\left(\sigma_\varepsilon\sqrt{2\pi e}\right).$$

Тогда ε–энтропия одного сечения гауссовского источника (ε–энтропия одного отсчета)

$$H\left(x\right) = \log(\sigma_x\sqrt{2\pi e}) - \log(\sigma_\varepsilon\sqrt{2\pi e}) = 0,5\cdot\log(\sigma_x^2/\sigma_\varepsilon^2).$$

Величина $\sigma_x^2/\sigma_\varepsilon^2$ показывает отношение мощности (дисперсии) сигнала $x(t)$ к мощности (дисперсии) ошибки, при котором среднеквадратическое отклонение сигналов $x(t)$ и $x^*(t)$ не превышает σ_ε.

Следовательно, *производительность источника непрерывных сообщений* можно определить как количество информации, которое необходимо передавать в единицу времени, чтобы восстановить сообщение с заданной точностью.

$$H'_{\varepsilon}(x)=v\cdot H_{\varepsilon}(x),$$

где $v = 1/\Delta t$ – скорость передачи отсчетов на выходе источника,
Δt – интервал между отсчетами.

Для стационарного сигнала с ограниченным спектром

$$\Delta t = 1/(2F_{max}),$$

тогда

$$H'_{\varepsilon}(x) = 2F_{max}\cdot H_{\varepsilon}(x).$$

Если, кроме того, источник является гауссовским, то

$$H'(x) = F_{\max}\cdot\log\left(\sigma_x^2/\sigma_\varepsilon^2\right).$$

Количество информации, выдаваемое гауссовским источником за время T_c, равно

$$T_c\cdot H'(x) = T_c\cdot F_{\max}\cdot\log\left(\sigma_x^2/\sigma_\varepsilon^2\right),$$

что совпадает с формулой для объёма сигнала, когда динамический диапазон сигнала

$$D_c = \log\left(\sigma_x^2/\sigma_\varepsilon^2\right).$$

Это значит, что объем сигнала на выходе источника равен количеству информации, которое содержится в сигнале для его воспроизведения с заданной точностью.

Определим информативность (эпсилон–энтропию) случайных величин, распределенных по некоторым наиболее известным законам распределения.

При нормальном законе распределения

$$f(a)=\frac{1}{D\sqrt{2\pi}}\exp\left[-\frac{(a-M)^2}{2D}\right];$$

$$H_{\varepsilon=1}=-\int_{-\infty}^{\infty}f(a)\log_2[f(a)]da=-\frac{1}{\ln 2}\int_{-\infty}^{\infty}f(a)\ln[f(a)]da=$$
$$=\frac{1}{\ln 2}\left[\int_{-\infty}^{\infty}f(a)\ln\left[D\sqrt{2\pi}\right]da+\int_{-\infty}^{\infty}f(a)\frac{(a-M)^2}{2D}da\right].$$

Учтем, что константа $\ln\left[D\sqrt{2\pi}\right]$ может быть вынесена из под знака интеграла в первом слагаемом, а сам интеграл в этом случае равен 1.

Константу $\frac{1}{2D}$ можно вынести из–под знака второго интеграла, а сам интеграл представляет собой дисперсию *D*.

Учитывая все это, получаем:

$$H_{\varepsilon=1}=\frac{1}{\ln 2}\left(\ln\left[D\sqrt{2\pi}\right]+1/2\right)=\frac{1}{\ln 2}\left(\ln\left[D\sqrt{2\pi}\right]+\ln\sqrt{e}\right)=$$
$$=\log_2 D\sqrt{2\pi e}$$

При равномерном законе распределения

$$f(a)=\begin{cases}\frac{1}{c-b} & при\ \ b\le a\le c\\ 0\ \ в\ \ остальных\ \ случаях\end{cases};$$

$$H_{\varepsilon=1}=-\int_{b}^{c}\frac{1}{c-b}\log_2\frac{1}{c-b}da=\log_2(c-b)\ \ .$$

При экспоненциальном законе распределения

$$f(a)=\begin{cases}\frac{1}{M}\exp\left(-\frac{a}{M}\right) & при \quad a\geq 0\\ 0 \quad при \quad a<0\end{cases};$$

$$H_{\varepsilon=1}=-\int_0^\infty \frac{1}{M}\exp\left(-\frac{a}{M}\right)\log_2\left[\frac{1}{M}\exp\left(-\frac{a}{M}\right)\right]da=$$

$$=-\frac{1}{M}\int_0^\infty \exp\left(-\frac{a}{M}\right)\cdot\left(-\frac{a}{M\ln 2}\right)da-\frac{1}{M}\int_0^\infty \exp\left(-\frac{a}{M}\right)\log_2\frac{1}{M}da\ .$$

После преобразований и взятия интегралов (там, где нужно при помощи таблиц интегралов) получим: $H_{\varepsilon=1}=\log_2 Me$.

Найдем значение энтропии, когда состояния элементов распределены по нормальному закону:

$$H=-\frac{1}{e\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{x^2}{2\sigma^2}}\log[\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{x^2}{2\sigma^2}}]dx-\log\Delta x=$$

$$=\log(\sigma\sqrt{2\pi}\frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{x^2}{2\sigma^2}}dx+$$

$$+\frac{\log e}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{x^2}{2\sigma^2}}dx-\log\Delta x)=\log(\frac{\sigma}{\Delta x}\sqrt{2\pi e})$$

Найдем значение энтропии, когда состояния элементов распределены внутри интервала их существования $a \leq x \leq b$ по равномерному закону, т. е.

$$P(x)=\begin{cases}\frac{1}{b-a} \ при\ a\leq x\leq b\\ 0 \ \ при\ x<a, x>b\end{cases}$$

$$H_p(x)=-\frac{1}{b-a}\int_a^b \log(\frac{1}{b-a}-\Delta x)dx=\log\frac{b-a}{\Delta x}$$

Дисперсия равномерного распределения

$$\sigma_p^2 = \frac{(b-a)^2}{12} \ ,$$

поэтому

$$b - a \ = \ 2\sqrt{3}\sigma_p .$$

С учетом этого можно записать

$$H_p(x) = \log(\frac{\sigma_p}{\Delta x} 2\sqrt{3})$$

Сравнивая между собой сообщения с равномерным и нормальным распределением вероятностей при условии $H_н(x) = H_p(x)$, получаем

$$\sigma_p^2 = \frac{e}{6}\sigma^2 \approx 1,42\sigma^2$$

Это значит, что при одинаковой информативности сообщений средняя мощность сигналов для равномерного распределения их амплитуд должна быть на 42% больше, чем при нормальном распределении.

Пример. При организации мешающего воздействия при передаче информации можно использовать источник шума с нормальным распределением плотности и источник, имеющий в некотором интервале равномерную плотность распределения.

Определить, какой источник шума применять экономичнее и каков при этом выигрыш в мощности.

Решение.

Очевидно, что сравнение источников следует проводить из условия обеспечения равенства энтропий, когда каждый источник вносит одинаковое мешающее воздействие при передаче информации.

Как было показано выше, значение энтропии, когда состояния элементов распределены по нормальному закону:

$$H'(x) = \log(\frac{\sigma}{\Delta x}\sqrt{2\pi e}) = \log(\sigma_{\partial}\sqrt{2\pi e}) \ ,$$

где Δx = 1 Ом,

$\frac{\sigma}{\Delta x} = \sigma_{\partial}$, т. е. $\sigma_{г}^{2}$ – дисперсия, характеризующая мощность, выделяемую на резистор с сопротивлением 1 Ом.

Для равномерного распределения

$$H_p(x) = H_p(x) = \log\frac{b-a}{\Delta x}.$$

Так как дисперсия равномерного распределения

$$\sigma_h^2 = \frac{(b-a)^2}{12}, \textit{то } b-a = 2\sqrt{3}\sigma_p, \textit{ и, следовательно,}$$

$$H_p(x) = \log(\frac{\sigma_p}{\Delta x}2\sqrt{3}) = \log(\sigma_{p\partial}2\sqrt{3}), \textit{ где}$$

$$\sigma_{p\partial} = \frac{\sigma_p}{\Delta x}, \Delta x = 1 \textit{ м}$$

Так как $H'(x) = H_p(x)$, то

$$\log(\sigma_{\partial}\sqrt{2\pi e}) = \log(\sigma_{p\partial}2\sqrt{3}),$$

$$2\pi e\sigma_{\partial}^2 = 12\sigma_{p\partial},$$

$$\sigma_{p\partial}^2 = \frac{\pi e}{6}\sigma_{\partial}^2 \approx 1,42\sigma_{\partial}^2$$

Поэтому следует выбирать источник шума с нормальным распределением плотности, т. к. при той же неопределенности, вносимой им в канал связи, можно выиграть в мощности 42%.

2.7.3. Энтропия дискретного сообщения, получаемого из непрерывного путем его квантования по уровню

Для всех трех случаев несложно получить энтропию дискретного сообщения, получаемого из непрерывного путем его квантования по уровню с шагом Δa. Вспомним, то энтропия непрерывной величины равна:

$$H_{непр} = -\int_b^c f(a)\log_2[f(a)]da - \lim_{\Delta a \to 0} \log_2\left[\Delta a\right]$$

При небольших шагах квантования Δa энтропию квантованного сообщения можно находить по формуле:

$$H_{квант} = H_{\varepsilon=1} - \log_2\left[\Delta a\right] \quad .$$

Поэтому выражения энтропии квантованного сообщения при различных законах распределения будут иметь вид:

1) нормальный закон

$$H = \log_2\left(\frac{D}{\Delta a}\sqrt{2\pi e}\right) \quad ;$$

2) равномерный закон:

$$H = \log_2 \frac{(c-b)}{\Delta a} \quad ;$$

3) экспоненциальный закон:

$$H = \log_2 \frac{Me}{\Delta a} \quad .$$

Таким образом, в радиотехнических системах носителями или переносчиками информации являются электрические сигналы, формируемые источниками этой информации. Даже в тех случаях, когда первичная информация носит неэлектрическую природу (речь, музыка, изображения, тексты, пакеты данных и т. д.), она в конечном итоге преобразуется в электрические сигналы и далее сохраняется или передается по каналам связи. Эти сигналы обычно носят непрерывный характер, то есть определены для любого момента времени или в бесконечном числе точек своего существования. Гораздо удобнее иметь дело с данными, имеющими конечный размер, – например, с массивами чисел конечного размера и ограниченной разрядности.

Контрольные вопросы

1. Каким трем основным направлениям в теории информации отвечают информационные меры? Охарактеризуйте эти направления.
2. В каких случаях применяют геометрическую меру информации? Дайте ее краткую характеристику.
3. Как определить количество информации при помощи комбинаторной меры?
4. Почему для определения количества информации в сообщении используется логарифмическая мера?
5. Какое сообщение содержит одну двоичную единицу информации?
6. Дайте определение количества информации равной 1 биту.
7. Что такое объем информации и как его найти?
8. Что понимается под семантикой информации?
9. Как определить целесообразность информации?
10. Как определяется количество информации в сообщении?
11. Какие события называют равновероятными и неравновероятными?
12. Как найти априорную и апостериорную энтропию?
13. Что такое энтропия?

14. Как определяется энтропия в разных областях знаний?

15. Что такое энтропия источника с независимым выбором сообщений?

16. Как определяется энтропия дискретного источника с независимым выбором сообщений?

17. Какова размерность энтропии источника?

18. Перечислите основные свойства энтропии.

19. Чему равна энтропия источника при укрупнении алфавита (при объединении букв в слова)?

20. Что такое эргодический дискретный источник?

21. Как вычисляется энтропия эргодического источника?

22. Что такое состояние эргодического дискретного источника, как вычислить количество состояний?

23. Что такое избыточность источника?

24. Какие факторы увеличивают избыточность источника?

25. Большая избыточность – это хорошо или плохо для систем передачи информации по каналам связи?

26. Что такое производительность дискретного источника, чему она равна?

27. Можно ли увеличить производительность источника путём укрупнения алфавита?

28. Как можно увеличить производительность источника при заданной длительности элементов сообщения?

29. Что такое совместная энтропия двух источников?

30. Что такое условная энтропия, её физический смысл?

31. Чему равна совместная энтропия двух независимых дискретных источников и двух полностью зависимых источников?

32. Что такое взаимная информация двух источников (или двух сообщений), чему она равна?

33. Что такое ненадежность канала передачи информации, как ее определить через условные вероятности?

34. Как определить информацию о помехах в канале передачи информации, используя условные вероятности перехода $p(y_j/x_i)$?

35. Как определяется дифференциальная энтропия непрерывной случайной величины?

36. Каковы разновидности энтропии непрерывной случайной величины?

37. Чему равна максимальная дифференциальная энтропия, если случайная величина ограничена в объёме?

38. При каком законе распределения максимальная дифференциальная энтропия максимальна?

39. Чему равна максимальная дифференциальная энтропия, если случайная величина не ограничена в объёме?

40. Как влияет математическое ожидание случайной величины на её энтропию?

41. Что такое энтропия непрерывного случайного процесса?

42. Как в энтропии проявляется свойство эргодичности случайного процесса?

43. Как определяется энтропия непрерывного эргодического случайного процесса, её размерность?

44. Что такое эпсилон– энтропия?

45. Как вычисляется эпсилон –энтропия источника непрерывных сигналов?

46. Как определяется эпсилон –производительность источника стационарных сигналов?

47. Как определяется эпсилон– энтропия и производительность гауссовского источника?

48. Что утверждает теорема Шеннона для канала связи, на входе которого подключен источник с известной эпсилон– производительностью?

Контрольные задания

49. Определите количество информации по Хартли числа, равного дате рождения в формате ДДММ.

50. Найдите аддитивную меру информации числа, равного году рождения вашей бабушки.

51. Телеграфист за день работы принимает 64 сообщения. Все сообщения разные. Наугад выбрали одно сообщение. Какое количество информации при этом было получено?

52. Ваш друг живет в шестнадцатиэтажном доме. Сколько информации содержит сообщение о том, что друг живет на 7 этаже?

53. Какое количество информации несет сообщение о том, что встреча назначена на июль?

54. Решите задачу: В 10–этажном доме вашего друга 6 подъездов. Сколько информации содержит сообщение о том, что друг живет в 5 подъезде на 3 этаже?

55. В двух залах студенческой столовой стоит 24 столика. Сколько информации несет сообщение о том, что во втором зале 1 столик свободен?

56. В аудитории 8 столов на 3-х человек. Сколько информации содержится в сообщении о том, что за 3-м столом сидит 1 человек?

57. Загадывают число в диапазоне от 1 до 750. Какое наименьшее количество вопросов надо задать, чтобы наверняка отгадать число. На вопросы можно отвечать только «Да» или «Нет».

58. В алфавите некоторого языка всего две буквы. Каждое слово этого языка состоит из m букв. Известно, что можно составить 2048 различных слов. Сколько букв в каждом слове?

59. Определить энтропию словосочетания, составленного из английских букв Lemon tea. Вероятность появления букв в английском тексте следующая: L = 0,29; E = 0,1105; M = 0,021; O = 0,065; N = 0,059; T = 0,072; A = 0,063

60. Сообщение составляется из алфавита a, b, c, d. Вероятность появления букв алфавита в текстах равна соответственно: $p_a = 0,2$; $p_b = 0,3$; $p_c = 0,4$; $p_d = 0,1$. Найти избыточность сообщений, составленных из данного алфавита.

61. Определить избыточность сообщений, построенных из алфавита со следующим распределением вероятностей появления символов в сообщениях: p_a = 0,03; p_b = 0,26; p_c = =0,09; p_d = 0,05; p_e = 0,16; p_f = 0,1; p_g=0,09; p_h = 0,22.

62. Определить общую условную энтропию сообщений, составленных из алфавита *A, B*, если вероятности появления символов в сообщении равны p_A=0,6; p_B=0,4. Условные вероятности переходов одного символа в другой равны $p(B/A)$=0,15; $p(A/B)$=0,1.

63. В результате статистических испытаний установлено, что при передаче каждых 100 сообщений длиной по 5 символов в сообщении символ *K* встречается 25 раз, а символ *T* – 60 раз. Вместе с

символом *К* символ *Т* встречается 15 раз. Определить условные энтропии *H*(*К*/*Т*) и *H*(*Т*/*К*).

64. Используя энтропию объединения, определить количество информации при передаче сообщений, построенных из алфавита 1, 2, 3, если априорные вероятности появления символов первичного алфавита равны между собой, а в результате действия помех 5% символов передаваемых сообщений могут с равной вероятностью перейти в любой другой символ данного алфавита.

65. Определить энтропию непрерывной случайной величины, распределенной по экспоненциальному закону с плотностью вероятности:

$$p(x) = \begin{cases} ce^{-cx}, & x \le 0 \\ 0, & x < 0 \end{cases}$$

66. Найдите энтропию случайной величины, распределенной по закону с плотностью вероятности

$$p(x) = \begin{cases} 0, & x \le 0 \\ x^2, & 0 < x \le 1 \\ 1, & x > 1 \end{cases}$$

Глава 3. Дискретизация и квантование сигналов

Непрерывные сигналы – носители информации – представляют собой непрерывные функции непрерывного аргумента – времени. Передача таких сигналов может выполняться при помощи непрерывных каналов связи, а обработка − при помощи аналоговых устройств. Непрерывные каналы связи и аналоговые устройства имеют множество недостатков, избежать которые можно, если преобразовать непрерывные сигналы в дискретно–непрерывные или, что еще лучше, в дискретные или, иными словами, в цифровые. Для их передачи и обработки используют более совершенные дискретно–непрерывные или цифровые каналы и устройства обработки соответствующего типа. Дискретизация или квантование позволяют уменьшить количество передаваемых данных, увеличить скорость передачи, использовать методы помехоустойчивого кодирования для ослабления вредного влияния помех.

Предположим, следует передать на расстояние информацию о форме некоторого непрерывного сигнала, причем, как можно точнее. Для этого выполняются следующие действия:

1) преобразование непрерывного сигнала в непрерывно–дискретную или дискретную (цифровую) форму;
2) кодирование эффективным кодом с целью устранения избыточности;
3) кодирование помехоустойчивым кодом с целью защиты от помех;
4) передача закодированного сигнала через канал связи;
5) декодирование после помехоустойчивого кодирования;
6) декодирование после эффективного кодирования;
7) обратное преобразование сигнала из непрерывно–дискретной или цифровой формы в непрерывную.

3.1. Виды дискретизации (квантования)

Наиболее простыми и часто используемыми видами квантования являются:

– квантование по уровню (будем говорить просто *квантование*) (рис. 3.1);

– квантование по времени (будем называть *дискретизацией*) (рис. 3.2);

– их сочетание (рис. 3.3).

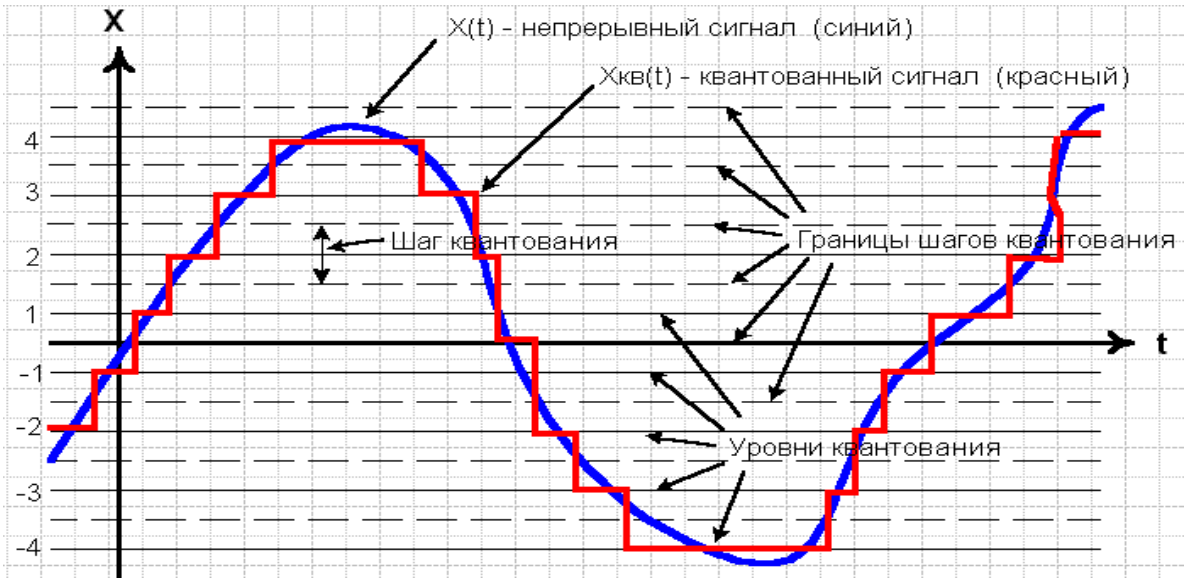


Рис. 3.1. Квантование по уровню (*квантование*)

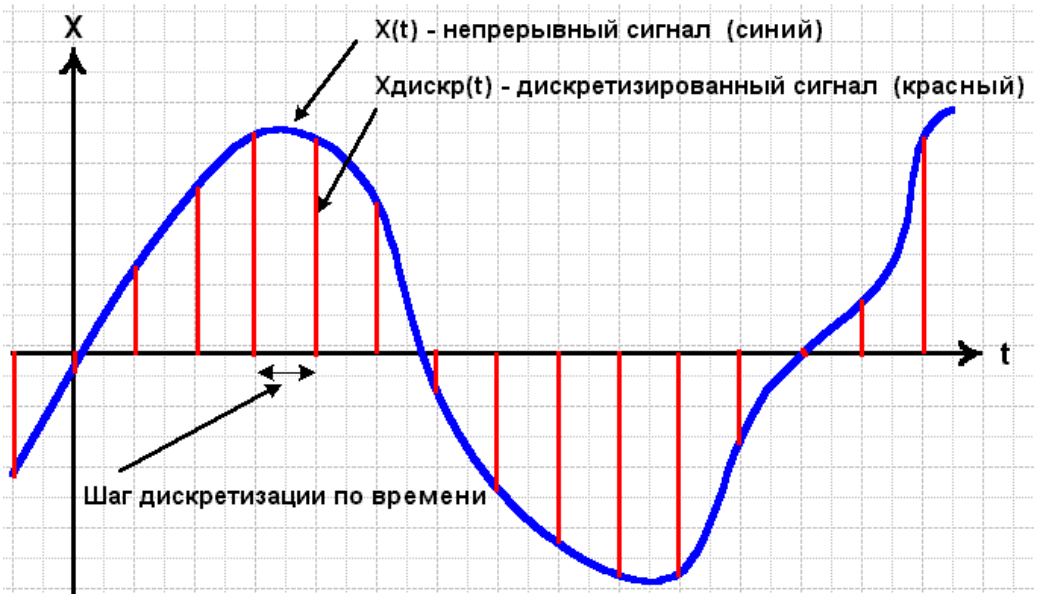


Рис. 3.2. Дискретизация по времени (*дискретизация*)

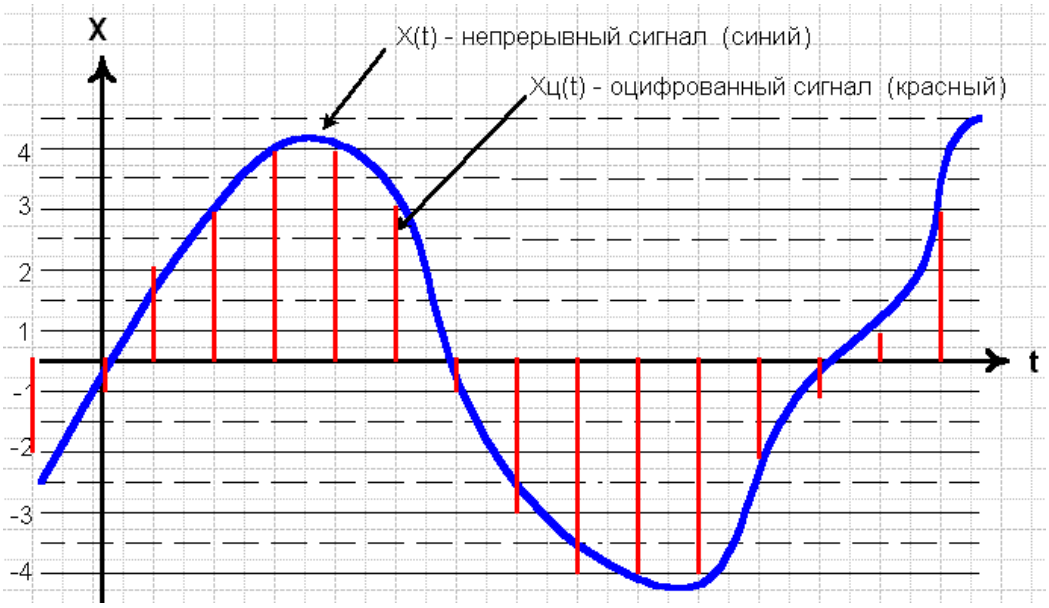


Рис. 3.3. Квантование и по уровню и по времени

Совокупность уровней и границ квантования называют *шкалой квантования.*

Формы квантования только по уровню или только по времени являются непрерывно–дискретными, т. к. в первом случае непрерывной величиной является время t_i перехода с одного уровня на другой, а во втором – значение $x(t_i)$ квантуемой величины в дискретный момент времени t_i.

При квантовании же по уровню и по времени одновременно и время и уровень принимают конечное количество значений, что с технической точки зрения выгоднее всего. Точность представления сигнала, однако, в этом случае наименьшая.

Кроме трех вышеупомянутых видов квантования существует и все шире используется четвертый вид – представление сигнала путем разложения его в ряд по некоторой системе функций:

$$x(t) \approx \sum_{k=0}^{n} c_k \varphi_k(t),$$

где $\varphi_k(t)$ – система функций, по которым выполняется разложение сигнала;

c_k – коэффициенты разложения.

Далее будет показано, что коэффициенты разложения C_k являются непрерывными величинами. Поэтому в данном случае име-

ем дело с непрерывно–дискретной формой представления сигнала. Для перехода к полностью дискретной форме эти коэффициенты нужно проквантовать по уровню.

Исследуем точность представления квантованного сигнала. Для этого прежде всего нужно определить критерии точности.

3.1.1. Критерии точности представления квантованного сигнала

В результате обратного преобразования из непрерывно–дискретной формы в непрерывную получается сигнал $\bar{x}(t)$, отличающийся от исходного $x(t)$ на величину ошибки $\varepsilon(t) = x(t) - \bar{x}(t)$. Сигнал $\bar{x}(t)$ называется *воспроизводящей функцией*.

Способы дискретизации и воспроизведения влияют на ошибку и ее параметры. Обычно, чем шире шаг квантования по уровню или по времени или чем меньше количество n членов разложения сигнала в ряд, тем больше ошибка и одновременно меньше данных нужно передавать через канал связи или меньше объем памяти, требуемый для хранения этого сигнала. Поэтому, зная связь между параметрами дискретизации и восстановления, надо выбирать компромиссное решение, удовлетворяющее как по точности, так и по объемам данных.

Ошибка $\varepsilon(t)$ является функцией времени и потому неудобна для использования в качестве критерия точности тракта дискретизация–восстановление. Поэтому в качестве такого критерия обычно используют какой–либо функционал ошибки.

1. Чаще всего в качестве такого функционала применяют *среднеквадратическую погрешность*, определяемую по формуле:

$$\varepsilon^2 = \frac{1}{T}\int_0^T \varepsilon^2(t)dt \quad ,$$

где T – некоторый временной интервал, на котором находится среднеквадратическая ошибка.

2. Иногда применяют другой критерий – *наибольшее отклонение*:

$$\varepsilon_m = \max\left|x(t) - \overline{x}(t)\right|, \quad 0 \le t \le T.$$

Однако его использование затруднено из–за необходимости априорного знания максимального значения сигнала и его производных.

3. Еще один критерий называется *интегральным*. Он находится по формуле

$$\varepsilon = \frac{1}{T}\int_0^T \varepsilon(t)dt \quad .$$

Интегральный критерий характеризует в основном отклонение среднего значения воспроизведенного сигнала от исходного. Его имеет смысл использовать тогда, когда целью передачи сигнала является передача именно его среднего значения. Критерий характеризуется минимальными объемами требуемых априорных знаний о передаваемом сигнале.

4. *Вероятностный* критерий задается формулой

$$p(\varepsilon(t) < \varepsilon_0) = p_0,$$

где ε_0 − ширина доверительного интервала,

p_0 − доверительная вероятность.

Вероятностный критерий показывает, с какой вероятностью отклонения воспроизведенного сигнала от исходного не выйдет за пределы доверительного интервала. Очевидно, что, чем ширина интервала меньше, а вероятность выше, тем точность воспроизведения сигнала будет больше. Однако отсутствие больших отклонений от исходного сигнала при этом не гарантируется.

5. *Информационный* критерий. При использовании этого критерия рассматривается количество информации, заключенной в воспроизведенном сигнале относительно исходного.

3.1.2. Элементы обобщенной спектральной теории сигналов

Обобщенная спектральная теория сигналов объединяет методы математического описания сигналов и помех. Эти методы позволяют обеспечить требуемую избыточность сигналов с целью уменьшения влияния помех, облегчают анализ и синтез систем передачи информации и т. п.

Сигналы в обобщенной спектральной теории описываются в виде

$$x(t) \approx \sum_{k=0}^{n} c_k \varphi_k(t), \quad t_1 < t < t_2,$$

где $x(t)$ – описываемый сигнал;

c_k – коэффициенты разложения сигнала $x(t)$ в ряд по системе функций $\{\varphi_k(t)\}$ (фигурные скобки обозначают множество);

$\varphi_k(t)$ − заранее известные функции, входящие в систему (множество) $\{\varphi_k(t)\}$; система в данном случае состоит из $n+1$ членов.

О приведенной формуле говорят, что она описывает разложение сигнала $x(t)$ в ряд по системе функций $\{\varphi_k(t)\}$. Функции, входящие в систему $\{\varphi_k(t)\}$, называют *базисными*, а саму систему $\{\varphi_k(t)\}$ − *базисом разложения*.

Если функции $\varphi_k(t)$ отвечают условию:

$$\int_{t_1}^{t_2} \varphi_k(t)\varphi_i(t)dx = \begin{cases} A_k & при \quad k = i \\ 0 & при \quad k \neq i \end{cases},$$

то такие функции называются *ортогональными*. Это очень удобное свойство. Поэтому для разложений сигналов чаще всего используют именно системы ортогональных функций.

Кроме того, выбирают функции, отвечающие еще одному условию (такие функции называются *нормированными*):

$$\int_{t_1}^{t_2} \varphi_k^2(t)dx = 1 \ (A_k = 1) \ .$$

Если функции отвечают обоим условиям одновременно, то их называют о*ртонормированными*.

Разложение сигналов, в котором используются ортогональные функции, называется *ортогональным,* а если используются ортонормированные функции, то ряд называется *ортонормальным*. Ряд Фурье является ортогональным, но не ортонормальным.

Ошибка разложения находится по формуле:

$$\varepsilon(t) = x(t) - \sum_{k=0}^{n} c_k \varphi_k(t) \ .$$

В качестве критерия ошибки будем использовать среднеквадратическую погрешность, которая в данном случае будет записываться в виде:

$$\varepsilon^2 = \frac{1}{t_2 - t_1} \int_{t_1}^{t_2} \left[x(t) - \sum_{k=0}^{n} c_k \varphi_k(t) \right]^2 dt \ .$$

Найдем выражение для коэффициентов c_k, минимизирующих среднеквадратическую погрешность ортогонального разложения. На коэффициенты при этом никаких условий не накладывается. Для минимизации следует решить систему уравнений $\frac{\partial \varepsilon^2}{\partial c_k^2} = 0$ при k = 0, 1,..., n. Если при этом окажется, что $\frac{\partial^2 \varepsilon^2}{\partial c_k^2} > 0$, то найденное решение минимизирует ε^2.

Для решения задачи представим ε^2 в виде:

$$\varepsilon^2 = \frac{1}{t_2 - t_1} \times$$

$$\times \left[\int_{t_1}^{t_2} x^2(t)dt - 2\sum_{k=0}^{n} \int_{t_1}^{t_2} x(t) c_k \varphi_k(t)dt + \sum_{k=0}^{n} \sum_{i=0}^{n} \int_{t_1}^{t_2} c_k c_i \varphi_k(t) \varphi_i(t)dt \right].$$

Третье слагаемое с учетом ортогональности равно:

$$\sum_{k=0}^{n} c_k^2 \int_{t_1}^{t_2} \varphi_k^2(t)\,dt$$

В результате

$$\varepsilon^2 = \frac{1}{t_2 - t_1}\left[\int_{t_1}^{t_2} x^2(t)dt - 2\sum_{k=0}^{n}\int_{t_1}^{t_2} x(t)c_k\varphi_k(t)dt + \sum_{k=0}^{n} c_k^2 \int_{t_1}^{t_2} \varphi_k^2(t)dt\right] .$$

Находим производную:

$$\frac{d\varepsilon^2}{dc_k} = \frac{1}{t_2 - t_1}\left[-2\int_{t_1}^{t_2} x(t)\varphi_k(t)dt + 2c_k\int_{t_1}^{t_2} \varphi_k^2(t)dt\right] = 0 .$$

Если предложить нормированность функций $\varphi_k(t)$, то второе слагаемое равно $2c_k$. В результате легко найти c_k:

$$c_k = \int_{t_1}^{t_2} x(t)\varphi_k(t)dt .$$

Несложно заметить, что вторая производная положительна. Таким образом, для того чтобы среднеквадратическая погрешность ортонормального разложения была минимальна, коэффициенты разложения должны вычисляться по приведенной формуле.

Существует бесконечно много базисных систем. Легко показать, что при одном и том же числе n членов разложения в общем случае среднеквадратическая погрешность при их применении будет различна.

Можно поэтому подыскать такую базисную систему, при которой погрешность будет наименьшей.

Примером такого разложения является разложение в ряд Котельникова.

3.2. Дискретизация по времени

3.2.1. Разложение в ряд Котельникова (Теорема Котельникова)

Описанный здесь результат Котельниковым был получен в 1933 г., но ранее подобные результаты были получены в 1915 г. Виттакером и в 1928 г. Найквистом.

Котельников доказал, что, если некоторый сигнал $x(t)$ имеет ограниченный сверху частотой f_m спектр, то его можно проквантовать по времени с периодом $\Delta t = \frac{1}{2f_m}$ и затем с абсолютной точность восстановить по формуле

$$x(t) = \sum_{k=-\infty}^{\infty} x(k\Delta t) \frac{\sin\left[\omega_m (t - k\Delta t)\right]}{\omega_m (t - k\Delta t)}$$

Такой ряд называется рядом Котельникова, а вышеуказанное утверждение – теоремой Котельникова.

По определению сигнал $x(t)$ и его спектр $S(j\omega)$ находятся в следующих отношениях:

$$S(j\omega) = \int_{-\infty}^{\infty} x(t) \exp(-j\omega t) dt;$$

$$x(t) = \frac{1}{2\pi} \int_{-\omega_m}^{\omega_m} S(j\omega) \exp(j\omega t) dt$$

Эти формулы образуют пару преобразований Фурье (прямое и обратное. Ограниченный интервал интегрирования– следствие ограниченности спектра, поскольку $|S(j\omega)| = 0$ *при* $|\omega| > \omega_m$. Здесь ω – круговая частота, а $j = \sqrt{-1}$.

Образуем новый спектр $S_n(j\omega)$ путем периодического (с периодом $2\omega_m$) продолжения спектра $S(j\omega)$ вдоль оси ω.

Тогда его, как периодическую функцию, можно разложить в ряд Фурье:

$$S_n(j\omega) = \sum_{k=-\infty}^{\infty} c_k \exp\left(\frac{2\pi kj\omega}{2\omega_m}\right)$$

где $c_k = \dfrac{1}{2\omega_m} \int\limits_{-\omega_m}^{\omega_m} S_n(j\omega) \exp\left(\dfrac{2\pi kj\omega}{2\omega_m}\right) d\omega$

В приведенных формулах можно подставлять как $S(j\omega)$, так и $S_n(j\omega)$, так как на участке интегрирования они равны.

Обозначим

$$\psi_k(t) = \frac{\sin[\omega_m(t - k\Delta t)]}{\omega_m(t - k\Delta t)} \quad .$$

Функция $\psi_k(t)$ называется *функцией отсчетов*.

Тогда формула восстановления принимает вид:

$$x(t) = \sum_{k=-\infty}^{\infty} x(k\Delta t)\psi_k(t)$$

На рисунке 3.4 изображена форма двух разных функций отсчетов.

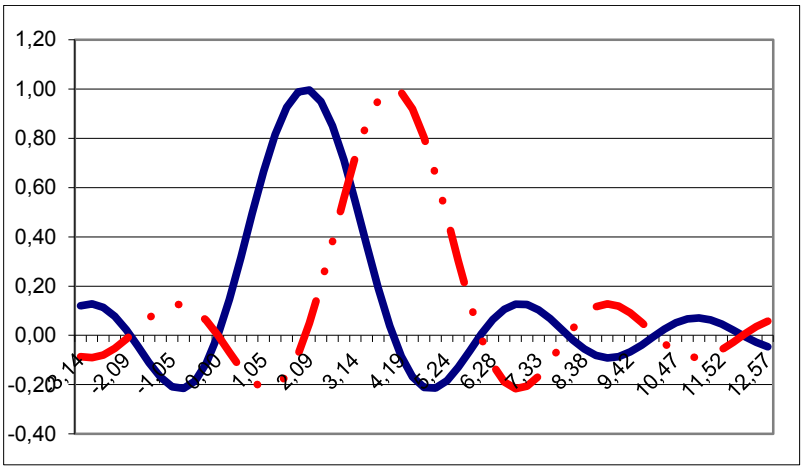


Рис. 3.4. Графики функций отсчетов

Свойства функции отсчетов заключаются в следующем:

1) функция отсчетов $\psi_k(t)$ имеет максимальное значение, равное 1, при $t = k\Delta t$;

2) $\psi_k(t)$ симметрична относительно точки $t = k\Delta t$;

3) функция отсчетов $\psi_i(t) = 0$ при $t = i\Delta t$ *и* $i \neq k$;

4) система функций $\{\psi_i(t)\}$ ортогональна:

$$\int_{-\infty}^{\infty} \psi_k(t)\psi_i(t)dt = \begin{cases} \Delta t \\ 0 \end{cases} при \begin{matrix} i = k \\ i \neq k \end{matrix} .$$

Возможную схему квантования–передачи–восстановления непрерывного сигнала можно представить в виде, изображенном на рисунке 3.5.

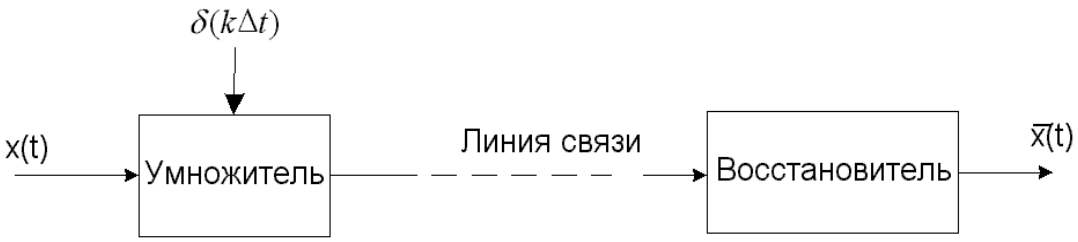


Рис. 3.5. Возможная схема квантования – передачи – восстановления непрерывного сигнала

Технически восстановитель исходной функции может быть реализован при помощи электронного устройства, называемого *идеальным фильтром нижних частот*. Этот фильтр отличается тем, что на импульсное воздействие, описываемое так называемой дельта–функцией $\delta(t)$, реагирует сигналом $h(t)$ в виде функции отсчетов $\psi_0(t)$.

Дельта–функция задается формулой

$$\delta(t) = \begin{cases} \infty \ при \ t = 0 \\ 0 \ при \ t \neq 0 \end{cases}; \ \int_{-\infty}^{\infty} \delta(t)dt = 1$$

Реакция $h(t)$ идеального фильтра нижних частот на входное воздействие в виде дельта–импульса равна

$$h(t) = \lim_{\tau \to \infty} \frac{\sin[\omega_m(t-\tau)]}{\omega_m(t-\tau)}$$

Фильтр нижних частот, как и широкий класс электрических схем, является линейным устройством. Эти устройства отличаются тем, что их реакция на сумму воздействий равна сумме реакций на каждое воздействие в отдельности. Таким образом, если на вход такого устройства, к примеру, подать последовательность импульсов, его реакция будет равна сумме реакций на каждый импульс в отдельности. Но результат все же отличается:

$$\bar{x}(t) = \lim_{\tau \to \infty} \sum_{k=-\infty}^{\infty} x(k\Delta t) \frac{\sin\left[\omega_m(t-k\Delta t-\tau)\right]}{\omega_m(t-k\Delta t-\tau)} = \lim_{\tau \to \infty} x(t-\tau)$$

Восстановленная функция оказывается смещенной по оси x на бесконечно большую величину.

Вторая принципиальная трудность практического использования теоремы Котельникова заключается в том, что все реальные сигналы конечны во времени. Математически доказывается, что в этом случае их спектр не ограничен максимальной частотой. Доказано, что ограниченный спектр могут иметь только сигналы бесконечно большой длительности.

По этой причине f_m приходится определять приближенно, руководствуясь в зависимости от конкретных условий различными рекомендациями.

Можно, например, f_m определить по доле энергии сигнала, сосредоточенной в диапазоне частот от 0 до f_m :

$$\frac{\textit{Энергия в интервале } [0;\omega_m]}{\textit{Полная энергия}} = \textit{Заданная величина} = \frac{\int_0^{\omega_m} S(j\omega)S^*(j\omega)d\omega}{\int_0^{\infty} S(j\omega)S^*(j\omega)d\omega}$$

Некоторые рекомендации позволяют сразу определить период квантования. Это, например, рекомендация, предложенная Железновым в 1960 г., согласно которой период дискретизации предлагается выбирать равным так называемому интервалу корреляции:

$$\Delta e = \tau_{кор} = \frac{1}{R_x(0)} \int_{-\infty}^{\infty} R_x(\tau) d\tau ,$$

где $R_x(\tau)$ – корреляционная функция.

Корреляционная функция связана со спектром следующей зависимостью:

$$R_x(\tau) = \int_{-\infty}^{\infty} S(j\omega) S^*(j\omega) \exp(j\omega\tau) d\omega$$

Этот способ согласуется с результатом, полученным Котельниковым, и, если спектр ограничен, дает полностью совпадающие результаты. Это означает, что отсчеты по Котельникову представляют собой ближайшие некоррелированные (независимые) значения квантуемого сигнала (если бы отсчеты брались с меньшим периодом, они были бы зависимы, т. е. несли бы частично совпадающую информацию, что излишне нагружало бы каналы связи).

В заключение заметим, что, хотя теорема Котельникова и базируется на идеализированной модели сигналов, она имеет большую практическую и теоретическую ценность.

3.2.2. Выбор периода дискретизации (квантования по времени) по критерию наибольшего отклонения

В результате квантования по времени функции $x(t)$ получается ряд значений $x(t_1), x(t_2), \ldots$ квантуемой величины $x(t)$ в дискретные моменты времени t_1, t_2, … Период квантования $\Delta t_i = t_i - t_{i-1}$ может быть постоянным, когда $\Delta t_1 = \Delta t_2 = \Delta t_3 = \ \ldots$, и переменным, когда $\Delta t_1 \neq \Delta t_2 \neq \Delta t_3 \neq \ldots$

Если *функция* $x(t)$ имеет ограниченный максимальной частотой f_m спектр, время ее воспроизведения и сложность воспроизводящего устройства не имеют существенного значения, хорошего воспроизведения можно достичь при помощи ряда Котельникова. Необходимая частота дискретизации при этом, как ранее было по-

казано, равна удвоенной максимальной частоте, а период дискретизации Δt, следовательно, находится по формуле $\Delta t = \frac{1}{2 f_m}$.

Чаще, однако, одно или несколько из вышеуказанных условий не выполняется и тогда используют другие принципы выбора периода дискретизации и способы воспроизведения. Период дискретизации при этом, как правило, уменьшается в 5 – 10 раз.

При воспроизведении исходной функции широко используют хорошо проработанные в математике методы интерполяции, приближения и экстраполяции. При интерполяции по уже полученным отсчетам восстанавливаются прошлые значения функции, причем чем больше этих отсчетов используется, тем точнее восстановление.

Методы экстраполяции позволяют предсказать будущие значения функции (прогнозировать), что для некоторых задач чрезвычайно важно (вспомним, например, прогноз погоды). И здесь, чем больше отсчетов используется для экстраполяции (предсказания), тем точнее предсказания выше.

Воспроизводящая функция в большинстве случаев рассчитывается по формуле

$$\bar{x}(t) = \sum_{k=0}^{n} x(t_k)\varphi_k(t),$$

где $\varphi_k(t)$ − некоторые функции.

Эти функции обычно стремятся выбрать так, чтобы

$$\varphi_k(t) = \begin{cases} 1 \ \ при \ \ t = t_k \\ 0 \ \ при \ \ t = t_{i \neq k} \end{cases}$$

В этом случае $\bar{x}(t_k) = x(t_k)$, т. е. значения воспроизводящей и исходной функций совпадают в моменты взятия отсчетов или, как принято говорить, в узлах интерполяции.

Функции, обладающие этим качеством, нашел выдающийся французский математик и механик Жозеф Луи Лагранж (1736–1813).

Функции Лагранжа L зависят от одного аргумента t и двух параметров n и k, здесь n – максимальный номер отсчета, а k – номер функции.

$$\varphi_k(t) = L_{nk}(t) = \frac{(t-t_0)(t-t_1)...(t-t_{k-1})(t-t_{k+1})...(t-t_n)}{(t_k-t_0)(t_k-t_1)...(t_k-t_{k-1})(t_k-t_{k+1})...(t_k-t_n)} = \prod_{\substack{i=0 \\ i\neq k}}^{n} \frac{t-t_i}{t_k-t_i}$$

Полином Лагранжа можно использовать для расчета воспроизводящей функции как при равномерной, так и при неравномерной дискретизации. Если же ограничиться только равномерной дискретизацией, полином Лагранжа можно преобразовать к виду

$$\bar{x}(t) = (-1)^n \frac{\Theta(\Theta-1)...(\Theta-n)}{n!} \sum_{k=0}^{n} (-1)^k \frac{C_n^k x(t_k)}{\Theta-k},$$

где $\Theta = \dfrac{t-t_0}{\Delta t}$, $C_n^k = \dfrac{n!}{k!(n-k)!}$.

Найдем погрешность интерполяции. Представим ее виде

$$\varepsilon(t) = x(t) - \bar{x}(t) = (t-t_0)(t-t_1)...(t-t_n)K(t)$$

где $K(t)$ – вспомогательная функция, которую надо найти.

Для произвольного t^* имеем:

$$x(t^*) - \bar{x}(t^*) - (t^*-t_0)(t^*-t_1)...(t^*-t_n)K(t^*) = 0.$$

Введем в рассмотрение еще одну функцию

$$\Phi(t) = x(t) - \bar{x}(t) - (t-t_0)(t-t_1)...(t-t_n)K(t^*).$$

Продиффенцируем ее $n+1$ раз по t.

Так как $\bar{x}(t)$ – полином n–й степени, то $\dfrac{d^{n+1}\bar{x}(t)}{dt^{n+1}} = 0$.

$K(t^*)$ –константа, значит

$$\frac{d^{n+1}\left[(t-t_0)(t-t_1)...(t-t_n)K(t^*)\right]}{dt^{n+1}}=(n+1)!K(t^*).$$

Следовательно

$$\frac{d^{n+1}\Phi(t)}{dt^{n+1}}=\frac{d^{n+1}x(t)}{dt^{n+1}}-(n+1)!K(t^*)$$

Функция $\Phi(t)$ пересекает ось t как минимум $n+2$ раз (при значениях аргумента t в точках t_1, t_2, …, t_n и t^*; рис. 3.5). Это значит, что первая производная имеет хотя бы $n+1$ нулей (по одному на нулю на каждом интервале), вторая производная – n нулей и т.д. $n+1$–я производная должна иметь хотя бы одно нулевое значение при некотором $t=\hat{t}$.

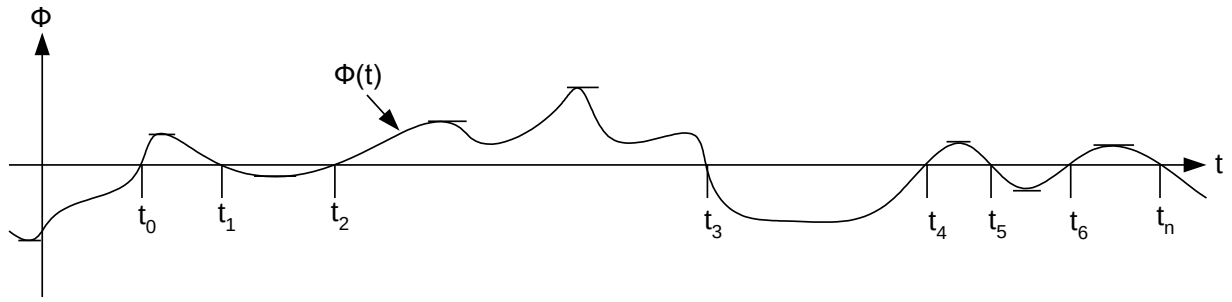


Рис. 3.5. Точки пересечения оси Φ функцией $\Phi(t)$

Отсюда в этой точке

$$\frac{d^{n+1}x(\hat{t})}{dt^{n+1}}=(n+1)!K(t^*)\ \ и\ \ K(t^*)=\frac{\frac{d^{n+1}x(\hat{t})}{dt^{n+1}}}{(n+1)!}$$

Тогда формула ошибки воспроизведения

$$\varepsilon(t)=x(t^*)-\bar{x}(t^*)=(t^*-t_0)(t^*-t_1)\ldots(t^*-t_n)\frac{\frac{d^{n+1}x(t)}{dt^{n+1}}}{(n+1)!}$$

Так как t^* произвольно, его можно заменить на t:

$$\varepsilon(t)=\frac{\prod_{i=0}^{n}(t-t_i)}{(n+1)!}\frac{d^{n+1}x(\hat{t})}{dt^{n+1}}$$

Аргумент $\hat{t}$ выбирался из условия $\frac{d^{n+1}\Phi(\hat{t})}{dt^{n+1}}=0$. На практике $\hat{t}$ найти сложно. Если вместо $\frac{d^{n+1}x(\hat{t})}{dt^{n+1}}$ подставить $M_{n+1}=\max\left|\frac{d^{n+1}x(\hat{t})}{dt^{n+1}}\right|$ и вместо $\prod_{i=0}^{n}(t-t_i)$ подставить $\max\left|\prod_{i=0}^{n}(t-t_i)\right|$, то получаем оценку абсолютной величины ошибки сверху

$$\left|\varepsilon(t)\right|<\frac{\max\left|\prod_{i=0}^{n}(t-t_i)\right|}{(n+1)!}M_{n+1}.$$

Таким образом, для определения наибольшего отклонения воспроизводящей функции, представленной полиномом Лагранжа, от исходной необходимо знать максимальное по абсолютной величине значение производной исходной функции.

Приступим теперь к решению обратной задачи: по заданному значению максимальной погрешности ε_m, порядку n полинома Лагранжа и максимальному значению производной M_n найти максимально возможный шаг равномерной дискретизации.

Предположим, что для дискретизации – передачи – восстановления сигнала используется схема, изображенная на рисунке 3.6.

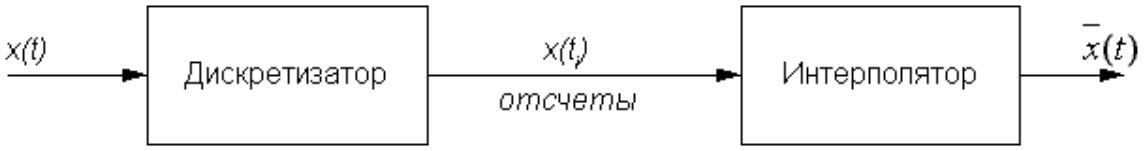


Рис. 3.6. Схема дискретизации–передачи–восстановления сигнала $X(t)$

Верхняя граница максимальной ошибки ε_m, как показано выше, задается формулой $\varepsilon_m = |\varepsilon(t)| < \dfrac{\max\left|\prod_{i=0}^{n}(t-t_i)\right|}{(n+1)!} M_{n+1}$.

Графики $\left|\prod_{i=0}^{n}(t-t_i)\right|$ на интервале дискретизации при разном порядке n полинома Лагранжа изображены на рисунке 3.7.

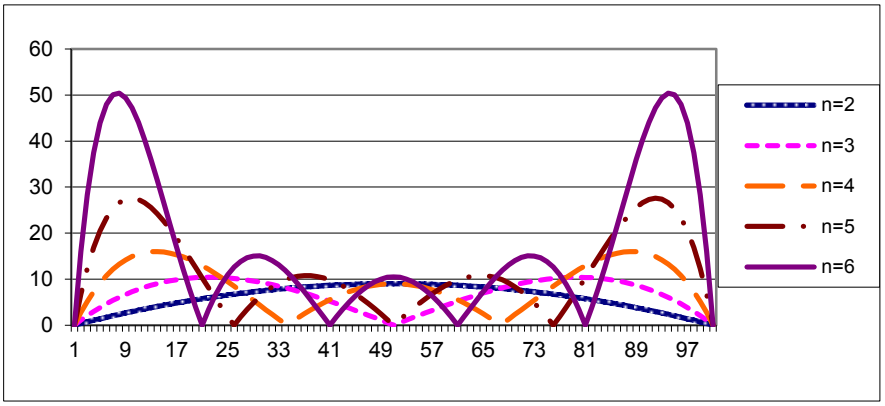


Рис. 3.7. График сомножителя $\left|\prod_{i=0}^{n}(t-t_i)\right|$ при разных n

Ясно, что наименьших значений $\max\left|\prod_{i=0}^{n}(t-t_i)\right|$ достигает на серединных шагах Δt. Следовательно, и восстанавливать сигнал лучше всего именно на этих серединных отрезках. Это позволит уменьшить ε_m.

Для этого предлагается использовать порядок восстановления, иллюстрируемый графиками, приведенными на рисунке 3.8.

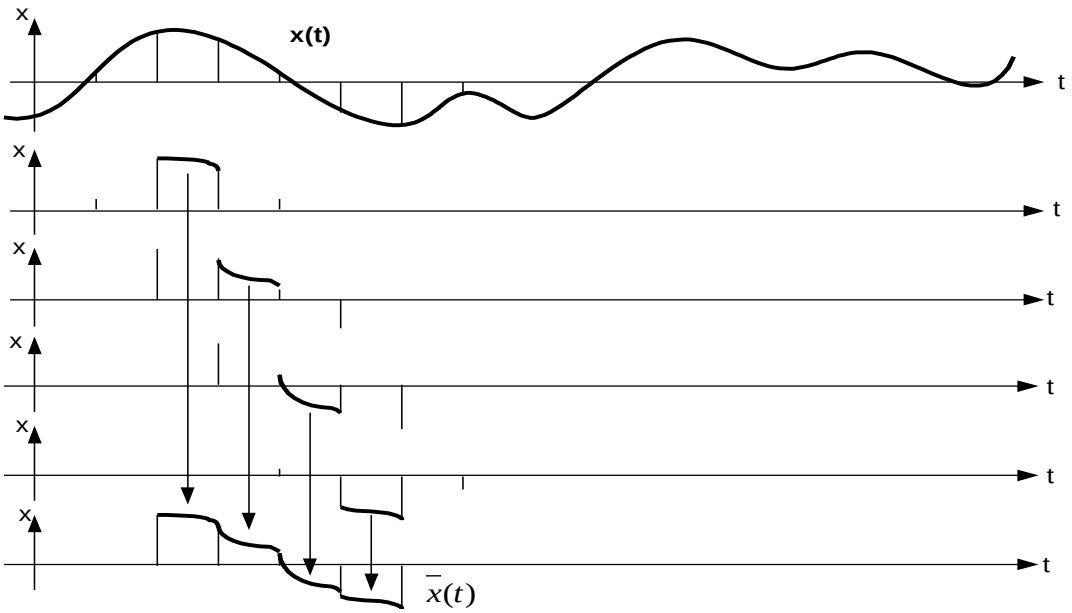


Рис. 3.8. Порядок восстановления исходной функции при использовании для этого полиномов Лагранжа 4 степени

$$\bar{x}(t) = \sum_{k=0}^{n} x(t_k)\varphi_k(t);$$

$$\varphi_k(t) = L_{nk}(t) = \frac{(t-t_0)(t-t_1)\ldots(t-t_{k-1})(t-t_{k+1})\ldots(t-t_n)}{(t_k-t_0)(t_k-t_1)\ldots(t_k-t_{k-1})(t_k-t_{k+1})\ldots(t_k-t_n)} =$$

$$= \prod_{\substack{i=0 \\ i\neq k}}^{n} \frac{t-t_i}{t_k-t_i}.$$

Рассмотрим несколько вариантов, отличающихся порядком n используемых полиномов Лагранжа.

Проведем расчет периода дискретизации при использовании для получения воспроизводящей функции полинома Лагранжа нулевого порядка.

Интерполяция полиномами нулевого порядка $n=0$.

Таких полиномов всего 1:

$L_{00}(t) = \frac{1}{1} = 1$, отсюда $\bar{x}(t) = x(t_0)$ при $t_0 - \Delta t \le t \le t_0$.

Воспроизводящая функция при использовании полиномов Лагранжа нулевого порядка имеет вид ступенчатой кривой (рис. 3.9).

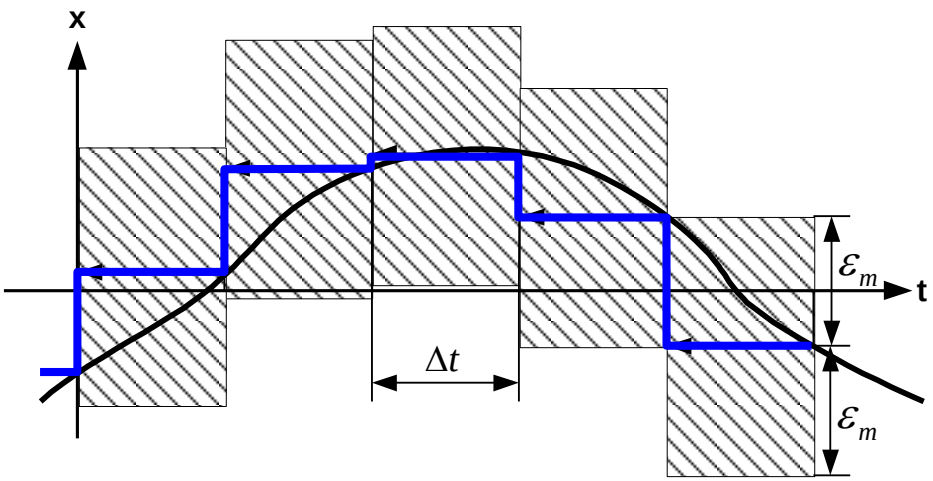


Рис. 3.9. Воспроизводящая функция при использовании для восстановления полиномов Лагранжа нулевого порядка

Подставив $n=0$ в формулу максимальной погрешности, по заданной погрешности найдем максимальный шаг квантования:

$\max\left|\prod_{i=0}^{0}(t-t_i)\right|=\Delta t$, отсюда $\varepsilon_m=\Delta t\cdot M_1$ и $\Delta t=\varepsilon_m/M_1$.

Выполним расчет периода дискретизации при использовании для получения воспроизводящей функции полинома Лагранжа нулевого порядка.

Интерполяция полиномами первого порядка $n=1$.

Таких полиномов оказывается всего 2:

$$\begin{cases} L_{10}=\dfrac{t-t_1}{t_0-t_1} \\ L_{11}=\dfrac{t-t_0}{t_1-t_0} \end{cases};$$

$$\bar{x}(t)=x(t_0)\frac{t-t_1}{\underbrace{t_0-t_1}_{-\Delta t}}+x(t_1)\frac{t-t_0}{\underbrace{t_1-t_0}_{\Delta t}}=\frac{-x(t_0)t+x(t_0)t_1+x(t_1)t-x(t_1)t_0}{\Delta t}=$$

$$=\frac{x(t_1)-x(t_0)}{\Delta t}t+\frac{t_1x(t_0)-t_0x(t_1)}{\Delta t}.$$

Видно, что последнее уравнение – уравнение прямой.

В графической форме такая интерполяция выглядит в виде ломаной линии (рис. 3.10).

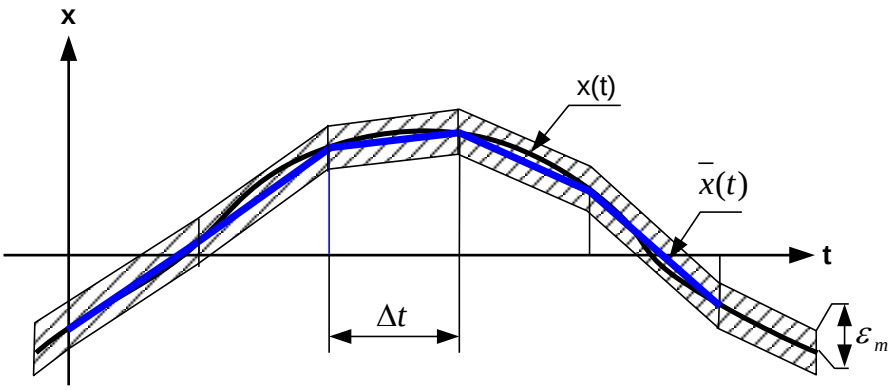


Рис. 3.10. Воспроизводящая функция при использовании для восстановления полиномов Лагранжа первого порядка

Найдем $\max\left|\prod_{i=0}^{1}(t-t_i)\right|$. Для этого приравняем производную этой величины по t к нулю:

$$\frac{d\left[(t-t_0)(t-t_1)\right]}{dt}=(t-t_0)+(t-t_1)=0,$$

отсюда $t=\dfrac{\Delta t}{2}$.

Вторая производная отрицательна. Это означает, что найден именно максимум.

Теперь подставим найденное t в формулу максимальной погрешности:

$$\max\left|\prod_{i=0}^{1}(t-t_i)\right|=\Delta^2 t/4;$$

$$\varepsilon_m=\frac{\Delta^2 t}{4\cdot 2}M_2;$$

$$\Delta t = \sqrt{8\varepsilon_m / M_2}$$

и найдем шаг дискретизации.

Выполним расчет периода дискретизации при использовании для получения воспроизводящей функции полинома Лагранжа второго порядка.

Интерполяция полиномами второго порядка $n=3$:

$$\begin{cases} L_{20} = \dfrac{(t-t_1)(t-t_2)}{(t_0-t_1)(t_0-t_2)} \\ L_{21} = \dfrac{(t-t_0)(t-t_2)}{(t_1-t_0)(t_1-t_2)} \\ L_{22} = \dfrac{(t-t_0)(t-t_1)}{(t_2-t_0)(t_2-t_1)} \end{cases}$$

$$\bar{x}(t) = x(t_0)\frac{(t-t_1)(t-t_2)}{\underbrace{(t_0-t_1)(t_0-t_2)}_{2\Delta t^2}} + x(t_1)\frac{(t-t_1)(t-t_2)}{\underbrace{(t_0-t_1)(t_0-t_2)}_{-\Delta t^2}} +$$

$$+x(t_2)\frac{(t-t_1)(t-t_2)}{\underbrace{(t_0-t_1)(t_0-t_2)}_{2\Delta t^2}} = at^2 + bt + c,$$

где $\bar{x}(t)$ − парабола.

Интерполяция полиномами n–го порядка рассматривается аналогично предыдущим случаям. При этом наблюдается значительное усложнение формул. Обобщение приводит к формуле следующего вида:

$$\Delta t = \sqrt[n+1]{v(n)\varepsilon_m / M_{n+1}},$$

Где $v(n)$ − некоторая функция n.

Кроме того, для M_{n+1} выведена оценка сверху:

$$M_{n+1} \le (2\pi f_m)^{n+1} \cdot \left|x(t)\right|_{\max},$$

где $\left|x(t)\right|_{\max}$ – максимальное значение x(t) на отрезке интерполяции.

При увеличении порядка полинома максимально допустимое значение Δt быстро растет. Максимально возможный период дискретизации возрастает при получении воспроизводящей функции полинома *n*–го порядка по сравнению со случаем использования для этих же целей полинома нулевого порядка при максимальной ошибке, равной 0,1% от X_{max}.

3.2.3. Выбор интервала дискретизации по критерию среднеквадратического отклонения

Рассмотрим случай дискретизации случайного стационарного эргодического процесса $x(t)$ с известной корреляционной функцией $R(\tau)$. Восстанавливать будем при помощи полиномов Лагранжа. Наиболее часто используются полиномы нулевого и первого порядка.

Начнем с полинома нулевого порядка.

Ошибка, как известно, вычисляется по формуле

$$\varepsilon(t) = x(t) - \overline{x}(t).$$

Дисперсия D_ε ошибки в случае стационарных эргодических процессов может быть вычислена по формуле:

$$D_\varepsilon = \lim_{T\to\infty}\frac{1}{T}\int_0^T\left[\varepsilon(t) - m_\varepsilon\right]^2 dt = S\left\{\left[\varepsilon(t) - m_\varepsilon\right]^2\right\},$$

где $S\{\varepsilon\}$ – оператор усреднения по времени;

m_ε – среднее значение (математическое ожидание) ошибки.

Найдем m_ε:

$$m_\varepsilon = M\left\{\varepsilon(t)\right\} = M\left\{x(t)\right\} - M\left\{x(t_i)\right\} = m_x - m_x = 0,$$

Где $M\{\varepsilon\}$ – оператор математического ожидания.

Значит,

$$D_\varepsilon = S\left\{\varepsilon^2(t)\right\} = S\left\{x^2(t)\right\} - 2S\left\{x(t)x(t_i)\right\} + S\left\{x^2(t_i)\right\}.$$

Найдем каждое из этих средних.

1) $S\left\{x^2(t)\right\} = S\left\{\left[x(t) - m_x + m_x\right]^2\right\} = S\left\{\left[x(t) - m_x\right]^2\right\} -$

$-2S\left\{\left[x(t) - m_x\right]\cdot m_x\right\} + S\left\{m_x^2\right\} = R_x(0) + m_x^2;$

2) $S\left\{x^2(t_i)\right\} = R_x(0) + m_x^2$ по аналогичной причине;

3) $S\left\{x(t)x(t_i)\right\} = \lim\limits_{n\longrightarrow 1}\dfrac{1}{n\Delta t}\sum\limits_{i=1}^{n}\int\limits_0^{\Delta t} x(t_i - \tau)x(t_i)d\tau =$

$= \dfrac{1}{\Delta t}\int\limits_0^{\Delta t}\left[\lim\limits_{n\longrightarrow 1}\sum\limits_{i=1}^{n} x(t_i - \tau)x(t)\right]d\tau = \dfrac{1}{\Delta t}\int\limits_0^{\Delta t} R_x(\tau)d\tau + m_x^2.$

В итоге:

$$D_s = 2\left[R_x(0) - \frac{1}{\Delta t}\int\limits_0^{\Delta t} R_x(\tau)d\tau\right] < 2\left[R_x(0) - R(\Delta t)\right],$$

отсюда

$$R_x(0) - R_x(\Delta t) \approx \frac{D_x}{2}.$$

Используя эту формулу, шаг дискретизации Δt можно найти графически так, как показано на рисунке 3.11.

В заключение отметим, что найдены также выражения для дисперсии ошибки и при восстановлении с помощью полиномов более высоких, чем нулевой, порядков.

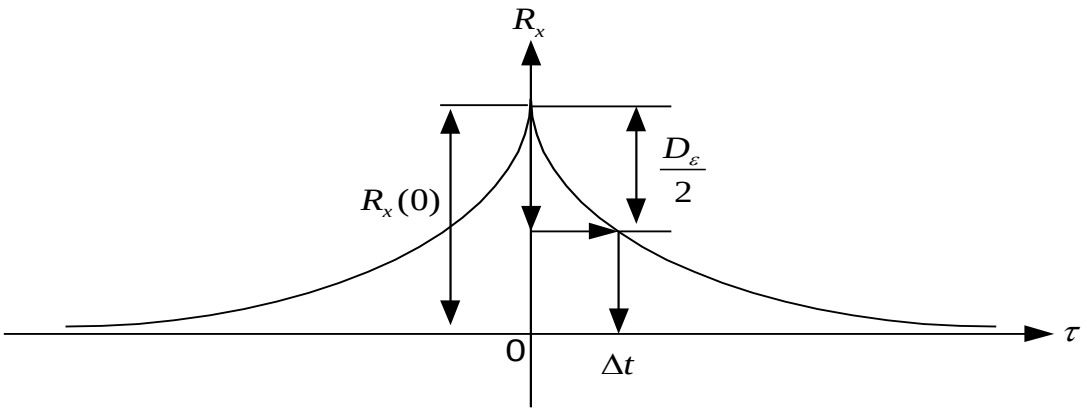


Рис. 3.11. Порядок графического определения шага дискретизации, если использовать критерий среднеквадратического отклонения

3.3. Квантование по уровню

3.3.1. Оптимальное квантование по уровню

Рисунком 3.12 иллюстрируется принцип квантования по уровню.

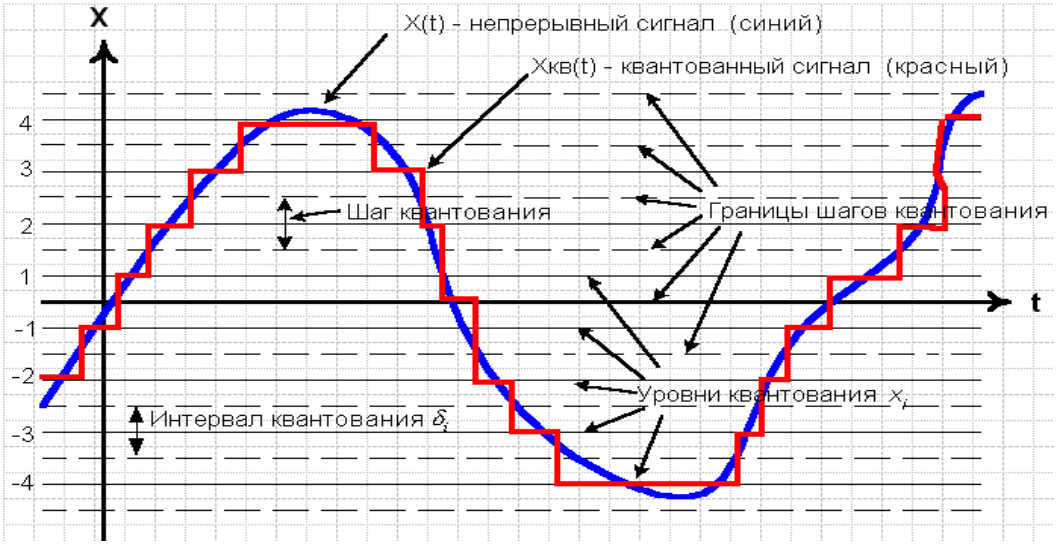


Рис. 3.12. Квантование по уровню

Это квантование сводится к замене значения исходного сигнала уровнем того шага, в пределы которого оно (это значение) попадает. Как уже говорилось, *шкала квантования по уровню* характеризуется совокупностью границ интервалов квантования δ_i и уровней квантования X_i.

Количество передаваемых или хранимых данных о сигнале прямо пропорционально количеству уровней квантования. Следовательно, чем этих уровней меньше, тем меньше число передаваемых или хранимых данных и тем экономичнее система передачи или хранения квантованного сигнала.

Совсем необязательно, чтобы шаги квантования по уровню были одинаковыми. Тогда, варьируя параметрами шкалы квантования, можно, к примеру, добиться минимально возможной при заданном числе шагов квантования дисперсии ошибки. Или, наоборот, задавшись дисперсией ошибки, найти параметры шкалы квантования (положение уровней и границ шагов квантования), при которых число уровней квантования минимально.

Вывод нужных соотношений будем вести, предполагая, что число уровней квантования достаточно велико, а скорость изменения квантуемого сигнала настолько мала, чтобы закон распределения ошибки внутри каждого из интервалов квантования можно было считать равномерным.

Это предположение иллюстрируется рисунком 3.13.

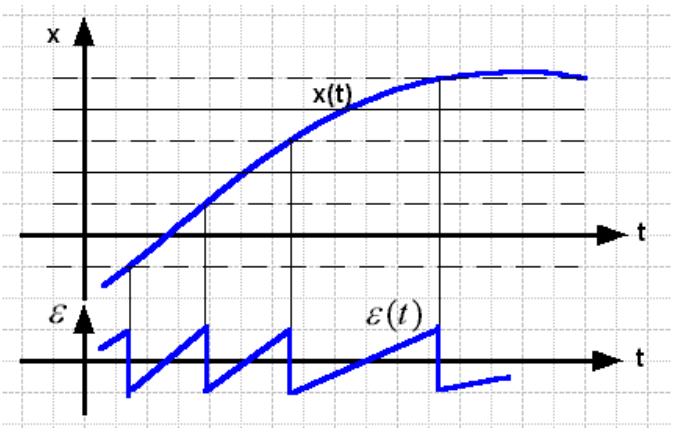


Рис. 3.13. График ошибки квантования $\varepsilon(t)$

Из этого следует, что функция ошибки $\varepsilon(t)$ имеет пилообразную форму и что чем меньше размер шага квантования, тем эта форма ближе к идеально пилообразной, это, в свою очередь, означает, что вероятности значений ошибки в пределах интервала квантования равны, т. е. закон распределения этих значений является равномерным.

Зададимся целью найти такую шкалу квантования по уровню, при которой *среднее значение квадрата ошибки минимально*.

Введем обозначения (рис. 3.14): $x_{нi}$ – нижняя граница шага квантования; $x_{вi}$ – верхняя граница шага квантования; x_i – уровень квантования; $\delta_i = x_{вi} - x_{нi}$ – размер i–го шага квантования.

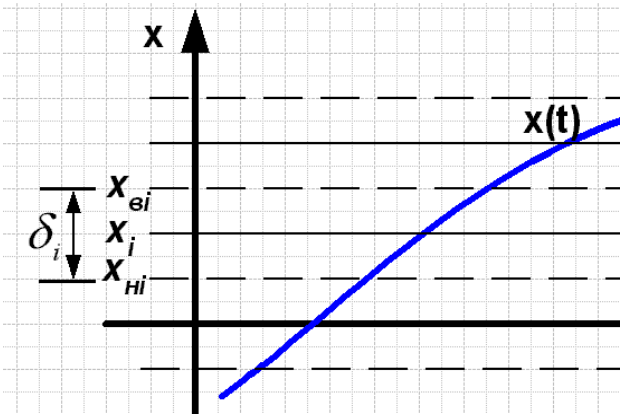


Рис. 3.14. Обозначения

Тогда ошибка $\varepsilon_i(t)$ на i–м интервале распределена по равномерному закону (рис. 3.15).

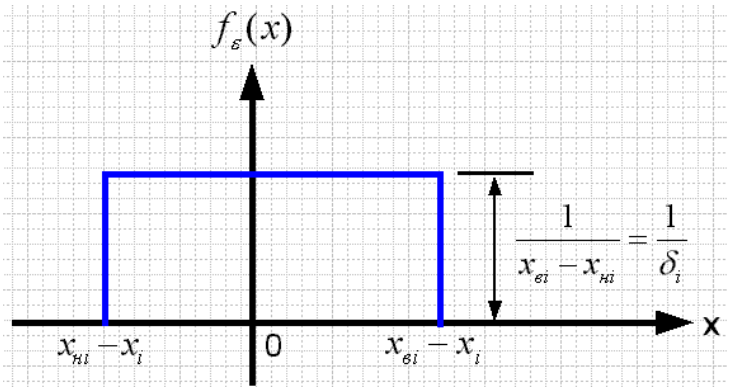


Рис. 3.15. Закон распределения ошибки

Выведем для этого случая формулы математического ожидания m_{ε_i} и дисперсии D_{ε_i} ошибки квантования на i–м интервале.

$$m_{\varepsilon_i} = \int_{-\infty}^{\infty} x \cdot f_{\varepsilon}(x)dx = \int_{x_{нi}-x_i}^{x_{вi}-x_i} x\frac{1}{\delta_i}dx = \frac{x^2}{2\delta_i}\Bigg|_{x_{вi}-x_i}^{x_{вi}-x_i} =$$

$$= \frac{x_{вi}^2 - 2x_{вi}x_i + x_i^2 - \left(x_{нi}^2 - 2x_{нi}x_i + x_i^2\right)}{2\delta_i} = \frac{x_{нi}+x_{вi}}{2} - x_i;$$

$$D_{\varepsilon_i} = \int_{-\infty}^{\infty} \left(x - m_{\varepsilon_i}\right)^2 f_{\varepsilon}(x)dx = \frac{\delta_i^2}{12}.$$

Среднее значение квадрата ошибки на i–м уровне квантования по определению находится по формуле

$$M\left\{\varepsilon_t^2(t)\right\} = M\left\{\left[\varepsilon_t(t) - m_{s_t} + m_{s_t}\right]^2\right\} = M\left\{\left(\varepsilon_i(t) - m_{s_t}\right)^2\right\} -$$

$$-2M\left\{\left(\varepsilon_i(t) - m_{s_t}\right)\cdot m_{s_t}\right\} + m_{s_t}^2 = D_{s_t} + m_{s_t}^2,$$

т. к.

$$M\left\{\left(\varepsilon_i(t) - m_{\varepsilon_i}\right)\right\} = 0.$$

Дисперсия D_{ε_i} от положения уровня квантования x_i не зависит, а

$$\min\left(m_{\varepsilon_i}^2\right) = 0 \;\; при \;\; x_i = \frac{x_{нi}+x_{вi}}{2}$$

Введем обозначения (рис. 3.16): $x_{нi}$ – нижняя граница шага квантования; $x_{вi}$ – верхняя граница шага квантования; $\delta_i = x_{вi} - x_{нi}$ – размер i–го шага квантования. Тогда получим, что для минимизации среднего значения квадрата ошибки уровни квантования должны помещаться в центре соответствующих шагов квантования.

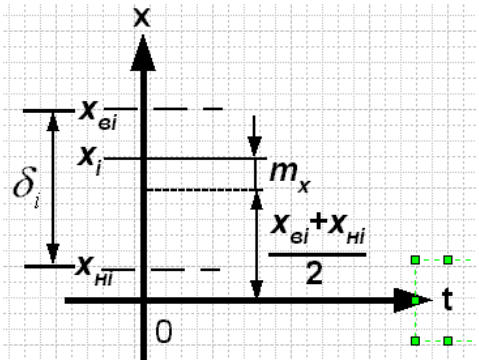


Рис. 3.16. Обозначения

При этом не надо забывать о предположении достаточно большого числа уровней квантования, из–за чего стало возможным предположение о равномерном законе распределения ошибки.

Доказано, что значения функции ошибки $\varepsilon(t)$ при различных i можно считать независимыми. Поэтому среднее значение квадрата ошибки по всем уровням ε_{cp}^2 можно выполнить по известной формуле:

$$\varepsilon_{cp}^2 = \sum_{i=1}^{n} p_i \cdot M\left\{\varepsilon_i^2(t)\right\},$$

где p_i – вероятность попадания квантуемого сигнала в i–й интервал квантования;

$M\left\{\varepsilon_i^2(t)\right\}$ – среднее значение квадрата ошибки на i–м интервале квантования.

Так как p_i от положения уровня квантования x_i внутри интервала квантования не зависит, то минимизация среднего значения квадрата ошибки на каждом из уровней приведет к минимизации среднего значения квадрата ошибки вообще.

При $m_{\varepsilon_i} = 0$ среднее значение квадрата ошибки $M\left\{\varepsilon_i^2(t)\right\} = D_{\varepsilon_i}$, поэтому

$$\varepsilon_{cp}^2 = \sum_{i=1}^{n} p_i \cdot M\left\{\varepsilon_i^2(t)\right\} = \sum_{i=1}^{n} p_i \cdot D_{\varepsilon_i} = \sum_{i=1}^{n} p_i \cdot \frac{\delta_i^2}{12} = D_{\varepsilon}.$$

При малых шагах δ_i можно считать $p_i = f_\varepsilon(i)\cdot\delta_i$ (рис. 3.17).

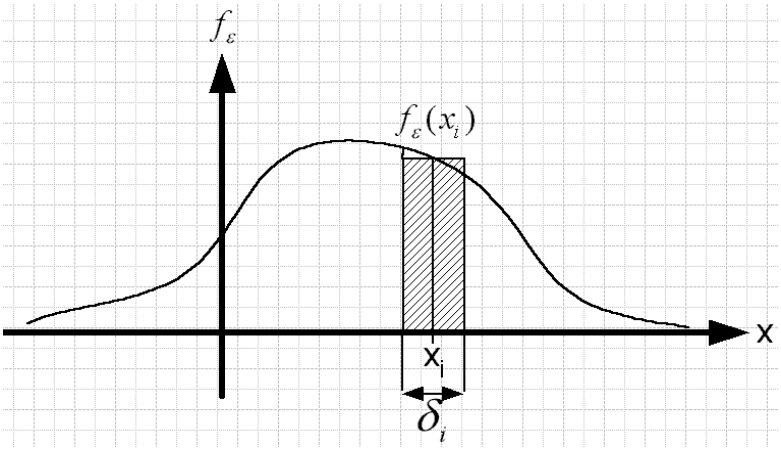


Рис. 3.17.К расчету вероятности попадания значения сигнала в пределы шага

Тогда

$$D_\varepsilon = \frac{1}{12}\sum_{i=1}^{n}\delta_i^3 f(x_i)$$

3.3.2. Дисперсия ошибки в случае использования равномерной шкалы квантования по уровню

Найдем дисперсию ошибки в случае использования равномерной шкалы квантования, т. е. равенства всех шагов квантования. В этом случае

$$\delta_i = \delta \qquad \text{и } D_\varepsilon = \sum_{i=1}^{n} p_i \cdot \frac{\delta^2}{12} = \frac{\delta^2}{12}\sum_{i=1}^{n} p_i = \frac{\delta^2}{12}\ , \ \textit{т. к.}\ \sum_{i=1}^{n} p_i = 1$$

Следовательно, при постоянном шаге квантования дисперсия ошибки квантования находится по формуле

$$D_\varepsilon = \frac{\delta^2}{12}.$$

Зададимся теперь числом шагов квантования n, границами интервала (x_{min}, x_{max}) значений квантуемой величины, покрываемой

шагами квантования (рис. 3.18), и найдем такое распределение шагов на интервале, которое минимизирует дисперсию ошибки или, что то же самое, среднее значение ее квадрата.

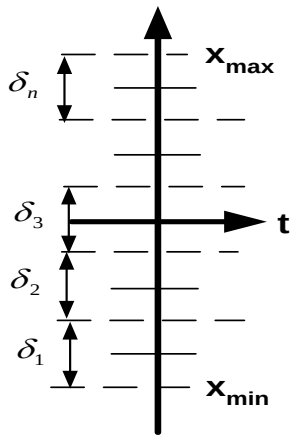


Рис. 3.18. Обозначения

Введем вспомогательную переменную

$$z_i = \delta_i \cdot \sqrt[3]{f(x_i)}.$$

Тогда, учитывая рисунок 3.19, дисперсия ошибки квантования:

$$D_\varepsilon = \sum_{i=1}^{n} \frac{z_i^3}{12}.$$

Заметим, что

$$\sum_{i=1}^{n} z_i = \sum_{i=1}^{n} \delta_i \cdot \sqrt[3]{f(x_i)} \approx \int_{-\infty}^{\infty} \sqrt[3]{f(x)} dx = S,$$

где S − функционал от плотности распределения вероятностей – некоторое число.

Теперь найдем такой набор z_i, при котором дисперсия ошибки минимальна. Для этого воспользуемся методом множителей Лагранжа.

Вспомогательная функция Φ в этом случае принимает вид:

$$\Phi = \sum_{i=1}^{n} \frac{z_i^3}{12} + \lambda \left(\sum_{i=1}^{n} z_i - S \right).$$

В результате получаем систему уравнений

$$\begin{cases} \dfrac{\partial \Phi}{\partial z_i} = \dfrac{z_i^2}{4} + \lambda = 0 \\ \sum\limits_{i=1}^{n} z_i = S \end{cases}$$

Из первого уравнения системы следует, что $z_i = -2\sqrt{\lambda}$, т. е. все z_i равны. Учитывая это, из второго уравнения системы получаем $z_i = S / n$.

После подстановки z_i получаем следующую формулу:

$$D_{\varepsilon\ \min} = \frac{S^3}{12n^2} \ ,$$

где $S = \int\limits_{-\infty}^{\infty} \sqrt[3]{f(x)}\ dx.$

Эту формулу используют для построения оптимальной шкалы квантования по уровню. Но прежде чем приступить к ее построению, рассмотрим одну из технических систем, в которой эта шкала применяется на практике.

На рисунке 3.19 изображена блок–схема варианта системы с использованием неравномерного квантования по уровню.

С помощью компрессора выполняется нелинейное преобразование входного сигнала $x(t)$ в выходной $y(t)$. Преобразование описывается формулой $y(t) = \varphi\left[x(t)\right]$.

При помощи экспандера выполняется обратное преобразование: $\overline{x}(t) = \varphi^{-1}\left[\overline{y}(t)\right]$.

Здесь $\varphi^{-1}(...)$ – функция, обратная функции $\varphi(...)$.

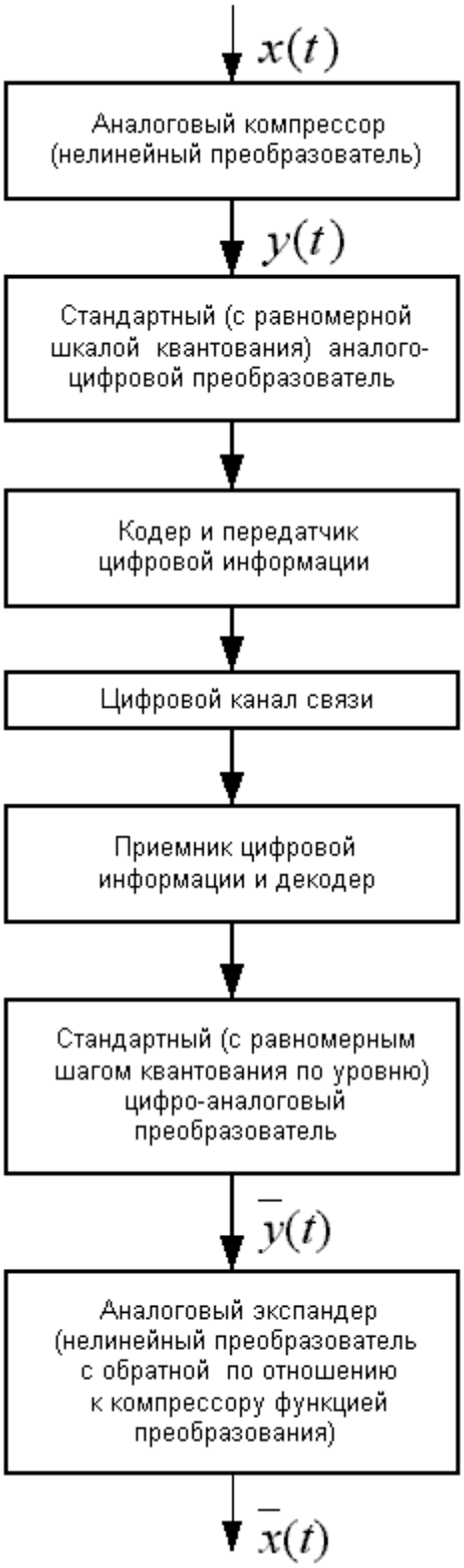


Рис. 3.19. Вариант системы с использованием неравномерного квантования

Связь между сигналами *x* и *y* и наоборот называется законом *компандирования* (от первых слогов слова *компрессор* и последних слогов слова *экспандер*).

Таким образом, закон компандирования связывает равномерную и неравномерную, в данном случае оптимальную, шкалы квантования по уровню. Это позволяет в технических решениях пользоваться стандартными равномерными аналого–цифровыми (АЦП) и цифро–аналоговыми (ЦАП) преобразователями.

График возможного варианта закона компандирования изображен на рисунке 3.20.

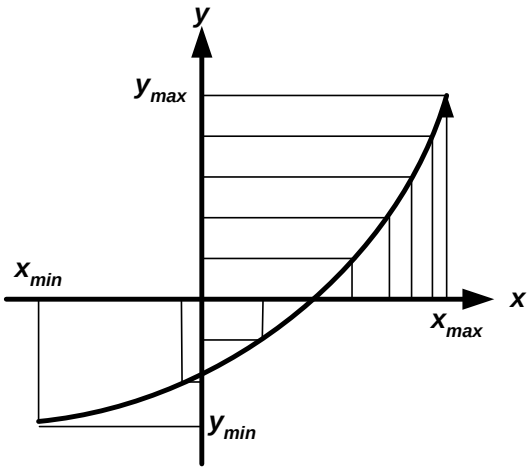


Рис. 3.20. График варианта закона компандирования

Найдем его формулу для случая оптимального квантования при следующих условиях:

$$y_{min} = x_{min};$$
$$y_{max} = x_{max};$$
$$y(x_i) = i \cdot \frac{y_{\max} - y_{\min}}{n} + y_{\min}.$$

Последнее условие отражает способность функции *y*(*x*) преобразовывать оптимальную неравномерную шкалу квантования сигнала *x*(*t*) в равномерную шкалу квантования сигнала *y*(*t*).

При больших *n* можно считать, что

$$x_i - x_{i-1} \approx \delta_i \approx dx \quad \text{и} \quad y_i - y_{i-1} = \frac{y_{\max} - y_{\min}}{n} = \frac{x_{\max} - x_{\min}}{n} \approx dy.$$

Тогда

$$\frac{dx}{dy} = \frac{\delta_i \cdot n}{x_{\max} - x_{\min}}.$$

Учитывая, что

$$\sum_{i=1}^{n} z_i = \sum_{i=1}^{n} \delta_i \cdot \sqrt[3]{f(x_i)} \approx \int_{-\infty}^{\infty} \sqrt[3]{f(x)}dx = S \quad \text{и} \quad z_i = S/n,$$

получим

$$\delta_i = \frac{S}{n\sqrt[3]{f(x_i)}},$$

тогда

$$\frac{dx}{dy} = \frac{S}{(x_{\max} - x_{\min})\sqrt[3]{f(x)}},$$

где $dy = \frac{(x_{\max} - x_{\min})}{S}\sqrt[3]{f(x)}\ dx$ – дифференциальное уравнение первого порядка с разделяющимися переменными.

Интегрируя левую его часть от $y_{\min}$ до y, а правую часть от $x_{\min}$ до x, получим

$$\int_{y_{\min}}^{y} dy = y - y_{\min} = y - x_{\min} = \frac{x_{\max} - x_{\min}}{S} \int_{x_{\min}}^{x} \sqrt[3]{f(\tau)}d\tau.$$

Отсюда получаем закон компандирования:

$$y = \frac{x_{\max} - x_{\min}}{S} \int_{x_{\min}}^{x} \sqrt[3]{f(\tau)}\, d\tau + x_{\min}.$$

Таким образом, закон компандирования описывается функцией $y = \varphi(x)$. К сожалению, пока не найдена простая обратная функция $x = \varphi^{-1}(y)$. Но именно эта обратная функция нам более всего и нужна, т.к. процедура построения оптимальной шкалы квантования предусматривает разбиение интервала $[y_{min}, y_{max}]$ на n равных шагов и определение границ интервалов неравномерной шкалы при помощи функции $x = \varphi^{-1}(y)$.

Поэтому процедура поиска границ шагов квантования неравномерной оптимальной шкалы квантования сейчас предусматривает использование компьютера для вычисления интеграла, входящего в правую часть формулы. Вычисление интеграла, как это обычно делается, заменяется вычислением суммы. По мере накопления слагаемых верхняя граница определенного интеграла растет. По мере пересечения значений y_i находятся соответствующие им значения x_i. В результате к концу интегрирования находятся все границы интервалов квантования неравномерной шкалы.

В заключение следует отметить, что в ряде случаев оптимальное квантование позволяет в десятки раз снизить дисперсию ошибки квантования или уменьшить число передаваемых отсчетов по сравнению со случаем применения равномерного квантования.

3.3.3. Расчет неравномерной оптимальной в смысле максимума количества информации в квантованном сигнале шкалы квантования

Другим, заслуживающим интерес критерием, является информационное содержание квантованного сигнала. Оно будет максимальным, если вероятности пребывания сигнала на всех уровнях будут равны, т. е.

$$f(x_i) \cdot \delta_i = 1/n,$$

отсюда

$$\delta_i = \frac{1}{n \cdot f(x)} \approx dx;$$

$$y_i - y_{i-1} = \frac{y_{\max} - y_{\min}}{n} = \frac{x_{\max} - x_{\min}}{n} \approx dy;$$

$$\frac{dx}{dy} = \frac{\cancel{n}}{(x_{\max} - x_{\min}) f(x) \cancel{n}};$$

$$dy = (x_{\max} - x_{\min}) f(x) dx;$$

$$\int_{y_{\min}}^{y} dy = y - x_{\min} = (x_{\max} - x_{\min}) \int_{x_{\min}}^{x} f(\tau) d\tau;$$

$$y = (x_{\max} - x_{\min}) \int_{x_{\min}}^{x} f(\tau) d\tau + x_{\min}.$$

Нелинейное квантование нашло применение, в частности, в цифровой телефонной связи и позволяет качественно передавать как слабые, так и мощные сигналы, уменьшить загрузку каналов связи.

3.3.4. Закон компандирования при условии равномерного закона распределения квантуемого сигнала

Закон распределения задается формулой

$$f(x) = \begin{cases} \dfrac{1}{x_{\max} - x_{\min}} & при \quad x_{\min} < x < x_{\max} \\ 0 \quad при \quad x < x_{\min} \quad и \quad x > x_{\max} \end{cases}$$

1. Критерий минимума дисперсии

$$y = \frac{x_{\max} - x_{\min}}{S} \int_{x_{\min}}^{x} \frac{1}{\sqrt[3]{x_{\max} - x_{\min}}} d\tau + x_{\min};$$

$$S = \int_{x_{\min}}^{x_{\max}} \frac{1}{\sqrt[3]{x_{\max} - x_{\min}}} d\tau .$$

После подстановки получаем

$$y = \frac{x_{\max} - x_{\min}}{x_{\max} - x_{\min}} (x - x_{\min}) + x_{\min} = x.$$

Таким образом, оптимальной в случае квантования равномерно распределенного сигнала является равномерная шкала квантования.

2. Критерий максимальной информативности:

$$y = \cancel{(x_{\max} - x_{\min})} \int_{x_{\min}}^{x} \frac{1}{\cancel{(x_{\max} - x_{\min})}} d\tau = x.$$

Таким образом, в случае использования критерия максимальной информативности оптимальной также является равномерная шкала квантования.

Контрольные вопросы

1. Какие способы представления моделей сигналов известны?

2. В чем заключаются преимущества частотного метода представления сигналов?

3. При каких условиях периодическая функция может быть представлена рядом Фурье?

4. Что понимается под спектром амплитуд и спектром фаз?

5. Каковы характерные особенности спектра периодического сигнала?

6. Как в спектре амплитуд отображается постоянная составляющая периодического сигнала?

8. Каков спектр гармонического сигнала?

9. Как можно энергетически истолковать спектр периодического сигнала?

10. Что понимается под практической шириной спектра периодического сигнала?

11. В чем состоит критерий выбора практической шириной спектра периодического сигнала?

12. Как выглядит спектр периодической последовательности прямоугольных импульсов?

13. Какой физический смысл имеет огибающая спектра амплитуд периодического сигнала?

14. Каков физический смысл ряда Котельникова?

15. В чем состоит физический смысл теоремы Котельникова?

16. Что такое функция отсчетов и каков ее вид?

17. В чем заключаются противоречия частотного критерия Котельникова?

18. Что собой представляет дискретное преобразование Фурье?

19. В чем состоит роль алгоритмов быстрого преобразования Фурье?

20. Какова реакция идеального фильтра нижних частот с граничной частотой ω_c на дельта–импульс?

21. Что такое координатная детерминированная функция времени?

22. Почему невозможно непосредственное приложение классического гармонического анализа к случайным процессам?

23. Каковы основные свойства корреляционной функции стационарного случайного процесса?

24. Каковы основные свойства спектральной плотности мощности стационарного случайного процесса?

25. Какой случайный процесс называется белым шумом?

26. Какова связь между спектральной плотностью мощности и корреляционной функцией случайного процесса?

27. Что такое интервал корреляции случайного процесса ?

28. Можно ли дискретные отсчеты по частотному критерию Котельникова считать статистически независимыми?

Глава 4. Кодирование информации

Кодирование − операция отождествления символов или групп символов одного кода с символами или группами символов другого кода.

Код (франц. *code*) − совокупность знаков (символов) и система определенных правил, при помощи которых информация может быть представлена (закодирована) в виде набора из таких символов для передачи, обработки и хранения (запоминания). Конечная последовательность кодовых знаков называется *словом*. Наиболее часто для кодирования информации используют буквы, цифры, числа, знаки [например, «−« (тире), «.« (точка)] и их комбинации].

Рассмотрим обобщенную схему передачи цифровой информации (рис. 4.1).

Рис. 4.1. Схема передачи цифровой информации

Рассматриваемые ниже принципы кодирования информации справедливы как для систем, основная задача которых – передача информации в пространстве (системы связи), так и для систем, основная задача которых – передача информации во времени (системы хранения информации). В последних линией связи считается среда, в которой хранится информация.

Если источник информации вырабатывает непрерывное сообщение, между ним и кодером источника следует установить квантователь по уровню и дискретизатор по времени.

Если же приемник требует представления информации в виде непрерывного сигнала, на выходе декодера источника ставится цифро–аналоговый преобразователь и, возможно, устройство интерполяции для уменьшения ошибки воспроизведения.

На вход кодера источника поступает последовательность символов, входящих во множество, называемое *первичным алфавитом*.

Кодер преобразует эту последовательность в другую, составленную чаще всего из других знаков, множество которых образует *вторичный алфавит*.

Если операции кодирования–декодирования выполняются компьютером и промежуточное представление информации человеку не требуется, эти алфавиты чаще всего состоят из 2 букв – 0 и 1 или *истина* и *ложь*. Это связано с особенностями современных цифровых устройств и их элементной базы.

Поскольку объем вторичного алфавита, как правило, меньше объема первичного алфавита и ввиду особенностей алгоритмов кодирования каждой букве или группе букв первичного алфавита ставится в соответствие группа букв вторичного алфавита, которую называют *кодовой комбинацией* или *кодовым словом*. Число символов (букв вторичного алфавита) в кодовой комбинации называют ее *значность*.

Коды, вторичный алфавит которых состоит из двух букв, обычно 0 и 1, называют *двоичными*.

Для кодовых слов двоичных кодов определено понятие веса кодового слова. Под *весом кодового слова* понимают количество символов 1, входящих в это кодовое слово.

Операция образования последовательности кодовых комбинаций по последовательности букв или комбинаций первичного алфавита называется *кодированием*. Выполняется кодирование при помощи *кодера*, который может быть устройством, программой или аппаратно–программным комплексом.

Обратная операция называется *декодированием* и выполняется *декодером*.

Кодирование выполняется для защиты, удобства хранения и обработки информации, что имеет место в компьютерной технике, и с теми же целями в системах связи.

Решающий вклад в теорию передачи информации в конце 40–х годов, как известно, внес американский ученый К. Шеннон. Он обосновал, в частности, эффективность ввода в систему передачи информации кодирующих–декодирующих устройств, основное назначение которых – согласование свойств источника сообщений со свойствами канала связи.

Одно из них – *кодер источника* – обеспечивает такое кодирование, при котором путем устранения избыточности существенно уменьшается объем данных, передаваемых по каналу связи.

Такое кодирование называется *эффективным* или *оптимальным*.

При наличии помех в канале связи это кодирование преобразует данные в форму, удобную для последующего помехоустойчивого кодирования.

Помехоустойчивое кодирование выполняется *кодером канала*. Его назначение следует из названия – борьба с искажениями, возникающими вследствие помех, действующих в канале связи.

Таким образом, при помощи кодера источника на основе статистических характеристик кодируемого сообщения уменьшается его объем, а при помощи кодера канала на основе статистических характеристик, действующих в канале помех, обеспечивается восстановление передаваемых через канал данных.

Такое разделение существенно упрощает исследование и проектирование кодеров–декодеров.

В случае малой избыточности можно отказаться от использования кодера–декодера источника. При малой же интенсивности помех в канале связи целесообразно отказаться от помехоустойчивого кодирования и, соответственно, кодера–декодера канала.

Носителями информации в конкретных линиях связи являются сигналы, которые для этого подвергаются модуляции. Модуляция и обратное преобразование – демодуляция – также выполняются с помощью предназначенных для этого устройств – модуляторов и демодуляторов. Подробнее вопросы, связанные с модуляцией будут рассмотрены в одном из следующих разделов курса. Отметим

только, что в случае хранения информации в оперативной памяти компьютера, модуляция не используется. Модуляция–демодуляция используется в компьютерных сетях, для чего предназначены устройства, называемые *модемами*.

4.1. Элементы теории кодирования

Некоторые общие свойства кодов рассмотрим на простых примерах. Предположим, что дискретный источник без памяти, т.е. дающий независимые сообщения – буквы – на выходе, имеет алфавит из k букв, поступающих с вероятностями $p_1, p_2, \ldots, p_k$.

Каждая буква кодируется с использованием вторичного алфавита из m букв. Обозначим n_i – число букв в кодовом слове, которое соответствует i–й букве источника. Нас будет интересовать средняя длина кодового слова n_{cp}, которая находится по известной формуле

$$n_{cp} = \sum_{i=1}^{k} p_i n_i \qquad (4.1)$$

В качестве примера рассмотрим возможные способы составления кодовых слов двоичного кода при объеме первичного алфавита $k = 4$ (таблица 4.1).

Заметим, что код 1 имеет неудачное свойство, состоящее в том, что буквы с номерами 1 и 2 кодируются одинаково – кодовым словом 0. Поэтому однозначное декодирование при использовании этого кода невозможно.

Таблица 4.1

Возможные способы составления кодовых слов двоичного кода при объеме первичного алфавита $k = 4$

i **(номер буквы первичного алфавита)**	p_i	**Код 1**	**Код 2**	**Код 3**	**Код 4**
1	0,5	0	0	0	0
2	0,25	0	1	10	01
3	0,125	1	00	110	011
4	0,125	10	11	111	0111

Код 2 обладает подобным же недостатком: последовательность 11 будет закодирована в 00, что совпадает с кодовым словом для буквы 3. Затруднений для декодирования в этом случае не будет только тогда, когда между кодовыми словами будет добавлен какой–либо разделительный символ. Но если такой символ ввести, то его можно понимать как дополнительную букву вторичного алфавита. Тогда каждое кодовое слово в конце или в начале просто дополняется этой буквой. В результате такой код можно рассматривать как частный случай кодов без разделительных символов. По этой причине коды с разделительными символами отдельно не рассматриваются.

Примером такого кода может служить код 4, где 0 можно понимать как разделительный символ.

Из вышесказанного следует, что код является *однозначно декодируемым*, если кодовые слова, соответствующие различным буквам первичного алфавита, различны. Это необходимое условие однозначной декодируемости, однако не всегда достаточное.

С точки зрения теории и практики наиболее важными среди всех однозначно декодируемых кодов являются так называемые *префиксные коды*.

Префиксными называют коды, в которых каждое кодовое слово не является начальной частью (префиксом) другого кодового слова. Это уже достаточное условие однозначной декодируемости.

Нетрудно заметить, что коды 1 и 2 не являются префиксными.

Коды удобно описывать графически с помощью графа, называемого *кодовым деревом*. Это дерево строится из одной вершины и имеет узлы нулевого, первого, второго и так далее порядков. Из каждого узла выходит m лучей – рёбер, каждому из которых соответствует определенная буква вторичного алфавита.

На рисунке 4.2 изображено кодовое дерево кода 3, описанного в таблице 4.1.

Каждому i–му кодовому слову длиной n_i ставится в соответствие узел x_i i–го порядка и определенный путь по кодовому дереву от корня до этого узла.

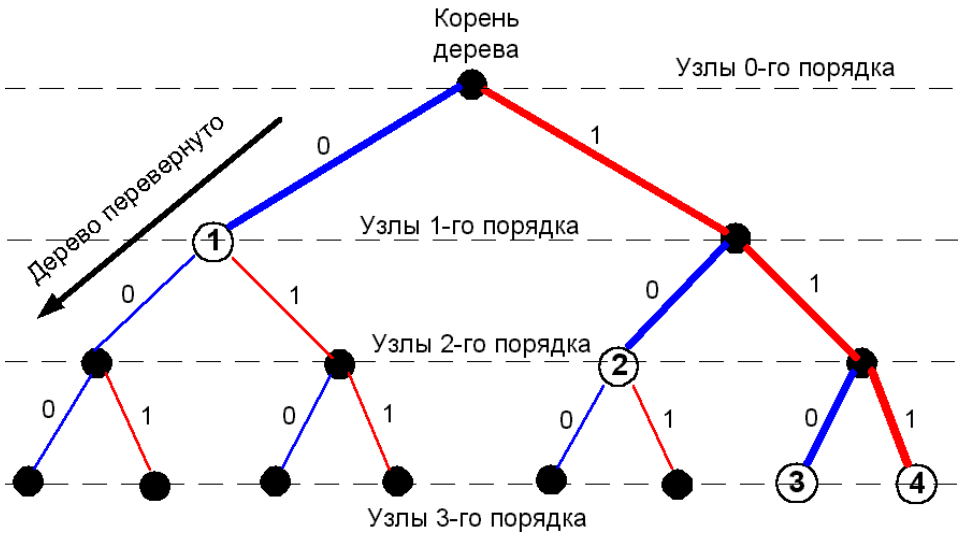


Рис. 4.2. Пример кодового дерева двоичного кода

Из графического представления следует, что код будет префиксным, если узлы, соответствующие кодовым словам, являются концевыми, т. е. через них не лежит путь к другим узлам, также соответствующим кодовым словам.

Заметим, что из каждого узла кодового дерева выходит m ребер и доля всех узлов любого порядка выше 0–го, приходящегося на один узел 1–го порядка равна $1/m=m^{-1}$. Далее, через 1 узел второго порядка проходят пути к $1/m^2=m^{-2}$ части всех узлов любого из узлов более высокого порядка и т. д. В общем случае через 1 узел i–го порядка проходят пути к m^{-i} части узлов любого из узлов более высокого порядка.

Учитывая последнее, докажем теорему, известную как *Неравенство Крафта*.

4.1.1. Неравенство Крафта

Теорема 1. Если целые числа $n_1, n_2, \ldots, n_k$ удовлетворяют неравенству

$$\sum_{i=1}^{k} m^{-n_i} \le 1,$$

то существует префиксный код с алфавитом объемом m, длины кодовых слов которого равны этим числам, и обратно, длины кодовых слов любого префиксного кода отвечают неравенству Крафта.

Доказательство

Предположим, что неравенство Крафта $\sum_{i=1}^{k} m^{-n_i} \le 1$ выполняется. Построим соответствующий префиксный код (первая часть теоремы).

Упорядочим длины n_i в порядке возрастания ($n_1 \le n_2 \le ... \le n_k$).

Выберем какой–либо узел x_1 порядка n_1 в полном кодовом дереве в качестве концевого. Это запретит использование m^{-n_1} части узлов n_1–го и более высокого порядка для выбора в качестве концевых для других кодовых слов.

Среди тех узлов, которые еще можно использовать в качестве концевых, выберем узел x_2 порядка n_2 в качестве концевого. Это запретит использование еще m^{-n_2} части узлов n_2–го и более высокого порядка для выбора в качестве концевых для других кодовых слов и т. д.

Сумма $\sum_{i=1}^{j} m^{-n_i}$ на j–м этапе меньше 1 и это значит, что для следующего $i+1$–го кодового слова еще есть возможность выбора узла.

Запрещенные к использованию узлы не пересекаются. Следовательно, общая доля запрещенных узлов равна сумме долей, запрещенных к использованию уже выбранными концевыми узлами. Эта доля равна $\sum_{i=1}^{k} m^{-n_i}$ по предположению меньше либо равно 1, а это означает, что остались некоторые узлы более высокого порядка, разрешенные к использованию в качестве концевых.

Для этого мы расположили n в порядке возрастания. Для пояснения нужно обратиться к кодовому дереву (рис. 4.2).

Для доказательства второй части теоремы о том, что длины любого префиксного кода отвечают неравенству Крафта, заметим, что префиксному коду соответствует определенный набор концевых узлов, каждый из которых закрывает m^{-n_i} долю узлов того же и

более высокого порядка для использования в качестве концевых. Эти доли не пересекаются, следовательно, их сумма не может превышать 1. Отсюда $\sum_{i=1}^{k} m^{-n_i} \le 1$, что и требовалось доказать.

4.1.2. Теорема об обобщении некоторых результатов, полученных для префиксных кодов, на все однозначно декодируемые коды

Докажем теперь теорему, позволяющую обобщать некоторые результаты, полученные для префиксных кодов на все однозначно декодируемые коды.

Теорема 2. Пусть задан код с длинами кодовых слов n_1, n_2, ..., n_k и с алфавитом объема m. Если код однозначно декодируем, неравенство Крафта $\sum_{i=1}^{k} m^{-n_i} \le 1$ удовлетворяется.

Доказательство.

Пусть z – произвольное положительное число.

Рассмотрим тождество

$$\left(\sum_{i=1}^{k} m^{-n_i}\right)^z = \sum_{i_1=1}^{k} \ldots \sum_{i_z=1}^{k} m^{-(n_{i_1}+n_{i_2}+\ldots+n_{i_z})} .$$

Заметим, что $n_{i_1} + n_{i_2} + \ldots + n_{i_z}$ можно понимать как сумму длин z кодовых слов. Обозначим эту сумму через j ($j = n_{i_1} + n_{i_2} + \ldots + n_{i_z}$), а количество комбинаций, дающих одинаковую сумму длин A_j. Тогда исходное тождество можно записать в виде

$$\left(\sum_{i=1}^{k} m^{-n_i}\right)^z = \sum_{j=1}^{zn_{\max}} A_j m^{-j} .$$

Если код однозначно декодируем, то все последовательности кодовых слов с суммарной длиной j кодовых букв (букв вторичного алфавита) являются *различными*. Поэтому $A_j \le m^j$, т. к. m^j – общее число всевозможных комбинаций из m букв по j. Из этого следует

$$\left(\sum_{i=1}^{k} m^{-n_i}\right)^z = \sum_{j=1}^{zn_{\max}} A_j m^{-j} \le \sum_{j=1}^{zn_{\max}} m^j m^{-j} = \sum_{j=1}^{zn_{mzx}} 1 = zn_{\max};$$

$$\sum_{i=1}^{k} m^{-n_i} \le (zn_{\max})^{1/z} .$$

Поскольку z – произвольное число, устремим его в бесконечность:

$$\lim_{z\to\infty}(zn_{\max})^{1/z} = 1,$$

что можно доказать, используя правило Лопиталя.

Отсюда следует, что $\sum_{i=1}^{k} m^{-n_i} \le 1$ – неравенство Крафта.

Согласно теореме 1 существует префиксный код с теми же $n_1, n_2, \ldots, n_k$. Поэтому все результаты, полученные для префиксных кодов *относительно средней длины кодового слова* n_{cp}, распространяются и на все однозначно декодируемые коды.

Теорема 2 доказана.

4.1.3. Основная теорема кодирования для канала связи без шума

Теорема 3. При заданной энтропии H источника и объеме m вторичного алфавита существует префиксный код с минимальной средней длиной $n_{cp\ min}$ кодовых слов, отвечающей следующему неравенству:

$$\frac{H}{\log_2 m} \le n_{cp\min} < \frac{H}{\log_2 m} + 1 .$$

Доказательство

Докажем сначала левую часть неравенства, для чего перепишем его в виде:

$$H - n_{cp\min}\log_2 m \le 0.$$

Далее выполним преобразования:

$$H - n_{cp\min}\log_2 m = -\sum_{i=1}^{k} p_i \log_2 p_i - \sum_{i=1}^{k} p_i n_i \log_2 m =$$
$$= \sum_{i=1}^{k}\left(p_i \log_2 \frac{1}{p_i} + p_i \log_2 m^{-n_i} \right) = \sum_{i=1}^{k} p_i \log_2 \frac{m^{-n_i}}{p_i}.$$

Если разложить логарифм $log_2 x$ в ряд Тэйлора (степенной ряд) и взять первые 2 члена этого ряда, получим следующее неравенство: $\log_2 x \le (x - 1)\log_2 e$.

Используя это неравенство, продолжим преобразование нашего выражения:

$$H - n_{cp\min}\log_2 m = \sum_{i=1}^{k} p_i \log_2 \frac{m^{-n_i}}{p_i} \le$$
$$\le \sum_{i=1}^{k} p_i \left(\frac{m^{-n_i}}{p_i} - 1 \right) \log_2 e = \log_2 e \cdot \left(\sum_{i=1}^{k} m^{-n_i} - \sum_{i=1}^{k} p_i \right).$$

Поскольку n_i – длины кодовых слов префиксного кода, согласно неравенству Крафта первая сумма в скобках $\sum_{i=1}^{k} m^{-n_i} \le 1$. Вторая сумма равна 1, т. к. является полной суммой вероятностей. Следовательно, уменьшаемое в скобках меньше вычитаемого и сама разность отрицательна. Логарифм $\log_2 e > 0$, следовательно, само выражение отрицательно: $H - n_{cp\min}\log_2 m \le 0$. Отсюда следует, что

$$\frac{H}{\log_2 m} \le n_{cp\min},$$

и, таким образом, левая часть неравенства доказана.

Докажем теперь правую часть неравенства.

Заметим, что при $p_i = m^{-n_{iид}}$ логарифм в правой части равенства

$$H - n_{ср\min} \log_2 m = \sum_{i=1}^{k} p_i \log_2 \frac{m^{-n_i}}{p_i}$$

становится равным 0. Тогда неравенство превращается в равенство

$$\frac{H}{\log_2 m} = n_{ср\min}.$$

Достичь этого можно было бы, если бы идеальные длины $n_{iид}$ кодовых слов находились из выражения $p_i = m^{-n_{iид}}$ ($n_{iид} = -\log_m p_i$), однако в общем случае они должны бы тогда быть дробными числами, что невозможно. Эти длины должны быть целыми числами и, кроме того, их значения должны соответствовать неравенству Крафта (напомним, что теорема доказывается для префиксного кода): $p_i < m^{-n_i}$. Значит, в качестве реальных длин кодовых слов можно выбирать ближайшее большее целое число. Максимальное же отличие реальной и идеальной длины при этом может приближаться к 1. Если предположить наихудший случай, когда все длины отличаются от идеальной на 1, то

$$p_i < m^{-n_i+1}.$$

Отсюда, логарифмируя левую и правую части неравенства, находим

$$n_i < \frac{-\log_2 p_i}{\log_2 m} + 1.$$

Умножая правую и левую части этого неравенства на p_i и суммируя результаты по всем i, получаем правую часть неравенства.

Таким образом, теорема 3 о минимальной средней длине кодового слова (основная теорема кодирования для канала без шума) доказана.

4.1.4. Теорема о минимальной средней длине кодового слова при поблочном кодировании

Рассмотрим теперь случай кодирования не отдельных букв источника, а последовательностей из L букв.

Теорема 4. Для данного дискретного источника без памяти (с независимыми буквами) и энтропией H и объемом вторичного алфавита m при кодировании блоками по L букв существует префиксный код со средней длиной кодового слова, отвечающей неравенству

$$\frac{H}{\log_2 m} \le n_{cp\min} < \frac{H}{\log_2 m} + 1/L \quad .$$

Доказательство

Последовательность из L букв можно рассматривать как новую супербукву. Тогда можно использовать теорему 3, согласно которой минимальная средняя длина кодового слова для такого сообщения ограничивается неравенством:

$$\frac{H_L}{\log_2 m} \le n_{cp\min L} < \frac{H_L}{\log_2 m} + 1 \ .$$

Так как отдельные буквы предполагаются независимыми, то энтропия последовательности L букв $H_L = L \cdot H$. Кроме того, средняя длина кода для последовательности L букв $n_{cp\min L} = n_{cp\min} \cdot L$, тогда

$$\frac{L \cdot H}{\log_2 m} \le n_{cp\min} \cdot L < \frac{L \cdot H}{\log_2 m} + 1 \ .$$

После деления всех частей неравенства на L получаем окончательный результат:

$$\frac{H}{\log_2 m} \le n_{cp\min} < \frac{H}{\log_2 m} + 1/L \ .$$

Из теоремы 4 следует, что если кодировать блоками, то, увеличивая размер блоков L, можно сколь угодно приблизиться к нижней границе средней длины кодового слова, т. е.

$$\lim_{L \to \infty} n_{cp\min} = \frac{H}{\log_2 m} \quad при \quad L \to \infty .$$

4.2. Оптимальные неравномерные коды

Неравномерными называют коды, кодовые слова которых имеют различную длину.

Оптимальность можно понимать по–разному, в зависимости от критерия. В данном случае таким критерием является средняя длина кодового слова. ***Оптимальность*** *с учетом этого критерия понимается в смысле минимальной длины средней длины кодового слова*.

Дальнейшие выводы будем делать при следующих условиях.

1. Пусть буквы источника $a_1, a_2, \ldots, a_k$ имеют вероятности появления $p_1, p_2, \ldots, p_k$ Упорядочим буквы в порядке убывания вероятностей их появления в сообщении и пронумеруем их в этом порядке. В результате $p_1 \ge p_2 \ge \ldots \ge p_k$. Для кодирования будем использовать вторичный алфавит, состоящий из 2 букв – 0 и 1, т. е. двоичный код.

2. Пусть $x_1, x_2, \ldots, x_k$ − множество кодовых слов, имеющих длину $n_1, n_2, \ldots, n_k$. Ограничимся также рассмотрением только префиксных кодов. Результаты, полученные в отношении длин кодовых слов для префиксных кодов, согласно теоремам 1 и 2 можно

распространять на весь класс однозначно декодируемых кодов, а результаты, полученные для двоичных кодов несложно обобщить на коды с любым объемом вторичного алфавита.

Докажем 2 леммы[1].

Лемма 1. О существовании оптимального кода с одинаковой длиной кодовых слов двух наименее вероятных кодируемых букв.

Формулировка. Для любого источника с $k>=2$ буквами существует оптимальный (в смысле минимума средней длины кодового слова) двоичный код, в котором два наименее вероятных слова x_k и x_{k-1} имеют одну и ту же длину и отличаются лишь последним символом.

Пример: если существует кодовое слово x_k = 0 1 1 0 0 0 1 0, то существует и кодовое слово x_{k-1} = 0 1 1 0 0 0 1 1

Доказательство

1. Докажем сначала, что для оптимального кода $n_i \le n_k$, где $i < k$, т. е. последнее кодовое слово, имеющее, кстати, наименьшую вероятность появления, имеет наибольшую длину.

Для доказательства этого вспомним, что $n_{cp} = \sum_{i=1}^{k} p_i n_i$.

Найдем на сколько изменится n_{cp}, если $n_k = n_{\max}$ и попытаемся поменять его с любым другим j–м словом длиной n_j, т. е. a_j кодировать x_k и a_k кодировать x_j.

После такой перестановки получим:

$$n_{cpnp} = \sum_{i=1;i\neq j}^{k-1} p_i n_i + p_j n_k + p_k n_j \ ,$$

$$\Delta n_{cp} = n_{cpnp} - n_{cp} = \left(\sum_{i=1,i\neq j}^{k-1} p_i n_i + p_j n_k + p_k n_j \right) - \sum_{i=1}^{k} p_i n_i =$$

$$= p_j n_k + p_k n_j - p_j n_j - p_k n_k = p_j (n_k - n_j) - p_k (n_k - n_j) =$$

$$= (p_k - p_j)(n_j - n_k)$$

[1] Лемма - вспомогательная теорема, необходимая для доказательства других теорем.

Если код оптимален, то любая перестановка кодовых слов должна приводить с росту средней длины кодового слова – к тому, что $\sum n_{cp} \geq 0$ и к превращению кода из оптимального в неоптимальный.

Докажем, что при принятых нами условиях действительно $\sum n_{cp} \geq 0$.

Поскольку, согласно постулированному[1] условию, буквы первичного алфавита упорядочены в порядке убывания вероятностей, разность $p_k - p_j \leq 0$. С другой стороны, $n_k - n_j \geq 0$, тогда, когда $n_k \geq n_j$, что и требовалось доказать.

2. Заметим далее, что в любом оптимальном префиксном коде, для которого $n_k = n_{max}$, должно быть другое кодовое слово, отличающееся от n_k только буквой в последнем разряде. Например, если x_k = 10110, то должно еще быть слово x_i = 10111. Это следует из того, что в противном случае этот последний разряд можно было бы отсечь без нарушения свойства префикса и n_{cp} уменьшилось бы, что для оптимального кода невозможно. Следовательно, кодовое слово, отличающееся только последним разрядом, должно существовать.

3. Наконец, если x_i отличается от x_k только в одной позиции, то n_i=n_k и, x_i и x_{k-1} можно поменять местами без увеличения n_{cp}.

Таким образом, нами доказано, что кодовые слова x_k и x_{k-1} оптимального префиксного кода имеют одинаковую и максимальную длину n_k=n_{k-1} и отличаются лишь одной буквой вторичного алфавита.

С помощью этой леммы решение задачи построения оптимального кода сводится к решению задачи построения кодовых слов x_1, x_2, … , x_{k-2} и отысканию первых n_{k-1} символов кодового слова x_k.

Следуя принятой терминологии, назовем совокупность $\{a\}$ букв первичного алфавита, упорядоченных в порядке убывания ве-

---

[1] *Постулат* (от лат. *postulatum* − требование) − утверждение (суждение), принимаемое в рамках какой-либо научной теории за истинное, хотя и недоказуемое ее средствами, и поэтому играющее в ней роль аксиомы.

роятностей появления и вероятностей $\{p\}$ их появления *ансамблем* U.

Определим теперь *редуцированный*[1] *ансамбль* U' как ансамбль, полученный из исходного нередуцированного ансамбля путем замены двух последних (с наименьшей вероятностью) букв исходного ансамбля одной новой буквой редуцированного ансамбля.

Таким образом, редуцированный ансамбль отличается от исходного нередуцированного меньшим на одну букву количеством букв первичного алфавита. Все буквы редуцированного ансамбля, кроме одной, совпадают с буквами исходного ансамбля. Одна же буква редуцированного ансамбля заменяет две последние (с наименьшей вероятностью появления) буквы исходного ансамбля. Вероятность появления этой новой буквы редуцированного ансамбля при такой замене, естественно, равна сумме вероятностей появления двух букв исходного ансамбля, которые она заменяет.

Буквы алфавита редуцированного ансамбля так же, как буквы алфавита нередуцированного ансамбля, упорядочиваются в порядке убывания вероятностей их появления в сообщении. В результате такого упорядочения положение даже одних и тех же букв в исходном и редуцированном ансамбле может быть различным.

Любой префиксный код редуцированного ансамбля U', можно превратить в код исходного ансамбля простым добавлением концевого символа 0 к кодовому слову x'_{k-1} символа 1 к кодовому слову x'_k. Это устанавливает взаимно однозначное соответствие между множеством префиксных кодов редуцированного ансамбля U' и множеством префиксных кодов исходного ансамбля U.

Лемма 2. Об оптимальности префиксного кода нередуцированного ансамбля, если префиксный код редуцированного ансамбля оптимален

Формулировка. Если некоторый префиксный код редуцированного ансамбля U' является оптимальным, то соответствующий ему префиксный код исходного ансамбля также является оптимальным.

Доказательство

---

[1] *Редукция* (от лат. *reductio* − возвращение, приведение обратно) − упрощение, сведение сложного к более простому, обозримому, понимаемому, более доступному для анализа или решения; уменьшение, ослабление чего-либо.

Заметим, что нередуцированный код от редуцированного отличается только заменой одного кодового слова редуцированного кода двумя кодовыми словами нередуцированного. Все же остальные кодовые слова этих кодов совпадают. Кроме того, поскольку согласно формулировке леммы редуцированный код оптимален, средняя длина его кодовых слов минимальна. Если удастся доказать, что замена одного кодового слова редуцированного кода двумя кодовыми словами нередуцированного приводит к минимальному увеличению средней длины кодового слова нередуцированного кода, то эта средняя длина также является минимально возможной и, следовательно, полученный таким образом нередуцированный код является оптимальным.

Для доказательства запишем выражения для средней длины кодовых слов редуцированного и нередуцированного кода.

Средняя длина нередуцированного кода

$$n_{cp} = \sum_{i=1}^{k} p_i n_i = \sum_{i=1}^{k-2} p_i n_i + \left(p_{k-1} + p_k \right) n_k ,$$

поскольку, согласно лемме 1, два последних кодовых слова двоичного оптимального кода имеют одинаковую длину и отличаются символом последнего разряда.

Средняя длина редуцированного кода

$$n'_{cp} = \sum_{j=1; j \neq r}^{k-1} p'_j n'_j + p_r n_r = \sum_{j=1; j \neq r}^{k-1} p'_j n'_j + \left(p_{k-1} + p_k \right) \left(n_k - 1 \right),$$

где r − номер кодового слова редуцированного кода, полученного путем объединения двух последних кодовых слов исходного нередуцированного кода.

Надо учесть, что после этого все кодовые слова редуцированного кода переупорядочиваются в порядке убывания вероятностей, а r – номер объединенного кодового слова после этого упорядочения.

Новое кодовое слово получается из кодовых слов, которые оно заменяет путем отбрасывания последнего разряда, которым они отличаются друг от друга. Поэтому его длина на 1 меньше.

Теперь найдем изменение средней длины Δn_{cp}

$$n'_{cp} = n_{cp} - n'_{cp} = \sum_{i=1}^{k-1} p_i n_i + \left(p_{k-1} + p_k\right) n_k -$$

$$- \sum_{j=1; jw}^{k-1} p'_j n'_j + \left(p_{k-1} + p_k\right)\left(n_k - 1\right) = p_{k-1} + p_k.$$

Это приращение минимально, т. к. p_{k-1} и p_k – наименьшие вероятности. Это и доказывает лемму 2.

Задача построения оптимального кода теперь сводится к построению последовательности редуцированных ансамблей:

$$U \to U' \to U'' \to U''' \to ....$$

При каждой редукции (уменьшении) число букв в первичном алфавите уменьшается на 1 букву и наступает момент, когда в результате редуцирования останутся только 2 буквы. Оптимальный код для такого ансамбля очевиден – 0 для одной буквы и 1 − для другой. Дальше, двигаясь в обратном направлении

$$... \to U''' \to U'' \to U' \to U,$$

переходя от оптимального кода редукции более высокого порядка к оптимальному коду редукции менее высокого порядка, что можно делать согласно лемме 2, можно дойти до оптимального кода исходного ансамбля U.

4.2.1. Коды Хаффмена

На этом алгоритме построена процедура построения оптимального кода, предложенная в 1952 г. Д. Хаффменом[1]:

1) буквы первичного алфавита выписываются в основной столбец в порядке убывания вероятностей;

2) две последние буквы объединяются в одну вспомогательную букву, которой приписывается суммарная вероятность;

3) вероятности букв, не участвовавших в объединении, и полученная суммарная вероятность снова располагаются в порядке убывания вероятностей в дополнительном столбце, а две последние объединяются;

4) процесс продолжается до тех пор, пока не получим единственную вспомогательную букву с вероятностью, равной единице.

Методика поясняется примером, представленным в таблице 4.2.

Для составления кодовой комбинации, соответствующей данному сообщению, необходимо проследить путь перехода вероятности сообщения по строкам и столбцам таблицы.

Для наглядности строится кодовое дерево. Из точки, соответствующей вероятности 1, направляются две ветви, причем ветви с большей вероятностью присваивается символ 1, а с меньшей − 0.

Такое последовательное ветвление продолжаем до тех пор, пока не дойдем до каждой буквы. Кодовое дерево для алфавита букв, рассматриваемого в таблице, изображено на рисунке 4.2.

Двигаясь по кодовому дереву сверху вниз, можно записать для каждой буквы соответствующую ей кодовую комбинацию:

m_1	m_2	m_3	m_4	m_5	m_6	m_7	m_8
01	00	111	110	100	1011	10101	10100

[1] Дэвид Хаффман (р. 9 августа 1925, Альянс, Огайо — 7 октября 1999, Санта Круз, Калифорния) — первопроходец в сфере информатики. В 1952 году создал алгоритм префиксного кодирования с минимальной избыточностью (известный как алгоритм или код Хаффмана). В 1999 году получил медаль Ричарда Хэмминга за исключительный вклад в Теорию информации

Таблица 4.2

Кодирование методом Хаффмана

Символы	Вероятности	Вспомогательные столбцы						
		1	2	3	4	5	6	7
m_1	0,22	0,22	0,22	0,26	0,32	0,42	0,52	1
m_2	0,20	0,20	0,20	0,22	0,26	0,32	0,42	
m_3	0,16	0,16	0,16	0,20	0,22	0,26		
m_4	0,16	0,16	0,16	0,16	0,20			
m_5	0,10	0,10	0,16	0,16				
m_6	0,10	0,10	0,10					
m_7	0,04	0,06						
m_8	0,02							

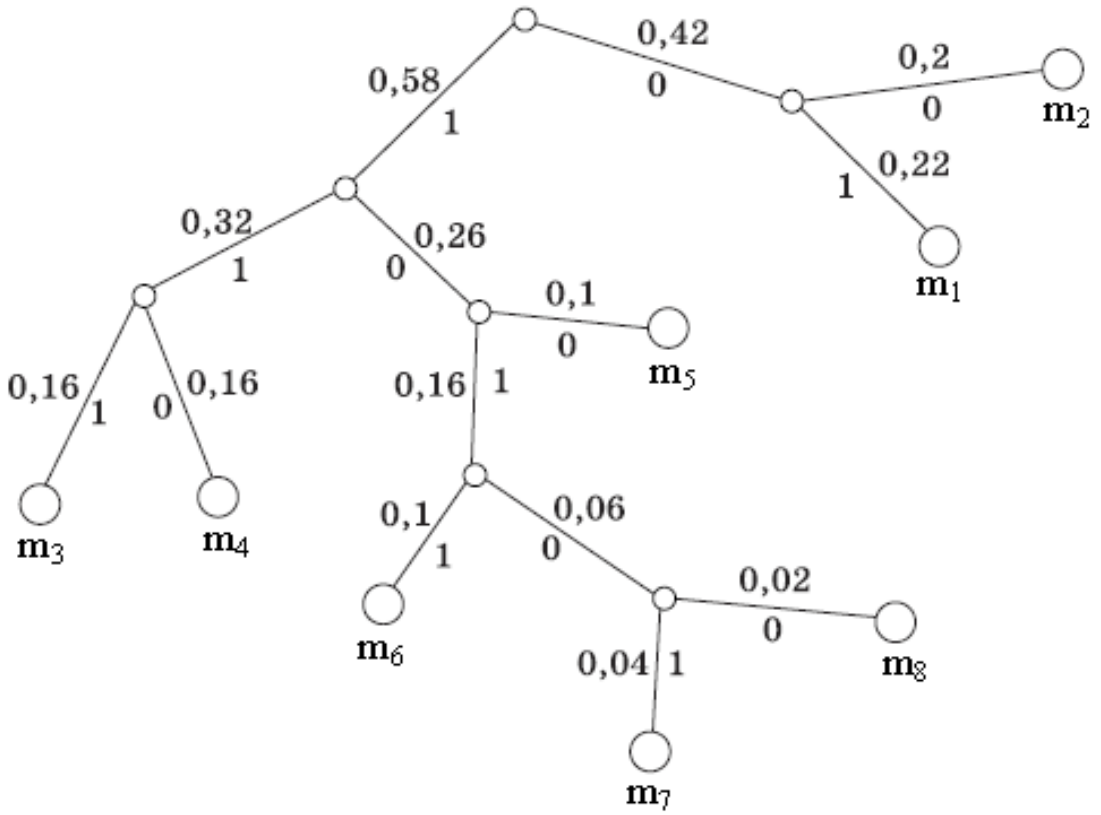


Рис.4.3. Кодовое дерево по алгоритму Хаффмана

4.2.2. Коды Шеннона − Фано

Ранее, независимо друг от друга Шенноном и Фано[1] была предложена другая процедура построения эффективного кода. Получаемый при ее помощи код называют *кодом Шеннона − Фано*.

Код Шеннона − Фано строится следующим образом:

1) буквы алфавита сообщений выписываются в таблицу в порядке убывания вероятностей;

2) затем они разделяются на две группы так, чтобы суммы вероятностей в каждой из групп были по возможности одинаковы;

3) всем буквам верхней половины в качестве первого символа приписывается 1, а всем нижним 0;

4) каждая из полученных групп, в свою очередь, разбивается на две подгруппы с одинаковыми суммарными вероятностями и т. д.;

5) процесс повторяется до тех пор, пока в каждой подгруппе останется по одной букве.

Рассмотрим алфавит из восьми букв. Ясно, что при обычном (не учитывающем статистических характеристик) кодировании для представления каждой буквы требуется три символа.

Наибольший эффект сжатия получается в случае, когда вероятности букв представляют собой целочисленные отрицательные степени двойки. Среднее число символов на букву в этом случае точно равно энтропии. Убедимся в этом, вычислив энтропию

$$H(Z) = -\sum_{i=1}^{8} p(z)\log p(z_i) = 1\frac{63}{64}$$

и среднее число символов на букву

$$l_{cp} = \sum_{i=1}^{8} p(z_i)n(z_i) = 1\frac{63}{64}$$

---

[1] Роберт Марио Фано, родился в 1917 г. в Италии в Турине. В период с 1950 по 1953 год он возглавлял группу изучения радиолокационных технологий в Лаборатории им. Линкольна, но в основном его работа была связана с МИТ. В начале 60-х он занимался разработкой компьютеров, работающих по принципу разделения времени. В 1976 г. Фано получил премию Клода Шеннона за вклад в развитие теории информации.

где $n(Zв)$ – число символов в кодовой комбинации[1], соответствующей букве Z_i. Характеристики такого ансамбля и коды букв представлены в таблице 4.3.

Таблица 4.3

Пример кодирования методом Шеннона–Фано

Буквы	Вероятности	Ступени разбиения							Кодовые комбинации
		1	2	3	4	5	6	7	
Z1	1/2	1							1
Z2	1/4	0	1						01
Z3	1/8	0	0	1					001
Z4	1/16	0	0	0	1				0001
Z5	1/32	0	0	0	0	1			00001
Z6	1/64	0	0	0	0	0	1		000001
Z7	1/128	0	0	0	0	0	0	1	0000001
Z8	1/128	0	0	0	0	0	0	0	0000000

В более общем случае для алфавита из восьми букв среднее число символов на букву будет меньше трех, но больше энтропии алфавита $H(Z)$.

В таблице 4.4, в отличие от таблицы 4.3, приведен пример кодирования произвольного ансамбля методом Шеннона - Фано. Для ансамбля букв, приведенного в таблице 4.4, энтропия равна 2,76, а среднее число символов на букву 2,84. Следовательно, некоторая избыточность в последовательностях символов осталась. Из теоремы Шеннона следует, что эту избыточность также можно уменьшить, если перейти к кодированию достаточно большими блоками.

[1] Число разрядов в кодовой комбинации здесь и в дальнейшем обозначено через n для того, чтобы избежать расхождения с общепринятой терминологией в области эффективного и помехоустойчивого кодирования.

Таблица 4.4

Пример кодирования методом Шеннона–Фано

Буквы	Вероятности	Ступени разбиения							Кодовые комбинации
		1	2	3	4	5	6	7	
Z1	0,22	1	1						11
Z2	0,20	1	0	1					101
Z3	0,16	1	0	0					100
Z4	0,16	0	1						01
Z5	0,10	0	0	1					001
Z6	0,10	0	0	0	1				0001
Z7	0,04	0	0	0	0	1			00001
Z8	0,02	0	0	0	0	0			00000

Рассмотрим сообщения, образованные с помощью алфавита, состоящего всего из двух букв Z_1 и Z_2, с вероятностями появления соответственно $p(Z_1)=0{,}9$ и $p(Z_2)=0{,}1$.

Поскольку вероятности не равны, то последовательность из таких букв будет обладать избыточностью. Однако при побуквенном кодировании никакого эффекта не получается.

Действительно, на передачу каждой буквы требуется символ либо 1, либо 0, в то время как энтропия равна 0,47.

При кодировании блоков, содержащих по две буквы, получим коды, показанные в таблице 4.5.

Таблица 4.5

Кодирование кодом Шеннона – Фано блоков по две буквы

Буквы	Вероятности	Ступени разбиения							кодовые комбинации
		1	2	3	4	5	6	7	
Z_1 Z_1	0,81	1							1
Z_2 Z_1	0,09	0	1						01
Z_3 Z_1	0,09	0	0	1					001
Z_4 Z_1	0,01	0	0	0					000

Так как буквы статистически не связаны, вероятности блоков определяются как произведение вероятностей составляющих букв.

Среднее число символов на блок получается равным 1,29; а на букву − 0,645.

Кодирование блоков, содержащих по три буквы, дает еще больший эффект. Соответствующий ансамбль и коды приведены в таблице 4.6.

Таблица 4.6

Кодирование кодом Шеннона – Фано блоков по три буквы

Буквы	Вероятности	Ступени разбиения							кодовые комбинации
		1	2	3	4	5	6	7	
$Z_1\ Z_1\ Z_1$	0,729	1							1
$Z_2\ Z_1\ Z_1$	0,081	0	1	1					01
$Z_1\ Z_2\ Z_1$	0,081	0	1	0					010
$Z_2\ Z_2\ Z_1$	0,081	0	0	1					001
$Z_1\ Z_1\ Z_2$	0,009	0	0	0	1	1			00011
$Z_2\ Z_1\ Z_2$	0,009	0	0	0	1	0			00010
$Z_1\ Z_2\ Z_2$	0,009	0	0	0	0	1			00001

Среднее число символов на блок равно 1,59; а на букву − 0,53; что всего на 12% больше энтропии. Теоретический минимум $H(Z) = 0{,}47$ может быть достигнут при кодировании блоков, включающих бесконечное число букв: $\lim\limits_{n\to\infty} l_{cp} = H(Z)$.

Следует подчеркнуть, что увеличение эффективности кодирования при укрупнении блоков, не связано с учетом все более далеких статистических связей, так как нами рассматривались алфавиты с некоррелированными буквами. Повышение эффективности определяется лишь тем, что набор вероятностей, получающийся при укрупнении блоков, можно делить на более близкие по суммарным вероятностям подгруппы.

Рассмотренная методика Шеннона – Фано не всегда приводит к однозначному построению кода. Ведь при разбиении на подгруппы можно сделать большей по вероятности как верхнюю, так и нижнюю подгруппу.

Множество вероятностей, приведенных в таблице 4.4, можно было бы разбить иначе, например, так, как это показано в таблице 4.7.

Таблица 4.7

Вариант кодирования методом Шеннона – Фано

Буквы	Вероятности	Ступени разбиения							кодовые комбинации
		1	2	3	4	5	6	7	
Z_1	0,22	1	1						11
Z_2	0,20	1	0						10
Z_3	0,16	0	0	1					001
Z_4	0,16	0	1	0					010
Z_5	0,10	0	0	1					001
Z_6	0,10	0	0	0	1				0001
Z_7	0,04	0	0	0	0	1			00001
Z_8	0,02	0	0	0	0	0			00000

При этом среднее число символов на букву оказывается равным 2,80. Таким образом, построенный код может оказаться не самым лучшим. При построении эффективных кодов с основанием $m>2$ неопределенность становится еще большей.

От указанного недостатка свободна методика Хаффмена. Она гарантирует однозначное построение кода с наименьшим для данного распределения вероятностей средним числом символов на букву.

4.2.3. Параметры эффективности оптимальных кодов

Таких параметров 2: коэффициент статистического сжатия и коэффициент относительной эффективности. Оба параметра характеризуют степень уменьшения средней длины кодового слова. При этом средняя длина кодового слова сравнивается с минимально возможной средней длиной кодового слова.

Коэффициент статистического сжатия определяется по формуле

$$Kcc = \frac{H\max}{(\log_2 m)\cdot n_{cp}} = \frac{\log_2 k}{\log_2 m \cdot \sum_{i=1}^{k} p_i n_i}, \qquad (4.2)$$

где k и m – объемы первичного и вторичного алфавитов; $\log_2 m$ учитывает объем вторичного алфавита. Для двоичных кодов (наиболее частый случай) $log_2 m = 1$.

Коэффициент статистического сжатия характеризует степень использования неодинаковости вероятностей появления букв для снижения средней длины кодового слова. Чем *Ксс* больше, тем лучше используется неодинаковость вероятностей для уменьшения средней длины кодового слова.

Коэффициент относительной эффективности рассчитывается по формуле

$$Коэ = \frac{H}{\left(\log_2 m\right) \cdot n_{ср}} = \frac{-\sum_{i=1}^{k} p_i \log_2 p_i}{\log_2 m \cdot \sum_{i=1}^{k} p_i n_i}. \qquad (4.3)$$

Данный коэффициент характеризует степень приближения $n_{ср}$ данного кода к минимально возможному значению, равному $H / log_2 m$.

Согласно основной теореме кодирования для канала связи без шума

$$\frac{H}{\log_2 m} \le n_{ср\min} < \frac{H}{\log_2 m} + 1, \qquad (4.4)$$

при кодировании отдельных символов и

$$\frac{H}{\log_2 m} \le n_{ср\min} < \frac{H}{\log_2 m} + 1/L \qquad (4.5)$$

при кодировании блоков по L символов первичного алфавита. Следовательно, $К_{оэ} \le 1$, и чем $К_{оэ}$ больше, тем $n_{ср}$ меньше и ближе к теоретически минимально возможному значению, тем код лучше.

4.2.4. Особенности эффективных кодов.

Особенности эффективных кодов заключаются в следующем.

Букве первичного алфавита с наименьшей вероятностью появления ставится в соответствие код с наибольшей длиной (лемма 1), т. е. такой код является неравномерным (с разной длиной кодовых слов). В результате, если моментами передачи сообщения от источника приемник управлять не может (например, кодовые слова поступают строго периодически с шагом Δt), через линию связи будут передаваться кодовые слова с разной длиной, т. е. количество передаваемых в единицу времени через линию связи букв вторичного алфавита будет меняться. Учитывая то, что любая линия связи характеризуется максимальной скоростью передачи информации (пропускной способностью), приходим к выводу, что при использовании такой схемы передачи информации пропускная способность линии связи будет использоваться не в полной мере. Избежать неэффективного использования линии связи можно, установив на ее входе и выходе буферные накопительные запоминающие устройства. Они позволяют сгладить неравномерность поступления букв вторичного алфавита как через линию связи, так и на вход декодирующего устройства. При этом во всей системе передачи информации возникают временные задержки и, чем объем буферного устройства выше, тем эти задержки выше.

Вторая особенность связана с временными задержками в передаче информации, возникающими при использовании кодирования блоков букв первичного алфавита, которые, как следует из теоремы 4, позволяют увеличить эффективность кода (уменьшить среднюю длину кодового слова). Кодирование блоков букв первичного алфавита требует их предварительного накопления. Отсюда и возникающие временные задержки.

Третья особенность заключается в том, что, как оказывается, эффективные коды не предназначены для использования в условиях помех. Если же все–таки по какой–либо причине (например, в результате электрической помехи от грозового разряда какой–то символ кодового слова исказится, правильное декодирование становится невозможным не только для этого кодового слова, но и для

целого ряда следующих за ним кодовых слов. Возникает так называемый *трек ошибки*).

Таким образом, улучшив одну из качественных характеристик кода, в данном случае среднюю длину кодового слова $n_{ср}$, ухудшается другая характеристика – устойчивость к действию помех. Эффективные коды надо использовать либо в условиях полного отсутствия помех, либо для устранения избыточности сообщения и подготовки к последующему помехоустойчивому кодированию.

Если $n_{ср} = \dfrac{H}{\log_2 m}$, то $H = n_{ср} log_2 m$. В этом случае энтропия, приходящаяся на одну букву вторичного алфавита $H_{1букву} = log_2 m$, совпадает с количеством информации по Хартли. Следовательно, буквы вторичного алфавита эффективного кода равновероятны и статистически независимы.

4.3. Помехоустойчивое кодирование

Как следует из названия, такое кодирование предназначено для устранения вредного влияния помех в каналах передачи информации. Уже сообщалось, что такая передача возможна как в пространстве, так и во времени, например, при хранении информации в виде документов в архивах или в запоминающих устройствах компьютеров.

Теория помехоустойчивого кодирования базируется на результатах исследований, проведенных К. Шенноном и сформулированных им в виде основной теоремы кодирования для дискретного канала с шумом: *«При любой скорости передачи символов вторичного алфавита меньшей, чем пропускная способность канала, существует такой код, при котором вероятность ошибочного декодирования будет сколь-угодно мала; вероятность ошибки не может быть сделана сколь-угодно малой, если скорость передачи больше пропускной способности канала»*.

Здесь под *скоростью передачи информации* понимается количество информации, передаваемой через канал связи за единицу времени, а под *пропускной способностью* – максимальная скорость

передачи информации через канал связи (не надо путать со скоростью передачи символов вторичного алфавита).

Подробнее эти вопросы будут рассмотрены ниже в главе «*Передача информации*».

Сейчас же при обсуждении методов помехоустойчивого кодирования важен следующий вывод из основной теоремы Шеннона: хотя теорема не указывает способ построения помехоустойчивых кодов, она утверждает, что *коды, способные полностью нейтрализовать вредное влияние помех при соответствующей скорости передачи информации, существуют*. Это мобилизовало усилия ученых на их разработку.

При рассмотрении процесса помехоустойчивого кодирования обычно исходят из того, что первичный и вторичный алфавиты совпадают. Предполагают, что они двоичные. Кроме того, предполагается равновероятность и статистическая независимость букв первичного алфавита. Это действительно так, если предварительно выполнить кодирование сообщения с помощью эффективного (оптимального) кода. Таким образом, предполагается, что энтропия букв первичного алфавита максимальна, а избыточность источника равна нулю.

В настоящее время существует большое количество разнообразных кодов. Выполнена их подробная классификация. Рассмотрим наиболее важные из классов кодов.

4.3.1. Классификация кодов

1. В зависимости от способности обнаруживать и исправлять ошибки различают коды, *обнаруживающие* ошибки, и коды, *исправляющие* ошибки. Первые позволяют только обнаруживать, но не исправлять ошибки. Вторые – не только обнаруживать, но и исправлять большую или меньшую долю обнаруженных ошибок.

2. По форме внесения избыточности различают *разделимые* и *неразделимые* коды.

В разделимых кодах кодовые слова четко разделены на две части. Одна часть состоит из исходных (кодируемых) разрядов, а другая из дополнительных разрядов, благодаря которым собствен-

но и обеспечивается возможность обнаружения и/или исправления ошибок.

Ясно, что эти дополнительные разряды создают в сообщении избыточность (вспомним, что предполагается нулевая избыточность исходного (кодируемого) сообщения).

В кодовых словах неразделимого кода такое разделение отсутствует.

3. В зависимости от способа образования дополнительных разрядов разделимые коды подразделяются на *систематические* и *несистематические*. В систематических кодах дополнительные разряды символов образуются с помощью различных линейных операций над информационными. Систематические коды – самая обширная и наиболее применяемая группа корректирующих кодов.

4. По способу кодирования различают *блочные* и *непрерывные* коды. При блочном кодировании входное кодируемое сообщение разбивается на блоки фиксированной длины k. Этим блокам ставятся в соответствие кодовые слова длиной n ($n > k$).

При непрерывном кодировании каждый символ b выходной последовательности (закодированного сообщения) определяется рекуррентным соотношением, связывающим его с соответствующими символами a входной последовательности (кодируемого сообщения), например:

$$b_{ij} = \mathrm{mod}_2\left(\sum_{k=0}^{s} c_{kj} a_{i-k}\right), \quad 1 \le j \le L,$$

где a – символы входной последовательности (кодируемого сообщения);

b – символы выходной последовательности (закодированного сообщения);

c_{kj} и s – константы, зависящие от кода;

i – номер символа входной последовательности (кодируемого сообщения);

j – номер символа в группе символов выходной последовательности (закодированного сообщения), создаваемой в результате поступления нового символа входной последовательности.

L – количество символов выходной последовательности, создаваемых в результате кодирования в ответ на поступления на вход кодера одного символа входной последовательности.

На рисунке 4.4 иллюстрируется кодирование непрерывным кодом.

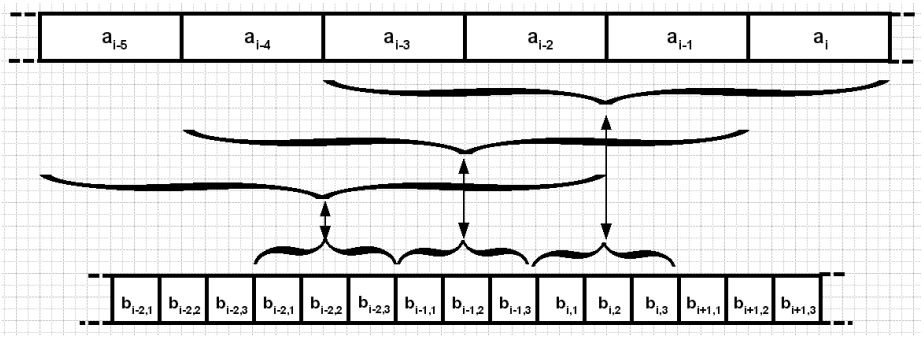


Рис. 4.5. Образование непрерывного кода

Таким образом, на каждый символ входной последовательности приходится L символов выходной.

Существуют и другие варианты формулы кодирования.

5. По способу описания выделяют *алгебраические* коды, т. е. коды, описываемые с помощью аппарата высшей алгебры.

6. В зависимости от длины кодового слова различают *равномерные* и *неравномерные* коды. Если длина кодовых слов постоянна, код *равномерный*, в противном случае – *неравномерный*.

Рассмотрим блочный равномерный код. Если произвольная линейная комбинация кодовых слов – также кодовое слово, такой код называется *линейным*.

Одни и тот же код может одновременно принадлежать нескольким классам, например, может существовать алгебраический разделимый неравномерный код.

4.3.2. Простейшие модели цифровых каналов связи с помехами

Свойство помехоустойчивых кодов обнаруживать и исправлять ошибки в сильной степени зависит от характеристик помех и канала передачи информации. В теории информации обычно рас-

сматривают две простые модели дискретного (цифрового) канала. Ими являются:

1) двоичный симметричный канал (ДСМК);
2) двоичный стирающий канал (ДСТК).

При передаче информации через ДСМК помеха способна с некоторой вероятностью P превратить 0 в 1 или 1 в 0. Передачу информации через ДСМК можно представить в виде графа, изображенного на рисунке 4.5.

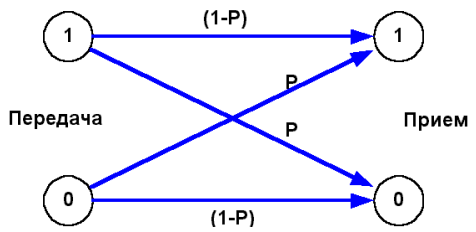


Рис. 4.5. Двоичный симметричный канал

Двоичный стирающий канал также описывается вероятностью P искажения передаваемого символа, однако в результате искажения передаваемый символ стирается, что отражается следующим графом, изображенном на рисунке 4.6.

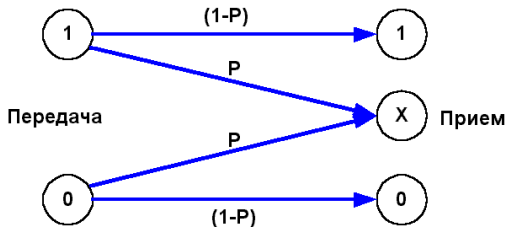


Рис. 4.6. Двоичный стирающий канал: X – нейтральный символ.

В дальнейшем воспользуемся моделью ДСМК. Его преимущество заключается в исключительной простоте.

4.3.3. Расчет вероятности искажения кодового слова в ДСМК

Положим, кодовое слово состоит из n двоичных символов. Вероятность неискажения кодового слова, как несложно доказать, равна:

$$p_0 = \underbrace{(1-p)\ldots(1-p)}_{n} = (1-p)^n .$$

Вероятность искажения одного символа (однократная ошибка):

$$p_1 = p\underbrace{(1-p)\ldots(1-p)}_{n-1} = p(1-p)^{n-1},$$

и любого из s символов (s – кратность ошибки):

$$p_s = p^s(1-p)^{n-s} \quad .$$

Нетрудно заметить, что если $P < 0{,}5$, то с ростом кратности ошибки s ее вероятность P_s уменьшается. Это следует из того, что P_s всегда представляет собой произведение n сомножителей, s из которых $P < 0{,}5$ а $n - s\cdot(1 - P) > 0{,}5$. С увеличением кратности ошибки s число больших сомножителей $(1-P)$ уменьшается, а меньших P – увеличивается. Это приводит к тому, что произведение P_s, т. е. вероятность ошибки, с ростом ее кратности в целом уменьшается.

Таким образом, в ДСМК наиболее вероятны ошибки с меньшей кратностью. Следовательно, такие ошибки следует исправлять в первую очередь.

4.3.4. Общие принципы использования избыточности

Для простоты рассмотрим блоковый код. С его помощью каждым k разрядам (буквам) входной последовательности ставится в соответствие n–разрядное кодовое слово. Количество разного вида k–разрядных информационных комбинаций равно 2^k. Каждой из них ставится в соответствие определенное кодовое слово. Их также 2^k. Эти кодовые слова называются *разрешенными*.

Однако возможное число n–разрядных комбинаций равно 2^n. Если из этого числа исключить все разрешенные кодовые слова, то останутся *запрещенные* кодовые слова.

Кодер посылает в линию связи только разрешенные кодовые слова, однако под действием помех они могут превратиться в любое другое из 2^n возможных − в другое разрешенное или в запрещенное. Все запрещенные слова данного кода заранее известны. Поэтому факт получения такого кодового слова легко устанавливается и является верным признаком наличия в кодовом слове ошибки – ошибка обнаруживается.

Возможность исправления ошибки связана со свойством кода, называемого *кодовым расстоянием*.

Определить *кодовое расстояние* невозможно без предварительного определения другого понятия – *расстояние между кодовыми словами*.

Под *расстоянием между кодовыми словами* понимают количество различий в одноименных разрядах двух кодовых слов.

Если, например, одно кодовое слово имеет вид 011101001, а другое – 111001100, то расстояние между ними можно вычислить, выполнив операцию XOR (исключенное или) и подсчитав количество единичных разрядов в результате:

0	1	1	1	0	1	0	0	1
XOR								
1	1	1	0	0	1	1	0	0
1	0	0	1	0	0	1	0	1

Расстояние между кодовыми словами, как видим, равно 4. Ниже будем расстояние между кодовыми словами обозначать буквой d. В данном примере $d = 4$.

Вспомним, что весом кодового слова двоичного кода называют число единичных разрядов в нем. Тогда кодовое расстояние равно весу поразрядной суммы по модулю 2 кодовых слов, расстояние между которыми определяется.

Выше мы показали, что наиболее вероятными в случае ДСМК являются однократные ошибки, и, чем выше кратность

ошибки, тем менее она вероятна. Возникновение s–кратной ошибки можно трактовать как преобразование передаваемого кодового слова в другое, находящееся на расстоянии s от передаваемого. Ясно, что переводы на меньшие расстояния более вероятны.

Геометрически множество разрешенных кодовых слов условно можно отобразить в виде точечного множества (рис. 4.7).

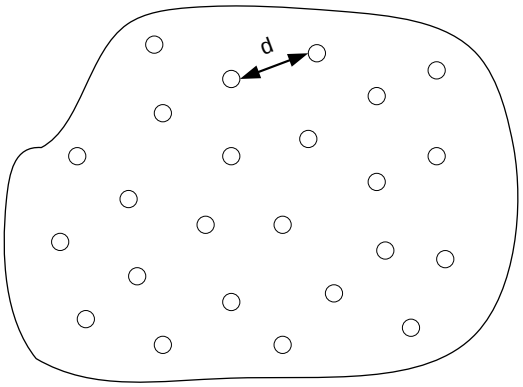


Рис. 4.7. Точечное множество

Теперь можно определить понятие кодового расстояния.

Кодовым расстоянием данного кода называют наименьшее расстояние между разрешенными кодовыми словами этого кода. Ниже оно обозначается $d_{\min}$.

В литературе по кодированию кодовое расстояние еще называют *Хэминговым* – по фамилии ученого, впервые его предложившего.

Продолжим изучение общих принципов использования избыточности.

Напомним, что вероятность определенной 2–кратной ошибки определяется формулой $p_2 = p^2(1-p)^{n-2}$. Но таких ошибок может быть несколько. Если, к примеру, кодовое слово состоит из 5 символов, то возможны ошибки в 1 и 5, 1 и 4, 1 и 3, 1 и 2, 2 и 5, 2 и 4, 2 и 3, 3 и 5, 3 и 4, 4 и 5 разрядах. Итого 10 различных 2–кратных ошибок. Их число равно числу сочетаний из 5 элементов по 2.

Если говорить не о конкретных, а о произвольных s–кратных ошибках, то вероятность $\overline{p_s}$ их возникновения находится как сумма всех возможных i–кратных конкретных ошибок:

$$\overline{p_s} = C_n^s p^s (1-p)^{n-s},$$

где $C_n^s = \dfrac{n!}{s!(n-s)!}$ – число сочетаний из *n* элементов по *ко*.

Следует заметить, что $\overline{p_s}$ в отличие от P_s не обязательно монотонно уменьшается с ростом *n*.

При получении искаженного кодового слова мы имеем дело с конкретной *s*–кратной ошибкой, вероятность появления которой – P_s. Эта вероятность монотонно уменьшается с ростом кратности ошибки *s*. Поэтому при обнаружении на выходе канала запрещенного кодового слова наиболее вероятным является то, что было передано разрешенное кодовое слово, отличающееся от полученного в наименьшем числе разрядов, т. е. находящееся от него на ближайшем расстоянии.

Если код имеет кодовое расстояние d_{min}, то гарантией обнаружения ошибки будет $s < d_{min}$. В этом случае помехи не смогут преобразовать одно разрешенное кодовое слово в другое разрешенное, поскольку расстояние до ближайшего разрешенного слова равно $d_{\min}$.

Если кратность ошибки превысит кодовое расстояние, она также может быть обнаружена, но не всегда.

Гарантией же правильного исправления ошибки будет условие $s < d_{min}/2$, т. к. согласно свойству вероятности ошибки наиболее вероятны ошибки с меньшей кратностью и, следовательно, исправление будет производиться в сторону ближайшего разрешенного кодового слова (рис. 4.9).

Для того, чтобы это исправление было верным, необходимо выполнение условия $s<d_{\min}/2$.

В связи с этим введено понятие *корректирующей способности кода*.

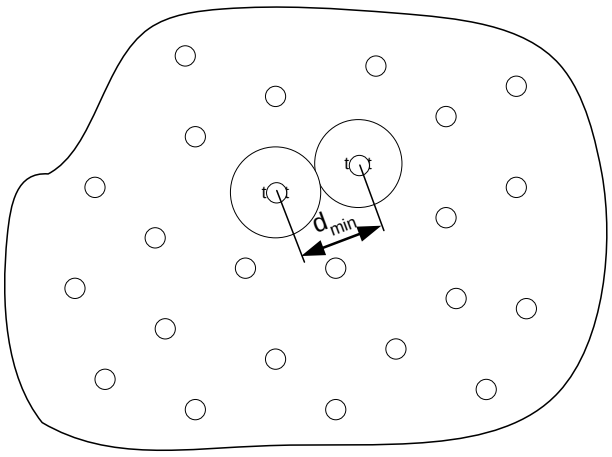


Рис. 4.9. Кодовое расстояние

Под *корректирующей способностью кода* понимается максимальная кратность исправляемой с применением этого кода ошибки.

Учитывая вышеприведенные рассуждения, корректирующую способность кода КСК можно было бы определить по формуле

$$КСК = \mathrm{int}\left(\frac{d_{\min}}{2}\right).$$

На практике же, учитывая увеличение вероятности ошибочного исправления ошибки с ростом ее кратности, реальную корректирующую способность уменьшают:

$$КСК < \mathrm{int}\left(\frac{d_{\min}}{2}\right).$$

В наших рассуждениях мы предполагали использование ДСМК. Если же реальный канал этой модели не соответствует, способы обнаружения и коррекции ошибок видоизменяются.

В общем же процедура обнаружения–исправления ошибок опирается на разбиение всего множества 2^n возможных кодовых комбинаций на 2^k подмножеств, каждое из которых состоит из одного разрешенного кодового слова и соответствующих ему запре-

щенных. Эти подмножества не пересекаются между собой, а их объединение не обязательно равно исходному множеству.

При приеме определяется, к какому подмножеству относится принятое кодовое слово и, если оно (кодовое слово) не является разрешенным, оно заменяется закрепленным за этим подмножеством разрешенным кодовым словом.

Если же принятое кодовое слово не принадлежит ни одному из подмножеств, фиксируется неисправимое искажение, за чем могут следовать какие–либо дополнительные меры, например, запрос на повторную передачу искаженной части сообщения.

4.3.5. Граница Хэмминга

Граница Хэмминга Q определяет максимально возможное количество разрешенных кодовых слов равномерного кода при заданных длине n кодового слова и корректирующей способности кода КСК.

Найдем связь между максимально возможной корректирующей способностью КСК, длиной n кодового слова и максимальным числом Q разрешенных кодовых слов в случае использования ДСК.

В круг (шар) диаметром КСК поместится $\sum_{i=1}^{\text{КСК}} C_u^i + 1 = \sum_{i=0}^{КСК} C_u^i$ кодовых слов.

Поскольку подмножества не пересекаются, то

$$Q = \text{int}\left(\frac{2^n}{\sum_{i=0}^{КСК} C_n^i} \right).$$

Коды, число кодовых слов которых равно Q, называются *оптимальными* или *плотноупакованными*.

В геометрической интерпретации это означает, что круги на плоскости или шары в объеме соприкасаются, не пересекаясь друг с другом, и любая кодовая комбинация обязательно входит в какой–либо круг или шар. Избыточность при этом используется мак-

симально. Однако, если учесть сложность аппаратной или программной реализации, использование оптимальных кодов не всегда выгодно.

4.3.6. Избыточность помехоустойчивых кодов

Одной из характеристик кода является его избыточность. Увеличение избыточности в принципе нежелательно, т. к. увеличивает объемы хранимых и передаваемых данных, однако для борьбы с искажениями избыточность вводят искусственно. Важно только, чтобы она использовалась эффективно.

Различают *абсолютную* и *относительную избыточность*.

Абсолютная избыточность $И_{абс}$ равномерных помехоустойчивых кодов находится по формуле:

$$И_{абс} = n - k$$

Относительная избыточность:

$$И_{отн} = И_{абс} / n = 1 - k / n$$

При определенных условиях избыточность помехоустойчивого кода связана с избыточностью сообщения.

Если предположить равновероятность появления разрешенных кодовых слов, то избыточность D сообщения в виде кодового слова, вычисляемая по формуле $D = H_{max} - H$, будет равна абсолютной избыточности кода $D = n - k = И_{абс}$.

Это легко доказать, учитывая, что

$$H_{\max} = -\sum_{i=1}^{2^n} \frac{1}{2^n} \log_2 \frac{1}{2^n} = \log_2 2^n = n \text{ и } H = -\sum_{i=1}^{2^k} \frac{1}{2^k} \log_2 \frac{1}{2^k} = \log_2 2^k = k.$$

4.3.7. Математическое введение к алгебраическим кодам

В настоящее время наиболее изучены, многочисленны, исследованы и широко используются корректирующие коды, описы-

ваемые при помощи аппарата высшей алгебры. Поэтому прежде, чем приступить к рассмотрению конкретных корректирующих кодов, ознакомимся с используемой при этом частью высшей алгебры.

Предметом высшей алгебры являются так называемые *алгебраические системы*, являющиеся совокупностью некоторых множеств и определенных на элементах этих множеств операций.

Под *двухместной операцией* здесь понимается однозначное сопоставление двум элементам множества с третьим по определенным правилам.

Под *одноместной операцией* понимается однозначное сопоставление одному элементу множества с другим по определенным правилам.

Обычно основные операции называются сложением $(a+b=c)$ и умножением $(ab=d)$, а обратные им – вычитанием и делением, даже если эти операции производятся не над числами (например, над оттенками цветов или геометрическими фигурами) и не идентичны соответствующим арифметическим операциям. С точки зрения высшей алгебры элементарная школьная алгебра − одна из бесконечного множества других возможных алгебраических систем.

Свойства и наименования систем зависят от законов, которые в них определены.

Пусть S – множество некоторых элементов, а a и b – элементы, принадлежащие множеству S (последнее записывается в виде $a \in S,\ b \in S$).

В таблице 4.8 определены некоторые важные для нас законы.

Система, в которой определены две основные операции (сложение и умножение) и которая соответствует всем отмеченным в таблице 4.8 свойствам, называется *полем*.

Система, в которой определены обе основные операции и выполняются законы A1−A5, M1, M2, D1 и D2, называется *кольцом*.

Таблица 4.8

Основные законы композиции

Законы композиции	Операция			
	Сложение		Умножение	
	обозн.	формула	обозн.	формула
Замкнутость: если $a \in S, b \in S$, то $c \in S$	A1	$a+b=c$	M1	$ab=c$
Ассоциативность	A2	$(a+b)+c=$ $=a+(b+c)$	M2	$(ab)c=a(bc)$
Коммутативность	A3	$a+b=b+a$	M3	$ab=ba$
Наличие нулевого для операции сложения и единичного для операции умножения элемента *e*.	A4	$a+e=$ $=e+a=a$	M4	$ae=ea=$ $=a$
Наличие для $a \in S$ обратного элемента $\bar{a} \in S$.	A5	$a+\bar{a}=\bar{a}+a=0$ $\bar{a}=-a$	M5	$a\bar{a}=\bar{a}a=1$ $\bar{a}=a^{-1}$
Дистрибутивность	D1	$a(b=c) \ =ab=ac$		
	D2	$(b+c)a \ = \ ba+bc$		

Если дополнительно в некотором кольце выполняется закон M3, то это кольцо – *коммутативное*.

Система, в которой задана лишь одна операция (сложение или умножение) и выполняются законы A1, A2, A4, A5 или M1, M2, M4, M5 называется *группой*.

Если в группе определена операция сложения, то такая группа называется *аддитивной*, если умножения, то *мультипликативной*.

Группа называется *коммутативной* или *абелевой*, если в ней выполняется закон коммутативности (A3 или M3) в зависимости от введенной операции.

Группы, имеющие конечное число элементов, называются *конечными*.

Построим алгебраическую систему, удобную для описания кодов равномерных двоичных кодов.

Предположим, что кодовые комбинации образуют множество S. Определим теперь операцию суммирования, обладающую свойствами А1, А2, А4, А5 и образуем таким образом группу. Операцию суммирования определим таким образом, чтобы удовлетворялось условие замкнутости А1. Следовательно, число разрядов при выполнении этой операции меняться не должно. Этому условию отвечает операция поразрядного сложения кодовых слов по модулю 2 (обозначается знаком $\oplus$). Результатом такого сложения будет 1, если число единиц в разряде слагаемых четно, и 0 − если нечетно.

Пример:

$$\begin{array}{r} 1\,0\,1\,1\,1\,0\,1 \\ \oplus\,0\,1\,1\,1\,1\,0\,1 \\ \underline{0\,0\,0\,1\,1\,1\,0} \\ 1\,1\,0\,1\,1\,1\,0 \end{array}$$

Выбранная операция коммутативна, поэтому группа будет абелевой.

Нулевым элементом является комбинация, состоящая из одних нулей (это легко проверить).

Противоположным элементом является сама кодовая комбинация, а обратная операция – вычитание – тождественна основной операции суммирования.

4.3.8. Линейные коды

Рассмотрим класс алгебраических кодов, называемых *линейными*.

Линейными называют блоковые коды, дополнительные разряды которых образуются путем линейных операций над определенными информационными разрядами.

Здесь используется понятие *линейная операция*.

Линейной операцией называется операция над объектами, описываемая выражением:

$$C_0 + C_1X_1 + C_2X_2 + \ldots,$$

где C_i – константы, X_i – объекты.

Слово *линейный* здесь используется в виду того, что вышеприведенное выражение описывает многомерную прямую линию. Следует только иметь в виду, что операции умножения и сложения, указанные в этом выражении, могут отличаться от привычных арифметических. В теории кодирования в качестве операции сложения используется сложение по модулю 2 ($\oplus$).

4.3.9. Упрощённый способ построения линейного кода

Упрощенный способ построения линейного кода имеет простой алгоритм, заключающийся в следующем.

1. Определение числа добавочных разрядов m.

Для определения числа добавочных разрядов можно воспользоваться формулой границы Хэмминга:

$$Q = \mathrm{int}\left(\frac{2^n}{\sum_{i=0}^{КСК} C_n^i} \right) \tag{4.5}$$

При этом можно получить плотноупакованный код, т. е. код с минимальной при заданных параметрах кода длиной *n* кодового слова и, следовательно, избыточностью.

К задаваемым параметрам кода относятся: длина информационной последовательности *k* и корректирующая способность кода КСК.

При *k* информационных двоичных разрядах может передаваться 2^k кодовых слов. Если приравнять $Q = 2^k$, то с учетом границы Хэмминга получаем:

$$2^k \le \frac{2^{k+m}}{\sum_{i=0}^{КСК} C_n^i}.$$

Если КСК = 1, т. е. строится код, исправляющий максимум однократные ошибки, то

$$\sum_{i=0}^{1} C_n^i = 1 + n = 1 + m + k,$$

т. к.

$$C_n^i = \frac{n!}{i!(n-i)!} \text{ и } 2^k \le \frac{2^{k+m}}{1+m+k};\ 2^m > m + k.$$

С учетом последней формулы ищется наименьшее *m*, при котором удовлетворяется это неравенство. Наименьшее потому, что при этом достигается наименьшая избыточность.

Пример: $k = 7$, тогда путем простого перебора легко найти, что $m = 4$.

2. *Построение образующей матрицы* $\|OM\|$

Линейные коды обладают следующим свойством: из всего множества 2^k разрешенных кодовых слов, образующих, кстати, группу, можно выделить подмножества из *k* слов, обладающих свойством линейной независимости.

Линейная независимость означает, что никакое из слов, входящих в подмножество линейно–независимых кодовых слов, нельзя получить путем суммирования (с помощью линейного выражения) любых других слов, входящих в это подмножество.

В то же время любое другое из разрешенных кодовых слов можно получить путем суммирования определенных линейно–независимых слов.

Таким образом, построение кодовых комбинаций линейного кода связано с линейными операциями. Для выполнения таких операций удобно пользоваться хорошо разработанным аппаратом матричных вычислений.

Для образования кодовых слов по информационной последовательности (кодирования) используют матрицу, называющуюся образующей.

Образующая матрица получается путем записи в столбец *k* линейно–независимых слов.

Обозначим кодируемую информационную последовательность X и будем записывать ее в виде матрицы–строки $\|X\|$ размерностью $1 \times k$, например: $\|X\| = \|11001\|$, где $k = 5$.

Один из способов построения образующей матрицы следующий. Она строится из единичной матрицы $\|I\|$ размерностью $k \cdot k$ и приписанной к ней справа матрицы добавочных (избыточных) разрядов $\|МДР\|$ размерности $k \cdot m$:

$$\|OM_{k\times n}\| = \|I_{k\times k} МДР_{k\times m}\|,$$

где при $k = 4$

$$\|I_{4\times 4}\| = \begin{Vmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{Vmatrix}.$$

Такая структура *ОМ* обеспечивает получение систематического кода.

Порядок построения матрицы МДР будет рассмотрен ниже.

3. *Порядок кодирования.* Кодовое слово КС получается путем умножения матрицы информационной последовательности $\|X\|$ на образующую матрицу $\|OM\|$:

$$\|KC_{1\times n}\| = \|X_{1\times k}\| \cdot \|OM_{k\times n}\|$$

Умножение выполняется по правилам матричного умножения (рис. 4.9):

$$a_{ij} = \sum_{k=1}^{h} b_{ik} c_{kj}.$$

Элементы i–й строки множимого $\|B\|$ умножаются на элементы j–го столбца множителя $\|C\|$ и складываются. В результате

такого сложения получается элемент i–й строки и j–го столбца a_{ij} результирующей матрицы $\|A\|$.

Надо только помнить, что сложение здесь ведется по модулю 2.

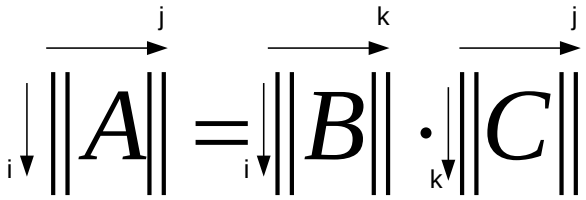


Рис. 4.9. Иллюстрация правила матричного умножения

Пример. Допустим, образующая матрица

$$\|OM\| = \begin{Vmatrix} 1000 & 100 \\ 0100 & 110 \\ 0010 & 011 \\ 0001 & 001 \end{Vmatrix}$$

и матрица информационной последовательности

$$\|X\| = \|1100\|.$$

Так как множимая матрица имеет всего одну строку, умножение упрощается. В этом случае следует поставить в соответствие строкам образующей матрицы $||OM||$ разряды матрицы информационной последовательности $||X||$ и сложить те строки образующей матрицы, которые соответствуют единичным разрядам матрицы $||X||$.

||X||							||OM||
1		1	0	0	0	1	1	0
1		0	1	0	0	1	1	1
0		0	0	1	0	0	1	1
0		0	0	0	1	1	0	1
||КС||	=	**1**	**1**	**0**	**0**	**0**	**0**	**1**

Заметим, что $\|KC\| = \|XДР\|$, где $\|X\|$ – информационная последовательность (т. к. умножается на единичную матрицу ||*I*||), а ||ДР|| – добавочные разряды, зависящие от матрицы добавочных разрядов ||МДР||*:*

$$\|ДР\| = \|X\| \cdot \|МДР\|$$

4.3.10. Порядок декодирования

В результате передачи кодового слова через канал оно может быть искажено помехой. Это приведет к тому, что принятое кодовое слово ||ПКС|| может не совпасть с исходным ||КС||.

Искажение можно описать с помощью следующей формулы:

$$\|ПКС\| = \|KC\| + \|BO\|,$$

где ||ВО|| − *вектор ошибки* – матрица–строка размерностью $1 \times n$, с единицей в тех разрядах, в которых произошли искажения.

Декодирование основано на нахождении так называемого опознавателя – матрицы–строки $\|ОП_{1\times m}\|$ длиной *m* разрядов (*m* – количество добавочных или избыточных разрядов в кодовом слове). Опознаватель используется для нахождения предполагаемого вектора ошибки.

Опознаватель находят по следующей формуле

$$\|ОП\| = \|ПКС\| \cdot \|ТПМ\|,$$

где ||ПКС|| – принятое и возможно искаженное кодовое слово;

||ТПМ|| – *транспонированная проверочная матрица*, которая получается из матрицы добавочных разрядов ||МДР|| путем приписывания к ней снизу единичной матрицы:

$$\|ТПМ\| = \left\| \begin{matrix} МДР_{k \cdot m} \\ I_{m \cdot m} \end{matrix} \right\|$$

Приведем пример транспонированной матрицы ||ТПМ||.

$$\left\| \begin{matrix} 1 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{matrix} \right\|$$

Поскольку

||ПКС|| = ||КС|| + ||ВО||,

последнюю формулу можно записать в виде:

||ОП|| = ||КС|| × ||ТПМ|| + ||ВО|| × ||ТПМ||

Рассмотрим первое слагаемое: ||КС|| − матрица–строка, причем первые *k* разрядов – информационные.

Докажем теперь, что произведение кодового слова ||КС|| на ||ТПМ|| приводит к получению нулевой матрицы ||0||.

Поскольку ||КС|| − матрица–строка, возможен упрощенный порядок умножения матриц, рассмотренный выше:

$$\|КС\| \cdot \|ТПМ\| = \left\| X \quad ДР \right\| \cdot \left\| \begin{matrix} МДР \\ I \end{matrix} \right\| = \|X\| \cdot \|МДР\| \oplus \|ДР\| \cdot \|I\| = \|ДР\| \oplus \|ДР\| = \|0\|.$$

Следовательно, первое слагаемое в

$$||ОП|| = ||КС|| \times ||ТПМ|| + ||ВО|| \times ||ТПМ||$$

всегда равно нулю и опознаватель полностью зависит от вектора ошибки ||ВО||.

Если теперь подобрать такую матрицу ТПМ, а значит и МДР, чтобы разным векторам ошибки соответствовали разные опознаватели ОП, то по этим опознавателям можно будет находить вектор ошибки ВО, а значит и исправлять эти ошибки.

Соответствие опознавателей векторам ошибки находится заранее путем перемножения векторов исправляемых ошибок на ТПМ:

$$||ОП|| = ||ВО|| \times ||ТПМ||$$

Таким образом, способность кода исправлять ошибки целиком определяется МДР.

Найдем способ построения МДР для кодов, исправляющих однократные ошибки.

Число столбцов m определяется, как показано выше, в виде минимального числа, удовлетворяющего неравенству $2^m > m + k$.

Так как ВО для случая исправления одной ошибки может иметь лишь одну единицу, таких ВО может быть только n (1 в различных разрядах). Учитывая возможность использования в данном случае вышеописанного упрощенного способа перемножения матриц, видим, что опознаватели в этом случае просто совпадают со строками ТПР, причем номер строки равен номеру искаженного разряда.

Для того чтобы ни одна строка МДР не совпадала со строкой, приписанной к ней ниже единичной матрицы, число единичных разрядов в МДР должно быть больше одного. Кроме того, чтобы строки не были одинаковыми (а они совпадают с опознавателями; опознаватели различных ошибок должны быть различны) они должны отличаться не менее, чем в одном разряде.

Эти правила можно получить и другим способом.

Кодовое расстояние для кода, исправляющего до КСК ошибок, вычисляется по формуле $КСК = int[(d_{min} - 1)/2]$. Несложно найти, что $d_{min} = 3$ при $КСК = 1$.

Для линейных кодов справедливо утверждение о том, что *кодовое расстояние линейного кода равно минимальному весу кодового слова этого кода.* Для доказательства вспомним определение линейного кода, согласно которому линейными называются коды любые линейные комбинации кодовых слов которых также являются кодовыми словами. Для определения расстояния между кодовыми словами нужно найти вес их суммы. Но эта сумма сама является кодовым словом. Кодовое же расстояние равно минимальному расстоянию между кодовыми словами. Это тоже кодовое слово.

При кодировании мы суммируем определенное количество (обозначим это количество буквой *i*) строк образующей матрицы ОМ. За счет единичной матрицы в полученном таким образом кодовом слове будет как минимум *i* единиц.

Если $i = 1$ (*i* не может быть меньше 1), то кодовое слово совпадает с одной из строк ОМ, а поскольку минимальный вес кодового слова в нашем случае равен 3, за счет МДР должно быть добавлено еще две единицы. Отсюда требование наличия *не менее чем двух единиц в каждой строке* МДР.

Если *i*=2, 2 единицы получаем за счет единичной матрицы и, чтобы получить еще одну единицу, нужно иметь *хотя бы одно различие между двумя любыми строчками* МДР.

Полученный нами код неудобен тем, что опознаватель, хотя и связан однозначно с номером искаженного разряда, как число не равен ему. Для поиска искаженного разряда нужно использовать дополнительную таблицу соответствия между опознавателем и этим номером. Коды, в которых такое соответствие присутствует, были найдены и получили название *кодов Хэмминга*.

Построение МДР для случая исправления многократных ошибок значительно усложняется. Разными авторами были найдены различные алгоритмы построения МДР для этого случая, а соответствующие коды называются их именами.

Несколько слов об обозначении кодов.

Равномерные блоковые систематические коды, имеющие n–разрядные кодовые слова с k информационными разрядами называют (n,k)–кодами.

В качестве примера рассмотрен код (8,4).

При использовании линейных кодов можно избежать применения операций над матрицами. Рассматривая процесс кодирования, нетрудно заметить, что определенные добавочные разряды получаются путем сложения по модулю 2 определенных информационных разрядов.

$$\|OM\| = \begin{Vmatrix} 1 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 0 & 1 \end{Vmatrix},$$

$$\|TПМ\| = \begin{Vmatrix} 1 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{Vmatrix}.$$

Обозначим кодовое слово $X_1X_2...X_kY_1Y_2...Y_m$, здесь X – информационные разряды, Y – добавочные разряды.

В нашем примере

$$k = 4,\ M = 3;$$
$$Y_1 = X_4, Y_2 = X_1 + X_2 + X_4;$$
$$Y_3 = X_1 + X_2 + X_3;$$
$$Y_4 = X_2 + X_3 + X_4.$$

Для расчета разрядов Z опознавателя также можно использовать соответствующие формулы:

$$Z_1 = F1 + F2 + F4 + F5;$$
$$Z_3 = F2 + F3 + F4 + F7;$$
$$Z_2 = F1 + F2 + F3 + F6,$$

где *Fi* – разряды ТПМ.

Расчеты по формулам широко используются при построении кодирующих и декодирующих устройств.

4.3.11. Двоичные циклические коды

Вышеприведенная процедура построения линейного кода имеет ряд недостатков. Она неоднозначна (МДР можно задать различным образом) и неудобна в реализации в виде технических устройств. Этих недостатков лишены линейные корректирующие коды, принадлежащие к классу *циклических*.

Циклическими называют линейные (n,k)–коды, обладающие следующим свойством: для любого кодового слова

$$\|KC\| = \|a_1, a_2, ..., a_{n-1}\|$$

существует другое кодовое слово:

$$\|\overline{KC}\| = \|a_{n-1}, a_1, a_2, ..., a_{n-2}\|,$$

полученное циклическим сдвигом элементов исходного кодового слова ||КС|| вправо или влево, и которое также принадлежит этому коду.

Для описания циклических кодов используют полиномы с фиктивной переменной. Будем обозначать эту переменную буквой *X*.

Например, кодовое слово ||КС|| = ||011010||.

Его можно описать полиномом

$$A(X) = 0 \cdot X^5 + 1 \cdot X^4 + 1 \cdot X^3 + 0 \cdot X^2 + 1 \cdot X^1 + 0 \cdot X^0 .$$

Таким образом, разряды кодового слова в описывающем его полиноме используются в качестве коэффициентов при степенях фиктивной переменной X.

Наибольшая степень фиктивной переменной X в слагаемом с ненулевым коэффициентом называется степенью полинома. В вышеприведенном примере получился полином 4–й степени.

Теперь действия над кодовыми словами сводятся к действиям над полиномами. Вместо алгебры матриц здесь используется алгебра полиномов.

Рассмотрим алгебраические действия над полиномами, используемые в теории циклических кодов.

Суммирование полиномов разберем на примере

$$C(X) = A(X) + B(X).$$

Пусть ||A|| = ||011010||, ||B|| = ||110111||. Тогда

$$\begin{array}{rrcccccccccccc}
 & A(X) & = & 0 & + & X^4 & + & X^3 & + & 0 & + & X^1 & + & 0 \\
\oplus & B(X) & = & X^5 & + & X^4 & + & 0 & + & X^2 & + & X^1 & + & X^0 \\
\hline
 & C(X) & = & X^5 & + & 0 & + & X^3 & + & X^2 & + & 0 & + & X^0
\end{array}$$

Таким образом, при суммировании коэффициентов при X в одинаковой степени результат берется по модулю 2. При таком правиле вычитание эквивалентно суммированию.

Умножение выполняется как обычно, но с использованием суммирования по модулю 2.

Рассмотрим умножение на примере умножения полинома $\left(X^3 + X^1 + X^0\right)$ на полином $\left(X^1 + X^0\right)$.

$$\begin{array}{rrcccccccc}
 & & & X^3 & + & 0 & + & X^1 & + & X^0 \\
 & & \otimes & & & & & X^1 & + & X^0 \\
\hline
 & & & X^3 & + & 0 & + & X^1 & + & X^0 \\
\oplus & X^4 & + & 0 & + & X^2 & + & X^1 & & \\
\hline
 & X^4 & + & X^3 & + & X^2 & + & 0 & + & X^0
\end{array}$$

Операция обратная умножению – деление. Деление полиномов выполняется как обычно, за исключением того, что вычитание выполняется по модулю 2 (вспомним, что вычитание по модулю 2 эквивалентно сложению по модулю 2).

Пример деления полинома $(X^6+X^4+X^3)$ на полином $(X^3+X^2+X^0)$:

$$
\begin{array}{l|l}
X^6 + 0 + X^4 + X^3 + 0 + 0 + 0 & X^3 + X^2 + 0 + X^0 \\
\cline{2-2}
X^6 + X^5 + 0 + X^3 \qquad\qquad \text{Результат} = & X^3 + X^2 \\
\cline{1-1}
\qquad X^5 + X^4 + 0 + 0 & \\
\qquad X^5 + X^4 + 0 + X^2 & \\
\cline{1-1}
\qquad\quad \text{Остаток} = \quad 0 + X^2 & .
\end{array}
$$

Циклический сдвиг влево на одну позицию коэффициентов полинома степени n – 1 получается путем его умножения на X с последующим вычитанием из результата умножения полинома X^n-1, если его порядок больше либо равен n. Последнее эквивалентно вычислению остатка от деления результата умножения на X^n-1.

Проверим это на примере.

Пусть требуется выполнить циклический сдвиг влево на одну позицию коэффициентов полинома $C(X) = X^5+ X^3+ X^2+ X^0$.

В результате должен получиться полином

$$\overline{C}(X) = X^4 + X^3 + X^1 + X^0$$

Это легко доказывается:

$$C(X)\times X-\left(X^6+1\right)=\left(X^{6+}X^4+X^3+X^1\right)+\left(X^6+X^0\right) \;=X^4+X^3+X^1+X^0.$$

В основе циклического кода лежит образующий полином m–го порядка (напомним, что m – число дополнительных разрядов). Будем обозначать его $g_m(X)$.

Образование кодовых слов (кодирование) КС выполняется путем умножения информационного полинома (информационный полином – полином с коэффициентами, являющимися информацион-

ной последовательностью) $И_i(X)$ порядка $i<k$ на образующий полином:

$$КС_{m+i}(X) = g_m(X) \cdot И_i(X).$$

Принятое кодовое слово может отличаться от переданного искаженными разрядами – результатом действия искажающих передаваемую информацию помех.

$$ПКС(X) = КС(X) + ВО(X),$$

где ВО(X) – полином вектора ошибки, а суммирование, как обычно, ведется по модулю 2.

Декодирование, как и раньше начинается с нахождения опознавателя, в данном случае в виде полинома ОП(X). Этот полином вычисляется как остаток от деления полинома принятого кодового слова ПКС(X) на образующий полином $g_m(X)$:

$$ОП(X) = остаток\left(\frac{ПКС(X)}{g_m(X)}\right) = остаток\left(\frac{КС(X)}{g_m(X)}\right) \oplus остаток\left(\frac{ВО(X)}{g_m(X)}\right).$$

Первое слагаемое остатка не имеет, т. к. кодовое слово было образовано путем умножения полинома информационной последовательности на образующий полином. Следовательно, и в данном случае опознаватель полностью зависит от вектора ошибки.

$$ОП(X) = остаток\left(\frac{ВО(X)}{g_m(X)}\right).$$

Образующий полином выбирается таким, чтобы при данном m как можно большее число отношений $ВО(X)/g_m(X)$ давало различные остатки.

Такому требованию отвечают так называемые *неприводимые полиномы*, которые не делятся без остатка ни на один полином степени m и ниже, кроме как сам на себя и на 1.

Приведенная здесь процедура образования кодового слова неудобна тем, что такой код получается несистематическим, т.е. таким, в кодовых словах которого нельзя выделить информационные и дополнительные разряды.

Этот недостаток был устранен.

Способ кодирования, приводящий к получению систематического линейного циклического кода, состоит в приписывании к информационной последовательности *И* дополнительных разрядов ДР.

Эти дополнительные разряды предлагается находить по следующей формуле:

$$ДР(X) = остаток\left(\frac{И_i(X)\cdot X^m}{g_m(X)}\right)$$

Порядок полинома ДР(X) гарантировано меньше m (поскольку это остаток).

Приписывание дополнительных разрядов к информационной последовательности, используя алгебру полиномов, можно описать формулой

$$КС(X) = И(X)\cdot X^m + ДР(X)$$

Одним из свойств циклических линейных кодов является то, что результат деления любого разрешенного кодового слова КС на образующий полином g также является разрешенным кодовым словом.

Докажем, что получаемые по вышеприведенному алгоритму кодовые слова являются кодовыми словами циклического линейного кода. Для этого нужно убедиться в том, что произвольное разрешенное кодовое слово делится на образующий полином без остатка:

$$\frac{КС(X)}{g(X)} = \frac{И(X)\cdot X^m}{g(X)} \oplus \frac{ДР(X)}{g(X)}$$

Рассмотрим первое слагаемое:

$$\frac{И(X)\cdot X^m}{g(X)} = d(X) \oplus \frac{ДР(X)}{g(X)} \quad ,$$

где $d(X)$ – целая часть результата деления.

Подставим полученную сумму на место первого слагаемого:

$$\frac{КС(X)}{g(X)} = d(X) \oplus \frac{ДР(X)}{g(X)} \oplus \frac{ДР(X)}{g(X)}$$

Суммирование последних двух слагаемых дает нулевой результат (напомним, что суммирование выполняется по модулю 2).

Значит $\frac{КС(X)}{g(X)} = d(X)$ − целая часть деления. Остатка нет. Это означает, что описанный выше способ кодирования соответствует циклическому коду.

4.3.12. Некоторые свойства циклических кодов

Все свойства циклических кодов определяются образующим полиномом.

1. Циклический код, образующий полином которого содержит более одного слагаемого, обнаруживает все одиночные ошибки.

Строго доказывать это не будем. Покажем это на примере простейшего образующего полинома $g(x) = X + 1$. Вектор однократной ошибки в i–м разряде описывается полиномом $BO(X) = X_i$.

Найдем опознаватель $ОП(X) = остаток\left(\frac{BO(X)}{g_m(X)}\right)$:

$$
\begin{array}{l|l}
X^i + 0 + 0 + \ldots + 0 & X + 1 \\
\cline{2-2}
X^i + X^{i-1} & X^{i-1} + X^{i-1} + \ldots + 1 \\
\cline{1-1}
\quad X^{i-1} + 0 & \\
\quad X^{i-1} + X^{i-2} & \\
\cline{1-1}
\qquad X^{i-2} & \\
\qquad \ldots & \\
\qquad\quad \ldots & \\
\qquad\qquad X & \\
\qquad\qquad X + 1 & \\
\cline{1-1}
\qquad\qquad\quad 1 \text{ - остаток} &
\end{array}
$$

Таким образом, поскольку ВО имеет вес 1 (не равен нулю) ошибка обнаруживается.

2. Докажем теперь, что циклический код, образованный при помощи полинома $g(X) = X + 1$, позволяет обнаруживать не только одиночные но и любые ошибки нечетной кратности.

Доказательство

Рассмотрим первый вариант кодирования – кодирование по формуле

$$КС(X) = И(X) \cdot g(X)$$

В нашем случае $g(X) = X + 1$.

После подстановки получим:

$$КС(X) = И(X) \cdot (X + 1) = X \cdot И(X) + И(X)$$

Если И(X) содержит L единичных коэффициентов, то и X×И(X) тоже содержит столько же L единичных коэффициентов, но при степенях X на единицу больших. Обозначим L' количество единичных коэффициентов при одинаковых степенях X. Оба полинома следует сложить. При сложении члены полиномов с одинаковыми степенями и единичными коэффициентами в сумме дадут слагаемые с нулевыми коэффициентами (сложение по модулю 2). В

общем, в сложении будет участвовать $2L$ слагаемых с единичными коэффициентами, $2L'$ из которых попарно взаимно уничтожаются.

В суммарном полиноме останется в итоге

$$2L - 2L' = 2(L - L')$$

членов с единичными коэффициентами – четное число. Таким образом, при использовании образующего полинома $X+1$ получаемые в результате кодовые слова обязательно имеют четное число единичных разрядов. Любое искажение с нечетным числом ошибок преобразует разрешенное кодовое слово с четным числом разрядов в запрещенное – с нечетным. Такое кодовое слово, поскольку оно является запрещенным, сразу же будет обнаружено по наличию остатка от его деления на образующий полином.

Надо заметить, что коды, обнаруживающие четное количество ошибок, используются довольно часто.

Найдем теперь для случая применения образующего полинома $X+1$ упрощенный вариант процедуры кодирования.

Ранее была приведена следующая формула получения кодового слова (случай систематического кода):

$$КС_n(X) = И_k(X) \cdot X^m \oplus ДР_m(X), \ \ где \ \ ДР_m(X) = остаток\left(\frac{И_k(X) \cdot X^m}{g_m(X)}\right)$$

Остатком от деления любого полинома на $X+1$ может быть либо 0 (остатка нет), либо 1. Следовательно, $m=1$, т. е. 1 разряд.

Учитывая вывод о том, что при использовании образующего полинома $X+1$ получаемые в результате кодовые слова обязательно имеют четное число единичных разрядов, делаем вывод, что этот один разряд должен дополнять количество единиц в информационной последовательности до четного числа. В этом и заключается упрощенный способ кодирования при использовании разделимого циклического кода с образующим полиномом $X+1$.

3. Циклический код, образованный полиномом $g_m(X)$, обнаруживает все двукратные и однократные ошибки, если значность кода n меньше или равна степени L полинома X^L+1, где L –

наименьшее число, при котором X^L+1 делится на $g_m(X)$ без остатка.

Доказательство не приводим.

4.3.13. Построение кода с заданной корректирующей способностью

Существует несложная процедура построения кода с заданной корректирующей способностью. Она состоит в следующем.

1. По заданному размеру информационной составляющей кодового слова длиной k определяется количество m дополнительных (избыточных) разрядов, обеспечивающее обнаружение и исправление 1–кратных ошибок. Сделать это можно при помощи формулы границы Хэмминга. Длина кодовых слов получающегося кода будет равной $n=k+m$. Строится систематический (n,k) код.

2. Рассматривая кодовые слова построенного (n,k)–кода как новую информационную последовательность длиной $k_1=n$, повторяем п.1.

В результате получаем новый (n_1,k_1)–код, способный также способный исправлять однократные ошибки.

Повторяя t раз п.п. 1 и 2, можно построить составной код, исправляющий t–кратные ошибки.

Кодирование при помощи это составного кода после этого сводится к последовательному t–кратному кодированию, а декодирование – к соответствующему t–кратному декодированию при помощи t различных кодов, каждый из которых способен исправлять только однократные ошибки.

Такой код, однако, оказывается неоптимальным с точки зрения числа дополнительных разрядов. Разработаны другие способы получения кодов с заданной корректирующей способностью, которые используют минимальное необходимое количество дополнительных разрядов. К ним, в частности, относятся широко известные среди лиц, занимающихся помехоустойчивым кодированием, коды Боуза – Чоудхури.

4.3.14. Матричное описание циклических кодов

Циклические коды можно, как и любые линейные коды, описывать с помощью матриц.

Вспомним, что

$$KC(X) = g_m(X) \cdot И(X).$$

Вспомним также на примере порядок умножения полиномов:

$$\begin{array}{lcl}
g_m(X) & = & 1\,0\,1\,0\,1\,1 \\
\underline{И(X)} & = & \underline{\qquad\quad 1\,0\,1\,1} \\
\oplus\ 1\,0\,1\,0\,1\,1 & & \times 1 \\
\oplus\ 1\,0\,1\,0\,1\,1 & & \times 1 \\
\oplus\ 1\,0\,1\,0\,1\,1 & & \times 0 \\
\underline{\oplus\ 1\,0\,1\,0\,1\,1} & & \underline{\times 1} \\
1\,0\,0\,1\,0\,0\,1\,0\,1 & &
\end{array}$$

Это соответствует описанному выше упрощенному способу умножения матриц:

$$\|KC\| = \|X\| \cdot \|OM\| = \begin{Vmatrix} 1 \\ 0 \\ 1 \\ 1 \end{Vmatrix} \begin{matrix} \otimes \\ \otimes \\ \otimes \\ \otimes \end{matrix} \begin{Vmatrix} 101011000 \\ 010101100 \\ 001010110 \\ 000101011 \end{Vmatrix}$$

Видно, что, если в качестве строк образующей матрицы взять наборы сдвинутых вправо коэффициентов образующего полинома, вычисление кодовых слов при помощи матриц полностью эквивалентно вычислению кодовых слов при помощи полиномов.

Согласно способу кодирования, позволяющему получить систематический код, кодовое слово находится по формуле

$$KC_n(X) = И_k(X) \cdot X^m \oplus ДР_m(X), \ \ где \ \ ДР_m(X) = остаток\left(\frac{И_k(X) \cdot X^m}{g_m(X)}\right)$$

Образующая матрица ||ОМ|| в этом случае должна состоять из 2 частей – единичной матрицы ||I|| и матрицы дополнительных разрядов ||МДР||:

$$\|ОМ\| = \|IМДР\|$$

Результат кодирования при помощи матриц будет совпадать с результатом кодирования при помощи полиномов в том случае, если строки матрицы дополнительных разрядов образовать по следующей формуле:

$$i - я \ \ строка \ \ МДР = ост\left(\frac{X^i}{g_m(X)}\right)$$

Далее по МДР легко построить ТПМ, найти опознаватели и пользоваться описанными выше для линейных кодов процедурами кодирования–декодирования.

4.3.15. Выбор образующего полинома

Ясно, что полиномы кодовых слов $KC(X)$ должны делиться на образующий полином $g(X)$ без остатка. Циклические коды относятся к классу линейных. Это означает, что для этих кодов существует m линейно–независимых кодовых слов и с их помощью путем суммирования в разных комбинациях можно получить все остальные разрешенные кодовые слова данного кода. Для того чтобы все они делились на образующий полином без остатка, достаточно, чтобы без остатка на него делились только m линейно независимых кодовых слов. Эти линейно–независимые кодовые слова можно получить путем m–кратного циклического сдвига вправо или влево на 1 позицию любого из разрешенных кодовых слов.

Напомним, что сдвигу влево соответствует умножение кодового слова на X с вычитанием, если в результате умножения получается полином порядка больше либо равном n, из результата полинома X^n+1. В формульном виде вышесказанное может быть отображено следующим образом:

$$g_i(X) = g(X)X^i - B(X^n+1),$$

где коэффициент $B = 1$, если степень

$$g(X)\cdot X^i < n$$

и $B = 0$, если степень $g(X)\cdot X^i \geq n$

Частное $g_i(X)/g(X)$ не имеет остатка, если полином (X^n+1) без остатка делится на образующий полином $g(X)$ (очевидно, первое слагаемое делится на $g(X)$ без остатка).

Таким образом, *образующий полином* $g_m(X)$ *циклического линейного кода должен делить полином* $X^n + 1$ *без остатка*.

Это свойство обеспечивает отсутствие остатка при делении разрешенных кодовых слов на образующий полином. Таких слов 2^k. Всего же различных кодовых слов, которые могут быть получены на выходе канала связи 2^n. Среди них 2^m-1 запрещенных, т.е. полученных в результате искажения разрешенных кодовых слов помехой. Ошибка обнаруживается, если остаток от их деления на образующий полином не равен нулю. Образующий полином имеет порядок m. Максимально возможное количество ненулевых остатков равно $2m-1$. Такое количество остатков могут дать только так называемые *неприводимые полиномы, т. е. полиномы, которые не делятся без остатка ни на какие другие полиномы, кроме* 1 *и себя самого*.

Контрольные вопросы

1. Что понимают под кодированием сообщения?
2. Приведите примеры простейших кодовых сообщений.
3. Какие коды называются равномерными?

4. Что называется двоичным кодом?

5. Как строится код Шеннона – Фано?

6. Как определяется число элементарных сигналов, приходящихся на одну букву сообщения?

7. Сформулировать основную теорему о кодировании.

8. Что называется декодированием сообщения?

9. Что называется блочным кодированием?

10. Представьте пример реализации блочного кодирования при построении оптимального неравномерного кода.

11. Назовите назначение и цели эффективного кодирования.

12. Поясните за счет чего, обеспечивается сжатие информации при применении эффективного кодирования.

13. Чем определяется минимальная длина кодовой комбинации при применении эффективного кодирования?

14. Какие проблемы возникают при разделении неравномерных кодовых комбинаций?

15. Объяснить принцип построения кода Хаффмана.

16. Какой код является самым выгодным?

17. За счет чего при эффективном кодировании уменьшается средняя длина кодовой комбинации?

18. До какого предела может уменьшиться длина кодовой комбинации при эффективном кодировании?

19. При каком распределении букв первичного алфавита оптимальный неравномерный код оказывается самым эффективным?

20. Какие коды называются помехоустойчивыми?

21. Что называется избыточностью?

22. Как образуются корректирующие коды?

23.Объясните методику построения кода Хэмминга.

24. Назовите основные параметры кода Хэмминга?

25. Как определить общее число элементов кодовых комбинаций кодов Хэмминга?

26. Как определить число проверочных и информационных элементов кода Хэмминга?

27. Как выбираются номера проверочных позиций кода Хэмминга?

28. По какому закону рассчитывают номера контрольных символов?

29. Объясните, в чем заключается правило четности.

30. Как происходит переход из двоичной системы счисления в десятичную?

31. Объясните особенности кода Хэмминга.

Контрольные задания

1. Как можно закодировать четыре сообщения a, b, c, d, используя только два сигнала 0 и 1?

2. Сообщение состоит из последовательности букв A, B и C, вероятности которых не зависят от предыдущего сочетания букв и равны P(A) = 0,7, P(B) = 0,2, P(C) = 0,1. Провести кодирование по алгоритму Шенно–Фано отдельных букв и двухбуквенных сочетаний. Сравнить коды по их эффективности и избыточности.

3. Закодировать методом Шеннона–Фано сообщение «Мы все учились понемногу, чему–нибудь и как–нибудь».

4. Задан алфавит из трех символов с вероятностями 0,75; 0,1; 0,15. Произвести кодирование отдельных букв и двухбуквенных сочетаний по методу Хаффмана. Для полученных кодов найти средние длины и коэффициенты оптимальности.

5. Закодировать сообщение «Ученье – свет, а неученье – тьма!»

6. Сколько информационных символов содержится в коде, исправляющем одиночную ошибку при числе информационных комбинации N = 32?

7. Определить избыточность корректирующего кода при общем числе кодовых комбинаций N = 256.

8. Постройте матрицу для группового кода, способного исправлять одиночную ошибку при передаче 16 символов первичного алфавита.

9. Используя метод образующих матриц, постройте циклически код, исправляющий одинарные ошибки при передаче комбинации четырехзначного двоичного кода на все сочетания (кроме нулевой комбинации).

Глава 5. Передача информации

Использование информации связано с транспортировкой ее в пространстве и во времени от источника к потребителю. При этом стремятся:

1) передавать как можно меньше данных в единицу времени, для чего, в частности, используется эффективное кодирование (сжатие);

2) повышать надежность передачи, что решается за счет применения помехоустойчивого кодирования;

3) использовать многоканальную систему передачи, как, например, реализовано в виде общей шины в современных компьютерных системах;

4) использовать новые перспективные линии связи, например оптические или микроволновые.

Итак, информация передается через каналы связи. Дадим определение понятию *канал передачи информации*.

Канал передачи информации − совокупность технических средств, предназначенных для передачи информации от источника к приемнику.

Дадим также определение понятию *линия связи*.

Линия связи – среда, в которой распространяются сигналы, несущие информацию.

В многоканальных системах передачи информации может использоваться одна или несколько линий связи.

5.1. Виды каналов передачи информации

Рассмотрим каналы, отличающиеся по типу используемых в них линий связи.

1. *Механические каналы*, в которых для передачи информации используется перемещение каких–либо твердых, жидких или газообразных тел. В первом случае могут использоваться рычаги или тросы (например, органы управления автомобилем), во втором – гидравлические системы (например, тормозная система автомобиля), в третьем – разного рода пневматические устройства (широко используются, например, в газовой промышленности).

2. *Акустические каналы*. Используют механические колебания звуковой и ультразвуковой частоты, особенно хорошо распространяющиеся в жидких средах. Широко применяются, например, для передачи информации людям или устройствам, находящимся под водой или в другой жидкой среде, а также при проведении медицинских исследований (УЗИ). Акустический канал в газовой среде – едва ли не основной для передачи информации между людьми (речь). Акустические сигналы низкой интенсивности безвредны для здоровья человека.

3. *Оптические каналы*. Используют оптические линии связи. Являются перспективными, т. к. обладают в настоящее время максимальной пропускной способностью. Обладают повышенной помехоустойчивостью к электромагнитным помехам. Отличаются низким уровнем паразитных излучений, что усложняет перехват сообщений, передаваемых по таким каналам, и делает их более безопасными. Физической средой для распространения света может быть газ (воздух), стекло (стекловолокно), вакуум (космос) и др. В качестве излучателя в настоящее время используются лазеры. Возможность точного выбора направления излучения в открытой среде (не в стекловолокне) обеспечивает скрытность и помехоустойчивость передачи информации.

4. *Электрические каналы*. Наиболее распространены в настоящее время при передаче информации на малые расстояния. Основа – проводные линии связи.

5. *Радиоканалы*. Как и оптические, используют для передачи информации электромагнитные волны, однако намного более низкой частоты. Благодаря способности таких волн огибать препятствия и отражаться от плазменных слоев, окружающих Землю, становится возможным передача информации на большие расстояния, в том числе в масштабе всей Земли. Эти преимущества, однако, являются источником недостатков. Радиоканалы сильно подвержены влиянию помех и менее скрытны. Радиоканал, наряду с оптическим, может использоваться для подключения к компьютерной сети Интернет в районах со слаборазвитой инфраструктурой проводной электросвязи.

5.2. Разделение каналов связи

Свойства канала передачи информации сильно зависят от используемой линии передачи информации. По разным причинам этих линий не может быть много. Например, в радиоканале она всегда одна – «эфир». Поэтому приходится каким–то образом выделять сигналы разных каналов, использующих одну и ту же линию передачи информации. Различные способы такого выделения принято называть *разделением каналов*.

1. *Частотное разделение каналов*. Весь спектр, пропускаемых линией связи частот делится на участки по количеству каналов. В каждом из каналов используются сигналы со спектром, лежащим в пределах своего участка. Через линию передачи информации передается сумма сигналов, каждый из которых имеет свой, не перекрывающийся с другими спектр. Это и позволяет на приемном конце выделить каждый из сигналов и воспринять переносимую им информацию. Для разделения сигналов используются так называемые полосовые фильтры.

Так работает радио– и телевещание, когда одновременно одна радиолиния связи (эфир) одновременно используется тысячами радиостанций для организации доставки информации миллионам жителей Земли (рис. 5.1).

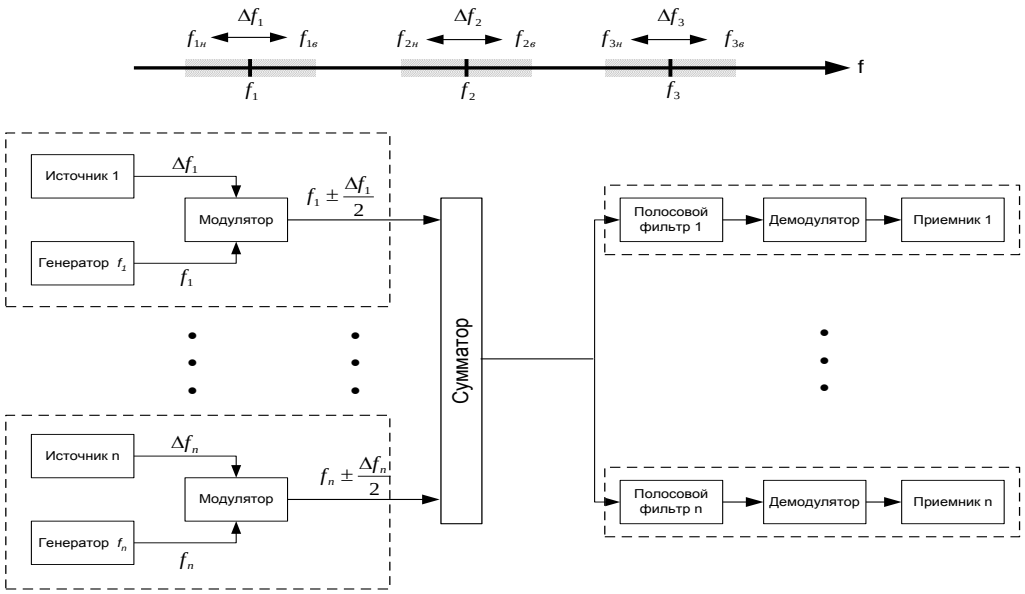


Рис. 5.1. Схема системы передачи информации, основанной на частотном разделении каналов.

Подобным же образом передача может вестись и по оптической линии передачи информации. Так по одному оптоволоконному кабелю одновременно может передаваться несколько миллионов телефонных разговоров, сотни телепрограмм и цифровых данных.

2. *Временное разделение каналов*. При таком виде разделения сигналы от источников передаются только в отведенные для них непересекающиеся отрезки времени длительностью Δt_i (рис. 5.2).

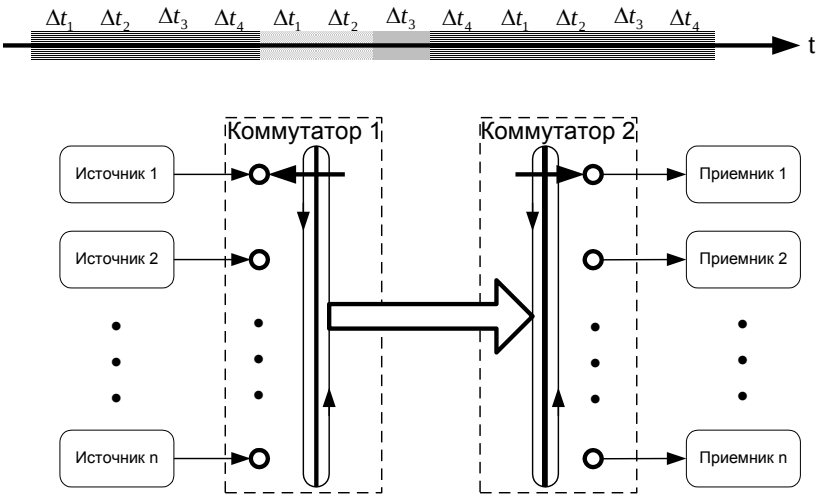


Рис. 5.2. Схема системы передачи информации, основанной на временном разделении каналов.

Такое использование линии связи в компьютерной технике называют *режимом разделения времени*.

3. *Кодовое разделение каналов*. Кодовое разделение похоже на временное. Отличие заключается в способе управления коммутатором или все чаще коммутаторами (на пути от источника к приемнику может оказаться множество коммутаторов). Передаваемые данные содержат адрес приемника. На основании его специальные компьютеры − коммутаторы − направляют сообщение к нужному приемнику. Это наиболее распространенный в настоящее время способ разделения каналов.

4. *Разделение каналов по уровню*. Предположим, что источники выдают сообщения, закодированные одноразрядным двоичным кодом. Эти сообщения можно использовать как разряды многоразрядного двоичного числа и подавать это число на вход цифроаналогового преобразователя. Сигнал с выхода этого преобразователя подается в линию связи для передачи на нужное расстояние. На

приемном конце устанавливается аналого–цифровой преобразователь. На его выходе образуются сигналы, соответствующие передаваемым (рисунок 5.3).

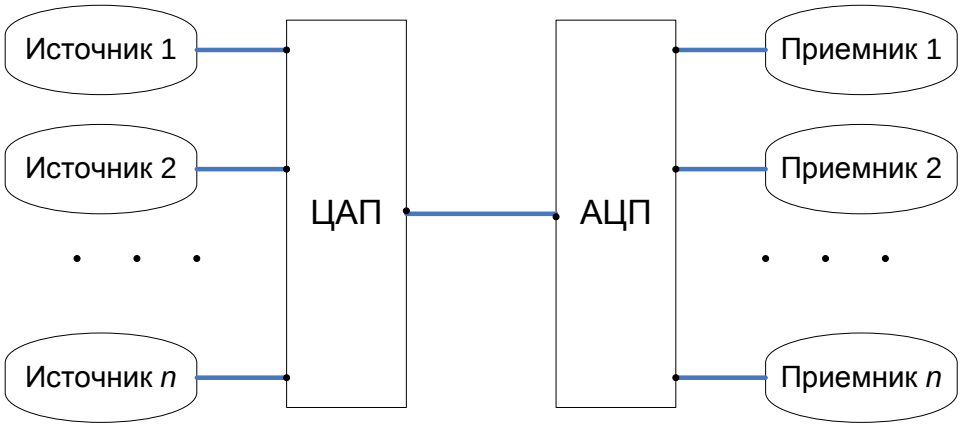


Рис. 5.3. Схема многоканальной системы передачи двоичной информации, основанной на разделении по уровню.

В таблице 5.1 приведен пример исходных данных для 4 источников *И1–И4* двоичных данных, а на рисунке 5.4 те же данные в графический форме, где *X* – сигнал с выхода 4–хразрядного аналого–цифрового преобразователя.

Таблица 5.1

Пример исходных данных для 4 источников *И1–И4* двоичных данных

№ опыта	*И1*	*И2*	*И3*	*И4*	***X***
1	1	0	1	0	**5**
2	1	0	1	0	**5**
3	1	0	1	0	**5**
4	1	0	1	0	**5**
5	0	1	0	0	**2**
6	0	1	0	0	**2**
7	0	1	0	0	**2**
8	0	1	0	0	**2**
9	0	0	0	0	**0**
10	0	0	0	0	**0**
11	0	0	0	0	**0**
12	0	0	0	0	**0**
13	0	0	0	1	**8**
14	0	0	0	1	**8**
15	0	0	0	1	**8**
16	0	0	0	1	**8**
17	0	1	0	0	**2**
18	0	1	0	0	**2**
19	0	1	0	0	**2**

Продолжение таблицы 5.1

№ опыта	*И1*	*И2*	*И3*	*И4*	***X***
20	0	1	0	0	**2**
21	1	0	0	1	**9**
22	1	0	0	1	**9**
23	1	0	0	1	**9**
24	1	0	0	1	**9**
25	0	1	1	1	**14**
26	0	1	1	1	**14**
27	0	1	1	1	**14**
28	0	1	1	1	**14**
29	1	0	1	1	**13**
30	1	0	1	1	**13**
31	1	0	1	1	**13**
32	1	0	1	1	**13**
33	0	1	0	1	**10**
34	0	1	0	1	**10**
35	0	1	0	1	**10**
36	0	1	0	1	**10**
37	1	0	0	1	**9**
38	1	0	0	1	**9**
39	1	0	0	1	**9**
40	1	0	0	1	**9**

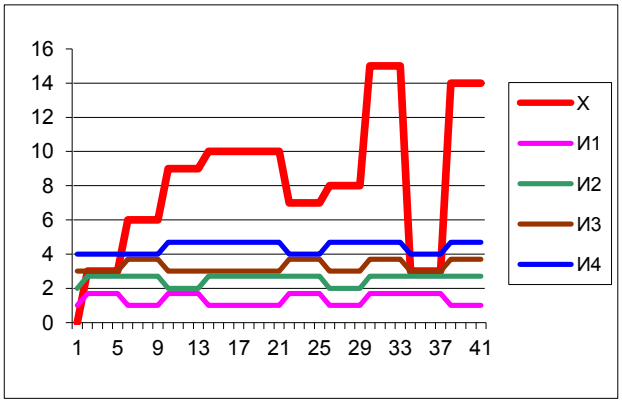


Рис. 5.4. Графики сигналов на входе и выходе АЦП

Сигнал *X* с выхода 4–хразрядного аналого–цифрового преобразователя и исходные данные для 4 источников *И1 −И4* связаны характерной для АЦП зависимостью:

$$X = И1 + 2 \times И2 + 4 \times И3 + 8 \times И4.$$

Ясно, что разрядность АЦП и ЦАП совпадает с числом разделяемых каналов. При наличии помех она не должна быть слишком большой.

Для безошибочной передачи разделяемых по уровню каналов нужно чтобы помеха не превосходила половины минимального шага уровня напряжения на выходе АЦП.

5. *Корреляционное разделение каналов*. В основу такого разделения положено использование системы ортогональных нормированных функций $\{\psi_i(t)\}$ на интервале (0; *T*).

В канал связи подается сумма

$$x(t) = \sum_{i=1}^{n} a_i \psi_i(t).$$

Информация с *i*–го канала здесь передается амплитудой сигнала a_i.

На приемном конце получают сигнал b_i:

$$b_i = \int_0^T x(t) \cdot \psi_i(t) dt = \int_0^T \sum_{j=1}^{n} a_j \psi_j(t) \cdot \psi_i(t) dt = \sum_{j=1}^{n} a_j \int_0^T \psi_j(t) \cdot \psi_i(t) dt = a_i,$$

т. к.

$$\int_0^T \psi_j(t) \cdot \psi_i(t) dt = \begin{cases} 1 \ \ при \ \ i = j \\ 0 \ \ при \ \ i \neq j \end{cases}$$

За отрезок времени *T* можно таким образом одновременно передать *n* величин a_i $(1 \le i \le n)$. Следующая порция *n* величин a_i передается за следующий отрезок времени длиной *T*.

На рисунке 5.5 изображена блок–схема системы корреляционного разделения каналов.

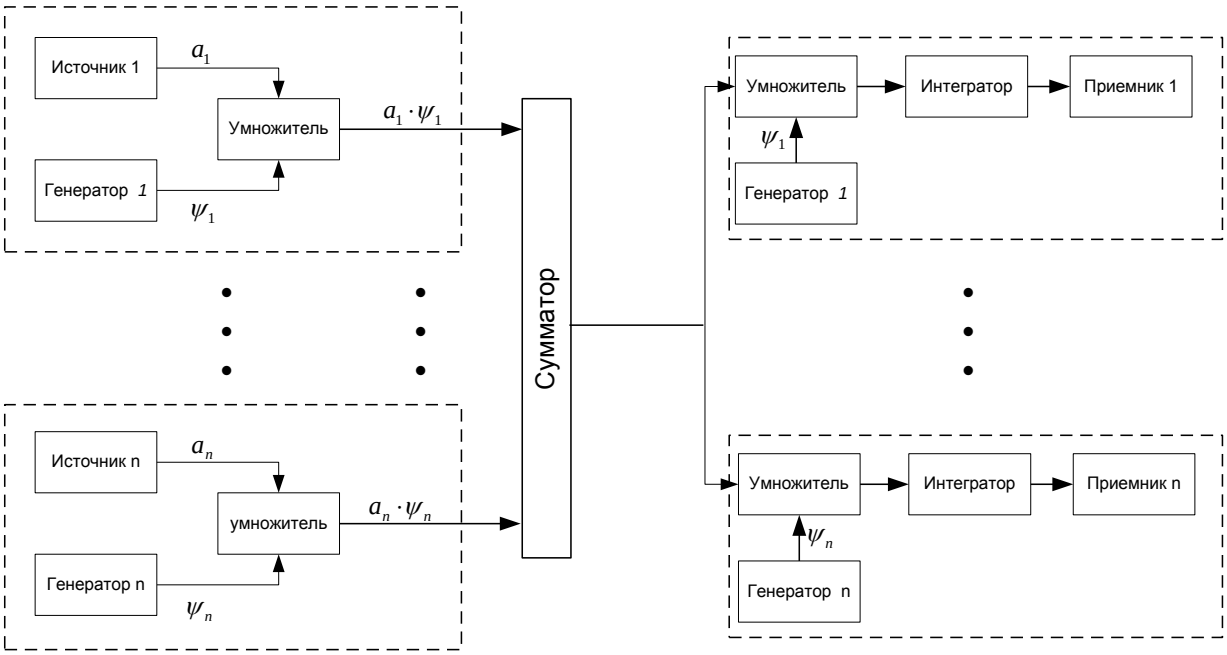


Рис. 5.5. Схема многоканальной системы передачи информации, основанной на корреляционном разделении каналов

Видно, что эта схема во многом подобна схеме частотного разделения. Это не случайно. Действительно, здесь есть сходство. Частотное разделение основано на использовании одного из видов ортонормированных функций – набора синусоидальных функций. Но они образуют лишь одну из множества возможных систем ортонормированных функций. Существуют еще системы функций Радемахера, Хара, Уолша, Лежандра и др., которые также можно использовать для разделения каналов и которые при определенных условиях обладают преимуществами по сравнению с остальными.

6. *Комбинированные методы разделения каналов*. Комбинированные методы разделения каналов позволяют увеличить число каналов и уменьшить их взаимное влияние.

Можно, например, сочетать временное или кодовое разделение с частотным, амплитудным или корреляционным.

В заключение отметим, что здесь перечислены далеко не все возможные методы разделения каналов связи. Благодаря непрерывающимся исследованиям, появляются новые и усовершенствуются ранее известные методы разделения каналов передачи информации.

5.3. Пропускная способность каналов связи

Эта тема является одной из центральных в теории информации. В ней рассматриваются предельные возможности каналов связи по передаче информации, определяются характеристики каналов, влияющие на эти возможности, исследуются в самом общем виде предельные возможности кодирования, обеспечивающие максимум помехоустойчивости и объема передаваемой информации.

Приведем основные определения

1. *Скорость передачи информации* – среднее количество информации, передаваемой через канал за единицу времени.

В случае канала без шума эта скорость равна $V_к \times H_к$, где $V_к$ – количество символов, передаваемых через канал в единицу времени, $H_к$ – средняя энтропия одного символа сообщения на входе и выходе канала.

2. *Производительность источника* – средняя скорость поступления информации от источника сообщений.

Производительность источника находится по формуле $V_и \cdot H_и$, где $V_и$ – количество символов, генерируемых источником в единицу времени; $H_и$ – средняя энтропия одного символа сообщения на выходе источника.

3. Пропускная способность канала связи – максимально возможная для данного канала скорость передачи информации. Будем обозначать ее $C_к$.

Отметим еще одну важную характеристику канала – максимальную скорость передачи символов $V_{кmax}$ через него. Она всегда ограничена. Поэтому максимальная скорость передачи информации достигается при использовании максимальной скорости передачи символов и максимальной средней энтропии $V_{кmax}$ передаваемого символа. Ранее доказывалось, что максимальная средняя энтропия в расчете на один символ достигается при равной вероятности и независимости их появления.

Поскольку источник информации совсем не обязательно выдает символы с такими характеристиками, для достижения максимально эффективного использования канала их необходимо кодировать. Ранее при изучении эффективного кодирования доказыва-

лось, что именно эффективное кодирование обеспечивает получение после кодирования символов с требуемыми параметрами. Энтропия символов вторичного алфавита в результате такого кодирования при кодировании бесконечно больших блоков информационной последовательности в пределе равна $log_2 m$, где m – объем вторичного алфавита, используемого на выходе кодирующего устройства.

Учитывая это

$$C_к = V_к \cdot H_{max} = V_к \cdot log_2 m$$

Если же $m = 2$ (для кодирования используется двоичный код), то энтропия одного символа на выходе кодера будет равна 1, т. е. каждый символ двоичного эффективного кода будет нести 1 бит информации, а сами символы будут равновероятны и статистически независимы.

В этом случае $C_к = V_к$.

При передаче информации через канал связи стремятся к наиболее эффективному (в смысле объема передаваемой информации) его использованию.

Найдем требования к источнику информации, при которых возможна максимальная скорость передачи информации через канал.

Будем описывать источник информации параметрами $V_и$ и $H_и$. Допустим, шум в канале связи отсутствует. Канал связи описывается своей пропускной способностью и объемом m алфавита.

Поскольку шума в канале нет, информация при передаче через него не искажается и не теряется. Поэтому скорости передачи информации на выходе источника $V_и \cdot H_и$ и на выходе канала будут совпадать. Наиболее эффективным будет такое использование канала, при котором производительность источника будет равна пропускной способности канала

$$C_к = V_{кmax} \cdot log_2 m = V_и \cdot H_и.$$

Таким образом, если известна средняя энтропия одного символа сообщения, поступающего с выхода источника, наиболее эффективного использования канала можно достичь, если скорость поступления этих символов от источника выбрать в соответствии с формулой

$$V_и = V_{кmax} \cdot log_2 m / H_и \text{ или } V_и = V_{кmax} / H_и$$

при использовании наиболее часто употребляемого двоичного кодирования.

Заметим, что эта формула предполагает использование эффективного кодирования информации, поступающей от источника перед передачей ее в канал связи без помех (шума).

Рассмотрим следующую модель канала связи с помехами (рис. 5.6).

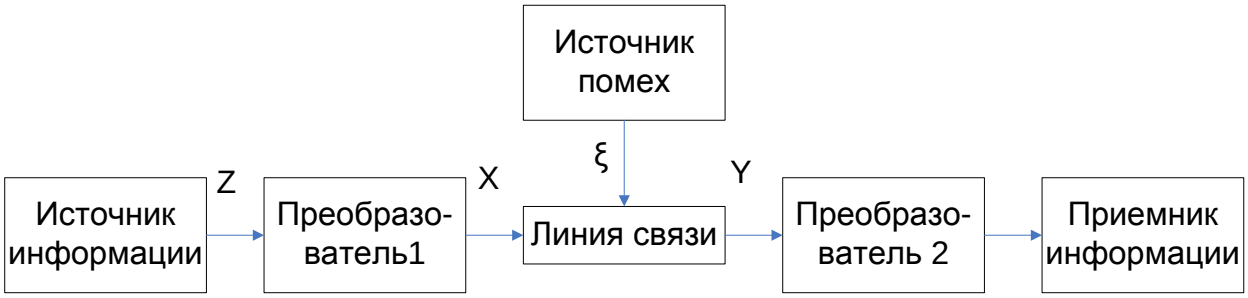


Рис. 5.6. Модель канала связи с помехами.

По виду передаваемых через канал сигналов различают дискретные и непрерывные каналы связи.

Важнейшей характеристикой канала является его *пропускная способность, определяемая как наибольшая скорость передачи информации через него*. Пропускная способность дискретного канала может быть рассчитана, например, по следующей формуле:

$$C = V_к \cdot I_{max},$$

где $V_к$ – скорость передачи символов алфавита через канал;

I_{max} – максимально возможное количество информации, приходящейся на один передаваемый через канал символ.

Количество информации, приходящееся на 1 передаваемый через канал символ зависит от энтропии (степени неопределенности получения символа) на входе и выходе канала. Согласно мере Шеннона

$$I = H_{апр} - H_{апст} = H(X) - H(X/Y)$$

Здесь $H_{апр} = H(X)$ *и* $H_{апст} = H(X/Y)$ – условная энтропия, характеризующая неопределенность о переданном на выход канала символе *X* по принятому символу *Y* на выходе. Наличие этой неопределенности – следствие действия на передаваемый через канал символ помех. Условная энтропия $H(X/Y)$ – характеристика канала.

5.3.1. Пропускная способность дискретного канала связи с шумом

Исследуем теперь пропускную способность дискретного канала связи с шумом.

Существует большое количество математических моделей таких каналов. Простейшей из моделей является канал с независимой от передаваемой информации помехой, имеющей независимые во времени значения. Последнее означает, что влияние помехи на различные буквы передаваемого сообщения не зависят друг от друга. Будем рассматривать канал без стирания с алфавитом объема *m* (не двоичный симметричный канал ДСМК, т. к. *m* не равно 2).

Если в канале без помех среднее количество информации, получаемой с выхода канала в единицу времени, соответствует среднему количеству информации, содержащейся во входном сообщении той же длительности, то в канале связи с шумом это соответствие *нарушается*. При наличии помехи скорость передачи информации через канал может быть меньше скорости ее поступления на его вход, т. к. часть информации за счет помех в канале может теряться.

Помехи нарушают взаимно–однозначное соответствие между символами на входе и выходе канала, что приводит к появлению на

выходе канала остаточной неопределенности об информации, поступившей на вход канала.

Количество переданной информации при этом вычисляется согласно формуле Шеннона как разность между априорной и апостериорной энтропиями:

$$I = H_{апр} - H_{апст},$$

$$H_{апр} = H_{ист} = -\sum_{i=1}^{m} p(b_i)\log_2(p(b_i)),$$

Где b_i – i–я буква из алфавита объемом m;

$H_{ист}$ - энтропия источника информации;

$H_{апст} = H(B/B')$ – средняя условная энтропия 1 буквы на входе канала связи с шумом относительно 1 буквы, принятой на выходе канала связи. Найдем эту энтропию.

В результате искажения передаваемые буквы b_i превращаются в другие b_i' из того же алфавита. Вероятность такого превращения можно описать графом, называемым *канальной матрицей* (рис. 5.7).

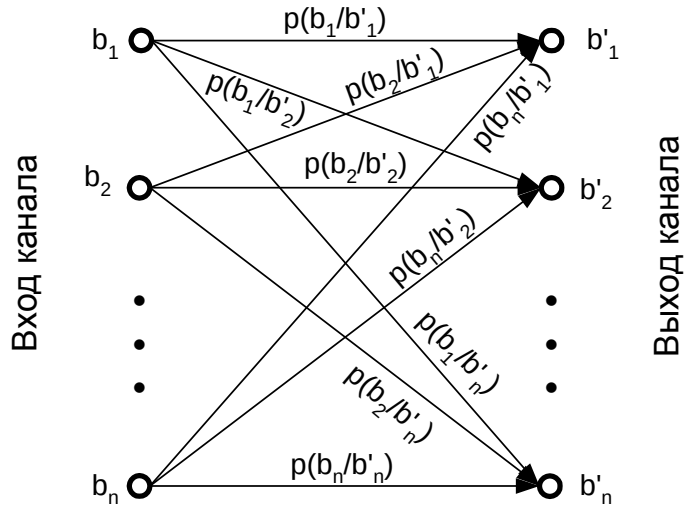


Рис. 5.7. Канальная матрица

Каждому ребру этого графа соответствует условная вероятность $p(b_i / b_j')$ – вероятность того, что на вход канала поступал

символ b_i при условии, что на выходе был получен символ b_j' (выходные символы обозначаются со штрихом).

Неопределенность (энтропия) того, какой символ был на входе канала связи по символу, полученному на выходе, находится согласно определению энтропии по формуле

$$H(b_i / b_j') = -log_2 p(b_i / b_j').$$

Нас интересует средняя неопределенность независимо от сочетания входной и выходной буквы. Вероятность такого сочетания – совместная вероятность $p(b_i / b_j')$.

Используя совместную вероятность, среднюю условную энтропию получаем в виде

$$H(B / B') = -\sum_{i=1}^{m}\sum_{j=1}^{m} P(b_i, b_j^{'}) \log_2 p(b_i / b_j^{'})$$

Учитывая, что $p(b_i, b_j^{'}) = \dfrac{p(b_i) p(b_j^{'} / b_i)}{p(b_j^{'})}$ (формула Байеса) и $p(b_j^{'}) = \sum_{i=1}^{m} p(b_i, b_j^{'}) = \sum_{i=1}^{m} p(b_i) p(b_j^{'} / b_i)$, получим:

$$H(B/B') = -\sum_{i=1}^{m}\sum_{j=1}^{m} p(b_i) p(b_j^{'}, b_i) \log_2 \frac{p(b_i) p(b_j^{'} / b_i)}{\sum_{i=1}^{n} p(b_i) p(b_j^{'} / b_i)} \qquad (5.1)$$

Значение условной энтропии $H(B / B')$ характеризует потери информации в канале и названо Клодом Шенноном *ненадежностью канала*.

Из формулы (5.1) следует, что *ненадежность канала* зависит от статистических характеристик входного сообщения *B* (вероятностей $p(b_i)$), а также от искажений, описываемых условными вероятностями $p(b_j' / b_i)$.

Подставив теперь выражения для априорной и апостериорной энтропии в формулу среднего количества информации на символ сообщения, получим

$$I = -\sum_{i=1}^{m} p(b_i)\log_2 p(b_i) + \sum_{i=1}^{m}\sum_{j=1}^{m} p(b_i, b_j^{'})\log_2 p(b_i / b_j^{'}).$$

После преобразований получим

$$I = \sum_{i=1}^{m}\sum_{j=1}^{m} p(b_i, b_j^{'})\log_2 \frac{p(b_i, b_j^{'})}{p(b_i)\,p(b_j^{'})} \qquad (5.2)$$

Заметим, что формула (5.2) симметрична относительно b_i и $b_i{}'$, поэтому

$$I = H(B) - H(B/B') = H(B') - H(B'/B), \qquad (5.3)$$

т. е., *во–первых*, среднее количество информации на символ равно разности энтропии на входе и неопределенности на выходе относительно входа; *во–вторых*, оно равно разности неопределенности (энтропии) на выходе и неопределенности того, что получено на выходе, если известно, что было передано на входе канала.

Пропускная способность получается путем максимизации скорости передачи информации по каналу связи путем подбора вероятностей $p(b_i)$.

$$C_к = V_{к\,max} \cdot I = V_{к\,max} \cdot \left[H(B) - H(B/B')\right].$$

Вообще, и уменьшаемое $H(B)$, и вычитаемое $H(B/B')$ в этой формуле зависят от $p(b_i)$, однако большинство авторов зависимостью вычитаемого $H(B/B')$ от $p(b_i)$ ввиду его малости пренебрегают и учитывают зависимость $C_к$ только от уменьшаемого $H(B)$.

Известно, что энтропия $H(B)$ максимальна при равенстве всех вероятностей $p(b_i)$. Это равенство возможно только при условии $p(b_i)=1/m$. В результате получаем, что пропускная способность дискретного канала связи с шумом равна

$$C_к = V_к[\log_2 m - H(B/B')] \ . \qquad (5.4)$$

Посмотрим теперь, как влияет шум на пропускную способность канала связи.

Ненадежность канала $H(B/B')$ будет минимальна и равна 0, если шума нет. Тогда символы на входе и выходе всегда совпадают. Это значит, что

$$p(b_i / b_j^{'}) = \begin{cases} 0 \ \ при \ \ i = j \\ 1 \ \ при \ \ i \neq j \end{cases}$$

В этом случае $H(B/B')=0$, а пропускная способность $C_{к\max}$ совпадает с пропускной способностью канала связи без шума, что и следовало ожидать.

В том же случае, когда шум настолько велик, что все символы на выходе могут появляться равновероятно не зависимо от того, какой символ подавался на вход канала связи, т. е. если $p(b_i / b_j')=1/m$ и $p(b_j')=1/m$, то $I=0$.

Это несложно доказать.

$$H(B/B') = -\sum_{i=1}^{m}\sum_{j=1}^{m} p(b_i, b_j^{'}) \log_2 p(b_i / b_j^{'}) =$$

$$= -\sum_{i=1}^{m}\sum_{j=1}^{m} p(b_j^{'}) p(b_i / b_j^{'}) \log_2 p(b_i / b_j^{'}) =$$

$$= \frac{1}{m^2} \log_2 m \sum_{i=1}^{m}\sum_{j=1}^{m} 1 = \log_2 m \ ,$$

а так как $I = H(B) - H(B/B')$, где $H(B) = \log_2 m$, то

$$I = \log_2 m - \log_2 m = 0$$

Для доказательства теоремы Шеннона для дискретного канала с шумом, введем понятие типичных последовательностей и рассмотрим их свойства.

5.3.2. Типичные последовательности и их свойства

Будем рассматривать последовательности статистически независимых букв. Согласно закону больших чисел, наиболее вероятными будут последовательности длиной n, в которых при $n \to \infty$ количества N_i появлений букв отвечают зависимости $N_i = p_i \cdot n$, где p_i – вероятность появления буквы i.

Последовательности, для которых выполняется равенство $N_i = p_i \times n$, называются *типичными*.

Согласно закону больших чисел при росте n все большее число последовательностей становятся типичными, а в пределе, при $n \to \infty$, все последовательности становятся типичными.

Вероятность появления любой типичной последовательности вычисляется по одной и той же формуле

$$p_{тип} = p_1^{p_1 n} \cdot p_2^{p_2 n} \cdot p_3^{p_3 n} \cdot \ldots \cdot p_m^{p_m n} = \prod_{i=1}^{m} p_i^{p_i n},$$

где m – число букв в алфавите (объем алфавита).

Ниже нам понадобится формула, связывающая среднюю энтропию H буквы сообщения и длину n сообщения с количеством типичных последовательностей $N_{\min}$.

Поскольку с ростом n все последовательности становятся типичными, то сумма вероятностей их появления

$$\sum_{j=1}^{N_{min}} p_{min\ j} = 1.$$

Учтем, кроме того, равенство этих вероятностей.

Отсюда находим энтропию $H_{тип}$ типичной последовательности

$$H_{тип} = -\log_2(1/N_{тип}) = \log_2 N_{тип}$$

Если H – энтропия буквы сообщения, то энтропия типичной последовательности $H_{тип} = n \cdot H = \log_2 N_{тип}$, то

$$N_{тип} = 2^{nH} . \qquad (5.4)$$

Теперь мы готовы к восприятию основной теоремы Шеннона для дискретного канала с шумом.

5.3.3. Основная теорема Шеннона для дискретного канала с шумом

Формулировка. Для дискретного канала с шумом существует такой способ кодирования, при котором может быть обеспечена безошибочная передача всей информации, поступающей от источника сообщений, если только пропускная способность канала превышает производительность источника сообщений, т.е. если выполняется условие

$$V_{к\,max}\left[\ \log_2 m\ -\ H\left(B/B'\right)\right]\ >\ V_u \cdot H\left(A\right),$$

где $H\left(A\right)$ – энтропия источника.

Здесь обозначения соответствуют схеме, изображенной на рисунке 5.8.

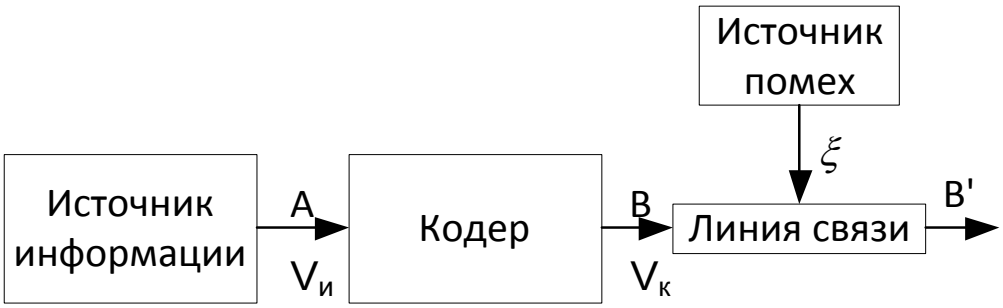


Рис. 5.8. Схема дискретного канала связи с шумом: *B, B'* – символы на входе и выходе канала; $V_и$ и $V_к$ – скорость поступления символов от источника на вход кодирующего устройства и с выхода кодирующего устройства на вход линии связи.

Доказательство.

Согласно формуле (5.4), число типичных последовательностей длительностью *T* равно

$$N_{тип}(A) = 2^{V \cdot и \cdot T \cdot H(A)}.$$

Максимально возможное число различных последовательностей на выходе кодера, которое сможет иметь ту же длительность *T*, равно

$$N(B) = m^{V \cdot к \cdot T}$$

где *m* – объем вторичного алфавита.

Эту формулу можно записать также в виде

$$N(B) = 2^{V_к \cdot T \cdot \log_2 m}$$

Предположим, что условие теоремы

$$V_к[\log_2 m - H(B/B')] > V_и \cdot H(A)$$

выполняется. Тогда, поскольку $H(B/B') > 0$, неравенство условия теоремы только усилится, если его записать в виде:

$$V_к \cdot \log_2 m > V_и \cdot H(A).$$

Домножим левую и правую части неравенства на T и сделаем их показателями двойки:

$$2^{T \cdot V_к \cdot \log_2 m} = N(B) \ > \ 2^{T \cdot V_и \cdot H(A)} = N_{тип}(A).$$

Таким образом, возможное число последовательностей длительностью T на выходе кодера больше числа типичных последовательностей той же длительности, которое может выдать источник.

Если установить однозначное соответствие между последовательностями на выходе кодера с последовательностями на входе, остаются последовательности, которые не могут использоваться для кодирования, т. е. *запрещенные* кодовые слова. Если все же такая последовательность получается с выхода канала связи – это верный признак ошибки.

Каждому способу кодирования будет соответствовать определенный вариант установки соответствия между каждой из множества типичных последовательностей, генерируемых источником и кодовым словом, взятым из множества разрешенных кодовых слов $N(B)$.

Требуется теперь доказать, что среди всех возможных способов установки этого соответствия существует такой, который обеспечивает однозначную идентификацию любого принятого кодового слова, даже если оно получилось в результате искажения помехой передаваемого слова.

Для доказательства найдем вероятность правильного приема, усредненную по всем возможным способам кодирования, т. е. установки соответствия между принятым и возможно искаженным кодовым словом и закодированной информационной последовательностью.

По условной энтропии $H(B/B')$ можно найти количество возможных типичных последовательностей на входе канала, кото-

рые после искажения в нем могли бы дать определенную последовательность на выходе канала

$$N_{тип}(B/B') = 2^{T \cdot V_к \cdot H(B/B')}.$$

Вывод аналогичен сделанному ранее при исследовании свойств типичных последовательностей. Только в данном случае энтропия символа на выходе канала – функция того, какой символ получен на входе канала – $H(B/B')$.

Правильный прием возможен только тогда, когда на $[N_{тип}(B/B')-1]$ не разрешенных кодовых слов приходится одно разрешенное, т. е. число возможных кодовых слов вообще должно быть в $N_{тип}(B/B')$ раз больше числа типичных $N_{тип}(A)$ последовательностей, которые может выдать источник.

Предположим теперь, что соответствия между типичными последовательностями, выдаваемыми источником, и кодовыми словами устанавливаются равновероятно. Тогда вероятность $p_{разр}$ того, что определенное кодовое слово среди $N_{тип}(A)$ является разрешенным, равна $p_{разр} = N_{тип}(A)/N(B)$ (кодируются только типичные последовательности).

Соответственно, вероятность того, что кодовое слово является запрещенным, равна

$$1 - p_{разр} = 1 - N_{тип}(A)/N(B).$$

Правильное декодирование возможно только тогда, если среди $N_{тип}(B/B')$ кодовых слов только одно выбрано в качестве разрешенного.

Вероятность этого равна:

$$p_{прав} = (1 - p_{разр})^{[N_{тип}(B/B')-1]} = \left[1 - \frac{N_{тип}(A)}{N(B)}\right]^{N_{тип}(B/B')-1}.$$

Так как основание степени $\left[1-\frac{N_{тип}(A)}{N(B)}\right]<1$, то, увеличивая показатель степени с $N_{тип}(B/B')-1$ до $N_{тип}(B/B')$, получим неравенство:

$$p_{прав}>\left[1-\frac{N_{тип}(A)}{N(B)}\right]^{N_{тип}(B/B')}.$$

Разложим правую часть этого неравенства в ряд по степеням $\frac{N_{тип}(A)}{N(B)}$, т. е. в ряд Тэйлора, который для функции *y(x)* одной переменной *x*, непрерывной и имеющей все производные при x=a, имеет следующий вид:

$$f(x)=f(a)+\frac{x-a}{1!}f'(a)+\frac{(x-a)}{2!}f''(a)+...+\frac{(x-a)^n}{n!}f^n(a)+...$$

и получим

$$p_{прав}>1-N_{тип}(B/B')\cdot\frac{N_{тип}(A)}{N(B)}+$$
$$+\frac{1}{2}\cdot N_{тип}(B/B')\cdot\left[N_{тип}(B/B')-1\right]\cdot\left[\frac{N_{тип}(A)}{N(B)}\right]^2+...$$

Поскольку $\frac{N_{тип}(A)}{N(B)}<1$ и члены ряда по абсолютной величине убывают, неравенство только усиливается, если отбросить все члены ряда выше первого порядка:

$$p_{прав}>1-\frac{N_{тип}(B/B')\cdot N_{тип}(A)}{N(B)}.$$

Получим

$$p_{ош} = 1 - p_{прав} > N_{тип}(B/B') \cdot \frac{N_{тип}(A)}{N(B)} =$$

$$= 2^{-T\left[V_к \cdot \log_2 m - V_к \cdot H(B/B') - V_и \cdot H(B)\right]} = 2^{-T\left[C_к - V_и \cdot H(A)\right]}$$

где $C_к = V_к \cdot \log_2 m - V_к \cdot H(B/B')$ – пропускная способность канала связи с учетом потерь информации в результате действия помех (шума).

Заметим, что, если $C_к > V_и \cdot H(A)$ (условие теоремы Шеннона), с ростом T вероятность ошибки $P_{ош}$ стремится к 0.

Помимо ошибки, возможной при приеме типичных последовательностей, вероятность которой определяется последней формулой, возможны ошибки за счет появления на выходе источника нетипичных последовательностей, кодирование которых вообще не предусматривается описанной процедурой. При вероятностях появления нетипичной последовательности δ и типичной 1-δ суммарная вероятность ошибочной передачи равна:

$$p_{\Sigma} = \delta + (1 - \delta) \cdot p_{ош}.$$

При $T \to \infty$, как было показано ранее, $\delta \to 0$.

Если $C_к - V_и H(A) > 0$ (т.е. при условии выполнения теоремы Шеннона), то при $T \to \infty$ и $p_{ош} \to 0$ суммарная ошибка $p_{\Sigma} \to 0$.

Поскольку среди всех вариантов кодирования обязательно существует хотя бы один, для которого вероятность ошибки не превышает среднего по всем вариантам кодирования значения p_{Σ}, теорему Шеннона можно считать доказанной.

5.3.4. Пропускная способность непрерывного канала при наличии аддитивного шума

Рассмотрим следующую модель канала:

1) канал способен пропускать колебания с частотами ниже F_m;

2) в канале действует помеха $n(t)$, имеющая нормальный (гауссовский) закон распределения с нулевым средним значением;

3) помеха $n(t)$ статистически не связана с полезным сигналом $x(t)$;

4) помеха аддитивна, т. е. сигнал $y(t)$ на выходе канала описывается формулой $y(t) = x(t) + n(t)$, где $x(t)$ – сигнал на входе канала;

5) мощность P_x полезного сигнала $x(t)$ ограничена.

Ограниченность полосы пропускания канала приводит к ограниченности спектра выходного сигнала $x(t)$. Поэтому согласно теореме Котельникова без потери информации сигналы $x(t)$, $y(t)$ и $n(t)$ можно представить в виде ряда независимых отсчетов, взятых с шагом $\Delta t \le \dfrac{1}{2F_m}$.

Как было доказано ранее, количество информации *I*, приходящейся на один отсчет, равно

$$I = H(x) - H(x/y) = H(y) - H(y/x) \qquad (5.5)$$

Ранее нами было показано, что энтропия $H(y)$ непрерывной случайной величины *y* находится по формуле:

$$H(y) = -\int_{-\infty}^{\infty} f(y)\log_2 f(y)dy + \lim_{\Delta y \to 0} \log_2 \Delta y.$$

Найдем $H(y/x)$. Перейдем к дискретному представлению *x* и *y* с шагом Δx и Δy.

$$H(y/x) = -\lim_{\Delta x \to 0; \Delta y \to 0} \sum_i \sum_j p(y_i, x_j) \cdot \log_2 p(y_i, x_j);$$
$$p(y_i, x_j) = p(x_j) \cdot p(y_i / x_j);$$
$$p(x_j) \approx f(x_j) \cdot \Delta x;$$
$$p(y_i / x_j) = f(y_i / x_j) \cdot \Delta y.$$

Найдем теперь дифференциальный закон $f(y/x)$. Если x фиксирован, а $y(t) = x(t) + n(t)$, то

$$f(y/x) = f(n - x).$$

График этой функции изображен на рисунке 5.9.

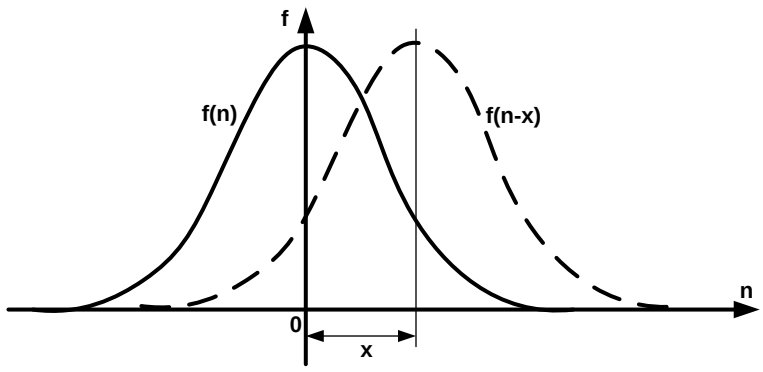


Рис. 5.9. График функции *f*(*y*/*x*)

Поскольку X фиксирован, то $\Delta y = \Delta n$.

Произведем подстановки:

$$H(y/x) = -\lim_{\Delta x \to 0, \Delta y \to 0} \sum_i \sum_j f(x_j) \Delta x \; f(y_i / x_j) \Delta y \log_2 \left[f(y_i / x_j) \Delta y \right] =$$

$$= -\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x) f(n - x) \log_2 \left[f(n - x) \Delta y \right] dx dn =$$

$$= -\int_{-\infty}^{\infty} f(x)dx \int_{-\infty}^{\infty} f(n)\log_2\left[f(n)\Delta y\right]dn + \log_2(\Delta y)\int_{-\infty}^{\infty} f(n)dn =$$

$$= -\int_{-\infty}^{\infty} f(n)\log_2\left[f(n)\right]dn + \log_2(\Delta y)$$

Подставив этот результат в формулу количества информации (5.5), получим:

$$I = -\int_{-\infty}^{\infty} f(y)\log_2\left[f(y)\right]dy + \left\{-\int_{-\infty}^{\infty} f(n)\log_2\left[f(n)\right]dn\right\} =$$

$$= H_{\varepsilon=1}(y) - H_{\varepsilon=1}(n).$$

Скорость передачи информации по теореме Котельникова будет равна:

$$\frac{dI}{dt} = V_к I,$$

где $\frac{dI}{dt} = V_к I$. Для нахождения пропускной способности непрерывного канала связи с шумом следует максимизировать скорость передачи информации $\frac{dI}{dt}$, варьируя параметрами передаваемого сигнала X;

$V_к$ – величина постоянная, связанная с параметром F_m канала;

I зависит от $H(y)$.

Поскольку x и n взаимно независимы,

$$D_y = D_x + D_n, m_y = m_x + m_n = m_x,$$

т. к. $m_n = 0$.

Ранее была выведена зависимость $H_{\varepsilon=1}(y) = \log_2 \sqrt[2]{2\pi e Dy}$, из которой следует, что $H_{\varepsilon=1}(y)$ пропорциональна D_y.

Отсюда следует, что $H_{\varepsilon=1}(y)$ максимальна, когда максимальна дисперсия D_y.

Но, поскольку $D_y = D_x + D_n$, D_y – максимальна, когда максимальна D_x.

Согласно условию, мощность P_x передаваемого сигнала X ограничена сверху.

А, поскольку $P_x = D_x + m_x{}^2$, максимальное значение дисперсии D_x достигается при нулевом математическом ожидании $m_x=0$. В это случае $P_x = D_x$.

Значит $P_y = D_y + m_y{}^2 = D_y$, т. к. $m_y = m_x + m_n$, $m_x = 0$, $m_n = 0$.

Ранее было доказано, что при ограничении мощности сигнала сверху его энтропия принимает максимальное значение, если он распределен по нормальному закону. Значит y должен иметь нормальный закон распределения. А какой в этом случае закон распределения должен иметь входной сигнал x?

Поскольку $y = x + n$, а y и n имеют нормальные законы распределения, x и n не зависимы, согласно теории вероятностей x тоже должен иметь нормальный закон распределения.

Следовательно:

$$f(x) = \frac{1}{\sqrt{2\pi D_x}} \exp\left[-\frac{x^2}{2D}\right],$$

$$H(y) = \log_2 \sqrt{2\pi e(D_x + D_n)},$$

$$H(n) = \log_2 \sqrt{2\pi e D_n},$$

отсюда:

$$I = H(y) - H(n) = \log_2 \sqrt{1 + \frac{D_x}{D_n}} = \frac{1}{2}\log_2\left(1 + \frac{D_x}{D_n}\right)$$

Поскольку m_x и $m_n = 0, D_x = P_x, D_n = P_n$,

$$I = \frac{1}{2}\log_2\left(1 + \frac{P_x}{P_n}\right).$$

Отношение P_x / P_n мощностей сигнала и шума называется *отношением сигнал–шум*.

В результате пропускную способность непрерывного канала связи с шумом можно записать в виде:

$$C_k = F_m \log_2\left(1 + \frac{P_x}{P_n}\right) \tag{5.6}$$

Таким образом, пропускная способность этого канала определяется максимальной частотой F_m пропускаемых через канал частот и отношением сигнал–шум. Пропускная способность этого типа канала тем выше, чем выше F_m и отношение сигнал–шум.

Приведенная формула описывает наихудший в смысле величины пропускной способности случай. Если шум имеет негауссовский закон распределения или имеет коррелированные (зависимые) отсчеты, то пропускная способность канала возрастает.

Контрольные вопросы

1. Что называется технической скоростью?
2. Что называется информационной скоростью?
3. Как определяется информационная скорость для равновероятных сообщений?
4. Как определяется пропускная способность канала без помех?
5. Как влияют помехи на величину пропускной способности?
6. Что утверждает теорема Шеннона для канала связи без помех?
7. Что утверждает теорема Шеннона для канала связи с помехами?
8. От чего зависит пропускная способность дискретного канала связи с помехами? Как она определяется?

9. Чему равна пропускная способность двоичного канала связи при вероятности искажения элемента сообщения, равной 0,5?

10. От чего зависит пропускная способность непрерывного канала связи?

11. Приведите формулу для нахождения пропускной способности непрерывного канала связи.

12. Как изменяется пропускная способность непрерывного канала связи при увеличении эффективной полосы пропускания канала?

13. Какие предъявляются требования к сигналу (его статистической структуре), чтобы приблизить скорость передачи непрерывного сигнала к пропускной способности канала?

Контрольные задания

1. Символы алфавита азбуки Морзе могут появиться в сообщении свероятностями: для точки – 0,57 ; для тире – 0,29 ; для промежутка между буквами – 0,1 и для промежутка между словами – 0,04 . Определить среднее количество информации в сообщении из 700 символов данного алфавита, считая, что связь между последовательными символами отсутствует.

2. Первичный алфавит состоит из трех знаков с вероятностями 0.2, 0.7, 0.1. Для передачи по каналу без помех использовался равномерный двоичный код. Частота тактового генератора 500 Гц. Какова пропускная способность канала и скорость передачи?

3. Для передачи сообщений используется код, состоящий из трех символов, вероятности появления которых равны 0.8, 0.1 и 0.1. Корреляция между символами отсутствует. Определить избыточность кода.

Заключение

Материал, представленный в настоящем учебном пособии, не претендует на полноту освещения всех областей теории информации, теории кодирования и передачи информации по каналам связи. Цель настоящего пособия – познакомить читателя с теоретическими основами прикладной теории информации и кодирования. Для закрепления изложенного материала после каждой главы приведены контрольные вопросы и задания, решение которых позволит читателю глубже понять теорию.

В учебном пособии достаточно подробно изложены вопросы количественной оценки информации. Однако параграф семантической оценки информации нуждается в более глубокой проработке. Авторами написан ряд работ в этой области, и в дальнейшем переиздании пособия возможно восполнение этого пробела.

В главе «Кодирование информации» достаточно подробно изложены методы построения большинства известных кодов. Вопросы декодирования изложены в достаточном для данного курса объеме. Более подробно, вопросы декодирования освящены в других учебных курсах.

Вопросы передачи информации по каналам связи изложены с точки зрения теории, практической реализации уделено мало внимания в силу того, что практические основы теории передачи информации по каналам связи более подробно изучаются в рамках других учебных курсов.

Изложенные материал апробирован при чтении лекций в течение нескольких лет в Юго–Западном государственном университете. Авторы выражают признательность своим студентам, которые помогали шлифовать в дискуссиях представление изложенного в учебном пособии материала.

Рекомендательный список литературы
Основной

1. Лукьянюк, С. Г. Теория электрической связи. Сигналы, помехи и системы передачи [Текст] : учеб. пособие / С. Г. Лукьянюк, А. М. Потапенко ; Юго–Зап. гос. ун–т. – Курск, 2012. – 223 с.

2. Тихонов, В.И. Случайные процессы. Примеры и задачи [Текст]: учеб. пособие. Т.5: Оценка сигналов и их параметров и спектров. Основы теории информации / В. И. Тихонов, Б. И. Шахтарин, В. В. Сизых; под ред. В. В. Сизых. – М.: Горячая линия – Телеком, 2009. – 400 с.: ил.

3. Баскаков, С. И. Радиотехнические цепи и сигналы [Текст]: учеб. / С. И. Баскаков. – 5–е изд., стер. – М.: Высш. шк., 2005. – 462 с.

4. Дмитриев, В.И. Прикладная теория информации [Текст]: учеб. пособие / В. И. Дмитриев. – М.: Высш. шк., 1999. – 320 с.

5. Бондарев, Б. Н. Основы теории передачи сигналов [Текст]: учеб. пособие / Б. Н. Бондарев, А. А. Макаров – Новосибирск: НЭИС, 1970. – 132 с.

6. Шавенько, Н. К. Основы теории информации и кодирования [Текст]: учеб. пособие / Н. К. Шавенько – М.: Изд–во МИИГАиК, 2012. – 125 с.

7. Березкин, Е. Ф. Основы теории информации и кодирования. [Текст]: учеб. пособие / Е. Ф. Березкин – М.: МИФИ, 2010.– 312 с.

8. Лидовский, В. В. Теория информации [Текст]: учеб. пособие / В. В. Лидовский – М.: Компания Спутник+, 2004. – 111 с.

9. Мардер, Н. С. Современные телекоммуникации [Текст] / Н.С. Мардер. – М.: ИРИАС, 2006.

Дополнительный

10. Потапенко, А. М. Проблема измерения информации в задачах развития инфотелекоммуникаций [Текст] / А. М. Потапенко, А. В. Хмелевская // Информационно–измерительные и управляющие системы. 2010. Т. 8. № 7. С. 16–22.

11. Цветков, В. Я. Семантика сообщений в телекоммуникационных системах [Текст] / В. Я. Цветков // Всероссийский конкурсный отбор обзорно–аналитических статей по приоритетному

направлению «Информационно–телекоммуникационные системы» - М.: Информика, 2008. - 34 с.

12. Потапенко, А.М. Меры качества информационного обеспечения [Текст] / А.М. Потапенко, А.В. Хмелевская, А.А. Волынец // Радиотехника. – 2010. – т.8, №11 – С. 47–50.

Приложение
Некоторые полезные сведения из теории вероятностей

1. Случайные события

Событием (U) называют всякий факт, который может произойти или не произойти.

Вероятностью события $P(U)$ называется численная мера степени объективной возможности этого события.

Достоверным событием называют событие (U), которое в результате опыта непременно должно произойти. Вероятность достоверного события принимается равной единице, т.е. $P(U)=1$.

Невозможным событием называют событие Q, которое в результате опыта никоим образом не может произойти. Вероятность невозможного события принимается равной нулю, т.е. $P(Q)=0$.

Вероятность любого реального события (A) заключена в интервале от 0 до 1, т. е. $0 \le P(A) \le 1$.

Несколько событий в данном опыте называют *несовместными*, если никакие два из них не могут произойти одновременно.

Несколько событий в данном опыте называют *равновозможными*, если по условиям симметрии опыта нет оснований считать какое–либо из них более предпочтительным или возможным.

Условной вероятностью события A при наличии события B называют вероятность события A, вычисленную при условии, что событие B произошло. Условная вероятность в этом случае обозначается как $P(A/B)$.

События называются *независимыми*, если появление одного из них никоим образом не меняет вероятности появления другого. Для независимых событий A и B справедливо:

$$P(A/B) = P(A), \qquad P(B/A) = P(B).$$

Полной группой событий называют несколько событий таких, что в результате опыта непременно произойдет хотя бы одно из них. Если события A_k (k=1, 2, ...,n) составляют полную группу событий и несовместны, то

$$\sum_k P(A_k) = 1,$$

т. к. $\sum_k A_k$ – достоверное событие.

Если несколько событий образуют полную группу событий; несовместны; равновозможны, то вероятность события *A* можно вычислить по формуле

$$P(A) = \frac{m}{n},$$

где n – общее число возможных событий (исходов опыта),

m – число событий, благоприятствующих событию *A*.

2. Алгебра событий

Случайные события обозначают большими буквами алфавита.

Равенство двух событий A и B (*A=B*) означает, что появление одного события непременно влечет за собой появление другого.

Суммой двух событий A и B называют событие *C* (*C=A=B*), заключающееся в появлении хотя бы одного из событий *A* и *B*.

Произведением двух событий A и B называют событие *C*(*C=A·B*), заключающееся в одновременном наступлении обоих событий *A* и *B*.

Если результат какого–либо опыта может иметь два взаимно исключающих события *A* и *B* (полная группа событий), одно из этих событий называют *противоположным событием* другому, например, событие *B*, противоположное событию *A*, и обозначается $\breve{A}$ (читается как «не *A*»). Для прямого события *A* и противоположного события $\breve{A}$ (как составляющих полную группу событий) справедливо тождество:

$$P(A) + P(\breve{A}) = 1.$$

Теорема сложения вероятностей. Вероятность суммы двух событий A и B равна сумме вероятностей этих событий без вероятности произведения этих событий, т. е.

$$P(A + B) = P(A) + P(B) - P(A \cdot B).$$

В случае, если A и B несовместны, то $P(A \cdot B) = 0$ и, следовательно,

$$P(A + B) = P(A) + P(B).$$

Теорема умножения вероятностей. Вероятность произведения двух событий A и B равна вероятности одного из них, умноженной на условную вероятность другого при наличии первого, т.е.

$$P(A \cdot B) = P(A \cdot B) = P(A) \cdot P(B/A) = P(B) \cdot P(A/B).$$

В случае, если события A и B независимы, то справедливоследующее соотношение

$$P(A \cdot B) = P(A) \cdot P(B) = P(B) \cdot P(A).$$

Полезное правило решения задач на нахождение вероятности случайных событий заключается в следующем. Для того чтобы найти вероятность интересующего события, необходимо:

а) с помощью алгебры событий описать интересующее событие через известные события, т. е. через такие события, вероятности которых известны;

б) с помощью теорем вероятности суммы и вероятности произведения найти вероятность интересующего события.

3. Случайные величины

Случайной величиной называют величину, которая в результате опыта может принимать то или иное численное значение, не известное заранее.

Дискретной (прерывной) случайной величиной называют случайную величину, которая может принимать только конечное число различных значений.

Непрерывной случайной величиной называют случайную величину, которая может принимать бесконечное число различных значений, заполняющих какой–либо промежуток.

Законом распределения случайной величины называют всякое соотношение, устанавливающее связь между всеми возможными значениями случайной величины и соответствующими им вероятностями.

Для дискретных случайных величин, в качестве закона распределения чаще всего используют ряд распределения. Ряд распределения представляет собой таблицу, в верхней части которой размещены все возможные значения дискретной случайной величины, а в нижней части – соответствующие им вероятности.

X	x_1	x_2	....	x_k	....	x_n
$P(x_i)$	$P(x_1)$	$P(x_2)$	....	$P(x_k)$	....	$P(x_n)$

Формальным признаком правильности составления ряда распределения является выполнение условия

$$\sum_{i=1}^{n} P(x_i) = 1.$$

Для непрерывных случайных величин в качестве закона распределения часто используют функцию распределения (функция распределения может быть использована в качестве закона распределения и для дискретных величин). *Функцией распределения случайной величины X* называют функцию $F(x)$, выражающую вероятность того, что случайная величина X примет значение меньше, чем x, т. е.

$$F(x) = P(X < x).$$

Из определения функции распределения вытекают ее основные свойства:

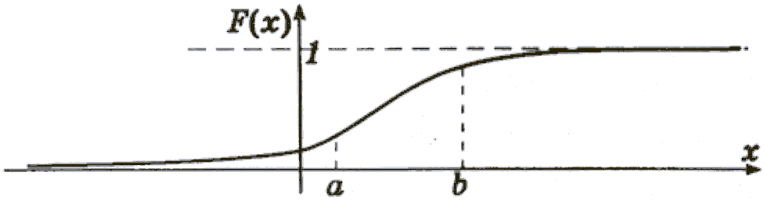
1). $0 \le F \le 1$;

2). $F\left(-\infty\right) = 0$;

3). $F\left(+\infty\right) = 1$;

4). $F(x)$ - неубывающая функция

В общем случае график функции распределения имеет вид



Из определения $F(x)$ вытекает важное соотношение, выражающее вероятность попадания случайной величины X в интервал от a до b:

$$P(a < X < b) = F(b) - F(a).$$

Наряду с функцией распределения, в качестве закона распределения непрерывных случайных величин используют и функцию плотности распределения вероятности $f(x)$, которая определяется как производная функции распределения, т. е.

$$f(x) = F'(x).$$

Из определения функции плотности распределения вероятности вытекают ее основные свойства:

1). $f\left(-\infty\right) = 0$;

2). $f\left(+\infty\right) = 0$;

3). $f\left(x\right) \ge 0$;

4). $\int\limits_{-\infty}^{+\infty} f\left(x\right)dx = 1$.

В общем случае график функции плотности распределения вероятности $f(x)$ имеет вид

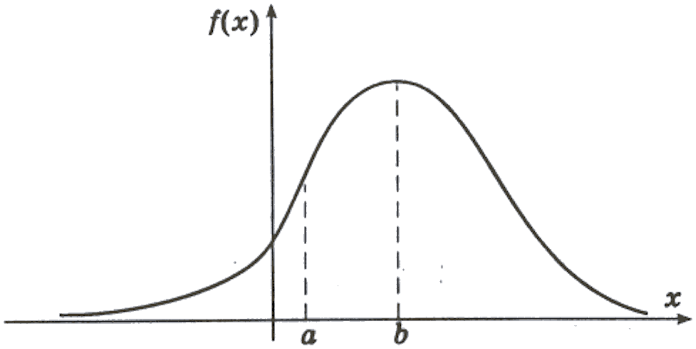


График функции $f(x)$ называют *кривой распределения.*

Элементом вероятности для случайной величины X называют величину $f(x)dx$, выражающую вероятность попадания случайной величины X в элементарный отрезок dx, примыкающий к точке x.

Вероятность попадания случайной величины X в конечный промежуток от a до b определяется соотношением:

$$P\left(a \le X \le b\right) = \int_a^b f\left(x\right)dx.$$

Можно получить выражение, связывающее функцию распределения $F(x)$ и функцию плотности распределения вероятности $f(x)$ случайной величины:

$$F\left(x\right) = \int_{-\infty}^{x} f\left(x\right)dx.$$

4. Статистические характеристики случайных величин.

Математическое ожидание случайной величины X (обозначают $M[X]$ или m_x) – это среднее значение случайной величины, вычисленное по следующим формулам:

1). Для дискретных случайных величин

$$M\left[X\right]=\sum_{i=1}^{n}x_i\cdot p\left(x_i\right);$$

2). Для непрерывных случайных величин

$$M\left[X\right]=\int_{-\infty}^{+\infty}x\cdot f\left(x\right)dx.$$

Центрированной случайной величиной $°X$ называют разность между самой случайной величиной X и ее математическим ожиданием $M[X]$, т. е.

$$X = X - M[X].$$

Дисперсия случайной величины X (обозначается $D[X]$ или d_x) — это есть математическое ожидание квадрата соответствующей ей центрированной случайной величины, т. е.

$$D[X] = M[°X^2],$$

которая вычисляется по следующим формулам:

1). Для дискретных случайных величин

$$D\left[X\right]=\sum_{i=1}^{n}\left(x_i-M\left[X\right]\right)^2\cdot p\left(x_i\right),$$

2). Для непрерывных случайных величин

$$D\left[X\right]=\int_{-\infty}^{+\infty}\left(x-M\left[X\right]\right)^2\cdot f\left(x\right)dx.$$

Дисперсия характеризует среднее отклонение значений случайной величины от её математического ожидания. Размерность дисперсии не совпадает с размерностью характеризуемой случай-

ной величины. Размерность дисперсии есть квадрат размерности соответствующей случайной величины.

Среднее квадратическое отклонение (σ) случайной величины X – это есть квадратный корень из ее дисперсии, т. е.

$$\sigma = \sqrt{D[X]}.$$

Среднее квадратическое отклонение иногда называют стандартом. Среднее квадратическое отклонение имеет ту же размерность, что и сама случайная величина.

Начальным моментом порядка k (v_k) случайной величины X называют математическое ожидание k–той степени этой случайной величины, т. е.

$$v_k = M\left[X^k\right].$$

Центральным моментом порядка k (μ_k) случайной величины X называют математическое ожидание k–той степени отклонения этой случайной величины от ее математического ожидания, т. е.

$$\mu_k = M\left[\left(X - M[X]\right)^k\right].$$

Эксцесс случайной величины E есть величина, вычисленная по формуле

$$E = \frac{\mu_4}{(\mu_2)^2} - 3.$$

Для нормального закона распределения $E = 0$, отличие эксцесса от нуля указывает на отклонение эмпирического закона распределения от нормального закона распределения.

Асимметрия характеризует симметричность кривой распределения случайной величины X. Показатель асимметрии S вычисляется по формуле:

$$S=\left(\frac{\mu_3}{(\sigma_x)^3}\right).$$

Для симметричных распределений $S = 0$.

5. Случайные функции

Случайной функцией $X(t)$ называют функцию, которая в результате опыта может принимать тот или иной конкретный вид, неизвестный заранее.

Конкретный вид, принимаемый случайной функцией, называют *реализацией* случайной функции.

Сечением случайной функции называют случайную величину $X(t_k)$, в которую обращается случайная функция $X(t)$ при фиксированном аргументе ($t = t_k$).

Одномерным законом распределения случайной функции $X(t)$ называют закон распределения $f(x,t_k)$ сечения $X(t)$ случайной функции.

Математическим ожиданием случайной функции $X(t)$ называют неслучайную функцию $m_x(t)$, которая при каждом значении аргумента t представляет собой математическое ожидание соответствующего сечения этой случайной функции.

Дисперсией случайной функции $X(t)$ называют неслучайную функцию $d_x(t)$, которая при каждом значении аргумента t представляет собой дисперсию соответствующего сечения этой случайной функции.

Корреляционной функцией случайной функции $X(t)$ называют неслучайную функцию двух аргументов $R_x(t_k,t_1)$, которая при каждой паре значений аргументов t_k и t_1 равна корреляционному моменту соответствующих сечений случайной функции, т. е.

$$R_x(t_k,t_1)=M\left[\dot{X}(t_k)\cdot\dot{X}(t_l)\right],$$

где $\dot{X}(t)=X(t)-m_x(t)$ – центрированная случайная функция.

Корреляционная функция характеризует статистическую связь между сечениями случайной функции, т. е. внутреннюю структуру случайной функции. При $t_k = t_1$ корреляционная функция обращается в дисперсию

$$D_x(t_k,t_k) = M\left[\dot{X}(t_k)\right]^2.$$

Нормированной корреляционной функцией случайной функции $X(t)$ называют неслучайную функцию двух аргументов $r_x(t_k,t_{k1})$, определяемую по формуле:

$$r_x(t_k,t_{k1}) = \frac{R_x(t_k,t_{k1})}{\sigma(t_k)\sigma(t_{k1})} = \frac{R_x(t_k,t_{k1})}{\sqrt{D(t_k)D(t_{k1})}},$$

при $t_k = t_1$, $r_x(t_k, t_{k1}) = 1$.

Стационарной случайной функцией называют случайную функцию, математическое ожидание которой постоянно $m_x(t) = const$, а её корреляционная функция зависит только от разности между аргументами:

$$R_x(t_1,t_2) = R_x(\tau),$$

где $\tau = t_1 - t_2$.

Учебное издание

Хмелевская Алёна Валентиновна
Потапенко Александр Михайлович

ОСНОВЫ ТЕОРИИ ИНФОРМАЦИИ И КОДИРОВАНИЯ

Учебное пособие

Редактор Н. В. Комардина
Компьютерная верстка и макет А. В. Хмелевская

Подписано в печать Формат 60×84 1/16
Усл. печ. л. 14. Уч.-изд. л. 12,68. Тираж экз. Заказ .
Юго-западный государственный университет
305040, г. Курск, ул. 50 лет Октября, 94
Отпечатано в ЮЗГУ.