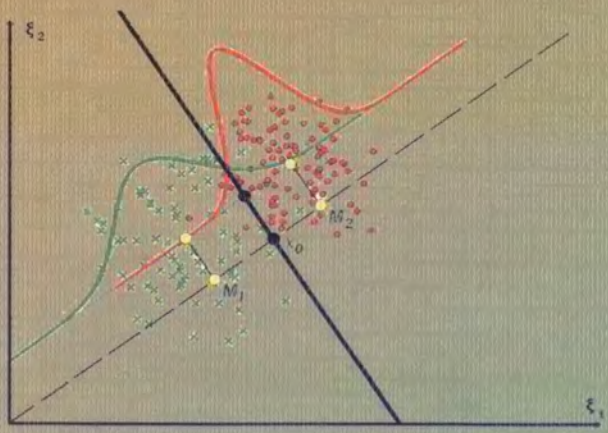
Д.В. Ломакин Л.С. Ломакина А.С. Пожидаева

ВЕРОЯТНОСТЬ. ИНФОРМАЦИЯ. КЛАССИФИКАЦИЯ



Нижний Новгород 2014

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ
УЧРЕЖДЕНИЕ ВЫСШЕГО ПРОФЕССИОНАЛЬНОГО ОБРАЗОВАНИЯ
«НИЖЕГОРОДСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ
им Р.Е. АЛЕКСЕЕВА»

Д.В. Ломакин, Л.С. Ломакина, А.С. Пожидаева

ВЕРОЯТНОСТЬ. ИНФОРМАЦИЯ. КЛАССИФИКАЦИЯ

Рекомендовано Ученым советом Нижегородского государственного технического университета им. Р. Е. Алексеева в качестве учебного пособия для студентов высших учебных заведений, обучающихся по направлению подготовки 230100 «Информатика и вычислительная техника» и 230400 «Информационные системы и технологии»

Нижний Новгород 2014

УДК 519.2
Л 74

Рецензент
доцент кафедры теоретической механики
Нижегородского государственного университета им. Н.И. Лобачевского,
кандидат физико-математических наук *А.Ф. Ляхов*

Д.В. Ломакин, Л.С. Ломакина, А.С. Пожидаева

Л74 Вероятность. Информация. Классификация: учеб. пособие / Д.В. Ломакин, Л.С. Ломакина, А.С. Пожидаева; Нижегор. гос. техн. ун-т им. Р.Е. Алексеева. – Н. Новгород, 2014. – 128 с.

ISBN 978-5-502-00480-0

Рассматриваются базовые понятия теории вероятностей, теории информации и использование вероятностных и информационных методов в задачах диагностики сложных систем и в задачах обработки многомерных данных на примере классификации состояний биоценоза.

Учебное пособие предназначено для студентов, обучающихся по направлениям: 230100 «Информатика и вычислительная техника» и 230400 «Информационные системы и технологии»

Рис. 42. Табл. 6. Библиогр.: 10 назв.

УДК 519.2

ISBN 978-5-502-00480-0

ОГЛАВЛЕНИЕ

ВВЕДЕНИЕ

В настоящее время широко используется системный подход к решению задач анализа и синтеза объектов и процессов различной физической природы. Объект описывается как система, т.е. как структурированный состав, при этом свойства объекта определяются свойствами построенной системы, которая выполняет функцию модели объекта при решении поставленной задачи. Исследование свойств модели, моделирование свойств с использованием современных информационных технологий, и синтез на основе результатов моделирования новых объектов, процессов и концепций являются основной частью научной и прикладной деятельности человека.

Описать состояние объекта как некоторой целостности минимальным количеством переменных (параметров, свойств) на ранних стадиях его изучения, как правило, не представляется возможным, не всегда удается даже определить их возможное количество. Поэтому на первом этапе возникает проблема с выделением наиболее информативной совокупности переменных, на основании которой можно было бы решить поставленную задачу, которая описывается заданной целевой функцией.

Кроме наблюдаемых переменных существуют еще скрытые переменные (компоненты), которые отражают структурные свойства объекта, законы, определяющие форму организации объекта. Совокупность значений наблюдаемых переменных называется многомерными данными в пространстве переменных.

К настоящему времени сформировалось несколько методов обработки многомерных данных, каждый из которых решает частную задачу. В настоящем пособии приведен обзор методов и подробно рассматривается метод классификации на примере анализа состояний биоценоза.

Для решения задач, связанных с обработкой многомерных данных, требуется соответствующий инструментарий, функции которого могут выполнять методы теории вероятностей, теории информации и математической статистики. Поэтому в пособии подробно описан их понятийный аппарат и способы решения конкретных задач.

В предлагаемом учебном пособии рассмотрены современные методы обработки многомерных данных с целью выявления скрытых параметров, которые характеризуют совокупность многомерных данных как некоторую целостность. Приведен обзор задач и методов их решения, которые сформировались к настоящему времени. Подробно рассмотрен метод проекций на примере классификации состояний биоценоза. В качестве инструментария используются классические методы теории вероятностей, теории информации и математической статистики, которые подробно изложены в соответствующих разделах.

1. ВЕРОЯТНОСТЬ

1.1. Основные понятия

Каждая наука начинается с определения объекта и предмета исследования, построения модели объекта, системы аксиом и с формирования основных понятий, на которых она базируется [1, 3, 4, 5, 7]. Поскольку определить понятие – это значит свести его к другим, более известным понятиям, то, очевидно, процесс должен где-то закончиться. Поэтому всегда существуют первичные понятия, которые строго не определяются, а только поясняются. Одним из таких понятий является понятие события.

В теории вероятностей объект исследования - это случайные явления различной физической природы.

Предметом теории вероятностей является математический анализ случайных явлений, выявление закономерностей в самих случайных явлениях независимо от их конкретной природы.

Событие – это любой факт, который может произойти при заданном комплексе условий D.

Достоверное событие – это событие, которое всегда происходит при заданном комплексе условий D.

Невозможное событие - это событие, которое никогда не происходит при заданном комплексе условий D.

Случайное событие – это событие, которое может произойти, а может и не произойти при заданном комплексе условий D.

Следует отметить, что достоверное, невозможное и случайное события остаются таковыми только при заданном комплексе условий D.

Комплекс условий D – это совокупность контролируемых физических величин или параметров, которые описывают эксперимент, испытание, опыт и т.д. Задать комплекс условий – это значит задать значения указанных физических величин или параметров.

Эмпирическим основанием для построения теории вероятностей послужила устойчивость относительной частоты появления события. Если при n испытаниях событие A появилось $n(A)$ раз, то его относительная частота появления равна отношению $n(A)/n$. Это свойство относительной частоты выражено в (одной из основных) аксиоме Колмогорова, согласно которой *вероятность* – это число p $(0 \le p \le 1)$, которое поставлено в соответствие данному событию. Значение вероятности иногда называют вероятностной мерой или весом события. Вероятность можно интерпретировать как степень возможности появления события.

Модель, лежащая в основе теории вероятностей, - это пространство элементарных событий, которое по определению представляет собой полную группу несовместных событий (исходов данного опыта, эксперимента) с заданной вероятностной мерой (законом распределения вероятностей).

События называются несовместными, если наступление одного из них исключает возможность наступления другого.

События образуют полную группу событий, если $\sum_{i=1}^{n} p_i = 1$, где p_i – вероятность i– го события, т.е. вероятность появления события, которое не принадлежит данной группе, равна нулю.

Математическая модель события (в отличие от приведенного ранее пояснения физического смысла события) – это любое подмножество в пространстве элементарных событий, чаще всего объединенных в подмножество по тому или иному свойству, признаку.

В теории вероятностей не исследуются причины, по которым события появляются с той или иной вероятностью. Основной ее задачей является разработка методов вычисления вероятностей, если известны вероятности элементарных событий или вероятности некоторых исходных событий.

Можно выделить четыре этапа вычисления вероятности события:

- построение пространства элементарных событий, которое определяется комплексом условий D в данной задаче;
- выделение подмножества, т.е. события, вероятность которого необходимо вычислить по условию задачи, и событий, которые участвуют в решении задачи;
- вычисление вероятностей элементарных событий, которые входят в выделенное подмножество;
- вычисление вероятности выделенного события как суммы вероятностей всех образующих его элементарных событий.

Указанные этапы желательно представлять при решении любой задачи, но это не значит, что нужно скрупулезно следовать им.

Разработанные в теории вероятностей методы (теоремы) позволяют найти более короткие способы вычисления вероятностей по сравнению с указанным общим методом.

1.2. Классический метод вычисления вероятностей

Классический метод вычисления вероятностей применим, когда все элементарные события равновероятны. В этом случае вероятность $P(\omega_i)$ отдельного элементарного события ω_i равна

$$p(\omega_i) = 1/N,$$

а вероятность события A равна

$$p(A) = m/N,$$

где N - общее количество элементарных событий, мощность пространства элементарных событий; m - количество элементарных событий, образующих событие A. События, образующие событие A, называются благоприятствующими для появления события A. Вероятность $p(A)$ можно интерпретировать как вес события A по отношению к весу всего пространства, причем $p(A)$ определяется только количеством благоприятствующих событий и не зависит от их состава.

Задача. Всего в урне N шаров, в том числе M - белых. Чему равна вероятность вынуть белый шар?

Решение. Пространство элементарных событий в данном случае представляет собой множество всех возможных исходов эксперимента, которое совпадает с множеством шаров в урне, если все шары считать различными, при этом количество благоприятствующих событий равно M. Поскольку все исходы можно считать равновероятными, то вероятность вынуть белый шар равна M/N.

Задача. Какова вероятность появления четного числа при бросании игральной кости (кубика)?

Решение. Пространство элементарных событий состоит из 6 равновероятных событий, причем 3 из них четные числа, следовательно, искомая вероятность равна 3/6.

Задача. Вычислить вероятность (события A) того, что при бросании двух игральных костей сумма выпавших на них цифр будет равна 5.

Решение. В данном случае результатом опыта, который состоит из двух испытаний, будут две цифры (x, y) соответственно на первом и втором кубиках. Этой упорядоченной паре цифр можно поставить в соответствие точку на плоскости. Легко проверить, что все 36 точек образуют пространство элементарных событий, причем это пространство представляет собой декартово произведение пространств соответственно для первого и второго кубиков. В пространстве элементарных событий можно выделить 4 события (точки), обладающие общим свойством, а именно, с суммой координат, равной 5 $(x + y = 5)$. Эти элементарные события образуют событие A, вероятность которого, очевидно, равна $4/36$.

1.3. Вычисление условной вероятности

Вероятность $P(A)$ события A при заданном комплексе условий D называется полной или безусловной вероятностью и может изменяться только при изменении комплекса условий. Если некоторое событие B так или иначе влияет на комплекс условий, накладывает на него ограничения, то вводится понятие условной вероятности $p(A/B)$ события A при условии B.

При вычислении условной вероятности следует обратить особое внимание на построение пространства элементарных событий. Оно должно содержать в себе события A и B как некоторые подмножества (рис 1.1). В этом случае условная вероятность

$$p(A/B) = p(AB)/p(B)$$

определяется на подмножестве B как новом пространстве элементарных событий, в котором событие A существует в виде пересечения AB событий A и B, при этом элементарные события в B образуют полную группу событий благодаря делению значений их вероятностей на $p(B)$. Условную вероятность $p(A/B)$ можно интерпретировать как вес события AB по отношению к весу события B. В частности, событие B может совпасть со всем пространством Ω ($B = \Omega, p(B) = 1$). В этом случае вероятность события A может зависеть от события B, так как реализация события B может привести к изменению закона распределения вероятностей элементарных событий и благоприятствующих для A в том числе.

Пример. Пусть произошло событие B, состоящее в том, что из двух возможных кубиков со смещенным центром массы и несмещенным случайным образом выбран кубик со смещенным центром массы. В этом случае событие B совпадает со всем пространством, а вероятность события A в B ($B = \Omega$) будет определяться законом распределения вероятностей появления цифр при выбранном кубике. В принципе, все события являются условными, так как появляются при том или ином комплексе условий.

Если все события в подмножестве B равновероятны, то применим классический метод вычисления вероятностей, который сводится к вычислению отношения количества элементарных событий в AB (благоприятствующих для события A в подмножестве B) к количеству элементарных событий в подмножестве B.

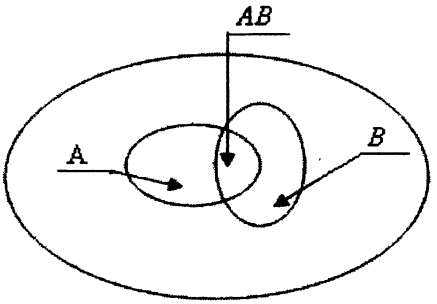


Рис. 1.1. Пространство элементарных событий

Например, если по условию задачи кубик бросают два раза, то исходом опыта будут две цифры, которые можно рассматривать как координаты точки на плоскости. В этом случае пространство состоит из 36 точек, а не из 6, как иногда ошибочно считают.

В данном случае опыт состоит из двух испытаний (бросаний кубика) и пространство элементарных событий (исходов опыта) представляет собой декартово произведение пространств отдельных испытаний.

Опыт может содержать произвольное количество испытаний. Этой терминологии мы будем придерживаться и в дальнейшем.

Задача. Студент выучил 10 билетов из 25. В каком случае вероятность вынуть выученный билет больше, когда студент вынимает билет первым или вторым (билеты не возвращаются)?

Решение. Результатом опыта являются два вынутых билета, т. е. их номера. Не нарушая общности решения, выученными билетами будем считать первые 10.

Исходы первого испытания будем откладывать по оси x, а второго - по оси y (рис. 1.2).

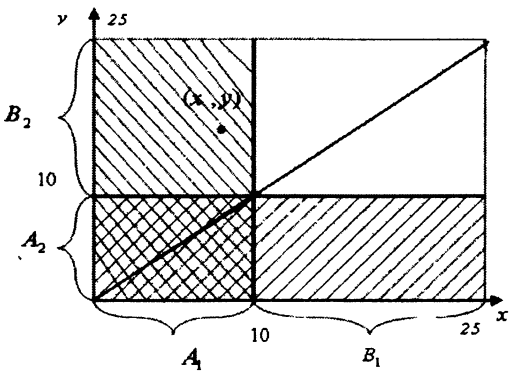


Рис. 1.2. Пространство элементарных событий

Точка с координатами (x, y) соответствует исходу опыта. Поскольку билеты не возвращаются, то события, которым соответствуют точки, лежащие на диагонали, являются невозможными и их следует исключить из пространства элементарных событий. Все остальные события, количество которых равно 25*25-25=25*24, можно считать равновероятными, и поэтому применим классический метод вычисления вероятности. Тогда вероятность события A_1, состоящего в том, что при первом испытании будет вынут выученный билет, равна $p(A_1) = (10*25-10)/(25*24) = 10/25$.

Аналогично вероятность события A_2, состоящего в том, что второй вынутый билет окажется выученным, равна

$$p(A_2) = (10*25-10)/(25*24) = 10/25.$$

Следовательно, $p(A_1) = p(A_2)$.

Кроме событий A_1 и A_2, в данном пространстве элементарных событий можно выделить много других событий и вычислить их вероятности. Например, вероятности событий B_1 (первый вынутый билет оказался не выученным) и B_2 (второй вынутый билет оказался невыученным) равны $p(B_1) = (15*25-15)/(25*24) = 15/25$, $p(B_2) = (15*25-15)/(25*24) = 15/25$.

Вероятность произведения событий A_1A_2 равна

$$p(A_1A_2) = (10*10-10)/(25*24) = 3/20.$$

Условные вероятности:

$$p(A_1/A_2) = (10*10-10)/(10*25-10) = 9/24 \text{ и}$$

$$p(A_2/A_1) = (10*10-10)/(10*25-10) = 9/24,$$

где в числителе стоит количество элементарных событий, которые входят в пересечение событий A_1 и A_2, а в знаменателе - количество элементарных событий, которые входят соответственно в A_2 и A_1. Поясним физический смысл условных вероятностей: $p(A_2/A_1)$ - это вероятность того, что при втором испытании появится выученный билет при условии, что первым был также выученный билет. Событие A_1 изменило комплекс условий, при котором совершается второе испытание; $p(A_1/A_2)$ - это вероятность того, что билет, вынутый первым, был выученным при условии, что второй билет оказался выученным, при этом предполагается, что мы не смотрели, каким был билет, вынутый первым. Вычислите условные вероятности для других событий.

Задача. Кубик бросают два раза. С какой вероятностью при первом испытании появится единица (событие A) при условии, что при втором испытании на кубике выпало значение больше, чем при первом (событие B).

Решение. Применим классический метод вычисления вероятностей, поскольку исходы опыта образуют пространство равновероятных элементарных событий (рис. 1.3).

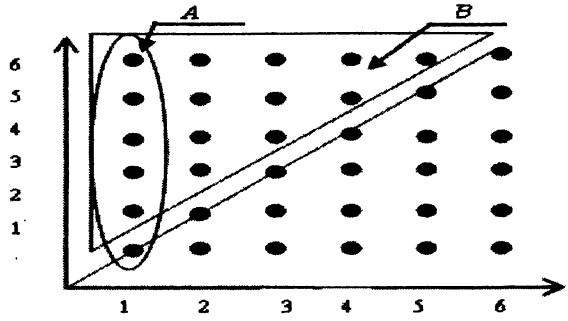


Рис. 1.3. Пространство элементариых событий

Необходимо вычислить условную вероятность

$$p(A/B) = n(AB)/n(B),$$

где $n(AB)$ - количество точек, входящих в пересечение событий A и B, равное 5; $n(B)$ - количество точек, входящих в подмножество B, равное 15. Подмножество B образовано точками, лежащими строго выше диагонали. Следовательно, $p(A/B) = 5/15 = 1/3$.

1.4. Формула полной вероятности и формула Байеса

В принципе, все вероятности событий являются условными, поскольку все события происходят при том или ином комплексе условий. Тем не менее, любой комплекс условий, который реализуется в данном эксперименте, можно считать полным начальным, т.е. без каких-либо ограничений, а соответствующие вероятности событий - полными или безусловными.

Часто полный комплекс условий можно представить как совокупность частных комплексов условий, которые образуют некоторую вероятностную структуру. Тогда вероятность некоторого события A при частном комплексе условий можно назвать частной или условной вероятностью.

Рассмотрим следующий способ вероятностной организации совокупности комплексов условий. Пусть имеется n комплексов условий, ка-

ждый из которых реализуется в данном эксперименте с некоторой вероятностью $p_i, i = 1...n$. Реализацию некоторого частного комплекса условий будем интерпретировать как появление события $B_i, i = 1..n$ с вероятностью, равной $p(B_i) = p_i$. Все события $B_i, i = 1..n$ образуют полную группу несовместных событий. Требуется вычислить полную вероятность события A, которое может наступить лишь при появлении одного из событий $B_i, i = 1..n$ с известной условной вероятностью $p(A/B_i)$ и при известных вероятностях $p(B_i)$. Для событий A и B_i можно записать формулу умножения вероятностей в виде

$$p(AB_i) = p(B_i)p(A/B_i) = p(A)p(B_i/A).$$

Тогда полная вероятность

$$p(A) = \sum_{i=1}^{n} p(A, B_i) = \sum_{i=1}^{n} p(B_i)p(A/B_i)$$

получается в результате суммирования двухмерного закона распределения $p(AB_i)$ по всем событиям $B_i, i = 1..n$, которые требуется исключить, т.е. понизить размерность распределения. Полученное выражение для вычисления вероятности $p(A)$ называется формулой полной вероятности, геометрическая интерпретация которой представлена в виде вероятностной диаграммы (рис. 1.4). Вероятность $p(A)$ равна сумме произведений вероятностей p_i на условные вероятности $p(A/B_i), i = 1...n$.

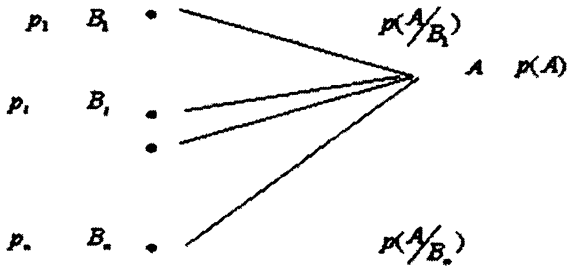


Рис. 1.4. Вероятностная диаграмма

Из формулы умножения вероятностей следует, что

$$p(B_i/A) = p(B_i)p(A/B_i)/p(A).$$

Полученное выражение называется формулой Байеса, где $p(A)$ вычисляется по формуле полной вероятности.

Формула Байеса позволяет вычислить вероятность события B_i при условии, что появилось событие A. В этом случае события B_i называются гипотезами и, как правило, обозначаются через H_i ($B_i = H_i$). Можно дать следующую интерпретацию формулы Байеса. В результате опыта реализуется ненаблюдаемое событие B_i с априорной (доопытной) вероятностью, равной $p(B_i)$, и наблюдаемое событие A, которое доставляет некоторое количество информации о реализованном событии B_i. На основании полученной информации вероятности $p(B_i)$ могут быть переоценены по формуле Байеса, т.е. может быть вычислена апостериорная (послеопытная) вероятность $p(B_i/A)$ события B_i.

Задача. Выше была задача про студента, который выучил 10 билетов из 25. Требовалось определить, в каком случае вероятность вынуть выученный билет больше, когда студент вынимает билет первым или вторым (билеты не возвращаются)? Задача была решена общим стандартным методом с построением полного пространства элементарных событий (рис. 1.2). Однако ее можно решить и с использованием формулы полной вероятности. Вероятностная диаграмма для этой задачи изображена на рис.5, где события A_1 и A_2 состоят в том, что был вынут выученный билет соответственно при первом и втором вынимании билета; событие B_1 - вынут невыученный билет при первом вынимании.

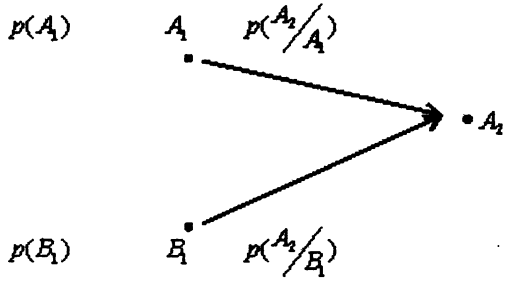


Рис. 1.5. Вероятностная диаграмма

События A_1 и B_1 образуют полную группу несовместных событий ($p(A_1) + p(B_1) = 1$). Вероятность $p(A_1) = 10/25$ вычисляется классическим методом, а вероятность $p(B_1) = 1 - 10/25 = 15/25$. Условные вероятности вычисляются следующим образом. Появление события A_1 изменяет комплекс условий, при котором наступает событие A_2, а именно: количество

выученных билетов уменьшается до 9, а общее количество билетов уменьшается до 24, отсюда $p(A_2/A_1) = 9/24$. Аналогично вычисляется вероятность $p(A_2/B_1) = 10/24$. По формуле полной вероятности вероятность $p(A_2) = 10/25 * 9/24 + 15/25 * 10/24 = 10/25$. Отсюда следует, что $p(A_1) = p(A_2)$.

Задача. Выше была решена классическим методом следующая задача. Кубик бросают два раза. С какой вероятностью при первом испытании появится единица (событие A) при условии, что при втором испытании выпало значение больше, чем при первом (событие B).

Пространство элементарных событий для этой задачи изображено на рис. 1.6. Эту задачу можно решить по формуле Байеса без построения полного пространства элементарных событий. В этом случае следует использовать вероятностную диаграмму.

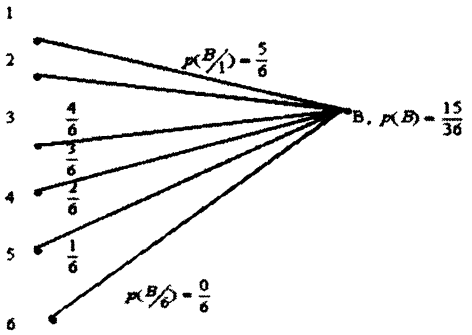


Рис.1.6. Вероятностная диаграмма

Слева изображены цифры, образующие пространство элементарных событий для первого кубика. Все события равновероятны ($p_i = 1/6, i = 1...6$). Условная вероятность $p(B/1) = 5/6$ события B вычисляется как вероятность того, что при втором испытании выпадет значение больше 1. Аналогично вычисляются остальные условные вероятности. По формуле полной вероятности находим:

$$p(B) = \frac{1}{6} \cdot (\frac{5}{6} + \frac{4}{6} + \frac{3}{6} + \frac{2}{6} + \frac{1}{6} + \frac{0}{6}) = \frac{15}{36}.$$

Условная вероятность $p(1/B)$ вычисляется по формуле Байеса:

$$p(1/B) = \frac{1/6 \; 5/6}{15/36} = \frac{1}{3}.$$

Задача. Рассмотрим пример оценки условной вероятности в случае непрерывной случайной величины. Мишень в виде круга радиуса R можно рассматривать как пространство элементарных событий Ω, если вероятность попасть в мишень принять равной единице (полная группа событий). Кроме этого, события можно считать несовместными, если размеры пули считать бесконечно малой величиной.

Стрелок делает n выстрелов, целясь в центр мишени, при этом пули будут распределены по всей мишени с некоторой плотностью, которую можно измерять количеством пуль (или весом пуль, поскольку все пули имеют одинаковый вес), приходящим на единицу площади. Выделим в мишени две фигуры (события) A и B, и оценим вероятности их поражения. Очевидно, $p(A) \approx n(A)/n$, $p(B) \approx n(B)/n$, $p(AB) \approx n(AB)/n$, где $n(A), n(B), n(AB)$ - количество пуль, попавших соответственно в A, в B и в AB - пересечение событий A и B. В частности, значение вероятности можно интерпретировать как вес соответствующего события по отношению к весу всего пространства Ω, равному n.

Кроме этого, можно ввести условную вероятность $p(A/B) \approx n(AB)/n(B)$, т.е. вес события A, которое появляется вместе с B (вес пересечения AB) по отношению к весу события B. Аналогично $p(B/A) \approx n(AB)/n(A)$.

Для условной вероятности $p(A/B)$ подмножество B является пространством элементарных событий с плотностью распределения вероятностей, равной $\omega(x,y)/p(B)$, где $\omega(x,y)$ - плотность распределения вероятностей в пространстве Ω. Благодаря делению на $p(B)$ подмножество B становится полной группой событий. Устойчивость относительной частоты $n(AB)/n(B)$ появления события A при условии B является эмпирическим основанием для введения по аксиоме Колмогорова понятия условной вероятности $p(A/B)$. Поскольку $n(AB)/n(B) = \frac{n(AB)/n}{n(B)/n}$, то, заменяя относительные частоты соответствующими вероятностями, получим

$$p(A/B) = p(AB)/p(B), \quad (p(B) \neq 0)$$

и аналогично

$$p(B/A) = p(AB)/p(A), \quad (p(A) \neq 0).$$

Отсюда очевидной становится формула умножения вероятностей: $p(AB) = p(A)P(B/A) = p(B)p(A/B)$.

Если имеет место равномерный закон распределения вероятностей в подмножестве B, то значение вероятности $p(A/B)$ можно вычислить как отношение площади пересечения событий A и B к площади B.

Задача. Известно, что в результате n испытаний событие C появилось один раз. Какова вероятность того, что оно появилось при втором испытании? Вероятность появления события C при отдельном испытании равна p.

Решение. Пространство элементарных событий для одного испытания состоит из событий C и $\overline{C}$, которые в дальнейшем заменим соответственно на 1 и 0, а пространство элементарных событий для опыта состоит из 2^n последовательностей. Необходимо вычислить условную вероятность $p(A/B)$, где A - событие, состоящее в том, что в результате опыта появится последовательность, содержащая единицу на втором месте. Это подмножество последовательностей, каждая из которых содержит 1 на втором месте. B - событие, состоящее в том, что последовательность будет содержать одну единицу. Пересечение событий A и B состоит из единственной последовательности 010000...0. Все n последовательностей в B равновероятны, поскольку вероятность каждой из них равна $p(1-P)^{n-1}$, так как испытания независимы. Поэтому применим классический метод вычисления вероятностей, согласно которому $p(A/B) = 1/n$.

1.5. Выборка без возвращения и с возвращением

При классическом способе вычисления вероятности, когда все элементарные события равновероятны, широко используется комбинаторика. Мы будем использовать комбинаторные понятия размещения, перестановки и сочетания.

Размещением из N элементов множества X по n элементам (местам) (коротко, размещением из N по n) назовем любой упорядоченный набор из n элементов множества X. Два размещения равны тогда и только тогда, когда равны элементы в соответствующих разрядах (позициях).

Сочетанием из N элементов множества X по n назовем любое подмножество, содержащее n элементов множества X, при этом сочетания различаются только составом.

Количество размещений A_N^n из N по n при $n=1$ равно N, при $n=2$ равно $N(N-1)$, поскольку на втором месте в последовательности (в размещении) может находиться только один из $N-1$ оставшихся элементов. Таким образом, по индукции $A_N^n = N(N-1)(N-2)...(N-n+1)$.

Количество размещений A_N^n при $n=N$ равно $N!$, т.е. равно количеству перестановок, которые можно получить из N элементов множества X.

Таким образом, пространство элементарных событий состоит из A_N^n размещений.

Количество сочетаний C_N^n из N по n можно вычислить следующим образом. Множество (пространство элементарных событий) всех последовательностей (размещений) с длиной, равной n, можно разбить на непересекающиеся подмножества, каждое из которых содержит последовательности, различающиеся только перестановкой элементов и не различающиеся составом (рис. 1.7).

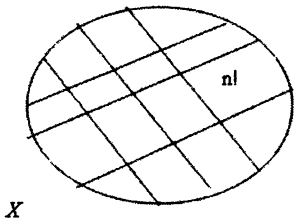


Рис. 1.7. Пространство элементарных событий

Каждое из подмножеств содержит $n!$ последовательностей. Поэтому количество сочетаний C_N^n равно количеству подмножеств в пространстве элементарных событий: $C_N^n = A_N^n / n! = N!/n!(N-n)!$.

Задача. Всего в урне находится N шаров, среди которых M белых и N-M черных. Какова вероятность того, что среди n вынутых без возвращения шаров белых окажется ровно m?

Решение. Все возможные размещения из N шаров по n местам, количество которых равно A_N^n, образуют пространство элементарных событий. Вычислим количество благоприятствующих событий, т.е. количе-

ство последовательностей (размещений), каждая из которых содержит m белых шаров.

1. Если все белые шары в последовательности считать неразличимыми, то количество способов, которыми можно получить все различающиеся между собой последовательности, равно количеству способов C_n^m, которыми можно выбрать m мест (подмножеств) для размещения белых шаров в последовательности с длиной, равной n. При этом состав шаров в последовательности не меняется, а меняется только состав выбранных мест для размещения белых шаров.

2. Если все белые шары в урне считать различными, то последовательности могут различаться в m выбранных для белых шаров местах A_M^m способами.

3. Следовательно, если все черные шары в урне считать различными, то последовательности могут различаться в n-m местах A_{N-M}^{n-m} способами. Тогда общее количество благоприятствующих последовательностей будет равно $C_n^m A_M^m A_{N-M}^{n-m}$, а вероятность появления m белых шаров среди n вынутых будет равна $p_n(m) = C_n^m A_M^m A_{N-M}^{n-m} / A_N^n$, где A_N^n - количество элементарных событий в вероятностном пространстве. Если в последнем равенстве сделать подстановку $C_n^m = n!/m!(n-m)!$ и учесть, что $\frac{A_M^m}{m!} A_M^m / m! = C_M^m$, $A_{N-M}^{n-m} / (n-m)! = C_{N-M}^{n-m}$, $A_N^n / n! = C_N^n$, то получим $p_n(m) = C_M^m C_{N-M}^{n-m} / C_N^n$.

Эту задачу можно решить, используя вероятностное пространство более высокого иерархического уровня, когда все последовательности с одинаковым составом рассматриваются как одно событие, т.е. образуют некоторое подмножество в пространстве элементарных событий.

Результат опыта состоит из m белых шаров и n-m – черных. Результаты опыта различаются составом белых шаров и составом черных шаров. Состав – это подмножество из m белых шаров, выбранных из M белых шаров, находящихся в урне, и аналогично подмножество из n-m черных шаров, выбранных из N-M черных шаров, находящихся в урне. Количество таких подмножеств соответственно равно C_M^m и C_{N-M}^{n-m}.

Общее количество возможных результатов опыта определяется как произведение $C_M^m C_{N-M}^{n-m}$.

В данном случае все указанные результаты опыта образуют подмножество благоприятствующих событий в пространстве элементарных событий, а все пространство содержит C_N^n исходов. Отсюда искомая вероятность будет равна $p_n(m) = C_M^m C_{N-M}^{n-m} / C_N^n$.

Это решение демонстрирует, как можно решать задачи с использованием вероятностных пространств на разных иерархических уровнях, но при этом нужно иметь в виду, что, продвигаясь вверх по иерархическим уровням, можно потерять часть информации, необходимой для решения конкретной задачи.

Задача. Вычислить вероятность того, что из урны, содержащей M белых шаров и $N-M$ чёрных, будет вынута заданная последовательность шаров. Например, белый, черный, белый (n=3).

Решение. В общем случае последовательность вынутых шаров можно записать как последовательность событий $A_1, A_2, ..., A_i, ..., A_n$, где

$$A_i = \begin{cases} 0, \textit{если } 0 \textit{ поставить в соответствие белому шару} \\ 1, \textit{если } 1 \textit{ поставить в соответствие черному шару} \end{cases}$$

Вероятность p (A_1, A_2,..., A_n) произведения (последовательности) в общем случае зависимых событий $A_1, A_2, ..., A_i, ..., A_n$ равна $p(A_1, A_2, ..., A_n) = p(A_1)p(A_2/A_1)p(A_3/A_2, A_1)...p(A_n/A_{n-1}, A_1)$, причем результат произведения событий и вероятность $p(A_1, A_2, ...A_n)$ не зависят от порядка сомножителей $A_1, A_2, ...A_n$, выбор которого иногда позволяет значительно упростить процедуру вычислений. Такая запись позволяет интерпретировать произведение событий как их последовательную реализацию, причем вероятность очередного события зависит от всех предыдущих реализованных событий. Все события, которые были реализованы перед очередным событием, изменяют комплекс условий, при котором появляется очередное событие, и тем самым влияют на вероятность его появления. В данном случае результат опыта можно записать в виде последовательности 010 ($n=3$), вероятность которой можно вычислить как вероятность произведения событий:

$A_1 = 0$, $A_2 = 1$, $A_3 = 0$, т.е. $p(010) = p(0)p(1/0)p(0/0,1)$, где

$p(0) = M/N$, $p(1/0) = N-M/N-1$, $p(0/1,0) = M-1/N-2$.

$$\text{Отсюда } p(010) = \frac{M}{N}\frac{N-M}{N-1}\frac{M-1}{N-2}.$$

Дадим интерпретацию решения этой задачи в пространстве элементарных событий. Поскольку все шары в урне мы считаем различными, то в пространстве элементарных событий последовательности 010 будет соответствовать подмножество последовательностей, которые различаются составом и порядком как белых, так и черных шаров в отдельности.

Белые шары будут различаться A_M^m способами, а черные - A_{N-M}^{n-m} способами. Общее количество последовательностей, которые являются бла-

гоприятствующими для события 010, будет равно $A_M^m \cdot A_{N-M}^{n-m}$, а вероятность появления заданной последовательности будет равна $A_M^m \cdot A_{N-M}^{n-m} / A_N^n$.

При $n = 3$ и $m = 2$ получим

$$P(010) = A_M^2 \cdot A_{N-M}^1 / A_N^3 = M(M-1)(N-M)/N(N-1)(N-2),$$

что совпадает с полученным ранее результатом. Следует отметить, что вероятность не зависит от порядка расположения нулей и единиц в заданной последовательности.

1.6. Нелинейное преобразование случайных величин

Пусть характеристика нелинейного элемента (НЭ) вход/выход задана зависимостью $\eta = g(\xi)$ имеющей однозначную обратную зависимость $\xi = g^{-1}(\eta) = h(\eta)$.

Вероятность того, что реализация x случайной величины ξ попадает в интервал $[\xi, \xi + d\xi]$ равна вероятности того, что реализация случайной величины η попадает в интервал $[\eta, \eta + d\eta]$, т.е. при преобразовании сохраняется вероятностная мера событий:

$$\begin{cases} p(\xi \le x \le \xi + d\xi) = \omega(\xi)d\xi \\ p(\eta \le y \le \eta + d\eta) = \omega(\eta)d\eta \end{cases}$$

$$\text{и} \quad \omega(\xi)d\xi = \omega(\eta)d\eta$$

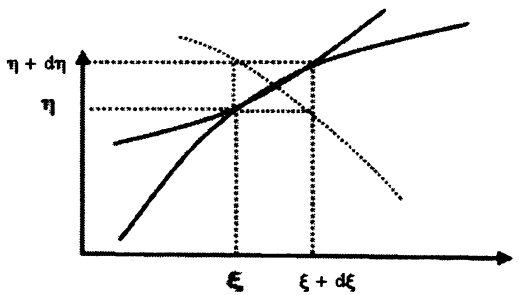


Рис. 1.8. Нелинейное преобразование $\eta = g(\xi)$

Отсюда $\omega(\eta) = \omega\big(h(\eta)\big) * \left|\frac{d\xi}{d\eta}\right| = \omega\big(h(\eta)\big)\left|\frac{dh(\eta)}{d\eta}\right|$.

В последнем используется значение модуля производной, поскольку результат не зависит от того, каким образом одна область отображается в другую.

Аналогично решается задача преобразования многомерных плотностей. Пусть известна n-мерная плотность $\omega_n(\xi_1 \dots \xi_n)$ случайных величин $\xi_1, \xi_2 \dots . \xi_n$ и нужно найти плотность $\omega_n(\eta_1 \dots \eta_n)$ для случайных величин

$$\eta_1 = g_1(\xi_1 \dots \xi_n)$$
$$\dots\dots\dots\dots\dots\dots\dots\dots\dots$$
$$\eta_n = g_n(\xi_1 \dots \xi_n),$$

где функции (нелинейные преобразования) g_i – кусочно-непрерывные функции. Предполагается, что существуют однозначные обратные функции

$$\xi_1 = h_1(\eta_1 \dots \eta_n)$$
$$\dots\dots\dots\dots\dots\dots$$
$$\xi_n = h_n(\eta_1 \dots \eta_n)$$

При попадании вектора $\vec{\xi}$ в область D_ξ

$$D_\eta = g(D_\xi).$$

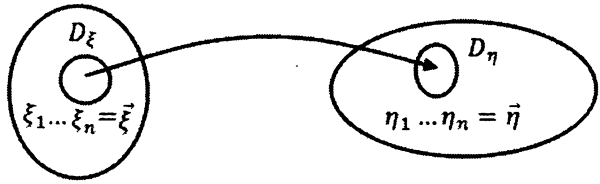


Рис. 1.9. Геометричесая интерпретация нелинейного преобразования

При этом справедливо равенство $p\{\vec{\xi} \in D_\xi\} = p\{\vec{\eta} \in D_\eta\}$, если $D_\eta = g(D_\xi)$. Если выбрать в качестве D_ξ элементарные объемы $d\xi_1 \dots d\xi_n = dV_\xi$, а в качестве D_η элементарный объем $d\eta_1 \dots d\eta_n = dV_\eta$, то $\omega(\xi_1 \dots \xi_n)|dV_\xi| = \omega(\eta_1 \dots \eta_n)|dV_\eta|$ и $\omega(\eta_1 \dots \eta_n) = \omega(h_1(\eta_1 \dots \eta_n) \dots h_n(\eta_1 \dots \eta_n)) \left|\frac{dV_\xi}{dV_\eta}\right|$.

Отношение элементарных объемов при преобразовании координат $\left|\frac{dV_\xi}{dV_\eta}\right| = \left|I\left(\frac{\xi}{\eta}\right)\right|$, где $I\left(\frac{\xi}{\eta}\right)$ - якобиан преобразования от случайных величин ξ к случайным величинам η, модуль которого равен:

$$\left|I\left(\frac{\xi}{\eta}\right)\right| = \left|det \begin{vmatrix} \frac{\partial \xi_1}{\partial \eta_1} & \dots & \frac{\partial \xi_1}{\partial \eta_n} \\ \dots & \dots & \dots \\ \frac{\partial \xi_n}{\partial \eta_1} & \dots & \frac{\partial \xi_n}{\partial \eta_n} \end{vmatrix}\right|$$

В общем случае значение якобиана может зависеть от координат $\xi_1 \dots \xi_n$. Если зависимость отсутствует, то преобразование называется линейным.

1.6.1. Закон распределения Релея

Пусть имеется вектор $\vec{\xi}(\xi_1, \xi_2)$. Координаты ξ_1 и ξ_2 – независимые центрированные случайные величины с одинаковым гауссовым законом распределения. Тогда двумерное распределение

$$\omega_\xi(\xi_1, \xi_2) = \frac{1}{2\pi\sigma^2} e^{\frac{\xi_1^2+\xi_2^2}{2\sigma^2}}.$$

Определим плотность распределения амплитуды a и фазы φ вектора с координатами ξ_1 и ξ_2.

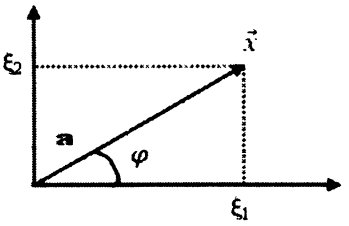


Рис. 1.10. Геометрическая интерпретация нелинейного преобразования

$$\left|\vec{\xi}\right| = a = \sqrt{\xi_1^2 + \xi_2^2}$$

$$\varphi = arctg\,\frac{\xi_2}{\xi_1} = \eta_1$$

$$\begin{cases} \xi_1 = a\cos\varphi & \eta_1 = a \\ \xi_2 = a\sin\varphi & \eta_2 = \varphi \end{cases}$$

$$\omega(a,\varphi) = \omega_\xi(a\cos\varphi, a\sin\varphi)\,|det\begin{vmatrix} \frac{\partial\xi_1}{\partial\eta_1} & \frac{\partial\xi_1}{\partial\eta_2} \\ \frac{\partial\xi_2}{\partial\eta_1} & \frac{\partial\xi_2}{\partial\eta_2} \end{vmatrix}| =$$

$$= \frac{1}{2\pi\sigma^2} e^{-\frac{a^2}{2\sigma^2}} |det\begin{vmatrix} \cos\varphi & -a\sin\varphi \\ \sin\varphi & a\cos\varphi \end{vmatrix}|;$$

$$\omega(a,\varphi) = \frac{a}{2\pi\sigma^2} e^{-\frac{a^2}{2\sigma^2}},$$

$$0 \le a \le \infty,\ 0 \le \varphi \le 2\pi.$$

$$\omega(a) = \int_0^{2\pi} \omega(a,\varphi)\,d\varphi = \frac{a}{\sigma^2} e^{-\frac{a^2}{2\sigma^2}}$$ - закон распределения Релея.

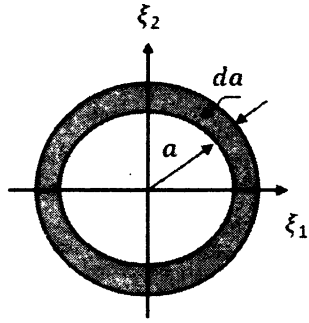


Рис. 1.11. Геометрическая интерпретация закона Релея

Вероятность $\omega(a)da$ равна вероятности того, что конец вектора $\vec{\xi}$ попадет в кольцо с шириной, равной da и радиусом, равным a (рис. 1.11).

Аналогично находится плотность распределения вероятностей фазы: $\omega(\varphi) = \int_0^\infty \omega(a, \varphi)da = \frac{1}{2\pi}$. Поскольку $\omega(a) \cdot \omega(\varphi)$, то a и φ независимые случайные величины.

Вероятность $\omega(\varphi)d\varphi$ равна вероятности того, что конец вектора $\vec{\xi}$ попадет в конус с углом, равным $d\varphi$. Этот результат широко используется при анализе узкополосного нормального шума.

Рассмотрим ещё один пример нелинейного преобразования. Пусть случайная величина ξ имеет закон распределения Релея $\omega(\xi) = \frac{\xi}{\sigma^2} e^{-\frac{\xi^2}{2\sigma^2}}$ при $\xi \geq 0$ и нулю при $\xi < 0$.

Нелинейное преобразование задается функцией $\eta = \xi^2$.

$P\{0 \leq y \leq \eta\} = F_y(\eta) = P\{0 \leq x \leq \sqrt{\eta}\} = F_x(\eta)$, где y и x – реализации случайных величин η и ξ.

Здесь область D определяется неравенством $0<\xi<\sqrt{\eta}$ а область D_η - интервалом $(0, \eta)$ (рис. 1.12).

Из равенства $P(\xi \in D_\xi) = P(\eta \in D_\eta)$ следует:

$$F_y(\eta) = \int_0^{\sqrt{\eta}} \frac{\xi}{\sigma^2} e^{-\frac{\xi^2}{2\sigma^2}} d\xi = = 1 - e^{-\frac{\eta}{2\sigma^2}}, \quad \eta > 0$$

и

$$\omega(\eta) = \frac{dF_y(\eta)}{d\eta} = \begin{cases} \frac{1}{2\sigma^2} e^{-\frac{\eta}{2\sigma^2}} \\ 0, \quad \eta < 0 \end{cases}, \quad \eta \geq 0.$$

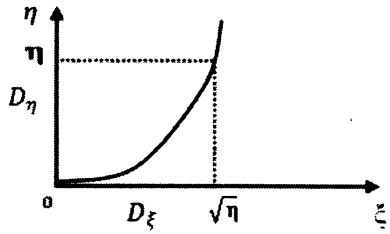


Рис. 1.12. Нелинейное преобразование $\eta = \xi^2$

1.6.2. Геометрическая интерпретация нелинейного преобразования случайной величины

На практике функцию *g(x)*, которая описывает нелинейное преобразование, удобно аппроксимировать линейно-ломаной функцией, которая представляет собой последовательность отрезков разной длины, при этом точность аппроксимации повышается с уменьшением длин отрезков (рис. 1.13).

Таким образом, нелинейное преобразование можно аппроксимировать последовательностью линейных преобразований, каждое из которых отображает некоторую область на оси *x* в соответствую область на оси *y*, например, интервал [*a b*] – в интервал [*d c*] (рис. 1.13).

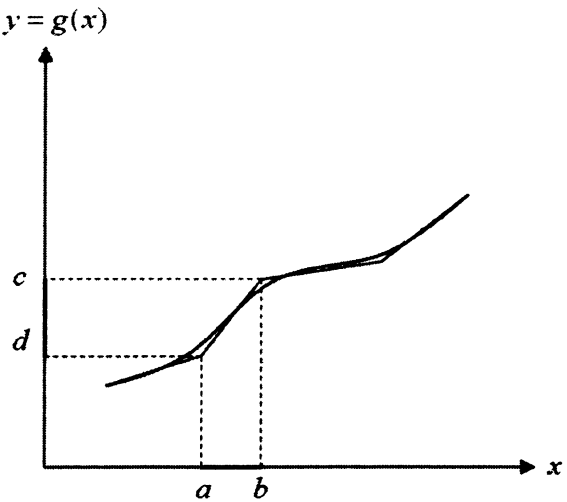


Рис. 1.13. Аппроксимация нелинейной функции g(x) линейно-ломаной функцией

Некоторые особенности линейного преобразования проявляются в зависимости от расположения отрезка: горизонтального, вертикального и под некоторым углом. Рассмотрим эти особенности на конкретных примерах.

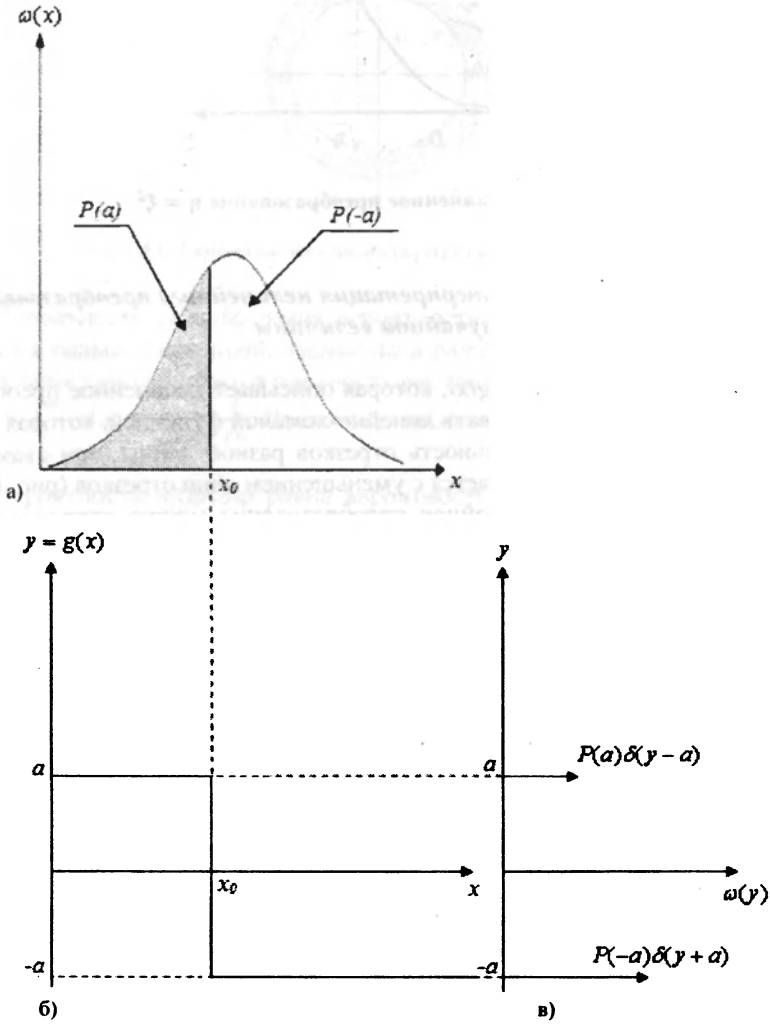


Рис. 1.14. Геометрическое представление нелинейного преобразования случайной величины:
а - исходный закон распределения; *б* - нелинейное преобразование в виде z-функции; *в* - закон распределения случайной величины y, полученной в результате нелинейного преобразования

На рис. 1.14 продемонстрирован процесс нелинейного преобразования непрерывной случайной величины x в дискретную случайную величину $y=\begin{cases} a \\ -a \end{cases}$.

В этом случае вероятность $p(-a) = p(x_0<x)$, а вероятность $p(a) = p(x<x_0)$, при этом вероятность $p(x_0<x)$ равна площади под плотностью распределения $\omega(x)$ правее x_0, а вероятность $p(x<x_0)$ – левее x_0.

Закон распределения дискретной случайной величины можно описать в виде плотности распределения вероятностей, если воспользоваться дельта-функцией: $\omega(y) = p(a)\delta(y-a) + p(-a)\delta(y+a)$, где дельта функция

$$\delta(x)=\begin{cases} 0 & при\ x \neq 0 \\ \infty & при\ x = 0 \end{cases} \text{ и } \int_{-\varepsilon}^{\varepsilon}\delta(x)dx = 1.$$

Геометрически δ-функция изображается стрелкой.

Дельта-функцию можно рассматривать как предел последовательности функций, площадь под которыми всегда равна единице, а значение в точке $x=0$ неограниченно растет. В частном случае δ-функцию можно получить как предел функции, изображенной на рис. 1.15, где значение ε является ее параметром.

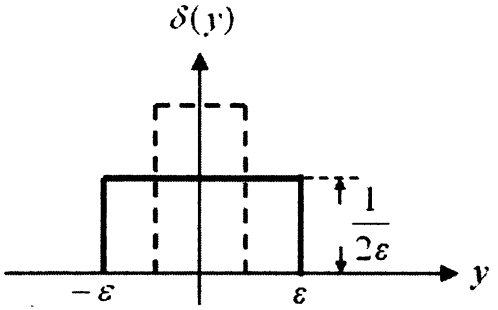


Рис. 1.15. δ-функция

При стремлении значения ε к нулю в пределе получается δ-функция. Таким образом, наличие горизонтальных участков в линейно-ломаной функции всегда приводит к появлению δ-функций в плотности распределения $\omega(y)$ преобразованной случайной величины.

Рассмотрим преобразование случайной величины, линейно-ломаная функция которого содержит отрезок, расположенный под некоторым углом (рис. 1.16)

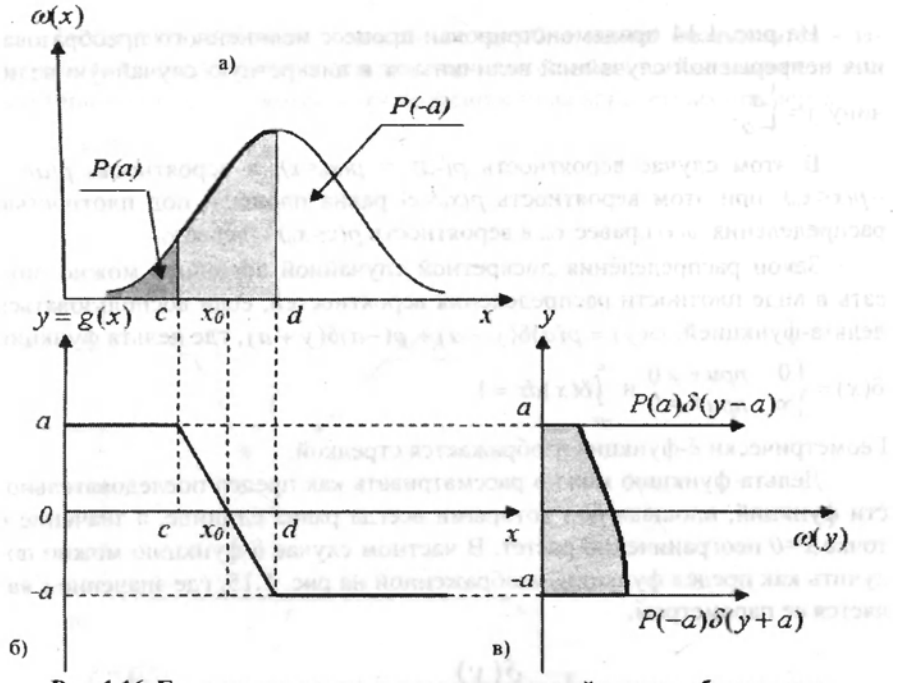


Рис. 1.16. Геометрическое представление нелинейного преобразования случайной величины:

а - исходный закон распределения; *б* - нелинейное преобразование; *в* - закон распределения случайной величины y, полученной в результате нелинейного преобразования

В этом случае вероятность $p(-a) = p(d < x)$, а вероятность $p(a) = p(x < c)$. Случайная величина x на интервале $[c,d]$ преобразуется линейно с масштабным коэффициентом k, равным $2a/(d-c)$, при этом вероятностная мера интервала $[c,d]$ сохраняется, то есть площадь под плотностью $\omega(x)$ на интервале $[c,d]$ равна площади под плотностью $\omega(y)$ на интервале $[-a,a]$. Если масштабный коэффициент равен единице $((d-c)=2a)$, то плотности $\omega(x)$ и $\omega(y)$ на этих интервалах совпадают при положительном значении тангенса угла наклона отрезка, а при отрицательном – совпадают плотности $\omega(x)$ и $\omega(-y)$.

Таким образом, если длина интервала, равная $2a$, увеличивается в k раз по сравнению с интервалом, равным $(d-c)$, то во столько же раз уменьшается масштабный коэффициент по оси $\omega(y)$, что обеспечивает сохранение вероятностной меры.

Рассмотрим случай, когда линейно-ломаная функция содержит отрезок с вертикальным расположением (рис. 1.17).

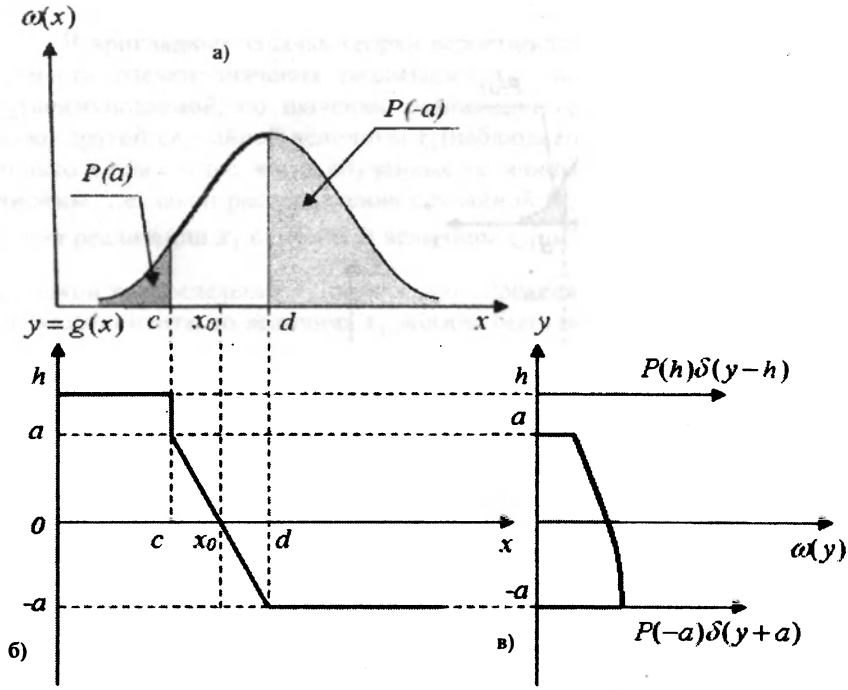


Рис. 1.17. Геометрическое представление нелинейного преобразования случайной величины (линейно-ломаная функция содержит отрезок с вертикальным расположением):
а -исходный закон распределения; *б* - нелинейное преобразование;
в - закон распределения случайной величины *y*; полученной в результате нелинейного преобразования)

Этот пример отличается от предыдущего только переносом δ-функции из точки *a* в точку *m*, при этом вероятностная мера отрезка (*a*, *m*) равна *0*, поскольку он является отображением всего одной точки с вероятностной мерой, равной нулю.

В предыдущих примерах линейные преобразования считались взаимно-однозначными. Рассмотрим пример с взаимно-неоднозначным преобразованием (рис. 1.18).

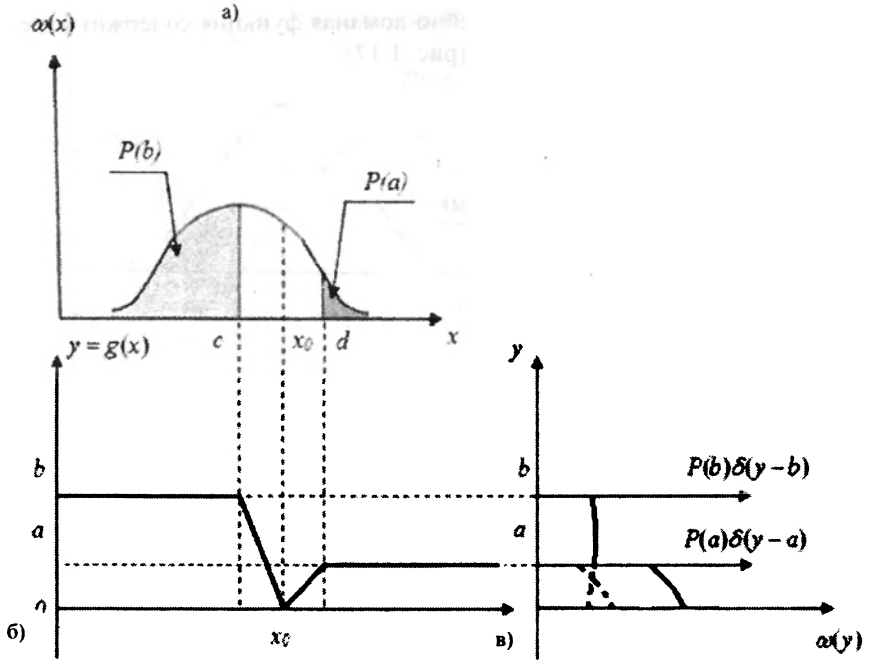


Рис. 1.18. Пример нелинейного взаимно-неоднозначного преобразования случайной величины:

а - исходный закон распределения; *б* - нелинейное взаимно-неоднозначное преобразование; *в* - закон распределения случайной величины *y*, полученной в результате нелинейного взаимно-неоднозначного преобразования)

В этом случае нелинейное преобразование необходимо представить в виде двух взаимно-однозначных преобразований, правее точки x_0 и левее, и для каждого из них в отдельности получить результат преобразования. Окончательный результат получается как сумма отдельных результатов, поскольку события правее x_0 и левее x_0 несовместны. Окончательный результат изображен сплошной линией.

1.7. Функция регрессии

1.7.1. Вывод выражения для функции регрессии

В прикладных задачах теории вероятностей часто возникает необходимость оценки значения реализации x_2 одной случайной величины ξ_2(ненаблюдаемой) по значению реализации x_1 (по выборочному значению) другой случайной величины ξ_1(наблюдаемой). Оценка имеет смысл только в том случае, когда случайные величины ξ_1 и ξ_2 статистически зависимы, т.е. закон распределения случайной величины ξ_2 зависит от значения реализации x_1 случайной величины $\xi_1(\omega(^{x_2}/_{x_1}) \neq \omega(x_2))$. Поскольку закон распределения ξ_2 полностью определяется комплексом условий D эксперимента, то величина x_1 должна быть параметром этого комплекса условий или так или иначе влиять на него. Условную плотность вероятностей $\omega\left(^{x_2}/_{x_1}\right)$ можно рассматривать как формальную математическую модель зависимости, которая отражает влияние значения x_1 на вероятность появления x_2 через комплекс условий D.

В качестве оценки значения x_2 выбирается значение $\widehat{x_2} = g(x_1)$ некоторой функции $g(x_1)$, вид которой определяет качество оценки. При отдельном наблюдении ошибка равна $x_2 - \widehat{x_2} = x_2 - g(x_1)$. При многократном наблюдении одного и того же значения x_1 случайная величина ξ_2, закон распределения которой $\omega(^{x_2}/_{x_1})$ зависит от x_1, будет принимать разные значения, то есть ошибка $\xi_2 - g(x_1)$ будет случайной величиной, средний квадрат $\sigma^2(x_1)$ которой можно выбрать в качестве критерия качества оценки:

$$\sigma^2(x_1) = \int_{-x}^{x}\left(x_2 - g(x_1)\right)^2\omega\left(^{x_2}/_{x_1}\right)dx_2 = \overline{(\xi_2 - g(x_1))^2}. \qquad (1.1)$$

Поскольку среднеквадратическая ошибка $\sigma^2(x_1)$ зависит от значения x_1 , то ее можно назвать условной. В качестве критерия чаще используется безусловная среднеквадратическая ошибка, которая получается в результате статистического усреднения $\sigma^2(x_1)$ по всем значениям x_1:

$$\sigma^2 = \int_{-x}^{x}\sigma^2(x_1)\omega(x_1)dx_1 = \int_{-x}^{x}\left(x_2 - g(x_1)\right)^2\omega(x_1)\omega\left(^{x_2}/_{x_1}\right)dx_1dx_2,$$

причем $\omega(x_1)\omega\left(^{x_2}/_{x_1}\right) = \omega(x_1, x_2)$.

Докажем, что σ^2 минимальна, когда функция $g(x_1)$ совпадает с условным средним значением случайной величины ξ_2:

$$g(x_1) = \int_{-x}^{x} x_2 \omega\left(x_2/x_1\right) dx_2 = m(x_1) = \overline{\xi_2(x_1)}.$$

Поскольку $\sigma^2(x_1) \geq 0$, то σ^2будет минимальна, если $\sigma^2(x_1)$ будет минимальна при каждом x_1. Преобразуя (1.1) при заданном значении x_1, получим:

$$\sigma^2(x_1) = \overline{(\xi_2 g(x_1))^2} = \overline{[\xi_2 - m(x_1) + m(x_1) - g(x_1)]^2} =$$
$$= \overline{(\xi_2 - m(x_1))^2} + \overline{(m(x_1) - g(x_1))^2} + 2\overline{(\xi_2 - m(x_1))}(m(x_1) - g(x_1)).$$

Сомножитель $\overline{\xi_2 - m(x_1)} = \overline{\xi_2(x_1)} - m(x_1) = 0$, и поэтому третье слагаемое равно нулю. Поскольку от $g(x_1)$ зависит только второе слагаемое и оно неотрицательно, то минимум суммы достигается при $g(x_1) = m(x_1)$, тоесть, когда второй член равен 0.

Условное среднее значение случайной величины ξ_2 при $\xi_1 = x_1$, $(m(x_1))$, рассматриваемое как функция переменного x_1, называется функцией регрессии ξ_2 на ξ_1.

Функция регрессии отражает зависимость одного из параметров (среднего значения) закона распределения случайной величины ξ_2 от значения x_1 (реализация) случайной величины ξ_1. Если зависимость среднего значения от x_1 отсутствует, то это еще не значит, что случайные величины ξ_1 и ξ_2 независимы – в общем случае от x_1 может зависеть не только среднее значение, но и другие параметры распределения. Если же случайные величины ξ_1 и ξ_2 независимы, то и каждый параметр распределения ξ_2 не будет зависеть от x_1, то есть форма функции распределения ξ_2 не будет зависеть от x_1.

1.7.2. Линейная функция регрессия

В некоторых случаях вводится ограничение на вид возможных функций $g(x_1)$, например, ограничиваются классом линейных функций $g(x_1)$, которые записываются в виде $g(x_1) = c_1 x_1 + c_2$. Выбор оптимальной функции из этого класса, тоесть той, которая дает оценку с минимальной среднеквадратической ошибкой, сводится к определению коэффициентов c_1, c_2.

Функция $m_y(x_1) = c_1 x_1 + c_2$, для которой среднеквадратическая ошибка

$$\sigma^2 = \int_{-x}^{x} \int (x_2 - c_1 x_1 - c_2)^2 \omega(x_1, x_2) dx_1 dx_2 = \overline{(\xi_2 - c_1 \xi_1 - c_2)^2}$$

минимальна, называется функцией линейной регрессии, а соответствующие коэффициенты c_1 и c_2 – коэффициентами регрессии.

Обозначив через a_1 и a_2 средние значения случайных величин ξ_1 и ξ_2, коэффициенты регрессии можно определить, если сделать следующие тождественные преобразования:

$$(\xi_2 - c_1 \xi_1 - c_2)^2 = [(\xi_2 - a_2) - c_1(\xi_1 - a_1) - (c_2 - a_2 + c_1 a_1)]^2 =$$

$= \left[\dot{\xi}_2 - c_1\dot{\xi}_1 - \gamma\right]^2$, где $\dot{\xi}_1 = \xi_1 - a_1$, $\dot{\xi}_2 = \xi_2 - a_2$, $\gamma = c_2 - a_2 + c_1 a_1$, $\dot{\xi}_1$ и $\dot{\xi}_2$ - центрированные случайные величины.

Тогда

$$\sigma^2 = \left[\dot{\xi}_2 - c_1\dot{\xi}_1 - \gamma\right]^2 = \dot{\xi}_2^{\,2} + c_1{}^2\dot{\xi}_1^{\,2} + \gamma^2 - 2\overline{\dot{\xi}_1\dot{\xi}_2}c_1 - 2\bar{\dot{\xi}}_2\gamma + \\ + c_1\bar{\dot{\xi}}_1\gamma = \sigma_2^2 + c_1^2\sigma_1^2 - 2\rho c_1 + \gamma^2, \qquad (1.2)$$

где $\overline{\dot{\xi}_1^{\,2}} = \sigma_1^2$ и $\overline{\dot{\xi}_2^{\,2}} = \sigma_2^2$ по определению дисперсии случайных величин ξ_1 и ξ_2, $\bar{\dot{\xi}}_1 = 0$ и $\bar{\dot{\xi}}_2 = 0$ как среднее значение от центрированных случайных величин.

$$\overline{\dot{\xi}_1\dot{\xi}_2} = \int\int_{-\infty}^{\infty}(x_1 - a_1)(x_2 - a_2)\,\omega(x_1, x_2)dx_1dx_2 = \rho,$$

среднее от произведения двух центрированных случайных величин $\dot{\xi}_1$ и $\dot{\xi}_2$ называется корреляцией между этими случайными величинами. Иногда удобнее использовать коэффициент корреляции $r = \frac{\overline{\dot{\xi}_1\dot{\xi}_2}}{\sigma_1\sigma_2} = \frac{\rho}{\sigma_1\sigma_2}$, который определяется как среднее значение от произведения центрированных и нормированных случайных величин $\frac{\dot{\xi}_1}{\sigma_1}$ и $\frac{\dot{\xi}_2}{\sigma_2}$.

С учетом введенных обозначений можно произвести следующие тождественные преобразования выражения (1.2):

$$\sigma^2 = \sigma_2^2 + c_1^2\sigma_1^2 - 2\sigma_1\sigma_2 r c_1 + r^2\sigma_2^2 - r^2\sigma_2^2 + \gamma^2 = \\ = \sigma_2^2(1 - r^2) + (c_1\sigma_1 - r\sigma_2)^2 + (c_2 - a_2 + c_1 a_1)^2.$$

Отсюда следует, что минимум среднеквадратической ошибки, равный $\sigma^2 = \sigma_2^2(1 - r^2)$ достигается при тех значениях c_1 и c_2, при которых последние два слагаемых (неотрицательные) равны нулю:

$$c_1\sigma_1 - r\sigma_2 = 0,$$
$$c_2 - a_2 + c_1 a_1 = 0.$$

Решая эту систему уравнений, получим:

$$c_1 = r\frac{\sigma_2}{\sigma_1}; \qquad c_2 = a_2 - c_1 a_1.$$

Соответствующая функция линейной регрессии имеет вид:

$$\widehat{x_2} = m_л(x_1) = a_2 + r\frac{\sigma_2}{\sigma_1}(x_1 - a_1). \qquad (1.3)$$

С целью наглядности геометрической интерпретации последнее равенство можно представить в виде:

$$\widehat{x_2} - a_2 = r\frac{\sigma_2}{\sigma_1}(x_1 - a_1) \qquad (1.4)$$

или

$$\frac{\widehat{x_2} - a_2}{\sigma_2} = r\frac{x_1 - a_1}{\sigma_1}. \qquad (1.5)$$

Для построения функции линейной регрессии достаточно знать только средние значения, дисперсии случайных величин ξ_1, ξ_2 и их коэффициент корреляции, который равен тангенсу угла наклона прямой, опре-

деляемой выражением (1.5), если по осям откладывать соответственно значения $\frac{x_1-a_1}{\sigma_1}$ и $\frac{\widehat{x_2}-a_2}{\sigma_2}$.

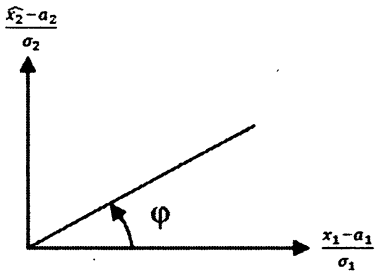


Рис. 1.19. Линейная функция регрессии, tgφ = *r*

Следует отметить, что в общем случае, то есть для произвольной плотности $\omega(x_1, x_2)$ функция линейной регрессии (1.9) может не совпадать с действительной функцией регрессии $m(x_1)$, поскольку условное среднее $m(x_1)$ может быть нелинейной функцией.

Поэтому, если r=0 , то говорят, что между ξ_2 и ξ_1 отсутствует линейная зависимость, но это еще не значит, что между условным средним значением случайной величины ξ_2 и значением x_1 вообще отсутствует какая-либо зависимость.

Пример. Рассмотрим классическую задачу измерения физической величины. До измерения физической величины обычно располагают некоторыми сведениями о ее значении, например, может быть известна область ее возможных значений и степень ожидания того или другого значения, которая характеризуется вероятностью. В рамках теории вероятностей измеряемая величина рассматривается как реализация случайной величины ξ_2 (ненаблюдаемой), априорная (до измерения) неопределенность которой ограничивается видом ее закона распределения $\omega(x_2)$ с дисперсией σ_2^2 и средним значением, равным a_2. Результат измерения рассматривается как реализация случайной величины (наблюдаемой) $\xi_1 = \xi_2 + \Delta$, где Δ – ошибка измерения с нулевым средним значением (Δ=0), которая представляет собой сумму двух независимых случайных величин ξ_2 и Δ и имеет дисперсию $\sigma_1^2 = \sigma_2^2 + \sigma_\Delta^2$ и среднее значение $\overline{\xi_1} = \overline{\xi_2} = a_2$.
Тогда корреляция между случайными величинами ξ_2 и ξ_1 равна

$$\rho = \overline{\dot{\xi}_1\dot{\xi}_2} = \overline{\dot{\xi}_2\,(\dot{\xi}_2 + \Delta)} = \overline{\dot{\xi}_2^{\,2}} + \overline{\dot{\xi}\Delta} = \overline{\dot{\xi}_2^{\,2}} = \sigma_2^2,$$

где $\dot{\xi}_2 = \xi_2 - a_2$, $\overline{\dot{\xi}_2\Delta} = \overline{\dot{\xi}_2}\,\overline{\Delta} = 0$.
Коэффициент корреляции

$$r = \frac{\sigma_2^2}{\sigma_1\sigma_2} = \frac{\sigma_2}{\sigma_1},$$

а равенство (1.9) запишется в виде

$$\widehat{x_2} - a_2 = \frac{\sigma_2^2}{\sigma_1^2}(x_1 - a_2)$$

или

$$\widehat{x_2} - a_2 = \frac{1}{1+\frac{\sigma_\Delta^2}{\sigma_2^2}}(x_1 - a_2).$$

Отсюда следует, что эффективная оценка $\widehat{x_2}$ значения измеряемой физической величины не совпадает с результатом измерения x_1.

Если дисперсия σ_Δ^2 ошибки измерения (неопределенность результата измерения или апостериорная неопределенность) много меньше дисперсии σ_2^2, которая характеризует априорную неопределенность, то $\widehat{x_2} \approx x_1$, а при $\sigma_\Delta^2 \gg \sigma_2^2$ оценка $\widehat{x_2}=a_2$, тоесть в этом случае результат измерения x_1 практически не несет информации о значении физической величины в дополнение к тому, что было о ней известно до измерения.

Минимальная среднеквадратическая ошибка равна

$$\sigma_2 = \sigma_2^2(1 - r^2) = \left(\frac{1}{\sigma_\Delta^2} + \frac{1}{\sigma_2^2}\right)^{-1}.$$

Из этого равенства следует, что в случае полного отсутствия априорных сведений об измеряемой величине $\sigma_2^2 = \infty$ и, следовательно, $\sigma^2 = \sigma_\Delta^2$. Когда же значение измеряемой величины заранее известно $(\sigma_2^2 = 0, \sigma^2 = 0)$, нет необходимости производить измерение. Если дисперсия σ_Δ^2=0 (абсолютно точное измерение), то $\sigma^2 = 0$, тоесть всю необходимую информацию о значении измеряемой величины доставляет результат измерения и поэтому априорными сведениями можно пренебречь.

Если $\sigma_\Delta^2 = \infty$, то $\sigma^2 = \sigma_2^2$, то есть результат измерения не доставляет какой-либо информации и неопределенность измеряемой величины определяется априорной дисперсией σ_2^2.

Интересно сравнить эту задачу с задачей предсказания результата измерения по известному значению измеряемой величины. По отношению к предыдущей задаче наблюдаемая и ненаблюдаемая величины поменяются местами и равенство (1.9) примет вид $\widehat{x_1} - a_1 = r\frac{\sigma_1}{\sigma_2}(x_2 - a_2)$, причем a_1=a_2. Коэффициент корреляции не изменится и будет равен $r=\frac{\sigma_2}{\sigma_1}$

Отсюда следует, что $\widehat{x_1}$=x_2 и ошибка предсказания $\sigma_1^2(1 - r^2) = \sigma_\Delta^2$.

Таким образом, чтобы эффективно предсказать результат измерения, в качестве его оценки всегда следует брать истинное значение измеряемой величины. В заключение следует отметить, что при оценке ненаблюдаемой величины по критерию среднеквадратической ошибки используются не все априорные сведения, содержащиеся в ее законе распределения, а только ее дисперсия и среднее.

2. МОДЕЛИРОВАНИЕ СЛУЧАЙНЫХ ВЕЛИЧИН

Описание законов распределения вероятностей случайных величин осуществляется при помощи функции распределения $F(x)$ (интегральной функции распределения) или плотности распределения вероятностей $W(x)$, которые связаны между собой соотношениями:

$$W(x) = \frac{dF(x)}{dx}; \qquad F(x) = \int_{-\infty}^{x} W(x)dx.$$

Напомним, что функция распределения вероятностей в точке x рана вероятности наступления события: «Значение x в результате реализации случайной величины ξ оказалось строго меньше значения Z»[1]:

$$F(Z) = P\{x < Z\}.$$

Плотность распределения вероятностей $W(x)$ характеризует вероятность попадания значения случайной величины в ту или иную область оси абсцисс. Вероятность попадания реализации аналоговой случайной величины в точку, очевидно, равна нулю, вероятность же попадания в интервал $[a,b]$ (пусть сколь угодно малый), имеет определенную величину[2]. С учетом определения плотности распределения вероятностей запишем:

$$P(a \le x \le b) = \int_{a}^{b} W(x)dx.$$

Включение или невключение конечных точек в интервал интегрирования несущественно в силу сказанного ранее.

Датчики (генераторы) случайных (псевдослучайных) чисел с заданным законом распределения вероятностей, т. е. с заданной функцией плотности распределения вероятностей $W(x)$ либо интегральной функцией распределения $F(x) = \int_{-\infty}^{x} W(x)dx$ применяются исключительно широко в криптографии, стеганографии, при математическом (имитационном) моделировании сложных технических систем, при реализации компьютерных игр и т.д. Особенно часто необходимость в подобных датчиках возникает при исследовании или проектировании информационных систем, систем массового обслуживания.

---

[1] Функция распределения вероятностей связывает между собой два понятия теории вероятностей: случайное событие и случайная величина – численное значение результатов опыта (эксперимента).

[2] Если плотность распределения вероятностей всюду на интервале [a,b] равна нулю, то и вероятность попадания реализации случайной величины на этот интервал тоже будет равна нулю.

Существует три способа получения реализаций последовательностей случайных чисел:

- выбор из таблиц случайных чисел (неудобство этого метода достаточно очевидно);
- использование физических датчиков (что также является сложной задачей, поскольку для каждого закона распределения потребуется изготовление собственного физического датчика; кроме того, в задачах защиты информации часто требуются псевдослучайные числа, а физические датчики обычно являются генераторами случайных чисел);
- генерирование псевдослучайных последовательностей чисел на ЭВМ.

Специфика всякой решаемой задачи защиты информации или моделирования требует своего закона распределения случайной величины, т.е. своей функции $W(x)$ либо $F(x)$. Разрабатывать датчики случайных чисел "на все случаи жизни" нерационально и практически невозможно. Отсюда необходимость создания некоторого универсального, простого в реализации генератора случайных чисел с достаточно простым законом распределения, используя который можно было бы получать случайные числа с законом распределения вероятностей, соответствующим решаемой задаче. В качестве такого генератора чаще всего используется датчик псевдослучайных чисел, равномерно распределенных в диапазоне (0, 1), реализуемый программно на ЭВМ.

Существуют три группы методов генерации случайных чисел с заданным законом распределения с использованием последовательности случайных чисел, равномерно распределенных в диапазоне (0,1):

1) точные — методы нелинейного преобразования и метод Неймана;

2) приближенные — метод кусочной аппроксимации, метод замены непрерывных распределений соответствующими дискретностями, метод численного решения уравнения при нелинейном преобразовании;

3) специальные — методы, основанные на свойствах преобразований случайных величин.

2.1. Моделирование случайной величины с произвольно заданным законом распределения посредством нелинейного преобразования случайной величины с равномерным законом распределения

Имеется датчик случайной (псевдослучайной) величины Y с функцией распределения вероятностей, равномерной в диапазоне [0,1],

$$W(y)=\begin{cases}1, & \text{при } 0\le y\le 1\\ 0, & \text{при } y<0,\ y>1\end{cases} \qquad (2.1)$$

Требуется получить датчик случайных чисел x с иным законом распределения вероятностей. Предполагается, что необходимый закон распределения известен точно, т.е. аналитическое выражение требуемой

плотности распределения вероятностей $W(x)$ либо функции распределения $F(x)$ задано.

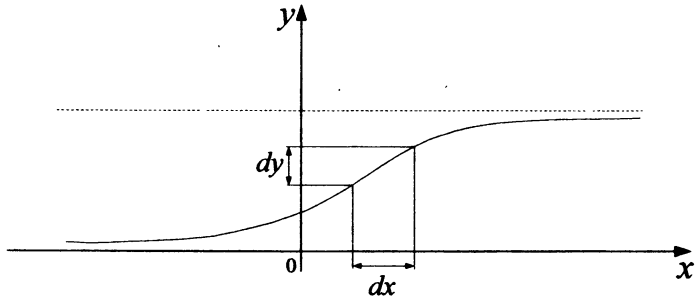


Рис. 2.1. Преобразование равномерно распределенной в диапазоне (0,1) случайной величины Y в случайную величину X с интегральной функцией распределения F(x)

Решение поставленной задачи сводится к определению вида нелинейного преобразования $x=g(y)$ (рис. 2.1) случайной величины Y с равномерным на интервале (0,1) законом распределения, в результате которого получается случайная величина X с заданной плотностью распределения вероятностей $W(x)$. Вероятность того, что реализация случайной величины y попадет в интервал dy, равна вероятности того, что значение $x=g(y)$ попадет в соответствующий интервал dx:

$$W(y)dy=W(x)dx.$$

Поскольку случайная величина y имеет равномерный закон распределения, то $W(y)=1$ во всей области определения [0,1]. Поэтому $dy=W(x)dx$. Интегрируя это дифференциальное уравнение, получим

$$y = \int_{-\infty}^{x} W(x)dx = F(x). \qquad (2.2)$$

Графическая иллюстрация проделанных преобразований приведена на рис.2.1.

В результате решения уравнения (2.2) относительно переменной x можно определить вид функции $g(y)$, которая совпадает с функцией, обратной к интегральной функции распределения $F(x)$: $x = F^{-1}(y)$.

Пример. Пусть требуется получить случайную величину с релеевским законом распределения, у которой

$$F(x)=1-\exp(-x^2/2\sigma^2), \qquad W(x)=(x/\sigma^2)\exp(-x^2/2\sigma^2).$$

Тогда, решая уравнение $y=1-\exp(-x^2/2\sigma^2)$, получим $x=\sigma\sqrt{-2\ln(1-y)}$.

Известно, что если случайная величина y распределена равномерно в диапазоне [0,1], то случайная величина $1-y$ имеет такой же закон распределения, поэтому последнее равенство целесообразно заменить статистически эквивалентным: $x=g(y)=\sigma\sqrt{-2\ln(y)}$. Таким образом, если датчик генерирует последовательность случайных чисел y_i, $i = 1, 2...$, с равномерным законом распределения, то последовательность чисел $x_i=\sigma\sqrt{-2\ln(y_i)}$ $i=1, 2...$ будет иметь закон распределения Релея.

2.2. Метод Неймана

Случайную величину с заданной плотностью распределения вероятностей $W(x)$ можно получить из базовой последовательности случайных (псевдослучайных) чисел, имеющих равномерную плотность распределения вероятностей на интервале [0,1] следующим образом.

Пусть имеются две случайные величины X, $x \in (0,a)$ и Y, $y \in (0,b)$ с равномерными законами распределения, которые образуют двумерное пространство элементарных событий (рис. 2.2).

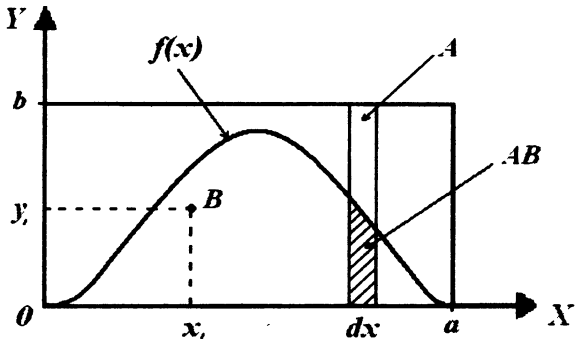


Рис. 2.2. Двумерное пространство элементарных событий:
B – область в пространстве элементарных событий между осью абсцисс и кривой, воспроизводящей в некотором масштабе плотность распределения вероятностей W(x); a - область прямоугольника с размерами: dx – по оси абсцисс, b - по оси ординат

Случайные величины X и Y можно получить в результате умножения чисел базовой последовательности соответственно на a и b. Функция

$f(x)$ изображает в общем случае в произвольном масштабе заданную плотность распределения вероятностей $W(x)$. Буквой A будем обозначать область прямоугольника с размерами: по оси абсцисс – dx, по оси ординат – b; B – область в пространстве элементарных событий между осью абсцисс и кривой $f(x)$.

Алгоритм

1. На первом шаге вырабатываются реализации x_1 и y_1 случайных величин X и Y (точка в пространстве элементарных событий).

2. Если точка с координатами (x_1,y_1) попала в область B под функцией $f(x)$, то y_1 отбрасывается, а x_1 используется в качестве выходного значения случайной величины. В противном случае отбрасываются оба значения x_1 и y_1 (холостой прогон).

3. Переход к пункту 1.

Таким образом, случайная величина с плотностью распределения $W(x)$ формируется посредством прореживания базовой последовательности.

Докажем, что получаемая таким образом случайная величина описывается плотностью распределения вероятностей $W(x)$, которая совпадает в определенном масштабе с функцией $f(x)$. В данном случае вероятность $W(x)dx$ попадания реализации выходной случайной величины в интервал dx равна условной вероятности $P(A/B)=P(AB)/P(B)=W(x)dx$, где события A и B определены как подмножества в пространстве элементарных событий. Поскольку плотность распределения вероятностей постоянна в пространстве элементарных событий, то применим геометрический метод вычисления вероятностей, т.е. условная вероятность $P(A/B)$ равна отношению площади пересечения событий A и B к площади $S(B)$ события B: $W(x)dx= P(A/B)=f(x)dx / S(B)$.
Отсюда следует, что $W(x) = f(x)/ S(B)$.

2.3. Моделирование случайной величины в случае приближенного задания ее закона распределения

Если аналитическое выражение нелинейного преобразования $x=q(y)$ не удается получить, или оно имеет сложный, "громоздкий" вид, а также если распределение случайной величины X точно не известно, в задачах моделирования применяют аппроксимацию. Пусть требуется получить случайную величину X с плотностью распределения вероятностей $W(x)$.

Аппроксимацию можно реализовать с помощью конечного ряда

$$W(x) \approx P_1 w_1(x)+P_i w_i(x)+P_n w_n(x)=\hat{W}(x), \qquad (2.3)$$

где $w_i(x)$, $i=1...n$ – законы распределения вероятностей n независимых случайных величин; P_i – коэффициенты разложения; $\hat{W}(x)$ – результат аппроксимации требуемой плотности распределения вероятностей. Точность аппроксимации обычно выбирается экспериментально, хотя можно использовать и теорию разложения функции в ряды. Проектируемый программный модуль (генератор) должен вырабатывать псевдослучайную величину с плотностью распределения $\hat{W}(x)$.

Блок-схема алгоритма генерации случайной величины с плотностью распределения $\hat{W}(x)$ представлена на рис. 2.3.

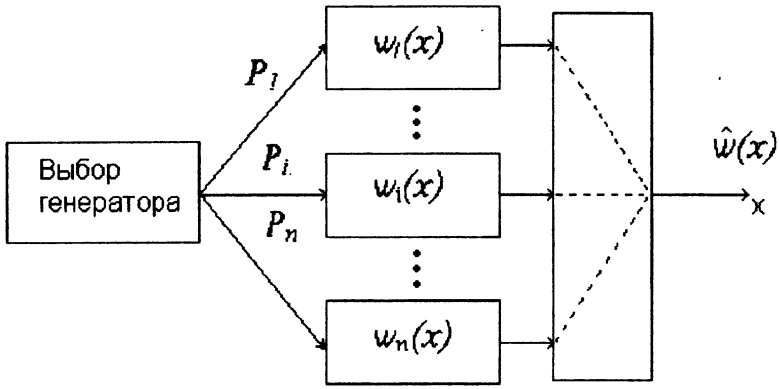


Рис. 2.3 Блок-схема алгоритма генерирования случайной величины с плотностью распределения $\hat{W}(x)$

На первом шаге с вероятностью P_i выбирается один из генераторов случайных величин с плотностью распределения вероятностей $w_i(x)$, $i=1...n$.

На втором шаге выбранный генератор вырабатывает реализацию x случайной величины X, которая используется как выходная.

Докажем, что на выходе получаем случайную величину с законом распределения вероятностей, равной $\hat{W}(x)$. Вероятность $\hat{W}(x)dx$ вычисляется по формуле полной вероятности $\hat{W}(x)dx=P_1 w_1(x),dx+ ...+P_i w_i(x)dx+...+ P_n w_n(x)dx$, где $w_i(x)dx$ интерпретируется как условная вероятность попадания реализации случайной величины X в интервал dx при условии, что выбран генератор с плотностью распределения вероятностей $w_i(x)$.

Обычно плотности $w_i(x)$ принадлежат одному типу и различаются только средними значениями и (при необходимости) – дисперсией.

2.3.1. Аппроксимация с помощью случайных величин с равномерным законом распределения

В случае аппроксимации с помощью случайных величин с равномерным законом распределения заданная плотность $W(x)$ аппроксимируется ступенчатой функцией $\hat{W}(x)$ (рис. 2.4). Отдельная ступенька описывается функцией $P_i w_i(x)$, где $w_i(x)=1/\Delta$ – равномерная на интервале, равном $\Delta=(x_i+\Delta/2)-(x_i-\Delta/2)$, плотность распределения вероятностей случайной величины X со средним значением, равным X_i. Высота прямоугольника, изображающего функцию $P_i w_i(x)$, равна $W(x_i)$, а площадь этого прямоугольника равна P_i, так как

$\int_{x_i-\Delta/2}^{x_i+\Delta/2} P_i w_i(x)dx = P_i = W(x_i)\Delta$, где интеграл $\int_{x_i-\Delta/2}^{x_i+\Delta/2} w_i(x)dx = 1$.

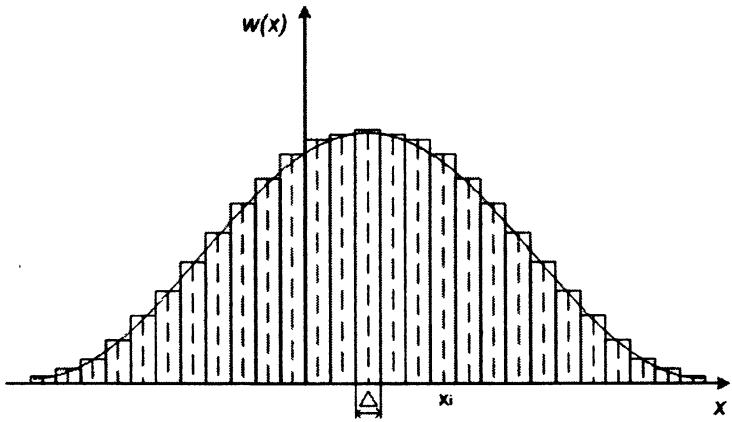


Рис. 2.4. Ступенчатая аппроксимация

Рассмотрим способ нахождения величин P_i.

Поскольку $\int_{-\infty}^{\infty}\hat{W}(x)dx = 1$, то $\sum_{i=-m}^{n} P_i = 1$, и, следовательно, $\sum_{i=-m}^{n} W(x_i)\Delta = 1$.

Отсюда $\Delta = 1 \Big/ \sum_{i=-m}^{n} W(x_i)$ и $P_i = W(x_i) \Big/ \sum_{i=-m}^{n} W(x_i)$.

Значения P_i инвариантны к масштабу функции $W(x)$ по оси ординат, т.е., если взять функцию $kW(x)$, где k произвольное положительное число,

то значения P_i не изменятся. Поэтому значения $W(x_i)$ можно измерять, например, количеством клеток подходящего размера на листе бумаги, на котором изображен график плотности $W(x)$.

Если в библиотеке ЭВМ имеется генератор случайной величины $У$ с равномерным законом распределения на интервале (0, 1), то случайную величину X с законом распределения $w_i(x)$, равномерным на интервале $(x_i - \Delta/2, x_i + \Delta/2)$ можно получить в результате следующего преобразования: $x = y\Delta + (i\text{-}1)\Delta$, где номер интервала i совпадает с номером генератора, случайно выбираемым с вероятностью P_i.

2.3.2. Аппроксимация с помощью случайной величины с треугольным законом распределения

Более точную (плавную) аппроксимацию можно получить, используя в качестве $w_i(x)$ в выражении (2.3) треугольные функции. При этом заданная плотность $w_i(x)$ аппроксимируется линейно-ломаной линией представленной на рис. 5. Вероятность P_i равна площади соответствующего треугольника, т. е.

$$P_i = W(x_i).$$

Так как $\sum_{i=-m}^{n} P_i = 1$, то, как и в первом случае, значения P_i можно вычислять по формуле

$$P_i = W(x_i) \Big/ \sum_{i=-m}^{n} W(x_i).$$

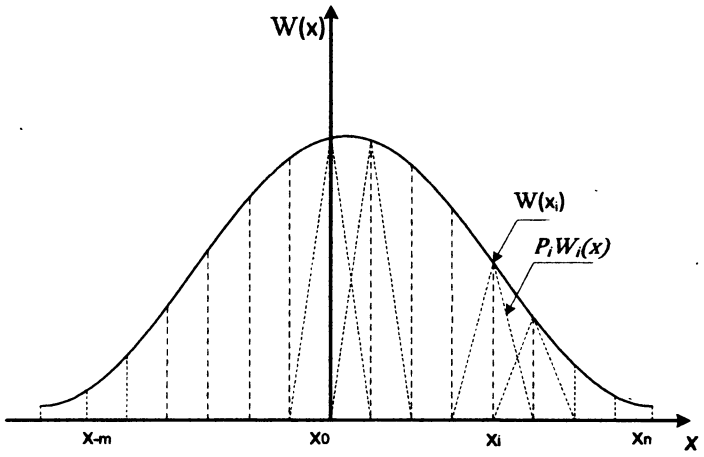


Рис. 2.5. Аппроксимация линейно-ломаной линией

Реализацию случайной величины с треугольным законом распределения, определенную на интервале (-1, 1), можно получить как разность реализаций y_1 и y_2 , двух независимых случайных величин с законом распределения (2.1). Реализация случайной величины с законом распределения $\hat{W}(x)$ можно вычислить по формуле: $x=(y_1-y_2)\ \Delta+i\Delta$.

2.3.3. Моделирование случайной величины с нормальным (гауссовым) законом распределения

Специальные методы применяются в тех случаях, когда удается использовать специфические свойства соответствующих случайных величин и их преобразований. Общих рекомендаций по специальным методам не существует. В каждом конкретном случае следует полагаться на свою квалификацию и научную интуицию.

Рассмотрим два способа моделирования случайной величины X, имеющей нормальный закон распределения с нулевым средним и дисперсией, равной единице. Напомним, что функция плотности распределения вероятностей такой случайной величины описывается выражением

$$f(x)=\frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{x^2}{2}\right\}.$$

Первый метод основан на центральной предельной теореме:
если Y_1, Y_2, Y_n — неизвестные случайные величины, имеющие одинаковый закон распределения с матожиданием m и дисперсией σ^2, то при неограниченном увеличении n закон распределения суммы

$$Z=\sum_{i=1}^{n} Y_i \qquad (2.4)$$

приближается к нормальному, с дисперсией $n\sigma^2$ и матожиданием nm.

Используя датчик независимых случайных чисел Y, с законом распределения вероятностей, определяемым выражением (2.1), следует получить n таких чисел и сложить их. Учитывая, что равномерно распределенная случайная величина Y имеет моменты m = 0,5 и σ^2 =1/12, определим, что случайная величина Z в выражении (2.4) будет иметь матожидание $0{,}5n$ и дисперсию $n/12$. Искомая нормированная случайная величина X получается из выражения

$$X=(Z-0{,}5n)\sqrt{12/n} \quad . \qquad (2.5)$$

Требуемое значение n определяется исходя из необходимой точности аппроксимации и быстродействия датчика. Известно, например, что вероятность отклонения гауссовой случайной величины от математического ожидания более чем на 3 не превышает 0,003. На практике обычно ограничиваются n = 6...12. Например, выбрав n=12 формула (2.5) примет весьма простой вид $X=Z-6$, что, безусловно, повысит быстродействие

генератора. Рассмотренный метод прост в реализации и позволяет получить закон распределения случайной величины, очень близкий к нормальному.

Второй метод основан на свойствах следующих преобразований.

В разделе 1.6.1 было установлено, что если ξ_1 и ξ_2 - две независимые гауссовы случайные величины с нулевыми средними значениями и дисперсией, равной σ^2, то имеет место взаимно однозначное преобразование:

$$\xi_1 = a\cos\varphi$$
$$\xi_2 = a\sin\varphi,$$

где a - случайная длина вектора с координатами ξ_1 и ξ_2, имеющая закон распределения Релея; φ - случайное значение угла между указанным вектором и осью абсцисс, с равномерной плотностью распределения вероятностей на интервале $[0, 2\pi]$, причем случайные величины a и φ независимы так же, как ξ_1 и ξ_2. Отсюда следует, что независимые случайные величины ξ_1 и ξ_2 можно получить в результате преобразования случайных величин a и φ, которые необходимо предварительно сгенерировать или получить в результате преобразования базовой последовательности с равномерной плотностью распределения на интервале [0, 1].

Моделирование случайной величины, распределенной по закону Релея, производится методом нелинейного преобразования и рассмотрено в примере 1. Моделирование случайной величины, равномерно распределенной в диапазоне $[0, 2\pi]$, осуществляется изменением масштаба. Отсюда пара независимых гауссовых случайных величин с параметрами ($m = 0$ и $\sigma^2 = 1$) получается путем следующего преобразования двух независимых равномерно распределенных на интервале [0, 1] случайных величин y_1, y_2:

$$x_1 = \sqrt{-2\ln y_1}\,\sin 2\pi y_2,$$
$$x_2 = \sqrt{-2\ln y_1}\,\cos 2\pi y_2,$$

где значения базовой последовательности с равномерной плотностью распределения на интервале [0, 1]

Этот метод позволяет получить точное нормальное распределение случайной величины, однако требует значительного времени из-за вычисления нелинейных функций. Обычно его используют, когда необходимо учитывать реализации гауссовых случайных величин с очень большим отклонением от математического ожидания, т.е. когда важны "хвосты" нормального закона распределения вероятностей.

3. МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

3.1. Сглаживание экспериментальных зависимостей по методу наименьших квадратов

Предположим, что точная зависимость y от x имеет вид $y=y(x)$. Тогда закон распределения у можно записать в виде

$$\omega_i(y_i) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{[y_i - y(x_i)]^2}{2\sigma^2}}.$$

Произведение всех $\omega_i(y_i)$ равно

$$\mathrm{k} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}[y_i - y(x_i)]^2}.$$

Функция правдоподобия максимальна, когда

$$\frac{1}{2\sigma^2}\sum_{i=1}^{n}[y_i - y(x_i)]^2 = \min,$$

или

$$\sum_{i=1}^{n}[y_i - y(x_i)]^2 = \min.$$

Если φ задать в параметрическом виде $\varphi(x; a, b, c \ldots)$, то экстремум находится из системы уравнений:

$$\sum_{i=1}^{n}[y_i - \varphi(x_i; a, b, c \ldots)]\left(\frac{\partial\varphi}{\partial a}\right)_i = 0$$

$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

$$\sum_{i=1}^{n}[y_i - \varphi(x_i; a, b, c \ldots)]\left(\frac{\partial\varphi}{\partial c}\right)_i = 0.$$

Пусть $y=ax+b=\varphi(x; a, b)$

$$\frac{\partial\varphi}{\partial a} = x \qquad \left(\frac{\partial\varphi}{\partial a}\right)_i = x_i$$

$$\frac{\partial\varphi}{\partial b} = 1 \qquad \left(\frac{\partial\varphi}{\partial b}\right)_i = 1.$$

Тогда уравнения примут вид

$$\sum_{i=1}^{n}[y_i - (ax_i + b)]x_i = 0,$$

$$\sum_{i=1}^{n}[y_i - (ax_i + b)] = 0.$$

Или раскрывая скобки и проводя суммирование получим

$$\begin{cases} \sum_{i=1}^{n} x_i y_i - a\sum_{i=1}^{n} x_i^2 - b\sum_{i=1}^{n} x_i = 0 \\ \sum_{i=1}^{n} y_i - a\sum_{i=1}^{n} x_i - bn = 0. \end{cases}$$

Разделив оба уравнения на n, имеем

$$\begin{cases} \frac{\sum x_i y_i}{n} - a\frac{\sum_{i=1}^{n} x_i^2}{n} - b\frac{\sum_{i=1}^{n} x_i}{n} = 0 \\ \frac{\sum_{i=1}^{n} y_i}{n} - a\frac{\sum_{i=1}^{n} x_i}{n} - b = 0, \end{cases}$$

$$\begin{cases} \alpha_{11}^*[x,y] - a\alpha_2^*[x] - bm_x^* = 0 \\ m_y^* - am_x^* - b = 0. \end{cases}$$

Выразим из второго уравнения и подставим в первое $b = m_y^* - am_x^*$

$$\alpha_{11}^*[x,y] - a\alpha_2^*[x] - (m_y^* - am_x^*)m_x^* = 0.$$

Решая последнее уравнение, получим

$$a = \frac{\alpha_{11}^*[x,y] - m_x^* m_y^*}{\alpha_2^*[x] - (m_x^*)^2} = \frac{k_{xy}^*}{D_x^*};$$

$$a = \frac{k_{xy}^*}{D_x^*} \quad ; b = m_y^* - am_x^*;$$

$$k_{xy}^* = \frac{\sum_{i=1}^{n}(x_i - m_x^*)(y_i - m_y^*)}{n};$$

$$D_x^* = \frac{\sum_{i=1}^{n}(x_i - m_x^*)^2}{n}.$$

Линейная зависимость между y и x равна

$$y = \frac{k_{xy}^*}{D_x^*}x + m_y^* - \frac{k_{xy}^*}{D_x^*}m_x^*$$

или

$$(y - m_x^*) = \frac{k_{xy}^*}{D_x^*}(x - m_x^*);$$

$$\left(\frac{(y-m_x^*)}{\sigma_y^*}\right)=\frac{k_{xy}^*}{\sigma_x^*\sigma_y^*}\left(\frac{x-m_x^*}{\sigma_x^*}\right);$$

$D_y^*=\frac{\sum_{i=1}^{n}(y_i-m_x^*)^2}{n}$; $\quad\sigma_y^*=\sqrt{D_y^*}\quad\sigma_x^*=\sqrt{D_x^*}$.

Пример

Пусть x - цена на нефть, y - индекс акций нефтяных компаний

Экспериментальные значения x и y приведены в табл. 3.1

Таблица 3.1

x	17,28	17,05	18,3	18,8	19,2	18,5	$\sum$ = 109,13
y	537	534	550	555	560	552	$\sum$ = 32,88

$\sum x_i y_i = 59847{,}06 \quad \sum x_i^2 = 1988{,}5209$

$1988{,}5209a + 109{,}13b = 59847{,}06$

$109{,}13a + 6b = 3288$

Решение: $a = 12{,}078\ , b = 328{,}32.$

Тогда $y = 12{,}078\,x + 328{,}32.$

3.2. Оценка параметров закона распределения

3.2.1. Метод максимального правдоподобия

Вид любого закона распределения $w(x)$ зависит только от комплекса условий D, изменение которого, вызванное изменением значения некоторого параметра θ, может привести к изменению вида закона распределения $w(x)$. В этом случае параметр θ можно рассматривать как параметр закона распределения $w(x/\theta)$ случайной величины x. Необходимо получить оценку θ^* значения параметра θ при условии, что тип закона распределения известен, но при этом не известно значение его параметра. Оценка параметра θ производится на основе выборки $(x_1, x_2, ...x_n) = \vec{x}$, которая представляет собой последовательность из n результатов испытаний при неизвестном, но одном и том же значении параметра θ. В качестве оценки выбирается такое значение параметра, которое чаще всего появляется при полученной в результате опыта выборке. Это значение называется оценкой по максимуму апостериорной (послеопытной) вероятности и вычисляется как $\max_{\theta} w(\theta/\vec{x})$. К сожалению, на практике возникают проблемы с определением вида функции $w(\theta/\vec{x})$ и поэтому чаще всего пользуются правдопо-

добной оценкой, которая строится следующим образом. Параметр θ и выборку $\vec{x}$ можно рассматривать как две зависимые случайные величины с двухмерным законом распределения $w(\theta,\vec{x})$, причем $\vec{x}$ принадлежит выборочному пространству, а θ пространству параметров. По формуле умножения вероятностей имеем: $w(\theta,\vec{x}) = w(\theta)w(\vec{x}/\theta) = w(\vec{x})w(\theta/\vec{x})$. Отсюда следует, что $w(\theta/\vec{x}) = \dfrac{w(\theta)w(\vec{x}/\theta)}{w(\vec{x})}$, где $w(\theta)$ - априорный (доопытный) закон распределения параметра θ.

Если параметр θ подчиняется равномерному закону распределения, то $w(\theta) = \text{const}$ на пространстве параметров и функции $w(\theta/\vec{x})$ и $w(\vec{x}/\theta)$ достигают своего максимального значения при одном и том же значении параметра θ. В этом случае функция $w(\vec{x}/\theta) = L(\theta)$ называется функцией правдоподобия и обозначается как $L(\theta)$, а значение параметра θ^*, которое доставляет максимум функции правдоподобия $L(\theta)$, называется правдоподобной оценкой. Следует напомнить, что выборка $\vec{x}$ постоянна в процессе вычисления оценки параметра.

Пример. Вычислить оценки максимального правдоподобия для дисперсии и математического ожидания гауссова закона распределения $w(x) = \dfrac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x-m)^2}{2\sigma^2}}$.

Решение. Поскольку элементы выборки считаем независимыми, то многомерный закон распределения равен произведению одномерных: $w(\vec{x}/m,\sigma^2) = (\dfrac{1}{\sqrt{2\pi\sigma}})^n \cdot e^{\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-m)^2} = L(m,\sigma^2)$. Это выражение является в то же время функцией правдоподобия $L(m,\sigma^2)$, если ее аргументами считать m и σ^2, при постоянном значении выборки $\vec{x}$. Функции $L(m,\sigma^2)$ и $\ln L(m,\sigma^2)$ достигают своего максимального значения при одних и тех же значениях аргументов, поскольку логарифм является монотонно возрастающей функцией, поэтому, с целью упрощения вычисления оценок, целесообразнее использовать функцию

$$\ln L(m,\sigma^2) = \ln w(\vec{x}/m,\sigma^2) = -n\ln\sqrt{2\pi\sigma^2} - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-m)^2.$$

Чтобы найти максимум, берем производные от $\ln L(m,\sigma^2)$ по m и σ^2 и приравниваем их к нулю:

$$\frac{\partial \ln L(m,\sigma^2)}{\partial m} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i - m) = 0,$$

$$\frac{\partial \ln L(m,\sigma^2)}{\partial \sigma^2} = -\frac{n}{\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}(x_i - m)^2 = 0.$$

После упрощения получим систему уравнений:

$$\begin{cases} \sum_{i=1}^{n}(x_i - m) = 0 \\ \frac{-n}{\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}(x_i - m)^2 = 0. \end{cases}$$

Решая эту систему уравнений, получим

$$m^* = \frac{1}{n}\sum_{i=1}^{n}x_i, \sigma^{2*} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - m^*)^2,$$

где m^* - выборочное среднее, σ^{2*} - выборочная дисперсия.

3.2.2. Метод моментов

Все параметры закона распределения можно разделить на две группы: моменты и все остальные. Считаем, что моменты можно оценить экспериментально по выборке $x_1, x_2 ... x_n$. Такие оценки называются выборочными. Тогда оценку параметра θ можно получить как функцию выборочного момента, например, $\theta^* = f(m,\sigma) = f(m_1^*, \sigma^2)$. Для этого достаточно теоретический момент приравнять эмпирическому моменту того же порядка, в результате чего получится уравнение, которое устанавливает связь между параметром θ и выборочным моментом. Если неизвестным является один параметр, то достаточно одного уравнения. В противном случае приходится решать систему уравнений, в которых участвуют разные моменты. Выбор моментов осуществляется экспериментально.

Пример. Вычислить точечные оценки неизвестных параметров a и b равномерного распределения, плотность которого $w(x) = \frac{1}{b-a}, a < x \le b.$

Решение. Поскольку неизвестных параметров два, то необходимо иметь два линейно независимых уравнения. Выражения для выбранных теоретических моментов дисперсии σ^2 и математического ожидания m имеют вид $m = \frac{a+b}{2}, \sigma^2 = \frac{(b-a)^2}{12}$ или $\begin{cases} a + b = 2m \\ b - a = 2\sqrt{3}\sigma \end{cases}$. Решая систему уравнений, получим $a = m - \sqrt{3}\sigma, b = m + \sqrt{3}\sigma$. Подставляя вместо m и σ их оценки, получим оценки параметров: $a^* = m^* - \sqrt{3}\sigma^*$ и $b^* = m^* + \sqrt{3}\sigma^*$.

Выборочные оценки $m^* = \frac{1}{n}\sum_{i=1}^{n} x_i, \sigma^* = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - m^*)^2}$. Следует отметить, что оценки, полученные методом моментов, не всегда являются оптимальными.

Выражение для дисперсии σ^2 можно получить следующим образом. Если случайную величину умножить на некоторый масштабный коэффициент k, то дисперсия изменится в k^2 раз. Дисперсия равномерного на отрезке $[0,1]$ закона распределения равна $\frac{1}{12}$. Поскольку дисперсия σ^2 не зависит от математического ожидания, то остается зависимость только от длины интервала $(b-a)$, которая является масштабным коэффициентом по отношению к случайной величине, определенной на единичном отрезке $[0,1]$. Поэтому дисперсия $\sigma^2 = \frac{(b-a)^2}{12}$.

3.2.3. Интервальные оценки параметров

Рассмотренные ранее оценки параметра Θ называются точечными оценками, поскольку результат оценки является единичным значением, которое не характеризует качество оценки. Точечная оценка является случайной величиной, так как вычисляется по случайным выборкам, которые образуют выборочное пространство с заданной вероятностной мерой. Поэтому свойства оценки должны описываться как свойства случайной величины.

Качество оценки характеризуется совокупностями следующих величин:

1. Точностью оценки $\Delta (0< \Delta)$; $\left|\Theta - \hat{\Theta}\right| \le \Delta$.
2. Вероятностью того, что отклонение значения оценки $\hat{\Theta}$ от истинного значения параметра Θ не превысит значение, равное $\Delta: P\left(\left|\Theta - \hat{\Theta}\right| \le \Delta\right) = \gamma$.
3. Размером выборки n.

В литературе величина 2Δ называется длиной доверительного интервала, γ – доверительной вероятностью или коэффициентом доверия.

Доверительная вероятность γ, точность оценки Δ и размер (объем) n выборки $\vec{x} = (x_1,...,x_n)$ связаны между собой. Если известны значения двух величин, то будет определена и третья. Определение зависимости между этими величинами является основной целью при решении задач на данную тему.

Неравенство $\left|\Theta - \hat{\Theta}\right| \le \Delta$ можно представить в виде двух эквивалентных неравенств:

$$\hat{\Theta} - \Delta \le \Theta \le \hat{\Theta} + \Delta,$$

$$\Theta - \Delta \le \hat{\Theta} \le \Theta + \Delta,$$

на основании которых можно дать следующее определение интервальной оценке.

Интервальной оценкой называется интервал, который с заданной вероятностью γ накрывает истинное значение параметра Θ. Границы интервала задаются как функции выборки $\vec{x} = (x_1, ..., x_n)$. В такой постановке задача не имеет единственного решения.

С целью получения единственного решения доверительный интервал строится следующим образом.

В качестве среденего значения интервала выбирается точечная оценка параметра $\hat{\Theta}$, при этом вся длина доверительного интервала выбирается равной 2Δ. В силу эквивалентности приведенных ранее неравенств вероятность γ также равна вероятности того, что точечная оценка $\hat{\Theta}$ попадет в интервал $[\Theta-\Delta, \ \Theta+\Delta]$.

Таким образом, для вычисления значения вероятности γ необходимо знать закон распределения оценки $\hat{\Theta}$, определение которого, чаще всего, является непростой задачей. В некоторых случаях удается аппроксимировать закон распределения оценки $\hat{\Theta}$ гауссовым законом и получить приближенные результаты.

В настоящее время получена интервальная оценка дисперсии гауссова закона распределения и интервальные оценки математического ожидания при известной и неизвестной дисперсии гауссова закона.

Интервальная оценка математического ожидания гауссова закона при известной дисперсии

Для математического ожидания доверительная вероятность $\gamma = p(\left|\hat{m} - m\right| \le \Delta)$, где m – истинное значение математического ожидания, $\hat{m} = \frac{1}{n}\sum_{i=1}^{n} x_i$ – его оценка, равная выборочному среднему. Оценка $\hat{m}$ имеет

гауссов закон распределения как сумма гауссовых случайных величин x_i и дисперсию, равную $\sigma^2_{\hat{m}} = \frac{\sigma^2}{n}$, где σ^2 - дисперсия случайных величин x_i. Вычислим значение доверительной вероятности γ. Вероятность того, что реализация $\hat{m}$ гауссовой случайной величины попадает в интервал $[a,b]$ записывается в виде: $P(a \le x \le b) = \Phi(\frac{b-m}{\sigma_{\hat{m}}}) - \Phi(\frac{a-m}{\sigma_{\hat{m}}})$, где $\Phi(\bullet)$ – интегральная функция распределения гауссовой случайной величины с нулевым средним значением и дисперсией, равной единице. В нашем случае $a = m - \Delta$, $b = m + \Delta$:

и поэтому $\gamma = \Phi(\frac{\Delta}{\sigma_{\hat{m}}}) - \Phi(-\frac{\Delta}{\sigma_{\hat{m}}}) = 2\Phi(\frac{\Delta}{\sigma_{\hat{m}}}) - 1$,

или с учетом равенства $\Phi(-\frac{\Delta}{\sigma_{\hat{m}}}) = 1 - \Phi(\frac{\Delta}{\sigma_{\hat{m}}})$ имеем $\gamma = 1 - \Phi(\frac{\Delta}{\sigma}\sqrt{n})$.

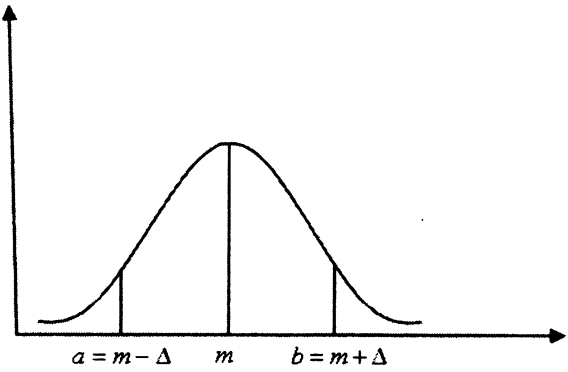


Рис. 3.1. Гауссов закон распределения

Введем следующее обозначение: $\frac{\Delta}{\sigma}\sqrt{n} = U_\gamma$. Отсюда находим $\Delta = U_\gamma \frac{\sigma}{\sqrt{n}}$.

Значение U_γ находится по таблицам из условия $\Phi(U_\gamma) = 0{,}5(1+\gamma)$.

Окончательно имеем:

$$P(\hat{m} - U_\gamma \frac{\sigma}{\sqrt{n}} \le m \le \hat{m} + U_\gamma \frac{\sigma}{\sqrt{n}}) = \gamma.$$

Доверительный интервал для среднего значения m нормального распределения при неизвестной дисперсии σ^2.

Пусть имеется гауссова случайная величина x с неизвестной дисперсией σ^2. Выборочной оценкой дисперии будет

$$S^2=\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\hat{m}_1)^2\ ;\ \hat{m}_1=\frac{1}{n}\sum_{i=1}^{n}x_i.$$

S - несмещенная оценка стандартного отклонения генеральной совокупности. Тогда случайная величина $t=\frac{\hat{m}_1-m}{S}\sqrt{n}$ имеет распределение Стьюдента (t - распределение) с числом степеней свободы, равным $n-1$.

Плотность этого закона распределения имеет вид

$$S_{n-1}(t)=\frac{\Gamma\left(\frac{n}{2}\right)}{\sqrt{(n-1)\pi}\,\Gamma\left(\frac{n-1}{2}\right)}\left(1+\frac{t^2}{n-1}\right)^{\frac{-n}{2}}.$$

$\Gamma(x)=\int_0^x U^{x-1}e^{-U}\,dU$ – гамма функция.

Построим доверительный интервал для математического ожидания

$$P(|\hat{m}_1-m|\le\Delta)=\gamma.$$

Умножим обе части неравенства на $\frac{\sqrt{n}}{s}>0$. Тогда $P(|\frac{\hat{m}_1-m)}{S}\sqrt{n}|\le\frac{\Delta\sqrt{n}}{S})=$ $=\gamma$ или $P(|t|\le t_\gamma)=\gamma$, где $t_\gamma=\frac{\Delta\sqrt{n}}{S}$; $\Delta=t_\gamma\frac{S}{\sqrt{n}}$.

Величина t_γ найдется из условия

$$P(|t|\le t_\gamma)=\int_{-t_\gamma}^{t_\gamma}S_{n-1}(t)\,dt=\gamma.$$

Так как S_{n-1} (t) – четная функция, то $2\int_0^{t_\gamma}S_{n-1}(t)\,dt=\gamma$. Это равенство определяет t_γ в зависимости от γ.

Значение t_γ определяется по таблицам распределения Стьюдента при заданном значении γ. Окончательно имеем:

$$P(\hat{m}_1-t_\gamma\frac{S}{\sqrt{n}}\le m\le\hat{m}_1+t_\gamma\frac{S}{\sqrt{n}})=\gamma.$$

Доверительный интервал указан в скобках.

Доверительный интервал для дисперсии σ^2 нормального распределения

Рассмотрим несмещенную оценку дисперсии

$$S^2=\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\hat{m}_1)^2,$$

от которой перейдем к случайной величине (статистике)

$$\chi^2=\frac{(n-1)S^2}{\sigma^2}>0.$$

Эта величина имеет χ^2- распределение (распределение Пирсона) с числом степеней свободы, равном $n-1$.

По таблице χ^2-распределения можно найти значение x_α, удовлетворяющее условию $P(\chi^2 > x_\alpha) = \alpha$ (рис. 3.2).

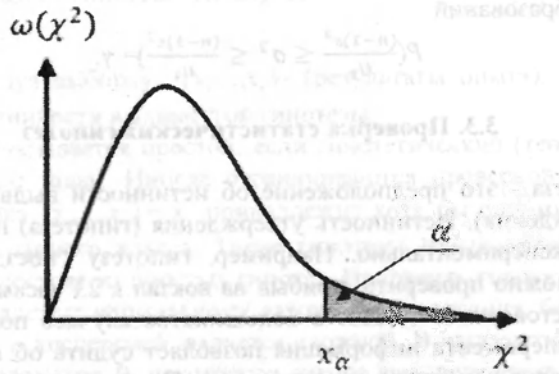


Рис. 3.2. Использование таблицы χ^2-распределения

По таблицам χ^2- распределения можно найти такие два числа U_1 и U_2, которые удовлетворяют условию

$$P(|U_1 \le \chi^2 \le U_2|) = \gamma.$$

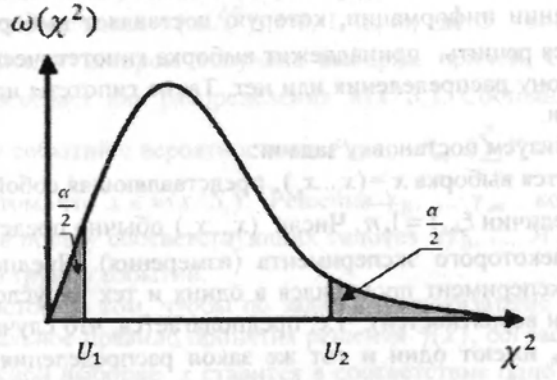


Рис. 3.3. Определение границ U_1 и U_2

Таких пар чисел U_1 и U_2 существует бесконечно много. Чтобы зафиксировать одну такую пару U_1 и U_2 введем дополнительные условия:

$$P(\chi^2 < U_1) = P(\chi^2 > U_2) = 0.5(1-\gamma).$$

Из таблицы находим U_2. Для нахождения U_1 используем вероятность противоположного события:

$$P(\chi^2 > U_1) = 1-0.5+0.5\gamma=0.5(1+\gamma).$$

Возвращаясь от χ^2 к s^2, имеем

$$P(U_1 \le \frac{(n-1)s^2}{\sigma^2} \le U_2) = \gamma.$$

После преобразований

$$P(\frac{(n-1)s^2}{U_2} \le \sigma^2 \le \frac{(n-1)s^2}{U_1}) = \gamma.$$

3.3. Проверка статистических гипотез

Гипотеза - это предположение об истинности выдвинутого утверждения (суждения). Истинность утверждения (гипотеза) проверяется чаще всего экспериментально. Например, гипотезу "поезд отправляется в 23 часа" можно проверить, прибыв на вокзал к 23 часам. Эта проверка является достоверной. Однако в большинстве случаев полученная в результате эксперимента информация позволяет судить об истинности гипотезы только с некоторой вероятностью. Поскольку мы рассматриваем задачу принятия решения об истинности гипотезы в рамках теории вероятностей, то класс гипотез, которые могут быть выдвинуты, ограничивается моделью теории вероятностей. Модель основного объекта, который изучает теория вероятностей, представляет собой случайную величину, т.е. множество возможных значений с заданной на нем вероятностной мерой. Поэтому гипотеза задается в виде определенного закона распределения случайной величины, а решение об истинности гипотезы выносится на основании информации, которую доставляет выборка $(x_1...x_n)=x$, т. е. требуется решить, принадлежит выборка гипотетическому (теоретическому) закону распределения или нет. Такие гипотезы называются статистическими.

Формализуем постановку задачи:

1. Имеется выборка $x=(x_1...x_n)$, представляющая собой n реализаций случайных величин ξ_i, $i=\overline{1,n}$. Числа $(x_1...x_n)$ обычно представляют собой результаты некоторого эксперимента (измерения). Предполагается, что все n раз эксперимент проводился в одних и тех же условиях (с точки зрения теории вероятностей), т.е. предполагается, что случайные величины ξ_1, ..., ξ_n имеют один и тот же закон распределения вероятностей $\overset{0}{W}_\xi(x)$. Закон $\overset{0}{W}_\xi(x)$ является истинным, но в рамках решаемой задачи неизвестен и подлежит оценке. Термин "оценка" имеет два значения. Первое отражает процесс оценивания, а второе – результат.

2. Задача исследователя состоит в определении закона $\overset{0}{W}_\xi(x)$.

3. Исходя из каких-либо предположений (физических, интуитивных и проч.) выносится предположение о законе распределения $\overset{0}{W}_{\xi}(x)$, а именно выдвигается гипотеза H: случайные величины ξ распределены по закону $\hat{W}_{\xi}(x)$.

4. Используя выборку $(x_1,...,x_n)$- (результаты опыта), оценивается вероятность истинности выдвинутой гипотезы.

Гипотеза называется простой, если гипотетический (теоретический) закон полностью задан. Иногда ограничиваются проверкой гипотезы о том, что выборка $(x_1,...,x_n)= x$ принадлежит хотя бы одному из законов распределения данного класса. Такие гипотезы называются сложными, поскольку они состоят из простых гипотез. Например, гипотеза о том, что выборка принадлежит нормальному закону распределения, безразлично с каким средним и дисперсией, является сложной. В частности, гипотезы о неизвестном параметре θ некоторого закона распределения могут быть простыми и сложными. Простая гипотеза утверждает, что параметр θ имеет одно конкретное значение $(\theta = \theta_0)$, а сложная гипотеза утверждает, что параметр θ имеет значение, принадлежащее некоторому множеству значений. Приведенные рассуждения показывают, что нет принципиальной разницы между проверкой гипотез и оценкой параметров.

Сформулируем задачу проверки $m + 1$ простых гипотез. Пусть известно, что выборка с вероятностью, равной единице, принадлежит одному из $m + 1$ распределений $w(x/S_i^{\text{Б}})$, i=0, 1, ..., m, где S_i - состояние комплекса условий при котором получена выборка, причем, состояние S_i полностью определяет вид распределения $w(x/S_i)$. Состояния образуют полную группу событий с вероятностями P_0, ..., P_m $(\sum_{k=0}^{m} P_k = 1)$. Гипотеза H_i состоит в том, что $x \in w(x/S_i)$. Решения γ_0, ..., γ_m, которые могут быть приняты в пользу соответствующих гипотез H_0, ..., H_m также образуют полную группу событий.

Задача состоит в том, чтобы по заданному показателю качества построить оптимальное правило принятия решения $\gamma(x)$, согласно которому каждой возможной выборке x ставится в соответствие одно из решений. Это означает, что пространство выборок X должно быть разделено на $m + 1$ непересекающихся областей X_0, X_k, X_m, каждой из которых ставится в соответствие одно из решений $\gamma_0,...,\gamma_m$. Следует отметить, что правило $\gamma(x)$ строится до наблюдения выборки x. Правило $\gamma(x)$, по которому каждой области X_k всегда ставится в соответствие определенное

решение $\gamma_k(\gamma(x)=\gamma_k, x\in X_k)$, называется детерминированным или нерандомизированным (рис. 3.4). Правило $\gamma(x)$ называется рандомизированным, если при попадании x в данную область X_j допускается выбор одного из нескольких решений в соответствии с некоторым распределением вероятностей.

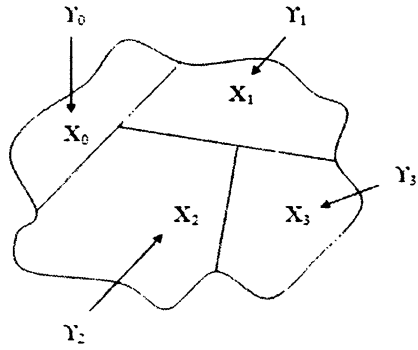


Рис. 3.4. Разделение пространства x на 4 области (m=3), γ_i - решение в пользу гипотезы H_i

Рассмотрим более подробно случай двух гипотез (m=1)

$$H_0: \quad x\in w(x/S_0),$$
$$H_1: \quad x\in w(x/S_1).$$

Гипотезу H_1 часто называют альтернативной гипотезой или просто альтернативой.

Решение γ_0 - принятие гипотезы H_0, решение γ_1 - принятие гипотезы H_1. На практике часто встречается задача, когда H_1: $x\in W(x/S)$, т.е. гипотеза состоит в выдвижении какого-либо предположения (в данном случае $x\in W(x/S)$, а альтернатива в отклонении (не подтверждении) этого предположения ($x\notin W(x/S)$)

Правило принятия решения эквивалентно разделению n-мерного пространства выборок $\boldsymbol{X}$ на две непересекающихся области X_0 и X_1 (рис. 5). Если $x\in X_0$, то принимается решение γ_0, если $x\in X_1$, то принимается решение γ_1. Область X_0 принятия гипотезы H_0 называется допустимой. Область X_1 отклонения гипотезы H_0 называется критической.

Рис. 3.5. Разбиение пространства X на две области (m=1)

В процессе принятия решения могут возникнуть ошибки, которые удобно изобразить в виде диаграммы, где q - априорная вероятность гипотезы H_0, p - априорная вероятность гипотезы H_1; α - вероятность ошибки первого рода или уровень значимости, т. е. вероятность отклонения гипотезы H_0, когда она истинна; β - вероятность ошибки второго рода, т. е. вероятность принятия гипотезы H_0, когда она ложна; $1-\beta$ - мощность правила принятия решения.

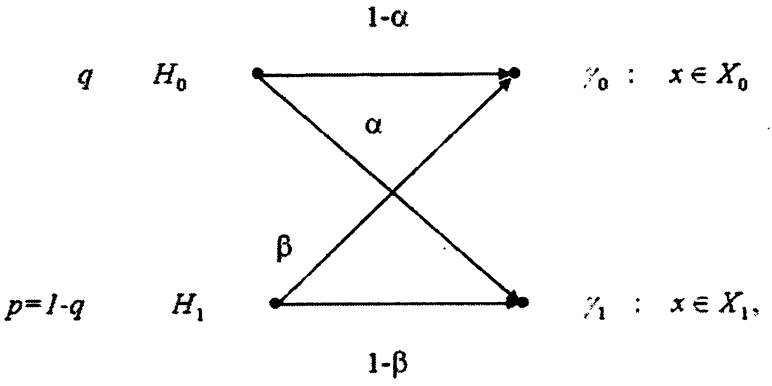


Рис. 3.6. Вероятностная диаграмма

По формуле полной вероятности можно вычислить безусловные вероятности

$$p(\gamma_0) = q(1-\alpha) + p\beta,$$

$$p(\gamma_1) = q\alpha + p(1-\beta) \ .$$

Перейдем к построению правила выбора решения $\gamma(x)$ (т.е. решающей схемы или способа разбиения пространства выборок X на допустимую X_0 и критическую X_1 области), которое будем считать оптимальным, если оно обеспечивает минимум средней вероятности ошибки (максимум вероятности правильного принятия решения). Этот критерий качества принятия (выбора) решения был предложен В.А.Котельниковым и назван критерием идеального наблюдателя, а приемник, работающий в соответствии с этим правилом, В.А. Котельников назвал идеальным.

В математической литературе правило выбора решения часто называют критерием. Можно показать, что средняя вероятность ошибки будет минимальна, если решение выносить в пользу гипотезы, имеющей наибольшую апостериорную вероятность $W(H_i / x)$ при заданной выборке $\vec{x}$. В нашем случае апостериорные вероятности для S_0 и S_1, что то же самое для H_0 и H_1, можно представить в виде:

$$p(S_0 / x) = \frac{qw(x/S_0)}{qw(x/S_0) + pw(x/S_1)} = \frac{1}{1 + \frac{p}{q} l(x)} \ ,$$

$$p(S_1 / x) = \frac{pw(x/S_1)}{qw(x/S_0) + pw(x/S_1)} = \frac{\left(\frac{p}{q}\right) l(x)}{1 + \left(\frac{p}{q}\right) l(x)},$$

где $l(x) = \dfrac{w(x/S_1)}{w(x/S_0)}$ - отношение правдоподобия.

Вывод представленных равенств становится более наглядным, если воспользоваться диаграммой вероятностных переходов (рис.3.7).

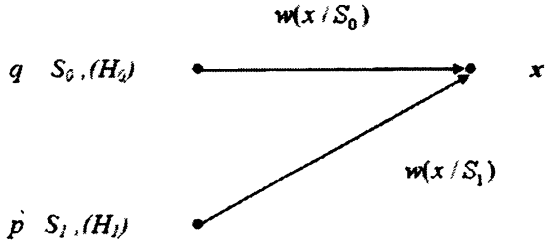


Рис. 3.7. Вероятностная диаграмма

Безусловная вероятность появления некоторой выборки x вычисляется по формуле полной вероятности:

$$p(x) = qw(x/S_0) + pw(x/S_1),$$

а апостериорные вероятности $p(S_i/x)$ (i=1, 2) по формуле Байеса:

$$p(S_0^{\text{Б}}/x) = \frac{qw(x/S_0)}{p(x)},$$

$$p(S_1^{\text{Б}}/x) = \frac{pw(x/S_1)}{p(x)},$$

где $p(x)$ нужно заменить на полученное ранее выражение.

Теперь можно сформулировать правило выбора решения: принимается гипотеза H_1 (отвергается гипотеза H_0), если для наблюдаемой выборки x выполняется неравенство

$$p(S_1/x) \geq p(S_0/x),$$

и принимается гипотеза H_0 (отвергается гипотеза H_1), если

$$p(S_0^{\text{Б}}/x) \geq p(S_1/x).$$

Для сравнения апостериорных вероятностей можно взять их отношение

$$\frac{p(S_1/x)}{p(S_0/x)} = \frac{p}{q} l(x).$$

Тогда условие принятия гипотезы H_1 можно записать в виде:

$$l(x) \geq \mu, \mu = \frac{q}{p}.$$

Таким образом, максимуму апостериорной вероятности соответствует такая критическая область пространства выборок, точки (x) которой удовлетворяют полученному неравенству. Вероятности p и q учитывают априорные сведения о гипотезах. Если гипотезы равновероятны, то μ=1 и правило выбора решения сводится к проверке неравенства $l(x) \geq 1$ для полученной выборки x.

Это правило называется правилом максимального правдоподобия и часто используется, когда вероятности гипотез неизвестны, т.е. гипотезы считаются равновероятными.

Принятие решения по правилу максимального правдоподобия совпадает с правилом максимума апостериорной вероятности удовлетворяющим критерию идеального наблюдателя, когда p=q. Итак, если для наблюдаемой выборки x выполняется неравенство

$$\ell(x) \geq \mu,$$

то принимается решение γ_1 (отвергается гипотеза H_0) и принимается решение γ_0 (гипотеза H_0 считается истинной) в противном случае. Скаляр-

ная величина $\ell(x)$, которая ставится в соответствие каждой выборке $\vec{x}$, называется пороговой статистикой. Выборки x, которые удовлетворяют неравенству $\ell(x) \geq \mu$, образуют критическую область X_1 данного правила выбора решения. Для вычисления вероятности ошибки первого рода

$$\alpha = p(\gamma_1 / H_0) = p(x \in X_1 / S_0) = \int_{X_1} w(x/S_0)dx = \int_{\mu}^{\infty} w(\ell / S_0)d\ell,$$

которая равна вероятности того, что случайная величина $\ell(x)$ превысит порог μ, необходимо знать ее условный закон распределения $w(\ell / S_0)$, который не всегда удается легко установить.

Аналогично вычисляется вероятность ошибки второго рода:

$$\beta = p(\gamma_0 / H_1) = p(x \in X_0 | S_1) = \int_{X_0} w(x|S_1)dx = \int_{-\infty}^{\mu} w(\ell | S_1^{\text{Б}})d\ell.$$

Можно доказать, что правило принятия решения, основанное на вычислении отношения правдоподобия, обеспечивает минимум средней (полной, безусловной) вероятности ошибки, которая равна

$$P_{\text{ош}} = q\alpha + p\beta.$$

Часто принятие гипотез основано на другом критерии - критерии Неймана - Пирсона, согласно которому правило выбора решения строится таким образом, чтобы обеспечить минимально возможную величину вероятности ошибки второго рода β при условии, что вероятность ошибки первого рода не больше заданной величины α. Иначе говоря, это правило имеет наибольшую мощность $(1-\beta)$ среди всех других правил, для которых уровень значимости не превосходит α. Критерий Неймана - Пирсона широко используется в радиолокации, в пожарном деле и проч., где α принято называть вероятностью ложной тревоги, а $(1-\beta)$ - вероятностью правильного обнаружения (цели, пожара ...).

Пороговой статистикой критерия Неймана - Пирсона также является отношение правдоподобия $\ell(x)$, которое сравнивается с некоторым пороговым значением C. Величина C выбирается таким образом, чтобы вероятность ее превышения пороговой статистикой $\ell(x)$ (вероятность выполнения неравенства $\ell(x)>C$) не превышала заданное значение α.

Таким образом, в случае простых гипотез H_0 и H_1 (альтернатива) алгоритм выбора решения сводится к вычислению отношения правдоподобия и сравнению его значения с порогом C.

Задача значительно усложняется, когда хотя бы одна из гипотез является сложной. Например, для проверки простой гипотезы H_0 о том, что случайная величина имеет заданный (гипотетический) закон распределения против сложной альтернативы H_1, о которой ничего не известно, разработаны специальные критерии, которые называются критериями согла-

сия. Эти критерии позволяют судить, насколько экспериментальные данные согласуются с предполагаемым (теоретическим, гипотетическим) законом распределения (гипотезой H_0).

Самым простым методом оценки является визуальное сравнение гистограммы, полученной по выборке x, и гипотетического закона распределения $\hat{w}(x)$. Однако этот метод является очень грубым, субъективным, не имеет четкой количественной характеристики и поэтому не может быть использован как критерий в строгих научных исследованиях. Его рекомендуется применять в качестве основы для выдвижения гипотез. Рассмотрим некоторые другие критерии, которые широко используются в настоящее время.

Критерий согласия Колмогорова основан на проверке близости между эмпирической (выборочной) функцией распределения - $F_n(x)$ и гипотетической функцией распределения - $F(x)$. В данном случае гипотеза H_0 предполагает, что истинная функция распределения $G(x)$ равна гипотетической функции распределения $F(x)$:

$$H_0: G(x)=F(x).$$

Для количественного выражения сходства функций $F_n(x)$ и $F(x)$ используется статистика Колмогорова:

$$D_n = \sup_{-\infty<x<\infty} |F_n(x) - F(x)|.$$

Функция $\sup_{x\in E}(\mu(x))$ означает нижнюю грань функции $\mu(x)$ на допустимом множестве аргумента $x \in E$.

Очевидно, что D_n - случайная величина, поскольку ее значение зависит от случайной функции $F_n(x)$. Если гипотеза H_0 справедлива, то при неограниченном возрастании n функция $F_n(x) \to F(x)$ при всяком x и $D_n \to 0$. Если же гипотеза H_0 не верна, то $F_n(x) \to G(x)$ и $G(x) \neq F(x)$, а $\sup_x |F_n(x) - F(x)| \to \sup_x |G(x) - F(x)|$. Последняя величина положительна, так как $G(x)$ не совпадает с $F(x)$.

Отсюда следует, что величина D_n имеет тенденцию к увеличению с ростом степени различия между истинной функцией распределения $G(x)$ и гипотетической функцией распределения $F(x)$. Это свойство позволяет использовать $\sqrt{n}D_n$ в качестве пороговой статистики критерия. Поскольку $D_n \to 0$ при истинности гипотезы H_0, то для стабилизации закона распределения пороговой статистики D_n умножается на неограниченно растущую величину $\sqrt{n}$. Случайная величина D_n обладает замечательным свойством, которое заключается в том, что её закон распределения оказывается одним и тем же для всех непрерывных функций $G(x)$. Он зависит только от объема выборки n. Доказательство этого факта основано на том,

что статистика D_n не изменяет своего значения при монотонных преобразованиях оси x. Таким преобразованием любое непрерывное распределение $G(x)$ можно превратить в равномерное на отрезке [0,1]. При этом $F_n(x)$ перейдет в функцию распределения выборки из этого равномерного распределения.

При малых n для статистики D_n при гипотезе H_0 составлены таблицы процентных точек, например, в [1] они доведены до n=100. При больших n распределение D_n (при гипотезе H_0) указывает найденная в 1933 г. А.Н. Колмогоровым предельная теорема, которая утверждает, что при справедливости H_0 (и если $G(x)$ непрерывна) вероятность $P(\sqrt{n}D_n < z)$ при неограниченном возрастании n имеет предел, и дает его выражение:

$$\lim_{n\to\infty} P(\sqrt{n}D_n < z) = 1 + 2\sum_{k=1}^{\infty}(-1)^k e^{-2k^2z^2} \quad ,$$

или

$$\lim_{n\to\infty} p(\sqrt{n}D_n \geq Z) = 1 - \sum_{k=-\infty}^{\infty}(-1)^k e^{-2k^2z^2} .$$

Для этих выражений имеются соответствующие таблицы, например, [1]. Таким образом, задав значение вероятности ошибки первого рода α, можно определить величину порога, с которой будет сравниваться значение пороговой статистики $\sqrt{n}D_n$ для вынесения решения об истинности гипотезы H_0. Статистика D_n вычисляется по формуле

$$D_n = \max_{1\leq k\leq n}\left[\frac{k}{n} - F(x_{(k)}), F(x_{(k)}) - \frac{k-1}{n}\right],$$

где через $x_{(1)}$, $x_{(2)}$, ... $x_{(n)}$ обозначены элементы вариационного ряда, построенного по исходной выборке, т.е., если элементы выборки расположить в порядке возрастания их значений, то $x_{(i)}$ обозначает элемент выборки, который стоит на i-м месте.

Критерий согласия омега-квадрат основан на статистике

$$\omega_n^2 = \int_{-\infty}^{\infty}[F_n(x) - F(x)]^2 dF(x),$$

которая измеряет расстояние между $F_n(x)$ и $F(x)$ в интегральной метрике.

Для вычисления ω_n^2 по реальной выборке можно использовать формулу:

$$n\omega_n^2 = \frac{1}{12n} + \sum_{i=1}^{n}\left[F(x_{(i)}) - \frac{2i-1}{2n}\right]^2 .$$

При справедливости гипотезы H_0: $F(x)=G(x)$ и непрерывности функции $G(x)$ закон распределения статистики ω_n^2, так же как закон распределения статистики D_n, зависит только от n и не зависит от G. Так же как для D_n,

для ω_n^2 при малых значениях n имеются таблицы процентных точек, а для больших значений n следует использовать предельный (при $n\to\infty$) закон распределения статистики $n\ \omega_n^2$. Предельный закон распределения был найден Н.В.Смирновым в 1939 г. [1].

Следует отметить, что критерии, основанные на D_n и ω_n^2, состоятельны против любой альтернативы $G(x)\neq F(x)$. Статистический критерий для проверки гипотезы H_0 называется состоятельным против альтернативы H_1, если его мощность стремится к единице при неограниченном увеличении объема выборки.

Рассмотренные критерии имеют ограниченную область применения, поскольку требуют непрерывности функции $F(x)$. Например, они не применимы для дискретных законов распределения. Поэтому полезно познакомиться с более универсальным критерием К.Пирсона (1900 г.). Он основан на сравнении гипотетического (теоретического) закона распределения $w(x)$ и гистограммы, построенной по выборке x.

Критерий К. Пирсона

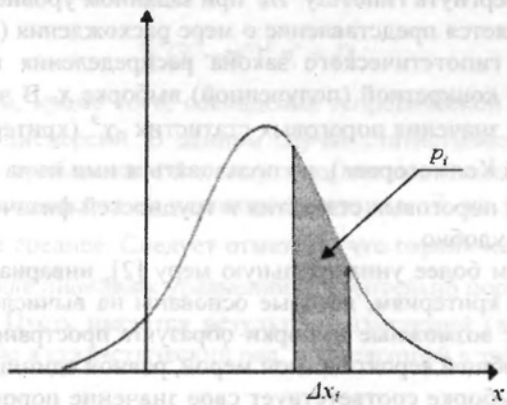


Рис. 3.8. Гипотетический закон распределения

В качестве пороговой статистики, которая называется статистикой хи-квадрат Пирсона для простой гипотезы, используется величина:

$$\lambda = n\sum_{i=1}^{k}\frac{(p_i^* - p_i)^2}{p_i},$$

(случайная величина λ имеет распределение вероятностей χ^2, поэтому в левой части последнего равенства пишут χ^2 и критерий называют крите-

рием хи-квадрат (χ^2)), где p_i - вероятность (теоретическая, гипотетическая) того, что случайная величина попадет в интервал Δx_i;

$p_i^* = \frac{m_i}{n}$ - оценка истинного значения вероятности (относительная частота);

n - общее количество испытаний (размер выборки);

m_i - количество испытаний, при которых значения случайных величин попали в интервал Δx_i;

k - количество интервалов на которое разделена область определения случайной величины.

На практике пороговую статистику удобнее вычислять по формуле

$$\chi^2 = \sum_{i=1}^{k} \frac{(m_i - np_i)^2}{nP_i}.$$

К.Пирсон доказал, что статистика χ^2 асимптотически подчиняется распределению χ^2 (хи-квадрат) с (k-1) степенями свободы.

Рассмотренные правила проверки статистических гипотез позволяют принять или отвергнуть гипотезу H_0 при заданном уровне значимости α, но при этом теряется представление о мере расхождения (или степени согласованности) гипотетического закона распределения и гистограммы, построенной по конкретной (полученной) выборке x. В частности, такой мерой являются значения пороговых статистик χ^2 (критерий Пирсона) и $\sqrt{n}D_n$ (критерий Колмогорова), но пользоваться ими из-за большого количества значений пороговых статистик и трудностей физической интерпретации не всегда удобно.

Рассмотрим более универсальную меру [2], инвариантную по отношению ко всем критериям, которые основаны на вычислении пороговой статистики. Все возможные выборки образуют пространство элементарных событий с общей вероятностной мерой, равной единице.

Каждой выборке соответствует свое значение пороговой статистики (χ^2 или $\sqrt{n}D_n$), характеризующее степень согласованности (совпадения) теоретического и статистического законов распределения, причем чем больше значение пороговой статистики, тем меньше степень согласованности.

Полученная (эмпирическая) выборка, которой соответствует некоторое значение пороговой статистики, например, χ_0^2, является граничной между выборками, которым соответствует значение $\chi^2 < \chi_0^2$, и выборками, которым соответствуют значения $\chi^2 < \chi_0^2$. Поэтому вероятностную меру

$p(\chi^2 > \chi_0^2)$ выборок, которые "хуже" полученной, можно использовать в качестве искомой меры согласованности законов распределения.

При вычислении вероятности $p(\chi^2 > \chi_0^2)$ необходимо учитывать зависимость закона распределения χ^2 от параметра r, называемого числом степеней свободы распределения. Число степеней свободы равно числу интервалов k минус число независимых ограничений (связей) наложенных на частоты p_i^*.

Примерами таких ограничений могут быть

$$\sum_{i=1}^{k} p_i^* = 1,$$

если мы требуем только того, чтобы сумма частот была равна единице (это требование накладывается во всех случаях);

$$\sum_{i=1}^{k} \tilde{x}_i p_i^* = m,$$

если мы подбираем теоретическое распределение таким образом, чтобы его математическое ожидание m совпало со статистическим средним значением;

$$\sum (\tilde{x}_i - m^*)^2 p_i^* = D_x,$$

если мы требуем, кроме того, совпадения теоретической дисперсии D_x и статистической дисперсий. В данном случае статистическая оценка дисперсии является приближенной в силу того, что все значения, попавшие в интервал Δx_i, заменяются на некоторое значение $\tilde{x}_i$ из этого интервала, m^* - выборочное среднее. Следует отметить, что ограничения на p_i^* задаются только в виде линейных уравнений относительно переменных p_i^*.

Пример. Пусть имеются результаты измерений (выборка размера n=500) сведенные в статистический ряд, приведенный в табл. 3.2.

Таблица 3.2

Δx_i	-4; -3	-3; -2	-2; -1	-1; 0	0; 1	1; 2	2; 3	3; 4
m_i	6	25	72	133	120	88	46	10
P_i^*	0,012	0,05	0,144	0,266	0,240	0,176	0,092	0,02

где $\Delta x_i (i = 1, \vec{k})$ интервалы, на которые разбита область определения случайной величины; m_i - количество реализаций (значений) случайной величины,

попавших в интервал Δx_i; $p_i^* = \frac{m_i}{n}$ - относительная частота попадания значений случайной величины в заданный интервал Δx_i; при этом $\sum_{i=1}^{k} m_i = n$.

Графически статистический ряд оформляется в виде гистограммы.

Проверим согласованность статистического распределения и теоретического (гипотетического), в качестве которого возьмем нормальный (гауссовский) закон распределения с параметрами, значения которых m = = 0,168 и σ = 1,448 совпадают с их статистическими оценками.

Зная теоретический закон распределения, можно найти теоретические вероятности попадания случайной величины в каждый из интервалов Δx_i:

$$p_i = \phi\left(\frac{x_{i+1} - m}{\sigma}\right) - \phi\left(\frac{x_i - m}{\sigma}\right),$$

где x_i, x_{i+1} - границы i-го интервала,

$$\phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{1}{2}t^2} dt.$$

В таблице приведены значения функции $\phi_0(Z) = \frac{1}{2\pi} \int_0^z e^{-\frac{1}{2}t^2} dt$,

причем $\phi(Z) = 0{,}5 + \phi_0(Z)$.

Пороговая статистика χ_0^2 в этом случае равна

$$\chi_0^2 = \sum_{i=1}^{8} \frac{(m_i - np_i)^2}{nP_i} = 3{,}94.$$

Если мы подбираем теоретическое распределение с тем условием, чтобы его математическое ожидание и дисперсия совпали с их статистическими оценками, то число степеней свободы r = 8-3 =5. По таблицам при r=5 находим, что

$$p(\chi^2 > \chi_0^2) \approx 0{,}56.$$

Если же на математическое ожидание и дисперсию не накладывать ограничений, а задать их значения априорно, то r = 8-1 = 7. Тогда при тех же значениях параметров (m=0,168 , σ=1,448) , но заданных априорно, $p(\chi^2 > \chi_0^2) \approx 0{,}8$.

4. ИНФОРМАЦИЯ

4.1. Модели, используемые в статистической теории информации

В основе любой теории лежит соответствующая модель подлежащей изучению части реального мира. Область применения результатов теории ограничена областью применения принятой модели. Мы рассмотрим модель, которая лежит в основе статистической теории информации [6]. Существуют и другие модели, на основе которых строятся невероятностные теории информации. Однако в данном курсе они рассматриваться не будут, за исключением прагматической (ценностной) теории информации, которая может быть построена в рамках статистической теории информации.

Понятие *информация* тождественно понятию *сведения* и ассоциирует с наличием по крайней мере двух взаимодействующих систем А и В, одна из которых В является наблюдаемой системой (приемником), а вторая А — источником информации. Вне указанной схемы понятие *информация* теряет смысл.

Любая система описывается совокупностью физических величин, которые могут зависеть от параметров. Состояния системы — это значения физической величины или параметра, которые ее описывают. Если эти значения дискретны, то система называется дискретной, а если непрерывны, то система называется системой с непрерывным множеством состояний. Таким образом, в рамках прикладной теории информации, информация – это сведения о состоянии системы.

Сообщение — это то, что можно сообщить, а сообщить можно только состояние системы. Следовательно, сообщение — это сообщаемое (передаваемое) состояние системы, которое сообщается в той или иной форме.

Система случайным образом с некоторой вероятностью может оказаться в том или другом состоянии (передатчик приходит в состояние, которое соответствует передаваемой букве). Следовательно, множество состояний системы можно рассматривать как множество случайных событий.

Две системы будем называть статистически зависимыми, если состояние одной из них влияет на вероятность состояния другой.

Множества состояний X и У соответственно систем A *и* B в зависимости от того, в каком отношении они рассматриваются, можно интерпретировать как множества состояний, сообщений и событий.

Два множества X и Y с заданным на них двумерным распределением $P(x_i,y_j)(x_i \in X, y_j \in Y, i=\overline{1,m_x}, j=\overline{1,m_y})$ представляют собой модель двух взаимодействующих систем. Эта модель лежит в основе построения статистической теории информации.

Сигнал — это материальный переносчик информации в пространстве и во времени.

Сигналы могут быть динамическими и статическими. Динамические сигналы предназначены для передачи информации в пространстве (электромагнитная волна). Статические сигналы (запоминающие устройства) предназначены для передачи информации во времени (магнитная лента, книга, кинофильм и т. д.). Точнее, сигналом является не сам материальный переносчик информации, а его состояние. Поэтому целесообразно конкретизировать определение сигнала. Сигнал — это значение физической величины, которое отображает состояние источника сообщений. Поскольку множество сообщений можно рассматривать как множество случайных событий, то отображающее значение физической величины также будет случайным.

Следовательно, случайную величину можно принять в качестве модели сигнала. В общем случае состояние системы (передаваемое сообщение) изменяется во времени, поэтому указанная случайная величина также будет изменяться во времени, зависеть от времени. Случайная величина, зависящая от времени (некоторого параметра), называется случайной функцией. Следовательно, случайная функция является моделью сигнала.

4.2. Установление количественной меры информации

4.2.1. Комбинаторное определение количества информации

Комбинаторное определение количества информации дано американским инженером Р.Хартли. Это определение предполагает модель с детерминированной связью (помехи отсутствуют) между дискретными состояниями двух систем без их вероятностного описания.

До получения сведений о состоянии системы имеется априорная неопределенность ее состояния. Сведения позволяют снять эту неопределенность, то есть определить состояние системы. Поэтому количество информации можно определить как меру снятой неопределенности, которая растет с ростом числа состояний системы.

Количественная мера информации устанавливается следующими аксиомами.

Аксиома 1. Количество информации, необходимое для снятия неопределенности состояния системы, представляет собой монотонно возрастающую функцию числа состояний системы.

В качестве количественной меры информации можно выбрать непосредственно число состояний системы m_x, которое является единственной характеристикой множества X.

Однако такое определение не удобно с точки зрения его практического применения. Поэтому в теории информации вводится несколько иная

количественная мера информации, которая является функцией m_x. Вид указанной функции позволяет установить аксиома 2.

Аксиома 2. Неопределенность состояния сложной системы, состоящей из двух подсистем, равна сумме неопределенностей подсистем.

Если для снятия неопределенности первой подсистемы, необходимо количество информации, равное $I(m_1)$, а для второй подсистемы количество информации, равное $I(m_2)$, то для снятия неопределенности сложной системы необходимо количество информации, равное:

$$I(m_1 m_2)=I(m_1)+I(m_2),$$

где m_1 — число состояний первой подсистемы; m_2 — число состояний второй подсистемы; $m_1 m_2$ — число состояний сложной системы.

Единственным решением полученного функционального уравнения является логарифмическая функция $I(m) = K\log_a m$, которая определяет количество информации как логарифм числа состояний системы. Произвольный коэффициент K выбирается равным единице, а основание логарифма a определяет единицу измерения количества информации. В зависимости от значения a единицы измерения называются двоичными $(a=2)$, троичными $(a=3)$ и в общем случае a -ичными. В дальнейшем под символом log будем понимать двоичный логарифм. Двоичная единица иногда обозначается *bit* (от английского *binary digit*— двоичный знак).

Каждое передаваемое слово из n букв, записанное в алфавите, содержащем m букв, можно рассматривать как отдельное «укрупненное» состояние источника сообщений. Всего таких состояний (слов) будет m^n.

Тогда количество информации, которое несет слово из n букв, равно $I=\log_a m^n = n\log_a m$. Отсюда следует, что одна буква несет $\log_a m$ a-ичных единиц информации. Если единица измерения информации $a=m$, то количество информации в слове $(I=n)$ измеряется количеством содержащихся в нем букв, а единица измерения информации определяется размером алфавита m. Таким образом, одна m-ичная единица содержит $\log_a m$ a -ичных единиц информации.

4.2.2. Определение количества информации по К. Шеннону

Согласно комбинаторному определению количества информации для установления записанного в регистр двоичного числа, имеющего n разрядов, требуется n двоичных единиц информации (по одной двоичной единице или по одному двоичному вопросу на выяснение содержания каждого разряда). Определить записанное в регистр число посредством задания меньшего числа вопросов, получив меньшее количество информации, невозможно, если мы об этом числе ничего, кроме того, что оно записано в регистр, не знаем. Количество необходимой информации можно уменьшить только в том случае, если мы будем распола-

гать некоторыми априорными сведениями о числе, в частности, о способе его записи (генерации).

Допустим, некоторое устройство вырабатывает (генерирует) число как независимую последовательность из единиц и нулей, которые появляются соответственно с вероятностями, равными p и $q=1—p$. В этом случае при неограниченном возрастании длины последовательности n с вероятностью, равной единице, появляются последовательности, количество единиц в которых незначительно отличается от среднего значения, равного np. Такие последовательности называются типичными. Они различаются между собой в основном только размещением единиц, а не их количеством. Поскольку количество типичных последовательностей Q меньше общего количества последовательностей, то имеется возможность уменьшить количество информации, необходимое для определения числа.

Последовательность назовем типичной для заданного источника, если количество единиц n в ней удовлетворяет неравенству

$$\left|\frac{n_1}{n}-p\right| < \varepsilon \text{ или } |n_1-np| < n\varepsilon \qquad (4.1)$$

и нетипичной в противном случае, то есть когда

$$|n_1-np| \geq n\varepsilon. \qquad (4.2)$$

Вероятность появления нетипичной последовательности равна вероятности, с которой n_1 удовлетворяет неравенству (4.2). Для оценки этой вероятности воспользуемся неравенством Чебышева, которое для произвольной случайной величины ξ, имеющей конечную дисперсию, при каждом $b>0$ записывается в виде

$$P\{|\xi-\bar{\xi}| \geq b\} \leq \frac{\sigma^2_\xi}{b^2},$$

где $\bar{\xi}$ и σ^2_ξ — соответственно математическое ожидание и дисперсия случайной величины ξ. Полагая $\xi=n_1$, $\bar{\xi}=\overline{n_1}=np$, $b=\varepsilon n$, $\sigma^2_\xi=\sigma^2_{n1}=npq$ ($q=(1—p)$), получим аналогичное неравенство для случайного числа единиц n_1

$$P\{|n_1-np| \geq n\varepsilon\} \leq \frac{pq}{\varepsilon^2 n}.$$

Следовательно, вероятность появления нетипичной последовательности

$$P_{нт} \leq \frac{pq}{\varepsilon^2 n},$$

а вероятность появления типичной последовательности

$$P_т = 1-P_{нт} > 1-\frac{pq}{\varepsilon^2 n}.$$

Вероятность $P_{нт}$ стремится к нулю, а вероятность $P_т$ стремится к единице при любом сколь угодно малом значении ε и неограниченном возрастании длины последовательности *n*. Интервал $[-n\varepsilon, n\varepsilon]$, которому принадлежит количество единиц в типичной последовательности, неограниченно увеличивается $(n\varepsilon \to \infty)$, хотя относительная величина интервала всегда меньше значения ε. Докажем, что одновременно с неограниченным увеличением длины последовательности *n* можно уменьшать значение $\varepsilon = \varepsilon(n)$ с такой скоростью, при которой относительная величина интервала будет стремиться к нулю, а вероятность появления типичной последовательности — к единице. При этом абсолютная величина интервала по-прежнему неограниченно возрастает. Вероятность $P_т$ стремится к единице, если величина $\varepsilon^2(n)n$ неограниченно увеличивается с ростом *n*. Пусть $\varepsilon^2(n)n = n^{2\delta}$, где $\delta > 0$ — некоторый параметр, определяющий скорость роста величины $\varepsilon^2(n)n$. Отсюда $\varepsilon(n) = n^{-0,5+\delta}$, $(0 < \delta < 0,5)$.

Величина $\varepsilon(n)$стремится к нулю с ростом *n* при $\delta < 0,5$. При этом абсолютная величина интервала $n\varepsilon(n)$ не может быть постоянной или стремиться к нулю одновременно с неограниченным увеличением величины $\varepsilon^2(n)n$, стремлением $\varepsilon(n)$ к нулю.

Определим количество типичных последовательностей *Q*.

Вероятность появления произвольной последовательности B_k равна

$$P(B_k)^{n_1} q^{n-n_1}.$$

В результате тождественных преобразований

$$P(B_k) = p^{np+(n_1-np)} q^{nq+(n-n_1-nq)} = p^{np} q^{nq} \left(\frac{p}{q}\right)^{(n_1-np)}.$$

Прологарифмировав последнее равенство, получим

$$\log_m P(B_k) = np \log_m p + nq \log_m q + (n_1 - np) \log_m \frac{p}{q} =$$

$$= -n[-p \log_m p - q \log_m q - \frac{n_1 - np}{n} \log_m \frac{p}{q}] = -n[H(x) - O(n)],$$

где величин $H(x) = -p \cdot \log_m p - q \log_m q$ является характеристикой источника сообщений и называется энтропией.

Покажем, что в случае типичных последовательностей остаточным членом *O(n)* по сравнению с величиной *H(X)* можно пренебречь.

Поскольку для типичных последовательностей справедливо неравенство $|n_1 - np| < n\varepsilon$, то

$$|O(n)| < \varepsilon(n)\log_m \frac{p}{q} = n^{-0.5+\delta} \log_m \frac{p}{q}.$$

Следовательно,

$$\lim_{n\to\infty} |O(n)| = 0 \qquad (p \neq 0, q \neq 0).$$

Таким образом, при достаточно большом n справедливо приближенное равенство $\log_m P(B_k) \approx -nH(x)$.

Отсюда вероятность появления отдельной типичной последовательности

$$P(B_k) \approx m^{-nH(X)}.$$

Поскольку правая часть равенства не зависит от номера типичной последовательности k, то все типичные последовательности примерно равновероятны. Вероятность появления типичной последовательности

$$Pm = \sum_{k=1}^{Q} P(B_k) = Qm^{-nH(X)} \approx 1,$$

где суммирование ведется по всему множеству типичных последовательностей. Отсюда:

$$Q \cdot m^{nH(X)},$$

причем единица измерения энтропии $H(X)$ совпадает с основанием степени m. Поскольку количество информации, необходимое для определения состояния регистра, равно $\log Q$, то энтропия $H(X) = \frac{\log Q}{n}$ равна количеству информации, которое необходимо для определения состояния одного разряда.

Аналогично определяется количество типичных последовательностей, вырабатываемых источником с алфавитом размера m_x, только в этом случае энтропия $H(X) = -\sum_{i=1}^{m_x} p(x_i)\log p(x_i)$.

4.2.3. Свойства энтропии

Энтропия $H(X) = -\sum_{i=1}^{m_x} p(x_i)\log p_i \geq 0$, поскольку p_i удовлетворяет неравенству $0 \leq p_i \leq 1$. Энтропия H(X)=0, когда система находится в одном из состояний с вероятностью, равной единице, и во всех остальных — с вероятностью, равной нулю. При этом имеется в виду, что

$$\lim_{p_i \to 0} p_i \log p_i = 0.$$

При равномерном распределении ($p_i = \frac{1}{m_x}$) энтропия $H(X) = \log m_x$.

Докажем, что это максимальное значение энтропии. Используя равенство $\sum_{i=1}^{m_X} p_i = 1$, можно выполнить следующие тождественные преобразования:

$$H(X)-\log m_X = \sum_{i=1}^{m_X} p_i \log \frac{1}{p_i} - \sum_{i=1}^{m_X} p_i \log m_X = \sum_{i=1}^{m_X} p_i \left(\log \frac{1}{p_i} - \log m_X\right) = \sum_{i=1}^{m_X} p_i \log \frac{1}{p_i m_X}.$$

Для оценки выражения $\log \frac{1}{p_i m_X}$ воспользуемся неравенством

$$\ln z \le z - 1, \text{ положив } z = \frac{1}{p_i m_X}.$$

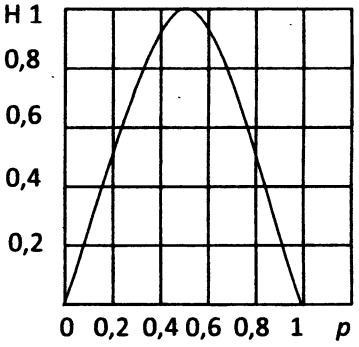


Рис. 4.1. Энтропия системы с двумя состояниями:
p – вероятность одного из состояний

Заменяя $\log \frac{1}{p_i m_x}$ на $\left(\frac{1}{p_i m_X} - 1\right)\log e$, получим

$$H(X) - \log m_X \le \sum_{i=1}^{m_X} p_i \left(\frac{1}{p_i m_X} - 1\right)\log e = \left(\sum_{i=1}^{m_X} \frac{1}{m_X} - \sum_{i=1}^{m_X} p_i\right) * \log e = (1-1)\log e = 0,$$

где log e —модуль перехода. Отсюда

$$H(X) \le \log m_X.$$

Пусть множество состоит из двух элементов, которые обозначим через единицу и ноль, причем единица появляется с вероятностью, равной p, а ноль — с вероятностью, равной $q=1-p$. Тогда

$$H(X) = -p\log p - (1-p)\log(1-p).$$

Указанная зависимость изображена на рис. 4.1. Максимум достигается при $p=q=0{,}5$.

4.2.4. Ценность информации

Все определения ценности информации связаны с понятием цели. Ценной считается та информация, которая способствует достижению поставленной цели.

Один из способов измерения ценности информации, сформулированный в рамках статистической теории информации, был предложен А.А.Харкевичем [3]. Ценность информации может быть выражена через приращение вероятности достижения цели. Если значение априорной вероятности достижения цели обозначить через p_1, а апостериорной — через p_2, то ценность полученной информации можно определить как $\log\frac{p_2}{p_1}$.

В системах передачи информации цель сводится к правильной передаче сообщений независимо от их конкретного содержания и формулируется относительно каждого символа множества X. Пусть целью является принятие решения в пользу x_i. Тогда относительно этой цели ценность сведений, содержащихся в принятом y_j, равна $\log\frac{p(x_i|y_j)}{p(x_i)}$, где $P(x_i)$ — априорная вероятность передачи x_i; $p(x_i|y_j)$ — вероятность того, что было передано x_i после принятия y_j. При такой формулировке цели ценность информации совпадает с обычным количеством информации, которое определено ранее.

Таким образом, количество информации, которое y_j несет об x_i, равно

$$I(y_j, x_i) = \log\frac{p(x_i|y_j)}{p(x_i)}.$$

Умножая числитель и знаменатель под логарифмом на $p(y_j)$ и учитывая равенства

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

получим

$$I(y_j, x_i) = \frac{p(x_i|y_j)}{p(x_i)} = \log\frac{p(x_i, y_j)}{p(x_i)p(y_j)} = \log\frac{p(y_j|x_i)}{p(y_j)} = I(x_i, y_j). \quad (4.3)$$

Отсюда следует, что y_j несет об такое же количество информации, какое x_i несет об y_j (свойство симметрии). Поэтому $I(x_i, y_j)$ называется взаимным количеством информации между i-м символом множества X и j-м символом множества Y. Взаимное количество информации $I(x_i, y_j)$, может быть положительным $(p(x_i|y_j) > p(x_i))$, отрицательным $(p(x_i|y_j) < p(x_i))$ и равным нулю $(p(x_i|y_j) = p(x_i))$. Отрицательная информация называется дезинформацией.

4.2.5. Собственное количество информации и энтропия

Пусть в канале отсутствуют помехи. Тогда между элементами множеств X и Y имеет место взаимно однозначное соответствие и $p(x_i|y_j)=1$ при $j=i$. В этих условиях количество информации, которое y_j $(j=i)$ доставляет об x_i, согласно (4.3) равно $I(x_i) = -\log p(x_i) = -\log p(y_j)$.

Эта величина называется собственным количеством информации, которое несет символ x_i, причем всегда $I(x_i)>0$. Усредняя $I(x_i)$ по всему множеству x_i, получим количество информации, которое в среднем несут сообщения множества X. Среднее значение совпадает с энтропией

$$H(X) = \sum_{i=1}^{X_m} p_i I(x_i) = -\sum_{i=1}^{m_X} p_i \log p_i,$$

где $p_i=p(x_i)$.

4.2.6. Взаимная информация

Источник информации и приемник можно рассматривать как подсистемы одной сложной системы. Взаимную информацию между состояниями подсистем, используя (4.3), можно записать в виде

$$\begin{aligned} I(y_j, x_i) &= -\log\ p(x_i) - [-\log\ p(x_i \mid y_j)] = \\ &= -\log\ p(x_i) - \log\ p(y_j) - [-\log\ p(x_i, y_j)] = \\ &= -\log\ p(y_j) - [-\log\ p(y_j|x_i)] = I(x_i, y_j). \end{aligned} \qquad (4.4)$$

Поскольку сложная система случайным образом приходит в то или иное состояние, определяемое парой чисел (x_i,y_j), то $I(xi,y_j)$ будет случайной величиной, которую можно усреднить по всему множеству состояний. В результате почленного усреднения (4.4) получим выражение для средней (полной) взаимной информации:

$$\begin{aligned} I(y,x) &= H(x) - H(x|y) = H(x) + H(y) - H(x,y) \\ &= H(y) - H(y|x) = I(x,y) \end{aligned} \qquad (4.5)$$

где $I(y,x) = \sum_{i=1}^{m_X}\sum_{j=1}^{m_Y} p(x_i, y_j) I(x_i, y_j)$;

$$H(X|Y) = -\sum_{i=1}^{m_X}\sum_{j=1}^{m_Y} p(x_i, y_j) \log p(y_j|x_i).$$

С точки зрения информационного описания системы связи безразлично, какую из подсистем рассматривать в качестве передатчика, а какую в качестве приемника.

Поэтому энтропии $H(X)$ и $H(Y)$ можно интерпретировать как информацию, которая поступает в канал связи, а условные энтропии $H(X|Y)$, $H(Y|X)$ как информацию, которая рассеивается в канале. В [1] доказано, что $I(X, Y)\geq 0$.

При выполнении указанного неравенства из (4.5) следует, что

$$H(X \mid Y) \leq H(X),$$

$$H(Y \mid X) \leq H(Y),$$

$$H(X,Y) \leq H(X) + H(Y).$$

Условную энтропию можно представить в виде

$$H(X \mid Y) = -\sum_{j=1}^{m_Y} p(y_j) \sum_{i=1}^{m_X} p(x_i \mid y_j) \log p(x_i \mid y_j) = \sum_{j=1}^{m_Y} p(y_j) H(X \mid y_j),$$

где величина $H(X \mid y_j) = -\sum_{i=1}^{m_X} p(x_i \mid y_j) \log p(x_i \mid y_j)$ называется частной условной энтропией. Она характеризует неопределенность состояния системы A в случае, когда известно состояние y_j наблюдаемой системы B. Зафиксировав состояние y_j системы B, мы тем самым изменяем комплекс условий, при которых может реализоваться событие x_i. Это обнаруживается как изменение вероятности реализации события x_i ($i=\overline{1,n}$) (имеет место статистическая зависимость). Если до изменения условий указанная вероятность была равна безусловной (полной) вероятности $p(X_i)$, то после изменения условий она стала равной условной вероятности $p(x_i|y_j)$. При отсутствии статистической зависимости $H(X|y_j)=H(X)$, поскольку $P(x_i|y_j)=p(x_i)$.

Таблица 4.1

y \ x	x_1	x_2	$p(y_i)$
y_1	0.5	0	0.5
y_2	0.25	0.25	0.5
$p(x_i)$	0.75	0.25	

Таблица 4.2

y \ x	x_1	x_2
y_1	1	0
y_2	0.5	0.5

При наличии статистической зависимости энтропия $H(X|y_j)$ может оказаться как меньше, так и больше $H(X)$. Напомним, что для энтропии $H(X|Y)$ всегда справедливо неравенство $H(X|Y)\leq H(X)$.

В качестве примера вычислим энтропии $H(X)$, $H(X|Y)$, $H(X|y_i)$ и взаимную информацию $I(X,Y)$, когда системы A и B описываются двумерным распределением $p(x_i, y_j)$, заданным в виде табл. 1.

Вычисленные значения условной вероятности $p(x_i \mid y_j) = \dfrac{p(x_i, y_j)}{p(y_j)}$ записаны в табл. 4.2.

Используя записанные в таблицах значения вероятностей, получим

$$H(X \mid y_1) = -p(x_1 \mid y_1)\log p(x_1 \mid y_1) - p(x_2 \mid y_1)\log p(x_2 \mid y_1) = -1\log 1 - 0\log 0 = 0,$$

$$H(X \mid y_2) = -p(x_1 \mid y_2)\log p(x_1 \mid y_2) - p(x_2 \mid y_2)\log p(x_2 \mid y_2) = -0.5\log 0.5 + 0.5\log 0.5 = 1,$$

$$H(X \mid Y) = -p(y_1)H(X \mid y_1) + p(y_2)H(X \mid y_2) = 0.5 \cdot 0 + 0.5 \cdot 1 = 0.5,$$

$$H(X) = -(0.25\log 0.25 + 0.75\log 0.75) \approx 0.811,$$

$$I(X,Y) = H(X) - H(X \mid Y) \approx 0.811 - 0.5 = 0.311 > 0.$$

Отсюда

$$H(X \mid y_1) < H(X) < H(X \mid y_2),$$

$$0 < 0.811 < 1,$$

$$H(X) > H(X \mid Y),$$

$$0.811 > 0.5.$$

4.3. Дискретные источники сообщений и их описание

4.3.1. Эргодические источники

Источник будем называть эргодическим, если его вероятностные параметры можно оценить по одной достаточно длинной реализации, которую он вырабатывает. При неограниченном возрастании длины реализации (n) оценка параметра (результат измерения) совпадает с его истинным значением с вероятностью, равной единице. Например, при бросании игральной кости можно оценить вероятность выпадания какой-либо цифры через относительную частоту ее появления в достаточно длинной серии испытаний. Указанная серия испытаний представляет собой ту самую реализацию, по которой осуществляется оценка вероятности (параметра). Реализации, по которым можно оценить закон распределения, являются типичными. Поэтому эргодическим источником можно назвать источник, который вырабатывает типичные последовательности. Типичная последовательность несет сведения о структуре источника, то есть является типичной для данного источника. Если два источника различаются своей

структурой (значением оцениваемого параметра), то, наблюдая реализацию, можно определить, какому из них она принадлежит. Источник, эргодический по одному параметру, может оказаться не эргодическим по другому параметру.

4.3.2. Производительность дискретного источника сообщений

Кодовое слово, которое вырабатывает источник, будем записывать в виде $x_{i,1}$, $x_{i,2}$,…,$x_{i,k}$,…, $x_{i,n}$, где X_{ik}-i_k- буква (символ) алфавита с k порядковым номером в слове. Например, пусть k=5, а i=3. Это значит, что пятой буквой в слове является третья буква алфавита. Обозначим через X_k множество букв (алфавит), из которых выбирается k-я буква слова. В нашем случае все множества $X_k(k=\overline{1.n})$ состоят из одних и тех же m_x букв. Когда не требуется указывать место буквы в слове, вместо X_k , будем писать X.

Количество информации, которое в среднем несет отдельное слово, равно энтропии $H(X_1...X_n) = -\sum p(x_{i1},...,x_{in})\log p(x_{i1},...,x_{in})$, где суммирование ведется по всему множеству слов. Определим производительность источника $H_и$ как предел отношения количества информации, которое в среднем несет отдельное слово, к числу букв в слове n при неограниченном возрастании n:

$$H_и = \lim_{n\to\infty}\frac{H(X_1,...,X_n)}{n}. \qquad (4.6)$$

Если буквы в слове статистически независимы (вероятность выбора очередной буквы не зависит от состава предшествующих ей букв), то

$H(X_1,...,X_n) = H(X_1)+...H(X_k)...+H(X_n)$, где $H(X_k) = -\sum_{i_k=1}^{m_X} p(x_{i.k})\log p(x_{i.k})$.

Источник со статистически независимыми буквами сообщений будет стационарным, если вероятность выбора i-й буквы алфавита не зависит от того, какое место в слове она занимает $(p(x_{i.k}) = p(x_i))$.

В этом случае $H(X_1,...,X_n) = nH(x)$ и производительность источника (бит/символ) $H_и = H(X) = -\sum_{i=1}^{m_X} p(x_i)\log p(x_i)$.

Часто производительность источника измеряется количеством информации $\nu H(x)$, которое он вырабатывает за одну секунду (ν — количество букв за одну секунду). Максимальная производительность источника достигается, когда все буквы алфавита появляются с равными вероятностями. В этом случае $H_п = \log m_x$.

4.3.3. Марковские источники сообщений

Рассмотренная модель дискретного источника сообщений имеет сравнительно узкую область применения, поскольку реальные источники вырабатывают слова при наличии статистической зависимости между буквами. В реальных источниках вероятность выбора какой-либо очередной буквы зависит от всех предшествующих букв. Многие реальные источники достаточно хорошо описываются марковскими моделями источника сообщений. Согласно указанной модели условная вероятность выбора источником очередной буквы $x_{i,k}$ зависит только от ν предшествующих. Математической моделью сообщений, вырабатываемых таким источником, являются цепи Маркова ν-го порядка. В рамках указанной модели условная вероятность выбора i_k-й буквы

$$p(x_{i,k}\,|\,x_{i,k-1},...,x_{i,k-\nu},x_{i,k-\nu-1},...,x_{i,1}) = p(x_{i,k}\,|\,x_{i,k-1},...,x_{i,k-\nu}).$$

Если последнее равенство не зависит от времени, то есть справедливо при любом значении k, источник называется однородным. Однородный марковский источник называется стационарным, если безусловная вероятность выбора очередной буквы не зависит от k ($p(x_{i,k})=p(x_i)$). В дальнейшем будем иметь дело только со стационарными источниками. Вычислим производительность источника для простой цепи Маркова (ν=1).

В этом случае вероятность $p(x_{i,1}...,x_{i,n}) = p(x_{i,1})p(x_{i,2}\,|\,x_{i,1})...p(x_{i,n}\,|x_{i,n-1})$.

Прологарифмировав последнее равенство, получим

$$-\log p(x_{i1},...,x_{in}) = -\log p(x_{i1}) - \log p(x_{i2}\,|\,x_{ix}) - ... - \log p(x_{in}\,|\,x_{in-1}).$$

Это равенство показывает, что индивидуальное количество информации, которое несет слово, равно количеству информации, которое несет первая буква, плюс количество информации, которое несет вторая буква при условии, что первая буква уже принята, и т. д.

Усредняя равенство по всем словам, получим количество информации, которое в среднем несет каждое слово:

$$H(X_1,...,X_n) = H(X_1) + H(X_2\,|\,X_1) + ... + H(X_n\,|\,X_{n-1}).$$

Поскольку источник стационарный, то энтропия не зависит от k и равна $H(X_1,...,X_n) = H(X) + (n-1)H(X_k\,|\,X_{k-1}) \le nH(X)$.

Подставляя полученный результат в (4.6) и учитывая, что всегда $H(X) \le \log m_x$, имеем $H_u = \lim\limits_{n\to\infty}\left(\frac{H(X)}{n} + \frac{n-1}{n}H(X_k\,|\,X_{k-1})\right) = H(X_k\,|\,X_{k-1})$.

В случае марковской цепи ν-го порядка H_u вычисляется аналогично и равна $H_u = H(X_{\nu+1}\,|\,X_\nu,...,X_1)$.

Таким образом, производительность марковского источника равна неопределенности выбора очередной буквы при условии, что известны ν предшествующих.

Для производительности марковского источника всегда справедливо неравенство $H_u \le H(X) \le \log m_x$.

Максимального значения, равного $\log m_x$, производительность источника достигает, когда отсутствует статистическая зависимость между буквами в слове и когда все буквы алфавита вырабатываются с равными вероятностями. Очевидно, максимальная производительность источника полностью определяется размером алфавита m_x.

Для того чтобы характеризовать, насколько полно использует источник возможности алфавита, вводится параметр $k = \dfrac{H_{\max}(X) - H_u}{H_{\max}(X)}$, называемый избыточностью.

Для передачи заданного количества информации, равного I, требуется $n=I/\mathrm{H}_u$ букв, если производительность источника равна H_u. В случае, когда производительность источника достигает своего максимального значения, равного $H_{\max}(X) = \log m_x$, для передачи того же количества информации I требуется минимальное количество букв, равное $n_0 = \dfrac{I}{H_{\max}(X)}$.

Отсюда $I=nH_u=n_0 H_{max}$ или $\dfrac{H_u}{H_{\max}(X)} = \dfrac{n_0}{n}$. Учитывая последнее равенство, выражение для избыточности можно записать в виде $k = 1 - \dfrac{H_u}{H_{\max}(X)} = \dfrac{n - n_0}{n}$

Таким образом, избыточность показывает, какая часть букв в слове не загружена информацией.

Пример 1. Определить избыточность источника, если он вырабатывает статистически независимую последовательность из единиц и нулей соответственно с вероятностями, равными p=0,3 и q=0,7.

Решение. Поскольку символы в последовательности статистически независимы, то производительность источника

$H_u = H(X) = -p\log p - q\log q \approx 0.88$ бит.

Максимально возможная производительность источника $H_{\max}(X) = \log m_x = 1$, поскольку m_x=2. При этом символы 1 и 0 должны вырабатываться с равными вероятностями (p=q=0.5). Отсюда $r = 1 - \dfrac{0.88}{1} = 0.12.$

Пример 2. Определить избыточность стационарного марковского источника, алфавит которого состоит из двух символов: 0 и 1. Вырабатываемая источником последовательность представляет собой простую цепь Маркова. Заданы следующие значения условных вероятностей

$p(x_{ik+1}|x_{ik})$ $(i_{k-1}=\overline{1,2},\ i_k=\overline{1,2})$:

$p(0|0)=0.3;\quad p(1|0)=0.7;\quad p(0|1)=0.1;\quad p(1|1)=0.9.$

Решение. Безусловную вероятность того, что $(k+1)$-м символом последовательности будет нуль, по формуле полной вероятности можно представить в виде $p_{k+1}(0)=p_k(0)p(0|0)+[1-p_k(0)]p(0|1)$.

В правую часть неравенства входит вероятность $p_k(0)$ того, что k-м символом последовательности будет нуль. В силу стационарности источника $p_{k+1}(0)=p_k(0)=p(0)$. Подставив в равенство значения $p(0|0)$ и $p(0|1)$, получим

$$p(0)=0.125,\quad p(1)=1-p(0)=0.875.$$

Производительность источника $H_и=p(0)H(X_{k+1}|0)+p(1)H(X_{k+1}|1)=$ $=-0.125\cdot(0.3\log 0.3+0.7\log 0.7)-0.875*(0.1\log 0.1+0.9\log 0.9)\approx 0.51$, а избыточность источника $k=1-\dfrac{H_и}{H_{\max}(X)}=1-\dfrac{0.51}{1}=0.49$,

где $H_{\max}(X)=1$.

Когда отношение ν/n стремится к нулю, при неограниченном возрастании n марковский источник вырабатывает типичные последовательности, количество которых $Q\approx 2^{(n-1)H_и}$ или более приближенно $Q\approx 2^{nH_и}$.

4.4. Кодирование сообщений при передаче по каналу без помех

4.4.1. Возможность оптимального (эффективного) кодирования

Под кодированием будем понимать отображение состояний некоторой системы (источника сообщений) с помощью состояний сложного сигнала, который представляет собой последовательность из n элементарных сигналов. Множество Y состояний элементарного сигнала образует алфавит кода размера m_y. При $m_y=2$ элементарный сигнал имеет два состояния, которые обозначим через 1 и 0. Состояние сложного сигнала описывается последовательностью из нулей и единиц, которая называется кодовым словом. Если кодовые слова имеют разную длину, то код называется неравномерным, а если одинаковую, то код называется равномерным.

Пусть источник сообщений вырабатывает последовательность из k букв, причем x_i буква в этой последовательности появляется n раз. Каждой букве x_i $(i=\overline{1,m_x})$ поставим в соответствие кодовое слово с длиной, равной l_i (посимвольное кодирование). Тогда длина соответствующей последовательности из кодовых слов будет равна $L=\sum_{i=1}^{m_x} n_i l_i$.

Это равенство можно представить в виде $L=K\sum_{i=1}^{m_x}\left(\dfrac{n_i}{K}\right)l_i$.

При неограниченном увеличении числа букв K в последовательности относительная частота появления x_i буквы с вероятностью, равной едини-

це, совпадает со значением вероятности p_i появления этой буквы. Поэтому с вероятностью, равной единице, выполняется равенство $L \approx K\sum_{i=1}^{m_x} l_i p_i = K\bar{l}$.

$\bar{l} = \sum_{i=1}^{m_x} l_i p_i$ - по определению средняя длина кодового слова. Длина последовательности кодовых слов L является случайной величиной, но значительные отклонения ее от среднего значения $\bar{L} = K\bar{l}$ маловероятны при неограниченном возрастании K.

Таким образом, источник сообщений с вероятностью, равной единице, вырабатывает типичные последовательности, длина которых мало отличается от их среднего значения $K\bar{l}$.

Поскольку время передачи сообщений определяется длиной L, то имеется возможность его сокращения за счет уменьшения средней длины кодового слова $\bar{l}$ ($L \approx K\bar{l}$). Задача оптимального кодирования заключается в определении однозначно декодируемых кодовых слов с такими длинами, при которых их средняя длина минимальна.

4.4.2. Префиксные коды

При неравномерном коде в длинной последовательности кодовых слов не всегда удается определить начало и конец переданной буквы.

Однако однозначное декодирование всегда имеет место в случае применения кодов, обладающих свойством префикса (приставки). Код обладает свойством префикса, если ни одно кодовое слово не является началом (приставкой) какого-либо другого кодового слова. Все множество кодовых слов, максимальная длина которых не превосходит число, равное l, геометрически удобно изобразить в виде узлов дерева (рис. 4.2).

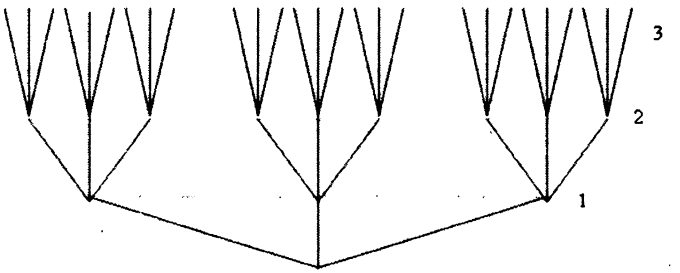


Рис. 4.2. Полное троичное (m_x=3) кодовое дерево третьего порядка (l=3): 1, 2, 3 узлы первого, второго, третьего порядков

Дерево представляет собой множество точек (узлов), соединенных отрезками, которые называются ребрами дерева. Из каждого узла выходят

m_γ ребер, каждое из которых изображает соответствующий символ алфавита кода. Слова, состоящие всего из одной буквы, изображаются узлами первого порядка. Их число равно m_γ. Слова, состоящие из двух букв, изображаются узлами второго порядка и т. д. Причем узлов (слов) K-го порядка в m_γ раз больше, чем узлов (K-1)-го порядка, поскольку каждый узел предыдущего порядка порождает m_γ узлов следующего порядка. Конкретный вид кодового слова определяется по изображающему его узлу как последовательность ребер (букв), которые соединяют основание дерева с указанным узлом. В случае префиксных кодов на пути, соединяющем основание дерева с изображающим узлом, не может быть промежуточных изображающих узлов.

Поскольку с помощью дерева (рис. 4.2) можно изобразить все кодовые слова с длиной, меньшей или равной l, то оно называется полным деревом порядка l с алфавитом объема m_γ.

4.4.3. Неравенство Крафта

Теорема. *Если целые числа* $l_1,...,l_i,...,l_N$ *удовлетворяют неравенству*

$$\sum_{i=1}^{N} m_\gamma^{-l_i} \le 1, \qquad (4.7)$$

то существует код, обладающий свойством префикса с алфавитом объема m_γ, *длины кодовых слов в котором равны, этим числам. Обратно, длины кодовых слов любого кода, обладающего свойством, префикса, удовлетворяют указанному неравенству.*

Теорема не утверждает, что любой код с длинами кодовых слов, удовлетворяющими (4.7), является префиксным. Так, множество двоичных кодовых слов 0,00,11 удовлетворяет (4.7), но не обладает свойством префикса. Теорема утверждает только существование префиксного кода, но не указывает его конкретный вид. Кодовые слова 0,10,11 удовлетворяют неравенству (4.7) и обладают свойством префикса.

Доказательство. Пусть числа $l_1,l_2,...,l_i,...,l_N$ удовлетворяют неравенству (4.7).

Покажем, как можно построить префиксный код с этими длинами кодовых слов, и тем самым докажем существование префиксного кода. Не нарушая общности доказательства, все числа можно перенумеровать в порядке возрастания их значений. Тогда будем иметь $l_1 \le l_2 \le ... \le l_N$. Построение будем вести на полном дереве порядка l_N.

Построение сводится к последовательному выбору узлов порядков $l_1,l_2,...,l_N$, но так, чтобы очередной выбираемый узел не был порожден каким-либо ранее выбранным узлом. Первый узел (кодовое слово) выбирается произвольно из числа узлов порядка l_1. Этот узел порождает

$m_y^{-l_1}$-ю часть узлов более высокого порядка, которые уже не могут быть использованы. После выбора следующего узла порядка l_2 уже $m_y^{-l_1} + m_y^{-l_2}$ часть узлов не может быть использована и т.д. После выбора (M-1)-го узла может быть использована $1 - \sum_{i=1}^{N-1} m_y^{-l_i}$ часть узлов порядка l_N.

Поскольку для чисел $l_1,...,l_N$ справедливо неравенство Крафта, которое можно записать в виде $\sum_{i=1}^{N-1} m_y^{-l_i} + m_y^{-l_N} \le 1$, то величина $\sum_{i=1}^{N-1} m_y^{-l_i}$ строго меньше единицы. Следовательно, существует часть узлов порядка l_N, из которых можно выбрать последний N-й узел.

Отметим еще одно свойство кодовых слов. Если код однозначно декодируется, то его кодовые слова удовлетворяют неравенству Крафта. Доказательство можно найти, например, в работе [4].

Таким образом, префиксные коды составляют часть однозначно декодируемых кодов, а последние составляют часть кодов, удовлетворяющих неравенству Крафта.

4.4.4. Предельные возможности оптимального кодирования

Определим границы для l_i и $\bar{l}$б, пользуясь эвристическими соображениями, основанными на количестве информации. Очевидно, код будет самым экономным, если каждый символ кодового слова будет переносить максимально возможное количество информации.

Поскольку собственное количество информации, содержащееся в сообщении $x_i \in X$, равно $-\log p_i$, информационной емкости соответствующего кодового слова будет достаточно, чтобы перенести указанное количество информации только в том случае, если его минимальная длина l_i будет находиться в пределах $\frac{-\log p_i}{\log m_y} \le l_i \le \frac{-\log p_i}{\log m_y} + 1$, где $\log m_y$ - максимальное количество информации, которое может перенести отдельный символ кодового слова.

Усредняя неравенство по всему множеству X сообщений, получим неравенство, определяющее границы для минимальной средней длины кодового слова: $\frac{H(X)}{\log m_y} \le \bar{l} \le \frac{H(X)}{\log m_y} + 1$.

В данном случае довольно большой интервал изменения возможных значений $\bar{l}$. Однако при кодировании блоков (слов из n букв x_i) средняя длина $\bar{l}$ приближается к значению $\frac{H(X)}{\log m_y}$ при неограниченном увеличе-

нии длины блока n, что следует из неравенства $\frac{H(X)}{\log m_y} \le \frac{l_\sigma}{n} < \frac{H(X)}{\log m_y} + \frac{1}{n}$,

где l_σ - среднее количество символов кодового слова, приходящееся на один блок, а $\frac{l_\sigma}{n}$ на одну букву в блоке.

Полученные неравенства справедливы для всех однозначно декодируемых кодов.

4.4.5. Алгоритмы эффективного кодирования

Алгоритм Шеннона-Фано

Пусть имеется множество сообщений $x_i \in X$.

1. Все сообщения располагаются в порядке убывания их вероятностей.

2. Упорядоченное множество сообщений делится на две части: верхнюю часть и нижнюю часть, причем так, чтобы разность между суммой вероятностей в верхней и нижней части была минимальной.

3. После этого сообщениям в верхней части ставится в соответствие 1, а в нижней – 0.

4. Далее аналогичные действия производятся с каждой из частей. Вновь полученные подмножества сообщений снова аналогичным образом делятся на две части и т.д., до получения по одному сообщению в каждом из подмножеств.

5. В результате каждому сообщению будет соответствовать своя последовательность из нулей и единиц, т.е. кодовое слово.

Пример:

Таблица 4.3

p_i	X			
0.5	x_1	1		
0.25	x_2	0	1	
0.125	x_3	0	0	1
0.125	x_4	0	0	0

Алгоритм Хаффмана

1. Все сообщения располагаются в порядке убывания их вероятностей.

2. Два самых нижних сообщения объединяются в одно событие, вероятность которого равна сумме вероятностей объединяе-

мых событий, причем верхнему событию ставится в соответствие 1, а нижнему 0.

3. Получился новый массив сообщений с количеством состояний на единицу меньшим по сравнению с предыдущим массивом, если два последних события считать одним более крупным событием. Далее преобразуем этот массив в соответствии с пунктами 1 и 2. Полученный массив вновь подвергаем указанным преобразованиям и так до тех пор, пока не будет исчерпан весь массив.

4. Геометрически результат можно представить в виде дерева, где кодовое слово, соответствующее x_i, выражается последовательностью ветвей с соответствующим сообщением x_i.

Пример:

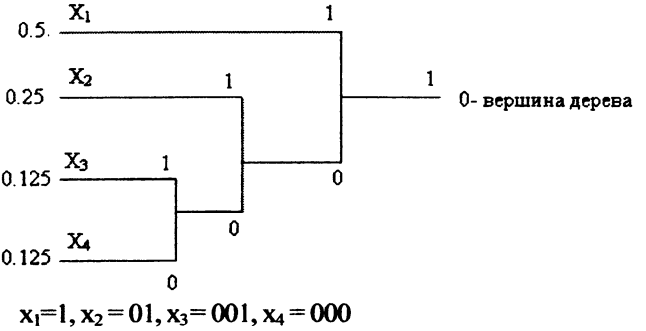


Рис. 4.3. Геометрическая интерпретация алгоритма Хаффмана

4.5. Пропускная способность дискретного канала связи

4.5.1. Определение пропускной способности канала

Пропускная способность канала – предельная скорость передачи информации, при которой может быть получена сколь угодно малая вероятность ошибки.

Для общего описания канала связи и построения теории информации используется одна и та же модель. Канал называется дискретным (непрерывным), если множества X и Y дискретны (непрерывны), и полунепрерывным, если одно из множеств дискретно, а другое непрерывно. Далее рассматриваются только дискретные каналы.

Канал полностью описывается условными вероятностями $p(y_{jk}|x_{i1},...,x_{ik})$ того, что k-м принятым символом будет j-й символ множества $Y(j_k=\overline{1,m_Y})$.

Указанную вероятность можно рассматривать как функцию y_{jk} и $x_{i1},...,x_{ik}$, вид которой отражает состояние канала, в частности, характер взаимодействия помехи и сигнала.

Если $p(y_{j,k}|x_{i,1},...,x_{i,k})=p(y_{j,k}|x_{i,k})$, $(j_k=\overline{1,m_Y},\ i_k=\overline{1,m_X})$ то соответствующий канал называется каналом без памяти. Если вероятность $p(y_{j,k}|x_{i,k})$ не зависит от k (от времени), то соответствующий канал называется стационарным. Ограничимся рассмотрением только стационарных каналов без памяти.

Определим скорость передачи информации как предел: $R=\lim\limits_{n\to\infty}\dfrac{I(\vec{X},\vec{Y})}{n}$, где $I(\vec{X},\vec{Y})$ средняя взаимная информация между переданным $x\in X$, и принятым $y\in Y$. В случае отсутствия помех $H(X|Y)=0$, следовательно, $R=H(X)$. Этот предел в случае канала без памяти равен взаимной информации: $R=I(X,Y)=H(X)-H(X|Y)=H(Y)-H(Y|X)$.

Скорость передачи информации R полностью определяется вероятностями $p(x_i)$ и $p(Y_j|X_i)$ $(i=\overline{1,m_X})$. Поэтому изменять величину R мы можем только за счет изменения вида распределения $p(x_i)$, поскольку $p(Y_j|X_i)$ - характеристика неуправляемого канала. Определим пропускную способность канала C как максимальную по $p(x_i)$ скорость передачи информации: $C=\max\limits_{p(x_i)}R=\max\limits_{p(x_i)}I(X,Y)$.

В случае отсутствия помех $C=\max\limits_{p(x_j)}H(X)=\log m_x$.

4.5.2. Вычисление пропускной способности симметричных каналов

Существует класс каналов, для которых пропускная способность C легко вычисляется. Канал полностью описывается так называемой стохастической матрицей

$$\begin{pmatrix} p(y_1|x_1) & \dots & p(y_{m_Y}|x_1) \\ \dots & \dots & \dots \\ p(y_1|x_{m_x}) & \dots & p(y_{m_Y}|x_{m_x}) \end{pmatrix},$$

в которой сумма всех элементов, образующих строку, равна единице.

Канал называется симметричным по входу, если строки матрицы различаются только порядком расстановки некоторого множества чисел $p_1,...,p_{mY}$.

Для симметричных по входу каналов частная условная энтропия

$$H(Y \mid x_i) = -\sum_{j=1}^{m_Y} p(y_j \mid x_i) \log p(y_j \mid x_i) = -\sum_{j=1}^{m_Y} p_j \log p_j .$$

Она не зависит от номера передаваемой буквы и может быть вычислена по любой строке матрицы. Поэтому условная энтропия

$$H(Y \mid X) = \sum_{i=1}^{m_X} p(x_i) H(Y \mid x_i) = -\sum_{j=1}^{m_Y} p_j \log p_j .$$

Канал называется симметричным по выходу, если столбцы матрицы различаются только порядком расстановки некоторого множества чисел $q_1, \ldots, q_{m_Y}$.

Если распределение источника равномерное ($p(x_i) = \frac{1}{m_X}$), то распределение $p(y_i)$ на выходе симметричного по выходу канала также будет равномерным. При этом энтропии $H(X)$ и $H(Y)$ достигают своего максимального значения. В этом легко убедиться, если доказать, что вероятность $p(y_i)$ не зависит от y_j. Представим вероятность $p(y_i)$ в виде $p(y_j) = \sum_{i=1}^{m_X} p(x_i) p(y_j \mid x_i)$.

Поскольку $p(x_i) = \frac{1}{m_X}$ и $p(y_j \mid x_i) = q_i$, то $p(y_j) = \frac{1}{m_X} \sum_{i=1}^{m_X} q_i$.

Сумма $\sum_{i=1}^{m_X} q_i$ не зависит от номера столбца j и в общем случае не равна единице. Поэтому вероятность $p(y_i)$ также не зависит от j и равна $p(y_j) = \frac{1}{m_Y}$. При этом $H(X) = \log m_X$, $H(Y) = \log m_Y$.

Канал называется симметричным, если он симметричен по входу и выходу. Для симметричного канала $H(Y|X)$ не зависит от распределения источника сообщений, поэтому пропускная способность

$$C = \max_{p(x_i)} [H(Y) - H(Y \mid X)] = \max_{p(x_i)} H(Y) - H(Y \mid X) = \log m_Y + \sum_{j=1}^{m_Y} p_j \log p_j .$$

В качестве примера вычислим пропускную способность симметричного канала, который описывается матрицей

$$\begin{pmatrix} 1-p_e & \frac{p_e}{m-1} & \ldots & \frac{p_e}{m-1} \\ \ldots & \ldots & \ldots & \ldots \\ \frac{p_e}{m-1} & \ldots & \ldots & 1-p_e \end{pmatrix},$$

где $m = m_X = m_Y$. В этом случае

$$p(y_j \mid x_i) = \begin{cases} 1 - p_e, & i = j, \\ \dfrac{p_e}{m-1}, & i \neq j. \end{cases}$$

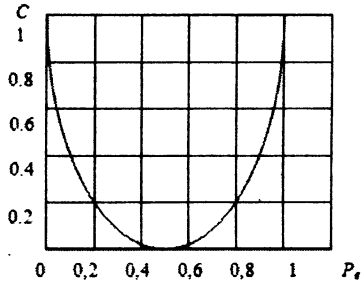


Рис. 4.4. Зависимость пропускной способности ДСК от вероятности ошибки p_e

Вероятность $1\text{-}p_e$ равна вероятности правильного приема символа. Вероятность ошибки p_e равна вероятности приема $y_j, j \neq i$ при условии, что было передано x_i. Тогда $C = \log m + (1 - p_e)\log(1 - p_e) + p_e \log \dfrac{p_e}{m-1}$.

Широкое распространение получил двоичный симметричный канал (ДСК) *(m = 2)*, для которого пропускная способность (рис. 4.4) $C = 1 + (1 - p_e)\log(1 - p_e) + p_e \log p_e$.

Максимальная скорость передачи информации, равная единице, получается при p_e=0 и при p_e=1. В этом случае множества X и $У$ находятся во взаимно однозначном соответствии, и по принятому y_j (j=1, 2) всегда можно определить с вероятностью, равной единице, переданную букву. К сожалению, это возможно только тогда, когда априори (до приема) известно значение вероятности p_e (нуль или единица).

4.5.3. Вычисление пропускной способности канала со стиранием

Приемное устройство на основе анализа реализации принятого сигнала выносит решение о том, какая буква (сообщение) была передана. В некоторых случаях создается такая помеховая обстановка, что вынести решение в пользу той или иной буквы не представляется возможным. В этом случае целесообразнее вообще не принимать решение о том, какая буква была передана. Указанный отказ от решения θ является тоже решением и входит в множество решений Y наравне со всеми остальными решениями.

Вынужденный отказ от решения часто возникает в проводных каналах связи при нарушении контакта (обрыве). В этом случае сигнал просто не проходит на выход канала и мы вынуждены отказаться от принятия решения даже в случае отсутствия внешних помех.

Вычислим пропускную способность канала, диаграмма переходных вероятностей которого изображена на рис. 4.5 ($m_x = 2, m_y = 3$).

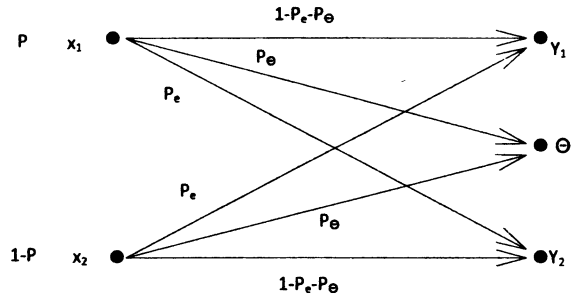


Рис. 4.5. Диаграмма переходных вероятностей для канала со стиранием

Указанный канал описывается матрицей:

$$\begin{pmatrix} 1-p_t-p_\theta, p_\theta, & p_t \\ p_t & p_\theta, 1-p_t-p_\theta \end{pmatrix}.$$

Поскольку строки матрицы различаются только перестановкой чисел p_t, p_θ, $1-p_t-p_\theta$, то соответствующий канал симметричен по входу.

Для симметричного по входу канала условная энтропия

$$H(Y/X) = -[p_t \log p_t - p_\theta \log p_\theta + (1-p_t-p_\theta)\log(1-p_t-p_\theta)] \quad (4.8)$$

не зависит от вида распределения источника сообщений и полностью определяется параметрами канала p_t и p_θ.

Поэтому пропускная способность канала можно записать в виде

$$C = \max_p H(Y) - H(Y/X). \quad (4.9)$$

Вычислим $\max H(Y)$, энтропия $H(Y)$ определяется через вероятности:

$$p(y_1) = p(1-p_t-p_\theta) + (1-p)p_t = p_1,$$
$$p(y_2) = pp_t + (1-p)(1-p_t-p_\theta) = p_2,$$
$$p(\theta) = pp_\theta + (1-p)p_\theta = p_\theta,$$

где обозначения p_1 и p_2 введены для сокращения записи. Из последнего равенства следует, что безусловная вероятность $p(\theta)$ равна условной вероятности $p(\theta)$.

Принимая во внимание, что $p_1 + p_2 + p_\theta = 1$, имеем:

$$H(Y) = -\left[p_1 \log p_1 + p_2 \log p_2 + p_\theta \log p_\theta\right] =$$
$$-\left[p_1 \log p_1 + (1-p_1-p_\theta)\log(1-p_1-p_\theta) + p_\theta \log p_\theta\right]$$

Отсюда следует, что $H(Y)$ зависит от p (от распределения источника) только через p_1.

Значение p, при котором $H(Y)$ достигает своего максимального значения, определяется из уравнения $\dfrac{\partial H(Y) \partial p_1}{\partial p_1 \partial p} = 0$,

которое для данного случая имеет вид: $(1 - p_\theta - 2p_\ell)\log\dfrac{1-p_1-p_\theta}{p_1} = 0$.

Последнее уравнение распадается на два уравнения:

$$1 - p_\theta - 2p_\ell = 0,$$

$$\log\frac{1-p_1-p_\theta}{p_1} = 0.$$

Если параметры канала p_θ и p_ℓ удовлетворяют первому уравнению, то есть p_ℓ=0.5 (1-p_θ), то пропускная способность канала C = 0. При этом $p_1 = p_2 = p_\ell$ независимо от значения вероятности p.

4.6. Помехоустойчивое кодирование

4.6.1. Теоремы К. Шеннона

Появление теории помехоустойчивого кодирования было вызвано необходимостью обеспечения надёжной передачи сообщений при наличии помех, которые всегда существуют в реальных системах.

Рассмотрим наипростейший способ повышения надёжности передачи сообщений посредством многократной передачи одного и того же сообщения.

Пусть по каналу связи каждое из сообщений x_1=1 и x_2=0, образующих множество X, передаётся по три раза. Тогда передачу сообщения x_1=1 (x_2=0) можно рассматривать как передачу по каналу связи соответствующего кодового слова 111=$\vec{x}_1$ (000=$\vec{x}_2$). Число различных последовательностей, которые при этом могут появиться на выходе канала связи, равно 2^3=8. Они образуют множество $\vec{Y}$, которое назовём множеством выходных последовательностей. Кодовые слова $\vec{x}_1$ и $\vec{x}_2$ также можно считать принадлежащими некоторому множеству X, которое назовём множеством

входных канальных последовательностей. Оно состоит из 2^3 последовательностей.

Для отображения характера взаимодействия помехи и сигнала вводится модель канала связи, с помощью которой устанавливается связь между передаваемой последовательностью $\vec{x}_k$ и принимаемой последовательностью $\vec{y}_j$ в виде:

$$\vec{y}_j = \vec{x}_k \oplus \vec{e}_q,$$

где $\vec{e}_q$ - вектор ошибок, который представляет собой последовательность нулей и единиц. Единица указывает место, где произошла ошибка, а ноль указывает на отсутствие ошибки. Суммирование производится по модулю два.

Пусть в канале действует единичная ошибка, т.е. вектор $\vec{e}_q$ может содержать только одну единицу. Из-за единичной ошибки кодовое слово $111=\vec{x}_1$ ($000=\vec{x}_2$) может перейти в такие кодовые слова множества Y, которые содержат один ноль (два нуля), что позволяет безошибочно определить, какое кодовое слово было передано, а, следовательно, и передаваемое сообщение. Таким образом, множество $\vec{Y}$ делится на два непересекающихся подмножества, которые находятся во взаимно однозначном соответствии с сообщениями $\vec{x}_1$ и $\vec{x}_2$.

Последовательности множества $\vec{X}'$, подлежащие передаче, называются разрешёнными, а все остальные – запрещёнными. Кодирование, в сущности, сводится к разбиению множества X на разрешённые и запрещённые последовательности (кодовые слова). Часто кодом называют множество разрешённых кодовых слов. Выбор разрешённых кодовых слов (выбор кода) определяет верность передачи сообщений.

Таким образом, верность передачи сообщения можно повысить за счёт увеличения числа повторений, т.е. за счёт существенного снижения скорости передачи информации. В рассматриваемом случае скорость передачи информации уменьшилась в три раза за счёт трёхкратного повторения. Однако в некоторых условиях ошибку декодирования можно сделать сколь угодно малой посредством выбора соответствующего кода, если производительность источника будет меньше пропускной способности канала на любую сколь угодно малую величину. Существование такого способа кодирования (кода) было доказано К. Шенноном и сформулировано им в виде следующих теорем.

Теорема 1. Если производительность источника H_u меньше пропускной способности канала C, то существует такой код и способ декодирования, при которых возможна передача сообщений со сколь угодно малой вероятностью ошибки, при неограниченном возрастании длины последовательности n.

Теорема доказывается при следующей мысленной организации способа передачи сообщений. Подлежащие передаче сообщения представляют собой последовательность символов, которые вырабатывает, в общем случае, марковский источник сообщений, производительность которого равна $H_и$. При неограниченном увеличении длины последовательности n источник вырабатывает с вероятностью, равной единице, типичные последовательности, количество которых равно $2^{nH_и}$. Будем передавать только типичные последовательности (сообщения), а нетипичные последовательности вообще не будем передавать, соглашаясь с тем фактом, что в случае передачи нетипичной последовательности всегда имеет место ошибка.

Таким образом, вероятность появления ошибки равна вероятности появления нетипичной последовательности и стремится к нуля при неограниченном увеличении *n*.

Специальный «генератор», производительность которого равна *H(X')*, вырабатывает входные канальные последовательности из *n* независимых символов. Используются не все, а только типичные входные канальные последовательности, количество которых равно $2^{nH(X')}$. Нетипичные последовательности нет смысла использовать, поскольку они появляются с вероятностью равной нулю, при неограниченном увеличении *n*. Множество X'_T входных типичных последовательностей разбивается на разрешённые и запрещённые последовательности. Кодирование осуществляется посредством установления взаимно однозначного соответствия между разрешёнными последовательностями и типичными последовательностями, которые вырабатывает источник сообщений.

В теореме утверждается, что при $H_и < C = H(X') - H(X'/Y)$ существует такое разбиение множества X'_T на разрешённые и запрещённые последовательности (существует такой код), при котором вероятность ошибки будет меньше любой сколь угодно малой наперёд заданной величины, если выбрать достаточно большую длину последовательности *n*. Докажем, что условие $H_и < C$ является необходимым для передачи сообщения со сколь угодно малой вероятностью ошибки.

С позиций внешнего наблюдателя передающую и приёмную части можно рассматривать как одну сложную схему, а передачу отдельного символа, как реализацию некоторого состояния (x_i, y_j) сложной системы.

Если количество переданных и принятых символов *n* достаточно велико, то с вероятностью, близкой к единице, будут появляться типичные последовательности $\vec{x}_k$ и $\vec{y}_j$ и типичные последовательности состояний сложной системы $(\vec{x}_k, \vec{y}_j)$, количество которых соответственно равно:

$$Q(\vec{X}') \approx 2^{nH(X')},$$
$$Q(\vec{Y}) \approx 2^{nH(Y)},$$

$$Q(\vec{X}',\vec{Y}) \approx 2^{nH(X',Y)}.$$

Указанные множества типичных последовательностей обозначим, соответственно, через X'_T, Y'_T, $(X', Y)_T$. С учётом равенства $H(X', Y) = H(Y) + H(X'/Y)$,
количество типичных последовательностей:

$$Q(\vec{X}',\vec{Y}) = 2^{nH(Y)}2^{nH(X'/Y)} = 2^{nH(X')}2^{nH(Y/X')} = Q(\vec{Y})2^{nH(X'/Y)} = Q(\vec{X}')2^{nH(Y/X')}. \qquad (4.10)$$

Если множества X' и Y статистически независимы, то

$$Q(\vec{X}',\vec{Y}) = 2^{nH(Y)}2^{nH(X')} = Q(\vec{Y})Q(\vec{X}').$$

Наличие статистической зависимости между множествами X' и Y уменьшает число типичных последовательностей $Q(X', Y)$, поскольку всегда

$$H(X') > H(X'/Y) \text{ и } H(Y) > H(Y/X').$$

Таким образом, в образовании типичной последовательности $(\vec{x}_k, \vec{y}_j) \in (\vec{X}, \vec{Y})_T$ состояний сложной системы могут участвовать все типичные последовательности $\vec{y}_j \in \vec{Y}_T$ и $\vec{x}_k \in \vec{X}_T$, но не во всех сочетаниях (рис. 4.6).

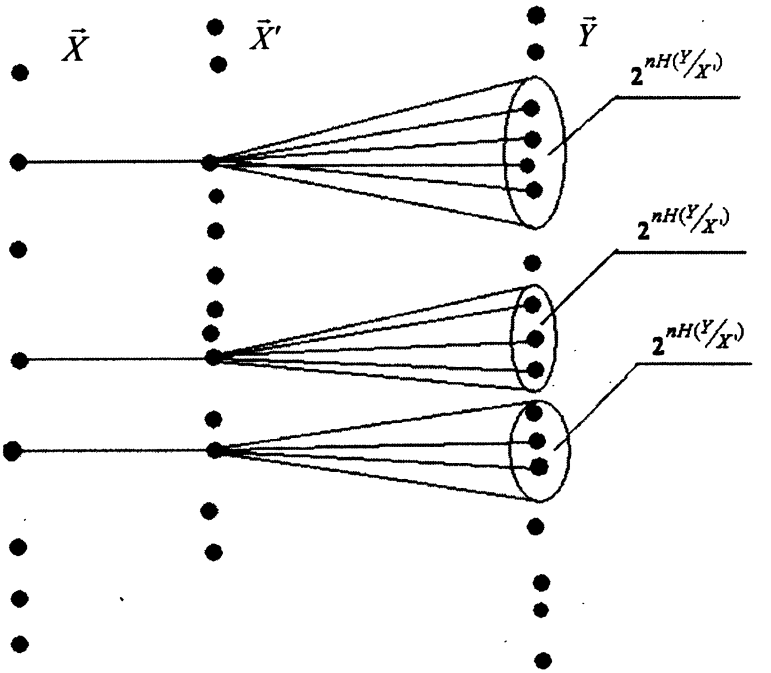


Рис. 4.6. Вероятностная диаграмма, описывающая модели системы связи К. Шеннона

Обозначим через $Q(\vec{Y}/\vec{x}_k)$ число типичных последовательностей $\vec{y}_j$, которые в сочетании с типичной последовательностью $\vec{x}_k$ могут образовывать типичные последовательности $(\vec{x}'_k, \vec{y}_j)$. Тогда общее число типичных последовательностей

$$Q(\vec{X}'\vec{Y}) = \sum_{k=1}^{Q(\vec{X}')} Q(\vec{Y}/\vec{x}'_k).$$

Можно доказать, что $Q(\vec{Y}/\vec{x}'_k)$ не зависит от номера типичной последовательности k. Поэтому (последнее равенство) можно переписать в виде

$$Q(\vec{X}'_k\vec{Y}) = Q(\vec{X}')Q(\vec{Y}/\vec{x}'_k).$$

Отсюда с учётом (4.10) получим

$$Q(\vec{Y}/\vec{x}'_k) = \frac{Q(\vec{X}',\vec{Y})}{Q(\vec{X}')} = 2^{nH(Y/X')}.$$

В результате аналогичных рассуждений можно получить, что

$$Q(\vec{X}'/\vec{y}_j) = \frac{Q(\vec{X}',\vec{Y})}{Q(\vec{Y})} = 2^{nH(X'/Y)}.$$

Таким образом, при передаче входной канальной последовательности $\vec{x}_k$ с вероятностью, равной единице, будет принята последовательность $\vec{y}_j$, принадлежащая подмножеству $\vec{Y}_T(\vec{x}_k)$, содержащему $2^{nH(Y/X')}$ последовательностей. Аналогично, если была принята последовательность $\vec{y}_j$, то с вероятностью, близкой к единице, была передана последовательность $\vec{x}_k$, принадлежащая подмножеству $\vec{X}'_T(\vec{y}_j)$, содержащему $2^{nH(X'/Y)}$ последовательностей (рис. 4.6). Декодирование с вероятностью, равной единице, будет безошибочным только в том случае, если разрешённые последовательности набирать таким образом, чтобы соответствующие подмножества $\vec{Y}(\vec{x}_k)$ не пересекались. Максимальное количество непересекающихся подмножеств равно

$$N \approx \frac{Q(\vec{Y})}{Q(\vec{Y}/\vec{x}_k)} \approx 2^{n[H(Y)-H(Y/X')]} \approx 2^{nC},$$

где пропускная способность канала

$$C = H(Y) - H(Y/X') = \frac{\log N}{n}$$

определяется максимальным количеством канальных последовательностей N (разрешённых последовательностей), которые безошибочно могут быть переданы по каналу связи. Очевидно, количество типичных последовательностей, которые вырабатывает источник сообщений должно быть меньше N:

$$2^{nH_u} < N = 2^{nC}.$$

Отсюда $H_u < C$.

Теорема 2. Если $H_u < C$, то среди кодов, обеспечивающих сколь угодно малую вероятность ошибки, существует код, при котором скорость передачи информации R сколь угодно близка к скорости создания информации H_u.

Идея доказательства состоит в следующем. Скорость передачи информации (на символ) определяется как

$$R = H_u - H(x/Y),$$

где $H(x/Y)$ - апостериорная энтропия переданного сообщения на символ, или количество рассеянной в канале информации. Доказывается, что энтропия $H(x/Y)$ может быть меньше сколь угодно малой величины, если выбрать достаточно большое значение n. Отсюда следует, что R сколь угодно мало отличается от H_u при неограниченном увеличении n.

Из теоремы следует, что C можно интерпретировать как предельную скорость передачи информации, при которой можно получить вероятность ошибки.

Теорема 3. (Обратная теорема)

Если скорость создания информации H_u больше пропускной способности канала C, то никакой код не может обеспечить сколь угодно малую вероятность ошибки. Минимальное рассеяние информации на символ, достижимое при $H_u > C$, равно $H_u - C$; никакой код не может обеспечить меньшего рассеяния информации.

В теоремах не затрагиваются вопросы построения оптимального кода. Тем не менее, значение их огромно, поскольку, обосновав принципиальную возможность такого кодирования, они мобилизовали усилия ученых на разработку конкретных кодов.

4.6.2. Линейные корректирующие коды

Надежность передачи сообщений по каналу связи при наличии помех зависит от вида выбранного кода. Под кодом будем понимать множество разрешенных (передаваемых) кодовых слов, которые выбираются из множества входных канальных последовательностей (кодовых слов). Разре-

шенные кодовые слова (код) следует выбрать таким образом, чтобы вероятность ошибки была минимальной. Вероятность ошибки в данном случае является критерием, т.е. количественной характеристикой качества функционирования системы связи. Она показывает, как часто ошибается приемник (код) в заданной помеховой обстановке и позволяет выбрать наилучший. Приемник (код), обеспечивающий минимум вероятности ошибки, называется оптимальным в заданной помеховой обстановке. Этот критерий был предложен академиком Котельниковым, а соответствующий оптимальный приемник был назван им «идеальным наблюдателем».

С целью уменьшения вероятности ошибки, разрешенные кодовые слова следует выбирать так, чтобы они как можно меньше были похожи друг на друга, что повышает надежность их различения на фоне помех.

В качестве меры различия кодовых слов выбирается расстояние Хемминга, т.е. количество разрядов, в которых кодовые слова различаются. Поэтому корректирующую способность кода можно характеризовать минимальным расстоянием d между разрешенными кодовыми словами, которое называется минимальным кодовым расстоянием.

Декодирование осуществляется по минимуму расстояния Хемминга между принятым кодовым словом и разрешенными кодовыми словами, которые хранятся в памяти приёмника, т.е. решение принимается в пользу разрешенного кодового слова, наиближайшего к принятому. В этом случае ошибки кратности $t \le d-1$ обнаружимы, а кратности $t < \frac{d}{2}$ или $2t+1 \le d$ исправимы. Кратность ошибки – это количество единиц в векторе $\vec{e}$.

При большой длине кодовых слов такой способ декодирования не удается реализовать в реальном масштабе времени из-за большого объема вычислений и, кроме этого, требуется большой объем памяти для хранения кодовых слов.

Преодолеть указанные трудности, в частности, позволяет использование линейных корректирующих кодов. В этом случае вместо запоминания всех разрешенных кодовых слов запоминается только система линейных уравнений, решениями которой являются разрешенные кодовые слова.

Основной задачей построения линейных корректирующих кодов является выбор системы линейных уравнений, которая записывается в виде:

$$h_{11}x_1 \oplus h_{12}x_2 \oplus \ldots \oplus h_{1n}x_n = 0$$
$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$
$$h_{m1}x_1 \oplus h_{m2}x_2 \oplus \ldots \oplus h_{mn}x_n = 0$$

или в матричном виде $H\vec{x} = 0$, где $H = \{h_{ij}\}$ - проверочная матрица, а кодовое слово размера n – $\vec{x} = (x_1, \ldots, x_n)$. Элементы матрицы и кодового

слова принимают значения только 0 или 1, при этом суммирование производится по модулю два ($\oplus$). В этой системе уравнений количество строк m выбирается меньше количества столбцов для того, чтобы система уравнений имела не единственное решение, каждое из которых является разрешенным кодовым словом. Каждое из уравнений уменьшает количество независимых переменных x_i на единицу. Поэтому количество независимых переменных, значения которых и место расположения можно выбирать произвольно, равно $n-m=k$. Если в качестве независимых переменных выбираются первые k символов $x_1,\ldots,x_k$ то код называется систематическим. Эти символы называются информационными в отличие от защитных $x_{k+1},\ldots,x_n$ причем защитные символы линейно выражаются через информационные. Таким образом, количество разрешенных кодовых слов равно 2^k, а общее количество входных канальных кодовых слов равно 2^n.

В случае линейных корректирующих кодов сумма двух разрешенных кодовых слов является разрешенным кодовым словом. Это основное свойство, которое можно использовать в качестве определения линейного корректирующего кода, следует из равенства

$$H(\vec{x}_i \oplus \vec{x}_n) = H\vec{x}_i \oplus H\vec{x}_k,$$

где $H\vec{x}_i = 0$ и $H\vec{x}_k = 0$ по определению разрешенного кодового слова.

Рассмотрим декодирование по синдрому, использующее согласно выбранной модели канала связи принятое кодовое слово $\vec{y} = \vec{x} \oplus \vec{e}$, где $\vec{x}$ - разрешенное кодовое слово, $\vec{e}$ - вектор ошибок.

Если $\vec{y}$ совпадает с разрешенным кодовым словом ($\vec{e} = 0$ или $\vec{e} = \vec{x}$ - случай необнаружимой ошибки), то правая часть системы уравнений равна нулю: $H\vec{y} = 0$, $\vec{y} = \vec{x}$, и отлична от нуля, когда $\vec{y}$ - запрещенное кодовое слово: $H\vec{y} \neq 0$. В общем случае $H\vec{y} = \vec{S}$, где вектор-столбец $\vec{S}$ называется синдромом.

Поскольку система уравнений линейна (H – линейный оператор), то справедливы следующие преобразования:

$$H\vec{y} = H(\vec{x} \oplus \vec{e}) = H\vec{x} \oplus H\vec{e} = \vec{S},$$

отсюда $H\vec{e} = \vec{S}$, т.к. $H\vec{x} = 0$ по определению разрешенного кодового слова $\vec{x}$.

Равенство $H\vec{e} = \vec{S}$ устанавливает связь между синдромом $\vec{S}$ и вектором ошибок $\vec{e}$, но, к сожалению, неоднозначную, поскольку не существует матрицы, обратной к H.

Каждому значению синдрома, количество которых равно 2^m, априори до приема кодового слова $\vec{y}$, соответствует некоторое подмножество векторов $\vec{e}$, из которого при декодировании выбирается наиболее вероятный вектор $\vec{e}$ с минимальным весом. (Одно из значений синдрома равно ну-

лю, когда вектор $\vec{e}$ совпадает с разрешенным кодовым словом, что соответствует случаю необна-ружимой ошибки).

Решение принимается в пользу кодового слова $\vec{x}^* = \vec{y} \oplus \vec{e}^*$, которое является разрешенным.

Таким образом, при декодировании по принятому вектору $\vec{y}$ вычисляется синдром, который используется при вычислении $\vec{e}^*$ и $\vec{x}^*$.

Декодирование по минимуму расстояния, по синдрому и по максимуму апостериорной вероятности $p\left(\vec{x}/\vec{y}\right)$ эквивалентны.

Эквивалентность декодирования по минимуму расстояния и по синдрому следует из равенства $\vec{e}^* = \vec{y} \oplus \vec{x}^*$, которое показывает, что расстояние между $\vec{y}$ и $\vec{x}^*$ измеряется весом (количеством единиц) вектора $\vec{e}^*$.

Вектор $\vec{x}^*$ является оценкой максимального правдоподобия, поскольку вектор ошибок с минимальным весом $\vec{e}^*$, однозначно определяющий $\vec{x}^*$, имеет максимальную вероятность при малых вероятностях ошибок в отдельном разряде вектора $\vec{e}$ ($\sim 10^{-2} - 10^{-3}$).

Рассмотрим некоторые свойства линейных кодов. Количество векторов $\vec{e}$ в каждом из подмножеств, соответствующих различным значениям синдрома, равно количеству разрешенных кодовых слов 2^k. Для любых двух векторов $\vec{e}_1$ и $\vec{e}_2$, принадлежащих одному и тому же подмножеству, справедливы равенства $H\vec{e}_1 = \vec{S}_i$ и $H\vec{e}_2 = \vec{S}_i$. Складывая эти равенства и преобразуя суммы с учетом линейности системы уравнений (линейности оператора Н), получим

$$H\vec{e}_1 \oplus H\vec{e}_2 = \vec{S}_i \oplus \vec{S}_i \quad , \quad H\left(\vec{e}_1 \oplus \vec{e}_2\right) = 0.$$

Отсюда следует, что сумма $\vec{e}_1 \oplus \vec{e}_2$ является разрешенным кодовым словом по его определению, т.е. все векторы подмножества различаются между собой только на разрешенные кодовые слова.

Таким образом, если один из векторов подмножества выбрать в качестве исходного (порождающего), то все остальные можно получить в результате его сложения последовательно с каждым разрешенным кодовым словом, количество которых равно 2^k.

Минимальное кодовое расстояние в случае линейных кодов совпадает с минимальным весом разрешенных кодовых слов.

Действительно, расстояние между двумя кодовыми словами измеряется весом их суммы. Для линейного кода эта сумма совпадает с одним из разрешенных кодовых слов, вес которого определяет расстояние между двумя выбранными разрешенными кодовыми словами.

Рассмотрим свойства матрицы H, которые гарантируют заданное значение минимального кодового расстояния. Для каждого кодового слова уравнение $H_{\vec{x}} = 0$ означает, что сумма некоторого подмножества столб-

цов матрицы H равна 0. Приумножении матрицы H на вектор-столбец $\vec{x}$ из матрицы выбираются столбцы, которым соответствуют единицы в векторе $\vec{x}$. Поскольку минимальное кодовое расстояние равно минимальному весу кодовых слов, то должно существовать, по крайней мере, одно подмножество, состоящее из d столбцов матрицы H, сумма которых равна 0. С другой стороны, не может существовать ни одного подмножества из d-1 или менее столбцов, сумма которых равна 0. Если рассматривать столбцы матрицы H как векторы, то можно сказать, что для кода с кодовым расстоянием, равным d, все подмножества из d-1 столбцов должны быть линейно независимыми. Это утверждение составляет одну из фундаментальных теорем о групповых кодах. Оно позволяет находить кодовое расстояние d группового кода, заданного матрицей H, а также строить матрицы H, гарантирующие заданное минимальное кодовое расстояние.

Декодированию по синдрому можно дать следующую физическую интерпретацию: вектор $\vec{y}$ подается на вход линейного дискретного фильтра, который описывается линейным оператором H, а на выходе фильтра наблюдается вектор $\vec{S}$. Фильтр не пропускает на выход разрешенные кодовые слова $(H\vec{x} = 0)$, а помеха $\vec{e}$, искаженная фильтром наблюдается на его выходе в виде синдрома $\vec{s}$.

По синдрому с некоторой точностью восстанавливается (оценивается) вид вектора ошибок на входе фильтра. Оценка $\vec{e}^*$ этого вектора используется для компенсации вектора ошибок $\vec{e}$, действующего в канале. Механизм компенсации можно пояснить с помощью равенства:

$$\vec{x}^* = \vec{y} \oplus \vec{e}^* = \vec{x} \oplus \vec{e} \oplus \vec{e}^*.$$

При декодировании ошибка отсутствует, когда $\vec{e} = \vec{e}^*$.

4.7. Передача непрерывных сообщений

4.7.1. Дискретизация непрерывных сообщений и сигналов

Произвольную кусочно-непрерывную функцию $x(t)$, изображающую сообщение или сигнал, можно разложить в обобщенный ряд Фурье по полной системе ортонормированных функций:

$$x(t) = C_0\varphi_0(t) + C_1\varphi_1(t) + \ldots + C_i\varphi_i(t) + \ldots ,$$

если энергия функции $E_x = \int_{-\infty}^{\infty} x^2(t)dt$ конечна [9].

Бесконечная система действительных функций $\varphi_0(t),\ldots,\varphi_i(t),\ldots$ называется ортогональной на отрезке $[a, b]$, если

$$\int_a^b \varphi_i(t)\varphi_m(t)dt = 0 \qquad \text{при} \quad i \neq m,$$

а отдельная функция $\varphi_i(t)$ называется нормированной, если

$$\int_a^b \varphi_i^2(t)dt = 1.$$

Система нормированных функций, в которой каждые две различающихся функции взаимно ортогональны, называется ортонормированной системой. При аппроксимации функции $x(t)$ ограничиваются, как правило, конечным числом членов ряда. При заданной системе функций $\varphi_i(t)$ и при фиксированном количестве членов ряда n значения коэффициентов C_i можно выбрать такими, при которых среднеквадратичная ошибка аппроксимации

$$\int_a^b \left[x(t) - \sum_{i=1}^{n} C_i\varphi_i(t) \right]^2 dt$$

достигает минимума. Ряд, обеспечивающий минимум среднеквадратической ошибки, называется рядом Фурье. Минимум среднеквадратичной ошибки достигается в том случае, когда коэффициенты ряда определяются по формуле

$$C_i = \int_a^b x(t)\varphi_i(t)dt \ ,$$

при этом они вычисляются независимо друг от друга при любом значении n и называются коэффициентами Фурье.

Ряд с определяемыми таким образом коэффициентами называется обобщенным рядом Фурье.

Ортогональная система называется полной, если путем увеличения количества членов в ряде среднеквадратичную ошибку можно сделать сколь угодно малой.

Таким образом, по счетному множеству коэффициентов C_i можно с определенной точностью восстановить соответствующую функцию $x(t)$, можно заменить передачей последовательности коэффициентов $C_0, C_1, ..., C_n$. Указанную последовательность можно интерпретировать как вектор в n - мерном Евклидовом пространстве с координатами $C_0, C_1, ..., C_n$, квадрат длины которого

$$\left|\vec{C}\right|^2 = \sum_{i=0}^{n} C_i^2 .$$

Последнее равенство является обобщением теоремы Пифагора на случай *n*-мерного пространства. Путем непосредственных вычислений легко установить, что энергия сигнала

$$E_x = \int x^2(t)dt \approx \sum_{i=0}^{n} C_i^2 .$$

Таким образом, дискретизацией называется замена непрерывной функции $x(t)$, определенной на интервале, равном T, последовательностью коэффициентов $C_0, C_1, ..., C_n ...$ (вектором).

Выбор системы ортогональных функций $\varphi_i(t)$ определяется целью и физической сущностью решаемой задачи, а не чисто математическими умозаключениями.

С целью передачи сигнала по каналу связи широко применяется разложение функции $x(t)$ в ряд Котельникова, которое позволяет существенно упростить определение коэффициентов C_i.

Согласно теореме Котельникова произвольная функция $x(t)$ с ограниченным спектром, может быть тождественно представлена счетным множеством ее значений, взятых через интервал времени $\Delta t = \frac{1}{2F}$, где F - верхняя граничная частота спектра сигнала. В этом случае функции

$$\varphi_i(t) = \frac{\sin 2\pi F(t - i\Delta t)}{2\pi F(t - i\Delta t)} ,$$

образующие систему ортогональных функций, отличаются друг от друга только сдвигом по оси времени t на величину кратную $\Delta t = \frac{1}{2F}$, при этом каждая из них достигает своего максимального значения в те моменты времени, когда значения всех остальных функций равны нулю. Коэффициенты разложения определяются по формуле

$$C_i = \frac{1}{\Delta t} \int_{-\infty}^{\infty} x(t)\varphi_i(t)dt ,$$

которую в результате тождественных преобразований можно привести к виду: $C_i = x(i\Delta t)$, то есть коэффициент C_i равен значению функции $x(t)$ в момент, когда функция $\varphi_i(t)$ достигает своего максимального значения.

Если дискретизации подлежит нормальный (гауссов) случайный процесс, энергетический спектр которого имеет прямоугольную форму, то коэффициенты C_i будут статистически независимыми случайными величинами, которые совпадают со значениями случайной функции $x(t)$, взятыми с шагом Δt [9].

Таким образом, непрерывные сообщения можно передавать в цифровом виде, то есть в виде последовательности чисел, при этом каждое число приближенно выражает величину соответствующего коэффициента C_i.

4.7.2. Энтропия системы с непрерывным множеством состояний

Система называется непрерывной по данному описывающему ее параметру, если этот параметр является непрерывной величиной. Состояния такой системы нельзя перенумеровать: они непрерывно переходят одно в другое, причем каждое отдельное состояние имеет вероятность, равную нулю, а распределение вероятностей характеризуется некоторой плотностью $P(x)$. Например, генератор шума, напряжение на выходе которого может принимать любое значение, является непрерывной системой по указанному параметру.

Попытаемся ввести количественную меру неопределенности непрерывной случайной величины X через энтропию $H(X')$ дискретной случайной величины X', которая получается в результате квантования непрерывной величины X по уровню. Математически квантование можно представить как нелинейное преобразование непрерывной величины X. Вся область возможных значений величины X разбивается на интервалы с длиной, равной Δx. Каждому интервалу ставится в соответствие некоторое значение x_i', принадлежащее дискретному множеству X'.

Вероятность появления значения x_i' равна вероятности попадания случайной величины X в соответствующий интервал. Чем меньше интервал квантования Δx, тем точнее дискретная величина X' отображает свойства непрерывной величины X. Поэтому в качестве количественной меры неопределенности случайной величины X логично использовать значение энтропии $H(X')$ при Δx стремящимся к нулю:

$$H(X')=-\lim_{\Delta x\to 0}\sum_{i=-\infty}^{\infty} w(x_i)\Delta x\log[w(x_i)\Delta x]= \\ =-\int_{-\infty}^{\infty} w(x)\log w(x)dx-\lim_{\Delta x\to 0}\log\Delta x. \qquad (4.11)$$

Первый член в (4.11) не зависит от Δx - степени точности, с которой определяется состояние системы. От Δx зависит только второй член $(-\log\Delta x)$, который стремится к бесконечности при Δx, стремящимся к нулю. Это и естественно, поскольку, чем точнее мы хотим задать состояние системы, тем большую неопределенность мы должны снять.

Таким образом, мы убедились, что система с непрерывным множеством состояний не допускает введения конечной абсолютной меры неопределенности. Однако можно ввести количественную меру неопределенности указанной системы по отношению к другой непрерывной системе, состояния которой описываются случайной величиной Y с некоторым стандартным распределением. В качестве последнего (эталонного) удобно использовать равномерное в некотором интервале d распределение

$$w(Y)=\begin{cases} \dfrac{1}{d}, -\dfrac{d}{2}<y<\dfrac{d}{2} \\ 0-\text{в противном случае} \end{cases}.$$

Энтропия $H(Y')$ вычисляется аналогично выражению (4.11)

$$H(Y')=\int_{-\frac{d}{2}}^{\frac{d}{2}}\frac{1}{d}\log\frac{1}{d}dy-\log\Delta x=\log d-\log\Delta x.$$

Относительной (дифференциальной) энтропией случайной величины X называется величина

$$H(x)=H(X')-H(Y')=-\int_{-\infty}^{\infty}w(x)\log[dw(x)]dx.$$

В частности, если интервал $d=1$, то

$$H(X)=-\int_{-\infty}^{\infty}w(x)\log w(x)dx\ .$$

Выясним физический смысл относительной энтропии $H(X)$.

Пусть источник сообщений вырабатывает последовательность значений случайной величины X. После квантования получим последовательность значений случайной величины X':

$$x'_{i,1},x'_{i,2},\ldots,x'_{i,k},\ldots,x'_{i,n}\ .$$

При неограниченном увеличении длины последовательности с вероятностью, равной единице, появляются только типичные последовательности, число которых

$$Q=2^{nH(X')}\approx 2^{nH(X)-n\log\Delta x}=2^{nH(X)-\log(\Delta x)^n}=2^{nH(X)}\cdot\frac{1}{\Delta V_n},$$

где $\Delta V_n=(\Delta x)^n$ - число элементарного n-мерного кубика. Конец вектора, изображающего типичную последовательность, является внутренней точкой этого кубика. Произведение $\Delta V_n\cdot Q=2^{nH(X)}$ равно объему некоторой области в n-мерном пространстве, внутренние точки которой изображают концы типичных векторов (последовательностей). При Δx, стремящихся к нулю, число типичных последовательностей стремится к бесконечности,

объем каждого элементарного кубика стремится к нулю. При этом объем V_T, занимаемый типичными последовательностями, остается постоянным, равным $2^{nH(X)}$.

Энтропию в дискретном случае можно было определить через число типичных последовательностей:

$$H(X)=\lim_{n\to\infty}\frac{\log Q}{n}.$$

Аналогично относительную энтропию можно определить через объем V_T, занимаемый типичными последовательностями:

$$H(X)=\lim_{\substack{n\to\infty\\ \Delta x\to 0}}\frac{\log V_T}{n}.$$

В отличие от дискретного случая относительная энтропия может быть не только положительной, но и отрицательной, а также равной нулю. Чем больше объем V_T, занимаемой типичными последовательностями, тем больше неопределенность того, какая из них появится. Единичному объему (V_T=1) соответствует энтропия (неопределенность), равная нулю ($H(x)$=0). Это значение принимается за начало отсчета относительной энтропии.

В частности, относительная энтропия случайной величины с равномерным на единичном интервале (d = 1) распределением равна нулю:

$$H(X)=\log d=0\ .$$

В этом случае область n-мерного пространства, занимаемая типичными последовательностями, примерно совпадает с областью определения всех последовательностей и имеет форму куба единичного объема ($V_T = d^n$ =1).

4.7.3. Экстремальные свойства энтропии

Одной из основных характеристик, используемых при проектировании информационных систем, является энтропия. Поэтому часто возникает необходимость определения закона распределения случайной величины X, при котором энтропия $H(X)$ имеет максимальное значение. Например, эффективность искусственно создаваемых помех тем выше, чем больше значение энтропии $H(X)$. Поэтому при заданной мощности генератора целесообразно выбирать тот закон распределения помехи, при котором значение энтропии $H(X)$ максимально.

Для дискретного множества X было установлено, что при равномерном распределении вероятностей энтропия $H(X)$ имеет максимальное зна-

чение, равное $\log m_x$. Однако в случае систем с непрерывным множеством состояний аналогичная задача не имеет решения, если на непрерывную случайную величину X априори не наложить некоторые ограничения. Например, можно ограничить дисперсию или область определения случайной величины.

Экстремальные свойства относительной энтропии удобно интерпретировать через объем подпространства, который занимают типичные последовательности. Согласно равенству $V_T \approx 2^{nH(x)}$, чем больше объем указанного пространства, тем больше относительная энтропия $H(X)$ (больше неопределенность того, какая из типичных последовательностей будет выбрана).

Приведем несколько примеров.

1. Пусть известно, что область X возможных значений случайной величины ограничена интервалом $[a,b]$, $(a \le x \le b)$. Найдем распределение, обладающее при этом максимальной относительной энтропией. В данном случае область определения всех возможных последовательностей представляет собой n- мерный куб, сторона которого равна b - a. Очевидно энтропия $H(X)$ будет иметь максимальное значение, если подобласть определения типичных последовательностей будет совпадать с областью определения всех последовательностей. Только в этом случае объем V_T имеет максимальное значение $(b-a)^n$, равное объему n - мерного куба, сторона которого равна $(b-a)$. При этом энтропия

$$H(x) = \lim_{n\to\infty} \frac{\log V_T}{n} = \log(b-a).$$

Приведенные рассуждения не являются строгим доказательством последнего равенства. Строгое доказательство можно найти в литературе.

2. Если на случайную величину X наложить следующие ограничения:

а) область возможных значений неограничена $(-\infty \le x \le \infty)$;

б) известно среднее значение величины X;

в) задана дисперсия σ_x^2 случайной величины X, то закон распределения, доставляющий максимум энтропии $H(X)$, будет нормальным.

3. В случае, когда x может принимать только положительные значения $(w(x) = 0$ при $x < 0)$ и первый момент X равен a, максимальное значение энтропии $H(X)$ достигается при

$$w(x) = \frac{1}{a}\exp\left(\frac{x}{a}\right).$$

4.7.4. Взаимная информация для систем с непрерывным множеством состояний

Представим непрерывные множества X и Y в виде суммы счетного множества непересекающихся интервалов Δx_i и Δy_j. Будем говорить, что система $A,(B)$ находится в i-ом (j-ом) состоянии, если её состояние описывается значением $x \in \Delta x_i (y \in \Delta y_j)$.

Проквантовав таким образом непрерывные множества X и Y, мы получим два дискретных множества X_Δ, Y_Δ, которые тем точнее описывают непрерывные множества, чем меньше интервалы Δx_i и Δy_j.

Вероятность нахождения системы $A,(B)$ в i-ом (j-ом) состоянии равна

$$P(x \in \Delta x_i) \approx w(x)\Delta x_i, (P(y \in \Delta y_j) \approx w(y)\Delta y_j ,$$

а совместная вероятность

$$P(x \in \Delta x_i, y \in \Delta y_j) = w(x,y)\Delta x_i \Delta y_j .$$

Теперь можно определить взаимную информацию между i- м элементом множества X_Δ и j- м элементом множества Y_Δ как

$$\log \frac{P(x \in \Delta x_i, y \in \Delta y_j)}{P(x \in \Delta x_i) P(y \in \Delta y_j}} .$$

Разделив числитель и знаменатель на произведение $\Delta x_i \Delta y_j$, получим:

$$\log \frac{\frac{1}{\Delta x_i \Delta y_j} P(x \in \Delta x_i, y \in \Delta y_j)}{\left[\frac{1}{\Delta x_i} P(x \in \Delta x_i)\right]\left[\frac{1}{\Delta y_j} P(y \in y_j)\right]} .$$

Переходя к пределу при Δx_i и Δy_j, стремящихся к нулю, получим взаимную информационную плотность

$$i(x,y) = \log \frac{w(x,y)}{w(x)w(y)} .$$

Информационная плотность определяет количество взаимной информации между бесконечно малыми интервалами Δx_i и Δy_j, которые стягиваются соответственно в точки x и y.

Взаимная информационная плотность является случайной величиной, среднее значение которой равно

$$I(X,Y) = \int_{-\infty}^{\infty} \int w(x,y) \log \frac{w(x,y)}{w(x)b(y)} dxdy .$$

Величину $I(X,Y)$ часто называют средней взаимной информацией или просто взаимной информацией.

В качестве примера вычислим взаимную информацию, которой обмениваются два абонента, разговаривающие по телефону. Напряжение, развиваемой на выходе микрофона, усиливается и с выхода усилителя подается в линию связи.

Это напряжение представляет собой нормальный случайный процесс $X(t)$ с дисперсией, равной σ_x^2.

В результате действия помех, слушающий абонент, вместо поступившей в линию реализации $x(t) \in X(t)$ примет реализацию

$$y(t) = x(t) + n(t),$$

где $n(t)$ - реализация нормального шума, дисперсия которого равна σ_n^2. Тогда случайный процесс $Y(t)$, которому принадлежит реализация $y(t)$, будет имеет нормальный закон распределения как сумма двух случайных процессов с нормальным (гауссовым) распределением. Поскольку процесс $X(t)$ и шум статистически независимы, то дисперсия $Y(t)$ будет равна $\sigma_Y^2 = \sigma_N^2 + \sigma_X^2$. Теперь можно вычислить взаимную информацию между двумя случайными величинами X и Y, которые равны значениям случайных процессов $X(t)$ и $Y(t)$ в некоторый фиксированный момент времени t.

При вычислении взаимной информации по формуле:

$$I(X,Y) = H(Y) - H(Y/X) \tag{4.12}$$

нам понадобятся следующие распределения случайной величины Y

$$w(y) = \frac{1}{\sqrt{2\pi}\sigma_Y} e^{-\frac{y^2}{2\sigma_Y^2}},$$

$$w(y/x) = \frac{1}{\sqrt{2\pi}\sigma_N} e^{-\frac{(y-x)^2}{2\sigma_N^2}}.$$

Энтропия

$$H(Y) = -\overline{\log w(y)} = \log\sqrt{2\pi}\sigma_y + \frac{\overline{y^2}}{2\sigma_Y^2}\log e = \frac{1}{2}\log 2\pi e\sigma_y^2.$$

При этом было использовано равенство $\sigma_Y^2 = \overline{y^2}$.

Условная энтропия

$$H(Y/X) = -\int_{-\infty}^{\infty} w(x)\left[\int_{-\infty}^{\infty} w(y/x)\log w(y/x)dy\right]dx$$

Усредняя

$$-\log w(y/x) = \log\sqrt{2\pi}\sigma_N + \frac{(y-x)^2}{2\sigma_N^2}\log e,$$

сначала по y, то есть вычисляя выражение в квадратных скобках, получим

$$-\overline{\log b(y/x)}^{y} = \log\sqrt{2\pi}\sigma_N + \frac{\overline{(y-x)^2}^{y}}{2\sigma_N}\log e = \frac{1}{2}\log 2\pi e\sigma_N^2.$$

При этом было использовано равенство

$$\overline{(y-x)^2}^{y} = \sigma_N^2.$$

Поскольку выражение в квадратных скобках не зависит от x, то

$$H(Y/X) = \frac{1}{2}\log 2\pi e\sigma_N^2 .$$

Подставляя в (4.12) вычисленные энтропии $H(Y)$ и $H(Y/X)$, получим

$$I(X,Y) = \frac{1}{2}\log(1+\frac{\sigma_X^2}{\sigma_N^2}).$$

4.7.5. Пропускная способность гауссова канала связи

Канал будем называть гауссовым, если он удовлетворяет следующим условиям: 1) полоса пропускания канала ограничена частотой F, 2) шум в канале имеет нормальный (гауссов) закон распределения, 3) энергетический спектр шума равномерен в полосе пропускания канала и имеет значение, равное N_0, 4) мощность сигнала $x(t)$ фиксирована и равна σ_X^2, 5) сигнал $x(t)$ и шум $n(t)$ статистически независимы, 6) канал аддитивен, т. е. выходной сигнал $y(t) = x(t) + n(t)$.

Пропускная способность канала $C = \max\limits_{w(x)}[H(Y) - H(Y/X)]$

Частная энтропия $H(Y/x)$ равна энтропии шума $H(N)$, поскольку из равенства $y = x + n$ следует, что значения y и n находятся во взаимно однозначном соответствии при фиксированном значении x. Энтропия $H(Y/X)$, которая получается в результате усреднения частной энтропии $H(Y/x)$ по всем значениям x, также равна $H(N)$. Отсюда $C = \max\limits_{P(X)} H(Y) - H(N)$. Максимальное значение энтропии $H(Y)$ доставляет гауссов закон распределения случайной величины Y при заданном значении дисперсии $\sigma_Y^2 = \sigma_X^2 + \sigma_N^2$. Дисперсия σ_Y^2 равна сумме дисперсий σ_X^2 и σ_N^2, поскольку сигнал и шум статистически независимы. При этом случайная величина X с необходимостью должна иметь гауссов закон

распределения как разность $X = Y - N$ случайных величин с гауссовыми законами распределения. Тогда значение пропускной способности будет равно значению взаимной информации $I(X,Y)$, которая определяется выражением

$$C = \frac{1}{2}\log(1+q^2) \quad [\text{бит/символ}],$$

где $q^2 = \frac{\sigma_X^2}{\sigma_N^2}$ называется отношением сигнал/шум.

Если отсчеты процесса $Y(t)$ на основании теоремы Котельникова выбрать с шагом $\Delta t = \frac{1}{2F}$, то за 1 секунду будет передано $2F$ отсчетов. Эти отсчеты в случае гауссова шума статистически независимы и поэтому пропускную способность канала можно определить следующим образом:

$$C_t = 2FC = F\log(1+q^2) \quad [\text{бит/символ}].$$

Заданного значения C_t можно достигнуть как за счет увеличения q^2 (мощности передатчика при заданной мощности шума) так и за счет увеличения полосы пропускания канала F. Однако, компенсировать возможное сокращение полосы пропускания канала за счет увеличения мощности передатчика, как правило, нецелесообразно, поскольку C_t зависит от q^2 по логарифмическому закону, а от F по линейному. Более эффективным является увеличение значения C_t за счет увеличения F. Однако, следует иметь в виду, что с увеличением полосы пропускания канала увеличивается и мощность шумов, попавших в канал, что приводит к снижению отношения сигнал/шум. Если шум в канале считать белым, то мощность шума $\sigma_N^2 = N_0 F$ и $C_t = F\log(1+\frac{\sigma_X^2}{N_0 F})$.

Пропускная способность C_t сначала быстро растет с ростом F, а затем асимптотически стремится к пределу $C_\infty = \frac{\sigma_X^2}{N_0}\log e$ [бит/символ].

Согласно второй теореме К.Шеннона надежная передача сообщений в принципе возможна при $R < C_\infty$. Умножая неравенство на время передачи сообщения T, получим

$$I = RT < C_\infty T = \frac{E_X}{N_0}\log e,$$

то есть количество информации I, которое надежно можно передать по каналу, должно быть меньше $\frac{E_X}{N_0}\log e$, где $E_X = \sigma_X^2 T$ - энергия сигнала. В

частности, для передачи 1 бита информации необходима энергия, равная $E_X > N_0 / \log e \approx 0{,}693 N_0$.

4.7.6. Эпсилон - энтропия

Проблема передачи непрерывного сообщения заключается в получении его копии на приемном пункте и, в сущности, сводится к процедуре воспроизведения сообщения на основе полученной информации. Очевидно, в данном случае не существует способа, позволяющего получить точную копию передаваемого сообщения, поскольку это требует бесконечной точности его воспроизведения, причем неограниченное увеличение точности требует неограниченного увеличения количества передаваемой информации. Например, нельзя получить два абсолютно совпадающих графика. Поэтому о передаче непрерывного сообщения имеет смысл говорить только в том случае, когда задана точность его воспроизведения.

Передача непрерывного сообщения $x(t)$ сводится к передаче последовательности его значений взятых в дискретные моменты времени. Все возможные значения функции $x(t)$ в некоторый момент времени образуют множество X.

Пусть случайная величина X имеет равномерное на отрезке $[a,b]$ распределение. Тогда распределение дискретной случайной величины X' также будет равномерным, если все интервалы на которые разбит отрезок $[a,b]$, имеют одну и ту же длину. При таком разбиении энтропии $H(X')$ достигает своего максимального значения, равного $\log N$, где N - число элементов в множестве X'.

Количество взаимной информации между множествами X и X', определяемое равенством

$$I(X,X') = H(X') - H(X'/X),$$

равно энтропии $H(X')$, поскольку неопределенность $H(X'/X)$, с которой значение случайной величины X определяет значение случайной величины X', равна нулю. Поэтому для воспроизведения значения случайной величины X с точностью Δx требуется количество информации, равное $H(X')$.

Пронумеруем все элементы множества X' числами от 1 до N и запишем их в m-ичной системе счисления. Минимальное количество разрядов n, некоторое при этом потребуется, удовлетворяет неравенству:

$$\log_m N \le n < \log_m N + 1 .$$

Если величина $\log_m N$ окажется целым числом, то количество информации, необходимое для воспроизведения значения случайной вели-

чины X с заданной точностью, непосредственно равно в m-ичных единицах количеству символов n (разрядов), передаваемых по каналу связи. Повышение точности требуют уменьшения длины интервала Δx, а, следовательно, и увеличения количества передаваемой информации.

Более универсальной мерой точности воспроизведения по сравнению с Δx является среднеквадратичная ошибка

$$\overline{[X-\varphi(X)]^2},$$

которую при неравномерном распределении случайной величины X можно минимизировать не только путем увеличения количества интервалов Δx, но и путем изменения их длины. При этом энтропия $H(X')$ также будет изменяться.

Таким образом, количество информации, которое требуется для воспроизведения значения случайной величины X с заданной точностью (с заданной среднеквадратичной ошибкой), зависит от выбранной меры точности и от характера статистической зависимости между множествами X и X' (от преобразования случайной величины X в случайную величину X'), в частности, от того, каким образом выбраны длины интервалов Δx_i и их количество. Канал, устанавливающий связь между множествами X и X', описывается условной вероятностью $p(X'/X)$ и, в сущности, представляет собой устройство квантования.

Определим ε - энтропию как минимальное по $p(X'/X)$ количество информации, необходимое для воспроизведения значения случайной величины X с заданной точностью при заданном распределении $p(x)$ источника:

$$H_\varepsilon(X)=\min_{p(X'/X)} I(X,X').$$

Следует напомнить, что пропускная способность канала C была определена как максимальное количество взаимной информации, но не по $p(X'/X)$, а по распределению источника $p(x)$ при заданной статистической связи между множествами X и X' (при заданном распределении $p(X'/X)$.

При квантовании непрерывной величины X указанная ε - энтропия равна энтропии $H(X')$. Поскольку $H(X') \le \log N$, то передача информации в виде чисел, записанных в m-ичной системе счисления и имеющих число разрядов, равное $(\log_m N)$, не будет экономной. Поэтому целесообразно предварительно осуществить экономное кодирование сообщений, в роли которых должны выступать элементы множества X'.

Отметим, что ε - энтропия также используется и в качестве количественной меры производительности источника непрерывных сообщений,

при этом, очевидно, нельзя говорить о производительности источника, не задав точность воспроизведения.

Таким образом, производительность источника можно определить как минимальное количество информации $\nu H_\varepsilon(X)$, необходимое в единицу времени для воспроизведения непрерывного сообщения $x(t)$ с заданной точностью, где ν - число отсчетов сообщения $x(t)$, передаваемых за единицу времени.

Кроме квантования существует много других способов преобразования непрерывной величины X. В частности, множество X' может представлять собой результаты измерений величины X. Пусть непрерывная величина X' с некоторой погрешностью η воспроизводит нормально распределенную случайную величину X: $x = x' + \eta$. Тогда при заданной дисперсии ε^2 погрешности (при заданной среднеквадратичной ошибке) ε - энтропия нормальной величины X равна [8]

$$H_\varepsilon(X) = \frac{1}{2}\log\frac{\sigma_X^2}{\varepsilon^2},$$

где σ_X^2 - дисперсия случайной величины X.

4.8. Применение теории информации при синтезе контролепригодных систем

4.8.1. Обобщённая вероятностно-структурная модель и стратегия определения состояния системы

Введём пространство состояний системы с заданной на нём вероятностной мерой. Если все блоки занумеровать в определённой последовательности от 1 до n и каждому блоку поставить в соответствие 0 или 1 в зависимости от того, исправен или не исправен блок, то получим последовательность из нулей и единиц, которая будет описывать состояние системы. Всего таких состояний (последовательностей) будет 2^n. Всё множество состояний S можно рассматривать как пространство элементарных событий $S_k \in S, (k = \overline{1,2^n})$, каждое из которых может наступить после эксплуатации системы в течение заданного времени с вероятностью, равной

$$p(S_k) = \prod_{j \in J} p_j \prod_{j \in J} (1 - p_j), \qquad (4.13)$$

где p_j - вероятность отказа j-го блока; J - множество номеров исправных блоков, I – множество номеров не исправленных блоков. При этом пред-

полагается, что отказ одного из блоков не влияет на вероятность отказа других блоков.

Возможна организация некоторой совокупности Z точек контроля, допустимое значение сигнала в каждой из которых обеспечивается определённым подмножеством блоков. В качестве исходной диагностической информации используется матрица проверок B_z [2], построенная на допустимом множестве точек контроля, где номер столбца совпадает с номером блока, а номер строки - с номером точки контроля. Проверка сигнала в каждой точке контроля позволяет судить о работоспособности всех блоков соответствующего подмножества, которое определяется совокупностью единиц в соответствующей строке матрицы проверок. Результат проверки принимается равным единице, если контролируемый сигнал вышел из допуска, и нулю в противном случае. Таким образом, при выходе из строя одного из блоков результат проверок совпадает с соответствующим столбцом матрицы. В случае выхода из строя нескольких блоков результаты проверок образуют вектор-столбец $y_i \in Y$, равный логической поэлементной сумме соответствующих столбцов матрицы, где Y - множество всех возможных исходов диагностического эксперимента, под которым будем понимать измерение сигналов в точке контроля.

Таким образом, диагностический эксперимент доставляет некоторый вектор y_i, который характеризует состояние системы с точностью до некоторого подмножества, причём вероятность $p(y_i)$ определяется как сумма вероятностей всех состояний, входящих в соответствующее подмножество.

Разработку стратегии определения состояния системы будем вести с учётом следующих свойств этих подмножеств.

Свойство 1. Подмножества, соответствующие разным векторам y_i, не пересекаются. Действительно, если бы они пересекались, то нашлось бы такое состояние системы, которому соответствовали два различных вектора y_i, чего быть не может.

Свойство 2. Подмножество состояний замкнуто относительно операции сложения состояний. (Под суммой состояний будем понимать состояние, определяемое как многократный дефект, объединяющий дефекты суммируемых состояний).

Свойство 3. Сумма всех состояний подмножества также принадлежит этому подмножеству.

Свойство 4. Существует максимальное подмножество блоков, через отказы которых определяется всё подмножество состояний системы, соответствующее данному y_i.

Свойство 5. Каждому вектору y_i соответствует подмножество подозреваемых в отказе блоков. Подмножества блоков, соответствующие разным векторам y_i, могут пересекаться в отличие от соответствующих подмножеств состояний.

Свойство 6. Максимальная кратность дефекта, который соответствует заданному y_i, определяется количеством блоков в соответствующем подмножестве.

Таким образом, стратегия определения состояния системы сводится к определению подмножества состояний, которое соответствует полученному вектору y_i, а следовательно, и к определению подмножества блоков. Стратегия определения действительно неисправных блоков в подмножестве может быть различной. В частности, можно проверить каждый блок в отдельности, начиная с блоков, дефекты которых определяют наиболее вероятные состояния, к которым чаще всего относятся состояния с однократным дефектом.

4.8.2. Информационная мера глубины диагностирования

В качестве количественной меры глубины поиска дефекта введём следующий коэффициэнт:

$$K = \frac{I(S,Y)}{H(S)}, \tag{4.14}$$

где

$$H(S) = -\sum_{k=1}^{2^n} p(S_k) \log p(S_k) \tag{4.15}$$

неопределённость состояния системы, равная количеству информации, которое необходимо получить, чтобы определить, в каком из состояний находится система;

$$I(S,Y) = H(Y) - H(Y/S) \tag{4.16}$$

количество информации, которое в среднем доставляет результат диагностического эксперимента;

$$H(Y) = -\sum_{y_i \in Y} p(y_i) \log p(y_i) \tag{4.17}$$

неопределённость исхода диагностического эксперимента. Условная неопределённость исхода диагностического эксперимента $H(Y/S)=0$, поскольку неопределённость появления y_i при заданном состоянии систеиы $S_k \in S$ равна нулю. Отсюда

$$I(S,Y)=H(Y) \text{ и } K = \frac{H(Y)}{H(S)}. \tag{4.18}$$

Поскольку состояния отдельных блоков считаются статистически независимыми, то энтропия системы равна сумме энтропий отдельных блоков:

$$H(S) = \sum_{j=1}^{n} H(X_j), \tag{4.19}$$

где X_j - множество состояний j-го блока, состоящее из нуля и единицы. Энтропия j-го блока

$$H(X_j) = -p_j \log p_j - (1-p_j)\log(1-p_j).$$

Для точного определения состояния системы необходимо, чтобы количество информации, которое в среднем доставляет диагностический эксперимент, определялось как $I(S,Y)=H(S)$. В этом случае коэффициент $K=1$. Когда вообще нельзя получить какие-либо сведения о состоянии системы, $I(S,Y)=0$ и коэффициент $K=0$.

4.8.3. Оптимизация глубины диагностирования

Под глубиной диагностирования будем понимать среднее количество состояний, с точностью до которых может быть локализован дефект на основании количества информации $I(S,Y)$, доставляемых диагностическим экспериментом. Очевидно, что глубина диагностирования растёт с увеличением количества информации $I(S,Y)$ или с увеличением коэффициента K.

Задача максимизации глубины диагностирования сводится к выбору из допустимого множества точек контроля такого подмножества заданного размера, при котором диагностический эксперимент доставляет максимальное количество информации о системе.

С целью получения конструктивных решений предлагается следующий эвристический алгоритм последовательного выбора точек контроля.

На первом шаге выбирается точка контроля, которая доставляет максимальное количество информации о системе. На каждом из последующих шагов выбирается та точка контроля, которая доставляет максимальное дополнительное количество информации. Процедура выбора точек контроля заканчивается после того, как будет выбрано заданное число точек контроля.

Аналогично решается задача выбора минимального множества точек контроля, которое обеспечивает заданную глубину диагностирования. В этом случае процедура выбора точек контроля заканчивается, когда количество снимаемой с них информации обеспечивает заданную глубину диагностирования.

Идея алгоритма очень проста, эффективность его реализации зависит от методики расчёта коэффициента K. Непосредственный (или точный) метод расчёта коэффициента K становится трудоёмким для систем большей размерности. Можно предложить, например, не рассматривать те состояния, которые маловероятны и вносят "малый" вклад в соответствующую вероятность $p(y_j)$, т.е. рассматривать лишь наиболее вероятные состояния из подмножеств $S_k \in S, (k = \overline{1,2^n})$. Таким образом, на вычисление коэффициента затрачивается значительно меньше машинного времени при незначительной потере точности вычислений.

5. КЛАССИФИКАЦИЯ МНОГОМЕРНЫХ ДАННЫХ

5.1. Постановка задачи

В природе практически все объекты (физические, биологические, технические, социальные) описываются множеством физических величин или параметров (переменных), которые в совокупности образуют модель объекта как структурированный состав, т.е. систему. Анализ свойств модели на основе системного подхода и экспериментальная проверка результатов анализа являются основной целью современных научных исследований. Объект (образец) обнаруживает себя во вне как совокупность наблюдаемых переменных, значения которых интерпретируются как свойства объекта, его состояния. Кроме непосредственно наблюдаемых переменных существуют скрытые (латентные) переменные, непосредственное наблюдение и измерение которых не представляется возможным, но которые существенно влияют на состояние системы и могут быть использованы в исследованиях как индикаторы её состояния. Например, скрытой переменной является здоровье человека, непосредственное измерение которого не представляется возможным. Тем не менее врачи выносят решение о его состоянии по совокупности результатов анализов и значений других наблюдаемых переменных на основе личного опыта.

Скрытые параметры являются интересующими нас структурными свойствами системы, которые представляют собой различного рода взаимодействия (отношения) компонентов (переменных) системы.

Скрытые переменные (параметры) системы могут иметь иерархическую организацию, в которой должна существовать количественная связь между совокупностью непосредственно измеряемых переменных и переменными более высокого уровня, т. е. такой скрытой переменной, которая нас интересует. Установление этих связей и выбор самих скрытых переменных является основной целью анализа системы, при этом анализ должен быть не формальным, а целесообразным, в контексте решаемой задачи. Наиболее интересным с практической точки зрения является обобщающий скрытый параметр (свойство), который описывает состояние системы как некоторой целостности, например, здоровье человека. В общем виде эту задачу не удается решить, но, тем не менее, здоровье человека можно оценить по некоторому промежуточному скрытому параметру (переменной) в соответствующей иерархии параметров. В качестве другого примера скрытого параметра можно выбрать код, который в форме определенной структуры присутствует в закодированном тексте, но непосредственное наблюдение кода в тексте не представляется возможным.

Таким образом, состояние объекта описывается значениями измеряемых переменных $\xi=(\xi_1, \xi_2, \ldots, \xi_n)$, последовательность которых можно рас-

сматривать как вектор в n-мерном евклидовом пространстве, пространстве переменных. В литературе переменные называют часто компонентами, поскольку в совокупности они образуют систему – модель объекта. Обнаружить структурные закономерности (свойства) системы, т.е. различного рода зависимости между компонентами, по единственному состоянию системы не представляется возможным, поскольку зависимость отражает характер совместного изменения значений переменных. Для решения поставленной задачи необходимо располагать, по крайней мере, множеством состояний, каждое из которых изображается точкой в n-мерном пространстве переменных, в совокупности образуя облако. Все множество многомерных состояний называют многомерными данными. Многомерные данные могут описывать как состояния одного объекта, в котором он может находиться в разные моменты времени, так и состояния разных объектов (образцов), принадлежащих одному и тому же виду, например, множество текстов (образцов), записанных в одном и том же коде (виде).

Состояния здоровья людей, образующих некоторое подмножество (репрезентативную выборку) всех людей, также можно представить в виде многомерных данных, при этом состояние отдельного человека описывается совокупностью значений, выбранных переменных $\xi=(\xi_1, \xi_2, \ldots, \xi_n)$. Переменные, которые описывают состояние объекта, могут быть как детерминированными так и случайными величинами в зависимости от их физической природы и содержания решаемой задачи. В зависимости от этого методы обработки данных можно разделить на детерминированные и статистические. Случайность возникает, когда объект случайным образом выбирается из некоторого множества объектов (образцов) или один объект случайным образом может оказаться в одном из возможных состояний.

Многомерные данные являются априорными данными, на основании которых в результате их соответствующей обработки делается оценка интересующего нас свойства объекта (скрытого параметра, переменной), например, здоров человек или болен.

Может быть неограниченное количество задач, связанных с обработкой многомерных данных, тем не менее, можно выделить наиболее важные задачи:

1. Кластеризация и классификация состояний объекта.

Кластеризация – это разбиение многомерных данных на подмножества (кластеры) состояний (образцов), обладающих некоторым общим свойством. Например, многомерные данные, описывающие состояния здоровья людей, можно разделить на подмножества здоровых и больных людей. Алгоритм кластеризации не требует какой-либо информации, кроме той, которая содержится в структуре многомерных данных, он работает без учителя. Кластеризация основана на сравнении состояний (образцов) между собой на основе выбранной метрики, которая, в сущности, есть

скрытый параметр (ненаблюдаемая переменная), которая вычисляется как функция наблюдаемых переменных. В зависимости от вида выбранной метрики результаты кластеризации могут получиться разными. Метрика выбирается экспериментально.

При классификации разбиение многомерных данных на подмножества задается априорно. Задача классификации заключается в определении, какому из подмножеств принадлежит наблюдаемое состояние (образец).

2. Обнаружение и исследование зависимостей (закономерностей) между переменными, в том числе, и скрытыми, которые позволяют эффективно решить поставленную задачу. В зависимости от содержания задачи модели, используемые при решении задачи и методы ее решения, могут быть как детерминированными, так и статистическими.

Одним из способов выбора, наиболее информативного для решения поставленной задачи, параметра (свойства) является метод проекций или метод главных компонент. Суть его заключается в вычислении проекций векторов, изображающих соответствующие состояния объекта (образцов), на специально выбранное направление в пространстве многомерных переменных. С выбранным направлением будет связана некоторая скрытая переменная, компонента, которая используется в качестве метрики в пространстве переменных.

Рассмотрим пример скрытого параметра (свойства). Пусть некоторый объект описывается вероятностной моделью, параметрами которой являются вероятности $p_1, \ldots, p_n$, а результатами наблюдения статистические оценки их значений $\overset{*}{p}_1 = \xi_1, \ldots, \overset{*}{p}_n = \xi_n$. Априори мы знаем, что сумма значений всех вероятностей равна единице, но непосредственные наблюдения оценок этих вероятностей не доставляют информацию об этом свойстве и поэтому оно является скрытым параметром. Если бы мы этого свойства не знали, то выявить (угадать) его можно, используя метод проекций, при соответствующем выборе направляющего вектора. Если в качестве направляющего вектора выбрать вектор с координатами, равными единице, то скалярное произведение вектора $\xi=(\xi_1, \xi_2, \ldots, \xi_n)$ и направляющего вектора будет равно $\sum_{i=1}^{n} \xi_i$. Эта сумма будет оценкой значения скрытого параметра, равного единице. В общем случае выбор направляющего вектора является отдельной проблемой, если априори мы не располагаем какими-либо сведениями о свойствах объекта кроме значений наблюдаемых параметров.

Рассмотрим пример классификации наблюдаемого состояния биоценоза на основе многомерных данных с вероятностным описанием его состояний.

5.2. Классификация состояний биоценоза[3]

В настоящее время актуальным является внедрение информационных технологий в процесс научных и прикладных исследований состояний биоценоза с целью повышения их эффективности. Биоценоз - это системно-организованная совокупность растений, животных или микроорганизмов, обитающих в определённой среде, на состояние которого можно влиять через среду и тем самым управлять развитием биоценоза или его уничтожением. Частным случаем биоценоза является микробиота желудочно-кишечного тракта (ЖКТ) человека. Это чувствительная индикаторная система, которая своими количественными и качественными изменениями реагирует на любые нарушения состояния здоровья человека. Правильная трактовка результата бактериологического исследования имеет исключительное значение в решении вопроса о природе кишечного заболевания и соответствующих методах лечения.

До настоящего времени в рамках научного исследования для этой цели были разработаны методы, основанные на сравнении результатов бактериологических исследований с эталонной группой здоровых людей, выбранных с помощью экспертных оценок, а также исследованы возможности применения нейросетевых технологий.

Цель данной работы заключается в разработке алгоритма классификации состояний микрофлоры ЖКТ пациента на основе статистического анализа априорных многомерных данных [2,8,9,10], полученных в результате бактериологических исследований качественного и количественного состава микрофлоры ЖКТ больных и здоровых пациентов.

Теоретический анализ

Базовая модель, описывающая состояния ЖКТ, построена на исходных данных, которые представляют собой результаты бактериологических исследований состояния ЖКТ отдельно для больных и здоровых людей. Результат исследования состояния ЖКТ отдельного пациента представлен в виде вектора $\xi=(\xi_1, \xi_2, \ldots, \xi_n)$ в n-мерном евклидовом пространстве, координатами которого являются скалярные величины, каждая из которых равна количеству микроорганизмов данного вида. Таким образом, базовая модель представляет собой n-мерное пространство признаков, которые априорно разделены на два класса, соответствующие здоровым и больным пациентам. Проблема заключается в принятии оптимального решения, какому из классов принадлежат результаты анализа пациента.

В общем случае количество микроорганизмов каждого из видов является случайной величиной с произвольным неизвестным законом рас-

[3] Данный раздел подготовлен совместно с С.А. Зеленцовым

пределения, что делает невозможным непосредственное использование классических методов оптимального статистического синтеза правил принятия решения о состоянии микробиоты. В работе предлагается перейти от случайной векторной величины к ее проекции x на специально выбранное направление, определяемое единичным вектором α. Проекция x оказалась достаточно информативной для решения проблемы классификации микробиоценоза, если вектор α параллелен прямой, проходящей через две точки n-мерного пространства, которые изображают математические ожидания многомерных случайных величин соответственно для «здоровых» и «больных» людей. Проекция x на данное направление вычисляется как скалярное произведение векторов $\xi=(\xi_1, \xi_2, \ldots, \xi_n)$ и $\alpha=(\alpha_1, \alpha_2, \ldots, \alpha_n)$ [3]

$$x = \sum_{i=1}^{n} \alpha_i \xi_i , \qquad (5.1)$$

где $\alpha_1, \alpha_2, \ldots, \alpha_n$ – косинусы углов, которые единичный вектор α образует с осями координат. Поскольку проекция x является весовой суммой случайных величин, то на основании центральной предельной теоремы теории вероятностей можно предположить, что ее закон распределения будет близок к гауссову. В этом случае законы распределения для здоровых и больных людей будут практически гауссовыми, и различаться будут только математическим ожиданием и дисперсией. В работе установлено, что гипотеза о гауссовом законе распределения по критерию согласия χ^2 не противоречит опытным данным с уровнем значимости 0,05 и поэтому ее можно считать правдоподобной.

Гистограммы соответственно для здоровых и больных пациентов представлены на рис. 5.1.

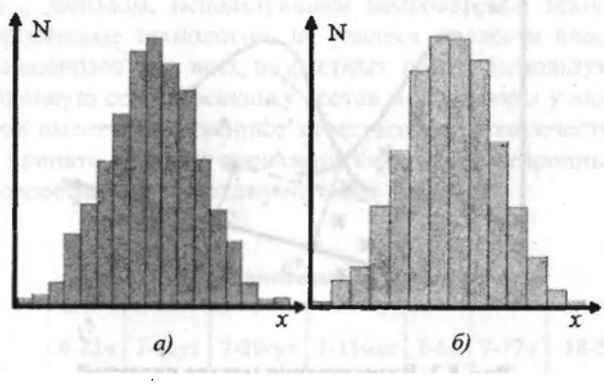


Рис. 5.1. Гистограмма для здоровых (*а*) и больных (*б*) пациентов.
x – значение проекции, *N* – относительная частота значения проекции x

Методика

Построение правила принятия решения о состоянии здоровья пациента начинается с формирования двух классов в пространстве признаков для здоровых и больных пациентов на основании анализа количественного и качественного состава микрофлоры ЖКТ и экспертных оценок. В каждом из классов по экспериментальным данным вычисляются оценки математического ожидания и дисперсии проекции случайного вектора ξ. Для решения поставленной задачи диагностирования используется критерий отношения правдоподобия L, который обеспечивает минимум средней вероятности ошибки. Пороговое значение x_0, которое делит пространство на критическую и допустимую области (рис. 5.2) при равенстве априорных вероятностей того здоров пациент или болен вычисляется при L=1 и $\sigma_1 \neq \sigma_2$ по формуле [2]

$$x_0 = \frac{\sigma_2^2 M_1 - \sigma_1^2 M_2 \pm \sigma_1 \sigma_2 \sqrt{(M_2 - M_1)^2 + (\sigma_2^2 - \sigma_1^2)\ln\frac{\sigma_2^2}{\sigma_1^2}}}{\sigma_2^2 - \sigma_1^2}, \qquad (5.2)$$

где M_1 и M_2 – математические ожидания проекций соответственно для групп здоровых и больных пациентов, σ_1 и σ_2 – среднеквадратические отклонения.

При $\sigma_1=\sigma_2$ пороговое значение вычисляется по формуле $x_0=(M_1+M_2)/2$.

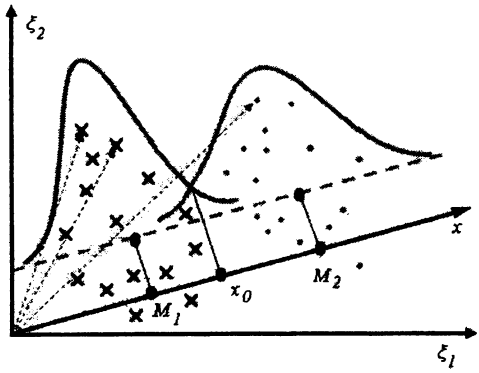


Рис. 5.2. Визуализация метода проекций

Получив результаты бактериологического анализа для диагностируемого пациента в виде многомерного вектора, вычисляется проекция x

этого вектора и, в зависимости от взаимного расположения получившейся проекции и заранее вычисленного значения величины x_0, принимается решение о том, болен пациент с данным результатом анализа или нет. Если значение проекции окажется очень близким к значению порога, то целесообразнее отказаться от указанных решений и отнести диагностируемого пациента к группе риска. С этой целью вводится некоторая окрестность (w_1,w_2) точки x_0. Пациент относится к группе риска, если значение проекции попало в этот интервал (рис. 5.3).

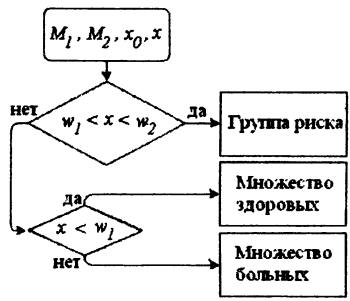


Рис. 5.3. Блок-схема алгоритма классификации

Результаты эксперимента

В качестве экспериментальных данных были отобраны и систематизированы результаты бактериологических исследований микрофлоры ЖКТ 2576 человек с n=29. Выполнен сравнительный анализ предложенного метода с методом, использующим нейросетевые технологии. Используя нейросетевые технологии, не удалось провести классификацию состояний биоценозов для всех возрастных групп, используя одну конкретную нейронную сеть. Поскольку состав микрофлоры у людей различных возрастов имеет существенные качественные и количественные различия, было принято решение производить обучение нейронных сетей для некоторых возрастных групп отдельно (табл. 5.1).

Таблица 5.1

Результаты сравнительных экспериментов

Возраст / Метод	0-23ч	1-6сут	7-29сут	1-11мес	1-6л	7-17л	18-59л	60л и >
Сеть Кохонена	-	-	+	+	+	+	+	+
Сеть Ворда	+	+	+	-	+	+	+	+
Метод проекций	+	+	+	+	+	+	+	+

Плюс в таблице обозначает работоспособность выбранного метода, а минус – невозможность отличить «норму» от «патологии». Лучшие результаты показал предлагаемый в работе метод проекций. Классификация с использованием сети Ворда, которая показала лучшие результаты среди нейронных сетей, обеспечило в среднем 85% правильных решений, а предлагаемый метод проекций - 96%.

Условные вероятности ошибок, правильного принятия решения и отнесение пациента к группе риска представлены в виде вероятностной диаграммы (рис. 5.4).

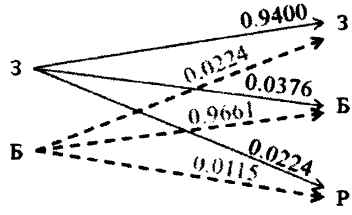


Рис. 5.4. Вероятностная диаграмма, З-здоровые, Б-больные, Р-риск

Таким образом, практические результаты подтвердили эффективность разработанного метода и достоверность теоретических выводов.

СПИСОК РЕКОМЕНДУЕМОЙ ЛИТЕРАТУРЫ

1. Боровков, А.А. Математическая статистика: оценка параметров, проверка гипотез, 2007. - 472 с.
2. Горелик, А.А. Методы распознавания / А.А. Горелик, В.А. Скрипкин — М.: Высш. шк., 2010. – 220 с.
3. Федоткин М.А. Основы прикладной теории вероятностей и статистики / М.А. Федоткин. М. : Высш. шк., 2008. – 368 с.
4. Вентцель Е.С. Теория вероятностей: учебник / Е.С. Вентцель. – 10-е изд.,стер. М. : Высш. шк., 2006.
5. Гмурман, В.Е. Руководство к решению задач по теории вероятностей и математической статистике: учеб. пособие / В.Е. Гмурман. – 9-е изд., стер М. : Высш.шк., 2004.
6. Кудряшов, Б.Д. Теория информации / Б.Д. Кудряшов. – СПб.: Питер, 2009. – 320 с.
7. Гмурман, В.Е. Теория вероятностей и математическая статистика. Базовый курс / В.Е. Гмурман. – М.: Юрайт, 2013. – 480 с.
8. Kim H.Esbensen. Multivariate Date Analysis In Practice. 5 th Edition.1994-2002 CAMO Process AS, Oslo, Norway.
9. Эсбенсен, К. Анализ многомерных данных. Избранные главы: [пер. с англ. С.В. Кучерявского] /; под ред. К. Эсбенсен, О.Е. Родионовой.- Черноголовка: Изд-во ИПХФ РАН, 2005. – 160 с.
10. Большаков, А.А. Методы обработки многомерных данных и временных рядов / А.А. Большаков, Р.Н. Каримов. – М.: Горячая Линия – Телеком, 2007. – 520 с.
11. Барсегян, А.А. Методы и модели анализа данных: OLAP и DATA Mining/А.А. Барсегян, М.С. Куприянов, В.В. Степаненко, И.И. Холод. – СПб.: БХВ-Петербург, 2004. – 336 с.

Ломакин Дмитрий Викторович
Ломакина Любовь Сергеевна
Пожидаева Анастасия Сергеевна

ВЕРОЯТНОСТЬ. ИНФОРМАЦИЯ. КЛАССИФИКАЦИЯ.

Редактор О.В. Пугина
Технический редактор Т.П. Новикова
Компьютерная верстка С.А. Зубкова

Подписано в печать 24.09.2014. Формат $60 \times 84^{1}/_{16}$
Бумага офсетная. Печать офсетная. Усл. печ. л. 8,0.
Тираж 100 экз. Заказ 617.

Нижегородский государственный технический университет им. Р.Е. Алексеева.
Типография НГТУ.
Адрес университета и полиграфического предприятия:
603950, г. Нижний Новгород, ул. Минина, 24

ISBN 978-5-502-00480-0